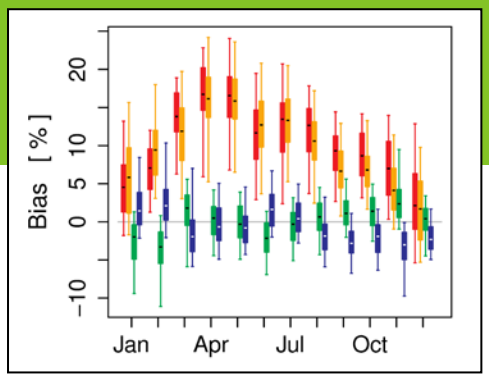
Wegener Center for Climate and Global Change
University of Graz



Scientific Report No. 56-2014

Evaluation of Mulit-Variable Quantile Mapping on Regional Climate Models

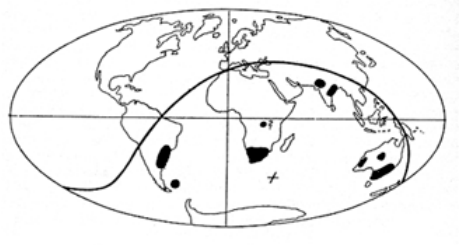
Renate Anna Irma Wilcke

January 2014

Financially supported by

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

The present report is the result of a PhD thesis work completed in December 2013. The work was funded by the Austrian Climate Research Programme (ACRP) under research grant A963645 (CC-Snow project) and K10AC0K00049 (CC-Snow II) and by the European Union Seventh Framework Programme (FP7/2007-2013) with the IMPACT2C project.



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

Wegener Center Verlag • Graz, Austria

ISBN 978-3-9503608-3-7

January 2014

Contact: Diplom Meteorologin Renate Anna Irma Wilcke
renate.wilcke@smhi.se

Wegener Center for Climate and Global Change
University of Graz
Brandhofgasse 5
A-8010 Graz, Austria
www.wegcenter.at

Dissertation
zur Erlangung des akademischen Grades einer Doktorin der Naturwissenschaften an der Naturwissenschaftlichen Fakultät der Karl-Franzens-Universität

EVALUATION OF MULTI-VARIABLE QUANTILE MAPPING BASED ON REGIONAL CLIMATE MODELS

RENATE ANNA IRMA WILCKE

Dezember 2013

Betreuer:
Univ.-Prof. Mag. Dr. Gottfried Kirchengast
Ass.-Prof. Mag. Dr. Andreas Gobiet

Wegener Center of Climate and Global Change
Karl-Franzens-Universität Graz

Renate Anna Irma Wilcke: *Evaluation of multi-variable quantile mapping based on regional climate models,* Diplom Meteorologin, © December 2013

SUPERVISORS:
Univ.-Prof. Mag. Dr. Gottfried Kirchengast
Ass.-Prof. Mag. Dr. Andreas Gobiet

LOCATION:
Graz, Austria

TIME FRAME:
April 2010 – December 2013

. . . The most naïve description of the scientific method is that you start with a theory and test it by experiments. . . . You might think that what she should be trying to do is to design an experiment that will prove her theory is correct. However, that's not good science. Good science consists of designing an experiment that will demonstrate that a theory is ***wrong*** *– if it is. So a large part of the scientist's job is not 'establishing truths', it is trying to shoot down the scientist's own ideas. And those of others scientists. This is what we meant when we said that science tries to protect us against believing what we want to be true, or what authority tells us is true. It doesn't always succeed, but that at least is the aim.*

— Terry Pratchett, Ian Stewart, Jack Cohen "The Science of Discworld II: The Globe"

ABSTRACT

Statistical downscaling and error correction methods are used to improve output from Regional Climate Models (RCMs) for the use in climate impact research. One such statistical empirical method is Quantile Mapping (QM), which is often used to correct for biases in RCM's temperature and precipitation, recently. In this thesis the performance of QM is investigated when applied to temperature and precipitation, but also to relative humidity, wind speed, global radiation, and surface pressure.

The main results are that QM strongly reduces the annual and monthly mean biases and corrects the distributions—independent of shape—for all variables and RCMs. Even non-stationarities in the biases cannot hinder the good results of QM. The temporal and inter-variable structure provided within the RCMs is not manipulated by QM, which is positive when the RCM's structure is correct. As strong deviations from observed inter-variable relations are found in RCMs, the conservation of those relations produces less convincing results of QM for combined indices.

The effect of individually quantile-mapped climate variables on derived variables, like number of snow days, snow depth, or temperature-humidity index (THI) is mostly positive. Derived variables are sensitive to the day-to-day structure of time-series. The improvement is not as strong as for directly corrected variables, as QM corrects for the mean statistics and preserves the day-to-day structure of the RCM. However, in terms of Nash-Sutcliffe-Efficiency (NSE) no improvement can be found in most cases.

Generally, QM corrects individual climate variables with convincing results, but needs further enhancements to improve the inter-variable relations in RCM output.

ZUSAMMENFASSUNG

Um Klimadaten aus Regionalen Klimamodellen (RCMs) für Klimafolgenforschung aufzubereiten werden statistische Skalierungs- und Fehlerkorrektur-Methoden verwendet. Diesbezüglich besonders geeignet ist Quantile Mapping (QM), welches in letzter Zeit häufig zur Korrektur simulierter Temperatur und Niederschlag verwendet wird. In dieser Arbeit wird zusätzlich untersucht wie gut QM Fehler in Relative Feuchte, Windgeschwindigkeit, Globalstrahlung und Bodendruck korrigieren kann.

Zu den Hauptergebnissen zählt die starke Reduktion jährlicher und monatlicher Fehler durch QM, wie auch die formunabhängige Korrektur der Dichteverteilungen für alle Variablen und RCMs. Sogar nicht stationäre Fehler werden mit QM reduziert. Es wird gezeigt, dass QM die vom RCM vorgegebene zeitliche Struktur und die Beziehungen zwischen Variablen nicht verändert. Allerdings werden zum Teil große Unterschiede zwischen den beobachteten und den simulierten Beziehungen zwischen Variablen gefunden. Durch die Erhaltung der fehlerhaften simulierten Beziehungen können größere Fehler in Indizes aus kombinierten Variablen erhalten bleiben.

Der Effekt von fehlerkorrigierten Klimavariablen auf davon abgeleitete Variablen ist überwiegend positiv, wie z.B. bei der Anzahl an Schneetagen, Schneehöhe oder Temperatur-Feuchte-Index (THI). Abgeleitete und kombinierte Variablen sind sensibel gegenüber der täglichen Struktur und den Beziehungen zwischen den Variablen. Da QM genau diese Strukturen in den Modeldaten nicht ändert, ist die Verbesserung nicht so großwie bei den direkt korrigierten Variablen. Gemessen an der Nash-Sutcliffe-Effizienz (NSE) kann keine Verbesserung gefunden werden.

Zusammengefasst korrigiert QM Klimavariablen sehr gut. Um die Beziehungen zwischen Variablen und zeitliche Strukturen in Modeldaten zu verbessern, muss QM allerdings weiterentwickelt werden.

PUBLICATIONS

Parts of this Ph.D. thesis have already been published. The references are:

Wilcke, R. A. I., Mendlik, T., and Gobiet, A. (2013). *Multi-variable downscaling and error-correction of regional climate models.* Clim. Change, 120 (4): 871–887, DOI: 10.1007/s10584-013-0845-x

Wilcke, R. A. I., Leuprecht, A., and Gobiet, A. (2012). *Effects of Climate Change on Future Snow Conditions in Tyrol and Styria (CC-Snow) - Final report Climate*, Scientific Report, Wegener Center Verlag Graz, August 2012, ISBN 978-3-9503112-5-9

. . . Without intelligence, we could never have got started on that path, but intelligence alone was not enough. We had to find a way to share our intelligence with others, and to store useful ideas and tricks for the benefit of the whole group, or at least, those in a position to make use of it. . . .

— Terry Pratchett, Ian Stewart, Jack Cohen "The Science of Discworld II: The Globe"

ACKNOWLEDGEMENT

There are quite some people, projects, and institutions that I would like to mention here and say thank you.

I will start with the people as they influenced my thesis and nearly four years of my life and accompanied me on the stony but exciting way towards being a PhD.

Thank you, Andreas Gobiet and Gottfried Kirchengast, for supervision of this thesis and the opportunity to work at the Wegener Center with ReLoClim in the first place. In particular, I want to thank Andreas Gobiet for the discussions and encouragement, in particular throughout the paper writing process. I think, I'm a quite self-critical person, but I really like our paper and am proud of our work. Gottfried, thank you for scientific discussions, in particular during my and other presentations, I really appreciated that.

For all the shared moments (hours) of fun, silliness, motivation, support, as well as frustration and despair during the last three years and nine months, for always being there for me to talk and discuss, scientifically as well as personal, for awesome lunch times at Stadtpark, archery-boccia (we should really publish our invention and fund a club), observing the dark side of Graz, and cycling adventures, thank you Bärbel Scherllin-Pirscher and Andi Prein. You made my time at the Wegener Center a great one, and gave me the feeling of having friends accompanying my path.

Thomas Mendlik, Martin Jury, and Georg Heinrich I'd like to thank you for interesting and inspiring discussions (with and without tea, Rioja, or beer) in various ways. All arguments and discussions lead to new ways of thinking, to opening up our minds, which is so very important in science and in life.

In general, I'd like to thank my room mates of the terrace office in old WEGC and the best office (010) in new WEGC, in particular the constants Andi Prein and Andrea Damm. I enjoyed working with you in a perfect mixture of quiet and chatty atmosphere.

I want to thank my brother, Robert! We rock the world! Many thanks to my parents, for enduring my bad moods when I'm stressed (which happened, of course, only once or twice during my studies),

and for making me believe that I can do *everything*, that I'm capable of doing *everything*. And . . . I am.

Thank you, Katrin Scheufele.

I want to thank COST Action VALUE and its participants, which was the best thing that could have additionally happened to me during my PhD. It gave me quite some opportunities for my future life as well as the chance to unfold myself in a (social) scientific way. I've met so many interesting and amazing scientists within this project and the discussions with them strengthened my decision to be a scientist. Thank you, Martin Widmann, you kind of were like a mentor to me lately. And I want to acknowledge VALUE for funding a short-term scientific mission to UCL to work with Richard Chandler for two months.

Many thanks to Richard Chandler for instructive two months, and for spending so much time discussing with me. It's been a great time which I like to repeat some day.

Part of this work was based on the collaboration established in the CC-Snow project, here I want to thank Florian Hanzer in particular, for running AMUNDSEN again and again and again. Of course, my thank is with Ulrich Strasser and Thomas Marke as well for joint discussions, and the CC-Snow Task Force, indeed.

This work has been funded by the ACRP projects CC-Snow (project number A963645) and CC-Snow II (project number K10AC0K00049) and by the European Union Seventh Framework Programme (FP7/2007–2013) project IMPACT2C.

The ENSEMBLES data used in this work was funded by the EU FP6 Integrated Project ENSEMBLES (Contract number 505539) whose support is gratefully acknowledged. I would like to acknowledge the Austrian Central Institute for Meteorology and Geodynamics (ZAMG) and the Swiss Federal Office of Meteorology and Climatology (MeteoSwiss) for providing observational data.

CONTENTS

ACRONYMS

ACF	autocorrelation function
AGCM	atmosphere general circulation model
all_v_cor	a simulated snow variable where only corrected climate variables have been used
all_uncor	where x is a snow variable, e. g., SWE simulated by using only uncorrected climate variables.
AMUNDSEN	**A**lpine **MU**ltiscale **N**umerical **D**istributed Simulation **EN**gine (Strasser, 2008)
AMUobs	AMUNDSEN results driven with observations
AOGCM	coupled atmosphere ocean general circulation model
BC	bias correction
BCM	Bergen Climate Model (Furevik et al., 2003)
BYF	bias of yes forecast
C4I	Community Climate Consortium for Ireland
CC-Snow	ACRP projects CC-Snow (project number A963645) and CC-Snow II (project number K10AC0K00049), both lead by Prof. Ulrich Strasser
CCS	climate change signal
CDF	cumulative distribution function
CMIP	Coupled Model Intercomparison Project
CORDEX	CORDEX is a project of the WCRP Group on Regional Climate (WGRC). CORDEX is providing global coordination of Regional Climate Downscaling for improved regional climate change adaptation and impact assessment. `http://wcrp-cordex.ipsl.jussieu.fr/`
DOY	day of year
ECDF	empirical cumulative distribution function
ECHAM5	ECHAM5 (Roeckner et al., 2003)
ECMWF	European Centre for Medium-Range Weather Forecasts

ENSEMBLES	The ENSEMBLES project (contract number GOCE-CT-2003-505539) is supported by the European Commission's 6th Framework Programme as a 5 year Integrated Project from 2004-2009 under the Thematic Sub-Priority "Global Change and Ecosystems" (`http://ensembles-eu.metoffice.com`)
ERA40	ECMWF 40-year re-analysis
FAR	false alarm ratio
GCM	General Circulation Model
HadCM3Q0	Hadley Centre Coupled Model, version 3, sensitiviy Q0 (Collins et al., 2001)
HadCM3Q16	Hadley Centre Coupled Model, version 3, sensitivity Q16 (Collins et al., 2001)
HadRM3Q0	Hadley Center Regional Model, version 3, sensitiviy Q0
HIRHAM	HIRHAM , acronym HIRHAM comes from the combination of HIRLAM and ECHAM4 (Christensen et al., 2007)
ICTP	The Abdus Salam International Centre for Theoretical Physics
IPCC	Intergovernmental Panel on Climate Change
KNMI	Koninklijk Nederlands Meteorologisch Instituut (Royal Netherlands Meteorological Institute)
MeteoSwiss	Swiss Federal Office of Meteorology and Climatology
METNO	Norwegian Meteorological Institute
METO-HC	Metoffice Hadley Center
MOS	model output statistics
NOAA	National Oceanic and Atmospheric Administration
NOS	number of snow days number of days with solid precipitation
NSE	Nash-Sutcliffe-Efficiency
PC	proportion correct
PDF	probabiliy density function
PP	perfect prognosis

PRUDENCE	Project EVK2-CT2001-00132 in the EU 5th Framework program for Energy, environment, and sustainable development. `http://prudence.dmi.dk/`
pr_uncor	where x is a snow variable, e. g., SWE simulated by using corrected climate variables but precipitation uncorrected.
QM	Quantile Mapping
RACMO2	Regional Atmospheric Climate Model version 2.1 (van Meijgaard et al., 2008)
RCA	Rossby Center Regional Climate Model (Samuelsson et al., 2011)
RCM	Regional Climate Model
RCP	Representative Concentration Pathway
RegCM3	RegCM3 , 3-dimensional, sigma-coordinate, primitive equation regional climate model (Winter et al., 2009)
RMSE	root mean squared error
SD	snow depth
SDM	statistical downscaling model
SF	snow fall amount
SMHI	Sveriges Meteorologiska och Hydrologiska Institut (Sweden's Meteorological and Hydrological Institute)
SRES	Special Report on Emission Scenarios
SWE	snow water equivalent
THI	temperature-humidity index
v_uncor	where x is a snow variable, e. g., SWE and v is an element of the group of uncorrected climate variables: air temperature, precipitation, relative humidity, global radiation, wind speed, and all.
WMO	World Meteorological Organisation
wss_uncor	where x is a snow variable, e. g., SWE simulated by using corrected climate variables but wind speed uncorrected.
ZAMG	Austrian Central Institute for Meteorology and Geodynamics

. . . Unfortunately, adult beliefs about causality are usually contaminated by the less sophisticated wish-fulfilment philosophy that we carry with us from the tinkly magic of our infancies. For example, scientists will object to alternative theories on the grounds that "if that was true, we wouldn't be able to do the sums". Why do they think that nature cares whether humans can do the sums? Because their own desire to to the sums, which lets them write papers for learned journals, contaminates their otherwise rational view. There's a feeling of feet being stamped; the Almighty should change Her laws so that we can do the sums.

— Terry Pratchett, Ian Stewart, Jack Cohen "The Science of Discworld II: The Globe"

1 INTRODUCTION AND MOTIVATION

Climate models use the current knowledge about physical processes in atmosphere, ocean, and other sub-systems (Solomon et al., 2007) to simulate historical as well as future climate on global and regional scale. Processes which are not yet understood, but observed, and too small to be resolved, are described in parameterisations within those models (Solomon et al., 2007). Naturally, parameterisations are approximations of reality which, thus, lead to deviations from observations together with model errors, discretisations, numerics and scale differences (Dolman and Gregory, 1992; Sellers et al., 1996). The resulting errors in the model output often makes it not directly usable for climate impact research (Fowler et al., 2007) (Chapter 2).

As impact modellers and policy makers need climate projections on regional scales, Regional Climate Models (RCMs) are used. In large projects like ENSEMBLES (van der Linden and Mitchell, 2009, `http://ensembles-eu.metoffice.com`) or CORDEX (Giorgi et al., 2009, `http://wcrp-cordex.ipsl.jussieu.fr/`) the uncertainty in large ensembles of model projections is investigated.

Especially impact researchers demand for high quality climate information. Impact models run on different resolutions and require preferably climate data on their corresponding scale. In particular, modelling on local scales (e. g., river catchments, mountain slopes) is done on high resolutions down to a couple of meters (e. g., Xu, 1999; Strasser, 2008). Thus, to bridge the gap in resolution and correct for errors in RCMs, statistical downscaling and error correction is applied (e. g., Maraun et al., 2010).

Error correction methods are various to be found and reach from a simple delta approach (Lehner et al., 2006) over multiple linear regression methods (Kilsby et al., 1998) to Quantile Mapping (QM)

(Brier and Panofsky, 1968). QM is a recently often used bias correction method and is part of empirical model output statistics (Déqué, 2007; Boé et al., 2007; Themeßl et al., 2012). As the name suggests, the correction is gained by mapping modelled distributions on observed distributions (quantile-wise). This accounts for different error characteristics of different quantiles and leads to better correction for errors than most other methods (e. g., Räisänen and Räty, 2012; Themeßl et al., 2011). Therefore, QM is the method of choice to be evaluated within this work, with focus on investigating bias correction of RCM output and of derived variables or indices.

To frame this work on bias correction of RCM output, I start with two review chapters. Chapter 2 gives a short review on climate modelling and downscaling, while in Chapter 3 errors in RCMs and their correction are reviewed. A description of the method (QM) can be found in Chapter 4.

Climate impact research needs RCM output of multiple meteorological variables. This typically includes temperature, precipitation, relative humidity, global radiation, surface pressure, and wind speed (e. g., Finger et al., 2012). These variables are therefore the focus of this study and described in Chapter 5.

Following the work lead by my colleague Matthias Themeßl (Themeßl et al., 2011, 2012), QM has then been evaluated in three studies as part of this PhD thesis.

Concerning bias correction in general and QM in particular, there are some questions arising, which are tackled in the following chapters:

- Can QM correct for bias in the main statistics and distributions? The biases of means on different temporal scales and the correction of distribution is validated in Chapter 6.
- Does QM degrade temporal characteristics? This question is answered by investigating the root mean squared error (RMSE) and autocorrelation for each variable and four RCMs in Chapter 6.
- Can QM be applied to climate scenarios as well as historical climate runs? A split sample test was performed throughout the first study in Chapter 6, whereby the calibration period is independent of the correction period.
- Does QM degrade inter-variable dependencies by correcting multiple variables independently? The correct representation of inter-variable relations is important for climate impact modellers as they use multiple climate variables as input. Therefore the correlation between simulated variable pairs is investigated and compared to their corrected version (first part of Chapter 7).
- How well do RCMs represent inter-variable relations? This question follows from the differences found in the correlations of

observed and simulated climate variables. By comparing conditional distributions of observed and simulated variable pairs, an attempt is made to find the status of the representation of inter-variable relations in RCM output (second part of Chapter 7).

- To what extent does QM on individual climate variables affect impact model output? Therefore, raw and corrected RCM output is used in impact models (a snow model and a heat index), and the derived indices are compared, respectively, and with observations (Chapter 8).

In each study chapter, conclusions are drawn at the end, which are summarised in an overall form in Chapter 9. Appendices with background information on methods and complementary figures and tables complete this thesis.

. . . what is important is the "World of If" point of view involved in the application of phase spaces . . . the real world alone is too limited to offer a convincing explanation . . . if you take the imaginative leap of considering unreal worlds, too, you can compare real world with all of those unreal worlds, and maybe find a principle that picks out the real one from all the others.

— Terry Pratchett

2

SHORT REVIEW ON CLIMATE MODELLING AND DOWNSCALING

Before writing about climate modelling it is good to understand the difference between weather and climate. The understanding will set the validation concepts and expectations on the models in the right relation. Here are their definitions:

Climate as it is defined by the World Meteorological Organisation (WMO):

"Climate in a narrow sense is usually defined as the "average weather", or more rigorously, as the statistical description in terms of the mean and variability of relevant quantities over a period of time ranging from months to thousands or millions of years. The classical period is 30 years, as defined by the WMO. These quantities are most often surface variables such as temperature, precipitation, and wind. Climate in a wider sense is the state, including a statistical description, of the climate system."

Weather as it is defined by the National Oceanic and Atmospheric Administration (NOAA).:

"The state of the atmosphere with respect to wind, temperature, cloudiness, moisture, pressure, etc. Weather refers to these conditions at a given point in time (e. g., today's high temperature), whereas Climate refers to the "average" weather conditions for an area over a long period of time (e. g., the average high temperature for today's date)."

2.1 CLIMATE MODELLING

The climate is determined by processes in the climate system which can be explained mainly by the interactions of and within five subsystems (Kraus, 2004, p. 12) (compare upper left panel in Figure 2.1):

- The *atmosphere*, which is a gas sphere enclosing the Earth and which consists of different sub-spheres reaching from the troposphere to the ionosphere.
- The *hydrosphere* is the water on Earth, which includes the oceans, rivers, lakes as well as groundwater. The oceans play the largest

role in the climate system by absorbing and emitting heat (ocean currents, e. g., golf-stream), and by evaporating water vapour into the atmosphere.

- the *cryosphere*, is the ice on Earth, which includes sea ice, land ice shields, and glaciers, but also snow.
- The *biosphere*, includes flora and fauna, as well humans.
- The *lithosphere* describes the solid Earth, in particular the land surface.

Climate modelling is about simulating the physics of those spheres to give an idealised picture (or time-series) of the world. The idea to model climate started with one-dimensional energy balance and radiation models over three-dimensional atmosphere general circulation models (AGCMs) to complex three dimensional coupled atmosphere ocean general circulation models (AOGCMs). The latter two are called General Circulation Models (GCMs), as they are built to simulate global circulation patterns in the atmosphere and ocean. They are based on the primitive equation system (Bjerknes, 1904) consisting of the Newton's law of motion, the hydrodynamic state equation, the thermodynamic energy equation, and the mass conservation. The physical equations are solved on a discretised model grid (see Figure 2.1).

Other physical processes which are too small or too complex like rainfall (Dolman and Gregory, 1992), land-surface interaction (Sellers et al., 1996), or heat transfer have to be parameterised. Parameterisations prescribe the statistics of a physical field (Intergovernmental Panel on Climate Change (IPCC) Solomon et al., 2007, 8.2.1.3). Their aim is to describe sub-grid-scale processes with the help of variables, which can be resolved on the model grid. Some of these parameterisations, like cloud or radiation processes, introduce major sources of uncertainties to climate models. As they are coupled to the physical model around, there are various approaches to parameterise sub-grid-scale processes and therefore they are an important source for differences between the different models (e. g., Lock, 2001; Martin et al., 2006).

Figure 2.2 shows the development in GCMs since the 1970s and how the modelled interaction between different parts of the system increased. It shows how more and more parameterisations, but also sub-models for e. g., vegetation or chemistry, were added to improve the modelling performance.

GCMs are either initialised with data (e. g., observational data) or spin up from a resting state till they reach an equilibrium state (Edwards, 2010). Then they are run freely simulating decades to 1000 years.

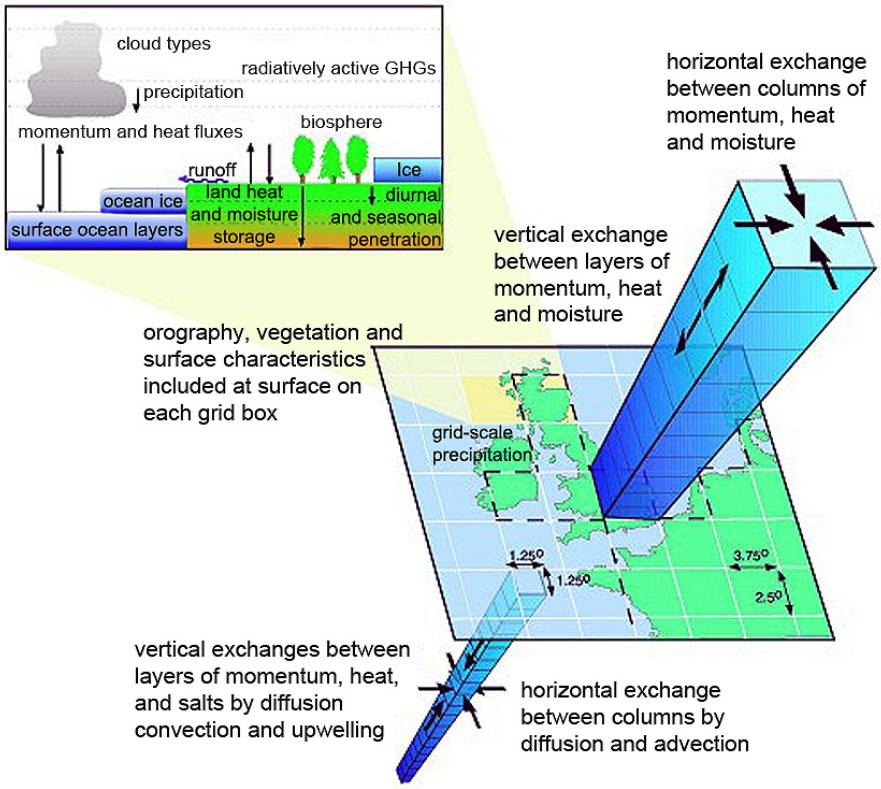


Figure 2.1: Schematic explanation of a climate model (Figure 2 in Viner et al., 1995) (with courtesy to `http://mycaveat.com/CAVEAT-background.htm` for a not blurry version).

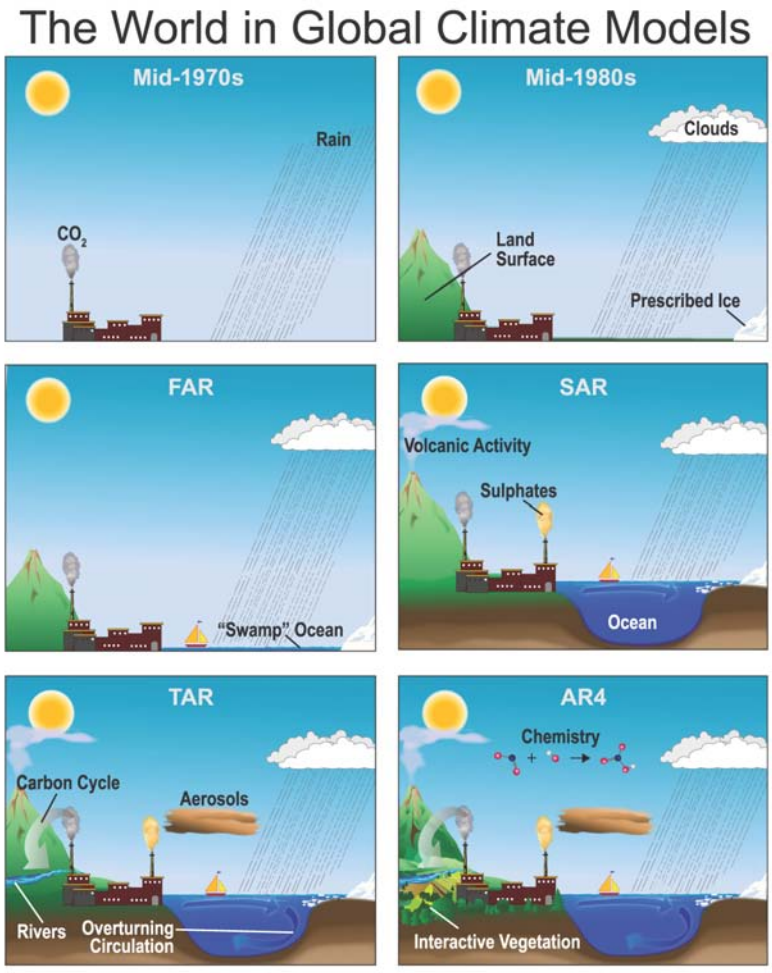


Figure 2.2: Sketch of the development of GCMs since mid-1970 showing the increase in complexity. Acronyms refer to the four assessment reports published by the IPCC. Image courtesy of IPCC AR4 WG 1 Figure 1.2.

The idea behind GCMs is to get statistical representations of future climate conditions, NOT to predict the weather on the e. g., 18th October 2064. However, they give a projection how the climate could be for the next, say, 100 years. As defined above, climate is the statistical description of weather over a longer time period. Thus it is not a "prediction" like a weather prediction for the far future. To be fully aware of this is important for validating climate models and using their output data for climate impact studies.

In climate modelling, the anthropogenic influence is part of external drivers of the climate system, thus, for future projections different greenhouse gas emission scenarios have been developed. Those emission scenarios (Special Report on Emission Scenarios (SRES), in the 3rd assessment report of the IPCC, Nakićenović et al. (2000)) are implemented as external drivers for the future projections of the GCMs. Four emission scenario families were defined: A1, A2, B1, B2. They cover a future world of rapid economic growth (A1), a divided world (national) (A2), an ecological friendly and integrated world (B1), and a world which is ecological friendly but divided (B2) (Nakićenović et al., 2000).

In 2011 new scenarios, regarding greenhouse gas emission and socio-economic possible futures, Representative Concentration Pathways (RCPs), were developed and used for the latest set of GCMs (CMIP5) and, thus, for the latest IPCC assessment report (2013/2014). The RCPs are defined and named by their level of radiative forcing by the end of the century (van Vuuren et al., 2011): $2.6\,W/m^2$, $4.5\,W/m^2$, $6\,W/m^2$ and $8.5\,W/m^2$. As extreme scenario, RCP8.5 shows a continuous increase in radiative forcing, whereas RCP2.6 levels out at about mid-century. The RCPs cover a period from 1850 to 2100, in which for 1850 to 2005 the pathways are fitted to observations.

GCMs run on a horizontal grid-spacing of 2.8° to 1.2°, which are about 300 km to 110 km, depending on the latitude. The effective resolution of the climate model is about times 4 to 8 higher (Grotch and MacCracken, 1991; Denis et al., 2002; Kapper, 2009; Prein et al., 2013b). The difference between grid-spacing and resolution of climate models, beside the literal meaning, results from the scale on which waves can be resolved. Small- and large-scale energy interactions are inaccurate in numerical models. Thus, to preserve the energy balance, waves of 2 times and 3 times the grid-spacing need to be removed. As one needs at least twice the grid-spacing to resolve two peaks of a wave (e. g., Grasso, 2000), this leads to an effective *resolution* of climate models of at least 4 times the grid spacing (Grasso, 2000; Pielke, 2002). More detailed information on this can be found in Kapper (2009).

Figure 2.3 shows meteorological phenomenon on their spatial and temporal scales. GCMs can simulate climate variations, seasonal cycles, planetary waves, high and large low pressure systems, and large phenomenons like teleconnection patterns (Handorf and Dethloff, 2012),

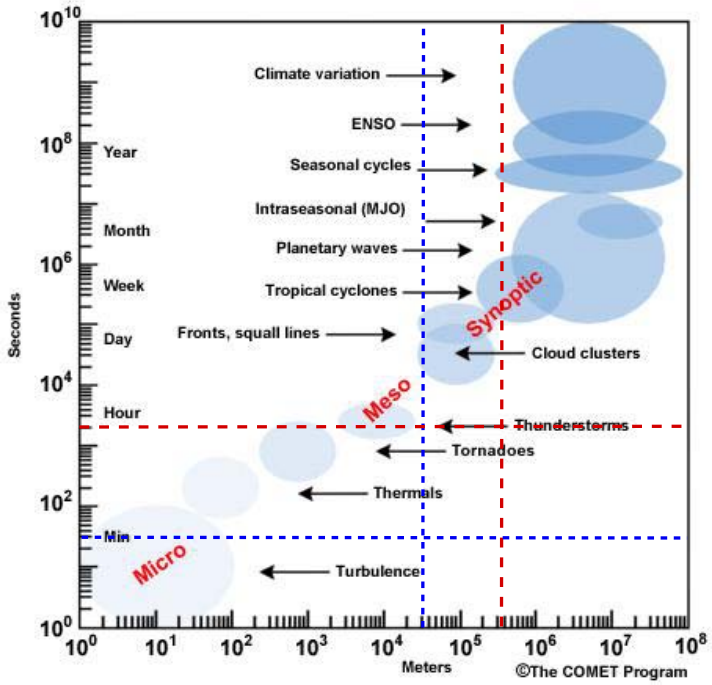


Figure 2.3: Meteorological spatial and temporal scales including GCM (red) and RCM (blue) resolution. Image courtesy to the COMET Program, modified by author.

but they do not cover smaller scales like cloud clusters or frontal systems or even turbulence which are, of course, also part of the climate system.

To gain information about higher resolved processes and e. g., to include more complex topography for e. g., investigating impacts of climate on river catchments, GCMs with variable resolutions have been developed and GCMs have been downscaled to local scales, which is described in the following section. Those called stretched grid GCMs run on high resolution for a specific region and coarser resolution elsewhere on the globe (Fox-Rabinovitz et al., 2001, 2006).

2.2 DOWNSCALING

There are two common strategies to derive higher resolved information from a GCM (Mearns et al., 1999): dynamical downscaling with RCMs and statistical downscaling with a broad variety of statistical downscaling models (SDMs). Both approaches have been compared several times e. g., over the European Alps (Hanssen-Bauer et al., 2005; Schmidli et al., 2006), or on monthly scale (Murphy, 1999). The outcome was that both improve to the raw GCM output on local scales.

2.2.1 *Dynamical downscaling*

Dynamical downscaling is synonym to regional climate modelling with RCMs. RCMs are very similar to GCMs but are able to simulate only limited areas instead of the entire globe. This makes them computationally more efficient and allows for higher resolutions in the selected region. The downscaling happens by using GCM or re-analysis[1] data as boundary conditions to drive the higher resolved RCM over the region of interest.

The idea of using RCMs for downscaling was suggested by Dickinson et al. (1989); Giorgi (1990). Those RCMs used similar parameterisations as the GCMs (McGregor, 1997).

Giorgi and Mearns (1999) are discussing an issue which still holds today: "Garbage In, Garbage Out". The longer an RCM runs and the larger the domain, the more degrees of freedom it gets, meaning it develops its own climatology which can be different from the GCM's (Giorgi and Bates, 1989). However, if the GCM produces wrong large-scale circulations, the RCM is not able to correct for those, and therefore will produce wrong small-scale fields.

Nowadays, RCMs are common tools to regionalise large-scale climate information of the GCMs and currently simulate the climate on grid spacing between 0.44° and 0.11° (e. g., Wang et al., 2004; Christensen and Christensen, 2007; Jacob et al., 2013) for regions like Europe. Experiments on convection permitting scale, below 4 km, with non-hydrostatic models are carried out for smaller regions, like Germany or the Alpine region (Hohenegger et al., 2008; Suklitsch et al., 2008, 2011; Langhans et al., 2012).

Modern RCMs still contain processes which are too small (spatial and temporal), too poorly understood, or too computationally costly to be explicitly represented. They are parameterised and include: radiation transfer in the atmosphere; planetary boundary layer, and surface layer (e. g., Hu et al., 2010); turbulence and diffusion (e. g., Han and Pan, 2011); cumulus convection (e. g., Kain and Fritsch, 1990; Kain, 2004); microphysics of clouds and precipitation (e. g., Chen and Sun, 2002; Jakob, 2003; Hong et al., 2004; Seifert and Beheng, 2006); interaction with Earth's surface (Knievel, 2008). For each of those processes various parameterisation schemes exist, e. g., for long-wave radiation there is the Rapid Radiative Transfer Model (Mlawer et al., 1997), or for short-wave radiation there is the Dodhia scheme (Dudhia, 1989). Another example is the microphysical scheme which describe the interactions among hydrometeors, vapour, and the environment (e. g., Knievel, 2008).

1 "Reanalysis (as well as analysis) is a process by which model information and observations of many different sorts are combined in an optimal way to produce a consistent, global best estimate of the various atmospheric, wave and oceanographic parameters." Dee et al. (2011)

The number of parameterisation schemes has increased and the schemes have improved over time, in particular regarding the land-surface interaction, atmospheric processes (Feser, 2006; Prein et al., 2013a,b) as well as radiative heating (Paeth and Mannig, 2013).

Since many institutes started to run RCM simulations, projects have been organised to co-ordinate the downscaling from GCMs with RCMs, e. g., in Europe PRUDENCE (Christensen and Christensen, 2007) with grid-spacing aim of 0.44°, the ENSEMBLES project (van der Linden and Mitchell, 2009, `http://ensembles-eu.metoffice.com`) with grid-spacing aims of 0.44° to 0.22° and recently CORDEX (Giorgi et al., 2009, `http://wcrp-cordex.ipsl.jussieu.fr/`) covering continental Earth with 14 domains with grid-spacing of 0.44° and 0.11° (Jacob et al., 2013; Vautard et al., 2013, `http://www.euro-cordex.net/`).

2.2.2 *Statistical Downscaling*

Statistical downscaling is a computational cheaper alternative to dynamical downscaling. Thus, statistical downscaling has been used to downscale from GCMs when only limited computer resources were available. As computers developed and more RCM runs were available, statistical downscaling is increasingly applied to dynamically downscaled products, which reduces the scale difference between the model grid and the target points. This combined approach uses the advantages of physical downscaling as far computational resources allow, then step in with computational cheaper statistical downscaling methods (e. g., Chen et al., 2012, for precipitation).

Statistical downscaling focuses on single meteorological variables as output, in particular precipitation (Maraun et al., 2010) and temperature. The main reason for that might be the dependence on observations for calibration, and precipitation and temperature are long observed variables. However, climate impact researchers, as hydrologists, ask for more variables as input for their models, thus more applications for variables like humidity, wind speed, and radiation can be found in recent literature (e. g., Boé et al., 2007; Hagemann et al., 2011).

Similar to the parameterisations in dynamical climate models, statistical models are calibrated with observations, though in the past, which leaves an unclear skill for applications on future climate. The skill of the SDMs varies for variables, seasons, and regions. I. e. Gutiérrez et al. (2013) found for Spain that regression methods perform best regarding correlation, while analogue and weather generators show more skill in reproducing the observed distributions for temperature in winter. Whereas the opposite picture is true for warm seasons. SDMs can be designed and calibrated for very local regions like special river catchments or mountain slopes. Therefore, their skill is high in that very region, but would only be applicable in that region.

There are various statistical downscaling techniques in use. An attempt to categorise those approaches has been made by e. g., Maraun et al. (2010) focusing on precipitation.

Statistical downscaling can be done two ways: one called perfect prognosis (PP) and the other model output statistics (MOS).

PERFECT PROGNOSIS: For the PP approach the statistical model is calibrated with observations for predictors as well as predictands. Predictors feed large-scale information to the model. During the application of the downscaling the predictors are gained from the dynamical model (GCM or RCM). Here, a realistic and reliable dynamical simulation is assumed, as physical processes are considered in a dynamical simulation.

Typical PP methods are e. g., deterministic linear models like *multiple linear regression* (e. g., Hertig and Jacobeit, 2008; Lutz et al., 2012) or *canonical correlation* analyses (e. g., Widmann, 2005; Busuioc et al., 2008), non-linear models like *artificial neural networks* and *machine learning* (e. g., Haylock et al., 2006), or analogue methods (Cubasch et al., 1996; Wetterhall et al., 2005; Matulla et al., 2008). Similar approaches exist for stochastic models like *vector Generalised Linear Models* (e. g., Yee and Stephenson, 2007) or *re-sampling techniques* (e. g., Beersma and Buishand, 2003). Weather generators are PP methods as well, which model random sequences of weather variables of unlimited length that are consistent with the key statistical properties of the observed meteorological records (e. g., Katz, 1996; Wilks and Wilby, 1999; Hirschi et al., 2012).

MODEL OUTPUT STATISTICS: MOS is based on statistical models that are calibrated using simulated predictors and observed predictands. The observed predictands are often on smaller spatial scale (e. g., observation stations) than the simulated predictors (i. e. RCM grid). In many applications an error correction is combined with the downscaling, which originates in weather forecasting (Wilks, 2006). The application can be done by calibrating on e. g., re-analysis driven RCMs or by downscaling whole distributions.

As with PP, MOS may follow a deterministic, stochastic or weather generator approach. One example for a deterministic non-linear model doing distribution-wise downscaling would be quantile mapping (e. g., Schmidli et al., 2006; Piani et al., 2010a; Themeßl et al., 2011; Wilcke et al., 2013) (more on that method later). Another example for MOS is a stochastic linear model like XCDF-t (Kallache et al., 2011), which is an extension of the non-parametric cumulative distribution function (CDF) method (CDF-t) developed by Michelangeli et al. (2009) for specific application to extremes. Here the CDFs of simulated and observed extreme precipitation is linked.

A more detailed overview on statistical downscaling methods can be found in Maraun et al. (2010); Fischer et al. (2013).

If you thought that science was certain - well, that is just an error on your part.

— Richard P. Feynman

3

ERROR CHARACTERISTICS OF REGIONAL CLIMATE MODELS AND BIAS CORRECTION

Regional Climate Model (RCM) output is the state-of-the-art climate data available for regional and local impact studies (Finger et al., 2012; Heinrich and Gobiet, 2012). As mentioned above, RCMs are nested into General Circulation Models (GCMs) and driven by their lateral boundary conditions. That, of course, means that RCMs adopt systematic errors of GCMs and, additionally, have biases themselves (Christensen et al., 2008; Giorgi and Coppola, 2010). Pielke and Wilby (2012), e. g., remark that RCMs amplify GCM errors, which will not be discussed here. Some errors can be corrected by the RCM, in particular orography related ones (Diaconescu et al., 2007; Laprise, 2008). Here, I will focus on the RCM errors but include the ones from GCMs where needed.

3.1 ERRORS IN RCMS

Errors and uncertainties in RCMs have various sources. In this section some of them will be discussed, whereby the presented analyses focus on Europe and on RCMs from the ENSEMBLES project (described in Section 2.2.1 and Section 5.1).

Systematic errors have been found from early on (e. g., Murphy, 1999; Frei et al., 2003; Hagemann et al., 2004). Jacob et al. (2007) investigated systematic errors in RCMs related to the ability to simulate the long-term mean climate and the inter-annual variability for temperature and precipitation. They found for summer months that temperature seemed to be systematically biased from the RCMs, while precipitation gains the bias from the lateral boundary conditions, i. e. inherited from the GCM. The bias in winter climate seemed to be mainly introduced by the GCM, while the transition seasons show no clear system. Those biases can be explained with, e. g., insufficient blocking frequencies, too zonal simulations, differences in model surface schemes. RCMs tend to overestimate temperature in summer months in southern Europe and underestimate winter precipitation sums in the North (Jacob et al., 2007; Christensen et al., 2008).

