Predictability of European Heat Waves

Zur Erlangung des akademischen Grades eines
DOKTORS DER NATURWISSENSCHAFTEN
von der KIT-Fakultät für Physik des
Karlsruher Instituts für Technologie (KIT)

genehmigte

DISSERTATION

von

M.Sc. Natalie Laube
aus Bad Pyrmont

Tag der mündlichen Prüfung: 24.05.2019

Referent: Prof. Dr. Christoph Kottmeier

Korreferent: Prof. Dr. Joaquim Pinto

All hope abandon, ye who enter here! - Dante Alighieri

Abstract

Periods like heat waves (HWs) and long-time warm spells are common during the summer months. However, these events have become more frequent and severe in the last two decades with devastating consequences (Meehl and Tebaldi, 2004; Della-Marta et al., 2007a; Russo et al., 2015). Mega-Heatwaves (Barriopedro et al., 2011) like 2003 in Central Europe or 2010 in Eastern Europe and Russia led to ten thousands additional heat related deaths (Fouillet et al., 2006; Robine et al., 2007) and had an extensive impact on the economy and environment. Events like wildfires and droughts frequently accompany the actual HW increasing the death toll and economic losses foremost due to crop failures. Therefore, seasonal to decadal predictions as well as climate projections are of great importance for warning systems and adaptation planning. Long-term observations as well as large ensembles of model simulations are applied to analyse the mechanism of HWs and to assess their predictability. This study considers three research topics: Decadal predictability, dependence to soil conditions and future changes of HWs.

The impact of changes in soil moisture and soil temperature fields on a decadal simulation for Europe is explored with the land surface model (LSM) VEG3D with special focus on summer temperatures. LSMs as the lower boundary of climate models are sensitive to inaccuracies in the initialisation data since they disturb the LSM from its balanced state. The fluctuations of the drift back to the balanced state are transferred to a certain degree into the atmosphere and are part of the bias in climate simulations. Therefore, stand-alone (SA) simulations are conducted with VEG3D to investigate its sensibility to the atmospheric forcing, model resolution and external data (soil types and land use). The SA results are evaluated against measurements from stations as well as satellite data and simulations with other LSM. The comparison revealed that ERA-Interim (Dee et al., 2011) forced simulations offer the highest correlations to the observations. With this forcing, balanced soil initial fields are produced by a $0.22°$ VEG3D SA simulation for Europe. A decadal prediction with a coupled COSMO-CLM_VEG3D system was performed using the new initial fields of VEG3D SA. The soil initialization show an evident impact on the spatial and temporal variability of summer temperatures in Europe. In particular, transition zones are dominated by the dependence of the evapotranspiration to the soil moisture, the effect was detectable even in the second pentad of the simulation. An added value was found for the Iberian Peninsula in the summertime related to the dryer soil in the run with the balanced fields. This simulation produced a more realistic drought stress in summer due to the lower soil water and higher 2 m temperatures. On the other hand, the impact is detected for less than four years in Scandinavia. There, an overall higher soil moisture is available limiting the land-atmosphere coupling, although traceable differences in the deeper soil layers still exist. It can be

concluded, that accurate soil initializations can be crucial for the development of extreme summer temperatures depending on dryer and warmer soils in model simulations. This finding is highly important in regard of the climate change and a possible enlarging of the transition zones.

The predictability of HWs is investigated with the extensive ensemble of global and regional decadal hindcasts conducted in the Bundesministeriums für Bildung und Forschung (BMBF) funded Mittelfristige Klimaprognose project (MiKlip Marotzke et al., 2016). In the first step, two reanalyses driven CCLM simulations are employed to estimate the ability of the regional model to reproduce HW characteristics for Europe derived from gridded daily observations (EOBS, ECA&D Haylock et al., 2008). A warm bias in Southern Europe and a cold bias in Northern Europe is found for the HW temperatures, whereas the number of HWs and their averaged mean and maximum lengths are quite well simulated. Decadal variations of the HW temperatures are detected with a moderate correlation (r ~0.5) to the Atlantic meridional oscillation (AMO). In addition, mediocre to high correlations (positive as well as negative) are found between HWs and the large-scale circulation anomalies (LCA) mostly located in the area of their centre of actions. An increased number of HW days is found for Central Europe during the positive upper third NAO index, while the days are reduced in Turkey and Southeastern Europe. Similar correlation patterns are detected in the reanalysis driven CCLM simulations indicating the model is able to reproduce the teleconnection of HWs and LCA. However, the correlations between monthly HW day sums and the LCA indices are not symmetrically distributed, when the index is filtered into its positive and negative phase. Depending on the region, higher (anti-)correlations are noticed for the different index phases. One phase seems to represent a dominant signal and mostly only smaller or no vice-versa signals are noticed for the contrary phase. For example, the correlation is 0 for the upper third of the EA pattern for the British Isles, whereas the negative lower third shows a moderate anticorrelation (r< -0.45). The yearly mean and maximum temperatures, as well as the lengths and HW per year numbers, are of similar quality in the regional hindcast ensemble as the ones from the reanalysis driven simulations. The global ensemble displays a larger cold bias for the temperatures and a positive bias for the HW lengths. A new approach to evaluate the ensemble on event basis was tested using the t-test to calculate fractional hit rates for each lead time and ensemble member. This method avoids a too strong smoothing of the events due to the established multi-year and ensemble-member averaging. In addition, HW days per year are introduced as a new user relevant variable. Furthermore, the evaluation showed, that they are predictable with a good skill overall in Europe.

A new ensemble of cloud-resolving climate projections was produced to bridge the gap between usual climate simulations with around 25 km resolution and impact models operating on several hundred meters and to fulfil the demand after very high resolved climate information. Forced by data from three global models, the ensemble is dynamically downscaled to $0.025°$ with three nesting steps (global to $0.44°$ to $0.0625°$ to $0.025°$). Besides the control (CTRL) period 1971-2000, two future period runs (2021-2050 and 2071-2100) using the RCP8.5 scenario were performed. Comparing the final nest, $0.025°$, to the coarser $0.0625°$ revealed an added value of the

higher resolution due to a decreased temperature bias. In contrast to this, higher precipitation errors are produced by the cloud-resolving run. Daily HW temperatures are better represented in the 0.025° simulations in the mean but their variance is too large compared to observations. An increase in the HW temperatures of nearly 7°C is detected for the future periods until 2100. In addition, an expansion of their occurrence time is noticed. HW conditions of the CTRL period are identified for nearly the whole summer half-year in the distant future. A reduction of precipitation was detected along with a decrease of relative humidity, soil moisture, latent heat flux and cloud cover and an increase of 2m and soil temperatures, sensible heat and incoming solar radiation. The lower soil moisture is one of the most key factors for HW development and amplification. More severe and longer lasting HWs can develop in the future periods due to the large reduction of the soil water content. Thus, conditions like 2003 will become more frequent rather than an exceptionally rare event.

Contents

List of Abbreviations

AA .... Arctic Amplification
AMO .... Atlantic Multidecadal Oscillation
*b*1 .... baseline1
CCLM .... COSMO-CLM
*CMIP*5 .... Coupled Model Intercomparison Project Phase 5
COSMO .... COnsortium for Small-scale MOdeling
CTRL .... Control period 1971-2000
DF .... Distant Future 2071-2100
EA .... East Atlantic
EAWR .... East Atlantic/Western Russia
ECMWF .... European Centre for Medium-Range Weather Forecasts
EPNP .... East Pacific/North Pacific
ESM .... Earth System Model
GCM .... global climate model
HW .... Heat wave
JJA .... June, July, August
LSM .... Land Surface Model
LT .... lead-time
MiKlip .... Mittelfristige Klimaprognose
MSESS .... mean square error skill score
NAO .... North Atlantic Oscillation
NF .... Near Future 2021-2050
NOA .... North Atlantic Oscillation
PDF .... probability density function
PNA .... Pacific/North American
POL .... Polar / Eurasia
QBO .... Quasi-Biennial-Oscillation
RAV .... relative added value
RCM .... regional climate model
RMSE .... root mean square error
RPCA .... rotated principle component analysis
SCAND .... Scandinavia
SNAO .... Summer North Atlantic Oscillation
SST .... Sea Surface Temperature

WP .... West Pacific

SHAI .... standardized heat area index

SHI .... standardized heat index

1. Introduction

Longer lasting hot phases with extreme temperatures are common during summer and often described as Heat Waves. Human health can be vitally affected by the anomalously high temperatures and economic losses and environmental damage are registered. Heat waves (HW) often occur consecutively or are preceded by more or less extensive drought, which can increase the consequences of the HW tremendously (Schubert et al., 2014). Over the last twenty years, very severe HWs were noted and the connection to the changing climate was pointed out by scientist (Meehl and Tebaldi, 2004; Della-Marta et al., 2007a; Fischer and Schär, 2010; Seneviratne et al., 2010; Russo et al., 2015). An unprecedented strong HW (HW) in intensity and affected area took place in 2003 (Schubert et al., 2014; Schär et al., 2004). In Germany temperatures around 40°C were registered, and the mortality rates increased up to 70.000 deaths across Europe during the HW in August alone (MunichRe NatCatSERVICE, 2018). Over the whole year, estimated 80.000 additional deaths (compared to the average 1998-2002) were counted because of the increased heat stress (Fouillet et al., 2006; Robine et al., 2007; García-Herrera et al., 2010) affecting mostly vulnerable groups like elderly or invalid people. Anomalously high temperatures could develop during the longlasting HW in East Europe and Russia in 2010 due to an excessive atmospheric blocking (Rex, 1950; Trenberth and Fasullo, 2012) with temperatures over 38°C in Moscow (Barriopedro et al., 2011). As in 2003, the death toll surpassed 50.000, while wildfires destroyed large areas and heavy losses were noted for the agricultural sector (~25 %) (MunichRe NatCatSERVICE, 2015). The next severe HW evolved in 2015, and although the maximum temperatures were lower than 2003, additional humidity increased the heat stress for the population (Muthers et al., 2017). 2017 - Another strong HW known as Lucifer by public reporting was documented in Southern Europe. Measurements at the Airport in Cordoba (Spain) registered maximum temperatures of 46.9°C in July and similar values in August (La Agencia Estatal de Meteorología (AEMET), 2017). New temperature records are set every year currently with no exception in 2018. Several countries installed heat health warning systems after 2003 because of the increased mortality rates during the HWs (Zacharias and Koppe, 2015; Muthers et al., 2017). However, long term planning dependents on the results of scientists and the knowledge about predictability and future changes about the extent of HWs (Muthers et al., 2017; Coughlan de Perez et al., 2018). It is likely HW will have an even stronger impact on human health, the economy and the environment in the future as a result of the globally rising temperature due to climate change (IPCC, 2013).

Still, uncertainties exist, how HWs develop and how their characteristics will change in the future (Schubert et al., 2014; Miralles et al., 2018), since they depend on particular conditions. HWs predominantly occur during the summertime in regions under stable anticyclones (Schu-

bert et al., 2014; García-Herrera et al., 2010). These enhance subsidence, which decreases cloudiness and higher incoming solar irradiation warming the surface. The stable anticyclones are linked to certain atmospheric conditions like blocking and other large-scale circulations anomalies (Stefanon et al., 2012; Vautard et al., 2007) which can trigger a HW. If the soil is drying-out or in the case of 2003 lacks moisture due to missing spring rains, sensible heat fluxes intensify and can increase the temperatures even further. Moreover, the soil conditions and the land-atmosphere feedback (Fischer et al., 2007a; Zampieri et al., 2009; Miralles et al., 2018) play a crucially important role during the development and for the strengths of a HW. Often discussed is the role of the ocean and to what extend northern hemisphere temperature extremes are connected to sea surface temperature (SST) changes (Folland et al., 2009; Sutton and Hodson, 2005; Hodson et al., 2010). Especially the slow varying changes of the Northern Atlantic ocean temperature offer a source of predictability for Europe (Müller et al., 2012; Pohlmann et al., 2013; Årthun et al., 2018). The bold red marked impact factors on HWs will be discussed further in section 2.2 in the next chapter.

In this thesis, HWs characteristics will be investigated across Europe and in very high-resolution climate simulations. In addition, a new basic HW definition is tested. With this definition detecting HWs will be possible for various time and spatial scales since the employed data sets vary largely in their spatial resolutions. Additionally, the definition is constructed to achieve the best possible applicability for other users. The following main research questions are going to be answered in this work:

1. How large is the impact of the soil initialisation on summer temperatures and with this on HWs?
2. Is there a decadal signal for HWs and is it possible to predict this with decadal hindcasts?
3. To what degree exists a dependence between large climate anomalies and HWs, and is it possible to simulate this with climate models?
4. What resolution is needed to produce reasonable projections of HWs for smaller evaluation areas and to what extent change the HWs in the Future?

In agreement with these questions, the thesis is separated into three main parts. In the first main chapter, the impact of the soil initialisations is investigated using a coupled simulation system. The land surface model (LSM) VEG3D is applied to provide the best available soil initialisation for this simulation; then a decadal prediction is performed. An estimation of the agreement of the produced soil water and temperature fields with observations will be given, and the impact of the changed simulation on the summer temperature will be assessed (chapter 3). In the second part, decadal predictions from the Mittelfristige Klimaprognose project (**MiKlip**) will be employed to examine the decadal signal of HWs (chapter 4). In addition, it will be investigated if this system is capable of reproducing the decadal variations and which weaknesses occur in the prediction of HWs. Furthermore, the prediction potential is investigated which originates from the connection of HWs to large climate anomalies like the NAO or other and the AMO. In the third central chapter, HWs are analysed on a local scale for Southern Germany. Although many

studies investigated HWs on the European scale, local aspects are still missing. HW characteristics are determined with a newly build ensemble of very high-resolution climate simulations for a control period of 1971-2000 and their future changes are estimated (chapter 5). Moreover, the question will be discussed in this chapter which resolution for climate simulations is indeed required. Besides this introduction, the thesis includes a chapter with HW definitions and the scientific background. A section describing the regional climate model COSMO-CLM is included in the first main chapter. Each chapter includes a short introduction, a summary of the main results and an extensive chapter discussion and conclusion. The research questions listed in the introduction are answered at the end of the thesis.

2. Definition of Heat Waves and scientific background

2.1. Heat Wave Definitions

Phases of intense heat are described as Heat Waves in summer, but no consensus about an explicit definition (Robinson, 2001) is reached so far. Summer or Hot days are defined as days with a daily maximum temperature $> 25/30°C$ (Zhang et al., 2011). Their absolute temperature threshold is adapted to the mid-latitudes only. Other definitions are based on a percentile threshold (90th or 95th) with varying lengths from three to six days. Perkins and Alexander (2013), for example, use the 95th percentile and a duration of 3 days, whereas Fischer and Schär (2010) apply a 90th percentile threshold to be exceeded for at least six consecutive days. In addition, the reference period for the percentiles differs considerably between the individual studies. In several, the percentile is calculated for a climatological period ((Fischer and Schär, 2010, 1961-1990) or (Lhotka et al., 2018, 1981-2010)), while others include the percentile of a moving window with varying lengths from five until 80 days (e.g. Barriopedro et al., 2011). Depending on the aim of the study, the main summer months June, July and August (JJA) are included in the percentile calculation (most of the previously named studies) or the whole summer half year. Lhotka et al. (2018) opted for an extension of the summer months to May until September due to the rising temperatures leading to potentially earlier or later occurring HWs. Furthermore, the difference between a warm spell and a HW is often discussed. Nairn and Fawcett (2015) proposed to define HWs as periods hot in an absolute sense, but not in a relative one. This would restrict the occurrence of HWs to the summer months and warmer climate zones, whereas a warm spell can be identified even in wintertime and in subpolar to polar climates. To detect HWs throughout the year, Barcena-Martin et al. (2018) offered an alternative definition using a daily based percentile as a threshold. On the other hand, Lhotka et al. (2018) employed a relative approach in their analysis and calculated a global intensity as a sum of the differences between the daily temperatures and a certain threshold. With this approach, HW signals can be detected in colder climates also. However, these HWs would be referred to as warm spells in most cases.

There are more complex definitions and indices combining the temperature with other variables besides ones based exclusively on the daily (minimum, mean or maximum) temperature. These joined indices can be extremely effective to estimate the impact on the economy or human health. The HUMIDEX, for example, proposed by Steadman (1979), combines temperature and the relative humidity. Another option is the physiologic equivalent temperature (Höppe, 1999; Matzarakis et al., 1999), which measures the thermal discomfort for average humans as it includes additional thermo-physiological parameters (Matzarakis, 2001). Furthermore, an optimum temperature was defined for the human body outside which health issues related to ex-

tremely low or excessively high temperatures appear (Honda et al., 2007; Ono, 2013; Zacharias and Koppe, 2015). If the body is unable to effectively cool down because of a too hot environment and/or a too intense strain like physical exertion, heat strokes are very likely. During a heat stroke, the body core temperature exceeds 40°C leading to moisture loss due to increased sweating, dizziness, headaches and in more severe cases to seizures, organ failures and death if not treated immediately (Ronaldson, 2016). Therefore, it was decided to include the daily maximum temperatures in the analysis. It helps to assess the risk for vulnerable or people working outdoors during HW without the chance of cooling. One has to keep in mind, that the combination of higher temperatures and ambient moisture like for HW 2015 is related to higher mortality rates as well (Muthers et al., 2017). If the body transpires not enough, the effective thermoregulation of the body is inhibited (Parsons, 2002; Havenith, 2002). As a consequence, mortality rates can increase, although the maximum temperatures were not as high as, e.g. in 2003.

A comparable definition for a heat wave is applied in this study similar to the one proposed in Perkins and Alexander (2013) and to indices developed from the Expert Team on Climate Change Detection and Indices (from ETCCDI: Warm spell/Heatwave duration index, Zhang et al. (2011)).
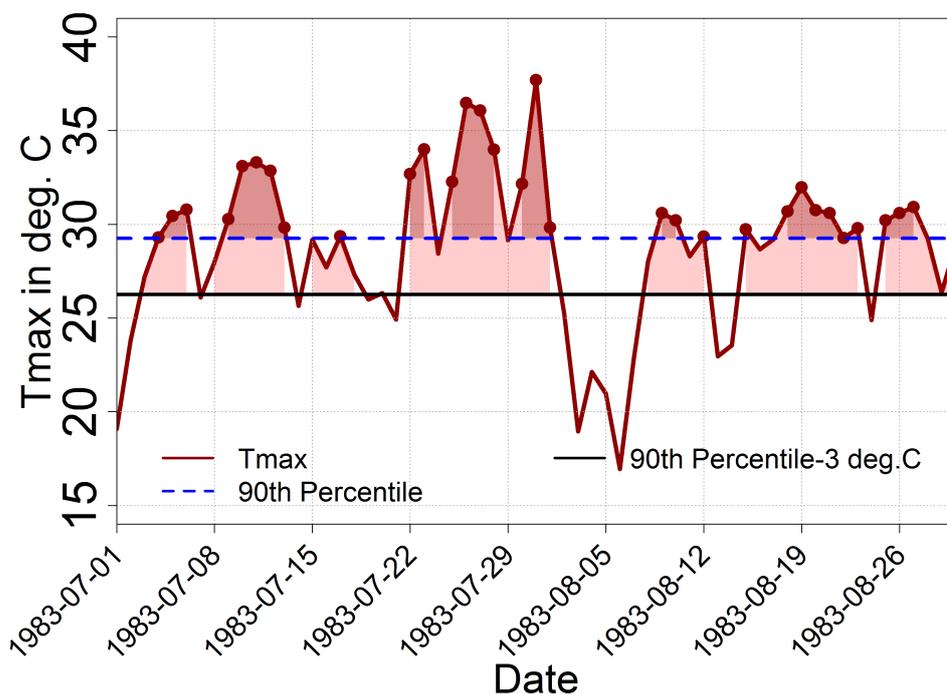
Figure 2.1.: Hypothetical daily maximum temperatures for July and August in 1983 to demonstrate the definition of an HW; daily maximum temperatures: dark red; blue dotted: the 90th percentile, black solid: the 90th percentile minus three deg. C.

Based on the daily maximum temperature, a heat wave is defined by a minimum of five consecutive days exceeding the long-term 90th percentile of the hydrological summer half-year

(May-October). The 90th percentile includes the 18.5 warmest days of summer and is calculated for a reference period from 1971-2000. Two criteria are implemented to handle smaller interruptions of the HW. The first is a relative break of the time series. If a HW is detected and the daily maximum temperatures of not more than two days are below the 90th percentile minus 3°C, a relative break is assumed but no interruption of the HW. If the temperatures are registered below the 90th percentile minus three degrees or the relative break is longer than two days, the HW is effectively interrupted (similar to the definition of Ouzeau et al. (2016)). Fig. 2.1 shows the break criterion applied to an artificial temperature time series. The deep red shadings mark the days exceeding the 90th percentile. Five consecutive days are detected from the 09.07.1983 until the 13.07.1983 identifying the first heat wave. The period from the 22.07.1983 until the 02.08.1983, on the other hand, includes two days with temperatures below the 90th percentile. However, this time is counted as one HW due to the first criterion. The simplicity of this procedure was chosen intentionally, as it can be easily applied to other data sets. In this study, most data consist of several ensemble members with deviating mean climate states and, therefore, different biases. Therefore, the index was computed for each ensemble member separately based on its specific 90th percentile. The annual mean and maximum duration, the number and the mean and maximum temperatures have been computed for the identified heat waves, and the ensemble means afterwards.

The aspects studied by researchers are often similar despite the numerous definitions: Onset and duration, as well as their spatial extent and intensity, the connections to large-scale circulations and/or local effects and small-scale processes affecting the HW. So-called "mega-heatwaves" are of particular interest (Barriopedro et al., 2011). These severe events possess anomalous strength and encompass areas with over 1 million square kilometres. The first occurrence of this category since the beginning of recording was 2003 (Schubert et al., 2014), the next event of this magnitude followed not much later in 2010 (mainly East Europe and Russia Barriopedro et al., 2011). Thus, HWs pose a possible threat to most of the world's population, especially for vulnerable groups in more temperate climate zones, like elderly and weak people, or those living in warmer climates.

The United Nations Office for Disaster Risk Reduction stated that about 75 % of the weather-related deaths in high-income countries are connected to temperature extremes (both cold and warms CRED et al., 2015). There is high confidence, that the frequency and intensity of HWs will increase due to global warming and with this the mortality as well (IPCC, 2013; Mora et al., 2017). Russo et al. (2015) detected at least five of ten record-breaking HWs after 2000 (see Table 2.1). This accumulation demonstrates the rapidly changing climate conditions, especially temperature extremes, and that the impacts are noticeable even nowadays (Donat et al., 2013). Also, a broadening and shift of their probability density function (PDF) was found, besides the rising mean temperatures (Lau and Nath, 2014; Zwiers et al., 2013). Both, the wider distribution as well as the sift, implicate a stronger increase of higher quantiles of the daily temperatures and in such the possibility for new record-breaking extreme heat events. Rahmstorf and Coumou (2011) estimated that the number of record temperatures increased fivefold due to climate

warming. The HW event 2010 in Moscow would not have occurred with a probability of 80 % without climate warming. The risks for the occurrence of HWs like 2003 has doubled due to the human influence (Stott et al., 2004). Thus, HWs are affected by climate change, but not only in the distant future as most of the studies investigating the change signal showed.

Year	Loc.	HWMId Peak	Area(%) HWMId $\geq$ 6	Area(%) HWMId $\geq$ 9	Area(%) HWMId $\geq$ 15	Area(%) HWMId $\geq$ 24
2010	Russia	71.9	36.38	29.13	22.54	14.07
2003	Cent. Eu	44.7	11.61	9.17	5.44	1.65
1972	Finland	38.2	26.42	18.35	6.57	0.96
1976	UK Brit.	35.8	4.55	2.98	1.21	0.23
1969	Norway	26.5	2.26	1.20	0.38	0.02
2015	Cent. Eu	26.0	11.94	5.67	0.56	0.01
2007	Greece	22.9	16.80	7.90	1.35	0
1994	Benelux	21.3	7.42	3.89	0.46	0
2014	Scandin.	21.2	11.58	3.65	0.3	0
1954	SW Rus.	19.7	9.3	1.9	0.05	0
2006	Cent. Eu	18.9	5.05	1.28	0.05	0

Table 2.1.: List of record-breaking heat wave events in the period 1950-2014 with E-OBS data including also data until September 2015 for the most recent heat wave. The latter is an additional information to the originally considered top 10 heat waves. For each specific event the spatial extent is estimated as the land area fraction exceeding a fixed HWMId value. The area fraction is expressed in percentage. The HWMId peak is the highest spatial HWMId value recorded during each specific event. Table and caption from Russo et al. (2015)

Meehl and Tebaldi (2004) calculated the changes in the HW duration and their occurrences times for Chicago and Paris and found an increase for both cities until the end of 2100. For Paris, the ensemble mean HW numbers increase from 1.18 (1961-1990) to 2.17 (2080-2099) and the HW duration from 11.39 days to 17.04 days. Della-Marta et al. (2007b) and Stefanon et al. (2012) determined similar results by analysing HWs over Europe for different geographical patterns, although their definition of HW and their analysis methods (canonical correlation vs clustering) differed. Fischer and Schär (2010) deduced a geographical dependence of the change signal, with the highest increase of HW temperatures and frequency for regions in Southern Europe. Additionally, a ten times higher chance was estimated for record-breaking events in Europe towards the end of the century (Bador et al., 2016).

Even more drastic is the effect of the upcoming HWs due to the climate change illustrated in the study of Mora et al. (2017). They identified life-threatening climatic conditions based on temperatures and relative humidity explicitly and pointed out the fraction of the human population experiencing it already. Fig. 2.2 shows that the land area and the population amount affected by deadly climatic conditions are rising significantly for all RCP scenarios (van Vuuren et al., 2011; Moss et al., 2010). In particular, the RCP8.5 reveals nearly half of the land area and over 70 % of the population will be exposed to these conditions until 2100. Therefore, knowledge and reliable prediction of such events are crucial due to the known consequences for

human health. Furthermore, estimates about the HW magnitudes from weather to seasonal to decadal time-scales become vital in case of the economic and ecologic adaptation strategies and the advancement of warning systems.
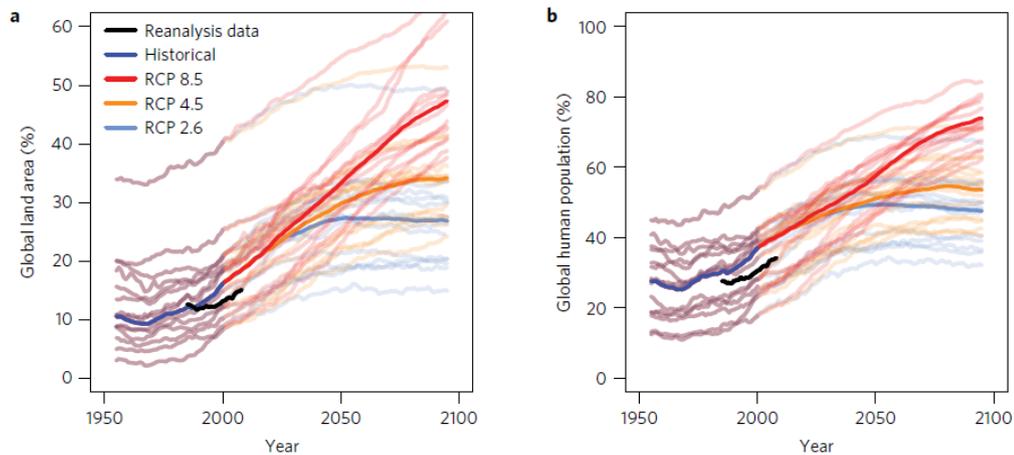
Figure 2.2.: Current and projected changes in deadly climatic conditions. Area of the planet (a) and percentage of the human population (b) exposed to climatic conditions beyond the 95% SVM deadly threshold (red line in Fig. 1b) for at least 20 days in a year under alternative emission scenarios. Bold lines are the multi-model medians, black lines are the results from reanalysis data and faded lines indicate the projections for each Earth System Model. Time series were smoothed with a 10-year-average moving window. Area of the planet and human population exposed to different lengths of time are shown in Supplementary Fig. 4. Results correcting for climatological mean biases between the reanalysis data and each Earth System Model are shown in Supplementary Figs. 8 and 10. Graphics and caption from Mora et al. (2017)

2.2. Ocean, large-scale atmospheric teleconnections and soil impacts

Three of the main influences influencing HWs are explored in this section beginning with the indirect influence of the ocean. As a source of predictive skill for temperature in Europe, it will most likely affect the predictability of HWs as well.

Cassou et al. (2005) established a link between European HWs and tropical Atlantic temperatures and with this the connection between the occurrence of HWs and certain atmospheric circulation regimes during the summer months. The leading mode in the North Atlantic, the North Atlantic Oscillation (NAO), is defined as a North-South Dipole of pressure anomalies between Iceland and the Azores (Barnston and Livezey, 1987). Its summer pattern, the Summer North Atlantic Oscillation (SNAO described by Hurrell et al. (2003) or Folland et al. (2009)), is located farther north than the winter one with a Southern Node over Northwest Europe and the Northern Node reaching into the Arctic (see Fig. A.1 in the Appendix). The mean sea level pressure the leading pattern of the SNAO explains about 22.1 % of its variance in JJA marking it as the leading mode in the Atlantic during the summer months too (July and August) Hurrell et al. (2003). In addition, the high correlations of the SNAO to the mean July, August 2 m temperatures for 1900-2007 (correlation around 0.5) for North and Western Europe indicated a possible connection of the SNAO to HWs. Cassou et al. (2005) confirmed this by determining

that "Blockings" or the positive phase of the SNAO favour HW days in France. Furthermore, higher numbers of hot days were detected during an extensive Atlantic Low with corresponding High over Central Europe. Another possibility to investigate the connection to HWs was tested by Chronis et al. (2011), who calculated the correlation between NCEP-DOE 700 hPa temperature data and the NAO monthly indices. The results are high positive correlations for Scandinavia and the British Isles in June and July, and higher negative correlations over the Mediterranean in July and August. The positive correlations are a result of reduced cloudiness during the positive phase dominated by a large anticyclone over Northwestern Europe, while vice-versa conditions occur in the Mediterranean. Folland et al. (2009) discussed the dependence to the SSTs in the Atlantic and with this to the Atlantic Multidecadal Oscillation (AMO Enfield et al., 2001). They concluded, along with studies from (Sutton and Hodson, 2005; Knight et al., 2006), that AMO warm phases frequently correspond with negative SNAO phases and vice-versa due to the changes in the mean sea level pressure. On the multidecadal scale, Ghosh et al. (2017) found a connection of the AMO and corresponding precipitation anomalies in Northwestern Europe. All this leads to the conclusion that the AMO is more or less linked to the enhancement or suppression of HWs and represents a possible source of predictability for HWs.

Another impact chain influencing HWs are the strength of the jet stream and the location of Rossby waves (Schubert et al., 2011; Lau and Nath, 2014; Trouet et al., 2018). Extreme events like floods and HWs can be linked to the position of the jet stream (Trouet et al., 2018). The velocity of the jet stream depends on the gradient between the temperatures in the Arctic and the lower latitudes (Woollings et al., 2010). If this gradient weakens due to a warmer Arctic, the velocity of the zonal winds is reduced, and ridges elongate farther to the North (Francis and Vavrus, 2012). Therefore, HWs and droughts are increased in Britain during a northward shift of the jet, whereas a southern shift leads to cold and wet summers. An opposing signal is noted in the Mediterranean corresponding to a higher/lower number of blocking events in this region. Furthermore, the propagation of the Rossby waves is slowed along the stronger meandering jet. If Rossby waves move with a small phase velocity or become quasi-stationary, extended episodes of unusually warm or cold weather are likely to occur in Europe Kyselý (2008). The stationary conditions increase the persistence of anticyclones and ridges which amplify the HW development and/or intensification. This was confirmed by results of Coumou et al. (2014), who also linked HWs/extreme weather to persistent quasi-stationary circulation regimes. Thus, the location and strength of the jet stream and the quasi-stationary Rossby wave trains can be indicators for the occurrence of HWs.

Summer 2018 was a live demonstration of the strength of this connection including persistent anticyclones, slower Rossby waves due to a weaker jet stream and very high temperatures across vast areas in Europe (Met Office, 2018). The question arises if such situations will occur more frequently due to climate change. The persistence of blocking could be increased by a weakening of the summer circulation in the Northern Hemisphere mid-latitudes leading to a more elongated jet with higher amplitudes Coumou et al. (2015). In particular, specific atmospheric circulation

patterns have become more persistent supporting longer and more severe HWs since the 1980s (Kyselý, 2002; Kyselý and Domonkos, 2006; Screen and Simmonds, 2014). One reason for the wavier jet streams is possibly the Arctic Amplification (AA) (Francis and Vavrus, 2012). The AA describes the higher rising temperatures in the Arctic over the last and in the upcoming years, which weaken the meridional temperature gradient. Thus, the AA increases the probability for wavier jet streams and in such of more persistent mid-latitude weather patterns (Francis and Skific, 2015; Petoukhov et al., 2013).

Besides the regions of the large circulation anomaly (LCA) themselves, also adjacent regions can have an impact on HWs. In particular, HWs can be triggered or intensified when warm and dry air is advected as, e.g. during the HW 2010 (Zampieri et al., 2009; Miralles et al., 2018). Mediterranean rainfall deficits in the preceding winter and spring season are another example of transport of anomalously warm and dry air from Southern Europe or the Mediterranean to Central Europe Vautard et al. (2007). The studies demonstrate the connection of HWs to foregoing or simultaneously occurring droughts and the soil conditions and soil memory. As Seneviratne et al. (2006a) noted, the soil is a slowly varying climate component due to its longer response time up to two months. This inertia offers some potential predictability for temperature extremes in seasonal to decadal forecast, especially due to spring surface-moisture deficits (Brabson et al., 2005; Koster et al., 2010; Della-Marta et al., 2007b; Weisheimer et al., 2011; Hirschi et al., 2011). Less soil water in spring and early summer leads to reduced evapotranspiration rates preventing the formation of clouds and trigger a further drying of the soil due to missing precipitation.

These processes are of particular interest in transition zones like the Mediterranean (transition between dry and wet climate Koster et al., 2004), where a strong land-surface coupling dominates the interactions between soil vegetation and atmosphere (Seneviratne et al., 2006b; Seneviratne and Stöckli, 2008). There, an univocal link is found between lower soil moisture and higher temperatures in the summertime. Its role for the development of the HW is investigated in case studies like for the summer 2003 (Black et al., 2004; Ferranti and Viterbo, 2006; Fischer et al., 2007b). Extreme temperatures are frequently connected to missing soil moisture and increased sensible heat fluxes during phases with stable positive pressure anomalies. In addition, the persistence of HWs corresponds with the soil moisture memory (higher inertia especially in the thicker and deeper soil layers) (Lorenz et al., 2010). The soil moisture-temperature interactions were investigated further with regional climate simulations (Fischer et al., 2007a). An increase in the daily maximum temperatures and HWs duration up to 50-80 % was found for the major HWs 1976, 1994, 2003 and 2005 due to the soil-atmosphere feedbacks. However, Weisheimer et al. (2011) showed that without a sufficient representation of the soil-vegetation-atmospheric transfer in a land surface model, the development of HWs can be inhibited or dampened in model simulations. Furthermore, the decrease of soil moisture increases the intraseasonal to interannual variability (Jaeger and Seneviratne, 2010). Thus, these interactions will be of higher importance for Europe in future scenarios because of the shifting of

the transition zones due to climate change. The land-surface coupling will increase for Central Europe and Eastern Europe (Seneviratne et al., 2006b; Jaeger and Seneviratne, 2010; Fischer and Schär, 2010) and with this also the chance of intense HWs, which have to be correctly simulated by the climate models.

3. Sensitivity of 2m summer temperatures to changes in the soil initialization

3.1. Introduction

Within a climate model, coupled processes between the soil, vegetation and atmosphere are simulated with a Land Surface Model (LSM). The interactions are described by, e.g. Seneviratne et al. (2010). Several attempts have been conducted to develop sophisticated LSMs realistically reproducing these processes. Some well-known examples of LSMs are the Community Land Model (Oleson et al., 2013), NOAH (Niu et al., 2011), JSBACH (Roeckner et al., 2003), TERRA_ML (Schrodin and Heise, 2002) and VEG3D (Schädler, 1990), which are implemented in earth system models (ESM) or can be coupled to regional climate models (RCM). The importance of and interest in these models is reflected by various model intercomparison projects like the AMMA land surface model intercomparison project (ALMIP, Boone et al. (2000)), the Global Land Atmosphere System Study (GLASS, van den Hurk et al. (2011)) or Weisheimer et al. (2011).

Model drifts are one of the many error sources in climate projections and predictions and can be reduced if realistic and balanced soil conditions are provided. LSMs are the lower boundary of the larger model systems and need to be initialised realistically especially the deeper layers. The time required to approach an equilibrium of soil and land surfaces in LSMs varies regionally reliant of land use and soil type, soil moisture availability and vegetation cover (Koster et al., 2004; Koster and Suarez, 2001; Khodayar et al., 2015). Initial conditions affect the simulation via heat and moisture fluxes depending on climate zone and soil type (Khodayar et al., 2015, 2013). Since information diffuses upward as well as downward in the soil, a wrong initialisation of the deeper layers may cause unrealistic soil moisture and temperature for years. Therefore, a balanced soil initialisation can be crucial for the quality of decadal predictions like the MiKlip project (see section 4.1) (Dirmeyer et al., 2013; Khodayar et al., 2015).

Seneviratne et al. (2010) pointed out the available soil water is decisive for the evapotranspiration rates and in such for the soil moisture-temperature/precipitation feedback. Thus, accurate soil initial fields can improve temperature predictions noticeably (Ardilouze et al., 2017). Ruosteenoja et al. (2018) found a source of predictability in the soil moisture and fluxes relationship for periods up to half a year after a soil drying. In particular, this can be attractive for adaptation measures in case of extreme events like heat waves and droughts.
Over the last years, climate extremes became more pronounced (for HWs e.g. (Russo et al., 2015)) and their understanding and prediction one focus of the current research. Soil moisture

conditions can be crucial for the development of a HW (Rasmijn et al., 2018). Higher amounts of soil water can constrain the development of temperature extremes, while lower amounts amplify them. Therefore, the soil conditions have to be represented correctly by climate models. In addition, Quesada et al. (2012) identified asymmetric feedbacks of dry and wet soils on heat wave development, which in itself is a challenge to simulate. The misrepresentation of land-atmosphere feedbacks is a possible error source in model simulations and can trigger too strong and persistent HWs (Vautard et al., 2013). Lorenz et al. (2010) confirmed this by linking the persistence of HWs to long-term soil moisture variations (soil memory). Soil moisture availability will be most likely reduced for extensive areas due to the climate change (Samaniego et al., 2018). As a result, the probability for stronger and longer lasting temperature extremes increases drastically.

Thus, the impact of changed soil initial conditions on summer temperatures and their time scales are investigated in this chapter. One approach to generate balanced and nature like soil water and soil temperature fields for predictions is by running a stand-alone (SA) LSM with the most realistic forcing. The SA simulations driven by reanalysis data sets as forcing are conducted and their performance is evaluated by comparing their results with observations. Following that, the simulated soil temperature and soil moisture conditions are used to initialise coupled RCM simulations driven by the reanalysis data set ERA-Interim (Dee et al., 2011). The results of the run with stable soil conditions and a reference run are compared to validate the impact of the initialisation and to identify a possible added value of the more realistic initial soil moisture and temperature data. The results presented in this chapter have been published in Breil et al. (2018a).

3.2. Model components and simulation setup

Two kinds of simulations were performed for this study; the first are a set of stand-alone simulations with the land surface model (LSM) VEG3D at 55 km, 25 km and 7 km resolution for validation and to provide balanced initial soil data for regional climate simulations. The second runs are coupled simulations with CCLM coupled with Veg3d to analyse a possible added value of the improved initial data. This section describes the models as well as the external and forcing data sets and the setup of the stand-alone and the coupled simulations.

3.2.1. The Land Surface Model VEG3D

The multilayer LSM VEG3D is applied in the stand-alone simulations to produce soil moisture and soil temperature initial data for Europe. VEG3D was developed at Karlsruhe Institute of Technology (KIT) (Schädler, 1990) and evaluated in several studies (Braun and Schädler, 2005; Kohler et al., 2012; Meissner, 2008). VEG3D solves the heat conduction equation for temperature and the Richards equation for soil water transport using finite-difference methods on ten non-equidistant soil layers. The layer thickness is variable and can be chosen as needed. The

heat conductivity is parametrised after Johansen (1977) for different soil types. The unsaturated soil hydraulic conductivity is based on van Genuchten (1980). A massless vegetation layer is implemented, the so-called "big leaf" approach, to calculate radiation and turbulent heat fluxes between the surface and the atmosphere. This layer has a canopy temperature and specific humidity derived iteratively from the canopy's energy balance. Based on these quantities, the vertical turbulent mixing is parameterised according to the Monin-Obukhov similarity theory. The root density parameterisation is described with an exponential distribution in older simulations and was adjusted to the root distribution of Schenk and Jackson (2003) as a new feature. The root zone has its maximum within 0-1.0 m and ends in a depth of about 2.0 m. The model has recently been adapted for parallel architectures and can be coupled via the OASIS3-MCT coupler to the RCM CCLM (Valcke (2013), cosmo interface: Will et al. (2017), tested in Breil and Schädler (2017)).

3.2.2. The Regional Climate Model COSMO-CLM

The impact of the changed initial fields is investigated in a coupled model system of the operational weather forecast model COSMO (COnsortium for Small-scale MOdeling (Rockel et al., 2008; Baldauf et al., 2011)) and VEG3D (Breil and Schädler, 2017). The regional climate model (RCM) COSMO-CLM (CCLM) is a three-dimensional, non-hydrostatic atmospheric model of the German Weather Service (DWD). The model uses a radiative transfer scheme based on Ritter and Geleyn (1992), the convection parameterisation of Tiedtke (1989) and solves the vertical turbulent diffusion with the prognostic TKE scheme from Raschendorfer (2001). The physical parameterisation is described in further detail in Doms et al. (2011) and the numerics and dynamics in Doms and Baldauf (2015). In this study, the simulations are performed with the version with cosmo 5.0 clm6. In the standard CCLM, the multilayer soil model TERRA-ML is included to describe the soil-vegetation-atmospheric interactions (Schrodin and Heise, 2002), which will be replaced by VEG3D.

3.2.3. Designing the stand-alone simulations

To operate the LSM Veg3d in a stand-alone (SA) mode, the minimal required input is the atmospheric forcing data (2 m air temperature and specific humidity, precipitation (snow and rain separately), incoming short- and longwave radiation, 10m wind speed and surface pressure) and the external data (land use class and soil type distribution). The SA simulations cover different spatial resolutions and the desired time from 1961–2010. All SA data runs are performed with a time step of 300 s and ten soil layers until a depth of 15 m (layer boundaries at 0.02 m, 0.05 m, 0.1 m, 0.2 m, 0.5 m, 1.0 m, 2.0 m, 5.0 m, 10.0 m, 15.0 m).

The first simulation is designed for the EURO-CORDEX domain with $0.5°$ (~55 km) horizontal resolution covering the time from 1960–2010. It was performed in two stages with forcing data from the WATCH project (Weedon et al., 2010). The WATCH forcing data sets contain a

composition of atmospheric variables needed to drive hydrological and or soil and vegetation models (Weedon et al., 2010, 2011, 2014) based on the ECMWF Reanalysis data sets ERA40 (WATCH WFD) and ERA-Interim (WATCH WFDEI). The first stage includes the time from 1960–2001 and is driven by WATCH WFD. The forcing was shifted to WATCH WFDEI in 1990 to simulate the years until 2010. A higher resolved simulation (0.22°) for the same domain was conducted using the twentieth-century reanalysis data ERA20C (Poli et al., 2016) for a period of 1955–2010. A third simulation was performed merely for Central Europe (see Fig. 3.1, Area 6) driven by interpolated ERA-Interim data with a very high resolution of 0.0625°for the period from 1979 until 2010. In the last two runs, a new root density distribution based on Schenk and Jackson (2003) is tested to include measured root depths. In the new profile, the maximum of the root density is located in shallower layers from 0.1 m until 0.5 m, while the old one has the maximum from 0.5 m until 1.9 m for most land use classes including coniferous forest and urban. The new profile required a redistribution of the soil layer. The differences in the vertical profile are not as relevant since means/sums over several layers are used in the comparison.

All simulations have a seven-year spin up period repeating their starting year allowing them to achieve nearly balanced soil conditions. The only exception is the WATCH WFDEI run. This simulation is re-initialised in 1990 with the balanced soil of the WATCH WFD, which was running for 37 years at this time. Therefore, only a two-year spin up was performed.
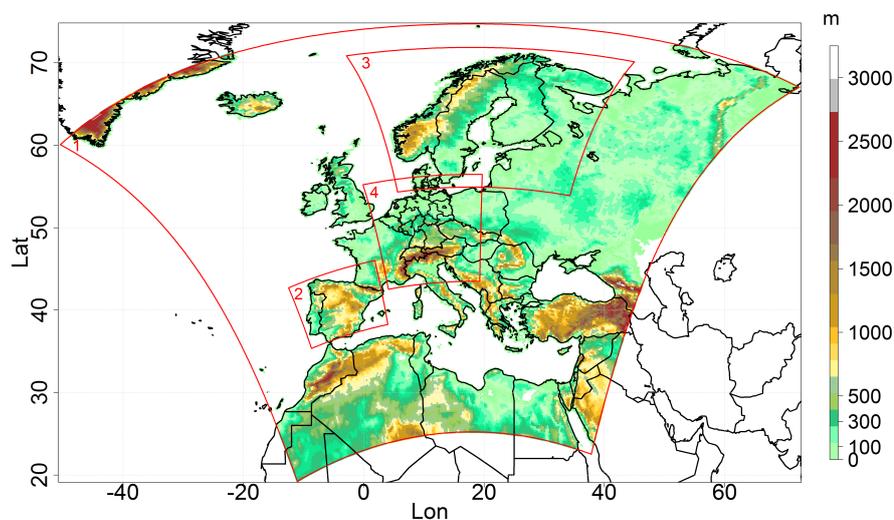
Figure 3.1.: Model domain for the all European VEG3D SA and coupled CCLM-VEG3D simulations (1) identical to the EUROCORDEX domain (Jacob et al. (2014)) and evaluation areas: (2) Iberian Peninsula, (3) Scandinavia and (4) Central Europe (also simulation domain of the 7km VEG3D SA ERA-Interim simulation).

3.2.3.1. Comparing the forcing data sets

The forcing data sets were evaluated against each other to test if large deviations exist between the WATCH data based on ERA40 and ERA-Interim data, the original reanalysis data and the newer ERA20C. The comparison of monthly field means of the three driving data sets is shortly

summarised in the following with special regard to the overlapping period 1990 - 2001 for Central Europe (area 4 in Fig. 3.1). Most of the variables demonstrated an excellent consistency for this period with a relative difference of less than 1 % to the WATCH data sets. These small deviations are presumably the result of the interpolation to different grids and resolutions. It was noted, that the WFD and WFDEI are slightly warmer than the ERA-Interim ($<$ -0.4°C) and the ERA20C ($<$ -0.7°C). More significant deviations are detected for shortwave radiation, precipitation and wind speed. The downward shortwave radiation is lower for the WATCH WFD data due to the general difference between ERA40 and ERA-Interim and a double correction of the ERA40 data within the WATCH WFD Weedon et al. (2014). The annual cycle of the downward shortwave radiation shows a deviation of -15 W/m^2 for the winter months and up to -50 W/m^2 in summer for Central Europe. On the other hand, the differences of the WFDEI, the ERA-Interim and the ERA20C are indeed smaller (-/+ 1 W/m^2). The precipitation is found to be higher in both WATCH data sets by nearly 10 % compared to the other reanalysis data sets. The reason for the deviation is the rain and snowfall correction during the production of the WATCH forcing data (Weedon et al., 2010). Additionally, higher monthly mean wind speeds were found in the WATCH data. In regard of the notable differences in the shortwave radiation of the WFD to the other forcing data sets, the results of WFD forced VEG3D SA run are not included in the following evaluation of the SA simulations.

3.2.3.2. External data sets

As external data sets for the SA simulations, 2D soil type fields are produced from the Harmonizes world soil data set (HWSD Fischer et al., 2008) aggregated to 0.5°, 0.22° and 0.0625° resolution. The land use classes are extracted from the GLC2000 land surface data set (Bartholomé and Belward, 2005).

(a) Soil Type   (b) Land use
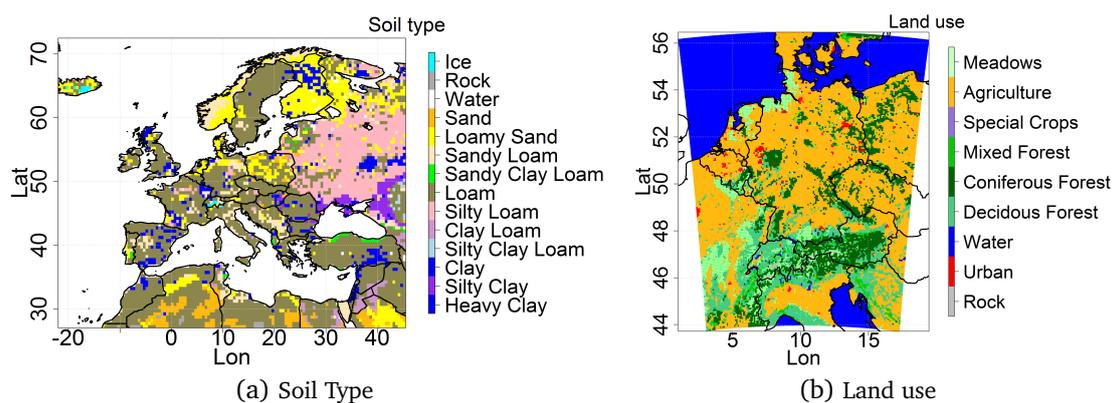
Figure 3.2.: HWSD soil types (resolution 0.5°) for the EURO-CORDEX area and GLC2000 land use classes (translated to the Veg3d classes) for Middle Europe with a resolution of 0.0625°.

The 23 classes from the GLC2000 data are translated into ten land use classes of Veg3d displayed in Fig. 3.2b for Middle Europe. The main land use class is agriculture in this area as in Southern

17

Europe, while the land is covered by forest in Northern Europe and mountainous regions. The topsoil from the HWSD data set is shown in Fig. 3.2a. Loam and clay loam soils dominate Europe with smaller areas of clay and loamy sand, in particular in Scandinavia. As Veg3d uses the same soil types, therefore, no further interpolation/translation is needed.

3.2.4. Setup of the coupled simulations

The impact of the changed initial soil conditions is investigated with two decadal simulations of the coupled CCLM-OASIS3-MCT-VEG3D. This coupled system is driven by ERA-Interim at the lateral and lower boundary conditions starting at January 1st 2001 (Dee et al., 2011). A possible added value due to the soil initialisation can be assessed more easily by using ERA-Interim reanalysis data as forcing since boundary conditions are applied including the most realistic climate states. In the first run, the soil is initialised by the soil moisture and temperature fields of ERA-Interim. These fields are most likely not fitting the balanced state of VEG3D as they are conducted with a different LSM. The comparison of the initial date, 01.01.2001, has confirmed this. Thus, we expect a certain model drift in this simulation. This simulation is taken as the reference, called CCLM_VEG3D_REF in the following sections. In the second run, the initial soil conditions of ERA-Interim are replaced by the simulated soil moisture and temperature values of the SA run at the starting date (CCLM_VEG3D_INIT), see Fig. 3.3. The model domain of the SA simulation and the coupled runs is identical to the EURO-CORDEX domain (see Fig. 4.1) with a horizontal grid spacing of 0.22°. The vertical direction of the atmosphere is discretised in 40 levels. The coupling time step, as well as the time steps of both models, are 150s.
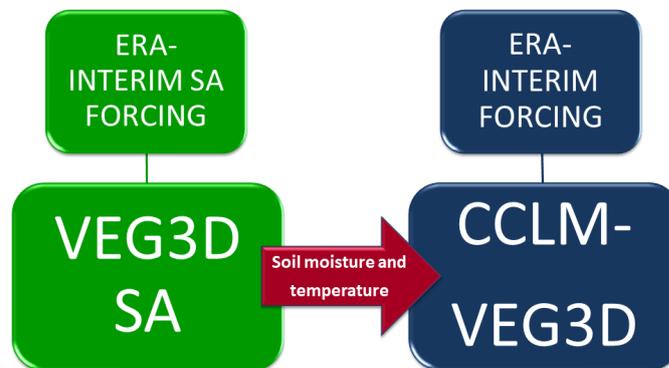
Figure 3.3.: Scheme of the interaction of VEG3D SA and CCLM-VEG3D coupled simulation

3.3. Validation data

Two kinds of data sets are applied for the validation: observations from the EFDC stations and spatial data derived from other LSMs or interpolated observations. The comparison is performed for soil water content, soil temperature as well as the latent and sensible heat flux for an overlap period 1990–2010. The data sets are described in more detail hereafter.

3.3.1. Site data: FLUXNET/EFDC Stations

The European Fluxes Data Cluster (EFDC) collects data from several databases, projects and measurement sites across Europe beginning in 1996 (EFDC, 2016). The stations included in the FLUXNET project (ORNL DAAC (2015), mainly from projects EuroFlux and CarboEuro), contribute data to the cluster as well as the CarboEuropeIP (successor of CarboEuro), IMECC and latest the GHG-Europe projects. The data coverage varies strongly between sites; the longest period of the measurements used for this study is 1996–2010. EFDC, as well as FLUXNET quality, controlled the data, but dealing with missing values remains unavoidable. In addition, crucial information about the measurement depths for soil moisture and soil water content are not clear and/or available. Assuming the first measuring depth of the sensors is < 30 cm, the model layers between the surface and 20 cm (second and fourth model layer) were averaged before the comparison to observations. Stations have been selected across Europe satisfying three criteria, longest available time, continuously data coverage and available variables. Some of these stations, e.g. the Italian Malga Arpaco which has a negligible number of missing values are problematic for other reasons. Altitude differences between the model and the station height induce significant deviations for mountain measurement sites in the temperatures. As a result for Malga Arpaco, which is located at 1730 m, a RMSE of $> 7.0°$C is produced by the SA simulation for the soil temperatures. Thus, stations under similar conditions are only considered as an alternative. A total of 34 stations has been selected for Europe finally. An additional complication remains the difference between the external data, land use and soil type, used in the model simulations and the actual land use class and soil at the station. Thus, differences between the SA simulations and the station can be expected due to the deviations of one or both external parameters.

3.3.2. Gridded data sets

Three data sets are employed for the spatial comparison: the ERA-Interim reanalysis data from the European Centre for Medium-Range Weather Forecasts (ECMWF), the ESA CCI soil moisture (ESA_CCI) satellite data and gridded soil data provided by the German weather service (DWD). The ERA-Interim and the DWD data represent products of other LSMs; therefore, they are not considered as observations.
The land surface scheme Tiled ECMWF Scheme for Surface Exchanges Over Land TESSEL (Dee et al., 2011; Viterbo and Beljaars, 1995) is employed for the production of the **ERA-Interim** data running on four soil layers from 0–0.07 m, 0.07–0.28 m, 0.28–1.00 m, 1.00–2.89 m (Albergel et al., 2013). The tile approach allows different vegetation types (van den Hurk et al., 2000) for one model grid box. On the other hand, it includes only a single soil texture (medium texture corresponding to a loamy soil) applied to all grid boxes, which was critically discussed by Balsamo et al. (2009). Monthly ERA-Interim data are interpolated to 0.22 °for the overlapping period of 1990-2010.

The German weather service (DWD) provides gridded soil moisture and evapotranspiration data using the agrarmeteorolocigal model **AMBAV**. A description can be found in Löpmeier (2014). AMBAV is driven by observations and simulates the soil water content (for 0-0.6m) and evaporation for weather stations. The model is set up with only one land use class, grass, and one soil type: sandy loam. The soil temperature is simulated by the soil-plant-atmosphere model **AMBETI** (Braden, 2012) in 0.05 m. These data sets are interpolated to a 1x1km Gauss Krüger grid available from 1991 onward. Monthly data were interpolated to the 0.0625 °grid for the validation of the high-resolution VEG3D run.

The ESA CCI daily soil moisture (**ESA_CCI**) data set is available with a spatial resolution of 25km from 1978–2014. It combines microwave sensor results of passive products from November 1978 and active products from August 1991 (D. Chung et al., 2015; Dorigo et al., 2015). The global data give information only about the uppermost centimetres of the soil and the data are quite incomplete during the first years. Nevertheless, this record is one of the most extended global data sets of observed soil moisture that exists.

3.4. Validation of VEG3D stand-alone simulations

Firstly, the simulations are evaluated against the EFDC stations; secondly, spatial comparisons to the available validation data sets are presented. One focus is on the quality of the forcing data, the resolution and the external data affecting the results of stand-alone simulations in this analysis.

3.4.1. Comparison to EFDC stations

The stations are sorted and averaged to condense the results of the 34 selected stations reasonably and to highlight the effect of the external data sets with a focus on the land use classes and the soil types. Four categories are established classifying the agreement between the external data of the SA run and the measurement site: Fits (land use class (LU) and soil type (ST) fit), Mixed (only one fit), Non-Fitting (both different) and All (including all stations). Six stations have matching land use and soil types in the 55km external data, while 17 are Non-Fitting and eleven missing one category. In the 25 km simulation (VEG3D ERA20C), 6 fits are registered, 17 are Mixed and 11 Non Fitting. The higher resolution of 7km increased the number of stations with fitting external conditions to seven but included a more balanced ratio between Mixed (14) and Non-Fitting (13) as well. The RMSE and the correlation between the VEG3D SA simulation and the station have been calculated for monthly data of the soil moisture and temperature as well as the latent and sensible heat flux for all categories.

The RMSE and correlation for the **soil moisture** are shown in Fig. 3.4a and Fig. 3.4b. For all stations, the highest RMSE is found for VEG3D ERA-Interim , the lowest for VEG3D WFDEI. On the other hand, the highest correlation is seen in VEG3D ERA-Interim , whereas VEG3D ERA20C

shows an anomalously low correlation value. The correlation ratio between the VEG3D simulations is similar for all categories, in contrast to the RMSE results which are very heterogeneous. One reason for the low correlations found for the ERA20C could be the higher soil moisture found in this simulation. Compared to the other SA simulations, its averaged soil water content is 30.8 % compared to VEG3D ERA-Interim with 25.5 % and VEG3D WFDEI with 27.4 %. However, the comparison of the forcing showed noticeable correlation differences between the ERA20C data and the WATCH data and the ERA-Interim. The trend of the forcing is transferred to the SA simulations and can reduce the correlation values in comparison with the stations. Since both impacts are not clear to distinguish, it remains unclear, which exerts the more considerable effect on the correlation.

(a) RMSE

(b) Correlation

(c) RMSE
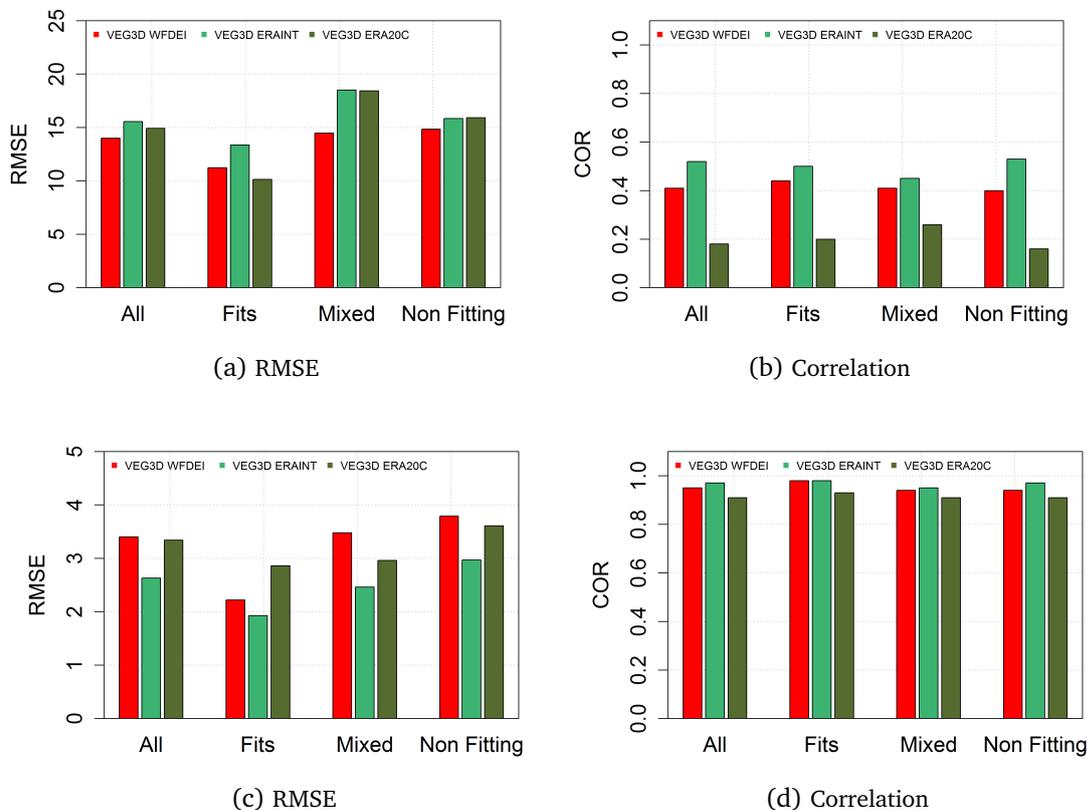
(d) Correlation

Figure 3.4.: Soil water content (upper row) and soil temperature (lower row) RMSE (left) and Correlation (right) of the SA simulations calculated to the EFDC stations for the whole year for the layer 0.01-0.2m on basis of monthly means.

The lowest RMSE and the highest correlation is found for the **soil temperature** in the VEG3D ERA-Interim simulation. As for the soil moisture, the lowest correlation across all categories is again seen in VEG3D ERA20C, while the RMSE for All stations is similar between this simulation and VEG3D WFDEI. Notably higher correlations (r > 0.8) exists for the soil temperatures compared to the soil water content, but the variation between and within the categories is lower.

The highest temperatures are found in VEG3D WFDEI. The lower temperatures of VEG3D ERA-Interim and VEG3D ERA20C are in better agreement with the observations demonstrated by the lower RMSE (see fig. 3.4c). An evaluation of the annual cycle revealed overly high annual amplitude for all SA simulations with overestimated summer temperatures, while the winter temperatures are underestimated by far in VEG3D ERA20C. The smallest RMSE is found for the Hits for VEG3D ERA-Interimincreasing with less fitting external data for all simulations.

In the following, the results of the comparison are displayed for the **latent** and **sensible** heat flux (Fig. 3.5a until Fig. 3.5d). The smallest RMSE for all categories is found for the latent heat flux for VEG3D WFDEI, while VEG3D ERA-Interim has the lowest RMSE for the sensible heat flux (except for the Non-Fitting). High correlations for the latent heat flux ($r > 0.8$) are noted for all simulations with highest values in VEG3D ERA20C. The sensible heat flux shows marginally lower correlation values and larger variations between the runs.

(a) Lat RMSE

(b) Lat Correlation

(c) Sen RMSE
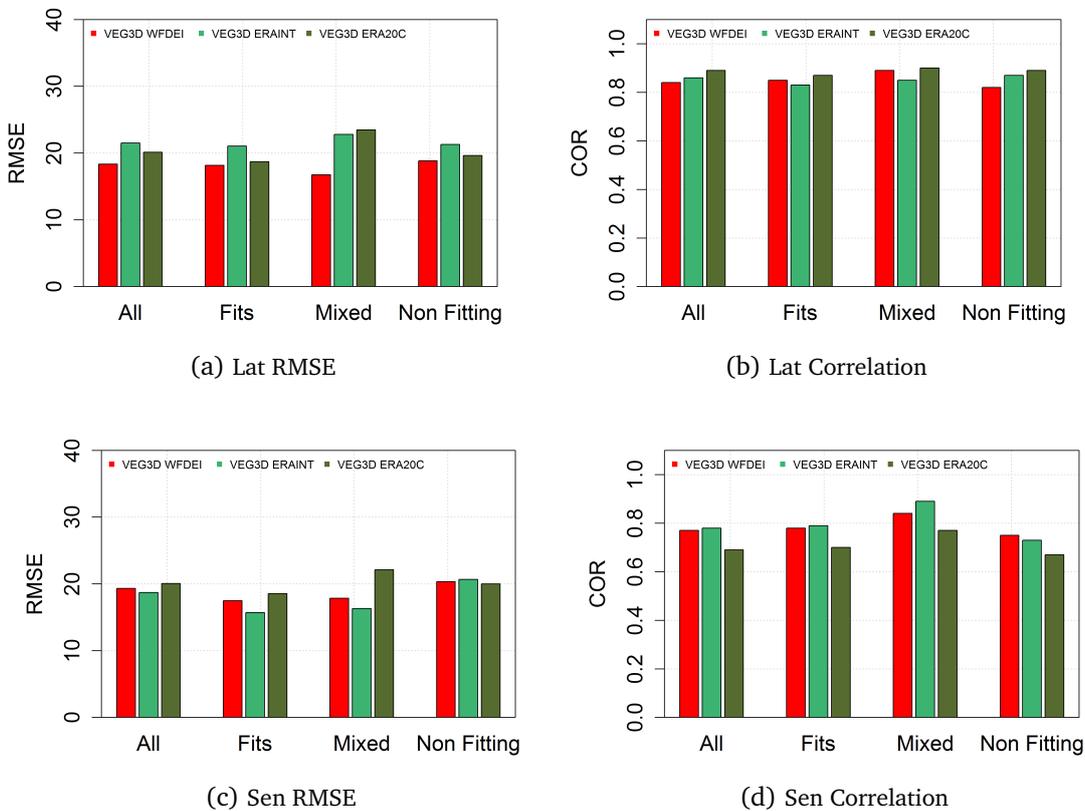
(d) Sen Correlation

Figure 3.5.: Latent (upper row) and sensible (lower row) heat flux RMSE and Correlation of the SA simulations calculated to the EFDC stations on basis of monthly means.

As for the soil moisture, the best result of the RMSE is achieved with VEG3D ERA-Interim for the sensible heat flux. Unlike the soil temperature, no increasing RMSE with less agreement to the external data is visible. This leads to the conclusion that the external data sets remain not the most critical factor for the fluxes.

In most cases, a reasonably good agreement with the stations has been found with correlation values > 0.7 and RMSE $< 25~W/m^2$ for the latent and sensible heat flux. This agrees with results from other studies like Stöckli et al. (2008) or Zink et al. (2017), who applied the Community Land model or the mesoscale meteorological model mHm in an analogue experimental design. Both studies had the advantage of using more fitting forcing data, like the higher resolved observation data with a resolution of 4 x 4 km or on-site measurement data, compared to the presented simulations with Veg3D.

The comparison with the 34 stations demonstrated no significant deviations between the number of fitting stations between the different horizontal resolutions. The considerable variation of the RMSE and the correlation indicate that the external data are not the only influence, although they are not to be underestimated. The knowledge about the correct land use class and soil type at the side can be crucial for the simulation results of the LSM beside the quality of the forcing data. To assess the impact of the atmospheric forcing and the external data statistically, a Kruskal-Wallis-Test is performed (also named H-test). The independence of groups can be quantified with this non-parametric rank sum test (in this case, the forcing and the external data). Different significance levels have been tested, and the result is shown in table 3.1.

	soil water content	soil temperature	latent heat flux	sensible heat flux
RMSE vs. forcing	-	-	90	-
RMSE vs. ext.	90	90	-	-
COR vs. forcing	95	90	-	-
COR vs. ext.	-	-	-	-

Table 3.1.: Significance level achieved from the Kruskal-Wallis Test testing the independence of the groups: forcing and external data (ext.) to the RMSE and the correlation for four variables.

The external data or the forcing have a significant impact on the simulation results depending on the variable. The external data are the critical factor for the RMSE and with this the bias of the soil variables. However, the forcing remains the most crucial factor for both soil variables to achieve a good correlation to the observations. The fluxes depend only indirectly on the external data. The forcing is not significantly essential for the RMSE and correlation; the only significant combination is RMSE and forcing for the latent heat flux.

3.4.2. Comparison to the gridded validation data sets

In this section, the stand-alone simulations are compared to the validation data sets ERA-Interim Reanalysis, ESA CCI and DWD AMBAV data for the area Central Europe (no. 4, see Fig. 3.1). Whereas ERA-Interim and AMBAV data are model data contain assimilated observations, ESA-CCI data are derived from satellites. The comparison is performed for the upper hydrological layer until 0.5m for 1991-2010 since the satellite data only take into account the first cm of the soil and the AMBAV and the reanalysis data do not possess deep levels as the VEG3D simulation.
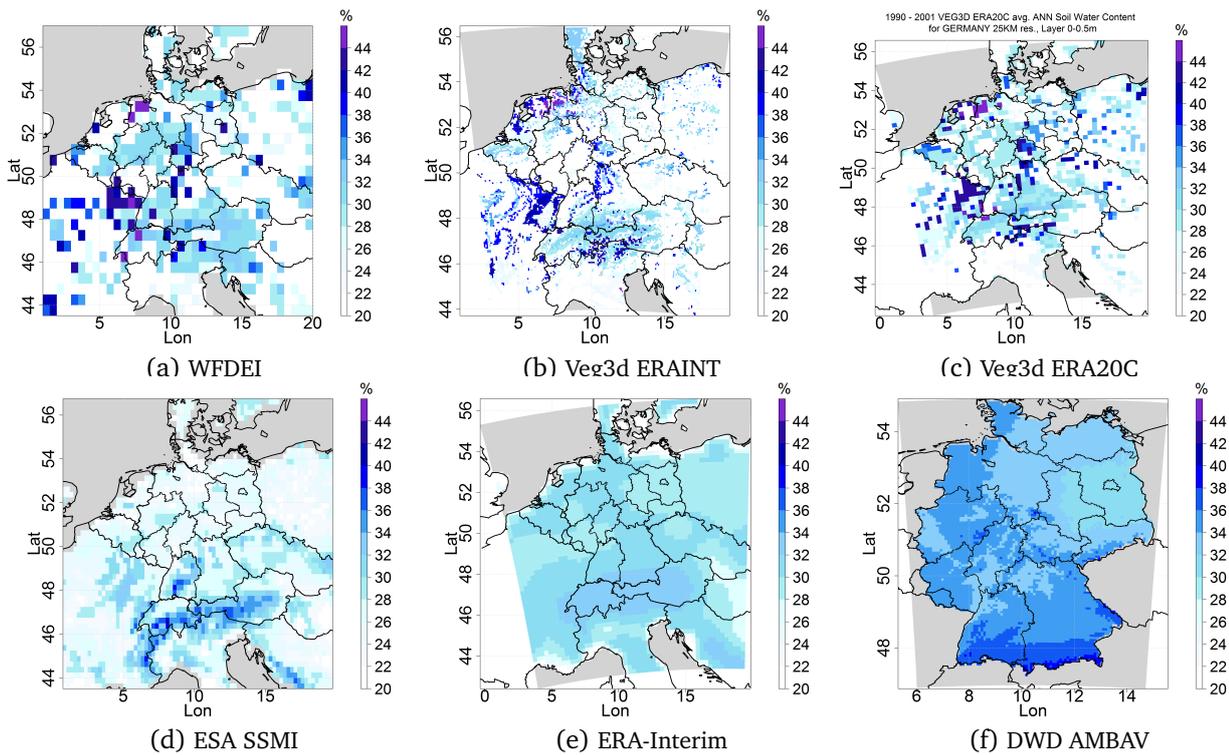
Figure 3.6.: Annual weighted mean Soil water content for 1991 – 2001 VEG3D SA simulations, ERA-Interim reanalysis data, ESA-SSC/I satellite data and DWD AMBAV data for 0-0.5m.

