List of relevant changes

Our answers on all questions, suggestions and remarks can be found on the next pages. Firstly, we summarize the major changes we will make to the revised version of the manuscript based on the comments of the different reviewers:

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

Author response to the review of Anonymous Referee #1

Central Asia is one of the least investigated CORDEX domains and any paper dealing with this area is more than welcome. I do not have an objection to this paper but have a general comment on the approach. Central Asia region has a rough topography and is sparsely populated. Therefore the station data in this region is not as reliable and dense as some other regions. Keeping this in mind, does it really make sense to use CRU data as the basis for model evaluation. In these regions, it may make much more sense to use Era-Interim as the basis and not take CRU (that much) into account.

Thank you for your positive comment. It is indeed true that the gridded datasets are not very reliable in some regions, as we stressed in our paper. We did not take ERA-Interim as reference since this product has some model dependency and might suffer from similar errors that are reproduced by our models which are forced by ERA-Interim for this evaluation study. The station observations also undergo manipulations to obtain a gridded dataset but these steps are not linked with any NWP model. Moreover, the relatively coarse (80km or 0.75 degrees) resolution of ERA-Interim makes it less suitable to serve as a reference for higher-resolution regional climate models due to the larger representativity issues. CRU is also quite coarse (0.5 degrees) but it has still a higher resolution than ERA-Interim.

Author response to the review of Anonymous Referee #2

General Comments

In this paper the authors present the results of an evaluation conducted over the CORDEX Central Asia domain for two different RCMs: REMO and ALARO-0. Comparing climatological seasonal and annual means obtained from two simulations covering the period 1980-2017 against gridded observational data-sets, they aim to assess the reliability of the models for the region of study, setting the basis for their use for future climate projections. The paper complements the results of other studies on RCMs for the same region, and is believed to be interesting for the regional climate modeling community. Nevertheless, in its current form the paper suffers from a series of major issues that need to be carefully addressed before it may be considered for publication for Geoscientific Model Development. In general, the quality of the paper is not very satisfactory. The text and the structure of the manuscript need a thorough revision, since information is many times not very clearly expressed or confusing. The presented analyses are too generic and not at all exhaustive. Explanations for evinced models behavior are often hypothesized without any appropriate investigation. Further, I think that different sources of uncertainty such as the error related to the use of different observational data-sets are not properly considered. I discuss the mentioned issues, together with additional ones, in more details below:

Specific comments

- The presented analyses are neither exhaustive nor accurate enough for a proper evaluation study. In particular, the analyses of the spatial correlation and of the spatial mean calculated over the entire domain are not very useful. First of all, the evinced conclusions for the mean of the spatial biases calculated over the entire domain might simply be the results of some compensating effect and could vary significantly from one area to another. At the same time, given the heterogeneity of the domain of study, spatial means and correlations calculated over the entire domain can hide model limitations specific to single regions characterized by different physical phenomena. Determining and understanding possible model limitations is one of the final goals of models evaluation and serves as the basis for models development. For these reasons, a quantitative analysis of model performances per sub-regions is therefore required.

We agree that there might be some compensating effects due to the spatial means over the large domain. In order to improve our analysis we will add a section evaluating the RCMs over subdomains that are defined by the IPCC6 report (Iturbide et al., 2020) and that are situated in the CAS-CORDEX domain.

- In the text there is lot of confusion between the different sections and their contents, with discussion performed in the results section and some of the results commented in the discussion part. Also, the authors discuss several variables not in the appropriate subsections. One example is the subsection with the discussion on precipitation results, where the results of temperatures are also partly discussed.

To account for this comment we will rearrange the text in the results and discussion section to improve the readability of the text.

- The authors somehow considered the effect of different observations on the comparison of the maps of climatological biases, as well as for the spatial mean calculated over the entire domain. Nevertheless, also the analysis of the spatial correlation, the ratio of standard deviation and the RMSE should take into account the effect of different sources of uncertainties, among which one of the most important is certainly the effect of different observations. In this context, sub-regions analyses assume even more importance. Additionally, other uncertainties could play a big role for the different regions, such as for example the effect of different boundaries. What happens when these sources of uncertainties are considered? The authors should acknowledge the possible effect of different uncertainty sources and all their analyses must at least take into account the effect of the observational uncertainty on the considered metrics.

In order to visualize the uncertainty in data of gridded observational datasets we will add graphs to the manuscript with curves that show the differences between the gridded observational datasets for the annual cycle of the different subregions. The spread of the curves of the different gridded datasets can be considered as a measure of uncertainty.

It is indeed true that the positioning of the boundaries might have an impact on the climate experiments (Rummukainen, 2009), but it is not the aim of this paper to investigate the effect of the domain choice on the resulting RCM data. This work is undertaken within the CORDEX framework which provides guidelines on domains, resolution,... in order to enable RCM intercomparisons between different modelling groups. Therefore, we used the CAS-CORDEX domain as described by the CORDEX project for our model experiments. Although running the same RCMs over different domains would be interesting, it does not fit the aim of our study that frames into the AFTER project. Moreover, it is impossible to realize such an investigation on a short timescale as it would necessitate writing new proposals to obtain computing time on the Tier-1 HPC infrastructure.

- The authors conducted their evaluation considering a single observational data-set for each variable. Then, they basically discussed in each case whether and when the model bias was related to the poor quality of the reference observations, by comparing these with two or three additional data-sets. I am quite critical with the approach they used. In fact, a simple comparison of three or four gridded observational data-sets does not allow to determine the best data for the different regions of the considered domain. For doing this a more robust analysis is needed, considering the initial observational stations of each data-set, their number, their precision and the uncertainty related to the employed interpolation methods. On the other hand, what the authors can do, given the considered data-sets, is to evaluate and take into account the reliability of the given observational data-sets for each point of the domain, by calculating for example the spread of the different observations. Instead of determining whether evinced model biases are due to the reference observational data-set (that in my opinion is not possible to conclude for all the points of the domain, given the available data), the authors should compare models results with the available data-sets, and then discuss whether those biases are within or outside of the range of the observations. In this way they could be able to affirm whether any conclusion on model performances can be drawn for a considered area.

We agree with the reviewer that the spread on the observational datasets is relevant when evaluating the performance of the RCMs. Therefore, we will add maps with the spread between the gridded datasets instead of using Fig. 10 and 11. Additionally, we will add graphs to the manuscript showing the annual cycle of each gridded dataset and each RCM for different subregions. The difference between the curves of the different gridded datasets shows the spread between the different gridded datasets which can be considered as a measure of the observational uncertainty and provides evidence of the performance of the RCMs.

- The authors only investigate climatological values, focusing on the mean bias and on spatial variability. I think that they should be more specific on the choices they made, discussing at least why they focused only on seasonal and annual means and why they did not decide to tackle temporal variability and the seasonal cycle. In particular, I would suggest to add some analysis on the mean seasonal cycle, since the authors claim in their manuscript that some of evinced model biases might likely be related to a wrong simulation of it.

We indeed focused too much on the spatial variability. We agree that the temporal aspects can not be fully understood with the current figures in the manuscript, therefore we opted to add graphs with annual cycles based on monthly data for different subregions.

- The authors focus their analyses on four variables, but then they discuss the results only for temperature and precipitation. Why the discussion is not conducted consistently among the different variables? Additionally, why are the analyses of tmin and tmax not carefully conducted in the same way as for precipitation and temperature, considering different observational data-sets?

We needed the minimum and maximum temperature together to bring the story about the limited diurnal cycle of the RCMs. That is the reason why we decided to split the discussion only in temperature and precipitation which indeed is different to the results section that had four subsections. We will merge the sections of minimum and maximum temperature in the results in a general section about the diurnal temperature range so it is clear that the different variables should be interpreted together to understand the processes later on in the discussion. In the discussion we will use the same structure of subtitles.

The evaluation of Tmin and Tmax is not conducted in the same way since the observational data was not available for all gridded datasets. The Matsuura and Willmott dataset of UDEL does not contain data about the Tmin and Tmax or the diurnal temperature range.

- The effect of the boundaries on the different variables can only be estimated by performing different simulations changing the boundary conditions. The authors should take into account this point whenever they claim that errors in the boundaries are the cause of evinced biases. They can eventually be able to (only partially) support these claims only by performing additional simulations with different boundary conditions.

As mentioned before, it is out of the scope of our research and the manuscript to do an in depth study of the effect of the boundaries due to the aim of the use of the CORDEX domain, the restricted computing time and the goals of the AFTER project which were the driver of these CAS-CORDEX simulations.

- In many cases the presented analyses are very superficial and most of the raised conclusions are mainly hypothesized without a proper demonstration. Additionally, sometimes the authors simply use the maps of the bias to interpret the results of the spatial means. Why should that be interesting and how such an analysis might help in evaluating models results? More in depth analyses are required, including the already mentioned investigation of the seasonal cycle and a quantitative comparison of model results and observations per sub-regions. Every hypothesis on the possible reason of evinced biases should be effectively supported by specific analyses or by bibliographic references.

As mentioned above, we will add a section with subregions that are lying within the CAS-CORDEX domain. We agree that an annual cycle based on monthly data improves the evaluation and insights. We will check which statements are not substantiated enough and we will add evidence that is forthcoming out of the added figures or we will refer to other scientific articles where needed.

- In the manuscript there is a tendency of justifying the bias of the model with the poor quality of CRU. For precipitation, for example, the authors state that CRU underestimates precipitation values: in this case, why do not you perform the comparison against the GPCC as reference then?

GPCC does not contain temperature data. Since it was important for us to refer for each variable to the same reference dataset in order to compare the performance of the different variables, we took CRU as a reference. By adding the analysis of the annual cycle over the subregions it will be easier to compare the RCM outcomes with the different datasets directly.

- I would be more careful stating that evinced results are in the range of the ones obtained for other studies and that indeed the models can be employed for climate projections. You affirm this only considering the reference of Kotlarski for Europe. Additional studies for more regions should be considered. Still, the authors must acknowledge the fact that extremely large biases are present over extensive parts of the domain. For these, either any conclusion on model reliability can be drawn due to high observational uncertainties or model results can not be considered very trustworthy.

We agree, for some parameters significant biases are present over parts of the domain for some seasons. The ALARO-0 RCM has a large positive temperature bias in winter over the northern part of the domain. The REMO model has difficulties in reproducing the observed precipitation patterns over the orography of Central-Asia. We agree that the biases observed in this study should be kept in mind when presenting future projections. We find it therefore important to publish an exhaustive evaluation study. In this evaluation study we saw that the main patterns are modelled correctly and therefore we concluded that we can move on towards climate projections. We will add to our conclusion that these large biases should be kept in mind when looking to the future projections. Additionally, to deal with the biases in impact studies, several bias adjustment methods have been tested within the AFTER project and the most suitable method will be applied before simulations for impact studies are done with these climate data. It is not in the scope of this evaluation study to explain the details about bias adjustments and impact modelling but to avoid misunderstandings we will add that bias adjustment is one of the possibilities when mentioning that the RCMs can be used for future projections.

In other scientific publications where models over the CAS-CORDEX domain were run there are as well large biases over certain parts of the domain (Ozturk et al., 2012; Ozturk et al., 2016; Russo et al., *2019) and even for RCMs run over subregions large biases were found (Wang et al., 2020; Zhu et al., 2020). There are not a lot of scientific articles to compare our results with and to refer to, however in the meantime some new studies are published with model evaluations over a subdomain of our domain and we will refer to them in the updated manuscript. We will thus rewrite the discussion and refer to more scientific articles.*

Minor Comments

- lines 35-37: How large are these ensembles? what about ensembles for the other CORDEX regions, such as for example North America?

These large ensembles consist all out of more than ten GCM-RCM combinations. For example, the ensemble of the EURO-CORDEX domain consists of 14 GCM-RCM combinations; 18 GCM-RCM combinations are available for CORDEX-Africa. North America contains as well a large ensemble of 13 GCM-RCM combinations for the 0.22° resolution but we did not want to list all the different CORDEX regions and the number of GCM-RCM combinations. In our submitted manuscript we mentioned EURO-CORDEX, CORDEX-Africa and MED-CORDEX but NA-CORDEX has indeed more GCM-RCM combinations at the 0.22° resolution. In the revised version we will therefore replace MED-CORDEX by NA-CORDEX. A detailed overview of the available ensembles over the different CORDEX regions can be found at the official CORDEX website: https://cordex.org/ and for each CORDEX domain there is a tab on this website with more information or a link to the website of that particular CORDEX domain.

- line 51: the term "validating" is normally considered not very appropriate when comparing climate models and observations, in particular in cases like this one, where large uncertainties in observations are present. "evaluating" would be a more appropriate term.

As suggested, we have changed "validating" to "evaluating".

- line 56: Same as above. Replace validation with evaluation everywhere in the text.

As suggested, we replaced "validation" with "evaluation" in the text.

- lines 62-63: delete "...that are sparsely populated" since it is repetitive (you already said that in the first line of the period).

As suggested, we removed it.

- line 65: "more extreme values": more extreme than what? just use "extreme values"

We agree and we changed it in the text.

- lines 67-68: The comparison against different observational datasets is useful only to address the reliability of observational datasets and does not help solving the problem of the lack of an ensemble. Please reformulate.

It is reformulated.

- lines 70-71: Similarly as expressed in my major concerns, you cannot directly prove that similar biases in the two models are due to observational errors.
In principle, uncertainty in the observational data-sets allow to say that over certain areas the observations are more or less reliable and whether robust conclusions can be drawn in this case. Please reformulate this part.

It is reformulated.

- line 80: complemented by

It has been corrected.

- lines 83-88: I think that this part would be more appropriate for the introduction rather than for the methods.

We agree, the text has been changed.

- line 106-107: "The outer domain consists of the inner domain plus a coupling zone of eight grid points in each direction.": This holds true for both domains, right? eventually specify.

Indeed, this is true for the domains of both RCMs. We specified this in the text so it is clear that we refer to both RCMs with this sentence.

- Fig.1: Where did the authors take the information on the topography from? the upper limit of the colorbar of 3000m seems not reasonable for the area.

The figure shows the values of the topography used in the regional climate model REMO [GTOPO30 global digital elevation model (DEM) 3 https://www.usgs.gov/centers/eros/science/usgs-eros-archive-digital-elevation-global-30-arc-second-elevation-gtopo30?qt-science_center_objects=0#qt-science_center_objects]. The explanation is added to the figure's caption.

We have increased the upper limit of the colorbar to the upper limit of the orography within the study area.

- line 131: what is the vertical extension in meters of the domain of study for each of the two models?

For REMO, with 27 levels, the top is approximately at 25 km height. The top of the uppermost gridbox is set equal to 0 hPa, but in reality the midpoint of the uppermost gridbox is ~25 km. ALARO-0 uses a vertically staggered grid and the top of the uppermost gridbox is also set equal to 0 hPa. The midpoint of the uppermost gridbox is situated at 67 km for a standard atmosphere.

- lines 139-140: Correct into: "...at the boundaries, up to the 31st of December 2017."

We corrected this sentence.

- lines 143-147: Is there any reason why in the case of ALARO-0 one year can be considered enough for spin-up with respect to the 31 years considered for REMO? Please specify.

Both RCMs are using a different soil model. The soil model used for REMO is using five layers with a mean rooting depth up to 5.7 m (Kotlarski, 2007), while there are only two layers in the ISBA model for ALARO-0. One year spin-up is enough for ALARO-0 since different variables reach their equilibrium after maximum one year. Most soil properties find their equilibrium after about one month. To reach an equilibrium state for the soil temperature and soil moisture, a warm spin-up period of ten years instead of thirty years was used for REMO. We will correct this in the text.

- lines 149-150: This sentence, in the way it is expressed, is not properly correct. In fact, you do a comparison of model results only against the CRU, while then you compare the different observations among them. Please better reformulate this sentence according to the comparison you will decide to perform.

We decided to add annual cycles of the different datasets and RCMs, thus this sentence should not be changed since in those new graphs the results of the RCMs are compared with the different datasets.

- lines 151-152: Again, given your analyses you can not tell whether the bias of the models is due to the observational uncertainty. What you can eventually say is that high uncertainties do not allow to draw robust conclusions.

It is reformulated.

- lines 160-165: I do not manage to find the reference of New et al. 2002 in your paper. Are not there any more up-to-date publications discussing problems of the latest CRU releases? Also, do not you think that the New et al. 1999 publication is more general and it might also apply to other observational datasets rather than simply the CRU? Please consider that all your considered data-sets are somehow characterized by uncertainties (Flaounas et al., 2012;Gómez-Navarro et al., 2012).

We checked and updated our references in this part of the text. Recently a new paper for the CRU data was published (Harris et al., 2020) and we updated our text taking this paper into account. Indeed, New et al. (1999) is rather describing general features about gridded datasets but they do focus on the first versions of CRU, that is why we mentioned this reference as well in the section about CRU. We agree that it is better to refer to more recent and concrete papers for the CRU dataset. Additionally, we will add a sentence in the general part about the reference datasets taking into account Gómez-Navarro et al. (2012). The study of Flaounes et al. (2012) (about the ECA&D gridded dataset over MED-CORDEX) is not general enough to be relevant for our text.

- lines 168-171: what about quality of UDEL for other variables than precipitation?

We added information about the variable temperature.

- lines 176-179: Please, make clear that Hu et al. 2018 only investigated the most central part of your domain of study. Also, the same study states that GPCC underestimates all seasonal means, not only but especially in spring.

Adaptations have been made in the text as suggested.

- lines 181-186: The original resolution of ERAInterim is not 25 Km but approximately 80 km. If you used the data provided by the ECMWF at 25 Km, be aware that these are interpolated data. Please specify this in the text.

As suggested, the explanation has been added to the text and adapted in Table 1.

- lines 181-186: an additional question concerns your choice of using ERAInterim data interpolated at 25 km: why you do not directly download ERAInterim data already onto a 50 Km grid?

We had ERA-Interim available at 25 km on our HPC infrastructure and a projection to the 50 km grid results to the same. The new graphs with annual cycles are not produced at the 0.50° resolution but at the resolution of each dataset and 0.22° for ALARO-0 and REMO.

- lines 186-188: First of all avoid saying initial errors in the boundary conditions, since it generates confusion. Then, how should the comparison of temperature derived from ERAInterim with the one of the models help you determining what is the effect of errors in the boundaries? The only way to assess the effect of the boundaries on the RCM results is to drive the same simulations with different boundaries.

We agree that it is confusing. To be certain about the effect of the errors at the boundaries, other boundaries should indeed be applied. We have deleted this part in the text.

- lines 189-190: The outputs of an RCM are dependent (but not univocally determined) on the values of several variables with which the model is forced at its boundaries. These variables will have an effect on several model variables. The temperature of the model is not only dependent on the values of temperatures provided at the boundaries, but other variables play a role. The same holds true for precipitation. If the model is forced with wrong temperatures it is very likely, at least from a theoretical perspective, that both model temperature and precipitation will be both badly reproduced.

We agree, the text at this line was deleted.

- lines 198-199: First of all UDEL and CRU have the same 0.5 degree resolution. Also GPCC is available at such resolution. Reformulate this period in a more accurate way, considering the fact that the "upscale" is only necessary for the models outputs.

We have changed "upscale" in the text as was suggested. The annual cycle graphs were created using the highest resolution of each dataset (0.50° for CRU and UDEL, 0.25° for GPCC and 0.80° interpolated to 0.25° for ERA-Interim).

- lines 207-209: reformulate this period.

We reformulated this part in the text.

- line 219: seasonal means of

We corrected this.

-line 227: what do you mean by limited bias? better specify.

We reformulated this part in the text.

- lines 228-229: First of all, you start discussing annual means but you put the relative figures at the bottom row of your image: move them up. Then, in my opinion, according to the scale you use in your plots, it seems that in both cases the absolute bias exceeds 3C over a very extensive part of the domain and not only over mountainous regions. Maybe the scale you are using does not help to clearly distinguish which areas are above or below a certain threshold. Try to change your scale.

We agree that the maps of annual means have to be placed at the top of the figure. It is indeed difficult to see the difference between each degree on the figure. We will change the scale.

- lines 229-230: Also the REMO exceeds the 3C range, in particular in winter. Please reconsider your sentence.

We included this REMO temperature bias a bit further in the text where we discuss the biases in the mountainous regions and say that REMO has a warm bias in winter over the Altai region. We agree that this might have been confusing and as suggested, the warm bias of REMO that exceeds 3 °C in winter over the north-western part of Mongolia has been added at this particular location in the text.

- lines 230-231: not totally correct. In fact, the biases ,when considering the entire domain, are particularly pronounced for ALARO-0 mainly in spring, over the northern part of the domain. In winter the most pronounced bias seems to be the one of REMO for north-western Mongolia. For summer and autumn the biases for the two models present a very similar range. The same holds true when considering only the eastern half of the domain. Reformulate this part.

We reformulated this part as suggested.

- lines 231-233: Actually you should really emphasize that the two models seem to have a completely different pattern of the bias of temperature in winter: one shows a bipolar behavior between North and South, while the other between East and West, with a peak in warmer simulated temperatures over north-western Mongolia. I think that it would be really important for the authors (and a very nice opportunity) to better investigate the causes of the two different behaviours. This could give us some clue on model limitations in the simulations of temperatures over the region, that seems to be a general issue for climate models.

As suggested, we will emphasize this different behavior of the models in the text. By including an additional subsection showing the yearly cycle of both temperature and precipitation of the observational datasets and model output over subdomains the reader gets more insight into these bias patterns.

- lines 233-234: What do you want to evince from this? why Scandinavia and not another region? Also, how is the bias similar in the two cases?

We moved this information to the discussion section. The climate in Scandinavia is similar to the climate in the northern part of the CAS-CORDEX domain. The reason why in both regions a warm bias is obtained for ALARO-0, is probably linked with a process that occurs in regions with a subarctic climate and not somewhere else. Deviations in snow related processes might explain the warm winter and cold spring temperature biases in the northern part of the domain and therefore we will add some additional information in the discussion part about this feature. We are currently investigating this.

- lines 238-239: Important biases are present in MAM also for REMO, for some regions such as the Western fringes of the Tibetan Plateau. Also, for both models biases exceed 3C over a large part of the domain in MAM. Reformulate.

As suggested, these sentences have been reformulated.

- lines 239-241: What do you mean by limited? you mean that biases are not very pronounced in summer? reformulate.

We reformulated this sentence as suggested.

- lines 239-241: also for REMO there are warm biases, even though they are inherent to a smaller portion of the domain, in particular with respect to ALARO. Be more precise.

We reformulated this sentence as suggested.

- Fig 2: Beside my previous comment on the figure colorbar, the quality of the image could be further improved by reducing white spaces in between rows and moving the names of the seasons on the left side of the figures. Additionally, units should be added to the colorbars, that should also be moved: the colorbar of the bias should be positioned in between the two columns for the bias of REMO and ALARO.

We decided not to change the location of the names of the seasons in the figure. By placing the names to the left side of the maps the maps would become smaller in order to fit the page. We want to present our figures as large as possible and that is why we structured it in this way. We will add the units to the colorbars and place the colorbar of the bias at the right side of the figure.

- lines 248-249: The mentioned gradient is not very clear, in particular in summer.

We removed this statement.

- line 249: "The outcomes of both RCMs for the mean temperature agree well with the CRU data in autumn (SON)": That is not totally true. In fact, performances of REMO in terms of simulated seasonal climatologies are very similar for autumn, but also for spring and summer.

We reformulated this sentence.

- lines 254-255: what do you mean by "should be placed in perspective"? in which perspective? please reformulate this period.

We reformulated this sentence to make clear that the uncertainty in observational gridded datasets is known to be larger at locations in mountainous areas.

- lines 258-259: "it is clear from Table 2 that the strong cold bias during spring in the north for the ALARO-0 model has a larger negative impact on the spatially averaged bias than the warm bias during winter": I would avoid talking about "negative impact" of the bias over some region on the calculation of the spatial mean bias. Instead, you could say that the spatial bias is largely influenced by the pronounced negative/positive bias over specific regions.

Thank you for the suggestion, we reformulated this sentence.

- lines 264-267: "However, the biases during summer are ... due to the smaller spatial variability in temperature during summer". I think that this period is not very clear and needs to be reformulated, eventually considering additional analyses supporting your conclusions. First of all in summer, in the observations, you have less spatial variability (more accurate than smaller spatial range) than in the other seasons. This is evident from the figure, even though it would be nice if you could support such conclusion with a more quantitative analysis of the CRU spatial variability. Additionally (and most importantly), in your analyses you do not effectively demonstrate that a lower correlation is due to a lower spatial variability in summer. Why can it not be simply due to a worse agreement in the spatial variations between the models and the observational dataset?

We agree that the sentence at line 264 is confusing and does not add any value, therefore we decided to remove this sentence. We have changed "smaller spatial range" into "less spatial variability". We will not include a more quantitative analysis of the CRU spatial variability to keep the document as concise as possible and since it is already visually clear from Fig. 2 that the spatial variability is smaller in summer. From Fig. 2 it is visually clear that the biases are lower in summer compared to winter and autumn, thus we assume that the lower spatial variability in summer is the reason for the lower correlation and not the worse agreement between the models and observational dataset.

