Stochastic Bias Correction of Dynamically Downscaled Precipitation Fields for Germany Through Copula-based Integration of Gridded Observation Data

Ganquan Mao

Angaben zur Veröffentlichung / Publication details:

Mao, Ganquan. 2016. "Stochastic Bias Correction of Dynamically Downscaled Precipitation Fields for Germany Through Copula-based Integration of Gridded Observation Data." Augsburg: Universität Augsburg.

Stochastic Bias Correction of Dynamically Downscaled Precipitation Fields for Germany Through Copula-based Integration of Gridded Observation Data

Dissertation zur Erlangung des Doktorgrades an der
Fakultät für Angewandte Informatik der
Universität Augsburg

vorgelegt von
M.Sc. Ganquan Mao
2015

Erstes Gutachten:	Prof. Dr. Harald Kunstmann
Zweites Gutachten:	PD Dr. Elke Hertig
Tag der Mündlichen Prüfung:	11. März 2016

Acknowledgements

Completing this dissertation has definitely been a challenge and could not have been accomplished without the assistance and support of many individuals and institutions. First and foremost I want to thank my supervisor Prof. Dr. Harald Kunstmann. His profound knowledge and expertise in this area of research has always led me throughout my work. The joy and enthusiasm he has for his research is contagious and motivational for me. What I have learned from him is not only the scientific knowledge, but also the way of thinking about research, which is invaluable for my future academic career. I appreciate all his contributions of time, ideas, and patience to make my Ph.D. experience productive and stimulating.

I would like to thank my dissertation reviewers PD Dr. Elke Hertig and PD Dr. Christoph Beck for their time, interest, and helpful comments.

I would like to give my special thanks to Dr. Patrick Laux, Dr. Stefanie Vogl and Dr. Sven Wagner for their kindness and generous helps. They give me assistance and guidance during my Ph.D study. I also greatly appreciate that they have been always available for discussions and answering questions. I would like to express my gratitude also to Christof Lorenz for helping me with the GMT plots.

I am pleased to acknowledge that the funding for this study was provided by a fellowship from China Scholarship Council. Without this financial assistance this study could not have been undertaken.

Additionally, I would like to thank all the members of the groups of Prof. Dr. Harald Kunstmann in IMK-IFU, KIT, for the friendly and pleasant environment they have created.

Finally, I am extremely grateful to my beloved family for their unconditional love and support over the last few years. I offer my regards and blessings to all of those who supported me in any respect during the completion of the dissertation.

Abstract

Dynamically downscaled precipitation fields from regional climate models (RCMs) often cannot be used directly for regional/local climate studies. Due to their inherent biases, i.e., systematic over- or underestimations compared to observations, correction approaches are usually required. Most of the bias correction procedures, such as the quantile mapping approach, employ a transfer function that is based on the statistical differences between RCM output and observations. Apart from such transfer function-based statistical correction algorithms, a stochastic bias correction technique, based on the concept of Copula theory, is developed in this thesis and applied to correct precipitation fields from the Weather Research and Forecasting (WRF) model. For dynamically downscaled precipitation fields, high-resolution (7 km, daily) WRF simulations for Germany driven by ERA40 reanalysis data for 1971–2000 were used. The REGNIE (REGionalisierung der NIEderschlagshöhen) data set from the German Weather Service (DWD) is used as gridded observation data (1 km, daily) and aggregated to 7 km for this application. The 30-year time series are split into a calibration (1971–1985) and validation (1986–2000) period of equal length. Based on the estimated dependence structure (described by a Copula function) between WRF and REGNIE data and the identified respective marginal distributions in the calibration period, conditional distribution functions are derived for each time step in the validation period. To generate bias corrected WRF precipitation, a random sample of possible outcomes is drawn from this conditional distribution. Thereby, this method does not only provide a single correction value for each time step, but rather estimates the range of possible values and the full probability density function (PDF).

The Copula-based correction is applied in two modes: The overall mode and seasonal mode. In the overall mode, for each grid cell all of the data from the calibration period are used to construct an overall Copula model. In the seasonal mode approach, the Copula models are estimated for each season separately. The results show that the Copula-based approach with both the overall and seasonal modes are able to correct most of the errors in WRF derived precipitation. The seasonal mode based correction is found to be more efficient. However, it is also found that the Copula-based correction in seasonal mode performs better for the wet bias correction than for the dry bias correction. The average relative bias of daily mean precipitation from WRF for the validation period is reduced from 10 % (wet bias) to −1 % (slight dry bias) after the application of the Copula-based correction seasonal mode. The bias in different seasons is

corrected from 32 % during March–April–May (MAM), −15 % during June–July–August (JJA), 4 % during September–October–November (SON) and 28 % during December–January–February (DJF) to 16 % (MAM), −11 % (JJA), −1 % (SON) and −3 % (DJF), respectively.

The Copula-based approach in seasonal mode is compared to the quantile mapping correction method. The root mean square error (RMSE) and the percentage of the corrected time steps that are closer to the observations are analyzed. The Copula-based correction derived from the mean of the sampled distribution reduces the RMSE significantly, while, e.g., the quantile mapping method results in an increased RMSE for some regions.

Finally, as outlook, the Copula-based stochastic bias correction is further extended to allow its use for climate projections, i.e. episodes in the future where no observation is available yet. This is achieved by a first-order hidden Markov model and the Viterbi algorithm. The extended framework is briefly described, applied for four selected pixels to validate its performance, and its potential is shown.

Zusammenfassung

Aus Regionalen Klimamodellen (RCMs) generierte Niederschlagsfelder können oftmals nicht direkt für lokale Klimaimpaktstudien verwendet werden. Wegen der inhärenten Fehler, d.h. systematische Über- oder Unterschätzungen im Vergleich zu den Beobachtungen, sind Fehlerkorrekturverfahren notwendig. Die meisten Fehlerkorrekturverfahren, wie beispielsweise das Quantile Mapping, leiten zur Korrektur eine Transferfunktion basierend auf statistischen Beziehungen zwischen RCM Feldern und Beobachtungen ab. Anders als solche auf Transferfunktionen beruhenden Korrekturverfahren wurde in dieser Doktorarbeit ein stochastisches Korrekturverfahren entwickelt und angewandt, das auf dem Konzept der Copula-Theorie basiert. Es wurde angewandt um Niederschlagsfelder des Weather Research and Forecasting (WRF) Modells zu korrigieren. In dieser Arbeit wurden dazu hochaufgelöste WRF Niederschlagsfelder (7 km horizontale Auflösung auf Tagesbasis) für Deutschland und die Periode 1971-2000 verwendet, angetrieben von ERA40 Reanalysen. In dieser Studie wurden REGNIE (REGionalisierung der NIEderschlagshöhen) Daten vom Deutschen Wetterdienst (DWD) als gerasterte Beobachtungsdaten (1 km horizontale Auflösung auf Tagesbasis) benutzt, die zu einer Auflösung von 7 km aggregiert wurden. Die Daten der Periode 1971–2000 wurden zu gleichen Teilen zur Kalibrierung (1971–1985) und zur Validierung (1986–2000) aufgeteilt. Auf der Grundlage der Abhängigkeitsstruktur zwischen WRF und REGNIE, die durch Copulas beschrieben werden kann, sowie den entsprechenden Marginalverteilungen, wurden konditionierte Verteilungs funktionen für jeden einzelnen Zeitschritt im Validierungszeitraum abgeleitet. Dieses Verfahren liefert daher nicht nur einen einzigen Korrekturwert für jeden Zeitschritt, sondern schätzt vielmehr einen Bereich an möglichen Werten und sogar die gesamte Wahrscheinlichkeitsdichtefunktion (PDF).

Die Korrektur wurde in zwei verschiedenen Modi durchgeführt, einerseits über den gesamten Untersuchungszeitraum (Gesamt-Modus) und andererseits separat für die einzelnen Jahreszeiten (Jahreszeiten-Modi). Beim Gesamt-Modus wird für jede Gitterzelle eine Copula Funktion aus allen Daten des Kalibrierungszeitraums abgeleitet, während bei den Jahreszeiten-Modi nur die Niederschlagsdaten der jeweiligen Jahreszeiten verwendet werden. Die Ergebnisse zeigen, dass sowohl das Modell im Gesamt-Modus als auch die Modelle mit den Jahreszeiten-Modi die meisten Fehler in den WRF Niederschlagszeitreihen effizient korrigieren. Die Modelle in den Jahreszeiten-Modi sind dabei aber noch effizienter. Außerdem zeigen sich die Modelle in den Jahreszeiten-Modi

performanter bei der Korrektur von Überschätzungen im RCM im Vergleich zu Unterschätzungen. Im Herbst und Winter führt die Fehlerkorrektur zu leichten Unterschätzungen im Nordwesten Deutschlands. Der mittlere relative Fehler kann durch die Fehlerkorrektur von 10 % (Überschätzung des RCMs) auf −1 % (leichte Unterschätzung des RCMs) reduziert werden. Der Fehler in den verschiedenen Jahreszeiten konnte von 32 % für März, April, Mai (MAM), −15 % für Juni, Juli, August (JJA), 4 % für September, Oktober, November (SON), und 28 % für Dezember, Januar, Februar (DJF) auf 16 % (MAM), −11 % (JJA), −1 % (SON) und −3 % (DJF) korrigiert werden.

Die Copula Korrekturmethode (Jahreszeiten-Modi) wurde mit der quantile mapping Methode verglichen. Der root mean square error (RMSE) und der Anteil korrigierter Zeitschritte, die besser mit den Beobachtungen übereinstimmen, wurden analysiert. Die Copula Korrektur konnte unter Verwendung des Erwartungswertes (Mittelwertes) den RMSE des gesamten Gebiets erheblich reduzieren, während das Quantile Mapping den RMSE für einige Regionen erhöht.

Im Ausblick wird schließlich die Weiterentwicklung der Copula Korrekturmethode zur Anwendung für Klimaprojektionen beschrieben. Dies wird mit einem Markov Modell erster Ordnung und dem Viterbi Algorithmus erreicht. Die Erweiterung wird eingeführt, und das Potential dieser Erweiterung wird anhand von 4 ausgewählte Gitterpunkten demonstriert.

Contents

List of Figures

List of Tables

Abbreviations

AIC	**A**kaike **I**nformation **C**riterion
ARW	**A**dvanced **R**esearch **W**RF
BIC	**B**ayesian **I**nformation **C**riterion
CDF	**C**umulative **D**istribution **F**unction
CP	**C**irculation **P**atterns
CV	**C**oefficient of **V**ariation
DEM	**D**igital **E**levation **M**odel
DJF	**D**ecember–**J**anuary–**F**ebruary
DWD	**D**eutscher **W**etter**d**ienst (German Weather Service)
FA	**F**requency **A**daptation
GCM	**G**lobal **C**irculation **M**odel
Gof	**G**oodness-**o**f-**F**it
HMM	**H**idden **M**arkov **M**odel
JJA	**J**une–**J**uly–**A**ugust
MAM	**M**arch–**A**pril–**M**ay
MLE	**M**aximum **L**ikelihood **E**stimation
PDF	**P**robability **D**ensity **F**unction
QM	**Q**uantile **M**apping

Q-Q	**Q**uantile-**Q**uantile
RCM	**R**egional **C**limate **M**odel
REGNIE	**REG**ionalisierung der **NIE**derschlagshöhen
RMSE	**R**oot **M**ean **S**quare **E**rror
SON	**S**eptember–**O**ctober–**N**ovember
WRF	**W**eather **R**esearch and **F**orecasting

Chapter 1

Introduction

1.1 Motivation

Assessing the impact of climate change on, e.g., the hydrological cycle or crop production, requires the knowledge of projections of climatological variables (e.g., temperature, precipitation) at regional or local scales. Global circulation models (GCMs), however, provide climatological information only on coarse scales, thus cannot be used directly for the assessment of regional consequences of climate change. In order to obtain fine-resolution information from GCM outputs, downscaling methods are usually employed. As dynamical downscaling, regional climate models (RCMs) are capable to bridge the gap between large-scale GCM data and local-scale information. RCMs use the output of the GCMs and provide climate variables at a finer spatial resolution. Nevertheless, the RCM simulations often show significant biases and do not agree well with the observations (Smiatek et al., 2009; Teutschbein and Seibert, 2010). These biases usually consist of two parts: (1) RCMs inherit some of the biases of the GCMs, (2) RCMs potentially introduce new biases due to the imperfect conceptualization, discretization and spatial averaging within grid cells. This makes the use of RCM simulations as direct input data for e.g. hydrological impact studies more complicated. Therefore, further bias correction is often required. The impacts of biases on hydrological and agriculture modeling has been studied extensively (Kunstmann et al., 2004; Baigorria et al., 2007; Ghosh and Mujumdar, 2009; Ott et al., 2013; Hertig and Jacobeit, 2013).

Precipitation is an important parameter in climate change impact studies. The simulation of precipitation is highly sensitive to the grid resolution, the numerical scheme and

physical parameterisations (Bachner et al., 2008; Fowler and Ekstr, 2009; Maraun et al., 2010). In climate models, the representation of precipitation is depending on a multitude of different processes such as cloud microphysics, radiative transfer, atmospheric dynamics and boundary layer interactions acting over a variety of space and time scales (Hagemann et al., 2011). The biases might be large for precipitation due to its highly nonlinear nature and large spatial variability (Fowler et al., 2007). Convection schemes are also considered as error sources, since they were primarily developed for the tropics with coarse resolutions (Hohenegger et al., 2008).

Typical biases of RCM derived precipitation are: RCMs usually tend to generate too many wet days with low-intensity rain (Ines and Hansen, 2006); RCMs often contain under- and overestimations of rainfall as well as incorrect representations of the seasonality (Schmidli et al., 2006; Terink et al., 2010); Even long-term means are also not reproduced well (Bárdossy and Pegram, 2012). The implementation of bias correction for precipitation (a discrete variable) is more complex than a bias correction of continuous variables, e.g.like temperature. The bias correction of precipitation is particularly challenging as it has to cope with the problem that the precipitation data is zero inflated.

1.2 State of the art

Bias correction procedures of precipitation usually employ a transformation algorithm to adjust RCM output. The underlying strategy is the identification of possible biases between observed and simulated climate variables and then used to correct the RCM runs. The assumption of these methods is that the applied correction procedure and parameters are assumed to remain constant over time, i.e. the bias behavior of the model does not change with time. Therefore the temporal errors of RCM runs cannot be corrected which is the major limitation of such bias correction methods. Recently, several bias correction methods for precipitation have been developed. These methods range from simple scaling approaches such as the linear scaling approach and local intensity scaling to methods like power transformation and quantile mapping (e.g., Ines and Hansen, 2006). A recent overview of bias correction methods for hydrological application is provided, e.g., by Themeßl et al. (2010); Teutschbein and Seibert (2012); Lafon et al. (2013). In the following, a brief description of those methods are listed.

One of the classical methods, often used in hydrology, is the delta change method (Hay et al., 2000). This approach simply adds the climate change signal of the climate model

to observations, before they are used to force hydrological models. With this method, it is assumed that the model error is the same for the control and future run. Thus, after subtraction, the pure climate signal remains. This method only accounts for changes in the mean but not for changes in variability.

The linear scaling approach operates with a linear transformation equation which considers also only changes in the mean of observed and simulated values (Shabalova et al., 2003; Horton et al., 2006; Lenderink et al., 2007). Precipitation is corrected with a factor based on the ratio of long-term mean or monthly mean of observed and simulated time series. By definition, corrected RCM simulations will perfectly agree in their mean or monthly mean values with the observations. The linear correction method belongs to the same family as the "factor of change" or "delta change" method (Hay et al., 2000). This method has the advantage of simplicity and modest data requirements: only monthly climatological information is required to calculate monthly correction factors. However, this method only considers the the first statistical moments (mean), and does not account for other statistical moments of the probability distribution of daily precipitations (Arnell, 2003; Diaz-Nieto and Wilby, 2005).

The local intensity scaling presented by Schmidli et al. (2006) then takes the linear scaling one step further and adjusts the mean as well as both wet-day frequencies and wet-day intensities of precipitation time series separately. First, an RCM specific precipitation threshold is calibrated such that the number of RCM simulated days exceeding this threshold matches the number of wet days of the observation. Then, the number of precipitation events for both control and scenario run are corrected by applying the calibrated RCM precipitation threshold so that all days with precipitation less than the RCM specific precipitation threshold are redefined to dry days with 0 mm precipitation. Finally, the same procedure as the linear scaling is applied to correct the precipitation intensities. This method corrects the precipitation mean as well as the dry probability of the precipitation time series.

Due to the fact that both the linear scaling and local intensity scaling are linear corrections. They are not able to correct the variance. Therefore, a non-linear correction in an exponential form of $P^* = aP^b$, is introduced and can be applied to specifically adjust the variance of a precipitation time series (Leander and Buishand, 2007; Leander et al., 2008). The scaling exponent, b, is calculated iteratively so that the coefficient of variation (CV) of the RCM precipitation time series matches that of the observed precipitation time series. Here, this is achieved using Brent's method (Press et al., 1993). The pre-factor, a, is then calculated so that the mean of the transformed precipitation values

is equal to the observed mean. This power transformation adjusts both the mean and the variance of the RCM simulations at the same time. However, biases in higher order moments are not removed by the nonlinear method. This method of bias correction also does not correct for the fraction of wet and dry days.

So far, all these methods correct the RCM simulations with a simple equation. They are easy to apply and are less computational demanding, but aim only at a few statistical characteristics. Extending the correction from mean and variance to the entire distribution, quantile mapping corrects for errors in the shape of the distribution and is therefore capable to correct all the statistical moments. This quantile-based approach originates from the empirical transformation of Panofsky and Brier (1968) and was successfully implemented in hydrological applications (Dettinger et al., 2004; Wood et al., 2004; Boé et al., 2007; Paxian et al., 2014) but recently also for error correction of RCMs (Dobler and Ahrens, 2008; Piani et al., 2010a). The idea of quantile mapping is to correct the distribution function of RCM-simulated climate values to agree with the observed distribution function. This is done by creating a transfer function to shift the occurrence distributions of precipitation (Sennikovs and Bethers, 2009):

$$P^* = F_{obs}^{-1}(F_{RCM}(P)) \tag{1.1}$$

where P^* is the corrected precipitation value, P is the original modelled precipitation value and F_{obs} is the CDF of the observations and accordingly F_{RCM} is the CDF of the RCM precipitation. F_{obs}^{-1} is the inverse function, which is named quantile function. Several other names can be found in the literature, such as 'probability mapping' (Block et al., 2009; Ines and Hansen, 2006), 'distribution mapping' (Teutschbein and Seibert, 2012), 'statistical downscaling' (Piani et al., 2010a) and 'histogram equalization' (Sennikovs and Bethers, 2009; Rojas et al., 2011).

The quantile mapping (QM) transformation can be grouped into three categories: 1) Empirical non-parametric transformation (Gutjahr and Heinemann, 2013); 2) Distribution derived parametric transformation (Ines and Hansen, 2006); 3) Parametric transformation with combination of two or more distributions (Gudmundsson et al., 2012).

The empirical non-parametric transformation uses empirical distribution functions (both for observed and modelled precipitation) to construct the transfer function and is hence referred to as empirical quantile mapping (eQM). To implement the empirical distribution correction method, the ranked observed precipitation distribution is divided into a

number of discrete quantiles. Following the procedure of Lafon et al. (2013), the empirical CDFs are approximated using tables of empirical percentiles. Values in between the percentiles are approximated using linear interpolation. For each quantile division, a linear correction factor was calculated by dividing the mean observation in that quantile by the RCM simulated mean precipitation in the same quantile, this being the transfer function. The number of quantile divisions controls the accuracy of the method: using fewer quantiles might smooth out the information contained within the observed record, while using too many quantiles might result in overfitting of the model to the data. This method is expected to produce the best correction due to the empirical fitting but also with a shortcoming of its inability of generating "new extremes". Alternatively, if new model values (e.g. from climate projections) are larger than the training values used to estimate the empirical CDF, the correction found for the highest quantile of the training period is used (Boé et al., 2007; Themeßl et al., 2011). Furthermore, this method depends on many degrees of freedom (two times the sample size) and hence may not be stationary for future time periods (Piani et al., 2010b).

