List of relevant changes

Dear Editor,

Thank you for giving us the opportunity to improve our paper. Based on the comments of the reviewer we have made the following major changes hoping to satisfy the reviewer with the more detailed text:

- *We have added mean absolute error (MAE) as an extra score.*
- *We have conducted the same analysis for the five subregions as was done for the complete CAS-CORDEX region, including: mean bias, MAE, spatial correlation, standard deviation and RMSE. In this way more detailed information is provided to the reader.*
- *Based on the additional analysis over the five subregions, we have substantially rewritten the result and discussion sections (Sect. 3 and Sect. 4), taking into account consistency over the different subsections.*
- *Moreover, we have updated the figures and text of the precipitation section.*
- *Finally, we have revised the conclusion and abstract based on the more detailed information given by the analysis over the subregions.*

The implementation of the changes requested by the reviewer has led to a significant extension of the manuscript. The level of detail is quite high, in accordance with the request for a more detailed analysis, even though this was not our intention when submitting the first version of our paper. We feel that this level of detail is required to address all of the concerns that were raised.

Author response to the review of the Anonymous Referee

Dear reviewer,

Thank you for your helpful comments. As you will read in the point-by-point answers, we have performed an analysis over the five subregions, similar to the complete CAS-CORDEX domain. As a consequence, our revised manuscript has a significantly higher level of detail.

General Comments

The authors surely put some efforts in trying to answer my comments on their previous version of the manuscript. Nevertheless I do not think that all these comments were exhaustively considered. In my opinion the paper still suffers from a series of major issues that need to be carefully addressed before it may be considered for publication for Geoscientific Model Development.

- The quality of the text and the structure of the manuscript are surely the points that have received more attention by the authors, but still some parts need revision. In particular you should check for consistency among the different subsections. For example in the subsections of the methods you specified what you used ERAInterim for, but you did not do the same for other data-sets such as GPCC. Another similar example can be found in the results subsection about DTR. Here you discuss the table with spatial means for maximum temperatures but not for minimum temperatures. Check for such inconsistencies throughout the text and correct them.

  We had another careful look at the different subsections and made them consistent. We added the used variables of the MW dataset that were not mentioned in an explicit way. Regarding which GPCC variables were used for this dataset, we stated: "In addition, the GPCC has no similar dataset for other variables and thus, only precipitation can be validated with this dataset." For ERA-Interim a more detailed explanation is provided since there are more options e.g. hourly, daily or monthly data, while the other datasets contain monthly data as highest time resolution for the studied variables. We mentioned in the text that monthly data is used for each gridded observational dataset, so it should be clear for the reader which data we used.

- You are not very accurate in the specification of the model behaviour and when you discuss the maps of the bias. I found a lot of inaccuracies in the text and I invite you to review it accordingly. Here a couple of examples:

  - **At the annual scale, the bias of the minimum temperature ranges mostly between -3C and 3C for REMO and between 0C and 5C for ALARO-0 (Fig. 5).**

    For REMo annual biases exceed the absolute value of 3 C over several areas such as Mongolia and the Himalayas. For Alaro a large part of the domain has a negative bias exceeding -5C in some case.

    This sentence should be placed in its context. The sentences following this sentence are indicating where the -3 °C- 3 °C range for REMO and the 0°C - 5 °C range for ALARO is exceeded, including Mongolia and the Himalayas. We understand this might not be clear when reading it for the first time and revised this paragraph to make sure it cannot be interpreted incorrectly.

  - l. 258-261 **Based on Fig. 3, both RCMs perform best during autumn and the spatial correlation is lowest during summer for ALARO-0 while, the biases during summer are smaller than during winter and spring for both RCMs (Table 2 and Fig. 2)**

    From figure 2 you cannot say that the biases in summer are smaller than in winter and spring for all the points of the domain.

    We agree, we meant that the biases are lower on average. We will revise this sentence.

- l. 452-454 **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.**
  This is not true for the entire points of the domain. When you propose such conclusions I invite you to first compare directly the map of the spread with the one of the bias (for example plotting their differences).

  The mean temperatures of the RCMs lying within the range of spread is indeed not true for all points of the domain, that is why we explicitly mention "for most parts of the domain". We added new maps with the difference between the absolute bias and the spread in the supplementary material. Based on these maps we can describe in more detail what we intended to say and in which areas the models do not perform that well.

- My main concern is that despite my previous comments, even though you added an analyses of the seasonal cycle for sub-region in the new version of the manuscript, you did not conduct the same analysis for the mean bias, spatial correlation, standard deviation and RMSE. The information of the mean bias calculated over the entire domain intuitively makes no sense, as already highlighted in my previous report. You should conduct the analyses of the bias per sub-domain too. One interesting thing that I would suggest you to do is to consider mean absolute bias instead of the bias (for both the entire domain and sub-regions, since spatial biases might compensate each other). Additionally, also Taylor diagrams should be calculated for every sub-region. For this I also think it would be important for you to independently check the values of the spatial correlations you obtained, since they seem to be too high given the spatial patterns of the bias. Since you use pre-defined functions (in R) for calculating the Taylor diagrams, I think it would make sense to double check the correctness of the results, independently. Finally, the plots of the seasonal cycle should be improved. In particular it was impossible to understand the ones drawn for temperature.

  We added the Taylor diagrams for the subregions including spatial correlation, standard deviation and RMSE. We agree with your comment that the mean bias over the entire domain does not provide much insight. We calculated instead this metric over the subregions, as it can give an insight into the regional dependence of the bias. We agree that the mean absolute error (MAE) is a better metric and we added MAE for the entire CAS-CORDEX domain and the subdomains.

  We calculated the spatial correlation, standard deviation and RMSE in two independent ways: once based on the formulas described by Kotlarski et al. (2014) and once with the pre-defined Taylor function in R. We obtained the same values for both approaches. The spatial correlations are based on the Pearson correlation method in the pre-defined Taylor function and can be described by the PACO formula (Kotlarski, et al.2014) as was mentioned in the manuscript. Calculations of the correlations based on the built in Pearson correlation in R and calculations with the PACO formula, both showed the same values as the values plotted with the built in Taylor function, so we are sure that the correlations are that high over the full domain.

  Figure 4 with the annual cycles of mean, minimum and maximum temperature was an incorrect figure, the lines of ERA-Interim for minimum and maximum temperature were incorrect. This might be the reason for not being able to understand the figure. We tried to put as much information on one figure to reduce the length of the manuscript and to be as concise as possible.

- The discussion part does not always result very clear and I would suggest you to carefully revise it while modifying it in consideration of my new comments.

  We revised this section.

- As a final remark, it seems that the authors are a bit too positive about the models performance for the region. I would suggest them to try to be more objective in their conclusions. Maybe new analyses might help in this sense. In my personal opinion the models results cannot be considered reliable over a large part of the domain. It is true that over some areas there is an issue with the poor reliability of observations, but over some regions the main issue is still the model. Having a bias above 10C does not make the model reliable. This is for example the case of temperatures simulated by ALARO over the north-western and north-central part of the domain, both in winter and in summer. Appropriate bias correction methods could be used to make the model more in-line with ”reality”, but this is not inherent to model evaluation and should be made clear in the text.

We reformulated the conclusions based on the information of the subdomains.

Additionally, we used an incorrect transformation of units for the CRU precipitation data in the results of the previous version of the manuscript. We have therefore updated the evaluation of the precipitation.

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[2,4]

[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. CORDEX Central Asia is one of these regions and this article describes the evaluation for two regional climate models (RCMs), REMO and ALARO-0, that were run for the first time at a horizontal resolution of 0.22° (25 km) over this region~~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 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 REMO model scores better for temperature, whereas the ALARO-0 model prevails for precipitation. Studying specific subregions provides a deeper insight into the strengths and weaknesses of both RCMs over the CAS-CORDEX domain. For example, ALARO-0 has difficulties in simulating the temperature over the northern part of the domain, particularly when a snow cover is present, while REMO poorly simulates the annual cycle of precipitation over the Tibetan Plateau. The evaluation of minimum and maximum temperature demonstrates that both models underestimate the daily temperature range. This study aims to evaluate whether REMO and ALARO-0 provide reliable climate information over the CAS-CORDEX domain for impact modelling and environmental assessment applications. Depending on the evaluated season and variable, it is demonstrated that the produced climate data can be used in several subregions e.g. temperature and precipitation over West Central Asia in autumn. At the same time, a bias adjustment is required for those regions where significant biases have been identified~~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 climate data can be used for in impact studies taking into account a bias correction for those regions where significant biases have been identified~~.

1 Introduction

There is a strong need for climate information at the regional-to-local scale that is useful and usable 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 conduct ~~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), Mediterranean (Ruti et al., 2016) 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 global-regional climate models (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 change simulations covering all major inhabited regions with a spatial resolution of about 25 km has been established within the WCRP CORDEX COmmon Regional Experiment (CORE) Framework to support 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 0.11° or 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°, insufficient for impact modelling and environmental assessment applications, 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. Thus higher-resolution climate data over the CAS-CORDEX region is needed (Kotova et al., 2018).~~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 run at 0.22° or 25 km resolution over the CAS-CORDEX region. This study aims to address the scarcity of reliable climate information over the CAS-CORDEX domain by evaluating two different RCMs based on multiple scores for temperature (mean, minimum and maximum) and precipitation over the longer period of 38 years.~~The current study significantly extends our knowledge of the CAS-CORDEX domain by evaluating 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 study is necessary 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 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 the absence of an ensemble of RCM runs over Central Asia. Additionally, in some regions the quality of gridded observational datasets, constructed through interpolation or area-averaging of station observations, is poor due to over-smoothing of 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 orographically complex regions such as the Himalayas. The current 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).

This study contains two assets: for the first time an in-depth evaluation of the RCMs ALARO-0 and REMO is performed at 0.22° spatial resolution over the CAS-CORDEX domain and we reflect on the impact of the observational datasets on the model evaluation. 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 and i~~. I~~n the final Sect. 5 we summarize the conclusions.

2 Methods

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 and~~,~~ 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. 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~~ poorly resolve the mountain ranges with a spatial resolution less than 0.50°~~ 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, such as permafrost in the north and the extremely hot regions and monsoon-driven climates with abundant precipitation linked to the Inter-Tropical Convergence Zone (ITCZ) passing in the south.~~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 model domain excluding the relaxation zone~~non-coupling zone~~. The CAS-CORDEX 0.22° ALARO-0 inner domain encompasses 333 by~~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 relaxation zone of eight grid points at every boundary.

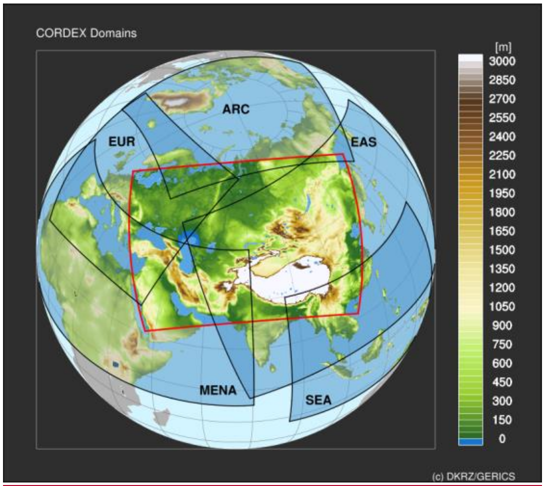
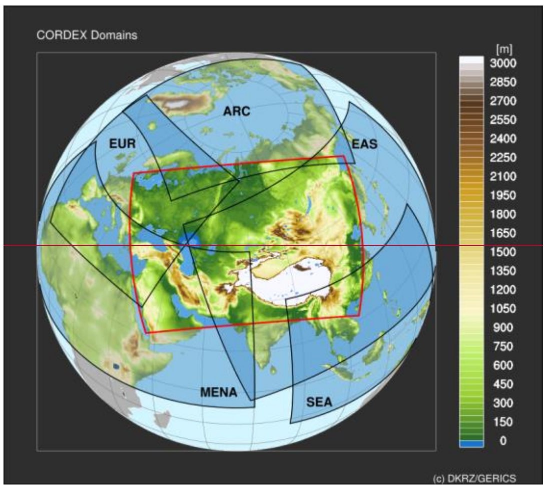


Figure 1: The CAS-CORDEX domain demarcated by a red contour and the main overlapping CORDEX domains (black contour lines): Europe (EUR), Arctic (ARC), South East Asia (SEA), East Asia (EAS) and MENA projected upon the topography of Eurasia (geopotential height [m] of the GTOPO30 global digital elevation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

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 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: 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).**~~

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 (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~~ uses 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 both RCMs, a run driven by a large-scale forcing taken from the ERA-Interim global reanalysis (Dee et al., 2011) wa~~i~~s 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 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 initiali~~s~~zed 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. Furthermore, constant climatological fields for some parameters are used and updated monthly following the methodology of Giot et al. (2016). 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 designated ~~taken~~ as spin-up year, REMO was spun-up for 10 years to produce an equilibrium for the soil temperature and soil moisture. These soil fields were then used as initial soil conditions when restarting the model from 1979. 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). 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 at a grid resolution of 0.50° covering the ~~complete~~ global land mass (excluding Antarctica) (Harris et al., 2020). 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, although~~however,~~ some issues have been reported (Harris et al., 2020), with the main concerns including ~~). Main concerns include~~ sparse coverage of measurement stations over certain regions, e.g. ~~the~~ Northern Russia and the dissimilarities in measurement methods that are used by different countries (Harris et al., 2020).

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 and Matsuura, 1995; Harris et al., 2020). Additionally, this dataset only contains monthly values of mean near surface air temperature and precipitation, which are used in this study. 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 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 slightly underestimated overall by GPCC~~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 also 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 0.7~~9~~0°~~ 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~~ to be used as forcing for both RCMs at a spatial resolution of 0.25°. Moreover, the ERA-Interim data is used to study the spread between observational gridded datasets and reanalysis data. To evaluate precipitation, t~~T~~otal monthly precipitation 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~~ at the 2 m temperature level are used for the mean temperature, while the minimum and maximum temperatures are retrieved by extracting the minimum and the maximum respectively from the 3-hourly ERA-Interim forecasts~~are used to study the spread between observational~~

~~gridded datasets and reanalysis data~~. Several studies have shown that ERA-Interim tends to have a warm bias in the northern part ~~over~~ of 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, Sun et al. (2018) showed that ERA-Interim generally overestimates precipitation~~Ozturk et al. (2016) showed that ERA-Interim tends to have a dry bias over the CAS-CORDEX region~~.