- lines 276-277: that is exactly one of the reasons why it would be better to consider the analyses per sub-regions.

We agree and will take into account a subregional analysis.

- lines 290-291: I think that your explanation on the reasons of a more negative bias for TMIN than for T2 is not exhaustive. Additionally, this needs to be moved to the discussion part.

We will move this to the discussion section, where we can explain it exhaustively.

- lines 297-298: "Following the main trend..": confusing, reformulate.

We reformulated this sentence.

- lines 299-301: "The warm minimum temperatures of the RCMs indicate that they underestimate the coldest diurnal temperatures or that the observational CRU dataset overestimates them." There are several issues in this period. First of all you need to reformulate your sentence because it is not the minimum temperature of the model that underestimates observation values but rather the model itself. Also, if the minimum temperatures are warmer than observations, it means that the model overestimates (and not underestimates) the coldest diurnal temperatures. Finally, from the comparison of model results against CRU you can only affirm that the models underestimate minimum diurnal temperatures. You can not prove that the observations overestimate them. The fact that CRU might overestimate them is a possibility, but still is not inherent to the behaviour of the model (nor it is evident from the figure you are commenting).

We reformulated this sentence according to the suggestions.

- lines 312-313: "except for the summer": why except if your are talking about annual values?

We agree and reformulated this part of the text.

- line 315: less good than what?

We reformulated this part of the text and moved it to the discussion section.

- line 323: you do not need to specify that temperature is a variable here

We agree and, according to the remarks that were made for minimum temperature, we moved this sentence to the discussion section.

- lines 323-324: You need to reformulate this sentence. In this case you have to specify that the negative TMAX bias is particularly remarkable in spring for the northern part of the domain, and also, to a less degree, in summer. In winter some other less extended parts of the domain, such as the north-eastern part, show a colder bias than REMO. In Autumn results are more similar between the two models.

We agree and we reformulated this text part.

- end of line 326: the cold bias

We corrected the typo.

- lines 326-328: Fig. 4 shows minimum temperatures. Then, how can we deduce from this figure that the bias in TMIN is due to maximum temperatures? please better explain and eventually reformulate this period.

We referred to the wrong figure, it should be Fig. 6. This sentence is describing what was earlier mentioned: "specify that the negative TMAX bias is particularly remarkable in spring for the northern part of the domain". We moved the sentence up and rewrote it a bit so it is clear what we are trying to say.

- line 342: "This means that ALARO-0 fails to reproduce the low nocturnal temperatures": This belongs to the discussion on minimum temperatures. Additionally, the model still fails in simulating warmer temperatures, despite the smaller bias when compared to TMIN.

As suggested, we moved the last paragraph of this section to the discussion section and we explain more clearly that ALARO-0 fails to reproduce temperature in general (including mean, minimum and maximum temperature) in the northern part of the domain.

-lines 344-346: You should discuss about minimum temperature in the appropriate section.

As suggested, we moved the last paragraph of this section to the discussion section.

- When you comment the maps of the bias, try to discuss the different seasons from up to down, consistently with the figures.

We agree.

- lines 364-366: This part should be moved to the discussion section.

As suggested, we moved this text to the discussion section.

- lines 365-366: why however? also, you did not discuss until this point the uncertainty of CRU: how can you claim that the reason for the wet bias is due to the observations?.

We agree that "However" at the beginning of this sentence is not suited here. It was not our intention that this sentence was interpreted as a shortcoming of the CRU dataset since this is the results section. We wanted to express that it is known from the observations that the amount of precipitation is low in certain regions as seen in Fig. 8 (< 5 mm/month), not that CRU contains precipitation amounts that are too low (this follows in the discussion). We reformulated this sentence, to overcome the confusion.

- lines 370-372: By whom is the bias turned into something else in summer? and how?

We reformulated this sentence to make it clear that we talk about summer, when the East Asian Monsoon takes place.

- lines 372-376: It would be nice if you could perform the analyses of the seasonal cycle to support your conclusions. This would make your evaluation more complete and exhaustive, while at the same time allowing to effectively confirm or deny your conclusions.

Thanks for this suggestion. We agree and did an analysis of the annual cycle over multiple subdomains.

- lines 390-391: "The dry biases for ALARO-0 in Table 5 are thus caused by the simulation of systematically less precipitation than the precipitation amounts in the CRU data.": Reformulate. It is obvious that if the model underestimates precipitation, it simulates less precipitation than observations.

We reformulated this sentence. We intended to say that there is no region that has a strong dry bias which is compensated with a wet bias in another subregion. This differs from the finding of temperature where the strong warm bias in the north is partly compensated by a cold bias in the southern part of the domain.

- line 393: systematically

We corrected the typo.

- lines 392-394: "The lower accuracy of simulated precipitation is due to the fact that precipitation is less systematic affected by land cover and topography compared to temperature": First of all that is quite a strong assumption given the extent and heterogeneity of the domain you are considering. Additionally, you did not perform (at least it is not reported in the paper) any analysis that supports your conclusion.

We agree, we did not perform an analysis on this topic but it is known that it is harder to simulate the spatial pattern of precipitation compared to temperature (Kotlarski et al., 2014) due to the reason we mentioned.

- lines 400-404: This is incorrect. In fact Russo et al. 2019 showed that uncertainty in observations is high over the north-eastern part of the domain, not that CRU overestimates the diurnal temperature range over the region.

We agree and will reformulate this text part.

- line 404: Why hence?

We agree that this is an incorrect cause-consequence structure and we will reformulate it.

- lines 404-406: Again, how can you surely state that the model underestimates values of the diurnal temperature range due to higher observation values?

We will rewrite this part.

- lines 407-408: why Czech Republic? what happens in other regions?

We agree that it would be better to refer to literature over Central Asia instead of referring to literature over EURO-CORDEX where ALARO-0 and REMO were already evaluated. We will refer to Russo et al. (2019) who obtained similar findings.

- lines 420-422: This is just an assumption that needs to be proven. Models develop their answer that is, to a certain degree, independent from the boundaries. To test your hypothesis, one easy experiment that could be conducted is to use different boundaries and compare the results.

We agree and we will remove this.

- line 425: "They related this warm bias already to shortcomings in the simulation of snow": this means that they explained the bias differently than with the boundary effect as you explained in the lines from 423 to 425.

Ozturk et al. (2012) explained the bias indeed with a shortcoming in the simulation of snow cover. We will remove the part about the boundary effect.

- lines 430-431: "Hence, we conclude that the warm forcing is the main reason for the warm bias over Eastern Russia during winter.": I further have to highlight that you can not make such conclusion, until you do not test different boundaries.

We agree and we will remove the part about the boundary effect.

- lines 435-436: As before, it would be nice if you could do the analyses of the seasonal cycle since you mention it for the interpretation of your results.

We agree and we will add annual cycle graphs as mentioned before.

- lines 440-442: How the fact that for Belgium there is some correlation between warm bias and cloud cover representation could explain the same for northeastern CAS. You could do some analyses on cloud cover to support your conclusion.

We mentioned this study over Belgium since it is the only study that investigated the relationship between temperature and cloud cover for ALARO. We agree that we cannot draw strong conclusions from this and that this previous paper only gives a clue that cloud cover might be one of the reasons why the temperature is not well estimated. Cloud cover is thus only one out of the many possible reasons, which should be further investigated. In the meantime we did some analysis on cloud cover and we will include our findings to the new version of the manuscript.

- lines 443-445: "Both could be due to too much cloud cover": according to whom? In theory it could be due to any reason.

We agree and we investigated this further to say something about it in the discussion.

- lines 448-451: These considerations are important: it would be nice to put them in a more objective context. Additionally, you say that New et al. show that CRU underestimates temperatures for Russia. Then you talk about Western Russia. If you state that temperatures from CRU are not good for Russia, then they can not be good for a part of it and bad for the rest. Reformulate.

We reformultated these sentences based on the additional analysis over the subregions.

- Fig. 10,11: To make the discussion easier I would suggest to plot the maps of the differences between different observational data-sets together, using the spread of the observations among the different data-sets. In this way you can easily know which areas are more reliable and which are not.

We agree and produced new figures.

- lines 465-467: the less reliable observations do not explain the bias, rather they do not allow to draw any conclusion.

We reformultated these sentences.

- line 471: "...winter and overestimate it during." During what?

The word "summer" is missing, we added it to the text.

- lines 482-484: what happens when you compare ALARO-0 with the other data-sets over the entire domain?

The precipitation of ALARO-0 is for most grid points within the range of the different gridded datasets during the different seasons. When averaging over the complete domain, then the output of both RCMs is within the range of the spread between the reference datasets for the different seasons. However, there are some subregions where the precipitation of ALARO-0 and/or REMO is lower or higher than the observational spread for a specific season. For example both RCMs slightly underestimate precipitation in summer over West Central Asia. We will add this information in the updated manuscript.

- lines 484-486: you mention two gridded data-sets: to which data-sets are you referring here? please better specify.

We are referring to GPCC and MW, these are observational gridded datasets. ERA-Interim is a reanalysis product, so we do not refer to it as an observational gridded dataset. We reformultated this sentence and we will make sure that this is clear throughout the complete manuscript.

- lines 486-489: again, you can claim that the bias is relative to the employed boundaries only performing a new simulation with different boundaries. Also, how can you be sure that ERA-Interim overestimates specific humidity?

We will reformulate these sentences since it was not intended to say that the boundary conditions of ERA-Interim affected our results. We just wanted to point to the similarities between the ERA-Interim data and the output of the RCMs. We agree to remove the suggestion of the overestimation of the specific humidity as we did not investigate it.

- lines 489-490: You are claiming this from the field means I guess. I think that plotting the maps of the bias of the models against all the different observational data-sets might help the discussion of your results.

We claim this based on Fig. 8, 11, S1 and S2 where the spatial patterns between ERA-Interim and REMO are visually very similar, while the patterns of ALARO-0 are similar to GPCC and MW. We agree that it can help to plot the maps of the bias of the models against the different observational datasets, however this will make the manuscript long.

- lines 489-490: How do ERA and REMO parameterize precipitation since you mention that they do it differently? Specify.

For REMO these specifications are included in Table S1. We will refer to this table at the end of this sentence and we will add that ERA-Interim uses a convection scheme modified from Tiedtke (1989) by Bechtold (2008; 2014) and the cloud scheme is based on Tiedtke (1993) with modifications made made by Forbes and Tompkins (2011), Forbes et al. (2011) and Tompkins et al. (2007) (https://www.ecmwf.int/en/research/modelling-and-prediction/atmospheric-physics). The similarities between ERA-Interim and REMO for precipitation are thus probably due to the fact that both use a modified scheme that is based on Tiedtke (1989). We did not further investigate this.

- lines 492-493: "This difference between ALARO-0 and REMO is related to the 3MT cloud microphysics scheme of ALARO-0": where did you demonstrate this?

We did not demonstrate or investigate this but it is an assumption since this is known to cause differences (Giot et al., 2016). We reformulated this statement so it is clear that it is an assumption that should be further investigated in the future.

- lines 496-497: Again, this is hard to affirm simply using three observational data-sets. The authors have to acknowledge the low number of observational data-sets. As I mentioned in many previous comments I personally think that it would be better to approach the differences between the observational datasets in terms of reliability rather than determining who is more correct.

We agree, we will mention the low number of observational datasets. We will reformulate the text so we focus on the reliability of the observations and the complications for our evaluation. It was not our intention that it looks like it is a research on which reference dataset is the best one.

- Fig. 11: In the colorbar of the bias, are units percentage? with respect to what? Please specify.

Indeed, the unit was wrong and we corrected it to the unit percentage. In Fig. 11 the precipitation of CRU is compared with the other datasets ERA-Interim, MW and GPCC. The relative values were obtained by dividing the difference by the value of CRU as already mentioned in section 2.4 Analysis methods. In order to make this clear in Fig. 11, we will specify this in the figure captation.

- lines 501-502: Not completely true. Specify that, as evinced from your maps, the wet bias in ERA (with respect to the other 3 data-sets) is only relative to the eastern part of the domain.

We reformulated this sentence as suggested.

- line 520: "that that". Correct.

We corrected this.

- lines 536-537: "REMO simulates the precipitation fairly well and ALARO-0 performs very well." How can you state that their performances are good?

The simulated precipitation of the RCMs is for most regions most of the time within the observational spread. We have clarified this in the manuscript.

- lines 539-540: "The warm temperatures obtained with REMO ... can be linked with the dry and wet bias in winter and spring respectively." Why and how can they be linked?

We agree that they cannot be linked without doing an in depth study on how they are exactly linked. We reformulated this sentence.

- lines 540-541: In which way the link between temperatures and precipitation should strengthen your hypothesis of a delay by REMO in the simulation of snow cover? Can you be more specific?

This was an assumption, we reformulated this sentence.

- lines 545-546: "The persistent warm bias over Pakistan and Northern India of both RCMs can be explained by the persistent underestimation in simulated precipitation over this region by both RCMs.": how can you state that given your analyses?

We agree that the warm temperature bias cannot be explained by an underestimation in precipitation. We reformulated this sentence.

- lines 547-550: You refer to the fact that your results are within the ranges of models for other domains, but then you only mention the results of Kotlarski et al. for Europe. You need more references.

We agree, we will add some papers that were recently published over parts of the CAS-CORDEX region.

- lines 562-563: That is arguable, given your analyses. How do you define an acceptable range?

An acceptable range is within the range of the observational spread. We will reformulate this so it is clearer.

- lines 565-567: You cannot state this, until you force the model with different boundaries and you conduct an analysis of snow cover (what you can eventually do for snow cover is to reference to the evidences from other studies).

We agree and we removed this sentence.

- Table 1: This table is not easily readable. Could you find a way to make the distinction between the different data-sets a bit clearer?

We will add a light gray background to the odd rows, so that the distinction between the information of the different rows is more clear.

Author response to the review of Anonymous Referee #3

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

<u>Major points</u>

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

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

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

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

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

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

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

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

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

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

<u>Minor points</u>

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

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

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

This would make the paper too long.

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

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

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

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

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

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

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

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

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

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

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

To overcome confusion we will use “relaxation zone”.

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

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

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

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

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

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

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

We agree, we have added the annual cycles.

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

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

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

We agree and we will add additional information.

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

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

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

Indeed, we corrected this.

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

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

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

We agree.

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

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

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

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

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

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

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

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

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

Indeed, we corrected this.

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

Evaluation of regional climate models ALARO-0 and REMO2015 at 0.22° resolution over the CORDEX Central Asia domain

Sara Top[1,2], Lola Kotova[3], Lesley De Cruz[4], Svetlana Aniskevich[5], Leonid Bobylev[6], Rozemien De Troch[4], Natalia Gnatiuk[6], Anne Gobin[7,8], Rafiq Hamdi[4], Arne Kriegsmann[3], Armelle Reca Remedio[3], Abdulla Sakalli[9], Hans Van De Vyver[4], Bert Van Schaeybroeck[4], Viesturs Zandersons[5], Philippe De Maeyer[1], Piet Termonia[2,4], Steven Caluwaerts[2]

[1]Department of Geography, Ghent University (UGent), Ghent, 9000, Belgium
[2]Department of Physics and Astronomy, Ghent University (UGent), Ghent, 9000, Belgium
[3]Climate Service Center Germany (GERICS), Helmholtz Zentrum Geesthacht, Hamburg, 20095, Germany
[4]Royal Meteorological Institute of Belgium (RMIB), Brussels, 1180, Belgium
[5]Latvian Environment, Geology and Meteorology Centre (LEGMC), Riga, LV - 1019, Latvia
[6]Nansen International Environmental and Remote Sensing Centre (NIERSC), St. Petersburg, 199034, Russia
[7]Flemish Institute for Technological Research (VITO), Mol, 2400, Belgium
[8]Department of Earth and Environmental Sciences, Faculty of BioScience Engineering, Heverlee, 3001, Belgium
[9]Iskenderun Technical University, Iskenderun, 31200, Turkey

Correspondence to: Sara Top (sara.top@ugent.be)

Abstract. To allow for climate impact studies on human and natural systems high-resolution climate information is needed. Over some parts of the world plenty of regional climate simulations have been carried out, while in other regions hardly any high-resolution climate information is available. This publication aims at addressing one of these regional gaps by presenting an evaluation study for two regional climate models (RCMs) (REMO and ALARO-0) at a horizontal resolution of 0.22° (25 km) over Central Asia. The output of the ERA-Interim driven RCMs is compared with different observational datasets over the 1980-2017 period. The ~~choice of~~ spread between the observational datasets has an impact on the scores but in general one can conclude that both models reproduce reasonably well the spatial patterns for temperature and precipitation. The ~~evaluation of minimum and maximum temperature demonstrates that both models underestimate the daily temperature range. More detailed studies of the annual cycle over subregions should be carried out to reveal whether this is due to an incorrect simulation in cloud cover, atmospheric circulation or heat and moisture fluxes. In general, the~~ REMO model scores better for temperature whereas the ALARO-0 model prevails for precipitation. Studying annual cycles over specific subregions enables to get deeper insight into the strengths and weaknesses of both RCMs over the CAS-CORDEX domain. The evaluation of minimum and maximum temperature demonstrates that both models underestimate the daily temperature range. This publication demonstrates that the REMO and ALARO-0 RCMs can be used to perform climate projections over Central Asia and that the ~~produced~~ climate data can be ~~applied~~ used for in impact studies taking into account a bias correction for those regions where significant biases have been identified~~modelling~~.

1 Introduction

There is a strong need for climate information at the regional-to-local scale that is useful and usable ~~to allow~~ for impact studies on human and natural systems (Giorgi et al., 2009). In order to accommodate for this, the World Climate Research Program (WCRP) Coordinated Regional Climate Downscaling Experiment (CORDEX) was initiated with the aim to design and gather several high-resolution experiments over prescribed spatial domains across the globe. CORDEX creates a framework to perform both dynamical and statistical downscaling, to evaluate these regional climate downscaling techniques and to characterize uncertainties of regional climate change projections by producing ensemble projections (Giorgi and Gutowski, 2015). Within CORDEX there are large ensembles of model simulations available at different resolutions for the Africa (Nikulin et al., 2012; Nikulin et al., 2018), Europe (Jacob et al., 2014; Kotlarski et al., 2014), ~~and the~~ Mediterranean (Ruti et al., 2016~~5~~) and North America (Diaconescu et al., 2016; Whan and Zwiers, 2017; Gibson, 2019) CORDEX regions (Gutowski et al., 2016). These large ensembles consist of more than ten different GCM-RCM combinations. In order to provide such ensembles over all CORDEX regions, coordinated sets of experiments were recently performed or are still ongoing for CORDEX regions such as South America (Solman et al., 2013), Central America (Fuentes-Franco et al., 2015; Cabos et al., 2019), South Asia (Ghimire et al., 2018), East Asia (Zou et al., 2016), South-East Asia (Tangang et al., 2018; Tangang et al., 2019; Tuyet et al., 2019), Australasia (Di Virgilio et al., 2019), Arctic (Koenigk et al., 2015; Akperov et al., 2018), Antarctic (Souverijns et al., 2019) and Middle East North Africa (Almazroui et al., 2016; Bucchignani et al., 2018). In addition, a new ensemble of ~~climate and~~ climate change simulations covering all major inhabited regions with a spatial resolution of about 25 km~~, within the WCRP CORDEX COmmon Regional Experiment (CORE) Framework,~~ has been established within the WCRP CORDEX COmmon Regional Experiment (CORE) Framework ~~in~~to support ~~of~~ the growing demands for climate services (Remedio et al., 2019). Furthemore, a number of high-resolution global simulations at climatic timescales, with resolutions of at least 50 km in the atmosphere and 28 km in the ocean, have been performed within the Coupled Model Intercomparison Project 6 (CMIP6) (Haarsma et al., 2016).

While high-resolution ensembles (up to 12.5 km spatial resolution) are available for certain regions, e.g. EURO-CORDEX (Jacob et al., 2014), for other regions such as Australasia (Di Virgilio et al., 2019) and the Antarctic (Souverijns et al., 2019) the first experiments were performed only recently. For the CORDEX Central Asia (CAS-CORDEX) domain only a single climate run with the regional climate model (RCM) HadRM3P (Gordon et al., 2000) of the Met Office Hadley Centre (MOHC) at a resolution of 0.44° was publicly available through the Earth System Grid Federation (ESGF) archive until 2019. In addition, climate projections with the RegCM model at 0.44° resolution for the 2071-2100 period and different emission scenarios were reported in Ozturk et al. (2012, 2016), however they are not available through the ESGF archive. Moreover, this resolution is insufficient for impact modelling and environmental assessment applications and thus higher-resolution climate data over the CAS-CORDEX region is needed (Kotova et al., 2018). Recently, Russo et al. (2019) presented model evaluation results of the COSMO-CLM 5.0 model ru~~a~~n at 0.22° or 25 km resolution over the CAS-CORDEX region. The current study significantly extends our knowledge ~~over~~ of the CAS-CORDEX domain by evaluating ~~validating~~ two different RCMs based on multiple scores for temperature (mean, minimum and maximum) and precipitation over a much longer period. In order to fill the knowledge gap over Central Asia two RCMs, ALARO-0 and REMO, were run over this region at 0.22° resolution in line with the CORDEX-CORE protocol (CORDEX Scientific Advisory Team, consulted on 01/03/2019). Here we present the model evaluation through the use of so-called “perfect boundary conditions” taken from the reanalysis data and by comparing the downscaled results to observed data for the period 1980-2017. Such a ~~validation~~ study is necessary ~~in order~~ to gain confidence in the RCM downscaling procedure before its application in the context of climate projections where the RCM is driven by a GCM (Giorgi and Mearns, 1999). The methodology for evaluation ~~validation~~ is partially based on Kotlarski et al. (2014) and Giot et al. (2016), that compared a large ensemble of RCMs over the EURO-CORDEX region with the high-resolution E-OBS observational dataset (Hofstra et al., 2009). However, in this study a slightly different approach is necessary due to ~~1)~~ the absence of an ensemble of RCM runs over Central Asia~~, and 2) the absence of a reliable observational dataset over this region~~. ~~While the Central Asian region is a vast area, the network of measurement stations is unevenly and sparsely distributed, especially the latter is problematic for several large subregions within the domain, that are sparsely populated.~~ ~~Therefore~~Additionally, in some regions the quality of gridded observational datasets, constructed through interpolation or area-averaging of station observations~~.~~, is poor due to ~~suffers, in some regions, from the small number of stations that leads to~~ over-smoothing ~~especially~~ of ~~more~~ extreme values (Hofstra et al., 2010)~~.~~ and/or because of station observations that are nonrepresentative for their large-scale environments. This is particularly the case for ~~Additionally, the measurements at existing stations may not be representative for their large-scale environments, in particular in~~ orographically complex regions such as the Himalayas. The current~~In order to account for the lack of a model ensemble and reliable observations, this~~ study compares the model simulations with different gridded observational datasets and reanalysis data. When the different datasets show large deviations and a large spread, then their uncertainty is high and no robust conclusions can be drawn (Collins et al., 2013; Russo et al., 2019).~~The model biases are compared with the differences among the observational datasets where the latter could be seen as estimates of the observational uncertainty (New et al., 1999). For instance, spatially-similar bias patterns among the two models could be caused by observational errors that might be revealed by large differences between the observational datasets.~~

This study contains two assets: for the first time an in-depth evaluation of the RCMs ALARO-0 and REMO~~, ran at 0.22° resolution,~~ is performed at 0.22° spatial resolution over the CAS-CORDEX domain and ~~in addition~~ we reflect on the impact of ~~choice of~~ the observational datasets on the model evaluation~~validation~~. Such an analysis is a prerequisite in order to be able to use the climate data in a sound way for later impact studies, e.g. for investigating climate change impacts on crop yields and biomass production in forest ecosystems, which will be done in the framework of the AFTER project (Kotova et al., 2018).

In the following section we describe the applied methodology for this study (Sect. 2). This section contains details about the study area, the model description, datasets used for the evaluation and the methodology of the analysis. In Sect. 3, we describe the annual cycle, seasonal and annual means, biases and variability of mean, minimum and maximum surface air temperature and precipitation. Further, we evaluate and provide a discussion of some remarkable anomalies in Sect. 4~~, complemented a brief outlook of the future plans of the ALARO-0 and REMO simulations~~. In the final Sect. 5 we summarize the conclusions.