Statistical transformations can be achieved by using theoretical distributions to solve Eq. 1.1, which is called the distribution derived parametric transformation. This approach has seen wide application for adjusting modelled precipitation (e.g. Ines and Hansen, 2006; Li et al., 2010; Piani et al., 2010a; Teutschbein and Seibert, 2012). Most of these studies assume that a two-parameter Gamma distribution is suitable to model precipitation intensities. In some literature, it is also called Gamma distribution correction method (Lafon et al., 2013). The Gamma distribution is defined as

$$f(x) = \frac{1}{b^a \Gamma(a)} x^{a-1} e^{\frac{-x}{b}} \qquad x, a, b \geq 0 \tag{1.2}$$

with b as the scale parameter, a as the shape parameter and Γ as the Gamma function. The Gamma distribution is not defined for $x = 0\,mm/day$. Therefore, the correction process will be a dual step (Piani et al., 2010b). First, the number of dry days is corrected by optimizing a threshold value, i.e. all values smaller than this threshold are set to zero, such that the number of dry days equals the observations. Afterwards, these fitted Probability Density Functions (PDFs) are integrated, and the resulting CDFs are used to replace the empirical CDFs in Eq. 1.1. The Gamma distribution correction depends only on two parameters and requires less computation, since it has a closed formula and no numerical calculation is needed. However, daily precipitation distributions are typically heavily skewed towards low-intensity values. When fitting a single Gamma distribution, the distribution parameters will be dictated by the most frequently occurring values, but

may then not accurately represent the extremes (Yang et al., 2010; Teng et al., 2015). Vlček and Huth (2009) also found that the Gamma distribution does not represent daily precipitation for every region in Europe adequately and should be tested a priori.

In practice, a two-parameter fit to the daily precipitation transform function was used as a good approximation for most of the cases (Hagemann et al., 2011). For some specific cases, three- or four-parameter transfer functions produced better results (Hagemann et al., 2011; Piani et al., 2010b). Therefore the parametric transformation with combined distributions is introduced and applied in many studies. Gutjahr and Heinemann (2013) used the distribution correction with a combination of a Gamma and a Generalized Pareto distribution to correct the the modelled precipitation field from COSMO-CLM (Consortium for Small-scale Modeling (COSMO)-Climate Limited-Area modeling (CLM)). The values smaller than the $95th$ percentile are assumed to follow a Gamma distribution, whereas values larger than this threshold are assumed to follow a General Pareto distribution:

$$P^* = \begin{cases} F^{-1}_{obs,Gamma}(F_{RCM,Gamma}(P)) & \text{if } x < 95th \text{ percentile,} \\ F^{-1}_{obs,GPD}(F_{RCM,GPD}(P)) & \text{if } x \geq 95th \text{ percentile.} \end{cases} \tag{1.3}$$

The General Pareto distribution has three parameters: a shape parameter, a scale parameter and a threshold parameter. Together with the two parameters from the Gamma distribution, there are five parameters to be estimated. Similar to Gutjahr and Heinemann (2013), Yang et al. (2010) applied the distribution correction with a combination of two Gamma distributions to correct two RCM projections, $R3E5A1B_1$ and $R3E5A1B_3$ based on the ECHAM5 GCM (Roeckner et al., 2006). To capture the main properties of normal precipitation as well as extremes, the precipitation distribution was divided into two partitions separated by the $95th$ percentile and two Gamma distributions are fitted for each partitions separately. Different to the previous two methods, which divide the precipitation distribution into segments and fit separate distributions to each segment, Teng et al. (2015) introduced a mixed distribution mapping method. Instead of introducing arbitrary cut-offs, this method mixed two Gamma distributions and is interpreted as a two-state distribution and the PDF is expressed as follows:

$$f(x) = \lambda \frac{x^{a_1-1}e^{-x/b_1}}{b_1^{a_1}\Gamma(a_1)} + (1-\lambda)\frac{x^{a_2-1}e^{-x/b_2}}{b_2^{a_2}\Gamma(a_2)} \tag{1.4}$$

with $0 < \lambda < 1$. The parameter λ is the relative occurrence of the states, and, fitted correctly, the two gamma distributions represent rainfall occurring in high and low rainfall

states. The advantage of this approach compared to segmenting the distribution is that all parameters can be estimated simultaneously using maximum-likelihood estimation.

The aforementioned different types of quantile mapping methods are widely applied in bias correction of RCM derived precipitation (Dosio and Paruolo, 2011; Gudmundsson et al., 2012). Piani et al. (2010a) applied the Gamma distribution based quantile mapping to simulated daily precipitation fields over Europe from the Danish Meteorological Institute RCM. Bias corrections were calculated for the whole decade from 1961 to 1970, and applied to the decade from 1991 to 2000, in order to maximize the lag between construction estimation period and the application evaluation period. Results showed that the methodology performed satisfactorily not only for mean quantities but even for time dependent statistical properties, such as the number of consecutive dry days and the cumulative amount of rainfall for consecutive heavy precipitation days. Bárdossy and Pegram (2012) found that performing quantile-quantile (Q-Q) transforms to adjust RCM data to the characteristics of contemporaneous observed rainfall was more successful when conditioning the transforms on circulation patterns (CPs) than not. Themeßl et al. (2010) applied the quantile mapping that is based on point-wise and daily constructed empirical cumulative distribution functions of modelled and observed datasets in the calibration period. This is in contrast to other bias correction studies where theoretical CDFs are estimated only from wet days. Themeßl et al. (2011) introduced a new extrapolation of the error correction function enables quantile mapping to reproduce new extremes without deterioration and mostly with improvement of the original RCM quality. This study also introduced a frequency adaptation (FA) method in order to account for a methodological problem, which occurs if the dry-day frequency in the model result is greater than in the observations.

In this study, a Copula-based stochastic bias correction method is applied to correct each individual time step of a RCM simulation. This is different to the traditional transfer function-based statistical correction approaches. The strategy of this method is the identification and description of the underlying dependence structures between observed and modeled climate variables (precipitation) and its application for bias correction. It is known that the traditional measures of dependence (e.g. Pearson's correlation coefficient) can only capture the strength of the linear dependence as a single global parameter. Alternatively, Copulas are able to describe the complex nonlinear dependence structure between variables (Bárdossy and Pegram, 2009). Based on the identified dependence structure between observed and modeled precipitation and the identified respective marginal distributions, a set of realizations is finally obtained through Monte

Carlo simulations.

Copulas are used for various applications in hydrometeorology (e.g., Dupuis, 2007; Grégoire et al., 2008; Serinaldi, 2008; Bárdossy and Pegram, 2009, 2012). Gao et al. (2007) performed Copula-derived observation operators for assimilating soil moisture from satellite remote sensing into land surface models. Salvadori and Michele (2007) used the Copulas for several hydrological application: a general theoretical framework for studying the return period of hydrological events; a trivariate model for the temporal structure of the sequence of storms; an explicit derivation of the storm volume statistics. Bárdossy (2006) demonstrated the application of Copulas for the investigation of groundwater quality parameters: chloride, sulfate, pH, and nitrate.van den Berg et al. (2011) has applied a Copula-based method for downscaling spatial rainfall from coarse resolution. Copula-based bias correction techniques have been originally introduced by Laux et al. (2011) and Vogl et al. (2012), and are extended in this study by investigating gridded precipitation fields instead of individual and unevenly distributed stations. The work is published in Mao et al. (2015).

1.3 Objectives

This study aims to develop a stochastic bias correction method for RCM derived precipitation (WRF) through Copula-based assimilation of REGNIE observation data over Germany. The Copula models are estimated for each grid cell and then are applied for bias correction. The overall objectives of this study are:

- Fitting distributions to both REGNIE and WRF precipitation and revealing the spatial differences between them
- Describing the dependence structures between REGNIE and WRF data by using Copulas
- Applying the Copula-based correction in the overall model and the seasonal mode separately to investigate seasonal variability
- Comparing the Copula-based correction to the quantile mapping method
- Developing a methodological framework to apply the Copula-based approach to climate projections, i.e. periods, for which no observations are available.

1.4 Innovation

This study is an extension of the studies of Laux et al. (2011) and Vogl et al. (2012) by applying the Copula-based bias correction technique to high resolution RCM precipitation output and a gridded observation product. Compared to those two studies, this study is based on a framework as follows:

- Working on a grid cell base and to estimate the Copula model (marginal distributions and Copula function) for each grid cell separately rather than selecting e.g. the most dominant model. Therefore, the statistical characteristics of observed (REGNIE) and modelled data (WRF) and their dependence structure is visualized spatially and analyzed for the first time.
- Implementing the Bayesian Information Criterion in addition to the Kolmogorov–Smirnov test for the marginal Goodness-of-fit test. From previous studies we found that very large sample sizes may bias the result of the K–S test, leading to the rejection of the null hypothesis (the sample comes from the selected distribution) most of the time.
- In addition to overall application, the Copula models are also estimated and applied for every season separately. Thus, different types of precipitation geneses are not masked by the same models. This, in general, leads to stronger dependencies and robuster models. The results are also improved.
- A Markov-based precipitation cases identification model is introduced. The method is able to identify the precipitation cases in the future, where only climate projections are available.

1.5 Structure of the dissertation

The dissertation is divided into 9 chapters with respect to the scope of this study. After the introduction and innovation, the study area and data resources are described in chapter 2. In chapter 3, an overview on Copulas theory is presented. The framework of the Copula-based stochastic bias correction is introduced in chapter 4. Afterwards, the Copula-based correction is applied in the overall mode in chapter 5 and in the seasonal mode in chapter 6. The Copula-based approach is compared to the quantile mapping

method in chapter 7. In chapter 8, the Markov-based precipitation cases identification model is introduced. The final chapter is devoted to summary and conclusions and recommendations for further research.

Chapter 2

Study area and data resources

In this chapter the data sources which are used for the application of the Copula-based stochastic bias correction method for gridded data sets is described. The developed stochastic approach is applied for Germany (see Fig. 2.1) for a 30-year time period from 1971 to 2000. The 30-year time series is split into a calibration (1971–1985) and a validation (1986–2000) period of equal length. For the application of bias correction, two types of data sources are used: the gridded observed precipitation fields and the dynamical downscaled precipitation fields.

As observations, we use the 1 km gridded, daily data set REGNIE (REGionalisierung der NIEderschlagshöhen) from the German Weather Service (DWD, 2011). The dynamical downscaled precipitation fields are the result of a high resolution (7km) WRF-ARW simulation over Central Europe driven by ERA40 reanalysis data (Berg et al., 2013). For the grid cell based bias correction the 1 km REGNIE data set is up-scaled and remapped to the 7 km WRF grid such that precipitation amounts are conserved. Hence, the data set consists of daily, gridded precipitation fields with 11710 grid cells for REGNIE and WRF-ERA40 for the time period 1971 to 2000.

2.1 RCM data

Dynamically downscaled precipitation fields over Germany from a regional climate model simulation (RCM) are used (Berg et al., 2013). The RCM used in this study is the Weather Research and Forecasting (WRF) model with the Advanced Research WRF (ARW) dynamics solver version 3.1.1 (Skamarock et al., 2008). WRF is a non-hydrostatic

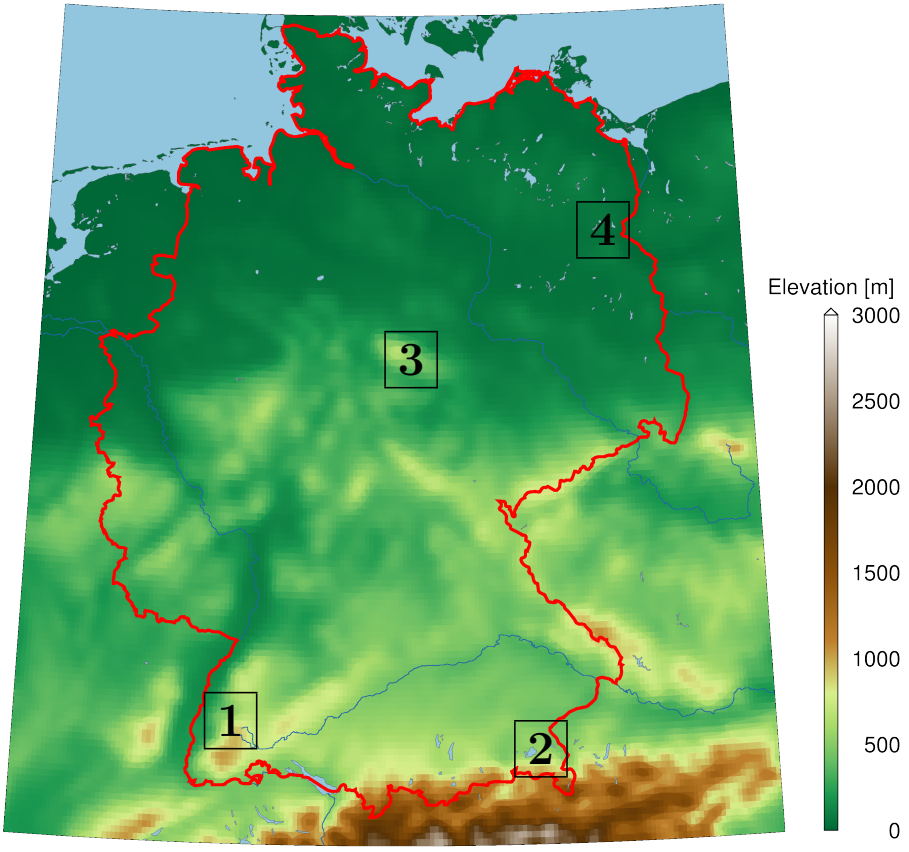

FIGURE 2.1: Terrain elevation of Germany (DEM). The numbers represent the position of the four specific grid cells for which the performance of the Copula-based algorithm is analyzed in Chapter 5.

model. For this data set, the WRF-ARW simulations are forced by ERA40 reanalysis data from 1971 to 2000 at the boundaries which implies large-scale circulation close to observations. ERA40 is an ECMWF (European Centre for Medium-Range Weather Forecasts) reanalysis of the global atmosphere and surface conditions for 45-years, over the period from September 1957 through August 2002 (Uppala et al., 2005).

The RCM is driven with atmospheric fields of temperature, wind and humidity, as well as sea surface temperature from the GCM. Due to the coarse resolution of the GCM, a double-nesting approach is applied in Lambert conformal map projection. The coarse nest extends over all of Europe (42 km) and the fine nest covers Germany and the near surroundings (7 km). The model uses 40 vertical levels for both nests. The applied setup uses the following main physical options:

- the WRF Single-Moment 5-class scheme (WSM5) microphysical parameterisation (Hong et al., 2004; Hong and Lim, 2006),

- the modified version of the Kain-Fritsch scheme for cumulus parameterisation (Kain, 2004),
- the Noah land surface model (Chen and Dudhia, 2001),
- the Yonsei University (YSU) parameterisation for the planetary boundary layer (Hong et al., 2006),
- the MM5-Dudhia SW scheme (Dudhia, 1989),
- the Rapid Radiative Transfer Model (RRTM) long-wave radiation scheme (Mlawer et al., 1997).

For further details on the applied WRF-ARW setup we refer to Berg et al. (2013) and the references listed therein.

2.2 Observational data

As observations, we used the 1 km gridded daily data set REGNIE from the German Weather Service (DWD) (DWD, 2011). The REGNIE product is available for complete Germany from 1951 to the present and the number of underlying stations is approximately 2000 stations. The statistical gridding approach of station data is based on the spatial interpolation of anomalies compared to long-term mean values. The statistical gridding approach for REGNIE can be described as follows:

- calculates a background climatological field with a multi-linear regression approach where the geographical position, elevation and wind exposure of the stations are taken into account,
- for the calculation of the daily precipitation fields, station values are first assigned to a grid point and divided by the background data to calculate anomalies,
- spatially interpolates the anomalies using inverse distance weighted interpolations,
- multiplying the results by the background field.

For the grid cell based bias correction the 1 km REGNIE data set is up-scaled and remapped to the 7 km WRF grid such that precipitation amounts are conserved.

Chapter 3

Copula theory

The word *Copula* is a Latin noun that means to join, connect or link. It is first employed in a mathematical or statistical sense by Sklar (1959) in Sklar's Theorem to describe the function that join together one-dimensional distribution functions to form a multivariate distribution function. Traditionally, the joint distribution between variables are modelled by classical multivariate distributions e.g. the normal and the log-normal distribution. The main limitation of such approaches is that the individual behavior of each variable as well as the joint dependence between them are characterized by the same parameter (Genest and Favre, 2007; Salvadori and Michele, 2007). The advent of Copula, however, allows us to avoid this restriction and also not limited by the Gaussian assumption.

3.1 Sklar's theorem

The Sklar's theorem (Sklar, 1959) is central of the theory of Copulas and is the foundation of many applications. It explicates the role that Copulas play in the relationship between multivariate distribution functions and their univariate marginal distributions (Nelsen, 1999). The Sklar's Theorem is as follows:

Let H be a joint distribution function with margins F and G. Then there exists a Copula C such that for all x, y in $\mathbb{R}$,

$$H(x, y) = C(F(x), G(y)) \tag{3.1}$$

If F and G are continuous, then C is unique; otherwise, C is uniquely determined on $RanF \times RanG$, where Ran denotes range. Conversely, if C is a Copula and F and G are distribution functions, then the function H defined by Eq. 3.1 is a joint distribution function with margins F and G.

It is important to remark that the Sklar's theorem is demonstrated by Sklar (1959) in a bivariate case, which is definitely possible to extent to multivariate cases. Therefore, a n-dimensional joint distribution function $H(x_1, \cdots, x_n)$ can be re-written as a Copula and its margins $F_1, \cdots, F_n$.

$$H(x_1, \cdots, x_n) = C(F_1(x_1), \cdots, F_n(x_n)) \tag{3.2}$$

Conversely, a Copula can also be presented by its joint distribution H and the corresponding inverse of the marginal distributions $F_1^{-1}, \cdots, F_n^{-1}$.

$$C(u_1, \cdots, u_n) = H(F_1^{-1}(u_1), \cdots, F_n^{-1}(u_n)) \tag{3.3}$$

As for general distribution functions, the probability density function (PDF) of a Copula is obtained by differentiating with respect to all variables (Melchiori, 2003):

$$c(u_1, \cdots, u_n) = \frac{\partial^n C(u_1, \cdots, u_n)}{\partial u_1 \cdot \partial u_2 \cdots \partial u_n} \tag{3.4}$$

The density function of the joint distribution H is then can be expressed in terms of a Copula PDF c and their marginal PDFs.

$$h(x_1, \cdots, x_n) = c(F_1(x_1), \cdots, F_n(x_n)) \prod_{i=1}^{n} f_n(x_n) \tag{3.5}$$

where $f_1, \cdots, f_n$ are the PDFs of the corresponding marginals $F_1, \cdots, F_n$.

As a consequence of Sklar's theorem, each multivariate joint distribution can be expressed in term of a Copula function and its margins. The Copula can be regarded as a functional link between the multivariate joint distribution and its univariate marginal distributions. As the marginal distributions give an exhaustive description of random variables taken separately, the joint dependence between these variables is fully and uniquely characterized by the Copula C. In other words, Copula describes the pure dependence structure between variables independent of its margins (Joe, 1997). Therefore

the fitting of the traditional joint distribution can now be separated into two independent steps: the fitting of the marginal distributions and the estimation of the Copula function. This provides higher flexibility to describe joint behavior between variables compared to traditional methods, e.g. multivariate Gaussian distribution.