The annual mean **soil water content** is displayed for the VEG3D simulations in Fig. 3.6a, Fig. 3.6b and Fig. 3.6c. All three show similar spatial structures with higher soil moisture along the coast of the North Sea, in the Vosges and the Alpine area. Much dryer regions are noticed for the Western and Eastern parts of Central Germany. It was expected, that regional structures would be better represented by increasing the spatial resolution. Only the 7km resolution shows topographical structures to a certain degree. However, the spatial distributions of the soil water content seem to represent the resolution of the land use classes and the soil types solely. Other geographical features as mountain ranges or else, found in the satellite data (Fig. 3.6d), are not resolved by the VEG3D simulations. This might a result of the coarser resolution of the ERA-Interim forcing (80 km) and the strong dependence of soil moisture variations to different land use classes and soil types in this model. Expected orographic structures, like the Swabian Alp or the Erzgebirge, are detected for the DWD AMBAV data since this model uses only one soil type and one land use class. Additionally, the forcing includes a much finer resolution of about 5km (Fig. 3.6f). Although the ERA-Interim and the AMBAV data have different soil moisture levels, their spatial patterns are strikingly similar. The TESSEL LSM has only one soil type as well. Thus, the forcing patterns become dominant in the spatial distribution of the soil water content in the ERA-Interim data. This shows the importance of proper representation of the soil and vegetation variety in LSMs, as Balsamo et al. (2009) noted also.

The AMBAV displays the wettest soil by far from all displayed models and the satellite data. AMBAV, as described earlier, uses the land use class "grass" and "loamy sand" as soil type merely, which is one reason for the higher mean soil water content. This is examined by filtering the other data sets into grid boxes with the land use class "grass/meadows". The result is presented in table 3.2. After the filtering, the mean soil water is higher than the mean soil water content for all grid points resulting in a reduced mean difference and RMSE to the DWD AMBAV. In particular, this increase leads to a reduction of the RMSE up to 30 % for VEG3D ERA20C and VEG3D ERA-Interim. In addition, the time series analysis revealed, that all data sets have a nearly identical trend and a very similar correlation to the DWD AMBAV data for the annual means (not shown). Also, years with very low soil moisture like 2003 are identified in nearly all data sets.

	WFDEI	ERA20C	ERAINT NR	ESA-CCI	ERA-INTERIM	DWD_AMBAV
ANN	21.6 ↘ 20.1	22.1 ↗ 27.4	18.2 ↗ 24.1	26.9 ↗ 27.4	29.4 ↗ 30.9	33.9 ↗ 34.6

Table 3.2.: 1990 – 2010 Mean Soil water content for the various data sets for the grid boxes with the land use class grass or meadows.

The interpolated DWD AMBETI data have the highest annual mean **soil temperatures** with > 10 °C for the upper layer. Since these data are obtained at a depth of 0.05cm, higher values and large amplitude of the annual cycle are expected. The VEG3D simulations deviate not much (< 0.7°C with VEG3D WFDEI $>$ VEG3D ERA-Interim $>$ VEG3D ERA20C) from the AMBETI mean and their trend is in good agreement with the DWD data. Analysis of the diurnal cycle showed that lower daily maximum temperatures occur in VEG3D ERA-Interim and VEG3D ERA20C (diff. $T_{max} > 2.0$ °C) explaining the lower mean annual soil temperatures. Additionally, VEG3D ERA20C has considerably lower soil temperatures during the winter, which are a result of the lower 2m temperatures of the ERA20C forcing. Overall, the spatial distributions of the soil temperature fields are very similar between the data sets with lower temperatures for mountainous ranges and higher temperatures in the lowlands. The spatial soil temperature distributions do not reflect variations due to the land use classes and the soil types as the soil water content.

The annual, summer and winter half-year means for the VEG3D simulations, the DWD AMBAV and the ERA-Interim reanalysis data are displayed for the **latent** and **sensible** heat flux in table 3.3 for 1991–2010. From all SA simulations, the highest latent heat flux is found for VEG3D ERA20C and VEG3D ERA-Interim in the summer half-year (SHY) ($> 60 \, W/m^2$). In the winter half-year (WHY), the differences between the runs are not as high since the fluxes are lower in general. The higher fluxes of VEG3D ERA-Interim and VEG3D ERA20C fit very well to the ERA-Interim reanalysis data. VEG3D WFDEI fits remarkably well to the lower flux from the DWD AMBAV data. The sensible heat flux differs not by much between VEG3D WFDEI and VEG3D ERA20C, unlike VEG3D ERA-Interim, which has a noticeable higher sensible heat flux. The sensible heat flux is moderately increased (mainly in the summertime) as the ERA-Interim forcing is warmer compared to the ERA20C combined with a dryer soil.

	VEG3D WFDEI	VEG3D ERA20C	VEG3D ERA-Interim	ERA-INTERIM	DWD
Latent heat flux					
ANN	33.7	42.5	42.3	47.8	36.7
Summer	49.2	63.8	62.9	71.6	53.0
Winter	18.0	21.2	21.6	24.1	20.4
Sensible heat flux					
ANN	9.8	11.7	12.1	14.2	
Summer	20.9	22.9	26.3	27.8	
Winter	-1.2	0.5	-2.0	0.6	

Table 3.3.: 1990 – 2010 Latent and sensible heat flux for the SA Veg3d simulations and the ERA-Interim reanalysis data averaged for Central Europe.

3.5. Results of coupled model runs

The focus of this section is on the possible improvements due to a balanced soil initialisation. Two decadal simulations driven by ERA-Interim have been performed, one with the soil fields from ERA-Interim (CCLM_VEG3D REF) and one with the initial fields conducted by a VEG3D SA simulation (CCLM_VEG3D INIT). Different evaluation domains have been selected to show a possible added value for diverse climates and soil conditions. Regarding the dependence between HWs and dryer soil conditions, the focus of this evaluation is on the summer months.

3.5.1. Changes in the initial fields after implementing the SA generated data

The field averaged soil water content derived from the ERA-Interim is homogeneous for all layers for Europe at the initial time, while the soil from the SA run is wetter in the upper layer and dryer in depths lower than 0.2 m. The minimum soil moisture is found at the depth with the highest root density in the new initial fields. The temperature is higher for most of the soil layers, except for the climatology layer. The spatial differences have been analysed for Europe for two layers, 0-1 m and 2-10 m. The SA simulations reveal a higher spatial variability than the interpolated ERA-Interim fields. The heterogeneity arises from a stronger impact and the higher number of soil types and land use classes in the SA run. The relative differences between the VEG3D SA and the ERA-Interim soil water content reach over 50 % for some regions, where the new initialisation is mostly dryer. The new soil temperature initialisation is > 2.5 °C warmer for the Central Europe mountainous regions and Scandinavia, while Eastern Europe shows no substantial differences. The new initialisation shows fewer variations for the soil water content and the soil temperature in the lower levels. In general, the new fields are dryer and warmer than the original ones.

(a) Soil water content   (b) Soil temperature
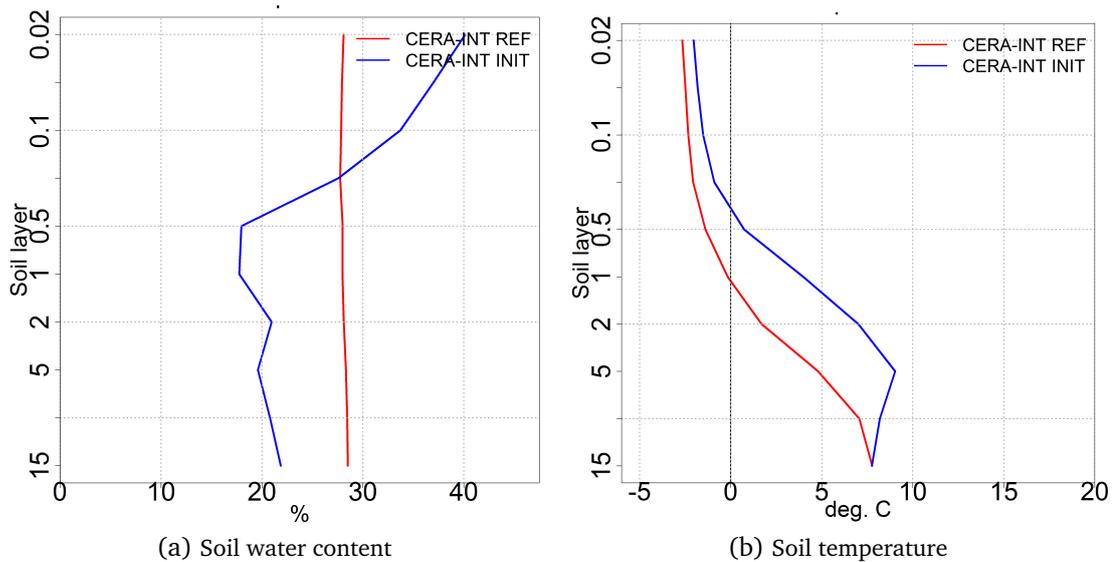
Figure 3.7.: Vertical soil profile for soil temperature and soil water content for the initialization time 01.01.2001 for Middle Europe.

3.5.2. Sensitivity to changed initial soil fields in decadal simulations for different European regions

A comparison revealed distinct deviations for the 2m temperature resulting from differences in the soil-vegetation-atmosphere-transfer triggered by the new initialisation fields for Central Europe (not shown). In addition, the effect of the initialisation varied on region and time scale, in particular, JJA 2m temperatures improved in Northern Italy and Southeastern France in the first three years. Therefore, the whole simulation domain is evaluated to identify hot spots with the most prominent impact of the initialisation. Also, the significant variability of the European simulation domain is a key factor since various climate zones are included and in such different soil equilibrium states. Beside the European area, the Iberian Peninsula (IP) and Scandinavia (SC) will be evaluated separately to demonstrate these differences explicitly. Additionally, the evaluation is performed for two periods 2001-2005 and 2006-2010 to differentiate between short-term and long-term changes.

The precipitation sums RMSE differences between CCLM_VEG3D INIT to EOBS minus CCLM_VEG3D REF to EOBS for JJA and DJF are displayed in Fig. 3.8a until Fig. 3.8f for the start year and the two averaging periods 2001-2005 and 2006-2010. The RMSE differences for the summer and the winter are quite contrary. While considerable RMSE differences with a mostly chaotic structure are visible in summer, especially in the initial year, the winter shows nearly no RMSE difference at all. The precipitation in the winter months is of synoptic origin; therefore, a change in the initial fields has a modest impact at the most. However, the summertime rainfall is highly related to convection and more sensitive to changes in the soil conditions.
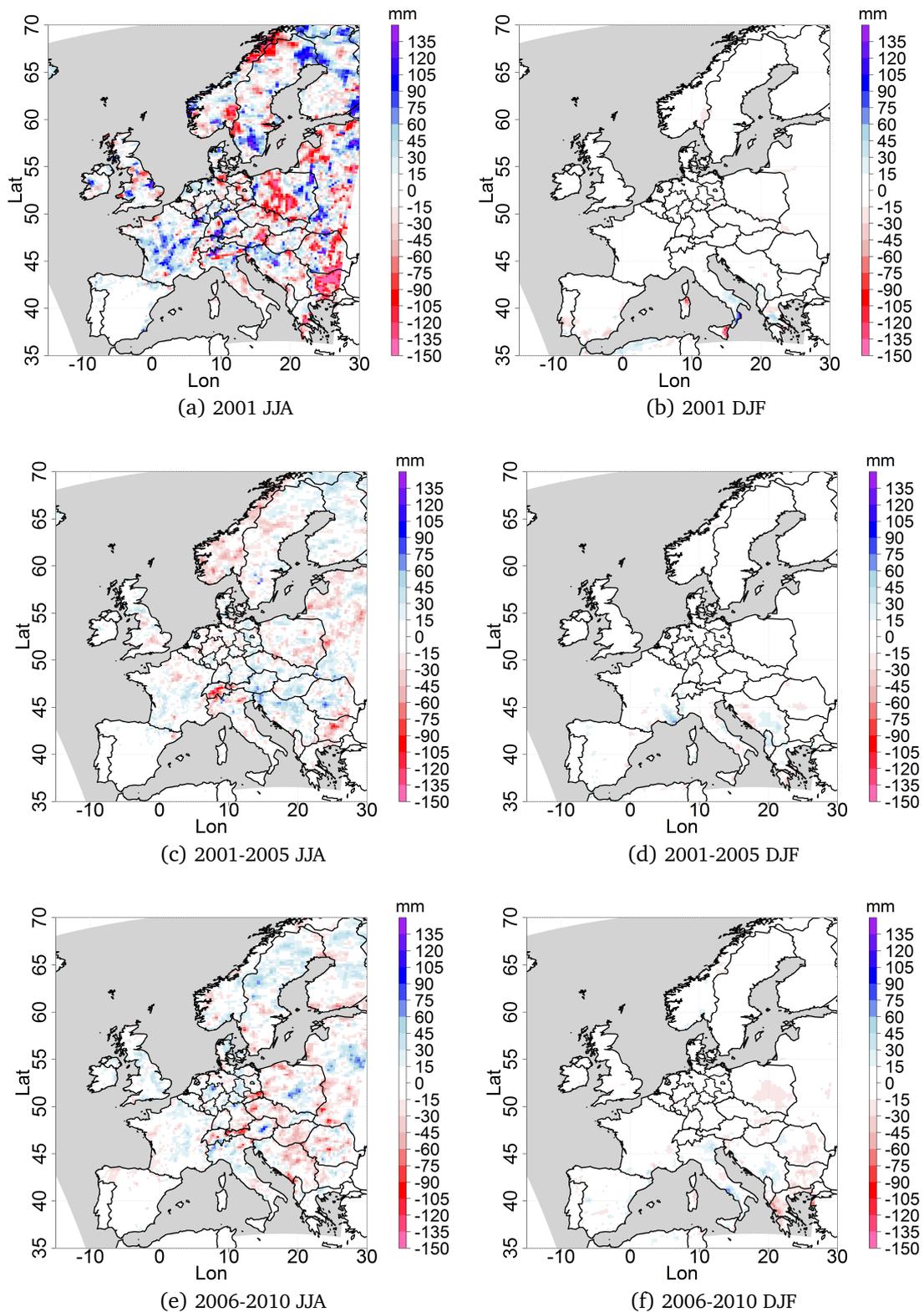
Figure 3.8.: 2001 (upper row), 2002-2005 (middle row) and 2006-2010 (lower row) JJA (left side) and DJF (right side) precipitation sum RMSE differences of CCLM_VEG3D INIT - EOBS to CCLM_VEG3D REF - EOBS for Europe.

The largest RMSE differences are found for the first year but as Fig. 3.8c and Fig. 3.8e show, still at least RMSE differences of about +/-50 mm occur in the long-term period as well. Overall, more or less randomly distributed differences show no clear improvement for the precipitation due to the soil initialisation.

The RMSE differences of the temperature show a reduced RMSE in Southern Scandinavia and the Mediterranean, including IP, for CCLM_VEG3D INIT in the year 2001 JJA (Fig. 3.9a). Contrarily, Central Europe remains nearly unaffected and no larger RMSE differences are noticed. The Balkan is the only region were the new initialisation lead to an increased RMSE. The RMSE differences depict a sign change and a marginally higher RMSE in the CCLM_VEG3D INIT simulation is found in nearly all Europe except for Northern Sweden for the long-term period (Fig. 3.9e). In this region and Southeastern Europe, improving 2 m temperatures are visible in DJF during the start year highlighted by a considerably reduced RMSE. 2001-2005 shows an increased RMSE for nearly all Europe (Fig. 3.9d), whereas such an increase is found only in Central Europe and Southern Finland in 2006-2010 (Fig. 3.9f).

Thus, improvements due to the changed soil initialisation towards a more balanced state are observed for the summertime but not in the winter months. As expected, IP and SC turned out to be sensitive to soil moisture and soil temperature changes. Additionally, the Balkan region showed a considerable impact, which was seen partly in the comparison for Central Europe. The cause for the notable RMSE increase cannot be thoroughly comprehended due to several reasons. It is known from other studies like Kotlarski et al. (2014), that CCLM and other RMCs produce an obvious warm bias in this region. Contrary to this, one can assume that fewer reliable measurements are included in the Eobs database in this region. This can lead to significant differences to the actual conditions during the interpolation to the European grid, since a lower number of stations are used as well, compared to Germany for example. Therefore, only IP and SC are evaluated in more detail henceforth.

The variations in the soil water content affect the 2m air temperature due to the coupling processes between soil vegetation and atmosphere. Lower heat capacity is found for a dryer soil due to the insulation effect of air in the pores of the soil. As a consequence, dry soil reacts thermally stronger to environmental changes like incoming precipitation or radiation changes (filling these pores by infiltrating water).

The soil variations reach a higher annual amplitude and with this higher summer and lower winter temperatures are possible in CCLM_VEG3D INIT. Additionally, the evapotranspiration rate is reduced due to soil moisture deficits. Less water is available for plants and the transpiration from the soil directly reducing the latent heat flux and in such the evaporative cooling effect significantly. In addition, more incoming energy is employed to increase the soil temperatures amplifying the sensible heat flux. The higher sensible heat flux leads to higher near surface temperatures as more energy heating the atmosphere is released. Since this process depends on the incoming radiation, it is more important during the summertime. The effects of named processes are demonstrated for the two evaluation areas IP and SC.
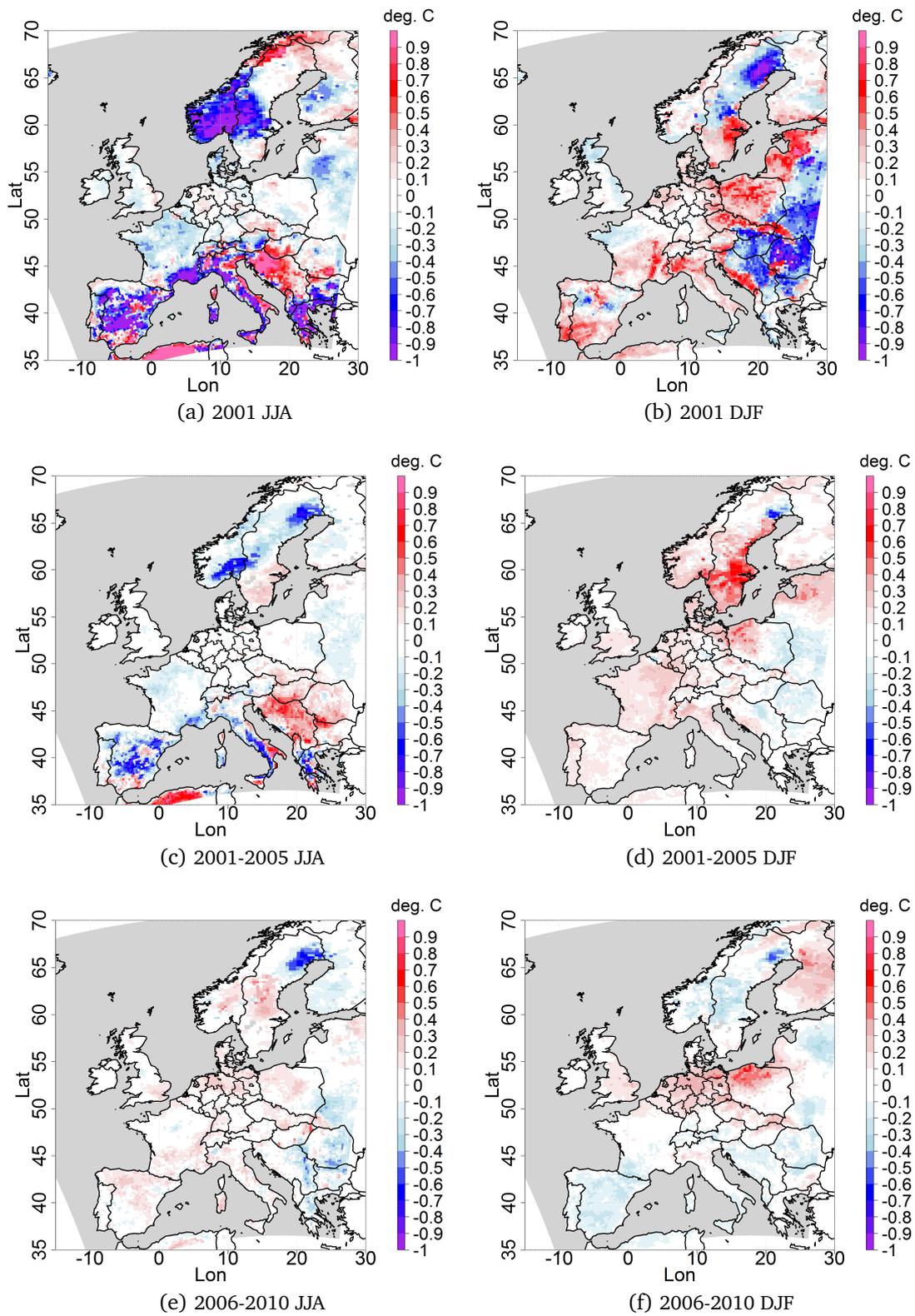
Figure 3.9.: 2001 (upper row), 2002-2005 (middle row) and 2006-2010 (lower row) JJA (left side) and DJF (right side) 2m temperature RMSE differences of CCLM_VEG3D INIT - EOBS to CCLM_VEG3D REF - EOBS for Europe.

The monthly soil water content for the layer 0-1m is displayed for 2001-2010 in Fig. 3.10a and Fig. 3.10b. Besides the monthly means, the difference is displayed with a green line. The simulations are considered as analogue and the soil nearly balanced if the differences are within their standard deviation (dashed lines). The trend of both models is in good agreement, although CCLM_VEG3D INIT is dryer in general. The soil water content of CCLM_VEG3D INIT is about 3.5 % lower for IP and 2.2 % for SC than the one of the CCLM_VEG3D REF in the decadal mean. Nearly no difference between both simulations exists at the end of the decade for IP as the green line shows.
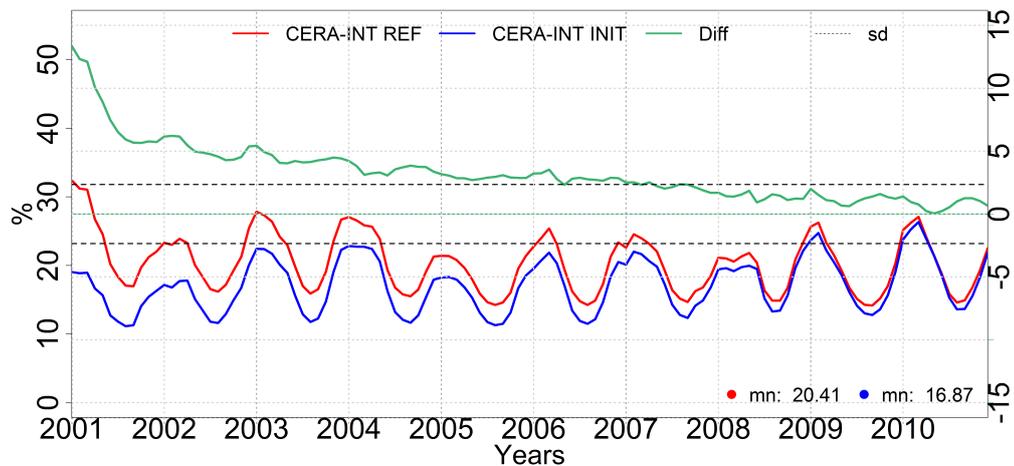
(a) soil water content IP
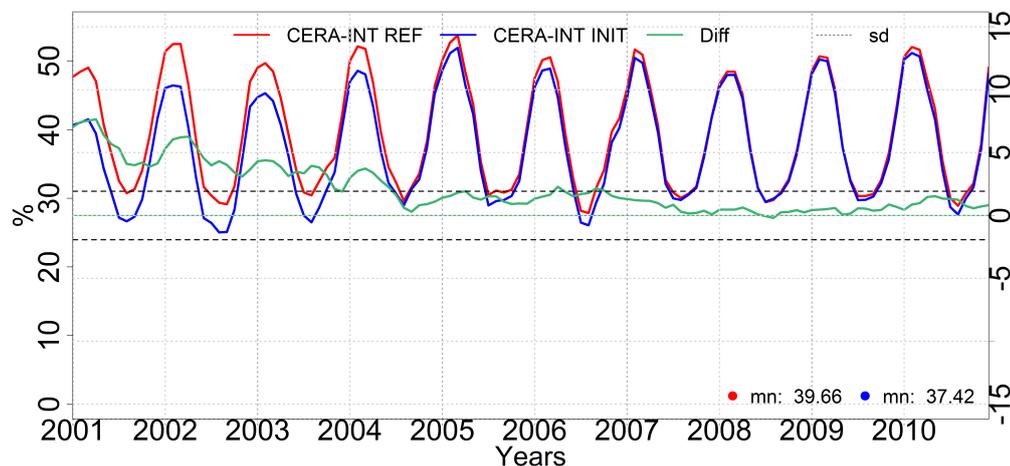
(b) soil water content SC

Figure 3.10.: 2001 - 2010 Monthly mean soil water content for 0-1m for the Iberian Peninsula (upper row) and Scandinavia (lower row) of CCLM_VEG3D INIT and CCLM_VEG3D REF (blue and red line) and their difference (green.

In particular, the drying of the CCLM_VEG3D REF is noticeable in IP during the first years as the model strives towards its quasi-equilibrium state. The soil moisture of the initial fields of

CCLM_VEG3D REF is about > 12 % higher than in the CCLM_VEG3D INIT. The soil dries during the simulation, while the opposite is observed in SC. The soil water content of CCLM_VEG3D INIT is increasing in the first years suggesting a re-wetting of the soil. In addition, the difference reaches its SD much earlier in SC (2005) than in the IP (2008). A dryer soil was found in area plots for most of Europe (not shown) with exceptions in smaller areas in Northern Russia. Also, the lower layers are noticeably dryer for nearly all Europe (except Southern Finland, BI and Eastern Europe). The reason for the notably lower soil water content in the deeper levels seems to be the calculation of the outflow of water in the lowest levels.

The boundary condition determines how much water is lost from the lowest soil layer into the groundwater. These boundary conditions were investigated and published in a sensitivity study: Breil et al. (2018b). A massive outflow of soil water was noticed in the lowest three levels which lead to an increased drying in the SA simulation. An overestimated hydraulic conductivity produced this large outflow. However, the shallow soil levels have a higher porosity with no density increase with depth due to the higher compacted material. Thus, a new depth-dependent function was tested altering the soil hydraulic conductivity. The new function increased the infiltration speed in the shallower level and dampened the outflow rate in the lower levels. This resulted in a changed Bowen ratio with lower latent and higher sensible heat fluxes and a reduced cloud cover due to the lower evapotranspiration.

A larger annual cycle of the 2m temperatures is observed. Fewer clouds enhance the outgoing thermal radiation in the wintertime; therefore, the 2m temperatures decrease. In summer, the incoming shortwave radiation and the temperatures are increased due to a lower cloud cover. As a result, the cold bias of the control simulation was reduced noticeably in summer. The study Breil et al. (2018b) rated, the positive impact of the new hydraulic function in the summertime outweighs the increase of the cold bias in winter.

The soil temperatures are shown for IP and SC in Fig. 3.11a and 3.11b. Increased amplitudes are found for both regions due to the dryer soils similar to the one in Central Europe. The increase of the summer temperatures is larger in the first year (+0.21 °C) in SC than for IP (0.16 °C). However, the differences decay a lot faster. The SD is reached for the SC in the fourth year, whereas this is not the case for IP even in 2006. For IP, the main changes are detected in the summer soil temperatures, while the models converge much faster in winter.
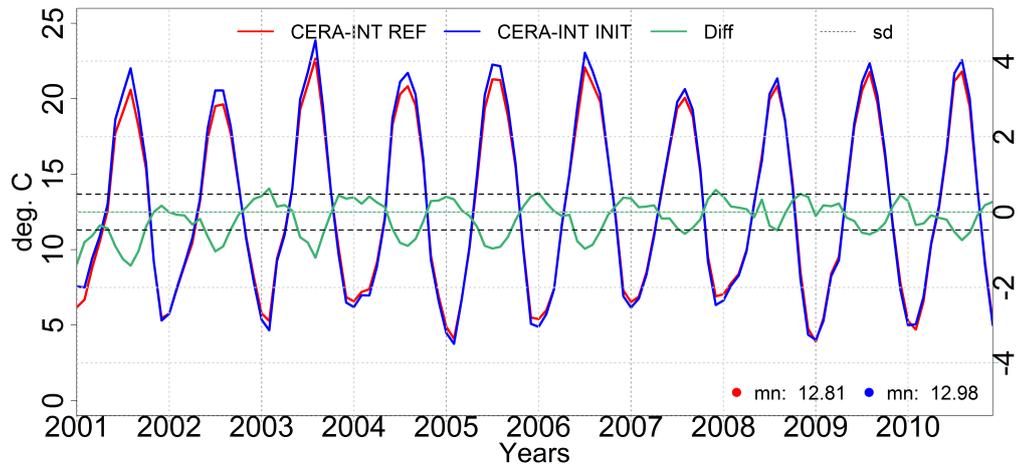
(a) soil temperature IP
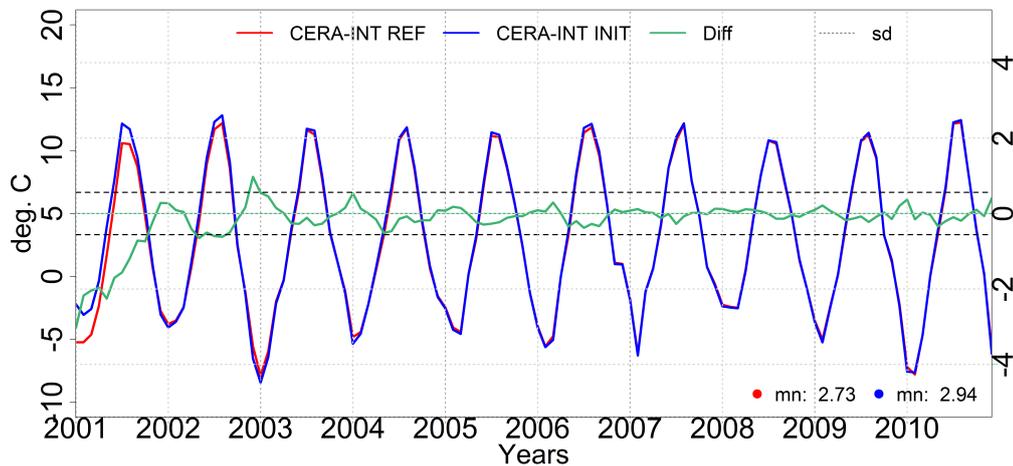
(b) soil temperature SC

Figure 3.11.: 2001 - 2010 Monthly mean soil temperatures for 0-1m for the Iberian Peninsula (IP, upper row) and Scandinavia (SC, lower row) of CCLM_VEG3D INIT and CCLM_VEG3D REF (blue and red line) and their difference (green.

The seasonal differences of the fluxes are displayed in Fig. 3.12a for IP and Fig. 3.12b for SC. In general, the fluxes are lower because of the lower incoming radiation and the larger amount of cloud cover in Scandinavia. Nevertheless, nearly no differences are noticeable, although the soil shows clear deviations for the first two years. In contrast, the IP experiences a considerable impact on the fluxes with a similar decay time due to the soil alteration. The lower water content affects the water cycle in this region in summer significantly and with this the evaporative cooling. In this transition zone, the evapotranspiration is highly dependent on the soil water availability but not as much on the incoming radiation (Seneviratne et al., 2006b, 2010). Furthermore, no corresponding effect on the 2m temperature and the fluxes is noticed for SC, although differences due to the soil initialisation are discovered in the annual cycle of

the soil variables. The fluxes are alternated in this region rather by the strength of the incoming radiation than by changes in the soil moisture as soil water is available during the whole year.
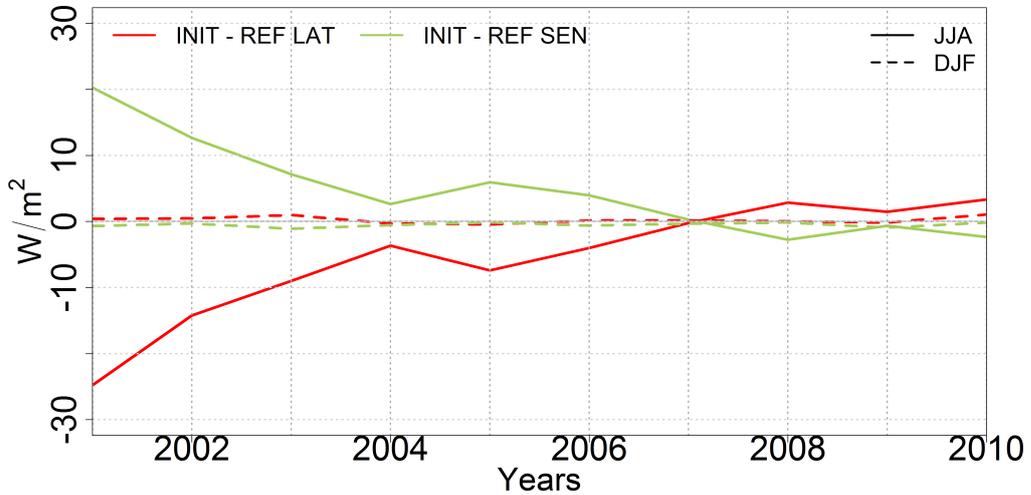
(a) heat fluxes IP
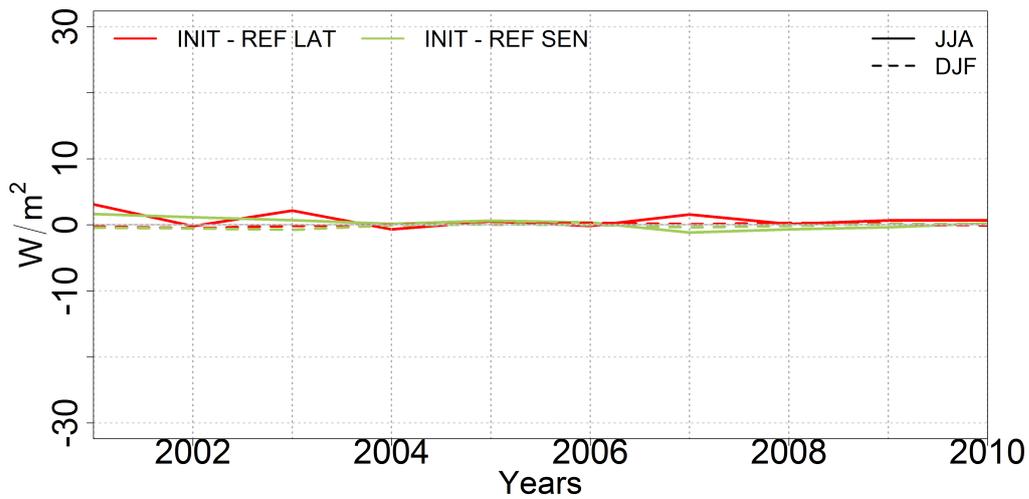
(b) heat fluxes SC

Figure 3.12.: 2001 - 2010 seasonal mean latent and sensible heat flux differences for the Iberian Peninsula (IP, upper row) and Scandinavia (SC, lower row) of CCLM_VEG3D INIT to CCLM_VEG3D REF displayed for the JJA (solid lines) and the DJF (dotted lines) separately.

Short summary

Stand-Alone Simulations:

The VEG3D ERA-Interim simulation had the highest agreement with the station data, although differences for the selected variables were noted. A Kruskal-Wallis test was conducted and emphasised the dependence of the variables on the forcing and the external data. The bias depends clearly on the external data, in particular for the soil water content. The forcing proved crucial for the correlation as it imprints the seasonal cycle on the model. The high correlations with the observations show that VEG3D ERA-Interim can realistically represent climate modes. This is essential for initial data since they nudge the climate model to these states during the initialisation. Therefore, this dataset was selected as forcing for the SA run to produce soil moisture and soil temperature initial fields for the decadal simulations with the coupled system. The quasi-equilibrium state of SA simulations is achieved by a spin up process (here seven years). A clear added value of the higher resolution 0.0625 °to the 0.22 °is not detected as the comparison with the station data showed. It was decided to use the slightly coarser 0.22 °resolution for the simulation for a European Domain due to the lower computational costs.

Coupled Simulations:

The new initialisation fields from the VEG3D SA simulations include dryer soils over most of Europe. This resulted in a reduced heat capacity, which induced a stronger reaction of the soil temperature and an enlarged annual cycle. Additionally, the evapotranspiration is reduced due to the lower plant available water. This not merely increased the thermal uptake of energy into the soil as less energy is used to evaporate soil moisture; also, it reduced transpiration by the plants effectively. As a consequence, less latent and more sensible heat is released to heat the atmosphere and lower cloud amounts are seen. Thus, increased temperatures are detected all over Europe in summer, with lower temperatures in the wintertime. In SC, the impact of the different initial fields was short-lived (< 4 years). Although larger differences occurred in the deeper soils on longer time scales, these had no impact on the atmosphere afterwards. A strengthening of the annual cycle is found in IP due to the lower soil moisture of the new initialisation. The limitation of the soil water availability reduced the evapotranspiration and hence, the cooling in summer. The higher sensible heat fluxes increased the summer temperatures agreeing better with the observations. The effects of the altered initialisation can be found noticeably longer even in the second pentad in the IP and other transition zones due to the dryer soils of these areas.

3.6. Chapter discussion and conclusions

In this chapter, the improvement and potential added value in climate simulations due to a balanced soil initialisation has been investigated for a decadal simulation for different regions in Europe. Several VEG3D SA simulations have been performed and were evaluated to iden-

tify the best possible forcing to produce balanced soil initial fields. Afterwards, two decadal CCLM-VEG3D simulations were produced driven by ERA-Interim, one as reference initialised with the original ERA-Interim soil conditions, the other initialised with the new soil moisture and temperature fields from the Veg3D SA run.

The comparison between the SA simulations and observational data (EFDC stations) revealed that **the SA run driven by ERA-Interim performed best regarding the correlation and the RMSE for the soil temperature and water content**. A Kruska-Wallis test was conducted and emphasised the dependence of the RMSE and the correlation to the forcing and the external data. In particular, the forcing represents a crucial factor since it modulates the seasonal cycle of the LSM. Additionally, the impact of the resolution of the external data was assessed. When at least one of the external parameters fit (land use class or soil type), the simulations agreed better with the observations for the soil temperature and the fluxes. Thus, soil water content and soil temperature fields were generated with **a resolution of 0.22 °for Europe to achieve the desired temporal and spatial coverage at minimum computational costs**. Other studies like Zink et al. (2017) and Stöckli et al. (2008) found similar results for RMSE and correlation of the latent and sensible heat flux, although the forcing, the model and other aspects of the experimental setup differed. The new root distribution implemented in the VEG3D ERA-Interim lead to a drying of the soil in the upper levels due to a shifting of the root density maximum. Plants can more efficiently utilise the infiltrating precipitation as more roots are located in shallower soil layers.

In the second step, two decadal simulations from 2001–2010 have been performed with the coupled system of CCLM and VEG3D. The evaluation disclosed that the changed initialisation had **no systematic impact or added value on the precipitation** so far, although differences were observed. Whereas the **summer temperatures demonstrated an added value due to the changed initialization**, none was found in the winter. The improvement for the summer was **longer lasting in semi-arid transition zones like the IP, whereas the effects in SC were short lived**. In regions with limited evapotranspiration like the IP, the changes towards the new dryer and balanced initial fields resulted in a significant added value for the summer season. These findings are in line with a similar study by Kothe et al. (2016) using the LSM TERRA-ML. The original ERA-Interim initial fields proved to be too wet for VEG3D. Therefore, soil drying was noticed in this region over the next seven years. The CCLM_VEG3D REF simulation **was unable to reproduce the necessary sensible heat fluxes to warm the atmosphere because most incoming energy was utilised for soil and plant evaporation in the wetter soils** . Additionally, the transpiration of the plants lasts longer in the summer, since more soil water is longer available. In the **winter seasons, the heat capacity differences for wet and dry soil become more crucial for the surface temperatures**, and the fluxes are lower. The arider soils have a larger annual cycle and underestimated winter temperatures. The **radiative forcing is more important for the evapotranspiration rates**in Northern Europe since these regions have wet enough soils over the whole year (Seneviratne et al., 2010). The annual cycle is increased in

Scandinavia also due to the higher heat capacity of the dryer soils. An improvement was noted for the summer temperatures as well as an increased bias in the wintertime. Furthermore, the simulations needed fewer years to become balanced and to converge than in the IP. Thus, an unbalanced soil initialisation causes a far longer-lasting effect for (semi-) arid regions than for humid regimes. The infiltration rate into the soil depends on the hydraulic conductivity and diffusion. As known from the Richards-equation, water can travel much faster through wet soils as through dry ones (Blume et al., 2010). Therefore, decay times are higher for locations with dryer soils, which is in line with Khodayar et al. (2015). It can be assumed, that the initial conditions are not as important for climate simulations with a spin up of three years running for time periods of 30 years and longer. However, they can have a massive effect on seasonal to decadal predictions. On global scale, large areas are considered to the transition zones sensitive to soil conditions (Dirmeyer, 2001; Seneviratne et al., 2006b; Miralles et al., 2012). In particular, the impact on extremes like heat waves is not negligible (Jaeger and Seneviratne, 2010). For example, mega-heatwaves like 2003 are highly connected to soil desiccation (Miralles et al., 2014) and with this to initial conditions as well as the LSM used for the simulation. Even if an accurate atmospheric forcing is applied, the quality of the LSM can be crucial for the development of heatwave conditions (Weisheimer et al., 2011). Furthermore, transition zones will likely expand or shift because of the climate change (Seneviratne et al., 2006b, 2010). Regions like Central Europe, currently classified as temperate humid, will likely become dryer and warmer in the next years (Schär et al., 2004). In this case, a non-fitting soil initialisation might have a much longer impact on the quality of predictions or even projections due to the increasing decay time of the dryer becoming soils.

4. Decadal predictability of Heat Waves and dependence on large scale circulation anomalies within the MiKlip predictions

4.1. Introduction

Decadal predictions fill the gap between seasonal predictions and long-term climate projections and aim to provide climate information for users with a planning range between 1 year and 5 to 10 years. On this timescale, these predictions are both initial and boundary value problem (Meehl et al., 2009). The predictability on interannual to decadal time scales arises from realistic initialisations of the slow climate components like the ocean, sea ice and soil, their long-term internal variability and the overall climate trend. A goal of the research program from the Bundesministerium für Bildung und Forschung (BMBF) / Forschung für Nachhaltige Entwicklung (fona) MiKlip is the development of a forecast system to produce skilful decadal predictions (Marotzke et al., 2016) and to explore its potential. In addition, the project includes sensitivity studies testing various approaches like the initialisation procedure or model grid resolution. These tests help to improve the robustness of the ensemble results and to find the best possible configuration for the operational prediction system. Up to now, several generations of global MiKlip ensembles have been testing new methods to improve the prediction system. This encompasses initialisation strategies (Pohlmann et al., 2013) as full field initialisation of the ocean instead of anomaly initialisation (Kröger et al., 2018; Paeth et al., 2018) or increased horizontal and vertical resolution of the global model for the ocean and atmosphere (Müller et al., 2018). Decadal predictions, so-called hindcasts, are performed to assess the quality of the prediction system from 1961 onwards. The decadal predictions are initialised every year to align models with actual conditions of the ocean, the soil and the atmosphere. This approach captures the internal climate variability and minimises the error due to deviations from the observed climate variability (see Marotzke et al., 2016).

The users of climate data often require high spatial resolutions for specific applications. Therefore, global simulations are dynamically downscaled with the regional climate model CCLM for Europe and Africa within MiKlip. Although uncertainty due to the downscaling is added to the data (Raäisaänen, 2007), measures have been taken to minimise it at least. In general, the quality of regional climate models depends on several factors, both external and internal ones. Internal ones are generated by the model itself, e.g. the used cloud microphysics scheme or the parametrisation of the land-surface-interactions. External ones include the driving data or the environmental parameter (e.g. soil types or vegetation factors depending on the land use

like the leaf area index or the stomata resistance). The external forcing, driving the RCM at its boundaries, can either be the output from a global climate model or a data set, e.g. reanalysis data. In addition, the quality of the model results depends on the initialisation of the ocean and the soil as lower boundaries, especially in the first years. To provide the best possible initialisation fields of soil moisture and temperature for the regional decadal predictions, so-called evaluation (EVA) runs are often conducted. In MiKlip, CCLM EVA b1 and CCLM EVA preop have been generated driven by reanalysis data, one of the most realistic forcings available. With the soil moisture and temperature fields of the EVA runs, a presumed initial shock is minimised in the soil and a possible model drift reduced during the first years.

In several studies, predictive skill has been found for the mean temperature (Meehl et al., 2009; Scaife et al., 2014; Boer et al., 2013) as well as for extreme temperatures of 5- and 10-year averages (Hanlon et al., 2013). It can be assumed that temperature based variables inherit parts of this forecast skill. The investigation of their potential predictability and forecast quality is vital due to the impact of HWs on the population and the environment. The predictability of European HWs and the dependence to large-scale climate anomalies (LCA) is examined in this study with the MiKlip data sets. In the first part of this chapter, the data sets are explained in further detail. The daily maximum temperatures, on which the HW calculation is based, are compared to the observations in the second part. Afterwards, the HWs produced by reanalysis driven simulations and decadal predictions will be analysed.

4.2. Data and Methods

For the investigation of the predictability of European heat waves, decadal hindcasts of the MiKlip ensemble are employed along with the two simulations driven by reanalysis data (EVA runs). The following setup describes the second generation of the MiKlip ensembles so-called baseline1 (b1). The b1 generation contains simulations of global (GCM) and regional climate models with yearly starting dates (initialised at the 1st January) and in addition a set of regional uninitialised simulations. The global simulations are performed with the earth system model MPI-ESM of the Max-Planck-Institute in various resolutions for the different ensemble generations. The MPI-ESM contains the atmospheric component ECHAM6 (Stevens et al., 2013) coupled to the ocean model MPI-OM (Jungclaus et al., 2013) and the JSBACH as land surface scheme (Reick et al., 2013). It is run in low resolution mode [atmosphere: T63/L47 ~1.9°, ocean: 1.5°L40, Müller et al. (2012), MPI-ESM-LR b1 in the following] and is initialized with temperature and salinity information from the ORA-S4 reanalysis data (Balmaseda et al., 2013). The ensemble of ten realisations is build using 1day lag, which generates new initial conditions for each start year (Müller et al., 2012).

The RCM CCLM is described in section 3.2.2; therefore only the configuration of the MiKlip b1 simulations is listed below. Additional to the yearly initialised simulations, GCM and RCM ensembles of uninitialized simulations are conducted. These so-called historical runs are produced with the same setup as for the RCM b1 but are continuously running from 1961 until 2010. The

comparison to the regional historical simulations is included to evaluate the added value due to the yearly initialisation. However, the uninitialized ensemble contains only seven realisations. Therefore, this comparison uses only the mutual seven ensemble members of each data set. It can be expected, that the results will show a possibly slightly lower skill in contrast to the full ensemble with ten members (Sienz et al., 2016; Reyers et al., 2019).

Besides the decadal predictions and the historical, the EVA runs are a valuable asset to investigate processes and to estimate the best possible results achieved with models. In addition, these runs can be held as substitute observations for variables measured scarcely or missing entirely. Along with the evolving MiKlip ensemble generations, two EVA simulations CCLM EVA b1 and CCLM EVA preop were produced (simulation domain Fig. 4.1) and will be analysed in the following.
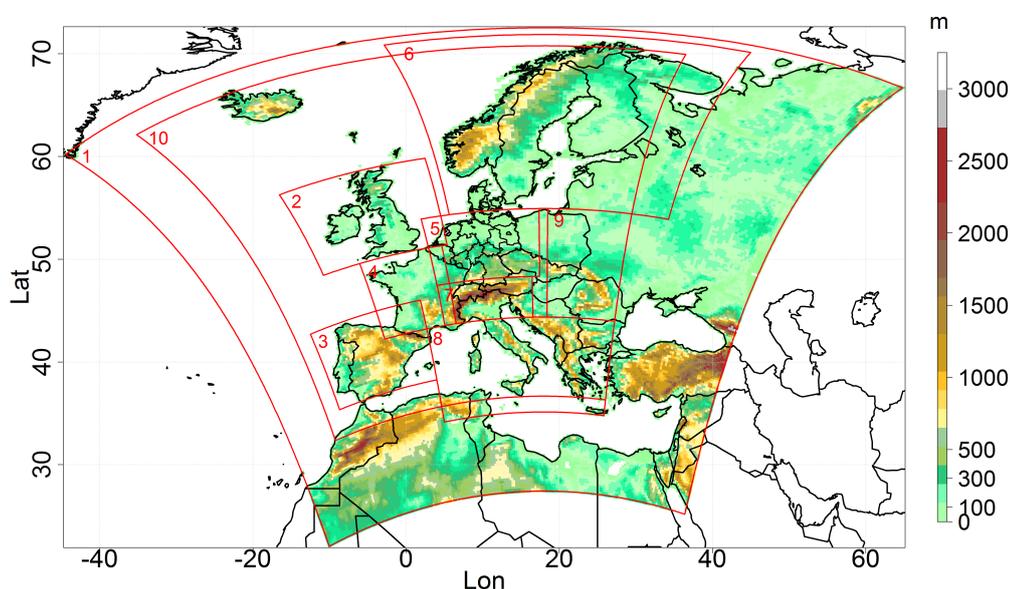
Figure 4.1.: Relief of Europe with PRUDENCE regions drawn into (2-9), the European evaluation area (10, EU_c) and the European simulation area without the relaxation boundary zone (1, EU). PRUDENCE regions after Christensen and Christensen (2007): (2) British Isles BI, (3) Iberian Peninsula IP, (4) France FR, (5) Middle Europe ME, (6) Scandinavia SC, (7) Alps AL, (8) Mediterranean MD, (9) Eastern Europe EA

The simulations have a nearly identical trend and differ mainly by the model version of the CCLM (CCLM 4.8 compared to CCLM 5.0), the spatial resolution ($0.44°$ to $0.22°$) and the preprocessor version INT2LM (preparing the global output for the CCLM) in the setup. The EVA simulations are produced as transient runs using the ERA40 data as forcing from 1960-1979 (Uppala et al., 2005) and the ERA-Interim from 1979 onward (Dee et al., 2011). The two simulations driven by ERA40 and ERA-Interim will be analysed for 1961-2010. This allows an estimation of the variability of the HW indices and the model performance with the best available forcing as said before. In a second step, the decadal predictions of the MiKlip generation baseline1 are analysed and the added value of the regionalisation is assessed. Daily maximum

temperatures of the Eobs V.14 data are used (Haylock et al., 2008) as observations with a spatial resolution of 0.22° (25 km). All model data are interpolated to this resolution and height corrected afterwards. The different PRUDENCE regions as well as the European evaluation domain and the simulation area are displayed in Fig. 4.1. If not mentioned otherwise evaluations for Europe are performed for region 10, while All Europe refers to the simulation area without the relaxation zone (1).

Beside the analysis of the absolute values and the differences between model and observations, metrics are applied to assess the skill. Correlations are calculated as well as the root mean square error (RMSE) and the mean square error skill score (MSESS). Based on the mean square error (MSE, Goddard et al. (2013)), the MSESS compares the results of hindcasts (H) and a reference (R, either the climatology or other hindcasts) to each other and the observations (definition after Wilks (2011) and Rust et al. (2014)).

$$MSE(H,R,O) = \frac{1}{n} \sum_{j=1}^{n} (H_j - O_j)^2 \qquad [4.1]$$

$$MSESS(H,R,O) = 1 - \frac{MSE_{H,O}}{MSE_{R,O}} \qquad [4.2]$$

The mean square error skill score of the hindcasts (H) and reference model (R) can be decomposed into correlation and conditional bias in the following form (Murphy, 1988; Rust et al., 2014):

$$MSESS(H,R,O) = \frac{r_{H,O}^2 - \left[r_{H,O} - \frac{s_H}{s_O}\right]^2 - r_{R,O}^2 + \left[r_{R,O} - \frac{s_R}{s_O}\right]^2}{1 - r_{R,O}^2 + \left[r_{R,O} - \frac{s_R}{s_O}\right]^2} \qquad [4.3]$$

The conditional bias contains information about the ratio of the variances of hindcasts and observations by comparing the standard deviations of both to the correlations (Rust et al., 2014). A further explanation of the scores can be found in Kadow et al. (2016).

$$Conditional\,bias : \left|r_{H,O} - \frac{s_H}{s_O}\right| - \left|r_{R,O} - \frac{s_R}{s_O}\right| \qquad [4.4]$$

In addition, the relative added value (RAV) is calculated comparing the correlations of the model simulations directly to the observations.

$$RAV(H,R,O) = \frac{r_{H,O} - r_{R,O}}{1 - r_{R,O}} \qquad [4.5]$$

In this study, the GCM or historicals are the reference data in the MSESS calculation. In the graphics for the correlation differences as well as the MSESS, positive values are marked by red colours and indicate an added value in CCLM b1 due to either the regionalisation or the initialisation. Blueish colours show that more reliable results are produced by global simulations or historicals. The application of the decomposition into the conditional bias and the correlation as an output of the evaluation system MurCSS (Rust et al., 2014) can be seen in the subsection

of the user relevant variables 4.4.2.4. Unlike the MSESS and the correlation, values around 0 show the best agreement with the observations for the conditional bias.

4.3. Simulating daily maximum temperatures for Europe

4.3.1. Evaluation runs

The daily maximum temperature of the summer half-year is evaluated briefly in this subsection as the HWs are based on it (see chapter 2.2). It can be expected, that biases of the daily temperatures will appear in the mean and maximum HW temperatures as well. In Fig. 4.2, the averaged summer half-year (SHY) temperatures are displayed. The Eobs data show a mean SHY temperature of 19.24 °C for the European evaluation domain. The EVA runs capture the trend very well with high mean correlations of 0.78 CCLM EVA b1 and 0.84 CCLM EVA preop. Furthermore, a low mean bias is seen in the last decade. CCLM EVA b1 has the highest mean temperature with 19.65 °C and deviates visibly from the observations except for the decade 1971-1980. In particular, temperatures of hot years are overestimated in the CCLM EVA b1, which is not recognisable in the CCLM EVA preop. This simulation fits quite well to the observations and shows only smaller differences after the decade 1961. This decade seems somewhat troublesome in the model results, but the forcing (ERA40) and the quality of the Eobs observations might as well be questionable for this period.
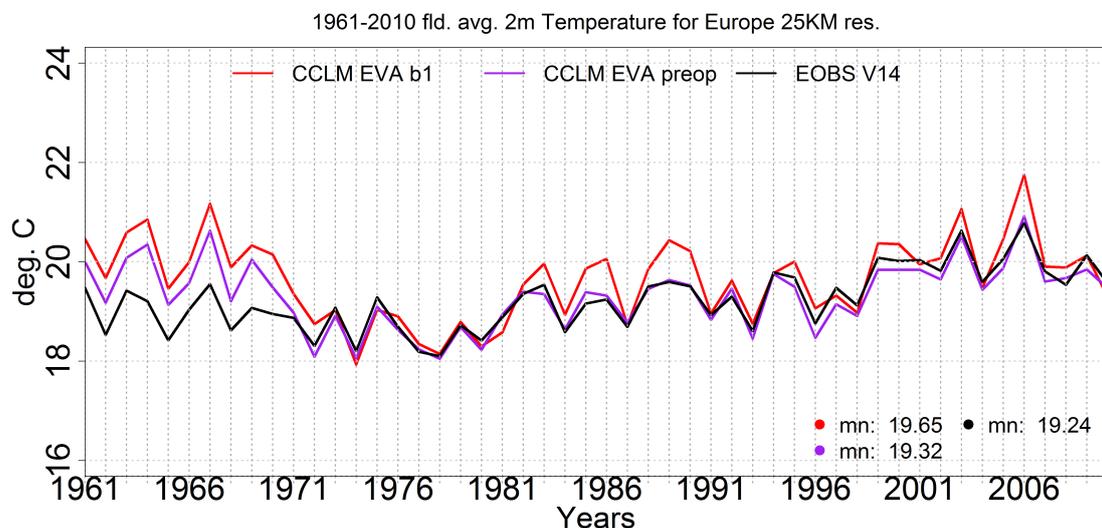
Figure 4.2.: Annual mean summer half-year temperatures for 1961 – 2010 EVA simulations and Eobs V14 data for Europe (10).

The averaged temperatures, the RMSE and the correlation are summarised in table A.1 in the appendix for the different PRUDENCE regions. Higher correlations for the CCLM EVA preop have been found (r> 0.8) in nearly all regions. CCLM EVA b1 has slightly lower correlations in Northwestern regions, while higher correlation coefficients are seen in the more continental and

southern ones. Additionally, a slightly higher RMSE is found for the CCLM EVA b1 simulation in most of Europe. The distribution of daily maximum temperatures averaged for summer half-years 1961-2010 for the Eobs data and the differences of the EVA runs to Eobs are displayed in the following graphics (Fig. 4.3a until Fig. 4.3e).

(a) Eobs V14 SHY   (b) CCLM EVA b1 –Eobs   (c) CCLM EVA preop –Eobs

(d) COR CCLM EVA b1 to Eobs   (e) COR CCLM EVA preop to Eobs   (f) CCLM EVA b1 –CCLM EVA preop
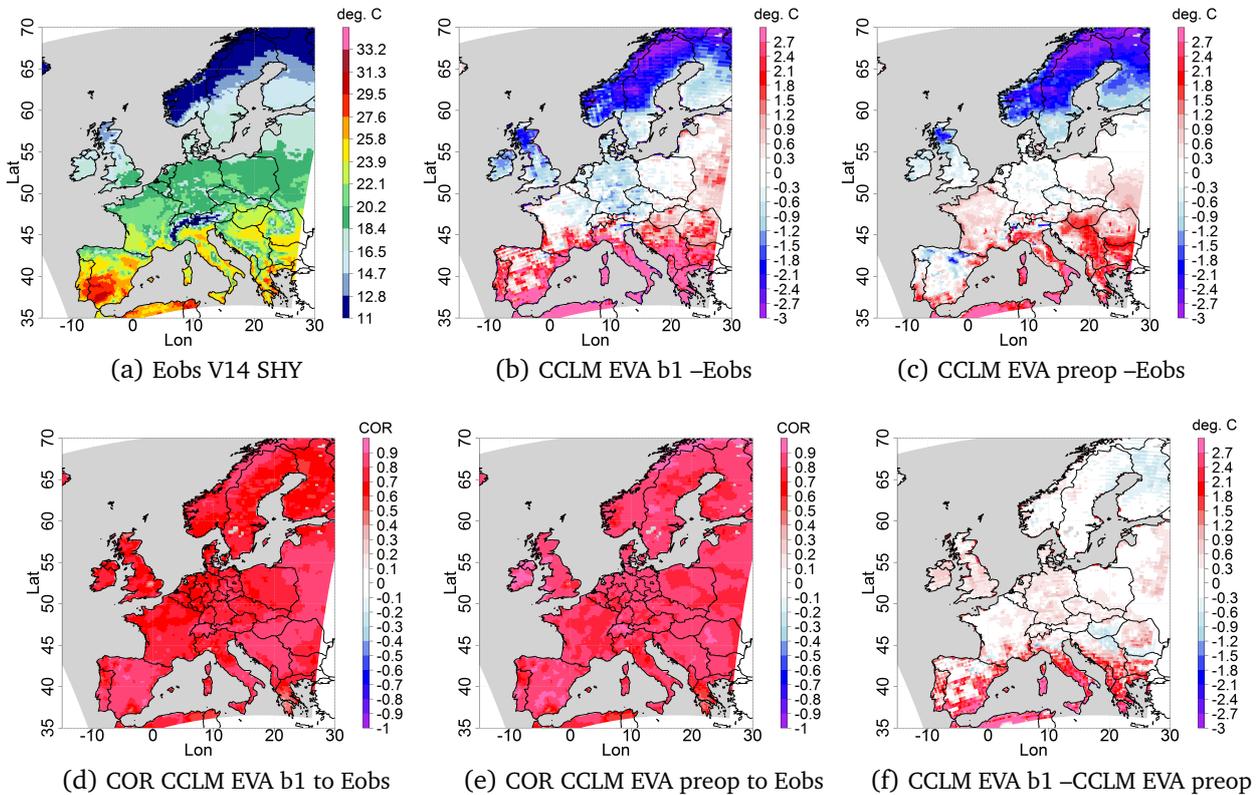
Figure 4.3.: Annual mean summer half-year temperatures for 1961 – 2010 Eobs V14 (upper left) and the differences of the EVA simulations (upper middle and right). Lower row: Correlations between EVA and Eobs (lower left and middle) and the RMSE difference of CCLM EVA b1 to Eobs and CCLM EVA preop to Eobs.

The temperatures range from below 11 °C in the North to around 30 °C in the South. Both simulations show a similar pattern, but the difference plots reveal too high temperatures in the South and a cold bias in Scandinavia. The CCLM EVA preop simulation is slightly colder in the mean than the CCLM EVA b1 as the cold bias in the North is larger (Scandinavia +0.29 K) and the overestimations in the South are smaller (Mediterranean -1.47 K). The correlation pattern are very similar with slightly higher values for the CCLM EVA preop by +0.06 in the mean (Fig. 4.3d and Fig. 4.3e). The RMSE difference shows clearly the overestimation of the CCLM EVA b1 in Southern Europe ($> +2.7$ °C). Thus, it can be assumed, that the new CCLM model version and the increase in resolution produce slightly lower mean temperatures (-0.33 °C) in the summer half-year.

4.3.2. Decadal predictions for Europe

The daily maximum temperatures of decadal hindcasts of the baseline1 generations are analysed as in the previous subsection. The correlation, the conditional bias and the MSESS (based on equation 4.2) are calculated for daily maximum temperature anomalies in the summer halfyear with Eobs V.14 as observations using the MurCSS plugin of the MiKlip Central Evaluation System (CES). The MurCSS results for the correlation and the MSESS are shown for different lead-times (LT) in the graphics(Fig. 4.4a to Fig. 4.4l. The figures include the correlation of the regional hindcasts (left), the MSESS (middle column) of the regional hindcasts compared to their climatology to the observations and MSESS of the regional hindcasts compared to global ones to Eobs (right column). The middle column represents the added value due to the initialisation, whereas the right column displays the added value due to the regionalisation.

The left column shows significant high correlations of r > 0.6 in the mean for nearly whole Europe for longer lead-times like 2-5 and 5-8 (Fig. 4.4g and Fig. 4.4j). These are with r < 0.2 noticeably lower in the European average in the first year (Fig. 4.4a). A nearly identical pattern is seen in the MSESS (middle column) with largely positive values in the longer LTs. An added value compared to the climatology is identified across Europe except for the first year in Eastern Europe (MSESS < -1, Fig. 4.4b). The large area with weak correlations (Fig. 4.4a or negative MSESS values in Eastern Europe in Fig. 4.4c shrinks for longer lead-times (Fig. 4.4f) or the exclusion of lead year 1 (Fig. 4.4i).

No noticeable differences between CCLM b1 and MPI-ESM-LR b1 hindcasts are found in the MSESS distribution for LT2-5 and LT5-8 (Fig. 4.4i and Fig. 4.4l). One reason for the similarity are the AMO phases dominating the trend of the temperature anomalies. This long-term variability, included in the SST data driving the GCM as lower ocean boundary, is inherited by the RCM. Therefore, quite high skill values existed beforehand and no large improvement can be expected by the downscaling for the correlation.

In particular, Eastern Europe is problematic during lead year 1 (Uhlig, 2016; Reyers et al., 2019). Year one is often ignored or evaluated separately, since it includes the strongest drift of the model system reacting to external factors and initialisation fields of atmosphere, soil and ocean (Meehl et al., 2009). The large differences between the start year and later lead-times induce the assumption that the soil initialisation may be erroneous or non-fitting for the regional hindcasts in this region. If the initial fields from the EVA simulations deviate strongly from the balanced state of the soils produced by the land-surface-scheme TERRA_ML of the regional model, then the soil would drift towards its balanced state mostly during the first year. The changing soil conditions can lead to varying latent and sensible heat fluxes in the regional model and with this to alternated temperatures (as seen in chapter 3). Fig. 4.4i and Fig. 4.4l seem nearly identical on the other hand because of the much weaker/disappearing drift. Uhlig (2016) and Reyers et al. (2019) showed similar results for the annual mean temperatures with high correlation values for LT2-5 and LT3-6. Furthermore, the averaging over longer lead-times is recommended by (Goddard et al., 2013), where intervals of four years like 2-5 or 6-9 were proposed.
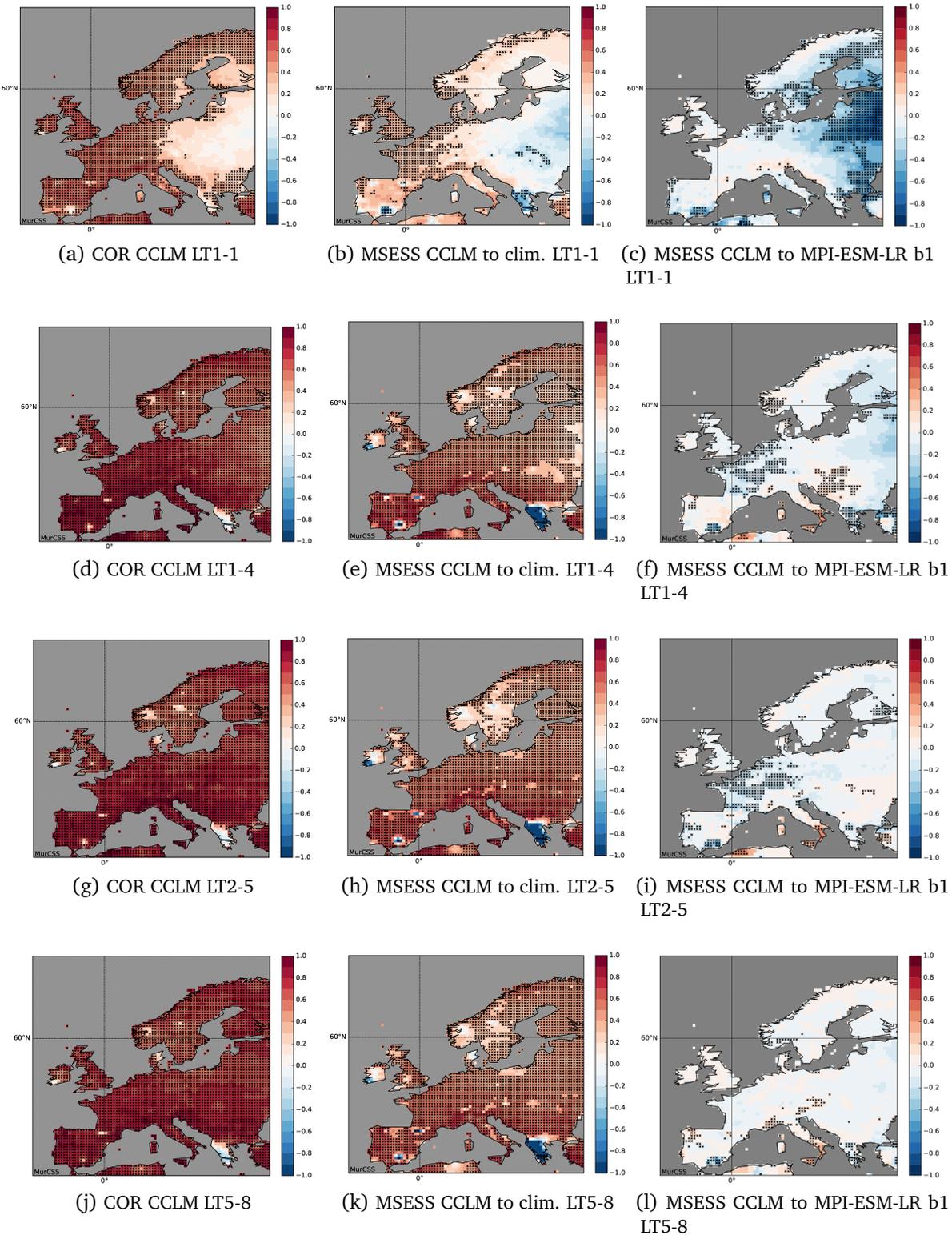
Figure 4.4.: MurCSS outputs for SHY daily maximum temperatures from left to right: Correlation of regional hindcasts to Eobs V.14, MSESS (added value) of the regional hindcasts to the climatology to Eobs and the MSESS of the regional and global hindcasts to Eobs. The x marks grid boxes with significant results. From top to bottom: Lead-time depended analysis for LT1, LT1-4, LT2-5 and LT5-8 with shifted time intervals to achieve an overlapping evaluation period for all lead-times. Significances are shown as black x for the plots in the left/middle row.

4.4. Comparing heat waves derived from simulations with observations

A heat wave occurs when on at least five consecutive days daily maximum temperatures exceed the 90ths percentile of the reference period as defined in 2.2. As in the prior section, the heat waves are evaluated separately for the EVA runs and the decadal hindcasts. The 90th reference percentile is examined; first, this allows a bias estimation for the HW temperatures. The percentiles are calculated for the climatological reference period 1981-2010 for the MiKlip data unlike the simulations for Southern Germany (chapter 5). The MiKlip project includes forecasts for the next decades, currently 2017-2027. Therefore, the reference period is shifted by 10 years to avoid a too large gap between the reference, the latest simulations and the recent forecasts.

In the following the results of the analysis for the EVA runs can be found including the 90th percentile, the HW indices and the assessment of the connection between HWs and large scale circulation anomalies. All evaluations are performed for the summer half year as before.

4.4.1. Evaluation runs

4.4.1.1. Extreme percentiles for the reference period 1981-2010

The 90ths, the 95ths and the 98ths percentile are compared for the two CCLM EVA simulations and the EOBS V14 data for the climatological reference time 1981-2010. The 90th percentile of daily maximum temperatures of the Eobs V14.0 exceeds $> 35\,^\circ$C in Southern Europe and decreases to 12.5-15 $^\circ$C in Scandinavia. In addition, a lower percentile is seen in mountain ranges like the Alps and the Carpathians. The spatial pattern of the 90th percentile is maintained in the 95th and 98th percentile, while the percentile values increase by $> +1.7\,^\circ$C for the 95th and $> 3.5\,^\circ$C for the 98th percentile in the mean. The percentiles are overestimated in both EVA runs in Southern and Eastern Europe ($> 3\,^\circ$C) intensifying for the higher percentile level (comparing Fig. 4.5b to 4.5e and 4.5h). In Scandinavia, however, an obvious cold bias exists ($< -2.5\,^\circ$C), while the deviations from the observations are quite small in Central Europe. This indicates a broader temperature distribution in the model simulations compared to the Eobs for high-level temperatures. The temperatures in the observations increase not as much for the higher percentiles resulting in a smaller distribution.