2 Methods

~~Coordinated sets of experiments were recently performed or are still ongoing for CORDEX regions such as South America (Solman et al., 2013), Central America (Fuentes-Franco et al., 2015), North America (Wang and Kotamarthi, 2015; Diaconescu et al., 2016; Whan and Zwiers, 2017), South Asia (Ghimire et al., 2018), East Asia (Zou et al., 2016), Southeast Asia (Tangang et al., 2018; Tangang et al., 2019; Tuyet et al., 2019), Australasia (Di Virgilio et al., 2019), Arctic (Koenigk et al., 2015), Antarctic (Souverijns et al., 2019) and Middle East North Africa (Almazroui et al., 2016; Bucchignani et al., 2018). In this paper we discuss the results of the simulations over the CAS-CORDEX domain.~~

2.1 CORDEX Central Asia domain and subdomains

The CAS-CORDEX domain as shown in Fig. 1 contains Eastern Europe, a large part of the Middle East (including: Saudi-Arabia, Jordania, Syria, Iraq, Iran) and Central Asia (including: Kazakhstan, Uzbekistan, Turkmenistan, Afghanistan, Pakistan, Tajikistan, Kyrgyzstan and Mongolia). The majority of Russia and China (excluding the most eastern provinces) and the northern part of India are included as well. This domain is an exceptional CORDEX domain in the sense that it barely covers any open ocean~~ or sea~~. It contains several important mountain ranges e.g. Ural, Caucasus, Altay and Himalaya, and deserts e.g. Arabian, Karakum, Thar, Taklamakan and Gobi desert. Mountainous environments are of special interest for regional climate modelling since global climate models do not resolve the mountain ranges with a spatial resolution less than 50 km and hence RCMs may have an added value here (Torma et al., 2015). In addition, the CAS-CORDEX domain contains a wide range of climatic and bioclimatic zones, with in the north permafrost and snow-driven processes and in the south extremely hot regions (e.g. Arabian Peninsula) and monsoon-driven climates with excessive convection linked to the Inter-Tropical Convergence Zone (ITCZ) passing.

In order to obtain simulations that are comparable, the CORDEX initiative prescribes the minimum inner domain of each CORDEX region that the RCM has to cover. While REMO uses the exact rotated lat-lon CAS-CORDEX grid (Jacob et al., 2007) described by the CORDEX community, ALARO-0 has adopted a conformal Lambert projection (Giot et al., 2016), which implies that the non-rotated boundary box should be applied in order to define the domain. The grids were set up in such a way that the CAS-CORDEX domain is completely covered by the non-coupling zone. The CAS-CORDEX 0.22° ALARO-0 inner domain encompasses 333 and 223 grid boxes, while REMO circumscribes 309 and 201 grid boxes in the east-west direction and north-south direction, respectively. The outer domain for both RCMs consists of the inner domain plus a ~~coupling~~ relaxation zone of eight grid points ~~in each direction~~at every boundary.

The CAS-CORDEX domain overlaps with eight other CORDEX domains, including the ones covering Europe, the Arctic, East Asia, South East Asia, South Asia, Africa/MENA and the Mediterranean. Both RCMs used in this study, ALARO-0 and

REMO, were already run and evaluated ~~validated~~ over the EURO-CORDEX region (Kotlarski et al., 2014; Giot et al., 2016) and additionally, REMO has been validated over five other overlapping CORDEX regions (Remedio et al., 2019).

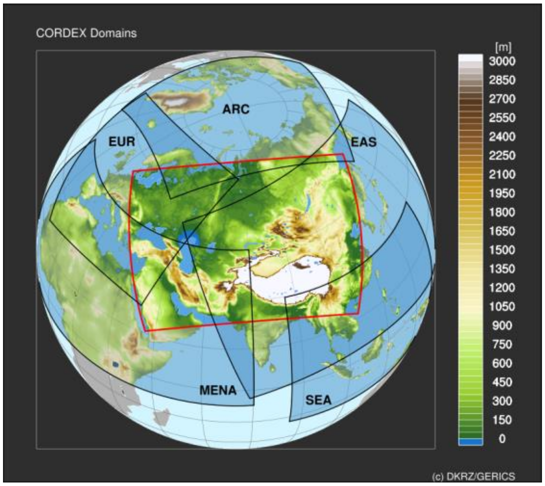


Figure 1~~1~~: The CAS-CORDEX domain demarcated by a red contour and main overlapping CORDEX domains (black contour lines): Europe (EUR), Arctic (ARC), South East Asia (SEA), East Asia (EAS) and MENA projected upon the topography of Eurasia(geopotential height [m] of the GTOPO30 global digital elevation model (DEM) 3).

Formatted: Font: 9 pt, Bold

Formatted: Font: 9 pt, Bold

Formatted: Font: 9 pt, Bold

The CAS-CORDEX domain was further subdivided into five subdomains according to the IPCC reference regions (Iturbide et al., 2020) named as: East Europe, West Siberia, East Siberia, West Central Asia and Tibetan Plateau. These subdomains, visualized in Fig. S1 of the supplementary material, were applied to evaluate the spatial differences in the study area and to investigate whether there were differences in the simulation of subcontinental processes.

2.2 Model description and experimental design

REMO and ALARO-0 are hydrostatic atmospheric circulation models aimed to run over limited areas. The ALARO-0 model is a configuration of the ALADIN model (ALADIN international team, 1997; Termonia et al. 2018a) which is developed, maintained and used operationally by the 16 countries of the ALADIN consortium. The dynamical core of the ALADIN model is based on a spectral spatial discretization and a semi-implicit semi-Lagrangian time stepping algorithm. The ALARO-0 configuration is based on the physics parameterization scheme 3MT (Modular Multiscale Microphysics and Transport (Gerard et al. 2009)), which handles convection, turbulence and microphysics. ALARO-0 has been used and validated for regional climate studies (Hamdi et al., 2012; De Troch et al., 2013; Giot et al., 2016; Termonia et al. 2018b).

The REMO model is based on the Europa Model, the former NWP model of the German Weather Service (Jacob, 2001). The model development was initiated by the Max-Planck-Institute for Meteorology and is further maintained and extended by the Climate Service Center Germany ~~German Institute for Climate Services~~ (HZG-GERICS). The physical parameterization originates from the global circulation model ECHAM4 (Roeckner et al., 1996), but there have been many further developments (Hagemann, 2002; Semmler et al., 2004; Pfeifer, 2006; Pietikäinen et al., 2012; Wilhelm et al., 2014). REMO is used in its most recent hydrostatic version, REMO 2015, and the dynamical core has a leap-frog time stepping with semi-implicit correction and Asselin-filter. For both RCMs, the vertical levels are based on hybrid normalized pressure coordinates which follow the orography at the lowest levels. For the ALARO-0 experiment 46 levels were used whereas the REMO run employs 27 levels. More details on the general setup of ALARO-0 can be found in Giot et al. (2016) and for REMO we refer to Jacob et al. (2001) and Jacob et al. (2012). An overview of the model specifications is given in Table S1 of the supplementary material.

In order to evaluate ~~validate~~ both RCMs, a ~~validation~~ run driven by a large-scale forcing taken from the ERA-Interim global reanalysis (Dee et al., 2011) is undertaken for the period 1980-2017. A one-way nesting strategy is applied to dynamically downscale the ERA-Interim data, having a horizontal resolution of about 0.70° (approximately 79~~80~~ km), to a higher ~~-~~resolution over the CAS-CORDEX domain (Denis et al., 2002). The ERA-Interim forcing data is prescribed at the lateral boundaries using the Davies (1976) relaxation scheme and the downscaling is performed to a horizontal resolution of 0.22° (approximately 25 km). Both model experiments are continuous runs initialised on the 1st of January 1979 and then forced every 6 hours at the boundaries up to December 31st 2017. Following the methodology of Giot et al. (2016), constant climatological fields for some parameters are used and updated monthly. These include sea surface temperatures (SSTs), surface roughness length, surface albedo, surface emissivity and vegetation parameters. A spin-up period is needed to allow the models and their surface fields to adjust to the forcing and internal model physics (Giot et al., 2016). While for ALARO-0 the year 1979 was taken as spin-up year, REMO~~The model~~ was spun-up for 1~~3~~0 years ~~from 1979 to 2008 for REMO~~ to produce an equilibrium for the soil temperature and soil moisture. ~~and t~~These soil fields were then used as initial soil conditions when restarting the model from 1979.~~, while for ALARO-0 the year 1979 was taken as spin-up year. Therefore, 1979 will not be used for the analysis in the subsequent sections.~~ The data produced by both models have been uploaded to the ESGF data nodes (website: http://esgf.llnl.gov/).

2.3 Reference datasets

In order to validate the model results, monthly, seasonally and annually averaged values for temperature and precipitation are compared with different reference datasets. Gridded datasets are based on interpolated station data and are used instead of station observations to overcome the scale difference between the model and observation field (Tustison et al., 2001). A multitude of datasets were considered to estimate the reliability of the gridded observational temperature and precipitation, since all gridded datasets are characterized by uncertainties (Gómez-Navarro et al., 2012) ~~(New et al., 1999). When these datasets show large differences amongst each other, then the obtained model biases could be in part attributed to the observational uncertainty~~. The reference datasets are briefly presented in Table 1 and in the next sections we give a more detailed overview of the different datasets used in this study.

2.3.1 Climatic Research Unit TS dataset

The gridded Climatic Research Unit (CRU) TS dataset (version 4.02) contains ten climate related variables for the period 1901-2018~~7~~ ~~(Harris et al., 2014)~~ at a grid resolution of 0.50° covering the complete global land mass (excluding Antarctica) (Harris et al., 2020~~New et al., 1999; New et al., 2000; Harris et al., 2014~~). Monthly values of minimum, maximum and mean near surface air temperature and precipitation are used in the current study. This dataset is widely used all over the world and in a wide range of disciplines ~~(Harris et al., 2014)~~, however, ~~there are also~~ some issues have been reported (Harris et al., 2020). Main concerns include sparse coverage of measurement stations over certain regions, e.g. the Northern ~~of~~ Russia ~~(New et al., 2002)~~ and the dissimilarities in measurement methods that are used ~~between and within~~by different countries (Harris et al., 2020)~~(New et al., 1999). New et al. (1999) indicate as well that the interpolation method is likely to produce warmer temperatures in sparsely covered mountainous areas and Hu et al. (2018) reported that the precipitation is underestimated in the centre of the CAS-CORDEX domain, especially in the mountainous areas~~.

2.3.2 Matsuura and Willmott gridded dataset

The Matsuura and Willmott (MW) (version 5.01) gridded dataset of the University of Delaware contains monthly values at a 0.5° resolution based on temperature and precipitation station observations. The main differences with the CRU dataset are the use of different measurement station networks and spatial interpolation methods (~~Willmott et al., 1985;~~ Willmott and Matsuura, 1995;~~-~~ Harris et al., 2020~~Willmott and Robeson, 1995~~). It is known that the MW dataset generally underestimates the precipitation in the central part of the CAS-CORDEX domain but especially during spring (Hu et al., 2018). The MW dataset contains globally up to 0.4 °C warmer temperatures for the latest decades compared to CRU (Harris et al., 2020).

2.3.3 Global Precipitation Climatology Centre dataset

The Global Precipitation Climatology Centre (GPCC) (version 2018) of the German Weather Service~~Deutscher Wetterdienst~~ is a monthly land surface precipitation dataset at 0.25° resolution based on rain gauge measurements. The GPCC full data monthly product version 2018 contains globally regular gridded monthly precipitation totals. This updated version is using "climatological infilling" to avoid interpolation artefacts for regions where an entire 5° grid is not covered by any station data (Schneider et al., 2018). Hu et al. (2018) concluded for the central part of our domain that GPCC is more in line with the observed station data in Central Asia compared to CRU and MW, however, precipitation is underestimated in mountainous areas and seasonal precipitation is underestimated, especially during spring. In addition, the GPCC has no similar dataset for other variables and thus, only precipitation can be validated with this dataset.

2.3.4 ERA-Interim

Reanalysis products like ERA-Interim are more continuous in space and time than station data, but they do contain biases as well. The ERA-Interim reanalysis of the European Centre for Medium-Range Weather Forecasts (ECMWF) is available from 1979 onwards. The spatial resolution of the dataset is approximately 79 km (T255 spectral) with 60 levels in the vertical direction from the surface up to 0.1 hPa (Dee et al., 2011). The ERA-interim data have been further interpolated and used as forcing for both RCMs at a spatial resolution of 0.25°. Total monthly precipitation ~~at a spatial resolution of 0.25°~~ was obtained from the Monthly Means of Daily Forecast Accumulations dataset by taking the mean over the precipitation amounts that are available for two time steps: 00:00 and 12:00. The Monthly Means of Daily Means data of 2 m temperature ~~at 0.25° is~~are used to study the ~~difference~~ spread between observational gridded datasets and reanalysis data. ~~In addition, the temperature of ERA-Interim can reveal if the deviations between the RCMs and observational datasets are due to initial errors in the boundary conditions or not, since the RCMs were driven by ERA-Interim. This is not the case for precipitation since the RCMs are not using the ERA-Interim precipitation as forcing. They simulate precipitation based on other variables which are forced by ERA-Interim such as temperature and specific humidity.~~ Several studies have shown that ERA-Interim tends to have a warm bias in the northern part over the CAS-CORDEX region, especially during winter (Ozturk et al., 2012 and 2016). Ozturk et al. (2012) relates this to the insufficient ability of ERA-Interim to produce a snow cover in winter. Additionally, Ozturk et al. (2016) showed that ERA-Interim tends to have a dry bias over the CAS-CORDEX region.

2.4 Analysis methods

~~Gridded datasets are based on interpolated station data and used instead of station observations to overcome the scale difference between the model and observation field (Tustison et al., 2001). Nevertheless, t~~The grids of the observational and reanalysis datasets generally differ from the model grid. Therefore, an interpolation to one common grid is needed in order to compare them (Kotlarski et al., 2014). The output of the RCMs was upscaled and bilinearly interpolated to the 0.50° resolution grid of the observational gridded datasets~~As the CRU dataset has the lowest spatial resolution, the other datasets (both modelled and gridded) are upscaled to this grid. For the interpolation to the CRU grid, bilinear interpolation is used~~.

For ALARO-0 and REMO, hourly values of 2 m temperature and convective and stratiform rain and snow are available. The precipitation variables were added up in order to obtain the hourly total precipitation which in turn was used to calculate monthly totals and seasonal and annual means. The diurnal temperature range was obtained by subtracting the minimum temperature from the maximum temperature and a~~The hourly temperature data are used to compute the daily minimum, mean and maximum temperatures. These daily values were then used to create monthly, seasonal and annual means of the mean, minimum and maximum temperature. Additionally, a~~ height correction was performed for mean, minimum and maximum temperature ~~using the topography of the CRU database and~~ assuming a uniform temperature lapse rate of 0.0064 K m^{-1}.

The model evaluation is done by calculating different evaluation metrics over the CAS-CORDEX domain for the 1980-2017 period. We computed the bias for the monthly, seasonal and annual climatological means of the evaluated variables to obtain graphs of the annual cycle ~~based on the monthly means of the datasets to get~~ maps that visualise the spatial patterns of the ~~differences~~ bias between the RCMs ~~or~~ and reference datasets. ~~and the CRU dataset.~~ The relative bias for precipitation is computed by subtracting the CRU value from the RCM or any other reference dataset~~value~~ and dividing it by the CRU value. These climatological means and biases were spatially averaged to obtain one mean value over the complete domain. Additionally, Taylor diagrams were produced in order to study the model performance for the different seasons and for annual means. Taylor diagrams supplement the bias analysis by visualizing in a concise way information about the correlation, centered root mean square error (RMSE) and ratio of spatial variability (RSV) between the model and the observational dataset (Taylor, 2001). The RSV is defined as the ratio of the model standard deviation and the standard deviation of the reference dataset, here CRU, over the spatial grid domain. In this study the Taylor diagrams represent the spatial pattern correlation between model and reference data, which is obtained by calculating correlations across the grid points of the CAS-CORDEX domain. For the used formulas we refer to appendix A of Kotlarski et al. (2014).

3 Results

In this section, the results of the model evalu~~id~~ation ~~results~~ are presented with a focus on evaluation metrics of seasonal means of~~in~~ mean, minimum and maximum near surface air temperature (henceforth denoted as temperature) and seasonal mean precipitation (henceforth precipitation). Limitations of the observational datasets should be kept in mind when interpreting the evaluation results (Kotlarski et al., 2014). These limitations are investigated by comparing the different observational datasets and their implications for the evaluation ~~validation as~~ will be described in Sect. 4.

3.1 Mean temperature

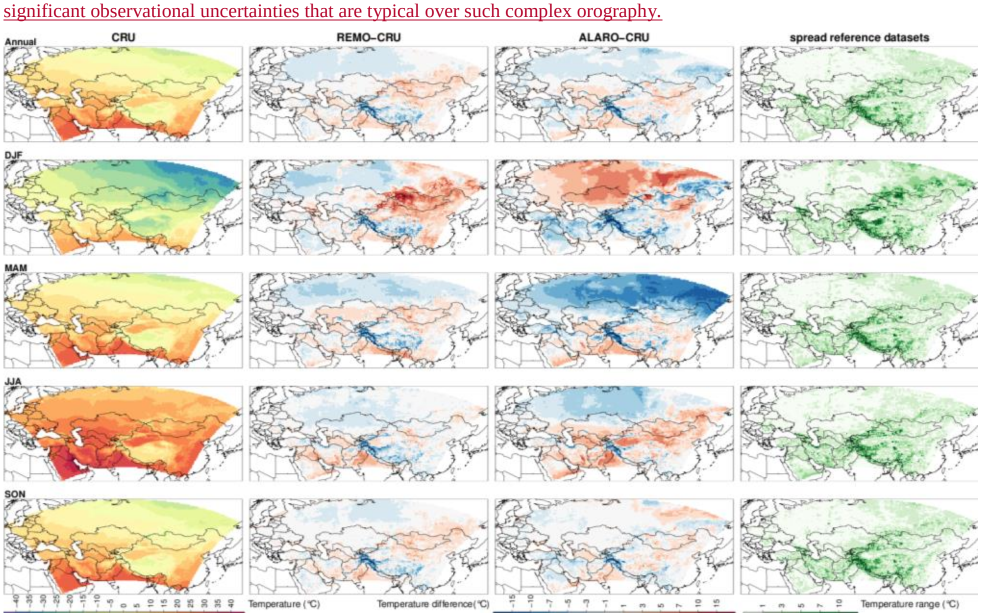
3.1.1 Annual and seasonal means over CAS-CORDEX domain

Formatted: Normal

In Fig. 2, the mean seasonal and annual temperature observations of CRU and the model biases with respect to CRU are shown for the 1980-2017 period. Moreover, the spread between the reference datasets (ERA-Interim, MW and CRU) is shown in the column at the right. Both RCMs are producing similar mean annual temperature patterns since they have similar biases with respect to CRU~~, except for the northeastern part of the domain, where REMO has a limited positive bias and ALARO-0 a limited negative one~~. At the same time a dipole pattern arises in the temperature bias of ALARO-0 between north and south and for REMO between east and west, with a peak in positively biased temperatures over north-western Mongolia. Annual biases vary between -3°C and 3°C for both RCMs, a~~A~~part from the orographically complex regions and some areas in North and East Siberia for ALARO-0. These regions exhibit a spread of 3 °C and more between the observational datasets and thus it is difficult to evaluate the models accurately in those regions.

On the seasonal timescale, biases over larger areas are mainly pronounced in winter (DJF) and spring (MAM), particularly for ALARO-0 with strong biases up to 10 °C and -15 °C respectively in the north-eastern part of the domain. In winter the most pronounced bias is found for REMO over the north-western part of Mongolia in the Altai mountains. Additionally, the REMO model has a cold bias in the western part of Russia during winter, while ALARO-0 shows a warm bias. During spring, cold biases are found for both models in the northern part of the domain, but the biases of ALARO-0 are more pronounced than those of REMO. For the summer (JJA) season, warm biases occur over the southern part of the domain and cold biases are more dominant in the north. These biases in summer are more pronounced for ALARO-0. Both models show modest bias patterns in autumn (SON), with in particular warm biases over the eastern part of the domain.

Biases in the high-altitude regions are largely persistent throughout the seasons. More specifically, both RCMs have large negative biases over the Pamir Mountains (Tadjikistan) and the Himalayas, while they also feature negative biases over the Tibetan Plateau, although this is to a lesser extent the case for ALARO-0 where this is only clearly visible for the winter season. As mentioned before and visualised in Fig. 2, the biases in mountainous regions should be placed in perspective to the significant observational uncertainties that are typical over such complex orography.



~~, annual biases vary between -3°C and 3°C for both RCMs. On the seasonal timescale this range is exceeded by the ALARO-0 data with a significant warm bias in winter and cold bias in spring in the northern part of the domain. Mean temperature biases are for both RCMs largest in the eastern half of the domain and are most outspoken for the ALARO-0 model. For the winter (DJF) period the REMO model shows a significant warm bias over Mongolia and the eastern part of the domain, whereas for ALARO-0 the warm bias is concentrated over Russia and Kazakhstan. A similar warm bias during winter was~~

~~found over Scandinavia in the EURO-CORDEX runs with ALARO-0 (Giot et al., 2016). Giot et al. (2016) suggested this could be due to the strong synoptic scale forcing in winter and stable boundary layer issues. A warm bias during winter in the northeastern part of the domain was found as well by Russo et al. (2019) and Ozturk et al. (2012 and 2016) for the COSMO-CLM 5.0 and RegCM models, respectively. Furthermore, Fig. 2 shows that the REMO model has a cold bias in northeastern Europe during winter, a feature previously found for REMO over different domains that include this region (Pietikäinen et al., 2018). During spring (MAM) only a modestly cold bias is found for REMO in the northern part of the domain, while ALARO-0 has a very strong cold bias. For the summer (JJA) season, biases are limited over most of the domain for the REMO model but for the ALARO-0 model there are warm biases, except for the cold biases in the northwest and over the mountain ranges. Similar biases to those of ALARO-0 in summer were found by Russo et al. (2019) with the RCM COSMO-CLM 5.0. In spring~~

~~and summer both RCMs show a pronounced warm bias~~

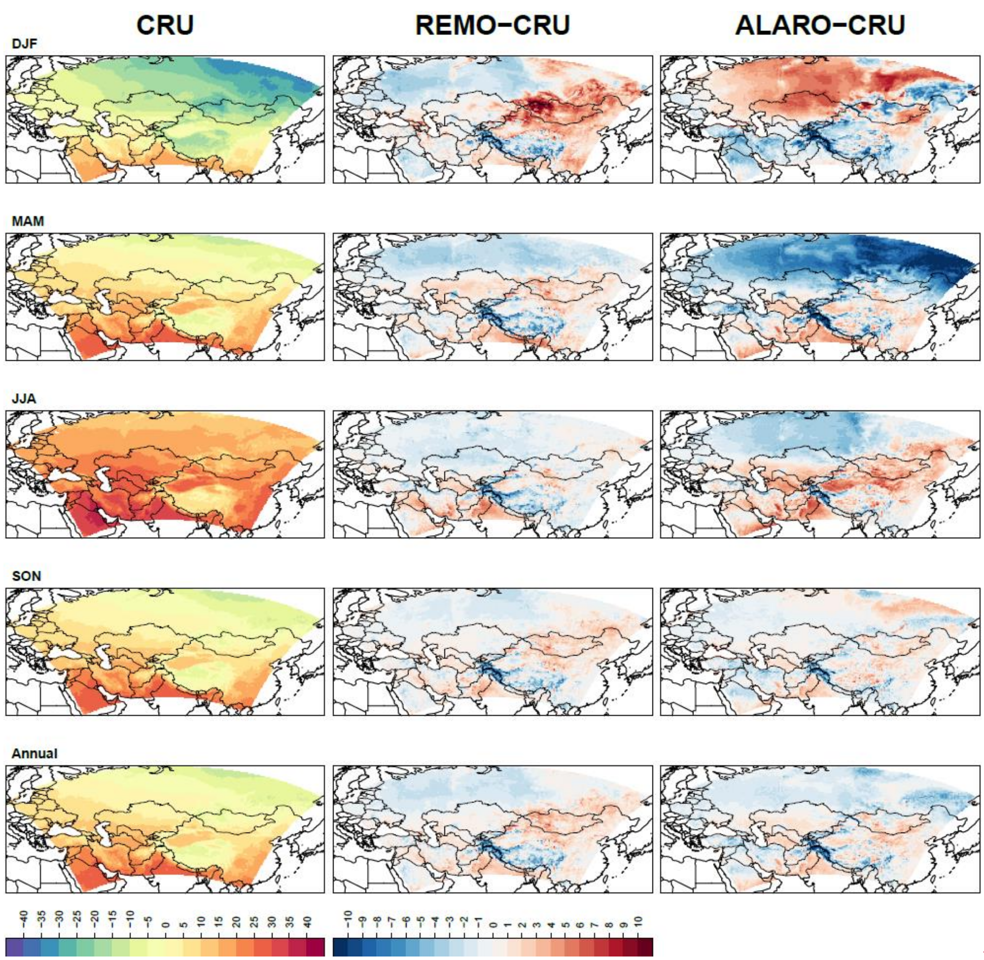


Figure 2: On the left: mean air temperature (°C) at 2 m height over the CAS-CORDEX domain based on the observational CRU dataset for the 1980-2017 period on annual level and for winter (DJF), spring (MAM), summer (JJA) and autumn (SON). In the middle columns: temperature difference (°C) between the simulated REMO mean temperature and the CRU mean temperature~~.~~, and ~~On the right:~~ temperature difference (°C) between the simulated ALARO-0 mean temperature and the CRU mean temperature. On the right: the range in temperature (°C) between the different reference datasets (CRU, MW and ERA-Interim).