3.2 Properties of Copulas

A n-dimensional Copula is defined as a multivariate probability distribution on the n-dimensional unit cube I^n with margins that are uniform on I:

$$C : [0,1]^n \to [0,1] \tag{3.6}$$

$$C(u_1, \cdots, u_n) = Pr(U_1 \leq u_1, \cdots, U_n \leq u_n) \tag{3.7}$$

where C is an n-dimensional Copula with multivariate random vector $U(U_1, \cdots, U_n)$ whose margins are $u(u_1, \cdots, u_n)$. The general properties of Copulas can be summarized as follows (Nelsen, 1999):

- $C(u_1, \cdots, u_n) = u_i$ whenever $\forall j \neq i, u_j = 1$;
- $C(u_1, \cdots, u_n) = 0$ whenever $0 \in u1, \cdots, u_n$;
- C is n–increasing, i.e., $\forall \mathbf{x}, \mathbf{y} \in [0,1]^n, x_i \leq y_i, i = 1, \cdots, n$, it holds

$$\sum_{J \subset \{1, \cdots, n\}} (-1)^{|J|} C(u_1^J, \cdots, u_n^J) \geq 0, \text{ where } u_i^J = \begin{cases} x_i & \text{if } i \in J, \\ y_i & \text{if } i \notin J. \end{cases} \tag{3.8}$$

Furthermore, another important property is that Copulas are invariant under any strictly monotonic increasing transformation of the variables. This means that the variables transformed by any monotonic increasing functions will not effect its Copula. That is, if $X_1, \cdots, X_n$ are continuous random variables with copula C and $\Psi_1, \cdots, \Psi_n$ are monotonic increasing functions on RanX_1, $\cdots$, RanX_n, then $\Psi_1(x_1), \cdots, \Psi_n(x_n)$ have the same copula C. This invariant property of Copulas is explored in more detail in the following as it is fundamental to the discussion in the following chapters.

Assume $F_1, \cdots, F_n$ to be the distribution functions of random variables $X_1, ..., X_n$, respectively. Consider monotonic transformations of the random variables $\Psi(X_1), \cdots, \Psi(X_n)$

with their corresponding marginals $G_1, \cdots, G_n$. Let C and C_Ψ be the copulas of $X_1, \cdots, X_n$ and $\Psi(X_1), \cdots, \Psi(X_n)$, respectively. The strictly increasing property of Ψ indicates that for any $x \in \mathbb{R}$ (Embrechts et al., 2003).

$$G(x) = Pr(\Psi(X) \leq x) = Pr(X \leq \Psi^{-1}(x)) = F(\Psi^{-1}(x)) \tag{3.9}$$

and thus:

$$\begin{aligned} C_\Psi(G_1(x_1), \cdots, G_n(x_n)) &= Pr(\Psi_1(X_1) \leq x_1, \cdots, \Psi_n(X_n) \leq x_n) \\ &= Pr(X_1 \leq \Psi_1^{-1}(x_1), \cdots, X_n \leq \Psi_n^{-1}(x_n)) \\ &= C(F_1(\Psi_1^{-1}(x_1)), \cdots, F_n(\Psi_n^{-1}(x_n))) \\ &= C(G_1(x_1), \cdots, G_n(x_n)) \end{aligned} \tag{3.10}$$

Equation (3.10) confirms that $C = C_\Psi \in I^n$. This is a great advantage in simulations as the variables may belong to different probability distributions and applying transformation functions may be required to obtain the right marginals.

3.3 The Fréchet-Hoeffding bounds for Copulas

As mentioned above, Copulas are multivariate probability distributions within the unit cube I^n. The Fréchet-Hoeffding Theorem (Fréchet, 1951; Fisher and Sen, 1994) states that for any Copula $C : [0,1]^n \to [0,1]$ and any $(u_1, \cdots, u_n) \in [0,1]^n$ the following bounds hold:

$$W(u_1, \cdots, u_n) \leq C(u_1, \cdots, u_n) \leq M(u_1, \cdots, u_n). \tag{3.11}$$

The function W is called the lower Fréchet-Hoeffdin lower bound that is given by

$$W(u_1, \cdots, u_n) = max\{0, \sum_{i=1}^{n} u_i - (n-1)\}, \tag{3.12}$$

The function M is called the upper Fréchet-Hoeffdin bound and is defined as

$$M(u_1, \cdots, u_n) = min\{u_1, \cdots, u_n\}. \tag{3.13}$$

The upper Fréchet-Hoeffdin bound M is also a Copula itself and describes the comonotone dependence of random variables $U_1, \cdots, U_n$. For the lower Fréchet-Hoeffdin bound

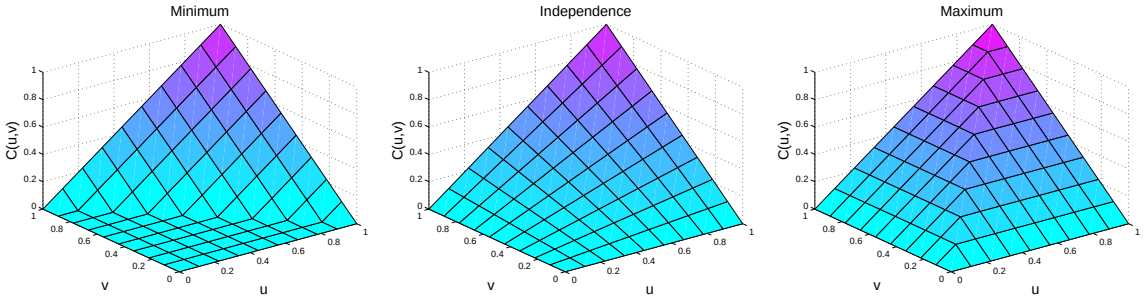

FIGURE 3.1: $3-D$ shaded surface graphs of the bivariate Fréchet-Hoeffding bounds and of the independence Copula (in the middle).

W, it is a Copula only if $n = 2$, and in that case describes the countermonotone dependence of random variables U_1 and U_2 (Nelsen, 1999). Therefore, in two dimensions the Fréchet-Hoeffdin bounds M and W are also known as the comonotonicity Copula and the countermonotonicity Copula, respectively.

For the bivariate case, the FréchetHoeffding Theorem states

$$max\{u + v - 1, 0\} \leq C(u, v) \leq min\{u, v\} \tag{3.14}$$

A third important Copula is the product Copula $\Pi(x_1, \cdots, x_n) = \prod_{i=1}^{n} x_i$ that describes their independence and it is also known as the independence Copula.

In order to illustrate these three fundamental Copulas in more detail, the $3-D$ shaded surface plots of them are shown in Fig. 3.1. The Fréchet-Hoeffdin bounds imply that all bivariate copulas lie between the surface represented by the minimum Copula (W) and the surface represented by the maximum Copula (M).

3.4 Empirical Copulas

Copulas are used to describe the dependence structure between variables. There are many parametric copula families available, which usually have parameters that control the strength of dependence. Due to the reason that the underlying theoretical Copula (i.e. dependence structure) between variables is general not known in advance, is necessary to study the empirical Copulas (Deheuvels, 1979). Since the bias correction frame used in this study is based on bivariate Copulas, therefore in the following the Copulas refer to bivariate Copulas only.

The empirical Copula $C_n(u, v)$, which is defined on the rank space, is an estimator for the unknown theoretical Copula. Let $\{r_1, \cdots r_n\}$ and $\{s_1, \cdots, s_n\}$ denote the rank values of the variables $\{x_1, \cdots, x_n\}$ and $\{y_1, \cdots, y_n\}$. Then the empirical Copula is defined as:

$$C_n(u, v) = 1/n \sum_{t=1}^{n} \mathbf{1}(\frac{r_i}{n+1} \leq u, \frac{s_i}{n+1} \leq v) \tag{3.15}$$

with $1(A)$ denoting the indicator function of set A. In some literature the empirical Copula is also described as the best sample-based representation of the theoretical Copula C, which is itself a characterization of the dependence in a pair (X, Y) (Genest and Favre, 2007).

3.5 Copula families

Copulas from different families describe different types of dependence structures. The two most commonly used Copula families are the Elliptical and the Archimedean Copula families. The Elliptical Copulas, e.g. the Gaussian Copula and the Student Copula, are constructed from elliptical distributions and show symmetrical upper/lower tail dependence structures.

In contrast, the Archimedean Copulas are able to describe asymmetrical tail dependence structures and are defined as following. Let φ be a continuous, strictly decreasing function from I to $[0, +\infty]$ such that $\varphi(1) = 0$, and let $\varphi^{[-1]}$ be the pseudo-inverse of φ defined as:

$$\varphi^{[-1]}(t) = \begin{cases} \varphi^{-1}(t) & \text{if } 0 \leq t \leq \varphi(0), \\ 0 & \text{if } \varphi(0) \leq t \leq +\infty, \end{cases} \tag{3.16}$$

then the function

$$C(u, v) = \varphi^{[-1]}(\varphi(u), +\varphi(v)) \tag{3.17}$$

defines a Copula only if φ is convex and φ is called the generator of the Archimedean Copula C. In the following, only the Copula families that are employed in this work are discussed. For additional information regarding different Copula families, the reader is referred to Nelsen (1999) and Joe (1997).

3.5.1 Gaussian Copula

The Gaussian copula (also known as normal Copula), derived from the multivariate normal distribution, is the most commonly used Copula family due to its simplicity. The two-dimensional multivariate Gaussian Copula is defined as:

$$C = \Phi_R(\Phi^{-1}(u), \Phi^{-1}(v)) \tag{3.18}$$

where R indicates the correlation matrix $\begin{bmatrix} 1 & \theta \\ \theta & 1 \end{bmatrix}$ with the linear correlation coefficient $\theta \in [-1, 1]$ and Φ^{-1} is the inverse of the univariate standard normal distribution function.

- If $\theta = 1$, then C equals the two-dimensional comonotonicity Copula M.
- If $\theta = -1$, then C equals the countermonotonicity Copula W.
- If $\theta = 0$, then C equals the independence Copula Π.

The Gaussian Copula can be thought of as a dependence structure that interpolates between perfect positive and perfect negative dependence, where $|\theta|$ represents the strength of the dependence. It is a symmetric Copula and is showing no tail dependence. If the marginal distributions $u = F(x), v = G(y)$ are normal, then the random vector (x, y) has a bivariate normal distribution.

3.5.2 Clayton Copula

The Clayton Copula is an asymmetric Archimedean Copula, exhibiting lower tail dependence. This Copula is given by:

$$C(u, v) = max[(u^{-\theta} + v^{-\theta} - 1), 0]^{-\frac{1}{\theta}} \tag{3.19}$$

where $\theta \in [-1, +\infty) \backslash \{0\}$. The generator of the Clayton Copula is:

$$\varphi(t) = \frac{1}{\theta}(t^{(-\theta)} - 1) \tag{3.20}$$

For our application $0 < \theta < +\infty$, i.e. only positive dependence could be found, The Clayton Copula can then be simplified to

$$C(u, v) = (u^{-\theta} + v^{-\theta} - 1)^{-\frac{1}{\theta}} \qquad \theta \in (0, +\infty) \tag{3.21}$$

In the limit as $\theta \to 0$, it approach the independence Copula Π, and as $\theta \to +\infty$ it approach the two-dimensional comonotonicity copula M. The Clayton Copula has a remarkable invariance under truncation, i.e. the truncated Copula on the sub-region is the same as the Copula for the entire area (Oakes, 2005).

3.5.3 Gumbel Copula

The Gumbel Copula (also known as Gumbel-Hougard Copula) is an asymmetric Archimedean Copula, exhibiting upper tail dependence. The Gumbel Copula is given by:

$$C(u,v) = e^{-((-\ln(u)^\theta)+(-\ln(v)^\theta))^{\frac{1}{\theta}}} \tag{3.22}$$

where $\theta \in [1, +\infty)$. The generator of the Gumbel Copula is:

$$\varphi(t) = (-\ln(t))^\theta \tag{3.23}$$

If $\theta = 1$, we obtain the independence Copula Π as a special case, and the limit of C as $\theta \to +\infty$ is the two-dimensional comonotonicity Copula M. Thus, the Gumbel Copula interpolates between independence and perfect positive dependence and the parameter θ represents the strength of the dependence.

3.5.4 Frank Copula

The Frank Copula is also an Archimedean Copula and is known to be symmetrical, i.e. it shows no tail dependence. The Frank Copula is given by:

$$C(u,v) = -\frac{1}{\theta}\ln(1 + \frac{(e^{-\theta u}-1)(e^{-\theta v}-1)}{e^\theta - 1}) \tag{3.24}$$

where $\theta \in (-\infty, +\infty)\backslash\{0\}$. The generator of the Frank Copula is:

$$\varphi(t) = -\ln(\frac{e^{-\theta t}-1}{e^\theta - 1}) \tag{3.25}$$

As θ approaches 0 the Frank Copula approaches the independence Copula Π, and as θ approaches $+\infty$ $(-\infty)$, it approaches the comonotonicity Copula M (the countermonotonicity Copula W).

3.6 Scatter plots of random sampling from Copulas

As mentioned above that the Copulas from different families are able to describe different dependence structures and the parameter controls the strength of the dependence. In order to illustrate it more intuitively, a few scatter plots of random sampling from different Copulas are shown in Fig. 3.2. Four different Copulas (Gaussian, Clayton, Frank and Gumbel) that are employed by this study are used for a random sampling of size = 1000. From the scatter plots (Fig. 3.2) it can be seen that the dependence structures represented by these Copulas are different. The Gumbel Copula is able to describe an upper tail dependence structure, while the Clayton Copula allows to express higher probability in the lower tail. The Frank Copula exhibits no tail dependence, and the Gaussian Copula describes a similar dependence as the Frank Copula, but with slightly higher densities in the lower and upper tails (Venter, 2002; Schmidt, 2007). While the parameter increases, the strength of the dependence also increases.

3.7 Dependence measures

In this section, three kinds of dependence measures for bivariate random vectors will be discussed: the Pearson's correlation coefficient, rank correlation coefficient and the coefficient of tail dependence. All of these dependence measures are sample-based non-parametric measure and yield a scalar measurement for a pair of random variables (X, Y), although the nature and properties of the measures are different in each case. The Pearson's correlation coefficients measures the dependence between two random variables in the data space, while the other kinds of dependence measures, i.e. rank correlations and tail dependence, are rank-based dependence measures and depend only on the underlying Copula of random variables. In contrast to ordinary correlation, these rank-based dependence measures are functions of the Copula only and can thus be used in the parametrization of Copulas.

3.7.1 Pearson's correlation coefficient

Pearson's correlation coefficient γ is well known and plays a central role in statistical theory. It measures the linear dependence between random variables (X, Y) and is

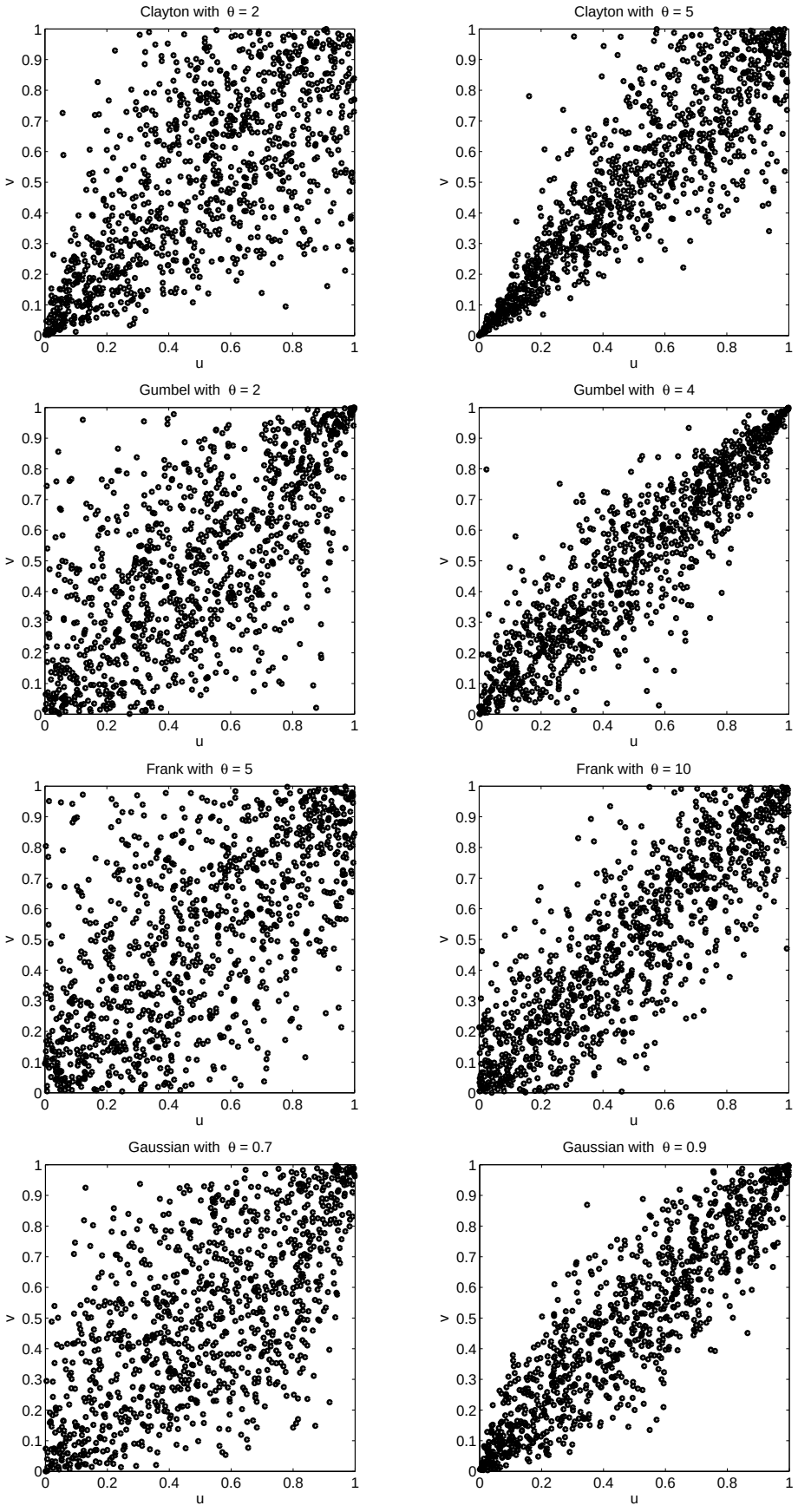

FIGURE 3.2: The scatter plots of random sampling from different Copulas with different parameters. The sample size is 1000 and the Copula functions are shown in Sect. 3.5

defined as:

$$\rho = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \sum_{i=1}^{n}(y_i - \overline{y})^2}} \tag{3.26}$$

where $\overline{x}$ and $\overline{y}$ are the sample means of X and Y,

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \tag{3.27}$$

It takes its values in $[1, 1]$ and, if X and Y are independent, then $\gamma(X, Y) = 0$, but the converse is not true because the correlation coefficient detects only linear dependence between two variables. The linear correlation is invariant under strictly increasing linear transformations. However, it is not invariant under nonlinear strictly increasing transformations. That is, if T_1 and T_2 are the strictly increasing functions, $\gamma(T_1(X), T_2(Y)) \neq \gamma(X, Y)$.

