Julia Lutz

IMPROVEMENTS AND APPLICATION
OF THE
STATISTICAL ANALOGUE RESAMPLING SCHEME STARS

Julia Lutz

IMPROVEMENTS AND APPLICATION
OF THE
STATISTICAL ANALOGUE RESAMPLING SCHEME STARS

Improvements and application of the STatistical Analogue Resampling Scheme STARS

DISSERTATION

ZUR

ERLANGUNG DES DOKTORGRADES (DR. RER. NAT.)

DER

MATHEMATISCH-NATURWISSENSCHAFTLICHEN FAKULTÄT

DER

RHEINISCHEN FRIEDRICH-WILHELMS-UNIVERSITÄT BONN

vorgelegt von
Dipl.-Met. Julia Lutz
aus
Tomsk in Russland

Bonn, 2016

Diese Arbeit ist die ungekürzte Fassung einer der Mathematisch-Naturwissenschaft-lichen Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn im Jahr 2016 vor-gelegten Dissertation von Julia Lutz aus Tomsk in Russland.

This paper is the unabridged version of a dissertation thesis submitted by Julia Lutz born in Tomsk in Russia to the Faculty of Mathematical and Natural Sciences of the Rheinis-che Friedrich-Wilhelms-Universität Bonn in 2016.

Anschrift des Verfassers: Address of the author:

Julia Lutz
Meteorologisches Institut der
Universität Bonn
Auf dem Hügel 20
D-53121 Bonn

jlutz@posteo.de

Selbstständigkeitserklärung

Hiermit erkläre ich, dass ich die vorliegende Dissertation selbstständig und ohne Benutzung anderer als der angegebenen Hilfsmittel verfasst habe. Die aus fremden Quellen direkt oder indirekt übernommenen Zitate sind als solche kenntlich gemacht.
Die Dissertation wurde bisher in gleicher oder ähnlicher Form keiner anderen Prüfungsbehörde vorgelegt oder veröffentlicht.

Julia Lutz

Für meine Liebsten

Zusammenfassung

In der Klimamodellierung kann man im Wesentlichen auf zwei Arten die Ergebnisse globaler Klimamodelle auf eine regionale Basis herunter brechen: entweder man benutzt dynamische Klimamodelle, die wie die globalen Modelle auf physikalische Gesetze basieren oder man verwendet statistische Methoden, die auf statistische Beziehungen zwischen den verschiedenen meteorologischen Variablen basieren. Beide Methoden bieten Vor- und Nachteile. Ein wesentlicher Vorteil der dynamischen Modellierung ist die Möglichkeit physikalische Prozesse zu analysieren und zu verstehen. Allerdings sind regionale Modelle mit einer hohen räumlichen Auflösung sehr langsam, was die Berechnungen aufwändig und teuer macht. Statistische Ansätze verbrauchen dagegen kaum Ressourcen und so sind viele Ensembleläufe möglich, die ein genaueres Abschätzen der Klimaveränderungen ermöglichen. Ihre Anwendung ist jedoch beschränkt auf die nahe Zukunft oder präziser ausgedrückt, auf die Simulation eines Klimas, das ähnlich ist zu dem aus den Eingabedaten.

In der vorliegenden Dissertation wird das statistische Modell STARS vorgestellt, das auf der Annahme basiert, dass bereits in der Vergangenheit beobachtetes Wetter so oder so ähnlich auch in der Zukunft auftreten wird. Die eigentliche Simulation besteht aus einer Umsortierung der meteorologischen Beobachtungen indem im ersten Schritt komplette Kalenderjahre und im zweiten Schritt 12-Tages-Blöcke zufällig gezogen und neu angeordnet werden, sodass eine neue Zeitreihe entsteht. Um die Klimaveränderung zu berücksichtigen, ist es möglich einen linearen Trend für eine ausgewählte meteorologische Größe (z.B. Temperatur) im Simulationszeitraum vorzugeben. Das Modell nimmt die Neusortierung der Beobachtungen dann anhand dieses Trends vor.

Diese Arbeit beschäftigt sich mit den folgenden Fragen und gibt darauf Antworten.

1. Ist es möglich STARS mit Gitterpunktsdaten anzutreiben (statt wie bisher nur mit Stationsdaten)?

   Dazu wird STARS mit einem Reanalysedatensatz für die Südspitze Afrikas angetrieben. Die guten Ergebnisse bestätigen, dass STARS auch mit Gitterpunktsdaten als Eingabe benutzt werden kann. Außerdem wird gezeigt, dass STARS auch ein großes Simulationsgebiet erfolgreich simulieren kann.

2. Kann das Simulationsgebiet ausgeweitet werden? Wenn ja, was sind die Herausforderungen?

   Mit der früheren STARS Version (STARS_2.1) ist es nicht möglich Simulation auf kontinentaler Ebene zu berechnen. Die Ergebnisse weisen einen unrealistischen Jahresgang der Temperatur und besonders der kurzwelligen Strahlung auf. Um dieses Problem zu lösen wurde eine neue Version (STARS_2.4) entwickelt. Diese Version liefert gute Ergebnisse für ganz Europa und ist daher gut geeignet um Simulationen auf kontinentaler Ebene mit einer hohen klimatischen Variabilität durchzuführen.

3. Wie ist die Qualität der STARS Projektionen? Berücksichtigen sie die Klimaerwärmung?

   Hierzu wird STARS mit den Daten eines dynamischen regionalen Klimamodells (CCLM) angetrieben, die Ergebnisse für die Zukunft werden dann mit Projektionen desselben Modells verglichen. Der Vergleich zeigt, dass STARS in der Lage ist die Temperatur und den Niederschlag für die Jahre 2030-2070 gut wiederzugeben. Im Falle der relativen Feuchte und der kurzwelligen Strahlung versagt STARS hingegen für diesen Zeitraum. Für die Jahre 2000-2040 gelingt es STARS alle diese Variablen zu reproduzieren. Das macht STARS zu einer echten Alternative zu dynamischen regionalen Klimamodellen für eine Zeitspanne von 40 oder 50 Jahren. Auf diese Weise könnte man die benötigten Ressourcen zur Bereitstellung von Klimadaten für diverse Klimafolgenanwendungen reduzieren ohne auf eine gute Qualität der Simulationsergebnisse für die nahe Zukunft zu verzichten.

Contents

Glossary

CCLM COSMO model in CLimate Mode.

CMIP5 Coupled Model Intercomparison Project Phase 5.

COSMO COnsortium for Small-scale MOdelling.

ECHAM5 is the GCM of the Max-Planck-Institute for Meteorology in Hamburg. ECHAM stands for ECMWF and Hamburg.

ECMWF European Centre for Medium-range Weather Forecasts.

ERA-40 reanalysis data is a global reanalysis data set of the ECMWF.

GCM Global Circulation Model.

IPCC Intergovernmental Panel on Climate Change.

LM Lokalmodell, predecessor of the COSMO model.

MPIOM Max-Planck-Institute Ocean Model.

NWP Nummerical Weather Prediction.

RCM Regional Climate Model.

RCP Representative Concentration Pathway.

REMO is the REgional MOdel of the Max-Planck-Institute for Meteorology in Hamburg.

SRES Special Report on Emission Scenarios.

STARS STatistical Analogue Resampling Scheme.

WATCH data set is a global data set of the Water and Global Change Programme.

1. Introduction

In the last decade, many studies have been published which conclude that there is a global warming (see IPCC, 2007 and IPCC, 2014 for a summary). These studies are mainly based on results from Global Circulation Models (GCMs). As the resolution of global climate simulations is too coarse for regional climate impact studies, two different methods are used to scale the GCM results to a higher resolution: dynamical and statistical downscaling.

Dynamical downscaling is based on physical laws and numerical approximations, it is carried out by Regional Climate Models (RCMs). Like GCMs, RCMs compute climate simulations by solving equations, e.g. conservation of momentum, thermal energy equation and conservation of mass. Thereby, the GCM output serves as forcing data and determines the boundary conditions.

Statistical downscaling is based on statistical relations between different meteorological variables and covers different methods. Examples for these methods are stochastic weather generators (Wilks, 2010; Wilks, 2012), regression models (Stoner et al., 2013) and weather analogue methods (Wilby et al., 1998; Zorita and Storch, 1999). Stochastic weather generators generate time series of single meteorological variables with respect to prescribed general time series statistics. Since this is done individually for each variable, it is likely that the results yield physical inconsistent combinations of variables. However, weather generators are independent of GCM output and associated biases. Regression models utilise statistically derived relations between regional weather and GCM output. Analogue methods look for the past circulation pattern that is most similar to a certain pattern in the future, simulated by a GCM, and assign the corresponding weather of the pattern in the past to that in the future. Both methods result in physically consistent combinations of meteorological variables in the future. However, since both methods downscale GCM output directly, they are dependent on the performance of these models.

Common to all statistical downscaling methods is the limitation to simulations of the near future or more general to simulations of a climate that is similar to the one in the input data. In this way, the stationary assumption which is implied in statistical downscaling is not overstrained. In contrast, dynamical downscaling is used to generate long time series (from decades to centuries). Different to dynamical downscaling, the aim of statistical methods is to provide climate projections but no insight into dynamic processes. This is done in a fast and simple way without consuming much computational costs.

Thus, it is possible to compute a large simulation ensemble, whereas RCMs generate only few simulation runs.

In this thesis the STatistical Analogue Resampling Scheme (STARS) is used. It is based on the general assumption that weather that has already been observed in the past is very likely to recur in the near future in the same or a very similar way. Hence, the method randomly chooses segments of past observations – in a first step entire calendar years and in a second step blocks of several consecutive days – and recomposes them to new time series. To take changes in climate into account, it is possible to prescribe a linear trend for a chosen variable (e.g. temperature) for the simulation time period. STARS then rearranges the observations according to the prescribed trend. For instance, in case of an increasing temperature, warmer years or blocks are favoured over cooler years or blocks. Several heuristic rules make sure that the resulting time series are realistic in terms of annual cycles, weather persistence, etc. The approach of STARS to use past observations and reassembling them according to a prescribed trend is similar to analogue methods described above. Its main advantage over other statistical methods is the fact that it can be used independently of GCM output and associated model biases while there is still a physical consistency between the different simulated variables.

To analyse the impacts of climate change on different regions of the world the role of climate impact studies increases. The basis of impact studies is meteorological data with the following requirements: coherent time series, physical consistency, small biases and high spatial resolution. These requirements apply not only to observations but especially to future projections of climate. Thus, there is a need for reliable climate simulations which are available in short time and which fulfil the above requirements. To estimate biases and possible future developments it is also important that the simulations cover a large ensemble size. STARS is able to fulfil these criteria. Its strength lies in the good estimation of the near future climate, and it regularly outperforms different dynamical RCMs on this time scale (see for example Orlowsky et al., 2008; Lutz et al., 2013). Thus, impact studies are an important and often used field of application for STARS (Hattermann et al., 2007; Hattermann et al., 2008; Huang et al., 2010; Liersch et al., 2012; Koch et al., 2012; Gädeke et al., 2013; Suckow et al., 2015). However, several questions arise after reading previous STARS studies. This thesis aims to give answers to these questions.

1. Is it possible to run STARS with gridded meteorological data?

   STARS has already been used successfully with observational station data (Orlowsky et al., 2008). As the use of stational data limits the application spectrum of STARS, it is desirable to run the model with gridded data, e.g. reanalysis data. In this thesis the application of

gridded data as input for STARS is tested for the first time. Therefor the WATCH data set, a global reanalysis data set, is used.

2. Can the simulation area be extended? If yes, what are the challenges?

   Until now, STARS was applied to small simulation regions and single river catchments, e.g. the Elbe river catchment (Orlowsky et al., 2008) and the Yangtze river catchment (Orlowsky et al., 2010). In this thesis the simulation area is extended up to continental scale. Thus, it is possible to meet new requirements on the model and make the model output available on spatial scales that are often demanded by data users. The extension of the simulation area implies a demanding setting for STARS because typically the climatological variability increases as well. This work elucidates the corresponding challenges and shows how the model deals with those.

   In a first experiment STARS is used to compute simulations for the southern tip of Africa. After this the model is applied to the whole of Europe. The simulation areas have been defined by research projects. However, they provide the opportunity to test the performance of STARS in climatological interesting regions where the model has not been used before.

3. How is the quality of STARS climate projections? Are they able to account for climate change?

   In previous studies, STARS simulations were always validated against past and/or present-day data. However, as the climate projections are mainly used for further climate impact studies it is interesting to verify their quality. To evaluate the climate projections made by STARS, a new approach is used in this thesis: The projections are compared to output of a dynamical RCM. In this way it is possible to make a qualitative statement about the STARS simulations of the future.

The thesis is structured as follows. Chapter 2 describes the model procedure focusing on the second step of the model process. It also gives an overview of the data used in this work. Chapter 3 deals with the usage of reanalysis data as input for STARS in southern Africa, answering the first question formulated above. The corresponding simulation results are compared to a dynamical RCM. Chapter 4 responds to the second question by extending the simulation area to continental scale (Europe). Running the STARS model on such a scale for the first time reveals some short-comings in the model. To eliminate these, the model process is improved. The improved version of STARS is used to compute climate projections for Europe. These results are shown as well.

Chapter 5 deals with the last question by using the data of a dynamical RCM to run STARS. The resulting climate projections are compared to the RCM data for the 21st century to evaluate the performance of STARS in the future. Finally, chapter 6 draws some overall conclusions and gives an outlook for future work with the STARS model.

2. STARS Model and input data

The basic idea of a statistical method to develop climate scenarios was already described in Werner and Gerstengarbe (1997). This idea was further developed by Orlowsky et al. (2008) to the present day STARS (Statistical Analogue Resampling Scheme). Since then, several studies have dealt with the usage of the model (Orlowsky and Fraedrich, 2009; Orlowsky et al., 2010; Lutz et al., 2013; Zhu et al., 2013; Feldhoff et al., 2014; Lutz and Gerstengarbe, 2014) in different regions of the world, approving its performance and advancing its implementation.

This chapter describes the basic principle of the STARS model and introduces the WATCH data set that was used in the studies in the following chapters. The STARS model (version 2.1) is already described in detail by Orlowsky (2007) and Orlowsky et al. (2008). Therefore, this chapter gives only a summary of the simulation process, which basically comprises two steps (cf. Fig. 2.1).

Note that parts of this chapter are published in Lutz and Gerstengarbe (2014).

2.1. STARS

In brief, the two modelling steps of STARS are:

1. Random rearranging of entire calendar years and selecting the rearrangement that fits best to the prescribed trend in a given variable

2. Improving the rearrangement by replacing blocks of continuous days according to a set of rules

Within statistics, the term "bootstrapping" (Efron, 1979; Efron and Tibshirani, 1994) is usually used for the random replacement in the second step. Chapter 4 describes how the bootstrapping is changed to improve the model results. This yields a new model version (referred to as version 2.4 in this thesis).

A daily time series of meteorological variables (observations or reanalysis data) and a linear trend prediction of a chosen variable for the time period to be simulated form the basis of the STARS resampling scheme. During the

2. STARS Model and input data
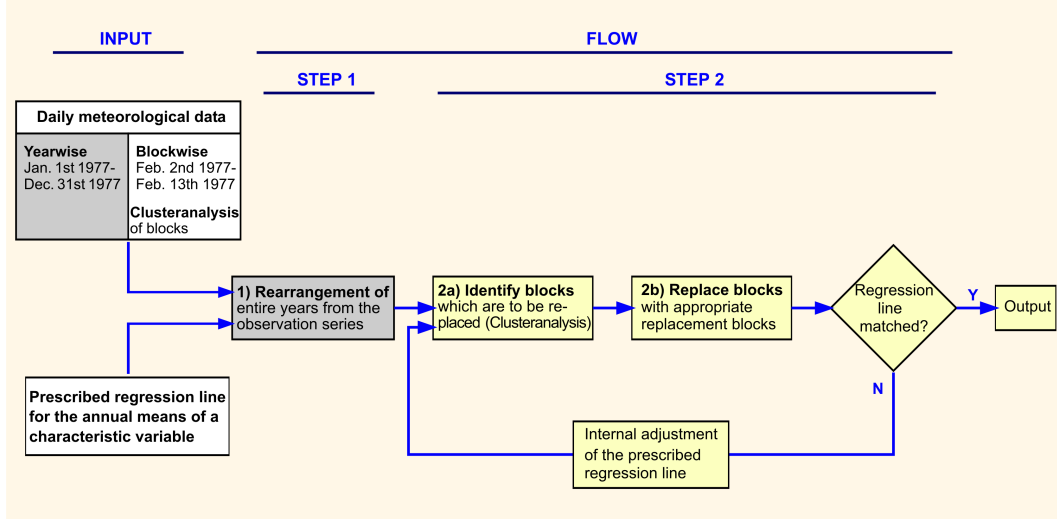
Figure 2.1. The simulation work flow of STARS. In the first step (shown in grey) the model rearranges entire calendar years randomly. The rearrangement that is closest to the prescribed trend is chosen for the second step. In this step (shown in yellow) blocks of consecutive days are replaced to improve the simple rearrangement from the first step.

simulation process, the time series is split into calendar years as well as into blocks of consecutive days. The length of these blocks reflects the weather persistence in the region of interest. Note that these blocks are running blocks starting each day to ensure a high number of blocks available. To identify blocks with similar weather conditions a combination of hierarchical and non-hierarchical cluster analyses for daily temperature is implemented in the model.

As mentioned above, at first STARS creates a large sample X (typically 100'000 elements) of possible time series x_i for the simulation time period by simply rearranging entire calendar years randomly. The years are drawn with replacement, the number of drawn years corresponds to the number of years N in the simulation period. As the sequences of days within the individual years are left unchanged, this procedure ensures realistic annual cycles and weather persistence for each time series x_i (which consists of several meteorological variables $v_1, ..., v_n$). The best rearrangement of the sample X (in STARS termination this is called the "first approximation") is chosen for a further step. In this context, "best" means that the linear regression lr of the annual means am of a chosen variable $y \in \{v_1, ..., v_n\}$ is closest to the given linear trend. Thus, the first approximation x_f is defined as the rearrangement with the minimum distance between the parameters of the corresponding regression line $\mathrm{lr}_a(y) := \mathrm{lr}(\mathrm{am}(y))$ and those of the given trend t:

$$x_f = x_k : \| \operatorname{lr}_a(y_k) - t \| \leq \| \operatorname{lr}_a(y_i) - t \| \quad \forall \, x_i = (v_{1i}, ..., y_i, ..., v_{ni}) \qquad (2.1)$$

It is possible, but not likely, that the regression parameters of the first approximation fit the given parameters within a certain tolerance already. In this case the simulation is finished. Else, the model continues with the second step. If the first approximation contains more days than the simulation period (due to leap years), the extra days are simply removed at the end of the rearrangement. If otherwise the first approximation is shorter than the simulation period, the lacking days are attached from the beginning of the last year of the first approximation.

The second step uses blocks of days instead of single days to preserve the weather persistence, at least within the blocks. The main goal of this step is to find and replace the blocks that contribute to the mismatch between the first approximation and the prescribed trend. STARS starts with generating an artificial time series y_f^* for the chosen variable y by simply in- or decreasing the means of the single years in the first approximation y_f until the resulting annual means match the prescribed regression line exactly:

$$y_f^* = y_f + a : \operatorname{lr}_a(y_f + a) = t \qquad (2.2)$$

Then, the model compares the cluster affiliation of the blocks from the first approximation with those from the artificial time series. The blocks from the first approximation are left unchanged if they belong to the same cluster class as those from the artificial time series. If their cluster affiliation differs, the blocks are replaced randomly with blocks that fulfil the following criteria:

- The blocks have to belong to the same class as the blocks from the artificial time series. This improves the regression line of the simulated time series.

- The blocks can be used only once. This avoids a too frequent reuse of blocks.

- The blocks must lie within a certain date window around the position of the block to be replaced. This ensures the seasonal matching of the blocks.

- The blocks have to connect well with the adjacent blocks.

The last criterion identifies suitable blocks by comparing the second half of the predecessor and the first half of the considered block with a block from the

observations. The observational block starts on the same day as the second half of the predecessor. They both have to belong to the same cluster class. The same applies to a successor, if it is already defined.

Potentially, the replacement results in a time series y_r that matches the prescribed trend within a given tolerance. However, if this is not the case the second step of the simulation process is repeated iteratively with slightly exaggerated regression parameters for the artificial time series. More precisely, the difference between given parameters and the parameters that were achieved in the previous iteration is added to the prescribed parameters such that the regression line of the artificial time series lies above (or in case of an overestimation below) the prescribed trend:

$$y_r^* = y_r + b : \mathrm{lr}_a(y_r + b) = t_1 = t + (t - \mathrm{lr}_a(y_r)) \qquad (2.3)$$

This leads to a more frequent replacement of blocks and ensures that the biases of the first approximation are compensated efficiently (Orlowsky et al., 2008).

At several stages of the model process STARS has to select segments out of the input data set, either years in the first step or blocks of consecutive days in the second step. As this is done randomly, any STARS simulation is an arbitrary sample from the universe of all possible simulations. The range of all possible simulations can be estimated using a large number of simulations (for example 100 as in chapters 4 and 5 or 1000 as in chapter 3).

The above description applies for a single weather station or data point, however the approach is similar for spatial simulations. In this case, the simulations take place in a parameter space of higher dimensionality. This leads to spatial consistent fields for every variable of the input data, as the spatial distributions consist of already observed data.

To limit the number of stations or data points in the simulation area that are used for the actual rearrangement, a hierarchical cluster analysis is carried out to identify climatological similar stations or grid points. This limitation avoids the use of redundant information (grid points with highly correlated data) and is often necessary due to practical specifications (e.g. memory limitations). The cluster analysis uses parameters like mean, variance and trend estimates of temperature and precipitation. The centre of mass of each cluster is chosen as the representative station of the respective cluster. The representative stations represent the spatial variability of the simulation area. The number of cluster classes C is chosen subjectively with regard to the climatology of the area. The specification of the prescribed trend is done individually for each representative station. Thus, different spatial developments are taking into account.

In the first step of the simulation process the model chooses the realisation

$X_k \in X_C$ for which the obtained regression lines for all representative stations $\mathrm{lr}_a(Y) = \mathrm{lr}_a((y_1, ..., y_C))$ are closest to the prescribed trends $T = (t_1, ..., t_C)$:

$$X_f = X_k : \| \mathrm{lr}_a(Y_k) - T \| \leq \| \mathrm{lr}_a(Y_i) - T \| \quad \forall \ X_i = (V_{1i}, ...Y_i, ..., V_{ni})$$
$$\wedge \ V_{ji} = (v_{j1}, ..., v_{jC}) \tag{2.4}$$

In the second step the blocks are characterised by the characteristic variables of all the representative stations (instead of just one as in the case of a single station or grid point). Hence, the artificial time series is generated for each representative station and the blocks of all these series merged together are compared with the combined first approximation from the first step. If several iterations of the second step are necessary, the exaggerated trends are calculated as for the single station case: for each representative station individually.

The source code of the STARS model is written in C++ and consists of 34 files with 12 000 lines of code. Generally, the STARS source code is available on request for research.

2.2. The WATCH data set

STARS rearranges observational station data or reanalysis data to generate ensembles of simulations. In this thesis the WATCH data set* (Weedon et al., 2011) is used mainly. This global data set is chosen because it combines the advantages of meteorological observations – covering a long time period and a good spatial resolution – and ERA-40 reanalysis data – availability, physical consistency of different variables and a good temporal resolution. It is available on a latitude-longitude grid with a resolution of $0.5° \times 0.5°$. It contains the following variables with a temporal resolution of three hours: wind speed at $10\,\mathrm{m}$, air temperature at $2\,\mathrm{m}$, surface pressure, specific humidity at $2\,\mathrm{m}$, downward longwave radiation flux, downward shortwave radiation flux, rainfall rate and snowfall rate. However, STARS works with daily data. Thus, daily means of air temperature, wind speed, surface pressure and shortwave radiation, as well as daily minimum and maximum air temperature are used. The specific humidity is calculated to relative humidity and the rainfall and snowfall rate are summed up to daily precipitation.

As shown by Weedon et al. (2011) the WATCH data set provides a good representation from (subdaily) meteorological events to climate trends for the period 1958–2001. As there is no ERA-40 reanalysis data (Uppala et al., 2005)

---

*see `http://www.eu-watch.org/data_availability` for further information

available prior to 1958, the WATCH data set does not reproduce particular events during the period 1901–1957, but subdaily to seasonal statistics are still well represented (Weedon et al., 2011). Furthermore, an observed temperature trend is imposed over the whole time period of the WATCH data set. Thus, the lack of ERA-40 reanalysis data does not hinder the application of the STARS model prior to 1958 using the WATCH data set.

3. STARS forced with gridded input data: an example for the Orange River basin in southern Africa

This chapter answers the question whether gridded data is applicable as input for STARS. Additionally, it analyses the performance of STARS in an extended simulation area. For this purpose, STARS is forced with the WATCH data set to compute climate projections for the Orange River basin in southern Africa. In addition, the results are compared to the climate projections of a dynamical regional climate model (CCLM). In order to determine the viability of both models in the region of interest, two cross-validations are carried out for the years 1976–2000.

In contrast to chapters 4 and 5, where a new STARS version has been developed and used, in this chapter the version STARS_2.1 is used. As shown below, this model version is able to produce reasonable results in southern Africa. This is due to the fact, that the temperature trend in the input data shows a small temporal variability in this simulation area. Additionally, the temperature trend for the projections is of the same magnitude.

This chapter is structured as follows. The data used and the experiment set-up are described in sections 3.3 and 3.4. Section 3.5 analyses the cross-validation experiment which demonstrates the applicability of STARS and CCLM to southern Africa. The results of the climate projections for the years 2011–2060 are illustrated in section 3.6. Finally, section 3.7 draws some conclusions.

Note that parts of this chapter have been published in Lutz et al. (2013).

3.1. Introduction

The Orange River basin is located in southern Africa (Fig. 3.1). It is the longest river in the region with a length of 2200 km. Its basin covers an area of almost $1\,000\,000\,\text{km}^2$ and affects the livelihood of the four nations South Africa, Namibia, Botswana and Lesotho. Its spring is located in the Drakensberg Mountains in Lesotho, from where it flows westward to the Atlantic Ocean, passing through vastly different areas such as mountains, dry grasslands and finally arid landscapes. There is a strong rainfall gradient from east to west: the mountain regions in Lesotho have an annual precipitation of up to 2000 mm,
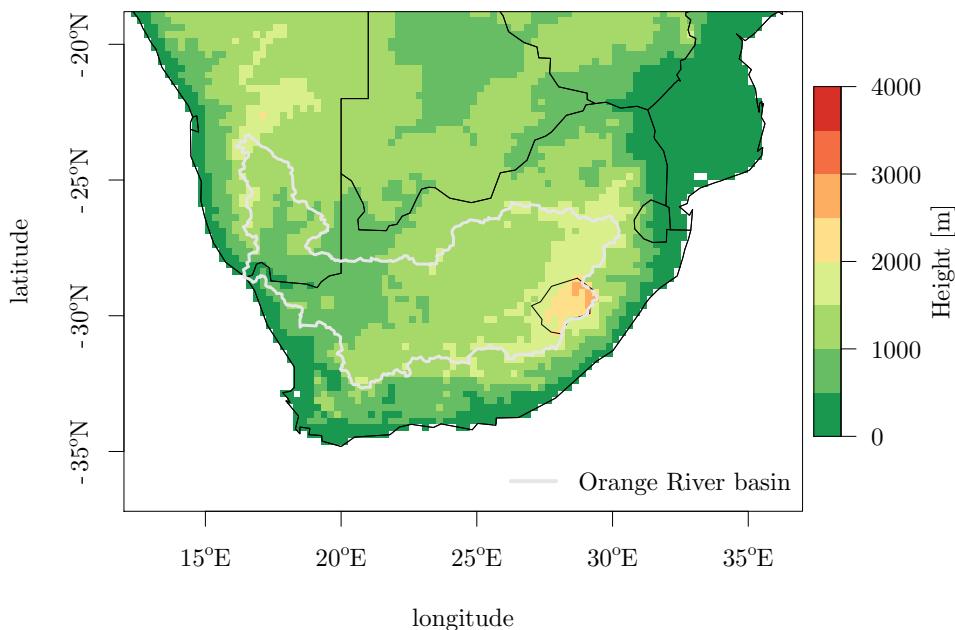
Figure 3.1. The location of the Orange River basin in southern Africa.

whereas the region around the mouth of the river hardly gets 50 mm per annum. The temperature gradient in the interior of southern Africa is strong as well. The daily mean temperatures range from approximately 10 °C in the Lesotho Highlands to more than 22 °C at the river mouth. This contrast is even more pronounced when looking at the extreme temperatures. The desert regions at the lower Orange River can reach very high temperatures of up to 50 °C while frost days are common in the mountain region.

