Statistical Downscaling of Variables with Hydrological Importance in the Alpine Region via Machine Learning Techniques

Severin Kaspar

Angaben zur Veröffentlichung / Publication details:

Kaspar, Severin. 2020. “Statistical Downscaling of Variables with Hydrological Importance in the Alpine Region via Machine Learning Techniques.” Augsburg: Universität Augsburg.

UNIVERSITY OF AUGSBURG

Statistical Downscaling of Variables with Hydrological Importance in the Alpine Region via Machine Learning Techniques

by

Severin Kaspar

Diplom-Geograph Univ.

A thesis submitted in partial fulfilment
for the degree of Doctor of Natural Sciences

in the
Faculty of Applied Computer Science
Department of Geography

Universität Augsburg
Institut für Geographie

Augsburg 2019

First Reviewer: PD Dr. Andreas Philipp

Second Reviewer: PD Dr. Christoph Beck

Thesis defence: 30.06.2020

Contents

List of Figures

List of Tables

List of Abbreviations

20Cv2	20th Century Reanalysis dataset version 2
AGCM	Atmospheric General Circulation Model
ALR	Adaptive learning rate training algorithm
AMO	Atlantic Multidecadal Oscillation
ANN	Artificial Neural Network
AOGCM	Coupled Atmosphere-Ocean General Circulation Models
CDD	Consecutive dry days
CDF	Cold day frequency
CMIP5	Coupled Model Intercomparison Project Phase 5
CRS	Coarse resolution subregions
CWD	Consecutive wet days
CTC	Circulation Type Classification
DD	Dynamical downscaling
DEM	Digital elevation model
DF	Distant future reference period: 2071-2100
DJF	Winter: December, January, February
DKM	Differentially initialised k-means clustering algorithm
DWD	Deutscher Wetterdienst
e.g.	Exempli gratia, for example
EDBD	Extended delta-bar-delta training algorithm
ENSO	El Niño and Southern Oscillation
ESM	Earth System Model
F1-F4	Kleinfleißgees snow depth measuring sites
Fig.	Figure
G1-G6	Goldbergkees snow depth measuring sites
GAR	Greater Alpine Region, $4°$ to $19°$ E and $43°$ to $49°$ N
GCM	General Circulation Model
GD	Gradient descent training algorithm
GDDE	Gradient descent with decaying eta
HFCs	Hydrofluorocarbons
Hist	Historical reference period: 1971-2000
HISTALP	Historical instrumental climatological surface time series of the greater Alpine region
IALR	Individual adaptive learning rate training algorithm
IPCC	Intergovernmental Panel on Climate Change
ISD	Interannual standard deviation

JJA	Summer: June, July, August
MAE	Mean Absolute Error
MAM	Spring: March, April, May
MMM	Multi model mean
MOS	Model Output Statistics
MSE	Mean Square Error
MSSS	Mean Square Skill Score
MT	Momentum Term version of Gradient Descent training algorithm
NAO	North Atlantic Oscillation
NF	Near future reference period: 2021-2050
NNC	A combined approach of Artificial **N**eural **N**etworks and Circulation Type **C**lassification
NM0	Number of snow free months
NWP	Numerical weather prediction model
OGCM	Oceanic General Circulation Model
P5max	Annual maximum five day precipitation sum
PaD	Partial Derivative sensitivity analysis for Artificial Neural Networks
PCA	Principal Component Analysis
Pf90	Frequency of precipitation days exceeding the observational $90\,\%$ quantile
PP	Perfect Prognosis
RCF	Reference Class Forecast
RCM	Regional Climate Model
RCP	Representative concentration pathway
RH1min	Annual minimum daily mean relative humidity
RHf95	Frequency of days exceeding the observational $95\,\%$ quantile of daily mean relative humidity
Rprop	Resilient Backpropagation
SANDRA	Simulated annealing and diversified randomization
SD	Statistical downscaling
SDI	Standard precipitation index
SON	Autumn: September, October, November
Son	Sonnblick
SRES	Special report on emissions scenarios
SST	Sea surface temperature
T1max	Annual maximum daily mean temperature
T1min	Annual minimum daily mean temperature
Tab.	Table

THC	Thermohaline circulation
WD	Weight decay training algorithm
WDF	Wet day frequency
WS1max	Annual maximum daily mean wind speed
WSf12	Frequency of days with an average wind speed above $12\,\mathrm{m\,s^{-1}}$
Z1	Zugspitze snow depth measuring site
ZAMG	Zentralanstalt für Meteorologie und Geodynamik
Zug	Zugspitze

Abstract

The European Alps are particularly vulnerable to the effects of climate change, as evidenced by an above-average increase in surface air temperature in recent decades. Besides the thermal changes impacts on the hydrological cycle can also be expected. Alpine water resources act as an important factor not only for the ecosystem but also for humanity, in the form of water security considering irrigation, hydrological energy production or winter tourism. Consequences are manifold but yet often not sufficiently investigated. This study is intended to give insight on modifications in the locally available water resources at the Zugspitze and Hoher Sonnblick area deducted from observed and simulated meteorological time series. Target variables include temperature and precipitation at both stations as well as relative humidity and wind speed at Zugspitze.

Future scenarios are provided by Earth System Model projections. To bridge the gap between the coarse spacial resolution of climate models, a statistical downscaling framework is developed to derive the corresponding local impact. The statistical transfer function between the large and local-scale is gained from non-linear methods, namely Artificial Neural Networks, Cluster Analysis and a novel approach combining both methods. Statistical downscaling models are calibrated and evaluated on a daily basis considering large-scale reanalysis datasets and local-scale observations. A main goal of this study is to find the best predictor setup with Artificial Neural Networks, whereby in climate studies rarely used sensitivity studies, in particular the Partial Derivative method, produced the most reliable results. Usually best modelling performance in validation is obtained by Artificial Neural Networks.

Future changes are investigated by transferring the statistical downscaling models on Earth System Model datasets. Hereby, the historical period is compared to time periods within future scenarios RCP 4.5 and RCP 8.5 and analysed for differences. Considering precipitation, high increases are found at both stations in winter, suggesting a future shift in the annual precipitation distribution. With low uncertainties, temperature is simulated to increase successively, with most intense warming taking place in summer and winter. Simultaneously, relative humidity at Zugspitze is found to continuously decrease. A weak trend in wind speed at Zugspitze is projected by the statistical downscaling models, showing a significant increase not before the end of the 21st century.

The resulting time series of the downscaling framework are applied in a statistical modelling application based on Artificial Neural Networks targeting monthly measurements of snow depths at both stations. In the annual average especially the RCP 8.5 based simulations expect partly drastically reduced snow depths of up to $50\ \%$.

1 Introduction

Preface

The Alpine region is considered to be highly vulnerable to climate change. Between the late 19th and early 21st century an overall temperature increase of $2\,^{\circ}\mathrm{C}$ is observed, which is more than twice the rate of the average warming of the northern hemisphere (Auer et al. 2007). Consequences are manifold but yet often not sufficiently investigated. Observed climate change impacts are, for example, the Alpine area wide melting of glaciers, changes in river runoffs or a shift in snow lines. The Alps contribute a large proportion of runoff to some of the main European river systems like the Danube, Rhine, Po and Rhone (Weingartner et al. 2007). Changes in water resources therefore affect not only agriculture, industry, hydro power production or winter tourism in the Alpine region but have a widespread influence on nature as well as water related management throughout Europe.

This thesis aims to contribute to the understanding of the Alpine water cycle, in particular, by analysing important influential meteorological variables at two high mountain stations, the Zugspitze located at the northern edge of the Alps in Southern Germany and the Hoher Sonnblick near the meridional centre of mountain range in Austria in a changing climate. Parts of this study result from investigations during the project Virtual Alpine Observatory (VAO II) founded by the Bavarian State Ministry of the Environment and Consumer Protection. Focus lies on the future development and characteristics of local temperature, precipitation, relative humidity and wind speed derived from Earth System Model (ESM) projections by non-linear statistical downscaling techniques. Statistical downscaling is a necessity to bridge the gap between the coarse resolution of Earth System Models, which make small scale impact analysis in complex terrain impossible, and the local climate. Statistical downscaling models are initially calibrated and evaluated on observed and reanalysis datasets and are, in a second step, transferred onto Earth System Model scenarios. In this thesis the focus on building transfer functions between the large and local-scale lies on non-linear techniques. In particular Artificial Neural Networks, Cluster Analysis and a novel approach of a combined application of both methodologies. The resulting time series of the statistical downscaling framework are then applied in an impact study on snow depths at Zugspitze and Hoher Sonnblick.

The highly elevated locations of Zugpitze and Hoher Sonnblick bear potential advantages and disadvantages. On the one hand, their summits are exposed to the free atmosphere resulting in a more direct impact of the above boundary layer circulation, which is usually better represented in Earth System Models. On the other hand, the complex terrain surrounding the summits can lead to the development of micro and meso climatic processes blurring the large-scale influence, e.g. turbulence or meso scale wind systems.

Motivation and Primary Aims of the Study

The major goals of this study can be summarised as follows:

- Future climatic conditions in the Alpine study areas Zugspitze and Hoher Sonnblick shall be derived from climate model scenarios in a statistical downscaling approach and the resulting local time series shall be analysed in their characteristics.
- In statistical downscaling so far underrepresented non-linear modelling techniques shall be evaluated in the scope of their applicability in a statistical downscaling framework.
- A major issue in the application of Artificial Neural Networks in climate studies is the identification of important input variables and their respective domain. A suitable approach needs to be identified.
- A novel approach of combining cluster analysis with Artificial Neural Networks shall be developed and investigated for a possible added value to the individual approaches.

Structure of this Thesis

Initially, an overview over key-factors and driving forces of climate change is given in Chapter 2 followed by a general description of the geographical and climatological aspects in the Alpine region in Chapter 3. In Chapter 4, the applied local- and large-scale datasets are summarised. Tools for statistical analyses and the statistical downscaling approaches are presented in Chapter 5, including a detailed insight on the difficulties on working with Artificial Neural Networks and the realisation of the novel combined approach of Cluster Analysis and Artificial Neural Networks. The large-scale climate model inputs and their climate signal are investigated in Chapter 6. Chapter 7 describes the development of the statistical downscaling approach with focus on the input variable selection and tuning of the statistical model parameters. Chapter 8 to 11 cover the analyses of the observed and simulated time series of precipitation, temperature, relative humidity and wind speed, respectively. In Chapter 12, the derived time series of the downscaling approach are applied in a climate

change impact study on snow depths at Zugspitze and Hoher Sonnblick. Finally, the major findings of this study are summarised in Chapter 13.

2
Climate Change

This chapter gives an introduction to the main drivers of climate change and the potentially resulting consequences on the recent and future climate conditions - worldwide and thus also for the investigation area of this thesis. Climatic changes originate from internal processes as well as external forcings according to Latif (2009). Internally induced natural climate variability modes are described by Jacobeit (2007): They result from oceanic variations, e.g. the Thermohaline Circulation (THC), from atmospheric variability modes such as the northern Atlantic Oscillation or from combined atmospheric and oceanic interactions like El Niño Southern Oscillation (ENSO). Changes in frequency and/or intensity of these modes affect the climate from regional to global extend.

Interrelated with the THC is the Atlantic multidecadal oscillation (AMO), which describes a periodical change in North Atlantic sea-surface temperature based on an underlying wavelength of approximately 60 to 90 years with high influence on the European climate and therefore most likely on the Alpine region as well (Knudsen et al. 2011; Knight et al. 2006; Schlesinger and Ramankutty 1994). Sutton and Hodson (2005) show that a warm AMO phase comes hand in hand with increased summer precipitation and temperatures in western Europe. Furthermore, Sutton and Dong (2012) suggest that during the 1990s, a substantial shift of European climate towards anomalously mild and wet summers in northern Europe and hot and dry summers in southern Europe is related to the AMO, switching to the recent warm phase. In addition, Knight et al. (2006) find positive near-surface air temperature anomalies in Europe and increased cyclonic pressure anomalies over the Atlantic and Europe at positive AMO conditions throughout all seasons. The impact of the AMO on Alpine climate has so far not been investigated very well. Zampieri et al. (2013) assess snowfall variations in the Swiss and French Alps occurring simultaneously with changes of the AMO phase. Transitions from cold to warm phases of the AMO can produce significant snowfall reductions up to $30\ \%$ with a more robust signal in the western Alps, especially in spring. In the case of warm to cold AMO transitions, the signal is not as strong, indicating a slight increase in precipitation. Moreover, Huss et al. (2010) show that glacier mass budget in the Swiss Alps varies in phase with the AMO in the past 250 years. In warm AMO phases glaciers are retreating,

while in cold phases glacier advance.

ENSO labels a coupled ocean (El Niño/La Niña) and atmospheric (Southern Oscillation) see-saw in the equatorial Pacific region (T. Li and Hsu 2018). The more frequent condition is characterised by a cold sea current (Humboldt current) at the South American coast near Equator and Peru in combination with a downward movement in the atmosphere and a very dry climate at the coastal region (Atacama desert). In contrast, the counter part near North Australia and Indonesia shows an upward movement in the atmosphere with an increased chance of precipitation and high sea surface temperatures (SSTs) in the equatorial West Pacific. An extreme occurrence of this state is called La Niña. In time scales of 2–8 years, the system shows inverted conditions, known as El Niño, resulting in high precipitation at the South American coast and drought in the western Pacific region (D'Arrigo et al. 2005; Moy et al. 2002). ENSO strongly influences climate in the Pacific region, but through teleconnections, ENSO signals can be detected worldwide (Alexander et al. 2002; Ott et al. 2015; B. Wang et al. 2000) and even impact European climate (Mariotti et al. 2002; Brönnimann et al. 2007; Greatbatch et al. 2004; Pozo-Vázquez et al. 2001; Fraedrich and K. Müller 1992). Nevertheless, a clear signal in the Alps is hard to identify. Studies by Quadrelli et al. (2001) and Wanner et al. (1997) and Fraedrich and K. Müller (1992) find no evidence for an ENSO related signal in the European Alps. Durand et al. (2009) investigated ENSO response on snow parameters in the French Alps for the later 20th century, but could not establish a link either. Yet, Moy et al. (2002) discovered a correlation between the south Alpine occurrence of extreme flood events and a record of moderate to strong El Niño events, which might indicate a teleconnection. Efthymiadis et al. (2007) associated ENSO indices with temperature and precipitation in the Alpine region finding only weak and non-stationary impacts during the late 19th and 20th century. Temperature increased, in most cases, in late autumn and early winter after an El Niño event in summer, except for the interval 1925 to 1972, where temperature decreased. Precipitation mainly showed a decrease following La Niña events.

The North Atlantic Oscillation (NAO) relates to the sea level pressure difference and respective location of the Icelandic low and the Azores high (Wanner et al. 2001). The NAO varies over time on interannual and multidecadal time scales. The NAO pressure see-saw is one of the main drivers of European weather and climate (Hurrell et al. 1995). During a positive NAO phase, one can observe opposite behaviour in southern and northern Europe precipitation, with wetter and warmer than normal conditions in northern Europe and drier and often colder than normal conditions in southern Europe and in the Mediterranean area. The opposite is true for periods when the NAO index is negative. Impact studies of the NAO phases on the Alpine climate were performed by various authors (Agrawala et al. 2007; Brunetti et al. 2006;

Casty et al. 2005; Beniston 2005; Hurrell et al. 1995). In general, the NAO index shows a negative correlation with precipitation, although Brunetti et al. (2006) find no significant relation in the northern parts of the Alps. Furthermore, Casty et al. (2005) depict the correlations as temporally unstable. The conclusions of this article indicate that the Alps are situated in a weak and probably varying forcing of the NAO, and therefore other atmospheric circulation modes may dominate during low correlation periods. Yet, all previous mentioned studies focusing temperature agree that the NAO index is positively correlated with Alpine temperatures.

The climate of our planet is not solely influenced by internal drivers, but is also affected by external forces. These are either of natural origin or caused or altered by mankind. Sorted by the speed of change from slowest to fastest, natural processes include for example plate tectonics, variations in Earth's orbital parameters, differences in solar radiation and volcanic activities (Latif 2009; Jacobeit 2007).

On long time scales of millions of years, plate tectonics change the Earth surface by shifting the position of continents or impact the atmospheric circulation by orogeny. Variations in the Earth's orbital parameters are often referred to as Milankovitch cycles and affect the orbital shape (eccentricity with a cycle duration of approximately 100000 years), the axial tilt (obliquity with a cycle duration of approximately 41000 years) and precession (axial precession and apsidal precession with a combined cycle duration of approximately 22000 years). The Milankovitch cycles are considered as the drivers of ice age cycles. The impact on climate from the Milankovitch cycles in small time periods is little. Cubasch and Kasang (2000) estimate changes in temperature to approximately $0.01\,^{\circ}\mathrm{C}$ per 100 years. Solar radiation varies with the periodicity of sunspots. One of the most well known cycles, the Schwabe cycle, has a duration of approximately eleven years and is therefore responsible for mainly short-term variability. Finally, the climate influence of volcano activities has short term impact on climate in the following couple of years. Thereby, particulate matter, transported into the stratosphere from strong eruptions, results in a temporal cooling.

The recently observed global temperature increase in the 20th century cannot be explained only by the described natural causes, as the impact of the natural influencers is either too slow or too fast or the effects on climate are in the considered time period are too low. Next to the possibility of a yet unknown natural cause, the most likely reason for the current climate change is of anthropogenic origin.

First of all, the greenhouse effect is anthropogenically altered by emitting climate affective trace gases. Initially, the greenhouse effect is of natural origin. Longwave infrared radiation emitted by the Earth, balancing the short wave solar radiation input, is absorbed by trace gases in the atmosphere and partly emitted back to the Earth surface as counter-radiation. Increasing the amount of trace gases in turn increases

Table 2.1: Characteristics of trace gases and anthropogenic influences. Units: dry-air mole fractions in parts per million (ppm), billion (ppb) and trillion (ppt). Sources: Trace gas origin and concentration in 1750 by Latif (2009); Resident time by IPCC (2007); Concentration from 2017 provided by ESRL (2018); Contribution to the anthropogenic green house effect computed from datasets provided by ESRL (2018).

Trace gas	**Primary anthropogenic origin**	**Concentration 1750**	**Concentration 2017**	**Atmospheric residence time in years**	**Contribution anthropogenic green house effect 2017**
CO_2	incineration	280 ppm	405 ppm	(?)	66 %
CH_4	agriculture, incineration	730 ppb	1850 ppb	9	17 %
N_2O	incineration, fertilizer	270 ppb	330 ppb	131	6 %
CFC-12	manufactured purely industrial	(0)	509 ppt	100	5 %
CFC-11	manufactured purely industrial	(0)	229 ppt	45	2 %

the amount of counter-radiation. Since the industrial revolution, the composition of gases in the air is heavily influenced by human activities. Usually, the year 1750 is used as reference, where atmospheric air composition is primarily developed naturally.

First, of particular interest are climate affective trace gases emitted by mankind with a long residence time in the atmosphere. In that case, turbulent mixing of the atmosphere and the planetary circulation lead to a global dispersion. Relevant gases with long residence in the atmosphere are carbon dioxide (CO_2), methane (CH_4), nitrous oxide (N_2O) and hydrofluorocarbons (HFCs). Table 2.1 lists the respective characteristics. CFC-12 and CFC-11 are selected as representatives for HFCs, as these two variations are considered the most harming. Anthropogenic sources of trace gases are combustion of fossil fuels or biomass and agriculture. Most trace gases have natural sources as well, with the exception of HFCs, which are produced synthetically. Comparing concentrations from 1750 and 2017 leads to the conclusion that climate relevant trace gas concentrations increased heavily. CO_2 concentration is raised by 45 %, CH_4 by 156 % and N_2O by 22 %. In contrast, the concentration of HFCs is currently regressive, as recently, due to the Montreal Convention, chemical

substitutes are used. CFC-12 reached its peak in 2002 with 543 ppt and CFC-11 in 1993/1994 with 268 ppt (ESRL 2018). Although concentrations of some trace gases are low in comparison to CO_2, the amplification of counter-radiation can be much higher. HFCs for example are considerably less present in air (ppt) than CO_2 (ppm) but still have a strong impact on the anthropogenic green house effect. Note that the residence time of CO_2 is discussed controversial. The reasons are the complex interactions with oceans and biosphere. Resident times between decades and millennials are suggested in the IPCC (2007). Additional trace gases of anthropogenic origin are tropospheric ozone (O_3) and water (H_2O). O_3 is primarily locally produced by photochemical reactions. Concentrations vary strongly as the residence time in the atmosphere is quite short. Nevertheless, O_3 has a much higher greenhouse gas potential than CO_2. H_2O is the most important natural green house gas. The anthropogenic influence on H_2O concentrations in air is hard to identify. H_2O can be, for example, emitted by incineration or air planes into the atmosphere. Comparing 1750 to 2017, the total additional radiative forcing caused by anthropogenic interference in the green house effect is suggested to be $3.1\,\mathrm{W\,m^{-2}}$ but bears large uncertainties (ESRL 2018).

Second, anthropogenic emitted particulate matter has an overall decreasing effect on solar radiation, often referred to as "global dimming" (Ramanathan et al. 2001). An increased amount of particulate matter leads to a compensation of anthropogenic green house effect, which was especially evident in the period 1950-1980 on the northern hemisphere (Cubasch and Kasang 2000).

Third, land use changes the Earth surface, which again leads to changing climatic conditions. Land use can either have positive or negative effects on solar radiation (Jacobeit 2007). On the one hand, the reduction or change of the vegetation coverage, e.g. clearing or degradation of vegetation, increases the Earth's albedo resulting in a cooling effect. On the other hand, fire clearing and various agricultural forms, e.g. rice cultivation, livestock farming or the usage of fertilizer, emit green house gases resulting in an enhanced counter-radiation.

In this thesis, the impact of climate change is evaluated for the high mountain regions Zugspitze and Hoher Sonnblick during the 20th century, as well as in future scenarios of the 21st century simulated by Earth System Models. To compute the future climate development, beside natural climate interactions, scenarios of anthropogenic influences are used to assess a broad spectrum of possible climatic trends. These so called "representative concentration pathways" (RCPs) include anthropogenic and natural climatic influences and the expected impact on radiative forcing. RCPs will be discussed in section 4.2.2.

3

Study Region: The European Alps

3.1 Geomorphology and Spatial Delineation

The Alpine mountain range, as described in Pfiffner (2015), is the result of the converging of primarily two tectonic plates: the Eurasian Plate and African Plate. The orogeny started approximately 125 million years ago and is still ongoing. A distinction is made between two main stages. First, during the Cretaceous period, the converging movement was primarily aligned in the east-west direction, while second, during the Cenozoic, the movement changed to a northern-southern one. The latter more direct encounter of the tectonic plates lead to a high gain in elevation. As soon as the surface rises in the orogeny process, erosion sets in (Zepp 2014). Beside a general loss in surface material, a more severe removal at geological weak points leads to the development of valleys usually caused by rivers.

The spacial expansion of the European Alps, in this study simply referred to as Alps, will in the following be defined by the greater Alpine region (GAR) ranging from $4°$ to $19°$ E and $43°$ to $49°$ N (Auer et al. 2007). The term and dimension is recently widely adapted by Alpine climate studies (Rubel et al. 2016; Smiatek et al. 2009; Agency 2009; Barry 2008; Efthymiadis et al. 2007; Brunetti et al. 2006; Matulla et al. 2005). The Alps are of arcuated form especially in the western parts with a primary extension parallel to the circle of latitude. Therefore, the Alps in general accentuate the climatic gradient between the Mediterranean climate and the climate of central Europe (Barry 2008). Figure 3.1 illustrates the spacial shape of the Alps by the snow cover of high elevated regions. The width of the Alps varies between $100\,\mathrm{km}$ in Switzerland and $250\,\mathrm{km}$ in the western parts (Barry 2008), the highest mountain being the Mont Blanc with an elevation of $4810\,\mathrm{m}$ above sea level. The Alps are subdivided by pronounced valleys such as the Rhône, Po, Rhine, Inn, Salzach, Enns, Mur or Drau. The heterogeneity in relief between high mountain summits and valleys lead to very specific regional weather and climate (Cebon 1998; Whiteman 2000; Barry 2008). The associated rivers secure water availability in large areas of Europe.

Figure 3.1: MODIS satellite image of the European Alps on March 19, 2016. The snow covered regions emphasise the Alpine mountain range. Source: Schmaltz (2016).

3.2 Atmospheric Influence on the Alps Considering Different Scales of Atmospheric Motion

Influencing factors on different atmospheric scales determine the Alpine climate. On the global scale, the main drivers are resulting from the energy gradient between the tropical and polar regions. The steep angle of solar incidence near the equator leads to an energy abundance compared to the polar regions, resulting in a (seasonally varying) pole-to-equator temperature difference. A direct compensation movement is prevented by the Coriolis effect. Trajectories of air streams are bent to the right on the northern hemisphere and bent to the left on the southern hemisphere. In the long term mean, a typical global pattern, referred to as planetary circulation, forms. A detailed description of the planetary circulation is beyond the scope of this study, but is given by various authors, for example Randall (2015) or Weischet and Endlicher (2008). Three major cells can be distinguished in the planetary circulation: The Hadley cell in the tropical latitudes (roughly $0°$ to $30°$ N/S), the Ferrel (or mid-latitude) cell (roughly $30°$ to $60°$ N/S) and the polar cell (roughly $60°$ to $90°$ N/S). The Alpine region is located in the westerlies of the Ferrel-cell.

On the synoptic scale, the westerlies are defined by transient weather systems like cyclones, anticyclones and the resulting fronts, which form, propagate and decay in the mid-latitudes. Different synoptic pattern determine the variations in the day-to-day weather. The particular daily setting can be categorised by similar weather types, which are typically generated from the respective atmospheric pressure field. Horizontal and vertical pressure gradients are closely related to the air stream velocity field, which is in turn highly related to the local climate and weather. This

was already suggested by Kirchhofer (1974) and has become a typical approach in statistical downscaling (Philipp et al. 2014b; Hofstätter et al. 2013; Beck et al. 2013; Jacobeit 2010; Huth et al. 2008a). A comprehensive overview regarding weather types of the European and Alpine specific region is provided by the COST 733 project - Harmonisation and Applications of Weather Type Classifications for European regions (Philipp et al. 2014b). Regarding the long-term mean, the sea level pressure fields over Europe show typical seasonal-depending characteristics. In winter, the Alps are embedded between the Icelandic Low in the north-west, the Azores High in the south-west, the Siberian High to the east and a weak low in the Mediterranean area to the south. Cebon (1998) points out that the Alps are located within the weakest pressure gradients in the extra tropical northern hemisphere. In winter, the main storm track trajectories start over the Atlantic Ocean and end in north-east Europe. The Alps are therefore mainly influenced by the southernmost frontal features. In summer, the Azores high greatly expands to the north becoming the major pressure influence of the Alps and further decreases the influence of the Atlantic storm track. In the consequence even less fronts pass the Alps during summer leading to a reduction of frontal induced precipitation.

The smallest meteorological scales mentioned in this section are the meso- and microscale features. Mesoscale weather events primarily affect the regional climate in the Alps on a spacial range of $2\,\mathrm{km}$ to $200\,\mathrm{km}$ (Whiteman 2000). Examples are thermally driven or diurnal wind systems as well as thunderstorms. The microscale meteorology (range below $2\,\mathrm{km}$) focuses on small-scale atmospheric phenomena like gusts, turbulence, thermals or certain cloud types (Whiteman 2000). The local impact of meso- and microscale atmospheric processes is described in more detail in chapters 8 to 12 in combination with their effect on the respective target variables.

Several aspects need to be kept in mind when considering the chain of spacial scales and their influence on the Alpine region (Cebon 1998). First, the influence of scale is bidirectional and not merely top to bottom. The energy transport from the equator to the pole, for instance, highly depends on synoptic weather systems in the mid-latitudes. Second, all four introduced scales show high temporal variability. Third, the impact of climate change can be transmitted through all scales. Changes in strength, frequency, location or trajectories track length of Atlantic cyclones have a particularly high influence on the regional Alpine climate. Fourth, the Alps themselves impact the weather and climate by affecting the air streams.

3.3 Alpine Regions with Similar Climate Variations

There are some attempts of collecting Alpine wide datasets and make them available to the public, one of the most well-known is the “historical instrumental climatologi-

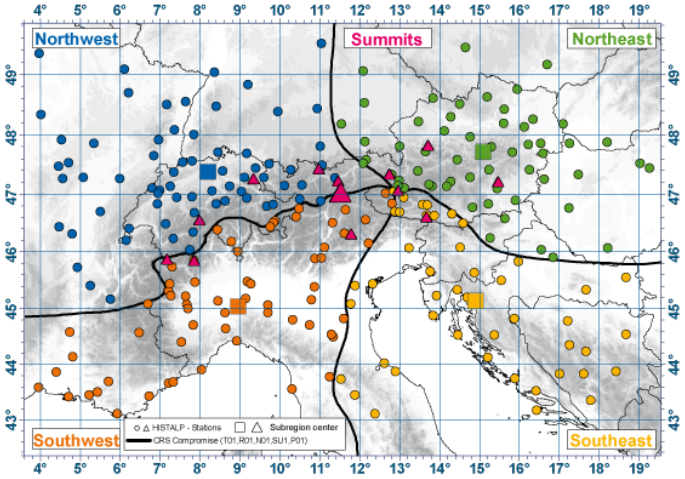


Figure 3.2: Coarse Resolution Subregions (CRS) of the Alps. Source: Auer et al. (2007)

cal surface time series of the Greater Alpine Region" (HISTALP) by Auer et al. (2007). The HISTALP database is consisting of monthly homogenised records of temperature, pressure, precipitation, sunshine and cloudiness for countries in the GAR. The HISTALP dataset can be used to identify subgroups of observational stations with similar climate variations. Auer et al. (2007) combine previous research results from Matulla et al. (2005) for temperature and Brunetti et al. (2006) for precipitation and added further analysis for air pressure, sunshine and cloudiness to define groups of stations with similar variations on an annual time scale via a varimax rotated principal component analysis in s-mode considering a time period from approximately 1930-2000. Areas were defined using the first four principal components. Figure 3.2 shows the final result in form of five Coarse Resolution Subregions (CRS) including a subgroup of summits. Boundaries (black lines) show a compromise between all investigated variables. More detailed figures for each individual variable can be found in Auer et al. (2007). A more or less zonal border following the main crest line separates the temperate westerly (north) from the Mediterranean subtropical climate (south). The meridional border line at approximately $12°$ longitude might be an indicator for a decreasing maritime influence of the Atlantic ocean in favour of more continental features of the Eurasian continent. The fifth CRS introduces vertical discrimination as it separates summits from the horizontal CRS, since especially winter temperature shows a strong decoupling of lower and higher altitudes. Precipitation was not considered for the summit cluster, as Auer et al. (2007) stated that there is almost no agreement between the stations. Both Zugspitze and Hoher Sonnblick belong to the fifth category of summits.

4 Data

4.1 Local-Scale Variables: Measuring Sites and Observational Records

The European Alps have a long history in meteorological observation and, today, represent one of the best monitored mountain ranges. According to Auer et al. (2007) some stations, for instance Basel, provide continuously measured data for air pressure and temperature since the 18th century. Frei and Schär (1998) estimate the number of rain gauges to approximately 6600. The high complexity of the terrain and the resulting heterogeneity in climatic conditions justify the large amount of measurement sites. Unfortunately, only a small number of datasets is internationally available. Auer et al. (2007) further point out that some observational records include a high number of inhomogeneities and outliers.

In this thesis, local variables of two high mountain weather stations in the Alps are investigated. The Bergwetterwarte Zugspitze, operated by the German weather service (Deutscher Wetterdienst, DWD), is located at $2964\,\mathrm{m}$ above sea level in the northern part of the Alps at $47.42°$ latitude and $10.99°$ longitude. Note that in the meta data in 2007 a relocation of the the measurement instruments to $10.98°$ longitude is mentioned, which may lead to inhomogeneities in the dataset. With $3105\,\mathrm{m}$ above sea level, the Sonnblick Observatory, operated by the Austrian weather service (Zentralanstalt für Meteorologie und Geodynamik, ZAMG), is found to be at a slightly higher elevation. The Sonnblick Observatory is located in the centre of the Alpine mountain ridge at $47.05°$ latitude and $12.96°$ longitude. In this study, the Bergwetterwarte Zugspitze is usually simply referred to as "Zugspitze" or abbreviated as "Zug", while the Sonnblick Observatory is called "Sonnblick" or "Son". Both stations are mapped in figure 4.1. Of particular interest are observed variables linked to the Alpine water balance. Table 4.1 shows the selected predictands and gives an overview over the dataset characteristics. A daily temporal resolution is required for all predictands to fulfil the needs of this study. Both stations offer a long-time observational record starting in the late 19th or early 20th century. At the end of World War II data recording at Zugspitze was cancelled from 01/05/1945 until 14/8/1945 resulting in the large amount of missing values considering precipitation, tempera-

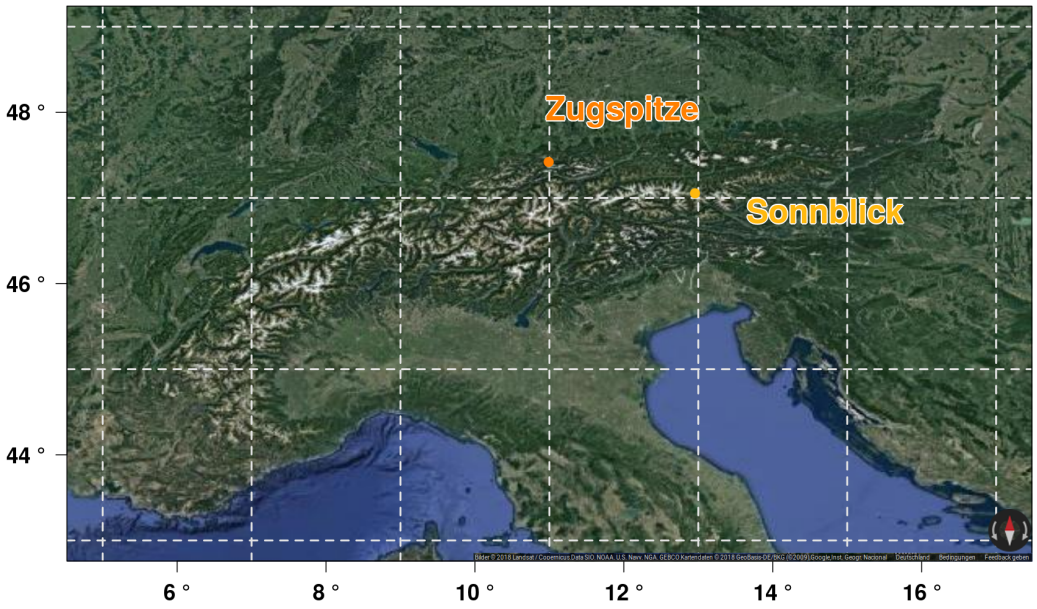


Figure 4.1: Satellite image of the Alpine relief with the marked locations of the Zugspitze and Hoher Sonnblick. The dashed grey grid indicates the 2° resolution of the large-scale predictors (see section 4.2). (Map based on Google 2018.)

ture and relative humidity. Unavailable measurements are usually omitted, except for the otherwise even shorter Zugspitze wind time series, where missing values are linearly interpolated.

One main point of this study is to analyse the future development of the local climate conditions at Sonnblick and Zugspitze. Scenario projections of future climate are derived from state of the art Earth System Models. Technical limitations of these models lead to a very coarse resolution (discussed in section 4.2). The common denominator of the spacial resolution considering all applied large-scale datasets in this study is a 2° longitude and latitude grid marked with dashed grey lines in figure 4.1. The Alpine mountain ridge and its topological complexity is poorly represented in the large-scale datasets. The Alps are located in an approximate total of ten grid cells. Each cell provides one single value intended to represent the average magnitude of a large-scale variable. If located in the same segment, peaks and valleys would be characterised by the same grid value. In the high complex terrain of the Alps it is therefore not feasible to analyse small scale climate directly from large-scale datasets, considering the coarse spacial resolution. To fulfil the aim of this study and generate local future projections for the summit stations Zugspitze and Sonnblick, downscaling procedures are indispensable. Several downscaling methodologies are evaluated and applied in this study (see chapter 7).

Table 4.1: Meteorological variables monitored at Zugspitze and Sonnblick.

Station	Predictand	Abbreviation	Recording period	Number of missing days
Zugspitze	daily precipitation sum	prc	01/01/1901 - 31/12/2015	108
	daily mean temperature	tmp	01/08/1900 - 31/12/2015	106
	daily mean relative humidity	rhum	01/08/1900 - 31/12/2015	140
	daily mean wind speed	wnd	01/01/1976 - 31/12/2015	22
Sonnblick	daily precipitation sum	prc	01/08/1890 - 31/12/2008	2
	daily mean temperature	tmp	01/10/1886 - 31/12/2008	0

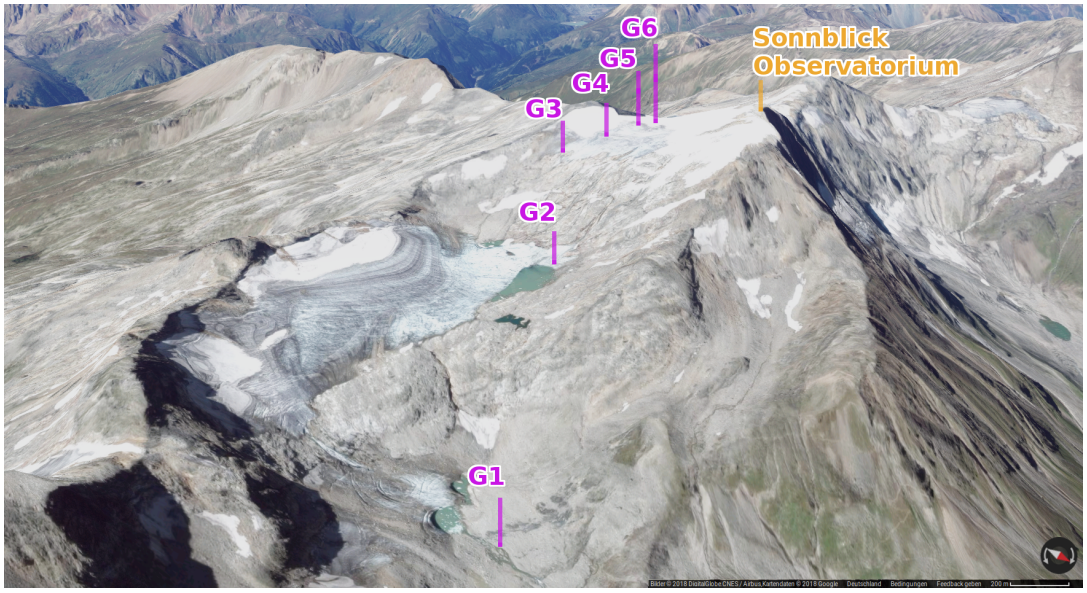


Figure 4.2: Measuring sites at the Goldbergkees. See table 4.2 for more details. (Map based on Google 2018.)

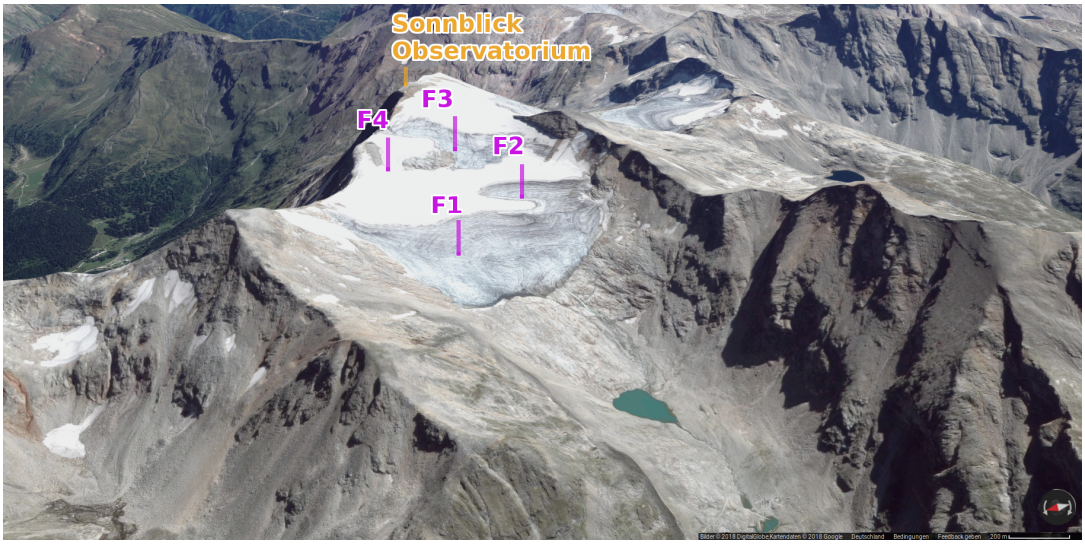


Figure 4.3: Measuring sites at the Kleinfleißgees. See table 4.2 for more details. (Map based on Google 2018.)

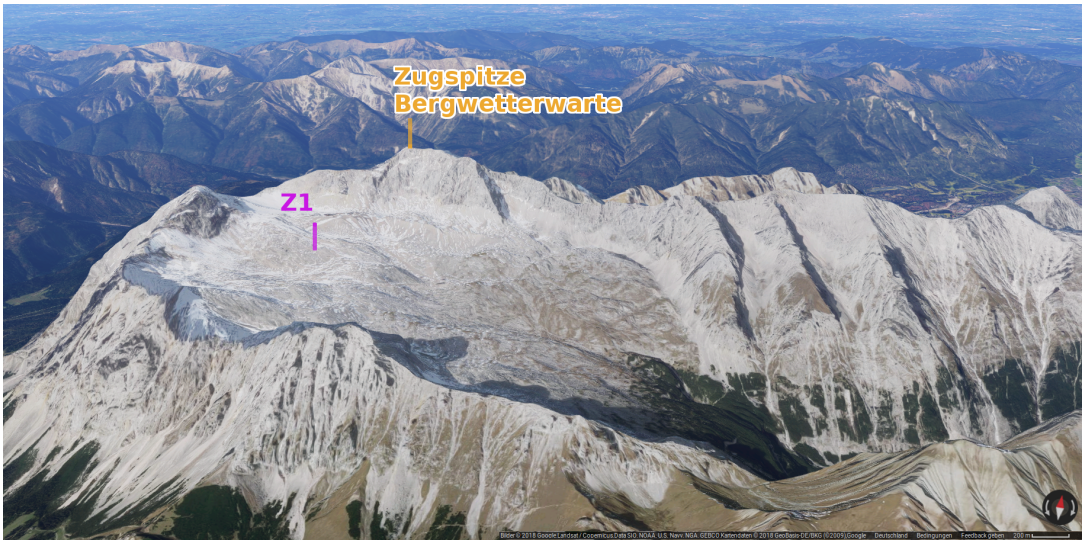


Figure 4.4: Measuring sites at the Zugspitze. (Map based on Google 2018.)

Table 4.2: Snow depth observations at Zugspitze and Sonnblick. The information is based on the meta dataset provided by DWD and ZAMG. ACR: Acronym for further usage in this thesis; NM: number of missing values, Ele: elevation in meter above sea level; Lat: latitude; Lon: longitude.

ZAMG Label	Acr	Data Record	NM	Ele	Lat/Lon
Unterer Goldbergkeesboden	G1	11/1927-10/2014	2	2400	47.0450/12.9779
Oberer Goldbergkeesboden	G2	11/1927-10/2014	0	2670	47.0469/12.9641
Steilhang	G3	11/1927-10/2014	0	2878	47.0477/12.9561
Untere Brettscharte	G4	11/1970-10/2014	0	2923	47.0491/12.9554
Obere Brettscharte	G5	11/1970-10/2014	0	2958	47.0502/12.9550
Fleißscharte	G6	11/1927-10/2014	0	2980	47.0510/12.9550
Fleißkees Zunge	F1	11/1975-10/2014	1	2820	47.0535/12.9430
Fleißkees unten	F2	11/1970-10/2014	2	2860	47.0513/12.9476
Fleißkees oben	F3	11/1970-10/2014	1	2940	47.0527/12.9516
Pilatusscharte	F4	11/1927-10/2014	0	2905	47.0549/12.9502
Zugspitze	Z1	01/1937-12/2013	0 (2872)	2964	47.42/10.98

In addition, time series of snow depths from ten monitoring stations operated by ZAMG in the surrounding Sonnblick area are investigated. The Sonnblick snow depth dataset provides monthly in situ measurements. Locations at Sonnblick are marked for the Goldbergkees (G1-G6) in figure 4.2 and Kleinfleißgees (F1-F4) in figure 4.3. From the Sonnblick summit, the Goldbergkees extends into the south-east direction, while the Kleinfleißgees is located on the western slope. At Zugspitze, one time series of daily snow depth observations is available. The location of the snow depth measuring point is marked in figure 4.4. Further description of the snow depth datasets at Sonnblick and Zugspitze is given in table 4.2. Note that the high number of missing values in the Zugspitze dataset does not affect the model framework in chapter 12, as only the first day of each month is used in the statistical modelling approach, where no missing values are recorded.

Monitoring in high mountain areas with heterogeneous terrain is susceptible to errors. On the one hand, random errors are caused by unknown and unpredictable changes in the experiment (Ren and M. Li 2007). Errors may occur from electronic noises in the circuit of an electrical instrument. On the other hand, systematic errors come from the measuring instrument itself. Precipitation suffers from systematic errors, in particular, induced by wind, wetting loss, evaporation loss or splashing loss (Isotta et al. 2014a; Auer et al. 2007; Ren and M. Li 2007). For a common rain gauge Sevruk (1986) estimates the average systematic error for rainfall in the range of $5\,\%$ to $15\,\%$ and for snow fall between $20\,\%$ to $50\,\%$ at which the largest errors arise

from wind and turbulence. For stations on mountain summits, where wind flow dynamics are complex, the error can be expected to be even higher. To reduce the error, effective protection from wind impact is beneficial (R. Rasmussen et al. 2012). Misinformation from wrong measurements can have a severe consecutive impact in hydrological assessment studies or on the calibration process of downscaling models (Stisen et al. 2012). Measuring snow depth is also affected by wind drift or different melting processes in shaded or non-shaded areas (NOAA 2013). To reduce the consequences of this error, typically several measurements are performed and averaged within a defined measuring field. Temperature, relative humidity or wind are less affected by systematic errors. Still, wrong calibration or neglected maintenance can cause discrepancies. For instance, by incorrect handling, the functionality of cup or vine anemometer can be impaired by icing due to the cold climatic conditions on summits. Furthermore, in a complex relief with heterogeneous microclimatic conditions even short distances can lead to a significant difference in climatic variables (Whiteman 2000). Measurements might not be representative even for nearby surroundings of the climate station. Beside monitoring issues, inhomogeneities can occur especially in long time series. The origin of inhomogeneities comes from further non-climatic impacts on the measuring timeline (Toreti et al. 2011). There are various reasons for inhomogeneities, for example relocation of the measurement instruments, varying time of measurement, replacement of the sensor, calibration of the sensor or changing in the surroundings of the measuring instruments: New buildings or changing vegetation, like tree growth, can influence micro-climatic conditions. Depending on the cause, inhomogeneities show a sharp change (relocation of the sensor) or effect in a creeping way (tree growth).

Following the meta data of the datasets listed in tables 4.1 and 4.2, data quality is routinely checked by both weather services DWD and ZAMG and are therefore not further investigated in this study.

4.2 Large-Scale Variables: Reanalysis and Climate Model Datasets

4.2.1 20th Century Reanalysis Dataset

As a substitute for observational data serves the 20th century reanalysis dataset version 2 (20Cv2) (Compo et al. 2011). Reanalysis data is produced with numerical weather prediction (NWP) models. The aim is to gain a regular grid from a finite set of possibly imperfect or irregularly distributed observations by a process called data-assimilation. The result is a complete and as homogeneous as possible global gridded dataset with a high spacial and temporal resolution (Dee et al. 2011). Re-

analysis datasets are often used in statistical downscaling frameworks, as there are several advantages compared to interpolated observational data. First, as previously mentioned, inhomogeneities or errors in measurements have a strong impact on interpolated data but can be cushioned by the NWP model. Second, reanalysis does not suffer too much from sparsity of observations, but data sparsity can be critical when interpolating observations. Third, variables, even those of which measurements are nearly absent (e.g. moisture flux, radiative heating or omega) or variables that are difficult to interpolate, can be derived from a NWP model on a quasi continuous, three or four dimensional grid (dimensions: longitude, latitude, if feasible vertical levels and time). Fourth, to investigate the impact of climate change with statistical downscaling approaches, "observations" need to be replaced with climate model data in the further progress. Statistical downscaling models calibrated with reanalysis datasets can often be better transferred to climate model data, as both datasets show a similar behaviour, which is typical for numerical models.

The 20Cv2 was generated with the coupled atmosphere–land model NCEP-GFS version 2008 of the Environmental Modeling Center (EMC) (Compo et al. 2011). Available are 6-hour, daily averaged and monthly time steps for the time period 1871 to 2012. The horizontal spatial resolution is $2°$ longitude to $2°$ latitude, with a vertical resolution based on pressure levels ranging from $1000\,\mathrm{hPa}$ to $10\,\mathrm{hPa}$ with a step size of $50\,\mathrm{hPa}$ followed by a reduced step size above the $100\,\mathrm{hPa}$ level. To be able to reconstruct such a long time period, a minimum of observation data is assimilated: Surface pressure, sea-surface temperature and sea-ice distributions are included as boundary conditions. The dataset can be downloaded from the Earth System Research Laboratory (ESRL) of the National Oceanic and Atmospheric Administration (NOAA).

In this thesis, the 20Cv2 is selected, as Zugspitze and Sonnblick provide long observational periods including the entire 20th century (see section 4.1), a similar time period is only provided by the 20Cv2 at the beginning of this study.

4.2.2 Climate Models

Climate modelling has come a long way, starting with zero dimensional energy balance models to the complex, now state-of-the-art, Earth System Models (ESM). The first more dimensional atmospheric models emerge from weather forecasting, steadily improved in line with the performance increase of computer systems. First term for these models was General Circulation Models (GCMs), sometimes also referred to as Global Circulation Models. At a time when only atmosphere or ocean could be considered due to limited computational power, different GCMs for atmosphere (AGCM) and ocean (OGCM) were developed. Later on, AGCMs and

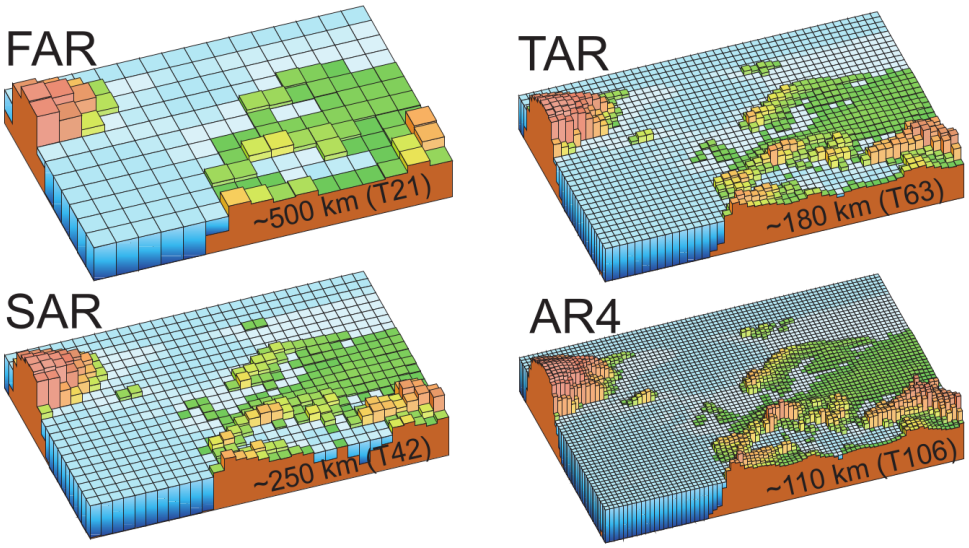


Figure 4.5: Development of the average climate model resolution from the first IPCC assessment report (FAR) over the second (SAR) and third (TAR) assessment report to the fourth assessment report (AR4). The climate model resolution in the current fifth report (AR5) did not substantially change. Source: IPCC (2007).

OGCMs were coupled to atmosphere-ocean general circulation models (AOGCM), a type still common in the climate modelling community, especially the improved versions with land surface modules and sea-ice models (IPCC 2013). Furthermore, the newest generation of ESMs includes biogeochemical cycles, like the carbon cycle, atmospheric chemistry, aerosol chemistry, dynamical vegetation components or land ice modules (Taylor et al. 2012). Internal feedback between the ocean, atmosphere and terrestrial biosphere reservoir is taken into account, e.g. allowing the model to adjust to greenhouse gas concentrations with sinks or natural emission sources. Over the past decades a quite large improvement in climate model performance can be noted with a much more realistic representation of the Earth system (Reichler and Kim 2008). Recent model runs are collected by the Working Group on Coupled Modelling (WGCM) of the World Climate Research Programme (WCRP) within the framework of the Coupled Model Intercomparison Project (CMIP). The current fifth phase (CMIP5) serves as database in this study. The aim of CMIP is to coordinated worldwide climate model experiments. CMIP5 is a popular database with free download access for non-commercial scientific or educational purpose. It includes a large ensemble of over 50 state-of-the-art climate models developed by over 20 working groups in over a dozen countries. A new phase 6 experimental design is currently in development.

To this day, uncertainties in climate models have kept the scientist community

busy. Origins of uncertainties are for example parametrisation techniques. Model parameters are a method to compensate missing subscale processes that cannot be physically described in the coarse temporal and spacial resolution of climate models. Examples for parameters are given by the IPCC (2013). In the atmospheric components of climate models for instance adaptions need to be made for clouds, convective precipitation, gravity waves, radiation (e.g. interaction between radiation and clouds) or aerosols. But the other modules need parameters as well, e.g. the distribution of vegetation, soil types or albedo in land models subscale currents like eddies in the ocean model or ice thickness in sea-ice modules. Parameters need to be set with caution, as they may be dependent on time or location of a specific region (Giorgi and Mearns 1999; Merz et al. 2011). By reason of spacial resolution, another limitation is the digital elevation model (DEM). Mountain ranges like the Alpine region and its heterogeneous relief are poorly represented in climate models. Figure 4.5 shows the horizontal resolution development since the first assessment report (FAR) to the previous fourth assessment report (AR4). The spacial resolution of the current fifth assessment report did not increase substantially in comparison to the AR4 due to the trade-off against higher model complexity in the established Earth system components (IPCC 2013). In addition in figure 4.5, the different relief complexity is indicated. In high complex terrain like in the Alpine region, resolution of climate models is still too coarse to include the heterogeneous climate conditions under the influence of the altitudes. Subsequently applied downscaling techniques can fill the gap by increasing the resolution or even by generating local-scale information. Further, uncertainties emerge from short control runs and inert systems like the ocean (Taylor et al. 2012). Oceans typically react in time periods of thousands of years. Control runs of several hundred years cannot ensure an equilibrium state within climate models, therefore these models might still be affected by climatic drift. Uncertainties can also result from the initial conditions of each model run. Fortunately, uncertainties from different initial conditions can be identified by running the model for several times with different starting patterns and evaluating the resulting climate trajectories. These runs are called realisations in the CMIP5 database (Taylor et al. 2012). The large amount of different model runs contributed to CMIP5 leads to a good representation of the expected future pathways and uncertainties in climate scenarios.

Several experiment families performed with climate models can be distinguished in the CMIP5 experimental design (Taylor et al. 2012). In the following, the experiments for long-term climate modelling used in this study are briefly described. Starting from pre-industrial climate in historical runs, climate models are forced with observed atmospheric composition changes, reflecting both anthropogenic and natural sources. Furthermore, in CMIP5, time-evolving land cover is included. Historical

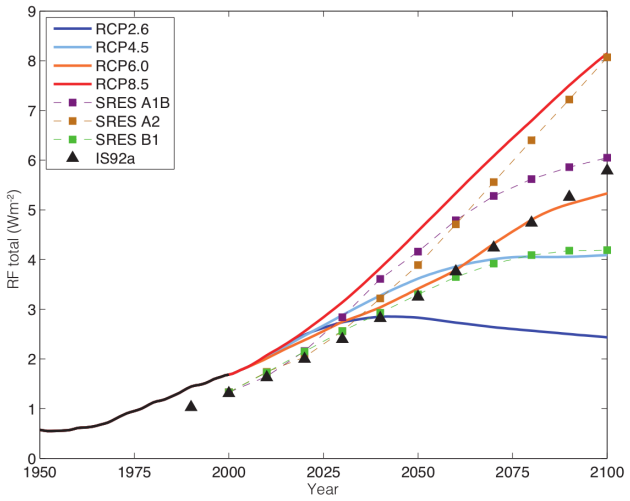


Figure 4.6: Comparison of RCPs and former SRESs including the first scenario IS92a of the first IPCC assessment report. Historical and projected total anthropogenic radiative forcing (RF) in [Wm^{-2}] relative to pre-industrial values (ca. 1765). Source: IPCC (2013).

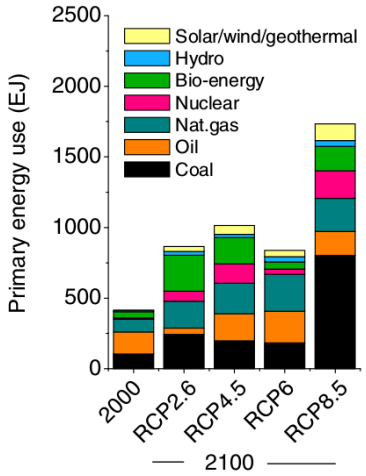


Figure 4.7: Estimated energy usage and sources at the end of the 21st century by different RCP scenarios. Source: Vuuren et al. (2011a).

runs typically cover the time period from 1850 to 2005. The historical runs are followed by future projections guided by RCPs. An overview of RCPs is given by Moss et al. (2010) and Vuuren et al. (2011b). Each RCP is assembled by independent working groups. Figure 4.6 shows the radiative forcing for four RCPs and compares them to selected previous Special Report on Emissions Scenarios (SRES) and the even older business as usual IPCC Scenario from 1992 (IS92a). The number behind RCP represents the resulting radiative forcing at stabilization after 2100 in Wm^{-2} relatively to pre-industrial values, therefore, currently the most pessimistic scenario is represented by RCP 8.5 and the most optimistic by RCP 2.6. The four RCPs illustrated in Figure 4.6 show the main representative pathways. RCP dataset provide primarily emissions and concentrations of greenhouse gases and associated radiative forcing covered in already existing literature analysing the 21st century (Vuuren et al. 2011a). RCPs should not be interpreted as forecasts or absolute bounds. In addition, different climate models using the same RCP can have varying results, as most current climate models can adjust greenhouse gas concentrations during runtime within integrated cycles of matter. Figure 4.7 splits the estimated energy usage into the main sources. Beside a high energy usage, the pessimistic RCP 8.5 scenario shows a heavy increase in non-renewable energy sources like coal. In comparison to the year 2000, the optimistic RCP 2.6 still features an almost doubled energy usage but with focus on renewable energy sources with low impact on radiative forcing.

Table 4.3: Description of Earth System Models applied in this study. In figures and tables of this thesis the model names are usually abbreviated with their first three letters, added in the model name column in brackets.

Model name	**Institute/Working group**	**Realisations**
ACCESS1-0 (ACC)	CSIRO (Commonwealth Scientific and Industrial Research Organisation) and BOM (Bureau of Meteorology), Australia	r1
CMCC-CMS (CMC)	CMCC (Centro Euro-Mediterraneo per i Cambiamenti Climatici), Italy	r1
HadGEM2-CC (HAD)	Met Office, United Kingdom	r1
IPSL-CM5A-LR (IPS)	IPSL (Institut Pierre Simon Laplace), France	r1
MPI-ESM-LR (MPI)	Max-Planck-Institut, Germany	r1, r2, r3

Timeline analysis of climate models needs to be done with care. The starting point of a model experiment (e.g. state of atmosphere and ocean) in the beginning of a scenario varies between different models or realisations. As a result, although most models provide a timeline, used datasets of this thesis cannot be compared directly on a daily basis, but need to be compared in larger, for example climatological, periods. This effect concerns both historical runs and future scenarios.

In this thesis, historical runs and the scenarios RCP $4.5\,\mathrm{W\,m^{-2}}$ and RCP $8.5\,\mathrm{W\,m^{-2}}$, referred to as RCP 4.5 and RCP 8.5, are applied in the statistical downscaling approaches. The three selected experiments belong to the core-set of long-term climate model runs (Taylor et al. 2012), hence are typically available in the CMIP5 database. At time of data acquisition, only five climate models and three different realisations of the MPI-ESM-LR (see Table 4.3) were available on a daily basis providing the required datasets for this study. Table 4.4 summarises included modules and respective categorisation of each climate model. All five models use at least one ESM module and are therefore referred to as ESMs in this thesis.

4.2.3 Selection and Processing of Large-Scale Predictor Variables

Model input variables (predictors) are selected by considering the current state-of-the-art of statistical downscaling in scientific literature as well as variables found to be suiting in this study. Nearly all predictor sets for the model target (predictand) precipitation include circulation based variables like sea level pressure, zonal or meridional wind velocity, geopotential height, atmospheric layer thickness of pressure fields, vorticity or divergence combined with moisture variables like specific humidity, relative humidity, precipitation itself or moisture flux (e.g. Carreau and

Table 4.4: Integrated modules in the selected climate models, which are classified with the new ESM standard by the IPCC 2013 (Physically based equation and two-way coupling of the respective module). In addition, all listed climate models include the AOGCM dedicated modules: Atmosphere, Land Surface, Ocean and Sea-Ice. source: IPCC (2013).

Model name	Aerosols	Atmosphere Chemistry	Land Carbon	Ocean Biogeochemistry
ACCESS1-0	x			
CMCC-CMS	x			
HadGEM2-CC	x		x	x
IPSL-CM5A-LR	x		x	x
MPI-ESM-LR	x	x	x	x

Vrac (2011), Cavazos and B. Hewitson (2002), Dobler et al. (2013), Duan and Mei (2014), Friederichs (2010), Gaitan et al. (2014), Hertig and Jacobeit (2013), Jarosch et al. (2012), Lutz et al. (2012), Maraun et al. (2010), Schmidli et al. (2007), Tareghian and P. F. Rasmussen (2013), Yang et al. (2012), and Wilby and Wigley (2000)). Some authors further include temperature (e.g. Dobler et al. (2013), Gaitan et al. (2014), Jarosch et al. (2012), Schmidli et al. (2007), and Yang et al. (2012)). There are also studies using simpler predictor setups, e.g. only sea level pressure (e.g. Schiemann and Frei (2010)). Predictor sets for temperature include less variety: Setups are often simply based on circulation based variables like sea level pressure or geopotetial height and temperature (e.g. (Hofer et al. 2015; Jarosch et al. 2012; Yang et al. 2012)). Table 4.5 gives a full list of applied predictor variables in this thesis, with respective abbreviations. Pressure level based variables are considered on atmospheric pressure levels $250\,\mathrm{hPa}$, $500\,\mathrm{hPa}$, $750\,\mathrm{hPa}$ and $850\,\mathrm{hPa}$, as these belong to the core-set of CMIP5 and are therefore usually available for all models. Although in stock, the near surface pressure level $1000\,\mathrm{hPa}$ is excluded as ACCESS1-0, HadGEM2-CC and IPSL-CM5A-LR datasets are characterised here by a large amount of missing values, which cannot reasonably be interpolated. No missing values are set in any model from $500\,\mathrm{hPa}$ and above. Note that in a pre study, level $300\,\mathrm{hPa}$ and $1000\,\mathrm{hPa}$ of the reanalysis dataset are used as well (see chapter 7).

In order to generate a superimposable grid, the ESMs are remapped to match the horizontal grid spacing of the lowest spacial predictor resolution, found in the 20Cv2 dataset (see chapter 4.2.1). Remaining missing values are replaced in the same turn, if needed. Most variables are linearly interpolated, apart from relative and specific humidity, where special attention is paid to conserve the water budget (see P. W. Jones (1999) for further information).

Variables in table 4.5 are computed manually, as these are not always provided by

Table 4.5: List of predictor variables applied in the statistical downscaling framework in this thesis.

Abbreviation	Name and/or description	Units	Manually calculated
air	Pressure level based gridded temperature	K	
slp	Sea level pressure	Pa	
hgt	Geopotential height	m	
thi	Atmospheric layer thickness between pressure level 500 hPa and 850 hPa	m	x
uwnd	Zonal wind velocity	$m\,s^{-1}$	
vwnd	Meridional wind velocity	$m\,s^{-1}$	
swnd	Wind speed	$m\,s^{-1}$	x
div	Divergence (indicates an expending or converging air mass tendency)	s^{-1}	x
vor	Vorticity (indicates the rotation of an air mass)	s^{-1}	x
omega	Vertical wind velocity	$Pa\,s^{-1}$	
hur	Relative humidity	%	
hus	Specific humidity	$kg\,kg^{-1}$	
umf	Moisture flux in zonal direction	$g\,m\,kg^{-1}\,s^{-1}$	x
vmf	Moisture flux in meridional direction	$g\,m\,kg^{-1}\,s^{-1}$	x

data sources. Atmospheric layer thickness is derived from geopotential height by calculating the distance between the $500\,\mathrm{hPa}$ and $850\,\mathrm{hPa}$ layer. Wind speed w can be computed using the Pythagorean Theorem from u and v wind vectors (uwnd and vwnd) following equation 4.1.

$$w = \sqrt{u^2 + v^2} \tag{4.1}$$

Detailed description on the computations for divergence and vorticity is given by Wallace and Hobbs (2006) or Bott (2016). Divergence D can be acquired by adding the partial derivative of zonal u and meridional v wind components in their respective direction (longitude: x and latitude: y) following equation 4.2.

$$D = \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \tag{4.2}$$

On a spherical grid, equation 4.2 needs to be distance weighted, resulting in equation 4.3, where r is the average radius of the Earth, λ the longitude and ϕ the latitude (Steeneveld 2018).

$$D = \frac{1}{r \cos\phi} \left(\frac{\partial u}{\partial \lambda} + \frac{\partial (v \cos\phi)}{\partial \phi} \right) \tag{4.3}$$

Analogically, vorticity (ζ) is computed by subtracting the partial derivatives of zonal u and meridional v wind velocity differentiate in their respective orthogonal direction (y and x) following equation 4.4.

$$\zeta = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} \tag{4.4}$$

Once again with spherical adaptions one gets:

$$\zeta = \frac{1}{r \cos\phi} \left(\frac{\partial v}{\partial \lambda} - \frac{\partial (u \cos\phi)}{\partial \phi} \right) \tag{4.5}$$

Last but not least, moisture flux in zonal (umf) and meridional (vmf) direction can be estimated by the product of each particular wind vector (u and v) and specific humidity multiplied by 1000 as units need to be converted from $\mathrm{kg\,kg^{-1}}$ to $\mathrm{g\,kg^{-1}}$ (see Table 4.5).

5 Methods

5.1 Working Environment

In statistical downscaling, a large amount of computation needs to be performed. A suitable computer-assisted software needed to be found, which includes a comprehensive tool set for Artificial Neural Networks, the predominantly used methodology within this thesis. Further requirements for the software are:

- to be capable of handling large meteorological datasets with a high quantity of input variables.
- to provide basic statistical tools, needed for statistical downscaling, e.g. data preprocessing.
- to provide a comprehensive tool set for Artificial Neural Networks, e.g adjustable net-topology, various training algorithms, etc.
- to provide a comprehensive tool set of evaluation metrics for the statistical downscaling results, in particular for Artificial Neural Networks.
- to feature open source code with the possibility to make adjustments if necessary.
- to provide an interface for a combined classification and Artificial Neural Network approach.
- fast runtime.
- to run in batch mode on a compute cluster.

At the beginning of the study no comprehensive software packages fulfilled all requirements. Therefore an R (R Core Team 2017) extension under the working title "gecco" is developed within the framework of this thesis. Parts of the package are based on the cost733-classification software (Philipp et al. 2014a). The R package is available under the terms of the GNU General Public License at the Institute of Geography of the University of Augsburg. All methodologies described in this chapter can be performed within the R framework.

5.2 Standardisation and Scaling of Datasets

Standardisation improves comparability of variables operating on different scales by adjusting the mean, thus the level, and the standard deviation, the scatter, to an equal value of 0 and 1, respectively (Bahrenberg et al. 1992). Formula 5.1 illustrates the approach by standardising variable x, equation 5.2 back-transforms standardised variables.

$$z_i = \frac{x_i - \overline{x}}{s_x} \tag{5.1}$$

$$x_i = s_x z_i + \overline{x} \tag{5.2}$$

z_i: standardised sample element i of variable x.
x_i: sample element i of variable x.
$\overline{x}$: arithmetical mean of x.
s_x: standard deviation of x.

In this study, each time series is standardised separately to ensure comparability. In statistical downscaling it is further often necessary to remove the annual cycle from input and output variables. This can be accomplished for daily datasets by building groups for each day and standardise each sample separately.

Moreover, Artificial Neural Networks expect a specific value range of input and output datasets. Equation 5.3 scales variable x in range r_{min} to r_{max}, while 5.4 back-transforms scaled values. The minimum and maximum value depends on the codomain of the applied transfer functions (see section 5.5.1).

$$sc_i = r_{min} + (r_{max} - r_{min}) \frac{x_i - x_{min}}{x_{max} - x_{min}} \tag{5.3}$$

$$x_i = \frac{sc_i - r_{min}}{r_{max} - r_{min}} (x_{max} - x_{min}) + x_{min} \tag{5.4}$$

sc_i: scaled sample element i of variable x.
x_i: sample element i of variable x.
r_{max}: maximum of the resulting value range.
r_{min}: minimum of the resulting value range.
x_{max}: maximum value of the variable x.
x_{min}: minimum value of the variable x.

In general, a small spacing to the outer bounds (r_{min} and r_{max}) of the transfer functions codomain is considered as good practice. In this study a gap of 0.1 seemed to be sufficient.

Standardisation and scaling parameters cannot directly be transferred from reanalysis data to ESM output as there is usually a bias between both datasets. However, a similar mean and variance is crucial as otherwise a climate signal can occur simply from different dataset characteristics. Under conditional acceptance that the variability is correctly represented in the ESMs one can handle the differences in variable properties by standardising and scaling both datasets separately (Wilby et al. 2004). As the climate variability modes of ESMs might differ from the reanalysis, the investigated time period should be of reasonable length in order to consider the climate variability comprehensively. In this study, an equal time period including the years 1970-2000 serves as baseline to derive standardisation and scaling parameters from historical EMS runs and reanalysis data. Finally, datasets of future ESM projections are standardised and scaled using parameters from each model's historical run to preserve the climate signal.

5.3 Evaluation Metrics

5.3.1 Error Measures

A comprehensive overview considering model error and accuracy measures is given by Wilks (2011). Here, the three most common types are briefly explained.

Mean Absolute Error (MAE)
The Mean Absolute Error is computed by considering the absolute differences between model and observations:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|x_i - y_i| \tag{5.5}$$

n: number of sample elements.
x: Simulated time series or test prediction series.
y: Observed time series.

The MAE has a similar scope of application as the Root Mean Square Error (see below), but is less affected by large differences between model and observations than the latter.

Mean Square Error (MSE)

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2 \tag{5.6}$$

Like the MAE the Mean Square Error is always non-negative. A characteristic of the MSE is the non-linear behaviour of the error: Larger discrepancies between

model and observations are pronounced by the square of the equation. In addition, the MSE value is characterised by the square of the input variable units, which makes the error magnitude difficult to interpret. The MSE is one of the most popular cost-functions, when training Artificial Neural Networks in regression mode.

Root Mean Square Error (RMSE)

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2} \tag{5.7}$$

The Root Mean Square Error is represented by the square root of the MSE, thus producing an error with the same units as the original quantity. The RMSE is easier to assess than the MSE and therefore one of the most common error measures considering model accuracy in statistical downscaling. Some studies consider the RMSE as misleading, as it depends on a combination of three error characteristics: The MAE, the distribution of error magnitudes (squared errors) and the square root of the number of elements (n) (Willmott and Matsuura 2005). The main issue of the RMSE is its high sensitivity to outliers in the error distribution. It is suggested using the MAE instead of the RMSE, as the MAE follows a more linear and therefore intuitive and comprehensible behaviour. Nevertheless, Chai and Draxler (2014) point out that the RMSE is still beneficial if errors are normal distributed and suggest an equal treatment of both measures in this case. Additionally, this characteristics makes the RMSE an interesting choice if the focus of the model evaluation is to identify models with unusual large errors (Chai and Draxler 2014; Brassington 2017).

5.3.2 Correlation and Explained Variance

Correlation measures the strength of dependence between two quantities. In case of model evaluation, the relationship between simulated (x) and observed (y) values can be analysed with correlation. Two types of correlation metrics will be described in the following, text and equations are based on Wilks (2011) and Bahrenberg et al. (1992). The codomain of the correlation coefficient lies between -1 and 1, where 0 shows no dependence, -1 and 1 ideal negative or positive dependence, respectively.

Pearson Product-Moment Coefficient

The Pearson product-moment coefficient (r_{xy}) of linear correlation requires parametrical datasets, data independence within each sample and a minimum sample number of 30 elements. The Pearson correlation coefficient is calculated by the fraction of covariance ($Cov_{x,y}$) and standard deviation of both quantities (s_x and s_y):

$$r_{xy} = \frac{Cov_{x,y}}{s_x s_y} = \frac{\dfrac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{n-1}}{\sqrt{\dfrac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{(n-1)}}\sqrt{\dfrac{\sum_{i=1}^{n}(y_i - \overline{y})^2}{(n-1)}}} \tag{5.8}$$

$Cov_{x,y}$: Covariance between sample x and y.
n: number of sample elements.
$\overline{x}$: arithmetical mean of x.
$\overline{y}$: arithmetical mean of y.
s_x: standard deviation of x.
s_y: standard deviation of y.

The Pearson correlation is unfortunately neither robust, as it probably cannot detect non-linear relationships, nor resistant as it is extremely sensitive to outliers. The Pearson correlation is closely related to linear regression. It can be shown that the explained variance of a linear regression model between observations y and simulations x can be computed as the squared Pearson correlation coefficient (r_{xy}^2). The explained variance is a popular tool to evaluate linear models. It shows how much of the total variance of the observations can be reproduced with the simulated time series. An ideal model would capture the total variance $r_{xy}^2 = 1$, while models of low quality explain none ($r_{xy}^2 = 0$).

Spearman Rank Correlation Coefficient

The Spearman rank correlation coefficient (r_s) is developed for non-parametrical situations:

$$r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)} \tag{5.9}$$

n: number of sample elements.
d: absolute rank difference of x_i and y_i.

The requirements are a monotonic dependence between two quantities x and y and a sample number of at least 20 elements. In this metric a rank is assigned to each sample element of the observational and simulated time series. Afterwards the differences in ranks are evaluated. In the presents of more than $25\,\%$ ties, the Spearman correlation needs to be adjusted. Spearman correlation is more robust and resistant compared to Pearson correlation.

5.3.3 Mean Square Error Skill Score

Mean square error skill score (MSSS) is a relative accuracy measure (Wilks 2011). The MSSS compares the MSE of a modelled time series (MSE_{sim}) with a reference prediction series (MSE_{ref}). As reference prediction serves usually the climatological mean of the observational variable (Goddard et al. 2013).

Each MSE is calculated using equation 5.6, while the MSSS is computed by:

$$MSSS = \frac{MSE_{sim} - MSE_{ref}}{0 - MSE_{ref}} = 1 - \frac{MSE_{sim}}{MSE_{ref}} \tag{5.10}$$

The resulting codomain of the MSSS varies between $]-\infty, 1]$, while best simulations show a value of 1. 0 indicates no improvement and negative values a worsening compared to the reference prediction.

5.4 Frequently Applied Statistical Tests

5.4.1 Mann-Whitney U-Test

In this thesis, the Mann-Whitney U-test is usually applied to compare two samples a and b for differences in their central tendency (Mann and Whitney 1947). The U-test is designed to verify the null hypothesis, whether or not it is equally likely that a randomly selected value from sample a will be less than or greater than a randomly selected value from sample b. As the main computation (see equation 5.11) investigates the ranks of values, the U-test can be used in a non-parametrical context. The U-test requires a sample size of at least ten elements and the both samples to underlay the same distribution. In the present of ties, T can be corrected, see Wilks (2011) for further information. The p-value is computed from the test statistic T using the normal distribution.

$$T = \frac{|U - \frac{n_a n_b}{2}|}{\sqrt{\frac{n_a n_b (n_a + n_b + 1)}{12}}} \tag{5.11}$$

$$U = min\{n_a n_b + \frac{n_a(n_a+1)}{2} - R_a; n_a n_b + \frac{n_b(n_b+1)}{2} - R_b\} \tag{5.12}$$

T: test statistic.
n_a: sample size of a.
n_b: sample size of b.
R_a: sum of the ranks in sample a.
R_b: sum of the ranks in sample b.

5.4.2 Fligner-Killeen Test

The Fligner-Killeen test investigates the homogeneity of variances of two or more groups of samples. The null hypothesis suggests that the variances of the latter are the same. The Fligner-Killeen test in this study is primarily used to identify changes in variability of a climatological time series, while different time periods determine the grouping. Conover et al. (1981) point out that Fligner-Killeen is one of the most robust tests for homogeneity of variance if departures from normality are expected. There are plenty different approaches on how to preprocess the input dataset for the Fligner-Killeen test. In this thesis, the R implemented median centring versions is used (R version 3.4.4). In the computational process, a time series x is median centred by subtracting from every group element x_i, j the median of the respective group $med(x_j)$:

$$a_{i,j}^{n} = x_{i,j} - med(x_j) \tag{5.13}$$

Second, the normal distribution quantiles q of the transformed ranks ($rank$) of the absolute differences to the median a_i^m, j are derived by:

$$a_i = q\left\{\frac{1 + rank(|a_{i,j}^{n}|)}{2(N+1)}\right\} \tag{5.14}$$

Finally, the test statistic T is computed by comparing the group means $\overline{a_j}$ of the quantiles to their overall average $\overline{a}$ (see Equation 5.15):

$$T = \frac{\sum_{j=1}^{k} (n_j(\overline{a_j} - \overline{a}))^2}{s^2} \tag{5.15}$$

T: test statistic.
k: number of groups.
N: size of the overall time series.
n_j: sample size of group j.
$med(x_j)$: median of the jth group.
a: quantiles.
$\overline{a_j}$: mean quantile of the jth group.
$\overline{a}$: mean of all quantiles.
s^2: variance of all quantiles.

The p-value is computed from the test statistic T using the χ^2 distribution.

5.4.3 Test of Proportions

In this study proportions are tested for significance based on the χ^2-test. The χ^2-test table is usually imagined as a matrix containing the frequency of two different types of categorisations specified in rows and columns. The χ^2 table in this thesis is typically defined by a matrix with two different time periods as rows and the occurrence or non-occurrence of an event in columns. Therefore, the χ^2 table represents a special case with only two rows and two columns. If the null hypothesis is accepted no significant change in the occurrence of an event is detected. Test requirements include $F^e_{i,j}$ to be equal or larger than five (Bahrenberg et al. 1999). The p-value is computed from the test statistic T using the χ^2 distribution.

$$T = \sum_{i=1}^{k} \sum_{j=1}^{l} \frac{(F^o_{ij} - F^e_{ij})^2}{F^e_{ij}} \tag{5.16}$$

T: test statistic.
k: number of first categories.
l: number of second categories.
n_j: sample size of group j.
F^o: observed absolute frequency.
F^e: expected absolute frequency. Product of the respective row and column sums divided by the total sum.
s^2: variance of all the normalization values.

5.4.4 Trend-Noise Ratio

The trend-noise ratio tests the significance of a linear trend in a time series. The trend value lt is computed by the difference of the first and last value of a fitted linear regression model and s represents the standard deviation of the sample (see e.g Schönwiese (2013b)).

$$T = \frac{lt}{s} \tag{5.17}$$

T: test statistic.
lt: linear trend.
s: standard deviation.

The null hypothesis suggests that no significant trend is observed. Test requirements include parametrical time series and a linear trend behaviour. In this thesis, the trend-noise ratio is primarily used to evaluate the within sample trend as most investigated time series are not normal distributed. It is therefore not possible to draw

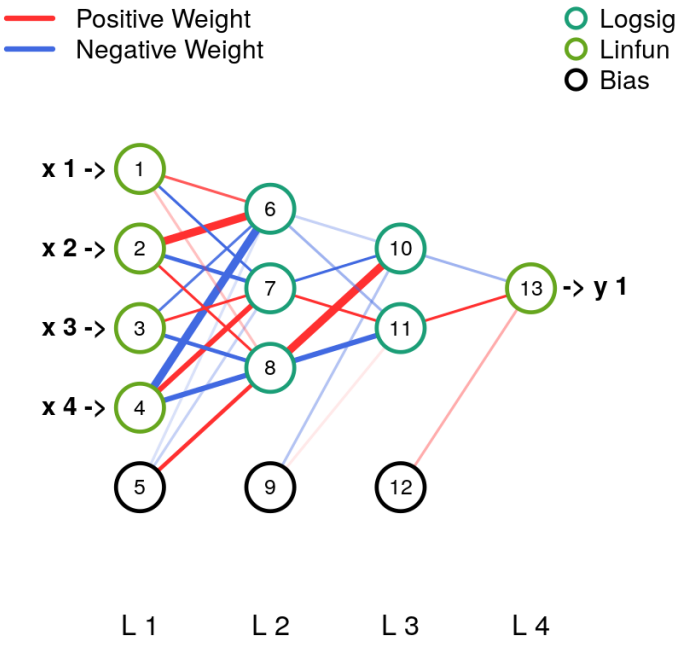


Figure 5.1: ANN topology of a fully connected feedforward network with two hidden layers.

conclusions onto the population. The p-value is computed from the test statistic T using the normal distribution.

5.5 Artificial Neural Networks (ANN)

5.5.1 Terms and Functionality

The basic principle of an Artificial Neural Network (ANN) is inspired by the biological structure of a brain, where neurons are connected via axons to store and process information. The structure of ANNs, referred to as architecture or topology, is usually illustrated in a schematic diagram similar to figure 5.1. Circles are called neurons and the lines between neurons are referred to as connections, synapses or axons. Every neuron can get information from an arbitrary number of input connections, but any neuron will only provide one output value, which can be passed on once again by an arbitrary number of connections. Information from predictor sets (here x_1 to x_4) enters the ANN via input neurons (here neuron number 1 to 4) and is processed by an user defined number of invisible hidden neurons (here neuron numbers 6 to 8, 10 and 11) organised in one or several hidden layers (here L2 and L3). Processed

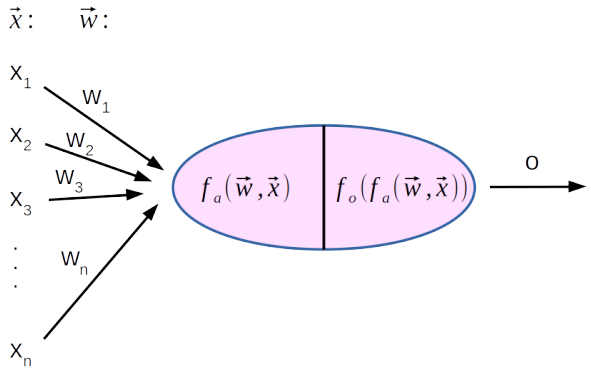


Figure 5.2: Functionality of an artificial neuron. Based on Lippe (2006).

data is extracted from the ANN at output neurons (neuron 13) of the output layer (L4). In practice, the output neurons represent e.g. the simulated predictand time series usually referred to as target variable. Special kinds of neurons are the optional bias neurons, which do not receive input. Bias neurons represent a threshold level (bias or offset), which controls the activation of connected neurons. Bias neurons are typically set constant to unity that means the bias itself will be simply defined by the optimised weight of the connection between bias neuron and linked neuron. Connections in feedforwad networks, as applied in this study, are purely organised in forward direction, beginning at the input layer and ending at an output neuron. Furthermore, synapses can skip certain layers, which are known as short cut connections.

Schematic diagram 5.2 shows the data processing at each ordinary hidden neuron (Lippe 2006). Typically, outputs of the n predecessor ($\overrightarrow{x}$) are weighted ($\overrightarrow{w}$) and summed up:

$$f_a(\overrightarrow{x}, \overrightarrow{w}) = \sum_{i=1}^{n} x_i w_i \tag{5.18}$$

The output of the sum function $f_a(\overrightarrow{x}, \overrightarrow{w})$, often referred to as the net input, is handed to the transfer or activation function $f_o(f_a(\overrightarrow{x}, \overrightarrow{w}))$. Theoretically transfer functions can be any kind of function as long as they are continuous and their derivative is known, as the latter is needed for the training process (Demuth et al. 2014). The choices of transfer function depends on the type of relationship, as only non-linear transfer functions can solve non-linear problems. Following the natural behaviour of neurons in a brain, transfer functions are typically "squashing functions" restricting the output value range of a neuron between a minimum and maximum limit. Figure 5.3 and table 5.1 compare a selection of frequently used "s"-shaped sig-

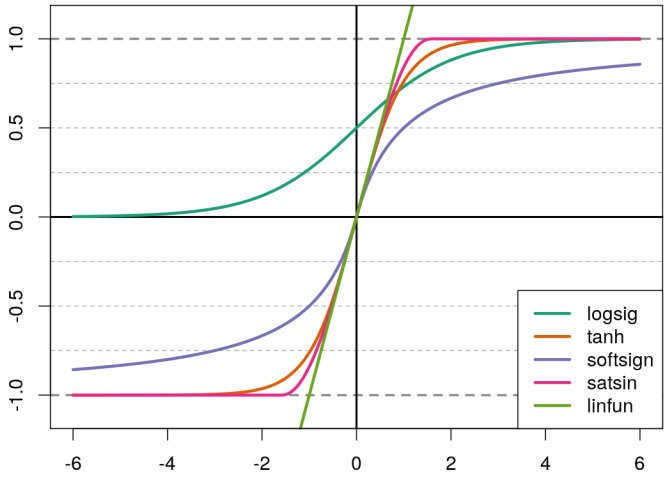


Figure 5.3: Graph of the value range of different transfer functions: Logistic transfer function (logsig), tangent hyperbolicus (tanh), SoftSign (softsign), saturating sinus (satsin), identity function (linfun). See also table 5.1.

moid transfer functions and the identity function used in regression tasks as output function. If the neuron output is close to the value limit of the transfer function this neuron is called saturated (Glorot and Bengio 2010). The output value o of a transfer function is referred to as activation of the respective neuron. Some training algorithms benefit from transfer functions, which are symmetrical to the origin. LeCun et al. (1998), for example, argue that second order training algorithms are slowed down by the logistic transfer function because of the non-zero mean. In this thesis, no second order training is performed. Therefore, the logistic transfer function can be used for hidden neurons as well. To save computational power the derivative of the transfer function, needed to train the ANN, is often a combination of the original function (see table 5.1). In regression-like approaches, the transfer function of the output neuron is the identity function. The value range between +/- infinity allowing the ANN to extrapolate and not to be stuck within bounds. Input neurons also rely on the identity function. Provided input and output datasets are scaled within the value range of hidden neuron transfer functions. This assures consistency and greatly enhances the training process (Nawi et al. 2013; Azadeh et al. 2011).

To receive a non-linear ANN, at least one hidden layer is necessary (Minsky and Papert 1969; Rumelhart et al. 1987). Following Lippe (2006), linear methods can only solve problems that are linearly separable via a hyperplane. The mathematical definition of linear separability of a n-dimensional space X into the subsets P and N is given, if a vector w_{n+1} exists with the property:

Table 5.1: Transfer functions.

Label	Function	First derivative
Identity function	$f_o(x) = x$	$f'_o(x) = 1$
Logistic transfer function	$f_o(x) = \frac{1}{1+e^{-x}}$	$f'_o(x) = f(x)(1-f(x))$
Tangent hyperbolicus	$f_o(x) = \tanh(x)$	$f'_o(x) = 1 - \tanh^2(x)$
SoftSign	$f_o(x) = \frac{x}{(1+\lvert x\rvert)}$	$f'_o(x) = \frac{1}{(1+\lvert x\rvert)^2}$
Saturating Sinus	$f_o(x) = \begin{cases} -1 & : \text{if } x < -\frac{\pi}{2} \\ sin(x) & : \text{else} \\ 1 & : \text{if } x > \frac{\pi}{2} \end{cases}$	$f'_o(x) = \begin{cases} 0 & : \text{if } x < -\frac{\pi}{2} \\ cos(x) & : \text{else} \\ 0 & : \text{if } x > \frac{\pi}{2} \end{cases}$

$$\sum_{i=1}^{n} x_i w_i = \begin{cases} \geq w_{n+1} & \forall \overrightarrow{x} = (x_1, ..., x_n) \in P \\ < w_{n+1} & \forall \overrightarrow{x} = (x_1, ..., x_n) \in N \end{cases} \tag{5.19}$$

An example often used to explain non-liner separability in the context of ANNs is the "exclusive or" or "XOR"-problem visualised in figure 5.4. The aim is to separate the binary respond values listed in the inner margins of the blue box in figure 5.4, while the axes represent the model input samples, e.g. an input of (1,1) results in a 1 in figure 5.4 a) and a 0 in figure 5.4 b). While the regular OR-problem in figure 5.4 a) can be solved by separating 0 from ones by a linear line and therefore a linear regression model, the correct answer to the XOR-problem in figure 5.4 b) can solely be given by non-linear methods like ANNs.

Different versions of ANN optimisation exist (Lippe 2006), whereby the calibration/optimisation process is commonly referred to as training or learning. The most frequently used training mechanic modifies weights and biases. Other methodologies can change the ANN topology, e.g. by adding or removing connections, neurons or layers. Some training algorithms are able to adapt themselves, e.g. by adjusting parameters to certain conditions. In this thesis, two different types of learning are used: Supervised learning trains the ANN with the aid of provided associated input and output pairs. The current ANN training state can be evaluated by comparing the simulated and desired output series. The aim of this learning type is to minimise the error between the latter. In contrast, unsupervised learning (or self-organised learning) procedures do not presume a certain output. The goal is to find patterns

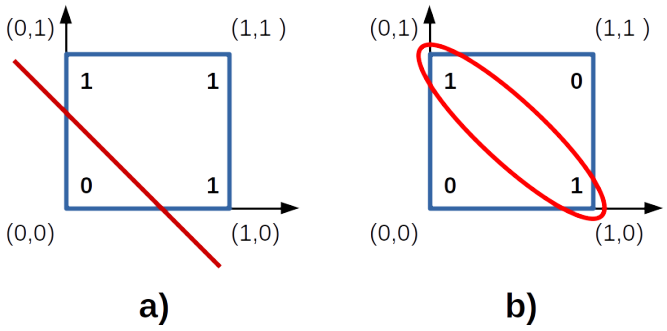


Figure 5.4: Example of a) linear and b) non-linear separability.

within the input dataset without external influence.

Learning can further be differed by the way the training dataset is presented during optimisation.

Online training, also known as stochastic or incremental training, updates the ANN after presenting one input output data pair (Demuth et al. 2014; LeCun et al. 1998). The advantage of online training is that some problems are faster to train, as ANN optimisation updates use less data. Specific application examples are models with a strong relationship between predictor and predictand or training with large datasets as well as high complex ANNs, which requires a high computational power. In addition, noisy updates reduce the risk of the training process to get stuck in local minima and can further help to prevent overfitting (see section 5.5.6). The downsides are that complex relationships are much harder to learn, as too frequent and noisy updates may cause the model error to jump with high variability. This makes it hard for the training algorithm to settle on an error minimum.

A trade-off between batch (see below) and stochastic mode is mini-batch learning, where a training dataset subsets of reasonable size are presented to the training algorithm. If all training pattern are presented to the algorithm, an "epoch" is finished.