The EVA runs with the CCLM versions 5.0. and 4.8 show a noticeable difference. CCLM EVA b1 has much larger positive deviations from Eobs in Southern and Eastern Europe, while the differences are not as high in CCLM EVA preop. On the other hand, a more prominent cold bias is found in CCLM EVA preop. As seen in the comparison of the daily maximum temperature, CCLM EVA preop is colder in the mean than CCLM EVA b1 fitting better to the observations for most regions. Thus, the result for the percentiles is in good agreement with the findings of the daily max temperatures. Vautard et al. (2013) found a similar bias distribution in CORDEX CCLM simulations. Additionally, he pointed out, that the increase of the resolution leads to a decrease of the warm bias, which is confirmed by the results of the EVA simulation so far.
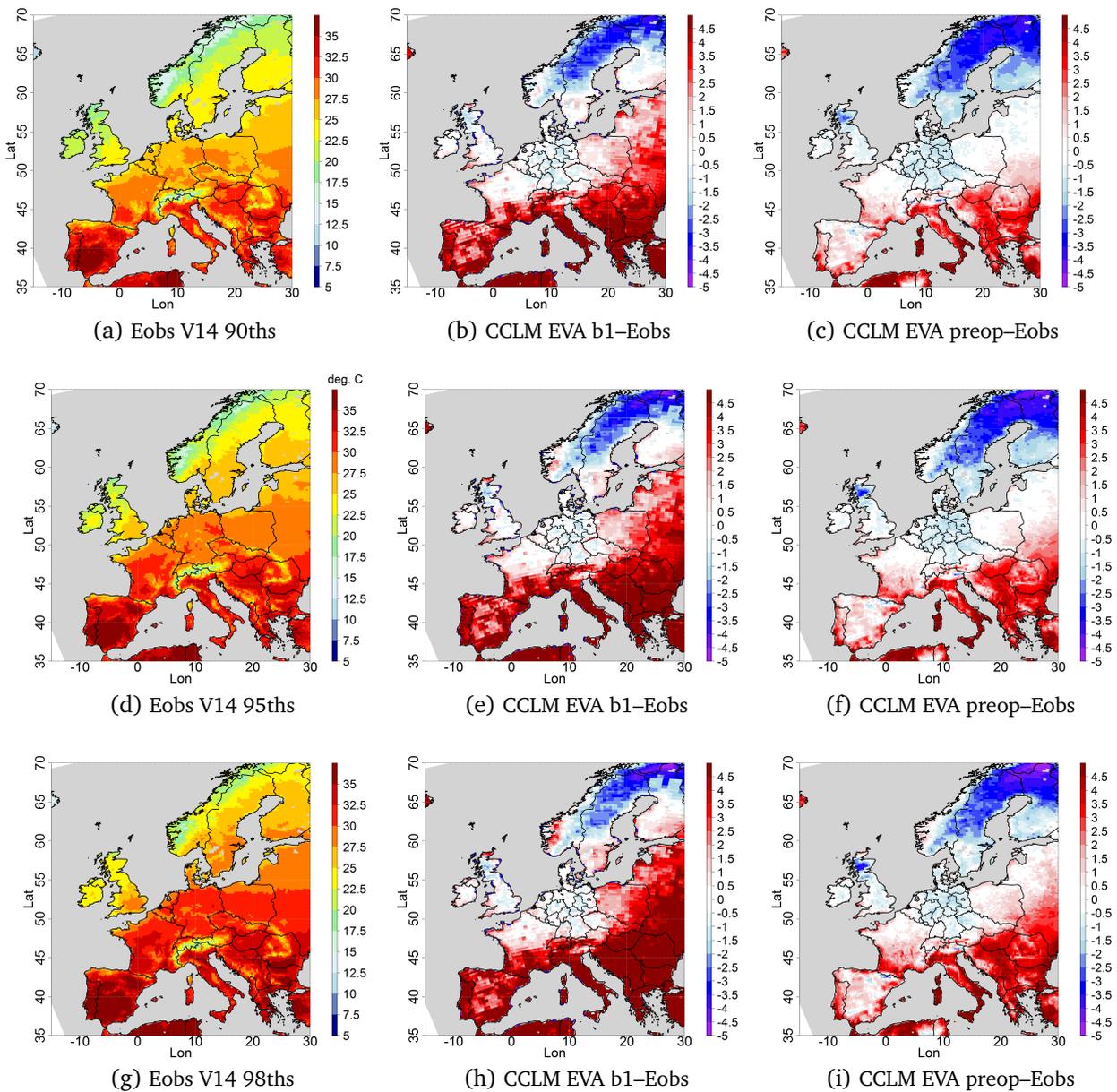
Figure 4.5.: 90ths (first), 95ths (middle) and 98ths (last row) percentile for the daily maximum temperature of the summer half years 1981-2010 for Eobs V14 and the differences of the CCLM Eval simulations to Eobs (middle and left column).

4.4.1.2. HW characteristics for Europe from 1961-2010

Different attempts have been conducted to characterise HWs in Europe in detail over the last years. In this study, an event set similar to Russo et al. (2015) is constructed to identify the years with the highest heat impact in form of intensity, spatial extent and duration. The spatial extent is calculated as the percentage of land GP affected by the HW and displayed for decadal slices from 1961-2010 in Fig. 4.6a until 4.6e. The Eobs data show the smallest numbers of affected land GP for the decade1980 in the European average. A visible increase in affected GPs is found

from the decade 1990 onward. The maximum GP sums are seen for the years 1994, 2003, 2006 and 2010. A HW affected over 50 % of the European area during these years, which agrees with the findings of Barriopedro et al. (2011) and exceeds the spatial range of their "mega-heatwaves" criterion.

(a) HW decade 1961-1970

(b) HW decade 1971-1980

(c) HW decade 1981-1990

(d) HW decade 1991-2000

(e) HW decade 2001-2010
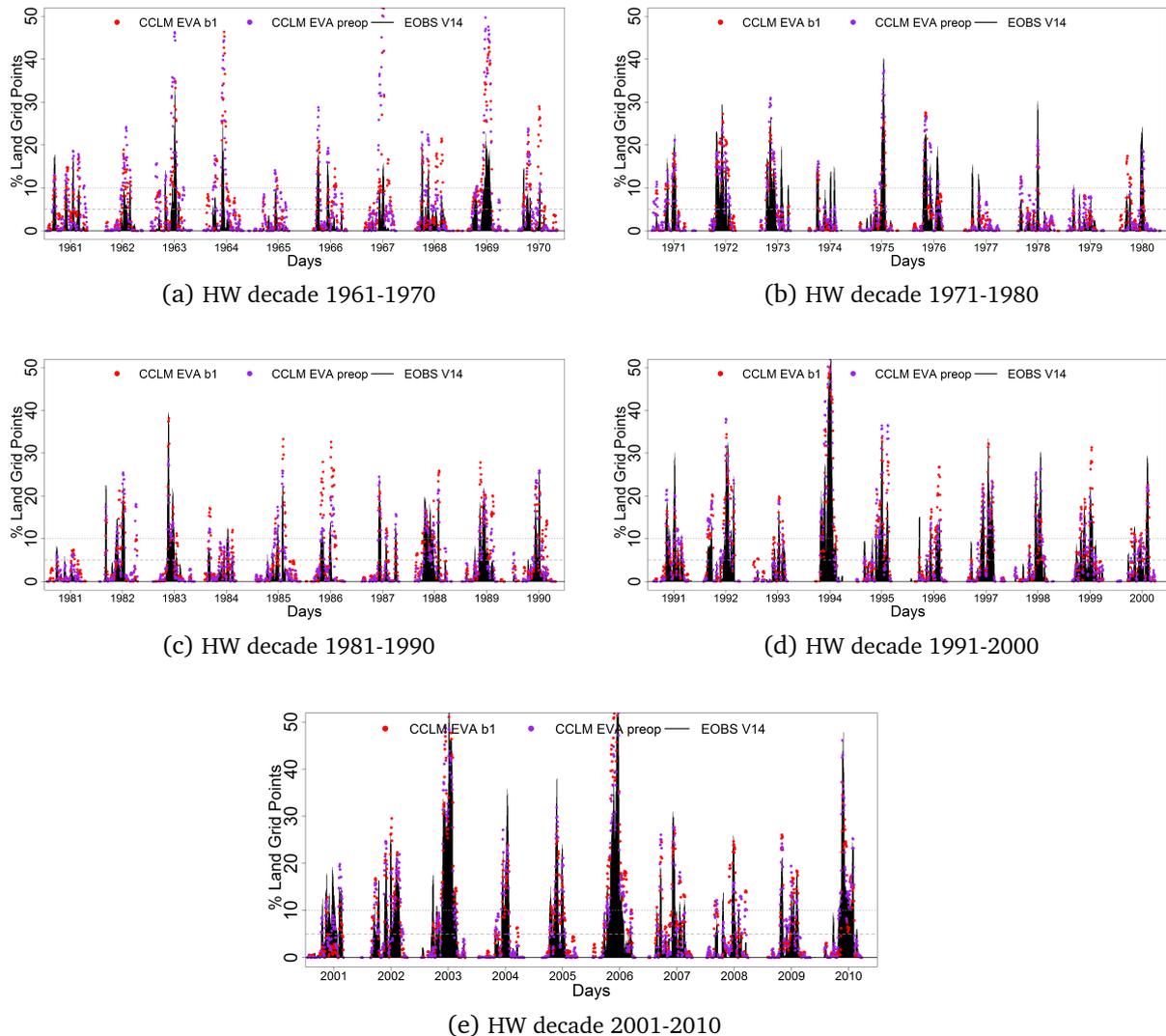
Figure 4.6.: Daily HW land GP percent for the decades from 1961-2010 for Eobs and the CCLM EVA simulations (middle and right column) for Europe (10).

Quite noticeable differences occur between the CCLM simulations and the results of the HW calculation from Eobs in the decade 1961-1970 (Fig. 4.6a). CCLM EVA b1 produces a considerably warm bias in this decade, which was seen in the daily maximum temperatures in the previous section. Several assumptions were attempted explaining this bias, e.g. lower quality of forcing or observational data. The EVA simulations show a quite good agreement with the observations for the other decades except for single years like 1985, 1986 or 1996. HWs cover on average 6.0% of the land GP of Europe in the Eobs observations for 1961-2010. However, this seems to

be a small number regarding the mega-heatwaves of the last years. The EVA runs show similar numbers with CCLM EVA b1 6.5% and CCLM EVA preop 6.3%. It was noticed, that most of the HWs affect less than 5 % of land points for the European evaluation area. Thus, the averages were calculated again with a low pass filter avoiding events with a too small impact and strong stochastic/noise character (HW affecting less than 5 % of the GP are excluded from the calculation of the mean). The mean GP area sum rises to 13.1 % for Eobs with this filtering. The EVA simulations increase in the same magnitude as the observations (CCLM EVA b1 13.5%, CCLM EVA preop 13.7%).
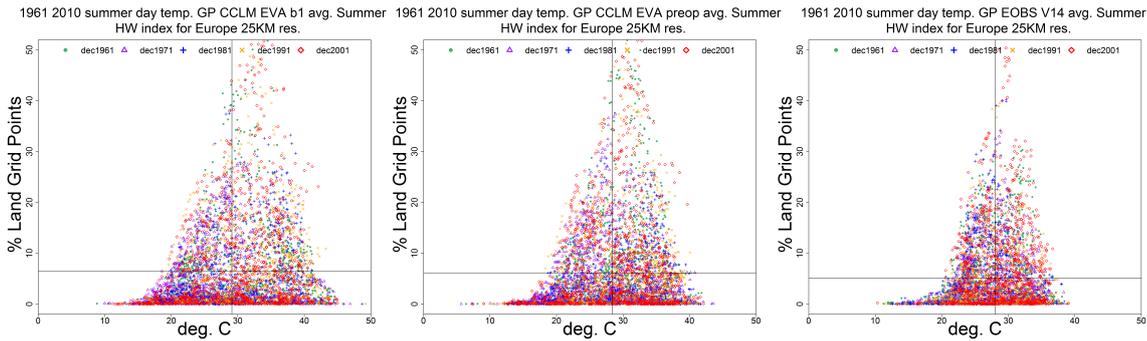
Figure 4.7.: Daily GP percentage in dependence to the HW temperatures [°C] for the decades from 1961-2010 for Eobs (right) and the CCLM EVA simulations (left and middle) for Europe (10). The different decades are displayed by coloured dots: 1961-1970 green, 1971-1980 purple, 1981-1990 blue, 1991-2000 yellow and 2001-2010 orange.

The scatter plots show the temperature distributions of the HW days in relation to the area sums for the three data sets (Fig. 4.7). The Eobs data show an averaged HW days temperature of 28.1 °C and a narrower distribution than the models. In addition, the distribution is not symmetric as can be seen by the much larger tail. European HWs associated with the broadest spatial coverage (> 40 % GPs) and temperatures ~ 30 °C are seen from decade 1981 onward. In addition, both EVA runs have much broader temperature distributions ranging at the base from 10 up to 45 °C compared to the Eobs data. Furthermore, a larger dispersion of HWs days exists for events with high magnitudes including temperatures around 40 °C and area sums from 25-30% (stronger in CCLM EVA b1 than in CCLM EVA preop). This indicates a larger variance of HW temperatures in the models than found in the observations.

A slight tilt towards higher temperatures can be noticed in the EVA distributions (left and middle panel from Fig. 4.7). This tilt is a result of too much HW events in the model simulations covering 40 to 50 % of the GP, especially in the CCLM EVA preop. The overestimation of the decade 1961 is visible for the largest HWs as well. Compared to Eobs, the CCLM EVA b1 mean HW days temperature is higher by +1.0 °C, while for CCLM EVA preop the differences are smaller with +0.2 °. Also, the CCLM EVA preop distribution of HW day temperatures is slightly narrower than the CCLM EVA b1. The graphics show that the model produces a general overestimation of area sums and HW day temperatures.

A slight gap is noticeable between normal HWs and extreme ones affecting the largest areas in Fig. 4.7. This gap can be important for the definition of thresholds for the creation of an event set of the most severe HWs or mega-heatwaves. Therefore, Fig. 4.8a shows a combination of three HW aspects (diagram after Ouzeau et al. (2016)) for the All Europe area. The gap between moderate and severe HWs is visible in all data sets. Thus, the HWs are easily classified by the existing gap. However, the gap seems to depend on the size of the chosen domain. The HWs are not as easy to distinguish for smaller areas like the PRUDENCE regions or the Europe evaluation area as the gap shrinks (compare Fig. 4.8a to Fig. 4.8b).

Mega-heatwaves like 2010 are clearly visible in the observations in the All Europe area. However, this HW is not dominant in the smaller evaluation area, since it affected mostly Eastern Europe and Russia. HWs like 1994 and 2003 are more pronounced being located mainly Western and Central Europe. The EVA simulations show the large gap for the larger domain as well and are in good agreement with the observations regarding the total number of HWs. The intensity is higher in the EVA simulations as in the observations, while the durations are very similar. This is true for the smaller European evaluation domain too.

The bubble graphics comprehensively present major aspects of HWs, and it is effortless to extract thresholds to evaluate future events. The duration threshold uses the 80th percentile of the duration of all HW events of the observations. The temperature threshold is defined as the mean HW intensity (axis of the ordinate) plus the temporal SD. Due to the broader distribution of the temperatures in the EVA simulations, the temperature threshold is calculated on each data set explicitly. These thresholds are marked by the vertical and horizontal lines in the figures. The intensity threshold of Eobs is with 2.2 °C visibly lower than for the EVA runs ($<$ -0.3/-0.6 °C). Furthermore, there is a noticeable difference between CCLM EVA b1 and CCLM EVA preop for the European evaluation domain. The higher intensity in CCLM EVA b1 is the result of the overestimation of the temperatures in Southern and Eastern Europe, which is not as large as in CCLM EVA preop.

The main fraction of HW events noted in (1) is usually included in the European evaluation area. However, 15 HWs more are counted in the Eobs data in the smaller European evaluation domain compared to the bigger All Europe. Therefore, a reasonable domain size for the evaluation has to be chosen depending on the area of interest. Due to the cropping of the area, HWs at the boundary zone in the East may be cut into two in the smaller domain. Thus, two parts of the same HW can exist simultaneously. If a temporal shift occurs during the extension, the event could be counted twice explaining the higher numbers in the smaller area. Schoetter et al. (2015) discussed a similar problem. They argued that most HWs are connected and in such not two independent events will occur in the same domain concurrently. Temporal traction of HW events could be a solution to prevent the double recording of the same event.
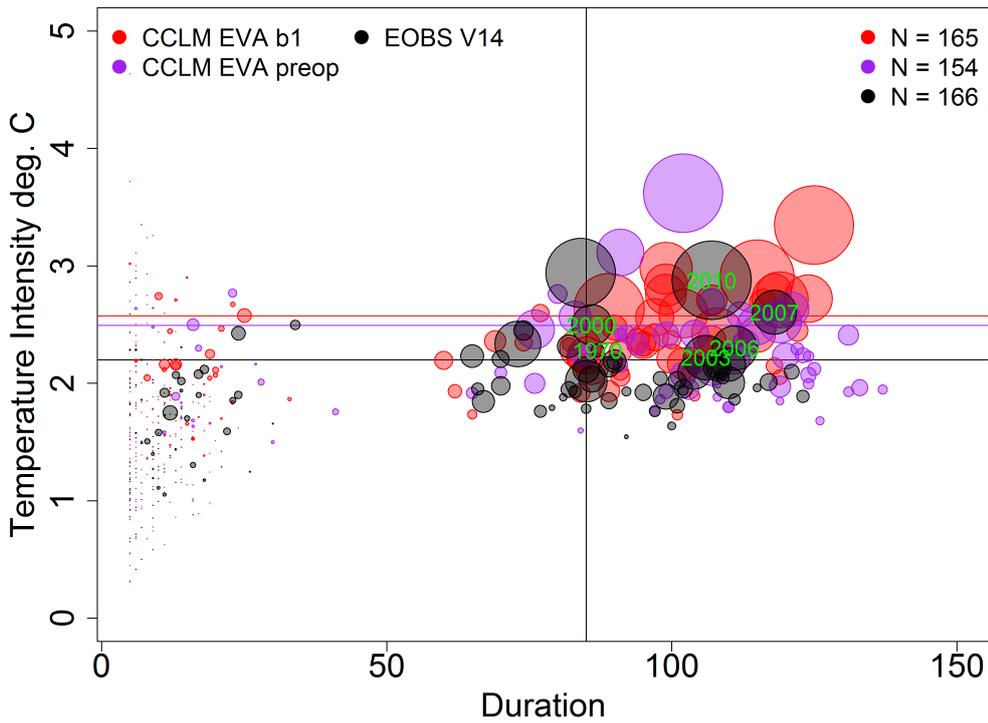
(a) Europe simulation domain (all Europe)
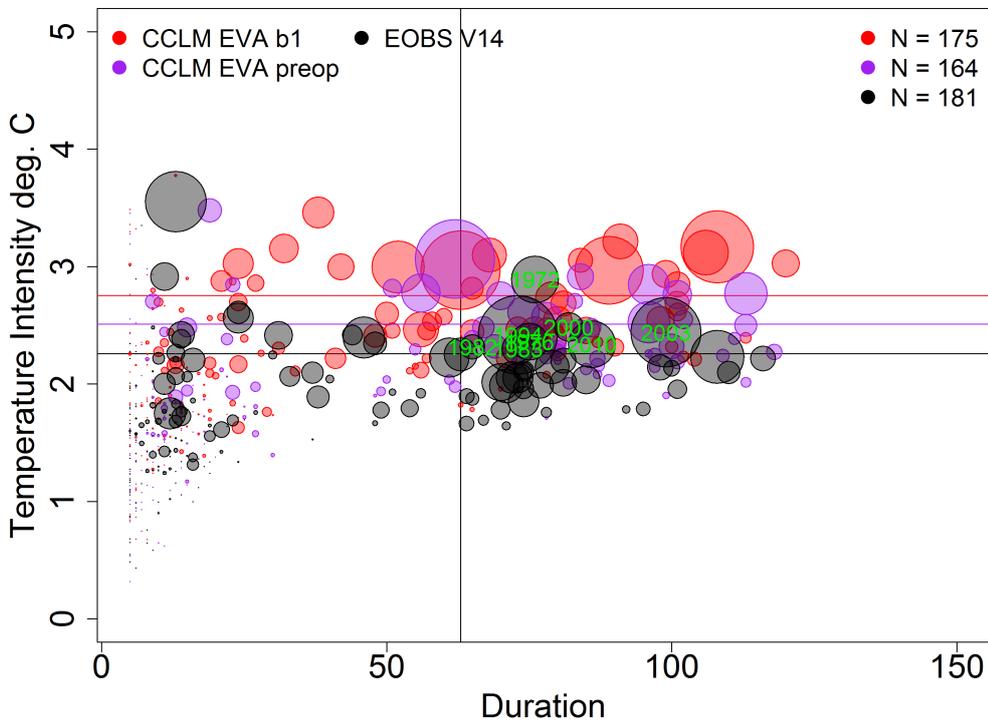
(b) Europe evaluation domain

Figure 4.8.: Bubble plot displaying three aspects (intensity sum: bubble size, duration in days and intensity mean in °C) of a HW for Eobs and the EVA simulations 1961-2010, areas averaged: All Europe (1, upper graph) and Europe evaluation domain (10, lowe4r graph). The total number of HWs is shown in the upper right corner.

The results indicate that the domain of the HWs calculation can be crucial for the building of an event set and the selection of the most intense events. Therefore, it was decided to use the simulation domain for the event set instead of the evaluation domain.

The event set is displayed in table 4.1 for six aspects of HWs with the highest magnitude for the European simulation domain (1). The intensity sum is standardised using the strongest event in the set; here, 2010 is the most severe HW from 1961-2010. All HWs after 2000 have a duration of over 100 days. Since the spatial mean was calculated for the daily data over the domain, the duration for one event is counted from the first onset until the last occurrence in the time series. The spatial fields are not checked for coherence before calculating the field means. Thus, the duration of the HWs increase. In addition, two events can be counted as one very long one, although they are not connected regionally. Temperatures exceeding 40 °C in the maximum are found for three out of the six selected HWs (2000, 2003 and 2007). HWs cover the largest area in 2003, 2006 and 2010. The selected events are in good agreement with the findings of Russo et al. (2015).

The number of events increases from six to eight if the smaller evaluation domain is used for the event set. Only the years 2003 and 2010 are included in both data sets. 1994 has a higher ranking in the smaller area; additionally, the years 1972, 1976, 1982, 1983 are listed. Most of these HWs were located in Central Europe as subsec. 5.4.1 shows later in this thesis. They are identified as the strongest HWs for the smaller Southern Germany evaluation area. Moreover, the exclusion of the Southern and Eastern regions leads to lower mean and maximum temperatures and shorter durations.

Year	Mean(Max.) temp.	Duration	Mean(Max.) Area	stand. int. sum
1970	28.0 (37.9)	87	0.03 (0.11)	0.17
2000	33.2 (41.6)	86	0.09 (0.21)	0.51
2003	31.5 (40.3)	106	0.11 (0.37)	0.57
2006	30.4 (36.7)	111	0.11 (0.33)	0.57
2007	31.9 (40.6)	118	0.10 (0.21)	0.57
2010	31.9 (37.7)	107	0.15 (0.40)	1.00

Table 4.1.: Event set for the major heat waves in the European simulation domain (1) presented for different HW aspects including the mean and maximum HW temperature [°C], duration [days], mean and maximum affected area [% GP] and the standardized intensity sum. The highest numbers for each category are marked with a red shading.

4.4.1.3. Spatial pattern of HW indices across Europe

The spatial variance across Europe is analysed for the selected HW indices. HWs mean and maximum temperatures/lengths and the number of HWs for each year are calculated and averaged for 1961-2010. The mean and maximum lengths and temperature are equal if only one HW occurs in a year. Mean and maximum HW temperatures and lengths show comparable geo-

graphical patterns; therefore, only the maximum temperatures and lengths are presented here. Fig. 4.9a displays the maximum HW temperatures per year for 1961-2010 for the observations.

(a) Eobs V14 Max HW Temp

(b) CCLM EVA b1–Eobs

(c) CCLM EVA preop–Eobs

(d) Eobs V14 Max HW Len

(e) CCLM EVA b1–Eobs

(f) CCLM EVA preop–Eobs

(g) Eobs V14 No. of HWs

(h) CCLM EVA b1–Eobs

(i) CCLM EVA preop–Eobs
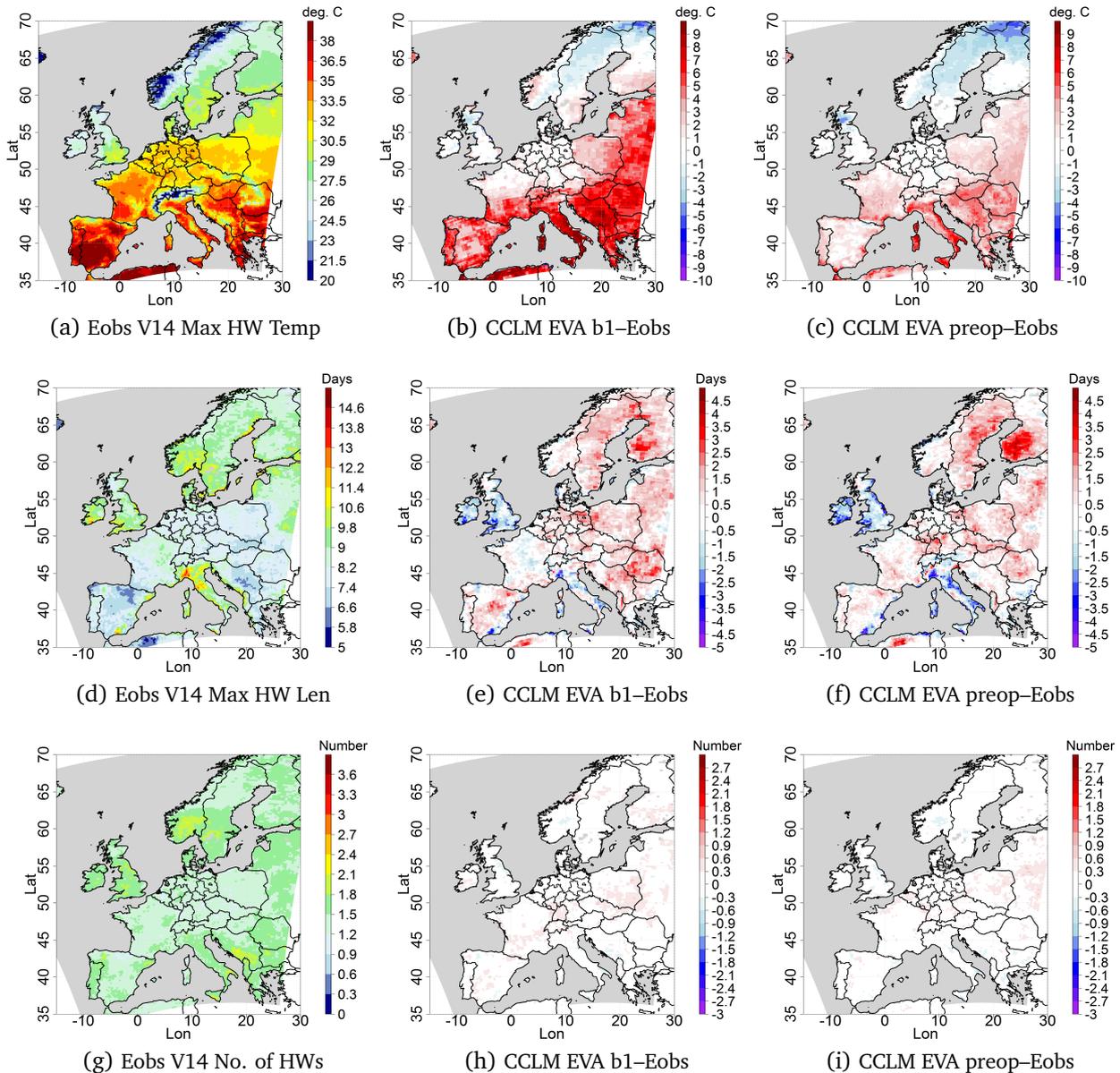
Figure 4.9.: Yearly maximum HW temperatures in °C (first), maximum HW lengths in days (middle) and numbers of HW per year (last row) averaged for 1961-2010 for Eobs V14 and the differences of the EVA simulations to Eobs (middle and left column).

A distinct South-North gradient is seen with values decreasing from over 38 °C in the Mediterranean to 20 °C in Scandinavia. Such a gradient is not found for the HW lengths (Fig. 4.9d), where the highest maximum lengths per year are noted in Northern Italy with over 13 days followed by values around 10 days for the British Isles and Scandinavia. The shortest maximum HW periods with only six days are found in the Iberian Peninsula and Southeastern Europe. These lengths are much shorter contrary to the event set since a GP based evaluation is applied

currently, whereas the event set includes area means. Thus, much longer lengths are acquired for HW meandering across Europe, whereas the same event may have a maximum length of 20 days at one GP for example. In addition, the time averaging shortens the maximum HW lengths. Therefore, more short events occur than mega-heatwaves. The number of HWs varies not very much across Europe with 1-2 HW per year in nearly all regions (Fig. 4.9g). The lowest number of HWs is found for regions like the Iberian Peninsula and France with only 1.2 HW per year and doubles for Southern Scandinavia and the Balkan States.

The number of HWs per year of the EVA simulations fits very well to the observations (Fig. 4.9h and 4.9i), they overestimate the number of HWs by +0.5 per year at most. The HW durations are often too long for the mainland of Europe. On the other hand, an underestimation of up to -4 days in the maximum HW lengths is noted for the British Isles and Italy. Both simulations are nearly identical in their spatial pattern (Fig. 4.9e and 4.9f). The larger overestimation of ~1.4 °C by the CCLM EVA b1 simulation is noticeable in the maximum temperature difference plots (Fig. 4.9b and 4.9c). This run tends to produce overly warm regions in Southern and Eastern as seen before in the 90th percentile (Fig. 4.5b). This warm bias is less pronounced in CCLM EVA preop. Furthermore, the maximum HW temperatures show a similar cold bias in the EVA simulations as for the 90th percentile in Scandinavia. All in all, a good agreement between the EVA simulations and the observations is found. These results fit well to the findings of Vautard et al. (2013). Also, they noticed too persistent and too strong HWs for models of the EURO-CORDEX ensemble driven by ERA-Interim. Jaeger and Seneviratne (2010) offer a possible explanation by pointing out the critical role of land-atmosphere processes (see chapter 2.2) and their representation in the models. The investigation showed that changes in these feedbacks could enhance the development of HWs (higher temperatures and numbers) or diminish them in the same measure. Furthermore, Vautard et al. (2013) found an increase in the frequency of hot days in case of reduced precipitation, which might be similar for the simulations in this subsection.

4.4.1.4. Temporal trends in HW indices

The largest events of the Eobs data are displayed in table 4.1 for the simulation domain (1). It can be seen, that the most intense events occur after 2000. Several studies like Barriopedro et al. (2011) and Russo et al. (2015) confirm this, demonstrating the importance of including the climate trend in temperature extremes. A 4-year running mean of yearly HW temperatures anomalies is shown in Fig. 4.10 spatially averaged for 1961-2010. The mean HW temperature anomalies of the Eobs data are lower in the decade 1971-1980 with a slight increase until the decade 2001-2010. Afterwards, they remain quite stable. The largest anomaly amplitude is found for CCLM EVA b1 (Fig. 4.10), while the CCLM EVA preop amplitude is much smaller (especially after the decade 1961-1970). In particular, the anomalies of CCLM EVA preop fit noticeably well to the observations from the mid of the 1980s. The correlations are calculated for the yearly HW temperatures. Correlation values of > 0.75 have been found for the HW temperatures for both EVA runs to Eobs similar to the high correlations of the daily maximum

temperatures (subsection 4.3.1). Thus, the dominating climate trend is captured well by the reanalysis runs as expected.
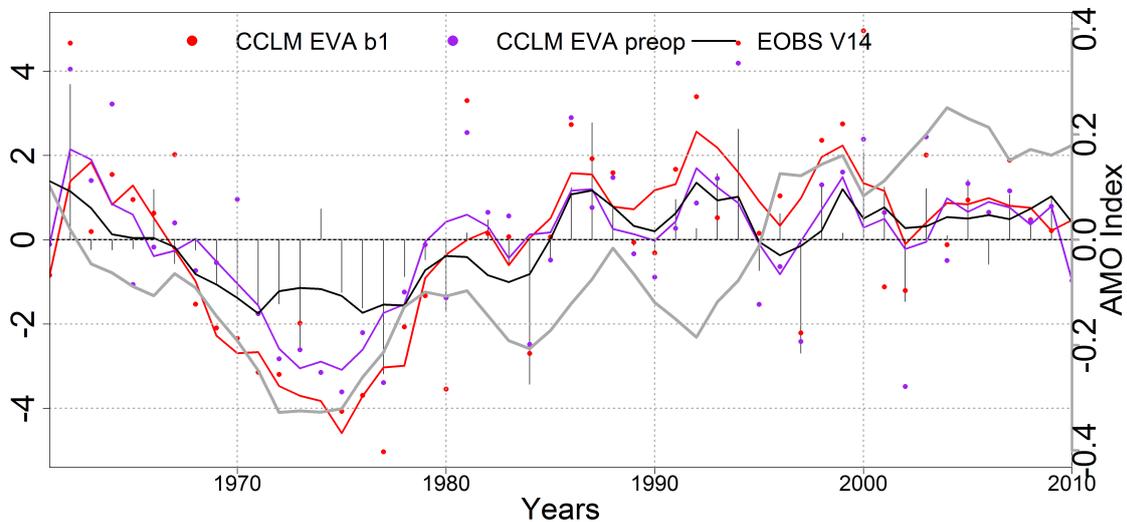
Figure 4.10.: 4 year running annual mean HW temperatures anomalies for 1961 – 2010 CCLM EVA simulations and Eobs V14 data for Europe (10). In grey, the 4 year running mean AMO index is displayed (Enfield et al., 2001).

Depending on the specific research question, it can be important to detrend variables. Therefore, the heat waves indices have been calculated with and without the linear climate trend. The trend of the mean HW length in the observations is +0.03 days per year for Europe and slightly higher for the maximum HW length with 0.05 days per year (see Fig. A.7 in the appendix). The climate trend for the HW temperatures is marginal lower than for the HW lengths with +0.02 °C for the mean and +0.03 °C for the maximum HW temperatures (Fig. A.6 in the appendix). On the other hand, the trends for the temperatures are quite noticeable between the decades (as seen in the previous figure) and not as monotonously increasing as for the length or the number of HW per year. Stronger trends have been found in the Mediterranean/Southern regions with +0.05/+0.08 days per year for the mean/maximum HW lengths. In contrast, the trends were near 0 and insignificant in Scandinavia (not shown). These results show that the trends in the HW variable are not negligible for parts of Europe. Other studies as Fischer and Schär (2010) found similar spatial trends with the strongest increase in Southern Europe. Within the MiKlip project, user relevant variables will be investigated and future periods simulated. Thus, the trend is included to give reliable information about HW temperatures and the change signal.

Long-term climate variations like the AMO can be connected to HW development as well. The AMO, as known from Schlesinger and Ramankutty (1994), is the leading mode of internal (natural) variability of the North Atlantic SST on decadal time-scale (65-70 years). Many studies investigated AMO phase changes and established that they are a source of predictability for variables like temperature and precipitation in Europe (Sutton and Hodson, 2005; Knight et al., 2006; Keenlyside et al., 2008; Folland et al., 2009). In Fig. 4.10, the detrended AMO index

can be seen as grey line. A connection of the AMO index to surface temperature extremes seem to exist in Europe (Fig. 4.10). Correlation values of r > 0.5 for yearly HW temperatures and the AMO index are found in all three data sets. However, it remains unclear, how much of the trend of the HW temperatures is related to phase changes in the AMO and how much can be attributed to external forcings like greenhouse gases and others. Part of the variability in the European averaged time-series is related to the increase of HW temperatures in Southern and Eastern Europe during the AMO positive phases. The Western domains, like BI, FR or ME, show a vice-versa increase of HWs during negative phases (not displayed). Nevertheless, the literature (Sutton and Hodson, 2005; Knight et al., 2006; Sutton and Dong, 2012) suggests a decrease of HW temperatures during the positive AMO phase in these regions. One can assume, that reduced HW temperatures are not seen explicitly in the data sets because of the superimposed climate trend (especially in the decade 2000). In general, the AMO is an essential factor to predict extreme temperatures and HWs on multi-year to multi-decadal scale beside the trend of external forcings like greenhouse gases (Kruschke et al., 2016).

4.4.1.5. Teleconnections to large scale atmospheric circulation patterns

One way to describe and determine the source of seasonal to inter-annual fluctuations (< 1 year) of the climate is to analyse the large-scale quasi-stationary circulation pattern (LCA). The variability of temperature or precipitation is strongly correlated to LCAs in corresponding regions over these time scales. Therefore, HWs are expected to be closely related too (Della-Marta et al., 2007b; Rust et al., 2015; Sousa et al., 2018). Teleconnections between selected indices and HWs in the Eobs V.14 data and the ERA driven CCLM simulations are analysed in the first part of this section. How well the MiKlip System can reproduce the LCA pattern is examined in the hindcast section (subsection 4.4.2.3).

Barnston and Livezey (1987) determined the most frequent pattern by applying the rotated principal component analysis (RPCA). Graphics of the pattern described in their study and the historical time series can be found at the National Oceanic and Atmospheric Administration (NOAA) Climate prediction center (NOAA / National Weather Service Climate Prediction Center, 2018b). For the RCPA method, the eigenvectors of the cross-correlation analysis based on 700 hPa daily geopotential height (aggregated to monthly data) are calculated, scaled and then rotated. Afterwards, the first ten components defining the main modes are selected (explaining 80 % of the winter and 70% of the summer variance). They implicate potential predictability due to their persistence and their reoccurrence. If model systems capture these LCAs correctly, it can be assumed, that they will simulate the connected short-term fluctuations more realistically as well. An alternative classification to describe the LCAs is the "Grosswetterlagen" after (Hess and Brezowsky, 1977; Gerstengarbe et al., 1999), which was employed by Kyselý (2008) in a similar study. Although the connection between HW days and highs or ridges over Central Europe was determined, the classification proved less effective for stations outside of Central/Western Europe.

Rust et al. (2015) found positive correlations (0.2 > r > 0.6) to NAO, EA, SCAND and POL (abbreviations in the following list) for July temperatures of the Eobs data in parts of Europe, while EAWR and WP were mostly negative correlated (-0.8 < r < -0.2). This study supports the assumption of connections between HW days to LCA patterns. An attempt is made in this thesis to link the monthly HW days sums to monthly LCA indices. The listed patterns are active during all months, except for the PT, which is restricted to August and September.

The following LCA pattern have been selected:
- EA: East Atlantic
- EAWR: East Atlantic/Western Russia
- EPNP: East Pacific/North Pacific
- NAO-CPC: North Atlantic Oscillation
- PNA: Pacific/North American
- POL: Polar/Eurasia
- SCAND: Scandinavia
- WP: West Pacific

Spatial correlation patterns between monthly HW day sums and the LCA patterns have been produced agreeing with the results of Rust et al. (2015) in general. However, the correlation values (both negative and positive) are lower for most patterns (not shown). Reason for the lower correlation values is probably the choice of the input data, Rust et al. (2015) used monthly mean Eobs temperatures contrary to the integrated monthly HW days sums in this thesis. The phase dependence is tested by filtering the LCA time series into their positive and negative phase. Thus, higher correlations might be achieved as well.

The index time series are correlated to the corresponding monthly HW day sums of the Eobs V.14 observations. The correlations are displayed in Fig. 4.11a and Fig. 4.11b for the PRUDENCE regions. The highest correlation values exceed +/- 0.4 for single LCAs (NAO, EA, EAWR and PT) mostly in the smallest and/or the westernmost regions. In addition, the strength of the signals in their different phases is not symmetrical, as it was expected. Vice-versa signals are found only for certain patterns like the EAWR and parts of the WP and PT. For others, the signal strength varies or signals are found just in one phase, like for the NAO-HPC (NAO), the PNA or the WP. In the positive phase, the highest positive correlations are found for the EA pattern in MD, AL, FR and ME, for the PNA pattern in MD and high anti-correlations for the PT pattern in ME and SC and the SCAND pattern in IP (Fig. 4.11a). However, correlation values are often < 0.2 for most of the other indices and regions.
The negative phase shows high positive correlations for the NAO in the Western regions like BI, FR and ME or MD (Fig. 4.11b). The highest negative correlations are noticed for the EA and PT patterns in BI, nearly all regions in Western/Central Europe and MD. Compared to the positive phase, stronger signals are detected in the Western/Central European regions. However, this is

mostly the case for Eastern/Southern areas like AL, EA and MD in the positive phase.
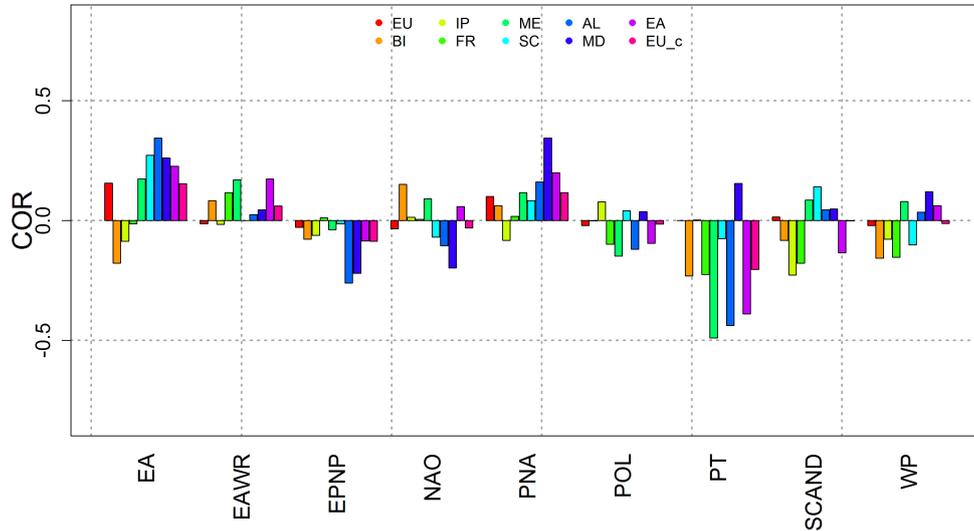
(a) pos. fraction
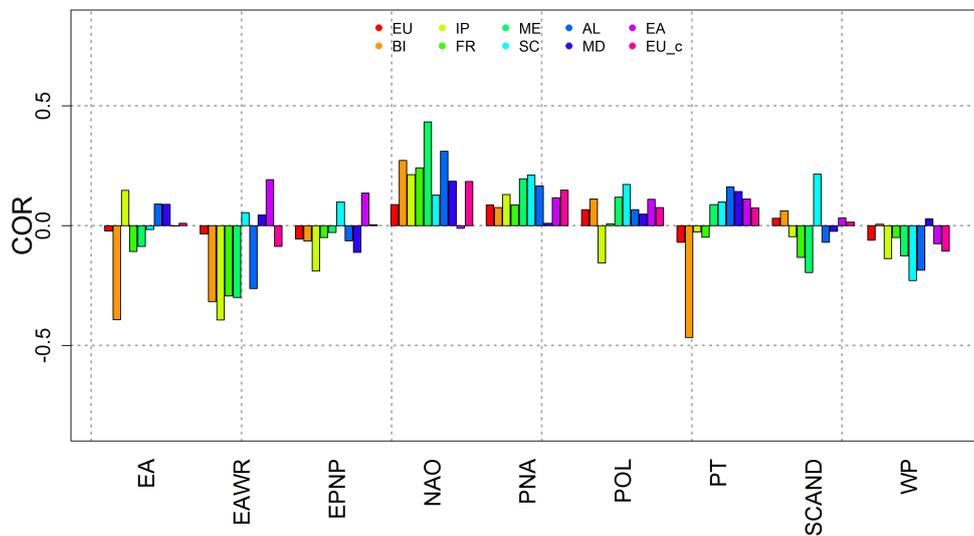
(b) neg. fraction

Figure 4.11.: Correlation for different PRUDENCE regions and the positive and negative phase of several indices to monthly number of HW days derived from Eobs V.14 data for Europe.

The filtering for positive and negative phases increased the correlations and gave a first insight into the workings of the pattern. The filter levels are shifted from $\pm\,0$ towards the upper and the lower third of the indices to derive more distinct signals for the positive and negative phase, see

Fig. 4.12a and Fig. 4.12b. Higher negative correlations are found for the patterns NAO, PNA, PT, POL and SCAND using only the upper third fraction of the index. In the negative phase, the increase of the correlation is similar, although it affects other patterns (mainly EA, PT and EPNP). A stronger differentiation between the PRUDENCE regions can be seen for the PT pattern in the negative phase and the PNA in the positive phase separating the Northwestern/Central regions (SC, FR, ME) and Southeastern (MD, EA) ones.
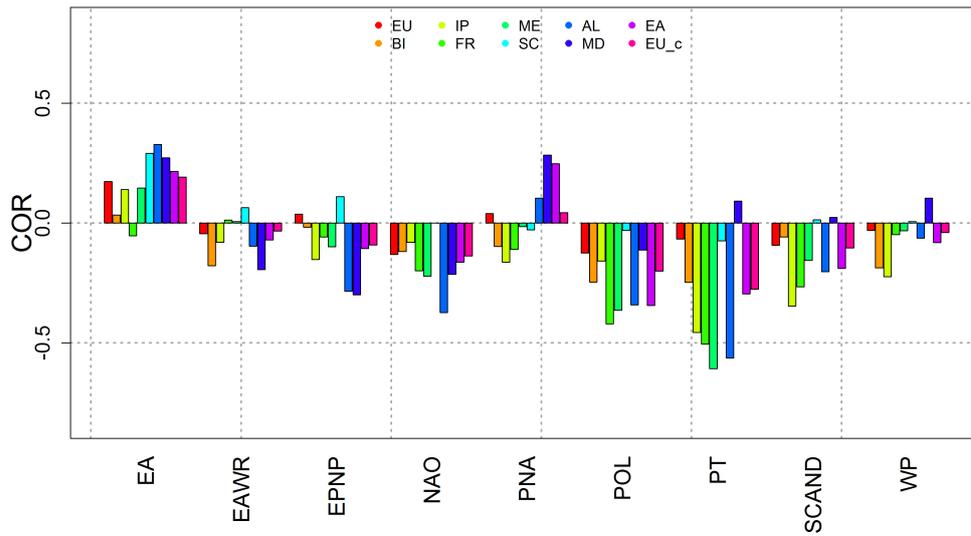
(a) pos. fraction
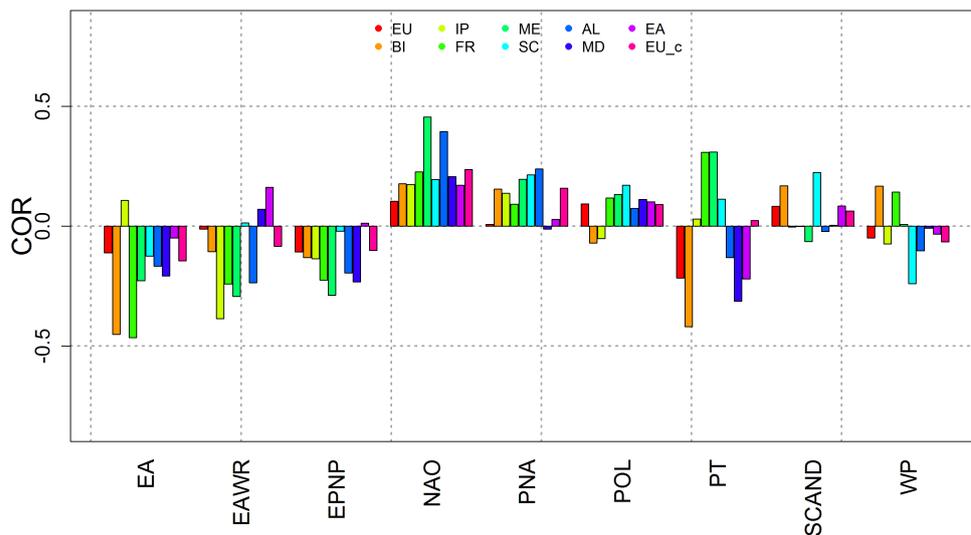
(b) neg. fraction

Figure 4.12.: Correlation for different PRUDENCE regions and the positive and negative upper third fraction of several indices to monthly number of HW days derived from Eobs V14 data for Europe.

LCAs often show correlations between each other or other circulation patterns like the El Niño Southern Oscillation in the Pacific or the equatorial Quasi-Biennial-Oscillation (QBO). Other studies, e.g. Pinto and Raible (2012) investigated the connection between the PNA and the NAO. Although their link is found to be variable, depending on their phases and the North Atlantic SST, a certain similarity in the signal seems to exist. Other systematic responses to phase changes may be the result of the structural analogues of the pattern like for example NAO and EA, which are both defined by a North-South dipole. NAO and EA are the most prominent pattern in the North Atlantic (Barnston and Livezey, 1987; NOAA / National Weather Service Climate Prediction Center, 2018b; Rust et al., 2015). One difference between these patterns is the location of their centres of action of the dipole.

This location difference can be one explanation for the connection of HW days to different phases of the corresponding index. The NAO centres shift northward during the summer months, while the EA is more stationary and located southwardly linking it to subtropical ridges. As Sousa et al. (2018) showed, an explicit connection exists between maximum temperatures and the area of ridges extending from the subtropics northwards. In addition, patterns, associated with the formations of blocking and/or ridges like the SCAND (Bueh and Nakamura, 2007; Sousa et al., 2018; Cassou et al., 2005) or the NAO in winter (Woollings et al., 2008), can be assigned to the positive or negative phase.

Four patterns and their possible impact on the HW days will be discussed as examples due to the diverse characteristics and mechanisms in more detail below. The complete European simulation domain except for the relaxation zone (1, see Fig. 4.1) is displayed to show the wide spatial extension of the pattern and the centres of action located in Eastern Europe and Northwestern Russia. Not all graphics can be found in this section, but are shown in the appendix in section A.3. The filtered HW day sums and the correlation can be seen there for the upper third and lower third.

NAO: The North Atlantic Oscillation (NAO) is the leading mode in the North Atlantic and defined by two pressure systems (Icelandic Low and Azores High). The dipole shifts northwards during the summertime, then, the southern node (the high) is located near Europe (Barnston and Livezey, 1987; NOAA / National Weather Service Climate Prediction Center, 2018b). Higher HW days numbers and positive correlations are expected because of this anticyclone in the positive phase. Additionally, the high-pressure system over Northwestern Europe in the positive phase is similar to the one associated with the positive phase of the Polar/Eurasia (POL) pattern (NOAA / National Weather Service Climate Prediction Center, 2018b). Regions directly below the mentioned anticyclone experience enhanced subsidence and continental warmer air masses are advected to Western Europe. In contrast to this, the negative phase is dominated by a large cyclonic structure in the same location (see Cassou et al., 2005, Fig. 1c) transporting cold and wet air towards Western and Central Europe.

(a)

(b)
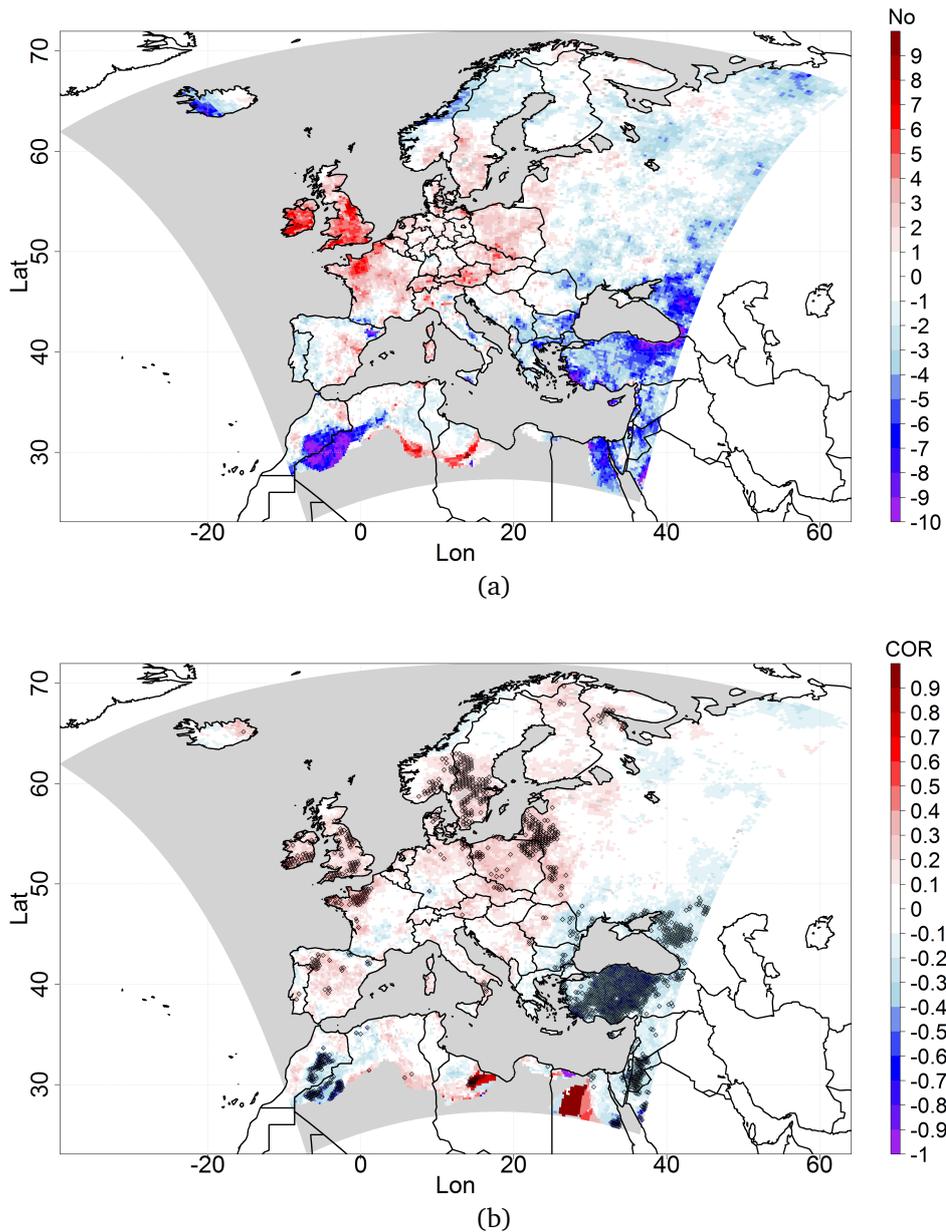
Figure 4.13.: Upper panel: Difference of No of monthly HW day sums between the highest third and the lowest third of the NAO index for Eobs V14 data for Europe. Lower panel: Correlation of the NAO index for monthly sums of HW days from Eobs V14 data for Europe. Marked by black circles are grid boxes with a significant correlation (95 % confidence interval).

Fig. 4.13a displays the difference of the monthly HW day sums between the upper third and the lower third of the NAO index (specific graphics of HW day sums for the positive third and the negative third in the appendix: Fig. A.2c and Fig. A.2d). The difference plot shows higher numbers of monthly HW days in Central and Northwestern Europe for the positive phase, whereas larger values are found in Southeastern Europe during the negative phase (Fig. 4.13a). Fig. 4.13b displays the correlation of the NAO index to the monthly HW day sums of the Eobs V.14 data. The correlation pattern has a very similar spatial distribution as the difference graphic

before. Weak positive and partly significant correlations (r < 0.4) between the HW day sums and the NAO are found for most of Europe, while a significant, negative correlation of equal strength is seen in Turkey (Fig. 4.13b).

Moderate positive correlations dominate Central and Western Europe during the negative phase, which can be found in the negative phases of the PNA (Fig. A.4f) and the POL (Fig. A.5b) in a similar fashion. On the other hand, only weak positive correlations in Scandinavia are found (r < 0.4) in the positive phase. Although a higher number of HW days in Northwestern Europe occur, no direct connection can be established between HW day and the NAO. Nevertheless, significant anti-correlations are found in Southern/Southeastern Europe in the positive phase aligning the lower number of HW days with the positive index values. The negative pressure anomaly over Europe seems to suppress the development of HWs stronger during the negative phase than the anticyclone enhances them in the positive phase. This finding is confirmed by Cassou et al. (2005), who noted mainly cold regimes in France during the NAO- phase. Therefore, no linear connection between the phases seems to exist explaining the non-symmetric signal seen in the barplots before.

EA: A dipole structure in the North Atlantic describes the East Atlantic pattern (EA) alike the NAO (Barnston and Livezey, 1987; NOAA / National Weather Service Climate Prediction Center, 2018b). The extension of a subtropical ridge creates the southern node enveloping Southern Europe and reaching up to Central Europe, while the corresponding northern low is located Northwest of Ireland. During the positive phase, more HW days are found in Southern Europe as the ridge elongates and intensifies. Rodríguez-Puebla et al. (2010) noticed a similar increase for warm days for IP during the positive EA phase. Along the ridge, warm air is advected into Eastern Europe resulting in an increase of HW days sums in this region (Fig. A.4a in the appendix). Significant positive correlations between HW days and the pattern index are found in the same area (Fig. 4.14b). In contrast, the British Isles are located northward/upstream of the ridge and under the influence of the corresponding low in the East Atlantic. Unlike Central Europe, colder and moister Atlantic air is transported into the UK. This leads to a decrease of HW days as can be seen in Fig. 4.14a during the positive phase and vice-versa to a higher number of HW days occurring during the negative phase (Fig. A.4b in the appendix).

PT: NOAA CPC (NOAA / National Weather Service Climate Prediction Center, 2018b) describes the Pacific Transition (PT) pattern as the leading mode of August and September. The pattern is associated with anomalous wave trains in 500 hPa over the Northern Pacific and the United States (also Barnston and Livezey, 1987). Downstream, ridges can be found in the North Atlantic reaching the Iberian Peninsula and extending into Europe in the positive phase. Along the ridge, colder and moist Atlantic air is transported towards the British Isles, Southern Scandinavia and Central Europe (Fig. A.3c in the appendix). This suppresses the development of hot temperature extremes in these regions. In the negative phase, the wave train location shifts and more warm extremes occur in the mentioned regions seen by negative differences in Fig. 4.15a.

(a)

(b)
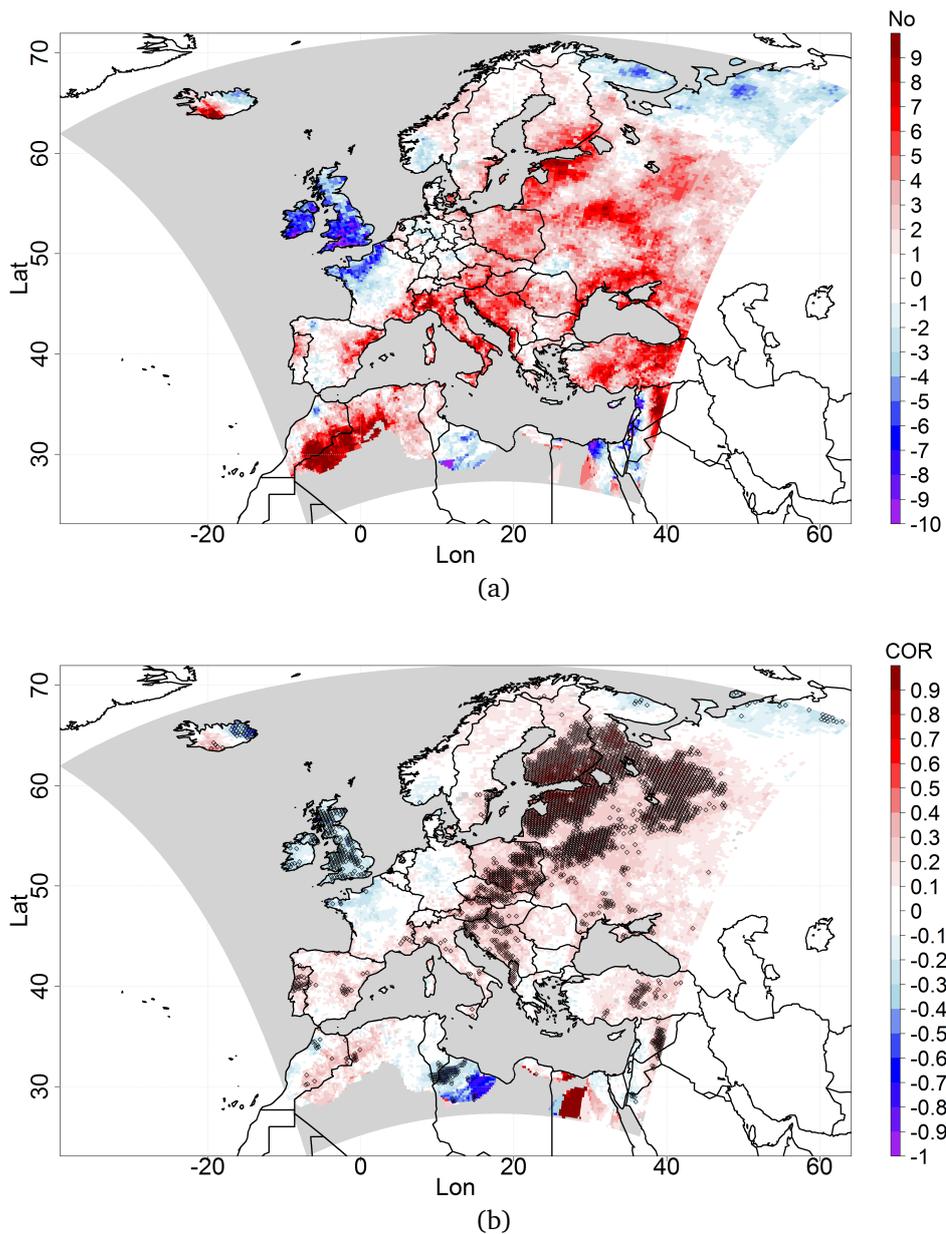
Figure 4.14.: Upper panel: same as Fig. 4.13a but for EA. Lower panel: same as Fig. 4.13b but for EA.

Since the PT pattern is described only for one or two months, a lower number of values is used for the correlation calculations. This leads to the higher correlation values compared to other pattern (Fig. A.5c and Fig. A.5d in the appendix) for both phases.

SCAND: The SCAND pattern represents the evolution of a large anticyclonic structure over Scandinavia and Eastern Russia. A higher number of HW days is found for Scandinavia under this anticyclone during the positive phase (Fig. 4.16a). There, the highest positive correlations between the HW day sums and the index are found as well (Fig. 4.16b). The linear connection between the HW day sums and the SCAND positive phase is not as strong as the weak positive, although significant, correlations in Northern Europe show (r < 0.4, Fig. A.5e).

(a)

(b)
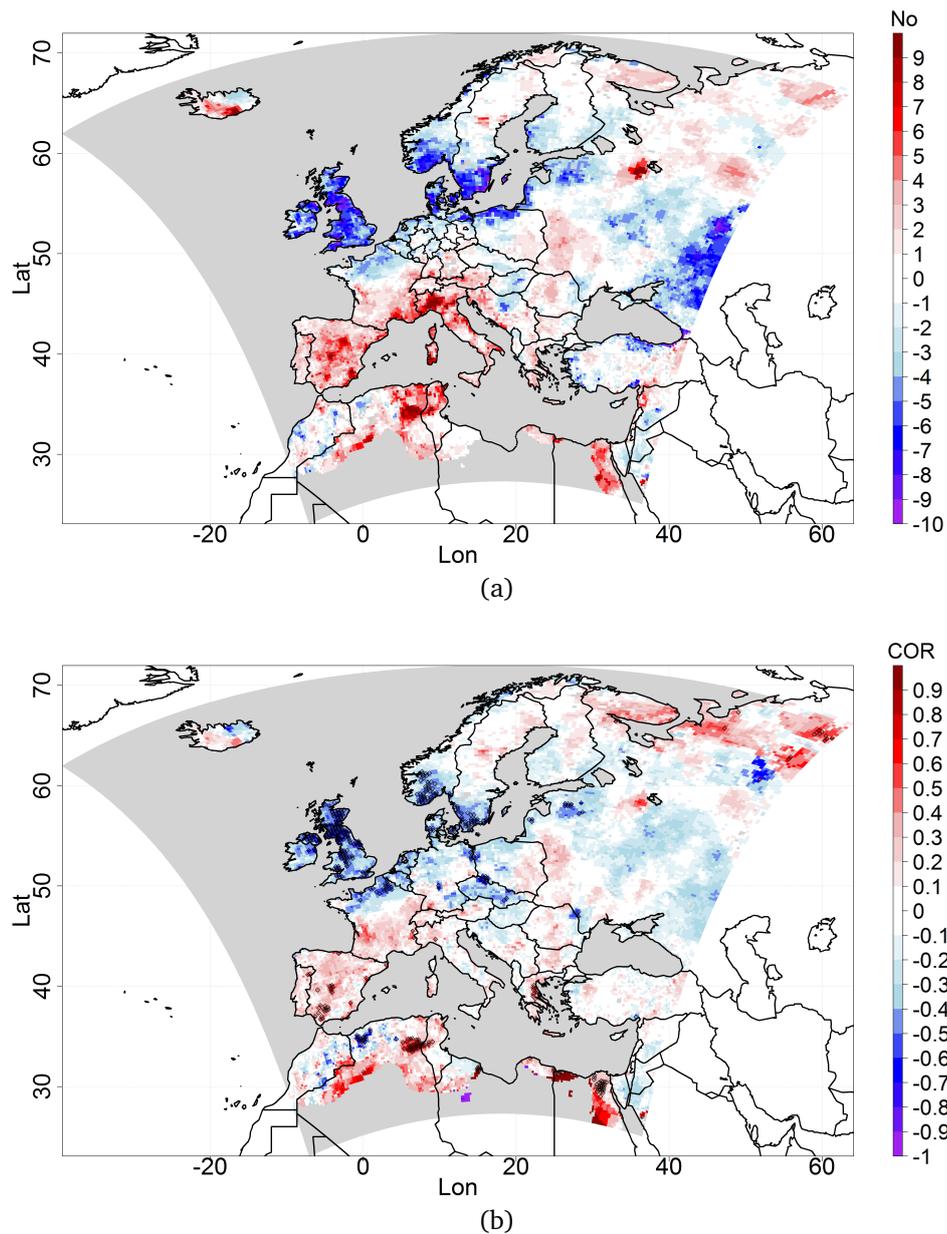
Figure 4.15.: Upper panel: same as Fig. 4.13a but for PT. Lower panel: same as Fig. 4.13b but for PT.

In addition, weak, but significant, negative correlation values are detected for the Iberian Peninsula during this phase. Reason for this is the cyclonic structure in Southwestern Europe opposing the Scandinavian high. When, along with the anticyclone over Scandinavia, the cyclone intensifies during the positive phase, it enhances the colder and wetter conditions found for the Iberian Peninsula. Therefore, a lower number of HW days is noted in Fig. A.5e. Studies as Della-Marta et al. (2007b) identified the opposing HW anomalies in SC and IP as well. The negative phase is associated with a lower number of HW days in Scandinavia (Fig. A.3f), which are positively correlated to the negative third of the SCAND index (Fig. A.5f). Different mechanism, as the shifting of anticyclones and ridges along with increased wave activity, influence the link of
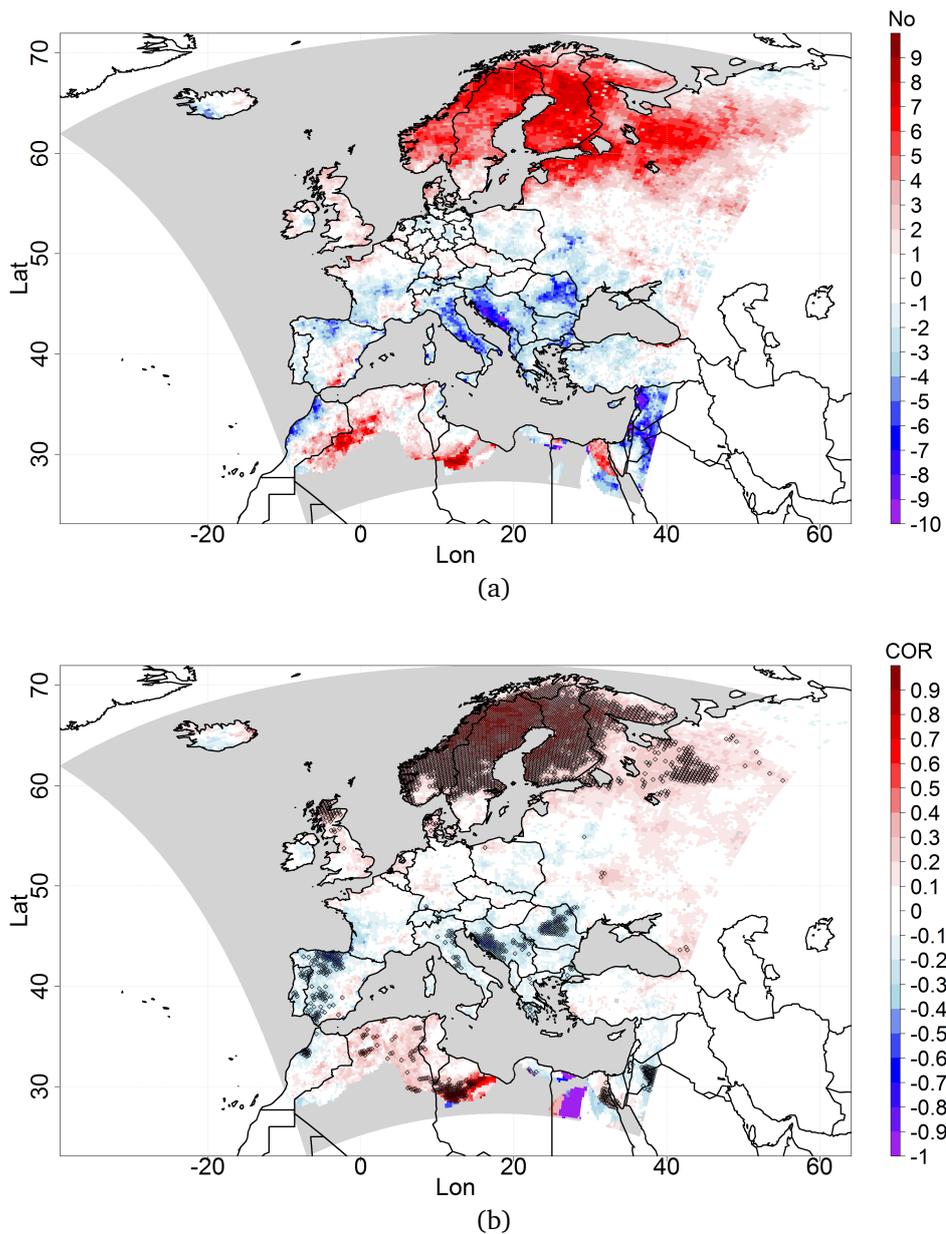
Figure 4.16.: Upper panel: same as Fig. 4.13a but for SCAND. Lower panel: same as Fig. 4.13b but for SCAND.

HW days to LCA pattern. However, this link is not symmetric in its strength for the different phases and regions as said before. In the following, correlations are shown to determine, if the EVA simulation can reproduce the links to the LCA pattern found in the Eobs data for the different PRUDENCE regions. Two (Mid-Europe and Eastern Europe) are selected as examples, due to the number of patterns and PRUDENCE regions, based on their correlation signal strength and location. Although the correlations vary quite largely between the three data sets for some pattern, values are found with analogue prefixes (Fig. 4.17a until Fig. 4.18b).

(a) ME

(b) EA
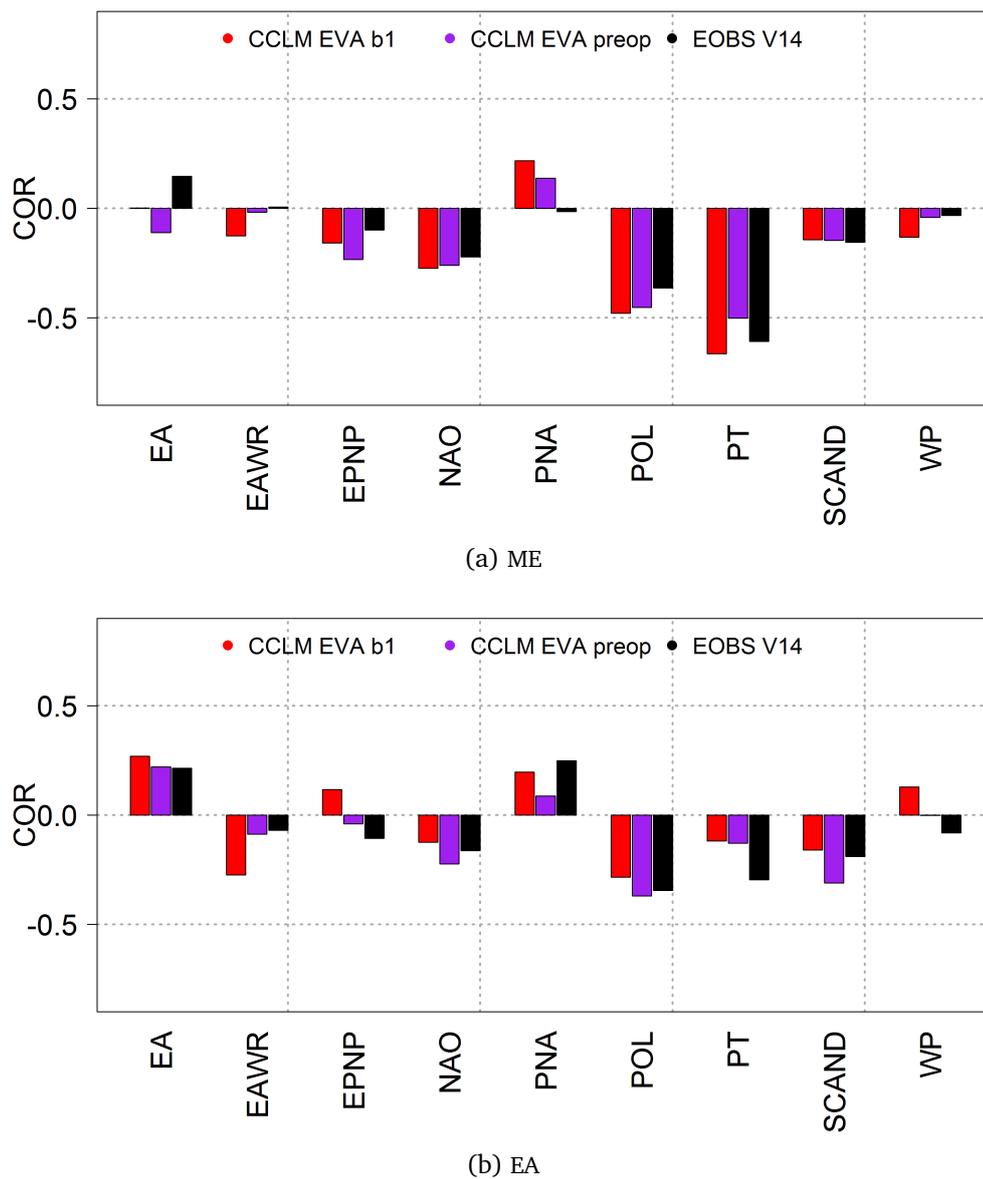
Figure 4.17.: Correlation for the Mid-Europe (upper barplot) and Eastern Europe (lower barplot) of the positive third of selected LCA to monthly number of HW days derived from CCLM EVA simulations and Eobs V14 data.