The natural water discharge is highly variable and dependent on climate. The Orange River plays an important role in the region's ecology and economy. The large Gariep and Vanderkloof dams provide hydro-electric power and water for irrigation, mining and industry. Water is even transferred through various river/tunnel systems beyond the basin boundaries. The water of the Orange River is in such demand that these transfer schemes are still being expanded (e.g. the Lesotho Highlands Water Project). Thus, it is important to investigate the possible hydrological developments in this region, which in turn depend on the future climate. This work is focused only on the climate projections though.

The high climatological variability in southern Africa is a challenge for every RCM because most models include physical parametrisations that are mainly tested and optimised for one specific climate regime; this is particularly true

for dynamical models. Although current RCMs are able to reproduce the main climatic features in different regions of the world, they tend to perform much better in their "home region" (Rockel and Geyer, 2008). STARS was developed and elaborately tested in Germany (Orlowsky et al., 2008). However, as STARS reproduces only spatial patterns which have been observed before, the simulated spatial variability is expected to be consistent with the observations. This feature has already been successfully demonstrated in Orlowsky et al. (2010) for station data in China. In this chapter, STARS is applied for the first time with gridded data to provide spatial fields instead of single station time series. Hence, it is possible to analyse the performance of STARS in this climatically interesting and varying region. Furthermore, the simulation results can be compared to those obtained by CCLM.

3.2. CCLM

Like most dynamical RCMs, CCLM originates from a Nummerical Weather Prediction (NWP) model. It is an offspring of Lokalmodell, predecessor of the COSMO model (LM), the predecessor of Deutscher Wetterdienst's current and operational NWP model COnsortium for Small-scale MOdelling (COSMO). In the early 2000s, features necessary for climate simulations were implemented on top of LM/COSMO, and in 2005 the model was named COSMO model in CLimate Mode (CCLM). In 2007/2008 both development branches, COSMO and CCLM, were merged back together, and CCLM became the official dynamical RCM of the German climate modelling community. Currently, CCLM is used by a growing community of more than 40 institutions in Asia, Europe and America.

CCLM was originally developed for applications in mid-latitude areas, such as Europe, but has been increasingly used in other regions of the world. For instance it was run for Asia, South America, Africa etc. In fact, results from a previous run over Africa are used here (see section 3.4.2). Generally, CCLM covers an area at the mesoscale, i.e. areas between $100\,\mathrm{km}^2$ and $10\,000\,000\,\mathrm{km}^2$. The resolution is typically in the range from a few kilometres up to, approximately, $50\,\mathrm{km}$. The time frame to be simulated varies from several years up to centuries; usually simulations are run for several decades.

CCLM, like other dynamical RCMs, requires data from a GCM to initialise the region to be simulated, and to drive it at the boundaries. The set-up for Africa is depicted in Fig. 3.2, where the typical size of simulation areas can be seen as well. As stated, values within the area in question, such as sea surface temperature, air temperatures, soil water content etc. are initialised with data from a GCM. Then, the equations from the physical description of the air, soil,
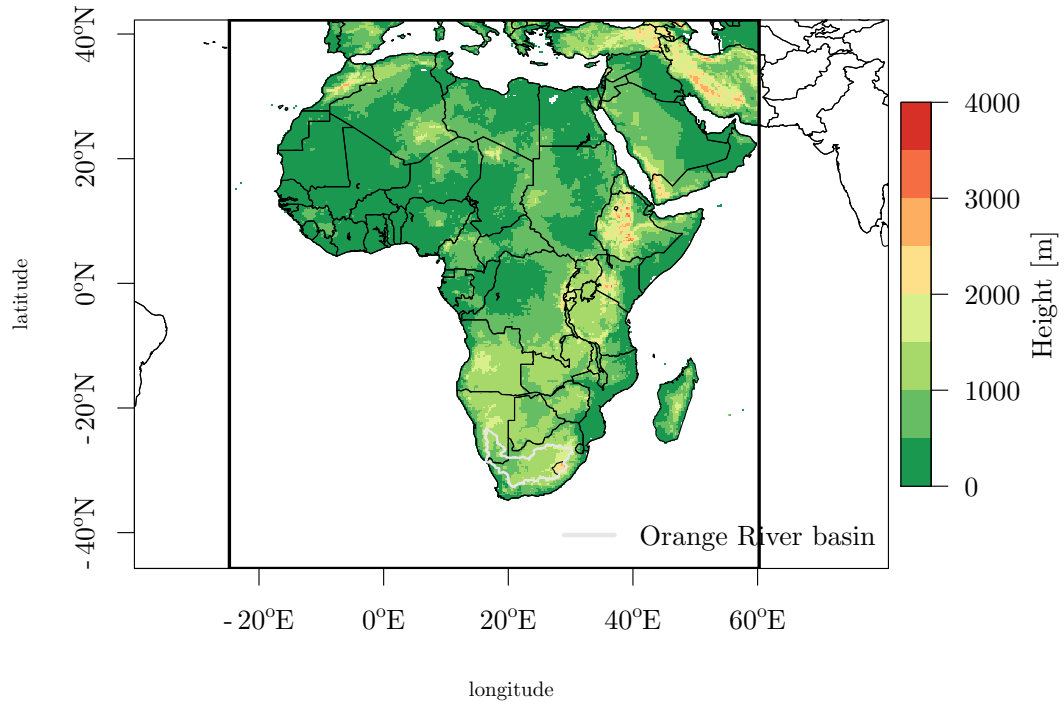
Figure 3.2. The simulation domain of CCLM in Africa, shown with a solid frame. All relevant fields are initialised with data from the driving GCM. The six hourly boundary conditions are obtained from the driving GCM as well.

and so forth are numerically solved. At certain intervals, in our case every six hours, boundary data from a GCM is fed into the model.

The fact that RCMs are based on physical principles/models allows researchers to qualitatively and quantitatively investigate the non-linear processes inherent to climate science. Furthermore, hypothetical scenarios, such as land use/land cover changes, can be examined. Dynamical models provide a lot of different variables at every point in space and time, unlike statistical models which are confined to the variables and positions of their input data. Potentially, a dynamical RCM can incorporate many submodels, such as urban models, lake models etc. However, a dynamical model can never escape its driving GCM, so any GCM bias is bound to show up in the RCM simulation as well. Due to the nesting procedure, there is no feedback to the driving GCM, so energy and momentum conservation is broken.
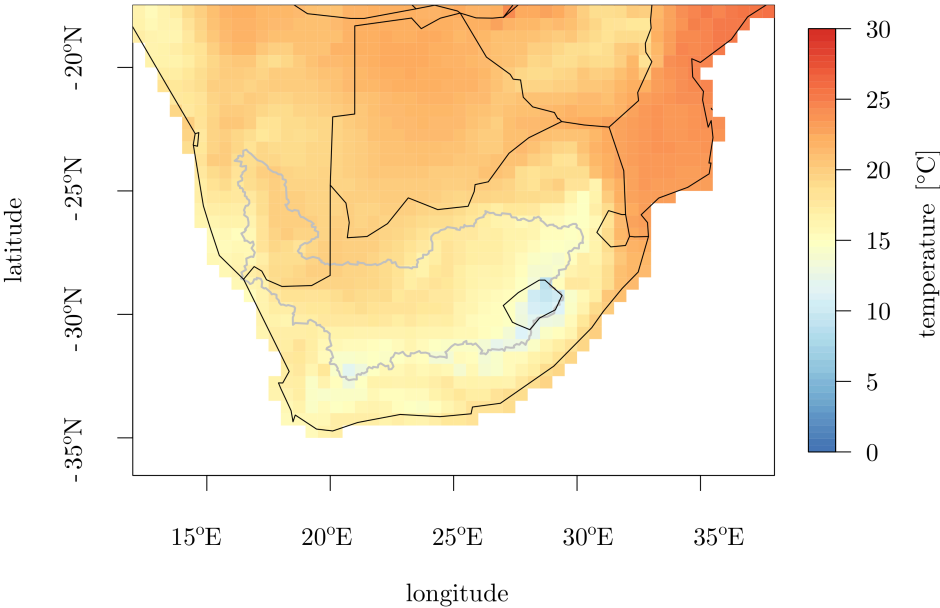
Figure 3.3. The observed annual average of the 2 m temperature from 1976–2000 (WATCH).
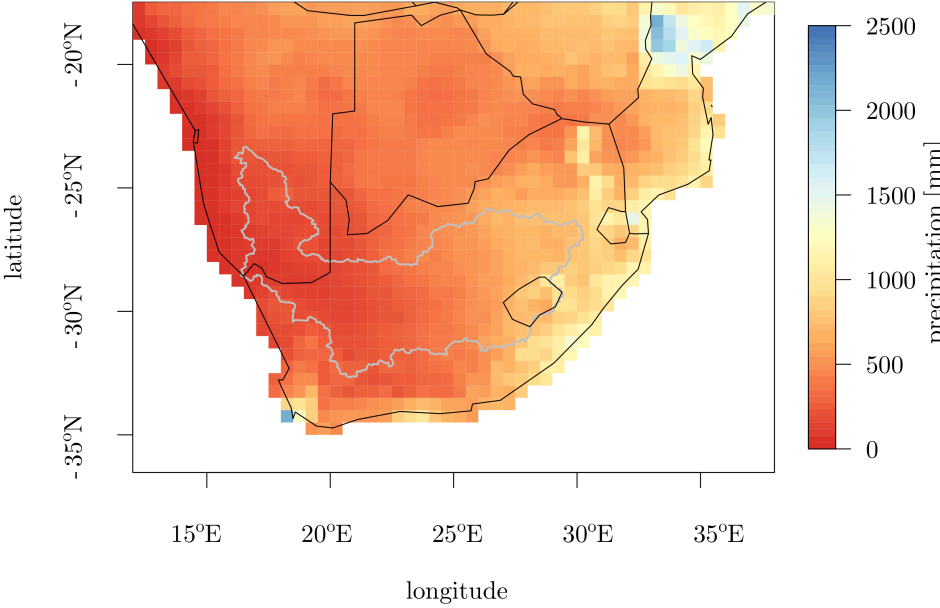
Figure 3.4. The observed annual average of total precipitation from 1976–2000 (WATCH).

3.3. Data

STARS rearranges observational station data to generate ensembles of future climate projections. In this study the WATCH data set (Weedon et al., 2011) from 1951 to 2000 is used. The daily time series of the following variables are used in the presented work: mean, minimum and maximum 2 m temperature, precipitation, relative humidity, shortwave radiation, wind and air pressure. The long term average of the 2 m temperature and of the precipitation are shown in Figs. 3.3 and 3.4.

As stated previously, climate projections from STARS are constrained by a linear temperature trend. In this work four different temperature trends are used. Thus, the projections have been carried out according to four different scenarios. Note that the temperature trend is not derived from a GCM in this study.

The simulations are carried out for the Orange River basin and its surroundings. They cover 904 grid points with a resolution of $0.5° \times 0.5°$, ranging from 21°S to 35°S and from 14°E to 36°E. The simulation area is split into five climatological sub-regions based on a hierarchical cluster analysis, as shown in Fig. 3.5.

3.4. Experiment set-up

3.4.1. STARS

To test the applicability of STARS to southern Africa, a cross-validation experiment is carried out. However, a positive cross-validation experiment is a necessary but not a sufficient condition for the calculation of climate projections. The input data is split into two periods of 25 years: the observation period 1951–1975, and the validation period 1976–2000. The climate of the validation period is simulated using the first 25 years and the observed trend of the annual mean temperature of the same period. This trend is derived from the data by a regression analysis of the annual mean temperature series at the five representative grid points. It ranges from 0.2 °C to 1.4 °C. The model reproduces the given trend with a user-defined tolerance. For the cross-validation this tolerance is chosen to be between 0.1 °C and 0.2 °C. The performance of STARS is evaluated comparing the simulated climatology with the observed one in the validation period. An ensemble of 1000 simulations is generated to quantify the model uncertainties.

Two of the representative grid points do not fulfil the "internal variability conservation" criterion (Orlowsky et al., 2010). This criterion says that only if the temperature anomalies of the input data and the simulated series can
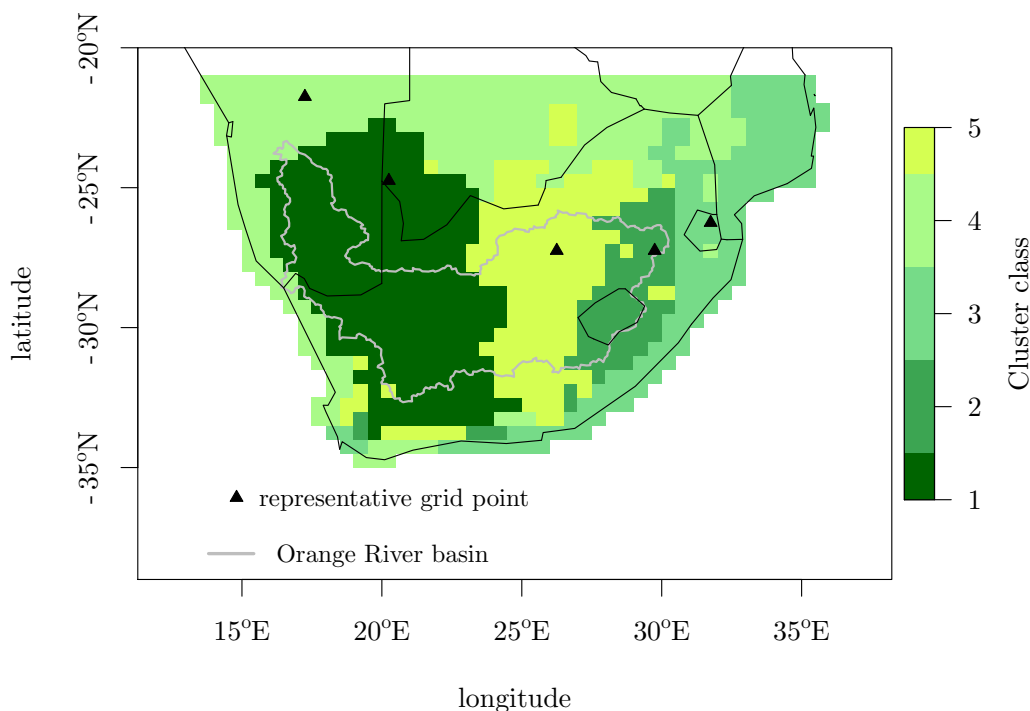
Figure 3.5. Result of the cluster analysis to determine regions with a similar climate (coloured areas represent the different clusters). The triangles mark the five representative grid points.

be seen as originating from the same distribution, the variability of the input data is large enough to generate series with a given temperature trend without a statistically visible reduction of variability. In various tests Orlowsky found that this criterion is only fulfilled when the warming in the simulated period continues with the same strength as in the training period. In case of the two grid points mentioned above, the annual mean temperature shows only a weak trend or even a temperature decrease during the first time period, while the annual means for the same grid points increase strongly during the valida-tion period. Therefore, the successful cross-validation despite this demanding setting gives strong evidence of the robustness of the projections.

After the cross-validation experiment future climate projections are com-puted for the time period 2011–2060. To avoid the dependency on GCMs and associated uncertainties, the projections are carried out using four different sce-narios: no temperature trend, a trend of 0.5 °C, a trend of 1 °C and a trend of 1.5 °C for the simulated time period. Note that the SRES A1B scenario (Naki-cenovic et al., 2000) results in a temperature trend of 1.6 °C for the simulation area and the simulation time period. This means that the projections done by CCLM are to be compared to the 1.5 °C scenario computed by STARS.

The simulations for each scenario contain 1000 ensemble members. From this large sample 100 simulations are drawn randomly without replacement to reduce the amount of data. Furthermore, only the ensemble median is chosen for a detailed analysis of the climate projections in this work.

3.4.2. CCLM

In this chapter CCLM runs are used that were originally carried out for the *Resilient Agro-landscapes to Climate Change in Tanzania* project (REACCT, 2012).

Since Africa is centred around the equator, there was no need for a rotated coordinate system. The simulation area stretches from 42.25°N to 45.75°S resp. from 24.75°W to 60.25°E. The resolution is 0.5°×0.5° which amounts to $56\,\text{km} \times 56\,\text{km}$.

For the validation the years 1976–2000 have been simulated. The simulations are driven by an ECHAM5 (Roeckner et al., 2006) run that was in turn driven by 20th century (C20) green house gas concentrations. The projection covers the years from 2001 to 2100. It is driven by ECHAM5 results as well, which are based on the SRES A1B (Nakicenovic et al., 2000) scenario.
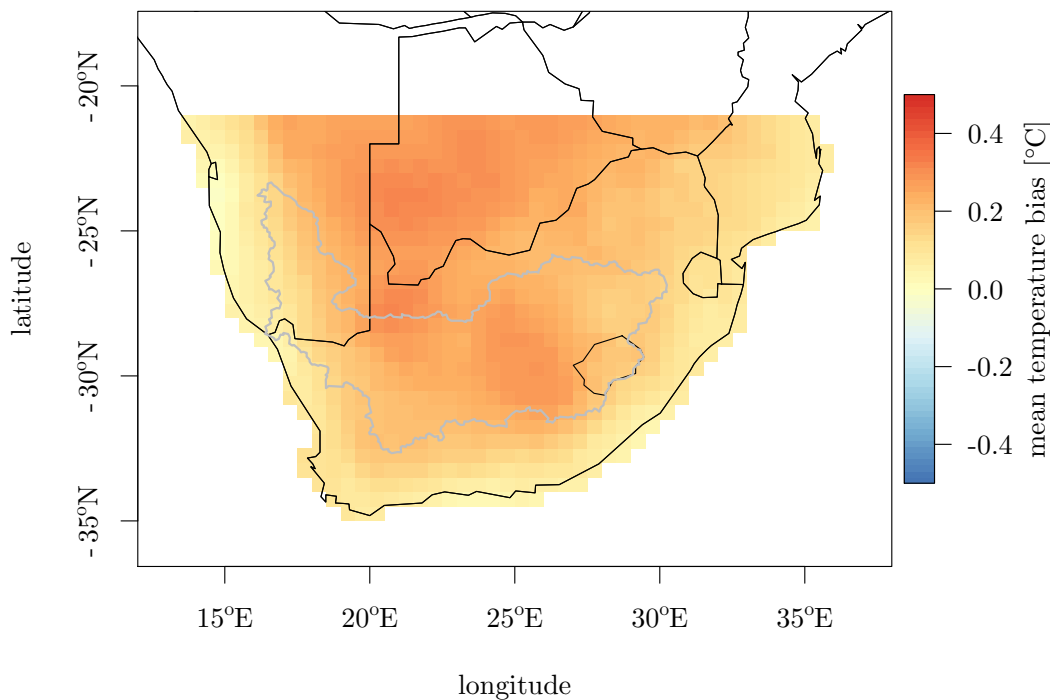
3.5. Validation

3.5.1. Validation STARS

The comparison between the observed and the simulated climate in the validation period (1976–2000) leads to the following results.

1. The agreement between the climatological means is very good for all statistics and variables analysed (not all are shown here).

2. The simulated spatial patterns of all variables are very similar to the observations.

3. The long time average of the mean, maximum and minimum temperature shows a slight overestimation of up to 0.4 °C. The strongest overestimation occurs in the interior of the simulation area (Fig. 3.6). However, the temperature bias was statistically tested using the Wilcoxon-Mann-Whitney test[*] (Hollander and Wolfe, 1999) with a significance level of 5 %: there is no significant difference between simulated and observed temperature for the whole simulation area.

---

[*]see the appendix for a short description

Figure 3.6. The absolute bias in the long term average of the 2 m temperature for the validation time period 1976–2000 as simulated by STARS (ensemble median).

In terms of precipitation the results are very good. Actually, precipitation is rather difficult to simulate, thus it is a challenge that many regional models cannot cope with (Kotlarski et al., 2005). The long term average of the precipitation (Fig. 3.7) indicates that STARS overestimates precipitation in general by approximately 10 % or even less. The overestimation reaches values of up to 50 % in the arid areas of Namibia. Despite the low bias, the difference between simulations and observations is significant (Wilcoxon-Mann-Whitney test* with a significance level of 5 %) for most parts of the simulation area, except Lesotho and Swaziland and surroundings. However, the mean annual cycle of the precipitation at the representative grid points shows that the model mostly underestimates precipitation in the rainy season (not shown). The precipitation in the rainy season is very well reproduced only for the representative grid point of the mountainous cluster, where Lesotho is located. Furthermore, the extreme rainfall (90th percentile of the precipitation and duration of dry days) at the representative grid points is reproduced well.

Figure 3.8 shows the spatial similarities for precipitation between the different realisations and the observations for the validation time period using

---

*see the appendix for a short description
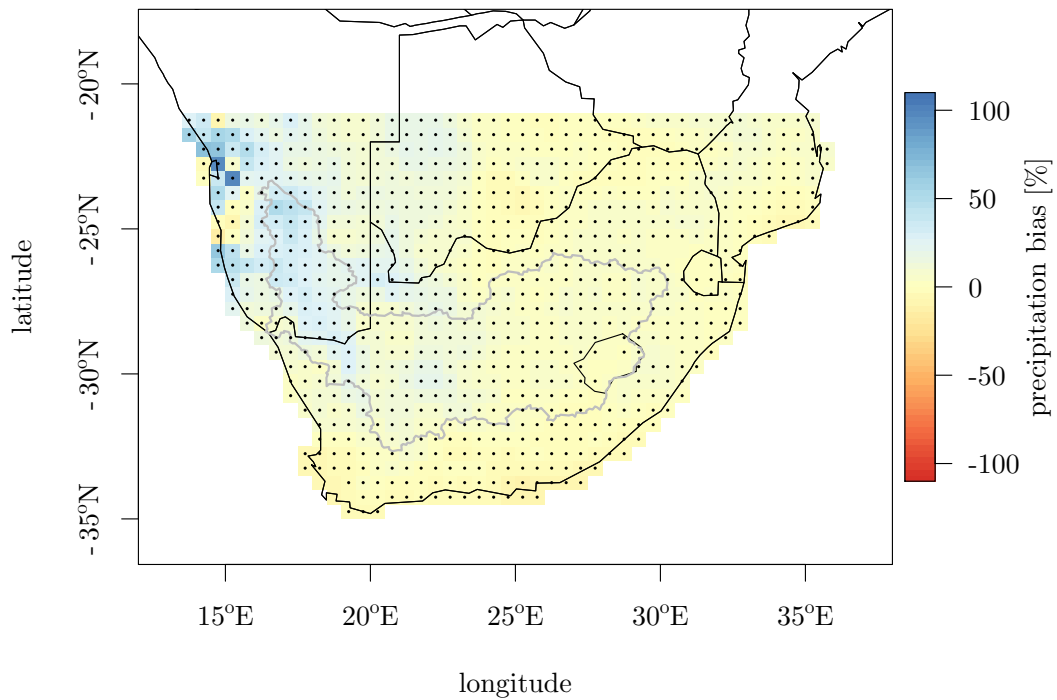
Figure 3.7. The relative bias in the long term average of the precipitation for the validation time period 1976–2000 as simulated by STARS (ensemble median). The points indicate grid points with a significant difference between simulations and observations ($\alpha = 0.05$). All values above 100 % are shown in the same colour.

a Taylor diagram (Taylor, 2001). This kind of illustration shows not only the spatial correlation between the simulations and observations, but also the similarity of spatial variability. It includes correlations, variability values (i.e. standard deviations) and the centred root-mean-square differences and is useful for comparing multiple data sets to a reference data set. As can be seen, the similarity between the different ensemble members (STARS) and the observations is very high and the standard deviations of the simulations are very close to the observations. CCLM, on the other hand, shows a comparatively lower correlation and standard deviation, and a larger root-mean-square difference.

Figures 3.9 and 3.10 illustrate the results for the mean annual cycles of mean, maximum and minimum temperature, precipitation, humidity, air pressure, radiation and wind. The annual cycle averaged over the validation period and the simulation area for each variable is illustrated in the upper parts of the plots. The observations are shown in red, while the spread of the simulations is shown in grey, including the ensemble mean in black. The inter-quartile range and the range between the 10th and the 90th percentile are also illustrated in grey. The two figures show that the spread of the ensemble is narrow and
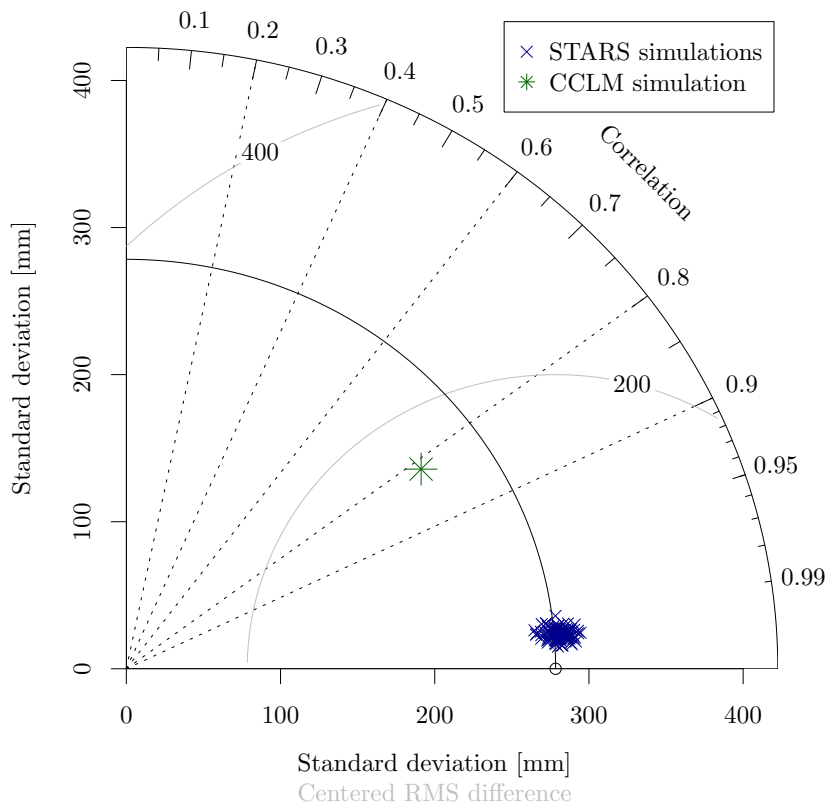
Figure 3.8. Similarities of the spatial precipitation pattern between the observations and the simulations for STARS and CCLM for the validation time period 1976–2000.

always contains the observed values.

3.5.2. Validation CCLM

The performance of CCLM is validated for the years 1976–2000. The variables investigated are the 2 m temperature and precipitation. As in the case of STARS, CCLM simulations are validated against the WATCH data set.