In contrast to the previous methods batch mode or offline training considers the complete calibration period. Thus, in stochastic or mini-batch training the number of iterations differs from epochs, while in batch mode training iterations and epochs are equal. In this thesis, batch mode training is preferred, as predictor predictand relationships can be difficult to estimate (e.g. precipitation as predictand), calibration datasets are not too large and ANN topologies are usually of low complexity. In this case batch mode leads to a faster and more stable convergence of the learning pro-

cess. In addition, algorithms adapting the parameter setup to the current situation of the cost-function, like the mostly used Rprop algorithm, cannot be satisfactorily trained in stochastic mode as the surface of the cost-function varies strongly depending on the input set making speed adaptions invalid.

5.5.2 Simulation of an ANN

The basic calculations of the simulation procedure are already partially described in section 5.5.1. In the following, the sequence of the processes is summarised with respect to a fully connected feedforward ANN applied in this thesis. Discussed shall be a network with a number of H ($h = 0, ..., (H-1)$) layers (U). The simulation process is realised layerwise until the output layer is reached.

Input layer: h=0
The input layer U_0 distributes a number of n_0 input values ($\overrightarrow{x}$) into the network. The output $o_{0,i}$ of the input neuron i is directly resulting from the input values x_i, as the transfer function is the identity function:

$$o_{0,i} = x_i, \quad (i = 1, ..., n_0) \tag{5.20}$$

Bias neurons have the constant value of unity:

$$o_{0,0} = 1.0 \tag{5.21}$$

Hidden layer: h=1,..,H-2
For each neuron i of the hidden layer h the net input $f_a = A_{h,i}$ is calculated from the weighted connections of n_{h-1} linked predecessor neurons of layer U_{h-1} (see Formula 5.18).

$$A_{h,i} = \sum_{j=0}^{n_{h-1}} w_{h,i,j} o_{h-1,j}, \quad (i = 1, ..., n_{h-1}) \tag{5.22}$$

The output of hidden neuron i of layer h is derived by integrating $A_{h,i}$ selected transfer function f_o:

$$o_{h,i} = f_o(A_{h,i}), \quad (i = 1, ..., n_h) \tag{5.23}$$

Once again the bias neurons are set to unity:

$$o_{h,0} = 1.0 \tag{5.24}$$

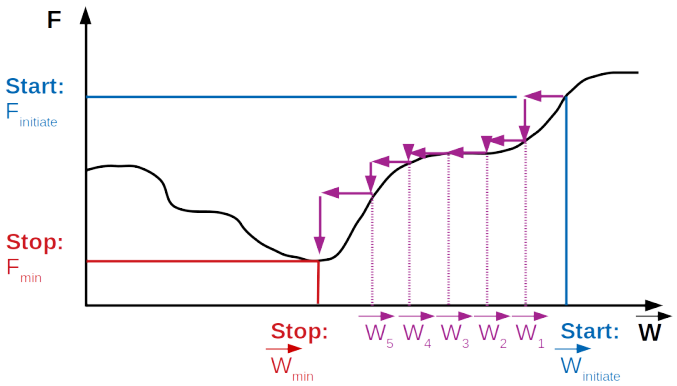


Figure 5.5: The initially unknown error curve (F) considering different weight vectors (w). The aim of gradient descent is to identify the gradient of the local position of the ANN on the error curve. The weight vector is adjusted by follow the error curve iteratively to its minimum.

Output layer: H-1

In most cases the identity transfer function is applied to output neurons $o_{H-1,i}$. Here, the calculation reduces to:

$$o_{H-1,i} = A_{H-1,i} = \sum_{j=0}^{n_{H-2}} w_{h,i,j} o_{h-1,j}, \quad (i = 1, ..., n_{H-2}) \tag{5.25}$$

Although common, it should be mentioned that the input and output layer transfer functions technically do not have to be restricted to the identity.

5.5.3 Backpropagation

Backpropagation was introduced by Werbos (1974) as a new method to optimise weights supervised in a feedforwad Neural Network and is now the most commonly used algorithm (Lippe 2006). In the procedure, the weights are adjusted by following the gradient of the error function until a minimum is reached. The procedure is therefore commonly known as "gradient descent". Figure 5.5 shows the behaviour of the error function F with respect to the weight vector $\overrightarrow{w}$. The curve of the error function F is a priori unknown, but Hecht-Nielsen (1992) proved that the gradient of F exists and is computable on the condition that the error function is continuous and can be differentiated with respect to $\overrightarrow{w}$ and further has a value range ≥ 0. The aim is to adjust the weight in order to follow the gradient into a minimum of the error function.

The Backpropagation procedure can be structured into three steps:

1. Forward pass.

2. Calculation of the current ANN error

3. Backward pass.

These three steps are repeated iteratively until a training goal is reached.

1) Forward Pass

Technically, the forward pass is equal to simulation (see section 5.5.2). The aim is to produce a simulated output series under consideration of the current training state. If no weights are defined yet, for example in the first iteration, they need to be initialised. Lippe (2006) points out that weights should never be initialised using the same value for all weights in a fully connected feedforward network. In this case, training will lead to an equal weight update for all weights between two layers. It is necessary to break this symmetry by using random initialisation procedures. Initial weights should not be too large, as those can result in high net input values leading in terms to saturated neurons with slow learning performance (see figure 5.3). Bengio (2012) recommends to generate uniformly distributed weights in the range of $[-r, r]$ following:

$$r = \sqrt{\frac{6}{con_{in} + con_{out}}} \tag{5.26}$$

for weighting the input connections of a hyperbolic tangent neuron and

$$r = \sqrt[4]{\frac{6}{con_{in} + con_{out}}} \tag{5.27}$$

for a neuron with logistic transfer function. con_{in} represents the number of incoming and con_{out} of outgoing connections of a neuron. Empirical tests showed that Equations 5.26 and 5.27 avoid saturated neurons efficiently. Further, Glorot and Bengio (2010) suggest to set bias weights initially to 0 to ensure a free development.

2) Calculation of the Current ANN Error

In case of a linear output transfer function the error can be measured by the MSE of the ANN:

$$F(\overrightarrow{w}) = \frac{1}{N} \sum_{k=1}^{N} (f(\overrightarrow{x}_k) - o_{net}(\overrightarrow{x}_k, \overrightarrow{w}))^2 \tag{5.28}$$

, while $f(\overrightarrow{x})$ is the function to generalise and $o_{net}(\overrightarrow{x}, \overrightarrow{w})$ the simulated output of the ANN. N represents the batch size of the presented training dataset. For example stochastic training would have a batch size of $N = 1$ in contrast to batch mode training where, in case of statistical downscaling, N would be the total number of observations.

The error function is also called cost or loss function. It should be mentioned that the cost function must be altered depending on the transfer functions of the output neurons. During the backward pass, the error function needs to be differentiated. Differentiating the MSE is quite simple if the transfer function of the output neurons is the identity, but can be a complex, computational expensive task considering other functions.

3) Backward Pass

Following Lippe (2006) the gradient of the error can be mathematically described as the partial derivatives of the error function F with respect to $\overrightarrow{w}$:

$$\nabla_{\overrightarrow{w}} F(\overrightarrow{w}) = \left(\frac{\partial F(\overrightarrow{w})}{\partial w_1}, ..., \frac{\partial F(\overrightarrow{w})}{\partial w_q} \right)^T \tag{5.29}$$

If the weight vector $\overrightarrow{w}$ is shifted in the opposite direction of the gradient, $-\nabla_{\overrightarrow{w}} F(\overrightarrow{w})$, the value of the cost function will be altered into the direction of the error functions minimum (see figure 5.5):

$$\overrightarrow{w}^{new} = \overrightarrow{w}^{old} - \eta \nabla_{\overrightarrow{w}} F(\overrightarrow{w}) \tag{5.30}$$

, while η (> 0) controls the step size at which the primary weight vector $\overrightarrow{w}^{old}$ is adjusted. The question now is how to handle the gradient $\nabla_{\overrightarrow{w}} F(\overrightarrow{w})$. Hecht-Nielsen (1992) shows that a single weight adjustment can be performed by:

$$\overrightarrow{w}^{new}_{h,i,j} = \overrightarrow{w}^{old}_{h,i,j} - \eta \frac{1}{N} \sum_{k=1}^{N} \delta^k_{h,i} o^k_{h-1,j} \tag{5.31}$$

The calculation of $\delta^k_{h,i}$ depends on the layer. At the output layer $h = H - 1$ the cost function, here the MSE, needs to be differentiated:

$$\delta^k_{h,i} = \frac{\partial}{\partial o_{h,i}} \frac{1}{N} \sum_{k=1}^{N} (y^k_i - o^k_{h,i})^2 = -\frac{2}{N} \sum_{k=1}^{N} (y^k_i - o^k_{h,i}) \tag{5.32}$$

,while y represents the desired output. In other layers ($h \neq H - 1$) $\delta^k_{h,i}$ can be computed as:

$$\delta^k_{h,i} = f'_o(A_{h,i}) \sum_{r=1}^{n_{h+1}} \delta^k_{h+1,r} w^k_{h+1,r,i} \tag{5.33}$$

Indices (h, i, j) are handled analogue to section 5.5.2. Equation 5.33 indicates that weights need to be adjusted backwards through the network, beginning with the output layer, as $\delta^k_{h,i}$ depends on $\delta^k_{h+1,r}$. When implementing the procedure into a computer program, it is useful to save $A_{h,i}$ or $o_{h,i}$ from the forward pass to avoid double computations as the derivatives of the transfer functions often depend on the original function (see section 5.5.2).

5.5.4 Drawbacks of Backpropagation

Following Lippe (2006), the standard gradient descent procedure suffers from various drawbacks. Most of the problems originate from an unlucky weight initialisation or a poor choice of the learning rate η. First of all, gradient descent can get stuck in a local minima of the cost function. Training should always be performed several times with different parameter setups to ensure a reasonable optimisation of the ANN. Second, global minima can be left under certain circumstances, for example, if the gradient is quite steep and the learning rate is too high. Third, the optimisation can oscillate. This happens if a deep minima with large gradients at each side of the curve causes the algorithm to jump from flank to flank. Fourth, one of major drawbacks considering training time are weak gradients. Several authors point out standard gradient descent can be highly inefficient, e.g. Demuth et al. (2014) and Wilamowski (2011) and LeCun et al. (1998). To counter these drawbacks, several adjustments of the gradient descent procedure are proposed. Some adaptions, used in this thesis are described in section 5.5.5.

5.5.5 Variants Based on the Gradient Descent Algorithm

Algorithms introduced in this section are based on heuristics. Lippe (2006) describes heuristics in Neural Network training as approximations of the original learning functions with the aim of speeding up the training process to reach convergence faster. Unfortunately, one can construct training samples, in which heuristical modifications work with either lower performance than the original algorithm or completely fails in learning the task. Nevertheless, in most cases, heuristical modifications are beneficial and lead to a significant increase in training speed while finishing with a comparable result. To simplify the description of modifications in the following, equation 5.31 will be rewritten as:

$$w_{i,j}(t+1) = w_{i,j}(t) + \Delta w_{i,j}(t) \tag{5.34}$$

, where $w_{i,j}$ is the connection weight between neuron j and i, t is the current training iteration and $\Delta w_{i,j}(t)$ the weight update at iteration t. Heuristical modifications will affect the weight update term $\Delta w_{i,j}(t)$. In addition the partial derivative of the cost function F at weight $w_{i,j}$ will be noted in its mathematical form $\frac{\partial F}{\partial w_{i,j}}$.

Gradient descent (GD)

Listed for the sake of completeness, the original gradient descent learning rule is now described as:

$$\Delta w_{i,j}(t) = -\eta \frac{\partial F(t)}{\partial w_{i,j}(t)} \tag{5.35}$$

, while the learning rate $\eta > 0$ must be specified by the user.

Momentum Term Version (MT)
Introduced by Rumelhart et al. (1986) the MT version modification increases the weight change at monotonic gradients and slows down training as soon as the sign of the gradient changes. Reason for a change in sign might be a jump over a minimum of the cost function. The strength of the MT lies in handling weak gradients with no change in sign.

$$\Delta w_{i,j}(t) = -(1-\alpha)\eta\frac{\partial F(t)}{\partial w_{i,j}(t)} + \alpha\Delta w_{i,j}(t-1) \tag{5.36}$$

Speed increase and decrease depends on the old weight update $\Delta w_{i,j}(t-1)$, which influences the current update under the constrains of factor $\alpha \in [0, 1[$ specified by the user. If $\alpha = 0$ the MT version is equal to the original GD algorithm.

Weight Decay (WD)
WD was first described by Werbos (1974). WD is an exception compared to the other algorithm described in this chapter, as it does not necessarily increase training speed. The aim of WD is to regulate the magnitude of weights, therefore belongs to the group of "regularisation" techniques. WD can greatly enhance the generalisation capability of an ANN. Two different versions of weight decay are described here. First, "L1" regularisation modifies the weight update with a constant factor $\lambda > 0$ specified by the user:

$$\Delta w_{i,j}(t) = -\eta\frac{\partial F(t)}{\partial w_{i,j}(t)} - \lambda * sgn(w_{i,j}(t)) \tag{5.37}$$

,while $sgn(w_{i,j}(t))$ is the sign function of weight $w_{i,j}(t)$. L1 regularisation is less effective against large weights but will strongly drive small weights to zero. As result only the most influential weights remain. Small weights with no noteworthy impact, but which might introduce noise into the ANN are suppressed.

In general, large weights should be avoided, as these are often linked with overfitting (see section 5.5.6). Here, "L2" regularisation comes in hand:

$$\Delta w_{i,j}(t) = -\eta\frac{\partial F(t)}{\partial w_{i,j}(t)} - \lambda w_{i,j}(t) \tag{5.38}$$

The reduction of $w_{i,j}$ depends on its current magnitude $w_{i,j}(t)$ and λ: The larger the weight, the more it will be downsized.

Resilient Backpropagation (Rprop)
Rprop represents the standard algorithm in this thesis, due to its quick convergence and its ability to reliably achieve models with high performance. Riedmiller and H. Braun (1993) argued that speed issues of gradient decent primarily result from its

dependence on the gradient. Therefore, they proposed a procedure, in which the magnitude of the weight update is independent from the gradient. However, the gradient still defines the direction of weight adjustment and increases and decreases the step size indirectly. The absolute, weight dependent update value $\Delta_{i,j}(t)$ is controlled by

$$\Delta_{i,j}(t) = \begin{cases} \eta^+ \cdot \Delta_{i,j}(t-1) & : \text{if } \frac{\partial F(t-1)}{\partial w_{i,j}(t-1)} * \frac{\partial F(t)}{\partial w_{i,j}(t)} > 0 \\ \eta^- \cdot \Delta_{i,j}(t-1) & : \text{if } \frac{\partial F(t-1)}{\partial w_{i,j}(t-1)} * \frac{\partial F(t)}{\partial w_{i,j}(t)} < 0 \\ \Delta_{i,j}(t-1) & : \text{else} \end{cases} \tag{5.39}$$

, while $0 < \eta^- < 1 < \eta^+$ are user specified values. Furthermore it is suggested to define a minimum and maximum value for $\Delta_{i,j}$ to prevent unrealistic weight magnitudes. In this thesis the standard values proposed by Riedmiller and H. Braun (1993) worked fine in all cases: $\eta^+ = 1.2$, $\eta^- = 0.5$, $\Delta_{i,j}^{max} = 50$ and $\Delta_{i,j}^{min} = 10^{-6}$. In the first iteration, $\Delta_{i,j}(1)$ needs to be initiated on a reasonable small value.

The procedure of the Rprop algorithm is determined by the gradient sign of the prior and current iteration (see equation 5.39). If the sign of the gradient did not change, $\Delta_{i,j}(t)$ is increased via η^+. The direction (sign) of the weight update $w_{i,j}(t)$ is depending on the current gradient:

$$\Delta w_{i,j}(t) = \begin{cases} -\Delta_{i,j}(t) & : \text{if } \frac{\partial F(t)}{\partial w_{i,j}(t)} > 0 \\ +\Delta_{i,j}(t) & : \text{if } \frac{\partial F(t)}{\partial w_{i,j}(t)} < 0 \\ 0 & : \text{else} \end{cases} \tag{5.40}$$

If the sign of the present gradient changed, it is assumed that the algorithm jumped over a minima of the cost function, in this case the update value is decreased by η^- and the weight update reverted. To avoid double punishment of the step size, the gradient of the current iteration is set to zero. In the next iteration, the third term of equation 5.39 will be selected.

Extended Delta-Bar-Delta (EDBD) Algorithm

The EDBD Algorithm proposed by Minai and Williams (1990) is based on the original Delta-Bar-Delta Algorithm by Jacobs (1988). The main difference consists of an additional momentum term in the extended version. EDBD leads to a similar speed increase as Rprop, but needs in total eleven parameters, which seem to be strongly dependent on the ANN configuration and learning task and moreover are rather sensitive to adjust. The time-consuming parameter search was mainly the reason for preferring the Rprop algorithm in this study. The weight update equation of the EDBD is similar to the momentum term approach (see Formula 5.36), but modifies the learning rate $\eta_{i,j}$ and the momentum value $\mu_{i,j}$ neuron and iteration dependent:

$$\Delta w_{i,j}(t) = -\eta_{i,j}(t)\frac{\partial F(t)}{\partial w_{i,j}(t)} + \mu_{i,j}(t)\Delta w_{i,j}(t-1) \tag{5.41}$$

Similar to Rprop, jumps over local minima are identified by a changing sign of the gradient. Here, not only the previous, but a decaying trace of predecessor gradient values ($\overline{\delta}_{i,j}$) is considered:

$$\delta_{i,j}(t) = \frac{\partial F(t)}{\partial w_{i,j}(t)} \tag{5.42}$$

$$\overline{\delta}_{i,j}(t) = (1-\theta)\delta_{i,j}(t) + \theta\overline{\delta}_{i,j}(t-1) \tag{5.43}$$

, where θ controls the influence of the current and old gradients.
Both the learning rate and MT are modified in a similar way:

$$\eta_{i,j}(t+1) = min[\eta_{max}, \quad \eta_{i,j}(t) + \Delta\eta_{i,j}(t)] \tag{5.44}$$

$$\mu_{i,j}(t+1) = min[\mu_{max}, \quad \mu_{i,j}(t) + \Delta\mu_{i,j}(t)] \tag{5.45}$$

$$\Delta\eta_{i,j}(t) = \begin{cases} k_l \exp\left(-\gamma_l|\overline{\delta}_{i,j}(t)|\right) & : \text{if } \overline{\delta}_{i,j}(t-1)\delta_{i,j}(t) > 0 \\ -\phi_l\Delta\eta_{i,j}(t) & : \text{if } \overline{\delta}_{i,j}(t-1)\delta_{i,j}(t) < 0 \\ 0 & : \text{else} \end{cases} \tag{5.46}$$

$$\Delta\mu_{i,j}(t) = \begin{cases} k_m \exp\left(-\gamma_m|\overline{\delta}_{i,j}(t)|\right) & : \text{if } \overline{\delta}_{i,j}(t-1)\delta_{i,j}(t) > 0 \\ -\phi_m\Delta\mu_{i,j}(t) & : \text{if } \overline{\delta}_{i,j}(t-1)\delta_{i,j}(t) < 0 \\ 0 & : \text{else} \end{cases} \tag{5.47}$$

,while the parameters θ, η_{max}, k_l,γ_l,ϕ_l,μ_{max}, k_m,γ_m and ϕ_m need to be specified by the user. One can see that if the trace of gradients and the current gradient point in the same direction, $\Delta\eta_{i,j}(t)$ and $\Delta\mu_{i,j}(t)$ are modified by a term similar to the law of exponential decay. If the gradient sign switches, both parameters are reduced by a constant ϕ_l and ϕ_m. The authors unfortunately only give examples but not an exhaustive study on how to configure the complex parameter setup.

Gradient Descent with Decaying Eta (GDDE)

The term $\Delta w_{i,j}(t)$ gives plenty of room to experiment. Some online blogs suggest reducing the learning rate during the training progress. $\eta(t)$, for example, could be updated by a factor d:

$$\eta(t) = \eta(t-1) * d \tag{5.48}$$

$$\Delta w_{i,j}(t) = -\eta(t)\frac{\partial F(t)}{\partial w_{i,j}(t)} \tag{5.49}$$

, while d ($0 < d < 1$) controls the annealing process and must be set by the user.

Individual Adaptive Learning Rate (IALR)
Some sources suggest that an adaptive learning rate might speed up the training process. In this thesis an attempt, based on the Rprop procedure was carried out. Here, a neuron and iteration dependent $\eta_{i,j}$ is derived from:

$$\eta_{i,j}(t) = \begin{cases} min[\eta_{max}, \eta^+ \cdot \eta_{i,j}(t-1)] & : \text{if } \frac{\partial F(t-1)}{\partial w_{i,j}(t-1)} * \frac{\partial F(t)}{\partial w_{i,j}(t)} > 0 \\ max[\eta_{min}, \eta^- \cdot \eta_{i,j}(t-1)] & : \text{if } \frac{\partial F(t-1)}{\partial w_{i,j}(t-1)} * \frac{\partial F(t)}{\partial w_{i,j}(t)} < 0 \\ \eta_{i,j}(t-1) & : \text{else} \end{cases} \tag{5.50}$$

, where the initial value of $\eta_{i,j}$, the alternation of the learnrate $0 < \eta^- < 1 < \eta^+$ and the maximum η_{max} and minimum η_{min} learnrates are user defined parameters. Once again, if there is no change in successive gradient signs, the learning rate is increased and the weight is update with:

$$\Delta w_{i,j}(t) = \eta_{i,j}(t)\frac{\partial F(t)}{\partial w_{i,j}(t)} \tag{5.51}$$

In contrast, if the sign changes, the last weight update is withdrawn. Double punishment of the magnitude of $\eta_{i,j}$ is avoided, by setting the gradient of the current iteration to zero.

Adaptive Learning Rate (ALR)
The ALR algorithm represents a simple implementation of a variable learning rate $\eta(t)$ based on gradient descent. Learning rate adjustments are performed by:

$$\eta(t) = \begin{cases} min[\eta_{max}, \eta(t-1) \cdot \eta^+] & : \text{if } F(t-1) > F(t) \\ max[\eta_{min}, \eta(t-1) \cdot \eta^-] & : \text{if } F(t-1) < F(t) \\ \eta(t-1) & : \text{else} \end{cases} \tag{5.52}$$

, where the initial value of η, the alternation of the learnrate $0 < \eta^- < 1 < \eta^+$ and the maximum η_{max} and minimum η_{min} learnrate are specified by the user. Weights are updated by using the gradient descent methodology (see equation 5.35). As long as the error of the cost function F decreases, learning is speed up by η^+ and slowed down by η^- if F increases. In the latter case, the current weight update is discarded.

5.5.6 Stopping the Training Process: How to Avoid Overfitting?

Theoretically, one could optimise an ANN using a certain training dataset until convergence, e.g. a minimum of the cost function, is reached and no further adjustments

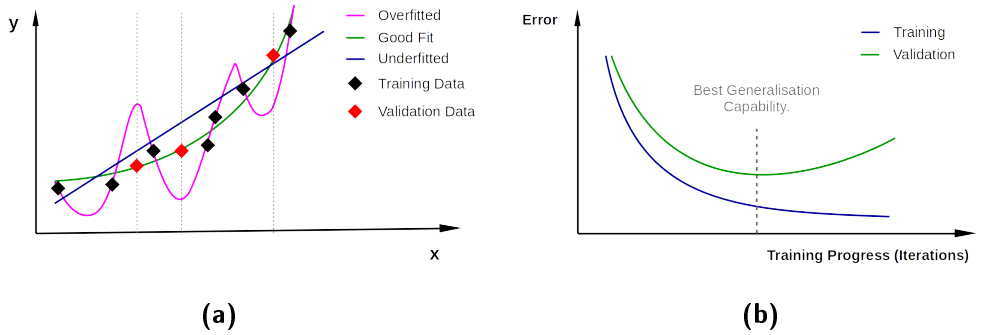


Figure 5.6: Theoretical example of model fits and how to prevent overfitting. a) different training states of a model. b) theoretical example of early stopping.

are made. The goal of training an ANN in this way is to establish a model, which approximates the training sample as well as possible. However, a model of this kind is very likely to suffer from overfitting. An overfitted model cannot be generalised, in terms, cannot be transferred onto unknown data pattern (Caruana et al. 2001; Zhang 2007). In statistical downscaling, the ANN needs to be able to deal with unknown input samples, which are slightly different from the training set: The ANN is required to have a high generalisation capability. To achieve this objective, the aim during training is to find a model, which is not affected by noise but capable of representing the main signal. Figure 5.6a visualises different model fits: The fit represented by the blue line can be considered as too simple or underfitted, the green model shows a good fit, while the purple curve displays an overfitted model. The misleading part is: the overfitted model would produce the lowest error during the training process. In contrast, when validated with slightly varied data (red dots), the overfitted model results in the largest error.

The chance to get an under- or overfitted model can be reduced by ensuring a sufficient number of calibration patterns, which ideally include all possible input-output combinations the model is applied to and by carefully choosing a suitable ANN configuration. There are additional tools which can help to prevent overfitting during the training process. In this context, Demuth et al. (2014) list two options, which may be beneficial to the generalisation capability of an ANN model. First, a procedure called "early stopping" helps to mark the point at which overfitting starts to occur during the training process. Early stopping requires a separated validation subset from the original training samples. The validation pattern must be different from the training samples and are not presented in any manner to the training algorithm. At each iteration, training and validation pattern are simulated and the respective performance is evaluated. Figure 5.6b visualises the expected behaviour during the optimisation procedure. At the beginning, both training and validation

set errors decrease, while at some point the validation error starts to increase again. It is assumed that from now on overfitting sets in. As in practice, the curve of the validation error is not as smooth as in figure 5.6b, a certain user defined amount of additional iterations, called checks, needs to be performed. The optimisation process is terminated if no further decrease of the validation error can be detected during these checks. In terms of generalisation, the optimal ANN training state is achieved at the iteration with the lowest validation error. Second "regularisation" techniques, like weight decay, reduce the risk of overfitting and can greatly enhance generalisation capability. In general, regularisation restricts the usually free development of weight magnitudes considering certain bounds. Weight decay is discussed in detail in section 5.5.5.

5.5.7 Sensitivity of ANNs to the Inputs

ANNs belong to the group of "black box" models. To the user, the internal data processing is hardly comprehensible. Sensitivity studies help to get a better insight of how the ANN handles the input-output relationship by assessing the importance of each element in the input vector on the output values. If negligible input is identified, it can be omitted to reduce the ANNs complexity. A drawback of most sensitivity analysis is that importance can only be estimated accounting to the presented input dataset.

Partial Derivatives (PaD)

One of the most popular sensitivity methods is evaluating the partial derivatives of the output with respect to the input (e.g. Kruse et al. (2016), Demuth et al. (2014), Shojaeefard et al. (2013), and Gevrey et al. (2003)). The gradient $s(i)$ of the ith input neuron x_i is computed as

$$s(i) = \frac{1}{N} \sum_{k=1}^{N} \sum_{u=1}^{n_{H-1}} \frac{\partial o_u^k}{\partial x_i^k} \tag{5.53}$$

, where N is the number of training pattern and u the uth output neuron o_u. The gradient $s(i)$ can be used to identify the affect of small changes in x_i. A strong gradient shows a high sensitivity to the input and vice versa. In addition, the sign gives hints on how the network output will respond to slight variations of x_i. For example, inputs characterised by a positive gradient will result in larger ANN outputs if the input value is slightly increased. The latter interpretation is only valid in a value range close to x_i, as in non-linear methods inputs can have different affect throughout their value range. The relative contribution of each input variable $s_r(i)$ to the output is given by

$$s_r(i) = \frac{\left|s_{(i)}\right|}{\sum_{i=1}^{n_0} |s_i|} \tag{5.54}$$

The $s_r(i)$ sums up to unity. The variable with the highest $s_r(i)$ has the greatest impact on the output calculation. In contrast to Shojaeefard et al. (2013), where the square of $s(i)$ is proposed, the absolute value is preferred here to keep the influence linear.

One-On-One Zeroing

The sensitivity of one input can also be identified by manipulating the original value. The approach suggested here zeros one by one each input of the input vector (Kaspar-Ott et al. 2019). Each time the ANN is simulated and the model error is monitored. Zeroing important inputs should result in a high simulation error in contrast to unnecessary input variables, which should barley have any impact on the error.

One-On-One Adding Noise

Similar to zeroing, noise is added one by one to each member of the input vector (Hunter et al. 2000). Once again, the simulation error is evaluated. Important inputs should result in a lower model performance. Noise is added uniformly distributed in a user defined range. The procedure is realised several times, as an unlucky choice of noise values might blur the resulting sensitivity.

5.5.8 Autoencoder and its Potential in Pre-Training Deep ANN Architectures

An autoencoder or replicator networks is a special version of an ANN (Haykin et al. 2009). Figure 5.7 shows the schematics. An autoencoder is trained by using the same dataset as predictor and predictand. The architecture of an autoencoder is typically symmetrical and the layers are linked fully connected in feed forward style. Depending on the purpose, autoencoders can have plenty of hidden layers with a reduced or increased number of hidden neurons. One main field of application of autoencoders is to reduce the dimensions of the input dataset. In this case, the input data is squashed trough a bottle neck layer (here L3) with a smaller amount of neurons. The input data is transferred through the network and should be able to reconstruct itself as perfect as possible. As the dataset is passing through layers with fewer neurons, it is assumed that all necessary information can be represented by the bottle neck layer. The progress of the input data through the autoencoder is referred to as encoding (beforehand of the bottle neck layer) and decoding (after the bottle neck layer). In a final step the decoding layers are omitted and the bottle neck

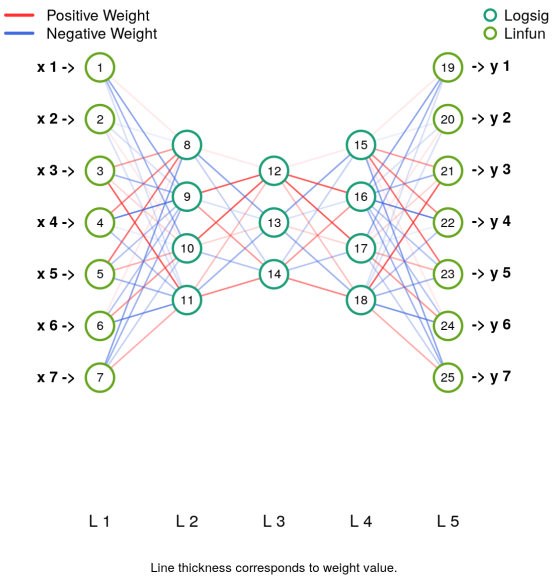


Figure 5.7: Architecture of an autoencoder. Layer L3 is the bottleneck layer.

layer is treated as the new output layer. If the training process finished successfully, the bottle neck layer should carry most of the input datasets information. Autoencoders are usually assigned to the group of unsupervised trained networks as the behaviour of each specific bottle neck neuron depends not only on internal features within the dataset but fundamentally on the initialisation of weights. The result of the autoencoder can be seen in a similar manner to extracting the first couple of principal components in a principal component analysis (PCA), which is a common technique to reduce dimensions in statistical downscaling (e.g. Hertig et al. (2014)). In comparison to the PCA, the autoencoder has the advantage that it can handle and discover non-linear relationships within and between predictors. The generalisation capability of autoencoders can be increased by a technique called denoising (Vincent et al. 2008). Denoising manipulates the training dataset with the aim to produce an autoencoder, which is robust to partial corruptions of the input dataset. The basic approach is to set a certain amount, typically between $10\,\%$ to $50\,\%$, of data points within each training step to zero.

Recently, a new topic in ANN modelling arises, which is commonly known as deep learning. Deep neural networks are currently widely used in image and speech recognition task (Goodfellow et al. 2016). ANNs with at least three hidden layers are considered as deep networks (Frochte 2019). The difficulty which comes in hand is to set and optimise an extremely high amount of weights correctly with limited available data, but without increasing the risk of over- or underfitting and to establish a stable and successful training. Problems with these requirements primarily occur

from the random initialisation of the weights. A new training technique proposed by Hinton et al. (2006), called greedy layer-wise pre-training, seems to have solved the issue and has lead to a significant increase in ANN and deep learning popularity in the last couple of years. In pre-training, each connection weight of the deep Neural Network is pre-trained after initialisation using an autoencoder. Layer weights are pre-trained successively starting from the input layer. After training of the first weight section, the bottle neck layer of the autoencoders is used as input for the next layer. When all weights are pre-trained, the final step is to fine-tune the ANN using common supervised training.

5.6 Weather Type Classifications

5.6.1 Classification Terms, Approaches and Implementation

Classification describes a method to separate data into groups (classes) under consideration of similarity between individual pattern as well as their differences (Wilks 2011). A comprehensive overview considering various classification methods in a climatological context is given by Philipp et al. (2014b). In this study, two non-hierarchical cluster algorithms are selected to construct the downscaling framework: Differentially initialised k-means clustering algorithm (DKM) and simulated annealing and diversified randomization (SANDRA).

Classification methodologies categorise observations into a number of predefined classes. The result is a catalogue with either a specific class for each observation or a membership function, which gives an estimation of the most likely class (Bardossy et al. 1995). DKM and SANDRA explicitly assign classes for each observation. The term centroid refers to the average predictor condition within each class. During simulation each sample of the training pattern is compared to every centroid and receive the best matching class. Similarity between a data point x and a centroid y is measured according to the euclidean distance (see Equation 5.55). In this regard, the dimension K represent the total number of input grid cells of the predictor at a single time step (Wilks 2011).

$$\parallel x - y \parallel = \parallel y - x \parallel = \sqrt{\sum_{i=1}^{K} (x_i - y_i)^2} \tag{5.55}$$

Some classification methods can be conditioned on a specific purpose. In statistical downscaling the separation in classes can be optimised by including the predictand in the calibration process (Lutz et al. 2012). Similar to ANNs, conditioning is prone to overfitting. The influence of the predictand should only guide the calibration process but not destroy the signal of the atmospheric circulation. Overfitting is controlled by

carefully choosing a suitable weight for the predictand. If the weight is too large, calibrating will suffer from overfitting, if the weight is too small predictand independent features might dominate the grouping of classes.

In statistical downscaling approaches, several atmospheric variables serve often simultaneously as predictor set. In his study, predictors with different domain sizes are weighted by the number of grid cells to ensure equal consideration during the optimisation. Otherwise, a large domain might annihilate the influence of a smaller one. Philipp et al. (2014a) suggest to weight each variable x by:

$$x_w = x\frac{\sqrt{w * N}}{N} \tag{5.56}$$

N: Number of elements of variable x at a time step.
w: Weight of the variable.

This weighting method is designed for euclidean distance, by accounting for the square (see Equation 5.55). If w is set to 1, all variables are treated equally important during calibration. Furthermore, areal distortions due to latitudinal effects on the sphere of the Earth can be adjusted. In this thesis, relatively small domain sizes at similar locations are considered, therefore, latitudinal weighting is neglected.

To derive simulated time series, reference values for classes are generated from the predictand after calibration in a similar way to centroids. A reference value is computed as the predictand average of each class. In the further, classifications with reference values for prediction are called reference class forecasts (RCF), while classification itself will be referred to as circulation-type classification (CTC). Note that a CTC can still be optimised conditioned.

5.6.2 Differentially Initialised K-Means Clustering Algorithm (DKM)

DKM represents an algorithm for non-hierarchical cluster analysis based on k-means clustering. The major difference to k-means is the initialisation of centroids. Introduced by Hartigan and Wong (1979), the standard k-means optimisation starts by selecting time steps randomly and combining them into classes. In an iterative process (1) centroids are computed, (2) each time step is checked for right assignment by comparing to all centroids and (3) is rearranged if not assigned to the most similar cluster centroid. Depending on the selection of the initial partitioning, the k-means cluster algorithm might get stuck in a local optima of the minimization function, which tries to reduce the within cluster variance. The resulting class assignment varies heavily between different runs. In contrast to the original k-means, DKM defines initial centroids by finding the most dissimilar pattern and allocates other observations to the most conform one (Philipp et al. 2014a; Enke and Spekat 1997).

The iterative optimisation of checking the right class assignment (1 to 3) is equal to the original k-means algorithm. The procedure does not depend on random effects, thus the resulting class assignment between different runs with identical predictors is equal. DKM was found to produce good classification results by low computational cost (Philipp et al. 2014b). Disadvantage of the DKM is that even if the results show good classification skill, it might be impossible to achieve the "perfect" grouping due to the fixed selection of starting partitions.

5.6.3 Simulated Annealing and Diversified Randomization (SANDRA)

Similar to the DKM, SANDRA is a non-hierarchical cluster analysis algorithm. Philipp et al. (2007) point out that unlike the k-means cluster analysis, the SANDRA algorithm is able to leave local minima of the cost function and has a high chance of resulting in a global optimum of the latter. As cost function the "within cluster sum of squares of deviations" is used to measure the degree of dissimilarity within the resulting clusters. SANDRA tries to avoid local minima by allowing a pattern to leave its cluster at any stage. Increases of the cost function, which often happen initially when leaving a local minimum, are tolerated with a certain probability. During the progress of the optimisation, the probability of wrong assignments is decreased until the chance is zero. The reduction is inter alia dependent on a "cooling" parameter, which needs to be defined by the user. The classification procedure is usually performed several times from different starting partitions allowing the usage of faster cooling rates to reduce the overall run time. The final classification product is selected as the one with the best cost-function result. If done conscientiously, slow cooling rates need to be chosen. This makes the SANDRA on the one hand a comparable slow clustering algorithm, but on the other hand, SANDRA is one of the few clustering methods, which are able to find the global optimum of the cost-function with a high probability.

5.7 A combined approach of Circulation Type Classification and Artificial Neural Networks (NNC)

In this thesis a novel approach of combined ANN and CTC usage is proposed, which will be referred to as NNC (Neural Network and Classification) in the following. The procedure is visualised in figure 5.8. First, the input dataset is separated into classes by a conditioned CTC. In a second step, ANNs calibrated particularly for each class simulate the respective output of the selected data points. In a third step, the ANN results are merged into one successive time series, the final product.

The advantage of conditioned pre-classification is that the successive ANNs can

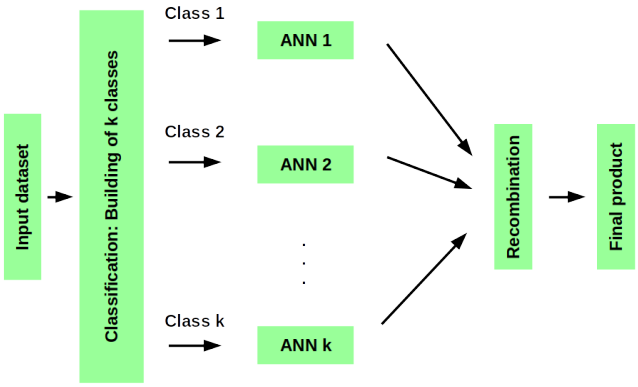


Figure 5.8: Schematics of the data flow in NNCs.

be specifically designed for a certain part of the predictand value distribution, which is objectively separated by criteria of similarity during the classification process. The downside of this methodology is the increased model complexity, as an appropriate number of classes needs to be defined, each with a suitable ANN architecture. Further, the overall procedure needs significantly more run time during training and simulation. In addition, the available data amount for calibrating the ANNs is reduced. The number of class members is determined by the classification algorithm, on which the user has hardly any influence. Therefore, one needs to check for classes with disproportional low number of elements to ensure a stable ANN calibration process.

6

Analysis of the Predictor Datasets

In the following the multi model mean (MMM) of the applied ESMs is analysed. The focus lies on the comparison between reanalysis and historical ESM datasets as well as changing conditions in the ESM results of future scenarios. ESM versions are listed in table 4.3 and variables are described in table 4.5. Evaluated is a domain ranging from $30°$ to $70°$ northern latitude and $-30°$ to $50°$ longitude. Size and location is chosen to cover the greater European area as well as near surroundings with the Alpine region in the centre. Predictors investigated must be included in at least one of the final statistical downscaling models (see section 7.5), while their most frequently picked level is selected as representative. The MMMs are computed as temporal mean of the reference period 1971-2000 (historical) as well as for the near (NF, 2021-2050) and distant (DF, 2071:2100) future of scenarios RCP 4.5 and RCP 8.5. Changing conditions are evaluated by the absolute difference between future periods and the historical experiment.

Historical MMMs and reanalysis patterns are displayed in figure 6.1. The patterns themselves match quite well. Note that biases in the value magnitudes between historical models and reanalysis datasets have no strong impact on the statistical downscaling results as both will be standardised separately in the modelling process. Atmospheric layer thickness (thi) and air temperature (air) show similar patterns with an almost latitudinal parallel gradient from high magnitudes in the south to lower ones in the north. The water vapour saturation of relative humidity (hur) in the Alps is lower than in the surrounding areas to the north and south. Moisture mainly originates from the seas, like the Atlantic ocean to the west or the Mediterranean sea in the south. Although in the southern border of the domain, low relative humidity levels are found, whereas the moisture content of specific humidity (hus) is high. This effect is due to the non-linear physical relationship between temperature and water vapour, as the warmer air masses in the south can carry a disproportional higher amount of water vapour by reduced saturation. High moisture content is usually found over large evaporation surfaces like the Mediterranean sea and the Atlantic ocean in the southern parts of the domain. Eastward wind speed (uwnd) in the frontal zone is characterised by a distinct band of westerlies. Highest uwnd speeds are found over the Atlantic ocean in the west with a leisurely decrease to the east

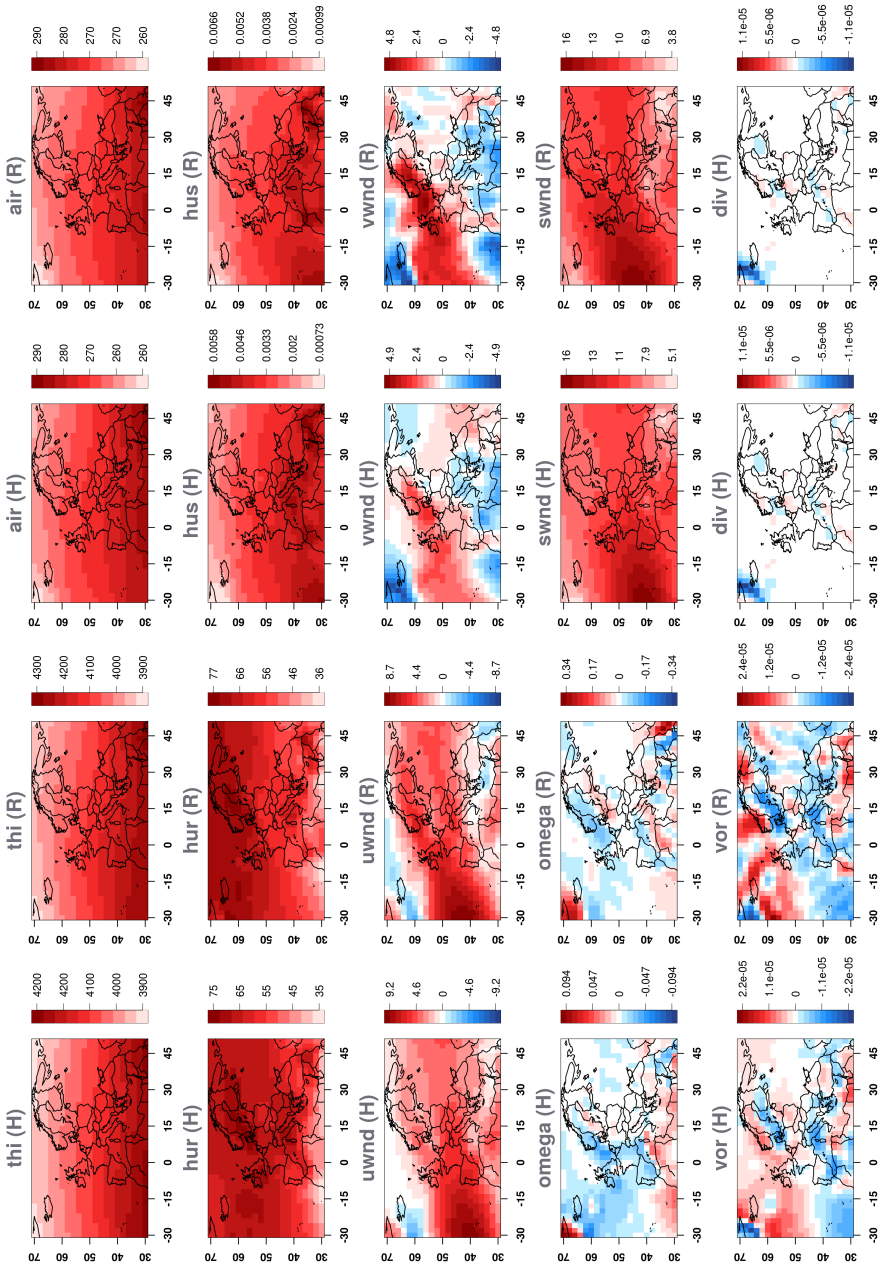


Figure 6.1: Climatological means (1971-2000) of applied predictor variables for the 20th century reanalysis dataset (R) and the multi model mean (MMM) of investigated ESMs (H). Predictor abbreviations are described in table 4.5. Only the most frequent level in the final predictor selection is analysed as representative (see table 7.4). 850 hPa: air, hus, uwnd, vwnd, vor, div; 700 hPa: hur, omega, swnd.

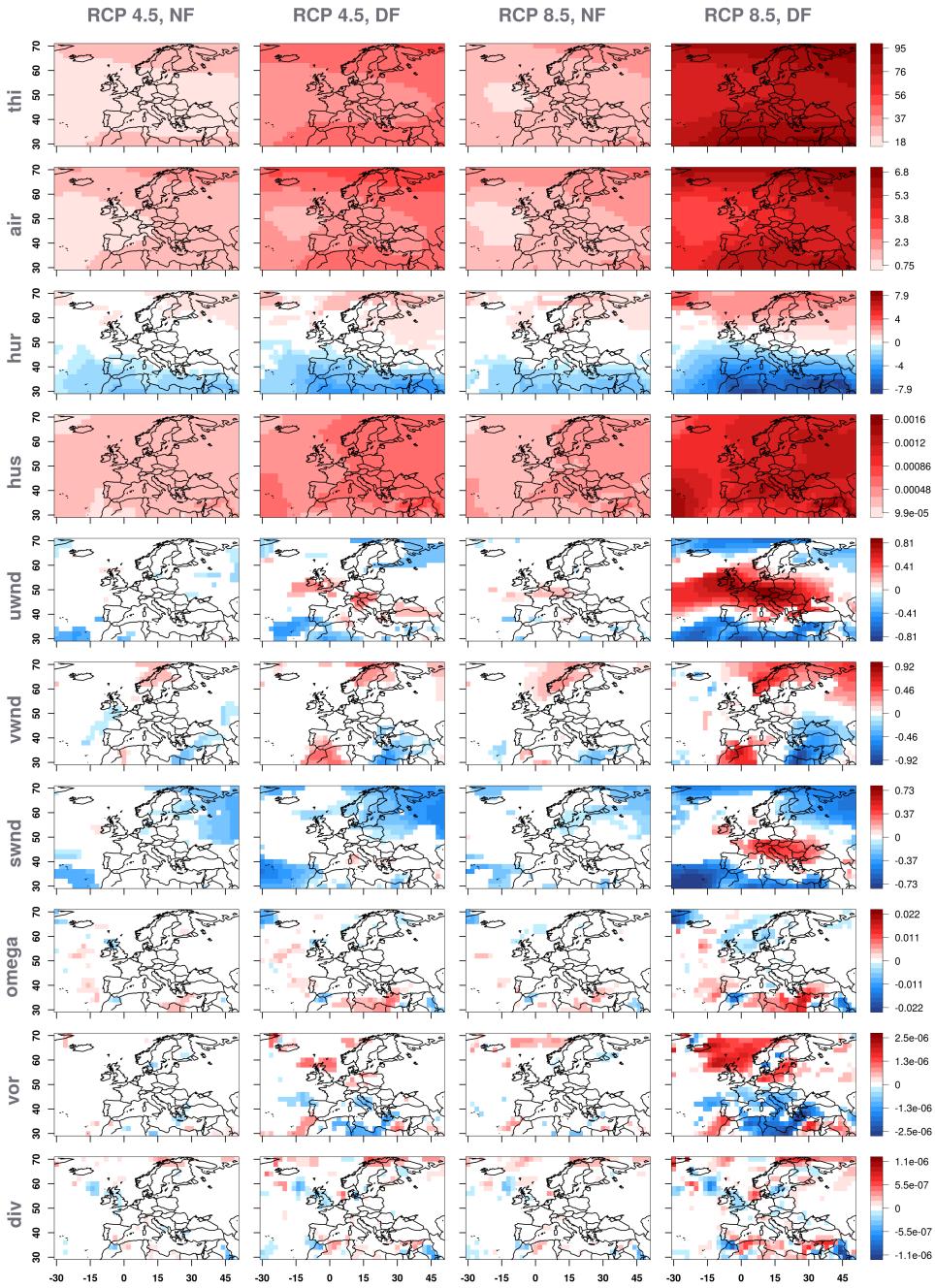


Figure 6.2: Absolute change in the multi model mean computed by subtracting the long-term temporal averages of the historical reference period (1971-2000) from the near (NF, 2021-2050) and distant future (DF, 2071-2100) of scenarios RCP 4.5 and RCP 8.5. Displayed predictors are applied in at least one statistical downscaling model, abbreviations are described in chapter 4. The most frequent level in the final predictor selection is analysed as representative (see table 7.4). 850 hPa: air, hus, uwnd, vwnd, vor, div; 700 hPa: hur, omega, swnd. Differences in the mean are statistically tested with an U-test (Significance level of 5 %. Only significant changes are shown.)

resulting from friction over the land surface. The Alps are located within the centre of this high speed uwnd band. In the northward wind (vwnd) pattern, the Alps lie in between a positive pattern to the north west and a negative one to the south east. General wind speed (swnd) is highly influenced by the uwnd pattern, again the Alps are found within a band of highest wind speed over the land surface. The omega field of vertical air mass motion shows downward tendencies in the south, as in this region the subtropical highs are located. Especially, the Atlantic parts in the west are characterised by upward motion, while over the land masses, vertical motion on average is close to zero. The Alps are located in the latter low omega area with light upward tendencies in the western parts of the Alps. In the southern region of the domain, the relative vorticity (vor) over seas is negative, while in the northern seas positive values dominate. An exception is the Baltic Sea where negative vorticity values are found. The pattern of historical runs is more homogeneous than of the reanalysis dataset especially in the northern and southern regions. The Alps lie in between a positive pattern to the north and a negative one in the south. Divergence (div) magnitudes on average are fairly low in the domain. To the north of the Alps, a slightly positive pattern is indicated, while to the south, a slightly negative one is detected.

Absolute changes in future ESM scenarios are illustrated in figure 6.2. The future development of atmospheric layer thickness and air temperature are characterised by similar patterns. Without exception, both variables show increasing magnitudes in the evaluated scenarios and time periods. In the centre of the domain, a west-east band with lower increases is simulated. Values in both scenarios are successively rising from NF to DF, while the DF of RCP 8.5 results in the highest increase. Changes in relative humidity show an almost latitudinal parallel gradient from south to north, while in the south, decreasing tendencies are simulated. In the northern parts of the domain, relative humidity is suggested to rise. In contrast, specific humidity is characterised by an overall increase, which follows the near surface trend in the last third of the 20th century (Willett et al. 2008). The higher water vapour content indicated by specific humidity is directly related to the expected warmer troposphere (IPCC 2013), while differences in relative humidity are more complex. Near the surface, relative humidity increases over oceans and decreases over land, due to the lack of sufficient water supply (O'Gorman and Muller 2010; Joshi et al. 2008). Sherwood et al. (2010) point out that relative humidity in higher tropospheric levels is rather decoupled from low level features, but strongly controlled by dynamical fields. It is further suggested that an upward and poleward expansion of the original distribution of relative humidity occurs resulting from a general northward shift of the climate zones circulation features. Amplified influence of the subtropical highs in the adjoining northern parts explains the decrease in relative humidity in the south-

ern regions of the domain. The rising tendencies of relative humidity magnitudes in the northern parts of the domain can be related to the stabilization of the lowermost atmosphere, which suppresses ventilation (Ruosteenoja and Räisänen 2013). The wind speed of westerlies is simulated to speed up over central Europe in the DF and NF of RCP 8.5. The Alpine region is usually affected by the strongest increase in the investigated scenarios and time periods. The change pattern of faster wind speeds is located slightly northward to the original high speed wind averages, which is once again most likely caused by the northward shift of the circulation (Woollings 2008; Barnes and Polvani 2013). The meridional wind speed directly over the Alps is stable in the future projections. Vwnd changes are primarily located at the border of the domain in the northern and southern parts. Accordingly, the main influence on swnd changes originates from uwnd, what leads in the DF of RCP 8.5 to a significant increase of swnd within the region of the Alps. Considering the other evaluated periods, swnd conditions over the Alps are usually stable. Omega does not show pronounced change signals over the Alpine region throughout the future scenarios. Stronger change signals are found to be mostly in the northern and southern parts of the domain. Considering future development of relative vorticity, the historical pattern is usually intensified in both negative and positive features. Divergence changes are rather low in magnitudes and are highly fragmented. No larger homogeneous change pattern are identified near the Alps.

Further analysed are the standardised change signals of the ESMs. Data processing is performed in the same manner as when applied to the statistical downscaling models (see section 5.2). The resulting pattern of standardised absolute differences (not shown here) are very similar to the results in figure 6.2. The benefit from standardised changes signals is that they can be compared in magnitudes. In the following, the focus lies on the highest changes within each predictor pattern. Large differences are found to be in air temperature, atmospheric layer thickness and specific humidity with local increases of up to two. Relative humidity, especially the reductions in the south, reaches magnitudes of -0.4. Lowest change signals are detected in the three wind velocity variables, as well as omega, relative vorticity and divergence. Considering the latter, change signals typical vary in equal magnitudes between approximately -0.15 and 0.15.

7

Development of Statistical Downscaling Models for Zugspitze and Sonnblick

7.1 Statistical Versus Dynamical Downscaling

Downscaling techniques help to increase the sparse spacial resolution of ESM outputs by generating regional-scale or local-scale information. In regions with heterogeneous relief, like the European Alps, ESM grid cells of approximately $110\,\mathrm{km}$ on average, containing the mean weather conditions, carry barely any regional or local-scale information. In addition, extreme values are often misrepresented in ESM output (Gutmann et al. 2012). To bridge the gap, dynamical and statistical downscaling methods are developed.

Dynamical downscaling (DD) is primarily based on physical equations and performed by regional climate models (RCMs). RCMs are nested into, for example, ESM results, using the latter as boundary conditions. The aim is to recalculate a limited area of interest with higher spacial and eventually temporal resolution usually under consideration of enhanced physics. Generally, there is no feedback from the RCM to the climate model (R. Jones et al. 1995). RCMs provide a significant added value compared to the driving GCMs, when it comes to describing mesoscale circulation or orographic effects e.g. in the European Alps (Buonomo et al. 2007; Frei et al. 2006; J. H. Christensen and O. B. Christensen 2007). RCMs increase the spacial resolution, but in the end, the resulting datasets are still grid cell based sparely lower than $12.5\,\mathrm{km}$ (Jacob et al. 2014).

In statistical downscaling (SD), a mathematical transfer function is established between large and local-scaled datasets. It is the responsibility of the user to ensure a consistent physical based relationship between predictor and predictand variables. Benestad et al. (2008) discuss several aspects, which makes SD interesting for climate impact studies and an useful alternative for dynamical downscaling. First of all, SD methods can be developed to derive local weather conditions. RCMs may still have a too low spatial resolution for some impact studies, e.g. analysing local weather extremes. Although RCM results can be adapted by applying further (statistical) downscaling techniques like MOS. Moreover, the possibility to costume-build

SD models for specific variables at certain locations makes SD models interesting for end users. The most time consuming part in SD is to find and calibrate a suitable model setup. Once calibrated, input datasets from different ESMs can be applied quickly, a property which comes in handy for example, when assessing ESM uncertainties. In contrast, RCMs need to be recalculated for every nesting experiment, a time consuming and computational expensive process. Last but not least, statistical methods are well suited to find relationships between climatic variables, and can hence improve the understanding of the climate system.

7.2 Statistical Downscaling Approaches

Statistical downscaling approaches are classified in various ways. Wilby and Wigley (1997), for example, distinguish methodologies, while Rummukainen (1997) suggested a categorisation based on the nature of the chosen predictors. Maraun et al. (2010) integrates these suggestions into three main approaches:

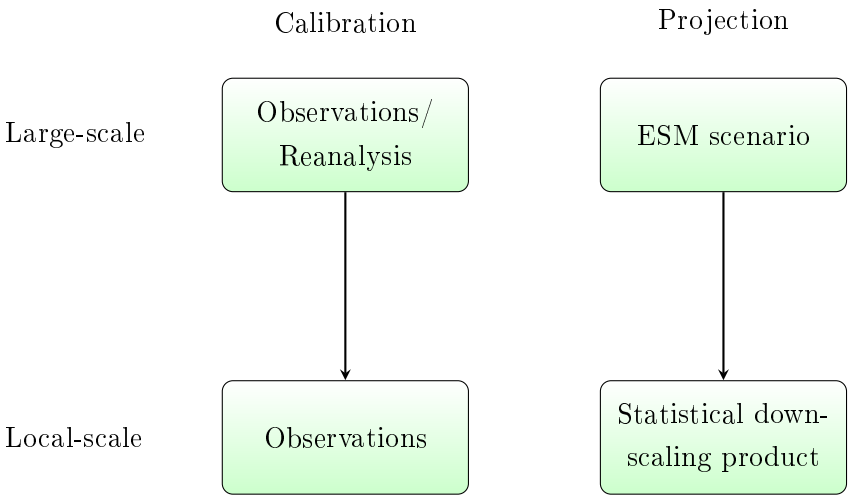
- Perfect prognosis (PP) establishes a direct link between large-scale datasets and local-scale observations.
- Model output statistics (MOS) establish a statistical relationship between RCM output and local-scale observations. MOS combines the benefits of both dynamical and statistical downscaling.
- Weather generators (WG), which are either conditioned on large-scale weather (similar to PP) or driven with changing factors derived from RCMs (similar to MOS).

In this thesis, the statistical downscaling framework is based on perfect prognosis. Figure 7.1 illustrates the schematics of PP. First, SD models are calibrated on large and local-scale observations. Recently, SD models are rather optimised with reanalysis data than observational data as already discussed in chapter 4.2.1. The SD model is now ready to downscale ESM datasets in step two.

Maraun et al. (2010) classify PP-suitable SD methodologies into linear models, generalised linear and additive models, vector generalised linear models, non-linear regression (here basically ANNs), weather/circulation type–based downscaling and Analog Methods.

In this study, non-linear statistical downscaling approaches from the group of machine learning techniques are applied, which can take non-linear predictor-predictand relationships into account. Improvements by non-linear methods are, for example, found by Huth et al. (2008b) considering the resulting shape of the statistical distribution of daily temperature series in Europe. Davy et al. (2010) suggest that the

Figure 7.1: Schematic diagram of the PP approach. Models are calibrated with observational-like datasets (left side). Afterwards climate scenarios can be projected (right side). Figure based on Maraun et al. (2010).



variability of wind can only be captured by non-linear methods. Maraun et al. (2010) and Harpham and Wilby (2005) emphasise the non-linear behaviour of precipitation. Nevertheless, non-linear methods still represent a side note in statistical downscaling, most likely due to the challenging tuning process and difficult interpretation of the optimised model structure. This study should contribute to a better understanding of non-linear models, especially ANNs.

7.3 Introduction in the Statistical Downscaling Framework

The focus of the statistical downscaling framework of this thesis lies on the possibilities and characteristics of ANNs and to evaluate their pros and cons. Additionally, it is tested if the combined usage in the NNC approach leads to an added value. The RCF methodology is primarily employed for comparative purposes. One of the greatest disadvantages of ANNs is the period of time needed to successfully finish training. Optimisation from unlucky initial conditions or bad network configurations may take a very long run time, further depending on the choice of the training algorithm or simply if the ANN exceeds a certain size, e.g. has to handle many input or output variables, hidden layers or hidden neurons. Although the algorithms in gecco are optimised for short run time, it sometimes takes days or even weeks to calibrate ANNs in certain complex configurations. Consequently, it is often not possible to perform exhaustive checks of different ANN configurations or predictor setups in reasonable time. Still, the aim is to examine each task as comprehensively as possible.

SD models in this chapter are calibrated considering daily predictor and predictand datasets. The model calibration period is defined from year 1970 to 2000 based on a running calibration procedure introduced in chapter 8. Variables and detailed description can be found in sections 4.2 and 4.1. Meteorological seasons are treated separately: SD models are build for winter (December, January, February; DJF), spring (March, April, May; MAM), summer (June, July,August; JJA) and autumn (September, October, November; SON). The split in seasons is necessary in extratropical climate because variables can have different impact throughout the annual cycle. For example, cloudless sky in summer leads to increased temperature values, as the days are long and the solar beam ankle is large, resulting in an energy balance dominated by insolation. In contrast, during winter, cloudless sky decreases temperature values, as days are short and the solar beam ankle is small and emission is the predominant factor in the energy balance. Thus, predictor variables closely related to cloudiness can affect the predictand temperature in opposite ways during the year. Another example is the moisture advection. Evaporation as well as the amount of water vapour an air mass can absorb, are highly related to temperature. Predictor variables only containing information on air flows can have a different effect on precipitation sums in winter and summer, as the quantity of advected water vapour may change. Such difficulties can often be prevented by ensuring a complete and physically consistent predictor-predictand relationship. Beside seasonal depending effects of the predictor on the predictand, varying genesis of the predictand variable may need a different predictor composition. Summer precipitation, for example, in the larger study region mainly emerges from convective processes, while winter precipitation primarily originates from frontal activities. A different predictor setup for each season regarding the selection of predictor variables, respective pressure level or an adjusted domain location/size may enhance model quality.

7.4 Generalisation of the Statistical Downscaling Models

In climate studies evaluating the transferability of models onto unknown data points is crucial. Beside the characteristics of the ANNs and conditioned classifications to be prone to overfitting, Storch et al. (2000) list requirements of the predictor variables to ensure a good generalisation capability of statistical downscaling models. First, predictor variables must be of relevance for the predictand. An extensive predictor screening procedure helps to exclude unnecessary input (see section 7.5). Second, the transfer function of the SD model must be valid under changing climatic conditions. The involved risk can be reduced, if the calibration period includes a wide range of variations, ideally all future conditions. Third, the predictand relevant climate change signal must be fully represented within the set of predictors (see section 7.5).

Last but not least, predictors must be well modelled by the ESM. Comparing ESM output to reanalysis or observational data ensures the dataset quality. Benestad et al. (2008) further suggest considering the "skilful scale" of ESMs. The skilful scale refers to the smallest spatial scale on which an ESM adequately represents a large-scale variable. The domain of a SD-predictor should not be smaller than the skilful scale. The skilful scale is based on findings from Grotch and MacCracken (1991), who pointed out that even if the mean of a large domain agrees perfectly between different climate models, large point-wise differences can exist. In application, a single grid cell should not be used as predictor variable in SD.

The generalisation skill of a fitted model can be evaluated by validating the model with new, unknown data pattern separated before calibration. Arlot and Celisse (2010) give an overview over validation techniques. The simplest method is called holdout, in which a single split into a calibration and validation period is performed. To estimate the dependence of the validation performance on the split, holdout is usually executed several times: The methodology is known under the term "cross-validation". Exhaustive cross-validation techniques (leave-one-out, leave-p-out) exclude one or a selection of p data pattern until every possible combination is evaluated, which can be a very time consuming process. Non-exhaustive cross-validation methods are generally faster. Many different approaches exist (e.g. k-fold cross-validation, balanced incomplete cross-validation, repeated learning-testing, Monte-Carlo cross-validation). In this thesis, two cross-validation methodologies, k-fold cross-validation (Picard and Cook 1984) and Monte-Carlo cross-validation (Geisser 1975), are used. K-fold cross-validation separates the committed dataset into k partitions. Models are calibrated k-times and at each iteration, a different partition represents the validation period, others the calibration pattern. Monte-Carlo cross-validation picks a percentage of validation values randomly, other time steps are considered as calibration pattern. As disadvantage can be seen that Monte-Carlo cross-validation allows the same split to be chosen several times, in case of occurrence negatively influencing the validation risk estimation.

To summarise: After the optimisation process, the main focus lies on how well the SD model performs on a validation dataset. Additionally, the performance differences between the calibration and validation datasets should not be large, as this as well indicates limited transferability. To lessen the effect of special cases, e.g. poor weight initialisation during ANN training or disadvantageous training or validation data, cross-validation procedures are applied.

Nevertheless, errors from extrapolation onto further unknown data pattern cannot be estimated in advance. The behaviour of non-linear methods in particular might be uncontrollable outside of the calibration data bounds. In contrast to ANNs and NNCs, RCFs have a fixed minimum and maximum reference value, which makes

this method unreliable when a strong shift in the value distribution is expected. For that reason, projected output needs to be checked carefully for irregularities. Considering extrapolation, linear methods seem superior, assuming that the linear relationship persists beyond the data limits of the calibration period.

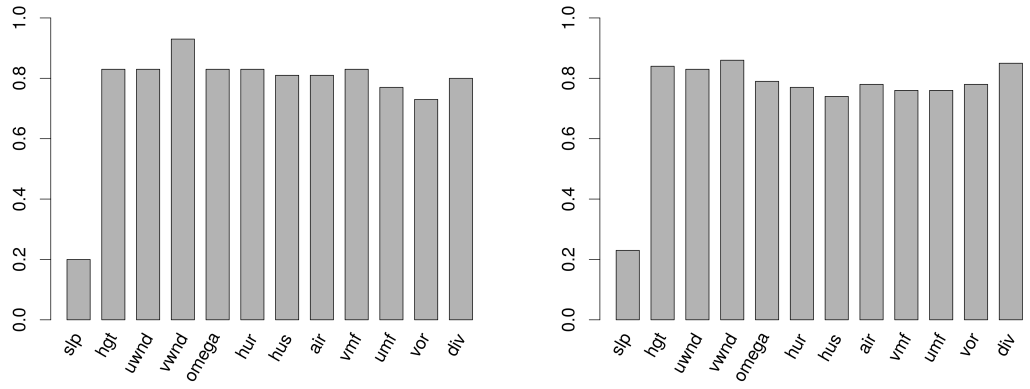
7.5 Predictor Screening

7.5.1 The Aim of Predictor Screening

One of the most important aspects in SD is to find a predictor setup with high explanatory degree. Questions hereby involve the best suited combination of different predictor variables as well as size and location of the most influential region in each predictor grid usually referred to as domain. The aim of predictor screening is to reduce the number of potential predictors to a small, but still sufficient set (Francis 2006). The Ozzacam's razor theorem, a famous philosophical concept in statistical modelling, suggests to always select the simplest model if several are carrying equal explanatory degree (Blumer et al. 1987). Models with reduced parameter setups are not only faster to optimise but can moreover be better in the task of generalisation. To perform climate impact analysis, the simple predictor combination must still include variables, which carry the climate signal (e.g. Maraun et al. (2017), Benestad et al. (2008), and Huth et al. (2014)). Further, it is necessary to find a good size and location for the domain (e.g. Beck et al. (2013) and Sauter and Venema (2011)).

7.5.2 Case Study: Comparison of Predictor Screening Approaches

Predictor screening techniques for ANNs in climatological context have not yet been studied very well in current literature. In the following, four different approaches are examined and compared considering ANN models with the intention to reproduce winter and summer precipitation at station Zugspitze. During predictor screening, the number of neurons is treated as of secondary importance. ANNs seem to perform well with eight hidden neurons for DJF and six for JJA in combination with the logistic transfer function. Rprop is chosen as training algorithm. This chapter summarises a preliminary study to provide an overview over advantages and difficulties one has to face to find a good predictor setup with/for ANNs. Thus, predictor configurations of this section are not included in the final downscaling result presented in section 7.5.3. Note that model performance resulting from every methodology in the following pre-study is collectively investigated at the end of this section.



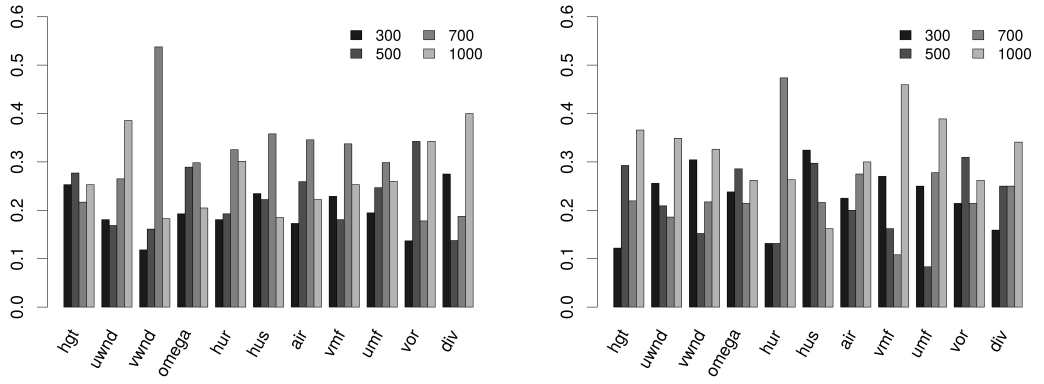
(a) DJF **(b)** JJA

Figure 7.2: Relative frequency with which each predictor is represented in the best 100 runs of random grid search, considering precipitation as predictand at Zugspitze station. **a)** DJF, **b)** JJA.

Random Grid Search

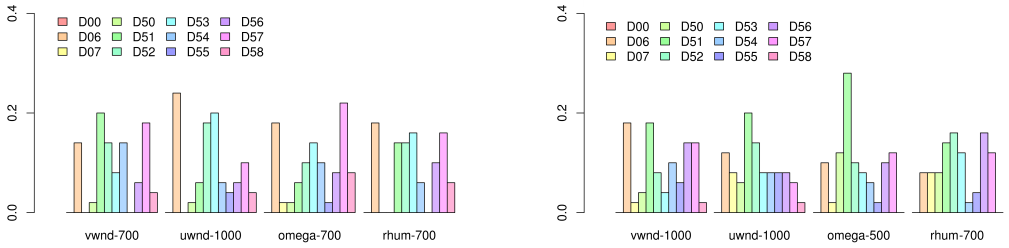
In a full grid search procedure, all possible input configurations are evaluated iteratively. Due to the long training time, resulting from a high number of possible predictor variations it is usually not feasible to perform a full grid search. Bengio (2012) suggests using a random grid search procedure to tune model hyperparameters, if a full grid search is not possible. In this study a modification of the original approach is investigated, where especially input specific properties are evaluated. In random grid search, for a large quantity of runs, input variable compositions (variables, levels and domain) are selected randomly from a pre-defined comprehensive set. Each configuration is used to calibrate and validate an ANN. The random grid search procedure assumes that setups including more suitable predictors perform relatively better. Setups of the best performing ANNs are assessed with the aim to find regularities. The number of top runs to be assessed and the required total number of runs must be subjectively defined depending on the available computational power and time frame. In this study a time limit of two weeks of computation is set to terminate the random grid search procedure. The approach is carried out for the time period 1970-2000 in winter and summer for Zugspitze precipitation. A total of 6957 runs were finished in winter and 7240 in summer.

Initially, the predictor variables are evaluated. Predictor variables are randomly selected with equal probability. In the total number of runs, the most frequently randomly selected variable in winter is slp (994 times) and the least frequently selected is umf (894 times). In summer, the most frequently selected variable is hgt (1019 times) and the least frequently selected is hur (962 times). Figure 7.2 shows the frequency each predictor is represented within the top 100 model configurations



(a) DJF **(b)** JJA

Figure 7.3: Relative frequency of the atmospheric levels 1000, 500, 700 and 300 (in hPa) of a predictor in the top 100 runs considering precipitation as predictand at Zugspitze station.



(a) DJF **(b)** JJA

Figure 7.4: Relative domain frequency in the first 50 occurrences of a specific predictor and level combination (e.g. vwnd-700: vwnd on a 700 hPa level). As predictand serves precipitation at the Zugspitze station. Compared to figure 7.2 and 7.3 the usage of the top 100 runs is considered to be insufficient to provide a comprehensive overview. The associated domains and acronyms are visualised in figure 7.5.

identified by the highest validation skill. The clearest signal is the low occurrence of slp in both seasons. No huge difference can be identified between the other predictors. This turns out as disadvantage of random predictor screening, as the user still has to derive the setup considering unambiguous information about the most influential predictors. Note that except for slp, the general high frequency implies that the top 100 configurations use simultaneous a large number of the provided predictors. A good setup for precipitation should be simple, but cover the air mass movement and its moisture content. Here, a similar variable setup is determined for both seasons: vwnd is selected as it is the most frequent predictor variable, in combination with uwnd, to include the horizontal state of the atmospheric circulation. The vertical motion is covered with omega and moisture with hur.

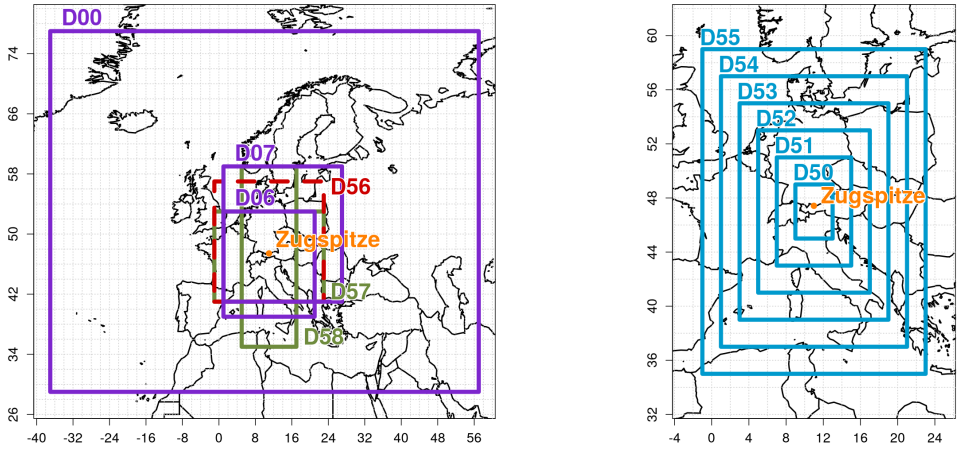


Figure 7.5: Subjectively defined domains for random grid search. Axis latitude and longitude in degrees. a) The domains Europe and surroundings (D00), Central Europe (D07) and Alpine region (D06) are delimited similar to Philipp et al. (2014b). D57 and D58 should satisfy predictors with strong zonal or meridional dependence, while D56 focuses the northern inflow direction. b) D50 to D55 are linearly increasing domains, with the target station Zugspitze in the centre.

Figure 7.3a visualises the respective atmospheric level of each predictor in the best 100 runs. Unlike to the predictor selection, the level study shows more evident tendencies. In winter $1000\,\mathrm{hPa}$ is identified for uwnd and $700\,\mathrm{hPa}$ for vwnd, omega and hur. In summer uwnd and vwnd both show good results on $1000\,\mathrm{hPa}$, hur on $700\,\mathrm{hPa}$ and omega on $500\,\mathrm{hPa}$. Selected levels are additionally listed in table 7.1.

One of the more difficult tasks is to find the best fitting domain. First of all, potential suitable domains are subjectively defined in advance, as random definition of borders is not feasible within an acceptable run-time (domains are displayed in figure 7.5). Second, the most frequent domains don't always include a proper representation of the characteristics of the predictor variable. For example, figure 7.4a implies D51 for vwnd. As vwnd should cover direction and origin of advecting air masses, D51 seems to be too small for that purpose (similar for uwnd in figure 7.4b). Hence, in some cases, the second or third frequent domain is selected (see Table 7.1). Conclusions of the perfect domains size are further difficult, as sometimes small and large domain seem to perform equally well (for example hur in figure 7.4a or vwnd in figure 7.4b). Last but not least, similar frequencies complicate the choice of the best domain. It is again up to the user to make an appropriate decision.

Table 7.1: Final predictor configuration derived with random grid search.

	Predictor	Level in hPa	Domain
DJF	uwnd	1000	D06
	vwnd	700	D57
	omega	700	D57
	hur	700	D06
JJA	uwnd	1000	D06
	vwnd	1000	D06
	omega	500	D51
	hur	700	D56

Correlation Based Predictor Screening

Predictors should be closely related to predictands, which can be quantified by correlation (e.g. Huth et al. (2014), Hertig et al. (2014), and Wilby and Wigley (2000)). Correlation based predictor screening is performed to confirm the importance of predictors identified by the random grid search process. In addition, domains can be specified more objectively by capturing high correlated areas. Spearman correlation coefficients are computed by comparing the time series of the predictand with each predictor grid cell of the 20th century reanalysis dataset. The predictor domain is defined to cover the European area and its surroundings ($-50°$ to $70°$ longitude and $26°$ to $80°$ latitude), while 1970 to 2000 is used as reference period.

Table 7.2 summarises the highest Spearman correlation coefficients between predictor and predictand for DJF and JJA sorted by the respective strength of relation. Before, each day of the time series is standardised separately to remove annual variability and detrended by applying a Gaussian high pass filter with eleven weights centred around the time step of interest.

Predictors selected during the random grid search procedure are found within the highest ranks. One could create a quite similar composition of predictors with minor modifications. First of all, other predictors could be preferred as for example in JJA pressure variables, like slp or hgt, show good correlation indices. Nevertheless, uwnd on $1000\,\mathrm{hPa}$ is top scorer in both cases and combined with vwnd to capture horizontal air-mass movement. vwnd, which is identified as top-predictor in random grid search, can still be located within the top five variables in both seasons. Predictor setups are completed with omega to cover vertical air movement and hur as moisture information carrier to satisfy the most important physical genesis factors of precipitation.

Additional considerations, which lead to the predictor selection: First of all, in

Table 7.2: Maximum Spearman correlation coefficient (r_s) between Zugspitze precipitation and the strongest related grid cell of a predictor-level combination in decreasing order (level in hPa). Left: DJF, right: JJA.

	Predictor	Level	r_s
1	**uwnd**	**1000**	**0.55**
2	umf	1000	0.54
3	**omega**	**1000**	**-0.53**
4	**vwnd**	**700**	**-0.47**
5	omega	700	-0.46
6	vor	1000	0.45
7	vmf	700	-0.45
8	slp		-0.43
9	hgt	700	-0.43
10	hgt	1000	-0.43
11	hgt	500	-0.42
12	div	1000	0.42
13	hgt	300	-0.41
14	vwnd	500	-0.38
15	**hur**	**700**	**0.38**
16	hur	1000	0.38
17	vmf	500	-0.36
18	uwnd	700	0.35
19	vor	700	0.35
20	umf	700	0.35
21	div	500	-0.34
22	vwnd	300	-0.33
23	omega	500	-0.33
24	vor	500	0.33
25	vmf	300	-0.33
26	air	700	0.32
27	vmf	1000	-0.32
28	uwnd	500	0.31
29	air	500	-0.31
30	vor	300	0.31
31	vwnd	1000	-0.30
32	uwnd	300	0.29
33	div	700	-0.29
34	hur	500	0.28
35	umf	500	0.27
36	hus	1000	0.26
37	umf	300	-0.24
38	air	300	-0.23
39	air	1000	0.23
40	omega	300	-0.22
41	hus	300	-0.22
42	hus	700	0.22
43	div	300	-0.19
44	hur	300	-0.16
45	hus	500	-0.16

Predictor	Level	r_s
uwnd	**1000**	**0.48**
umf	1000	0.47
slp		-0.42
hgt	1000	-0.42
vwnd	**500**	**-0.41**
vwnd	1000	-0.41
hur	**700**	**0.41**
air	700	-0.41
air	1000	-0.41
vor	500	0.41
vmf	1000	-0.41
vor	300	0.40
hgt	500	-0.39
vwnd	300	-0.39
vwnd	700	-0.39
omega	**500**	**-0.39**
div	300	0.39
hgt	300	-0.38
omega	300	-0.38
hur	1000	0.38
vmf	300	0.38
vmf	500	-0.38
vmf	700	-0.37
hgt	700	-0.36
omega	1000	-0.36
hur	500	0.36
div	1000	-0.36
vor	700	0.36
vor	1000	-0.36
air	500	-0.35
uwnd	700	0.32
omega	700	-0.32
uwnd	500	0.30
hus	700	0.29
umf	300	-0.29
umf	700	0.29
uwnd	300	0.28
umf	500	-0.27
hus	300	-0.26
hus	500	0.26
hus	1000	-0.25
div	700	-0.25
hur	300	0.23
div	500	0.22
air	300	-0.19

both seasons, umf and in winter additionally vmf are ranked higher than hur. The reason could be that moisture flux is computed from u- and vwnd, thus a large amount of relationship can be traced back to the wind vectors. For this reason, hur is selected in both seasons, as wind vectors are already included in the predictor setup. Second, although identical in most cases, levels slightly change in comparison to random grid search. The $1000\,\mathrm{hPa}$ level of omega in winter is higher correlated than $700\,\mathrm{hPa}$ and in summer the $500\,\mathrm{hPa}$ level of vwnd seems to be slightly superior to $1000\,\mathrm{hPa}$. Third, equal ranks are oftentimes assigned to predictors, due to similar correlation coefficients. This is particularly problematic in cases where different atmospheric levels of a predictor share one rank, e.g. vwnd $500\,\mathrm{hPa}$ and $1000\,\mathrm{hPa}$ in JJA. In these cases, a cleaner spacial pattern of the correlation coefficients (not shown here) is used as subjective criterion to decide, which level to choose. The final predictor setup for the correlation based screening is marked in bold in table 7.2.

Furthermore, to include temporarily delayed relationships between predictor and predictand time lags of -2 to +2 days are investigated in a cross-correlation like approach, but no higher correlation coefficient could be identified.

The next step is to find a suitable domain. Figure 7.6 and figure 7.7 visualise the results for winter and summer, respectively. Domains are identified using quantiles of the correlation indices considering a base region of $-50°$ to $70°$ longitude and $26°$ to $80°$ latitude. Domains are generated fully automated by delimiting regions above $97\,\%$ quantile for positive correlation indices and below $3\,\%$ quantile for negative ones. However, quantiles are adjusted under certain circumstances: If the resulting domain includes insignificant correlation coefficients (level of significance $5\,\%$), quantiles are shifted iteratively by $0.1\,\%$ to their outer bounds ($-0.1\,\%$ for negative and $0.1\,\%$ for positive correlation indices). The domain size is not further reduced if it would fall below a minimum of nine grid cells. The latter condition should preserve the skilful scale criterion. In case the starting domain is already smaller than the minimum grid cell number, the quantile range is enlarged. Final quantiles are listed in table 7.3. Two different starting partitions are compared: First, only the strongest correlation pattern is engridded, second, the strongest positive and negative correlation indices are considered. For the sake of simplicity, the outer bounds of the quantile border is transferred into a rectangular domain. In the second case, where positive and negative correlation indices are considered, both centres are integrated into one large domain. Insignificant values are impossible to exclude in the resulting large domains, the criterion is therefore not applied.

The pattern of highest correlation lies usually close to the measuring stations. The general characteristics in winter and summer are similar, although correlation features are usually more pronounced in winter. Strong west wind speed and a high moisture content result in high precipitation. The omega pattern in winter is scat-

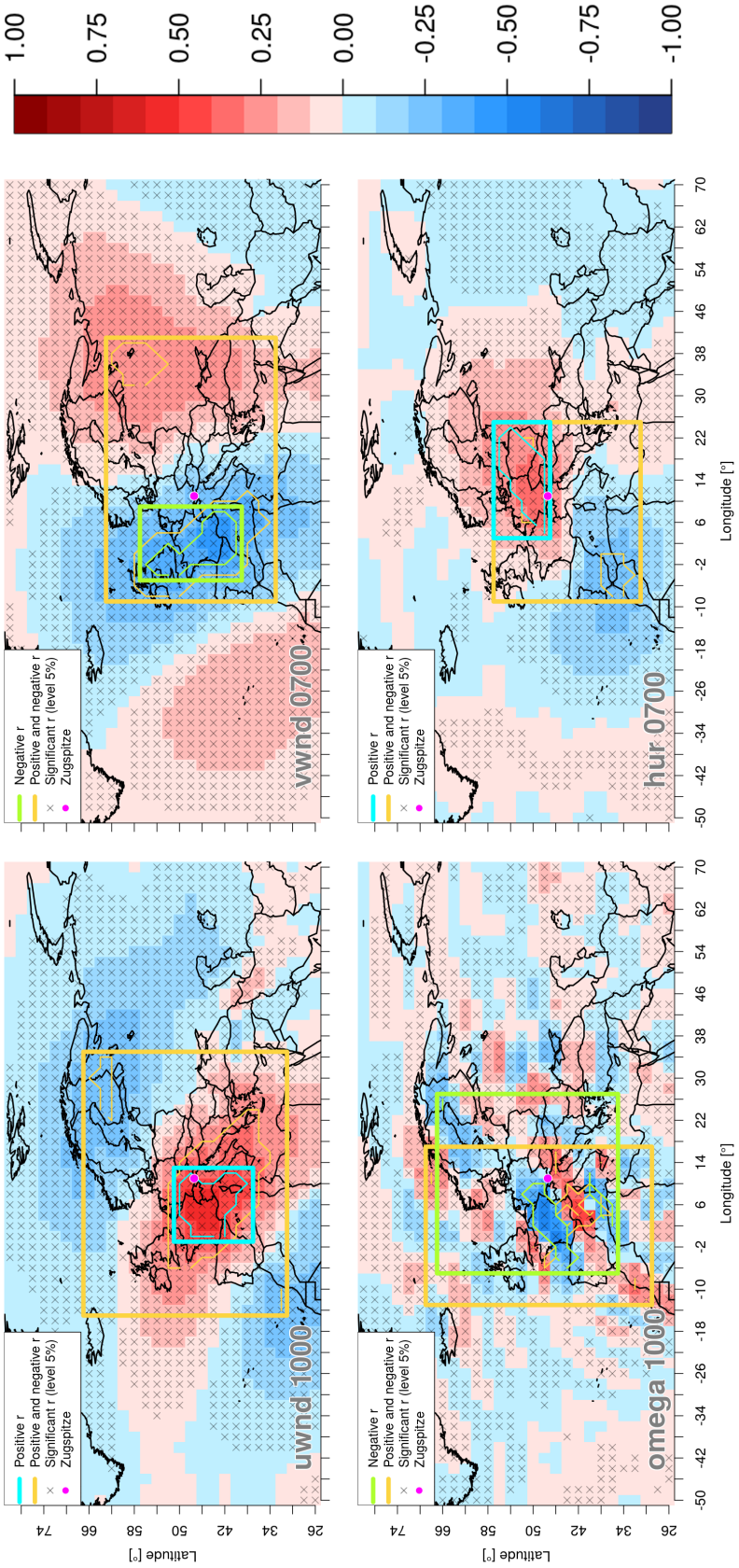


Figure 7.6: Predictor domains defined by correlation (r) for Zugspitze precipitation in winter. The correlation map displays Spearman correlation coefficients between the predictand and the respective grid cells of the selected predictors from the 20th century reanalysis dataset (see table 7.2) in the reference time period 1970–2000. Table 7.3 lists the corresponding quantile values. Thick polygon lines visualise the resulting domain, thin lines represent the actual threshold.

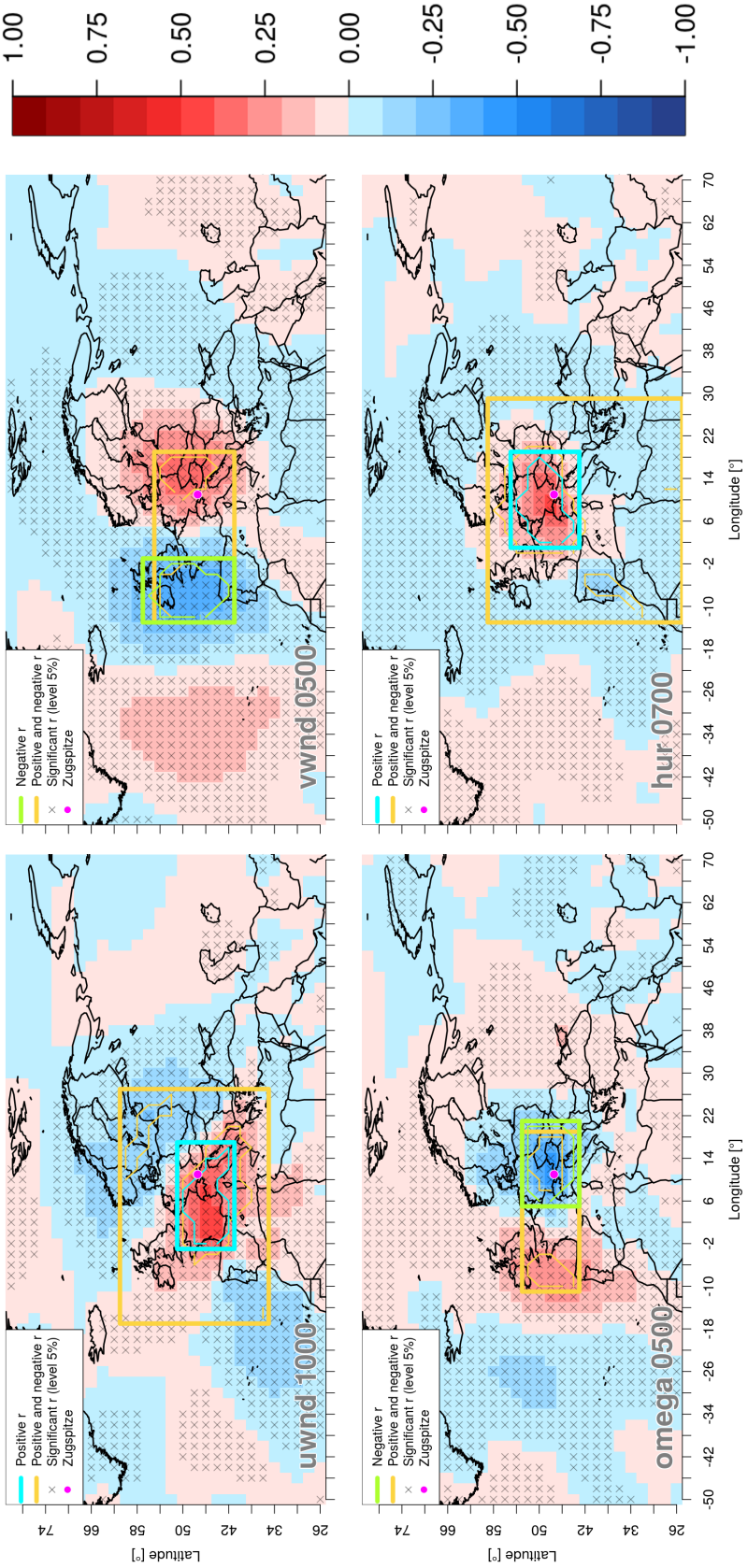


Figure 7.7: Predictor domains defined by correlation (r) for Zugspitze precipitation in summer. The correlation map displays Spearman correlation coefficients between the predictand and the respective grid cells of the selected predictors from the 20th century reanalysis dataset (see table 7.2) in the reference time period 1970-2000. Table 7.3 lists the corresponding quantile values. Thick polygon lines visualise the resulting domain, thin lines represent the actual threshold.

Table 7.3: Quantiles of the distribution of Spearman correlation coefficients identified to delimit the domain for each predictor-level combination. Quantile str represents the quantile including the strongest correlation coefficient, positive or negative, while quantile +/- focuses on the best positive and negative correlation coefficients.