3.7.2 Spearman's Rho

Spearman's rho is one of the two well-known measures of rank correlation. Given a random sample $\{(x_1, y_1), \cdots, (x_n, y_n)\}$ from (X, Y), $\{(r_1, s_1), \cdots, (r_n, s_n)\}$ are the corresponding rank values. Spearman's rho is given by

$$\rho = \frac{\sum_{i=1}^{n}(r_i - \overline{r})(s_i - \overline{s})}{\sqrt{\sum_{i=1}^{n}(r_i - \overline{r})^2 \sum_{i=1}^{n}(s_i - \overline{s})^2}} \tag{3.28}$$

where

$$\overline{r} = \frac{1}{n}\sum_{i=1}^{n} r_i = \frac{n+1}{2} = \frac{1}{n}\sum_{i=1}^{n} s_i = \overline{s} \tag{3.29}$$

Mimicking the familiar approach of Pearson to the measurement of dependence, Spearman's rho also computes the correlation between the samples but with their rank values. In other words, Spearman's rho is simply the linear correlation of ranks of random variables (or the probability transformed random variables). However, Spearman's rho (ρ) is theoretically far superior to Pearson's classical correlation coefficient (γ). Spearman's

rho is a rank correlation that depend only on the underlying Copula of a bivariate distribution and not on the marginal distributions, unlike linear correlation which depends on both. It is expressible in terms of underlying Copula $C : I^2 \to I = [0, 1]$ as follows (Schweizer and Wolff, 1981):

$$\rho = 12 \iint_{I^2} C(u, v) \mathrm{d}u\mathrm{d}v - 3 = 12 \iint_{I^2} uv \mathrm{d}C(u, v) - 3 \tag{3.30}$$

where $\mathrm{d}C$ denotes the doubly stochastic measure induced on I^2 by C (Fredricks and Nelsen, 2007). If the Copula family (C_θ) is given, the Copula parameter then can be directly calculated from the Spearman's rho (ρ). Furthermore, Spearman's rho has more appealing properties than the linear correlation γ (Embrechts et al., 2002; Genest and Favre, 2007):

1. $E(\rho) = \pm 1$ occurs if and only if X and Y are functionally dependent, i.e., whenever their underlying copula is one of the two FréchetHoeffding bounds, M or W
2. In contrast, $E(\gamma) = \pm 1$ if and only if X and Y are linear functions of one another, which is much more restrictive
3. ρ estimates a population parameter that is always well defined, whereas there are heavy-tailed distributions (e.g. Cauchy distribution) for which a theoretical value of Pearson's correlation does not exist.

3.7.3 Kendall's Tau

Another well-known rank correlation measure is the Kendall's tau. Again let $\{(x_1, y_1), \cdots, (x_n, y_n)\}$ denote a random sample of n observations from a vector (X, Y) The second well-known nonparametric measure of dependence is Kendall's tau (τ). It is given by

$$\tau = \frac{P_n - Q_n}{P_n + Q_n} = (P_n - Q_n) / \binom{n}{2} = \frac{4}{n(n-1)} P_n - 1 \tag{3.31}$$

where P_n denote the number of concordant pairs and Q_n the number of discordant pairs. Here, two pairs $(x_i, y_i), (x_j, y_j)$ are said to be concordant when $(x_i x_j)(y_i - y_j) > 0$, and discordant when $(x_i x_j)(y_i - y_j) < 0$.

Similarly, Kendall's tau can also be expressed in terms of Copula C (Nelsen, 1999):

$$\tau = 4 \iint_{I^2} C(u,v)\mathrm{d}C(u,v) - 1 = 1 - 4 \iint_{I^2} \frac{\partial C}{\partial u}(u,v)\frac{\partial C}{\partial v}(u,v)\mathrm{d}u\mathrm{d}v \tag{3.32}$$

Therefore, the Copula parameter can also be directly calculated from Kendall's tau (τ).

For many joint distributions these two measures (ρ and τ) have different values, as they measure different aspects of the dependence structure. In terms of dependence properties, Spearman's rho is a measure of average quadrant dependence, while Kendall's tau is a measure of average likelihood ratio dependence (Nelsen, 1992).

3.7.4 Tail dependence

The two nonparametric measures of dependence (ρ and τ) in ranks introduced above measure the average of the dependence. Another measure of the dependence is so-called tail dependence, which measures the dependence between the variables in the upper-right quadrant and in the lower-left quadrant of I^2.

Let X and Y be continuous random variables with distribution functions F and G, respectively. The upper tail dependence parameter λ_U is the limit (if it exists) of the conditional probability that Y is greater than the $100t$-th percentile of G given that X is greater than the $100t$-th percentile of F as t approaches 1, i.e.

$$\lambda_U = \lim_{x \to +1^-} P[Y > G^{(-1)}(t) | X > F^{(-1)}(t)] \tag{3.33}$$

Similarly, the lower tail dependence parameter λ_L is the limit (if it exists) of the conditional probability that Y is less than or equal to the $100t$-th percentile of G given that X is less than or equal to the $100t$-th percentile of F as t approaches 0, i.e.

$$\lambda_L = \lim_{x \to +0^+} P[Y \leq G^{(-1)}(t) | X \leq F^{(-1)}(t)] \tag{3.34}$$

These parameters (λ_U and λ_L) are also nonparametric and depend only on the Copula of X and Y, since they are rank based measure of the dependence. Therefore, the upper and lower tail dependence parameters of the random vector (X, Y) with the Copula C, can be defined as follows (Joe, 1997):

$$\lambda_U = \lim_{x \to +1^-} \frac{1 - 2u + C(u,u)}{1 - u} \tag{3.35}$$

and

$$\lambda_L = \lim_{x \to +0^+} \frac{C(u,u)}{u} \tag{3.36}$$

The upper tail dependence expresses the probability occurrence of positive large values (outliers) at multiple locations jointly, while the lower tail dependence expresses the the probability occurrence of positive small values.

Chapter 4

Copula-based stochastic bias correction framework

The bias correction framework used in this study is based on Copula theory. A bivariate Copula model forms the basis of this stochastic bias correction algorithm. The Copula model consists of two respective marginal distributions and a bivariate Copula function and is then used to generate bias corrected WRF data by conditional stochastic sampling.

As already mentioned above in Sec. 3.1, Sklar's theorem allows to separate the multivariate joint distribution estimation into individual marginal distribution estimation and the Copula (dependence structure) estimation independently. Which is rather flexible to describe the joint behavior between variables with full freedom to the choice of the univariate marginal distributions and the Copulas. This is especially advantageous in cases where the dependence structure between the variates is too complex to be modelled by e.g. a multivariate Gaussian distribution, as it is often the case for hydrometeorological variables (Salvadori and Michele, 2007; Dupuis, 2007). In this study, following Sklar's theorem, a so called bivariate Copula model is structured to describe the joint behavior between REGNIE and WRF data. It is then used to generate bias corrected WRF data by Copula based conditional stochastic sampling.

4.1 A bivariate Copula model

Suppose that the realizations $(x_1, y_1), \cdots, (x_n, y_n)$ are given from a pair of random variates (X, Y), and that it is desired to identify the bivariate distribution $F_{XY}(x, y)$

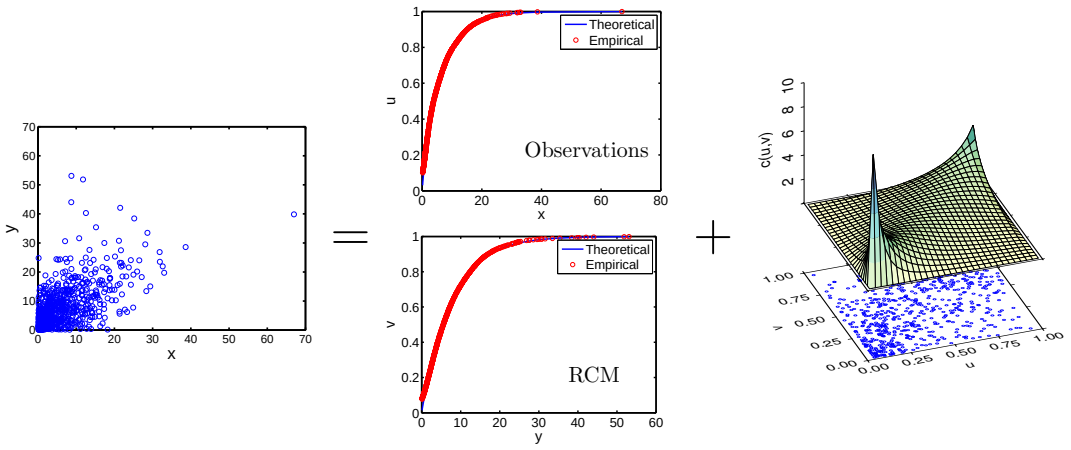

FIGURE 4.1: Visualization of a bivariate Copula model consisting of two marginal distributions and a theoretical Copula function that describes the pure dependence.

that characterizes their joint behavior. In a view of Sklars theorem, a bivariate Copula model can be applied. The bivariate Copula model of the variates X and Y consists of two univariate marginal distributions ($F_X(x)$ and $F_Y(y)$) and a Copula function $C(u, v)$. The marginal distributions describe the statistical properties of the variates (X and Y) and the Copula captures the dependence structure between them. The Copula model ($F_X(x)$, $F_Y(y)$ and $C(u, v)$) can be estimated separately based on the realizations x, y. Figure 4.1 visualizes the process of estimating a Copula model with a bivariate exemplary data set, i.e. realizations (x, y) of the two random variates X and Y.

A scatter plot of the two realizations (x, y) is shown in Fig. 4.1 (left). To model the joint behavior by using a Copula model, the first step is to fit a marginal distribution function for the two variates X and Y, respectively (see Fig. 4.1, middle). The realizations (x, y) are then transformed from the data space to the rank space (u, v) based on the fitted marginal distributions. The next step is to estimate the Copula function C from the ranked values (u, v) (see Fig. 4.1, right). Here a Copula PDF is used instead of Copula CDF as the PDF is more illustrative. Finally, the unknown joint distribution $F_{XY}(x, y)$ is fully determined by the marginal distributions and the Copula function, i.e. the dependence structure itself (Grégoire et al., 2008). Figure 4.1 visualizes the fact that different marginal distributions and Copula functions can be combined independently allowing to model highly complex dependence structures between the variables X and Y. This is especially beneficial if these dependence structures are non-linear, asymmetric or the data show heavy-tail behavior.

4.2 Marginal distributions estimation

The Copula based modeling of the joint behavior between the variates X and Y requires the fitting of suitable marginal distributions for both data sets (REGNIE and WRF) for each grid cell. Generally, both non-parametric and parametric fitting approaches for the local precipitation distribution are found in the literature (Dupuis, 2007; Gao et al., 2007; Bárdossy and Pegram, 2009; van den Berg et al., 2011). The difference between them is that the non-parametric fitting can be regard as the perfect parametric fitting, but also with the shortcoming that it is not able to sample the new extremes outside the range of the calibration period (Themeßl et al., 2011). Another drawback of the empirical distribution fitting is that it depends on many degrees of freedom and may not be stationary due to possible overfitting. In this study, a parametric fitting of the precipitation distribution is applied as it allows also an illustration of the spatially distributed differences (provided as the fitted marginal distribution family maps) between WRF and REGNIE. This gives additional valuable information about differences in their statistical properties.

4.2.1 Theoretical marginal distribution candidates

In this study, five different theoretical distribution functions are selected as the candidates and their probability density functions are listed below.

1. The Weibull distribution with scale parameter $a > 0$ and shape parameter $b > 0$.

$$f(x) = \frac{b}{a}\left(\frac{x}{a}\right)^{b-1} e^{-(x/a)^b} \qquad -\infty \leq x + \infty \tag{4.1}$$

2. The Gamma distribution with shape parameter $a > 0$ and scale parameter $b > 0$, where $\Gamma(\cdot)$ is the Gamma function.

$$f(x) = \frac{1}{b^a \Gamma(a)} x^{a-1} e^{\frac{-x}{b}} \qquad x \geq 0 \tag{4.2}$$

3. The Normal distribution with mean μ and standard deviation σ.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \qquad -\infty \leq x + \infty \tag{4.3}$$

4. The Generalized Pareto distribution with shape parameter k, scale parameter σ and the threshold parameter θ.

$$f(x) = \begin{cases} (\frac{1}{\sigma})(1 + k\frac{(x-\theta)}{\theta})^{-(\frac{1}{k+1})} & \text{if } k > 0, x > \theta \\ (\frac{1}{\sigma})(1 + k\frac{(x-\theta)}{\theta})^{-(\frac{1}{k+1})} & \text{if } k < 0, \theta < x < -\frac{\sigma}{k} \\ (\frac{1}{\sigma}e^{-\frac{(x-\theta)}{\sigma}}) & \text{if } k = 0, x > \theta \end{cases} \tag{4.4}$$

5. The Exponential distribution with mean $\mu > 0$.

$$f(x) = \frac{1}{\mu}e^{\frac{-x}{\mu}} \qquad x > 0 \tag{4.5}$$

Their cumulative distribution function (CDF) can then be simply derived by taking the integration of the PDFs over X.

4.2.2 The goodness-of-fit testing of marginal distributions

The Goodness-of-fit (GoF) is evaluated in a two-stage process. Firstly, a Kolmogorov–Smirnov test (K–S test) is applied (Massey, 1951). As the K–S test is highly sensitive due to the large sample sizes (Serinaldi, 2008), the null hypothesis (the sample comes from the selected distribution) is rejected in some cases for all of the candidates. In other cases there might be more than one possible candidate for the best fit. For that reason, all candidates which are accepted by the K–S test are further inspected by using the Bayesian Information Criterion (BIC) (Weakliem, 1999). If all of the candidates are rejected by the K–S test, only the BIC is relevant for the selection of the best fit. When fitting the distribution to the data (REGNIE and WRF), a standard maximum likelihood estimation (MLE) is applied to estimate the parameters of the respective distribution functions (Myung, 2003).

The Kolmogorov–Smirnov test (Massey, 1951) is a nonparametric test of the null hypothesis that the population CDF of the data is equal to the hypothesized CDF (i.e. the sample data comes from a population with the hypothesized distribution). The test statistic quantifies a distance between the empirical distribution function of the sample and the cumulative distribution function of the reference distribution, which is the maximum absolute difference between them:

$$D = \max_x(|F_n(x) - F(x)|) \tag{4.6}$$

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(X_i \leq x) \tag{4.7}$$

where $\mathbf{1}(A)$ denoting the indicator function of set A and $F_n(x)$ is the empirical distribution function of X. $F(x)$ is the hypothesized theoretical distribution function of variable X. The K–S test then compares the distance statistic D to the critical values that computed by using an approximate formula or by interpolation in a table at a given level of significance (Massey, 1951). If the statistic D exceed the critical value, the hypothesis that the sample data comes from this theoretical distribution is rejected. In contrast the rejection failed.

The Bayesian Information Criterion is a measure of the relative quality of a statistical model for a given set of data. It is based on the likelihood function and closely related to the Akaike information criterion (AIC), but with a increased penalty term that avoids the overfitting. BIC provides a means for model selection within a finite set of models and the formula is as follows:

$$\mathrm{BIC} = k\ln(n) - 2\ln(L), \tag{4.8}$$

where k denotes the number of the free parameters of the model, n is the sample size and L is the maximized value of the likelihood function of the estimated model. The smallest value of the BIC suggests the best fitting of the distribution.

4.2.3 Evaluation of the quality of the Goodness-of-fit tests

In this study it was found that the K–S test rejects the null hypothesis for some of the gird cells in the domain and only the BIC is actually contributed to the results of the Goodness-of-fit test. Therefore, a further step is made to evaluate the quality of this Goodness-of-fit test. First, the residual sum of squares (RSS) is calculated for each grid cell. Then, the grid cells that with relatively high RSS are inspected visually by picturing and comparing their empirical CDFs and fitted theoretical CDFs. Furthermore, the quantile-quantile plot is also inspected.

The RSS is also known as the sum of squared residuals (SSR) or the sum of squared errors of prediction (SSE) in some statistical literature. It is a measure of the discrepancy between the data and an estimation model. A small RSS indicates a tight fit of the model

to the data. The RSS is given by

$$RSS = \sum_{i=1}^{n} (F_n(x_i) - F(x_i))^2 \tag{4.9}$$

where $F_n(x)$ is the empirical CDF of X (see Equation 4.7) and $F(x)$ is the fitted theoretical marginal distribution.

The quantile-quantile or Q-Q plot is an exploratory graphical device used to check the validity of a distributional assumption for a data set. In general, the basic idea is to compute the theoretically expected value for each data point based on the distribution in question. If the data indeed follow the assumed distribution, then the points on the q-q plot will fall approximately on a straight line. If the points in a q-q plot depart from a straight line, then the assumed distribution is called into question.

4.3 Copula function estimation

As mentioned in Sect. 3.5, the Copulas from different families represent different dependence structures. To increase the accuracy of the description of the dependence structures, different types of Copulas are considered, since one common Copula might be incapable to capture the dependence structure for all grid cells over the entire study area and for all seasons. In this study, four different one-parametric Copulas are selected as the candidates (see Sect. 3.5). They are the Gumbel, Frank, Clayton Copulas which are from the Archimedean Copula family and the Gaussian Copula that is from Elliptical Copula families.

The goodness-of-fit (GoF) test for Copulas are applied by comparing theoretical Copulas to the empirical Copulas such that the type of Copula whose dependency structure best characterizes the training data will be selected. There are different Goodness-of-fit tests available. A review and comparison of goodness-of-fit procedures is given by Genest et al. (2009). The Goodness-of-fit test used in this study is one of the so-called "blanket" tests, that is, rank-based procedures requiring no parameter tuning or other strategic choices such as kernel, bandwidth, etc. The test is based on the Cramér-von Mises statistic which is a good combination of power and conceptual simplicity (Genest and Favre, 2007).

$$S_n = \int_{[0,1]^2} \mathbb{C}_n(u,v)^2 \mathrm{d}C_n(u,v) \tag{4.10}$$

where $\mathbb{C}_n = \sqrt{n}(C_n - C_\theta)$. This statistic measures how close the fitted copula C_θ is from the empirical copula C_n. The distribution of this statistic depends on the unknown value of θ under the null hypothesis that C is from the class C_θ. The specific parametric bootstrap procedure to obtain the approximate P-value is described as follows (Genest et al., 2009).

1. Computer C_n and estimate the parameter θ for C_θ.
2. If there is an analytical expression for C_θ (e.g. Clayton, Gumbel and Frank), compute the value of S_n as defined in Equation 4.10. Otherwise, proceed by Monte Carlo approximation. Specifically, choose $m > n$ (n indicates the size of the data samples) and carry out the following extra steps:
   (a) Generate a random sample $(U_1^*, V_1^*), \cdots, (U_m^*, V_m^*)$ from distribution C_θ.
   (b) Approximate C_θ by
   $$B_m^* = \frac{1}{m}\sum_{i=1}^{m} \mathbf{1}(U_i^* \le u, V_i^* \le v), \qquad (u, v) \in [0,1]^2. \tag{4.11}$$
   (c) Approximate S_n by
   $$S_n = \sum_{i=1}^{n} \{C_n(U_i, V_i) - B_m^*(U_i.V_i)\}^2. \tag{4.12}$$
3. For some large integer N, repeat the following steps for every $K \in \{1, \cdots, N\}$:
   (a) Generate a random sample $(X_{1,k}^*, Y_{1,k}^*), \cdots, (X_{n,k}^*, Y_{n,k}^*)$ from Copula distribution C_θ, and compute their associated rank vectors $(R_{1,k}^*, T_{1,k}^*,), \cdots, (R_{n,k}^*, T_{n,k}^*,)$.
   (b) Compute $U_{i,k}^* = R_{i,k}^*/(n+1), V_{i,k}^* = T_{i,k}^*/(n+1)$ for $i \in \{1, \cdots, n\}$ and let
   $$C_{n,k}^*(u, v) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(U_{i,k}^* \le u, V_{i,k}^* \le v) \qquad (u, v) \in [0,1]^2 \tag{4.13}$$
   and estimate parameter θ for Copula distribution $C_{\theta_{n,k}^*}$.
   (c) If there is an analytical expression for C_θ, let
   $$S_{n,k}^* = \sum_{i=1}^{n} \{C_{n,k}^*(U_{i,k}^*, V_{i,k}^*) - C_{\theta_{n,k}^*}(U_{i,k}^*, V_{i,k}^*)\}. \tag{4.14}$$
   Otherwise, proceed as follows:

(i) Generate a random sample $(X^{**}_{1,k}, Y^{**}_{1,k}), \cdots, (X^{**}_{m,k}, Y^{**}_{m,k})$ from Copula distribution $C_{\theta^*_{n,k}}$.