Again, the non-symmetric correlation distribution is noticed for the EVA runs for most patterns. A visual comparison revealed generally fitting spatial distributions between the EVA simulations and the Eobs derived patterns. In the EVA runs, slightly higher correlations to the pattern are found for the positive phase for ME. In the negative phase, opposing signals are noteworthy for NAO and PT. The newer CCLM EVA preop simulation has a slightly better agreement with the observations fitting in 34 of 50 cases (considering all cases and the PRUDENCE regions, 5 pattern states∗10 Regions). Seldom, the simulations show a signal contrary to the observations. Opposing signals within the EVA runs are rare as well. The negative phase of the patterns seems to have a less distinct effect in Eastern Europe, according to Fig. 4.18b. Heterogeneous

correlation patterns including many contrary signals are noticed for various LCAs. In such, the boundaries of the evaluation area Eastern Europe or the region itself could be inappropriate for the time series analysis as it often includes correlation values compensating themselves. In addition, this is the case for Central Europe as well, but to a lesser degree.
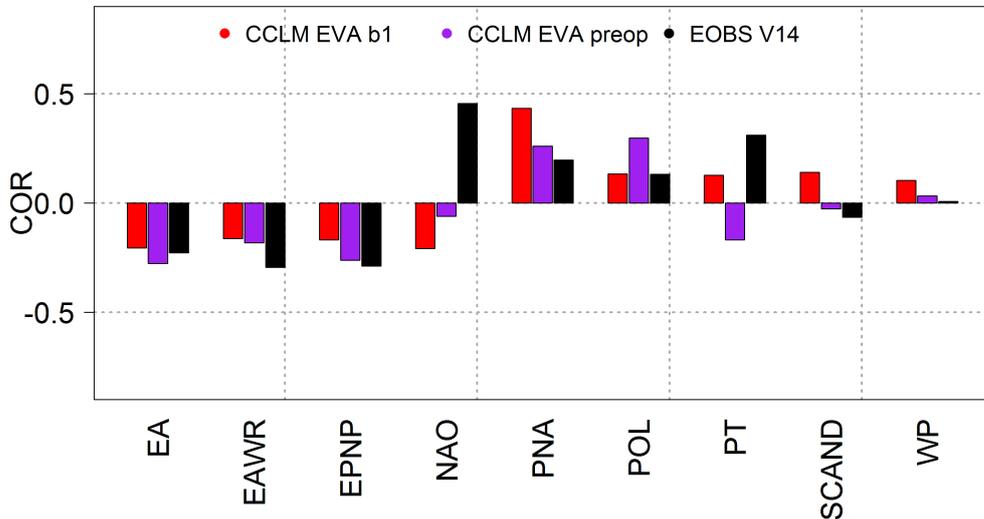
(a) ME
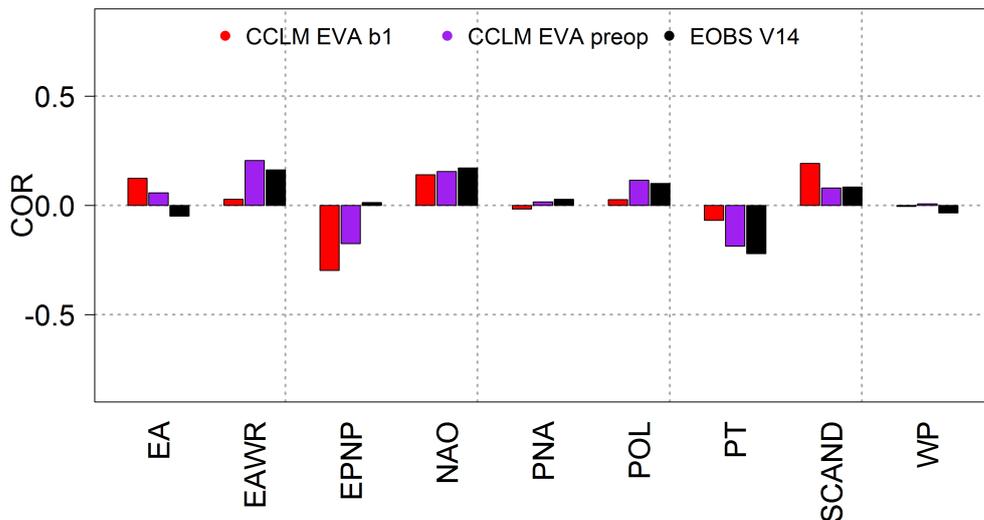
(b) EA

Figure 4.18.: Correlation for Mid-Europe (upper barplot) and Eastern Europe (lower barplot) of the negative third of selected LCA to monthly number of HW days derived from CCLM EVA simulations and Eobs V14 data.

4.4.2. Decadal predictions for Europe

The decadal hindcasts of the b1 ensemble are evaluated to determine the quality of these predictions in terms of HWs. The lead-time dependency of the reference percentile is investigated, because of the strong drift in the first year found in subsection 4.3.2 for the daily maximum temperatures. Alternative techniques are explored to evaluate HWs on event-level due to the various ensemble members and the lead-time interval.

4.4.2.1. Lead-time dependency of the 90th reference percentile

Studies like Reyers et al. (2019) and Uhlig (2016) investigated the dependency of the skill scores to lead-times, starting years and ensemble size for multi-year temperature means and precipitation sums. In general, the skill scores increased for higher lead-times like 2-5 years or 3-6 years and for yearly starting years compared to decadal initialisations. After the first year, large temperature deviations in Eastern Europe are reduced, resulting in higher MSESS values for longer lead-times. Due to the drift in the initialisations year in the mentioned regions, the percentile threshold is expected to differ between the lead-times. It has to be determined, how large the drift of the percentile is. Zhang et al. (2005) mentioned the risk of inhomogeneity in percentile-based indices due to discontinuities in time series. Therefore, the 90th percentile is calculated for different lead-times regarding the strong drift during the first year. The investigated lead-time intervals range from lead year 1 to 7-10. Periods longer than four years were not considered because of the long averaging time. The analysis of the observation data has shown, that HWs indeed have large year-to-year as well as intraseasonal deviations (previous subsection). Thus, the four-year interval is a compromise between the averaging of the lead-times (reduction of the signal-to-noise ratio) and the visibility of the strongest HWs exceeding the multi-year variability. In table 4.2, the differences between the selected lead-times can be seen derived from the b1 RCM ensemble.

tasmax [deg. C]	Min	Max	Mean	Diff to LT1-1
LT 1-1	4.34	54.59	31.00	
LT 1-2	4.39	54.66	30.87	-0.13
LT 1-3	4.43	54.64	30.76	-0.24
LT 1-4	4.45	54.63	30.69	-0.31
LT 2-5	4.46	54.64	30.54	-0.46
LT 3-6	4.47	54.63	30.46	-0.54
LT 4-7	4.47	54.64	30.43	-0.57
LT 5-8	4.49	54.63	30.4	-0.60
LT 6-9	4.51	54.62	30.4	-0.61

Table 4.2.: The 90ths percentile for the reference period of different lead-times of CCLM b1 hindcasts and the difference to the 90ths percentile of lead year 1.

The initial year and lead-time averages including it show a larger drift for the minimum temperatures (+0.17°C), which is lower in the maximum values (+0.03 °C). A stabilisation is noticeable

for years 2-5 and 3-6. The deviations between the lead-times are negligible for the intervals at the end of the decade. Thus, the HW of the RCM and GCM ensemble will be computed for the lead-time 2-5 considering these results and prior studies. Unlike the baseline1 simulations, the historical simulations ended in 2010. Therefore, the evaluation period has to be cropped to 1962-2007 to include complete intervals until 2010 for the seven corresponding members of the regional and global hindcasts and the historicals.

Fig. 4.19a to Fig. 4.19c show the difference in the 90th percentile for lead-time year 2-5 of the global model MPI-ESM-LR b1, the regional CCLM b1 and the historicals to the Eobs V.14 data. The spatial distribution of the absolute values of the 90th percentile from the observations is shown in Fig. 4.5a. In the average, the 90th percentile is 26.5 °C in the Eobs V.14 data for Europe and with this exceeds the definition of a summer day (daily maximum temperature > 25 °C Zhang et al., 2011). The lowest difference is found for CCLM b1 hindcasts with -1.3 °C, while the regional historicals and MPI-ESM-LR b1 show a higher underestimation of -2.4 °C and -3.2 °C. Northern Europe, as seen in the figures, depicts a cold bias in all model simulations, while a small warm bias is noticed in the RCM simulations in Southern Europe (CCLM b1 and historicals). In the global data, the coastal regions show the highest underestimations of the temperature, which is often a result of the interpolation to the finer grid mixing land and water grid points.

(a) MPI-ESM-LR b1–Eobs   (b) CCLM b1–Eobs   (c) historicals–Eobs
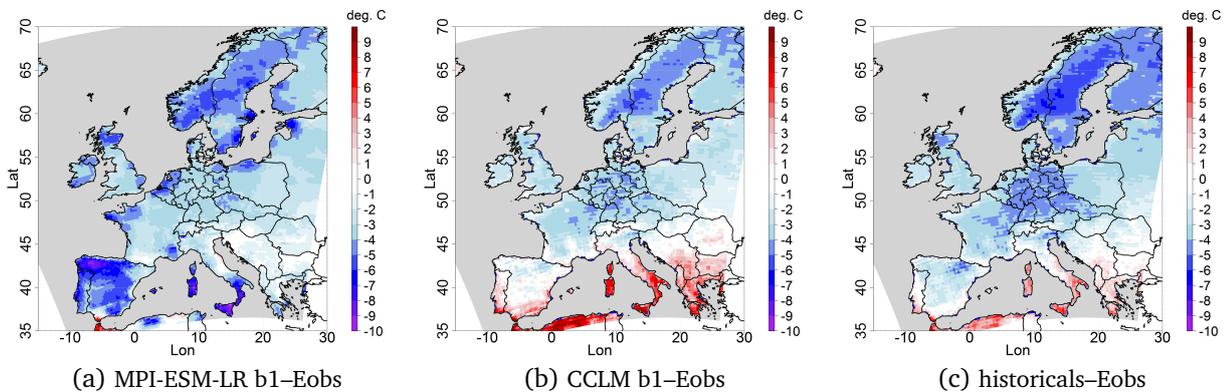
Figure 4.19.: Differences of the 90ths percentile for the daily maximum temperature of the summer half years 1981-2010 LT2-5 of the MPI-ESM-LR b1 (left), the CCLM b1 (middle) and the historicals (right) to Eobs V.14.

Reasons for this very different behaviour concerning the extremes are likely 1) the resolution and/or 2) the land-atmosphere feedback. This thesis shows that very high temperatures can only be achieved with the CCLM by increasing the resolution (Fig. 5.2b). An improvement of the model performance for simulations with higher resolutions is also known for global models from different studies. There, the enhancement of skill could be traced back to a better representation of circulation pattern like the QBO (Hertwig et al., 2015) or blockings (Müller et al., 2018). The second important point is the impact of the LSMs (see chapter 3 Vautard et al.,

2013; Weisheimer et al., 2011; Miralles et al., 2018). As Weisheimer et al. (2011) stated, the simulation of the mega-heatwave 2003 was only possible for the European Centre for Medium-Range Weather Forecasts (ECMWF) forecasting system by improving the used LSM H-TESSEL (Balsamo et al., 2009). This highlights the crucial role of sophisticated LSMs for the development of heat waves and droughts. Thus, depending on the land-atmosphere coupling developed by either LSM, JS-BACH in the GCM or TERRA-ML in the RCM, temperature extremes can be slightly under/overestimated or might seem realistic because of the wrong reasons (Vautard et al., 2013).

4.4.2.2. HW within decadal predictions

In the following section, five earlier defined HW indices are evaluated for the hindcasts. As in subsection 4.4.1.3, the HW temperatures and lengths (mean/max.) are calculated for the evaluation time and the selected region. The differences for the maximum HW temperatures and lengths as well as the numbers of HW per year are shown. In addition, correlation and RAV are displayed. The RAV is calculated by subtracting the correlation to the observation data of CCLM b1 and historicals to show the added value of the initialisation. Additionally, this score highlights a possible added value of the regionalisation, if the RCM is tested against the GCM.

The upper row (Fig. 4.20a to Fig. 4.20a) displays the temperature bias known from the 90th percentile. Likewise, spatial bias distributions are noticed in the maximum HW temperatures with an obvious cold bias in the global model, whereas the RCM tends to be too warm in Southern Europe. These bias differences are visible in Fig. A.6, in particular, the lower deviation of CCLM b1 in the average. The overestimation in Southern Europe balances the underestimated temperatures in Northern Europe in the regional hindcasts efficiently, resulting in the best agreement with the observations (+0.08 °C). On the contrary, MPI-ESM-LR b1 underestimates the maximum HW temperatures in general, leading to an averaged difference of -3.4 °C. The correlation of the hindcasts and the historicals to the observations are presented in the second row. Almost the same magnitudes and patterns are found for all models ranging from -0.5 to 0.7 with the lowest and highest values in the historical data set. CCLM b1 shows significantly high correlations south of 50 °North, whereas significant positive values are found all over Europe in the historicals. The RAV (Fig. 4.20g and Fig. 4.20h) resembles the very heterogeneous differences of the correlation fields in particular between CCLM b1 and the historicals. While reddish areas often indicate a weak added value against either MPI-ESM-LR b1 or the historicals ($>$ +40 %), Northern France and Southern Finland show a noticeably negative RAV ($<$ -60 %).

The HW lengths across Europe vary from 5 days up to 13 days in the Eobs data for the 46-year averaging period. Fig. 4.21a and Fig. 4.21c show the difference of the maximum HW length of the regional simulations to the observations. They reveal striking similarities of the CCLM runs for the spatial pattern and the size of the over- and underestimation of the HW lengths.

(a) Diff CCLM b1–Eobs   (b) Diff MPI-ESM-LR b1–Eobs   (c) Diff hist–Eobs

(d) Cor CCLM b1–Eobs   (e) Cor MPI-ESM-LR b1–Eobs   (f) Cor hist–Eobs

(g) RAV CCLM b1–MPI-ESM-LR b1   (h) RAV CCLM b1–hist
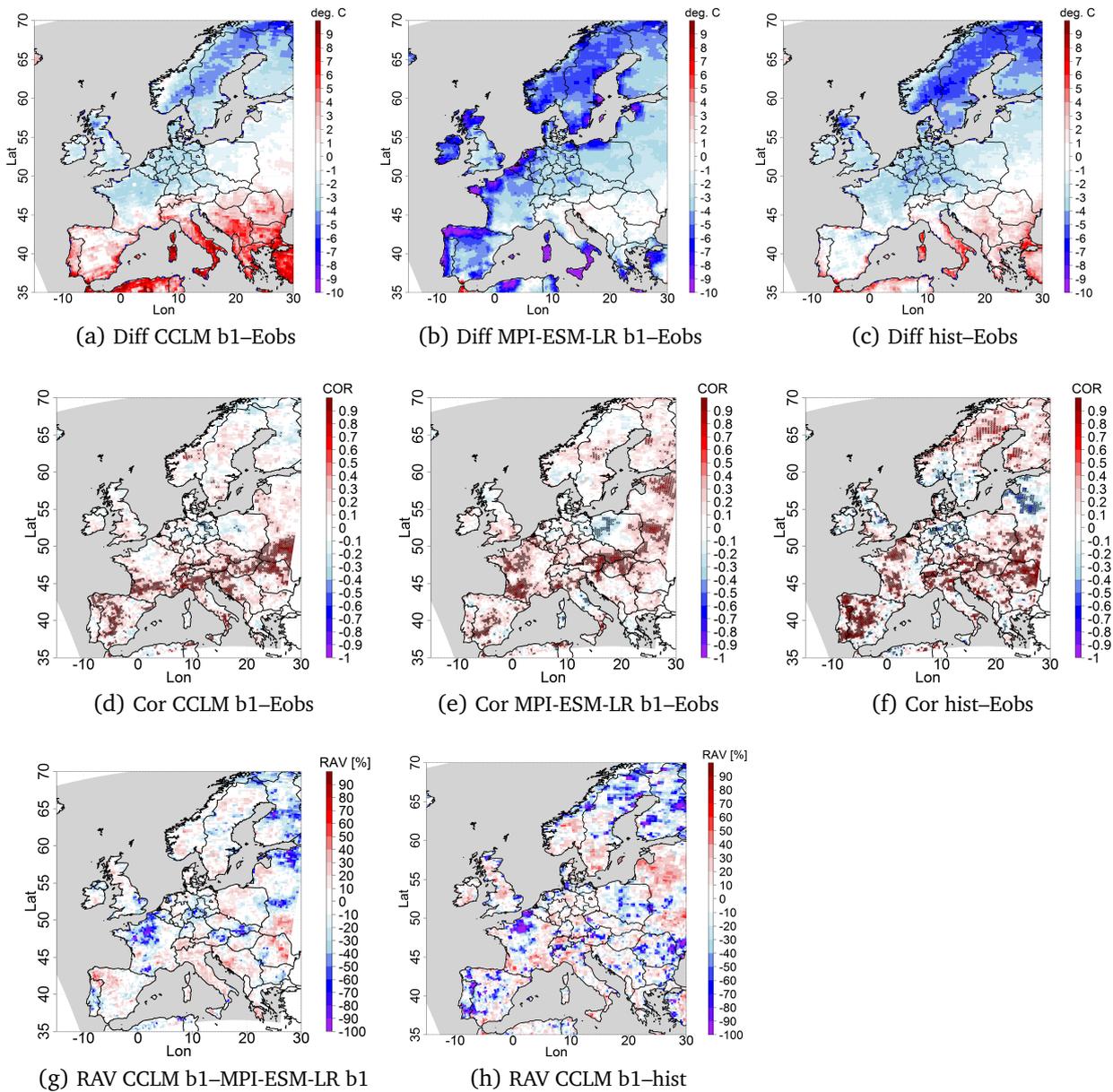
Figure 4.20.: Maximum HW temperature differences of the CCLM b1, MPI-ESM-LR b1 and hist to Eobs for 1962-2010 LT2-5 (upper row), correlation of the simulations to Eobs (middle row) RAV (after equ. 4.5) of CCLM b1 to MPI-ESM-LR b1 to Eobs (lower row, left) and RAV of CCLM b1 to historicals to Eobs (lower row, right). Grid boxes with significant correlations are marked with x in the graphics.

Both RCM generated data sets have an averaged length of 7.6 days (CCLM b1) and 7.53 days (historicals) and are in good agreement with the Eobs V.14 (7.43 days). However, the MPI-ESM-LR b1 overestimates the Eobs data by +0.84 days and the spatial differences show overestimations of over four days of the HW length in Scandinavia and Southern Europe (Fig. 4.21b).

(a) Diff CCLM b1–Eobs   (b) Diff MPI-ESM-LR b1–Eobs   (c) Diff hist–Eobs

(d) RAV CCLM b1–MPI-ESM-LR b1   (e) RAV CCLM b1–hist
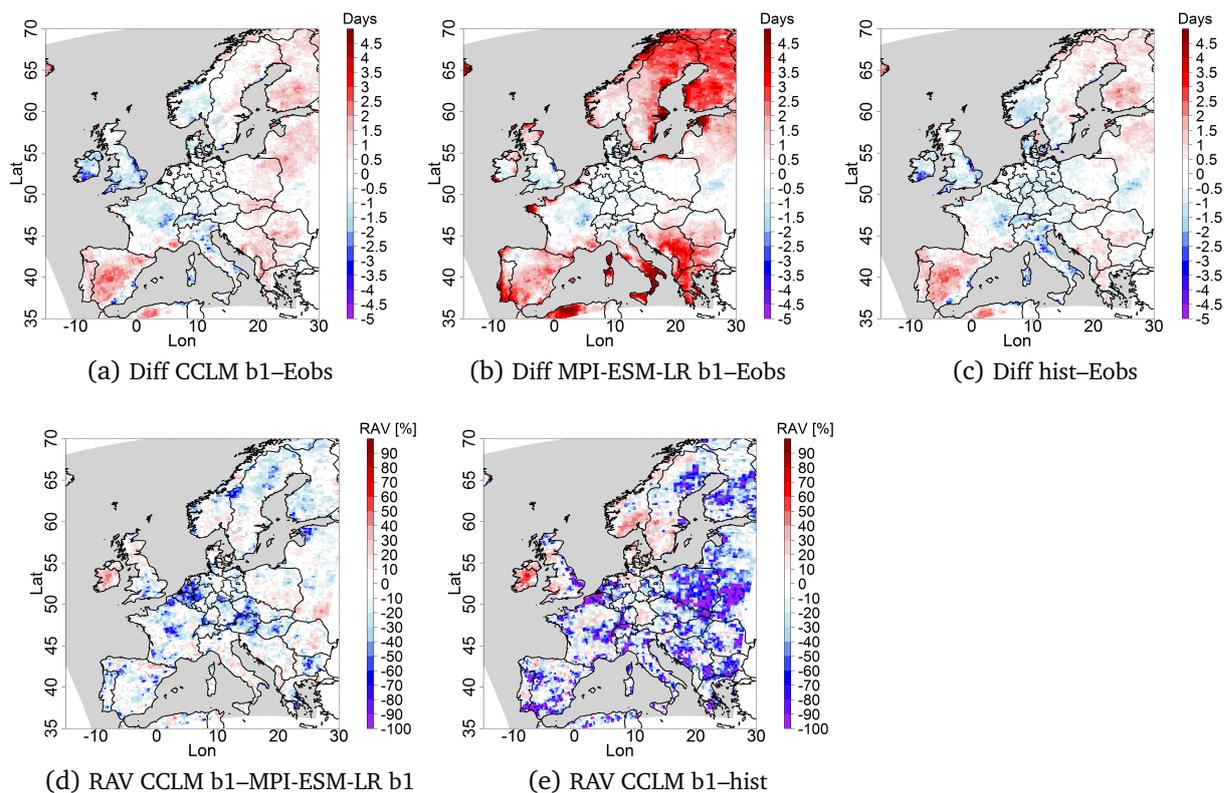
Figure 4.21.: Maximum HW length differences of the CCLM b1, MPI-ESM-LR b1 and hist to Eobs for 1962-2010 LT2-5 (upper row), RAV (after equ. 4.5) of CCLM b1 to MPI-ESM-LR b1 to Eobs (lower row, left) and RAV of CCLM b1 to historicals to Eobs (lower row, right).

Additionally, the coastlines show inaccuracies due to the interpolation to the finer resolved RCM grid. The correlation patterns are very similar for CCLM b1 and MPI-ESM-LR b1 with weak to moderate positive correlations (r: 0 - 0.5) for all Europe except Eastern Europe and Poland (therefore not shown). Nevertheless, the correlation of MPI-ESM-LR b1 to Eobs is higher in Central Europe (+0.22). Therefore, the RAV is often negative (Fig. 4.21d) and gives no hint of an added value. Notably higher correlations are found in the historicals (+0.31) compared to CCLM b1, which leads to a generally negative RAV except for Ireland and Southern Scandinavia. The RAV often reaches values < -60 % in Fig. 4.21e, indicating, that the initialization of the hindcasts is no asset in this case.

The number of HW per year in the Eobs V.14 is 1.45 for Europe. MPI-ESM-LR b1 overestimates them by +0.14 HWs per year, which is similar in CCLM b1 (+0.1 HWs per year). On the other hand, the historicals show a marginal smaller number with 1.47 HWs per year. In general, the spatial distribution is very heterogeneous as seen before. The number of HWs per year has a smaller SD of the model simulations compared to the observations as for the other variables before. This is probably the result of the averaging procedure (see Fig. 4.22). In addition, the smallest trend is found for the number of HW per year for Europe. The RAV graphics are very heterogeneous as seen for the other variables and therefore not shown. Only the Alps, BI and

parts of Eastern Europe show weak added values due to the regionalisation, whereas Northern France and Lithuania have noticeable higher correlations for MPI-ESM-LR b1. A slight advantage of the initialisations is found for BI, SC and IP, while Central and Southeastern Europe display obvious negative RAV values.
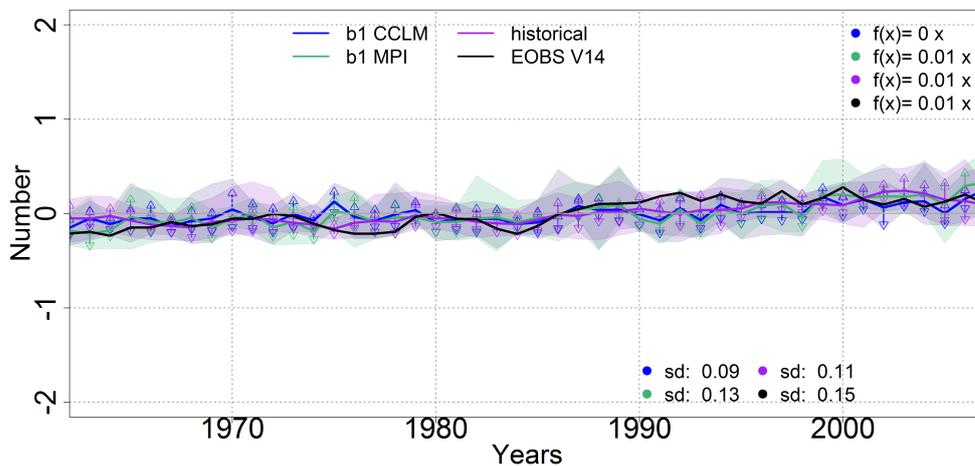
Figure 4.22.: 1962-2010 LT2-5 Number of HWs per year anomalies for Europe for the model simulation and the Eobs V.14 data (upper row). Also displayed is the interquartile range for the model simulations.

This analysis highlights many similarities between the data sets. However, it seems, that the lower bias is the only advantage of the RCM. Kruschke et al. (2016) stated, that most of the skill in decadal hindcast is derived from the long-term trend when analysing winter storms. The results confirm that this is true for HWs as well. The correlation values of the historicals, which also include the external forcing like greenhouse gases, are nearly identical or better than the hindcasts ones. One other reason for the strong resemblance between the data sets can be the averaging procedure applied in this analysis, as stated earlier. The three averaging steps (averaging over larger regions, the ensemble mean and the lead-time mean) nearly completely diminish the internal dynamic of the data sets for the temperatures (see Fig. A.6 in the appendix). Comparisons for the smaller PRUDENCE regions showed that the internal variance is much higher in these regions and the effect of the starting years is slightly visible. Still, the temporal standard deviation of the ensemble mean is much lower for the model data compared to the observations (marked by sd in Fig. A.6). The calculation of the ensemble means and the lead-times further flattens the time series.

Evaluation and preparation of an event set like in Fig. 4.8b seems not possible with the usual procedure, because of the averaging of the lead years and the ensemble members. Although, the evaluation for multi-year means is the state-of-the-art approach for decadal hindcasts (Goddard et al., 2013; Boer et al., 2016), this method is unsuitable for evaluating event-based variables like HWs in hindcasts. Hence, different ways have to be established to assess ensemble hindcasts on the European scale. A possible approach can be the probabilistic analysis presented in Kruschke et al. (2016). They obtained skill for winter storm frequencies in the North Atlantic

by separating the forecasts into three categories. A similar attempt is tested in this thesis to evaluate the HW events using daily data.

In the first step, the HW events per lead-time are calculated for each ensemble member separately. The following variables are stored for each event: Maximum temperature, intensity, duration, mean affected area, time of occurrence and the number of events per lead-time. Afterwards, the results of the hindcasts and historicals are compared to the observations and the significance is determined via the Wilcoxon-Mann-Witney-Test (also named rank sum test, in the following Wilcoxon-Rank-Test) (Wilks, 2011). If a certain level of confidence is achieved (in this case 80 %), a hit is counted for the lead-time and ensemble member. The fractional parts of the integrated hits sums are calculated including all events, starting years and ensemble members. For the variable time, a hit is counted, if a tolerance range of ± 1 year is not exceeded between events in the model data and the observations. If e.g. a HW occurs in the observations in the year 2003 and the model simulates the HW in the year 2002, it is noted as a hit. This relatively large tolerance range is chosen because of the earlier results for the yearly averaged values.
The result of this method is displayed in Fig. 4.23 for Europe. While the number of HWs is reproduced by the simulations with a hit fraction of only < 80 % (within the tolerance), the other variables range from nearly 80 to > 90 %. The historicals show quite good results with the highest hit fraction in the categories duration, time and number. While CCLM b1 has a higher agreement for maximum temperature (88 %) and scores similar to the historicals in the category area (91 %), MPI-ESM-LR b1 surpassed the other simulations only for the intensity ($+10$ %). To summarise the results of this assessment, the hit rate of each simulation for the different PRUDENCE regions and the two bigger European domains (ALL EU (1) and EU_c (10)) are shown in Fig. 4.24.

A systematic better representation of the maximum HW temperatures is found for CCLM b1 simulations for most PRUDENCE regions (Fig. 4.24). The absolute temperatures are notably improved because of the downscaling with the RCM and in particular areas like the Alps benefit from the increased resolution and the better representation of the orography (hits in duration and time). MD and IP show the temperature overestimation of CCLM b1 and ALL EU is negatively affected by the warm bias. The number of HWs is better represented by MPI-ESM-LR b1, the historical show a higher agreement with the observations for time, area and intensity. Since the intensity is not bias-dependent, the cold bias of the global model poses no disadvantage in this category in contrast to the maximum HW temperatures. The area index is disputable as no coherent areas are evaluated. In this analysis, every grid point is counted including scattered grid boxes. Therefore, this variable is difficult to rate in comparison to the other categories. In addition, one has to keep in mind, that the duration is crucial for calculating correct global intensity values.
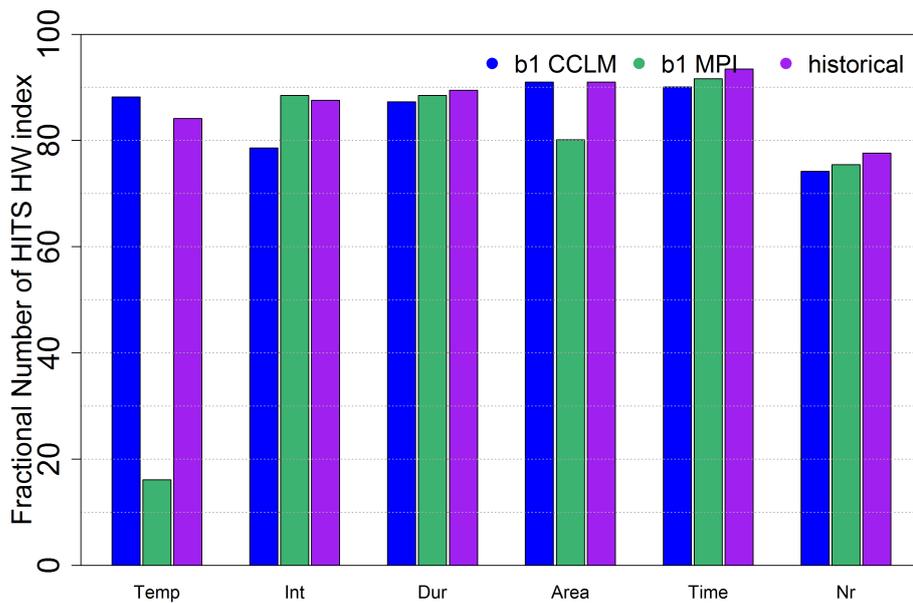
Figure 4.23.: 1962-2010 LT2-5 Maximum HW temperatures, HW intensity, duration, averaged area, time and number of events evaluated via Wilcoxon rank sum test with an 80 % confidence interval for Europe of the model simulation to Eobs V.14 data

If the duration deviates strongly, the intensity will possibly be wrong, even when the difference of HW temperature to 90th percentile was properly captured or vice-versa. In general, Fig. 4.23 shows a higher number of hits for the downscaled simulations, but no benefit of the yearly initialisations. Another issue of this evaluation is the input for the significance test. The lead-time based evaluation nearly always produces a time series of different lengths between the observations and the simulations. If, for example, the Eobs data contain a vector with ten HWs in one lead-time, while CCLM b1 and the historicals show more and MPI-ESM-LR b1 less than ten, the time series are cropped to the shortest mutual duration of all simulations and the observation. These cropped time series are used then to calculate the Wilcoxon-Rank-Test. This approach accepts that each lead-time and realisation is evaluated for varying lengths and neglects events exceeding the mutual time. On the other hand, the results show the best possible outcome of the comparison and give an estimation of the quality of the prediction on their most basic level. In particular, extreme events can have a large impact on the yearly averages presented earlier in this chapter and are ignored because of the respective lengths. For example, MPI-ESM-LR b1 seen in Fig 4.24 captures the duration best in Scandinavia, although the yearly averaged maximum HW lengths are overestimated by the GCM (Fig. 4.21b).

Thus, both analysis methods have some (dis)advantages. The bias is more realistically estimated in the yearly averages including all events, whereas single HW events can be evaluated with the just presented method without blurring them too much.
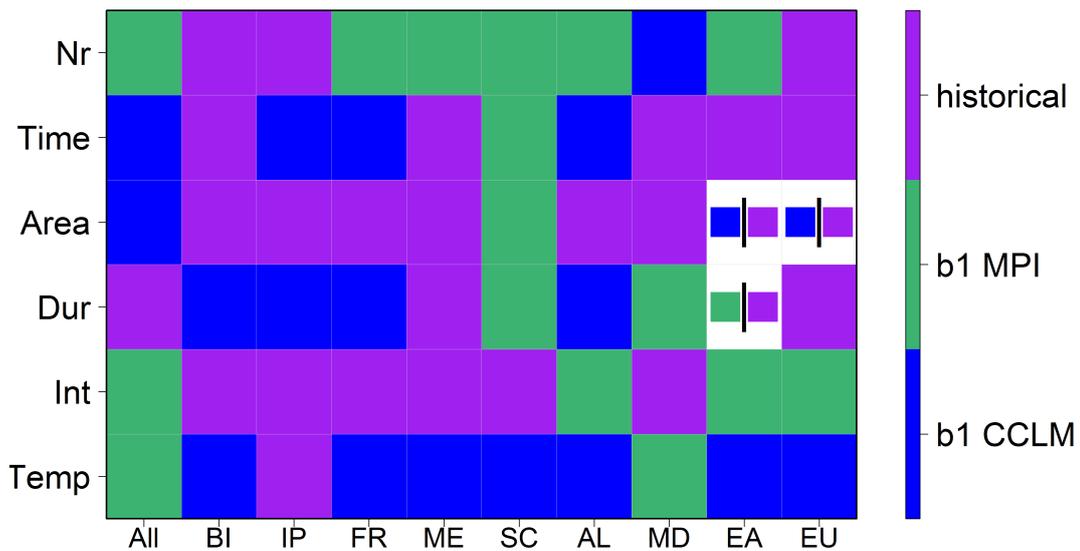
Figure 4.24.: 1962-2010 LT2-5 Maximum HW temperatures, HW intensity, duration, averaged area, time and number of events hit rates for Europe of the model simulation to Eobs V.14 data based on Fig. 4.23. In unequivocal boxes, both simulation colours are shown.

However, using a statistical test which requires input vectors of the same lengths is difficult at most. The cropping of the surplus values compresses the distribution and alters the mean values. One option to overcome this problem might be to use fitting functions for the event distributions and compare these instead of the actual events. In future tests, a smoothing spline function could be applied to generate surrogate values fitting model to observations. The events are very anomalously distributed, which leads to significant deviations between the actual values and the fits. Thus, a new approach is needed for this evaluation to counterbalance the unequal event numbers.

4.4.2.3. Teleconnections to large scale atmospheric circulation pattern

The ability of the CCLM EVA runs to simulate the connections between the HW days and the LCA pattern in most cases (see subsection 4.4.1.5) leads to the following two assumptions. In cases of a correct reproduction of the LCA pattern by the GCM, some predictability, originating from the LCA pattern, can be expected for the simulation of HWs. If these patterns can be passed to the RCM, some added value in the simulations of HW days should be found because of the dynamical downscaling. The skill of the RCM simulations depends on the correct representation of patterns in the global model. They are not able to generate these patterns themselves, because of the different spatial scales. The large-scale patterns can encompass at least half of the RCM simulation domain and/or the centres of action often are outside the regional domain.

Various teleconnection indices and their predictive skill can be calculated with the VADYtele plugin by Benjamin Lang of the standardised data and evaluation system framework (MiKlip CES) (Lang, 2015). The plugin uses the RPCA method of Barnston and Livezey (1987) to compute

the indices either from MiKlip GCM products or observations. In addition, the plugin includes an analysis part to assess, e.g. the correlation for different lead-times and skill scores like the MSESS. The correlation of the LCA indices is displayed in table 4.3 derived from the MPI-ESM-LR b1 to the ones of the ERA-Interim data. Additionally, the historical GCM simulations are included for the lead year one.

Correlation	EA	EAWR	NAO-HPC	NAO-HST	SCAND	POL
lead year 1						
historicals GCM	0.572	0.117	-0.274	0.062	0.066	0.381
b1 GCM	0.641	0.189	-0.139	-0.122	-0.227	0.122
lead year 2						
b1 GCM	0.573	0.122	-0.105	-0.3	-0.036	0.131
lead year 2-5						
b1 GCM	0.565	0.058	-0.159	-0.107	-0.131	0.033
lead year 6-9						
b1 GCM	0.598	-0.102	0.015	0.001	-0.050	0.034

Table 4.3.: Correlation of the LCA index of b1 GCM data to ERA-Interim observations for the JJA of different lead years of the LCA pattern derived from VADYtele.

The pattern with the highest correlation between ERA-Interim observations and MPI-ESM-LR b1 data is the EA with r > 0.5. The other patterns show only weak correlations to the observations. The correlation is around 0 for most patterns and longer time intervals in table 4.3. The hindcasts with annual starting dates produce higher pos./neg. correlations for most of the patterns, while the historical GCM simulations show higher agreement with the observations only for the POL with r=0.381. The higher negative correlations indicate a systematical failure to reproduce the LCA index. Two possible causes are presented, but the exact reasons are not examined further in this study.

Independent of the simulation strategy (initialised or uninitialized), the global model does not seem to contain LCAs fitting the observations, which can be due to the model is not able to a) reproduce the pattern at all or b) to simulate patterns in phase with the observations. These problems (a and b) were noticed by Scaife et al. (2010), who investigated conditions associated with the NAO or blocking. One option enhancing the quality of the MPI-ESM-LR b1 simulations was the increase of the resolution, which lead to an improvement for blocking frequency and location (Müller et al., 2018). HWs associated with blocking would be possibly better represented by the model system as well.

The only pattern well reproduced by the GCM is the EA pattern for which moderate to high correlations were found independent of the lead-times. Barnston and Livezey (1987) described the EA as the second leading mode in the North Atlantic. The northern centre of action of the EA pattern is found in the Northern/Northeastern Atlantic near 55°N and 20-35 °W in the sea level

pressure maps, the southern centre near 25-35°N and 0-10 °W. Three possible reasons, why this pattern shows noteworthy correlation values are given in the following.

1. Possible impact of the climate trend on EA: The index shows three long time intervals (> 10 years) in the evaluation period, two with a dominating cold/warm phase and one transition phase. Therefore, the EA index has lower decadal variations than every other index in this study. The EA standardised index time series indicates a shift from a mainly negative phase before 1980 to a mainly positive phase afterwards (NOAA / National Weather Service Climate Prediction Center, 2018a). The positive trend in the standardised EA index is likely connected to the climate trend of the North Atlantic SST (increase of the SSTs, (Knudsen et al., 2011, Fig. 1b) and Rayner et al. (2003)) and the rising temperatures over Northern Africa. The North Atlantic is well represented in the model and major source of skill for the temperatures (Marotzke et al., 2016). This skill might be partly found in the geopotential and as such in the EA index as well.

2. Similar patterns are captured well by the model: Cattiaux et al. (2013) analysed a pattern similar to the EA described as the Atlantic Low (Fig. 6). The comparison for models participating in the Coupled Model Intercomparison Project Phase 5 (CMIP5) revealed, that MPI-ESM-LR b1 reproduced the frequency of the Atlantic Low with only a low bias. However, the Atlantic Low in Cattiaux et al. (2013) misses the Southern Centre, which is essential in the EA. Nevertheless, the dominant low west of Britain exists in both pattern and is well reproduced by the GCM as well as the EA probably.

3. Good representation of long-term circulation patterns: Noticeable warming in the SSTs of the North Atlantic occurred in the early 1990s signalling the beginning of the next AMO warm phase (Sutton and Dong, 2012). The positive phase of the AMO in recent years further increases the SSTs of the North Atlantic (warming due to the multidecadal variability ~60 % since 1970) (Sutton and Hodson, 2005; Polyakov et al., 2010). Ruprich-Robert and Cassou (2015) found a time lag of EA-conditions to the next AMO+ phase of -30 to +10 years in a 1000-year climate simulation. This period mostly fits the phase change of the EA pattern and the recent AMO phase change. Furthermore, Polyakov et al. (2010) performed an empirical orthogonal function analysis and showed that its second component (representing the AMO) is associated with sea level anomalies of the EA pattern. MPI-ESM-LR b1 reproduced the AMO structure as second best of the participating models as known from a study with models participating in CMIP5 (Ruiz-Barradas et al., 2013). Thus, if the model simulates this associated structure, it is likely, that the pattern powering it is captured as well.

The reasons for the high correlation of the EA pattern remain unclear as all the named points are part of ongoing research questions. However, long-term stable phases seem to be well captured by MPI-ESM-LR b1 as the high correlation shows.

In the following, the MSESS of the global model hindcasts compared to the historical simulations to observations is displayed for the six patterns. No skill values exceeding $\pm$ 0.4 are found for

the selected LCAs in the global b1 hindcasts compared to the historicals (Fig. 4.25b – Fig. 4.25f). Small positive skill is found for the MPI-ESM-LR b1 for the EA pattern during the first year (Fig. 4.25a). The NAO pattern shows skill with contrary prefixes either calculated from a principal component analysis (Fig. 4.25d) or from station data (Fig. 4.25e). The other three patterns (EAWR, SCAND and POL) show no advantage of the starting years. On the contrary, negative skill values are found for all indices especially pronounced in POL (Fig. 4.25f).
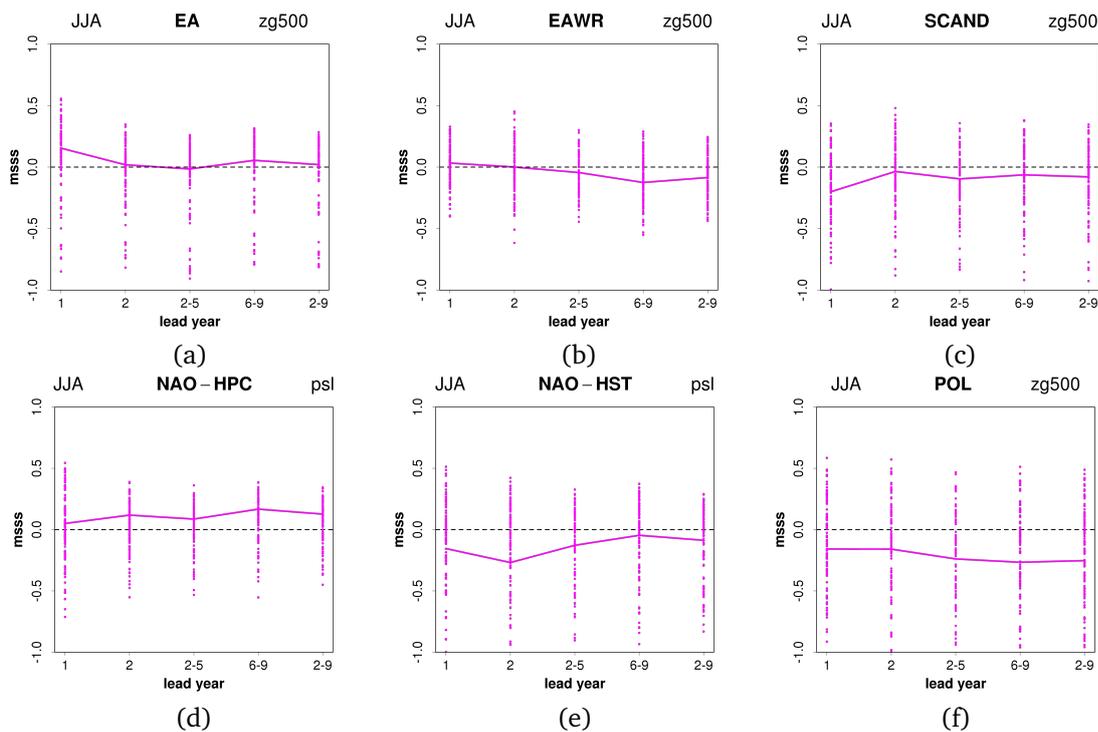
Figure 4.25.: MSESS of the GCM b1 hindcasts to the GCM historicals to observations for different lead years.

4.4.2.4. User-relevant decadal climate information

The previous results made clear that it is not possible to predict single events exactly, including their intensity and extension, with the MiKlip hindcasts. The number of HW events per lead-time interval showed some skill as well as yearly HW indices. A new integrated variable counting the days per year exceeding the 90ths percentile is established to estimate the magnitude of heat stress per year. The HWDS (sum of heat wave days per year) is implemented into the MURCSS evaluation system and results show high correlations similar to the daily maximum temperatures for LT2-5 (subsection 4.3.2). Additionally, a new recalibration tool from the Freie Universität Berlin is tested for this variable providing the user with the option to re-calibrate the bias and perform a drift correction (Pasternack, 2018; Pasternack et al., 2018).

The MSESS, the correlation and the conditional bias are displayed in Fig. 4.26a to Fig. 4.26d for the uncalibrated and calibrated CCLM b1 HWDS calculated with the MurCSS plugin.

(a) Con. Bias b1

(b) Con. Bias b1 cal

(c) COR b1

(d) COR b1 cal

(e) MSESS b1 cal to b1 LT1-1

(f) MSESS b1 cal to b1 LT2-5
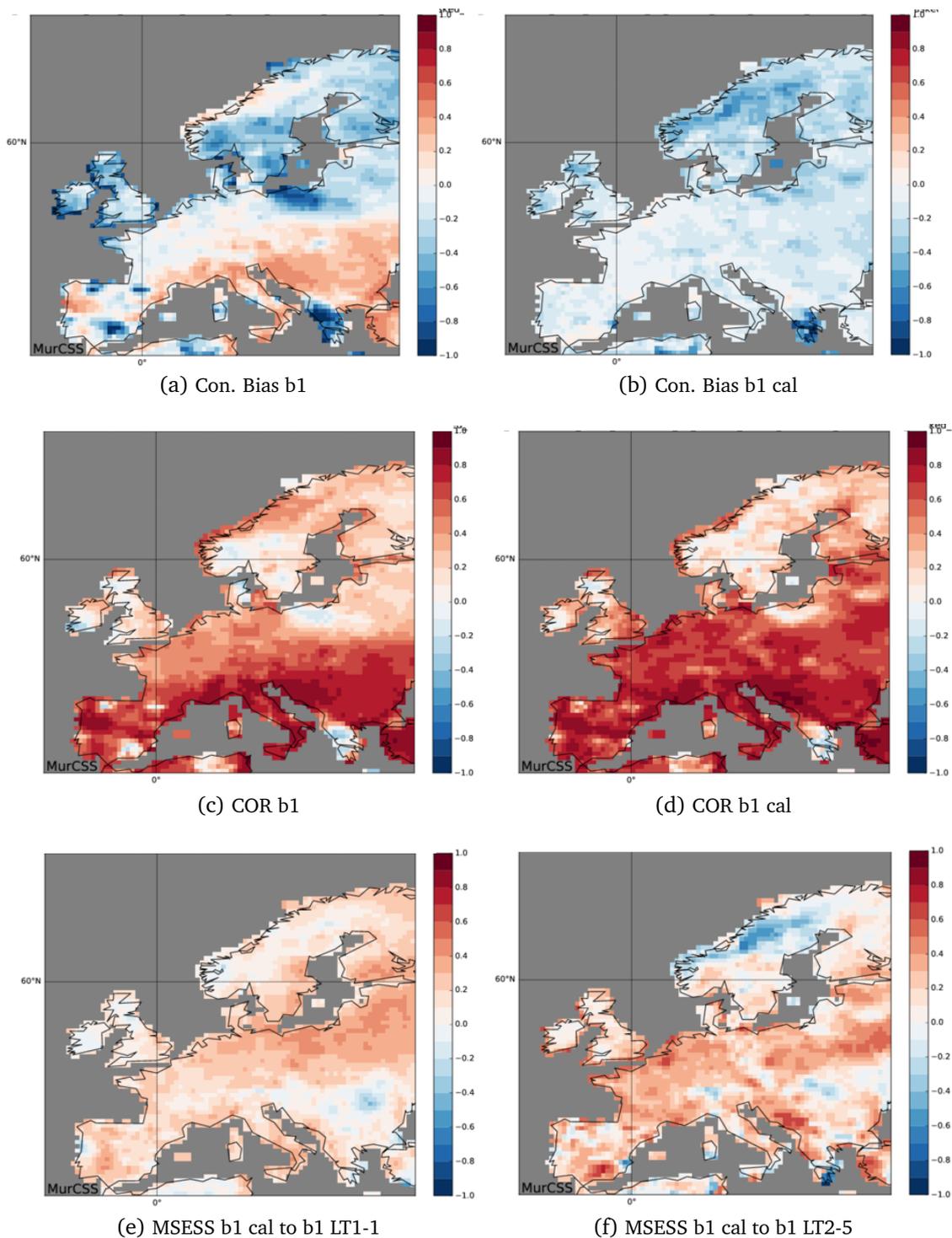
Figure 4.26.: MURCSS results LT2-5 for the HWDS: Conditional bias (upper row) of regional hindcasts b1 to Eobs V.14 for the HWDS for LT2-5 (left) and calibrated b1 (right), same constellation in the middle row for correlation and the MSESS (lower row) of the calibrated regional hindcasts to uncalibrated b1 to observations for LT1-1 (left) and LT2-5(right).

The calibration reduces visibly the positive conditional bias for Southern Europe (Fig. 4.26a compared to Fig. 4.26b). The best agreement to the observations is detected for values around

0 for the conditional bias unlike the MSESS and the correlation. However, the averaged conditional bias is higher for the calibrated (-0.18) than in the uncalibrated hindcasts (-0.1). On the other hand, the spatial spread is largely improved for the calibrated (b1: -1.38 to 0.54; b1 cal: -0.95 to 0.13). The largest improvement due to the recalibration is found for the correlation in Northern Central and Northeastern Europe. Weak negative correlation values, seen in Fig. 4.26c, vanish in the calibrated data (Fig. 4.26d) and higher correlations are found throughout Europe (b1: 0.42; b1 cal: 0.52). Fig. 4.26e reveals a drift and bias correction for lead year 1 improving the negative skill values in Eastern Europe. It remains unclear if this area is affected by an incorrect initialisation. Soil conditions strongly deviating from the balanced state of the land-surface scheme TERRA_ML in CCLM could lead to the negative MSESS in LT1 as seen in (Fig. 4.4c). The effects of unbalanced soil initial values were investigated in chapter 3 and timescales of 1 until five years match the results of the actual analysis. Although, the calibration seems to improve the skill over large areas in Europe, Fig. 4.26f shows that a small skill reduction occurs in Scandinavia. The conditional bias becomes slightly larger and the correlation is reduced in this area during the recalibration for yet unknown reasons. Nevertheless, the HW days sum per year seems to be a successfully predictable variable and easily communicatable to potential users due to its simplicity.

Short summary

Daily maximum temperature:

The summer daily maximum temperatures of the EVA simulations agree well with the observations. Temperatures are overestimated in Southern Europe and mostly underestimated in the North by all models, revealing a North-South gradient. The older EVA simulation overestimates temperatures stronger in Southern Europe leading to a slightly higher mean bias ($+0.41°C$). For longer lead-times, the decadal predictions show high correlations for the SHY daily maximum temperatures for most of Europe ($r > 0.5$). The higher lead-times are unaffected by the problematic drift recognised in the first year in Eastern Europe. Thus, longer lead-times excluding the first year seem to be more meaningful.

Heat Waves:

In the Eobs data, the 90th percentile of the daily maximum temperatures ranges from over $35°C$ in Southern Europe to below $15 °C$ in coastal regions in Scandinavia. The reanalysis driven simulations show a warm bias in Southern Europa and a cold one in Scandinavia compared to Eobs in the 90th percentile. Maximum HW temperatures of $> 38°C$ are found in Southern Europe in Eobs, whereas parts of Northern Europe show temperatures below $20.1°C$. The bias pattern known from the 90th percentile is also noticeable in the mean and maximum HW temperature. 1-2 HW per year with durations from 6 days up to 14 days in the annual averages are found in the observations. The EVA runs mostly overestimate the lengths and numbers per year.

Six mega-Heatwaves occurred from 1961-2010 in the larger European simulation domain with up to 50 % of the European evaluation domain covered by events in 1994, 2003 or 2010. A stronger climate trend in HW temperatures and the annual number of HWs was discovered in Southern Europe, while the trends were negligible small in the Northern regions. A decadal signal is detected aligning the HW temperatures with the AMO index. Connections of HWs and selected LCAs could be determined within the EOBS data; however, the correlation patterns are very heterogeneous in space and strength. Areas with significant correlations are restricted to certain regions frequently associated with the centres of action of the LCA or in the downstream of their wave pattern. For example, the EA is a dipole with a subtropical ridge and a corresponding low in the North Atlantic. HWs are positively correlated in Southeastern to Northeastern Europe due to the elongating subtropical ridge, while the British Isles show negative correlations influenced by the Atlantic low. Correlations between the negative and positive phase of selected LCA to monthly HW days are derived. The magnitude of the correlation signal between the phases is not symmetric and dependents on the PRUDENCE region. It was noted, that correlations often are considerably higher when HWs are dampened or suppressed in colder years. In general, the spatial distribution and the phase dependent signals are well captured by the reanalysis driven models. In the hindcasts, similar spatial mean and maximum HW temperature bias patterns are found as in the EVA preop simulations.

The CCLM b1 overestimates the maximum temperatures by $> 4.5°$C for parts of Southern Europe and underestimates them in Northern Europe ($<$ -2.7°C). The MPI-ESM-LR b1 underestimates the max. temperatures in nearly all Europe (the only exception is the Balkans). Moderate to good correlations of r ≤ 0.6 are noted for the max. HW temperatures in all model simulations all over Europe, although the patterns are very heterogeneous. No clear RAV of the CCLM b1 was observed neither for the regionalisation nor the yearly initialisation. The maximum HW length is slightly overestimated in the hindcasts as well as the number of HWs per year. A new approach has been tested applying a Wilcoxon-Rank-Test with an 80% confidence interval to evaluate event-based HWs in the decadal hindcasts. Nearly all HW variables are well captured (> 70 % of hits), but the yearly initialisation indicates no clear advantage against the uninitialized. Nevertheless, the regionalisation shows an added value for the HW area and the temperature. The LCAs are not adequately reproduced by the MPI-ESM-LR b1, moderate to high correlations have been noted solely for the EA pattern($r > 0.56$). A new user relevant variable, HW days per year, has been tested and skill all over Europe was identified for LT2-5. A recalibration increased this skill even further, in particular for Central and Eastern Europe (increase from 0.42 to 0.52).

4.5. Chapter discussion and conclusions

Summer temperatures and HWs in model simulations and observations were evaluated in this chapter and their potential decadal predictability assessed on a European scale. Two simulations driven by ECMWF reanalysis data were employed to investigate the connection of large-scale circulation anomalies to HWs for Europe and the PRUDENCE regions beside the b1 decadal hindcasts of the MiKlip project. The lead-time 2-5 years was chosen to investigate HWs in the decadal hindcasts based on the 90th percentile and the skill found for the summer maximum temperatures. A drift in the first year affected large areas in Eastern Europe leading to substantially reduced skill. Therefore, the longer interval is not only more adequate but in accordance with other studies e.g. Goddard et al. (2013) as well.

Events like the Mega-heatwaves 2003 and 2010, associated to thousands of premature heat-related deaths during the summertime, enhanced the awareness of stakeholders (politics, economy a.s.o.) and in the population leading to the establishment of warning systems across Europe. These systems are limited by the quality of reliable short-term forecasts. Prediction systems like the MiKlip project can help to fulfil the need for information of stakeholders to overcome problems concerning health, water resources and infrastructure on time periods up to 10 years. Coughlan de Perez et al. (2018) noted good short-term predictability of HWs over Europe and suspected the prediction time could be extended towards the sub-seasonal scale if other factors like soil moisture variability are included. However, the benefits and limitations of decadal predictions are discussed extensively within the scientific community (Meehl et al., 2009; Smith et al., 2012; Eade et al., 2014; Marotzke et al., 2016). The dynamical downscaling with RCMs is seen critically since it is often doubted that the benefits are worth the higher computational costs. Furthermore, the downscaling adds uncertainty to the data (Raäisaänen, 2007). On the other hand, clear advantages are known foremost the physical consistency of variables. In addition, not every variable requires a specific method unlike the statistical downscaling. Thus, large data sets can be created and applied for investigations like statistical analyses of insurance companies or as forcing data for models (hydrological, economic or others).

Factors as North Atlantic SST and precipitation can be potentially used as predictors with time lags up to half a year, because of their connection to European HWs (Della-Marta et al., 2007b). In particular, the slowly varying SST changes of the AMO are a source of predictability for temperature and precipitation in Europe (Sutton and Hodson, 2005; Knight et al., 2006; Müller et al., 2012; Ghosh et al., 2017). The evaluation revealed a close connection of the AMO index to HW temperatures derived from Eobs observations and EVA simulations with moderate correlations of r >0.5 for Europe. Depending on the region, periods with lower HW temperatures can be linked to the positive or negative AMO phase. A dampening of HWs temperatures e.g. is noted in the Western regions like BI or FR during the AMO+ phase, while higher temperatures are detected in the AMO- phase in the 1970th. The decrease of HWs during the positive AMO phase is most likely a result of the increased cyclonality (Sutton and Hodson, 2005) and pre-

cipitation (Knight et al., 2006). However, other areas superimpose the HW temperature signal from the westernmost regions in the European average. An anticipated decrease in the HW temperatures during the AMO- phase is found only for the decade 1971. The long-term climate trend and its temperature increase dominate the HW temperature progression, explaining why the uninitialised "historicals" are able to reproduce annual HW variables relatively well (similar to the results from Kruschke et al. (2016) for winter storms). Unlike the hindcasts, the uninitialised ensemble is not nudged to the actual climate state (ocean, soil) and the source of their skill is primarily related to an accurate representation of the mean climate state and the general climate trend. This seems to be no disadvantage in this study and for the selected variables.

Beside the AMO, HWs are connected to LCAs like the NAO, EA, SCAND and other. These teleconnections are investigated with the reanalysis forced EVA simulations and Eobs observations for Europe. The EVA runs include the most realistic forcing and are the most likely ones to capture the connection correctly. Calculations for different phases of the indices revealed phase specific as well as spatial variances with significant positive/negative correlations between monthly HW day sums and the LCA index. An unsymmetrical response of the HWs to the patterns was identified after the filtering into the index phases. Vice-versa responses to the phase changes were expected but this is not the case for most indices and PRUDENCE regions. One reason could be the size and the chosen boundaries of the PRUDENCE regions. The spatial scale of the teleconnection patterns is e.g. much larger than the typical size of a PRUDENCE region (especially smaller ones like the Alps). Therefore, the affected areas do not necessarily fit the spatial range of the LCAs. Additionally, their centres of action are not necessarily located directly within a PRUDENCE domain. This can lead to contradicting correlation signals as noticeable gradients occur within a PRUDENCE region. This can potentially diminish an otherwise clear signal in the averages. Since monthly means of HW day sums and the summer half-year are used, the correlation values are lower as in Rust et al. (2015), but a general agreement was identified for the spatial pattern response in the observations. The phase filtering proved to be an effective method which helps to understand the development of HWs in the separate phases of the LCA indices. The EVA simulations are in good agreement with the observation for most of the patterns. CCLM EVA preop achieved marginally better results using the newer model version and the higher resolution. This study showed that model simulations are able to reproduce most of the connections between HWs and LCA pattern if a realistic forcing is provided. Thus, the deviations of HWs in model simulations can be a result of the inability of the model to capture the connected LCAs correctly.

An analysis of the global hindcasts emphasised that the MPI-ESM-LR b1 failed to capture the patterns, except the EA for which a noteworthy correlation exists. This pattern has dominant long-term stable phases (> 10 years) following the long-term climate trend. This might be one reason for the higher correlation as the trend is well captured by the model (Kruschke et al., 2016). Furthermore, teleconnections to other long-term variations like the AMO and ability of

the model to simulate similar circulation pattern increased the chance of an adequate representation of EA in the GCM simulation (Cattiaux et al., 2013; Ruprich-Robert and Cassou, 2015; Polyakov et al., 2010). Other patterns like NAO, SCAND or POL often associated with blocking show no skill at all. The reasons for this, either a mistiming of the phase, the spatial extent or the ability to simulate the pattern at all, is not investigated here.

Although a certain skill has been identified for long-term averages, the misrepresentation of the LCA can be one of the reasons why nearly no added value is found in the hindcasts with yearly initialization. In addition, large problems to simulate circulation anomalies like blockings are known for the GCMs from other studies. Frequently, LCA patterns are simulated by the global models, but differences in spatial extent and variability are noticed (Masato et al., 2013; Scaife et al., 2010; Wójcik, 2014; Eade et al., 2014). The development of temperature extremes is associated with blocking over the whole year, and their frequency is often underestimated by GCMs (Buehler et al., 2011; Masato et al., 2013). The GCM MPI-ESM-LR b1 used in the MiKlip project underestimates the blocking days largely (Müller et al., 2018). An improvement for the simulation of blocking situations and alike weather regimes has been found for GCMs, including the MPI-ESM-HR, with an increased resolution (Berckmans et al., 2013; Dawson and Palmer, 2015; Müller et al., 2018). The newer generations of the MiKlip system dcppA and preop are computed using the MPI-ESM-HR but were not completed before this thesis was written.

A new approach was tested to evaluate HWs on an event basis. Using the Wilcoxon-Rank-Test, the HW variables derived from all model data are compared to the HWs from the observations tested for significant differences with an 80 % confidence interval. The CCLM runs showed a higher agreement to the observations for the maximum HW temperatures, the area, the intensity and the timing. The more explicitly adjusted orography in regions like the Alps is clearly one of the advantages of the RCM and leads to compensation of the cold bias of the GCM (Fig. 4.24). Thus, an added value of the regionalization can be demonstrated. MPI-ESM-LR b1 and the uninitialized (historicals) seem to have a higher agreement with the Eobs data for mean HW lengths on event-basis and the number of HW. This result is partly in conflict with the outcome of the evaluation of the annual averages. There, the MPI-ESM-LR b1 overestimated both variables in Northern Europe (length: +2 days and number: +0.13 HWs per year), while in the event-based evaluation the GCM has similar hit rates as the uninitialized.

No unequivocal proof of an added value of the yearly initializations was detected because of the large hit number of the uninitialized. One critical issue is the small sample sizes for the estimation of significance. Furthermore, the different number of events of simulations and observations is problematic as a large number of events are not considered. Thus, additional investigations are needed to test the robustness of the results.

Possible end users of climate data are often interested in easily understandable, useful products with a high spatial resolution. However, usefulness and usability depend on either producer or user (Lemos et al., 2012). Variables like the 2m temperature and precipitation are often

investigated, nevertheless, parameter interacting stronger with the society or economy are more relevant for stakeholders (Hackenbruch et al., 2017; Lemos et al., 2012). Climate variables as defined by Zhang et al. (2011) or even more specific ones are often asked for e.g. the growing season lengths, which is a useful indicator for the agriculture as the heating degree days are for the energy sector (Isaac and van Vuuren, 2009). Since the event-based HWs are not explicitly predictable by the model, an integrated value has been designed to fulfil possible user needs about HW information. HW days per year (HWDS) are easy to communicate and understand. Additionally, HWDS are predictable with good skill by the hindcasts with a similar spatial distribution as the ones derived from the daily maximum temperatures. The application of a recalibration tool increased the skill values in most of the European regions in particular in Eastern Europe in LT1 (only exception: SC) by improving the conditional bias and correlation.

5. Heat waves in very high resolution projections and their climate change signal

5.1. Introduction

Temperature extremes like heat waves (HW) have become more intense and frequent in the last century (Della-Marta et al., 2007a; Russo et al., 2015). 76 % of the weather-related disaster deaths from 1995-2015 are related to extreme temperature events (mainly HWs) for high-income countries (CRED et al., 2015). Mortality rates increase strongly during HW days affecting foremost vulnerable population groups like elderly, children and outdoor working people Muthers et al. (2017). Thus, predictions of severe HW like 2003 are of particular interest for adaptation and emergency planning. Most of the planning is conducted for small-scale areas with the size of districts and high spatial resolution. The resolution of climate predictions and projections has increased over the lasts years continuously to fulfil the need of potential users of model data. In this chapter, a newly built ensemble with a resolution of $0.025°$ (~2.8 km) is evaluated which is currently one of the largest ensembles of high-resolution climate projections at least for Europe. The added value compared to the next coarser nest is assessed in this analysis. In addition, the benefit of the ensemble approach will be highlighted and systematical differences resulting from the global models are investigated. Beside averaged variables, extremes like HWs and in short heavy precipitation events will be analysed and compared to observations. The question will be tackled, how spatial and temporal pattern HWs will change in future periods for Southern Germany.

Very high resolved simulations are able to reproduce complex structures in precipitation fields far better than conventional climate simulations usually conducted at scales of 10 to 50km like in the CORDEX project Jacob et al. (2014). In particular, heavy precipitation events in complex terrain are underestimated by simulations with coarser resolution Frei et al. (2003). Hence, highly resolved and physically consistent data sets are of special interest for impact or hydrological studies conducted at scales below 1 km. However, these simulations not only require massive amounts of computational power, but scientists are confronted with new challenges (technical as well as scientific) running models at convection resolving scale. For example, Ban et al. (2014) found higher precipitation errors for mountainous regions in simulations with very high resolutions. Although, decreasing deviations were expected because of the better representation of orographic structures and external parameters like the land use classes and others in the model Knote and Heinemann (2010). On the other hand, small scale temporal processes like the daily cycle often show a higher agreement with the observations for the very high simulations [Dirmeyer et al. (2012), Kendon et al. (2012), Ban et al. (2014), Fosser et al. (2015)].

Whereas simulations with parametrised convection frequently underestimate convection intensity and include a timing error Brockhaus et al. (2008). Furthermore, previous studies indicated heterogeneous climate change signals for areas characterised by complex topography, especially for high intense precipitation events (Maraun et al., 2010; Ford et al., 2013; Kendon et al., 2012, 2017).

Fosser et al. (2017) raised the question if an actual benefit of very high resolution ($<$ 4 km) exists for climate simulations and a gain of information. Answering this question is not trivial, currently not enough studies have been conducted and experience is missing. Contrary to Fosser et al. (2017), Hackenbruch et al. (2016) identified an added value for selected variables due to a reduction of the model bias in a higher resolved simulation driven by ERA40 reanalysis data (2.8km compared to 7km). However, the bias reduction is not consistent for the various RCMs included in the EURO-CORDEX comparison Vautard et al. (2013) and depends partly on the forcing Keuler et al. (2016). Therefore, ensembles are built to overcome the disadvantages of single models and to obtain more reliable and robust projections.

5.2. Methods and Data

5.2.1. Model Setups and Nesting Strategy

The regional climate model COSMO-CLM (CCLM) (Baldauf et al., 2011; Rockel et al., 2008) is used to perform very high resolved climate simulations for Southern Germany. At the cloud-resolving scale, the non-hydrostatic model CCLM explicit calculates deep convection instead of using the parameterisation applied for coarser resolutions. Three model versions are employed (because of the time of the simulation performance): cosmo_5.0_clm3a and cosmo_5.0_clm6 for the slightly older runs, and cosmo_5.0_clm9 for the newer ones.

Three global climate models have been selected (MPI-ESM-LR, HadGEM2-ES and EC-EARTH) as forcing for the small ensemble included in the Coordinated Regional Climate Downscaling Experiment (CORDEX, (Jacob et al., 2014)). In addition, two simulations have been performed with the ERA40 Reanalysis (Dee et al., 2011) as driving data. With these runs, the impact of the lower boundary by changing the standard land surface scheme (LSM) TERRA_ML to VEG3D was assessed. Also, they provide additional validation data for the evaluation. This comparison is not part of this thesis and therefore, the ERA40 driven runs are not included in the following evaluation.

The global input has been dynamically downscaled using three nesting steps. The coarser resolution provides the input for the next finer step to achieve a final resolution of 0.025° (2.8 km). The nests are sub-divided as follows: i) global to 0.44° (50km), ii) 0.44° to 0.0625° (7km) iii) 0.0625° to 0.025° (2.8km). The first and the second nest cover a period from 1971 until 2100, including a three-year spin-up, as a transient simulation using the RCP8.5 as emission scenario (van Vuuren et al., 2011) from 2005 ongoing. Because of the computational costs, the third nest is set up as time-slice experiments containing three 30-year periods (Control,

CTRL: 1971-2000, Near Future, **NF**: 2021-2050; Distant Future, **DF**: 2071-2100). All time-slices include a three-year spin-up period. The modelling domain for the 0.025° is displayed in Fig. 5.1a; the evaluation area, for which all analyses in the next chapters are conducted, is marked with the solid red line. Additionally, the names for the regional model simulations are shortened to CEH6 (CCLM_MPI-ESM-LR), CEC-EARTH (CCLM_EC-EARTH) and CHadGEM2 (CCLM_HadGEM2-ES) and used as such hereafter.

(a) Relief

(b) Land use

(c) Soil types TERRA_ML
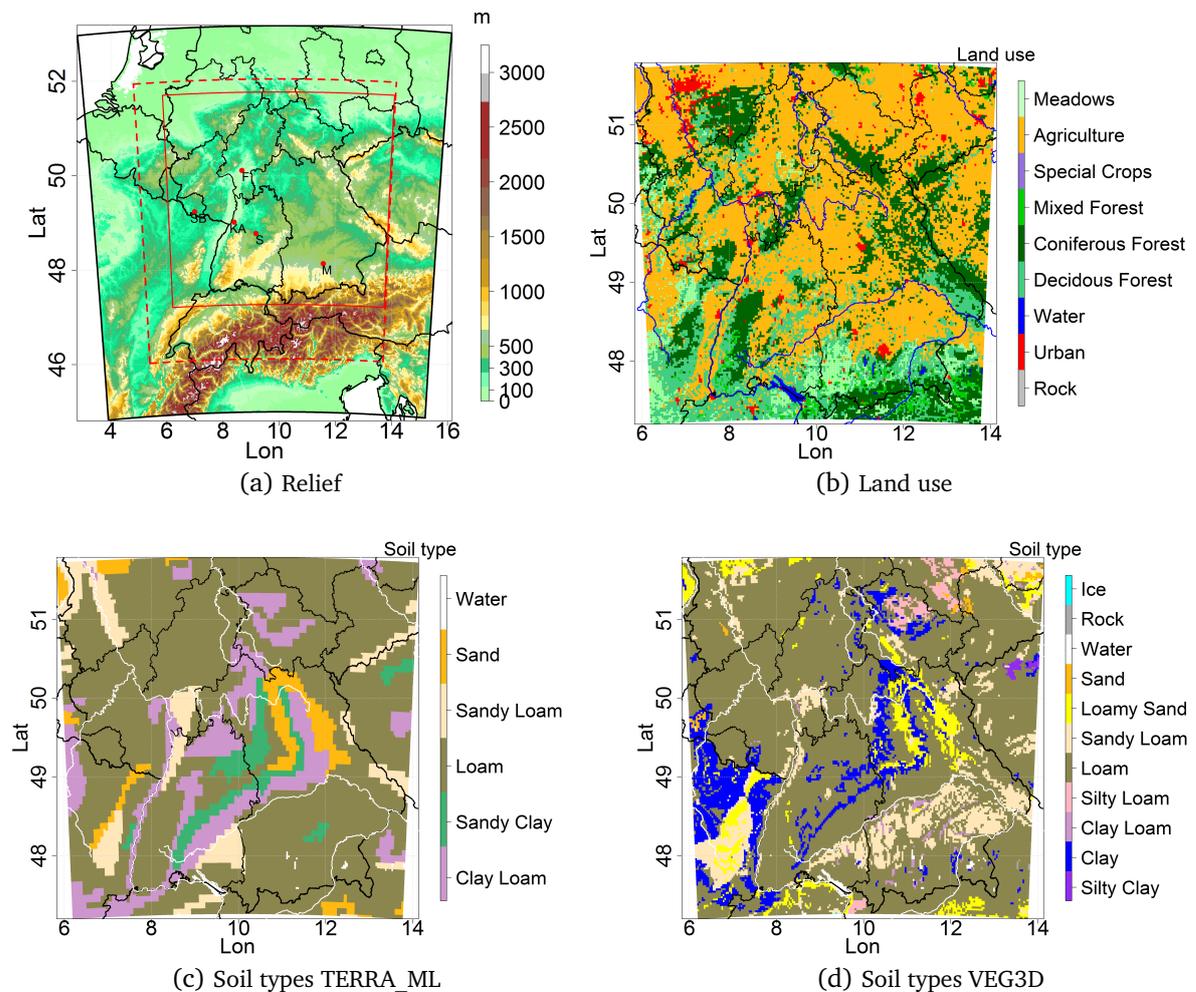
(d) Soil types VEG3D

Figure 5.1.: In the upper left graphic the whole simulation domain is shown with the evaluation area in solid red; the upper right shows the land use classes prevailing in the region. The lower row displays the soil types for TERRA_ML (left) and VEG3D (right).

