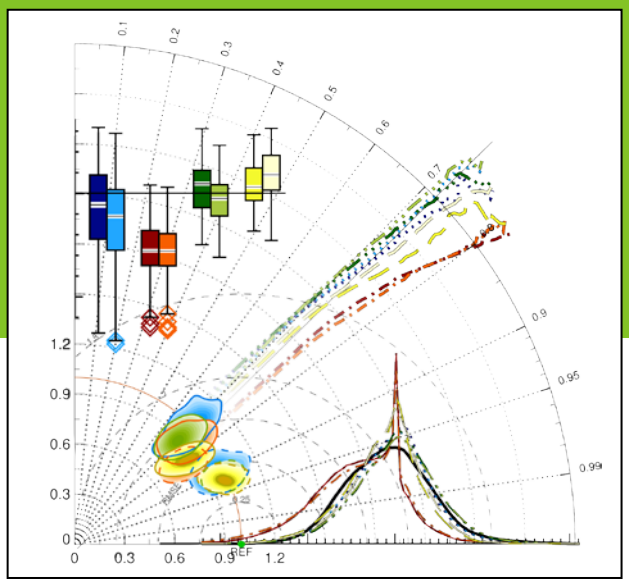
Wegener Center for Climate and Global Change
University of Graz



Scientific Report No. 41-2011

NHCM-1: Non-hydrostatic climate modelling

Part III

Evaluation of the Local Climate Model Intercomparison Project (LocMIP) simulations

Andreas F. Prein
Heimo Truhetz
Martin Suklitsch
Andreas Gobiet

February 2011

Wegener Center
www.wegcenter.at

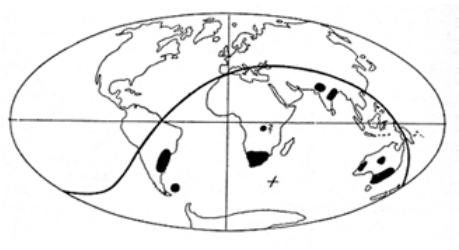
UNI GRAZ

kindly supported by:

FWF
Der Wissenschaftsfonds.

The **Wegener Center for Climate and Global Change** combines as an interdisciplinary, internationally oriented research center the competences of the University of Graz in the research area „Climate, Environmental and Global Change“. It brings together, in a dedicated building close to the University central campus, research teams and scientists from fields such as geo- and climate physics, meteorology, economics, geography, and regional sciences. At the same time close links exist and are further developed with many cooperation partners, both nationally and internationally. The research interests extend from monitoring, analysis, modeling and prediction of climate and environmental change via climate impact research to the analysis of the human dimensions of these changes, i.e., the role of humans in causing and being effected by climate and environmental change as well as in adaptation and mitigation. (more information at www.wegcenter.at)

The research documented in this report was supported by the Austrian Science Fund.



Alfred Wegener (1880-1930), after whom the Wegener Center is named, was founding holder of the University of Graz Geophysics Chair (1924-1930) and was in his work in the fields of geophysics, meteorology, and climatology a brilliant, interdisciplinary thinking and acting scientist and scholar, far ahead of his time with this style. The way of his ground-breaking research on continental drift is a shining role model — his sketch on the relationship of the continents based on traces of an ice age about 300 million years ago (left) as basis for the Wegener Center Logo is thus a continuous encouragement to explore equally innovative scientific ways: *paths emerge in that we walk them* (Motto of the Wegener Center).

Wegener Center Verlag • Graz, Austria

ISBN 978-3-9502940-8-8

February 2011

Contact: Andreas F. Prein.
andreas.prein@uni-graz.at

Wegener Center for Climate and Global Change
University of Graz
Leechgasse 25
A-8010 Graz, Austria
www.wegcenter.at

NHCM-1:
Non-Hydrostatic Climate Modelling

Part III

Evaluation of the Local Climate Model Intercomparison Project (LocMIP) simulations

Andreas Prein[1], Heimo Truhetz[1], Martin Suklitsch[1], Andreas Gobiet[1]

[1] Wegener Center for Climate and Global Change and Insitute for Geophyics, Astrophysics and Meteorology, Institute of Physics, University of Graz, Graz, Austria

February 2011

Contents

Preface

This document describes progress and results of the FWF funded project NHCM-1 (project ID P19619). It is part three of a three document collection. The other parts are:

- A. Prein and A. Gobiet (2011): *Defining and Detecting Added Value in Cloud Resolving Climate Simulations*. WegCenter Report, **39**, WegCenter Verlag, Graz, Austria, ISBN 978-3-9502940-6-4.
- M. Suklitsch, H. Truhetz, A. Prein, A. Gobiet (2011): *Current state of selected cloud resolving regional climate models and their error characteristics*, WegCenter Report, **40**, WegCenter Verlag, Graz, Austria, ISBN 978-3-9502940-7-1.

Further results have been or will be published in peer reviewed literature:

- Awan, N. K., A. Gobiet, M. Suklitsch, 2011: *Simulating climate extremes: Performance of high resolution regional climate models in European Alpine region*, J. Geophys. Res., submitted
- Awan, N. K., A. Gobiet, H. Truhetz, 2010: *Parameterization induced error-characteristics of MM5 and WRF operated in climate mode over the Alpine Region: An ensemble based analysis*, J. Climate, accepted
- Suklitsch, M., A. Gobiet, H. Truhetz, N. K. Awan, H. Göttel, D. Jacob, 2010: *Error Characteristics of High Resolution Regional Climate Models over the Alpine Area*, Clim. Dyn., doi: 10.1007/s00382-010-0848-5, published "online first" (http://www.springerlink.com/content/qj813x0255712826/)
- Suklitsch, M., A. Gobiet, A. Leuprecht, C. Frei, 2008: *High Resolution Sensitivity Studies with the Regional Climate Model CCLM in the Alpine Region*, Meteorol. Z., **17**, 4, 467 – 476

Additionally, three peer reviewed articles covering the topics in the three reports are currently prepared.

Abstract

Climate change is a global problem but the phenomena that affects the society or ecosystems have very often small scale features. Especially small scale extreme events and hydrological processes have large impacts and therefore the climate impact community has a high demand in reliable small scale climate projections. Dynamical downscaling with regional climate models (RCMs) has proved to add value especially on regional scale compared to General Circulation Models (GCMs). Currently, RCMs are usually operated with horizontal grid spacing of 10 km or larger, which is much too coarse to resolve local scale phenomena. However, simulating local scales (below 10 km) is far from being established in climate modelling.

Therefore, in the Local Climate Model Intercomparison Project (LocMIP) an ensemble of non hydrostatic RCM simulations with ∼10 km and ∼3 km horizontal grid spacing is compared to a highly resolved reference dataset within the time period June, July, and August (JJA) 2007 and December, January, and February (DJF) 2008. The simulations are evaluated over the entire Eastern Alps and, in more detail, in two smaller focus regions, where one lies in the south-eastern foothills of the Alps and one is centered on the highest peaks of Austria.

The major aim of LocMIP is to enable comparable local scale climate simulations (also called convection-resolving climate simulations (CRCSs)) performed with different models operated by different research groups. This initiative is a starting point to establish reliable and internationally accepted information on the error characteristics of specific convection resolving climate models and on the quality of CRCS in general. Within LocMIP, two groups operated three different models, with further groups expressing interest to contribute, but not within the time frame of the project Non-Hydrostatic Climate Modelling (NHCM-1). This is of course not representative for the international CRCS community, but LocMIP establishes a starting point, it provides first valuable results and – even more importantly – methods, datasets, and formats that can now be easily used and expanded for larger initiatives.

The results of LocMIP show that each RCM has its strengths and weaknesses and that no model clearly outperforms all other models. However, deficiencies in some models are very clearly identified. Very often the two model versions (4.0 and 4.8) of COSMO model in CLimate Mode (CCLM) show similar error characteristics. However, the newer version outperforms the older one on a wide range of statistics. Also the Weather Research and Forecasting Model (WRF) and Fifth-Generation Mesoscale Model (MM5) simulations show similar errors, but MM5 has problems especially in terms of temporal

correlation which do not occur in WRF.

Generally, the quality of the 3 km simulations is mostly dominated by the quality of their 10 km mother simulations. Added value can especially be found in the spatial root mean squared errors (RMSEs) and correlation coefficients of the 2 m temperature field, which is mainly caused by the better representation of orography in the highly resolved simulations. Additionally, the 3 km simulations are able to produce rainfall with higher intensities and considerably improve the skill of CCLM to generate heavy precipitation events.

Acknowledgement

The authors acknowledge the data providers at Central Institute for Meteorology and Geodynamics (ZAMG) for data of the Integrated Nowcasting through Comprehensive Analysis (INCA) dataset and the data providers at Wegener Center for Climate and Global Change (WegC) for data of the WegenerNet, which made this detailed evaluation possible in the first place.

The authors are also thankful to Nauman Khurshid Awan from the Wegener Center for Climate and Global Change (WegCenter) for contributing his WRF simulation and to Goran Georgievski and Klaus Keuler CCLM from the Brandenburg University of Technology Cottbus (BTU) for their contribution with CCLM.

This study heavily relied on computing resources. The authors are therefore grateful to the Jülich Supercomputing Centre (JSC) for providing these resources. The European Centre for Medium-Range Weather Forecasts (ECMWF) receive thanks for providing operational Integrated Forecast System (IFS) analysis and short term forecasts which served as lateral boundary conditionss (LBCs) for the simulations.

Last but not least many thanks go to the Austrian Science Fund (FWF) who funded this research under project ID P19619.

1 Introduction

During the last decade, regional climate models (RCMs) have become sophisticated tools for the simulation of present day and future climate with grid spacings between 50 km and 25 km (e.g., Christensen and Christensen (2007), Linden and Mitchel (2009)). At this resolution many important physical processes cannot be resolved and have to be parameterized. These parameterizations are only partly based on fundamental physical laws and carry therefore potential errors, particularly when applied outside the range they have been developed for. For example, there is large uncertainty in the summertime precipitation change over Europe, because it strongly depends on the model formulation (Déqué et al. 2007). The reason for this is the weak synoptic scale forcing, the prevalent moist convection, and the importance of surface atmosphere interactions (Seneviratne et al. 2006). Thereby, convection parameterizations which where mainly developed to mimic convection over the tropics at relatively coarse grid spacings of 50 km to 100 km might not be applicable over Europe, and hence they are a major source of uncertainty.

Convection-resolving climate simulations (CRCSs) with horizontal grid spacings of ∼3 km and below have a high potential to reduce those kinds of uncertainties by abandoning convection parameterizations. In general, CRCSs have two major advantages: First, a higher resolution enables a better representation of the orography, and hence surface fields are expected to be improved. Second, deep moist convection can be resolved explicitly. Some studies (e.g., Grell et al. (2000), Hohenegger et al. (2008)) have already shown the advantages of CRCSs but climate modeling at the scale below 10 km grid spacing is far from being established.

The Local Climate Model Intercomparison Project (LocMIP) addresses convection-resolving issues and evaluates the performance and the error ranges of non-hydrostatic RCM simulations on regional and local scale with respect to mean climate, temporal, and spatial characteristics. Four models (or versions of models) (COSMO model in CLimate Mode (CCLM) 4.0, Fifth-Generation Mesoscale Model (MM5), and Weather Research and Forecasting Model (WRF) at the Wegener Center for Climate and Global Change (WegCenter) and CCLM 4.8 at Brandenburg University of Technology Cottbus (BTU)) were used to make 4 simulations, each with 10 km and 3 km horizontal grid spacing. The major difference between the simulations with different resolutions is that the 10 km simulations use some kind of convection parameterization, while in the corresponding 3 km simulations deep convection is resolved explicitly (i.e., the parameterization of convection is turned off).

The simulations were performed with lateral boundary conditionss (LBCs) derived

from the Integrated Forecast System (IFS) of the European Centre for Medium-Range Weather Forecasts (ECMWF) (Bechtold et al. 2008) on 25 km horizontal grid spacing in two seasons (June, July, and August (JJA) 2007 and December, January, and February (DJF) 2007/08) on a central European (for the 10 km simulations) and an eastern Alpine domain (for the 3 km simulations). As reference the Integrated Nowcasting through Comprehensive Analysis (INCA) dataset of the Central Institute for Meteorology and Geodynamics (ZAMG) (Haiden et al. 2011) is used, which provides hourly fields of 2 m temperature, precipitation, relative humidity, and global radiation concerning the entire Austrian territory with 1 km grid spacing.

Three domains where considered for the evaluation: Domain eastern Alpine Region (D3) covers the whole eastern Alpine region from Innsbruck to the Austrian Hungarian boarder. The other two domains are focus regions within D3 and are centered on a hilly region in south east Styria (domain south-east Styria (D4a)) and on the Hohe Tauern National park, which covers the highest peaks in Austria (e.g., Großglockner, 3798 m altitude; Hohe Tauern National Park (D4b)).

The evaluation methods originate from the Non-Hydrostatic Climate Modelling (NHCM-1) project evaluation (Prein and Gobiet 2011) and can bee seen as best practice for analyzing the skill of very high resolution RCM simulations.

2 Materials and Methods

2.1 Study Regions

The considered domains cover the entire eastern Alpine region (see Figure 2.1) from Innsbruck in the west to the Austrian Hungarian boarder in the east. This region will be later shortly called domain D3. Additionally, a special focus lies on two sub regions of D3 where one is in the Alps (Hohe Tauern National Park (D4b)) and the second is in the Alpine foreland (south-east Styria (D4a)).

Domain D4a: D4a focuses on a hilly region in eastern Styria on the foothills of the Alps where the highest summit is roughly 1500 m and located in the south western corner of the domain. The climate of this region is characterized by the predominant influence of Mediterranean cyclones and turbulent processes especially in summer. From the north and the west it is shielded by the Alps. Characteristic for summertime are convective precipitation events and partly long-lasting droughts. In winter typically dry conditions are predominant.

Domain D4b: The domain D4b is centered on the highest peaks of the Austrian Alps which are in the Hohe Tauern National Park. The Großglockner, with an elevation of 3798 m, is the highest summit in this region, and the valleys are roughly on a height of 550 m. To get along with this high vertical expansion is a challenging task for regional climate models (RCMs). Overall, there are 73 glaciers in this region with a wide range of sizes.

2.2 Synoptic Overview

In this chapter the authors give an overview of how the two seasons that were simulated in the course of the experiment (June, July, and August (JJA) 2007 and December, January, and February (DJF) 2007/08) compare to the long-term climate from the European observation dataset (E-OBS) dataset (Haylock et al. 2008) (Subsection 2.2.1). This is followed by a more detailed description of outstanding weather events that occurred during the two periods in Subsection 2.2.2. The focus in this synoptic overview lies on the model domain which later on will be evaluated.

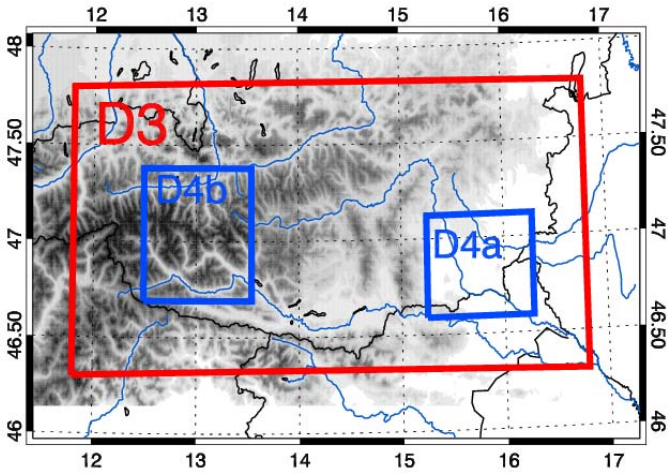


Figure 2.1: Investigated domains in the eastern Alpine region. The domain D3 is shown as red rectangle while the domains D4a and D4b are surrounded by blue rectangles.

2.2.1 Climatic Characteristics

Summer 2007: Unusually warm but at the same time nearly normally precipitated. These are the keywords for the summer season of 2007. The summer season of 2007 was interesting in terms of the succession of weather events which was rather uncommon for a mid European summer. Usually, sometime in June — or July the latest — a stable high pressure system develops around the Azores which subsequently reaches out over the Alpine region. In summer 2007, however, this high pressure system never grew strong enough to do that, which most prominently resulted in the more autumn-like weekly passage of upper air troughs and associated fronts across the evaluation domain between mid June and mid July. As a result temperature also rose and fell with this roughly weekly frequency. The last passage of a strong upper air trough occurred around July 12 and was followed by a long lasting southwesterly flow. This resulted in the warmest period of that summer between July 15 and July 20, with the absolute maximum of measured temperature (in Austria) being 39.5 °C (which was only 0.2 °C less than the Austrian all-time maximum temperature). The temperature level of that summer was very high in general. Even the mentioned upper air troughs in mid July could not sustainably drop this temperature level, although each trough advected comparably cool air masses to the Alps. The deviations from the climate mean therefore lie between +1.2 °C and +2.5 °C.

In terms of precipitation the summer 2007 was rather normal. Most parts of

the evaluation domain received their usual amount of precipitation. Only a few small regions received notably more precipitation, particularly in the southwestern parts of the evaluation domain and two areas in eastern Austria where thunderstorms locally brought very much precipitation (described in more detail in Subsection 2.2.2).

Winter 2007/08: The period from December 2007 to February 2008 was very warm and very dry, especially the last two months. The polar front was shifted pole wards, which enabled a for this time of the year unusual advection of subtropical, warm air mass to the Alps. At the same time the advected air mass was very dry, which resulted in a virtually precipitation free period between December 18 and December 30 in the whole evaluation domain. Especially the southeastern parts of the evaluation domain did further not receive noteworthy precipitation throughout January and February, but also in the rest of the evaluation domain days with notable precipitation amount were scarce.

This resulted in temperature deviations between +1.7 °C and +2.5 °C averaged over the whole season (maximum deviations in a single month partially exceeded +4.5 °C). The differences between climate mean precipitation amount and the one received in that particular season lie between -10 % and -70 %.

2.2.2 Striking Weather Events

Summer 2007: During the summer of 2007 several heavy precipitation events occurred in combination with thunderstorms including hail. Two of them are of major interest for the evaluation region D3. On June 20 and June 21 severe thunderstorms developed in a very warm and moist air mass advected in front of a strong low pressure system far off the evaluation domain (southwest of Ireland). Temperature in 850 hPa partly exceeded 18 °C. The thunderstorms occurred in the warmest parts of the air mass closely to the upper air ridge.

For several days in late August the Alpine region was loosely affected by a high pressure system centered over the East Atlantic. Temperatures widely rose above the 30 °C mark, and only local weak thunderstorms occurred. On August 28 an upper air trough with cool air mass attached reached the Alps and induced the development of widespread thunderstorms, which were severe especially in the southern and eastern parts of the evaluation domain. During the following two days the region was affected by a cold front passage which spawned further heavy precipitation events in its progress. The precipitation sums at many stations exceeded 50 mm within 24 hours, maximum temperature decreased from around 30 °C to about 20 °C.

A third weather event that is of special interest occurred on July 31. In the days before, the Alpine region was embedded in a steady westerly flow with accordingly

high temperatures. Even at night temperature did not fall below 10 °C at low altitude stations. In the night to July 31, however, an upper air trough crossed the Alps. This trough advected a cold air mass to the Alps and rendered the coldest night of that summer.

Winter 2007/08: In this season there are two weather events which are of major interest, both occurred in the second half of January. The first event occurred between January 22 and January 24. Before this event the Alpine region was covered by a very warm air mass, upper air temperature (at the 850 hPa level) reached +5 °C in most regions, especially in the southern part it even exceeded the +10 °C mark. On January 22, a rapid moving upper air trough approached the Alpine region from the northwest and crosses it in overnight to January 23. That night, temperature at 850 hPa widely drops below the -5 °C mark within a few hours. 48 hours later the Alps were again in a southwesterly flow and temperatures reached +5 °C.

Two days later, on January 26, the eastern Alps were affected by the passage of a strong low pressure system, the core of which had its path along the southern North Sea and Baltic Sea further to the East. On January 27 the highest measured wind speed was 165 km/h at an elevated station in D4b. But also stations in valleys measured wind speeds of up to and partly exceeding the 100 km/h mark.

2.3 Reference Dataset and Parameters

A major problem in evaluating high resolution RCM simulations is to find accurate highly resolved reference datasets. Generally used datasets for RCM evaluation, like E-OBS (Haylock et al. 2008), have grid sizes larger than 20 km. To find added value in convection-resolving climate simulations (CRCSs) the resolution of such reference datasets is normally much too coarse.

A pioneering observational dataset is provided by the Wegener Center for Climate and Global Change (WegCenter) at the University of Graz, which addresses exactly the above mentioned needs. The WegenerNet (Kirchengast et al. 2008) contains more than 150 climate stations in a mean distance of 1.4 km which are located in a region around Feldbach in south-eastern Styria within domain D4a. Gridded datasets with a resolution of 1 km × 1 km are available online on the WegenerNet dataportal (`http://www.wegenernet.org/`) since 2006. Since the WegenerNet covers only a small sub region of D4a it can not be used directly for the Local Climate Model Intercomparison Project (LocMIP) simulation evaluation.

For the evaluation in this study the Integrated Nowcasting through Comprehensive Analysis (INCA) dataset, provided by the Central Institute for Meteorology and Geodynamics (ZAMG), is used (Haiden et al. 2011). It has a 1 km × 1 km resolution and covers entire Austria. The INCA dataset is derived via a combination of numerical weather

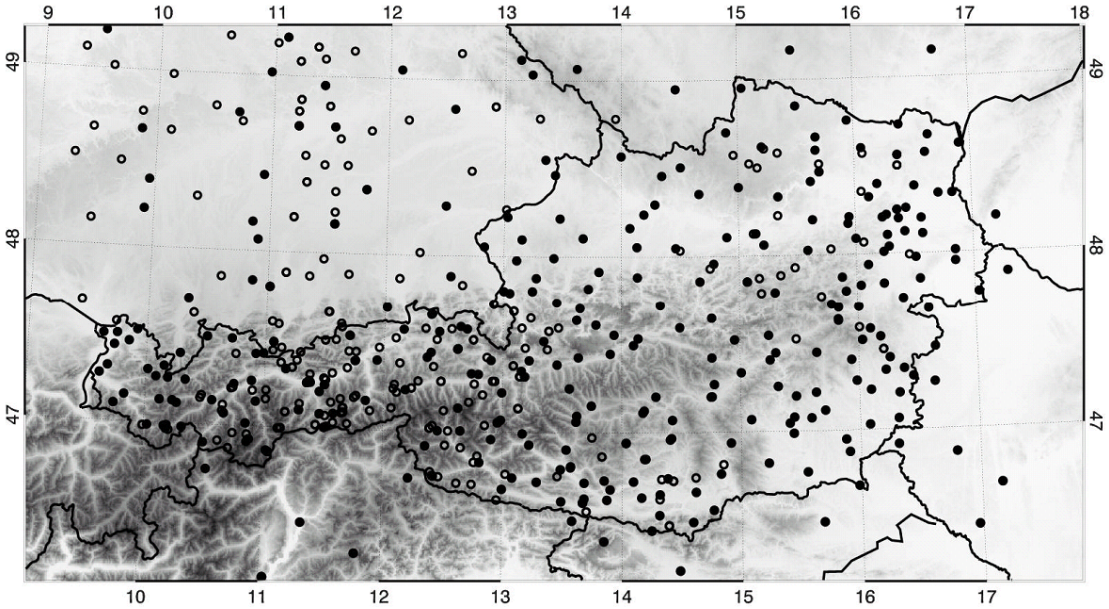


Figure 2.2: Full circles show the TAWES and SYNOP stations which are used in the hourly temperature and humidity analysis in INCA while open circles show hydrological stations (Haiden et al. 2011).

predictions (NWPs) (ALADIN, ECMWF) with current observation data from stations, radar, and satellites, and is further refined with highly resolved orographic information. The location of the stations which are integrated into INCA can be seen in Figure 2.2. INCA is based on a dense station network, especially in the mountainous region of domain D4b.

In this report the following four parameters are used for the evaluation: 2 m air temperature, precipitation, relative humidity and global radiation. With these parameters it is possible to get an insight in important physical processes of the RCMs and to give an overview on the overall performance of the simulations.

In order to estimate the feasibility of INCA for model evaluation it is compared to WegenerNet data in the corresponding area. The mean temperature bias in summer 2007 (panel a) in Figure 2.3 is nearly zero. However, there is an orography dependent bias with too warm temperatures in the valleys and too cold temperatures on the hills. The same is visible in winter 2007/08 (panel b) where the mean bias is also vanishing but the orography-dependent error is similarly structured than in summer. In both seasons the biases are very small in the closer area of the two meteorological stations which are integrated in the INCA dataset.

Figure 2.4 depicts the bias maps (INCA minus WegenerNet) for precipitation. In

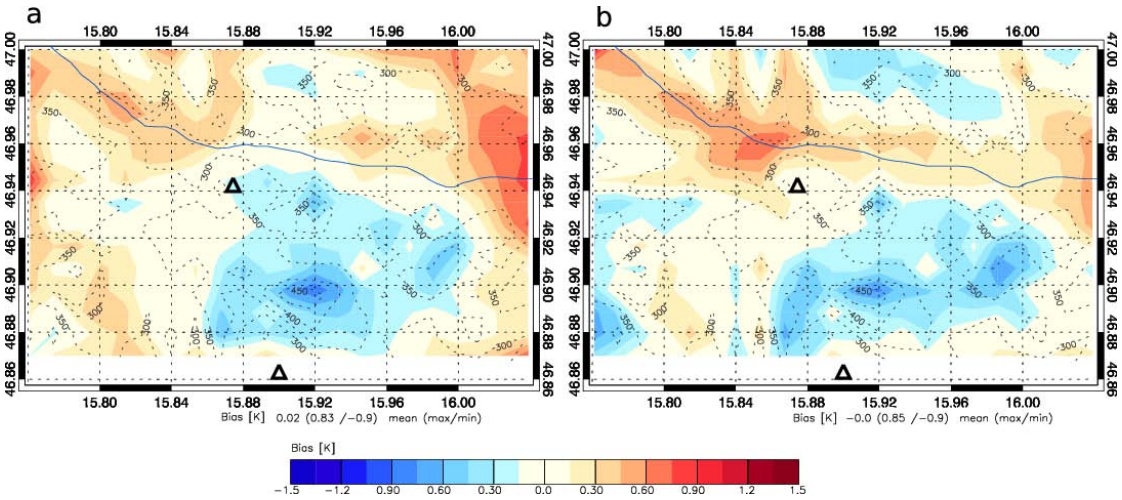


Figure 2.3: Bias maps of INCA (derived from WegenerNet) in the WegenerNet domain for air temperature in summer 2007 (panel a) and winter 2007/08 (panel b). Dashed contour lines are shown in the background to visualize the orography in this region. The two triangles display the locations of stations which are integrated in INCA in this region.

summer 2007 (panel a) INCA has a mean wet bias of 1.2 mm/d. Especially in the northern and in the eastern part of the domain, INCA has too much precipitation with a maximum of 2.7 mm/d. In particular, a small bias can be found in the region around Bad Gleichenberg (46.86°N, 15.90°O) and also in Feldbach (46.94°N, 15.87°O) where data of the observation stations are implemented in the INCA dataset. Also in winter 2007/08 (panel b) INCA has a wet bias of 0.8 mm/d. However, parts of this bias might be due to measurement errors in the WegenerNet, which uses unheated rain gauges.

As shown above 2 m temperature fields are in a good agreement between the INCA and the WegenerNet dataset. Following Haiden et al. (2011), typical errors are less than 0.3 K. Major problems can occur during nighttime stable conditions especially in deep Alpine valleys where no station are in the vicinity. The precipitation patterns should be well captured in INCA because radar data are used. The exact amount of precipitation can have errors of up to 50 % in summer and 100 % in winter if short time periods (15 minutes) and grid points are considered (Haiden et al. 2011). However, on scales of small catchments (100 km^2) the spatial averages are more reliable. Further comparisons with the WegenerNet data covering the complete year 2008, which have been conducted in a further research project funded by the Styrian Government (Gobiet et al. 2010), show similar results increasing the evidence for systematic errors of INCA with respect to the vertical temperature gradients.

It can be concluded, that INCA (a) underestimates air temperature in higher regions and overestimates temperature in valleys and (b) precipitation is overestimated in gen-

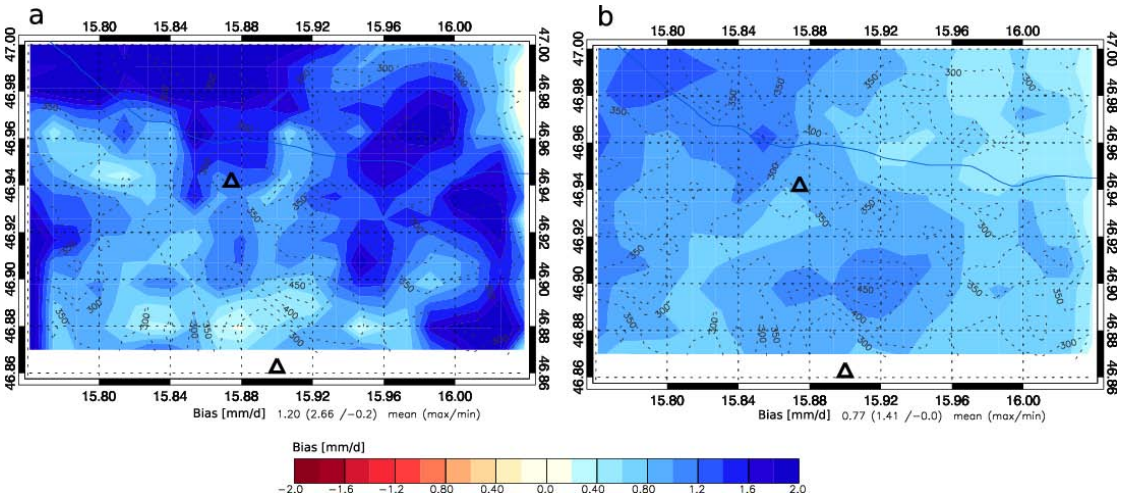


Figure 2.4: Same as figure Figure 2.3 but for precipitation [mm/d].

eral, but underestimations with notable extent occur during July and JJA (Gobiet et al. 2010) when convective conditions occur most frequently. A thorough evaluation of the INCA dataset lies outside the Non-Hydrostatic Climate Modelling (NHCM-1) project and is therefore set aside. It is just pointed out that INCA carries certain errors and that these errors have to be taken into account in any further evaluation where INCA is used as a reference dataset.

In Haiden et al. (2011) there is no information on how reliable the fields of global radiation or relative humidity are, but as long as the considered grid points are not too far away from observation sites, INCA is assumed to be sufficient reliable for the evaluations in this study.

2.4 Models and Simulation Setup

The lateral boundary conditionss (LBCs) for the LocMIP RCMs are taken from the Integrated Forecast System (IFS) of the European Centre for Medium-Range Weather Forecasts (ECMWF). Those data have a T799L91 resolution (roughly 25 km horizontal grid spacing and 91 vertical levels). A temporal resolution of three hours is achieved by combining IFS analysis and short term forecast fields. It is assumed that these boundary conditions represent the real weather conditions adequately, and hence the RCMs' performance can be judged apart from the quality of the LBC.

The initial conditions for the surface boundary conditions (SBC) for WegCenter COSMO model in CLimate Mode (CCLM) and Fifth-Generation Mesoscale Model (MM5) are taken from long term simulations, which started at the beginning of January 2007. This has the

Table 2.1: Overview of the RCMs, institutes, and contact persons contributing to the LocMIP ensemble.

RCM	Version	Institute	Responsible	Contact
CCLM	4.0	Wegener Center	M. Suklitsch	martin.suklitsch@uni-graz.at
CCLM	4.8	BTU Cottbus	G. Georgievski	goran.georgievski@tu-cottbus.de
MM5	3.7.4	Wegener Center	H. Truhetz	heimo.truhetz@uni-graz.at
WRF	2.2.1	Wegener Center	N. K. Awan	nauman.awan@edu.uni-graz.at

advantage that the soil, which has a long term memory for initial conditions (Seneviratne et al. 2006), is in a more balanced state at the beginning of the simulations. The Weather Research and Forecasting Model (WRF) and the Brandenburg University of Technology Cottbus (BTU) CCLM simulations have one month spin-up time.

The simulation periods consist of the two seasons DJF 2007/2008 and JJA 2007 plus one additional month for model spin-up. That means that the simulation periods range from 01-06-2007 to 01-09-2007 and from 01-12-2007 to 01-03-2008 (00 UTC each).

According to Table 2.1, four different RCMs (or versions of an RCM) (CCLM, MM5, and WRF) from two institutions contributed to the LocMIP ensemble (Table 2.1).

The key settings of the different RCM simulations are listed in Table 2.2. One major difference between the 10 km simulations and the 3 km simulations is switching off convection parameterizations and resolving convection by the RCMs explicitly due to the high resolution (3 km grid spacing).

Table 2.2: Listing of all LocMIP simulations and their key set-ups.

RCM/Acronym	Domain	Key settings
Weg-C CCLM / W_CLM10	D2	3rd Runge-Kutta numerics (time step: 80 s); Kain-Fritsch moist convection (Kain and Fritsch 1993; Kain 2003); cloud ice scheme with prognostic cloud water and cloud ice, prognostic rain and snow and transport of rain/snow; TKE-based turbulence scheme including sub-grid scale effects of condensation/evaporation, LBC update every 3 hours from mixed IFS T799 analysis and short-term forecasts

Table 2.2 – continued from previous page

RCM/Acronym	Domain	Key settings
Weg-C CCLM / W_CLM3	D3	3rd Runge-Kutta numerics (time step: 25 s); shallow convection; cloud ice scheme with prognostic cloud water and cloud ice, prognostic rain and snow and transport of rain/snow, and with graupel as additional prognostic variable; TKE-based turbulence scheme including subgrid scale effects of condensation/evaporation, LBC update every hour from W_CLM10 output
BTU CCLM / B_CLM10	D2	Leapfrog numerics (time step: 60 s); Tiedtke scheme convection parameterization; cloud ice scheme with prognostic cloud water and cloud ice, prognostic rain and snow and transport of rain/snow
BTU CCLM / B_CLM3	D3	2 time-level Runge-Kutta time-split scheme (time step: 25 s); shallow convection; cloud ice scheme with prognostic cloud water and cloud ice, prognostic rain and snow and transport of rain/snow, and with graupel as additional prognostic variable, three dimensional domain mask for horizontal diffusion; dynamical bottom boundary condition;
Weg-C MM5 / W_MM510	D2	2nd order leapfrog scheme (Grell et al. 1995) (time steps: 6.66 *second*) with time splitting scheme (Klemp and Wilhelmson 1978) for sound wave treatment; shortwave radiation scheme from Dudhia (1989) and the Rapid Radiative Transfer Model (RRTM) from Mlawer et al. (1997) for long wave radiation; microphysics scheme (REISNER 2) for phase transition of water adapted from Reisner et al. (1998); Kain-Fritsch cumulus scheme from Kain and Fritsch (1993) and Kain (2003); MRF boundary layer scheme from Hong and Pan (1996) together with Zilitinkevich formulation for stable stratification; atmosphere / soil interaction is treated by the NOAH land surface model (Chen and Dudhia 2001); the LBCs between finer and coarser model domains are updated with the model-internal time steps due to two-way coupling

Table 2.2 – continued from previous page

RCM/Acronym	Domain	Key settings
Weg-C MM5 / W_MM53	D3	2nd order leapfrog scheme (Grell et al. 1995) (time steps: 2.22 *second*) with time splitting scheme (Klemp and Wilhelmson 1978) for sound wave treatment; shortwave radiation scheme from Dudhia (1989) and the Rapid Radiative Transfer Model (RRTM) from Mlawer et al. (1997) for long wave radiation; microphysics scheme (REISNER 2) for phase transition of water adapted from Reisner et al. (1998); MRF boundary layer scheme from Hong and Pan (1996) together with Zilitinkevich formulation for stable stratification; atmosphere / soil interaction is treated by the NOAH land surface model (Chen and Dudhia 2001); the LBCs between finer and coarser model domains are updated with the model-internal time steps due to two-way coupling
Weg-C WRF / W_WRF10	D2	3rd order Runge-Kutta split-explicit time integration scheme (time steps: 20 *second*; Kain-Fritsch cumulus scheme from Kain (2003), no convection; shortwave radiation scheme from Dudhia (1989) and the Rapid Radiative Transfer Model (RRTM) from Mlawer et al. (1997) for long wave radiation; Eta grid-scale cloud and precipitation (2001) scheme (Rogers et al. 2001); Yonsei State University PBL scheme (Hong and Pan 1996) (plus Monin-Obukhov surface layer scheme after Skamarock et al. (2007)); atmosphere / soil interaction is treated by the NOAH land surface model (Chen and Dudhia 2001); the LBCs between finer and coarser model domains are updated with the model-internal time steps due to two-way coupling

Table 2.2 – continued from previous page

RCM/Acronym	Domain	Key settings
Weg-C WRF / W_WRF3	D3	Third-order Runge-Kutta split-explicit time integration scheme (time steps: 6.67 *second*; shortwave radiation scheme from Dudhia (1989) and the Rapid Radiative Transfer Model (RRTM) from Mlawer et al. (1997) for long wave radiation; Eta grid-scale cloud and precipitation (2001) scheme (Rogers et al. 2001); Yonsei State University PBL scheme (Hong et al. 2006) (plus Monin-Obukhov surface layer scheme after Skamarock et al. (2007)); atmosphere / soil interaction is treated by the NOAH land surface model (Chen and Dudhia 2001); the LBCs between finer and coarser model domains are updated with the model-internal time steps due to two-way coupling

2.5 Evaluation Methods

The here performed evaluation of the LocMIP ensemble is based on three basic error categories in which the performance of the RCMs might differ and added value in the convection resolving 3 km grid spacing simulations may become visible. The three error categories are concerning:

- mean climate,
- spatial characteristics, and
- temporal characteristics.

All methods are based on gridded datasets which share a common grid. Therefore, all simulations and the reference dataset are resampled onto a 3 km grid which is identical with the grid of the W_CLM3 simulation. The applied resampling technique is described in Riegler (2007) and accounts for conservation of area-averaging properties of gridded data. Temporally the reference and simulated data are investigated on an hourly basis which, together with the high spatial resolution, enables to evaluate the skill of the simulations in short time ranges and small scales. The evaluation methods used in each error category are discussed briefly in the following subsections and in more detail in Prein and Gobiet (2011).

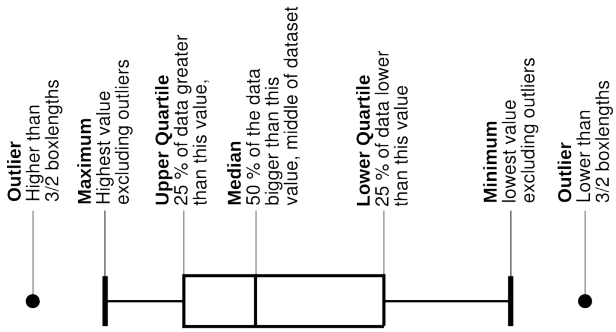


Figure 2.5: Statistical values behind the box-whisker plot. The line within the box shows the median of the considered data sample. The upper end of the box shows the 25 % and the lower end the 75 % quantile of the sample. The whiskers have a maximum length of one and a half of the length of the box. All points which are outside this distance are outliers and are displayed separately.

2.5.1 Evaluating the Performance in Mean Climate

Bias Maps are simple, but very informative to display the spatial variability of the mean bias. The reference data and the simulations are temporally averaged, and the differences between these averages are plotted for each grid point.

Box-Whisker Plots are used to display the error ranges of the bias on a quarter-daily basis. First, the data is spatially averaged over the model domain and six hourly data are calculated for the daytimes: night (22 UTC till 04 UTC), morning (04 UTC till 10 UTC), midday (10 UTC till 16 UTC), and evening (16 UTC till 22 UTC). This means that every daytime consists of 92 data points (number of days) in summer 2007 and 91 data points in winter 2007/08. From the distributions of these points the boxes and whiskers are generated. The background statistics for the here used box-whisker plot are shown in Figure 2.5.

Conditional-Quantile Plots: In this kind of plots certain aspects of the joint distribution of the reference and the simulation data are shown. The plot is separated in two parts, where the first one shows the conditional distribution and the second one shows the unconditional distribution of the data.

The conditional distribution is derived by separating the data range of a considered variable of the reference dataset into 100 parts of equal distant bins. Then, for each bin those grid points of the reference are considered that lie within the range of this bin (e.g., for 2 m temperature between 10 °C and 11 °C). Afterwards, the values of the simulation at these grid points are taken and the 50 % quantile is calculated and plotted against the mean value of the considered bin of the reference

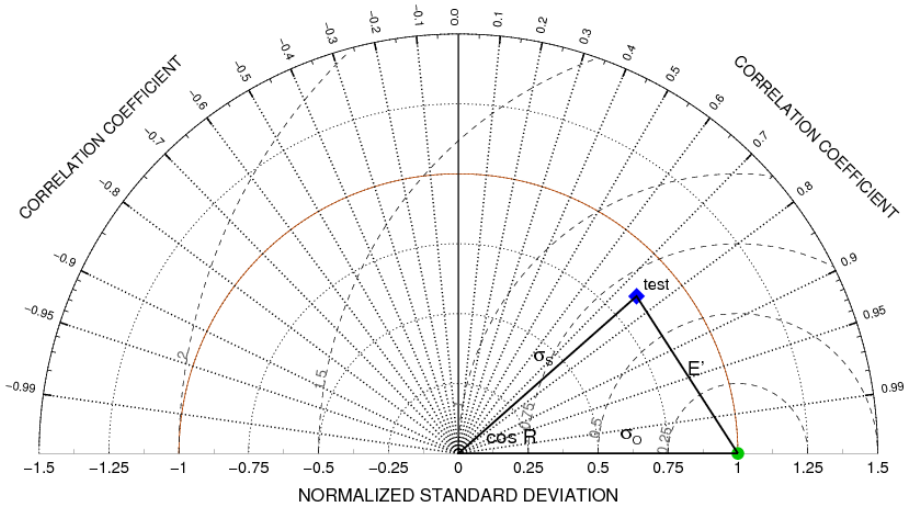


Figure 2.6: Example Taylor diagram for a test field. The grey triangle shows the relation between the correlation coefficient R, the standard deviations of the test σ_s and the reference σ_o, and the centered root mean square error E'.

dataset. This is applied to each bin. The 50 % quantile of a perfect simulation will therefore exactly lie on the 1:1 diagonal line.

The second part of the plot is the unconditional distribution and shows the density functions of the reference and the simulations.

A more detailed description of conditional-quantile plots can be found in Prein and Gobiet (2011), Wilks (2005), or Stevenson (2006).

2.5.2 Evaluating the Spatial Performance

Spatial Taylor Diagrams: Taylor diagrams are displaying four statistical properties in one single plot as in Figure 2.6: the x-axis refers to the normalized standard deviation of the observation (σ_o, which by design always equals 1), the y-axis refers to the normalized standard deviation of the modeled data (σ_s), the correlation coefficient R is the arccosine between the x-axis and the test data point, and the distance between the test and the observed data point shows the centered root mean squared error E'. For a detailed description of the Taylor diagram see Taylor (2001).

In this report, two kinds of Taylor diagrams are shown. The first, displays the spatial statistical parameters of the temporal averaged fields. The second contains the spatial statistical values of each hourly time slice. The density of these points in the Taylor diagram is then shown as contour where dark colors mark regions

with a high density. This method not only shows the spatial performance of the simulations but also contains information on the RCMs' ability to capture spatial characteristics on an hourly basis.

Fractional Skill Score (FSS): Evaluating the spatial characteristics of precipitation in a highly temporal and spatial resolution is one of the hardest tasks in verifying atmospheric simulations, because precipitation fields have a high spatial variability and are partly discontinuous and continuous within a region. Looking at very high resolutions, precipitation features on grit-point basis may get non-deterministic and unpredictable (e.g., intermittent convective rainfall).

Therefore, evaluating spatial precipitation fields with traditional statistical methods often leads to a 'double penalty' problem, because modeled and observed precipitation may not exactly match in space and time. E.g., even if spatial patterns are captured very well, the temporal correlation is worse when the pattern is shifted in time. To account for such evaluation problems special methods where developed (especially in the NWP community) which are summarized in Prein and Gobiet (2011).

In this report the Fractional Skill Score (FSS), a method which is often used in evaluating numerical weather prediction forecasts, is applied. It was published by Roberts and Lean (2008) and is based on the assumption that a useful simulation has a similar frequency of precipitation events as the observation. In the FSS method multiple neighborhood sizes and precipitation thresholds are taken into account to evaluate the dependency of skill on spatial scales and rainfall intensity.

2.5.3 Evaluating the Temporal Performance

Mean Diurnal Cycles: Many physical processes which are represented in RCMs have different characteristics during different daytimes. For example, the surface radiation budget is very different between day and night. Therefore, diurnal cycles can give valuable insights in the physical processes in a RCM and help to detect deficits in their representation. Here, the mean diurnal cycle of each simulation is plotted together with the mean diurnal cycle of the reference dataset.

Temporal Taylor Diagrams: Taylor diagrams can also be used for temporal evaluation. In this report, two kinds of Taylor diagrams are calculated for temporal evaluation. The first shows the spatially averaged temporal statistical values and can be seen as the ability to reproduce the mean time series of the whole domain. The second displays the temporal statistical values for each grid cell as a contour plot. It gives an insight on how accurate the RCMs are able to simulate temporal characteristics in different parts of the considered domain.

2.5.4 Relative Performance Index (RPI)

To give a holistic view on the overall performance of the simulations taking account for multiple statistical quantities, the relative performance index (RPI) is introduced, which shows the performance of the simulations in relation to the best and worst simulation in the ensemble. The RPI aggregates statistical quantities and maps them onto an interval between zero and one. The RPI is calculated for the following statistical values:

- bias and root mean squared error (RMSE) to represent the mean climate performance,
- spatial correlation coefficient, relative variability, and the mean FSS to show the spatial performance,
- and temporal correlation coefficient and relative variability to indicate the temporal characteristics.

For the bias, the RMSE, the correlation coefficient, and the FSS the RPI is calculated as follows:

$$RPI_i = 1 - \frac{ABS(S_i) - min(ABS(\vec{S}))}{max(ABS(S_i) - min(ABS(\vec{S})))} \tag{2.1}$$

where S_i is the considered statistical quantity for a certain simulation i, ABS means the absolute value and max and min, respectively, refers to the maximum and minimum of the statistical value in the simulated ensemble $\vec{S}$.