~~over Pakistan and the northern part of India and there is also a north-south gradient from cold to warm biases over the Arabian Peninsula. The outcomes of both RCMs for mean temperature agree well with the CRU data in autumn (SON). Biases in the main high-altitude regions are largely persistent throughout the seasons. More specifically, both ALARO-0 and REMO have large negative biases over the Pamir Mountains (Tadjikistan) and the Himalayas, while they also feature negative biases over the Tibetan Plateau, although this is to a lesser extent the case for ALARO-0 where this is only clearly visible for the winter season. Additionally, REMO contains large positive biases over the Altai, especially in winter, while this is not the case for ALARO-0. As mentioned before these biases should be placed in perspective since there are uncertainties in the observational dataset as well, especially in the mountainous regions where observations are sparse.~~

The spatially averaged mean temperatures of CRU for the different seasons during the 1980-2017 period are given in Table 2, accompanied by the mean bias over the domain for the RCMs. In agreement with Fig. 2 the biases are very small for both RCMs during autumn. ~~Furthermore, it is clear from Table 2 that the strong cold bias during spring in the north for the ALARO-0 model has a larger negative impact on the spatially averaged bias than the warm bias during winter.~~

Figure 3 shows a~~ spatial~~ Taylor diagram for the mean temperature of both RCMs for the different seasons and for the annual mean value. Both models have in general a good model performance for temperature over the CAS-CORDEX domain for the different seasons and on the annual level since the spatial correlation between the model output and the CRU data is high (> 90 %), while the centred RMSE is small (< 0.5) and the normalized RSV is mostly close to 1. Based on Fig. 3, both RCMs perform best during autumn and the spatial correlation is lowe~~r~~st during summer for ALARO-0~~. However~~ while, the biases during summer are ~~for both RCMs~~ smaller than during winter and spring for both RCMs (Table 2 and Fig. 2). This is related to ~~the smaller~~less spatial ~~range~~ variability in temperatures during summer compared to the other seasons, as can be seen in Fig. 2 for CRU. An equal ~~deviation~~ bias in temperature for each season would~~ill~~ lead to a less good correlation in summer due to the smaller spatial variability in temperature during summer. During autumn and winter, both RCMs do simulate the normalized standard deviation of the temperature very well~~. However~~, although there was a clear warm bias observed during winter (Table 2 and Fig. 2). ~~but indicates that the RCMs capture the spatial variability well.~~ During spring the cold bias in the north is limited to -5 °C for the REMO model but not for ALARO-0, which leads to a clear overestimation of the normalized RSV during spring. Both RCMs overestimate the normalized RSV during summer and spring, while in winter they underestimate it slightly. The underestimation of the spatial variation by the RCMs in winter is due to the warmer temperatures in the northern part of the domain, where the coldest temperatures are observed for CRU (Fig. 2 and 3). In spring and summer, the spatial variation is overestimated since colder temperatures are simulated by the RCMs in the coldest part of the domain.

The small mean bias for ALARO-0 during summer (JJA) (Table 2) is obtained by averaging the warm biases in the south and the cold biases in the north (Fig. 2) and does not result in a very good overall performance of the modelled temperature (Fig. 3). ~~Based on Fig. 2, Fig. 3 and Table 2, ALARO-0 has a slightly better performance during autumn than REMO.~~ Comparing the metrics of the RCMs (Fig. 2, Fig. 3 and Table 2) shows that REMO is better in simulating the variability in temperature ~~, except for autumn,~~ and has smaller biases compared to ALARO-0, except for the autumn. On the other hand ALARO-0 better captures spatial temperature patterns since the spatial pattern correlation is slightly higher than for REMO, except during summer.

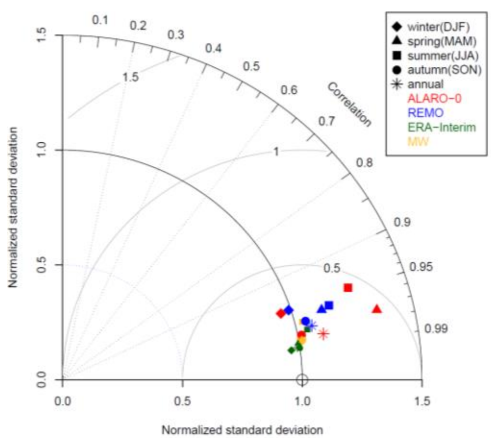


Figure 3: Normalized Taylor diagram representing the performance of mean temperature for seasonal and annual means for both RCMs (ALARO-0 and REMO), the ERA-Interim reanalysis and MW observational data with respect to CRU.

3.1.2 Annual cycles over subdomains

When analysing the seasonal cycle of the mean temperature for the different subdomains (Fig. 4), it is indeed observed that the RCMs simulate the mean temperature very well during the autumn months (months 9, 10 and 11). In the northern subdomains East Europe and West Siberia, there is on average a strong warm bias in December and January for ALARO-0, reaching a maximum of respectively 4.1 °C and 5.8 °C during December. During winter months (months 12, 1 and 2) REMO simulates temperatures within the uncertainty range for West Siberia and underestimates the temperatures on average by 1.4 °C in January over East Europe. REMO simulates warm biases around 2 °C in December and January over East Siberia. On average there is no strong warm bias observed for ALARO-0 during the winter months in East Siberia due to the compensation effect of cold biases, both in time (Fig.4) and space (Fig. 2). Furthermore, there is a remarkable cold bias observed for ALARO-0 during spring (months 3, 4 and 5) and June in the northern subdomains East Europe, West Siberia and East Siberia, reaching up to -7.3 °C over East Siberia during April. REMO is performing well during spring months over the northern subdomains. Compared to the northern subdomains, ALARO-0 simulates the annual cycle better for the southern subdomains West Central Asia and Tibetan Plateau but slightly overestimates the amplitude of the annual temperature cycle. REMO simulates the mean temperature very well over the West Central Asian subdomain with only a slight overestimation of the temperatures in July and August. In the mountainous area of the Tibetan Plateau REMO underestimates the temperatures, except for January and

Formatted: Normal, Space After: 10 pt, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

December. The better results in spring, summer and autumn for ALARO-0 over the subdomain Tibetan Plateau are due to spatial averaging of cold biases in the northern Himalayas and warm biases over the Taklamakan Desert and the opposite is true for REMO during winter (Fig. 2).

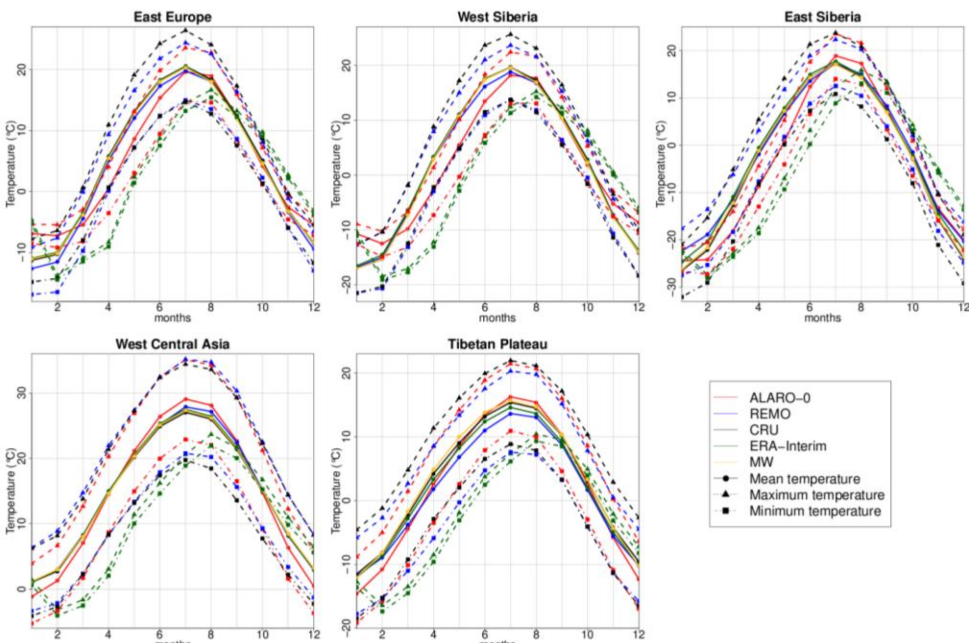


Figure 4: Annual cycles of the mean, minimum and maximum temperature for both RCMs (ALARO-0 and REMO) compared to the ERA-Interim reanalysis, MW and CRU observational data over five subdomains.

Formatted: Normal, Space After: 10 pt, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

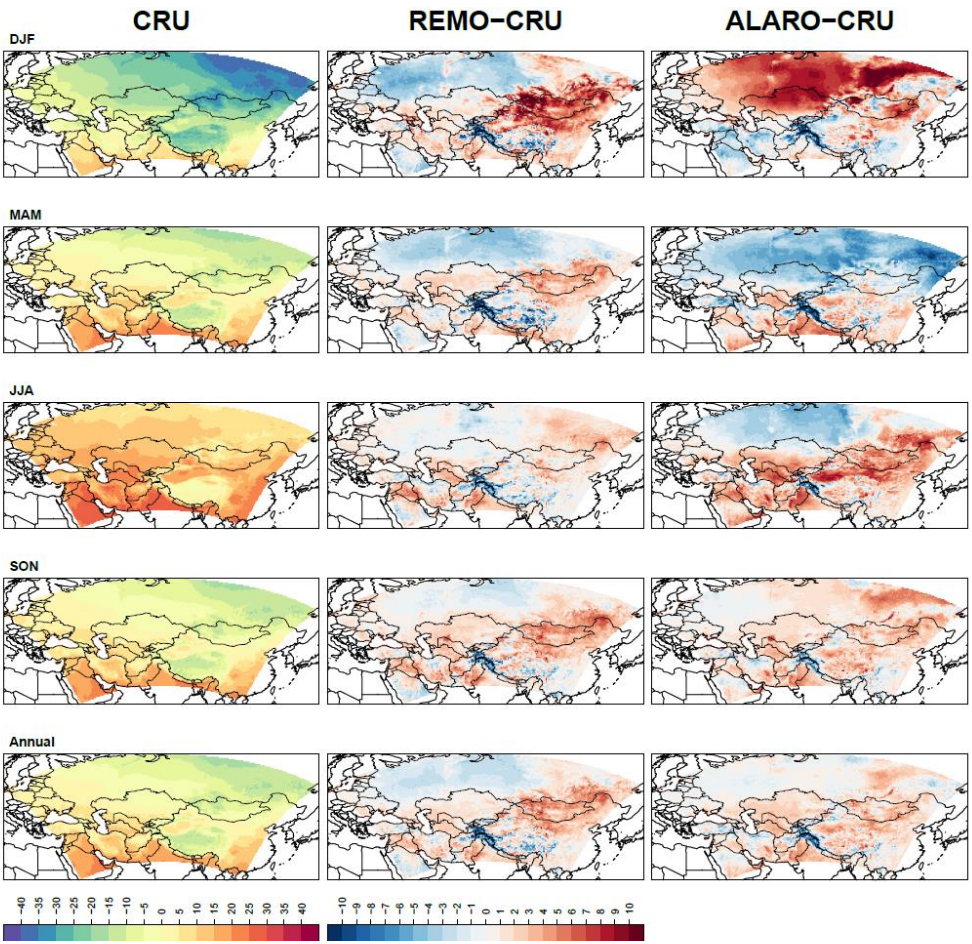
3.2 Diurnal temperature range

3.2.1 Annual and seasonal means over CAS-CORDEX domain~~Minimum temperature~~

The diurnal temperature range is found by subtracting the minimum temperature from the maximum temperature. Therefore, minimum and maximum temperature are first discussed and then conclusions for the diurnal temperature range are deduced. Similar ~~as for~~to the mean temperature, the modelled daily minimum temperature averaged over the different seasons and years during 1980-2017 is compared with the observational CRU data. At the annual scale, the bias of the minimum temperature ranges mostly between -3 °C and 3 °C for REMO and between 0 °C and 5 °C for ALARO-0 (Fig. 5). Compared to ALARO-0, the REMO model shows larger warm biases over Mongolia during all seasons, except for summer. These warm biases are most pronounced during winter. ALARO-0 shows as well large biases up to 15 °C, but they cover the northern part of the domain while the warm biases for REMO cover the eastern part of the domain. Moreover, strong cold biases are present in the north during spring for both models, but they are more pronounced for the ALARO-0 model with biases up to -10 °C in the north-eastern part of the domain. ~~Spatially averaged biases are larger for the minimum temperature than those of the mean temperature, except for the spring season, indicating that the model outputs are deviating more from the CRU data (Tables 2 and 3). This is due to the fact that both RCMs produce seasonal and annual means over the domain which are generally warmer for the minimum temperature than it was the case for the mean temperature. This causes a stronger warm bias in winter for the minimum temperature, which is especially visible in the northern part of the domain for the ALARO-0 model (Fig. 4). The REMO model also shows warmer biases over Mongolia during winter and spring when compared to the mean temperature (Fig. 2 and 4). Moreover, the cold bias in the north during spring for the ALARO-0 model is weaker for the minimum temperature than it was the case for the mean temperature.~~ During the summer season the biases for the REMO model are limited between -5 °C and 7 °C except for the Himalayan mountain range~~small~~, while the ALARO-0 model output has, except for the Himalayas, a cold bias up to -7 °C in the north-western part of Russia and warm bias up to 10 °C in the southern and eastern part of the domain ~~other regions~~ (Fig. 5~~4~~). ~~Following the main trend, these warm biases have a larger magnitude for minimum temperature when compared to the mean temperature.~~ In autumn, both models have a warm bias over almost the entire domain, except for the cold biases in the mountainous areas, the Arabian Peninsula, northern Iran and for REMO also in the central northern part of the domain. The increased minimum temperatures obtained with the RCMs indicate that they do not capture the coldest diurnal temperatures~~which was not the case for mean temperature. The warm minimum temperatures of the RCMs indicate that they underestimate the coldest diurnal temperatures or that the observational CRU dataset overestimates them. Although the magnitude of the biases is different for mean and minimum temperature, the spatial patterns are maintained for each of the RCMs. This means that these two variables are spatially highly correlated with each other in both, models and observations~~.

The metrics in Fig. 6 show that the RCMs simulate the minimum temperature spatially well for annual and seasonal means. ALARO-0 has at annual and seasonal scale, except for summer, a slightly better spatial pattern correlation with the minimum temperatures of the CRU dataset than REMO. On the other hand, REMO better simulates the variability and mean minimum temperature, except for autumn where ALARO-0 simulates the variability better (Fig. 6 and Table 3).

The maximum temperatures are underestimated by both RCMs and this underestimation is more pronounced for ALARO-0 than for REMO at the annual scale and for all seasons (Fig. 7 and Table 4). Figure 7 shows that the cold bias is especially present in the northern part of the domain during spring and to a lesser extent during summer for both RCMs. In autumn the cold bias in the north is limited to -3 °C, but some stronger biases up to -7 °C appear in the north-east for the ALARO-0 model. The warm biases during autumn are limited to 5 °C and excluding the Himalayas, the smallest range in biases is obtained for both RCMs during this season. During winter, a negative spatially averaged bias of -0.77 °C is obtained for the mean maximum temperature of ALARO-0 and a small positive bias of 0.08 °C for REMO (Table 4). These limited spatial biases are obtained by biases with an opposite sign in different parts of the domain. REMO has cold biases in the north-west and warm biases in the east, except for the Tibetan Plateau, while ALARO-0 produces warm biases in the north and cold biases in the south-west and north-east.


CRU
REMO−CRU
ALARO−CRU
DJF
MAM
JJA
SON
Annual
−40 −35 −30 −25 −20 −15 −10 −5 0 5 10 15 20 25 30 35 40
−10 −9 −8 −7 −6 −5 −4 −3 −2 −1 0 1 2 3 4 5 6 7 8 9 10

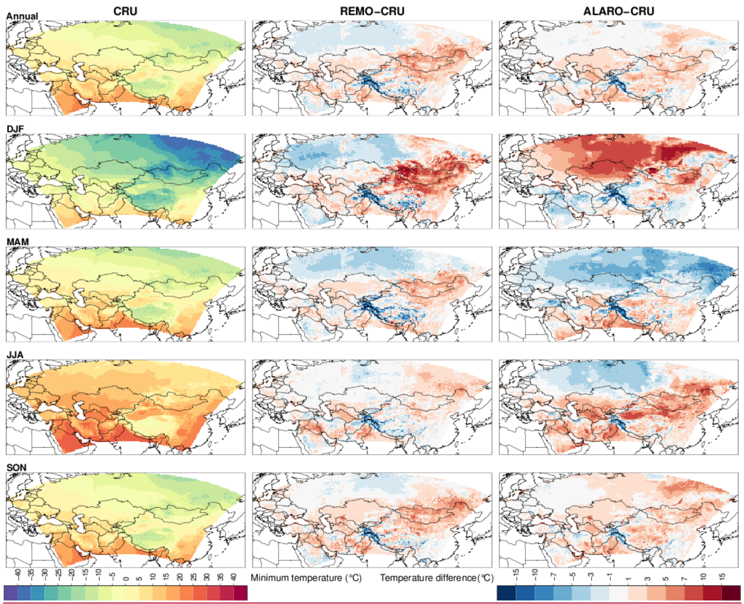


Figure 54: On the left: minimum air temperature (°C) at 2 m height over the CAS-CORDEX domain based on the observational CRU dataset for the 1980-2017 period on annual level and for winter (DJF), spring (MAM), summer (JJA) and autumn (SON). In the middle: temperature difference (°C) between the simulated REMO minimum temperature and the CRU minimum temperature. On the right: temperature difference (°C) between the simulated ALARO-0 minimum temperature and the CRU minimum temperature.

~~The metrics in Fig. 5 show that the RCMs simulate the minimum temperature spatially well for annual and seasonal means. When comparing them to those of the mean temperature (Fig. 3), then it is seen that the metrics of both variables are similar for both RCMs during the different seasons. Similar as was found for the mean temperature, ALARO-0 has on annual level a slightly better spatial pattern correlation with the minimum temperatures of the CRU dataset when compared to REMO, except for the summer for which the correlation deviates even more for minimum temperature (Fig. 5). On the other hand, REMO simulates better the variability and mean minimum temperature (Fig. 5 and Table 3). Similar as for mean temperature, ALARO-0 simulates for the minimum temperature the variability less good during summer and spring (Fig. 4 and 5).~~

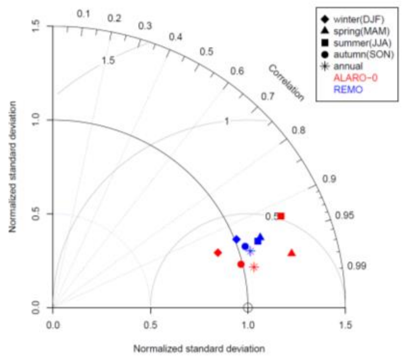
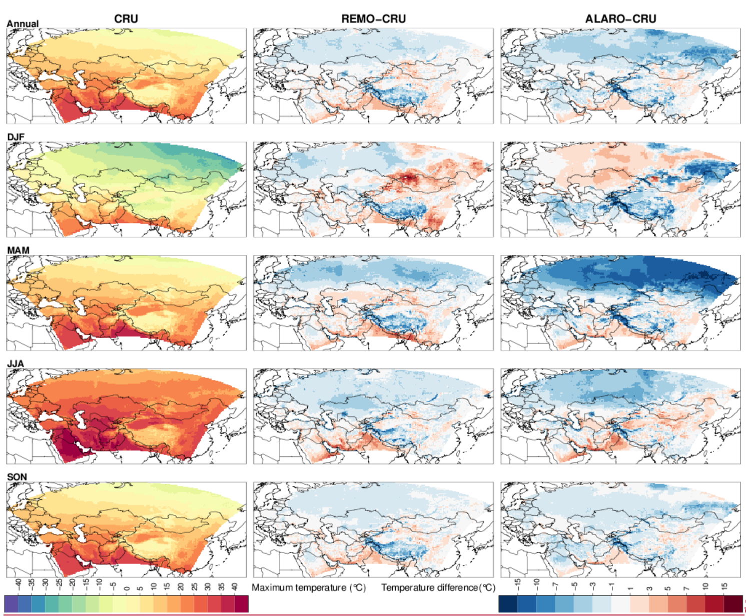


Figure ~~5~~6: Normalized Taylor diagram representing the model performance of the minimum temperature for seasonal and annual means for both RCMs (ALARO-0 and REMO) with respect to CRU.



~~3.3 Maximum temperature~~

~~For the maximum temperature (Fig. 6), similar spatial patterns are found in the biases as for the mean temperature and the minimum temperature (Fig. 2 and 4) over the different seasons and for the annual mean. However, the biases are generally~~

~~colder than it was the case for the variables mean temperature and minimum temperature (Tables 2, 3 and 4). This underestimation of the maximum temperatures is more pronounced for ALARO-0 than for REMO. During winter it counters the warm bias that was obtained for mean and minimum temperature, resulting in a negative spatially averaged bias for the mean maximum temperature of ALARO-0 and a small positive one for REMO (Table 4). In Fig. 4 it is seen that he cold bias present in the northern part of the domain during spring is more pronounced for both RCMs due to the underestimation in maximum temperatures, which results especially for ALARO-0 in a strong deviation from the observational data. In autumn the smallest range in biases is obtained for both RCMs, which was the case as well for minimum and mean temperature.~~

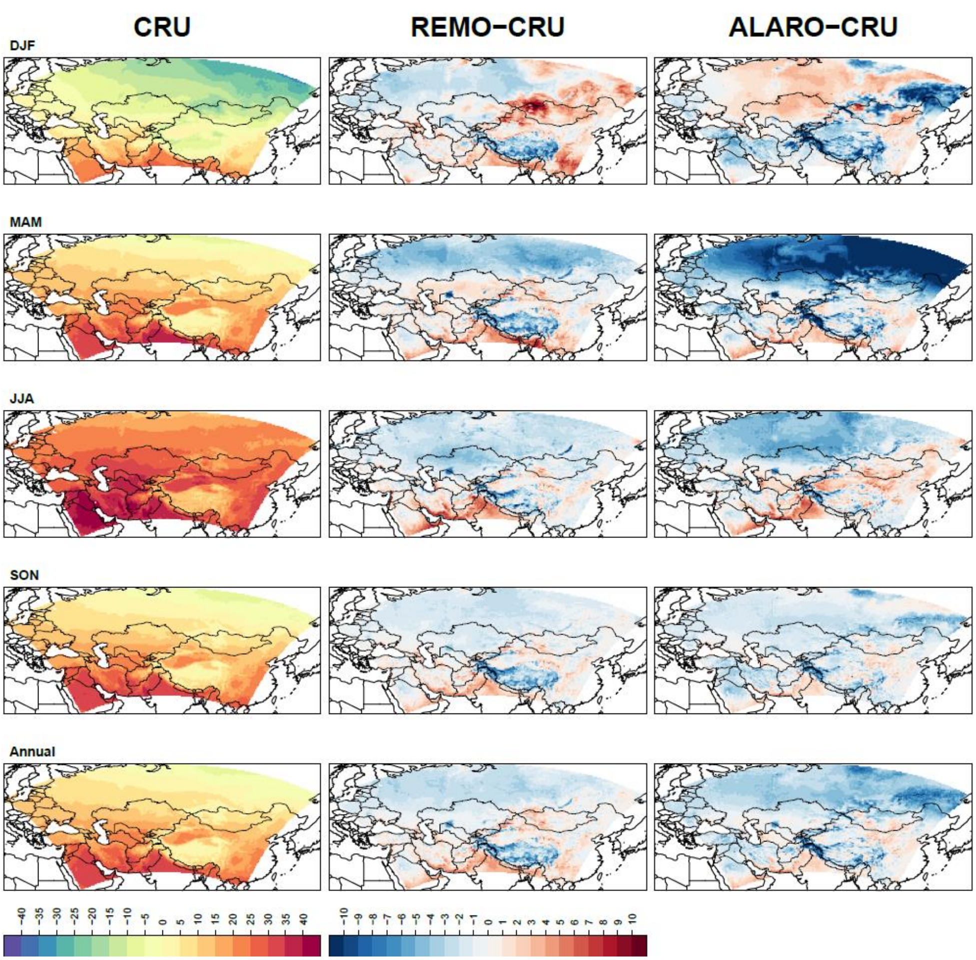


Figure ~~6~~7: On the left: maximum air temperature (°C) at 2 m height over the CAS-CORDEX domain based on the observational CRU dataset for the 1980-2017 period on annual level and for winter (DJF), spring (MAM), summer (JJA) and autumn (SON). In the middle: temperature difference (°C) between the simulated REMO maximum temperature and the CRU maximum temperature. On the right: temperature difference (°C) between the simulated ALARO-0 maximum temperature and the CRU maximum temperature.

