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

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

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

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

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

5   Regional Climate Models (RCMs) play a fundamental role.

In this paper, the results of a set of sensitivity experiments with the regional climate model COSMO-CLM version 5.0, for the Central Asia CORDEX domain, are presented. Starting from a reference model setup, general model performance is evaluated for present-days, testing the effects of a set of singular physical parameterizations and their mutual interaction on the simulation of monthly and seasonal values of three variables that are important for impact studies: 2-meter temperature, precipitation and

10 diurnal temperature range. The final goal of this study is two-fold: having a general overview of model performance and its uncertainties for the considered region and determining at the same time an optimal model configuration.

Results show that the model presents remarkable deficiencies over different areas of the domain. The combined change of the albedo taking into consideration the ratio of forest fractions and the soil conductivity taking into account the ratio of liquid water and ice in the soil, allows to achieve the best improvements in model performance in terms of climatological means.

15 Importantly, the model seems to be particularly sensitive to those parameterizations that deal with soil and surface features, and that could positively affect the repartition of incoming radiation. The results for the mean climate appear to be independent of the observational dataset used for evaluation and of the boundary data employed to force the simulations. On the other hand, due to the large uncertainties in the variability estimates from observations, the use of different boundaries and the model internal variability, it has not been possible to rank the different simulations according to their representation of the monthly

20 variability.

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

1 Introduction

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

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

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

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

such as Mongolia, are particularly prone to the harmful effects of climate change, due to their limited water resources. These regions have already been affected by long and extreme droughts in the last decades, with alarming risks for agricultural areas. Future climatic conditions will likely be drier and warmer, with a significant impact on water resources, food and biodiversity (Chuluunkhuyag, 2008). Interestingly, for Mongolia, studies have suggested that beside the effect on water availability, climate

5 change will also affect local herders, with the number of people migrating due to enviromentally-induced economic reasons increasing in the near future (Diniega, 2012). Another region of the CORDEX Central Asia domain whose unique and delicate ecosystem has already been highly affected by recent climate changes is Siberia. In particular, in the last decades, the eurasian tundra of western Siberia has seen a large spread and growth of shrub cover, due to climate warming (Macias-Fauria et al., 2012). Future warming could further foster these changes. All the reported studies confirm the harmful effects that climate

10 change could have on the region. In this context, reliable information on potential future changes with high spatial resolution acquire significant importance for the development of effective adaptation and mitigation strategies.

The recognized prerequisite that every climate model has to satisfy, in order to provide reliable future climate projections, is the ability of realistically simulating present-day climate (Kim et al., 2014; Nature-Editorial-Board, 2010; Kim and Lee, 2003). Assessing the ability of a climate model to simulate the current climate is defined as model evaluation (Airey and Hulme, 1995).

15 Model evaluation consists in an assessment of model quality and deficiencies originating from different modeling assumptions, conducted through the comparison of model outputs and observations (Kim et al., 2014; Kim and Lee, 2003; Flato et al., 2013; Lenderink, 2010; Overpeck et al., 2011; Bellprat et al., 2012a, b). Model evaluation is an essential part of regional model development (Kotlarski et al., 2014). Evaluation experiments normally consist in a set of present-days simulations conducted in a perfect boundary setting, i.e., using reanalysis products as lateral boundary forcings. This "modus operandi" allows for

20 the separation of possible model biases from biases due to erroneous large-scale forcings, thus highlighting specific model deficiencies (Kotlarski et al., 2014). These may be related to the model formulation and to choices in model configuration (Awan et al., 2011; Bellprat et al., 2012a, b; de Elía et al., 2008; Evans et al., 2012). In the second case, it should be possible to improve model performances by testing different model configuration setups and choosing the one that better agrees with observations. This approach might be conceived as an optimization step. Nevertheless, it is important to emphasize the fact

25 that a specific model configuration could produce better results, by simply compensating for some deficiencies in the model formulation (Hourdin et al., 2017).

A series of different aspects have to be considered for the configuration of a climate model simulation. In climate models, the complexity and small spatial scales of the physical processes involved, requires the so-called parameterization of many of these processes: this basically consists in summarizing physical phenomena and their interaction across different spatial and

30 temporal scales (Fernández et al., 2007; Rummukainen, 2010; McFarlane, 2011; Hourdin et al., 2017), which is associated with substantial uncertainties. The same processes may be described through different parameterizations, with a different degree of complexity. Consequently, the outcomes of a climate model might largely differ, depending on the parameterizations used. Additionally, the use of different forcings datasets, for example for greenhouse gases, aerosols or land cover changes, might have a significant effect on the results. Further, other details that need to be considered when configuring a climate

35 model simulation for a defined domain are the configuration and spatial resolution of the model grid (both horizontally and

vertically) and the coupling with different models representative of other components of the climate system. For regional climate models, all these aspects are domain dependent (Jacob et al., 2007, 2012; Rockel et al., 2008). This means that a regional climate modeler should always evaluate different model configurations, isolating the one that leads to a better agreement with observations, for each investigated region and employed model. In doing so, several sources of uncertainties should be taken

5 into consideration: the fact that performances of the RCM for a specific region might vary according to the boundary conditions, the model internal variability and observational datasets should be acknowledged when evaluating model performances.

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

10 experiment, we need to investigate model performances and deficiencies for the area and propose optimal model configurations.

In the light of the upcoming phase of the Coordinated Regional climate Downscaling EXperiment (CORDEX) (Giorgi et al., 2009), denominated CORDEX-CORE (Gutowski Jr et al., 2016), in this paper the results of a set of sensitivity experiments with the regional climate model COSMO-CLM version 5.0 (Rockel et al., 2008), for the Central Asia CORDEX domain, are presented. Starting from a reference model setup, general model performance is evaluated, testing the effects of a set of singular

15 physical parameterizations and their mutual interaction as well as different forcing datasets on the simulation of monthly and seasonal values of three variables that are important for impact studies. These are near surface temperature (T2M), precipitation (PRE) and diurnal temperature range (DTR), the latter representing the daily excursion between maximum and minimum temperature, which is particularly important in terms of human body adaptability and stress. The final goal of this study is two-fold: having a general overview of model performance and its uncertainties for the considered region and determining at

20 the same time a "best" suitable model configuration.

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

2 Methods

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

2.1 Model and Experiments Description

The Consortium for Small-Scale Modeling in Climate Mode (COSMO-CLM (Rockel et al., 2008)) is a non-hydrostatic regional climate model developed by the CLM-community. The model version employed in this study is the COSMO-CLM 5.0_clm9.

30 Many studies have been conducted in the recent years over different CORDEX regions, using the COSMO-CLM (Panitz et al., 2014; Dobler and Ahrens, 2010; Bucchignani et al., 2016; Smiatek et al., 2016; Jacob et al., 2014; Kotlarski et al., 2014; Zhou et al., 2016). This study represents the first application of the COSMO-CLM to the CORDEX Central Asia domain.

The simulations presented in this study are performed with a spatial resolution of 0.22^o, as specified in the new CORDEX-CORE directives (Gutowski Jr et al., 2016), on a rotated geographical grid. The initial simulation domain extends over 326 points in longitudes and 220 points in latitudes. It includes a model relaxation zone, consisting of 10 additional points on each domain side and used to "relax" the model variables towards the driving data (Køltzow, 2012; Davies, 1976). Results of the

5   simulation for this area are excluded from the analysis, with a final domain extent of 306×200 points, as shown in Fig. 1. If not differently specified, all the simulations are run over a 15 year-long period from 1991 to 2005, with the first 5 years excluded from the analysis and considered as spinup time.

In a set of sensitivity experiments labeled from **a** to **q** in the first section of Tab. 1, the effects on model performance of different physical parameterizations are tested, first individually and then combining them with each others. The setup of

10 experiment **a** is the reference from which the other experiments are configured, by implementing the modifications specified in the table. The model configuration used for the reference simulation is the same used for the CORDEX East Asia domain for the COSMO-CLM model version 5.0, available on the CORDEX page of the CLM-community website (https://www.hzg. de/ms/clm-community/076384/index.php.en). This was considered as a good reference for the purposes of this study, since the two regions share a large part of their domains. A general description of the setup of the reference simulation is provided in

15 Tab. 2.

All the performed simulations are driven by the NCEP version 2 reanalysis data (Kanamitsu et al., 2002), provided as boundary and initial conditions. The boundaries have a temporal resolution of 6 hours and a spectral resolution of T62 ($\sim 1.89^o$ lon). NCEP2 data have been selected as boundary data, instead of commonly employed ERAInterim reanalyses (Dee et al., 2011), since their spatial resolution is closer to the one of Global Circulation Models (GCMs) normally employed in CORDEX

20 simulations. In order to estimate the effects of the driving data on the simulations results and to support possible conclusions on an optimal setup, two additional simulations are performed, driven by ERAInterim reanalysis data (second section of Tab. 1), which have a spectral resolution of T255 ($\sim 0.7^o$ lon) .