(ii) Approximate $C_{\theta^*_{n,k}}$ by

$$B^{**}_{m,k}(u,v) = \frac{1}{m}\sum_{i=1}^{m} \mathbf{1}(X^{**}_{i,k} \leq u, Y^{**}_{i,k} \leq v), \qquad (u,v) \in [0,1]^2 \tag{4.15}$$

and let

$$S^*_{n,k} = \sum_{i=1}^{n} \{C^*_{n,k}(U^*_{i,k}, V^*_{i,k}) - B^{**}_{m,k}(U^*_{i,k}, V^*_{i,k})\}^2. \tag{4.16}$$

An approximate P -value for the test is then given by $\sum_{k=1}^{N} \mathbf{1}(S^*_{n,k} > S_n)/N$. When Fitting the parametric Copulas to the data, the standard maximum likelihood estimation (MLE) is used to estimate the unknown parameter θ of each Copula candidate.

4.4 Copula-based conditional simulation

The Copula-based bias correction applied for this study is based on the estimation of a Copula model for each pair of observed (X) and modelled (Y) rainfall for each grid cell. As soon as this Copula model (i.e. $F_X(x), F_Y(y)$ and $C(u,v)$) is estimated, both unconditional and conditional random samples can be generated through Monte Carlo simulations (Gao et al., 2007; Salvadori et al., 2007). Both simulations (or predictions) are based on the conditional distributions of the Copula, given by

$$C_{V|U=v}(v) = P[V \leq v|U = u] = \frac{\partial C(u,v)}{\partial u} \tag{4.17}$$

$$C_{U|V=v}(u) = P[U \leq u|V = v] = \frac{\partial C(u,v)}{\partial v} \tag{4.18}$$

The unconditional simulation of (x, y) is divided into three steps:

(1) Generate random samples of u uniformly from $[0,1]$, remembering that $u = F_X(x)$;

(2) Given a sample value of u, generate a random sample of $v|u$ using the inverse conditional Copula CDF $C^{-1}_{V|U}(v)$;

(3) Generate the corresponding x and y variates by inverting their marginal CDFs from u and v: $x = F_X^{-1}(u)$ and $y = F_Y^{-1}(v)$.

The conditional simulation of $y|x$ is given as follows:

(1) Compute $u = F_X(x)$

(2) Draw random samples of $v|u$ from the inverse conditional Copula CDF $C^{-1}_{V|U}(v)$

(3) Invert from v to obtain y: $y = F_Y^{-1}(v)$.

The conditional simulation of $x|y$ can simply derived by applying the procedure described above and switching the variables x and y.

In this study, the conditional simulation is applied and it is the critical step of this bias correction approach, as it forces a certain variable (e.g. the observation) to take a value when another variable (e.g. the RCM value) is given. The predictions of the observation that are conditioned on the RCM precipitation are then taken as the bias corrected precipitation.

The complete Copula-based bias correction algorithm consists of the following steps:

1. Estimate the theoretical marginal distributions $F_X(x)$ and $F_Y(y)$ for observation and RCM data respectively
2. Transform the time series $x_1, \cdots, x_n$ and $y_1, \cdots, y_n$ to the rank space by taking $u = F_X(x)$ and $v = F_Y(y)$
3. Calculate the empirical Copula $C_n(u, v)$ as a rank based estimator for the theoretical Copula function $C_\theta(u, v)$
4. Estimate the Copula parameter θ and perform Goodness-of-fit tests to identify the best theoretical Copula function $C_\theta(u, v)$
5. Calculate the Copula distribution conditioned on the variate v representing the RCM time series in the rank space
6. Generate the pseudo-observations in the rank space for each time step by using the conditional Copula distribution
7. Transform back the random samples to the data space by using the integral transformation.

To assess the uncertainty associated with this prediction, the conditional prediction process (step 6 and 7) must be repeated for a large number of times, which can be considered as the the Monte Carlo simulation. This provides the possibility to obtain a large set of random realizations and additionally gives the information of a probability density function (PDF) for each corrected time step. From the PDF the spread of the distribution in form of the interquantile range can e.g. be provided as an additional uncertainty criterion for the bias correction.

4.5 Copula-based regression

The Copula-based stochastic bias correction method gives a full ensemble and the empirical predictive distribution of corrected WRF precipitation. For practical reasons and the typical needs of subsequent modelers, e.g. in hydrology, a single corrected value is usually required instead of a complete distribution. One can choose, e.g. the expectation, median or mode of the derived predictive distribution to get a single corrected value. This can be regarded as a Copula-based regression by taking such a "typical" value as the estimator of the derived empirical predictive distribution of corrected WRF precipitation.

4.6 Correction strategy for continuous time series

The implementation of a bias correction for precipitation (a discrete variable) is more complex than a bias correction of a continuous variable, e.g. temperature. In general four cases have to be distinguished, namely (0,0), (0,1), (1,0), and (1,1), where 0 denotes a dry day and 1 indicates a wet day (see Fig. 4.2). A threshold of rainfall amount of 0.1 mm per day was used to identify a wet day with respect to the usual precision of rain gauges (Dieterichs, 1956; Moon et al., 1994). This can also remove the drizzle behavior of the RCM precipitation. Therefore, the four cases are defined as follows:

1. (1,1): REGNIE and WRF precipitation $\geq$ 0.1 mm
2. (0,1): REGNIE $<$ 0.1 mm, while WRF $\geq$ 0.1 mm
3. (1,0): REGNIE $\geq$ 0.1 mm and WRF $<$ 0.1 mm
4. (0,0): Both REGNIE and WRF $<$ 0.1 mm

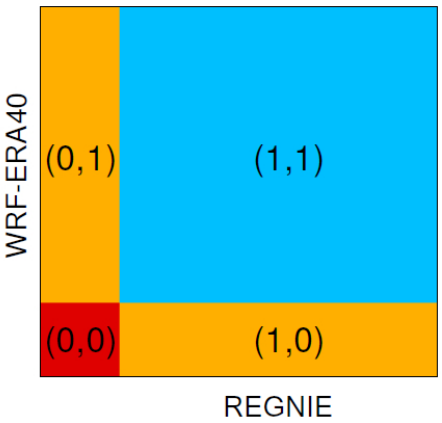

FIGURE 4.2: Illustration of the four cases: (0,0) indicates that both REGNIE and WRF show no rain, (0,1) stands for an observation with no precipitation but the RCM model shows a rain event, while (1,0) indicates the opposite of (0,1), (1,1) implies that both are wet.

To cope with this zero inflated data, various methods are established. The simplest method is the local intensity scaling method (LOCI) (see Sect. 1.2). This method also introduces a precipitation threshold (LOCI threshold), which is not the same as the aforementioned precipitation threshold for the wet-day identification. The LOCI threshold is calibrated such that the number of RCM simulated days exceeding this threshold matches the number of wet days of the observation. Therefore, the number of precipitation events for both control and scenario runs are corrected by applying the calibrated LOCI threshold. The days with precipitation less than the LOCI threshold are redefined to dry days with 0 mm precipitation. This method allows for the correction of the fraction of dry days in time series. However, it does not guaranty the correction for each single time step. In order to illustrate this issue, a small example using one randomly selected grid cell from our domain is shown in the following.

Firstly, the proportion of the four cases for the selected pixel between REGNIE and WRF precipitation is calculated. Then, the LOCI correction is applied for WRF precipitation. Finally, the proportion of the four cases between REGNIE and corrected WRF precipitation is computed and compared with the proportion of the four cases before correction. The results are shown in Table 4.1.

In Table 4.1, the tabular in the top is the the proportion of the four cases before the LOCI correction, while the tabular in the bottom is that after the LOCI adjustment. REGNIE_0 indicates the dry-day in REGNIE, while REGNIE_1 indicates the wet-day in REGNIE. Accordingly, WRF_0 represents the dry-day in WRF and WRF_1 implies the wet-day in WRF. Before the correction, the dry probabilities of REGNIE and WRF in

TABLE 4.1: The proportion of the four cases before (top) and after (bottom) adjustment of dry probabilities following the local scaling approach.

	WRF_0	WRF_1
REGNIE_0	0.07	0.21
REGNIE_1	0.06	0.66

	WRF_0	WRF_1
REGNIE_0	0.14	0.14
REGNIE_1	0.14	0.58

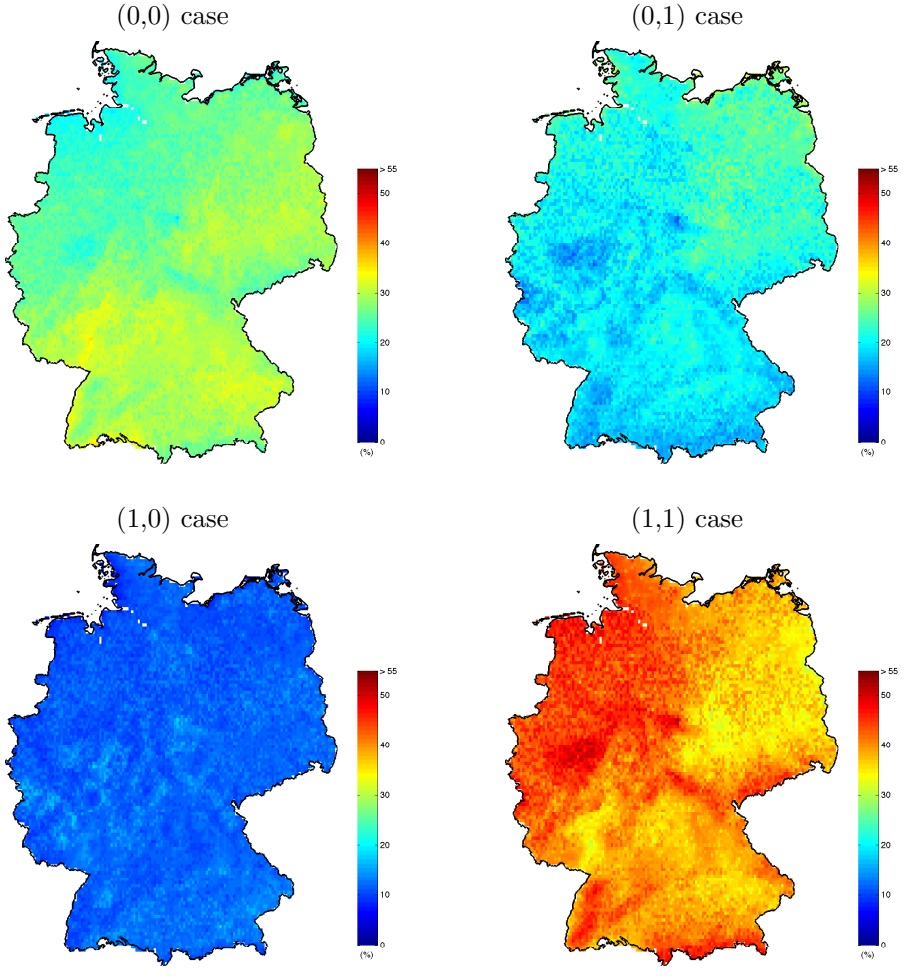

FIGURE 4.3: The proportion of the four cases over the study area for the validation time period (from 1986 to 2000).

the selected example grid cell are $0.07+0.21 = 0.28$, and $0.07+0.06 = 0.13$, respectively. After the local scaling adjustment, the dry probabilities are $0.14 + 0.14 = 0.28$, and $0.14+0.14 = 0.28$, respectively. However, from the table we can still see nearly the same proportion of (0,1) and (1,0) errors ($0.14 + 0.14 = 0.28$) after adjustment of the dry probability if compare to that ($0.21 + 0.06 = 0.27$) before the adjustment. This means that the dry probabilities are adjusted, but the error fractions still remain.

There are also other different approaches exist in the literature to account for the intermittent nature of rainfall. For example the truncated Copula suggested in Bárdossy and Pegram (2009) and the Copula-based mixed model described in Serinaldi (2008). Both methods are able to produce time series that statistically hold the same proportion of the four different cases (0,0), (0,1), (1,0), and (1,1). Similar to the LOCI correction, these methods allow for the correction of the total number of dry days, but do not allow to correct individual events in the (0,1) and (1,0) cases.

In this study, we aim to an event-based correction as described in the following: the Copula-based concept focuses on the correction of the (1,1) cases, i.e. the positive pairs, since the Copula works on the (1,1) cases only. In order to generate a complete bias corrected time series of WRF output, the events that are not covered by the (1,1) case are left unchanged. For the (0,0) cases, there is no error. The errors that come from the (0,1) and (1,0) cases are not corrected by this correction and must be corrected separately. To justify this strategy, we investigated the proportion of the four cases in the study area (see Fig. 4.3): the (1,1) cases take the highest proportion, followed by the (0,0) cases. The proportion of both the (0,1) and (1,0) cases are comparatively low. The average proportion of these cases are 40% for the (1,1) cases, 29% for the (0,0) cases, 19% for the (0,1) cases and 12% for the (1,0) cases, respectively.

Chapter 5

The application of the Copula-based bias correction

5.1 Introduction

In this chapter, the Copula-based bias correction is applied for a WRF simulated precipitation field in Germany. As mentioned in Chapter 4, a bivariate Copula model forms the basis of the bias correction technique. Details about the estimated Copula models in the calibration period (1971–1985) are presented, which include information about the fitting of the marginal distributions and the identification of the theoretical bivariate Copula functions. Since the marginal distributions reflect the statistical characteristics of precipitation, the differences between estimated marginal distribution of WRF and REGNIE precipitation are analyzed spatially. The identified Copula functions are also analyzed over space as they imply the dependence structure between WRF and REGNIE precipitation. The fitted Copula models are then applied for the validation period (1986–2000) to bias correct the WRF precipitation. Finally the performance of the Copula-based correction is validated by investigating the relative biases of daily mean precipitation after the bias correction. Furthermore, to investigate typical situations in detail, the monthly mean precipitation for four specific pixels are analyzed.

5.2 Estimated marginal distributions

The marginal distributions estimation is the first step of the Copula model building. For both REGNIE and WRF data five different distribution functions (see Sect. 4.2.1) are employed for each grid cell separately: Generalized Pareto distribution (gp); Gamma distribution (gam); Exponential distribution (exp); Weibull distribution (wbl) and Normal distribution (norm). This guarantees the great flexibility in selecting the most appropriate distribution for each grid cell. Since the proposed bias correction framework is focused on the (1,1) cases only (see Sect. 4.6). Therefore, only pair-wise recorded modelled and observed data (i.e. positive pairs) are used for the marginal distribution estimation. The Goodness-of-fit test follows a two-stage process (K–S test and the Bayesian information criterion) which is described in Sect. 4.2.2. The fitted marginal distributions for both datasets in the study area are shown in Fig. 5.1. It can be seen that the Goodness-of-fit tests reject the Normal distribution in all cases, while the Generalized Pareto distribution is accepted most frequently for both REGNIE and WRF. The result shows a reasonable agreement of selected marginal distribution between REGNIE and WRF mainly in the east and south parts of Germany. It is also found that the patterns of the selected types follow the topography of Germany (see Fig. 2.1). In the northwest of Germany, the Weibull distribution function prevails as well as in the low mountain ranges. In general, this effect is stronger for WRF while the patterns are more patchy for REGNIE. In the northeast a few discrete grid cells with fitted Gamma distribution are found in REGNIE, while this patchy pattern is not shown in WRF. A possible reason could be that the REGNIE data is interpolated from the point based station rainfall and the interpolated precipitation field might be effected by the intensity of the stations as there are not so many stations available in the northeast of Germany.

In order to investigate this coincidence between REGNIE and WRF marginal distributions in more detail, the confusion matrix (Stehman, 1997) of them is calculated. Each row of the matrix represents the distribution types of REGNIE, while each column represents that of WRF (in %). The major diagonal shows the fraction of concurring marginal distribution types. The confusion matrix of REGNIE and WRF for the calibration period is shown in Table 5.1. It is found that for 42 % of grid cells, the Generalized Pareto distribution is selected for both data sources concordantly. For the Weibull distribution this holds true for 16 % of the grid cells. Since the total number of grid cells where Gamma and Exponential distribution are fitted is very low, the percentage of hits in the diagonal of the confusion matrix is small. Summing up the major diagonal

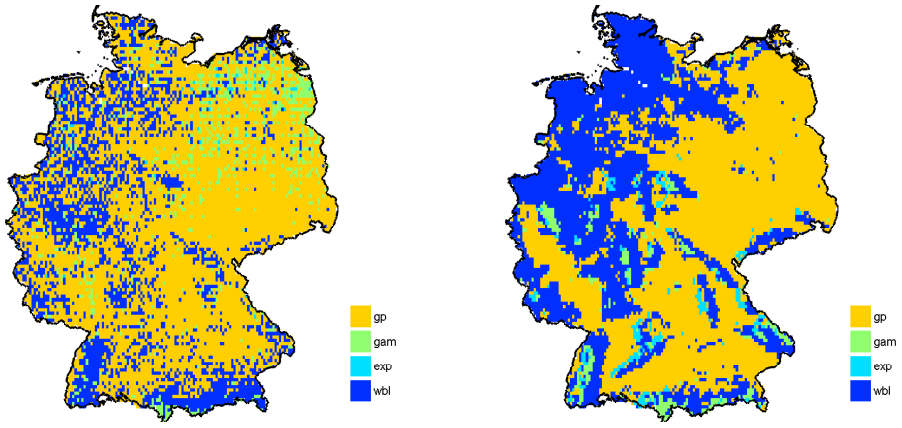

FIGURE 5.1: Estimated marginal distributions of precipitation for Germany for both REGNIE (left) and WRF (right). The results are shown for the calibration period (1971–1985) and positive pairs only.

TABLE 5.1: Confusion matrix between REGNIE and WRF for the different distribution types.

		WRF			
		gp	gam	exp	wbl
REGNIE	gp	42.04%	1.27%	1.55%	20.79%
	gam	4.92%	0.5%	0.18%	2.44%
	exp	0.27%	0%	0%	0.23%
	wbl	7.14%	1.94%	0.79%	15.93%

gives a measure for the overall agreement. For the complete calibration series about 59 % correspond. The failures of 21 % of grid cells, where REGNIE follows the Generalized Pareto distribution and WRF follows the Weibull distribution, are predominately located in the Northwest of Germany (Fig. 5.1).