2.4 Analysis methods

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.

For ALARO-0 and REMO, hourly values for temperature at 2 m ~~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. Seasons are defined as meteorological seasons, where winter includes: December, January and February (DJF), spring: March, April and May (MAM), summer: June, July and August (JJA), and autumn: September, October and November (SON).

The diurnal temperature range was obtained by subtracting the minimum temperature from the maximum temperature and a height correction was performed for mean, minimum and maximum temperature assuming a uniform temperature lapse rate of 0.0064 K m^{-1}.

The model evaluation wa~~i~~s done by calculating different evaluation metrics over the CAS-CORDEX domain and the defined subdomains 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 and maps that visuali z~~s~~e the spatial patterns of the bias between the RCMs and reference datasets. The spread between the different reference datasets (observational datasets and ERA-Interim reanalysis dataset) is calculated for each grid point by taking the maximum value of the different reference datasets subtracted by the minimum value, and this for every 3-month period (season) averaged over the 1980-2017 period. The relative bias for precipitation is computed by subtracting the CRU value from the RCM or any other reference dataset and dividing it by the CRU value. These climatological means and biases were spatially averaged to obtain one mean value over the complete domain and subdomains. Additionally, the mean absolute error (MAE) was calculated to account for compensating errors. Moreover, Taylor diagrams were produced in order to study the model performance for the different seasons and for annual means. These~~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(CRU in this case) over the domain~~, 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 evaluation are presented with a focus on evaluation metrics of seasonal means of mean, minimum and maximum near surface air temperature (henceforth denoted as temperature) and seasonal mean precipitation (henceforth precipitation). This is done for the complete CAS-CORDEX domain and for the five subregions. 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 will be described in Sect. 4.

3.1 Mean temperature

~~**3.1.1 Annual and seasonal means over CAS-CORDEX domain**~~

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, which will be discussed in Sect. 4. Both RCMs are producing similar mean annual temperature patterns in the western part of the domain since they have similar biases with respect to CRU (Fig. 2).~~Both RCMs are producing similar mean annual temperature patterns since they have similar biases with respect to CRU.~~ 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, apart from the orographically complex regions and some areas in North and East Siberia for ALARO-0. The spatially averaged mean temperatures of CRU at the annual level and for the different seasons during the 1980-2017 period are given in Table 2, accompanied by the mean bias and MAE over the domain for the RCMs.

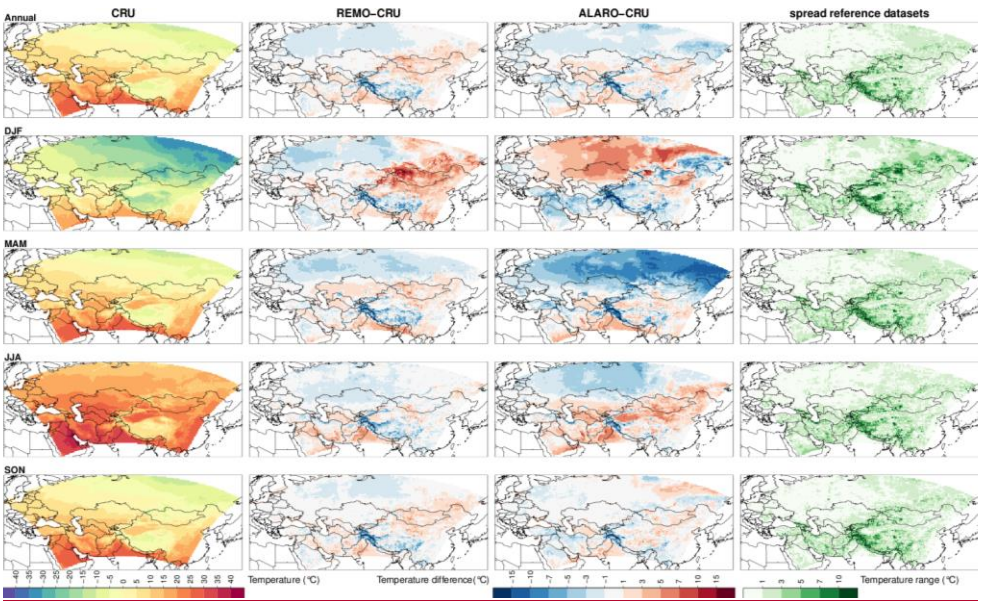


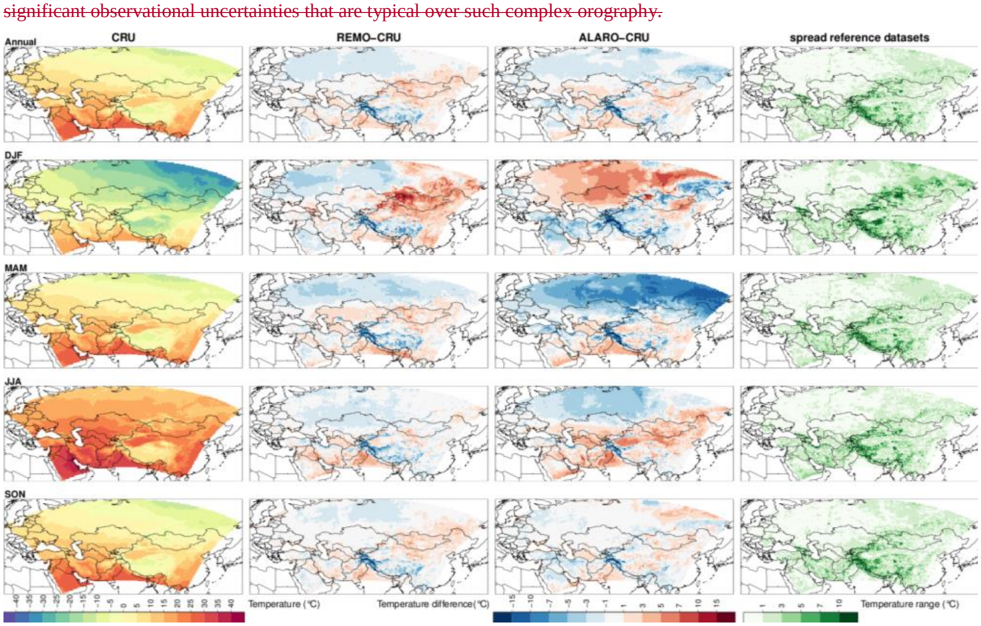
Figure 2: Left column: mean air temperature (°C) at 2 m height over the CAS-CORDEX domain based on the observational CRU dataset for the 1980-2017 period on annual level and for winter (DJF), spring (MAM), summer (JJA) and autumn (SON). In the middle columns: difference in mean temperature between models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

On the seasonal timescale, biases over larger areas are mainly pronounced in winter (DJF) and spring (MAM), particularly in the north-eastern part of the domain for ALARO-0 with strong biases up to 10 °C and -15 °C, respectively. These large biases for ALARO-0 are reflected by the large spatially averaged biases and MAE for the northern subdomains EEU, WSB and ESB as presented in Table 2. In winter the most pronounced bias is found for REMO over the north-western part of Mongolia in the Altai mountains, resulting in a large MAE of 3.40 °C over the ESB domain. 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 (Fig. 2 and Table 2). For the summer (JJA) season, warm biases occur over the southern part of the domain for both RCMs, with exception of some regions such as the Himalayas, south-eastern China and the northern border of Iran, which exhibit cold biases. Contrary, cold biases in summer are overall more dominant in the north. These biases in summer are more pronounced for ALARO-0 and both models have the smallest biases and MAE over the ESB region in this season (Table 2). Both models show modest bias patterns in autumn (SON), with notably modest warm biases over the eastern part of the domain (Fig. 2). In agreement with Fig. 2 the spatial averaged biases and MAE in Table 2 are small for both RCMs during autumn, especially for

East Europe (EEU), the west and central Russian region and Kazachstan (WSB).~~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.~~

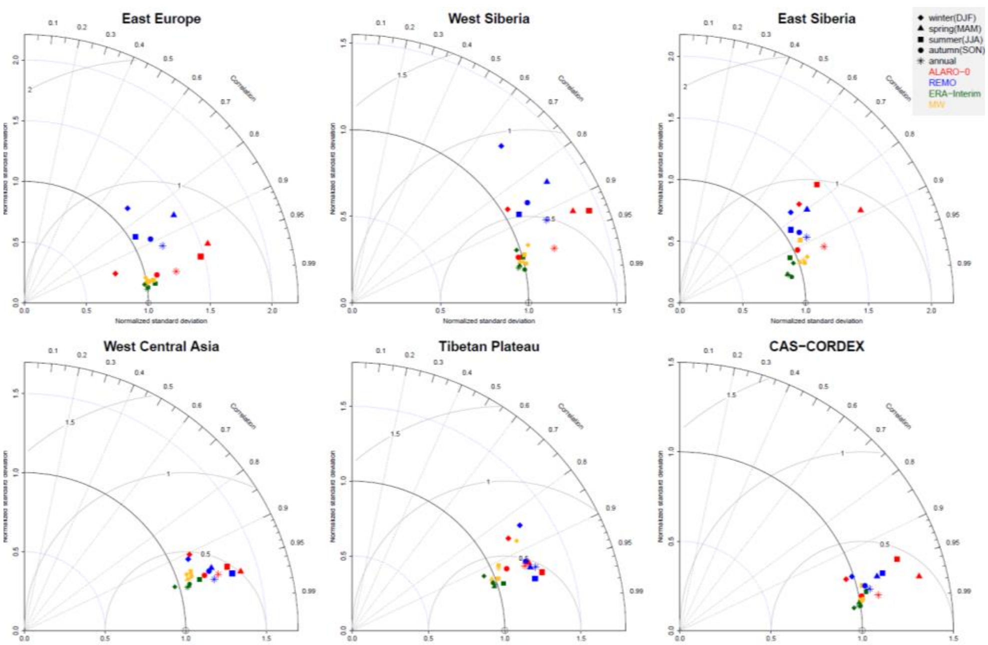


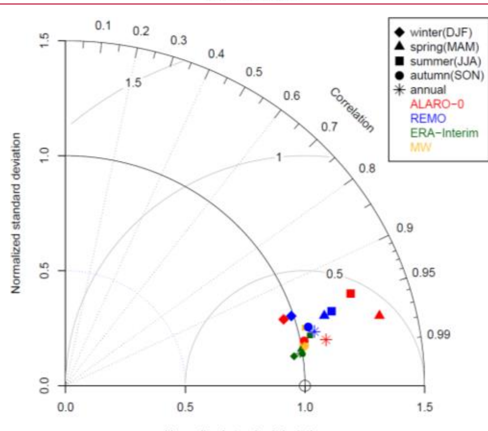
~~**Figure 2: On the left: mean air temperature (°C) at 2 m height over the CAS-CORDEX domain based on the observational CRU dataset for the 1980-2017 period on annual level and for winter (DJF), spring (MAM), summer (JJA) and autumn (SON). In the middle columns: temperature difference (°C) between the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

~~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.~~ Figure 3 shows a Taylor diagram for the mean temperature of both RCMs for the different seasons and for the annual mean value in the subdomains and the complete CAS-CORDEX domain. 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. However, the Taylor diagrams for the subdomains illustrate how scores calculated over the complete CAS-CORDEX domain can hide underlying regional trends. When considering the different subdomains, both RCMs perform generally best over the WCA subdomain and the RCMs perform best during autumn, except for the REMO simulations in the subregions WSB and TIB. During the other seasons both RCMs simulate the temperature clearly worse in the northern part of the CAS-CORDEX domain (EEU, WSB, ESB). In general, both RCMs simulate the normalized standard deviation of the temperature well during autumn and winter. Additionally, REMO simulates the normalized standard deviation well during summer for the northern subdomains.~~Based on Fig. 3, both RCMs perform best during autumn and the spatial correlation is lowest during summer for ALARO-0 while, the biases during summer are smaller than during winter and spring for both RCMs (Table 2 and Fig. 2). This is related to less spatial variability in temperatures during summer compared to the other seasons, as can be seen in Fig. 2 for CRU. An equal bias in temperature for each season would 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, although there was a clear warm bias observed during winter (Table 2 and Fig. 2).~~ 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 for the northern subregions. In general, both RCMs overestimate the normalized RSV but ALARO-0 underestimates it in winter over the EEU subdomain and in autumn over the WCA subdomain. The small mean bias during summer (JJA) for ALARO-0 over ESB and the complete domain (Table 2) is obtained by averaging the warm biases in the south and the cold biases in the north (Fig. 2) and results in a poorer overall performance of the modelled temperature since the spatial correlation is lower and the normalized standard deviation is higher (Fig. 3). The spatial pattern correlation is lowest during winter for both RCMs, except for the ESB subdomain where ALARO-0 shows a lower spatial correlation during summer. Moreover, the spatial correlation is extremely high (> 90 %) for ALARO-0 over all subdomains on the annual level. Annual mean temperatures of REMO have slightly lower spatial correlations with CRU when compared to those of ALARO-0 but they are still high (> 90 %), excluding the ESB subregion. ~~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).~~ Comparing the metrics of the RCMs (Fig. 2, Fig. 3 and Table 2) shows that REMO is better in simulating the variability in temperature ~~and has smaller biases~~ compared to ALARO-0, except for the autumn in all subdomains and winter in the WSB and TIB subdomain. On the other hand ALARO-0 often better captures spatial temperature patterns since the spatial pattern correlation is slightly higher than for REMO, except during winter and summer over the ESB and WCA subregions and spring and summer over the TIB subregion~~summer~~.