Figure 3.3 shows the observed annual average temperature in the region according to the WATCH data set. The CCLM bias is plotted in Fig. 3.11. As one can see, the performance of CCLM varies considerably throughout the region. Some areas, in particular deserts like the Kalahari or the Namib, are simulated with statistically significant excessive temperatures (Wilcoxon-Mann-Whitney
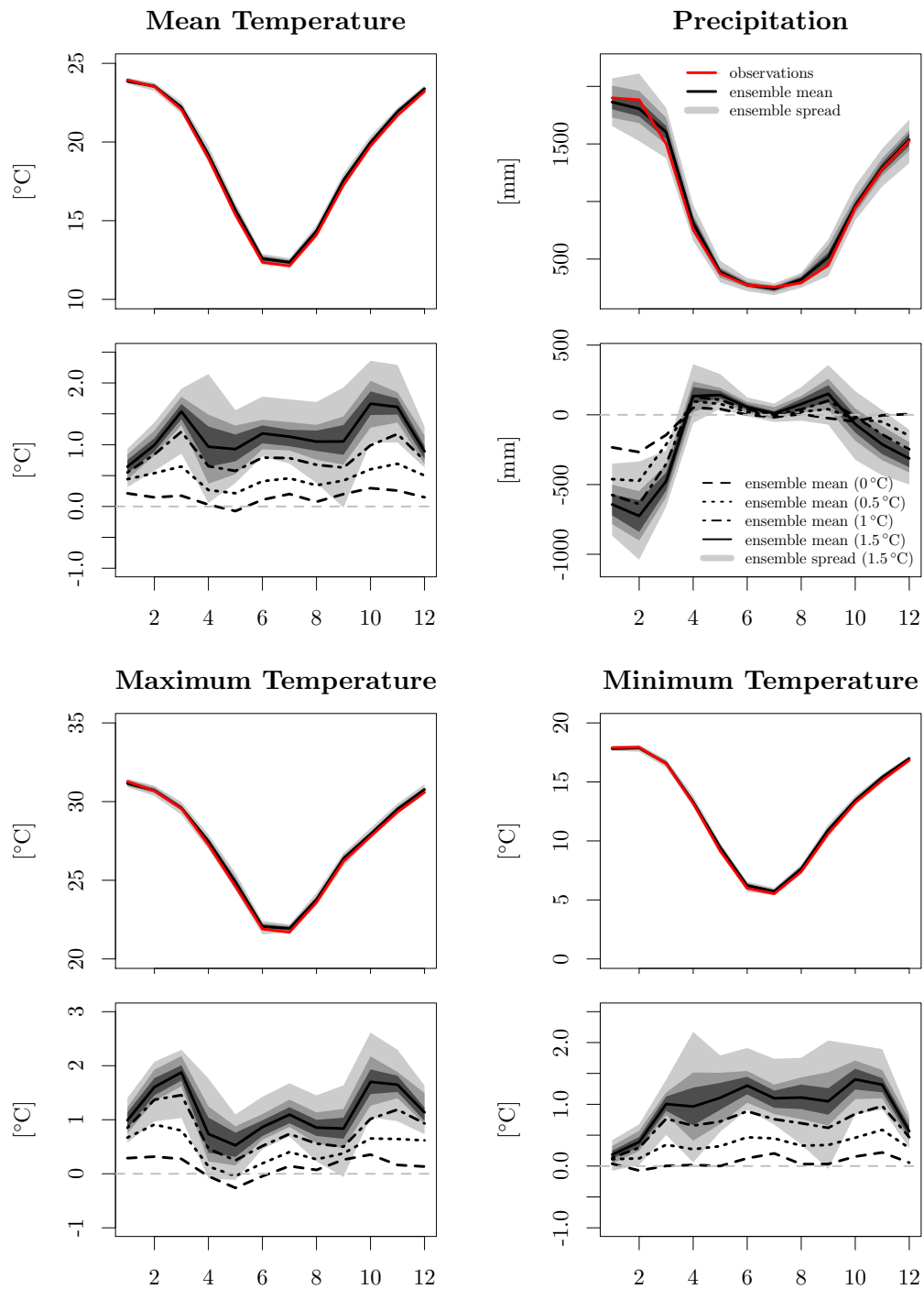
Figure 3.9. Mean annual cycle of temperature (mean, maximum and minimum) and precipitation. The upper parts show absolute values for the validation period 1976–2000 (black and grey: simulations by STARS, red: observations), the lower parts show the differences between the last 25 years of the projections (2036–2060) as simulated by STARS and the observations of the validation period. The legend applies to all plots.
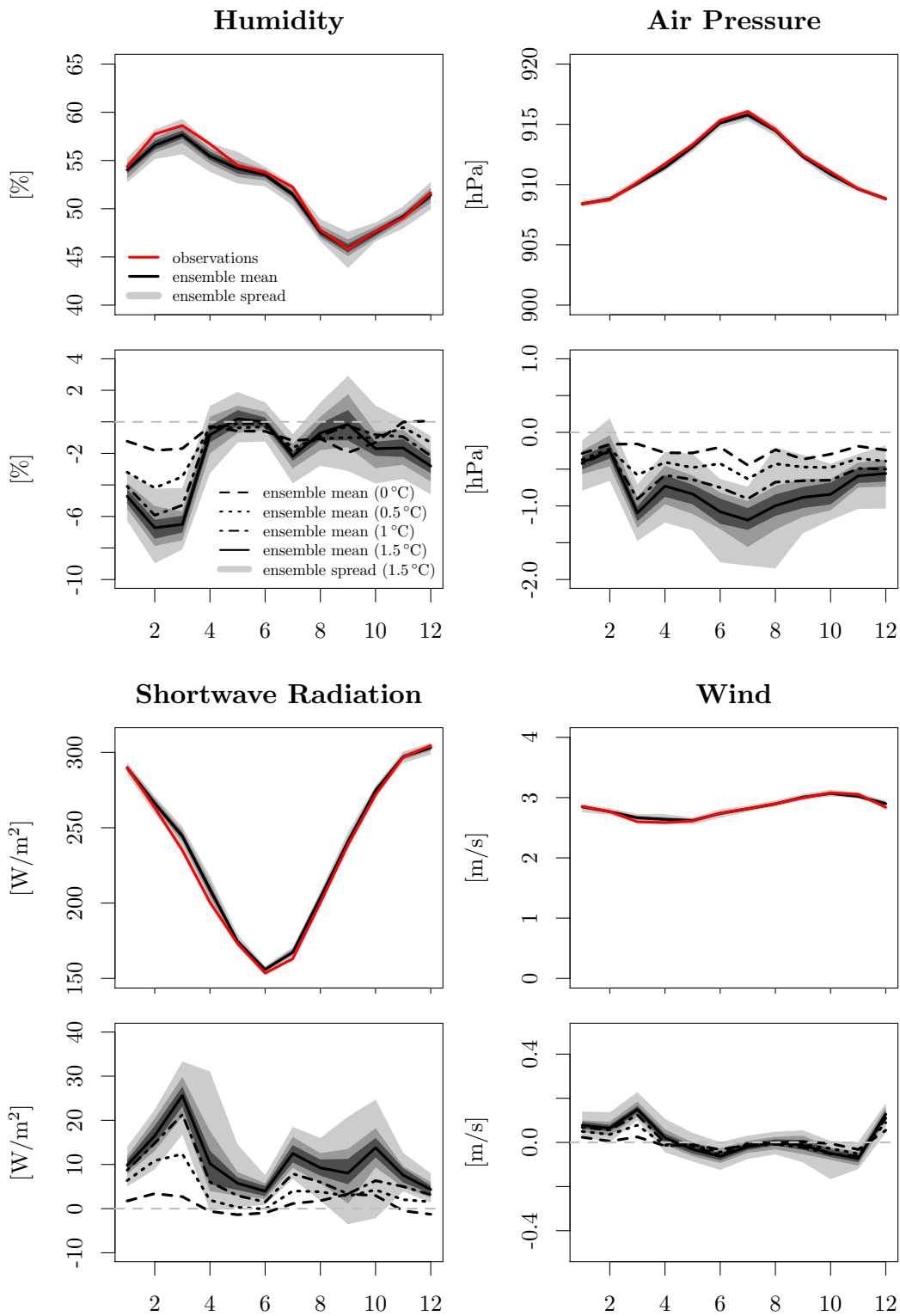
**Figure 3.10.** As for Fig. 3.9 but for relative humidity, air pressure, shortwave radiation and wind.
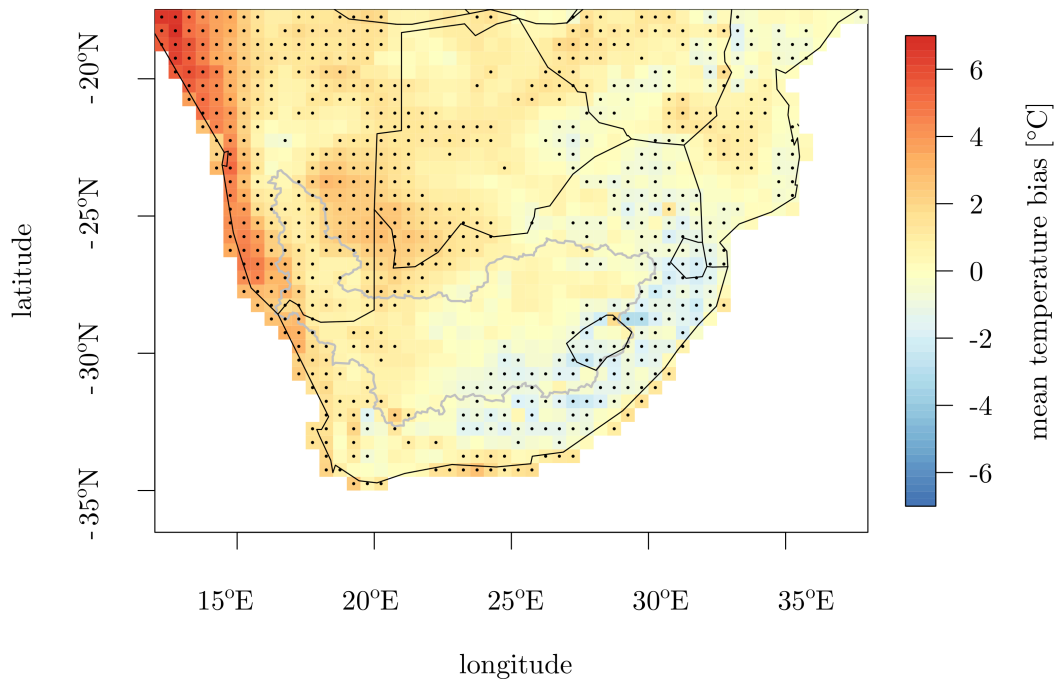
Figure 3.11. The bias in the annual average 2 m temperature from 1976–2000 as simulated by CCLM. The points indicate grid points with a significant difference between simulations and observations ($\alpha = 0.05$).

test* (Hollander and Wolfe, 1999), $\alpha = 0.05$). In the Namib Desert the annual 2 m temperature bias amounts to as much as 6 °C. Other areas, especially the mountainous areas around Lesotho, the Drakensberg Mountains and the eastern coast show an underestimation of temperature. For the remaining areas, such as the Orange River basin, the temperature is simulated reasonably well.

Similarly to the case of the 2 m temperature, observations (Fig. 3.4) and the corresponding bias (Fig. 3.12) for the total annual precipitation are shown. In Fig. 3.12 one can see that the CCLM performance in reproducing the southern African precipitation is rather poor. Throughout most of the region precipitation is too high, the overestimation for large parts of South Africa, Botswana and Namibia amounts to 100 %. On the Namibian coast the relative overestimation is even higher, but here precipitation values are very low in the first place. In southern Mozambique, on the other hand, CCLM shows a rather large negative bias, reaching occasionally values of -50 %. The large bias in Lesotho is particularly unfortunate, since most runoff for the Orange River is

---

*see the appendix for a short description
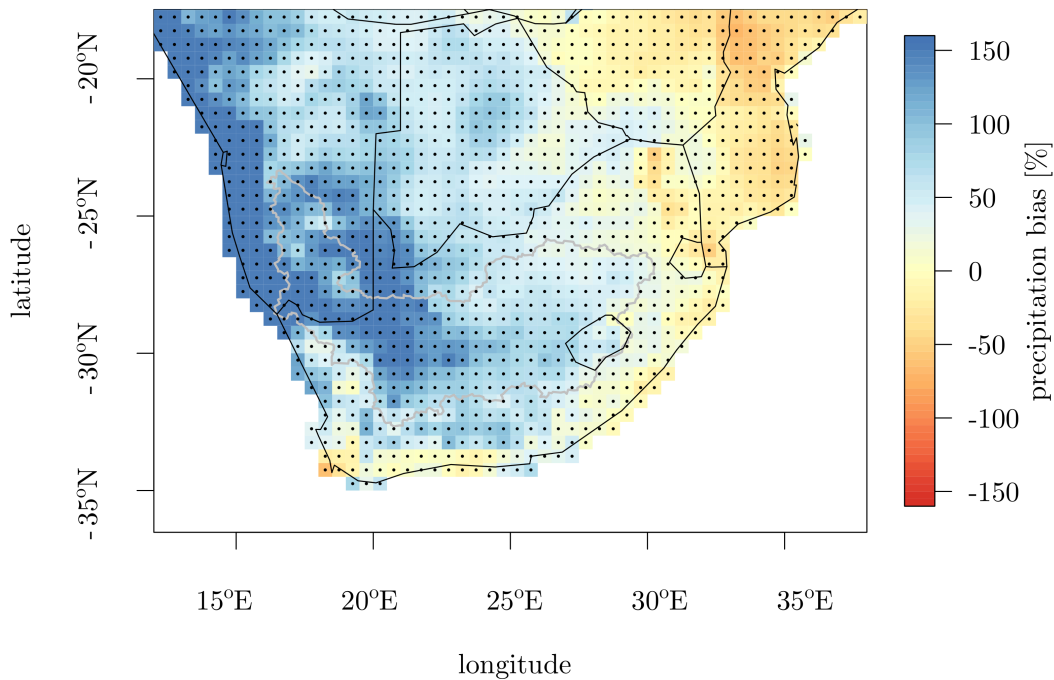
Figure 3.12. The relative bias in the annual average of total precipitation from 1976–2000 as simulated by CCLM. The points indicate grid points with a significant difference between simulations and observations ($\alpha = 0.05$). All values above 150 % are shown in the same colour.

generated in the Lesotho highlands. This renders CCLM results unsuitable for hydrological impact modelling in this area. These deviations exemplify the need to fine-tune CCLM better to the study area and to carry out a bias-correction with the simulation results.

3.6. Climate Projections

3.6.1. Climate projections with STARS

The 0 °C trend scenario of the climate projections is a control run. As expected, the comparison of the corresponding simulations and the observations provide very good results, underlining the good performance of STARS in southern Africa (not shown).

In Figs. 3.9 and 3.10 the differences between the last 25 years of the future projections (2036–2060) and the observations of the validation time period for the mean annual cycles are shown below the mean annual cycles for the

validation period (1976–2000). For the 1.5 °C trend scenario the ensemble spread is again shown in gray (the same applies to the inter-quartile range and the range between the 10th and the 90th percentile, the solid black line indicates the ensemble mean). Only the ensemble mean is shown for the other scenarios.

The differences between the future time period and the validation time period for the mean, maximum and minimum temperature show an increase for the future. But while the mean and maximum temperature rise especially in March and November, the increase of the minimum temperature has its maximum from March to November, the period of lowest temperatures in southern Africa. The temperature increase coincides with an increase in shortwave radiation that is most profound in March, but also in austral spring.

The results for precipitation indicate a strong decrease of over 25 % in austral summer, which is the rain period in southern Africa. However, there is a slight precipitation increase in autumn and spring. A similar pattern can be found for the humidity, but in this case there is no pronounced increase in May and September. The changes in wind and air pressure are very small or non-existent.

Figures 3.13 and 3.14 show the difference between the future projections and the observations for the long-term averages of temperature and precipitation. The change between projections and observations in the past is again statistically tested using a Wilcoxon-Mann-Whitney test* (Hollander and Wolfe, 1999) with a significance level of 5 %. To compare the future projections with the observations for the validation time period, only the last 25 years of the simulations (2036–2060) are considered. Of course, the strongest difference can be found in the 1.5 °C scenario. However, the spatial patterns described below hold just as well for the 0.5 °C and the 1 °C scenario, except the changes are not as strong as in the 1.5 °C scenario (not shown).

Figure 3.13 indicates a strong, significant temperature increase for the interior of southern Africa (especially Namibia) with up to 2 °C for the austral summer. However, the temperature increase is very moderate and not significant along the coastlines of Mozambique and South Africa and at the Orange River mouth (below 0.5 °C). In austral winter, the strongest temperature increase also occurs in the inland regions of the simulation area. But this increase is less than in summer with values up to 1.5 °C. Due to a more uniform warming, the coastal areas get relatively warmer than in summer.

The inland warming coincides with a general decrease of precipitation in the centre of the simulation area. As illustrated in Fig. 3.14 the decrease of precipitation is strongest in summer (DJF) during the rain period. This is
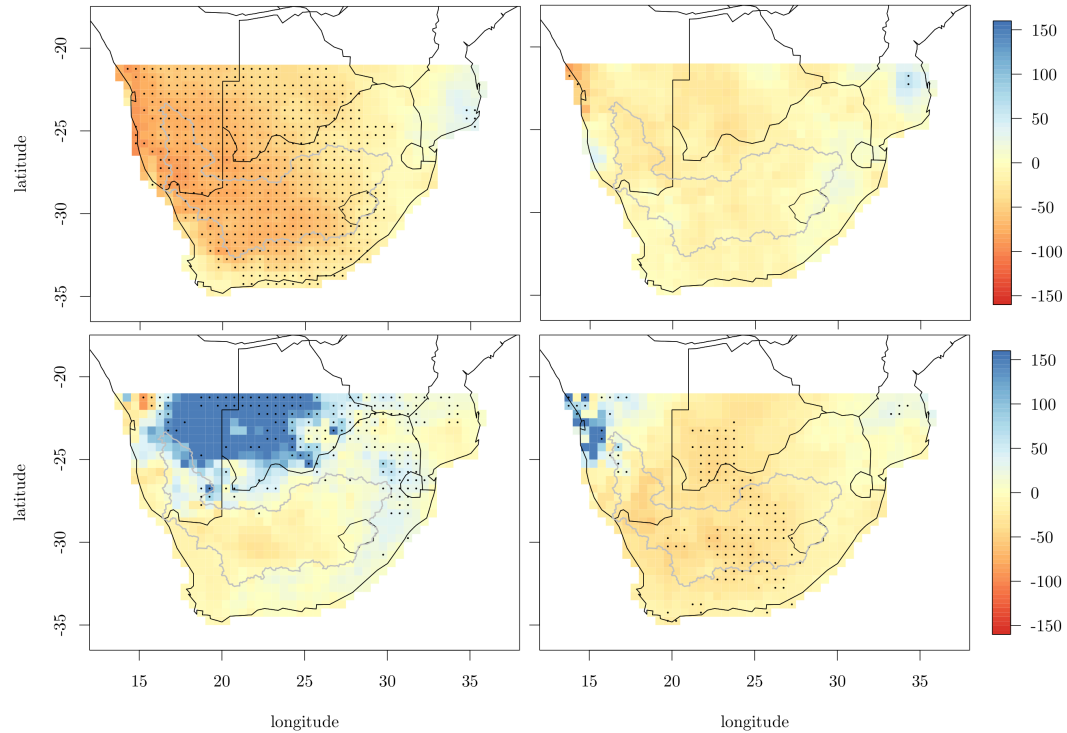

*see the appendix for a short description
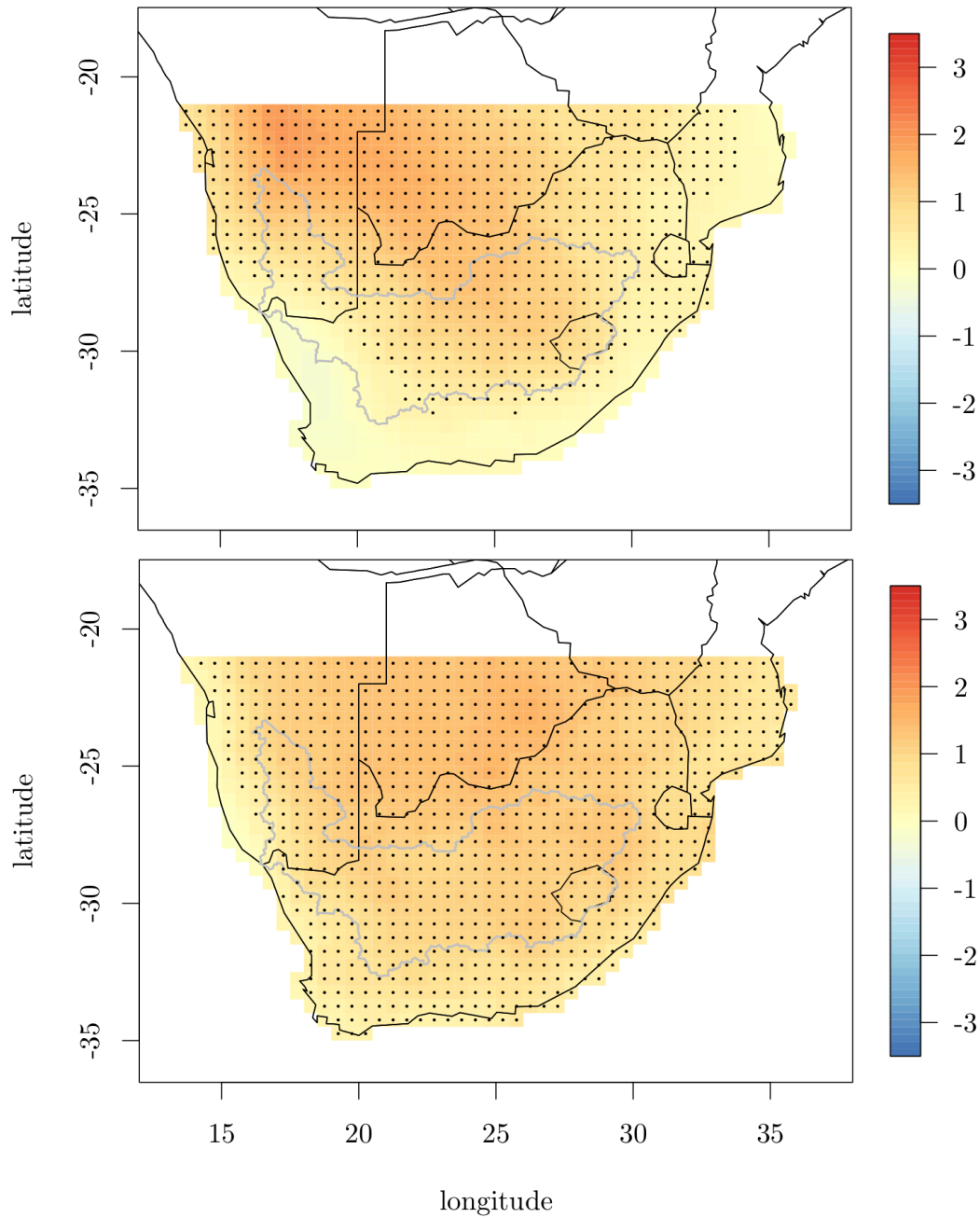
Figure 3.13. The projected change in the mean temperature (2036–2060 compared to 1976–2001) is shown. The 1.5 °C scenario (STARS) is illustrated. Top: DJF, bottom: JJA. The points indicate grid points with a significant difference between simulations and observations ($\alpha = 0.05$).

Figure 3.14. The projected change in the precipitation (2036–2060 compared to 1976–2001) is shown. The 1.5 °C scenario (STARS) is illustrated. Top-left: DJF, top-right: MAM, bottom-left: JJA, bottom-right: SON. The points indicate grid points with a significant difference between simulations and observations ($\alpha = 0.05$). All values above 150 % are shown in the same colour.

also the time when the change in precipitation is predominantly significant. In the arid areas (central South Africa, Namibia and Botswana) the monthly mean of the summer precipitation decreases by over 50 %, whereas the summer precipitation increases by up to 40 % in southern Mozambique. In this context it is important to look at Lesotho, where the predominant water volume of the Orange River originates. There, the summer precipitation decreases significantly by up to 30 %, which would result in a severe change for the Orange River. The right part of Fig. 3.14 shows that the changes are small and mostly not significant for the austral spring and fall. However, there is a precipitation increase in winter (JJA), when the precipitation typically reaches its lowest values. Especially Lesotho, eastern South Africa and Swaziland expect a (partly significant) increase of up to 40 %. The amounts of precipitation in the north-western parts of the simulation area are very low in this time of year, so that the values of the relative change are neglected.
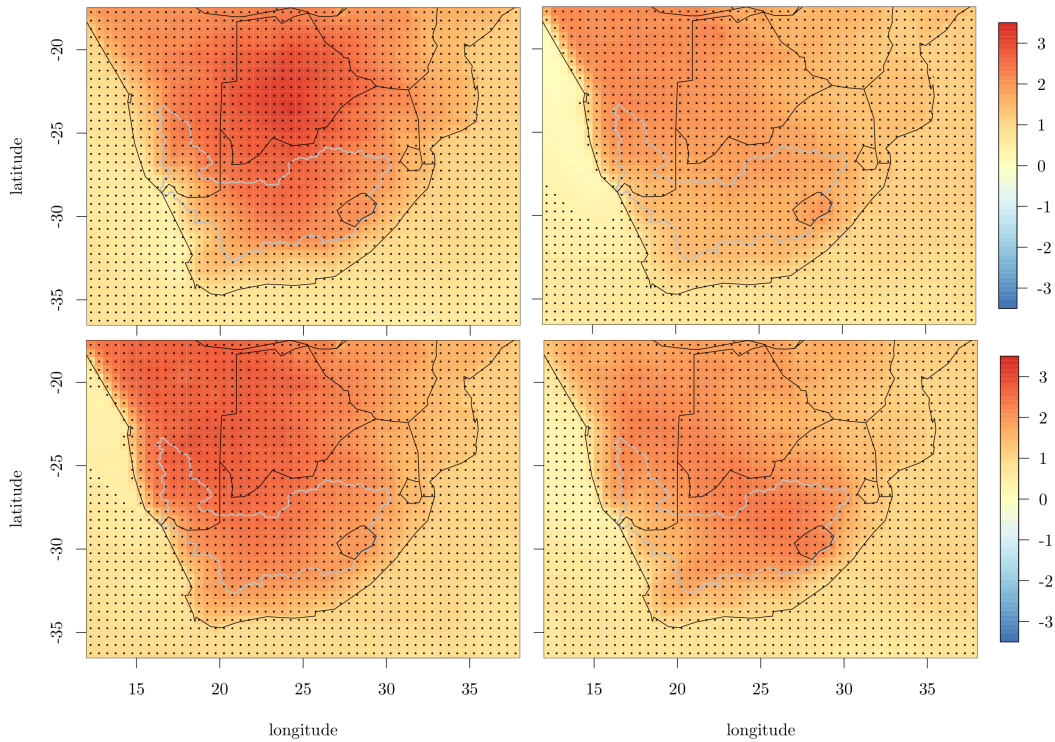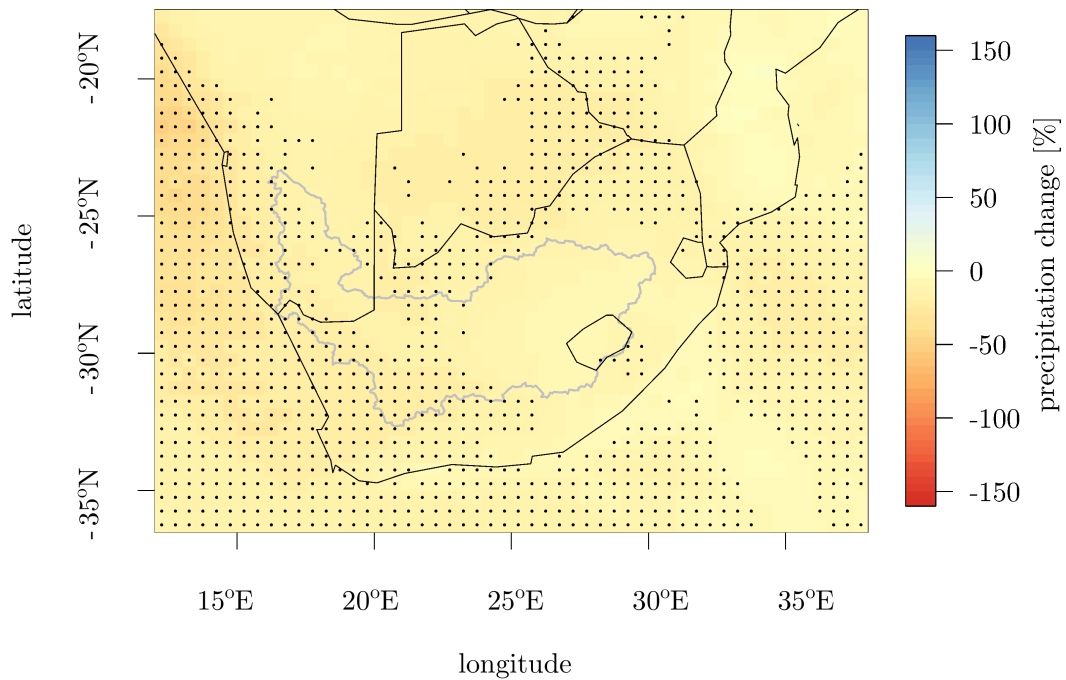
Figure 3.15. The CCLM-projected change in the 2 m temperature (2036–2060 compared to 1976–2000) is shown. Top-left: DJF, top-right: MAM, bottom-left: JJA, bottom-right: SON. The points indicate grid points with a significant difference between simulations and observations ($\alpha = 0.05$).