For each variable, the model results are compared to observations for each model itself and the ensemble (ENS) mean on the grid of the final resolution of 0.025°. For the 2m temperature, precipitation and relative humidity the model data are compared to HYRAS observations (Rauthe et al., 2013) interpolated from 5 km to 0.025°. For the global radiation, a data set produced by the Landesanstalt f̈ur Umwelt Baden-Württemberg (LUBW) called HYDRON afterwards is included Landesanstalt für Umwelt Baden-Württemberg (LUBW) (2012). Unfortunately, HYDRON is only available for a smaller area and a shorter time (1983-2000). Furthermore, climate in-

dices are calculated (Zhang et al., 2011), the reference is climate data produced by the German Weather service Deutsche Wetterdienst (2018a). The Eobs V.14 (Haylock et al., 2008) is used for the comparison of the HW indices interpolated from 25 km to 0.025°. Therefore, interpolation inaccuracies are accepted due to the lack of a finer resolved data set of daily maximum temperatures.

Beside the HW variables presented earlier, two indices are tested to measure the magnitude of HWs. With this, powerful heat waves can be detected, and estimation is given, if it is possible to simulate severe HW events with CCLM for Southern Germany (similar to mega-heatwaves, e.g. (Barriopedro et al., 2011; Bador et al., 2017; Sanchez-Benitez et al., 2018)). The mentioned studies analysed HWs for larger areas like Europe, but the detection of HWs for smaller areas is not mentioned further. The first index, the standardised heat index (SHI), is the product of the annual maximum HW temperature and the sum of the affect grid points of the area.

$$SHI = \frac{(HW_{temp} * \sum GP) - \overline{HW_{temp} * \sum GP}}{SD(HW_{temp} * \sum GP)} \qquad [5.1]$$

The second, the standardised heat area index (SHAI), uses the intensity instead of the actual temperatures combined with the area sum. The HW intensity is the exceeding difference of the maximum heat wave temperature to the 90th percentile. Using this measure instead of the actual temperatures has the advantage, that the model bias is less relevant for the results. This study attempts to define suitable thresholds to build an event set of mega-heatwaves similar to Barriopedro et al. (2011). Barriopedro et al. (2011) defined a mega-heatwaves as one exceeding at least 3 SD the mean temperature anomalies 1970-1999 and with an area of ~1 million km^2 (10 % of the whole Europe). In this study, mayor or severe heat waves are noted, when the differences of the yearly maximum heat wave temperature to the 90ths percentile exceed 4.5 °C (nearly 3 SD) and at least two-thirds of the evaluation area is affected (~350x350km^2).

5.3. A very high-resolution climate ensemble evaluated for various variables and climate indices for the control and future periods

5.3.1. Evaluation against observations and added value to coarser resolution for the control period

In this comparison, four variables have been evaluated: 2 m temperatures, precipitation, 2 m relative humidity and the global radiation. The first three variables are compared to HYRAS for the CTRL period, while the global radiation is compared to the HYDRON data. Additionally, ETCCDI climate indices are presented.

5.3.1.1. 2 m Temperature

The 2 m temperature is better represented in the 0.025° simulations due to the smaller cold bias. In the annual ENS mean, the differences to the observations are reduced from -1.3 °C in the 2ndNest to -0.8 °C in the 3rdNest. In this simulation, a smaller error is detected for the summer half-year (-0.4 °C) compared to the winter months (-1.2 °C). The error to the observations is displayed in Fig. 5.2a for the summer and Fig. 5.2c for the winter half year. The generally higher temperatures reduced the deviations to the observations for the finer resolved simulations during the whole year. This decreased bias is more prominent in the summer months (+0.67 °C) than in winter (+0.34 °C). The different ensemble members show varying annual amplitudes which enlarges the model spread substantially. CHadGEM2includes the largest amplitude, while CEH6 and CEC-EARTH have a similar range. However, the observations are not completely reproduced even by the higher resolved ensemble, especially in the winter half year.

The annual cycle of the 2 m temperature is displayed in Fig. 5.3. As said before, the winter temperatures are underestimated by all models; also, CHadGEM2 overestimates them in the summertime noticeably. In the graphics 5.4a until 5.4f show the spatial distribution of the 2 m temperatures, the differences of the ENS to the observations and the MSESS comparing the 2nd and 3rd Nest.

In Fig. 5.4a, the highest temperatures in HYRAS are found in the Rhine Valley and along Rivers like the Danube for the summertime (> + 18 °C). In the winter, the mountainous regions like the Black Forest, the Swabian and Frankonian Alps or the Ore Mountains show the lowest temperatures (< + 2 °C), while the Rhine Valley is warmer again (> + 4 °C). The ENS underestimates the observations in the summer and winter half years as seen in the annual cycle. The largest differences are located in the Northern part of the evaluation domain and in mountain areas in the summertime (< - 0.9 °C). Although the very high resolutions simulations include a more realistic topography, these deviations indicate the model is not able to fully reproduce the observations in areas with fast-changing orography. The MSESS of the 3rd Nest to the 2nd Nest to observations is displayed in Fig. 5.4c and Fig. 5.4f. The overall reddish colours show the higher skill of the 3rd Nest compared to the 2nd Nest.

(a) ENS 0.0625° vs ENS 0.025° Summer

(b)

(c) ENS 0.0625° vs ENS 0.025° Winter

(d)
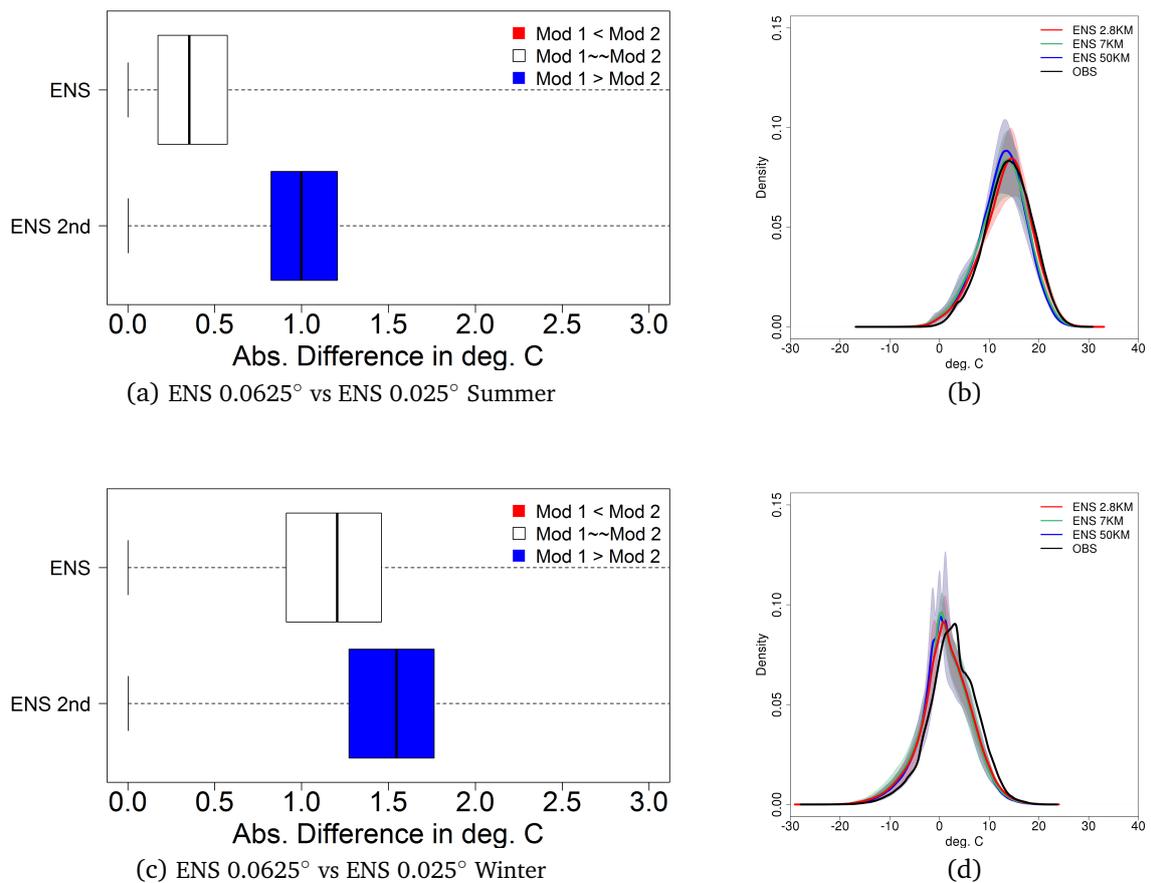
Figure 5.2.: 1971 – 2000 Absolute Differences of the ENS mean of the 2ndNest and the 3rdNest to HYRAS observations in [°C]. With a Wilcoxon rank test, significant differences are tested and marked with coloured boxes (blue: error is significantly larger, white: not significant and red: significant smaller). On the right side, the density distributions for the ensembles at the three nesting stage (0.44°, 0.0625° and 0.025°) and the observations are displayed.
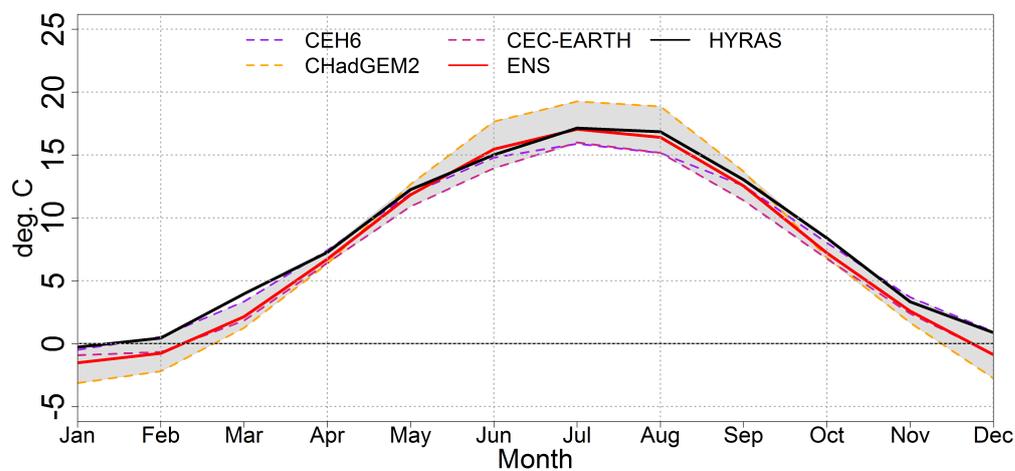
Figure 5.3.: Annual cycle of the 2 m temperatures in [°C] for the KLIWA area 1971 - 2000; shown are the single ensemble members (dotted lines), the ensemble mean (the red line), the ensemble spread (grey area) and the HYRAS observation data (black line).

(a) HYRAS Summer   (b) 3rd Nest ENS–HYRAS Summer   (c) MSESS 3rd to 2nd Nest Summer

(d) HYRAS Winter   (e) 3rd Nest ENS–HYRAS Winter   (f) MSESS 3rd to 2nd Nest Winter
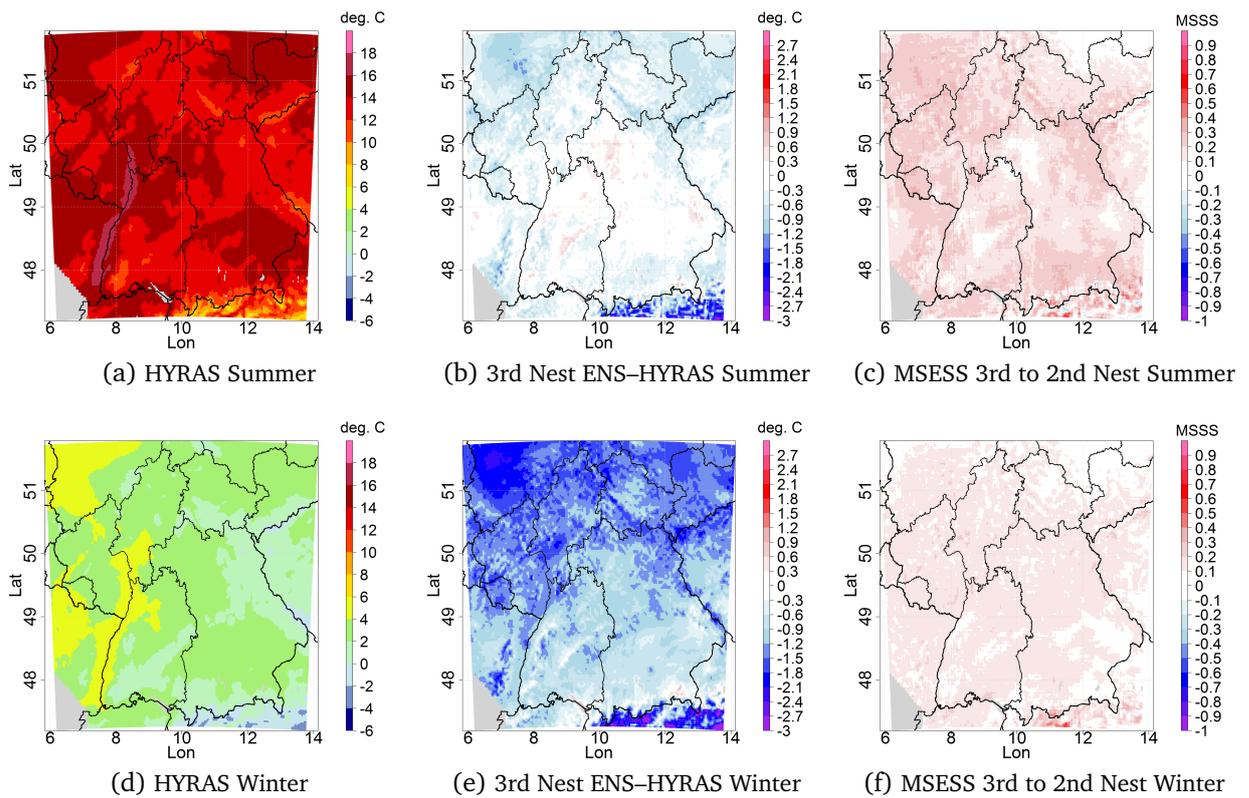
Figure 5.4.: 1971 – 2000 summer (upper row) and winter (lower row) half year mean temperatures of the HYRAS observations in [°C], differences of the 3rdNest (middle) and MSESS of the 2nd Nest compared to the 3rd Nest to the observations (reddish coloors indicate higher skill for the 3rd Nest, blueish coloors lower skill).

5.3.1.2. Precipitation

The mean annual precipitation sums are increased in the 3rd Nest (1040 mm) by nearly 20 % compared to the 2nd Nest (912 mm). Thus, the very high-resolution ENS overestimated the annual sums of the observations largely (rel. difference of +22 %). Fig. 5.5b and Fig. 5.5d show the density distributions for the ensembles compared to the observations for the summer and winter half-year. The 0.0625° ensemble fits noticeably better to the observations in both seasons. Foremost, the larger precipitation sums in the head of the distribution are overestimated by the higher resolved 3rd Nest. The boxplots of the absolute differences in Fig. 5.5a and Fig. 5.5c show that the 2nd Nest has a significantly lower error compared to the observations (red colours). The size of the error differs notably between the height level. In particular, the highest deviations are detected at the highest altitudes. GP over 800 m show a rel. difference of 0.0625°: +25 % which is by far smaller than for the 0.025°: +45 %. Although this layer has the highest annual precipitation sums in the observations, the values are exceeded excessively because of the exaggerated rainfall amounts in the model. Two effects seem to contribute to the development of these large differences. One is the topography, the other the resolution of the observations. The model appears to be unable to develop such excessively high amounts of precipitation using

the orography of the coarser resolved 2nd Nest (as seen in several studies, e.g. Fosser et al. (2015), Ban et al. (2014)). The HYRAS data have a very high resolution of 5km. Nevertheless, precipitation amounts can be underestimated especially in mountain regions (due to a lack of stations and the gridding process). Berg et al. (2012) estimated an error range of 10 to 50 % of HYRAS to the REGNIE data (Algorithm: (Rauthe et al., 2013; Deutsche Wetterdienst, 2018d)) due to a missing undercath correction in HYRAS. This undercath correction reduces the error of heated precipitation gauges made for winter precipitation and snowfall measurements due to evaporation and wind. It could be presumed, that errors in the 3rd Nest might actually be lower than seen in this evaluation. Ban et al. (2014) confirmed this partly as their higher resolved simulations fitted better to the higher resolved observations (or station data).

(a) ENS 0.0625°vs ENS 0.025°Summer   (b)

(c) ENS 0.0625°vs ENS 0.025°Winter   (d)
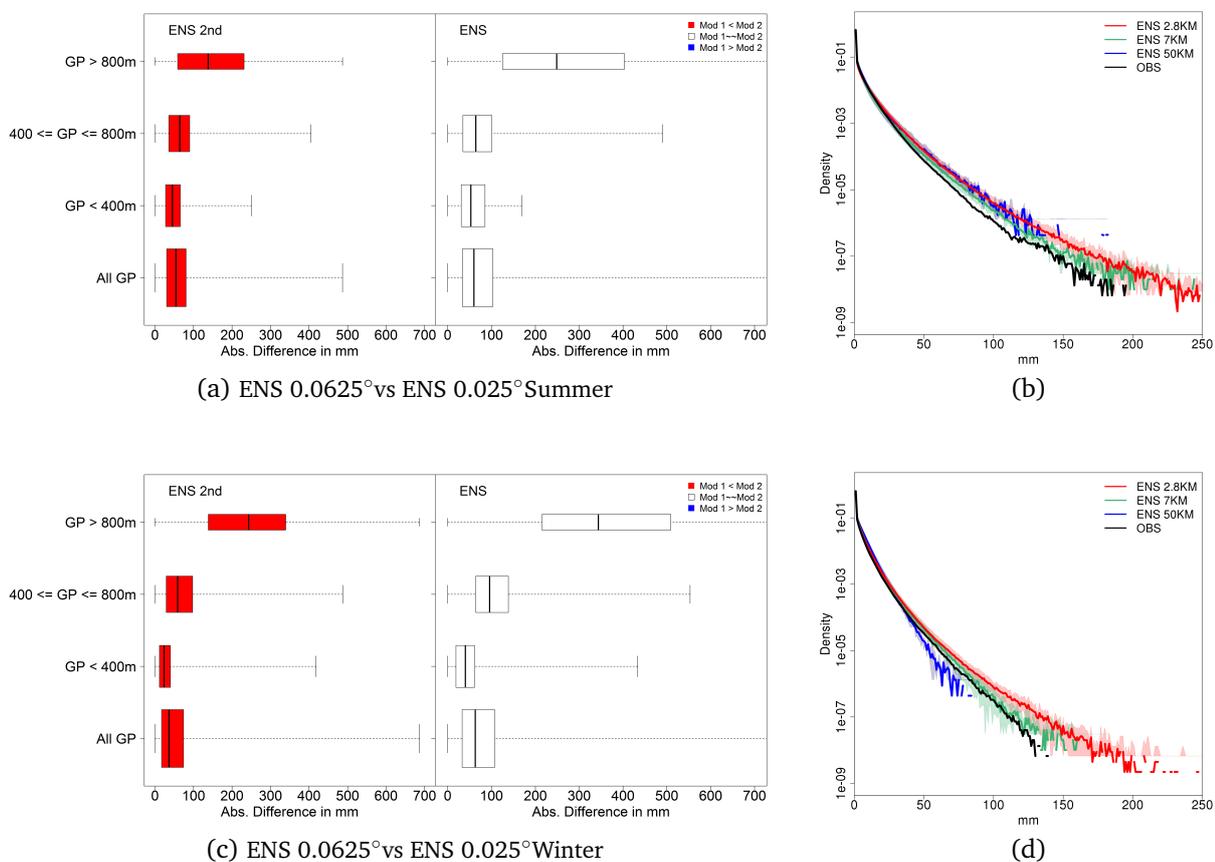
Figure 5.5.: 1971 – 2000 Absolute Differences of the ENS mean of the 2ndNest and the 3rdNest to HYRAS observations in [°C]; colour coding same as Fig. 5.2a but for different height level. On the right side, the density distributions for the ensembles at the three nesting stage (0.44°, 0.0625° and 0.025°) and the observations are displayed.

The error of both nests is larger in the wintertime. The annual cycle of the 3rd Nest is displayed in Fig. 5.6. All ensemble members overestimate the precipitation in the winter season, while CHadGEM2 underestimates the rainfall of the same magnitude as CEH6 overestimates it during the summer. Thus, the average ENS error is reduced in the summer season, whereas the spread is increased. This result is alike in the 2nd Nest and therefore not displayed here.
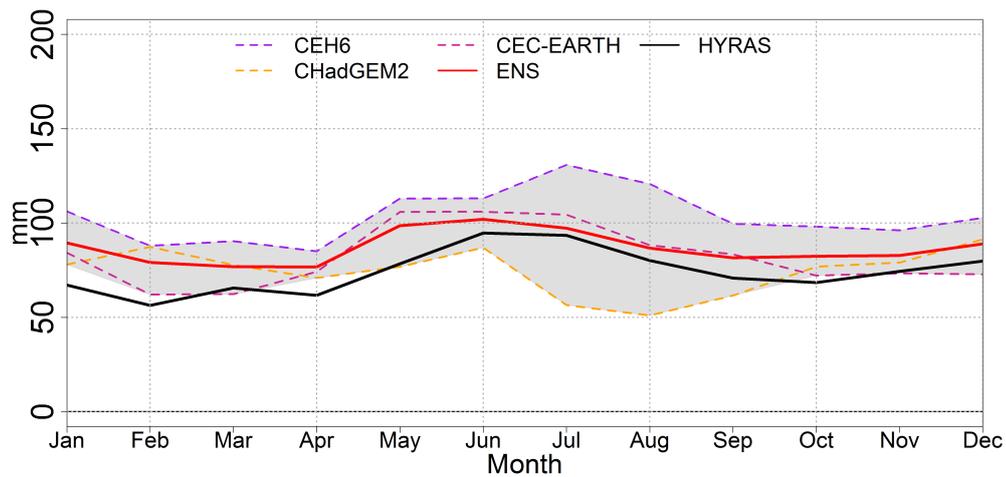
Figure 5.6.: Annual cycle of the precipitation sums in [mm] for the KLIWA area 1971 - 2000; shown are the single ensemble members (dotted lines), the ensemble mean (the red line), the ensemble spread (grey area) and the HYRAS observation data (black line).

(a) HYRAS Summer   (b) 3rd Nest ENS–HYRAS Summer   (c) MSESS 3rd to 2nd Nest Summer

(d) HYRAS Winter   (e) 3rd Nest ENS–HYRAS Winter   (f) MSESS 3rd to 2nd Nest Winter
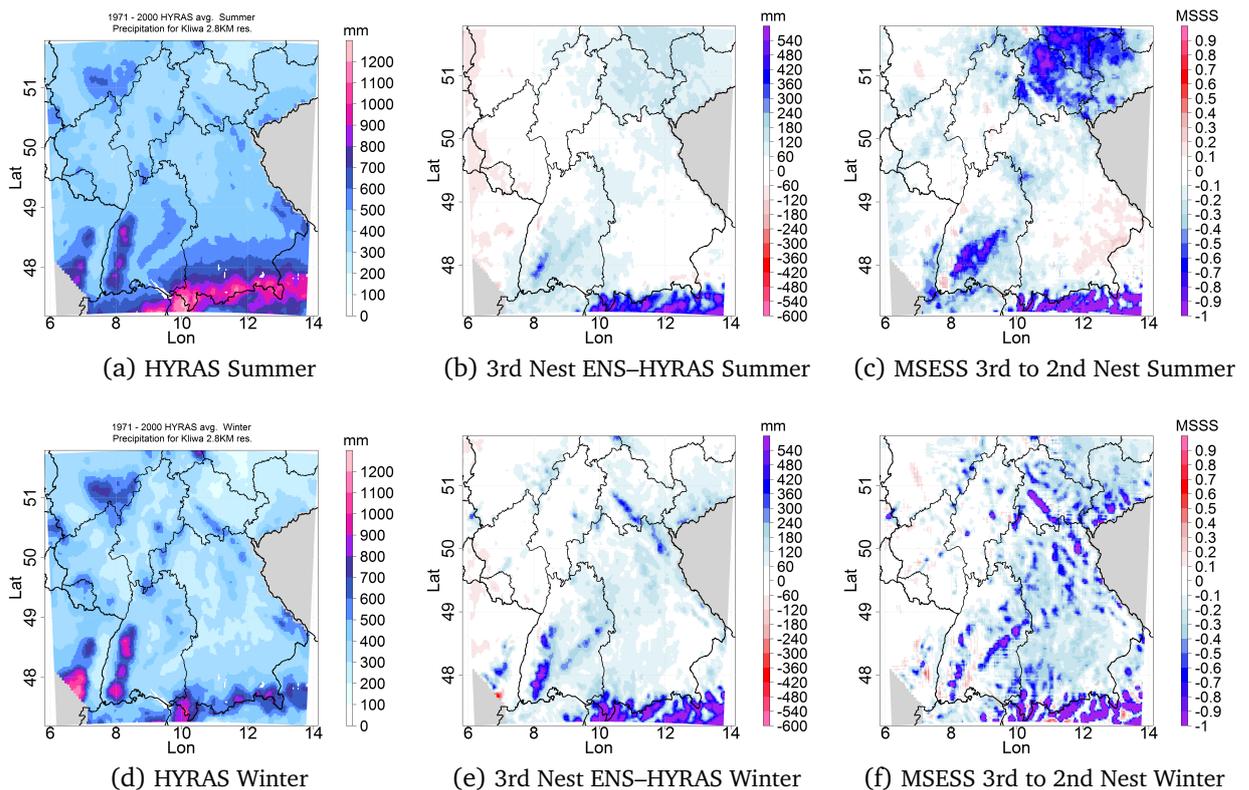
Figure 5.7.: 1971 – 2000 summer (upper row) and winter (lower row) half year mean precipitation of the HYRAS observations in [°C], differences of the 3rdNest (middle) and MSESS of the 2nd Nest compared to the 3rd Nest to the observations.

The summer and winter spatial precipitation sums are displayed in Fig. 5.7a and Fig. 5.7d for the evaluation domain. The highest rainfall values are detected in the Black Forest, the Vosges

and the Alps, while the plains rarely reach > 700mm per season (with lower values in the winter time). In particular, the peaks of the mountains display noticeable precipitation amounts in the winter, while the precipitation is generally higher in the summer half-year. Fig 5.7b and Fig. 5.7e show the differences of the 3rd Nest to the observations.

The ENS overestimates the precipitation in both seasons; additionally, a weak west-east gradient is noted. In winter, the overestimation of the rainfall amounts is most prominent in the mountainous regions, where corresponding negative MSESS values are detected. The mostly negative MSESS indicates the better agreement of the 2nd Nest to the observations as a result of the often lower precipitation of the 2nd Nest.

5.3.1.3. Global irradiation, fluxes, relative humidity and cloud cover

The annual averages of the 2 m relative humidity of the 3rd Nest ENS mean are in excellent agreement with the observations (annual means HYRAS: 78.5 %, absolute difference to ENS: +0.2 %). The coarser 2nd Nest shows a slightly larger error with +0.8 %. The ensemble members differ largely, the spread (maximum–minimum) of the annual means exceeds 8 %. Additionally, a high seasonality is found for the spread, in the summer it nearly doubles mainly due to the largely reduced relative humidity in CHadGEM2(62 % compared to 83 % in winter). The extreme low summer values of CHadGEM2 can be seen clearly in the annual cycle, see Fig. 5.6.
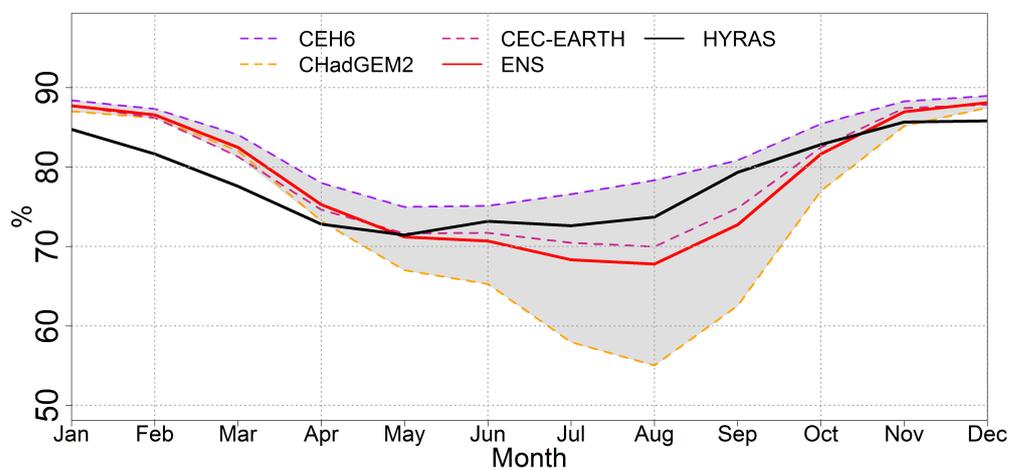
Figure 5.8.: Annual cycle of the 2 m relative humidity in [%] for the KLIWA area 1971 - 2000; shown are the single ensemble members (dotted lines), the ensemble mean (the red line), the ensemble spread (grey area) and the HYRAS observation data (black line).

The 2nd Nest has overall higher relative humidity values in the summer half year and with this a higher agreement to the observations (difference ENS 2nd: +0.85 %, ENS 3rd: -3.37 %). Therefore, the MSESS has mostly negative values (except for the Northeastern part of the domain), while the signal becomes noisier and heterogeneous in the winter (not shown).

The global irradiation is compared to the HYDRON data set for 1983-2000. As mentioned, the data set contains a much smaller area (only Baden Württemberg) and half-year sums. For this domain, the annual global irradiation sums are nearly 5 % higher in the 3rd Nest compared to the 2nd Nest. The higher values of the 3rd Nest fit better to the observations (ENS: 1117 kWh/m^2, ENS 2nd: 1043 kWh/m^2, obs: 1140 kWh/m^2), especially in the summer half-year. In the summertime, the largest incoming radiation is found in CHadGEM2 simulations (869 kWh/m^2), whereas CEH6 has the lowest incoming radiation with only 703 kWh/m^2 (see Fig. 5.9). Because of the temporal resolution of the observations, the MSESS is not of interest for this variable.
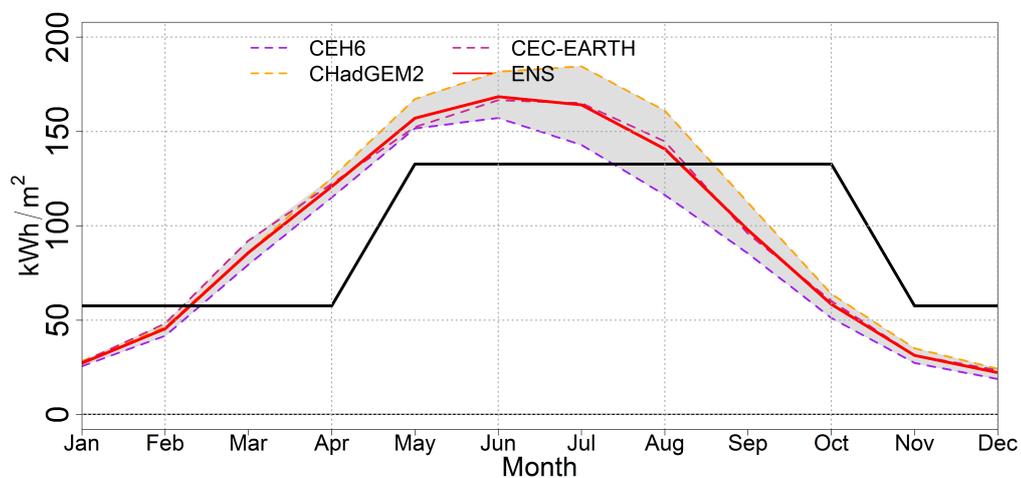
Figure 5.9.: Annual cycle of the global irradiation sums in [kWh/m^2] for the KLIWA area 1971 - 2000; shown are the single ensemble members (dotted lines), the ensemble mean (the red line), the ensemble spread (grey area) and the HYDRON observation data (black line).

The annual cycles of the cloud cover, the latent and the sensible heat flux are displayed in the appendix (see Fig. A.8a, Fig. A.8b and Fig. A.8c). The higher 2 m temperatures in CHadGEM2 in the summer are a result of processes responding to the incoming radiation forcing. CHadGEM2 has lower soil moisture in the upper level 0 until 0.5m (M: 28.2 %, E: 26.2 %, H: 21.2 %) leading to reduced evapotranspiration. Higher sensible flux (M: 43 W/m^2, E: 60 W/m^2, H: 84 W/m^2) and lower latent heat flux (M: 112 W/m^2, E: 106 W/m^2, H: 93 W/m^2) are detected and shift the Bowen ratio for the ensemble members. As a consequence, the alternated radiation balance and the reduced relative humidity induce less cloudiness (M: 77.8 %, E: 72.3 %, H: 62.8 %), lower precipitation and higher incoming irradiation. Comparing the 2nd Nest to the third, slightly different conditions can be assumed. More drizzle leads to higher soil moisture amounts and increased latent heat flux. This results in a lower 2 m temperature, higher cloudiness and lower incoming irradiation. The drizzle assumption was presented firstly by Ban et al. (2014) and seems realistic for the presented ensembles, but it was not explicitly investigated in this thesis.

5.3.1.4. Climate indices

Based on Zhang et al. (2011), four temperature climate indices (frost and icing days, summer and hot days) and three precipitation indices (heavy precipitation I and II and dry days) have been analysed. Additional, a drought index was calculated with the following definition: A drought is measured, when at least eleven consecutive days with less than 1 mm precipitation occur. The number of droughts per year, their mean and maximum duration derived with this definition from the model data are evaluated against DWD observational data (Deutsche Wetterdienst, 2018a). The results of the comparison are summarized in table A.2 for the temperature based indices and in table A.3 for the precipitation based indices in the appendix.

Table A.2 affirms the expected characteristics of the model results. Summer and heat days are underestimated in the ENS mean, while a too large number of frost and icing days occur. CEH6 and CEC-EARTH contain reduced numbers of summer as well as hot days due to their cold bias. On the other hand, the large annual amplitude of CHadGEM2 leads to overly high numbers of summer and hot days as well as to too much icing days. In most cases, the heavy precipitation events are adequately represented by the ENS, whereas the dry days are slightly underestimated in the ENS mean. CHadGEM2 and CEC-EARTH have a good agreement for the dry days with EOBS as can be seen in Table A.3. However, CEH6 underestimates them noticeably. This result implies, that even in the newer model version too much drizzle is produced resulting in too less dry days for simulations with ECHAM6 forcing. Nevertheless, the three dry period indices are quite well simulated by the ENS mean. A higher number and longer dry periods are found for CHadGEM2 also, while heavy precipitations events with more than 10 mm per hour do not differ by considerably between the three models.

5.3.2. Assessing the climate change signal

In the first part, differences between two future times NF: 2021-2050 and DF: 2071-2100 and CTRL will be shown for the 2 m temperature and the precipitation. Afterwards, a summary is presented for other variables of the previous section.

5.3.2.1. 2 m temperature

The 2 m temperature increases by +1.7 °C in the ENS mean for the NF period, which triples by +4.2 °C for the DF future. The lowest temperature increase is found for CEH6 NF: +0.8 °C and DF: +3.0 °C, CHadGEM2 contains the largest rising of the temperatures with NF: +2.9 °C and DF: +6.0 °C. The mean spatial changes are displayed in Fig. 5.10a until Fig. 5.10d for the two future periods. The models show a very robust change signal (all models agree on the sign of the change with a significance of 0.95 % for all grid points), although the strength of the temperature increase varies for the different ensemble members.

(a) NF–CTRL Summer

(b) DF–CTRL Summer

(c) NF–CTRL Winter

(d) DF–CTRL Winter
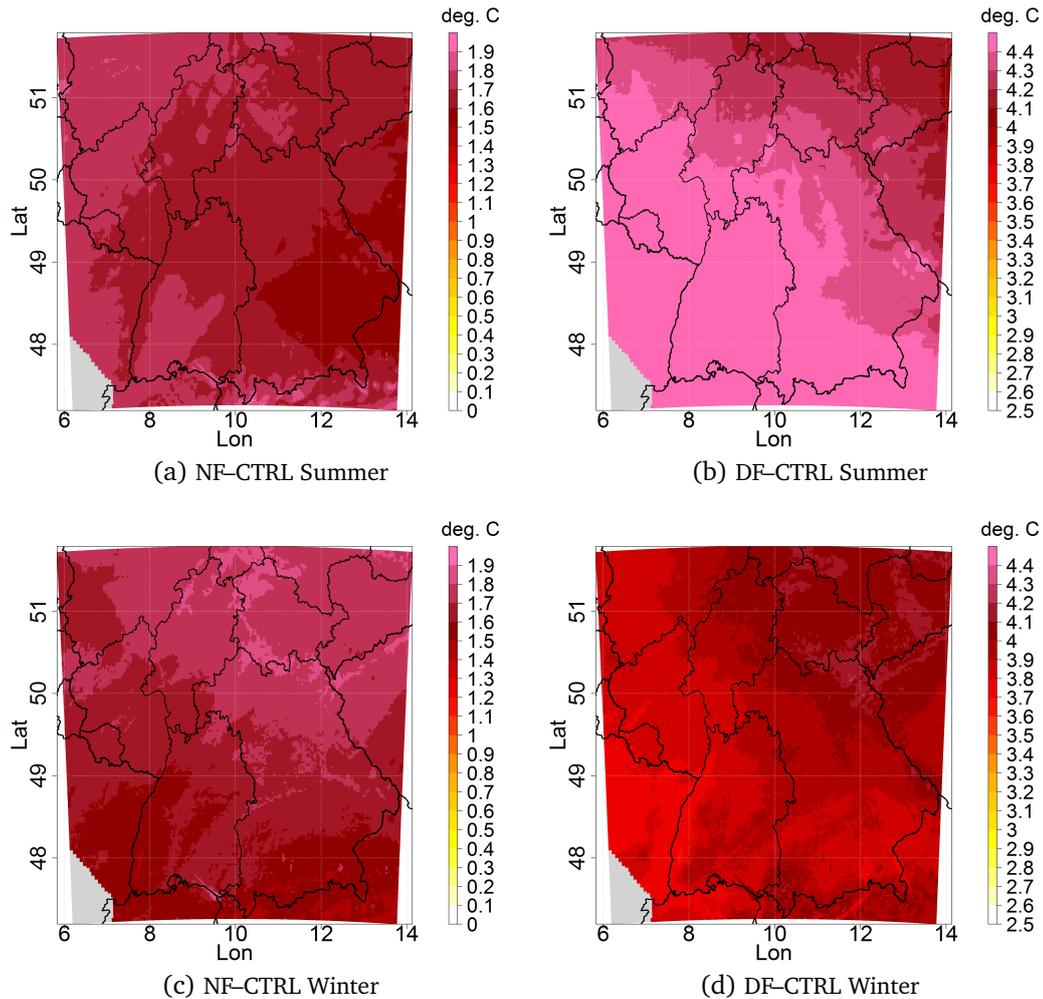
Figure 5.10.: Differences for the summer (upper row) and winter (lower row) half year temperatures [°C] of the two future time slices (NF left; DF right) to CTRL. Marked with grey squares: grid points with a significant change signal. Sum of significant grid points: Summer NF/DF 100 %, Winter NF/DF 100 %, if the grid point sum is below 15 % or above 85 % no grey squares will be displayed due to presentation reasons.

In summer, a stronger rising of the temperature occurs in the western half of the domain as well as in the higher altitudes. Strong temperature increases are detected in the Northeastern part in Hesse, Thuringia, Saxony and the mountains in winter. The largest change occurs in the summer and winter months as can be seen in the annual cycle in Fig. 5.11, whereas spring shows a noticeable smaller signal. In addition, the significance of the change signal has been computed for the four seasons. The only weaker change signal was found in CEH6in Spring (March-April-May), which was less robust with < 42 % significant grid points.

5.3.2.2. Precipitation

The ENS annual precipitation sums show nearly identical average values for the DF compared to the CTRL, while a slight precipitation increase is noticed in the NF;
Difference: $+3.4$ %. The main change occurs in the DF in the summer months, where a consid-

erable reduction of nearly $> -13\,\%$ in the ENS mean can be seen clearly (Fig. 5.12). In addition, a time shift of the months with the maximum precipitation from July to May is noticed in the future periods.
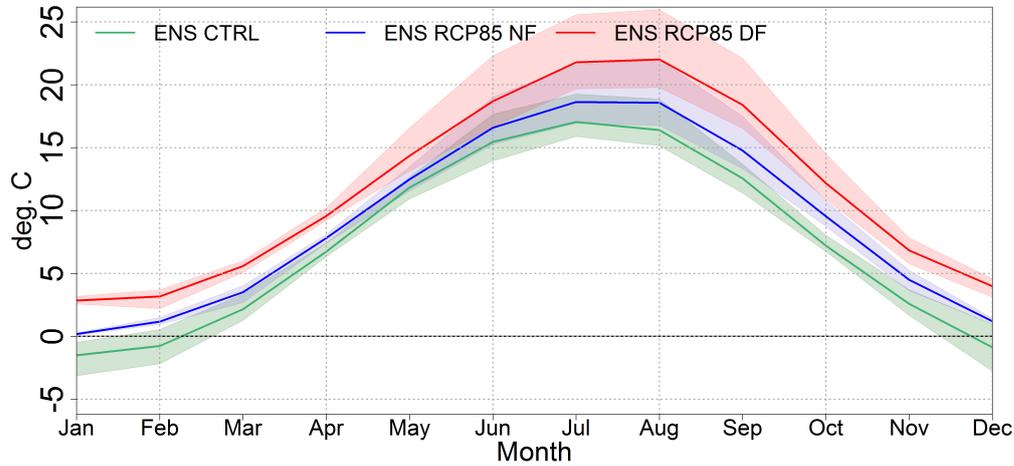
Figure 5.11.: Annual Cycle for the three climate periods (green, blue, red) of the ensemble mean (solid line) and the spread of the ensemble members (max-min, shaded areas).
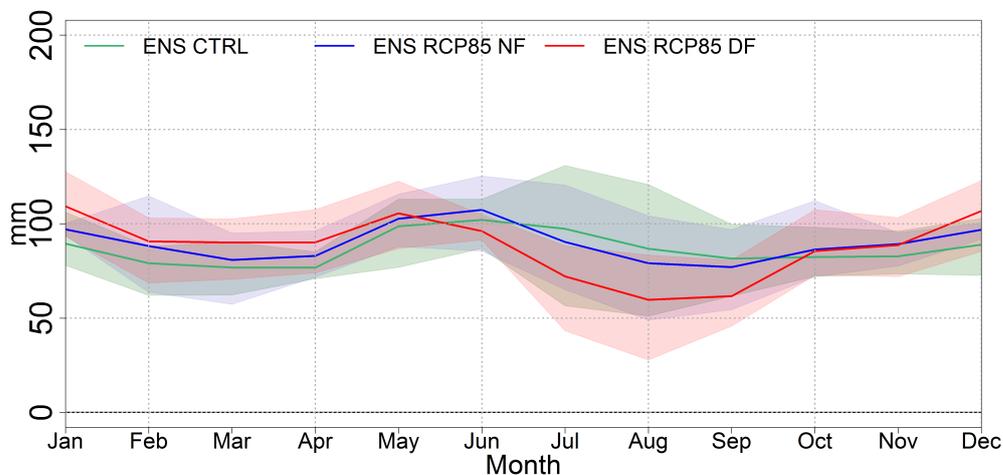
Figure 5.12.: Annual Cycle for the three climate periods (green, blue, red) of the ensemble mean (solid line) and and the spread of the ensemble members (max-min, shaded areas).

The winter precipitation increases by $< 15\,\%$ in the DF as well as the model spread. The change signals range from an increase of $+6.5\,\%$ in CEH6 to $> 25\,\%$ in CHadGEM2 . The signal of the change sign is not as robust as for the temperature in the annual sums of the NF (significance changes for E: 2.0 %, M: 1.2 %, H: 68 % GP). However, the DF shows a more consistent signal for all models (significance changes for E: 29.5 %, M: 4.3 %, H: 40.2 % GP). Additionally, the sums of significant GPs increase when summer and winter half are analysed separately.

The largest differences are found in mountainous regions and the Northern part of the evaluation domain in the NF summer (Fig. 5.13a), while an overall precipitation reduction is noticed for the DF summer (Fig. 5.13b). In the wintertime, the largest increase occurs in the Alps and surrounding areas (Fig. 5.13c and 5.13d), whereas a stronger relative increase is detected in the flatter regions like the Rhine Valley (not shown).

(a) NF–CTRL Summer

(b) DF–CTRL Summer

(c) NF–CTRL Winter

(d) NF–CTRL Winter
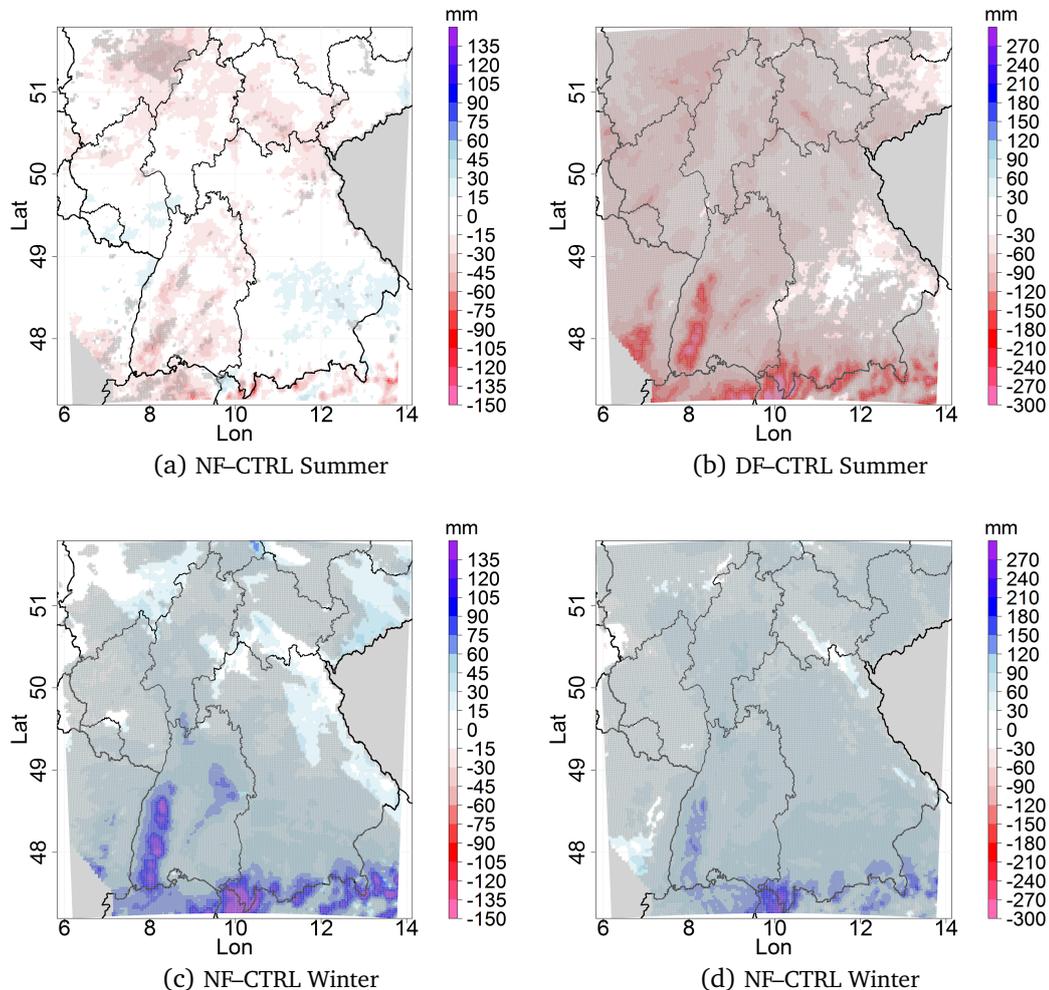
Figure 5.13.: Differences for the summer (upper row) and winter (lower row) half year precipitation sums [mm] of the two future time slices (NF left; DF right) to CTRL. Number of grid points with a significant change signal: Summer NF/DF 7/91 %, Winter NF/DF 80/97 %

5.3.2.3. Short summary of the climate change signal of other discussed variables and climate indices

The NF and DF changes for the global irradiation, the relative humidity and the cloud cover are displayed in the appendix in table A.4. Similar to the precipitation, a weak mixed signal is found for the ENS members in the NF. Despite this, all models agree on the relative humidity decreases in the DF, in particular in summer (see Fig. A.9a). The incoming global irradiation increases for most of the summer months (July until September) as a consequence of the reduction of

rainfall and the relative humidity. The global irradiation is slightly reduced from January until June fitting the shifted annual precipitation maximum in CEH6 and CEC-EARTHin the DF. The decrease of the cloud cover is visible for both future periods (Fig. A.9c), although it is most pronounced in JJA due to the low values in CHadGEM2 . The results are more robust for the change signal of the cloud cover and the global irradiation in the DF than in the NF with stronger variations of the relative humidity in both periods (similar to the precipitation).

The changing soil conditions, amongst others, amplify the increase of the temperature and the reduction of the relative humidity and precipitation. The soil moisture decreases largely in the summertime (Fig. A.10c), which leads to a reduction of the latent heat flux, in particular in the second half of the year (Fig. A.10a). A strong rising of soil temperatures and sensible heat flux (Fig. A.10b) is noticed as a result of the decrease of the soil moisture and the latent heat flux (Fig. A.10e). This shift in the fluxes leads to the lower relative humidity and less transported water vapour into the atmosphere, especially in July and August. Thus, the cloud and precipitation are reduced as seen in the DF.

The change signal for the climate indices, included in section 5.3.1.4, has been evaluated. An increase in the ENS mean for the summer days was revealed by +12 days in the NF and up to +41 days the DF. The number of hot days (T $> 30\,^{\circ}$C) rises from 4 days per year to 9 days (NF) and 25 days (DF). Along with this, the frost days were reduced notably from 104 days in the CTRL to 39 days in the DF. Nearly no icing days were found for the later future period (CTRL: 42 days, DF: 9 days). More dry days occurred (CTRL: 225 days, DF: 238 days) as the mean precipitation is reduced. However, extreme precipitation indices increase slightly. One more day with $>$25 mm precipitation was registered, the number of hours with more than 10 mm rose even stronger (CTRL: 3 days, DF: 5 hours). Thus, the probability of stronger heavy precipitation events rises for the future, although the mean precipitation decreases.

Summary

Current climate conditions:

The 2 m temperatures of the 3rd Nest fit very well to the observations in the summertime (-0.4 °C), whereas a slightly larger error is identified in the winter half-year (-1.2 °C). The increased resolution leads to marginally higher temperatures compared to the previous nesting step, which reduces the mean bias. A moderately positive MSESS shows the higher skill of the 3rd Nest. A large North-South gradient in the deviations to HYRAS is found for the 3rd Nest ENS in the winter. This gradient is not observed in the summer half, where mostly patterns emerge based on the orography. The mean precipitation is overestimated in the 3rd Nest ENS mean with higher relative errors in the winter half (+21 %) than in the summer (+13 %). The spread of the summer is noticeably larger because of the underestimation of the rainfall in CHadGEM2. On average, the 2nd Nest includes lower precipitation sums fitting better to the observations. Negative MSESS are evident, highlighting the lower skill of the 3rd compared to the 2nd Nest in particular in mountainous areas. The relative humidity resembles the results of the precipitation comparison quite well. Furthermore, overestimations of the winter half-year are found, while a considerable ensemble spread dominates the summer. Comparing the 3rd Nest to the HYDRON data revealed too less incoming radiation in all models in the winter, whereas the summer half-year is overestimated. However, the annual cycle can only be estimated as the HYDRON data set contains only half year sums. The fluxes represent the annual precipitation cycle and spread. During the summer, decreased latent and increased sensible heat fluxes are found in CHadGEM2, while CEH6 shows a reduced sensible heat flux due to its higher precipitation amounts and lower incoming radiation. The cloud cover is in line with these results so far and displays a large spread in the summer.

Projection for two future periods:

Change signals for the precipitation, relative humidity and cloud cover are mixed in the NF; however, a small rising of 1.7 °C is noticed for the temperature. The ENS mean shows a significant temperature increase of +4.2 °C in the DF with a considerable spread between the ensemble members ranging from +2.9 °C to +6.0 °C. A significant reduction of the summer precipitation is found of -13 % in the ENS mean, whereas the winter precipitation increases in the similar range. Climate indices based on temperature and precipitation show increasing summer and hot days, decreasing frost and cold days and more dry periods. In particular, the changes are consistent in the summer half-year and describe a chain of effects impacting each other. Decreasing soil moisture, latent heat flux, relative humidity and cloud cover are detected beside the decreasing precipitation values. In contrast, higher values are seen for the soil temperature, the sensible heat flux and incoming shortwave radiation.

5.4. Simulations of extremes and their climate change signal within the very high-resolution ensemble

5.4.1. Evaluation of the CTRL period - Heat Waves

HWs in Southern Germany are evaluated using in the previous section presented very high-resolution ensemble in the first part of this section. Beside the temperature differences, the mean and maximum lengths are analysed as well as the number of HW per year. The definition of HWs is based on the 90th percentile for the climate period of 1971–2000 in this work; this percentile is shown in Fig. 5.14a for the EOBS data. The highest percentile values up to 30 °C are found along major rivers like Rhine and Danube and their tributary streams like Main, Neckar or Isar. The mountainous areas show the lowest temperatures with slightly over 22.5 °C, like seen for the Harz, Erzgebirge or the Black Forest. The CEH6 , as well as the CEC-EARTH, underestimate the EOBS data by -3.1/-3.0 °C in the average, while CHadGEM2has an obvious warm bias with +1.94 °C (ENS mean: -1.5 °C).

(a) 90th perc EOBS V14   (b) 90th perc ENS–EOBS

(c) 90th perc CEH6–EOBS   (d) 90th perc CEC-EARTH–EOBS   (e) 90th perc CHadGEM2–EOBS
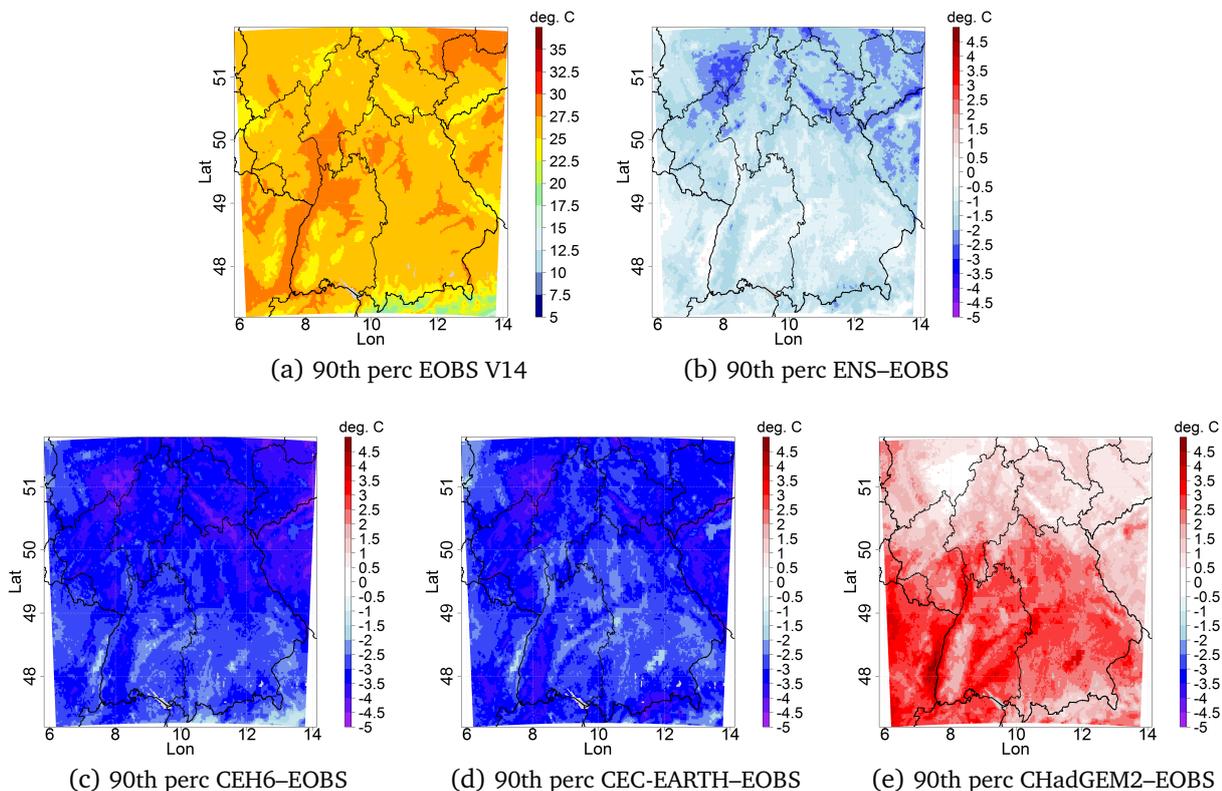
Figure 5.14.: 1971 – 2000 90th percentile Daily Summer maximum temperature of the EOBS V14 data and the differences of the CCLM simulations and the ensemble mean (upper right) to EOBS.

In Fig. 5.15, HW events with their duration, maximum temperatures and global intensity are displayed for the CTRL period. The bubble size represents the global intensity, which is calculated as the sum of differences between the HW temperatures and the 90th percentile. The severest events can be recognised in the upper right corner, separated by two lines (horizontal: mean maximum HW temperature +1 SD; vertical: duration of 15 consecutive days). In Eobs

V.14, 73 Events in the 30 years are detected (slightly more than two per year in the average) with five exceptionally strong HWs in 1976, 1983, 1992, 1994 and 1998. The identified years agree well with other studies (like Russo et al., 2015) and DWD data (Deutsche Wetterdienst, 2018b). Since the correlation for the 2m temperatures is weak (-0.07<r<0.26, not shown), high correlations between the HW temperatures of the simulations and the ones from the observations are not expected (-0.07<r<0.18). Therefore, it is not necessary to create a set of historical events but to evaluate the characteristics of the simulated events like duration, spatial coverage and intensity (maximum length and temperatures). Nearly identical trends can be observed in the maximum temperatures within a HW similar to the bias of the 90th percentile (e.g. shown in Fig. 5.15). The averaged maximum temperatures are over 30 °C in the observations, whereas CEH6 and CEC-EARTH show temperatures < 28 °C (Fig. A.12 in the appendix). A positive bias is found for the HadGEM2-ES, which can be seen in the bubble plot by the higher located orange bubbles. Thus, the ensemble reproduces the maximum temperatures well enough; however, too many HW are simulated (N> 80).

The durations of HWs are shown in Fig. 5.15 for the CTRL period (additionally, the yearly maximum lengths are displayed in Fig. A.13 in the appendix). The HWs lengths range from the 5-day minimum to the longest events with over 30 days in 1976 and 1994 in the EOBS data. The averaged maximum length per year is 6.86 days in the observations in the CTRL period. The maximum length is half a day longer than the mean duration, although single years like 1976(1994) exceed this value by 4(2) days in the mean and up to 8(6) days in the maximum length. HWs are too long in the ENS mean compared to the observations, mostly, due to the large overestimation in CHadGEM2. This simulation has noticeably more events with longer durations as can be seen in Fig. 5.15). Hws are identified nearly a day longer in the mean in CHadGEM2, while CEC-EARTH is in good agreement with the observations. However, CEH6 does not simulate long enough HWs and has a noticeably cold bias.

The global intensity reached in the observations can be found likewise in the simulations; however, the difference between the most severe HWs and the average ones is more substantial in the observations. A more continuous distribution is noticed for the simulations indicating a lower interannual variability in the simulations than in the observations. This might be a result of the mostly longer HWs in the simulations (CEC-EARTH, CHadGEM2), but it was also seen in CEH6 which generally has shorter maximum HW lengths. Nevertheless, it was possible to detect HW events independent from the temperature bias of the model due to the percentile threshold. In particular, CEC-EARTH, having the largest summer mean temperature bias, displays HWs with characteristics similar to the other ensemble members.

These results show that HWs are slightly overestimated in length and number per year in the ensemble, while the HW temperatures of the ENS mean are principally in good agreement with the observations. These findings support the need for a temperature anomaly based index or a minimum area threshold to identify the strongest events. Indices based on the absolute HW temperatures can be critical when derived from huge domains or ones with strong orography variations. Lower HW temperatures are frequently observed, averaging such regions, which can
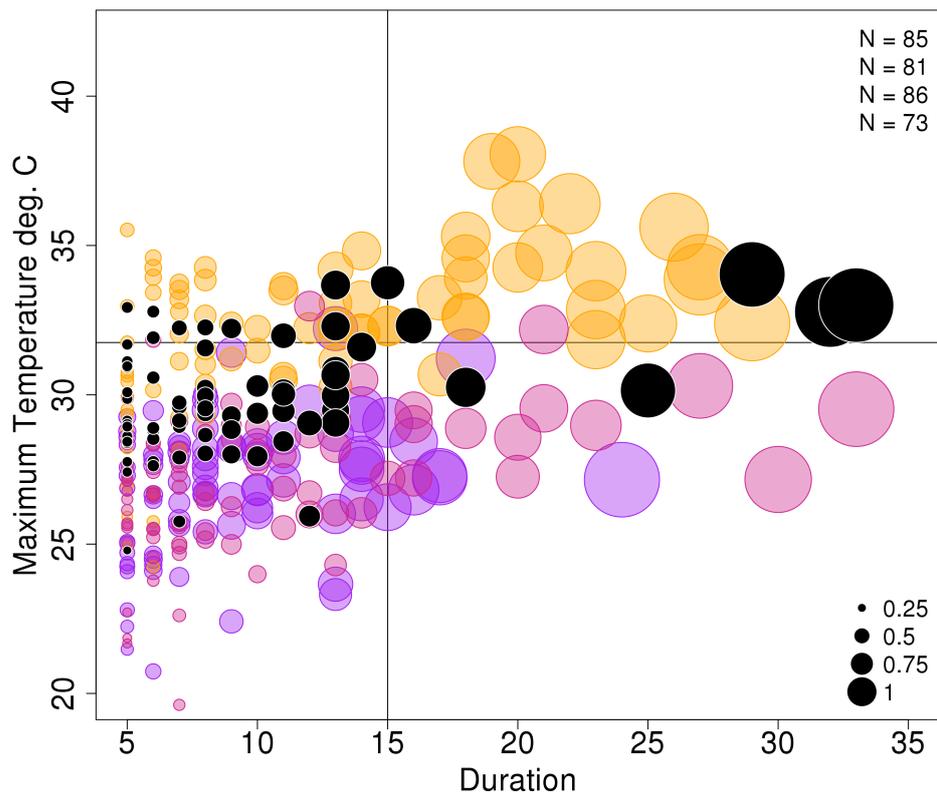
Figure 5.15.: Bubble plot displaying HW temperatures, duration and global intensity for the KLIWA area 1971 - 2000; shown are the single ensemble member (CEH6 purple, CEC-EARTH magenta, CHadGEM2 orange) and the HYRAS observation data (black bubbles). Also the number of the HW in the CTRL period on the right side.

lead to an underestimation of the HW severity. If HWs are detected only for the foothills of the Alps, lower HW mean (maximum) temperatures can be the result of certain years, e.g. for the CHadGEM2 in the year 1989 or other in CEH6 and CEC-EARTH in Fig. A.12. A solution might be an index combining affected area with the temperatures or exceedance of the 90th percentile. Thus, the standardised heat area index (SHI) index is tested employing a combination of heat wave temperatures and affected area/grid points (see chapter 2.2). The annual SHI is shown in Fig. 5.16. For Eobs V.14, six major HWs are identified for six years in the CTRL period (SHI > 0.5), seven in CEH6, six in CEC-EARTHand seven in CHadGEM2. Additionally, the graphic reveals a striking non-symmetric distribution of yearly SHI (range from -3 to +1). The SHI is either in the upper or lowest 20 % of the SHI range in the observations for more than half of the years. This ratio is shifted to 24 of 30 years for the ENS mean.

Less intense heat waves affecting only small areas produce an overly strong negative shift in the distribution of the index for the small evaluation domain. It is assumed, that this index is not symmetrically distributed because of the used absolute temperatures. Significant spatial variability of the HW temperatures due to orographic differences seems to lead to strong negative signals in the SHI if only certain areas like mountain ranges are affected by the HW. Thus,

filtering of HW events affecting only a minor percentage of the domain is sensible as HWs are associated with large-scale circulation patterns (Rust et al. (2015), Sousa et al. (2018), Della-Marta et al. (2007b)). In addition, an index based on anomalies instead of absolute values could improve the results. Therefore, a variation of the SHI, the SHAI is tested in this study. The SHAI requires the exceedance of certain thresholds, which makes it less sensitive to weak events (minimum temperature, minimum affected area). Major heat waves are detected with the SHAI for the years 1976, 1983, 1990, 1992, 1994 and 1998 (also Fig. A.16).
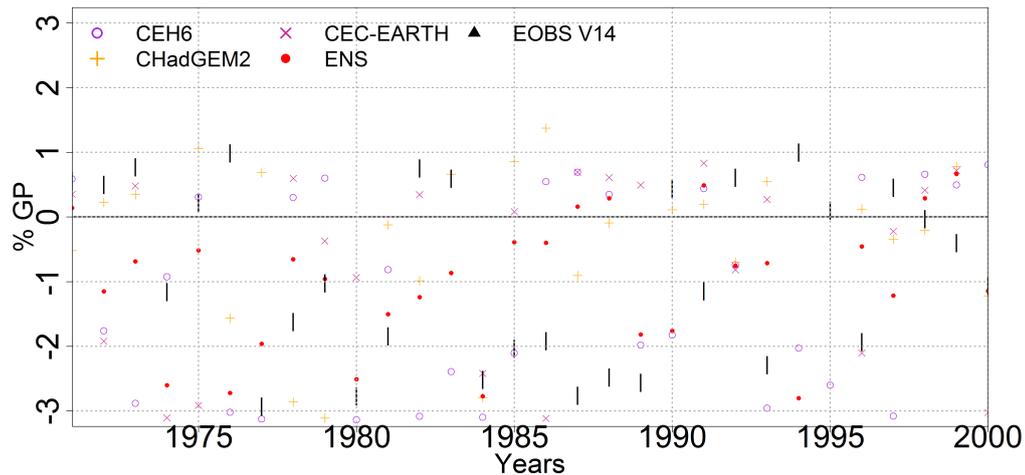
Figure 5.16.: Annual SHI for the KLIWA area 1971 - 2000; shown are the means of the single ensemble member (symbols) and the Eobs observation data (black line).

These years are nearly identical to the ones determined from Fig. A.11. Thus, the most severe HW can be identified by either of them in small enough areas. Major HWs are included even in the simpler maximum HW temperature; nevertheless, years like 1998 are emphasised stronger in the combined indices. If the interest is only in the identification of major events like 1976 and 1994, the non-symmetry of the SHI index can be ignored mostly.

5.4.1.1. Spatial characteristics

The spatial distribution of mean/max. HW length and mean/max. HW temperature and the numbers per year are displayed in Fig. 5.17a until Fig. 5.17e. Local maxima of HWs length up to 9/11 days are found for the mean/max. length in the Rothaargebirge, the Ruhrgebiet and the area of Frankfurt am Main. The shortest HWs are detected in the lower alpine area in Bavaria and in Saxony with durations lower than six days. This distribution seems independent of height level or land use as the Black Forest and the Erzgebirge show vastly different lengths as well as agricultural areas or forest. No apparent dependences to external parameters are recognised for the number of HWs per year also. The spatial SD is low in the Eobs data with more notable variations from the mean (1-2 HWs per year) only in very small areas. The spatial pattern in the mean and max. HW temperatures are nearly identical, the largest temperatures are found along major rivers like the Rhine (especially in the area of Frankfurt and along the Main and Neckar),

parts of Danube and in the area around Wurzburg. Temperatures over 30 °C (mean) or 33.5 °C (max.) are seen in these areas. Lower values are located in mountainous regions like the Black Forest or the Erzgebirge (mean temp. < 25 °C and max. temp. < 29 °C). Fig. 5.18a until Fig. 5.18e display the difference of the ENS mean compared to Eobs for the five variables.

(a) mean HW len   (b) max HW len   (c) No per year

(d) mean HW temp   (e) max HW temp
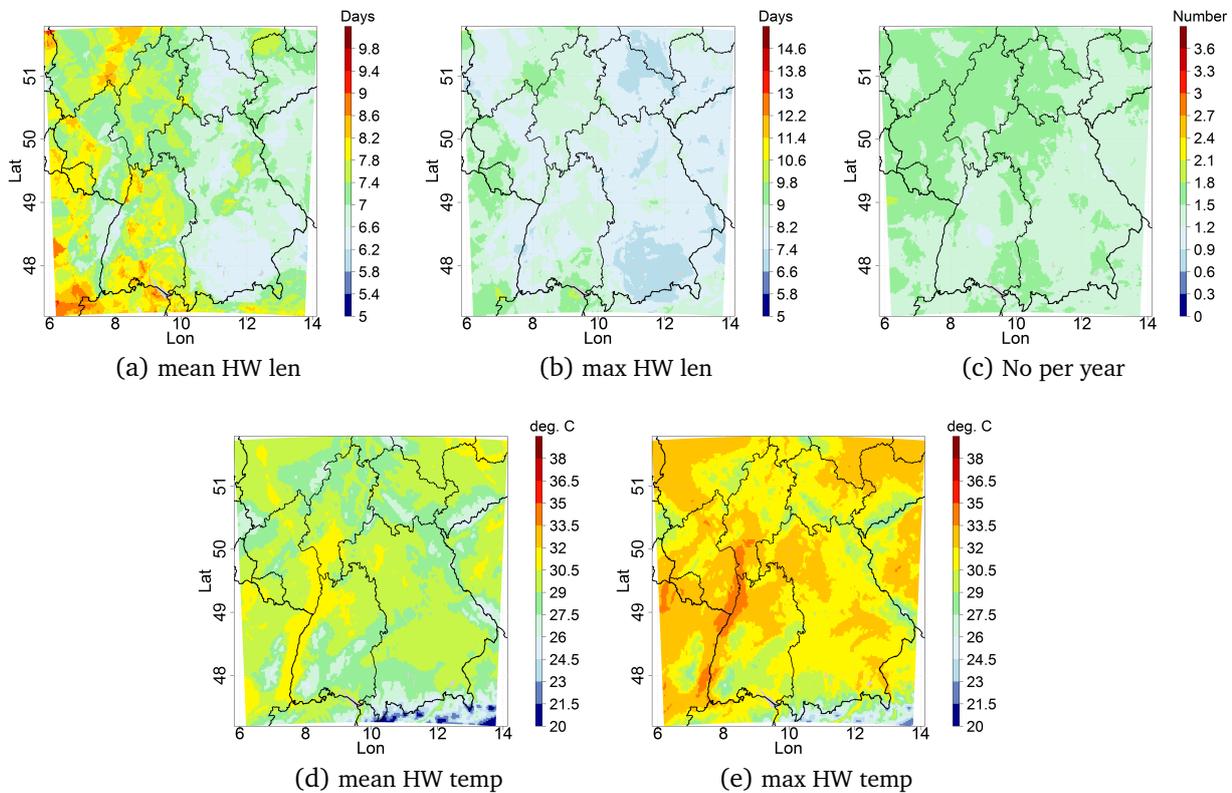
Figure 5.17.: 1971 – 2000 Mean and Maximum HW length and temperatures and number of HW per year for the EOBS V14.