	Predictor	**Level**	**Quantile str**	**Quantile +/-**
DJF	uwnd	1000	$x > 0.98$	$0.07 < x > 0.93$
	vwnd	700	$x < 0.02$	$0.06 < x > 0.94$
	omega	1000	$x < 0.01$	$0.03 < x > 0.97$
	hur	700	$x > 0.98$	$0.05 < x > 0.95$
JJA	uwnd	1000	$x > 0.98$	$0.05 < x > 0.95$
	vwnd	500	$x < 0.02$	$0.03 < x > 0.97$
	omega	500	$x < 0.02$	$0.02 < x > 0.98$
	hur	700	$x > 0.98$	$0.05 < x > 0.95$

tered, which can be caused by the low atmospheric level of omega ($1000\,\mathrm{hPa}$), with positive and negative influences near the station. During summer, upward air movement tendencies are connected to high precipitation. During winter, a strong north wind increases precipitation due to the orographic uplift at the slope of the mountain range. During summer, southern winds are related to high precipitation. This can be an indirect measure for cyclones moving from west to east-north of the Alps. These cyclones suck air from the southern parts of the Alps resulting in a southern wind. Precipitation at Zugspitze therefore origins from the cyclone and not from the usually dry southern wind due to the Foehn effect (see section 11.1).

Several drawbacks occur when only relying on correlation based predictor screening. Similar to random grid search, the combination of predictor variables must be defined by the user. Different characteristics of correlation indices in space and time need to be considered, ideally combined. Further, relationships between predictors should be accounted for (e.g. Kaspar-Ott et al. (2019)). Variables on different levels or of similar type (e.g. hgt $1000\,\mathrm{hPa}$ and slp) are often ranked closely to each other, which makes simple selections, like the top "x" ranks, unfeasible (see table 7.2). Dependent predictors usually do not add any additional value but instead add redundancy and noise to a model and complicate the optimisation process. Last but not least, SD models themselves have no influence on the predictor configuration in this procedure. Although a strong relationship between predictor and predictand is necessary, correlation on its own is only an indication and no guarantee for good SD model performance. Slp and vor, for example, in winter or umf in summer are highly correlated but less frequently found in the top 100 runs of the grid search procedure.

Automated Generation of a Domain

The aim of automated domain generation is to find a non-generic domain under consideration of the SD models needs. Sauter and Venema (2011), for example, propose that an optimised domain can be beneficial for the SD model quality. In the following, a simple approach of an automated domain generation procedure is suggested. The process is realised separately for each predictor and level, which must be combined afterwards into a physically consistent setup. Here, as basis serves the previously identified predictor combination via correlation. The results are visualised in figure 7.8 and 7.9 for winter and summer, respectively. The domain search procedure is typically started at the grid point location of the strongest correlation index, with two exceptions: For vwnd in summer the closer positive correlation centre was selected as the negative centre seems to be quite far away from the observation station. In addition, the starting partition for omega in winter is also defined closer to the station comprising significant negative and positive correlation indices. The correlation map of omega is fairly fragmented. Starting from vicinity of the Zugspitze appeared to be a saver assumption. In all other cases the strongest correlated and three neighbour grid cells, which are selected in alignment with the direction of the meteorological station are defined as a 2x2 starting partition, labelled 0. At each step of the procedure the SD model is calibrated and validated with extended domains in the four cardinal directions. Every new domain configuration is ten times Monte-Carlo cross-validated. Domain size is increased afterwards following the direction with the lowest mean error. Progression of enlargement is marked with step numbers in figures 7.8 and 7.9. The iterative process is carried out until a border of the regional excerpt ($-50°$ to $70°$ Longitude and $26°$ to $80°$ Latitude) is reached. Lowest mean error of the cross-validation process at each step is displayed by the MSE in figure 7.10. The final domain (dashed purple in figures 7.8 and 7.9) is selected from one of the first visually pre-eminent local minima (marked with a red box in figure 7.10) to keep the domain size small and thus the SD model as simple as possible.

The development of domain extend shows that generated domains do not necessarily follow regions with the strongest correlation indices. The domain of vwnd in winter, for example, is enlarged eastwards, although higher correlation areas are found in the southern direction. Similar for uwnd in summer: A higher correlated pattern ranges from west to east, while the procedure follows a secondary southern outgrowth. A reason could be redundancy: Similar correlated areas might carry equivalent information, adding no additional value to the SD model. Nevertheless, the final tendency of the domain generation procedure is to build domains, which cover the highest correlated areas. This characteristic approves the correlation based domain specification. The subjective alternation of the starting partition for vwnd in summer in the correlation based approach is questioned, as the algorithm extends

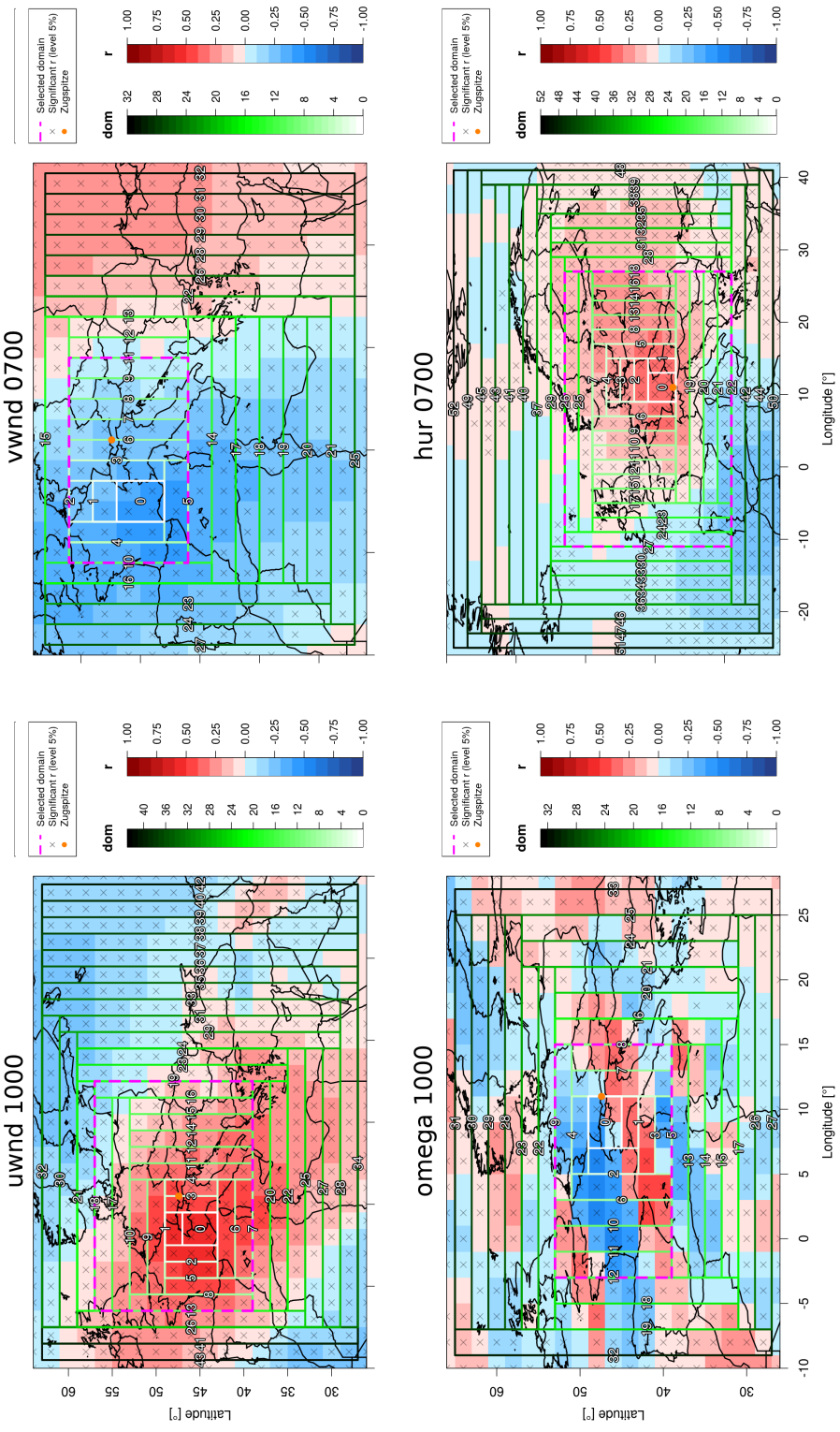


Figure 7.8: Development of the domain for each predictor in DJF. Target: Zugspitze precipitation. The iterative process begins at a starting domain 0, which is defined by the strongest Spearman correlation (r) between predictor and predictand. White numbers mark the respective direction and step of the domain enlargement.

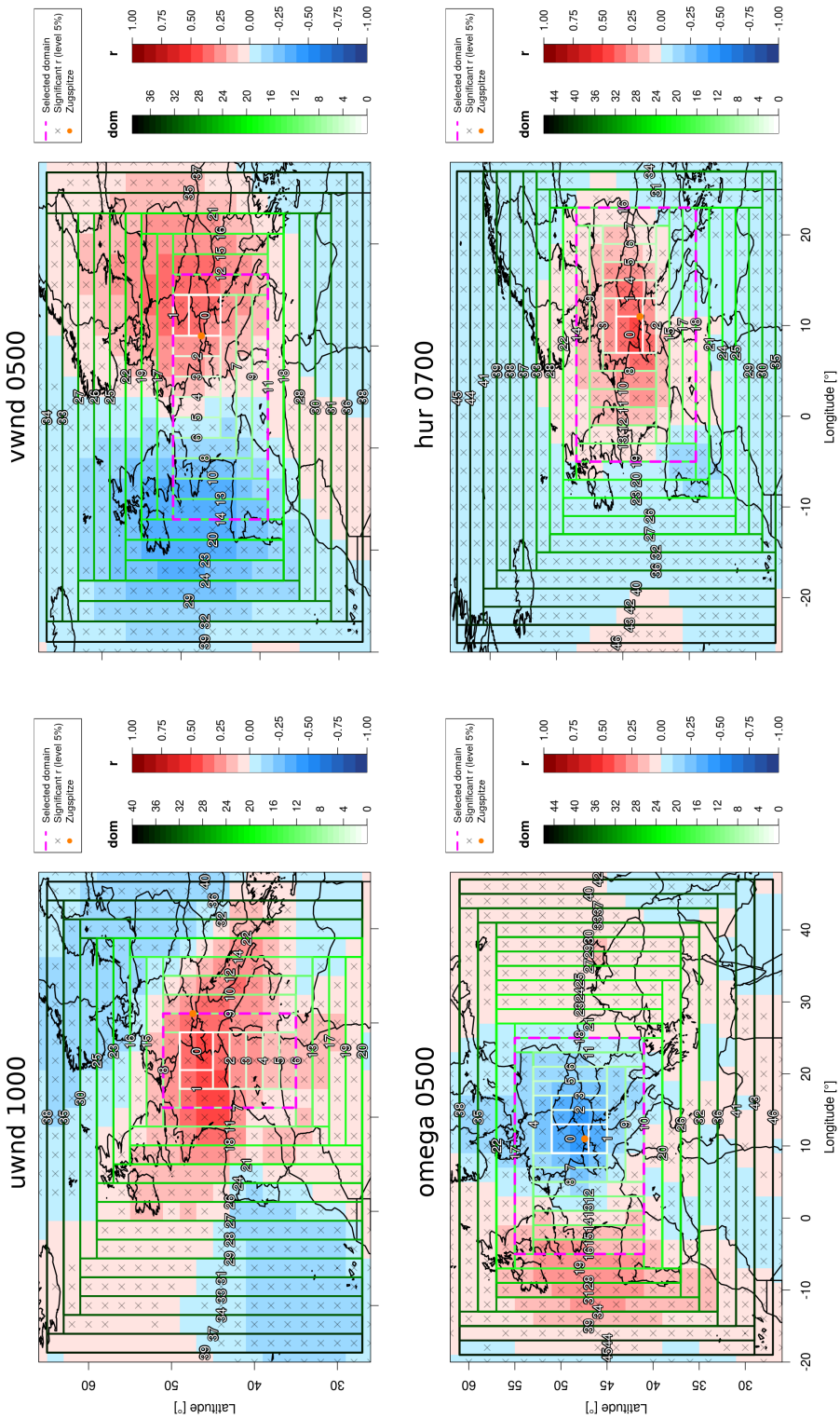
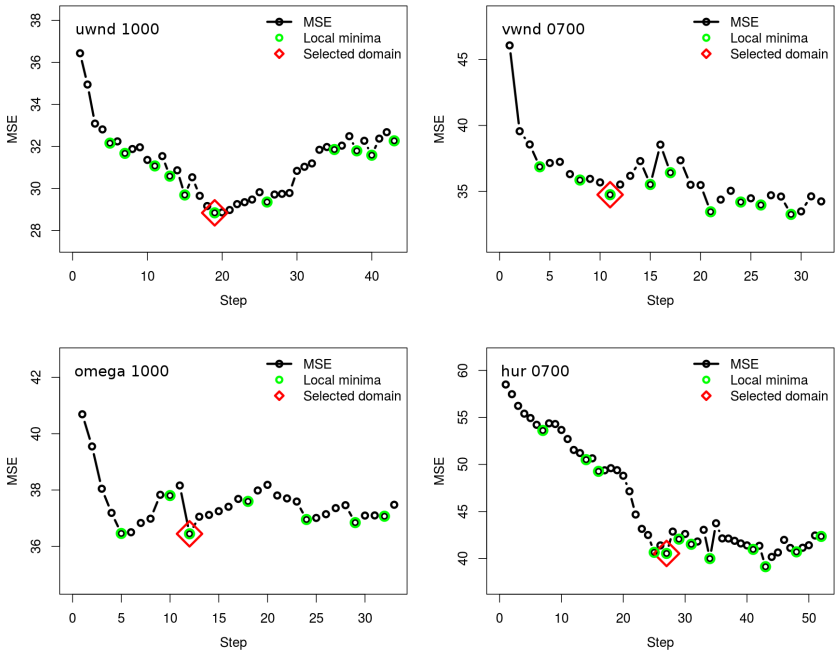
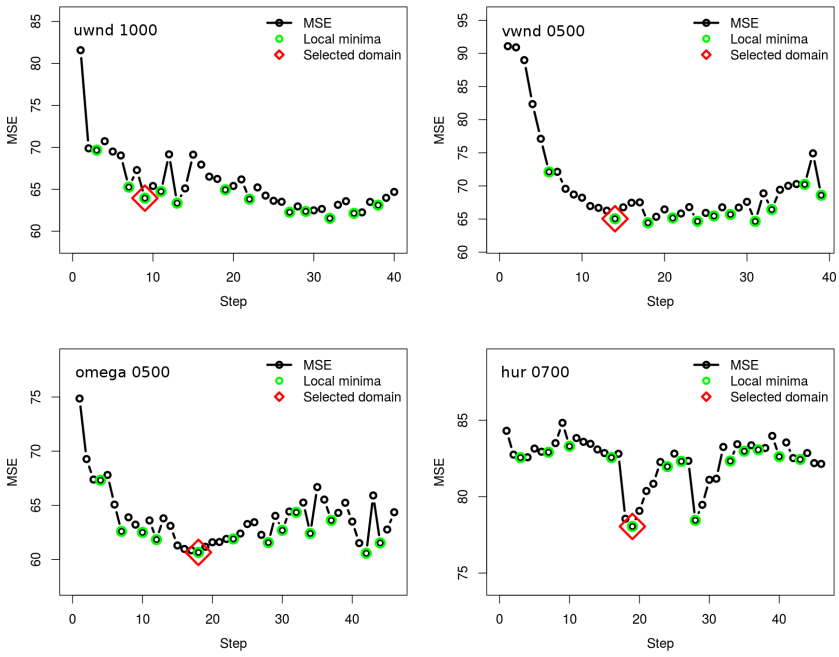


Figure 7.9: Development of the domain for each predictor in JJA. Target: Zugspitze precipitation. See figure 7.8 for detailed description.



(a) DJF



(b) JJA

Figure 7.10: MSE progression of the domain search procedure, **a)** of figure 7.8 and **b)** figure 7.9. To keep the domain's size as small as possible, but as explanatory as necessary, one of the first suitable local minima is subjectively selected as domain (red rhombus shaped box).

the domain primarily into the west. Similarly, domains enlarged of omega in winter point into the western direction, where the area with the highest correlation indices is located as well.

The major drawback of this approach is the long run time, as a large amount of ANNs needs to be trained. Further, the final selection of the resulting domain based on the error is in the final step done again subjectively. In addition, the procedure is carried out for one predictor at a time and therefore does not take inter-predictor relationships into account. Finally, the procedure does not help to find a suitable predictor composition, but needs further information from other approaches like correlation based predictor screening or random grid search. Last but not least, the domain is extended by only considering the immediate neighbouring grid cells. An enlargement into the optimal domain can be prohibited by a trench of unfitting grid cells separating suitable ones from the current domain.

The approach can be modified in various ways. An initialisation from random seeds, for example, is conceivable. Disadvantage of random initialisation is that not all possible starting partitions will lead to a satisfactory result, e.g. when outer boundaries are selected. Domain expanding can also be performed by adding single grid cells. However, these modifications would lead to a considerably higher runtime and were therefore discarded. Furthermore, by downsizing a large domain, the domain optimisation process could also be reversed. Last but not least, a setup of several predictors could be optimised simultaneously by updating each domain consecutively by one step in a round, which would take inter-predictor relationships better into account.

Analysis of Sensitivity

A popular way to investigate the impact of ANN inputs on the predictand are sensitivity studies (e.g. Shojaeefard et al. (2013), Gevrey et al. (2003), Dreiseitl and Ohno-Machado (2002), Zeng and Yeung (2001), and Hunter et al. (2000)). The methodological background of different sensitivity analysis approaches, which are used in the following, are described in chapter 5.5.7. The sensitivity studies of this section serve three goals: First, to assess the importance of input time series encircled by the domains defined during previous approaches. Second: Is it possible to specify "ideal" domains with sensitivity studies? Sensitivity studies of ANNs are rarely used to evaluate predictors in statistical downscaling (e.g. Nourani and Fard (2012) and Castellanos et al. (1994)) and so far not applied to explore domain size and location. Third, how coinciding are different sensitivity studies in their results?

Figures 7.11 and 7.12 compare the sensitivity of previously identified predictor domains (rows, blue border) by predictor screening techniques (columns) with the help of PaD (see section 5.5.7). White areas have no/low influence on the simulated

predictand time series, in contrast to high impact areas coloured in dark red. Note that each predictor screening method (columns) has a different scale. The values of PaD sum up to unity, thus the magnitude of sensitivity highly depends on the total number of grid cells. Equal scales would make setups with small and large predictor domains visually hardly comparable. Scales in this context are defined starting at the minimum and ending at the maximum sensitivity with an equal amount of intervals.

Although the domains are generally not completely congruent, identical regions feature similar sensitivity tendencies: In winter (figure 7.11) uwnd (a) shows a sensitive region in advance of the western Alpine ridge, which fits in the flow direction of the westerlies. The selected domains of vwnd (b) unfortunately varies considerably, but one can still see a high sensitive region west to Switzerland extending to the southern parts of France, a second region near the English Channel and a third one near Croatia. Surprisingly, the direct northern inflow, which is responsible for large amounts of precipitation at the Zugspitze, is only captured within the random grid search procedure (h). The sensitivity of omega (c) surrounds the Alps arc like shaped, starting in south-west Germany, leading through central France and ending south of the Po Valley. Omega uses a different level in the random grid search procedure and is therefore not shown in column (h). The main sensitive regions considering rhum (d) are located in the direct neighbourhood of the observation station, which is typical for moisture predictors. Large domains (e.g. column (f) or rhum in (g)) tend to have a high number of low influential grid cells. These should be excluded as they add unnecessary information or noise, which handicaps the optimisation of the ANN and, in the worst case, limits the generalisation capability. Similar single isolated but sensitive outliers clearly represent effects caused by the optimisation and need to be excluded from the model. Causal relationship of these aparted grid cells is very unlikely, e.g. rhum near Morocco or Algeria in column (f). In summer (figure 7.12), uwnd (a) shows a similar sensitivity pattern like in winter. vwnd (b) and rhum (d) have both low influence with no specific centres, but a strong signal is eminent close to the Zugspitze in omega (c). Again, if domains include superimposable areas, similar sensitive regions can be identified throughout predictor screening methods. In contrast to winter, the summer setup has fewer but more pronounced sensitive areas. Winter and summer sensitivities indicate the different precipitation genesis factors. In winter, the focus lies on the horizontal, advective, air movement and although uwnd is still a sensitive predictor. In summer, important core areas are found in omega, which represents vertical movement or convective processes.

Taking into account the similar sensitivity pattern found by the different predictor screening approaches, the question arises if domains or even predictor setups can be defined on the basis of sensitivity studies. The latter will be discussed in chapter 7.5.3, here the specification of each domain to the most important areas is described.

Column (i) in figure 7.11 and figure 7.12 shows the selective domain setting. In addition, the domain of column (i) is displayed in purple in the other columns for comparison purpose. The pattern within the fitted domains in column (i) goes along with previous findings. Here, again similar regions are identified as most sensitive.

Figure 7.11 and 7.12 compare different sensitivity methods (described in section 5.5.7). The noise range of one-on-one adding noise is set from -0.5 to 0.5 and the respective sensitivity is calculated from the mean of 30 runs. Note that the sensitivity quantities have different meanings and therefore scales. Sensitivities of PaD sum up to unity, while one-on-one zeroing and adding noise result in MSE values. This could be a reason for a higher agreement in the pattern of zeroing and adding noise. Similar to the figures above, to keep as much comparability as possible, the scales are constructed equally by identical intervals from the minimum to maximum value. In all three cases, white areas are of no importance to the predictand, whereas dark red areas have strong influence. Again the pattern of each predictor (rows) are quite similar in both seasons, with strong analogy especially of the most important grid cells. PaD (e) seems to be more sensitive to influential inputs, although this effect could also origin from different scales. Furthermore, PaD is the fastest methodology, as it is computed with little effort from one forward pass (simulation) of the ANN, while zeroing needs a simulation for each grid cell, and adding noise, in addition to the run time of zeroing, should be performed several times. For these reasons, PaD is primarily chosen in the following to evaluate ANN sensitivities. This comes hand in hand with the rating of Gevrey et al. (2003) or Shojaeefard et al. (2013), where PaD is described as one of the most reliable tools to analyse ANN sensitivity.

How does the ANN exclude unnecessary input? It is clear that this characteristic is originating from the weight settings optimised by the training algorithm. Unfortunately, several studies, which try to derive the sensitivity of inputs with ANN weights found that the magnitudes of the weights are not a reliable measure of importance for ANN inputs (e.g. Shojaeefard et al. (2013), Montano and Palmer (2003), W. Wang et al. (2000), and Garson (1991)). The reason might be that the impact of very small input values is still small even if the weight is large. To reduce the risk of misinterpretation, it is helpful to transform input variables in equal magnitude ranges, for example by scaling or standardisation of predictors. Predictor distributions with non-zero skewness can blur variable importance and to prevent this one could check e.g. for standard normal distribution. Nevertheless, in theory, various weight situations are able to exclude input. If a weight, for example, connecting an input grid point with the next neuron layer is zero, the respective proportion to the net input is zero (see equation 5.18). In figures 7.13 and 7.14 inputs are labelled with a pink cross, when their connection weights to the next layer are close to zero. Weights are defined as close to zero if they range from -0.1 to 0.1 (large weights can reach val-

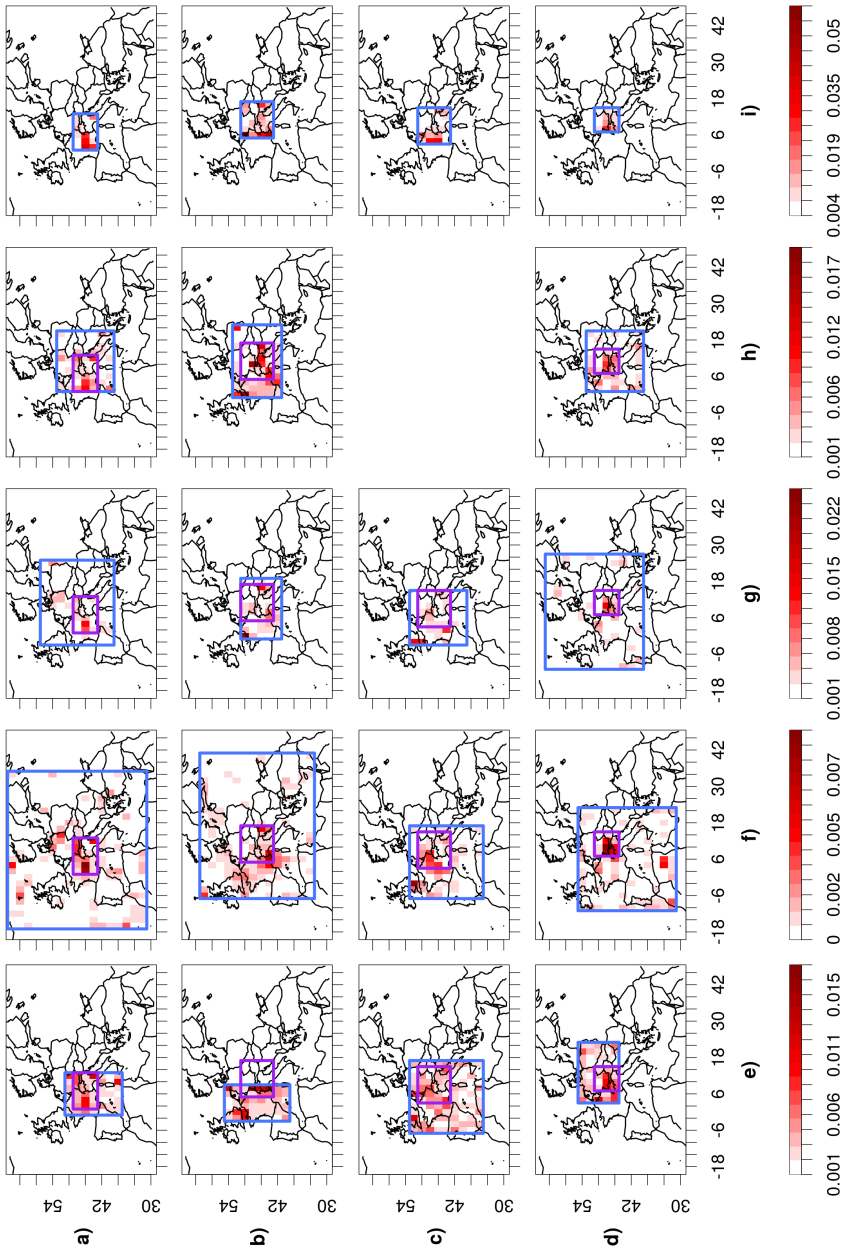


Figure 7.11: Sensitivity analysis of predictor setups for Zugspitze precipitation in winter. Each row represents a different predictor: a) uwnd 1000 hPa, b) vwnd 700 hPa, c) omega 1000 hPa and d) hur 700 hPa. Columns separate predictor selection procedures: e) strongest correlation, f) min and max correlation, g) Domain search, h) random grid search and i) reduced to most sensitive regions. A dissimilar level of omega is selected during random grid search, thus sensitivity pattern are not included.

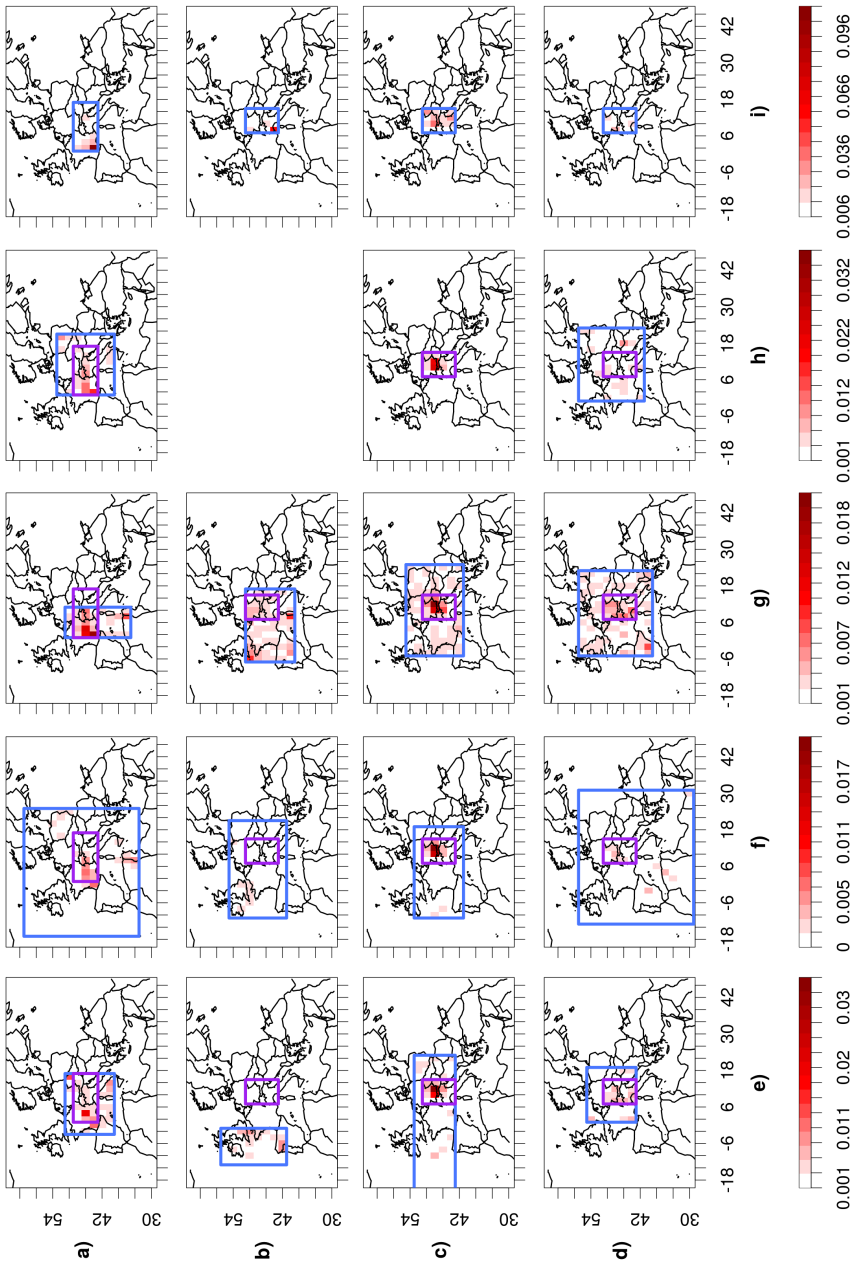


Figure 7.12: Sensitivity analysis of predictor setups for Zugspitze precipitation in summer. Each row represents a different predictor: a) uwnd 1000 hPa, b) vwnd 500 hPa, c) omega 500 hPa and d) hur 700 hPa. Columns separate predictor selection procedures: e) strongest correlation, f) min and max correlation, g) Domain search, h) random grid search and i) reduced to most sensitive regions. A dissimilar level of vwnd is selected during random grid search, thus sensitivity pattern cannot be included.

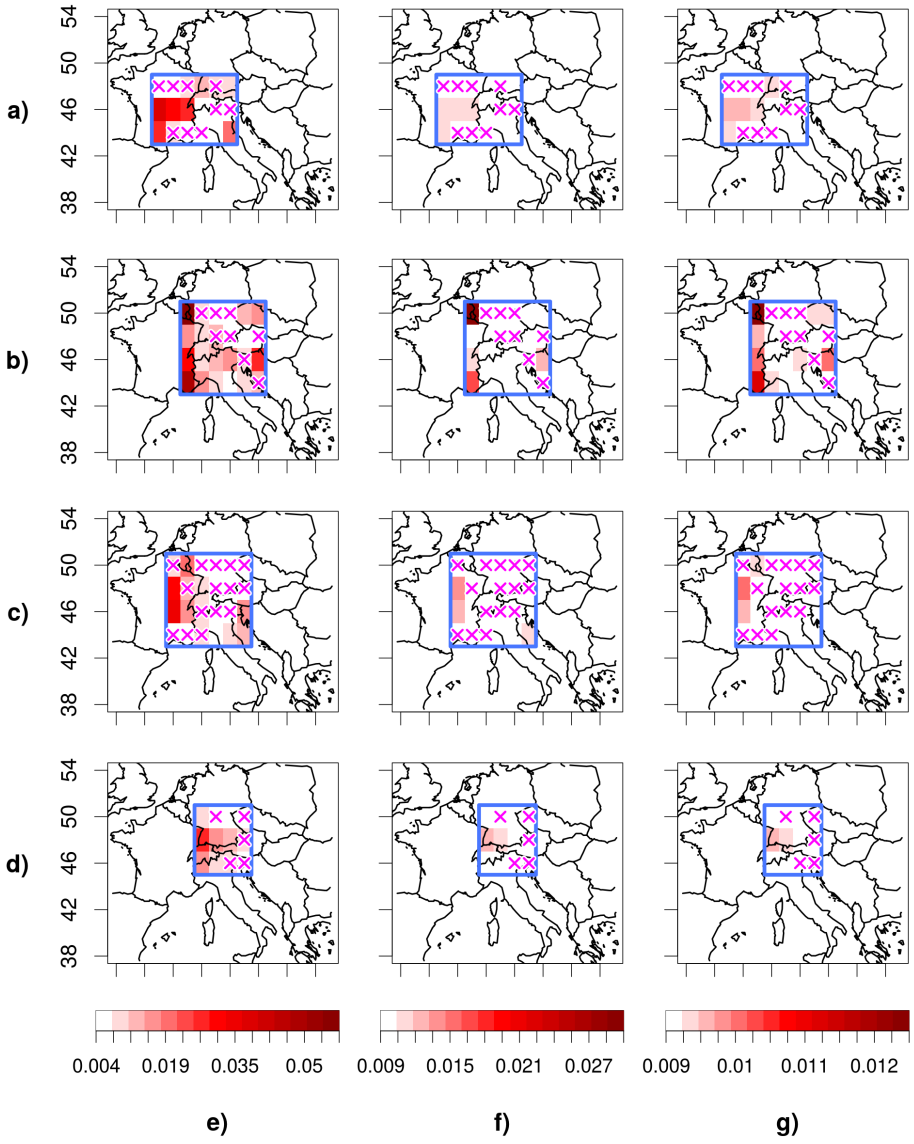


Figure 7.13: Comparing different sensitivity methods considering the predictor setup reduced to the most sensitive regions applied to Zugspitze precipitation in DJF (see column i) in figure 7.11). Each row represents a different predictor: a) uwnd 1000 hPa, b) vwnd 700 hPa, c) omega 1000 hPa and d) hur 700 hPa. Columns separate sensitivity methods: e) Partial derivatives, f) One-on-one zeroing and g) One-on-one adding noise.

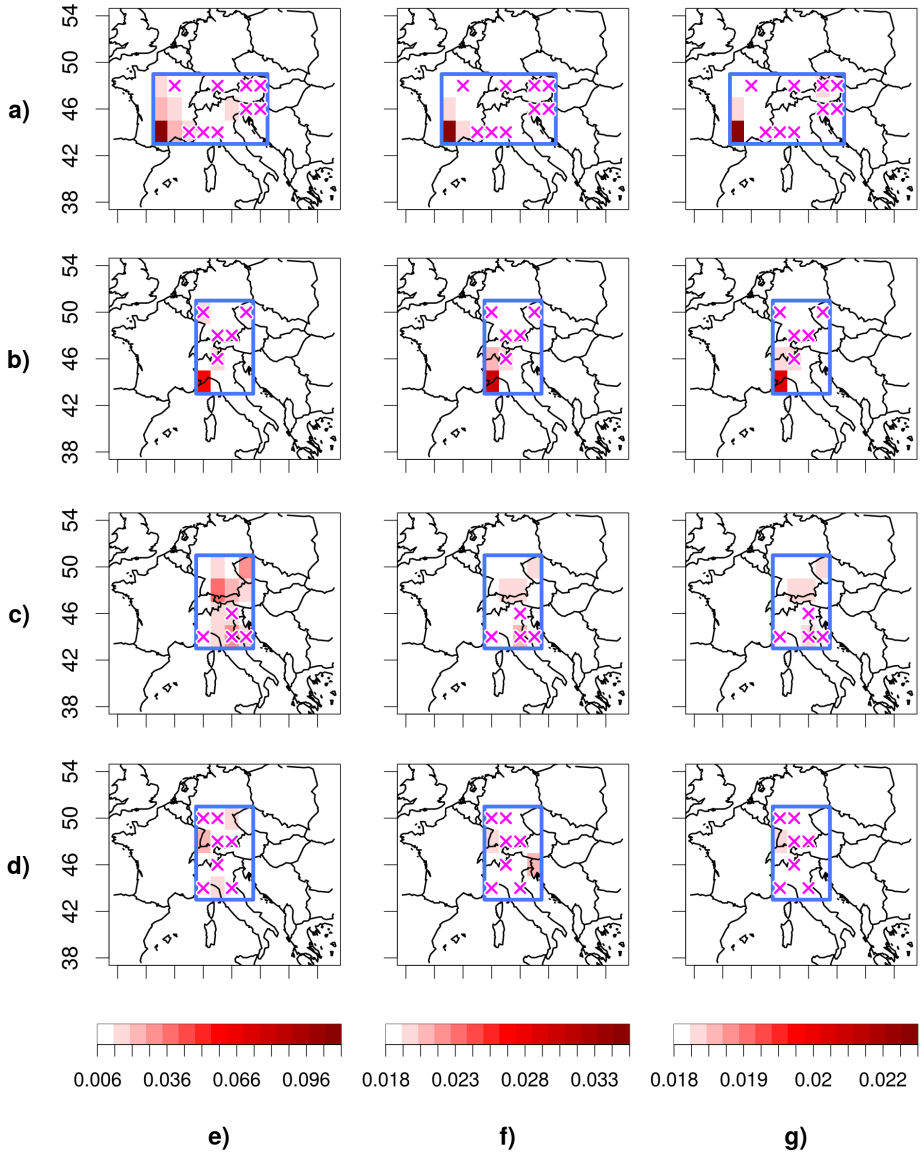


Figure 7.14: Comparing different sensitivity methods considering the predictor setup reduced to the most sensitive regions applied to Zugspitze precipitation in JJA (see column i) in figure 7.12). Each row represents a different predictor: a) uwnd 1000 hPa, b) vwnd 500 hPa, c) omega 500 hPa and d) hur 700 hPa. Columns separate sensitivity methods: e) Partial derivatives, f) One-on-one zeroing and g) One-on-one adding noise.

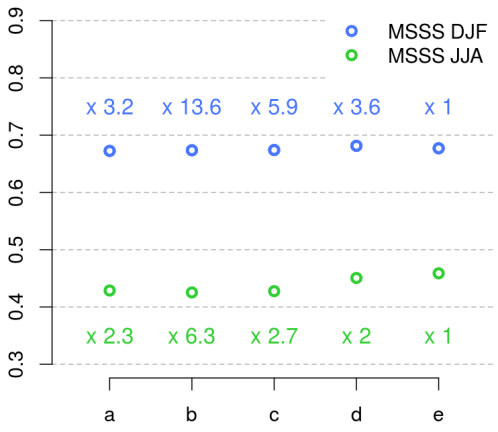


Figure 7.15: Comparing the model performance of selected predictor combinations considering the mean MSSS of a 10-fold Monte-Carlo cross-validation for both seasons DJF and JJA. a) strongest correlation, b) min and max correlation, c) Domain search, d) random grid search and e) reduced to most sensitive regions. Further, the multiplier of the number of weights is inscribed, which must be optimised in each combination compared to e.

ues up to +/- 15). Generally less sensitive grid points are marked. Still, there are plenty of cases which do not coincide: First of all, many insensitive areas, especially in summer, cannot be identified by low weighted connections. Second, some grid points with modest sensitivity in all three methods are tagged as low weighed (e.g. in vwnd or omega in summer). Here, other ANN internal weight settings and interdependency must be responsible for the sensitivity.

Sensitivity analysis comes with disadvantages, the fact that sensitivity can only be checked after the training process being one of them. A suitable predictor setup must be defined in advance and can be checked afterwards. Further, the best domain is difficult to be objectively obtained as sensitivity studies only give hints on important inputs. A stand-alone approach of how to find a proper predictor setup is proposed in section 7.5.3. In addition, sensitivities are computed considering a certain input dataset and are only valid for this pattern.

Resulting Model Performance

Figure 7.15 shows the MSSS obtained by the above-stated predictor screening methods. There is hardly any difference in model quality between the setups in both winter and summer. The reason might be the high accordance in the respective sensitive regions. Even when using large domains (e.g. column (f) in figure 7.11 and 7.12) the

network seems to be quite efficient in excluding unnecessary input and focus on the most sensitive grid cells. This theory is supported by the almost identical SD model quality if only the most sensitive regions are used as domain (e). The biggest advantage of small domain sizes is the extremely reduced number of model parameters. In winter, for example, the 13.6-fold and in summer the 6.3-fold number of weights needs to be trained when using the domain based on positive and negative correlation. The possibility to find specific domain sizes and locations in particular, which comes hand in hand with a minimised number of ANN parameters to be optimised, leads to the decision to build predictor setups exclusively from sensitivity studies. Castellanos et al. (1994) describe the characteristics of resulting ANN skills in a comparable context: Different predictor combinations in this study lead to similar model performances. Castellanos et al. (1994) recommend using sensitivity studies to exclude unnecessary predictor variables. Note that the model quality in figure 7.15 is hardly influenced by the different levels of omega in winter and vwnd in summer when comparing the model quality. An explanation can be the almost identical correlation between the vwnd levels and the predictand in summer (see table 7.2). Slightly larger differences in correlation can be found between the omega levels and the predictand in winter.

7.5.3 Final Result of Predictor Setups Derived from Sensitivity Studies

This section describes the creation of the final domain and predictor combination for the six predictand variables: precipitation and temperature at Sonnblick and Zugspitze, as well as relative humidity and wind speed at the Zugspitze station.

Note that the predictor screening procedure in this section is redone from scratch. Concluding from the previous section ANNs are efficient and consistent in finding the most important inputs. Therefore, not only the domain on preselected predictors, like in the previous section, but also the right predictors will be selected by sensitivity studies in the following.

All models are again developed within the time period 1970-2000, while $60\,\%$ of the time steps are used for training, $20\,\%$ for early-stopping validation (with 200 checks) and $20\,\%$ for testing the SD model. Each input setup is trained ten times cross-validated with eight hidden neurons by the Rprop algorithm.

The predictor screening procedure is performed iteratively. At each iteration, an ANN is trained with a certain predictor configuration while the importance of each predictor is assessed with PaD and the model quality is evaluated by the MSSS. The process is further divided into two major tasks: 1) The first aim is to reduce a large number of possible predictors and levels to a sufficient selection. The focus of the first steps lies on finding the best level for each predictor, only highly insensitive

ones are removed completely. Towards the end of the first objective, the key aspect switches to the elimination of predictors. The best predictor combination is identified as the one previously of a severe drop in model performance. In addition, a minimum setup is defined for each predictand type. The minimum setup is represented by predictors, which are mandatory to adequately describe the physical genesis of the predictand (see below). The minimum setup defines the hard limit, at which the predictor elimination process is cancelled in any case. 2) The second objective is to reduce the domain of each predictor member of the best configuration to the most sensitive regions over two steps. The final predictor configuration should not result in a strong reduction of model performance compared to the initial setup.

The initial predictor selection, as well as the minimum setup both depend on the predictand variable and are based on common setups found in recent literature (see section 4.2.3). The initial setup for the predictand precipitation covers all acquired predictor variables except for horizontal wind speed (swnd), as there is no physical or advective link between horizontal speed and precipitation. The minimum setup should at least include one example of the three major large-scale atmospheric components, which determine precipitation received at an observational station (Tuller 1973): 1) a circulation type component (any predictor counts except for temperature, which is optional); 2) the quantity of available water is taken into account by specific humidity (hus); 3) the proximity to the dew point at which water will condense represented by relative humidity (hur). The initial setup for the predictand relative humidity is equal to precipitation, however the minimum setup is simply hur. The initial setup of temperature excludes moisture variables (hur, hus, umf and vmf) and wind speed. The link, carried in moisture variables, between absorption and emission of energy during phase transition of water is considered to be too complex for a SD model approach of this type and neglected to reduce the input dataset and thereby training time of the ANN. The minimum setup of temperature is simply the atmospheric temperature field (air). The initial setup of the predictand wind speed consists primarily of circulation type variables. The minimum setup is solely atmospheric wind speed (swnd).

Further, the selection of the atmospheric layer levels is adjusted within this final predictor screening approach. During the ESM data acquisition and inspection, discrepancies occurred, which made an adjustment of predictor levels within this final predictor screening procedure necessary. The $300\,\mathrm{hPa}$ level is rarely provided, as it is not part of the CMIP5 core-set and is therefore replaced by the $250\,\mathrm{hPa}$ level (Taylor et al. 2012). In addition, a high amount of missing values is detected within the $1000\,\mathrm{hPa}$ level of some ESMs, which is therefore also excluded from the selection process.

The final predictor selection trees for each ANN model approach are summarised

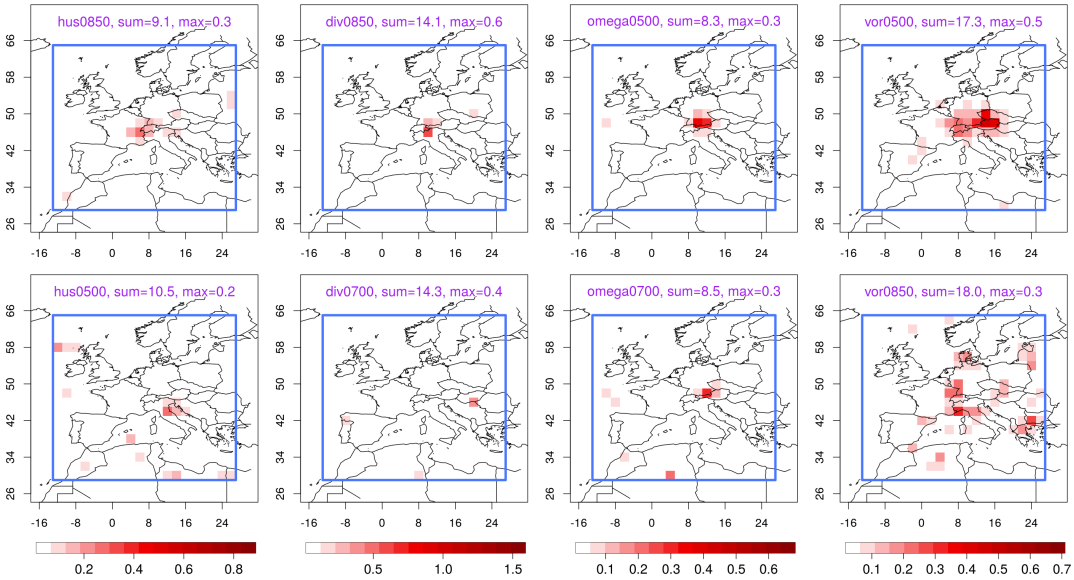


Figure 7.16: Pattern, which could not be rated satisfactorily in an automatic manner. One needs to decide, which predictor field, first or second row, should be kept. The first row shows the subjectively preferred variable, the second row displays the variable which would be selected by an automated routine based on the total sums of PaD sensitivities. Examples are taken from the predictor screening process. Hus: Step two of Sonnblick winter precipitation; div Step two of Sonnblick autumn temperature; omega: Step two Zugspitze summer precipitation; vor: Step two Zugspitze spring temperature.

in Tables A.1, A.2, A.3, A.4, A.6 and A.5 in the appendix. At the beginning of task 1, the selection of sensitive predictors, a standard domain of $-12°$ to $26°$ longitude and $30°$ to $64°$ latitude is defined for every predictor field. The importance of each predictor field is primarily assessed by the sum of the respective PaD sensitivities. The sum in this case is comparable, since an equal sized standard domain is used. Unfortunately, the process cannot be performed automatically as in some cases subjective user intervention is needed.

The first disadvantage of this approach lies in the character of predictor variables. As some variables show large sensitive areas, the sum of the PaD sensitivity can be high. Other variables are only sensitive on small regions close to the study area resulting in small sensitivity sums. Nevertheless, the latter can still add value to the SD model. Considering the maximum value instead of the areal sum can sometimes be a solution. Another disadvantage is that occasionally variables show an unlikely or physical inconsistent sensitivity pattern. This can only be identified by a subjective decision of the user. Figure 7.16 shows examples of pattern-depended decisions on the selection of the atmospheric layer. The first row represents the subjectively preferred solution, as the sensitivity pattern is more continuous and arranged at a plausible location. In all examples, except for omega, considering the maximum value would help during decision making.

Due to the long training time, especially in the beginning of the approach, several predictor fields are deleted simultaneously at one step. The elimination process is slowed down to the end, close to each minimum setup. This procedure is justified, as throughout the predictor selection process, the sensitive region of the final predictor combination faces barely any changes. Figures A.1 and A.2 show exemplary that the main focus stays on the same areas, especially at the beginning. At the end of the predictor elimination process and during the reduction of the domain size, the areas of interest can slightly change.

The final predictor combination is in addition dependent on the SD model performance. Figures A.3 and A.4 monitor the MSSS during the predictor selecting procedure. In most cases, at some point, there is a more or less severe drop in model skill or an increase in the variability of the model quality measure, indicating the loss of important information. At this point, the predictor configuration of the previous step is considered as optimal.

Finally, the domain size is decreased to the most sensitive regions following two steps. The final result of the predictor screening procedure is summarised in table 7.4. Beside hur and hus, which are mandatory, frequently selected predictors for precipitation are vorticity, divergence and omega. At Zugspitze station, wind vectors, especially vwnd seems to be beneficial to explain precipitation originating from orographic lift at the northern edge of the mountain range. Considering the predic-

Table 7.4: Final predictor setups for each predictand by seasons (left side). Predictors are specified by atmospheric level [hPa], longitude (lon, [°]) and latitude (lat, [°]) as well as influence (cont, [%]) on the ANN output. Predictor variables in columns.

			slp	thi	hgt	uwnd	vwnd	swnd	omega	air	hus	hur	umf	vmf	vor	div
Son prc	DJF	level							700		850	700			850	700
		lon							8:20		4:16	8:16			4:18	4:22
		lat							46:52		44:52	44:52			36:50	40:52
		cont							11.6 %		32.7 %	8.0 %			32.7 %	15.0 %
	MAM	level							500		700	850			850	850
		lon							2:18		6:14	4:16			-2:14	2:18
		lat							42:50		44:52	44:50			38:56	40:50
		cont							8.3 %		27.0 %	15.7 %			41.9 %	7.1 %
	JJA	level				850			700		700	700				850
		lon				-4:16			-4:20		2:16	6:16				-4:16
		lat				36:50			40:54		44:54	44:52				40:52
		cont				22.3 %			27.6 %		30.1 %	7.9 %				12.1 %
	SON	level									850	700			850	700
		lon									0:16	4:18			-8:20	-4:22
		lat									38:52	40:52			38:56	42:52
		cont									27.3 %	12.7 %			16.2 %	43.8 %
Zug prc	DJF	level					850		700		850	700				
		lon					0:14		0:16		-2:14	4:14				
		lat					42:52		42:52		38:52	40:56				
		cont					27.3 %		24.3 %		19.4 %	28.9 %				
	MAM	level				850	850				700	700				
		lon				-2:16	-4:16				-4:16	6:12				
		lat				40:50	38:50				40:52	44:52				
		cont				27.9 %	37.7 %				26.5 %	7.9 %				
	JJA	level							500		850	500			850	850
		lon							-12:16		0:14	0:14			0:20	0:14
		lat							44:52		42:52	42:52			42:52	40:54
		cont							12.5 %		27.5 %	14.2 %			22.3 %	23.5 %
	SON	level				850			700		850	700			850	700
		lon				2:12			6:20		2:12	6:16			-4:24	6:16
		lat				44:48			44:52		42:52	42:52			38:52	40:50
		cont				9.2 %			8.2 %		17.7 %	8.8 %			40.6 %	15.5 %
Son tmp	DJF	level								850					850	850
		lon		8:18						6:14					-4:20	8:20
		lat		44:52						42:50					38:50	42:50
		cont		21.2 %						16.7 %					15.0 %	47.2 %
	MAM	level								850					500	850
		lon		4:18						4:16					4:20	8:16
		lat		42:54						44:50					42:54	44:52
		cont		25.1 %						32.2 %					23.8 %	18.8 %
	JJA	level								850					500	
		lon		6:18						4:18					8:24	
		lat		44:54						42:52					40:50	
		cont		25.1 %						35.7 %					39.3 %	
	SON	level								850					850	
		lon								4:18					6:22	
		lat								42:50					36:50	
		cont								62.0 %					38.0 %	
Zug tmp	DJF	level					850			700						850
		lon		6:16			2:20			6:16						2:14
		lat		46:52			40:52			44:52						42:50
		cont		19.0 %			41.3 %			18.1 %						21.6 %
	MAM	level								850					500	
		lon		4:14						0:14					4:18	
		lat		44:52						42:52					42:52	
		cont		24.3 %						28.8 %					46.9 %	
	JJA	level								850					850	850
		lon		4:16						0:14					0:22	0:16
		lat		46:52						44:52					40:52	42:52
		cont		17.9 %						21.4 %					21.3 %	39.3 %
	SON	level					850			850					500	
		lon					2:16			-4:14					0:18	
		lat					42:50			38:56					44:56	
		cont					22.8 %			29.9 %					47.3 %	
Zug rhum	DJF	level				850	850				850	700				
		lon				0:20	-4:16				0:16	4:20				
		lat				38:58	38:54				40:54	40:56				
		cont				26.0 %	25.0 %				26.4 %	22.6 %				
	MAM	level				850	850				700	700				
		lon				-2:12	0:12				6:18	6:18				
		lat				40:48	40:50				44:54	44:52				
		cont				24.5 %	28.4 %				25.7 %	21.5 %				
	JJA	level										700			850	
		lon										4:14			-4:16	
		lat										46:52			42:52	
		cont										23.9 %			76.1 %	
	SON	level										700			850	
		lon										4:16			-6:18	
		lat										44:50			38:52	
		cont										26.0 %			74.0 %	
Zug wnd	DJF	level						850								
		lon						0:16								
		lat						42:52								
		cont						100.0 %								
	MAM	level						700								
		lon						4:16								
		lat						42:52								
		cont						100.0 %								
	JJA	level						700								
		lon						2:18								
		lat						42:52								
		cont						100.0 %								
	SON	level						700								
		lon						2:16								
		lat						42:52								
		cont						100.0 %								

tand temperature, atmospheric layer thickness, vorticity and divergence represent the most important predictor combination in addition to the obligatory temperature field. Relative humidity at Zugspitze station is divided into two main setups: Winter and spring setups use vwnd, uwnd and hus, while summer and autumn simply add divergence to the mandatory hur field. Finally, for wind speed at Zugpitze, large-scale wind speed seemed to be sufficient as predictor field. For almost any setup of both stations, the primarily level fields of interest are the lower ones: $850\,\mathrm{hPa}$ and $700\,\mathrm{hPa}$ level are most frequently selected. Considerably seldom the $500\,\mathrm{hPa}$ level asserts. The upper tropospheric level at $250\,\mathrm{hPa}$ can be considered as least influential: It is not used in any final combination and was usually excluded during the predictor selection in one of the first steps. Further noticeable is the quick elimination of slp and hgt during the predictor selection process. Both are the most frequently used predictors in statistical downscaling studies (see section 4.2.3). The fast exclusion however, does not necessarily mean that pressure fields are bad predictors, however it seems to be beneficial to process the raw pressure fields to gain additional information. After all, tropospheric wind vectors, divergence and vorticity directly depend on or are calculated from the pressure field. Further, the preference of wind velocity, divergence or vorticity can emerge from ANNs to be able to detect non-linear relationships. These predictors can bear an added value, which the often used linear models are not able to access.

7.6 Tuning of the Statistical Downscaling Models

7.6.1 Artificial Neural Networks

Comparison of Different Training Algorithms

In general one can differ two main branches of training algorithms: Methods based on the first-order or second-order partial derivatives (Haykin et al. 2009). First-order optimisation methodologies are represented by the standard back-propagation algorithm as well as its heuristics, all relying on the first-order derivative of the of the cost function. In contrast, second-order optimisation methods use the second-order derivative of the cost-function, called Hessian matrix. The Hessian is complex and computational expensive to calculate and can only be approximated under certain circumstances. The advantage of second-order optimisation methods is a very fast training progress typically finishing within a couple of iterations (Burney et al. 2004). The fast progress has to be counterbalanced with higher computational demands. Although it is still not clear which version is preferable, recent comprehensive studies, e.g. Bengio (2012), still recommend gradient descent variants as these outperform second-order optimisation methods in most cases. Therefore, in this the-

sis, primarily first-order optimisation methods are selected. A detailed description on the computation of each algorithm can be found in section 5.5.5.

An overview over the characteristics of different training algorithm is given in the appendix for two learning tasks: tables A.7 to A.15 use Zugspitze precipitation in winter as target, while in tables A.16 to A.24, the training algorithms are applied to Zugspitze temperature in winter. To the user, the most important aspects are a fast execution time, a high resulting model quality and an easy configuration of the algorithm parameters. Note that in a study like this only a reasonable number of parameters can be evaluated. There can be better configurations, which are not taken into account here. In addition, some algorithms need to be configured by a large amount of parameters (e.g. the Extended Delta-Bar-Delta algorithm or Rprop) leaving the user to work with a complex parameter relationship, which is beyond this study to explore. If at all necessary, the focus at these algorithms lies on the most important parameter.

Rprop is a well balanced algorithm, with the fastest run time and excellent performance results, only slightly excelled by some configurations of the momentum term version and weight decay/regularisation. No noteworthy improvement of the default training parameters as those proposed by Riedmiller and H. Braun (1993) ($\Delta_{i,j}(1) = 0.01, \eta^+ = 1.2, \eta^- = 0.5, \Delta_{i,j}^{max} = 50, \Delta_{i,j}^{min} = 10^{-6}$; see also section 5.5.5 for more details) could be found in this study. Surpassing most other algorithms in usage, speed and performance, Rprop is considered as the standard training algorithm in this study. Critical seen can be the conspicuous high weights, which occur especially in the precipitation example. The Extended Delta-Bar-Delta algorithm belongs to the fastest group of algorithms, which can still reach high performance measures if the initial learnrate is small enough. The algorithm has eleven adjustable parameters, here only the learning rate η is varied, other parameters are taken from an example of Minai and Williams (1990) ($\alpha = \mu = 0.01, \theta = 0.7, k_l = 0.05, \gamma_l = 20, \phi_l = 0.3, k_m = 0.1, \gamma_m = 5, \phi_m = 0.5, \eta_{max} = 10, \mu_{max} = 0.9$). Further experiments with different momentum term start values α (in the original literature μ) usually failed (not shown here), suggesting the algorithm as rather sensible to its parameters. Due to the high failure rate, additional experiments are aborted and finally lead to the preferation of the easier to handle Rprop algorithm. Standard gradient descent is considered rather inefficient in batch mode training, leading to a large number of heuristical adaptions to improve the training process (see section 5.5.4 and 5.5.5 for a discussion on the topic). One can see the disadvantages mainly in a quite long run time (the fastest run with a comparable performance to Rprop (11 seconds run time) takes 4 minutes in the precipitation example and 1 minute (13 seconds Rprop) on temperature. Training can be sped up with a large learning rate, however, these models tend to have less predictive skill. Small learning rates rapidly increase the training time:

Considering a η of 0.001 in the precipitation example, seven of ten runs did not finish within the 3 hour training time limit. None of them did in the temperature example. The interrelation between the learning rate η, run time and model quality shows the necessity of a well tuned η. In contrast, adaptive learning does not strongly depend on the initial η as there is no clear difference in performance or run time found (other parameters are set to $\eta^+ = 1.2, \eta^- = 0.5, \eta_{min} = 10 * *(-6), \eta_{max} = 50$). The resulting model skill reaches almost or equally as high quality as Rprop. Moreover, adaptive learning is characterised by rather quick run times. Overall, adaptive learning can be viewed as a reliable choice with easy handling and can be categorised close to Rprop. Individual learning rates tend to act sensibly and with no clear tendency when comparing performance and initial learning rates especially in the example temperature (other parameters are set to $\eta^+ = 1.2, \eta^- = 0.5, \eta_{min} = 10 * *(-6), \eta_{max} = 50$). In addition, when starting from unsuited learning rates, the model quality can drop quite heavily. The advantage of individual learning rates is a very fast run time, in some cases comparable to Rprop. Similar to Rprop using individual learning rates leads to quite large weights in comparison to the other algorithms. The most important aspect of training with decaying η is to select a slow cooling rate in combination with a fast learning rate. Fast cooling rates seem to prevent a successful training process as resulting model performance drops rapidly, in some cases training even fails. In comparison to other algorithms the final model quality is not competitive within the tested parameter combinations. In addition, the run time of the algorithm in the more successful cases can be quite slow. To get a final conclusion on the algorithm, further parameter testing would be necessary. Modifications of the algorithm, for example with a non-linear decrease in η, for example by using the law of (radioactive) decay, might also improve the training process. Although an extensive parameter search is needed and quite long training times are common of all tested parameter combinations, the momentum term version is a good alternative to Rprop, when it comes to model performance. The momentum version is one of the few algorithm in this survey, able to outperform Rprop. Best values for η in both examples are found starting at 0.1. Too small learning rates (already setting in from $\eta = 0.01$) increase the run time heavily. α should not be larger than 0.5 to ensure high quality models. Both regularisation techniques L1 and L2 are able to produce models with slightly higher predictive skill than Rprop. This is not surprising, as regularisation is developed to enhance the predictive skill of models. In general, high model skill comes hand in hand with comparable long run times when using regularisation techniques. Both algorithms show best results when the decay parameter d is very small (0.001 or lower), similar to the other algorithms, η should not be too large (below 0.5). The most interesting aspect of this algorithm is the resulting weight magnitudes. A very low fraction of weights can be viewed as low weights (swf is typically below $10\,\%$).

This behaviour indicates that weight decay results in a large fraction of meaningful weights/connections, as a weight close to 0 often means no fundamental information transfer by this connection (see e.g. section 5.5.3). Further concluding, regularisation makes use of the whole input pattern while other algorithms tend to focus on certain areas. This might be beneficial in the task of projecting ESMs, as SD models might more likely fulfil the requirement of the "skilful scale". On the other hand, this characteristic can reduce the possibility to find a good predictor setup and domain with sensitivity studies, where the delineation of certain most influential regions and/or predictors is the aim. This conclusion must be verified in future studies. In addition, the minimum and maximum magnitude of weights are comparable small. ANNs based on small weights are less prone to overfitting (see section 5.5.6).

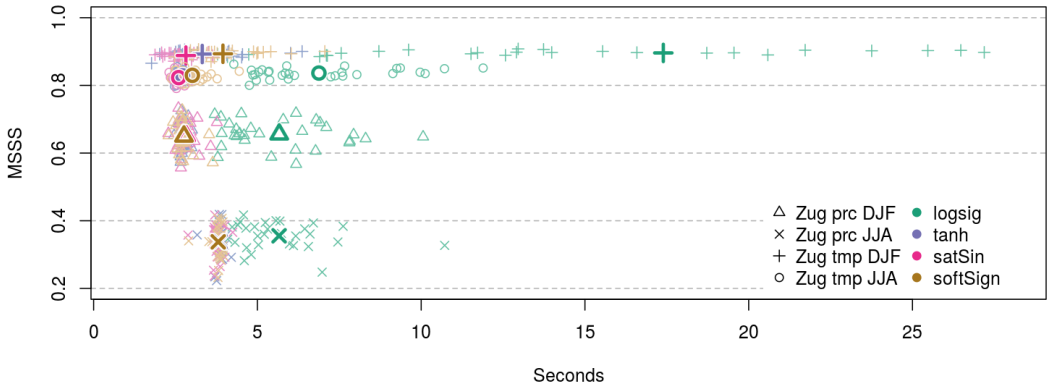
In summary, when comparing the nine algorithms of this thesis, the most simple to use, fast, but still leading in resulting model skill is the Rprop algorithm. Various studies support these findings: Fernandes et al. (2013), Günther and Fritsch (2010), Almeida et al. (2010), and Adamowski and Karapataki (2010) and Ahmad et al. (2008) emphasise the high speed and resulting model quality of Rprop. Although most training methods reach similar high model performances in certain parameter settings, either extensive parameter testing or long run times need to be considered. In this study, the only weaknesses of Rprop are the resulting large weights and the associated risk for overfitting. Rprop is usually located at the top ranks of studies comparing learning algorithms (e.g. Azami et al. (2011)). Most authors provide no/limited information on parameter settings, Adamowski and Karapataki (2010) suggest a initial learning rate of 0.01 as best starting condition in their case, a value, which also showed good performance in this study. Beside the thesis-specific R package gecco, a package by Günther and Fritsch (2010) called "neuralnet" is available on the official CRAN sources, whose training methodology is based on different implementations of Rprop. The best parameter combinations found in this survey are not necessarily transferable on different learning tasks, but need to be optimised for each case separately (e.g. Lippe (2006)). One can see in comparison of both training targets, temperature and precipitation that the best parameter configuration shows a similar tendency but is not always identical. The survey should therefore rather provide a general starting point from which one can begin the parameter grid search in a statistical downscaling framework. It can be necessary to expand the parameter grid to find the ideal setup. Last but not least, an interesting aspect can be to combine different training algorithms. For example, one can initiate the training procedure with a fast algorithm like Rprop, but finish the training with regularisation to reduce the weight magnitude, which in return could further improve the generalisation capability. This could be an interesting topic for future research.

Transfer Functions

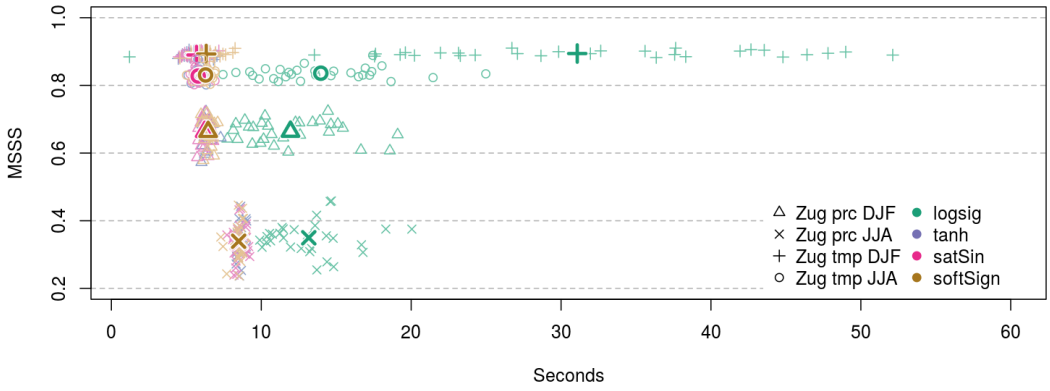
Figure 7.17 compares the effect of different transfer functions on training speed and model quality (see section 5.5.1 for more details on the applied transfer functions). On the one hand, transfer functions symmetrical to the origin (tangent hyperbolicus (tanh), saturating sinus (satSin), and the soft sign (softSign) function) show barley any difference in the mean training speed. Largest variations can be identified modelling winter temperature. Here, saturating sinus finishes first, followed by the tangent hyperbolicus, while softsign is slightly slower than the others. On the other hand, a clear difference in training speed is found considering the logistic transfer function (logsig). This issue is also described in literature. Glorot and Bengio (2010), for instance, advise against the usage of the logistic transfer function if the aim is a fast training speed. LeCun et al. (1998) showed that the slowdown of the training progress by logistic neurons is caused by their non-zero mean (the codomain of the logistic transfer function lies between]0,1[). Nevertheless, the logistic transfer function outcomes are characterised by a slightly higher MSSS. In this study, a longer training time was accepted with the objective to gain a higher model quality. Therefore, the logistic transfer function is selected.

Early Stopping

In addition to the theoretical examples figure 7.18 displays the actual error behaviour during the ANN training progress. The error (MSE) of the training, validation (early stopping sample) and testing set is calculated during optimisation using standardised and scaled (range 0 to 1) data pattern. The MSE curves of Zugspitze winter temperature and precipitation are characterised by a smooth and rapid decrease during the first optimisation steps. After the first few hundred iterations, the MSE reduction lessens and further progress is hardly noticeable. More apparent is the further development of the MSE in the precipitation example. Subsequently, after the validation minimum is reached, the validation error is slightly increasing again, while the training error is still decreasing. In early stopping, this characteristic is treated as evidence for overfitting of the model onto the training samples. Since the validation minimum can be a local one, the training is continued for an user defined amount of checks, here 300. When no further decrease of the validation error is found in the check iterations, the training state of the ANN is set back to where the MSE of the validation period was in its minimum. An adequate number of checks depends on various conditions, for example the adjustment speed and certainty of the algorithms updates. Fast learning methods with stable updates need fewer checks. Considering the error development of various examples, including runs with up to $3 \cdot 10^4$ checks (not shown here), a number of 300 checks is identified to be sufficient for the Rprop algorithm. Figure 7.18 already indicates in both cases a flat plateau with no further

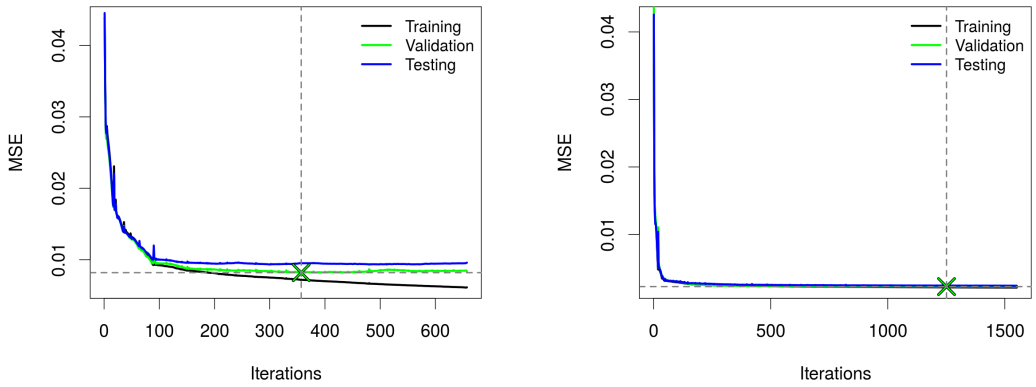


(a) 8 hidden neurons



(b) 20 hidden neurons

Figure 7.17: Scatter plot comparing four different Sigmoid transfer functions in training speed (in seconds) and model quality by the MSSS of the validation period. Predictands are Zugspitze (Zug) temperature (tmp) and precipitation (prc) in winter (DJF) and summer (JJA). Thick marks represent the mean of the 30 fold cross-validation approach, thin marks each element. Samples of the cross-validation procedure are trained with equal starting conditions (identical training set, identical initial weights), only the transfer function changes. Examples in figure a) is trained with eight hidden neurons, examples in figure b) with twenty.



(a) Precipitation **(b)** Temperature

Figure 7.18: Early stopping examples with the learning tasks: Zugspitze winter a) precipitation and b) temperature. Data separation: 60 % training, 20 % validation (early stopping) and 20 % testing. Final predictor configuration is trained respectively with 8 neurons. Minimum of the validation MSE is marked by the green cross. Number of checks: 300.

noticeable gradient of the MSE curve after a few hundred iterations. Even though it cannot be ruled out completely, a strong additional reduction in the validation MSE is not expected.

Number Of Hidden Layer

In a standard fully-connected feed-forward ANN topology two major structure elements need to be defined. First of all, the number of hidden layers and second, the number of neurons in each hidden layer. Theoretically it can be proven that an ANN with one hidden layer and non-linear transfer functions in the hidden neurons can approximate any common equation in a very precise and satisfactory sense (Hornik et al. 1989). Thus, most downscaling studies rely on one hidden layer, e.g. Gaitan et al. (2014), Moya Quiroga et al. (2013), Coulibaly et al. (2005), Trigo and Palutikof (1999), and Wilby et al. (1998) and Castellanos et al. (1994). Huth et al. (2008b) and Tolika et al. (2007) explicitly point out that no improvement in model quality is found with more than one hidden layer in their SD framework. Only few studies, e.g. Chadwick et al. (2011) or Azadeh et al. (2011) use more than one hidden layer. None of these studies uses greedy pre-training, a fairly new technique, which is applied with great success in deep learning tasks. Greedy pre-training allows a stable training of ANNs with several hidden layers. In this study, the possibilities of ANNs with up to three hidden layers are exemplarily evaluated. As predictand serves winter precipitation of the Zugspitze station, the only differences between the ANNs is the number of hidden layers and respective hidden neurons. Figure 7.19 summarises the results of the experiment. The best model quality is reached during training as

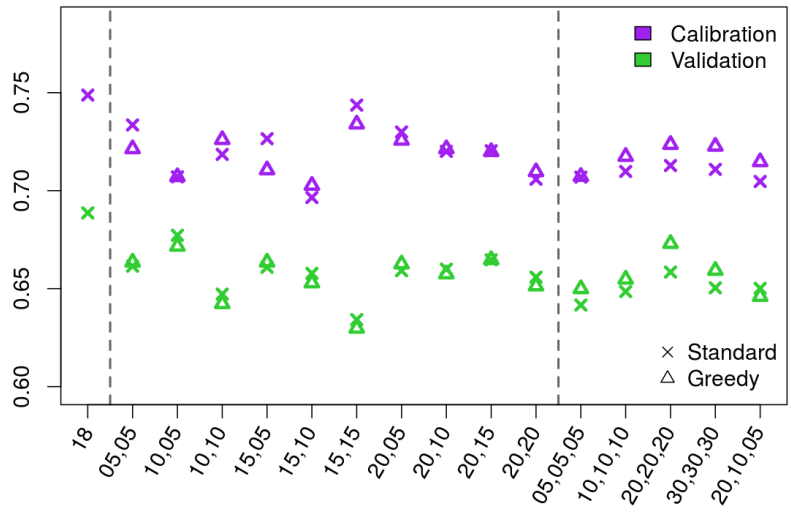


Figure 7.19: Comparison of different numbers of hidden layers with a varying amount of hidden neurons by their model skill, here the MSSS. Number of hidden neurons are listed in the bottom for each layer. First number represents the neurons of the first hidden layer, additional layers are appended, separated by commas. ANNs are trained in standard and greedy mode (20 % denoising). As reference serves a single hidden layer network with 18 hidden neurons, which represents the selected final ANN configuration for the learning task of Zugspitze winter precipitation (see next section as well as table 7.5). Each MSSS notation in the figure represents the mean of a 10 fold cross-validated approach.

well as validation by a standard single hidden layer ANN with 18 hidden neurons. Although, in theory, with access to much more predictive power, larger ANNs perform worse. The main reason for the low quality of the larger ANNs might be the limited amount of training data. This is indicated by the lower skill during calibration in comparison to the single-hidden-layer-ANN. Further, in some cases, a large gap between calibration and validation performances can be noticed, which is a sign for overfitting. Greedy pre-training has little effect on ANNs with two hidden layers. The resulting model quality is not necessarily better. In contrast, the largest ANNs with three hidden layers benefit from pre-training in both calibration and validation performance in any but one case. This characteristics suggest that models with two hidden layers do not yet strongly benefit from pre-training the weights. Considering the results, in the following, ANNs are constructed with one hidden layer.

Number Of Hidden Neurons

To find the optimal amount of hidden neurons, different numbers are iteratively tested. An example from a preliminary study, visualised in figure 7.20, shows the general behaviour of an ANN with increasing numbers of hidden neurons. ANNs are trained individually for every month with different numbers of hidden neurons. The aim is to reproduce the Zugspitze temperature time series, as predictor serves hgt $500\,\mathrm{hPa}$. On the one hand, with more hidden neurons a better performance is measured considering the training period, as the training algorithm has more possible degrees of freedom to adapt the network to the training dataset. On the other hand, the generalisation capability of the ANN is suffering from large numbers of hidden neurons, when assessing the validation period. An explanation is the increasingly higher risk of overfitting with larger numbers of hidden neurons. However, too small numbers of hidden neurons are insufficient to capture the complexity of the problem. The right amount of hidden neurons can be identified by the peak of the validation performance, which typically lies between six and fifteen hidden neurons. Although there is a clear tendency, the correlation differences between the evaluated and reasonable numbers of hidden neurons are not very high in the validation period.

In regard to the previous example numbers of hidden neurons are searched for the final ANN model approaches. Detailed evaluation of ANN performance in regard to the number of hidden neurons is given in the appendix in figure A.5. A sequence of hidden neurons from two to 30 with step size one and from 35 to 50 with step size 5 is evaluated during 30 cross-validated runs in the time period 1970 to 2000 for the seasonal models. The number of tested hidden neurons goes hand in hand with various downscaling studies, where optimal numbers of hidden neurons are identified below 30, in most cases between 10 and 20 (e.g. Gaitan et al. (2014), Azadeh et al.

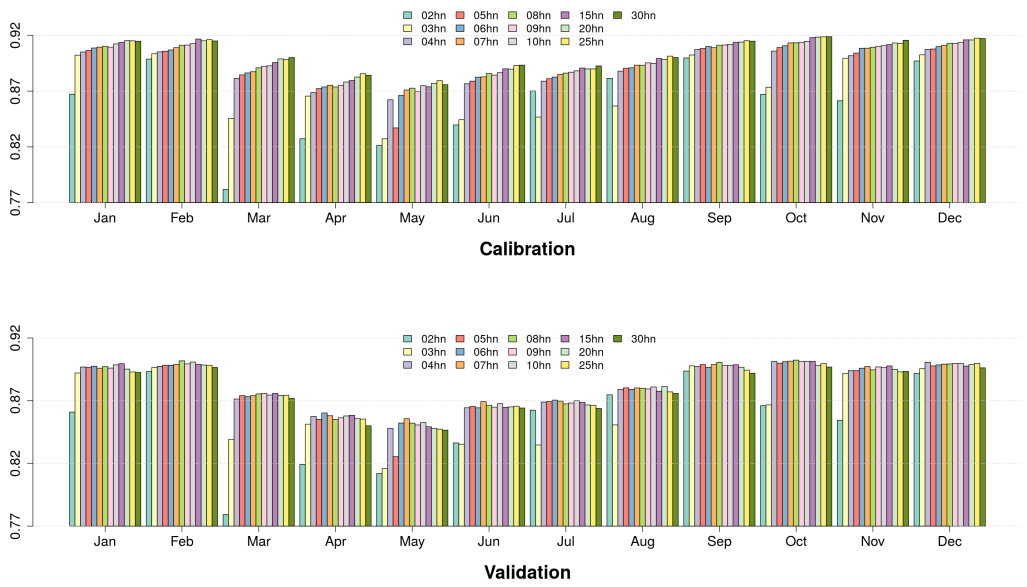


Figure 7.20: Mean Spearman correlation coefficient between observation and simulation of 30 cross-validated ANNs reached by different numbers of hidden neurons (hn) arranged in one hidden layer. As predictand serves Zugspitze temperature and as predictor hgt on the 500 hPa level. SD Models are calibrated and validated for each month separately. Time period is 1901-2010, 20 % of the data samples are taken for validation.

Table 7.5: Optimal number of hidden neurons for an ANN with one hidden layer under consideration of the final predictor setup (see Table 7.4), selected by an automated screening process.

	DJF	MAM	JJA	SON
Son prc	17	24	29	6
Zug prc	18	6	10	9
Son tmp	21	13	26	15
Zug tmp	28	10	22	5
Zug rhum	11	8	30	27
Zug wnd	17	29	12	8

(2011), Huth et al. (2008b), Tolika et al. (2007), Coulibaly et al. (2005), and Castellanos et al. (1994)). In this study, the best number of hidden neurons is automatically evaluated and selected during the optimisation process by comparing the mean of the Spearman correlation coefficient of the cross-validated runs. The final selection is summarised in table 7.5. Despite being tested, large numbers of hidden neurons above 30 are excluded from the selection process, as the high amount of parameters that comes in hand, is estimated to be too large for the limited amount of training samples to ensure a complete elimination of overfitting. This hard limit was only triggered once for winter temperature at Zugspitze, where 50 hidden neurons perform slightly better as the selected 28. In general, during the validation period, the differences in mean or median are very small with regard to different hidden neurons. In contrast, the variability of performances is quite large, which results e.g. from the cross-validation process or different ANN parameter initiations. There is no clear predictor set (predictor type or input size), seasonal-or predictand-specific tendency on the number of hidden neurons. A similar behaviour is also be witnessed by Huth et al. (2008b), where the best hidden layer neuron configuration varies between seven and eighteen without any clear predictor dependency.

7.6.2 Reference Class Forecast

Number Of Classes

The number of classes in a reference class forecast approach defines the number of events, which can be distinguished and thus, the number of reference values that can be assigned. The number of classes should be neither too low, as a low number is not able to represent a comprehensive overview of circulation types, nor too high, as differences between classes begin to vanish. Considering only the classification process, the aim is to find a number of classes with maximum internal homogeneity and at the same time maximum external dissimilarity (Beck and Philipp 2010; Huth

et al. 2008a). An additional, and further, the most important requirement in a RCF approach is a good forecast skill of the model. Therefore, in this study, the optimal number of classes was identified by the best model forecast performance during the validation procedure. Figures A.6, A.7 and A.8 in the appendix give an overview of the validation MSSS derived from the mean of fifteen Monte-Carlo sampled cross-validation runs, considering two, three, nine, 18 and 27 classes. The latter are, with little variations, common class numbers in literature and cover the typical class range in weather type applications (e.g. Philipp et al. (2014b), Beck et al. (2013), Lutz et al. (2012), Demuzere et al. (2011), Kysely (2007), and Goodess and P. Jones (2002)). The low class numbers (two and three) of this study are not suited for a RCF approach, but are included in the graphics, as these will come in hand in the NNC approach (see chapter 7.6.3). The highest model skill is reached with 27 classes throughout all predictands and seasonal subsets. A maximum limit was set to 27 classes as differences between centroids were starting to fade. Therefore, in the RCF approach, the number of classes/types is set to 27.

Conditioning Weights Of The Predictand

The feature selected by the cluster optimisation to discriminate classes within the predictor set might not always be the perfect choice, when the aim is to reproduce local climate. Lutz et al. (2012), for example, point out that unconditioned clustering often produces predictand-independent circulation pattern, which can have greater predictand variability within such pattern than across different pattern. To point the algorithm in the right direction, the classification process can be conditioned by including the weighted predictand (Beck et al. 2014; Lutz et al. 2012). During simulation the predictand is excluded, new datasets are compared to a centroid purely built from the predictor set, thus the atmospheric conditions. The weight needs to be selected carefully, as too large weights induce the cluster algorithm to basically ignore the atmospheric components, which in return greatly decreases the predictive power of the model, as the model must perform the simulation step solely based on the state of the atmosphere. In that case, analogue to above, the predictor variability of single observations within each class increases, while differences across classes vanish. In this thesis, weights are evaluated during an extensive grid search procedure considering the following magnitudes and step sizes: step size 0.001 between 0 and 0.2; step size 0.01 between 0.2 to 1; step size 0.1 between 1 to 10; step size 0.5 between 10 to 20; step size 5 between 20 and 30; step size 10 between 30 and 50. The top five weights for every predictand and season are marked in figures A.6,A.7 andA.8 in the appendix. The final selection of weights is listed in table 7.6.

Precipitation models for both stations, all seasons and all class numbers show best validation results with low predictand weights, in most cases below one. A simi-

Table 7.6: Best predictand weights for the optimisation process.

	Son prc	Zug prc	Son tmp	Zug tmp	Zug rhum	Zug wnd
DJF	0.1	0.09	4.9	3.9	0.124	0.031
MAM	0.014	0.008	0.67	0.7	0.012	0.016
JJA	0.051	0.067	0.046	50	0.016	0.007
SON	0.042	0.087	5.3	30	0.055	0.034

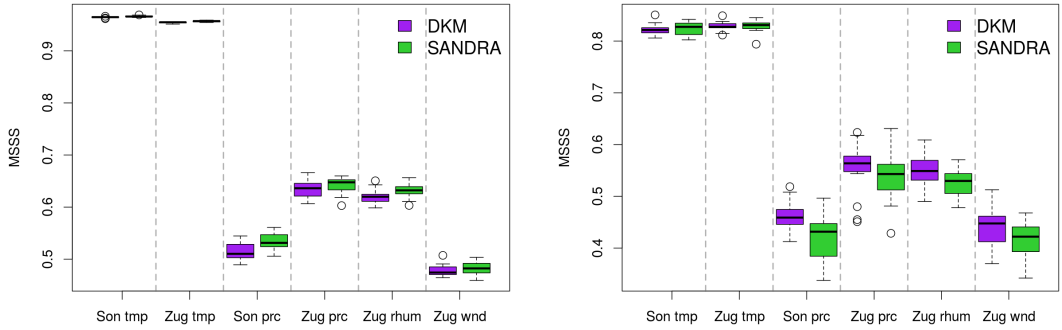
lar behaviour can be witnessed for wind speed, where only very small weights are applicable. Relative humidity, although equal in the overall tendency, shows an exception in winter, where in general larger weights are selected, especially in classifications with lower type numbers. The selection of primarily small weights indicates that the classification of the predictand only would lead to a quite different grouping of the observational time steps, than classifications exclusively build on the predictor sets. That means that features, responsible for the classification of types in both predictand and predictor sets are kind of dissimilar. This is also confirmed by the low skill of models for these predictands. This problem is not only restricted to the RCF methodology, but caused by the complex relation between the large-scale atmospheric condition and their local influence on the mentioned predictands.

A different picture is drawn considering predictand weights for temperature models. In general, the primary skill gradient is determined by the number of classes. Within each class, good model skills can be gained with small and very large weights, which means that the relationship between predictors and predictand is very good and similar features would be selected when classifying both separately. Including the predictand has no large impact on the assignment of classes. This is not surprising, as one of the predictor variables is large-scale temperature, which has a good relationship with local temperature conditions. The good relationship between the large-scale atmosphere and local-scale temperature leads to a comparatively high model skill.

Further can be noticed that in most modelling tasks, the optimum weight magnitude tends to decrease with the number of classes.

Predictor Weights

In this thesis, every predictor is weighted by the number of input grid cells, to ensure an equal consideration within the euclidean distance during the optimisation process (see section 5.6.1). A second approach, investigated in this study, is to weight each predictor by the respective cumulative sensitivity derived from the partial derivative sensitivity method (see 7.5.3). There was no significant increase in model quality witnessed (not shown here), therefore this practice was discarded.



(a) Calibration **(b)** Validation

Figure 7.21: Comparison of the DKM and SANDRA cluster analysis during calibration and validation for winter (DJF). A cooling rate of 0.99 is defined for SANDRA. Boxplots show the distribution of the MSSS within a 15-fold cross-validation approach. All classification tasks are optimised including the weighted predictand (see table 7.6 for weight magnitudes).

Optimisation Algorithm

Properties of DKM and SANDRA are compared in their optimisation characteristics and resulting skill. Large differences are found in the computer run time. The clustering process of the DKM in comparison finishes fast, while SANDRA is a rather slow algorithm, if done conscientiously. For example, in this study, the implemented version of SANDRA needed on average 1.5 hours (cooling rate: 0.99, number of repeats: 1000) to optimise one sample of Zugspitze precipitation, while DKM finishes in one second. Figure 7.21 compares both clustering algorithms in their model skill, during calibration and validation of the Zugspitze and Sonnblick predictands in winter. On the one hand, in the calibration period, SANDRA shows in general a higher model skill, but on the other hand, the validation performance, except for winter temperature, is lower. The reasons can be that the weights derived from the DKM grid search are not optimal for the SANDRA cluster algorithm. An indication is the higher skill during calibration and lower skill during validation of SANDRA, which might be a sign for overfitting. Further, Beck and Philipp (2010) conclude that high performance in separating classes does not necessarily go hand in hand with high predictive skill. Considering the fast run time and good validation skill, in this thesis, DKM is selected as the standard clustering algorithm.

7.6.3 Tuning of the NNC Approach

Model combinations are calibrated considering $80\,\%$ of the input dataset time steps, while $20\,\%$ are separated randomly for validation. To assess the generalisation capability, 15 cross-validation runs are realised. Classification is performed using the DKM algorithm. ANNs are optimised by Rprop. $20\,\%$ of the class elements are re-

served for early stopping with 300 checks. A maximum of three types is defined for the CTC to reduce the risk of an insufficient number of training pattern within a class for the successive ANNs and a number of two to 30 hidden neurons, arranged in one hidden layer, is evaluated in a full grid search procedure. The number of hidden neurons with the highest MSSS for each class in the validation period is selected. The best CTC and ANN combination is identified by the highest MSSS in the validation period of the recombined time series, basically deciding between the configuration with two or three classes. Classification can be either performed unconditioned or conditioned. In case of unconditioned classification, most similar circulation types will serve as ANN input, though a higher withintype variability of the predictand can be expected. In case of conditioned classification, typically a higher withintype variability of the predictors follow suit with a more narrow predictand distribution (see section 7.6.2). Accordingly, the main effects are different link conditions and learning tasks for the ANN. In this thesis, the conditioned approach is used.

Table 7.7 lists parameters and properties of the best model combinations. Classes are sorted by their magnitude of the reference value from large (class one) to small (highest class number two or three). In general, predictor weights are rather small for both two and three type classifications, in most cases below unity. Larger weights need to be applied at Sonnblick temperature or Zugspitze temperature in autumn. One outstanding exception, with a tendency to very large weights, is relative humidity at Zugspitze in winter (see figure A.8). A comprehensive discussion on predictor weights is given in section 7.6.2. The number of elements per class varies between 437 (Sonnblick, precipitation, winter, class 1) and 1756 (Zugspitze, precipitation, winter, class 2). No conspicuous features in model skill occur during the calibration or validation period, therefore the number of training pattern is considered to be sufficient. Similar to section 7.6.1, no clear pattern can be found for the optimum amount of hidden neurons. Noteworthy are the low amounts of hidden neurons in several runs. In these examples, the relationship between predictors and predictand seems to be simplified, so that ANNs with low complexity are sufficient, only Zugspitze temperature in spring selects the predefined hidden neuron cap. The skill of weather type specific ANNs does not vary too strong between classes. On the one hand, ANNs for Zugspitze precipitation or wind speed show often better skills for classes with higher magnitudes. On the other hand, considering relative humidity, classes with lower values are modelled better. In these cases it seems to be apparent that classes, which most frequently include a fixed value limit, e.g. no precipitation or a relative humidity of $100\,\%$, have the tendency to show a lower skill as the value distribution is more complex. Examples, which do not accord to this hypothesis are Sonnblick precipitation (winter and autumn) as well as relative humidity and wind speed during winter time at Zugspitze. ANNs for temperature at both stations tend

Table 7.7: Parameters and properties of NNCs: Predictor weight (wgt) of the conditioned classification, number of elements per class (cl), hidden neurons and MSSS of the validation period for the class-specific ANNs.

		Wgt	Elements per Cl			Hidden neurons			ANN MSSS (val)		
			Cl 1	Cl 2	Cl 3	Cl 1	Cl 2	Cl 3	Cl 1	Cl 2	Cl 3
Son prc	**DJF**	1.1	437	915	886	5	9	5	0.2	0.3	0.4
	MAM	0.1	643	857	781	4	7	23	0.5	0.4	0.3
	JJA	0.2	589	805	887	15	27	25	0.1	0.4	0.1
	SON	0.2	653	776	827	17	23	2	0.2	0.2	0.6
Zug prc	**DJF**	0.6	482	1756	x	2	2	x	0.5	0.4	x
	MAM	0.2	734	757	790	17	26	3	0.3	0.4	0.1
	JJA	0.3	466	925	890	27	5	2	0.2	0.2	0.2
	SON	0.7	1019	1237	x	28	2	x	0.5	0.3	x
Son tmp	**DJF**	8.9	1411	827	x	3	6	x	0.6	0.7	x
	MAM	2.5	1280	1001	x	15	9	x	0.8	0.9	x
	JJA	1.9	1262	1019	x	2	9	x	0.6	0.7	x
	SON	0.2	841	898	517	2	19	25	0.8	0.8	0.9
Zug tmp	**DJF**	0.1	1361	877	x	2	11	x	0.7	0.8	x
	MAM	0.8	1235	1046	x	2	30	x	0.8	0.9	x
	JJA	0	1349	932	x	2	9	x	0.7	0.7	x
	SON	4.5	1380	876	x	2	15	x	0.8	0.9	x
Zug rhum	**DJF**	40	1502	736	x	6	24	x	0.5	0.4	x
	MAM	0.1	729	893	659	21	2	21	0.3	0.4	0.5
	JJA	0.2	1226	1055	x	4	13	x	0.4	0.5	x
	SON	0.1	1245	1011	x	22	29	x	0.4	0.5	x
Zug wnd	**DJF**	0.1	653	1152	x	20	18	x	0.2	0.3	x
	MAM	0.2	677	1162	x	27	7	x	0.1	0.1	x
	JJA	0.1	811	1028	x	23	6	x	0.1	0.1	x
	SON	0.1	690	1129	x	22	20	x	0.3	0.1	x

to perform better with lower values. With the exception of Sonnblick temperature in autumn and Zugspitze relative humidity in spring, combinations based on three classes prevailed primarily in precipitation tasks. In most instances, a grouping into two classes is superior.