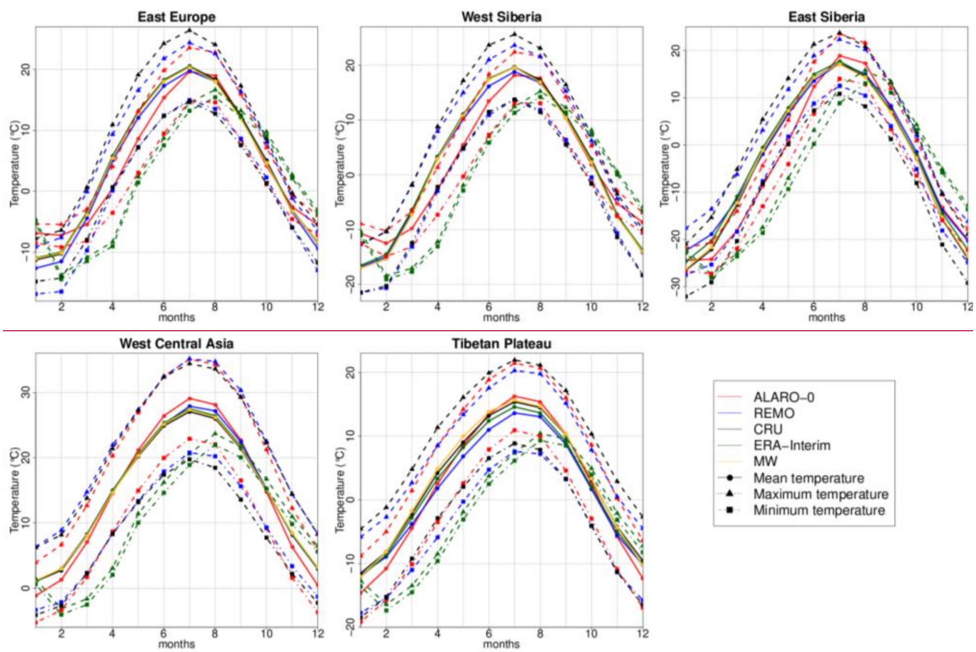


Formatted: English (United States)

Figure 3: Normalized Taylor diagram expressing the spatial 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 for the five subdomains and the complete CAS-CORDEX domain~~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 analyz~~s~~ing the seasonal cycle of the mean temperature for the different subdomains (Fig. 4), it is indeed observed that the RCMs simulate the mean temperature extremely ~~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 respectively during December. During winter months (months 12, 1 and 2) REMO simulates temperatures within the uncertainty range for West Siberia and for East Europe underestimates the temperatures on average by 1.4 °C in January. ~~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 (Table 2) 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 extremely ~~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 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). This compensating effect is reflected by the large MAE over this subdomain during the mentioned seasons (Table 2).


East Europe
West Siberia
East Siberia
West Central Asia
Tibetan Plateau
Temperature (°C)
months
ALARO-0
REMO
CRU
ERA-Interim
MW
Mean temperature
Maximum temperature
Minimum temperature

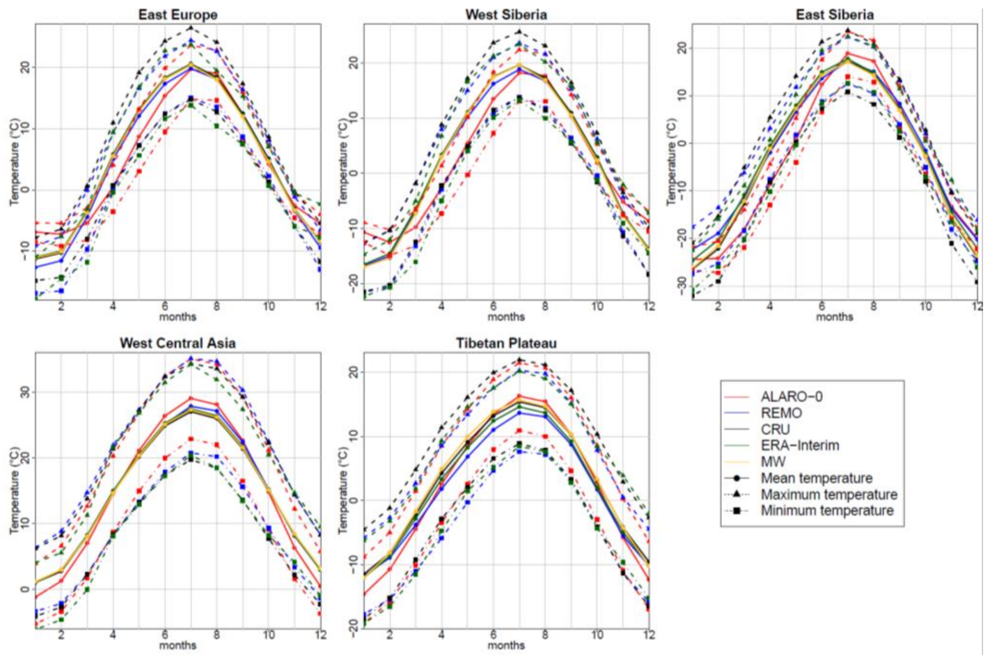


Figure 4: Annual cycles of the mean, minimum and maximum temperature for both RCMs (ALARO-0 and REMO) compared to the ERA-Interim reanalysis, MW and CRU observational data over five subdomains.

3.2 Diurnal temperature range

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

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 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 in Fig. 5. Annual biases of the minimum temperature over Russia vary mostly between -3 °C and 3 °C for REMO and between -1 °C and 5 °C for ALARO-0, excluding the orographically complex regions e.g. the Stanovoy Range and Central Siberian Plateau. ~~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. The~~se~~ warm biases for REMO in the eastern part of the domain are most pronounced during winter. ALARO-0 also 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. During the summer season the biases for the REMO model are limited between -5 °C and 7 °C except for the Himalayan mountain range, 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 (Fig. 5). 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, western Russia 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.

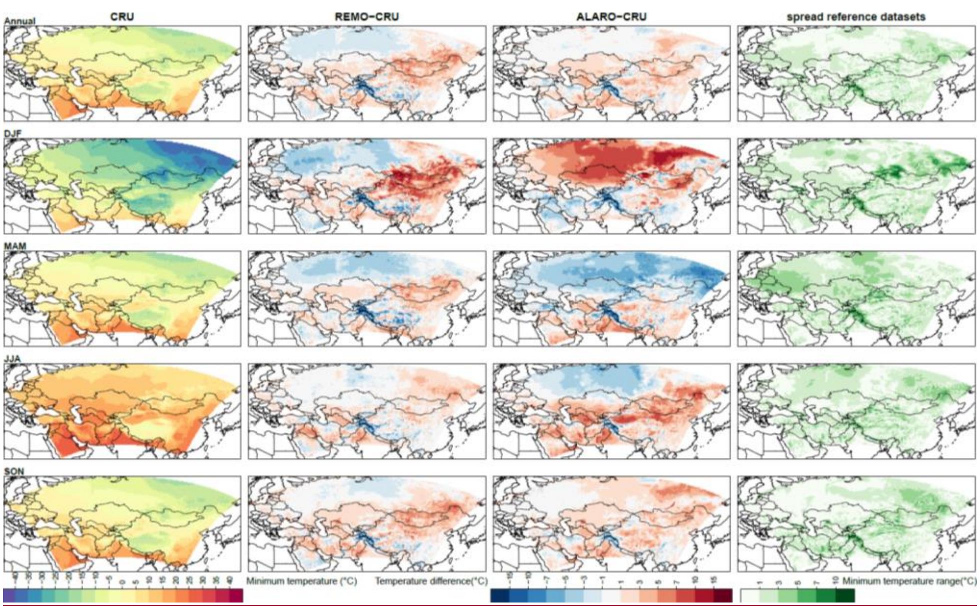


Figure 5: Left column: minimum air temperature (°C) at 2 m height over the CAS-CORDEX domain based on the observational CRU dataset for the 1980-2017 period on annual level and for winter (DJF), spring (MAM), summer (JJA) and autumn (SON). In the middle columns: difference in minimum temperature between models and CRU. Right column: the range in minimum temperature (°C) between the different reference datasets (CRU and ERA-Interim).

Table 3 shows the spatially averaged biases and MAE for minimum temperature during the 1980-2017 period of both RCMs and ERA-Interim compared to the minimum temperatures of CRU for the different seasons over the CAS-CORDEX domain and subregions. These scores confirm that the RCMs ALARO-0 and REMO are not able to reproduce the minimum temperature over the northern and eastern part of the domain during winter. During winter and spring, both models simulate minimum temperature best over the subregion WCA, while during summer and autumn they both perform best over the EEU region. REMO is able to simulate the minimum temperature accurately over the EEU and WSB subdomains during summer since the errors are small (MAE < 1 °C). In general ALARO-0 has difficulties in simulating the minimum temperature correctly in any season and is only able to simulate the minimum temperature well over the EEU region during autumn.

The metrics in Fig. 6 confirm that the RCMs struggle to simulate the spatial pattern of minimum temperature well over the north-eastern part of the domain (ESB). The RCMs simulate the spatial pattern of minimum temperature well over the WCA region.

Formatted: Justified

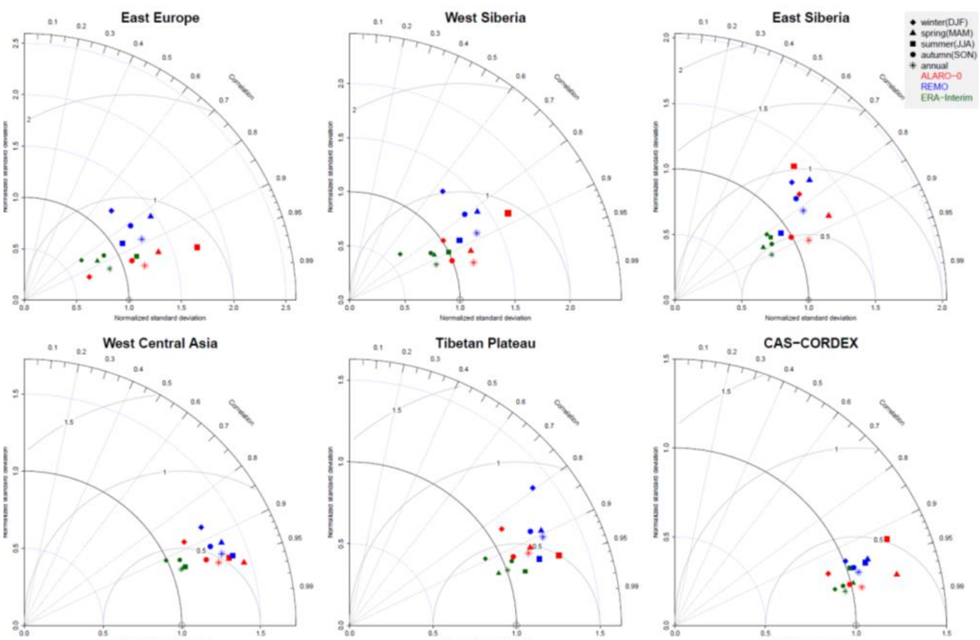


Figure 6: Normalized Taylor diagram expressing the model spatial performance of the minimum temperature for seasonal and annual means for both RCMs (ALARO-0 and REMO) and ERA-Interim reanalysis with respect to CRU for the five subdomains and the complete CAS-CORDEX domain.

Additionally, ALARO-0 produces minimum temperatures with a high spatial correlation to CRU over the EEU subdomain. At annual and seasonal scale, except for summer in WSB, ESB and TIB, ALARO-0 has 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 during summer, except for the WCA region (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 generally 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). During winter, 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. Biases in Figure 7 and Table 4 show that a pronounced cold bias is present~~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.

Formatted: Justified

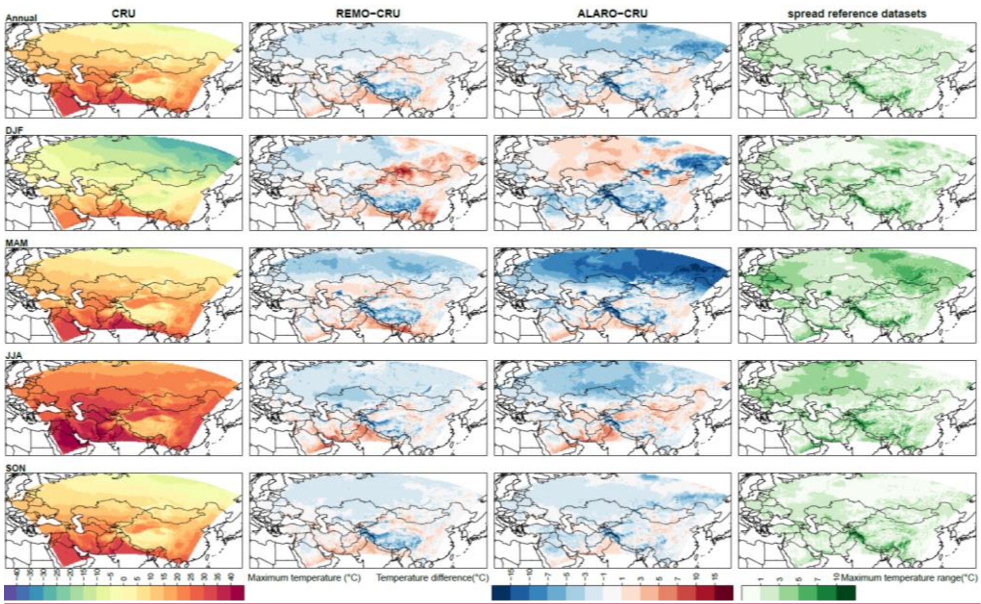
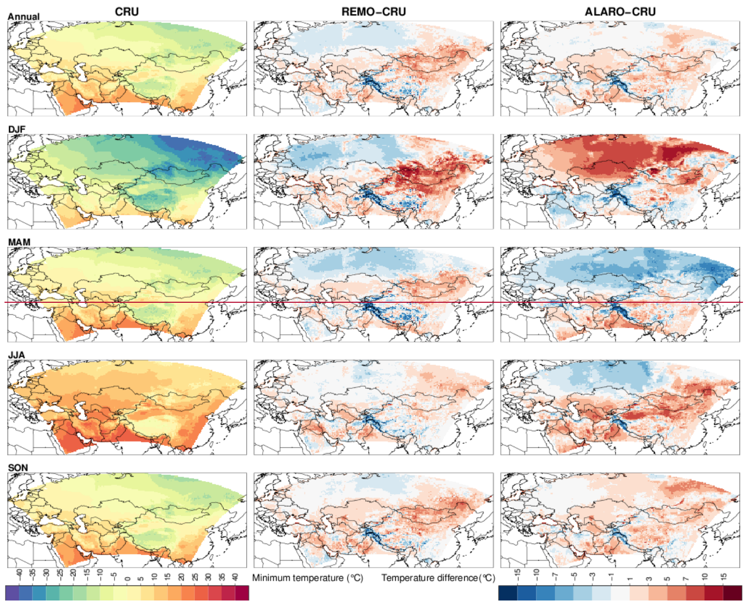
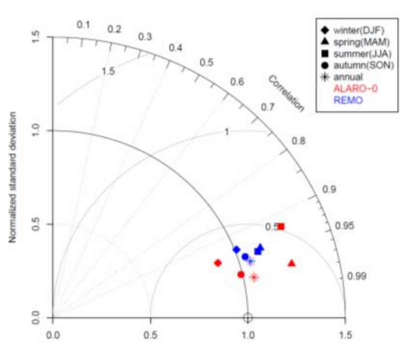


Figure 7: : Left column: maximum air temperature (°C) at 2 m height over the CAS-CORDEX domain based on the observational CRU dataset for the 1980-2017 period on annual level and for winter (DJF), spring (MAM), summer (JJA) and autumn (SON). In the middle columns: difference in maximum temperature between models and CRU. Right column: the range in maximum temperature (°C) between the different reference datasets (CRU and ERA-Interim).