For the relative standard deviation the RPI is calculated slightly differently, because the best obtainable value is not zero as for the above mentioned statistical values but 1:

$$RPI_i = 1 - \frac{ABS(S_i - 1) - min(ABS(\vec{S} - 1))}{max(ABS(S_i - 1) - min(ABS(\vec{S} - 1)))} \tag{2.2}$$

3 Results and Discussion

In this chapter the evaluation of the Local Climate Model Intercomparison Project (LocMIP) regional climate models (RCMs) is presented and discussed. The evaluation concentrates on the performance of single simulations compared to the reference dataset Integrated Nowcasting through Comprehensive Analysis (INCA) as well as on the comparison of different simulations to each other (e.g., model spread, relative performance of single simulations compared to the ensemble). Furthermore, error ranges for 2 m temperature, precipitation, relative humidity and global radiation are described for the whole LocMIP ensemble, focussing on the comparison of 10 km and 3 km grid spacing simulations to show possible 'added value' in the convection-resolving RCM simulations.

All evaluations are either done on an hourly or quarter daily basis or on seasonal mean fields as described in Section 2.5. The individual sections in this chapter focus on different aspects which are important for showing the holistic skill of the simulations and to quantify possible added value of convection-resolving simulations. In Section 3.1 evaluations on temporal mean values are performed, while Section 3.2 concentrates on the spatial and Section 3.3 on the temporal characteristics of the simulations. Each of these sections closes with concluding remarks, summing up and discussing the most important findings. A summary of the performance evaluations is given in Section 3.4.

3.1 Mean Climate

To simulate the mean climate state on regional and local scales as accurately as possible is the main objective of RCMs. A better representation of small scale features (e.g., orography, surface fields, atmospheric flows, ...) by reducing the horizontal grid spacing is not only hoped to improve small scale but also large scale characteristics. For example, more realistic precipitation shadowing effects of mountains can occur because of a better representation of the orography (Laprice 2010). Furthermore, by decreasing the grid spacing beyond ∼3 km it gets possible to explicitly resolve deep convection which has high potential to improve common errors occurring through cumulus convection schemes (e.g., Grell et al. (2000), Hohenegger et al. (2008)).

In this section the simulations are analyzed with respect to their mean climate performance. The term 'mean climate' should be understood in a conceptional way, meaning that the methods used are able to verify climatic mean performance of a RCM simulation, even though only seasonal simulations are considered.

3.1.1 Bias Maps

In this subsection temporal mean biases are shown grid point wise in the form of maps containing the domains: eastern Alpine Region (D3), south-east Styria (D4a), and Hohe Tauern National Park (D4b). The bias maps of all eight simulations are arranged in one single figure, whereby figures are shown for every season and parameter.

2 m Air Temperature

Figure A.1 shows the bias maps for 2 m air temperature during June, July, and August (JJA) 2007. Even if there is no improvement in the mean bias between the 10 km (panel a, c, e, and g) and the 3 km (panel b, d, f, and h) simulations, all 3 km runs show improved maxima and minima biases and root mean squared error (RMSE) values (not shown). This can mainly be explained by the better representation of the orography in the 3 km simulations. Nonetheless, results are better in the hilly region D4a than in the mountainous region D4b.

In domain D3, the Fifth-Generation Mesoscale Model (MM5) simulations (panel c and d) have nearly no mean bias, whereas the Wegener Center for Climate and Global Change (WegCenter) COSMO model in CLimate Mode (CCLM) (panel g and h) simulations have a cold bias of -1.3 K and the Weather Research and Forecasting Model (WRF) runs have a warm bias of 0.3 K and 0.4 K, respectively.

In general the bias is smaller in D4a than in D4b (with exception of B_CLM3 which has a warm bias in this region). In D4b the bias structure is much more inhomogeneous because of the strong influence of the orography. Often warm biases (especially on mountain tops) and cold biases (especially in valleys) are cancelling out, so that the resulting mean bias is close to zero.

Bias maps for December, January, and February (DJF) 2007/2008 are shown in Figure A.2. In this season the RCMs are performing very differently. The simulations W_CLM10 and W_CLM3 have a strong cold bias of -3.4 K which is most pronounced in the inner Alpine region. Even the better representation of the orography in W_CLM3 does not improve the biases. The B_CLM10 simulation has also a slight cold bias which is improved in the B_CLM3 simulation. The WRF and MM5 simulations show an inhomogeneous bias field whereby, regions with warm biases and regions with cold biases are nearly cancelling out in the MM5 runs. In the WRF simulations the warm bias dominates and results in about 0.8 K in D3.

Results are better in D4a than in D4b. A remarkable improved bias field can be seen between B_CLM10 and B_CLM3 in D4b: in winter the mean bias decreases from -1.4 K in the 10 km simulation to -0.7 K in the 3 km simulation, and also the maximum and minimum biases are getting smaller.

Precipitation

There is a general tendency to overestimate precipitation in mountainous regions and underestimate it in flat and hilly areas in JJA 2007 (see Figure A.3). In the CCLM simulations the increasing resolution and the thereby explicit resolved deep convection causes an increase in precipitation in the entire domain D3, particularly in mountainous areas. Thereby, large wet biases of 1.1 mm/d in W_CLM3 (panel h) and 2.2 mm/d in B_CLM3 (panel f) appear. No major differences occur between the two simulations of MM5 and WRF. In all runs, the maximum precipitation biases are increasing in the 3 km simulations.

Considering the sub-domains, there is generally too little precipitation generated (except in B_CLM3) in D4a, while in D4b there is too much precipitation (except in the WRF simulations).

Figure A.4 depicts that there is a clear overestimation of precipitation in the alpine area in DJF 2007/08 which gets even more pronounced in the 3 km simulations. This is most probably caused by orographic precipitation, which increases when the orography is better represented, and by the micro physics scheme which is responsible for the phase transition of water. All simulations overestimate the precipitation amount in D3. Thereby, the typical mean biases are beneath 1 mm/d.

Similar to summer 2007, there is generally too much precipitation in domain D4b whereas the mean precipitation in domain D4a is very well represented (within a range of ± 0.2 mm/d).

Relative Humidity

As shown in Figure A.5 the typical mean biases for relative humidity in summer in domain D3 are between 0.4 % in W_CLM3 and -10.2 % in W_WRF10. All MM5 and

WRF simulations show a dry bias which is strongest in the southern part of D3. The valleys and basins like the Drau valley, in W_MM510 and W_WRF3 are particularly dry. Such effects are not visible in the CCLM simulations. In CCLM there is a general north-west, south-east gradient. On the north-western corner of D3 there is a wet bias and on the south-eastern side a dry bias. This structure is especially pronounced in W_CLM10 and W_CLM3.

Relative humidity in domain D4a is generally too low. In domain D4b all CCLM simulations are too humid and all MM5 and WRF simulations are too arid.

In winter the mean biases in relative humidity are between 10.8 % in W_CLM10 and -10.3 % in W_WRF10 (shown in Figure A.6). The spatial structure looks similar to summer 2007, since MM5 and WRF are too dry and all CCLM simulations have a strong wet bias in the north-eastern corner of D3. Also the dry bias in valleys and basins in MM5 and WRF occurs in winter and gets even more pronounced in the 3 km simulations.

Relative humidity is generally well represented in D4a in winter (mean biases between -6.1 % and 1.9 %). Biases in domain D4b are smallest in B_CLM10 and B_CLM3. Strong dry biases occur in MM5 and WRF while a heavy wet bias is present in W_CLM10 and W_CLM3.

Global Radiation

The spatial bias characteristics for global radiation in summer 2007 are very different between the RCMs (Figure A.7). CCLM has a negative bias which is larger in the WegCenter simulations than in the Brandenburg University of Technology Cottbus (BTU) simulations. The negative bias is decreasing in both CCLM 3 km simulations. The MM5 and the WRF simulations show very similar bias characteristics. Both models perform well in the 10 km grid spacing simulations, but have a positive global radiation bias in the 3 km simulations (W_MM53 and W_WRF3). The reason for this behavior might be related to a wrong representation of cloud forming processes and a therefore increased global radiation at the surface. The high variable spatial structure of the bias, where areas with high biases are close to bias free areas (especially in mountainous regions, e.g., domain D4b) indicates a strong relationship between mountain tops and cloud-cover in W_MM53 and W_WRF3. In domain D4a the spatial bias structure is more homogeneous and the biases are generally smaller.

In winter (Figure A.8) biases in global radiation are much smaller than in summer, which is probably caused by the prevalence of stratiform clouds (compared to convective clouds which often occur in summer) and the lower amount of incoming solar radiation. Only W_CLM10 and W_CLM3 show stronger biases in large areas of D3, whereas the other simulations perform very well. The WegCenter CCLM simulations have a well pronounced positive global radiation bias in the mountains and a slightly negative bias in flat areas. Differences between the 10 km and the 3 km grid spacing simulations are marginal.

W_CLM10 and W_CLM3 show a positive global radiation bias in the domain D4b. Around the highest peaks of Austria, the Großglockner massif, a strong positive radiation bias occurs in particular in W_CLM10 and W_CLM3 but also in the other simulations. This might be caused by a too low cloud top where the highest peaks are breaking through. However, especially in winter, the cloud top height in INCA is also error prone. In the hilly region of domain D4a all models, except the WRF simulations, show a good performance with respect to the mean bias.

3.1.2 Error Ranges in the Diurnal Circle

RCMs often have different error characteristics at different times of a day since physical processes can alter strongly during daytimes. In this chapter, biases are shown for four six hourly mean time slices (night (22 UTC till 04 UTC), morning (04 UTC till 10 UTC), midday (10 UTC till 16 UTC) and evening (16 UTC till 22 UTC). From this six-hourly data box-whisker plots are generated. This means, every box-whisker is built out of 92 values in JJA and 91 in DJF corresponding to the number of days in the season. As a measure for the error spread, generally the inter whisker distance is considered.

2 m Temperature

MM5 has the largest error ranges in domain D3 (Figure A.9 panel a) with a maximum whisker to whisker spread of about 10 K at midday and evening. The other RCM simulations have a typical error range between 5 K and 6 K. There is no general difference between the 10 km grid spacing simulations and the 3 km simulations.

Median biases of MM5 are slightly negative during night and morning and positive during midday and evening. WegCenter CCLM simulations are generally too cold but especially in the evening and night, while WRF performs especially well during these times but is too warm during morning and midday. The BTU CCLM simulations show best results during morning, midday and evening, but they are too warm at night.

In domain D4a (panel b) the uncertainties are slightly increased by about 30 %. The mean biases of MM5, WRF and BTU CCLM stay the same, but the WegCenter CCLM simulations have a warm bias during midday. In domain D4b (panel c) uncertainties are also slightly increased by ∼30 % compared to the spread in D3. The median biases look very similar to those in D3 for all simulations.

As in summer (Figure A.9) also in winter Figure A.10 MM5 has the largest error range with ∼10 K (upper to lower whisker distance) in D3. The other RCMs have typically an error range of 5 K to 7 K. WegCenter CCLM has a significant cold bias of ∼-3 K in median during the whole day. The other simulations are within a median bias range of ±2 K.

In domain D4a (panel b) the error ranges are almost doubling in the CCLM and WRF simulations. The cold bias of WegCenter CCLM is smaller compared to those in D3. The error ranges are smaller in D4b than those in D4a and are partly even smaller than those

in D3 (e.g., WRF at night and morning). The cold bias of WegCenter CCLM with the range between -4 K and -5 K is larger than in D3. Thereby, W_CLM3 is even colder than W_CLM10.

Precipitation

In the box-whisker plots for precipitation only hourly precipitation values above 1 mm/d are considered. This means that the mean biases which are shown in the former section (see Figure A.3 and Figure A.4) differ from the biases here, because only hours which are either rainy in the observation or in the simulation are considered in these plots. Hence, a direct comparability of the precipitation bias in the bias maps in Figure A.3 with the biases represented in the box-whisker plots is not possible.

Figure A.11 shows biases for summertime precipitation. There are many underestimated (or missed) and many overestimated (or falls alarm) events which is reflected in the large error ranges (inter-whisker distance) and the outliers. In domain D3 the largest error ranges occur in the evening, because between 16 UTC and 22 UTC there are the most thunderstorms in summer and the diurnal cycle has its maximum here.

In domain D4a (panel b) most RCM simulations are too dry. This becomes especially visible when the mean bias is considered. Only B_CLM3 has a good representation of the median and mean bias. In D4b (panel c) the error ranges are largest. Notably is the overestimation of precipitation by the 3 km CCLM simulations at midday and in the evening compared to their 10 km parent simulations. This effect is most probably caused by an overshoot of orographically induced convective rainfall in the 3 km simulations.

Precipitation biases are generally smaller in winter than in summer because there is less convective rainfall and synoptic scale phenomena are dominating (Figure A.12). The error ranges in domain D3 (panel a) are lowest during nighttime. There are only minor differences between the 10 km and the corresponding 3 km simulations. However, the large error ranges in the MM5 simulations remain.

In domain D4a (panel b) error ranges are generally smaller than in D4b because there is very little rainfall during winter in this region. In domain D4b the error ranges are largest compared to the other domains.

Relative Humidity

Biases of relative humidity during JJA 2007 are shown in Figure A.13. In domain D3 (panel a) the error ranges are smallest at night and largest in the evening. This is most probably related to the summertime convective precipitation events in the afternoon, which have a significant effect on relative humidity in 2 m height and also on temperature and global radiation. The differences between 10 km and 3 km grid spacing simulations are generally small. The largest differences occur between the B_CLM10 and the B_CLM3 simulation whereby B_CLM10 performs very well in the night and

morning and B_CLM3 has a smaller median bias during midday and evening. MM5 and WRF are both too dry during the whole day.

All RCMs simulate too low relative humidity values in D4a (panel b) with the exception of the B_CLM10 simulation during midday and evening. In domain D4a the CCLM simulations are generally too humid and the MM5 and WRF simulations are too dry.

Comparing Figure A.13 with Figure A.14 the error ranges are generally larger in winter than in summer. In D3 (panel a) MM5 has the largest error ranges with a more than 30 % higher inter whisker distance than the other simulations. WegCenter CCLM simulations are too humid with the largest median biases at midday while WRF is generally too dry.

In domain D4a the error ranges are generally larger than in D3 but the median biases are smaller. The largest error ranges occur in D4b where for example MM5 has an inter-quantile distance of more than 80 % relative humidity in the morning and in the evening. The median biases are similar to those in D3.

Global Radiation

Since global radiation has a strongly pronounced diurnal cycle with a maximum at high noon and a minimum from dusk till dawn, the error ranges show a similar characteristic (Figure A.15). WegCenter CCLM simulations have generally a negative bias in global radiation with a minimum median of ∼-200 W/m^2 at midday. The W_CLM3 simulation performs thereby slightly better than the W_CLM10 simulation. Generally, the 3 km simulations are showing higher global radiation median values than the corresponding 10 km simulations.

In domain D4a (panel b) the negative biases of WegCenter CCLM are smaller but the main characteristics are similar to the results of D3 (panel a). In domain D4b (panel c) error ranges are larger than in D3. Especially, in W_MM510 the inter-whisker distance is very large (1100 W/m^2).

Figure A.16 depicts the box-whisker plot for global radiation in DJF 2007/08. The error ranges are much smaller in winter because of convective processes in summer and the generally lower amount of global radiation during DJF. The largest error ranges are found in the MM5 simulations with 400 W/m^2 inter-whisker distance during midday in D3 (panel a). Larger differences between the 10 km and the 3 km grid spacing simulations occur only in the B_CLM10 and the B_CLM3 simulation, whereby the higher grid spacing improves the median biases.

Results in D4a (panel b) are similar to those in D3 with the exception of the WegCenter CCLM simulations which have a positive bias in D3 and a negative in D4a. In domain D4b (panel c) the positive bias of the WegCenter CCLM simulations is even larger than in D3 while the other median biases stay relatively constant.

3.1.3 Conditional Quantile Plots

As described in Section 2.5, conditional quantile plots show properties of the joint distribution of the simulated and the reference space-time field. With this method two important features can be validated: first, weather the simulation shares the same distribution than the reference and second, the performance of the simulations at specific values of the parameter space (e.g., between -10 °C and -5 °C).

The basic structure of this chapter is the same as in Section 3.1.1: all parameters are discussed separately for all seasons and domains.

2 m Temperature

Conditional quantile plots are shown for 2 m temperature in JJA 2007 in Figure A.17. In domain D3 (panel a) all simulations show a good performance since the median temperatures are close to the diagonal. Comparing the 10 km with the 3 km simulations, there is no remarkable difference in major parts of the temperature spectrum. However, especially at low temperatures the 3 km grid spacing simulations stay closer to the diagonal and are therefore more consistent with the reference. The density function on the bottom of panel a shows that the maximum is slightly shifted by -3 K in W_CLM10 and W_CLM3 and that the distribution is broader than the reference. The best fit compared to the reference distribution have the WRF simulations W_WRF10 and W_WRF3.

In domain D4a (panel b) W_CLM3 has a warm bias between 1 °C and 10 °C. Above 20 °C the B_CLM10 has a cold bias of several degrees.

Looking at the density functions both BTU simulations are too peaked and have a shift in maxima of ∼2 K.

A general cold bias at warm temperatures and a warm bias at cold temperatures is depicted for domain D4b (panel c). The strongest warm bias at cold temperatures occurs in W_WRF10 while the strongest cold bias at warm temperatures shows the W_CLM3 simulation. Similar to D3 the maxima of the density functions of W_CLM10 and W_CLM3 are shifted by -2 °C. Those two simulations also show a remarkable peak at zero degrees.

Compared to JJA 2007, in DJF 2007/08 larger differences between the different models occur (see Figure A.18 panel a). While both WRF and MM5 simulations perform well over the whole temperature spectrum, the BTU CCLM simulations have a cold bias above 0 °C and the WegCenter CCLM simulations are too cold in the whole temperature spectrum but especially above 0 °C. Also the density functions of MM5 and WRF look much similar to those of the reference than those of the CCLM simulations which all have a zero degree peak. This peak is particularly strong and sharp in W_CLM10 and W_CLM3.

The performances in domain D4a and D4b (panel b and c) is similar to those in D3. The zero degree peak is sharply pronounced in the W_CLM10 and W_CLM3 simulations in D4a, while in D4b the density functions of those two simulations are shifted by several

degrees too colder temperatures. The B_CLM10 and B_CLM3 simulations start to get a stronger cold bias above 0 °C which is slightly improved in the 3 km grid spacing simulation.

Precipitation

Figure A.19 shows the conditional quantile plots for precipitation in JJA 2007. In domain D3 and its sub domains, simulations were not able to produce (in median) precipitation at the same time and place than the reference. A few exceptions can be seen for example in domain D4a where a few simulated precipitation events matched the position and time of the reference precipitation between ∼15 mm/d and ∼25 mm/d (even if they underestimated the corresponding precipitation amount).

Looking at the density functions of domain D3 (lower plot in panel a), 3 km grid spacing simulations are producing higher precipitation amounts and higher intensities than the 10 km simulations. Especially W_CLM3 shows much more precipitation and higher intensities than the W_CLM10 simulation. However, these events do not match the place and time of the reference data, which is visible in the conditional part of the distribution (upper plot in panel a).

In domain D4a, nearly all simulation overestimate the amount of intensive precipitation (above 6 mm/d). However, they underestimate precipitation between 1 mm/d and 6 mm/d. In domain D4b orographic induced precipitation is strongly overestimated. Only MM5 shows an underestimation between 1 mm/d and 4 mm/d.

In winter RCMs have more skill in simulating precipitation at the same place and time than the reference dataset (Figure A.20). Nevertheless, the median amount of precipitation is generally too low. While all simulations show similar results in D3 (panel a) and D4b (panel c), B_CLM10 and B_CLM3 are clearly better in representing the precipitation field in D4a (panel b). Especially the B_CLM3 simulation has a good performance across the whole precipitation spectra while B_CLM10 is worse between 4 mm/d and 6 mm/d.

Concerning the density functions in domain D3, all simulations show too high precipitation frequencies. Only the MM5 and WegCenter CCLM simulations have too little intensive rainfall. In domain D4a the RCMs are able to capture the density function of the reference to a great extent, while in domain D4b there is again an overestimation of rainfall.

Relative Humidity

Figure A.21 displays the conditional quantile plot of D3 (panel a in Figure A.21). MM5 and WRF are showing too dry relative humidity values and CCLM simulations are too humid across large parts of the spectra. A difference between the CCLM simulations is, that B_CLM10 is more humid than B_CLM3 while W_CLM10 is generally drier than

the W_CLM3 simulation.

In domain D4a (panel b) all simulations are too dry, especially at higher relative humidity values. Only the B_CLM10 simulation is too humid between 20 % and 80 %. The situation in D4b (panel c) is similar than in D3. CCLM is too humid below 90 % relative humidity and MM5 and WRF is too dry above 40 % relative humidity.

Most simulations are not able to represent the density functions properly. Especially the characteristic sudden increase at the right end of distribution representing fog, low level clouds, and precipitation events is only visible in W_CLM10 and W_CLM3 and partly in B_CLM3 in domain D4b.

As shown in Figure A.22, RCMs have more problems to reproduce relative humidity in winter than in summer (Figure A.21). Especially, the W_CLM10 and W_CLM3 simulations show nearly constant high median relative humidity values (between 80 % and 90 %) in D3 and D4b. B_CLM10 and B_CLM3 perform slightly better, but also suffer from too humid conditions. WRF and MM5 simulations are looking similar and are too humid under 50 % relative humidity and too dry above.

As in summer the RCMs are not able to generate the sudden increase at the right end of the density function of relative humidity. The only exceptions are the W_CLM10 and the W_CLM3 simulations. However, these simulations show a significant peak at ∼90 % relative humidity, which is not visible in the reference dataset. The best representation of the density function is given by the WRF and MM5 simulations.

Global Radiation

Figure A.23 displays the conditional quantile plots for global radiation in JJA 2007. The wave-like structure in the median of the conditional distribution occurs most likely because of the update frequency in the radiation schemes of the models (hourly in CCLM and half hourly in WRF and MM5). Furthermore, all RCMs show a negative median above ∼900 W/m^2. However, this causes no large average biases, because there is only a small amount of values above 900 W/m^2 as shown in the distribution functions. A negative median value of global radiation can be seen in all CCLM simulations, whereby the 3 km grid spacing simulations are slightly improving the results (e.g., compare W_CLM10 to W_CLM3 in D4b panel c).

The density functions of W_CLM10 and W_CLM3 show too few values from 50 W/m^2 onwards. In these simulations values above 850 W/m^2 do actually not exist. Too much values above 850 W/m^2 can however be found especially in the 3 km simulations of MM5 and WRF. For example in domain D4a (panel b) a strong local maximum is visible at ∼950 W/m^2 which is most pronounced in W_MM53 and W_WRF3.

In DJF 2007/08 RCMs have problems with representing high global radiation values above ∼600 W/m^2 (Figure A.24). However, they have an overall good performance in the conditional distribution of D3 (panel a). Especially the MM5 and WRF simulations are close to the diagonal line in D3 and the sub domains. W_CLM10 and W_CLM3 show an

underestimation of global radiation in D4a above 150 W/m^2 and a general overestimation in D4b.

The density functions are worse, especially at low global radiation values. The strong increase in the reference density function below 150 W/m^2 in D3 is only partially captured by W_MM510 and W_MM53 and missed by the other simulations. A similar behaviour is shown in domain D4a (panel b).

3.1.4 Concluding Remarks: Mean Climate

Concerning mean climate performance, differences between RCMs are much larger than differences between RCM simulations with different horizontal grid spacings. Thereby, WRF and MM5 share similar characteristics in their performance and also the two versions of CCLM do so. However, in case of CCLM the newer version CCLM 4.8 outperforms the older version CCLM 4.0 in multiple ways (e.g., cold bias in winter, global radiation bias in summer, 0 °C peak). The reason for the better performance of the B_CLM simulations can be found in model development (including multiple bug fixes especially during the work in the project cold bias (Keuler et al. 2010)) and the different settings of the W_CLM and the B_CLM simulations (see Table 2.2). In particular the usage of different time integration schemes in the 10 km simulations (in W_CLM10 3rd order Runge-Kutta and in B_CLM10 leapfrog) are likely to have major impacts on the performance differences.

Added value of the convection-resolving 3 km simulations compared to their 10 km parent simulations can be found in the representation of 2 m temperature fields where the 3 km simulations improve the maxima and minima biases and the area mean RMSE in D3 and its sub domains.

More intense precipitation can generally be found in the 3 km simulations, in particular in mountainous areas where the higher resolved model surface enhances orographic precipitation. In hilly areas like the sub domain D4a, almost all simulations show too little precipitation. Thereby, especially medium precipitation between 1 mm/d and 6 mm/d is underestimated while intense rainfall events are overestimated in particular by the 3 km simulations. Convective precipitation events in summer lead too large error ranges in the evening between 16 UTC and 22 UTC. In the winter season errors are generally smaller because of the predominance of precipitation related to frontal passages which are easier to simulate than summertime convective precipitation.

Relative humidity values are underestimated by WRF and MM5 while CCLM shows too humid conditions in the north east part of D3 and too dry in the south west. In summer the error ranges of relative humidity are largest in the evening caused by convective precipitation events. Generally, error ranges are larger in winter than in summer, whereby RCMs have in particular an overestimation of dry conditions below ∼35 % relative humidity. Furthermore, in winter a connection between 2 m temperature and relative humidity exists: areas with negative temperature biases typically correspond with areas with too

high relative humidity and vice versa (see Figure A.2 and Figure A.6).

The 10 km simulations seem to produce more widespread cloud cover compared to the 3 km simulations in which small convective cells are prevalent. Thereby, global radiation is generally increased in the 3 km simulations. In winter biases in global radiation are generally smaller than in summer. The major reason is most probably related to the predominance of stratiform clouds in winter which are easier to simulate by RCMs than convective clouds (which typically occur in summer) and the generally smaller amount of incoming solar radiation. In the conditional-quantile plots of global radiation a wave-like structure becomes visible, which might be related to the update frequency of the radiation scheme (in CCLM hourly and in WRF and MM5 half-hourly).

The bad performance of both W_CLM simulations in mountainous regions in winter (especially the cold bias in the 2 m temperature and the constant high relative humidity values) can probably be drawn back to a problem in the snow parameterization scheme. It is possible that much of the incoming solar energy is consumed by melting or sublimating the snow cover, whereby the thereby produced moisture input lead to constant high relative humidity values. Furthermore, the consumption of solar incoming energy through phase change processes of snow can also explain the severe zero degree Celsius peak in the W_CLM simulations.

3.2 Spatial Characteristics

In this chapter the spatial characteristics of the simulations are analyzed. For all parameters, Taylor diagrams are shown in two different versions. The first shows the spatial correlation coefficients, the centered pattern RMSE differences, and the normalized standard deviations calculated for every time-step (every hour) as a density contour. The second shows the above mentioned values for the temporal averaged field.

Since, for precipitation small displacement errors (shifts in time or space) can have big impacts on values like the spatial RMSE or the correlation ('double penalty problem'), the Fractional Skill Score (FSS) is used to analyze the performance of the simulation of precipitation.

3.2.1 Spatial Taylor Diagrams

2 m Temperature

Figure A.25 shows the Taylor diagram for the temporal averaged 2 m temperature field in summer 2007. In domain D3 (panel a) all simulations have a correlation coefficient above 0.9. The 3 km simulations have correlation coefficients $\geq$0.97, which is larger than the correlation coefficient of their corresponding 10 km simulations. This indicates added value due to higher resolution. However, since averaged fields are analyzed positive and negative errors might cancel each other. Concerning the relative standard deviation, the simulations are within a range of 0.91 (W_WRF10) and 1.17 (W_CLM10).

In domain D4a (panel b), the 3 km grid spacing simulations also have a higher correlation coefficient with the exception of W_CLM3. However, the spread of the standard deviations is larger compared too D3 and there is no clear improvement in the 3 km simulations. While the W_CLM10 is overestimating the spatial variability by $\sim$40 %, W_CLM3 is underestimating it by the same amount. Very good performances have the W_MM53 and W_WRF3 simulations.

Looking at domain D4b (panel c), there is a clear improvement from the 10 km simulations to the 3 km simulations in terms of correlation coefficient. Also the spatial variability is getting better for all 3 km simulations (except W_CLM3).

Figure A.26 shows the same data which is presented in Figure A.25 but not as a temporal averaged field but for every time slice. These hourly data-points are then used to build the density expressing contours. Using this method, it is possible to estimate the temporal variability of the spatial correlation coefficient, the normalized spatial variances, and the centered RMSE.

As shown in panel a) in Figure A.26 the higher densities of the contours are approximately at the same position than in Figure A.25 indicating, that spatial structures on hourly basis and seasonal basis are approximately equally well represented. The 3 km grid spacing contours generally have a higher correlation coefficient than the 10 km simulations. However, the contours are overlapping to a great extend, indicating that the

added value in spatial correlation might not be significant.

Looking at the results of D4a (panel b), area covered by the contours get larger compared to D3. Especially the MM5 simulations have a large variability in its statistical parameters. In D4b (panel c), contours are much denser than in D4a. Here, the 3 km simulations of MM5, WRF, and BTU CCLM are significantly better (meaning the contours are totally separated) than the 10 km simulations, indicating a robust improvement due to higher resolution.

In DJF 2007/08 (Figure A.27) the RCMs have worse performance than in summer (Figure A.25). In domain D3 all simulations have correlation coefficients above 0.79. The spatial variability is commonly beneath 0.98 (W_MM510) and 1.25 (B_CLM10) except for the WegCenter CCLM simulations. In the W_CLM10 simulation the spatial variability is overestimated by 52 % and in W_CLM3 by 70 %. The 3 km grid spacing simulations show a better correlation coefficient then the 10 km simulations, except the WegCenter CCLM simulations.

In domain D4a (panel b), the 10 km simulations of WRF and MM5 have very low correlation coefficients. However the 3 km simulations are improving the correlation remarkably. For the WegCenter CCLM simulations it is the other way around: W_CLM10 has a correlation of 0.86 while W_CLM3 has 0.17. Furthermore W_CLM10 overestimates while W_CLM3 underestimates the standard deviation significantly.

Increasing correlation coefficients along higher resolutions are also shown in domain D4b (panel c). The only exception is again W_CLM3. The overall best performance in D4b has the B_CLM3 simulation with a correlation coefficient of 0.92 and a relative spatial standard deviation of 1.08.

The contours are overlapping to a great extend in the Taylor plot for 2 m temperature at DJF 2007/08 in domain D3 (Figure A.28 in panel a) indicating that the higher correlation coefficients in the 3 km simulations might not be significant. Compared to the mean Taylor plot in Figure A.27 the WegCenter CCLM simulations show a better representation of the spatial variability here.

In domain D4a (panel b) the spread of the RCMs ensemble is increased compared to D3. Like in Figure A.27, the W_CLM3 has a much smaller correlation coefficient than the W_CLM10 simulation.

The RCMs results are much more clustered in D4b (panel b) than on the other domains, pronouncing the strong influence of orography on 2 m temperature also on hourly time scale. There is a clear separation between the 3 km simulations and the 10 km simulations and hence the 3 km simulations have significantly higher correlation coefficients. Only in W_CLM3 the correlation coefficients are not improved.

Precipitation

The Taylor plots for the temporal mean precipitation field in summer 2007 are shown in Figure A.29. In domain D3 (panel a) the correlation coefficients are between 0.39

(W_MM53) and 0.66 (W_WRF10). The 3 km grid spacing simulations have generally a higher spatial standard deviation than the corresponding 10 km simulations. The MM5 simulations have a too low spatial variability while the other RCMs have too high variability.

WRF and BTU CCLM have higher correlation coefficients in D4a (panel b) than in D3 while MM5 simulations and W_CLM10 show similar results. W_CLM3 has a very good representation of the spatial variability, but only a correlation coefficient of 0.4.

In domain D4b (panel c) the performance of the RCMs is generally worse than in the other domains. Best correlation coefficients have been found in the MM5 simulations, but they overestimate the spatial variability by ∼100 %. All 3 km simulations have higher spatial variabilities than their correspondent 10 km simulations.

The hourly Taylor plots of the spatial precipitation fields in summer 2007 (Figure A.30) look very different to the temporal mean plots (Figure A.29) because the RCMs often miss precipitation events or, simulate precipitation at the wrong time or place. This double penalty problem can drastically be seen in Figure A.30. In all three domains the correlation coefficients are between ∼-0.15 to ∼0.4. In domain D3 (panel a) the W_CLM3 and also the W_CLM10 simulation in D4b have some skill in simulating the spatial variability. There are many scattered data points outside the borders of the plot (with relative spatial variabilities greater than 3). They appear because there is a lot of simulated precipitation where none or few is observed (e.g., if only one precipitation cell occurs in the reference data but multiple cells are simulated, the division by the very low reference spatial variability yield too large values in the normalized variability).

The temporal mean spatial variability in precipitation is overestimated by all RCMs during winter 2007/08 Figure A.31. Especially the CCLM simulations show an overestimation in domain D3 (panel a) (70 % in B_CLM10, 120 % in W_CLM3, 285 % in W_CLM3, and even 320 % in W_CLM10). Thereby the correlations are relatively good and within the range of 0.82 (W_WRF10) and 0.85 (B_CLM10).

In domain D4a (panel b) the CCLM simulations show a very high spatial variability. The best simulation is the W_MM53 with a correlation coefficient of 0.82 and a relative spatial variability of 1.26.

In domain D4b (panel c) only the CCLM simulations show a well performance in terms of spatial variability (except B_CLM3). The other simulations but especially the WRF simulations have an overestimated spatial variability.

The RCMs have slightly more skill in simulating the hourly spatial precipitation field in DJF (Figure A.32) than in the JJA. In domain D3 (panel a) maximal correlation coefficients of 0.5 are reached. Especially the WRF simulations have also some performance in simulating the spatial variability.

In domain D4a and D4b (panel b and c) spatial variabilities of all simulations are generally much too low or much too high (many data points are not visible because they are outside the plot area).

Relative Humidity

All simulations are overestimating the spatial variability of relative humidity in summer 2007 in domain D3 (panel a in Figure A.33). Thereby, the 3 km grid spacing simulations have a larger spatial variability than the 10 km simulation (except in the WegCenter CCLM simulations which have the highest overestimation of variability). The spatial correlation coefficients are within a range of 0.82 (W_WRF10) and 0.86 (W_CLM3).

In domain D4a (panel b) the overestimation of the spatial variability by the CCLM simulations is even stronger than in D3 and they also have lower correlation coefficients.

Worst performance in terms of correlation is found in D4b (panel c) where the best simulation is B_CLM3 with a correlation coefficient of 0.38. Also here the spatial variability is overestimated (except in the BTU CCLM simulations). Only for the BTU CCLM simulations the 3 km grid spacing simulation shows an improvement.

Considering the spatial properties on an hourly time-scale it gets visible, that the performance in terms of correlation of the RCMs are widely spread between -0.15 to 0.88 in domain D3 (panel a in Figure A.34). The different simulations are overlapping to a great extend. Only the results of the W_CLM10 simulation are mainly outside the cluster since it has a much too high spatial variability as already seen in Figure A.33. Also the W_CLM3 simulation has many time-slices with too high spatial variability.

Analysis in sub-domains D4a and D4b show the same main characteristics and do not allow a separation between complex and hilly terrain.

Compared to JJA 2007, the RCMs have generally lower correlation coefficients in domain D3 (panel a in Figure A.35) in DJF 2007/08. However, the representation of the spatial variability is better compared to summer. The different RCM simulations with 10 km and 3 km grid spacings are clustered closely together. The BTU CCLM simulations have a very similar performance when compared to the MM5 and WRF simulations, but they underestimate the spatial variability. WegCenter CCLM simulations have with ∼-0.3 a negative spatial correlation.

Improving correlation coefficients can be seen for all 3 km grid spacing simulations on domain D4a (panel b). Especially the W_CLM3 simulation has with 0.55 a much higher correlation than the W_CLM10 with -0.11.

In domain D4b (panel c) the correlations are lower than in D3 and D4a. Differing from domain D4a, the correlations become gradually weaker in the 3 km simulations compared to the correspondent 10 km simulations.

Similar results than in the mean Taylor plots in (Figure A.35) can be found in the hourly Taylor diagrams in Figure A.36. Compared to the results in summer (Figure A.34), in D3 in winter the contours of single RCMs simulations are gradually lesser overlapping. An especially large spread of the data-points can be seen in MM5 and WRF simulations.

In domain D4a (panel b) the contour lines are overlapping to a great extend. The situation in D4b (panel c) is similar. Remarkable is the underestimated spatial variability

and the low correlation of the WegCenter CCLM simulations in this area.

Global Radiation

The mean Taylor plots for global radiation in JJA 2007 (Figure A.37) in domain D3 show a strongly overestimated spatial variability of the 10 km grid spacing simulations (with exception of B_CLM10). However, the 3 km simulations are improving the variability to a great extend (except W_CLM3), but the MM5 and WRF 3 km simulations also have lower correlation coefficients.

In domain D4a (panel b) the correlation coefficients are higher than in D3 (except W_MM510 and W_WRF10). Especially the 3 km grid spacing simulations show fair results in terms of correlation and variability.

The worst performance can be found in domain D4b (panel c) where the best simulation has a correlation coefficient of 0.37 (B_CLM3). Also the spatial variability is overestimated by most of the simulations, except the BTU CCLM simulations.

Figure A.38 shows the spatial Taylor plot for global radiation in summer 2007 on hourly basis. Results on all three considered domains look very similar. There are a lot of time-slices which have a very low and a very high (out of range of Figure A.38) spatial variability. Only a minority of data-points in domain D3 (panel a) show a correct spatial variability, but along with low correlation coefficients. A possible reason for this is an incorrect representation of clear-sky conditions relations in the RCMs compared to INCA.

RCMs are showing similar results in winter 2007/08 as in summer 2007. In domain D3 (panel a) the MM5, WRF, and BTU CCLM simulations are clustering within a correlation range of 0.5 to 0.74 and a relative spatial variability of 0.85 to 1.11. Only the WegCenter CCLM simulations are overestimating the spatial variability largely but are within the same correlation coefficient range than the other simulations.

In domain D4a (panel b) the 3 km simulations of MM5 and WRF are improving their corresponding 10 km simulations in terms of correlation. However, all 3 km simulations show a decreased spatial variability. No considerable correlation can be found in the W_CLM3 simulation.

Correlation coefficients below 0.3 are found in domain D4b (panel c). Similar to domain D3, the MM5, WRF, and BTU CCLM simulations are clustering while the WegCenterCCLM simulations have too high spatial variabilities. However, the higher grid spacing improves the variability essentially in the W_CLM3 simulation.

Figure A.40 shows the Taylor plots for hourly spatial statistics in winter 2007/08. The characteristics are similar to those in summer (Figure A.38), since most data-points can be found near the origin or out of the plot range meaning they have either too small or too large spatial variability. In domain D3 (panel a) CCLM is the only model which is able to simulate at least a few hours with the correct spatial variability.

Since global radiation is highly influenced by the formation of clouds the upper analysis

indicates misleading representation of clouds in general, regardless of the complexity of the terrain. However, conclusions about single cloud-forming processes in relation to model configuration/parameterization cannot be drawn from this analysis.

3.2.2 Fractional Skill Score

As shown in Section 3.2.1, traditional statistical methods like the correlation coefficient or the RMSE may give bad results for simulated precipitation fields, if they are considered in a spatial and temporal high resolution. This problem occurs because even simulations with only small shifts of precipitation events in time and/or space have no skill, if they are evaluated with traditional statistical methods.

To avoid this 'double penalty' problem, many alternative methods were developed, in particular in the numerical weather prediction community. In this sub-chapter, the precipitation fields of the LocMIP simulations are evaluated with the FSS method (introduced in Subsection 2.5.2), which is very popular in verifying precipitation forecasts in numerical weather prediction.

WegCenter CCLM In Figure A.41 the temporal averaged FSSs for the WegCenter CCLM simulations are shown for JJA 2007. The W_CLM10 simulation in domain D3 (left plot in panel a) shows FSSs above 0.5 on a wide range of spatial scales (in terms of the extension of the neighborhood size) and for precipitation thresholds below 2 mm/h. For threshold above 5 mm/h the simulation has almost no skill any more.

The W_CLM3 simulation (middle plot in panel a) has higher FSSs across all thresholds and spatial scales. Especially for precipitation events above 5 mm/h, the FSS of the 3 km grid spacing simulation is increased (right plot in panel a). This means, that in particular heavy precipitation events are better represented by using the 3 km grid spacing in CCLM 4.0.

In domain D4a (panel b) the improvement in the W_CLM3 is even more pronounced than in D3. Here even smaller scales and light rain event have higher FSS.

In domain D4b (panel c) the improvement is smaller than in D3 and D4a. Nevertheless, heavy precipitation events are still better represented. Since the improvement is stronger pronounced in D4a than in D4b, local precipitation based on convection seems to be more properly computed when horizontal grid resolution is increased than orographically triggered precipitation.

In winter (Figure A.42), the FSSs show similar magnitudes. The precipitation thresholds however, are much lower because there are less intensive precipitation events in winter. For the W_CLM10 simulation in domain D3 (panel a left) there are FSSs above 0.5 for a precipitation threshold of 0.5 mm/h. This is improved in W_CLM3 (middle) where FSSs are larger than 0.5 at thresholds of 1 mm/h. The 3 km grid spacing improves the FSSs below 1.5 mm/h, as shown in the difference plot to the right. Above 1.5 mm/h the FSSs are slightly deteriorated.

In domain D4a (panel b) the FSSs are generally lower than in D3. Both WegCenter CCLM simulations have no skill above 1 mm/h. The 3 km grid spacing simulation shows increasing skill in scales above 20 km and decreasing skill for low thresholds on the scale of 3 km.

Larger FSSs can be found for both simulations in domain D4b (panel c) compared to those in D3. The 3 km simulation have higher FSSs for thresholds below 1 mm/h, similar to D3. Above 1 mm/h there are lower FSSs compared to W_CLM10. The highest decrements are found in large scales and above precipitation threshold of 1 mm/h.

BTU CCLM The BTU CCLM simulations in domain D3 in summer 2007 (panel a in Figure A.43) have similar FSSs than the WegCenter CCLM simulations. The B_CLM10 simulation (left plot in panel a) has FSSs above 0.5 for precipitation thresholds of 2 mm/h. The B_CLM3 simulation has higher FSSs for all thresholds and spatial scales (right plot in panel a). For heavy precipitation events between 2 mm/h and 10 mm/h and on scales above 35 km the increase is especially high.

FSSs are slightly worse in domain D4a (panel b) than in D3. Again, the 3 km simulation improves the FSSs in a similar way than in D3.

In domain D4b the FSSs are generally higher than on the other two domains. The highest improvements between the 10 km and the 3 km simulation can be found for precipitation thresholds above 2.5 mm/h and below 10 mm/h. But also for light rainfalls (below 0.5 mm/h) notably enhancements can be found. An increased improvement in D4a compared to D4b due to increasing resolution (as it is visible for WegCenter CCLM) cannot be found in the BTU CCLM simulation.

Compared to the WegCenter CCLM simulations, the BTU CCLM simulations have slightly higher FSSs in winter 2007/08. Thereby the B_CLM3 simulation shows higher values than the W_CLM10 simulation. The score is especially improved for precipitation thresholds of 2.5 mm/h in medium and large scales.

In domain D4a (panel b) the FSSs for low rainfall thresholds are smaller than those in D3 but for thresholds above 1.5 mm/h they are notably higher. Both, the 10 km and the 3 km simulations, have the maximum FSS not at the lowest precipitation threshold (as the other simulations), but at thresholds between 1.5 mm/h and 3 mm/h. There are nearly no differences in the FSSs between the 10 km and the 3 km simulation for thresholds below 1 mm/h. Above 1 mm/h the 10 km simulation has higher scores.

High FSSs at low precipitation thresholds and nearly no skill above 2 mm/h is found in domain D4b (panel c). The 3 km simulation has higher FSS than the 10 km resolution at a threshold of 0.1 mm/h. Especially for thresholds between 1.5 mm/h and 3 mm/h the B_CLM10 simulation performs better.

MM5 The FSSs of the MM5 simulations are much smaller than those found in the CCLM simulations. The maximum FSS in domain D3 (panel a) is 0.31 for the W_MM510

simulation. For W_MM53 it is with 0.29 even smaller. As shown in Figure A.3, MM5 has also the largest dry bias of all RCMs in summer which is consistent with the small FSSs found here. The 3 km grid spacing simulation shows very similar FSSs than the 10 km simulation. Only at thresholds between 0.1 mm/h and 0.2 mm/h the 10 km simulation has slightly higher FSSs values.

In D4a (panel b) the FSSs are even smaller than in D3. The W_MM510 simulation performs similar than the W_MM53 simulation. Only for low precipitation thresholds the W_MM510 performs slightly better.

Slightly higher FSSs can be found in domain D4b (panel c). In this domain the dry bias is also smaller than those in D3 and D4a (see Figure A.3). Again the 10 km simulation has a little higher FSSs than the 3 km simulation for precipitation thresholds below 0.5 mm/h.

In winter 2007/08 the MM5 simulations show similar FSSs than the CCLM simulations. In domain D3 (panel a) they are approximately twice as large as those found in summer 2007. The FSS differences between the W_MM53 minus the W_MM510 simulations (left plot in panel a) are below 0.01.

As in summer, the FSSs are smallest in domain D4a (panel b). The FSSs are below 0.1 for all scales above a threshold of 0.5 mm/h. Differences between the 10 km and the 3 km simulation are very small.

In domain D4b (panel c) the FSSs are by ∼0.1 higher than in D3. Nevertheless, they are small compared to the FSSs of the CCLM simulations. Differences between the FSSs of the B_MM53 and the B_MM510 simulation (right column in panel c) are small in general except for strong precipitation above 3.5 mm/d.

WRF The W_WRF10 simulation (Figure A.47 panel a, left column) has slightly higher FSSs than the CCLM 10 km simulations in domain D3 in summer 2007. However, the 3 km simulation (middle column in panel a) has lower values than the CCLM 3 km simulations for thresholds above 0.2 mm/h. The differences are largest for heavy precipitation events around thresholds of 5 mm/h. This is because the 3 km simulations of CCLM have higher skills especially for heavy precipitation than their 10 km simulations while in case of WRF there is nearly no difference between the two grid spacings (right column in panel a).

The same is true for the results in domain D4a and D4b (panel b and c). The FSSs of the WRF 10 km simulations are very similar to those of the CCLM 10 km simulations but the added skill for intensive rainfall which is shown in the 3 km CCLM simulations is missing in the 3 km WRF simulation. In domain D4a the 10 km simulation has higher FSSs for thresholds below 0.5 mm/h while in domain D4b there is some slight improvement visible in the 3 km simulation for precipitation between 0.5 mm/h and 2 mm/h.

WRF has with a maximum FSS of 0.9 the best performance of all RCMs in domain D3 in winter 2007/08. It has higher FSSs than CCLM for all thresholds and scales. The W_WRF3 simulation shows only minor differences compared to W_WRF10 (right

column in panel a). There are some decreased FSSs at low and medium intensive rain events and some increased for thresholds above 2.5 mm/h.

Lower FSSs are found in domain D4a (panel b) compared to D3 (or compared to the CCLM simulations). In this region the 10 km simulation performs better for precipitation thresholds between 0.1 mm/h and 2.5 mm/h on nearly all spatial scales.