7.7 Calibration and Evaluation of the Final Statistical Downscaling Models

SD models are finally trained in a 15-fold Monte-Carlo cross-validation procedure using the identified predictor setups and model/training configurations. The final model building results will be henceforward referred to with the keyword "final" (e.g. final predictor setup).

An overview of each SD model performance and predictive skill is given in figure 7.22. Evaluation is performed considering Spearman correlation, MSSS, MAE and RMSE of the final time series product with original variable units. Focus lies on the validation period, as in climate change studies SD models need to have a good generalisation capability. All models are calibrated and validated on the 20th century reanalysis dataset and local observations between 1970-2000, except wind speed, for which the available dataset starts in 1976. The calibration period is primarily determined by a running calibration procedure, whose results will be described in detail in the respective predictand chapters (e.g. section 8.3 for precipitation). In general, in the validation period, with few exceptions, ANNs perform best followed by NNCs, while RCF show the lowest skill. Nevertheless, as the RCF is by far the most simple one of the three methods, its performance measures can keep up quite well in most cases. RCF models for wind speed at Zugspitze are often superior to NNCs, most obvious in correlation and MSSS. The ANN approach gets only slightly outperformed at Zugspitze precipitation and relative humidity in autumn by the combined approach in both cases. In addition, the difference between calibration and validation should not be too large, as this could be a sign for overfitting. ANNs usually show only little differences. In comparison, RCF models can exhibit quite large gaps, especially in terms of temperature. Typically, Zugspitze precipitation and temperature models show better correlations and MSSS in the validation period than Sonnblick in all instances. MAE and RMSE show contrary results, but error measures for different tasks are hardly comparable. Most models show seasonal variations in skill, except for wind speed, whose performance does not strongly alter throughout the year. Precipitation models perform best in winter and worst in summer. Differences likely occur from variations in precipitation genesis, which is primarily advective in winter and convective in summer. Smaller convective cells cannot be represented in

large-scale datasets, leading to a decline in the relationship between predictor and predictand. This effect is frequently described in downscaling articles, e.g. Maraun et al. (2010). In between lies the model quality of the transitional seasons spring and autumn. Models for both seasons show similar predictive skill. Considering temperature, there is a slight, but recognisable decrease in predictive skill in summer. Model quality for relative humidity can be divided into winter and autumn, which bear nearly equally good results, while spring and summer perform a little worse. Comparing the performance of the different predictands, temperature achieves the highest MSSS, followed in the order by relative humidity, precipitation and wind speed. Correlation largely follows the characteristics of the MSSS results, although wind speed variability is better represented than precipitations in summer and equally in autumn. An in-depth evaluation of specific time series characteristics and their representation in the SD model output will be given in chapter 8 for precipitation, in chapter 9 for temperature, in chapter 10 for relative humidity and in chapter 11 for wind speed.

7.8 Summary

Statistical downscaling models are developed seasonally for four predictand time series at Zugspitze and two at Sonnblick, covering temperature and precipitation at both stations and additionally relative humidity and wind speed at Zugspitze. To ensure time period, and in case of ANNs in addition starting partition independent, configurations are at least 10 times Monte-Carlo cross-validated. In climate studies, a good generalisation capability of the SD models is crucial. Therefore, a high predictive skill in the validation period is of major concern. The two main parts of model construction are: First, the application of predictor screening techniques help to find an informative predictor setup. Second, SD model parameters are tuned for each predictor setup.

Predictor screening is performed with ANNs. In a pre-study, four different predictor screening techniques are evaluated, as there are no satisfying recommendations in current literature: Random grid search, a correlation based method, automated domain generation and three methods with the aim to assess the sensitivity of ANNs to the provided input variables. Initially, predictor variables are selected following random grid search and correlation coefficients. Although resulting predictor variable combinations are quite similar, partly large differences in domain size and location occur. An issue, which is attempted to be solved by an automated domain generation procedure. Unexpectedly, the diversities in domain did not strongly affect the model performance in the presented examples. Investigations with tools to analyse sensitivity suggest that ANNs keep focus on the main area of interest and are

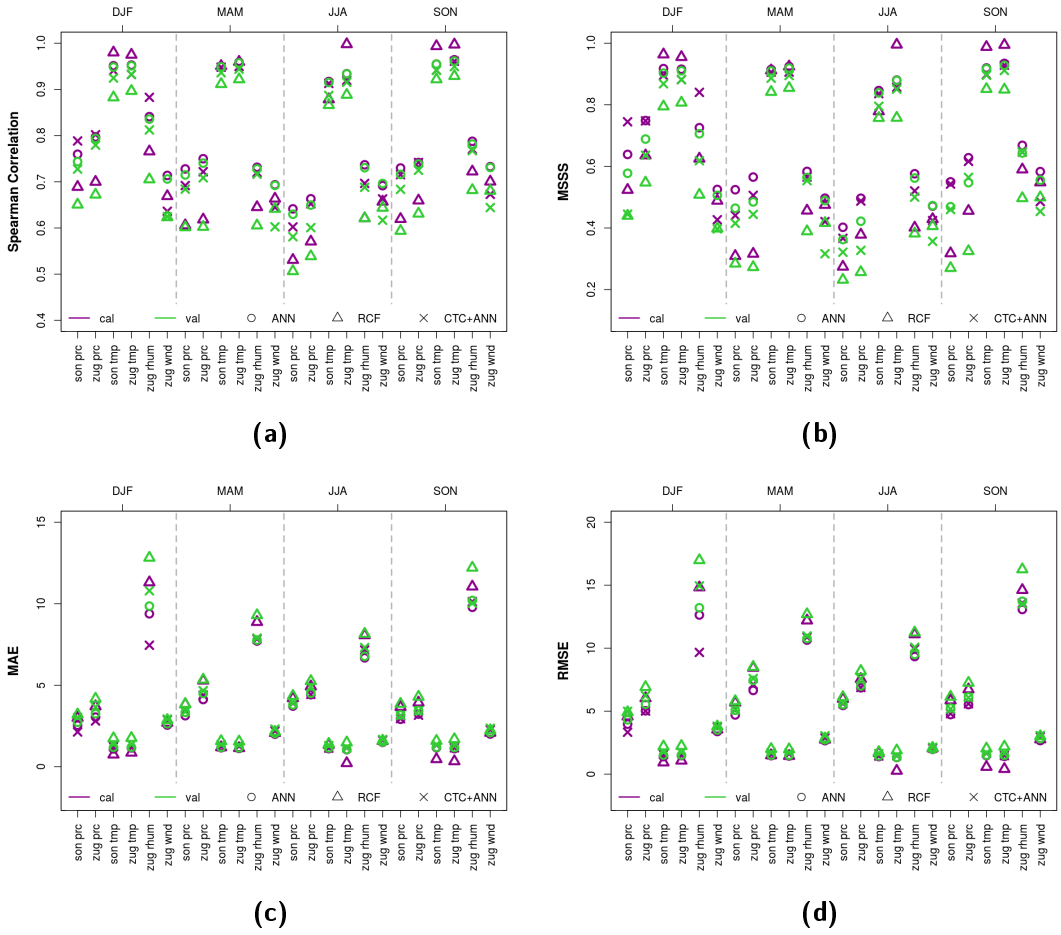


Figure 7.22: Summarised evaluation metrics for every final SD model configuration considering calibration and validation using the 20th century reanalysis dataset and local observations. Units: precipitation in mm, temperature in degree Celsius, relative humidity in percent and wind speed in m/s. Time period: 1970 to 2000, 20 % of the time steps is reserved for validation. Marks represent the mean of a 15-fold Monte-Carlo cross-validation procedure.

efficient in eliminating input time series with no added value to the model. Based on this results, a framework to find a good predictor setup is developed using primarily sensitivity studies. In addition, it is taken care to receive a physical consistent predictor setup, which carries the climate signal. From the three methodologies zeroing, adding noise and partial derivatives, the latter is primarily applied, as all identify similar sensitive regions, but PaD exhibits by far the fastest run time. During the predictor screening process, insensitive predictors and levels are removed successively from a starting point including all appropriate predictors. Defined as the best predictor combination, is the setup preceding a rapid drop in model performance. Subsequently, the domain of each predictor is cut to the most sensitive regions.

In order to find good SD model configurations for the final predictor combination, large sets of parameters are evaluated for the three downscaling methods: ANNs, RCFs and NNCs. Considering ANNs, a fast and in high performance consistent training algorithm is found to be the Rprop algorithm. Although not the fastest transfer function, logsig is preferred for its slightly better performance. If a short training time is crucial, one would be better off with transfer functions symmetrically to the coordinate origin. Training is stopped by Early Stopping, a number of 300 checks was found to be in general sufficient for Rprop. In the presented examples, ANNs with higher numbers of hidden layers do not lead to an increase in model performance, even if trained by greedy layer-wise pre-training, thus, one hidden layer is selected in all tasks. Finally, the right amount of hidden neurons needs to be identified. Various numbers of hidden neurons are successively evaluated, while the best performing number is chosen for the final model configuration. A large variance can be witnessed within the best number of hidden neurons. One of the most influential parameters on RCFs model skill is the number of classes. 27 types showed the best results. RCFs are optimised conditioned by including the weighted predictand into the classification procedure. With the exception of temperature, RCFs prefer small predictand weights, considerably below one. RCFs are optimised by the DKM algorithm, which is characterised by a considerable lower run time and slightly better results in the validation period in the investigated examples compared to the SANDRA algorithm. The model setup of NNCs is greatly influenced by the findings of the preceding ANN and RCF studies. Input datasets are divided by a conditioned DKM-based classification approach into a maximum of three types. The successive ANNs for each dataset, separated by classes, are trained by the Rprop algorithm. NNCs come in hand with a large effort, as a distinct larger amount of parameters needs to be evaluated.

SD models are finally trained in a 15-fold Monte-Carlo cross-validation procedure using the identified predictor setups and model/training configurations. Calibrated models are evaluated by the Spearman correlation coefficient, the MSSS, the MAE

and the RMSE. Comparing the validation period of the different SD approaches, the exclusively ANN based models lead to top results. Highest model performance is typically achieved for temperature followed in order by relative humidity, precipitation and wind speed. The resulting SD models are applied on ESM datasets in the climate change studies in the following sections.

8 Precipitation

8.1 An Overview over Origin and Distribution of Precipitation in the Alpine Region

Figure 8.1 illustrates the annual precipitation sums of the GAR for the climatic period 1961-1990. The highest precipitation amounts can be found in the northern and southern flanks of the Alpine mountain range and in a transitional zone near the Gotthard Pass in Switzerland. The south-eastern precipitation pattern is extended to the south following the Julian and Dinaric Alps. Moisture is advected towards the Alps in the northern part from the Atlantic Ocean or in the southern part from the Mediterranean Sea. Reaching the mountain foothills, air masses underlie orographic lift and cool down adiabatically, when gaining altitude. Therefore, on the Luv side, relative humidity is raised, which can create clouds and, under the right conditions, precipitation resulting in high precipitation amounts at the flanks of the mountain range. The air masses passing the central Alps are characterised by lower moisture content leading to relatively lower precipitation sums in the inner regions, especially in valleys (Barry 2008). This process chain often goes hand in hand with the Foehn effect.

Typically, Foehn occurs in the Alps as a north or south wind. Related processes are schematically illustrated in figure 8.2. Moist air advect at the Luv side of the mountain barrier. Orographic lift cools the air masses dry adiabatically (A) until the condensation point is reached. In the further progress, cooling is performed following the moist adiabatic curve, reducing the temperature loss due to the release of latent heat (B). In addition, clouds form and precipitation may be the result. After passing the peak on the Lee side, air temperature increases again dry adiabatically (C), resulting in a warm dry Foehn wind on the low level mountain flank with higher temperature then at the outset. Near the peak on the Lee side dissolution of the cloud can occur, leading to a so called Foehn wall. The elevated air stream on the Lee side oscillates forming Lee or gravity waves often with Lenticular clouds. Furthermore, wind rotors can develop on the Lee side, a turbulent vortex, with its axis of rotation parallel to the mountain range. The topside of the rotation follows the high level air stream direction.

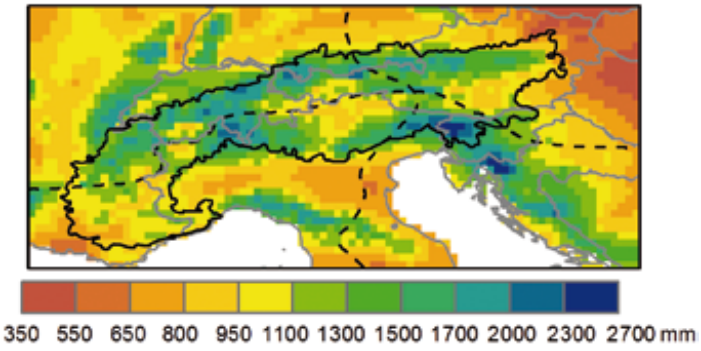


Figure 8.1: Annual precipitation sum of the European Alps for the period 1961-1990 based on the data from Auer et al. (2008). Regional statistics: GAR = 1097 mm, Alps (thick black line) = 1317 mm, summits higher than 1500 m = 1292 mm. Source: Agency (2009).

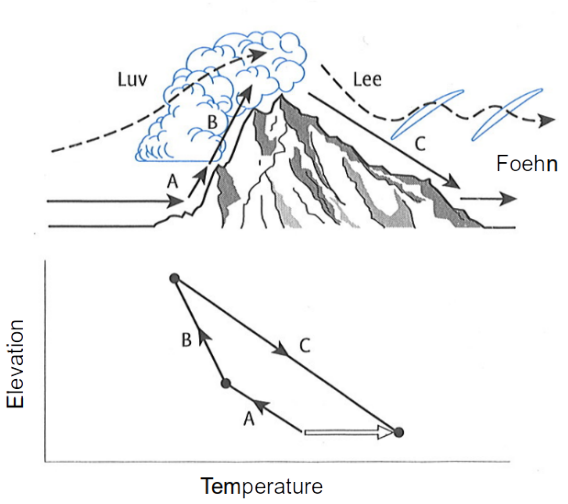


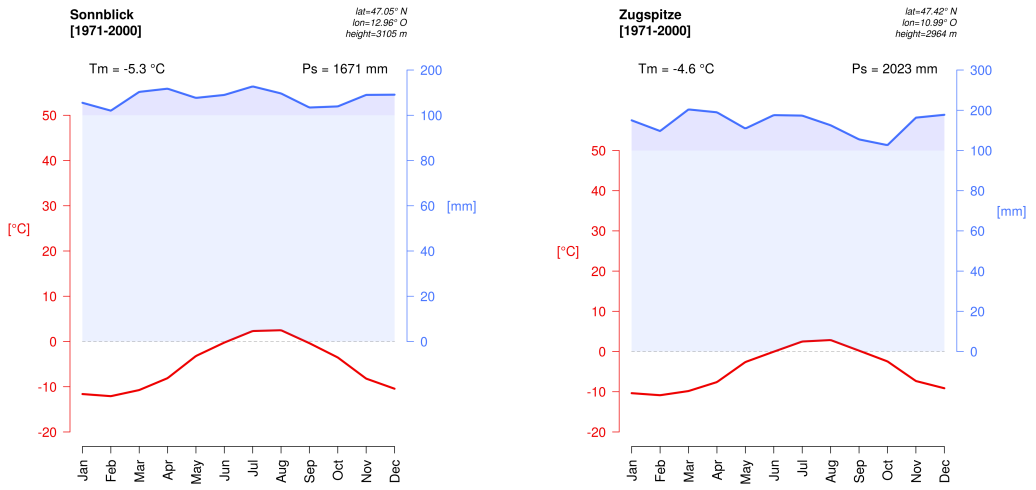
Figure 8.2: Foehn effect. Figure based on Schönwiese (2013a).

The annual precipitation amounts often origin with considerable parts from moderate- and heavy precipitation events. Cebon (1998) pointed out that in some areas, $4\,\%$ of the wettest days contribute $40\,\%$ of the overall precipitation totals. 1960, for example, one of the wettest years measured at the Sonnblick Observatory shows a frequency minimum of precipitation events. Most of these high-intensity precipitation events are triggered by local atmospheric conditions like thunderstorms, whereas precipitation time series in the Alps, particularly of high mountain stations, show weak correlations. The dependency of the yearly precipitation sum on rare heavy events further leads to a pronounced inter-annual variability.

Barry (2008) gives an overview of the annual precipitation cycle in various regions. In the northern and central parts of the Alps, a precipitation maximum can be found in the summer months caused by a higher probability of occurrence of thunderstorms. In the northern parts, a second maximum can be identified in the winter months induced by a higher moisture advection due to an increased cyclone occurrence in this season. The central parts show a winter or spring minimum. The southern slopes are characterised by a precipitation maximum in autumn and minimum in winter. No clear peak can be found in the western parts. To the east, the Alps precipitation totals decrease with continentality.

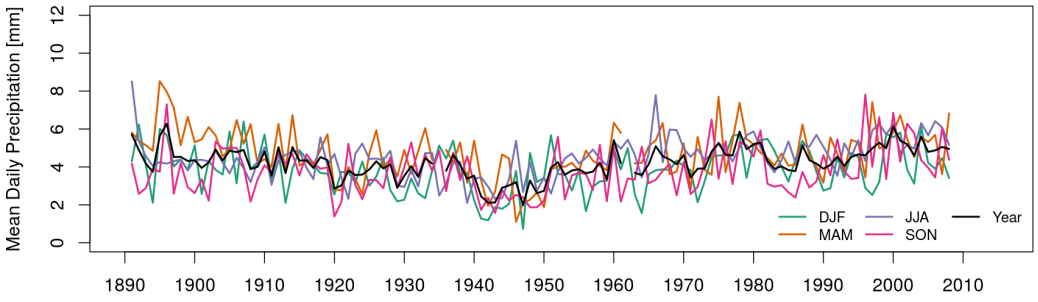
8.2 Analysis of the Observational Datasets

Figure 8.3 shows the characteristic climate observed at Sonnblick and Zugspitze of the time interval 1971-2000, which will be used as reference period in the following sections. Considering the locations of both stations in figure 8.1, the Zugspitze is positioned in the heavy precipitation band on the northern flanks of the Alps, while Sonnblick is found within the latitudinal centre of the Alpine range. Although generally lower precipitation occurs in this region, Sonnblick is located within a pattern of slightly higher precipitation ranging from Salzburg (Austria) to Triest (Italy). Nevertheless, the annual precipitation sum at Sonnblick Observatory ($1671\,\mathrm{mm}$) is $17\,\%$ lower than at Zugspitze ($2023\,\mathrm{mm}$). Both station time series do not underlie a pronounced annual precipitation cycle. The monthly rainfall distribution is similar: A noticeable reduction in precipitation can be observed in early summer (May) as well as autumn (September/October). A further minimum in February might be caused by the lower number of days contributing precipitation to the monthly sum. In the reference period, the highest monthly average precipitation sum occurs at Zugspitze in March ($202\,\mathrm{mm}$) and at Sonnblick in July ($163\,\mathrm{mm}$). The minimum precipitation sum at Zugspitze is measured in October ($113\,\mathrm{mm}$) and at Sonnblick in February ($110\,\mathrm{mm}$) although, as mentioned before, this value can be biased by the number of days (next in line are September ($117\,\mathrm{mm}$) and October ($120\,\mathrm{mm}$)). Following Walter

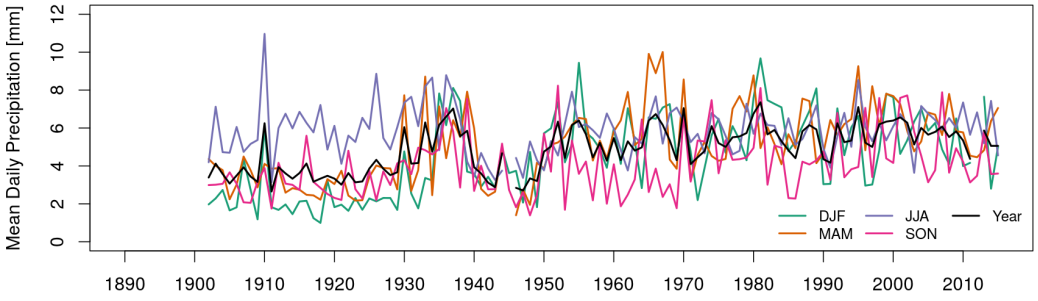


(a) Sonnblick **(b)** Zugspitze

Figure 8.3: Climograph following Walter and Lieth (1967). Displayed are the monthly mean temperatures alongside the monthly average precipitation sums of the local observation in the climatological reference period 1971-2000. Tm: Average annual temperature mean. Ps: Average annual precipitation sum.



(a) Sonnblick



(b) Zugspitze

Figure 8.4: Annual and seasonal time series of the averaged local precipitation measurements computed from the total observational record.

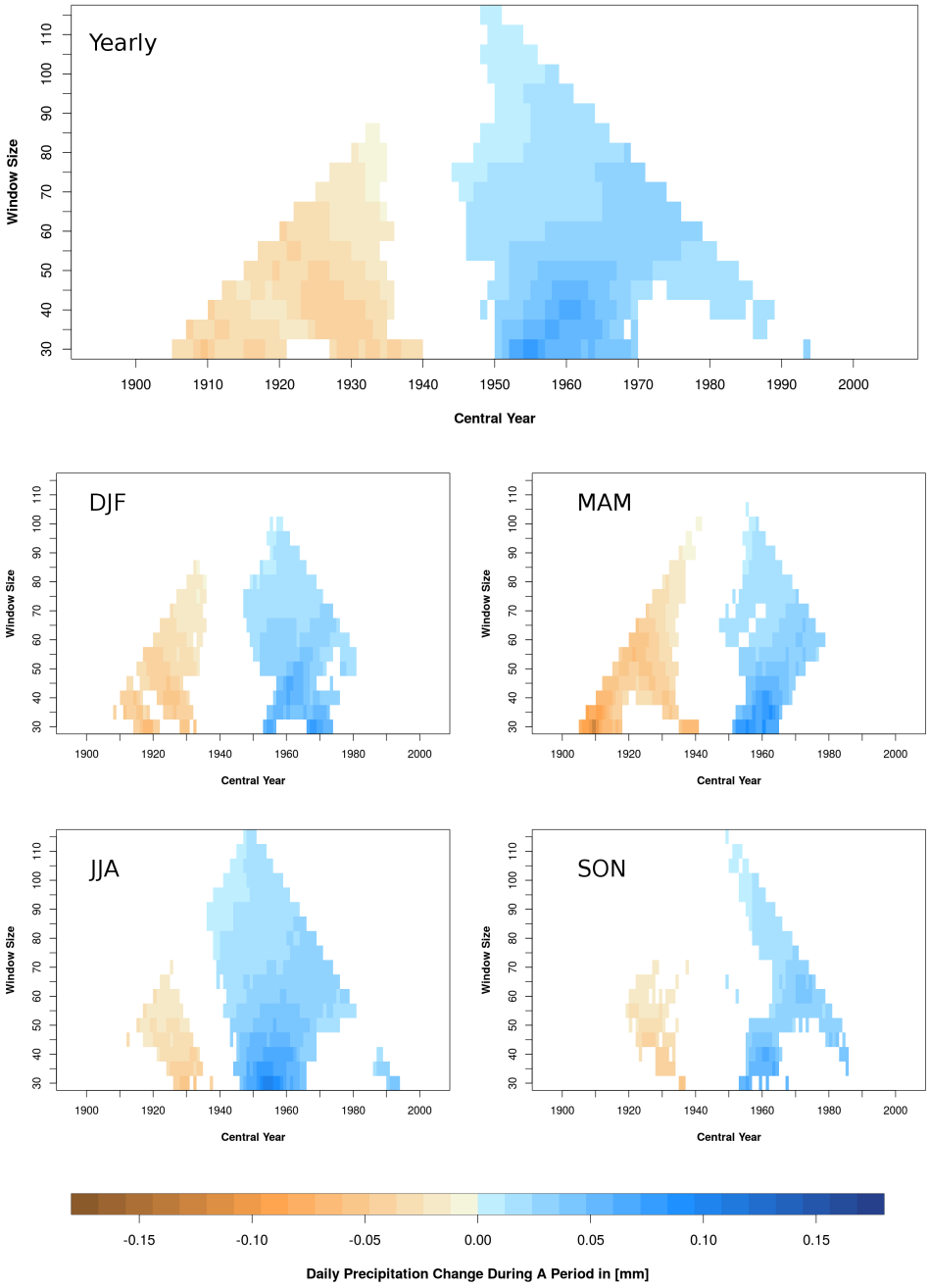


Figure 8.5: Trends in the observational time series of the seasonally and annually aggregated Sonnblick precipitation record. The x-axis represents the central year of the evaluated time period, the y-axis lists corresponding time window sizes in years. Trends are displayed if tested significantly by trend-noise ratio (level of significance 5 %).

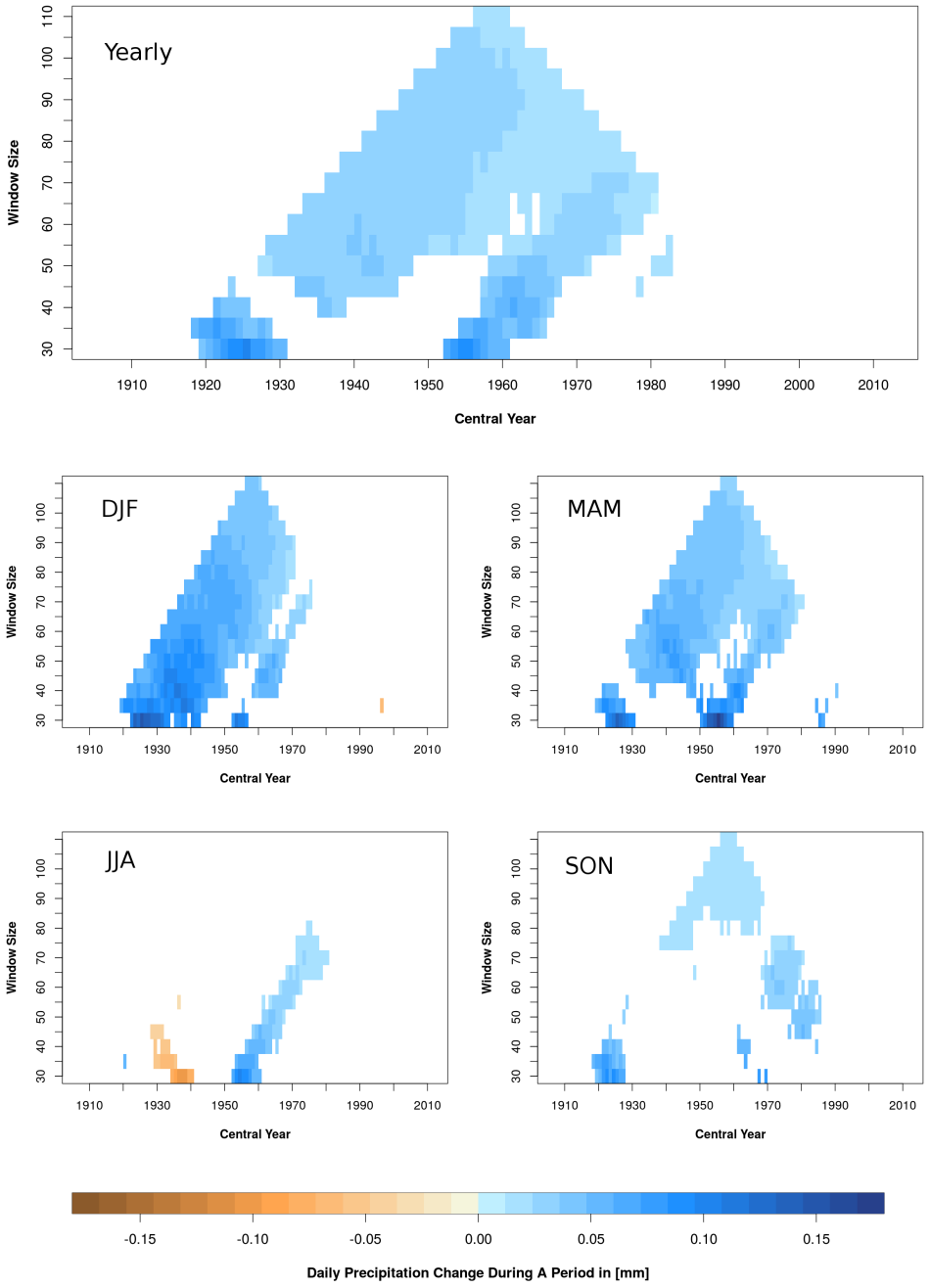


Figure 8.6: Trends in the observational time series of the seasonally and annually aggregated Zugspitze precipitation record. For a detailed description see figure 8.5.

and Lieth (1967), both stations can be classified as humid. The inter-annual variability at both stations is quite large. Zugspitze is characterised by a standard deviation of $315\,\mathrm{mm}$ and Sonnblick of $238\,\mathrm{mm}$. In the period 1971-2000, magnitudes of annual precipitation sums at Zugspitze range from a minimum in 1971 with $1494.4\,\mathrm{mm}$ to the maximum of $2724.4\,\mathrm{mm}$ in 1981, at Sonnblick from a minimum in 1971 with $1233.8\,\mathrm{mm}$ to a maximum of $2250.3\,\mathrm{mm}$ in 2000.

Unless explicitly described otherwise, investigations in the following sections and chapters concerning annual averages or sums are based on the meteorological year, starting from the first of December of the previous calendric year. The observed precipitation time series in figure 8.4 indicate no large difference between seasonal or annual averaged daily precipitation sums, except for the early 20th century at Zugspitze, where a more obvious difference between seasons occurs. Here, the wettest season is summer, while winters are relatively dry. Starting from the 1960s, winter precipitation heavily increases to the level of summer. Beside circulation induced reasons, a more sophisticated measurement system with an improved catchment of solid precipitation may lead to higher and more representative results in winter (see section 4.1 for a discussion on precipitation measurement difficulties). Both datasets show no conspicuous features in the model building period 1970-2000. Trends of the time series are analysed in figures 8.5 and 8.6. Trends are computed for different time periods successively increased by five years from a starting value of 30 years and finishes with the largest possible window size, approximately covering the whole available dataset (see y-aches). The x-axis represent the central year splitting the considered time window in half. Windows start at the beginning of the dataset and are shifted by one year until the latest measurement is reached. The trend value of each window is computed by linear regression and printed if successfully checked for significance by the trend-noise ratio (see 5.4.4). A level of significance of $5\,\%$ is selected as threshold. At both stations smaller window sizes typically show larger trend values than bigger ones, indicating that changes in the precipitation behaviour occur rather rapid. For the Sonnblick figures 8.4a and 8.5 feature decreasing precipitation values in the first half of the 20th century, in contrast to the second half starting at 1950, where positive trends are dominant. This tendency can be identified throughout the seasons, although in autumn trends are less frequently significant. Precipitation at Zugspitze primarily increases significantly in the 20th century, except for summer, where on the short scales from centre years 1930 to 1940 precipitation decreases. Especially the increasing precipitation values in winter and spring at 1930 and 1950 stand out in figures 8.4b and 8.6. The evaluation of the Zugspitze record aligns quite well with the literature. In the 20th century, authors usually find increases in winter precipitation in the northern parts of the Alps and decreasing tendencies south of the main ridge (Brunetti et al. 2006; Schmidli et al. 2002; Quadrelli

Table 8.1: Spearman correlation between Sonnblick and Zugspitze observed daily precipitation of the maximal congruent (first line) and reference time period (second line). Indices are computed for the overall period and seasons, time steps with missing values are omitted. All coefficients are significant on a significance level of 5 percent.

	Total Period	**DJF**	**MAM**	**JJA**	**SON**
01.01.1901 - 31.12.2008	0.65	0.65	0.65	0.64	0.66
01.01.1971 - 31.12.2000	0.72	0.76	0.71	0.66	0.74

et al. 2001). The seasonal trends determine the annual precipitation leading to wetter conditions in the north and dryer conditions in the south of the Alps (Brunetti et al. 2006).

Oftentimes, precipitation of summit stations in the Alps show no strong match in temporal variability (Auer et al. 2008). Table 8.1 lists Spearman correlation coefficients between Zugspitze and Sonnblick. Reaching values of up to 0.76 in DJF during the reference period, the relationship between both stations is substantial. In comparison to the overall period, the reference period is characterised by higher magnitudes in all time periods. Lowest correlation is detected in summer, where local weather events, like thunderstorms, increase in frequency. In periods where precipitation origins from large-scale circulation, like for example in winter, the relationship intensifies. Further investigations of cross-correlations of the example periods in table 8.1 did not show increased relationships.

8.3 Statistical Model Results with Focus on the 20th Century

In figure 8.7, time period dependent changes in model quality throughout the measurement period are visualised. Monitored is the skill of the ANN models reached during calibration. Changes of skill can be caused by inhomogeneities of the dataset, non-stationarities of the statistical relationship (Hertig and Jacobeit 2013) and/or data quality issues. Non-stationarities can have various origins, for example from short training periods, which do not include all possible atmospheric conditions. Modes of long term variabilities, like the NAO, may not be sufficient enough represented in calibration period. Hertig and Jacobeit (2008) are pointing out the importance of long calibration periods to avoid the impact of non-stationarities. Another non-stationarity source can be an inadequate predictor combination and/or missing important deterministic factors (Razavi et al. 2015). Model quality increases towards the end of the 20th century at both stations, beginning approximately from 1950, except for summer, where the peak is shifted towards a more recent time. In some cases, the model skill increases drastically to the end of the 20th century, for exam-

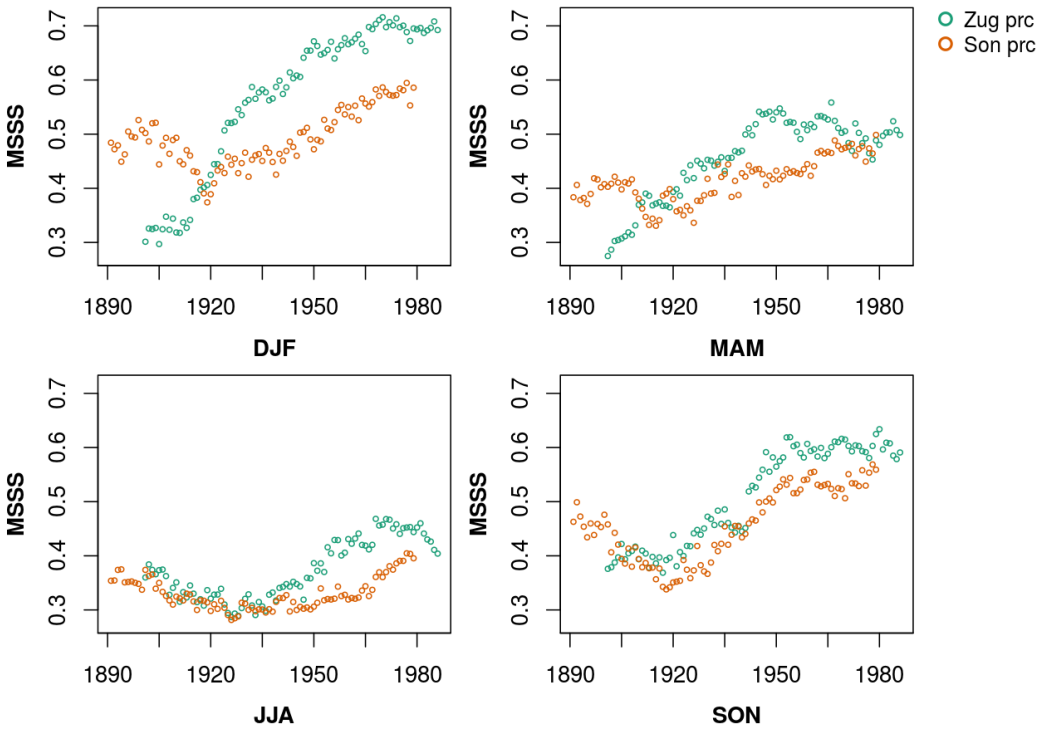


Figure 8.7: Seasonally dependent analysis of the temporal altering model skill of the final ANN configuration evaluated by the MSSS. Models are calibrated using observed precipitation (prc) at Zugspitze (Zug) and Sonnblick (Son) and the 20th century reanalysis dataset. Training is repeated from 15 different ANN initialisations, the resulting skills scores are averaged. Models are calibrated on a 30-year period, starting at each mark of the graphic.

ple in Zugspitze winter by roughly 0.4. The low skill at the beginning of the 20th century coincides with the suspicious low precipitation in the predictand time series (see figure 8.4a). Around 1920, a noticeable temporal decrease in model skill occurs at Sonnblick station, which can be witnessed slightly alleviated at Zugspitze in summer and autumn as well. In the calibration period, starting from 1970, all seasons are characterised with among the highest model skill, which makes this time period seem to be the most appropriate one for calibrating the final SD models.

A possible drawback of selecting the calibration period by this approach can be linked to non-stationarities. Merkenschlager (2017) points out that SD models tend to prefer certain atmospheric modes. For example, if higher model performance is reached while the NAO is positive one could falsely exclude the negative complementary, resulting in an overall limited explanatory degree of the SD model. In the selected calibration period 1970 to 2000, the seasonal winter mean of the NAO index indicate a mainly positive NAO phase, while the AMO index suggests a primarily negative AMO phase (ESRL 2018). However, the pronounced increase in model quality does not show an AMO and NAO index oscillation dependent curve progression that other influences, e.g. the quality of reanalysis and observational data, seem to be a more likely reason here.

In the following the transfer of the SD models from reanalysis to ESMs data input is evaluated. On a short-time scale, ESMs do not feature a congruent timeline with reality (see section 4.2.2). ESMs and observed time series can therefore only be compared by long-term statistical values or by focusing on distributions. An analogue segment of 30 years is usually considered to be sufficient to represent the climate variability. QQ-plots of figures 8.8 and 8.9 compare the distribution of observations and reanalysis and ESM data driven models by the quantiles in the reference period 1971-2000. Quantiles of the best models are located close to the dotted grey equality line. In the overall picture, the SD models tend to overestimate frequency and amount of low precipitation days and underestimate heavy precipitation events and work best in the centre of the value distribution. This decrease of accuracy at the tails of the distribution is typical for daily precipitation models (Cavazos and B. C. Hewitson 2005). To reduce the so called “drizzle effect” at the lower parts of the distribution and improve the frequency of wet days, very small simulated precipitation values below a certain threshold, here the lowest value of the observed time line $0.1\,\mathrm{mm}$, can be set to 0 (Maraun et al. 2010). This correction procedure is applied to all SD model time series, including future scenarios. Although the effect is still present in the figures, larger correction values tend to decrease the overall model skill. Considering the accuracy of the models at the upper tails of the value distribution, the ANNs and oftentimes NNCs show better result as RCFs. A reason can be the fixed maximum reference value, which needs to cover a larger value range

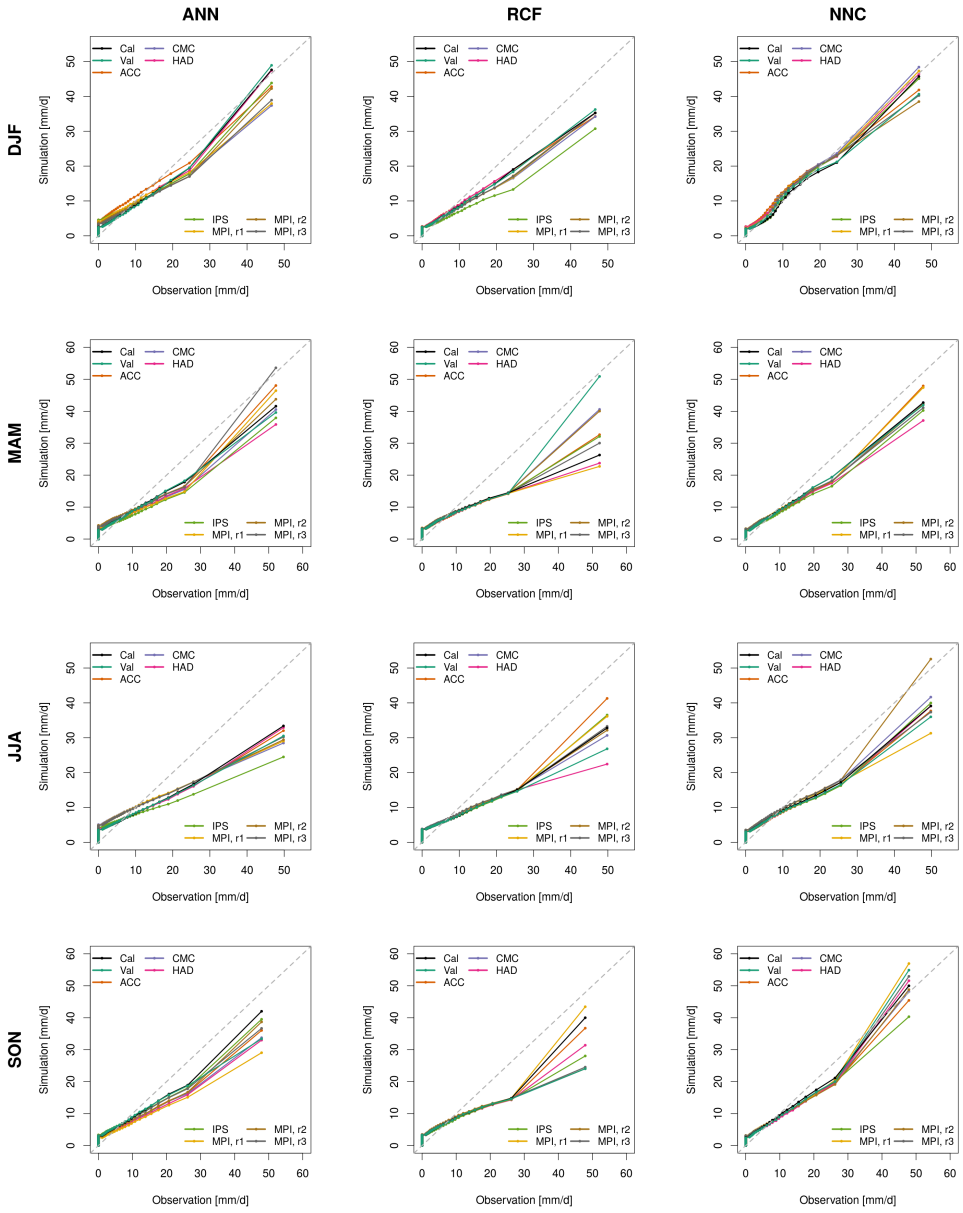


Figure 8.8: QQ-plots for Sonnblick precipitation considering the seasonal models in the reference period 1971-2000. 50 quantiles of the observed dataset are compared with their simulated counterparts from reanalysis and historical ESM data-based downscaling products derived by the final SD models. Reanalysis series are separated by calibration (Cal) and validation (Val) periods. The fifteen Cross-validation runs are treated as one sample.

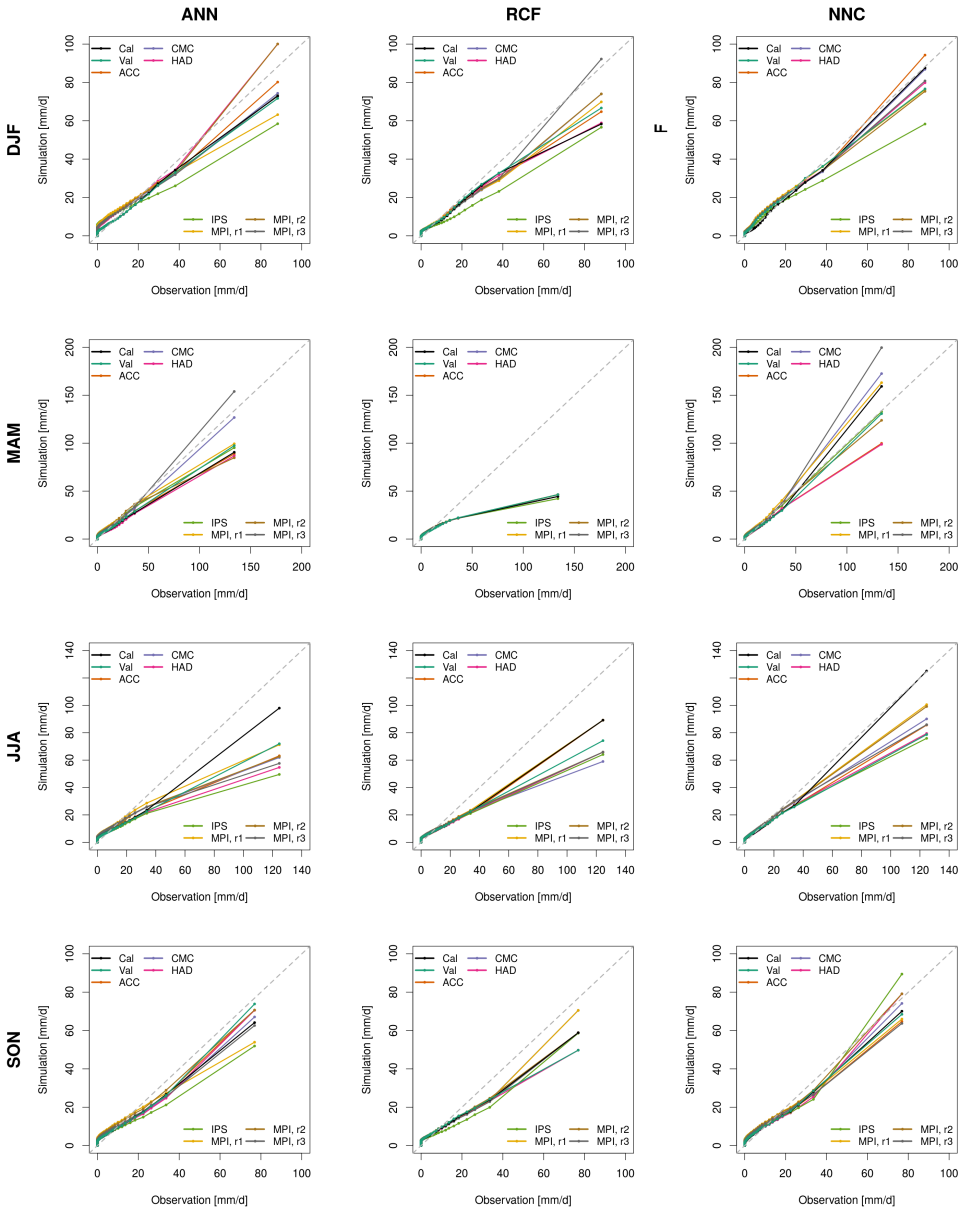
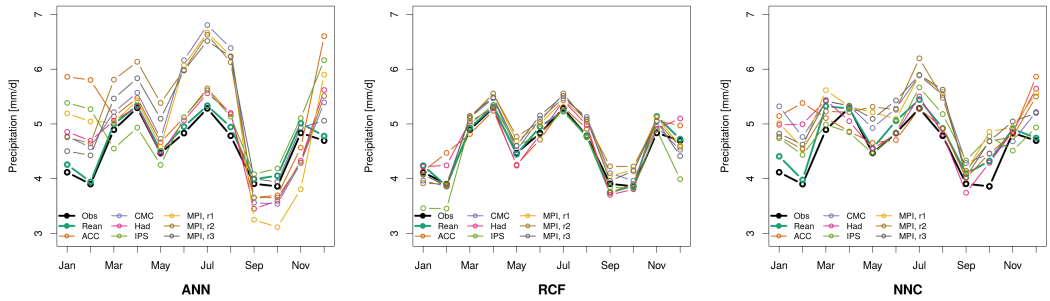


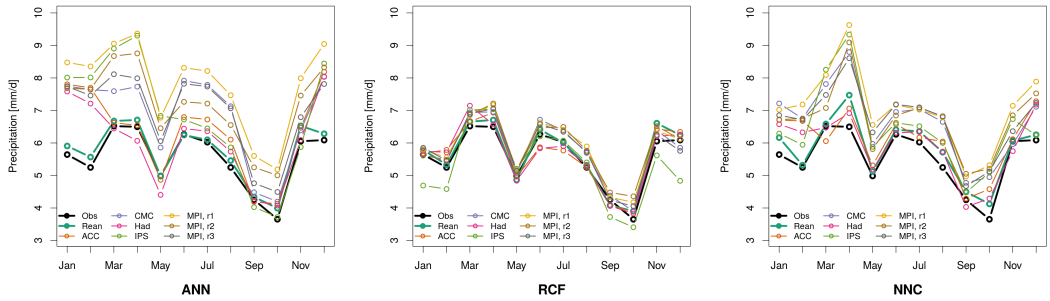
Figure 8.9: QQ-Plots for Zugspitze precipitation: See figure 8.8 for detailed description.

due to increased variability in the upper parts of the distribution. This effect is particularly apparent at Zugspitze in spring: With $133.9\,\mathrm{mm}$, an outstandingly heavy precipitation event occurred in spring on the 21th of May in 1999 (second highest value in the time series: $81.0\,\mathrm{mm}$ on the 17th of March in 2000). The associated QQ-plot of RCF implies a failure of the model only producing reliable value below $50\,\mathrm{mm}$ caused by a maximum reference value of approximately $45\,\mathrm{mm}$. In general, quantiles of the SD model simulations differ little, neither between reanalysis and ESMs, nor in between ESMs with the only exception of the last upper quantile, where partly quite large spreads occur. SD model quantiles visually match better at Sonnblick than Zugspitze. Comparing ESM results, there is no wetness specific order, except for IPSL, which tends to be dryer than other models in several cases. The within variation of MPI-ESM realisations (r1,r2,r3) show similar magnitudes than between the other ESMs, indicating that although produced by the same model, no systematic dependency is noticeable. NNCs sometimes exhibit a slight bulge in the centre of the QQ-Plot, most likely occurring from overlapping value ranges of the successor ANNs. This effect is also common in mixture models and should be reduced as far as possible (Marin et al. 2005). Beside winter, the effect could be efficiently reduced, by carefully optimising the CTC and ANN combination.

A correct representation of monthly means is particularly important for successive modelling task, like hydrological models, but often a shift in the annual cycle can be noticed (e.g. Smiatek et al. (2013) and Dai (2006)). In figure 8.10, the observed and simulated monthly averaged (1971-2000) daily precipitation sums are compared. Results from reanalysis fit the observational dataset quite well, with largest differences in a few cases of the NNC approach. In general, the course of the year is well represented by all models. The ANN models tend to overestimate precipitation at Sonnblick especially in winter and summer, while in summer mainly the MPI-ESM realisations and CMCC are affected. Zugspitze precipitation averages are mainly overestimated as well. Largest differences between observations and ESMs is found in winter. Except for winter, once again good agreement exits between observations and ACCESS and HadGEM. Best accordance between modelled and observed monthly averages show the RCFs. Again, the largest discrepancies are identified in winter, especially noticeable is the IPSL model, which underestimates the observational magnitudes. Similar to ANNs, NNCs overestimate the measurements. Larger differences between observations and realisation-based simulations are carried into ESM driven time series, leading to a dryer April and wetter October at Sonnblick and a wetter April at Zugspitze.

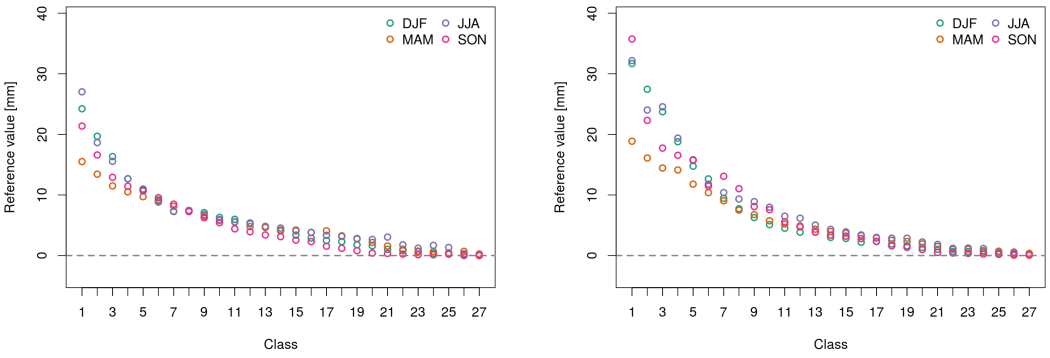


(a) Sonnblick



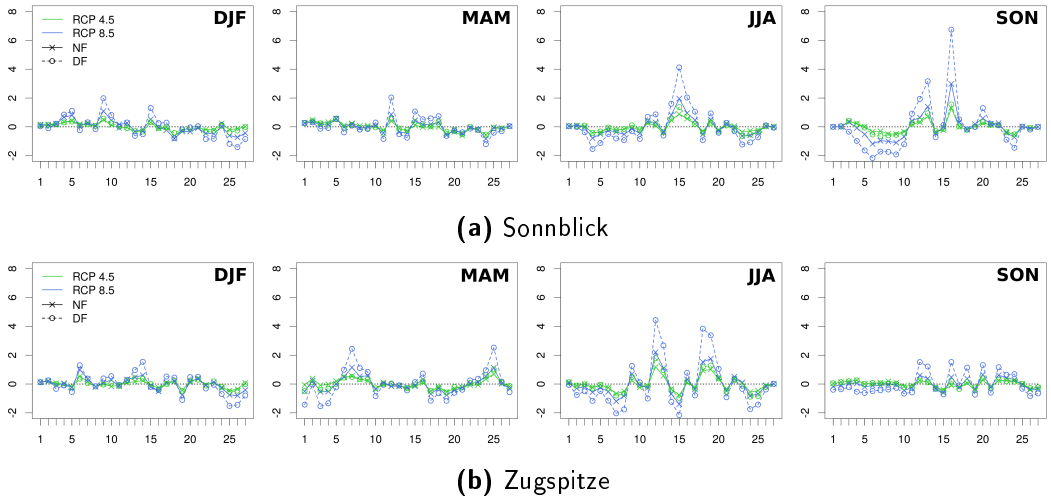
(b) Zugspitze

Figure 8.10: Observed (Obs) and simulated (complete reanalysis (Rean) and climate model driven results) monthly averaged daily precipitation sums considering the reference period 1971-2000 for each SD model type.



(a) Sonnblick **(b)** Zugspitze

Figure 8.11: Circulation type specific mean daily precipitation sum reference values in [mm d^{-1}] for each of the 27 classes of the seasonal RCFs at Zugspitze and Sonnblick derived from the calibration period 1971-2000.



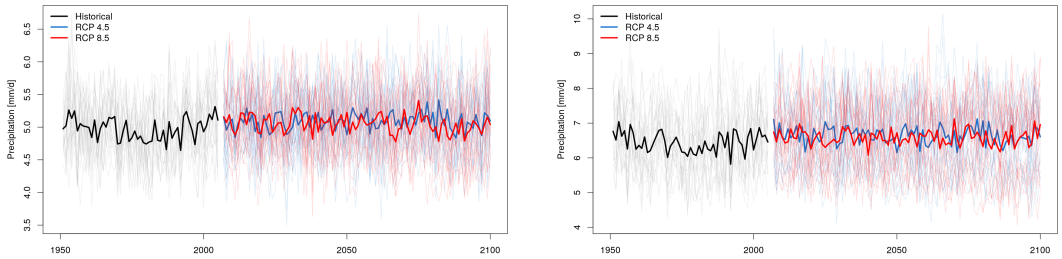
(a) Sonnblick

(b) Zugspitze

Figure 8.12: Average future changes of type frequency in [%] comparing the historical reference (1971-2000) of all ESMs with both scenarios (RCP 4.5 and RCP 8.5) in the near (NF) and distant future (DF) of the seasonal RCF precipitation models at Sonnblick and Zugspitze.

8.4 Climate Change Signals: 20th vs. 21st Century

At first the characteristics and behaviour of RCFs in a climate change scenario shall be discussed. Figure 8.11 displays the average reference values of daily precipitation sums considering the 15 cross-validation results for each class and season. Reference values and respective classes are sorted by their magnitudes. However, in the mean of the 15 cross-validation reference values slight order discrepancies can occur. An exponentially decaying value height can be observed from lower to upper class numbers. Largest differences between seasons are observed in the circulation types leading to high precipitation. Reference values of heavy precipitation reach higher values at Zugspitze. Considering heavy precipitation types, with the exception of class 1 at Zugspitze, reference values are higher for summer and winter and lowest for spring. Starting approximately at class nine at Sonnblick and class 13 at Zugspitze, seasonal differences become fairly small. In figure 8.12, the expected future change in frequencies is visualised. The development of RCP 4.5 and RCP 8.5 frequency changes is similar in the tendency, but RCP 8.5 amplitudes in the equal future period are typically higher. Similar, a slight difference in NF is gradually stronger in the DF. It is noticeable that mostly a few types, orientated towards the centre of the reference value range, show a strong increase in frequency, whereas plenty of them are characterised by slight decreases. The general decrease of many types in favour of the increase of a couple ones, explains the detected reduction in future precipitation variability (see table 8.3). The effect is especially pronounced at Zugspitze and Sonnblick in the DF of RCP 8.5 in summer and autumn. Important



(a) Sonnblick **(b)** Zugspitze

Figure 8.13: Annually averaged daily precipitation sums resulting from the ESM downscaling product of the total available time period 1950 to 2100 (historical and scenario dataset). Thick lines represent the overall average of ESM results. Thin lines in the background display the mean of the cross-validation procedure considering a single SD model and ESM combination.

to evaluate in a climate change study is the frequency change of the tails of the reference value range. Predictor classes in RCFs show quite similar behaviour like the corresponding reference values. For example, a class with low hus, hur and wind velocity typically bears low predictand reference values and vice versa. Considering precipitation, in particular the frequency of the heavy event class 1 is of interest. A strong increase in type 1 can be an indication for higher projected values of hus or wind in the future. Therefore, the reference value range might be insufficient and one or more new classes of higher magnitudes are necessary to get reliable results. In this study, no large increase is detected in any case in the top three classes, making the necessity of higher reference values from this perspective unlikely. Still, it gives only an indication. SD model problems originating from this matter cannot be ruled out completely. However, the strong rise in frequency of single classes in the centre of the value range can indicate the necessity of a finer subclassification near the peak. In general, RCFs are not suited for evaluating changes in the magnitudes of extreme values, as the latter are defined fixed during the calibration process and cannot adapt to changing conditions.

In the following, the climate change analysis results of ANNs, RCFs and NNCs are discussed. The overall development of annual precipitation means resulting from all SD models driven by ESM data throughout the historical period and scenarios are displayed in figure 8.13, while trends are further listed numerically in table 8.2. Considering all ESMs and SD models, a significant slight positive trend is projected for both stations, Zugspitze and Sonnblick and the scenarios, RCP 4.5 and RCP 8.5, as well as the time periods near future (NF, 2021-2050) and distant future (DF, 2071-2100). This tendency origins primarily from ANNs and NNCs. RCFs feature predominantly negative changes, but rarely significant ones. The contrary development of RCFs might result from the general difficulties of this methodology to deal with

Table 8.2: Historical (Hist: 1971 to 2000) annually averaged precipitation [mm] at Zugspitze and Sonnblick as well as projected relative changes to the historical period [%] considering different ESMs, SD models and scenarios. Changes are computed for the near future (NF: 2021 to 2050) and distant future (DF: 2071 to 2100). Positive numbers are marked in green, negative ones in red, while insignificant values (U-test, significance level 5 %) are labelled in grey. (Observational average of Sonnblick: 4.2 mm d^{-1}; Zugspitze: 5.6 mm d^{-1}.)

		Sonnblick					Zugspitze				
			RCP 4.5		RCP 8.5			RCP 4.5		RCP 8.5	
		Hist	NF	DF	NF	DF	Hist	NF	DF	NF	DF
ACC, r1	ANN	5.1	5.0	10.7	8.3	8.4	6.3	5.5	13.2	9.0	10.8
	RCF	4.6	-0.8	2.9	2.3	-2.2	5.6	-0.4	3.2	2.7	-3.4
	NNC	4.9	2.0	7.8	5.1	2.1	6.1	6.5	15.3	10.4	13.3
CMC, r1	ANN	5.1	1.0	2.6	3.2	3.2	6.9	2.2	1.8	0.8	2.6
	RCF	4.7	-1.8	-1.3	0.2	-4.0	5.8	-1.3	-4.1	-2.6	-6.8
	NNC	5.1	-0.9	-1.0	0.4	-2.9	6.7	3.6	3.3	2.0	5.3
HAD, r1	ANN	4.8	6.5	8.0	6.7	10.7	6.0	10.1	6.7	6.2	12.9
	RCF	4.6	3.2	1.6	3.1	0.6	5.6	3.0	-2.1	0.1	-3.8
	NNC	4.9	6.9	6.5	5.0	6.9	5.9	11.4	8.4	7.4	16.3
IPS, r1	ANN	5.0	7.1	9.6	7.9	9.3	6.8	10.1	5.4	5.0	4.8
	RCF	4.4	4.5	3.1	4.2	-1.3	5.3	4.6	-1.7	1.1	-7.8
	NNC	4.8	6.3	6.9	6.2	3.0	6.4	11.7	6.4	6.2	6.6
MPI, r1	ANN	5.0	4.2	1.6	1.1	2.9	7.8	2.7	1.5	1.4	-1.3
	RCF	4.7	3.3	-1.4	-1.1	-1.8	5.9	0.3	-2.9	-1.3	-7.6
	NNC	5.1	5.0	0.6	1.0	0.1	7.0	2.6	1.6	1.1	-0.1
MPI, r2	ANN	5.2	3.7	4.3	4.7	6.4	7.2	1.8	3.4	3.7	6.1
	RCF	4.7	1.7	1.3	2.3	0.8	5.9	-0.9	-0.6	1.3	-2.7
	NNC	5.1	2.1	3.4	4.9	4.0	6.7	1.4	4.3	3.5	8.9
MPI, r3	ANN	5.2	5.7	6.1	4.2	2.9	7.0	8.3	8.1	5.2	6.9
	RCF	4.7	4.0	2.6	1.2	-3.3	5.8	5.0	2.5	0.3	-3.4
	NNC	5.1	5.6	4.0	2.1	-0.4	6.7	8.1	7.4	4.7	8.3
All ESMs	ANN	5.1	4.7	6.1	5.1	6.2	6.9	5.6	5.5	4.3	5.8
	RCF	4.7	2.0	1.2	1.7	-1.6	5.7	1.4	-0.8	0.2	-5.0
	NNC	5.0	3.8	4.0	3.5	1.8	6.5	6.3	6.5	4.9	8.1
All ESMs	All SDs	4.9	3.6	3.8	3.5	2.2	6.4	4.6	4.0	3.3	3.4

shifts in the input distribution.

Of all ESMs, CMCC (r1) and MPI-ESM (r1) are characterised by a disproportionately high number of insignificant values. Comparing scenarios in the overall mean, a higher increase in daily precipitation is expected in the RCP 4.5 scenario, although the result is highly influenced by RCFs. The effect might be caused by the lower relative humidity input values in the RCP 8.5 scenario (see figure 6.2).

Figures 8.14 and 8.15 investigate the change signals by the distribution of seasonal means, whereby ESMs are considered combined. At both stations, the SD models show the strongest and steady increase in precipitation during winter. This result is supported by for example Smiatek et al. (2009) or Schmidli et al. (2007) who found strong increases in winter precipitation in the northern parts of the Alps. The expected wetter winter months could lead to a change in the annual cycle with a pronounced precipitation maximum in winter. In the DF considering, all SD models scenario RCP 8.5 shows an increase of $11.5\,\%$ for Sonnblick and $15.8\,\%$ at Zugspitze, although the standard deviation between models is quite high in both cases. Considering SD models combined, spring and autumn are characterised by an initial rise in the NF and a subsequent drop in DF. Nevertheless, except for the DF of RCP 8.5 in autumn, which is highly dominated by a strong reduction simulated by the RCFs at both stations, daily precipitation means are still higher than they were during the historical period. In comparison to the historical period, in summer a precipitation decrease is shown at Sonnblick, while at Zugspitze a first increase in the NF is followed by a decrease in the DF in both scenarios. Similar developments of future precipitation in summer are identified in literature. Smiatek et al. (2009) find decreasing tendencies of precipitation in summer in the Alpine region. A comparing study of RCMs and SD models by Schmidli et al. (2007) suggest that especially RCMs tend to simulate drier conditions in summer, while there is a large spread within the group of SD models. The latter are found to including rising tendencies as well. Comparing different SDs, the best agreement in the direction of change is found in the precipitation increase in winter, even if the relative change can differ quite largely. At Zugspitze, for example, the highest increase in winter is found by NNCs with $19.7\,\%$ (DF, RCP 8.5), while RCFs estimate a rise by only $5.5\,\%$ (DF, RCP 8.5). This is similar at Sonnblick: Here, ANNs project an increase of $19.6\,\%$ (DF, RCP 8.5), while with $5.3\,\%$, NNCs values are considerably lower. Largest discordance between direction of change in SD models can be found at Zugspitze in summer: NNCs suggest an overall future increase in any case, while in contrast, RCFs project decreases. Already indicated by the weak model performance in summer, these uncertainties in the change signal further emphasis the complex genesis of precipitation in summer and the challenge for SD models to derive reliable signals from the large-scale predictors (see also section 7.7). RCFs stand out to disagree quite often with ANNs

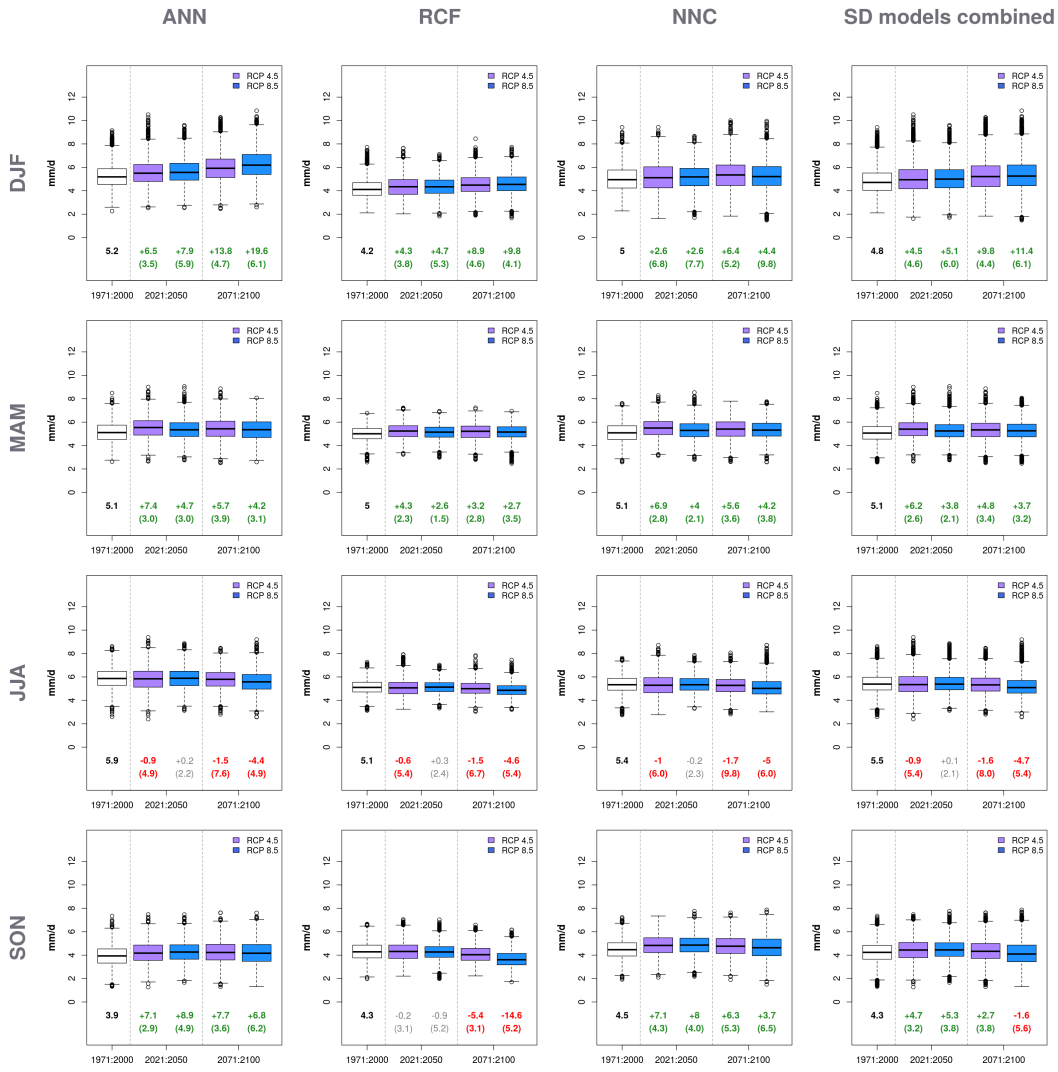


Figure 8.14: Seasonal dependent comparison of climate change signals separated by SD models considering all ESM downscaling products for daily precipitation sums at Sonnblick. Differences of the overall mean between the historical reference period and future scenarios are given relatively [%]. In addition, the standard deviation of the diverse change signals originating from different ESMs is listed in parenthesis. Significant increases are coloured in green, decreases in red. Significance is tested using the U-test (level of significance 5 %). Observational references: DJF: 3.9 mm d^{-1}, MAM: 4.8 mm d^{-1}, JJA: 4.5 mm d^{-1}, SON: 3.8 mm d^{-1}.

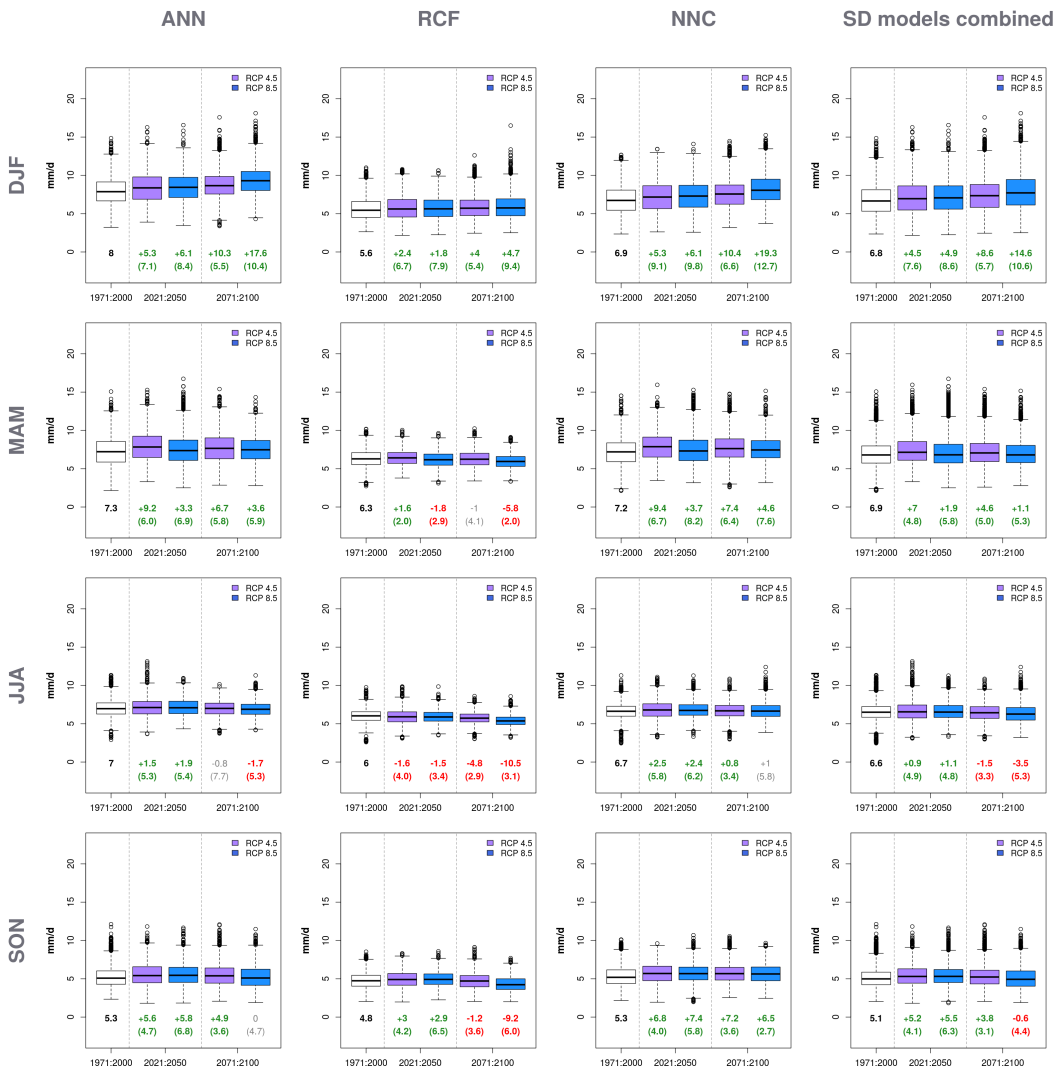


Figure 8.15: Expected seasonal change of daily precipitation sums at Zugspitze. Observational references: DJF: 5.6 mm d^{-1}, MAM: 6.0 mm d^{-1}, JJA: 5.9 mm d^{-1}, SON: 4.7 mm d^{-1}. See figure 8.14 for detailed description.

Table 8.3: Changes in standard deviation of daily precipitation at Zugspitze and Sonnblick considering the observational record (Obs) as well as downscaling products based on the reanalysis dataset (Rean, calibration and validation) and ESM experiments (historical: Hist; Scenarios: RCP 8.5 and RCP 4.5) separated by season and annually. Results from cross-validation folds and ESMs are averaged. Number in brackets in Hist: standard deviation between ESMs. Number in brackets in scenarios: standard deviation of the change signal. Reference periods: 1971-2000 (Obs, Rean, Hist); 2021-2050 (NF); 2071-2100 (DF). Changes: Increasing values are marked in green, declining ones in red and insignificant ones in grey (more than two ESMs show no change signal or a different sign.). Absolute change signals are computed by comparing historical and scenario results.

		Sonnblick							**Zugspitze**						
					RCP 4.5		RCP 8.5					RCP 4.5		RCP 8.5	
		Obs	Rean	Hist	NF	DF	NF	DF	Obs	Rean	Hist	NF	DF	NF	DF
DJF	ANN	6.7	5.1	4.6 (0.4)	12.0 (3.2)	19.2 (3.2)	13.5 (2.0)	24.6 (4.3)	10.1	8.6	8.1 (0.8)	8.3 (5.4)	8.2 (5.6)	8.3 (4.8)	11.4 (6.0)
	RCF	6.7	4.8	4.4 (0.5)	4.3 (3.3)	4.7 (4.6)	3.6 (2.7)	2.9 (2.9)	10.1	8.1	7.3 (0.8)	2.6 (4.7)	1.5 (5.0)	1.9 (3.4)	2.5 (4.8)
	NNC	6.7	5.7	6.2 (0.2)	3.1 (4.5)	4.2 (4.9)	2.8 (3.3)	4.6 (5.8)	10.1	9.3	9.2 (0.7)	5.1 (5.7)	4.8 (4.8)	5.6 (4.9)	8.0 (6.2)
MAM	ANN	6.9	4.7	4.0 (0.2)	12.5 (3.6)	10.1 (1.9)	10.1 (2.7)	9.9 (2.4)	10.2	7.2	8.1 (0.9)	12.3 (7.7)	10.4 (7.9)	7.0 (3.6)	5.9 (6.5)
	RCF	6.9	3.8	3.8 (0.0)	2.0 (2.6)	1.2 (2.2)	1.1 (1.8)	0.2 (2.2)	10.2	5.8	5.9 (0.1)	0.7 (2.6)	-1.7 (2.7)	-1.8 (1.8)	-6.4 (2.6)
	NNC	6.9	5.1	4.6 (0.1)	8.3 (2.7)	6.6 (2.1)	5.8 (2.6)	5.1 (2.4)	10.2	8.4	9.3 (0.7)	13.9 (8.6)	14.1 (10.0)	10.1 (4.4)	8.4 (7.8)
JJA	ANN	7.0	4.3	4.1 (0.4)	5.6 (5.0)	4.1 (3.9)	5.6 (3.0)	0.8 (2.8)	9.6	6.4	6.1 (0.7)	8.7 (6.9)	3.4 (6.2)	6.6 (4.5)	1.0 (7.7)
	RCF	7.0	3.7	3.7 (0.1)	0.0 (5.1)	-1.9 (3.0)	-0.2 (2.1)	-8.2 (2.8)	9.6	5.7	5.6 (0.3)	0.8 (4.7)	-4.5 (4.0)	-1.3 (3.8)	-13.9 (2.8)
	NNC	7.0	4.6	4.5 (0.2)	0.9 (4.5)	-1.7 (3.2)	0.4 (1.6)	-6.7 (2.9)	9.6	7.1	7.4 (0.5)	9.2 (8.6)	5.8 (8.3)	4.8 (5.1)	6.3 (7.0)
SON	ANN	7.1	5.0	4.4 (0.3)	10.1 (3.0)	12.5 (6.7)	13.7 (2.0)	14.0 (5.1)	9.0	7.0	6.6 (0.6)	8.5 (6.2)	10.8 (7.8)	10.9 (5.3)	6.1 (8.2)
	RCF	7.1	4.1	4.0 (0.1)	1.0 (1.7)	-1.3 (3.6)	1.1 (1.0)	-8.9 (4.4)	9.0	6.2	5.9 (0.4)	3.9 (5.3)	1.2 (6.5)	5.1 (3.6)	-6.6 (7.3)
	NNC	7.1	5.6	5.2 (0.1)	9.0 (4.8)	8.6 (5.8)	10.8 (3.6)	7.1 (6.5)	9.0	7.3	7.2 (0.4)	7.9 (6.0)	9.4 (6.7)	9.5 (3.1)	4.7 (5.7)
Annual	ANN	7.0	4.8	4.4 (0.2)	9.6 (1.8)	11.8 (3.9)	10.5 (1.7)	13.5 (4.1)	9.8	7.3	7.4 (0.5)	9.7 (5.3)	8.8 (4.6)	8.1 (3.2)	7.9 (4.7)
	RCF	7.0	4.1	4.0 (0.1)	2.1 (2.4)	1.2 (2.0)	1.6 (1.0)	-2.2 (1.4)	9.8	6.5	6.2 (0.3)	2.1 (3.5)	-0.3 (2.4)	1.1 (1.7)	-4.1 (2.0)
	NNC	7.0	5.3	5.2 (0.1)	5.4 (2.0)	4.9 (2.5)	5.0 (1.0)	3.7 (2.3)	9.8	8.1	8.4 (0.4)	9.4 (5.9)	9.1 (4.1)	7.6 (3.6)	7.7 (4.3)

and NNCs in magnitude and direction of change. In numerous cases, RCFs tend to project precipitation reductions, where ANNs and NNCs concordantly simulate increases. This effect can once again emerge from the difficulty of RCFs to deal with changing conditions.

In table 8.3, the standard deviation of daily precipitation is listed separated by season and annually. Change signals are computed comparing the historical results with future scenarios of the downscaled ESMs. Future changes are considered significant if five of seven ESMs show positive statistical test results and in addition point in the same direction of change (increase or decrease). Changes are tested for significance using the Fligner-Killeen test considering a significance level of $5\,\%$. The observed variability of daily precipitation sums is under-represented in simulations (rean and hist). The standard deviation of SD model time series driven by historical ESM datasets is typically further reduced if compared to the reanalysis based ones. ANNs and NNCs suggest significant increases in variability for the future with the exception of summer. In contrast, time series originating from RCFs often show a reduction in variability in summer, whereby significant values are primarily found in the DF. The reason is an unilateral redistribution of type frequencies.

Additionally, precipitation indices of the observed and modelled times series are evaluated with respect to precipitation intensity and temporal behaviour (Duan and Mei 2014; Isotta et al. 2014b; Smiatek et al. 2013; Schmidli et al. 2007; Marengo et al. 2009). Some of these indices need to separate wet from dry days. A threshold of $> 1\,\mathrm{mm\,d^{-1}}$ defines a wet day in this thesis.

Time series features referring to precipitation intensities are summarised in the appendix in table B.1 for Zugspitze and table B.3 for Sonnblick. The tables include the standard precipitation index (SDI), which is computed from the mean of wet days, giving the average event intensity. Two metrics are focusing on extreme events: On the one hand, the mean annual maximum five day sum (P5max) focuses on short-time steady rain, for example originating from frontal systems and on the other hand the frequency of events exceeding the observational $90\,\%$ (Pf90) quantile shift in individual heavy precipitation days. As statistical test, the U-test is chosen to detect changes in SDI and P5max, while Pearson's χ^2-test is used to analyse future Pf90 exceedance. Comparing the SDI of observational datasets and downscaled precipitation simulated by reanalysis or historical ESM inputs, the simulated time series underestimate intensities, which seems to be a typical problem in the Alpine region (see Frei et al. (2003)). On the other hand, the SDI of reanalysis and historical results matches quite well, suggesting a good transferability of the calibrated models onto the ESM datasets. Although the absolute differences at Zugspitze are higher, the relative underestimation is similar. At both stations and across all SD models, scenarios indicate a future increase in precipitation intensities in winter and further

at Sonnblick in spring, although especially RCFs tend not to be significant. Smiatek et al. (2009) also find strong increases in winter precipitation intensity in the Alpine region. ANNs and NNCs in spring and autumn at both stations are characterised by an increasing SDI as well. In contrast, RCFs suggest an overall future decline of SDI magnitudes at Sonnblick as well as in the DF of Zugspitze in autumn. In summer, primarily RCFs show decreasing SDIs throughout future scenarios. In contrast, ANNs and NNCs usually show increases at Sonnblick, which coincides with the findings of Smiatek et al. (2009), and an indifferent signal at Zugspitze. Annually ANNs and NNCs suggest an overall future increase in SDI, while RCF's change signals are usually either insignificant or decreasing. Considering the extreme value statistics P5max and Pf90, all SD models underestimate the observational time series. Strong deficits are found especially in summer. This indicates that models have trouble reproducing heavy precipitation events, which in summer often originate from local convection processes resulting in thunderstorms. With few exceptions, there is not much difference between SD models driven by historical ESM runs and reanalysis data inputs. If significant, ANNs as well as NNCs suggest throughout all seasons and annually an increase in either P5max or the Pf90. Once again, RCFs often show contradictory results in particular in the DF of both scenarios although the results, especially of P5max, need to be treated with caution due to the fixed value range. Dobler et al. (2013) witnessed a similar issues in the context of downscaling extreme precipitation in the Alpine region. Statistical models in this study, too, tend to point in different directions at times.

Further, precipitation characteristics focusing on the temporal behaviour are listed in the appendix in table B.2 for Zugspitze and table B.4 for Sonnblick. The wet day frequency (WDF) points at changes in the occurrence of precipitation. Further, the consecutive dry days (CDD) and consecutive wet days (CWD) are counted to identify changes in wet and dry spells. Pearson's χ^2-test is applied to evaluate future WDF changes, CDD and CWD are checked for significance using the U-test. Comparing the simulations based on reanalysis or ESM historical runs with observations, the WDF at both stations is much too high. This typical effect in statistical downscaling of precipitation oftentimes compensates the deficit of heavy events in the total precipitation sums (Maraun et al. 2010). WDF of reanalysis and ESM driven results are again similar. A strong disagreement between the SD models is witnessed in future simulations. Annually, at Sonnblick, the WDF is estimated to increase by ANNs, decrease by RCFs, while no significant change is detected by NNCs. This sequence is pre-eminent in seasons as well with minor discrepancies. A similar picture is drawn at Zugspitze annually, although less significant changes are found and NNCs show one reduction in frequency as well (NF RCP 8.5). In DJF, primarily higher WDFs are estimated, in the other months, lower WDFs are predominant. The in general high

number of insignificant change values and discrepancies between the models indicate a high uncertainty in the future development of WDFs at both stations with best agreement in an increase in winter. Comparing CDDs and CWDs, observed and simulated (based on reanalysis and ESM historical datasets) time series, following the previous results, there are too little CDDs and too many CWDs. As expected, CDD and CWD show diametrically opposed behaviour in a model series, but the counterpart does not necessarily need to be significant. At both stations and throughout the seasons, detached small significant changes are identified by single models with bearing very low differences. Best model agreement is found at Sonnblick in the DF of RCP 8.5 with a decrease in CDD and increase in CWD.

8.5 Summary: Precipitation

In comparison to the foothills of the Alps, summit stations show higher annual precipitation sums. Additionally, at the edge of the mountain range, where the Zugspitze is located, higher amounts can be expected than in the centre, where Sonnblick is located. This is also evident in the reference period from 1971 to 2000, where the annual precipitation sum from Zugspitze surpasses that of Sonnblick. The annual sums of both stations are characterised by a large interannual variability. The relationship between the daily sums of Zugspitze and Sonnblick, especially in the reference time slot, is quite high, with highest correlation coefficients in winter and lowest in summer. In the first half of the 20th century, annual precipitation sums at Sonnblick show a gradual reduction followed by a steady increase in the second half, valid for entire years and seasonally. At Zugspitze, except for summer, especially the long-period time windows are characterised by positive trends. Summer is, in addition, the only season with a systematically significant decrease around the central years 1930 to 1940. Except for Sonnblick in winter, where an increase is detected, both stations show no significant trends during the calibration period 1971 to 2000.

The search for variations in the statistical relationship between predictors and predictand point out that best relationships are generally found at the end of the 20th century, which amongst others lead to the definition of the calibration period. The comparison of the quantiles in the reference period shows that calibrated SD models typically produce too many small precipitation events and underestimate the magnitudes of heavy events. This is also evident from the lower values in the indices SDI, P5max, Pf90 and CDD of the simulated runs. In contrast values of WDF and CWD are overestimated by the statistical models. Best representation of heavy precipitation is simulated with NNCs. The locally observed annual course of mean monthly precipitation sums is well reproduced by reanalysis driven SD models. The transfer of SD models on historical ESM datasets works best with RCFs, while ANNs and

NNCs tend to overestimate the monthly means.

To investigate changes in climatic conditions, the downscaled ESM scenarios RCP 4.5 and RCP 8.5 are compared with historical runs, considering the near (2021 to 2050) and distant future (2071 to 2100). The overall annual averages of daily precipitation sums projected by ESMs tend to increase at both stations. Thereby, the rise in RCP 4.5 is in most cases larger than in RCP 8.5. Individually evaluated, the RCFs changes are standing out: In contrast to ANNs and NNCs, ESM-specific RCF results tend to be mostly insignificant and point in a different direction. Apart from RCFs, in general an increase in precipitation intensity can be expected, as shown by higher SDI, P5max and Pf90 magnitudes. Annual CWD and CDD changes are however barely significant. In addition, there is a large disagreement between the SD models on the WDF. Separated by seasons, the strongest SD model consent is achieved in winter, where large increases in seasonal means, SDI, P5max and Pf90 appear. A similar signal is simulated at Sonnblick in spring. Seasonally, precipitation reductions are detected as well, in particular in summer. Between the SD models, good conformity in the evaluated statistics and indices, at least in the direction of the climate signal, is achieved by ANNs and NNCs. In contrast, RCFs often tend to point in the exact opposite direction.

9 Air Temperature

9.1 An Overview over Drivers and Distribution of Air Temperature in the Alpine Region

In general, the air temperature in the Alps at a specific time depends on basic parameters like seasonally dependent solar radiation, the diurnal cycle or mesocale wind systems, like the mountain breeze and valley breeze or travelling weather systems, which can advect cold or warm air masses.

The annual average temperature in the GAR is shown in figure 9.1 for the period 1961-1990. Temperatures range from $-6\,^{\circ}\mathrm{C}$ in the central Alps to nearly $14\,^{\circ}\mathrm{C}$ in the southern foothills. Towards higher altitude, the temperature decreases approximately by a long-term mean lapse rate of $0.65\,\mathrm{K}$ per $100\,\mathrm{m}$ following WMO (1970), which can be seen as a standard magnitude for Europe. Proxies for temperature like tree line and snow line occur at higher elevations in the central Alps. This incident is called "mass elevation effect" and was introduced by Quervain (1904) but is still investigated in recent literature (Han et al. 2012). The mass elevation effect can be traced back to either wind streams (Richter 2000) and/or heating surfaces (Weischet and Endlicher 2008). One important aspect of wind stream influence in the Alps is the Lee side of the Foehn (detailed description of the Foehn in section 8.1) with relatively warmer air temperatures in comparison to the Luv side. In addition, the mass elevation effect can be triggered or increased by highly elevated heating surfaces. On the one hand, during the insolation period, latent heat can be transferred from the surface to the atmosphere leading to a higher temperature than in regions outside of the mountain range at equal altitudes. On the other hand, during periods with higher energy emission, gravity wind descends from summits through the valleys. The mass deficiency will be replaced by a downward wind stream heating up on the way by the adiabatic process, once again leading to a higher temperature when reaching the surface. This process was already described for the Sonnblick by Steinhauser (1938).

If the increase of temperature with height in the atmosphere is higher than the particular adiabatic curve, a temperature inversion occurs. In this case, air exchange of low level areas, for example of valleys, is prevented. High fog is a common result

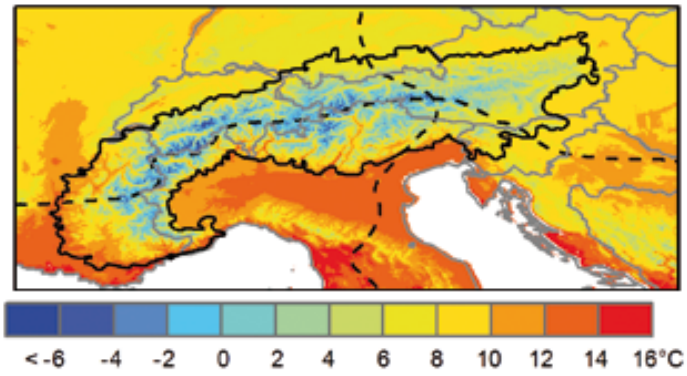
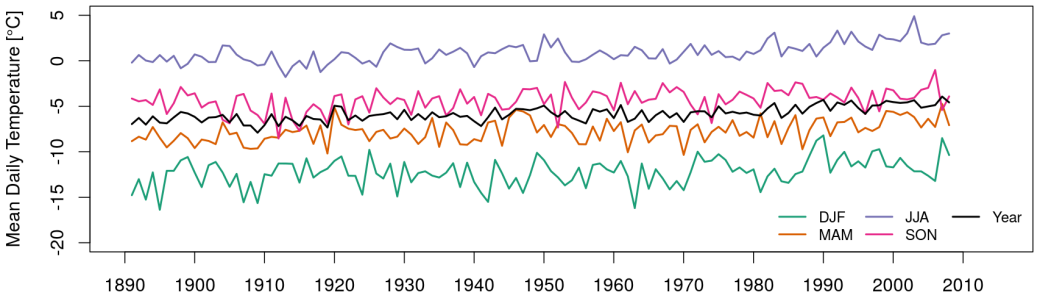


Figure 9.1: Annual mean temperature of the European Alps for the period 1961-1990. Regional statistics: GAR = 8.5 °C, Alps (thick black line) = 5.5 °C, summits higher than 1500 m = 1.1 °C. Source: Agency (2009) based on data from Auer et al. (2008).