On the other hand, Kjellström et al. (2010) found a warm bias in northern Europe for daily minimum temperature. Lenderink (2010) shows positive systematic biases in extreme precipitation for whole

Europe, which is in agreement with van der Linden and Mitchell (2009) and Kjellström et al. (2010) for extreme conditions. The wet-day frequency over Europe is known to be overestimated by RCMs as well (Leander and Buishand, 2007; Piani et al., 2010b).

There are different sources for systematic errors in RCMs. Examples are the necessary physical parameterisations, ill-posed lateral boundary conditions, and non-implemented processes. Those are described below.

Parameterisations, e. g., deep and shallow convection, micro physics in cloud processes, which are the least understood (Solomon et al., 2007), radiation parameterisation, small-scale turbulence, which have been already mentioned in Section 2.2.1. Two major sources of error herein are conceptual shortcomings in schemes themselves and if the data fed to the schemes are incomplete or inaccurate (Knievel, 2008). Awan et al. (2011), using a non-hydrostatic model over complex terrain, points out that errors are dependent on the simulated sub-regions. Thus, improvements can be made by adjusting the configurations to the region of interest.

Giorgi et al. (1994); Evans (2003) show that the land surface processes are poorly represented by RCMs which results in poor agreement between simulated and observed surface runoff (Rojas et al., 2011).

Kanamaru and Kanamitsu (2007) found noise in RCMs due to wave reflection and instabilities. Those are introduced by mathematically ill-posed lateral boundary conditions, the driving coarse model field, and inconsistencies between the model resolutions. Within a buffer zone of several grid points along the RCM boundary, a lateral boundary relaxation method can deal with those wave reflections (Davies, 1976; Marbaix et al., 2003).

Castro et al. (2005) states that if the variability of synoptic features is underestimated or there is a consistent bias in the large model, no increased skill would be gained by dynamical downscaling. This is the "garbage in, garbage out" effect, mentioned in Section 2.2.1

The higher resolution of the RCMs, compared to the GCMs, can correct for orographic errors, but also does not erase them. In particular over complex terrain like the European Alps or Scandinavia the biases are considerable (Christensen et al., 2008; Suklitsch et al., 2011; Kjellström et al., 2010; Kotlarski et al., 2010).

For hydrological applications, like stream-flow simulations, run off or flood analysis, raw RCM output is problematic (Kidson and Thompson, 1998; Murphy, 1999; Wilby et al., 2000; Wood et al., 2004). E. g., precipitation can be overestimated by 20 % to 50 %, which leads to an overestimation of, e. g., run-off (e. g., Wood et al., 2004). This makes bias correction important and necessary (Wood et al., 2004; Ashfaq et al., 2010).

3.2 REVIEW ON BIAS CORRECTION

Motivated by the errors mentioned in the former section, different types of empirical statistical bias correction methods were and still are developed. Here, it should be emphasised that all empirical statistical methods assume a stationary climate or at least a stationary bias when applied to future periods.

Even though non of the methods completely correct biases or errors of climate model output, and despite ongoing discussions on the appropriate phrasing, here, I stick to the naming *bias correction* (BC) in the following. Most of the following methods have been developed and applied for temperature or precipitation or both.

The most simple approach is a delta correction, which removes an average bias (delta) for a certain period (e. g., Lehner et al., 2006; Lenderink et al., 2007). This can be done either as one delta for the whole period considered, or for different steps like seasonal or monthly deltas. Engen-Skaugen (2007) uses empirical factors to tailor the RCM outputs considering normalisation which adjusts the standard deviation and calculation of residuals.

Another approach calculates monthly correction factors which are based on the ratio between observed and simulated values in the past (e. g., Durman et al., 2001).

Multiple linear regression is used to produce linear transformation functions between one or more predictors and the predictand (e. g., Kilsby et al., 1998; Hay and Clark, 2003; Horton et al., 2006). Those adjust changes in the mean and variance of the observed and simulated time-series. To cope with the reduced variability, von Storch (1999) introduced a noise term, which randomly adds variability and leads to better results (Themeßl et al., 2011).

Local intensity scaling (e. g., Widmann et al., 2003; Leander and Buishand, 2007; Dobler and Ahrens, 2008) for precipitation sums corrects for rain day frequency and intensity biases as well as inaccurate orography (Schmidli et al., 2006). "*The difference between the evaluation period model precipitation and model rain day threshold is multiplied by the scaling factor and added to the observation rain day threshold (1 mm/d)*" (Dobler and Ahrens, 2008). Here, the scaling factor is the ratio of the mean rain day intensity of observed to simulated precipitation.

Power transformation for precipitation corrects not only the mean but also the variance by applying a non-linear correction, e. g., exponential, (Leander and Buishand, 2007; Leander et al., 2008). Here the observed long-term monthly mean is mapped on the monthly mean of the corrected daily simulated precipitation series.

Variance scaling corrects the mean and the variance of temperature time-series (Chen et al., 2011). It is a step-wise approach by first applying a linear scaling. In a second step the modified control and scenario time-series are shifted on to a monthly zero mean.

The third step scales the standard deviations to the ratio of observed and control-run standard deviations. The last step shifts the standard-deviation-corrected time-series back using the corrected mean of the first step.

Analogue methods are advanced re-sampling, where a historical atmospheric state is searched that closely re-samples the synoptic situation on a given day to be simulated (e. g., Cubasch et al., 1996; Zorita and von Storch, 1999; Wetterhall et al., 2005; Matulla et al., 2008). Here, no assumptions about the statistical distributions of the variables, spatial and temporal structure of the field, and mutual dependencies between the variables need to be made. If the synoptic situations are randomly chosen from the historical pool, the method is called nearest neighbour analogue method (e. g., Brandsma and Buishand, 1998).

A rather modern approach is called *Quantile Mapping* (QM), *quantile matching*, *distribution mapping*, or *histogram equalisation*. QM was first mentioned by Brier and Panofsky (1968) as empirical transformation. Déqué (2007); Boé et al. (2007) first used it for downscaling and error correction. Quantile based methods are getting more popular lately and have been applied to downscale and correct temperature and precipitation data from RCMs (e. g., Dobler and Ahrens, 2008; Piani et al., 2009, 2010b,a; Ashfaq et al., 2010; Dosio and Paruolo, 2011; Themeßl et al., 2012), in particular in hydrological studies. The applications differ in the details, e. g., Piani et al. (2009) only considered wet days, e. g., Ashfaq et al. (2010) calculated monthly transfer functions, and e. g., Themeßl et al. (2012) used daily transfer functions. Rojas et al. (2011) used it as histogram equalisation method to correct precipitation and mean, minimum, and maximum temperatures, assuming a stationary error model. As QM is the main topic of this work, it will be explained in more detail in Chapter 4.

Themeßl et al. (2011) compared most of the above mentioned downscaling and error correction methods applied to daily precipitation sums of an RCM for the region of the European Alps. They showed that a QM method performs best for daily precipitation. All methods correct the median bias to zero, as well as the variability is corrected by most methods, but multiple linear regression and local intensity scaling. The analogue methods and multiple linear regression reduce the correlation, but none of the other methods increased the correlation. The drizzling (e. g., Gutowski et al., 2003) is corrected by QM and local intensity scaling.

Teutschbein and Seibert (2012) compared similar BC methods as Themeßl et al. (2011) but with focus on hydrological applications, here streamflow simulations. They aggreed on the success of each method, but also found differences in the level of each success. Only the power transformation and distribution mapping were able to cor-

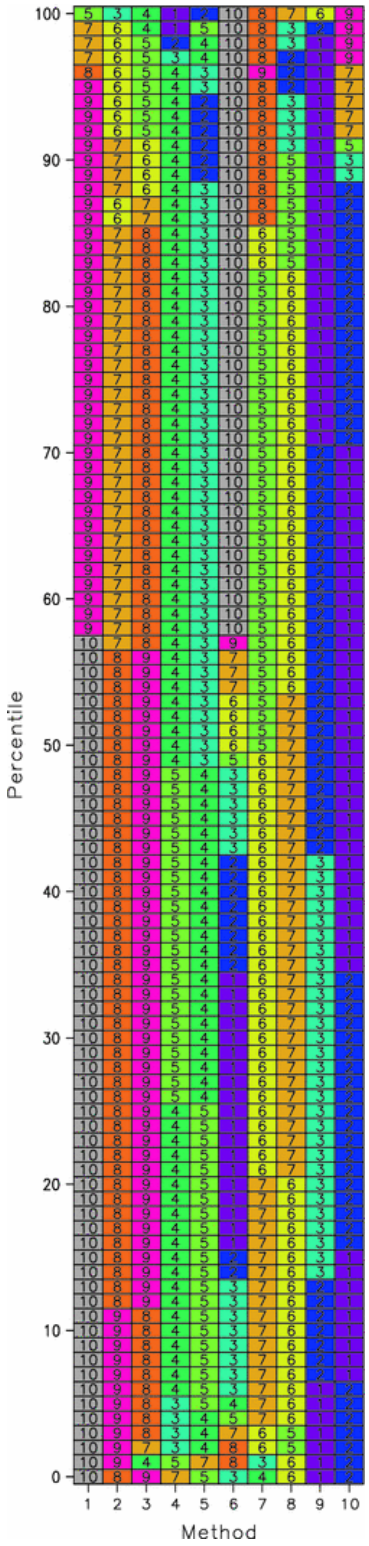


Figure 3.1: Ranking of ten methods for cross-validated mean squared error of different percentiles of the temperature distribution in 2069 to 2098 (Figure 8 in Räisänen and Räty, 2012). Cells numbered 1 (dark blue) indicate best performance, cells numbered 10 (grey) worst. M 9 is QM with running means and M 10 is QM with linear fit.

rect for more than the mean. As Themeßl et al. (2011), they agree on QM as the best BC method of those tested.

Räisänen and Räty (2012) compared five delta change methods and five bias correction methods for application on future climate temperature in a pseudo-reality set up. The BC methods are (1) just correction for the mean bias, (2) correcting for the mean bias and standard deviation, (3) correcting for the mean, standard deviation, and skewness, (4) correcting with QM and smoothing (running averages) for the tails of the distribution to avoid noisy signals there, (5) correction with QM and linear fit. Figure 8 in Räisänen and Räty (2012) (see Figure 3.1) shows impressively the good performance of the two QM approaches throughout the distribution, with exception of the higher percentiles for the linear fitted approach.

Chen et al. (2011) found that the uncertainty from the choice of decade(s) for BC method calibration is minor compared to uncertainties from the choice of climate model or choice of greenhouse gas concentration scenario (SRES/RCPs). Piani et al. (2010b) found the choice of decade(s) for calibrating the BC method (QM) the major source of uncertainty, compared to fit error, choice of transfer function, or observational uncertainty.

Bordoy and Burlando (2013) approved the statement of strong improvement of RCM output by applying QM over the complex terrain of the Swiss Alps as Themeßl et al. (2011, 2012) for the Austrian Alps.

A possible drawback is the influence of BC on the climate change signal (CCS) (Haerter et al., 2010). Themeßl et al. (2012) analysed the impact of QM on the CCS (temperature and precipitation) and found moderate changes. They found that by correcting the bias, the trends in the scenarios are modified. The question of how to judge the modifications of the CCS by BC methods, is tackled in on-going studies (e. g., Gobiet et al., 2013).

Happy is he who gets to know the reasons for things.

— Virgil

4

BIAS CORRECTION — QUANTILE MAPPING

The bias correction (BC) method Quantile Mapping (QM), how it is used in this work, has been proposed by Themeßl et al. (2011) to correct RCM simulations and is based on Déqué (2007). Themeßl et al. (2012) demonstrated its successful application to future scenarios for precipitation. The process combines downscaling aspects with model BC. The correction of the altitude differences between the model and actual orography of the observational data-set is implicitly included.

The distributions used for the mapping can be gained empirically or parametrically. Studies like Gutjahr and Heinemann (2013) showed the better skill of empirical QM over parametric QM.

The QM method used in this PhD thesis is purely empirical (i. e. no assumption about the distributions of the meteorological variables are made) and is based on Themeßl et al. (2012). In the following Section 4.1 the application on temperature is described. Adaptation to the specific requirements of different variables are described in Section 4.2.

4.1 METHOD DESCRIPTION

This implementation of QM fits modelled daily empirical cumulative distribution functions (ECDFs) to corresponding observational ECDFs of one station or grid-point. For each day of year (DOY) in the calibration period, ECDFs are constructed using a sliding window of 31 days (i. e. ± 15 days around the day of interest), which results in, e. g., 620 values for 20 years of calibration. The daily ECDFs sensibilise the correction to varying error characteristics throughout the year. The ECDFs are calculated by sorting the values into bins with adjustable width. The bin width is set to the resolution of the observational data (mostly 0.1) and a linear interpolation is applied between two percentiles (bins) (Déqué, 2007). In some cases, particularly in the case of low-quality observational data, this interpolation can lead to inadequate approximations, as will be demonstrated in Section 6.2.2 and Section 7.2.

I want to emphasise that daily ECDFs were calculated and they were applied on daily data, thus there is no temporal scale discrepancy. Spatially, the discrepancy was minimised by using RCMs instead of GCMs.

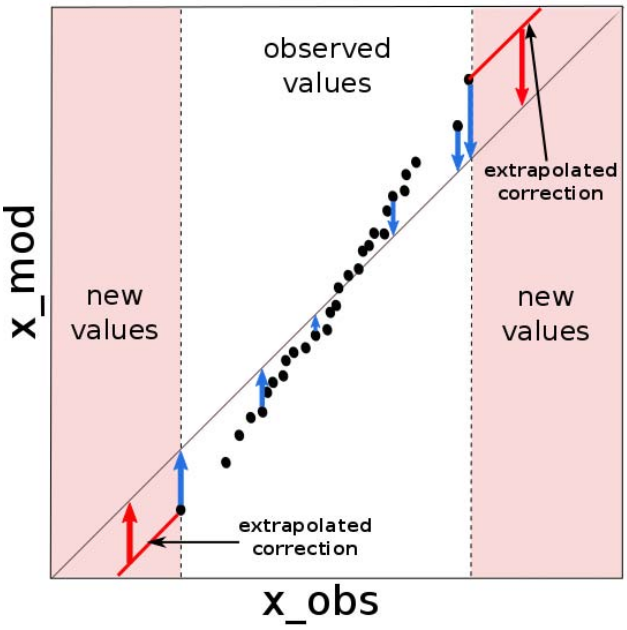


Figure 4.1: Sketched Q-Q plot for any simulated and observed variable x_mod and x_obs. Red line indicating the parallel extrapolated correction value, red region indicating possible future new extremes, blue arrows showing schematical correction for values in calibration sample.

For the correction of temperature, which is described here exemplarily, the cumulative probability p for a modelled value is calculated at point i and day t (Equation 4.1). The correction term $\Delta x_{t,i}$ is calculated as a difference of the inverse ECDFs (quantile) of the observation ($\text{ECDF}^{\text{obs,cal}^{-1}}$) and model ($\text{ECDF}^{\text{mod,cal}^{-1}}$) for probability p at a certain day of a station (Equation 4.2).

$$p_{t,i} = \text{ECDF}^{\text{mod,cal}}_{\text{DOY},i}(x^{\text{mod,raw}}_{t,i}) \tag{4.1}$$

$$\Delta x_{t,i} = \text{ECDF}^{\text{obs,cal}^{-1}}_{\text{DOY},i}(p_{t,i}) - \text{ECDF}^{\text{mod,cal}^{-1}}_{\text{DOY},i}(p_{t,i}) \tag{4.2}$$

$$x^{\text{mod,cor}}_{t,i} = x^{\text{mod,raw}}_{t,i} + \Delta x_{t,i} \tag{4.3}$$

The correction term $\Delta x_{t,i}$ is then added to the raw model value $x^{\text{raw}}_{t,i}$ (Equation 4.3).

For correcting simulations into the future, the procedure is very similar. In Equation 4.1 $x^{\text{mod,raw}}_{t,i}$ is exchanged with $x^{\text{mod,raw}_{\text{future}}}_{t,i}$. That means, the ECDFs of the calibration period (in the past) are used to gain $p_{t,i}$ and the $\Delta x_{t,i}$ is applied to future values.

In order to avoid the suppression of new extremes in the future periods (i. e. values outside the calibration range), an extrapolation

of the correction has been implemented, by keeping the correction term of minimum and maximum values constant outside the observational range (compare Figure 4.1). This approach for extremes has been recently confirmed by Bellprat et al. (2013).

4.2 TECHNICAL DETAILS

Dealing with precipitation, a frequency adaptation is implemented to parry a deficiency of QM leading to a wet bias. That bias occurs if the dry day frequency in the raw model output ($ECDF^{mod,cal}$) is larger than in the observations ($ECDF^{obs,cal}$), which would lead to a strong positive bias after the correction (Themeßl et al., 2012). Thus, the model data below 0.1 mm/d is divided to finer bins with width of 0.01. Dry days are generated by randomly sampling the observational distribution into the first bin (0 mm/d to 0.01 mm/d). However, this bias is a rare case. More often the model overestimates the light precipitation frequency ("drizzling-effect"; e. g., Gutowski et al., 2003), which is caught by QM automatically.

The correction of relative humidity required minor adaptations, since the interpolation can lead to values outside the physically reasonable range, in particular to values above 100 %. These non-physical values were set to the maximum value observed on the corresponding ECDF. This adaptation, however, only takes effect in very rare cases and the overall effect can hardly be noticed in climatological evaluation.

For wind speed, global radiation, and surface pressure QM works straightforward, as it has been implemented for temperature.

However, the observed values of wind speed, which have been used in this work, are of poor quality (compare Section 5.3), which leads to a small systematic bias even after applying QM. In a test noise was added, which removed this residual bias, but adding noise is no robust solution for producing time-series for further climate studies. Thus, no changes to the code were made, and the residual bias was accepted so far.

It is a capital mistake to theorise before one has data. Insensibly one begins to twist facts to suit theories instead of theories to suit facts.

— Sherlock Holmes

5 DATA USED IN THIS WORK

5.1 RCM OUTPUT

Daily mean Regional Climate Model (RCM) data were derived from the multi-model data-set of the ENSEMBLES project (Goodess et al., 2009). The RCMs have a horizontal grid-spacing of 0.22° and cover entire Europe (Figure 5.1). ENSEMBLES provides *hindcast simulations*, i. e. the lateral boundary conditions for the RCMs are given by re-analysis data. For this project the ECMWF 40-year re-analysis (ERA40) data-set (Uppala et al., 2005), has been used. The hindcast simulations cover the historical period from 1961 to 2000.

In the ENSEMBLES project, transient time-series from 1951 to 2100 have been simulated. For the historic years (1951 to 2000) the General Circulation Models (GCMs) are driven with observed emissions. RCM runs driven by the historical period of the GCMs are called *historical runs*. For future projections different greenhouse gas emission scenarios (SRES) are integrated in the GCMs. Here, the SRES A1B scenario (van der Linden and Mitchell, 2009) was used, which is part of the A1 emission family (see Section 2.1). The RCM runs from 2001 to 2100 are the *scenario runs*.

Table 5.1: Selected RCMs (historical, SRES A1B, and re-analysis) from ENSEMBLES for this thesis. Full institute names and information on specific models can be found in the Acronyms.

INSTITUTE	RCM	GCM	RE-ANALYSIS
ICTP	RegCM3	ECHAM5-r3	ERA40
SMHI	RCA	BCM	ERA40
METNO	HIRHAM	HadCM3Q0	ERA40
C4I	RCA3	HadCM3Q16	ERA40
KNMI	RACMO2	-	ERA40
METO-HC	HadRM3Q0	-	ERA40

Out of the ensemble of 22 RCMs from the ENSEMBLES project, running transient time-series on A1B, six simulations (c. f. Table 5.1) have been selected for use in this work. The selection criterion for the

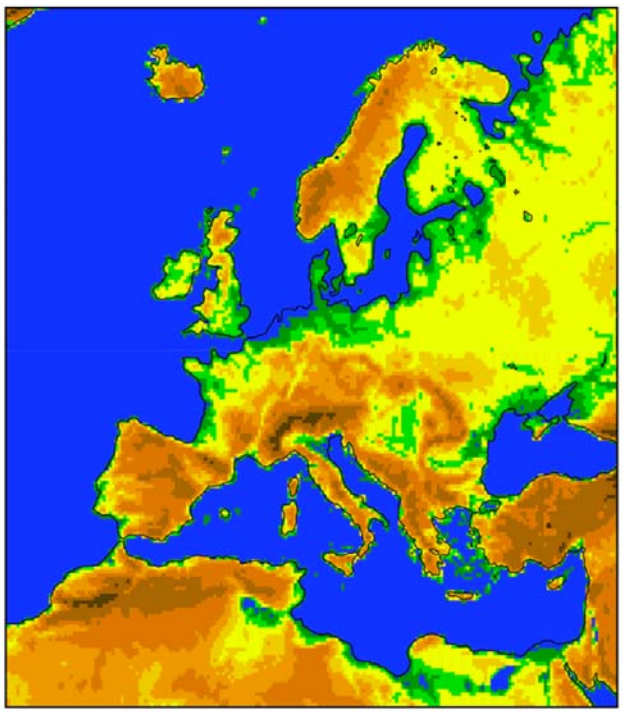


Figure 5.1: ENSEMBLES domain. Figure 1.2.1 of the ENSEMBLES deliverable report D2B.9 (C4I, 2008) to show domain covered by climate model data.

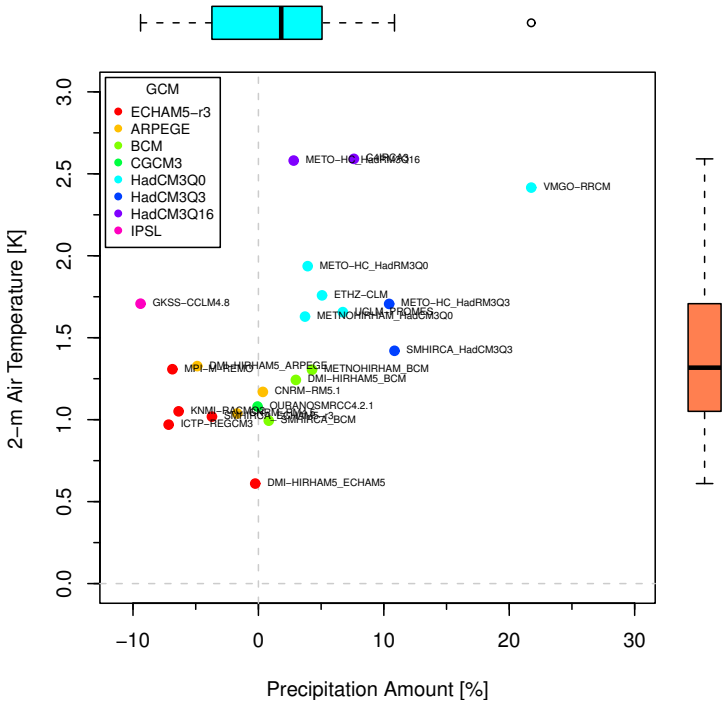


Figure 5.2: Climate change signal of temperature and precipitation for Tirol for winter season 2021 to 2050 minus 1971to 2000 simulated by ENSEMBLES A1B RCMs (Figure 2.1 of Wilcke et al., 2012).

first four was based on two projects (CC-Snow & CC-SnowII) funding this study. The selection criterion was to cover to full uncertainty range of mid-century climate change of temperature and precipitation (Déqué et al., 2007, 2011). Figure 5.2 shows the climate change signal (CCS) for a county in Austria (Tirol) for the winter season (as the projects were related to winter tourism). Other regions in Austria show a similar picture (Wilcke et al., 2012). Compared to the other ENSEMBLES simulations Community Climate Consortium for Ireland (C4I)-Rossby Center Regional Climate Model (Samuelsson et al., 2011) (RCA)3-Hadley Centre Coupled Model, version 3, sensitivity Q16 (HadCM3Q16) shows a strong warming and wetter conditions in this study region, The Abdus Salam International Centre for Theoretical Physics (ICTP)-RegCM3 (RegCM3)-ECHAM5 (ECHAM5)-r3 shows little warming and drier conditions, Norwegian Meteorological Institute (METNO)-HIRHAM (HIRHAM)-Hadley Centre Coupled Model, version 3, sensitiviy Q0 (HadCM3Q0) shows moderate warming and moderate change in precipitation, and Sveriges Meteorologiska och Hydrologiska Institut (SMHI)-RCA-Bergen Climate Model (BCM) shows little warming and wetter conditions in the future (Wilcke et al., 2012).

Six meteorological variables of those RCMs were obtained:

- 2 m air temperature [°C] (tas)
- precipitation amount [mm/d] (pr)
- relative humidity at 2 m [%] (hurs)
- wind speed at 10 m [m/s^2] (wss)
- surface air pressure [hPa] (ps)
- surface down-welling shortwave radiation (here called global radiation) [W/m^2] (rsds)

Those variables were used in 3-hourly (as input for the snow model (see Section 5.2) and in daily resolution for this study.

5.2 SNOW MODEL DATA

The following description of the model is partly from Strasser (2008).

AMUNDSEN is a modular, physically based, distributed modelling system, which is applied to simulate snow cover. It has been designed especially for mountainous regions under climate change conditions (Strasser, 2004; Strasser et al., 2008). AMUNDSEN is driven with subdaily data from five meteorological variables:

- 2 m temperature
- precipitation amount
- 2 m relative humidity

- down-welling shortwave radiation (global radiation)
- 10 m wind speed

As AMUNDSEN runs on sub-daily meteorological data, the RCM output needed not only correction and spatial downscaling, but temporal downscaling also. I basically applied Quantile Mapping (QM) on the daily RCM output and added the diurnal cycle given by the 3-hourly RCM output (details in Section A.1).

Other variables are simulated, like shortwave and longwave radiation including topographical effects (e. g., shadowing) and clouds (Corripio, 2003; Greuell et al., 1997). Snow-melt is simulated on basis of an energy balance model (Strasser et al., 2008). The snow albedo is parameterised depending on snow age and temperature (Rohrer, 1992), as well as the snow density (Anderson, 1976; Jordan, 1991). To account for the shadowing effects the snow model simulated a 25 km $\times$ 25 km domain over the station of interest, with a grid spacing of 100 m.

From AMUNDSEN the following output was used on daily basis and on station locations:

- snow depth (SD)
- By distinguishing the densities of new and old snow the snow water equivalent (SWE) was derived
- snow fall amount (SF), which is solid precipitation and defined with a wet-bulb temperature threshold of 2 °C.

5.3 OBSERVATIONS

Observational data for this work were obtained from the Austrian Central Institute for Meteorology and Geodynamics (ZAMG) and the Swiss Federal Office of Meteorology and Climatology (MeteoSwiss). The data are on point-scale and result from meteorological measurements on observation stations. For the different studies in this work about 80 stations (depending on the variable) distributed over entire Austria were used (Figure 5.3). For Switzerland only 18 stations were used, which are situated in the Rhone catchment. The Austrian data cover a 50-year period from 1961 to 2010, while the data from MeteoSwiss range from 1981 to 2010.

To ensure a sufficiently large sample size only stations with more than 80 % (compromise from experiences with missing values) data coverage for a period from 1961 to 2010 are included (Swiss data had no gaps). When instruments of a station were moved, like a couple of meters in any direction, the station got a new ID. As the study depends on sufficiently long time-series, those shifted stations were merged. To consider inhomogenities in time-series due to vertical shifts of the instruments, which is not intended, a Wilcoxon rank

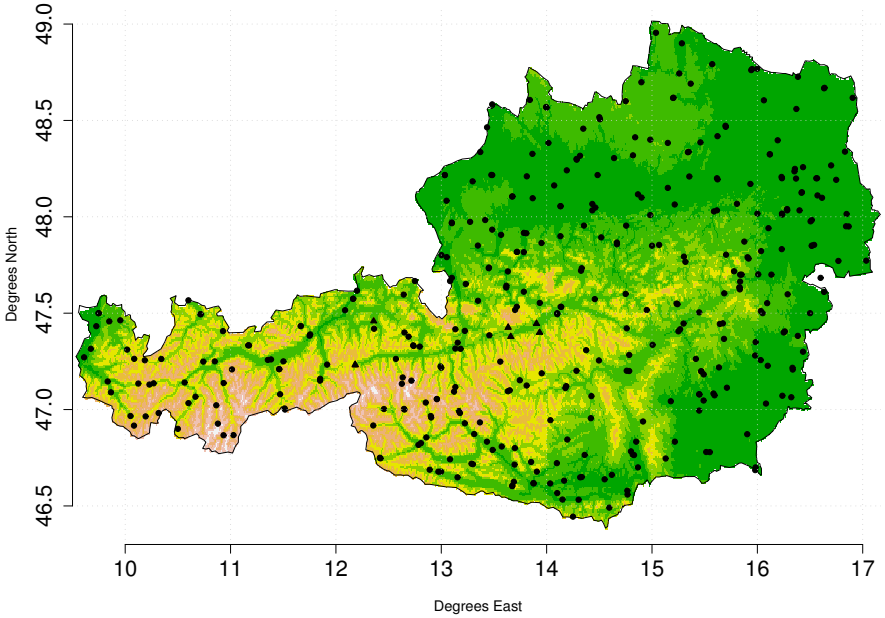


Figure 5.3: Observation station locations of the ZAMG network.

test (e. g., Wilks, 2006) has been applied to the time series of interest. In case the test fails, the stations were not merged.

AGGREGATION FROM POINT-SCALE TO RCM GRID: For the validation of inter-variable relations in Section 7.3, single grid cell time-series have been evaluated against averaged time-series of observation stations inside those grid cells. This has been done to avoid uncertainties introduced by further downscaling to point-scale (Maraun, 2013). To ensure that the station averages can be regarded as reasonably comparable with the gridded RCM output, attention was restricted to the seven grid cells containing the highest number (9 to 13) of stations. The coordinates of the chosen grid cell centres are listed in Table 5.2 and visualised as asterisks in Figure 5.4. In whole 71 stations have been included (see Table C.2). Figure 5.4 shows the distribution of the grid cells in space. The colours of points and boxes indicate their relation. The stations cover a wide range of geographical characteristics, compare box 3 (violet), which is located over high mountains, and box 7 (light blue), which is located in a flat region.

VARIABLES: In different parts of this study different combinations of meteorological variables are used. However, in sum there were eight variables as daily mean or daily sums employed from the ZAMG for the period 1961 to 2010 and MeteoSwiss for the period 1981 to 2010:

- 2 m air temperature [°C] (tas)
- precipitation amount [mm/d] (pr)
- relative humidity at 2 m [%] (hurs)

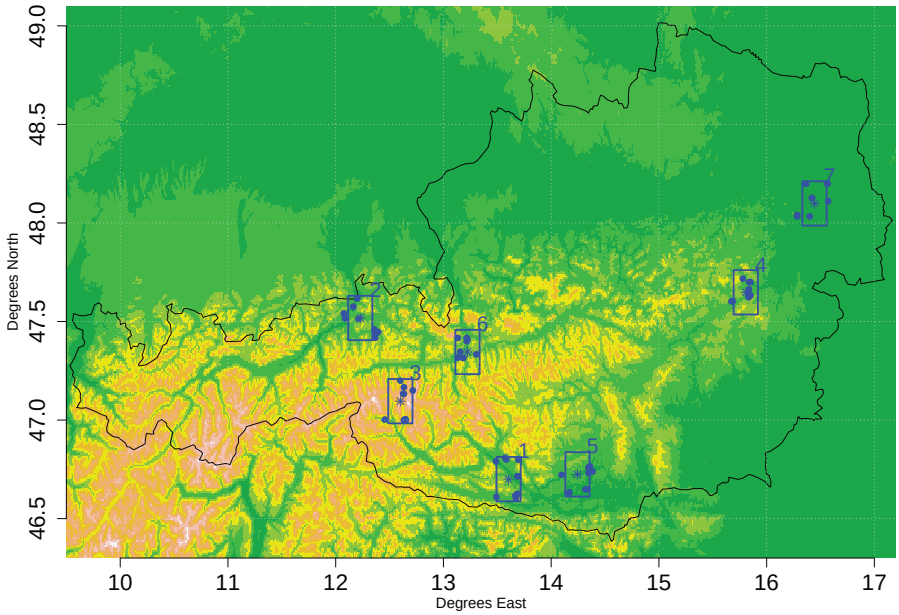


Figure 5.4: Evaluated grid cells and observation stations in Austria. (1) Villach region, (2) Kitzbühel Alps, (3) Hohe Tauern Nationalpark, (4) Semmering, (5) Basin of Klagenfurt, (6) Pongau, (7) Vienna.

Table 5.2: Coordinates of chosen grid cell centres in Austria.

no.	**grid cell name**	**lat East**	**lon North**
1	Villach region	46.7003	13.6047
2	Kitzbühel Alps	47.5179	12.2284
3	Hohe Tauern Nationalpark	47.0957	12.6004
4	Semmering	47.6481	15.8059
5	Basin of Klagenfurt	46.7245	14.2440
6	Pongau	47.3454	13.2232
7	Vienna	48.0991	16.4443

- wind speed at 10 m [m/s^2] (wss)
- surface air pressure [hPa] (ps)
- global radiation [W/m^2] (rsds)
- snow water equivalent [mm/d] (SWE)
- snow depth [cm] (snow)

While part of this work aims for correcting errors in RCMs, errors in observations are not covered. Many studies analyse the quality of meteorological observations (e. g., Auer et al., 2001; Caussinus and Mestre, 2004; Della-Marta and Wanner, 2006; Frei and Schär, 1998; Schmidli et al., 2002) and the author is aware of those shortcomings. In particular the wind induced undercatch of precipitation (e. g., Frei et al., 2003; Guo et al., 2001; Nešpor and Sevruk, 1999) needs to be considered. The influence of observation's quality on model and method evaluation is strong. However, in this work the observations obtained by ZAMG and MeteoSwiss are defined as truth. Nevertheless, this limitation is remembered through out this work.

5.4 TEMPERATURE-HUMIDITY INDEX

Any derived index is like a small impact model. As an example I examined the influence of QM on RCM output via the temperature-humidity index (THI). The THI tries to describe the effect of high temperature and humidity on a human body, how it "feels" relative to a dewpoint of 14 °C (Steadman, 1979). Steadman (1979) published a THI which was based on several biometeorological studies. The studies included the effect of clothing, solar radiation, wind chill, activity, and build of a body on the felt temperature. Each effect can be described by an equation or empirical value, which are listed in, e. g., Rothfusz (1990). Various multiple regression models, considering relative humidity and dry-bulb temperature, were developed to calculate the THI (e. g., Rothfusz, 1990). Schoen (2005) found relative humidity and dry-bulb temperature too dependent on each other and proposed to use "true independent" variables: dry-bulb temperature (T) and absolute humidity. As absolute humidity is rarely reported by observation stations, the dewpoint (D, see Equation A.3) is close enough to cover for absolute moisture.

The empirical model to calculate the THI developed by Schoen (2005) is given by

$$\mathrm{THI} = T - 1.0799 \exp(0.03755T) \left[1 - \exp(0.0801(D - 14))\right], \quad (5.1)$$

where THI, T, and D are all in degrees Celsius. For details on calculating the dewpoint see Section A.2.

Nothing in life is to be feared, it is only to be understood. Now is the time to understand more, so that we may fear less.

— Marie Curie

6 DIRECT VALIDATION OF QUANTILE MAPPING

6.1 OUTLINE OF VALIDATION CONCEPT

The bias correction method Quantile Mapping (QM) is applied on individual RCM output variables. The "direct" validation, here, evaluates how well QM corrects those variables, thus the "direct" effects of QM on RCM output. Climate change impact research needs regional climate scenarios of multiple meteorological variables. This typically includes temperature, precipitation, relative humidity, global radiation, and wind speed (e. g., Finger et al., 2012), which are therefore the focus of this study. They are usually available from RCMs, but affected by considerable biases.

The question, however, whether error correction degrades temporal characteristics is an open issue. As many climate impact models use multiple variables at the same time, consistent time-series for each are important.

In this study the performance of QM has been investigated, when applied to a multi-variable output of four RCMs with a grid-spacing of 0.22° (Section 5.1) on daily basis. The particular focus of this study is on biases, frequency distributions, and temporal structure of time-series in present climate.

A split-sample evaluation approach has been used to mimic the application to future climate as far as possible. The calibration period has no overlap with the evaluation period. The available observation period from 1971 to 2010 (1981 to 2010 for MeteoSwiss stations) is divided in halves, taking 1971 to 1990 as calibration period and 1991 to 2010 as application period, and vice versa. Of course, the split-sample approach can only give rough indication about the performance of QM in far future periods. However, since calibration and evaluation periods are independent and climate variability and change results in different climate characteristics in both periods, severe deficiencies can be expected to be detected.

In addition to the split-sample approach described above, evaluation was done also with equal calibration and application period, which is further on denoted as technical evaluation. This allows judging the performance of QM in an idealized world by neglecting climate variability and climate change. To keep the comparability with

Table 6.1: Observation stations of ZAMG and MeteoSwiss used for direct validation.

station name	**lon East**	**lat North**	**height** [m]
Wien-Hohewarte	16.35638889	48.24861111	198
Sonnblick	12.9575	47.05416667	3105
Innsbruck Uni	11.38416667	47.26	578
Zermatt	7.752468	46.029282	1608

the split-sample approach the technical test was performed on 20-year periods.

For the evaluation of the GCM-driven RCM simulations two basic statistics were inspected: bias and density distribution. Bias and density distribution describe the performance of QM on single variables and are analysed on the past only, since observations are indispensible. For the re-analysis driven RCM simulations, the temporal structure of time-series was analysed by the root mean squared error (RMSE) and autocorrelation.

Four stations (Table 6.1) with different climatological characteristics were analysed in depth. Hohe Warte in Vienna is located in a flat region between the north-eastern deviating veins of the Alps, in the north-western part of the Vienna basin. Sonnblick is a meteorological observatory on top of the mountain Hoher Sonnblick at 3105 m above sea level and is exposed to the free atmosphere. Innsbruck University maintains an observation station at 578 m height that represents a valley (Inn valley) with open ends on both sides. The valley lies in southwest-northeast direction and has a width of about 7 km. The station of Zermatt at 1608 m height lies at the inner end of the Matter valley, which extends in south-north direction and has a width of about 1.5 km.

6.2 RESULTS & DISCUSSION OF DIRECT VALIDATION

6.2.1 *Bias*

The bias is defined as long-term average difference between model and observation. For the ICTP-RegCM3 model the biases of temperature, precipitation, relative humidity, and wind speed are shown in Figure 6.1 and Figure 6.2, exemplarily. Figure 6.1 shows the annual mean bias of the raw and split-sample corrected model for stations in Austria. The biases for the other three models are presented in Figure B.1 to Figure B.3.

Figure 6.2 shows the annual cycle of monthly biases. In this case, the spatial error variability is indicated by box-and-whisker plots.

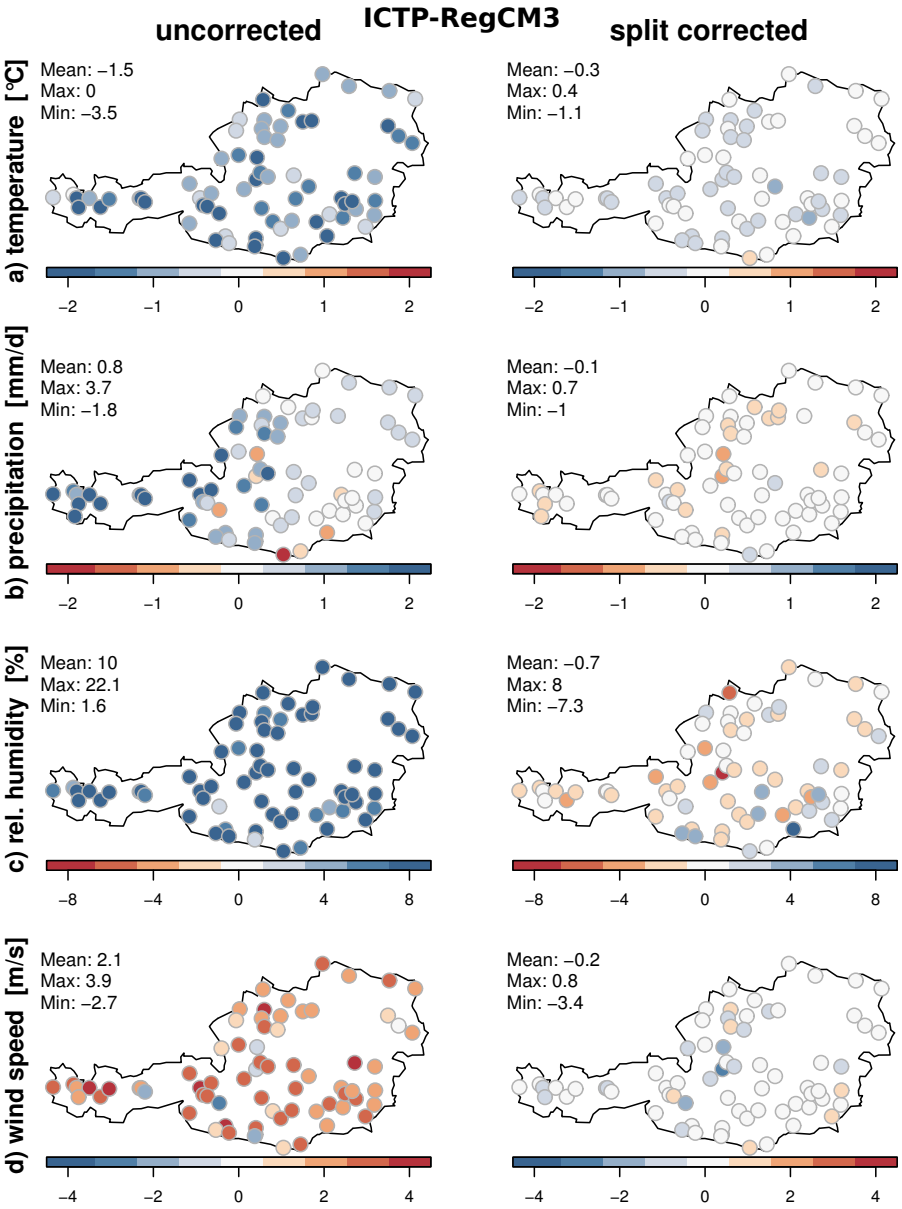


Figure 6.1: Annual mean RCM (ICTP-RegCM3) bias at observation stations (1991 to 2010) for a) temperature, b) precipitation, c) relative humidity, and d) wind speed (top down) for the raw RCM (for temperature altitude corrected), the error-corrected RCM with split-sample evaluation.

This analysis shows both versions of the split-sample approach, one evaluated in the period 1991 to 2010 and calibrated in the period 1971 to 1990 (blue), the other one vice-versa (green), together with the raw RCM bias of each period (red and orange). The results for the other models are presented in Figure B.4 to Figure B.6.