In domain D4b (panel c) the WRF simulations have, similar than in D3, the highest FSSs of all RCMs and they are even larger than those in D3. There are only minor differences between the W_WRF10 and the W_WRF3 simulation for precipitation thresholds below 2 mm/h. However, above 2.5 mm/h the 3 km simulation performs much better than the 10 km simulation.

3.2.3 Concluding Remarks: Spatial Performance

In this report the central statistics to evaluate spatial performance of the LocMIP simulations are the spatial correlation coefficient, the centered pattern RMSE, and the normalized standard deviation. This three values can be displayed in one singe diagram like described in Taylor (2001). Thereby, these statistics are calculated for the seasonally averaged spatial fields and for each time-step (hourly). While in the first method biases can cancel out, the second gives insights in the ability of the RCMs to reproduce spatial characteristics at the right time. However, RCMs are not able to simulate the spatial patterns of highly discontinuous parameters like precipitation (or partly also global radiation which strongly depends on cloud cover) at the same time and place than the reference dataset. This 'double penalty' problem can clearly be seen in Figure A.30, A.32, A.38, and A.40, where traditional metrics (like the correlation coefficient or the centered RMSE) fail to display the real skill of the simulations. Therefore, precipitation is evaluated additionally with the FSS method (Roberts and Lean 2008) (see Subsection 2.5.2) which does not depend on an exact match between simulation and observation and permits spatial offsets in the simulation.

Added value in the spatial characteristics of the 3 km simulations can be found in the correlation coefficients of 2 m temperature especially in the mountainous region of D4b. This effect can mainly be traced to the better representation of orography in the 3 km simulations and can already be seen in the bias maps in Figure A.1 and A.2. The spatial structures of 2 m temperature are also well represented on an hourly basis (see Figure A.26 and A.28).

For precipitation the correlation coefficients of the seasonal averaged field (between 0.6 and 0.7 in summer and 0.8 and 0.9 in winter) are relatively high in D3. Smaller correlations are generally found in D4b indicating, that steep orography complicates the simulation of precipitation. The normalized spatial standard deviations are generally higher in the 3 km simulations compared to the 10 km simulations.

The temporal-averaged FSSs show added value in both 3 km simulations of CCLM compared to their 10 km parent simulations, especially in JJA 2007. However, even if the

precipitation frequencies (tested with the FSS) in summer are improved, the CCLM 3 km simulations have a strong wet bias in this season. In winter 2008 the improvements in the FSSs in CCLM is more restricted to precipitation thresholds below 1 mm/h. Above 1 mm/h the FSSs are slightly reduced in the 3 km simulations (except for B_CLM3 in D3). MM5 shows generally low FSSs compared to the other RCMs, especially in summer which is consisted with the dry bias in both MM5 simulations in this season (see Figure A.3). There are only minor differences between the performance of the 10 km and the 3 km simulation. The highest FSSs of all 10 km simulations can be found in W_WRF10. However, there is no improvement in the W_WRF3 simulation (with exception of precipitation values above 2 mm/h in D4b inDJF). The W_WRF3 simulation performs approximately as good as the W_CLM3 and the B_CLM3 simulations.

Relative humidity has an enlarged normalized spatial standard deviation in summer, in particular in the 3 km simulations. In winter correlations are worse than in summer but the spatial standard deviation is improved.

The temporal averaged spatial properties of global radiation are better represented in D4a than in D4b, indicating more problems in cloud forming processes in mountainous areas than in flat and hilly regions. Especially in winter the correlation coefficients in D4b are below 0.3. On an hourly basis, the simulations show worse results in terms of correlations and standard deviations, which probably means, that the RCMs are not able to produce a cloud cover similar to INCA.

3.3 Temporal Characteristics

In the former sections the mean climate (Section 3.1) and the spatial characteristics (Section 3.2) of the LocMIP simulations were analyzed. In this section the focus lies on the temporal performance of the RCMs. Therefore, the mean diurnal cycle of each simulation is analyzed and compared to the reference dataset. This method can give valuable insights in deficiencies in different physical processes within the RCMs.

As a second evaluation method two different types of Taylor diagrams are used. The first shows the hourly data of temporal correlation, the centered RMSE, and the normalized temporal standard deviation for the spatially averaged fields. The second displays the same properties, but for every grid-point as density contour plot which gives insights in the spatial distribution of the above mentioned statistics.

3.3.1 Diurnal Cycles

2 m Temperature

In JJA 2007 a pronounced diurnal cycle is shown for 2 m temperature in the reference and all RCM simulations with a maximum in the afternoon and a minimum before sunrise (Figure A.49). In domain D3 (panel a) the ensemble error range is roughly constant with ±1.5 K. Thereby the simulations scatter around the reference, meaning that the ensemble mean is nearly bias free. The temperature maximum between 13 UTC and 14 UTC is most misplaced within the WegCenter CCLM simulations. Their maxima are at 12 UTC and therefore two hours too early. A possible reason for this shift is the underestimation of global radiation (probably caused by an overestimation of cloud cover) in the W_CLM simulations. In the MM5 simulations, the diurnal cycle is amplified, meaning that the minimum temperature is too cold and the maximum temperature is too warm along with a delayed maximum. WRF captures the temperatures between 20 UTC and 4 UTC very well but has a warm bias during 4 UTC and 17 UTC. In the B_CLM10 simulation the day-night amplitude is too small. The B_CLM3 diurnal cycle is shifted by ∼0.5 K, captures the temperature between 6 UTC and 19 UTC very well, but has a larger warm bias than the B_CLM10 at night.

The diurnal cycle is more pronounced in domain D4a but the main characteristics are similar than those in D3. From 7 UTC to 17 UTC a majority of the RCMs are overestimating the temperature. Differences to D3 occur also in the WegCenter CCLM simulations, which have no cold bias during night any more but a warm bias between 7 UTC and 15 UTC.

In domain D4b (panel c) the amplitude of the diurnal cycle is smaller than on the other domains. The WegCenter CCLM simulations have a cold bias during the whole day. The BTU CCLM simulations are capturing the diurnal cycle within ∼1 K, while the WRF simulations are too warm during daytime and the MM5 simulations are too cold during nighttime.

The ensemble spread in 2 m temperature in winter 2007/08 (Figure A.50) is larger than in summer 2007 (Figure A.49). The WegCenter CCLM simulations have a ∼3 K cold bias during the whole day in domain D3 (panel a). Similar to summer, the temperature maxima for these simulations are one hour too early. The BTU CCLM simulations are too cold during daytime. The MM5 and the WRF simulations are too warm during 20 UTC and 11 UTC.

Similar results found in D3 can also be seen in D4a (panel b). The cold bias of WegCenter CCLM is smaller but the maxima temperature occurs still too early. Also the maxima of the WRF and MM5 simulations are misplaced by -1 hour. After the maxima of these simulations the temperatures decrease too fast which leads to a cold bias between 14 UTC and 17 UTC in WRF. The diurnal cycle of the BTU CCLM simulations are too weak and so they have a warm bias at night and a cold bias around midday.

The cold bias of WegCenter CCLM is especially strong in domain D4b (panel c) and strongest pronounced in the W_CLM3 simulation. The diurnal cycle of BTU CCLM simulations are too weak, which results in a cold bias especially during daytime. The B_CLM3 simulation improves this bias by ∼1 K. The WRF and MM5 simulations are too warm at night and during the first half of the day. In these simulations there is a strong temperature decrease between 14 UTC and 16 UTC. After 16 UTC the temperature stays nearly constant until sunrise.

Precipitation

In summer 2007 INCA shows a clear precipitation maximum in domain D3 (panel a in Figure A.51) between 15 UTC and 17 UTC which is caused by convective precipitation. Four simulations are able to capture this maximum at the right time (W_WRF10, W_WRF3, W_CLM3, and B_CLM3). The best performance have the WRF simulations which resembles the reference data fairly during the whole day. In the W_CLM10 simulation the diurnal cycle is well captured during 20 UTC till 13 UTC but the afternoon maximum is too weak. The W_CLM3 simulation represents the maximum precipitation better but overestimates the rainfall amount especially during nighttime. The 3 km grid spacing simulation improves the location of the maximum significantly in the BTU CCLM simulations. However, the B_CLM3 simulation overestimates the precipitation in general but especially the maximum. No differences can be seen between the two MM5 simulations. In MM5 the precipitation maximum occurs one to two hours too late, but the magnitude is well captured. During the rest of the day there is too less precipitation.

Domain D4a (panel b) is much drier than D3 and there is no clear afternoon precipitation maximum. Both 3 km CCLM simulations show an improvement in the diurnal cycle especially in the afternoon and in the first half of the night.

The most pronounced afternoon precipitation maximum can be found in D4b (panel c) where the orography facilitate the development of convective precipitation. Beside the maximum in the afternoon, there is a second precipitation maximum in the morning

at 5 UTC. Most simulations are able to simulate the precipitation maximum in the afternoon. However, the WRF simulations as well as the 3 km CCLM simulations greatly overestimate the magnitude of the maximum. Furthermore, the maxima are shifted by minus one to minus two hours in both 3 km CCLM simulations. In the B_CLM10 simulation, the maximum occurs at 15 UTC and is therefore shifted by -2 hours. The best performance in D4b have the MM5 simulations. Also the W_CLM10 simulation shows a remarkable good fit compared to the other simulations and also captures the secondary rainfall maximum in the morning.

In winter the mean diurnal precipitation cycle in domain D3 has no special characteristic because this season is dominated by frontal precipitation events (see panel a in Figure A.52). All simulations are overestimating the precipitation except of W_CLM10 and W_CLM3 between 10 UTC and 15 UTC. There are no differences between the 10 km and the 3 km grid spacing simulations in WRF and MM5. For CCLM the 3 km simulations slightly decrease the wet bias.

In domain D4a (panel b) there is generally less precipitation in INCA, which the RCMs are able to capture nicely.

A general overestimation of precipitation can be seen in domain D4b (panel c). The best performance has the W_CLM3 simulation which improves the W_CLM10 remarkably. A particularly high wet bias of approximately 50 % have the BTU CCLM simulations. WRF and MM5 simulations overestimate the precipitation especially during the first half of the day and show slightly better results in the second one.

Relative Humidity

Relative humidity is directly related to air temperature because warm air can hold more water vapor than cold air, like it is empirically formulated in the Clausius–Clapeyron equation. Therefore, the shape of the diurnal cycle of relative humidity is inverse to the shape of the diurnal cycle of air temperature.

All RCMs except MM5 simulate the minimum relative humidity too early. However, the MM5 simulations are too dry, especially during daytime. WRF has a similar dry bias but the shape of the diurnal cycle is worse. WegCenter CCLM simulations have a relatively good performance between 19 UTC to 12 UTC but are too humid in the afternoon. In both BTU CCLM simulations the diurnal cycle is too weak. The 10 km simulation shows a better performance during nighttime while the 3 km simulation has lower biases during daytime.

The diurnal cycle is more amplified in domain D4a (panel b) than in D3. All RCMs have a dry bias during the night and most of them (except B_CLM10) also during the day. This might be related to a very dry soil in this region, because nearly all simulations (except B_CLM3) simulate also too less precipitation.

In domain D4b (panel c) WRF and MM5 simulations are too dry while CCLM simulations are too humid. In the BTU CCLM simulations the 3 km grid spacing improves the positive

humidity bias remarkably.

Also in winter, relative humidity has a well pronounced diurnal cycle which is inverse to the diurnal cycle of air temperature. In the simulations in domain D3 (panel a) this characteristic diurnal cycle is only partially represented since it is too weak and also misplaced. WegCenter CCLM simulations are generally too humid while the BTU CCLM simulations have a good fit during the night but no reduction of relative humidity during the day. MM5 simulates too high relative humidity values during the day but even higher during night while WRF has generally the largest dry bias.

The diurnal cycles of the MM5 and WRF simulations in D4a (panel b) are more pronounced but the minima occur too early. In the CCLM simulations the diurnal cycle is too weak and instead of increasing relative humidity values during the night the values of the simulations stay constant between 18 UTC and 6 UTC.

The worst performance can be found in domain D4b. Not one RCM is able to simulate a realistic diurnal relative humidity cycle. MM5 simulates even a maximum during the day and has a dry bias during the night. WRF performs similar but the daytime maximum is less pronounced than in MM5. The BTU CCLM simulations are able to reproduce the relative humidity during the night but have also a maximum during the day. The WegCenter CCLM simulations are much too humid and have no diurnal cycle at all.

Global Radiation

Global radiation is generally increased in all 3 km grid spacing simulations compared to the corresponding 10 km simulations. In domain D3 (panel a) WegCenter CCLM has a large underestimation of global radiation which is most probably caused by too much cloud cover. Also the BTU CCLM simulations have a negative global radiation. The B_CLM3 simulation is shifted by plus one hour compared to the B_CLM10 simulation. The best performance has the W_WRF10 simulation while the W_WRF3 simulation overestimates the maximum by $\sim$150 W/m^2. The maxima in both MM5 simulations are misplaced by one hour and are overestimated by 50 W/m^2 in W_MM510 and by 140 W/m^2 in W_MM53.

The situations in domain D4a and D4b (panel b and c) are very similar to those in D3. However, in D4b the negative biases of the WegCenter CCLM simulations are smaller and in D4b they are larger.

In winter the diurnal cycle of the global radiation in domain D3 is generally better represented than in summer (Figure A.56). The maximum of the B_CLM10 simulation is shifted by -1 hour while the B_CLM3 simulation has no shift. MM5 has a very good fit while WegCenter CCLM and WRF are overestimating the global radiation maximum.

In domain D4a (panel b) the WegCenter CCLM simulations and W_MM510 have nearly a perfect fit, however the W_MM53 overestimates the maximum slightly. The BTU CCLM simulations are very similar and overestimate the global radiation as well. Again, the B_CLM10 is shifted by -1 hour. WRF simulations have the highest overestimation

of the global radiation maximum.

WegCenter CCLM simulations overestimate the global radiation in D4b (panel c) by ∼100 W/m^2. The other simulations perform well and stay within a range of 50 W/m^2.

3.3.2 Temporal Taylor Diagrams

2 m Temperature

In the spatial mean temporal Taylor diagrams, the 10 km grid spacing simulations and the corresponding 3 km simulations show generally very similar results (Figure A.57). In domain D3 (panel a) the WRF and CCLM simulations are clustering at a correlation coefficient of ∼0.95 and a relative temporal variability between 0.85 (W_CLM10) and 1.03 (W_WRF3). Only the MM5 simulations have a lower correlation coefficient of 0.83 and a slightly overestimated variability.

In domain D4a (panel b) CCLM and WRF simulations are still clustering but have a little bit lower correlation coefficients. The MM5 simulations are again outside the cluster with a correlation of 0.82. Very similar results can also be found in domain D4b (panel c).

The grid-point wise temporal Taylor diagrams for 2 m temperature in summer 2007 are shown in Figure A.58. The main characteristics are similar to those in Figure A.57 on all domains. MM5 has always worse correlation coefficients, while the contours of the other RCMs are overlapping to a great extend. There are only minor differences between the 10 km grid spacing simulations and the 3 km simulations. Only in domain D4a (panel b) the 3 km simulations of CCLM have both an improved temporal variability.

All RCMs show very similar results in winter 2007/08 (Figure A.59). The correlation coefficients in domain D3 (panel a) are between 0.9 (W_MM510 and W_MM53) and 0.95 (W_CLM3). The simulations slightly underestimate the temporal mean variability between -20 % (B_CLM3) and -3 % (W_WRF3).

Temporal variability is a little bit more underestimated in domain D4a (panel b) than in D3. Especially the CCLM simulations have low values which is in good agreement with their weak diurnal cycle in winter (Figure A.50).

In domain D4b (panel c) the temporal variability is well captured by all simulations. The correlation coefficients are slightly decreased in the MM5 and increased in the BTU CCLM simulations compared to D3.

The RCM simulations are clustering also when temporal values are considered on grid-points basis. In domain D3 (panel a) a bipolar structure with two maxima can be seen for the contours of the CCLM simulations. The reason for this is that the temporal variability is underestimated in the hilly domain D4a (panel b) while it is well captured in the mountainous region of domain D4b (panel c). The area of the contour of W_CLM3 is very large in domain D4b (panel c).

Precipitation

The temporal correlation coefficients of summertime precipitation in domain D3 (panel a in Figure A.61) are between 0.13 (MM5 simulations) and 0.78 (W_CLM3). All simulations except W_CLM10 overestimate the temporal variability. The temporal variability is increased especially in the CCLM 3 km grid spacing simulations compared to their corresponding 10 km simulations.

On both sub-domains the correlation coefficients are smaller than in D3 because in D3 a much larger area is averaged and so large displacements (larger than the sub-domain sizes) are not penalized in the temporal evaluation. The lowest correlations can be found in D4a (panel b). The MM5 simulations have zero correlation while W_CLM3 has the highest with 0.42. Like in domain D3 especially the CCLM 3 km simulations have a higher temporal variability compared to their correspondent 10 km simulations. In domain D4b correlation coefficients are again a little bit higher than in D4a. MM5 simulations have correlation coefficients below 0.1 while the B_CLM3 simulation has the best correlation with 0.62. All simulations except the 10 km simulations of CCLM have too high temporal variabilities.

The performance of the RCM simulations is much worse by considering grit-point values because there has to be precipitation not only at the right time, but also exactly at the right grid box (Figure A.62).

In domain D3 (panel a) the contours of WegCenter CCLM, B_CLM10, and WRF are overlapping to a great extend. The MM5 simulations have nearly no correlation and the B_CLM3 has much too high temporal variabilities and the majority of its data points lies outside the plot range.

The situation in domain D4a (panel b) is similar than in D3. Clearly displayed is an increase of temporal variability in the 3 km simulations. This increase is even more pronounced in D4b (panel c) where the main parts of the contours of the 3 km simulations are outside the plot range.

In winter the RCMs are overestimating the mean temporal variability of precipitation in domain D3 (panel a in Figure A.63). The highest overestimation of variability and also the highest correlation coefficients can be found in the BTU CCLM simulations. MM5 has the worst correlation coefficient and also a very high temporal variability.

In domain D4a (panel b) only the BTU CCLM simulations overestimate the variability. All simulations have lower correlation coefficients than in D3. Especially the MM5 simulations show nearly no correlation at all.

Correlations of WRF, WegCenter CCLM, and especially those of MM5 are higher in D4b (panel c) compared to D4a. The general overestimation of the temporal variability is similar then in D3.

On the grid-point basis the correlation coefficients are generally higher in winter (Figure A.64) as they are in summer (Figure A.62). In winter in domain D3 (panel a) the best performance in terms of correlation has the W_WRF10 simulation while the MM5

simulations have the worst. All contours are stretched to the origin and have their center of mass at too low variabilities. However, there are also points outside the plot range which are not shown. Especially the B_CLM10 simulation has much too high temporal variabilities.

In domain D4a (panel b) the RCMs generally underestimate the variability. The best simulations in terms of correlation are the BTU CCLM simulations.

In domain D4b (panel c) the areas covered by contours are large and many data-points are outside the plot ranges because of their high temporal variability.

Relative Humidity

The correlation coefficients for relative humidity in summer 2007 in domain D3 (panel a in Figure A.65) are between 0.88 (B_CLM3) and 0.78 (MM5 and WegCenter CCLM simulations). The relative variability is between 0.68 in the BTU CCLM simulations and 1.18 in the W_MM53 simulation. The overall best performance have the WRF simulations with a high correlation coefficient and a nearly perfect temporal variability.

Results in domain D4a (panel b) are similar to those in D3 but the correlation coefficients are slightly lower. The same is true for domain D4b (panel c) where the correlation coefficients are lowest compared to the other two domains.

The contour Taylor plots for relative humidity in summer 2007 (Figure A.66) are similar compared to the mean Taylor plots (Figure A.65). The best performance in domain D3 (panel a) have also the WRF simulations with a high correlation and a nicely centered temporal variability. The WegCenter CCLM has many data points with too low variability.

In domain D4a (panel b) the extend of the contours is much smaller than in D3 for all simulations. This is different in D4b (panel c) where the contours are wide spread and overlapping (except WegCenter CCLM).

The RCMs show similar good results in the mean Taylor diagrams of relative humidity in winter (Figure A.67) than they show in summer (Figure A.65). WRF is the overall best performing model with correlation coefficients of ∼0.8 and a good representation of temporal variability. In D4a (panel b) the MM5 simulations show a similar good performance but have too high variabilities. WegCenter CCLM simulations have a much too low variability especially in D4b (panel c). The BTU CCLM simulations perform good in domain D4b but bad in D4a.

The maxima in the contours in Figure A.68 are in similar locations than the data points in the mean Taylor plots (Figure A.67) for relative humidity in winter 2007/08. A bipolar structure can be seen in the BTU CCLM simulations in domain D3 (panel a) because in the hilly region of D4a (panel b) the simulations underestimate the temporal variability by ∼40 % while in the mountainous region D4b (panel c) it has a higher variability and also a higher correlation.

Global Radiation

The correlation coefficients of global radiation in summer 2007 in domain D3 (panel a in Figure A.69) are within a range of 0.97 (W_WRF3) and 0.86 (W_MM510 and W_MM53). The values of the relative temporal variability are in good agreement with the overestimation or underestimation of the maximum in the mean diurnal cycle of global radiation (see Figure A.55). CCLM simulations generally underestimate the variability (the WegCenter simulations stronger than the BTU simulations) and WRF and MM5 are overestimating the variability. Thereby, the 3 km grid spacing simulations always have higher temporal variabilities than the 10 km simulations.

The results in domain D4a and D4b (panel b and c) are similar to those in D3. The WegCenter CCLM simulations however, have a better representation of temporal variability in D4a than in D4b which can be a result of too much cloud cover especially over mountainous regions.

The results in Figure A.70 are similar to those in Figure A.69. A shift to higher variabilities can be seen in nearly all 3 km grid spacing simulations compared to the corresponding 10 km simulations. A slight improvement in terms of correlation can in particular be found in the CCLM 3 km simulations compared to their correspondent 10 km simulations.

In winter the RCMs have an especially good performance in domain D3 (panel a) with correlations between 0.95 (W_MM510) and 0.99 (W_WRF10 and W_CLM3). Thereby, also the variability is well captured. In domain D4a (panel b) the correlation coefficients are slightly lower than in D3 and the variability in the WRF simulations is ∼20 % too large. The correlation in D4b (panel c) is better than in D4a but here the temporal variability of the WegCenter CCLM simulations is ∼30 % too high.

In domain D3 (panel a) in Figure A.72) the contours of the different simulations are overlapping for global radiation in winter 2007/08. There are bipolar structures displayed especially in the WegCenter CCLM but also in the WRF simulations. This is because in the hilly region of domain D4a (panel b) the WegCenter CCLM runs have a lower temporal variability and correlation than in the mountainous region D4b (panel c). For WRF the correlation is also higher in D4b than in D4a but the variability is lower in the mountains.

3.3.3 Concluding Remarks: Temporal Performance

The diurnal cycle of 2 m temperature is well covered by all simulations in summer whereby the biases are scattering around the observation within a range of $\leq \pm 1.5$ K. The maximum temperatures are displaced especially in the WegCenter CCLM simulations. In these simulations the maxima occur at 12 UTC which is consisted with the underestimation of global radiation (probably caused by too much cloud cover).

In winter the biases are generally larger. Especially the two WegCenter CCLM simulations have a severe cold bias like already seen in the mean climate evaluations in

Section 3.1.

No general added value can be found in the temporal Taylor plots of 2 m temperature between the 3 km and the corresponding 10 km simulations. The correlations are typically high (above ∼0.9, except MM5) in summer and winter.

The diurnal cycle of precipitation in D3 has a distinct maximum in the afternoon (caused by convective precipitation events). This maximum is captured by W_WRF10, W_WRF3, W_CLM3, and B_CLM3. In D4a no precipitation maximum can be found either in the simulations nor in the reference dataset. In D4b however, the afternoon precipitation maximum is even more amplified, caused by the strong orographic forcing in the Alps. All 3 km simulations (except the W_WRF3) overestimate this maximum to a great extend. Especially, in the 3 km CCLM simulations, there is much more intense precipitation in the afternoon compared to their 10 km parent simulations, indicating that convective events are either too strong or convection is triggered too fast. From the conditional-quantile plot in Figure A.19 it can bee seen that the first is definitively true and the second is likely because of the overestimated mean temporal standard deviation (see Figure A.61). Furthermore, the overestimation of the afternoon maximum is the major reason for the inner Alpine precipitation bias in the 3 km CCLM simulations (see Figure A.3).

Wintertime precipitation is overestimated in the Alps but not in the hilly region of D4a. Unlike in summer, the 3 km simulations does not increase the precipitation amounts compared to the 10 km simulations. This means that the improved orography in the 3 km simulations do not lead to an increase in orographic triggered precipitation.

The temporal correlations of relative humidity are between 0.78 and 0.88 and therefore well captured by all RCMs. However, the diurnal cycle of relative humidity is much too weak in winter (especially in D4a). This problem can also bee seen in the conditional-quantile plots in Figure A.20.

In summer all 3 km simulations have higher maximum values of global radiation compared to their parent simulations. Consistent results are found for normalized standard deviation, where all 3 km simulations have larger values than their corresponding 10 km simulations.

In winter, the diurnal cycles, the temporal correlation coefficients, and the standard deviations of global radiation are improved compared to summer. This is most probably caused by the predominance of stratiform clouds in winter, which are easier to simulate by RCMs than convective clouds which often occur in summer.

3.4 Overall Performance

In this section a comprehensive overview of the performance of the LocMIP simulations is given by considering seven statistical values from Section 3.1, Section 3.2, and Section 3.3 in one plot by showing the performance in terms of the relative performance index (RPI) (see Subsection 2.5.4).

The RPI values in domain D3 are shown as colors in Figure A.73. Generally, every simulation has its strengths and weaknesses. However, looking at the mean performance (right outermost column) differences in the model skill can be seen.

In summer (panel a), the best performing simulation is B_CLM10. It has a very good representation of the temporal correlation, the mean spatial variability, and the biases in all considered parameters. Only the mean temporal variability for 2 m temperature and relative humidity are worse represented compared to the other simulations. The second best mean RPIs have the WRF simulations with an excellent performance in terms of temporal correlation. The lowest mean RPI can be found in the W_CLM10 simulation. This simulation has especially high RMSEs (except for precipitation), and a general overestimation of the mean spatial variability. Nevertheless, it has the best performance in the precipitation RMSE and the temporal variability of 2 m temperature.

One clear added value of the 3 km grid spacing simulations can be seen in the RMSE values and the spatial correlations of 2 m temperature. Furthermore, the FSS of the 3 km CCLM simulations are higher than those of the corresponding 10 km simulations. Multiple improvements in the finer gridded simulations can be found in various statistical parameters but only for single simulations. Remarkable, is the improvement in the mean RPI of 9 % between the W_CLM10 simulation and the W_CLM3 simulation.

In winter (panel b) the best simulation is W_WRF10 with its excellent performance in temporal correlation and temporal variability. It also has the highest FSS in the whole ensemble. Only a marginal smaller mean RPI can be found in the B_CLM3 simulation which is especially good in representing the biases and RMSEs of 2 m temperature, relative humidity, and global radiation. The W_CLM10 has again the smallest mean RPI especially because of its high biases and RMSEs in 2 m temperature, relative humidity, and global radiation.

The best performing simulation in terms of mean RPI in summer in domain D4a (panel a in Figure A.74) is the W_WRF3. Especially the temporal and spatial correlation and the spatial variability is over average. The W_WRF10 has only slightly smaller RPI values. The lowest mean RPI can be found (as in domain D3) for the W_CLM10 simulation. However, the corresponding 3 km simulation (W_CLM3) shows a clear improvement and has, for example, the highest mean FSS in the whole ensemble.

In winter (panel b) the best performance can be found in the W_MM53 simulation. It has an over average performance especially in terms of bias, RMSE, spatial correlation, and temporal variability. However, it has the lowest mean FSS which might be caused by its bad temporal variability of rainfall. The lowest mean RPI can be found in W_CLM3

which has over average RPIs for nearly all statistics and parameters.

In the mountainous domain D4b WRF and BTU CCLM simulations have the best performance during summer (panel a in Figure A.75). Especially the 3 km simulations of WRF has high RPI values. MM5 simulations have a good representation of the biases and the RMSEs but show low RPIs for the FSSs and the temporal correlations. The worst mean performance has the W_CLM10.

In winter (panel b) the WRF simulations show the best performance even thou they have a under average performance in precipitation (bias, RMSE, and spatial correlation). Again, the lowest RPI values can be found in the WegCenter CCLM simulations.

4 Conclusion

The regional climate models (RCMs) which contributed to the Local Climate Model Intercomparison Project (LocMIP) ensemble are able to simulate regional and local atmospheric fields with some notable accuracy. The models show different performance for different variables and for different types of error characteristics (mean climate, spatial and temporal characteristics). Thereby, the simulations have generally more skill in simulating mean climate characteristics because random errors (e.g., small displacements of convective precipitation) are canceling out by averaging and only systematically biases are evaluated. If spatial and/or temporal highly resolved data are considered, systematic and random errors adding up and the performance of model outputs is typically worse especially for temporal and spatial highly fluctuating parameters like precipitation. Furthermore, also reference datasets (like Integrated Nowcasting through Comprehensive Analysis (INCA)) often have relatively large errors when grid point values are considered on short timescales (e.g., Haiden et al. (2011)).

The performance of the cloud resolving 3 km grid spacing simulations is in general strongly determined by the performance of their 10 km parent simulations. Added value of the high resolved simulations can especially be found in the 2 m temperature fields, where the ensemble mean spatial root mean squared error (RMSE) is decreasing by ∼0.5 K in June, July, and August (JJA) 2007 in the entire eastern Alpine Region (D3). Furthermore, a general improvement of the spatial correlation of 2 m temperature occurs. Both effects are mainly caused by the better resolved orography in the 3 km simulations.

The convection resolving simulations are able to produce more intensive rainfall and higher spatial variability than the 10 km simulations, in which convective precipitation is parameterized. While traditional statistical methods (e.g., correlation, bias, RMSE) do not indicate added value in precipitation evaluations, the Fractional Skill Score (FSS) analysis depict a notable improvement of the rainfall fields in summer in both model version 3 km COSMO model in CLimate Mode (CCLM) simulations. The improvements are especially high for intense rainfall above a threshold of 2 mm/h. However, in these simulations the total precipitation amount is overestimated especially in the inner Alpine region which indicates a too strong forcing of orographically induced convection. All simulations are too wet in the Hohe Tauern National Park (D4b) (except those of the Fifth-Generation Mesoscale Model (MM5) in summer). In JJA especially the 3 km CCLM, but also the Weather Research and Forecasting Model (WRF) simulations produce too much precipitation which is mainly caused by an overestimation of the afternoon rainfall maximum by ∼100 %.

Global radiation is generally increased in all 3 km simulations compared to the corresponding 10 km simulations during summer referring to a reduced cloud cover in the fine gridded simulations. The increased global radiation improves the CCLM results but deteriorated those of the WRF and MM5 simulations in terms of bias and temporal variability. Since also precipitation intensity is increased in the 3 km simulations, it is likely that there are more small scale convective clouds when deep convection is resolved explicitly.

Concerning the mean relative performance index (RPI), differences between the RCMs are generally larger than differences between the two resolutions. The best performing RCMs are WRF and CCLM version 4.8 used by Brandenburg University of Technology Cottbus (BTU). The MM5 simulations have generally a bad performance of temporal correlation and therefore a decreased mean RPI value. The simulation with the lowest mean RPI is the 10 km CCLM version 4.0 run. However, the corresponding 3 km simulation is notably improving the RPI of the W_CLM3 on domain D3.

The differences between the two model versions of CCLM are mainly caused by two reasons. First, because of combined effects of numerous modifications (e.g., introduction of a reference atmosphere and a sub-grid scale orography scheme) and corrections (multiple bug fixes) in CCLM version 4.8 compared to version 4.0 (Keuler et al. 2010). Second, because of differences in the model set-up, whereby the most striking is probably the choice of Leapfrog numerics in the B_CLM10 simulation and Runge-Kutta 3rd order numerics in W_CLM10. The newer version outperforms the older in nearly all considered statistics (e.g., CCLM 4.8 has a much weaker cold bias in winter, a better representation of global radiation in summer, and a more accurate representation of relative humidity). The typical zero degree peak can, however, still be found in the CCLM 4.8 simulations even if it is distinctly improved compared to CCLM 4.0.

A similar improvement can be seen between the MM5 and WRF simulations (the development of MM5 stopped in 2004 while WRF is steadily further developed) whereby WRF especially outperforms MM5 through a better performance in temporal correlation and variability and in the representation of precipitation. The common history of the two models can still be seen in the error characteristics of their simulations which are more similar to each other (e.g., in terms of spatial error patterns) than to the CCLM simulations.

A major problem in evaluating high resolved climate simulations is the availability and accuracy of reference datasets. Also, the here used INCA dataset can have large errors especially in regions with a low density of meteorological stations and if short temporal and grid point values are considered. Further limitations in this evaluation are the rather short time periods and the small domain sizes. A simulation of climatic time horizons and larger regions would have the advantage of a better statistical evaluability and a higher accuracy for the evaluation of model performance. Also the evaluation of extreme events would become possible. Extremes are very promising to improve in convection-resolving climate simulation (CRCS) which also can be seen from the here derived results (e.g., precipitation intensities).

Nevertheless, the findings of the LocMIP are giving important information for future CRCS studies. They show the typical behaviour and error ranges of CRCS compared to coarser gridded simulations and highlight important topics for future model developments. Furthermore, the used evaluation methods are presenting a guideline for evaluating the performance of very high resolved RCM simulations holistically.

A Figures of the Results

In this appendix the figures of Chapter 3 are displayed. To reduce the memory size of this report the here depicted images are of low quality. A separated document, containing high quality images for printing, can be downloaded from http://www.uni-graz.at/igam3www/WegCenterVerlag/WCV-SciRep-No41-Appendix-APreinEtAl-nhcm1-iii-Feb2011.pdf.

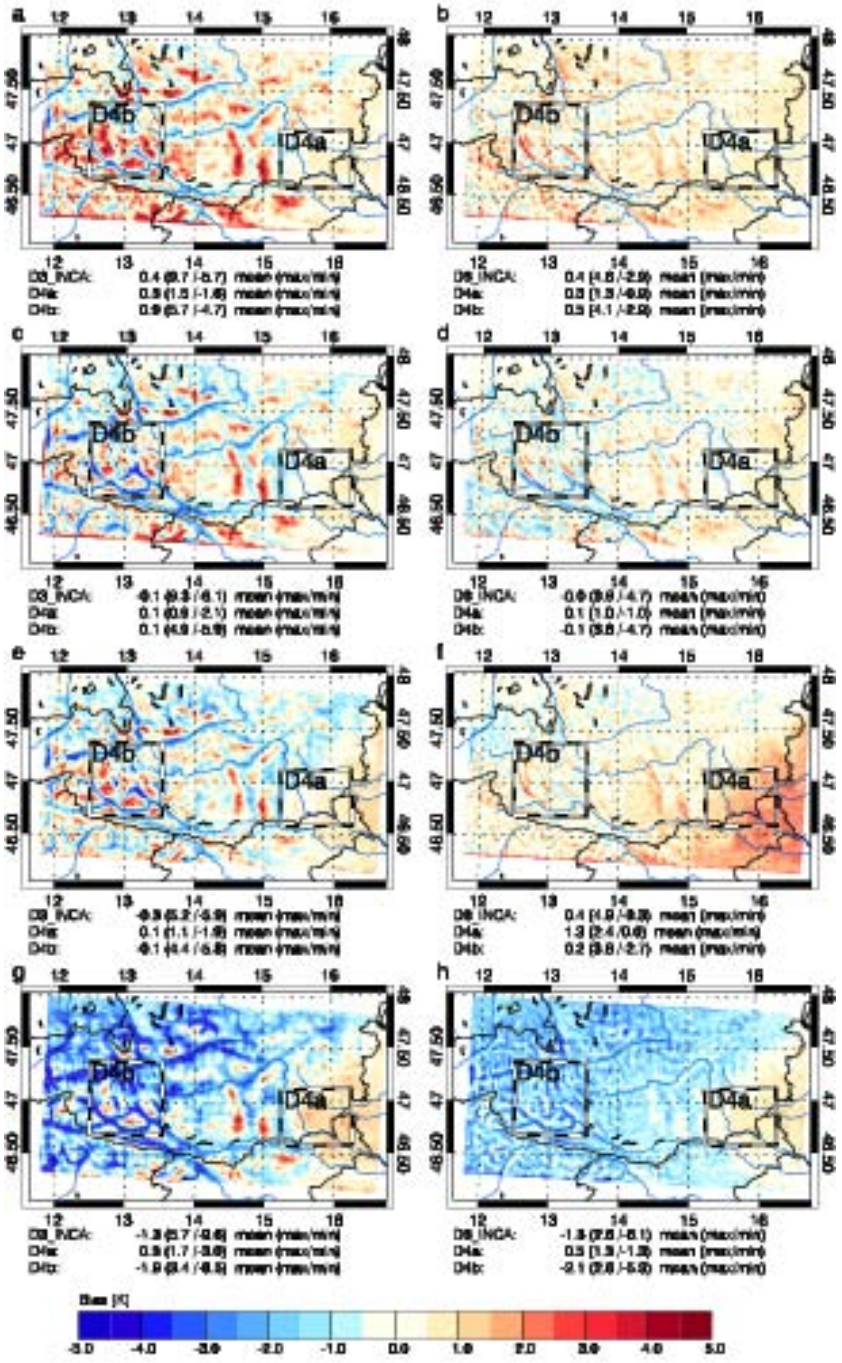


Figure A.1: Bias maps of JJA 2007 for 2 m air temperature [K]. Underneath each map the mean, maximum (max) and minimum (min) biases are shown for the entire domain D3 and the focus regions D4a and D4a. The panels show the following simulations: a) W_WRF10, b) W_WRF3, c) W_MM510, d) W_MM53, e) B_CLM10, f) B_CLM3, g) W_CLM10, h) W_CLM3.

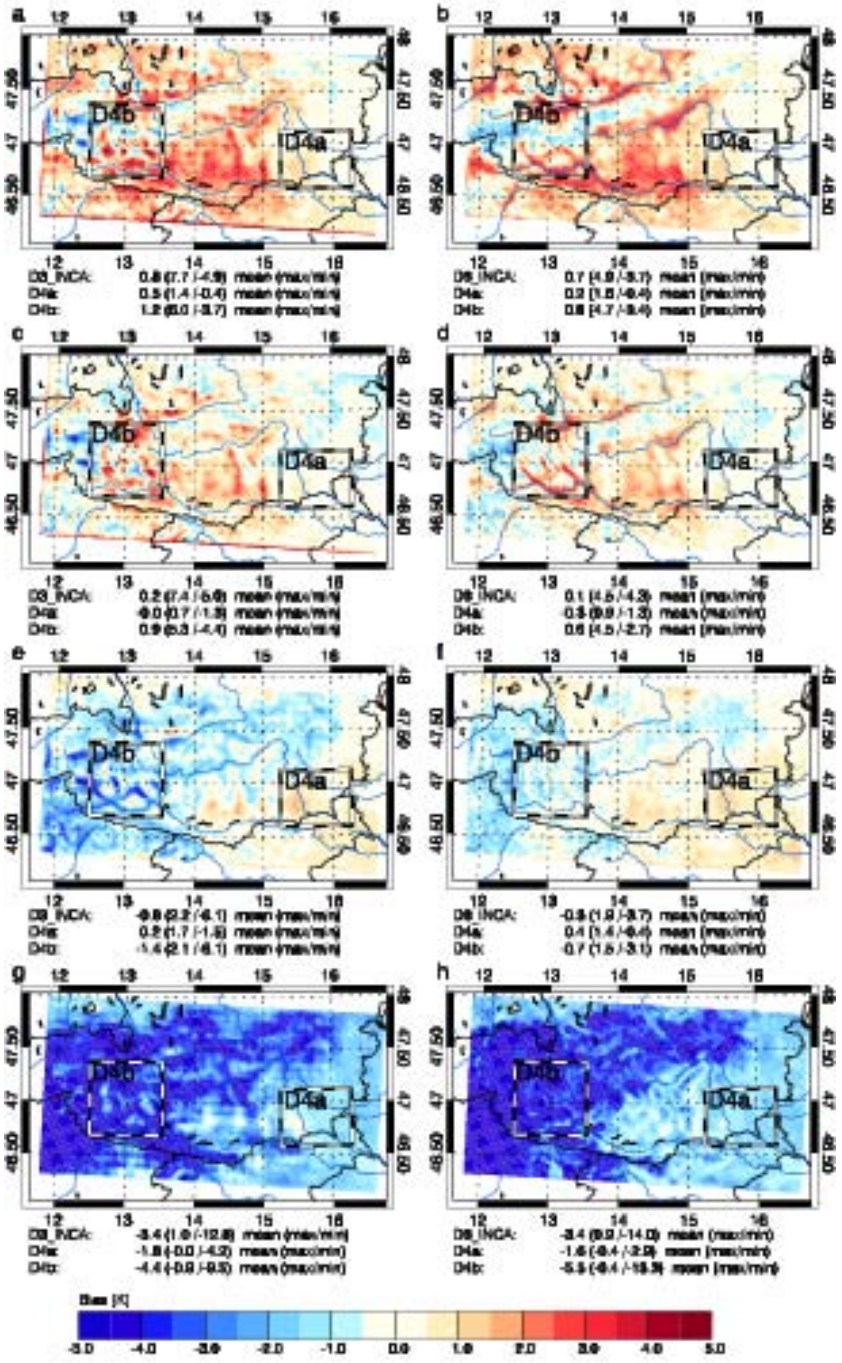


Figure A.2: Same as Figure A.1 but for winter 2007/08.

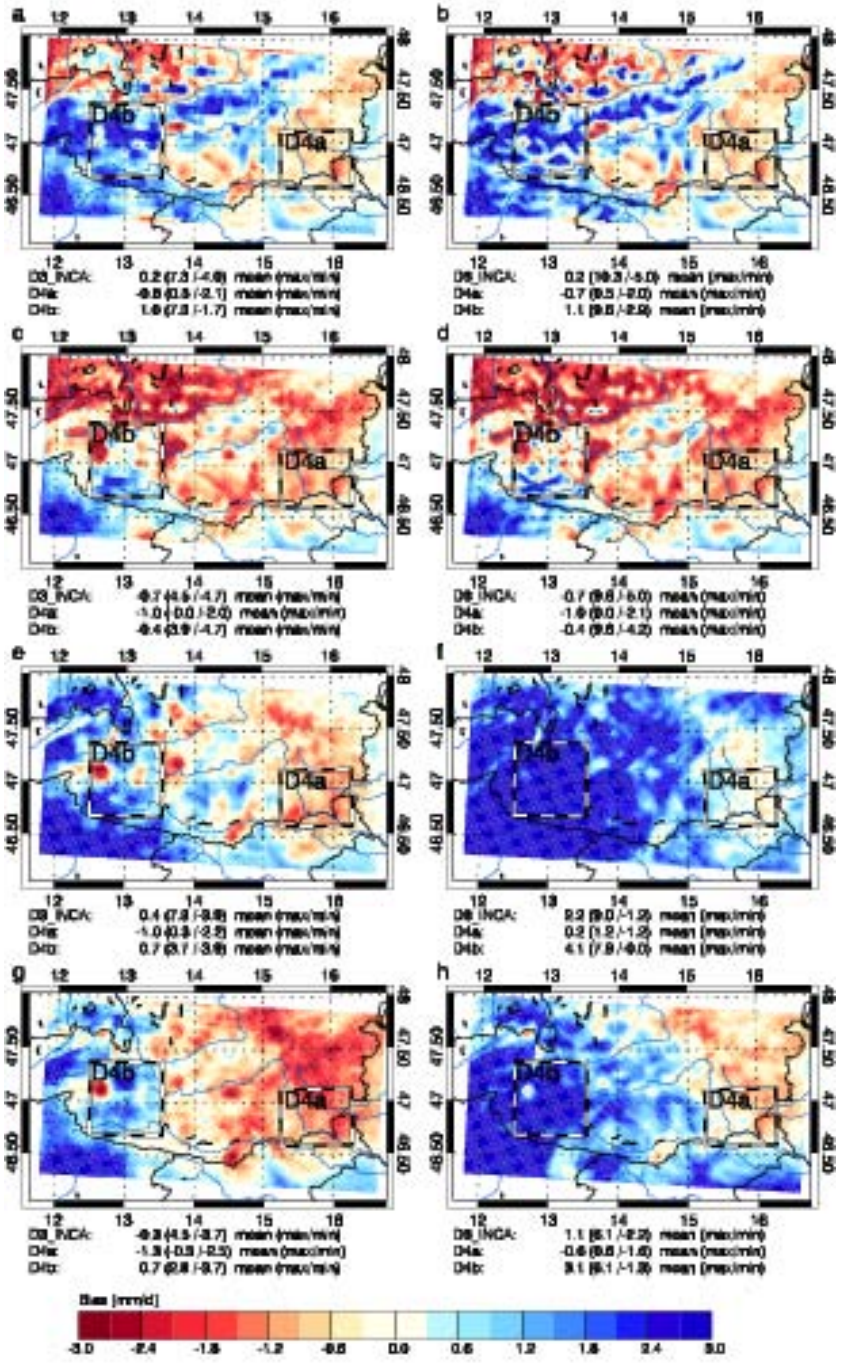


Figure A.3: Same as Figure A.1 but for precipitation [mm/d] in summer 2007.