The mean HW lengths show an overestimation in the eastern part of the evaluation area of < +2 days. However, the patterns of the lengths differences are more heterogeneous in the western half with over- and underestimations of from -2 to +1 days. The spatial patterns of the mean and maximum values are remarkably similar, although the error for the maximum HW temperatures is slightly larger than in the mean. In particular, CHadGEM2 overestimates the eastern regions with +3 days considerably. The models overestimate the number of the HW with up to < 30 % nearly in the whole domain (Fig. 5.18c). The error in the mean and maximum HW temperatures shows a slight northwest-southeast gradient. Mountainous regions are highlighted as most of the models produce too cold HW temperatures in these regions. The underestimation of CEC-EARTH (-3.0 °C) and CEH6 (-3.2 °C) is pronounced for most of the evaluation domain, whereas CHadGEM2 is too warm with +2.2 °C for the mean HW temperatures. The cold bias is slightly reduced (-0.2 °C) in the maximum HW temperatures as the models show higher averaged values than for the mean HW temperatures (difference to observations: -1.4 °C).

(a) mean HW len (b) max HW len (c) No per year

(d) mean HW temp (e) max HW temp
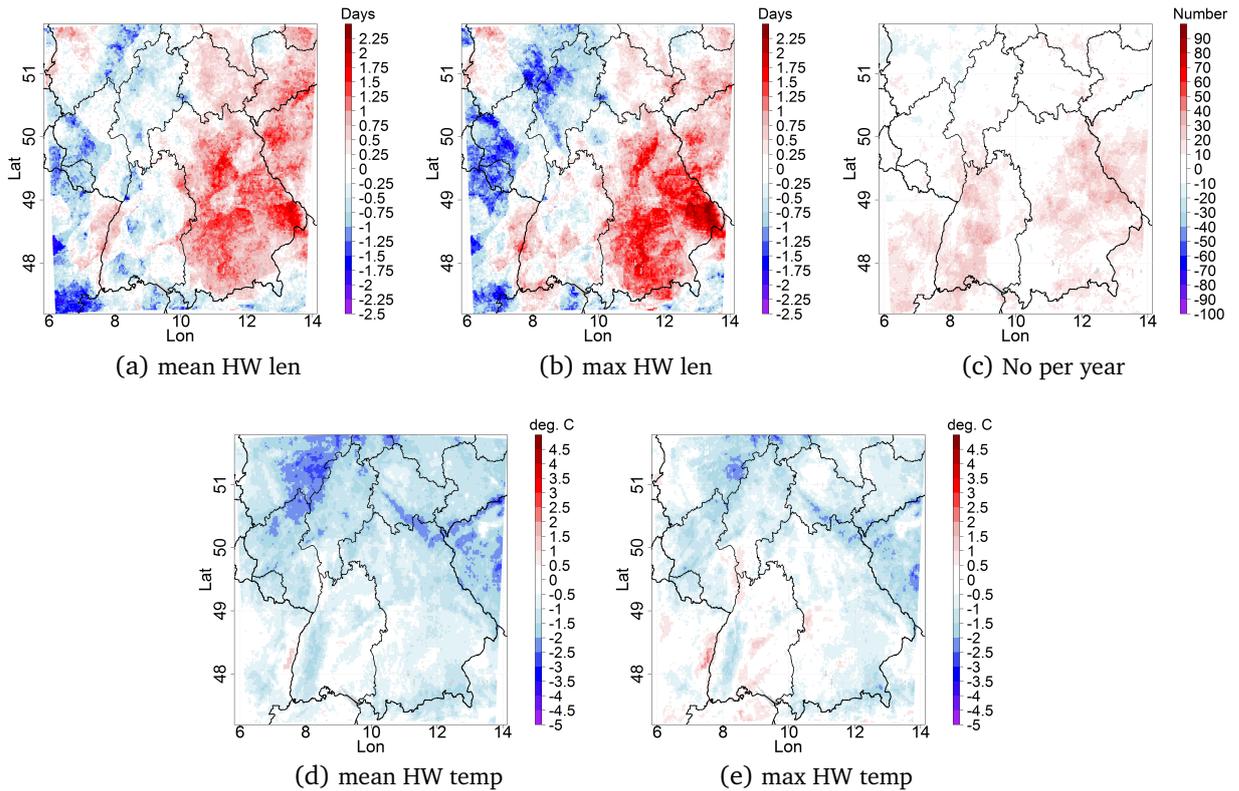
Figure 5.18.: 1971 – 2000 Mean and Maximum HW length and temperatures differences between the ENS means and EOBS V14, for the number of HW per year the relative difference is displayed.

5.4.1.2. Annual cycle

The averaged annual cycle for the CTRL period is displayed in Fig. 5.20 and Fig. 5.19 for the heat wave temperatures and the affected area (sums of the grid points). In these graphics, the onset and the decaying periods of HWs can be identified for the evaluation area. The onset and decay are marked by GP sums exceeding 50 % for larger scale HW. Three phases are noted for the Eobs data with GP sums larger than 20 % in both figures. The first short heat period begins late in May and ends in mid-June, while the longer one follows one to two weeks afterwards. The last heat phase is found in early September. No extensive area with heat conditions is seen after the mid of September in the average. The ensemble mean may not be the best way to encompass the annual cycle for this value, since large differences of the onset of the HW are recognised for the ensemble members. CEH6 and CEC-EARTH reach area sums of 50 % earlier in the Mid of May. However, the HW onset in CHadGEM2 occurs in the Mid of June far later than in the other simulations or the observations. In addition, the decay is later for the CHadGEM2 than for the others, although none of the models shows HWs after the beginning of September like the Eobs data. The time shift of the HWs derived from CHadGEM2 might be a result of the occurrences of certain weather pattern. Additionally, onset and decay time could be influenced by the larger annual cycle of the CHadGEM2 and the level of the 90th percentile. Interval periods, when only smaller areas or none at all are affected by HWs, are shorter
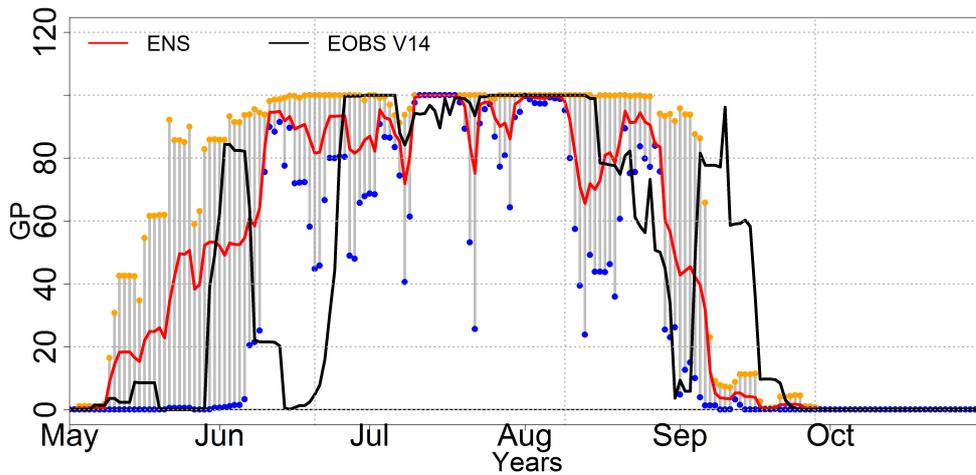
Figure 5.19.: Annual cycle of grid points sums of heat wave events for the KLIWA area 1971 - 2000; shown are the ensemble minimum and maximum (blue and yellow dots) with the ensemble spread (grey area), the ENS mean (red line) and the HYRAS observation data (black bars).
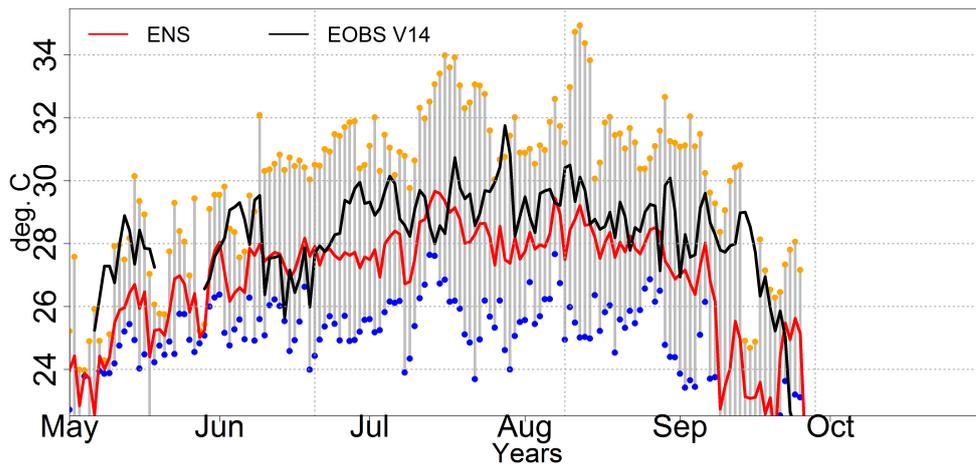
Figure 5.20.: Annual cycle of HW temperatures for the KLIWA area 1971 - 2000; shown are the ensemble minimum and maximum (blue and yellow dots) with the ensemble spread (grey area), the ENS mean (red line) and the HYRAS observation data (black bars).

for the ensemble. In particular, this is observed for the CHadGEM2, whereas CEH6 shows a larger break in the mid of August. The annual cycle of the HW temperatures is displayed in Fig. 5.20. Temperatures exceeding the 90th percentile of the summer half-year are found in the observations for a short phase in May and from June until the end of September. Although the temperatures reach $> 28\,°C$ in May, this affects only limited areas as seen before in Fig. 5.19. During the more powerful events in July and August, the mean HW temperatures exceed $30\,°C$ in the area averages; the highest values are often detected at the end of July and the beginning of August. Similar values are recognised in the ENS mean. The ensemble spread is distinctly pronounced in the annual cycle reaching $8\,°C$ for several days deviations (max-min ENS mean

to member). The CHadGEM2 simulation displays the highest temperatures, while for CEH6 and CEC-EARTH values are found lower than the EOBS V14 data. Furthermore, the location of the maximum HW temperature varies between the ensemble members. The maximum occurs in the CHadGEM2 data in mid-July and the early August days, whereas the highest HW temperatures are noted for CEC-EARTH at the beginning of June and at the end of July and in CEH6 in the mid of June and July (Fig. 5.20).

5.4.1.3. Added value of the ensemble approach

Reliable climate projections have to be able to represent all possible values found for the observations. The mean square error skill score (MSESS) is calculated for the ensemble members to the ENS mean to the observations to assess the added value of the ensemble approach (see equation 4.2). The perfect fit would be an MSESS of 0, whereas a MSESS ranging from 0 to 1 shows a higher skill. The patterns of the maximum temperatures and lengths are intensified but spatially similar to the mean ones. Therefore, the mean HW lengths and the max. HW temperatures are displayed as examples (Fig. 5.21a to 5.21f). The MSESS is negative in the average; the single members have no higher agreement than the ENS mean with the observations for the most areas and variables. Exceptions for the maximum HW lengths, where the MSESS is around 0 or positive, are found for CEC-EARTH in most of Bavaria and for CEH6 in the Northern Part of the evaluation domain (Fig. 5.21d - 5.21f). CHadGEM2 only includes MSESS values around 0 in the Southeastern but shows no distinctly higher skill.
Foremost negative values are noticed for the MSESS for the max. temperatures. It seems that, to a certain degree, contrary patterns can be detected for CEH6 and CEC-EARTH. Areas of higher skill in the lengths correspond with areas of lower skill in the maximum temperatures and vice-versa. The North, for example, shows positive skill for the length in CEH6, whereas weak negative skill is discovered in the max. temperatures (Fig. 5.21a to 5.21d). The large bias of CHadGEM2dominates Fig. 5.21f. At least two of three models are able to simulate an accurate number of HWs in good agreement to the observations. However, their spatial coverage and maximum intensity (lengths and temperature) vary strongly. The large biases of the ensemble members are considerably reduced in the ensemble mean. Therefore, an evident advantage of the ensemble approach is shown for the HW.

5.4.1.4. Added value compared to the 2nd Nest

The MSESS has been calculated to estimate the possible added value of the higher resolution for the control period. Reddish colours show an added value of the 2nd Nest, blueish colors the better agreement of the 3rd Nest with the observations. The MSESS is mostly negative for the HW temperatures (Fig. A.14c and Fig. A.14d), which shows the decreased bias in the 3rd Nest. The added value of the HW lengths is undetermined as the mixed spatial MSESS patterns show (Fig. A.14a and Fig. A.14b in the appendix).
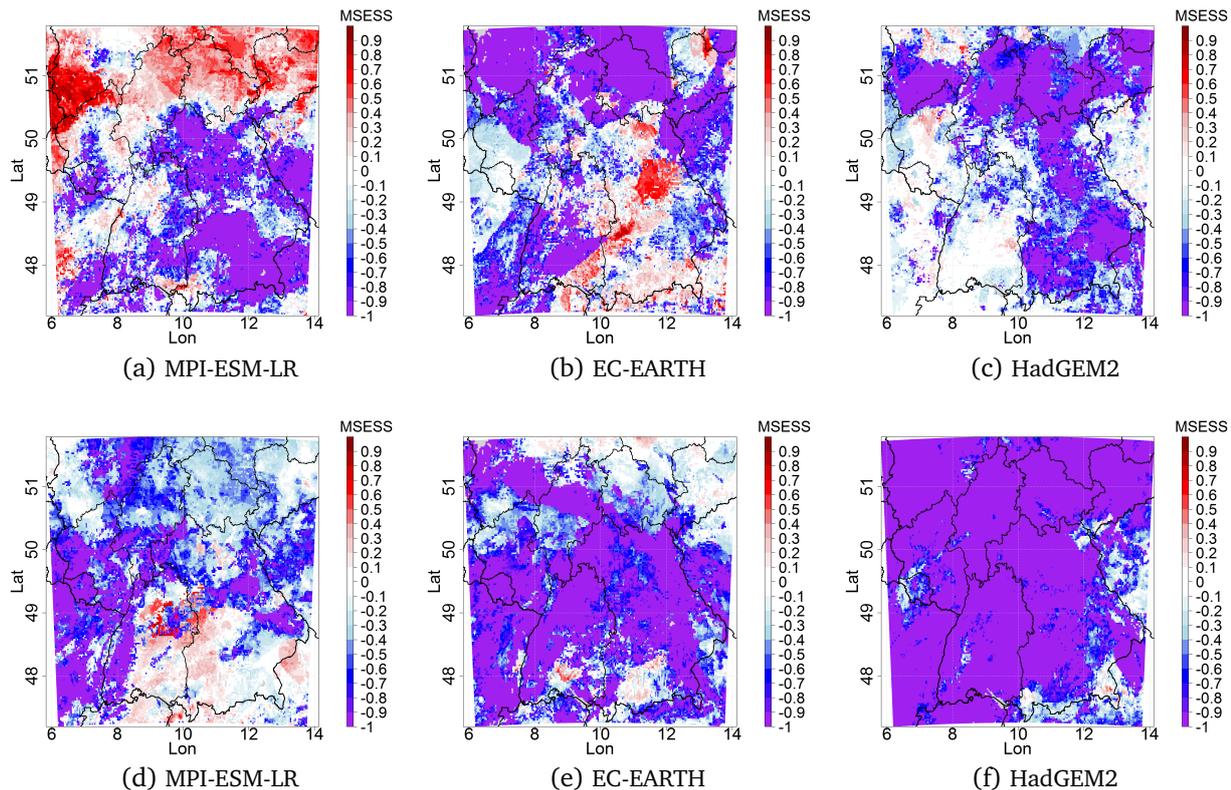
Figure 5.21.: 1971 – 2000 MSESS of the mean HW length (upper row) and the max HW temperature (lower row) of the ensemble member to the ENS mean and EOBS V14.

Normal density distributions have been fitted to model and observation data (see Fig. 5.22). The ability of the models to develop very high temperatures is assessed by Comparing the 25th, 50th, 75th, 95th percentile and the maximum of the daily HW temperatures. The quantiles of the normal distribution fits are displayed in table A.5 (in the Appendix). The median of the 3rd Nest ENS (28.8 °C) is marginally higher than the observations (28.2 °C), the 2nd Nest is slightly too cold (26.9 °C). However, the 2nd Nest has a better agreement to the observations for the higher quantiles (> 75th percentile), while two of three members are noticeably too warm in the cloud-resolving ensemble. The slight overestimation can be seen in 5.22. On the other hand, one has to consider, that values exceeding the 75th occur only on four days per year in the average.

Additionally, the full width at half maximum (FWHM) has been calculated for the density distributions. The FWHM is the length between the x values on the left and right hand of the distribution, where the function values f(x) are only half of the functions maximum f_{max}. This value can be used as a measure to estimate if the spread and variability of the HW temperatures are accurately captured. If the FWHM of the models and observation have the same range, the spread of the ensemble fits well to the observations. A higher FWHM has been found for all members of both ensembles, indicating a broader distribution of HW temperatures compared to the observations. The FWHM of the 2nd Nest is smaller with 5.7 °C (3rd Nest: 6.25 °C), which

fits better to Eobs V.14. Reason for this is for one the smaller FWHM of the ensemble members, but also the different medians of the members. The difference between the 2nd Nest ensemble member medians (5.0 °C) is lower than for the 3rd Nest (5.3 °C). Combining these two effects, the FWHM becomes noticeably larger in the 3rd Nest compared to the 2nd.

Additionally, the interpolation of the 0.25° gridded EOBS V.14 to a 0.025° resolution increases the uncertainty in the observational data. The gridded Eobs data sets possess uncertainty due to their production from sparsely covered by stations (Kyselý and Plavcová, 2010). Although the height correction may have preserved most of the quality of the observations, discrepancies between can exist the measurements at the stations and the data at the nearby grid points. Thus, the EOBS data might slightly under- or overestimate the daily maximum temperatures. DWD stations data DWD Climate Data Center (CDC) (2018) are compared to the nearest EOBS V.14 grid point for twelve stations and percentiles for the summer daily maximum temperatures have been calculated. Scatter plots for the stations are displayed in the appendix (see from Fig. A.15a until A.15l). Six out of 12 stations are in good agreement for the gridded data and the station measurements. The minimum data fits quite well for Karlsruhe, while the maximum (95th percentile) was slightly underestimated. This is recognised for Rheinstetten as well, but here the deviations are even larger (> 0.45 °C). On the other hand, mountain stations like the Hornigsrinde are overestimated by over $+1.7$ °C. The distributions are slightly smaller because of the interpolation routine which produces the 0.22° grid data as expected.
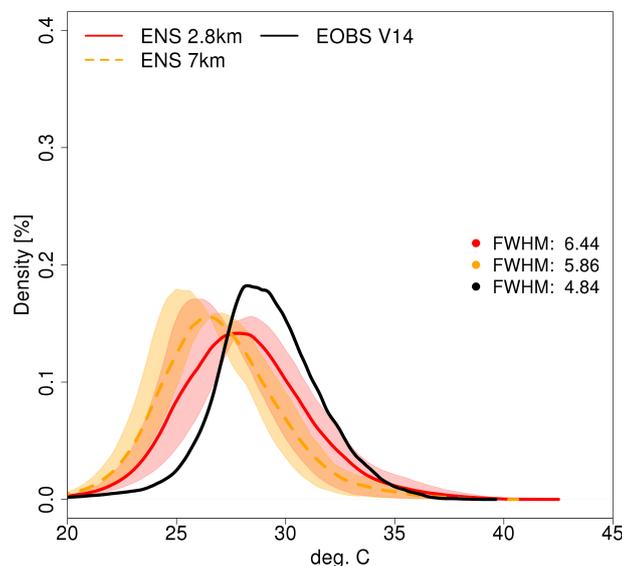
Figure 5.22.: Density Distributions for the CTRL time periods for CCLM simulations with the 0.0625° and 0.025°. The full width at half maximum (FWHM) is displayed on the right side.

5.4.1.5. Weather Patterns

The connection of the daily HW index to certain weather patterns is determined by applying the objective weather pattern classification after (Dittmann et al., 1995; Bissolli and Dittmann, 2001). The classification after Dittmann and Bissolli takes three variables into account, the wind

direction, the cyclonality and humidity, and is optimised for Germany. The wind components are computed for one layer in 700 hPa, whereas the temperature, relative humidity and the geopotential are taken into account for at least two at most five height level from 950 hPa to 300 hPa. The humidity index is calculated from the vertically integrated precipitable water (derived from temperature and relative humidity). Five letters characterise the weather pattern: first and second describe the wind direction, the third and fourth letter includes the cyclonality at 950 hPA and 550 hPa and fifth letter the humidity. A total of 40 weather patterns is derived from these variables from July 1979 onward with a daily resolution (available online at Deutsche Wetterdienst (2018e)). In addition, the connection of HWs to weather pattern in the ensemble is evaluated. The weather patterns of the simulations are calculated for the first nest (0.44° ~50 km) fitting best to the 55 km used by the DWD and due to the required area (greater than Germany).

The percentage of each weather pattern in the DWD data for days in the summer half-year is displayed from 1980 until 2000 in Fig. 5.23a. The AA pattern (anticyclonic near the ground and in the higher atmospheric levels) occurred for Southwesterly and Northwesterly wind directions with over 30 %. Thus, it is the dominant weather pattern in the summer. Additional, the AZ pattern (anticyclonic near the ground and cyclonic in the higher atmospheric levels) are found for the same wind directions with a fraction of nearly 20 %.
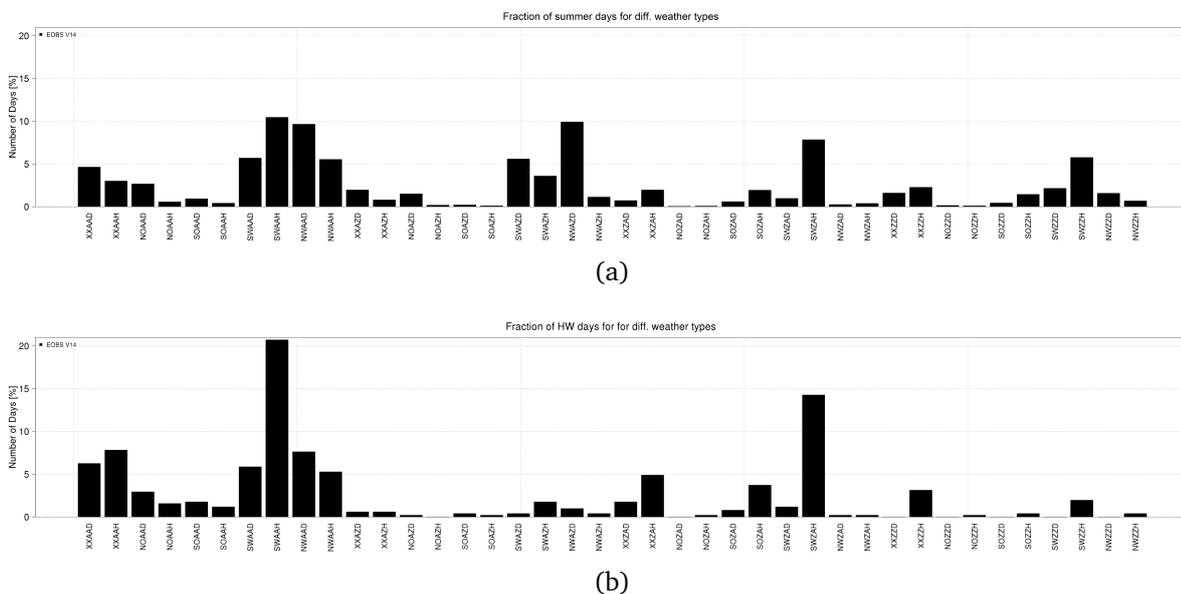
Figure 5.23.: Fractional occurrence of weather patterns at all days (a) and of the HW days (b) of the summer half year for the KLIWA region for 1980-2000 derived from the DWD data.

When the weather patterns are filtered for the HW days, the distribution reveals two prevailing patterns favouring HW conditions (Fig. 5.23b). Combined, the SWAAH and the SWZAH occur on over 35 % of the HW days. The patterns XXAAH, XXAAD, NWAAD, NWAAH and SWAAD are related to HW days with smaller percentages: 5 % < WP < 15 %. These six classes are detected

on $>$ 50 % of the HW days. Thus, a dominating influence of anticyclonic conditions in the troposphere on HW days mostly with Southwesterly winds was shown. Anticyclones enhance subsidence and transport of warm air into Germany favouring the development of HWs.

The "Wetterlagenklassen" (WLK) derived from the simulations are sorted for HW days as well. The differences to the observations are displayed as fractions filtered for wind directions in Fig. 5.24a and vorticity in Fig. 5.24b. The ensemble overestimates the AA pattern fraction noticeably ($<$ +12 %). Whereas too less ($>$ -10 %) mixed patterns with a cyclone at the surface and an anticyclone in the upper height, ZA, are produced by the models. The weather patterns derived from the models contain much more dryer fractions ($>$ 20 %) except for CHadGEM2 , which is in good agreement with the DWD (not shown). More D (dry) patterns indicate, that CEH6 , as well as CEC-EARTH, are dryer in the upper atmospheric levels than the DWD data. Concerning the wind directions, higher XX fractions are seen in all simulations ($<$ +10 %) and too less SW fractions (Fig 5.24a).

(a)

(b)
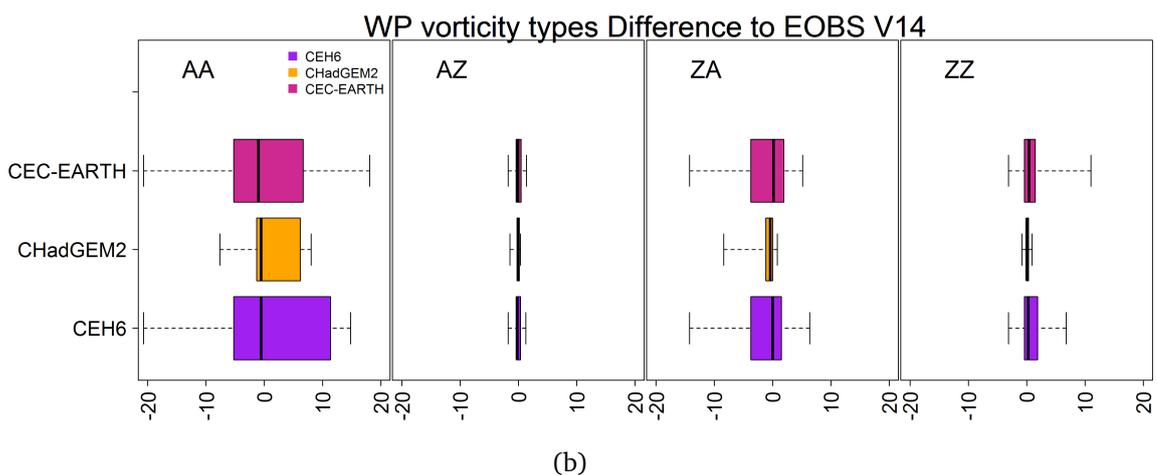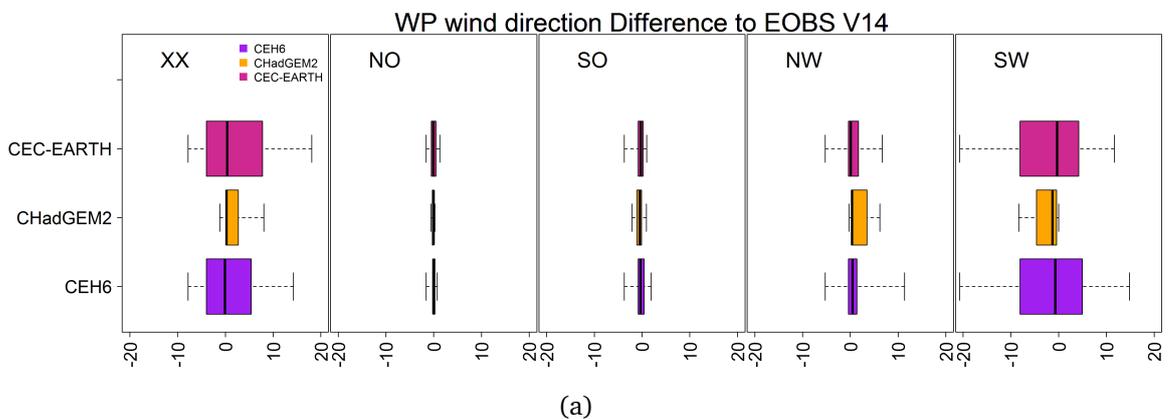
Figure 5.24.: Difference for the fractional occurrence of weather patterns in the simulations compared to EOBS (sorted by wind direction (a) and vorticity (b)) for HW days of the summer half year for the KLIWA region for 1980-2000.

In particular, the anticyclonic/anticyclonic vorticity fraction with southwesterly winds and humid in underestimated (SWAAH) in all members (< -20 %). Contrarily, the same vorticity fraction but with an ambiguous wind direction XXAA is overestimated only by CEH6 (+15 %) and CEC-EARTH (+18 %).

In general, CHadGEM2 achieves the best agreement to the observations. The similarity between CEH6 and CEC-EARTH are obvious, while CHadGEM2 produces different patterns, as seen for other variables. CHadGEM2 shows a higher number of NW weather patterns, unlike the others. van Ulden and van Oldenborgh (2006) estimated that HadGEM1 produces stronger easterly flow bringing continental air masses into Central Europe. In summer, this dryer and warmer air lead to the known warm bias in the model. However, the easterly wind fraction is not noticeably larger for the HadGEM2 used in this thesis unlike for the predecessor HadGEM1. Unknown reasons have to exist for the production of the overly large annual amplitude. It can be assumed, that the land-atmosphere feedbacks enhance the bias since strongly deviating soil conditions are detected between CHadGEM2 and the other ensemble members. Furthermore, the HW days are not connected to easterly wind directions as the observations showed. No higher number of South/Northeasterly winds are counted for CHadGEM2, yet, stronger Northwesterlies are noticed in the seasonal mean 10 m wind directions (not shown). A stronger westerly flow is found for ECHAM5 which can induce a cold bias in Central Europe because of the colder and moister air advected from the Atlantic van Ulden and van Oldenborgh (2006). Additionally, weaker wind speeds are seen in all model simulations. These winds can be associated with blocking situations when weather systems become stationary. Still, it remains unclear if this connection is simulated. On the other hand, the weaker winds could be of a westerly direction but are not identified as such because of the wind direction criterion of the WLK. This seems to be a possible explanation for CEH6 and CEC-EARTH at least.

5.4.2. Evaluation of the CTRL period - Precipitation Events

In this subsection, the precipitation distribution and extreme events including strength and durations for certain return periods are analysed for the ensemble and the observations. Fig. 5.5b and Fig. 5.5d in subsection 5.3.1.2 showed the density distributions of the daily precipitation intensities in summer and winter. The tails of the distributions display the considerably overestimated daily precipitation within the higher resolved ensemble. Therefore, extreme precipitation events are evaluated to assess the extent of this overestimation further. The tool ZR_STAT has been employed to calculate precipitation sums for different durations and return periods using a peak-over-threshold method with the 95th percentile as limit (Früh et al., 2010). Since HYRAS data are available as daily sums, only the duration of 24h is displayed in the following. Additionally, shorter durations of 12h and 1h are evaluated for the model data (not shown).

The precipitation sums are overestimated for different return periods of summer and winter half-year events in all model simulations (Table 5.1). The summer events include up to 30 % too much rainfall in the ENS mean, whereas this number is significantly lower in winter (< 5 %).

The model spread is remarkably large, due to the over the top precipitation sums in CEH6. The mean summer precipitation in CHadGEM2 is much lower than in the observations; nevertheless, intense events are overestimated even in this simulation. Thus, the mean precipitation is no indicator for the adequate simulation of extreme events. In general, the patterns for the different return periods seem stable but intensify for longer return periods.

precipitation	Mean	SD	Min	Max
Summer				
CEH6	43.0	28.1	-29.4	196.2
CHadGEM2	20.2	22.3	-34.2	158.6
CEC-EARTH	27.0	22.5	-32.6	117.6
ENS	30.1	21.8	-27.7	142.5
Winter				
CEH6	14.2	18.3	-34.8	136.5
CHadGEM2	-0.3	16.6	-47.3	115.2
CEC-EARTH	0.7	16.9	-50.5	92.0
ENS	4.9	15.4	-44.1	104.9

Table 5.1.: CTRL period: Relative differences in % between the models and HYRAS for precipitation events in the summer and winter half-year with a return period of 10 years and a duration of 24h.

The relative difference in the precipitation sums of events with a ten year return period and duration of 24 hours between the ENS mean and HYRAS are displayed in Fig. 5.25a and Fig. 5.25b. Since similar spatial patterns are identified for all models, especially in the summer half-year, only the ensemble mean difference to HYRAS is shown. Too strong events are detected for nearly the whole evaluation area in summer, except for the foothills of the Alps. In addition, local hot spots are noted in the Northern and Western areas. In winter, the difference between the observations and the models is considerably smaller with +5 % in the ENS mean. This might be a result of the generally lower precipitation amounts for the different return periods and duration intervals.

(a) SHY

(b) WHY
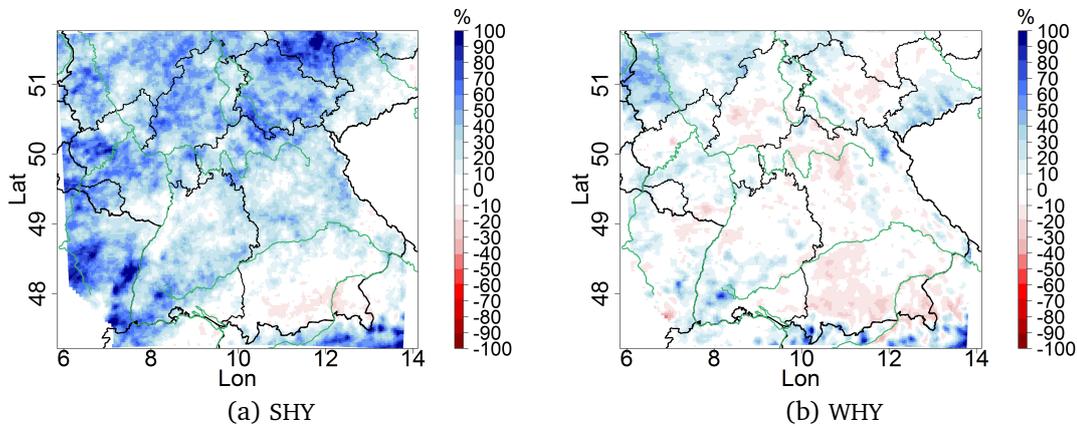
Figure 5.25.: CTRL period: Relative differences between the ensemble mean and HYRAS for precipitation events in the summer and winter half year with a return period of 10 years and a duration of 24h.

In the foothills of the Alps and parts of Northern Bavaria, the precipitation sums of the winter events are underestimated contrary to the summer season. They are overestimated only in small regions like along the Thuringian Forest or the Southern Black Forest (Fig. 5.25b). Although spatially varying, all models show slightly underestimated areas. In particular, the overestimation seen in the mean seasonal precipitation sums for mountainous regions is not found for extreme events in winter.

5.4.3. Assessing the climate change signal

In the following subsection, the difference between the CTRL period and the two future periods are evaluated for the ensemble. The HWs of the high-resolved 0.025°ensemble has been compared to the data of the lower resolved 0.0625°for the CTRL period in subsection 5.4.1.4. The result was a good agreement of mean HW length and temperatures, but an overestimation of the maximum temperatures for the severest events (mainly in CHadGEM2). The lower resolved 2nd Nest fits better to the observations for the most extreme HWs temperatures; however, it underestimates them in the average. Therefore, the 2nd Nest need to be part of the analysis of the future periods to achieve a complete assessment of the climate change signal and to avoid an overemphasis of its strength.

5.4.3.1. Heat Waves

The averaged ENS mean temperature increases for the NF by $+1.7\,°C$ and by $+4.3\,°C$ in the DF. The rise of the 90th percentile is of a similar range for the NF, while it is notably higher for the DF with $+5.47\,°C$. This leads to the conclusion, that the increase of the extremes is accelerated compared to the general rising of the mean temperature. In the following, this finding is further confirmed.

The density plots of the daily HW temperatures for the ensemble members, two resolutions and three periods are displayed in Fig. 5.26a until Fig. 5.26c. The large deviations in the HW temperatures are obvious for the three members. While CEH6 and CEC-EARTH predict only an increase of $5.5\,°C$, the HW temperatures of CHadGEM2 are $> 2\,°C$ higher. Therefore, the temperature rises by $+6.2\,°C$ in the ENS mean, which is considerably larger than for the 90th percentile. All members produce similar changing rates in both resolutions for the NF which can be seen in the values of the FWHMs. Larger FWHMs are found for the distant future indicating a broadening of the density distributions. In addition, partly significantly differences in the temperature increase rate are detected between the second and the third nest. Nearly no deviations are found between NF and DF in the FWHMs of the 0.0625°simulations in the graphics, especially for CHadGEM2 and CEC-EARTH. However, the FWHMs increases for CHadGEM2 from $6.93\,°C$ in the NF to $7.47\,°C$ in the DF.

The mean differences between the two future periods and the CTRL are shown for the five HW variables in table 5.2. To estimate the change rates, the HWs are calculated using the specific reference period of each time slice. An increase is detected for the HW temperatures (max. $<$

(a) MPI-ESM-LR

(b) EC-EARTH

(c) HadGEM2
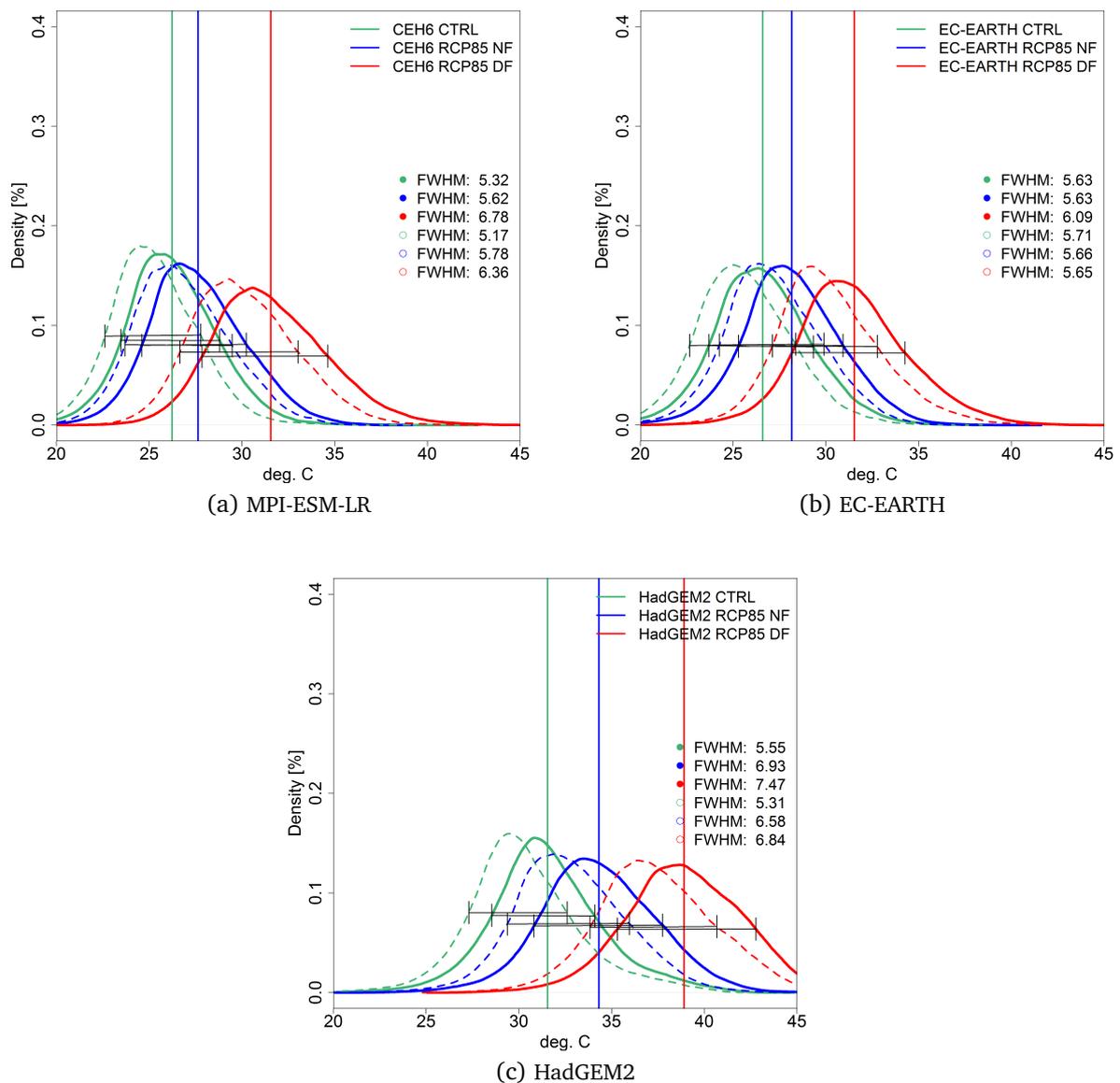
Figure 5.26.: Density Distributions for the three time periods for CCLM simulations with the 0.0625°and 0.025°. As in Fig. 5.22, the FWHM is shown on the right side.

6.6 °C in the ENS mean), while number and lengths of the HWs remain mostly stable. Since yearly and spatial averages are contained in the table, timing changes within the year or spatial shifts might be balanced or are not visible in the means. Therefore, the spatial changes and the annual cycles will be shown afterwards.

The spatial differences between the future periods and the CTRL are displayed in Fig. 5.27a and Fig. 5.27b. The temperature rises for both periods, but the strength of the signal varies locally. The strongest temperature increase occurs in several sub-regions as in the East of Bonn and Cologne or the foothills of the Alps and seems partly independent to external factors like the orography. On the other hand, a clear west–east gradient is noticeable for the DF (Fig. 5.27b) with significant differences for nearly 100 % of the GP. The lengths of the HW change

	Mean	Min	Max
mean HW temperature [°C]			
NF	2.0	1.4	2.9
DF	6.2	5.5	7.7
max. HW temperature [°C]			
NF	2.2	1.5	3.3
DF	6.7	5.9	7.9
mean HW length [days]			
NF	0.0	-0.2	0.2
DF	-0.1	-0.7	0.3
max. HW length [days]			
NF	0.0	-0.2	0.5
DF	0.0	-0.6	0.4
number of HW			
NF	0.1	0.0	0.1
DF	0.1	0.0	0.1

Table 5.2.: Changes in mean and maximum HW temperatures and lengths and number of HW per year for the 0.025° resolution between the NF and DF to CTRL period

not significantly (< 1 % of the GP show a p-value of 0.05). Nevertheless, a slight increase in the length is found for most of the domain (Fig. 5.27c) in the NF, while shorter HW of a similar range is seen in the Eastern part (Fig. 5.27d) in the DF.

The following three graphics show the annual cycle of the ENS mean and the ensemble members for the three periods. The area affected by the HW reaches earlier higher GP sums in the NF (Fig. 5.28b). In addition, HW signals of CHadGEM2 are seen about two weeks earlier although this includes a break of a few days at the end of May. The model shows a limited extent into September also. CEH6 only affects a larger area but displays no change in the onset date. A distinct shift towards a more delayed onset of the HW is seen in the DF (Fig. 5.27c). The onset defined as first occurrences of larger affected areas is seen in the mid of June instead of the mid of May. Additionally, the HWs end less than a week later as in NF. One reason for the shift towards a later onset and decay of the HW can be the shift of the annual precipitation maximum from July to May (see Fig. 5.12). Very warm years like 2003 are often coupled to lower precipitation in spring resulting in a strong soil drying in summer (Ferranti and Viterbo (2006), Vautard et al. (2007), Fischer et al. (2007b), García-Herrera et al. (2010)). Thus, the shift of the main precipitation to May and June could prevent the development of HW during this months due to the higher relative humidity (Fig. A.9a), could cover (Fig. A.9c) and the reduction of intense incoming radiation (Fig. A.9b). The reason for the shift of the precipitation

(a) Max. HW temp. NF-CTRL

(b) Max. HW temp. DF-CTRL

(c) Max. HW length NF-CTRL

(d) Max. HW length DF-CTRL
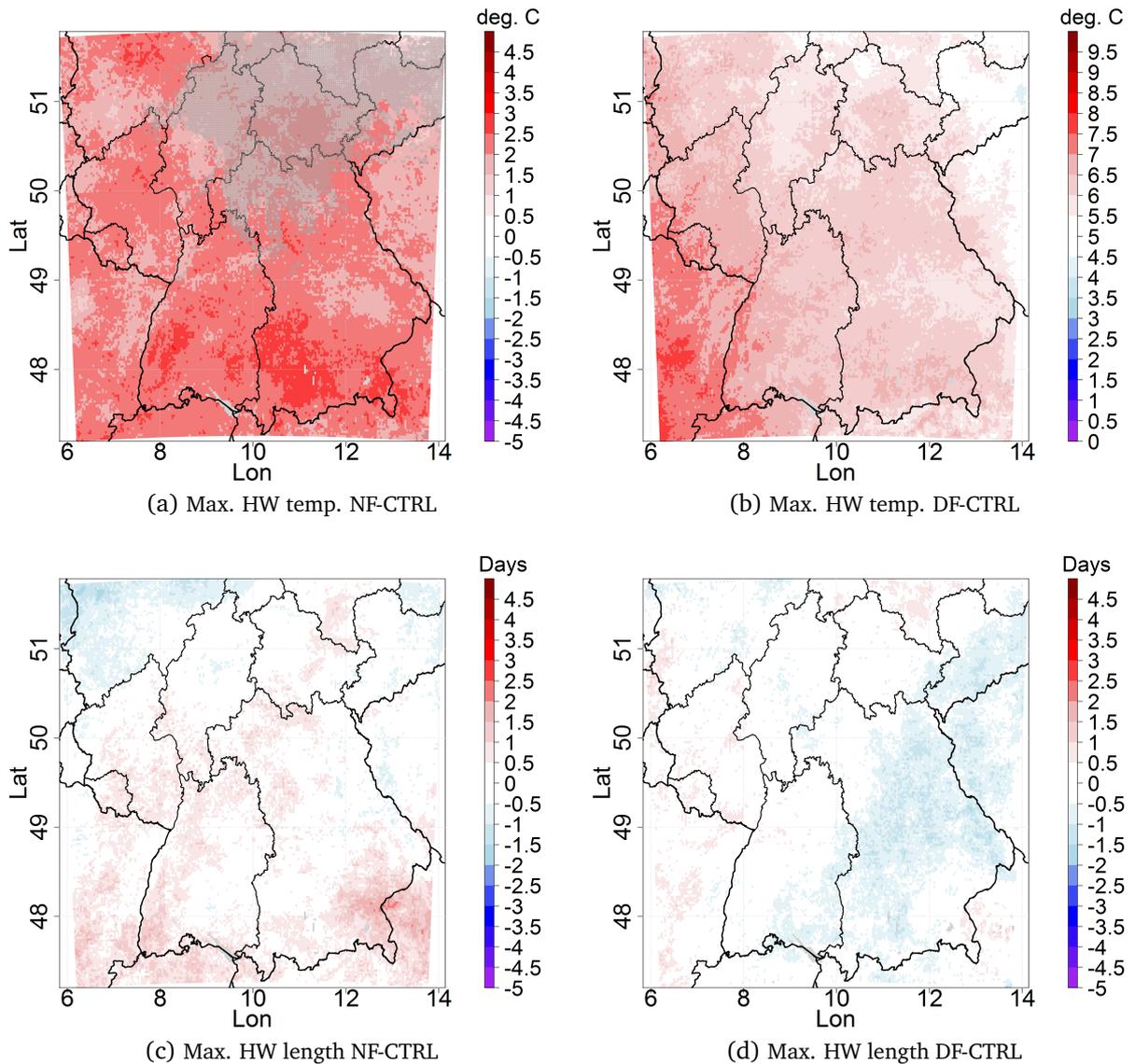
Figure 5.27.: HW max. temperature and lengths differences between the future periods and the CTRL for the 0.025°ENS mean. Grey shaded areas show significant changes as before (see e.g. Fig. 5.10d).

maximum can merely be suspected. One option could be a change in the general circulation and the weather patterns associated with intense rainfall events due to the changing land-sea temperature contrast. This theory remains to be proven for the DF and is only one estimation as Sedlmeier (2015) observed no difference for the weather types in the near future. No obvious changes in the weather patterns seem reasonable regarding the similarities between CTRL and NF. Nevertheless, it is doubtful, that the occurrence of weather pattern continues in the same fashion for the DF.

The absolute strength and the impact of HWs on health and economy are dependent on the level of the 90th percentile relative to their location. HWs in Scandinavia are less severe for human health than ones in Spain. The same is applicable concerning larger temporal variations of the 90th percentile. Zhang et al. (2005) investigated the inhomogeneity due to percentile variability
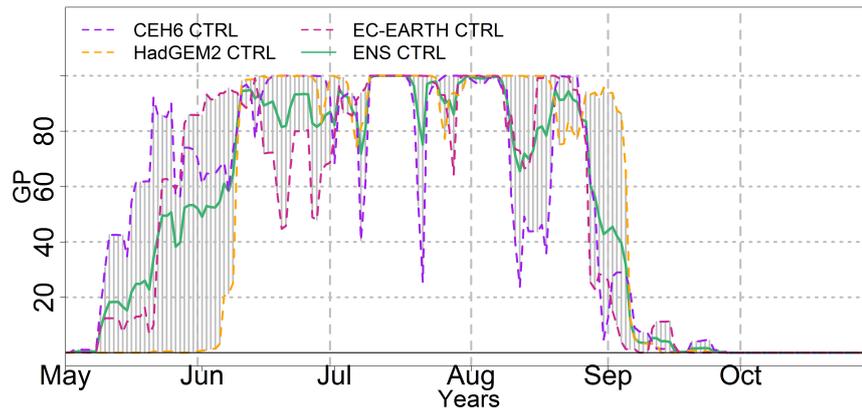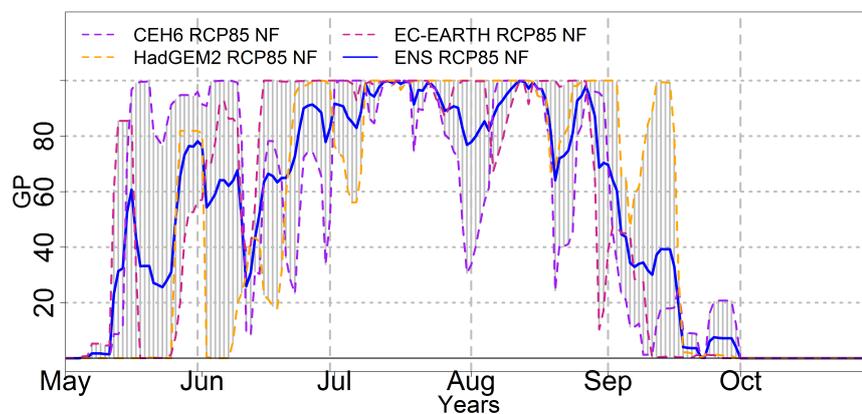
(a) CTRL

(b) NF
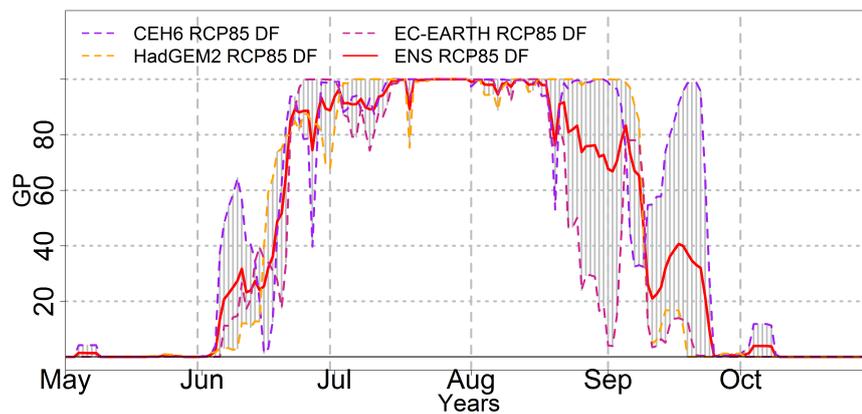
(c) DF

Figure 5.28.: Annual cycle of grid points sums of heat wave events for the KLIWA area CTRL (green line), NF (blue line) and DF (red line); shown are the ensemble members (orange, magenta, purple lines) with the ensemble spread (grey area), the ENS mean (solid line).

and demonstrated ways to reduce them. However, these methods concern the calculation of percentiles based on a moving daily window. They differ clearly from the climatological percentile used in this study. Therefore, the HW of the future periods will be calculated with the 90th

percentile of the CTRL period to assess the "jump" in the percentile. There exists doubtlessly such an inhomogeneity in the percentile due to the large temperature increase.

Fig. A.17a and Fig. A.17b in the appendix show the annual cycle with the percentile of the CTRL period. The HWs affected area reaches earlier 100 % of the GP and encompasses nearly the whole summer half year for NF and DF. In the DF, HWs are detected even earlier than May and extend to the mid of October, which was not seen graphics with the relative percentile (Fig. 5.28c). Combining the two findings of the annual cycle of HWs, it can be noted, that the most intense HW will occur later in the year. However, heat stress will probably affect the population nearly during the whole summer half-year. It is likely for intense HWs to arise in late spring/earlier summer in DF due to the temperature increase, which occurred only in the main summer month in the CTRL. The strongest HWs in DF will possibly develop later in summer than in the CTRL. The differences of HWs between NF/DF and CTRL calculated with the CTRL reference period are summarised in table A.6 in the appendix. The maximum HW temperature shows no deviation between both tables, which was expected since both approaches capture the most intense day. The increase in the mean HW temperature is lower than before (DF: $+2.3$ °C $< +6.2$ °C), as a consequence of including days, which would not be counted as a HW day in the NF and DF before. The mean and the maximum lengths increase strongly with $+16.6$ days in the maximum in the DF ENS mean.

5.4.3.2. Precipitation Events

Precipitation signal changes in the future are very variable and differ strongly for seasons and regions. The differences to the recent period are summarised as followed in the IPCC (2013): Wet regions will become wetter, while dry regions will get dryer. For Europe, trends are less clear in the continental regions, whereas Southern Europe will likely experience a decrease in precipitation and Northern Europe an increase. The density distributions of the daily precipitation sums are displayed for the ensemble members (2nd and 3rd Nest) for summer and winter of the three time-slices in Fig 5.29a until Fig. 5.29f. Higher daily sums are detected in the summer half increasing for the NF and DF in all members and resolutions. A continuous rising of precipitation is found for CHadGEM2 (Median CTRL: 1.76 mm/day, NF: 1.93 mm/day, DF: 2.09 mm/day), while CEH6 (Median CTRL: 2.53 mm/day, NF: 2.61 mm/day, DF: 2.34 mm/day) and CEC-EARTH (Median CTRL: 2.14 mm/day, NF: 2.11 mm/day, DF: 2.08 mm/day) show a slight decrease in DF.

In addition, the climate change signal differs for precipitation events depending on their intensity (heads of the density distributions). While the mean precipitation sum of the summer half-year decreases, higher extreme values are noted in all models and both future periods (Fig. 5.29a to Fig. 5.29e). In winter, the daily precipitation sums are significantly smaller than in summer; however, the mean change signal is more distinct and larger in winter. The median of the summer ENS mean changes by $+0.05$ mm/day in NF and -0.02 mm/day in DF, whereas the winter daily precipitation increases by $+0.17$ mm/day in NF and $+0.4$ mm/day in DF. The extreme high daily precipitation reveals notable differences between the ensemble members.

(a) CEH6 Sum

(b) CEH6 Win

(c) CEC-EARTH Sum

(d) CEC-EARTH Win

(e) CHadGEM2 Sum

(f) CHadGEM2 Win
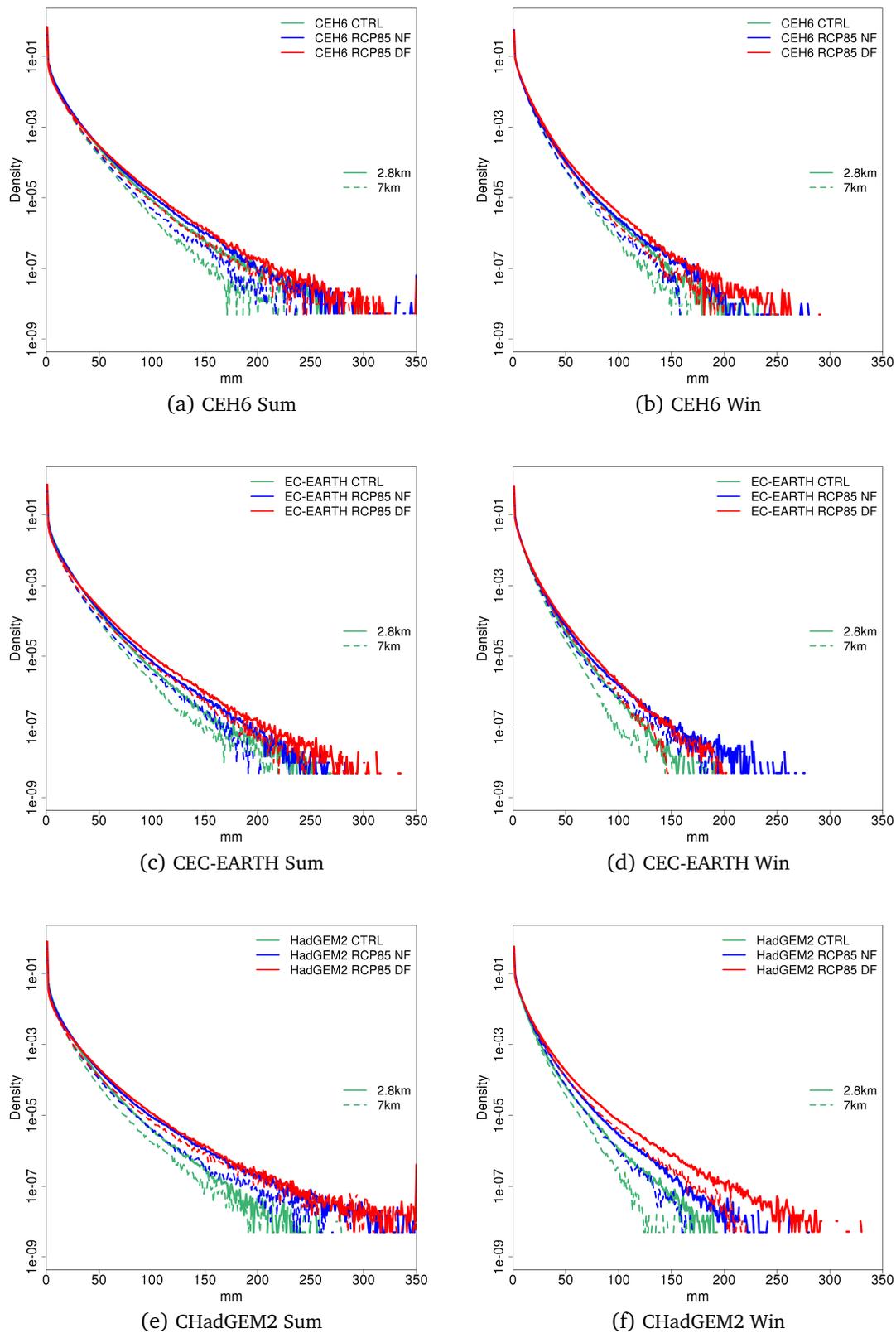
Figure 5.29.: Density Distributions of daily precipitation sums for the three periods for CCLM simulations with the 0.0625°(dotted lines) and 0.025°(solid lines).

The strongest increase are seen in CHadGEM2 (Fig. 5.29f), while CEH6 shows a similar increase as in the summer (Fig. 5.29b). The extreme precipitation sums of CEC-EARTH are notably reduced for the DF (Fig. 5.29c)), although the median of the daily precipitation sums increases. With this temporal development, CEC-EARTH deviates strongly from the other two members. Additional comparison with the 2nd Nest revealed lower increase rates of the median in the summer half than the 3rd Nest.

The changes in the density distributions fit well to the difference of precipitation events calculated via ZR_STAT. With increasing return period, the 12 hour and 24-hour events include up to > 29/40 % more rainfall in the NF/DF compared to the CTRL. As an example, the precipitation sums of daily events with a ten-year return period are shown in Fig. 5.30a until Fig. 5.30d. Although the ENS mean displays a general increase, considerable spatial variations are found highlighting the varying strength of the change signal. This result confirms that intense precipitation events become stronger even if the mean precipitation generally shows a reduction. For weaker events with shorter return periods (duration of one hour with a return period of 1 until 10 years), a reduction of the precipitation sums (1 year: < -40 % SHY and 10 year: < -20 % Year) is noticed (not shown).

(a) SHY NF - CTRL

(b) WHY NF - CTRL

(c) SHY DF - CTRL

(d) WHY DF - CTRL
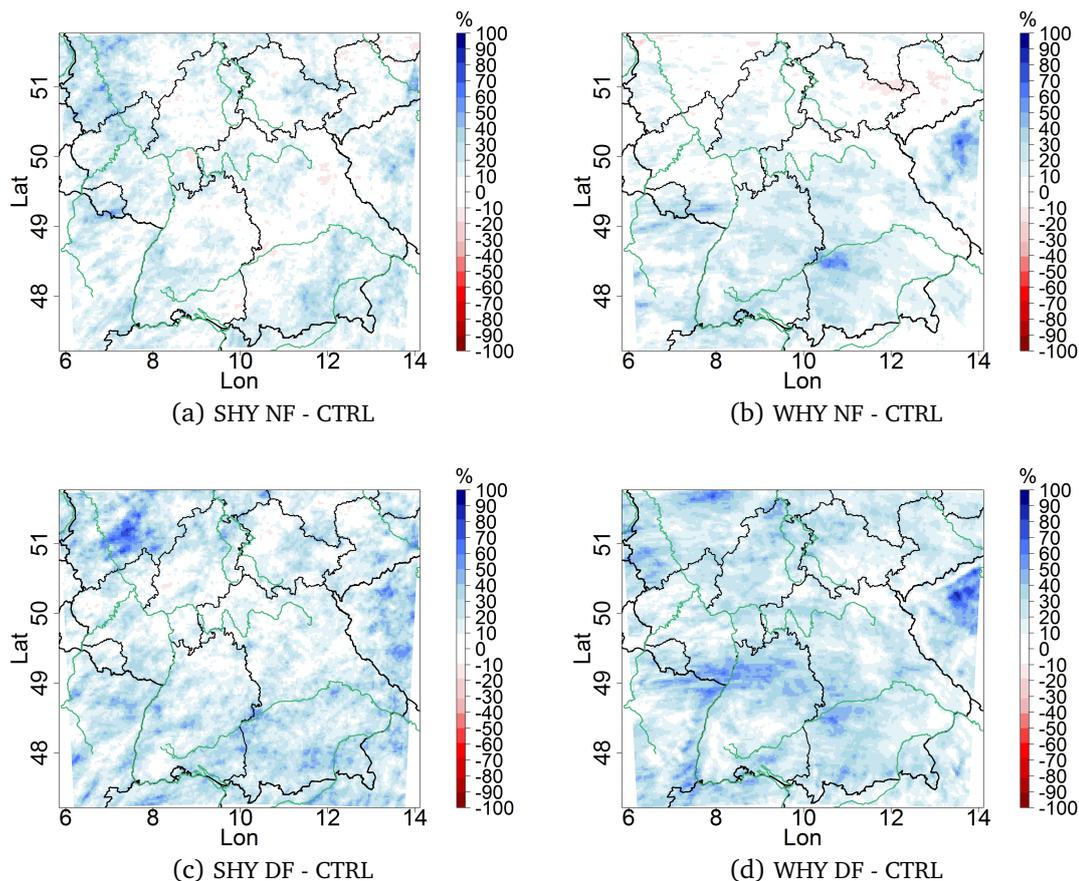
Figure 5.30.: Relative differences between the ENS mean of the NF/DF to the CTRL period for precipitation events in the summer and winter half year with a return period of 10 years and a duration of 24h.

Table A.7 in the appendix shows the changes in the precipitation-based climate indices. A clear trend towards a higher number of high precipitation events is detected in contrast to rising numbers of dry days. In particular, intense precipitation events in the summer-time are frequently associated with fast developing thunderstorms due to the strong convection and are difficult to predict. Their resolution of a few hundred meters up to a few kilometres (with this they encompass only a minimum number of GP due to the model resolution) makes it often nearly impossible to be resolved correctly in a numerical weather prediction (resolution of the COSMO-D2: 2.2 km, Deutsche Wetterdienst (2018c)).

> **Short summary**
> **Current climate conditions:**
> The 3rd Nest ENS mean has a cold bias compared to the Eobs V.14 observations for the mean HW temperatures (-1.4 °C). The cold bias is noticeably higher for CEH6 (-3.2 °C) and CEC-EARTH (-3.0 °C) , while CHadGEM2 has a pronounced warm bias (+2.2 °C). The mean HW lengths of the ENS agree well with the observations (+0.1 days) but also for the members (range: -0.2 to +0.8 days). The 2nd Nest is too cold in the mean (-2 °C), while the most extreme HW temperatures show lower deviations to the observations than in the 3rd Nest. Therefore, the assessment of the 3rd vs 2nd Nest is partly inconclusive. Pronounced spatial differences are observed for the HW temperatures, which mainly arise from orographic deviations. Skill scores proved the ensemble approach advantageously compared to the members. An added value of the 3rd Nest could be demonstrated for HW temperatures for significant parts of the domain as well. The onset of the HW often occurs in the models in June later than in the observations, where the first phase in May was seen. The cold bias of CEC-EARTH and CEH6 and the warm bias of CHadGEM2 seemed partly explained by the large variations in the distribution of weather patterns. Overall, fewer circulation patterns with southwesterly winds and anticyclonic conditions in the middle troposphere and at the surface are found in the model results (up to -20 %) compared to the observations, whereas CHadGEM2 overestimates patterns with northwesterly winds marginally. Although CHadGEM2 underestimates the mean summer rainfall, intense precipitation events are mostly overestimated by all models in the ZR_STAT results (10 year return period of daily events: summer: +30.1 %, winter: +4.9 %). Large spatial differences are noted with hot spots of higher precipitation often overestimated in the models in summer.
> On the other hand, winter events are partly underestimated and have a better agreement with the observations. However, the winter precipitation is noticeably too high in the model simulations. Thus, the averaged precipitation sums are no indicator of the correct simulation of extreme precipitation events. In addition, connections are recognised between topographic structures and the largely overestimated precipitation.

> **Projection for two future periods:**
> The future periods show a moderate rising of the mean HW temperatures for the NF (+2.0 °C) and a steep increase in the DF (+6.2 °C), especially in CHadGEM2(+7.7 °C). Thus, the HW temperature rise by > +6 °C in the DF exceed the mean temperature change by far (+4.2 °C). Longer (+8 days in the mean length) and more HWs per year (+2 HW per year) are noticed relative to the CTRL periods. The time interval with most intense HWs (temperature and affected area) shifts to later in summer, although the heat stress becomes more severe for the whole summer half-year because of the overall increasing temperatures. A decrease of the precipitation was noted in the DF summer months (section 5.3.2.2), contrary to this; heavy precipitation events intensify (+ 40 % higher precipitation sums in DF summer). A shift is identified towards more extended dry periods reducing the mean precipitation with more severe extreme events. The increase of precipitation is very heterogeneous over the evaluation domain since most of the events in the summertime are of a convective origin. Thus, the accuracy of change pattern is only an estimation due to the difficulty to predict small-scale intense precipitation events correctly.

5.5. Chapter discussion and conclusions

In this chapter, a new very high resolved ensemble of climate projections has been evaluated for selected variables with focus on temperature extremes. The ensemble is built by downscaling three GCMs to the final resolution of 0.025°. The ensemble has been compared to observations for the CTRL period 1971-2000, while for two future periods (2021-2050 and 2071-2100) changes for the selected variables were analysed. As known from other studies, the correlation between climate projections and observations is often poor, and skill is derived due to the general climate trend (chapter 4 and Kruschke et al. (2016)). Evident improvements for the higher resolution are expected due to an adequate representation of the orography and better resolved small scale processes like the diurnal precipitation cycle.

In the CRTL period, the 0.025° ENS was able to capture the 2m temperature well (-0.4 °C) and to reduce the cold bias from the 0.0625° ENS by-0.8 °C. An added value was found for the temperature with the increase of the resolution for all seasons. Although the second nest was height corrected for this comparison, a Wilcoxon test demonstrated the significant improvement for the higher resolution for all height levels. On the other hand, the precipitation bias was considerably increased with the higher resolution from +23 mm ENS 0.0625° to +151 mm ENS 0.025°in the annual sums, and no clear added value was detected. The analyses of the density distribution showed the ENS 0.0625° has the highest agreement for daily precipitation intensities of all nesting steps.
This result contradicts Hackenbruch et al. (2016), who noticed at least a decreasing precipitation error for the 0.025° simulation compared to the 0.0625° run for grid points < 400 m. Contrary to the GCM driven simulations in this study, Hackenbruch et al. (2016) used ERA40

reanalysis data as forcing which improved the model results largely being already dryer and warmer. Overall, precipitation sums increase for the third nest, in particular, the ENS 0.025° is too wet (+90 mm) and overconfident in the winter season. The overestimation of the precipitation occurs mostly over mountain peaks in this season (+378 mm), while no direct connection to the orography was recognised in the error patterns in summer. Furthermore, the ENS produces a massive spread in the summer season (range from 408 mm to 673 mm in the season sum). CHadGEM2 displays a pronounced dry bias in the summertime (+77 mm) along with a large positive temperature bias (+1.0 °C). Whereas CEC-EARTH is too cold (-1.4 °C) and slightly too wet (+74 mm), while CEH6 has a marginal cold (-0.7 °C) and pronounced wet bias (+188 mm).

van Ulden and van Oldenborgh (2006) tracked the deviations in the model biases back to systematically different wind fields and found stronger easterly winds as a possible source for the warm bias of the HadGEM1 in the early summertime. In contrast, ECHAM5 displays too strong westerly flows leading to the cold and wet bias in summer. In this thesis, westerly winds are found for CEH6 as well, but no reversal to easterly winds in May or June is noticed for CHadGEM2. There are some possible reasons for this, for one, the model development from HadGEM1 to HadGEM2-ES. It can be assumed, this has some effect on the large-scale circulation (Martin et al., 2011; Collins et al., 2011; Perez et al., 2014). Secondly, the testing procedure and the evaluation domain are notably different between van Ulden and van Oldenborgh (2006) and this study. The larger annual temperature amplitude is likely a result of the different soil conditions in CHadGEM2 as this simulation has a much dryer soil compared to the other ensemble members. In seen in chapter 3, the dryer soil accounts for warmer summer and colder winter soil temperatures as the soil heat conductivity is altered due to the lower soil water content.