3.6.2. Climate projections with CCLM

Figure 3.15 shows the CCLM projection results for the individual seasons. More precisely, the difference between the average 2 m temperature during the projection period 2036–2060 and the reference period 1976–2000 is depicted. All CCLM projection changes are measured against the CCLM C20 simulation, instead of the WATCH data set, since this might even out model biases. According to CCLM, the increase is unevenly distributed across the seasons. In summer (DJF) and fall (MAM) the temperature is projected to increase by more than 2.5 °C in the northern Orange basin, and in Namibia in general. The largest increase is predicted for the Kalahari in summer, and amounts to 3 °C. In fall and winter, an increase in the 2 m temperature is projected as well; however the increases are not as severe as in summer and fall. They range between 1 °C and 2 °C. Overall, the temperature change projected by the A1B scenario amounts to about 1.6 °C, but compared to STARS for the 1.5 °C scenarios, CCLM sees more drastic temperature increases in some areas.

Figure 3.16. The CCLM-projected change in the annual average precipitation from 2036–2060 when compared to 1976–2000. The points indicate grid points with a significant difference between simulations and observations ($\alpha = 0.05$). All values above 150 % are shown in the same colour.

For instance, while STARS projects an increase of about 2 °C for the Kalahari in summer, CCLM sees an increase of up to 3.5 °C there.

For precipitation only the result for the annual average projection is shown, because the validation turned out to be rather poor. The projection changes are shown in Fig. 3.16. CCLM projects a general decrease in precipitation, while STARS projects a decrease in the centre of the simulation area and no change or increases in the east (not shown). In the areas where CCLM projects the largest decreases, the biases were largest, though, particularly in Namibia.

3.7. Conclusions

The statistical model STARS and the dynamical model CCLM were used to generate future climate projections for southern Africa. A validation experiment affirmed the applicability of STARS to southern Africa, yielding excellent results despite the demanding simulation area with high climatological variability. CCLM, on the other hand, performed reasonably well in the reproduction

of the 2 m temperature, but for precipitation the validation results turned out to be poor.

STARS generated ensembles of future climate projections based on daily gridded data for the time period from 1951 to 2000 for four different scenarios, given as a mean temperature trend over the time period from 2011 to 2060. In these projections the mean, maximum and minimum temperature showed a significant increase, especially in the interior of southern Africa. Such an increase is also projected by CCLM, but the spatial and temporal patterns in the projected temperature increase differ from STARS. While STARS projects seasonal increases of up to 2 °C, CCLM sees seasonal increases of up to 3.5 °C.

The area-averaged precipitation, as simulated by STARS, decreases by over 25 % in the rainy season (DJF). However, there is a slight increase in autumn and spring. The long term average of the spatial distribution (STARS) shows that the precipitation decreases mainly in the inland regions of southern Africa. This decrease is especially profound in austral summer with values of up to 50 %. In Lesotho, where the Orange River has its spring, and where its main runoff is generated, the precipitation decreases significantly by approximately 30 % in summer. In contrast, the winter precipitation increases significantly by approximately 40 % in Swaziland and the northeastern part of the simulation area. CCLM results with respect to precipitation seem not to be reliable.

Due to the projected precipitation changes, the effects on the Orange River catchment and associated consequences for the ecology and economy should be analysed using a hydrological model. As a consequence of its design, STARS is not able to simulate a climate that is essentially different from the present climate (unlike CCLM), since it is based on the assumption that statistical properties of the different variables are the same in the observation and the future time period. This is why the climate projections in this study cover only the years from 2011 to 2060.

This chapter shows that it is possible to run STARS with gridded data. Thus, it is possible to use reanalysis data instead of station data if the latter is not or only scarcely available in certain regions of the world. The very good results of the validation experiment in southern Africa indicate that STARS is not limited to small simulation areas or single river catchments. Thus, the simulation area is extended even to continental scale.

4. Continental scale projections with STARS

This chapter deals with the question whether the simulation area for the STARS model can be extended to the continental scale and how involving obstacles can be overcome.

Applications at continental scale currently lie in the focus of modelling and impact studies. As climate projections obtained by STARS play an important role in several impact studies it is crucial to provide results that are reasonable in terms of physical consistency. This also includes the annual cycle of the different variables. However, due to the modelling processes in the current STARS version (STARS_2.1) this may yield a seasonal mismatching, especially in shortwave radiation. To prove this, simulations have been carried out for the whole of Europe using the WATCH data set as input (see section 2.2). This setting is very demanding as it includes a high number of different climate regions and a large prescribed trend. In earlier simulations with STARS using less demanding settings (like in the previous chapter), the mismatch did not emerge. In the first part of this chapter (4.2) the mismatching was resolved by improving the current STARS version.

In the second part of this chapter (4.3) the new STARS version (STARS_2.4) is used to carry out ensembles of climate projections according to three different future scenarios for the whole of Europe.

Although the availability of dense and reliable station data for Europe has increased in recent years, it is still poor or restricted for certain regions. Thus, the WATCH data set is used instead. Furthermore, this makes the STARS results more comparable to dynamical RCMs or even GCMs than using station data.

The focus of the analyses of the climate projections lies in this chapter on the weather extremes associated with heat stress (hot days, dry days and dry periods). This is due to the subject of the research project the simulations have been carried out for. The project investigates how climate change affects forest fires in Europe and other fire-affected areas in the world. Naturally, a particular focus lies on weather extremes and climate phenomena, such as droughts and heat waves, since they have a large impact on fires.

Section 4.1 presents the input data and the experiment set-up for both parts of this chapter. The model development and the corresponding results are shown in section 4.2. Previous to the climate projections for Europe, a cross-validation experiment is carried out to analyse the performance of STARS_2.4.

The results of both simulations can be found in section 4.3. Finally, some conclusions are drawn in Sect. 4.4.

Note that parts of this chapter have been published in Lutz and Gerstengarbe (2014).

4.1. Data and experiment set-up

4.1.1. Data

In both parts of this chapter, STARS is forced with the WATCH data set. The years 1950–2001 are used in the first part, whereas the time period is extended to 1942–2001 in the second part to receive a time series of sixty years. By this means the output time period for the climate projections can be extended to sixty years as well, because the output time period should have at least the same length as the input time period to ensure a high variability in the simulations. The simulations are carried out for Europe and northern Africa. They cover 17492 grid points with a resolution of $0.5° \times 0.5°$, ranging from $25.25°N$ to $71.25°N$ and from $-14.75°E$ to $44.75°E$ (see Fig. 4.1).

To reduce the complexity and the computational time the model process explained before is not applied on all 17492 grid points but the simulation area is split into seven climatological sub-regions for the first section and into six sub-regions for the second section. These sub-regions are identified using a standard k-means clustering method (Hartigan, 1975) based on daily temperature and precipitation that is implemented in the model. The results for both sections are shown in Fig. 4.1. In this chapter, the sub-regions are called "Western Europe and Mediterranean (WEM)", "Western Europe (WE)", "Northern Europe (NE)", "Mountains (M)", "Eastern Europe (EE)", "Northern Africa (NA)" and "Central and Eastern Europe (CEE)". The grid points that are closest to the centre of mass of a cluster represent this particular cluster/sub-region.

The choice of the sub-regions is done subjectively, based on the principle to choose as many as necessary but as few as possible. The analysis of the clustering results suggests to choose seven sub-regions for the first section of this chapter, whereas only six sub-regions are sufficient in the second part of this chapter.

The rearrangements are carried out for one characteristic variable and for the corresponding representative grid points but the resulting date-to-date mapping prevails for all other grid points and variables, as the model assigns each day in the simulation time period a day from the input time period. The results are simulations that are physically consistent in terms of the relations between all variables and spatial patterns.
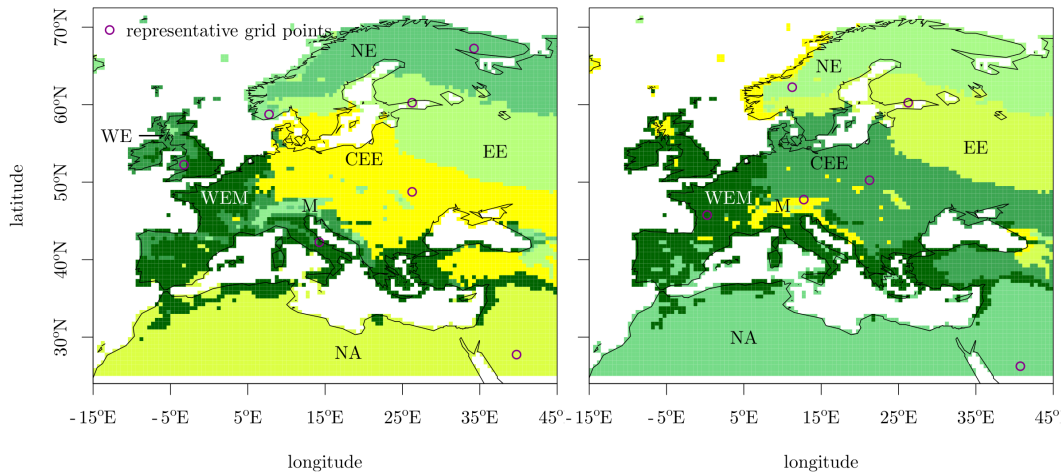
Figure 4.1. Result of the cluster analysis to determine regions with a similar climate (left: applies to 4.2, right: applies to 4.3). Coloured areas represent the different clusters, where circles mark the grid points closest to the centre of mass (representative grid points). The abbreviations of the regions are explained in the text.

4.1.2. Experiment set-up

To evaluate the applicability of the different model versions in Europe, a cross validation experiment is carried out for both parts of this chapter. Therefor, the corresponding time period of the WATCH data set is divided into two parts. The first part is used to compute simulations for the second part. The temperature trend for the second part (the actual model forcing) is derived from the WATCH data set by a regression analysis of the annual mean temperature series at the representative grid points. The performance of STARS is evaluated comparing the simulated climatology with the one derived from the WATCH data set. An ensemble of 100 simulations is generated for both computations to quantify the model uncertainties.

The model reproduces the given trend with a user-defined tolerance for each representative grid point. In section 4.2 the temperature trends ranges from $1.0\,°C$ in 26 years for WE to $2.0\,°C$ in 26 years for EE. The tolerance was chosen to be between $0.25\,°C$ and $0.35\,°C$, depending on the temperature trend at the different grid points. In section 4.3 the lowest temperature trend can be found in CEE with a value of $0.9\,°C$ for 30 years. The region NA has the highest trend with a value of $1.5\,°C$ for 30 years. The tolerance was chosen to be between $0.15\,°C$ and $0.2\,°C$.

Following the cross validation in section 4.3, climate projections for the years 2001–2060 are carried out with STARS_2.4 according to the RCP scenarios 2.6, 4.5 and 8.5 (Moss et al., 2010). The linear temperature trend prescribed for this time period is calculated from the mean of 22 GCMs from the CMIP5

project. These scenarios were chosen to cover a light, a moderate and a strong temperature trend for the future. The trend from each scenario is assigned for each representative grid point individually. As the average resolution of the GCMs is coarser than that of the WATCH data set, the trend of the grid box that embeds the representative grid point is used. The temperature trend (averaged over the representative grid points) for the RCP 2.6 scenario amounts to 0.9 °C during the simulation time period. For the RCP 4.5 scenario it amounts to 1.6 °C and for the RCP 8.5 scenario the mean temperature trend is 2.4 °C.

Please note that the prescription of a future temperature trend does not depend necessarily on GCM data but could also be based on other assumptions (e.g., assuming 1.0 °C, 1.5 °C and 2.5 °C).

4.2. Improving seasonal matching in STARS

4.2.1. Model development

Note that the general modelling procedure is explained in section 2.1. This section focuses on the second step of the process, namely the replacing of the 12-days-blocks.

The workflow of STARS_2.1 on replacing the blocks in the second step (cf. Fig. 2.1) was organised as illustrated in Fig. 4.2. To generate a pool of possible candidates of blocks for replacement to improve the first approximation, the model firstly chooses blocks from the same cluster class as the blocks in the artificial time series. This ensures that the linear regression of the first approximation moves closer to the prescribed trend. Secondly, only unused blocks are kept in this pool. This avoids a too frequent use of certain blocks. Thirdly, the model checks which blocks lie within the user-defined ordinal date window. Blocks that do not lie in this ordinal date window, and thus are responsible for a possible seasonal mismatching, are removed from the pool. If no block is left in the pool afterwards, the window is extended by 2 days. This step is repeated as long as the ordinal date window is not larger than 40 days. The procedure up to this point is shown in blue in Fig. 4.2.

If the ordinal date window has already been extended to 40 days and there is still no block left in the pool, the model uses all unused blocks that belong to the same cluster as the artificial time series, not considering the ordinal date window. This approach is highlighted in orange in Fig. 4.2. Finally, STARS rejects the blocks that do not match their preceding and succeeding (if already defined) block.

This means that all blocks that belong to the appropriate cluster class, match with the predecessor and successor and are not used yet can be used for re-
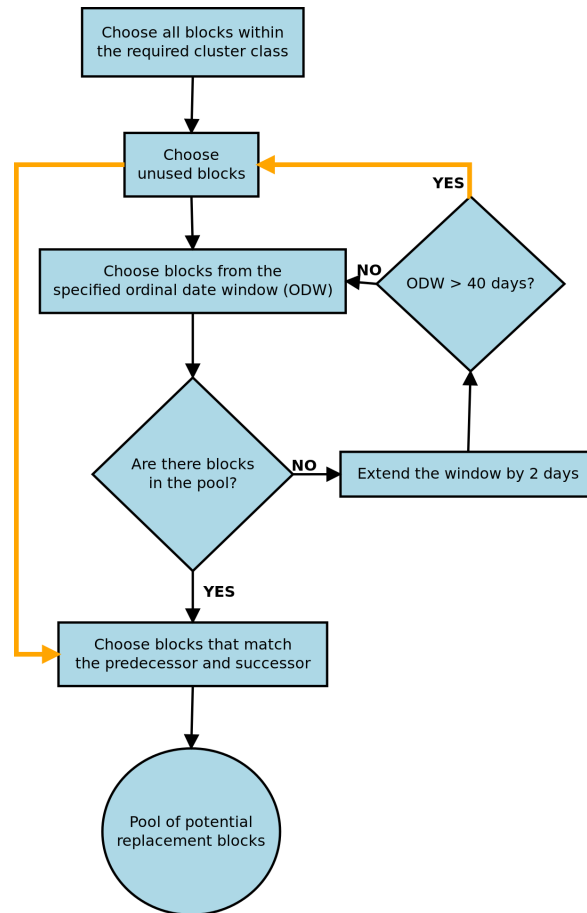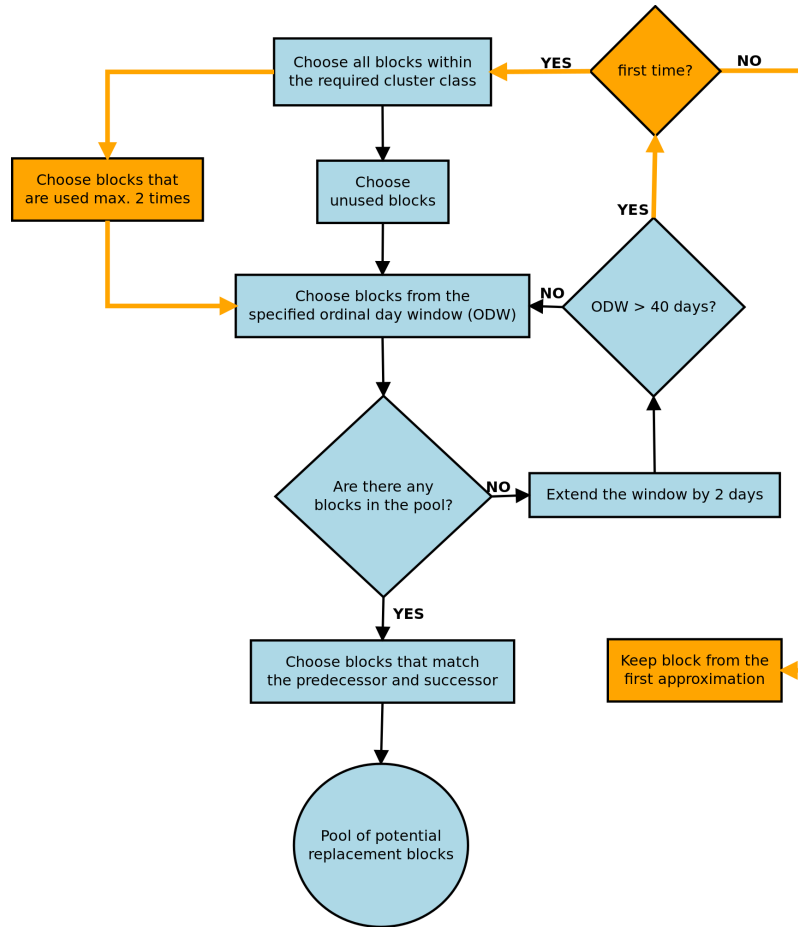
Figure 4.2. Process chart to illustrate how a pool of potential candidates for replacement (step 2 in the model process) is generated in STARS_2.1. The orange arrows are carried out if the ordinal date window (ODW) is greater than 40 days.

placement. They do not have to lie within the ordinal date window of 40 days around the block to be replaced. In this case the model neglects the seasonal matching, more precisely it is possible that the model picks blocks from the entire calendar year to replace a certain block.

The new version of STARS (STARS_2.4; illustrated in Fig. 4.3) accounts for the seasonal matching. To this end, an additional step (shown in orange) is introduced after the check of the ordinal date window. If there is no block in the pool after the extension of the window to more than 40 days, the model returns to the whole sample of the same cluster class instead of considering the unused blocks only. Hence, unused and already used blocks are accounted for in the new version. However, STARS_2.4 can use the same block only up to three times to avoid a too frequent use of blocks. Then, the model checks

Figure 4.3. Process chart to illustrate how a pool of potential candidates for replacement (step 2 in the model process) is generated in STARS_2.4. The orange part is carried out if the ordinal date window (ODW) is greater than 40 days. This part has been changed in STARS_2.4.

again if the blocks belong to the user-defined ordinal date window. Note that STARS_2.1 skips this step after returning to the unused blocks. The window is repeatedly extended by 2 days if no block can be found. In the rare event of finding no block that lies within the 40 days window, the original block is kept from the first approximation, i.e., from the rearrangement of entire calendar years that fit best the given linear trend.

4.2.2. Simulation results

STARS_2.1 was used to simulate the years 1976-2001. The evaluation of the simulation results reveals the weak point of this model version. It is not able to reproduce the seasonality of the shortwave radiation and the temperature.

To demonstrate this, the simulated and observed mean annual cycles of daily shortwave radiation and maximum temperature for the representative grid points are shown in Figs. 4.4 and 4.5. The effect is the same for daily mean and minimum temperature, however the maximum temperature is shown because the effect is most prominent in this case. The other variables do not exhibit such a seasonal dependency, this is why they are not examined. The spread of all 100 realisations for both model versions is displayed.

To emphasise the difference between the two different model versions a statistical test has been carried out. As the old model version is likely to pick not only neighbouring blocks for replacement, but also to choose blocks from outside a 40-day-window around the block to be replaced, it is expected that the variance is significantly larger in the old version than in the new one, especially for radiation. Thus, an F-test* (Snedecor and Cochran, 1989) to compare the variances of both distributions has been applied. To confirm the test assumptions, the data for the single days was tested for normality (using a Shapiro-Wilk normality test*; Shapiro and Wilk, 1965) showing almost normally distributed data for all days in both versions. This is supported by the fact that the non-parametric Levene's test* (Levene, 1960) for equality of variances did come up with mostly the same days of differing variances (not shown). The results of the F-test are shown in Figs. 4.4 and 4.5.

The mean annual cycle of the radiation computed with STARS_2.1 shows a very high overestimation in autumn. In fact, this overestimation is a further maximum in the annual radiation curve that is not existent in the WATCH data set. Furthermore, a slight decrease of shortwave radiation can be seen in spring (approximately at day 100), especially for the grid points WEM, EE, CEE and NA. The results of the F-test confirm the assumption that the variance is mostly significantly smaller for the new STARS version. This coincides with a much larger spread in the STARS_2.1 simulations, especially in winter.

The second maximum in the radiation curve has disappeared in the simulations with STARS_2.4. Additionally, the simulated radiation does not decreases in spring. This version is able to reproduce a realistic annual cycle as observed in the WATCH data set. However, it still overestimates the shortwave radiation in spring and autumn. In winter and summer the accordance of simulated and observed radiation is very good for all representative grid points.

The simulated (STARS_2.1) overestimation of daily shortwave radiation in autumn coincides with an overestimation of daily temperature, especially of daily maximum temperature (see Fig. 4.5). The overestimation is most promi-

---

*see the appendix for a short description
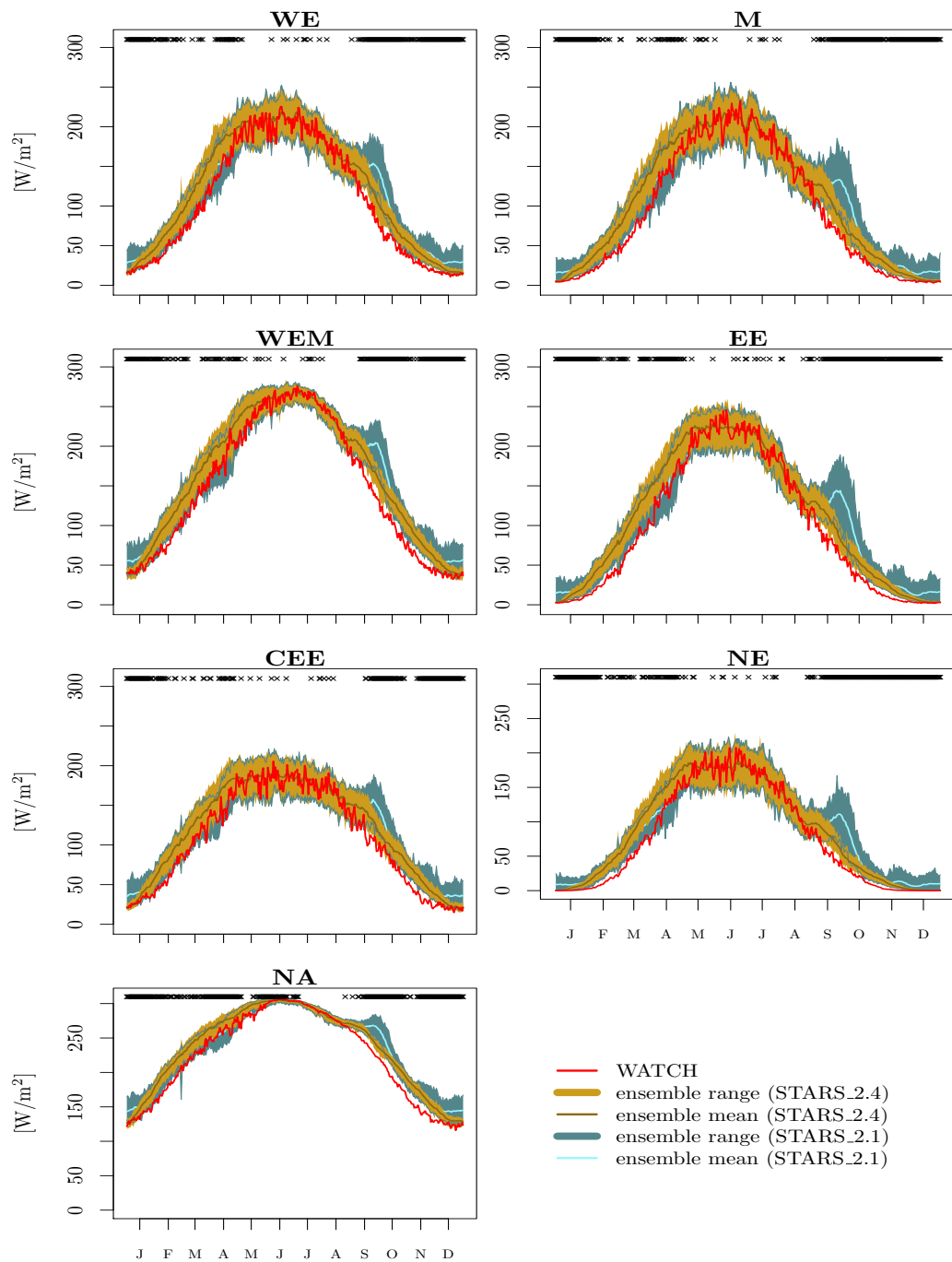
Figure 4.4. Mean annual cycle of daily shortwave radiation for each representative grid point (see Fig. 4.1), computed with STARS_2.1 and with STARS_2.4. Averaging was applied for the time period 1976–2001. The crosses indicate days on which the variance within the ensemble of 100 simulations of STARS_2.4 is significantly smaller (with significance level 0.1) than with STARS_2.1.
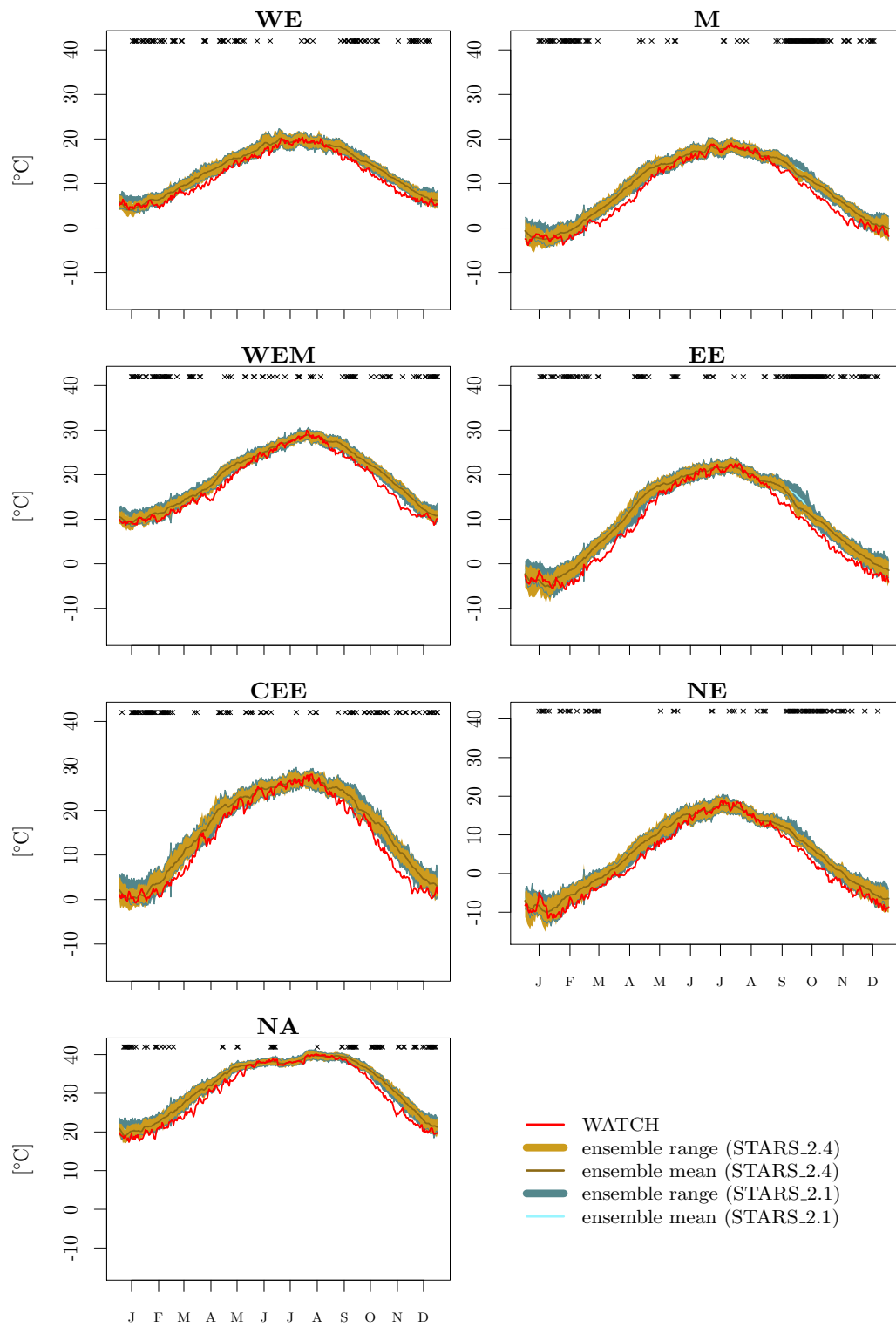
Figure 4.5. As Fig. 4.4 but for daily maximum temperature.