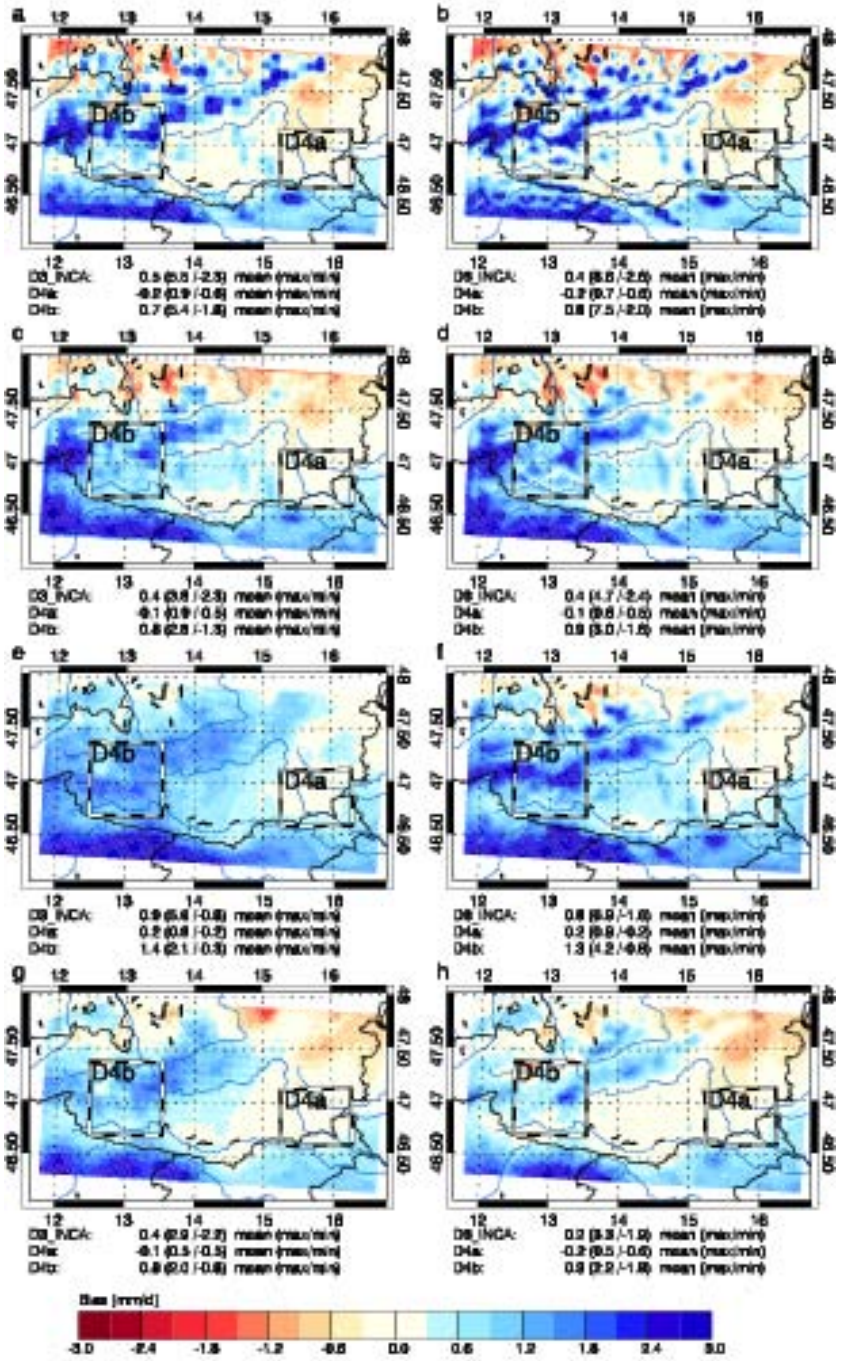


Figure A.4: Same as Figure A.1 but for precipitation [mm/d] in winter 2007/08.

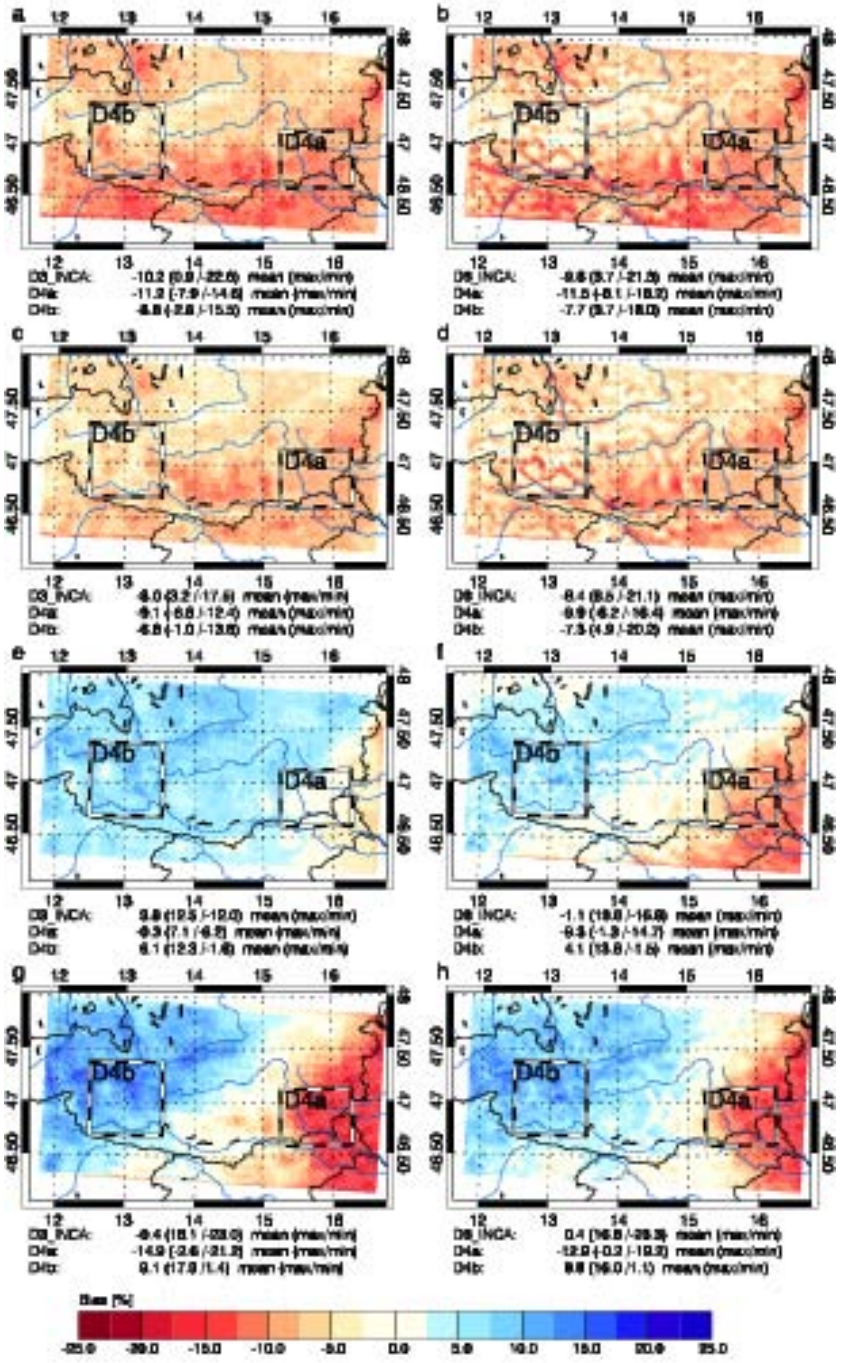


Figure A.5: Same as Figure A.1 but for relative humidity [%] in summer 2007.

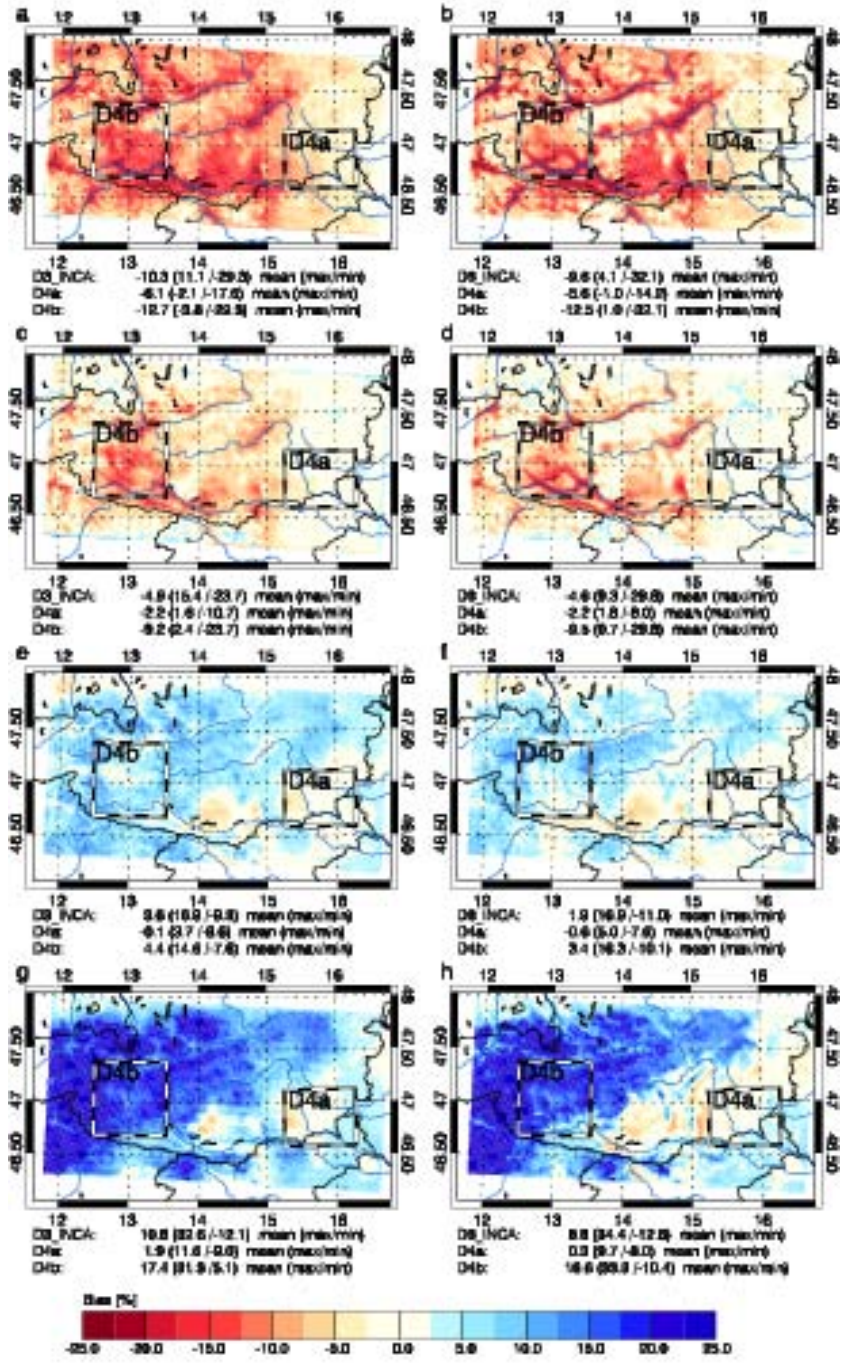


Figure A.6: Same as Figure A.1 but for relative humidity [%] in winter 2007/08.

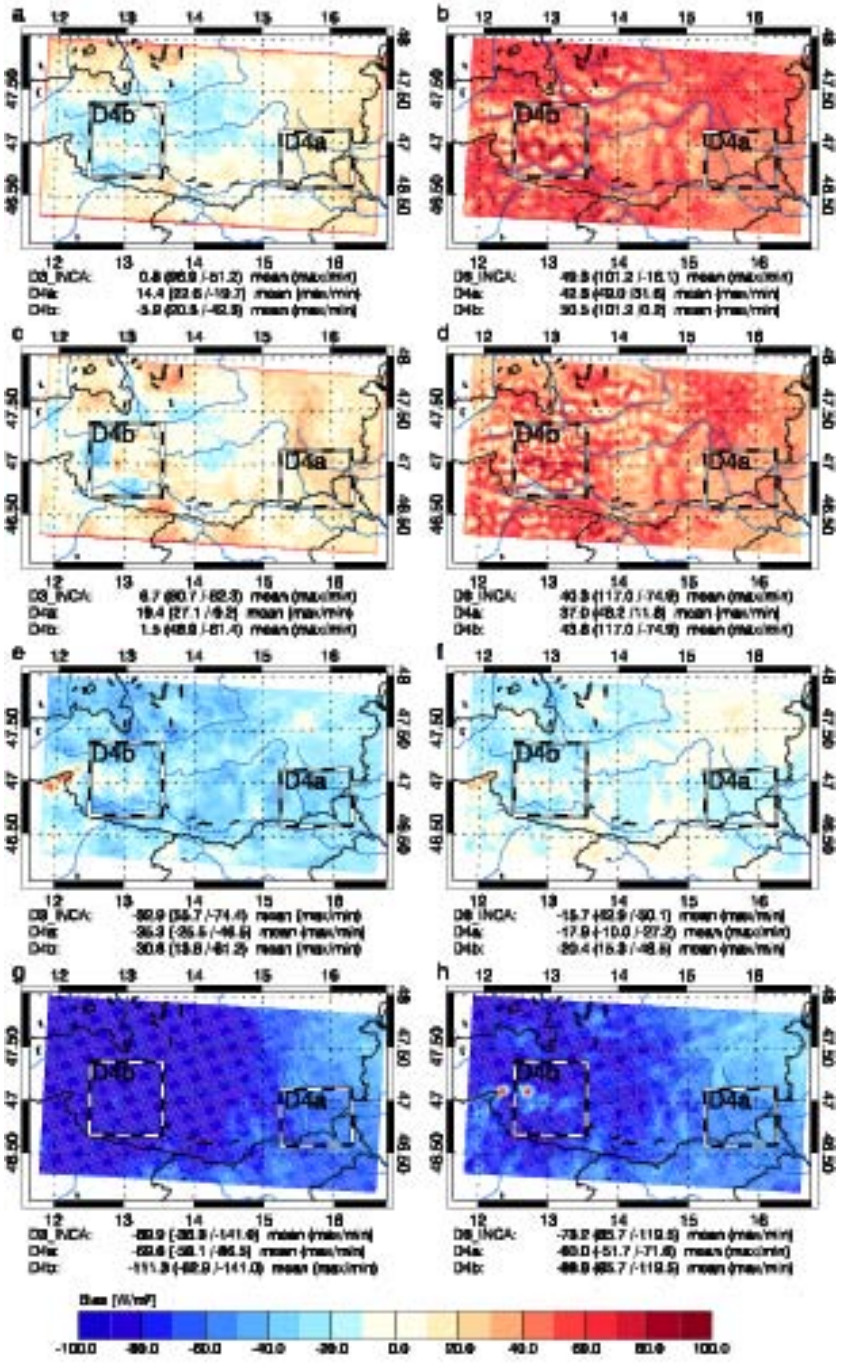


Figure A.7: Same as Figure A.1 but for global radiation [W/m²] in summer 2007.

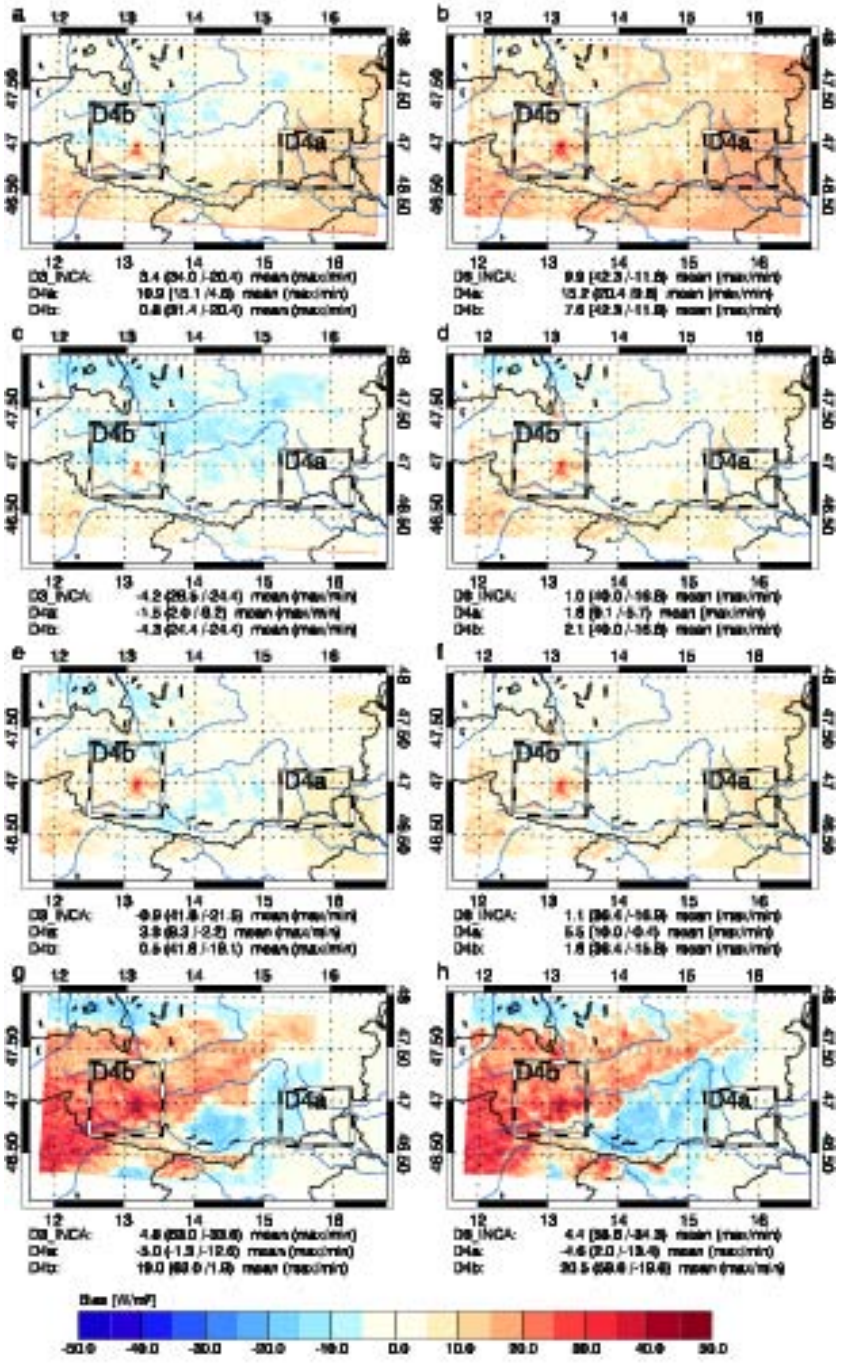


Figure A.8: Same as Figure A.1 but for global radiation [W/m^2] in winter 2007/08. Attention has to be paid to the differences in the error ranges when comparing the bias maps of global radiation in summer 2007 (Figure A.7, $\pm$ 100 W/m^2) with those shown here ($\pm$ 50 W/m^2).

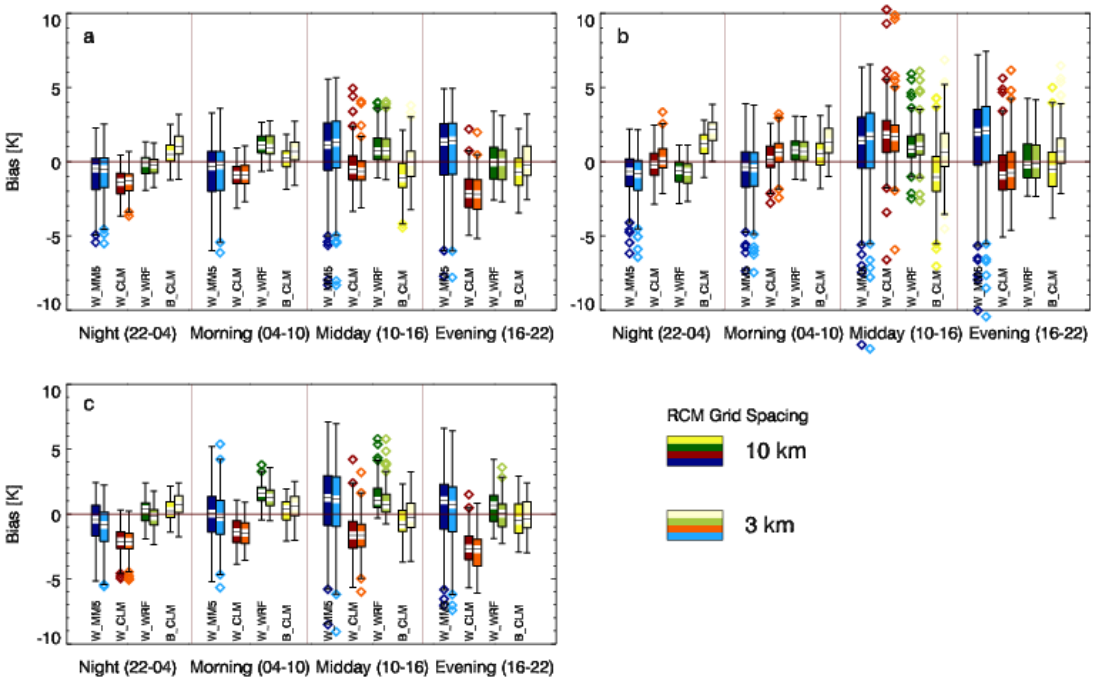


Figure A.9: Box-whisker plots for 2 m temperature [K] in summer. Each plot is divided into four parts correspondent to four quarters of a day. Displayed are quarter daily domain averaged biases for 92 days in summer. The different RCM runs are shown in different colors. 10 km grid spacing simulations are shown in dark colors while 3 km simulations are shown in bright colors. Results are shown for domain D3 in panel a, D4a panel b, and D4b in panel c.

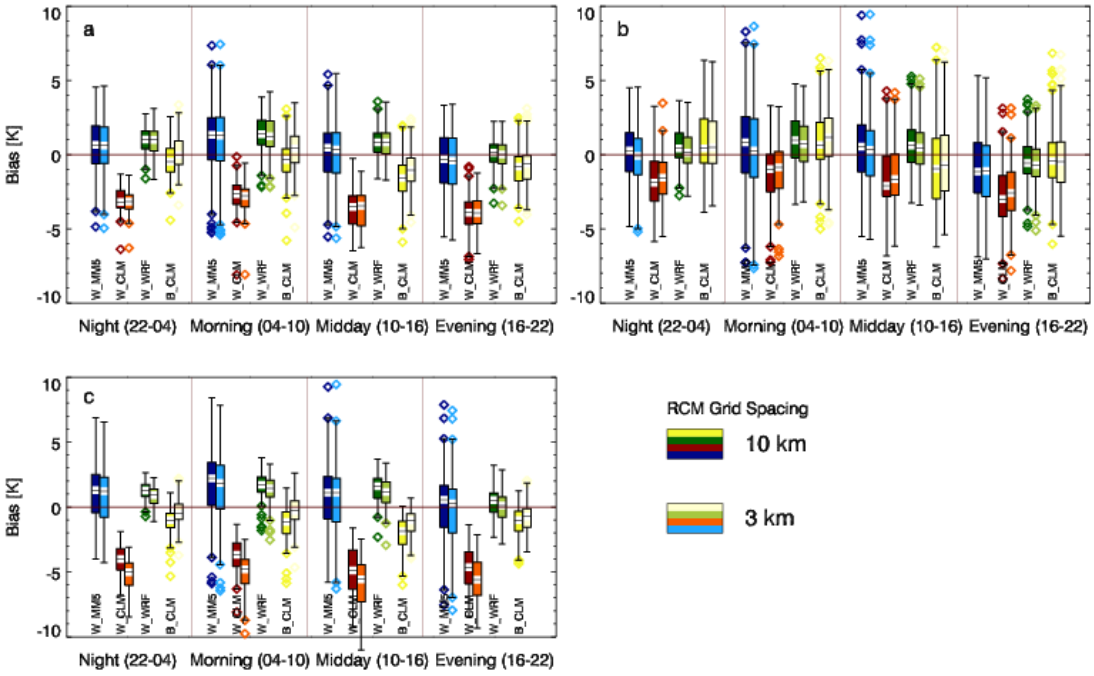


Figure A.10: Same as for Figure A.9 but for 2 m temperature [K] in winter 2007/08.

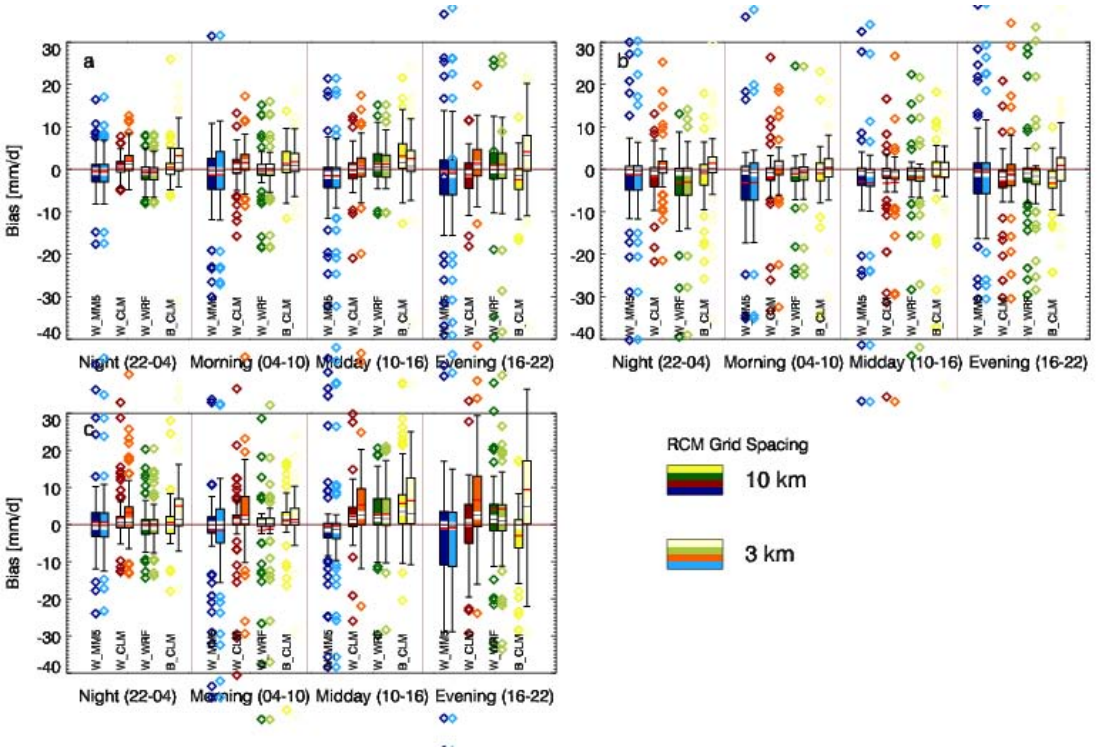


Figure A.11: Same as for Figure A.9 but for precipitation [mm/d] in summer 2007. Additionally, to the median the red line shows the mean bias.

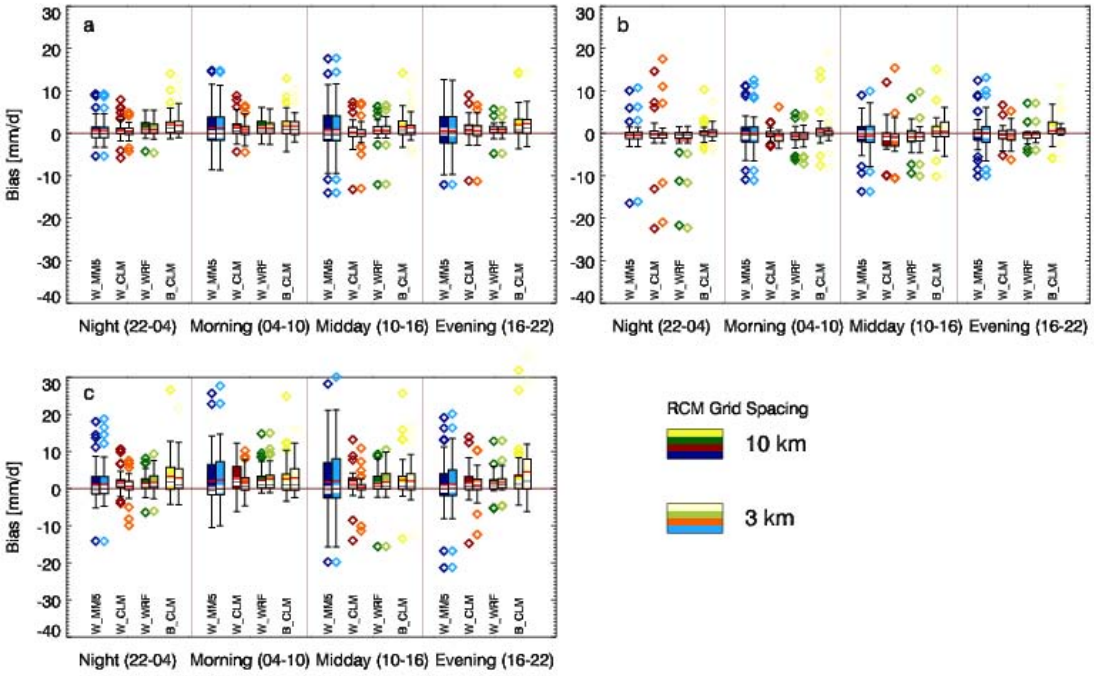


Figure A.12: Same as for Figure A.9 but for precipitation [mm/d] in winter 2007/08.

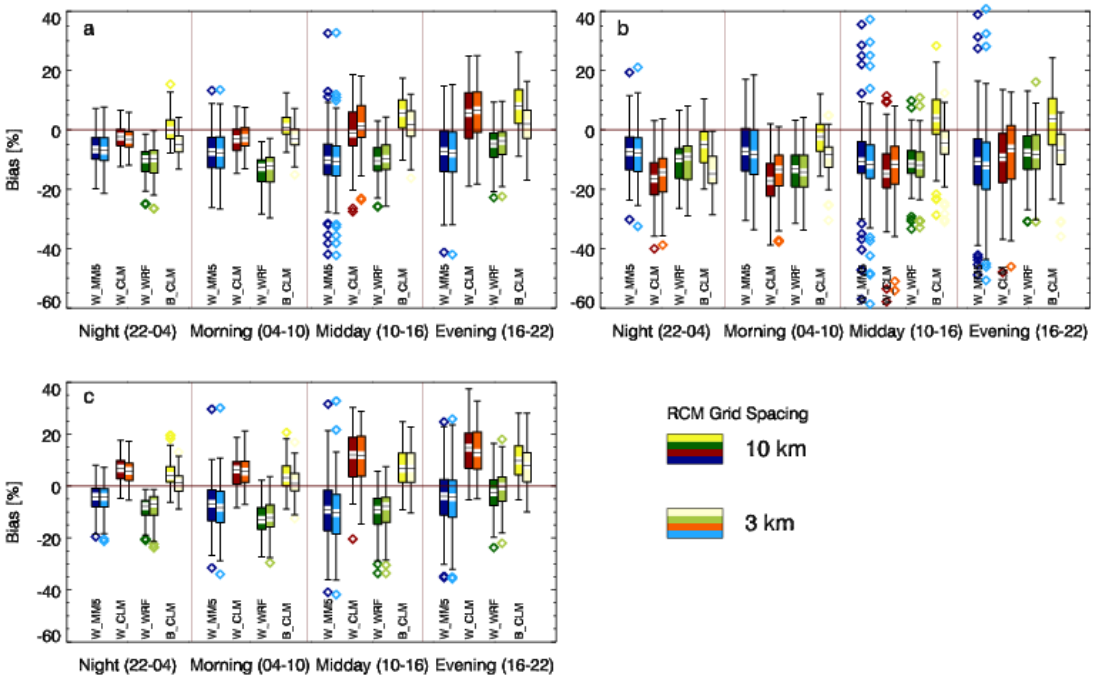


Figure A.13: Same as for Figure A.9 but for relative humidity [%] in summer 2007.

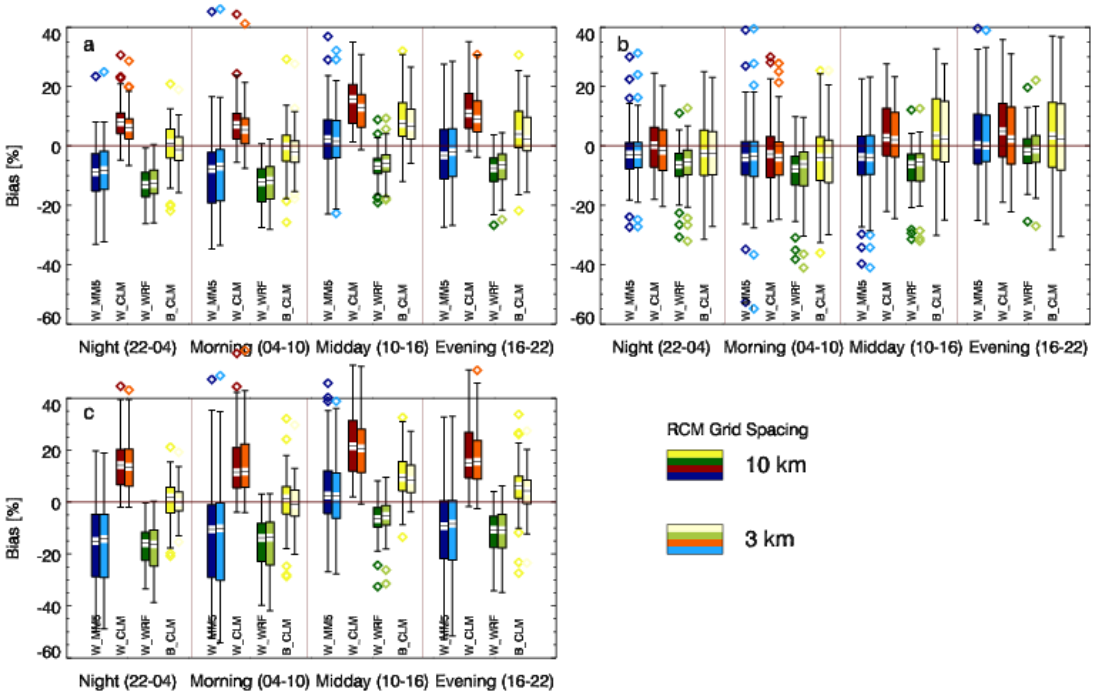


Figure A.14: Same as forFigure A.9 but for relative humidity [%] in winter 2007/08.

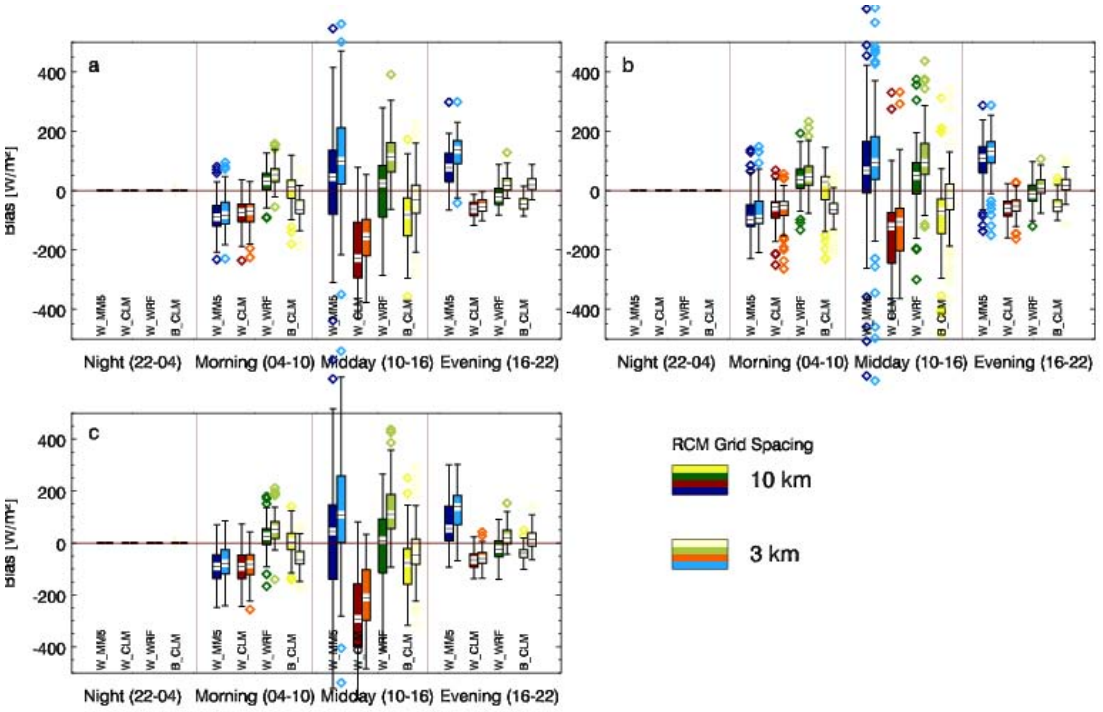


Figure A.15: Same as for Figure A.9 but for global radiation [W/m²] in summer 2007.

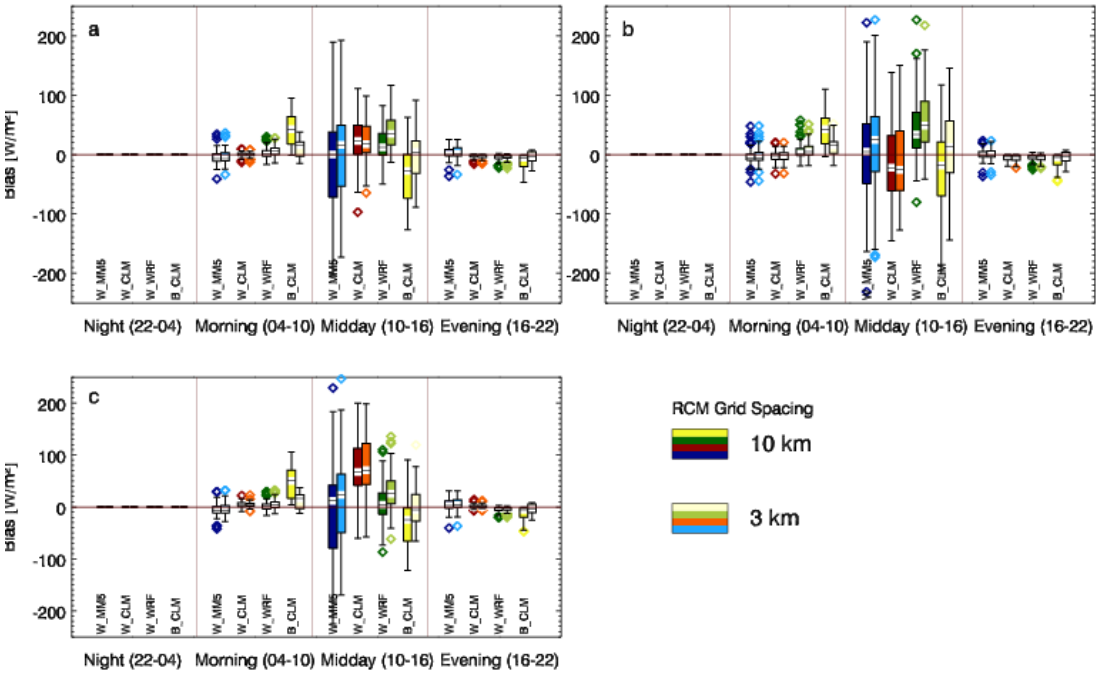


Figure A.16: Same as for Figure A.9 but for global radiation [W/m²] in winter 2007/08.

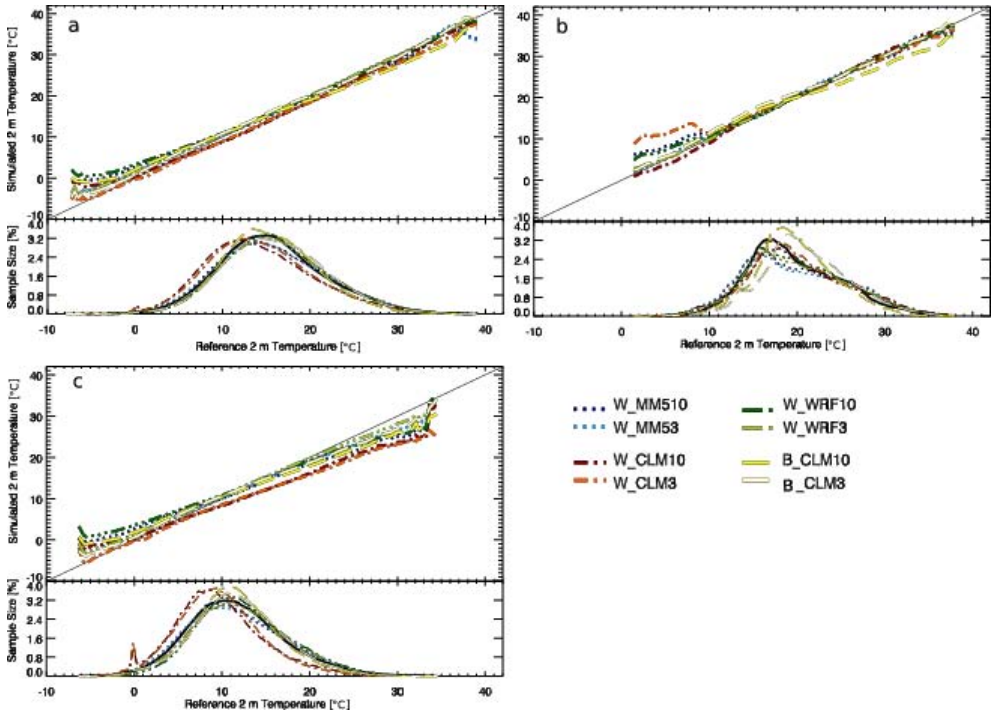


Figure A.17: Conditional quantile plot for the 2 m temperature [K] in summer 2007. Different RCMs results are shown in different colors and line styles in accordance with the legend. 10 km grid spacing simulations are displayed in dark colors while 3 km simulations are shown in light colors. Each subpanel is divided into two parts: the upper part shows the median values of the conditional distribution and the lower part shows the unconditional (density) distribution of the simulations and the reference. Results are given for D3 in panel a, D4a panel b, and D4b in panel c.

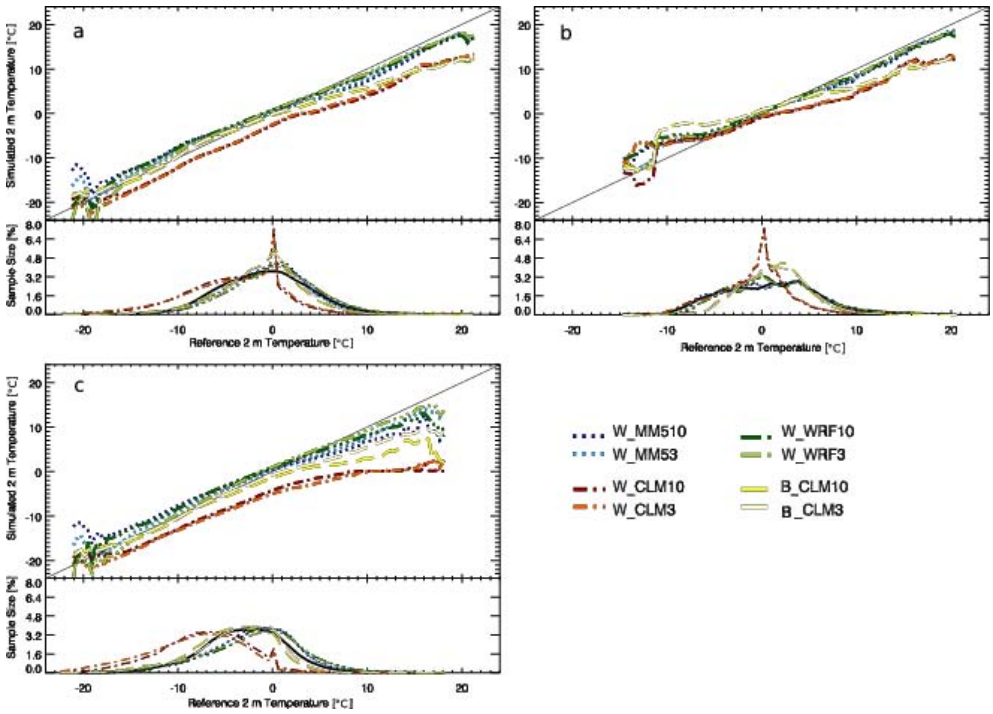


Figure A.18: Same as in Figure A.17, but for 2 m temperature [K] in winter 2007/08.

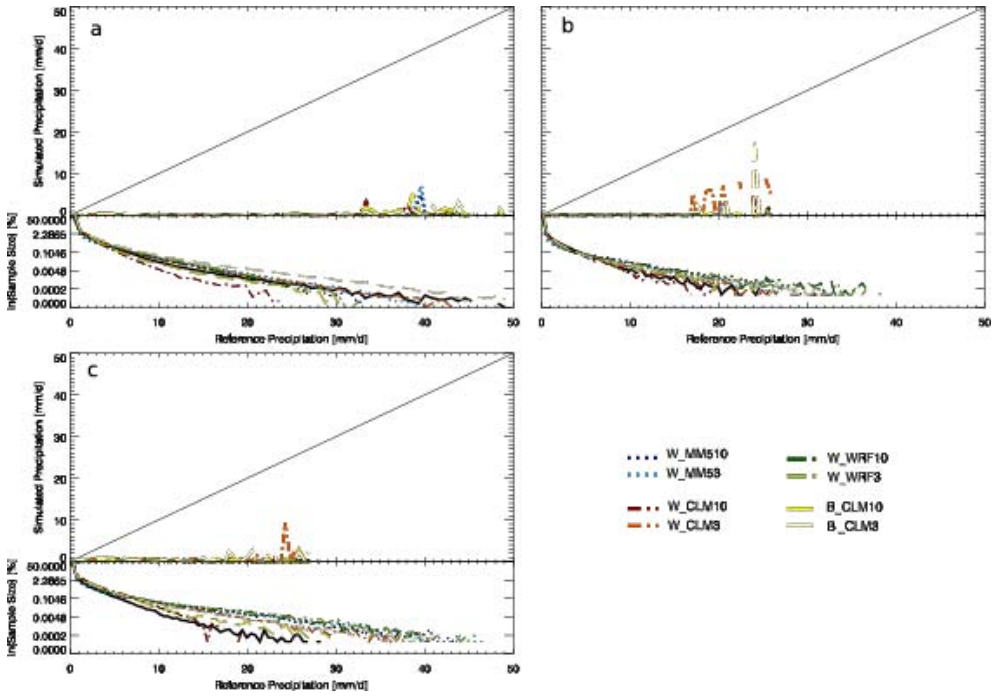


Figure A.19: Same as in Figure A.17, but for precipitation [mm/d] in summer 2007. The y-axis of the density function plots is logarithmically scaled.

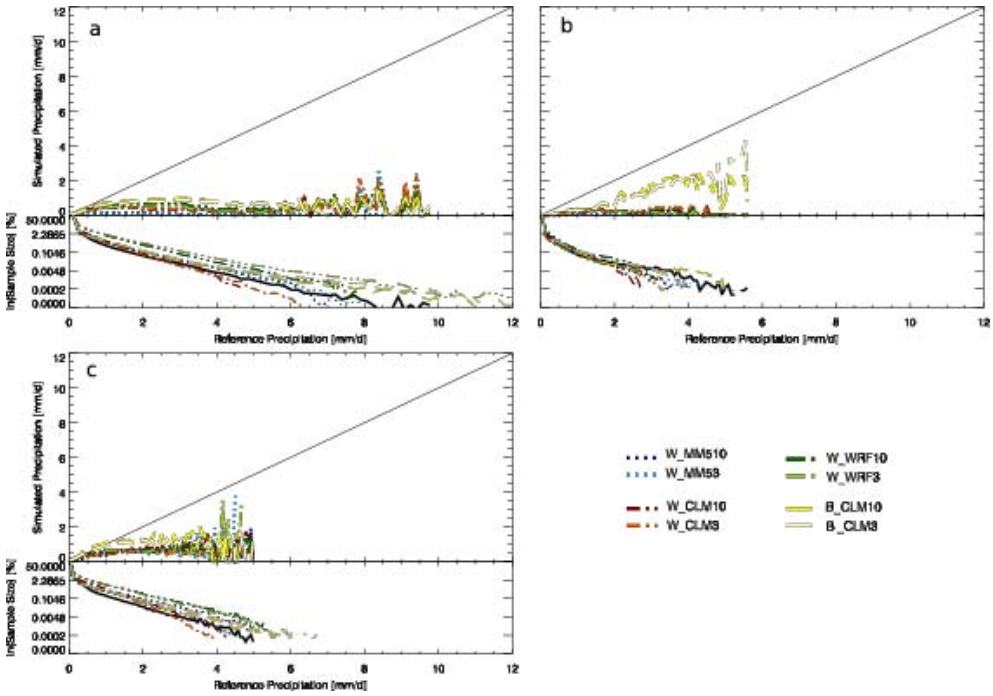


Figure A.20: Same as in Figure A.17, but for precipitation [mm/d] in winter 2007/08. The y-axis of the density function plots is logarithmically scaled.