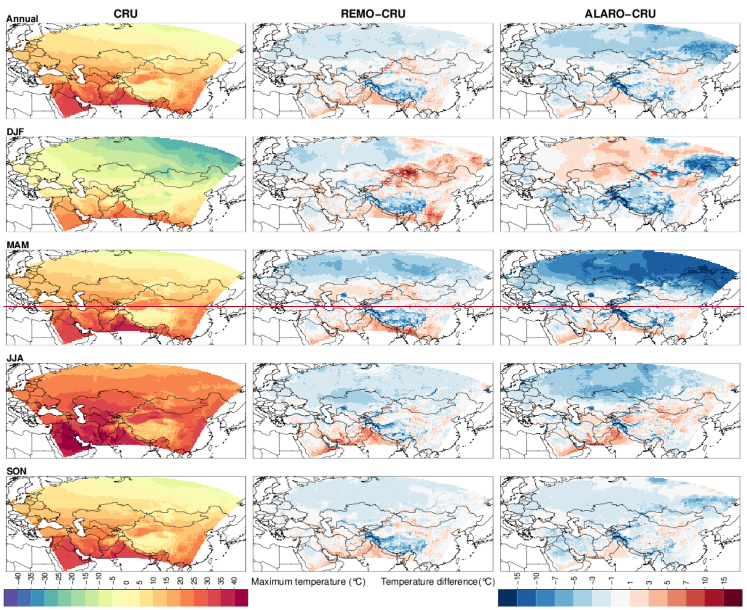
During winter, ALARO-0 has the best performance over the EEU domain, while REMO has the best performance over the WCA subdomain (Table 4 and Fig. 7). The numbers in table 4 confirm that during spring, the maximum temperature over the northern part of the domain deviates strongly (MAE > 2.50 °C) from CRU for both RCMs. Based on the MAE, both RCMs show the best performance for maximum temperature during autumn over all subdomains, except for REMO over the TIB subdomain. From this we can conclude that ALARO-0 simulates the maximum temperature poorly in any season~~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~~.



~~**Figure 5: 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.**~~

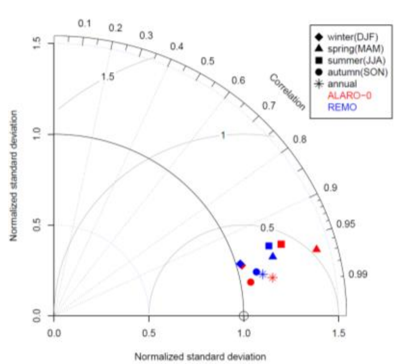


~~**Figure 6: Normalized Taylor diagram representing the model performance of the minimum temperature for seasonal and annual means for both RCMs (ALARO-0 and REMO) with respect to CRU.**~~



~~Figure 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: Normal, Tab stops: 8,85 cm, 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 shows that for all seasons, both RCMs have a high spatial correlation (> 90%) and a normalized RSV close to 1 for maximum temperature over the WCA subdomain. This is the case as well for the TIB subdomain, excluding the winter season. ALARO-0 has a high spatial correlation over the EEU subdomain during all seasons and over the WSB subdomain except for winter. Both RCMs struggle the most with reproducing the spatial patterns over the ESB subdomain. ALARO-0 has higher spatial pattern correlations with CRU compared to REMO, except for autumn over the TIB subdomain and winter over the ESB and WCA subdomains.

Formatted: Justified

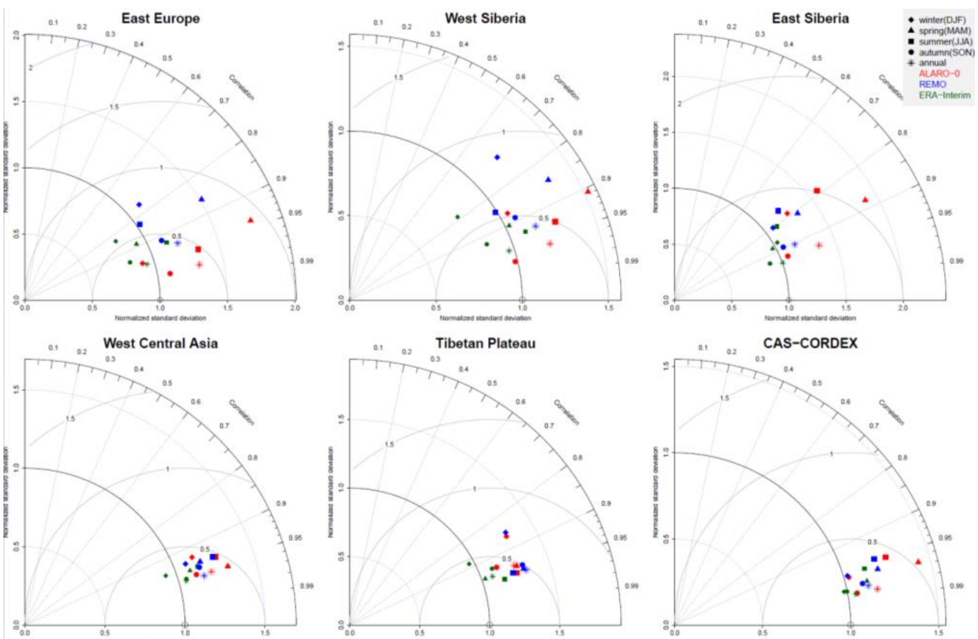


Figure 8: Normalized Taylor diagram expressing the model spatial performance of the maximum temperature for seasonal and annual means for both RCMs (ALARO-0 and REMO) and ERA-Interim reanalysis with respect to CRU for the five subdomains and the complete CAS-CORDEX domain.

REMO has more often a normalized RSV value closer to 1 than ALARO-0, for the different subdomains and seasons. Additionally, it is seen that both RCMs overestimate the normalized RSV of the maximum temperature for each subdomain and season, except for winter in EEU and summer and autumn in WSB ~~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. 8). Based on Fig. 7 and 8, both RCMs simulate the maximum temperature best during autumn.

~~-~~In general the minimum temperature (Table 3 and Fig. 4) shows warmer biases than the mean temperature (Table 2 and Fig. 2) over the different seasons, excluding winter in EEU and WSB and spring in WSB and TIB, while~~) and~~ the maximum temperature (Table 4 and Fig. 6) shows colder biases compared with the mean temperature , excluding winter and spring in WCA and summer in TIB. The increased minimum temperatures obtained with the RCMs indicate that they do not capture the coldest diurnal temperatures, while they do not capture the warmest diurnal temperatures based on the decreased maximum temperatures. ~~over the different seasons.~~ From this one 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 spatial averaged minimum temperatures as they are observed, resulting in better model results during autumn.~~.~~ ~~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 of~~for~~ CRU for the entire annual cycle over the Tibetan Plateau subregion. ALARO-0 overestimates minimum temperatures ~~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.

3.3 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 over the full domain and subdomains. The relative biases and MAE of the RCMs with respect to CRU during the different seasons and on an annual level are presented as well. For both RCMs the overall bias for precipitation is wet, except for spring and summer in the WCA subdomain and for ALARO-0, during summer in WSB, winter in WCA and spring and summer in the ESB subdomain. At the annual level, the REMO model mainly shows a wet bias in the northern and the eastern part of the domain and a dry bias in the south-western part of the domain, while ALARO-0 has a wet bias in the north-west and south-east (Fig. 9). Furthermore, a strong wet bias is persistent over the annual cycle for both RCMs over the East Asian monsoon region TIB, with a less notable wet bias during summer. Next to these wet biases in the monsoon region, both models show dry biases over the Taklamakan desert, except for winter. During winter both RCMs have an extremely strong wet bias in the eastern part of the domain (Fig. 9 and Table 5). This is ~~. 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 region~~s~~. The largest relative biases can be found in relatively dry regions and therefore the absolute biases are presented in the supplementary material Fig. S5~~2~~ and Table S2. When the absolute bias during winter is examined, then it is seen that both RCMs only simulate an extremely small absolute overestimation in precipitation (< 5 mm per month)

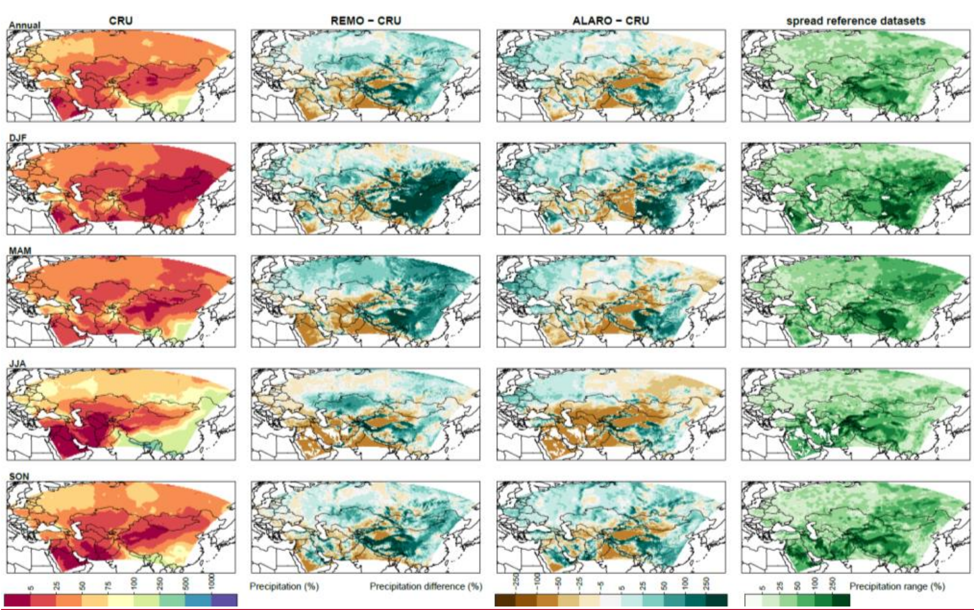
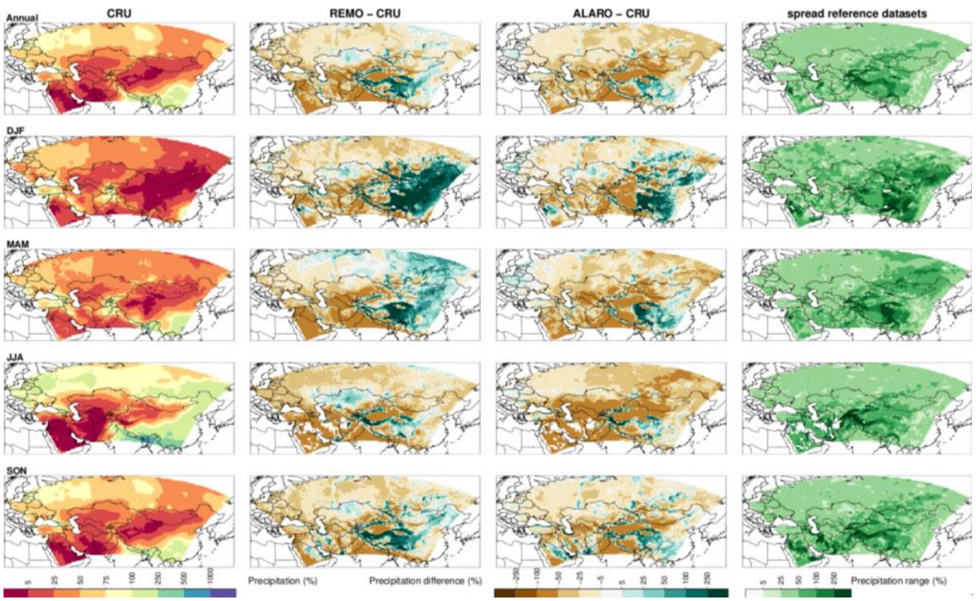


Figure 9: Left column: mean monthly precipitation amounts (mm month-1) over the CAS-CORDEX domain based on the observational CRU dataset for the 1980-2017 period on annual level and for winter (DJF), spring (MAM), summer (JJA) and autumn (SON). In the middle columns: relative difference between the average annual and seasonal CRU precipitation and the precipitation

simulated by the models (%). Right column: the range in precipitation (%) between the different reference datasets (CRU, MW, GPCC and ERA-Interim).