The improved temperature bias is explained as follows: The increased temperatures in the convection permitting simulations are produced due to a reduction of the soil moisture. The dryer soils are a consequence of the reduction of frequent drizzle (compared to the parametrised scale Ban et al., 2014). Lesser rewetting of the soil layers decreases the evapotranspiration and in such the Bowen ratio is shifted. Ban et al. (2014) found higher sensible heat fluxes in the high-resolved simulation and a reduced cloud cover. This mechanism, it was concluded, lead to higher 2m temperatures and in case of the GCM forced simulations to a reduced temperature bias. However, no complete explanation exists, why the precipitation is overestimated largely in the cloud-resolving ensemble. However, an added value is detected on sub-daily scale (Kendon et al., 2012; Ban et al., 2014; Chan et al., 2014; Kendon et al., 2017). The comparison of the relative humidity reflects the results from the precipitation. The winter humidity is overestimated (+3.1 %), while the ENS means underestimates the rel. humidity in summer (-3.4 %) due to the large spread (CHadGEM2: -11.3 %, CEH6 : +3.3 %). All things considered, the ensemble members produced the expected spread and similar results as in the study of Keuler et al. (2016) for the 12 km CORDEX ensemble. Thus, the ensemble showed its added value as it produces a

robust and mostly reliable signal with minor exceptions.

The large ensemble spread is reflected in the climate change signals. The change rate nearly triples for the temperature between the NF and DF in the ENS mean. Compared to the more considerable temperature increase in the CHadGEM2 (+6.0 °C), the other members indicate a more moderate increase of the annual mean temperatures (CEH6: 3.0 °C, CEC-EARTH: 3.6 °C). Overall, the tendency of the single models to under/overestimate a variable and the relations between the ensemble members match the findings of Mauritzen et al. (2017). Models with a higher climate sensitivity tend to have a more significant temperature change signal like the CHadGEM2, whereas the CEC-EARTH and CEH6 show lower climate sensitivity and the smaller change signal. Thus, the selection of these three particular GCMs results a reliable ENS mean, although the mixing of models with different climate sensitivities is seen critically in determining the size of climate change signals (Mauritzen et al., 2017). Unlike the mean temperature and mean precipitation, heavy precipitation events are not strictly dependent on climate sensitivity as the comparison for the climate indices and certain return periods have shown.
Notably, the shift of the precipitation maximum towards earlier months speeds up the drying-out of the soil during the summertime, which is accompanied by higher sensible and lower heat fluxes, lower relative humidity and cloud cover to enhanced temperatures and incoming radiation. Ambiguous for the NF, the DF differences of temperature and precipitation to the CTRL are significant for nearly the whole evaluation area and imprint on the climate indices calculated from these variables. A widespread drying is found for Germany in the summertime, whereas the winter precipitation increases (similar results are observed in studies like Seneviratne et al., 2006b; Christensen and Christensen, 2007; IPCC, 2013; Scoccimarro et al., 2015; Ruosteenoja et al., 2018). With the general warming, warm and hot days and heat waves increase in magnitude and number. The quantity of events with smaller rain amounts decreases, whereas heavy precipitation (daily and sub-daily) events intensify (higher number and stronger events, this study but also Fosser et al., 2017; Kendon et al., 2014).

Heat waves and their change signal are investigated for Southern Germany in addition to the climate indices. The definition with a percentile based threshold is one of the advantages of the chosen HW definition. The mean bias of each ENS member varies some HW indices dramatically as seen from the temperature based climate indices. As an alternative, the HW could be based on a temperature anomaly or be defined as a combination of the daily maximum and minimum temperature or be related to the relative humidity. In particular, indices assessing the heat stress on the population or combined indices as the HUMIDEX are missing in this evaluation as well as the connection to mortality rates. As known from Muthers et al. (2017), HWs with lower maximum temperatures but increased humidity can have similar consequences as dryer but hotter HWs. The proposed definition is usable all over Europe and overcomes the problem with region-specific HW definitions. In addition, the usage of the maximum heat temperature is advantageous as it can be used as a possible predictor for heat strokes and can be connected

to mortality rates (Honda et al., 2007; Ono, 2013; Muthers et al., 2017). With HW indices like the mean and maximum HW temperature, duration, area and global intensity an event set was derived for the observations.

The 0.025° ENS mean has a better agreement for HW temperatures with the observations due to the lower the summer cold bias. Thus, general added values are found for the very high-resolution (bubble graphic after Ouzeau et al., 2016). A gap between average HWs and mega-heatwaves was identified for the smaller Southern Germany area, as it was seen for the European evaluation. This is an advantage of the event analysis based on the chosen HW definition in this study. Most of the HW days are associated with anticyclonic and southwesterly weather patterns as an analysis of weather patterns defined after Dittmann et al. (1995) showed. Similar results were determined for hot events from the study of Sedlmeier (2015) and for HW frequency in Kyselý (2008)(Central European high and ridge). Higher numbers of the XX (unspecified wind direction) are found in the simulations. A possible reason for this can be the increasing variety of wind directions as a result of the lower wind speeds occurring in the model runs. Thus, the threshold defining a prevailing wind direction may not as easily reached with the model data and a higher number of XX days (undefined wind direction) are noted. Lower wind speeds can also be connected to the centre area of large-scale structures like blocking or similar stationary systems enhancing the occurrence of HWs. If models overestimate these (duration or strength) or underestimate the wind speeds within these systems, also a higher number of XX days can occur.

Wind directions and vorticity of the ENS members follow the already known similarities and differences with similar results for CEC-EARTH and CEH6 and deviations for CHadGEM2. The main difference is a higher number of NWs weather patterns in CHadGEM2. Evaluations for the 10 m seasonal mean wind directions showed a stronger Northwesterly flow in CHadGEM2 indicating higher pressure possibly due to a strong Atlantic ridge in Southwestern Europe extending further into Central Europe (Dittmann et al., 1995).

van Ulden and van Oldenborgh (2006) also assumed positive feedback due to the pronounced soil desiccation in the HadGEM1, which may not exclusively exists in future changes but to a certain degree to the CTRL period. The soil drying is accompanied with a reduction of precipitation and an increase of the temperature due to the reduced evapotranspiration and alters the land-sea-contrast which in turn leads to changes in the circulation patterns and stronger continental winds. A similar assumption was introduced by Lhotka et al. (2018), they proposed that lower evaporation rates and lower soil moisture amplify HWs in CHadGEM2 runs further. This can be confirmed for the models in this study as lower latent and by far higher sensible heat fluxes and lower soil water content are detected for CHadGEM2. However, circulation changes could not be investigated further. Similar to the discrepancies between the ensemble members for the temperature, a tendency towards longer mean HWs lengths is noticed for the CHadGEM2, while the other two show nearly equal values. On the other hand, lengths of single events are often

underestimated by the models, which might be a result of the simulation of blocking events connected to HWs. If the blocking duration or frequency is underestimated as it is frequently noted for the driving GCMs (Scaife et al., 2010), the weak HW events might be the result.

In the two future periods, two approaches have been followed to assess the change to the CTRL. Depending on the research question, the 90th percentile from the CTRL period or the respective percentile of NF or DF has been used. Implementing the 90th percentile of the CTRL period, the time for which HW conditions occur extends nearly to the whole summer half year. With a mean HW temperature increase up to +2.3 °C in the DF in the ENS mean, the increase of this variable is even lower than the rise of the averaged 2m temperature (+4.6 °C). This is not overly surprising since a large part of the general temperature increase is related to the rise of the winter temperatures. Furthermore, the summer half-year mean as it is employed here includes May, September and October, months for which the temperature increase is not as large and which are included in the mean difference as the HW periods expand. With this percentile, the number of HWs per year doubles and an even stronger increase is found for the mean and maximum HW lengths.

As the respective percentile is chosen, no increase is noticed in the number as well as the lengths. Compared to the length, a more exact range of the temperature change signal can be determined, if this percentile is applied. The temperature increase triples up to +6.2 °C relative to CTRL and exceeds the changes seen for the mean 2 m temperatures. Similar change rates are found for other studies like Orlowsky and Seneviratne (2012), who noted a factor of two between extreme quantile changes over land and the global annual mean average increase (depending on the season and region). In addition, an apparent temporal shift is detected towards later summer months. The precipitation maximum is located two months earlier in the DF than in the CTRL period, possibly due to changes in the large-scale circulation. Thus, longer into the summer soil water is available and major HWs occur later at the end of July and August. Whereas in the CTRL simulation earlier soil moisture depletion in May and June can lead to HWs before August if in the particular year also a lack of precipitation coincides. As seen in the earlier chapter 3, in particular, the availability of soil water for evapotranspiration can exert a crucial effect on summer temperatures due to the reduced cooling effect and the stronger rising of the soil temperatures.

As mentioned in the main introduction, many studies found an increase in HW temperatures, lengths and numbers, which confirm the results of this study. The general increase is likely connected to shifts in the land-atmosphere feedback (as noted in chapter 3 and the main introduction, mainly (Seneviratne et al., 2006b; Fischer and Schär, 2010)) and more persistent atmospheric patterns producing an increase of severity and HW duration (Kyselý, 2008; Trouet et al., 2018). Blockings will play a significant role in the future climate as their frequency increases in the GCM simulations for the CMIP5 models in summer (Cattiaux et al., 2013). Additionally,

Dunn-Sigouin and Son (2013) revealed a slight increase in blocking events in Eastern Europe and Russia (Ural blocking). (Francis and Vavrus, 2012; Francis and Skific, 2015) proposed a mechanism leading to more persistent blocking events. As a result of the Arctic amplification, the Rossby waves become stationary with increased amplitude. A slower progression of these weaves due to a lower zonal winds velocity, as it occurs when the gradient between the Arctic and the tropics weakens, can enhance the chance of more and longer extreme events including HWs (Coumou et al., 2014).

This matches the occurring increase of heavy precipitation events. This study found little change in the annual sums, but a decrease of summer rainfall and an increase of winter precipitation. The higher winter precipitation is possibly a product of the warmer atmosphere, which can contain more water (Clausius-Claperon rate, Trenberth et al. (2003)). A higher number of dry days is discovered in the DF, along with more heavy precipitation events. The reduction of low intense precipitation events, as well as the increase of the very intense events, is known from previous studies like Schär et al. (2016), although these ensembles often lack the very high-resolution similar trends are found. The strongest de-/increase was seen in mountainous regions, which might lead to stronger flooding events for hillside regions. The regional patterns derived for different return periods also display very heterogeneous signals particularly in the summertime, when these events are connected to convection. In the winter half year, the change signal seems more zonal leading to the conclusion, that it might be connected to changes in the large-scale circulation.
Similar results are stated in other studies like (IPCC, 2013; Lehmann et al., 2015; Berg et al., 2013), who demonstrated that primarily intensifying convection causes extreme precipitation events as it responds stronger to the temperature rise. Another possible reason could be a higher number of blocking events during the winter time. Blocking events can lead to heavy precipitation at their flanks (Lau and Kim, 2012) and are frequently associated with NAO- phases in the winter months and a wavier jet stream (Woollings et al., 2008; Luo, 2005). As Francis and Vavrus (2012) noted, the wavier jet stream enhances the chance of cold spells and prolonged precipitation as it increases the persistence of weather patterns.

6. Discussion and Conclusions

The HWs of the last 20 years were the most severe ever accounted for hot summer temperature extremes affecting health, economy and the environment. It is very likely, that this trend will continue and potentially accelerate for the coming decades. Therefore, the predictability of HWs, the factors impacting them and their possible future change were investigated in this study. In this last chapter, the research questions defined in the introduction are answered summarising the conclusions drawn from the three main chapters.

How large is the impact of the soil initialisation on summer temperatures and with this on HWs? The experimental setup in this study led to the conclusion that changes in the soil moisture and soil temperature fields can be found even after a decade depending on the climatologic conditions. The impact on the 2 m temperature becomes smaller in time for most of Europe (less < 3 years), but it remains detectable in transition zones like the Iberian Peninsula up to a decade. The decay time, which the soil needs to balance itself after the new initialisation, is reduced with increasing soil moisture content. For example, no noticeable 2 m temperature difference is found in Scandinavia after three years. Although large deviations exist in the deeper soil level, they are not transferred into the atmosphere. The evapotranspiration depends more on the radiation than on the water availability for regions in high latitudes. In contrast, the soil moisture availability is crucial for the evapotranspiration in transition zones and the reason, why a noticeable drift of the 2 m temperature is detected in Iberia even after three years. Furthermore, the ratio of latent and sensible heat fluxes is shifted towards a higher sensible heat flux with decreasing soil moisture. Thus, more energy is available to heat the surface near the atmosphere. In addition, lower amounts of clouds are noticed and higher soil temperatures. Therefore, the initial fields do not only have an impact on seasonal to decadal predictions but affect climate projections to a certain degree depending on the region and magnitude of the deviation from the balance level of the LSM.

Is there a decadal signal for HWs and is it possible to predict this with decadal hindcasts? The definition employed in this thesis is based on the daily maximum temperature. A predominant decadal signal is found for the HW temperature indices for Europe, but not for the lengths or the numbers per year. In the hindcasts, yearly mean or maximum temperatures and lengths are quite well reproduced by the models. However, a pronounced bias in Northern and Southern Europe is found in the temperatures. A newly defined user variable, continuous and cumulative, showed high correlations between the RCM b1 simulations and the observations. This implies a higher predictability for the newly defined variable. The skill of the decadal hindcasts is pri-

marily connected to the climate trend, and no apparent improvement due to the yearly starting dates is seen. An added value of the RCM compared to the GCM is noted only for the HW temperatures or annual anomalies. Multi-year means are state-of-the-art for the evaluation of decadal hindcast. Nevertheless, they proved inept for the comparison of HWs on an event basis, when more than the general trend is to be analysed. The single HWs from the lead-times and the ensemble members are superimposed, and the variability is lost.

To what degree exists a dependence between large climate anomalies and HWs and is it possible to simulate this dependence with climate models? The annual mean HW temperatures values are closely related to the AMO signal as the analysis showed. The connection between monthly HW day sums and LCA pattern is determined for the Eobs V.14 data and two reanalyses driven simulations. HW days sums and the LCA pattern displayed considerably spatial heterogeneous signals and a phase dependency. For example, a higher number of HW days was detected for the positive phase of the SCAND pattern in Scandinavia, while a notably lower number of HW days occurred on the Iberian peninsula. Correlation values are increased after filtering into the upper and lower third of the pattern index. However, the correlation means were blurred as opposing signs are frequently included in the average for the same evaluation area. The signal strength is asymmetrical between the phases, often higher negative correlations are noted, but no analogue positive correlations in the opposing phase are observed. Depending on the LCA, the phase and the region, enhancement or suppression of HWs dominates the signal. In most cases, the connection between the LCA and HWs is well represented in the reanalysis data driven model simulations, although variations in strength and spatial extent exist. However, it is not possible to reproduce the patterns correctly with the GCM MPI-ESM-LR b1. Merely the EA pattern seemed to be simulated adequately with moderate to high correlations.

What resolution is needed to produce reasonable projections of HWs for smaller evaluation areas and to what extent change the HWs in the Future?
The existing cold bias of the current climate conditions in the ensemble means of the 0.0625° nest is mostly reduced in the high-resolved 0.025° nest. However, existing warm biases are further increased as in CHadGEM2. Furthermore, CHadGEM2 displays a considerably larger annual amplitude and therefore overestimated HW temperatures and lengths. In general, the high-resolution ensemble is in good agreement with the from Eobs V.14 derived HW temperatures and lengths. The maximum daily HW temperatures are slightly overestimated high-intense days in the third nest. This overestimation is i) a result of the larger variability compared to the second nest and the observations, ii) due to marginally underestimated observations as a consequence of the interpolation method and iii) because of the large spread between the models. The broader density distribution of the daily HW temperatures is detected for the two future periods and is possibly related to the different climate sensitivity of the models as well. Regarding future climate projections, HW conditions are identified for nearly the whole summer half-year, whereas the severest events are restricted to late July and August. This shift seems to be a result

of circulation changes and/or changes in the water cycle and is noticeable in several variables like the precipitation sums, soil water content, heat fluxes and cloud cover.

The holistic approach followed in this study enabled to capture the manifold characteristics of European HWs. Not only were the occurrence and characteristics of HWs statistically quantified, but also the tremendous impact of soil conditions was discussed in regard of their HW development and strength. The different time and spatial scales included in this study permitted substantial advances to our understanding of HW events. In particular, the presented analyse of the future change of these events on the cloud-resolving scale with the ensemble approach contains valuable information not only for scientist but also for stakeholders.

A. Appendix

A.1. Introduction
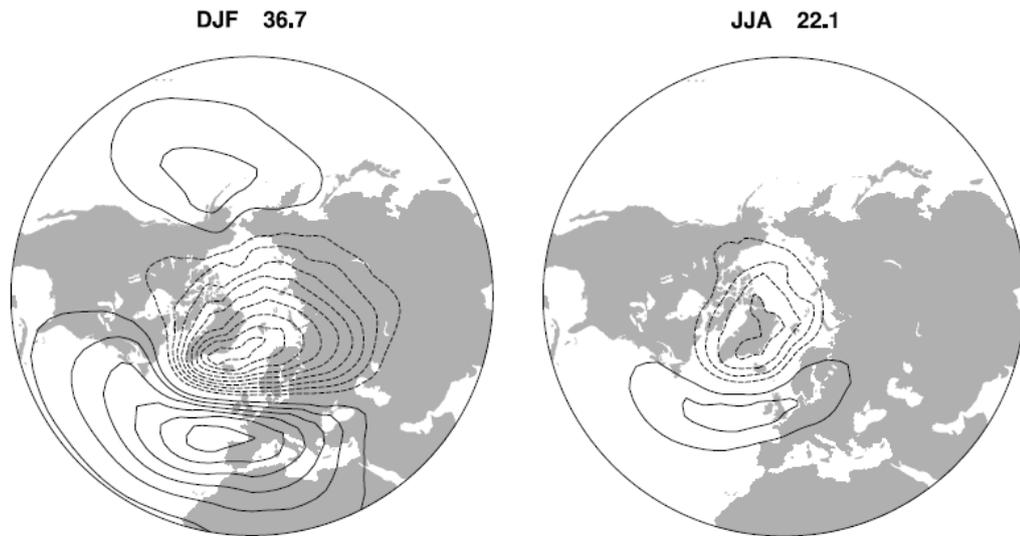
Figure A.1.: Leading empirical orthogonal functions (EOF 1) of the seasonal mean sea level pressure anomalies in the North Atlantic sector (20°-70°N, 90°W-40°E), and the percentage of the total variance they explain. The patterns are displayed in terms of amplitude (hPa), obtained by regressing the hemispheric sea level pressure anomalies upon the leading principal component time series. The contour increment is 0.5 hPa, and the zero contour has been excluded. The data set cover 1899-2001 [see Trenberth and Paolino (1980)]. Graphic and caption from Hurrell et al. (2003)

A.2. Simulating daily maximum temperatures for Europe

tasmax [deg. C]	Mean	RMSE	RMSE SPAT	COR
British Isles				
CCLM EVA b1	15.7	1.21	1.21	0.71
CCLM EVA preop	16.3	0.57	0.66	0.90
Eobs V14	16.8			
France				
CCLM EVA b1	21.3	0.88	1.05	0.76
CCLM EVA preop	21.5	0.78	0.76	0.83
Eobs V14	21.0			
Iberian Peninsula				
CCLM EVA b1	21.3	1.67	1.88	0.87
CCLM EVA preop	21.5	0.54	0.84	0.90
Eobs V14	21.0			
Mid-Europe				
CCLM EVA b1	19.0	0.84	0.88	0.77
CCLM EVA preop	19.5	0.48	0.52	0.86
Eobs V14	19.5			
Scandinavia				
CCLM EVA b1	13.4	1.14	1.42	0.78
CCLM EVA preop	13.1	1.31	1.41	0.89
Eobs V14	14.3			
Alps				
CCLM EVA b1	19.8	0.93	1.30	0.80
CCLM EVA preop	19.8	0.71	0.90	0.85
Eobs V14	19.3			
Mediterranean				
CCLM EVA b1	28.0	3.46	3.72	0.80
CCLM EVA preop	26.6	2.00	1.89	0.80
Eobs V14	24.7			
Eastern Europe				
CCLM EVA b1	21.3	1.00	1.06	0.86
CCLM EVA preop	21.3	0.82	0.84	0.86
Eobs V14	20.7			
Europe				
CCLM EVA b1	19.7	0.65	1.74	0.78
CCLM EVA preop	19.3	0.37	1.19	0.84
Eobs V14	19.2			

Table A.1.: Annual mean summer half year temperatures for 1961 – 2010 for the different PRUDENCE regions of the CCLM EVA simulations and Eobs V14 and the temporal RMSE and Correlation and the spatial mean RMSE of the CCLM Eval simulations and CRU to Eobs.

A.3. Comparing heat waves derived from simulations with observations

(a) pos. EA

(b) neg. EA

(c) pos. NAO

(d) neg. NAO

(e) pos. PNA

(f) neg. PNA
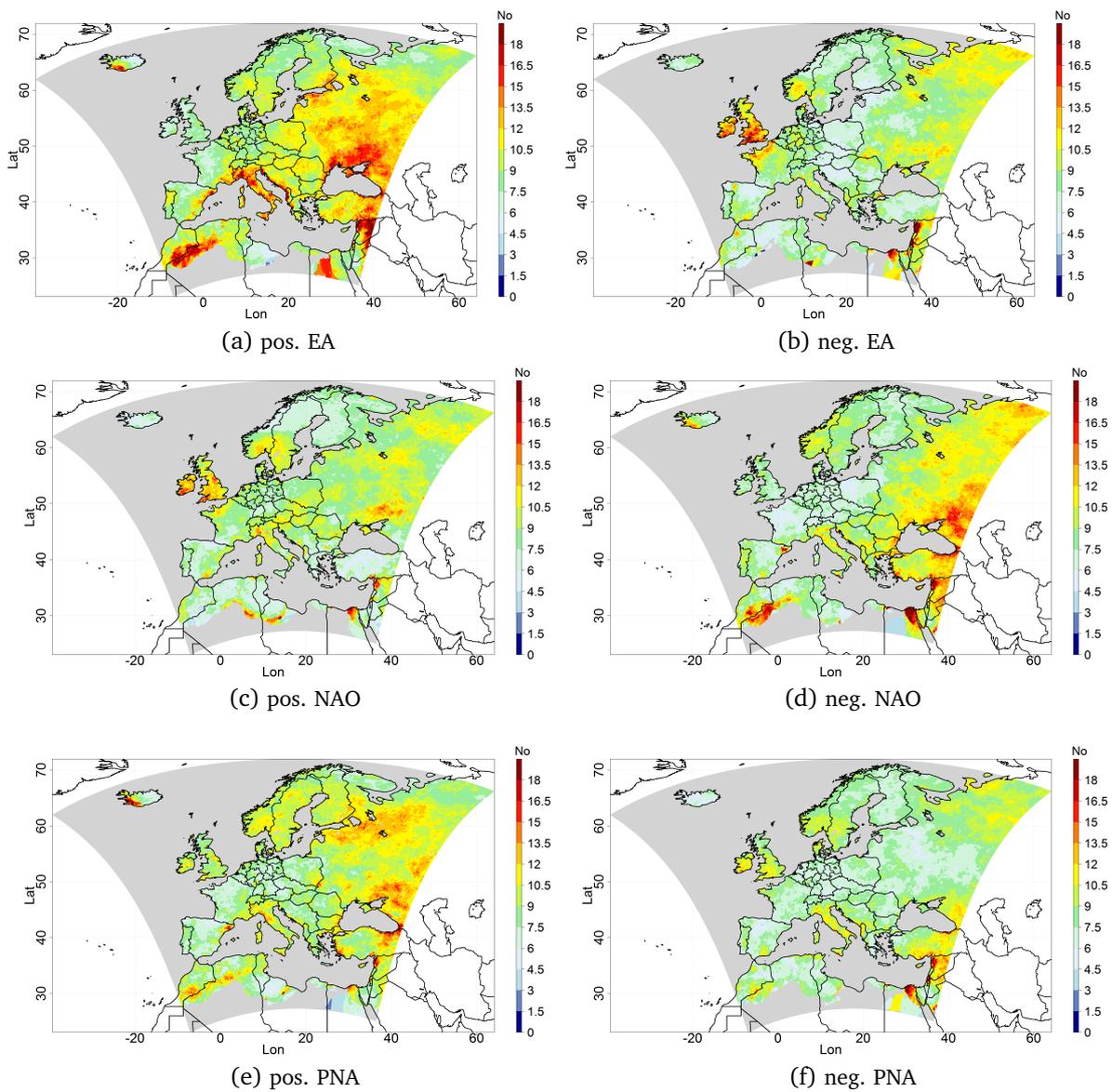
Figure A.2.: Monthly sums of HW days of positive (left) and negative (right) third phases of EA, NAO and PNA from Eobs V14 data for Europe.

(a) pos. POL

(b) neg. POL

(c) pos. PT

(d) neg. PT

(e) pos. SCAND

(f) neg. SCAND
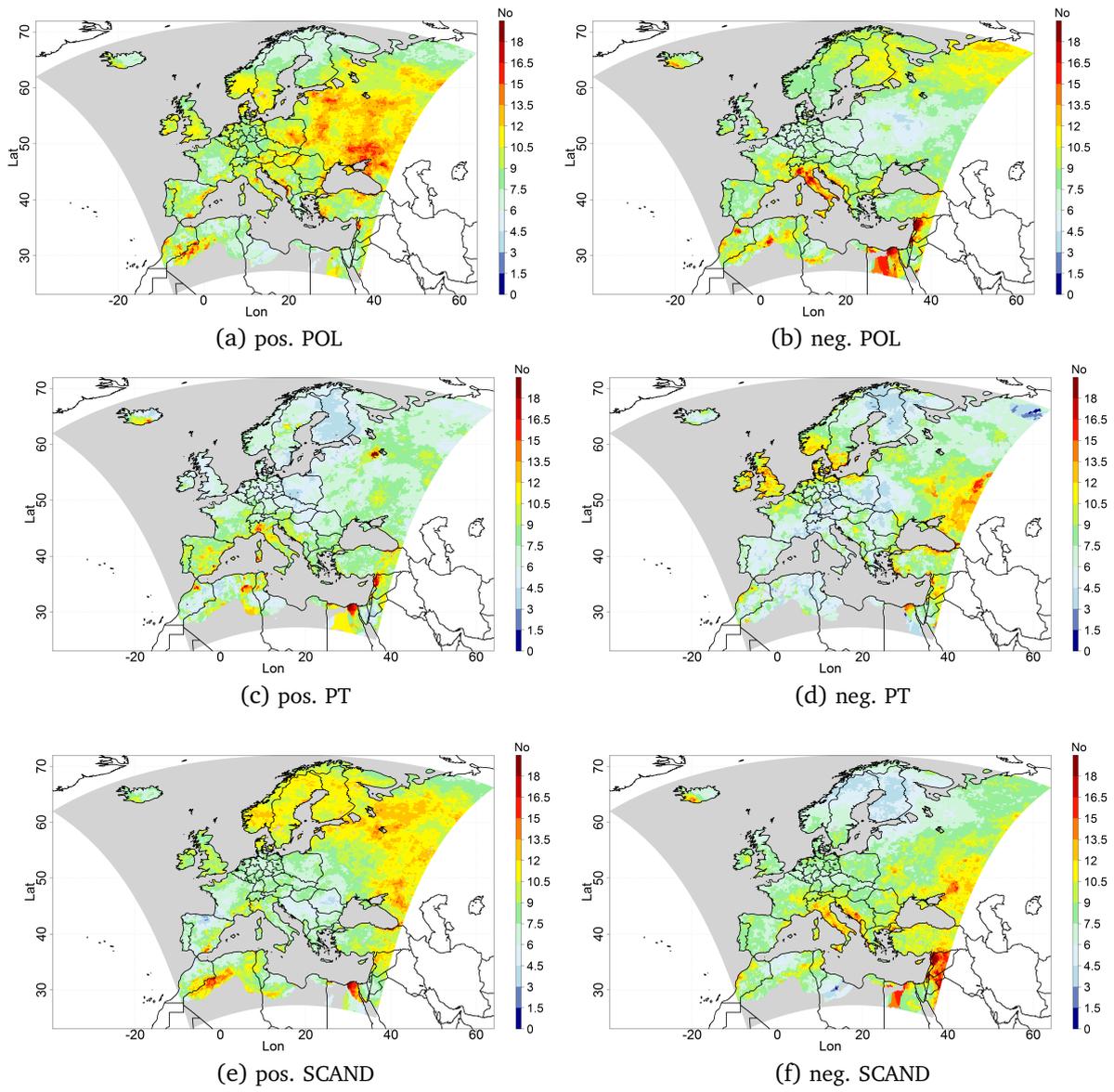
Figure A.3.: Monthly sums of HW days of positive (left) and negative (right) third phases of POL, PT and SCAND from Eobs V14 data for Europe.
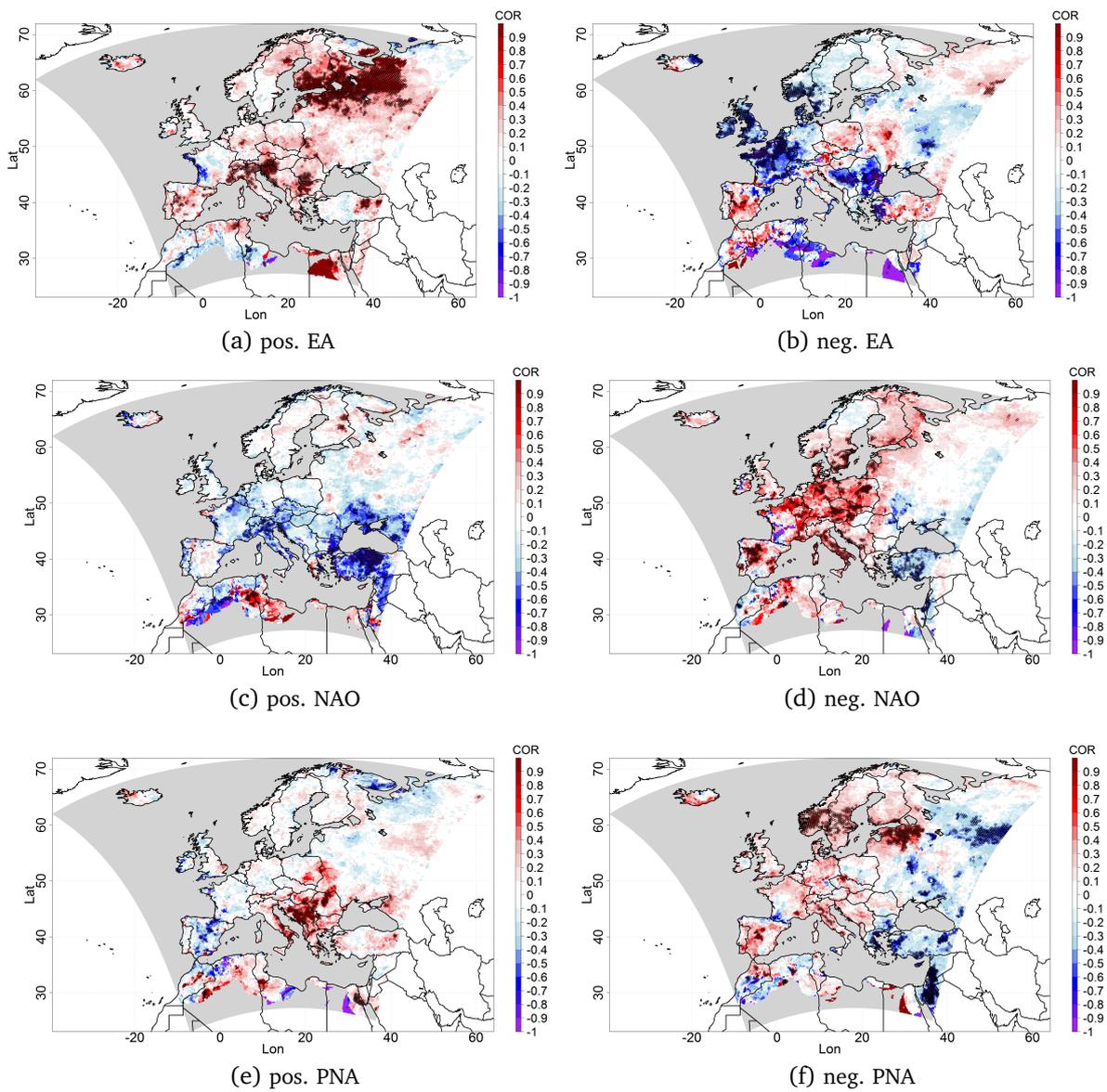
Figure A.4.: Correlation of positive (left) and negative (right) third phases of EA, NAO and PNA to monthly sums of HW days from Eobs V14 data for Europe. The dots in the graphics show significances > 0.95%.
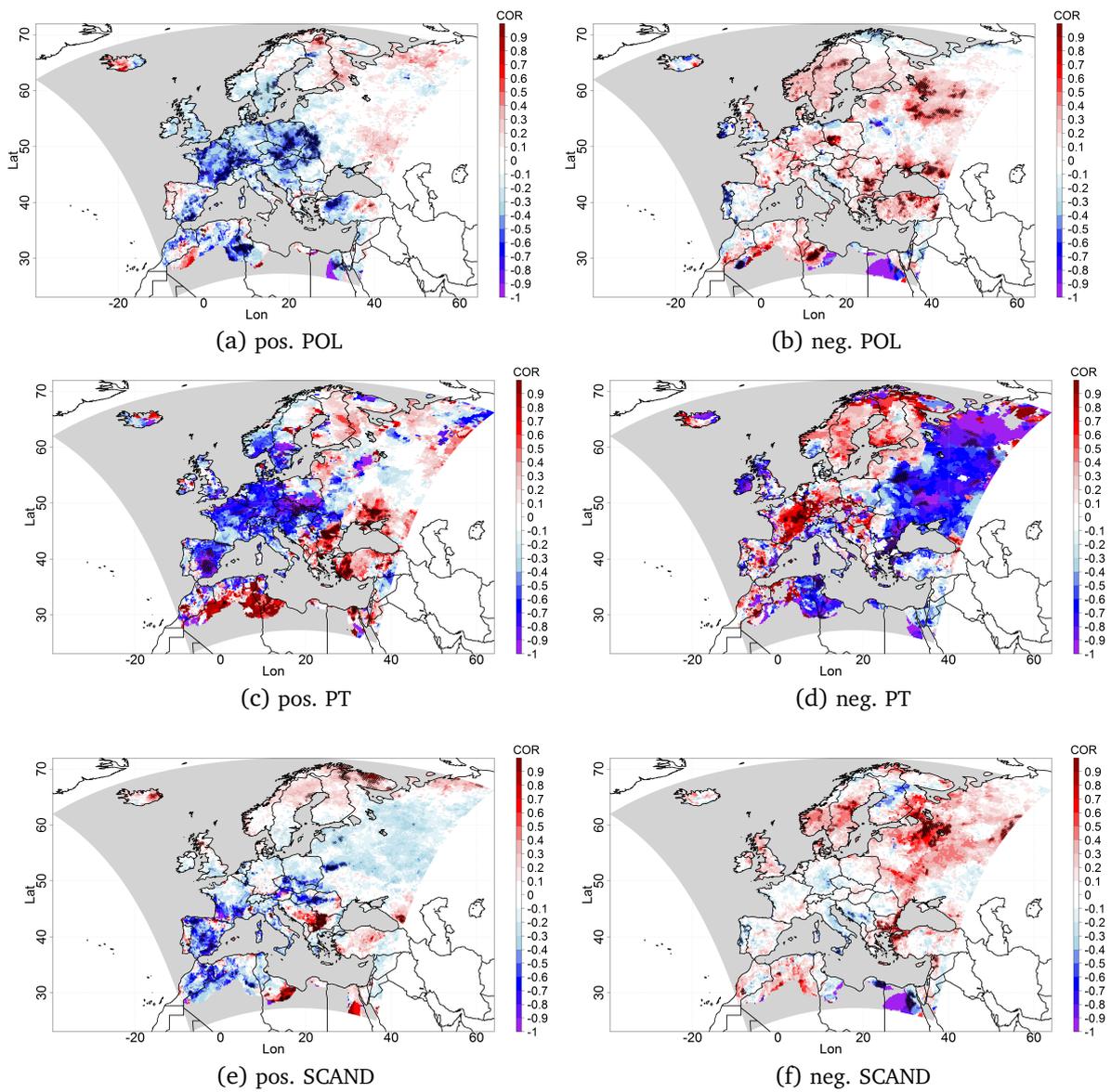
Figure A.5.: Correlation of positive (left) and negative (right) third phases of POL, PT and SCAND to monthly sums of HW days from Eobs V14 data for Europe. The dots in the graphics show significances > 0.95%.
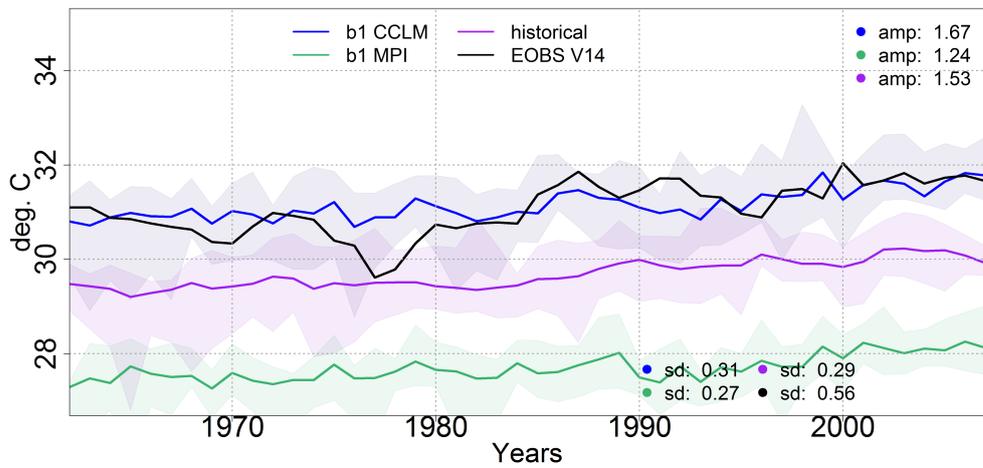
Figure A.6.: 1962-2010 LT2-5 Maximum HW temperatures anomalies for Europe for the model simulation and the Eobs V.14 data. Also displayed is the interquartile range for the model simulations.
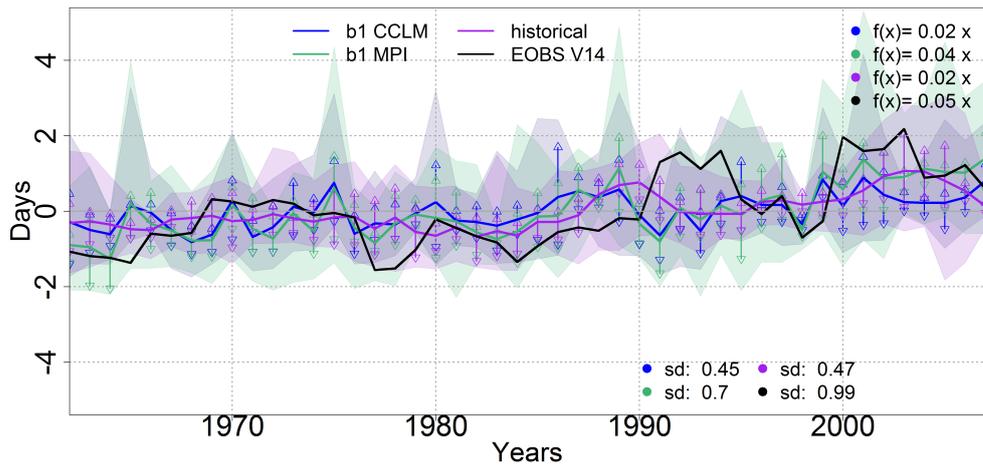
Figure A.7.: 1962-2010 LT2-5 Maximum HW length anomalies for Europe for the model simulation and the Eobs V.14 data. Also displayed is the interquartile range for the model simulations.
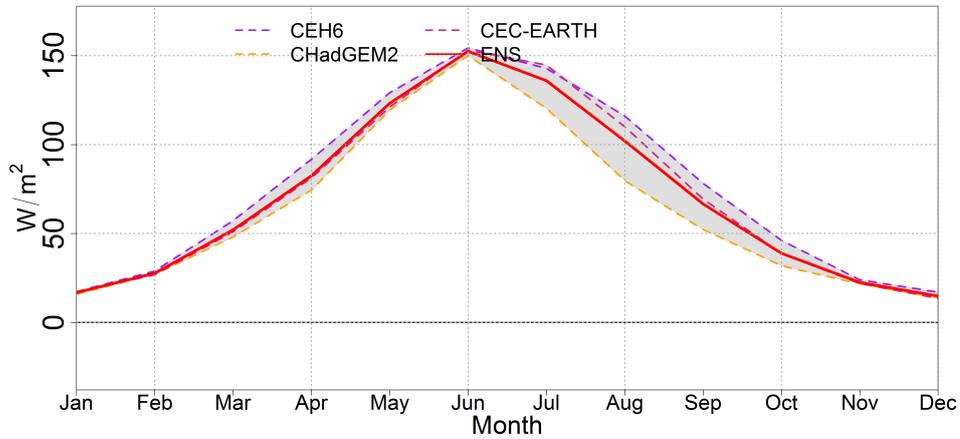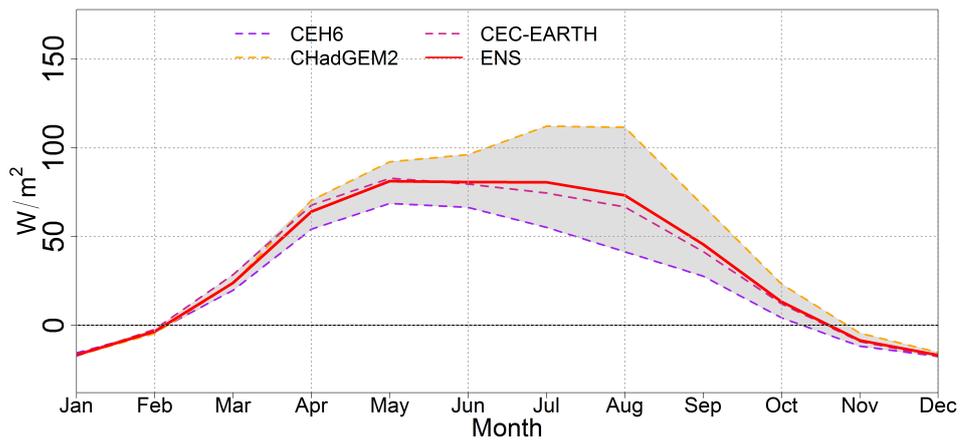
(a) Latent heat flux

(b) Sensible heat flux

(c) Cloud Cover

Figure A.8.: Annual cycle of the latent (upper row) and sensible (middle) heat flux [W/m^2] for 06-18.00 o'clock and the cloud cover [%] (lower row) for the KLIWA area 1971 - 2000; shown are the single ensemble members (dotted lines), the ensemble mean (the red line), the ensemble spread (grey area) and the HYRAS observation data (black line).

	Mean	SD	Min	Max
Summer Days	Temperature > 25 °C/day			
CEH6	10	5	0	31
CEC-EARTH	11	5	0	31
CHadGEM2	41	13	1	78
ENS	20	7	0	45
DWD	32	9	0	58
Hot Days	Temperature > 30 °C/day			
CEH6	1	1	0	4
CEC-EARTH	1	1	0	5
CHadGEM2	10	6	0	34
ENS	4	2	0	14
DWD	5	2	0	16
Frost Days	min. Temperature < 0 °C/day			
CEH6	81	19	42	209
CEC-EARTH	107	19	59	233
CHadGEM2	123	15	80	208
ENS	104	17	62	217
DWD	93	22	37	270
Icing Days	max. Temperature < 0 °C/day			
CEH6	29	12	11	138
CEC-EARTH	40	14	15	158
CHadGEM2	58	14	30	156
ENS	42	13	20	151
DWD	25	11	6	180

Table A.2.: 1971–2000 temperature based climate indices with mean, standard deviation, minimum and maximum values (Zhang et al. (2011)).

	Mean	SD	Min	Max
Dry Days	Precipitation < 1 mm/day			
CEH6	208	19	138	258
CEC-EARTH	232	17	163	271
CHadGEM2	235	19	161	279
ENS	225	18	154	269
HYRAS	234	14	186	277
Heavy Precipitation I	Precipitation > 25 mm/day			
CEH6	6	5	2	69
CEC-EARTH	4	3	1	46
CHadGEM2	3	3	1	45
ENS	5	4	2	53
HYRAS	3	3	1	27
Heavy Precipitation II	Precipitation > 10 mm/hour			
CEH6	3	1	2	14
CEC-EARTH	3	1	1	8
CHadGEM2	3	1	1	8
ENS	3	1	2	10
HYRAS	–	–	–	–
mean dry length	days			
CEH6	13	1	8	16
CEC-EARTH	14	1	10	17
CHadGEM2	15	1	13	20
ENS	14	1	11	17
HYRAS	15	1	0	22
max dry length	days			
CEH6	29	4	18	51
CEC-EARTH	36	9	22	120
CHadGEM2	45	13	22	189
ENS	30	5	18	76
HYRAS	30	5	0	101
number of dry periods	number			
CEH6	3	1	1	5
CHadGEM2	5	1	2	8
CEC-EARTH	4	1	1	7
ENS	4	1	1	7
HYRAS	4	1	0	7

Table A.3.: 1971–2000 precipitation based climate indices with mean, standard deviation, minimum and maximum values (Zhang et al. (2011)); definition dry period: at least 11 consecutive days.

Rel. humidity [%]	CTRL	NF-CTRL	DF-CTRL
MPI-ESM-LR	78.6/85.9	+0.5/+0.3	-1.9/+0.6
HadGEM2-ES	64.1/83.5	+0.4/+1.8	-3.3/+1.6
EC-EARTH	73.5/84.3	-0.8/+0.3	-4.5/+0.3
ENS	72.1/84.6	0.0/+0.8	-3.1/+0.9

Global irradiation [W/m^2]	CTRL	NF-CTRL	DF-CTRL
MPI-ESM-LR	116/52	-3.7/-2.4	-3.5/-6.3
HadGEM2-ES	142/57	-1.2/-5.4	+2.2/-8.0
EC-EARTH	128/57	-1.0/-2.0	+2.2/-4.3
ENS	128/56	-2.0/-3.3	+0.1/-6.2

Cloud Cover [%]	CTRL	NF-CTRL	DF-CTRL
MPI-ESM-LR	78/84	-0.4/-0.6	-4.1/+1.8
HadGEM2-ES	62/80	-3.6/+1.9	-10/+0.4
EC-EARTH	72/78	-1.3/+0.6	-6.6/+0.1
ENS	71/81	-1.8/+0.6	-6.7/+0.9

Table A.4.: CTRL and future period differences to CTRL for summer and winter half (red,blue) for relative humidity, global irradiation and cloud cover for the KLIWA region
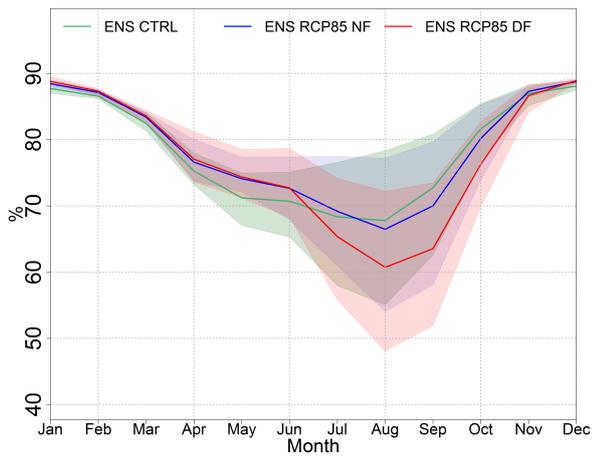
(a) Relative humidity
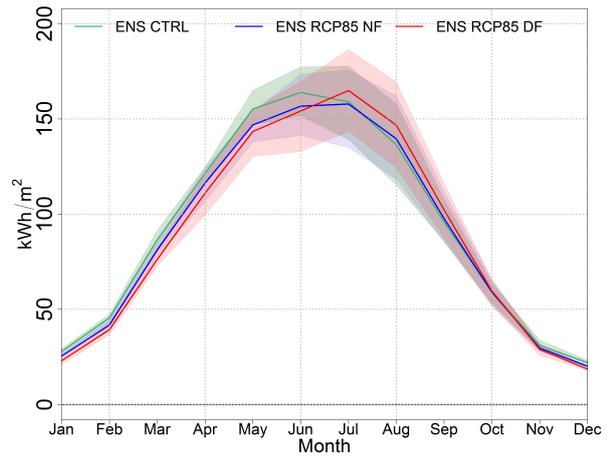
(b) Global Irradiation
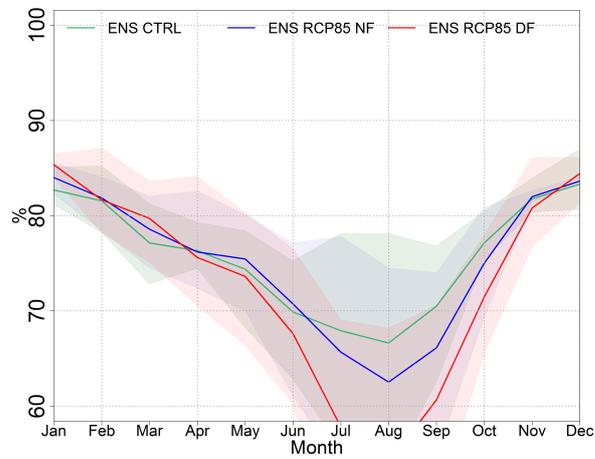
(c) Cloud Cover

Figure A.9.: Annual cycle of the relative humidity [%] (upper row) and global irradiation [W/m^2] (middle) and the cloud cover [%] (lower row) for the KLIWA area, shown are the ensemble means for the three periods and the ensemble spread (shaded areas).
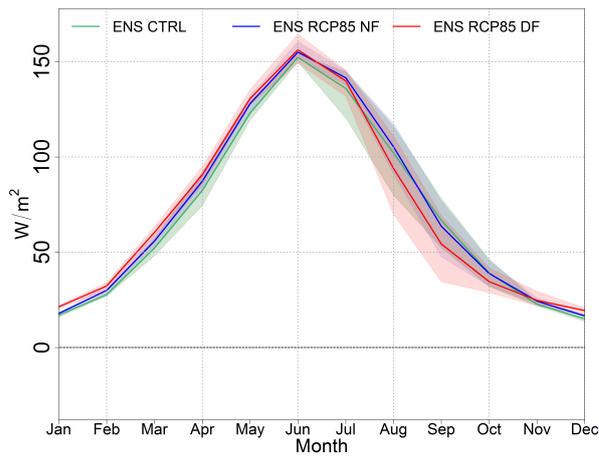
(a) Latent heat
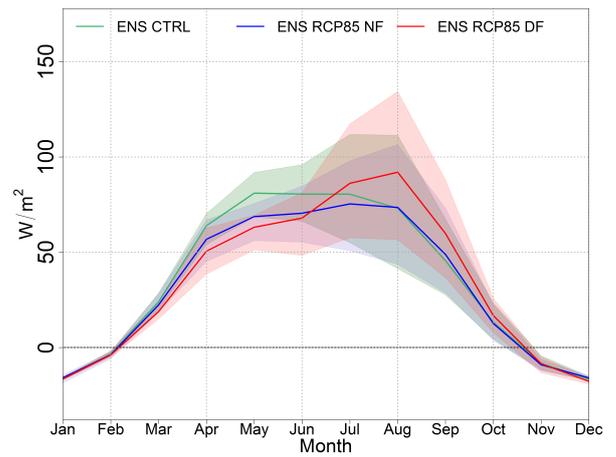
(b) Sensible heat
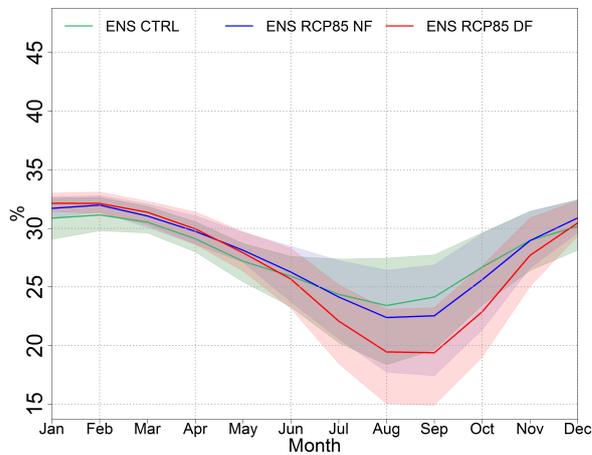
(c) Soil moisture 0-50cm
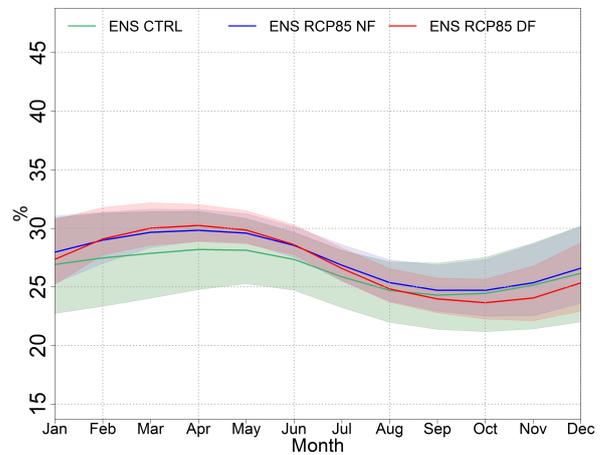
(d) Soil moisture 1-2m
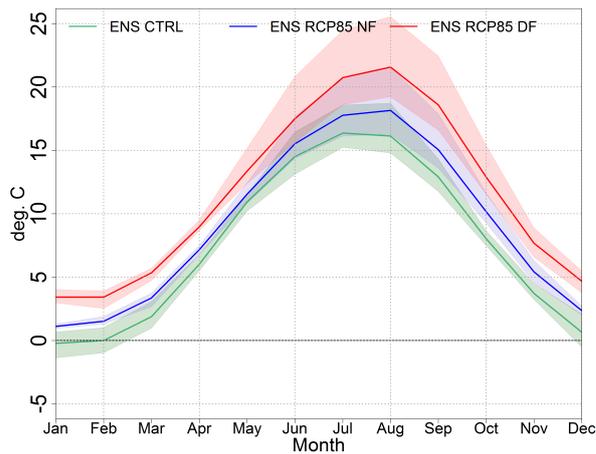
(e) Soil temperature 0-50cm
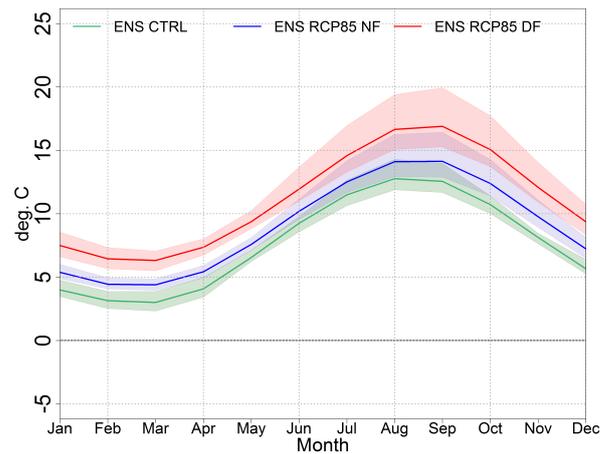
(f) Soil temperature 1-2m

Figure A.10.: Annual cycle of the latent and sensible heat flux [W/m^2] for 06-18.00 o'clock (upper row), the soil water content in [%] for the layers 0-50cm and 1-2m (middle row), and the soil temperature for the same level [°C] (lower row) for the KLIWA area, shown are the ensemble means for the three periods and the ensemble spread (shaded areas).

A.5. Simulations of extremes and their climate change signal within the very high-resolution ensemble
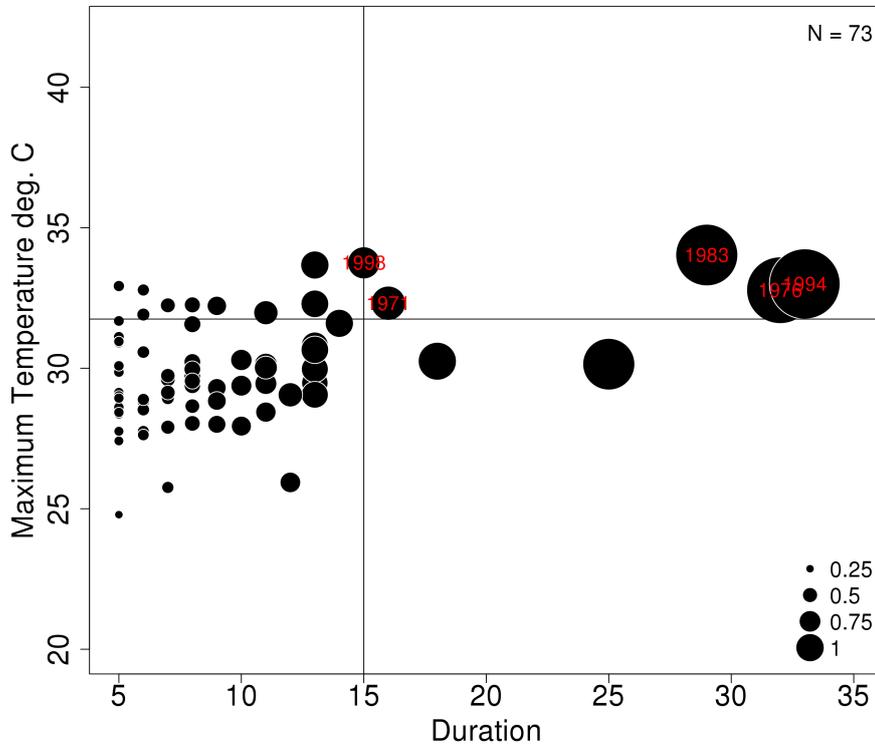
Figure A.11.: Bubble plot displaying HW temperatures, duration and global intensity for the KLIWA area 1971 - 2000; shown are the Eobs observation data (black bubbles). Also the number of HWs in the CTRL period on the right side and the year of the most severe HW in red.
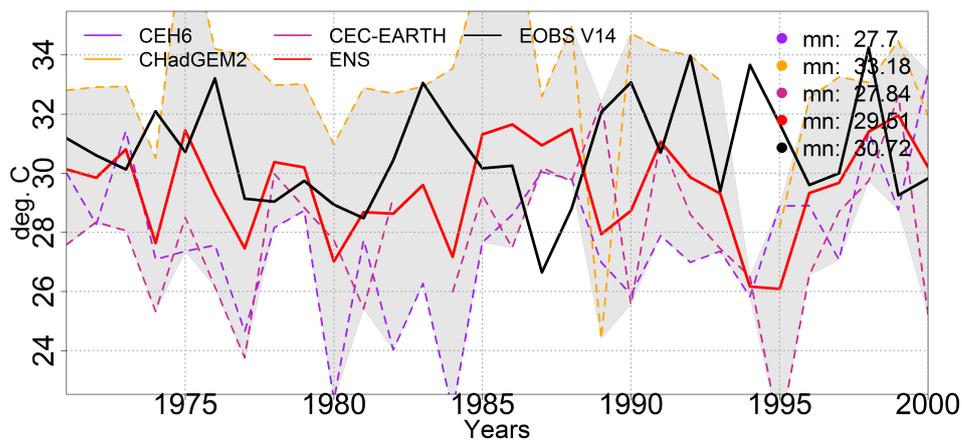
Figure A.12.: Annual max. HW temperatures for the KLIWA area 1971 - 2000; shown are the means of the single ensemble member (dotted lines), the ensemble mean (the red line), the ensemble spread (grey area) and the HYRAS observation data (black line). Also, the mean of the period is plotted in the graph on the right side.
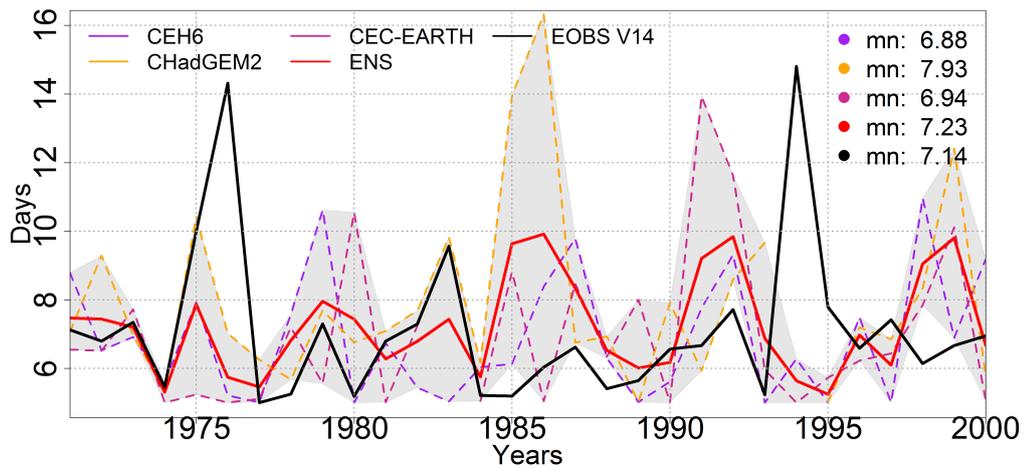
Figure A.13.: Annual max. HW length for the KLIWA area 1971 - 2000; shown are the means of the single ensemble member (dotted lines), the ensemble mean (the red line), the ensemble spread (grey area) and the HYRAS observation data (black line). Also, the mean of the period is plotted in the graph on the right side.

HW temperatures	25 per	Median	75 per	Maximum	FWHM
3rd Nest					
MPI-ESM-LR	21.2	28.4	35.6	42.7	5.3
HadGEM2-ES	24.6	31.3	38.1	44.8	5.6
EC-EARTH	19.8	26.6	33.3	40.0	5.6
2nd Nest					
MPI-ESM-LR	20.7	27.1	33.5	39.9	5.2
HadGEM2-ES	24.0	30.4	36.9	43.3	5.3
EC-EARTH	19.8	26.1	32.4	38.7	5.7
EOBS V14					
EOBS	22.5	28.2	33.9	39.6	4.8

Table A.5.: CTRL period: Quantiles and full width at half maximum (FWHM) of the normal distribution fit for the EOBS V14 observations and the model simulations in [°C], upper block: 3rd Nest, lower block: 2nd Nest
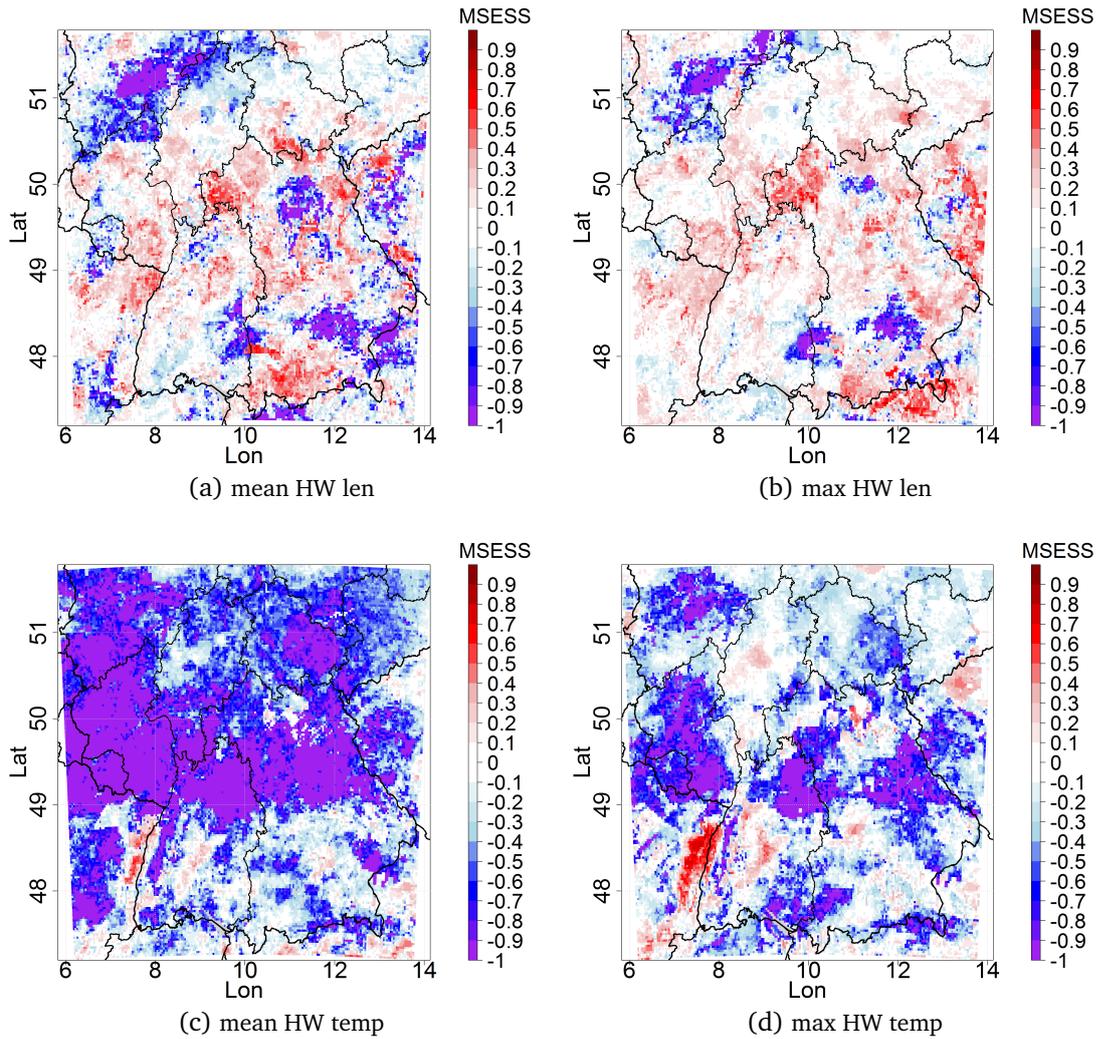
Figure A.14.: 1971 – 2000 Mean and Maximum HW length and temperatures MSESS of the 2nd Nest ENS mean to the 3rd Nest ENS mean to EOBS V14, reddish colors indicate the smaller errors in the 3rd Nest, blueish colors a larger error compared to the 2nd Nest.

(a) Bad Mergentheim-Neunkirchen

(b) Crailsheim-Alexandersreut

(c) Hechingen

(d) Hornisgrinde

(e) Karlsruhe

(f) Münsingen–Apfelstetten

(g) Pfullendorf

(h) Rheinstetten

(i) Rottenburg-Kiebingen

(j) Schwäbisch Gmünd–Straßdorf

(k) Wertheim–Eichel

(l) Wolfach
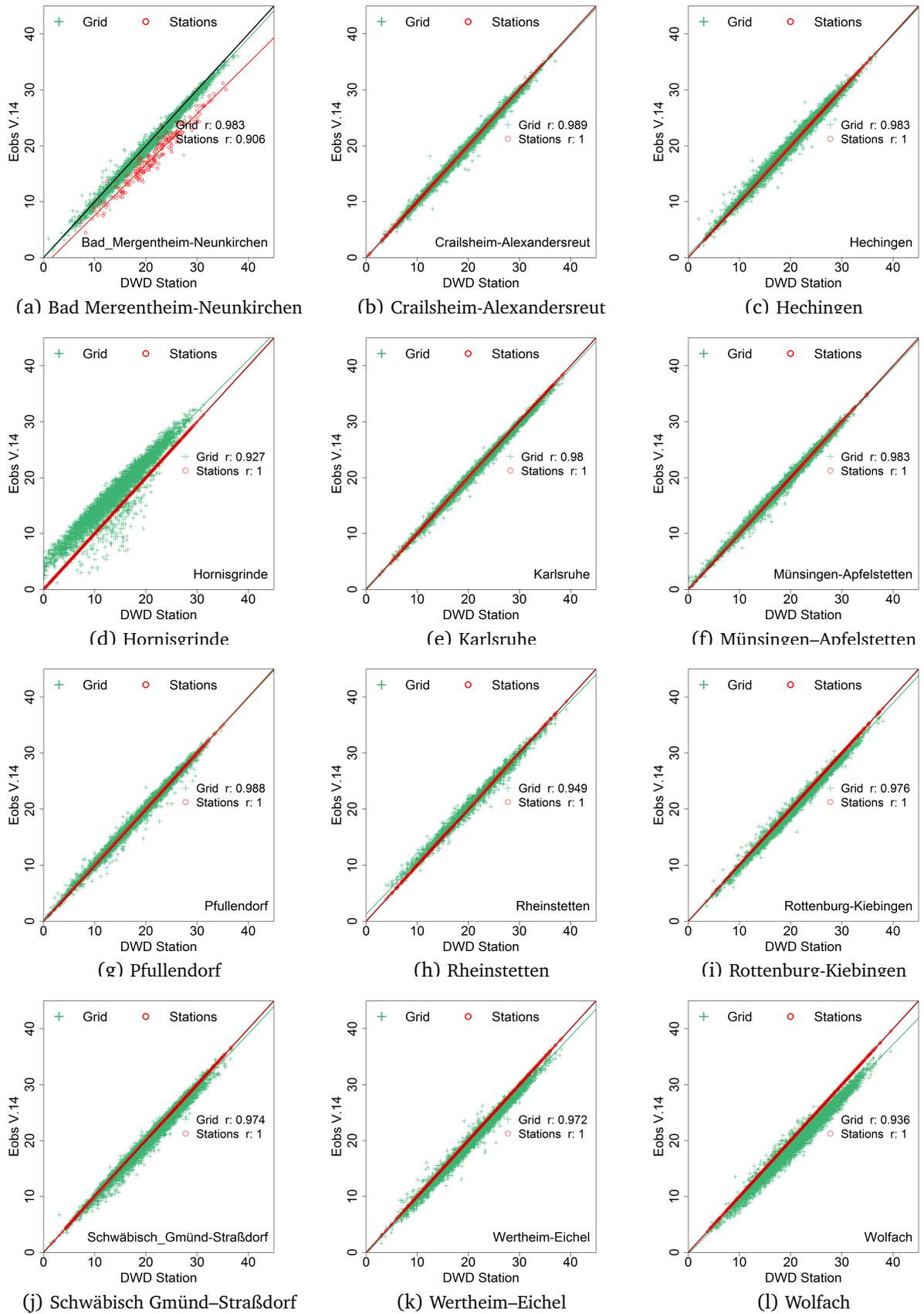
Figure A.15.: Eobs V.14 Grid points evaluations for twelve stations, compared are the EOBS V.14 grid points values and stations data to the DWD station data. In the right, the gradient is displayed.
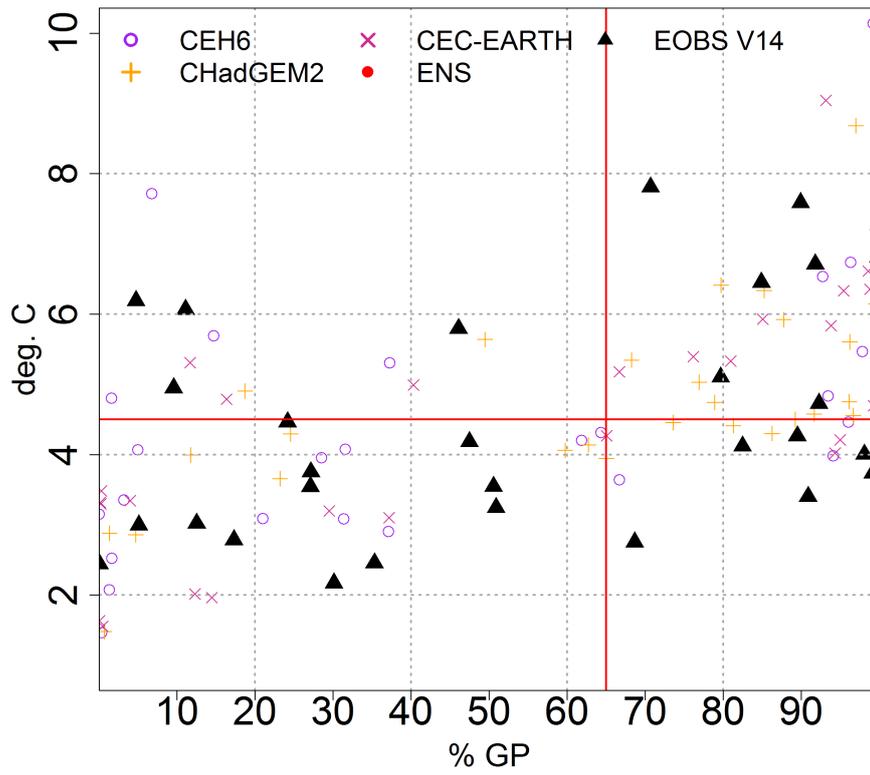
Figure A.16.: Maximum HW temperatures to HW affected grid points sum for the 3rd Nest simulations and the observations EOBS V.14. Additionally marked are lines for 65 % percent gp and 4.5 °C as threshold marks.
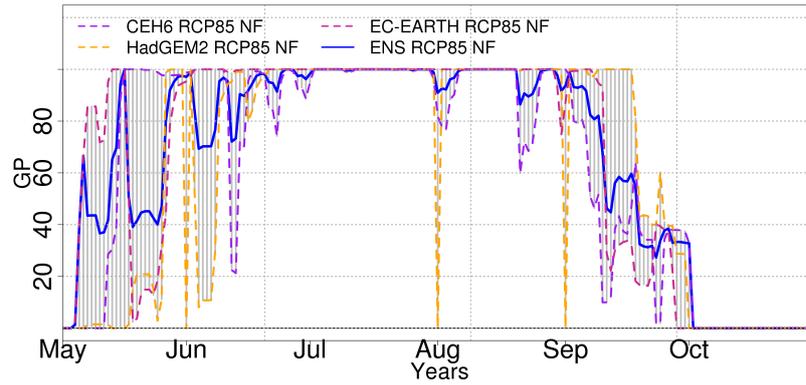
(a) NF
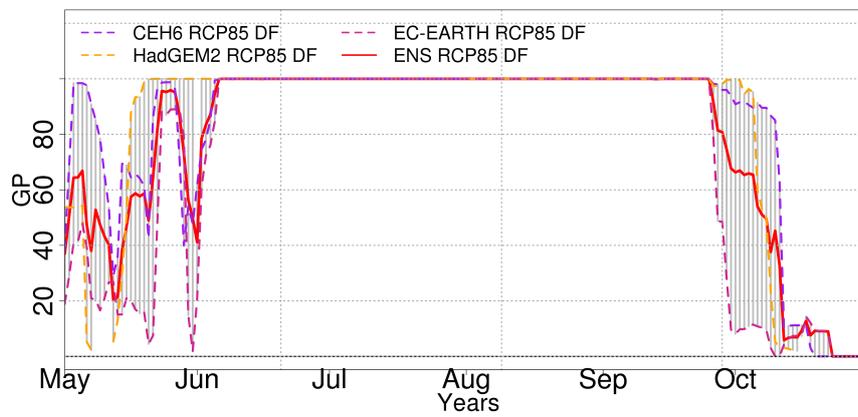
(b) DF

Figure A.17.: Annual cycle of grid points sums of heat wave events for the KLIWA area NF (blue line) and DF (red line); shown are the ensemble members (orange, magenta, purple lines) with the ensemble spread (grey area), the ENS mean (solid line).