Table 4.1. Comparison between simulated dates (rows) and dates that were picked from the observations (columns) by STARS_2.1 to generate the simulated time series. The values indicate the number of agreements for 100 simulations.

Month	J	F	M	A	M	J	J	A	S	O	N	D
J	34483	15230	9061	1329	0	0	0	0	0	76	6379	14042
F	2318	34236	30699	2454	17	0	0	0	0	690	2713	373
M	0	1194	34264	37185	240	0	0	0	136	5990	1519	72
A	0	0	211	35085	30780	144	0	0	528	10665	587	0
M	0	0	0	195	50169	28530	499	749	287	163	8	0
J	0	0	0	0	671	57196	19375	683	75	0	0	0
J	0	0	0	0	0	5577	65554	9468	1	0	0	0
A	0	0	0	0	0	403	7459	70617	2121	0	0	0
S	0	0	0	0	5219	6834	1465	34568	29340	574	0	0
O	0	0	0	4	5271	5600	90	160	21908	46846	721	0
N	0	5	974	2464	157	22	0	0	82	37639	35996	661
D	2126	414	5457	2785	36	0	0	0	0	974	36098	32710

nent for the grid points M and EE, where the annual temperature variability is particularly high. However, the effect is by far not as strong as for radiation.

The annual cycle of the maximum temperature simulated with STARS_2.4 fits better to the WATCH data set as well. The results of the F-test emphasise the smaller variance in the new STARS version in spring and autumn. Although, the model still overestimates the temperature values in spring and autumn but the bias is very small.

Table 4.2. Comparison between simulated dates (rows) and dates that were picked from the observations (columns) by STARS_2.4 to generate the simulated time series. The values indicate the number of agreements for 100 simulations.

Month	J	F	M	A	M	J	J	A	S	O	N	D
J	39138	22658	2530	0	0	0	0	0	0	0	1574	14700
F	1687	33876	36266	1661	0	0	0	0	0	0	0	10
M	0	873	35060	43648	1019	0	0	0	0	0	0	0
A	0	0	189	37518	39973	320	0	0	0	0	0	0
M	0	0	0	158	45126	35031	285	0	0	0	0	0
J	0	0	0	0	682	56968	20331	19	0	0	0	0
J	0	0	0	0	0	6841	63407	10349	3	0	0	0
A	0	0	0	0	0	39	8224	70147	2190	0	0	0
S	0	0	0	0	0	0	57	36314	40521	1108	0	0
O	0	0	0	0	0	0	0	142	29508	50548	402	0
N	0	0	0	0	0	0	0	0	128	43988	33284	600
D	2382	0	0	0	0	0	0	0	0	1028	43163	34027

Looking at the annual cycles it is evident that the high values, particularly of shortwave radiation, in autumn must come from another time of the year, namely either from some weeks earlier in autumn or even from spring or summer. This means that during the simulation process, STARS_2.1 assigns a block from summer, spring or early autumn to one in late autumn. A deeper evaluation confirms this suspicion. In Table 4.1 the days of the simulation time period are compared to the days that were picked from the observations by STARS_2.1. The rows indicate the months of the simulation time period and the columns indicate the associated months of the observations. Typically, STARS_2.1 picks most days from the same month but it is possible that this

version assigns days e.g. from April to days in November or the other way round. However, the latter possibility is not as striking as the first because it is rare. For instance, to simulate days in November the model picked 2464 days from April, whereas it only picked 587 days from November to simulate days in April.

The comparison of the simulated dates and the dates picked from the observation time period for STARS_2.4 are shown in Table 4.2. It confirms the considerations described above. With the new model version, it is not possible that the model picks blocks that are outside a 40 day window around the block to be replaced. For the example from above it means that no days from April were picked to simulate days in November and no days in November were picked to simulate days in April.

4.2.3. Discussion

The seasonal mismatching problem was found to have its seeds in the replacement of blocks in the second step of the modelling process. A huge climatologically differing region requires a high number of climatological sub-regions and corresponding representative grid points. This makes it difficult for STARS to find blocks that follow the prescribed trend at all representative grid points and fulfil the necessary criteria described in 2.1. A high prescribed trend adds to this difficulty. In such a demanding simulation setting, the criterion to use blocks only once outweighs the criterion that the blocks have to lie within a defined date window. However, the second criterion is much more important for a successful simulation of the annual cycle. Note that for less demanding settings with fewer climatological sub-regions, replacements within the date window are much easier to achieve. Thus, earlier simulations with STARS do not face this problem so rigorously.

The problem was solved successfully, resulting in a new version of the model (STARS_2.4). Now, the blocks must lie in a date window of at most 40 days but they can be used up to three times. The simulation results computed with STARS_2.4 show a realistic annual cycle of the shortwave radiation and the maximum temperature, even for the high demanding setting, whereas the simulations computed with STARS_2.1 result in a second maximum in autumn and a slight decrease in spring, especially for shortwave radiation. The new model version prevents the possibility to choose blocks from outside a 40-days window. This leads to a significantly smaller variance on most days, generally improving the simulation ensemble. Definitely, it results in more realistic annual cycles.

The seasonal mismatch is very profound because of the imbalance between the shortwave radiation and the weather conditions during the year. Although,

considering the meteorological definitions, spring and autumn have very similar weather conditions, the potential shortwave radiation is higher in the spring months than in the autumn months due to the fact that the equinoxes, on which the potential shortwave radiation has the same values, are in March and September. As STARS tends to pick warmer blocks to achieve the prescribed positive temperature trend, STARS_2.1 often uses blocks from spring to replace those in autumn as the weather conditions are similar. Thus, the higher values of the shortwave radiation in spring lead to an unrealistic second maximum in the annual cycle if they are used in autumn. This problem happens also the other way round: blocks from autumn are chosen to replace blocks in spring. This happens less frequent as the probability to find a warmer neighbouring block in spring or early summer is higher than to find it in autumn. Hence, the decrease of the shortwave radiation in spring is very small (but still visible).

The above mentioned tendency to choose warmer blocks is a characteristic of the STARS model (and similar models) when it tries to achieve the prescribed increasing temperature trend. The probabilities are high to choose blocks which have either a higher ordinal date (in spring) or a lower ordinal date (in autumn), because of their generally higher temperature values. This leads to overestimated values in spring and autumn.

In this case and this particular region the problem is limited to shortwave radiation and temperature as these variables exhibit a strong seasonal dependency. In other regions other variables like precipitation or relative humidity could be affected as well. In particular cases the date window could be reduced. The local weather and climate conditions should thereby be taken into account.

As can be seen in the results of STARS_2.4, the maximum of the shortwave radiation has not completely disappeared. This is due to the fact that the date window can still be extended to 40 days, which means that the model can pick blocks from September to replace blocks in early November, for example. This again happens especially for simulations with a demanding setting. Thus, it is important not to overstress the simulation conditions.

4.3. Evaluation and application of STARS_2.4 in Europe

4.3.1. Results of the cross-validation experiment

The validation is carried out for the time period 1972–2001. In general, the observed climate is very well reproduced by STARS (not all results are shown
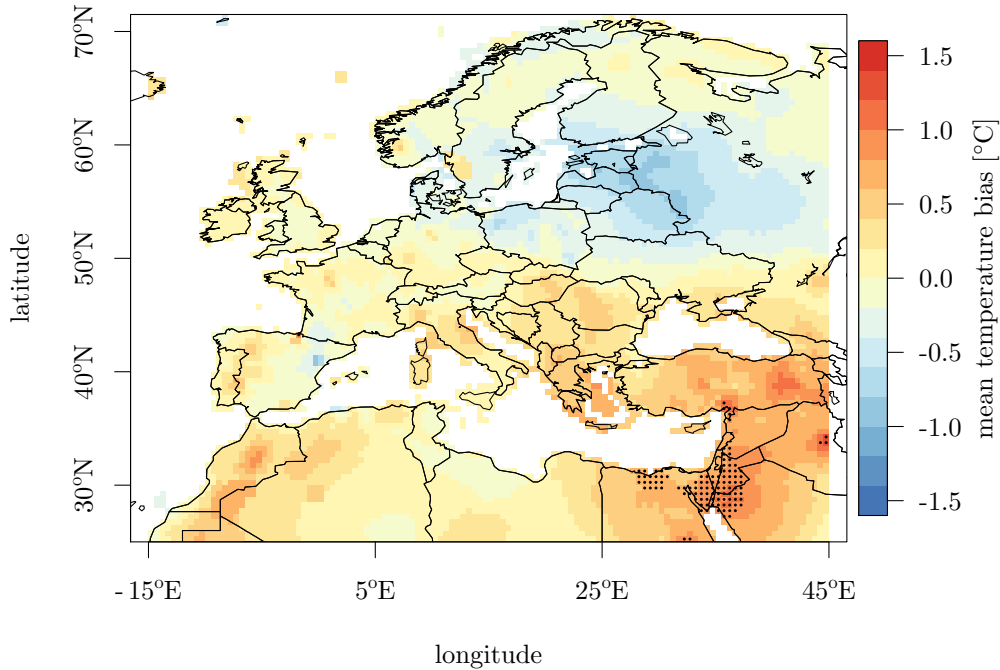
Figure 4.6. Absolute long-term difference between simulated and observed (WATCH) mean temperature for 1972–2001. Black dots indicate areas with a significant difference ($\alpha = 0.05$).

here) despite the demanding setting of a high prescribed trend and a large simulation area.

The temperature and precipitation biases are shown in Figs. 4.6 and 4.7. The 100 STARS realisations were sorted by the climatic water balance and the median was chosen for these diagrams. The climatic water balance is the difference between precipitation and potential evapotranspiration. It characterises the water availability in a certain region and thus provide an indication of the predominant vegetation. In this case, it classifies the different simulation runs in rather dry or rather wet realisations.

As can be seen the biases are small, especially for the precipitation which is on average below 20 %. The high values of the relative bias in the arid northern African region are a result of the small observed values. For temperature, the bias is positive (which means that the simulated temperature is higher than the observed) in southern Europe, whereas there is a negative bias in the north-eastern parts of Europe.

The differences are tested for significance using the Wilcoxon-Mann-Whitney test* (Hollander and Wolfe, 1999) with a significance level of 5 %. According

---

*see the appendix for a short description
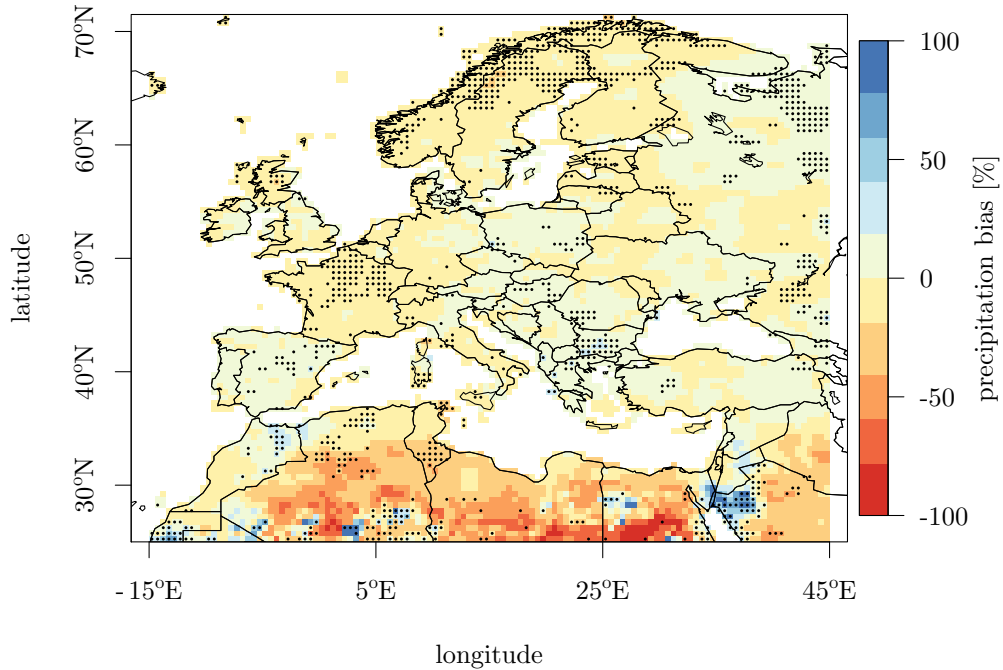
Figure 4.7. Relative long-term difference between simulated and observed (WATCH) precipitation for 1972–2001. Black dots indicate areas with a significant difference ($\alpha = 0.05$).

to this test the simulated and observed mean temperatures differ significantly only at some grid points in the south-eastern part of the simulation area, where the simulated temperature is approximately $1\,°\mathrm{C}$ higher than the observed. In most places, the simulated and observed precipitation does also not differ significantly. Exceptions can be found in northern and western Europe, where the model is approximately $20\,\%$ too dry.

Figure 4.8 shows the number of dry periods as observed in the WATCH data set and the absolute difference between the simulated and observed number of dry periods. A dry period is hereby defined as more than five consecutive days on which the daily precipitation is less than $1\,\mathrm{mm}$. As can be seen, STARS tends to overestimate the number of dry periods in most parts of the simulation area. However, in Italy and south-eastern Europe (where the number of dry periods is highest), as well as in the north west, the model underestimates the number of dry periods. For the most parts the bias is approximately $10\,\%$. Due to little precipitation or no precipitation at all, the number of dry periods is very low in northern Africa.

Figure 4.9 shows whisker plots for the number of dry periods during the summer months (JJA) at the six representative grid points. Shown is the variability of the 100 realisations, where the boxes cover the interquartile range.
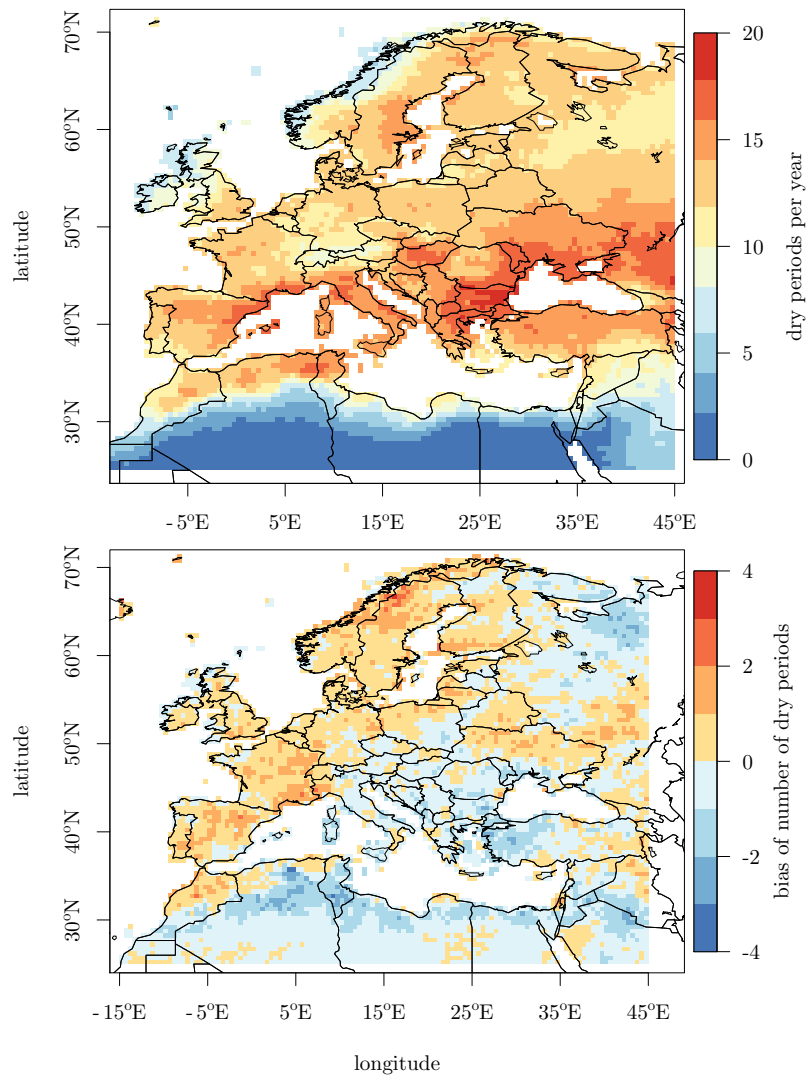
Figure 4.8. Top: Number of dry periods per year as observed in WATCH, averaged over the years 1972–2001. Bottom: Bias of the number of dry periods per year between simulations and WATCH.

As can be seen the model simulates approximately 10 % more dry periods in WEM and NE during the validation time period (first box from left). These results correspond to the simulated precipitation in northern Europe, where STARS is too dry on average.

The 90th percentile of the daily maximum temperature is shown in Fig. 4.10. This value is often used as a threshold to identify hot days. The value is well reproduced in NE and M, whereas STARS overestimates the 90th percentile in CEE and underestimates it in WEM, NA and EE.
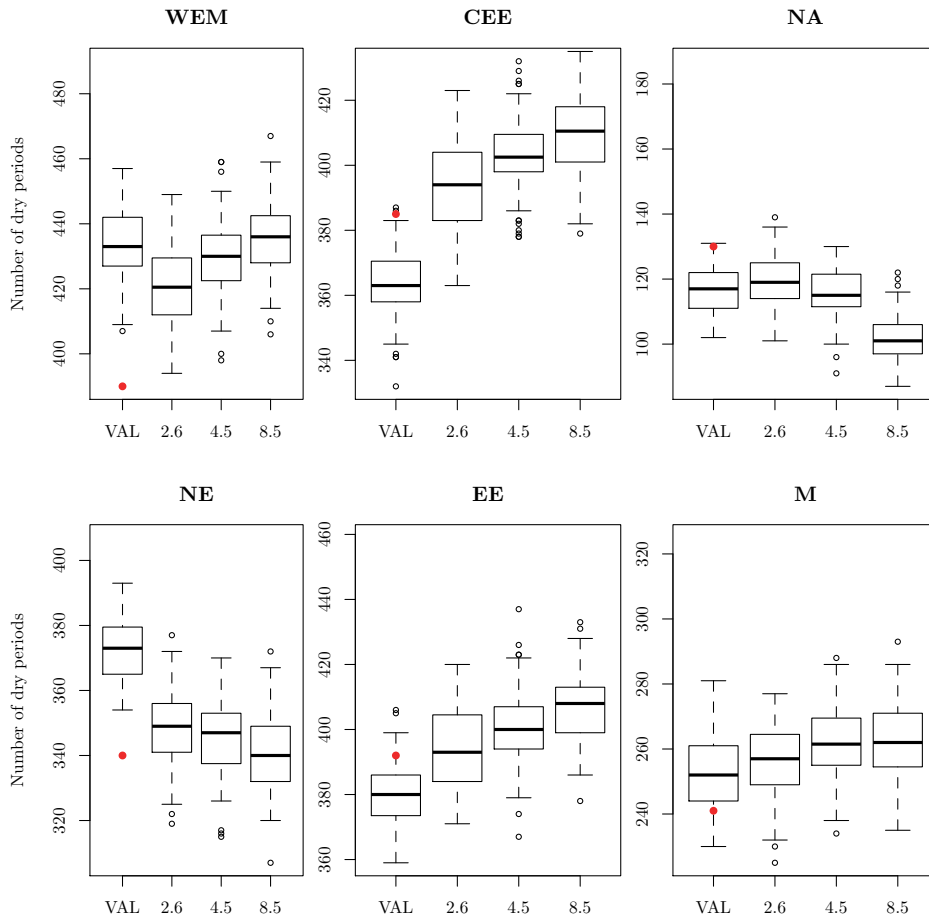
Figure 4.9. Number of dry periods for the six representative grid points in JJA (see Fig. 4.1 for the abbreviations). The validation time period (VAL, 1972-2001) and the last 30 years of the three future scenarios are shown (2031-2060). The red dot represents the observations (WATCH). The whiskers cover the range between 10th and 90th percentile of the 100 realisations and outliers are represented by circles.

Figure 4.11 presents the 90th percentile of the precipitation. The values are reasonably well or even very good reproduced by STARS for all the representative grid points. The values for NA are of course very small, thus a qualitative evaluation is not possible.

The precipitation trend is illustrated in Fig. 4.12. Note that the same attributes as for the other whisker plots applied for this plot, except that all months were taken into account, not only JJA. The trend is very good reproduced for CEE, NA, NE and EE. However, STARS underestimates the trend for WEM and M. These grid points have also the largest simulation spread.
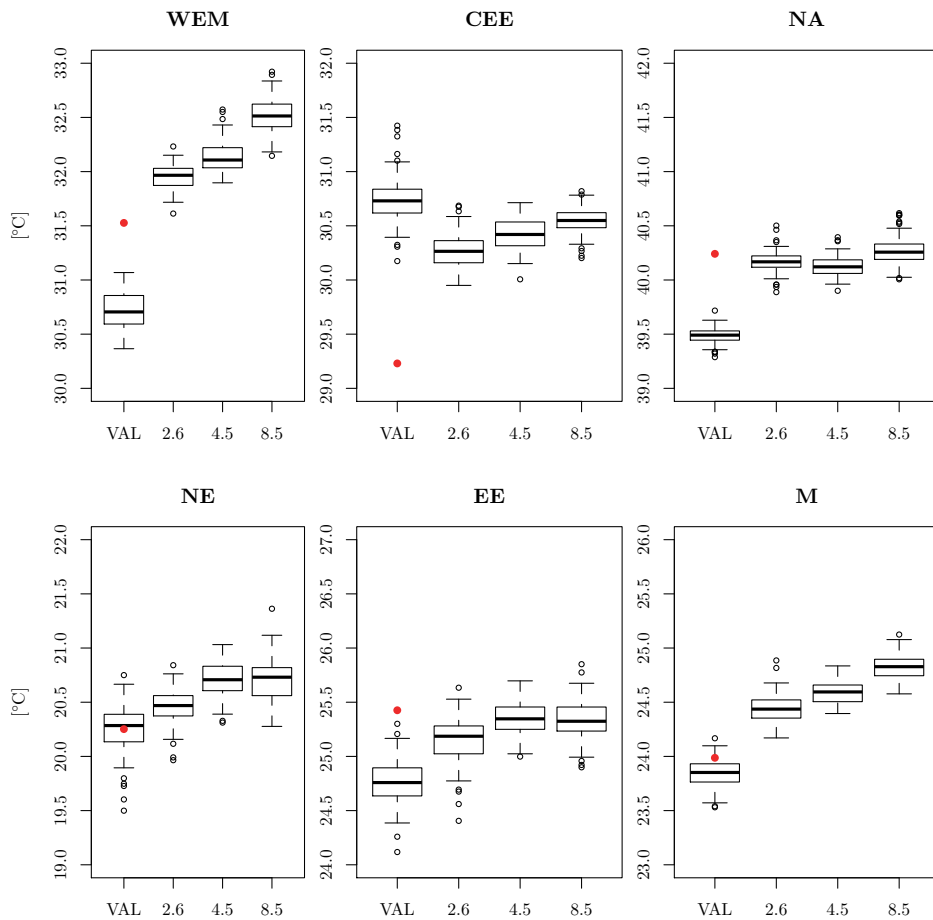
Figure 4.10. As Fig. 4.9 but for the 90th percentile of daily maximum temperature.

4.3.2. Results of the climate projections

To show future climate developments, the last thirty years of the projections (2031-2060) for the RCP 2.6, RCP 4.5 and RCP 8.5 scenarios are compared to the STARS simulations for the years 1972-2001. This is done to remove possible constant model biases.

Figure 4.13 shows the spatial distributions of changes in mean temperature and precipitation compared to the validation time period for all three RCP scenarios. Again, the 100 STARS realisations of every scenario were sorted by the climatic water balance and the median realisation is illustrated. All three scenarios project a general temperature increase with a maximum in north-eastern Europe. The temperature changes are largest for the RCP 8.5 scenario, where the temperature increases by about 2 °C in central Europe and
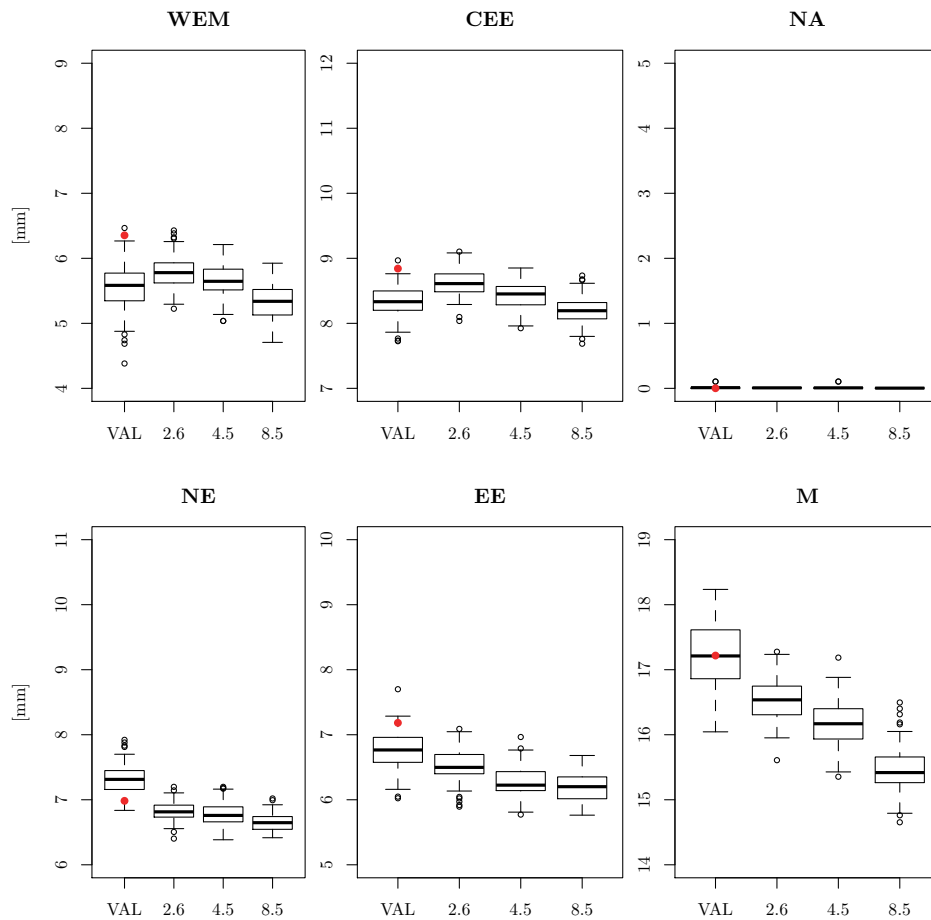
Figure 4.11. As Fig. 4.9 but for the 90th percentile of precipitation.

up to 4 °C in eastern and northern Europe.