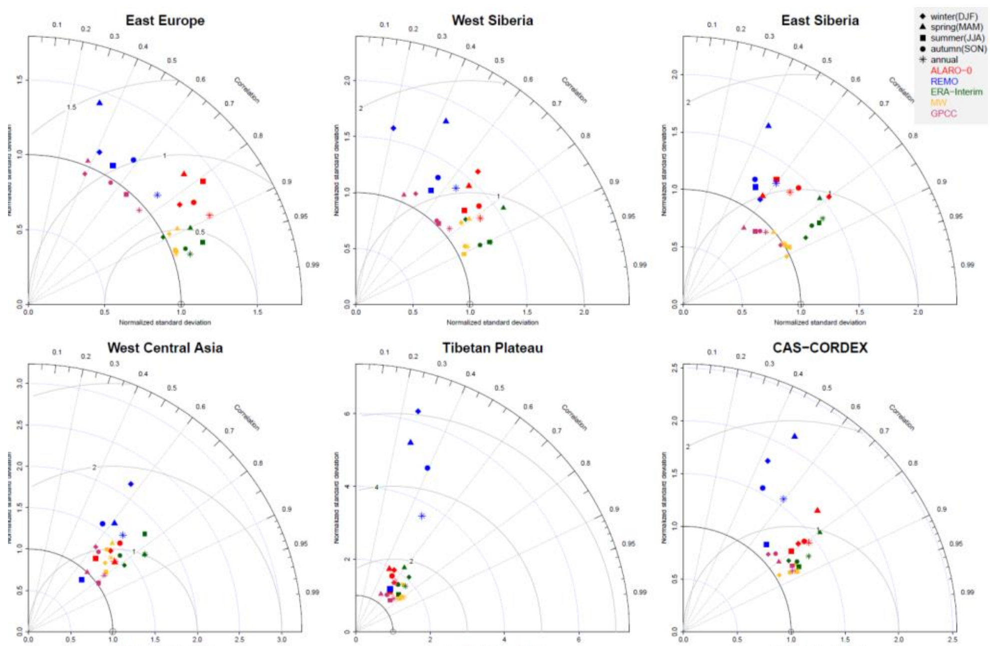
over Mongolia and the northern part of China. In spring, a clear wet bias is also present for REMO over the complete northern part of the domain and for ALARO-0 over the north-western part, while a strong dry bias is present in the south-western part of the domain for both RCMs (Fig. 9). The wet bias over East Siberia during spring is low in absolute values when compared to the subdomain Tibetan Plateau (Fig. 11 and S2). In summer, both RCMs also have a dry bias over the south-western part of the domain. The Taklamakan and Arabian deserts are located in these areas with a dry bias. In Fig. S5, the absolute dry biases over these regions are less pronounced (< -25 mm per month). The dry biases over the south-western part of the domain result in spatially averaged negative biases for precipitation over the WCA subdomain in spring and summer for both RCMs (Table 5). Additionally, a smaller relative wet bias is present over the East Asian monsoon region during summer compared to the other seasons (Fig. 9). This is related to the higher precipitation rates in the south-eastern part of the domain during summer due to the East Asian Monsoon. Moreover, both RCMs have a dry bias in the northern part of the domain during summer (Fig. S5). For REMO this dry bias is situated in the north-western part of the domain and for ALARO-0, a stronger dry bias is situated in the north-eastern part of the domain, resulting in a significant dry bias over the ESB subdomain (Table 5). Furthermore, the dry bias over the Taklamakan desert is more outspoken in summer. In autumn, both RCMs mainly produce a wet bias over the CAS-CORDEX domain, excluding some areas with low precipitation rates that have dry biases e.g. the Taklamakan desert. In absolute numbers these dry biases are limited (< -25 mm per month).~~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).~~



~~**Figure 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).**~~

~~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 it can be deduced that REMO is only able to reliably reproduce the precipitation over the TIB subdomain during summer and not during the other seasons~~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, except for the summer precipitation over WCA. Despite the substantial ALARO-0 biases shown in Table 5 over most parts of the domain, the spatial patterns are thus well represented (Fig. 9 and 11). Both RCMs overestimate the variability in precipitation for all seasons and subregions, except for REMO in summer over WCA (Fig. 10). This excessive spatial variation is due to an overestimation of the precipitation in the wettest regions combined with an underestimation in the driest regions (Fig. 9).~~. 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).~~



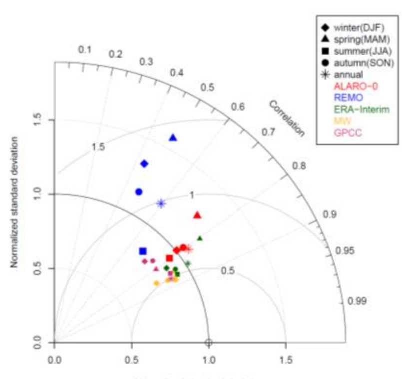
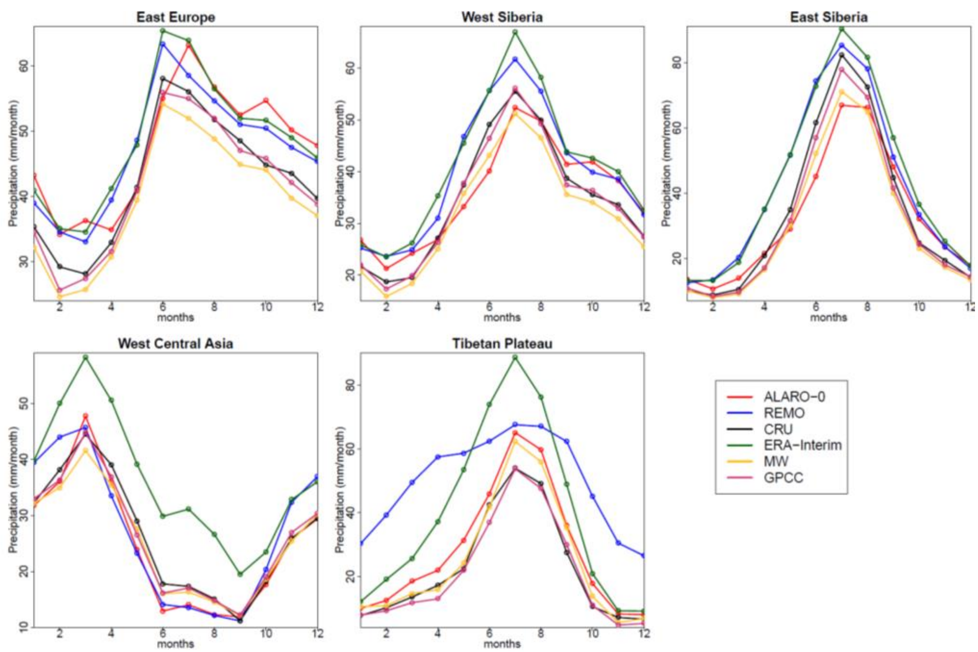


Figure 10: Normalized Taylor diagram expressing ~~representing~~ the model performance of precipitation for seasonal and annual means for both RCMs (ALARO-0 and REMO), gridded observational datasets (MW, GPCC) and the ERA-Interim reanalysis data with respect to CRU for the five subdomains and CAS-CORDEX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

~~3.3.2 Annual cycles over subdomains~~

The annual cycles over the subdomains show that ALARO-0 and REMO indeed mostly ov~~und~~erestimate 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). However, 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 March and June. As mentioned before~~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. As seen in Fig. 11 and Table 5 the spatially averaged precipitation rate of REMO is slightly closer to the observations than ALARO-0 over the EEU subdomain during winter and autumn. In addition, the annual cycle and MAE show that REMO better captures the precipitation over the ESB region than ALARO-0 during summer.


East Europe
West Siberia
East Siberia
West Central Asia
Tibetan Plateau
Precipitation (mm/month)
months
ALARO-0
REMO
CRU
ERA-Interim
MW
GPCC

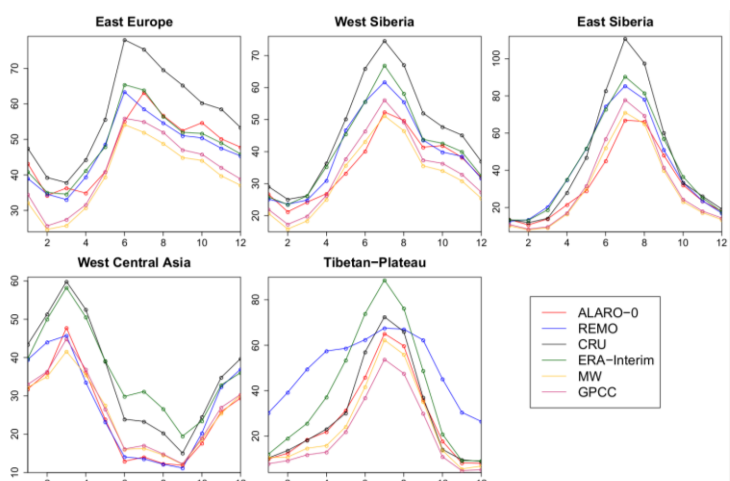

East Europe
West Siberia
East Siberia
West Central Asia
Tibetan-Plateau
ALARO-0
REMO
CRU
ERA-Interim
MW
GPCC

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 Performance of ALARO-0 and REMO with respect to observational spread and other RCMs

When considering the temperature biases of the RCMs with respect to CRU, larger values are partly located in regions where the range of the different reference datasets is large (> 3 °C) (Fig. 2). Some regions where ALARO-0 and REMO show a bias over 3 °C also exhibit a spread of at least 3 °C between the reference datasets (CRU, MW and ERA-Interim), resulting in an insignificant bias when compared to the spread (Fig. 2 and Fig. S2). This is for example the case over mountainous regions such as the Himalayas and Stanovoy Range, which makes it difficult to evaluate the models accurately over these mountainous regions. The observational temperature spread is larger for the East Siberian subdomain compared to East Europe and West Siberia, indicating there is a larger uncertainty for temperature evaluation over East Siberia. Significant observational uncertainties are typical over complex orography, but this does not explain why there is a larger uncertainty over the complete East Siberian subdomain. New et al. (1999) mentioned that CRU contains colder temperatures in winter over Russia, which could explain this larger spread.

However, not all RCM biases are located within the spread of the reference datasets. The strong biases in the north-eastern part of the domain for ALARO-0 during winter and spring exceed for example the range in temperatures between the different reference datasets, indicating that ALARO-0 is not able to simulate the temperatures accurately over this region (Fig. S2). Furthermore, the smaller biases for both RCMs over East Europe (< 3 °C) are not situated within the small (< 1°) range of the reference datasets (Fig. S2). The biases over western Siberia are not within the range of the reference datasets either, except for ALARO-0 during autumn. Fig. S2 shows that for the majority of grid points the mean temperatures of ALARO-0 and REMO lie within the range of spread between the reference datasets during autumn. From this we conclude that both RCMs simulate temperatures fairly well in autumn. During winter and spring none of the RCMs are able to reproduce temperatures that can be completely explained by the observational uncertainty over a large part of the CAS-CORDEX domain, while this is also the case for ALARO-0 during summer (Fig. 2 and Table 2).

When comparing the mean spatial biases and MAE for the 1980-2017 period (Table 2), it is seen that in most cases the differences between the observational datasets are smaller than the differences between the RCMs and CRU. However, the MAE and spatially averaged bias are smaller for both RCMs than for MW during autumn over the West Siberia subdomain since both RCMs perform well over Kazakhstan with grid points with biases between -1 °C and 1°C. Moreover, REMO has lower MAE values than MW over the East Siberian subdomain during summer and autumn and over the West Central Asian subdomain during winter. ALARO-0 has lower MAE values than MW during autumn over the Tibetan Plateau subdomain.

Formatted: Not Highlight

The Taylor diagrams of temperature (Fig. 3) show that the normalized standard deviation of ERA-Interim and MW differs less from CRU than the RCMs, except for REMO over the EEU and ESB subdomains during summer and for ALARO-0 over ESB during autumn and WSB and TIB during winter. This smaller difference between the reference datasets 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 lower than or close to those between CRU and the RCMs for the subdomains WCA and TIB, which indicates that the RCMs are able to reproduce the spatial temperature patterns within the range of observational uncertainty, even though they slightly deviate from the spatial temperature patterns in the CRU data. It is seen that the observed spatial patterns are less reliable during summer over East Siberia since the MW and ERA-Interim both show a lower spatial correlation (< 90 % for ERA-Interim) with CRU during summer compared to the other seasons. However, the lower spatial correlation of the RCMs during summer over East Siberia can only partly be explained by the observational uncertainty in spatial correlation of temperatures.

Similar to our findings, Ozturk et al. (2016) reported a lower spatial correlation during summer over the complete CAS-CORDEX domain with RegCM4.3.5 at 0.50° horizontal resolution. Additionally, similarly 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). For summer temperatures, Russo et al. (2019) found that COSMO-CLM 5.0 produces 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 are similar to ALARO-0.

In general both ALARO-0 and REMO produce biases within a similar order of magnitude as those 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). Zhu et al. (2020) conducted model runs with different land cover schemes in the WRF model over a smaller domain than CAS-CORDEX containing Kazakhstan, Uzbekistan, Kyrgyzstan, Turkmenistan and Tajikistan. None of their experiments produced biases over Kazakhstan as small as those of REMO in winter and at the annual level, while they obtained biases with different signs and similar magnitude in summer. However, it should be mentioned that they used the observational dataset from the Climate Prediction Center (CPC) which makes comparison difficult. ALARO-0 has biases with the same magnitude at the annual level as the WRF runs, but the absolute value of the biases is larger during winter and summer.

Similar to our findings, 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 partly due to the fact that observational gridded data, such as MW and CRU, are based on measurements of meteorological stations in the valleys (New et al., 1999). The gridded observations are thus less reliable over the Himalayas and Tibetan Plateau, creating a larger observational uncertainty, and resulting in large biases of the RCMs that lie within the range of observational uncertainty in most of the grid points(Fig S2). 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).