A set of 4 simulations are additionally performed for the investigation of the model internal variability (third section of Tab. 1). These simulations have the same setup as the reference simulation **a**, but are initialized at four different dates, shifted by

25 +/- 1 and 3 months with respect to the reference one.

Finally, two additional 25-year long simulations, covering the period 1991-2015, are performed for testing different configurations that could help in reducing model biases over areas characterized by the presence of permafrost in winter. The two simulations, labeled **SOIL** and **SNOW** in the bottom part of Tab. 1, are performed, respectively, increasing the number of soil layers from 10 to 13, together with their total depth from approximately 15 m to more than 130 m, and using the multi-layer

30 snow model of COSMO-CLM (Machulskaya, 2015). These simulations cover a longer period than the others, since a longer spinup time is necessary in order to account for more and deeper soil layers. Their results, excluded from the direct comparison with the other simulations, are discussed in the results and concluding sections of this paper.

2.2 Observations

Gridded observational datasets are used to compare model results against observational data on a similar scale. These gridded datasets are obtained through statistical extrapolations of surface observations. In addition to uncertainties related to the original measurements, these datasets also contain important uncertainties due to the statistical extrapolation procedure (Flaounas et al.,
5 2012; Gómez-Navarro et al., 2012). For climate model evaluation studies, these uncertainties are usually taken into account by using a range of different datasets (Collins et al., 2013; Gómez-Navarro et al., 2012; Bellprat et al., 2012a, b; Flaounas et al., 2012; Lange et al., 2015; Zhou et al., 2016; Solman et al., 2013).

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

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

2.3 Analysis Details and Evaluation Metrics

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

Since in the context of CORDEX simulations the main interest is often on the comparison of the mean climate between
25 two different periods in time, the primary focus of the proposed analyses is on climatological monthly and seasonal means of the considered variables. In addition, the results are supported by the investigation of the simulated variability. In the latter case, since the model is not expected to exactly match the observed temporal evolution of the investigated variables point by point (Gleckler et al., 2008; Wilks, 2006), regional mean anomalies are considered. For each grid point in the domain, monthly anomalies are first calculated by subtracting the climatological mean from each monthly value. The variability is then analysed
30 based on these anomalies averaged over sub-regions characterized by similar climatic conditions.

The decomposition of the domain into a set of sub-regions is obtained by means of a k-means clustering (Steinhaus, 1956; Ball and Hall Dj, 1965; MacQueen et al., 1967; Lloyd, 1982; Jain, 2010) of quantile-normalized (q-normalized) monthly climatologies of the investigated variables. K-means is a clustering technique using the concept of Euclidean distance from

the centroids of a pre-determined group of clusters, for separating similar data into groups. For the purposes of this paper, following several tests and the results of other studies (Mannig et al., 2013), a total number of eleven clusters is selected. The k-means clustering algorithm is reiterated over 3000 times in order to achieve the presented results, using q-normalized values of monthly climatologies of 2-meter temperature and diurnal temperature range derived from the CRU dataset and precipitation

5 values derived from the GPCC as input. Fig. 2 shows the results of the k-means clustering. The mean climatologies of the considered regions for the three investigated variables are also reported in Tab. 4.

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

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

2.3.1 Climatology

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

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

where the Mean Absolute Error is given by:

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

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

20 **2.3.2 Variability**

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

There is no reason to expect models and observations to agree on the phasing of internal (unforced) variations. Hence metrics such as MAE are not appropriate for characterizing the model performance of interannual variability (Gleckler et al., 2008).

25 Here, for an overall evaluation of the simulated variance in the different cases, the ratio of simulated to observed variance is considered:

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

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

3   Results

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

3.1   Mean Climate

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

10   are first presented.

Fig.3 shows that for temperature, the largest biases are evident in winter (central panels), with warmer simulated conditions over the northeastern part of the domain, where the bias in some case exceeds 15^oC. The two simulations (**SOIL** and **SNOW**) specifically designed for testing the effects of changes in soil depth and the use of a multi-layer snow model on the COSMO-CLM simulation of near surface temperature over areas characterized by the presence of permafrost in winter do not present

15   significantly different results (not shown). In summer (Fig.3, right panels), a positive bias (ranging from $+5^oC$ to $+10^oC$), is present over the central and south-western part of the domain, in arid and desertic areas such as the Arabian Peninsula and the Taklamakan desert. Conversely, a cold bias is present over Siberia in this case, with values rarely below -5^oC. Biases of annual mean values (Fig.3, left panels) are smaller than in the seasonal cases, except for the Tibetan Plateau. Here a similar particularly pronounced cold bias is evident with respect to all observational datasets, with values sometimes smaller than

20   -10^oC. For all seasons, the simulation results are in better agreement with the MERRA2 dataset than with the CRU and the UDEL. Nevertheless, despite differences in the amplitude, the pattern of the bias is similar for all datasets.

Concerning precipitation (Fig.4), remarkable biases are present in the winter and summer as well as in the annual mean for all the observational datasets. The biases in this case are expressed as percentage with respect to the values of observational estimates. In summer (Fig.4, right panels), a particularly pronounced negative bias, with values down to -100%, is visible over

25   arid regions and the monsoon area. Over the Tibetan Plateau the bias in this case is positive, with values larger than 100%. In winter (Fig.4, central panels), this positive bias becomes even larger, and extends further over adjacent regions. Over the central part of the domain, a different behaviour is evident between winter and summer. While in winter the model simulates wetter conditions (+20% to +100%), summers are drier ($\sim$ -50 %) than in observations. In the annual mean (Fig.4, left panels), the simulated climate is wetter over a large part of the eastern domain (with values exceeding +100%) and drier over desert

30   zones (with rare values smaller than -80%). Over the central part of the domain, winter and summer biases compensate each other.

Interestingly, in all the cases, the simulated DTRs are smaller than the observed ones over almost the entire domain (Fig.5). A positive bias in DTR, rarely exceeding +5°C, is evident only over isolated parts of the southern domain, in particular over the southern borders of the Tibetan Plateau. The differences compared with CRU observarions are more pronounced than with reanalysis data, with biases lower than -10°C in some cases. The pattern of the bias is similar for all the three considered

5   datasets, with some differences over southern regions in summer. Over the northernmost part of the domain, characterized by particularly cold conditions (minimum temperature under -30°C in winter, see Tab 4), a strong bias is evident only with respect to the CRU in all seasons. The smaller bias over this area arising from the comparison against reanalysis data is most likely due to the nature of these datasets, which combine model predictions and observations.

3.1.1   SS - Seasonal Cycle

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

Fig.6 (upper row) shows that for temperature, among the experiments for which single changes to the reference model configuration are applied (left side of the dotted vertical line), the ones with changes in the albedo treatment (**c+d**) lead to a noticeable improvement of the results (ranging between +4.5% and +7%). Nevertheless, in this case, the largest improvements

15   (greater than 5% for all the observational datasets) are obtained for experiment **j**, in which the type of the hydraulic lower boundary accounts for ground water with drainage and diffusion. In the combined configurations of different experiments (right side of Fig. 6) the results for temperature are considerably improved, whenever either one of the configuration changes of experiments **d** or **j** are used, with values of SS larger than 4% in almost all the cases. Other "combined" experiments do not have an important effect on the results.