	Mean	Min	Max
mean HW temperature [°C]			
NF	0.8	0.5	1.1
DF	2.3	2.0	2.7
max. HW temperature [°C]			
NF	2.2	1.5	3.3
DF	6.7	5.9	7.9
mean HW length [days]			
NF	1.2	0.5	1.9
DF	7.8	4.4	11.0
max. HW length [days]			
NF	2.5	0.9	4.5
DF	16.6	9.0	24.5
number of HW			
NF	0.6	0.3	0.9
DF	2.0	1.8	2.3

Table A.6.: Changes in mean and maximum HW temperatures and lengths and number of HW per year for the 0.025 °resolution between the NF and DF to CTRL period

A.5.1. Highres - Precipitation extremes

	Mean	Min	Max
Dry Days	Precipitation < 1 mm/day		
NF	2.4	-0.1	5.1
DF	13.3	11.4	15.8
Heavy Precipitation I	Precipitation > 25 mm/day		
NF	0.7	0.4	1.4
DF	16.7	7.5	46.0
Heavy Precipitation II	Precipitation > 10 mm/hour		
NF	0.7	0.4	1.1
DF	1.5	1.0	2.0

Table A.7.: precipitation based climate indices differences NF/DF to CTRL with mean, minimum and maximum values.

List of Figures

List of Tables

Bibliography

Albergel, C., and Coauthors, 2013: Skill and global trend analysis of soil moisture from re-analyses and microwave remote sensing. *J. Hydrometeor*, **14 (4)**, 1259–1277, doi:10.1175/JHM-D-12-0161.1.

Ardilouze, C., and Coauthors, 2017: Multi-model assessment of the impact of soil moisture initialization on mid-latitude summer predictability. *Climate Dynamics*, **49 (11)**, 3959–3974, doi:10.1007/s00382-017-3555-7.

Bador, M., L. Terray, and J. Boé, 2016: Emergence of human influence on summer record-breaking temperatures over Europe. *Geophysical Research Letters*, **43 (1)**, 404–412, doi:10.1002/2015GL066560.

Bador, M., L. Terray, J. Boe, S. Somot, A. Alias, A.-L. Gibelin, and B. Dubuisson, 2017: Future summer mega-heatwave and record-breaking temperatures in a warmer France climate. *Environmental Research Letters*, **12 (7)**, doi:10.1088/1748-9326/aa751c.

Baldauf, M., A. Seifert, J. Förstner, D. Majewski, and M. Ras, 2011: Operational convective-scale numerical weather prediction with the COSMO model: Description and sensitivities. *Monthly Weather Review*, **139**, 3887–3905, doi:10.1175/MWR-D-10-05013.1.

Balmaseda, M. A., K. Mogensen, and A. T. Weaver, 2013: Evaluation of the ECMWF ocean reanalysis system ORAS4. *Quarterly Journal of the Royal Meteorological Society*, **139 (674)**, 1132–1161, doi:10.1002/qj.2063, https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.2063.

Balsamo, G., P. Viterbo, A. Beljaars, B. v. d. Hurk, M. Hirschi, A. K. Betts, and K. S., 2009: A revised hydrology for the ECMWF model: Verification from field site to terrestrial water storage and impact in the integrated forecast system. *Journal of Hydrometeorology*, **10**, 623–643, doi:10.1175/2008JHM1068.1.

Ban, N., J. Schmidli, and C. Schär, 2014: Evaluation of the convection–resolving regional climate modeling approach in decade–long simulations. *Journal of Geophysical Research: Atmospheres*, **119 (13)**, 7889–7907, doi:10.1002/2014JD021478, https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2014JD021478.

Barcena-Martin, E., J. Molina, and J. D. Ruiz-Sinoga, 2018: Issues and challenges in defining a heat wave: A Mediterranean case study. *International Journal of Climatology*, **0 (0)**, doi:10.1002/joc.5809, https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/joc.5809.

Barnston, A. G., and R. E. Livezey, 1987: Classification, seasonality and persistence of low-frequency atmospheric circulation patterns. *Monthly Weather Review*, **115 (6)**, 1083–1126, doi:10.1175/1520-0493(1987)115<1083:CSAPOL>2.0.CO;2, https://doi.org/10.1175/1520-0493(1987)115<1083:CSAPOL>2.0.CO;2.

Barriopedro, D., E. M. Fischer, J. Luterbacher, R. M. Trigo, and R. Garcia-Herrera, 2011: The hot summer of 2010: redrawing the temperature record map of Europe. *Science*, **332**, 220–224, doi:10.1126/science.1201224.

Bartholomé, E., and A. S. Belward, 2005: Glc2000: A new approach to global land cover mapping from Earth observation data. *International Journal of Remote Sensing*, **26 (9)**, 1959–1977, doi:10.1080/01431160412331291297.

Berckmans, J., T. Woollings, M.-E. Demory, P.-L. Vidale, and M. Roberts, 2013: Atmospheric blocking in a high resolution climate model: influences of mean state, orography and eddy forcing. *Atmospheric Science Letters*, **14 (1)**, 34–40, doi:10.1002/asl2.412, https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/asl2.412.

Berg, P., H. Feldmann, and H.-J. Panitz, 2012: Bias correction of high resolution regional climate model data. *Journal of Hydrology*, **448**, 80–92, doi:10.1016/j.jhydrol.2012.04.026.

Berg, P., C. Moseley, and J. O. Haerter, 2013: Strong increase in convective precipitation in response to higher temperatures. *Nature Geoscience*, **6**, 181–, doi:10.1038/ngeo1731.

Bissolli, P., and E. Dittmann, 2001: The objective weather types classification of the German Weather Service and its possibilities of application to environmental and meteorological investigations. *Meteorologische Zeitschrift*, **10 (4)**, 253–260, doi:10.1127/0941-2948/2001/0010-0253.

Black, E., M. Blackburn, G. Harrison, B. Hoskins, and J. Methven, 2004: Factors contributing to the summer 2003 European heatwave. *Weather*, **59 (8)**, 217–223, doi:10.1256/wea.74.04.

Blume, H.-P., G. W. Brümmer, R. Horn, E. Kandeler, I. Kögel-Knabner, R. Kretzschmar, K. Stahr, and B.-W. Wilke, 2010: *Scheffer/Schachtschabel: Lehrbuch der Bodenkunde*. Spektrum Akademischer Verlag Heidelberg.

Boer, G. J., V. V. Kharin, and W. J. Merryfield, 2013: Decadal predictability and forecast skill. *Climate Dynamics*, **41**, doi:10.1007/s00382-013-1705-0.

Boer, G. J., and Coauthors, 2016: The decadal climate prediction project (DCPP) contribution to CMIP6. *Geoscientific Model Development*, **9 (10)**, 3751–3777, doi:10.5194/gmd-9-3751-2016.

Boone, A., V. Masson, T. Meyers, and J. Noilhan, 2000: The influence of the inclusion of soil freezing on simulations by a soil-vegetation-atmosphere transfer scheme. *Journal of Applied Meteorology*, **39 (9)**, 1544–1569, doi:10.1175/1520-0450(2000)039<1544:TIOTIO>2.0.CO;2.

Brabson, B. B., D. H. Lister, P. D. Jones, and J. P. Palutikof, 2005: Soil moisture and predicted spells of extreme temperatures in Britain. *Journal of Geophysical Research: Atmospheres*, **110 (D5)**, n/a–n/a, doi:10.1029/2004JD005156, d05104.

Braden, H., 2012: The model AMBETI. Tech. rep., Deutscher Wetterdienst. Available online at ftp://ftp-cdc.dwd.de/pub/CDC/grids_germany/monthly/soil_temperature_5cm/AMBETI.pdf, last accessed: 19.09.2010.

Braun, F. J., and G. Schädler, 2005: Comparison of soil hydraulic parameterizations for mesoscale meteorological models. *Journal of Applied Meteorology*, **44**, 1116–1132, doi:10.1175/JAM2259.1.

Breil, M., N. Laube, J. G. Pinto, and G. Schädler, 2018a: The impact of soil initialization on regional decadal climate predictions in Europe. *Climate Research*, in revision.

Breil, M., and G. Schädler, 2017: Quantification of the uncertainties in soil and vegetation parameterizations for regional climate simulations in Europe. *Journal of Hydrometeorology*, doi:10.1175/JHM-D-16-0226.1, accepted, http://dx.doi.org/10.1175/JHM-D-16-0226.1.

Breil, M., G. Schädler, and N. Laube, 2018b: An improved soil moisture parametrization for regional climate simulations in Europe. *Journal of Geophysical Research: Atmospheres*, **123 (14)**, 7331–7339, doi:10.1029/2018JD028704, URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018JD028704, https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2018JD028704.

Brockhaus, P., D. Lüthi, and C. Schär, 2008: Aspects of the diurnal cycle in a regional climate model. *Meteorologische Zeitschrift*, **17 (4)**, 433–443, doi:10.1127/0941-2948/2008/0316.

Bueh, C., and H. Nakamura, 2007: Scandinavian pattern and its climatic impact. *Quarterly Journal of the Royal Meteorological Society*, **133 (629)**, 2117–2131, doi:10.1002/qj.173.

Buehler, T., C. C. Raible, and T. F. Stocker, 2011: The relationship of winter season north atlantic blocking frequencies to extreme cold or dry spells in the ERA-40. *Tellus A: Dynamic Meteorology and Oceanography*, **63 (2)**, 174–187, doi:10.1111/j.1600-0870.2010.00492.x, https://doi.org/10.1111/j.1600-0870.2010.00492.x.

Cassou, C., L. Terray, and A. S. Phillips, 2005: Tropical atlantic influence on European heat waves. *Journal of Climate*, **18 (15)**, 2805–2811, doi:10.1175/JCLI3506.1, https://doi.org/10.1175/JCLI3506.1.

Cattiaux, J., H. Douville, and Y. Peings, 2013: European temperatures in CMIP5: origins of present-day biases and future uncertainties. *Climate Dynamics*, **41 (11)**, 2889–2907, doi:10.1007/s00382-013-1731-y.

Chan, S. C., E. J. Kendon, H. J. Fowler, S. Blenkinsop, N. M. Roberts, and C. A. T. Ferro, 2014: The value of high-resolution met office regional climate models in the simulation of multihourly precipitation extremes. *Journal of Climate*, **27 (16)**, 6155–6174, doi: 10.1175/JCLI-D-13-00723.1, https://doi.org/10.1175/JCLI-D-13-00723.1.

Christensen, J. H., and O. B. Christensen, 2007: A summary of the PRUDENCE model projections of changes in European climate by the end of this century. *Climatic Change*, **81 (1)**, 7–30, doi: 10.1007/s10584-006-9210-7.

Chronis, T., D. E. Raitsos, D. Kassis, and A. Sarantopoulos, 2011: The Summer North Atlantic Oscillation influence on the Eastern Mediterranean. *Journal of Climate*, **24 (21)**, 5584–5596, doi:10.1175/2011JCLI3839.1, https://doi.org/10.1175/2011JCLI3839.1.

Collins, W. J., and Coauthors, 2011: Development and evaluation of an earth-system model - HadGEM2. *Geoscientific Model Development*, **4 (4)**, 1051–1075, doi:10.5194/gmd-4-1051-2011.

Coughlan de Perez, E., and Coauthors, 2018: Global predictability of temperature extremes. *Environmental Research Letters*, **13 (5)**, doi:10.1088/1748-9326/aab94a.

Coumou, D., J. Lehmann, and J. Beckmann, 2015: The weakening summer circulation in the northern hemisphere mid-latitudes. *Science*, **348 (6232)**, 324–327, doi:10.1126/science.1261768.

Coumou, D., V. Petoukhov, S. Rahmstorf, S. Petri, and H. J. Schellnhuber, 2014: Quasi-resonant circulation regimes and hemispheric synchronization of extreme weather in boreal summer. *Proceedings of the National Academy of Sciences*, **111 (34)**, 12 331–12 336, doi:10.1073/pnas.1412797111.

CRED, EM-DAT, and UNISDR, 2015: The human cost of weather related disasters 1995-2015. Tech. rep., Center for Research on the Epidemiology of Disasters (CRED), The United Nations Office for Disaster Risk Reduction (UNISDR). Available online at https://www.unisdr.org/2015/docs/climatechange/COP21_WeatherDisastersReport_2015_FINAL.pdf.

D. Chung, D., R. Kidd, W. Wagner, W. Dorigo, and R. De Jeu, 2015: Product specification document (psd). Tech. rep., European Space Agency (ESA).

Dawson, A., and T. N. Palmer, 2015: Simulating weather regimes: impact of model resolution and stochastic parameterization. *Climate Dynamics*, **44 (7)**, 2177–2193, doi:10.1007/s00382-014-2238-x.

Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, **137 (656)**, 553–597, doi:10.1002/qj.828.

Della-Marta, P. M., M. R. Haylock, J. Luterbacher, and H. Wanner, 2007a: Doubled length of western European summer heat waves since 1880. *Journal of Geophysical Research: Atmospheres*, **112 (D15)**, doi:10.1029/2007JD008510, d15103.

Della-Marta, P. M., J. Luterbacher, H. von Weissenfluh, E. Xoplaki, M. Brunet, and H. Wanner, 2007b: Summer heat waves over western Europe 1880-2003, their relationship to large-scale forcings and predictability. *Climate Dynamics*, **29 (2)**, 251–275, doi:10.1007/s00382-007-0233-1.

Deutsche Wetterdienst, 2018a: CDC Climate Data Center FTPServer. Last accessed 07.11.2018, ftp://ftp-cdc.dwd.de/pub/CDC/.

Deutsche Wetterdienst, 2018b: Markante Hitzewellen seit 1950. Available online at https://www.dwd.de/DE/klimaumwelt/aktuelles/170619_markante_hitzewellen.html.

Deutsche Wetterdienst, 2018c: Numerische Modellvorhersagedaten. [Available online at https://www.dwd.de/DE/leistungen/modellvorhersagedaten/modellvorhersagedaten.html?nn=16102].

Deutsche Wetterdienst, 2018d: REGNIE - Regionalisierte Niederschläge Verfahrensbeschreibung und Nutzeranleitung. Available online at https://www.dwd.de/DE/leistungen/regnie/download/regnie_beschreibung_pdf.pdf?__blob=publicationFile&v=3.

Deutsche Wetterdienst, 2018e: WLK-Daten. Available online at https://www.dwd.de/DE/leistungen/wetterlagenklassifikation/online_wlkdaten.txt?view=nasPublication&nn=16102, last accessed 08.11.2018.

Dirmeyer, P. A., 2001: An evaluation of the strength of land-atmosphere coupling. *Journal of Hydrometeorology*, **2 (4)**, 329–344, doi:10.1175/1525-7541(2001)002<0329:AEOTSO>2.0.CO;2, https://doi.org/10.1175/1525-7541(2001)002<0329:AEOTSO>2.0.CO;2.

Dirmeyer, P. A., S. Kumar, M. J. Fennessy, E. J. Altshuler, T. DelSole, Z. Guo, B. A. Cash, and D. Straus, 2013: Model estimates of land-driven predictability in a changing climate from CCSM4. *Journal of Climate*, **26 (21)**, 8495–8512, doi:10.1175/JCLI-D-13-00029.1.

Dirmeyer, P. A., and Coauthors, 2012: Simulating the diurnal cycle of rainfall in global climate models: resolution versus parameterization. *Climate Dynamics*, **39 (1)**, 399–418, doi:10.1007/s00382-011-1127-9.

Dittmann, E., S. Barth, G. Müller-Westermeier, and J. Lang, 1995: Objektive Wetterlagenklassifikation. Tech. Rep. 197, Selbstverlag des Deutschen Wetterdienstes, Offenbach am Main.

Doms, G., and M. Baldauf, 2015: A description of the nonhydrostatic regional COSMO-Model. Part I: Dynamics and numerics. Tech. rep., Deutscher Wetterdienst (DWD), Offenbach, 225–235 pp. Available online at https://www.dwd.de/SharedDocs/downloads/EN/model_

description/nwp/lm_parts/lm_docu_1_dynamics_200211_en.pdf?__blob=publicationFile&v=4.

Doms, G., and Coauthors, 2011: A description of the nonhydrostatic regional COSMO-Model. Part I: Physical parameterization. Tech. rep., Consortium for Small-Scale Modelling.

Donat, M., L. Alexander, H. Yang, I. Durre, R. Vose, and J. Caesar, 2013: Global land-based datasets for monitoring climatic extremes. *Bulletin of the American Meteorological Society*, **94 (7)**, 997–1006, doi:10.1175/BAMS-D-12-00109.1, https://doi.org/10.1175/BAMS-D-12-00109.1.

Dorigo, W., and Coauthors, 2015: Evaluation of the ESA CCI soil moisture product using ground-based observations. *Remote Sensing of Environment*, **162**, 380 – 395, doi:10.1016/j.rse.2014.07.023.

Dunn-Sigouin, E., and S.-W. Son, 2013: Northern hemisphere blocking frequency and duration in the CMIP5 models. *Journal of Geophysical Research: Atmospheres*, **118 (3)**, 1179–1188, doi:10.1002/jgrd.50143, URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/jgrd.50143, https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/jgrd.50143.

DWD Climate Data Center (CDC), 2018: Historische tägliche Stationsbeobachtungen (Temperatur, Druck, Niederschlag, Sonnenscheindauer, etc.) für Deutschland. Version v006, Available online at ftp://ftp-cdc.dwd.de/pub/CDC/observations_germany/climate/daily/kl/historical/.

Eade, R., D. Smith, A. Scaife, E. Wallace, N. Dunstone, L. Hermanson, and N. Robinson, 2014: Do seasonal-to-decadal climate predictions underestimate the predictability of the real world? *Geophysical Research Letters*, **41 (15)**, 5620–5628, doi:10.1002/2014GL061146, https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2014GL061146.

EFDC, 2016: EFDC Data Portal. Available online at http://www.europe-fluxdata.eu/home/.

Enfield, D. B., A. M. Mestas-Nuñez, and P. J. Trimble, 2001: The atlantic multidecadal oscillation and its relation to rainfall and river flows in the continental U.S. *Geophysical Research Letters*, **28 (10)**, 2077–2080, doi:10.1029/2000GL012745, https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2000GL012745.

Ferranti, L., and P. Viterbo, 2006: The European summer of 2003: Sensitivity to soil water initial conditions. *Journal of Climate*, **19 (15)**, 3659–3680, doi:10.1175/JCLI3810.1, http://dx.doi.org/10.1175/JCLI3810.1.

Fischer, E. M., and C. Schär, 2010: Consistent geographical patterns of changes in high-impact European heatwaves. *Nature*, **3**, 398–403, doi:doi:10.1038/ngeo866.

Fischer, E. M., S. I. Seneviratne, D. Lüthi, and C. Schär, 2007a: Contribution of land-atmosphere coupling to recent European summer heat waves. *Geophysical Research Letters*, **34 (6)**, doi:10.1029/2006GL029068, l06707.

Fischer, E. M., S. I. Seneviratne, P. L. Vidale, D. Lüthi, and C. Schär, 2007b: Soil moisture-atmosphere interactions during the 2003 European summer heat wave. *Journal of Climate*, **20 (20)**, 5081–5099, doi:10.1175/JCLI4288.1, https://doi.org/10.1175/JCLI4288.1.

Fischer, G., F. F. Nachtergaele, S. Prieler, H. van Velthuizen, L. L. Verelst, and D. Wiberg, 2008: Global Agro-ecological Zones Assessment for Agriculture (GAEZ 2008). Tech. rep., IIASA, Laxenburg, Austria and FAO, Rome, Italy.

Folland, C. K., J. Knight, H. W. Linderholm, D. Fereday, S. Ineson, and J. W. Hurrell, 2009: The Summer North Atlantic Oscillation: Past, Present, and Future. *Journal of Climate*, **22 (5)**, 1082–1103, doi:10.1175/2008JCLI2459.1, https://doi.org/10.1175/2008JCLI2459.1.

Ford, K. R., A. K. Ettinger, J. D. Lundquist, M. S. Raleigh, and J. Hille Ris Lambers, 2013: Spatial heterogeneity in ecologically important climate variables at coarse and fine scales in a high-snow mountain landscape. *PloS one*, **8 (6)**, doi:10.1371/journal.pone.0065008.

Fosser, G., S. Khodayar, and P. Berg, 2015: Benefit of convection permitting climate model simulations in the representation of convective precipitation. *Climate Dynamics*, **44 (1)**, 45–60, doi:10.1007/s00382-014-2242-1.

Fosser, G., S. Khodayar, and P. Berg, 2017: Climate change in the next 30 years: What can a convection-permitting model tell us that we did not already know? *Climate Dynamics*, **48 (5)**, 1987–2003, doi:10.1007/s00382-016-3186-4.

Fouillet, A., and Coauthors, 2006: Excess mortality related to the August 2003 heat wave in France. *International Archives of Occupational and Environmental Health*, **80 (1)**, 16–24, doi:10.1007/s00420-006-0089-4.

Francis, J. A., and N. Skific, 2015: Evidence linking rapid Arctic warming to mid-latitude weather patterns. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, **373 (2045)**, doi:10.1098/rsta.2014.0170, URL http://rsta.royalsocietypublishing.org/content/373/2045/20140170, http://rsta.royalsocietypublishing.org/content/373/2045/20140170.full.pdf.

Francis, J. A., and S. J. Vavrus, 2012: Evidence linking Arctic amplification to extreme weather in mid-latitudes. *Geophysical Research Letters*, **39 (6)**, doi:10.1029/2012GL051000.

Frei, C., J. Hesselbjerg Christensen, M. Déqué, D. Jacob, R. G. Jones, and P. L. Vidale, 2003: Daily precipitation statistics in regional climate models: Evaluation and intercomparison for the European Alps. *Journal of Geophysical Research: Atmospheres*, **108 (D3)**, doi:10.1029/2002JD002287, https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2002JD002287.

Früh, B., H. Feldmann, H.-J. Panitz, G. Schaedler, D. Jacob, P. Lorenz, and K. Keuler, 2010: Determination of precipitation return values in complex terrain and their evaluation. *Journal of Climate*, **23 (9)**, 2257–2274, doi:10.1175/2009JCLI2685.1.

García-Herrera, R., J. Díaz, R. Trigo, J. Luterbacher, and E. Fischer, 2010: A review of the European summer heat wave of 2003. *Critical Reviews in Environmental Science and Technology*, **40**, 267–306, doi:10.1080/10643380802238137.

Gerstengarbe, F.-W., P. Werner, and U. Rüge, 1999: Katalog der Großwetterlagen Europas (1881 - 1998) Nach Paul Hess und Helmuth Brezowsky. Tech. rep., Potsdam-Institut für Klimafolgenforschung, Deutscher Wetterdienst Offenbach a. M. 5., verbesserte und ergänzte Auflage.

Ghosh, R., W. A. Müller, J. Baehr, and J. Bader, 2017: Impact of observed North Atlantic multidecadal variations to European summer climate: a linear baroclinic response to surface heating. *Climate Dynamics*, **48 (11)**, 3547–3563, doi:10.1007/s00382-016-3283-4.

Goddard, L., and Coauthors, 2013: A verification framework for interannual-to-decadal predictions experiments. *Climate Dynamics*, **40**, 245–272, doi:10.1007/s00382-012-1481-2.

Hackenbruch, J., T. Kunz-Plapp, S. Müller, and J. W. Schipper, 2017: Tailoring climate parameters to information needs for local adaptation to climate change. *Climate*, **5 (2)**, 25, doi:10.3390/cli5020025, 12.01.02; LK 01.

Hackenbruch, J., G. Schädler, and J. W. Schipper, 2016: Added value of high-resolution regional climate simulations for regional impact studies. *Meteorologische Zeitschrift*, **25 (3)**, 291–304, doi:10.1127/metz/2016/0701.

Hanlon, H. M., G. C. Hegerl, S. F. B. Tett, and D. M. Smith, 2013: Can a decadal forecasting system predict temperature extreme indices? *Journal of Climate*, **26 (11)**, 3728–3744, doi: 10.1175/JCLI-D-12-00512.1, URL https://doi.org/10.1175/JCLI-D-12-00512.1, https://doi.org/10.1175/JCLI-D-12-00512.1.

Havenith, G., 2002: Interaction of clothing and thermoregulation. *Exogenous Dermatology*, **1 (5)**, 221–230, doi:10.1159/000068802.

Haylock, M. R., N. Hofstra, A. M. G. Klein Tank, E. J. Klok, and P. D. Jones, 2008: A European daily high-resolution gridded dataset of surface temperature and precipitation. *Journal of Geophysical Research*, **113 (113)**, doi:10.1029/2008JD10201.

Hertwig, E., J.-S. von Storch, D. Handorf, and K. Dethloff, 2015: Effect of horizontal resolution on ECHAM6-AMIP performance. *Climate Dynamics*, doi:10.1007/s00382-014-2396-x.

Hess, P., and H. Brezowsky, 1977: Katalog der Großwetterlagen Europas; 2. verb. u. erg. Aufl. Tech. Rep. 113, Deutscher Wetterdienst, Offenbach am Main.

Hirschi, M., and Coauthors, 2011: Observational evidence for soil-moisture impact on hot extremes in southeastern Europe. *Nature Geoscience*, **4**, 17–21, doi:10.1038/ngeo1032.

Hodson, D. L. R., R. T. Sutton, C. Cassou, N. Keenlyside, Y. Okumura, and T. Zhou, 2010: Climate impacts of recent multidecadal changes in Atlantic Ocean Sea Surface Temperature: a multimodel comparison. *Climate Dynamics*, **34 (7)**, 1041–1058, doi:10.1007/s00382-009-0571-2.

Honda, Y., M. Kabuto, M. Ono, and I. Uchiyama, 2007: Determination of optimum daily maximum temperature using climate data. *Environmental health and preventive medicine*, **12 (5)**, 209–216, available online at https://www.ncbi.nlm.nih.gov/pubmed/21432083, last accessed: 23.10.2018.

Höppe, P., 1999: The physiological equivalent temperature - a universal index for the biometeorological assessment of the thermal environment. *International Journal of Biometeorology*, **43 (2)**, 71–75, doi:10.1007/s004840050118.

Hurrell, J. W., Y. Kushnir, G. Ottersen, and M. Visbeck, 2003: *An Overview of the North Atlantic Oscillation*, 1–35. American Geophysical Union (AGU), doi:10.1029/134GM01.

IPCC, 2013: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1535 pp., doi:10.1017/CBO9781107415324, available online at www.climatechange2013.org.

Isaac, M., and D. P. van Vuuren, 2009: Modeling global residential sector energy demand for heating and air conditioning in the context of climate change. *Energy Policy*, **37 (2)**, 507 – 521, doi:10.1016/j.enpol.2008.09.051.

Jacob, D., and Coauthors, 2014: EURO-CORDEX: new high-resolution climate change projections for European impact research. *Regional Environmental Change*, **14 (2)**, 563–578, doi:10.1007/s10113-013-0499-2.

Jaeger, E. B., and S. I. Seneviratne, 2010: Impact of soil moisture–atmosphere coupling on European climate extremes and trends in a regional climate model. *Climate Dynamics*, **36 (9)**, 1919–1939, doi:10.1007/s00382-010-0780-8.

Johansen, O., 1977: Thermal conductivity of soils. Ph.D. thesis, University of Trondheim, Norway, (CRREL Draft Translation, 1977), ADA 044002.

Jungclaus, J. H., and Coauthors, 2013: Characteristics of the ocean simulations in the Max Planck Institute Ocean Model (MPIOM) the ocean component of the MPI-Earth system model. *Journal of Advances in Modeling Earth Systems*, **5 (2)**, 422–446, doi:10.1002/jame.20023, https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/jame.20023.

Kadow, C., S. Illing, O. Kunst, H. W. Rust, H. Pohlmann, W. A. Müller, and U. Cubasch, 2016: Evaluation of forecasts by accuracy and spread in the MiKlip decadal climate prediction system. *Meteorologische Zeitschrift*, **25 (6)**, 631–643, doi:10.1127/metz/2015/0639.

Keenlyside, N. S., M. Latif, J. Jungclaus, L. Kornblueh, and E. Roeckner, 2008: Advancing decadal-scale climate prediction in the North Atlantic sector. *Nature,* **453**, 84–, doi: 10.1038/nature06921.

Kendon, E. J., N. M. Roberts, H. J. Fowler, M. J. Roberts, S. C. Chan, and C. A. Senior, 2014: Heavier summer downpours with climate change revealed by weather forecast resolution model. *Nature Climate Change,* **4**, 570–, doi:10.1038/nclimate2258.

Kendon, E. J., N. M. Roberts, C. A. Senior, and M. J. Roberts, 2012: Realism of rainfall in a very high-resolution regional climate model. *Journal of Climate,* **25 (17)**, 5791–5806, doi: 10.1175/JCLI-D-11-00562.1, https://doi.org/10.1175/JCLI-D-11-00562.1.

Kendon, E. J., and Coauthors, 2017: Do convection-permitting regional climate models improve projections of future precipitation change? *Bulletin of the American Meteorological Society,* **98 (1)**, 79–93, doi:10.1175/BAMS-D-15-0004.1, https://doi.org/10.1175/BAMS-D-15-0004.1.

Keuler, K., K. Radtke, S. Kotlarski, and D. Lüthi, 2016: Regional climate change over europe in COSMO-CLM: Influence of emission scenario and driving global model. *Meteorologische Zeitschrift,* **25 (2)**, 121–136, doi:10.1127/metz/2016/0662.

Khodayar, S., N. Kalthoff, and G. Schädler, 2013: The impact of soil moisture variability on seasonal convective precipitation simulations. part i: validation, feedbacks, and realistic initialisation. *Meteorologische Zeitschrift,* **22 (4)**, 489–505, doi:10.1127/0941-2948/2013/0403.

Khodayar, S., A. Sehlinger, H. Feldmann, and C. Kottmeier, 2015: Sensitivity of soil moisture initialization for decadal predictions under different regional climatic conditions in europe. *International Journal of Climatology,* **35 (8)**, 1899–1915, doi:10.1002/joc.4096.

Knight, J. R., C. K. Folland, and A. A. Scaife, 2006: Climate impacts of the Atlantic Multidecadal Oscillation. *Geophysical Research Letters,* **33 (17)**, doi:10.1029/2006GL026242, https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2006GL026242.

Knote, C., and G. R. Heinemann, 2010: Changes in weather extremes: Assessment of return values using high resolution climate simulations at convection-resolving scale. *Meteorologische Zeitschrift,* **19 (1)**, 11–23, doi:10.1127/0941-2948/2010/0424.

Knudsen, M. F., M.-S. Seidenkrantz, B. H. Jacobsen, and A. K., 2011: Tracking the Atlantic Multidecadal Oscillation through the last 8,000 years. *Nature Communications,* **2**, 178–, doi: 10.1038/ncomms1186.

Kohler, M., G. Schädler, L. Gantner, N. Kalthoff, F. Königer, and C. Kottmeier, 2012: Validation of two SVAT models for different periods during the West African monsoon. *Meteorologische Zeitschrift,* **21**, 509–524, doi:10.1127/0941-2948/2012/0490.

Koster, R. D., and M. J. Suarez, 2001: Soil moisture memory in climate models. *Journal of Hydrometeorology*, **2 (6)**, 558–570, doi:10.1175/1525-7541(2001)002<0558:SMMICM>2. 0.CO;2.

Koster, R. D., and Coauthors, 2004: Regions of strong coupling between soil moisture and precipitation. *Science*, **305 (5687)**, 1138–1140, doi:10.1126/science.1100217, http://science. sciencemag.org/content/305/5687/1138.full.pdf.

Koster, R. D., and Coauthors, 2010: Contribution of land surface initialization to subseasonal forecast skill: First results from a multi-model experiment. *Geophysical Research Letters*, **37 (2)**, doi:10.1029/2009GL041677, URL https://agupubs.onlinelibrary.wiley. com/doi/abs/10.1029/2009GL041677, https://agupubs.onlinelibrary.wiley.com/doi/pdf/10. 1029/2009GL041677.

Kothe, S., J. Tödter, and B. Ahrens, 2016: Strategies for soil initialization of regional decadal climate predictions. *Meteorologische Zeitschrift*, **25 (6)**, 775–794, doi:10.1127/metz/2016/ 0729.

Kotlarski, S., and Coauthors, 2014: Regional climate modeling on European scales: a joint standard evaluation of the EURO-CORDEX RCM ensemble. *Geoscientific Model Development*, **7 (4)**, 1297–1333, doi:10.5194/gmd-7-1297-2014.

Kröger, J., and Coauthors, 2018: Full-field initialized decadal predictions with the MPI earth system model: an initial shock in the North Atlantic. *Climate Dynamics*, **51 (7)**, 2593–2608, doi:10.1007/s00382-017-4030-1.

Kruschke, T., H. W. Rust, C. Kadow, W. A. Müller, H. Pohlmann, G. C. Leckebusch, and U. Ulbrich, 2016: Probabilistic evaluation of decadal prediction skill regarding Northern Hemisphere winter storms. *Meteorologische Zeitschrift*, **25 (6)**, 721–738, doi:10.1127/metz/2015/0641.

Kyselý, J., 2002: Temporal fluctuations in heat waves at Prague-Klementinum, the Czech Republic, from 1901-97, and their relationships to atmospheric circulation. *International Journal of Climatology*, **22 (1)**, 33–50, doi:10.1002/joc.720, https://rmets.onlinelibrary.wiley.com/doi/ pdf/10.1002/joc.720.

Kyselý, J., 2008: Influence of the persistence of circulation patterns on warm and cold temperature anomalies in Europe: Analysis over the 20th century. *Global and Planetary Change*, **62 (1)**, 147 – 163, doi:https://doi.org/10.1016/j.gloplacha.2008.01.003.

Kyselý, J., and P. Domonkos, 2006: Recent increase in persistence of atmospheric circulation over Europe: comparison with long-term variations since 1881. *International Journal of Climatology*, **26 (4)**, 461–483, doi:10.1002/joc.1265, https://rmets.onlinelibrary.wiley.com/doi/ pdf/10.1002/joc.1265.

Kyselý, J., and E. Plavcová, 2010: A critical remark on the applicability of E-OBS European gridded temperature data set for validating control climate simulations. *Journal of Geophysical Research: Atmospheres*, **115 (D23)**, doi:10.1029/2010JD014123, https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2010JD014123.

La Agencia Estatal de Meteorología (AEMET), 2017: Extreme values. Córdoba Aeropuerto. Available online at http://www.aemet.es/en/serviciosclimaticos/datosclimatologicos/efemerides_extremos*?w=0&k=and&l=5402&datos=det&x=5402&m=7&v=todos.

Landesanstalt für Umwelt Baden-Württemberg (LUBW), 2012: Wasserhaushaltssimulationen mit COSMO-CLM-Daten Version 4.8 (run1, linear skaliert) für das Rhein-Einzugsgebiet bis zum Pegel Worms. Unpublished data set, HYDRON GmbH im Auftrag der Landesanstalt für Umwelt Baden-Württemberg.

Lang, B., 2015: VADYtele. Tech. rep., Institut für Geographie, Universität Augsburg. URL https://www-miklip.dkrz.de/about/vadytele/, plugin of the Freva (Freie Universität Berlin Evaluation) System.

Lau, N.-C., and M. J. Nath, 2014: Model simulation and projection of European heat waves in present-day and future climates. *Journal of Climate*, **27 (10)**, 3713–3730, doi:10.1175/JCLI-D-13-00284.1, https://doi.org/10.1175/JCLI-D-13-00284.1.

Lau, W. K. M., and K.-M. a. Kim, 2012: The 2010 Pakistan Flood and Russian heat wave: Teleconnection of hydrometeorological extremes. *Journal of Hydrometeorology*, **13 (1)**, 392–403, doi:10.1175/JHM-D-11-016.1.

Lehmann, J., D. Coumou, and K. Frieler, 2015: Increased record-breaking precipitation events under global warming. *Climatic Change*, **132 (4)**, 501–515, doi:10.1007/s10584-015-1434-y.

Lemos, M. C., C. J. Kirchhoff, and V. Ramprasad, 2012: Narrowing the climate information usability gap. *Nature Climate Change*, **2**, 789, doi:10.1038/nclimate1614.

Lhotka, O., J. Kyselý, and E. Plavcová, 2018: Evaluation of major heat waves' mechanisms in EURO-CORDEX RCMs over Central Europe. *Climate Dynamics*, **50**, 4249–4262, doi:10.1007/s00382-017-3873-9.

Löpmeier, F.-J., 2014: Agrarmeteorologisches Modell zur Berechnung der aktuellen Verdunstung (AMBAV). Tech. rep., Deutscher Wetterdienst. Available online at ftp://ftp-cdc.dwd.de/pub/CDC/grids_germany/monthly/soil_moist/AMBAV.pdf.

Lorenz, R., E. B. Jaeger, and S. I. Seneviratne, 2010: Persistence of heat waves and its link to soil moisture memory. *Geophysical Research Letters*, **37 (9)**, doi:10.1029/2010GL042764, l09703.

Luo, D., 2005: Why is the North Atlantic block more frequent and long-lived during the negative NAO phase? *Geophysical Research Letters*, **32**, 20 804–, doi:10.1029/2005GL022927.

Maraun, D., and Coauthors, 2010: Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user. *Reviews of Geophysics*, **48**, doi:10.1029/2009RG000314.

Marotzke, J., and Coauthors, 2016: Miklip: A national research project on decadal climate prediction. *Bulletin of the American Meteorological Society*, **97 (12)**, 2379–2394, doi:10.1175/BAMS-D-15-00184.1.

Martin, G. M., and Coauthors, 2011: The HadGEM2 family of Met Office Unified Model climate configurations. *Geoscientific Model Development*, **4 (3)**, 723–757, doi:10.5194/gmd-4-723-2011.

Masato, G., B. J. Hoskins, and T. Woollings, 2013: Winter and Summer Northern Hemisphere Blocking in CMIP5 Models. *Journal of Climate*, **26 (18)**, 7044–7059, doi:10.1175/JCLI-D-12-00466.1, https://doi.org/10.1175/JCLI-D-12-00466.1.

Matzarakis, A., 2001: Die thermische Komponente des Stadtklimas. Tech. rep., Berichte des Meteorologischen Institutes der Universität Freiburg.

Matzarakis, A., H. Mayer, and M. G. Iziomon, 1999: Applications of a universal thermal index: physiological equivalent temperature. *International Journal of Biometeorology*, **43 (2)**, 76–84, doi:10.1007/s004840050119.

Mauritzen, C., T. Zivkovic, and V. Veldore, 2017: On the relationship between climate sensitivity and modelling uncertainty. *Tellus A: Dynamic Meteorology and Oceanography*, **69 (1)**, 1327 765, doi:10.1080/16000870.2017.1327765.

Meehl, G. A., and C. Tebaldi, 2004: More Intense, More Frequent, and Longer Lasting Heat Waves in the 21st Century. *Science*, **305**, 994–997, doi:10.1126/science.1098704.

Meehl, G. A., and Coauthors, 2009: Decadal Prediction. *Bulletin of the American Meteorological Society*, **90 (10)**, 1467–1486, doi:10.1175/2009BAMS2778.1, https://doi.org/10.1175/2009BAMS2778.1.

Meissner, C., 2008: High-resolution sensitivity studies with the regional climate model COSMO-CLM. Ph.D. thesis, Karlsruher Institut für Technolgie - Institut für Meteorologie und Klimaforschung.

Met Office, 2018: Summer temperature 2018 - the 'new normal'? Available online at https://blog.metoffice.gov.uk/2018/07/27/summer-temperature-2018-the-new-normal/, last accessed: 24.09.2018, Press Office.

Miralles, D. G., M. J. den Berg, A. J. Teuling, and R. A. M. Jeu, 2012: Soil moisture-temperature coupling: A multiscale observational analysis. *Geophysical Research Letters*, **39 (21)**, doi:10.1029/2012GL053703, https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2012GL053703.

Miralles, D. G., P. Gentine, S. I. Seneviratne, and A. J. Teuling, 2018: Land-atmospheric feedbacks during droughts and heatwaves: state of the science and current challenges. *Annals of the New York Academy of Sciences*, doi:10.1111/nyas.13912.

Miralles, D. G., A. J. Teuling, C. C. van Heerwaarden, and J. Vila-Guerau de Arellano, 2014: Mega-heatwave temperatures due to combined soil desiccation and atmospheric heat accumulation. *Nature Geosci*, **7 (5)**, 345–349, doi:10.1038/ngeo2141.

Mora, C., and Coauthors, 2017: Global risk of deadly heat. *Nature Climate Change*, **7**, 501–506, doi:10.1038/nclimate3322.

Moss, R. H., and Coauthors, 2010: The next generation of scenarios for climate change research and assessment. *Nature*, **463**, 747–, URL http://dx.doi.org/10.1038/nature08823.

Müller, W. A., and Coauthors, 2012: Forecast skill of multi-year seasonal means in the decadal prediction system of the Max Planck Institute for Meteorology. *Geophysical Research Letters*, **39**, doi:10.1029/2012GL053326.

Müller, W. A., and Coauthors, 2018: A higher-resolution version of the Max Planck Institute Earth System Model (MPI-ESM1.2-HR). *Journal of Advances in Modeling Earth Systems*, **10 (7)**, 1383–1413, doi:10.1029/2017MS001217, https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2017MS001217.

MunichRe NatCatSERVICE, 2015: Heat wave, drought, wildfires in Russia (Summer 2010) MR Touch Natural hazards - Event report. Tech. rep., Geo Risks Research, Münchener Rückversicherungs-Gesellschaft. Last accessed 05. Sep. 2018 , available online at https://www.munichre.com/site/touch-naturalhazards/get/documents_E-132629599/mr/assetpool.shared/Documents/5_Touch/_NatCatService/Catastrophe_portraits/event-report-hw-dr-wf-russia-touch-en-update.pdf.

MunichRe NatCatSERVICE, 2018: NatCatSERVICE Online Tool Report: Heatwave / wildfire events worldwide 2003. Last accessed 21.02.2019.

Murphy, A. H., 1988: Skill scores based on the mean square error and their relationships to the correlation coefficient. *Monthly Weather Review*, **116 (12)**, 2417–2424, doi:10.1175/1520-0493(1988)116<2417:SSBOTM>2.0.CO;2, https://doi.org/10.1175/1520-0493(1988)116<2417:SSBOTM>2.0.CO;2.

Muthers, S., G. Laschewski, and A. Matzarakis, 2017: The summers 2003 and 2015 in South-West Germany: Heatwaves and heat-related mortality in the context of climate change. *Atmosphere*, **8**, doi:10.3390/atmos8110224.

Nairn, J. R., and R. J. B. Fawcett, 2015: The excess heat factor: A metric for heatwave intensity and its use in classifying heatwave severity. *International Journal of Environmental Research and Public Health*, **12**, 227–253, doi:10.3390/ijerph120100227.

Niu, G.-Y., and Coauthors, 2011: The community Noah land surface model with multipa-rameterization options (Noah-MP): 1. model description and evaluation with local-scale measurements. *Journal of Geophysical Research: Atmospheres*, **116 (D12)**, n/a–n/a, doi: 10.1029/2010JD015139, d12109.

NOAA / National Weather Service Climate Prediction Center, 2018a: East Atlantic (EA) pattern standardized 3-month running mean index. Available online at http://www.cpc.ncep.noaa.gov/data/teledoc/ea_ts.shtml, last accessed 21.11.2018.

NOAA / National Weather Service Climate Prediction Center, 2018b: Northern hemisphere teleconnection patterns. Available online at http://www.cpc.ncep.noaa.gov/data/teledoc/telecontents.shtml.

Oleson, K., and Coauthors, 2013: Technical description of version 4.5 of the Community Land Model (CLM). Tech. rep., National Center for Atmospheric Research, Boulder, Colorado. doi: doi:10.5065/D6RR1W7M., nCAR Technical Note NCAR/TN-503+STR.

Ono, M., 2013: Heat stroke and the thermal environment. *Japan Medical Association Journal*, **56 (3)**, 199–205, available online at https://www.med.or.jp/english/journal/pdf/2013_03/199_205.pdf, last accessed: 23.10.2018.

Orlowsky, B., and S. I. Seneviratne, 2012: Global changes in extreme events: regional and seasonal dimension. *Climatic Change*, **110 (3)**, 669–696, doi:10.1007/s10584-011-0122-9.

ORNL DAAC, 2015: FLUXNET Web Page. Oak Ridge, Tennessee, U.S.A., available online at fluxnet.ornl.gov.

Ouzeau, G., J.-M. Soubeyroux, M. Schneider, R. Vautard, and S. Planton, 2016: Heat waves analysis over France in present and future climate: Application of a new method on the EURO-CORDEX ensemble. *Climate Services*, **4**, 1–12, doi:10.1016/j.cliser.2016.09.002.

Paeth, H., J. Li, F. Pollinger, W. A. Müller, H. Pohlmann, H. Feldmann, and H.-J. Panitz, 2018: An effective drift correction for dynamical downscaling of decadal global climate predictions. *Climate Dynamics*, doi:10.1007/s00382-018-4195-2.

Parsons, K., 2002: *Human Thermal Envirnoments: The Effects of Hot, Moderate, and Cold Environments on Human Health, Comfort and Performance*. CRC Press.

Pasternack, A., 2018: Recalibration. Available online at https://www-miklip.dkrz.de/plugins/recalibration/detail/, accessed last: 28.09.2018.

Pasternack, A., J. Bhend, M. A. Liniger, H. W. Rust, W. A. Müller, and U. Ulbrich, 2018: Parametric decadal climate forecast recalibration (DeFoReSt 1.0). *Geoscientific Model Development*, **11 (1)**, 351–368, doi:10.5194/gmd-11-351-2018.

Perez, J., M. Menendez, F. J. Mendez, and I. J. Losada, 2014: Evaluating the performance of CMIP3 and CMIP5 global climate models over the north-east Atlantic region. *Climate Dynamics*, **43**, 2663–2680, doi:10.1007/s00382-014-2078-8.

Perkins, S. E., and L. V. Alexander, 2013: On the measurement of heat waves. *Journal of Climate*, **26 (13)**, 4500–4517, doi:10.1175/JCLI-D-12-00383.1, https://doi.org/10.1175/JCLI-D-12-00383.1.

Petoukhov, V., S. Rahmstorf, S. Petri, and H. J. Schellnhuber, 2013: Quasiresonant amplification of planetary waves and recent Northern Hemisphere weather extremes. *Proceedings of the National Academy of Sciences*, **110 (14)**, 5336–5341, doi:10.1073/pnas.1222000110, http://www.pnas.org/content/110/14/5336.full.pdf.

Pinto, J. G., and C. C. Raible, 2012: Past and recent changes in the North Atlantic oscillation. *Wiley Interdisciplinary Reviews: Climate Change*, **3 (1)**, 79–90, doi:10.1002/wcc.150.

Pohlmann, H., and Coauthors, 2013: Improved forecast skill in the tropics in the new MiKlip decadal climate predictions. *Geophysical Research Letters*, **40 (21)**, 5798–5802, doi:10.1002/2013GL058051, https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2013GL058051.

Poli, P., and Coauthors, 2016: ERA-20C: An atmospheric reanalysis of the twentieth century. *Journal of Climate*, **29 (11)**, 4083–4097, doi:10.1175/JCLI-D-15-0556.1, https://doi.org/10.1175/JCLI-D-15-0556.1.

Polyakov, I. V., V. A. Alexeev, U. S. Bhatt, E. I. Polyakova, and X. Zhang, 2010: North Atlantic warming: patterns of long-term trend and multidecadal variability. *Climate Dynamics*, **34 (2)**, 439–457, doi:10.1007/s00382-008-0522-3.

Quesada, B., R. Vautard, P. Yiou, M. Hirschi, and S. I. Seneviratne, 2012: Asymmetric European summer heat predictability from wet and dry southern winters and springs. *Nature Climate Change*, **2**, 736–, doi:10.1038/nclimate1536.

Raäisaänen, J., 2007: How reliable are climate models? *Tellus A: Dynamic Meteorology and Oceanography*, **59 (1)**, 2–29, doi:10.1111/j.1600-0870.2006.00211.x.

Rahmstorf, S., and D. Coumou, 2011: Increase of extreme events in a warming world. *Proceedings of the National Academy of Sciences*, **108 (44)**, 17 905–17 909, doi:10.1073/pnas.1101766108.

Årthun, M., E. W. Kolstad, T. Eldevik, and N. S. Keenlyside, 2018: Time scales and sources of european temperature variability. *Geophysical Research Letters*, **45 (8)**, 3597–3604, doi:10.1002/2018GL077401, https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2018GL077401.

Raschendorfer, M., 2001: The new turbulence parameterization of LM. *COSMO Newsletter*, **1**, 89–97, URL http://www.cosmo-model.org/content/model/documentation/newsLetters/newsLetter01/newsLetter_01.pdf.

Rasmijn, L. M., G. van der Schrier, R. Bintanja, J. Barkmeijer, A. Sterl, and W. Hazeleger, 2018: Future equivalent of 2010 Russian heatwave intensified by weakening soil moisture constraints. *Nature Climate Change*, **8 (5)**, 381–385, doi:10.1038/s41558-018-0114-0.

Rauthe, M., H. Steiner, U. Riediger, A. Mazurkiewicz, and A. Gratzki, 2013: A Central European precipitation climatology - part i: Generation and validation of a high-resolution gridded daily data set (HYRAS). *Meteorologische Zeitschrift*, **22**, 235–256, doi:10.1127/0941-2948/2013/0436.

Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan, 2003: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research: Atmospheres*, **108 (D14)**, doi:10.1029/2002JD002670, https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2002JD002670.

Reick, C. H., T. Raddatz, V. Brovkin, and V. Gayler, 2013: Representation of natural and anthropogenic land cover change in MPI-ESM. *Journal of Advances in Modeling Earth Systems*, **5 (3)**, 459–482, doi:10.1002/jame.20022, https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/jame.20022.

Rex, D. F., 1950: Blocking action in the Middle Troposphere and its effect upon regional climate. *Tellus*, **2**, 275–301, doi:10.1111/j.2153-3490.1950.tb00339.x.

Reyers, M., and Coauthors, 2019: Development and prospects of the regional MiKlip decadal prediction system over Europe: Predictive skill, added value of regionalization and ensemble size dependency. *Earth System Dynamics Discussions*, **2017**, 1–27, doi:10.5194/esd-2017-70, URL https://www.earth-syst-dynam-discuss.net/esd-2017-70/.

Ritter, B., and J.-F. Geleyn, 1992: A comprehensive radiation scheme for numerical weather prediction models with potential applications in climate simulations. *Monthly Weather Review*, **120**, 303–325, doi:10.1175/1520-0493(1992)120<0303:ACRSFN>2.0.CO;2.

Robine, J.-M., K. Cheung, S. Le Roy, H. Oyen, and F. Herrmann, 2007: Report on excess mortality in Europe during summer 2003. Tech. rep., EU Community Action Programme for Public Health, Grant Agreement 2005114.

Robinson, P. J., 2001: On the definition of a heat wave. *Journal of Applied Meteorology*, **40 (4)**, 762–775, doi:10.1175/1520-0450(2001)040<0762:OTDOAH>2.0.CO;2, https://doi.org/10.1175/1520-0450(2001)040<0762:OTDOAH>2.0.CO;2.

Rockel, B., A. Will, and B. Rockel, 2008: The regional climate model COSMO-CLM (CCLM). *Meteorologische Zeitschrift*, **4 (4)**, doi:10.1127/0941-2948/2008/0309, special issue, ISSN 0941-2948.

Rodríguez-Puebla, C., A. H. Encinas, L. A. García-Casado, and S. Nieto, 2010: Trends in warm days and cold nights over the Iberian Peninsula: relationships to large-scale variables. *Climatic Change*, **100 (3)**, 667–684, doi:10.1007/s10584-009-9721-0.

Roeckner, G., and Coauthors, 2003: The atmospheric general circulation model ECHAM 5. PART I: Model description. Tech. rep., Max-Planck-Institut für Meteorologie, Bundesstr. 55, D-20146 Hamburg, Germany.

Ronaldson, K. J., 2016: Chapter 11 - heat stroke and rhabdomyolysis. *Life-Threatening Effects of Antipsychotic Drugs*, P. Manu, R. J. Flanagan, and K. J. Ronaldson, Eds., Academic Press, San Diego, 241 – 251, doi:10.1016/B978-0-12-803376-0.00011-3.

Ruiz-Barradas, A., S. Nigam, and A. Kavvada, 2013: The Atlantic Multidecadal Oscillation in twentieth century climate simulations: uneven progress from CMIP3 to CMIP5. *Climate Dynamics*, **41 (11)**, 3301–3315, doi:10.1007/s00382-013-1810-0.

Ruosteenoja, K., T. Markkanen, A. Venäläinen, P. Räisänen, and H. Peltola, 2018: Seasonal soil moisture and drought occurrence in Europe in CMIP5 projections for the 21st century. *Climate Dynamics*, **50 (3)**, 1177–1192, doi:10.1007/s00382-017-3671-4.

Ruprich-Robert, Y., and C. Cassou, 2015: Combined influences of seasonal East Atlantic Pattern and North Atlantic Oscillation to excite Atlantic multidecadal variability in a climate model. *Climate Dynamics*, **44 (1)**, 229–253, doi:10.1007/s00382-014-2176-7.

Russo, S., J. Sillmann, and E. M. Fischer, 2015: Top ten European heatwaves since 1950 and their occurrence in the coming decades. *Environmental Research Letters*, **10 (12)**, 124 003, doi:10.1088/1748-9326/10/12/124003.

Rust, H. W., M. Fischer, C. Kadow, S. Illing, and O. Kunst, 2014: Assessing accuracy and ensemble spread with MurCSS. Available online at https://www-miklip.dkrz.de/about/murcss/, last accessed: 12.09.2018.

Rust, H. W., A. Richling, P. Bissolli, and U. Ulbrich, 2015: Linking teleconnection patterns to European temperature - a multiple linear regression model. *Meteorologische Zeitschrift*, **24 (4)**, 411–423, doi:10.1127/metz/2015/0642.

Samaniego, L., and Coauthors, 2018: Anthropogenic warming exacerbates European soil moisture droughts. *Nature Climate Change*, **8 (5)**, 421–426, doi:10.1038/s41558-018-0138-5.

Sanchez-Benitez, A., R. Garcia-Herrera, D. Barriopedro, P. M. Sousa, and R. M. Trigo, 2018: June 2017: The Earliest European summer mega-heatwave of reanalysis period. *Geophysical Research Letters*, **45 (4)**, 1955–1962, doi:10.1002/2018GL077253, https://agupubs. onlinelibrary.wiley.com/doi/pdf/10.1002/2018GL077253.

Scaife, A. A., T. Woollings, J. Knight, G. Martin, and T. Hinton, 2010: Atmospheric blocking and mean biases in climate models. *Journal of Climate*, **23 (23)**, 6143–6152, doi:10.1175/ 2010JCLI3728.1, https://doi.org/10.1175/2010JCLI3728.1.

Scaife, A. A., and Coauthors, 2014: Skillful long-range prediction of European and North American winters. *Geophysical Research Letters*, **41 (7)**, 2514–2519, doi:10.1002/2014GL059637, https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2014GL059637.

Schädler, G., 1990: Numerische Simulationen zur Wechselwirkung zwischen Landoberfläche und atmosphärischer Grenzschicht. Ph.D. thesis, Karlsruher Institut für Technologie.

Schär, C., P. L. Vidale, D. Lüthi, C. Frei, C. Haberli, M. A. Liniger, and C. Appenzeller, 2004: The role of increasing temperature variability in European summer heatwaves. *Nature*, **427 (6972)**, 332–336, doi:10.1038/nature02300.

Schär, C., and Coauthors, 2016: Percentile indices for assessing changes in heavy precipitation events. *Climatic Change*, **137**, 201–216, doi:10.1007/s10584-016-1669-2.

Schenk, H. J., and R. B. Jackson, 2003: Global distribution of root profiles in terrestrial ecosystems. Oak Ridge, Tennessee, U.S.A, data set, Available online at https://daac.ornl.gov/VEGETATION/guides/global_root_profiles.html, doi:doi:10.3334/ORNLDAAC/660.

Schlesinger, M. E., and N. Ramankutty, 1994: An oscillation in the global climate system of period 65-70 years. *Nature*, **367**, 723–, doi:10.1038/367723a0.

Schoetter, R., J. Cattiaux, and H. Douville, 2015: Changes of western European heat wave characteristics projected by the CMIP5 ensemble. *Climate Dynamics*, **45 (5)**, 1601–1616, doi:10.1007/s00382-014-2434-8.

Schrodin, E., and E. Heise, 2002: A new multi-layer soil model. *COSMO newsletter*, **2**, 149–151, available online at http://www.cosmo-model.org/content/model/documentation/newsLetters/newsLetter02/newsLetter_02.pdf.

Schubert, S., H. Wang, and M. Suarez, 2011: Warm season subseasonal variability and climate extremes in the Northern Hemisphere: The role of stationary rossby waves. *Journal of Climate*, **24 (18)**, 4773–4792, doi:10.1175/JCLI-D-10-05035.1.

Schubert, S. D., H. Wang, R. D. Koster, M. J. Suarez, and P. Y. Groisman, 2014: Northern Eurasian heat waves and droughts. *Journal of Climate*, **27 (9)**, 3169–3207, doi:10.1175/JCLI-D-13-00360.1, https://doi.org/10.1175/JCLI-D-13-00360.1.

Scoccimarro, E., G. Villarini, M. Vichi, M. Zampieri, P. G. Fogli, A. Bellucci, and S. Gualdi, 2015: Projected changes in intense precipitation over Europe at the daily and subdaily time scales. *Journal of Climate*, **28 (15)**, 6193–6203, doi:10.1175/JCLI-D-14-00779.1, https://doi.org/10.1175/JCLI-D-14-00779.1.

Screen, J. A., and I. Simmonds, 2014: Amplified mid-latitude planetary waves favour particular regional weather extremes. *Nature Climate Change*, **4**, 704–, doi:10.1038/nclimate2271.

Sedlmeier, K., 2015: Near future changes of compound extreme events from an ensemble of regional climate simulations. Ph.D. thesis, Karlsruher Institute for Technology, doi:10.5445/IR/1000050912.

Seneviratne, S., and Coauthors, 2006a: Soil moisture memory in AGCM Simulations: Analysis of Global Land-Atmosphere Coupling Experiment (GLACE) Data. *Journal of Hydrometeorology*, **7 (5)**, 1090–1112, doi:10.1175/JHM533.1, https://doi.org/10.1175/JHM533.1.

Seneviratne, S. I., T. Corti, E. L. Davin, M. Hirschi, E. B. Jaeger, I. Lehner, B. Orlowsky, and A. J. Teuling, 2010: Investigating soil moisture-climate interactions in a changing climate: A review. *Earth-Science Reviews*, **99 (3)**, 125–161, doi:https://doi.org/10.1016/j.earscirev.2010.02.004.

Seneviratne, S. I., D. Luthi, M. Litschi, and C. Schär, 2006b: Land-atmosphere coupling and climate change in Europe. *Nature*, **443 (7108)**, 205–209, doi:10.1038/nature05095.

Seneviratne, S. I., and R. Stöckli, 2008: *The Role of Land-Atmosphere Interactions for Climate Variability in Europe*, 179–193. Springer Netherlands, Dordrecht, doi:10.1007/978-1-4020-6766-2_12.

Sienz, F., W. A. Müller, and H. Pohlmann, 2016: Ensemble size impact on the decadal predictive skill assessment. *Meteorologische Zeitschrift*, **25 (6)**, 645–655, doi:10.1127/metz/2016/0670.

Smith, D. M., S. A. A., and B. P. Kirtman, 2012: What is the current state of scientific knowledge with regard to seasonal and decadal forecasting? *Environmental Research Letters*, **7 (1)**, doi:10.1088/1748-9326/7/1/015602.

Sousa, P. M., R. M. Trigo, D. Barriopedro, P. M. M. Soares, and J. A. Santos, 2018: European temperature responses to blocking and ridge regional patterns. *Climate Dynamics*, **50 (1)**, 457–477, doi:10.1007/s00382-017-3620-2.

Steadman, R. G., 1979: The assessment of sultriness. part i: A temperature-humidity index based on human physiology and clothing science. *Journal of Applied Meteorology*, **18 (7)**, 861–873, doi:10.1175/1520-0450(1979)018<0861:TAOSPI>2.0.CO;2.

Stefanon, M., F. D'Andrea, and P. Drobinski, 2012: Heatwave classification over Europe and the Mediterranean region. *Environmental Research Letters*, **7 (1)**, 014 023, doi:10.1088/1748-9326/7/1/014023.

Stevens, B., and Coauthors, 2013: Atmospheric component of the MPI-M Earth System Model: ECHAM6. *Journal of Advances in Modeling Earth Systems*, **5 (2)**, 146–172, doi:10.1002/jame.20015, https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/jame.20015.

Stöckli, R., and Coauthors, 2008: Use of FLUXNET in the Community Land Model development. *Journal of Geophysical Research: Biogeosciences*, **113 (G1)**, doi:10.1029/2007JG000562, g01025.

Stott, P. A., D. A. Stone, and M. R. Allen, 2004: Human contribution to the European heatwave of 2003. *Nature*, **432**, 610–, doi:10.1038/nature03089.

Sutton, R. T., and B. Dong, 2012: Atlantic ocean influence on a shift in European climate in the 1990s. *Nature Geoscience*, **5**, 788–, doi:10.1038/ngeo1595.

Sutton, R. T., and D. L. R. Hodson, 2005: Atlantic ocean forcing of North American and European summer climate. *Science*, **309 (5731)**, 115–118, doi:10.1126/science.1109496, http://science.sciencemag.org/content/309/5731/115.full.pdf.

Tiedtke, M., 1989: A comprehensive mass flux scheme for cumulus parameterization in large-scale models. *Monthly Weather Review*, **117 (8)**, 1779–1800, doi: 10.1175/1520-0493(1989)117<1779:ACMFSF>2.0.CO;2, URL http://dx.doi.org/10. 1175/1520-0493(1989)117<1779:ACMFSF>2.0.CO;2, http://dx.doi.org/10.1175/ 1520-0493(1989)117<1779:ACMFSF>2.0.CO;2.

Trenberth, K. E., A. Dai, R. M. Rasmussen, and D. B. Parsons, 2003: The changing character of precipitation. *Bulletin of the American Meteorological Society*, **84 (9)**, 1205–1218, doi:10. 1175/BAMS-84-9-1205, https://doi.org/10.1175/BAMS-84-9-1205.

Trenberth, K. E., and J. T. Fasullo, 2012: Climate extremes and climate change: The Russian heat wave and other climate extremes of 2010. *Journal of Geophysical Research: Atmospheres*, **117 (D17)**, doi:10.1029/2012JD018020.

Trenberth, K. E., and D. A. Paolino, 1980: The Northern Hemisphere sea level pressure data set: Trends, errors and discontinuities. *Monthly Weather Review*, **108**, 855–872, doi:10.1175/ 1520-0493(1980)108<0855:TNHSLP>2.0.CO;2.

Trouet, V., F. Babst, and M. Meko, 2018: Recent enhanced high-summer North Atlantic jet variability emerges from three-century context. *Nature communications*, **9 (1)**, 180–180, doi: 10.1038/s41467-017-02699-3.

Uhlig, M., 2016: Regional decadal climate predictions for Europe – feasibility m& skill. Ph.D. thesis, Karlsruher Institut für Technologie (KIT), doi:10.5445/IR/1000069100, 12.01.02; LK 01.

Uppala, S. M., and Coauthors, 2005: The ERA-40 re-analysis. *Quaterly Journal of the Royal Meteorological Society*, **131 (612)**, 2961–3012, doi:10.1256/qj.04.176.

Valcke, S., 2013: The OASIS3 coupler: a European climate modelling community software. *Geoscientific Model Development*, **6 (2)**, 373–388, doi:10.5194/gmd-6-373-2013.

van den Hurk, B., M. Best, P. Dirmeyer, A. Pitman, J. Polcher, and J. Santanello, 2011: Acceleration of land surface model development over a decade of glass. *Bulletin of the American Meteorological Society*, **92 (12)**, 1593–1600, doi:10.1175/BAMS-D-11-00007.1.

van den Hurk, B., P. Viterbo, A. Beljaars, and A. Betts, 2000: Offline validation of the ERA40 surface scheme. Technical Memorandum 295, European Centre for Medium-Range Weather Forecasts.

van Genuchten, M., 1980: A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Science Society of America Journal*, **44 (5)**, 892–898, doi:10.2136/sssaj1980.03615995004400050002x.

van Ulden, A. P., and G. J. van Oldenborgh, 2006: Large-scale atmospheric circulation biases and changes in global climate model simulations and their importance for climate change in Central Europe. *Atmospheric Chemistry and Physics*, **6 (4)**, 863–881, doi:10.5194/acp-6-863-2006.

van Vuuren, D. P., and Coauthors, 2011: The representative concentration pathways: an overview. *Climatic Change*, **109 (1-2)**, 5–31, doi:10.1007/s10584-011-0148-z.

Vautard, R., and Coauthors, 2007: Summertime European heat and drought waves induced by wintertime Mediterranean rainfall deficit. *Geophysical Research Letters*, **34 (7)**, doi:10.1029/2006GL028001, l07711.

Vautard, R., and Coauthors, 2013: The simulation of European heat waves from an ensemble of regional climate models within the EURO-CORDEX project. *Climate Dynamics*, **41 (9)**, 2555–2575, doi:10.1007/s00382-013-1714-z.

Viterbo, P., and A. Beljaars, 1995: An improved land surface parameterization scheme in the ECMWF model and its validation. *Journal of Climate*, **8**, 2716–2748, doi:10.1175/1520-0442(1995)008<2716:AILSPS>2.0.CO;2.

Weedon, G., G. Balsamo, N. Bellouin, S. Gomes, M. Best, and P. Viterbo, 2014: The WFDEI meteorological forcing data set: WATCH Forcing Data methodology applied to ERA-Interim reanalysis data. *Water Resources Research*, **50**, 7505–7514, doi:10.1002/2014WR015638.

Weedon, G., S. Gomes, P. Viterbo, H. Österle, J. Adam, N. Bellouin, O. Boucher, and M. Best, 2010: The WATCH Forcing Data 1958-2001: a meteorological forcing dataset for land surface- and hydrological models. *WATCH Technical Report*, **22**, available online at http://www.eu-watch.org/publications/technical-reports.

Weedon, G., and Coauthors, 2011: Creation of the WATCH Forcing data and its use to assess global and regional reference crop evaporation over land during the twentieth century. *Journal of Hydrometeorology*, **12**, 823–848, doi:10.1175/2011JHM1369.1.

Weisheimer, A., F. J. Doblas-Reyes, T. Jung, and T. N. Palmer, 2011: On the predictability of the extreme summer 2003 over Europe. *Geophysical Research Letters*, **38 (5)**, doi:10.1029/2010GL046455, URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2010GL046455, https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2010GL046455.

Wilks, D. S., 2011: *Statistical methods in the atmospheric sciences*, Vol. 3rd edition. Academic Press, San Diego, CA.

Will, A., and Coauthors, 2017: The COSMO-CLM 4.8 regional climate model coupled to regional ocean, land surface and global earth system models using OASIS3-MCT: description and performance. *Geoscientific Model Development*, **10 (4)**, 1549–1586, doi:10.5194/gmd-10-1549-2017.

Wójcik, R., 2014: Reliability of CMIP5 GCM simulations in reproducing atmospheric circulation over Europe and the North Atlantic: a statistical downscaling perspective. *International Journal of Climatology*, **35 (5)**, 714–732, doi:10.1002/joc.4015, https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/joc.4015.

Woollings, T., A. Hannachi, and B. Hoskins, 2010: Variability of the North Atlantic eddy-driven jet stream. *Quarterly Journal of the Royal Meteorological Society*, **136 (649)**, 856–868, doi:10.1002/qj.625, https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.625.

Woollings, T., B. Hoskins, M. Blackburn, and P. Berrisford, 2008: A new rossby wave-breaking interpretation of the North Atlantic Oscillation. *Journal of the Atmospheric Sciences*, **65 (2)**, 609–626, doi:10.1175/2007JAS2347.1.

Zacharias, S., and C. Koppe, 2015: Einfluss des Klimawandels auf die Biotropie des Wetters und die Gesundheit bzw. die Leistungsfähigkeit der Bevölkerung in Deutschland. Tech. rep., Umweltbundesamt. UMWELT & GESUNDHEIT 06/2015.

Zampieri, M., F. D'Andrea, R. Vautard, P. Ciais, N. de Noblet-Ducoudré, and P. Yiou, 2009: Hot European summers and the role of soil moisture in the propagation of Mediterranean drought. *Journal of Climate*, **22 (18)**, 4747–4758, doi:10.1175/2009JCLI2568.1, https://doi.org/10.1175/2009JCLI2568.1.

Zhang, X., L. Alexander, G. C. Hegerl, P. Jones, A. K. Tank, T. C. Peterson, B. Trewin, and F. W. Zwiers, 2011: Indices for monitoring changes in extremes based on daily temperature and precipitation data. *Wiley Interdisciplinary Reviews: Climate Change*, **2 (6)**, 851–870, doi:10.1002/wcc.147, https://onlinelibrary.wiley.com/doi/pdf/10.1002/wcc.147.

Zhang, X., G. Hegerl, F. W. Zwiers, and J. Kenyon, 2005: Avoiding inhomogeneity in percentile-based indices of temperature extremes. *Journal of Climate*, **18 (11)**, 1641–1651, doi:10.1175/JCLI3366.1, https://doi.org/10.1175/JCLI3366.1.

Zink, M., R. Kumar, M. Cuntz, and L. Samaniego, 2017: A high-resolution dataset of water fluxes and states for Germany accounting for parametric uncertainty. *Hydrology and Earth System Sciences*, **21 (3)**, 1769–1790, doi:10.5194/hess-21-1769-2017.

Zwiers, F. W., and Coauthors, 2013: *Climate Extremes: Challenges in Estimating and Understanding Recent Changes in the Frequency and Intensity of Extreme Climate and Weather Events*, 339–389. Springer Netherlands, Dordrecht, doi:10.1007/978-94-007-6692-1_13.

Acknowledgments

A mighty flame followeth a tiny spark. - Dante Alighieri