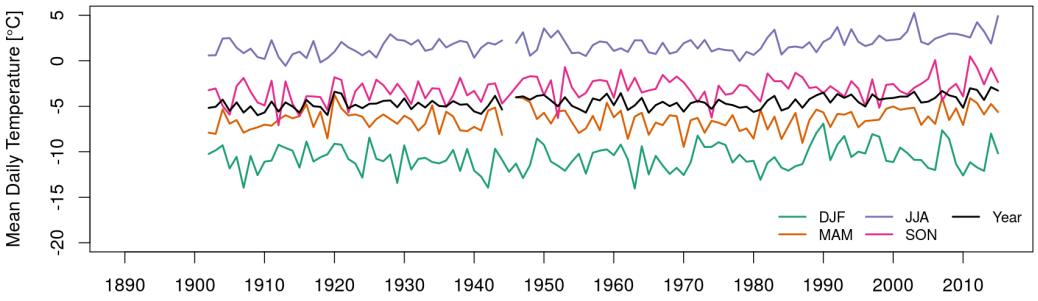
of this atmospheric stratification. According to Whiteman (2000), such layers can form, when the ground cools faster than the air above (typical in winter or during night time), if cold air is brought into valleys by gravity winds or cold air currents as well as when warm air is advected by winds aloft. It can be a difficult task to include this effect in statistical downscaling as the affected areas can be rather small.

The inner Alps are characterised with less cloud cover in comparison to the northern and southern margins, resulting in a greater sunshine duration in the central zone. This goes hand in hand with higher totals of absorbed global solar radiation at daytime in summer but also high emission at night time in winter, leading to larger annual amplitudes of mean daily maximum temperature: Barry (2008) specifies the spacial contrast with a range of $23.8\,\mathrm{K}$ in the centre to $20.8\,\mathrm{K}$ at the edges at $500\,\mathrm{m}$ elevations and $17\,\mathrm{K}$ and $15\,\mathrm{K}$, respectively, at $2000\,\mathrm{m}$ elevation.

In higher latitudes, the aspect of hillsides is becoming increasingly important for temperature. A noticeable difference in temperature can be found comparing sunny southern slopes (often called adret) and shaded northern slopes (often called ubac) in the pronounced east-west valleys of the Alps (see section 3.1). This effect is most apparent in the winter months, when the sun angle is low. According to Barry (2008) the difference can reach magnitudes of up to $3\,\mathrm{K}$ which, among other, impacts snow cover or even land use on the respective flank. Southern slopes, for example, show a reduced snow cover of approximately 130 days in comparison to 160 days on the northern sides.



(a) Sonnblick



(b) Zugspitze

Figure 9.2: Annual and seasonal time series of the averaged observed air temperature.

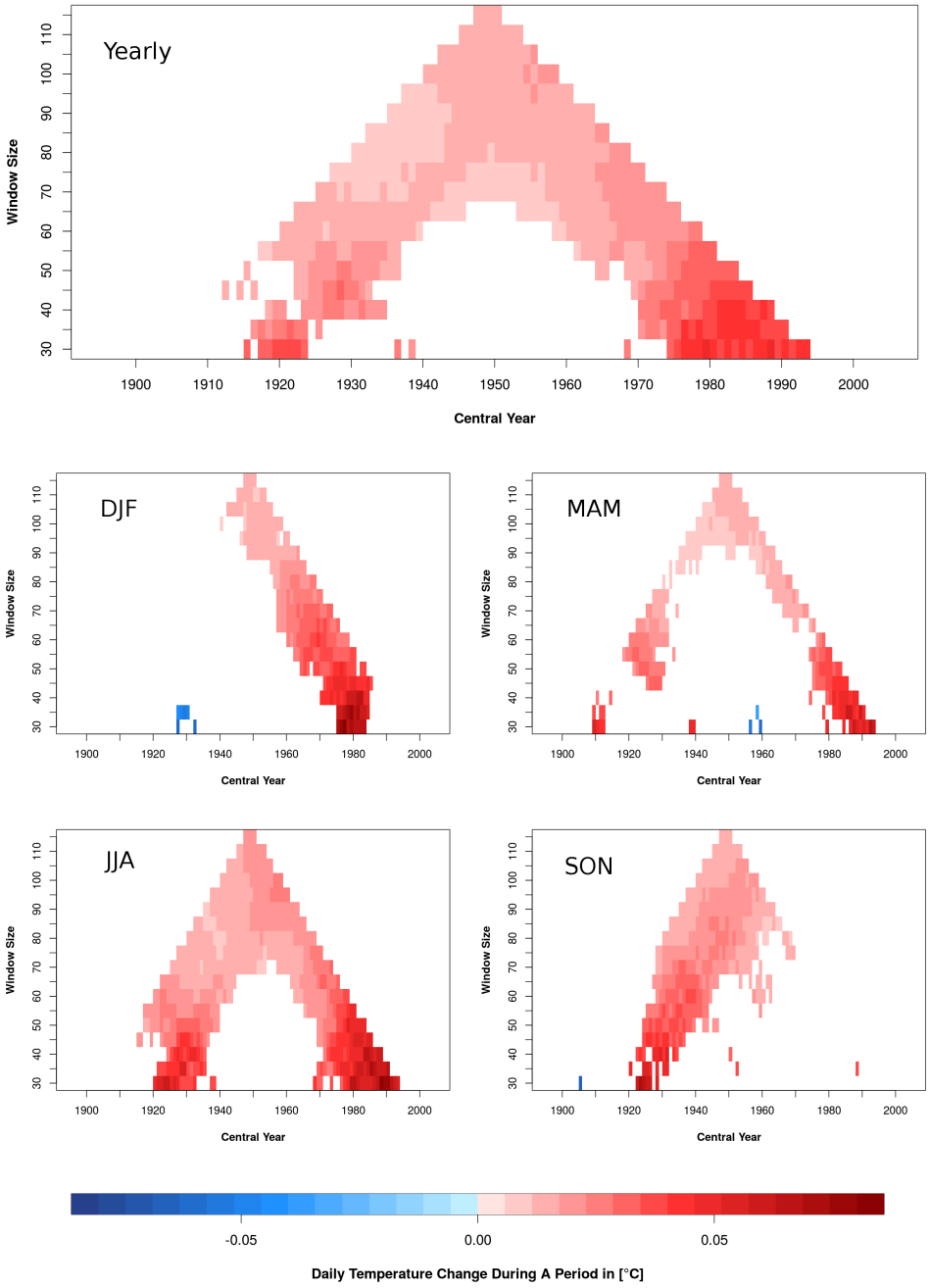


Figure 9.3: Trends in the observational time series of the seasonally and annually aggregated Sonnblick temperature record. The x-axes represents the central year of the evaluated time period, the y-axes lists corresponding time window sizes in years. Trends are displayed for significance by the trend-noise ratio (level of significance 5 %).

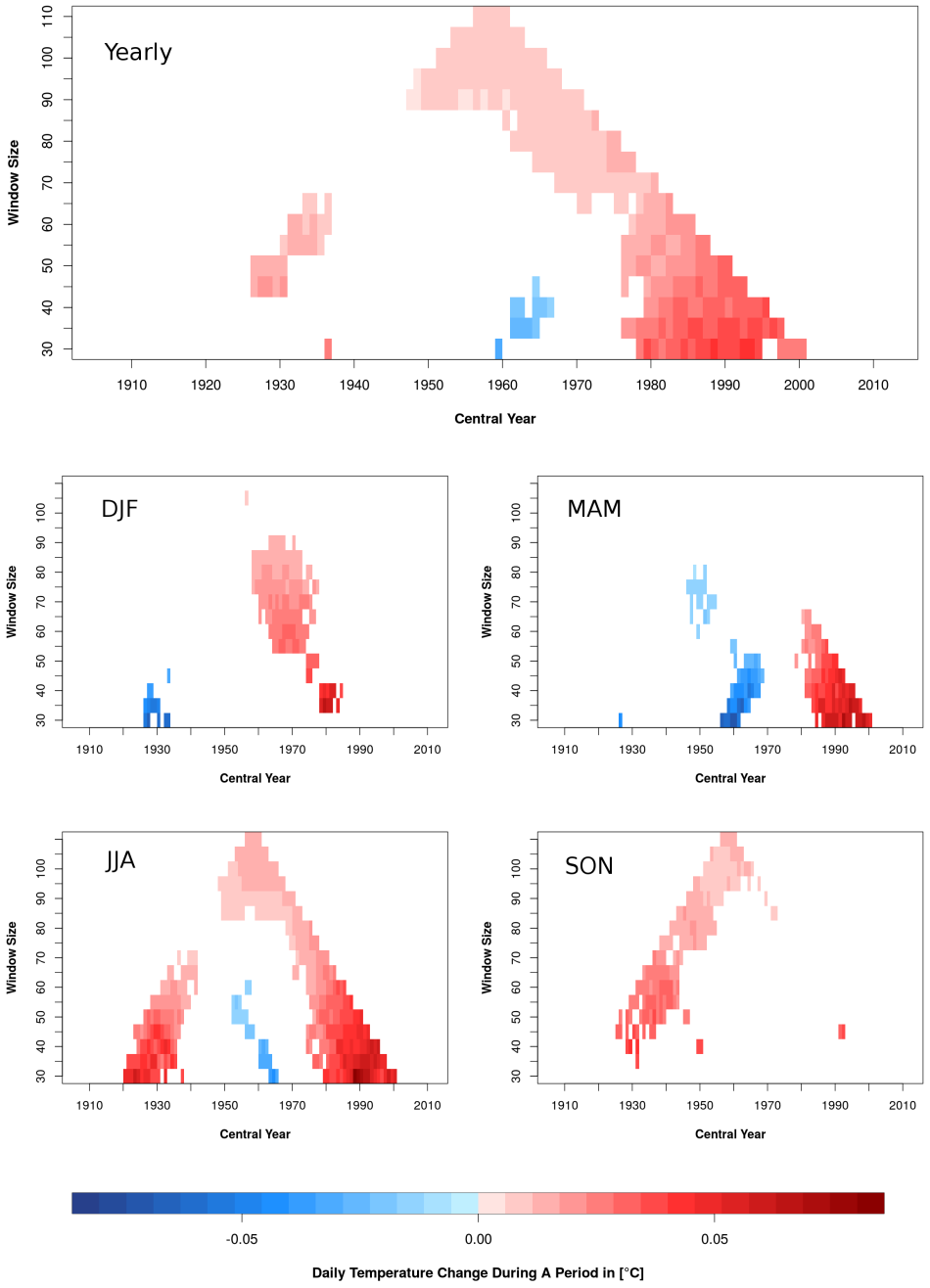


Figure 9.4: Trends in the observational time series of the seasonally and annually aggregated Zugspitze temperature record. For a detailed description, see figure 9.3.

Table 9.1: Spearman correlation between Sonnblick and Zugspitze daily mean temperature of the maximal congruent (first line) and reference time period (second line). Indices are computed for the overall period and seasons, time steps with missing values are omitted. All coefficients are significant on a significance level of 5 %. In addition, a cross-correlation procedure is carried out. Correlation coefficients marked with $x^{(t_s-1)}$ are found to be higher with a temporal offset. In this example, often times, Sonnblick time steps (t_s) shifted by one day backwards ($t_s - 1$) compared to Zugspitze (t_z) in the shorter reference period show higher coefficients.

	Total Period	**DJF**	**MAM**	**JJA**	**SON**
1901-2008	0.96	0.89	0.92	0.89	0.93
1971-2000	$0.96^{(t_s-1)}$	$0.90^{(t_s-1)}$	0.92	$0.9^{(t_s-1)}$	0.91

9.2 Analysis of the Observational Datasets

In the reference period from 1971 to 2000, the observed average annual daily mean temperature at Sonnblick is with a value of $-5.3\,°\mathrm{C}$ by $-0.7\,°\mathrm{C}$ colder than at the Zugspitze ($-4.6\,°\mathrm{C}$). Coldest month at both stations is February, reaching mean temperatures of $-12.1\,°\mathrm{C}$ at Sonnblick and $-10.9\,°\mathrm{C}$ at Zugspitze. The warmest month is August with an average temperature of $2.5\,°\mathrm{C}$ at Sonnblick and $2.8\,°\mathrm{C}$ at Zugspitze. The amplitude of the annual cycle is therefore more pronounced at Sonnblick. The interannual variability is rather low at both stations, reaching values of $0.64\,°\mathrm{C}$ at Zugspitze and $0.62\,°\mathrm{C}$ at Sonnblick. Annual minimum and maximum mean temperatures are $-5.8\,°\mathrm{C}$ (1980) and $-3.4\,°\mathrm{C}$ (1989) at Zugspitze and $-6.2\,°\mathrm{C}$ (1980) and $-4.2\,°\mathrm{C}$ (2000) at Sonnblick. Figure 9.2 displays the observed seasonal and annual time series. Of the transitional seasons at both stations, autumn shows slightly higher temperatures than spring. A detailed analysis of trends in the observational period is given in figures 9.3 and 9.4 (identical methodology to figures 8.5 and 8.6). Considering the annual averages at Sonnblick, if significant, a positive trend is witnessed. On the medium and short scale, particularly periods centred at 1940 to 1970, show a stagnation or even decreasing temperature magnitudes at both stations. This effect is mainly caused by anthropogenic particulate matter leading to a temporal cooling (see chapter 2). In the recent history, strong increases in temperature are found, starting approximately from periods centred around 1980, with the exception of autumn, where at both stations no strong trends are displayed. The strong positive trend starting from 1980 is observed in the whole Alpine area (Agrawala et al. 2007; R. O. Weber et al. 1997) and can be traced back to the enhanced green house effect. In total, an overall increase of $2\,°\mathrm{C}$ is observed during the 20th century in the Alpine region (Beniston 2016).

In comparison to precipitation, the relationship of temperature at Zugspitze and Sonnblick is very high especially in all year view (see table 9.1), which is typical for

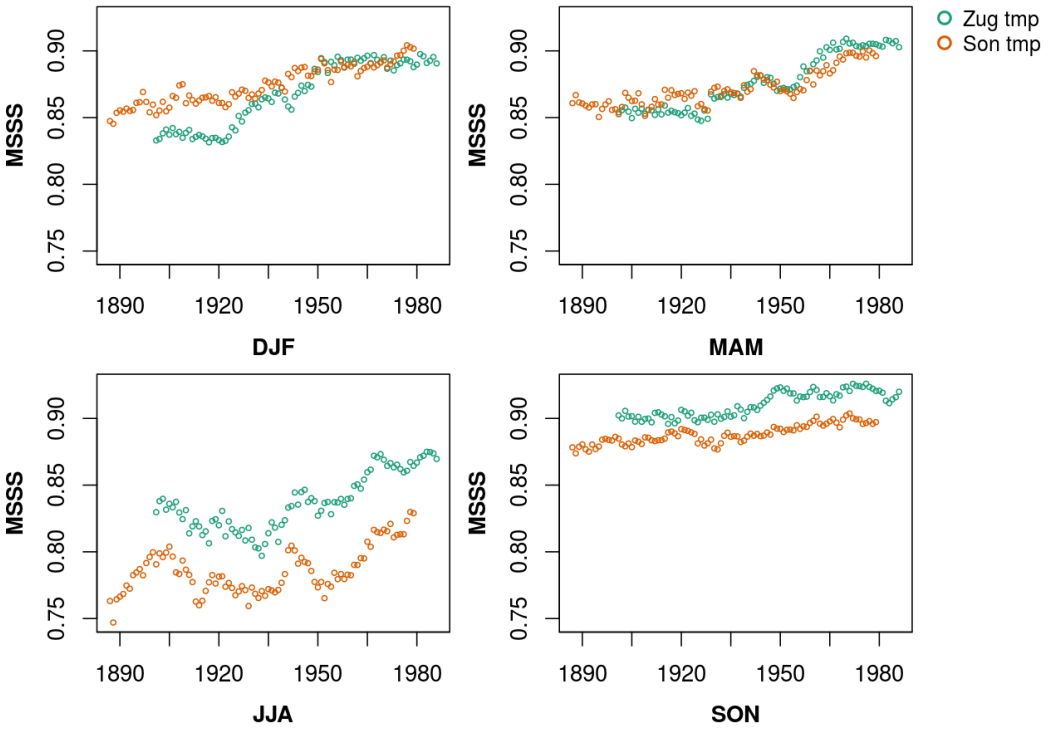


Figure 9.5: Seasonally dependent analysis of the temporal altering model skill of the final ANN configuration evaluated by the MSSS. Models are calibrated using observed temperature (tmp) at Zugspitze (Zug) and Sonnblick (Son) and the 20th century reanalysis dataset. Training is repeated from 15 different ANN initialisations, the resulting skills scores are averaged. Models are calibrated on a 30-year period, starting at each mark of the graphic.

Alpine summits (R. O. Weber et al. 1997). A likely reason is the direct exposure to the temperature advection from the free atmosphere. Seasonally, transitional periods are characterised by higher correlation indices than winter and summer, whereas the differences between the two periods 1901 to 2008 and 1971 to 2000 are quite low. Especially in the reference period, a cross-correlation approach suggests to shift the time series of Sonnblick (t_s) one day backwards ($t_s - 1$). In practice, Zugspitze temperature time series is correlated with the respective values of the next day at Sonnblick.

9.3 Statistical Model Results with Focus on the 20th Century

The experimental design of the search for best predictor-predictand relationship in figure 9.5 is equal to figure 8.7. Throughout the 20th century, a rise in model skill is observed in all seasons, only in summer, the MSSS shows a higher rate of fluctuations in the first half of the 20th century. It is again suggested to restrain the calibration

period onto the end of the available datasets, where the highest MSSS values are found. Between the stations, the MSSS in winter and spring is either rather similar at first hand, or approaches an equal level until the end of the century. In contrast, MSSS values of Sonnblick in summer and autumn are lower than at Zugspitze.

The overall replication of the observed temperature quantiles at Sonnblick (figure 9.6) and Zugspitze (figure 9.7) by reanalysis and ESM based ANN, RCF and NNC simulations is quite accurate. Congruent QQ-plots at both stations display typically similar behaviour. Already indicated by the high model performance in section 7.7, calibration and validation quantiles fit the observational quantiles to a very high degree. Only in the uppermost and lowest quantiles, slight differences occur. The transfer of the SD models onto historical ESM datasets works satisfyingly, although a larger spread at the tails of the distribution is observed. Considering the lower quantiles, in winter, summer and autumn colder temperatures than observed are simulated and warmer ones in spring. The highest quantile is often overestimated, except for winter. In particular in the centre of the distribution, ESM-series-based differences are very low, beside single outliers, largest disagreement is found in summer, where MPI-ESM and CMCC based simulations tend to underestimate the observed quantiles. NNC products show the best conformity with observations throughout the distribution and further perform best in the context of extremes. RCFs have a good accordance in the centre of the quantile range, but show the typical problems at the tails of the distribution. Largest differences are resulting considering the ANN time series. Figure 9.8 shows that the annual cycle is well reproduced by all SD models, with minor differences to the observations. The best fit is found to be achieved by RCFs, closely followed by NNCs, while especially in summer and winter ANNs are characterised by the largest spread. NNC downscaling products are in general too low, while ANNs scatter around the observed means.

9.4 Climate Change Signals: 20th vs. 21st Century

At first, the RCF characteristics and their behaviour in the climate change analysis is discussed. The average reference values of the RCFs derived from the observational dataset in the calibration period are displayed in figure 9.9 for Sonnblick and Zugspitze. In most seasons, the type-dependent decrease in temperature is rather linear in the centre. As large-scale temperature is an applied predictor in all RCFs, the linear dependency between the local and large-scale is not surprising. On the other hand, in the higher class numbers, an usually more rapid decrease is found. This indicates that the large-scale temperature, which has to satisfy a mixture of local temperature values in a widespread area, does not reproduce or include the low local temperature magnitudes at Zugspitze and Sonnblick. Although less pronounced,

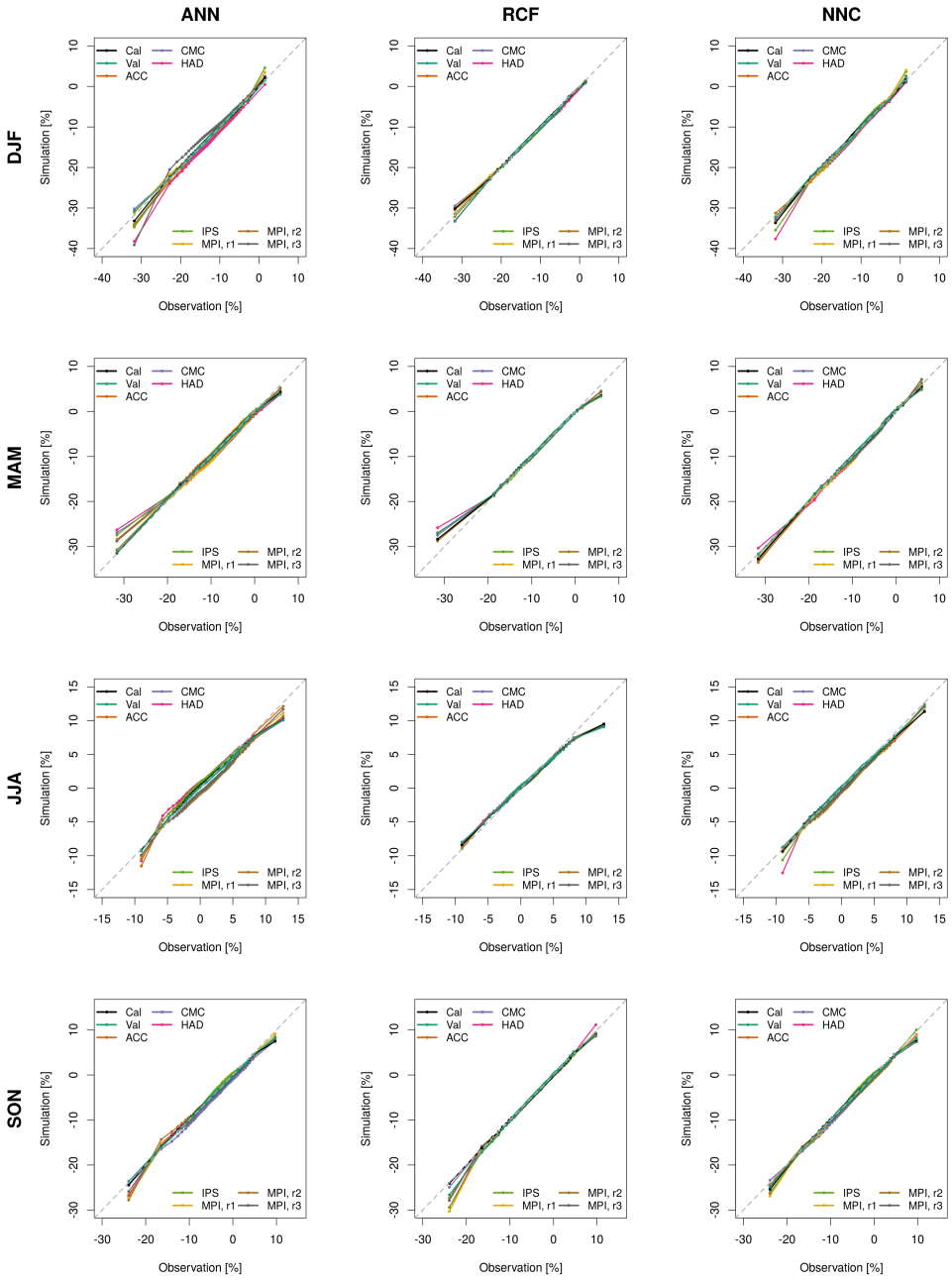


Figure 9.6: QQ-plots for Sonnblick temperature considering the seasonal models in the reference period 1971-2000. 50 quantiles of the observed dataset are compared with their simulated counterparts from reanalysis and historical ESM data-based downscaling products derived by the final SD models. Reanalysis series are separated by calibration (Cal) and validation (Val) periods. The fifteen Cross-validation runs are treated as one sample.

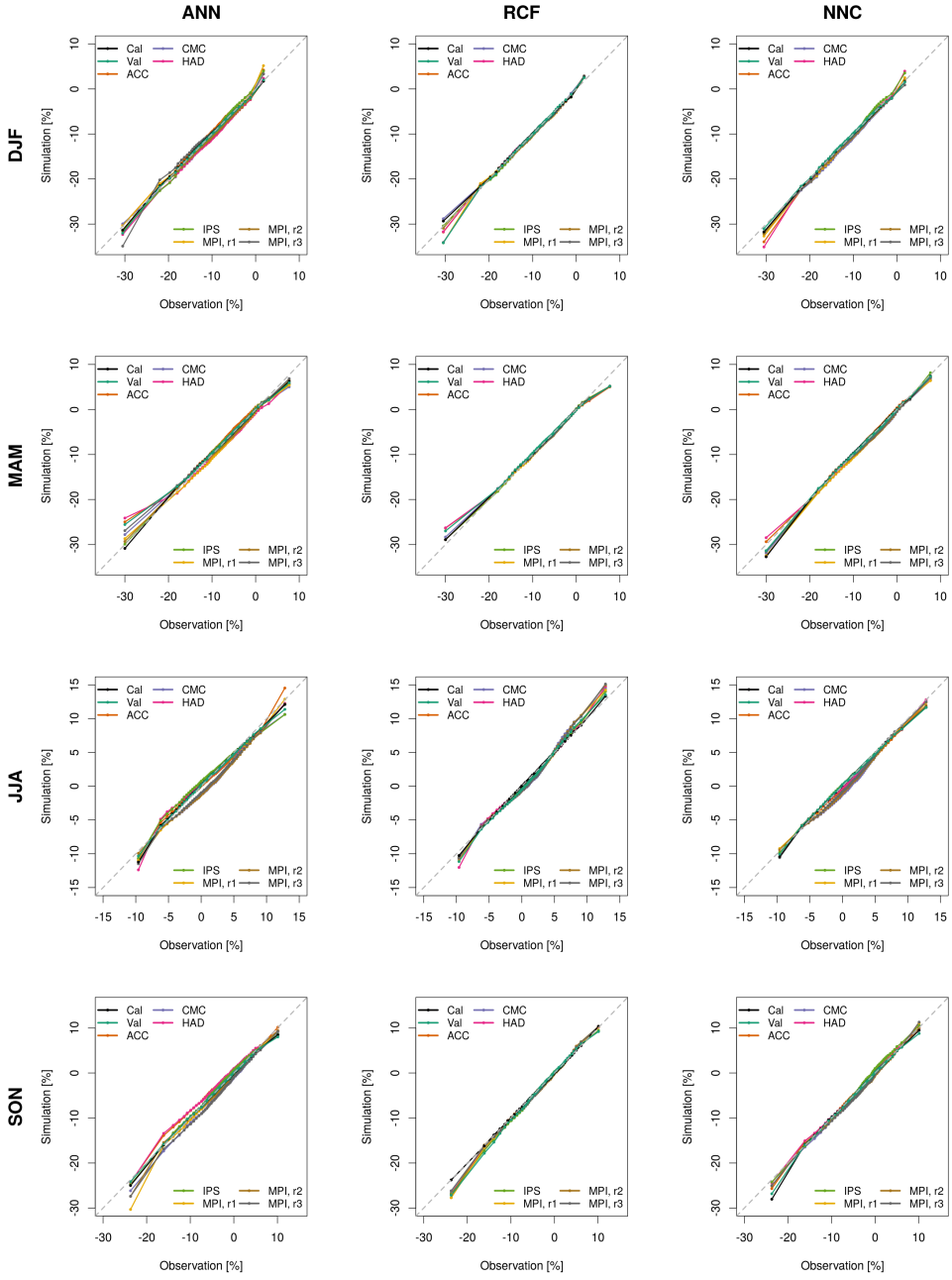
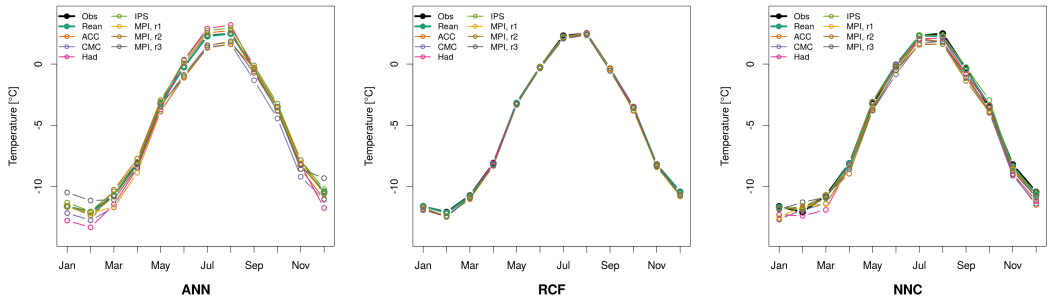
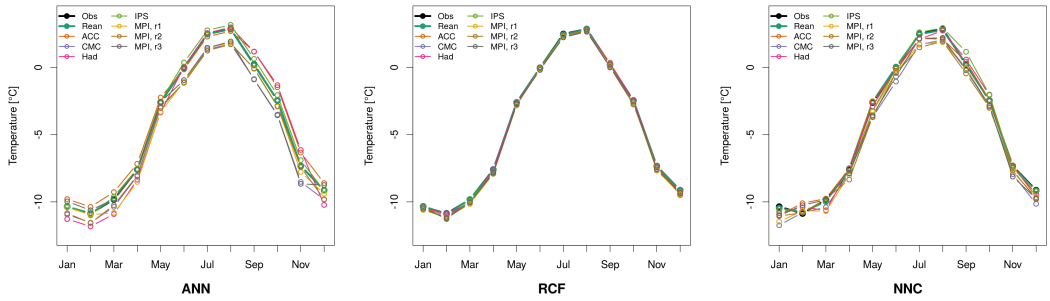


Figure 9.7: QQ-plots for Zugspitze temperature. See figure 9.6 for detailed description.

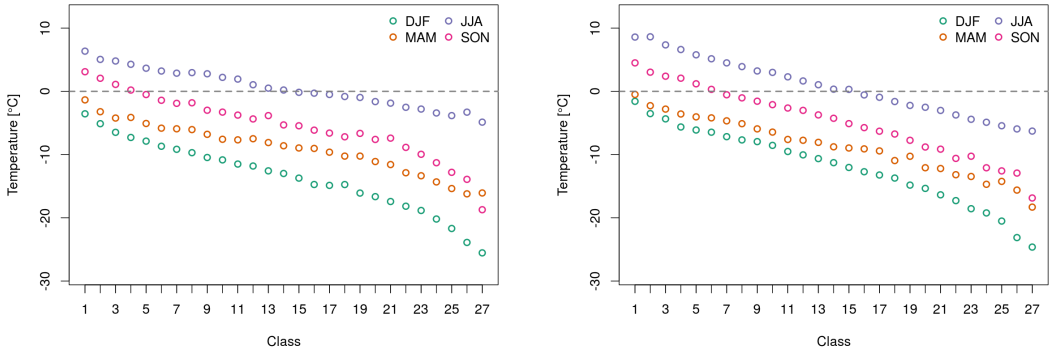


(a) Sonnblick



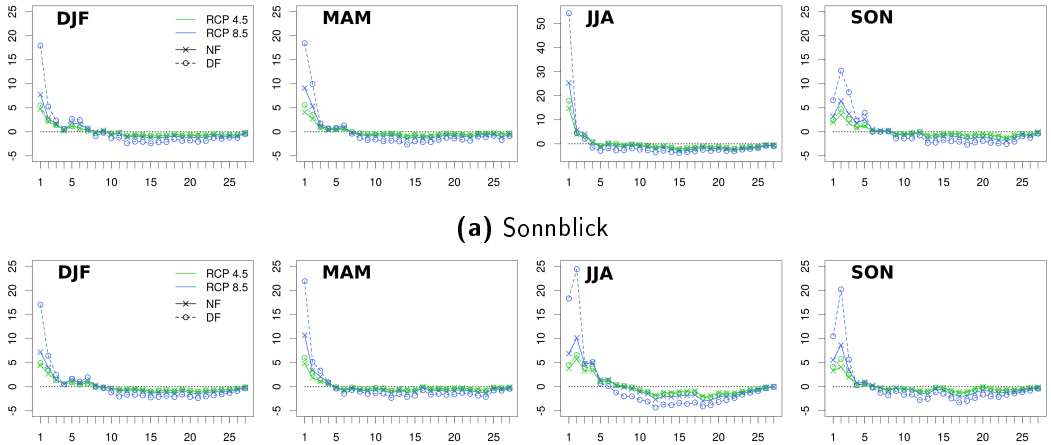
(b) Zugspitze

Figure 9.8: Observed (Obs) and simulated (complete reanalysis (Rean) and climate model driven results) monthly averaged daily mean temperatures considering the reference period 1971-2000 for each SD model type.



(a) Sonnblick **(b)** Zugspitze

Figure 9.9: Circulation type specific mean temperature reference values in [°C] for each of the 27 classes of the seasonal RCFs at Zugspitze and Sonnblick derived from the calibration period 1971-2000.



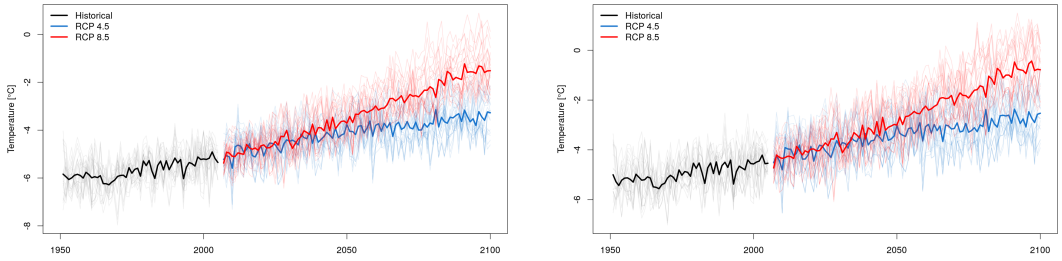
(a) Sonnblick

(b) Zugspitze

Figure 9.10: Average future changes of type frequency in [%] comparing the historical reference (1971-2000) of all ESMs with both scenarios (RCP 4.5 and RCP 8.5) in the near (NF) and distant future (DF) of the seasonal RCF temperature models at Sonnblick and Zugspitze.

a higher non-linear increase in temperature is found in the lower class numbers. The seasonal reference values of identical class numbers follow the annual temperature cycle. Highest values are found in summer, followed by autumn, spring and winter. Only one intersection is observed at type 27 at Sonnblick in the transitional seasons. Changes in frequency on type occurrence in figure 9.10 point out a strong increase of lower class numbers, thus the warmest reference temperatures. Frequency variations are usually pointing in the same direction in scenarios and future time slices, while RCP 8.5 and DF show stronger changes than RCP 4.5 and NF. Usually, the lowest classes are subjected to a sharp increase indicating that a class with higher reference values might be necessary to produce reliable temperature magnitudes in future scenarios. The very low temperature types, in particular type 27, show only very small decreasing tendencies. This means that low reference values are still in usage and needed. Biggest frequency losses are primarily found near the centre and up to higher numbered types.

In the following the climate change analysis results of ANNs, RCFs and NNCs are discussed. The curves in figure 9.11 visualise the temperature development of the ESM downscaling products between 1950 and 2100. Additional statistics of the overall mean as well as the impact from different ESM influences is described in table 9.2. Individually analysed, as well as in the multi model mean approaches, a similar tendency is pictured by the SD models. At both stations a continuously significant increase of the annual mean temperature in the selected periods is projected, while the resulting rise of the RCP 8.5 scenario is considerably higher than of the RCP 4.5. At both stations, the change signals of the scenarios are fairly equal, in the NF even



(a) Sonnblick **(b)** Zugspitze

Figure 9.11: Annually averaged daily temperature means resulting from the ESM downscaling product of the total available time period 1950 to 2100 (historical and scenario dataset). Thick lines represent the overall average of ESM results. Thin lines in the background display the mean of the cross-validation procedure considering a single SD model and ESM combination.

Table 9.2: Historical (Hist: 1971 to 2000) annually averaged temperature [°C] at Zugspitze and Sonnblick as well as projected absolute changes to the historical period considering different ESMs, SD models and scenarios. Changes are computed for the near future (NF: 2021 to 2050) and distant future (DF: 2071 to 2100). Positive numbers are marked in green, negative ones in red, while insignificant values (U-test, significance level 5 %) are labelled in grey. (Observational average Sonnblick: −5.9 °C; Zugspitze: −4.5 °C)

		Sonnblick					Zugspitze				
			RCP 4.5		RCP 8.5			RCP 4.5		RCP 8.5	
		Hist	NF	DF	NF	DF	Hist	NF	DF	NF	DF
ACC, r1	ANN	-5.3	1.5	2.5	1.7	4.7	-4.1	1.2	2.1	1.4	4.1
	RCF	-5.4	1.4	2.2	1.5	3.7	-4.6	1.5	2.4	1.6	4.4
	NNC	-5.7	1.4	2.3	1.6	4.3	-4.6	1.2	2.0	1.4	3.9
CMC, r1	ANN	-6.1	1.4	2.1	1.3	3.8	-5.4	1.4	2.0	1.2	3.6
	RCF	-5.4	1.3	1.9	1.2	3.2	-4.7	1.5	2.2	1.3	3.8
	NNC	-5.9	1.4	2.1	1.3	3.7	-5.2	1.4	2.0	1.2	3.6
HAD, r1	ANN	-5.6	1.4	2.6	2.1	4.6	-4.7	1.2	2.3	1.8	4.2
	RCF	-5.4	1.3	2.4	1.9	3.8	-4.6	1.5	2.7	2.1	4.6
	NNC	-5.7	1.4	2.5	2.0	4.3	-4.7	1.2	2.3	1.8	4.1
IPS, r1	ANN	-5.0	1.5	2.5	1.7	4.5	-4.3	1.3	2.3	1.5	4.2
	RCF	-5.4	1.4	2.3	1.6	3.8	-4.7	1.5	2.6	1.8	4.5
	NNC	-5.3	1.5	2.4	1.7	4.3	-4.5	1.3	2.2	1.5	4.1
MPI, r1	ANN	-5.7	0.8	1.5	1.0	3.1	-5.1	0.8	1.5	1.0	3.0
	RCF	-5.5	0.8	1.4	1.0	2.7	-4.7	0.9	1.6	1.1	3.2
	NNC	-5.7	0.9	1.5	1.0	3.1	-5.1	0.8	1.4	1.0	2.9
MPI, r2	ANN	-5.5	0.8	1.6	1.0	3.0	-5.1	0.8	1.5	0.9	2.8
	RCF	-5.4	0.8	1.5	0.9	2.6	-4.7	0.8	1.7	1.0	3.0
	NNC	-5.7	0.9	1.6	1.0	2.9	-5.1	0.8	1.5	1.0	2.8
MPI, r3	ANN	-5.3	0.7	1.5	1.2	3.3	-5.0	0.6	1.4	1.1	3.2
	RCF	-5.4	0.6	1.4	1.1	2.9	-4.7	0.7	1.6	1.3	3.4
	NNC	-5.6	0.7	1.6	1.3	3.3	-5.0	0.6	1.5	1.2	3.1
All ESMs	ANN	-5.5	1.2	2.0	1.4	3.9	-4.8	1.0	1.9	1.3	3.6
	RCF	-5.4	1.1	1.9	1.3	3.2	-4.7	1.2	2.1	1.5	3.9
	NNC	-5.7	1.2	2.0	1.4	3.7	-4.9	1.1	1.9	1.3	3.5
All ESMs	All SDs	-5.5	1.1	2.0	1.4	3.6	-4.8	1.1	1.9	1.4	3.7

identical. Noticeable are the lower change signals deriving from the three MPI-ESM realisations. The results of SD models coincide well, uncertainties of the projected temperatures primarily originate from the different ESMs. The RCP 8.5 scenario differs from the RCP 4.5 mainly in the DF, in the NF only a small difference of $0.3\,°C$ appears. In the DF a deviation of $1.6\,°C$ at Sonnblick and $1.8\,°C$ at Zugspitze suggests a growing impact from RCP 8.5. An accelerated warming of high impact scenarios in the Alpine region is also described by Gobiet et al. (2014). Annually, a strong increase of $3.6\,°C$ at Sonnblick and $3.7\,°C$ at Zugspitze is projected by the RCP 8.5 scenario in the DF. In the European average, an increase between $2.6\,K$ to $4.8\,K$ is expected (Agency (2016); compared is 1986–2005 with 2081-2100 of RCP 8.5).

The expected high temperature increase can have a severe impact on various aspects. Glaciers will retreat and Alpine permafrost melt if the maximum preserving temperature is exceeded. As an result, soil would be destabilised and landslides can occur. The climatological snow line will be found at higher elevations. Following Cebon (1998) a snow line lift of about $100\,m$ to $200\,m$ can be expected considering a temperature increase of $1\,K$. Therefore, in the study area, an upwards shift of approximately $350\,m$ to $700\,m$ is projected in the DF of RCP 8.5. In addition, if precipitation switches from snowfall to rainfall due to the higher temperatures water storage capacity of the region will change. Runoff will be more oriented to a pluvial river regime, with peaks earlier in the year. A reduced water storage capacity can also lead to water scarcity, making an adapted water management necessary.

Seasonal distributions of the simulated daily temperatures are displayed in figures 9.12 and 9.13. At both stations, a future increase in temperature is projected by the SD models. The continuous warming is found throughout the Alpine region (Gobiet et al. 2014). Strongest change signals are found in summer, while the weakest warming is projected in spring. The maximum temperature increase in summer is confirmed by other studies, e.g. Smiatek et al. (2009). In winter, increases are usually only slightly lower than in summer. RCFs tend to produce similar results in winter and autumn. In contrast to the RCP 4.5 scenario, the impact of the RCP 8.5 strengthens in the DF. The spread within each boxplot is rather low. Table 9.3 lists differences in the inter seasonal and annual standard deviation of the respective temperature means. Usually, simulated time series are characterised by a slightly lower annual and seasonal variability, although differences are rather small. Only RCFs driven by historical ESM datasets tend to overestimate the standard deviation of the observed time series. Changes in variability are fairly low, often insignificant and do not show a clear pattern. Seasonally, most significant changes suggest a decrease in variability, in particular in winter and autumn. At Sonnblick in spring and summer, as well as at Zugspitze in summer increases can be found in the ANN products. Change signal magnitudes between the SD models are mostly alike. Largest differences occur in

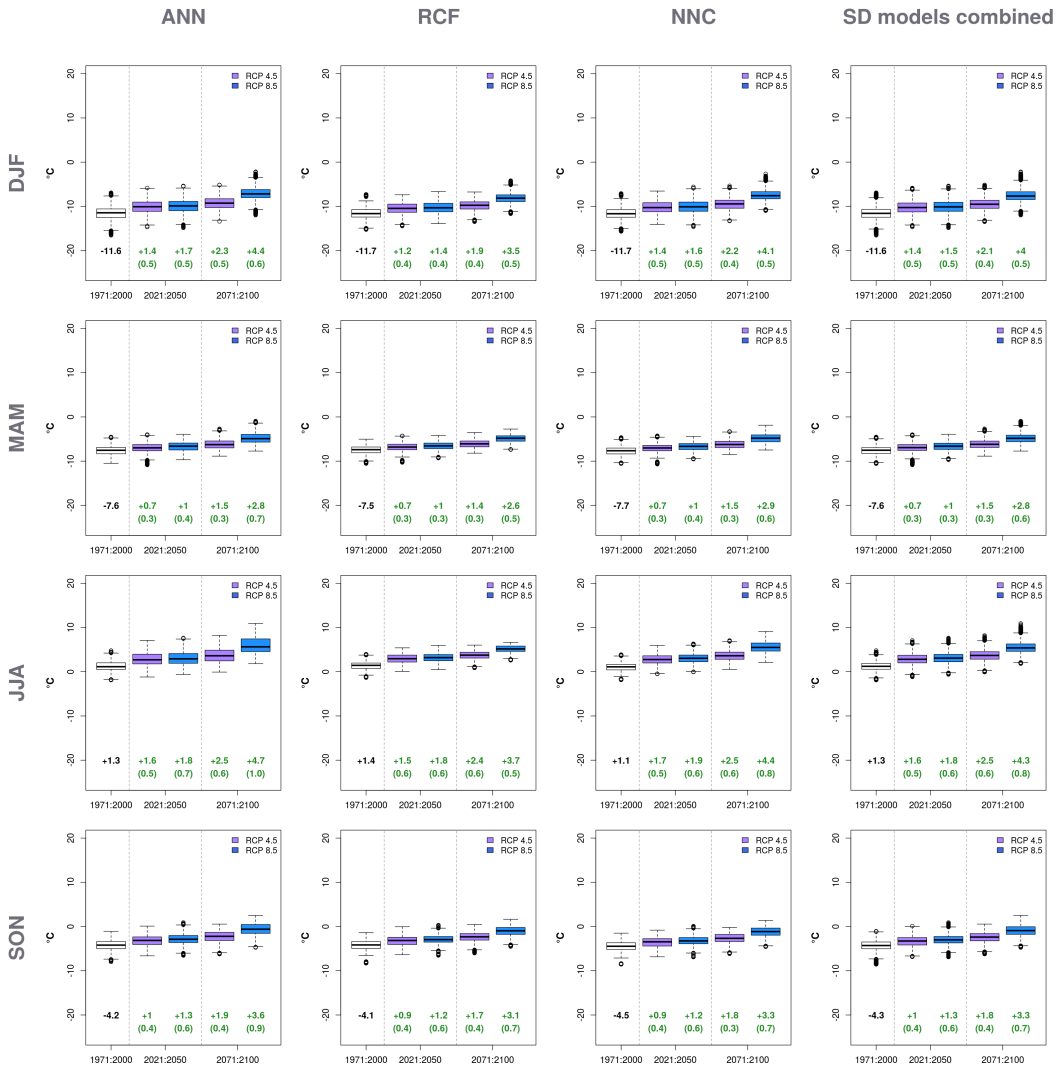


Figure 9.12: Seasonal dependent comparison of climate change signals separated by SD models considering all ESM downscaling products for daily mean temperature at Sonnblick. Differences of the overall mean between the historical reference period and future scenarios are given in absolutes [°C]. In addition, the standard deviation of the diverse change signals originating from different ESMs is listed in parenthesis. Significant increases are coloured in green, decreases in red. Significance is tested using the U-test (level of significance 5 %). Observational references: DJF: −12.3 °C, MAM: −7.8 °C, JJA: 0.9 °C, SON: −4.4 °C.

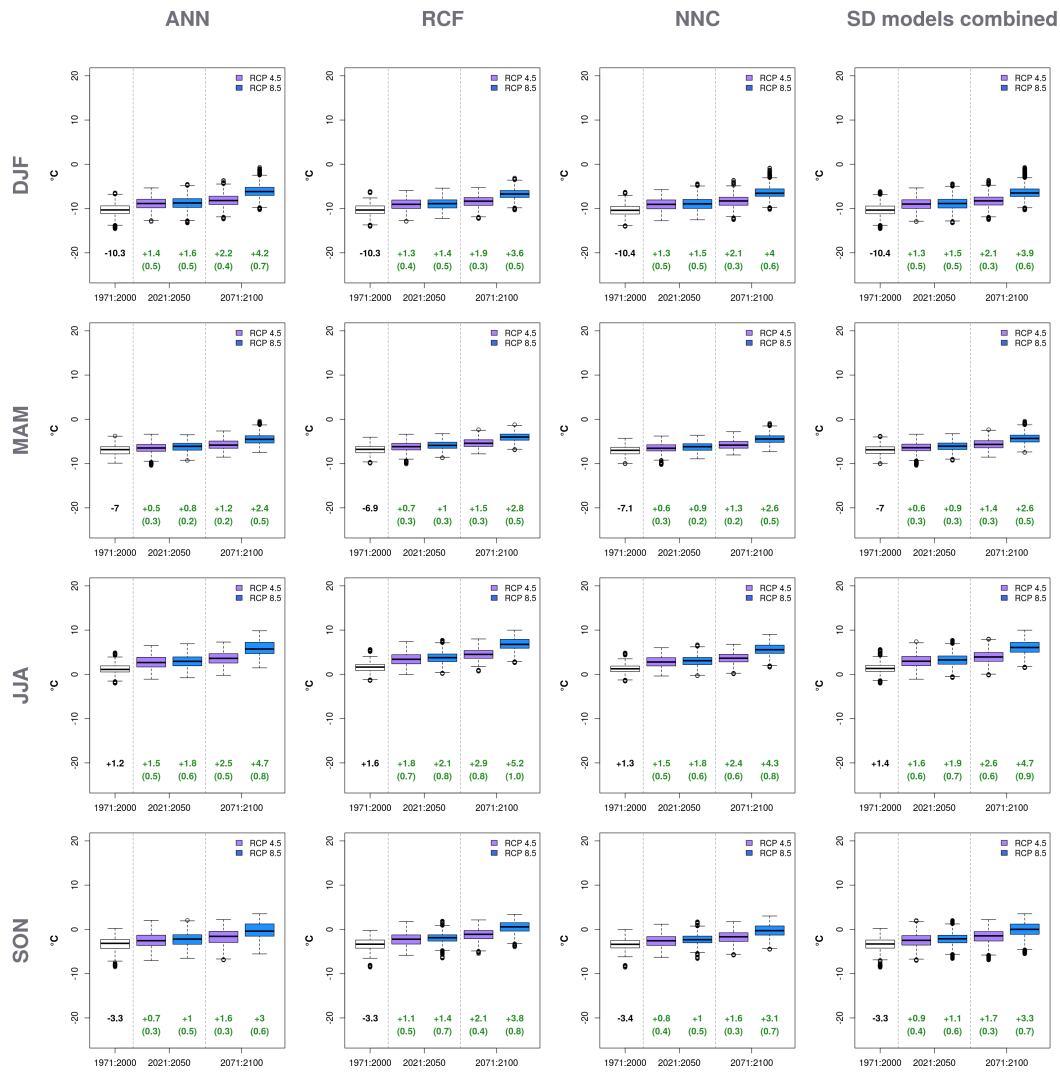


Figure 9.13: Expected seasonal change of daily mean temperatures at Zugspitze. Observational references: DJF: −10.1 °C, MAM: −6.7 °C, JJA: 1.8 °C, SON: −3.2 °C. See figure 9.12 for detailed description.

Table 9.3: Changes in standard deviation of daily temperature at Zugspitze and Sonnblick considering the observational record (Obs) as well as downscaling products based on the reanalysis dataset (Rean, calibration and validation) and ESM experiments (historical: Hist; Scenarios: RCP 8.5 and RCP 4.5) separated by season and annually. Results from cross-validation folds and ESMs are averaged. Number in brackets in Hist: standard deviation between ESMs. Number in brackets in scenarios: standard deviation of the change signal. Reference periods: 1971-2000 (Obs, Rean, Hist); 2021-2050 (NF); 2071-2100 (DF). Changes: Increasing values are marked in green, declining ones in red and insignificant ones in grey (more than two ESMs show no change signal or a different sign.). Absolute change signals are computed by comparing historical and scenario results.

		Sonnblick							**Zugspitze**						
					RCP 4.5		RCP 8.5					RCP 4.5		RCP 8.5	
		Obs	Rean	Hist	NF	DF	NF	DF	Obs	Rean	Hist	NF	DF	NF	DF
DJF	ANN	1.37	1.32	1.38 (0.17)	-0.02 (0.17)	-0.12 (0.21)	-0.02 (0.33)	-0.05 (0.27)	1.46	1.35	1.37 (0.14)	-0.03 (0.12)	-0.12 (0.20)	-0.02 (0.35)	-0.03 (0.23)
	RCF	1.37	1.29	1.32 (0.13)	-0.06 (0.14)	-0.17 (0.19)	-0.09 (0.31)	-0.20 (0.26)	1.46	1.39	1.31 (0.11)	-0.02 (0.11)	-0.13 (0.20)	-0.05 (0.32)	-0.14 (0.23)
	Comb	1.37	1.32	1.41 (0.13)	-0.05 (0.17)	-0.17 (0.20)	-0.04 (0.33)	-0.20 (0.26)	1.46	1.37	1.35 (0.11)	-0.02 (0.13)	-0.12 (0.20)	0.00 (0.34)	-0.08 (0.24)
MAM	ANN	1.08	0.98	0.97 (0.13)	0.01 (0.15)	-0.02 (0.15)	-0.07 (0.23)	-0.01 (0.20)	1.06	0.93	1.04 (0.14)	-0.01 (0.18)	-0.03 (0.16)	-0.08 (0.22)	-0.05 (0.19)
	RCF	1.08	0.96	0.95 (0.13)	-0.03 (0.17)	-0.09 (0.14)	-0.10 (0.23)	-0.16 (0.21)	1.06	1.01	1.07 (0.14)	-0.01 (0.20)	-0.05 (0.17)	-0.09 (0.27)	-0.14 (0.22)
	Comb	1.08	0.99	0.98 (0.14)	0.01 (0.17)	-0.05 (0.17)	-0.07 (0.24)	-0.06 (0.21)	1.06	0.97	1.05 (0.13)	0.00 (0.18)	-0.02 (0.18)	-0.08 (0.24)	-0.07 (0.19)
JJA	ANN	0.91	0.82	0.85 (0.08)	0.10 (0.18)	0.02 (0.15)	0.09 (0.12)	0.10 (0.14)	0.88	0.81	0.88 (0.08)	0.13 (0.23)	0.03 (0.16)	0.07 (0.12)	0.10 (0.13)
	RCF	0.91	0.80	0.85 (0.10)	-0.01 (0.17)	-0.18 (0.15)	-0.05 (0.12)	-0.35 (0.21)	0.88	0.87	1.03 (0.11)	0.15 (0.27)	-0.01 (0.18)	0.08 (0.16)	-0.10 (0.17)
	Comb	0.91	0.83	0.91 (0.08)	0.00 (0.16)	-0.13 (0.15)	-0.03 (0.14)	-0.15 (0.16)	0.88	0.80	0.93 (0.07)	0.08 (0.24)	-0.06 (0.15)	0.00 (0.11)	-0.10 (0.13)
SON	ANN	1.07	0.96	1.06 (0.11)	-0.03 (0.25)	-0.09 (0.21)	-0.07 (0.10)	-0.11 (0.23)	1.13	1.05	1.20 (0.18)	-0.01 (0.28)	-0.09 (0.15)	-0.11 (0.17)	-0.14 (0.23)
	RCF	1.07	1.04	1.09 (0.13)	-0.06 (0.25)	-0.11 (0.20)	-0.12 (0.12)	-0.22 (0.22)	1.13	1.11	1.34 (0.15)	-0.05 (0.28)	-0.18 (0.17)	-0.15 (0.17)	-0.30 (0.22)
	Comb	1.07	0.97	1.07 (0.13)	-0.07 (0.26)	-0.13 (0.20)	-0.11 (0.10)	-0.20 (0.22)	1.13	1.06	1.23 (0.15)	-0.04 (0.27)	-0.13 (0.14)	-0.12 (0.15)	-0.16 (0.21)
Annual	ANN	0.59	0.58	0.57 (0.07)	0.04 (0.11)	0.02 (0.06)	0.07 (0.13)	0.13 (0.08)	0.61	0.57	0.59 (0.07)	0.03 (0.10)	0.01 (0.09)	0.05 (0.15)	0.12 (0.11)
	RCF	0.59	0.58	0.57 (0.06)	0.00 (0.10)	-0.04 (0.05)	0.01 (0.13)	-0.03 (0.09)	0.61	0.61	0.62 (0.07)	0.03 (0.10)	-0.01 (0.10)	0.06 (0.16)	0.04 (0.09)
	Comb	0.59	0.58	0.59 (0.06)	0.00 (0.10)	-0.02 (0.07)	0.04 (0.13)	0.03 (0.08)	0.61	0.59	0.60 (0.07)	0.02 (0.09)	0.00 (0.10)	0.03 (0.14)	0.07 (0.10)

the DF of RCP 8.5, whereby usually RCFs project a considerably larger reduction in variability. Annually, significant changes are rare, unlike to the seasonal findings usually increases in variability are simulated in the DF of RCP 8.5, except for RCFs, which show a reduction at Sonnblick.

Additional temperature indices are evaluated to further investigate the future development of temperature (see appendix tables C.1 and C.2). Shifts in the extremes are evaluated by the average of the yearly maximum (T1max) and minimum (T1min) daily mean temperature. The annual analysis of T1max and T1min change signals is omitted, as the values are equal to winter and summer, respectively. In addition, the frequency of "cold days" (CDF) is examined. Cold days are here defined as days with a mean temperature below $0\,°\mathrm{C}$. Cold days are meant to give hints on how likely snow fall occurs: In the observational record of the reference period at the Zugspitze a rise in snow depth is measured in $89\,\%$ of days, if the daily mean temperature was below $0\,°\mathrm{C}$ and the daily precipitation sum above $100\,\mathrm{mm\,d^{-1}}$.

Reanalysis and historical ESM based simulations of T1min and T1max fit the observational dataset fairly well. SD models match by a very high degree in the change signal direction and magnitudes of T1min. As change signals are positive and therefore within the known reference value range, RCFs align with results of ANNs and NNCs. In contrast, as T1max changes are positive as well, the reference value range of the RCFs is exceeded. Therefore, RCFs show lower T1max change signals, in particular in the DF and in RCP 8.5, where larger increases can be expected. The known issue of the RCF functionality makes the accuracy of the T1max change signal in this application questionable, especially the magnitudes must be treated with caution. In the following, RCF changes of T1max will not be examined. At both stations in the NF of RCP 4.5, no significant change of T1min is found in spring and autumn. Otherwise, an increase in both T1min and T1max is usually found. This indicates a general shift in the value distribution of temperature, which is also suggested for the overall Alpine region by Gobiet et al. (2014). In the investigated time periods of both the scenarios, T1min stays negative at Zugspitze and Sonnblick, even in summer. In the SD model mean, the strongest T1min increase is found to be in DF in winter, while the change signal of the other seasons is characterised by similar magnitudes. Except for summer, the variation between seasons of NF change signals of T1min are low as well. T1max at Zugspitze is projected to switch from negative to positive in all analysed future scenarios. At Sonnblick in winter, which has an in general lower T1max historical base value, T1max is still found to be negative in the NF, but will be increasingly positive in the DF, especially in the RCP 8.5 scenario. Strongest T1max increases are expected in winter and summer. Comparing the future T1min and T1max development, except in the DF of RCP 8.5 (and RCFs), the warming in T1max is usually projected to be stronger than of T1min at both stations.

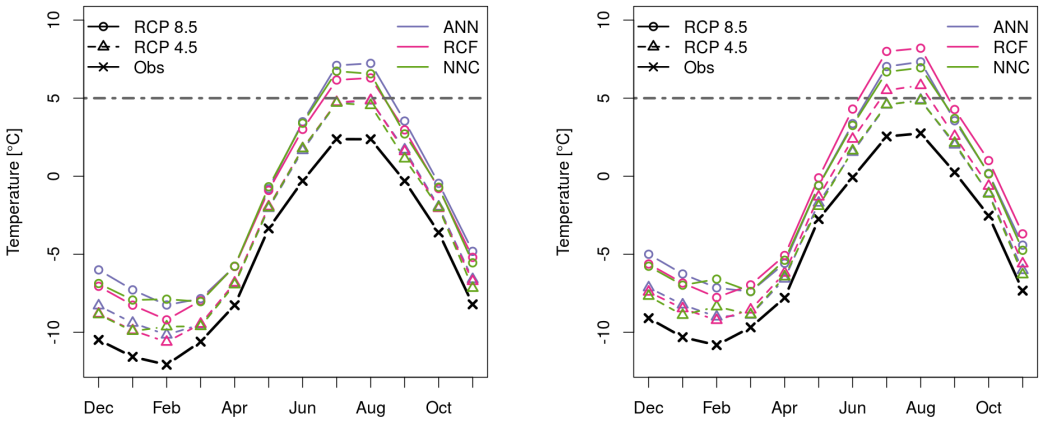
Table 9.4: Annual changes of the growing season length considering the observational record (Obs) as well as downscaling products based on the reanalysis dataset (Rean, calibration and validation) and ESM experiments (historical: Hist; Scenarios: RCP 8.5 and RCP 4.5). Results from cross-validation folds and ESMs are averaged. Number in brackets in Hist: standard deviation between ESMs. Number in brackets in scenarios: standard deviation of the change signal. Reference periods: 1971-2000 (Obs, Rean, Hist); 2021-2050 (NF); 2071-2100 (DF). Changes: Increasing values are marked in green, declining ones in red and insignificant ones in grey (more than two ESMs show no change signal or a different sign.). Relative change signals [%] are computed by comparing historical and scenario results.

	Sonnblick							Zugspitze						
				RCP 4.5		RCP 8.5					RCP 4.5		RCP 8.5	
	Obs	Rean	Hist	NF	DF	NF	DF	Obs	Rean	Hist	NF	DF	NF	DF
ANN	5.1	4.1	3.8	4.8	7.6	5.6	15.6	6.7	5.9	5.0	4.9	8.0	5.7	15.7
			(0.9)	(2.4)	(3.4)	(2.8)	(4.9)			(1.0)	(2.3)	(3.3)	(2.7)	(4.7)
RCF	5.1	3.9	4.1	4.7	7.6	5.6	13.6	6.7	6.6	6.9	5.8	9.6	6.8	17.6
			(0.2)	(2.3)	(3.0)	(2.7)	(3.2)			(0.4)	(2.6)	(3.5)	(2.9)	(4.7)
Comb	5.1	4.3	3.6	4.2	6.8	4.9	14.0	6.7	6.1	5.8	4.5	7.5	5.3	14.8
			(0.4)	(2.0)	(2.7)	(2.1)	(4.0)			(0.6)	(2.1)	(2.9)	(2.3)	(4.2)

Observations and simulations driven by reanalysis and historical ESM datasets accord well in CDF. Bigger differences are only found at Zugspitze in summer, where the historical reference value is by approximately $4\,\%$ higher than in the observed time series. SD models show good agreement in the CDF change signals. Similar to previous results, an overall decrease of the frequency is expected in the future. Strong decreases are found especially during summer. In the DF of the severe scenario RCP 8.5, the CDF is reduced from approximately $37\,\%$ to $9\,\%$ at Zugspitze and from $35\,\%$ to $7\,\%$ at Sonnblick. The change in autumn is almost as strong as in summer, although the reference value is remarkably higher than in summer. In winter and spring, the occurrence of cold days is still high. In winter, the CDF values do not fall below approximately $89\,\%$ at Zugspitze and $95\,\%$ at Sonnblick, even in the DF of RCP 8.5. At both stations the CDF stays usually above $80\,\%$ in spring.

Further analyses investigate changes in the growing season (GS) at Zugspitze and Sonnblick. Following EPA (2014), the GS is defined as the number of days with a mean temperature above $5\,°\mathrm{C}$, a temperature, where plant growth is possible. The $5\,°\mathrm{C}$ threshold is meant to satisfy the majority of native plants in Europe. For certain species, the temperature level needs to be adjusted, as some plants need higher temperatures or can already grow with lower ones (Menzel et al. 2003). The GS here focuses on temperature, while other important on site conditions like frost days, soil, precipitation or daylight hours are disregarded.

The GS in the observational reference period reaches average values of 18.6 days/year at Sonnblick and 24.5 days/year at Zugspitze (see table 9.4). Future change signals in frequency of days above $5\,°\mathrm{C}$ in future scenarios is statistically evaluated by the χ^2-test, whereby on a level of significance of $5\,\%$ all change signals are significant. In



(a) Sonnblick **(b)** Zugspitze

Figure 9.14: Monthly mean temperatures of the observational period (Obs, 1971-2000) compared to the distant future (2071-2100) by the RCP 4.5 and RCP 8.5 scenarios applied to the SD models. Months above the horizontal grey line at 5 °C are contributing to the growing season.

the following the relative change signal of the GS derived from the ESM scenarios is transferred onto the observational record. This step is necessary as some ESMs do not use the Gregorian calendar leading to different base number of days per year. The relative change signals of the SD models are averaged in the following computations. The large increases of RCP 8.5 in the DF results into a GS of 71.2 days/year at Sonnblick and 83.0 days/year at Zugspitze, which almost quadruples the current GS lengths. In the DF of RCP 4.5, the gain in GS is more moderate: 45.4 days/year at Sonnblick and 55.0 days/year at Zugspitze. In the NF the difference between scenarios is small: Sonnblick shows a GS of 35.3 days/year in RCP 4.5 and 38.2 days/year in RCP 8.5, while the number of GS days at Zugspitze sum up to 42.9 in RCP 4.5 and 46.1 in RCP 8.5.

There are different definitions of the GS, which also account for other variables than temperature. Schultz (2016), for example, characterises the GS by the number of humid months with a monthly average temperature above $5\,°C$. Zugspitze and Sonnblick can be considered humid throughout the year in the historical period as well as in future scenarios. Figure 9.14 shows the monthly average temperatures for the observational time series in the reference period and the development in the DF of scenarios RCP 4.5 and RCP 8.5. The $5\,°C$ threshold is marked by the dotted grey line. At the summit stations, no month of the observational period can be classified as contributor to the GS, which also accounts for the NF (not shown in the figures). In the DF of RCP 4.5 solely RCFs exceed the $5\,°C$ threshold at Zugspitze in July and August. However, in the RCP 8.5 scenario, where higher temperatures

are projected, July and August are consistently specified as GS by all downscaling approaches. Schultz (2016) postulates that tree growth requires at least three months with an average temperature above $5\,°C$ (or one month above $10\,°C$, which is not reached by any examined subset). Following Oliver (2014), the current timberline in the Wetterstein mountains, the mountain group where Zugspitze is located, lies at approximately $1600\,m$. In particular in the DF of RCP 8.5 at Zugspitze and considering only the temperature limitation, the elevation of the timberline could rise slightly below the summit.

9.5 Summary: Temperature

Daily mean temperature series at both stations are characterised by a pronounced annual cycle. In the consequence of the cold high mountain climate, monthly averaged temperatures mostly lie below the freezing point. Monthly temperatures above zero are found in July and August as well as September at Zugspitze. In long time slices of the 20th century, positive trends reveal an increase in temperature. In short and medium-term periods cooling can be found as well, in particular from 1940 to 1970.

Analysing the temporal dependency of the SD model predictor-predictand relationship, the gain in performance near the end of the 20th century is still observed, but not as highly pronounced as in terms of precipitation. In any case, the choice to restrict the available training samples onto the reference period at the end of the 20th century is approved again. The observed quantiles and the annual cycle are well reproduced by the SD models, which is already indicated by the high SD model performances. Differences comparing observed and reanalysis or historical ESM-based downscaling products where usually very small, suggesting a good quality and transferability of the SD models.

Downscaled ESM scenarios project a strong increase in temperature annually as well as seasonally at both stations. Strongest temperature increases are observed in summer, weakest in spring. The projected difference between RCP 8.5 and RCP 4.5 is intensified at the end of the 21st century. In the investigated scenarios and time periods, T1min increases but stays negative at Zugspitze and Sonnblick. In winter, the only negative season of T1max in the reference period, is projected to switch from negative to positive at Zugspitze. In the NF of Sonnblick in winter, T1max is still found to be negative, but will be increasingly positive in the DF. Apart from RCFs and the DF of RCP 8.5, the absolute increase values of T1max are higher than of T1min. CDF is subjected to decrease heavily in summer and autumn, and considerably less severe in winter and spring, as the latter on average still lies substantially below the freezing point. The GS is featured by a sharp increase in the DF of RCP

8.5, where relative changes are expected to be twice as high as in RCP 4.5. From the perspective of temperature, the vegetational life zones will relocate considerably towards the summits. In the DF of RCP 8.5, on average, two months surpass the $5°$ threshold. Nevertheless, at the end of the 21st century, the potential timberline at Zugspitze could already be slightly below the mountain top in this scenario.

The expected future shift in the temperature distribution poses problems for the RCF approach. An oftentimes sharp increase in the frequency of the warm temperature classes, indicates the necessity of more types focusing on the high temperature reference values. This limitation of RCFs is especially reflected in the extreme value analysis of hot days. T1max values are often conspicuously lower in RCFs than simulated by ANNs or NNCs, what can result from this systematical problem. However, the direction of change should be properly exhibited in RCF products.

10 Relative Humidity

10.1 An Overview over Origin and Distribution of Atmospheric Moisture in the Alpine Region

Atmospheric humidity at a specific point within the Alpine region is closely related to the atmospheric moisture flux. On the Luv side of a Foehn event, for example, relative humidity is increased as air masses are orographically forced to ascend and decreased on the Lee-side (Panziera and Germann 2010). A detailed description of the Foehn wind and its effects is given in section 8.1. Locally, relative humidity further depends on precipitation amounts and the soil moisture-atmosphere feedback (Gobiet et al. 2014). High precipitation saturates soil with water, which in turn intensifies evaporation, resulting in higher moisture content of the near-surface atmosphere. Therefore, it can be expected that regions characterised by high precipitation sums show high relative humidity magnitudes as well. Beniston and Jungo (2002) further point out the relationship of humidity and the NAO index, especially in highly elevated areas. There is a general reduction of moisture in the atmosphere when the NAO index is strongly positive. This further suggests that large amounts of moisture are advected from the surrounding seas, for example the Atlantic ocean. M. Müller and Kaspar (2011) find different seasonal dependencies between air moisture content and large-scale advection in the southern Alps. A similar behaviour cannot be ruled out in the study region of this thesis, which confirms the decision of building seasonal models.

10.2 Analysis of the Observational Datasets

In this study, relative humidity is only available for Zugspitze. The long term observational mean of relative humidity in the reference period 1971-2000 at Zugspitze is $80.9\,\%$. This quite high value indicates that near surface air masses at Zugspitze are usually close to the saturation point. The minimum annual average of $77.0\,\%$ is observed in 1989, which coincides with the warmest year in the reference period (see section 9.2). Highest relative humidity is measured in 1974 with an annual average of $86.5\,\%$. The annual variability of $2.4\,\%$ can be considered as small. Relative hu-

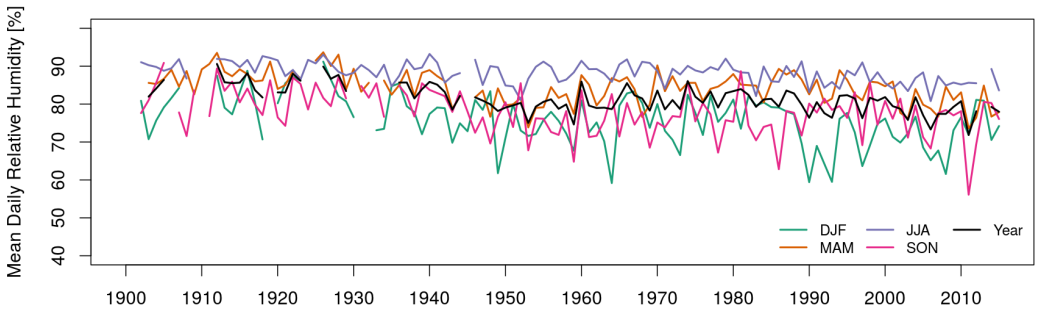


Figure 10.1: Annual and seasonal time series of the averaged local relative humidity measurements computed from the total observational record.

midity is characterised by a pronounced annual cycle (see Obs in figure 10.5). In the reference period, June shows a sharp peak with a monthly average of $90.2\,\%$. From June a rapid decrease follows both flanks of the curve leading to similar low magnitudes from October to February, ranging between the minimum of $73.3\,\%$ in January and $75.0\,\%$ in February.

In the long term, annual means of relative humidity are decreasing throughout the 20th century (see figures 10.1 and 10.2). In short and medium periods, a temporal increase is observed around 1965, which is most pronounced in spring. Strong reductions are primarily found short to medium term in the first part of the 20th century near 1940 and in addition in winter and spring in the late 20th century. Investigations of Beniston (2005) on moisture in winter confirm decreasing tendencies in the entire Alpine region and link the reductions to the mostly positive NAO pattern in the late 20th century. In summer, only comparable weak trends are found. Autumn shows no trend in the medium to short term periods in the second half of the 20th century. Comparing the previous results of precipitation and temperature with relative humidity in spring, a cooling comes alongside with increasing precipitation amounts and relative humidity, which can be expected from a physical point-of-view. In general, trends of temperature and relative humidity match well in summer, autumn and in the second half of the 20th century in spring. Warming goes hand in hand with decreasing relative humidity trends and vice versa. In contrast, beside parts of the century in spring, the pattern of precipitation and relative humidity usually show no conform developments.

10.3 Statistical Model Results with Focus on the 20th Century

The experimental setup to analyse time-dependent model skill between predictors and relative humidity is identical to precipitation in section 8.3. Similar to precipi-

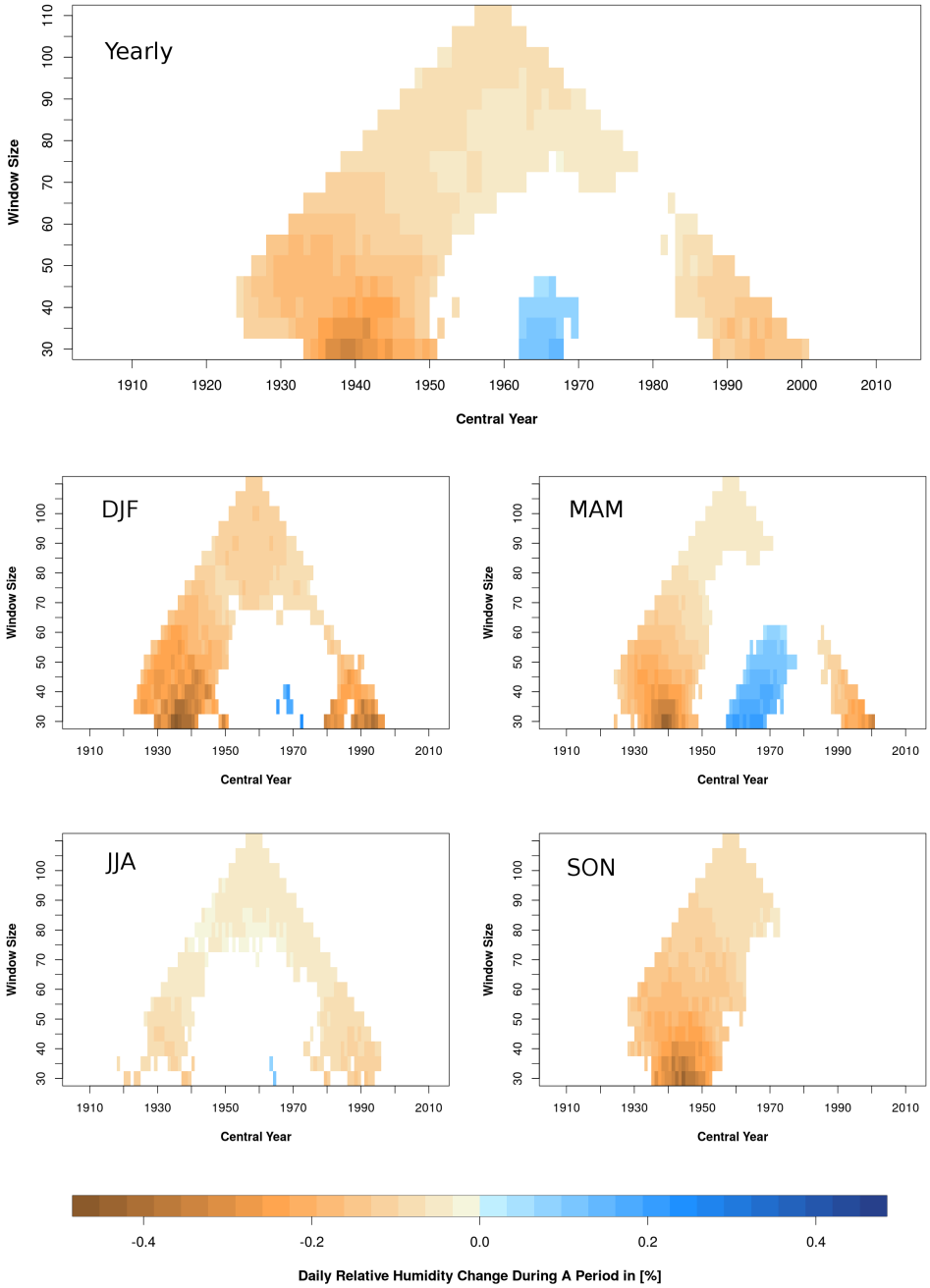


Figure 10.2: Trends in the observational time series of the seasonally and annually aggregated Zugspitze relative humidity record. The x-axes represents the central year of the evaluated time period, the y-axes lists corresponding time window sizes in years. Trends are displayed if tested significantly by the trend-noise ratio (level of significance 5 %).

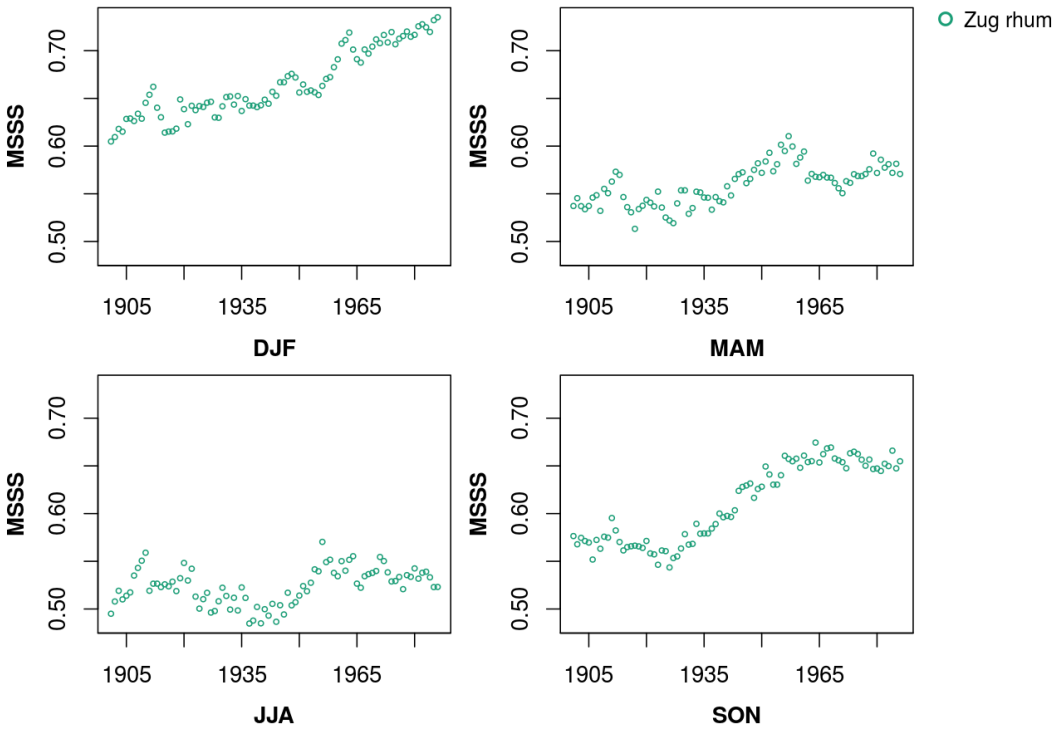


Figure 10.3: Seasonally dependent analysis of the temporal altering model skill of the final ANN configuration evaluated by the MSSS. Models are calibrated using observed relative humidity (rhum) at Zugspitze (Zug) and the 20th century reanalysis dataset. Training is repeated from 15 different ANN initialisations, the resulting skills scores are averaged. Models are calibrated on a 30-year period, starting at each mark of the graphic.

tation and temperature, the SD models show higher quality at the end of the 20th-century (see figure 10.3). The gain in MSSS is more pronounced in winter and autumn. In autumn a similar sharp increase in MSSS is found almost simultaneously to precipitation. The MSSS of the reference period is amongst the highest in winter, summer and autumn. In spring, although still on a high level, earlier time windows provide better model performance. As in the final turn, seasonal model results are merged into one time series, a compromise needed to be found. For the sake of comparability, the choice fell on the congruent period to precipitation and temperature.

QQ-plots of figure 10.4 compare the distribution of observations and SD models by the quantiles of the reference period. The value range of relative humidity from $0\,\%$ to $100\,\%$ with two hard limits as well as the unusual value distribution is a complex task to learn for SD models. The quantiles of calibration and validation series usually lie close together. Throughout seasons and SD-methodologies, the higher quantiles, approximately $80\,\%$ and above are underrepresented in the SD models. In the central value range between $20\,\%$ to $80\,\%$ typically an overestimation of the observed quantiles is found. The lower ($20\,\%$ and below) observational and simulated quantiles converge again, although often times SD models tend to overshoot and underestimate the lower quantiles. Solely, RCFs in spring and summer stay in the overestimation mode in the lower quantiles. NNCs in winter show similar problems like in the precipitation approach. The rapid increase in quantile values at $80\,\%$ emerges from the overlapping value ranges of the successor ANNs. The transfer of the calibrated models onto the historical ESM datasets shows little discrepancies by RCFs. The only larger differences are found in the lower quantiles of summer, where ACCESS and MPI-ESM (r1) show considerably lower quantiles than the other ESMs in calibration as well as validation. In winter and summer, ESMs, downscaled by ANNs and NNCs, are usually characterised by a strong overestimation of the central value range between $20\,\%$ to $80\,\%$. This effect is also clearly visible in the monthly averages displayed in figure 10.5. While RCFs fit the monthly means of the annual cycle well, NNCs and especially ANNs tend to overestimate October to March. Closest to the observed averages in the critical months from October to February in terms of ANNs are ACCESS and HadGEM. Both ESMs show good results with NNCs from October to December as well. IPSL downscaled by ANNs and NNCs differs by far the most from the observed means from October to February. Applied to ANNs, IPSL does not follow the shape of the annual cycle very well. In terms of ANNs, ESMs show also difficulties at the kink between November and December. In contrast, RCFs represent the annual course quite well. Also, NNCs come close to the observed values and follow generally the yearly curve, only June is relatively underestimated.

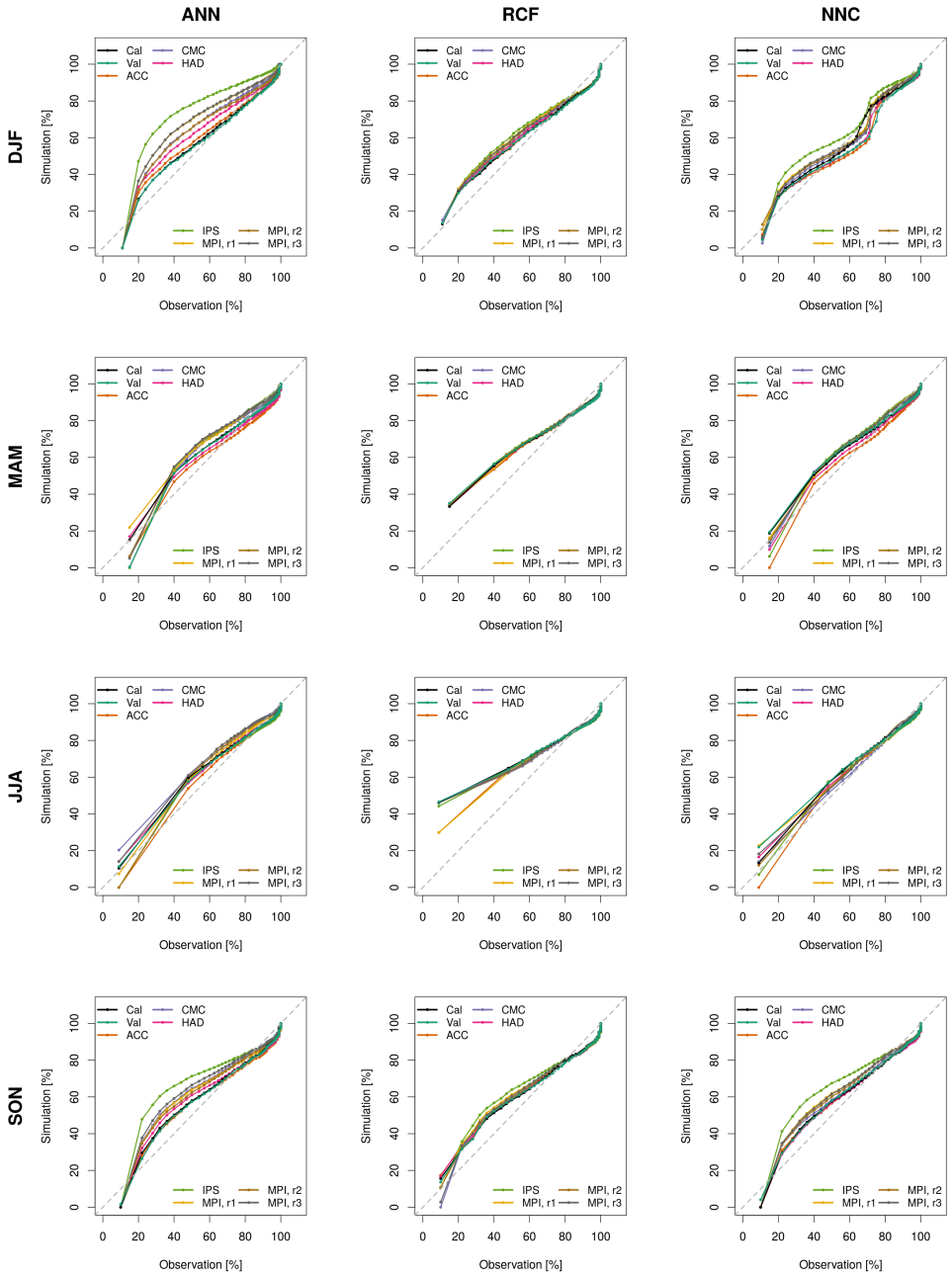


Figure 10.4: QQ-plots for Zugspitze relative humidity considering the seasonal models in the reference period 1971-2000. 50 quantiles of the observed dataset are compared with their simulated counterparts from reanalysis and historical ESM data-based downscaling products derived by the final SD models. Reanalysis series are separated by calibration (Cal) and validation (Val) periods. The fifteen Cross-validation runs are treated as one sample.

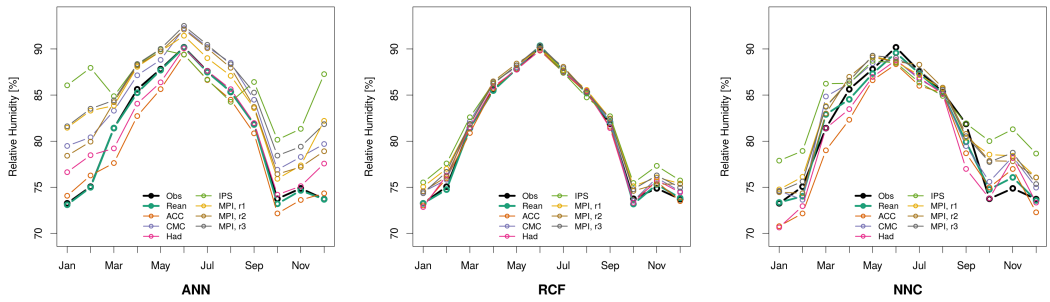


Figure 10.5: Observed (Obs) and simulated (complete reanalysis (Rean) and climate model driven results) monthly averaged daily mean relative humidity considering the reference period 1971-2000 for each SD model type.

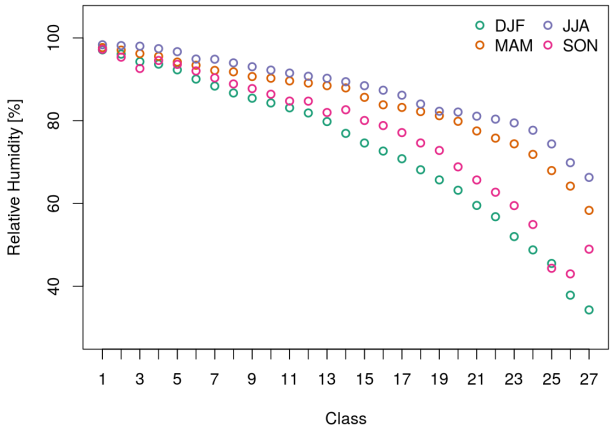


Figure 10.6: Circulation type specific mean relative humidity reference values in [%] for each of the 27 classes of the seasonal RCFs at Zugspitze derived from the calibration period 1971-2000.

10.4 Climate Change Signals: 20th vs. 21st century

Introductory, an overview over RCF characteristics in the climate change study is given. The relative humidity means of each class of the RCFs in this study are displayed in figure 10.6. Similar magnitudes and development of the curves are found in spring and summer, as well as autumn and winter, whereby the latter are considerably lower. The first decrease is rather linear while towards the end a progressively rapid drop occurs. Consequently, the first half of the value range is handled by approximately less than 2/3 of the types, leading to a lower resolution of the smaller relative humidity magnitudes. Changing frequencies of weather types are shown in figure 10.7. There is barely any difference in RCP 4.5 between the NF and DF. In contrast, in RCP 8.5, especially the higher changes in frequencies, tend to further intensify in the DF. Summer months and the RCP 8.5 in autumn are characterised by a

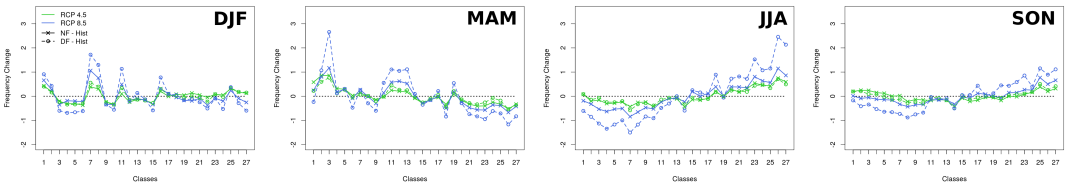


Figure 10.7: Average future changes of type frequency in [%] comparing the historical reference (1971-2000) of all ESMs with both scenarios (RCP 4.5 and RCP 8.5) in the near (NF) and distant future (DF) of the seasonal RCF relative humidity models at Zugspitze.

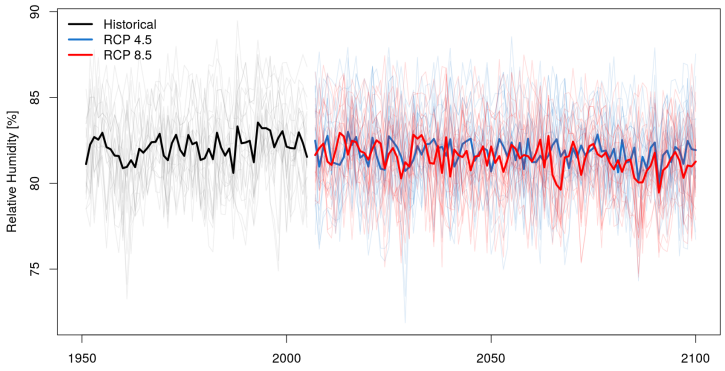


Figure 10.8: Annually averaged daily relative humidity means resulting from the ESM downscaling product of the total available time period 1950 to 2100 (historical and scenario dataset). Thick lines represent the overall average of ESM results. Thin lines in the background display the mean of the cross-validation procedure considering a single SD model and ESM combination.

decrease in the first half of the reference value range and an increasing second half. In winter and spring, frequency changes are less organised, but still affect the upper and lower tail of the reference values. In terms of relative humidity with a fixed upper limit an increase of the $100\,\%$ type is unproblematic as there is a reference value included. In contrast, the lower limit of $0\,\%$ relative humidity is not present in the calibration datasets value range (lowest reference values are found in winter slightly below $40\,\%$). Changes here can lead to inaccurate magnitudes.