Formatted: Centered

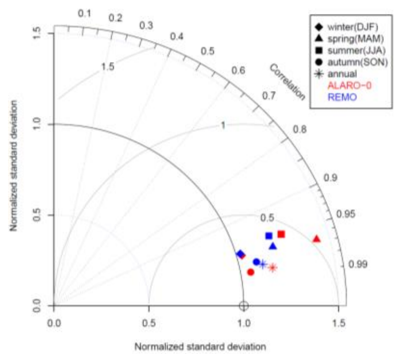


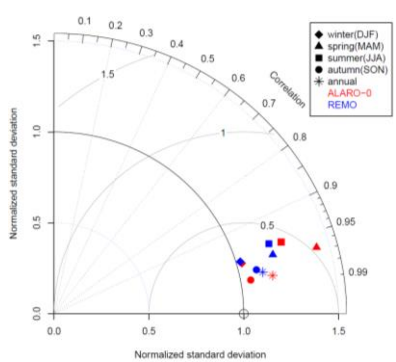
Figure 8: Normalized Taylor diagram representing the model performance of the maximum temperature for seasonal and annual means for both RCMs (ALARO-0 and REMO) with respect to CRU.

Figure 8~~7~~ shows that both models have an acceptable model performance for maximum temperature over the CAS domain, since the spatial pattern correlation is high and the normalized RSV is mostly close to 1. Additionally, it is seen that both RCMs overestimate the normalized RSV of the maximum temperature (Fig. ~~7~~8). ~~This differs from the mean temperature where both models underestimated the normalized RSV during winter (Fig. 3).~~ Based on Fig. 7~~6~~ and 8~~7~~, both RCMs simulate ~~best~~ the maximum temperature best during autumn.

~~The strong warm bias in the mean temperature over Russia for ALARO-0 during winter (Fig. 2) is mostly caused by the warm bias in minimum temperatures (Fig. 4), since the warm bias is larger for minimum temperatures than for maximum and mean temperatures (Fig. 6 and 2). This means that ALARO-0 fails to reproduce the low nocturnal temperatures. On the other hand, the large negative bias in spring over Russia is mostly caused by the cold bias in maximum temperatures (Fig. 6), meaning that ALARO-0 fails to reproduce the daytime temperatures in spring. In general the minimum temperature (Table 3 and Fig. 4) shows warmer biases than the mean temperature (Table 2 and Fig. 2) and the maximum temperature (Table 4 and Fig. 6) shows colder biases compared with the mean temperature over the different seasons. From this can be concluded that both cold and warm extreme temperatures are simulated less extremely by the models over most of the domain compared to the extreme temperatures in the observational CRU dataset. In other words, the daily temperature range is generally underestimated by both RCMs.~~ In general the minimum temperature (Table 3 and Fig. 4) shows warmer biases than the mean temperature (Table 2 and Fig. 2) and the maximum temperature (Table 4 and Fig. 6) shows colder biases compared with the mean temperature over the different seasons. From this can be concluded that both minimum and maximum temperatures are simulated less extremely by the models over most of the domain compared to the observational CRU dataset. In other words, the daily temperature range is generally underestimated by both RCMs.

3.2.2 Annual cycles over subdomains

Moreover, the annual cycles in Fig. 4 show that both minimum and maximum temperatures are overestimated by ALARO-0 during winter in the northern part of the domain, while they are underestimated during spring. In summer the model is able to restore its balance and to simulate temperatures as they are observed. For REMO the maximum temperature is underestimated during winter, spring and summer in East Europe, while the minimum temperature is only strongly underestimated during winter. REMO overestimates the minimum temperatures during the complete annual cycle for East Siberia, while the maximum temperatures in East Siberia are only overestimated during winter and underestimated during spring and summer. Both RCMs underestimate the maximum temperatures for the entire annual cycle over the Tibetan Plateau subregion. ALARO-0 underestimates the minimum temperatures during the winter months and overestimates them during the summer months, while REMO slightly overestimates winter and underestimates summer minimum temperatures.



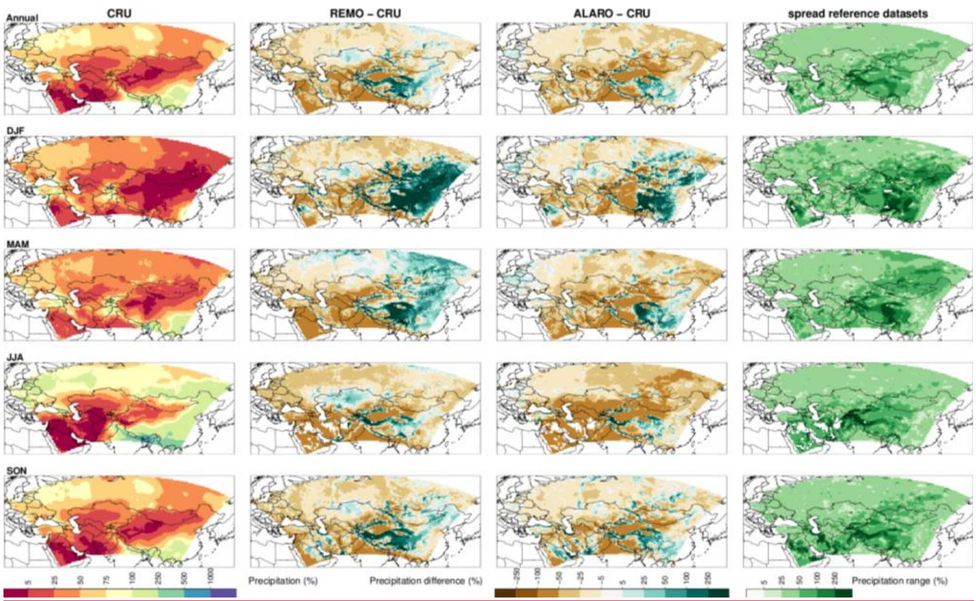
~~**Figure 7: Normalized Taylor diagram representing the model performance of the maximum temperature for seasonal and annual means for both RCMs (ALARO-0 and REMO) with respect to CRU.**~~

3.3~~4~~ Precipitation

3.3.1 Annual and seasonal means over CAS-CORDEX domain

In Table 5, the spatially averaged precipitation over the 1980-2017 period is given for CRU. ~~and t~~The relative biases of the RCMs with respect to CRU during the different seasons and on annual level are presented as well. For both RCMs the overall bias for precipitation is dry, except for REMO in spring. Figure 9 shows that the annual precipitation for both models lies mostly within the spread of the different reference datasets. Furthermore, a strong wet bias is present during winter for both RCMs over the south-eastern region and for REMO this wet bias extends even further up north to the Russian-Mongolian border. This large wet bias during the winter is partly due to the low precipitation quantities in several regions e.g. less than 5 mm per month in the Taklamakan and Gobi desert regions. The largest relative biases can be found in relatively dry regions and therefore the absolute biases are presented in the supplementary material Fig. S2 and Table S2. When the absolute bias during winter is examined, then it is seen that REMO only simulates a very small absolute overestimation in precipitation over Mongolia and the northern part of China, but both RCMs do overestimate the precipitation in the South-East Asian monsoon region during winter and spring (Fig. 9). The wet bias of the ALARO-0 model in the south-eastern CAS-CORDEX region is situated within the spread of the different reference datasets (Fig. 9 and 11). In summer, when most rain falls due to the East Asian Monsoon, a dry bias is present (Fig. 9 and S2).

~~In Fig. 8, it is shown that this wet bias for REMO during spring is caused~~


CRU
REMO – CRU
ALARO – CRU
spread reference datasets
Annual
DJF
MAM
JJA
SON
Precipitation (%)
Precipitation difference (%)
Precipitation range (%)

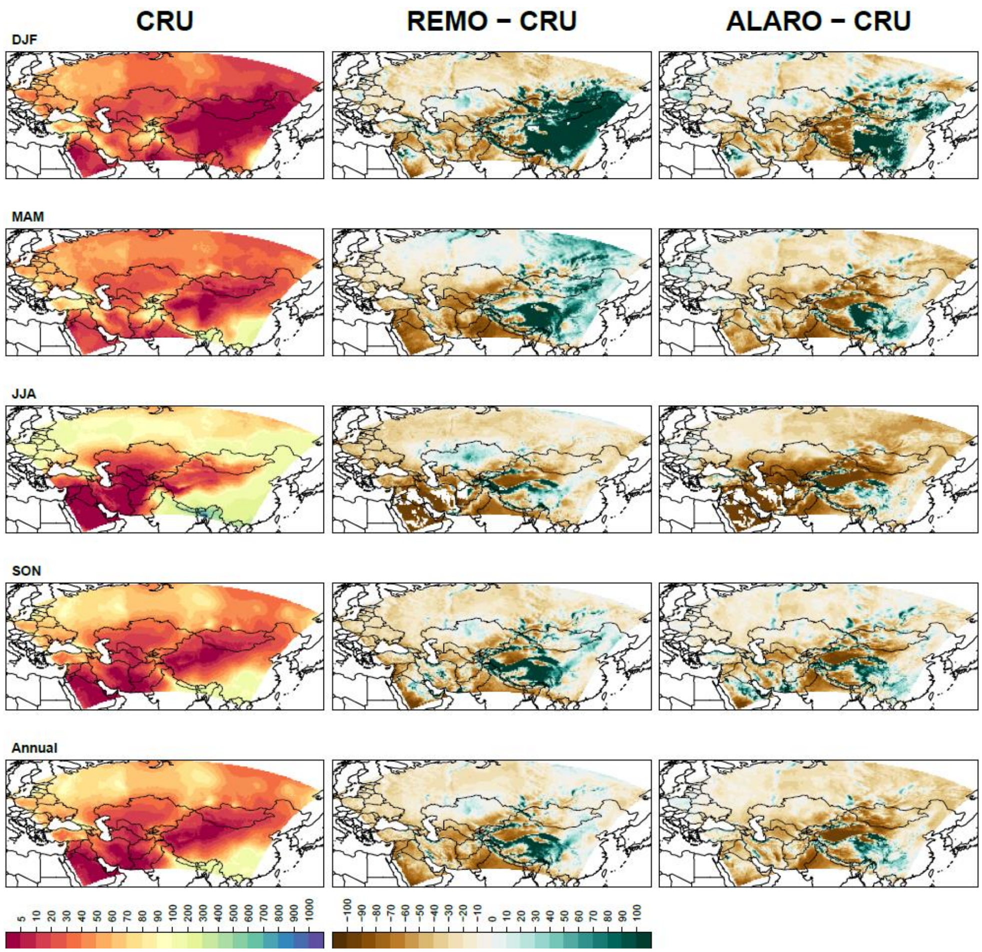


Figure ~~8~~9: Left: mean monthly precipitation amounts (mm month-1) over the CAS-CORDEX domain based on the observational CRU dataset for the 1980-2017 period on annual level and for winter (DJF), spring (MAM), summer (JJA) and autumn (SON). In the middle: relative difference between the average annual and seasonal CRU precipitation and the precipitation simulated by REMO and ALARO-0 (%). Right: the range in precipitation (%) between the different reference datasets (CRU, MW, GPCC and ERA-Interim)~~Relative difference between the average seasonal and annual CRU precipitation (mm month-1) and the precipitation simulated by REMO and ALARO-0 over the 1980-2017 period~~.

Next to these biases in the monsoon region, both models show dry biases over the Tarim basin and the south-western part of the domain during spring and summer. The Taklamakan and Arabian deserts are located here, which are already dry regions in the CRU dataset (Fig. 9). The absolute biases over this region are less pronounced in Fig. S2. In addition, both RCMs have a dry bias in the northern part of the domain during summer, which is the strongest dry bias in this region over the different seasons in absolute precipitation deficiency (Fig. S2).

From Fig. 10 can be deduced that ALARO-0 is better than REMO in capturing the annual and seasonal variations in precipitation since the RSVs are closer to 1. Additionally, ALARO-0 better captures the spatial patterns since the correlations are larger than those for REMO. The dry biases for ALARO-0 in Table 5 are thus caused by the simulation of systematically less precipitation compared to CRU over most parts of the domain (Fig. 9 and 11). Both RCMs show the largest error in normalized RSV during spring. This too large spatial variation is due to an overestimation of the precipitation in the wettest region combined with an underestimation in the driest region of the CAS-CORDEX domain (Fig. 9). During summer, both RCMs underestimate the variability in precipitation (Fig. 10).

~~by an overestimation of precipitation in the eastern part of the CAS-CORDEX domain. During spring, ALARO-0 shows only an extended wet bias in the southeastern part of the domain. A very strong relative wet bias is present during winter for both RCMs over this southeastern region and for REMO this wet bias even extents further up north to the Russian-Mongolian border. Russo et al. (2019) found a similar spatial pattern of a wet bias with their COSMO-CLM model as presented here for REMO. However, this large relative wet bias during the winter is partly due to the low rainfall quantities in the observational CRU dataset. The largest relative biases can be found in relatively dry regions, therefore the absolute biases are presented in the supplementary material Fig. S1 and Table S2. When the absolute bias during winter is examined (supplementary material Fig. S1), then it is seen that REMO does not simulate a large absolute overestimation in precipitation in Mongolia and the northern part of China, but both RCMs do overestimate the precipitation in the Southeast Asian monsoon region during winter and spring. This wet bias over the southeastern monsoon region during winter and spring is almost completely turned into a weak dry bias in summer, when most rain falls, except for Northern India and Pakistan where there is even a strong dry bias (Fig. 8 and S1). The wet bias during winter and spring and dry bias in summer between CRU and both RCMs over the southeastern part of the CAS-CORDEX domain can be linked to an early onset of the monsoon. However, this should be further investigated with an annual cycle over the region. This feature is more pronounced for REMO and was already highlighted by Remedio et al. (2019), who saw the same shift for REMO with different CORDEX experiments over the subtropical region where the Asian monsoon takes place.~~

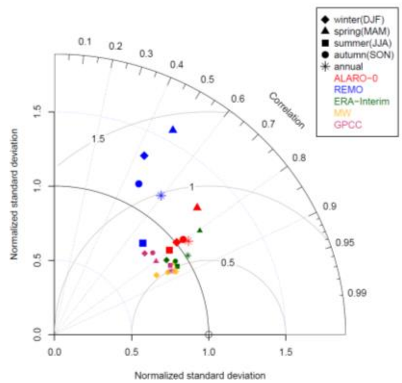
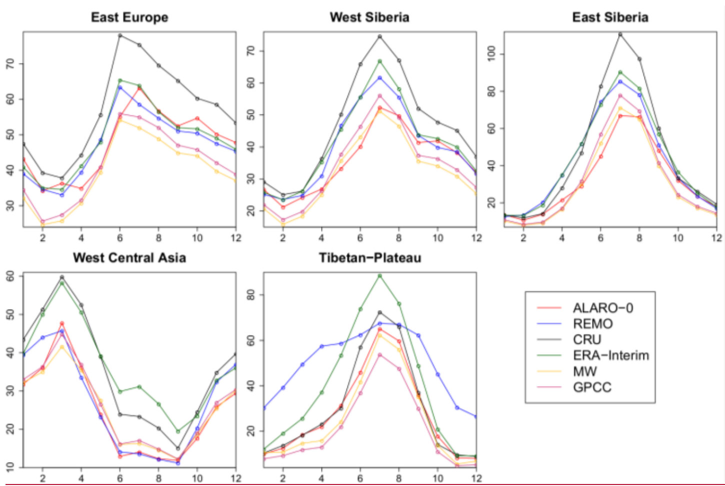


Figure 10~~9~~: Normalized Taylor diagram representing the model performance of precipitation for seasonal and annual means for both RCMs (ALARO-0 and REMO), gridded observational datasets (MW, GPCC) and the ERA-Interim reanalysis data with respect to CRU.

~~Next to these biases in the monsoon region, both models show during spring and summer dry biases over the Tarim basin and the southwestern part of the domain, where respectively the Taklamakan and Arabian desert are located, which are already dry regions in the CRU dataset and therefore show a strong dry bias in Fig. 8. The absolute biases over this region are less pronounced in Fig. S1. In addition, both RCMs have a dry bias in the northern part of the domain during summer, which is the strongest dry bias in this region over the different seasons in absolute precipitation deficiency, causing the strongest dry spatially averaged bias for this season (Table 5).~~

~~From Fig. 9 can be deduced that ALARO-0 is better than REMO in capturing the seasonal variation in precipitation since the RSVs are closer to 1. Additionally, ALARO-0 captures for all seasons better the spatial patterns since the correlations are larger than those for REMO. The dry biases for ALARO-0 in Table 5 are thus caused by the simulation of systematically less precipitation than the precipitation amounts in the CRU data. Both models are worse in simulating the spatial correlation of precipitation (Fig. 9) compared to the mean, minimum and maximum temperature (Fig. 3, 5 and 7). The lower accuracy of simulated precipitation is due to the fact that precipitation is less systematic affected by land cover and topography compared to temperature (Kotlarski et al., 2014). Both RCMs show the largest error in normalized RSV during spring. This overestimation of the spatial variation is due to the overestimation and underestimation of the precipitation amount in respectively the wettest and driest areas of the domain (Fig. 8). During summer both RCMs overestimate the variability in temperature (Fig. 3), while they underestimate the variability in precipitation (Fig. 9).~~3.3.2 Annual cycles over subdomains

The annual cycles over the subdomains show that ALARO-0 and REMO indeed mostly underestimate the precipitation values of CRU in the different subdomains, but for East Europe and the Tibetan Plateau the precipitation amounts are higher than those of MW and GPCC and are thus within the range of observational spread (Fig. 11). ALARO-0 does underestimate the precipitation slightly in May and June over West Siberia and in June and July over East Siberia. For the West Central Asian subdomain, both RCMs underestimate the precipitation in spring and summer. REMO overestimates the precipitation slightly over the East Siberian subdomain in spring. Additionally, it is seen that REMO is unable to simulate the annual cycle of precipitation correctly over the subdomain of the Tibetan Plateau (Fig. 11). The precipitation rates are too high, except during the summer when the Asian Monsoon takes place.



Formatted: Centered

Figure 11: Annual cycles of precipitation (mm/month) for both RCMs (ALARO-0 and REMO) compared to the ERA-Interim reanalysis, MW, GPCC and CRU observational data over five subdomains.

4 Discussion

4.1 Temperature

4.1.1 CAS-CORDEX domain

Formatted: Normal

~~The underestimation of the diurnal range over the CAS-CORDEX domain was also observed by Russo et al. (2019) for the winter and summer season. Their RCM produced smaller diurnal ranges compared to different observational datasets and the comparison between the observational datasets pointed out that CRU overestimates the diurnal range in the northeastern part of the domain. This explains why both RCMs show the largest shift between biases in minimum and maximum temperature over this area. Hence, the RCMs underestimate the diurnal range, which is similar to the findings over other regions (Laprise et al., 2003; Kyselý and Plavcová 2012), but in the northeast of the Central Asia domain the more pronounced underestimation is due to an overestimation by CRU. For the Czech Republic (Europe), Kyselý and Plavcová (2012) stated that this~~

~~underestimation is probably caused by an incorrect simulation of atmospheric circulation, cloud cover or heat and moisture fluxes between land surface and atmosphere.~~

When comparing ~~we compare~~ the above results ~~for~~ of temperature with the other reference datasets (Fig. 3), ~~then~~ the normalized standard deviation of ERA-Interim and MW ~~deviate~~ differs less from CRU than the RCMs ~~do~~ during spring and summer ~~(Fig. 3)~~. This implies that the deviation in spatial variation of temperature between the RCMs and CRU cannot be completely explained by the observational uncertainty, meaning that the data of the RCMs deviates from the observations and can be improved. The spatial correlations between CRU and ERA-Interim or MW are close to those between CRU and the RCMs, which indicates that the RCMs are able to reproduce the spatial temperature patterns ~~very~~ well, even though they were slightly deviating from the spatial temperature patterns in the CRU data. The latter can also be explained by the spread of the reference datasets in Fig. 2: larger biases between the RCMs and CRU are especially located in regions where the spread between the different reference datasets is high, which means that there is a large observational uncertainty at those locations. Figure 3 shows that the larger RSVs of the RCMs during summer are partly due to an underestimation of the variability in the CRU dataset since the ERA-Interim and MW data show both a slight overestimation compared to CRU. In addition, it is seen that the observed spatial patterns are less reliable during summer since the two other reference datasets both show a lower spatial correlation with CRU during summer compared to the other seasons. The lower performance of the RCMs during summer can thus partly be explained by the observational uncertainty in spatial variation of temperatures. This is more pronounced for the summer season since the spatial variation in temperature is lower during this season. Ozturk et al. (2016) reported a lower spatial correlation during summer with RegCM4.3.5 at 0.50° horizontal resolution. Additionally, similar high spatial correlations are obtained during the different seasons for ALARO-0 and REMO at 0.22° horizontal resolution when compared to the results of Ozturk et al. (2016). Zhu et al. (2020) obtained spatial correlations that are slightly lower than those obtained for ALARO-0 and REMO. They obtained a slightly larger spatial variation during winter and similar results for the spatial variation in summer and on annual level. Although, it should be mentioned that their domain is smaller than the CAS-CORDEX domain and they used a different observational dataset which makes comparison difficult.

Larger differences between temperatures of the reference datasets in the region of the Tibetan Plateau (Fig. 2) were also observed by Ozturk et al. (2012 and 2016) and Russo et al. (2019) and this is due to the fact that gridded data is based on measurements of meteorological stations in the valleys (New et al., 1999). This is the case for the gridded observational data of CRU and MW (Fig. 4). The gridded observations are thus less reliable over the Himalayas and Tibetan Plateau and cause a bias of the RCMs within the range of observational uncertainty. Further, the amplification of the biases over the mountainous regions for the RCMs can be attributed to the used assumption of the lapse rate of 0.0064 K m-1 for the elevation correction (Kotlarski et al., 2014).

When comparing the mean spatial biases for the 1980-2017 period (Table 2), then it is seen that the differences between the observational datasets are smaller than the differences between the RCMs and CRU, except for autumn for both RCMs and for REMO on the annual level. Additionally, Fig. 2 and 4 show that for most parts of the domain the mean temperatures of ALARO-0 and REMO are lying within the range of spread between the reference datasets during autumn. From this we conclude that both RCMs simulate temperatures in autumn within the range of observational uncertainty. During winter, spring and summer none of the RCMs are able to reproduce temperatures that can be completely explained by the observational uncertainty (Fig. 2 and Table 2).

In the following subsection the temperature biases over snow covered areas during winter and spring are explained. For summer temperatures, Russo et al. (2019) found with COSMO-CLM 5.0 a spatial pattern with a cold temperature bias in the north and warm biases in the southern part of the domain except for some locations on the Tibetan Plateau, which is similar to ALARO-0. In general both ALARO-0 and REMO produced biases within a similar order of magnitude as was obtained with other RCMs over the CAS-CORDEX region (Russo et al., 2019) and Central Asian subdomains (Wang et al., 2020; Zhu et al., 2020).

4.1.2 Spring and winter biases in northern subdomains

The cold bias for REMO during winter over the East European subdomain is likely due to the surface treatment of the model when there is snow (Pietikäinen et al., 2018). Pietikäinen et al. (2018) already reported that the thermodynamics of the snow layer plays an important role in the cold bias that appears over East Europe during the months when snow cover is present. Although this cold bias occurs in the north-west, both models are producing on average temperatures that are too warm in winter and too cold in spring (Table 2).

New et al. (1999) mentioned that CRU contains colder temperatures in winter over Russia. The range between the different reference datasets is larger for the East Siberian subdomain, indicating that there is a larger uncertainty for this subdomain during winter. This observational uncertainty could explain the warm biases for both RCMs over the mountain ranges Altai, Yablonovy and Stanovoy since the spread between the reference datasets is larger than the obtained biases (Fig. 2). However, this is not the main reason for the warm bias over Russia since the spread between the reference datasets is smaller than the obtained biases.

Moreover, during winter the RCMs simulate warm biases in different regions, while in spring they both show a cold bias over the north (Fig. 2 and 4). Compared to the northern part in the CAS-CORDEX region, a similar warm bias during winter was found over Scandinavia in the EURO-CORDEX runs with ALARO-0 (Giot et al., 2016). Both regions have a similar climate which suggests that similar physical processes might be at the basis of these biases. The warm bias during winter and cold bias during spring in the north-eastern part of the domain are not due to a shift in the annual cycle in the northern part of the domain, although there is a delay in warming temperatures during spring.

A limited warm bias arises in the north during autumn, when the first snow cover appears over this region. This bias increases when the snow covered region expands. ALARO-0 seems to underestimate cooling above snow cover during stable conditions (Fig. 4). Mašek (2017) linked too warm temperatures above snow to the used single layer snow scheme (Douville et al., 1995). REMO is using a multi-layer snow scheme and does not encounter this problem.

A similar strong warm bias in the north, as found for ALARO-0 in winter, was also found by Ozturk et al. (2012) and Russo et al. (2019) for the RegCM and COSMO-CLM 5.0 models, respectively. Ozturk et al. (2012) related this warm bias to shortcomings in the simulation of snow, whereas Russo et al. (2019) found that changes in the snow scheme did not affect the simulation results significantly and did not reduce the warm bias in the north-east during winter. This shows that a complexer multi-layer snow scheme might not be enough to solve the warm bias for ALARO-0 during winter. Therefore, further investigation should be done to see whether the warm bias in winter over the northern part of the domain is due to the inability of the current snow scheme to reproduce the heat conductivity of snow.