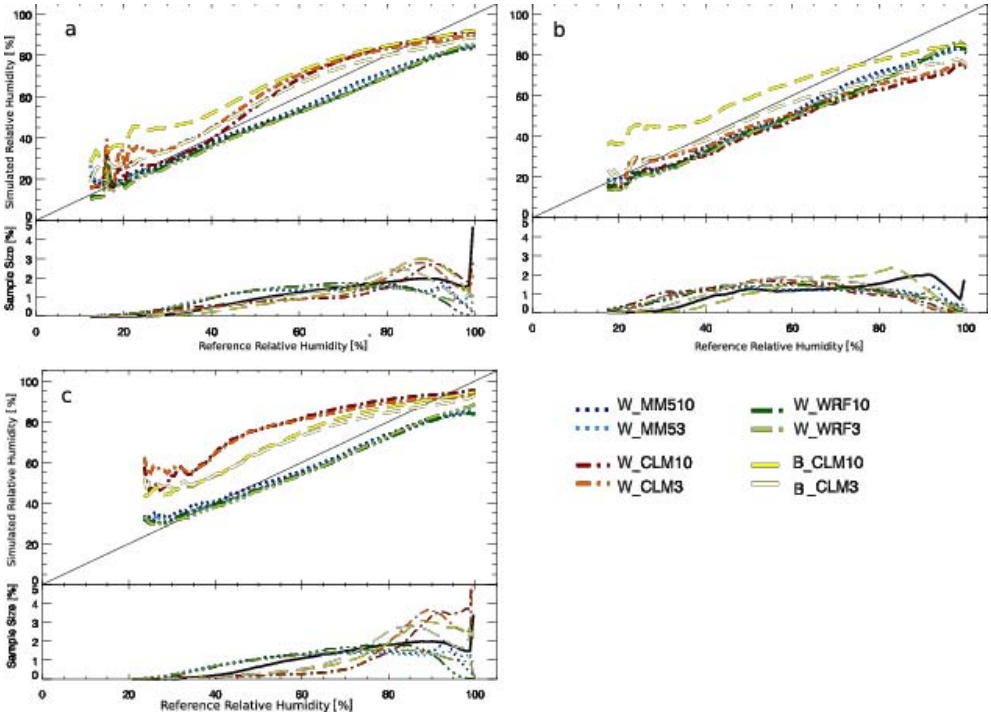


Figure A.21: Same as in Figure A.17, but for relative humidity [%] in summer 2007.

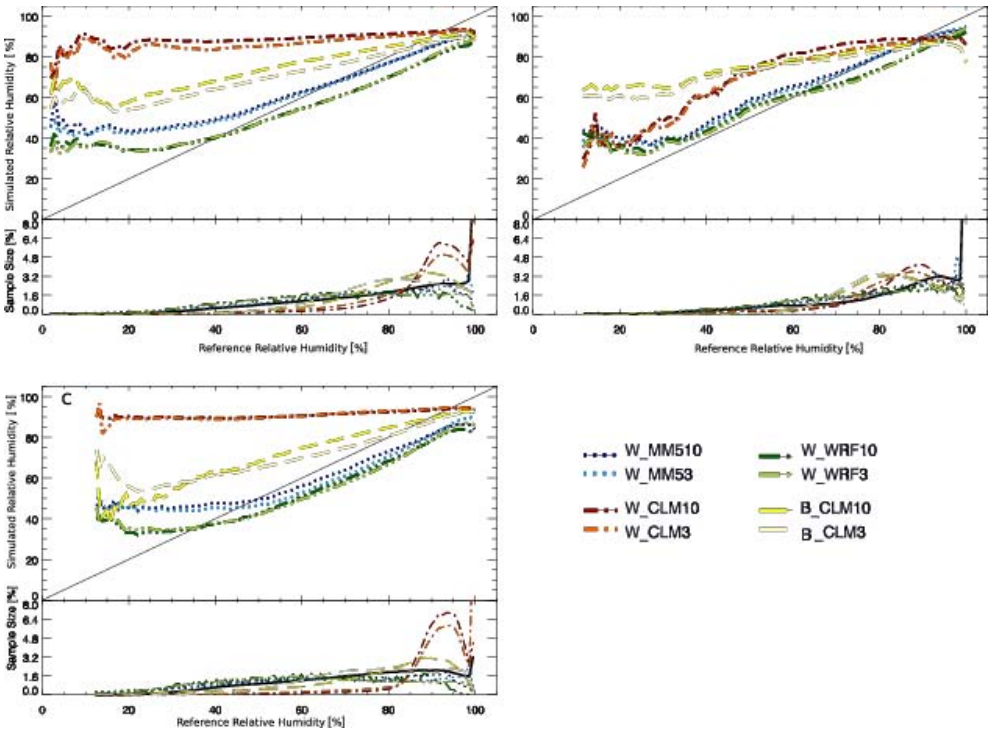


Figure A.22: Same as in Figure A.17, but for relative humidity [%] in winter 2007/08.

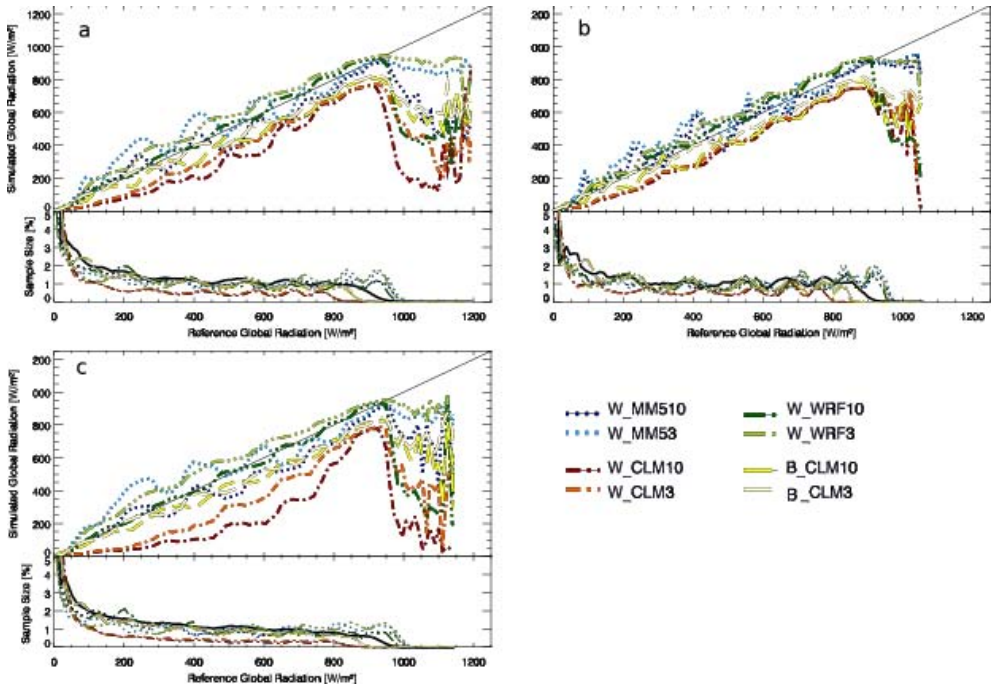


Figure A.23: Same as in Figure A.17, but for global radiation [W/m^2] in summer 2007.

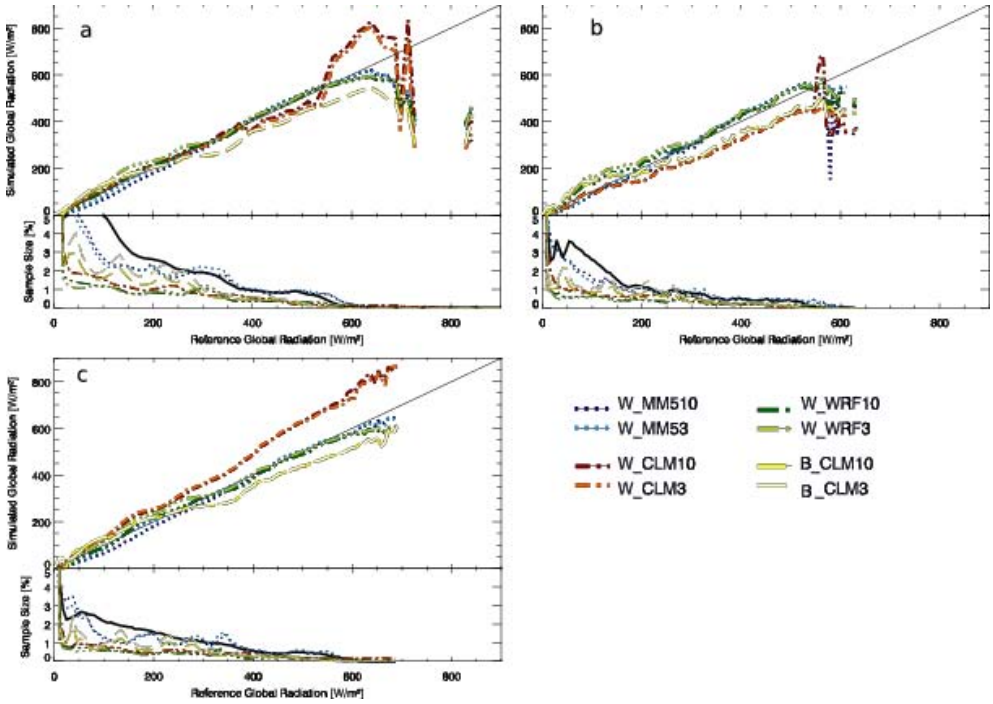


Figure A.24: Same as in Figure A.17, but for global radiation [W/m²] in winter 2007/08.

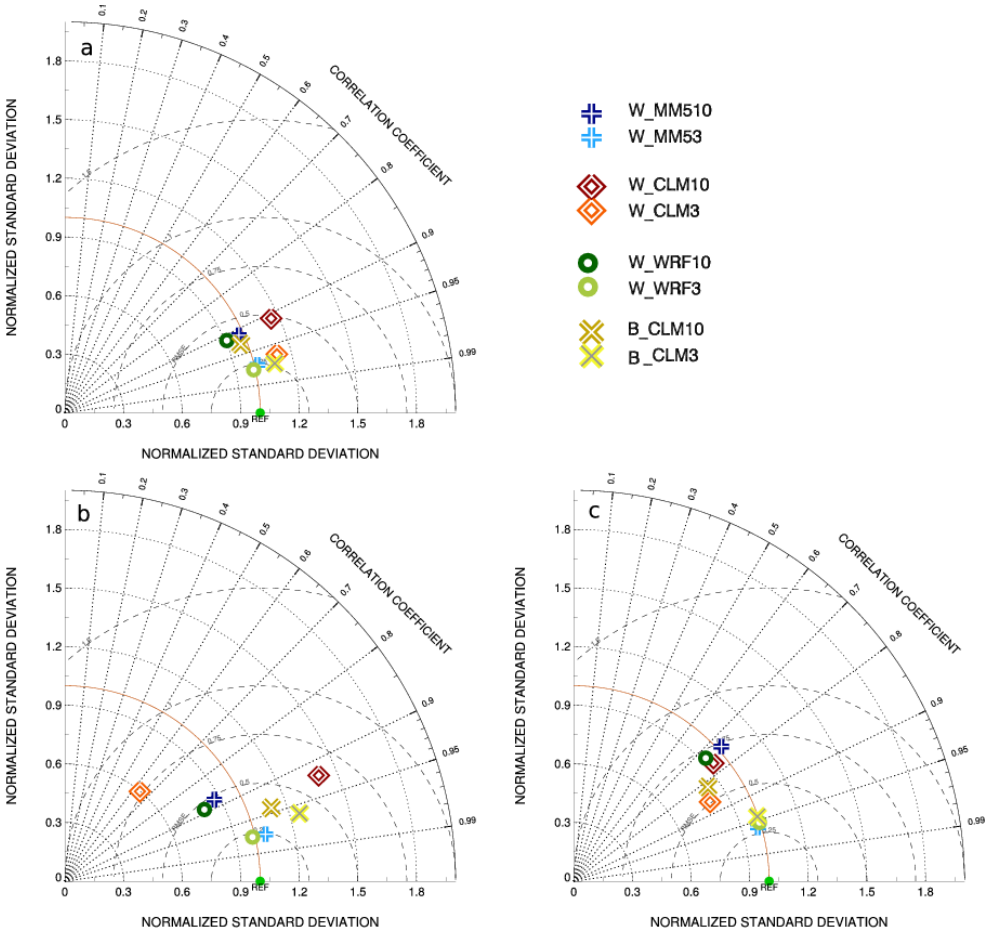


Figure A.25: Taylor plot which shows the spatial values of the temporal averaged field of 2 m temperature in summer 2007. Different RCMs have different symbols and colors. 10 km grid spacing simulations have dark colors while 3 km simulations have bright colors. Panel a displays the results on D3, while panel b and c are showing the results for D4a and D4b.

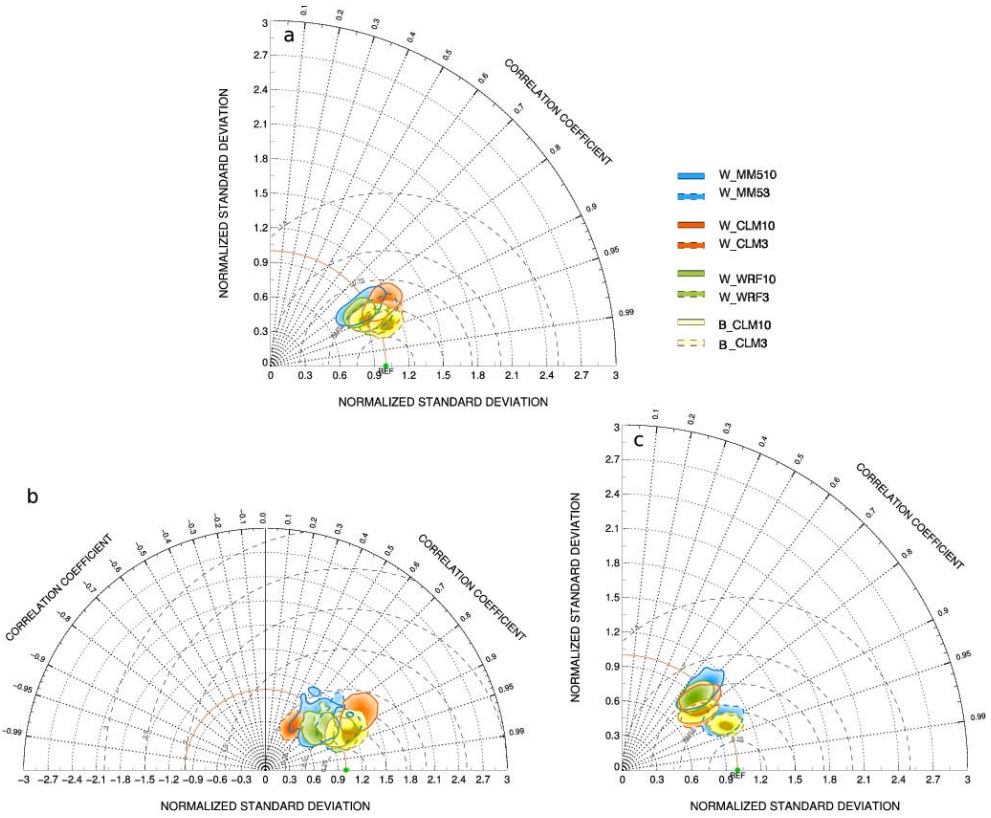


Figure A.26: Taylor plot which shows the spatial values of the hourly fields of 2 m temperature in summer 2007 as contours. Different RCMs have different colors. The brightness of the color represents the density of data points in this area. 10 km grid spacing simulations have solid lines around the outermost contour while 3 km simulations have dashed lines. Panel a) displays the results on D3, while panel b) and c) are showing the results for D4a and D4b.

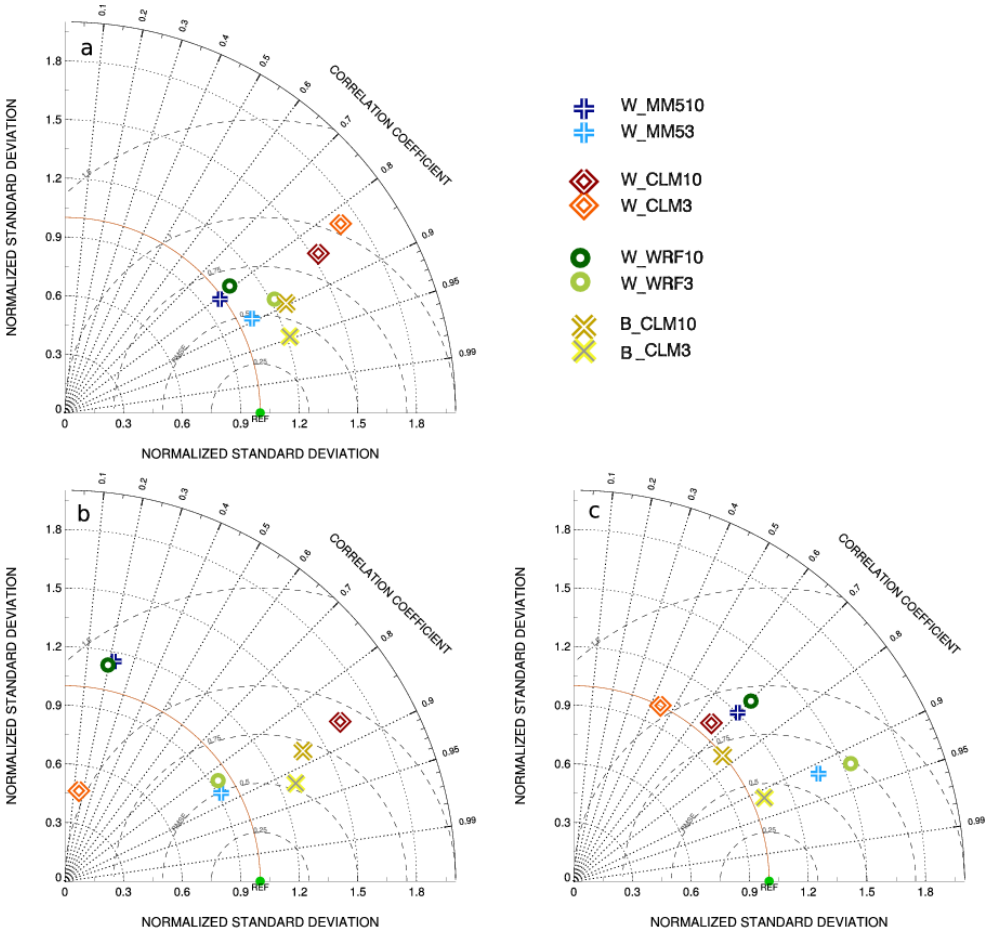


Figure A.27: Same as in Figure A.25 but for 2 m temperature in winter 2007/08.

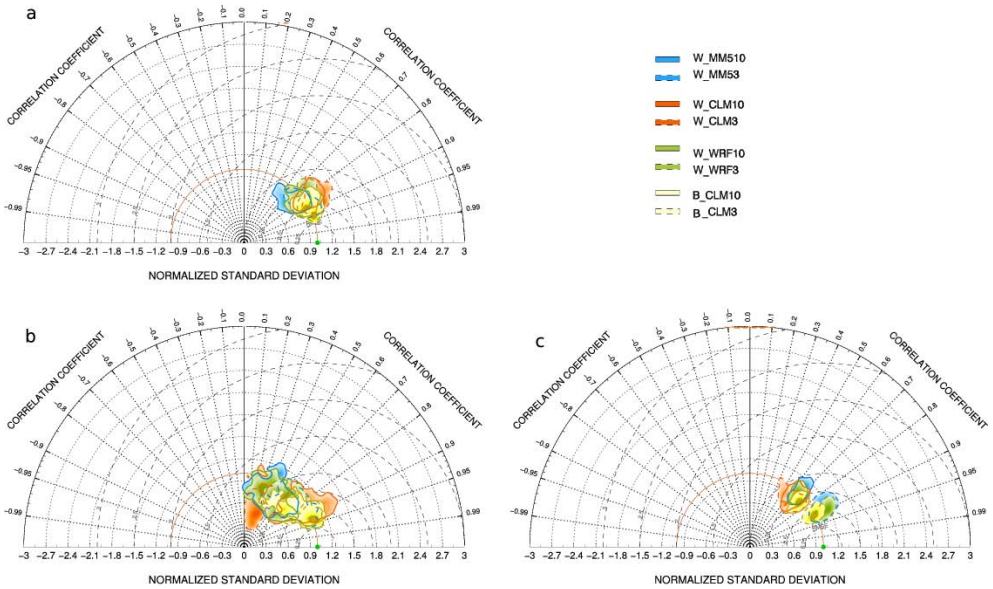


Figure A.28: Same as in Figure A.26 but for 2 m temperature in winter 2007/08.

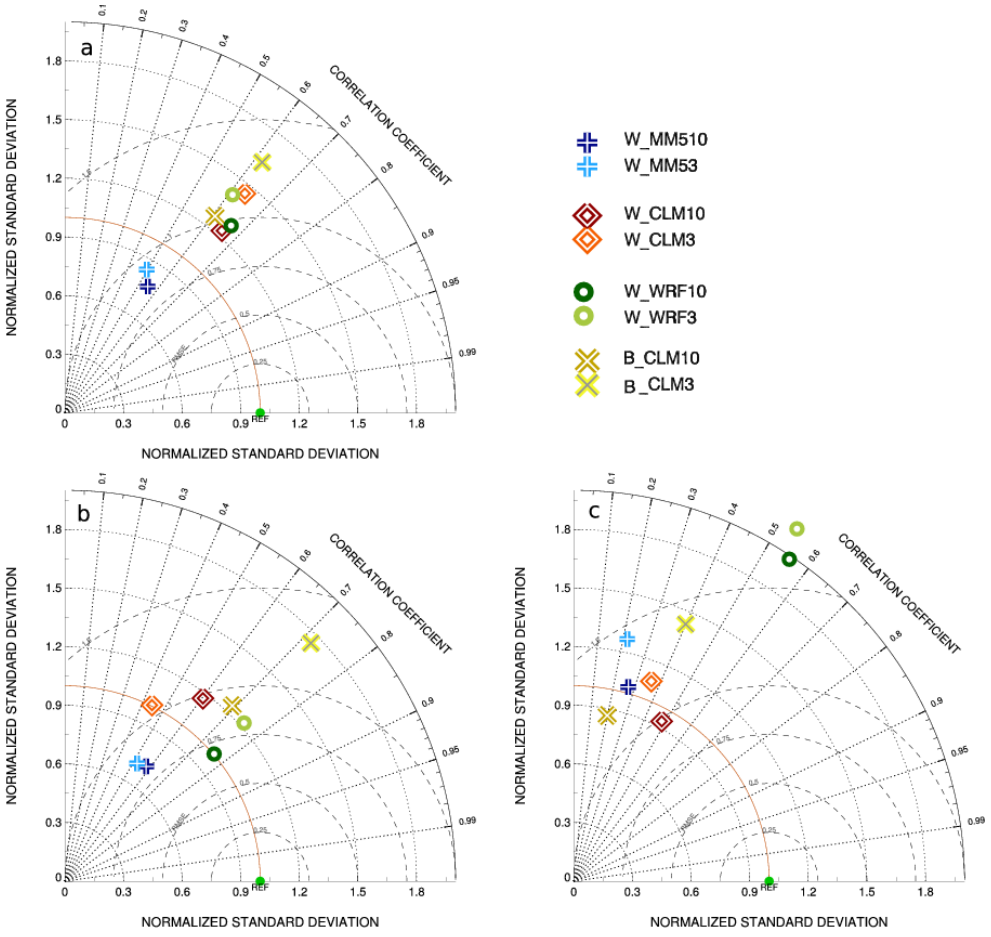


Figure A.29: Same as in Figure A.25 but for precipitation in summer 2007.

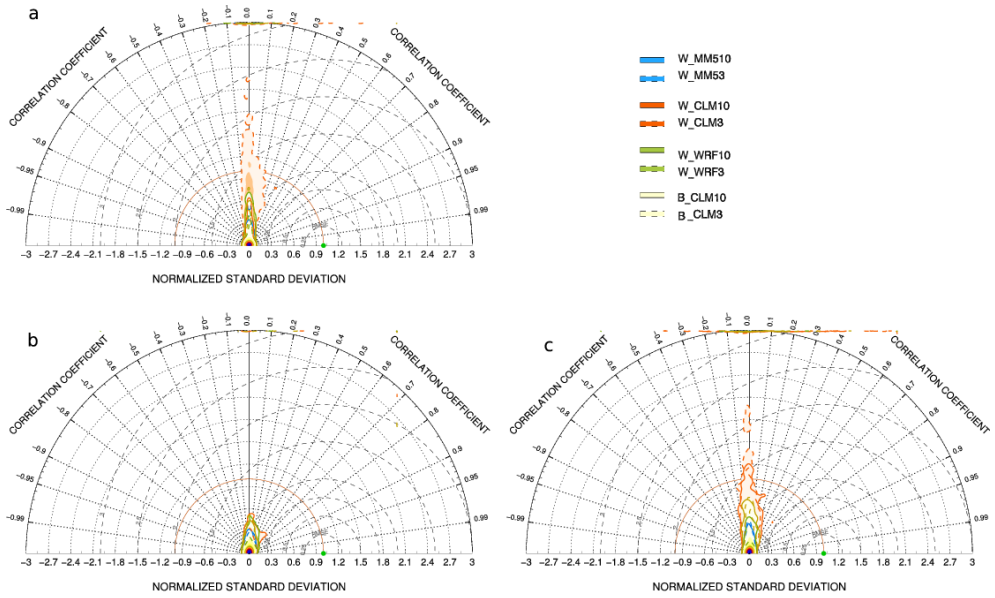


Figure A.30: Same as in Figure A.26 but for precipitation in summer 2007.

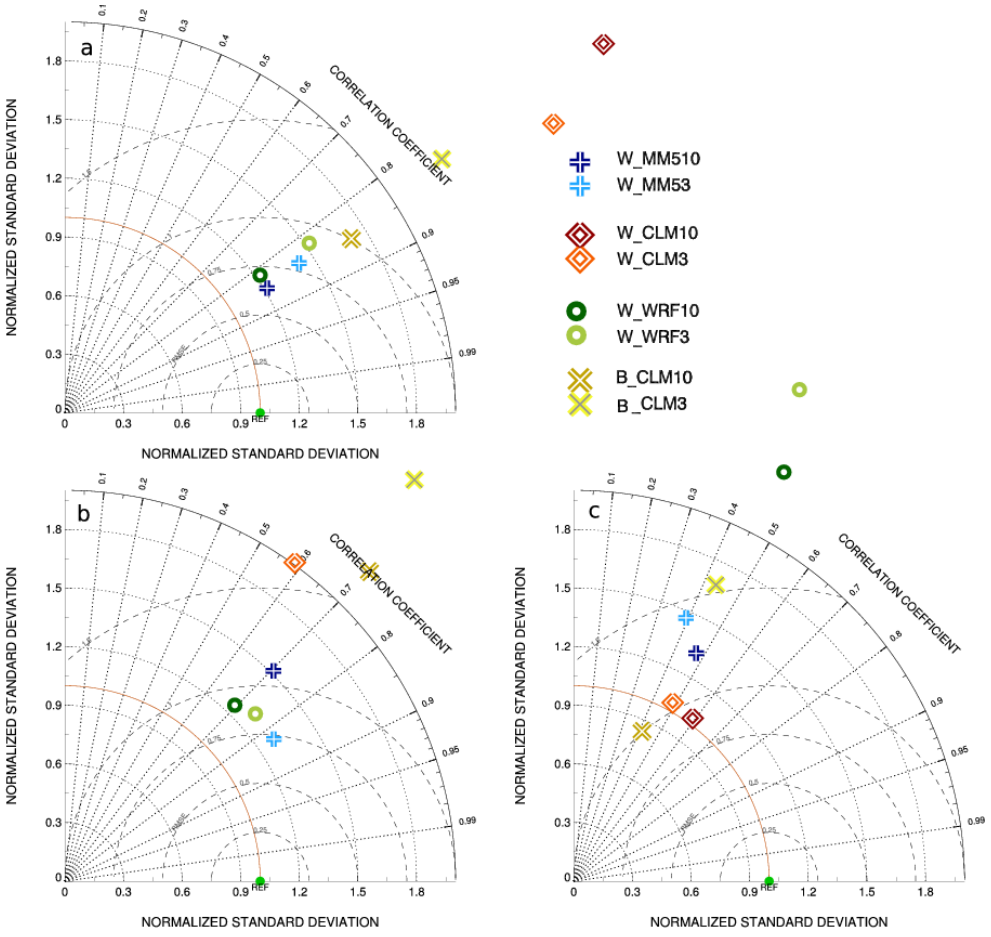


Figure A.31: Same as in Figure A.25 but for precipitation in winter 2007/08.

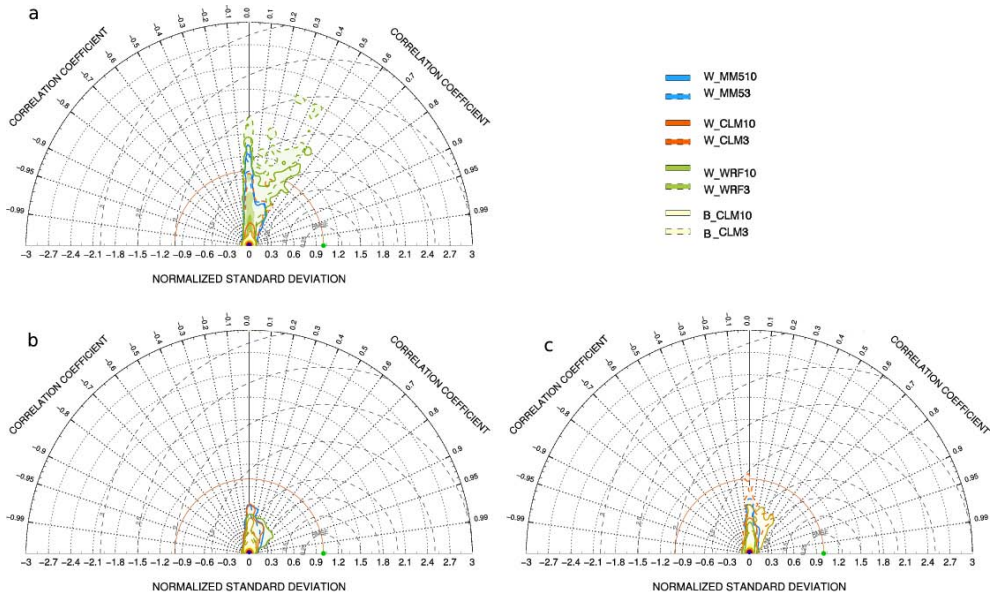


Figure A.32: Same as in Figure A.26 but for precipitation in winter 2007/08.

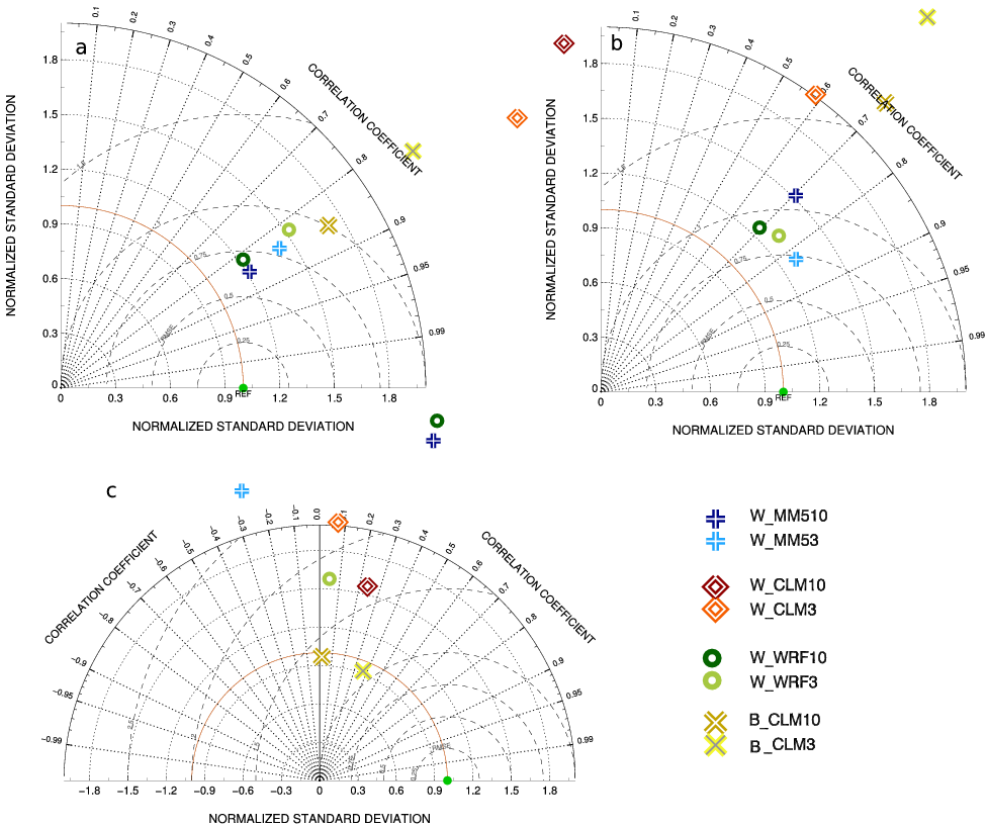


Figure A.33: Same as in Figure A.25 but for relative humidity in summer 2007.

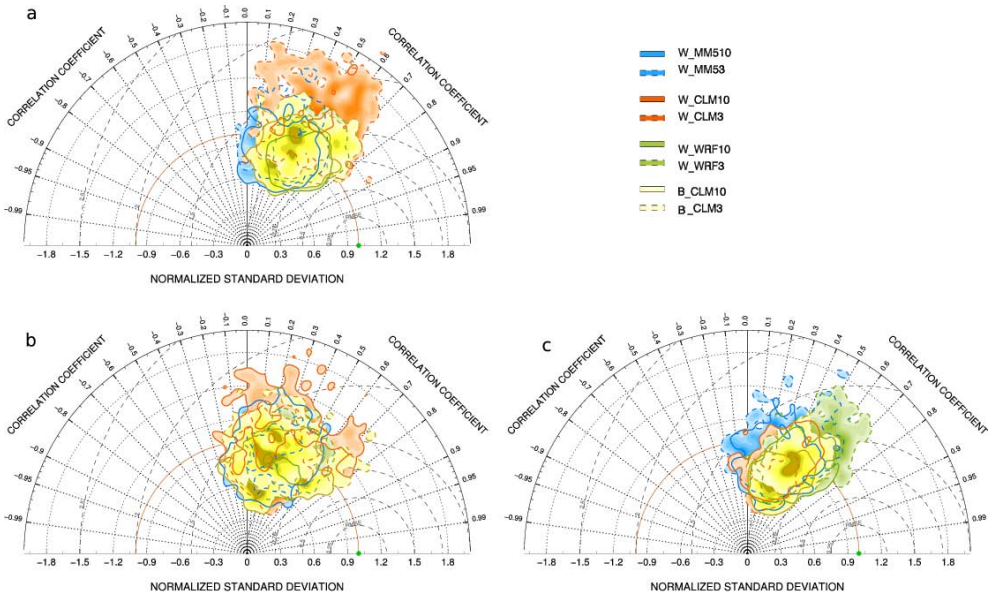


Figure A.34: Same as in Figure A.26 but for relative humidity in summer 2007.

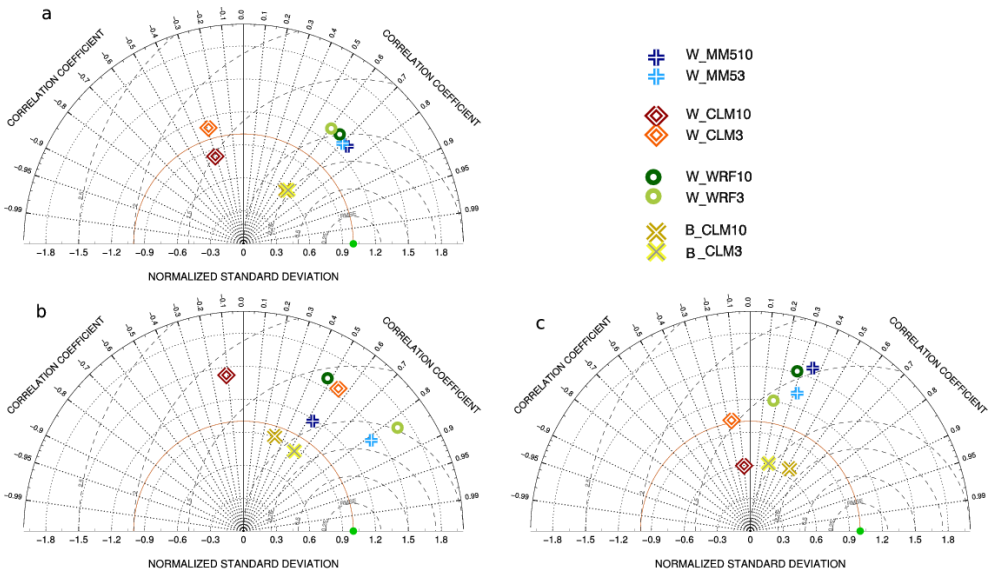


Figure A.35: Same as in Figure A.25 but for relative humidity in winter 2007/08.

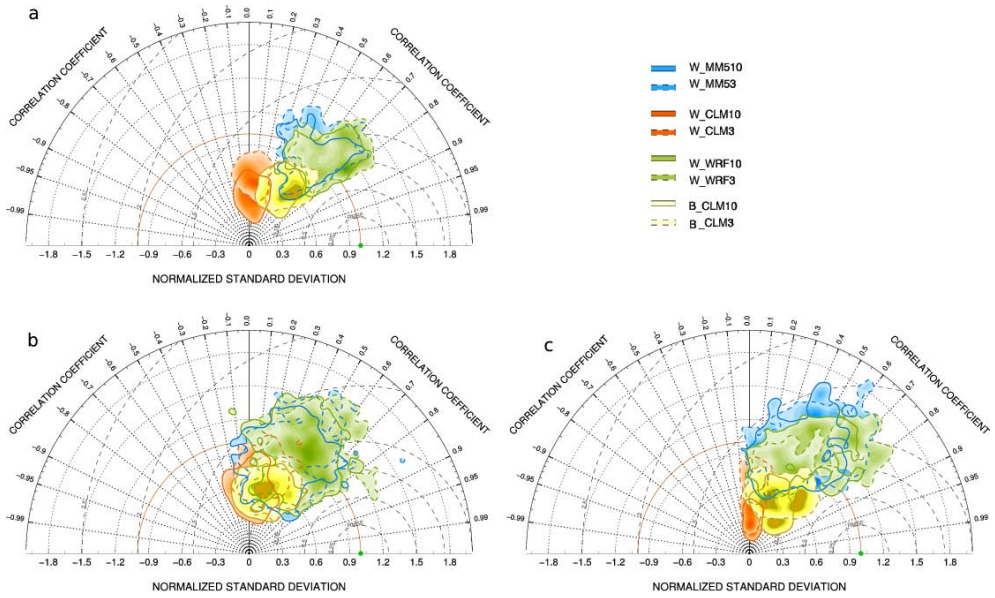


Figure A.36: Same as in Figure A.26 but for relative humidity in winter 2007/08.

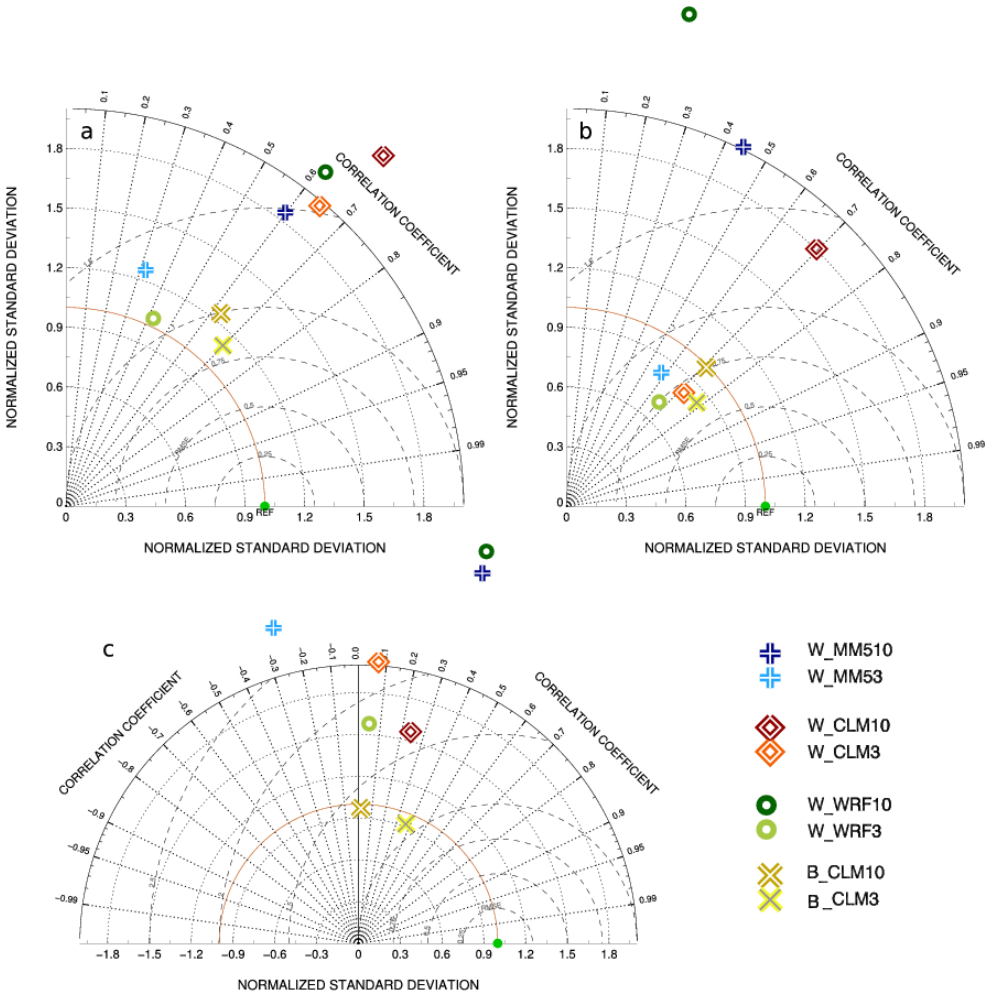


Figure A.37: Same as in Figure A.25 but for global radiation in summer 2007.

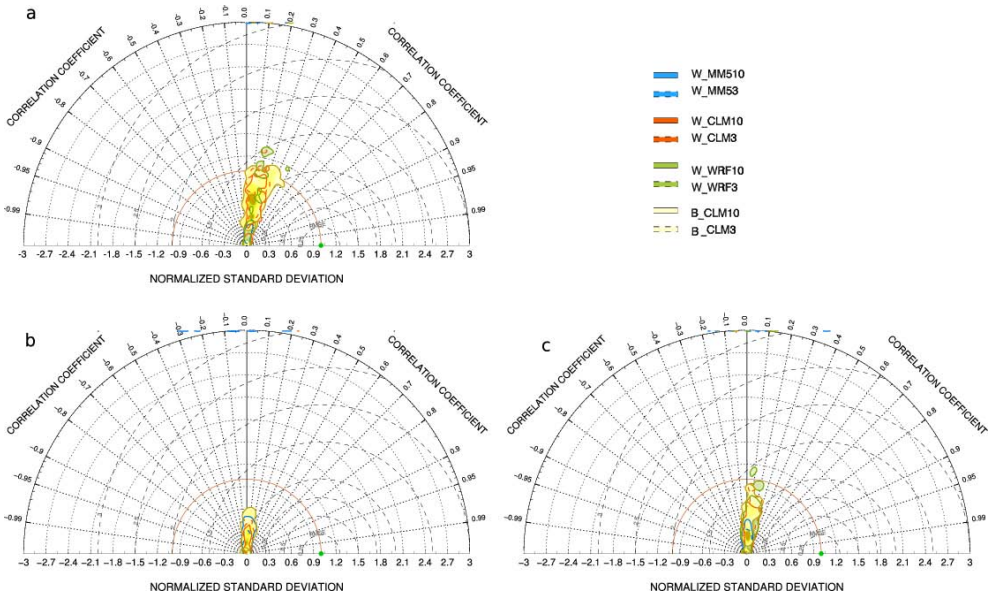


Figure A.38: Same as in Figure A.26 but for global radiation in summer 2007.