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

As for precipitation, also for diurnal temperature range (Fig.6, bottom row) only one experiment seems to sensibly improve

25   over the results of the reference simulation: experiment **i** (SS ranging between +4% and +5%). In this experiment, the soil heat conductivity takes into account the ratio of soil moisture to soil ice. For DTR, two experiments, **d** and **j** lead to particularly negative skills (SS between -4% and -5%), which also affect the combined experiments including these configurations. The unique exception is the combined experiment **q**: in this case, the negative effects on the simulation of DTR of the parameterizations employed in experiment **d**, seem to be compensated by the positive ones of experiment **i**, resulting in positive values of

30   SS, varying between +1% and +2%.

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

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

3.1.2 SS - Single Seasons

5 In order to better understand the reasons for the variations in model performances due to specific changes to the model config-uration and to give more weight to the results of Sec. 3.1.1, the same SS analysis is conducted for individual seasons. In this case, the MAE is computed for the monthly climatologies of each season rather than of the entire year. Results of SS for the entire seasonal cycle might be biased by extremely high/low values over single seasons. Therefore, seasonal analyses could help in discriminating simulations presenting good and coherent performances over more periods of the year. In this case,

10 analyses focus on a single observational dataset for each variable: CRU for T2M and DTR, and GPCC for PRE.

The results of Fig. 7 show that, for all variables, in winter the changes in model performances among the different experi-ments are substantially smaller than in the other seasons. This indicates that none of the investigated parameters is particularly important for the model performances in winter over the region.

Fig. 7 (upper row) shows that for temperature, considering only the experiments in which single changes in the model

15 configuration are tested, the largest variations in the calculated SS are evident in summer and fall for simulation **j**. In this case, values of SS reach +20%. This suggests the importance of processes related to soil-atmosphere interaction for the simulation of summer temperatures over the region. Effects of changes in the treatment of albedo of experiment **c** and **d** also seem to be particularly important during the same seasons, with SS values of up to +8% and +13%, respectively in summer and in fall. Interestingly, experiment **d**, in which the vegetation albedo is modified according to forest fraction, has particularly strong

20 positive effects on temperature for the entire growing season, including spring, leading to an improvement of almost +5% in this case. Winter values are slightly negative in almost all the cases. Considering the combined simulations (right part of the upper panel of Fig.7), even higher positive values are obtained for those simulations using the setup of experiments **d** and **j**. An improvement of almost +30%, with respect to the reference simulation, is obtained for temperatures in summer for experiments **n** and **p**. These represent the highest values of SS obtained from all simulations, variables and seasons. Nonetheless, SS values

25 from these experiments are positive only in two seasons and sensibly negative, down to -10%, in the others. Other experiments, such as **m** and **q**, although not resulting in similarly high SS high values in summer, have more similar positive values also in spring.

For precipitation, relevant changes are evident for the single experiments **d** and **j** and for a series of combined configurations including the same changes (Fig. 7, middle row). In most of the cases, remarkable improvements are obtained only in summer.

30 The highest absolute values of SS for precipitation, with respect to the reference simulation, are obtained for experiments **n**, **o** and **p** in summer, exceeding +10%. Experiment **d** is the only one yielding distinctly positive values for all seasons. This is reflected in the simulation **q**, showing a similar behaviour. Indeed, an improvement in the representation of the albedo, with a better repartition of surface incoming radiation, allows to better simulate not only near surface temperatures but also precipitation. Conversely, despite their high SS values in summer, the combined configurations **n**, **o**, **p**, sharing the use of the

setup of **j**, produce negative skills in the other seasons. The range of changes in SS among all the different experiments for precipitation is smaller than for temperature, varying in between -10% and +15%.

For DTR, no remarkable improvements are evident for any of the experiments testing the use of single changes throughout all the seasons, except for experiment **i** (Fig. 7 lower row). In this case, values of SS vary in between +2% and +7%. Accordingly,

5 the treatment of soil heat conductivity taking into account soil moisture and soil ice separately, seems to be the only relevant factor, among the ones considered, leading to an improvement of the simulation of seasonal values of daily temperature ranges. Due to the fact that large parts of the domain are dry or semi-arid, a better consideration of soil moisture could improve the simulation of surface fluxes, positively affecting the daily temperature range. For DTR the range of changes in SS is smaller than for the other two variables, with values varying by less than +/-10%. The combined use of the configuration of experiments

10 **d** and **i** (experiment **q**), leads to an improvement of the results in all seasons (except winter), in a range of +1% and +4%.

An important consideration that can be drawn from the presented analysis is that the largest changes in the seasonal values of SS are obtained for summer, for variables and processes related to the representation of surface and soil properties.

Considering the results of the analysis proposed in this section and in Sec. 3.1.1, we conclude that the most important and consistent improvements in the simulated climatological mean of the considered variables with respect to the reference

15 simulation are obtained for experiment **q**.

3.1.3   SS - NCEP Vs ERAInterim

For the experiments performed using NCEP2 reanalysis data as boundary conditions, the best results for the simulation of the mean climatological values of the considered variables are obtained for experiment **q**. To test if these results also hold with different initial and boundary conditions, the SS of the seasonal cycle is calculated for two additional simulations conducted

20 using ERA-Interim reanalysis data to drive the model. The simulations are performed using the same configuration of the reference experiment **a** and the one of experiment **q**. The results of the two new simulations are presented in Tab. 4. The ranges of improvement obtained with configuration **q** with respect to **a** when using ERA-Interim as boundary conditions, are similar as when using NCEP2 as drivers. The changes in SS reach $\sim +10\%$ for temperatures, $\sim +4\%$ for precipitation, and vary between $+1.5$ and $+2\%$ in the case of DTR. This indicates that the obtained improvements in model performances for the

25 climatological seasonal mean do not depend on the driving dataset.

3.2   Variability

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

30 Fig. 8 shows the ratio of variance of the different CCLM experiments with respect to the one of the observations. The variances are calculated from monthly anomalies values of the three considered variables averaged over the subregions shown in Fig. 2 (see also Sec. 2.3). Values closer to 1 indicate a better agreement with the observations. Values between 0 and 1 show that the model variability is smaller than the one of observations. Finally, values greater than 1 indicate that the model results

have a greater variance than observations. A single observational dataset is used for each of the considered variables in this case: CRU for T2M and DTR, and GPCC for PRE.

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

For T2M, the best results in terms of simulated variance are obtained: the model is able to reproduce the interannual variability of the observations particularly well. In particular, a good agreement between simulated data and observations is evident for subregions **WSH**, **IMO** and **ARC**. The largest underestimation of the observed variance of temperature is obtained for cluster **CSA**. Therefore, the model is not only unable to simulate the mean temperatures for areas characterized by particularly low
10 climatological values, as demonstrated in Sec. 3.1, but it also shows a very low variability for the same regions when compared to observations. A negative value of the variance ratio is also evident for the sub-domains **DSS**, **SAR** and **STE** throughout almost all the experiments. These regions are all charachterized by a large range between minimum and maximum monthly temperatures (see Tab. 4).

For precipitation, in general, the values of the ratio of simulated-to-observed variance are considerably larger than 1 for
15 almost all the experiments and subdomains. Values are closer to 1 only for the domains **WSC** and **DHS** throughout all the experiments. In the domains **CSA**, **DSS** and **TIB** variance ratios are particularly large, reaching +3 in some cases. Over these domains, characterized by high topography, results from Sec. 3.1 have shown that the model simulates significantly wetter conditions. Hence, for montainous areas of the domain the model overestimates both mean values and variability of precipitation.
20 Values of variance ratios for DTR are smaller than 1 over almost all the subdomains and simulations. This indicates that the model, beside underestimating climatological values of the observed temperature diurnal cycle over the entire Central Asia domain as demonstrated in Sec. 3.1, it also undererestimates the amplitude of variations in the monthly means.

3.2.1 Uncertainties in the Investigation of Simulated Variability

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

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

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

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

4 Conclusions

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

Concerning the simulation of the mean climate, the model shows remarkable deficiencies in simulating the three considered variables (2-meter temperature, precipitation and diurnal temperature range) over different areas of Central Asia and different
15 seasons.