6.2.1.1 *Temperature*

QM performs very well in removing the annual mean temperature bias from the ICTP-RegCM3 model. In the split-sample evaluation (Figure 6.1 a), the mean bias over all stations is reduced from $-1.5\,°C$ to $-0.3\,°C$ and the biases of the individual stations range to below $1.1\,°C$ at maximum, compared to $3.5\,°C$ before correction. The results for the other models confirm these results.

The monthly temperature biases of the ICTP-RegCM3 model are exemplarily shown in Figure 6.2 a. The biases and their spatial variability are generally strongly reduced. In some months, however, considerable errors remain after the correction (e. g., in summer), which is caused by different model error characteristics in the calibration and the application period (i. e. by non-stationarity). Similar analysis for other models (Figure B.4 to Figure B.6) show that ICTP-RegCM3 is rather extreme in this respect and that bias correction of other models is mostly less affected by non-stationarity. Non-stationarity, however, is a limitation of bias correction (and empirical-statistical methods in general) and narrows their application to periods not too far in the future. Maraun (2012) investigated non-stationarities of a very simple bias correction method on a seasonal time scale and found that at the end of the 21st century bias correction only partly improves RCM results. Here, it could be demonstrated that the improvement in case of non-stationarity is still large for near future (20 years) applications, particularly when averaged over the year.

6.2.1.2 *Precipitation*

The ICTP-RegCM3 bias of precipitation is strongly improved in the split-sample evaluation with a bias reduction from $0.8\,mm/d$ to $-0.1\,mm/d$ (Figure 6.1 b). The results for other RCMs are very similar. On the monthly timescale (Figure 6.2 b), the split-sample evaluation generally results in smaller biases than the raw model and in a smaller bias range. The influence of non-stationarity is smaller than for temperature in this case. Bias correction of other models yields very similar results (Figure B.1 to Figure B.6).

6.2.1.3 *Relative humidity*

With regard to relative humidity the mean bias of ICTP-RegCM3 is reduced to close zero; the maximum of biases over the stations is reduced from 22 % to 8 % in the split-sample approach (Figure 6.1 c).

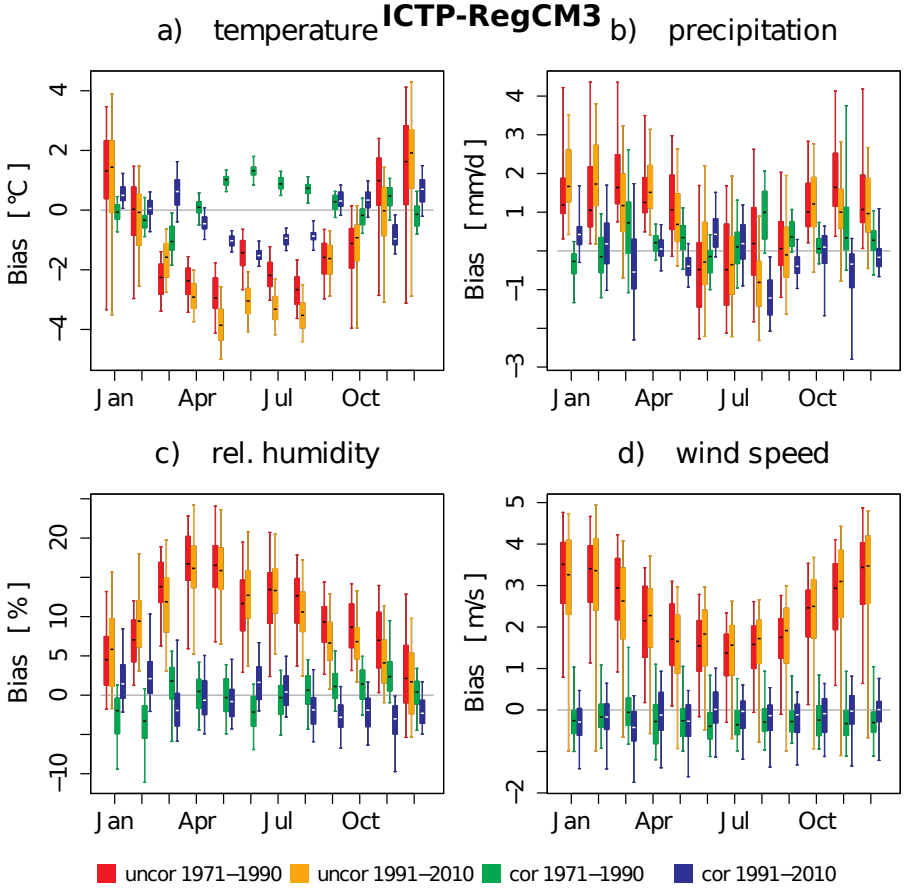


Figure 6.2: Monthly bias of a) temperature, b) precipitation, c) relative humidity, and d) wind speed as box-whisker plots for the raw RCM (for temperature altitude corrected) split in two periods (red and orange), the bias corrected RCM (ICTP-RegCM3) with split-sample evaluation (green and blue, see Section 6.2.1 for detailed description). Box and whiskers indicate the (spatial) variability of errors at the different Austrian stations. Boxes indicate the first (q25) and the third (q75) quantile, the whiskers extend to q5 and q95, and the black horizontal line indicates the median.

Figure 6.2c reflects similar results on the monthly scale. The median biases in both periods are close to zero in all months and the error range over the stations is smaller than for the raw RCM. The analysis of the other RCMs confirm these results, with some additional indication for non-stationarity in the SMHI-RCA model.

6.2.1.4 *Wind speed*

The error correction of wind speed leads to a reduction of the mean bias from 2.1 m/s to −0.2 m/s in the split-sample approach (Figure 6.1 d). The range of biases over the stations is only reduced from 6.6 m/s to 4.2 m/s. However, this small reduction is caused only by one single station. The mean bias of 0.2 m/s also is visible in the monthly bias of the split-sample analysis (Figure 6.2 d) and the technical evaluation (not shown). A similarly small remaining bias can

also be found in the other three RCMs. These remaining biases are mainly caused by lacking quality of the observational data: As mentioned in Section 5.3 and Section 4.1, an interpolation error occurs if the resolution of the observational data is much lower than the resolution of the RCM. Wind measurements sometimes have gaps in the data distribution of up to 0.6 m/s, already at small velocities. The vertical stripes in the scatter-plots of Figure 7.1 a (see Section 7.2) for wind speed demonstrates that. For testing, random noise was added to fill those gaps, which results in a mean bias of zero. Nevertheless, also with a remaining bias of 0.2 m/s after correction, QM strongly improves raw RCM output.

6.2.1.5 *Surface air pressure*

The main bias of surface air pressure results from the different altitudes of the model grid and the station. This effect is fully corrected by QM, and results in a mean bias of zero for the technical as well as for split-sample evaluation, for the entire year as well as for single months (not shown).

6.2.2 *Density distribution*

The correction of density distributions from all four RCMs is presented for the period from 1991 to 2010 (calibration: 1971 to 1990) on seasonal scale (Figure 6.3 for summer), and the entire year in Figure B.7. Four selected stations are evaluated (see Table 6.1). The density distributions of observation (black fat curve) were compered with raw (red thin lines) and error-corrected (green thin lines) RCMs for temperature, precipitation, relative humidity, wind speed, and global radiation.

Figure 6.3 demonstrates that the distributions of all models and variables are nicely adjusted to the observed distribution. Various distortions of the temperature distributions are corrected and the over-pronounced frequency of light and medium precipitation ("drizzling effect") is adjusted. More details on the performance of QM for daily precipitation including extremes are discussed in Themeßl et al. (2012). With regard to relative humidity, the overestimated frequency of near-100 % values and with regard to wind speed and global radiation, underestimated frequency of higher values are corrected.

Such adjustment of the distributions would be trivial if the calibration and evaluation periods were the same, but is quite remarkable in the split-sample analysis with two independent periods. In particular, some variables (e. g., relative humidity) and models (e. g., ICTP-RegCM3 for wind speed) feature distributions that are very different from the observation. Such strong modification of the distribution by error correction raises the question, whether the corresponding time-series and inter-variable relationships are still plausible after cor-

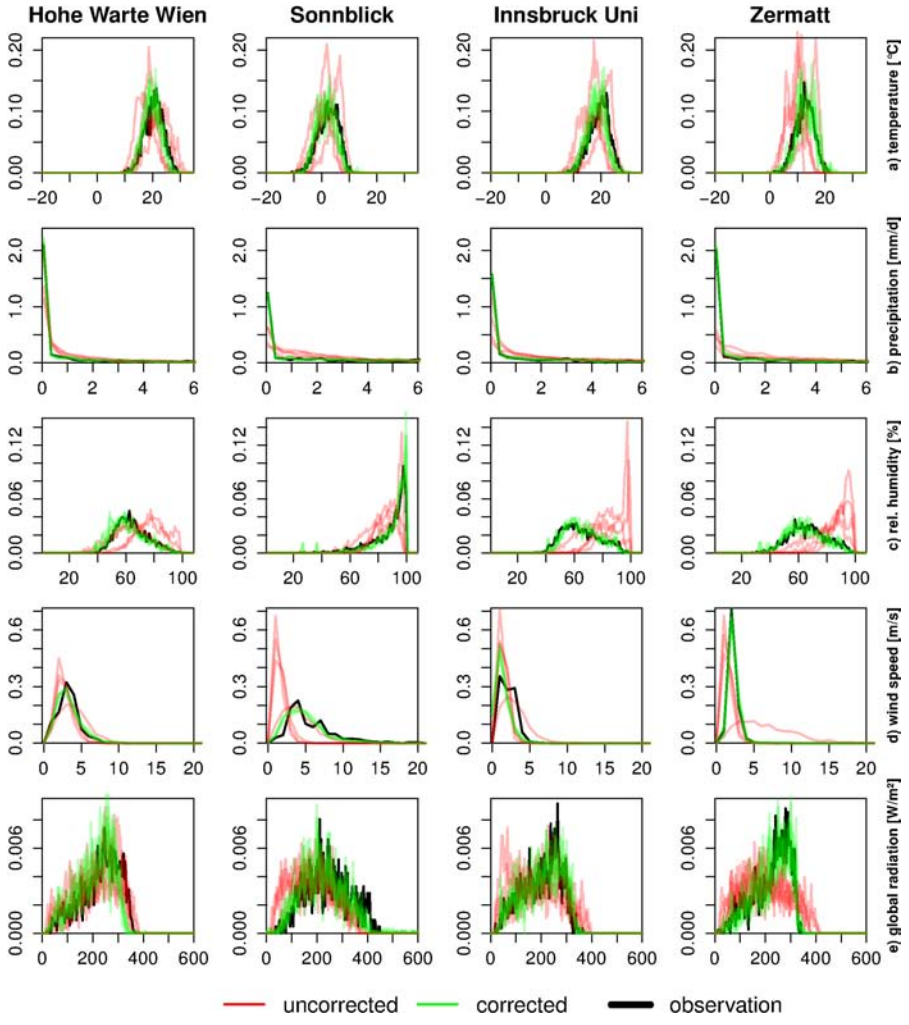


Figure 6.3: Density distributions of the four raw RCMs (red thin lines), the split-sampled error-corrected RCMs (green thin lines), and the observations (black fat line) for summers (JJA) for the period 1991 to 2010 of Hohe Warte Vienna, Sonnblick, Innsbruck University, Zermatt for a) temperature, b) precipitation, c) relative humidity, d) wind speed, and e) global radiation.

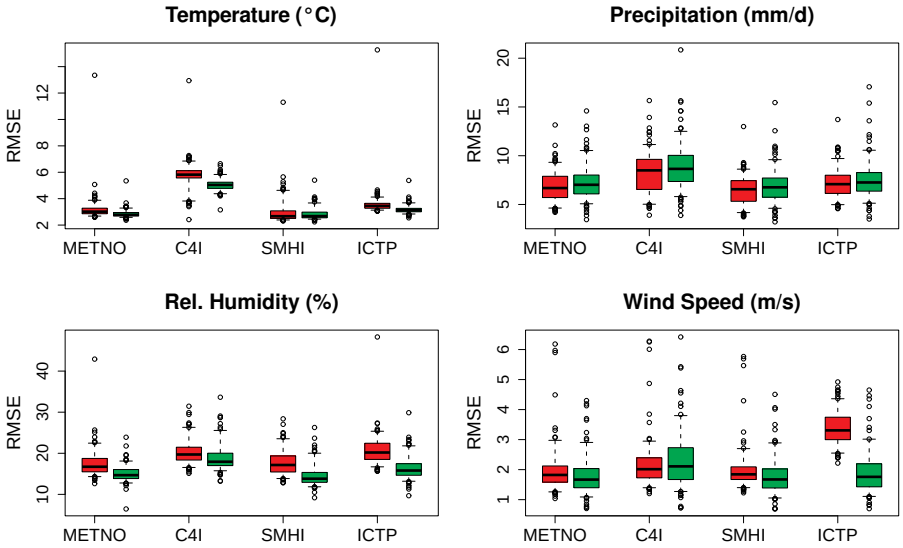


Figure 6.4: RMSE of ERA40 driven raw RCMs (red) and split-sample corrected (val: 1981 to 2000) (green) for whole years. Boxes contain RMSE of about 60 stations in Austria. Box and whiskers indicate the (spatial) variability of errors at the different Austrian stations. Boxes indicate the first (q25) and the third (q75) quantile, the whiskers extend to q5 and q95, and the black horizontal line indicates the median.

rection. These issues are analysed in the following Section 6.2.3 and Section 7.2.

6.2.3 *Temporal structure*

In order to analyse a potential distortion of the RCM's temporal structure by QM, the RMSE and the autocorrelation of corrected and uncorrected time-series from re-analysis driven RCM simulations were regarded. In Figure 6.4 the RMSE for the four RCMs averaged over the period 1981 to 2000 (calibration period 1961 to 1980) are displayed. The box-and-whisker plots show spatial variability of 60 stations in Austria.

The RMSE of temperature is generally improved by QM, with larger improvement for models with larger RMSE (e. g., C4I-RCA3-ERA40). For precipitation, QM has no clear effect on the RMSE. For relative humidity the RMSE is improving, comparable to temperature. For wind speed, the RMSE is only clearly affected for the model with the largest RMSE and the worst distribution (ICTP-RegCM3-ERA40). Improvements in the RMSE are related to the correction of strong biases and shifts in distributions.

Figure 6.5 shows the autocorrelation of precipitation of the observation (black), the raw (red), and corrected (blue) four RCMs in each panel for lags of up to 6 days. Autocorrelation of the precipitation time-series is predominantly visible at lag-1 (around 0.3) with very small values after that. The RCMs generally feature larger autocorrela-

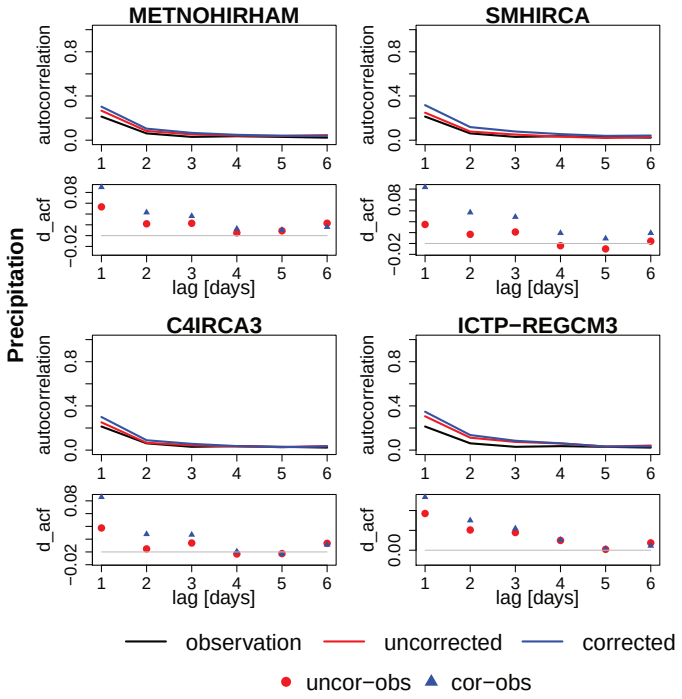


Figure 6.5: Upper panels shows autocorrelation function for 20 year period (validation period 1981 to 2000) of precipitation, respectively, for observations (black), raw RCM-ERA40 (red), and split sample corrected RCM-ERA40 (blue) for four RCMs. The difference in autocorrelations of raw (red) and corrected (blue) RCMs to observations is shown in the lower panels. Here the grey line represents the observation (zero).

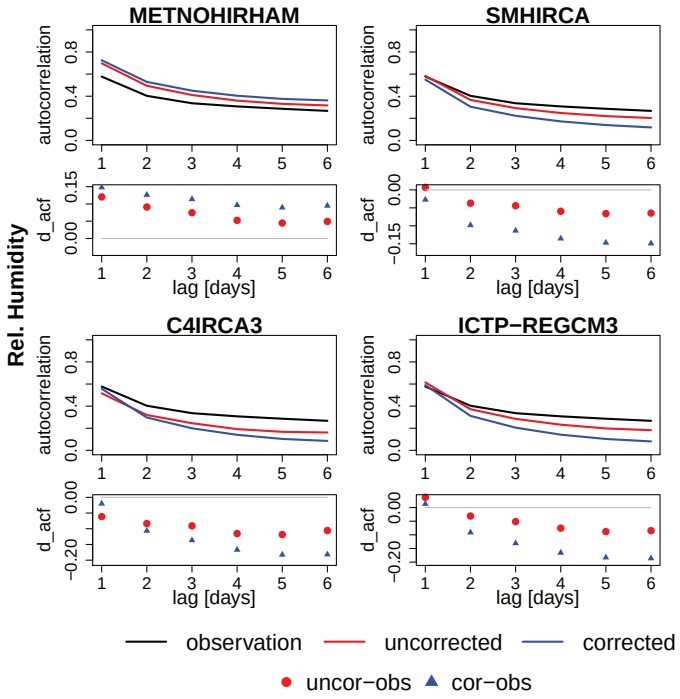


Figure 6.6: Autocorrelation function and delta autocorrelation as in Figure 6.5 but for relative humidity.

tion than the observation, with slightly lower values of the corrected than the raw model.

All four RCMs catch this autocorrelation well, which is not seriously disturbed by error correction. The differences in autocorrelation coefficients (lower panels) between corrected and uncorrected RCM are very small (about 0.01 to 0.15). With regard to relative humidity (Figure 6.6), the autocorrelation is partly over- and partly underestimated by the RCMs, depending on the RCM. For wind speed, like for precipitation, the inter-daily dependency is weak (Figure 6.7). However, the RCMs show a stronger autocorrelation than the observations do, and the corrected RCMs are always closer to the observation than the raw ones.

In summary, improvement or no change could be found in RMSE and autocorrelation due to error correction. It should be emphasised, that this improvement cannot be interpreted as an improvement of the temporal structure of the time-series in a strict sense, but is rather caused by correction of intensity and distribution. An actual improvement of the temporal structure is out of the scope of the presented error correction method. The results mainly demonstrate, that QM conserves the temporal structure of RCM time-series, including their strengths and weaknesses.

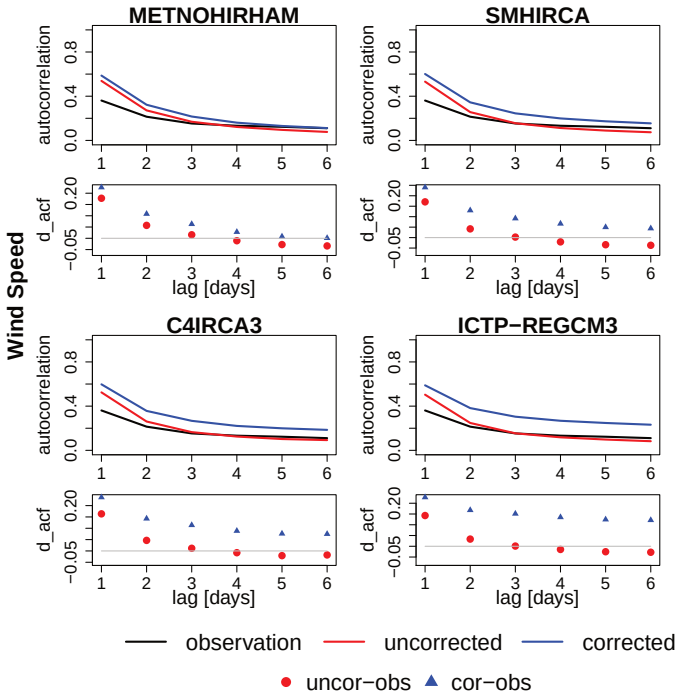


Figure 6.7: Autocorrelation function and delta autocorrelation as in Figure 6.5 but for wind speed.

6.3 SUMMARY OF DIRECT VALIDATION

In the direct validation study the application of QM has been evaluated for a small ensemble of RCMs and six meteorological variables: temperature, precipitation, relative humidity, wind speed, global radiation, and surface air pressure. The evaluation included biases and measures for temporal consistency.

Annual and monthly biases were strongly reduced by QM for all variables in most cases. Exceptions were found, when non-stationarities of the model's error characteristics occured. Those non-stationarities were not restricted to highly variable variables like precipitation and one particularly prominent case was found for temperature. Even in the worst cases of non-stationarity, QM still clearly improved the biases of the raw RCM. Independent calibration and evaluation periods have been used, which are affected by climate variability and change. Thus, these results give some indication for the performance of QM applied to future scenarios. However, the effect of non-stationarity can be expected to be larger in far future, which limits the scope of these results. Maraun (2012) demonstrated that for periods at the end of the 21st century, a simple bias correction method only partially improves the raw RCM results.

The purely empirical implementation of QM successfully corrected variables with very different probability density distributions, which makes it highly flexible and applicable to various meteorological vari-

ables and regions. The drawback is the necessity to interpolate between values of the empirical cumulative distribution function (ECDF) and to extrapolate for new extremes (Chapter 4), which leads to small systematic errors in some cases with low-quality observational data (for this studay wind speed).

Considerable differences were found between the distributions of the uncorrected RCMs and observations in some variables (e. g., relative humidity) and models (e. g., ICTP-RegCM3 for wind speed). QM successfully adjusts also these distributions. Such strong modifications raise the question whether the time-series are still plausible after correction. This question was examined by analysing the autocorrelation and RMSE of raw and corrected hindcast simulations. When applying QM to RCM output, improvement or no clear effect were found in RMSE and autocorrelation.

These results demonstrate that QM strongly reduces biases, adjusts distributions, and retains the quality of the temporal structure of the time-series of RCMs. However, the retainment of the temporal structure is not an improvement as deficiencies of the RCMs in those features are retained as well.

The reality of scientific endeavour is profoundly messy, often illogical, deeply emotional, and driven by the individual personalities involved as they sleepwalk their way to a temporary scientific truth.

— Jim Baggott

7 VALIDATION OF INTER-VARIABLE RELATIONS

7.1 OUTLINE OF VALIDATION CONCEPTS

The topic of physical consistency and inter-variable relations of climate model output, in particular using statistical downscaling, is widely discussed in the climate impact modelling and climate impact community. Climate data are commonly used as input for climate impact models (e. g., snow models, hydrological models, economical models). A reasonable correct output of those models depends on the correct representation of physics in climate data.

Nevertheless, most studies "only" evaluate the downscaling model output uni-variate (e. g., Gilliam and Pleim, 2010; Vautard et al., 2012), not considering the representation of inter-variable relations. Hoffmann and Rath (2012) have done one of few studies analysing the *meteorologically consistency* after applying a bias correction (BC) method. They use *standard quantile mapping* as proposed by Piani et al. (2010a) (fitting cumulative distribution function (CDF)) and introduce a modified *consistent quantile mapping* to correct *hourly* precipitation and radiation for a couple of stations in Germany. The improved consistency is gained by creating two conditioned empirical cumulative distribution functions (ECDFs) instead of one unconditioned ECDF for each variable. Additionally, a weighting factor is used to find the combined optimisation between optimised bias correction and optimised consistency. As consistency measure the inverse bi-variate ECDF is calculated.

A study on RCM output used for air quality simulations (Gilliam et al., 2006) evaluates the inter-variable correlation of direct RCM output (wind speed, wind direction, and temperature) as well as derived variables (dewpoint, surface pressure, precipitation, and cloud fraction). A very similar approach has been done for this PhD thesis and is presented in Section 7.2.

Guenni et al. (1990) investigate the inter-variable correlations of temperature (mean, minimum, maximum) to fractional cloudiness and sunshine duration. They analysed only observations to gain information for a crop model, not simulated variables.

Few other studies (e. g., Lenderink and van Meijgaard, 2008; Berg et al., 2009; Piani and Haerter, 2012) investigate temperature and precipitation relations. The main interest is on the assumption that pre-

cipitation intensities increase with increasing temperature (Clausius-Clapeyron relation). Therefore, the focus of those studies was on the top end percentiles for precipitation (e. g., Berg et al. (2009), 70th, 90th, 99th, 99.9th percentiles of precipitation for 2 K bins). I. e. Berg et al. (2009) investigated a distribution conditioned on a threshold value of another variable, which is similar to my approach, discussed in Section 7.3, where I investigate the distribution of the variable conditioned on a distribution of another variable.

Here, the topic of inter-variable relations is accessed with six meteorological variables (Chapter 5) in two steps. The first approach tests the effect of Quantile Mapping (QM) on the inter-variable correlation given by an RCM. Bi-variate correlations in raw and corrected RCMs were analysed regarding their linearity and rank in Section 7.2.

As most variable pairs are not strongly linear correlated and the rank can hardly be disturbed by QM, a more differentiated evaluation is needed. Therefore, inter-variable relations in RCMs were examined by comparing joint distributions of pairs of RCM output variables. Climate models use different parameterisations, numerical solvers, grid, etc. and therefore behave different for single variables (see, e. g., Figure 6.3 in Section 6.2.2). Therefore, in approaching this analysis method an ensemble of RCMs (driven by re-analysis) was considered.

7.2 INTER-VARIABLE CORRELATION

QM acts on each variable separately, so concern exists, whether inter-variable dependencies are distorted by QM. The focus here is not on the physical consistency of raw RCMs, but on changes of inter-variable relations in RCMs due to QM.

One way to validate the inter-variable correlations is by visual display. Friendly (2002) presented a possibility to combine the magnitude and sign of correlation in a plot together with various forms of display, e. g., a scatterplot or pie chart like in this study. Considering more than two variables one gains a matrix of plots (Figure 7.1), which he called *corrgram*. The individual correlation plots can be re-ordered to group "similar" variables. The re-ordering can be done by applying a principal component analysis. The grouping assists the exploration of similar variables and correlations. In this study the grouping was not included. Correlation matrices of three data sets were compared with each other.

Here, a corrgram of five meteorological variables were created for observations, raw RCM, and bias corrected RCM output. The RCMs are driven by GCMs and have been corrected for the 128-year period, 1971 to 2098. The validation is done for the historical 40-year period, 1971 to 2010, as well as for two 30-year periods in the future: near future 2021 to 2050 and far future 2069 to 2098. As inter-variable correlation can be regionally different, four stations with representative locations

have been chosen, three in Austria (Sonnblick, Hohe Warte Vienna, and Innsbruck) and one in Switzerland (Zermatt). For more details on their locations and exhibitions see Section 6.1.

7.2.1 *Results of validation of inter-variable correlations*

The correlations of five variables are analysed pairwise before and after correction, As for comparison the observed correlation is shown. Figure 7.1 shows 12 corrgrams, which illustrate scatterplots and correlation coefficients of each pair of variables at the four selected stations. The correlations are discussed exemplarily for the ICTP-RegCM3 model in summer. Results for further seasons are shown in Figure B.9 to Figure B.23. Three more RCMs have been investigated and show the same results (not shown).

Since most variables are not linearly correlated, the Spearman rank correlation coefficient as been chosen, which is based on the ranks and not on linear relation like the Pearson correlation coefficient (Wilks, 2006). QM is a transformation that conserves ranks. This, however, is only valid for a specific day of year (DOY) in our implementation of QM, since for each day of the year, single ECDFs are created. The Spearman coefficient regards ranks of the entire time-series under consideration which can be modified by QM indeed.

The coefficients are shown as pie-charts in Figure 7.1 and additionally for Zermatt in Table C.1. The correlation coefficients are calculated for the period 1971 to 2010 (1981 to 2010 for Zermatt). For the historical analysis, the technical approach (c. f. Section 6.2, calibration period equals evaluation period) has been used. In addition, a similar analysis for future scenarios is shown in Figure 7.2 for near and far future for the station of Innsbruck (for other stations see Figure B.9 to Figure B.23).

Comparing the scatterplots and the Spearman coefficients of the raw (Figure 7.1 b) and error-corrected RCM (Figure 7.1 c), no big differences are visible. Table C.1 underlines this for the station of Zermatt. The correlation of temperature with global radiation shows small differences, same as precipitation with wind speed, and precipitation with relative humidity. This also counts for the other seasons. Nevertheless, no systematic degradation of the RCM's correlation by applying QM can be detected. Differences in the scatterplots are related to the mapping of the value range towards the observations. This compresses or stretches the scatterplots without changing the correlation itself.

Considerable differences are visible when comparing the correlations of observation (Figure 7.1 a) and raw RCMs. The observed correlations, e. g., of wind speed and relative humidity for Innsbruck and Sonnblick show opposite signs and different shapes of scatterplots. The correlation coefficient of wind speed and temperature is

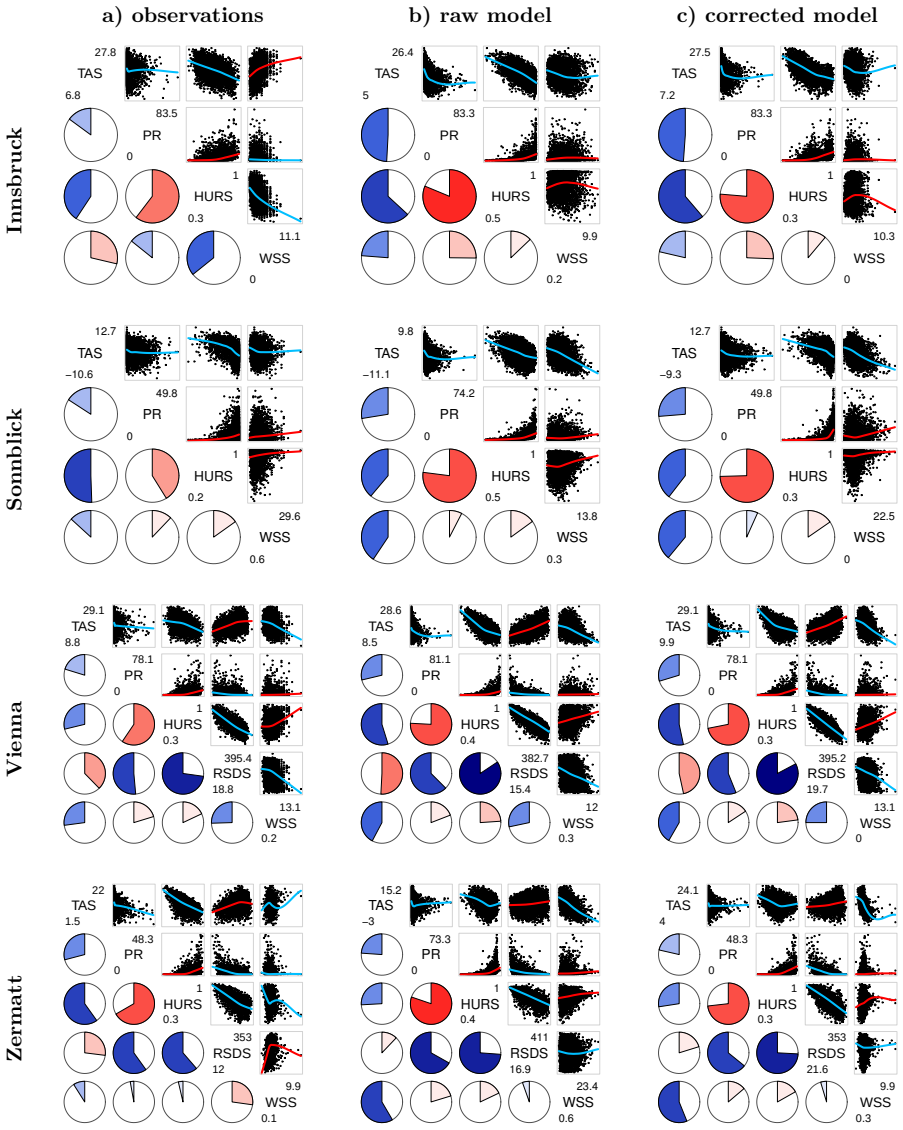


Figure 7.1: Correlation matrices for the period 1971 to 2010 (Zermatt 1981 to 2010) including temperature (tas), precipitation (pr), relative humidity (hurs), global radiation (rsds) (only for Vienna and Zermatt), and wind speed (wss) for selected stations in Austria and Switzerland for a) observed, b) modelled, c) error-corrected modelled data in summer (JJA). Pie charts show Spearman correlation coefficients, indicated with counterclockwise (negative correlation, blue) and clockwise (positive correlation, red) pie slices. Lines in scatterplots are the Loess fit. The values above and below the variable names give the range of the data. The model shown here is the ICTP-RegCM3.

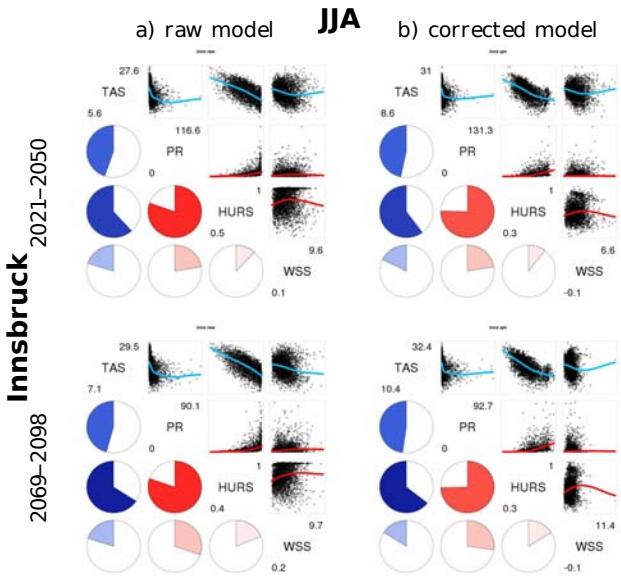


Figure 7.2: Correlation matrices for Innsbruck for summer (JJA) of the periods (top) 2021 to 2050 and (bottom) 2069 to 2098 for (a) raw and (b) corrected RCM. Same layout as for Figure 7.1.

much higher in the RCM than in the observation. Those differences can be caused by e. g., model parameterisation, observational errors, or local effects which RCMs cannot capture due to scale discrepancies. QM does not correct such effects, it rather conserves the inter-variable correlations of the RCMs.

For future periods (Figure 7.2), as for past periods, the correlations given by the RCM are not systematically changed by QM. This is a pleasant result, given the RCM provides a correct representation of the physical processes. If this is not the case a bias correction like QM does not correct those deficiencies.

7.3 INTER-VARIABLE RELATIONS IN RCM OUTPUT AS CONDITIONAL DENSITIES

As mentioned in the outline (Section 7.1), this study approaches the inter-variable relations of five RCMs (c. f. Section 5.1) by evaluating conditional densities of pairs of meteorological variables. The variables are air temperature, precipitation amount (only wet days, defined as precipitation $> 0.1\,\mathrm{mm/d}$), relative humidity, global radiation, wind speed, and surface air pressure. For evaluation, again, the point-scale observations of ZAMG were used, but aggregated to RCM grid cells (Section 5.3 for details on the aggregation).

A 20-year period in the past has been chosen for evaluation, 1981 to 2000. As all variables show different characteristics in distribution depending on the season or even month considered, daily time-series were used and validated on monthly basis. For a 20-years period i. e.

one time-series connects, e. g., 20 Januaries, another time-series contains 20 Aprils.

7.3.1 *Method — conditional density*

Conditional density is the probabiliy density function (PDF) of one variable (say precipitation) at a specific value of another variable (say a temperature of 0 °C). There is a wide range of interesting temperature values or other variables to condition on.

To avoid the quandary of choosing the "right/interesting" threshold (and maybe forgetting an even more interesting one) and to reduce the amount of graphs produced, conditional density estimates were calculated. This results in a set of PDFs, one for each value of the conditioning variable; and these can be displayed as a perspective plot such as shown in Figure 7.3.

Picturing the conditional behaviour of variables can be done by simply drawing a histogram of frequencies of variable x at a certain value of variable y. To get a clearer view on the underlying structure of this distribution, one can use density estimates instead of histograms.

Density estimates of single variables can be calculated by applying a smooth kernel function (Bowman and Azzalini, 1997)

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^{n} w(x - x_i; h), \tag{7.1}$$

with $\hat{f}(x)$ being the density estimate, w is a smooth *kernel function*, which is symmetric about zero and which satisfies $\int_{-\infty}^{\infty} w(x; h)dx = 1$. This restriction ensures that the density estimate itself integrates to 1 and hence is a valid PDF. The restriction also amounts to a requirement that the kernel function is itself a PDF: the quantity h is a measure of the spread of this kernel function and, analogous to the choice of bin width in a histogram, controls the smoothness of the resulting density estimate. The degree of smoothing is defined by h and needs to be set individually (see Section 7.3.1.1).

Given two variables X and Y, with PDFs $f_X(x)$ and $f_Y(y)$, respectively, and joint density $f_{X,Y}(x, y)$ (Equation 7.2),

$$\hat{f}_{X,Y}(x, y) = \frac{1}{n} \sum_{i=1}^{n} w(x - x_i; h_x) w(y - y_i; h_y) \tag{7.2}$$

the conditional density of X when $Y = y$ is defined as

$$f(X|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)} \quad \left(\text{or} \quad f(Y|x) = \frac{f_{X,Y}(x, y)}{f_X(x)} \right), \tag{7.3}$$

where f is the density function, X the variable of interest, and Y the variable to be conditioned on (e. g., Loader, 1999; Pagan and Ullah, 1999; Li and Racine, 2007).

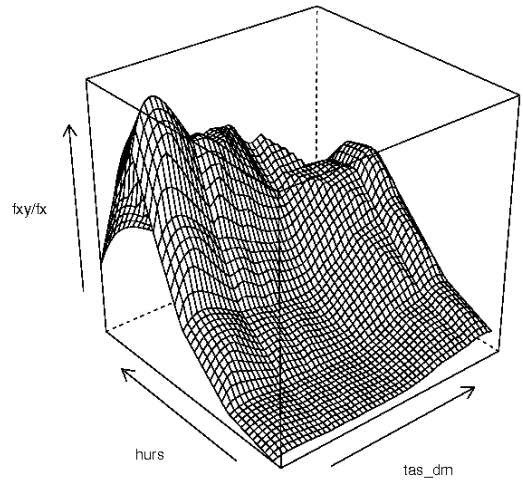


Figure 7.3: Conditional density estimate ($f(Y|x) = f_{X,Y}(x,y)/f_X(x)$) for relative humidity (Y) conditioned on temperature (X), created with *condens1* function of the *aratio* package in R.

The densities of, e. g., relative humidity conditioned on temperature results in a 3-dimensional graph with relative humidity on the x-axis, temperature on the y-axis and the conditional density as z-values (e. g., like in Figure 7.3).

To calculate the density functions/estimates of $f_{X,Y}(x,y)$ and $f_Y(y)$ (see Equation 7.1 and Equation 7.2) the R-package "sm" (Bowman and Azzalini, 1997) was used, which includes a function "sm.density", to calculate density estimates. For the purpose of calculating a conditional density as in Equation 7.3, it is important that the uni-variate density $f_Y(y)$ is compatible with the joint density $f_{X,Y}(x,y)$ in the sense that $f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\mathrm{d}x$. To ensure this in our density estimation, we use sm.density() to estimate the bi-variate density $f_{X,Y}(x,y)$, and then derive the marginal density estimates by integration as appropriate.

7.3.1.1 *Implementation*

For meaningful comparison of the conditional densities obtained from observations and from the RCM, it is important to ensure that they are estimated in exactly the same way so that any discrepancy between them cannot be attributed to differences in the estimation procedure. In particular, the choice of smoothing parameters h_x and h_y must be the same for the RCM as for the observations. By default, the sm.density() routine uses the data to determine an appropriate choice of smoothing parameters; but to ensure that the choices in this work

are appropriate, for each pair of variables the means of the smoothing parameters for the observations and for the RCM have been used:

$$h_x = \frac{h_{x,\text{obs}} + h_{x,\text{mod}}}{2}, \quad h_y = \frac{h_{y,\text{obs}} + h_{y,\text{mod}}}{2} \tag{7.4}$$

For any pair of variables, meaningful comparison between the conditional density estimates for the RCM and observations also requires that the conditional densities are estimated on the same grid of values. To ensure this, for each pair of variables the conditional densities are evaluated over a regular grid ranging from the smallest to the largest value of each variable in the combined sample of RCM and observed values.

In the case of strongly skewed distributions, like precipitation, single outliers can occur with a distance bigger than the smoothing parameter to the bulk of values. This occurrence leads to discontinuities in the density estimate. To avoid such, one could use a variable smoothing parameter (higher value for outlier and smaller value for values close together), which would make a comparison between two data-sets impossible (or at least senseless). Another solution is to work with transformed precipitation data to reduce the skewness. This is simpler and more appropriate for this application. Thus, the logarithm of precipitation was used.

RCM data is evaluated against observation, i. e. conditional densities have been calculated for modelled and observed data. As a visual evaluation (compare two graphs with naked eye) is not as accurate as a single value or a single graph, the ratio φ of the simulated and observed conditional density estimates were calculated:

$$\varphi = \frac{\hat{f}_{\text{mod}}(x|y)}{\hat{f}_{\text{obs}}(x|y)} \tag{7.5}$$

A potential disadvantage of this expression is that it can range over several orders of magnitude due to sampling variation, especially where $\hat{f}_{\text{obs}}(x|y)$ is close to zero. This variation can make visualisation (as in Figure 7.3) difficult. To overcome this problem, Equation 7.5 has been logarithmised. If the observed and modelled conditional densities are equal, that means $\log \varphi = 0$.

7.3.1.2 *Uncertainty*

To answer the questions "What does a ratio of 0.2 or of 4 mean?" "Is 0.2 still *O.K.*?", to support the evaluation of the ratio plots and give a hint on how variable the results are, the standard deviation σ has been calculated (Equation 7.6):

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N} (\log(\varphi_i) - \log(\overline{\varphi}))^2}{N}} \tag{7.6}$$

φ is the ratio (Equation 7.5) calculated before, and N are the samples gained by bootstrapping[1]. σ is smallest in regions where the log-ratio can be estimated relatively precisely (e. g., in regions with large numbers of data points) and is larger elsewhere.

To interpret the standard deviations, it is helpful to bear in mind that if the sampling distribution of the estimated log-ratio (i. e. the logarithm of Equation 7.5) is approximately normal then the estimate should be within two standard deviations of the true log-ratio around 95 % of the time.

The standard deviation is applied under the assumption of normal distribution and homoscedasticity. Thus, the ratio has been logarithmised, which tones down the violation against those assumptions.