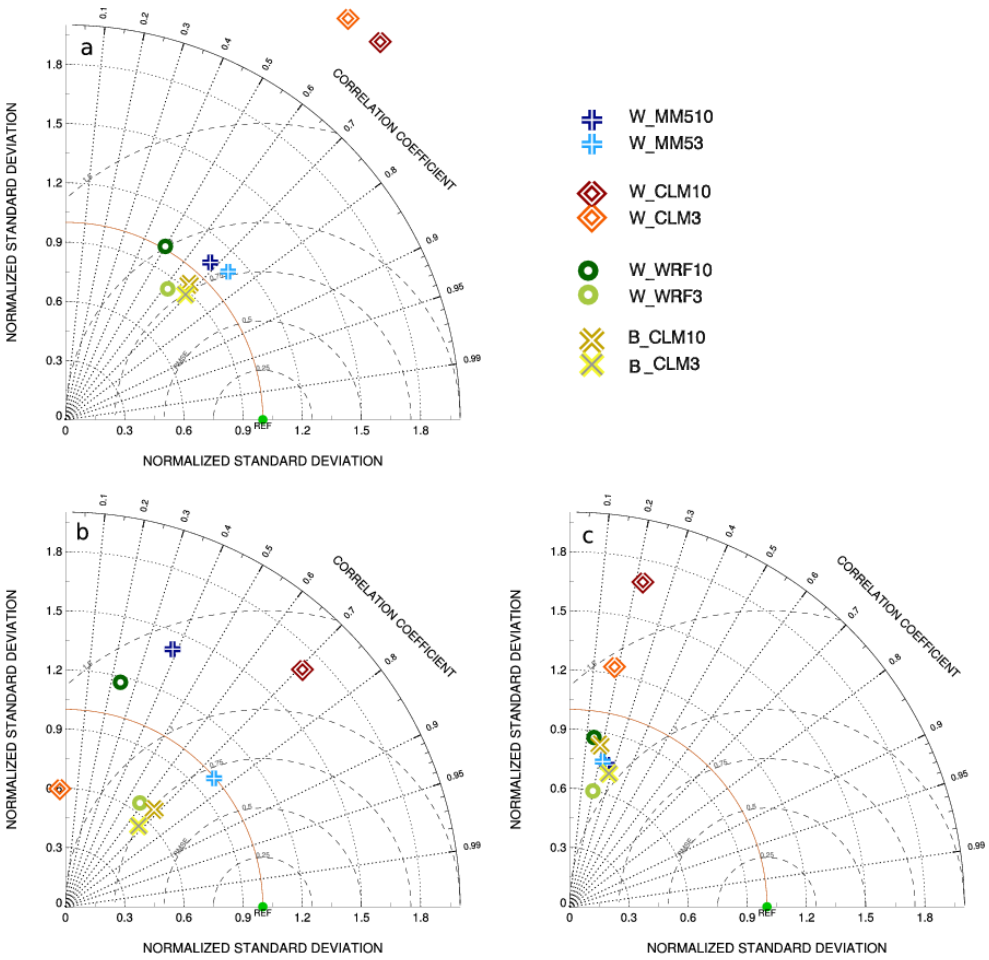


Figure A.39: Same as in Figure A.25 but for global radiation in winter 2007/08.

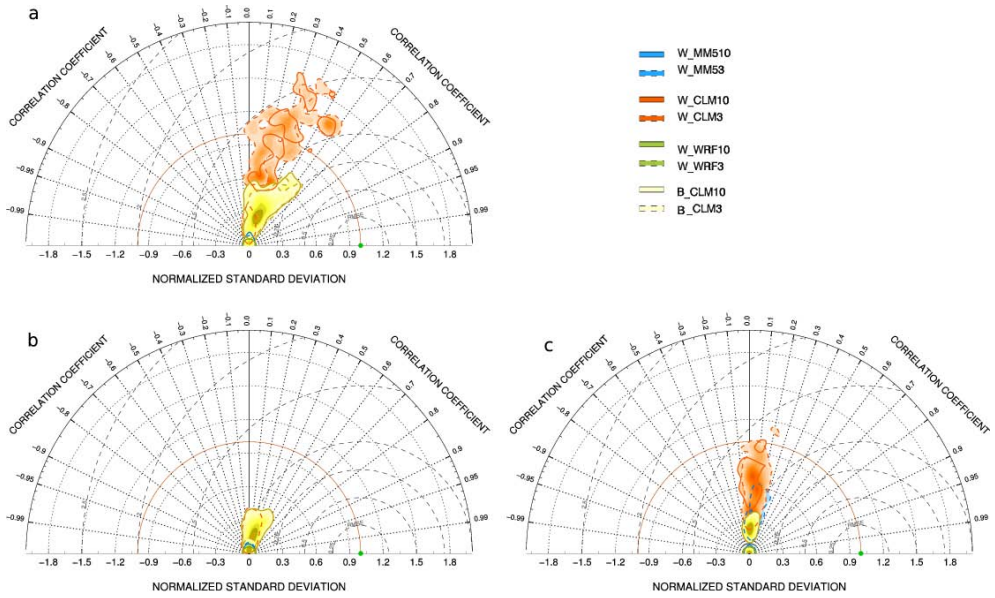


Figure A.40: Same as in Figure A.26 but for global radiation in winter 2007/08.

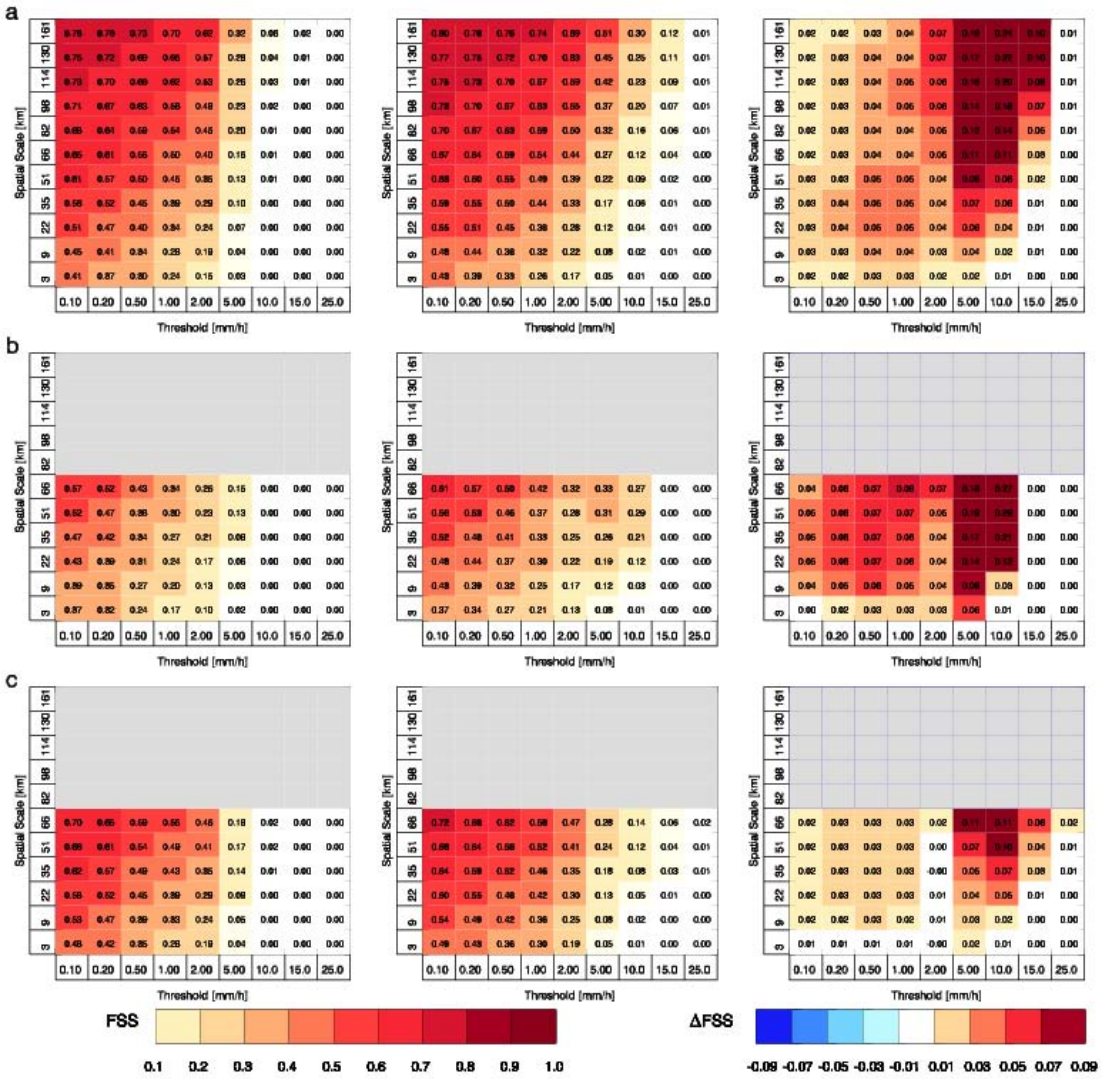


Figure A.41: Temporal averaged FSSs for WegCenter CCLM simulations in summer 2007. The left column shows the FSSs for the W_CLM10 simulations and the middle one those for the W_CLM3 simulation. The right column displays the difference between the 3 km grid spacing simulation and the 10 km simulation. Thereby red colors indicate an improvement in the 3 km simulation, while blue colors show a deterioration. In panel a the results for domain D3, in line b those for D4a, and in line c the results for D4b are shown. On the x-axis of the plots the precipitation threshold in mm/h, and on the y-axis the length of the neighborhood square in km are displayed.

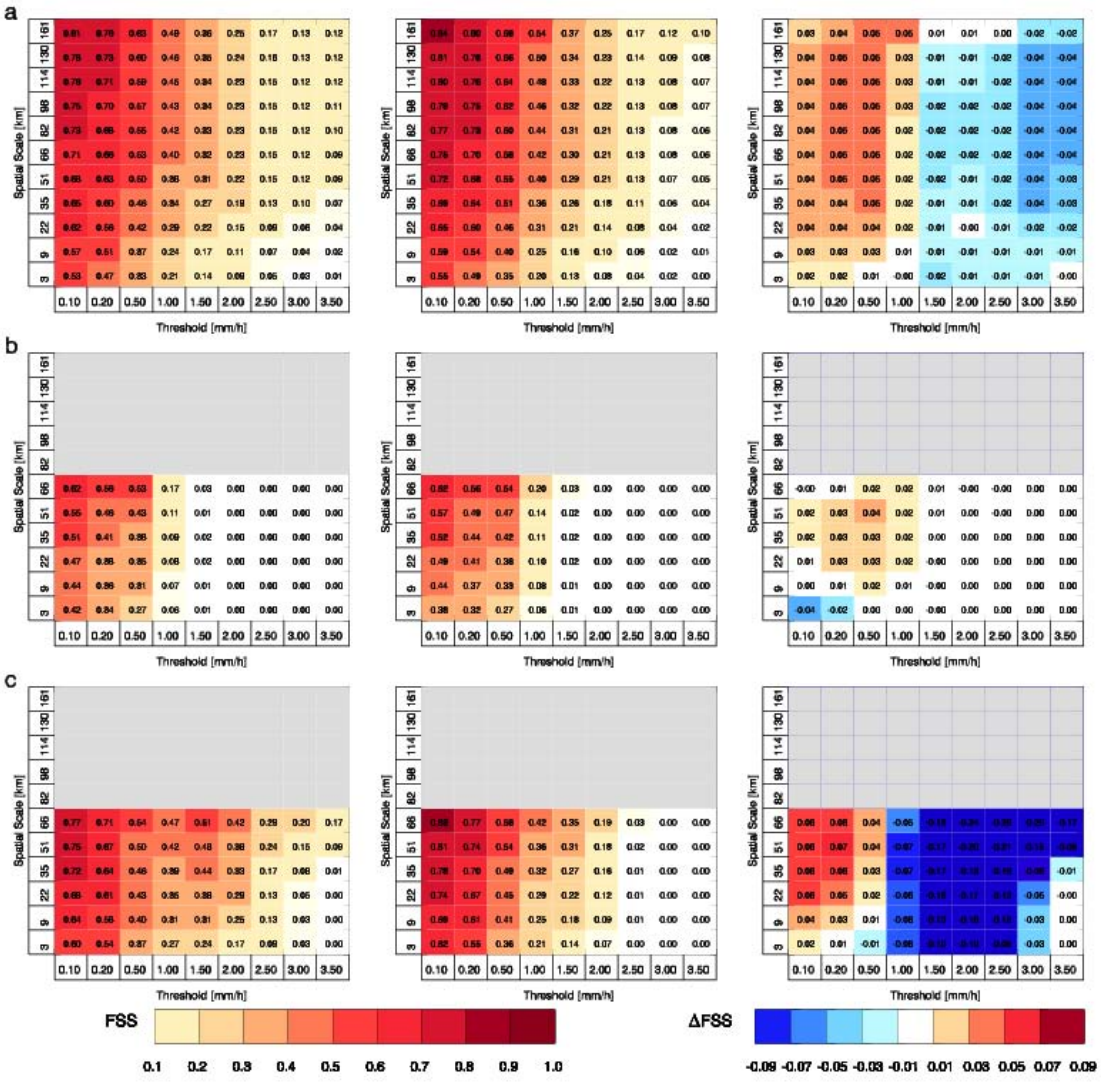


Figure A.42: Same as in Figure A.41 but for the WegCenter CCLM simulations in winter 2007/08.

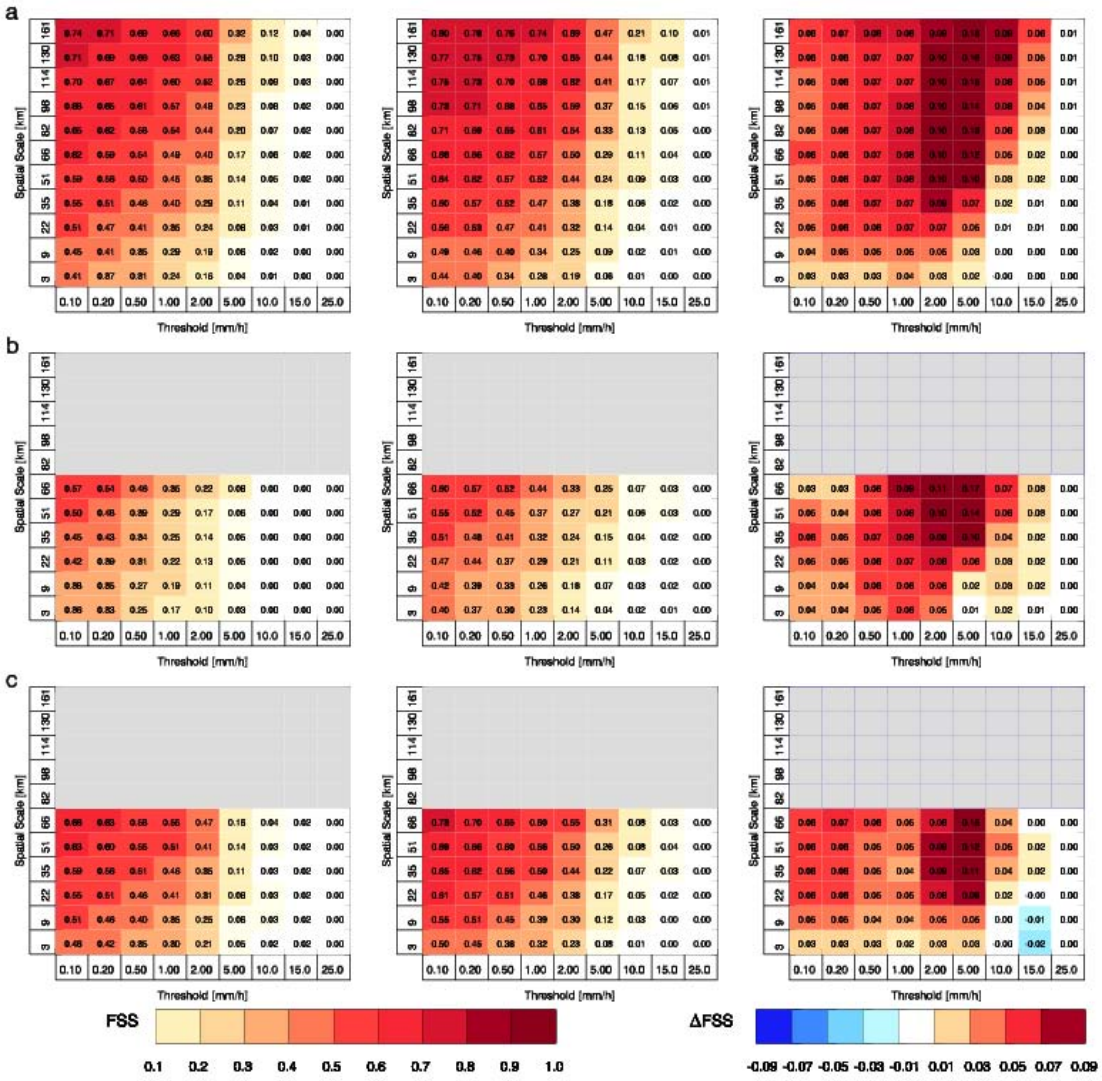


Figure A.43: Same as in Figure A.41 but for the BTU CCLM simulations in summer 2007.

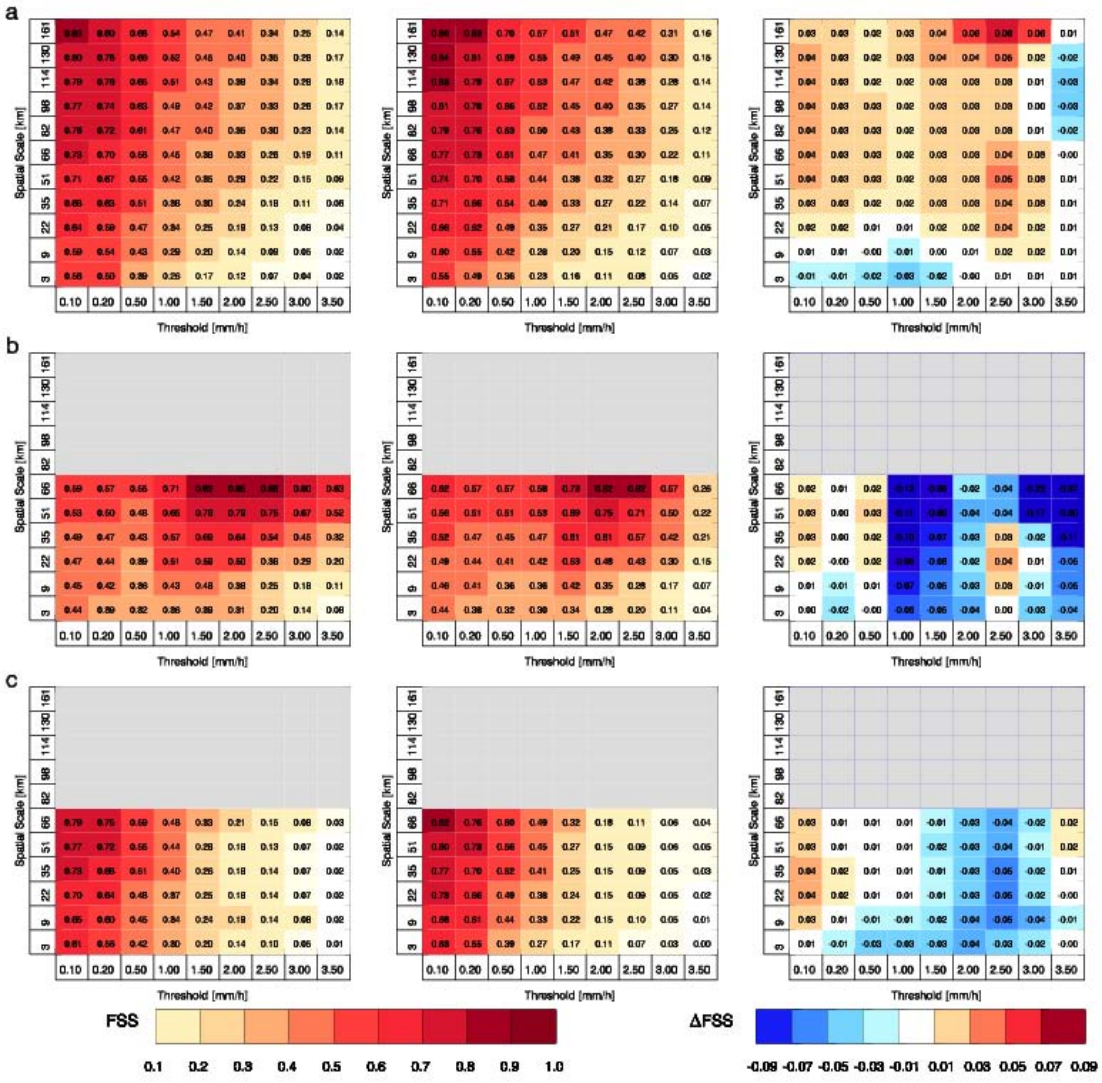


Figure A.44: Same as in Figure A.41 but for the BTU CCLM simulations in winter 2007/08.

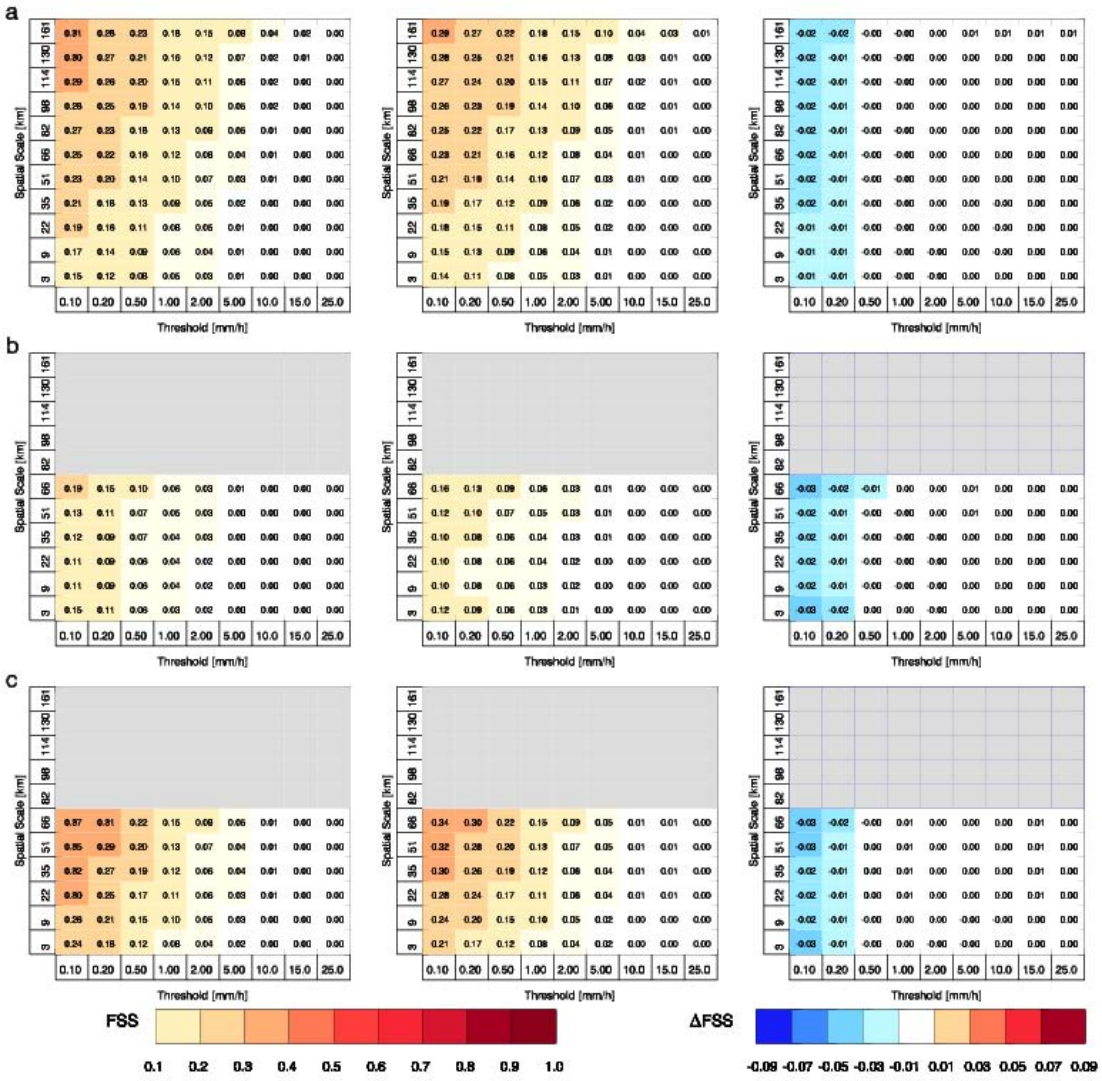


Figure A.45: Same as in Figure A.41 but for the MM5 simulations in summer 2007.

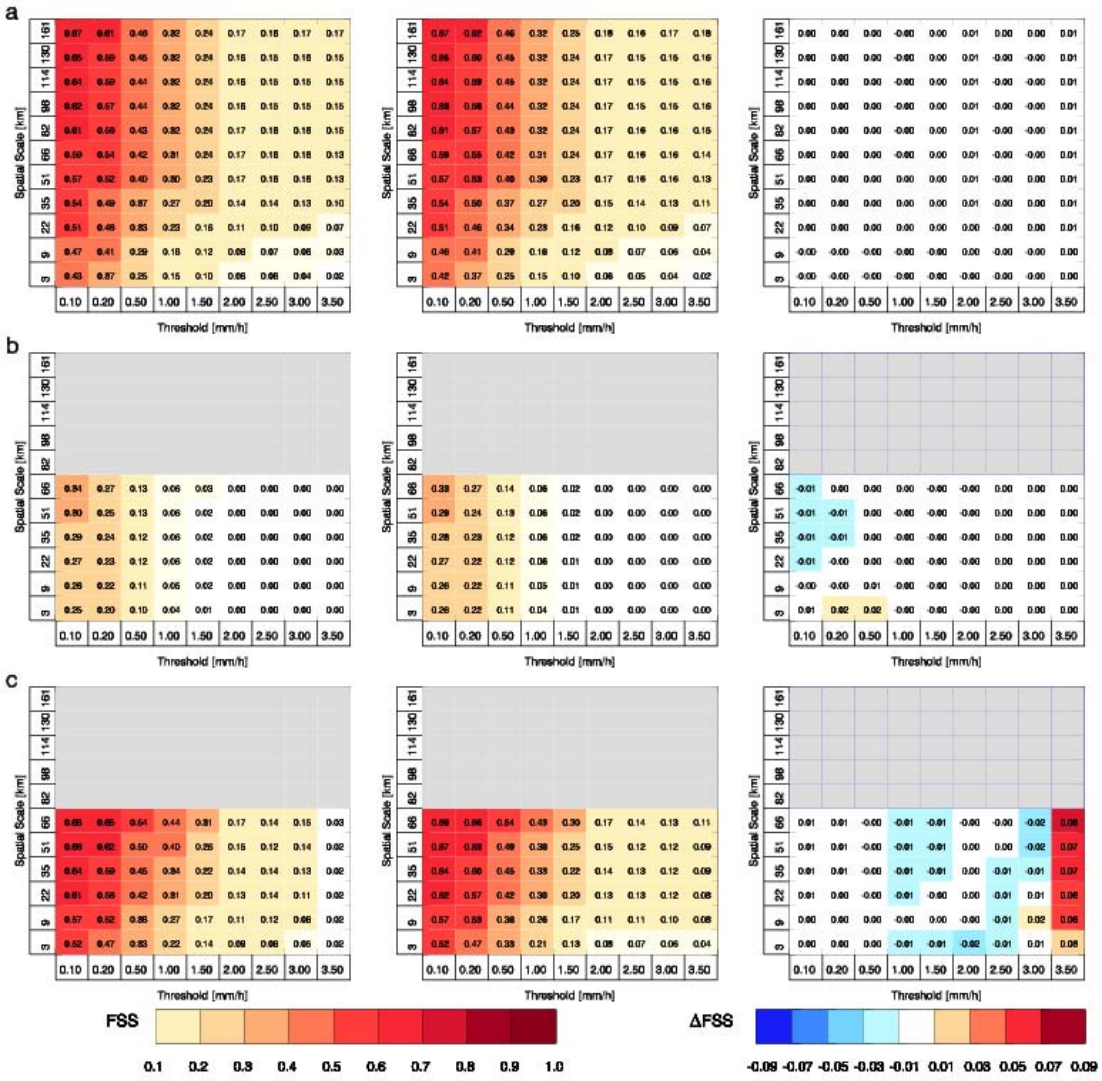


Figure A.46: Same as in Figure A.41 but for the MM5 simulations in winter 2007/08.

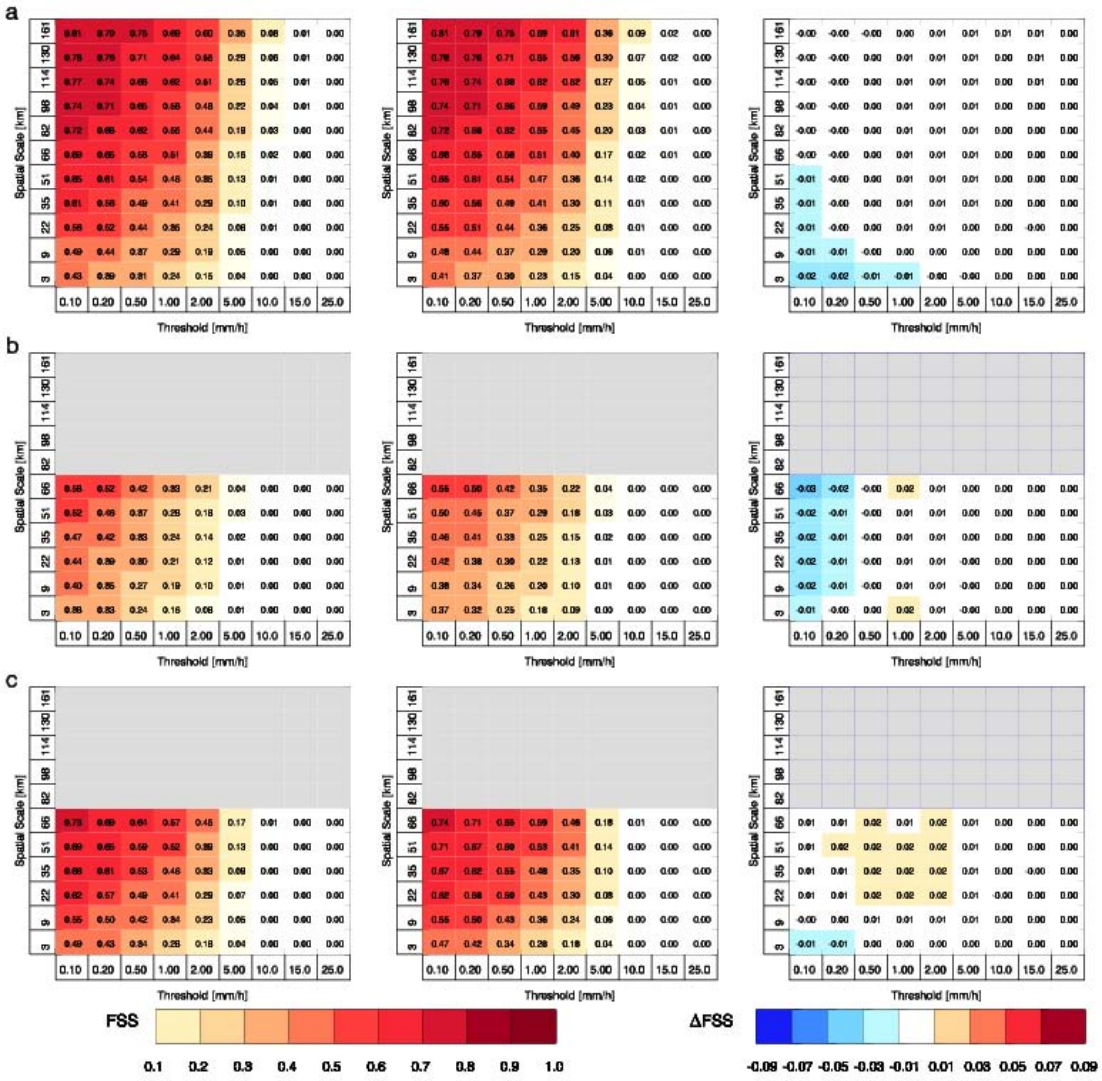


Figure A.47: Same as in Figure A.41 but for the WRF simulations in summer 2007.

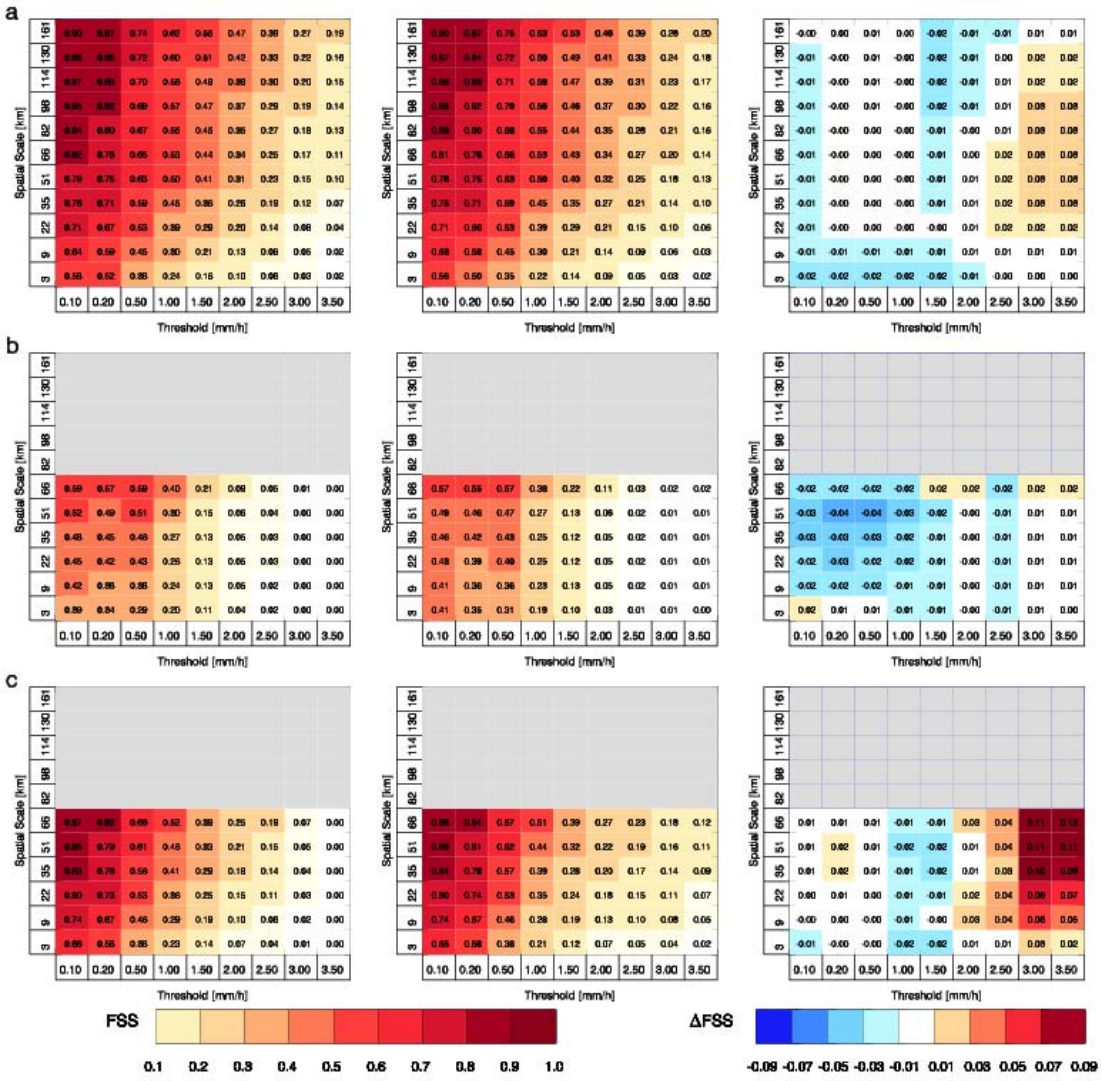


Figure A.48: Same as in Figure A.41 but for the WRF simulations in winter 2007/08.

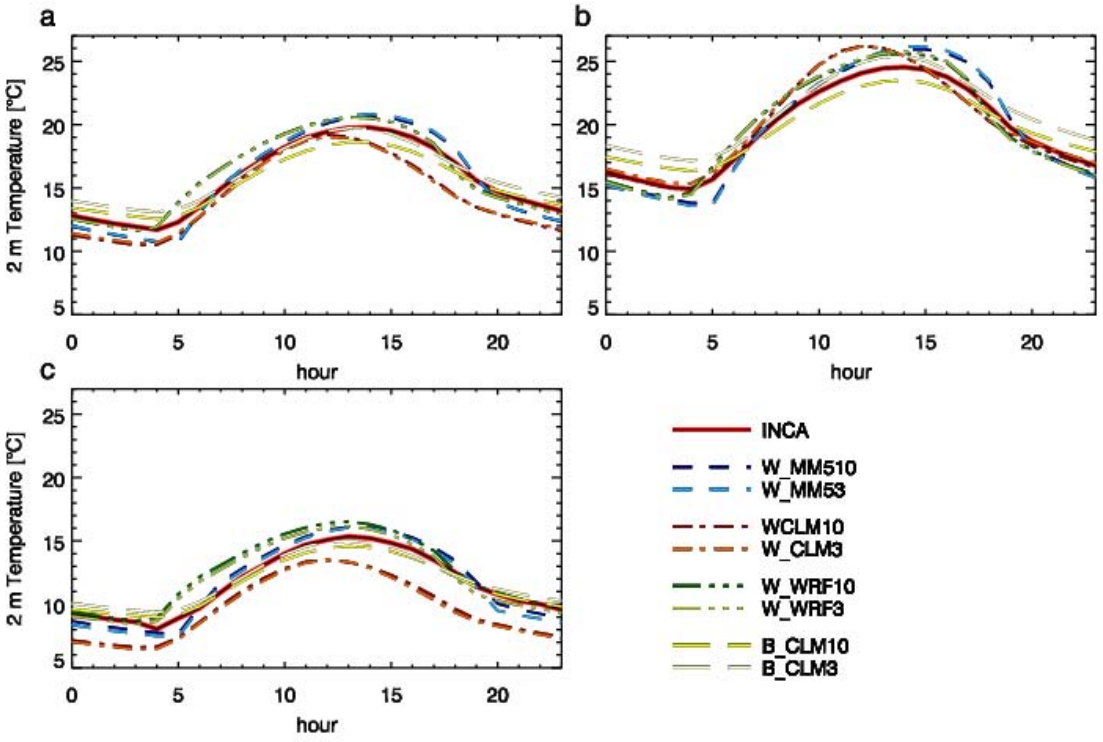


Figure A.49: Seasonal mean diurnal cycles for hourly values of 2 m temperature [°C] in summer 2007. Different RCMs have different colors and line-styles. The 10 km grid spacing simulations have intensive colors, while the 3 km simulations have bright colors. The reference dataset INCA is shown in the background as a black solid red framed line. Results for domain a) D3, b) D4a, and c) D4b are shown.

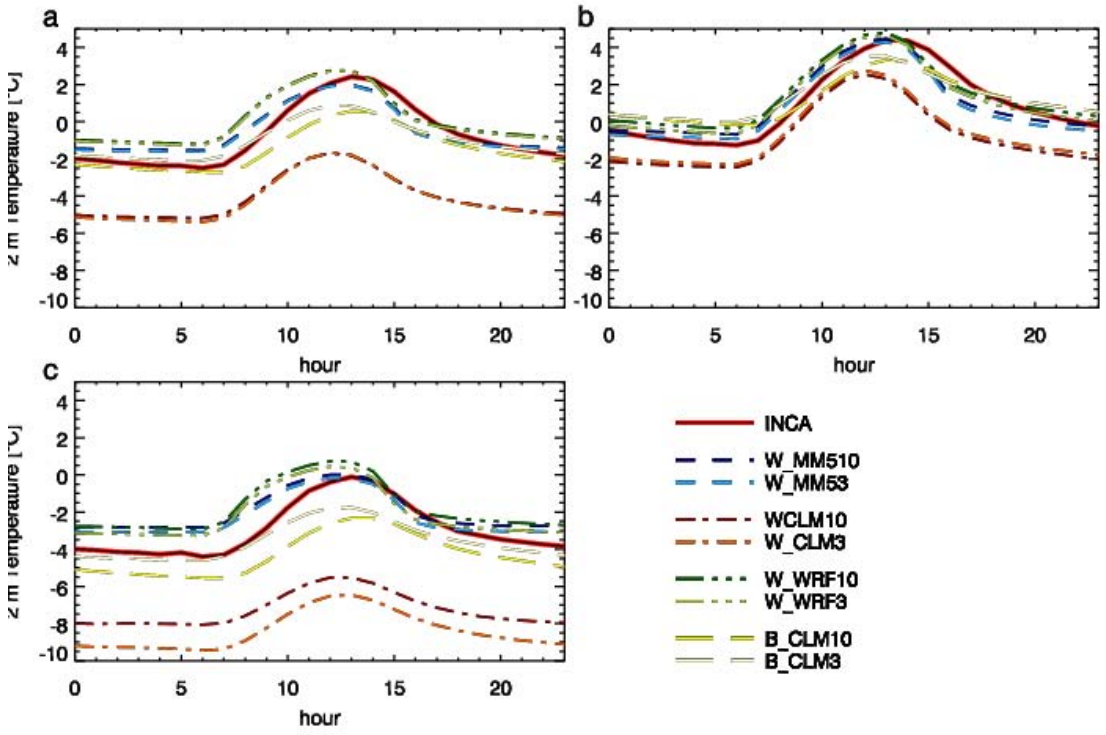


Figure A.50: Same as in Figure A.49 but for 2 m temperature [°C] in winter 2007/08.

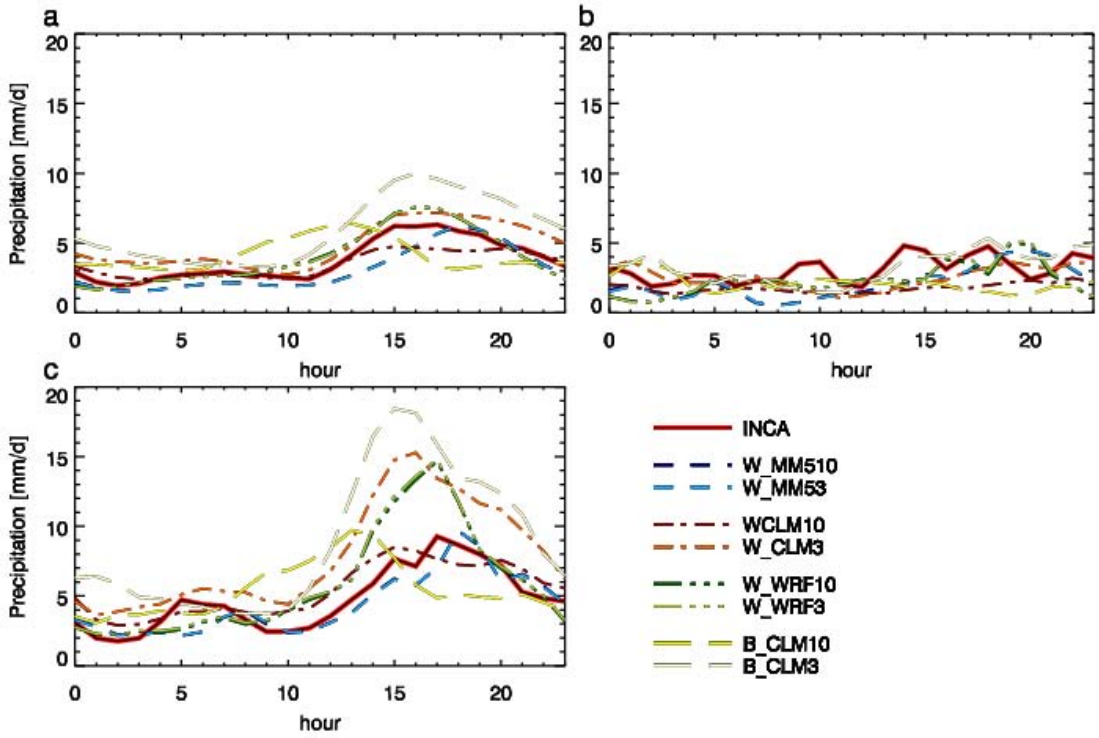


Figure A.51: Same as in Figure A.49 but for precipitation [mm/d] in summer 2007.

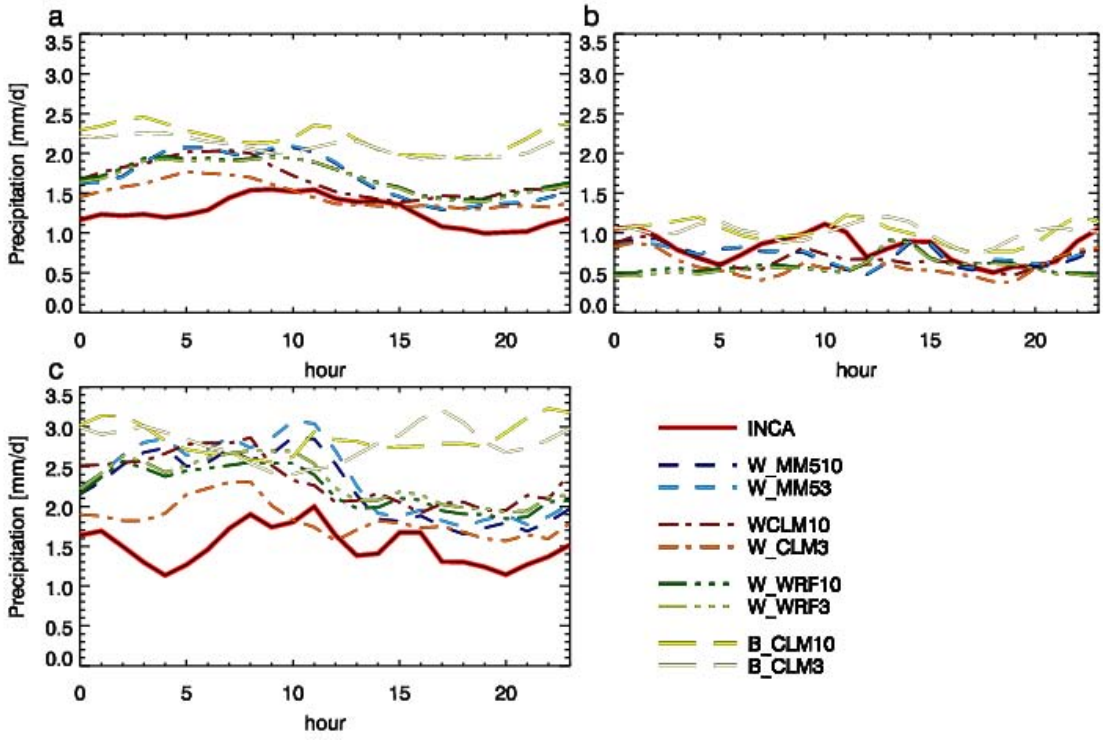


Figure A.52: Same as in Figure A.49 but for precipitation [mm/d] in winter 2007/08.