For temperature, the largest model biases are present in winter over Siberia, with positive biases exceeding $+15^{o}C$ in some cases. There are two likely reasons for these biases: an unsatisfactory representation of snow cover and soil permafrost. In fact, both these factors have a significant impact on heat transport within the soil and heat flux between soil and atmosphere, with important effects on near surface temperatures (Frauenfeld et al., 2004; Lachenbruch and Marshall, 1986; Saito et al.,
20 2007; Klehmet, 2014). Siberian permafrost often exceeds a depth of 100 meters, reaching values of up to 1km (Yershov, 2004). Therefore, many studies (Alexeev et al., 2007; Dankers et al., 2011; Nicolsky et al., 2007; Lawrence et al., 2008; Saito et al., 2007; Klehmet, 2014) highlight the importance of an adequate depth of model soil layers for the proper representation of processes related to permafrost. At the same time, other studies (Saito et al., 2007; Waliser et al., 2011; Klehmet, 2014) suggest that a better representation of the vertical stratification of the snow pack could have a significant effect on the simulated energy
25 budget and, consequently, on near surface temperatures over the area. Following these hypotheses, two 25-year long additional simulations have been conducted during this study, with an increase of the total model soil depth and with the use of a multi-layer snow model. Results indicate that, for the part of Siberia included in the domain of study, no significant changes are evident in the two cases and further tests are indeed necessary. Importantly, an additional cold bias, in some cases lower than -10°C, is present for every season over the Tibetan Plateau. Other regional climate models suffer from a similar bias (GUO et al.,
30 2018; Meng et al., 2018). This could likely be related to a bad representation of the albedo for highly complex topographies. In fact, a study by Meng et al. (2018) showed that changes in the albedo over the region have led to an important improvement of the results of an RCM. Another possible explanation for this cold bias might be the parametrization of surface fluxes (Zhuo et al., 2016). Consequently, further analyses should focus on improving the mode representation of these processes.

For precipitation, particularly wet conditions are simulated by the COSMO-CLM over the Tibetan Plateau. Again, this bias seems to be common to several RCMs for areas characterized by complex topography (GUO et al., 2018; Gao et al., 2015; Feng and Fu, 2006) and is likely related to an overestimation of orographic precipitation enhancement in the models (Gerber et al., 2018) and/or to an incorrect simulation of the planetary boundary layer (Xu et al., 2016). Additionally, in the COSMO-CLM

5   simulations a significant dry bias occurs over arid and desertic regions, especially in summer. A similar COSMO-CLM bias has already been seen for other semi-arid and dry regions of the world, such as the Mediterranean region. In this case, it was connected with an incorrect simulation of soil-atmosphere interactions by the model (Fischer et al., 2007; Seneviratne et al., 2010; Russo and Cubasch, 2016), which is likely the case also for Central Asia.

For DTR, the model underestimates the climatological mean of the diurnal cycle of temperatures, for all seasons and sub-

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

In order to reduce these biases, sensitivity experiments have been performed to study the effect of different physical param-

15   eterizations of COSMO-CLM and their mutual interaction. In this way, an optimal model setup with respect to the simulated mean climatologies has been determined. The combined change of the albedo taking into consideration the ratio of forest fractions and the soil conductivity taking into account the ratio of liquid water and ice in the soil, leads to the best results in simulated climatological means of the three considered variables. Importantly, the model seems to be particularly sensitive to those parameterizations that deal with soil and surface features, and that could positively affect the repartition of incoming

20   radiation.

An analysis of the model performance for seasonal climatologies confirms these results. Interestingly, for all the analyzed variables, winter is the season for which no substantial improvements in model results could be achieved with the set of investigated configurations. This points to an important role of other factors for the winter climatology, such as the simulation of snow cover, that are not affected by the investigated parameters. The results for the mean climatologies appear to be independent

25   of the observational dataset used for evaluation and of the boundary data employed to force the simulations.

Finally, the observed variability of temperature is relatively well represented in the model simulations for different sub-regions of the domain. For precipitation, the model overestimates the variability of observations. On the contrary, the model underestimates the variability in the diurnal cycle of temperatures over the entire region. Among the three investigated variables, only for precipitation there are significant changes in the simulated variance throughout all conducted experiments. However,

30   due to the large uncertainties in the variability estimates from observations, the use of different boundaries and the model internal variability, it has not been possible to rank the different simulations according to their representation of the monthly variability.

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

References

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

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

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

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

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

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

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

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

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

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

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

Davies, H.: A laterul boundary formulation for multi-level prediction models, Quarterly Journal of the Royal Meteorological Society, 102,
25   405–418, 1976.

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

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

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

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

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

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

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

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

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

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

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

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

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

Gerber, F., Besic, N., Sharma, V., Mott, R., Daniels, M., Gabella, M., Berne, A., Germann, U., and Lehning, M.: Spatial variability in snow

20 precipitation and accumulation in COSMO–WRF simulations and radar estimations over complex terrain, The Cryosphere, 12, 3137–3160, 2018.

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

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

25 113, 2008.

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

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

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

Harris, I. and Jones, P.: CRU TS4.01: Climatic Research Unit (CRU) Time-Series (TS) version 4.01 of high-resolution gridded data of month-by-month variation in climate (Jan. 1901- Dec. 2016)., available from

35 http://catalogue.ceda.ac.uk/uuid/58a8802721c94c66ae45c3baa4d814d0, 2017.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Geoscientific
Model Development
Discussions