In the following the climate change analysis results of ANNs, RCFs and NNCs are discussed. The overall development of annual averages of relative humidity resulting from all SD models driven by ESM data throughout the historical period and scenarios are displayed in figure 10.8, while trends are further listed numerically in table 10.1. All significant change signals agree in a decreasing tendency of relative humidity compared to the historical time series. In general the differences between historical runs and scenarios are projected to be rather small. Similar decreases are usually found in the DF of RCP 4.5 and the NF of RCP 8.5. In RCP 4.5, the more

Table 10.1: Historical (Hist: 1971 to 2000) annually averaged relative humidity [%] at Zugspitze as well as projected absolute changes to the historical period [%] considering different ESMs, SD models and scenarios. Changes are computed for the near future (NF: 2021 to 2050) and distant future (DF: 2071 to 2100). Positive numbers are marked in green, negative ones in red, while insignificant values (U-test, significance level 5 %) are labelled in grey. (Observational average: 80.9 %.)

			RCP 4.5		RCP 8.5	
		Hist	NF	DF	NF	DF
ACC, r1	ANN	79.8	-0.6	-0.2	-0.3	-1.5
	RCF	80.9	-0.1	0.4	0.2	-0.2
	NNC	79.5	-0.3	0.2	0.0	-0.8
CMC, r1	ANN	84.1	-1.4	-1.2	-1.3	-2.1
	RCF	81.4	-0.6	-0.5	-0.6	-0.9
	NNC	81.5	-1.2	-0.9	-1.1	-1.5
HAD, r1	ANN	81.4	-0.6	-0.6	-0.6	-0.7
	RCF	81.0	0.1	0.2	0.2	0.3
	NNC	79.9	-0.2	-0.1	-0.1	0.1
IPS, r1	ANN	86.0	-0.7	-1.4	-0.9	-2.5
	RCF	81.9	-0.2	-0.8	-0.3	-1.3
	NNC	83.3	-0.4	-1.1	-0.5	-2.2
MPI, r1	ANN	84.4	-0.8	-1.5	-1.3	-2.0
	RCF	81.7	0.0	-0.8	-0.7	-1.0
	NNC	82.1	-0.4	-1.2	-1.1	-1.5
MPI, r2	ANN	83.9	-0.4	-0.3	-0.5	-0.6
	RCF	81.5	-0.1	0.2	0.2	0.0
	NNC	81.9	-0.2	0.0	-0.3	-0.1
MPI, r3	ANN	85.4	-0.5	-0.8	-1.1	-1.9
	RCF	81.7	0.1	-0.2	-0.5	-1.2
	NNC	82.1	-0.2	-0.5	-0.8	-1.7
All CMs	ANN	83.6	-0.7	-0.9	-0.9	-1.6
	RCF	81.4	-0.1	-0.2	-0.2	-0.6
	NNC	81.5	-0.4	-0.5	-0.6	-1.1
All CMs	All SDs	82.2	-0.4	-0.5	-0.5	-1.1

severe reduction takes place in the NF, the DF shows only slightly lower values. In contrast, in RCP 8.5, the difference between historical and NF is almost identical to the gap between NF and DF. Individually, HadGEM, MPI-ESM (r2) and ACCESS usually show stable conditions. Significant decreases are mostly found in the DF of both scenarios as well as in the NF of RCP 8.5. In the multi model mean of every ESM and SD model combination, decreasing trends are to be found throughout the analysed future scenarios and time periods.

Seasonally, there are increasing tendencies simulated as well (see figure 10.9). NNCs and RCFs suggest an initial decrease followed by an increase of relative humidity in the DF of winter. This characteristic in the Alpine region is also described by Gobiet et al. (2014). In spring, solely RCFs suggest a comparable strong overall future rise in relative humidity, which in the further determines the SD-method overall in this season. In summer and autumn ANNs, RCFs and NNCs agree in a reduction throughout the evaluated future periods. A consistently falling trend in the summer months of the 21st century is also suggested by Gobiet et al. (2014). The negative trend in summer follows the future relative humidity development of southern Europe (Ruosteenoja and Räisänen 2013). The theory behind the winter and summer developments is based on the interaction between evaporation and the atmosphere. The expected strong increase of precipitation in the DF in winter makes more water available to evaporate, overcompensating the higher water vapour capacity of warmer air. In summer, at Zugspitze, where only small positive precipitation trends but high temperature rises are simulated, the available soil-water from precipitation cannot satisfy the potential demand, leading to an overall decline in relative humidity. Similar to the annual findings, especially in summer and autumn, a much stronger drop in relative humidity is found between the NF and DF in RCP 8.5, than RCP 4.5.

Interannual and seasonal variabilities of relative humidity is very small in comparison to the annual and seasonal means (see table 10.2). The largest standard deviations of the reference period are observed in winter. Variability is underestimated in the downscaled time series. Change signals in the evaluated future periods are rather small and often insignificant. Largest changes are found in winter in the NF of RCP 4.5, where an increase of $0.7\,\%$ is detected by ANNs. In spring, summer, autumn and annually, the maximum absolute changes rarely reach $0.3\,\%$. Clearest tendencies of the direction of change are found in summer and winter, where if significant, all SD-methodologies suggest an increase in variability throughout the considered future periods. The only significant annual change is found in the DF of RCP 4.5 where an concordantly decrease is projected by SD models.

Further evaluations include the mean of the annual minimums (RH1min) and the frequency of events exceeding a relative humidity of $95\,\%$ (RHf95). Change signals

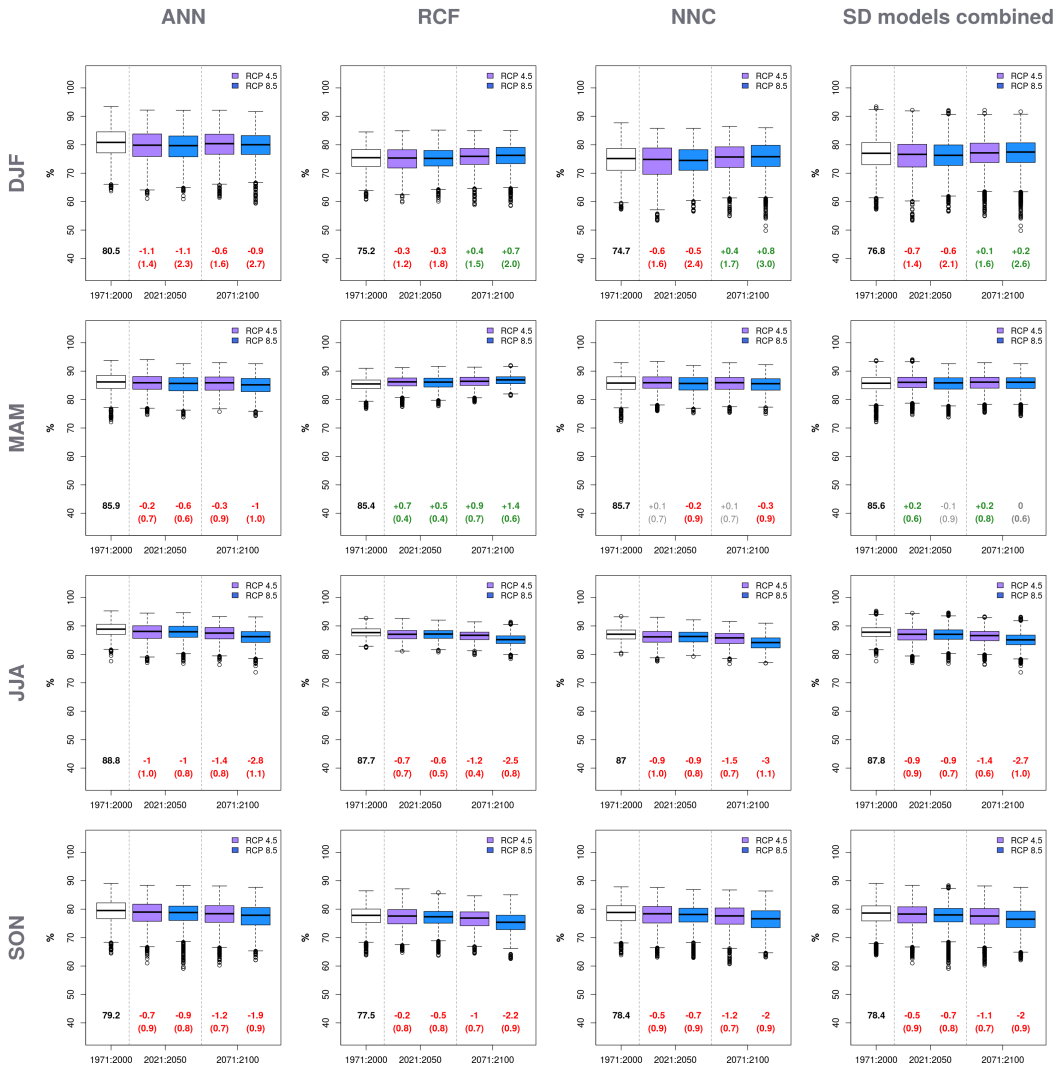


Figure 10.9: Seasonal dependent comparison of climate change signals separated by SD models considering all ESM downscaling products for daily relative humidity means at Zugspitze. Differences of the overall mean between the historical reference period and future scenarios are given in absolutes [%]. In addition, the standard deviation of the diverse change signals originating from different ESMs is listed in parenthesis. Significant increases are coloured in green, decreases in red. Significance is tested using the U-test (level of significance 5 %). Observational references: DJF: 73.9 %, MAM: 85.0 %, JJA: 87.9 %, SON: 76.8 %.

Table 10.2: Changes in standard deviation of daily relative humidity at Zugspitze considering the observational record (Obs) as well as downscaling products based on the reanalysis dataset (Rean, calibration and validation) and ESM experiments (historical: Hist; Scenarios: RCP 8.5 and RCP 4.5) separated by season and annually. Results from cross-validation folds and ESMs are averaged. Number in brackets in Hist: standard deviation between ESMs. Number in brackets in scenarios: standard deviation of the change signal. Reference periods: 1971-2000 (Obs, Rean, Hist); 2021-2050 (NF); 2071-2100 (DF). Changes: Increasing values are marked in green, declining ones in red and insignificant ones in grey (more than two ESMs show no change signal or a different sign.). Absolute change signals are computed by comparing historical and scenario results.

					RCP 4.5		RCP 8.5	
		Obs	Rean	Hist	NF	DF	NF	DF
DJF	ANN	6.6	5.2	4.0 (0.8)	0.7 (0.9)	0.5 (0.8)	0.4 (0.6)	0.5 (1.0)
	RCF	6.6	4.6	4.1 (0.7)	0.3 (0.8)	0.0 (0.8)	0.0 (0.6)	-0.2 (0.8)
	NNCs	6.6	5.7	5.1 (0.9)	0.6 (1.1)	0.4 (0.8)	0.2 (0.8)	0.2 (1.2)
MAM	ANN	2.6	1.7	2.6 (0.4)	0.2 (0.6)	0.2 (0.7)	0.3 (0.3)	0.1 (0.4)
	RCF	2.6	1.5	2.2 (0.3)	0.0 (0.4)	-0.1 (0.5)	0.0 (0.2)	-0.4 (0.3)
	NNCs	2.6	1.8	2.7 (0.4)	0.1 (0.6)	0.1 (0.8)	0.2 (0.4)	-0.1 (0.4)
JJA	ANN	2.5	1.8	1.9 (0.3)	0.3 (0.6)	0.2 (0.4)	0.2 (0.2)	0.2 (0.4)
	RCF	2.5	1.4	1.8 (0.2)	0.2 (0.4)	0.1 (0.3)	0.0 (0.2)	0.1 (0.2)
	NNCs	2.5	1.8	2.2 (0.4)	0.2 (0.6)	0.1 (0.4)	0.1 (0.2)	0.1 (0.3)
SON	ANN	5.5	5.0	3.4 (0.7)	0.1 (1.0)	-0.1 (1.1)	-0.1 (1.2)	0.0 (0.6)
	RCF	5.5	4.5	3.5 (0.8)	0.2 (1.0)	-0.1 (1.0)	-0.2 (1.2)	0.1 (0.5)
	NNCs	5.5	4.9	3.7 (0.7)	0.2 (1.1)	0.0 (1.2)	-0.2 (1.3)	0.1 (0.5)
Annual	ANN	2.4	1.8	1.7 (0.2)	0.0 (0.2)	-0.1 (0.3)	0.0 (0.3)	0.1 (0.3)
	RCF	2.4	1.7	1.6 (0.2)	-0.1 (0.3)	-0.2 (0.2)	-0.1 (0.2)	0.0 (0.2)
	NNCs	2.4	2.0	1.9 (0.2)	0.0 (0.3)	-0.1 (0.3)	0.0 (0.3)	0.0 (0.3)

are listed in the appendix in table D.1. The focus of RH1min lies on the development of extremes, while RHf95 is simply an index for days of high moisture, as RHf95 is quite frequently exceeded. Already indicated in figure 10.4, RH1min is overestimated by approximately $10\,\%$ in most cases in the ESM and reanalysis-driven experiments, whereby the latter usually fit the observational characteristics slightly better. In the future, if significant, usually a decreasing trend is identified throughout the seasons and annually. The high number of insignificant values and comparable low changes of RCFs point out to the repeatedly discussed problem of RCFs and their strict reference value range. Thus, RH1min development found by RCFs should be treated with caution. Drops in RH1min simulated by ANNs and NNCs are rather severe. The largest decline is usually simulated between the historical and NF subset of both scenarios. Comparing the NF and DF, the differences are small, often times again a rise of relative humidity is detected (e.g. RCP 8.5 in spring, autumn and annually). RHf95 frequencies are partly heavily underestimated by the simulations in the reference period. In autumn, for example, a difference of nearly $20\,\%$ is simulated by RCFs. RHf95 is strongly underestimated by the models in comparison to the characteristics of the observational time series. A threshold of $95\,\%$ relative humidity is surpassed least frequently in $27.5\,\%$ of cases in winter and most frequently by $47.2\,\%$ of days in summer. The seasonal behaviour is split into two parts. In winter and spring, primarily higher frequencies are detected (except for ANNs in spring in DF of RCP 8.5), while in summer and in the DF of autumn, reduced probabilities are simulated. Annually, although often insignificant, a slightly higher frequency of RHf95 is found in the NF and a lower one in the DF of both scenarios.

10.5 Summary: Relative Humidity

In the observational reference period, relative humidity at Zugspitze is found to be high throughout the annual cycle as all months show averages above $70\,\%$. In the long term trends, relative humidity is observed to be reduced in the 20th century. In short and medium term periods, positive trends are witnessed as well, most pronounced in the time interval 1960 to 1975 in spring. The analysis of the time dependent model quality in the 20th century shows that the predictor-predictand relationship improves usually at the end, thus showing a similar picture as in the precipitation and temperature task. However, strong increasing skills are only found in winter and autumn, considerably lower ones in spring and summer. QQ-plots comparing the observational dataset with the simulated time series hint on relative humidity to be a complex downscaling target. A reason can be the hard limits of the relative humidity value range. Best fit is found at values around approximately $80\,\%$. Discrepancies between observational and simulated quantiles show strong

non-linear characteristics. In comparison to precipitation and temperature, a worse transferability of the SD models onto ESM datasets is detected. Large differences are particularly noticeable in the context of ANNs and NNCs from October to February.

Considering future changes by the multi model mean of all SD and ESM combinations, annually, relative humidity is expected to decline in the investigated scenarios and time periods. Seasonally, the strongest change signal occurs in summer. SD models agree best in a reduction of relative humidity means in summer and autumn, while mixed signals are found in winter and spring. RH1min is projected to decrease annually and for all seasons. Considering RHf95, in winter and spring primarily higher frequencies are detected, while in summer and in the DF of autumn, reduced probabilities are simulated. Annually, a slightly higher frequency of RHf95 is found in the NF and a lower one in the DF of both scenarios.

RCFs are again subjected to the restriction of a fixed value range, in terms of simulating low reference values. This matter is most evident in the comparable small or insignificant RH1min changes. Frequency changes of weather types suggest an increase of the low reference classes in spring, summer and autumn.

11 Wind Speed

11.1 An Overview over Origin and Distribution of Wind in the Alpine Region

Mainly two factors determine wind speed in mountain regions (Barry 2008): On the one hand, wind speed changes with altitude. In the mid-latitudes, where the Alps are located, average wind velocity typically increases with height in the atmosphere. The main reason for this is the reduction of friction with distance to the surface and no overlaying circulation system compensating this effect (Lauer and Bendix 2006). Isolated peaks and exposed ridges to the free atmosphere experience high average and extreme wind speeds.

On the other hand, interaction of air streams with topography can change wind velocity. Two topographic effects are operating against each other. Vertical compression of the airflow over mountains causes acceleration, in contrast to friction, which causes retardation. Friction can be further divided into skin friction, shear stress from small-scale elements below $10\,\mathrm{m}$ dimension, and form drag of large topographic obstacles in the range of $0.1\,\mathrm{km}$ to $1\,\mathrm{km}$. In mountainous regions, the latter is more important. Different suggestions can be found in literature until which altitude surface influences impact wind velocity. F. Müller et al. (1980) describe drag influences to extend up to about $1\,\mathrm{km}$ in the central Alps, while Ohmura (1990) points out that momentum transfer between the atmosphere and the mountains takes place up to $4\,\mathrm{km}$, what coincides with findings from the ALPEX (Alpine Experiment) program investigating the central Swiss Alps (Barry 2008). Schumacher (1923) analysed wind speeds of the Alpine stations Säntis, Sonnblick and Zugspitze already in the early 20th century and summarised that the deceleration of the mean annual wind speeds average to about 0.8 times of those in the free atmosphere on comparable levels. A more extensive study by Wahl (1966) for European summits concludes that wind speeds average approximately to 0.5 times the corresponding free-air values. Barry (2008) points out that wind velocity measured on summits compared to the free atmosphere heavily depends on the precise anemometer location on the investigation area. At Sonnblick, for example, southerly winds are measured stronger than those in the free air, whereas westerly and north-east to easterly winds tend to be weaker.

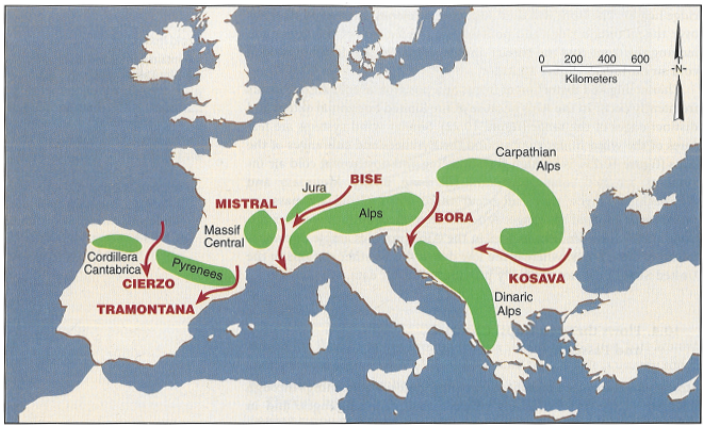


Figure 11.1: Flows between major mountain barriers in Europe. Source: Whiteman (2000).

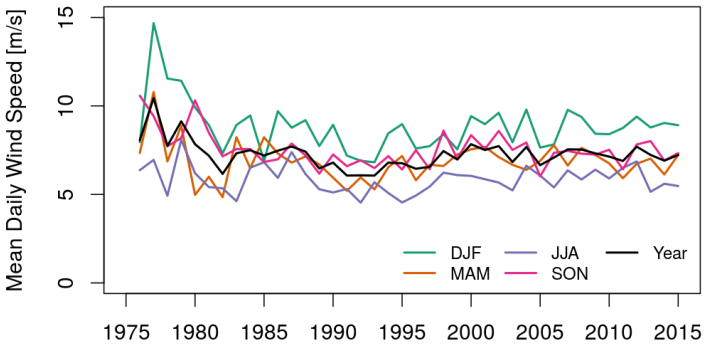


Figure 11.2: Annual and seasonal time series of the averaged wind speed observations.

Similar on the Zugspitze, southern winds can even exceed free-air values. A reason for acceleration is compression, the process can be described with the Bernoulli-Effect (Davidson et al. 1964). Assuming the mass transfer of an air stream is conserved, wind velocity needs to increase when facing constriction and decrease when channels widen out. Narrowing of the flow-through area can be found not only at ridges or summits but also at erosional openings like channels, gaps or passes, which penetrate the mountain range. Some of the flows between the mountain barriers in Europe have been given special names like Mistral, Bise or Bora, marked in figure 11.1.

Airflows crossing the Alps typically underlie the Foehn effect on which a detailed description was given in section 8.1.

11.2 Analysis of the Observational Datasets

The available observational time period of wind speed at Zugspitze starts in the year 1976 and is therefore the shortest predictand time series in this study. In terms of wind speed, the reference period is adjusted to the time span from 1976 to 2000. Figure 11.2 illustrates that there are minor differences between the seasons. Strongest wind speeds are typically found in winter, followed by autumn and spring, while the weaker ones are observed in summer. Split into monthly averages (see figure 11.5) a similar low level is monitored from May to August, with a minimum of $5.6\,\mathrm{m\,s}$ in August. The peak of monthly averages with $9.4\,\mathrm{m\,s}$ is observed in December. Thus, monthly averages feature an annual cycle. Variations between the years are rather low, from 1976 to approximately 1985, a larger spread between the years and their respective seasons is observed. The average annual wind speed of $7.3\,\mathrm{m\,s}$ in the reference period is characterised by a standard deviation of $0.8\,\mathrm{m\,s}$.

Equally to precipitation, temperature and relative humidity, a running trend analysis is performed of wind speed, see section 8.2 for methodological details. This study refrains from going into too much detail, as the observational record is rather short and in the overall period 1976 to 2015 barely any significant trends are identified. Annually on the short term scale there is a significant drop of wind speed detected in the first two 30 year windows, which is also visually apparent in figure 11.2. This trend is primarily resulting from a strong reduction in winter in addition with a pronounced, but less severe one, in autumn. Spring and summer are not subjected to significant trends. Further, no long-term trends are identified annually as well as seasonally.

11.3 Statistical Model Results with Focus on the 20th Century

The investigation of period-dependent model performance of the rather short time period reveals, similar to precipitation, temperature and relative humidity, a rise in wind speed model quality over the time (see figure 11.3). Unfortunately, the lack of data limits the scope of action. In particular NNCs but also ANNs require a high amount of training data to ensure a stable training process, which leads to the usage of the best possible congruent time interval to the other predictand calibration periods ranging from 1976 to 2000.

The quantiles of observed and simulated time series of the reference period are compared in figure 11.4. Calibration and validation quantiles of ANNs, RCFs and NNCs fit quite well, with larger discrepancies in the uppermost quantile of high wind speeds. This heavy scattering of high wind speed quantiles is particularly strong considering ESMs. Best fit of the upper quantiles is found by RCFs and NNCs

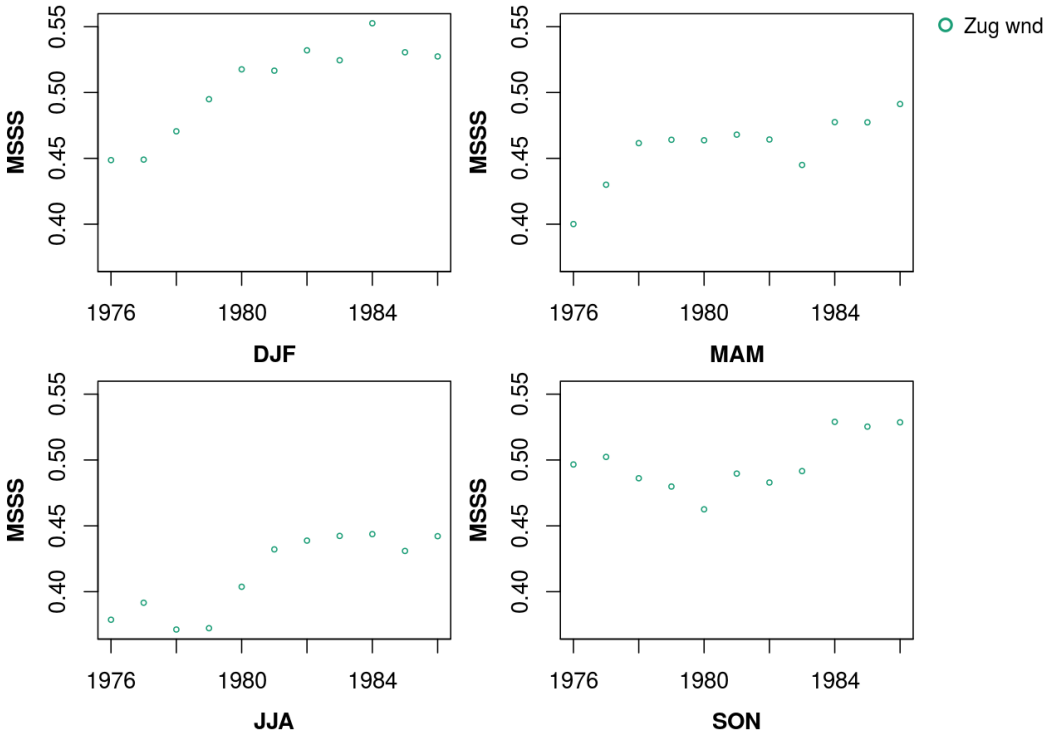


Figure 11.3: Seasonally dependent analysis of the temporal altering model skill of the final ANN configuration evaluated by the MSSS. Models are calibrated using observed wind speed (wnd) at Zugspitze (Zug) and the 20th century reanalysis dataset. Training is repeated from 15 different ANN initialisations, the resulting skills scores are averaged. Models are calibrated on a 30-year period, starting at each mark of the graphic.

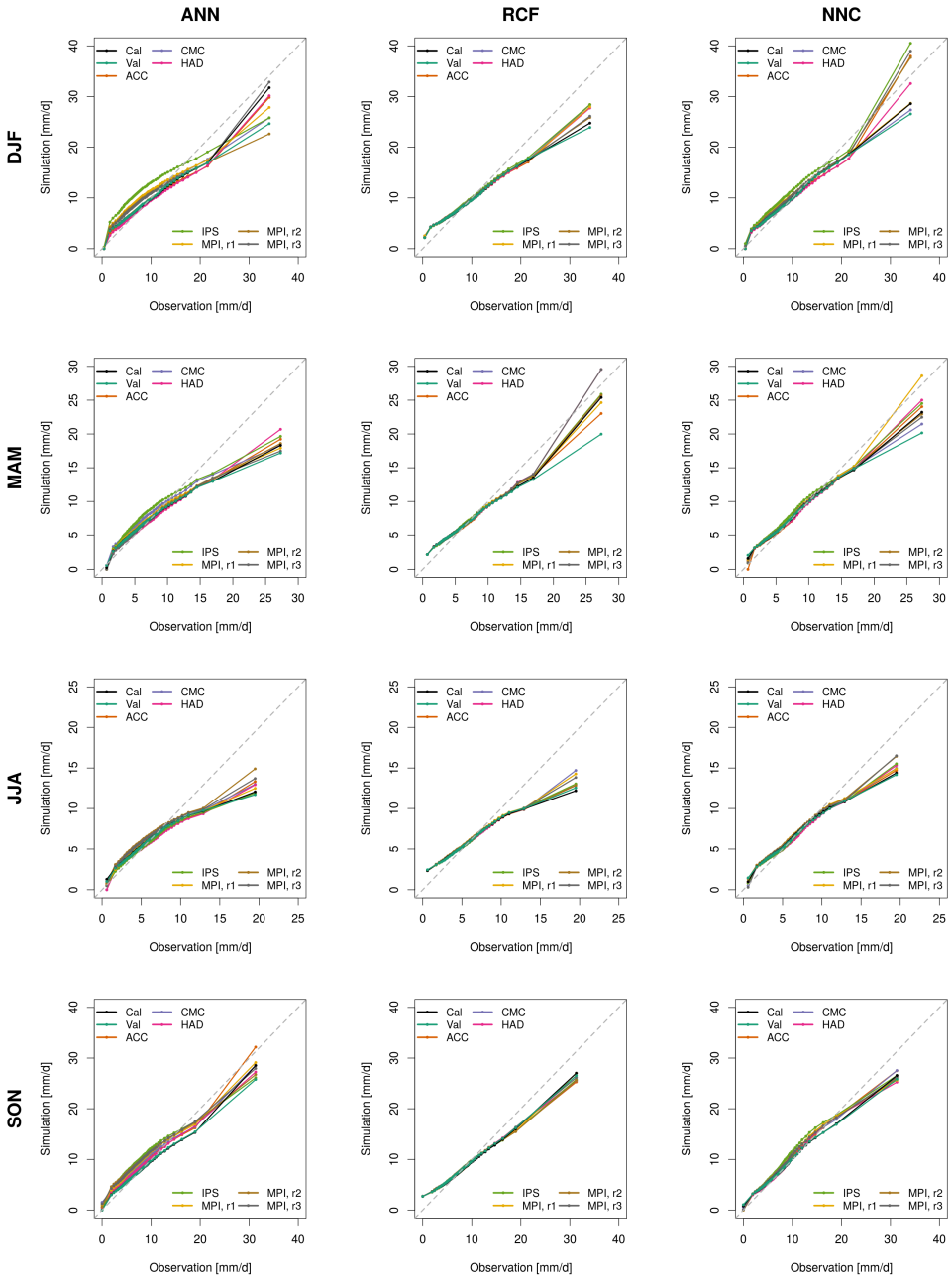


Figure 11.4: QQ-plots for Zugspitze wind speed considering the seasonal models in the reference period 1971-2000. 50 quantiles of the observed dataset are compared with their simulated counterparts from reanalysis and historical ESM data-based downscaling products derived by the final SD models. Reanalysis series are separated by calibration (Cal) and validation (Val) periods. The fifteen Cross-validation runs are treated as one sample.

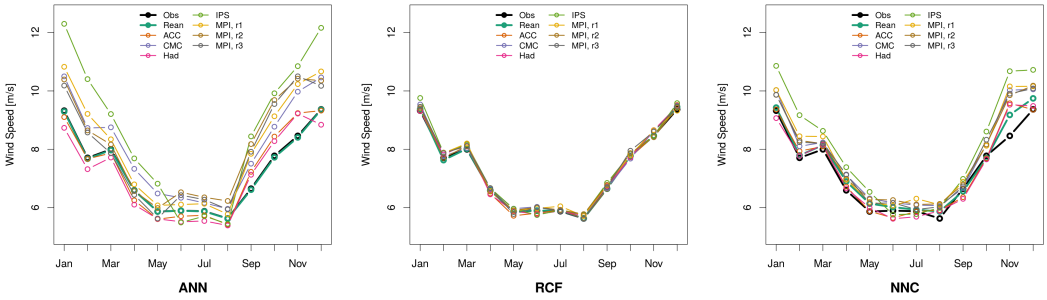


Figure 11.5: Observed (Obs) and simulated (complete reanalysis (Rean) and climate model driven results) monthly averaged daily mean wind speeds considering the reference period 1971-2000 for each SD model type.

in autumn. Good accordance between the observed and simulated wind speeds is found in the central part of the value distribution. ANNs and NNCs include a good representation of the low-speed winds. The low occurrence of weak wind speeds, lead to no specialised low wind speed classes. As a result, the reference values of RCFs for low wind speeds are too high. SD model transfer onto historical ESM datasets exhibits largest differences in terms of ANNs in the central parts, while RCFs and NNCs agree well with the quantiles of the reanalysis driven runs.

ANNs, RCFs and NNCs represent the annual cycle of monthly averages well (see figure 11.5). The reanalysis based results of the three SD models are almost identical to the observed reference with two minor exceptions by NNCs in November and December. Analogously to the predictands precipitation, temperature and relative humidity, the transfer of RCFs onto ESM data shows high accordance to the observations. Slight but noticeable differences emerge from NNCs, the largest discrepancies are produced by ANNs. There is no clear tendency, which ESM fits the reanalysis runs best, although downscaling products derived from ACCESS and HadGEM by ANNs and NNCs, especially from October to February, are noticeably closer to the reanalysis driven simulations than other ESMs.

11.4 Climate Change Signals: 20th vs. 21st Century

Preliminary, the characteristics and behaviour of RCFs in a climate change scenario shall be discussed. The reference values of RCF types, displayed in figure 11.6, underlie a slight regressive reduction from class 1 to 27. In autumn and spring, a more rapid decrease is found in the first reference values. The seasonal order of reference values considering equal type numbers typically starts from winter, with the strongest magnitudes, autumn, spring and ends in summer, where the weakest ones

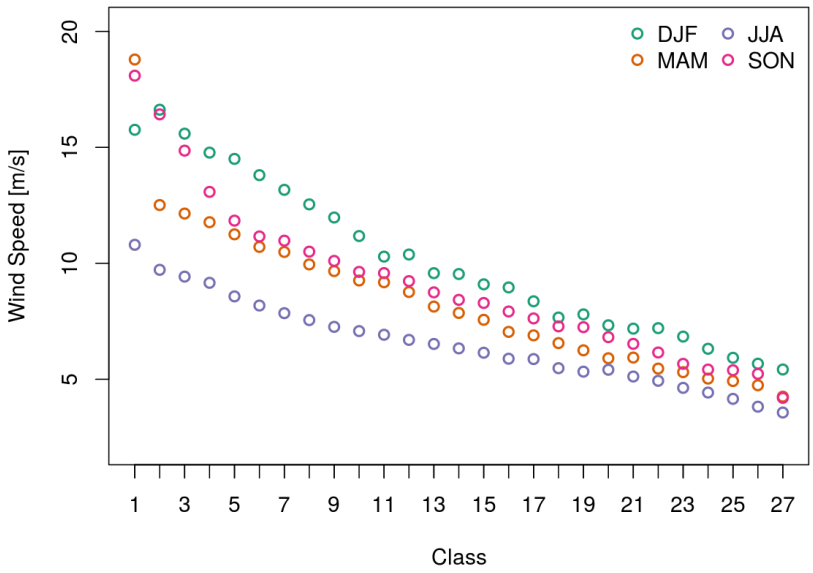


Figure 11.6: Circulation type specific mean wind speed reference values in [ms^{-1}] for each of the 27 classes of the seasonal RCFs at Zugspitze and Sonnblick derived from the calibration period 1971-2000.

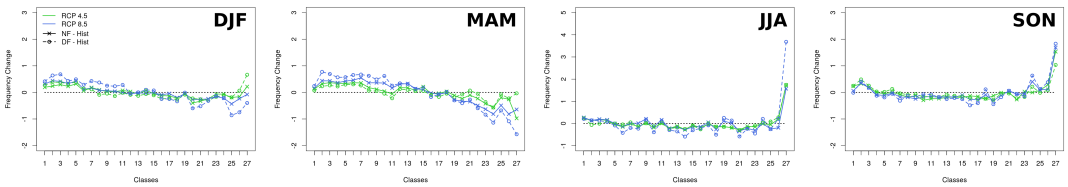


Figure 11.7: Average future changes of type frequency in [%] comparing the historical reference (1971-2000) of all ESMs with both scenarios (RCP 4.5 and RCP 8.5) in the near (NF) and distant future (DF) of the seasonal RCF wind speed models at Zugspitze.

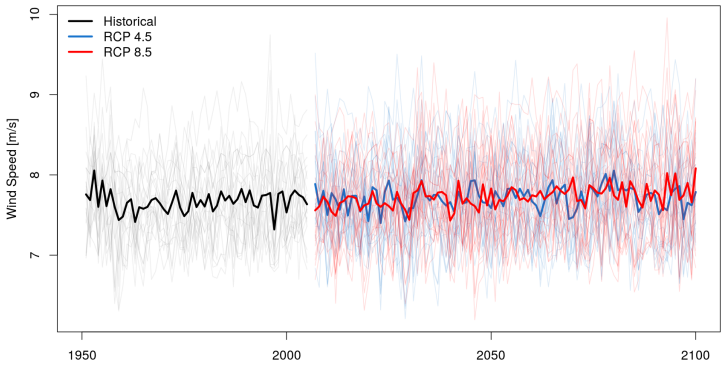


Figure 11.8: Annually averaged daily wind speed means resulting from the ESM downscaling product of the total available time period 1950 to 2100 (historical and scenario dataset). Thick lines represent the overall average of ESM results. Thin lines in the background display the mean of the cross-validation procedure considering a single SD model and ESM combination.

are found. There is no weather type representing calmness, the lowest reference values of class 27 in summer is $3.6\,\mathrm{m\,s^{-1}}$.

Future frequency changes of classes in winter and spring are primarily positive at the lower class numbers and negative at the higher ones, with the exception of RCP 4.5 in winter (see figure 11.7). In summer and autumn, a rapid frequency increase of class 27 occurs, representing low wind speeds, while stronger frequency reductions are found in the central classes. In addition, in summer and autumn, slight positive trends are witnessed in the strong wind section. Especially the change in spring, summer and autumn of the higher weather type numbers can be an indicator for the necessity of a reference class exhibiting weaker winds. In general, in RCP 8.5 the change is continued in the DF. In RCP 4.5, the development cannot easily be generalised, as in the DF often a reverse tendency of the change signals is simulated in comparison to the NF.

In the following the climate change analysis results of ANNs, RCFs and NNCs are discussed. Individually investigated, ESM driven future projections of the SD models seldomly project significant changes, as shown in table 11.1. Most of the models hinting on changes in wind speed conditions belong to the MPI-ESM group. The suggested significant change signals by the MPI-ESMs are restricted to the DF and occur more frequently in the RCP 8.5 scenario. All significant results show an increase in the annual wind speed, in comparison to RCP 4.5, the change values in DF of RCP 8.5 are typically larger. Most likely the trend signal of the MPI-ESMs determines the multi model mean of all ESMs and SDs hinting on slight increases in the DF of RCP 4.5 ($1.1\,\%$) and RCP 8.5 ($1.6\,\%$). In a similar vein, Gobiet et al. (2014) suggest that the overall wind speed in the Alpine region is stable, with no

Table 11.1: Historical (Hist: 1971 to 2000) annually averaged wind speed [$m\,s^{-1}$] at Zugspitze as well as projected relative changes to the historical period [%] considering different ESMs, SD models and scenarios. Changes are computed for the near future (NF: 2021 to 2050) and distant future (DF: 2071 to 2100). Positive numbers are marked in green, negative ones in red, while insignificant values (U-test, significance level 5 %) are labelled in grey. (Observational average Zugspitze: 7.2 $m\,s^{-1}$).

			RCP 4.5		RCP 8.5	
		Hist	NF	DF	NF	DF
ACC, r1	ANN	7.3	0.3	2.0	1.4	-0.3
	RCF	7.3	1.1	2.6	2.4	0.7
	NNC	7.4	0.2	1.9	1.4	-0.2
CMC, r1	ANN	8.1	0.9	-1.2	-0.5	0.5
	RCF	7.3	1.8	-0.6	0.2	1.4
	NNC	7.8	0.7	-1.3	-0.5	0.3
HAD, r1	ANN	7.1	0.5	0.6	-0.3	-0.9
	RCF	7.3	0.2	0.3	-0.1	-1.0
	NNC	7.3	0.5	0.7	-0.2	-0.7
IPS, r1	ANN	8.7	-0.2	0.2	-1.0	2.1
	RCF	7.4	-0.4	0.3	-0.9	2.0
	NNC	8.1	-0.4	0.6	-1.0	2.3
MPI, r1	ANN	8.1	0.1	0.5	-0.7	1.7
	RCF	7.4	0.6	1.1	-0.1	2.7
	NNC	7.8	0.0	0.5	-0.7	1.7
MPI, r2	ANN	8.1	-0.1	1.5	1.6	2.5
	RCF	7.3	0.5	2.3	2.7	3.5
	NNC	7.8	-0.3	1.4	1.4	2.6
MPI, r3	ANN	8.0	1.3	2.6	0.3	3.2
	RCF	7.3	2.2	3.6	1.6	4.5
	NNC	7.7	1.7	3.1	0.7	3.5
All CMs	ANN	7.9	0.4	0.9	0.1	1.3
	RCF	7.3	0.9	1.4	0.8	2.0
	NNC	7.7	0.3	1.0	0.1	1.4
All CMs	All SDs	7.6	0.5	1.1	0.4	1.6

large annual changes.

Separated by seasons, the boxplots in figure 11.9 show divers tendencies in the future development of wind speeds throughout the year. In winter and spring, an increase in wind speed is suggested by all SD models, while in summer and autumn weaker ones are simulated. Gobiet et al. (2014) confirm decreasing tendencies in summer and autumn at some locations in the Alps. Overall, in the seasonal view, a high number of stable climatic conditions is suggested. Similar to the annual evaluation, changes usually further intensify in the respective DF with some exceptions in winter (NNCs, RCP 4.5) and autumn (RCFs and RCP 4.5 of NNCs). Between the scenarios, the affect of changes are intensified in the DF in RCP 8.5, when compared to RCP 4.5, although in the NF it is often inverted. The relative increases in spring

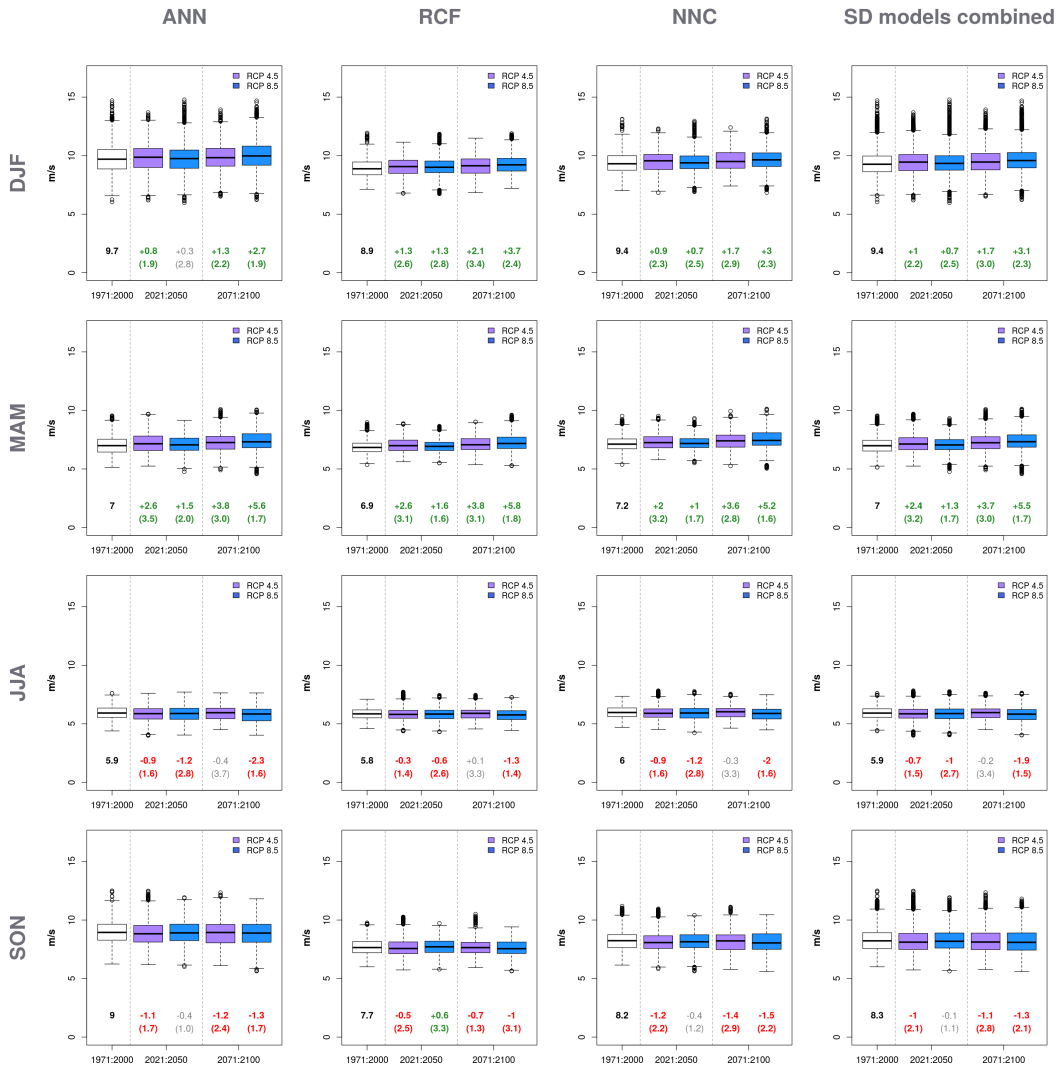


Figure 11.9: Seasonal dependent comparison of climate change signals separated by SD models considering all ESM downscaling products for daily wind speed means at Zugspitze. Differences of the overall mean between the historical reference period and future scenarios are given relatively [%]. In addition, the standard deviation of the diverse change signals originating from different ESMs is listed in parenthesis. Significant increases are coloured in green, decreases in red. Significance is tested using the U-test (level of significance 5 %). Observational references: DJF: 8.8 m s^{-1}, MAM: 6.8 m s^{-1}, JJA: 5.8 m s^{-1}, SON: 7.6 m s^{-1}.

are higher than in winter. In summer and autumn, no clear ranking can be obtained.

The observed standard deviation of annual and seasonal wind speed is underestimated by the SD models. Results in the reference period from reanalysis-driven and ESM-driven SD models are similar, solely in autumn, ESMs tend to show larger standard deviations than of reanalysis experiments. In general, the observed and simulated standard deviations are rather small and so are their changes in future scenarios. In winter and spring, primarily an increasing variability is detected. In summer, no change is observed, while in autumn, especially in the DF of RCP 4.5, the SD model consents shows a reduction in magnitudes. The latter is mainly overcompensated in annual terms, as significant changes suggest an increase throughout the scenarios and future periods. Annually, in RCP 4.5, an initial increase between hist and NF is followed by a slight reduction between NF and DF, while in RCP 8.5 a further increase in variability is found.

In addition, indices of the average annual maximum daily mean wind speed (WS1max) and the frequency of days above $12\,\mathrm{m\,s^{-1}}$ (WSf12) are investigated. Statistics are listed in the appendix in table E.1. Due to the exposed location, the annual extreme value of strong winds WS1max can reach quite high magnitudes, in the observational period with a maximum of $22.3\,\mathrm{m\,s^{-1}}$ in winter and a minimum of $13.1\,\mathrm{m\,s^{-1}}$ in summer. SD models underestimate the observed magnitudes. Reanalysis- and ESM-driven experiments of the reference period match well. All significant changes hint towards an increase in WS1max. Seasonally, significant future changes scatter strongly between the SD models. The only agreement of ANNs, RCFs and NNCs is found in the DF of RCP 8.5. Noticeable is that despite the limitations of a fixed value range, RCFs frequently suggest significant changes. Even the magnitudes are quite similar to ANNs, except in winter, where the RCF changes are considerably lower. In contrast to ANNs and RCFs, NNCs project steady wind speed conditions in spring, summer and autumn, throughout the future scenarios and periods. Biggest changes are found in the DF of RCP 8.5 in spring closely followed by winter. Annually, exclusively significant changes are found in the considered RCP 4.5 time periods, whereby the increase in WS1max from hist to NF is stronger than from the NF to DF. In RCP 8.5, ANNs and NNCs suggest only a significant increase in the NF, in contrast to RCFs, where only the DF is significant.

Windy days are further evaluated by the WSf12 index. Although an average daily wind speed of $12\,\mathrm{m\,s^{-1}}$ seems moderate, short duration wind gusts with much higher wind speeds, can be expected on these days as well. A statistical analysis of maximum daily wind speeds, which are available from 1980, shall demonstrate the effect. From 1980 to 2000, the average maximum daily wind speed of WSf12 days reaches $38.3\,\mathrm{m\,s^{-1}}$, which is classified in the highest Beaufort number "12" also referred to as "Hurricane force" (Weischet and Endlicher 2008). The daily maximum

Table 11.2: Changes in standard deviation of daily wind speed at Zugspitze considering the observational record (Obs) as well as downscaling products based on the reanalysis dataset (Rean, calibration and validation) and ESM experiments (historical: Hist; Scenarios: RCP 8.5 and RCP 4.5) separated by season and annually. Results from cross-validation folds and ESMs are averaged. Number in brackets in Hist: standard deviation between ESMs. Number in brackets in scenarios: standard deviation of the change signal. Reference periods: 1971-2000 (Obs, Rean, Hist); 2021-2050 (NF); 2071-2100 (DF). Changes: Increasing values are marked in green, declining ones in red and insignificant ones in grey (more than two ESMs show no change signal or a different sign.). Absolute change signals are computed by comparing historical and scenario results.

					RCP 4.5		RCP 8.5	
		Obs	Rean	Hist	NF	DF	NF	DF
DJF	ANN	1.79	0.77	0.73 (0.08)	0.04 (0.13)	0.10 (0.11)	0.00 (0.09)	0.04 (0.14)
	RCF	1.79	0.73	0.73 (0.08)	0.05 (0.13)	0.11 (0.09)	0.02 (0.12)	0.08 (0.10)
	Comb	1.79	0.77	0.74 (0.09)	0.03 (0.11)	0.09 (0.10)	-0.02 (0.08)	0.04 (0.12)
MAM	ANN	1.33	0.54	0.58 (0.08)	0.04 (0.09)	0.10 (0.19)	0.00 (0.13)	0.10 (0.11)
	RCF	1.33	0.55	0.58 (0.07)	0.03 (0.10)	0.09 (0.17)	-0.03 (0.13)	0.09 (0.11)
	Comb	1.33	0.56	0.61 (0.08)	0.03 (0.10)	0.08 (0.19)	-0.02 (0.13)	0.07 (0.11)
JJA	ANN	0.92	0.44	0.47 (0.03)	0.03 (0.11)	0.01 (0.11)	0.06 (0.07)	0.08 (0.10)
	RCF	0.92	0.43	0.47 (0.03)	0.02 (0.10)	-0.01 (0.11)	0.05 (0.07)	0.03 (0.10)
	Comb	0.92	0.47	0.49 (0.03)	0.01 (0.12)	-0.01 (0.11)	0.03 (0.09)	0.05 (0.11)
SON	ANN	1.00	0.56	0.74 (0.10)	0.03 (0.12)	-0.03 (0.14)	-0.04 (0.07)	-0.03 (0.08)
	RCF	1.00	0.51	0.68 (0.08)	0.05 (0.09)	-0.02 (0.14)	-0.01 (0.08)	-0.02 (0.09)
	Comb	1.00	0.55	0.78 (0.11)	0.01 (0.11)	-0.04 (0.16)	-0.06 (0.10)	-0.05 (0.10)
Annual	ANN	1.00	0.27	0.30 (0.04)	0.08 (0.05)	0.05 (0.06)	0.05 (0.04)	0.07 (0.05)
	RCF	1.00	0.25	0.30 (0.03)	0.07 (0.05)	0.04 (0.05)	0.04 (0.03)	0.06 (0.05)
	Comb	1.00	0.27	0.32 (0.04)	0.07 (0.05)	0.04 (0.05)	0.04 (0.03)	0.06 (0.05)

wind speed values of WSf12 days range from $22.1\,\mathrm{m\,s^{-1}}$ to $93\,\mathrm{m\,s^{-1}}$. WSf12 days therefore lead to increasing impairments up to life threatening conditions for people on-site. SD models driven by reanalysis underestimate the observational record of WSf12. The difference between reanalysis- and ESM-driven SD results in the historical reference period is small in spring and summer. RCFs fit also well in winter and autumn, but ANNs and NNCs are characterised by much larger WSf12 values using ESM input than in terms of reanalysis, mostly even surpassing the observational record. The SD model specific future changes show a little more consents like in WS1max, but still prohibit a clear congruent tendency. Seasonally as well as annually, primarily a rise in the occurrence of WSf12 is detected, with the exception of autumn, where reductions are suggested as well, but without model consents. Best SD model agreement is projected in spring and in the NF of RCP 4.5 in winter. Changes in summer and autumn are marginal, slightly larger ones are projected in winter and spring. The difference between NF and DF in RCP 4.5 is again much smaller than between hist and NF. Annually, all SD models project a rise in WSf12 in the DF, whereby a higher increase is found in RCP 8.5. Annual changes in the NF are similar between the scenarios but often remain insignificant.

11.5 Summary: Wind Speed

The observational record of wind speed starts in 1976. For that reason, the reference period of wind speed is adjusted to 1976-2000. Due to the exposed location, wind speeds at Zugspitze can be comparably high. In the overall period, barely any calm days are observed. In monthly averages, strongest winds are found in winter and lowest ones in summer, the time series therefore show an annual cycle. In the observed time series no long term trends are identified annually as well as seasonally.

SD models are able to reproduce the monthly wind speed averages in their annual cycle. Driven by ESMs, the resulting magnitudes of monthly averages are usually overestimated by ANNs and NNCs but show low discrepancies by RCFs. Comparing the quantiles of observed and simulated time series, largest differences are found predominantly in the uppermost quantile, which is usually underestimated and also shows the highest spread in the downscaling products.

Annually, barely any future changes are identified by single ESM products, solely in the multi model mean of the overall ensemble a slight increase in wind speeds is detected in the DF of both scenarios. Seasonally, experiments in winter and spring are characterised by an increasing average wind speed, while summer and autumn show mostly weaker ones. WS1max values are often steady, if significant exclusively increases are projected. Best SD model agreement is found in RCP 4.5 of the annual evaluation, otherwise consents of significant values between the statistical methods

are rare. Seasonally, as well as annually, primarily a rise in the occurrence of WSf12 is detected. Best SD model agreement is thereby projected in spring. Annually, in WSf12, best model agreement is found to be in the DF of both scenarios, where a higher future occurrence is simulated.

Frequency changes of RCFs show the typical problem, analogous to temperature and relative humidity, but the impact on time series statistics is rather marginal. No distinct dissents between RCFS and ANN/NNCs are identified. In terms of wind speed, RCFs fit well into the general picture, even in the evaluations of extremes.

12

Climate Impact Analysis on Snow Depths

12.1 Snow in the Alpine Region and the Relationship to Glacier Ice

Height, duration and water-equivalent of snow in the Alpine region is primarily influenced by temperature, precipitation or regional effects (Cebon 1998; Barry 2008). The term snow water equivalent refers to the amount of water that would be obtained upon complete melting of a specific snowpack. In details, following Warscher et al. (2013), snow accumulation is dependent on snow precipitation, resublimation and lateral snow transport, while snow ablation is affected first of all by snowmelt itself, but also sublimation and again lateral transport. Lateral snow redistribution is mainly caused by gravitational and wind driven processes. In addition, resulting snow depths depend on the type of falling snow (Sturm et al. 2010): Although carrying the same snow water equivalent, accumulation of wet or dry snow, thus snow with different density, leads to different snow depths. The average climatic snow line varies between $2400\,\mathrm{m}$ to $3000\,\mathrm{m}$, usually tends to be lower in the northern parts of the Alps caused by latitudinal effects on insolation and on shaded northern (ubac) slopes in general (Cebon 1998). In the centre of the Alps, the snow line elevation is increased by the mass elevation effect. Climate signals in snow cover are hard to identify, as annual snow characteristics alternate heavily. Scherrer et al. (2013) further describe the difficulties in understanding change signals, as various local influences, e.g. temperature, precipitation, radiation, wind and humidity, and variabilities in the atmospheric circulation interfere. Snowmelt in summer leads to a nival runoff regime peaking in summer. If glaciers are involved, the runoff peak shifts, or is extended, towards autumn (glacial runoff regime). Foehn events can accelerate snowmelt on the Lee side. Snow accumulation depends heavily on temperature. Cebon (1998) in accordance with Hantel and Maurer (2011) estimate a $100\,\mathrm{m}$ to $200\,\mathrm{m}$ rise of the snowline altitude per $1\,\mathrm{K}$ warming of the annual mean temperature. If snow cover persists permanent, glaciers can form (Fierz et al. 2009). Fresh snow is transformed via snow metamorphosis and packing to Firn, an at least one year old, well bonded and compact snow. Mainly by further packing, Firn can finally be converted into glacier ice, a process which can take several years.

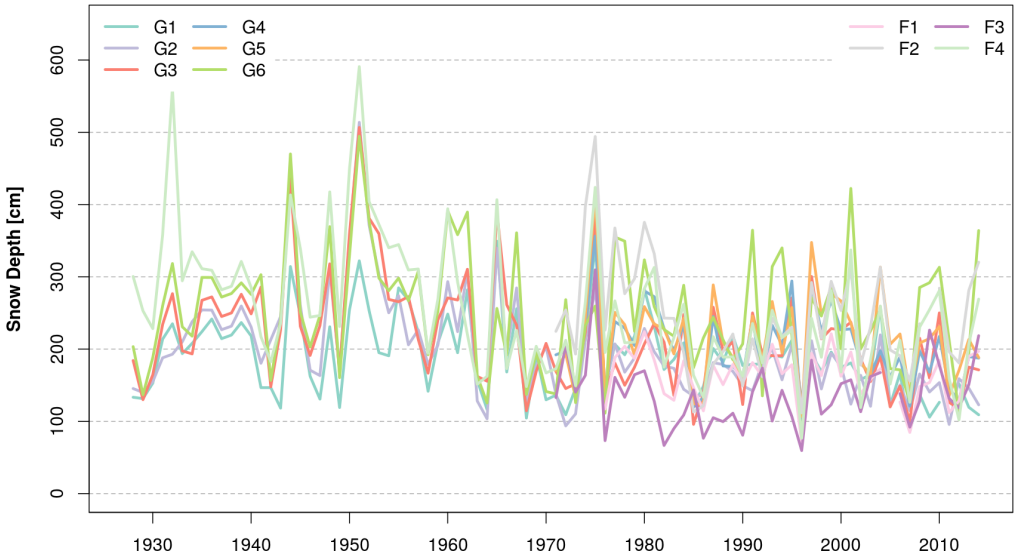
Table 12.1: Relative (rel) and absolute (abs) changes in average snow depths [cm] of the long term observational series at Sonnblick and Zugspitze. Changes are computed from the difference in means of time periods 1938-1967 (mean) and 1971-2000. Marked in grey are insignificant changes (U-test, level of significance 5 %). At Zugspitze, an additional experiment Z1* comparing 1955-1984 and 1985-2014 is carried out. Measuring site IDs and respective locations are listed in table 4.2.

	G1	G2	G3	G6	F4	Z1	Z1*
mean	213	254	270	286	301	175	216
rel	-11.7	-32.5	-26.4	-17.6	-28.3	22.3	-17.2
abs	-25	-83	-71	-50	-85	39	-37

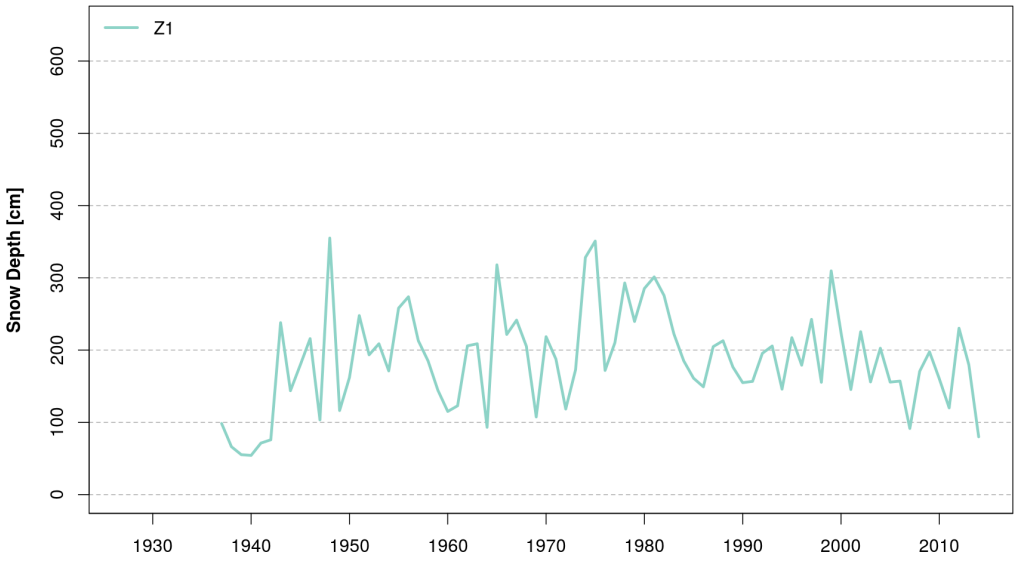
12.2 Analysis of the Observational Datasets

The development of annual snow depth averages is displayed in figure 12.1. Annual averages are computed considering the hydrological year starting from November. The definition is oriented on Schöner et al. (2009) who investigated the snow depth development in the Alps during the 20th century. The annual cycle is further described: The core winter starts in February and the maximum snow depth is reached in May. August is considered as core summer. The snow depth minimum consisting of remaining firn is observed in October. Monthly snow depth averages of the reference period 1971 to 2000 are illustrated for Sonnblick in figure 12.3 and for Zugspitze in figure 12.4. In the reference period, snow depth maximums are conform to the definition in May at all locations, but the month of minimum snow depth can vary from September to November, depending on the measuring site. The hydrological year at Zugspitze in this study is defined equally to Sonnblick, as the overall annual cycles are similar: Zugspitze snow depth maximum is found in May and the minimum in October. In May, most snow gauges at Sonnblick and Zugspitze, show monthly averages above $350\,\mathrm{cm}$, except G2, F1 and F3. Snow gauges are located in or near the glacier region. Snow-free months occur at all measuring sites, most frequently in September and October. The interannual snow depth variability of yearly means is very high. The average standard deviation of Zugspitze and Sonnblick gauges is $58\,\mathrm{cm}$. Annual snow depth averages range from $132\,\mathrm{cm}$ at F3 to $244\,\mathrm{cm}$ at F2. See also table 12.4 for more details on annual averages and standard deviation.

To evaluate changing conditions of the average snow depth in the 20th century a 30 year period at the start of the measurement is compared to the reference period 1971 to 2000. The first period was selected to be congruent available at all stations but not overlapping with the second period. Observational records, which do not fulfil the requirements are omitted. Results are listed in table 12.1. Measuring sites



(a) Sonnblick



(b) Zugspitze

Figure 12.1: Annual snow depth averages from the observational records at Sonnblick and Zugspitze. Measuring site IDs and respective locations are listed in table 4.2.

G2, G3 and F4 are characterised by a reduction of snow depths by one fourth to one third compared to the reference time period. G6 shows a weaker but still strong decrease. The change signals at G1 and Z1 are insignificant. Visually, the highest snow depths at Sonnblick are found in 1940 to 1950, afterwards an initial rapid decrease is observed until 1980, where snow depths started to stagnate at a lower level. At Zugspitze snow depth reduction set in later, beginning in approximately 1980 with subsequent declining observation values. This is confirmed if the time slices to be compared are shifted onto the end of the observational record. The successive periods 1955 to 1984 and 1985 to 2014 exhibit a reduction of $17.2\,\%$ in Z1*. The low levels of snow depth from 1970 on are Alpine-wide described (Schöner et al. 2009; Beniston 2006; Laternser and Schneebeli 2003). The cause can partly be traced back to the mainly positive NAO phase, which resulted in unusual warm temperatures in combination with high pressure fields leading to low precipitation sums. However, Zampieri et al. (2013) point out that the reduction cannot solely be explained by the NAO phase.

12.3 Modelling Snow Depth: Statistical Model Framework

The aim of this chapter is to calibrate statistical models which reproduce the snow depths at Sonnblick and Zugspitze. ANNs are used to build the transfer function between predictors and predictand. The ANNs are calibrated on the observational record, while in the future, impact analysis time series from the statistical downscaling framework are applied. Local temperature and precipitation serve as predictors on which snow processes largely depend on (Schöner et al. 2009). Note that this two predictor setup cannot directly care for some important aspects like snow relocation. Mott et al. (2008) showed for Sonnblick that additional modules for relocation improve the understanding of snow depth related processes at least in dynamical models. In this thesis, the limited access and availability of long term observational datasets, beside temperature and precipitation, which could serve as predictors impedes the building of complexer snow depth models.

In comparison to the previous SD-framework, some adjustments to the model development process are necessary. First of all, the Zugspitze snow depth record as well as predictor observations and statistical downscaling products are reduced to a monthly time scale to get an identical structuring to the Sonnblick snow depth dataset for the purpose of comparability. The low amount of observations in a monthly dataset can be critical for ANNs. To provide a maximum amount of training pattern, the whole available record is used during calibration ($80\,\%$ of the total record) and validation ($20\,\%$ of the total record). In addition, one model is calibrated using the entire year. Exemplary ANNs of a preliminary study, calibrated on

seasonal or half-year splits showed lower model skill, thus the annual approach is preferred. At Zugspitze, the years 1937 to 1942, which are characterised by unusually low snow depths, are omitted, as a considerable gain in model performance is achieved without them.

The data preprocessing includes several steps. First, standardisation and scaling of the predictor and predictand datasets is performed in the given order. The procedure is executed for each month individually to eliminate the annual cycle. In terms of the ESM based runs, standardisation and scaling parameters are derived from the overall historical period (1950-2005) to ensure a as much as possible congruent segment to the observational record. Afterwards, the historical parameters are projected onto the future scenarios. Second, recent snow depths can be accumulated throughout several months (e.g. Schöner et al. (2009)). A method to allow an extended retrospective of previous weather conditions is to aggregate preceding months by means or sums (Smiatek et al. 2013). Hence, up to three previous months are examined for an added value to the statistical model. Third, cross-correlations are evaluated to identify a possible time offset of the predictor influence on the predictand. Hereby, predictor and predictand time series are displaced one relative to the other to find a peak in the relationship by considering time lags. Lags of each final predictor combination are summarised in table 12.2. Usually, a shift of one is detected, which indicates that snow depths primarily depend on the climatic conditions of the previous month. Next in line is to find a good predictor setup. In this study, only one temperature and one precipitation time series is included into an ANN at a time, in order to avoid redundancy. A full grid search procedure is performed to find the best combination. ANNs are trained with two hidden neurons using $80\,\%$ of the observational record in an 150-fold Monte Carlo cross-validation approach. The predictor combination with the highest average Spearman correlation coefficient in the test sample (remaining $20\,\%$ of the dataset) is selected as final configuration. Best predictor setup for each predictand is found in the corresponding row of table 12.2, marked by a lag value. Usually long-term aggregates are identified as best predictors, supporting the thesis that snow depth is stacking up over a substantial time in winter.

In the next step, the number of hidden neurons is optimised for each target. A number of two to 50 hidden neurons is evaluated, while $80\,\%$ of the observational record in a 30-fold Monte Carlo cross-validation approach is used for calibration. Configurations with the highest Spearman correlation coefficient in the validation period are selected. Usually, small numbers of hidden neurons are sufficient, however, in some cases large numbers are selected as well in the automated process. Further analysis showed a negligible small improvement of runs using high hidden neuron numbers: Highest gains in the average validation Spearman correlation coef-

Table 12.2: Time lag peak of the cross-correlation process between predictors and predictands specified for the final predictor combination. A value of one suggests to shift the predictand dataset along the time line by one month into the positive direction. Predictand acronyms are listed in table 4.2. Predictor acronyms are compounds from precipitation (P) and temperature (T) monthly aggregated to sums (s) or means (m). In addition, if several months are taken into account, the computational type, sums (S) or means (M), and the total number of months are denoted. For example TmM2: Temperature average of the current and preceding month.

	Ps	**PsS2**	**PsS3**	**PsS4**	**Tm**	**TmM2**	**TmM3**	**TmM4**
G1				1			0	
G2	1						0	
G3				1			0	
G4	1						1	
G5			1		1			
G6				1		1		
F1			1				0	
F2		1						0
F3			1					0
F4				1			0	
Z1	1							1

Table 12.3: Number of hidden neurons for each predictand-specific ANN.

G1	G2	G3	G4	G5	G6	F1	F2	F3	F4	Z1
5	4	2	5	2	2	2	2	2	5	3

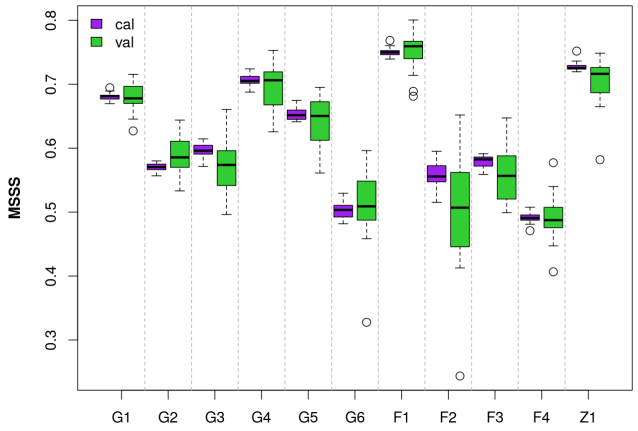


Figure 12.2: Final MSSS results from the calibration and validation procedure of Sonnblick and Zugspitze snow depth models.

ficient, for example, are found at snow gauge G1 with 0.005 (5 vs. 46 hidden neurons) and G2 with 0.004 (4 vs. 44 hidden neurons). In simple structure ANNs with only two input variables the high amount of hidden neurons is non-credible. In combination with the low increase in model skill, a restriction to a maximum number of five hidden neurons is decided. Results are listed in table 12.3.

Finally, the ANN models are calibrated considering the selected predictor combination and optimised number of hidden neurons. A total of 15 Monte-Carlo cross-validated models are generated from 80 % of the observational dataset. The resulting model skill is displayed in figure 12.2. Both stations indicate an overall good model performance with medians usually above an MSSS of 0.5. Some models even surpass a MSSS of 0.7 (F1, Z1, G4).

In a final assessment of the ANN snow depth framework, the resulting models are compared to the much simpler and easier to handle multiple linear regressions technique. Using ANNs to generate a suitable transfer function between two inputs and one output variable can be rated as quite of an overkill, especially if the relationship between predictors and predictands are not of highly non-linear character. The final predictor combinations are therefore fitted by linear regressions and the skill of the regression models is compared to the ANNs. 15 regression models are calibrated in a similar manner as the ANNs and the difference in the resulting model skills is evaluated by an U-test (level of significance 5 %). Considering the calibration results,

ANNs performed better but usually no significant differences are detected. In terms of validation, no significant differences between regression models and ANNs are detected at all. Taken into account the high expense in calibrating ANNs and finding suitable ANN parameters as well as the general disadvantages of extrapolation with non-linear methods (see section 7.4), in this case of snow models with low complexity, using regression analysis is a good alternative if the statistical requirements of transferability are fulfilled.

12.4 Statistical Model Results in the 20th and 21st Century

In the following, the calibrated ANNs are used to derive future changes in snow depth development from the SD products of ANNs, RCFs and NNCs discussed in the previous chapters. Figures 12.3 and 12.4 display the monthly snow depth averages of observed and simulated time series at Sonnblick and Zugspitze, respectively. In general the annual cycle of monthly snow depths is well represented in the simulations. The Zugspitze model shows slightly stronger discrepancies and scatter around the observational record. Larger differences at both stations are often found near the peak, although no systematical over- or underestimations take place. Further, in September and October, the statistical models often do not trail the observed declining snow depth, but show a less strong reduction instead. Optimisation results and ESM-driven output magnitudes lie closely together.

The future snow depth development is illustrated in figures 12.5 and 12.6 for Sonnblick and Zugspitze, respectively. The time series in the observational period as well as future changes are statistically evaluated in table 12.4 by annual means, inter annual standard deviation (ISD) and number of months with no snow cover (NM0). The observational average annual snow depth is well represented by the models during training and when transferred on historical ESM datasets. There are no systematics in over- or underestimation, however depending on the target series both can occur. In future scenarios, a decreasing snow depth tendency is simulated. G5 is the only station showing stable conditions in the NF of the two scenarios, in the DF the decreasing tendencies at G5 are comparable moderate as well. Similarly, F1 is characterised by a less strong climate impact. Snow depth reductions are in general projected to further increase in the DF and to be more severe in the RCP 8.5 scenario. Especially in the DF of RCP 8.5, typically strong losses between $30\,\%$ and $45\,\%$ are simulated. In comparison, RCP 4.5 shows less intense reductions between approximate $15\,\%$ and $25\,\%$. Strong reductions are usually simulated for the stations at lower altitudes, as in future scenarios sites at higher elevation still show low temperature values suitable for the genesis and storage of snow. This characteristic can also be generalised for the Alpine region (Steger et al. 2013; Beniston et al. 2003). The accor-

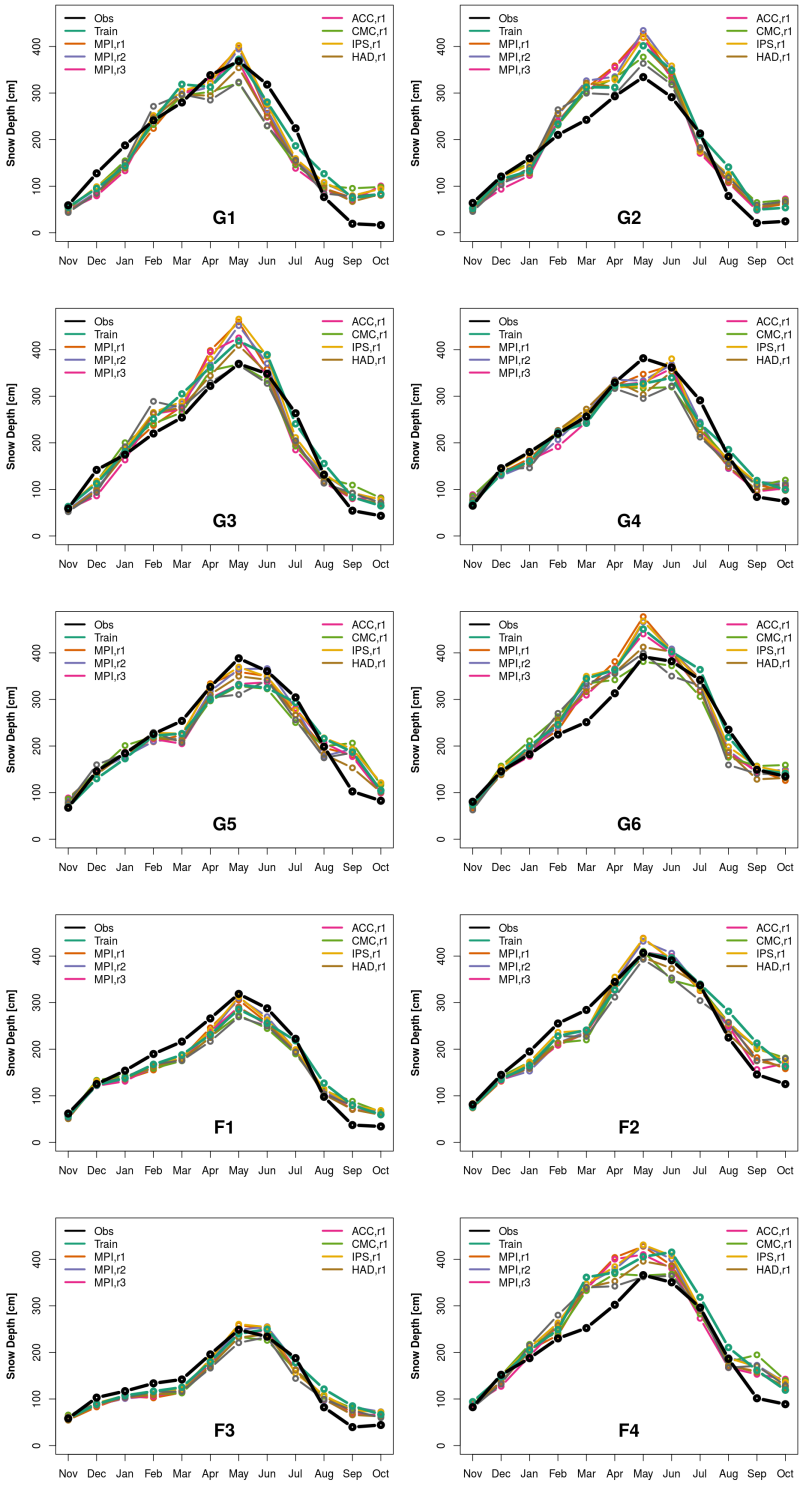


Figure 12.3: Monthly means of the reference period 1971-2000 computed from observations (Obs) as well as simulated model output from the training period (Train, combined calibration and validation) and ESM driven experiments at Sonnblick. In simulations, cross-validation and/or input from different downscaling products (ANN, RCF and NNC) are averaged.

Table 12.4: Statistics of snow depth indices at Zugspitze and Sonnblick considering the observational record (Obs) as well as downscaling products based on the observational record (Train, calibration and validation) and ESM experiments (historical: Hist; Scenarios: RCP 8.5 and RCP 4.5) separated by season and annually. Table includes the annual means (cm), interannual standard deviation (ISD) and number of months with a snow depth of 0 cm (NM0). Results from cross-validation folds and ESMs are averaged. Number in brackets in Hist: standard deviation between ESMs. Number in brackets in scenarios: standard deviation of the change signal. Reference periods: 1971-2000 (Obs, Train, Hist); 2021-2050 (NF); 2071-2100 (DF). Changes: Increasing values are marked in green, declining ones in red and insignificant ones in grey (more than two ESMs show no change signal or a different sign.). Change signals are computed by comparing historical and scenario results. Relative change values: annual means, NM0. Absolute change values: ISD.

	Annual averages							**Interannual standard deviation**							**Number of months with snow depth of 0 cm**						
				RCP 4.5		RCP 8.5					RCP 4.5		RCP 8.5					RCP 4.5		RCP 8.5	
	Obs	Train	Hist	NF	DF	NF	DF	Obs	Train	Hist	NF	DF	NF	DF	Obs	Train	Hist	NF	DF	NF	DF
G1	188	191	181	-13.2 (4.9)	-24.9 (4.8)	-17.2 (5.0)	-45.5 (6.3)	46	19	24	0.3 (4.3)	-1.2 (5.0)	-0.2 (4.9)	-3.0 (5.0)	45	0	0	6.0 (6.2)	20.5 (13.1)	8.6 (7.3)	83.3 (28.8)
G2	172	197	196	-16.7 (4.4)	-29.7 (5.1)	-20.9 (5.4)	-51.1 (6.0)	39	17	21	0.6 (3.9)	-1.9 (4.3)	0.1 (4.3)	-4.8 (3.8)	40	0	0	1.9 (2.6)	10.8 (10.1)	3.7 (4.0)	75.0 (30.8)
G3	199	219	209	-16.4 (5.8)	-30.2 (5.7)	-21.4 (5.7)	-52.4 (6.5)	61	27	28	1.0 (6.1)	-2.3 (6.9)	-0.6 (6.4)	-5.4 (6.1)	27	0	0	1.4 (2.5)	11.2 (10.9)	3.2 (3.9)	74.5 (35.1)
G4	213	206	204	-9.5 (3.2)	-17.9 (3.4)	-12.2 (3.5)	-33.6 (4.6)	57	16	17	0.3 (3.2)	-1.1 (3.8)	-0.1 (4.0)	-1.0 (3.1)	16	0	0	0.1 (0.4)	1.0 (1.5)	0.0 (0.0)	18.5 (15.9)
G5	220	215	217	-2.1 (3.1)	-6.1 (2.8)	-3.7 (2.4)	-15.1 (2.7)	67	24	20	0.0 (3.3)	-0.9 (3.7)	-1.0 (3.8)	-0.3 (2.9)	18	0	0	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.2 (0.7)
G6	236	258	248	-7.2 (3.5)	-14.7 (3.1)	-10.0 (3.2)	-29.0 (4.3)	74	23	25	-0.2 (4.9)	-2.3 (5.3)	-1.4 (5.1)	-4.1 (4.7)	2	0	0	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
F1	167	161	157	-7.4 (2.9)	-14.3 (2.8)	-9.8 (2.8)	-26.4 (3.6)	35	13	11	-0.1 (2.0)	-0.9 (2.3)	-0.4 (2.3)	-1.1 (1.7)	28	0	0	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
F2	244	248	241	-8.8 (3.3)	-16.7 (3.1)	-11.4 (3.3)	-31.0 (4.0)	86	21	22	0.7 (3.5)	-1.9 (4.6)	-0.3 (4.3)	-3.2 (3.1)	11	0	0	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.2 (0.6)
F3	132	135	129	-11.1 (4.0)	-21.0 (3.9)	-14.5 (4.0)	-37.5 (4.8)	50	14	13	-0.1 (2.4)	-1.4 (2.6)	-0.5 (2.5)	-2.1 (2.1)	37	0	0	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	1.3 (2.8)
F4	216	255	246	-14.1 (4.3)	-25.5 (4.7)	-17.8 (4.8)	-46.2 (6.0)	64	20	23	0.7 (4.1)	-1.5 (4.8)	0.4 (5.1)	-2.0 (4.3)	16	0	0	0.2 (0.6)	1.5 (1.9)	0.4 (0.9)	28.7 (23.4)
Z1	214	191	188	-12.0 (4.5)	-23.8 (5.0)	-16.5 (4.1)	-40.6 (5.8)	60	25	28	-2.1 (7.7)	-8.5 (7.0)	-5.2 (7.1)	-10.5 (5.0)	9	0	0	1.0 (1.6)	4.7 (6.3)	1.5 (2.6)	28.6 (21.1)

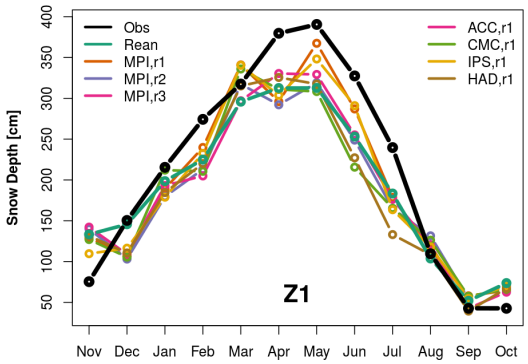


Figure 12.4: Monthly means of the reference period 1971-2000 computed from observations (Obs) as well as simulated model output from the training period (Train, combined calibration and validation) and ESM driven experiments at Zugspitze. In simulations, cross-validation and/or input from different downscaling products (ANN, RCF and NNC) are averaged.

dance of the change signal between the snow depth models driven by different input datasets (ESMs and statistical downscaling products) is high, which is indicated by a low standard deviation among them. The development of snow depth can furthermore affect the regional glaciers. A reduced snow amount in combination with higher temperatures can inhibit the regeneration in winter. Springer et al. (2013) investigated mass balance of the Sonnblick glacier and found clear recessions in the 21st century. The ISD is heavily underestimated between the snow depth models in the training period and during the historical experiments. In the NF, change signals are quite low and come in hand with as well increases as decreases approximately equally frequent. The discordance between stations is less pronounced in the NF of RCP 8.5 where solely G2 and F4 indicate higher variability. Good agreement between the stations is found in the DF, where all predictand series are characterised by a reduced variability. The NM0 is poorly represented within the models during training as well as in the historical experiment. In the typically concerned snow free months, mostly September and October, a certain amount of snow cover remains. This effect seems to be similar to the drizzle error in terms of precipitation downscaling (see e.g. section 8.3). Unfortunately, in the snow depth model results, the amount of remaining snow is too high to be empirically corrected. In scenarios, first NM0 months are detected in the NF, but changes are usually found to be insignificant. In the DF, more and more snow free months are simulated, especially at measuring points at lower altitudes. G1 and G3 show a significant increase in the DF of both scenarios, others only in DF of RCP 8.5. Solely the highest locations of measurement sites are characterised by mostly yearlong snow cover in the scenarios.

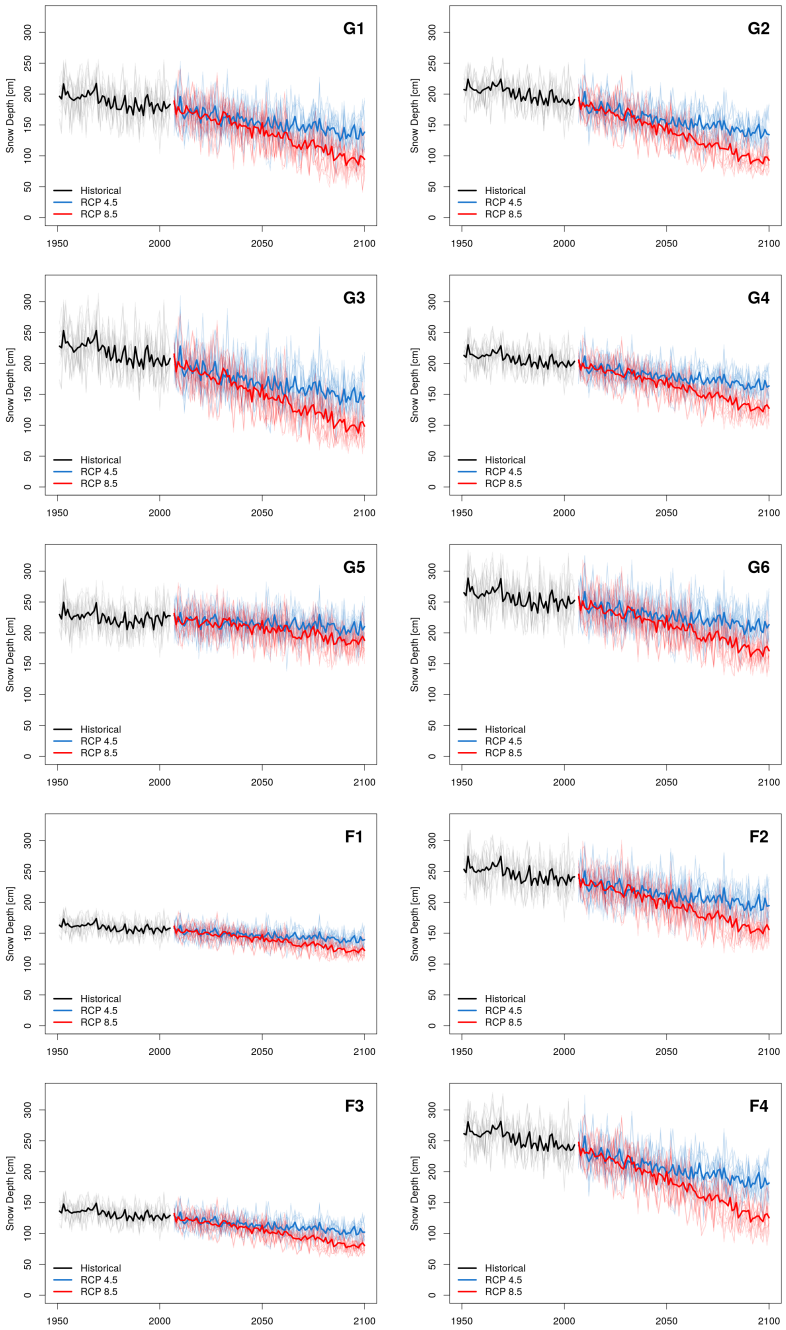


Figure 12.5: Time series of annual snow depths resulting from ANNs driven by statistical downscaling inputs from ANNs, RCFs and NNCs at Sonnblick. Thick lines incorporate the overall experiment mean. Thin lines represent one snow model and SD model combination, whereby cross-validation output is averaged.

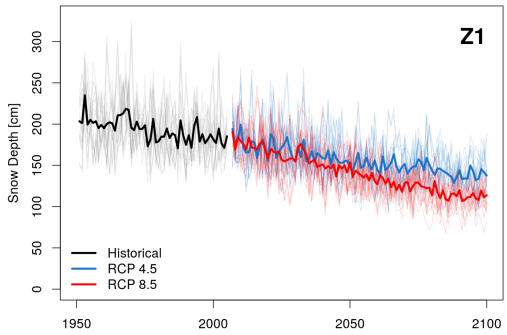


Figure 12.6: Time series of annual snow depths resulting from ANNs driven by statistical downscaling inputs from ANNs, RCFs and NNCs at Zugspitze. Thick lines incorporate the overall experiment mean. Thin lines represent one snow model and SD model combination, whereby cross-validation output is averaged.

12.5 Summary: Snow Depth

All in all, ten observational snow depth time series at Sonnblick and one at Zugspitze are evaluated with focus on the impact of changing climatic conditions. All time series show a pronounced annual cycle with a peak in May and a minimum typically between September and November. All monthly averages in the reference period are characterised by a certain amount of snow cover, although single months can be snow free. The interannual variability is large and can differ up to 38 % from the annual means. If early and late 20th century time periods are compared, long term records of annual averages are characterised by a significant decline.

To analyse the impact of climate change on snow depth, ANNs are calibrated in a low complex framework, with the aim to use the statistical downscaling products from previous chapters as input. Predictors are taken from the precipitation and temperature experiments, as only these variables are available for Sonnblick. Nevertheless, both variables have a high influence on snow depths. Predictors and predictands are harmonised into monthly aggregates. In addition, predictor aggregates are constructed from a maximum of three preceding months. Evaluating the suitability of predictors, usually the larger window sizes prevailed. In this simple setup, mostly small numbers of hidden neurons turned out to be sufficient. An additional experiment with multiple linear regression models was carried out. Although no major differences in model quality are identified in comparison to the ANNs, the focus of this study lies on non-linear methods and therefore ANNs are selected to remain true to this principle.

In the impact study, snow depth models are driven by the statistical downscaling products from ANNs, RCFs and NNCs. The annual snow depth cycle as well as the long term annual average is well reconstructed by the snow depth models at

all measuring sites. In the 21st century mostly significant decreases are projected in both scenarios, whereby stronger decreases are usually found at locations with lower altitude. Considering the latter, in the DF of RCP 8.5, reductions of over $50\ \%$ are projected in some cases. Best agreement in changing interannual variability conditions is simulated in decreasing tendencies in the DF of both scenarios. Snow free months (NM0s) are poorly represented in snow depth simulations. It can be expected that the NM0 signal bears large uncertainties. Most pronounced are increasing NM0s in the DF of RCP 8.5 at lower elevation locations. Although models seem to have trouble with NM0s, taken into account the reduction of annual snow depth, an increase in frequency of snow free months can be expected.

13
Summary and Outlook

Summary

This thesis focused on changing climate condition at the Zugspitze in the northern Alps in southern Germany and the Hoher Sonnblick located in the meridional centre of the Alpine mountain range of Austria. In a statistical downscaling approach, models for meteorological variables of hydrological importance were developed. At both locations, observational records of precipitation and temperature served as statistical downscaling targets, while at Zugspitze further models for relative humidity and wind speed were calibrated. In addition, an impact study on snow depths at both regions was performed.

Methodological focus of this study were statistical models based on non-linear approaches: Artificial Neural Networks (ANN), a reference class forecast procedure (RCF) based on cluster analysis and a novel approach of combining cluster analysis with ANNs (NNC) were evaluated. Statistical models were seasonally calibrated using daily reanalysis and observational time series.

In the context of the statistical downscaling framework, initially the best predictor setup needed to be found. Evaluated in a pre-study were the random grid search procedure, correlation based predictor selection, the automated generation of domains and, an in this context novel approach based on analysing the input sensitivity of ANNs. It turned out to be challenging to get definite predictor setups from the random grid search procedure or correlation. In both methods, the domain size and location was based on assumptions. This gap was intended be filled by an automated domain search procedure. In a reasonable time, the latter can only be performed for each predictor separately, where interconnections between predictors were not considered. In contrast, sensitivity studies of calibrated ANNs take into account both criteria simultaneously: to identify the best predictors and the areas most important to the ANN in an initially large domain. Therefore, the final predictor setup was derived from sensitivity studies. All predictor screening methods led to similar model quality as ANNs seemed to be highly effective in excluding unnecessary input during the training process. Nevertheless, from the investigated methodologies, solely sensitivity studies provided the required information to reduce the number of inputs to the most influential ones and thus leading to a minimum of model parameters.

In the next step, the calibration algorithm and hyperparameters of each statistical downscaling methodology were analysed. ANNs and RCFs can be optimised by different algorithms. In terms of ANNs, nine training algorithms were evaluated. Although most algorithms performed equally well in certain configurations, an extensive grid search of parameters was usually required beforehand. Additionally, the run time of some algorithms was comparably high. A fast and simple-to-configure algorithm was found in the Rprop method, which was henceforward the standard ANN training method. Various ANN topologies were evaluated including different transfer functions, hidden neurons and several hidden layers. The latter were trained by state-of-the-art layer-wise pre-training procedures. However, a simple ANN structure with one hidden layer was found to be sufficient. In terms of RCFs, two cluster analysis algorithms were compared. In this thesis, the DKM algorithm was preferred, as it was significantly faster and usually showed slightly better validation performance than SANDRA in the examined examples. Model performance increased with the number of types, but was limited to 27, as otherwise differences between type centroids were expected to vanish. Classification procedures were performed conditionally by including the target variable into the classification process. The influence of the predictand was controlled by usually very small weights below 0.1 except for the temperature models, where weights can be comparable high. NNCs were constructed considering the findings during the evaluation of ANNs and RCFs. A low number of types was classified by the DKM algorithm to ensure a sufficient amount of training data for the follow up ANN, which consisted of one hidden layer and was trained by the Rprop algorithm. Evaluating the model performance, usually ANNs outperformed RCFs and NNCs in model quality measures of accuracy and relationship.