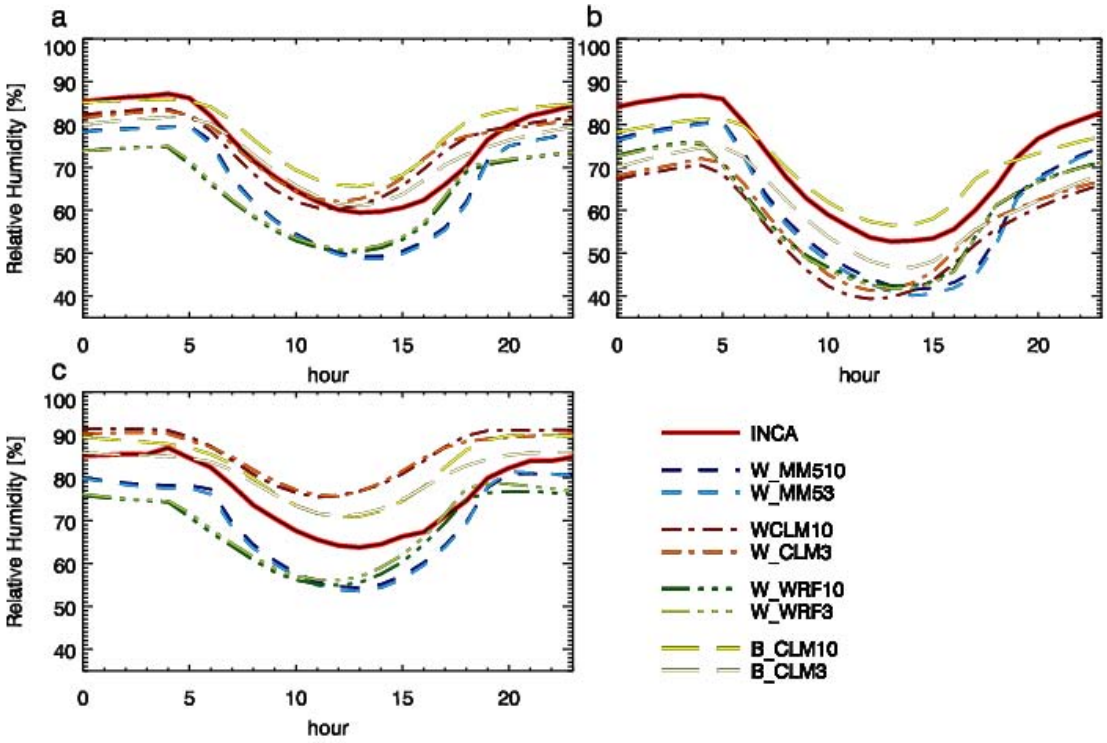


Figure A.53: Same as in Figure A.49 but for relative humidity [%] in summer 2007.

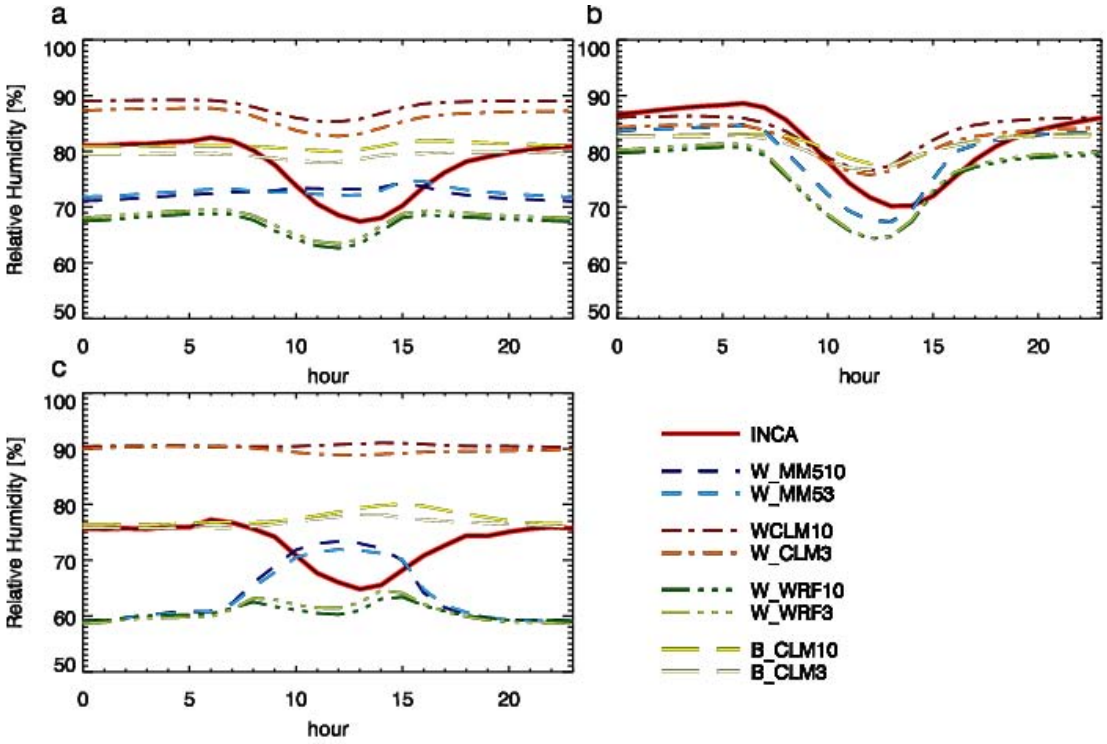


Figure A.54: Same as in Figure A.49 but for relative humidity [%] in winter 2007/08.

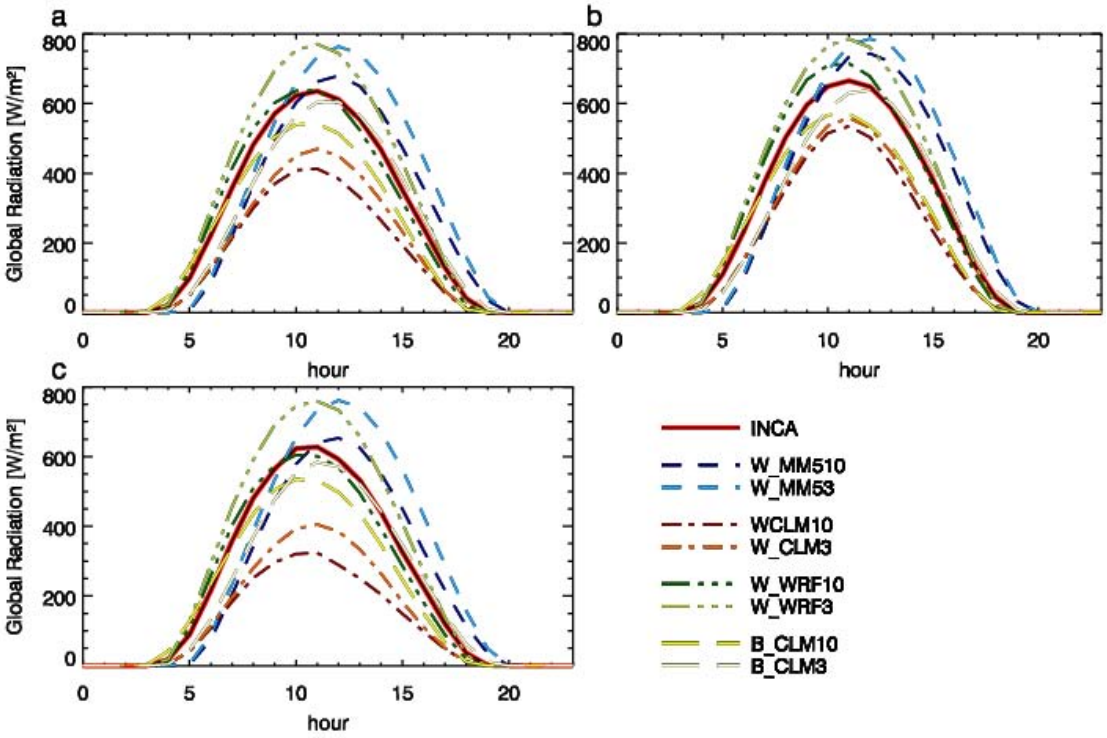


Figure A.55: Same as in Figure A.49 but for global radiation [W/m²] in summer 2007.

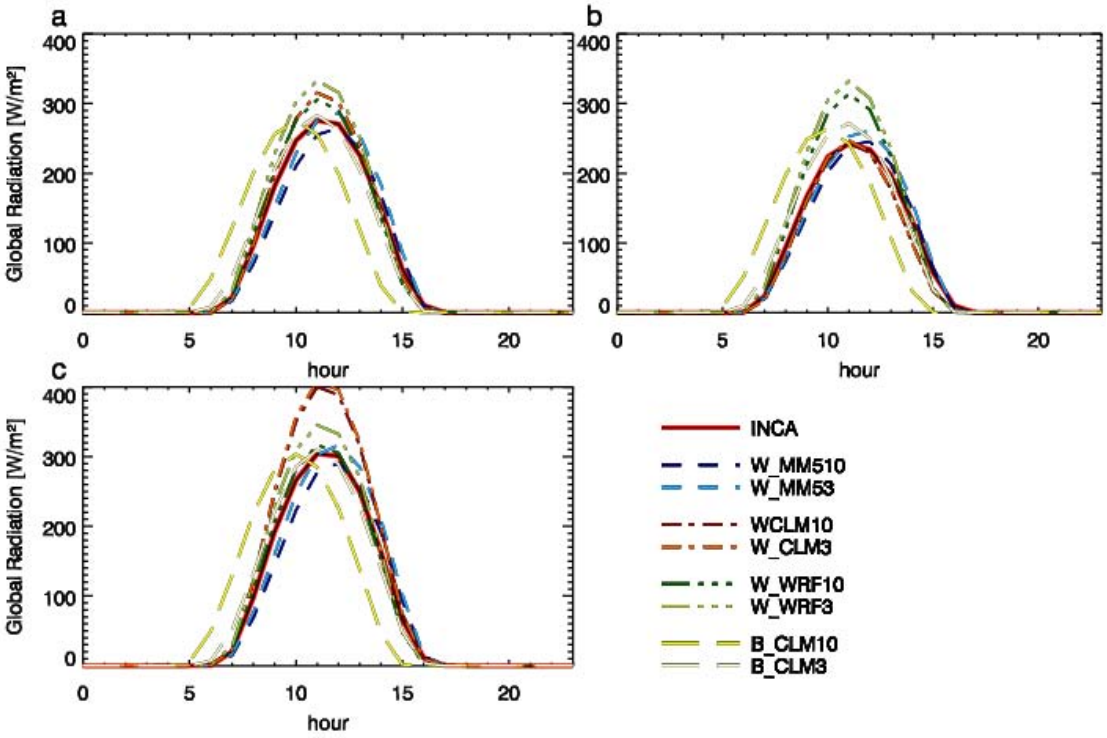


Figure A.56: Same as in Figure A.49 but for global radiation [W/m²] in winter 2007/08.

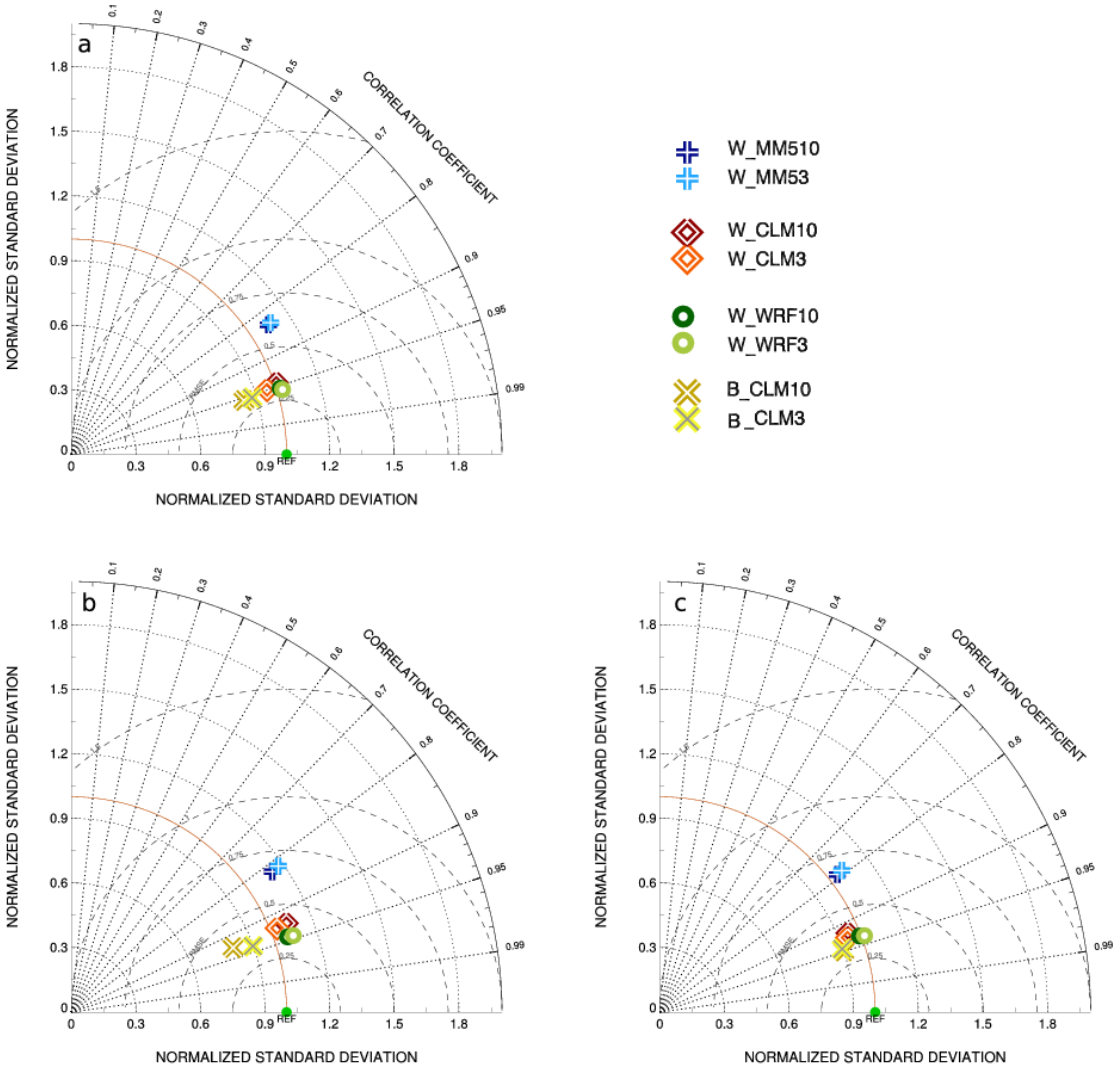


Figure A.57: Taylor diagram for spatially mean hourly temporal values for 2 m temperature in summer 2007. Different RCMs have different colors and symbols. The 10 km grid spacing simulations have intensive colors, while the 3 km simulations have bright colors. In panel a the results on domain D3, in panel b those on D4a, and in panel c those on domain D4b are presented.

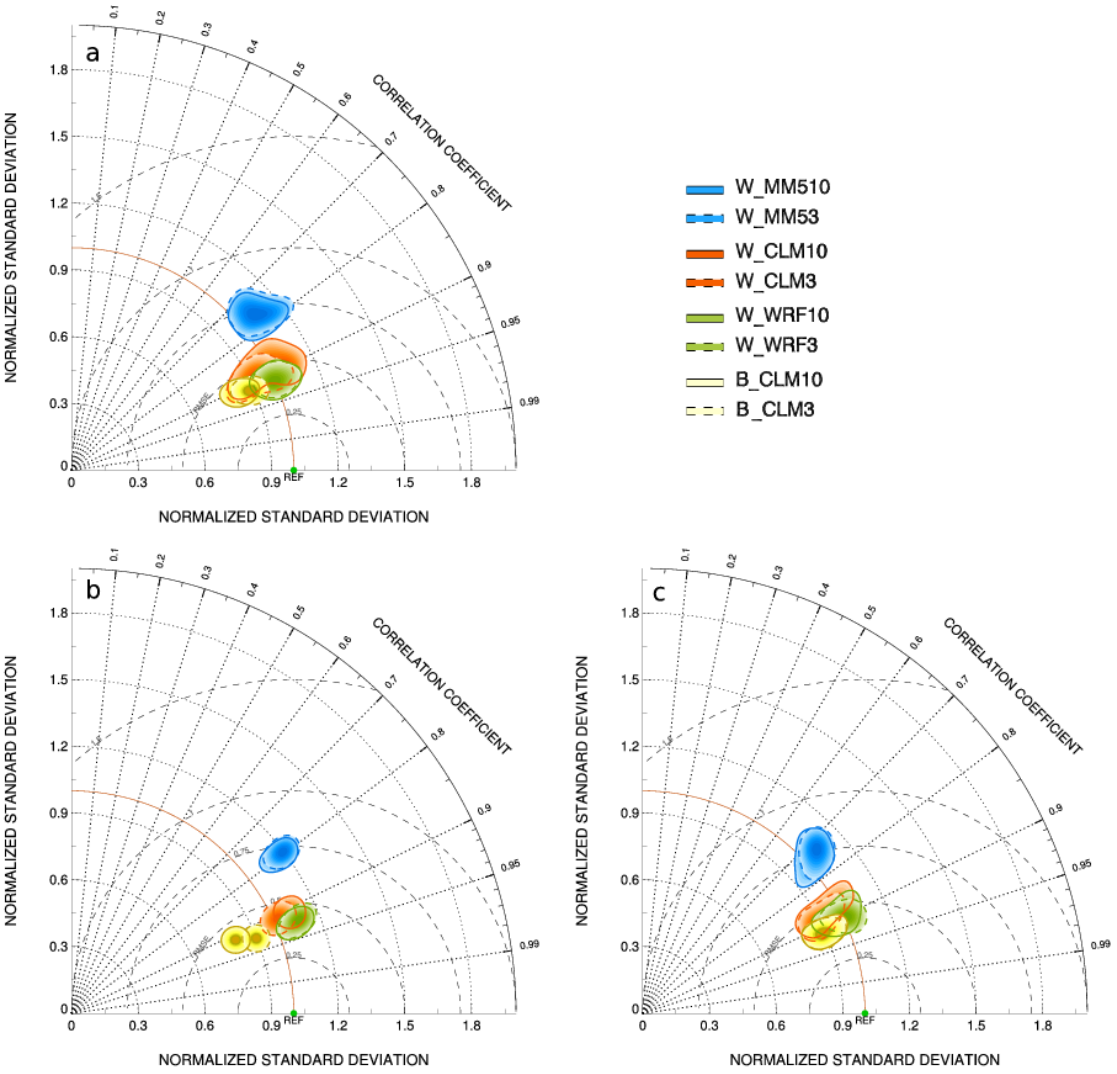


Figure A.58: Taylor diagram for grid-point wise hourly temporal values for 2 m temperature in summer 2007. Different RCMs have different colors. The 10 km grid spacing simulations have solid contour lines while the 3 km simulations have dashed lines. In panel a the results on domain D3, in panel b those on D4a, and in panel c those on domain D4b are presented.

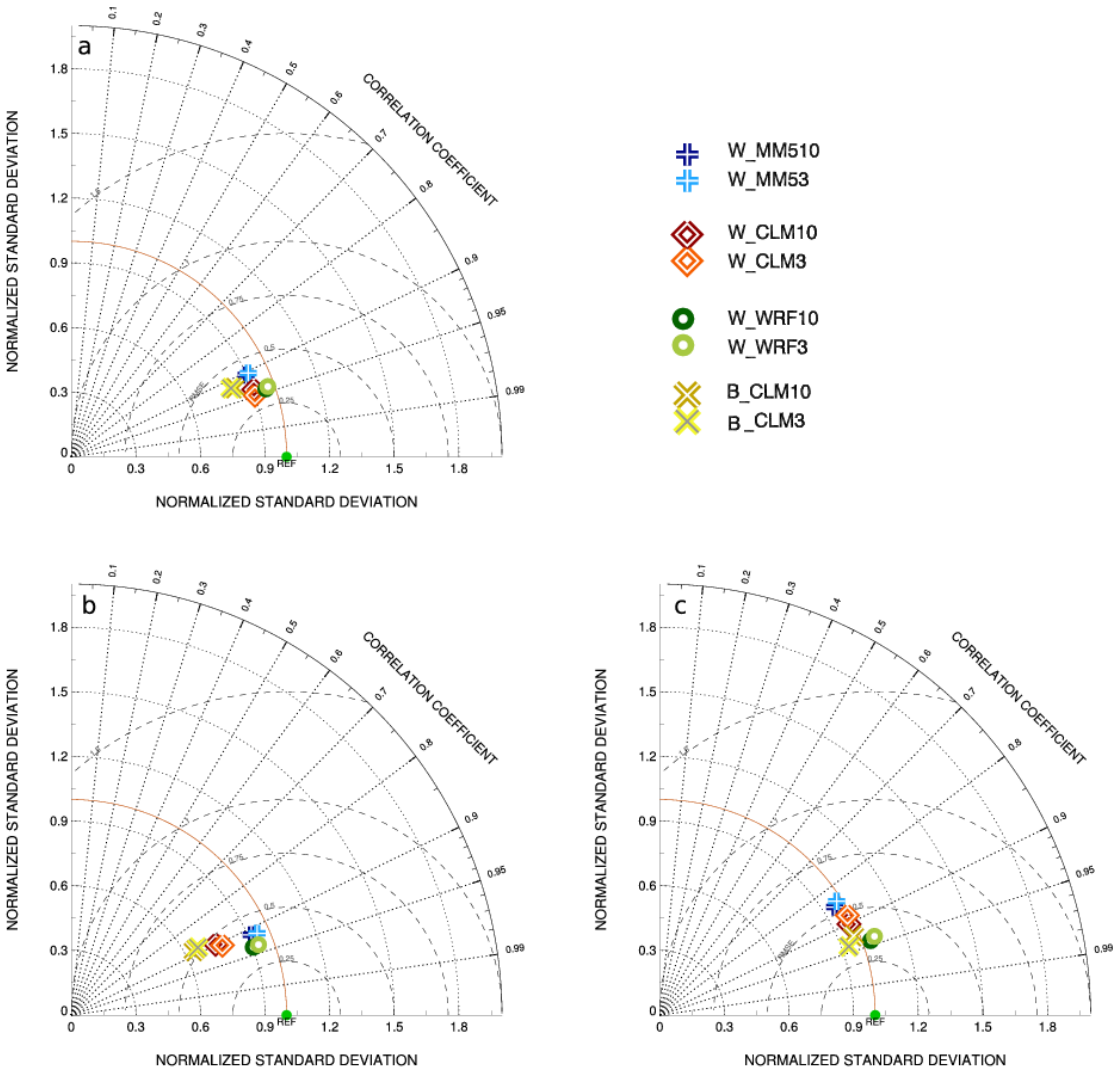


Figure A.59: Same as in Figure A.57 but for 2 m temperature in winter 2007/08.

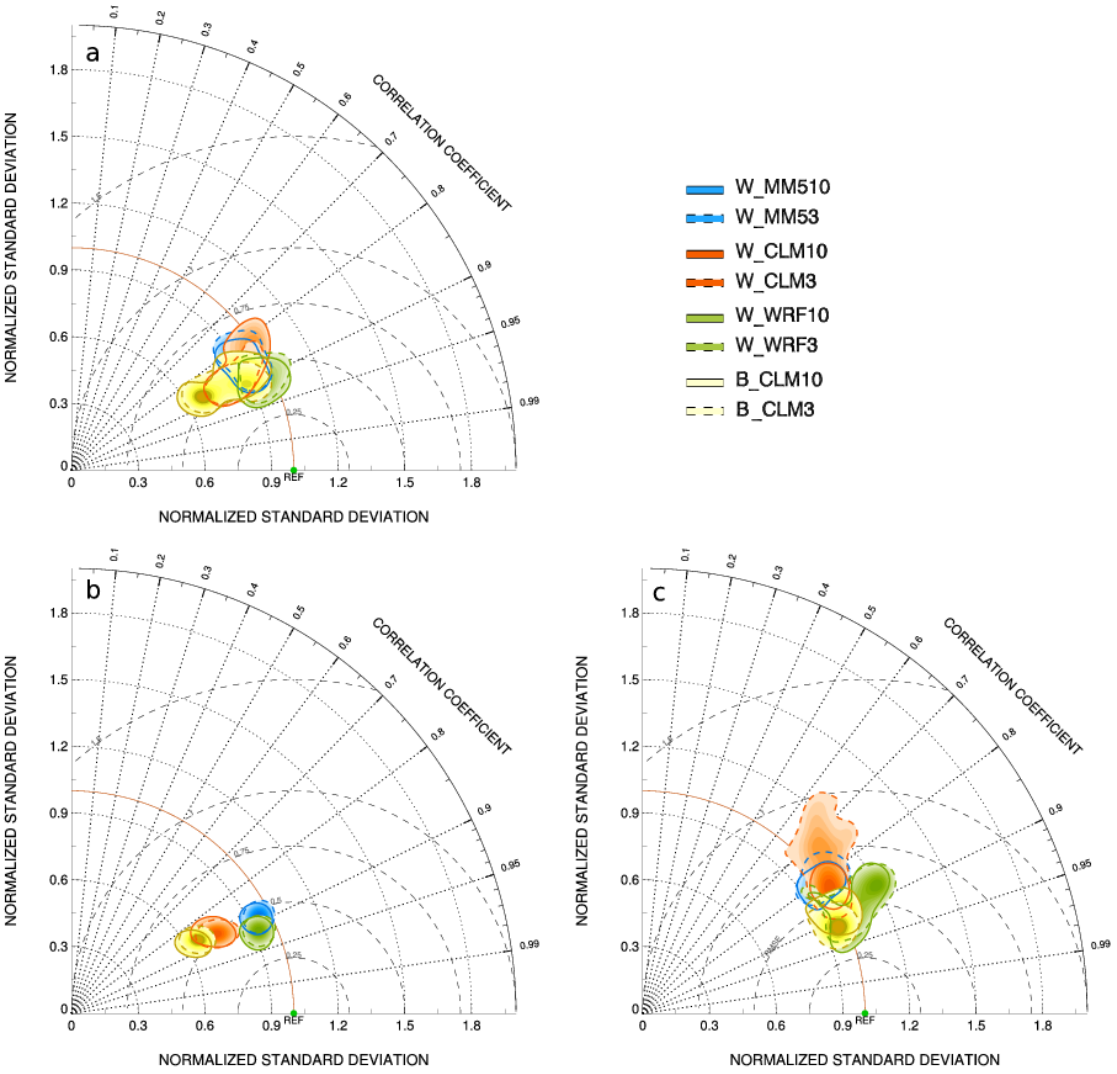


Figure A.60: Same as in Figure A.58 but for 2 m temperature in winter 2007/08.

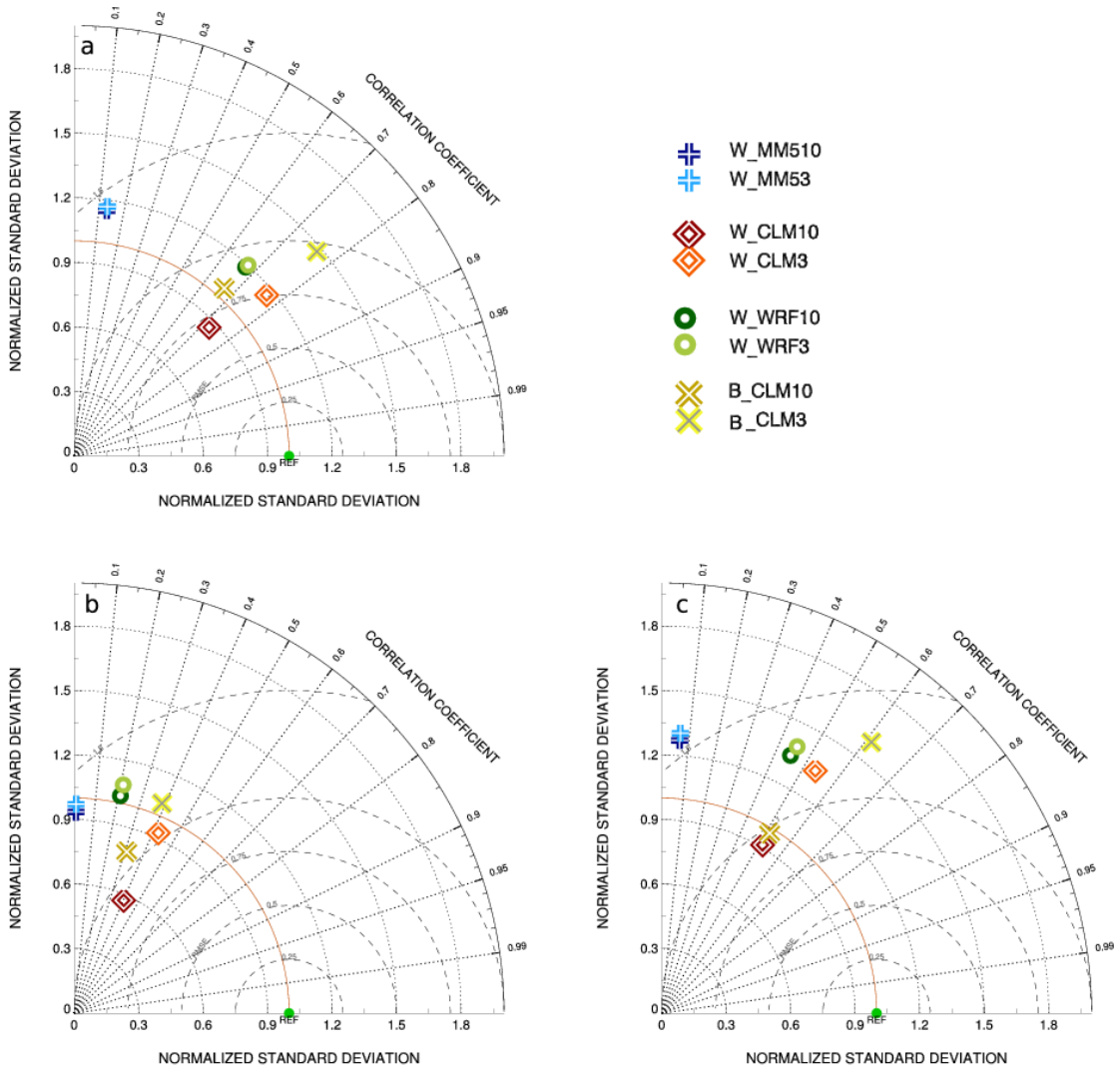


Figure A.61: Same as in Figure A.57 but for precipitation in summer 2007.

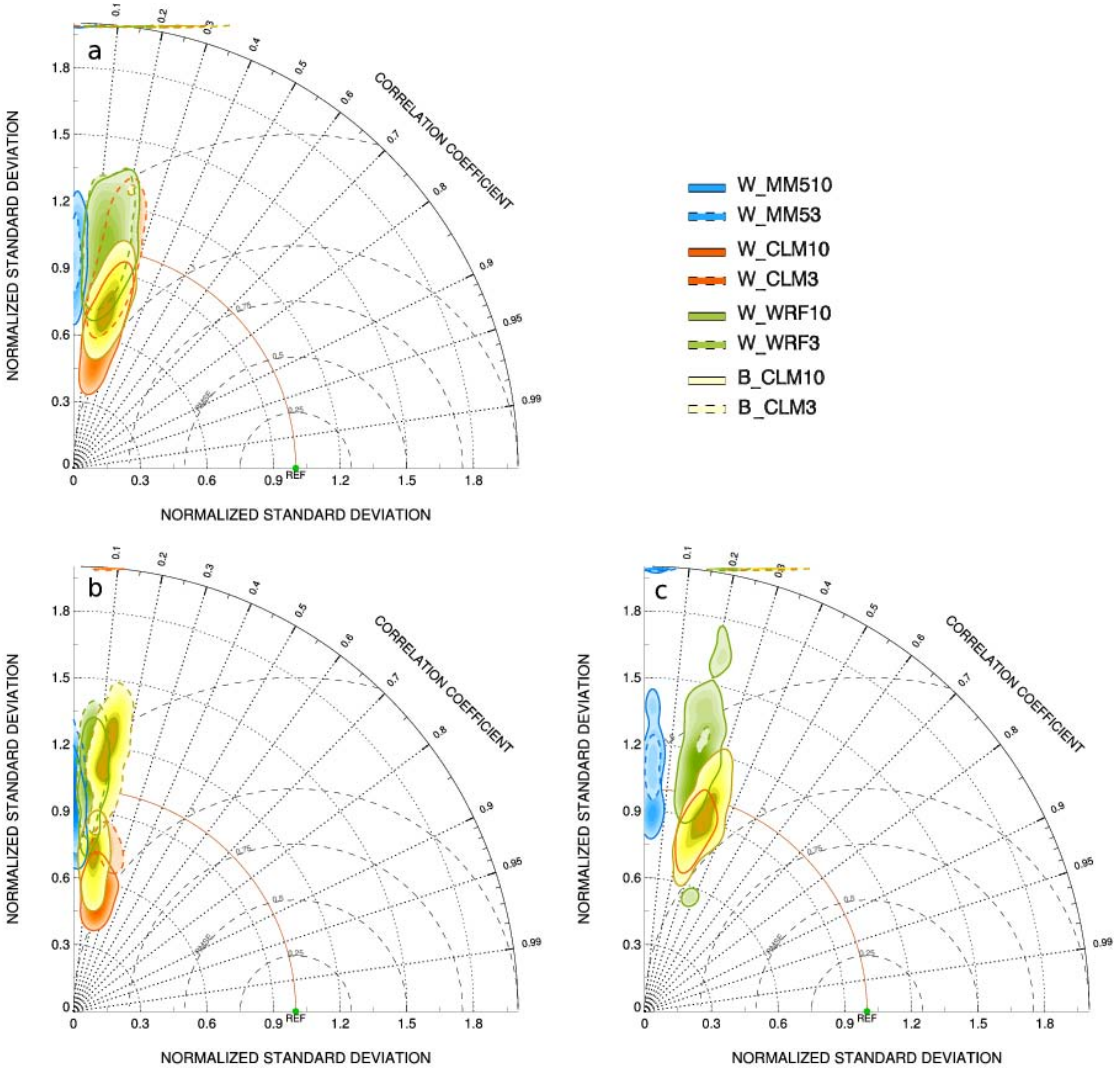


Figure A.62: Same as in Figure A.58 but for precipitation in summer 2007.

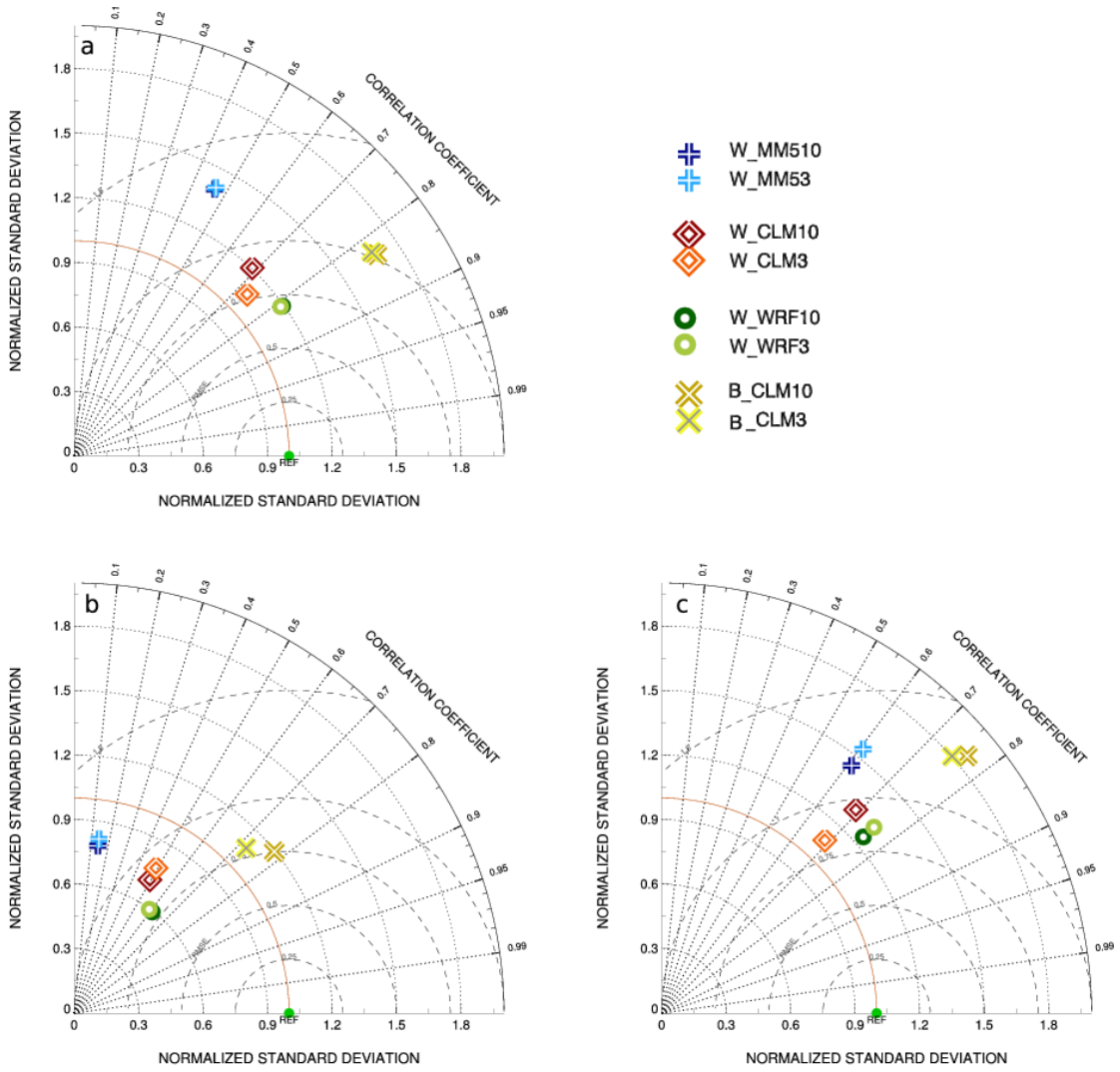


Figure A.63: Same as in Figure A.57 but for precipitation in winter 2007/08.

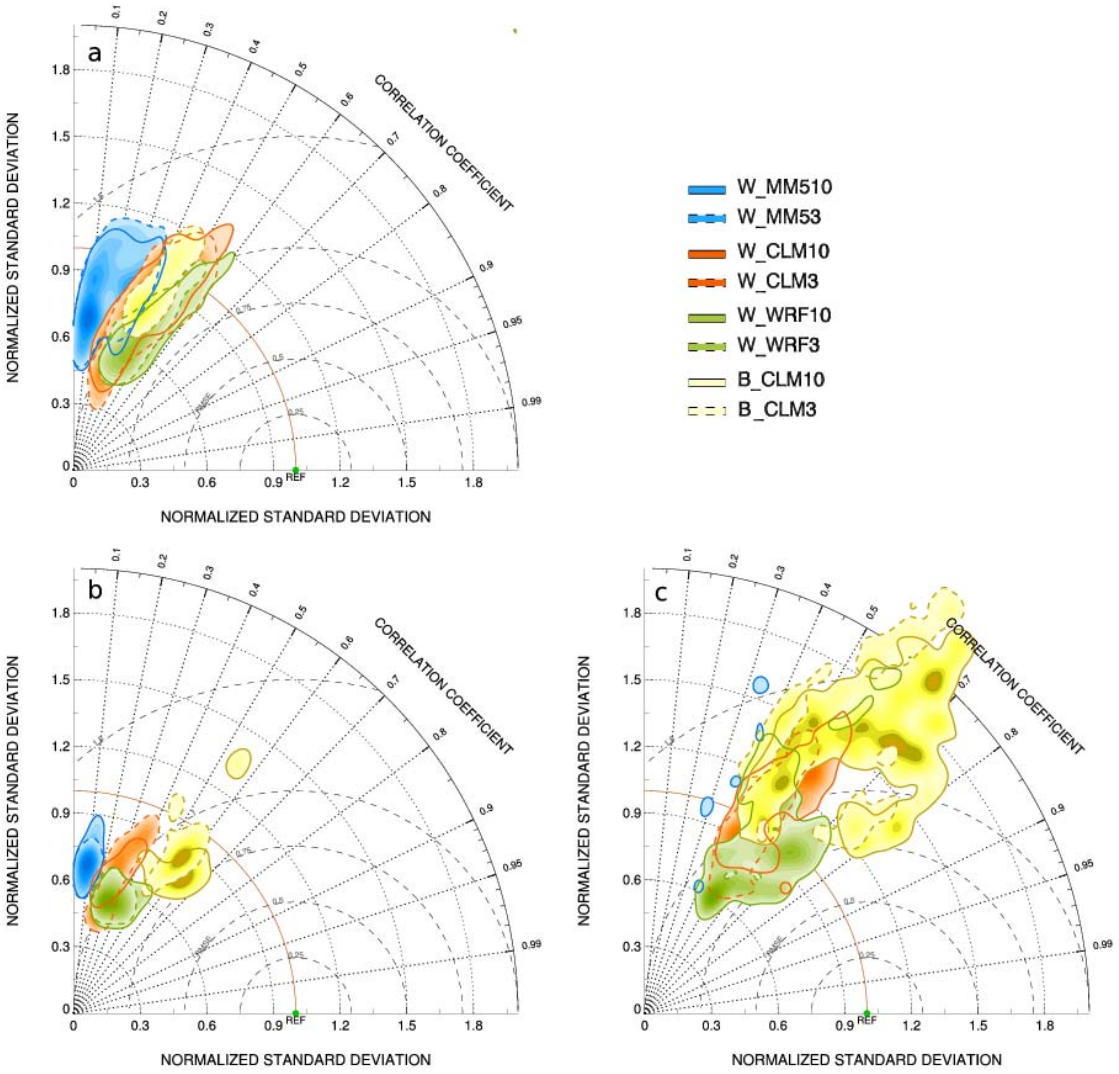


Figure A.64: Same as in Figure A.58 but for precipitation in winter 2007/08.

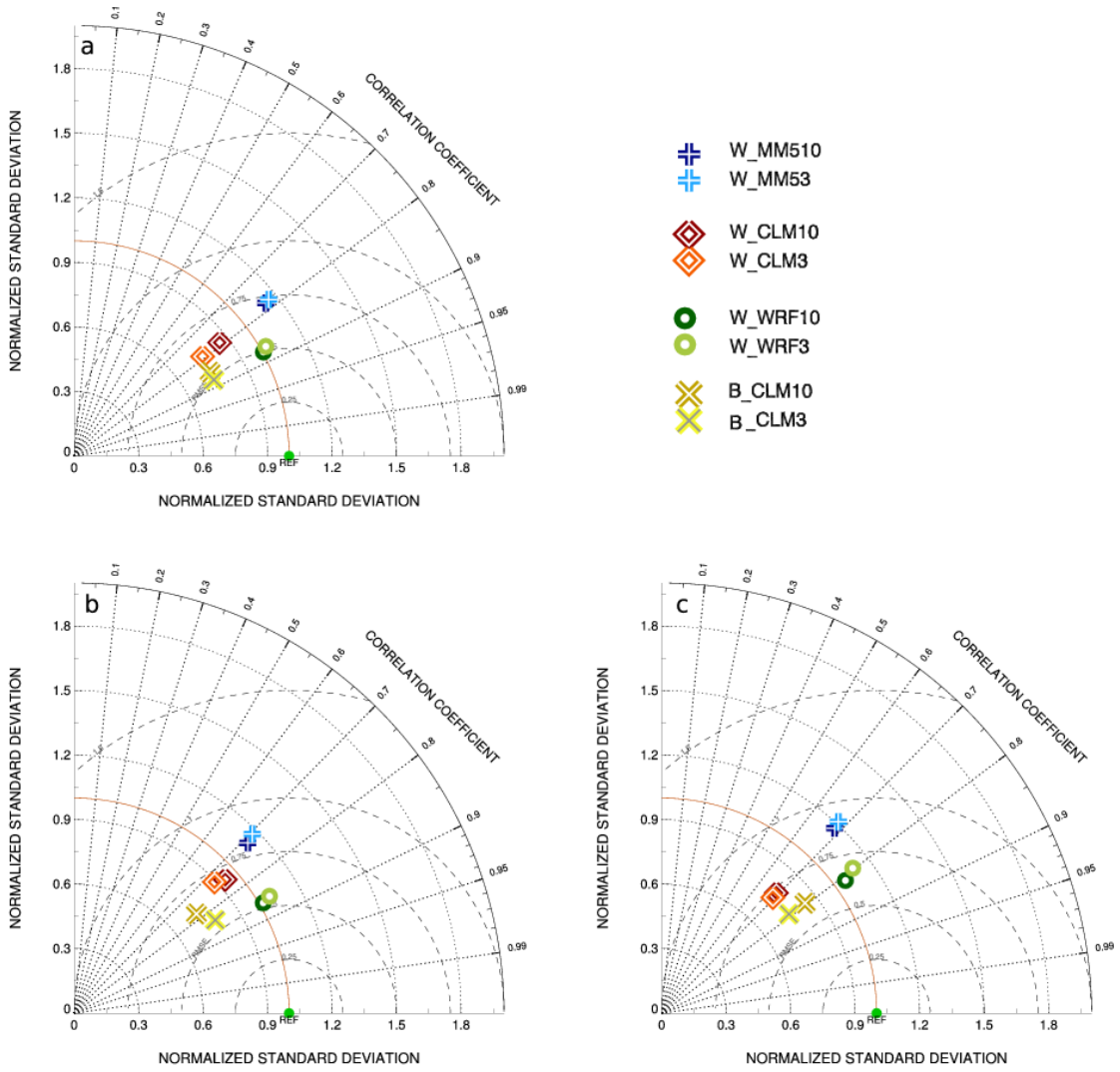


Figure A.65: Same as in Figure A.57 but for relative humidity in summer 2007.