Open Access

EGU

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

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

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

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

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

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

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

Mearns, L., Giorgi, F., McDaniel, L., and Shields, C.: Analysis of variability and diurnal range of daily temperature in a nested regional climate model: Comparison with observations and doubled CO $_2$ results, Climate Dynamics, 11, 193–209, 1995.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Zhuo, H., Liu, Y., and Jin, J.: Improvement of land surface temperature simulation over the Tibetan Plateau and the associated impact on circulation in East Asia, Atmospheric Science Letters, 17, 162–168, 2016.
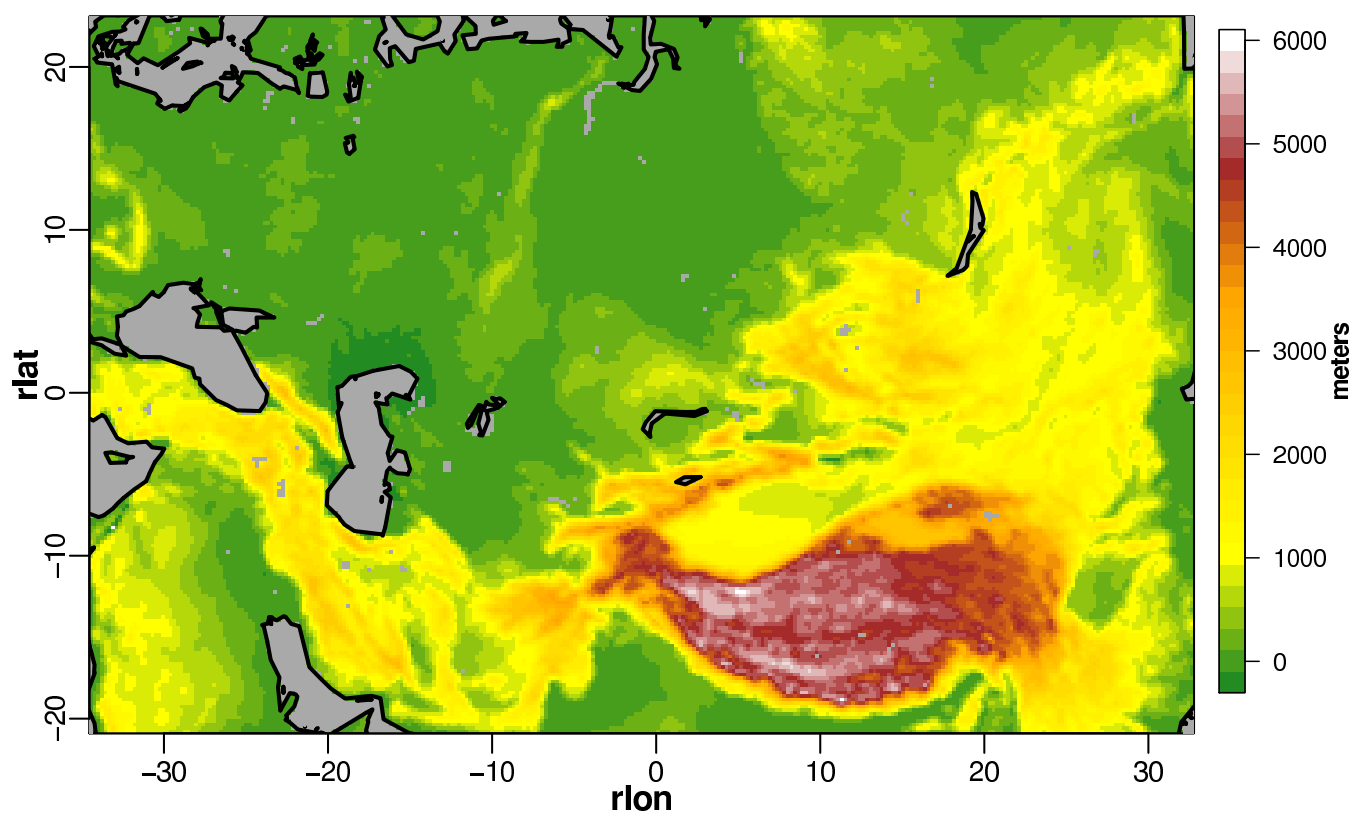
Figure 1. Orography map of the Central Asia simulation domain in rotated coordinates, at a spatial resolution of 0.22^o lon.
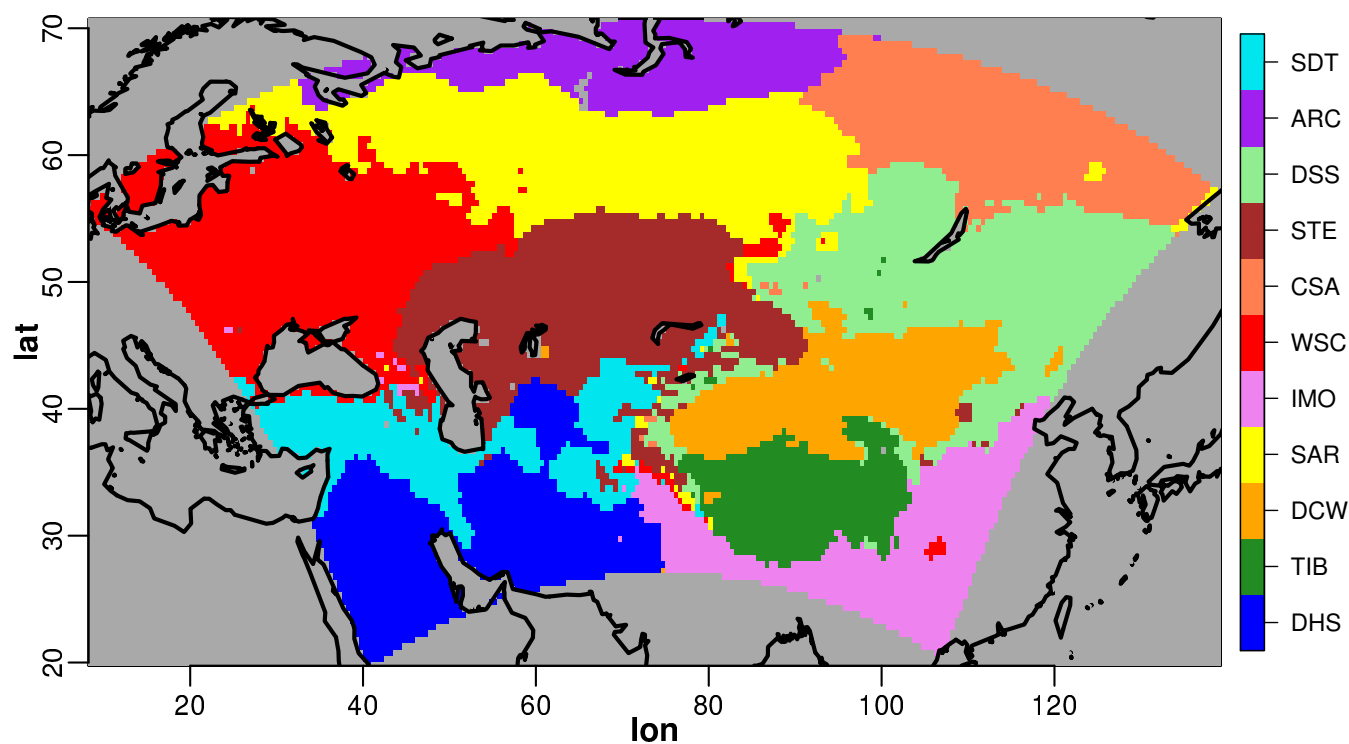
Figure 2. Map of the 11 subdomains obtained through k-means clustering of the q-normalized monthly climatologies of the three considered variables over the period 1996-2005.
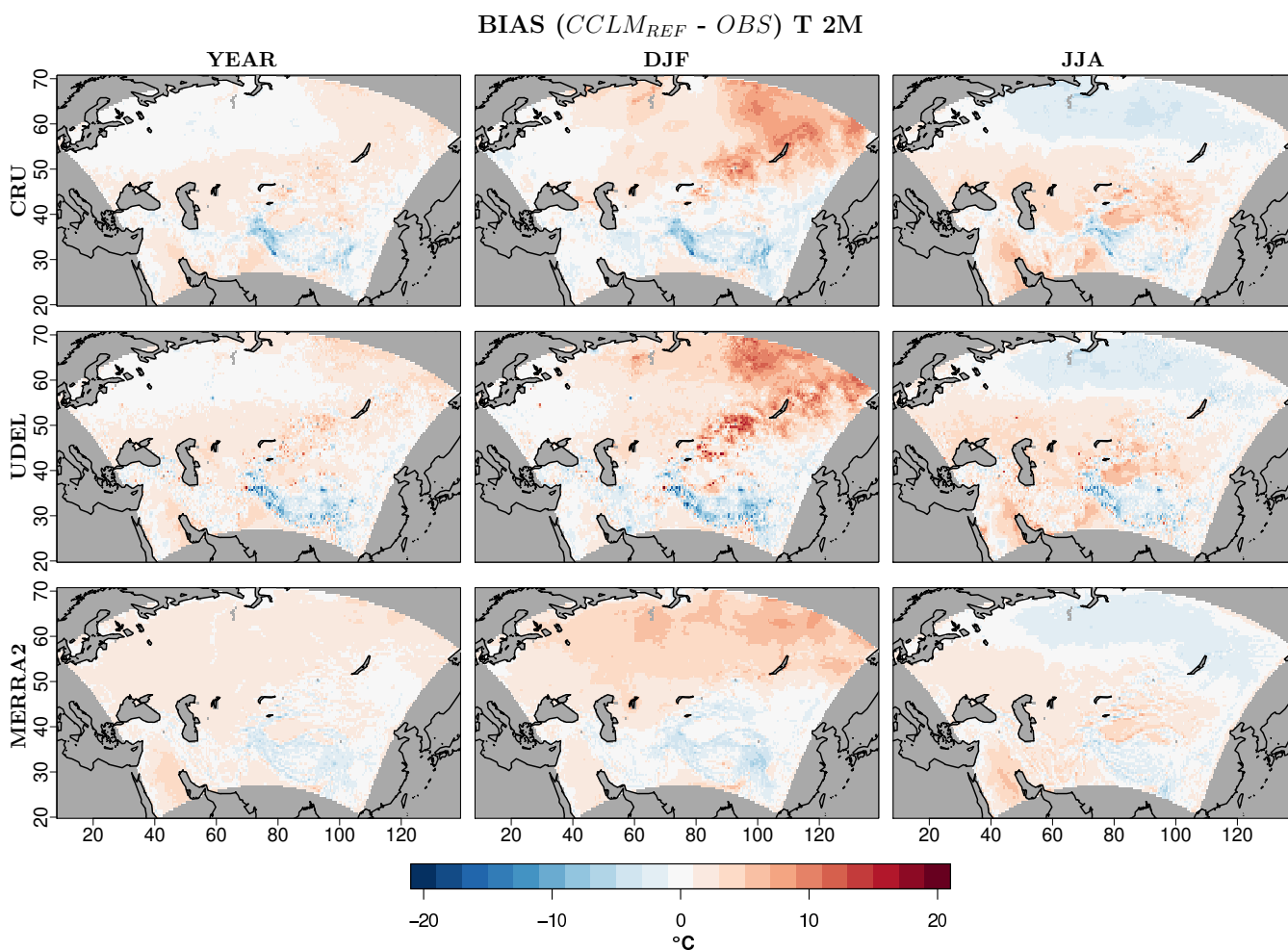
Figure 3. Mean bias of annual mean (*left*), winter mean (*middle*) and summer mean (*right*) near surface temperature (T2M,^{o}C), of the reference COSMO-CLM simulation (**a**) with respect to three observational datasets (from top to bottom: CRU, UDEL and MERRA2), for the period 1995-2005.
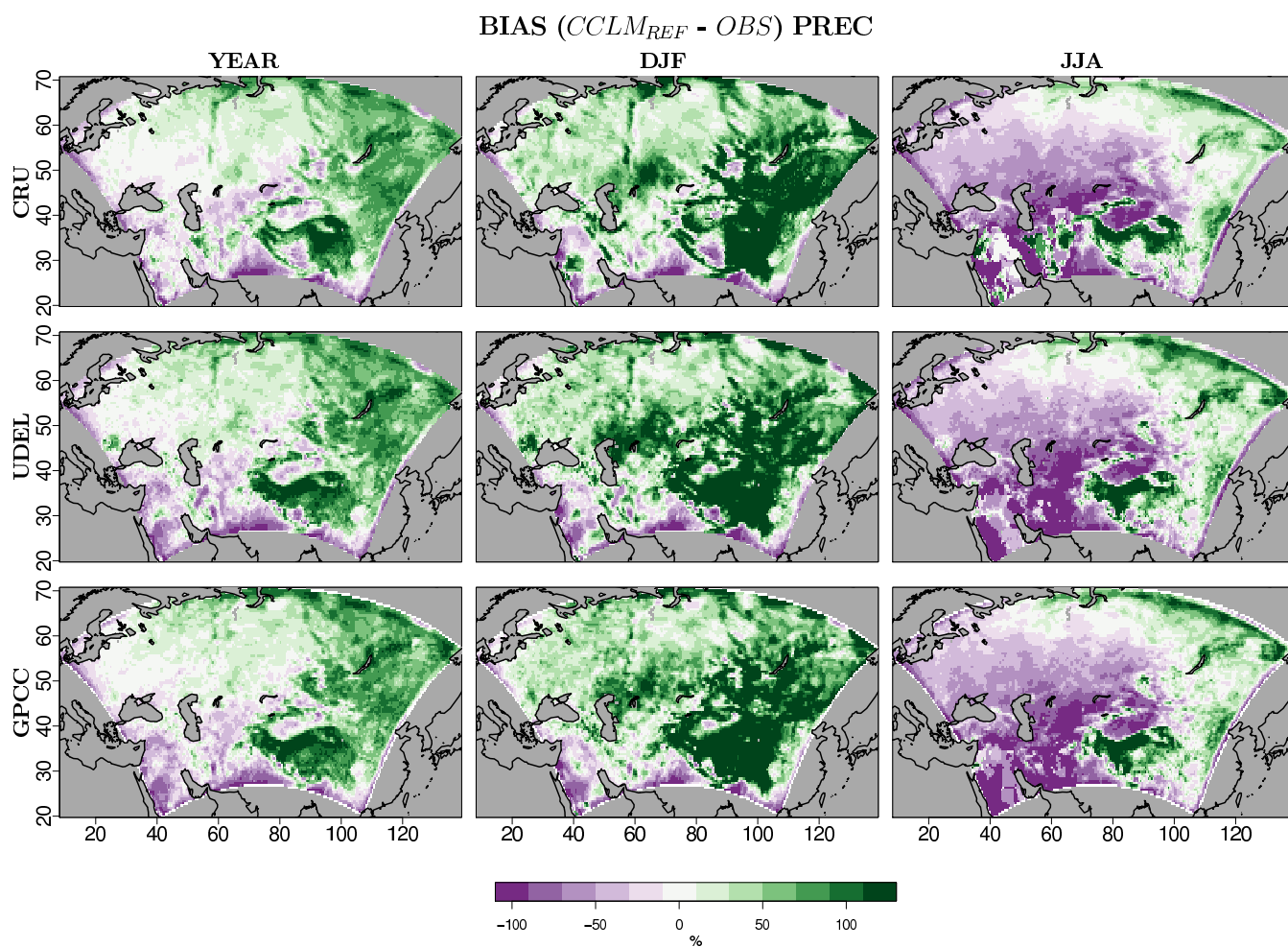
Geoscientific
Model Development
Discussions