As mentioned above in Sect. 4.2.2 the Goodness-of-fit tests follow a two-step process due to the fact that the K–S test is highly sensitive to large sample sizes. For the marginal distribution identification in this study, for 99 % of the grid cells the K–S test fails and only the BIC is used for REGNIE, while the number for WRF is 68 %. To justify this Goodness-of-fit test, a visual inspection is further applied (see Sect. 4.2.3). First, the RSS between the empirical distribution and the fitted theoretical marginal distribution is computed for both REGNIE and WRF for each grid in the study area. Then, three grid cells of the highest RSS are selected for both REGNIE and WRF data and their empirical CDF, theoretical CDF and the corresponding quantile-quantile plots are shown for a visual inspection. The calculated RSS map are shown in Fig. 5.2. It

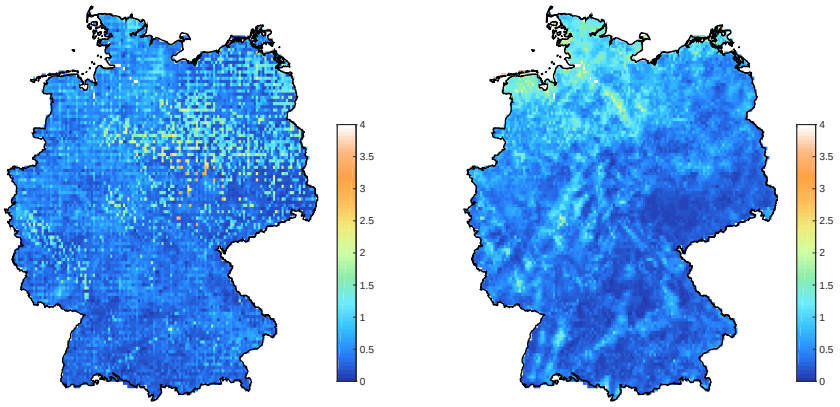

FIGURE 5.2: The residual sum of squares (RSS) between the empirical distributions and the fitted theoretical marginal distributions. The RSS for REGNIE is shown in left side and for WRF it is shown in right side. The results are shown for the calibration period (1971–1985).

can be seen that the RSS for both REGNIE and WRF precipitation are low in most of the study area. For the REGNIE data, the RSS varies between 0 and 3.7 and the high values are mainly located in the middle of the domain. For the WRF precipitation field, the RSS varies between 0 and 3.3 and the high values are dominated in the southwest of the study area. Furthermore, the RSS pattern of REGNIE are much more patchy than that of the WRF data, especially in the northeast of the domain. This might be caused by the patchy distributed marginal distribution (see Fig. 5.1). Low density of the rain gauges in the northeast of Germany effects the interpolated precipitation field (REGNIE) and therefore causing this patchy pattern.

The empirical distributions, fitted theoretical distributions and the corresponding Q-Q plots for three selected grid cells with highest RSS value are shown in Fig. 5.3 and Fig. 5.4 for REGNIE and WRF, respectively. It can be seen that the theoretical distributions are quite close to the empirical distributions which indicates a good agreement between them even the RSS are high compared the other gird cells (see Fig. 5.3 and Fig. 5.4, left panel). To further investigate on it, the Q-Q plots are then analyzed. For REGNIE all of the three selected grid cells seem to follow the assumed distributions reasonably well. Most of the points in the Q-Q plots fall on a straight line ($y = x$) except the extreme parts. In the tail part, the Q-Q plot of sample data versus distribution are showing either left-skewed or right-skewed patterns (see Fig. 5.3, right panel). The Q-Q plot skewed left from the straight line ($y = x$) indicates that the fitted theoretical distribution has a fatter PDF than the empirical distribution, while the right-skewed pattern implies the opposite. The grid cell with the highest RSS (RSS = 3.7) is fitted by a gamma distribution (see Sect. 4.2.2) with the parameters $a = 0.95, b = 3.89$ and the

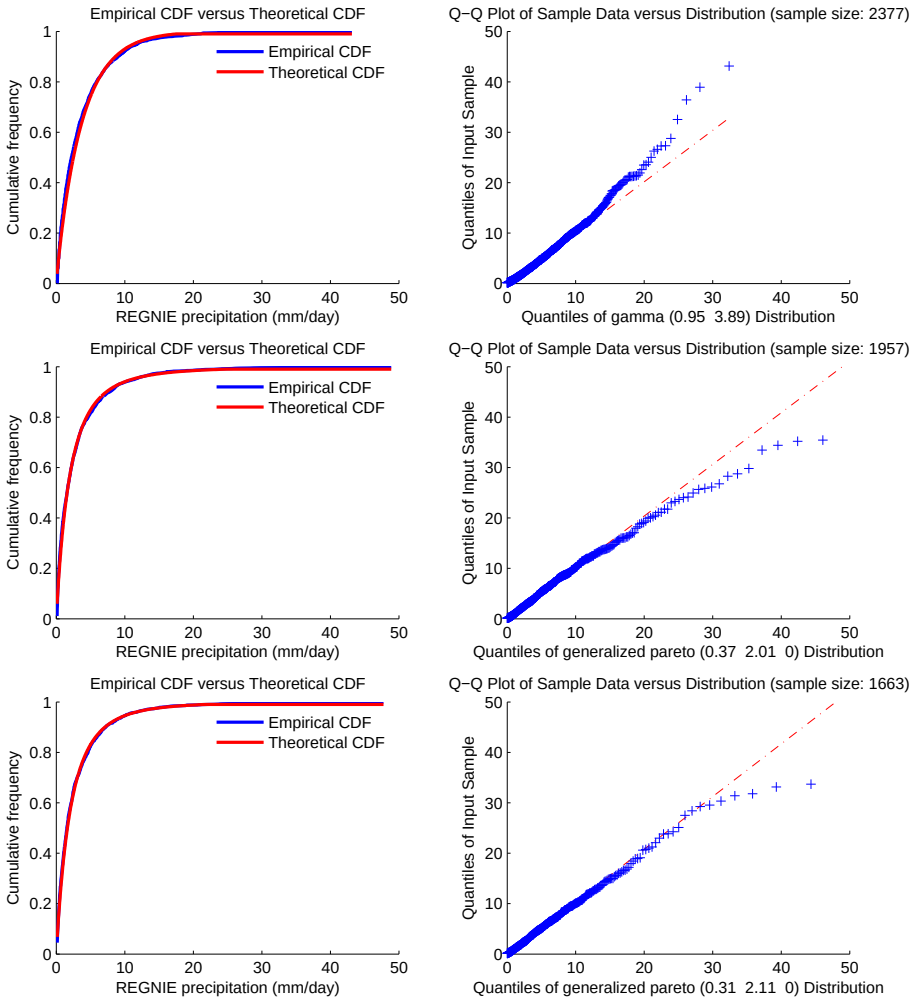

FIGURE 5.3: The empirical and fitted theoretical distributions of REGNIE precipitation (left panel) and the corresponding quantile-quantile plots (right panel) for the selected grid cells. These three grid cells are selected from Fig. 5.2 (left) with highest RSS value. From top to bottom, they are listed in descending order by RSS value.

corresponding Q-Q plot has a left skew. The grid cells with the second (RSS = 3.6) and third highest (RSS = 3.5) RSS are both fitted by generalized pareto distributions and are showing both right skews. For WRF precipitation, the Q-Q plots indicate a great agreement between the fitted distribution and the sample data. Only a few outliers are shown in the data. This visual inspection of the distribution fitting implies that the precipitation are reasonably well fitted for both REGNIE and WRF even the RSS are relatively high in the entire domain. Therefore the two-stage Goodness-of-fit testing is justified.

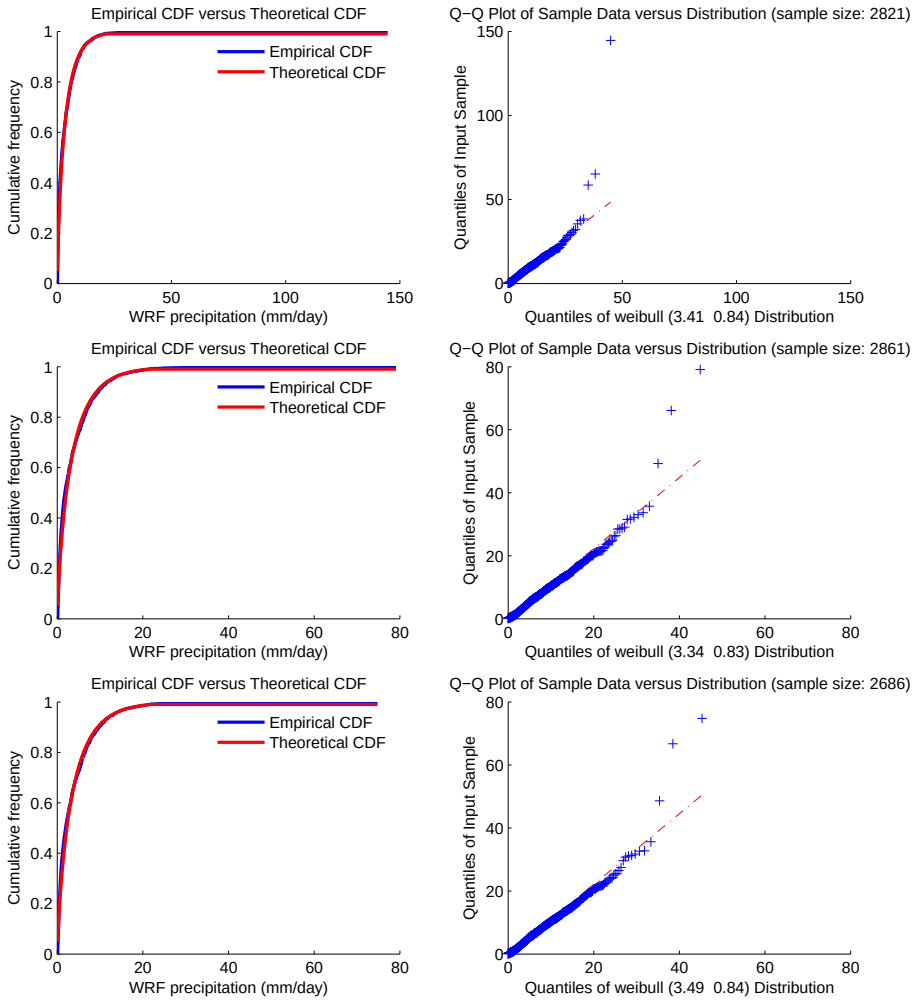

FIGURE 5.4: The empirical and fitted theoretical distributions of WRF precipitation (left panel) and the corresponding quantile-quantile plots (right panel) for the selected grid cells. These three grid cells are selected from Fig. 5.2 (right) with highest RSS value. From top to bottom, they are listed in descending order by RSS.

5.3 Identified Copula functions

For each grid cell the theoretical Copula function that characterizes the dependence structure between REGNIE and WRF data is identified separately. The identification of the Copula functions is based on the ranked values that are already transformed by the fitted marginal distributions of REGNIE and WRF. Four different types of Copulas (Clayton, Frank, Gumbel and Gaussian) are investigated by applying the Goodness-of-fit tests described in Sect. 4.3. Three of them (Clayton, Frank and Gumbel Copula) are from the Archimedean Copula families and the Gaussian Copula is from the Elliptical Copula families (see Sect. 3.5). These four Copulas are able to capture different dependence

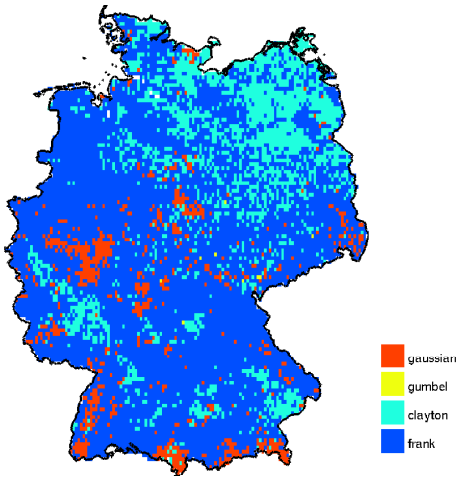

FIGURE 5.5: Identified Copula functions between REGNIE and WRF precipitation in the calibration period (1971 to 1985) with positive pairs.

structures. The Gumbel Copula is able to describe an upper tail dependence structure, while the Clayton Copula allows to express higher probability in the lower tail. The Frank Copula exhibits no tail dependence, and the Gaussian Copula describes a similar dependence as the Frank Copula, but with slightly higher densities in the lower and upper tails (Venter, 2002; Schmidt, 2007).

The appropriate Copula function is selected for each grid cell by applying the Goodness-of-fit testing that is described in Sect. 4.3. Figure 5.5 shows the results of the Goodness-of-fit tests for the calibration period for the complete study area. It is found that for most of the grid cells in the study area, the Frank Copula can capture the dependence structure best, while for the Northeast of Germany the Clayton Copula provides the best fit. In total the dependence structure of 72 % of the grid cells is modelled by the Frank, 20 % by the Clayton, 7 % by the Gaussian and only 0.09 % by the Gumbel Copula. As mentioned above, these four Copulas describe different dependence structures. Therefore from the fitted Copula family map we can see that for most of the study area they show no tail dependence and the grid cells in the northeast show a lower tail dependence. The grid cells that showing a upper tail dependence are rarely found in the domain.

5.4 Validation of the Copula-based bias correction

Based on the estimated Copula model (the marginal distributions and the Copula function) the conditional distribution of REGNIE conditioned on WRF is derived for each

grid cell separately (see Sect. 4.4). To generate bias-corrected WRF precipitation, random samples of possible outcomes are drawn from this conditional distribution through the Monte Carlo simulations. We use a sample size of 100. The result can be interpreted as an empirical predictive distribution for corrected WRF (pseudo-observations) that is determined for all conditioning WRF precipitation values for each time step. While this stochastic bias correction method gives a full ensemble and the empirical predictive distribution of corrected WRF precipitation, for practical reasons a "typical" value can be taken as the estimator of the predictive distribution to get single corrected values (see Sect. 4.5).

Figure 5.6 exemplarily shows WRF (red), REGNIE (green) and the bias corrected WRF (the box-plot, while blue indicates the median of the realizations) data for pixel 1 in Fig. 2.1 during wintertime 1986–1987 (positive pairs only). For each corrected time step, 100 possible outcomes of bias corrected precipitation are generated. The box plot visualizes the spread of the generated random sample (100 members) indicating the uncertainty of the predicted bias-corrected precipitation, while the blue line shows the median of the respective empirical predictive distribution. It can be seen from Fig. 5.6 that for most of the time steps the proposed Copula-based approach can successfully correct for biases in the modelled precipitation compared to observed values by e.g., taking the median of the sampled bias corrected values.

To investigate the spatial performance of the correction algorithm, the relative bias of RCM modelled mean daily precipitation (WRF) compared to gridded observations (REGNIE) is compared to that of the bias corrected model data (B.C. WRF) for Germany.

A comparison of corrected WRF data derived by the expectation, median and mode of the predictive distribution with observations are shown in Fig. 5.7. Figure 5.7 (top-left) shows the relative bias between REGNIE and WRF, indicating wet biases in most of the study area. These wet biases are most prominent in high elevation areas following the topography of Germany. Wet biases are also detected in the Northeast of Germany, where the elevation is low. Dry biases are found in the alpine and pre-alpine areas in the Southeast of Germany as well as in the West of Germany. Figure 5.7 (top-right) is the relative bias map between REGNIE and bias corrected WRF by taking the mean of the stochastic sampling. It can be seen that the wet biases are corrected for most of the domain, except for a very small region in the Northeast. It is also found that the dry bias can also be significantly reduced, but small dry biases are introduced in some areas in the West of the domain. The average of the bias for the whole study area is reduced

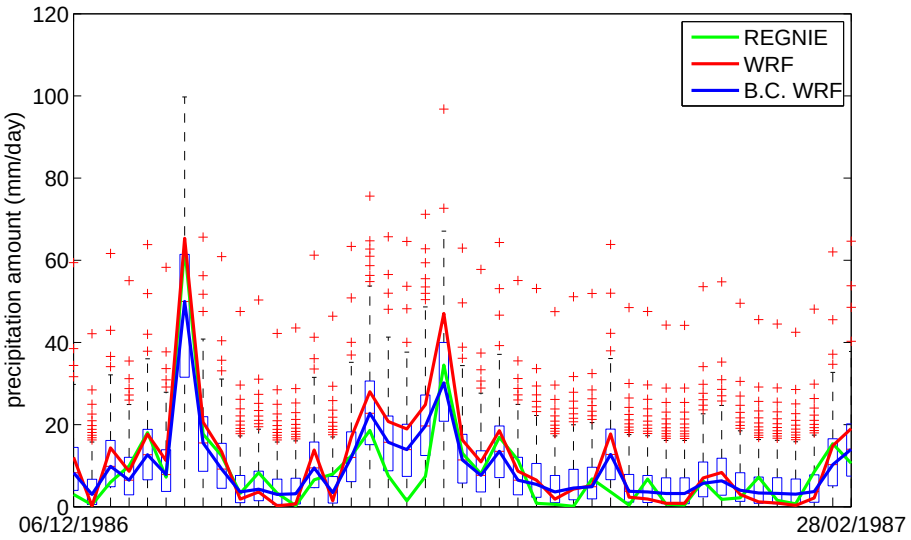

FIGURE 5.6: Comparison of bias-corrected WRF data by taking the median regression (blue) with the original WRF data (red) and REGNIE (green) in winter 1986–1987 (positive pairs only) for pixel 1 in Fig. 2.1. For each time step 100 realizations are drawn from the conditional distribution visualized by the box-whiskers (boxes are defined by the lower $Q1$ and the upper quartile $Q3$). The length of the whiskers is determined by $1.5 \cdot (Q3 - Q1)$ and outliers, i.e. data values beyond the whiskers are marked by crosses.

from 10 to $-1\,\%$. The bottom-left relative bias map is calculated between REGNIE and bias corrected WRF taking the median of the realizations, while the bottom-right is the one computed between REGNIE and mode based bias corrected WRF. Both simulations tend to underestimate the precipitation values, thus causing a dry bias over the domain. Therefore the expectation of the sampled pseudo-observations is then taken as the estimator of the bias corrected precipitation. For the following illustration, the results are shown and analyzed for the expectation only.

In order to evaluate the performance of the bias correction in different seasons, the relative bias maps of mean daily precipitation are also calculated seasonally and are shown in Fig. 5.8. The seasonal index in this study are defined by grouping the calendar months in the following way:

- Spring - the three transition months March, April and May.
- Summer - the three hot months June, July and August.
- Autumn - the transition months September, October and November.
- Winter - the three cold months December, January and February.

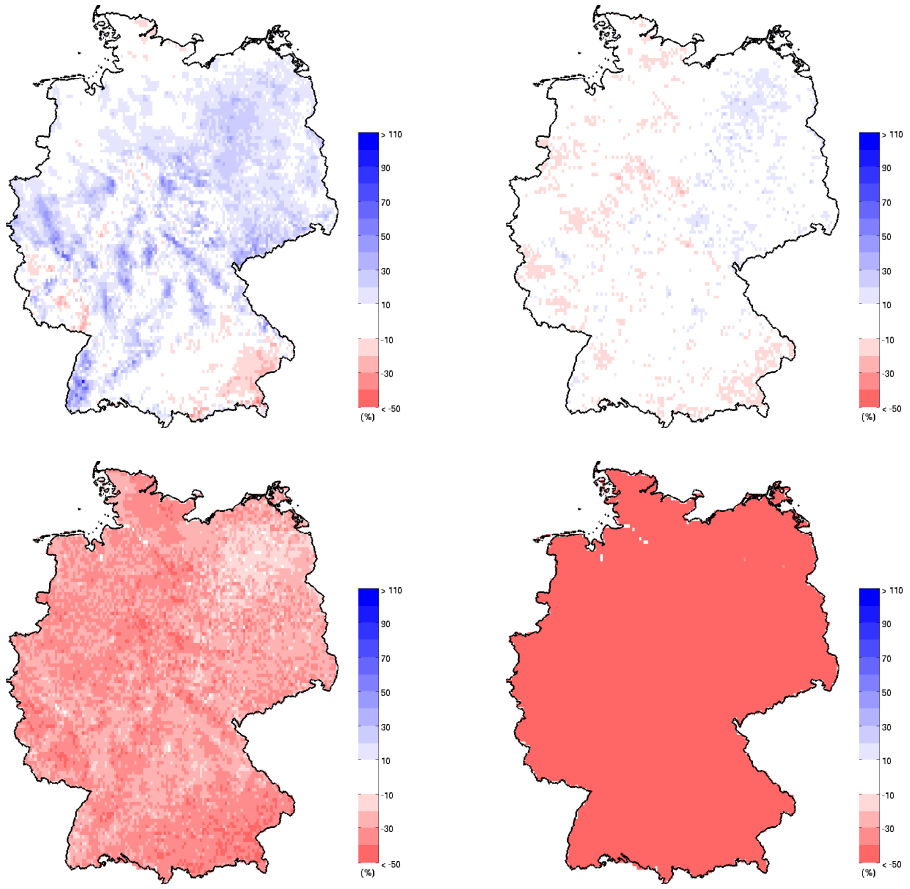

FIGURE 5.7: Relative bias map of mean daily precipitation for uncorrected precipitation field (top-left), corrected WRF precipitation field by taking the expectation (top-right), median (bottom-left) and the mode (bottom-right) as the estimator of the sampled distribution. The results are based on the validation period 1986-2000.