In spring, the warm temperature bias of the ALARO-0 simulation over the northern subdomain evolves into a significant cold bias. This remarkable evolution is probably related to another issue related to the snow scheme as we find a delay in the springtime melting of the snowpack (not shown). Additionally, ALARO-0 simulates too high pressure values over the northern area (not shown). Further research is needed to clarify whether this overestimation of the Siberian High in the ALARO-0 simulations is coupled to the difficulties with the snow cover.

4.2 Diurnal temperature range

Spatially averaged biases are warmer for the minimum temperature and colder for the maximum temperature, when compared to those of the mean temperature (Tables 2, 3 and 4). This is due to the fact that both RCMs produce seasonal and annual means over the domain which are generally warmer for the minimum temperature and colder for the maximum temperature than it was the case for the mean temperature. This causes a stronger warm bias in winter for the minimum temperature and a stronger cold bias for maximum temperature in spring, which is especially visible in the northern part of the domain for the ALARO-0 model (Fig. 2, 5 and 7). Moreover, the cold bias in the north during spring for the ALARO-0 model is weaker for the minimum temperature than for the mean temperature. The REMO model shows warmer biases over Mongolia during winter and spring for minimum temperature and colder biases in maximum temperature in the north during spring when compared to the mean temperature.

Although the magnitude of the biases is different for mean, minimum and maximum temperature, similar spatial patterns are found in the biases of both RCMs over the different seasons and for the annual mean (Fig. 2, 5 and 7). This means that these variables are spatially highly correlated with each other in both models and observations. When comparing the metrics in Fig. 6 and 8 to those of the mean temperature (Fig. 3), then it is seen that the metrics of mean, minimum and maximum temperature are similar for both RCMs during the different seasons. However, both RCMs overestimate the normalized RSV of the maximum temperature for all seasons (Fig. 8), which differs from the mean temperature where ALARO-0 and REMO underestimated the normalized RSV during winter (Fig. 3). This indicates that there is a slightly larger spatial variation in winter maximum temperatures simulated by the RCMs with respect to CRU, while for mean temperatures a smaller spatial variation is simulated. Additionally, both minimum and maximum temperatures have a similar temporal pattern as for the mean temperature, e.g. the smallest range in mean, minimum and maximum temperature biases is obtained in autumn for both RCMs (Fig. 4). Moreover, the underestimation of the minimum and maximum temperatures in spring is more pronounced for ALARO-0 than for REMO (Fig. 4).

The RCMs underestimate the diurnal range, which is similar to the findings over other regions (Laprise et al., 2003; Kyselý and Plavcová 2012). The underestimation of the diurnal range over the CAS-CORDEX domain was also observed by Russo et al. (2019) for the winter and summer seasons. Their RCM produced smaller diurnal ranges compared to different observational datasets and the comparison between the observational datasets pointed out that the observational uncertainty is high for the diurnal range in the north-eastern part of the domain, which makes it difficult to evaluate the diurnal range accurately over this area. In particular ALARO-0 shows a very small range in the diurnal cycle of temperatures due to very high minimum temperatures (Fig. 4) and this could be due to the inability of the model to simulate temperatures correctly over snow cover during stable conditions (Mašek, 2017).

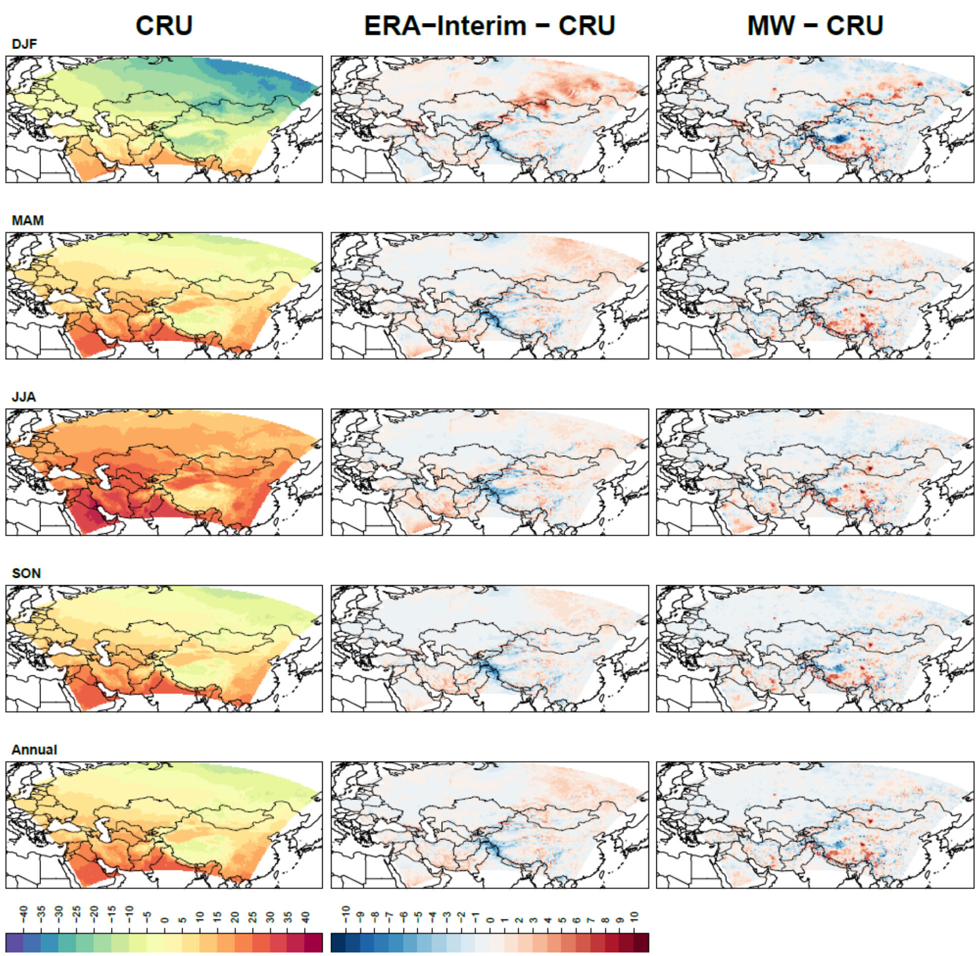
The evaluation of temperature and its diurnal cycle shows that a bias adjustment is essential before the climate data is applied in impact modelling for some regions, e.g. Tibetan Plateau and East Siberia. The current research is done within the AFTER project (Kotova et al, 2018). Within this project different bias-adjustment techniques are applied to the set of climate simulations. This will enable impact modellers to optimally use our climate data in their models for crop production, biomass production, etc.

~~When we compare the mean spatial biases for the 1980-2017 period (Fig. 2, Fig. 10 and Table 2), then it is seen that the differences between MW and CRU are smaller than the differences between the RCMs and CRU, except for the autumn and for REMO on the annual level. From this we conclude that both RCMs are able to simulate temperatures in the autumn that are within the range of observational uncertainty. During winter, spring and summer none of the validated RCMs are able to reproduce temperature means that can be completely explained by the observational uncertainty. Hence, in winter both models are producing on average temperatures that are too warm and in spring they are too cold. Figure 10 shows that the driving force ERA-Interim has a warm bias in winter over the northeastern part of the domain and thus, the warm bias that is produced by both RCMs in winter can be assigned to this forcing. It must be noted that the spatial pattern of the warm bias in the ERA-Interim data is more similar to the warm bias pattern created by REMO. This warm bias in winter for the driving force ERA-Interim and the reflection of it in the RCM data was also found by Ozturk et al. (2012 and 2016). They related this warm bias already to shortcomings in the simulation of snow. Contrary to Ozturk et al. (2016) but similar to Ozturk et al. (2012), the warm bias in the ERA-Interim forcing during winter is amplified by the two RCMs evaluated in this study. Additionally to the influence of the warm forcing New et al. (1999) mentioned that CRU contains colder temperatures in winter over Russia, although this is not the main reason of the warm bias over Russia since there are only patches of warm biases observed between the two observational datasets MW and CRU, namely over the mountain ranges Yablonovyy and Stanovoy in the southeastern part of Russia (Fig. 10). Hence, we conclude that the warm forcing is the main reason for the warm bias over Eastern Russia during winter. In contrast to winter, a cold bias is obtained in the northeast during spring for both RCMs (Fig. 2), although a weak warm bias is still present in the ERA-Interim forcing (Fig. 10). This feature was also presented for RCMs at 0.50° horizontal resolution in Ozturk et al. (2012 and 2016), but they did not explain it. The warm bias during winter and cold bias during spring in the northeastern part of the domain could be due to an incorrect simulation of snow related processes or a delay in the simulation snow cover. This should be further investigated by looking if there is a temperature delay in the annual~~

~~cycle over Eastern Russia and if there is a similar delay in one of the processes. Russo et al. (2019) found, however, for their RCM that changes in the snow scheme did not affect the simulation results significantly and it did not reduce the warm bias in the northeast during winter. Cloud cover is another process that might explain the pronounced temperature biases in the north. Ozturk et al. (2012) obtained significant better temperature results in the northern part of the CAS-CORDEX domain when using a cloud cover correction. Hamdi et al. (2012) found a strong correlation between a warm bias and cloud cover representation over Belgium (Europe) for ALARO-0, so this could be the reason why there are some large temperature biases in the north especially for ALARO-0. From Fig. 4 and 6 was deduced that ALARO-0 overestimates the nocturnal winter temperatures, while the diurnal temperatures in spring are underestimated. Both could be due to too much cloud cover and this could explain as well why the RCMs underestimate the diurnal range. Therefore, the relation between the temperature biases in the north and the cloud cover should be further investigated by studying this specific region more comprehensively. Another possibility is that the RCMs calculate the temperature incorrectly during stable circumstances.~~

~~New et al. (1999) found that CRU overestimates the temperatures in summer over Russia. When ERA-Interim and MW are both compared to CRU, then it is seen that these two datasets contain lower temperatures over Western Russia during all seasons except for winter (Fig. 10) and thus, the weak cold bias over Western Russia during these seasons for both RCMs (Fig. 2) can be attributed to temperatures in the CRU dataset that are too warm. However, the cold bias in the northwest during winter for REMO cannot be explained by this feature since a small warm bias is found between ERA-Interim and CRU during~~

~~winter. As mentioned before this feature was already described by Pietikäinen et al. (2018). The north-south gradient in the~~



~~Figure 10: CRU temperature (left), difference between ERA-Interim temperature and CRU temperature (middle) and difference between MW temperature and CRU temperature (right) for the winter (DJF), spring (MAM), summer (JJA), autumn (SON) and annual (ANN) mean of the 1980 - 2017 period over the CAS-CORDEX domain.~~

Formatted: Normal

~~temperature bias over the Arabian Peninsula during spring and summer for both RCMs (Fig. 2) can be explained by a sparse coverage of observational stations in the CRU dataset over this region (New et al., 1999), since for both ERA-Interim and MW a similar bias is found with respect to CRU (Fig. 10). The warm bias over Pakistan and Northern India, present for both RCMs during spring, summer and autumn, cannot be explained by the ERA-Interim forcing or differences between the observational datasets and thus a process in the RCMs is likely to overestimate the temperatures in this region. There is as well a significant cold bias between the ERA-Interim and CRU data over the Himalayas and Tibetan Plateau during the different seasons. The latter was also observed by Ozturk et al. (2012 and 2016) and is due to the fact that gridded data is based on measurements of meteorological stations in the valleys (New et al., 1999). This is the case as well for the gridded observational data of WM. Similar as Russo et al. (2019) concluded for COSMO-CLM 5.0, the cold bias of the RCMs over the Himalayas and Tibetan Plateau is mainly due to the gridded observations that are less reliable. The amplification of the biases over the mountainous regions for the RCMs can be attributed to the used assumption of the spatially and temporally uniform lapse rate of 0.0064 K m~~$^{-1}$ ~~for the elevation correction (Kotlarski et al., 2014) or by an amplification induced by the RCMs.~~

~~There is the same tendency as mentioned by Kotlarski et al. (2014) for the European domain that the RCMs underestimate the spatial variation slightly during winter and overestimate it during. In Fig. 3, it is seen that the larger RSVs of the RCMs during summer are due to an underestimation of the variability in the CRU dataset since the ERA-Interim and MW data show both a slight overestimation compared to CRU. In addition it is seen that the spatial patterns during summer are not completely captured by the CRU data since, the two other reference datasets both show a lower spatial correlation with CRU during summer, compared to the other seasons. The lower performance of the RCMs during summer can thus be explained by the uncertainty in spatial variation of temperatures within the observational CRU dataset. As mentioned before this is more pronounced for the summer season since the spatial variation in temperature is lower during this season. Ozturk et al. (2016) found as well a lower spatial correlation during summer with their RCM RegCM4.3.5 at 0.50° horizontal resolution. Additionally, similar high spatial correlations are obtained during the different seasons for ALARO-0 and REMO at 0.22° horizontal resolution when compared to the results of Ozturk et al. (2016).~~

4.3~~2~~ Precipitation

The precipitation of ALARO-0 and REMO is for the majority of the grid points situated within the spread of the different gridded datasets during the different seasons (Fig. 9). However, there are some subregions where the precipitation of ALARO-0 and/or REMO exceeds the observational spread for a specific season. For example, both RCMs show slightly lower precipitation amounts in summer over West Central Asia compared to the different reference datasets (Fig. 11). Ozturk et al. (2012) and Russo et al. (2019) obtained similar seasonal patterns in precipitation, with their model simulations at a horizontal resolution of 0.50° and 0.22°, respectively. They also obtained a dry bias in summer over the north-western and south-western part of the domain. Additionally, an excess of precipitation was simulated over the mountainous areas of the Asian monsoon region during winter, spring and autumn, while in summer a dry bias was observed in mountainous areas except for some parts of the Tibetan Plateau (Ozturk et al., 2012; Russo et al., 2019).

ALARO-0 and REMO produce smaller spatially averaged precipitation biases over the CAS-CORDEX region at a horizontal resolution of 0.22° than the RegCM4.3.5 model at a resolution of 0.50°, except during summer (Ozturk et al., 2016). The spatial correlations between CRU and REMO are similar to the values obtained with RegCM4.3.5, except for winter where REMO has a higher spatial correlation (Fig. 10). ALARO-0 obtains higher values for the spatial correlations and they are close to those of the other observational datasets.

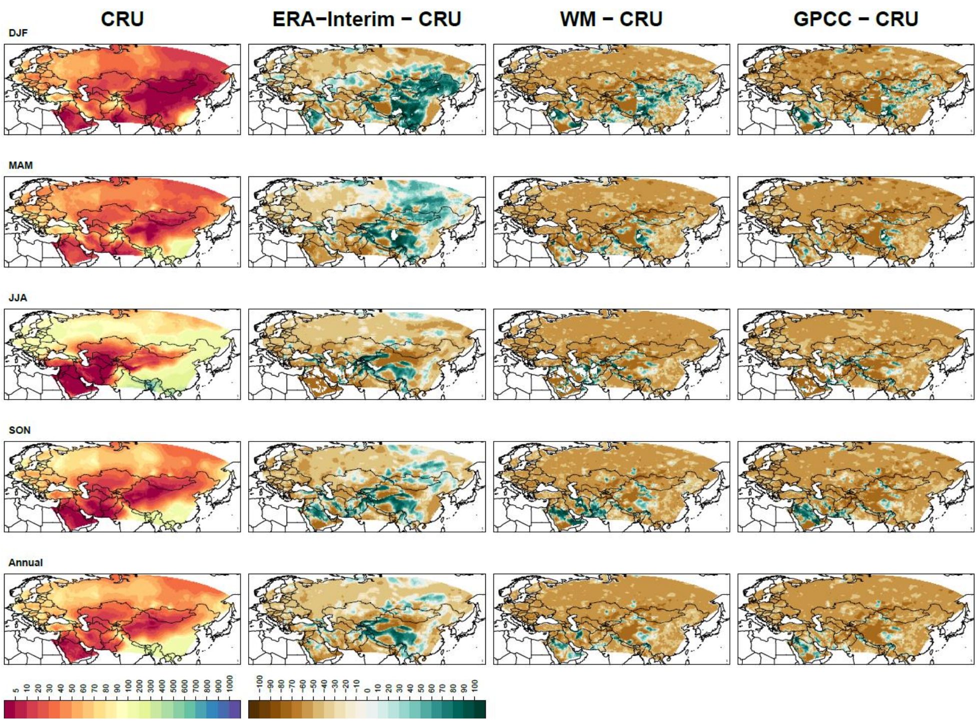
The overestimation of precipitation by the RCMs over the Himalaya, Altay, Tian Shan and Kunlun Mountains on annual level is partly due to the fact that gridded observational datasets CRU, MW and GPCC underestimate the precipitation over these mountainous regions. It is a known feature that the accuracy of gridded precipitation datasets decreases with elevation, especially when the altitude of 1500 m is reached (Zhu et al., 2015). This explains as well why the gridded observational datasets show a drier environment than the ERA-Interim reanalysis dataset in the eastern part of the domain (East Siberia and Tibetan Plateau), particularly during spring (Fig. 11) (Hu et al., 2018). Moreover, this pronounced difference during spring between the observational gridded datasets on the one hand and the RCMs and ERA-Interim reanalysis data on the other hand explains why the scores with respect to CRU are worse during spring (Fig. 10). This difference between the observational and reanalysis datasets makes it difficult to draw sound conclusions over the south-eastern part of the domain during spring, when the monsoon takes place.

It is known that CRU data shows higher precipitation rates at most of the grid points in eastern Russia due to poor station coverage (New et al., 1999). This overestimation of precipitation in the CRU data causes a larger spread in variability, which explains why the RCMs underestimated the spatial variation only during summer (Fig. 10). When averaging over the complete domain, then the output of both ALARO-0 and REMO is within the range of the spread between the reference datasets for the different seasons (Table 5).

Table 5 and Fig. 11 show that CRU contains higher precipitation amounts compared to the two other observational datasets, MW and GPCC. This explains the systematic dry bias that was found for ALARO-0 during all seasons when compared to CRU (Table 5). The underestimation in precipitation by ALARO-0 during spring in the north-eastern part of the domain might be related to the Siberian High that remains too strong during spring (not shown). REMO simulated wetter circumstances with respect to all reference datasets over East Siberia during spring and over the Tibetan Plateau during all seasons except for summer (Fig. 11). The wet bias over East Siberia during spring is in absolute values very low when compared to the subdomain Tibetan Plateau (Fig. 11 and S2). Russo et al. (2019) found a similar spatial pattern of a wet bias during winter (autumn and spring were not discussed) over the south-eastern region with their COSMO-CLM model as presented here for REMO (Fig. 9).

The precipitation amounts of REMO tend in the north more towards those of ERA-Interim (Fig. 11). The similarities between ERA-Interim and REMO for precipitation are probably due to the fact that both use a modified convection scheme that is based on Tiedtke (1989) (Table S1; www.ecmwf.int, consulted on 07/07/2020), while ALARO-0 uses the 3MT cloud microphysics scheme and shows a different behaviour. For example, the weak wet bias which was observed in the north-eastern part of the domain during spring for REMO and not for ALARO-0 is also visible in the ERA-Interim data, but not in the MW and GPCC data (Fig. 11). Additionally, REMO is not able to reproduce the annual cycle of precipitation over the Asian monsoon region. Remedio et al. (2019) found as well a shift in precipitation for REMO over the subtropical region where the Asian monsoon takes place with wetter winter and spring seasons and a drier summer season.

It can be concluded that REMO and ALARO-0 simulated precipitation for the different subregions and seasons mostly within the range of the observational spread, although it should be mentioned that the observational uncertainty is large. MW, GPCC and ERA-Interim deviate more from CRU than it was the case for temperature, resulting in a larger observational uncertainty for precipitation. Russo et al. (2019) showed additionally that the influence of observational datasets on the RSV is larger for precipitation than for temperature. Moreover, both models are worse in simulating the spatial correlation of precipitation (Fig. 10) compared to the mean, minimum and maximum temperature (Fig. 3, 6 and 8). The lower accuracy of simulated precipitation is due to the fact that precipitation is less systematically affected by land cover and topography compared to temperature (Kotlarski et al., 2014). Furthermore, the uncertainty range and error in the observational products should be restricted in the future to improve the evaluation of precipitation (Russo et al., 2019).~~Table 5 and Fig. 11 show that CRU overestimates the precipitation amounts since the two other observational datasets, MW and GPCC, have a strong dry bias over almost the complete domain when compared to CRU. This can explain the systematic dry bias that was found for ALARO-0 during all seasons (Table 5). The small patches with wet biases in the southeastern part of the domain for these two gridded datasets, however, do not explain the extensive wet bias in the southeast during winter and spring which was observed for both RCMs (Fig. 8). This wet bias is present in the ERA-Interim data (Fig. 11) and thus the RCMs might produce this wet bias due to an overestimation in specific humidity of the ERA-Interim forcing, although ALARO-0 is able to reduce the excessive amount of precipitation visible in the ERA-Interim data to a certain extent (Fig. 8, 11, S1 and S2). The biases of REMO tend more towards those of ERA-Interim, although REMO and ERA-Interim parameterize precipitation differently, and the biases of ALARO-0 tend more towards those of MW and GPCC. For example, the weak wet bias which was observed in the northeastern part of the domain during spring for REMO and not for ALARO-0 is also visible in the ERA-Interim data, but not in the WM and GPCC data. This difference between ALARO-0 and REMO is related to the 3MT cloud microphysics scheme of ALARO-0, which is known for its good performance (Giot et al., 2016). Another similarity between the ERA-Interim data and the output of the RCMs is seen in Fig. 9, where both RCMs are worst in simulating the spatial variation during spring and the ERA-Interim data has a similar overestimation in spatial variation during spring when compared to CRU. Except for ERA-Interim in spring, the other reference datasets have a lower spatial variation in precipitation during all seasons, which means that CRU generally overestimates the spatial variation in precipitation (Fig. 9).~~



~~Figure 11: Relative difference between the average seasonal and annual CRU precipitation (mm month^{-1}) and the precipitation in the ERA-Interim, MW and GPCC datasets over the 1980-2017 period.~~

Formatted: Normal

~~In Table 5 and Fig. 11, it is seen that the gridded observational datasets, CRU, MW and GPCC, show a drier environment than the ERA-Interim reanalysis dataset during spring, which is a known feature (Hu et al., 2018). The observed relative wet bias in the east for the ERA-Interim data during winter is in absolute values not that outspoken as the wet bias in spring, which is due to the low precipitation quantities over this region during winter, as was mentioned before (Fig. 11 and S2). The wet bias in winter for ERA-Interim is not reflected by a positive value for the spatial mean bias in Table 5, since it is completely compensated by a dry bias in the northwestern part of the domain (Table 5 and Fig. 11 and S2). The dry bias in the southwest of the domain during spring and summer, which was observed for both RCMs, is also seen for the ERA-Interim, MW and GPCC data. From this is concluded that this dry bias is due to a small overestimation in precipitation by CRU which leads to large relative biases since the precipitation quantities are low (Fig. 13). Harris et al. (2013) mentioned that the Middle East is sparsely covered with precipitation measurements, which leads to uncertainties and errors in the CRU data. Both RCMs have~~

~~the driest spatially mean bias compared to CRU in summer due to a dry bias over Russia and Northern India (Table 5 and Fig. 8 and S1). Similar patterns are found for the observational datasets MW and GPCC when looking to the absolute differences with CRU (Fig. S2). It is known that CRU data shows higher precipitation rates at most of the grid points in eastern Russia due to poor station coverage (New et al., 1999). The dry biases over this regions during summer for the RCMs and observational datasets MW and GPCC are thus due to an overestimation of precipitation in the CRU data (Fig. 8, 11, S1 and S2). This overestimation of precipitation in the CRU data causes a larger spread in variability, which explains why the RCMs underestimated the spatial variation only during summer (Fig. 9). The overestimation in precipitation by both RCMs over the eastern part of Tibetan Plateau and the Altay, Tianshan and Kunlun Mountains on the annual level is according to Zhu et al. (2015) due to the fact that that gridded datasets underestimate the precipitation over these mountainous regions. It is a known feature that the accuracy of gridded precipitation datasets decreases with elevation, especially when the altitude of 1500 m is reached (Zhu et al., 2015). Table 5, Fig. 9 and Fig. 11 show that the observational gridded datasets and ERA-Interim deviate more from CRU than it was the case for temperature, resulting in a larger observational uncertainty for precipitation. Russo et al. (2019) showed additionally that the influence of observational data sets on the RSV is larger for precipitation than for temperature.~~