Future climatic conditions at Zugspitze and Sonnblick were investigated by transferring the calibrated models onto ESM scenarios. Historical datasets and the scenarios RCP 4.5 and RCP 8.5 of seven ESM versions were used as SD model drivers. Changing conditions were identified comparing the historical reference period 1971-2000 with the near (2021-2050) and distant (2071-2100) future. In terms of precipitation, the positive trend, found especially in long-terms of the observational time series in the second half of the 20th century, was continued in the NF and DF of both scenarios at Zugspitze and Sonnblick. The highest precipitation increase in the combined view of all SD models was expected in winter, carrying the annual mean, as precipitation in summer is, in particular in the DF, suggested to decrease. Best coincidence between the SD models was witnessed considering temperature at both stations. With low differences, all SD models simulated successively increasing temperatures from NF to DF in all seasons, with most intense warming in summer and winter. Simultaneously, relative humidity at Zugspitze was found to continuously

decrease, most likely depending on the warmer temperatures and the associated increase in water absorption capacity of the air, but limited available water resources to evaporate water vapour. A low trend in wind speed at Zugspitze was projected by the SD models, showing a significant increase not before the DF of both scenarios. Increases in winter and spring compensate annual decreases in summer and autumn.

In the climate change analysis, RCFs usually showed good results when transferred from reanalysis to ESM datasets in the historical period. In future scenarios under changing conditions causing a shift in the value distribution, RCFs oftentimes pointed to different directions in comparison to ANNs and NNCs or solely showed no trend. In addition, in the comparison of simulated and observed quantiles in the historical period, RCFs were often not able to represent the extremes of unlimited upper or lower boundaries. Most likely, the reason was the fixed reference value range, making the RCF results questionable in the context of climate change or extremes value analysis.

Finally, an impact study on monthly snow depths at Zugspitze and ten measurement sites at Sonnblick was performed using future projections of temperature and precipitation of the statistical downscaling approach. Predictors were aggregated in different window sizes including up to three previous months. ANNs were calibrated with Rprop and an optimised number of hidden neurons in one hidden layer to capture the relationship between predictors and snow depth. In this application, the performance results of the ANNs were usually quite low. In the annual average, partly drastically snow depth reductions were expected, especially at low level stations in the DF of RCP 8.5.

To summarise the main scientific issues:

- Future climatic conditions were evaluated for the Alpine summits of Zugspitze and Hoher Sonnblick.

- The non-linear statistical modelling techniques showed high model skills during the training and evaluation process. No major problems or unexpected behaviour occurred within the framework.

- A well-suited method in deriving good predictor combinations and an associated domain and location was found by studying the sensitivity of ANNs.

- Unfortunately, the novel NNC approach was typically outperformed by standard ANNs. A reason could have been the reduced amount of data available to train the follow up ANNs in addition with the in general high complexity of the NNC construction process and hyperparameter tuning.

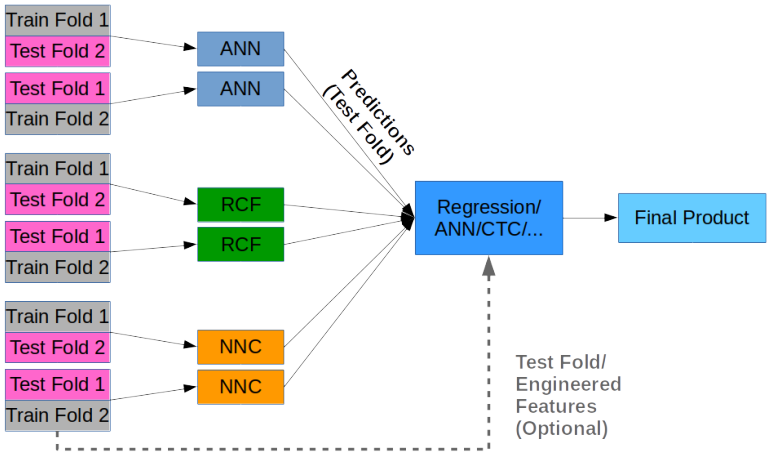


Figure 13.1: Schematics of a stacked model approach based on the statistical downscaling framework of this thesis.

Outlook

The model framework and results provide various possible starting points for further scientific research. A selection will be listed in the following.

- The resulting time series of the statistical downscaling framework are available at the University of Augsburg and can be applied in various follow up studies. One example of planned usage is the application in the Cold Regions Hydrological Model (CRHM) by VAO project cooperation partners (see M. Weber et al. (2017) for a preliminary study).
- The developed R software package of this study can be derived from the Institute of Geography at the University of Augsburg. The package offers a comprehensive library on tools for non-linear and linear analysis as well as data pre- and post-processing. It will be used in further studies at the Geographical department of the University of Augsburg, e.g. by analysing particulate matter.
- At the time of this study only seven ESM runs could be derived from CMIP5 providing the required predictors on a daily resolution. To capture a large spectrum of future projections and thus analyse ESM uncertainties, a large ensemble is essential. If more climate models or newer versions of the latter are available, the developed statistical downscaling framework can be easily used to extend the ensemble of future projections.
- A handy feature of ANNs is that only parts of the predictor set can be used to update the net weights in one iteration in stochastic or minibatch training.

ANNs in this study are optimised in batch mode, which uses all provided predictor time steps for one weight update. The best fit of the ANN simulations will most likely be in the value range, in which the distribution of predictand values has its maximum. To put additional weight on rarely occurring events, the optimisation algorithm can be adjusted in stochastic and minibatch training to select pattern at the minimum of the predictand distribution more frequently. The resulting ANN might show an improved behaviour in these parts of the predictand distribution, which can be beneficial in studies on extremes.

- A recently well established approach in machine learning is to build one "strong" model from several "weak" models. First application examples in natural science are published, e.g. Zhai and Chen (2018). Figure 13.1 visualises the schematics of a possible implementation considering the model framework of this thesis. For the sake of simplicity, a reduced amount of cross-validation folds is displayed. Initially, the available dataset is split into training and test sets to calibrate an ensemble of weak models, here ANNs, RCFs and NNCs. The models produced within this study already represent a good basis. In addition, models with other hyperparameter configurations can be beneficial, for example different number of hidden neurons, hidden layers, number of circulation types or models calibrated by different optimisation algorithms. In this approach, one model by itself does not need to show the highest possible performance. In a second step, a stacked ensemble model is calibrated. The aim of the stacked model is to identify strengths and weaknesses within each weak model prediction and to choose the best weak model under a certain given condition. When simulating extremes, for example, the drawbacks of RCFs can be compensated by focusing on ANNs or NNCs. Stacking can be performed by various regression or classification based methods depending on the task. The input of the stacked model can include additional features of the original predictors. Further, the stacking process can be performed in several stages, e.g first the cross-validation ensemble of ANNs, RCFs and NNCs are combined separately, before in a second step the final model is built.

Bibliography

Adamowski J. and Karapataki C. (2010): Comparison of multivariate regression and artificial neural networks for peak urban water-demand forecasting: evaluation of different ANN learning algorithms. *Journal of Hydrologic Engineering* 15.10, pp. 729–743.

Agency E. E. (2009): Regional climate change and adaptation. The Alps facing the challenge of changing water resources. *EEA Report No 8* 8, p. 143.

– (2016): Climate change, impacts and vulnerability in Europe 2016. Tech. rep. European Environment Agency.

Agrawala S. et al. (2007): Climate change in the European Alps: adapting winter tourism and natural hazards management. Organisation for Economic Cooperation and Development (OECD).

Ahmad I., Ansari M., and Mohsin S. (2008): Performance comparison between backpropagation algorithms applied to intrusion detection in computer network systems. *Proceedings of the 7th WSEAS International Conference on Applied Computer and Applied Computational Science as ACM guide*, pp. 47–52.

Alexander M. A., Bladé I., Newman M., Lanzante J. R., Lau N.-C., and Scott J. D. (2002): The atmospheric bridge: The influence of ENSO teleconnections on air–sea interaction over the global oceans. *Journal of Climate* 15.16, pp. 2205–2231.

Almeida C., Baugh C., Lacey C., Frenk C., Granato G., Silva L., and Bressan A. (2010): Modelling the dusty universe–I. Introducing the artificial neural network and first applications to luminosity and colour distributions. *Monthly Notices of the Royal Astronomical Society* 402.1, pp. 544–564.

Arlot S. and Celisse A. (2010): A survey of cross-validation procedures for model selection. *Statistics surveys* 4, pp. 40–79.

Auer I., Böhm R., Hiebl J., Schöner W., Spinoni J., Lentini G., Brunetti M., Percec Tadic M., Dolinar M., Bissolli P., et al. (2008): High Resolution Temperature Climatology in Complex Terrain: demonstrated in the test area Greater Alpine Region GAR: realisation and Presentation of the Alpine Temperature Climatology. *Meeting ECSN-EAC*.

Auer I., Böhm R., Jurkovic A., Lipa W., Orlik A., Potzmann R., Schöner W., Ungersböck M., Matulla C., Briffa K., Jones P., Efthymiadis D., Brunetti M., Nanni T., Maugeri M., Mercalli L., Mestre O., Moisselin J.-M., Begert M., Muller-Westermeier G., Kveton V., Bochnicek O., Stastny P., Lapin M., Szalai S., Szentimrey T., Cegnar T., Dolinar M., Gajic-Capka M., Zaninovic K., Majstorovic Z., and Nieplova E. (2007): HISTALP-historical instrumental climatological surface time series of the Greater Alpine Region. *International Journal of Climatology* 27.1, pp. 17–46.

Azadeh A., Sheikhalishahi M., Tabesh M., and Negahban A. (2011): The effects of pre-processing methods on forecasting improvement of artificial neural networks. *Australian Journal of Basic and Applied Sciences* 5.6, pp. 570–580.

Azami H., Sanei S., and Mohammadi K. (2011): Improving the neural network training for face recognition using adaptive learning rate, resilient back propagation and conjugate gradient algorithm. *Journal of Computer Applications* 34.2, pp. 22–26.

Bahrenberg G., Giese E., and Nipper J. (1999): Statistische Methoden in der Geographie 1. Vol. 1.

Bahrenberg G., Giese E., and Nipper J. (1992): Statistische Methoden in der Geographie. Vol. 2. Teubner Stuttgart.

Bardossy A., Duckstein L., and Bogardi I. (1995): Fuzzy rule-based classification of atmospheric circulation patterns. *International journal of climatology* 15.10, pp. 1087–1097.

Barnes E. A. and Polvani L. (2013): Response of the midlatitude jets, and of their variability, to increased greenhouse gases in the CMIP5 models. *Journal of Climate* 26.18, pp. 7117–7135.

Barry R. G. (2008): Mountain weather and climate.

Beck C. and Philipp A. (2010): Evaluation and comparison of circulation type classifications for the European domain. *Physics and Chemistry of the Earth, Parts A/B/C* 35.9, pp. 374–387.

Beck C., Philipp A., and Streicher F. (2013): The effect of domain size on the relationship between circulation type classifications and surface climate. *International Journal of Climatology*.

Beck C., Weitnauer C., and Jacobeit J. (2014): Downscaling of monthly PM10 indices at different sites in Bavaria (Germany) based on circulation type classifications. *Atmospheric Pollution Research* 5.4, pp. 741–752.

Benestad R. E., Hanssen-Bauer I., and Chen D. (2008): Empirical-statistical downscaling. World Scientific Publishing Company.

Bengio Y. (2012): Practical recommendations for gradient-based training of deep architectures. *Neural networks: Tricks of the trade*. Springer, pp. 437–478.

Beniston M. (2005): Mountain climates and climatic change: an overview of processes focusing on the European Alps. *Pure and Applied Geophysics* 162.8, pp. 1587–1606.

– (2006): Mountain weather and climate: a general overview and a focus on climatic change in the Alps. *Hydrobiologia* 562.1, pp. 3–16.

– (2016): Environmental change in mountains and uplands. Routledge.

Beniston M. and Jungo P. (2002): Shifts in the distributions of pressure, temperature and moisture and changes in the typical weather patterns in the Alpine region in response to the behavior of the North Atlantic Oscillation. *Theoretical and Applied Climatology* 71.1, pp. 29–42.

Beniston M., Keller F., Koffi B., and Goyette S. (2003): Estimates of snow accumulation and volume in the Swiss Alps under changing climatic conditions. *Theoretical and Applied Climatology* 76.3, pp. 125–140.

Blumer A., Ehrenfeucht A., Haussler D., and Warmuth M. K. (1987): Occam's razor. *Information processing letters* 24.6, pp. 377–380.

Bott A. (2016): Synoptische Meteorologie: Methoden der Wetteranalyse und-prognose. Springer-Verlag.

Brassington G. (2017): Mean absolute error and root mean square error: which is the better metric for assessing model performance? *EGU General Assembly Conference Abstracts*. Vol. 19, p. 3574.

Brönnimann S., Xoplaki E., Casty C., Pauling A., and Luterbacher J. (2007): ENSO influence on Europe during the last centuries. *Climate Dynamics* 28.2-3, pp. 181–197.

Brunetti M., Maugeri M., Nanni T., Auer I., Böhm R., and Schöner W. (2006): Precipitation variability and changes in the greater Alpine region over the 1800–2003 period. *Journal of Geophysical Research: Atmospheres* 111.D11.

Buonomo E., Jones R., Huntingford C., and Hannaford J. (2007): On the robustness of changes in extreme precipitation over Europe from two high resolution climate change simulations. *Quarterly Journal of the Royal Meteorological Society* 133.622, pp. 65–81.

Burney S. M. A., Jilani T. A., and Ardil C. (2004): A Comparison of First and Second Order Training Algorithms for Artificial Neural Networks. *International Conference on Computational Intelligence*, pp. 12–18.

Carreau J. and Vrac M. (2011): Stochastic downscaling of precipitation with neural network conditional mixture models. *Water Resources Research* 47.10.

Caruana R., Lawrence S., and Giles C. L. (2001): Overfitting in neural nets: Backpropagation, conjugate gradient, and early stopping. *Advances in neural information processing systems*, pp. 402–408.

Castellanos J., Pazos A., Rios J., and Zafra J. (1994): Sensitivity analysis on neural networks for meteorological variable forecasting. *Neural Networks for Signal Processing [1994] IV. Proceedings of the 1994 IEEE Workshop*. IEEE, pp. 587–595.

Casty C., Wanner H., Luterbacher J., Esper J., and Böhm R. (2005): Temperature and precipitation variability in the European Alps since 1500. *International Journal of Climatology* 25.14, pp. 1855–1880.

Cavazos T. and Hewitson B. (2002): Relative performance of empirical predictors of daily precipitation.

Cavazos T. and Hewitson B. C. (2005): Performance of NCEP-NCAR reanalysis variables in statistical downscaling of daily precipitation. *Clim. Res* 28.2, pp. 95–107.

Cebon P. (1998): Views from the Alps: Regional perspectives on climate change. MIT Press.

Chadwick R., Coppola E., and Giorgi F. (2011): An artificial neural network technique for downscaling GCM outputs to RCM spatial scale. *Nonlinear Processes in Geophysics* 18.6.

Chai T. and Draxler R. R. (2014): Root mean square error (RMSE) or mean absolute error (MAE)?–Arguments against avoiding RMSE in the literature. *Geoscientific model development* 7.3, pp. 1247–1250.

Christensen J. H. and Christensen O. B. (2007): A summary of the PRUDENCE model projections of changes in European climate by the end of this century. *Climatic change* 81.1, pp. 7–30.

Compo G. P., Whitaker J. S., Sardeshmukh P. D., Matsui N., Allan R. J., Yin X., Gleason B. E., Vose R., Rutledge G., Bessemoulin P., et al. (2011): The twentieth century reanalysis project. *Quarterly Journal of the Royal Meteorological Society* 137.654, pp. 1–28.

Conover W. J., Johnson M. E., and Johnson M. M. (1981): A comparative study of tests for homogeneity of variances, with applications to the outer continental shelf bidding data. *Technometrics* 23.4, pp. 351–361.

Coulibaly P., Dibike Y. B., and Anctil F. (2005): Downscaling precipitation and temperature with temporal neural networks. *Journal of Hydrometeorology* 6.4, pp. 483–496.

Cubasch U. and Kasang D. (2000): Anthropogener Klimawandel. Klett-Perthes.

D'Arrigo R., Cook E. R., Wilson R. J., Allan R., and Mann M. E. (2005): On the variability of ENSO over the past six centuries. *Geophysical Research Letters* 32.3.

Dai A. (2006): Precipitation characteristics in eighteen coupled climate models. *Journal of Climate* 19.18, pp. 4605–4630.

Davidson B., Gerbier S., Papagionakis S., and Rijkoort P. (1964): Sites for wind power installations. *WMO Technical Note* 63.

Davy R. J., Woods M. J., Russell C. J., and Coppin P. A. (2010): Statistical downscaling of wind variability from meteorological fields. *Boundary-layer meteorology* 135.1, pp. 161–175.

Dee D. P., Uppala S., Simmons A., Berrisford P., Poli P., Kobayashi S., Andrae U., Balmaseda M., Balsamo G., Bauer P., et al. (2011): The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the royal meteorological society* 137.656, pp. 553–597.

Demuth H. B., Beale M. H., De Jess O., and Hagan M. T. (2014): Neural network design (2nd Edition). Martin Hagan.

Demuzere M., Kassomenos P., and Philipp A. (2011): The COST733 circulation type classification software: an example for surface ozone concentrations in Central Europe. *Theoretical and applied climatology* 105.1-2, pp. 143–166.

Dobler C., Bürger G., and Stötter J. (2013): Simulating future precipitation extremes in a complex Alpine catchment. *Natural Hazards and Earth System Sciences* 13.2, pp. 263–277.

Dreiseitl S. and Ohno-Machado L. (2002): Logistic regression and artificial neural network classification models: a methodology review. *Journal of biomedical informatics* 35.5-6, pp. 352–359.

Duan K. and Mei Y. (2014): A comparison study of three statistical downscaling methods and their model-averaging ensemble for precipitation downscaling in China. *Theoretical and applied climatology* 116.3-4, pp. 707–719.

Durand Y., Giraud G., Laternser M., Etchevers P., Mérindol L., and Lesaffre B. (2009): Reanalysis of 47 years of climate in the French Alps (1958–2005): climatology and trends for snow cover. *Journal of applied meteorology and climatology* 48.12, pp. 2487–2512.

Efthymiadis D., Jones P. D., Briffa K. R., Böhm R., and Maugeri M. (2007): Influence of large-scale atmospheric circulation on climate variability in the Greater Alpine Region of Europe. *Journal of Geophysical Research: Atmospheres* 112.D12.

Enke W. and Spekat A. (1997): Downscaling climate model outputs into local and regional weather elements by classification and regression. *Climate Research*, pp. 195–207.

EPA (2014): Climate change indicators in the United States. Tech. rep. EPA 430-R-14-004. Environmental Protection Agency.

ESRL (2018): National Oceanic and Atmospheric Administration, Earth System Research Laboratory. https://www.esrl.noaa.gov/.

Fernandes P. O., Teixeira J. P., Ferreira J., and Azevedo S. (2013): Training neural networks by resilient Backpropagation algorithm for tourism forecasting. *Management Intelligent Systems*. Springer, pp. 41–49.

Fierz C., Armstrong R. L., Durand Y., Etchevers P., Greene E., McClung D. M., Nishimura K., Satyawali P. K., and Sokratov S. A. (2009): The international classification for seasonal snow on the ground. Vol. 25. UNESCO/IHP Paris.

Fraedrich K. and Müller K. (1992): Climate anomalies in Europe associated with ENSO extremes. *International Journal of Climatology* 12.1, pp. 25–31.

Francis R. C. (2006): Measuring the strength of environment–recruitment relationships: the importance of including predictor screening within cross-validations. *ICES Journal of Marine Science* 63.4, pp. 594–599.

Frei C., Christensen J. H., Déqué M., Jacob D., Jones R. G., and Vidale P. L. (2003): Daily precipitation statistics in regional climate models: Evaluation and

intercomparison for the European Alps. *Journal of Geophysical Research: Atmospheres* 108.D3.

Frei C. and Schär C. (1998): A precipitation climatology of the Alps from high-resolution rain-gauge observations. *International Journal of climatology* 18.8, pp. 873–900.

Frei C., Schöll R., Fukutome S., Schmidli J., and Vidale P. L. (2006): Future change of precipitation extremes in Europe: Intercomparison of scenarios from regional climate models. *Journal of Geophysical Research: Atmospheres* 111.D6.

Friederichs P. (2010): Statistical downscaling of extreme precipitation events using extreme value theory. *Extremes* 13.2, pp. 109–132.

Frochte J. (2019): Maschinelles Lernen: Grundlagen und Algorithmen in Python. Carl Hanser Verlag GmbH Co KG.

Gaitan C. F., Hsieh W. W., and Cannon A. J. (2014): Comparison of statistically downscaled precipitation in terms of future climate indices and daily variability for southern Ontario and Quebec, Canada. *Climate Dynamics* 43.12, pp. 3201–3217.

Garson G. D. (1991): Interpreting neural-network connection weights. *AI expert* 6.4, pp. 46–51.

Geisser S. (1975): The predictive sample reuse method with applications. *Journal of the American statistical Association* 70.350, pp. 320–328.

Gevrey M., Dimopoulos I., and Lek S. (2003): Review and comparison of methods to study the contribution of variables in artificial neural network models. *Ecological modelling* 160.3, pp. 249–264.

Giorgi F. and Mearns L. O. (1999): Introduction to special section: Regional climate modeling revisited. *Journal of Geophysical Research: Atmospheres* 104.D6, pp. 6335–6352.

Glorot X. and Bengio Y. (2010): Understanding the difficulty of training deep feedforward neural networks. *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, pp. 249–256.

Gobiet A., Kotlarski S., Beniston M., Heinrich G., Rajczak J., and Stoffel M. (2014): 21st century climate change in the European Alps - a review. *Science of the Total Environment* 493, pp. 1138–1151.

Goddard L., Kumar A., Solomon A., Smith D., Boer G., Gonzalez P., Kharin V., Merryfield W., Deser C., Mason S. J., et al. (2013): A verification framework for interannual-to-decadal predictions experiments. *Climate Dynamics* 40.1-2, pp. 245–272.

Goodess C. and Jones P. (2002): Links between circulation and changes in the characteristics of Iberian rainfall. *International Journal of Climatology* 22.13, pp. 1593–1615.

Goodfellow I., Bengio Y., Courville A., and Bengio Y. (2016): Deep learning. Vol. 1. MIT press Cambridge.

Greatbatch R., Lu J., and Peterson K. (2004): Non-stationary impact of ENSO on European winter climate. *Geophys Res Lett* 31.

Grotch S. L. and MacCracken M. C. (1991): The use of general circulation models to predict regional climatic change. *Journal of Climate* 4.3, pp. 286–303.

Günther F. and Fritsch S. (2010): neuralnet: Training of neural networks. *The R journal* 2.1, pp. 30–38.

Gutmann E. D., Rasmussen R. M., Liu C., Ikeda K., Gochis D. J., Clark M. P., Dudhia J., and Thompson G. (2012): A comparison of statistical and dynamical downscaling of winter precipitation over complex terrain. *Journal of Climate* 25.1, pp. 262–281.

Han F., Yao Y., Dai S., Wang C., Sun R., Xu J., and Zhang B. (2012): Mass elevation effect and its forcing on timberline altitude. *Journal of geographical Sciences* 22.4, pp. 609–616.

Hantel M. and Maurer C. (2011): The median winter snowline in the Alps. *Meteorologische Zeitschrift* 20.3, pp. 267–276.

Harpham C. and Wilby R. L. (2005): Multi-site downscaling of heavy daily precipitation occurrence and amounts. *Journal of Hydrology* 312.1-4, pp. 235–255.

Hartigan J. A. and Wong M. A. (1979): Algorithm AS 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 28.1, pp. 100–108.

Haykin S. S., Haykin S. S., Haykin S. S., and Haykin S. S. (2009): Neural networks and learning machines. Vol. 3. Pearson Upper Saddle River, NJ, USA:

Hecht-Nielsen R. (1992): Theory of the backpropagation neural network. *Neural networks for perception*. Elsevier, pp. 65–93.

Hertig E., Seubert S., Paxian A., Vogt G., Paeth H., and Jacobeit J. (2014): Statistical modelling of extreme precipitation indices for the Mediterranean area under future climate change. *International Journal of Climatology* 34.4, pp. 1132–1156.

Hertig E. and Jacobeit J. (2008): Assessments of Mediterranean precipitation changes for the 21st century using statistical downscaling techniques. *International Journal of Climatology* 28.8, pp. 1025–1045.

– (2013): A novel approach to statistical downscaling considering nonstationarities: application to daily precipitation in the Mediterranean area. *Journal of Geophysical Research: Atmospheres* 118.2, pp. 520–533.

Hinton G. E., Osindero S., and Teh Y.-W. (2006): A fast learning algorithm for deep belief nets. *Neural computation* 18.7, pp. 1527–1554.

Hofer M., Marzeion B., and Mölg T. (2015): A statistical downscaling method for daily air temperature in data-sparse, glaciated mountain environments. *Geoscientific Model Development* 8.3, pp. 579–593.

Hofstätter M., Beck C., Chimani B., Ganekind M., Homan M., Jacobeit J., and Phillip A. (2013): WETRAX: WEather Patterns, Cyclone TRAcks and related precipitation EXtremes. *EGU General Assembly Conference Abstracts*. Vol. 15.

Hornik K., Stinchcombe M., and White H. (1989): Multilayer feedforward networks are universal approximators. *Neural networks* 2.5, pp. 359–366.

Hunter A., Kennedy L., Henry J., and Ferguson I. (2000): Application of neural networks and sensitivity analysis to improved prediction of trauma survival. *Computer methods and programs in biomedicine* 62.1, pp. 11–19.

Hurrell J. W. et al. (1995): Decadal trends in the North Atlantic Oscillation: regional temperatures and precipitation. *Science-AAAS-Weekly Paper Edition* 269.5224, pp. 676–678.

Huss M., Hock R., Bauder A., and Funk M. (2010): 100-year mass changes in the Swiss Alps linked to the Atlantic Multidecadal Oscillation. *Geophysical Research Letters* 37.10.

Huth R., Beck C., Philipp A., Demuzere M., Ustrnul Z., Cahynová M., Kysely J., and Tveito O. E. (2008a): Classifications of atmospheric circulation patterns. *Annals of the New York Academy of Sciences* 1146.1, pp. 105–152.

Huth R., Kliegrová S., and Metelka L. (2008b): Non-linearity in statistical downscaling: does it bring an improvement for daily temperature in Europe? *International Journal of Climatology* 28.4, pp. 465–477.

Huth R., Mikšovský J., Štěpánek P., Belda M., Farda A., Chládová Z., and Pišoft P. (2014): Comparative validation of statistical and dynamical downscaling models on a dense grid in central Europe: temperature. *Theoretical and Applied Climatology*, pp. 1–21.

IPCC (2007): Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Ed. by H. Le Treut, R. Somerville, U. Cubasch, Y. Ding, C. Mauritzen, A. Mokssit, T. Peterson, and M. Prather. Cambridge University Press.

– (2013): Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Ed. by T. Stocker, D. Qin, G.-K. Plattner, M. M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. M. Midgley. Cambridge University Press.

Isotta F. A., Frei C., Weilguni V., Perčec Tadić M., Lassegues P., Rudolf B., Pavan V., Cacciamani C., Antolini G., Ratto S. M., Munari M., Micheletti S., Bonati V., Lussana C., Ronchi C., Panettieri E., Marigo G., and Vertačnik G. (2014a): The climate of daily precipitation in the Alps: development and analysis of a high-resolution grid dataset from pan-Alpine rain-gauge data. *International Journal of Climatology* 34.5, pp. 1657–1675.

Isotta F. A., Vogel R., and Frei C. (2014b): Evaluation of European regional reanalyses and downscalings for precipitation in the Alpine region. *Meteorol Z. doi* 10, p. 1127.

Jacob D., Petersen J., Eggert B., Alias A., Christensen O. B., Bouwer L. M., Braun A., Colette A., Déqué M., Georgievski G., et al. (2014): EURO-CORDEX: new high-resolution climate change projections for European impact research. *Regional environmental change* 14.2, pp. 563–578.

Jacobeit J. (2007): Zusammenhänge und Wechselwirkungen im Klimasystem.

– (2010): Classifications in climate research. *Physics and Chemistry of the Earth, Parts A/B/C* 35.9, pp. 411–421.

Jacobs R. A. (1988): Increased rates of convergence through learning rate adaptation. *Neural networks* 1.4, pp. 295–307.

Jarosch A. H., Anslow F. S., and Clarke G. K. (2012): High-resolution precipitation and temperature downscaling for glacier models. *Climate dynamics* 38.1-2, pp. 391–409.

Jones P. W. (1999): First-and second-order conservative remapping schemes for grids in spherical coordinates. *Monthly Weather Review* 127.9, pp. 2204–2210.

Jones R., Murphy J., and Noguer M. (1995): Simulation of climate change over europe using a nested regional-climate model. I: Assessment of control climate, including sensitivity to location of lateral boundaries. *Quarterly journal of the Royal meteorological society* 121.526, pp. 1413–1449.

Joshi M. M., Gregory J. M., Webb M. J., Sexton D. M., and Johns T. C. (2008): Mechanisms for the land/sea warming contrast exhibited by simulations of climate change. *Climate Dynamics* 30.5, pp. 455–465.

Kaspar-Ott I., Hertig E., Severin K., Pollinger F., Ring C., Paeth H., and Jacobeit J. (2019): Weights for general circulation models from CMIP3 / CMIP5 in a statistical downscaling framework and the impact on future Mediterranean precipitation. *International Journal of Climatology*.

Kirchhofer W. (1974): Classification of European 500 mb patterns. Tech. rep. Swiss Meteorological Institute.

Knight J. R., Folland C. K., and Scaife A. A. (2006): Climate impacts of the Atlantic multidecadal oscillation. *Geophysical Research Letters* 33.17.

Knudsen M. F., Seidenkrantz M.-S., Jacobsen B. H., and Kuijpers A. (2011): Tracking the Atlantic Multidecadal Oscillation through the last 8,000 years. *Nature communications* 2, p. 178.

Kruse R., Borgelt C., Braune C., Mostaghim S., and Steinbrecher M. (2016): Computational intelligence: a methodological introduction. Springer.

Kysely J. (2007): Implications of enhanced persistence of atmospheric circulation for the occurrence and severity of temperature extremes. *International Journal of Climatology* 27.5, pp. 689–695.

Laternser M. and Schneebeli M. (2003): Long-term snow climate trends of the Swiss Alps (1931–99). *international Journal of climatology* 23.7, pp. 733–750.

Latif M. (2009): Klimawandel und Klimadynamik. UTB Ulmer.

Lauer W. and Bendix J. (2006): Klimatologie–Das Geographische Seminar.

LeCun Y., Bottou L., Orr G. B., and Müller K.-R. (1998): Efficient backprop. *Neural networks: Tricks of the trade*. Springer, pp. 9–50.

Li T. and Hsu P.-c. (2018): Fundamentals of Tropical Climate Dynamics. Springer.

Lippe W.-M. (2006): Soft-computing: Mit neuronalen Netzen, Fuzzy-logic und evolutionären Algorithmen. Springer-Verlag.

Lutz K., Jacobeit J., Philipp A., Seubert S., Kunstmann H., and Laux P. (2012): Comparison and evaluation of statistical downscaling techniques for station-based precipitation in the Middle East. *International Journal of Climatology* 32.10, pp. 1579–1595.

Mann H. B. and Whitney D. R. (1947): On a test of whether one of two random variables is stochastically larger than the other. *The annals of mathematical statistics*, pp. 50–60.

Maraun D., Huth R., Gutiérrez J. M., Martin Benistonn D. S., Dubrovsky M., Fischer A., Hertig E., Soares P. M., Bartholy J., Pongrácz R., et al. (2017): The VALUE perfect predictor experiment: evaluation of temporal variability. *International Journal of Climatology*.

Maraun D., Wetterhall F., Ireson A., Chandler R., Kendon E., Widmann M., Brienen S., Rust H., Sauter T., Themeßl M., et al. (2010): Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user. *Reviews of Geophysics* 48.3.

Marengo J. A., Jones R., Alves L. M., and Valverde M. C. (2009): Future change of temperature and precipitation extremes in South America as derived from the PRECIS regional climate modeling system. *International Journal of Climatology: A Journal of the Royal Meteorological Society* 29.15, pp. 2241–2255.

Marin J.-M., Mengersen K., and Robert C. P. (2005): Bayesian modelling and inference on mixtures of distributions. *Handbook of statistics* 25, pp. 459–507.

Mariotti A., Zeng N., and Lau K.-M. (2002): Euro-Mediterranean rainfall and ENSO-a seasonally varying relationship. *Geophysical research letters* 29.12.

Matulla C., Auer I., Böhm R., Ungersböck M., and Schöner (2005): Outstanding past decadal-scale climate events in the Greater Alpine Region analysed by 250 years data and model runs. *GKSS-Forschungszentrum*.

Menzel A., Jakobi G., Ahas R., Scheifinger H., and Estrella N. (2003): Variations of the climatological growing season (1951–2000) in Germany compared with other countries. *International Journal of Climatology* 23.7, pp. 793–812.

Merkenschlager C. (2017): Abschätzung zukünfiger saisonaler Niederschlagsextreme im Mittelmeerraum unter Berücksichtigung variierender Prädiktor-Prädiktand-Beziehungen. PhD thesis. University of Augsburg.

Merz R., Parajka J., and Blöschl G. (2011): Time stability of catchment model parameters: Implications for climate impact analyses. *Water Resources Research* 47.2.

Minai A. A. and Williams R. D. (1990): Back-propagation heuristics: a study of the extended delta-bar-delta algorithm. *Neural Networks, 1990., 1990 IJCNN International Joint Conference on*. IEEE, pp. 595–600.

Minsky M. and Papert S. (1969): Perceptrons: Anlntroduction to Computational Geometry.

Montano J. and Palmer A. (2003): Numeric sensitivity analysis applied to feedforward neural networks. *Neural Computing & Applications* 12.2, pp. 119–125.

Moss R. H., Edmonds J. A., Hibbard K. A., Manning M. R., Rose S. K., Van Vuuren D. P., Carter T. R., Emori S., Kainuma M., Kram T., et al. (2010): The next generation of scenarios for climate change research and assessment. *Nature* 463.7282, pp. 747–756.

Mott R., Faure F., Michael L., Löwe H., Hynek B., Michlmayer G., Prokop A., and Schöner W. (2008): Simulation of seasonal snow-cover distribution for glacierized sites on Sonnblick, Austria, with the Alpine3D model. *Annals of Glaciology* 49, pp. 155–160.

Moy C. M., Seltzer G. O., Rodbell D. T., and Anderson D. M. (2002): Variability of El Niño/Southern Oscillation activity at millennial timescales during the Holocene epoch. *Nature* 420.6912, pp. 162–165.

Moya Quiroga V., Mano A., Asaoka Y., Kure S., Udo K., and Mendoza J. (2013): Snow glacier melt estimation in tropical Andean glaciers using artificial neural networks. *Hydrology and Earth System Sciences* 17.4, pp. 1265–1280.

Müller F., Ohmura A., Schroff K., Funk M., Pfirter K., Bernath A., and Steffen K. (1980): Combined ice, water and energy balances of a glacierized basin of the Swiss Alps–the Rhonegletscher Project. *Geogr. Helvet* 35.5, pp. 57–69.

Müller M. and Kaspar M. (2011): Association between anomalies of moisture flux and extreme runoff events in the south-eastern Alps. *Natural Hazards and Earth System Sciences* 11.3, p. 915.

Nawi N. M., Atomi W. H., and Rehman M. (2013): The effect of data pre-processing on optimized training of artificial neural networks. *Procedia Technology* 11, pp. 32–39.

NOAA (2013): Snow Measurement Guidelines for National Weather Service Surface Observing Programs. U.S. Department of Commerce National Oceanic and Atmospheric Administration. Office of Climate, Water and Weather Services, Silver Spring, MD 20910.

Nourani V. and Fard M. S. (2012): Sensitivity analysis of the artificial neural network outputs in simulation of the evaporation process at different climatologic regimes. *Advances in Engineering Software* 47.1, pp. 127–146.

O'Gorman P. and Muller C. (2010): How closely do changes in surface and column water vapor follow Clausius–Clapeyron scaling in climate change simulations? *Environmental Research Letters* 5.2, p. 025207.

Ohmura A. (1990): On the wind profile over the Alps. *Proc. xxi Internat. Tagung f. Alpine Meteorologie*. Vol. 48. Schweiz. Met. Anstalt, pp. 102–5.

Oliver K. (2014): Untersuchungen zu Flora und Vegetation des Zugspitzplatts (Wettersteingebirge, Bayerische Alpen): rezente Vegetationsdynamik unter besonderer Berücksichtigung klimatischer und anthropo-zoogener Prozesse. PhD thesis. University of Augsburg.

Ott I., Romberg K., and Jacobeit J. (2015): Teleconnections of the tropical Atlantic and Pacific Oceans in a CMIP5 model ensemble. *Climate Dynamics* 44.11-12, pp. 3043–3055.

Panziera L. and Germann U. (2010): The relation between airflow and orographic precipitation on the southern side of the Alps as revealed by weather radar. *Quarterly Journal of the Royal Meteorological Society* 136.646, pp. 222–238.

Pfiffner O. A. (2015): Geologie der Alpen. UTB.

Philipp A., Beck C., Esteban P., Kreienkamp F., Krennert T., Lochbihler K. U., Lykoudis S. P., Pianko-Kluczynska K., Post P., Alvarez10 D. R., et al. (2014a): cost733class-1.2 User guide. Augsburg Germany.

Philipp A., Beck C., Huth R., and Jacobeit J. (2014b): Development and comparison of circulation type classifications using the COST 733 dataset and software. *International Journal of Climatology*.

Philipp A., Della-Marta P.-M., Jacobeit J., Fereday D. R., Jones P. D., Moberg A., and Wanner H. (2007): Long-term variability of daily North Atlantic–European pressure patterns since 1850 classified by simulated annealing clustering. *Journal of Climate* 20.16, pp. 4065–4095.

Picard R. R. and Cook R. D. (1984): Cross-validation of regression models. *Journal of the American Statistical Association* 79.387, pp. 575–583.

Pozo-Vázquez D., Esteban-Parra M., Rodrigo F., and Castro-Diez Y. (2001): The association between ENSO and winter atmospheric circulation and temperature in the North Atlantic region. *Journal of climate* 14.16, pp. 3408–3420.

Quadrelli R., Lazzeri M., Cacciamani C., and Tibaldi S. (2001): Observed winter Alpine precipitation variability and links with large-scale circulation patterns. *Climate Research* 17.3, pp. 275–284.

Quervain A. de (1904): Die Hebung der atmosphärischen Isothermen in der Schweizer Alpen und ihre Beziehung zu deren Höhengrenzen. *Gerlands Beitr. Geophys.* 6, pp. 481–533.

R Core Team (2017): R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria.

Ramanathan V., Crutzen P., Kiehl J., and Rosenfeld D. (2001): Aerosols, climate, and the hydrological cycle. *science* 294.5549, pp. 2119–2124.

Randall D. (2015): An introduction to the global circulation of the atmosphere. Princeton University Press.

Rasmussen R., Baker B., Kochendorfer J., Meyers T., Landolt S., Fischer A. P., Black J., Thériault J. M., Kucera P., Gochis D., et al. (2012): How well are we measuring snow: The NOAA/FAA/NCAR winter precipitation test bed. *Bulletin of the American Meteorological Society* 93.6, pp. 811–829.

Razavi S., Elshorbagy A., Wheater H., and Sauchyn D. (2015): Toward understanding nonstationarity in climate and hydrology through tree ring proxy records. *Water Resources Research* 51.3, pp. 1813–1830.

Reichler T. and Kim J. (2008): How well do coupled models simulate today's climate? *Bulletin of the American Meteorological Society* 89.3, pp. 303–311.

Ren Z. and Li M. (2007): Errors and correction of precipitation measurements in China. *Advances in Atmospheric Sciences* 24.3, pp. 449–458.

Richter M. (2000): A hypothetical framework for testing phytodiversity in mountainous regions: the influence of airstreams and hygrothermic conditions. *Phytocoenologia* 30.519-42.

Riedmiller M. and Braun H. (1993): A direct adaptive method for faster backpropagation learning: The RPROP algorithm, pp. 586–591.

Rubel F., Brugger K., Haslinger K., and Auer I. (2016): The climate of the European Alps: Shift of very high resolution Köppen-Geiger climate zones 1800–2100. *Meteorologische Zeitschrift.*

Rumelhart D. E., Hinton G. E., and Williams R. J. (1986): Learning representations by back-propagating errors. *nature* 323.6088, p. 533.

Rumelhart D. E., McClelland J. L., Group P. R., et al. (1987): Parallel distributed processing. Vol. 1. MIT press Cambridge, MA.

Rummukainen M. (1997): Methods for statistical downscaling of GCM simulations. *SMHI Rapporter. Meteorologi och Klimatologi (Sweden). no. 80.*

Ruosteenoja K. and Räisänen P. (2013): Seasonal changes in solar radiation and relative humidity in Europe in response to global warming. *Journal of Climate* 26.8, pp. 2467–2481.

Sauter T. and Venema V. (2011): Natural three-dimensional predictor domains for statistical precipitation downscaling. *Journal of Climate* 24.23, pp. 6132–6145.

Scherrer S. C., Wüthrich C., Croci-Maspoli M., Weingartner R., and Appenzeller C. (2013): Snow variability in the Swiss Alps 1864–2009. *International journal of climatology* 33.15, pp. 3162–3173.

Schiemann R. and Frei C. (2010): How to quantify the resolution of surface climate by circulation types: an example for Alpine precipitation. *Physics and Chemistry of the Earth, Parts A/B/C* 35.9, pp. 403–410.

Schlesinger M. E. and Ramankutty N. (1994): An oscillation in the global climate system of period 65–70 years. *Nature* 367.6465, pp. 723–726.

Schmaltz J. (Mar. 2016): Spring Snow Cover in The Alps. `https://earthobservatory.nasa.gov/IOTD/view.php?id=87753&src=ve`. Accessed: 2017-10-26.

Schmidli J., Goodess C. M., Frei C., Haylock M. R., Hundecha Y., Ribalaygua J., and Schmith T. (2007): Statistical and dynamical downscaling of precipitation: An evaluation and comparison of scenarios for the European Alps. *Journal of Geophysical Research: Atmospheres (1984–2012)* 112.D4.

Schmidli J., Schmutz C., Frei C., Wanner H., and Schär C. (2002): Mesoscale precipitation variability in the region of the European Alps during the 20th century. *International Journal of Climatology* 22.9, pp. 1049–1074.

Schöner W., Auer I., and Böhm R. (2009): Long term trend of snow depth at Sonnblick (Austrian Alps) and its relation to climate change. *Hydrological Processes: An International Journal* 23.7, pp. 1052–1063.

Schönwiese C.-D. (2013a): Klimatologie. 4th ed.

– (2013b): Praktische statistik für meteorologen und geowissenschaftler.

Schultz J. (2016): Die Ökozonen der Erde. UTB.

Schumacher C. (1923): Der Wind in der freien Atmosphäre und auf Säntis, Zugspitze und Sonnblick. *Beitr. Phys. frei Atmos.* 11, pp. 20–42.

Sevruk B. (1986): Correction of precipitation measurements. Vol. 24. Proceedings of the International Workshop on the Correction of Precipitation Measurements, Instruments and Observing Methods. World Meteorological Organization.

Sherwood S. C., Ingram W., Tsushima Y., Satoh M., Roberts M., Vidale P. L., and O'Gorman P. A. (2010): Relative humidity changes in a warmer climate. *Journal of Geophysical Research: Atmospheres* 115.D9.

Shojaeefard M. H., Akbari M., Tahani M., and Farhani F. (2013): Sensitivity analysis of the artificial neural network outputs in friction stir lap joining of aluminum to brass. *Advances in Materials Science and Engineering* 2013. DOI: `http://dx.doi.org/10.1155/2013/574914`.

Smiatek G., Kunstmann H., Knoche R., and Marx A. (2009): Precipitation and temperature statistics in high-resolution regional climate models: Evaluation for the European Alps. *Journal of Geophysical Research: Atmospheres* 114.D19.

Smiatek G., Kaspar S., and Kunstmann H. (2013): Hydrological climate change impact analysis for the Figeh Spring near Damascus, Syria. *Journal of Hydrometeorology* 14.2, pp. 577–593.

Springer C., Matulla C., Schöner W., Steinacker R., and Wagner S. (2013): Downscaled GCM projections of winter and summer mass balance for Central European glaciers (2000–2100) from ensemble simulations with ECHAM5-MPIOM. *International Journal of Climatology* 33.5, pp. 1270–1279.

Steeneveld G.-J. (2018): Atmospheric Practical. Lecture notes.

Steger C., Kotlarski S., Jonas T., and Schär C. (2013): Alpine snow cover in a changing climate: a regional climate model perspective. *Climate dynamics* 41.3-4, pp. 735–754.

Steinhauser F. (1938): Die Meteorologie des Sonnblicks: Beiträge zur Hochgebirgsmeteorologie nach Ergebnissen sojähriger Beobachtungen des Sonnblick-observatoriums, 3106 m. Springer-Verlag.

Stisen S., Højberg A., Troldborg L., Refsgaard J., Christensen B., Olsen M., and Henriksen H. (2012): On the importance of appropriate precipitation gauge catch correction for hydrological modelling at mid to high latitudes. *Hydrology and Earth System Sciences* 16.11, p. 4157.

Storch H. von, Hewitson B., and Mearns L. (2000): Review of empirical downscaling techniques. *Regional climate development under global warming. General Technical Report* 4.

Sturm M., Taras B., Liston G. E., Derksen C., Jonas T., and Lea J. (2010): Estimating snow water equivalent using snow depth data and climate classes. *Journal of Hydrometeorology* 11.6, pp. 1380–1394.

Sutton R. T. and Dong B. (2012): Atlantic Ocean influence on a shift in European climate in the 1990s. *Nature Geoscience* 5.11, pp. 788–792.

Sutton R. T. and Hodson D. L. (2005): Atlantic Ocean forcing of North American and European summer climate. *science* 309.5731, pp. 115–118.

Tareghian R. and Rasmussen P. F. (2013): Statistical downscaling of precipitation using quantile regression. *Journal of hydrology* 487, pp. 122–135.

Taylor K. E., Stouffer R. J., and Meehl G. A. (2012): An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society* 93.4, pp. 485–498.

Tolika K., Maheras P., Vafiadis M., Flocas H. A., and Arseni-Papadimitriou A. (2007): Simulation of seasonal precipitation and raindays over Greece: a statistical downscaling technique based on artificial neural networks (ANNs). *International Journal of Climatology* 27.7, pp. 861–881.

Toreti A., Kuglitsch F.-G., Xoplaki E., Della-Marta P. M., Aguilar E., Prohom M., and Luterbacher J. (2011): A note on the use of the standard normal homogeneity test to

detect inhomogeneities in climatic time series. *International Journal of Climatology* 31.4, pp. 630–632.

Trigo R. M. and Palutikof J. P. (1999): Simulation of daily temperatures for climate change scenarios over Portugal: a neural network model approach. *Climate Research* 13.1, pp. 45–59.

Tuller S. E. (1973): Seasonal and annual precipitation efficiency in Canada. *Atmosphere* 11.2, pp. 52–66.

Vincent P., Larochelle H., Bengio Y., and Manzagol P.-A. (2008): Extracting and composing robust features with denoising autoencoders. *Proceedings of the 25th international conference on Machine learning*. ACM, pp. 1096–1103.

Vuuren D. P. van, Edmonds J., Kainuma M., Riahi K., Thomson A., Hibbard K., Hurtt G. C., Kram T., Krey V., Lamarque J.-F., Masui T., Meinshausen M., Nakicenovic N., Smith S. J., and Rose S. K. (Aug. 2011a): The representative concentration pathways: an overview. *Climatic Change* 109.1, p. 5. ISSN: 1573-1480. DOI: 10.1007/s10584-011-0148-z.

Vuuren D. P. van, Edmonds J. A., Kainuma M., Riahi K., and Weyant J. (Aug. 2011b): A special issue on the RCPs. *Climatic Change* 109.1, p. 1. ISSN: 1573-1480. DOI: 10.1007/s10584-011-0157-y.

Wahl E. W. (1966): Windspeed on Mountains. Final Report AFCRL-66-280. University of Wisconsin.

Wallace J. M. and Hobbs P. V. (2006): Atmospheric science: an introductory survey. Vol. 92. Academic press.

Walter H. and Lieth H. (1967): Klimadiagramm-Weltatlas: etwa 8000 Klimastationen (etwa 9000 Diagramme), 33 Hauptkarten, 22 Nebenkarten; Herausgegeben in 3 Lieferungen. Fischer.

Wang B., Wu R., and Fu X. (2000): Pacific–East Asian teleconnection: how does ENSO affect East Asian climate? *Journal of Climate* 13.9, pp. 1517–1536.

Wang W., Jones P., and Partridge D. (2000): Assessing the impact of input features in a feedforward neural network. *Neural Computing & Applications* 9.2, pp. 101–112.

Wanner H., Brönnimann S., Casty C., Gyalistras D., Luterbacher J., Schmutz C., Stephenson D. B., and Xoplaki E. (2001): North Atlantic Oscillation–concepts and studies. *Surveys in geophysics* 22.4, pp. 321–381.

Wanner H., Rickli R., Salvisberg E., Schmutz C., and Schüepp M. (1997): Global climate change and variability and its influence on alpine climate-concepts and observations. *Theoretical and Applied Climatology* 58.3, pp. 221–243.

Warscher M., Strasser U., Kraller G., Marke T., Franz H., and Kunstmann H. (2013): Performance of complex snow cover descriptions in a distributed hydrological model system: A case study for the high Alpine terrain of the Berchtesgaden Alps. *Water resources research* 49.5, pp. 2619–2637.

Weber M., Bernhardt M., and Ludwig R. (2017): Hydrologische Modellierung und Schneedeckenmonitoring auf dem Zugspitzplatt. *Geographica Augustana*.

Weber R. O., Talkner P., Auer I., Böhm R., Gajić-Čapka M., Zaninović K., Brazdil R., and Faško P. (1997): 20th-century changes of temperature in the mountain regions of central Europe. *Climatic Change at High Elevation Sites*. Springer, pp. 95–112.

Weingartner R., Viviroli D., and Schädler B. (2007): Water resources in mountain regions: a methodological approach to assess the water balance in a highland-lowland-system. *Hydrological Processes: An International Journal* 21.5, pp. 578–585.

Weischet W. and Endlicher W. (2008): Einführung in die allgemeine Klimatologie. Schweizerbart'sche Verlagsbuchhandlung.

Werbos P. (1974): Beyond regression: new fools for prediction and analysis in the behavioral sciences. *PhD thesis, Harvard University*.

Whiteman C. D. (2000): Mountain meteorology: fundamentals and applications. Oxford University Press.

Wilamowski B. M. (2011): How to not get frustrated with neural networks. *Industrial Technology (ICIT), 2011 IEEE International Conference on*. IEEE, pp. 5–11.

Wilby R. L., Charles S. P., Zorita E., Timbal B., Whetton P., and Mearns L. O. (2004): Guidelines for use of climate scenarios developed from statistical downscaling methods. *Supporting material of the Intergovernmental Panel on Climate Change, available from the DDC of IPCC TGCIA* 27.

Wilby R. L. and Wigley T. (1997): Downscaling general circulation model output: a review of methods and limitations. *Progress in Physical Geography* 21.4, pp. 530–548.

– (2000): Precipitation predictors for downscaling: observed and general circulation model relationships. *International Journal of Climatology* 20.6, pp. 641–661.

Wilby R. L., Wigley T., Conway D., Jones P., Hewitson B., Main J., and Wilks D. (1998): Statistical downscaling of general circulation model output: a comparison of methods. *Water resources research* 34.11, pp. 2995–3008.

Wilks D. S. (2011): Statistical methods in the atmospheric sciences. Vol. 100. Academic press.

Willett K. M., Jones P. D., Gillett N. P., and Thorne P. W. (2008): Recent changes in surface humidity: Development of the HadCRUH dataset. *Journal of Climate* 21.20, pp. 5364–5383.

Willmott C. J. and Matsuura K. (2005): Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate research* 30.1, pp. 79–82.

WMO (1970): Climatic Atlas of Europe I: Maps of Mean Temperature and Precipitation.

Woollings T. (2008): Vertical structure of anthropogenic zonal-mean atmospheric circulation change. *Geophysical Research Letters* 35.19.

Yang T., Li H., Wang W., Xu C.-Y., and Yu Z. (2012): Statistical downscaling of extreme daily precipitation, evaporation, and temperature and construction of future scenarios. *Hydrological Processes* 26.23, pp. 3510–3523.

Zampieri M., Scoccimarro E., and Gualdi S. (2013): Atlantic influence on spring snowfall over the Alps in the past 150 years. *Environmental Research Letters* 8.3, p. 034026.

Zeng X. and Yeung D. S. (2001): Sensitivity analysis of multilayer perceptron to input and weight perturbations. *IEEE Transactions on Neural Networks* 12.6, pp. 1358–1366.

Zepp H. (2014): Geomorphologie: eine Einführung. Vol. 2164. UTB.

Zhai B. and Chen J. (2018): Development of a stacked ensemble model for forecasting and analyzing daily average PM 2.5 concentrations in Beijing, China. *Science of The Total Environment* 635, pp. 644–658.

Zhang G. P. (2007): Avoiding pitfalls in neural network research. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 37.1, pp. 3–16.

A

Appendix

Table A.1: Predictor screening process separated by seasons for precipitation at the Sonnblick. The bold vertical line separates the predictor selection process from the domain reduction steps (step number at the top). The initial setup (step 0) is listed on the left hand side, by predictor and level in hPa. The total number of predictors is reduced to a minimum setup (see section 7.5.3), the step number of the selected final predictor configuration is labelled in red. The green number represents the average sum of sensitivities and the purple number the average maximum of the respective predictor sensitivity derived from PaD during a 10-fold cross validated approach. Green numbers should sum up to approximately 100, although round-off errors are possible.

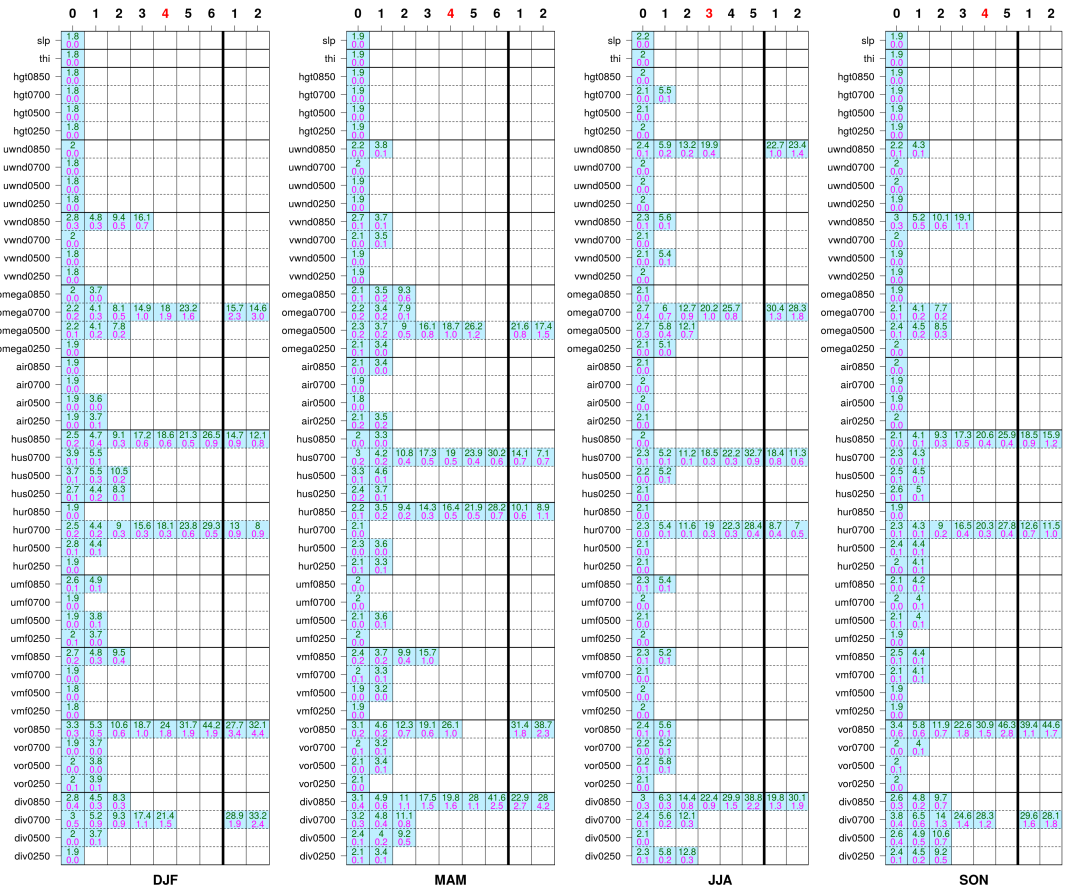


Table A.2: Predictor screening process separated by seasons for precipitation at the Zugspitze. For further explanation see table A.1.

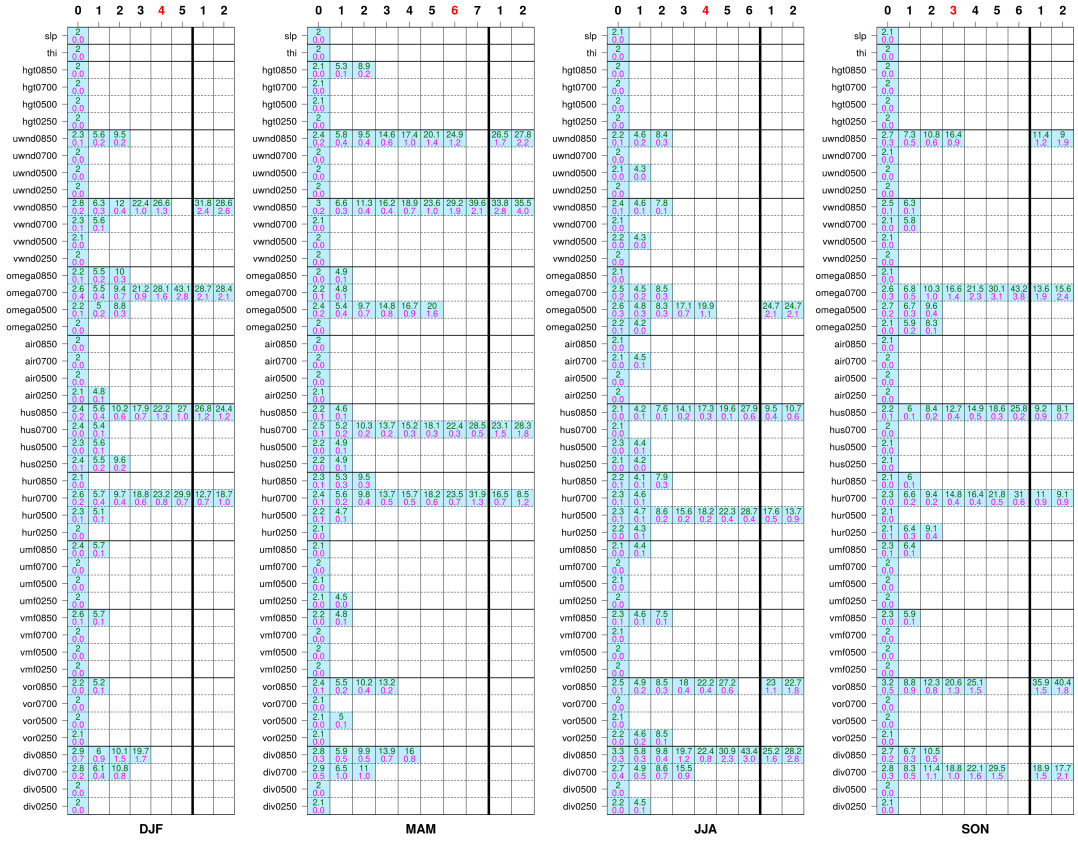


Table A.3: Predictor screening process separated by seasons for temperature at the Sonnblick. For further explanation see table A.1.

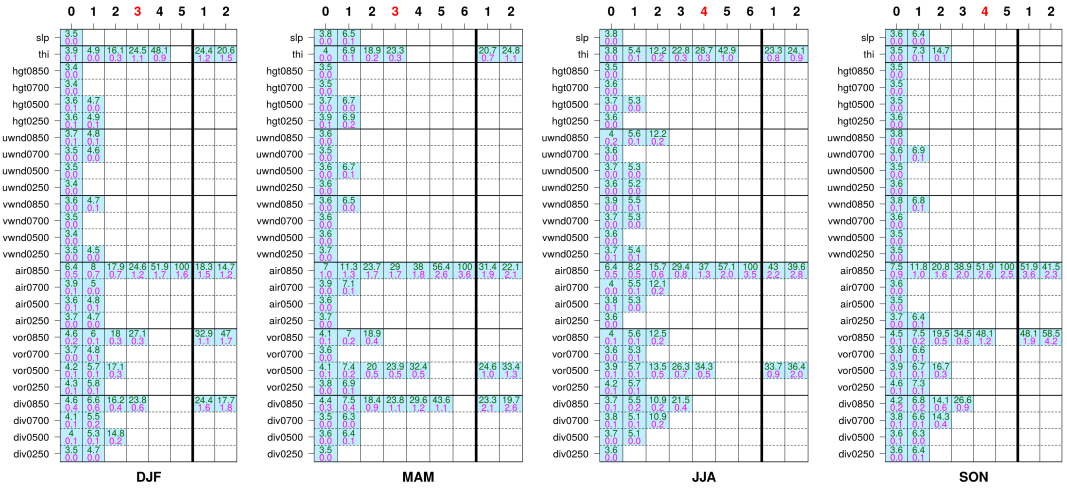


Table A.4: Predictor screening process separated by seasons for temperature at the Zugspitze. For further explanation see table A.1.

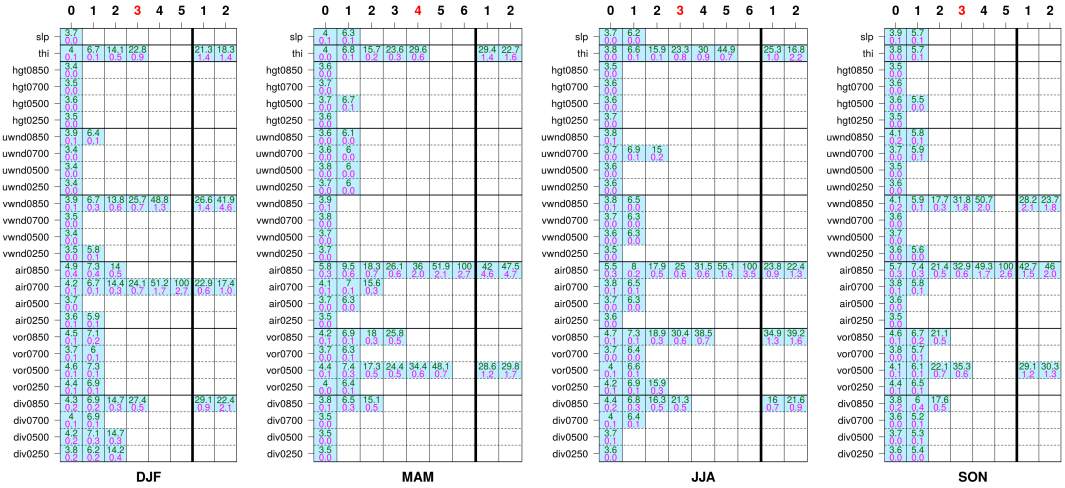


Table A.5: Predictor screening process separated by seasons for wind speed at the Zugspitze. For further explanation see table A.1.

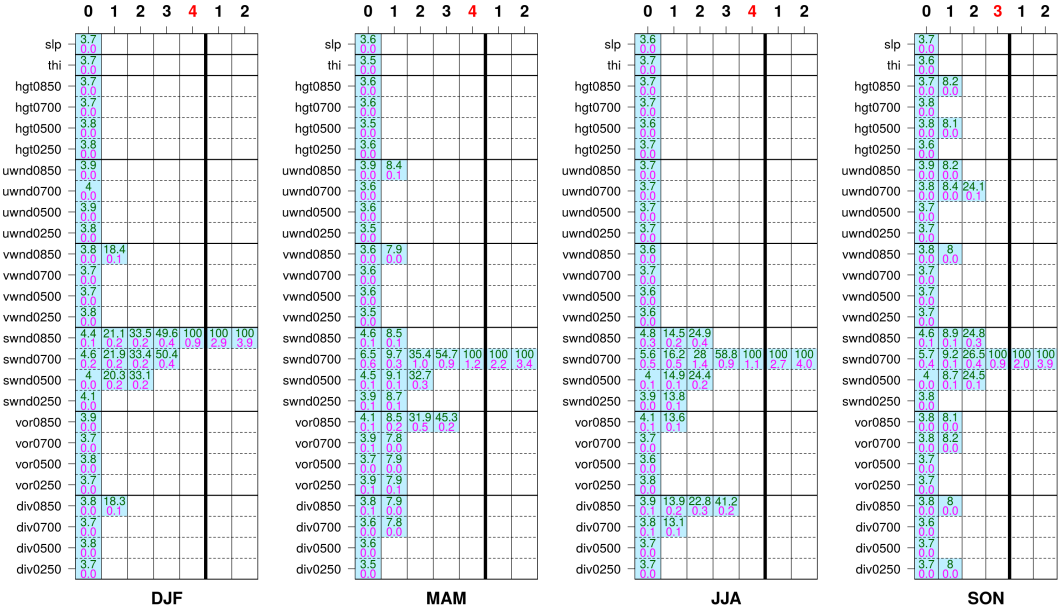
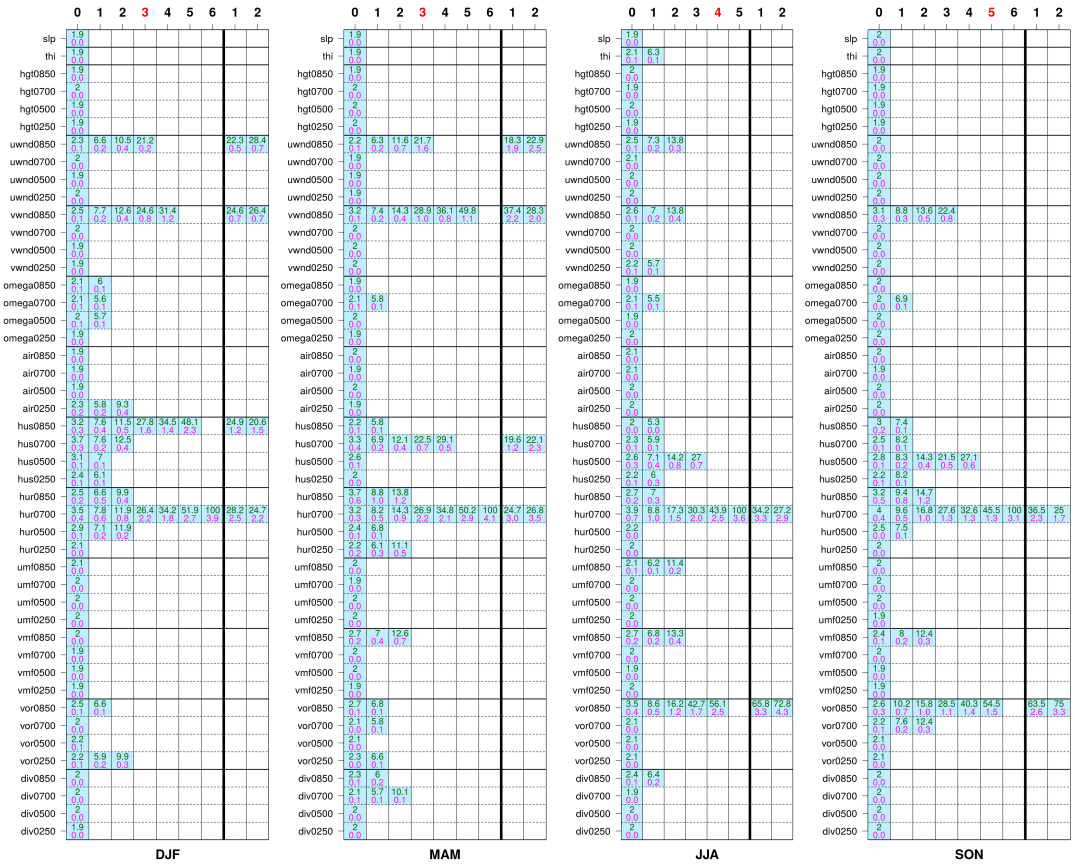


Table A.6: Predictor screening process separated by seasons for relative humidity at the Zugspitze. For further explanation see table A.1.



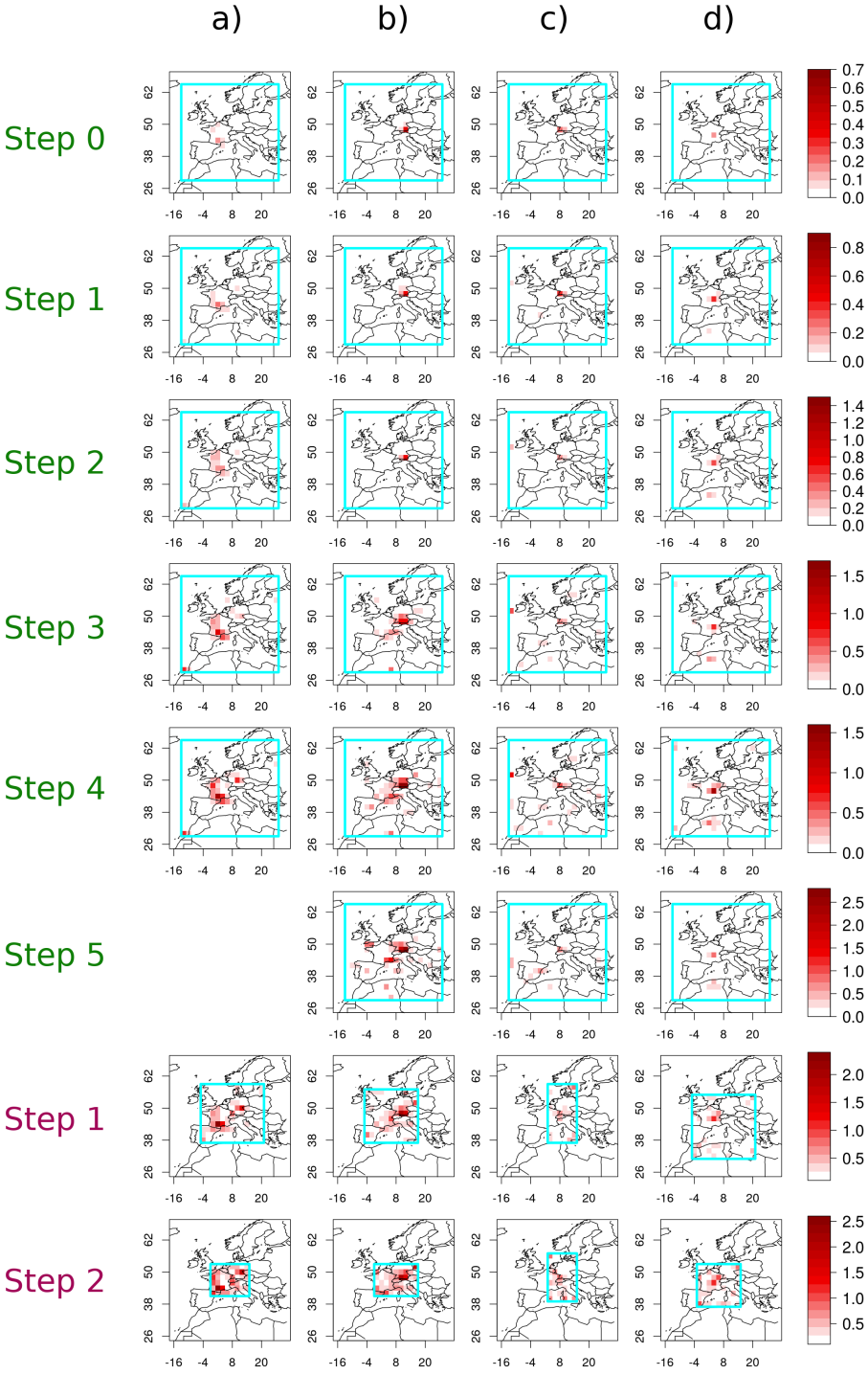


Figure A.1: Sensitivity pattern development during predictor screening for the final predictors considering Zugspitze precipitation in winter. Each row is one selection step (see also figure A.4), while columns represent the final predictor combination. Column a) vwnd 850 hPa; b) omega 700 hPa; c) hur 700 hPa; d) hus 850 hPa.

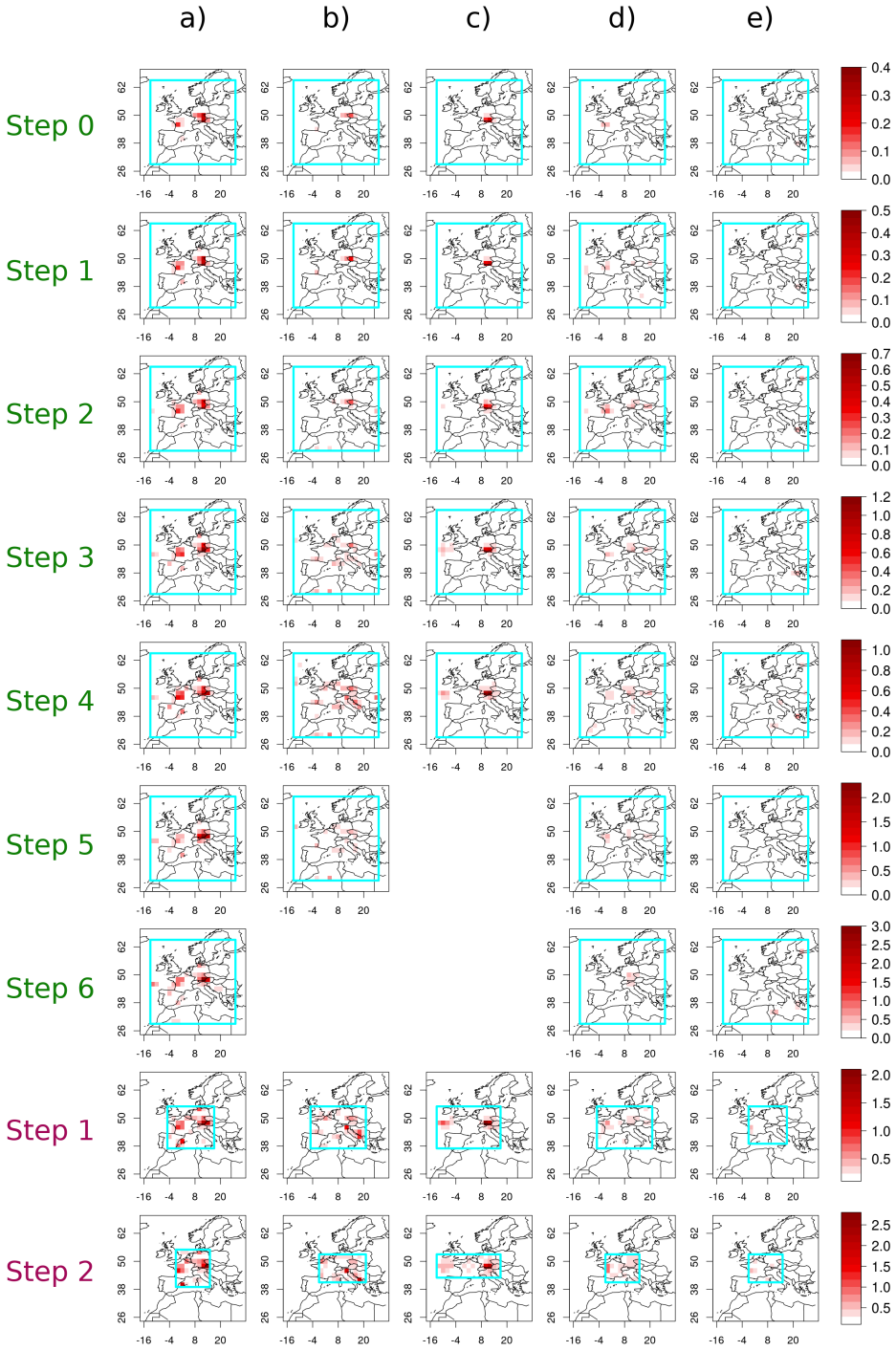


Figure A.2: Sensitivity pattern development during predictor screening for the final predictors considering Zugspitze precipitation in summer. Each row is one selection step (see also figure A.4), while columns represent the final predictor combination. Column a) div 850 hPa; b) vor 850 hPa; c) omega 500 hPa; d) hur 500 hPa; e) hus 850 hPa.

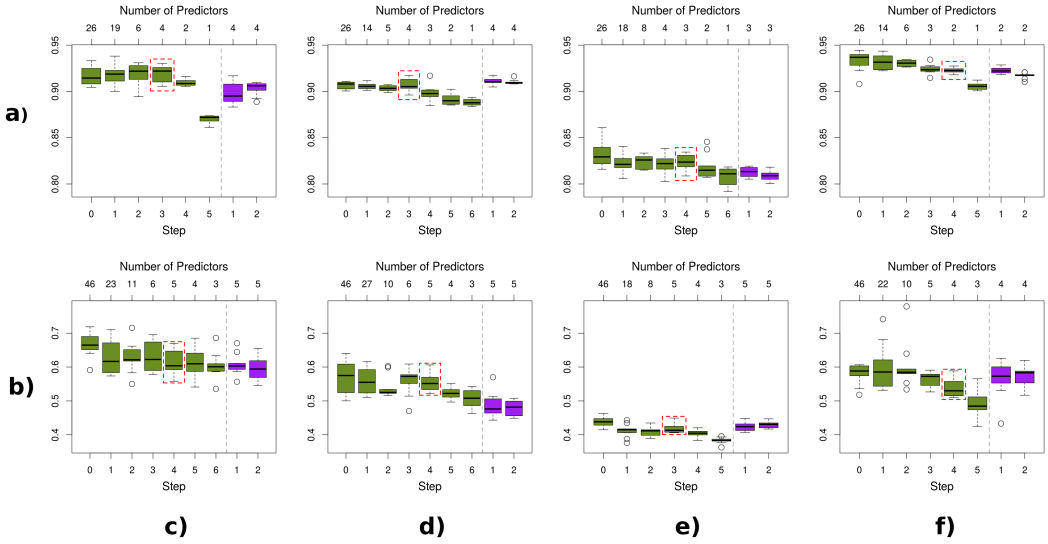


Figure A.3: Sonnblick ANN model performance during predictor screening measured by the MSSS for each predictand and season. Row a) temperature and row b) precipitation. Column c) DJF d) MAM e) JJA and f) SON. The dashed vertical line separates the predictor selection process (green boxes) from the domain reduction steps (purple boxes, step number at the bottom). Labels on top list the number of predictors at each step. The dashed red box marks the selected final predictor configuration.

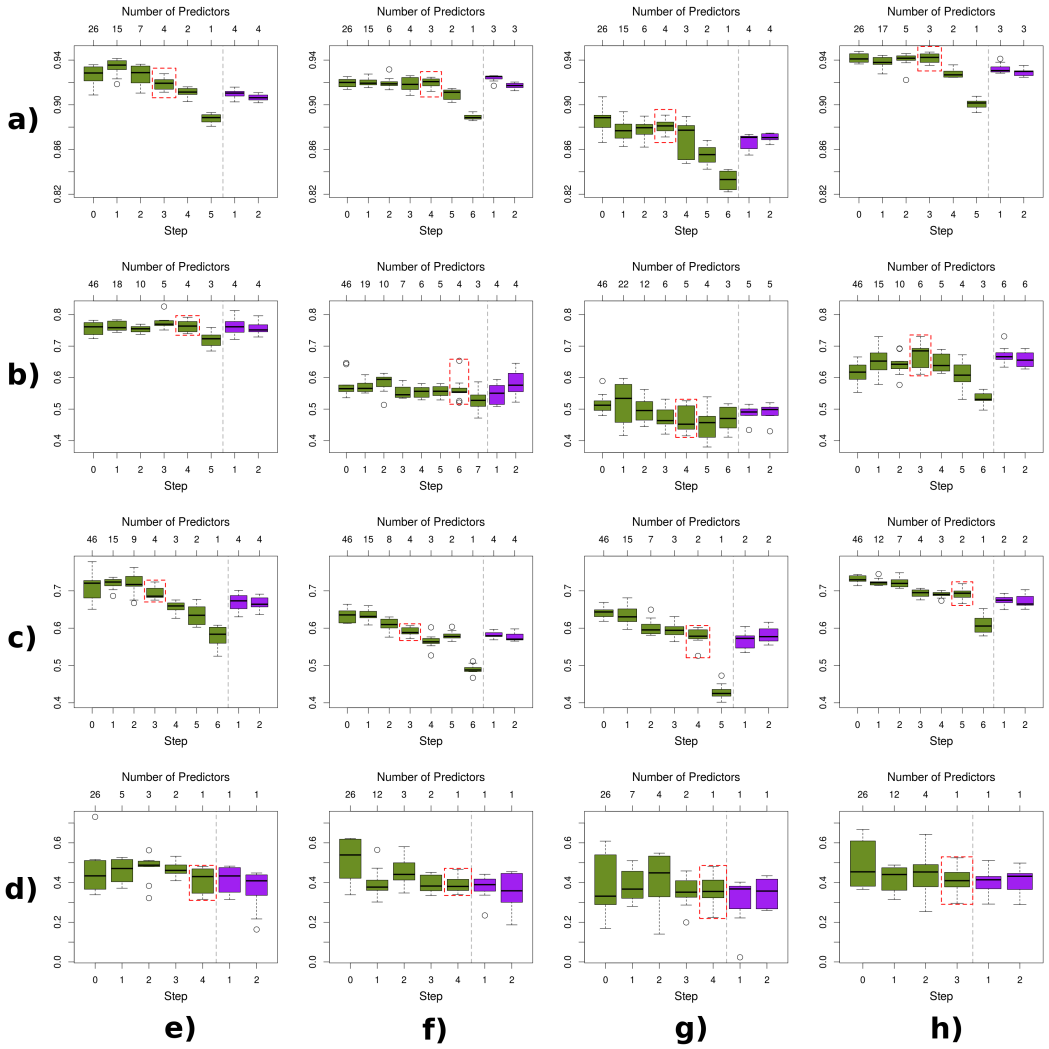


Figure A.4: Zugspitze ANN model performance during predictor screening measured by the MSSS for each predictand and season. Row a) temperature, row b) precipitation, row c) relative humidity and row d) wind speed. Column e) DJF f) MAM g) JJA and h) SON. For further explanation see figure A.3.

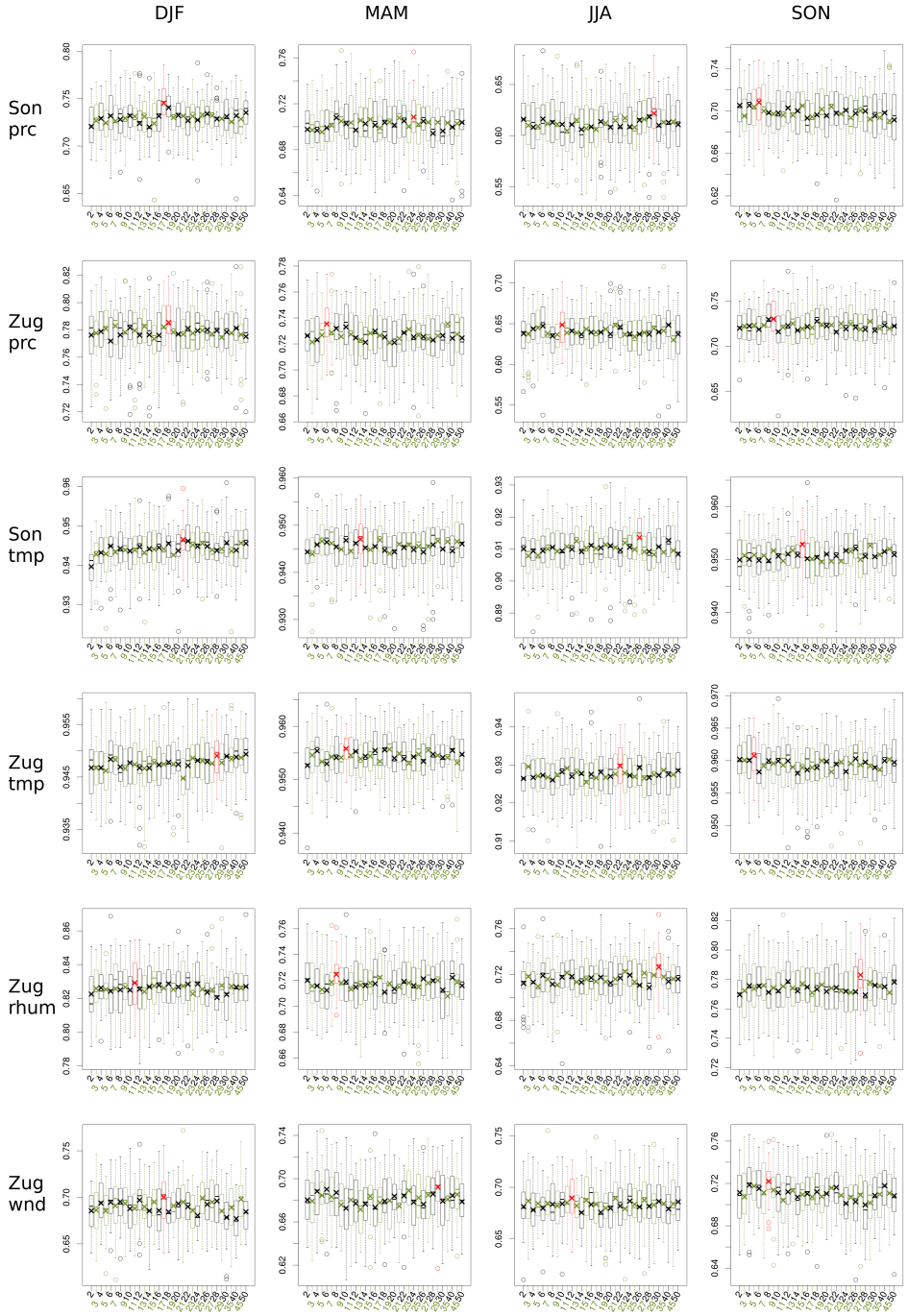
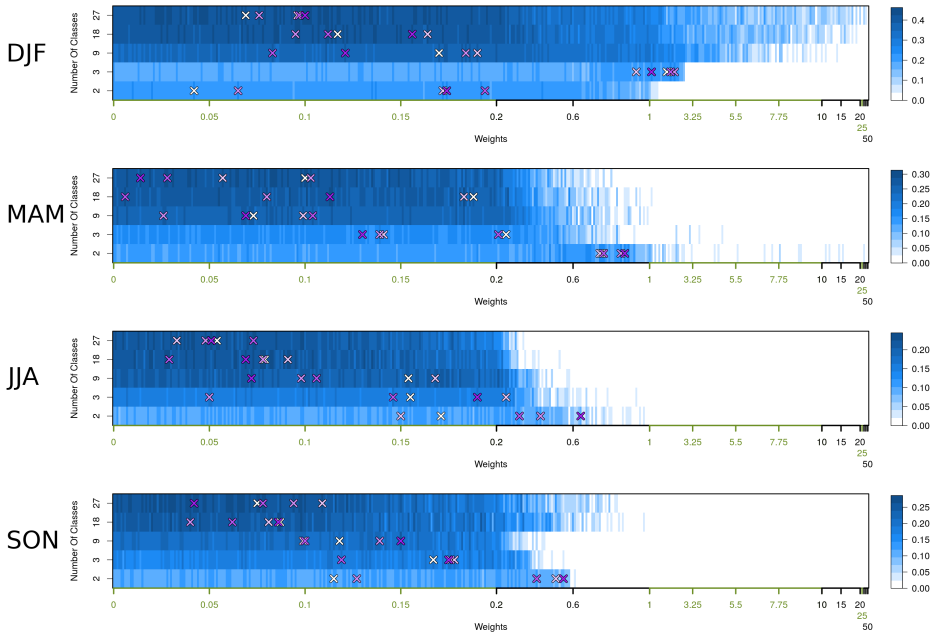
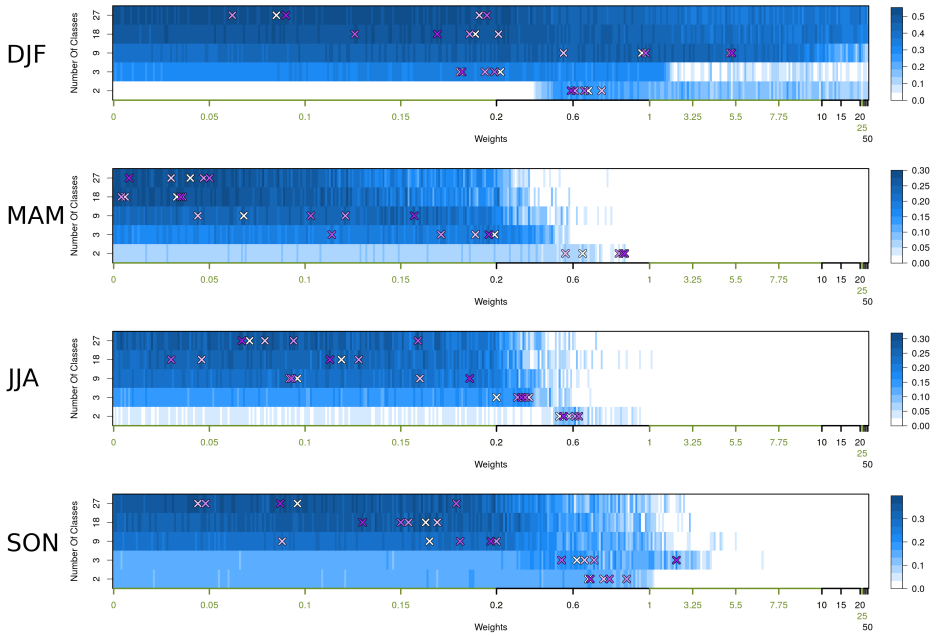


Figure A.5: Spearman correlation coefficient between observations and reanalysis based simulations considering different numbers of hidden neurons in an ANN with one hidden layer for the final predictor configurations (see Table 7.4). Boxplots show the distribution of the Spearman correlation coefficient reached by 30 cross-validated runs in the validation period. Selection of the best number is performed in an automated routine by comparing the arithmetical mean (crosses). The final choice is marked in red. Number of hidden neurons are coloured alternated in green and black for better readability.