Formatted: Not Highlight

~~4.1.1 CAS-CORDEX domain~~

~~When comparing the above results of temperature with the other reference datasets (Fig. 3), the normalized standard deviation of ERA-Interim and MW differs less from CRU than the RCMs during spring and summer. 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 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

In this section the temperature biases over snow covered areas during winter and spring will be explained. As mentioned in the previous sections, both RCMs have large temperature biases in the northern part of the domain that are not within the range of the reference datasets during winter and spring (Fig. S2). During winter, ALARO-0 simulates warm biases over the northern part of the domain and REMO simulates cold biases over the north-western part of the domain, while in spring they both show a cold bias over the north (Fig. 2 and 4).

~~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 exceedingly ~~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 more complex ~~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 connected ~~related~~ to the snow scheme as we find a delay in the springtime melting of the snowpack (not shown). Additionally, ALARO-0 simulates exceedingly ~~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 related ~~coupled~~ to the difficulties with the snow cover.

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.

Formatted: Highlight

4.2 Diurnal temperature range

Similar to the mean temperature the observational spread for minimum and maximum temperature is larger in the orographically complex regions (Fig. 5 and Fig. 7). ALARO-0 and REMO are not able to reproduce the minimum and maximum temperature since they produce biases that are outside this significant observational range (e.g. the range for maximum temperature is 5 °C to 7 °C in the north-eastern part of the domain in spring) (Fig. S3 and S4). However, during summer REMO simulates minimum and maximum temperatures within the observational range over western Russia. The MAE of REMO for minimum and maximum temperatures is acceptable during summer over East Europe and the West Siberia subdomains since the MAE between ERA-Interim and CRU is larger (Table 3 and 4). Moreover, the MAE of REMO for maximum temperature is lower than the MAE of ERA-Interim over the WCA domain, indicating that REMO is able to produce maximum temperatures over this subregion within the range of the reference datasets.

Both RCMs produce a smaller daily temperature range, resulting in biases that are generally warmer for the minimum temperature and colder for the maximum temperature, when compared to those of the mean temperature (Fig. 2, 5, 7 and Tables 2, 3 and 4). The smaller daily temperature range causes a stronger warm bias in winter for the minimum temperature and a stronger cold bias for maximum temperature in spring, which is notably visible in the northern part of the domain for the ALARO-0 model (Fig. 2, 5, 7 and Table 2, 3 and 4). Additionally, it is seen that the cold bias in the north during spring for the ALARO-0 model is weaker for the minimum temperature than for the mean temperature, while 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. Moreover, the smaller daily temperature range causes larger MAE scores for minimum temperature during winter and for maximum temperature during spring, except for ALARO-0 over the WCA and TIB subdomain (Table 3 and 4). This indicates that minimum temperatures are less accurately simulated by both RCMs compared to temperature during winter, while maximum temperatures are simulated less accurately during spring.

The underestimation of the diurnal range is similar to the findings over other regions (Laprise et al., 2003; Kyselý and Plavcová 2012) and was also observed over the CAS-CORDEX domain by Russo et al. (2019). Their RCM produced smaller diurnal ranges compared to different observational datasets. In particular ALARO-0 shows a smaller 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).

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. Additionally, both minimum and maximum temperatures have a similar temporal pattern as the mean temperature (Fig. 4).

The metrics in Fig. 6 and 8 show that spatial pattern correlations of ERA-Interim deviate more from CRU for minimum and maximum temperature compared to mean temperature (Fig. 3). This larger uncertainty makes it harder to draw sound conclusions from the lower spatial pattern correlations of ALARO-0 and REMO.

The evaluation of temperature and its diurnal cycle shows that a bias adjustment is essential before the climate data is applied in impact modelling. However, REMO simulates mean and maximum temperatures well over the West Central Asia subdomain when the observational range is taken into account~~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~~.

4.3 Precipitation

Compared to the RegCM4.3.5 model (Ozturk et al., 2016) ALARO-0 has lower RMSEs over all seasons and REMO has higher RMSEs, excluding summer (Fig. 10). 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.

For the majority of the grid points, the precipitation of ALARO-0 and REMO is situated within the spread of the different gridded datasets for all seasons (Fig. S6). However, there are some subregions where the precipitation of ALARO-0 and/or REMO exceeds the observational spread during one or more seasons. For example, both RCMs show slightly lower precipitation amounts in summer over West Central Asia compared to the different reference datasets (Fig. 11 and S6). Additionally, the overestimation in precipitation by both RCMs in the East Asian monsoon region exceeds the observational spread, especially in winter and spring for REMO and in spring and autumn for ALARO-0, indicating that the models do not completely capture the East Asian monsoon system. Moreover, the ALARO-0 model overestimates the precipitation significantly over the East European subdomain during all seasons when compared to the spread of the reference datasets (Fig S6 and 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. For example, an extreme excess of precipitation was simulated over the East Asian monsoon region, with a smaller relative wet bias in summer. Additionally, they obtained a dry bias in summer over the western part of the domain which is similar for REMO, while ALARO-0 shows only a dry bias in the south-western part of the domain. Moreover, ALARO-0 produces a dry bias over the north-eastern part of the domain during summer, while this is not the case for the other RCMs (REMO, COSMO-CLM 5.0 and RegCM4.0) (Ozturk et al., 2012; Russo et al., 2019). The underestimation in precipitation by ALARO-0 during spring and summer in the north-eastern part of the domain might be related to the Siberian High that remains too strong (not shown).

Table 5 and Fig. 11 show that on average, CRU contains higher precipitation amounts than the two other observational datasets, MW and GPCC. As mentioned before, it is known that the MW and GPCC datasets generally underestimate the seasonal precipitation over Central Asia, especially during spring for the central part of the CAS-CORDEX domain (Hu et al., 2018).

The overestimation of the annual precipitation by the RCMs over the Himalaya, Altay, Tian Shan and Kunlun Mountains is partly due to the fact that gridded observational datasets CRU, MW and GPCC underestimate the precipitation over these mountainous regions. It is known that the accuracy of gridded precipitation datasets decreases with elevation, especially over an altitude of 1500 m (Zhu et al., 2015). By contrast, ERA-Interim generally overestimates the precipitation, particularly over mountainous regions (Sun et al., 2018). Moreover, a similar pattern of an underestimation by gridded observational datasets and overestimation by reanalysis data is present over the Tibetan Plateau (Sun et al., 2018), causing larger biases (Fig. 9 and Fig. 11). The discrepancy between the observational gridded datasets and the ERA-Interim reanalysis data (Fig. 9 and Fig. 11) explains why the strong wet biases of the RCMs compared to CRU over the mountainous areas and Tibetan Plateau are not significant (Fig. S6). The pronounced difference between the observational and reanalysis datasets makes it difficult to draw sound conclusions over these regions.

Even when taking into account the large spread between the reference datasets, REMO is not able to reproduce the annual cycle of precipitation over the Asian monsoon region. Remedio et al. (2019) also found a wet bias for REMO at the annual level over the subtropical region where the Asian monsoon takes place.

In the north, the precipitation amounts of REMO bear more resemblance to those of ERA-Interim and COSMO-CLM 5.0 described by Russo et al. (2019) (Fig. S7). This similarity is probably due to the fact that they all use a 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.

It can be concluded that for the different subregions and seasons, REMO and ALARO-0 simulated precipitation 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 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). This lower correlation 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 reduced in the future to improve the evaluation of precipitation (Russo et al., 2019).~~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).~~

5 Conclusion

The evaluation over the CAS-CORDEX domain of ALARO-0 and REMO, run at 0.22° resolution, showed that in general both RCMs reproduced realistic spatial patterns for temperature since there is a high spatial correlation with observational data. Additionally, the values of spatial variation for mean temperature of both RCMs correspond closely to the values obtained with other reference datasets. When evaluating the modelled precipitation, poorer scores were obtained for these metrics but the spread between the different observational datasets is also larger for precipitation as compared to temperature.

Both RCMs performed best during autumn for temperature and precipitation, showing biases within the range of the observational uncertainty for the majority of the CAS-CORDEX domain. 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 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 behavior is probably linked to limitations of the used snow scheme. REMO produced excessive precipitation amounts over the Tibatian Plateau subregion during all seasons and incorrectly simulated the annual cycle of the East Asian Monsoon system. In general, REMO was better than ALARO-0 in reproducing the seasonal mean temperatures, except during autumn, whereas ALARO-0 estimated the precipitation well.

Additionally, 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.

We conclude that REMO and ALARO-0 can be used for climate modelling over Central Asia e.g. for precipitation and temperature over West Central Asia. However, the deficiencies of both models over Central-Asia described in this evaluation study should be kept in mind. Climate data produced by both RCMs can only be used for impact studies if a suitable bias adjustment is applied for those subregions where the RCMs perform less well e.g. temperature over East Siberia and precipitation over the Tibetan Plateau~~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~~.

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.

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, 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.

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.

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.

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., Türkeş, 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.

Sun, Q., Miao, C., Duan, Q., Ashouri, H.,Sorooshian, S., and Hsu, K.-L.: A review of global precipitation data sets: Data sources, estimation, and inter-comparisons.Reviews of Geophysics, 56, 79–107. doi:10.1002/2017RG000574, 2018.

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.

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.

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 - 2018	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	0.70°	2 m mean air temperature, precipitation	monthly	1979 - 2017	global

Formatted: No

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

Formatted: Le
Bottom: 1,65 c

	~~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~~

	EEU					WSB					ESB				
	DJF	MAM	JJA	SON	Annual	DJF	MAM	JJA	SON	Annual	DJF	MAM	JJA	SON	Annual
CRU	-10.01	5.09	19.08	4.77	4.8	-15.44	2.39	18.16	2.13	1.89	-24.29	-2.34	15.35	-3.66	-3.64
REMO - CRU	-1.53	-1.42	-1.06	-0.46	-1.11	-0.40	-0.94	-1.22	-0.52	-0.77	3.11	-0.42	-0.13	0.90	0.86
MAE REMO CRU	1.85	2.06	1.11	0.72	1.31	1.94	1.95	1.33	0.86	1.28	3.40	1.78	0.71	1.25	1.40
ALARO - CRU	3.27	-4.35	-1.56	-0.44	-0.79	4.57	-5.26	-2.16	-0.14	-0.77	1.26	-6.90	0.63	0.57	-1.12
MAE ALARO CRU	3.28	4.36	2.32	0.66	1.22	4.87	5.31	2.79	0.51	1.18	3.97	6.99	2.09	1.45	1.65
ERA-Interim - CRU	0.24	-0.10	-0.15	-0.23	-0.06	0.41	0.06	-0.19	-0.29	-0.01	1.68	1.04	0.49	0.41	0.91
MAE ERA-Interim CRU	0.41	0.3	0.43	0.31	0.25	0.85	0.53	0.62	0.49	0.43	1.94	1.25	0.83	0.80	1.10
MW - CRU	0.01	-0.42	-0.39	-0.49	-0.32	-0.20	-0.46	-0.36	-0.65	-0.42	0.08	0.12	-0.14	-0.26	-0.05
MAE MW CRU	0.46	0.52	0.56	0.46	0.46	0.88	0.78	0.73	0.88	0.88	1.55	0.96	0.94	1.55	1.55

	WCA					TIB					CAS-CORDEX				
	DJF	MAM	JJA	SON	Annual	DJF	MAM	JJA	SON	Annual	DJF	MAM	JJA	SON	Annual
CRU	2.25	14.34	25.98	14.89	14.42	-9.79	3.69	14.36	3.05	2.88	-9.35	5.87	19.23	5.72	5.44
REMO - CRU	-0.11	-0.05	0.57	0.22	0.16	-0.07	-1.49	-1.16	-0.90	-0.90	0.48	-0.56	-0.33	0.01	-0.11
MAE REMO CRU	1.48	1.64	2.03	1.46	1.47	3.31	2.76	2.50	2.37	2.59	2.33	1.82	1.34	1.20	1.43
ALARO - CRU	-2.13	-0.38	1.70	-0.41	-0.29	-2.57	-1.04	1.29	-0.28	-0.63	0.83	-3.19	0.02	-0.03	-0.60

MAE ALARO CRU	2.77	2.38	2.79	1.59	1.81	3.24	2.92	3.25	1.94	2.32	3.16	4.20	2.42	1.24	1.56
ERA-Interim - CRU	-0.03	0.11	0.32	0.07	0.12	-0.46	-0.62	-0.60	-0.82	-0.62	0.42	0.21	0.16	-0.02	0.19
MAE ERA-Interim CRU	1.26	1.27	1.58	1.21	1.17	1.77	1.95	2.02	1.80	1.77	1.16	1.02	0.98	0.85	0.87
MW - CRU	-0.09	-0.23	0.08	-0.09	-0.08	-0.46	0.75	0.56	0.14	0.26	-0.41	-0.19	-0.09	-0.43	-0.28
MAE MW CRU	1.53	1.38	1.48	1.53	1.53	2.78	2.22	2.12	2.78	2.78	1.32	1.10	1.07	1.32	1.32