7.3.2 *Results & discussion of conditional densities in raw RCMs*

To keep this study focused and to avoid huge amounts of plots (without gaining more information) results for selected months, grid cells, and variables are presented. The months have been cut to four representatives for the four seasons: January, April, July, and October. Region-wise three very different grid cells have been selected. For high mountains grid cell 3 (Hohe Tauern Nationalpark) with station heights between 1041 m and 2315 m is chosen, for flat terrain grid cell 7 (Vienna) with station heights between 153 m and 222 m, and for a region with mountains and valleys grid cell 6 (Pongau) with station heights between 550 m and 1503 m (c. f. Figure 5.4) were selected.

The following discussion focuses on special combinations of variables. Hereby the physical relation of variables is considered. That gives six pairs of variables:

(a) Wind speed conditioned by pressure as a large-scale process in winter over flat terrain. An RCM should be able to catch this relation quite easily.

(b) While the same pair of variables but evaluated in autumn over more difficult terrain is more difficult for the RCM. Wind speed is related to convective processes as well, which are not resolved by the model.

(c) Temperature conditioned on global radiation over flat terrain in summer, which has a strong relation.

(d) Temperature conditioned on precipitation.

1 Simple bootstrapping means random sampling with replacement of homogeneous samples. The time-series of RCM and observation have been pair-wise sampled (quadruple, because two variables and two time-series) 1000 times, as is recommended for density estimates (Davison and Hinkley, 1997) (remember here that the time-series are for single months). The pair-wise sampling preserves the connection of RCM and observations to be evaluated.

(e) Relative humidity conditioned on temperature has a strong physical relation as well.

(f) Precipitation conditioned on temperature is considered more difficult for the RCM as the correlation is not strong, furthermore, precipitation amount is difficult to simulate (convection resolving simultions in Chapter 2.

The log-ratio of the conditional density estimates and its standard deviation for a variable with all combinations of conditioning is combined in one graph (e. g., Figure 7.4) by showing the log-ratio as heat map overlaid by the standard deviation as contour plot. The colour intensity shows the discrepancy of log-ratio from zero. The white colour is set to ± 0.25.

Figure 6.3 in Section 6.2.2 shows uni-variate density distributions. To have those in mind helps evaluating the conditional densities presented below. Also, remember from above that the ratio φ (Equation 7.5) is calculated from $\varphi = \hat{f}_{\text{mod}}/\hat{f}_{\text{obs}}$, i. e. a small φ indicates a bigger value for the observed conditional density than for the simulated one, which then results in a negative $\log \varphi$. That means:

- $\hat{f}_{\text{mod}} > \hat{f}_{\text{obs}} \rightarrow \log \varphi > 0$
- $\hat{f}_{\text{mod}} < \hat{f}_{\text{obs}} \rightarrow \log \varphi < 0$

A) WIND SPEED | SURFACE PRESSURE (FLAT, WINTER) Wind speed and surface pressure are large-scale climate variables which are strongly related. Figure 7.4 shows a good match between simulated and observed conditional density for low to medium wind speed intensities for all RCMs but ICTP.

High wind speeds are underestimated by the models (negative $\log \varphi$). This structure is visible for all pressure situations which indicates an overall bias towards lower wind speeds, independent of surface pressure. Even though the uncertainty range for higher wind speeds is wider (larger standard deviation), it excludes a good match for RCM and observation. As an example in Figure 7.4 for Sveriges Meteorologiska och Hydrologiska Institut (SMHI), this means the calculated log-ratio has a value of 8 and the standard deviation is about 2.8. Here the uncertainty would only allow a value of 5.2 ($8 - 2.8$) which is still far off zero, the perfect match. Higher standard deviations for low pressure values show a more variable density distribution of wind speed.

The ICTP model shows a bad performance for $\log \varphi(\text{wss}|\text{ps})$ which was expected when comparing with Section 6.2.2 d, where the simulated distribution of wind speed is very different from the observed one. Even though not named there, the ICTP model is the one model with a wrong representation of the density distribution for wind speed. $\log \varphi(\text{wss}|\text{ps})$ for the ICTP model is underestimated for low

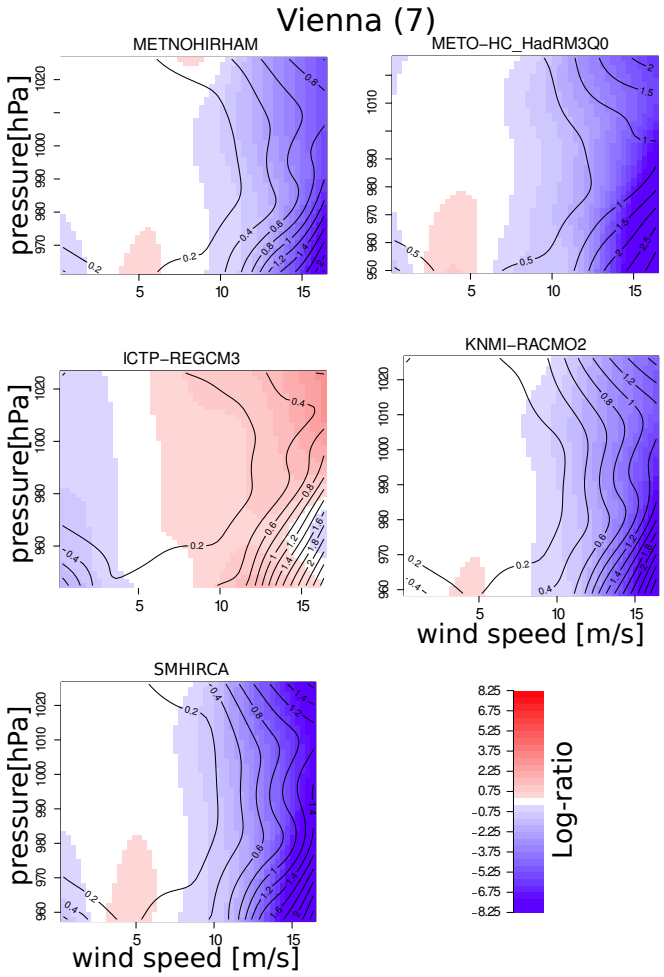


Figure 7.4: Log-ratio (colour bar) of wind speed conditioned on surface pressure for the region Vienna (7) in January 1981 to 2000, for five RCMs. Overlaid by the standard deviation as contour plot.

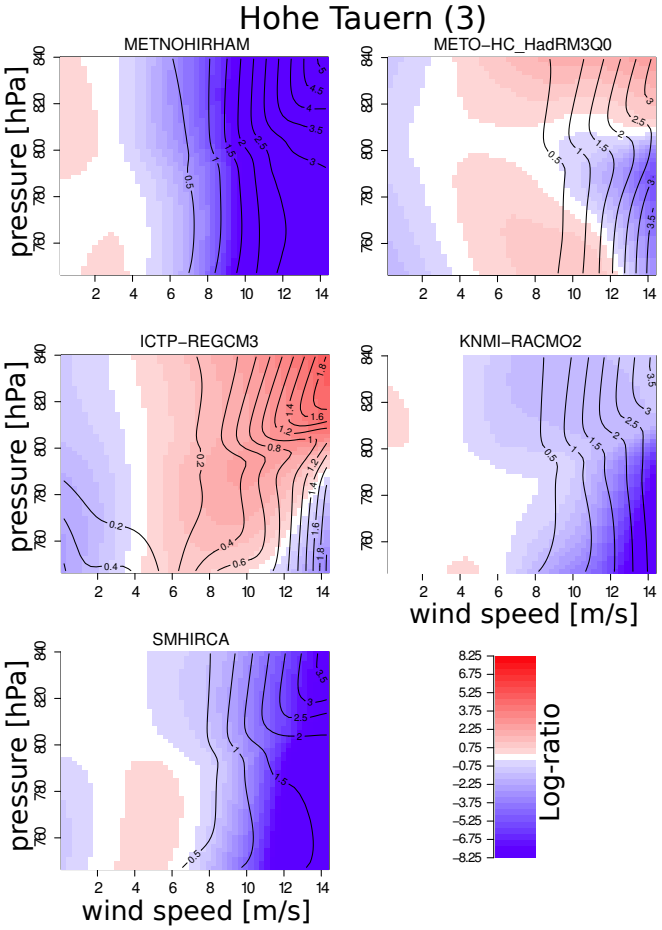


Figure 7.5: Log-ratio of wind speed conditioned on surface pressure for the region Hohe Tauern (3) in October 1981 to 2000.

wind speed and overestimated for medium to high wind speed. The white region, here, indicating $\hat{f}_{\text{mod}} = \hat{f}_{\text{obs}}$ is the transition zone from under- to overestimation, at $\approx 5\,\text{m/s}$ for all surface pressure values.

B) WIND SPEED | SURFACE PRESSURE (MOUNTAINOUS, AUTUMN) In autumn one experiences low pressure systems with strong winds in middle Europe. Over the mountains the intensities are even higher. As pressure systems are large-scale phenomenons, the strong winds should be well presented by the RCMs. Figure 7.5 shows such relationship for the Hohe Tauern region in October. A similar picture as in Figure 7.4 is presented for SMHI-RCA, KNMI-RACMO2, and ICTP-RegCM3, but with higher standard errors at high wind speeds and higher values for $\log \varphi(\text{wss}|\text{ps})$ already for medium wind speeds. In autumn over high mountains METO-HC-HadRM3Q0 shows a different $\log \varphi(\text{wss}|\text{ps})$ structure than in winter over flat terrain. Low wind speeds at medium pressure are well represented, also considering the low uncertainty. Like the other RCMs, METO-HC-HadRM3Q0 has a high uncertainty for high wind speeds, which makes a clear

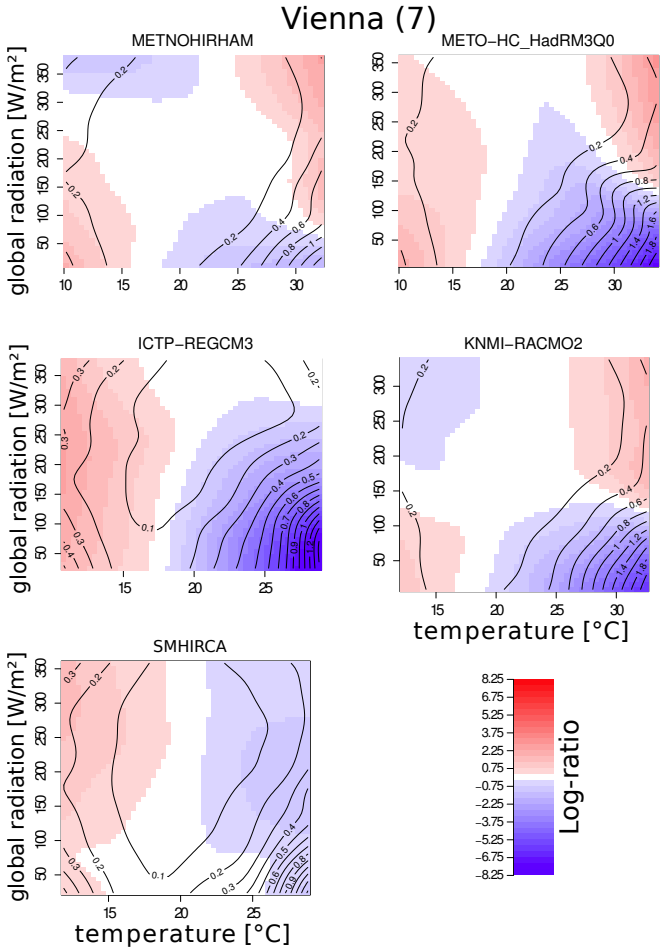


Figure 7.6: Log-ratio of temperature conditioned on global radiation for the region Vienna (7) in July 1981 to 2000.

interpretation difficult. For extremes like with METNO-HIRHAM where $\log \varphi(\mathrm{wss}|\mathrm{ps})$ is outside of the standard deviation, one can say that signal from the RCM is a clear and strong underestimation of the inter-variable relation of wind and pressure.

c) AIR TEMPERATURE | GLOBAL RADIATION Air temperature depends on global radiation, which is visible in the seasonal and diurnal cycle of temperature. Figure 7.6 shows a slightly biased density distribution of the RCMs for air temperature for Vienna region in July. The RCMs overestimate low temperatures and underestimate high temperatures for low global radiation, in particular ICTP-RegCM3. For very low radiation values the agreement between SMHI-RCA and observation is good, which is indicated by the white area and low standard deviation. The other RCMs show an overestimation for high radiation values and extreme temperatures. Thus, ignoring extreme temperatures, $\hat{f}_{\mathrm{mod}}$ represents $\hat{f}_{\mathrm{obs}}$ quite well for all shown RCMs.

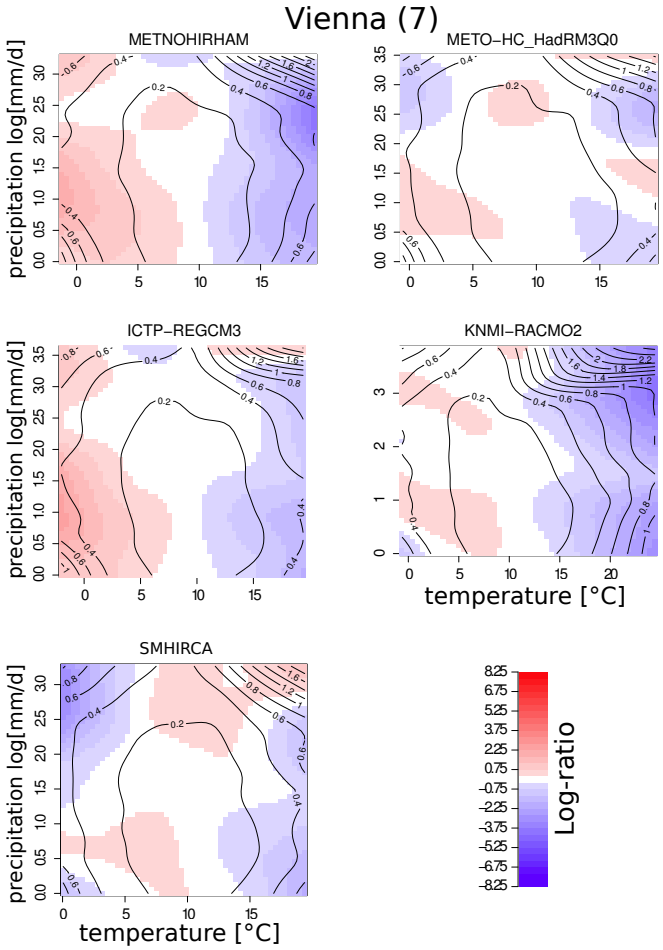


Figure 7.7: Log-ratio of temperature conditioned on precipitation amount for the region Vienna (7) in April 1981 to 2000.

D) AIR TEMPERATURE | PRECIPITATION Temperature is affected by precipitation amount, e. g., during small-scale phenomena like convective events or large-scale systems like cold fronts. During convective events single stations can experience a sudden cooling introduced by a thunderstorm, where cool upper level air is pushed downwards with the rain and hail. This happens on sub-daily scale, but results in a lower daily mean temperature. As RCMs usually do not simulate on convection resolving scale (which is smaller than $0.22° \times 0.22°$), one would expect a mismatch for those medium precipitation amounts (as daily sums are investigated, not hourly). Precipitation introduced cooling due to cold fronts is related to large-scale meteorology, i. e. low pressure systems. In fact it works inversely, as the cool air propagates and pushes under the warm air, lifts it, which results in precipitation. But such meteorological differences are not distinguishable in this statistical analysis. However, large-scale processes should be well presented by the RCMs.

Figure 7.7 shows the log-ratio for temperature and precipitation amount for the Vienna region in April. Even though the best part

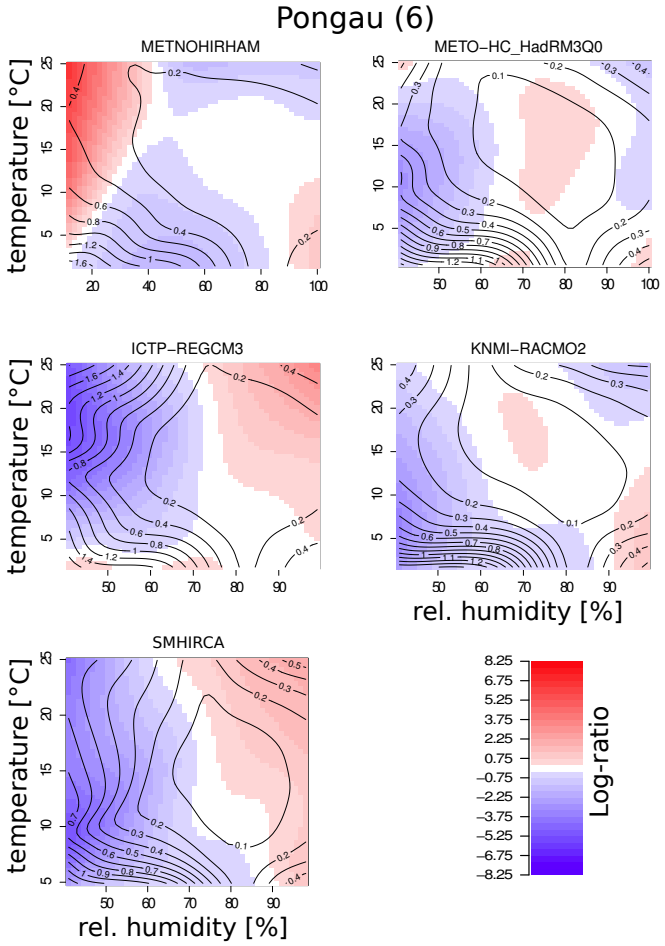


Figure 7.8: Log-ratio of relative humidity conditioned on near surface temperature for the region Pongau (6) in July 1981 to 2000.

of the temperature distribution seems to be well simulated (white regions in the middle) by all RCMs, the variability is quite high for ratio values that are small on the edges of these graphs. E. g., for extreme high temperatures and extreme precipitation amounts the uncertainty range is larger than ± 3.6, so the conditional density of the RCMs can be higher than observed (as shown here), but also smaller.

E) RELATIVE HUMIDITY | AIR TEMPERATURE Air temperature affects relative humidity directly by expanding or shrinking of the volume of an air parcel by heating or cooling, which affects the amount of water that can be contained. On small scales, like shading of mountains, this is observed in diurnal cycle, which does not count in this study as daily mean values are evaluated. Beside local small-scale effects, the same relation between temperature and relative humidity is observed on large-scales with, e. g., frontal processes.

In Figure 7.8 the ratio for the Pongau region (medium high mountains and valleys) in July is shown. The ratio shows a strong linear signal for SMHI-RCA and ICTP-RegCM3. The RCMs underestimate the

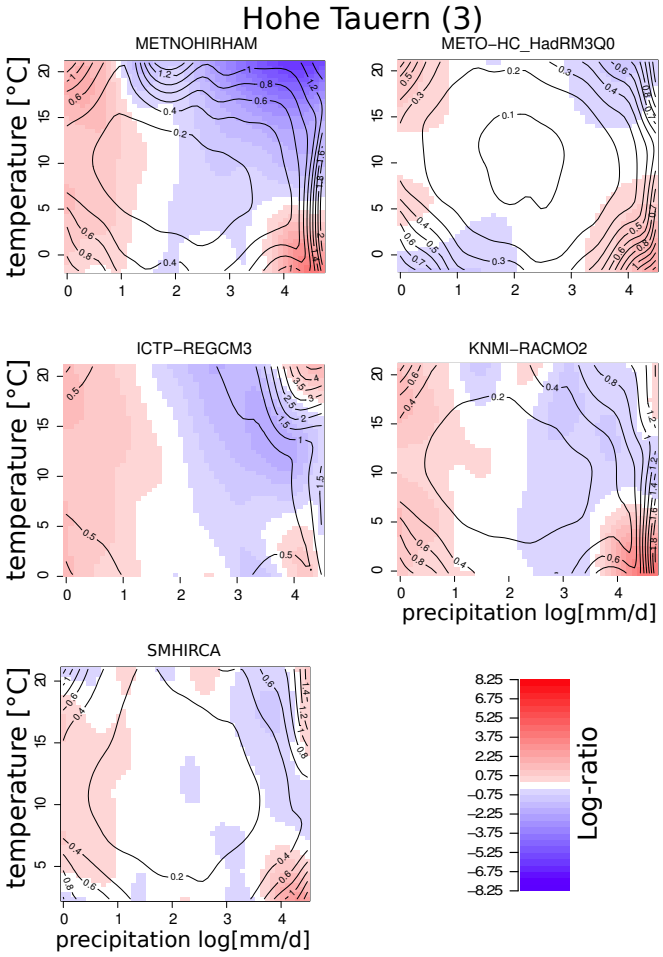


Figure 7.9: Log-ratio of precipitation amount conditioned on near surface temperature for the region Hohe Tauern (3) in July 1981 to 2000.

conditional density for low relative humidities and overestimate it for high humidity values, in particular for high temperatures. As this structure is nearly parallel, a clear statement about the conditional densities of RCMs and observation is difficult. The parallel structure indicates a bias in the RCMs for relative humidity (Figure 6.3 c as well as Figure 6.1 c and Figure B.3 c, Figure 6.2 c and Figure B.6 c) which is stronger than the differences in conditional densities of RCM to the observations. The strong bias in relative humidity is visible for all regions. However, there is apparently some temperature dependence to this bias, as the white stripe has slightly off-vertical alignment.

A similar structure is present for the KNMI-RACMO2. The off-vertical alignment of the underestimation of conditional densities can be found here as well. On the other hand much more confidence can be given into the representation of medium and high humidity values for medium and high temperature values for KNMI-RACMO2, METO-HC-HadRM3Q0, and METNO-HIRHAM.

F) PRECIPITATION | AIR TEMPERATURE In summer most precipitation in mountainous regions is induced by convection. Therefore, the relation of daily precipitation amount and daily mean air temperature on RCM-scale is not very strong. The conditional density of the RCMs seems to match well the observed density for most parts of the distributions for the METO-HC and SMHI model (c. f. Figure 7.9). On the other hand the uncertainty range for extremes of both variables is high. The parellel structure visible in the results for METNO, ICTP, and KNMI model indicate a strong bias in the precipitation field which hides the influence of temperature on the distribution of precipitation.

7.4 SUMMARY OF INTER-VARIABLE RELATIONS IN RCMS

Many climate impact models need multiple climate variables as driving data. Thus, the physical relationship between those variables need to be correct, i. e. as observed in reality, to get physical reasonable model output. The bias correction of single variables (which is the general topic of this work) raised the question whether inter-variable relationships are still plausible after correction. Therefore, inter-variable correlations have been validated for four example stations and four RCMs for historical and future periods. When applying QM to RCM output, no clear effect on the correlation between meteorological variables could be found. The main result of this part is that QM retains the quality and also the deficiencies of the inter-variable dependencies of RCMs.

In a comparison of correlations gained by raw RCM output and by observations, considerable differences were found. RCMs are based on physics but use many parameterisations to descibe unresolved and complex processes. Those parameterisations—together with model errors, numerics, discretisation, scale discrepancies—not only introduce errors (e. g., Feser, 2006; Prein et al., 2013b) but also lead to incorrect inter-variable relations.

To estimate the inter-variable relations conditional densities of pairs of variables have been calculated for five example RCMs driven by ERA40. The densities of an RCM have then been compared to the conditional densities of observed variables.

The main results in the second part of the study (Section 7.3.2), showed a mismatch of the conditional densities of RCMs and observations. The mismatch varies in intensity for regions, months, and variable pairs. Variables with a high bias in the simulation (e. g., relative humidity) were more difficult to interpret, as the inter-variable signal in the ratio of RCM and observation was weaker than the bias of single variables (like relative humidity). On the other hand, if the physics in the RCM was correct then any bias in one variable would translate into a corresponding bias in the other: but in this case the

conditional densities would still be the same in the RCM and in the observations. Thus, more investigation is need to be carried out.

I also intend to applying the analysis of conditional densities on bias corrected (quantile mapped) data, in particular for variables with strong biases one should gain a better picture of inter-variable relations. As QM changes the distributions of each variable, a different output can be expected for the conditional densities. In case of relative humidity the correction of the marginal distributions might change the conditional distribution towards the observed, which might count as improving the inter-variable relation in the RCM. More thoughts and time would be needed for that extended approach.

No amount of experimentation can ever prove me right; a single experiment can prove me wrong.

— Albert Einstein

8 INDIRECT VALIDATION OF QUANTILE MAPPING

8.1 OUTLINE OF VALIDATION CONCEPT

Evaluating impacts of future climate on regional scale is an ongoing task for climate and impact researchers (Varis et al., 2004). Projects like CC-Snow cover that topic related to future snow development regarding ski tourism in the Austrian Alps. Therefore, high-quality climate data are essential for a reasonable interpretation of impact model results (Christensen et al., 2008).

Impact models run on different resolutions and require preferably driving climate data on their corresponding scale. In particular modelling on local scales (e. g., river catchments, mountain slopes) is done on high resolutions down to a couple of meters (e. g., Xu, 1999; Strasser, 2008). Thus, to bridge the gap in resolution, statistical downscaling and error correction is applied (e. g., Maraun et al., 2010) in this validation study. In Chapter 6 the good skill of Quantile Mapping (QM) is shown for correcting biases of RCMs for different meteorological variables, relevant for impact studies.

Impact researchers are interested, not only in statistics of RCM output, but also in the effects of downscaling and error correction on their impact model output (e. g., Graham et al., 2007). Most studies use simple bias correction methods on GCM data (e. g., Sharma et al., 2007; O'Connor et al., 2009; Ye et al., 2012) and focus on evaluating the bias and root mean squared error (RMSE) (Dirmeyer et al., 2011; Hagemann et al., 2011). Compared to climate focused studies, however, they consider variables like radiation (Berg et al., 2003; Haddeland et al., 2012), humidity (O'Connor et al., 2009), wind, and soil moisture (Dirmeyer et al., 2011) beside the standard variables temperature and precipitation.

The indirect validation is a concept to investigate the effects of corrected RCM output on impact model output. Here, this is done in two ways: one simple approach is an index, the second is a snow model. The index is calculated from at two meteorological variables and count as simple model (see Section 5.4). The snow model is a physically based model (AMUNDSEN), developed by Strasser (2008), which uses five meteorological variables (see Section 5.2)

The snow model is driven with different sets of RCM output:

1. all variables uncorrected
2. all variables corrected, and
3. all variables corrected, but one single variable

Using a set of corrected climate variables but one variable uncorrected means using a physical inconsistent data-set. In Chapter 7 it was shown that QM does not destroy the inter-variable relations of variables corrected independently. But not correcting a single variable does. Error correction methods rely on reference data (observations) for calibration (Chapter 4). This is a practical problem which must often be faced by impact modellers. In some applied cases it is not possible to correct single climate variables (which are important for the impact model) due to missing or short observation periods. Using uncorrected climate data would result in a biased impact.

Even though the observation network in European countries is quite dense, this does not account for, e. g.,the southern hemisphere. Considering variables like radiation, the observation network is sparse even for Europe. Wind speed is covered well, but the length of high quality measurements is short (< 10 years). The uncertainty in the observations transfers directly to the error correction and thus to the corrected climate data (Chapter 4).

One intention of this study is to investigate the effect of inconsistent climate data on impact models, using AMUNDSEN exemplarily. Beside that, the effect of error correction on snow model output at different locations in Austria was investigated.

The study region is Austria with the very complex terrain of the Alps and flatter regions in the East. Figure 8.1 shows the station considered here. In black (and red) one finds 12 stations with long snow measurements, considered for validating with the snow model. As the indices are related to heat and drought, five of the warmest stations in Austria were chosen, indicated in blue (and red) (also Table C.3).

Parts of the study described in this chapter were carried out together with partners from the CC-Snow project. (Florian Hanzer[1], Dr. Thomas Marke[1], and Prof. Dr. Ulrich Strasser[1]), who developed and ran the snow model AMUNDSEN.

The impact of bias corrected RCM output on derived variables (snow water equivalent (SWE), snow depth (SD), and snow fall amount (SF)) is evaluated on seven components which are shortly introduced in the following. The snow-day frequency (wet-day frequency for solid precipitation only) in winter shows the difference to the fully corrected climate data (Section 8.2.1). The effect on the snow model variable SWE is investigated in Section 8.2.2. Both validations were carried out for METNO-HIRHAM (driven by HadCM3Q0) and C4I-RCA3 (driven by HadCM3Q16).

1 Institute of Geography, University of Innsbruck, Austria

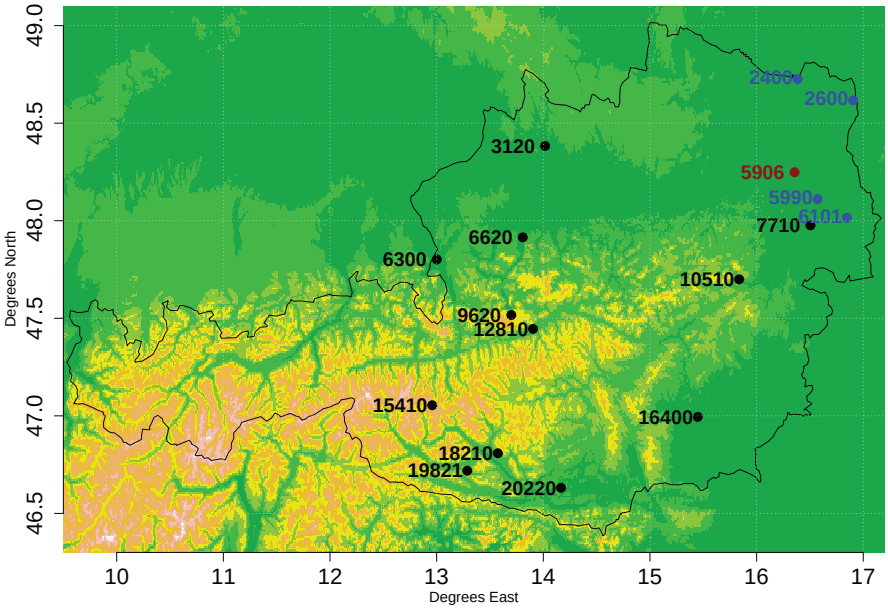


Figure 8.1: Observation station locations of the ZAMG used for indirect validation, for the snow model (black & red) and for the THI (blue and red). Numbers indicate the station ID. Station coordinates and heights can be found in Table C.3.

In Section 8.2.3 the snow model efficiency is described by evaluating the Nash-Sutcliffe-Efficiency (NSE) (Section A.3) for SD related to different reference data. A 2×2 contingency table describes the three dimensions of a yes-no forecast of SF in Section 8.2.4. The persistence is evaluated by calculating the autocorrelation function (ACF) of SF and the maximum number of days with snow cover (Section 8.2.5). These measures are validated for METNO-HIRHAM and C4I-RCA3, both driven by re-analysis (ERA40) and winter seasons only. The winter season is defined as November to March.

All evaluation on SF is done only for wet days. To keep track of the uncorrected and corrected climate variables used as input for the snow model in each validation step, some definitions are given here:

- "x_{v_uncor}", where x is a snow variable, e. g., SWE and v is an element of the group of uncorrected climate variables: air temperature, precipitation, relative humidity, global radiation, wind speed, and all. E. g., SWE_{pr_uncor} is SWE simulated by using corrected variables but precipitation was not corrected.

- "$x_{all\_v\_cor}$", which means a simulated snow variable where only corrected climate variables have been used

Considering the summer season the THI, which is a precipitation un-related measure, was calculated for four RCMs (c. f. Section 5.1) re-analysis driven on six observation stations indicated blue in Figure 8.1. In Austria a heat index is only relevant for the hottest summer months. Thus, the THI was estimated for July and August in a

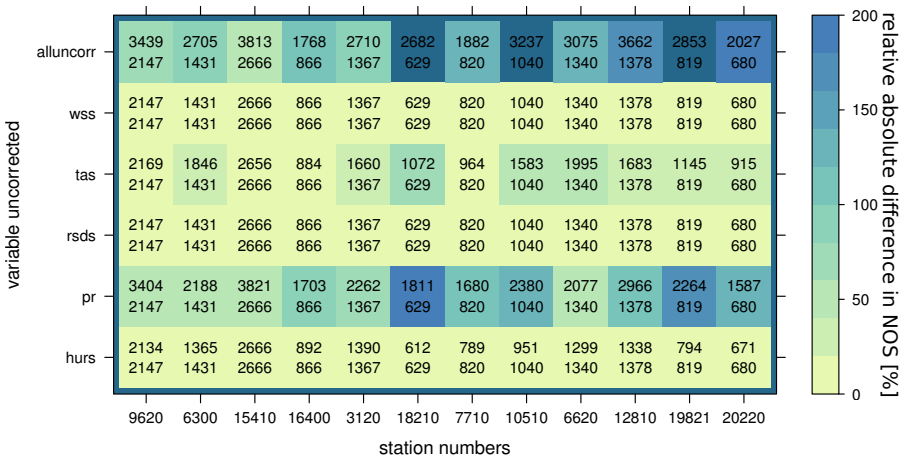


Figure 8.2: Relative absolute difference in [%] (colourkey) in number of snow days (NOS)$_{v\_uncor}$ overlaid by absolute values of NOS$_{v_uncor}$ (upper value) and absolute values of NOS$_{all_v_cor}$ (lower value), used METNO-HIRHAM (driven by HadCM3Q0), accumulated for November to March in 1970 to 2000.

20-year period, 1981 to 2000. Here, bias corrected temperature and humidity was used as well as uncorrected temperature and humidity. In Section 8.2.6 the bias and frequency distribution of THI is validated.

8.2 RESULTS & DISCUSSION OF INDIRECT VALIDATION

8.2.1 *Effect on number of snow days*

The difference in number of snow days (NOS) (days with solid precipitation) has been investigated for the winter months, November to March, in 30 years (1970 to 2000) for two GCM-driven RCMs (METNO-HIRHAM-HadCM3Q0 and C4I-RCA3-HadCM3Q16). The NOS relates to the representation of temperature and humidity. In Figure 8.2 the relative absolute difference of the NOS is shown with reference to NOS$_{all_v_cor}$ (colourkey) for METNO. The stations are indicated on the x-axis and y-axis shows each v of NOS$_{v_uncor}$. Each cell shows the absolute NOS for the respective v_uncor (upper value) and of NOS$_{all_v_cor}$ (lower value), respectively.

As could be expected, wind speed and radiation do not effect the solid phase and amount of precipitation. Thus, using those variables uncorrected within an otherwise corrected data-set works well for this snow model. Uncorrected precipitation amount yields the largest difference in NOS. While a bias in temperature (in particular a warm bias) as well as—to a smaller extent—in humidity influences the phase of precipitation. Using the C4I model produces a very similar picture, with an even more pronounced signal for uncorrected precipitation (c. f. Figure B.25).

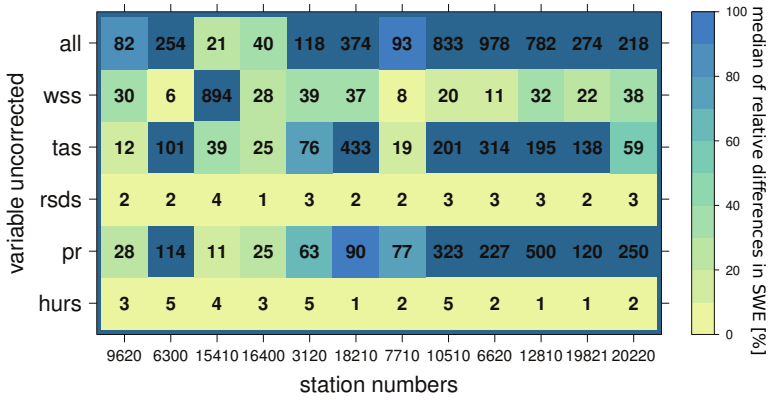


Figure 8.3: Median of relative differences [%] in $\text{SWE}_{\text{v\_uncor}}$ compared to $\text{SWE}_{\text{all\_v\_cor}}$ as values and colourkey RCM: METNO-HIRHAM (driven by HadCM3Q0), period: November to March, 1970 to 2000.

8.2.2 *Effect on snow water equivalent*

Furthermore, the relative and absolute differences in SWE was calculated. It is the difference of $\text{SWE}_{\text{v\_uncor}}$ to $\text{SWE}_{\text{all\_v\_cor}}$, which is described in Equation 8.1.

$$\Delta\text{SWE}_{\text{abs,v}} = |\text{SWE}_{\text{v\_uncor}} - \text{SWE}_{\text{all\_v\_cor}}|, \tag{8.1a}$$

$$\Delta\text{SWE}_{\text{rel,v}} = \frac{|\text{SWE}_{\text{v\_uncor}} - \text{SWE}_{\text{all\_v\_cor}}|}{\text{SWE}_{\text{all\_v\_cor}}}, \tag{8.1b}$$

where v_uncor means that all climate variables used in the snow model were corrected but the v^{th}. $\Delta\text{SWE}_{\text{abs,v}}$ and $\Delta\text{SWE}_{\text{rel,v}}$ were calculated for each station, respectively, for the winter period and are shown as levelplots in Figure 8.3 (relative difference) and Figure 8.4 (absolute difference) for the METNO model. Individual winter months have been evaluated as well (not shown) They show the same characteristics as the plot for the whole season, despite the seasonal cycle of snow.

Regarding the correction or not correction of precipitation and temperature Figure 8.3 shows a similar signal as for NOS in Figure 8.2. Not correcting those two variables, respectively, yields to a large difference to $\text{SWE}_{\text{all\_v\_cor}}$. This result is quite intuitive as precipitation and temperature have a large impact on SF as well as snow melt or accumulation. Compared to $\text{SWE}_{\text{all\_uncor}}$, not correcting precipitation and temperature shows a still much smaller difference to $\text{SWE}_{\text{all\_v\_cor}}$. For the C4I model the temperature signal is less dominant and for some stations even comparable to not correcting wind speed, i. e. $\text{SWE}_{\text{wss\_uncor}}$ (c. f. Figure B.26 and Figure B.27).

The interesting result in Figure 8.3 is the effect of not correcting wind speed on simulating SWE. The effect is smaller than for precipitation and temperature for most stations, but not negligible. Here, one should notice that for the snow model, wind plays only a role for sublimation. For Sonnblick (station #15410) not correcting wind

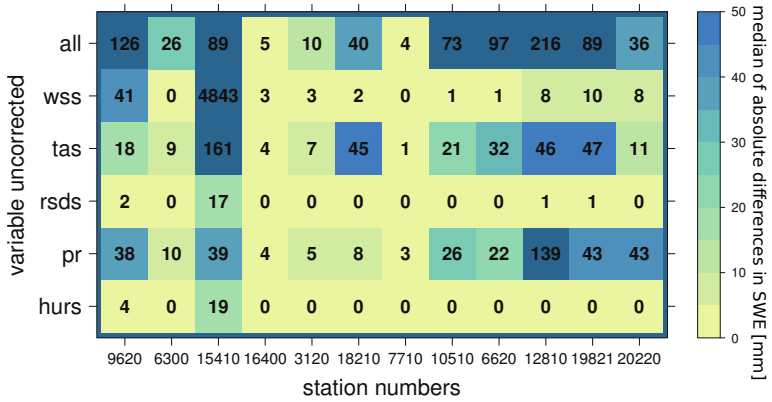


Figure 8.4: As in Figure 8.3 but showing the median of absolute difference in SWE [mm].

speed but all other variables has a tremendous effect, it increases the SWE in the median about 900 % (very similar for C4I model) which translates to 5 m SWE as absolute difference. The observation station of Sonnblick is located on a mountain top with exposure to the free atmosphere and extreme weather conditions (Section 5.3).

Two facts have to be considered for wind speed. One is, the RCM with a grid-spacing of 0.22° and a mean grid cell height, is not able to catch the climate on a mountain peak of 3000 m (which, of course, counts for all variables). Second is, for QM one needs good quality observation data for calibration and wind speed is a difficult variable to measure (Section 5.3). Only recently (about 15 years ago) the instruments were changed to ultra-sonic meters. Before cup anemometers were commonly used which have a start velocity of 0.4 m/s and which mechanics can freeze in winter, thus, in particular on mountain stations. As QM depends on observations, this is an open issue to consider.

Not correcting global radiation and relative humidity does not show any relevant effect on the resulting SWE throughout the winter months.

8.2.3 *Effect on Nash-Sutcliffe-Efficiency*

One measure of linear correlation is the Nash-Sutcliffe-Efficiency (NSE) (for detailed definition see Section A.3). Considering a scatter plot of observed to simulated values, the NSE indicates the agreement to a 1:1 line (perfect correlation) (Moriasi et al., 2007). The NSE is defined between 1 and $-\infty$, where 1 is the perfect match of observation and simulation, whereas negative values are interpreted as no match. Moriasi et al. (2007, Table 4) defines values ≤ 0.5 as unsatisfactory. In this study the focus is on the change in NSE for different v_uncor.

In Figure 8.5 three different references were used: a) snow depth $(\mathrm{SD})_{\mathrm{all\_v\_cor}}$, b) $\mathrm{SD}_{\mathrm{obs}}$, and c) $\mathrm{SD}_{\mathrm{AMUobs}}$ (AMUNDSEN results driven with observations). Taking AMUobs as reference should reduce the ef-

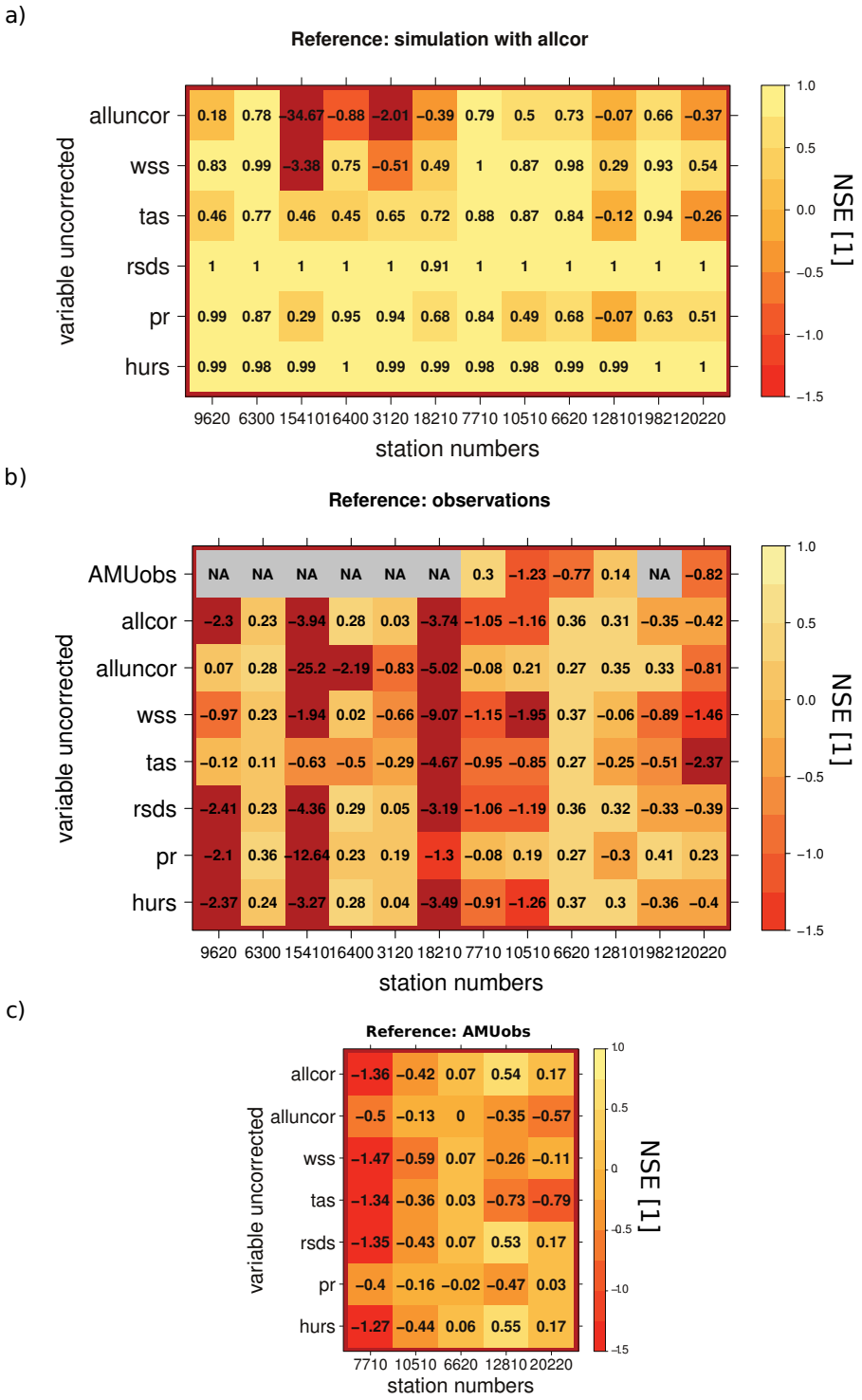


Figure 8.5: NSE of SDv_uncor with reference a) $SD_{all\_v\_cor}$, b) SD_{obs}, c) SD_{AMUobs}, for November to March 1982 to 2000, METNO-HIRHAM (ERA40 driven).

fect of observed undercatched precipitation (e. g., Guo et al., 2001) and the uncertainty in the snow model output. Here, the re-analysis driven RCMs were used for winter months of 20 years 1982 to 2000.