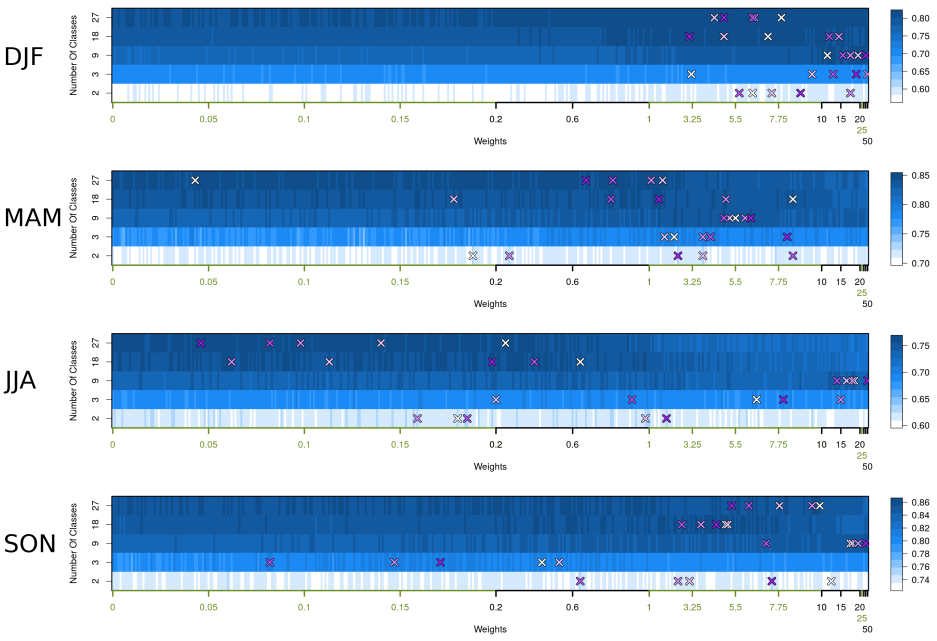


(a) Sonnblick precipitation

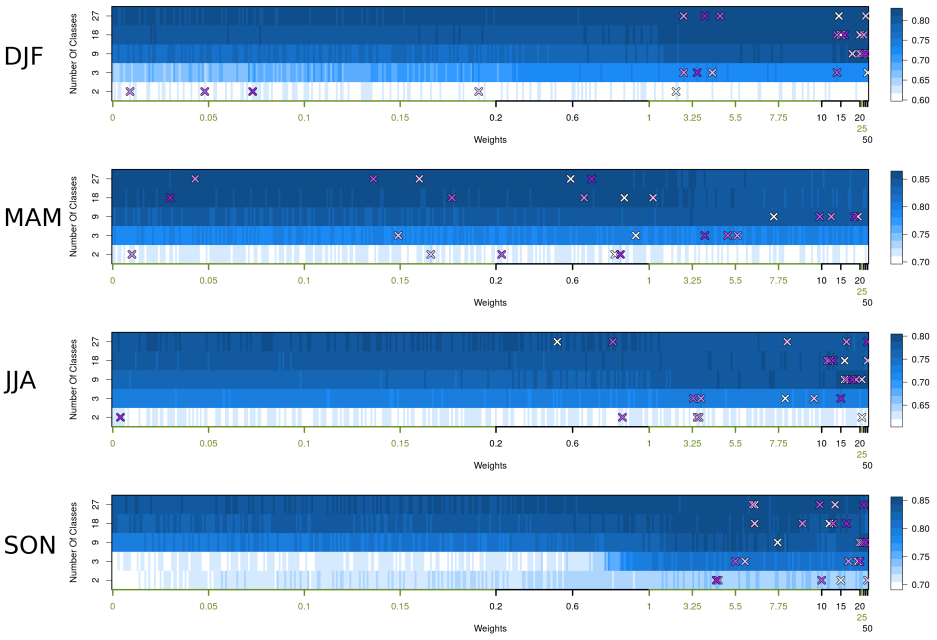


(b) Zugspitze precipitation

Figure A.6: Resulting MSSS of reanalysis based precipitation simulations in the validation period with respect to the number of classes and predictand weights in the RCF approach. Top five weights of each class are marked with crosses coloured from purple (highest skill score) to white (fifth-highest skill score). Weights are altered in different step sizes indicated by a switch in axis colour. Step sizes are: 0.001 between 0 and 0.2; 0.01 between 0.2 to 1; 0.1 between 1 to 10; 0.5 between 10 to 20; 5 between 20 and 30; 10 between 30 and 50.

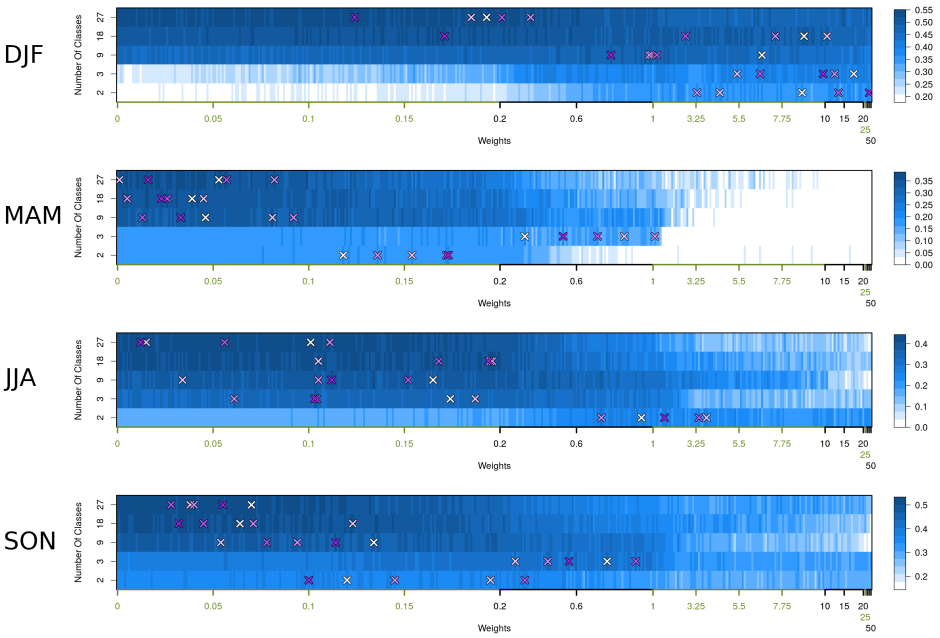


(a) Sonnblick temperature

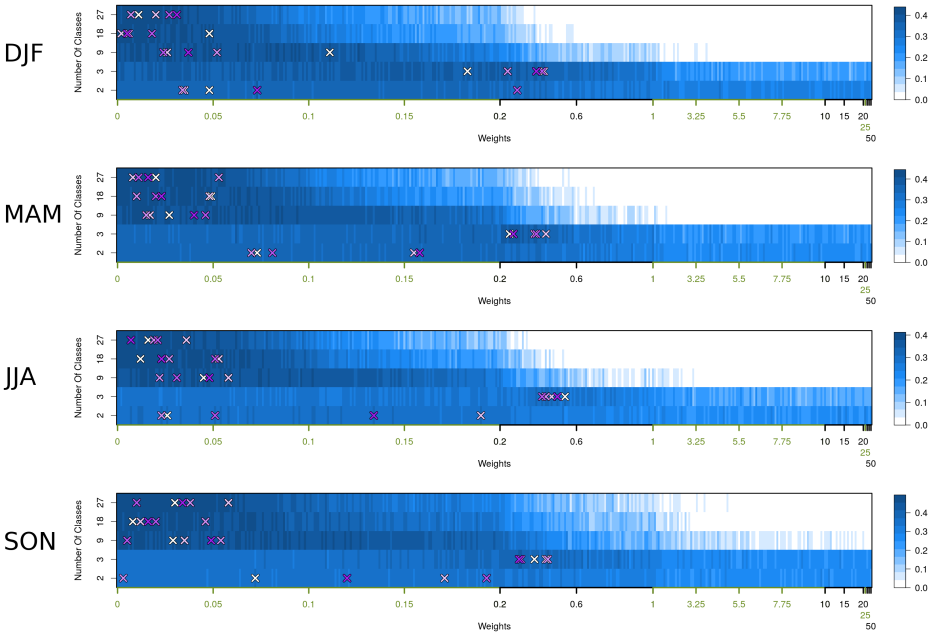


(b) Zugspitze temperature

Figure A.7: Resulting MSSS of reanalysis based temperature simulations in the validation period with respect to the number of classes and predictand weights in the RCF approach. See figure A.6 for plot description.



(a) Zugspitze relative humidity



(b) Zugspitze wind speed

Figure A.8: Resulting MSSS of reanalysis based relative humidity and wind speed simulations in the validation period with respect to the number of classes and predictand weights in the RCF approach. See figure A.6 for plot description.

Table A.7: Characteristics of ANN training with Rprop considering the final SD model for Zugspitze winter precipitation (time period 1970-2000). Training-specific values represent the mean of a 10-fold Monte-Carlo cross-validation approach, while 20 % of the dataset is separated for validation. Training is performed with 10 hidden neurons in one hidden layer in batch mode and terminated at 300 early stopping checks (20 % of the dataset is used for the evaluation of early stopping). Altered hyperparameter are listed in the table header. Line keywords: learning rate ($\boldsymbol{\eta}$), momentum term ($\boldsymbol{\alpha}$), decay parameter (**d**), η decay factor (**de**), not successful trained (**nt**) and not finished (**nf**) training examples, iterations here **Epochs**, training run time (**rt**; hours:minutes:seconds), min and max value of the final ANN weights (**wr**), frequency [0,1] of small weights (small weight value range [-0.1,0.1]), Mean Square Skill Score (**MSSS**) and Spearman correlation coefficient ($\boldsymbol{r_s}$) of the validation period. Note that some properties only apply to specific training algorithms. ANNs are considered as "successfully" trained if at least one weight adjustment is carried out. A hard limit of 10^6 epochs is set, which corresponds to an approximate training run time of 3 hours on the used computer cluster. Performance statistics (MSSS and r_s) are solely computed from trained and finished ANNs. If unfinished ANNs are included, an overall performance is given in parenthesis, while untrained ANNs are excluded. In this thesis, primarily the Rprop algorithm is applied. For the purpose of comparison, algorithms resulting in an equal or higher model performance as the Rprop, are printed in bold.

nt/nf	0/0
Epochs	839
rt	00:00:11
wr	-6.2/11.6
swf	0.22
MSSS	**0.67**
$\boldsymbol{r_s}$	**0.73**

Table A.8: Characteristics of training with the Extended Delta-Bar-Delta algorithm considering the final SD model for Zugspitze winter precipitation. See table description A.7 for detailed information on setup and keywords.

$\boldsymbol{\eta}$	0.5	0.1	0.01	0.001
nt/nf	0/0	0/0	0/0	0/0
Epochs	369	1317	991	2442
rt	00:00:04	00:00:13	00:00:10	00:00:23
wr	-0.2/1	-0.6/0.9	-0.6/0.8	-0.6/1
swf	0.02	0.19	0.21	0.2
MSSS	0.06	**0.67**	**0.67**	**0.67**
$\boldsymbol{r_s}$	0.07	**0.73**	**0.73**	**0.73**

Table A.9: Characteristics of training with Gradient Descent considering the final SD model for Zugspitze winter precipitation. See table description A.7 for detailed information on setup and keywords.

$\boldsymbol{\eta}$	0.5	0.1	0.01	0.001
nt/nf	0/0	0/0	0/0	0/7
Epochs	3221	18913	247533	649298
rt	00:00:42	00:04:03	00:49:60	02:08:32
wr	-0.7/0.9	-0.5/1	-0.7/0.9	-0.5/0.8
swf	0.35	0.19	0.2	0.2
MSSS	0.62	**0.67**	**0.67**	**0.67** (0.53)
$\boldsymbol{r_s}$	0.71	**0.73**	**0.73**	**0.73** (0.6)

Table A.10: Characteristics of training with the Adaptive Learning Rate algorithm considering the final SD model for Zugspitze winter precipitation. See table description A.7 for detailed information on setup and keywords.

η	0.5	0.1	0.01	0.001
nt/nf	0/0	0/0	0/0	0/0
Epochs	7861	6437	7378	7162
rt	00:01:16	00:01:03	00:01:11	00:01:11
wr	-0.4/0.9	-0.4/0.8	-0.5/0.8	-0.6/0.8
swf	0.25	0.26	0.26	0.25
MSSS	0.66	0.66	0.66	**0.67**
r_s	**0.73**	**0.73**	**0.73**	**0.73**

Table A.11: Characteristics of training with Individual Adaptive Learning Rate algorithm considering the final SD model for Zugspitze winter precipitation. See table description A.7 for detailed information on setup and keywords.

η	0.5	0.1	0.01	0.001
nt/nf	0/0	0/0	0/0	0/0
Epochs	1385	1120	1421	1979
rt	00:00:14	00:00:11	00:00:14	00:00:20
wr	-2.8/2.4	-1.7/2.1	-3.6/5.1	-2.9/5.7
swf	0.16	0.18	0.18	0.19
MSSS	0.66	0.65	0.6	**0.67**
r_s	**0.73**	**0.73**	0.62	**0.73**

Table A.12: Characteristics of training with Decaying Eta considering the final SD model for Zugspitze winter precipitation. See table description A.7 for detailed information on setup and keywords.

η	0.5	0.5	0.5	0.5	0.5	0.1	0.1	0.1	0.1	0.1
de	0.9999	0.999	0.99	0.9	0.5	0.9999	0.999	0.99	0.9	0.5
nt/nf	0/0	1/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
Epochs	32952	23117	3358	535	353	180097	29975	3505	604	348
rt	00:05:24	00:03:43	00:00:34	00:00:06	00:00:04	00:29:05	00:05:05	00:00:37	00:00:07	00:00:04
wr	-0.8/1	-0.6/1	-0.5/0.4	-0.4/0.3	-0.2/0.2	-0.4/0.8	-0.4/0.5	-0.3/0.3	-0.2/0.3	-0.2/0.3
swf	0.35	0.36	0.35	0.31	0.22	0.2	0.19	0.19	0.19	0.19
MSSS	0.66	0.64	0	0	-1.1	**0.67**	0.57	0.26	0	-0.02
r_s	0.72	0.71	0.16	0.08	0.21	**0.73**	0.71	0.61	0.13	0.11
η	0.01	0.01	0.01	0.01	0.01	0.001	0.001	0.001	0.001	0.001
de	0.9999	0.999	0.99	0.9	0.5	0.9999	0.999	0.99	0.9	0.5
nt/nf	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
Epochs	270138	30606	3318	616	349	274268	28509	3494	609	346
rt	00:46:06	00:05:09	00:00:34	00:00:07	00:00:04	00:45:51	00:04:43	00:00:35	00:00:06	00:00:04
wr	-0.3/0.5	-0.2/0.3	-0.2/0.3	-0.2/0.2	-0.3/0.3	-0.3/0.3	-0.2/0.3	-0.2/0.2	-0.2/0.3	-0.3/0.2
swf	0.19	0.19	0.19	0.19	0.19	0.19	0.19	0.19	0.19	0.19
MSSS	0.57	0.23	0.03	-0.79	-4.69	0.21	0.01	-1.12	-2.09	-5.1
r_s	0.71	0.58	0.18	0.03	-0.12	0.58	0.13	-0.05	-0.02	-0.21

Table A.13: Characteristics of training with Momentum Term considering the final SD model for Zugspitze winter precipitation. See table description A.7 for detailed information on setup and keywords.

η	0.5	0.5	0.5	0.5	0.5	0.5	0.5
α	0.99	0.5	0.1	0.05	0.01	0.001	0.0001
nt/nf	0/0	0/0	0/0	1/0	0/0	0/0	0/0
Epochs	6531	4609	8652	6141	2550	2793	1854
rt	00:01:21	00:00:57	00:01:42	00:01:13	00:00:30	00:00:33	00:00:22
wr	-0.4/0.8	-0.5/0.9	-0.8/1	-0.7/1	-0.6/0.9	-0.5/1.1	-0.7/1
swf	0.25	0.21	0.32	0.31	0.33	0.36	0.37
MSSS	0.53	**0.67**	0.65	0.66	0.62	0.57	0.59
r_s	0.65	**0.73**	0.72	0.72	0.72	0.65	0.71
η	0.1	0.1	0.1	0.1	0.1	0.1	0.1
α	0.99	0.5	0.1	0.05	0.01	0.001	0.0001
nt/nf	0/0	0/0	0/0	0/0	0/0	0/0	0/0
Epochs	14762	26732	23878	18672	28644	21604	19453
rt	00:02:55	00:05:12	00:04:39	00:03:40	00:05:34	00:04:16	00:03:49
wr	-0.4/0.7	-0.7/0.9	-0.6/0.9	-0.6/0.9	-0.7/0.8	-0.6/0.9	-0.5/0.9
swf	0.22	0.21	0.2	0.2	0.2	0.19	0.19
MSSS	0.48	**0.67**	**0.67**	**0.67**	**0.67**	**0.67**	**0.67**
r_s	0.64	**0.73**	**0.73**	**0.73**	**0.73**	**0.73**	**0.73**
η	0.01	0.01	0.01	0.01	0.01	0.01	0.01
α	0.99	0.5	0.1	0.05	0.01	0.001	0.0001
nt/nf	0/0	0/0	0/0	0/0	0/0	0/1	0/0
Epochs	178021	234174	271386	260847	352454	192615	222678
rt	00:35:45	00:46:15	00:52:58	00:50:44	01:08:58	00:37:50	00:43:37
wr	-0.5/0.9	-0.6/0.9	-0.6/0.9	-0.7/0.9	-0.7/1	-0.5/0.9	-0.6/0.9
swf	0.2	0.2	0.19	0.2	0.2	0.19	0.19
MSSS	**0.67**	**0.67**	**0.67**	0.66	**0.67**	**0.67** (0.14)	**0.67**
r_s	**0.73**	**0.73**	**0.73**	**0.73**	**0.73**	**0.73** (0.15)	**0.73**
η	0.001	0.001	0.001	0.001	0.001	0.001	0.001
α	0.99	0.5	0.1	0.05	0.01	0.001	0.0001
nt/nf	0/5	0/6	0/6	0/6	0/7	0/7	0/7
Epochs	222240	787528	678544	718377	658283	581725	782884
rt	00:43:52	02:34:54	02:11:51	02:19:53	02:09:14	01:53:51	02:30:21
wr	-0.3/0.4	-0.7/0.6	-0.6/0.7	-0.5/0.8	-0.4/0.8	-0.4/0.8	-0.4/0.7
swf	0.19	0.19	0.19	0.2	0.19	0.19	0.2
MSSS	0.25 (0.35)	**0.68** (0.46)	0.66 (0.46)	**0.68** (0.46)	**0.67** (0.53)	**0.67** (0.52)	**0.67** (0.53)
r_s	0.3 (0.41)	**0.74** (0.53)	**0.73** (0.53)	**0.75** (0.53)	**0.74** (0.6)	**0.74** (0.6)	**0.74** (0.6)

Table A.14: Characteristics of training with Weight Decay (L1) considering the final SD model for Zugspitze winter precipitation. See table description A.7 for detailed information on setup and keywords.

η	0.5	0.5	0.5	0.5	0.5	0.1	0.1
d	0.5	0.1	0.01	0.001	0.0001	0.5	0.1
nt/nf	1/0	0/0	0/0	0/0	0/0	0/0	0/0
Epochs	506	637	352	582	976	528	450
rt	00:00:06	00:00:08	00:00:05	00:00:07	00:00:12	00:00:07	00:00:06
wr	-33.9/4.8	-0.6/0.1	-0.5/0.3	-0.6/0.5	-0.5/0.7	-0.2/0.2	-0.2/0.2
swf	0.47	0.12	0.13	0.14	0.25	0.06	0.08
MSSS	-0.19	0.01	0.02	0.19	0.5	0.02	0.02
r_s	0.3	0.36	0.15	0.57	0.69	0.31	0.25
η	0.1	0.1	0.1	0.01	0.01	0.01	0.01
d	0.01	0.001	0.0001	0.5	0.1	0.01	0.001
nt/nf	0/0	0/0	0/0	0/0	0/0	0/0	0/2
Epochs	3128	35491	22583	586	396	1105	190558
rt	00:00:36	00:06:58	00:04:26	00:00:07	00:00:05	00:00:13	00:36:42
wr	-0.3/0.3	-0.7/0.8	-0.6/1.1	-0.2/0.2	-0.2/0.2	-0.2/0.2	-0.6/0.7
swf	0.07	0.07	0.07	0.06	0.07	0.06	0.07
MSSS	0.2	0.63	**0.67**	-0.13	0	0.1	0.62 (0.2)
r_s	0.59	0.72	**0.74**	0	0.18	0.49	0.72 (0.22)
η	0.01	0.001	0.001	0.001	0.001	0.001	
d	0.0001	0.5	0.1	0.01	0.001	0.0001	
nt/nf	0/0	0/0	0/0	0/0	0/5	0/7	
Epochs	202159	626	1060	98531	277171	798406	
rt	00:39:60	00:00:08	00:00:13	00:19:20	00:53:30	02:34:34	
wr	-0.6/1	-0.2/0.2	-0.2/0.2	-0.2/0.2	-0.4/0.6	-0.4/1	
swf	0.07	0.06	0.08	0.07	0.07	0.08	
MSSS	**0.67**	-0.34	-0.02	0.14	0.56 (0.33)	**0.68** (0.5)	
r_s	**0.73**	0	0.11	0.5	0.72 (0.44)	**0.75** (0.6)	

Table A.15: Characteristics of training with Weight Decay (L2) considering the final SD model for Zugspitze winter precipitation. See table description A.7 for detailed information on setup and keywords.

η	0.5	0.5	0.5	0.5	0.5	0.1	0.1
d	0.5	0.1	0.01	0.001	0.0001	0.5	0.1
nt/nf	0/0	0/0	0/0	0/0	0/0	1/5	0/2
Epochs	370	12549	15668	16456	2895	303	2736
rt	00:00:05	00:02:21	00:03:17	00:03:11	00:00:35	00:00:04	00:00:33
wr	-0.2/0.2	-0.9/0.9	-0.6/0.7	-0.8/1	-0.7/0.8	-0.2/0.2	-0.2/0.6
swf	0.1	0.11	0.11	0.22	0.31	0.15	0.07
MSSS	0.02	0.48	0.52	0.59	0.58	0.04 (0.03)	0.25 (0.09)
r_s	0.29	0.71	0.71	0.71	0.7	0.23 (0.15)	0.6 (0.19)
η	0.1	0.1	0.1	0.01	0.01	0.01	0.01
d	0.01	0.001	0.0001	0.5	0.1	0.01	0.001
nt/nf	0/5	0/0	0/0	0/5	0/3	0/10	0/6
Epochs	185750	158200	24718	439	1176	(10^6)	433654
rt	00:35:54	00:28:41	00:04:32	00:00:05	00:00:13	(03:03:01)	01:24:08
wr	-0.8/1.7	-0.8/1.2	-0.6/0.9	-0.2/0.2	-0.2/0.2	(-1.06/1.14)	-0.6/0.8
swf	0.08	0.08	0.16	0.1	0.06	(0.09)	0.07
MSSS	0.65 (0.38)	**0.67**	**0.67**	0.03 (0.16)	0.07 (0.02)	(0.08)	**0.68** (0.42)
r_s	**0.73** (0.44)	**0.74**	**0.73**	0.21 (0.23)	0.53 (0.19)	(0.54)	**0.74** (0.51)
η	0.01	0.001	0.001	0.001	0.001	0.001	
d	0.0001	0.5	0.1	0.01	0.001	0.0001	
nt/nf	0/1	0/2	0/2	0/10	0/9	0/7	
Epochs	158594	1312	9232	(10^6)	926791	617438	
rt	00:27:45	00:00:14	00:01:36	(02:58:28)	02:59:35	01:56:41	
wr	-0.4/0.9	-0.2/0.2	-0.2/0.2	(-0.49/0.49)	-0.4/0.9	-0.4/0.7	
swf	0.16	0.11	0.06	(0.06)	0.07	0.18	
MSSS	**0.67** (0.14)	0.05 (0.01)	0.07 (0.03)	(0.06)	**0.67** (0.53)	**0.67** (0.52)	
r_s	**0.73** (0.15)	0.28 (0.09)	0.52 (0.17)	(0.61)	**0.73** (0.72)	**0.74** (0.61)	

Table A.16: Characteristics of training with the Rprop algorithm considering the final SD model for Zugspitze winter temperature. See table description A.7 for detailed information on setup and keywords.

nt/nf	0/0
Epochs	1488
rt	00:00:13
wr	-1.6/2
swf	0.25
MSSS	**0.89**
r_s	**0.94**

Table A.17: Characteristics of training with the Extended Delta-Bar-Delta algorithm considering the final SD model for Zugspitze winter temperature. See table description A.7 for detailed information on setup and keywords.

η	0.5	0.1	0.01	0.001
nt/nf	0/0	0/0	0/0	0/0
Epochs	328	8735	12283	11364
rt	00:00:03	00:01:02	00:01:26	00:01:21
wr	-0.4/1.7	-0.8/0.5	-0.7/0.5	-0.7/0.6
swf	0	0.24	0.23	0.23
MSSS	0	**0.89**	**0.89**	**0.89**
r_s	-0.07	**0.94**	**0.94**	**0.94**

Table A.18: Characteristics of training with Gradient Descent considering the final SD model for Zugspitze winter temperature. See table description A.7 for detailed information on setup and keywords.

η	0.5	0.1	0.01	0.001
nt/nf	0/0	0/0	0/8	0/10
Epochs	661	87934	579123	(10^6)
rt	00:00:05	00:11:20	01:16:59	(02:12:14)
wr	-1/0.9	-0.7/0.6	-0.7/0.7	(-0.51/0.48)
swf	0.43	0.23	0.23	(0.24)
MSSS	0.84	**0.89**	**0.89** (0.81)	(0.88)
r_s	0.91	**0.94**	**0.94** (0.85)	(0.94)

Table A.19: Characteristics of training with the Adaptive Learning Rate algorithm considering the final SD model for Zugspitze winter temperature. See table description A.7 for detailed information on setup and keywords.

η	0.5	0.1	0.01	0.001
nt/nf	0/0	0/0	0/0	0/0
Epochs	19545	19802	22924	17689
rt	00:02:19	00:02:21	00:02:42	00:02:05
wr	-0.5/0.9	-0.5/0.5	-0.5/0.5	-0.5/0.4
swf	0.29	0.28	0.28	0.27
MSSS	**0.89**	**0.89**	**0.89**	**0.89**
r_s	**0.94**	**0.94**	**0.94**	**0.94**

Table A.20: Characteristics of training with Individual Adaptive Learning Rate algorithm considering the final SD model for Zugspitze winter temperature. See table description A.7 for detailed information on setup and keywords.

η	0.5	0.1	0.01	0.001
nt/nf	0/0	0/0	0/0	0/0
Epochs	4041	4207	2668	2742
rt	00:00:29	00:00:30	00:00:19	00:00:19
wr	-2.8/2.2	-1.6/2.3	-1.4/2.1	-1.5/1.8
swf	0.19	0.22	0.18	0.22
MSSS	0.8	**0.89**	0.71	0.69
r_s	0.76	**0.94**	0.87	0.91

Table A.21: Characteristics of training with Decaying Eta considering the final SD model for Zugspitze winter temperature. See table description A.7 for detailed information on setup and keywords.

η	0.5	0.5	0.5	0.5	0.5	0.1	0.1	0.1	0.1	0.1
de	0.9999	0.999	0.99	0.9	0.5	0.9999	0.999	0.99	0.9	0.5
nt/nf	0/0	0/0	0/0	0/0	4/0	0/0	0/0	0/0	0/0	0/0
Epochs	25047	27425	3604	611	356	258714	29938	3545	606	349
rt	00:02:46	00:03:13	00:00:26	00:00:05	00:00:03	00:30:29	00:03:31	00:00:25	00:00:05	00:00:03
wr	-0.9/1	-1.2/1	-0.9/0.6	-0.7/0.5	-0.3/0.2	-0.4/0.5	-0.4/0.4	-0.2/0.3	-0.2/0.3	-0.2/0.3
swf	0.39	0.43	0.42	0.4	0.35	0.25	0.23	0.23	0.24	0.24
MSSS	0.85	0.86	0.36	0.15	-9.97	0.88	0.86	0.53	0.2	-0.12
r_s	0.92	0.92	0.83	0.47	0.76	0.93	0.92	0.86	0.53	0.34
η	0.01	0.01	0.01	0.01	0.01	0.001	0.001	0.001	0.001	0.001
de	0.9999	0.999	0.99	0.9	0.5	0.9999	0.999	0.99	0.9	0.5
nt/nf	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
Epochs	267150	30791	3401	626	351	278080	29073	3553	619	348
rt	00:31:22	00:03:36	00:00:24	00:00:05	00:00:03	00:32:34	00:03:24	00:00:25	00:00:05	00:00:03
wr	-0.4/0.4	-0.3/0.3	-0.2/0.3	-0.2/0.3	-0.3/0.3	-0.3/0.4	-0.2/0.3	-0.2/0.3	-0.2/0.3	-0.3/0.2
swf	0.23	0.24	0.24	0.24	0.24	0.24	0.24	0.24	0.23	0.24
MSSS	0.86	0.51	0.25	-4.02	-16.1	0.52	0.14	-2.73	-11.8	-13.62
r_s	0.92	0.85	0.59	0.37	0.23	0.83	0.37	0.29	-0.32	0.25

Table A.22: Characteristics of training with Momentum Term considering the final SD model for Zugspitze winter temperature. See table description A.7 for detailed information on setup and keywords.

η	0.5	0.5	0.5	0.5	0.5	0.5	0.5
α	0.99	0.5	0.1	0.05	0.01	0.001	0.0001
nt/nf	0/0	0/0	0/0	0/0	0/0	0/0	0/0
Epochs	33075	14657	1438	871	749	657	711
rt	00:04:24	00:01:56	00:00:12	00:00:07	00:00:06	00:00:06	00:00:06
wr	-0.8/0.9	-0.7/0.8	-0.9/1	-1/0.9	-1.2/0.9	-1.1/0.9	-1.2/0.9
swf	0.3	0.26	0.39	0.44	0.43	0.42	0.44
MSSS	0.67	**0.89**	0.87	0.85	0.84	0.83	0.84
r_s	0.86	**0.94**	0.93	0.92	0.92	0.91	0.92
η	0.1	0.1	0.1	0.1	0.1	0.1	0.1
α	0.99	0.5	0.1	0.05	0.01	0.001	0.0001
nt/nf	0/0	0/0	0/0	0/0	0/0	0/0	0/0
Epochs	121263	185996	162049	121662	127121	120637	115724
rt	00:15:49	00:23:53	00:20:31	00:15:30	00:16:12	00:15:18	00:14:53
wr	-0.7/1	-0.9/0.8	-0.8/0.8	-0.8/0.8	-0.8/0.7	-0.8/0.8	-0.7/0.8
swf	0.27	0.25	0.23	0.24	0.24	0.23	0.24
MSSS	0.75	**0.9**	**0.9**	**0.89**	**0.9**	**0.89**	**0.89**
r_s	0.9	**0.94**	**0.94**	**0.94**	**0.94**	**0.94**	**0.94**
η	0.01	0.01	0.01	0.01	0.01	0.01	0.01
α	0.99	0.5	0.1	0.05	0.01	0.001	0.0001
nt/nf	0/3	0/7	0/4	0/4	0/5	0/4	0/4
Epochs	243701	655582	654497	753127	697936	603505	754998
rt	00:30:20	01:23:02	01:23:00	01:35:53	01:29:19	01:16:55	01:37:07
wr	-0.3/0.4	-0.6/0.7	-0.6/0.6	-0.6/0.7	-0.6/0.7	-0.6/0.6	-0.7/0.6
swf	0.24	0.24	0.23	0.24	0.24	0.23	0.23
MSSS	0.41 (0.19)	**0.9** (0.72)	**0.9** (0.45)	**0.9** (0.45)	**0.89** (0.54)	**0.89** (0.45)	**0.9** (0.45)
r_s	0.77 (0.33)	**0.94** (0.76)	**0.94** (0.47)	**0.94** (0.47)	**0.94** (0.57)	**0.94** (0.47)	**0.94** (0.47)
η	0.001	0.001	0.001	0.001	0.001	0.001	0.001
α	0.99	0.5	0.1	0.05	0.01	0.001	0.0001
nt/nf	0/4	0/10	0/10	0/10	0/10	0/10	0/10
Epochs	575	(10^6)	(10^6)	(10^6)	(10^6)	(10^6)	(10^6)
rt	00:00:05	(02:10:22)	(02:10:20)	(02:13:21)	(02:11:18)	(02:14:11)	(02:14:19)
wr	-0.2/0.3	(-0.53/0.5)	(-0.46/0.48)	(-0.47/0.52)	(-0.5/0.51)	(-0.51/0.46)	(-0.46/0.53)
swf	0.24	(0.24)	(0.24)	(0.24)	(0.24)	(0.23)	(0.23)
MSSS	0.28 (0.15)	(0.33)	(0.33)	(0.33)	(0.33)	(0.33)	(0.33)
r_s (0.81)	0.75 (0.4)	(0.82)	(0.82)	(0.82)	(0.82)	(0.82)	(0.82)

Table A.23: Characteristics of training with Weight Decay (L1) considering the final SD model for Zugspitze winter temperature. See table description A.7 for detailed information on setup and keywords.

η	0.5	0.5	0.5	0.5	0.5	0.1	0.1
d	0.5	0.1	0.01	0.001	0.0001	0.5	0.1
nt/nf	0/0	0/0	0/0	0/0	0/0	0/0	0/0
Epochs	753	423	590	514	586	2107	1732
rt	00:00:06	00:00:04	00:00:05	00:00:05	00:00:05	00:00:17	00:00:14
wr	-1.3/0.6	-0.8/0.6	-0.8/0.8	-1/0.9	-0.9/0.9	-0.4/0.6	-0.4/0.7
swf	0.44	0.17	0.17	0.27	0.39	0.09	0.08
MSSS	0.55	0.33	0.65	0.77	0.83	0.75	0.76
r_s	0.81	0.77	0.87	0.9	0.92	0.88	0.9
η	0.1	0.1	0.1	0.01	0.01	0.01	0.01
d	0.01	0.001	0.0001	0.5	0.1	0.01	0.001
nt/nf	0/0	0/0	0/0	0/0	0/0	0/0	0/0
Epochs	5828	7713	81864	9671	10658	28965	81660
rt	00:00:47	00:01:03	00:11:06	00:01:19	00:01:24	00:03:51	00:10:46
wr	-0.5/0.7	-0.5/0.6	-0.8/0.9	-0.5/0.7	-0.4/0.4	-0.3/0.5	-0.5/0.6
swf	0.08	0.09	0.08	0.09	0.12	0.08	0.08
MSSS	0.81	0.86	**0.89**	0.67	0.41	0.52	0.86
r_s	0.92	0.92	**0.94**	0.85	0.72	0.89	0.92
η	0.01	0.001	0.001	0.001	0.001	0.001	
d	0.0001	0.5	0.1	0.01	0.001	0.0001	
nt/nf	0/2	0/0	1/0	0/0	0/0	0/9	
Epochs	312715	53276	991	45227	105196	991391	
rt	00:41:55	00:07:11	00:00:09	00:06:07	00:14:02	02:12:33	
wr	-0.7/0.7	-0.4/0.4	-0.2/0.3	-0.3/0.4	-0.4/0.5	-0.5/0.5	
swf	0.07	0.08	0.12	0.08	0.09	0.05	
MSSS	**0.89** (0.27)	-0.43	0.16	0.42	0.85	0.87 (0.88)	
r_s	**0.94** (0.28)	0.65	0.71	0.83	0.91	**0.94** (0.93)	

Table A.24: Characteristics of training with Weight Decay (L2) considering the final SD model for Zugspitze winter temperature. See table description A.7 for detailed information on setup and keywords.

η	0.5	0.5	0.5	0.5	0.5	0.1	0.1
d	0.5	0.1	0.01	0.001	0.0001	0.5	0.1
nt/nf	0/0	0/0	0/0	0/0	0/0	0/1	0/1
Epochs	323	802	874	731	645	40882	47626
rt	00:00:03	00:00:07	00:00:07	00:00:06	00:00:06	00:05:34	00:06:19
wr	-0.5/0.5	-0.9/0.6	-0.7/0.9	-1/0.9	-1/0.9	-0.9/0.9	-0.9/0.7
swf	0.13	0.15	0.13	0.37	0.39	0.09	0.06
MSSS	0.18	0.66	0.83	0.84	0.84	0.71 (0.16)	0.75 (0.16)
r_s	0.49	0.88	0.92	0.92	0.92	0.87 (0.18)	0.93 (0.19)
η	0.1	0.1	0.1	0.01	0.01	0.01	0.01
d	0.01	0.001	0.0001	0.5	0.1	0.01	0.001
nt/nf	0/1	0/1	0/1	0/6	0/6	0/6	0/7
Epochs	41027	95599	111639	228686	208763	526727	485698
rt	00:05:31	00:12:52	00:14:53	00:30:51	00:26:02	01:08:30	01:04:07
wr	-0.9/1	-0.8/0.9	-0.8/0.7	-0.5/0.5	-0.4/0.4	-0.9/0.9	-0.5/0.7
swf	0.07	0.07	0.09	0.12	0.04	0.04	0.03
MSSS	0.88 (0.18)	**0.89** (0.18)	**0.9** (0.18)	0.35 (0.21)	0.41 (0.23)	0.85 (0.55)	0.88 (0.72)
r_s	0.93 (0.19)	**0.94** (0.19)	**0.94** (0.19)	0.84 (0.58)	0.88 (0.61)	0.93 (0.64)	0.93 (0.75)
η	0.01	0.001	0.001	0.001	0.001	0.001	
d	0.0001	0.5	0.1	0.01	0.001	0.0001	
nt/nf	0/9	0/7	0/7	0/10	0/10	0/10	
Epochs	761009	952	3297	(10^6)	(10^6)	(10^6)	
rt	01:36:36	00:00:08	00:00:26	(02:06:53)	(02:03:11)	(02:01:10)	
wr	-0.9/0.8	-0.2/0.2	-0.2/0.2	(-0.66/0.63)	(-0.45/0.53)	(-0.52/0.49)	
swf	0.1	0.15	0.16	(0.06)	(0.06)	(0.22)	
MSSS	**0.9** (0.9)	0.16 (0.2)	0.33 (0.23)	(0.37)	(0.37)	(0.37)	
r_s	**0.94** (0.94)	0.74 (0.67)	0.87 (0.69)	(0.84)	(0.84)	(0.84)	

B

Appendix

Table B.1: Statistics of precipitation indices at Zugspitze considering the observational record (Obs) as well as downscaling products based on the reanalysis dataset (Rean, calibration and validation) and ESM experiments (historical: Hist; scenarios: RCP 8.5 and RCP 4.5) separated by season and annually. The table includes the wet day (> 1 mm) mean (SDI), mean annual maximum five day sum (P5max) and frequency [%] of values > 90 % Obs quantile (Pf90). Results from cross-validation folds and ESMs are averaged. Number in brackets in Hist: standard deviation between ESMs. Number in brackets in scenarios: standard deviation of the change signal. Reference periods: 1971-2000 (Obs, Rean, Hist); 2021-2050 (NF); 2071-2100 (DF). Changes: Increasing values are marked in green, declining ones in red and insignificant ones in grey (more than two ESMs show no change signal or a different sign.). Change signals are computed by comparing historical and scenario results. Relative change values: SDI, P5max. Absolute change values: Pf90.

		Wet day (> 1 mm) mean (SDI) [mm]							**Mean annual maximum five day sum (P5max) [mm]**							**Frequency [%] of values > 90 % Obs quantile (Pf90)**						
					RCP 4.5		RCP 8.5					RCP 4.5		RCP 8.5					RCP 4.5		RCP 8.5	
		Obs	Rean	Hist	NF	DF	NF	DF	Obs	Rean	Hist	NF	DF	NF	DF	Obs	Rean	Hist	NF	DF	NF	DF
DJF	ANN	12.4	8.6	9.2 (0.3)	6.7 (5.8)	9.2 (5.8)	6.3 (3.1)	14.6 (7.2)	129.4	117.8	112.3 (11.0)	8.8 (5.8)	8.3 (6.7)	8.7 (5.4)	11.0 (8.6)	10.0	8.2	9.9 (1.5)	1.6 (1.7)	2.0 (1.5)	1.5 (0.9)	3.2 (1.9)
	RCF	12.4	7.6	7.2 (0.6)	2.0 (5.9)	1.7 (5.8)	0.8 (2.9)	0.2 (6.6)	129.4	108.4	93.4 (11.4)	4.7 (5.8)	2.9 (6.4)	2.9 (5.3)	3.3 (7.3)	10.0	8.4	7.3 (1.5)	0.4 (1.1)	0.2 (1.1)	0.2 (0.7)	0.1 (1.2)
	Comb	12.4	9.5	9.8 (1.0)	4.0 (6.9)	4.0 (4.2)	3.4 (2.9)	7.9 (6.3)	129.4	120.9	112.9 (10.2)	7.5 (6.0)	5.9 (5.5)	6.7 (4.5)	8.7 (8.3)	10.0	9.8	11.6 (1.2)	1.0 (1.5)	1.3 (1.9)	1.0 (1.0)	2.4 (2.0)
MAM	ANN	11.6	7.8	8.8 (1.0)	10.3 (5.6)	8.3 (5.8)	5.9 (4.5)	6.6 (5.7)	127.1	95.9	118.1 (15.6)	7.5 (9.6)	7.0 (10.3)	1.4 (7.6)	0.0 (8.5)	10.0	6.3	8.3 (2.2)	2.3 (1.3)	1.8 (1.2)	1.2 (0.8)	1.5 (1.1)
	RCF	11.6	7.0	7.3 (0.1)	2.1 (2.2)	-0.3 (2.6)	-0.7 (2.6)	-3.7 (2.8)	127.1	79.9	82.5 (1.0)	-0.8 (2.9)	-1.6 (3.8)	-2.3 (2.3)	-7.8 (2.8)	10.0	4.6	4.7 (0.2)	0.1 (0.5)	-0.3 (0.5)	-0.3 (0.2)	-1.0 (0.5)
	Comb	11.6	8.1	9.0 (0.6)	8.5 (5.6)	6.4 (5.9)	4.4 (4.0)	2.9 (5.8)	127.1	107.2	125.8 (11.0)	9.1 (7.7)	10.0 (11.0)	3.3 (8.5)	1.2 (8.5)	10.0	7.1	9.1 (1.3)	1.5 (1.0)	1.2 (1.4)	0.6 (0.9)	0.6 (1.3)
JJA	ANN	10.8	7.1	7.7 (0.8)	2.4 (4.6)	-0.1 (4.8)	2.5 (3.1)	-1.9 (6.6)	108.3	84.6	80.7 (5.0)	8.5 (6.0)	5.9 (5.9)	9.0 (5.3)	4.1 (8.5)	10.0	4.8	6.7 (2.2)	0.7 (0.8)	0.0 (0.9)	0.6 (0.5)	-0.3 (1.3)
	RCF	10.8	6.5	6.7 (0.3)	-2.1 (3.1)	-5.5 (2.6)	-2.5 (1.9)	-12.6 (2.7)	108.3	76.9	71.8 (1.3)	1.4 (6.1)	-2.1 (5.6)	1.7 (5.9)	-8.3 (3.3)	10.0	4.4	4.7 (0.7)	-0.1 (0.4)	-0.7 (0.4)	-0.2 (0.4)	-1.7 (0.5)
	Comb	10.8	7.3	7.6 (0.5)	5.2 (5.3)	4.2 (5.2)	4.9 (3.0)	6.2 (3.8)	108.3	89.9	87.5 (2.4)	8.6 (8.6)	7.3 (7.2)	5.9 (6.2)	7.6 (7.6)	10.0	7.1	8.4 (1.2)	0.8 (1.0)	0.5 (1.1)	0.7 (0.7)	0.6 (0.9)
SON	ANN	11.6	7.3	7.5 (0.6)	6.9 (4.5)	7.7 (6.4)	7.7 (3.6)	4.0 (6.5)	107.2	97.6	86.6 (10.6)	8.8 (8.7)	9.4 (10.0)	9.9 (8.5)	4.8 (11.7)	10.0	7.0	7.4 (1.7)	1.3 (0.8)	1.2 (1.2)	1.4 (0.7)	0.6 (1.0)
	RCF	11.6	6.6	6.3 (0.4)	2.7 (4.4)	-0.4 (6.8)	3.2 (3.0)	-8.5 (7.0)	107.2	85.3	77.9 (6.8)	4.1 (5.6)	0.4 (7.3)	5.3 (5.1)	-5.8 (8.4)	10.0	7.1	6.4 (1.1)	0.4 (0.7)	0.0 (1.1)	0.6 (0.5)	-1.1 (1.0)
	Comb	11.6	8.2	8.5 (0.4)	6.9 (4.5)	9.0 (5.0)	8.3 (2.7)	8.3 (3.9)	107.2	96.5	89.7 (8.5)	7.5 (8.6)	7.2 (10.9)	8.1 (5.4)	3.6 (10.4)	10.0	7.5	8.5 (1.2)	1.0 (1.1)	1.1 (1.1)	1.3 (0.8)	0.6 (0.8)
Annual	ANN	11.6	7.7	8.3 (0.5)	6.6 (3.6)	6.4 (3.2)	5.5 (1.8)	6.5 (3.8)	168.8	136.2	137.0 (8.7)	9.0 (5.4)	8.1 (6.5)	7.9 (5.5)	5.1 (6.8)	10.0	6.6	8.2 (1.4)	1.5 (0.8)	1.3 (0.7)	1.2 (0.4)	1.3 (0.9)
	RCF	11.6	6.9	6.9 (0.3)	1.1 (2.8)	-1.1 (2.4)	0.0 (1.1)	-6.1 (1.7)	168.8	117.3	104.5 (7.1)	4.9 (4.4)	2.7 (3.7)	4.8 (4.3)	1.8 (4.0)	10.0	6.1	5.9 (0.7)	0.2 (0.5)	-0.2 (0.4)	0.0 (0.2)	-0.9 (0.4)
	Comb	11.6	8.2	8.7 (0.4)	6.4 (3.9)	6.2 (3.0)	5.3 (1.9)	6.9 (3.2)	168.8	144.2	142.2 (8.7)	8.9 (5.8)	8.8 (6.0)	6.8 (6.4)	4.8 (5.8)	10.0	8.0	9.4 (1.0)	1.1 (0.8)	1.0 (0.9)	0.9 (0.5)	1.1 (0.8)

Table B.2: Statistics of precipitation indices at Zugspitze, see table B.1 for details on structure and abbreviations. Table includes the wet day frequency (WDF), consecutive dry days (CDD) and consecutive wet days (CWD). Absolute change values: WDF, CDD, CWD.

		Wet day (> 1 mm) frequency [%]							Consecutive dry days (CDD)							Consecutive wet days (CWD)						
					RCP 4.5		RCP 8.5					RCP 4.5		RCP 8.5					RCP 4.5		RCP 8.5	
		Obs	Rean	Hist	NF	DF	NF	DF	Obs	Rean	Hist	NF	DF	NF	DF	Obs	Rean	Hist	NF	DF	NF	DF
DJF	ANN	45.7	68.1	86.0	-0.4	1.3	0.3	3.1	12.5	6.5	3.3	0.2	-0.1	-0.1	-0.5	9.9	16.9	29.4	-0.8	0.9	-0.4	3.2
				(3.9)	(1.8)	(2.4)	(2.0)	(2.7)			(0.8)	(0.6)	(0.5)	(0.5)	(0.6)			(4.9)	(1.6)	(2.3)	(1.8)	(3.4)
	RCF	45.7	75.0	77.7	0.9	1.8	0.9	4.1	12.5	5.1	4.8	0.0	-0.2	-0.2	-0.8	9.9	20.0	20.4	0.3	1.3	0.1	2.5
				(1.1)	(1.6)	(2.4)	(2.5)	(2.9)			(0.4)	(0.6)	(0.6)	(0.8)	(0.7)			(1.1)	(0.8)	(1.0)	(0.8)	(1.7)
	Comb	45.7	61.6	69.1	1.6	4.4	2.0	7.9	12.5	8.6	7.1	-0.4	-1.1	-0.7	-1.6	9.9	15.4	16.6	0.5	1.7	0.6	3.1
				(5.1)	(2.5)	(4.0)	(3.3)	(4.7)			(1.7)	(1.2)	(0.9)	(1.2)	(1.5)			(1.0)	(0.5)	(1.3)	(1.1)	(2.3)
MAM	ANN	51.8	78.0	82.0	-0.7	-1.0	-1.8	-2.2	9.0	4.5	4.7	0.1	0.1	0.4	0.3	9.5	21.9	27.7	-1.4	-1.4	-3.0	-2.9
				(3.5)	(1.2)	(1.5)	(2.1)	(1.7)			(0.2)	(0.3)	(0.5)	(0.4)	(0.4)			(4.2)	(1.4)	(1.7)	(2.9)	(1.8)
	RCF	51.8	85.4	85.7	-0.4	-0.6	-1.1	-1.9	9.0	3.6	3.7	0.1	0.0	0.2	0.2	9.5	26.7	29.1	-0.8	-0.9	-1.6	-3.6
				(0.6)	(0.9)	(1.2)	(1.7)	(1.8)			(0.2)	(0.3)	(0.4)	(0.2)	(0.3)			(0.4)	(1.0)	(1.6)	(2.4)	(1.9)
	Comb	51.8	78.2	79.6	0.8	0.9	-0.4	1.6	9.0	4.3	5.0	-0.1	-0.1	0.2	-0.3	9.5	20.7	24.8	0.1	0.5	-0.9	0.8
				(4.3)	(1.6)	(2.0)	(2.6)	(2.3)			(0.5)	(0.3)	(0.5)	(0.5)	(0.4)			(3.9)	(1.0)	(2.8)	(2.3)	(1.9)
JJA	ANN	53.6	83.0	90.6	-0.9	-0.7	-0.5	0.2	7.2	3.0	2.1	0.1	0.1	0.1	0.0	9.2	21.3	32.2	-0.8	-1.3	0.1	0.9
				(1.5)	(1.1)	(1.2)	(1.1)	(1.1)			(0.2)	(0.2)	(0.2)	(0.2)	(0.2)			(4.7)	(1.9)	(1.7)	(3.7)	(3.7)
	RCF	53.6	90.0	89.7	0.5	0.6	1.0	2.1	7.2	2.4	2.3	0.0	-0.1	-0.1	-0.3	9.2	34.2	30.9	1.0	1.5	2.3	6.3
				(0.6)	(1.0)	(1.1)	(1.0)	(1.2)			(0.1)	(0.4)	(0.3)	(0.3)	(0.4)			(3.0)	(1.4)	(1.9)	(3.7)	(3.3)
	Comb	53.6	82.9	86.7	-2.3	-3.0	-2.1	-4.4	7.2	3.2	2.7	0.4	0.6	0.4	0.8	9.2	21.4	25.6	-2.3	-3.2	-2.5	-4.3
				(0.7)	(1.3)	(1.5)	(1.4)	(2.1)			(0.1)	(0.2)	(0.3)	(0.3)	(0.4)			(2.4)	(1.4)	(0.9)	(1.5)	(2.1)
SON	ANN	40.1	66.9	69.1	-0.9	-1.9	-1.2	-2.8	13.1	7.2	6.9	0.3	0.4	0.0	0.2	7.6	14.9	16.1	-0.2	-0.6	-0.6	-1.8
				(3.1)	(1.7)	(1.5)	(1.3)	(1.9)			(0.6)	(0.7)	(0.4)	(0.5)	(0.5)			(2.0)	(1.0)	(0.8)	(0.8)	(0.8)
	RCF	40.1	72.2	74.2	0.1	-0.8	-0.2	-0.9	13.1	7.2	6.4	0.2	0.4	0.0	0.0	7.6	19.6	19.3	0.2	0.1	-0.2	-0.4
				(1.7)	(1.0)	(0.6)	(1.7)	(2.1)			(0.7)	(0.6)	(0.4)	(0.5)	(0.7)			(1.1)	(0.8)	(0.7)	(1.2)	(1.4)
	Comb	40.1	58.3	61.9	-0.1	-1.1	-0.4	-1.0	13.1	9.0	8.4	0.2	0.3	-0.1	-0.2	7.6	12.9	13.2	0.2	-0.3	-0.2	-0.5
				(3.9)	(1.3)	(1.5)	(1.4)	(1.6)			(1.0)	(0.9)	(0.6)	(0.7)	(0.7)			(1.3)	(0.8)	(1.0)	(0.9)	(1.2)
Annual	ANN	47.8	74.1	81.9	-0.7	-0.6	-0.8	-0.4	16.5	8.8	7.6	0.4	0.4	0.1	0.3	13.1	28.0	43.6	-0.9	-0.9	-1.1	1.1
				(2.3)	(0.5)	(0.9)	(1.0)	(1.2)			(0.5)	(0.5)	(0.5)	(0.4)	(0.4)			(7.1)	(1.3)	(2.1)	(2.8)	(3.7)
	RCF	47.8	80.7	81.9	0.3	0.2	0.2	0.8	16.5	8.1	7.3	0.3	0.3	0.0	-0.2	13.1	40.5	39.7	0.8	1.1	1.6	4.4
				(1.0)	(0.3)	(0.7)	(0.9)	(0.9)			(0.5)	(0.5)	(0.5)	(0.5)	(0.6)			(2.6)	(1.4)	(1.7)	(2.4)	(3.3)
	Comb	47.8	70.3	74.4	0.0	0.3	-0.2	1.0	16.5	11.1	10.1	-0.1	-0.4	-0.5	-0.9	13.1	27.7	33.7	-1.4	-1.0	-2.2	-1.2
				(3.3)	(0.8)	(1.2)	(1.2)	(1.8)			(1.3)	(0.8)	(0.8)	(0.4)	(0.7)			(3.4)	(1.1)	(2.0)	(1.2)	(1.9)

Table B.3: Statistics of precipitation indices at Sonnblick, see table B.1 for details on structure and abbreviations. Table includes the wet day (> 1 mm) mean (SDI), mean annual maximum five day sum (P5max) and frequency [%] of values > 90 % Obs quantile (Pf90). Relative change values: SDI, P5max. Absolute change values: Pf90.

		Wet day (> 1 mm) mean (SDI) [mm]							**Mean annual maximum five day sum (P5max) [mm]**							**Frequency [%] of values > 90 % Obs quantile (q90)**						
					RCP 4.5		RCP 8.5					RCP 4.5		RCP 8.5					RCP 4.5		RCP 8.5	
		Obs	Rean	Hist	NF	DF	NF	DF	Obs	Rean	Hist	NF	DF	NF	DF	Obs	Rean	Hist	NF	DF	NF	DF
DJF	ANN	8.6	6.0	6.2 (0.4)	8.6 (2.0)	14.9 (2.9)	9.5 (2.3)	20.6 (2.7)	77.2	67.0	59.3 (6.3)	10.2 (3.0)	16.7 (3.7)	12.6 (3.0)	21.0 (4.2)	10.0	6.8	7.1 (1.7)	2.2 (0.5)	4.0 (1.3)	2.6 (1.1)	5.6 (1.1)
	RCF	8.6	5.6	5.3 (0.4)	4.0 (3.7)	6.0 (4.2)	3.5 (2.2)	4.7 (2.6)	77.2	61.1	55.0 (6.7)	4.8 (4.2)	4.4 (5.7)	5.2 (3.1)	4.2 (5.9)	10.0	5.8	4.9 (1.4)	0.6 (0.6)	0.8 (0.9)	0.5 (0.6)	0.7 (0.6)
	Comb	8.6	6.6	6.9 (0.4)	4.2 (5.7)	5.8 (5.3)	3.7 (4.0)	4.5 (6.8)	77.2	67.6	71.0 (4.1)	3.3 (3.3)	4.8 (5.1)	3.3 (3.1)	4.9 (6.0)	10.0	11.0	13.4 (0.8)	0.8 (1.6)	1.4 (1.7)	0.8 (1.1)	1.2 (2.2)
MAM	ANN	8.6	6.1	5.8 (0.3)	8.5 (2.4)	7.3 (2.3)	6.2 (2.8)	6.7 (1.1)	83.4	64.1	58.3 (3.3)	9.4 (5.7)	7.5 (3.6)	7.5 (4.2)	4.8 (5.8)	10.0	4.9	3.3 (0.7)	1.6 (0.4)	1.3 (0.4)	1.1 (0.4)	1.3 (0.4)
	RCF	8.6	5.6	5.7 (0.1)	3.4 (1.8)	2.1 (1.7)	1.9 (2.4)	1.3 (2.1)	83.4	54.1	53.0 (1.0)	1.7 (2.5)	1.6 (3.8)	1.0 (3.1)	0.7 (3.0)	10.0	2.4	2.3 (0.1)	0.2 (0.4)	0.2 (0.4)	0.1 (0.2)	0.2 (0.4)
	Comb	8.6	6.4	6.1 (0.2)	6.8 (1.6)	5.5 (1.3)	4.6 (2.2)	4.4 (1.9)	83.4	66.3	60.7 (2.3)	8.1 (3.0)	7.2 (3.3)	5.1 (5.2)	3.8 (6.0)	10.0	6.1	5.1 (0.6)	1.5 (0.5)	1.1 (0.4)	1.0 (0.5)	0.8 (0.5)
JJA	ANN	8.9	5.9	6.5 (0.6)	0.9 (4.2)	-0.3 (4.6)	1.6 (1.8)	-3.1 (5.6)	75.0	57.4	58.3 (3.6)	3.9 (5.6)	3.6 (4.5)	5.3 (5.6)	-0.2 (5.6)	10.0	3.2	3.7 (1.3)	0.6 (0.8)	0.4 (0.7)	0.7 (0.5)	-0.1 (0.6)
	RCF	8.9	5.4	5.5 (0.1)	-1.3 (4.6)	-2.4 (4.0)	-0.6 (1.9)	-6.3 (4.7)	75.0	51.6	52.8 (1.0)	-0.1 (4.6)	-1.1 (2.1)	1.2 (2.9)	-5.2 (2.6)	10.0	2.1	2.2 (0.2)	0.1 (0.5)	0.0 (0.3)	0.1 (0.2)	-0.4 (0.3)
	Comb	8.9	6.0	6.2 (0.3)	0.0 (4.8)	-1.3 (5.6)	0.3 (1.8)	-4.7 (6.7)	75.0	58.0	57.7 (1.0)	0.1 (5.0)	-0.2 (3.6)	1.0 (4.1)	-4.5 (5.4)	10.0	3.8	4.0 (0.8)	0.2 (0.7)	-0.2 (0.6)	0.2 (0.3)	-0.8 (0.7)
SON	ANN	9.4	6.1	5.6 (0.3)	8.8 (2.0)	10.3 (4.7)	10.7 (1.8)	11.5 (5.5)	72.3	61.1	53.0 (3.4)	9.8 (3.9)	9.8 (6.8)	13.6 (3.9)	12.8 (3.1)	10.0	6.2	4.1 (0.9)	1.3 (0.6)	1.6 (0.8)	1.7 (0.5)	1.9 (0.9)
	RCF	9.4	5.4	5.4 (0.2)	-0.4 (3.1)	-4.8 (4.4)	-1.1 (1.9)	-14.6 (4.8)	72.3	53.2	53.7 (0.7)	1.0 (3.7)	-1.2 (5.0)	1.6 (2.6)	-7.4 (5.6)	10.0	2.8	2.7 (0.3)	0.3 (0.3)	0.2 (0.4)	0.3 (0.1)	-0.3 (0.2)
	Comb	9.4	6.7	6.2 (0.1)	6.4 (3.8)	5.3 (3.1)	7.3 (3.2)	0.3 (5.9)	72.3	62.7	60.7 (0.8)	9.5 (3.1)	9.4 (4.8)	13.6 (5.3)	9.6 (4.5)	10.0	7.4	6.3 (0.3)	1.6 (0.8)	1.7 (0.8)	2.0 (0.8)	1.1 (1.2)
Annual	ANN	8.8	6.0	6.1 (0.2)	6.4 (1.8)	7.6 (2.7)	6.6 (1.5)	8.5 (2.4)	101.4	77.5	70.0 (3.6)	7.7 (2.4)	10.6 (4.1)	9.1 (2.6)	11.2 (5.1)	10.0	5.2	4.4 (0.5)	1.4 (0.2)	1.7 (0.5)	1.5 (0.3)	1.9 (0.5)
	RCF	8.8	5.5	5.5 (0.1)	1.4 (2.3)	0.2 (1.6)	0.9 (1.0)	-3.6 (1.4)	101.4	67.2	64.0 (2.5)	2.6 (3.4)	2.0 (4.4)	3.7 (3.2)	-0.2 (5.2)	10.0	3.2	3.0 (0.3)	0.3 (0.3)	0.2 (0.3)	0.2 (0.1)	-0.1 (0.2)
	Comb	8.8	6.4	6.3 (0.2)	4.3 (2.5)	3.8 (2.2)	3.9 (1.0)	1.2 (2.0)	101.4	79.1	76.6 (2.8)	5.1 (2.0)	7.1 (3.8)	6.7 (1.9)	5.7 (2.9)	10.0	6.9	7.0 (0.5)	1.0 (0.5)	0.9 (0.4)	0.9 (0.2)	0.5 (0.5)

Table B.4: Statistics of precipitation indices at Sonnblick, see table B.1 for details on structure and abbreviations. Table includes the wet day frequency (WDF), consecutive dry days (CDD) and consecutive wet days (CWD). Absolute change values: WDF, CDD, CWD.

		Wet day (> 1 mm) frequency [%]							**Consecutive dry days (CDD)**							**Consecutive wet days (CWD)**						
					RCP 4.5		RCP 8.5					RCP 4.5		RCP 8.5					RCP 4.5		RCP 8.5	
		Obs	Rean	Hist	NF	DF	NF	DF	Obs	Rean	Hist	NF	DF	NF	DF	Obs	Rean	Hist	NF	DF	NF	DF
DJF	ANN	49.2	70.6	83.9 (4.4)	-1.4 (2.0)	-0.6 (2.7)	-0.8 (2.2)	-0.3 (3.1)	11.1	6.3	3.9 (0.9)	0.2 (0.6)	0.1 (0.6)	0.0 (0.5)	0.0 (0.7)	9.9	17.6	26.4 (4.6)	-1.6 (1.1)	-0.6 (1.7)	-1.0 (1.6)	-0.7 (2.4)
	RCF	49.2	74.6	77.5 (1.2)	0.5 (1.9)	2.5 (2.4)	1.3 (2.6)	4.5 (2.6)	11.1	5.9	5.4 (0.6)	0.0 (0.9)	-0.4 (0.7)	-0.3 (0.9)	-0.8 (0.7)	9.9	21.2	20.9 (1.1)	0.6 (1.0)	2.2 (2.2)	1.2 (1.4)	3.8 (1.9)
	Comb	49.2	65.3	71.6 (2.2)	-0.9 (2.1)	0.5 (2.8)	-0.4 (2.5)	0.3 (3.5)	11.1	8.1	6.2 (0.7)	0.1 (0.7)	-0.1 (0.5)	0.1 (0.8)	0.0 (0.9)	9.9	15.7	17.0 (2.3)	-0.2 (0.9)	0.1 (1.5)	-0.4 (0.9)	-0.1 (0.9)
MAM	ANN	57.3	80.5	87.2 (1.9)	-0.8 (1.1)	-1.2 (1.0)	-1.2 (1.6)	-2.1 (2.0)	8.0	4.3	3.4 (0.4)	0.1 (0.3)	0.3 (0.3)	0.3 (0.4)	0.3 (0.3)	12.1	23.3	33.4 (3.3)	-2.2 (1.6)	-1.2 (3.2)	-1.4 (2.8)	-2.9 (3.6)
	RCF	57.3	86.4	87.4 (0.6)	0.8 (1.0)	1.1 (0.8)	0.7 (1.5)	1.3 (1.6)	8.0	3.3	3.1 (0.2)	-0.1 (0.3)	0.0 (0.4)	0.0 (0.5)	-0.1 (0.4)	12.1	31.0	33.5 (0.5)	0.7 (1.2)	1.6 (1.8)	1.0 (3.2)	2.6 (3.5)
	Comb	57.3	77.9	82.6 (1.5)	0.2 (1.3)	0.2 (1.2)	-0.3 (2.1)	-0.1 (1.9)	8.0	4.8	4.2 (0.4)	0.0 (0.5)	0.0 (0.5)	0.2 (0.7)	0.0 (0.4)	12.1	21.8	25.7 (1.3)	-0.9 (0.7)	0.1 (2.0)	-0.4 (1.6)	-0.8 (2.5)
JJA	ANN	56.2	85.0	90.1 (1.5)	-1.8 (1.1)	-1.2 (1.5)	-1.3 (0.8)	-1.5 (2.0)	6.8	3.2	2.7 (0.2)	0.2 (0.3)	0.1 (0.4)	0.2 (0.3)	0.2 (0.4)	9.6	26.4	34.8 (5.0)	-2.4 (2.3)	-1.9 (3.3)	-1.6 (2.2)	-2.1 (4.3)
	RCF	56.2	92.1	91.8 (0.3)	0.6 (1.0)	0.9 (1.4)	0.9 (0.5)	1.7 (2.2)	6.8	2.2	2.3 (0.1)	-0.3 (0.2)	-0.2 (0.3)	-0.2 (0.1)	-0.4 (0.4)	9.6	39.2	38.0 (1.7)	0.7 (3.1)	0.6 (3.7)	1.8 (3.3)	4.9 (7.7)
	Comb	56.2	85.4	86.6 (1.0)	-0.9 (1.4)	-0.4 (1.7)	-0.4 (0.6)	-0.4 (3.1)	6.8	3.0	3.1 (0.2)	0.0 (0.4)	-0.1 (0.3)	0.1 (0.3)	-0.1 (0.5)	9.6	25.6	27.0 (3.0)	-0.6 (2.6)	-0.1 (3.1)	-0.1 (2.0)	0.7 (5.4)
SON	ANN	45.3	69.9	68.4 (3.7)	-0.9 (1.6)	-1.5 (1.0)	-0.9 (1.6)	-2.7 (1.8)	12.6	6.7	7.1 (1.0)	0.1 (0.5)	0.3 (0.3)	0.0 (0.6)	0.0 (0.6)	8.7	17.7	15.7 (1.2)	-0.5 (0.7)	-0.5 (1.0)	-0.1 (1.0)	-1.5 (1.1)
	RCF	45.3	76.6	78.6 (0.6)	0.1 (1.3)	-0.6 (0.9)	0.1 (1.6)	-0.3 (2.1)	12.6	5.7	5.3 (0.5)	-0.1 (0.4)	0.0 (0.5)	-0.4 (0.6)	-0.4 (0.8)	8.7	22.7	21.8 (0.9)	0.2 (0.7)	-0.4 (0.8)	0.1 (1.3)	-1.2 (1.5)
	Comb	45.3	65.0	70.9 (1.8)	0.7 (1.1)	0.8 (0.8)	0.6 (2.1)	2.5 (2.8)	12.6	7.9	6.4 (0.8)	-0.1 (0.4)	-0.1 (0.3)	-0.1 (0.5)	-0.8 (0.8)	8.7	15.0	16.3 (1.0)	0.5 (1.2)	0.9 (1.2)	0.5 (1.7)	1.5 (2.3)
Annual	ANN	52.0	76.6	82.5 (1.9)	-1.2 (0.5)	-1.1 (1.0)	-1.1 (1.0)	-1.7 (1.3)	15.7	8.3	7.7 (0.8)	0.2 (0.3)	0.3 (0.3)	0.0 (0.5)	0.0 (0.6)	14.9	32.7	47.4 (4.6)	-2.9 (2.2)	-2.1 (3.1)	-1.9 (2.7)	-2.2 (4.2)
	RCF	52.0	82.5	83.9 (0.6)	0.5 (0.4)	1.0 (0.9)	0.7 (1.0)	1.8 (1.2)	15.7	7.6	6.9 (0.6)	0.0 (0.4)	-0.3 (0.7)	-0.5 (0.7)	-0.8 (0.8)	14.9	46.2	47.2 (1.5)	1.4 (2.9)	1.8 (2.6)	2.6 (2.6)	6.3 (6.6)
	Comb	52.0	73.5	78.0 (1.0)	-0.2 (0.7)	0.3 (1.1)	-0.1 (1.3)	0.6 (1.6)	15.7	10.3	8.3 (0.7)	0.0 (0.4)	-0.1 (0.3)	0.0 (0.3)	-0.3 (0.7)	14.9	30.3	34.8 (1.6)	-0.3 (2.5)	0.6 (3.2)	0.2 (2.6)	1.5 (4.6)

C
Appendix

Table C.1: Statistics of temperature indices at Zugspitze, see table B.1 for details on structure and abbreviations. Table includes the annual minimum daily mean temperature (T1min), annual maximum daily mean temperature (T1max) and cold day frequency (CDF). Absolute change values: T1min, T1max, CDF.

		Annual minimum daily mean temperature [°C]							**Annual maximum daily mean temperature [°C]**							**Cold day (< 0 °C) frequency [%]**						
					RCP 4.5		RCP 8.5					RCP 4.5		RCP 8.5					RCP 4.5		RCP 8.5	
		Obs	Rean	Hist	NF	DF	NF	DF	Obs	Rean	Hist	NF	DF	NF	DF	Obs	Rean	Hist	NF	DF	NF	DF
DJF	ANN	-22.9	-22.8	-23.1	0.4	2.1	1.0	4.2	-0.6	-1.4	-0.9	2.3	3.0	2.3	4.8	99.5	99.7	99.4	-2.2	-3.7	-2.5	-10.2
				(0.8)	(1.1)	(1.0)	(1.2)	(1.5)			(0.8)	(0.7)	(0.5)	(0.5)	(0.6)			(0.4)	(0.9)	(1.7)	(0.9)	(3.6)
	RCF	-22.9	-22.8	-23.1	0.6	1.8	1.0	3.6	-0.6	-0.5	-0.5	1.2	1.6	1.2	2.0	99.5	99.3	99.2	-1.7	-2.6	-1.8	-6.0
				(0.3)	(0.6)	(0.9)	(0.9)	(1.2)			(0.2)	(0.4)	(0.3)	(0.3)	(0.2)			(0.2)	(0.8)	(0.8)	(0.7)	(1.8)
	NNC	-22.9	-23.5	-24.1	0.1	1.8	0.7	3.7	-0.6	-1.2	-1.1	1.5	2.1	1.7	3.4	99.5	99.7	99.7	-1.3	-2.3	-1.5	-6.5
				(0.5)	(1.1)	(1.0)	(1.2)	(1.3)			(0.4)	(0.4)	(0.4)	(0.4)	(0.4)			(0.2)	(0.7)	(1.4)	(0.8)	(3.2)
MAM	ANN	-18.4	-18.1	-18.9	0.1	1.0	0.9	2.9	3.2	2.6	2.5	0.9	1.4	1.0	2.4	91.9	92.0	92.3	-2.3	-5.9	-3.8	-11.6
				(0.6)	(0.6)	(1.0)	(0.4)	(0.5)			(0.4)	(0.6)	(0.5)	(0.3)	(0.4)			(1.8)	(0.4)	(2.3)	(1.4)	(3.4)
	RCF	-18.4	-18.5	-19.1	0.2	1.2	1.0	3.3	3.2	2.6	2.7	0.6	1.0	0.8	1.5	91.9	92.2	91.8	-2.8	-6.9	-4.5	-13.0
				(0.4)	(0.6)	(1.0)	(0.4)	(0.6)			(0.1)	(0.4)	(0.4)	(0.2)	(0.3)			(0.3)	(1.0)	(3.0)	(2.1)	(3.8)
	NNC	-18.4	-18.7	-19.8	0.0	1.0	0.8	3.2	3.2	3.2	3.2	0.9	1.5	1.1	2.6	91.9	91.4	92.2	-2.1	-5.7	-3.7	-11.1
				(0.4)	(0.8)	(1.1)	(0.5)	(0.6)			(0.4)	(0.5)	(0.6)	(0.4)	(0.3)			(1.1)	(0.7)	(2.2)	(2.0)	(3.2)
JJA	ANN	-6.8	-6.8	-6.5	0.5	1.4	1.0	3.2	9.6	9.0	8.5	1.9	2.8	2.2	4.7	33.1	30.9	37.5	-12.6	-18.8	-13.8	-27.9
				(0.7)	(0.6)	(0.6)	(0.7)	(1.3)			(0.2)	(0.7)	(0.7)	(0.7)	(0.9)			(7.9)	(3.3)	(3.4)	(3.6)	(3.6)
	RCF	-6.8	-6.9	-6.7	0.5	1.4	0.9	3.1	9.6	9.9	10.6	1.4	1.8	1.4	2.2	33.1	32.9	37.3	-12.8	-19.4	-14.3	-28.6
				(0.1)	(0.6)	(0.7)	(0.7)	(1.3)			(0.8)	(0.9)	(0.9)	(0.8)	(0.9)			(2.3)	(3.9)	(3.6)	(4.5)	(2.8)
	NNC	-6.8	-6.8	-6.6	0.6	1.3	1.0	3.0	9.6	9.1	9.2	1.3	2.1	1.6	3.0	33.1	31.2	38.1	-12.8	-19.0	-14.3	-28.7
				(0.2)	(0.5)	(0.4)	(0.6)	(1.1)			(0.4)	(0.6)	(0.5)	(0.5)	(0.6)			(4.6)	(3.7)	(3.7)	(4.3)	(2.8)
SON	ANN	-17.3	-16.9	-16.9	0.0	0.9	0.5	3.0	6.6	6.0	6.1	1.2	1.6	1.2	2.5	69.5	69.0	69.0	-6.1	-11.8	-7.0	-21.2
				(1.5)	(1.4)	(1.1)	(1.1)	(1.4)			(0.3)	(0.5)	(0.5)	(0.5)	(0.7)			(5.7)	(2.3)	(3.7)	(2.9)	(5.2)
	RCF	-17.3	-17.8	-18.0	0.4	1.7	1.1	4.2	6.6	6.7	7.1	0.8	1.0	0.9	1.3	69.5	68.7	68.7	-8.0	-14.7	-9.2	-26.8
				(0.6)	(1.3)	(1.1)	(1.0)	(1.7)			(0.2)	(0.5)	(0.4)	(0.4)	(0.4)			(1.0)	(3.1)	(4.5)	(3.5)	(5.9)
	NNC	-17.3	-17.3	-17.0	-0.2	0.9	0.3	2.6	6.6	6.3	6.9	1.1	1.6	1.3	2.6	69.5	68.0	69.0	-6.0	-11.8	-7.2	-20.9
				(0.7)	(1.0)	(1.0)	(0.9)	(1.4)			(0.2)	(0.5)	(0.5)	(0.6)	(0.6)			(3.0)	(2.7)	(3.9)	(2.9)	(4.9)
Annual	ANN	-	-	-	-	-	-	-	-	-	-	-	-	-	-	73.4	72.8	74.5	-5.8	-10.1	-6.8	-17.7
																		(3.5)	(1.5)	(2.0)	(1.7)	(2.5)
	RCF	-	-	-	-	-	-	-	-	-	-	-	-	-	-	73.4	73.2	74.1	-6.3	-10.9	-7.5	-18.6
																		(0.8)	(2.0)	(2.6)	(2.4)	(3.3)
	NNC	-	-	-	-	-	-	-	-	-	-	-	-	-	-	73.4	72.5	74.7	-5.6	-9.7	-6.7	-16.9
																		(2.1)	(1.7)	(2.2)	(2.1)	(2.8)

Table C.2: Statistics of temperature indices at Sonnblick, see table B.1 for details on structure and abbreviations. Table includes the annual minimum daily mean temperature (T1min), annual maximum daily mean temperature (T1max) and cold day frequency (CDF). Absolute change values: T1min, T1max, CDF.

		Annual minimum daily mean temperature [°C]							**Annual maximum daily mean temperature [°C]**							**Cold day (< 0 °C) frequency [%]**						
					RCP 4.5		RCP 8.5					RCP 4.5		RCP 8.5					RCP 4.5		RCP 8.5	
		Obs	Rean	Hist	NF	DF	NF	DF	Obs	Rean	Hist	NF	DF	NF	DF	Obs	Rean	Hist	NF	DF	NF	DF
DJF	ANN	-24.0	-23.8	-23.7	0.7	2.2	1.0	4.7	-2.2	-2.7	-2.3	2.1	2.8	2.2	4.8	99.8	99.9	99.9	-0.9	-1.7	-1.0	-5.4
				(1.3)	(0.9)	(1.1)	(1.3)	(1.5)			(0.6)	(0.6)	(0.7)	(0.5)	(0.6)			(0.1)	(0.3)	(0.6)	(0.5)	(1.5)
	RCF	-24.0	-23.8	-23.7	0.8	1.9	1.1	3.9	-2.2	-2.2	-2.3	1.0	1.3	1.0	1.9	99.8	99.9	99.9	-0.1	-0.2	-0.2	-0.5
				(0.2)	(0.6)	(0.9)	(0.9)	(1.1)			(0.3)	(0.4)	(0.3)	(0.4)	(0.3)			(0.0)	(0.1)	(0.1)	(0.1)	(0.2)
	NNC	-24.0	-23.8	-24.5	0.8	2.2	1.1	4.7	-2.2	-2.7	-2.4	1.3	1.8	1.4	3.1	99.8	99.9	99.9	-0.3	-0.5	-0.4	-1.8
				(0.5)	(1.1)	(1.0)	(1.2)	(1.4)			(0.6)	(0.5)	(0.5)	(0.5)	(0.5)			(0.1)	(0.1)	(0.3)	(0.2)	(0.7)
MAM	ANN	-18.8	-19.0	-19.3	0.2	1.3	1.0	3.3	1.8	1.2	1.3	1.0	1.7	1.1	2.7	94.7	95.6	95.6	-2.7	-6.2	-3.8	-12.4
				(0.4)	(0.7)	(1.0)	(0.5)	(0.8)			(0.4)	(0.5)	(0.5)	(0.2)	(0.6)			(1.1)	(1.1)	(2.9)	(2.0)	(5.1)
	RCF	-18.8	-19.2	-19.8	0.3	1.2	1.0	3.4	1.8	1.4	1.5	0.6	1.0	0.7	1.4	94.7	95.5	95.4	-2.2	-5.2	-3.5	-9.8
				(0.4)	(0.5)	(0.9)	(0.6)	(0.6)			(0.1)	(0.4)	(0.4)	(0.2)	(0.3)			(0.3)	(0.8)	(2.7)	(2.0)	(3.0)
	NNC	-18.8	-19.7	-20.5	0.4	1.4	1.1	3.9	1.8	2.1	2.3	0.9	1.6	1.2	2.7	94.7	93.9	94.2	-2.4	-5.6	-3.7	-11.1
				(0.7)	(0.9)	(1.2)	(0.8)	(0.6)			(0.2)	(0.5)	(0.4)	(0.3)	(0.4)			(0.7)	(1.0)	(2.8)	(2.0)	(3.7)
JJA	ANN	-6.4	-6.3	-5.7	0.5	1.3	1.0	3.2	8.3	7.7	7.7	1.8	2.8	2.1	4.7	35.2	32.1	36.0	-13.1	-19.1	-14.5	-27.7
				(0.7)	(0.6)	(0.8)	(0.7)	(1.5)			(0.1)	(0.6)	(0.7)	(0.7)	(1.0)			(8.6)	(2.7)	(2.7)	(2.9)	(3.7)
	RCF	-6.4	-5.9	-5.5	0.8	1.6	1.2	3.6	8.3	7.4	7.6	0.8	1.2	0.9	1.4	35.2	31.7	33.3	-14.5	-20.8	-16.1	-28.2
				(0.2)	(0.6)	(0.8)	(0.6)	(1.4)			(0.3)	(0.5)	(0.5)	(0.5)	(0.4)			(1.5)	(3.8)	(3.5)	(4.6)	(2.2)
	NNC	-6.4	-6.3	-6.5	0.8	1.8	1.3	4.0	8.3	8.0	8.2	1.4	2.4	1.7	3.9	35.2	33.5	37.4	-16.5	-23.4	-18.0	-32.0
				(0.2)	(0.6)	(0.9)	(0.6)	(1.5)			(0.3)	(0.5)	(0.6)	(0.6)	(0.8)			(3.3)	(3.9)	(3.2)	(4.6)	(1.1)
SON	ANN	-17.5	-17.0	-16.9	0.4	1.6	1.2	4.1	5.2	4.7	4.3	1.7	2.2	1.7	3.7	76.5	77.3	78.1	-7.9	-14.9	-9.7	-26.2
				(0.8)	(1.0)	(1.1)	(1.0)	(1.7)			(0.3)	(0.8)	(0.8)	(0.8)	(1.2)			(3.2)	(3.9)	(5.2)	(4.3)	(6.7)
	RCF	-17.5	-17.8	-17.7	0.2	1.7	1.4	4.0	5.2	5.3	5.3	1.2	1.5	1.2	1.9	76.5	76.1	77.0	-7.3	-13.5	-8.9	-23.2
				(0.4)	(1.0)	(1.2)	(1.1)	(1.7)			(0.3)	(0.7)	(0.6)	(0.6)	(0.5)			(0.4)	(3.4)	(5.0)	(3.8)	(6.5)
	NNC	-17.5	-17.7	-17.6	0.2	1.5	0.9	3.8	5.2	4.9	4.7	1.4	1.9	1.5	3.0	76.5	76.8	78.9	-6.6	-13.1	-8.4	-23.3
				(0.7)	(0.9)	(1.2)	(0.8)	(1.6)			(0.2)	(0.7)	(0.7)	(0.8)	(0.9)			(2.8)	(3.6)	(4.8)	(3.8)	(5.8)
Annual	ANN	-	-	-	-	-	-	-	-	-	-	-	-	-	-	76.5	76.1	77.3	-6.2	-10.5	-7.3	-18.0
																		(2.7)	(1.7)	(2.1)	(2.0)	(2.6)
	RCF	-	-	-	-	-	-	-	-	-	-	-	-	-	-	76.5	75.7	76.3	-6.1	-10.0	-7.2	-15.5
																		(0.4)	(1.9)	(2.5)	(2.3)	(2.8)
	NNC	-	-	-	-	-	-	-	-	-	-	-	-	-	-	76.5	75.9	77.5	-6.5	-10.7	-7.7	-17.1
																		(1.4)	(2.0)	(2.4)	(2.3)	(2.5)

D
Appendix

Table D.1: Statistics of relative humidity indices at Zugspitze, see table B.1 for details on structure and abbreviations. Table includes the average annual minimum (RH1min) and frequency of values > 95 % (RHf95) . Absolute change values: RH1min, RHf95.

		Average annual minimum [%]							**Frequency [%] of values > 95 %**						
					RCP 4.5		RCP 8.5					RCP 4.5		RCP 8.5	
		Obs	Rean	Hist	NF	DF	NF	DF	Obs	Rean	Hist	NF	DF	NF	DF
DJF	ANN	19.4	22.7	30.0	-6.8	-7.1	-7.7	-8.2	27.5	14.0	20.8	1.5	1.5	0.7	2.0
				(4.7)	(4.7)	(5.0)	(5.5)	(5.7)			(7.0)	(2.1)	(3.8)	(2.6)	(3.9)
	RCF	19.4	27.3	28.5	-0.2	0.2	-0.5	0.9	27.5	10.1	10.2	0.2	0.3	0.0	0.3
				(0.6)	(1.8)	(1.4)	(2.2)	(1.7)			(0.5)	(1.2)	(1.5)	(1.0)	(1.9)
	NNCs	19.4	25.5	26.9	-5.1	-5.0	-5.1	-5.3	27.5	13.5	15.7	1.7	1.8	1.0	2.9
				(2.2)	(3.8)	(4.1)	(3.8)	(4.7)			(3.0)	(1.7)	(2.6)	(1.5)	(3.1)
MAM	ANN	30.6	42.7	44.8	-5.6	-4.0	-6.5	-6.0	39.8	21.4	25.4	2.3	1.2	0.5	-0.2
				(3.7)	(3.1)	(2.4)	(3.4)	(3.1)			(7.0)	(1.4)	(1.3)	(1.7)	(1.9)
	RCF	30.6	48.0	50.1	-0.2	0.9	0.8	1.4	39.8	16.5	18.3	2.6	2.0	1.5	3.0
				(1.0)	(1.4)	(1.7)	(2.7)	(1.7)			(0.9)	(1.1)	(1.2)	(1.1)	(1.0)
	NNCs	30.6	42.5	43.0	-4.9	-3.0	-4.8	-4.3	39.8	21.9	26.9	2.5	1.4	0.6	0.3
				(1.9)	(3.0)	(2.1)	(5.0)	(2.5)			(5.1)	(1.6)	(1.4)	(1.5)	(1.9)
JJA	ANN	40.6	51.1	50.1	-4.1	-4.5	-5.2	-6.2	47.2	28.7	34.1	-2.2	-3.9	-1.5	-8.5
				(3.1)	(3.0)	(1.3)	(3.7)	(3.0)			(7.7)	(3.3)	(2.2)	(1.6)	(2.9)
	RCF	40.6	60.1	58.8	-0.8	-1.5	-1.1	-2.2	47.2	23.5	24.1	-1.3	-2.9	-0.8	-6.2
				(0.7)	(1.2)	(1.0)	(1.3)	(1.7)			(2.1)	(2.8)	(1.5)	(1.6)	(2.3)
	NNCs	40.6	49.8	45.5	-3.9	-5.1	-5.2	-5.7	47.2	31.2	32.4	-2.4	-4.6	-2.3	-9.4
				(1.3)	(2.8)	(1.4)	(2.4)	(2.4)			(3.4)	(3.6)	(2.5)	(2.0)	(3.0)
SON	ANN	19.2	24.7	28.2	-4.5	-5.0	-5.9	-4.2	29.9	13.4	14.0	0.3	-0.4	0.5	-2.0
				(5.4)	(2.6)	(2.8)	(3.3)	(3.6)			(2.3)	(0.9)	(0.8)	(0.7)	(0.9)
	RCF	19.2	28.7	29.3	-0.4	-1.9	-1.3	-1.6	29.9	10.8	10.2	0.2	-0.5	0.3	-1.8
				(1.0)	(1.4)	(1.4)	(1.5)	(2.5)			(1.1)	(1.0)	(0.9)	(0.9)	(0.6)
	NNCs	19.2	26.4	27.9	-4.0	-5.9	-5.2	-4.5	29.9	12.5	11.6	0.3	-0.2	0.4	-1.2
				(3.5)	(2.5)	(2.7)	(2.4)	(2.7)			(1.2)	(0.6)	(0.5)	(0.6)	(0.9)
Annual	ANN	15.8	16.7	20.8	-6.5	-6.7	-8.4	-7.2	36.1	19.4	23.6	0.5	-0.4	0.1	-2.2
				(2.9)	(1.8)	(1.9)	(2.1)	(3.7)			(4.4)	(0.9)	(0.8)	(0.6)	(0.9)
	RCF	15.8	23.7	24.8	-0.3	-0.4	-0.8	-0.1	36.1	15.3	15.8	0.4	-0.3	0.3	-1.2
				(0.5)	(0.5)	(0.9)	(1.1)	(1.4)			(1.0)	(1.1)	(0.6)	(0.5)	(0.7)
	NNCs	15.8	20.5	22.1	-7.0	-7.3	-7.3	-7.2	36.1	19.8	21.7	0.5	-0.4	-0.1	-1.9
				(1.9)	(2.3)	(2.4)	(2.2)	(3.5)			(2.6)	(1.0)	(0.6)	(0.3)	(0.9)

E
Appendix

Table E.1: Statistics of wind speed indices at Zugspitze, see table B.1 for details on structure and abbreviations. Table includes the frequency of strong wind days (WSf12) and annual maximum daily mean wind speed (WS1max). Absolute change values: WSf12. Relative change values: WS1max.

		Frequency of strong wind days [$dailymean > 12m/s$]							**Annual maximum daily mean wind speed** [m/s]						
					RCP 4.5		RCP 8.5					RCP 4.5		RCP 8.5	
		Obs	Rean	Hist	NF	DF	NF	DF	Obs	Rean	Hist	NF	DF	NF	DF
DJF	ANN	22.6	18.3	25.4 (9.7)	1.5 (1.9)	1.9 (2.1)	1.2 (2.9)	3.7 (2.2)	22.3	18.3	18.6 (1.0)	3.4 (3.1)	3.6 (4.0)	3.0 (4.5)	4.6 (4.5)
	RCF	22.6	17.7	18.2 (0.9)	1.6 (2.3)	1.9 (2.5)	1.5 (3.0)	3.4 (2.2)	22.3	18.8	18.9 (0.3)	1.8 (1.9)	2.4 (2.8)	1.2 (2.7)	2.6 (3.0)
	Comb	22.6	20.9	23.5 (3.7)	0.9 (1.9)	1.5 (2.1)	0.7 (2.5)	2.6 (1.8)	22.3	20.8	20.7 (0.6)	3.3 (3.4)	3.5 (3.1)	3.7 (3.5)	4.2 (4.7)
MAM	ANN	10.4	4.4	5.5 (1.6)	1.2 (1.3)	1.2 (1.0)	0.8 (1.5)	1.8 (1.0)	17.6	14.1	14.5 (0.5)	6.3 (3.3)	4.3 (4.0)	4.6 (4.8)	5.8 (3.2)
	RCF	10.4	4.7	5.5 (0.2)	0.7 (1.3)	0.9 (0.7)	0.4 (1.6)	1.5 (0.9)	17.6	14.6	14.9 (0.4)	2.3 (3.4)	2.7 (4.2)	1.5 (5.7)	4.6 (5.3)
	Comb	10.4	8.4	8.9 (0.9)	0.7 (1.3)	1.1 (0.5)	0.3 (1.1)	1.6 (0.7)	17.6	16.5	16.9 (0.3)	1.3 (2.4)	2.6 (3.0)	1.0 (3.0)	2.3 (1.7)
JJA	ANN	2.5	0.0	0.1 (0.1)	0.0 (0.1)	0.0 (0.1)	0.0 (0.1)	0.1 (0.1)	13.1	10.3	10.3 (0.3)	2.3 (3.3)	1.6 (1.7)	1.2 (2.0)	2.5 (3.8)
	RCF	2.5	0.0	0.1 (0.0)	0.1 (0.1)	0.1 (0.1)	0.1 (0.1)	0.1 (0.1)	13.1	10.5	10.5 (0.1)	2.3 (1.6)	2.1 (2.5)	2.1 (1.7)	2.2 (2.1)
	Comb	2.5	0.4	0.7 (0.1)	0.1 (0.2)	0.0 (0.2)	0.1 (0.1)	0.1 (0.3)	13.1	11.6	12.0 (0.2)	0.5 (3.7)	0.6 (1.9)	0.3 (1.7)	0.6 (2.3)
SON	ANN	11.8	9.1	17.1 (4.0)	-0.4 (2.1)	0.1 (2.1)	0.1 (0.9)	0.0 (1.5)	20.0	17.5	18.4 (0.5)	1.3 (3.7)	2.6 (4.3)	2.5 (5.6)	1.0 (5.4)
	RCF	11.8	8.5	9.0 (0.3)	0.3 (1.6)	0.4 (1.8)	0.8 (1.3)	0.3 (1.6)	20.0	18.7	17.9 (0.4)	1.8 (4.5)	2.7 (5.7)	4.3 (5.0)	1.6 (5.5)
	Comb	11.8	12.8	16.2 (2.0)	-0.4 (1.8)	-0.2 (2.1)	0.1 (0.9)	-0.2 (1.5)	20.0	18.9	20.4 (0.3)	-0.4 (3.5)	2.4 (5.2)	1.3 (5.3)	1.2 (4.9)
Annual	ANN	11.8	7.9	12.0 (3.6)	0.6 (0.8)	0.8 (0.6)	0.5 (0.8)	1.4 (0.9)	23.6	19.1	19.9 (0.6)	3.0 (2.7)	3.5 (3.1)	2.9 (4.0)	2.9 (3.7)
	RCF	11.8	7.7	8.1 (0.2)	0.7 (0.8)	0.8 (0.7)	0.7 (0.9)	1.3 (0.9)	23.6	20.5	20.0 (0.4)	1.8 (2.9)	2.6 (3.8)	2.0 (3.4)	2.3 (3.9)
	Comb	11.8	10.6	12.3 (1.7)	0.3 (0.6)	0.6 (0.6)	0.3 (0.7)	1.0 (0.7)	23.6	21.4	22.1 (0.3)	2.6 (2.9)	4.1 (2.7)	2.8 (3.8)	3.8 (3.3)