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

	~~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~~

	EEU					WSB					ESB				
	DJF	MAM	JJA	SON	Annual	DJF	MAM	JJA	SON	Annual	DJF	MAM	JJA	SON	Annual
CRU	-13.56	-0.03	13.3	0.99	0.24	-20	-3.26	12.24	-2.48	-3.3	-30.12	-9.47	8.78	-9.27	-9.93
REMO - CRU	-2.21	-1.29	0.05	0.35	-0.77	-0.67	-1.16	-0.32	0.47	-0.42	3.64	0.87	1.77	2.48	2.18
MAE REMO CRU	2.73	2.17	0.56	0.90	1.42	2.38	2.24	0.82	1.37	1.49	4.13	2.40	1.86	2.66	2.49
ALARO - CRU	5.10	-3.21	-0.79	0.45	0.37	7.15	-4.02	-1.51	1.26	0.69	4.74	-3.92	2.18	2.79	1.43
MAE ALARO CRU	5.11	3.26	2.45	0.67	0.88	7.24	4.07	2.78	1.36	0.97	5.35	4.10	3.00	2.86	1.73
ERA-Interim - CRU	0.24	-2.21	-1.38	-0.23	-0.90	0.81	-2.53	-1.19	0.86	-0.52	2.32	-0.83	1.85	2.18	1.38
MAE ERA-Interim CRU	1.35	2.24	1.50	0.56	1.00	1.60	2.60	1.42	0.96	0.88	2.73	1.38	2.02	2.25	1.62
	WCA					TIB					CAS-CORDEX				
	DJF	MAM	JJA	SON	Annual	DJF	MAM	JJA	SON	Annual	DJF	MAM	JJA	SON	Annual
CRU	-3.02	7.93	18.54	7.84	7.87	-16.76	-3.35	7.76	-4.03	-4.04	-14.43	-0.22	13.18	0.40	-0.20
REMO - CRU	0.68	0.00	1.07	1.57	0.83	1.00	-1.70	-0.61	0.55	-0.19	0.77	-0.25	0.60	1.09	0.55
MAE REMO CRU	2.4	2.10	2.56	2.60	2.29	4.31	3.44	2.29	2.90	2.98	3.02	2.22	1.52	1.96	1.97
ALARO - CRU	-1.00	0.34	3.05	1.27	0.92	-0.26	0.09	2.44	1.32	0.91	2.85	-1.71	1.10	1.42	0.90
MAE ALARO CRU	2.43	2.60	3.82	2.30	2.31	2.80	3.06	3.86	2.55	2.71	4.07	3.21	2.93	1.88	1.59
ERA-Interim - CRU	-0.84	-0.98	0.22	0.80	-0.19	-0.13	-1.44	-0.46	0.47	-0.39	0.39	-1.46	0.00	0.79	-0.08
MAE ERA-Interim CRU	1.95	1.70	1.68	1.89	1.46	2.11	2.30	2.18	2.14	1.90	1.90	1.96	1.63	1.46	1.33

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

	~~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~~

	EEU					WSB					ESB				
	DJF	MAM	JJA	SON	Annual	DJF	MAM	JJA	SON	Annual	DJF	MAM	JJA	SON	Annual
CRU	-6.50	10.23	24.91	8.57	9.38	-10.94	8.04	24.13	6.74	7.08	-18.52	4.78	21.97	1.93	2.64
REMO - CRU	-1.58	-2.27	-2.42	-1.25	-1.89	-0.87	-1.77	-2.44	-1.59	-1.67	2.03	-2.42	-1.62	-0.74	-0.70
MAE REMO CRU	1.67	2.77	2.43	1.27	1.90	2.03	2.61	2.50	1.69	1.91	2.50	2.81	1.77	1.01	1.33
ALARO - CRU	1.34	-6.06	-3.36	-1.47	-2.41	1.97	-7.10	-3.83	-1.68	-2.69	-1.85	-9.87	-1.28	-1.51	-3.64
MAE ALARO CRU	1.40	6.06	3.47	1.49	2.46	2.54	7.14	3.97	1.71	2.74	3.90	9.94	2.22	1.78	3.80
ERA-Interim - CRU	-0.48	-2.65	-3.02	-1.33	-1.88	-0.47	-2.13	-2.63	-0.39	-1.41	-0.65	-4.17	-1.14	-0.64	-1.66
MAE ERA-Interim CRU	0.92	2.65	3.04	1.36	1.88	1.21	2.20	2.75	0.90	1.55	1.78	4.20	1.40	0.99	1.77

	WCA					TIB					CAS-CORDEX				
	DJF	MAM	JJA	SON	Annual	DJF	MAM	JJA	SON	Annual	DJF	MAM	JJA	SON	Annual
CRU	7.53	20.8	33.47	21.98	21.01	-2.86	10.73	21.00	10.13	9.81	-4.29	11.97	25.34	11.06	11.09
REMO - CRU	-0.04	0.18	0.26	0.07	0.11	-1.13	-1.90	-1.15	-1.77	-1.49	0.08	-1.24	-1.07	-0.71	-0.74
MAE REMO CRU	1.49	1.66	2.00	1.43	1.44	3.22	3.23	2.56	2.88	2.84	2.15	2.49	2.08	1.48	1.75
ALARO - CRU	-2.31	-1.24	0.15	-1.32	-1.18	-3.68	-2.20	-0.07	-1.47	-1.85	-0.77	-4.84	-1.46	-1.24	-2.08
MAE ALARO CRU	2.73	2.28	2.16	1.87	1.89	3.96	3.12	2.74	2.24	2.59	2.63	5.54	2.79	1.70	2.61
ERA-Interim - CRU	-1.25	-1.10	-1.14	-1.33	-1.21	-0.93	-2.03	-1.90	-0.90	-1.45	-0.84	-2.43	-1.77	-0.80	-1.46
MAE ERA-Interim CRU	2.02	1.76	1.86	1.66	1.61	2.01	2.61	2.93	2.28	2.23	1.53	2.86	2.34	1.36	1.82

Table 5: Climatological mean CRU precipitation (mm month^{-1}) for the 1980-2017 period over the CAS-CORDEX domain and subdomain, and relative biases (%) and MAE (%) 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~~

	EEU					WSB					ESB				
	DJF	MAM	JJA	SON	Annual	DJF	MAM	JJA	SON	Annual	DJF	MAM	JJA	SON	Annual
CRU	34.91	34.16	55.26	45.62	42.51	22.74	27.99	51.53	35.94	34.6	11.13	22.10	72.28	29.62	33.90
REMO - CRU	12	20	7	9	11	16	25	13	14	16	30	63	8	21	22
MAE REMO CRU	18	22	21	13	14	33	34	28	26	25	133	74	17	37	28
ALARO - CRU	21	12	10	18	15	20	3	-4	17	7	35	-1	-19	21	-3
MAE ALARO CRU	25	17	22	19	16	28	17	22	22	15	65	24	28	30	19
ERA-Interim - CRU	13	19	10	9	12	18	27	16	15	18	29	57	11	31	24
MAE ERA-Interim CRU	18	20	11	10	13	25	29	19	19	21	79	66	16	36	26
MW - CRU	-11	-7	-7	-6	-7	-8	-5	-8	-6	-7	-4	-15	-13	-9	-12
MAE MW CRU	14	10	10	14	14	17	14	15	17	17	33	23	16	33	33
GPCC - CRU	-24	-15	-7	-11	-13	-12	-11	-4	-8	-8	-7	-21	-9	-13	-12
MAE GPCC CRU	24	17	11	24	24	23	18	10	23	23	30	26	12	30	30
	WCA					TIB					CAS-CORDEX				
	DJF	MAM	JJA	SON	Annual	DJF	MAM	JJA	SON	Annual	DJF	MAM	JJA	SON	Annual
CRU	33.18	37.52	16.74	18.45	26.46	8.12	17.73	48.56	15.02	22.45	22.60	32.34	64.75	35.50	38.88
REMO - CRU	17	-10	-19	18	2	259	194	31	187	110	29	39	4	20	18
MAE REMO CRU	45	46	66	43	39	1169	638	243	240	137	205	107	52	53	39
ALARO - CRU	-2	-5	-18	9	-4	26	36	14	38	23	22	19	1	22	13

MAE ALARO CRU	32	33	78	44	33	260	279	185	107	84	73	54	49	42	30
ERA-Interim - CRU	21	29	77	38	36	59	117	63	73	75	22	38	19	21	24
MAE ERA-Interim CRU	32	33	123	51	34	267	384	340	131	104	80	72	63	40	32
MW - CRU	-4	-8	-2	7	-3	14	3	9	20	10	-6	-4	-3	-2	-3
MAE MW CRU	32	28	81	32	32	104	100	64	104	104	39	27	31	39	39
GPCC - CRU	0	-7	-7	-2	-4	-9	-17	-4	-2	-7	-7	-8	-1	-5	-4
MAE GPCC CRU	31	24	55	31	31	88	90	61	88	88	39	27	28	39	39

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/)

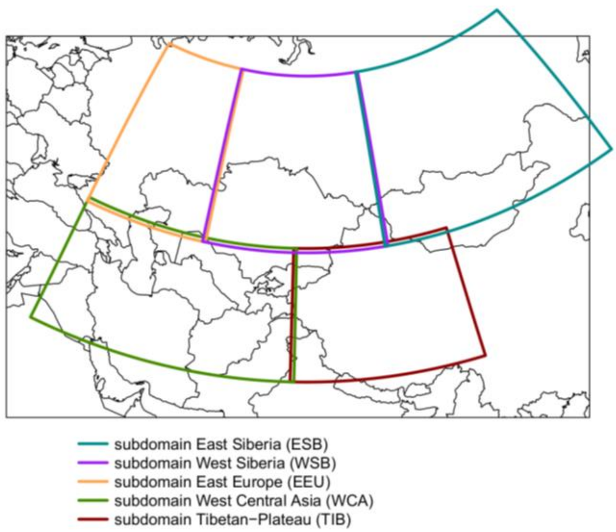


Figure S1: IPCC6 subdomains projected on the CAS-CORDEX region.

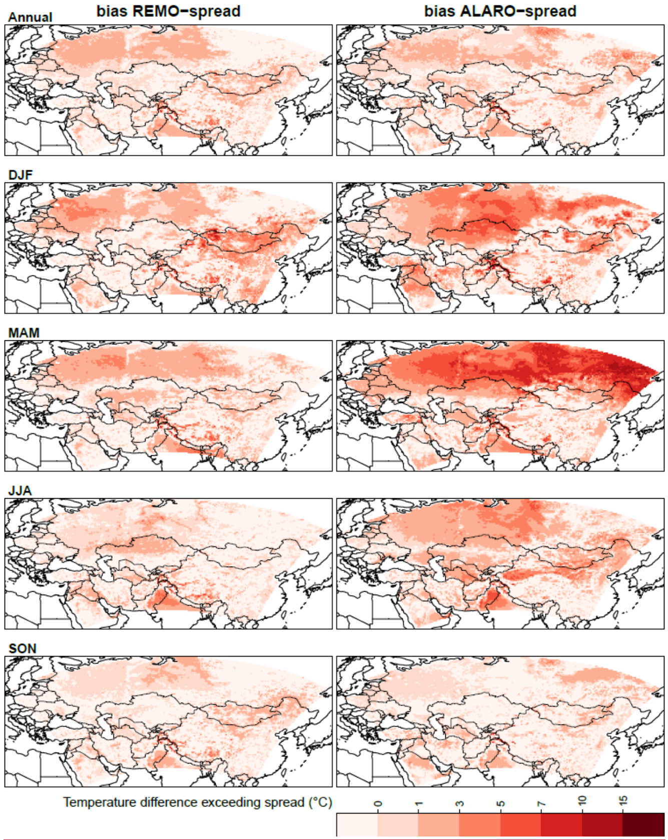
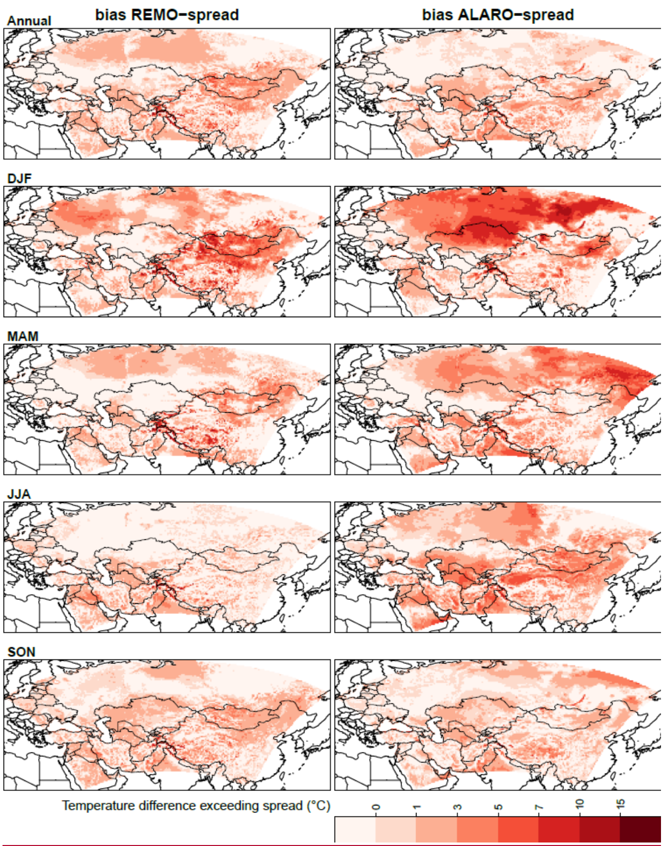


Figure S2: Difference between absolute value of bias and observational spread for the variable mean temperature (°C) of RCMs REMO and ALARO-0.