In case of precipitation, all three scenarios project a significant increase in northern Europe and a significant decrease in some parts of southern Europe. This pattern is especially pronounced for the RCP 8.5 scenario, where the precipitation decreases in parts of Spain, the Balkan, Bulgaria, Turkey and the Middle East, while it increases in northern parts of Scandinavia. However, the changes are small, hardly reaching values of 20 %.

Figure 4.14 shows the relative difference in summer precipitation between the projections and the simulations for the past using the RCP 4.5 scenario as an example. The results for the other scenarios are similar and thus are not illustrated. It shows a significant decrease for Great Britain, central, northern and north-eastern Europe, whereas the precipitation significantly increases around the Black Sea.
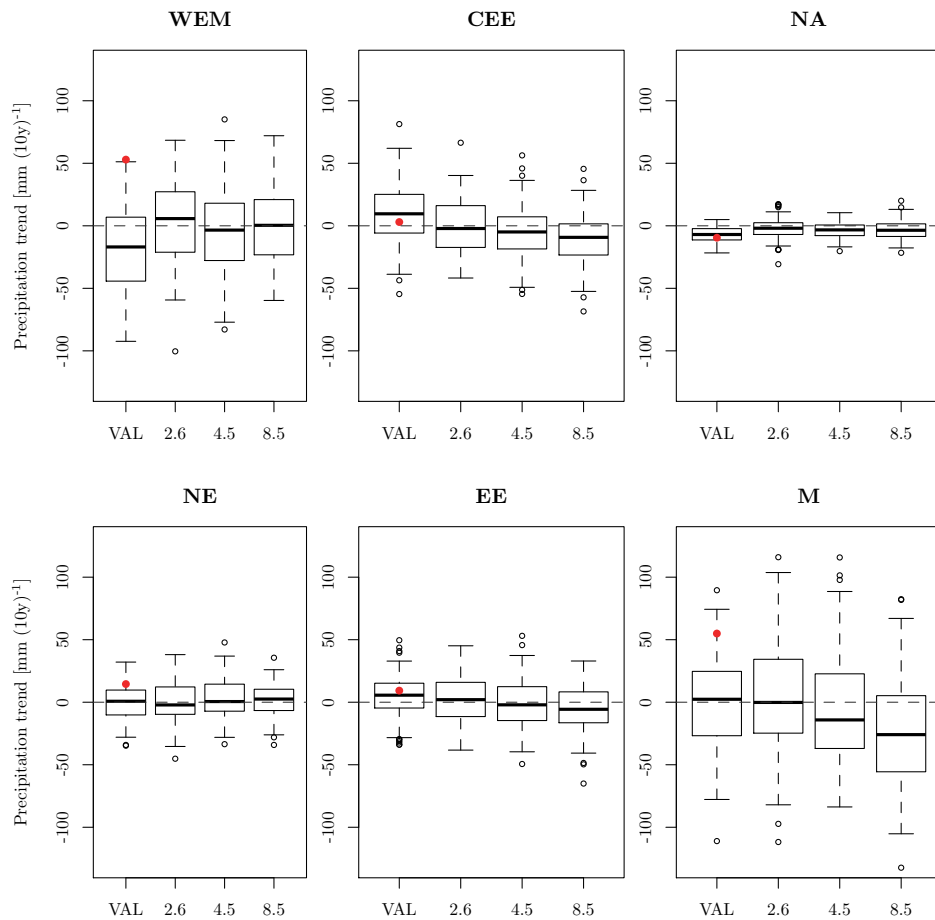
Figure 4.12. As Fig. 4.9 but for the precipitation trend (in mm per decade) and covering the whole years, not only JJA.

Figure 4.9 shows the changes in the number of dry periods at the six representative grid points in JJA. For WEM and M there are almost no changes. This result is in accordance with the not significant precipitation change in these regions. NA and NE show a decreasing trend. Although there is a significant precipitation increase in northern Europe for the yearly average (see Fig. 4.13), the summer precipitation decreases significantly for this region (see Fig. 4.14). The decrease of dry periods and the decrease of summer precipitation in this region suggests that despite the less amount of precipitation, the amount of precipitation events in summer will increase. The number of dry periods increases in EE and especially in CEE. This increase can be explained by looking at the summer precipitation (see Fig. 4.14), that will decrease significantly in both regions.
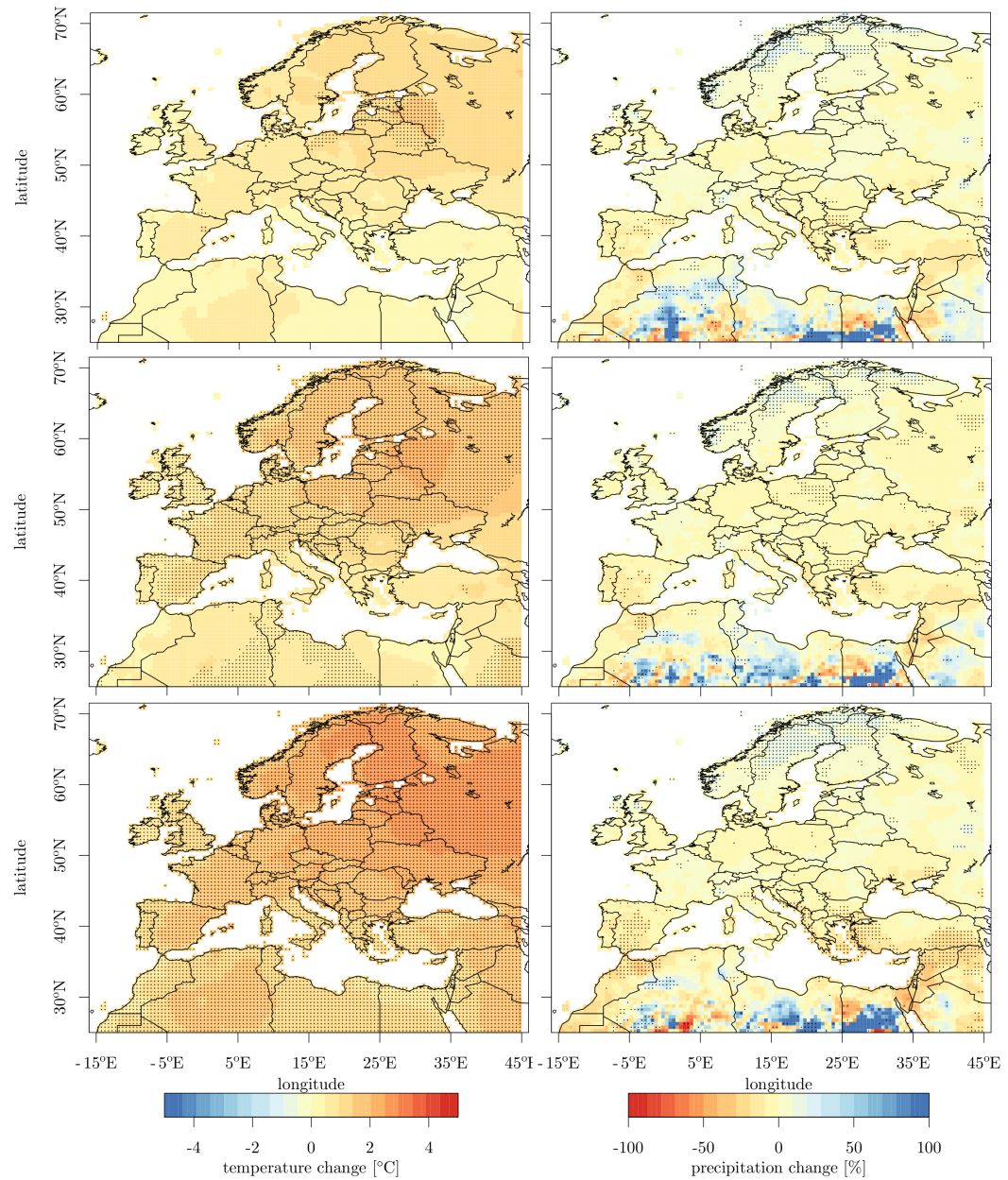
Figure 4.13. Difference between projected and simulated mean temperature (left) and precipitation (right), 2031–2060 vs. 1972–2001. Black dots indicate grid points with a significant difference ($\alpha = 0.05$). Top: RCP 2.6, middle: RCP 4.5, bottom: RCP 8.5.
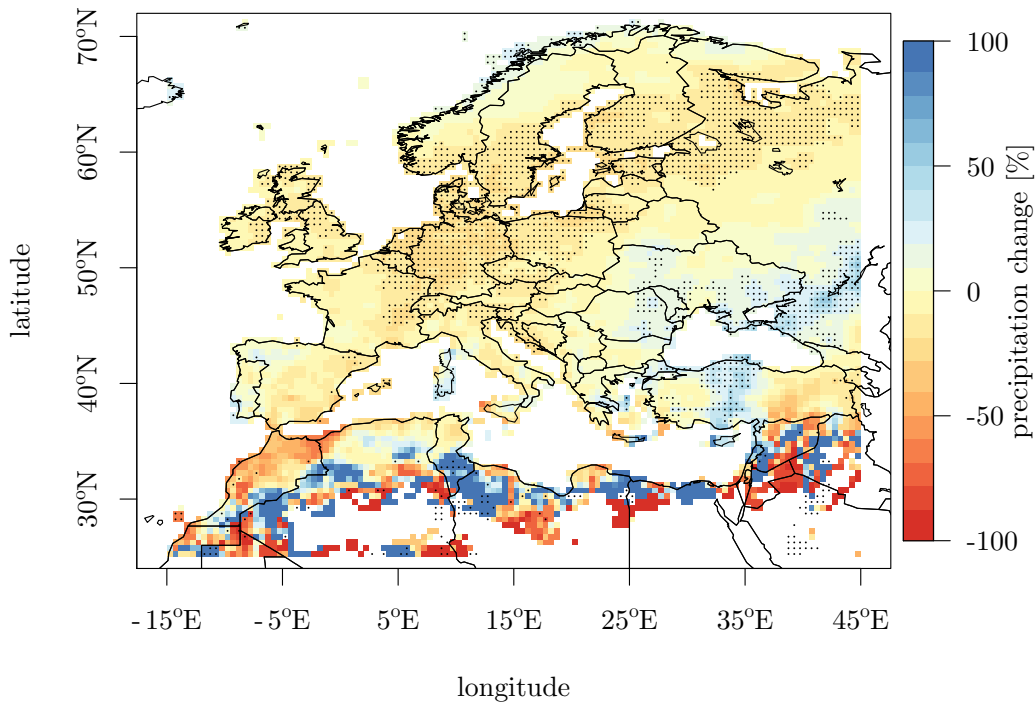
Figure 4.14. Relative difference between projected and simulated precipitation for JJA, 2031–2060 vs. 1972–2001. Black dots indicate grid points with a significant difference ($\alpha = 0.05$). The RCP 4.5 scenario is shown.

The 90th percentile of maximum temperature is shown in Fig. 4.10. There is a significant increase for WEM, NA and M. For EE and CEE, where the temperature increase is largest, the 90th percentile does not change much. However, it is evident that the 90th percentile increases with the different scenarios, which means that in the projections there will be more hot days according to the 90th percentile of the same reference period as in the past.

Figure 4.11 shows the 90th percentile of the precipitation for the six representative grid points in summer. In general the changes are very small, which means that the threshold for heavy precipitation events does not increase in the STARS projections. Only for M there is a decreasing trend.

The precipitation trend for the validation time period and the future projections is illustrated in Fig. 4.12. As can be seen the changes for all the grid points are very small. For WEM, CEE, NA, NE and EE STARS projects no clear precipitation trend at all. Note that the WATCH data set already shows no trend for CEE, NA, NE and EE during the validation time period. The trend for M is slightly negative. However, this result is statistically not significant.

Despite temperature and precipitation, the following variables were used in

this chapter: relative humidity, air pressure, shortwave radiation and wind speed. The results for these variables are not shown here. However, a short summary can be given at this point.

- The relative humidity decreases significantly in some parts of northern Europe and Spain and in central and eastern Europe. The values increase significantly in Great Britain, Sardinia and Corsica. However, the differences are below 5 %.

- The air pressure increases significantly in central and western Europe, but the differences do not even reach 2 hPa.

- The shortwave radiation increases throughout the whole simulation area by up to 8 W m^{-2}. However, the changes are not significant.

- The changes in the wind speed are very small, hardly reaching 1 m s^{-1}.

4.4. Conclusions

The application of the former STARS version (STARS_2.1) to a large simulation area such as the whole of Europe leads to a seasonality mismatch for shortwave radiation due to its way of replacing blocks during the simulation process. Thus, it is not possible to use this version on continental scale. The improvement of the resampling technique results in a new model version, STARS_2.4. This version was used to generate climate projections for Europe on the basis of the WATCH data set. A cross-validation experiment showed very good results and confirmed the applicability of the model in Europe. Hence, the new version of STARS is able to simulate large areas with a high climatological variability. The climate projections computed in this work provide a new STARS data set covering the whole of Europe and will be used in different impact studies.

The climate projections, that were computed following the RCP 2.6, RCP 4.5 and the RCP 8.5 scenarios, show a clear positive temperature trend, especially in the north-east of the simulation area. This trend increases with the different scenarios. In case of precipitation the model projects a significant increase in northern Europe, whereas it decreases in some parts of southern Europe. This pattern can be seen in all three scenarios analysed.

This general pattern of the changes in mean temperature and precipitation obtained with STARS is similar to the results of Jacob et al., 2014. Although Jacob et al., 2014 analyses the changes of mean temperature and precipitation at the end of the 21st century and in this work the focus lies on the mid-century,

it can be assumed that the general pattern is the same for the mid-century but that it has less impact.

The 90th percentile of the maximum temperature shows a significant increase for WEM, NA and M, which means that the model predicts more days that are nowadays declared as hot (i.e., the maximum temperature is higher than the 90th percentile of the maximum temperature for a certain reference period) for the future.

The number of dry days in summer increases in eastern and north-eastern parts of Europe, while the summer precipitation decreases in these parts. Together with the large temperature increase projected by STARS, these parts are very likely to suffer from heat stress in the future.

However, the validation at the representative grid points also shows that STARS fails to reproduce the number of dry periods and the 90th percentile of maximum temperature at some of the grid points. This could be improved by a new approach of modelling. Instead of modelling the whole time series at once with only one linear trend for the whole simulation period, it is possible to prescribe a trend only for the individual seasons. Then, the different seasonal developments could be accounted for.

5. Evaluation of STARS performance in climate projections

Finally, this chapter deals with the last questions formulated in the introduction: How is the quality of STARS climate projections? Are they able to account for climate change?

In the previous chapters and in several papers cited afore (for example Orlowsky et al. (2008) or Zhu et al. (2013)), the standard procedure of simulating with STARS is as follows. First, a cross-validation experiment is carried out to analyse the performance of STARS for the region of interest. Typically, the data for this experiment is taken from the past and current climate so that the STARS simulations can be validated against observations or reanalysis data. Then, the same data and a temperature trend for a future time period is used to compute climate projections with STARS. In general, the validation results are good or even very good, STARS is even able to outperform regional climate models like CCLM or REMO.

This chapter presents a further validation method which allows to test the performance of STARS in the future and to show its limits compared to dynamical regional climate models (if any). The regional climate model CCLM is used as a "real world", thus it is possible to use data from the past as well as from the future as pseudo observations based on physical laws. Comparing the STARS simulations to CCLM projections allows to evaluate the reproducibility of a changing climate. STARS depends only on the CCLM data and a temperature trend for the future time period of the forcing GCM, in short it is a closed system. Hence, this experiment could be carried out with other RCMs as well. Note however, that it has to be the same model for the past and the future.

This chapter is composed of an elaborate section about the data used and the organisation of the simulation experiment (section 5.1). Section 5.2 provides the simulation results and section 5.3 concludes this chapter.

5.1. Data and experiment set-up

5.1.1. Data

Differently to the other chapters where simulations with STARS has been computed using a reanalysis data set, this time another climate model is used as the
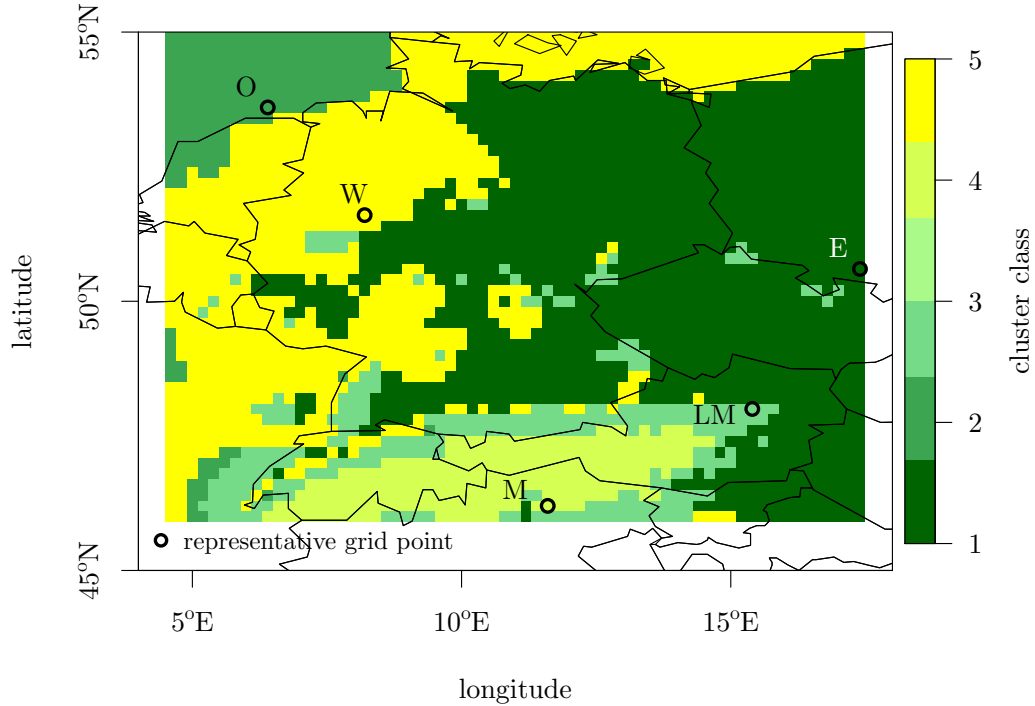
Figure 5.1. Result of the cluster analysis to determine regions with a similar climate. Coloured areas represent the different clusters, where circles mark the grid points closest to the centre of mass (representative grid points). The abbreviations of the regions are explained in the text.

input data for STARS. This model, CCLM, is already presented in section 3.2. In this study the consortial simulations that were computed with CCLM for the whole of Europe for the time period 1960-2100 are used. The spatial resolution is 0.2°. The regional model was forced by the GCM ECHAM5/MPIOM (where the latter is the Max-Planck-Institute Ocean Model). Although the temporal resolution of CCLM is very high and many different climate variables are available, the range of data in this experiment is limited to daily means of temperature, precipitation, relative humidity and shortwave radiation. In addition to the daily mean temperature, the daily minimum and maximum temperature is included.

In CCLM, three realisation runs simulate the current climate of the 20th century. They are all based on the same control run, but set off at different initialisation times. The climate of the 21st century is simulated with respect to two IPCC climate scenarios (A1B and B1, see Nakicenovic et al. (2000)) describing different assumptions regarding the development of global greenhouse gas concentrations. In total, four transient experiments were computed: each of the future scenario connects to two realisation runs of the 20th century
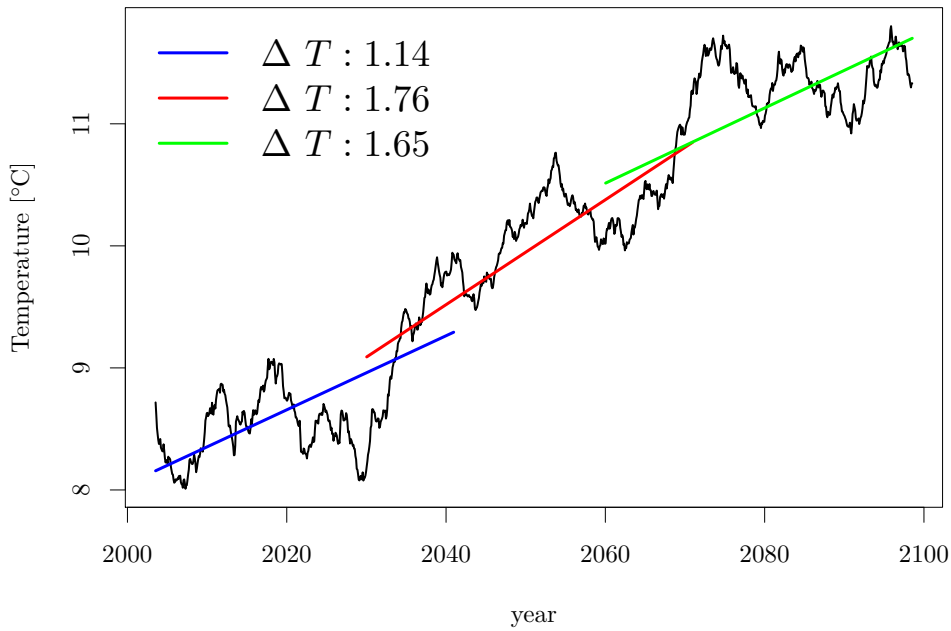
Figure 5.2. Illustration of mean temperature for the 21st century, following the A1B scenario, in the ECHAM5 model (5 year running means) and corresponding linear trends. Blue: 2001–2041, red: 2030–2070 and green: 2060–2100. The area average of the simulation area is shown.

climate. But in this chapter, only the second realisation run for the years 1960-2000 and the simulations that follow scenario A1B and are based on the second realisation run for the years 2001-2100 are used.

In this study the simulation area is limited to 46°N–56°N and 4.6°E–17.4°E with 3315 grid points. As it is impossible to compute the resampling for each of these grid points, the simulation area is split into five climatological sub-regions. These sub-regions are identified using a standard k-means clustering method based on daily temperature and precipitation that is implemented in the model. The results are shown in Figure 5.1. The sub-regions are called "Ocean (O)", "West (W)", "Mountains (M)", "Low Mountains (LM)" and "East (E)". The grid points that are closest to the centre of mass of a cluster represent this particular cluster/sub-region.

In addition to the input data, STARS also needs a linear trend of a forcing variable, in this case the daily mean temperature. The linear trend for the 20th century is derived from the CCLM simulations, whereas it is taken from ECHAM5 for the 21st century. This is done to ensure an experiment setting which is identical to those where "real" climate projections are computed. In these experiments, it is not possible to derive the future trend from the observations (which are the CCLM simulations in our case), it must be estimated

either using a GCM or different temperature increases as in section 3.4.

The mean temperature derived from ECHAM5, corresponding to the A1B scenario, is shown in Figure 5.2. It is averaged over the simulation area. As there are only 41 years of data available for the 20th century, the simulation time periods for the 21st century are limited to 41 years as well. This is why the century is split into three time periods: 2001–2041, 2030–2070 and 2060–2100. The corresponding linear temperature trend is shown as well. Note that the strongest trend ($\Delta T = 1.8°C$) can be found in the second time period, while the trend is smallest ($\Delta T = 1.1°C$) in the first 41 years.

5.1.2. Experiment set-up

To test the general viability of the method, the 20th century data of CCLM is divided in two parts: 1960–1980 and 1981–2000. The first part is used as input data for STARS to simulate the second part. These simulations are compared to the years 1981–2000 of the CCLM data. The linear temperature trend extracted from the CCLM data for the years 1981–2000 reaches from $0.6\,°C$ for M to $1.3\,°C$ for W, whereas the temperature trend for the years 1960–1980 lies between $-0.1\,°C$ for O and $0.6\,°C$ for LM. Note that the temperature trend for the years 1981–2000 is used as the forcing variable for the STARS simulations of the same years. The individual numbers for the different representative grid points can be found in Table 5.1.

The climate projections were computed for the time periods 2001–2041 and 2030–2070. It was not possible to simulate the years 2060–2100 as the model did not converge with the input requirements for this time period. Whereas the CCLM data for the 20th century was used as input for 2001–2041, the data had to be extended to 1970-2010 for the second future time period. This was done because of the high temperature trend in the middle of the 21st century (see Fig. 5.2). For the years 2001–2041 the temperature trend extracted from the ECHAM5 model reaches from $0.9\,°C$ for W and O to $1.3\,°C$ for E. This temperature trend was used to force the STARS simulations for the same years. The temperature trend in detail is shown in Table 5.1 as well. To compute simulations for the years 2030–2070, with the years 1970–2010 as input, it is necessary to add the temperature trend from the years 2010-2030 to the actual temperature trend for 2030–2070. This means, STARS does not only simulate trends from $1.5\,°C$ for M to $1.9\,°C$ for O but the warming adds up to $2.2\,°C$ and to $2.8\,°C$ in the second time period (see Table 5.1).

For each time period (1981–2000, 2001–2041 and 2030–2070) an ensemble of 100 realisations was computed.

Table 5.1. Overview over the temperature trends for the 20th and 21st century. The columns show the different time periods and in parentheses whether the trend is taken from CCLM (C) or ECHAM5 (E), the rows show the different regions (compare with Fig 5.1) and their average.

	1960–1980 (C)	1981–2000 (C)	1960–2000 (C)	1970–2010 (C)	2001–2041 (E)	2030–2070 (+2010–2030) (both E)
O	-0.1	0.9	0.3	0.8	0.9	1.9 (+0.9)
W	0.3	1.3	0.6	1.2	0.9	1.7 (+0.8)
M	0.6	0.6	0.3	1.0	1.0	1.5 (+0.7)
LM	0.6	1.1	0.3	0.8	1.2	1.6 (+0.8)
E	0.5	1.2	0.6	1.0	1.3	1.8 (+0.9)
mean	**0.4**	**1.0**	**0.4**	**1.0**	**1.0**	**1.7 (+0.8)**

5.2. Results

5.2.1. Validation

The comparison of the STARS simulations and the CCLM data for the years 1981–2000 provides good results so that it can be assumed that this method is generally applicable. The results are not shown in detail, but note that the absolute difference between the simulated long-term average of the daily mean temperature and the CCLM "observations" is mainly around 0.7 °C. It is overestimated by STARS. The long-term average of the precipitation is underestimated by STARS, the bias reaches values of up to 15 %. The mean annual cycle of all the variables is well reproduced.

5.2.2. Projections

Below, the projections for the years 2001–2041 and 2030–2070 are compared to the corresponding time periods of the CCLM data. Thus it is possible to estimate the performance of the STARS model. The differences between STARS simulations and CCLM data for temperature, precipitation, shortwave radiation and relative humidity are tested for significance using the Wilcoxon-Mann-Whitney test *(Hollander and Wolfe, 1999) with a significance level of 5 %.