Figure 4. Mean bias of annual (*left*), winter (*middle*) and summer mean (*right*) relative precipitation (PRE, %), of the reference COSMO-CLM simulation (**a**) with respect to three observational datasets (from top to bottom: CRU, UDEL and GPCC), for the period 1995-2005.
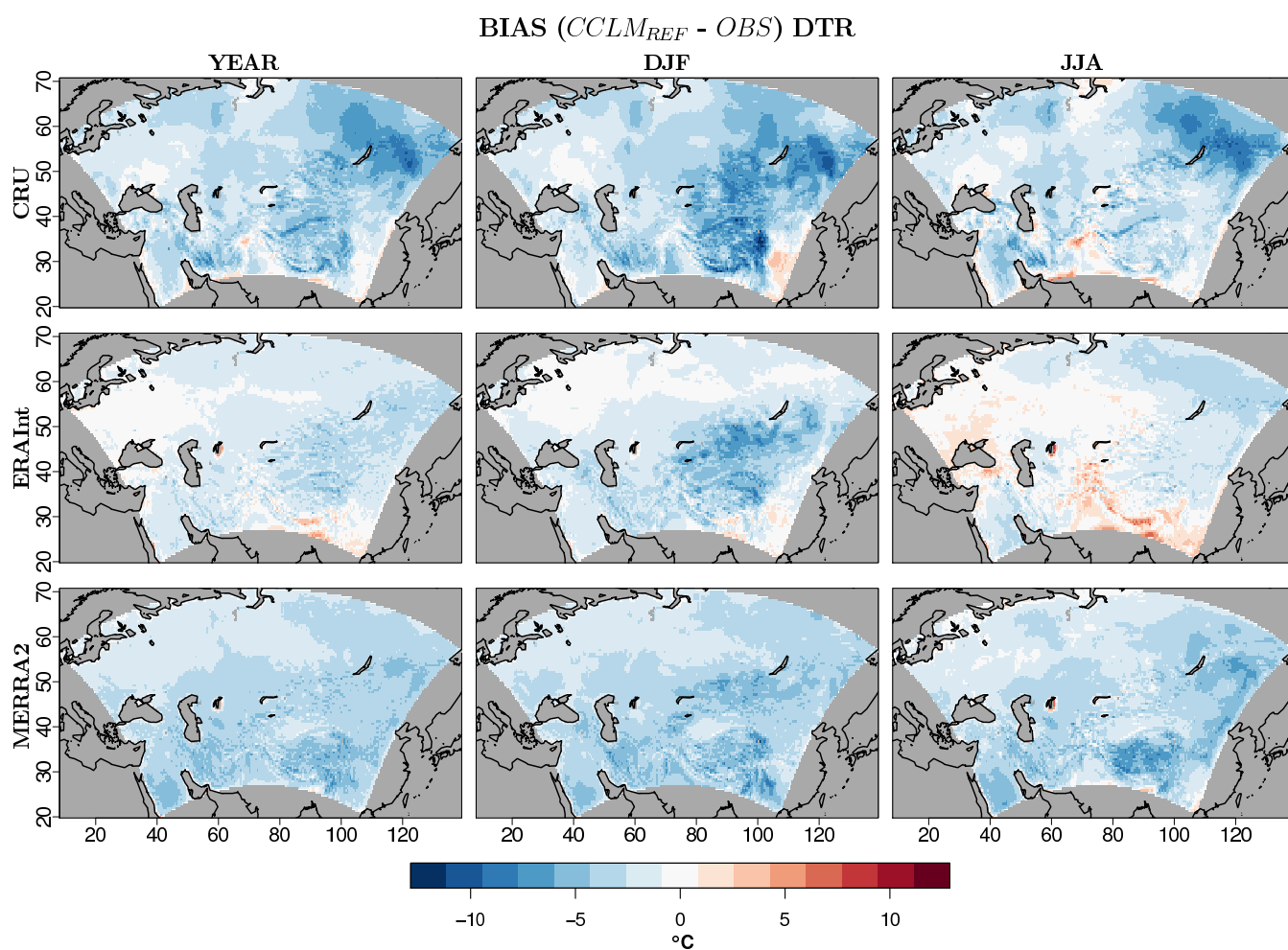
Figure 5. Mean bias of annual (*left*), winter (*middle*) and summer mean (*right*) diurnal temperature range (DTR,^{o}C), of the reference COSMO-CLM simulation (**a**) with respect to three observational datasets (from top to bottom: CRU, MERRA2 and ERAInterim), for the period 1995-2005.
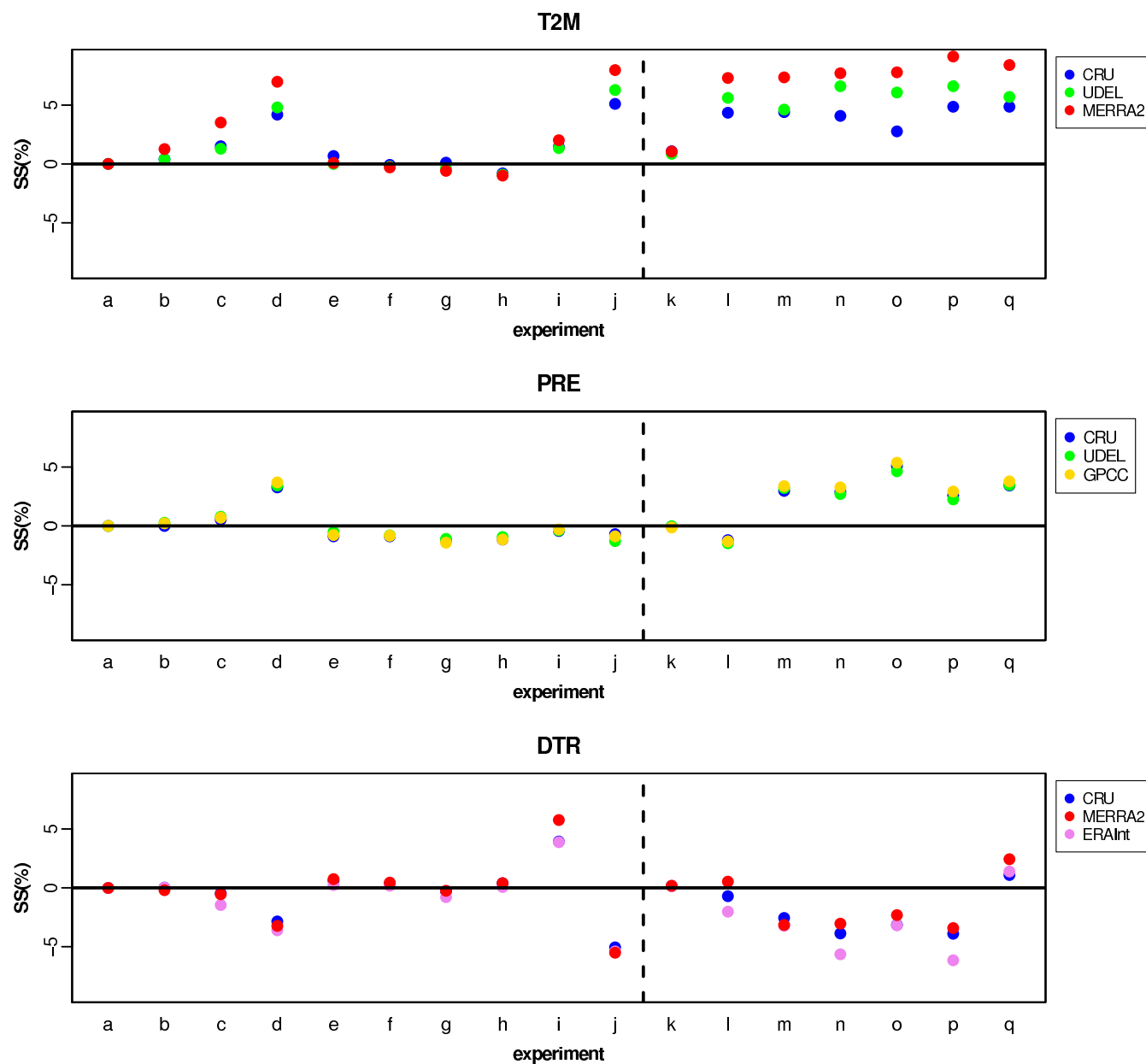
Figure 6. Skill Score (SS) derived from the MAE calculated over the monthly climatological values of the seasonal cycle of different COSMO-CLM simulations and observational datasets. From top to bottom, the SS for each variable is displayed. The dotted vertical black line divides the simulations with the same configuration of the reference simulation plus a single change in the model setup (on the left) and the ones obtained through the combinations of the previous ones (on the right). Positive (negative) values indicate better (worse) performance of the considered simulations compared to the reference one.
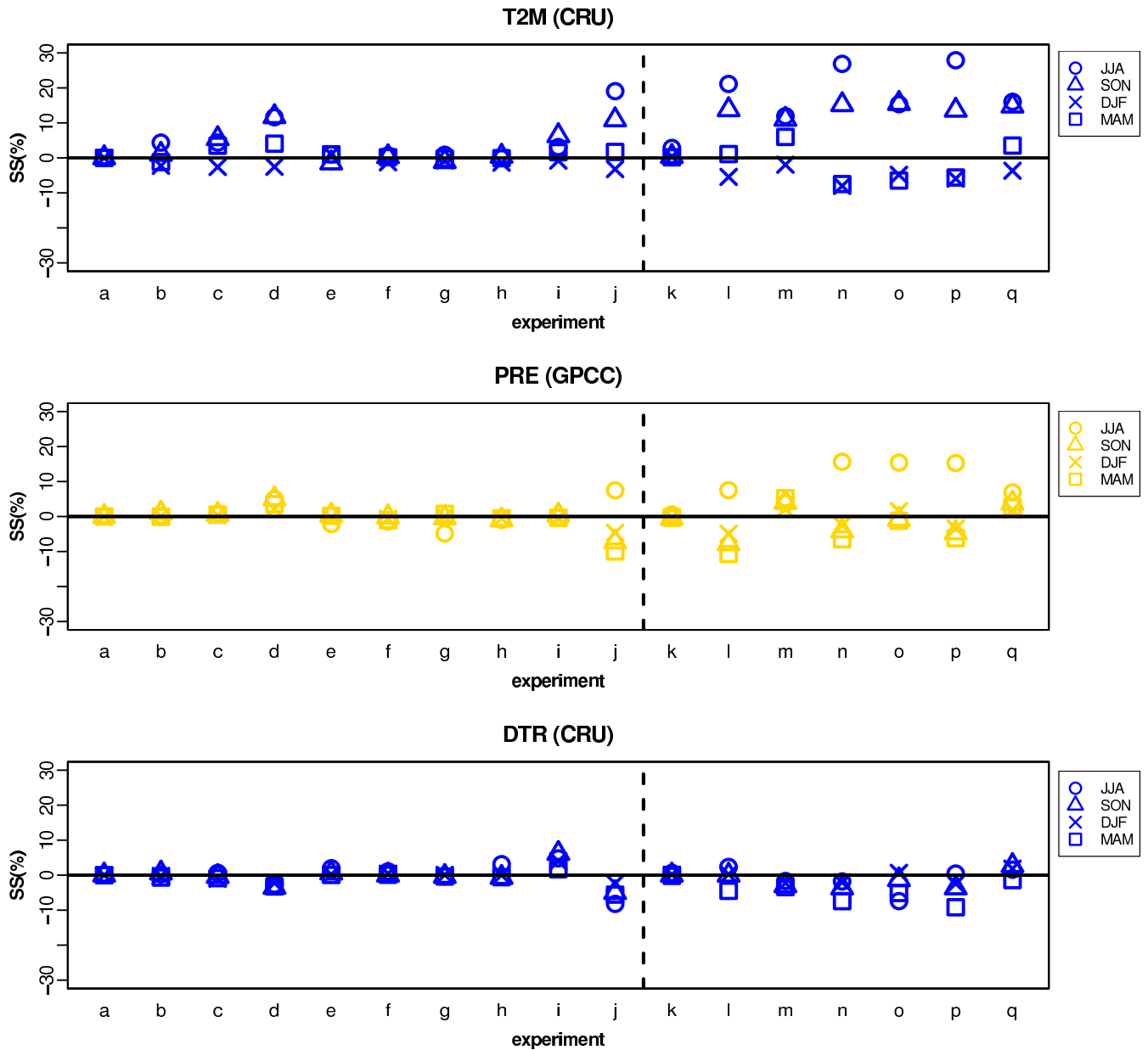