Figure 6.7 (left) shows the relative bias between uncorrected WRF mean daily precipitation and the REGNIE data set for the different seasons (spring – MAM, summer – JJA, autumn – SON, winter – DJF, from top to bottom). It can be seen that the relative biases between REGNIE and uncorrected WRF are even larger for different seasons compared to the biases for the complete fifteen years in calibration period from 1986 to 2000 (see Fig. 5.7). The WRF model tends to generate too much precipitation in spring and winter for the majority of grid cells in the study area. For summer and autumn, there are also regions found, where the model is too dry. These regions are mostly located in the North and in the South of Germany. This effect is found to be strongest in summer while in autumn areas with an overestimation of precipitation are still found in the Northeast and Southwest of Germany. In all cases, the bias is influenced by the underlying terrain showing an overestimation especially in regions with higher altitude. The average of the biases from spring to winter are 32, -15, 4 and 28 %, respectively.

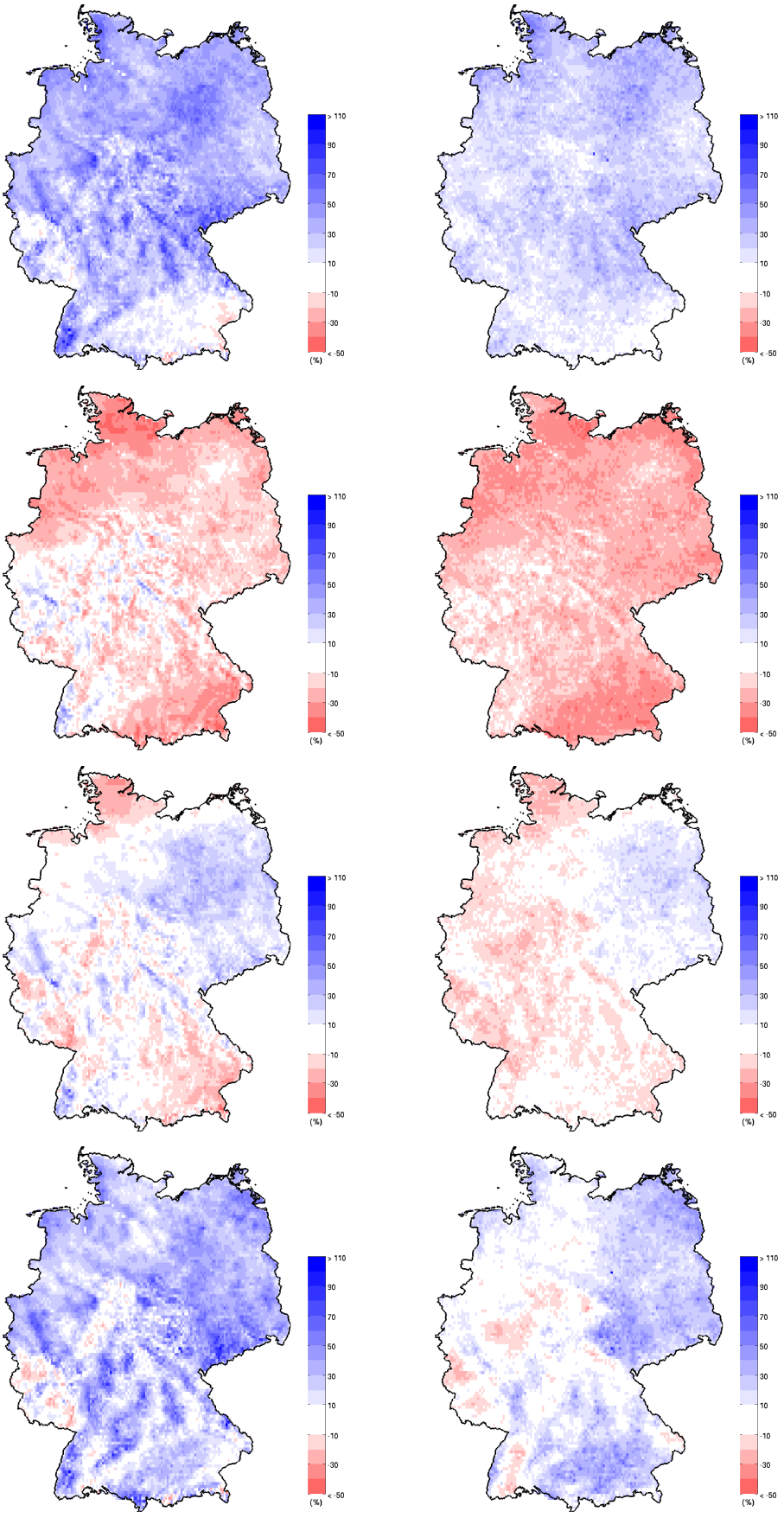

FIGURE 5.8: Relative bias between uncorrected (left) and corrected (right) WRF mean daily precipitation and the REGNIE data set in Germany for the different seasons (spring–MAM, summer–JJA, autumn–SON, winter–DJF, from top to bottom). The results are derived for the validation time period (1986-2000).

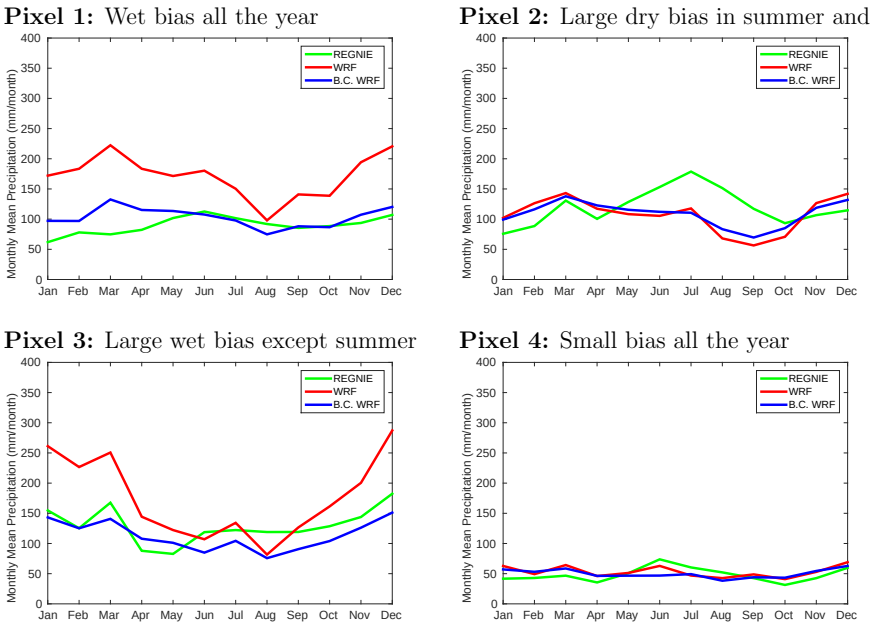

FIGURE 5.9: Comparison of bias corrected WRF mean monthly precipitation (blue) with REGNIE (green) and original WRF data (red) for the selected four pixel 1–4 in the validation period from 1986 to 2000. The number of the respective grid cell is noted in the upper left corner of each plot.

Figure 6.7 (right) shows the relative bias between corrected WRF mean daily precipitation and the REGNIE. It shows that in spring the wet biases are reduced for the entire study area even though still some biases are remain. In summer the Copula-based bias correction reduced too much precipitation thus increased the dry biases for the domain. In autumn and winter, the wet biases are significantly reduced. However, in both seasons the correction introduced some dry biases for the west of Germany, especially in the autumn. The average biases after the bias correction are 21, −23, −4 and −13 % respectively for different seasons from spring to winter.

In the following, it is further analyzed how well the model can reproduce the intra-annual variability of observed precipitation and how the performance for the different seasons is influenced by the Copula-based correction algorithm.

To investigate typical situations in detail, the results are shown for four specific pixels (grid cells) in the study area (see Fig. 2.1): pixel 1 and pixel 3 are selected as they show the highest wet bias between WRF and the REGNIE. Pixel 2 is located in the region where a dry bias was generated by the WRF in summer and autumn and a wet bias was generated in winter. Pixel 4 represents a case where the agreement between uncorrected model data and REGNIE observations is already good.

Figure 5.9 shows mean monthly precipitation derived for the validation period (1986–2000) for the selected grid cells 1–4 (see Fig. 2.1 for their exact locations). The number of the respective pixel is noted in the upper left corner.

The results for grid cell 1 in Fig. 5.9 confirm the fact that the RCM model results strongly overestimate the precipitation amount in that case. The annual variability of the observations is in general reproduced, except for a strong increase of the mean precipitation in August that is not found in the observations. This behavior is found also for grid cell 3 indicating a relatively too dry summer season. For grid cells 1 and 3, the Copula-based correction is found to be able to correct for the overestimation of precipitation amounts as well as for the effect of a too strong decrease of precipitation in August. However, the correction is introducing a slight underestimation mainly during summer and autumn instead. For grid cell 2, the correction shows only a slight improvement while in summer still large dry biases remains. The same performance are also found for grid cell 4 in which the performance of WRF was already satisfactory. The correction slightly changes the monthly mean precipitation but decrease too much in June, thus introduces dry biases.

5.5 Summary and discussion

The proposed stochastic bias correction technique is based on a bivariate Copula model which consists of two marginal distributions (marginal distribution of REGNIE precipitation and marginal distribution of WRF derived precipitation) and a Copula function. The marginal distributions reflect the statistical characteristics of precipitation and the Copula function captures the dependence structure between WRF and REGNIE precipitation. The bivariate Copula models are estimated for each grid cells within the study area in the calibration period (1971–1985). For construction of Copula models only the positive pairs of REGNIE and WRF precipitation (i.e. (1,1) cases) are used, since the Copula only works on the (1,1) cases.

In the calibration time period from 1971 to 1985, the REGNIE and WRF precipitation for each pixel are fitted to a theoretical marginal distribution separately through a two-stage Goodness-of-fit test (K–S test and the Bayesian information criterion). Five different theoretical distribution functions are selected as the candidates of Goodness-of-fit tests: Generalized Pareto distribution; Gamma distribution; Exponential distribution; Weibull distribution and Normal distribution. This guarantees the great flexibility in

selecting the most appropriate distribution for each grid cell. The fitted marginal family maps of REGNIE and WRF show different patterns that indicates the deficiencies of representing the precipitation distribution by WRF. It is found that the patterns of fitted marginal families follow the topography of Germany, which means that the precipitation distribution changes over elevation. The fitted marginal distribution map for REGNIE are also found to be more patchy than that for WRF precipitation field, especially in the northwest of Germany.

Due to the fact that the K–S test is highly sensitive to large sample sizes, for plenty of the grid cells (99 % for REGNIE and 68 % for WRF) the KS test fails and only the BIC is used for the selection of the marginal distribution. To justify this Goodness-of-fit test, a visual inspection is further applied, which is based on the empirical CDF, theoretical CDF and the corresponding quantile-quantile plots. The pixels that have relative high RSS are further inspected and the results show good agreements between fitted theoretical marginal distributions and the empirical distributions. This means that the two-stage Goodness-of-fit testing is justified.

After the marginal functions for REGNIE and WRF are fitted, the precipitation values are then transformed to rank space where a Copula function can fit. The Copula functions are also fitted for each pixel separately and four different types of Copulas are investigated: Clayton, Frank, Gumbel and Gaussian. The Goodness-of-fit testing for the Copulas are based on the Cramér-von Mises statistic. The fitted Copula family map shows that the Copulas are different in different areas. Since different Copulas which are applied in this study represent different kinds of dependence structure, the dependence structures between REGNIE and WRF are found to vary over space. It is also found that for most of the study area the dependence structures between REGNIE and WRF show no tail dependence, where the Frank and Gaussian Copula are fitted.

Based on the estimated Copula model (the marginal distributions and the Copula function) the conditional distribution of REGNIE conditioned on WRF is derived for each grid cell separately. To generate bias-corrected WRF precipitation, random samples of possible outcomes are drawn from this conditional distribution through the Monte Carlo simulations. The random sample size is taken as 100 in this study. Which means for each corrected time step 100 realizations (pseudo-observations) are generated as the corrected WRF precipitation. Therefore, actually an empirical predictive distribution for corrected WRF for each time step is derived. For practical reasons, e.g. spatial illustration, the expectation of these 100 realizations is taken as the estimator of the corrected WRF precipitation. Since the Copula model is established only based on the

positive pairs (i.e. (1,1) cases), the correction is also applied only on the (1,1) cases in the validation period (1986–2000). In order to generate a complete time series, the time steps that are not belong to the (1,1) cases are then kept the same as the raw WRF data.

By investigating on the relative bias of mean daily precipitation, the Copula-based bias correction is found to be able to reduce the biases from WRF derived precipitation significantly. However, when looking at the seasonal relative bias of mean daily precipitation, the performance of the correction drops. The biases are only slightly reduced and in summer the dry biases are even increased. In order to evaluate the performance of the proposed method regarding to the intra-annual variability, four specific pixels in the study area are again selected and further analyzed by looking at their monthly mean precipitation. Results show that corrected monthly mean WRF precipitation are much closer to the observations (REGNIE) compared to the uncorrected WRF precipitation.

Since the method is found to be inefficient for seasonal bias correction, the Copula-based correction is then applied in a seasonal mode and the results are analyzed in the following chapter.

Chapter 6

The application of Copula-based bias correction with a seasonal mode approach

6.1 Introduction

In Chapter 5, the Copula-based bias correction was applied for a WRF simulated precipitation field in Germany. The bivariate Copula models (two marginal distributions and one Copula function) for each gird cells are established from a 15–year calibration period from 1971 to 1985 and are then applied for another 15–year validation period (1986–2000) to bias corrected the WRF simulated precipitation fields. The results showed that the Copula-based bias correction (the mean of the Monte Carlo simulated realizations is taken as the estimator) is able to significantly reduce the biases of daily mean precipitation for the entire study area. However, when analyzed for different seasons, the biases can only be slightly reduced. In summer, the biases are even increased after the correction. It is also found that not only the dependence structures, but also the marginal distributions for both the REGNIE and WRF vary intra-annually. Therefore, in this chapter the Copula-based bias correction is applied seasonally. The Copula models are estimated for different seasons (spring – MAM, summer – JJA, autumn – SON, winter – DJF) in the calibration period separately and are then applied also separately for different seasons in validation period to correct the WRF precipitation fields. The

fitted marginal distribution of WRF and REGNIE precipitation as well as the identified Copula functions are analyzed spatially and seasonally. The performance of the Copula-based bias correction is validated by investigating the relative biases of daily mean precipitation, monthly mean precipitation.

6.2 Estimated marginal distributions

As the first step of the Copula model building, the marginal distributions are estimated for different seasons (spring – MAM, summer – JJA, autumn – SON, winter – DJF). Following the same process in Chapter 5, for both REGNIE and WRF data five different distribution functions (see Sect. 4.2.1) are employed for each grid cell to guarantee the great flexibility in selecting the most appropriate distribution. For each season, the pairwise recorded modelled and observed data are thrown into a two-stage goodness-of-fit testing process to fit the appropriate theoretical distributions (see Sect. 4.2.2).

For both REGNIE and WRF data, the seasonal representation of the different distribution types is shown in Fig. 6.1. For all the seasons, the Normal distribution is rejected by the goodness-of-fit process for each grid cells. The fitted seasonal marginal distribution maps indicate that the choice of the optimal marginal distribution clearly depends on the season. For WRF, the winter (summer) season is dominated by Exponential (Generalized Pareto). The differences for REGNIE are not that obvious since the dominant distribution type is the Generalized Pareto distribution for all seasons. For WRF data the effect of the underlying elevation on the identified distribution type is most prominent during winter and fall. In the low mountain regions the favorite marginal distribution change from fall (Weibull, Generalized Pareto) to winter (Exponential, Weibull).

The seasonal confusion matrices for different seasons are shown in Table 6.1. The major diagonal shows the fraction of concurring marginal distribution types and the sum of the major diagonal indicates the overall agreement. In spring, for 40 % the pixels the Generalized Pareto distribution is selected for both data sources concordantly, while for other three distributions the numbers are less than 4 %. In summer, the concurrence are mainly from the Generalized Pareto distribution and the Weibull distribution with the fraction of 42 % and 14 %, respectively. The same trend is found in autumn, where the agreements between WRF and REGNIE are contributed by the Generalized Pareto distribution with a number of 15 % and by the Weibull distribution with a number of 18 %. For the Gamma distribution and the Exponential distribution, the number of

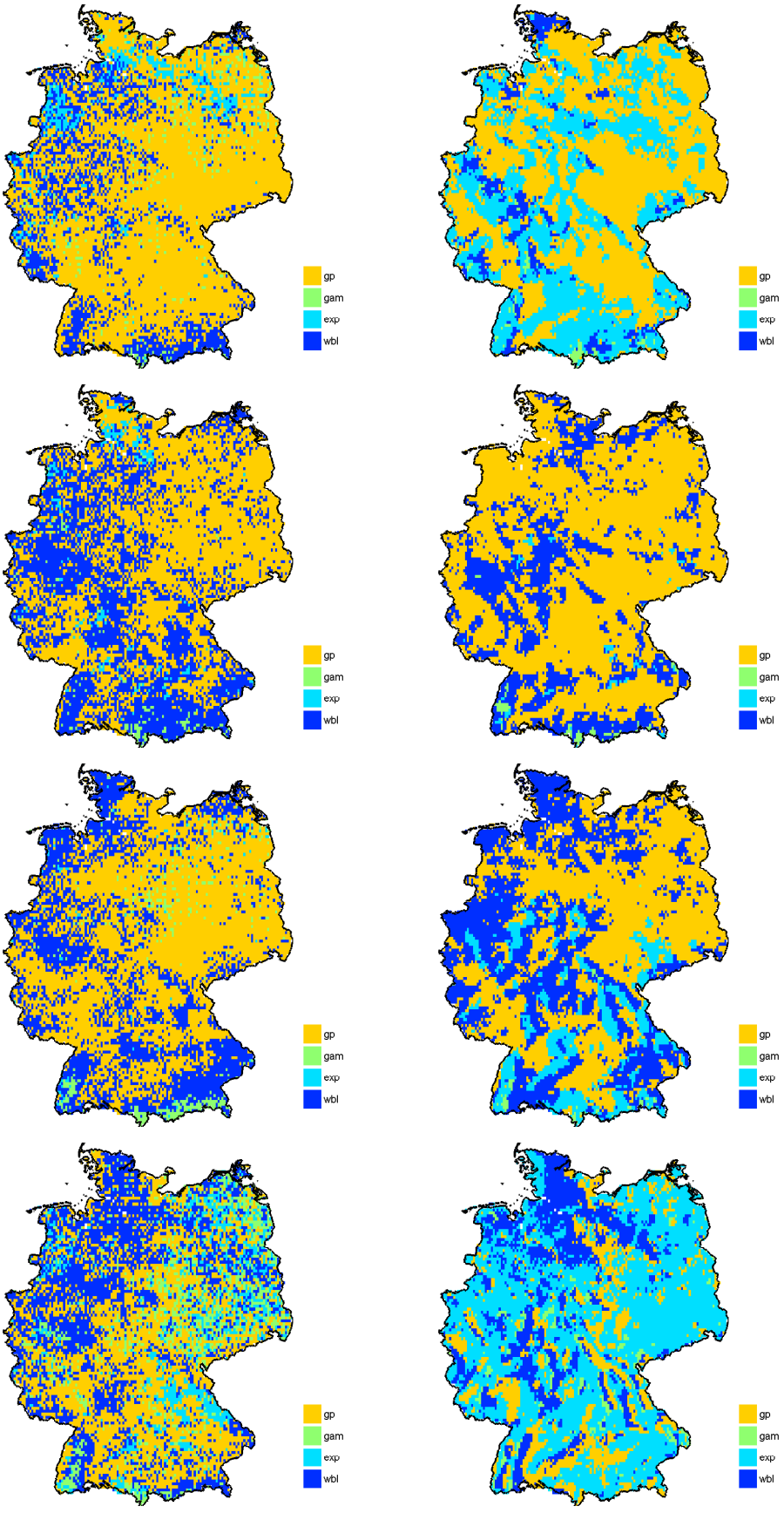

FIGURE 6.1: Estimated marginal distribution of precipitation for the different seasons for REGNIE (left panel) and WRF (right panel) in Germany. The results are shown for the calibration period (1971–1985) for positive pairs only. Spring (MAM), summer (JJA), autumn (SON) and winter (DJF) are illustrated from top to bottom.

coincidence are less than 1 %. In winter, the percentage of hits in the diagonal of the confusion matrix are nearly uniform except for the Gamma distribution, which has the concurrence number around 0. By summing up the numbers in the major diagonal, the results show the best agreement between WRF and REGNIE (approximately 56 % of the grid cells) in summer, while in wintertime only approximately 30 % of the types agree.