Considering SD_{v_uncor} compared with the optimal case of fully corrected climate variables ($SD_{all\_v\_cor}$), decreases the NSE only little for relative humidity and global radiation (Figure 8.5 a). For SD_{v_uncor} with v from the variables wind speed, temperature, and precipitation the NSE strongly depends on the station, as for some stations the $NSE(SD_{v\_uncor})$ is close to 1 while for other stations, like the mountain top station Sonnblick, the $NSE(SD_{v\_uncor})$ is negative. Not correcting any climate variable results in expected low values for $NSE(SD_{all\_uncor})$.

$SF_{all\_v\_cor}$, which depends mainly on the correct representation of temperature and precipitation amount shows a quite high NSE (not shown).

The SD_{obs} as reference is demanding for the NSE (Figure 8.5 b), because the observed precipitation has not been corrected for wind shift (c. f. Section 5.3, Guo et al. (2001)). The undercatch of observed precipitation produces a systematic bias in SD through the calibration with observations for QM. I. e. by bias correcting towards an untrue observed precipitation, the snow model simulates a biased SD. Figure 8.5 b shows poor results of the NSE for SD throughout most stations, even if all climate variables were corrected. Only for four stations the NSE is mostly positive. Additionally, the NSE of SD_{AMUobs} is calculated with observations as reference (first row in Figure 8.5 b), showing unsatisfying results. This confirms the difficulties with precipitation undercatch, but also can be explained by snow model errors.

Figure 8.5 c shows the NSE with SD_{AMUobs} as reference, with similar bad results as before. Bias correcting all climate variables improves only the NSE of 3 out of 5 stations.

Neglecting the snow model uncertainty, the NSE was calculated using the snow model driven with observations (Figure 8.5 c). The structure, related to which data were used within the snow model, is comparable to Figure 8.5 a and b. The NSE for each station differs from Figure 8.5 b, e. g., Gmunden (station #6620) shows positive values in both plots, but using observations in the snow model leads to values very close to zero, Pörtschach (station #20220) shows better results when compared with observation driven simulations, than with observations.

As mentioned in the beginning of this subsection, the NSE shows the distance to a perfect linear correlation. The results presented above for METNO, and for C4I as well, show a mostly positive NSE with reference $SD_{all\_v\_cor}$. Using SD_{obs} as reference, the NSE results are very mixed but overall rather poor, when considering the effect of bias correction. Beside the strong station dependency, only half of the sta-

tions show an improvement in the NSE for $SD_{all\_v\_cor}$ compared to SD_{all_uncor}. This result is quite disappointing.

One explanation is the precipitation undercatch in the observations, which has not been corrected. This is a conceptual problem of this study. When the study was designed, the focus was on the comparison of uncorrected and corrected climate model input and the effects on the snow model output. Later, the comparison with the snow observations became more relevant, as that is the reference for impact modellers. It is planned to repeat parts of this study with undercatch correction applied to the observed precipitation.

Another explanation (additionally) are the RCMs themselves. The inter-variable correlations of RCM output compared to those of observations is rather poor (Section 7.2). It has been shown in Section 7.2 that QM does not change the inter-variable correlations given by the RCM. Derived variables, like the snow model output, depend on the correct representation of the inter-variable relations, and therefore the correction of single variables cannot improve all statistics of the derived variable.

8.2.4 *Effect on accuracy, bias, reliability*

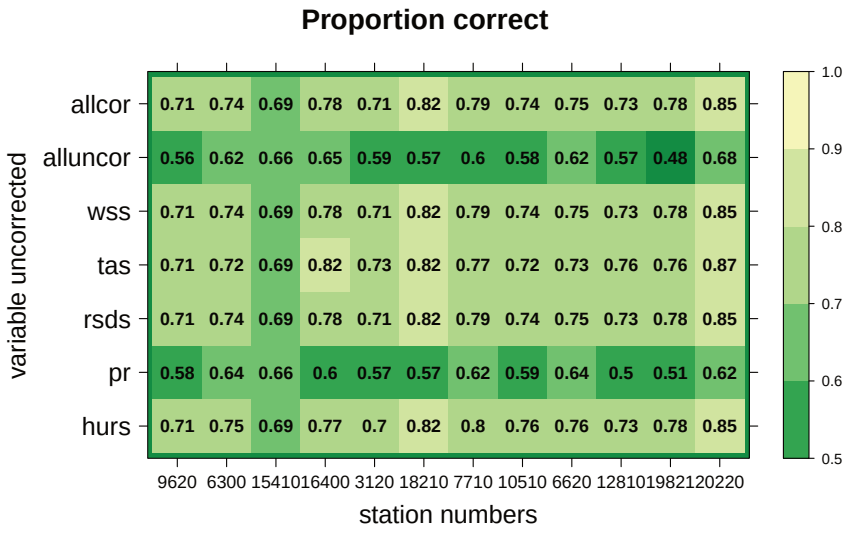
Investigating not only if the amount of snow is changed by bias correcting climate variables, a simple dichotomous forecast for snow fall amount (SF) is evaluated for each station and correction situation. Therefore, the three dimensions of a 2×2 contingency table were calculated: proportion correct (PC), which is an accuracy measure; bias of yes forecast (BYF), which is "the ratio of the number of yes forecasts to the number of yes observations" (Wilks, 2006, p. 264); and false alarm ratio (FAR), which shows the reliability of the forecast as fraction of yes forecasts that turn out to be wrong (c. f. Section A.4). The reference here is observed SF, which is why the re-analysis driven RCM output was used.

The PC for $SF_{all\_v\_cor}$ events can be improved (see light colour in Figure 8.6 a, first two rows). Regarding the occurrence of SF the not-correction of single variables does not have any impact on the $PC(SF_{v\_uncor})$, besides for not correcting precipitation.

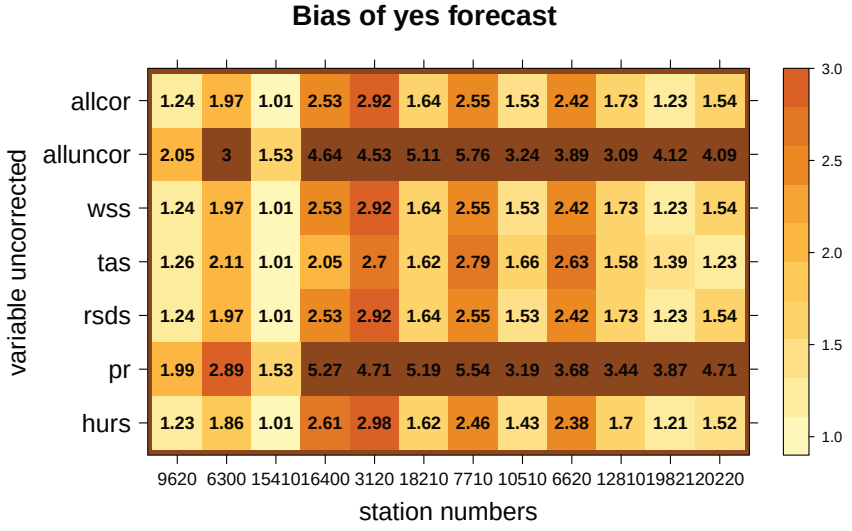
The frequency of forecasted SF events is clearly overestimated by the RCMs which is shown as BYF in as values > 1 (over-forecasting) in Figure 8.6 b (for METNO). Even for $SF_{all\_v\_cor}$ the NOS is too high, but strongly reduced. This confirms the result in Section 8.2.1, where the absolute NOS_{v_uncor} (upper values in Figure 8.2) is mostly higher than the absolute $NOS_{all\_v\_cor}$ (lower values). Obviously, the correction of precipitation has the strongest impact on the number of SF events.

The station dependency of the results is most prominent for BYF. Stations like Aschach (#3120) can be improved regarding BYF, but still

a)



b)



c)

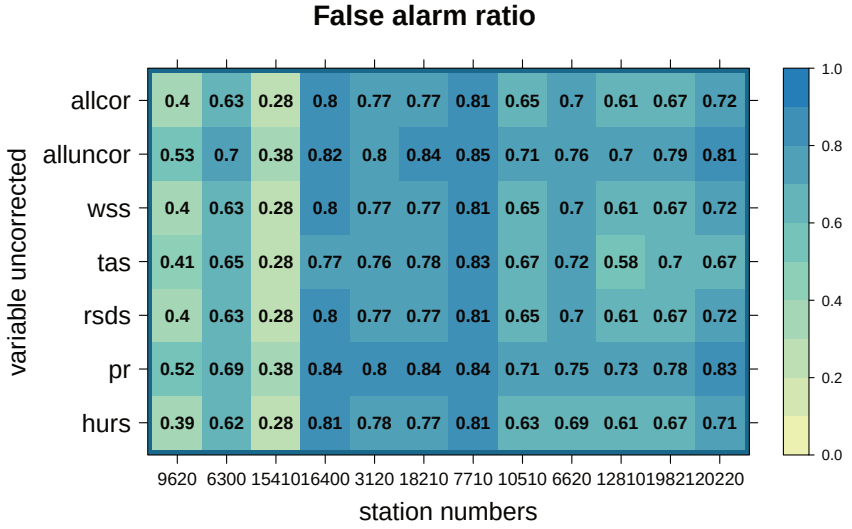


Figure 8.6: The three of a yes/no forecast for SF are proportion correct (PC) (a), bias of yes forecast (BYF) (b), and false alarm ratio (FAR) (c). The months November to March are included for 1970 to 2000, METNO-HIRHAM-ERA40.

have a SF over-forecasting (BYF of about 3), whereas for stations like Krippenstein (#9620) and Weissensee-Gatschach (#19821) the over-forecasting can be further reduced (BYF with values just above 1). The Sonnblick station (#15410) shows a perfect forecast after bias correction, which relates to the already good performance of SF_{all_uncor} and the high elevation. Liquid precipitation is quite rare during winter months anyway, which means it is a pure shift of the distribution in precipitation amount.

The overall performance for the FAR is rather poor (0 is good, 1 is bad), beside for the two more elevated stations, Krippenstein (#9620) and Sonnblick (#15410). The results for the proportion of forecasted SF events to those which did not happen relates to the strong over-forecasting found in Figure 8.6 b.

Figure 8.6, in particular b), shows the difference in performance for different stations, which has been observed with NSE already (Section 8.2.3). Here, we did not investigate where the difference comes from, but as the difference shows up when using uncorrected climate variables as well as corrected ones, I suggest this uncertainty originates from the snow model.

8.2.5 *Effect on persistence*

One measure of persistence is the autocorrelation function (ACF). Here, Figure 8.7 shows the ACFs of the simulated SF are plotted with the ACF of the observed SF (black) and the ACF(SF_{AMUobs}) (purple). The autocorrelation of $SF_{all\_v\_cor}$ (red) shows an improvement over the autocorrelations of SF_{v_uncor} when compared to the observed ACF for most stations. For, e. g., station Reichenau the ACF of SF_{all_uncor} (orange) is closer to the observed ACF, whereas the ACF of $SF_{all\_v\_cor}$ is closer to the one of SF_{AMUobs}. This example is even more pronounced for the C4I model (Figure B.29). The difference of SF_{obs} and SF_{AMUobs} is attributed to the snow model. Being aware of the different monthly characteristics, the ACFs were investigated for single months as well (not shown), with very similar results as for the whole winter period.

As a second measure of persistence the maximum snow cover period per winter season is investigated. Here the snow cover is defined as a SD of more than 10 cm for more than one day[2]. Validated were the maximum number of consecutive days with SD > 10 cm for each season. The threshold here is not related to any impact related number, but to a visible snow cover which still leads to multi-day periods even for stations with low elevation.

2 In literature one finds thresholds of > 30 cm accumulated per month as minimum requirement for skiing (e. g., Witmer, 1986; Koenig and Abegg, 1997; Scott and McBoyle, 2003).

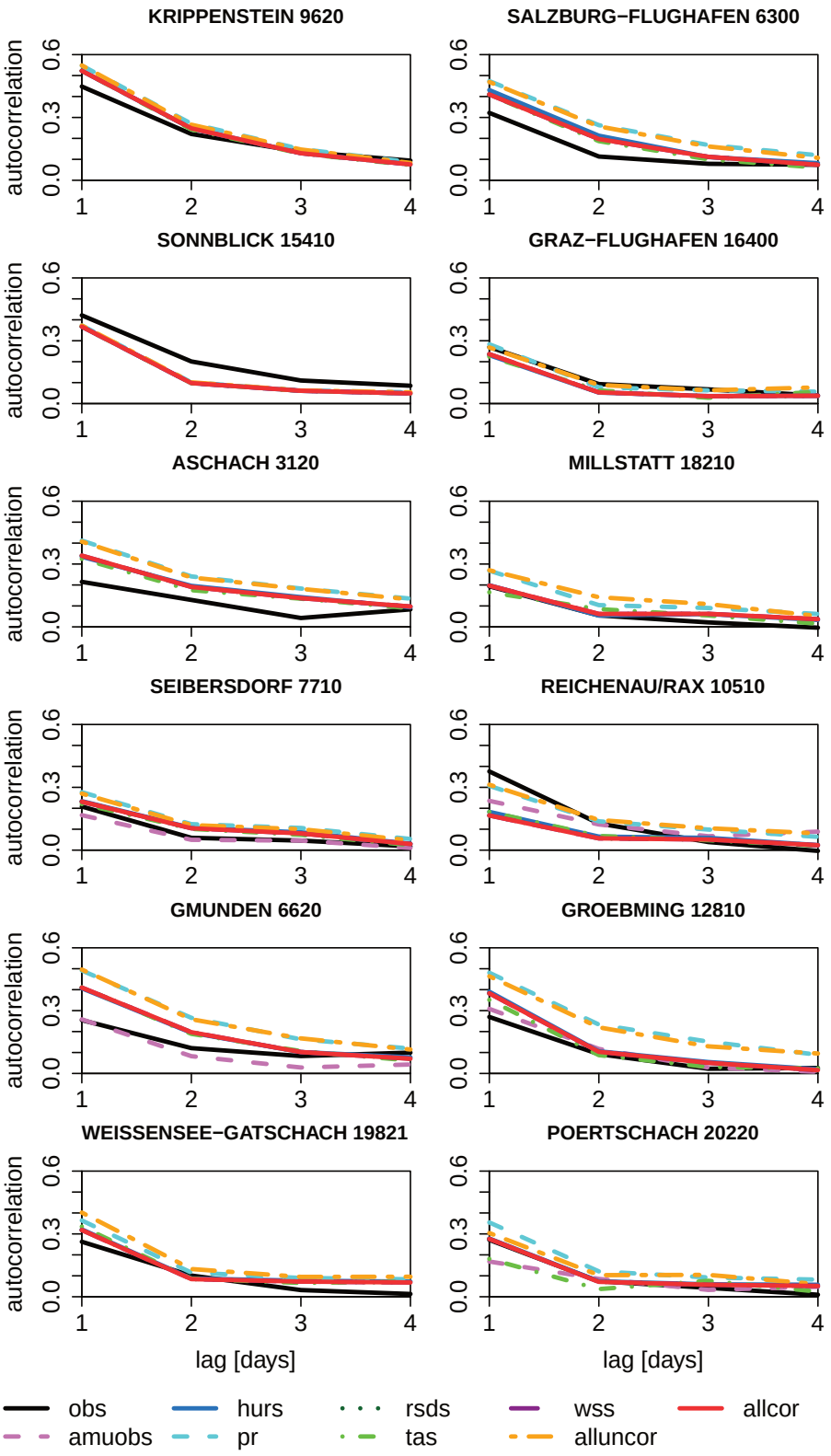


Figure 8.7: Autocorrelation of SF inclusive simulation driven by observations, different colours/line types indicate v_uncor and all_v_cor. Observations are in black, AMUobs is in orchid. For November to March, 1981 to 2000, METNO-HIRHAM-ERA40.

.

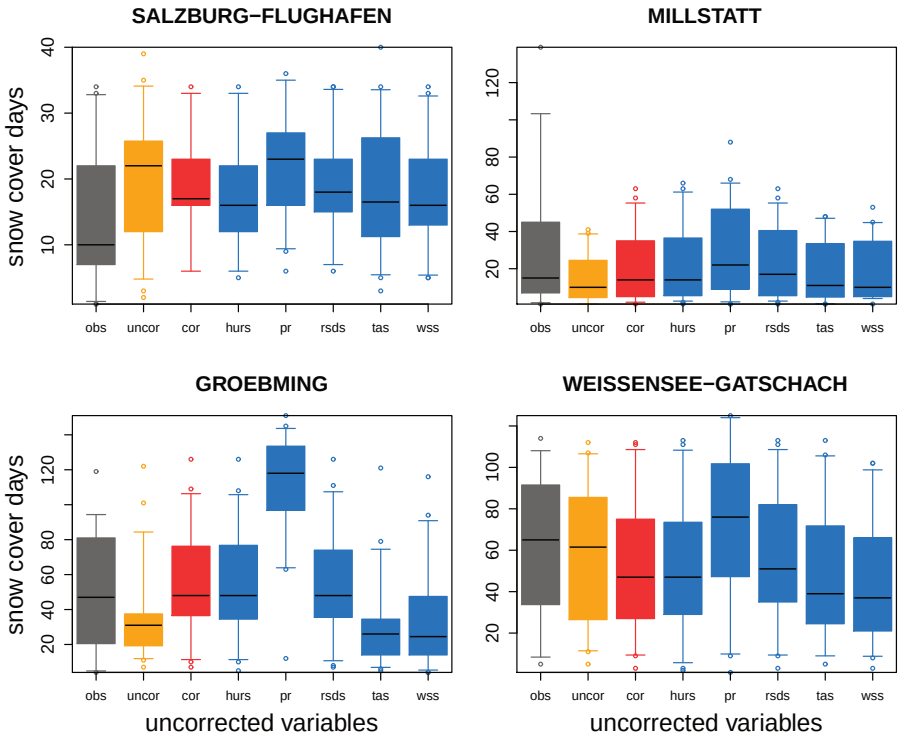


Figure 8.8: Maximum number of consecutive snow cover days (≥10 cm) per winter in 1970 to 2000, METNO-HIRHAM-ERA40.

Figure 8.8 shows the maximum snow cover periods for six selected stations (with enough snow cover to evaluate) for the METNO model. Comparing the observed snow cover days (grey box) with simulated snow cover days, using allcor variables (red box) shows an improvement for most stations compared to not correcting the climate variables at all (orange box). The effect of v_uncor (blue boxes) is, again, strongly v and station dependent.

8.2.6 *Effect on temperature-humidity index*

Till here, only the winter season has been investigated as well as only strong precipitation dependent indices. To cover the summer season as well the temperature-humidity index (THI) has been investigated. The THI as defined by Schoen (2005) is a heat index which considers humidity as additional stress factor for the human body (c. f. Section 5.4).

The THI was calculated for July and August (hottest months in Austria) for 20 years (1981 to 2000) using all four RCMs (hindcast driven). For validation the density distributions and monthly mean biases were derived on six station locations. The locations were chosen regarding the maximum mean temperature in those 20 years. Reference is the THI calculated from observed temperature and humidity.

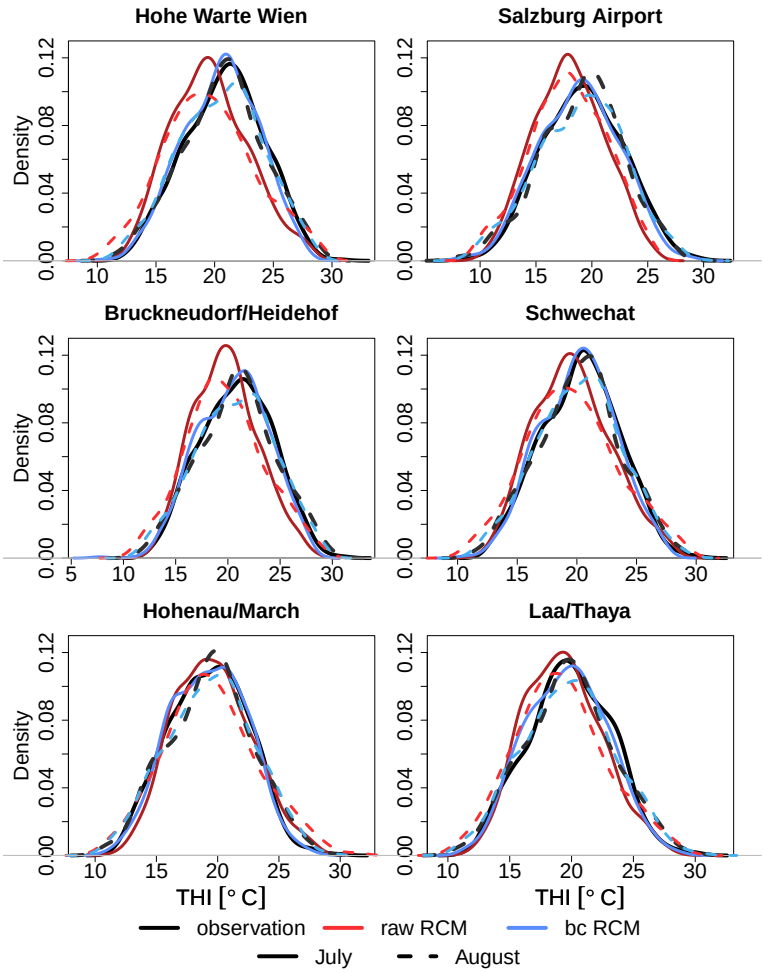


Figure 8.9: Density distribution of the temperature-humidity index (THI) for six stations in Austria for observation (black), raw RCM (red), bias corrected RCM (blue); for July (solid) and August (dashed). RCM shown: SMHI-RCA-ERA40; period: 1981 to 2000. Note the different ranges of the x-axis.

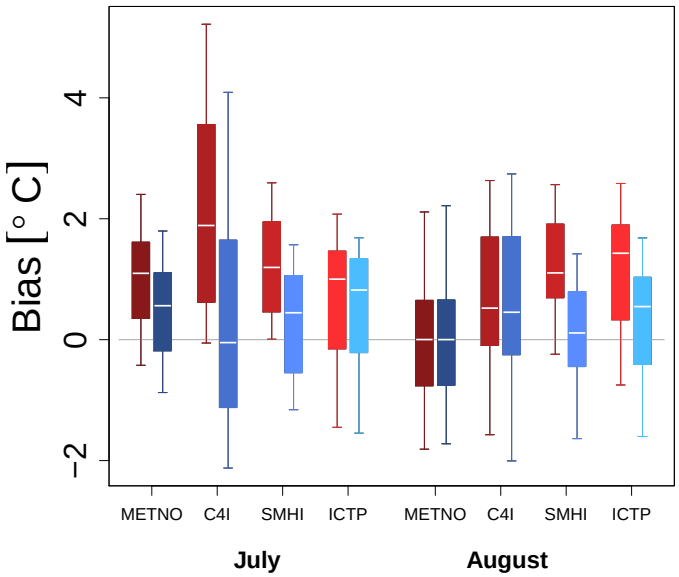


Figure 8.10: Box-whisker plot for bias in the THI for Salzburg for raw RCMs (red), bias corrected RCMs (blue); for July (left) and August (right) for period 1981 to 2000. Details for box-whisker definition see Figure 6.2.

For most stations the THI density distribution in terms of raw RCM is shifted to smaller THI values compared to the observed THI distribution, which can be reduced by using bias corrected RCM output (Figure 8.9 for SMHI). The results for the other three RCMs are similar (c. f. Figure B.31 to Figure B.33), even though the THI distribution of the The Abdus Salam International Centre for Theoretical Physics (ICTP) model does not catch the height of the distribution at all stations for July (c. f. Figure B.33). However, the overall picture shows an improvement in the distributions of the THI after bias correction temperature and relative humidity.

The monthly mean bias of the THI is improved for both summer months at most stations for all RCMs. Figure 8.10 shows as a representative example the station Salzburg where the bias in THI is reduced for all models. The bias of other stations (c. f. Figure B.34) is reduced for the C4I and SMHI model for both months. The results for METNO and ICTP show a good improvement for August. For July the THI of th two RCMs is sometimes over-corrected, i. e. the bias in the THI from corrected RCM output has an opposite sign than the bias from raw RCM output. In Chapter 6 it was shown that QM strongly reduces the mean bias for all variables. It was also shown that the monthly biases for temperature and relative humidity show the effect of non-stationarities in the error. That results in an over-correction of the bias, which would explain the heterogenity in the results of the monthly mean bias of the THI for July.

8.3 SUMMARY OF INDIRECT VALIDATION

The indirect validation of QM via snow model output and THI showed the relevance and performance which is important for impact modellers. One motivation for this study was to evaluate the pure effect of bias correcting climate variables on derived indices. A second motivation have been situations when not all requested meteorological variables can be provided with long enough observed time-series. To understand the effects of single uncorrected variables on the quality of impact model output, an attempt was made by investigating snow model output driven by five climate variables, which have been uncorrected, partly and fully corrected. As the snow model strongly depends on precipitation, a second model has been investigated, the THI, which only depends on temperature and humidity. Here, either both variables have been corrected or not corrected.

Investigating the pure difference between using corrected or uncorrected RCM output, GCM driven models have been used. The comparison of snow fall SF_{v_uncor} with $SF_{all\_v\_cor}$ has been investigated by calculating the number of snow days (NOS). The SWE_{v_uncor} was evaluated by calculating absolute and relative differences to the $SWE_{all\_v\_cor}$. Both indices showed a strong relevance of the correction of precipitation, which is intuitive as that is the main dependence for SWE (SD) and SF. The no-correction of temperature made a difference as well, in particular for SWE. Exposed stations, like Sonnblick, are very sensitive to the correction of wind speed.

To carry out a fair validation against observations and AMUNDSEN results driven with observations (AMUobs), hindcast-driven RCMs have been used as input for the snow model. The Nash-Sutcliffe-Efficiency (NSE) as measure for correlation showed a rather poor model performance for nearly all stations. This can be explained partly by the not-corrected undercatch in observed precipitation. As the snow model depends on five meteorological variables, their relation needs to be well represented in the input data. In case this is not true for the given RCM, applying QM does not correct a missing inter-variable relation.

Other measures like the proportion correct (PC), bias of yes forecast (BYF), false alarm ratio (FAR), and the persistence measures showed an improvement when fully corrected RCM data were used for the snow model. All measures showed the obvious importance of precipitation. Another strong feature was the station dependency of the skill. One explanation can be that the snow model's performance is location dependent.

For the snow model output validation, regarding most measures, there was an improvement when using fully corrected climate data compared to fully uncorrected ones. The difficulties with the correlation measure, NSE, need to be further investigated by applying a correction for the undercatch problem of observed precipitation.

The THI results depend on the RCM. The monthly bias of THI could be improved by bias-correcting temperature and relative humidity. However, the bias correction had a smaller effect on the heat index, or even changed the sign of the bias, for some stations in July when calculated with output from METNO and ICTP. However, the distribution of the THI was corrected quite well for all stations and models.

Nothing has such power to broaden the mind as the ability to investigate systematically and truly all that comes under thy observation in life.

— Marcus Aurelius

9 SUMMARY & CONCLUSIONS

In this work the prominent bias correction method Quantile Mapping (QM) is evaluated. After introductory chapters reviewing climate modeling, downscaling, error characteristics, and bias correction (Chapter 2 and 3), the QM method and data used in this work were introduced (Chapter 4 and 5).

The evaluation then starts with a first study answering basic questions regarding the main statistics and the temporal structure after the correction was applied to multiple variables. A second study focused on the effect of QM on inter-variable correlations and on the inter-variable relations of raw Regional Climate Model (RCM) output compared to observations. While in those two studies the effect of QM was directly validated on the variables applied to, a third study investigated the effect of bias correction on derived variables like snow cover.

In the first study (Chapter 6) the bias correction of multiple climate variables (2 m air temperature, precipitation amount, relative humidity, global radiation, wind speed, and surface air pressure) was validated regarding mean annual and mean monthly biases, distributions, and temporal structure. Four RCMs were used to cover the uncertainty. The RCMs were driven by hindcast data (ERA40) as well as General Circulation Models (GCMs). The outputs of these four models were downscaled to point scale and bias corrected. This has been done for station locations over Austria and parts of Switzerland and for 40 years in the past (1971 to 2010).

Validating QM directly on the variables corrected showed strong improvements for correcting the bias and distribution. A split sample approach was applied, i. e. the calibration period had no overlap with the corrected period. Under the assumption of stationarity in the error characteristics, this approach tested the application of QM in an unknown climate. Due to observational data availability for a period of 40 years, the climates of the split periods are not independent as they follow each other in time.

The results for monthly mean biases showed the annual cycle of the errors as well as the violation of the assumption of stationarity, in particular for temperature. In all cases the non-stationarity was not big enough to interfere with the positive effect of bias correction. In some rare cases, the error in the calibration period was bigger than in

the correction period, which led to an over-correction and to a change in the sign of the error.

The purely empirical implementation of QM successfully corrected variables with very different probability density distributions, which makes it highly flexible and applicable to various meteorological variables and regions. Considerable differences were found between the distributions of the uncorrected RCMs and observations in some variables, e. g., relative humidity and wind speed (in particular for ICTP-RegCM3). QM successfully adjusts also these distributions.

If strong modifications in the distribution of a variable are needed, the question whether the time-series are still plausible after correction demands an answer. Therefore, the autocorrelation and the root mean squared error (RMSE) of raw and corrected hindcast simulations were analysed. When applying QM to RCM output, improvement or no clear effect were found in RMSE and autocorrelation, as QM corrects for statistics but does not copy the observed time series. These results demonstrate that QM retains the quality of the temporal structure of the time-series of RCMs. This is not an improvement since also deficiencies of the RCMs in those features are retained. A similar situation arises regarding fine-scale spatial variability (which was outside the scope of this study). Maraun (2013) demonstrated that spatial and temporal variability of QM-corrected data show considerable deficiencies compared to observations. Those limitations are important to be aware of for the application of error-corrected model results in climate change impact studies.

The bias correction of single variables raised the question whether inter-variable relationships are still plausible after bias correction of individual variables, which was addressed in a second study and discussed in Chapter 7 (Section 7.2) by validating the correlation of pairs of variables. As in Chapter 6 five output variables (without surface pressure) of four RCMs (GCM driven) were used. With the future change in climate variables and their not linear correlation, a possible change in their correlation could be expected. Therefore the application of QM to future simulations was of interest and three periods were analysed, a historical period (1971 to 2010), a near future period (2021 to 2050), and a far future period (2069 to 2098).

When applying QM to RCM output, no clear effect on the correlation between meteorological variables could be found. These results demonstrate that QM retains the quality and also deficiencies of the inter-variable dependencies of RCMs.

In a comparison of correlations gained by *raw* RCM output and by *observations*, considerable differences were found, which questions the representation of inter-variable relations in RCM output. In Section 7.3 the conditional densities of variable pairs were compared for observations and RCM output. That is, the conditional densities of RCM output

variables were then compared to the conditional densities of observed variables.

The results showed a mismatch of the conditional densities of RCM and observations. The mismatch varies in intensity for regions, months, and variable pairs. Variables with a high bias in the simulation (e. g., relative humidity) are more difficult to interpret, as the inter-variable signal in the ratio of RCM and observation is weaker than the bias of single variables (like relative humidity). On the other hand, if the physics in the RCM was correct then any bias in one variable would translate into a corresponding bias in the other: but in this case the conditional densities would still be the same in the RCM and in the observations.

In a third study (Chapter 8) an indirect validation of QM applied to RCM output was performed. Thereby, a simple model (temperature-humidity index (THI)) and a complex impact model (snow model, AMUNDSEN) were driven with corrected and uncorrected RCM output. The THI was calculated for the months July and August in a 20-year period (1981 to 2000). The THI was calculated with corrected temperature and humidity as well as both uncorrected, from four RCMs.

For the more complex model a physical snow model (AMUNDSEN, Strasser, 2008) was used. The AMUNDSEN model was driven with seven different climate data sets of two RCMs, and additionally with observations. The climate data sets were: fully corrected RCM output (all_v_cor), fully uncorrected RCM output (all_uncor), and all variables corrected but one (v_uncor). The use of v_uncor investigated the common situation of lack of observations, which then lead to the impossibility of bias correcting single variables, while others could be corrected. With the snow model output (snow water equivalent (SWE), snow depth (SD), and snow fall amount (SF)) six further indices were validated for winter months (November to March) in the period 1971 to 2000.

Investigating the pure difference between using corrected or uncorrected RCM output in a snow model (AMUNDSEN), GCM driven models were used. The comparison of snow fall SF_{v_uncor} with $SF_{all\_v\_cor}$ was investigated by calculating the number of snow days (NOS). The SWE_{v_uncor} was evaluated by calculating absolute and relative differences to the $SWE_{all\_v\_cor}$. Both indices showed a strong relevance of the correction of precipitation, which is intuitive as that is the main dependence for SWE (SD) and SF. The no-correction of temperature made a difference as well, in particular for SWE. Exposed stations, like Sonnblick (mountain top), are very sensitive to the correction of wind speed.

To carry out a fair validation against observations and AMUNDSEN results driven with observations (AMUobs), hindcast-driven RCMs were used as input for the snow model. The Nash-Sutcliffe-Efficiency (NSE) as measure for correlation showed a rather poor model performance

for nearly all stations. This can be explained partly by the not-corrected undercatch in observed precipitation. As the snow model depends on five meteorological variables, their relation needs to be well represented in the input data. In case this is not true for the given RCM, applying QM does not correct a wrong inter-variable relation.

Other measures like the proportion correct (PC), bias of yes forecast (BYF), false alarm ratio (FAR), and the persistence measures showed an improvement when fully corrected RCM data were used for the snow model. All measures showed the obvious importance of precipitation, while another strong feature was the station dependency of the skill. One possible explanation is that the snow model's performance is location dependent.

For the snow indices validation, regarding most measures, there was an improvement when using fully corrected climate data compared to fully uncorrected ones. The difficulties with the correlation measure, NSE, need to be further investigated by applying a correction for the undercatch problem of observed precipitation.

The THI results depend on the RCM. The monthly bias of THI could be improved by bias correcting temperature and relative humidity. However, the bias correction had a smaller effect on the heat index, or even changed the sign of the bias, for some stations in July when calculated with output from METNO and ICTP. However, the distribution of the THI was corrected well for all stations and models.

QM can, by construction, map any distribution onto an arbitrary other distribution. This, however, does not necessarily indicate that the mapping is sensible in a physical way. The overall assumption for error-correcting RCMs is that RCMs represent the regional climate in a physical correct way in space and time. The on-going discussion if this assumption is justified was not a core topic in this PhD thesis. It is known that the performance of RCMs depends on region, season, and meteorological variable (e. g., Christensen et al., 2008). Nonetheless, in case of essentially no correlation between re-analysis driven RCM and observation (e. g., Appendix B, Figure B.24), one should reconsider or at least be aware of this when applying QM to this specific RCM (Widmann et al., 2003; Eden et al., 2012). If simulation and observation are not correlated, there is also no confidence about possible future trends projected by the simulations.

However, the retainment of the RCM's temporal structure together with large improvements with regard to biases, qualifies QM as a valuable, though not perfect, method at the interface between climate models and climate change impact research.

Future improvements of QM with regard to multi-variable error correction can be particularly expected from multi-variate approaches, which might lead to improved inter-variable dependencies (Piani and Haerter, 2012). In turn, improved inter-variable relations would improve impact model performance. However, such approaches are not

straightforward to implement due to limitations in the sample sizes usually available to build or estimate the distributions. In addition, more sophisticated inter- and extrapolation techniques could mitigate the effect of low-quality observational data and improve the representation of extremes. Particularly promising are also stochastic approaches, which could be implemented as add-on to QM and could lead to improvements with regard to spatial and temporal variability.

For the proper representation of new extremes, an extrapolation of the empirical cumulative distribution function (ECDF) outside the calibration range is necessary, which was not the focus of this thesis. Themeßl et al. (2012) found that a simple constant extrapolation leads to satisfying results also for precipitation extremes that are outside of the calibration range, which was confirmed recently by Bellprat et al. (2013). One might circumvent these issues by fitting theoretical distributions or some functions to the data (e. g., Piani et al., 2009; Rojas et al., 2011), but this would lead to less flexibility and the need for a specific implementation for each variable and probably also for each climate regime.

I also intend to apply the analysis of conditional densities on bias corrected (quantile mapped) data; in particular for variables with strong biases one should gain a better picture of inter-variable relations. As QM changes the distributions of each variable, a different output can be expected for the conditional densities. In case of relative humidity the correction of the marginal distributions might change the conditional distribution towards the observed, which might count as improving the inter-variable relation in the RCM. More thoughts and time will be needed for such an extended approach.

Overall, the present work has shown that QM corrects individual climate variables with convincing results, but needs further enhancements to improve the inter-variable relations in RCM output.

A

APPENDIX METHODS

A.1 APPLICATION OF DIURNAL CYCLE

To fit the demands of the project partners, the corrected data were brought to a finer time scale (3-hourly). Therefore the diurnal variation of 3-hourly uncorrected RCMs was calculated and applied to the corrected daily data. For temperature, relative humidity, and surface pressure, the difference in 3-hourly RCM data and their daily mean values were added (summed) to the corresponding corrected daily values (see Figure A.1). For precipitation, global radiation, and wind speed ratios of 3-hourly RCM data and their daily sums were multiplied by the corrected daily value. In case of relative humidity values can unrealistically exceed 100 %. All values of that very day are multiplied by a factor, shrinking the daily maximum value to 100 %. This is a simple method which relies on the uncorrected sub-daily data of the RCMs. By using post-processed daily values major biases are corrected, but model errors in the diurnal cycle (i. e. in temperature range or in timing of precipitation) are not corrected.

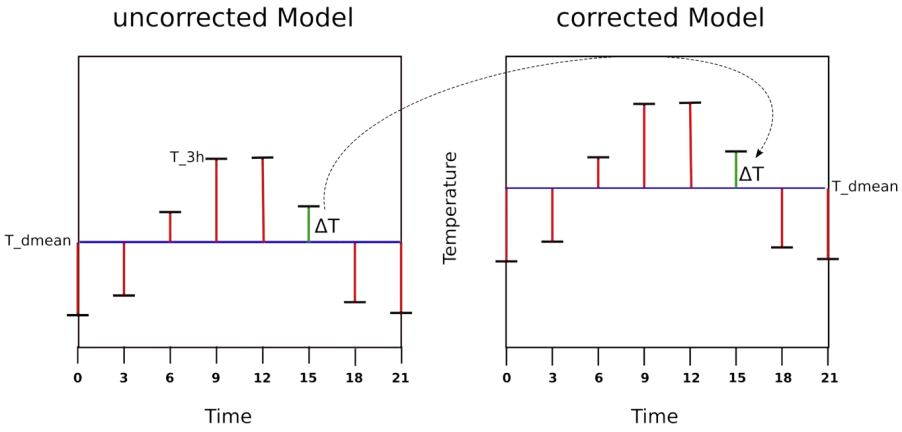


Figure A.1: Schematic description of the method to calculate 3-hourly data from daily error-corrected data for air temperature. The deviation of the 3-hourly time steps from the daily mean of the uncorrected model is added to the error-corrected daily mean.

A.2 TEMPERATURE-HUMIDITY INDEX

Steadman (1979) published a temperature-humidity index (THI) which was based on several biometeorological studies. Schoen (2005) found relative humidity and dry-bulb temperature too dependent on each

other and proposed to use "true independent" variables: dry-bulb temperature (T) and absolute humidity. As absolute humidity is rarely reported by observation stations, the dewpoint (D, see Equation A.3) is close enough to cover for absolute moisture.

The dewpoint can be estimated using the Magnus formula for saturated water vapour over water

$$e^*_{\mathrm{W}} = 9.1078 \exp\left(\frac{17.27T}{237.3+T}\right) \tag{A.1}$$

and the relation for e^*_{W} to relative humidity (f)

$$f = \frac{e}{e^*_{\mathrm{W}}(T_{\mathrm{a}})} \tag{A.2}$$

where e is the water vapour and T_{a} the air temperature. The dewpoint can be derived like

$$D = \frac{237.3\left(\frac{17.27T}{237.3+T} + \ln(f)\right)}{17.27 - \left(\frac{17.27T}{237.3+T} + \ln(f)\right)} \tag{A.3}$$

The new empirical model to calculate the THI developed by Schoen (2005) is given by

$$\mathrm{THI} = T - 1.0799 \exp(0.03755T)\left[1 - \exp(0.0801(D-14))\right], \tag{A.4}$$

where THI, T, and D are all in degrees Celsius.

A.3 NASH-SUTTCLIFFE-EFFICIENCY

The Nash-Sutcliffe-Efficiency (NSE) is defined as the residual variance (F^2) compared to the initial variance (F_0^2) (Nash and Sutcliffe, 1970) and is computed as shown in Equation A.5

$$\mathrm{NSE} = 1 - \frac{F^2}{F_0^2} = 1 - \frac{\sum_{i=1}^{n}(q_{\mathrm{obs},i} - q_{\mathrm{sim},i})^2}{\sum_{i=1}^{n}(q_{\mathrm{obs},i} - \overline{q_{\mathrm{obs}}})^2} \tag{A.5}$$

where q_{obs} and q_{sim} are the observed and simulated snow values and n the number of time steps. The NSE ranges from 1 to $-\infty$ with 1 indicating that observed and simulated values fitting a 1:1 line. Values > 0 are considered commonly acceptable, while Moriasi et al. (2007, Table 4) calls values ≤ 0.5 unsatisfactory.