Figure 7. Skill Score (SS) derived from the MAE calculated over the monthly climatological values of each season of the different COSMO-CLM simulations and the observational datasets. A single observational data-set is considered for each variable in this case: CRU for T2M and DTR, and GPCC for PRE. From top to bottom, the SS for each variable is displayed. The dotted vertical black line divides the simulations with the same configuration of the reference simulation plus a single change in the model setup (on the left) and the ones obtained through the combinations of the previous ones (on the right). Positive (negative) values indicate better (worse) performances of the considered simulations compared to the reference one.
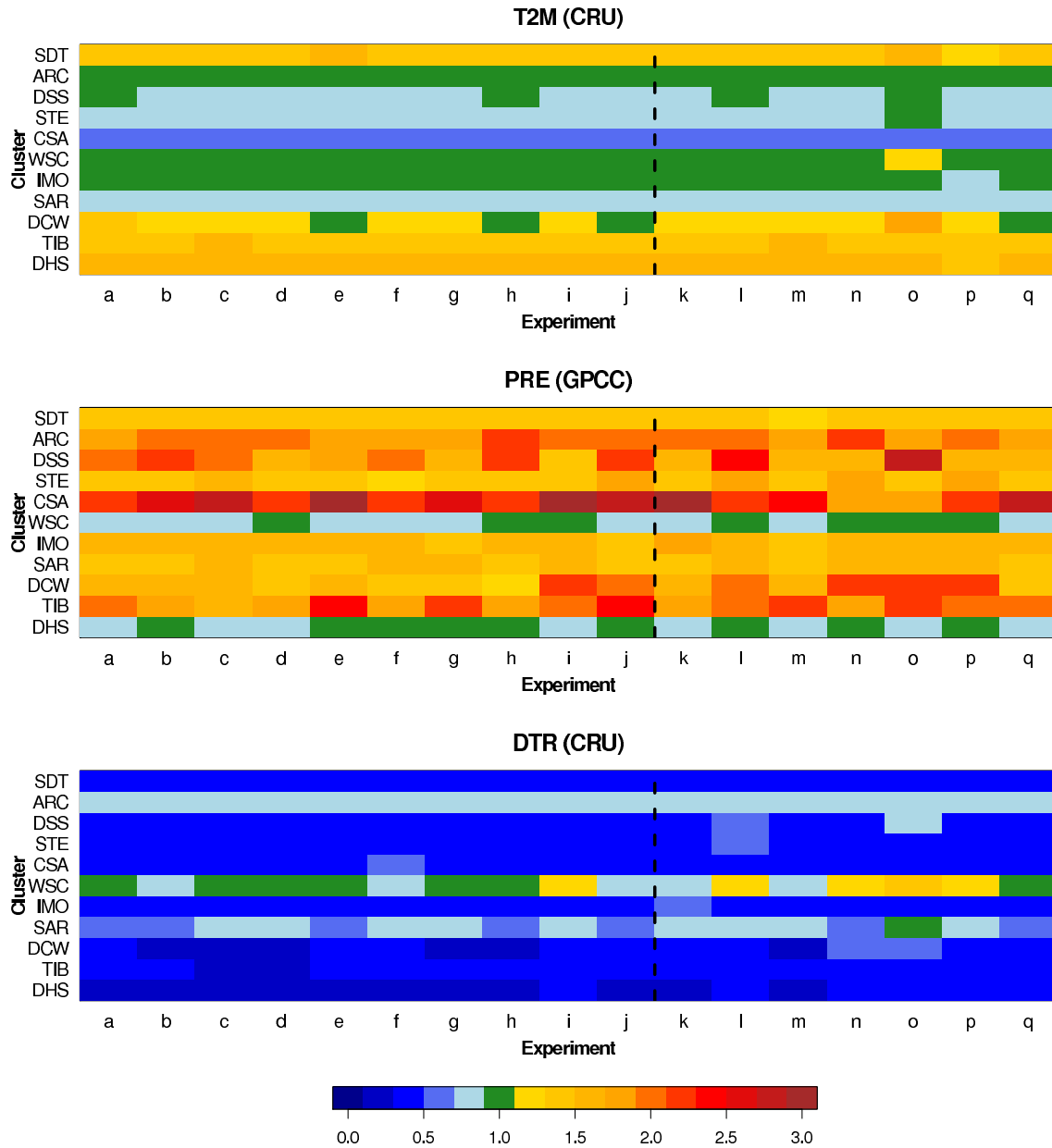
Figure 8. Fraction of Variance calculated between the monthly anomalies over the period 1996-2005 derived from the different COSMO-CLM simulations and a single observational dataset, for (top to bottom) 2-meter temperature, precipitation and diurnal temperature range, for each of the 11 sub-regions obtained by k-means clustering. The dotted vertical black line divides the simulations with the same configuration of the reference simulation plus a single change in the model setup (on the left) and the ones obtained through the combinations of the previous ones (on the right). Values larger (smaller) than one indicate better (worse) performances of the considered simulations with respect to the reference one, in the representation of observed variability.