Figure 5.3 shows the absolute difference between the mean temperature simulated by STARS and taken from the CCLM data for both future time periods. In the time period 2001–2041, the temperature simulated by STARS is slightly

---

*see the appendix for a short description
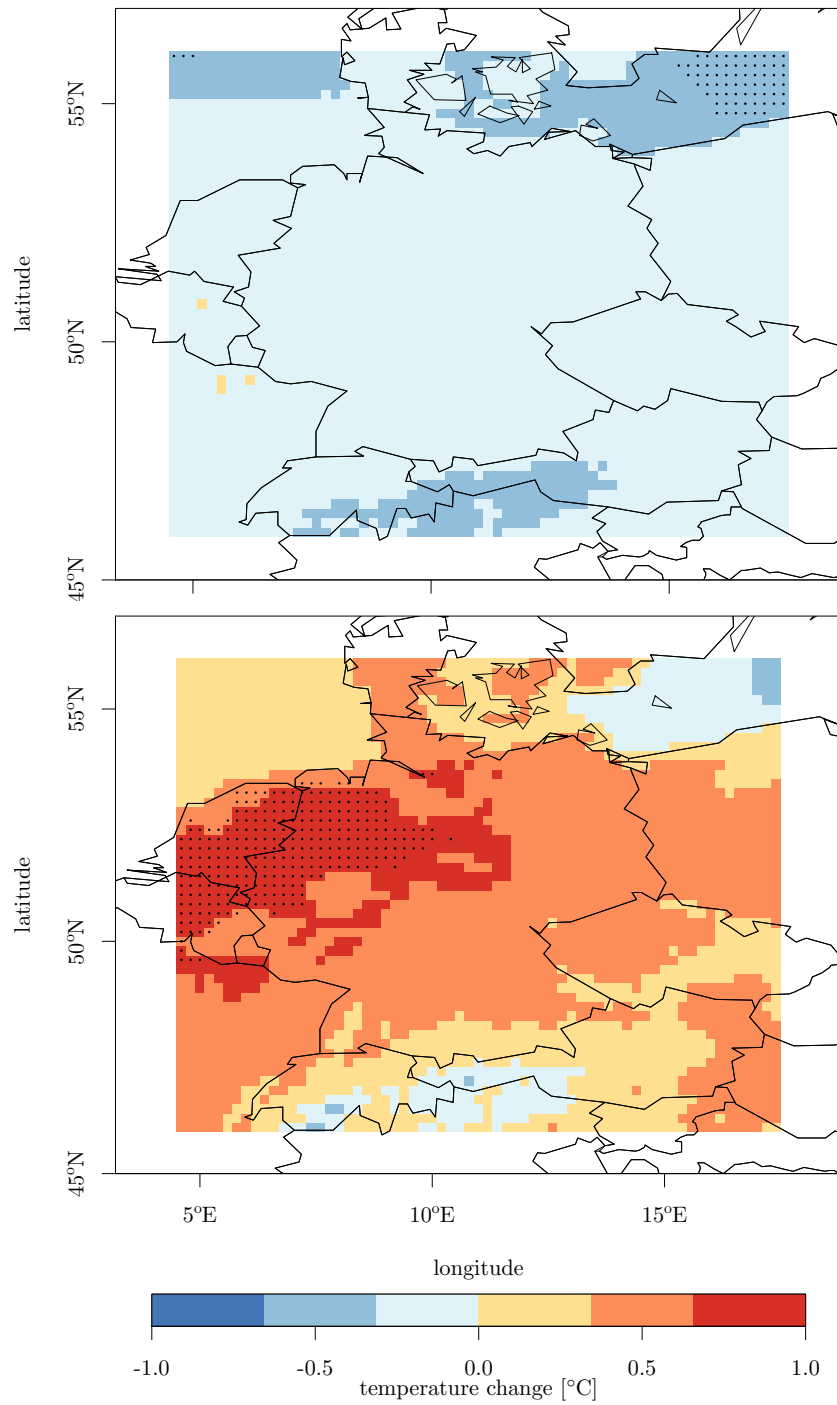
Figure 5.3. Absolute long-term difference between the mean temperature as simulated with STARS and the CCLM data. The points indicate grid points with a significant difference between the two simulations ($\alpha = 0.05$). Top: 2001–2041, Bottom: 2030–2070.

cooler than the CCLM temperature. However, the bias is significant only in the eastern Baltic Sea with values below 0.6 °C. In the second time period, STARS is mainly warmer than the CCLM data. The bias is higher, reaching values of up to 1 °C, but still it is limited to a small area in the west of the simulation area.

Figure 5.4 shows the relative difference between the STARS precipitation and the CCLM precipitation for both future time periods. For both time periods STARS is mainly drier than the CCLM data. For the first time period the difference is significant only in certain parts in the southern simulation area. However, the bias is low, hardly reaching 10 %. The bottom part of the figure shows that the number of grid points with a significant difference is considerably higher than in the upper figure. The bias is higher as well, reaching values of up to 25 %. Yet, the bias is quite low compared to other model biases (for example the precipitation bias of CCLM in Southern Africa, cf. chapter 3).

The mean annual cycles for temperature and precipitation are shown in Fig. 5.5. The representative grid point in the west of the simulation area (marked with a "W" in Fig. 5.1) was chosen for this illustration but the results are very similar for the other representative grid points: The seasonal reproduction of the CCLM temperature and precipitation is very good in the first time period. In the second time period, the performance of STARS is weaker, especially in autumn and spring, where the simulated temperature is considerably lower than the CCLM data.

The absolute difference between the simulated shortwave radiation and the CCLM data is illustrated in Fig. 5.6. For both time periods, the radiation simulated by STARS is higher than that taken from CCLM. Whereas there are still parts without a significant difference in the upper figure and the bias is low, reaching values of 5 to $10 \, \mathrm{W \, m^{-2}}$, the difference is significant throughout the whole simulation area with values between $15 \, \mathrm{W \, m^{-2}}$ and $20 \, \mathrm{W \, m^{-2}}$ in the lower figure.

The increase in the shortwave radiation simulated by STARS is accompanied with a decrease in the relative humidity. The bias of the latter is shown in Fig. 5.7. The decrease is again more profound in the second time period, where the significant difference has still small values of up to 3 % only.

However, the difference between the two time periods is even clearer when looking at the mean annual cycles of the shortwave radiation and the relative humidity, which are illustrated in Fig. 5.8. For the years 2001–2041, the shortwave radiation is well reproduced except for spring and autumn, where the simulated shortwave radiation is slightly higher than that taken from CCLM. Furthermore, STARS underestimates the relative humidity by up to 5 % in summer.
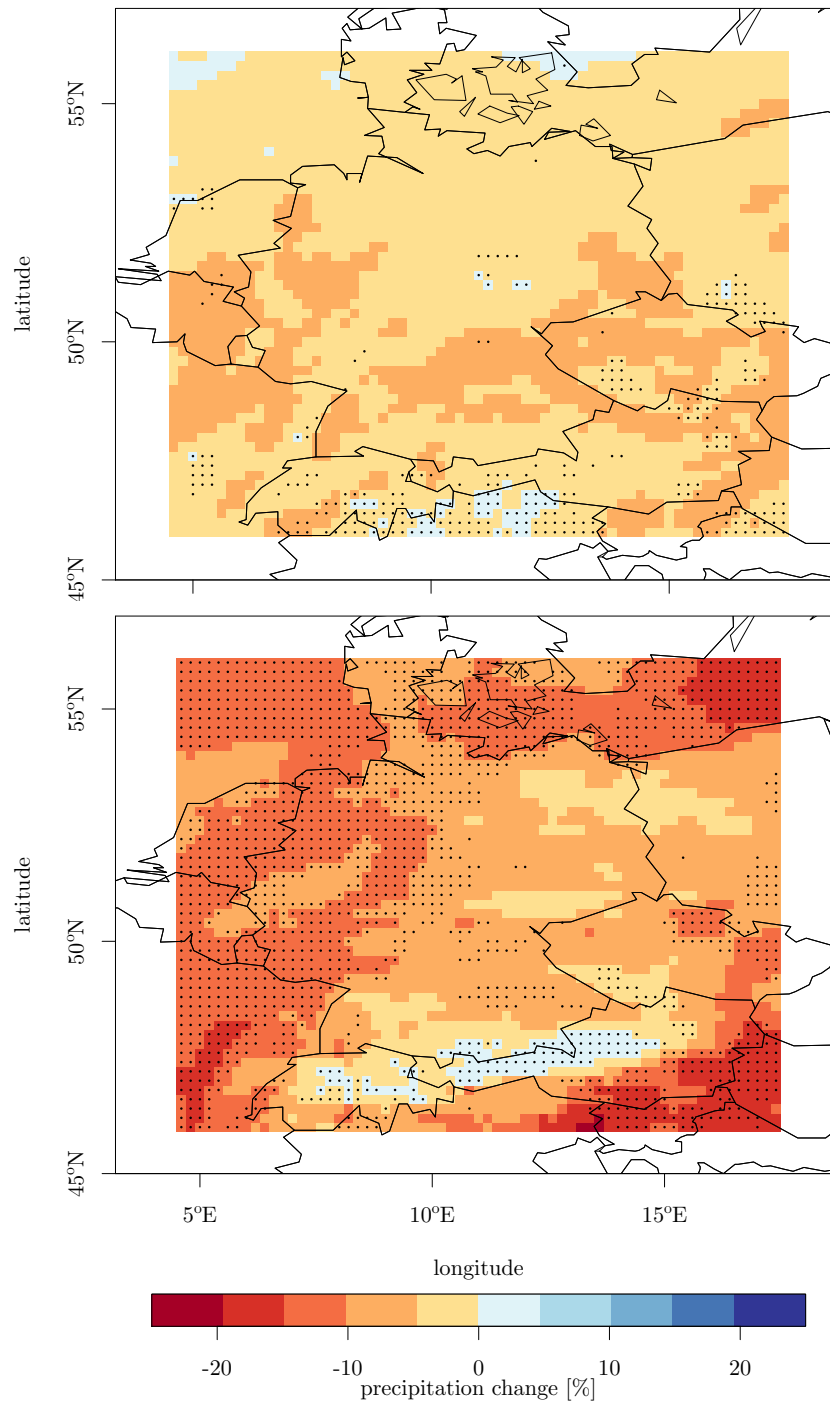
Figure 5.4. Relative long-term difference between the precipitation as simulated with STARS and the CCLM data. The points indicate grid points with a significant difference between the two simulations ($\alpha = 0.05$). Top: 2001–2041, Bottom: 2030–2070.
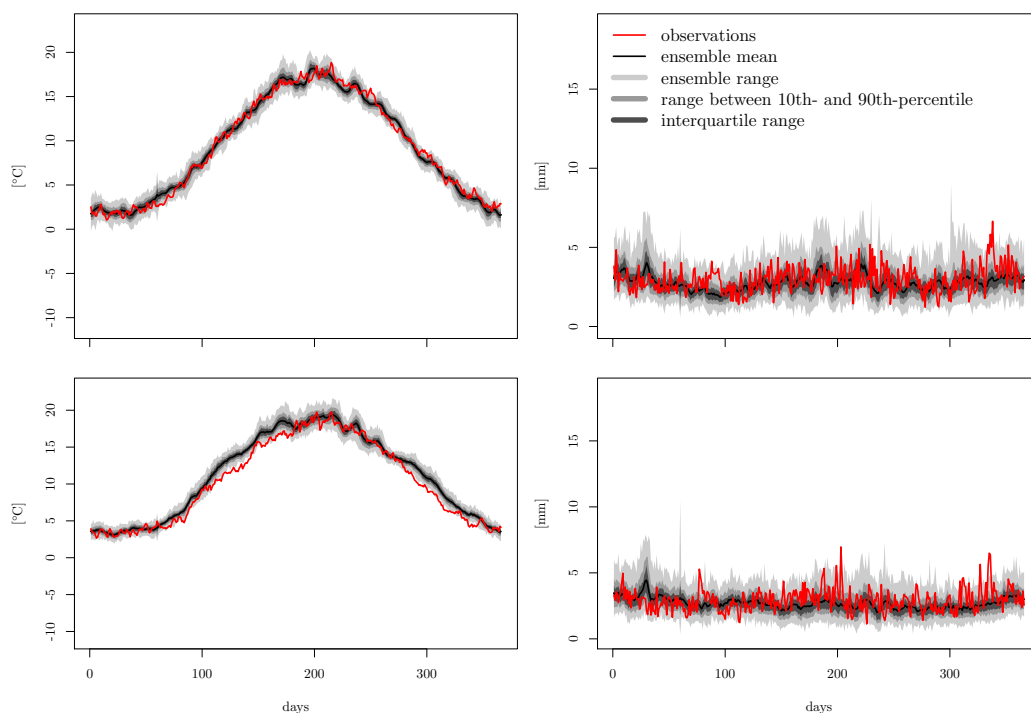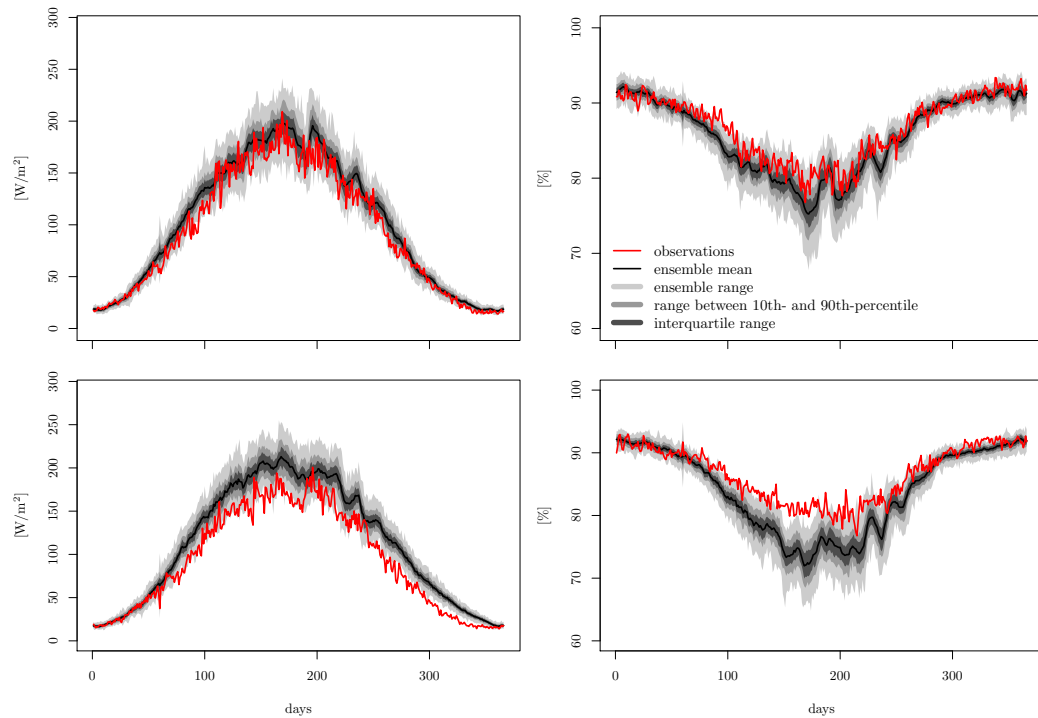
Figure 5.5. Mean annual cycles of temperature (left) and precipitation (right) for the representative grid point in the west of the simulation area (marked with "W" in Fig. 5.1). The CCLM data is shown in red, while the range of all STARS realisations is shown in grey and the mean over all STARS realisations is shown in black. Top: 2001–2041, Bottom: 2030–2070.

In contrast, STARS is not able to reproduce both, the shortwave radiation and the relative humidity, correctly for the years 2030–2070. It overestimates the shortwave radiation throughout the year but especially in summer, when the difference between the CCLM data and the mean over all STARS realisation reaches values of up to $50\,\mathrm{W\,m^{-2}}$. The difference between the STARS-humidity and the CCLM-humidity is low in the winter month, while STARS underestimates it for the rest of the year. Again, the difference is high, reaching values of up to $10\,\%$.

5.3. Conclusions

In summary, STARS is able to reproduce all variables analysed very well for the validation time period. In case of projections, the results for the mean temperature (for the minimum and maximum temperature as well, although the results are not shown) and for the precipitation are very good in the first
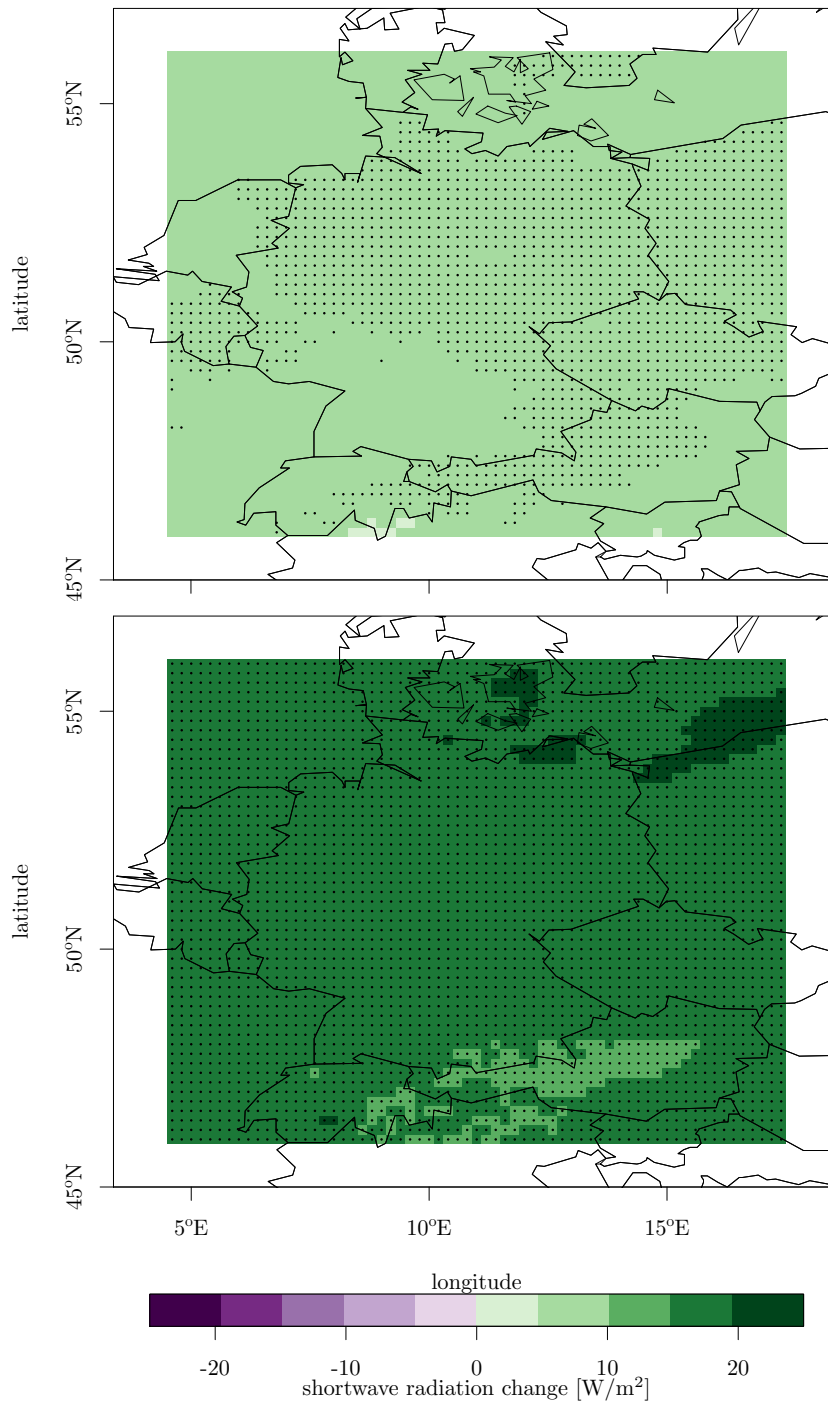
Figure 5.6. Absolute long-term difference between the shortwave radiation as simulated with STARS and the CCLM data. The points indicate grid points with a significant difference between the two simulations ($\alpha = 0.05$). Top: 2001–2041, Bottom: 2030–2070.
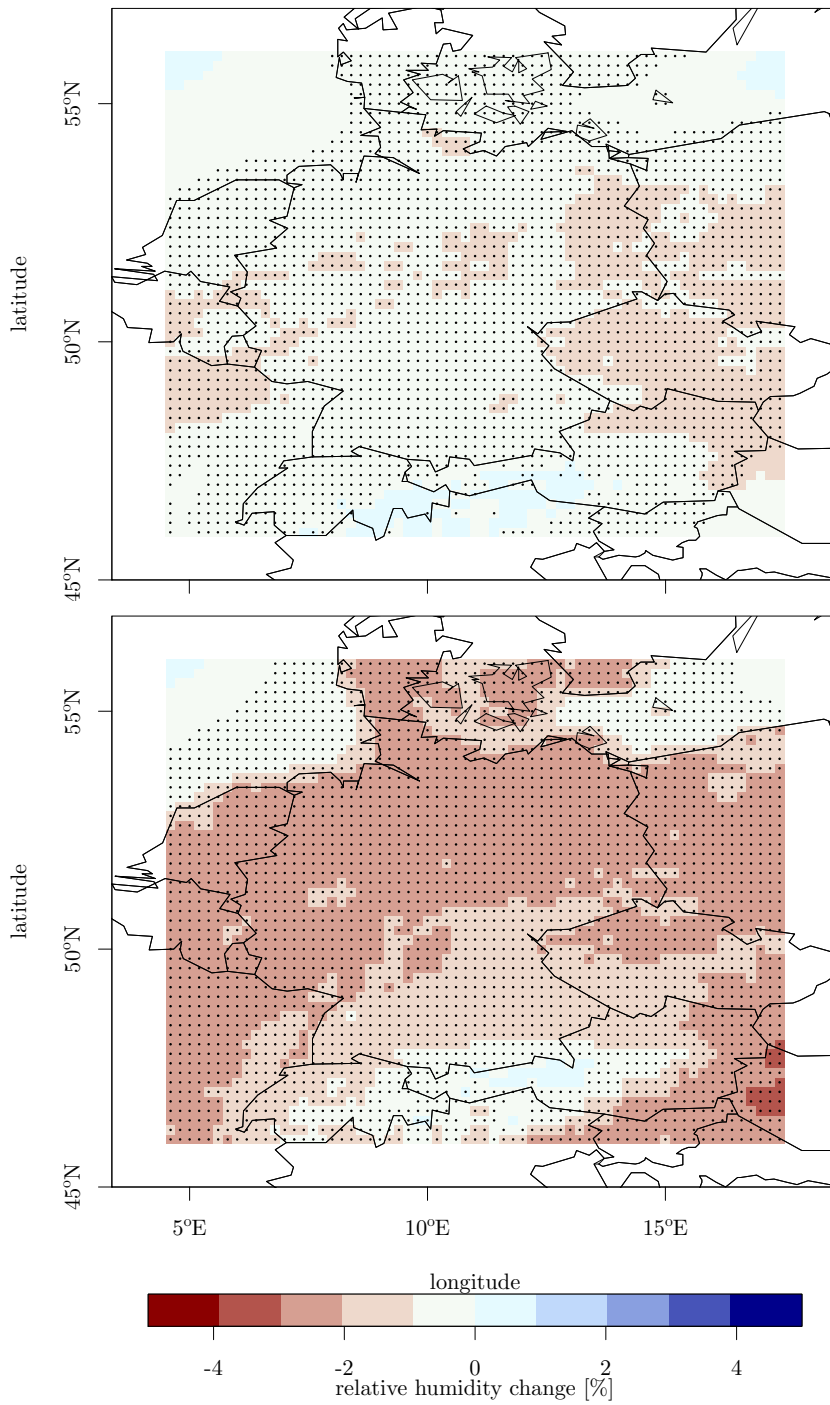
Figure 5.7. Absolute long-term difference between the relative humidity as simulated with STARS and the CCLM data. The points indicate grid points with a significant difference between the two simulations ($\alpha = 0.05$). Top: 2001–2041, Bottom: 2030–2070.

Figure 5.8. Mean annual cycles of shortwave radiation (left) and relative humidity (right) for the representative grid point in the west of the simulation area (marked with "W" in Fig. 5.1). The CCLM data is shown in red, while the range of all STARS realisations is shown in grey and the mean over all STARS realisations is shown in black. Top: 2001–2041, Bottom: 2030–2070.

time period (2001–2041) and quite good in the second time period (2030–2070). Note however, that the differences between STARS and CCLM for temperature and precipitation are higher in the validation time period than for the first 41 years of the 21st century. For the years 2030–2070, the shortwave radiation and relative humidity simulated by STARS are significantly different from the CCLM data, the difference is particularly large when looking at the annual cycles.

These results show that STARS is able to reproduce well the time series of different variables even for future time periods. However, it makes a difference whether the simulation period lies in the near or in the far future. For 2030–2070, STARS is able to reproduce temperature and precipitation quite well, but the model fails for shortwave radiation and relative humidity. However, these variables play a significant role in climate impact research, where STARS projections are often used. The reason lies of course in the higher temperature trend for the second future time period. To achieve this trend STARS is forced

to choose blocks with higher temperatures. As can be seen in Fig. 5.5, this is mainly done in spring, summer and autumn. At these times, the warmer blocks typically involve lower cloudiness and hence higher shortwave radiation and mostly lower relative humidity. This very characteristic can be observed in the results for 2030–2070.

The better results for the years 2001-2041 than the ones for the years 1981–2000 can be explained by the different temperature trends in the input data. For the validation time period, only the first 21 years (1960–1980) of the input data is used. This time period has a very low temperature trend while the trend of the validation time period is similar to the one in the first future time period (see values in Table 5.1). As it is more difficult to simulate a strong temperature trend using input data with a low trend, the results for the validation time period are not as good as for the years 2001–2041. For this reason, the input data for the second future time period is shifted by ten years to 1970–2010. However, the method does not help to improve the results for the second future time period.

The results in this chapter underline the "internal variability conservation" that was already formulated in Orlowsky et al. (2010). It says that only if the temperature anomalies of the input data and the simulated time series can be seen as originating from the same distribution, the variability of the input data is large enough to generate time series with a given temperature trend without a statistically visible reduction of variability. In several tests Orlowsky found out that this criterion is only fulfilled when the warming in the simulated period continues with the same strength as in the input period. This rule should be followed for future STARS simulations, otherwise the model is very likely to provide incorrect climate projections. However, in case of very demanding input requirements the model is not able to converge and find suitable time series so that it is not possible to create projections under this circumstances. In this chapter this was the case for the time period 2060–2100.

The good reproduction of the years 2001–2041 leads to the conclusion that STARS is able to provide good climate projections for the near future. This makes STARS an alternative to dynamical regional climate models for a time span of 40 or 50 years in the future. In contrast to STARS these models require high computational costs, whereas STARS can produce a large ensemble in a short time. Thus, the resources to provide climate data for several climate impact applications could be reduced without decreasing the quality of the resulting simulations for the near future.

6. Overall conclusions and Outlook

6.1. Conclusions

In this thesis the statistical model STARS is used in three different regions with different climate conditions to test the performance of the model. The results are briefly summarised in the following. Afterwards, this section gives answers to the questions which have been formulated in the introduction of this work and which form the basis of this work.

In chapter 3, STARS is used with gridded data (WATCH data set) in southern Africa for the first time. This is also the first time that STARS simulates such a large area. A validation experiment yields excellent results despite the high climatological variability in this region. It even outperforms the regional climate model CCLM (Böhm et al., 2006).

The subsequent climate projections for the years 2011–2060 are computed based on four different mean temperature trends. They agree on a significant increase for the mean, maximum and minimum temperature. In case of precipitation, the model projects a decrease of over 25 % in the rainy season (DJF) for the simulation area. This decrease is especially high in the inland of the simulation area, where the annual precipitation is very low anyway. However, it is also significant in Lesotho where the main runoff for the most important river in southern Africa, the Orange River, is generated.