Nash and Sutcliffe (1970) shows that the efficiency can be used to compare model modifications also. In this case Equation A.5 would be the proportion of residual variance

$$\mathrm{NSE}_{\mathrm{clim}} = 1 - \frac{F_2^2}{F_1^2} \tag{A.6}$$

where the suffixes 1 and 2 denote before and after the model modification and the suffix clim indicates the comparison inside the climate model world. In our study 1 relates to using only corrected climate variables, while 2 relates to climate data set with single uncorrected variables.

A.4 3 DIMENSIONS OF CONTINGENCY TABLE

The contingency table of a simple dichotomous forecast has 3 dimensions: accuracy, bias, and reliability. Those are defined and explained in the following on the basis of Wilks (2006, p 261–265).

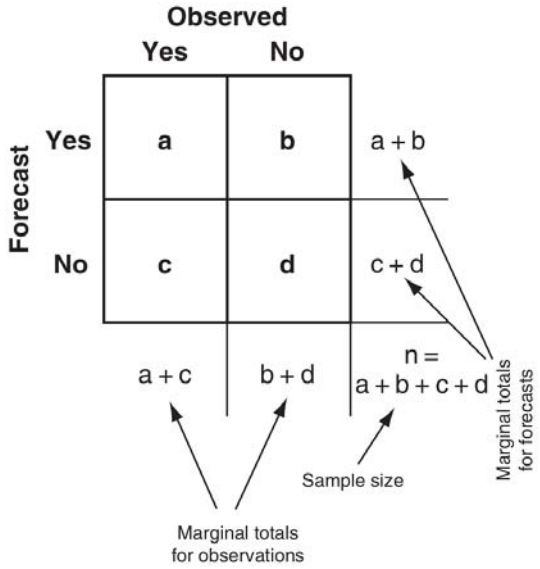


Figure A.2: 2×2 contingency table (Figure 7.1 a in Wilks, 2006). Bold a to d indicate the counts of forecast/event pairs of forecasts and observations, with the marginal totals and marginal distributions on the bottom and right.

A.4.1 *Accuracy*

Measures of accuracy describe the relationship between pairs of forecasts and the state of fulfilment. That means, a perfect forecast would have $b = c = 0$ in Figure A.2. The proportion correct (PC) is a direct and intuitive measure of accuracy proposed by Finley (1884). It is defined as

$$\mathrm{PC} = \frac{a + d}{n}. \tag{A.7}$$

in terms of Figure A.2. Correct yes and no forecasts count equally in PC, which can be problematic in cases with rare correct yes forecasts. However, likewise false alarms and misses are penalised equally. PC is defined between 0 (worst) and 1 (best).

A.4.2 *Bias*

The bias of yes forecast (BYF) of a contingency table measures the frequency of event forecasted to frequency of event observed and is defined like

$$\mathrm{BYF} = \frac{a+b}{a+c}. \tag{A.8}$$

A bias greater than 1 means the event has been overforecasted, whereas a BYF lower than 1 indicates an underforecasting of an event. When the event has been forecasted with the same frequency as it has been observe BYF equals 1.

A.4.3 *Reliability*

Reliability can be measured by the conditional relative frequencies of yes or no forecasts. The false alarm ratio (FAR) is the probability of not occurrence (o_2) conditioned on yes forecast (y_1), which is calculated like

$$\mathrm{FAR} = p(o_2|y_1) = \frac{b}{a+b}. \tag{A.9}$$

Wilks (2006) writes "FAR is ... that proportion of the forecast events that fail to materialize". FAR is defined between zero, which would be best, and one, which would be worst.

B

APPENDIX FIGURES

B.1 FIGURES: DIRECT VALIDATION OF QM

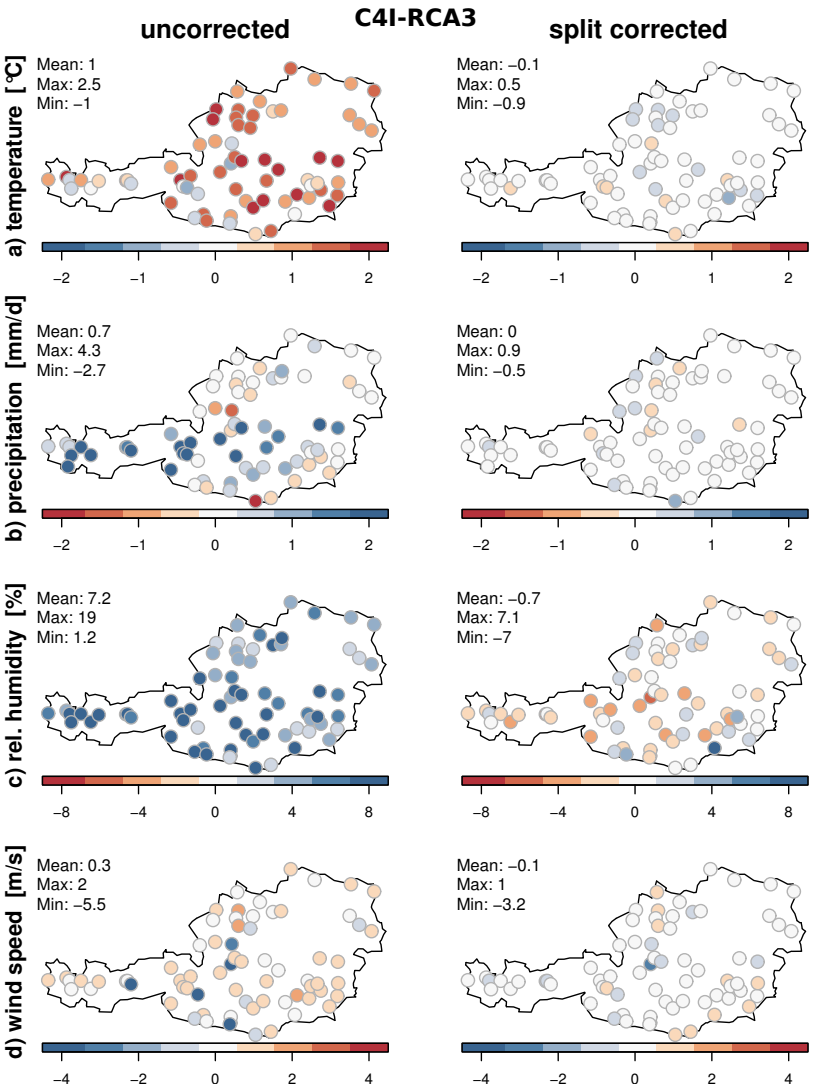


Figure B.1: Annual mean RCM (C4I-RCA3) bias at observation stations (1991 to 2010) for a) temperature, b) precipitation, c) rel. humidity, and d) wind speed (top down) for the raw RCM (for temperature altitude corrected), the error-corrected RCM with split sample evaluation, and the error-corrected RCM in technical evaluation (same calibration and application period).

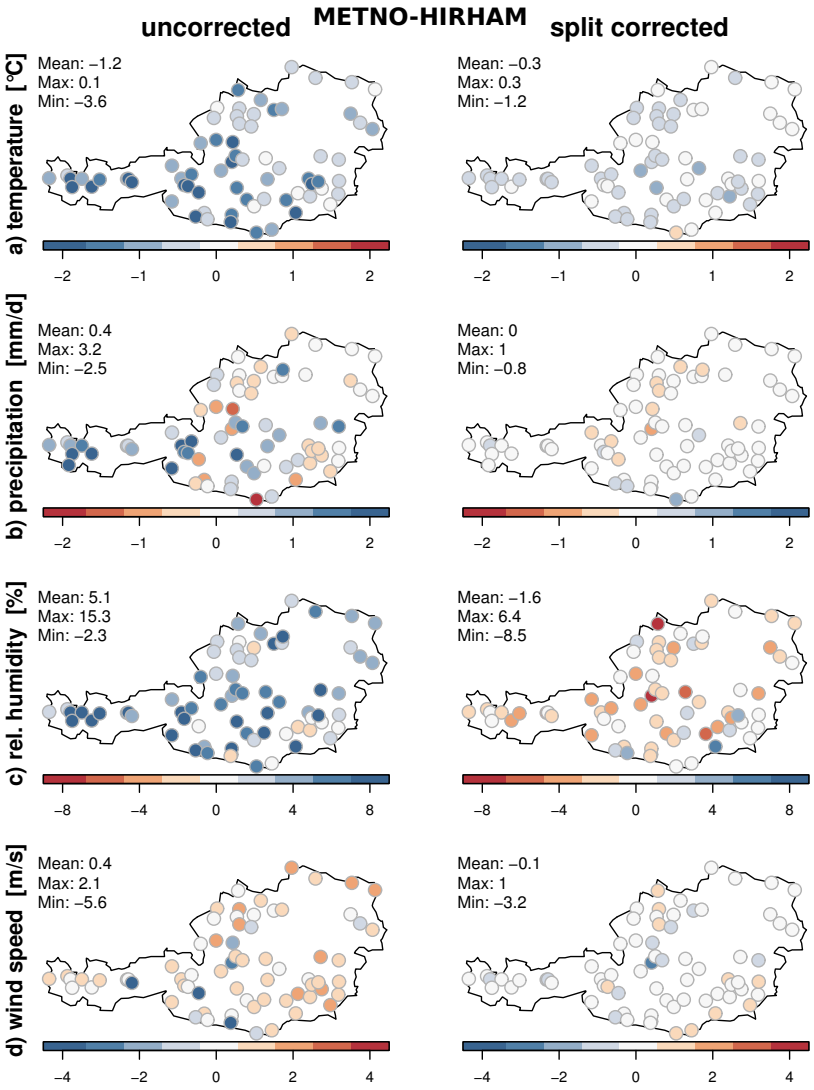


Figure B.2: Same as in Figure B.1 for METNO-HIRHAM

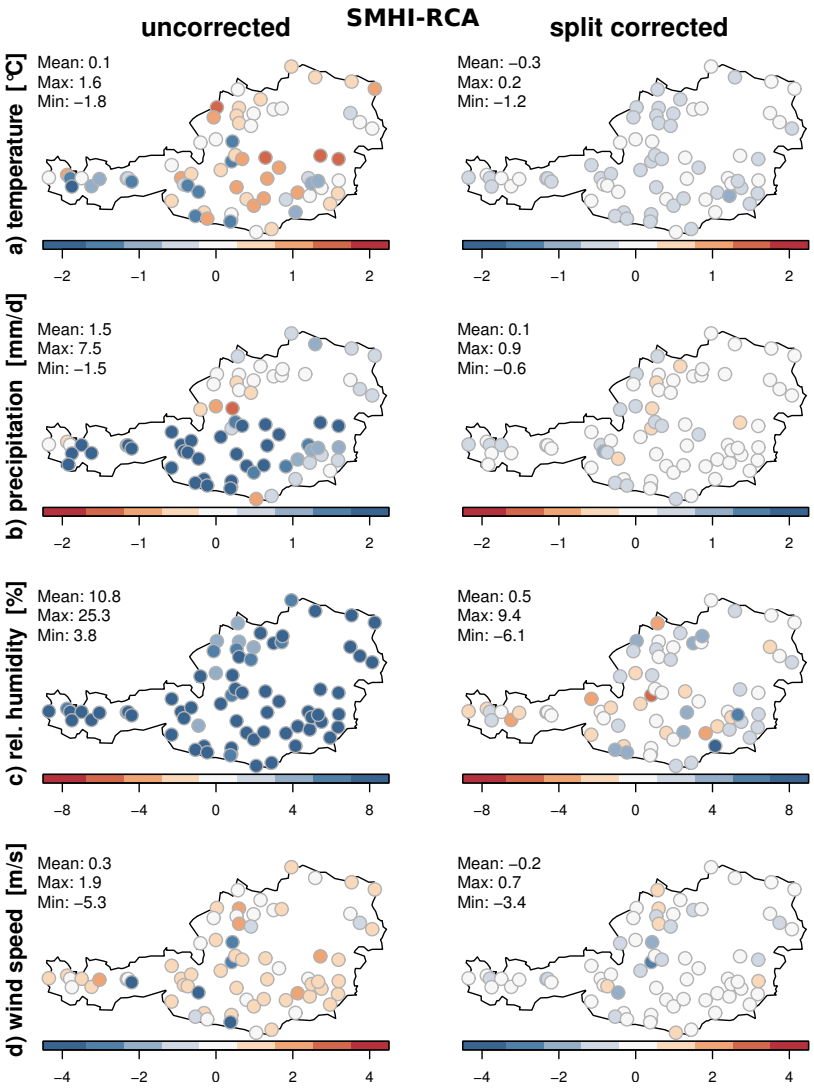


Figure B.3: Same as in Figure B.1 for SMHI-RCA

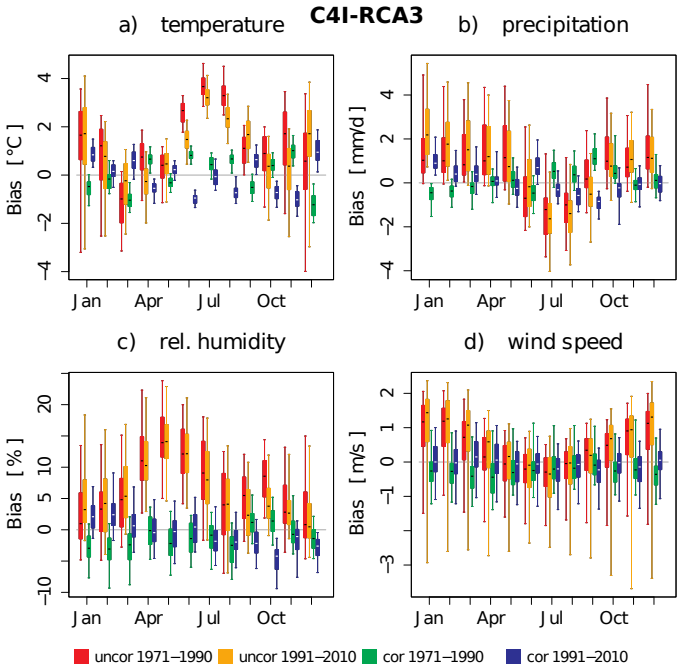


Figure B.4: Monthly bias of a) temperature, b) precipitation, c) rel. humidity, and d) wind speed as box-whisker plots for the uncorrected RCM splited in two periods (red and orange), the error-corrected RCM (C4I-RCA3) with split sample evaluation (green and blue), see Section 6.2.1 and Figure 6.2 for detailed description.

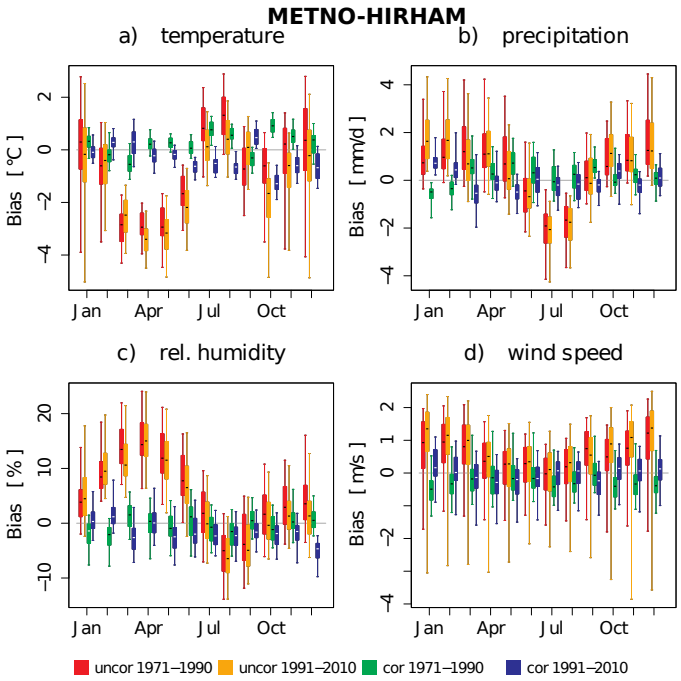


Figure B.5: Same as Figure B.4 but for METNO-HIRHAM

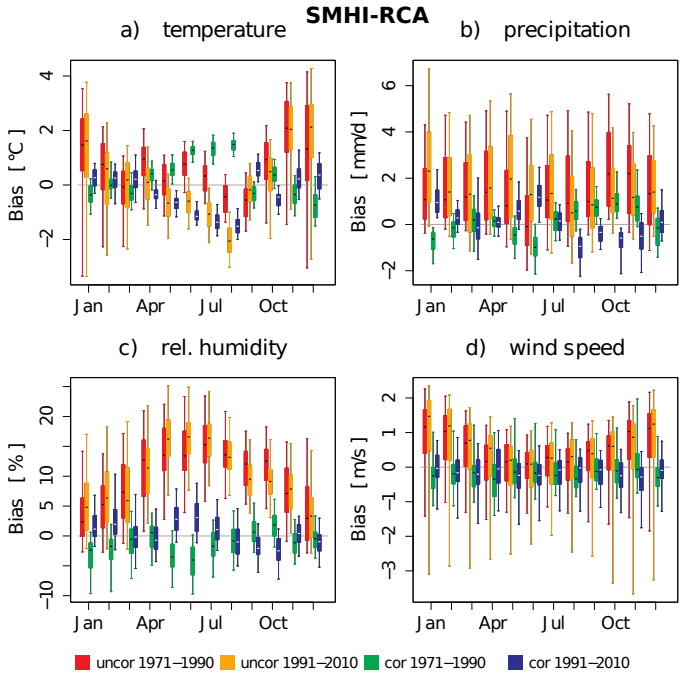


Figure B.6: Same as Figure B.4 but for SMHI-RCA

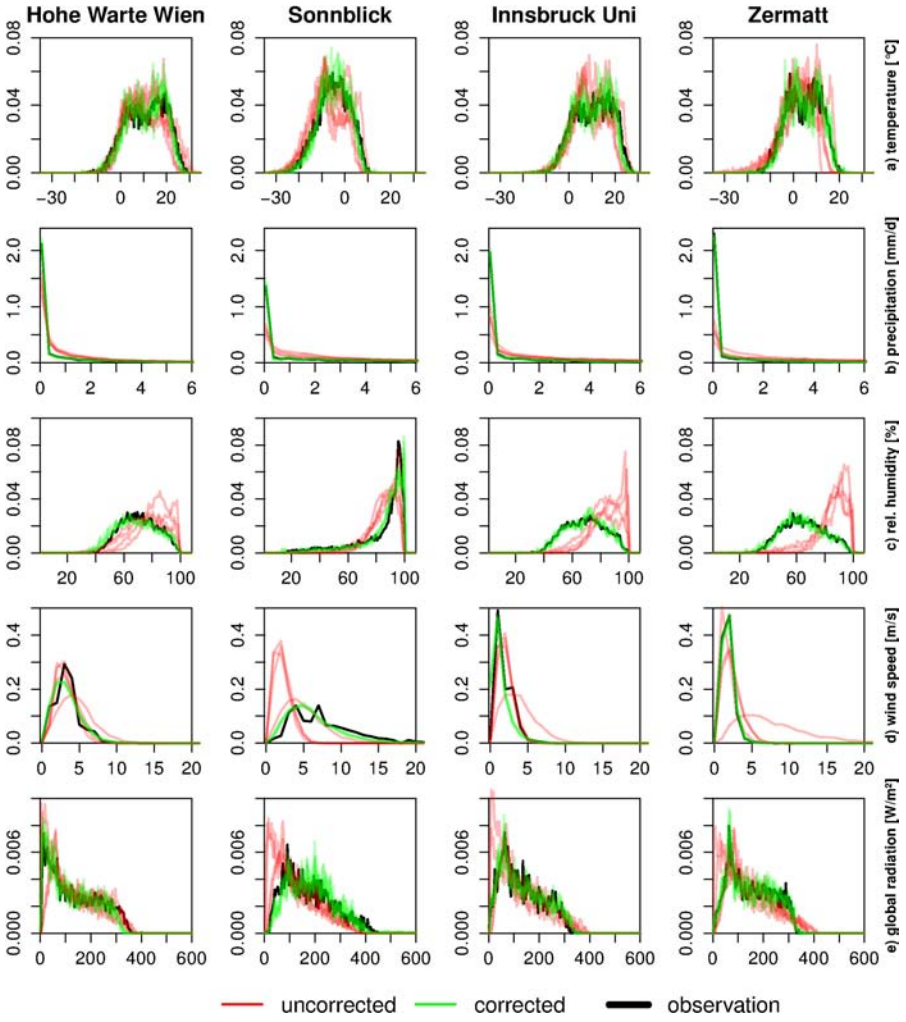


Figure B.7: Density distributions of the four raw RCMs (red thin lines), the split sampled error-corrected RCMs (green thin lines), and the observations (black fat line) for 1991 to 2010 of Hohe Warte Vienna, Sonnblick, Innsbruck University, Zermatt for a) temperature, b) precipitation, c) rel. humidity, d) wind speed, and e) gl. radiation.

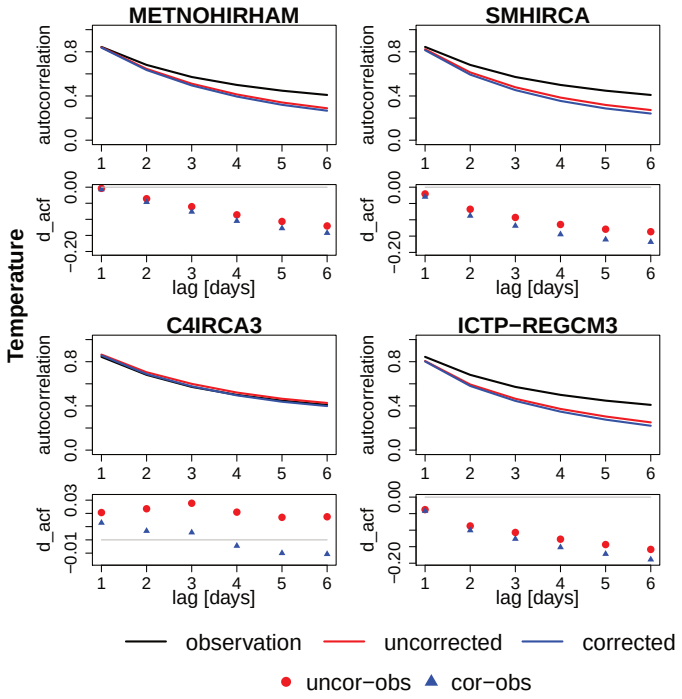


Figure B.8: Upper panel shows autocorrelation function for 20 year period (validation period 1981 to 2000) of temperature for observations (black), raw RCM-ERA40 (red), and split sample corrected RCM-ERA40 (green). Delta autocorrelation of raw (red) and corrected (green) to observations in the lower panels. For temperature the annual cycle has been removed for this figure. Panel pairs are shown for all four RCMs.

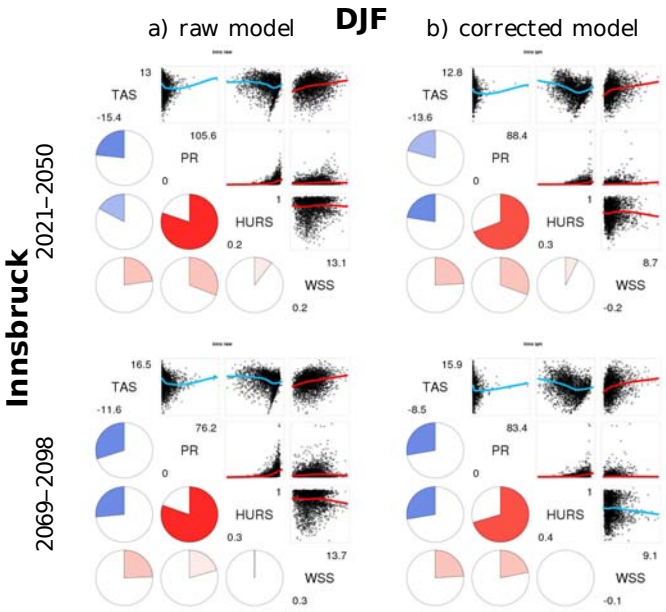


Figure B.9: Correlation matrices for **Innsbruck** for winter (DJF) of the periods a) 2021 to 2050 and b) 2069 to 2098. Including temperature (tas), precipitation (pr), relative humidity (hurs), and wind speed (wss) for selected stations in Austria and Switzerland for a) modelled, b) error-corrected modelled data. Pie charts show Spearman correlation coefficients, indicated with counterclockwise (negative correlation) and clockwise (positive correlation) pie slices. Lines in scatter plots are the Loess fit. The values above and below the variable names give the range of the data. The model shown here is the ICTP-RegCM3.

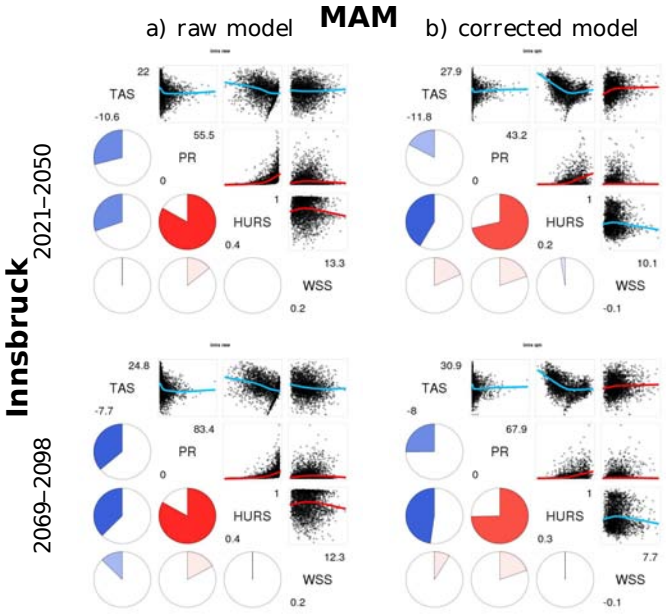


Figure B.10: Correlation matrices for **Innsbruck** for spring (MAM) of the periods a) 2021 to 2050 and b) 2069 to 2098. Same layout as in Figure B.9.

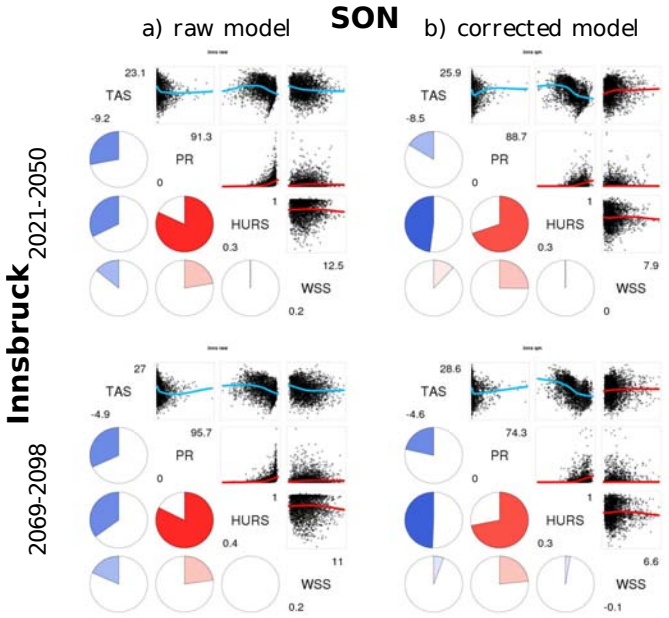


Figure B.11: Correlation matrices for **Innsbruck** for autumn (SON) of the periods a) 2021 to 2050 and b) 2069 to 2098. Same layout as in Figure B.9.

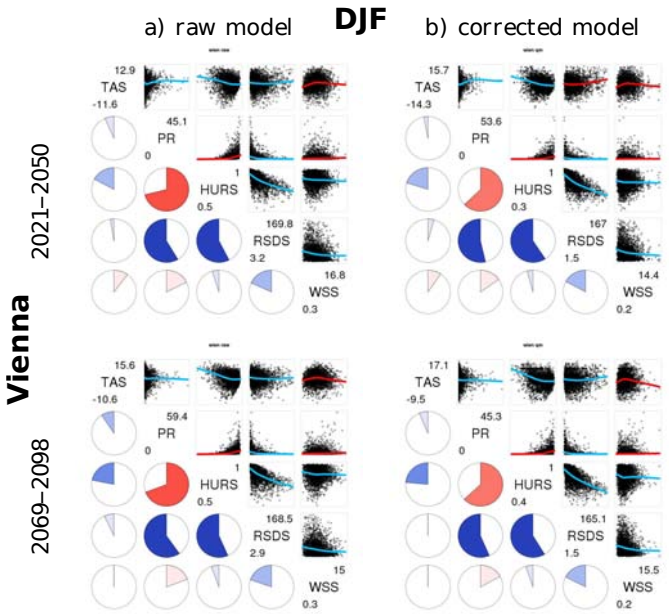


Figure B.12: Correlation matrices for **Vienna** for winter (DJF) of the periods a) 2021 to 2050 and b) 2069 to 2098. Same layout as in Figure B.9 but including global radiation (rsds).

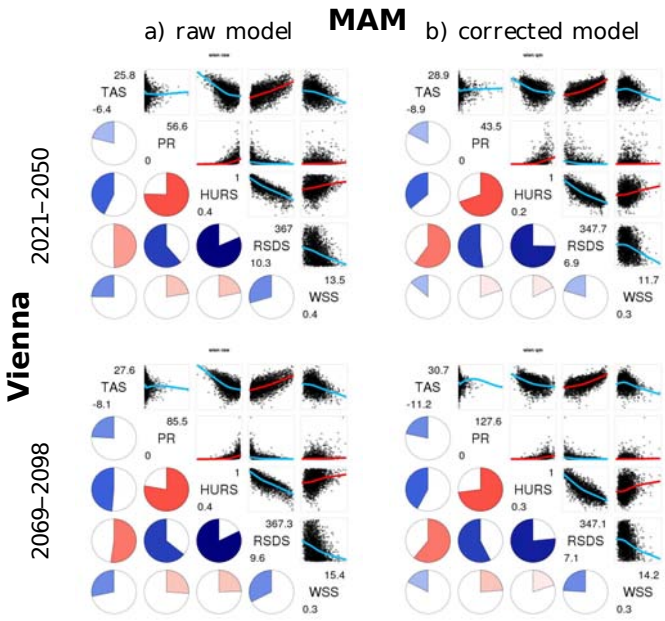


Figure B.13: Correlation matrices for **Vienna** for sprint (MAM) of the periods a) 2021 to 2050 and b) 2069 to 2098. Same layout as Figure B.9 but including global radiation (rsds).

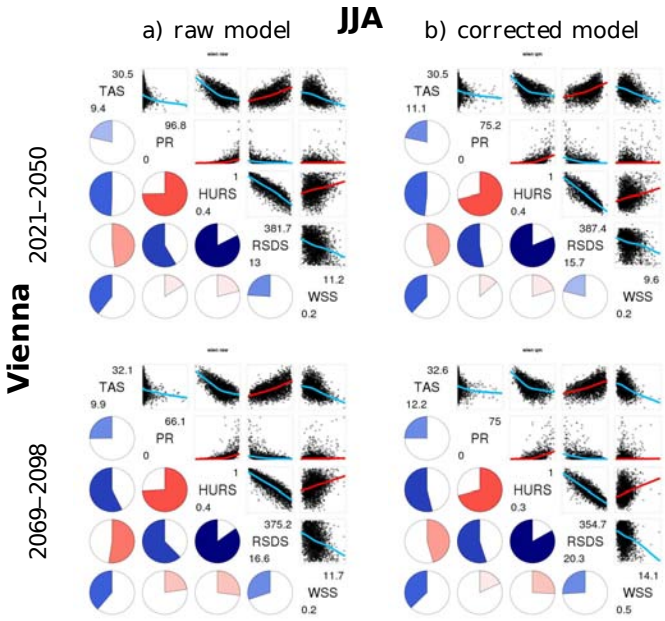


Figure B.14: Correlation matrices for **Vienna** for summer (JJA) of the periods a) 2021 to 2050 and b) 2069 to 2098. Same layout as in Figure B.9 but including global radiation (rsds).

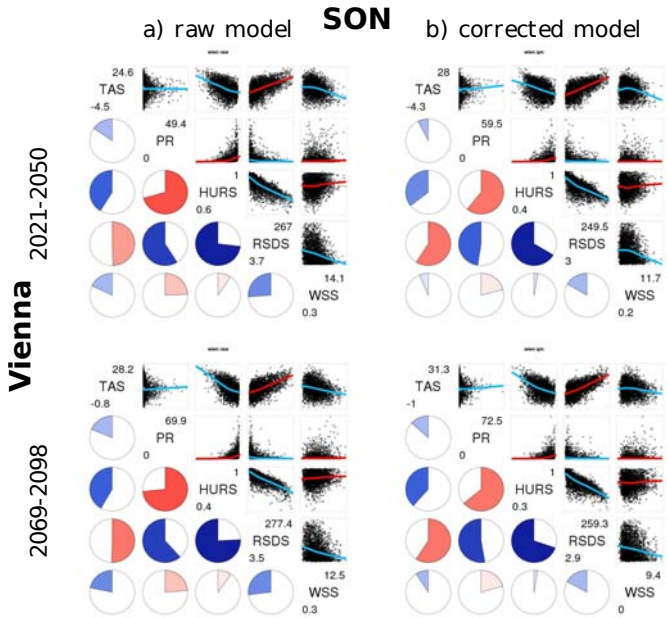


Figure B.15: Correlation matrices for **Vienna** for autumn (SON) of the periods a) 2021 to 2050 and b) 2069 to 2098. Same layout as in Figure B.9 but including global radiation (rsds).

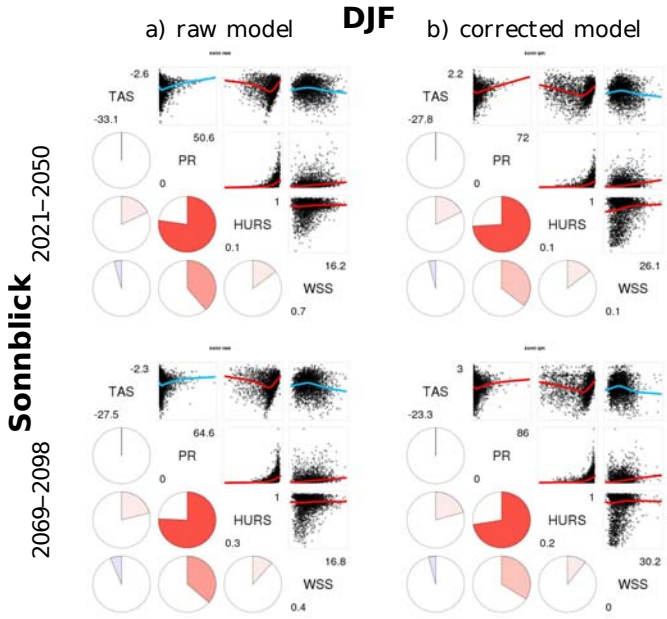


Figure B.16: Correlation matrices for **Sonnblick** for winter (DJF) of the periods a) 2021 to 2050 and b) 2069 to 2098. Same layout as in Figure B.9

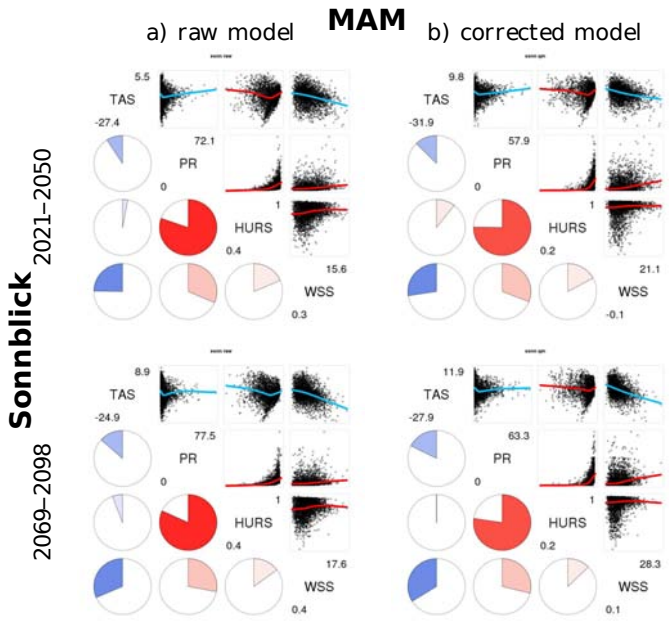


Figure B.17: Correlation matrices for **Sonnblick** for sprint (MAM) of the periods a) 2021 to 2050 and b) 2069 to 2098. Same layout as in Figure B.9

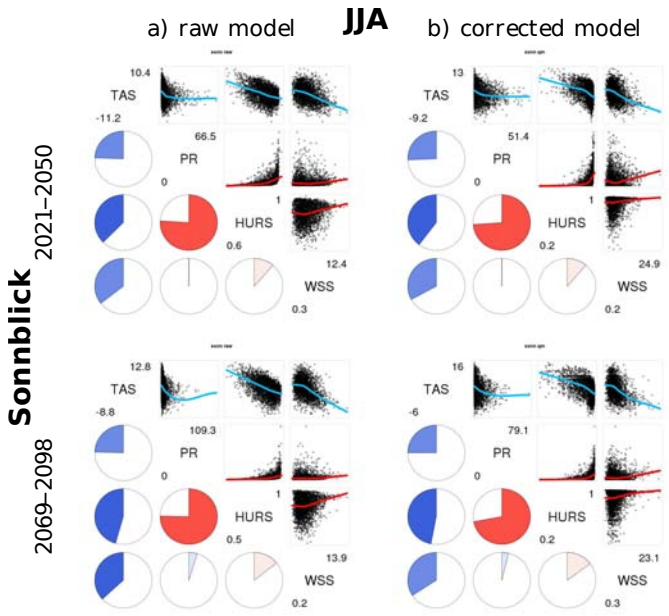


Figure B.18: Correlation matrices for **Sonnblick** for summer (JJA) of the periods a) 2021 to 2050 and b) 2069 to 2098. Same layout as in Figure B.9

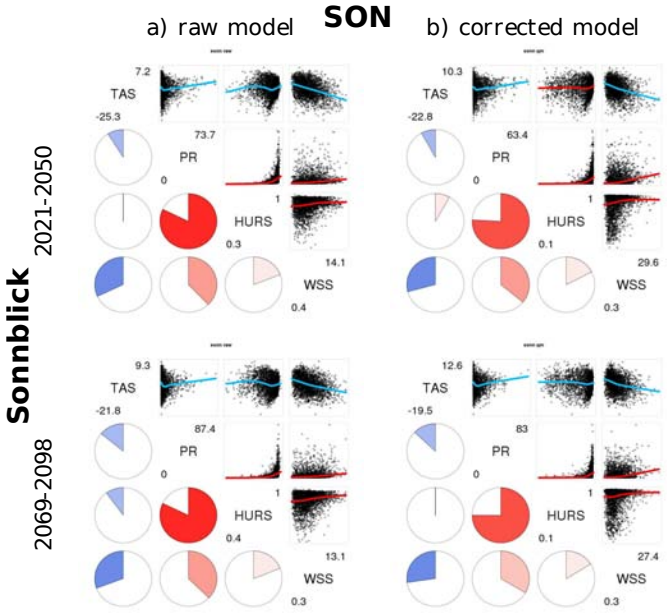


Figure B.19: Correlation matrices for **Sonnblick** for autumn (SON) of the periods a) 2021 to 2050 and b) 2069 to 2098. Same layout as in Figure B.9

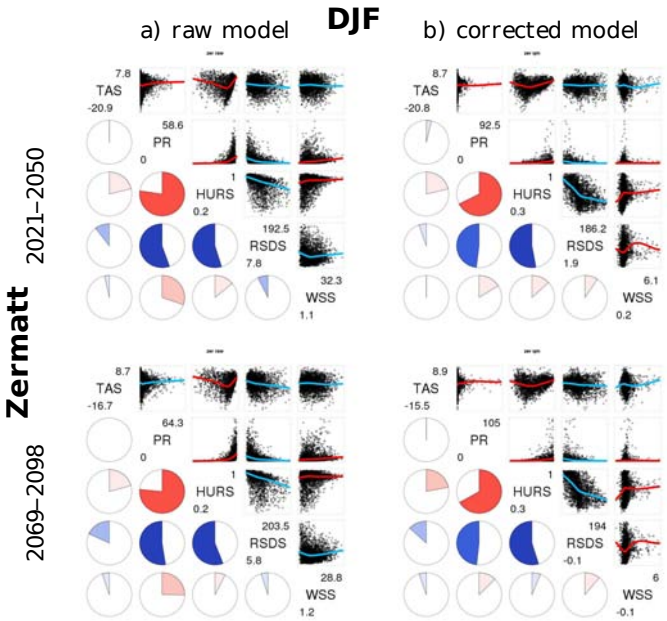


Figure B.20: Correlation matrices for **Zermatt** for winter (DJF) of the periods a) 2021 to 2050 and b) 2069 to 2098. Same layout as in Figure B.9 but including global radiation (rsds).

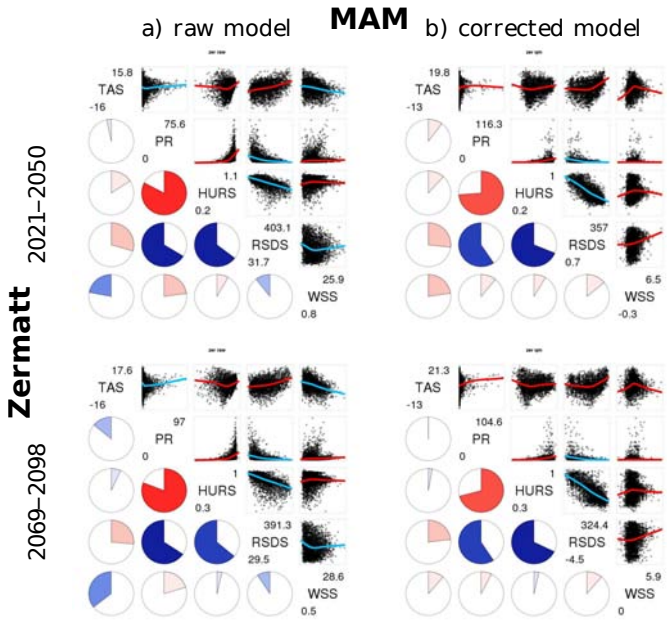


Figure B.21: Correlation matrices for **Zermatt** for spring (MAM) of the periods a) 2021 to 2050 and b) 2069 to 2098. Same layout as in Figure B.9 but including global radiation (rsds).