~~Ozturk et al. (2012 and 2016) and Russo et al. (2019) obtained similar seasonal patterns in precipitation, with their model simulations at a horizontal resolution of 0.50° and 0.22°, respectively. An excess of precipitation was simulated over the mountainous areas of the Asian monsoon region during winter, spring and autumn, while in summer a dry bias was observed. Additionally, they obtained as well a dry bias in summer over the northwestern and southwestern part of the domain. The ALARO-0 and REMO models produce at a horizontal resolution of 0.22° smaller spatially averaged precipitation biases over the CAS-CORDEX region than was obtained with the RegCM4.3.5 model at a resolution of 0.50° (Ozturk et al., 2016). ALARO-0 and REMO have similar values for spatial correlations of precipitation (Fig. 9) as for regions in the EURO-CORDEX domain which range between 40 % and 90 % (Kotlarski et al., 2014). The spatial correlations between CRU and REMO are similar to the values obtained with RegCM4.3.5, except for winter where REMO has a higher spatial correlation. ALARO-0 obtains higher values for the spatial correlations and they are close to those of the other observational datasets (Fig. 9). Although the observational uncertainty is quite large, we can conclude that REMO simulates the precipitation fairly well and ALARO-0 performs very well. However, the uncertainty range and error in the observational products should be restricted to improve the evaluation of precipitation.~~

~~The warm temperatures obtained with REMO in winter and the cold temperatures in spring over the northeastern part of the domain can be linked with the dry and wet bias in winter and spring respectively. This strengthens our hypothesis that there is a delay by REMO in simulating snow or snow cover. As stated before for temperature this should be further analysed by plotting the annual cycles of precipitation and temperature for this region. For ALARO-0 this link between an overestimation (underestimation) of temperature and an underestimation (overestimation) in precipitation during winter (spring) is not seen. Therefore, it is likely that some processes affecting the temperature are not simulated well by ALARO-0 over the northeastern~~

~~part of the domain. The persistent warm bias over Pakistan and Northern India of both RCMs can be explained by the persistent underestimation in simulated precipitation over this region by both RCMs.~~

~~When we compare the results of temperature (Fig. 3) and precipitation (Fig. 9) with other domains e.g. Fig. 9 and 10 in Kotlarski et al. (2014), then we can conclude that these RCMs have a similar model performance as the RCMs have over other domains. However, one should be aware that the CAS-CORDEX domain as a whole is a larger domain and thus the result might be more smoothed because of the larger amount of grid points which was taken into account to create the Taylor diagram.~~

~~4.3 Outlook~~

Formatted: Normal

~~In the near future a similar evaluation over several subregions will be undertaken, since this evaluation over the large domain highlighted some specific regions where there might be deficiencies in the RCMs e.g. the warm bias in winter over Eastern Russia and wet bias over the East Asian monsoon region. By looking into more detail to subregions we hope to understand which processes in the RCMs cause the deficiencies e.g. shift in snow related processes and monsoon. In addition, we ran both RCMs up to 2100 driven by different GCMs under the scenarios of representative concentration pathways (RCPs) 2.6, 4.5 (only for ALARO) and 8.5, which will be used to investigate the climate sensitivity over Central Asia and to study the evolution of extreme events. Further, we plan to perform a bias adjustment on the model data by using observations. To select the optimal bias adjustment method, a comparison of different approaches will be made. This will enable impact modellers to optimally use our climate data in their models for crop production, biomass production, etc.~~

5 Conclusion

The evaluation over the CAS-CORDEX domain of ALARO-0 and REMO, run at 0.22° resolution, showed that both RCMs reproduced in general realistic spatial patterns for temperature and precipitation. Both RCMs perform best during autumn, showing biases within the range of observational uncertainty for temperature and precipitation. Nevertheless, there are significant biases in several regions during several seasons e.g. a warm bias in the north during winter and a wet bias in spring over the Asian monsoon region. For ALARO-0 the northern part of the CAS-CORDEX domain is subject to significant positive temperature biases in winter, followed by large negative temperature biases in spring. This behaviour is probably linked to limitations of the used snow scheme. The evaluation of minimum and maximum temperatures showed that the RCMs underestimate the daily temperature range. This illustrates the added value of taking more evaluation variables into account than only the commonly evaluated variables mean temperature and precipitation.

The values of spatial variation and pattern correlation for mean temperature of both RCMs correspond closely to the values obtained with other reference datasets. These metrics indicated a less good performance for precipitation data of the RCMs since they deviated more from the reference datasets than it was the case for temperature. However, the different reference datasets deviated more for precipitation from CRU, than for temperature which indicates that there is a larger uncertainty in the spatial patterns of precipitation.

We conclude that REMO and ALARO-0 can be used to perform climate projections over Central Asia since they perform similarly to experiments with other models over the same domain. REMO is better than ALARO-0 in reproducing the seasonal mean temperatures over the entire domain except during autumn, while ALARO-0 is very good in estimating the precipitation. However, deficiencies described in this evaluation study should be kept in mind. Climate data produced by both RCMs can be used for impact studies if a suitable bias adjustment is applied for those subregions where the RCMs perform less well e.g. East Siberia and Tibetan Plateau.

~~The first validation results over the CAS-CORDEX domain of ALARO-0 and REMO, ran at 0.22° resolution, showed that both RCMs reproduced realistic spatial patterns for temperature and precipitation with biases within an acceptable range, except for the temperature of ALARO-0 during spring. However, there are large biases in several regions during several seasons e.g. a warm bias in the north during winter and a wet bias in spring over the Asian monsoon region. The comparison between CRU, ERA-Interim and the other gridded observational datasets showed that the warm bias in winter is induced by the warm ERA-Interim forcing and a delay in the simulation of snow and snow cover for REMO. For ALARO-0 the temperature delay could not be explained by a delay in precipitation and thus it is likely that some processes which affect the temperature in this region are not captured well by ALARO-0. A similar validation over subregions should be done, to examine the shift in the annual cycle and the processes that are lacking or simulated incorrectly over those particular regions where a less good performance was found. Negative precipitation biases for both RCMs during all seasons are due to an overestimation of precipitation in the CRU data since the other reference datasets show dry biases. For all variables large biases are observed over the mountainous areas but these are mainly attributed to the observational error.~~

~~Both RCMs perform very well during the autumn, showing biases within the range of observational uncertainty for temperature and precipitation. Additionally, the values for spatial variation and pattern correlation of both RCMs are very close to the values obtained with other reference datasets for the mean temperature. For precipitation these metrics indicated a less good performance of the RCMs since they deviated more from the reference datasets than it was the case for temperature. However, the different reference datasets deviated more for precipitation from CRU, than for temperature which indicates that there is a larger uncertainty in the spatial patterns of precipitation. The precipitation biases of both RCMs are within the range of observational uncertainty and the precipitation is simulated similar for REMO or better for ALARO-0 when compared to other CORDEX simulations. REMO is better than ALARO-0 in reproducing the temperatures based on the biases and spatial variability, except during autumn, while ALARO-0 is very good in estimating the precipitation.~~

~~The evaluation of minimum and maximum temperatures showed that the RCMs simulate these variables less pronounced over most of the domain compared to the observational CRU dataset, which is generally caused by an underestimation of the daily temperature range. This shows the advantage of taking more evaluation variables into account than only the ordinary mean temperature and precipitation. These findings are important for regional impact modelling. Since the RCMs perform as well over the CAS-CORDEX domain as other RCMs do, we finally conclude that these RCMs can be used to perform climate projections and the produced climate data can be applied in impact modelling.~~

Code availability

The R code used for the analysis is available through: http://doi.org/10.5281/zenodo.3659717 (Top et al., 2020). For the code of the ALARO-0 model we refer to the Code availability section in Termonia et al. (2018). More information about the REMO model is available on request by contacting the Climate Service Center Germany (contact@remo-rcm.de).

Data availability

The climate data produced by ALARO-0 and REMO2015 have been uploaded to the ESGF data nodes (website: http://esgf.llnl.gov/). In order to obtain the data, one of the nodes must be chosen. Thereafter, click on 'CORDEX' or search for 'CORDEX' and then select the domain 'CAS-22' and the RCM model in the left column. The exact identifiers can be found in Table S2 of the supplementary material.
The CRU data is available through (http://www.cru.uea.ac.uk). The MW data is freely available at: http://climate.geog.udel.edu/~climate/html_pages/download.html and NetCDF files can be found here: https://www.esrl.noaa.gov/psd/data/gridded/data.UDel_AirT_Precip.html: air.mon.mean.v501.nc and precip.mon.total.v501.nc. The GPCC data can be accessed through: doi: 10.5676/DWD_GPCC/FD_M_V2018_025.

Author contribution

Modelling and performing simulations: C.S., D.C.L., D.T.R., K.L., K.A, R.R.A. ;Post-processing: D.C.L., D.T.R., K.L., K.A, R.R.A. ;Visualization: K.L., T.S. ;Writing - original draft: T.S. ;Writing - review & editing: A.S., B.L., C.S., D.C.L., D.M.P., D.T.R., G.N., G.A., H.R., K.A, K.L., R.R.A., S.A., T.P., T.S., V.D.V.H., V.S.B., Z.V. ;Supervision: C.S., D.M.P., T.P. ;Funding acquisition: A.S., B.L., D.M.P., G.A., K.L., T.P.

Acknowledgement

The AFTER project is granted by the ERA.Net RUS Plus Initiative, ID 166. HZG-GERICS received funding from the Federal Ministry of Education and Research (BMBF). Ghent University and VITO received funding from the Research Foundation - Flanders (FWO), grant G0H6117N. NIERSC received funding from the Russian Foundation for Basic Research (RFBR) - grant № 18-55-76004. LEGMC received funding from the State Education Development Agency (SEDA) and ISTE received funding from the scientific and technological research council of Turkey (TUBITAK agreement nr: 2017O394).
The computational resources and services for the ALARO-0 regional climate simulations were provided by the Flemish Supercomputer Center (VSC), funded by the Research Foundation - Flanders (FWO) and the Flemish Government department EWI. The CORDEX-CORE REMO simulations were performed under the GERICS/HZG share at the German Climate Computing Centre (DKRZ).

We would like to thank Ján Mašek for his insights on the warm bias above snow cover.

Competing interests

The authors declare that they have no conflict of interest.

References

Akperov, M., Rinke, A., Mokhov, I.I., Matthes, H., Semenov, V.A., Adakudlu, M., Cassano, J., Christensen, J.H., Dembitskaya, M.A., Dethloff, K. and Fettweis, X.: Cyclone activity in the Arctic from an ensemble of regional climate models (Arctic CORDEX). Journal of Geophysical Research: Atmospheres, 123, 2537-2554, https://doi.org/10.1002/2017JD027703, 2018.

ALADIN International Team: The ALADIN project: Mesoscale modelling seen as a basic tool for weather forecasting and atmospheric research, WMO bull., 46, 317–324, 1997.

Almazroui, M., Islam, M. N., Alkhalaf, A. K., Saeed, F., Dambul, R. and Rahman, M. A.: Simulation of temperature and precipitation climatology for the CORDEX-MENA/Arab domain using RegCM4, Arab. J. of Geosci., 9, 13, doi:10.1007/s12517-015-2045-7, 2016.

Bucchignani, E., Mercogliano, P., Panitz, H. J. and Montesarchio, M.: Climate change projections for the Middle East–North Africa domain with COSMO-CLM at different spatial resolutions, Advances in Climate Change Research, 9, 66–80, doi:10.1016/j.accre.2018.01.004, 2018.

Collins, M., AchutaRao, K., Ashok, K., Bhandari, S., Mitra, A.K., Prakash, S., Srivastava, R. and Turner, A.: Observational challenges in evaluating climate models, Nature Climate Change, 3, 940–941. https://doi.org/10.1038/nclimate2012, 2013.

CORDEX Scientific Advisory Team: The WCRP CORDEX Coordinated Output for Regional Evaluations (CORE) Experiment Guidelines, Available online: http://www.cordex.org/experiment-guidelines/cordex-core (accessed on 1 March 2019).

Cabos, W., Sein, D.V., Durán-Quesada, A., Liguori, G., Koldunov, N.V., Martínez-López, B., Alvarez, F., Sieck, K., Limareva, N. and Pinto, J.G.: Dynamical downscaling of historical climate over CORDEX Central America domain with a regionally coupled atmosphere–ocean model. Climate dynamics, 52, 4305-4328, https://doi.org/10.1007/s00382-018-4381-2, 2019.

Davies, H.C.: A lateral boundary formulation for multi-level prediction models, Quart. J. R. Meteor. Soc., 102, 405–418, doi:10.1002/qj.49710243210, 1976.

Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M.A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A.C., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A.J., Haimberger, L., Healy, S.B., Hersbach, H., Hólm, E.V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., Mcnally, A.P., Monge-Sanz, B.M., Morcrette, J.J., Park, B.K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.N. and Vitart, F.: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, Quarterly Journal of the Royal Meteorological Society, 137, 553–597, doi:10.1002/qj.828, 2011.

De Troch, R., Hamdi, R., Van de Vyver, H., Geleyn, J. F., and Termonia, P.: Multiscale performance of the ALARO-0 model for simulating extreme summer precipitation climatology in Belgium, Journal of Climate, 26, 8895–8915, doi:10.1175/JCLI-D-12-00844.1, 2013.

Denis, B., Laprise, R., Caya, D. and Côté, J.: Downscaling ability of one-way nested regional climate models: the Big-Brother Experiment. Climate Dynamics, 18, 627–646, doi:10.1007/s00382-001-0201-0, 2002.

Diaconescu, E. P., Gachon, P., Laprise, R. and Scinocca, J. F.: Evaluation of precipitation indices over North America from various configurations of regional climate models, Atmosphere-Ocean, 54, 418–439, doi:10.1080/07055900.2016.1185005, 2016.

Di Virgilio, G., Evans, J. P., Di Luca, A., Olson, R., Argüeso, D., Kala, J., Andrys, J., Hoffmann, P., Katzfey, J. J. and Rockel, B.: Evaluating reanalysis-driven CORDEX regional climate models over Australia: model performance and errors, Climate Dynamics, 53, 2985–3005, doi:10.1007/s00382-019-04672-w, 2019.

Douville, H., Royer, J-F. and Mahfouf., J-F.: A new snow parameterization for the Meteo-France climate model, Climate Dynamics, 12, 21-35, 1995.

ECMWF: Atmospheric physics, https://www.ecmwf.int/en/research/modelling-and-prediction/atmospheric-physics, (accessed on 7 July 2020).

Fuentes-Franco, R., Coppola, E., Giorgi, F., Pavia, E. G., Diro, G. T. and Graef, F.: Inter-annual variability of precipitation over Southern Mexico and Central America and its relationship to sea surface temperature from a set of future projections from CMIP5 GCMs and RegCM4 CORDEX simulations. Climate Dynamics, 45, 425-440, doi:10.1007/s00382-014-2258-6, 2015.

Gerard, L., Piriou, J. M., Brožková, R., Geleyn, J. F. and Banciu, D.: Cloud and precipitation parameterization in a meso-gamma-scale operational weather prediction model, Monthly Weather Review, 137, 3960–3977, doi:10.1175/2009MWR2750.1, 2009.

Ghimire, S., Choudhary, A. and Dimri, A. P.: Assessment of the performance of CORDEX-South Asia experiments for monsoonal precipitation over the Himalayan region during present climate: part I, Climate dynamics, 50, 2311–2334, doi:10.1007/s00382-015-2747-2, 2018.

Gibson, P.B., Waliser, D.E., Lee, H., Tian, B. and Massoud, E.: Climate model evaluation in the presence of observational uncertainty: precipitation indices over the Contiguous United States. Journal of Hydrometeorology, 20,1339-1357, 2019.

Giorgetta, M. and Wild, M.: The water vapor continuum and its representation in ECHAM4, MPI for Meterolo., report no. 162, Hamburg, 1995.

~~Giorgetta, M. and Wild, M.: The water vapor continuum and its representation in ECHAM4, MPI for Meterolo., report no. 162, Hamburg, 1995.~~

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

Giorgi, F. and Gutowski Jr, W. J.: Regional dynamical downscaling and the CORDEX initiative, Annual Review of Environment and Resources, 40, 467–490, 2015.

Giorgi~~o~~, F. and Mearns, L. O.: Introduction to special section: Regional climate modeling revisited, J. Geophys. Res., 104, 6335–6352, doi:10.1029/98JD02072, 1999.

Giot, O., Termonia, P., Degrauwe, D., De Troch, R., Caluwaerts, S., Smet, G., Berckmans, J., Deckmyn, A., De Cruz, L., De Meutter, P., Duerinckx, A., Gerard, L., Hamdi, R., Van den Bergh, J., Van Ginderachter, M. and Van Schaeybroeck, B.: Validation of the ALARO-0 model within the EURO-CORDEX framework, Geosci. Model Dev., 9, 1143–1152, doi:10.5194/gmd-9-1143-2016, 2016.

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?, Geophys. Res. Lett., 39, L24701, https://doi.org/10.1029/2012GL054206, 2012.

Gordon, C., Cooper, C., Senior, C. A., Banks, H., Gregory, J. M., Johns, T. C., Mitchell, J. F. B. and Wood, R. A.: The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments, Climate dynamics, 16, 147–168, doi:10.1007/s003820050010, 2000.

Gutowski Jr., W. J., Giorgi, F., Timbal, B., Frigon, A., Jacob, D., Kang, H.-S., Raghavan, K., Lee, B., Lennard, C., Nikulin, G., O'Rourke, E., Rixen, M., Solman, S., Stephenson, T., and Tangang, F.: WCRP COordinated Regional Downscaling EXperiment (CORDEX): a diagnostic MIP for CMIP6, Geosci. Model Dev., 9, 4087–4095, doi:10.5194/gmd-9-4087-2016, 2016.

Haarsma, R. J., Roberts, M. J., Vidale, P. L., Senior, C. A., Bellucci, A., Bao, Q., Chang, P., Corti, S., Fučkar, N. S., Guemas, V., von Hardenberg, J., Hazeleger, W., Kodama, C., Koenigk, T., Leung, L. R., Lu, J., Luo, J.-J., Mao, J., Mizielinski, M. S., Mizuta, R., Nobre, P., Satoh, M., Scoccimarro, E., Semmler, T., Small, J., and von Storch, J.-S.: High Resolution Model Intercomparison Project (HighResMIP v1.0) for CMIP6, Geosci. Model Dev., 9, 4185-4208, doi:10.5194/gmd-9-4185-2016, 2016.

Hagemann, S.: An improved land surface parameter data set for global and regional climate models, Max Planck Institute for Meteorology report series, Hamburg, Germany, Report No. 336, 2002.

Hamdi, R., Van de Vyver, H. and Termonia, P.: New cloud and microphysics parameterisation for use in high-resolution dynamical downscaling: application for summer extreme temperature over Belgium, Int. J. Climatol., 32, 2051–2065, doi:10.1002/joc.2409, 2012.

Harris, I., Osborn, T.J., Jones, P.D. and Lister, D.H.: Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset, Scientific Data, 7 (109), doi:10.1038/s41597-020-0453-3, 2020.

~~Harris, I., Jones, P.D., Osborn, T.J. and Lister, D.H.: Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 Dataset, Int. J. Climatol., 34, 623–642, doi:10.1002/joc.3711, 2014.~~

Hofstra, N., Haylock, M., New, M. and Jones, P. D.: Testing E-OBS European high-resolution gridded data set of daily precipitation and surface temperature, Journal of Geophysical Research: Atmospheres, 114, doi:10.1029/2009JD011799, 2009.

Hofstra, N., New, M. and McSweeney, C.: The influence of interpolation and station network density on the distributions and trends of climate variables in gridded daily data, Climate dynamics, 35, 841–858, doi:10.1007/s00382-009-0698-1, 2010.

Hu, Z., Zhou, Q., Chen, X., Li, J., Li, Q., Chen, D., Liu, W. and Yin, G.: Evaluation of three global gridded precipitation data sets in central Asia based on rain gauge observations, International Journal of Climatology, 38, 3475–3493, doi:10.1002/joc.5510, 2018.

Iturbide, M., Gutiérrez, J. M., Alves, L. M., Bedia, J., Cimadevilla, E., Cofiño, A. S., Cerezo-Mota, R., Di Luca, A., Faria, S. H., Gorodetskaya, I., Hauser, M., Herrera, S., Hewitt, H. T., Hennessy, K. J., Jones, R. G., Krakovska, S., Manzanas, R., Marínez-Castro, D., Narisma, G. T., Nurhati, I. S., Pinto, I., Seneviratne, S. I., van den Hurk, B., and Vera, C. S.: An update of IPCC climate reference regions for subcontinental analysis of climate model data: Definition and aggregated datasets, Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2019-258, in review, 2020.

Jacob, D.: A note to the simulation of the annual and inter-annual variability of the water budget over the Baltic Sea drainage basin. Meteorol. Atmos. Phys., 77, 61–73, doi:10.1007/s007030170017, 2001.

Jacob, D., Bärring, L., Christensen, O. B., Christensen, J. H., De Castro, M., Déqué, M., Giorgi, F., Hagemann, S., Hirschi, M., Jones, R., Kjellström, E. Lenderink, G., Rockel, F., Sánchez, E., Schär, C., Seneviratne, S. I., Somot, S., van Ulden, A. and van den Hurk, B.: An inter-comparison of regional climate models for Europe: model performance in present-day climate, Climatic change, 81, 31–52, doi:10.1007/s10584-006-9213-4, 2007.

Jacob, D., Elizalde, A., Haensler, A., Hagemann, S., Kumar, P., Podzun, R., Rechid, D., Remedio, A. R., Saeed, F., Sieck, K., Teichmann, C. and Wilhelm, C.: Assessing the transferability of the regional climate model REMO to different coordinated regional climate downscaling experiment (CORDEX) regions, Atmosphere, 3, 181-199, doi:10.3390/atmos3010181, 2012.

Jacob, D., Petersen, J., Eggert, B., Alias, A., Christensen, O. B., Bouwer, L. M., Braun, A., Colette, A., Déqué, M., Georgievski, G., Georgopoulou, E., Gobiet, A., Menut, L., Nikulin, G., Haensler, A., Hempelmann, N., Jones, C., Keuler, K., Kovats, S., Kröner, N., Kotlarski, S., Kriegsmann, A., Martin, E., van Meijgaard, E., Moseley, C., Pfeifer, S., Preuschmann, S., Radermacher, C., Radtke, K., Rechid, D., Rounsevell, M., Samuelsson, P., Somot, S., Soussana, J.-F., Teichmann, C., Valentini, R., Vautard, R., Weber, B. and Yiou, P.: EURO-CORDEX: new high-resolution climate change projections for European impact research, Regional environmental change, 14, 563–578, doi:10.1007/s10113-013-0499-2, 2014.

~~Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., Zimmermann, N. E., Linder, H. P. and Kessler, M.: Climatologies at high resolution for the earth's land surface areas. Scientific data, 4, 170122, doi:10.1038/sdata.2017.122, 2017.~~

Kotlarski, S.: A subgrid glacier parameterisation for use in regional climate modelling, PhD thesis, Reports on Earth System Science, Max Planck Institute for Meteorology, Hamburg, 2007.

Kotlarski, S., Keuler, K., Christensen, O. B., Colette, A., Déqué, M., Gobiet, A., Goergen, K., Jacob, D., Lüthi, D., van Meijgaard, E., Nikulin, G., Schär, C., Teichmann, C., Vautard, R., Warrach-Sagi, K., and Wulfmeyer, V.: Regional climate modeling on European scales: a joint standard evaluation of the EURO-CORDEX RCM ensemble, Geosci. Model Dev., 7, 1297–1333, doi:10.5194/gmd-7-1297-2014, 2014.

Kotova L., Aniskevich S., Bobylev L., Caluwaerts S., De Cruz L., De Troch R., Gnatiuk N., Gobin A., Hamdi R., Sakalli A., Sirin A., Termonia P., Top S., Van Schaeybroeck B. and Viksna A.: A new project AFTER investigates the impacts of climate change in the Europe-Russia-Turkey region, Climate Services, 12, 64–66. doi:10.1016/j.cliser.2018.11.003, 2018.

Koenigk, T., Berg, P. and Döscher, R.: Arctic climate change in an ensemble of regional CORDEX simulations, Polar Research, 34, 24603, doi:10.3402/polar.v34.24603, 2015.

Kyselý, 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, doi:10.1007/s00382-011-1200-4, 2012.

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, doi: 10.1007/s00382-003-0342-4, 2003.

Lohmann, U. and Roeckner, E.: Design and performance of a new cloud microphysics scheme developed for the ECHAM general circulation model. Climate Dynamics, 12, 557–572, doi:10.1007/BF00207939, 1996.

Mašek, J.: Problem with screen level temperatures above snow in ISBA scheme, report RC LACE, 2017.

Morcrette, J.-J., Smith, L. and Fouquart, Y.: Pressure and temperature dependence of the absorption in longwave radiation parameterizations, Atmos. Phys., 59, 455–469, 1986.

New, M., Hulme, M. and Jones, P.: Representing twentieth-century space–time climate variability. Part I: Development of a 1961–90 mean monthly terrestrial climatology, Journal of climate, 12, 829–856, doi:10.1175/1520-0442(1999)012<0829:RTCSTC>2.0.CO;2, 1999.