Figure 9. Absolute differences in variance ratio. The left column of each panel shows the absolute differences in the variance ratio of the same experiment **a** when considering different observational datasets. The right column of each panel shows the range of absolute differences betweeen the variance ratio of experiment **a** and the variance ratios of the other experiments, from **b** to **q**, when using a single observational dataset for each variable. Each of the three panels, from left to right, presents, respectively, the results of the comparison for T2M, PRE and DTR. The different rows show the results for each of the clusters.

Table 1. List of experiments and main selected parameters.

Experiment	Changes in Model Configuration
a	Reference Simulation - CORDEX South East Asia setup
b	**a**+Tegen Aerosol Dataset
c	**a**+Surface albedo determined as a weighted mean of two external fields for dry and saturated soil
d	**a**+Vegetation albedo modified considering forest fraction.
e	**a**+Type surface-atmosphere transfer based on prognostic TKE in the surface layer
f	**a**+Cloud representation taking into account subgrid-scale condensation; cloud cover and water content calculated according to a statistical closure.
g	**a**+Equal to **f** but cloud cover and water content calculated according to a relative-humidity criterion
h	**a**+Exponential root distribution
i	**a**+Soil heat conductivity taking into account soil moisture/soil ice
j	**a**+Hydraulic lower boundary considering ground water with drainage and diffusion
k	**a+e+f+g**
l	**a+h+i+j**
m	**a+b+d**
n	**a+d+h+i+j**
o	**a+b+d+e+f+g+i**
p	**a+b+d+e+f+g+h+i+j**
q	**a+d+i**
a_ERAInterim	**a** driven by ERAInterim
q_ERAInterim	**q** driven by ERAInterim
a2	**a** with initial date shifted by +1 month
a3	**a** with initial date shifted by +3 month
a4	**a** with initial date shifted by -1 month
a5	**a** with initial date shifted by -3 month
SOIL	increased soil layers number and depth (**25-year long**)
SNOW	increased soil layers number and depth + use of multi-layer snow model (**25-year long**)

Geoscientific
Model Development
Discussions

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

Spatial Resolution	$\sim 0.22^{o}$ lon
Domain Extent	342×220 points
Convection	Tiedke
Time Integration	Runge-Kutta,
Lateral Relaxation Layer	250km
Soil Model	TERRA-ML SVAT
Rayleigh Damping Layer (rdheight)	≥ 18km
Active Soil Layers	9
Active Soil Depth	5.74m
Atmospheric Vertical Layers	45

Geoscientific
Model Development
Discussions

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

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

Table 4. Changes of the SS calculated for the experiment **q** with respect to reference simulation **a**, when both simulations are driven by ERA-Interim reanalysis data.

Dataset	T2M	PRE	DTR
CRU	+7.5	+3.5	+1.8
UDEL	+7.3	+3.8	x
GPCC	x	+4.1	x
MERRA2	+11.1	x	+3.1
ERAINT	x	x	+1.5