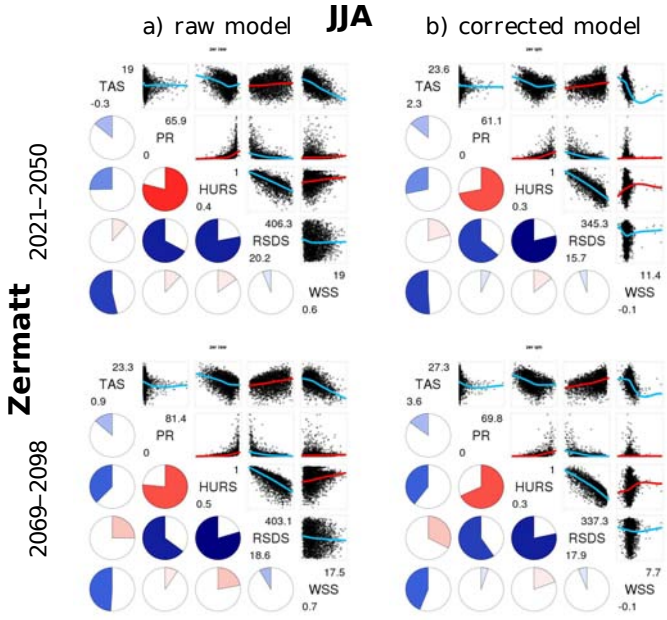


Figure B.22: Correlation matrices for **Zermatt** for summer (JJA) of the periods a) 2021 to 2050 and b) 2069 to 2098. Same layout as in Figure B.9 but including global radiation (rsds).

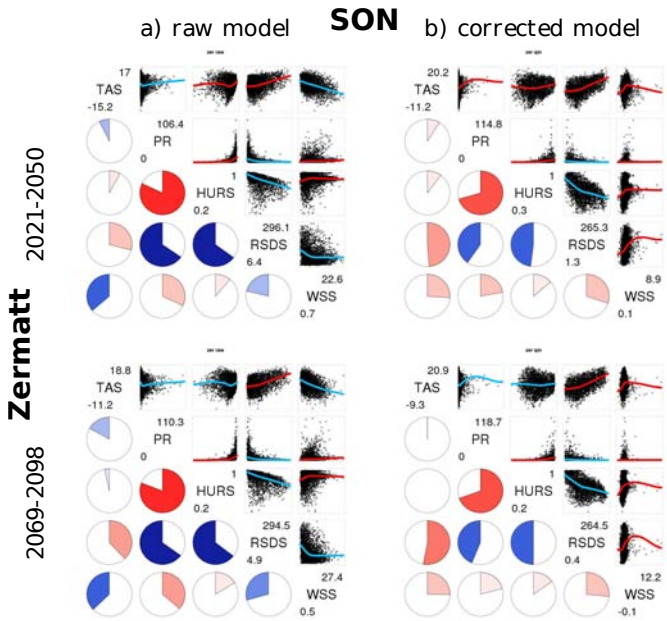


Figure B.23: Correlation matrices for **Zermatt** for autumn (SON) of the periods a) 2021 to 2050 and b) 2069 to 2098. Same layout as in Figure B.9 but including global radiation (rsds).

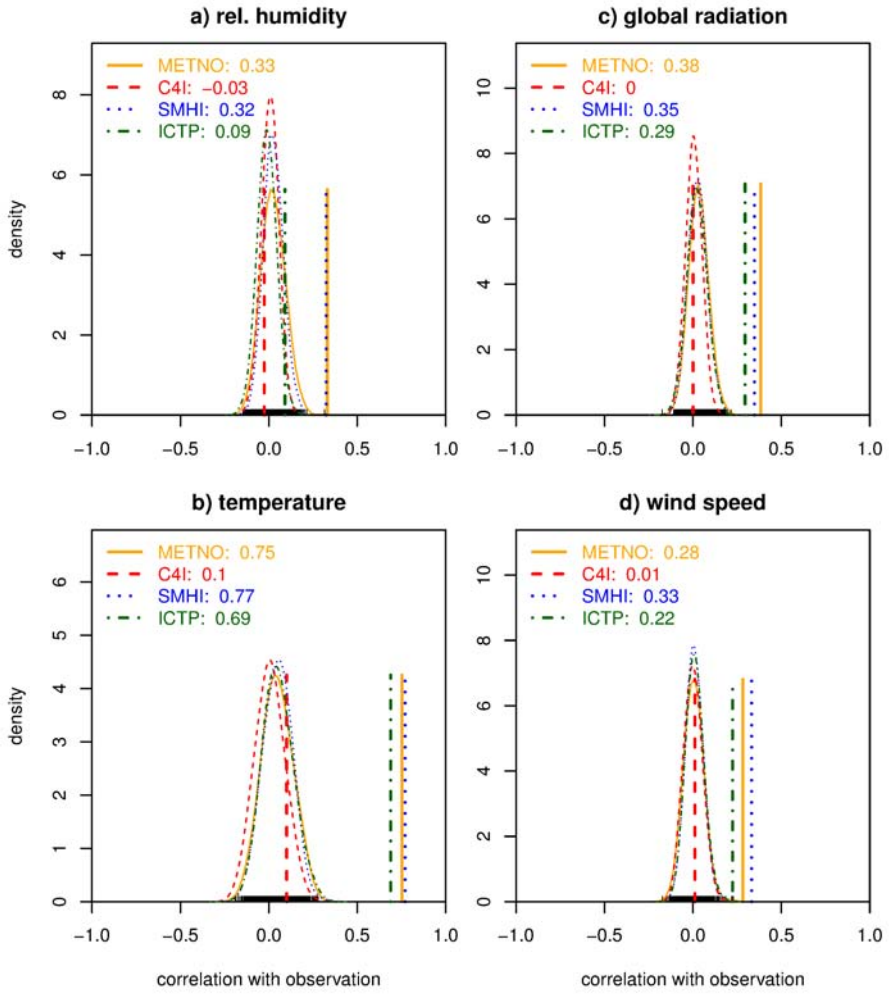


Figure B.24: Pearson correlation coefficient for 4 RCMs for a) rel. humidity, b) air temperature, c) global radiation, d) wind speed. The coefficient is shown beside the RCM name and as vertical line. Also is shown the distribution of sampled RCM correlations. A correlation coefficient in the tail of the distribution denotes a relevant correlation.

B.2 FIGURES: INDIRECT VALIDATION OF QM

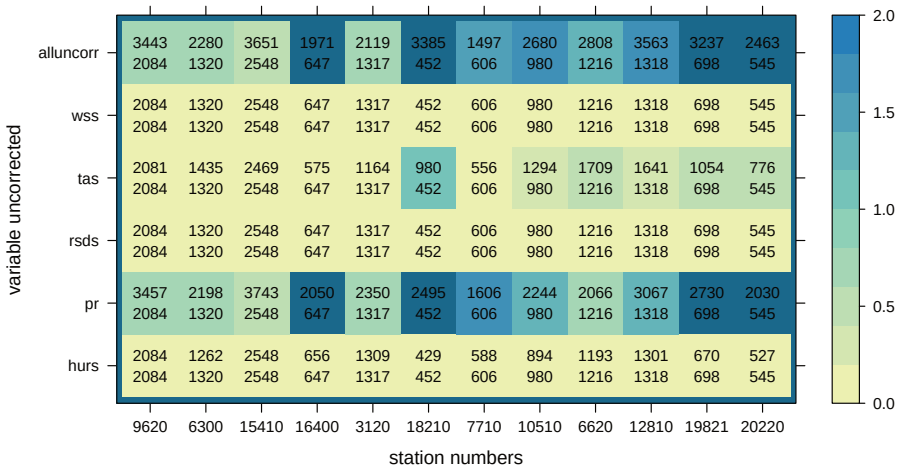


Figure B.25: Relative (colorkey) difference in number of snow days overlayed by absolut values (upper value) and absolute reference value (lower value), used C4I-RCA3 (driven by HadCM3Q16), accumulated for November to March in 1970 to 2000.

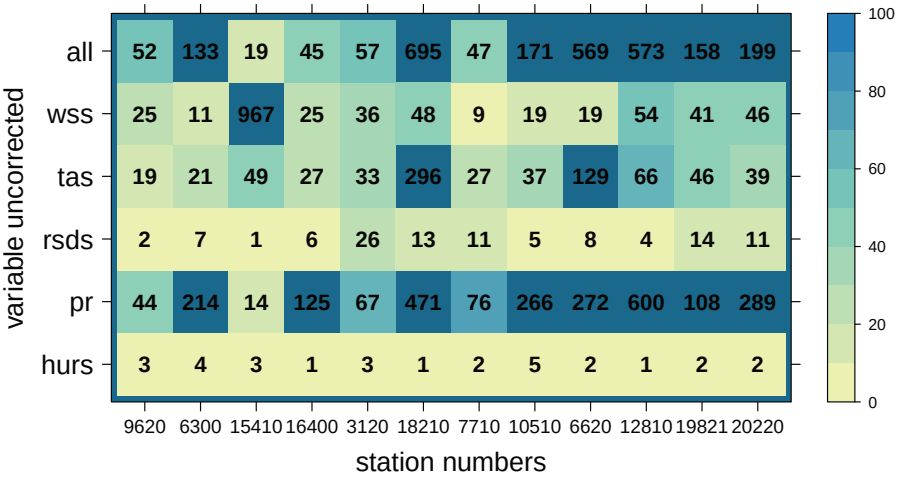


Figure B.26: Median of relative difference in SWE compared to all climate variables corrected. RCM: C4I-RCA3 (driven by HadCM3Q16), period: November to March, 1970 to 2000

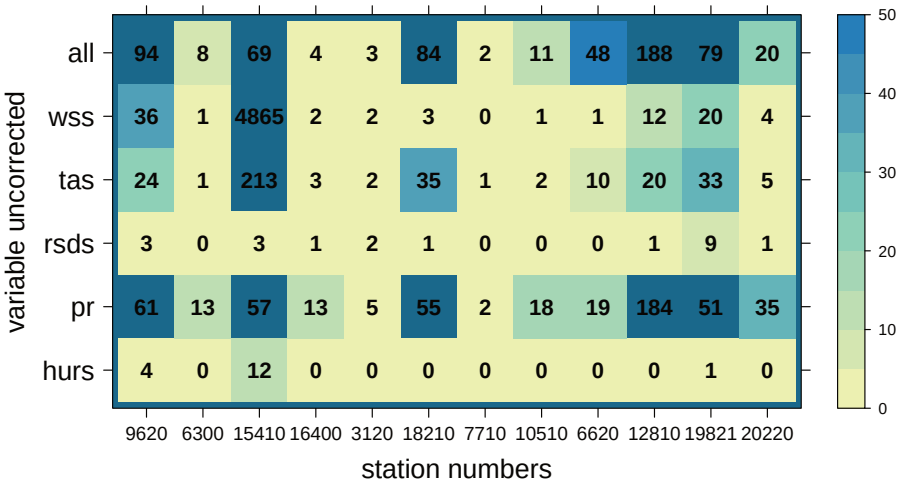
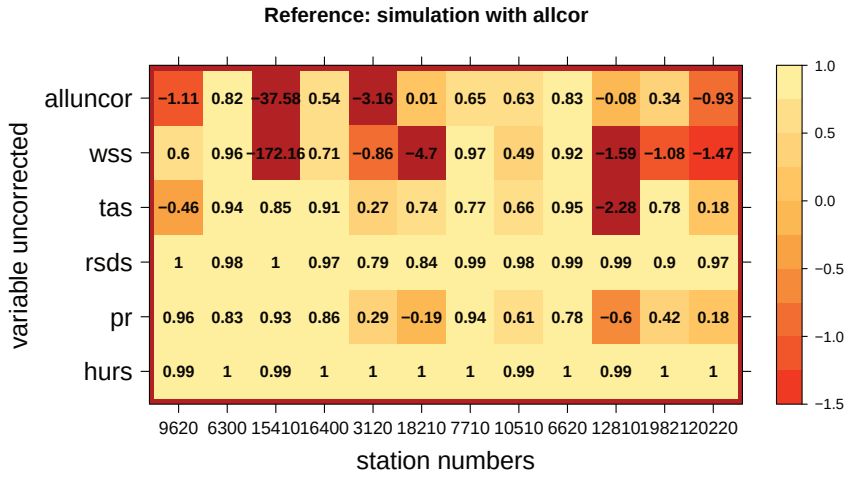
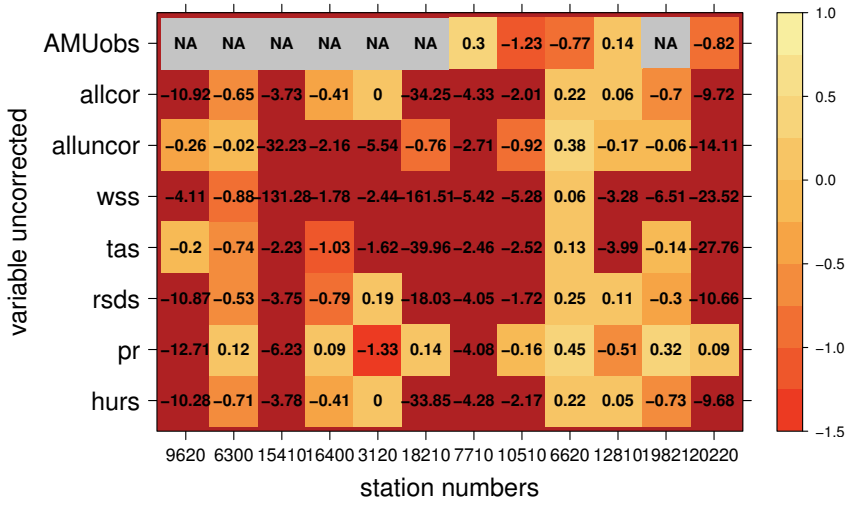


Figure B.27: As in Figure B.26 but showing the median of absolute difference in SWE [mm]

a)



b)



c)

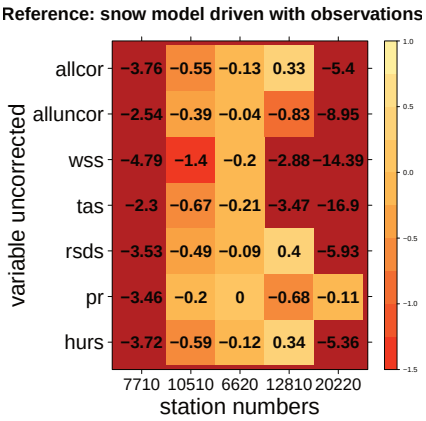


Figure B.28: NSE of snow depth with reference a) simulation with all corrected variables, b) observations, c) simulation driven with observations, for November to March 1982 to 2000, C4I-RCA3 (ERA40 driven)

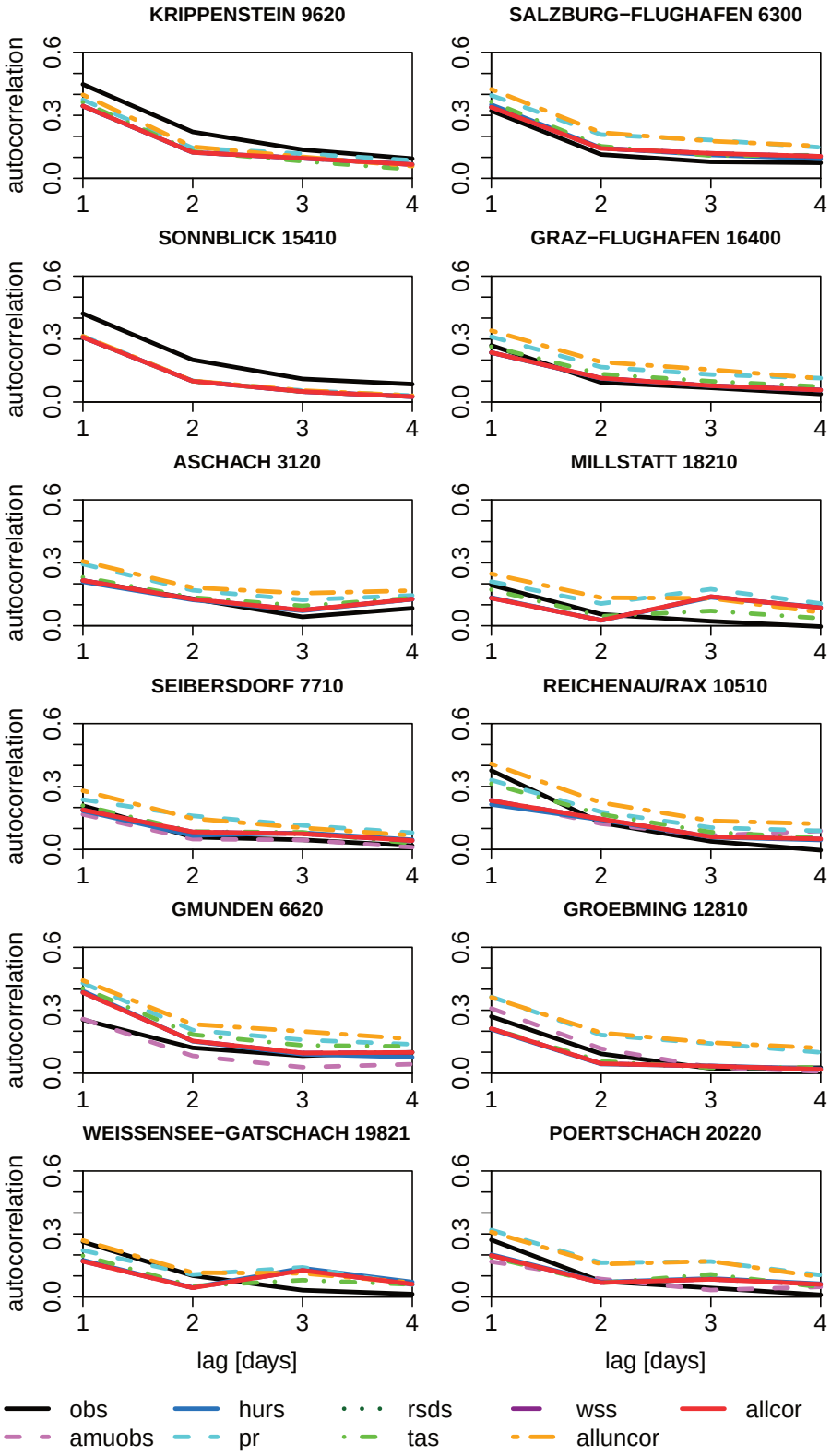


Figure B.29: Autocorrelation of snow fall inclusive simulation driven by observations, for November to March, 1981 to 2000, C4I-RCA3-ERA40

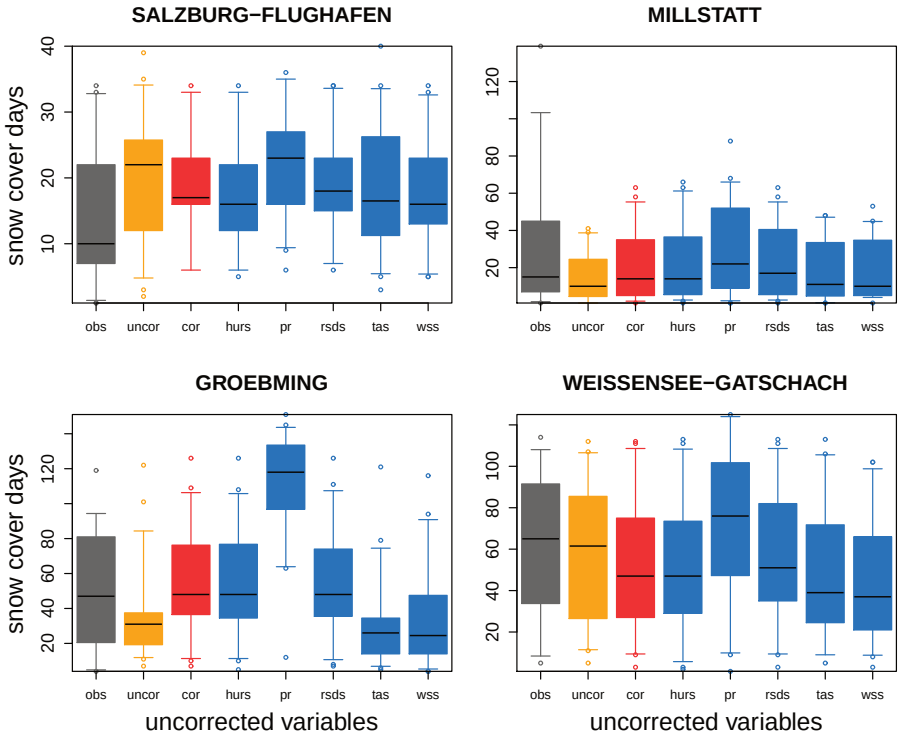


Figure B.30: Maximum consecutive snow cover (≥10 cm) length per winter in 1970 to 2000, METNO-HIRHAM-ERA40

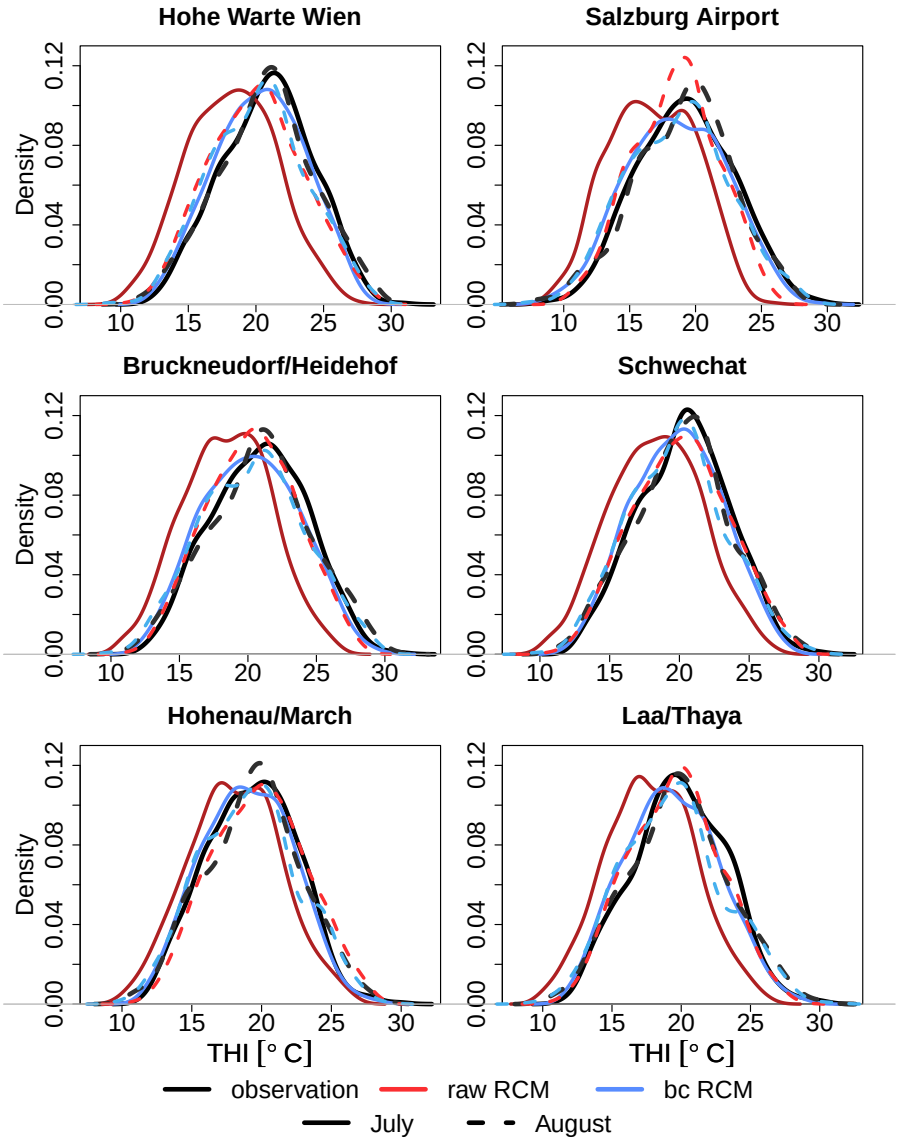


Figure B.31: Density distribution of THI for six stations in Austria for observation (black), raw RCM (red), bias corrected RCM (blue); for July (solid) and August (dashed). RCM shown: C4I-RCA3-ERA40; period: 1981 to 2000

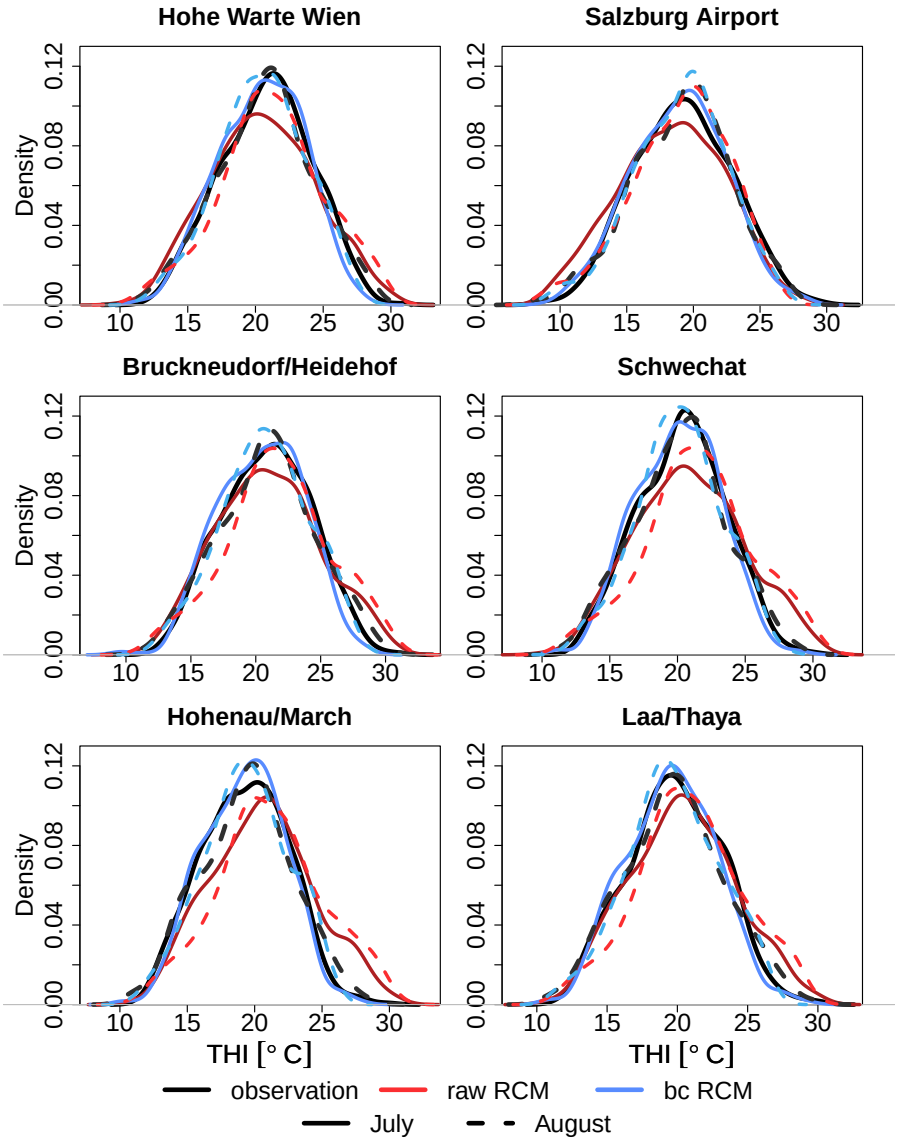


Figure B.32: Same layout as in Figure B.31 but for METNO-HIRHAM

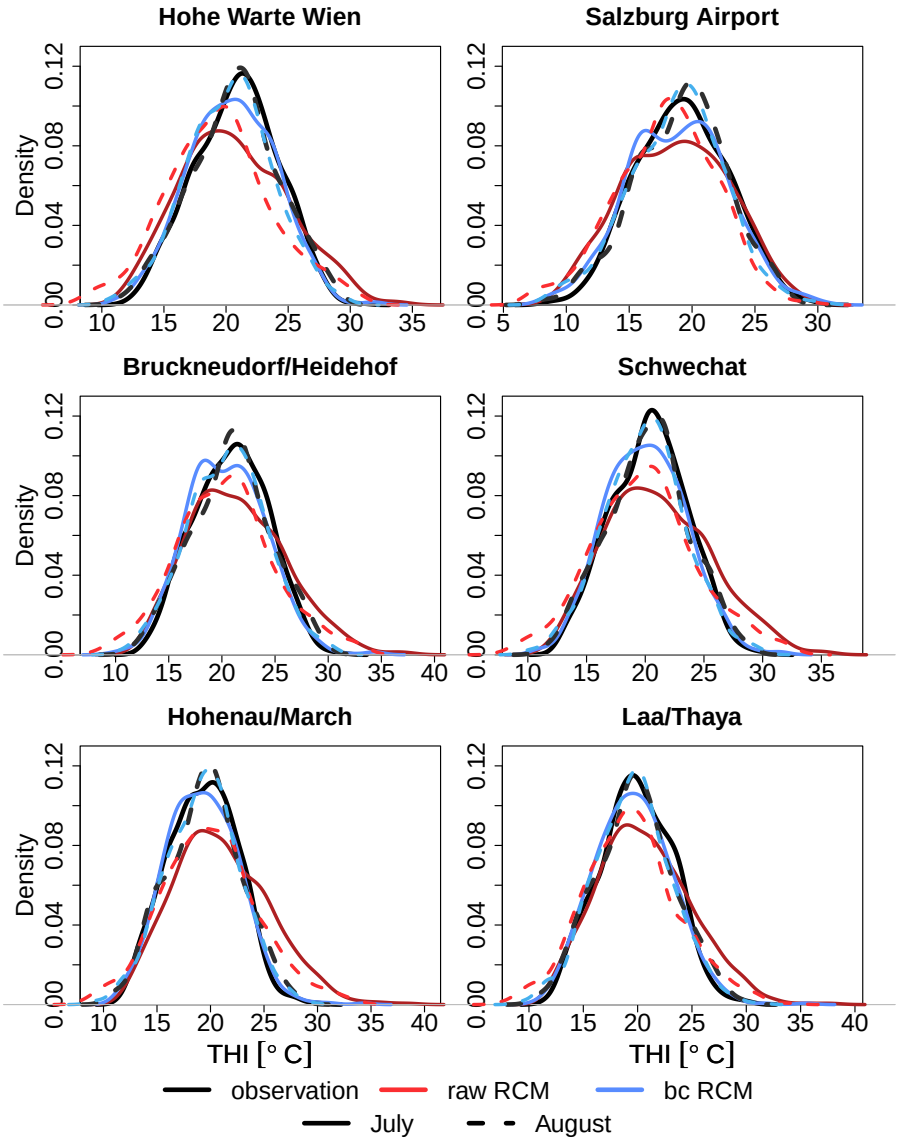


Figure B.33: Same layout as in Figure B.31 but for ICTP-RegCM3

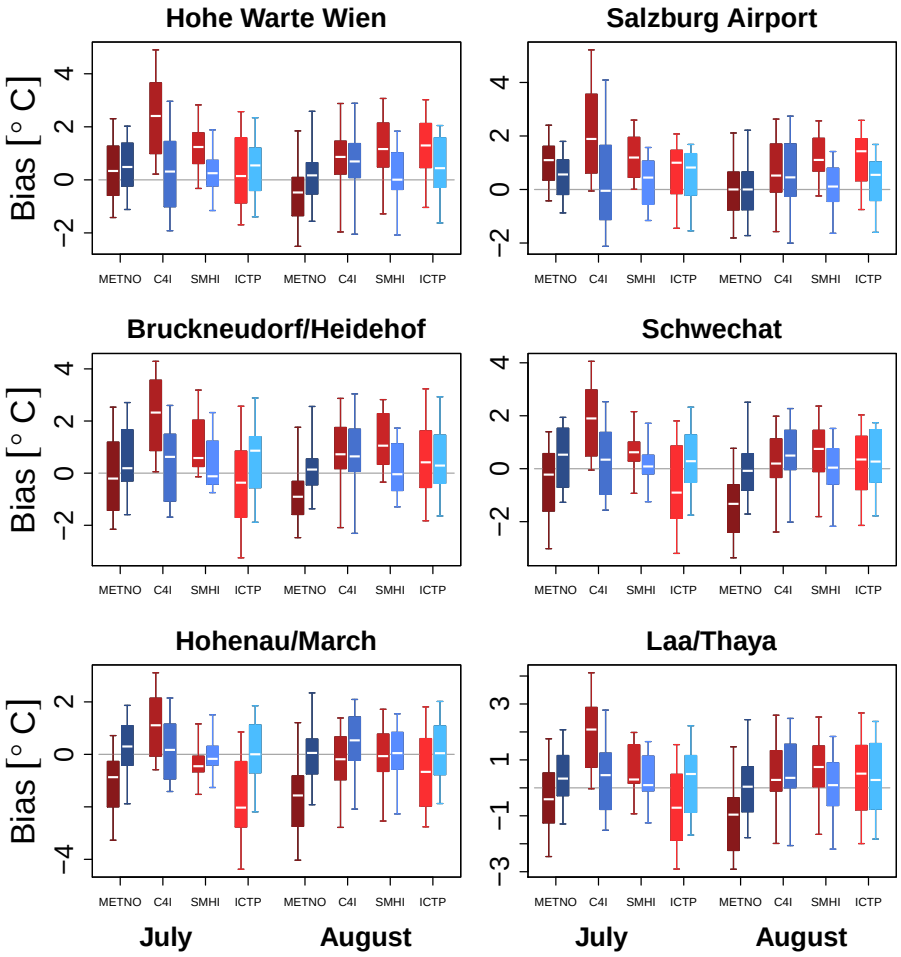


Figure B.34: Box-whisker plot of bias in THI for six stations in Austria for raw RCMs (red), bias corrected RCMs (blue); for July (left) and August (right) for period 1981 to 2000.

C

APPENDIX TABLES

Table C.1: Spearman correlation coefficient matrix of Zermatt for the period 1981 to 2010 of ICTP-RegCM3 described in Section 7.2; including temperature (tas), precipitation amount (pr), relative humiditiy (hurs), global radiation (rsds), and wind speed (wss).

observations					
	tas	pr	hurs	rsds	wss
tas	1.00				
pr	-0.29	1.00			
hurs	-0.60	0.67	1.00		
rsds	0.27	-0.60	-0.61	1.00	
wss	-0.09	-0.03	-0.04	0.27	1.00
raw RCM					
	tas	pr	hurs	rsds	wss
tas	1.00				
pr	-0.24	1.00			
hurs	-0.26	0.80	1.00		
rsds	0.12	-0.67	-0.74	1.00	
wss	-0.58	0.20	0.18	-0.05	1.00
error-corrected RCM					
	tas	pr	hurs	rsds	wss
tas	1.00				
pr	-0.21	1.00			
hurs	-0.27	0.73	1.00		
rsds	0.20	-0.64	-0.74	1.00	
wss	-0.56	0.14	0.17	-0.05	1.00

Table C.2: Stations in Austria aggregated to RCM grid cells for direct validation study of inter-variable relations in Chapter 7.

grid-cell	stat. ID	name	lon East	lat North	height
1 Villach region	18210	Millstatt	13.5736	46.8083	719
	18211	Millstatt-Kleindombra	13.5833	46.8000	754
	18225	Spittal/Drau	13.4872	46.7906	542
	18305	Radenthein	13.6981	46.7981	685
	18306	Radenthein	13.6983	46.8003	688
	19911	Hermagor	13.4922	46.6097	562
	20001	Bad Bleiberg	13.6889	46.6250	907
	20002	Bad Bleiberg	13.6844	46.6253	909
	20005	Kreuth	13.6667	46.6167	835
	20010	Fresach	13.6950	46.7144	748
	20011	Fresach	13.6853	46.7153	681
	20020	Villacher Alpe	13.6733	46.6036	2140
	20021	Villacher Alpe	13.6733	46.6036	2164
2 Kitzbühel Alps	8900	Kreit	12.0756	47.5411	625
	9000	Kirchbichl	12.0867	47.5156	498
	9010	Kufstein	12.1639	47.5742	492
	9016	Kufstein	12.1628	47.5753	490
	9011	Oberndorf/Ebbs	12.2000	47.616	474
	9018	Soellmt	12.2128	47.5133	656
	12201	Kitzbuehel	12.3583	47.4597	791
	12202	Kitzbuehel	12.3942	47.4472	744
	12210	Hahnenkamm	12.3667	47.4333	1665
	12212	Hahnenkamm-Ehrenbachhoehe	12.3667	47.4167	1760
	12215	Hahnenkamm-Ehrenbachhoehe	12.3619	47.4192	1790
3 Hohe Tauern Nationalpark	15210	Virgen	12.4558	47.0028	1212
	15300	Enzingerboden	12.6333	47.1667	1480
	15310	Mooserboden	12.7167	47.1500	2036
	15321	Rudolfshuette	12.6256	47.1339	2304
	15322	Rudulfshuette/ Weisssee	12.6333	47.1333	2315

Continued on next page

Table C.2 – *continued from previous page*

grid-cell	stat. ID	name	lon East	lat North	height
	15330	Schneiderau	12.6000	47.2000	1041
	15340	Kals	12.6500	47.0000	1347
	15343	Kals	12.6461	47.0033	1338
	15344	Kals	12.6464	47.0047	1352
	15341	Kals-Grossdorf	12.6333	47.0000	1350
4 Semmering	10400	Muerzzuschlag	15.6858	47.6031	758
	10401	Muerzzuschlag	15.6728	47.6022	705
	10415	Semmering	15.8283	47.6331	985
	10412	Rax/Seilbahn Bergstation	15.7786	47.7175	1547
	10416	Semmering/ Suedbahn-Hotel	15.8333	47.6333	1000
	10505	Semmering-Mariaschutz	15.8500	47.6333	855
	10508	Breitenstein	15.8361	47.6636	825
	10511	Reichenau/Rax	15.8500	47.7000	485
	10510	Reichenau/Rax	15.8369	47.6997	488
	10550	Hirschenkogel	15.8333	47.6233	1318
5 Basin of Klagenfurt	18620	St.Veit/Glan	14.3606	46.7647	475
	18621	St.Veit/Glan	14.3458	46.7597	475
	18622	St.Veit/Glan	14.3856	46.7392	463
	20211	Klagenfurt-Flughafen	14.3333	46.6500	447
	20212	Klagenfurt-Flughafen	14.3183	46.6483	450
	20219	Poertschach	14.1667	46.6333	441
	20220	Poertschach	14.1664	46.6311	452
	20270	Feldkirchen	14.0969	46.7219	546
	20305	Hoerzendorf/ St.Veit	14.3464	46.7322	492
6 Pongau	12501	Bischofshofen	13.2167	47.4167	556
	12503	Bischofshofen	13.2239	47.4081	550
	12506	Bischofshofen-Buchberg	13.2167	47.4000	733

Continued on next page

Table C.2 – *continued from previous page*

grid-cell	stat. ID	name	lon East	lat North	height
	12505	St.Johann im Pongau	13.1836	47.3153	634
	12530	St.Johann im Pongau	13.1836	47.3153	634
	12510	Schwarzach	13.1333	47.3167	600
	12513	St.Veit im Pongau	13.1553	47.3469	750
	12520	Mitterberg	13.1333	47.4167	1503
	12620	Wagrain	13.3050	47.3336	880
7 Vienna	5880	Gumpoldskirchen	16.2833	48.0333	222
	5881	Gumpoldskirchen	16.2822	48.0403	219
	5916	Wien-Unterlaa	16.4228	48.1275	200
	5917	Wien/Unterlaa	16.4192	48.1247	201
	5925	Wien-Innerestadt	16.3669	48.1983	177
	5970	Gross-Enzersdorf	16.5667	48.2000	153
	5972	Gross-Enzersdorf	16.5592	48.1997	154
	5980	Muenchendorf	16.4000	48.0333	184
	5990	Schwechat	16.5708	48.1108	184

Table C.3: Observation stations and their coordinates, used for indirect validation in Chapter 8.

station name	**id**	**lon East**	**lat North**	**height** [m]
Krippenstein	9620	13.7000	47.5167	2050
Sonnblick	15410	12.9581	47.0544	3105
Graz-Flughafen	16400	15.4478	46.9947	337
Aschach	3120	14.0167	48.3833	282
Millstatt	18210	13.5736	46.8083	719
Seibersdorf	7710	16.5050	47.9764	185
Reichenau/Rax	10510	15.8369	47.6997	488
Gmunden	6620	13.8075	47.9147	424
Groebming	12810	13.9036	47.4456	766
Weissensee-Gatschach	19821	13.2867	46.7192	945
Poertschach	20220	14.1664	46.6311	452
Salzburg-Flughafen	6300	13.0017	47.8014	430
Wien-Hohewarte	5906	16.3564	48.2486	198
Bruckneudorf/Heidehof	6101	16.8467	48.0147	167
Schwechat	5990	16.5708	48.1108	184
Hohenau/March	2600	16.9044	48.6172	155
Laa/Thaya	2400	16.3853	48.7261	187

LIST OF FIGURES

LIST OF TABLES

BIBLIOGRAPHY

Anderson, E. (1976). A point energy and mass balance model of a snow cover. Technical Report 19.

Ashfaq, M., Bowling, L. C., Cherkauer, K., Pal, J. S., and Diffenbaugh, N. S. (2010). Influence of climate model biases and daily-scale temperature and precipitation events on hydrological impacts assessment: A case study of the United States. *J. Geophys. Res.: Atmospheres*, 115(D14).

Auer, I., Böhm, R., and Schöner, W. (2001). *Austrian long-term climate 1767–2000*. Zentralanst. für Meteorologie u. Geodynamik.

Awan, N. K., Truhetz, H., and Gobiet, A. (2011). Parameterization-Induced Error Characteristics of MM5 and WRF Operated in Climate Mode over the Alpine Region: An Ensemble-Based Analysis. *J. Climate*, 24:3107–3123.

Beersma, J. J. and Buishand, T. A. (2003). Multi-site simulation of daily precipitation and temperature conditional on the atmospheric circulation. *Clim. Res.*, 25:121–133.

Bellprat, O., Kotlarski, S., Lüthi, D., and Schär, C. (2013). Physical constraints for temperature biases in climate models. *Geophys. Res. Lett.*, 40(15):4042–4047.

Berg, A. A., Famiglietti, J. S., Walker, J. P., and Houser, P. R. (2003). Impact of bias correction to reanalysis products on simulations of North American soil moisture and hydrological fluxes. *J. Geophys. Res. D: Atmos.*, 108(D16).

Berg, P., Haerter, J. O., Thejll, P., Piani, C., Hagemann, S., and Christensen, J. H. (2009). Seasonal characteristics of the relationship between daily precipitation intensity and surface temperature. *J. Geophys. Res.: Atmospheres*, 114(D18).

Bjerknes, V. (1904). Das Problem der Wettervorhersage, betrachtet vom Standpunkte der Mechanik und der Physik. *Meteorol. Z.*, pages 1–7.

Boé, J., Terray, L., Habets, F., and Martin, E. (2007). Statistical and dynamical downscaling of the Seine basin climate for hydro-meteorological studies. *Int. J. Climatol.*, 27(12):1643–1655.

Bordoy, R. and Burlando, P. (2013). Bias Correction of Regional Climate Model Simulations in a Region of Complex Orography. *J. Appl. Meteor. Climatol.*, 52:82–101.