~~New, M., Hulme, M., and Jones, P.: Representing twentieth-century space–time climate variability. Part II: Development of 1901–96 monthly grids of terrestrial surface climate, Journal of climate, 13, 2217–2238, doi:10.1175/1520-0442(2000)013<2217:RTCSTC>2.0.CO;2, 2000.~~

Nikulin, G., Jones, C., Giorgi, F., Asrar, G., Büchner, M., Cerezo-Mota, R., Christensenf, O. B., Déquég, M., Fernandezh, J., Hänsleri, A., van Meijgaardj, E., Samuelssona, P., Syllab, M. B. and Sushamak, L.: Precipitation climatology in an ensemble of CORDEX-Africa regional climate simulations, Journal of Climate, 25, 6057–6078, doi:10.1175/JCLI-D-11-00375.1, 2012.

Nikulin, G., Lennard, C., Dosio, A., Kjellström, E., Chen, Y., Hänsler, A., Kupiainen, M., Laprise, R., Mariotti, L., Fox Maule, C., van Meijgaard, E., Panitz, H.-J., Scinocca, J. F. and Somot, S.: The effects of 1.5 and 2 degrees of global warming on Africa in the CORDEX ensemble. Environ. Res. Lett., 13, 065003, doi:10.1088/1748-9326/aab1b1, 2018.

Ozturk, T., Altinsoy, H., Türkeș, M. and Kurnaz, M. L.: Simulation of temperature and precipitation climatology for the Central Asia CORDEX domain using RegCM 4.0, Climate Research, 52, 63–76, doi:10.3354/cr01082, 2012.

Ozturk, T., Turp, M. T., Türkeș, M., and Kurnaz, M. L.: Projected changes in temperature and precipitation climatology of Central Asia CORDEX Region 8 by using RegCM4. 3.5, Atmospheric Research, 183, 296–307, doi:10.1016/j.atmosres.2016.09.008, 2016.

Pfeifer, S.: Modeling cold cloud processes with the regional climate model REMO, PhD thesis, Reports on Earth System Science, Max Planck Institute for Meteorology, Hamburg, 2006.

Pietikäinen, J.-P., O'Donnell, D., Teichmann, C., Karstens, U., Pfeifer, S., Kazil, J., Podzun, R., Fiedler, S., Kokkola, H., Birmili, W., O'Dowd, C., Baltensperger, U., Weingartner, E., Gehrig, R., Spindler, G., Kulmala, M., Feichter, J., Jacob, D., and Laaksonen, A.: The regional aerosol-climate model REMO-HAM, Geosci. Model Dev., 5, 1323–1339, doi:/10.5194/gmd-5-1323-2012, 2012..

Pietikäinen, J.-P., Markkanen, T., Sieck, K., Jacob, D., Korhonen, J., Räisänen, P., Gao, Y., Ahola, J., Korhonen, H., Laaksonen, A. and Kaurola, J.: The regional climate model REMO (v2015) coupled with the 1-D freshwater lake model FLake (v1): Fenno-Scandinavian climate and lakes, Geosci. Model Dev., 11, 1321–1342, doi:10.5194/gmd-11-1321-2018, 2018.

Rechid, D.: On biogeophysical interactions between vegetation phenology and climate simulated over Europe, PhD thesis, Reports on Earth System Science, Max Planck Institute for Meteorology, Hamburg, 2009.

Remedio, A.R., Teichmann, C., Buntemeyer, L., Sieck, K., Weber, T., Rechid, D., Hoffmann, P., Nam, C., Kotova, L. and Jacob, D.: Evaluation of New CORDEX Simulations Using an Updated Köppen–Trewartha Climate Classification, Atmosphere, 10, 726, doi:10.3390/atmos10110726, 2019.

Roeckner, E., Arpe, K., Bengtsson, L., Christoph, M., Claussen, M., Dümenil, L., Esch, M., Giorgetta, M., Schlese, U. and Schulzweida, U.: The Atmospheric General Circulation Model Echam-4: Model Description and Simulation of the Present Day Climate, Report No. 218, Max-Planck-Institute for Meteorology: Hamburg, Germany, 1996.

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., 12, 5229–5249, doi:/10.5194/gmd-12-5229-2019, 2019.

Ruti, P. M., Somot, S., Giorgi, F., Dubois, C., Flaounas, E., Obermann, A., Dell'Aquila, A., Pisacane, G., Harzallah, A., Lombardi, E., Ahrens, B., Akhtar, N., Alias, A., Arsouze, T., Aznar, R., Bastin, S., Bartholy, J., Béranger, K., Beuvier, J., Bouffies-Cloché, S., Brauch, J., Cabos, W., Calmanti, S., Calvet, J.-C., Carillo, A., Conte, D., Coppola, E., Djurdjevic, V., Drobinski, P., Elizalde-Arellano, A., Gaertner, M., Galàn, P., Gallardo, C., Gualdi, S., Goncalves, M., Jorba, O., Jordà, G., L'Heveder, B., Lebeaupin-Brossier, C., Li, L., Liguori, G., Lionello, P., Maciàs, D., Nabat, P., Önol, B., Raikovic, B., Ramage, K., Sevault, F., Sannino, G., Struglia, M. V., Sanna, A., Torma, C. and Vervatis, V.: MED-CORDEX initiative for Mediterranean climate studies, Bulletin of the American Meteorological Society, 97, 1187–1208, doi:10.1175/BAMS-D-14-00176.1, 2016.

Schneider, U., Becker, A. Finger, P. Meyer-Christoffer, A. and Ziese, M.: GPCC Full Data Monthly Product Version 2018 at 0.25°: Monthly Land-Surface Precipitation from Rain-Gauges built on GTS-based and Historical Data, doi: 10.5676/DWD_GPCC/FD_M_V2018_025, 2018.

Semmler, T., Jacob, D., Schlünzen, K. H. and Podzun, R.: Influence of sea ice treatment in a regional climate model on boundary layer values in the Fram Strait region, Mon. Weather Rev., 132, 985–999, doi:10.1175/1520-0493(2004)132<0985:IOSITI>2.0.CO;2, 2004.

Solman, S. A., Sanchez, E., Samuelsson, P., da Rocha, R. P., Li, L., Marengo, J., Pessacg, N. L., Remedio, A. R. C., Chou, S. C., Berbery, H., Le Treut, H., de Castro, M. and Jacob, D.: 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, doi:10.1007/s00382-013-1667-2, 2013.

Souverijns, N., Gossart, A., Demuzere, M., Lenaerts, J. T. M., Medley, B., Gorodetskaya, I. V., Vanden Broucke, S. and van Lipzig, N. P. M.: A New Regional Climate Model for POLAR-CORDEX: Evaluation of a 30-Year Hindcast with COSMO-CLM2 Over Antarctica, Journal of Geophysical Research: Atmospheres, 124, 1405–1427, doi:10.1029/2018JD028862, 2019.

Tangang, F., Santisirisomboon, J., Juneng, L., Salimun, E., Chung, J., Cruz, F., Ngai, S. T., Ngo-Duc, T., Singhruck, P., Narisma, G., Santisirisomboon, J., Wongsaree, W., Promjirapawat, K., Sukamongkol, Y., Srisawadwong, R., Setsirichok, D., Phan-Van, T., Gunawan, D., Aldrian, E., Nikulin, G. and Yang, H.: Projected future changes in mean precipitation over Thailand based on multi-model regional climate simulations of CORDEX Southeast Asia, Int. J. Climatol., 39, 5413–5436, doi:10.1002/joc.6163, 2019.

Tangang, F., Supari, S., Chung, J. X., Cruz, F., Salimun, E., Ngai, S. T., Juneng, L., Santisirisomboon, J., Santisirisomboon, J., Ngo-Duc, T., Phan-Van, T., Narisma, G., Singhruck, P., Gunawan, D., Aldrian, E., Sopaheluwakan, A., Nikulin, G., Yang, H., Remedio, A.R.C., Sein, D. and Hein-Griggs, D.: Future changes in annual precipitation extremes over Southeast Asia under global warming of 2 C. APN Science Bulletin, 8, 3–8, doi:10.30852/sb.2018.436, 2018.

Taylor, K. E.: Summarizing multiple aspects of model performance in a single diagram, Journal of Geophysical Research: Atmospheres, 106, 7183–7192, doi:10.1029/2000JD900719, 2001.

Termonia, P., Fischer, C., Bazile, E., Bouyssel, F., Brožková, R., Bénard, P., Bochenek, B., Degrauwe, D., Derková, M., El Khatib, R., Hamdi, R., Mašek, J., Pottier, P., Pristov, N., Seity, Y., Smolíková, P., Španiel, O., Tudor, M., Wang, Y., Wittmann, C., and Joly, A.: The ALADIN System and its canonical model configurations AROME CY41T1 and ALARO CY40T1, Geosci. Model Dev., 11, 257–281, doi:10.5194/gmd-11-257-2018, 2018a.

Termonia, P., Van Schaeybroeck, B., De Cruz, L., De Troch, R., Caluwaerts, S., Giot, O., Hamdi, R., Vannitsem, S., Duchêne, F., Willems, P., Tabari, H., Van Uytven, E., Hosseinzadehtalaei, P., Van Lipzig, N., Wouters, H., Vanden Broucke, S., van Ypersele, J.-P., Marbaix, P., Villanueva-Birriel, C., Fettweis, X., Wyard, C., Scholzen, C., Doutreloup, S., De Ridder, K., Gobin, G., Lauwaet, D., Stavrakou, T., Bauwens, M., Müller, J.-F., Luyten, P., Ponsar, S., Van den Eynde, D. and Pottiaux, E.: The CORDEX.be initiative as a foundation for climate services in Belgium, Climate Services, 11, 49–61, doi:10.1016/j.cliser.2018.05.001, 2018b.

Tiedtke, M. (1989). A comprehensive mass flux scheme for cumulus parameterization in large-scale models. Mon. Wea. Rev., 117, 1779-1800.

Top, S., Kotova, L., De Cruz, L., Aniskevich, S., Bobylev, L., De Troch, R., Gnatiuk, N., Gobin, A., Hamdi, R., Kriegsmann, A., Remedio, A. R., Sakalli, A., Van De Vyver, H., Van Schaeybroeck, B., Zandersons, V., De Maeyer, P., Termonia, P. and Caluwaerts, S.: R code validation analysis ALARO-0 and REMO2015 climate data Central Asia Top et al. 2020, Zenodo, doi:10.5281/zenodo.3659717.

Torma, C., Giorgi, F. and Coppola, E.: Added value of regional climate modeling over areas characterized by complex terrain—Precipitation over the Alps, Journal of Geophysical Research: Atmospheres, 120, 3957–3972, doi:10.1002/2014JD022781, 2015.

Tustison, B., Harris, D. and Foufoula-Georgiou, E.: Scale issues in verification of precipitation forecasts. Journal of Geophysical Research: Atmospheres, 106, 11775–11784, doi:10.1029/2001JD900066, 2001.

Tuyet, N. T., Thanh, N. D., and van Tan, P.: Performance of SEACLID/CORDEX-SEA multi-model experiments in simulating temperature and rainfall in Vietnam, Vietnam Journal of Earth Sciences, 41, 374–387, doi:10.15625/0866-7187/41/4/14259, 2019.

Wang, Y., Feng, J., Luo, M., Wang, J. and Yuan, Q.: Uncertainties in simulating Central Asia: sensitivity to physical parameterizations using WRF, International Journal of Climatology, doi:10.1002/joc.6567, 2020.

~~Wang, J. and Kotamarthi, V. R.: High-resolution dynamically downscaled projections of precipitation in the mid and late 21st century over North America, Earth's Future, 3, 268–288, doi:10.1002/2015EF000304, 2015.~~

Whan, K. and Zwiers, F.: The impact of ENSO and the NAO on extreme winter precipitation in North America in observations and regional climate models, Climate Dynamics, 48, 1401–1411, doi:10.1007/s00382-016-3148-x, 2017.

Wilhelm, C., Rechid, D., and Jacob, D.: Interactive coupling of regional atmosphere with biosphere in the new generation regional climate system model REMO-iMOVE, Geosci. Model Dev., 7, 1093–1114, doi:10.5194/gmd-7-1093-2014, 2014.

Willmott, C.J. and Matsuura, K.: Smart interpolation of annually averaged air temperature in the United States, Journal of Applied Meteorology, 34, 2577–2586, doi:10.1175/1520-0450(1995)034<2577:SIOAAA>2.0.CO;2, 1995.

Zhu, X., Wei, Z., Dong, W., Ji, Z., Wen, X., Zheng, Z., Yan, D. and Chen, D.: Dynamical downscaling simulation and projection for mean and extreme temperature and precipitation over central Asia, Climate Dynamics, 54, 3279-3306, doi:10.1007/s00382-020-05170-0, 2020.

Formatted: Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)

~~Willmott, C.J. and Robeson S.M.: Climatologically aided interpolation (CAI) of terrestrial air temperature, International Journal of Climatology, 15, 221–229, doi:10.1002/joc.3370150207, 1995.~~

~~Willmott, C.J., Rowe, C.M. and Philpot, W.D.: Small-scale climate maps: a sensitivity analysis of some common assumptions associated with grid-point interpolation and contouring. American Cartographer, 12, 5-16, doi:10.1559/152304085783914686, 1985.~~

Zhu, X., Zhang, M., Wang, S., Qiang, F., Zeng, T., Ren, Z. and Dong, L.: Comparison of monthly precipitation derived from high-resolution gridded datasets in arid Xinjian, central Asia, Quaternary International, 358, 160-170, doi:10.1016/j.quaint.2014.12.027, 2015.

Zou, L., Zhou, T. and Peng, D.: Dynamical downscaling of historical climate over CORDEX East Asia domain: A comparison of regional ocean-atmosphere coupled model to stand-alone RCM simulations, Journal of Geophysical Research: Atmospheres, 121, 1442–1458, doi:10.1002/2015JD023912, 2016.

Table 1: Overview of the used reference datasets.

Dataset	Short name	Type	Resolution	Used variables	Frequency	Temporal coverage	Domain
gridded Climatic Research Unit TS dataset (version 4.02)	CRU	gridded station data	0.50°	2 m mean air temperature, 2 m maximum air temperature, 2 m minimum air temperature, precipitation	monthly	1901 - 201~~7~~8	global land mass (excluding Antarctica)
Matsuura and Willmot, University of Delaware (version 5.01)	MW	gridded station data	0.50°	2 m mean air temperature, precipitation	monthly	1900 - 2017	global land mass
Global Precipitation Climatology Centre gridded dataset (version 2018)	GPCC	gridded station data	0.50° or 0.25°	precipitation	monthly	1891 - 2016	global land mass (excluding Antarctica)
ERA-Interim	ERA-Interim	reanalysis data	0.~~25~~70°	2 m mean air temperature, precipitation	monthly	1979 - 2017	global

Table 2: Climatological mean CRU temperature (°C) for the 1980-2017 period over the CAS-CORDEX domain and biases (°C) of the RCMs (REMO and ALARO-0) and the other reference datasets (ERA-Interim and MW) against those CRU means.

	DJF	MAM	JJA	SON	Annual
CRU	-9.35	5.87	19.23	5.72	5.44
REMO - CRU	0.48	-0.56	-0.33	0.01	-0.11
ALARO - CRU	0.83	-3.19	0.02	-0.03	-0.60
ERA-Interim – CRU	0.42	0.21	0.16	-0.02	0.19
MW - CRU	-0.41	-0.19	-0.09	-0.43	-0.28

Formatted: Not Highlight

Formatted Table

Formatted: Not Highlight

Formatted Table

1255 **Table 3: Spatial average over the CAS-CORDEX domain of climatological mean CRU minimum temperature (°C) for the 1980-2017 period and biases (°C) against those CRU means for REMO and ALARO-0.**

	DJF	MAM	JJA	SON	Annual
CRU	-14.43	-0.22	13.18	0.40	-0.20
REMO - CRU	0.77	-0.25	0.60	1.09	0.55
ALARO - CRU	2.85	-1.71	1.10	1.42	0.90

Formatted Table

Table 4: Spatial average over the CAS-CORDEX domain of climatological mean CRU maximum temperature (°C) for the 1980-2017 period and biases (°C) against those CRU means for REMO and ALARO-0.

	DJF	MAM	JJA	SON	Annual
CRU	-4.29	11.97	25.34	11.06	11.09
REMO - CRU	0.08	-1.24	-1.07	-0.71	-0.74
ALARO - CRU	-0.77	-4.84	-1.46	-1.24	-2.08

Formatted Table

Table 5: Climatological mean CRU precipitation (mm month^{-1}) for the 1980-2017 period over the CAS-CORDEX domain and relative biases (%) against those CRU means for the RCMs (REMO and ALARO-0), and the other reference datasets (ERA-Interim, MW and GPCC).

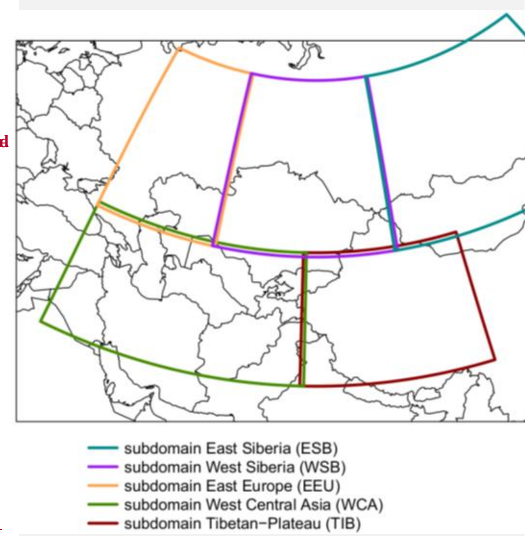
	DJF	MAM	JJA	SON	Annual
mean CRU	30.38	43.46	87.03	47.72	52.26
REMO - CRU	-4	3	-23	-11	-12
ALARO - CRU	-9	-11	-25	-9	-16
ERA-Interim - CRU	-10	3	-11	-10	-8
MW - CRU	-30	-28	-28	-27	-28
GPCC - CRU	-31	-32	-27	-30	-29

Formatted: Not Highlight
Formatted Table
Formatted: Not Highlight
Formatted Table
Formatted: Not Highlight
Formatted Table

Table S1: Overview of the model specifications for the ALARO-0 and REMO RCM experiments used for this study.

	ALARO-0	REMO
projection resolution	Lambert conical projection 0.22°	rotated pole 0.22°
horizontal spatial discretisation	spectral on collocated grid	2^{nd} order finite differences on staggered C-grid
vertical coordinate levels	46 hybrid levels	27 hybrid levels
temporal discretisation	semi-implicit semi-Lagrangian	leap-frog with semi-implicit correction and Asselin filter, semi-Lagrangian advection
time step	450 s	120 s
convective scheme	3MT scheme	Tiedtke with modifications after Nordeng and Pfeifer (Pfeifer, 2006)
radiation scheme	The Action de Recherche Petite Echelle Grande Echell (ARPEGE) Calcul Radiatif avec Nebulosité (ACRANEB) scheme for radiation	Morcrette et al. (1986) and Giorgetta and Wild (1995)
turbulence vertical diffusion	A pseudoprognostic turbulent kinetic energy (pTKE) scheme (i.e., a Louis-type scheme for stability dependencies, but with memory, advection, and autodiffusion of the overall intensity of turbulence)	Louis-type with a higher order closure scheme for the transfer coefficients of momentum, heat, moisture and cloud water within and above the planetary boundary layer. Eddy diffusion coefficients are calculated as functions of the turbulent kinetic energy.
cloud microphysics scheme	A statistical sedimentation scheme for precipitation within a prognostic-type scheme for microphysics.	The cloud microphysical scheme by Lohmann and Roeckner (1996).
land surface scheme	The Interaction Sol-Biosphère-Atmosphère (ISBA) scheme	Based on the surface runoff scheme (Hagemann, 2002), inland glaciers (Kotlarski, 2007), and vegetation phenology (Rechid, 2009)
institute	RMIB-UGent	HZG-GERICS (https://remo-rcm.de/)

Figure S1: IPCC6 subdomains projected on the CAS-CORDEX region. Absolute difference between the average seasonal and annual CRU precipitation (mm month^{-1}) and the precipitation simulated by REMO and ALARO-0 over the 1980-2017 period



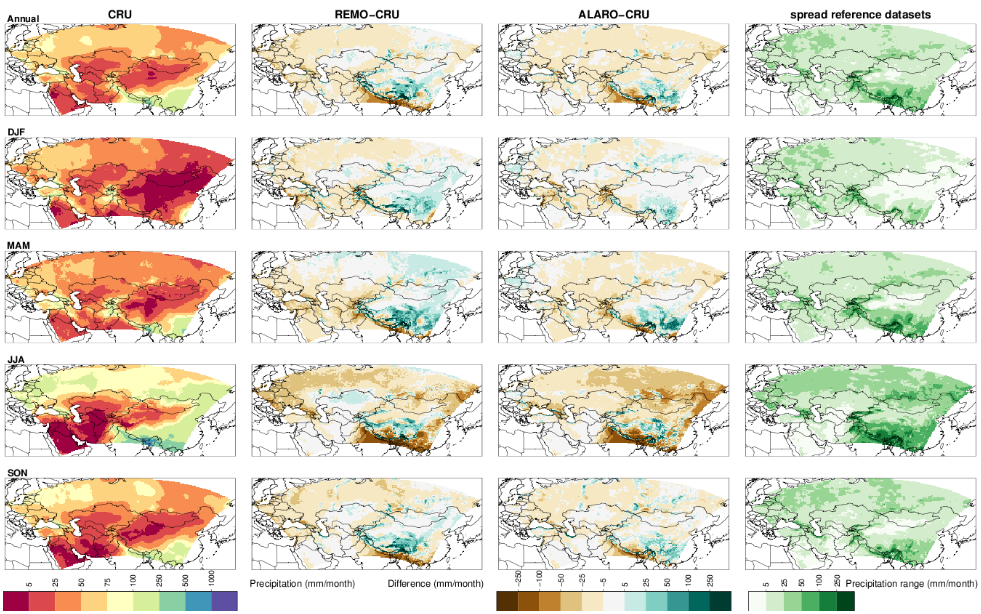


Figure S2: Absolute difference between the average seasonal and annual CRU precipitation (mm month-1) and the precipitation simulated by REMO and ALARO-0 over the 1980-2017 period.

Table S2: Climatological mean CRU precipitation (mm month^{-1}) for the 1980-2017 period over the CAS-CORDEX domain and absolute biases (mm month^{-1}) against those CRU means for the RCMs (REMO and ALARO-0), and the other reference datasets (ERA-Interim, MW and GPCC).

	DJF	MAM	JJA	SON	Annual
CRU	30.38	43.46	87.03	47.72	52.26
REMO - CRU	-1.23	1.33	-19.81	-5.24	-6.26
ALARO – CRU	-2.74	-4.98	-21.54	-4.40	-8.45
ERA-Interim - CRU	-2.90	1.25	-9.78	-4.61	-4.01
MW - CRU	-9.06	-12.37	-24.03	-13.06	-14.66
GPCC - CRU	-9.43	-13.77	-23.20	-14.11	-15.15

Formatted: Not Highlight
Formatted Table
Formatted: Not Highlight
Formatted Table
Formatted: Not Highlight
Formatted Table

~~**Figure S2: Absolute difference between the average seasonal and annual CRU precipitation (mm month^{-1}) and the precipitation in the ERA-Interim, MW and GPCC datasets over the 1980-2017 period.**~~

Data	Identifier	PID
ALARO-0		
precipitation	cordex.output.CAS-22.RMIB-UGent.CNRM-CERFACS-CNRM-CM5.historical.r1i1p1.ALARO-0.v1.mon.pr	/
temperature	cordex.output.CAS-22.RMIB-UGent.CNRM-CERFACS-CNRM-CM5.historical.r1i1p1.ALARO-0.v1.mon.tas	/
minimum temperature	Not available on the ESGF platform. Data can be downloaded with the key “userGMDpaper1” from: https://cloud.meteo.be/s/gRP2NFSfAWJas4g	/
maximum temperature	Not available on the ESGF platform. Data can be downloaded with the key “userGMDpaper1” from: https://cloud.meteo.be/s/8YEg4LY9DmX4EGF	/
REMO		
precipitation	cordex.output.CAS-22.GERICS.ECMWF-ERAINT.evaluation.r1i1p1.REMO2015.v1.day.pr	hdl:21.14103/2ecffe86-b5e4-359c-8c34-e7152de17a43
temperature	cordex.output.CAS-22.GERICS.ECMWF-ERAINT.evaluation.r1i1p1.REMO2015.v1.day.tas	hdl:21.14103/bf8468cf-b15c-3a20-ae42-4c42b14e749c
minimum temperature	cordex.output.CAS-22.GERICS.ECMWF-ERAINT.evaluation.r1i1p1.REMO2015.v1.day.tasmin	hdl:21.14103/74aa90a5-c99b-35f9-888e-acc0115dfc4d
maximum temperature	cordex.output.CAS-22.GERICS.ECMWF-ERAINT.evaluation.r1i1p1.REMO2015.v1.sem.tasmax	hdl:21.14103/a72e5ea1-533d-3685-b04d-5e4ab162e065