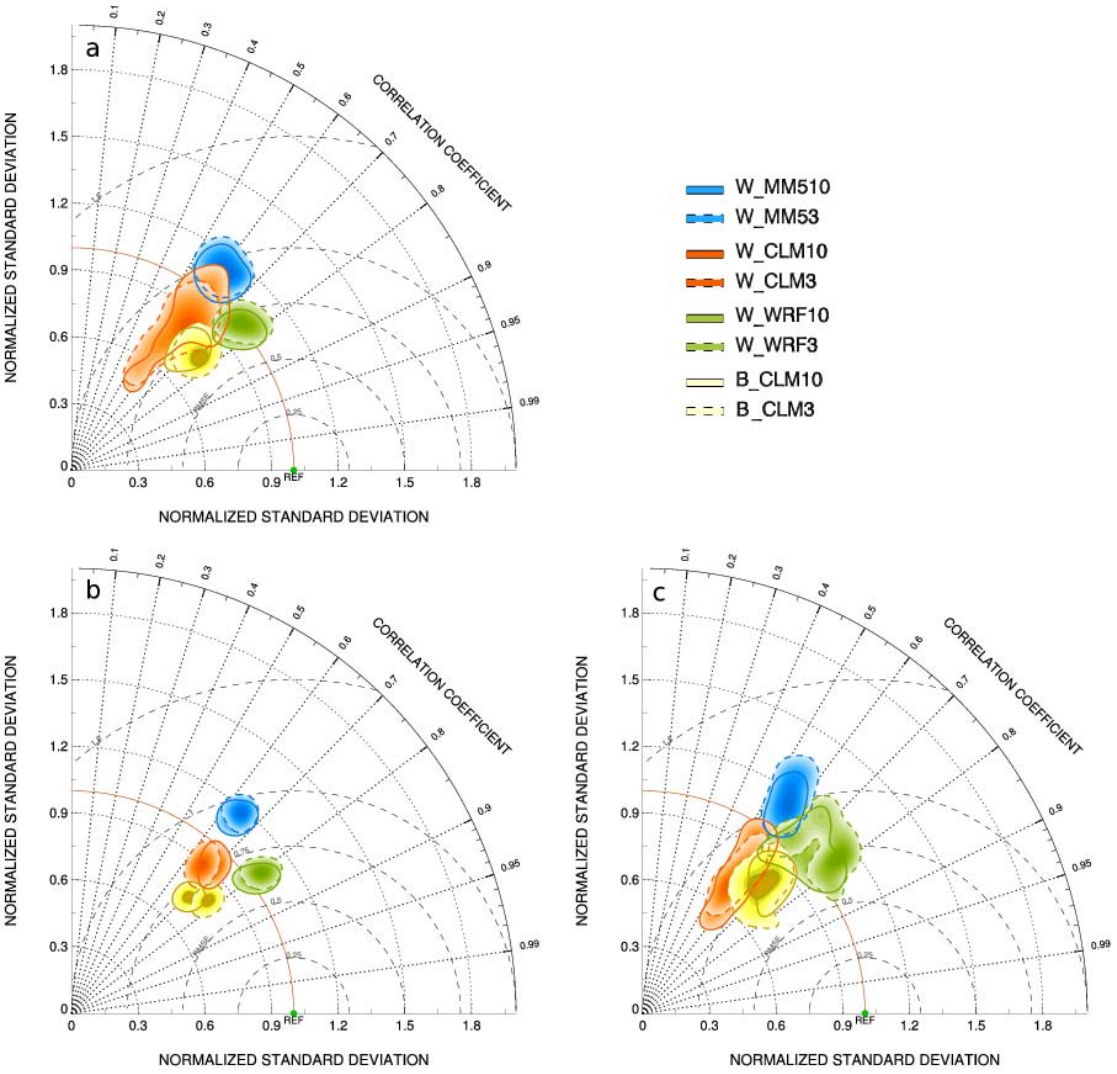


Figure A.66: Same as in Figure A.58 but for relative humidity in summer 2007.

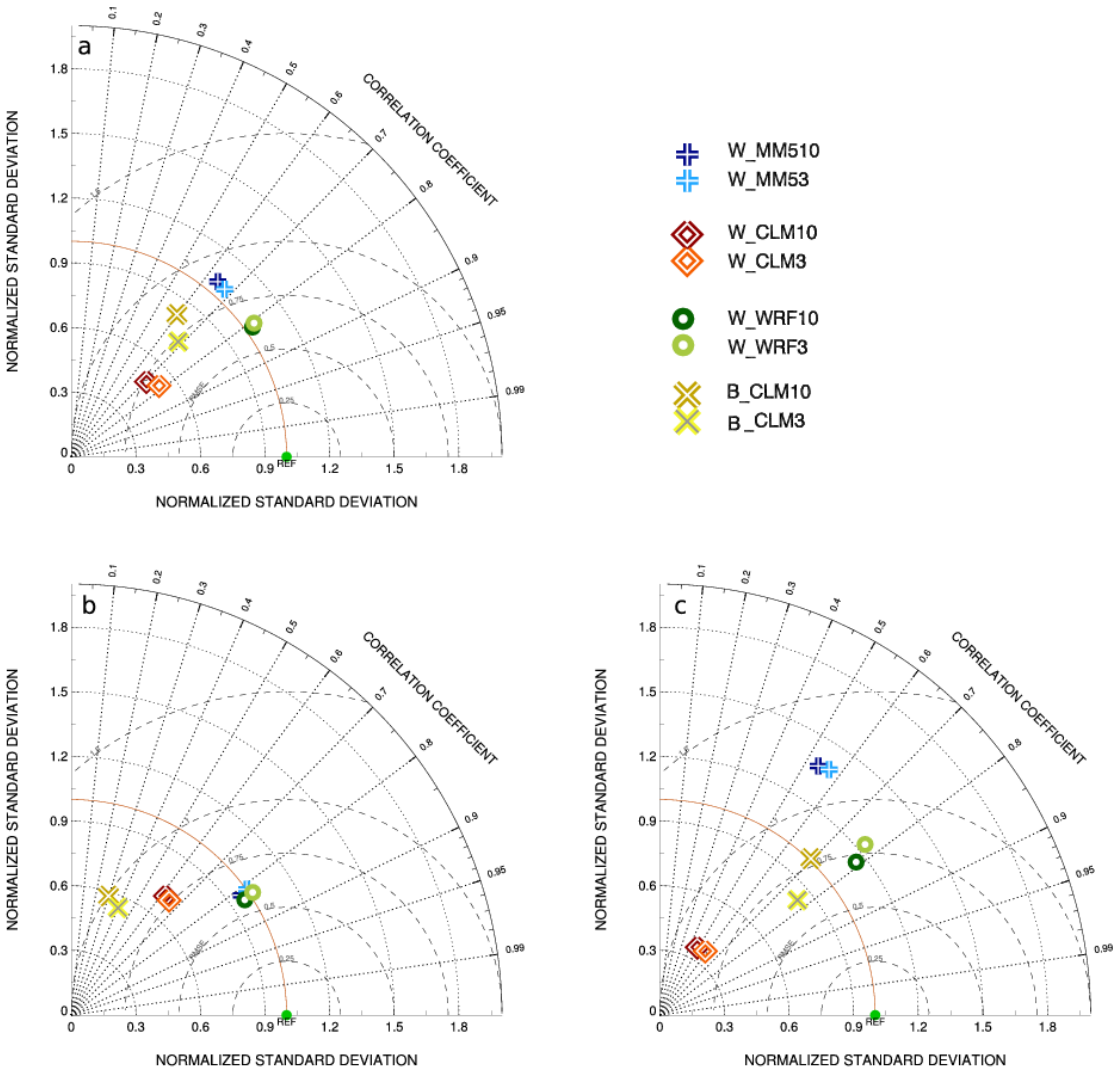


Figure A.67: Same as in Figure A.57 but for relative humidity in winter 2007/08.

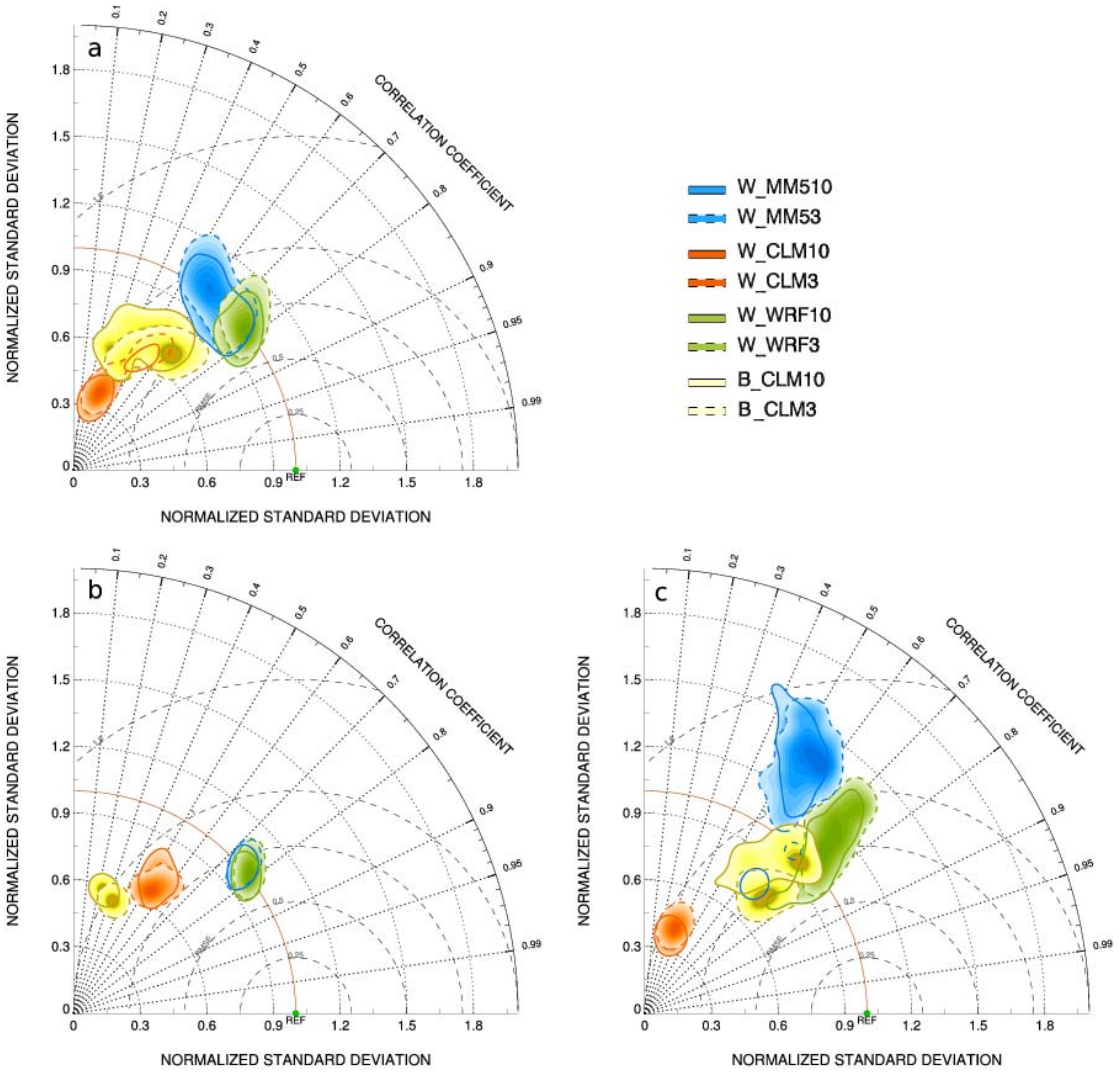


Figure A.68: Same as in Figure A.58 but for relative humidity in winter 2007/08.

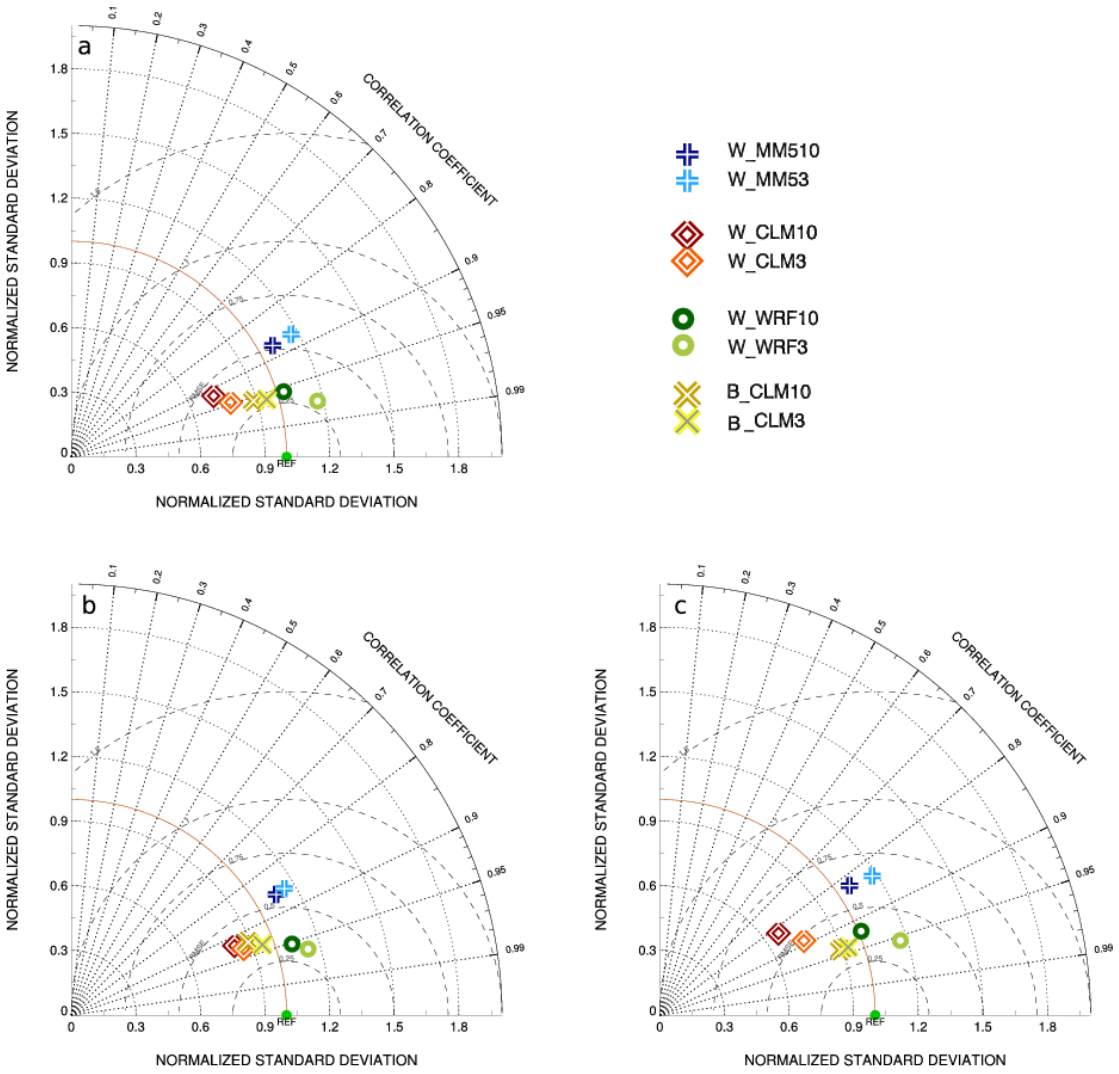


Figure A.69: Same as in Figure A.57 but for global radiation in summer 2007.

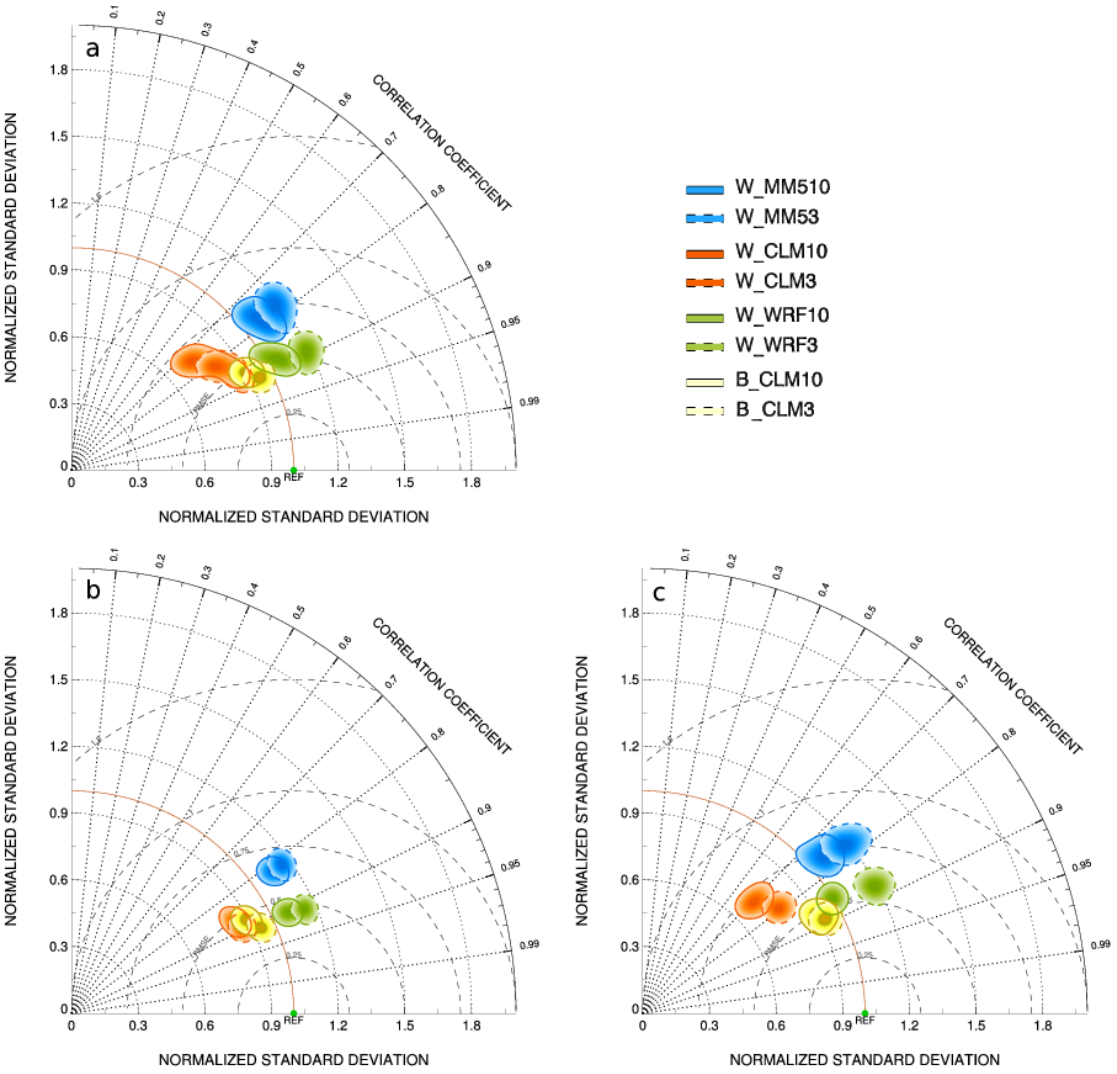


Figure A.70: Same as in Figure A.58 but for global radiation in summer 2007.

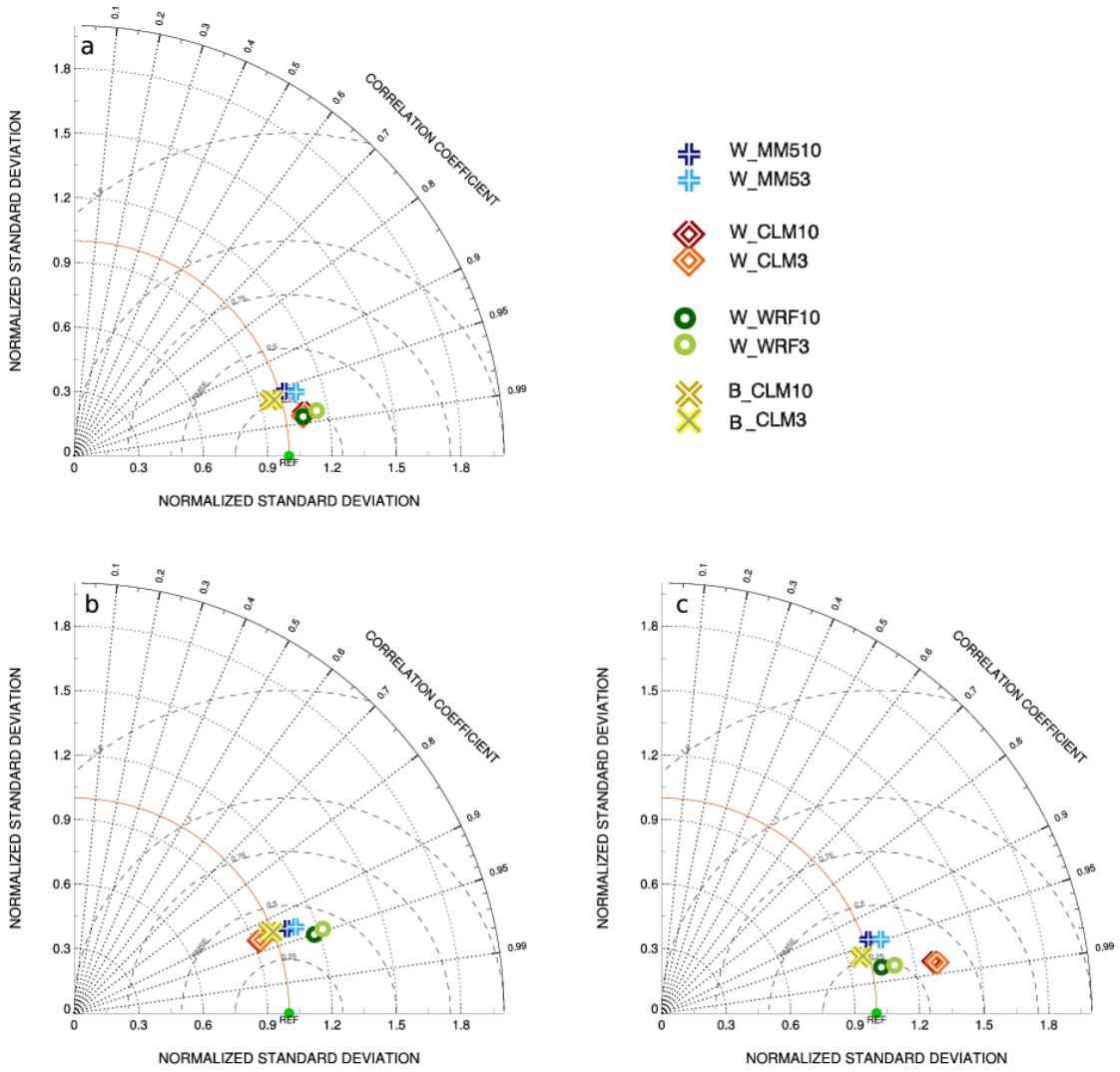


Figure A.71: Same as in Figure A.57 but for global radiation in winter 2007/08.

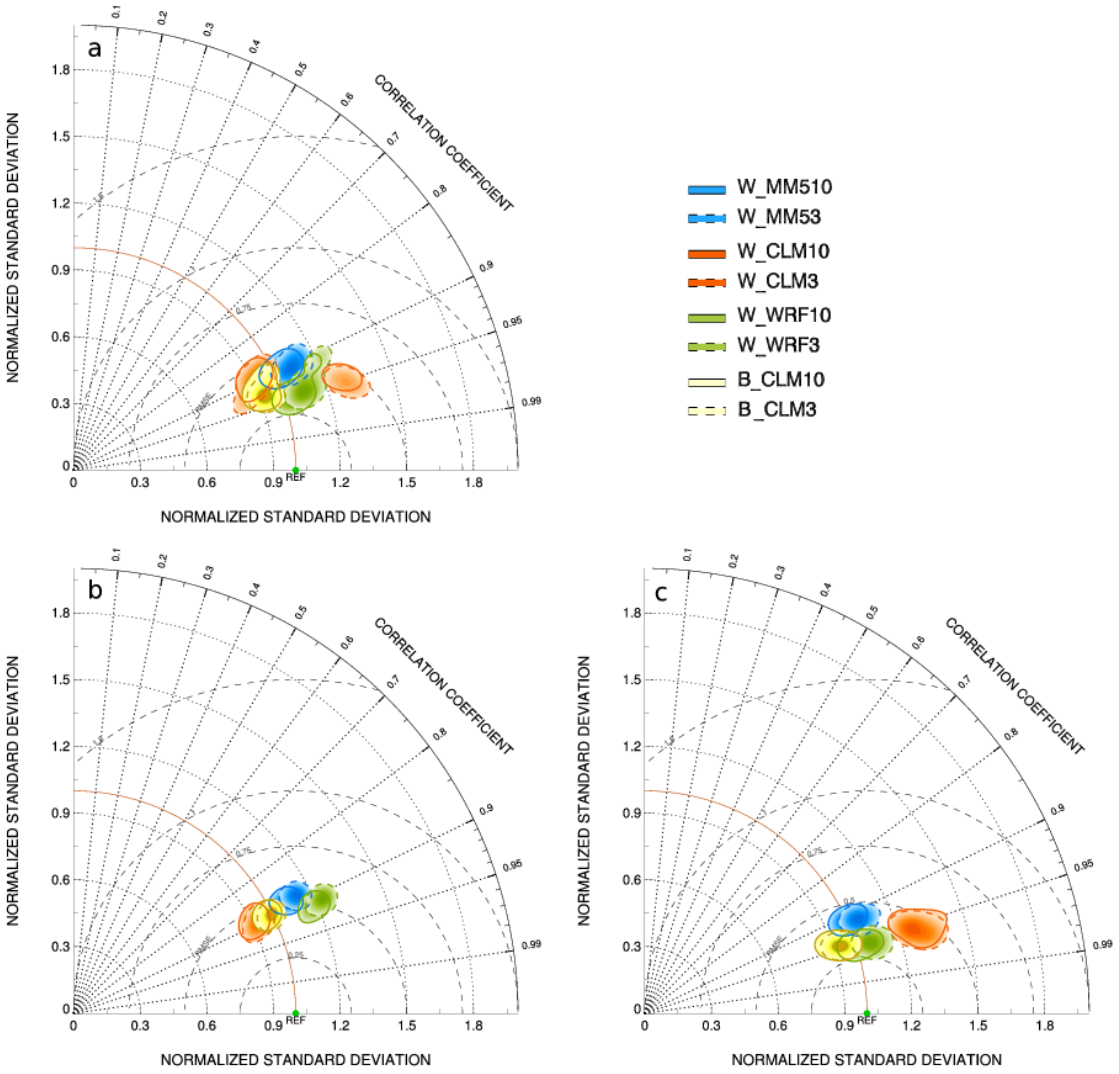


Figure A.72: Same as in Figure A.58 but for global radiation in winter 2007/08.

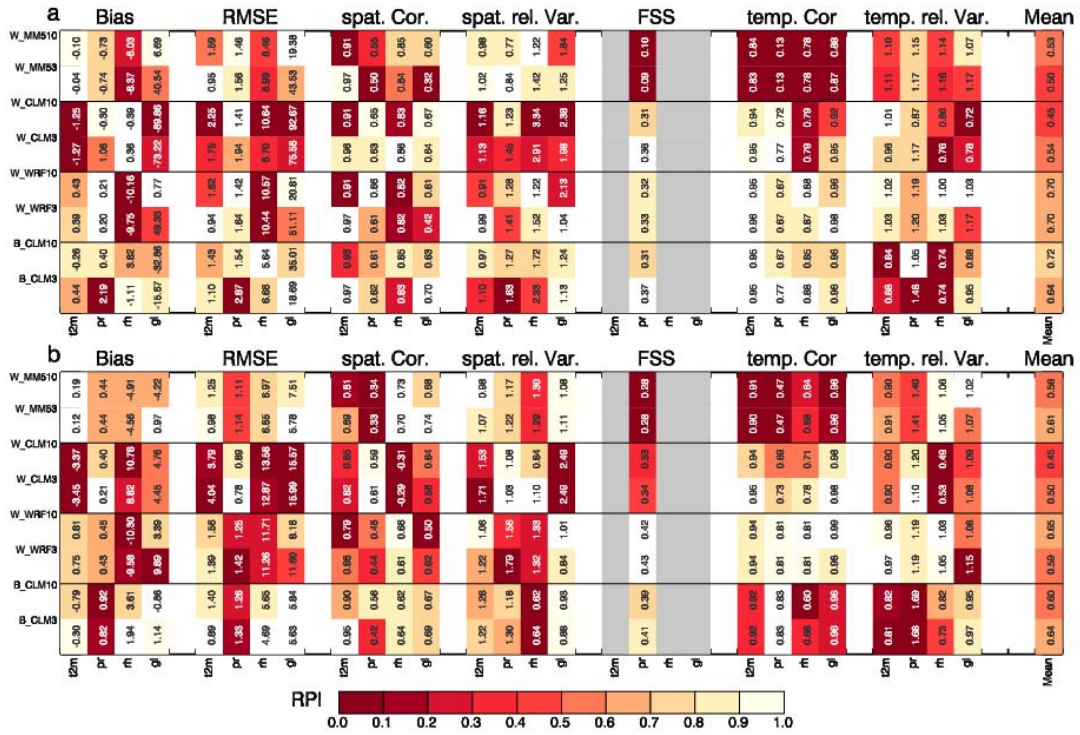


Figure A.73: RPI values for seven statistical values (Bias, spatial mean RMSE, spatial correlation (spat. Cor.)spatial relative standard deviation (spat. rel. Var.), FSS, temporal correlation (temp. Cor.), and temporal relative standard deviation (temp. rel. Var.) as the overall mean for all simulations and all atmospheric parameters. The RPI is shown as color corresponding to the legend. RPI values close to one are shown in white while in dark red worse performing models with RPIs close to zero are shown. The Numbers in the boxes show the absolute statistical value except in the last column (Mean) where the mean RPI is shown. For the bias and the RMSE the values of 2 m temperature (t2m) is shown in K, precipitation (pr) in mm/d, relative humidity (rh) %, and global radiation (gl) in W/m^2. In panel a results for JJA 2007 are presented while in panel b results for DJF 2007/08 are shown.

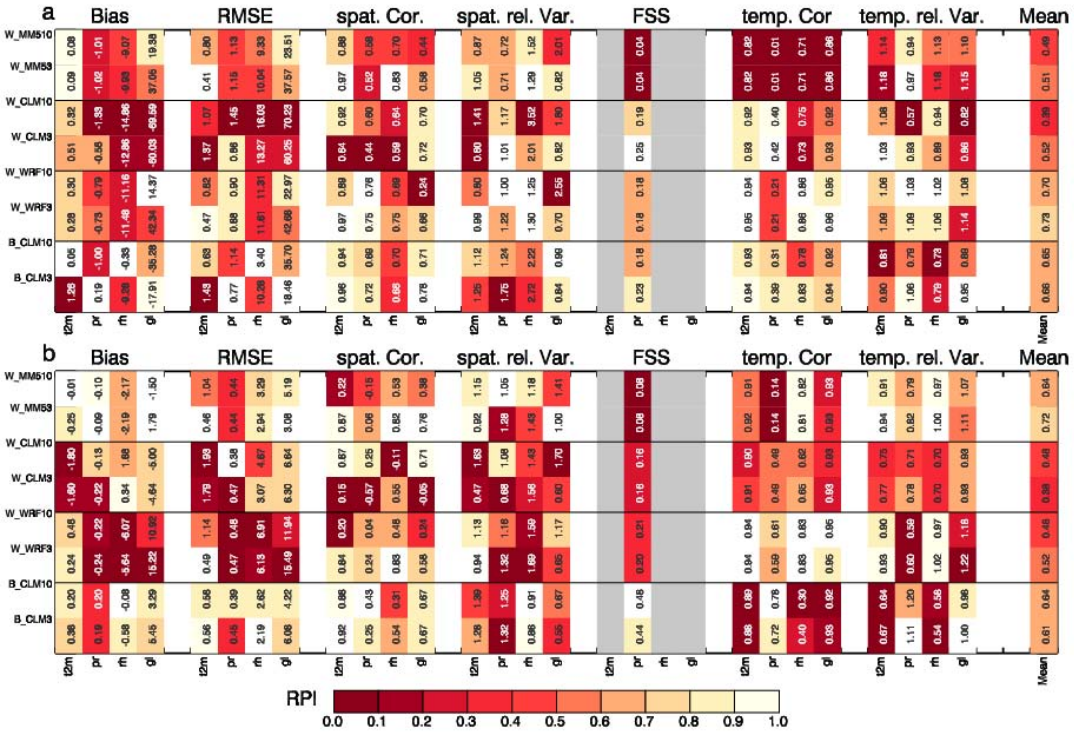


Figure A.74: Same as in Figure A.73 but for domain D4a.

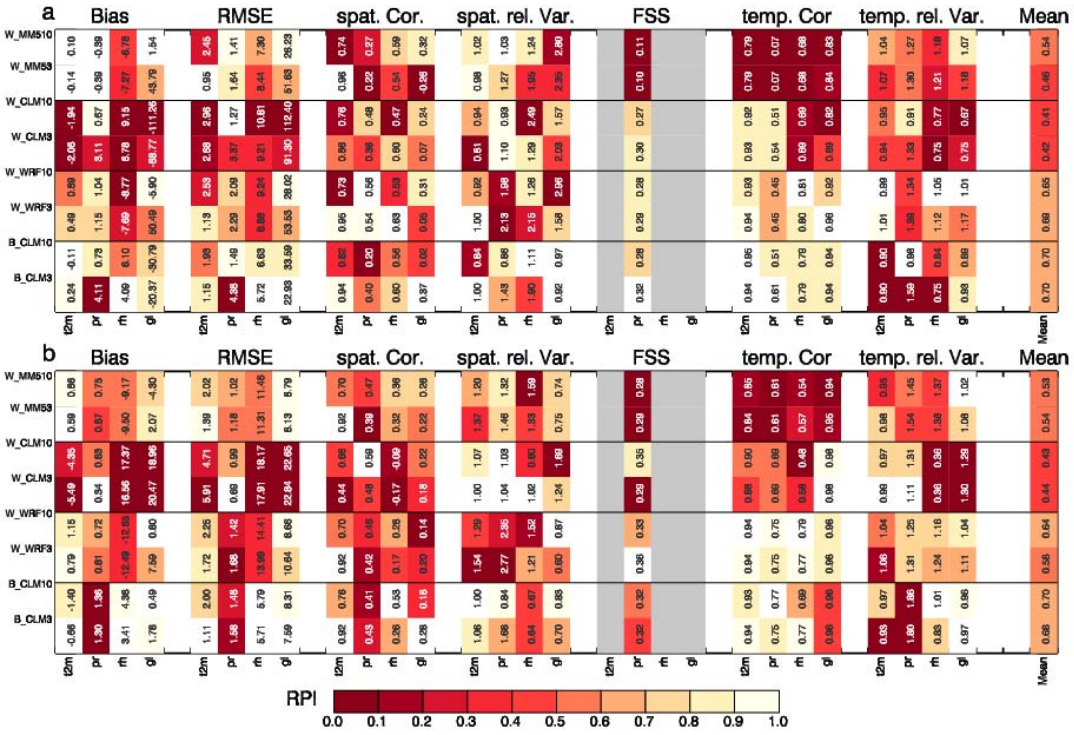


Figure A.75: Same as in Figure A.73 but for domain D4b.

List of Figures

Acronyms

Symbols

NHCM-1 Non-Hydrostatic Climate Modelling. iii, 10, 17

B

BTU Brandenburg University of Technology Cottbus. III, 9, 29, 30, 33, 39–42, 44, 48–55, 58, 59, 103, 104

C

CCLM COSMO model in CLimate Mode. v–vii, III, 9, 17, 18, 27–37, 39–60, 101–104

CRCS convection resolving climate simulation. 60

D

D3 eastern Alpine Region. I–IV, 10–13, 27–36, 38–57, 59, 60, 62, 70, 78, 86, 87, 101, 109, 117, 118

D4a south-east Styria. I–V, 10–12, 14, 27–36, 38–57, 62, 70, 78, 86, 87, 101, 109, 117, 118, 134

D4b Hohe Tauern National Park. I–V, 11, 12, 14, 15, 27–35, 38–56, 58, 59, 70, 78, 86, 87, 101, 109, 117, 118, 135

DJF December, January, and February. v, V, 10, 11, 18, 28, 30, 32, 33, 35, 39–41, 47, 133

E

E-OBS European observation dataset. 11, 14

ECMWF European Centre for Medium-Range Weather Forecasts. 9, 17

F

FSS Fractional Skill Score. v, III, V, 24, 25, 38, 43–47, 57–59, 101, 133

FWF Austrian Science Fund. iii, vii

I

IFS Integrated Forecast System. vii, 9, 17

INCA Integrated Nowcasting through Comprehensive Analysis. vii, I, IV, 10, 14–17, 26, 30, 42, 47, 60, 109

J

JJA June, July, and August. v, I, V, 10, 11, 17, 18, 27, 28, 30, 31, 33–35, 40–43, 47, 48, 59, 62, 133

L

LBC lateral boundary conditions. 17

LocMIP Local Climate Model Intercomparison Project. v, vii, 9, 14, 17, 18, 21, 26, 43, 46, 48, 57, 59, 60

M

MM5 Fifth-Generation Mesoscale Model. v, III, 9, 17, 18, 27–37, 39–42, 45, 47–56, 58–60, 105, 106

N

NWP numerical weather prediction. 24

R

RCM regional climate model. v, II–IV, 9–11, 14, 16–21, 23, 24, 26–32, 34–37, 39–42, 45–48, 50–56, 59, 60, 70, 78, 86, 87, 109, 117, 118

RMSE root mean squared error. v, V, 25, 27, 36, 38, 43, 46, 48, 57–59, 133

RPI relative performance index. V, 24, 25, 57–60, 133

S

SBC surface boundary conditions. 17

W

WegCenter Wegener Center for Climate and Global Change. III, 9, 14, 17, 27, 29, 30, 32–34, 39, 41–44, 48–51, 53–55, 58, 101, 102

WRF Weather Research and Forecasting Model. v, vii, III, 9, 18, 27–30, 32–37, 39–42, 45, 46, 48–55, 57–60, 107, 108

Z

ZAMG Central Institute for Meteorology and Geodynamics. 14

Bibliography

Bechtold, P. et al. (2008). 'Advances in simulating atmospheric variability with the ECMWF model: From synoptic to decadal time-scales'.
In: *Quart. J. Roy. Met. Soc.* 134 (634 2008), 1337–1351. DOI: 10.1002/qj.289.
See p. 10.

Chen, F. and J. Dudhia (2001).
'Coupling an Advanced Land Surface-Hydrology Model with the Penn State-NCAR MM5 Modeling System. Part I: Model Implementation and Sensitivity'.
In: *Mon. Wea. Rev.* 129.4, pp. 569–585. See pp. 19–21.

Christensen, J. and O. Christensen (2007). 'A summary of the PRUDENCE model projections of changes in European climate by the end of this century'.
In: *Clim. Change* 81 (0 2007). 10.1007/s10584-006-9210-7, pp. 7–30.
URL: http://dx.doi.org/10.1007/s10584-006-9210-7. See p. 9.

Déqué, M. et al. (2007). 'An intercomparison of regional climate simulations for Europe: assessing uncertainties in model projections'. In: *Clim. Change* 81 (0 2007). 10.1007/s10584-006-9228-x, pp. 53–70.
URL: http://dx.doi.org/10.1007/s10584-006-9228-x. See p. 9.

Dudhia, J. (1989). 'Numerical study of convection observed during the winter monsoon experiment using a mesoscale two-dimensional model'. In: *J. Atmos. Sci* 46, pp. 3077–3107. See pp. 19–21.

Gobiet, A. et al. (2010). *Integriertes Modellsystem Steiermark-Umwelt: Pilotprojekt Integrierte Wetter-Klima-Luftgüte-Land-Hydrologie Modellierung, IMOS-U v1.0 Final Report (FinRep).* Tech. Rep.
University of Graz, Graz, Austria: Wegener Center for Climate and Global Change, p. 141. See pp. 16, 17.

Grell, G. A. et al. (1995). *A Description of the Fifth-Generation Penn State/NCAR Mesoscale Model (MM5), NCAR/TN-398+STR.* Tech. rep.
Available at: http://www.mmm.ucar.edu/mm5/mm5-home.html (December 2010).
Boulder, USA: National Center for Atmospheric Research (NCAR). See pp. 19, 20.

Grell, G. A. et al. (2000).
'Nonhydrostatic climate simulations of precipitation over complex terrain'.
In: *J Geophys Res-Atmos* 105.24 (Dec. 2000), 29595–29608. See pp. 9, 27.

Haiden, T. et al. (2011). 'The Integrated Nowcasting through Comprehensive Analysis (INCA) system and its validation over the Eastern Alpine region'. In: *Weather Forecast, in Press.* DOI: 10.1175/2010WAF2222451.1. URL: http://journals.ametsoc.org/doi/abs/10.1175/2010WAF2222451.1. See pp. 10, 14–17, 59.

Haylock, M. R. et al. (2008). 'A European daily high-resolution gridded data set of surface temperature and precipitation for 1950-2006'. In: *J. Geophys. Res.* 113.D20 (Oct. 2008). See pp. 11, 14.

Hohenegger, C. et al. (2008). 'Towards climate simulations at cloud-resolving scales'. In: *Meteorol. Z.* 17.4, Sp. Iss. SI, 383–394. See pp. 9, 27.

Hong, S. Y. et al. (2006). 'A new vertical diffusion package with an explicit treatment of entrainment processes'. In: *Mon. Wea. Rev.* 134, pp. 2318–2341. See p. 21.

Hong, S. and H. Pan (1996). 'Nonlocal Boundary Layer Vertical Diffusion in a Medium-Range Forecast Model'. In: *Mon. Wea. Rev.* 124.10, pp. 2322–2339. See pp. 19, 20.

Kain, J. S. (2003). 'The Kain-Fritsch convection parameterization: an update'. In: *J. Appl. Meteor.* 43, pp. 170–181. See pp. 18–20.

Kain, J. S. and J. M. Fritsch (1993). 'Convective parameterization for mesoscale models: the Kain-Fritsch scheme'. In: *Meteorol. Monogr.* 24, pp. 165–170. See pp. 18, 19.

Keuler, K. et al. (2010). *Final Report of Project Group 'Cold-Bias'; Sensitivity studies with the new COSMO-CLM version 4 with respect to cold bias and time dependence of simulation results.* Tech. Rep. P. 16. See pp. 36, 60.

Kirchengast, G. et al. (2008). *Pionierexperiment WegenerNet Klimastationsnetz: Ein neuartiges Messnetz in der Region Feldbach (Steiermark/Österreich) zur Beobachtung von Wetter und Klima mit sehr hoher Auflösung.* 1st ed. Wegener Center Verlag. ISBN: 139783950261509. See p. 14.

Klemp, J. B. and R. B. Wilhelmson (1978). 'Simulations of three-dimeensional convective storm dynamics'. In: *J. Atmos. Sci* 35, pp. 1070–1096. See pp. 19, 20.

Laprice, R. (2010). *Where and when should one hope to find added value from dynamical downscaling of GCM data.* See p. 27.

Linden, P. van der and J. F. B. Mitchel (2009). *Climate Change and its Impacts: Summary of research and results from the ENSEMBLES project.* 1st ed. Met Office Hadley Center, Exeter. See p. 9.

Mlawer, E. J. et al. (1997). 'Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave'. In: *J. Geophys. Res.* 102.D14, pp. 16663–16682. See pp. 19–21.

Prein, A. F. and A. Gobiet (2011). *NHCM-1: Non-hydrostatic climate modelling Part 1: Defining and Detecting Added Value in Cloud Resolving Climate Simulations.* 1st ed. Wegener Center Verlag. ISBN: 978-3-9502940-6-4. See pp. 10, 21, 23, 24.

Reisner, J. et al. (1998). 'Explicit forecasting of supercooled liquid water in winter storms using the MM5 mesoscale model'. In: *Quart. J. Roy. Met. Soc.* 124.B, pp. 1071–1107. See pp. 19, 20.

Riegler, A. (2007). *Evaluierung eines hochaufgelösten regionalen Klimamodells im Alpenraum.* Wegener Center Verlag, 132 pp. ISBN: 13 978-3-9502308-7-1. See p. 21.

Roberts, N. M. and H. W. Lean (2008). 'Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events'. In: *Mon. Wea. Rev.* 136.1 (Jan. 2008), pp. 78–97. See pp. 24, 46.

Rogers, E. et al. (2001). *Changes to NCEP meso ETA analysis and forecast system: increase in resolution, new cloud microphysics, modified precipitation assimilation, modified 3DVAR analysis.* NWS technical procedures bulletin. NOAA. See pp. 20, 21.

Seneviratne, S. I. et al. (2006). 'Land-atmosphere coupling and climate change in Europe'. In: *Nature* 443, pp. 205–209. See pp. 9, 18.

Skamarock, W. C. et al. (2007). *A Description of the Advanced Research WRF version 2.* Tech. rep. 468. Mesoscale and Microscale Meteorology Division at NCAR. See pp. 20, 21.

Stevenson, M. (2006). 'Forecast verification: A practitioner's guide in atmospheric science'. In: *Int J Forecasting* 22.2, pp. 403–405. See p. 23.

Taylor, K. E. (2001). 'Summarizing multiple aspects of model performance in a single diagram.' In: *J. Geophys. Res.* 106.D7, pp. 7183–7192. See pp. 23, 46.

Wilks, D. S. (2005). *Statistical Methods in the Atmospheric Sciences, Volume 91, Second Edition (International Geophysics).* 2nd ed. Academic Press. ISBN: 0127519661. See p. 23.

Abstract:

Climate change is a global problem, but its impacts on society and ecosystems are predominantly occurring on rather small scales. Especially small scale extreme events and hydrological processes have large impacts, which generates high demand for reliable small scale climate projections. However, simulating climate at local scales is currently neither established nor reliable. Therefore, in the Local Climate Model Intercomparison Project (LocMIP) was initiated to enable comparable local scale climate simulations (also called convection resolving climate simulations, CRCS) performed with different models operated by different research groups. This initiative is a starting point to establish reliable and internationally accepted information on the error characteristics of specific convection resolving climate models and on the quality of CRCSs in general. Within LocMIP, two groups operated three different models, with further groups expressing interest to contribute in future. LocMIP provides first valuable results and methods, datasets, and formats that can now be easily used and expanded for larger initiatives.

The results of LocMIP show that each RCM has its strengths and weaknesses and that no model clearly outperforms all other models. However, model deficiencies and “bad” models can be clearly identified. Generally newer models or newer model versions outperform the older ones. Added value of CRCS compared to simulations on a 10 km grid can be found in the representation of spatial patters of the 2 m temperature field, which is mainly caused by the better representation of orography. Additionally, the 3 km simulations are able to produce rainfall with higher intensities often show more skill in generating heavy precipitation.

Zum Inhalt:

Der Klimawandel ist ein globales Problem, jedoch sind seine Effekte auf die Gesellschaft und Ökosysteme am stärksten lokal spürbar. Da vor allem relativ kleinskalige Extremereignisse gravierende Auswirkungen haben, besteht hoher Bedarf an verlässlichen kleinskaligen (konvektionsauflösenden) Klimaszenarien. Klimamodelle, die solche Szenarien liefern können sind aber noch in der Versuchsphase. Das „Local Climate Model Intercomparison” Projekt (LocMIP) wurde initiiert um vergleichbare konvektionsauflösende Klimasimulationen mit unterschiedlichen Modellen von unterschiedlichen Forschungs-gruppen zu ermöglichen. Es stellt einen Startpunkt für die Etablierung von international akzeptierter Bewertung konvektionsauflösender Klimamodelle dar. In LocMIP habe zwei Forschungsgruppen drei unterschiedliche Modelle betrieben und weitere Gruppen haben Interesse bekundet sich dieser Initiative in Zukunft anzuschließen.

Die Ergebnisse zeigen, dass unterschiedlich Modell unterschiedlich Schwächen und Stärken aufweisen und es schwer ist ein „bestes“ Modell zu finden. Allerdings können Defizite in Modellen und „schlechte“ Modelle sehr deutlich identifiziert werden. Generell zeigen neuere Modelle oder Modellversionen bessere Performance als ihre Vorläufer und es konnte gezeigt werden, das konvektionsauflösende Simulationen vor allem in räumlichen Mustern des Temperaturfeldes und in der Darstellung von Starkniederschlägen Vorteile bringen.

Wegener Center for Climate and Global Change
University of Graz
Leechgasse 25
A-8010 Graz, Austria
www.wegcenter.at
ISBN 978-3-9502940-6-4