Formatted: Centered



10 **Figure S3: Difference between absolute value of bias and observational spread for the variable minimum temperature (°C) of RCMs REMO and ALARO-0.**

Formatted: Centered

Formatted: Caption

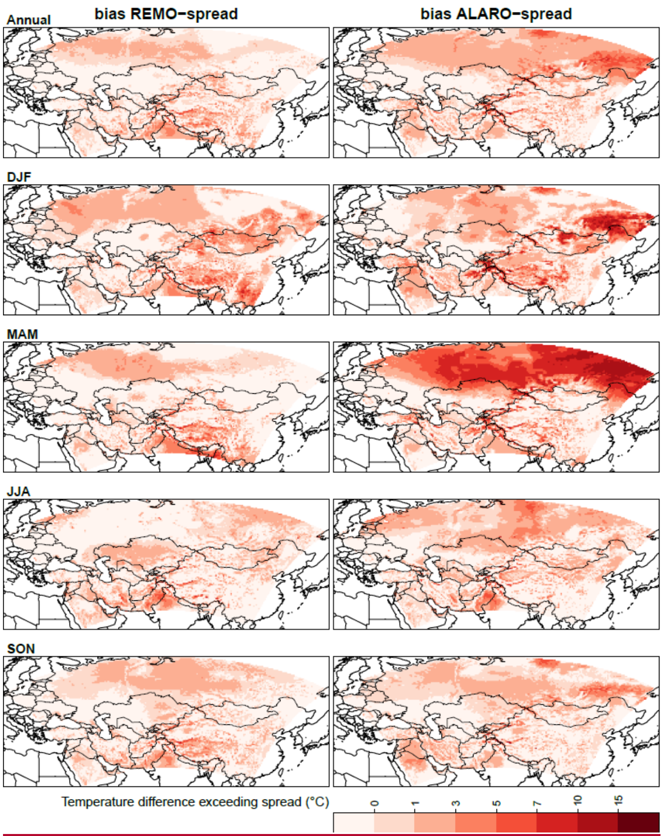
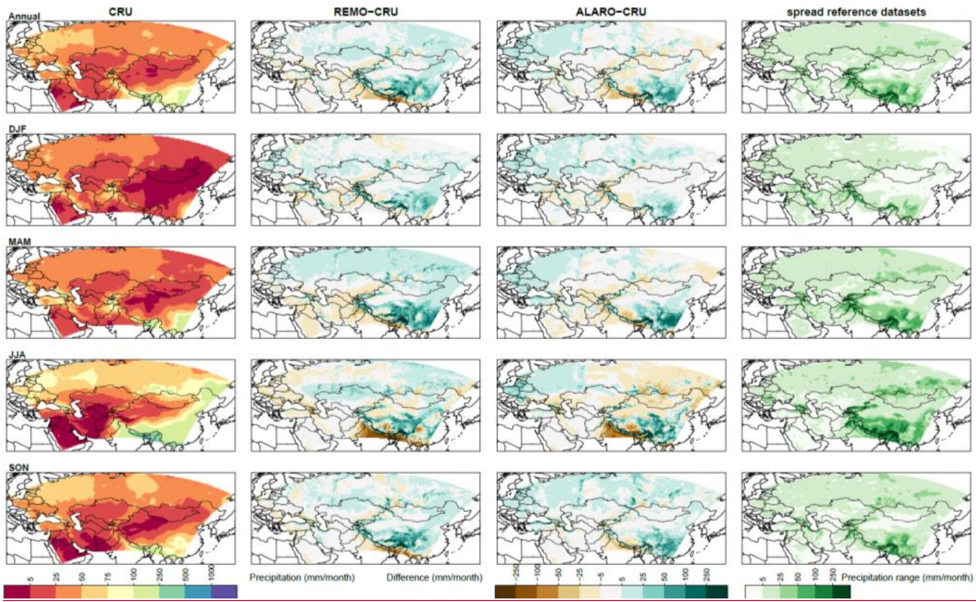


Figure S4: Difference between absolute value of bias and observational spread for the variable maximum temperature (°C) of RCMs REMO and ALARO-0.

Formatted: Centered

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



CRU
REMO−CRU
ALARO−CRU
spread reference datasets
Annual
DJF
MAM
JJA
SON
5
25
50
75
100
250
500
1000
Precipitation (mm/month)
Difference (mm/month)
−250
−100
−50
−25
−5
5
25
50
100
250
Precipitation range (mm/month)

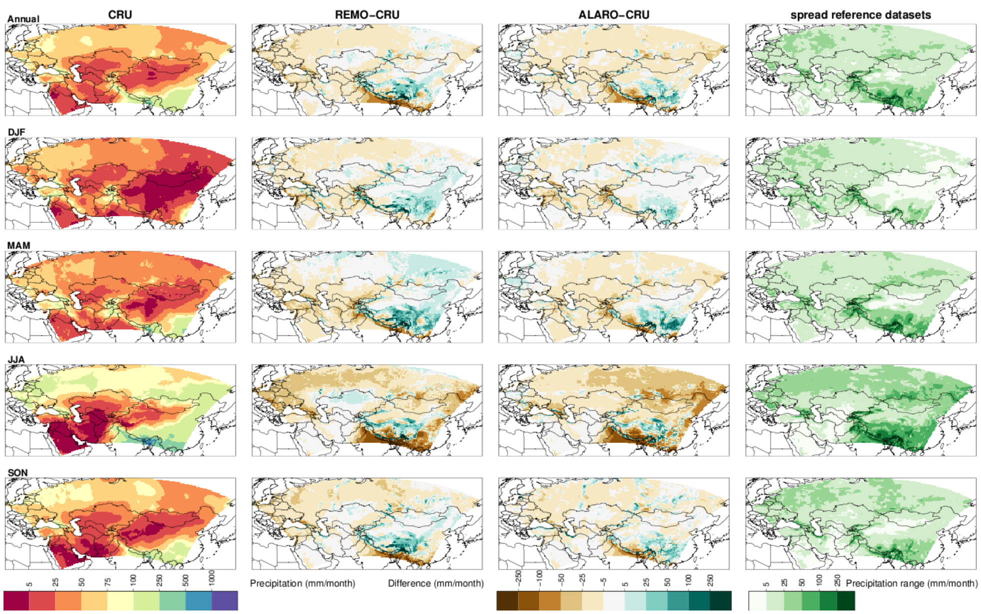


Figure S5~~2~~: Absolute difference between the average seasonal and annual CRU precipitation (mm month^{-1}) and the precipitation simulated by REMO and ALARO-0 over the 1980-2017 period.

Formatted: Superscript

Table S2: Climatological mean CRU precipitation (mm month^{-1}) for the 1980-2017 period over the CAS-CORDEX domain and subdomains, and absolute biases (mm month^{-1}) and MAE (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: En

	EEU					WSB					ESB				
	DJF	MAM	JJA	SON	Annual	DJF	MAM	JJA	SON	Annual	DJF	MAM	JJA	SON	Annual
CRU	34.91	34.16	55.26	45.62	42.51	22.74	27.99	51.53	35.94	34.60	11.13	22.10	72.28	29.62	33.90
REMO - CRU	4.18	6.83	4.02	4.31	4.84	3.73	7.10	6.72	4.96	5.64	3.33	14.01	5.73	6.35	7.38
MAE REMO CRU	5.62	7.86	8.64	5.61	5.62	6.23	9.83	10.73	7.69	7.31	5.18	15.00	11.37	9.09	8.88
ALARO - CRU	7.45	3.95	5.50	8.33	6.29	4.55	0.90	-2.21	6.05	2.30	3.91	-0.26	-13.52	6.08	-0.99
MAE ALARO CRU	8.04	5.72	10.58	8.73	7.18	5.93	4.66	8.11	7.59	4.85	5.24	5.14	18.38	8.65	5.96
ERA-Interim - CRU	4.49	6.53	5.75	3.98	5.19	3.99	7.46	8.35	5.50	6.34	3.26	12.61	8.19	9.07	8.30
MAE ERA-Interim CRU	5.18	6.59	6.62	4.46	5.33	4.80	8.47	9.41	6.10	6.97	4.08	12.88	11.21	9.71	8.84
MW - CRU	-3.69	-2.33	-3.69	-2.89	-3.14	-1.75	-1.47	-4.13	-2.00	-2.34	-0.42	-3.42	-9.59	-2.72	-4.05
MAE MW CRU	4.49	3.07	4.44	4.49	4.49	3.48	3.69	6.10	3.48	3.48	2.09	4.42	11.04	2.09	2.09
GPCC - CRU	-8.21	-5.23	-4.05	-5.19	-5.65	-2.70	-3.19	-1.81	-2.81	-2.63	-0.81	-4.59	-6.57	-3.72	-3.94
MAE GPCC CRU	8.82	5.68	5.85	8.82	8.82	4.88	5.02	4.70	4.88	4.88	2.38	5.15	8.24	2.38	2.38

	WCA					TIB					CAS-CORDEX				
	DJF	MAM	JJA	SON	Annual	DJF	MAM	JJA	SON	Annual	DJF	MAM	JJA	SON	Annual
CRU	33.18	37.52	16.74	18.45	26.46	8.12	17.73	48.56	15.02	22.45	22.60	32.34	64.75	35.50	38.88
REMO - CRU	5.77	-3.59	-3.20	3.24	0.53	21.07	34.40	15.23	28.07	24.70	6.55	12.45	2.47	6.98	7.12
MAE REMO CRU	17.57	16.96	7.87	8.51	11.32	24.04	39.38	32.92	30.72	30.47	10.85	18.88	18.13	12.80	13.56
ALARO - CRU	-0.71	-1.75	-3.00	1.61	-0.96	2.15	6.37	6.85	5.64	5.26	5.04	6.14	0.74	7.82	4.93

MAE ALARO CRU	11.24	11.63	11.06	8.09	9.13	7.83	16.36	32.96	12.83	16.29	8.31	12.75	19.82	11.69	11.41
ERA-Interim - CRU	6.90	10.79	12.85	7.02	9.41	4.76	20.81	30.60	10.98	16.86	4.88	12.37	12.50	7.61	9.37
MAE ERA-Interim CRU	10.46	14.20	15.02	8.85	11.13	7.88	23.20	39.94	13.36	19.98	6.85	14.00	17.27	9.50	11.08
MW - CRU	-1.22	-2.98	-0.33	1.20	-0.83	1.11	0.53	4.14	3.07	2.21	-1.28	-1.25	-1.75	-0.84	-1.28
MAE MW CRU	9.75	9.57	5.00	9.75	9.75	6.13	9.50	20.30	6.13	6.13	4.56	6.08	11.10	4.56	4.56
GPCC - CRU	0.08	-2.50	-1.21	-0.38	-1.01	-0.74	-3.08	-2.11	-0.30	-1.57	-1.65	-2.65	-0.92	-1.89	-1.77
MAE GPCC CRU	10.44	8.73	5.03	10.44	10.44	5.88	9.81	20.44	5.88	5.88	5.82	6.52	10.64	5.82	5.82

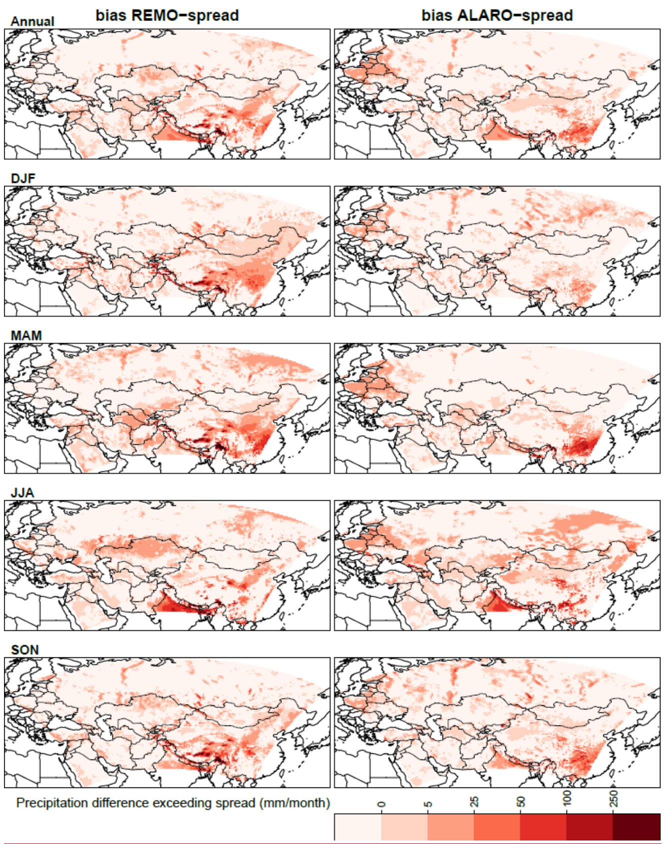


Figure S6: Difference between absolute bias and observational spread for the variable precipitation of RCMs REMO and ALARO-0.

Formatted: Centered

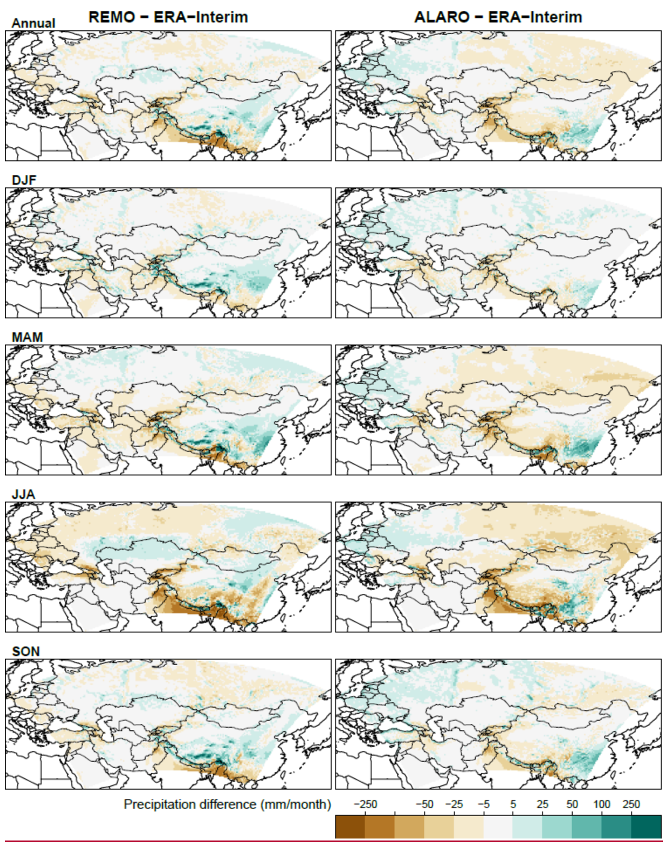


Figure S7: Absolute difference between the average seasonal and annual ERA-Interim precipitation (mm month^{-1}) and the precipitation simulated by REMO and ALARO-0 over the 1980-2017 period.

Formatted: Centered

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

Formatted: Superscript

Formatted: Normal

Table S3~~2~~: Overview of the identifiers on the ESGF data platform (data node: esgf1.dkrz.de) of the used ALARO-0 and REMO RCM climate data.

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