Bowman, A. and Azzalini, A. (1997). *Applied Smoothing Techniques for Data Analysis: The Kernel Approach with S-Plus Illustrations: The Kernel Approach with S-Plus Illustrations*. OUP Oxford.

Brandsma, T. and Buishand, T. (1998). Simulation of extreme precipitation in the Rhine basin by nearest-neighbour resampling. *Hydrol. Earth. Syst. Sc.*, 2:195–209.

Brier, W. and Panofsky, H. (1968). *Some applications of statistics to meteorology*. Mineral Industries Extension Services, School of Mineral Industries, Pennsylvania State College.

Busuioc, A., Tomozeiu, R., and Cacciamani, C. (2008). Statistical downscaling model based on canonical correlation analysis for winter extreme precipitation events in the Emilia-Romagna region. *Int. J. Climatol.*, 28(4):449–464.

C4I (2008). D2B.9: Report on the RCA3 dynamical downscaling of an ECHAM5 A1B or A2 GCM simulation for the period 1950-2050 and evaluation of regional climate model simulations with HOAPS satellite data. deliverable report D2B.9, C4I, ENSEMBLES project.

Castro, C. L., Pielke Sr., R. A., and Leoncini, G. (2005). Dynamical downscaling: Assessment of value retained and added using the Regional Atmospheric Modeling System (RAMS). *J. Geophys. Res.*, 110(D5):1–21.

Caussinus, H. and Mestre, O. (2004). Detection and correction of artificial shifts in climate series. *J. R. Stat. Soc. Ser. C Appl. Stat.*, 53(3):405–425.

Chen, C., Haerter, J. O., Hagemann, S., and Piani, C. (2011). On the contribution of statistical bias correction to the uncertainty in the projected hydrological cycle. *Geophysical Research Letters*, 38(20).

Chen, J., Brissette, F., and Leconte, R. (2012). Coupling statistical and dynamical methods for spatial downscaling of precipitation. *Clim. Change*, 114(3-4):509–526.

Chen, S.-H. and Sun, W.-Y. (2002). A One-dimensional Time Dependent Cloud Model. *J. Meteorol. Soc. Jpn.*, 80(1):99–118.

Christensen, J. H., Boberg, F., Christensen, O. B., and Lucas-Picher, P. (2008). On the need for bias correction of regional climate change projections of temperature and precipitation. *Geophys. Res. Lett.*, 35(20).

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. *Clim. Change*, 81(S1):7–30.

Christensen, O., Drews, M., Christensen, J., Dethloff, K., Ketelsen, K., Hebestadt, I., and Rinke, A. (2007). *The HIRHAM Regional Climate Model. Version 5 (beta)*. Denmark. Danish Meteorological Institute. Technical Report. Danish Climate Centre, Danish Meteorological Institute.

Collins, M., Tett, S., and Cooper, C. (2001). The internal climate variability of HadCM3, a version of the Hadley Centre coupled model without flux adjustments. *Clim. Dyn.*, 17:61–81.

Corripio, J. (2003). Vectorial algebra algorithms for calculating terrain parameters from DEMs and solar radiation modelling in mountainous terrain. *Int. J. Geogr. Inf. Sci.*, 17(1):1–23.

Cubasch, U., von Storch, H., Waszkewitz, J., and Zorita, E. (1996). Estimates of climate change in Southern Europe derived from dynamical climate model output. *Clim. Res.*, 7(2):129–149.

Davies, H. C. (1976). A lateral boundary formulation for multi-level prediction models. *Q. J. R. Meteorol. Soc.*, 102:405–418.

Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap Methods and Their Applications*. Cambridge University Press, Cambridge.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F. (2011). The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.*, 137(656):553–597.

Della-Marta, P. M. and Wanner, H. (2006). A method of homogenizing the extremes and mean of daily temperature measurements. *J. Climate*, 19(17):4179–4197.

Denis, B., Côté, J., and Laprise, R. (2002). Spectral Decomposition of Two-Dimensional Atmospheric Fields on Limited-Area Domains Using the Discrete Cosine Transform (DCT). *Mon. Wea. Rev.*, 130:1812–1829.

Déqué, M. (2007). Frequency of precipitation and temperature extremes over France in an anthropogenic scenario: Model results and statistical correction according to observed values. *Glob. Planet. Change*, 57(1–2):16–26.

Déqué, M., Rowell, D. P., Lüthi, D., Giorgi, F., Christensen, J. H., Rockel, B., Jacob, D., Kjellström, E., Castro, M., and Hurk, B. (2007).

An intercomparison of regional climate simulations for Europe: assessing uncertainties in model projections. *Clim. Change*, 81(S1):53–70.

Déqué, M., Somot, S., Sanchez-Gomez, E., Goodess, C. M., Jacob, D., Lenderink, G., and Christensen, O. B. (2011). The spread amongst ENSEMBLES regional scenarios: regional climate models, driving general circulation models and interannual variability. *Clim. Dyn.*

Diaconescu, E. P., Laprise, R., and Sushama, L. (2007). The impact of lateral boundary data errors on the simulated climate of a nested regional climate model. *Clim. Dyn.*, 28:333–350.

Dickinson, R. E., Errico, R. M., Giorgi, F., and Bates, G. T. (1989). A regional climate model for the western United States. *Clim. Change*, 15(3):383–422.

Dirmeyer, P. A., DelSole, T., and Zhao, M. (2011). Limits to the Impact of Empirical Correction on Simulation of the Water Cycle. *J. Hydrometeor*, 12:650–662.

Dobler, A. and Ahrens, B. (2008). Precipitation by a regional climate model and bias correction in Europe and South Asia. *Meteorol. Z.*, 17(4):499–509.

Dolman, J. and Gregory, D. (1992). The parametrization of rainfall interception in GCMs. *Q. J. R. Meteorol. Soc.*, 118:455–467.

Dosio, A. and Paruolo, P. (2011). Bias correction of the ENSEMBLES high-resolution climate change projections for use by impact models: Evaluation on the present climate. *J. Geophys. Res.*, 116(D16):1–22.

Dudhia, J. (1989). Numerical Study of Convection Observed during the Winter Monsoon Experiment Using a Mesoscale Two-Dimensional Mode. *J. Atmos. Sci.*, 46:3077–3107.

Durman, C. F., Gregory, J. M., Hassell, D. C., Jones, R. G., and Murphy, J. M. (2001). A comparison of extreme European daily precipitation simulated by a global and a regional climate model for present and future climates. *Q. J. R. Meteorol. Soc.*, 127(573):1005–1015.

Eden, J., Widmann, M., Grawe, D., and Rast, S. (2012). Skill, Correction, and Downscaling of GCM-Simulated Precipitation. *J. Clim*, 25(11):3970–3984.

Edwards, P. N. (2010). *A Vast Machine*. MIT Press.

Engen-Skaugen, T. (2007). Refinement of dynamically downscaled precipitation and temperature scenarios. *Clim. Change*, 84(3-4):365–382.

Evans, J. (2003). Improving the characteristics of streamflow modeled by regional climate models. *J. Hydrol.*, 284(1):211–227.

Feser, F. (2006). Enhanced Detectability of Added Value in Limited-Area Model Results Separated into Different Spatial Scales. *Mon. Wea. Rev.*, 134:2180–2190.

Finger, D., Heinrich, G., Gobiet, A., and Bauder, A. (2012). Assessing future water resources and its uncertainty in a glaciated Alpine catchment and its subsequent effects on hydropower operations during the 21st century. *Water Resour. Res.*, 48:20 pp.

Finley, J. (1884). Tornado prediction. *Amer. Meteor. J.*, 1:85–88.

Fischer, A., Wilcke, R. A. I., Widmann, M., Wibig, J., and Eden, J. (2013). Report on downscaling inventory, WG 3–5. deliverable report, COST Action ES1102, VALUE.

Fowler, H. J., Blenkinsop, S., and Tebaldi, C. (2007). Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling. *Int. J. Climatol.*, 27(12):1547–1578.

Fox-Rabinovitz, M., Côté, J., Dugas, B., Déqué, M., and McGregor, J. L. (2006). Variable resolution general circulation models: Stretched-grid model intercomparison project (SGMIP). *J. Geophys. Res.: Atmospheres*, 111(D16):pp. 21.

Fox-Rabinovitz, M. S., Takacs, L. L., Govindaraju, R. C., and Suarez, M. J. (2001). A Variable-Resolution Stretched-Grid General Circulation Model: Regional Climate Simulation. *Mon. Wea. Rev.*, 129:453–469.

Frei, C., Christensen, Jens, H., Déqué, M., Jacob, D., Jones, Richard, G., and Vidale, Pier, L. (2003). Daily precipitation statistics in regional climate models: Evaluation and intercomparison for the European Alps. *J. Geophys. Res.*, 108(D3):1–19.

Frei, C. and Schär, C. (1998). A precipitation climatology of the Alps from high-resolution rain-gauge observations. *Int. J. Climatol.*, 18(8):873–900.

Friendly, M. (2002). Corrgrams: Exploratory displays for correlation matrices. *The American Statistician*, 56(4):316–324.

Furevik, T., Bentsen, M., Drange, H., Kindem, I., Kvamstø, N., and Sorteberg, A. (2003). Description and evaluation of the bergen climate model: ARPEGE coupled with MICOM. *Clim. Dyn.*, 21(1):27–51.

Gilliam, R. and Pleim, J. (2010). Performance assessment of new land surface and planetary boundary layer physics in the WRF-ARW. *J. Appl. Meteorol. Clim.*, 49(4):760–774.

Gilliam, R. C., Hogrefe, C., and Rao, S. (2006). New methods for evaluating meteorological models used in air quality applications. *Atmos. Environ.*, 40(26):5073–5086.

Giorgi, F. (1990). Simulation of regional climate using a limited area model nested in a general circulation model. *J. Climate*, 3:941–963.

Giorgi, F. and Bates, G. T. (1989). The Climatological Skill of a Regional Model over Complex Terrain. *Mon. Wea. Rev.*, 117:2325–2347.

Giorgi, F. and Coppola, E. (2010). Does the model regional bias affect the projected regional climate change? An analysis of global model projections. *Clim. Change*, 100(3-4):787–795.

Giorgi, F., Hostetler, S. W., and Brodeur, C. S. (1994). Analysis of the surface hydrology in a regional climate model. *Q. J. R. Meteor. Soc.*, 120(515):161–183.

Giorgi, F., Jones, C., and Asrar, G. (2009). Addressing climate information needs at the regional level: The CORDEX framework. *WMO Bulletin*, 58(3):175–183.

Giorgi, F. and Mearns, L. O. (1999). Introduction to special section: Regional Climate Modeling Revisited. *J. Geophys. Res.: Atmospheres*, 104(D6):6335–6352.

Gobiet, A., Suklitsch, M., and Heinrich, G. (2013). How can model error characteristics be utilised to improve climate projections? In *EGU General Assembly Conference Abstracts*, volume 15 of *EGU General Assembly Conference Abstracts*, page 8961.

Goodess, C., Jacob, D., Déqué, M., Guttiérrez, J., Huth, R., Kendon, E., Leckebusch, G. C., Lorenz, P., and Pavan, V. (2009). Downscaling methods, data and tools for input to impacts assessments. In van der Linden, P. and Mitchell, J. F., editors, *ENSEMBLES: Climate Change and its Impacts: Summary of research and results from the ENSEMBLES project*, pages 59–78. Met Office Hadley Centre, FitzRoy Road, Exeter EX1 3PB, UK.

Graham, L. P., Andréasson, J., and Carlsson, B. (2007). Assessing climate change impacts on hydrology from an ensemble of regional climate models, model scales and linking methods – a case study on the Lule River basin. *Clim. Change*, 81(1):293–307.

Grasso, L. (2000). The Difference Between Grid Spacing and Resolution and Their Application to Numerical Modeling. *Bull. Amer. Meteor. Soc.*, 81:579–580.

Greuell, W., Knap, W. H., and Smeets, P. C. (1997). Elevational changes in meteorological variables along a midlatitude glacier during summer. *J. Geophys. Res.*, 102(D22):25941–25954.

Grotch, S. L. and MacCracken, M. C. (1991). The Use of General Circulation Models to Predict Regional Climatic Change. *J. Climate*, 4:286–303.

Guenni, L., Rose, C., Hogarth, W., Braddock, R., and Charles-Edwards, D. (1990). Seasonal changes in interrelationships between climatic variables. *Agr. Forest. Meteorol.*, 53:45–58.

Guo, J., Urbonas, B., and Stewart, K. (2001). Rain Catch under Wind and Vegetal Cover Effects. *J. Hydrol. Eng.*, 6(1):29–33.

Gutiérrez, J. M., San-Martín, D., Brands, S., Manzanas, R., and Herrera, S. (2013). Reassessing Statistical Downscaling Techniques for Their Robust Application under Climate Change Conditions. *J. Climate*, 26:171–188.

Gutjahr, O. and Heinemann, G. (2013). Comparing precipitation bias correction methods for high-resolution regional climate simulations using COSMO-CLM. *Theor. Appl. Climatol.*, 114(3-4):511–529.

Gutowski, Jr., W. J., Decker, S. G., Donavon, R. A., Pan, Z., Arritt, R. W., and Takle, E. S. (2003). Temporal Spatial Scales of Observed and Simulated Precipitation in Central U.S. Climate. *J. Climate*, 16:3841–3847.

Haddeland, I., Heinke, J., Voß, F., Eisner, S., Chen, C., Hagemann, S., and Ludwig, F. (2012). Effects of climate model radiation, humidity and wind estimates on hydrological simulations. *Hydrol. Earth Syst. Sci.*, 16(2):305–318.

Haerter, J. O., Hagemann, S., Moseley, C., and Piani, C. (2010). Climate model bias correction and the role of timescales. *Hydrol. Earth Syst. Sci. Discussions*, 7(5):7863–7898.

Hagemann, S., Chen, C., Haerter, J. O., Heinke, J., Gerten, D., and Piani, C. (2011). Impact of a Statistical Bias Correction on the Projected Hydrological Changes Obtained from Three GCMs and Two Hydrology Models. *J. Hydrometeor.*, 12(4):556–578.

Hagemann, S., Machenhauer, B., Jones, R., Christensen, O., Déqué, M., Jacob, D., and Vidale, P. (2004). Evaluation of water and energy budgets in regional climate models applied over Europe. *Clim. Dyn.*, 23(5):547–567.

Han, J. and Pan, H.-L. (2011). Revision of Convection and Vertical Diffusion Schemes in the NCEP Global Forecast System. *Wea. Forecasting*, 26:520–533.

Handorf, D. and Dethloff, K. (2012). How well do state-of-the-art atmosphere-ocean general circulation models reproduce atmospheric teleconnection patterns? *Tellus A*, 64.

Hanssen-Bauer, I., Achberger, C., Benestad, R., Chen, D., and Førland, E. (2005). Statistical downscaling of climate scenarios over Scandinavia. *Clim. Res.*, 29:255–268.

Hay, L. and Clark, M. (2003). Use of statistically and dynamically downscaled atmospheric model output for hydrologic simulations in three mountainous basins in the western United States. *J. Hydrol.*, 282:56–75.

Haylock, M. R., Cawley, G. C., Harpham, C., Wilby, R. L., and Goodess, C. M. (2006). Downscaling heavy precipitation over the United Kingdom: a comparison of dynamical and statistical methods and their future scenarios. *Int. J. Climatol.*, 26(10):1397–1415.

Heinrich, G. and Gobiet, A. (2012). The future of dry and wet spells in Europe: a comprehensive study based on the ENSEMBLES regional climate models. *Int. J. Climatol.*, 32(13):1951–1970.

Hertig, E. and Jacobeit, J. (2008). Assessments of Mediterranean precipitation changes for the 21st century using statistical downscaling techniques. *Int. J. Climatol.*, 28(8):1025–1045.

Hirschi, M., Stoeckli, S., Dubrovsky, M., Spirig, C., Calanca, P., Rotach, M. W., Fischer, A. M., Duffy, B., and Samietz, J. (2012). Downscaling climate change scenarios for apple pest and disease modeling in Switzerland. *Earth Sys. Dyn.*, 3(1):33–47.

Hoffmann, H. and Rath, T. (2012). Meteorologically consistent bias correction of climate time series for agricultural models. *Theor. Appl. Climatol.*, 110:129–141.

Hohenegger, C., Brockhaus, P., Bretherton, C., and Schär, C. (2008). Towards climate simulations at cloud-resolving scales. *Meteorol. Z.*, 17(4):383–394.

Hong, S.-Y., Dudhia, J., and Chen, S.-H. (2004). A Revised Approach to Ice Microphysical Processes for the Bulk Parameterization of Clouds and Precipitation. *Mon. Weather Rev.*, 132:103.

Horton, P., Schaefli, B., Mezghani, A., Hingray, B., and Musy, A. (2006). Assessment of climate-change impacts on alpine discharge regimes with climate model uncertainty. *Hydrol. Process.*, 20:2091–2109.

Hu, X.-M., Nielsen-Gammon, J. W., and Zhang, F. (2010). Evaluation of Three Planetary Boundary Layer Schemes in the WRF Model. *J. Appl. Meteor. Climatol.*, 49:1831–1844.

Jacob, D., Bärring, L., Christensen, O. B., Christensen, J. H., Castro, M., Déqué, M., Giorgi, F., Hagemann, S., Hirschi, M., Jones, R., Kjellström, E., Lenderink, G., Rockel, B., Sánchez, E., Schär, C., Seneviratne, S. I., Somot, S., Ulden, A., and Hurk, B. (2007). An intercomparison of regional climate models for Europe: model performance in present-day climate. *Clim. Change*, 81(1):31–52.

Jacob, D., Petersen, J., Eggert, B., Alias, A., Christensen, O. B., Bouwer, L. M., Braun, A., Colette, A., Déqué, M., Georgievski, G., Georgopoulou, E., Gobiet, A., Menut, L., Nikulin, G., Haensler, A., Hempelmann, N., Jones, C., Keuler, K., Kovats, S., Kröner, N., Kotlarski, S., Kriegsmann, A., Martin, E., Meijgaard, E., Moseley, C., Pfeifer, S., Preuschmann, S., Radermacher, C., Radtke, K., Rechid, D., Rounsevell, M., Samuelsson, P., Somot, S., Soussana, J.-F., Teichmann, C., Valentini, R., Vautard, R., Weber, B., and Yiou, P. (2013). EURO-CORDEX: new high-resolution climate change projections for European impact research. *Reg. Environ. Change.*, pages 1–16.

Jakob, C. (2003). An improved strategy for the evaluation of cloud parameterizations in GCMS. *Bull. Amer. Meteor. Soc.*, 84:1387–1401.

Jordan, R. (1991). A one-dimensional temperature model for a snow cover: Technical documentation for SNTHERM. 89. Technical report, Hanover, NH.

Kain, J. S. (2004). The Kain Fritsch Convective Parameterization: An Update. *J. Appl. Meteor.*, 43:170–181.

Kain, J. S. and Fritsch, J. M. (1990). A One-Dimensional Entraining/Detraining Plume Model and Its Application in Convective Parameterization. *J. Atmos. Sci.*, 47:2784–2802.

Kallache, M., Vrac, M., Naveau, P., and Michelangeli, P.-A. (2011). Nonstationary probabilistic downscaling of extreme precipitation. *J. Geophys. Res.: Atmospheres*, 116(D5).

Kanamaru, H. and Kanamitsu, M. (2007). Scale-Selective Bias Correction in a Downscaling of Global Analysis Using a Regional Model. *Mon. Wea. Rev.*, pages 334–350.

Kapper, K. L. (2009). *Scale Sensitive Analysis of Regional Climate Models*. PhD thesis, University Graz, Austria. (cit. on p. 81).

Katz, R. (1996). Use of conditional stochastic models to generate climate change scenarios. *Clim. Change*, 32:237–255.

Kidson, J. W. and Thompson, C. S. (1998). A Comparison of Statistical and Model-Based Downscaling Techniques for Estimating Local Climate Variations. *J. Climate*, 11:735–753.

Kilsby, C. G., Cowpertwait, P. S. P., O'Connell, P. E., and Jones, P. D. (1998). Predicting rainfall statistics in England and Wales using atmospheric circulation variables. *Int. J. Climatol.*, 18(5):523–539.

Kjellström, E., Boberg, F., Castro, M., Christensen, J. H., Nikulin, G., and Sánchez, E. (2010). Daily and monthly temperature and precipitation statistics as performance indicators for regional climate models. *Clim. Res.*, 44(2–3):135–150.

Knievel, J. (2008). Physical parameterizations in the WRF Model. Presented at ATEC Forecasters' Conference, July 2007 and February 2008.

Koenig, U. and Abegg, B. (1997). Impacts of Climate Change on Winter Tourism in the Swiss Alps. *J. Sustain. Tour.*, 5(1):46–58.

Kotlarski, S., Paul, F., and Jacob, D. (2010). Forcing a Distributed Glacier Mass Balance Model with the Regional Climate Model REMO. Part I: Climate Model Evaluation. *J. Climate*, 23:1589–1606.

Kraus, H. (2004). *Die Atmosphäre Der Erde: Eine Einführung in Die Meteorologie*. Springer.

Langhans, W., Schmidli, J., and SchÃ€r, C. (2012). Bulk Convergence of Cloud-Resolving Simulations of Moist Convection over Complex Terrain. *J. Atmos. Sci.*, 69:2207–2228.

Laprise, R. (2008). Regional climate modelling. *J. Comput. Phys.*, 227(7):3641–3666.

Leander, R. and Buishand, T. A. (2007). Resampling of regional climate model output for the simulation of extreme river flows. *J. Hydrol.*, 332(3–4):487–496.

Leander, R., Buishand, T. A., van den Hurk, B. J., and de Wit, M. J. (2008). Estimated changes in flood quantiles of the river Meuse from resampling of regional climate model output. *J. Hydrol.*, 351(3–4):331–343.

Lehner, B., Döll, P., Alcamo, J., Henrichs, T., and Kaspar, F. (2006). Estimating the Impact of Global Change on Flood and Drought Risks in Europe: A Continental, Integrated Analysis. *Clim. Change*, 75(3):273–299.

Lenderink, G. (2010). Exploring metrics of extreme daily precipitation in a large ensemble of regional climate model simulations. *Clim. Res.*, 44:151–166.

Lenderink, G., Buishand, A., and van Deursen, W. (2007). Estimates of future discharges of the river Rhine using two scenario methodologies: direct versus delta approach. *Hydrol. Earth Syst. Sci.*, 11(3):1145–1159.

Lenderink, G. and van Meijgaard, E. (2008). Increase in hourly precipitation extremes beyond expectations from temperature changes. *Nature Geosci.*, 1(8):511–514.

Li, O. and Racine, J. S. (2007). *Nonmarametric Econometrics: Theory and Practice*. Princeton University Press.

Loader, C. (1999). *Local Regression and Likelihood*. Springer.

Lock, A. P. (2001). The Numerical Representation of Entrainment in Parameterizations of Boundary Layer Turbulent Mixing. *Mon. Wea. Rev.*, 129:1148–1163.

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. *Int. J. Climatol.*, 32(10):1579–1595.

Maraun, D. (2012). Nonstationarities of regional climate model biases in European seasonal mean temperature and precipitation sums. *Geophys. Res. Lett.*, 39.

Maraun, D. (2013). Bias Correction, Quantile Mapping and Downscaling. Revisting the Inflation Issue. *J. Climate*, 26:2137–2143.

Maraun, D., Wetterhall, F., Ireson, A. M., Chandler, R., Kendon, E., Widmann, M., Brienen, S., Rust, H., Sauter, T., Themeßl, M. J., and et al. (2010). Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user. *Rev. Geophys.*, 48(3):1–34.

Marbaix, P., Gallée, H., Brasseur, O., and van Ypersele, J.-P. (2003). Lateral Boundary Conditions in Regional Climate Models: A Detailed Study of the Relaxation Procedure. *Mon. Wea. Rev.*, 131(3):461–479.

Martin, G. M., Ringer, M. A., Pope, V. D., Jones, A., Dearden, C., and Hinton, T. J. (2006). The Physical Properties of the Atmosphere in the New Hadley Centre Global Environmental Model (HadGEM1). Part I: Model Description and Global Climatology. *J. Climate*, 19(7):1274–1301.

Matulla, C., Zhang, X., Wang, X., Wang, J., Zorita, E., Wagner, S., and von Storch, H. (2008). Influence of similarity measures on the performance of the analog method for downscaling daily precipitation. *Clim. Dyn.*, 30:133–144.

McGregor, J. (1997). Regional climate modelling. *Meteorol. Atmos. Phys.*, 63(1-2):105–117.

Mearns, L. O., Bogardi, I., Giorgi, F., Matyasovszky, I., and Palecki, M. (1999). Comparison of climate change scenarios generated from

regional climate model experiments and statistical downscaling. *J. Geophys. Res.: Atmospheres*, 104(D6):6603–6621.

Michelangeli, P.-A., Vrac, M., and Loukos, H. (2009). Probabilistic downscaling approaches: Application to wind cumulative distribution functions. *Geophys. Res. Lett.*, 36(11).

Mlawer, E. J., Taubman, S. J., Brown, P. D., Iacono, M. J., and Clough, S. A. (1997). Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *J. Geophys. Res.: Atmospheres*, 102(D14):16663–16682.

Moriasi, D., Arnold, J., Van Liew, M., Bingner, R., Harmel, R., and Veith, T. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE*, 50(3):885–900.

Murphy, J. (1999). An Evaluation of Statistical and Dynamical Techniques for Downscaling Local Climate. *J. Climate*, 12:2256–2284.

Nakićenović, N., Alcamo, J., Davis, G., de Vries, B., Fenhann, J., Gaffin, S., Gregory, K., Grübler, A., Jung, T. Y., Kram, T., La Rovere, E. L., Michaelis, L., Mori, S., Morita, T., Pepper, W., Pitcher, H., Price, L., Riahi, K., Roehrl, A., Rogner, H.-H., Snkovski, A., Schlesinger, M., Shukla, P., Smith, S., Swart, R., van Rooijen, S., Victor, N., and Dadi, Z. (2000). *IPCC: Special Report on Emissions Scenarios*. Cambridge University Press, UK, The Edinburgh Building Shaftesbury Road, Cambridge CB2 2RU ENGLAND.

Nash, J. and Sutcliffe, J. (1970). River flow forecasting through conceptual models part I – A discussion of principles. *J. Hydrol.*, 10(3):282–290.

Nešpor, V. and Sevruk, B. (1999). Estimation of Wind-Induced Error of Rainfall Gauge Measurements Using a Numerical Simulation. *J. Atmos. Oceanic Technol.*, 16(4):450–464.

O'Connor, F. M., Johnson, C. E., Morgenstern, O., and Collins, W. J. (2009). Interactions between tropospheric chemistry and climate model temperature and humidity biases. *Geophys. Res. Lett.*, 36(16).

Paeth, H. and Mannig, B. (2013). On the added value of regional climate modeling in climate change assessment. *Clim. Dyn.*, 41(3-4):1057–1066.

Pagan, A. and Ullah, A. (1999). *Nonparametric Econometrics*. Cambridge University Press.

Piani, C., Haerter, J., and Coppola, E. (2010a). Statistical bias correction for daily precipitation in regional climate models over Europe. *Theor. Appl. Climatol.*, 99(1-2):187–192.

Piani, C. and Haerter, J. O. (2012). Two dimensional bias correction of temperature and precipitation copulas in climate models. *Geophys. Res. Lett.*, 39.

Piani, C., Haerter, J. O., and Coppola, E. (2009). Statistical bias correction for daily precipitation in regional climate models over Europe. *Theor. Appl. Climatol.*, 99(1–2):187–192.

Piani, C., Weedon, G., Best, M., Gomes, S., Viterbo, P., Hagemann, S., and Haerter, J. (2010b). Statistical bias correction of global simulated daily precipitation and temperature for the application of hydrological models. *J. Hydrol.*, 395(3–4):199–215.

Pielke, R. (2002). *Mesoscale Meteorological Modeling*. International geophysics series. Academic Press.

Pielke, R. A. and Wilby, R. L. (2012). Regional climate downscaling: What's the point? *Eos, Transactions American Geophysical Union*, 93(5):52–53.

Prein, A., Gobiet, A., Suklitsch, M., Truhetz, H., Awan, N., Keuler, K., and Georgievski, G. (2013a). Added value of convection permitting seasonal simulations. *Clim. Dyn.*, pages 1–23.

Prein, A. F., Holland, G. J., Rasmussen, R. M., Done, J., Ikeda, K., Clark, M. P., and Liu, C. H. (2013b). Importance of Regional Climate Model Grid Spacing for the Simulation of Heavy Precipitation in the Colorado Headwaters. *J. Climate*, 26:4848–4857.

Räisänen, J. and Räty, O. (2012). Projections of daily mean temperature variability in the future: cross-validation tests with ENSEMBLES regional climate simulations. *Clim. Dyn.*, pages 1–16.

Roeckner, E., Bäuml, G., Bonaventura, L., Brokopf, R., Esch, M., Giorgetta, M., Hagemann, S., Kirchner, I., Kornblueh, L., Manzini, E., Rhodin, A., Schlese, U., Schulzweida, U., and Tompkins, A. (2003). The atmospheric general circulation model ECHAM5, Part I: Model description. Technical Report Report No. 349, Max-Planck-Institute for Meteorology.

Rohrer, M. B. (1992). Die Schneedecke im schweizerischen Alpenraum und ihre Modellierung. *Züricher Geographische Schriften*, 49.

Rojas, R., Feyen, L., Dosio, A., and Bavera, D. (2011). Improving pan-European hydrological simulation of extreme events through statistical bias correction of RCM-driven climate simulations. *Hydrol. Earth Syst. Sci.*, 15(8):2599–2620.

Rothfusz, L. P. (1990). The Heat Index "Equation" (or, More Than You Ever Wanted to Know About Heat Index). Tech. Attachment SR 90–23, NWS Southern Region Headquarters.

Samuelsson, P., Jones, C. G., Willén, U., Ullerstig, A., Gollvik, S., Hansson, U., Jansson, C., Kjellström, E., Nikulin, G., and Wyser, K. (2011). The Rossby Centre Regional Climate model RCA3: model description and performance. *Tellus A*, 63(1):4–23.

Schmidli, J., Frei, C., and Vidale, P. (2006). Downscaling from GCM precipitation: a benchmark for dynamical and statistical downscaling methods. *Int. J. Climatol.*, 26(5):679–689.

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. *Int. J. Climatol.*, 22(9):1049–1074.

Schoen, C. (2005). A New Empirical Model of the Temperature Humidity Index. *J. Appl. Meteor.*, 44:1413–1420.

Scott, D. and McBoyle, G. (2003). Climate change and the skiing industry in southern Ontario (Canada): exploring the importance of snowmaking as a technical adaptation. *Clim. Res.*

Seifert, A. and Beheng, K. D. (2006). A two-moment cloud microphysics parameterization for mixed-phase clouds. Part 1: Model description. *Meteorol. Atmos. Phys.*, 92(1-2):45–66.

Sellers, P., Randall, D., Berry, J., Field, C., Dazlich, D., Zhang, C., Collelo, G., and Bounoua, L. (1996). A revised land surface parameterization (sib2) for atmospheric gcms. part i: model formulation. *J. Climate*, 9:676–705.

Sharma, D., Das Gupta, A., and Babel, M. S. (2007). Spatial disaggregation of bias-corrected GCM precipitation for improved hydrologic simulation: Ping River Basin, Thailand. *Hydrol. Earth Syst. Sci.*, 11(4):1373–1390.

Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K., Tignor, M., and Miller, H. (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*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Steadman, R. G. (1979). The Assessment of Sultriness. Part I: A Temperature-Humidity Index Based on Human Physiology and Clothing Science. *J. Appl. Meteor.*, 18:861–873.

Strasser, U. (2004). Spatial and temporal variability of meteorological variables at Haut Glacier d'Arolla (Switzerland) during the ablation season 2001: Measurements and simulations. *J. Geophys. Res.*, 109(D3):D03103.

Strasser, U. (2008). Modelling of the mountain snow cover in the Berchtesgaden National Park. Technical Report 55, Berchtesgaden.

Strasser, U., Bernhardt, M., Weber, M., Liston, G. E., and Mauser, W. (2008). Is snow sublimation important in the alpine water balance? *The Cryosphere*, 2:53–66.

Suklitsch, M., Gobiet, A., Leuprecht, A., and Frei, C. (2008). High Resolution Sensitivity Studies with the Regional Climate Model CCLM in the Alpine Region. *Meteorol. Z.*, 17(4):467–476.

Suklitsch, M., Gobiet, A., Truhetz, H., Awan, N., Göttel, H., and Jacob, D. (2011). Error characteristics of high resolution regional climate models over the Alpine area. *Clim. Dyn.*, 37:377–390.

Teutschbein, C. and Seibert, J. (2012). Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods. *J. Hydrol.*, 456–457(0):12–29.

Themeßl, M. J., Gobiet, A., and Heinrich, G. (2012). Empirical-Statistical Downscaling and Error Correction of Regional Climate Models and its Impact on the Climate Change Signal. *Clim. Change*, 112(2):449–468.

Themeßl, M. J., Gobiet, A., and Leuprecht, A. (2011). Empirical-statistical downscaling and error correction of daily precipitation from regional climate models. *Int. J. Climatol.*, 31(10):1530–1544.

Uppala, S. M., Kållberg, P. W., Simmons, A. J., Andrae, U., Bechtold, V. D. C., Fiorino, M., Gibson, J. K., Haseler, J., Hernandez, A., Kelly, G. A., Li, X., Onoi, K., Saarinen, S., Sokka, N., Allan, R. P., Andersson, E., Arpe, K., Balmaseda, M. A., Beljaars, A. C. M., Berg, L. V. D., Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Hólm, E., Hoskins, B. J., Isaksen, L., Janssen, P. A. E. M., Jenne, R., Mcnally, A. P., Mahfouf, J.-F., Morcrette, J.-J., Rayner, N. A., Saunders, R. W., Simon, P., Sterl, A., Trenberth, K. E., Untch, A., Vasiljevic, D., Viterbo, P., and Woollen, J. (2005). The ERA-40 re-analysis. *Q. J. R. Meteorolog. Soc.*, 131(612):2961–3012.

van der Linden, P. and Mitchell, J. (2009). *ENSEMBLES: Climate Change and its Impacts: Summary of research and results from the ENSEMBLES project*. Met Office Hadley Centre.

van Meijgaard, E., van Ulft, L., van de Berg, W., Bosveld, F., van den Hurk, B., Lenderink, G., and Siebesma, A. (2008). The knmi regional atmospheric climate model racmo, version 2.1. Technical Report KNMI Technical Report 302, KNMI, Postbus 201, 3730 AE, De Bilt, The Netherlands.

van Vuuren, D. P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., Hurtt, G. C., Kram, T., Krey, V., Lamarque, J.-F., Masui, T., Meinshausen, M., Nakićenović, N., Smith, S. J., and Rose, S. K. (2011). The representative concentration pathways: an overview. *Clim. Change*, 109(1–2):5–31.

Varis, O., Kajander, T., and Lemmelä, R. (2004). Climate and Water: From Climate Models to Water Resources Management and Vice Versa. *Clim. Change*, 66(3):321–344.

Vautard, R., Gobiet, A., Jacob, D., Belda, M., Colette, A., Déqué, M., Fernández, J., García-Díez, M., Goergen, K., Güttler, I., Halenka, T., Karacostas, T., Katragkou, E., Keuler, K., Kotlarski, S., Mayer, S., Meijgaard, E., Nikulin, G., Patarčić, M., Scinocca, J., Sobolowski, S., Suklitsch, M., Teichmann, C., Warrach-Sagi, K., Wulfmeyer, V., and Yiou, P. (2013). The simulation of European heat waves from an ensemble of regional climate models within the EURO-CORDEX project. *Clim. Dyn.*, 41(9-10):2555–2575.

Vautard, R., Moran, M. D., Solazzo, E., Gilliam, R. C., Matthias, V., Bianconi, R., Chemel, C., Ferreira, J., Geyer, B., Hansen, A. B., Jericevic, A., Prank, M., Segers, A., Silver, J. D., Werhahn, J., Wolke, R., Rao, S., and Galmarini, S. (2012). Evaluation of the meteorological forcing used for the Air Quality Model Evaluation International Initiative (AQMEII) air quality simulations . *Atmos. Environ.*, 53:15–37.

Viner, D., Hulme, M., and Raper, S. (1995). Climate Change Scenarios for the Assessments of the Climate Change on Regional Ecosystems. *J. Therm. Biol.*, 20(1–2):175–190.

von Storch, H. (1999). On the Use of "Inflation" in Statistical Downscaling. *J. Climate*, 12(12):3505–3507.

Wang, Y., Leung, L. R., McGregor, J. L., Lee, D. K., Wang, W. C., Ding, Y. H., and Kimura, F. (2004). Regional climate modeling: Progress, challenges and prospects. *J. Meteor. Soc. Japan*, 82:1599–1628.

Wetterhall, F., Halldin, S., and yu Xu, C. (2005). Statistical precipitation downscaling in central Sweden with the analogue method. *J. Hydrol.*, 306(1–4):174–190.

Widmann, M. (2005). One-Dimensional CCA and SVD, and Their Relationship to Regression Maps. *J. Climate*, 18(14):2785–2792.

Widmann, M., Bretherton, C., and Salathé Jr., E. (2003). Statistical Precipitation Downscaling over the Northwestern United States Using Numerically Simulated Precipitation as a Predictor. *J. Climate*, 16(5):799–816.

Wilby, R. L., Hay, L. E., Gutowski, W. J., Arritt, R. W., Takle, E. S., Pan, Z., Leavesley, G. H., and Clark, M. P. (2000). Hydrological responses

to dynamically and statistically downscaled climate model output. *Geophys. Res. Lett.*, 27(8):1199–1202.

Wilcke, R. A. I., Leuprecht, A., and Andreas, G. (2012). Effects of Climate Change on Future Snow Conditions in Tyrol and Styria (CC-Snow), Final report – Climate. Wegener Center Verlag, Graz, Austria.

Wilcke, R. A. I., Mendlik, T., and Gobiet, A. (2013). Multi-Variable Downscaling and Error-Correction of Regional Climate Models. *Clim. Change*, 120(4):871–887.

Wilks, D. (2006). *Statistical Methods in the Atmospheric Sciences*. Academic Press, Amsterdam, 2nd edition.

Wilks, D. S. and Wilby, R. L. (1999). The weather generation game: a review of stochastic weather models. *Prog. Phys. Geog.*, 23(3):329–357.

Winter, J. M., Pal, J. S., and Eltahir, E. A. (2009). Coupling of Integrated Biosphere Simulator to Regional Climate Model Version 3. *J. Climate*, 22.

Witmer, U. (1986). *Erfassung, Bearbeitung und Kartierung von Schneedaten in der Schweiz*. Geographica Bernensia G25, Bern.

Wood, A. W., Leung, L. R., Sridhar, V., and Lettenmaier, D. P. (2004). Hydrologic Implications of Dynamical and Statistical Approaches to Downscaling Climate Model Outputs. *Clim. Change*, 62:189–216.

Xu, C.-Y. (1999). From GCMs to river flow: a review of downscaling methods and hydrologic modelling approaches. *Prog. Phys. Geog.*, 23(2):229–249.

Ye, B., Yang, D., and Ma, L. (2012). Effect of precipitation bias correction on water budget calculation in Upper Yellow River, China. *Environ. Res. Lett.*, 7(2):025201.

Yee, T. W. and Stephenson, A. G. (2007). Vector generalized linear and additive extreme value models. *Extremes*, 10(1-2):1–19.

Zorita, E. and von Storch, H. (1999). The Analog Method as a Simple Statistical Downscaling Technique: Comparison with More Complicated Methods. *J. Climate*, 12:2474–2489.

Abstract:
Statistical downscaling and error correction methods are used to improve output from Regional Climate Models (RCMs) for the use in climate impact research. One such statistical empirical method is Quantile Mapping (QM), which is often used to correct for biases in RCM's temperature and precipitation, recently. In this thesis the performance of QM is investigated when applied to temperature and precipitation, but also to relative humidity, wind speed, global radiation, and surface pressure.
The main results are that QM strongly reduces the annual and monthly mean biases and corrects the distributions—independent of shape—for all variables and RCMs. Even non-stationarities in the biases cannot hinder the good results of QM. The temporal and inter-variable structure provided within the RCMs is not manipulated by QM, which is positive when the RCM's structure is correct. As strong deviations from observed inter-variable relations are found in RCMs, the conservation of those relations produces less convincing results of QM for combined indices.
The effect of individually quantile-mapped climate variables on derived variables, like number of snow days, snow depth, or temperature-humidity index (THI) is mostly positive. Derived variables are sensitive to the day-to-day structure of time-series. The improvement is not as strong as for directly corrected variables, as QM corrects for the mean statistics and preserves the day-to-day structure of the RCM.
However, in terms of Nash-Sutcliffe-Efficiency (NSE) no improvement can be found in most cases.
Generally, QM corrects individual climate variables with convincing results, but needs further enhancements to improve the inter-variable relations in RCM output.

Zum Inhalt:
Um Klimadaten aus Regionalen Klimamodellen (RCMs) für Klimafolgenforschung aufzubereiten werden statistische Skalierungs- und Fehlerkorrektur-Methoden verwendet. Diesbezüglich besonders geeignet ist Quantile Mapping (QM), welches in letzter Zeit häufig zur Korrektur simulierter Temperatur und Niederschlag verwendet wird. In dieser Arbeit wird zusätzlich untersucht wie gut QM Fehler in Relative Feuchte, Windgeschwindigkeit, Globalstrahlung und Bodendruck korrigieren kann.
Zu den Hauptergebnissen zählt die starke Reduktion jährlicher und monatlicher Fehler durch QM, wie auch die formunabhängige Korrektur der Dichteverteilungen für alle Variablen und RCMs. Sogar nicht stationäre Fehler werden mit QM reduziert. Es wird gezeigt, dass QM die vom RCM vorgegebene zeitliche Struktur und die Beziehungen zwischen Variablen nicht verändert. Allerdings werden zum Teil große Unterschiede zwischen den beobachteten und den simulierten Beziehungen zwischen Variablen gefunden. Durch die Erhaltung der fehlerhaften simulierten Beziehungen können größere Fehler in Indizes aus kombinierten Variablen erhalten bleiben.
Der Effekt von fehlerkorrigierten Klimavariablen auf davon abgeleitete Variablen ist überwiegend positiv, wie z.B. bei der Anzahl an Schneetagen, Schneehöhe oder Temperatur-Feuchte-Index (THI). Abgeleitete und kombinierte Variablen sind sensibel gegenüber der täglichen Struktur und den Beziehungen zwischen den Variablen. Da QM genau diese Strukturen in den Modeldaten nicht ändert, ist die Verbesserung nicht so großwie bei den direkt korrigierten Variablen. Gemessen an der Nash-Sutcliffe-Effizienz (NSE) kann keine Verbesserung gefunden werden.
Zusammengefasst korrigiert QM Klimavariablen sehr gut. Um die Beziehungen zwischen Variablen und zeitliche Strukturen in Modeldaten zu verbessern, muss QM allerdings weiterentwickelt werden.

Wegener Center for Climate and Global Change
University of Graz
Brandhofgasse 5
A-8010 Graz, Austria
www.wegcenter.at
ISBN 978-3-9503608-3-7