In chapter 4, the former STARS model version revealed a seasonality mismatch for shortwave radiation due to its way of replacing blocks during the simulation process. When using a demanding simulation setting (e.g. a large climatologically differing region with many climatological sub-regions such as a continent, a high prescribed trend or both), the criterion to use blocks only once outweighs the criterion that the blocks have to lie within a defined date window. However, the second criterion is much more important for a successful simulation of the annual cycle. Earlier simulations with STARS do not face this problem so rigorously since replacements within the date window are much easier to achieve for a less demanding setting.

The problem was solved successfully, resulting in a new version of the model. Now, the blocks must lie in a date window of at most 40 days but they can be used up to three times. The simulation results for Europe computed with the new model version show a realistic annual cycle of the shortwave radiation in contrast to the simulations computed with the former version.

6. Overall conclusions and Outlook

The new version of the STARS model is used to generate climate projections for Europe on the basis of the WATCH data set (Weedon et al., 2011). A cross-validation experiment shows very good results. The climate projections, that were computed following the RCP 2.6, RCP 4.5 and the RCP 8.5 scenarios (Moss et al., 2010), show a clear positive temperature trend, especially in the north-east of Europe. This trend increases with the emissions of the different scenarios. In case of precipitation the model projects a significant increase in northern Europe, whereas precipitation decreases in some parts of southern Europe. This pattern can be seen in all three scenarios analysed. The climate projections provide a new STARS data set covering the whole of Europe and will be used in different impact studies.

The "internal variability conservation" (Orlowsky et al., 2010) says that only if the temperature anomalies of the input data and the simulated time series can be seen as originating from the same distribution, the variability of the input data is large enough to generate time series with a given temperature trend without a statistically visible reduction of variability. In several tests Orlowsky found out that this criterion is only fulfilled when the warming in the simulated period continues with the same strength as in the input period. To test this theoretical assumption in a real case application and to analyse the performance of STARS in future projections, a new validation method is used in chapter 5: the data of the regional climate model CCLM serves as model input for STARS. Thus, it is possible to compare the simulation results of STARS with the CCLM data not only for the past but also for the future.

The results for the mean temperature and for the precipitation are very good in the first time period (2000–2040) and quite good in the second time period (2030–2070). For the years 2030–2070, the short wave radiation and relative humidity simulated by STARS is significantly different from the CCLM data, the difference is particularly large when looking at the annual cycles.

These results show that STARS is able to well reproduce the time series of different variables even for future time periods. However, it makes a difference how far the simulation period lies in the future. For 2030–2070, STARS is not able to reproduce shortwave radiation and relative humidity. However, these variables play a significant role in climate impact research, where STARS projections are often used. As indicated by Orlowsky et al. (2010), the reason for failure lies in the higher temperature trend for the second future time period.

All these results allow the following answers to the questions formulated in the introduction:

1. Is it possible to run STARS with gridded meteorological data?

   The good performance of STARS in chapter 3 shows that it is possible to use not only station data but also gridded data as model input. Additionally, the application for a large simulation area has been successfully tested.

2. Can the simulation area be extended? If yes, what are the challenges?

   The former model version of STARS (STARS_2.1) is not able to compute simulations on a continental scale. It provides unrealistic annual cycles of temperature and especially short wave radiation. A new version (STARS_2.4) has been developed to solve this problem. This version provides good results for the whole of Europe and hence is able to simulate large areas on continental scale with a high climatological variability.

3. How is the quality of STARS climate projections? Are they able to account for climate change?

   The comparison of the future climate projections computed with STARS and the CCLM data shows that STARS is able to reproduce temperature and precipitation quite well for the years 2030–2070. However, the STARS model fails for shortwave radiation and relative humidity in this time period. For 2000–2040, STARS is able to reproduce all variables analysed. This makes STARS an alternative to dynamical regional climate models for a time span of 40 or 50 years into the future. Thus, the resources to provide climate data for several climate impact applications can be reduced without decreasing the quality of the resulting simulations for the near future.

6.2. Outlook

Regarding further development, there are two points in the model code which are currently considered most favourable to have a closer look at and possibly improve. The first one deals with future extreme values in the simulated variables. The corresponding part of the model deals with the physical consistency of the different simulated variables. Here, the basic assumption of the STARS model is that any simulated set of variables for a specific time step $\vec{s}_t$ must lie closer to the closest set of observations $\vec{o}_x$ than this set to its closest set of observations $\vec{o}_{y \neq x}$, i.e.

6. Overall conclusions and Outlook

$$|\vec{s}_t - \vec{o}_x| \leq |\vec{o}_x - \vec{o}_{y \neq x}| \qquad (6.1)$$

where $||$ is some reasonable distance metric between sets of observations or simulations (e.g. the Euclidean distance). Currently, this is solved by the use of simulated sets of variables which already have been observed, i.e.

$$|\vec{s}_t - \vec{o}_x| = 0 \quad \Rightarrow \quad |\vec{s}_t - \vec{o}_x| \leq |\vec{o}_x - \vec{o}_{y \neq x}| \qquad (6.2)$$

While this guarantees the physical consistency of the simulated variables, it also hinders the simulated values to be larger than the already observed ones. Thus, no new extreme values can be simulated for single time steps. However, for longer time spans like consecutive days of precipitation or draughts, the STARS model is already able to generate new extremes. For many climate impact models the aspect of physical consistency is usually very important. Nevertheless, since there is a general agreement that a warmer climate will also yield heavier precipitation events an improvement of STARS in this point is desirable.

The second point deals with the distribution of the ensemble around the prescribed trend. Currently, the prescribed trend ΔT must be met with a given tolerance ϵ. The ensemble members tend to be uniformly distributed between the lower and upper tolerance level if the trend is relatively small, while with a large trend the ensemble members are usually closer to the lower tolerance level, as the upper parts are seldom reached. In any case, the distribution of the ensemble members around the prescribed trend does not follow any given distribution, but is a result of the fitting prescription. It would be desirable to prescribe not only a given trend but also the distribution of the ensemble members around this trend. This would enable STARS to follow the spread of temperature changes of a whole GCM ensemble instead of the ensemble mean only. This means that the distribution of the simulated ensemble must lie within a certain tolerance ϵ_F of the prescribed distribution:

$$\int |F_{\Delta T}(x) - F_B(x)| \, dx \leq \epsilon_F \qquad (6.3)$$

where $F_{\Delta T}$ and F_B are the cumulative distribution functions of the pre-scribed trends and of the regression parameters of the simulated ensemble members, respectively.

Technically, this could for instance be done by simulating a very large ensemble with changing prescribed trends and drawing from the final distribution those simulations that minimise the left hand side of equation 6.3.

At the moment STARS is typically used as a stand alone tool for climate modelling. As already discussed above, the applicability of STARS is limited

to the near future. The simulation of the far future is mainly carried out by dynamical RCMs. Due to high computational costs the number of simulation runs for dynamical RCMs does not exceed a few runs per scenario. To combine the advantage of STARS (producing a large ensemble in a short time) and a dynamical RCM (simulating on a physical basis) a new approach could be to run STARS with RCM climate projections. Thus, a single simulation run could be enlarged to an ensemble without exhausting the computational resources. Of course, it is also possible to use already existing RCM simulations for the 20th century as input for the STARS model to provide climate projections for the near future. These RCM simulations could for instance be driven by reanalysis data.

A. Statistical tests

A.1. Wilcoxon-Mann-Whitney test

The Wilcoxon-Mann-Whitney test (Hollander and Wolfe, 1999), also called Mann–Whitney U test, is a nonparametric test. It is used to test whether two independent samples are drawn from the same or identical distributions based on the rank sums of the two samples.

To compute the test statistic, both samples are combined and ranked in ascending order. Equal values receive the average of their ranks. Then the sum of ranks for each sample is calculated. The test statistic U is defined as

$$U_1 = R_2 - \frac{n_1(n_1 + 1)}{2},$$

where n_1 is the sample size for sample 1 and R_2 is the sum of ranks in sample 2. It does not make any difference which of the two samples is considered sample 1. An equivalent notation is

$$U_2 = R_1 - \frac{n_2(n_2 + 1)}{2}.$$

The smaller value of U_1 and U_2 is used to consult significance tables. The sum of these two values is given by

$$U_1 + U_2 = n_1 n_2.$$

For large samples, U is approximately normal distributed. In this case the standardised variable can be written as

$$z = \frac{U - m_U}{\sigma_U},$$

where m_U is the mean and σ_U is the standard deviation of U. The significance of z can be checked in tables for the normal distribution.

A.2. F-test

The purpose of the F-test (Snedecor and Cochran, 1989) is to check if the variances of two samples are equal. There are two versions of this test: a one-tailed or a two-tailed. While the two-sided test tests against the alternative hypothesis that the variances are not equal, the one-sided test can only be used in one direction. This means the variance of the first sample is either greater than or less than (but not both) the variance of the second sample.

The hypotheses are then formulated as

$$
\begin{aligned}
H_0 : \ & \sigma_1{}^2 = \sigma_2{}^2 \\
H_a : \ & \sigma_1{}^2 \neq \sigma_2{}^2 \ \text{ for a two-tailed test} \\
& \sigma_1{}^2 < \sigma_2{}^2 \ \text{ for a lower one-tailed test} \\
& \sigma_1{}^2 > \sigma_2{}^2 \ \text{ for an upper one-tailed test,}
\end{aligned}
$$

where H_0 is the null hypothesis, H_a is the alternative hypothesis and $\sigma_1{}^2$ and $\sigma_2{}^2$ are the sample variances.

The test statistic is defined as

$$
\mathbf{F} = \frac{\sigma_1{}^2}{\sigma_1{}^2}.
$$

The more this ratio deviates from 1, the stronger the evidence for unequal sample variances.

The null hypothesis is rejected if

$$
\left.
\begin{aligned}
& \mathbf{F} < \mathbf{F}_{1-\alpha/2, N_1-1, N_2-1} \\
& \text{or} \\
& \mathbf{F} > \mathbf{F}_{\alpha/2, N_1-1, N_2-1}
\end{aligned}
\right\} \quad \text{for a two-tailed test}
$$

$$
\begin{aligned}
& \mathbf{F} > \mathbf{F}_{1-\alpha, N_1-1, N_2-1} \ \text{ for a lower one-tailed test} \\
& \mathbf{F} > \mathbf{F}_{\alpha, N_1-1, N_2-1} \quad \text{ for an upper one-tailed test,}
\end{aligned}
$$

where $\mathbf{F}_{\alpha, N_1-1, N_2-1}$ is the critical value of the F distribution with $N_1 - 1$ and $N_2 - 1$ degrees of freedom and a significance level α.

A.3. Shapiro-Wilk Test

The Shapiro-Wilk test (Shapiro and Wilk, 1965) is used to find out whether a sample $x_1, ..., x_n$ comes from a normally distributed population. Thus, the

null hypothesis is that the population is normally distributed.

The test statistic is defined as

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_{(i)}\right)^2}{\sum_{i=1}^{n} (x_i - \overline{x})^2},$$

where $x_{(i)}$ is the i-th order statistic (i.e. the i-th smallest number in the sample) and $\overline{x}$ is the sample mean. The a_i are constants generated from the means, variances and covariances of the order statistics of a sample of size n from a normal distribution.

The null hypothesis is rejected if W is below a predetermined threshold.

A.4. Levene's test

This test (Levene, 1960) is used to test the homogeneity of variances for a variable calculated for two or more groups. This means, it tests the null hypothesis that the variances of the populations from which different samples are drawn are equal. The Levene's test is able to handle small deviations from normality.

Given a variable Y with a sample of size N divided into k subgroups, where N_i is the sample size of the ith subgroup, the test statistic W is defined as

$$W = \frac{(N-k) \sum_{i=1}^{k} N_i (\overline{Z_{i.}} - \overline{Z_{..}})^2}{(k-1) \sum_{i=1}^{k} \sum_{j=1}^{N_i} (Z_{ij} - \overline{Z_{i.}})^2}.$$

$\overline{Z_{i.}}$ are the group means of the Z_{ij} and $\overline{Z_{..}}$ is the overall mean of the Z_{ij}. Z_{ij} can have one of the following definitions.

$$Z_{ij} = \begin{cases} |Y_{ij} - \overline{Y_{i.}}|, & \overline{Y_{i.}}\text{is the mean of the i-th group} \\ |Y_{ij} - \widetilde{Y_{i.}}|, & \widetilde{Y_{i.}}\text{is the median of the i-th group} \end{cases}$$

The use of the median instead of the mean is an extension of the original formulation of the Levene's test. The choice of a definition for Z_{ij} is determined by the underlying distribution.

The test rejects the null hypothesis if $W > F_{\alpha, k-1, N-k}$, where F is a quantile of the F-test distribution with $k-1$ and $N-k$ its degrees of freedom and α is the chosen level of significance.

Bibliography

Böhm, U., M. Kücken, W. Ahrens, A. Block, D. Hauffe, K. Keuler, B. Rockel and A. Will (2006). 'CLM - The Climate Version of LM: Brief Description and Long-Term Applications'. In: *COSMO Newsletter* 6, pp. 225–235 (cit. on p. 83).

Efron, B. (1979). 'Bootstrap Methods: Another Look at the Jackknife'. In: *The Annals of Statistics* 7.1, pp. 1–26 (cit. on p. 17).

Efron, B. and R. J. Tibshirani (1994). *An Introduction to the Bootstrap, Monographs on Statistics and Applied Probability, Vol. 57.* Chapman and Hall/CRC (cit. on p. 17).

Feldhoff, J. H., S. Lange, J. Volkholz, J. F. Donges, J. Kurths and F.-W. Gerstengarbe (2014). 'Complex networks for climate model evaluation with application to statistical versus dynamical modeling of South American climate'. In: *Climate Dynamics*, pp. 1–15 (cit. on p. 17).

Gädeke, A., H. Hölzel, H. Koch, I. Pohle and U. Grünewald (2013). 'Analysis of uncertainties in the hydrological response of a model-based climate change impact assessment in a subcatchment of the Spree River, Germany'. In: *Hydrological Processes* 28 (12), pp. 3978–3998 (cit. on p. 14).

Hartigan, J. A. (1975). *Clustering Algorithms.* New York, NY, USA: John Wiley & Sons, Inc. (cit. on p. 46).

Hattermann, F. F., H. Gömann, T. Conradt, M. Kaltofen, P. Kreins and F. Wechsung (2007). 'Impacts of global change on water-related sectors and society in a trans-boundary central European river basin – Part 1: project framework and impacts on agriculture'. In: *Advances in Geosciences* 11.11, pp. 85–92 (cit. on p. 14).

Hattermann, F. F., J. Post, V. Krysanova, T. Conradt and F. Wechsung (2008). 'Assessment of Water Availability in a Central-European River Basin (Elbe) Under Climate Change'. In: *Advances in Climate Change Research* 4, pp. 42–50 (cit. on p. 14).

Hollander, M. and D. A. Wolfe (1999). *Nonparametric statistical methods.* NY John Wiley & Sons (cit. on pp. 30, 36, 38, 57, 73, 89).

Bibliography

Huang, S., V. Krysanova, H. Österle and F. F. Hattermann (2010). 'Simulation of spatiotemporal dynamics of water fluxes in Germany under climate change'. In: *Hydrological Processes* 24.23, pp. 3289–3306 (cit. on p. 14).

IPCC (2007). *Climate Change 2007: Synthesis Report. Contribution of Working Groups I, II and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, Pachauri, R.K , A. Reisinger (eds.)]* Geneva, Switzerland: IPCC (cit. on p. 13).

– (2014). *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]* Geneva, Switzerland: IPCC (cit. on p. 13).

Jacob, D., J. Petersen, B. Eggert, A. Alias, O. B. Christensen, L. M. Bouwer, A. Braun, A. Colette, M. Déqué, G. Georgievski, E. Georgopoulou, A. Gobiet, L. Menut, G. Nikulin, A. Haensler, N. Hempelmann, C. Jones, K. Keuler, S. Kovats, N. Kröner, S. Kotlarski, A. Kriegsmann, E. Martin, E. Meijgaard, C. Moseley, S. Pfeifer, S. Preuschmann, C. Radermacher, K. Radtke, D. Rechid, M. Rounsevell, P. Samuelsson, S. Somot, J.-F. Soussana, C. Teichmann, R. Valentini, R. Vautard, B. Weber and P. Yiou (2014). 'EURO-CORDEX: new high-resolution climate change projections for European impact research'. In: *Regional Environmental Change* 14.2, pp. 563–578 (cit. on p. 66).

Koch, H., S. Vögele, M. Kaltofen and U. Grünewald (2012). 'Trends in water demand and water availability for power plants – scenario analyses for the German capital Berlin'. In: *Climatic Change* 110.3-4, pp. 879–899 (cit. on p. 14).

Kotlarski, S., A. Block, U. Böhm, D. Jacob, K. Keuler, R. Knoche, D. Rechid and A. Walter (2005). 'Regional climate model simulations as input for hydrological applications: evaluation of uncertainties'. In: *Advances in Geosciences* 5, pp. 119–125 (cit. on p. 31).

Levene, H. (1960). *Contributions to probability and statistics: Essays in honor of Harold Hotelling.* Ed. by I. O. et al. Stanford University Press (cit. on pp. 51, 91).

Liersch, S., J. Cools, B. Kone, H. Koch, M. Diallo, J. Reinhardt, S. Fournet, V. Aich and F. F. Hattermann (2012). 'Vulnerability of rice production in the Inner Niger Delta to water resources management under climate variability and change'. In: *Environmental Science & Policy* 34, pp. 18–33 (cit. on p. 14).

Lutz, J. and F.-W. Gerstengarbe (2014). 'Improving seasonal matching in the STARS model by adaptation of the resampling technique'. In: *Theoretical and Applied Climatology* (cit. on pp. 17, 46).

Lutz, J., J. Volkholz and F.-W. Gerstengarbe (2013). 'Climate projections for southern Africa using complementary methods'. In: *International Journal of Climate Change Strategies and Management* 5.2, pp. 130–151 (cit. on pp. 14, 17, 23).

Moss, R. H., J. A. Edmonds, K. A. Hibbard, M. R. Manning, S. K. Rose, D. P. van Vuuren, T. R. Carter, S. Emori, M. Kainuma, T. Kram et al. (2010). 'The next generation of scenarios for climate change research and assessment'. In: *Nature* 463.7282, pp. 747–756 (cit. on pp. 47, 84).

Nakicenovic, N., J. Alcamo, G. Davis, B. de Vries, J. Fenhann, S. Gaffin, K. Gregory, A. Grubler, T. Y. Jung, T. Kram et al. (2000). *Special report on emissions scenarios: a special report of Working Group III of the Intergovernmental Panel on Climate Change*. Tech. rep. Pacific Northwest National Laboratory, Richland, WA (US), Environmental Molecular Sciences Laboratory (US) (cit. on pp. 29, 30, 70).

Orlowsky, B., O. Bothe, K. Fraedrich, F.-W. Gerstengarbe and X. Zhu (2010). 'Future Climates from Bias-Bootstrapped Weather Analogs: An Application'. In: *Journal of Climate* 23.13, pp. 3509–3524 (cit. on pp. 15, 17, 25, 28, 81, 84).

Orlowsky, B., F.-W. Gerstengarbe and P. C. Werner (2008). 'A resampling scheme for regional climate simulations and its performance compared to a dynamical RCM'. In: *Theoretical and Applied Climatology* 92 (3), pp. 209–223 (cit. on pp. 14, 15, 17, 20, 25, 69).

Orlowsky, B. (2007). 'Setzkasten Vergangenheit – ein kombinatorischer Ansatz für regionale Klimasimulationen'. PhD thesis. Universität Hamburg (cit. on p. 17).

Orlowsky, B. and K. Fraedrich (2009). 'Upscaling European surface temperatures to North Atlantic circulation-pattern statistics'. In: *International Journal of Climatology* 29.6, pp. 839–849 (cit. on p. 17).

REACCT (2012). *http://www.reacctanzania.com/* (cit. on p. 30).

Rockel, B. and B. Geyer (2008). 'The performance of the regional climate model CLM in different climate regions, based on the example of precipitation'. In: *Meteorologische Zeitschrift* 17, pp. 487–498 (cit. on p. 25).

Roeckner, E., R. Brokopf, M. Esch, M. Giorgetta, S. Hagemann, L. Kornblueh, E. Manzini, U. Schlese and U. Schulzweida (2006). 'Sensitivity of simulated

climate to horizontal and vertical resolution in the ECHAM5 atmosphere model'. In: *Journal of Climate* 19, pp. 3771–3791 (cit. on p. 30).

Shapiro, S. S. and M. B. Wilk (1965). 'An analysis of variance test for normality (complete samples)'. In: *Biometrika* 52.3 and 4, pp. 591–611 (cit. on pp. 51, 90).

Snedecor, G. W. and W. G. Cochran (1989). *Statistical Methods, Eighth edition.* Iowa State University Press (cit. on pp. 51, 90).

Stoner, A. M. K., K. Hayhoe, X. Yang and D. J. Wuebbles (2013). 'An asynchronous regional regression model for statistical downscaling of daily climate variables'. In: *International Journal of Climatology* 33.11, pp. 2473–2494 (cit. on p. 13).

Suckow, F., P. Lasch-Born, F.-W. Gerstengarbe, P. C. Werner and C. P. O. Reyer (2015). 'Climate change impacts on a pine stand in Central Siberia'. In: *Regional Environmental Change*, pp. 1–13 (cit. on p. 14).

Taylor, K. E. (2001). 'Summarizing multiple aspects of model performance in a single diagram'. In: *Journal of Geophysical Research* 106.D7, pp. 7183–7192 (cit. on p. 32).

Uppala, S. M., P. W. Kållberg, A. J. Simmons, U. Andrae, V. D. C. Bechtold, M. Fiorino, J. K. Gibson, J. Haseler, A. Hernandez, G. A. Kelly, X. Li, K. Onogi, S. Saarinen, N. Sokka, R. P. Allan, E. Andersson, K. Arpe, M. A. Balmaseda, A. C. M. Beljaars, L. van de Berg, J. Bidlot, N. Bormann, S. Caires, F. Chevallier, A. Dethof, M. Dragosavac, M. Fisher, M. Fuentes, S. Hagemann, E. Hólm, B. J. Hoskins, L. Isaksen, P. A. E. M. Janssen, R. Jenne, A. P. Mcnally, J.-F. Mahfouf, J.-J. Morcrette, N. A. Rayner, R. W. Saunders, P. Simon, A. Sterl, K. E. Trenberth, A. Untch, D. Vasiljevic, P. Viterbo and J. Woollen (2005). 'The ERA-40 re-analysis'. In: *Quarterly Journal of the Royal Meteorological Society* 131.612, pp. 2961–3012 (cit. on p. 21).

Weedon, G. P., S. Gomes, P. Viterbo, W. J. Shuttleworth, E. Blyth, H. Österle, J. C. Adam, N. Bellouin, O. Boucher and M. Best (2011). 'Creation of the WATCH Forcing Data and Its Use to Assess Global and Regional Reference Crop Evaporation over Land during the Twentieth Century'. In: *Journal of Hydrometeorology* 12.5, pp. 823–848 (cit. on pp. 21, 22, 28, 84).

Werner, P. C. and F.-W. Gerstengarbe (1997). 'Proposal for the development of climate scenarios'. In: *Climate Research* 8, pp. 171–182 (cit. on p. 17).

Wilby, R. L., T. M. L. Wigley, D. Conway, P. D. Jones, B. C. Hewitson, J. Main and D. S. Wilks (1998). 'Statistical downscaling of general circulation model output: A comparison of methods'. In: *Water Resources Research* 34.11, pp. 2995–3008 (cit. on p. 13).

Wilks, D. S. (2010). 'Use of stochastic weather generators for precipitation downscaling'. In: *Wiley Interdisciplinary Reviews: Climate Change* 1.6, pp. 898–907 (cit. on p. 13).

Wilks, D. S. (2012). 'Stochastic weather generators for climate-change downscaling, part II: multivariable and spatially coherent multisite downscaling'. In: *Wiley Interdisciplinary Reviews: Climate Change* 3.3, pp. 267–278 (cit. on p. 13).

Zhu, X., W. Wang and K. Fraedrich (2013). 'Future climate in the Tibetan Plateau from a statistical regional climate model'. In: *Journal of Climate* 26 (24), pp. 10125–10138 (cit. on pp. 17, 69).

Zorita, E. and H. von Storch (1999). 'The analog method as a simple statistical downscaling technique: Comparison with more complicated methods'. In: *Journal of Climate* 12.8, pp. 2474–2489 (cit. on p. 13).

List of Tables

100

List of Figures

Acknowledgements

104

BONNER METEOROLOGISCHE ABHANDLUNGEN

Herausgegeben vom Meteorologischen Institut der Universität Bonn durch Prof. Dr. H. FLOHN (Hefte 1-25), Prof. Dr. M. HANTEL (Hefte 26-35), Prof. Dr. H.-D. SCHILLING (Hefte 36-39), Prof. Dr. H. KRAUS (Hefte 40-49), ab Heft 50 durch Prof. Dr. A. HENSE.

Heft 1-63: siehe `http://www.meteo.uni-bonn.de/bibliothek/bma`

64-77: open access, verfügbar unter `https://uni-bn.de/kpSDaQffe1`

Heft 64: ***Michael Weniger***: Stochastic parameterization: a rigorous approach to stochastic three-dimensional primitive equations, 2014, 148 S. + XV.

Heft 65: ***Andreas Röpnack***: Bayesian model verification: predictability of convective conditions based on EPS forecasts and observations, 2014, 152 S. + VI.

Heft 66: ***Thorsten Simon***: Statistical and Dynamical Downscaling of Numerical Climate Simulations: Enhancement and Evaluation for East Asia, 2014, 48 S. + VII. + Anhänge

Heft 67: ***Elham Rahmani***: The Effect of Climate Change on Wheat in Iran, 2014, [erschienen] 2015, 96 S. + XIII.

Heft 68: ***Pablo A. Saavedra Garfias***: Retrieval of Cloud and Rainwater from Ground-Based Passive Microwave Observations with the Multi-frequency Dual-polarized Radiometer ADMIRARI, 2014, [erschienen] 2015, 168 S. + XIII.

Heft 69: ***Christoph Bollmeyer***: A high-resolution regional reanalysis for Europe and Germany - Creation and Verification with a special focus on the moisture budget, 2015, 103 S. + IX.

Heft 70: ***A S M Mostaquimur Rahman***: Influence of subsurface hydrodynamics on the lower atmosphere at the catchment scale, 2015, 98 S. + XVI.

Heft 71: ***Sabrina Wahl***: Uncertainty in mesoscale numerical weather prediction: probabilistic forecasting of precipitation, 2015, 108 S.

Heft 72: ***Markus Übel***: Simulation of mesoscale patterns and diurnal variations of atmospheric CO_2 mixing ratios with the model system TerrSysMP-CO_2, 2015, [erschienen] 2016, 158 S. + II

Heft 73: ***Christian Bernardus Maria Weijenborg***: Characteristics of Potential Vorticity anomalies associated with mesoscale extremes in the extratropical troposphere, 2015, [erschienen] 2016, 151 S. + XI

Heft 74: **Muhammad Kaleem**: A sensitivity study of decadal climate prediction to aerosol variability using ECHAM6-HAM (GCM), 2016, 98 S. + XII

Heft 75: **Theresa Bick**: 3D Radar reflectivity assimilation with an ensemble Kalman filter on the convective scale, 2016, [erschienen] 2017, 96 S. + IX

Heft 76: **Zied Ben Bouallegue**: Verification and post-processing of ensemble weather forecasts for renewable energy applications, 2017, 119 S.

Heft 77: **Julia Lutz**: Improvements and application of the STatistical Analogue Resampling Scheme STARS, 2016, [erschienen] 2017, 103 S.