Due to the fact that the K–S test is highly sensitive to large sample sizes, our Goodness-of-fit tests follow a two-step process which is described in Sect. 4.2.2. For the annual marginal distribution identification, for 99 % of the grid cells the K–S test fails and only the BIC is used for REGNIE, while the number for WRF is 68 % (see Sect. 5.2). Since the sample size is reduced in seasonal analysis, the failures of K–S test are decreased dramatically. The results are shown in Table 6.2. It can be seen that for REGNIE precipitation distribution fitting, the K–S test failure appears highly in winter with a number of 56 %. In other seasons, the number of the K–S test failure are relatively low. Especially in summer, only 11 % of the grid cells are rejected by the K–S test and only the BIC is used for the Goodness-of-fit test. For WRF precipitation, the K–S test failure number are below 15 % for all the seasons. The largest number is 12 % in autumn and the smallest number is nearly zero in spring. Nevertheless, to justify this Goodness-of-fit test, a visual inspection is necessary (see Sect. 4.2.3).

The RSS between the empirical distribution and the fitted theoretical marginal distribution is computed for both REGNIE and WRF for each grid in the study area. The computed RSS are shown for different seasons in Fig. 6.2. It can be seen from the RSS maps that for WRF precipitation field the RSS between the empirical distribution and the fitted theoretical marginal distribution are nearly zero for all the seasons in the entire study area. For REGNIE data the high RSS value appears in winter, while it still less than 3. The same as the annual case in Sect. 5.2, the RSS pattern of REGNIE are much more patchy than that of the WRF data and this may due to the fact that the interpolation of REGNIE data effects the precipitation distribution.

For a further inspection of the performance of the Goodness-fit-tests, the empirical distributions, fitted theoretical distributions and the corresponding Q-Q plots for grid cells with highest RSS value in different seasons are shown in Fig. 6.3 and Fig. 6.4 for REGNIE and WRF precipitation, respectively. For both REGNIE and WRF precipitation, the fitted theoretical distributions show great agreements with their corresponding empirical distributions for selected pixels in each seasons (see Fig. 6.3 and Fig. 6.4, left panel). From the Q-Q plots of those pixels more details can be seen. The right panel

TABLE 6.1: Seasonal confusion matrix of fitted REGNIE and WRF precipitation distribution.

MAM		WRF			
		gp	gam	exp	wbl
REGNIE	gp	39.57%	0.29%	25.68%	3.89%
	gam	2.32%	0.12%	1.32%	0.18%
	exp	2.68%	0.02%	3.03%	0.14%
	wbl	8.88%	0.56%	7.81%	3.51%

JJA		WRF			
		gp	gam	exp	wbl
REGNIE	gp	42.3%	0.09%	0.39%	11.58%
	gam	0.72%	0.14%	0.04%	0.83%
	exp	1.74%	0%	0%	0.81%
	wbl	26.4%	0.62%	0.61%	13.73%

SON		WRF			
		gp	gam	exp	wbl
REGNIE	gp	35.43%	0.08%	6.36%	18.83%
	gam	1.55%	0.29%	0.95%	1.14%
	exp	0.51%	0%	0.15%	0.41%
	wbl	11.23%	0.29%	4.88%	17.9%

DJF		WRF			
		gp	gam	exp	wbl
REGNIE	gp	8.92%	1.25%	24.66%	7.12%
	gam	2.18%	0.27%	7.65%	1.21%
	exp	1.44%	0.48%	8.08%	1.12%
	wbl	6%	0.89%	16.42%	12.31%

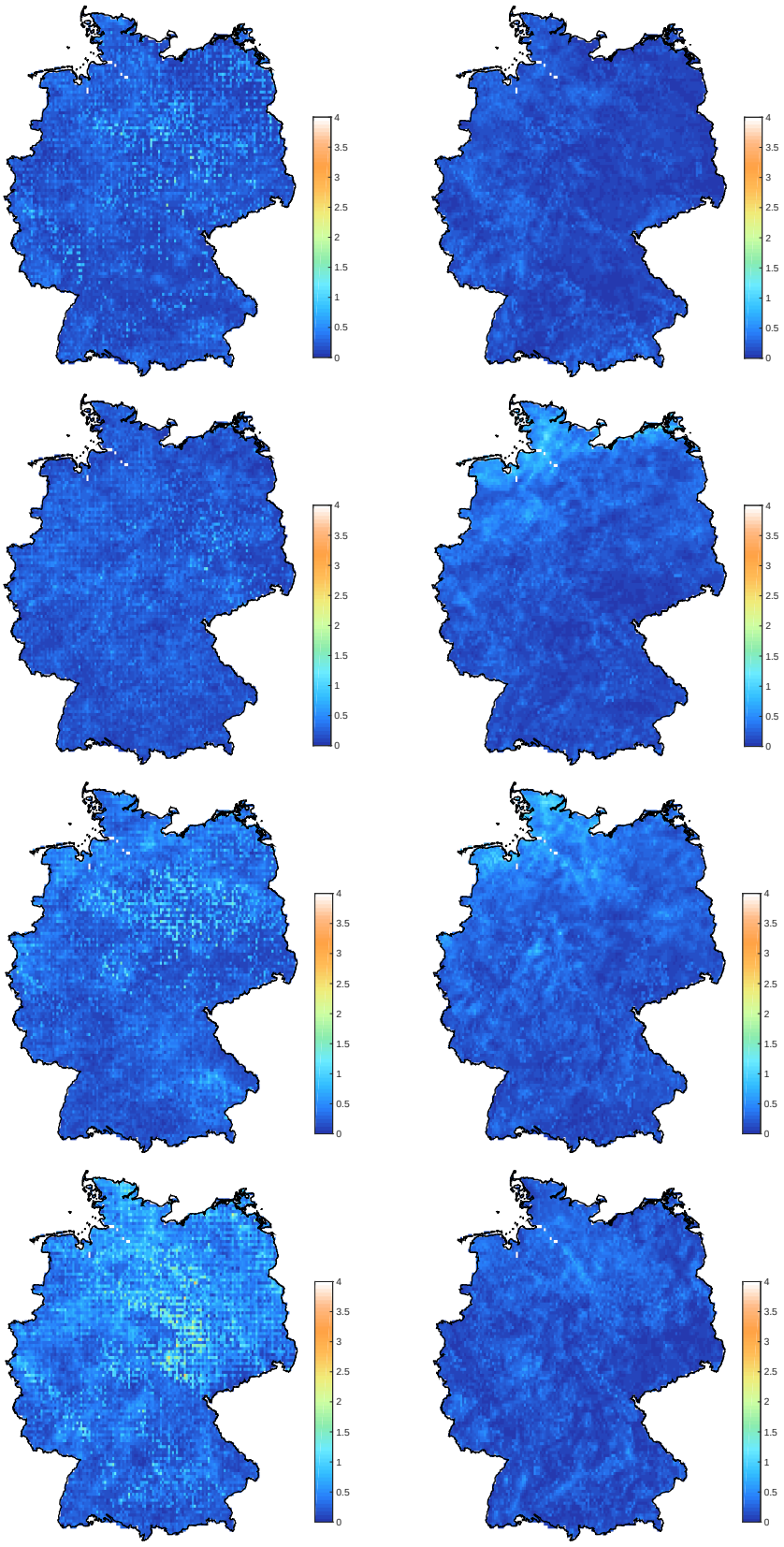

FIGURE 6.2: The residual sum of squares (RSS) between the empirical distributions and the fitted theoretical marginal distributions in each grid for both REGNIE and WRF data. The maps in left panel are the RSS for REGNIE data and the right panel indicate the RSS for WRF data. From top to bottom, the maps indicate RSS for spring (MAM), summer (JJA), autumn (SON) and winter (DJF), respectively. The results are shown for the calibration period (1971–1985).

TABLE 6.2: The proportion of grid cells for both REGNIE and WRF that K–S test failed and only BIC is used in Goodness-of-fit procedure.

	Spring	Summer	Autumn	Winter
REGNIE	25.83 %	10.86 %	38.38 %	56.13 %
WRF	0.31 %	10.61 %	12.26 %	3.88 %

of Figure 6.3 show the Q-Q plots between fitted theoretical and empirical distributions of REGNIE precipitation for selected pixel in each seasons. In spring, summer and autumn the pixels with the highest RSS are all fitted by the generalized Pareto distribution, while in winter it is fitted by the Weibull distribution. The sample quantiles match the theoretical quantiles quite well as most of the points (around 95 % of the points) on the Q-Q plot fall on a straight line (the major diagonal) except a few points which are slightly right-skewed in the extreme parts. This means in each selected pixels for 95 % of the quantiles the theoretical distribution matches the empirical distribution and for the rest 5 % of the quantiles the theoretical distribution has a thinner PDF than the empirical distribution. The Q-Q plot for WRF precipitation fields in selected pixels show similar patterns as that for REGNIE precipitation but with higher proportion of the matched quantiles. For all of the selected pixels, more than 98 % of the points in Q-Q plots fall on the straight line. In spring, the Q-Q plot shows an outlier in the left side while for other seasons the Q-Q Plots show also slight right-skewed patterns which indicates a slightly thinner PDF from the fitted theoretical distribution.

6.3 Identified Copula functions

While the marginal distributions are fitted, the precipitation are then transformed to rank space where the Copula function can be fitted to describe the dependence structure between REGNIE and WRF. In order to assess for the annual variability of the dependence structures between REGNIE and WRF precipitation time series, the Copula functions are also identified for the different seasons separately. In each season for each pixel, the REGNIE and WRF pair-wised precipitation are firstly transformed to the rank value based on the seasonal fitted marginal distribution. The pair-wised rank value are then fitted to a theoretical Copula function through the Goodness-of-fit testing which is described in Sect. 4.3. Four Copulas (Gaussian, Frank, Gumbel and Clayton) are selected as the candidates for Goodness-of-fit testing. The fitted Copula family maps for each season are shown in Fig. 6.5. It is easy to see that the patterns of the Copula

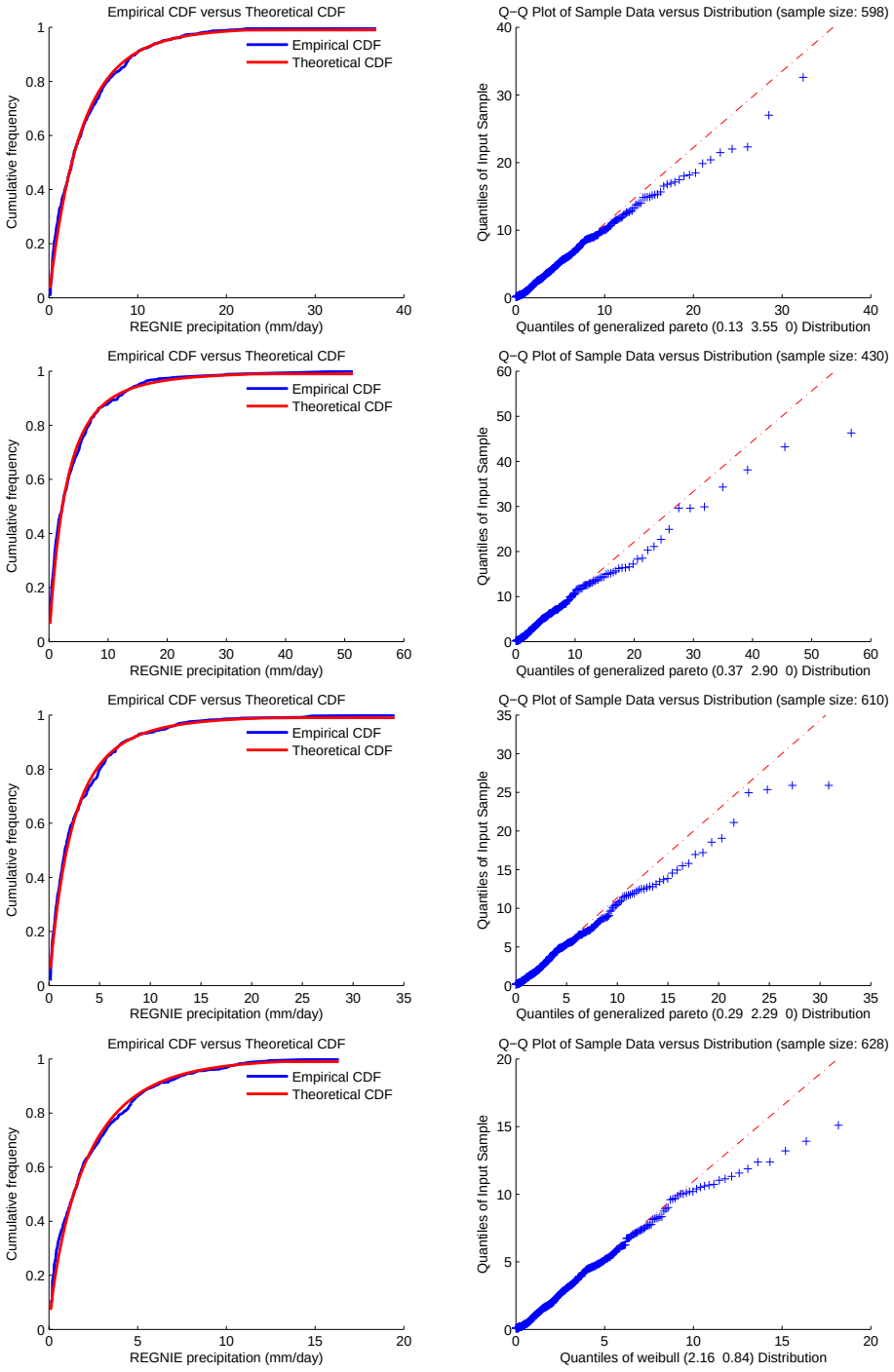

FIGURE 6.3: The empirical and fitted theoretical distributions of REGNIE precipitation (left panel) and the corresponding quantile-quantile plots (right panel) for the selected grid cells. From top to bottom, they represent the highest RSS value pixel in spring, summer, autumn and winter respectively.

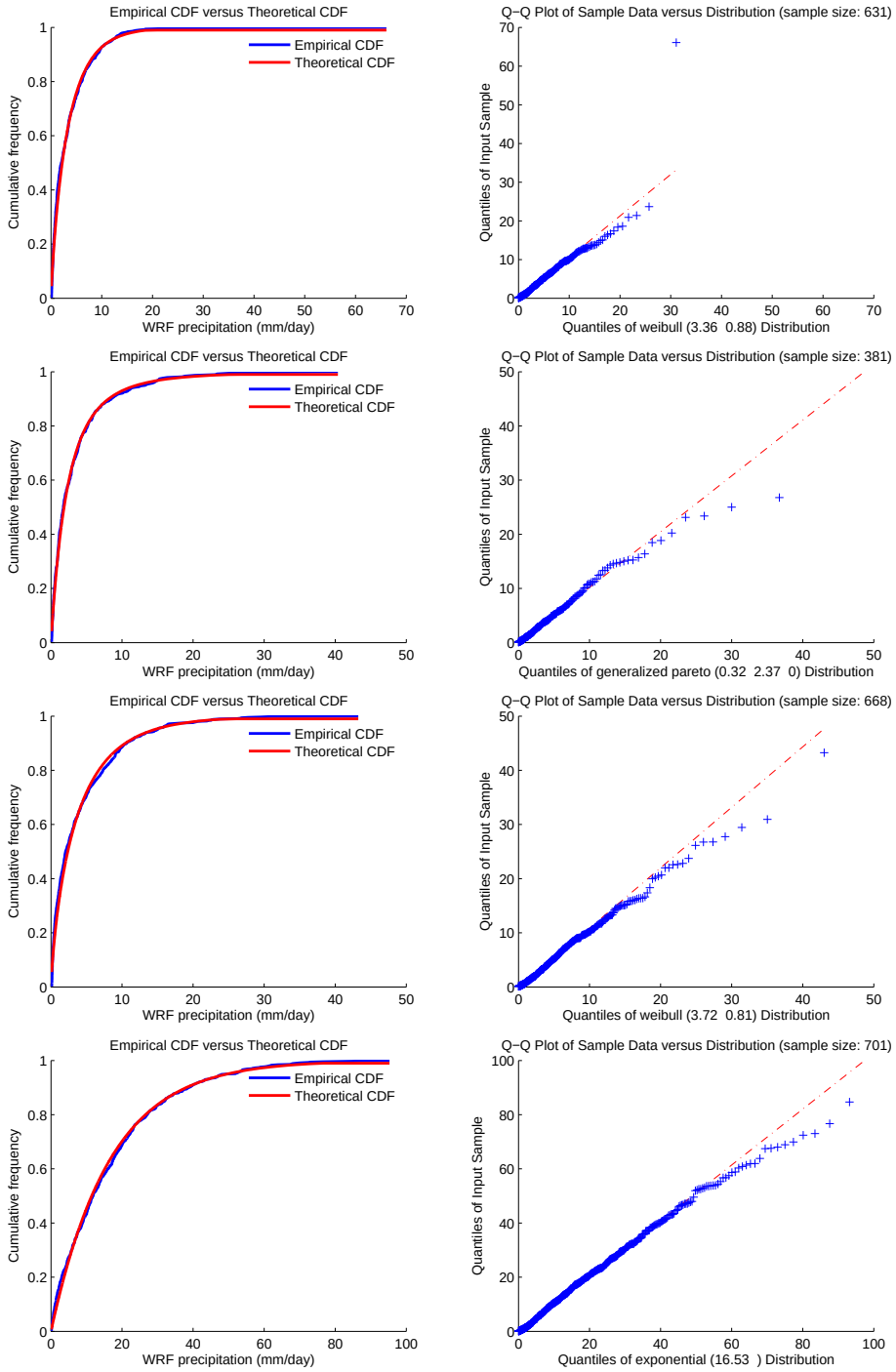

FIGURE 6.4: The empirical and fitted theoretical distributions of WRF precipitation (left panel) and the corresponding quantile-quantile plots (right panel) for the selected grid cells. The seasonal visual inspection of marginal distribution for WRF precipitation. From top to bottom, they represent the highest RSS value pixel in spring, summer, autumn and winter respectively.

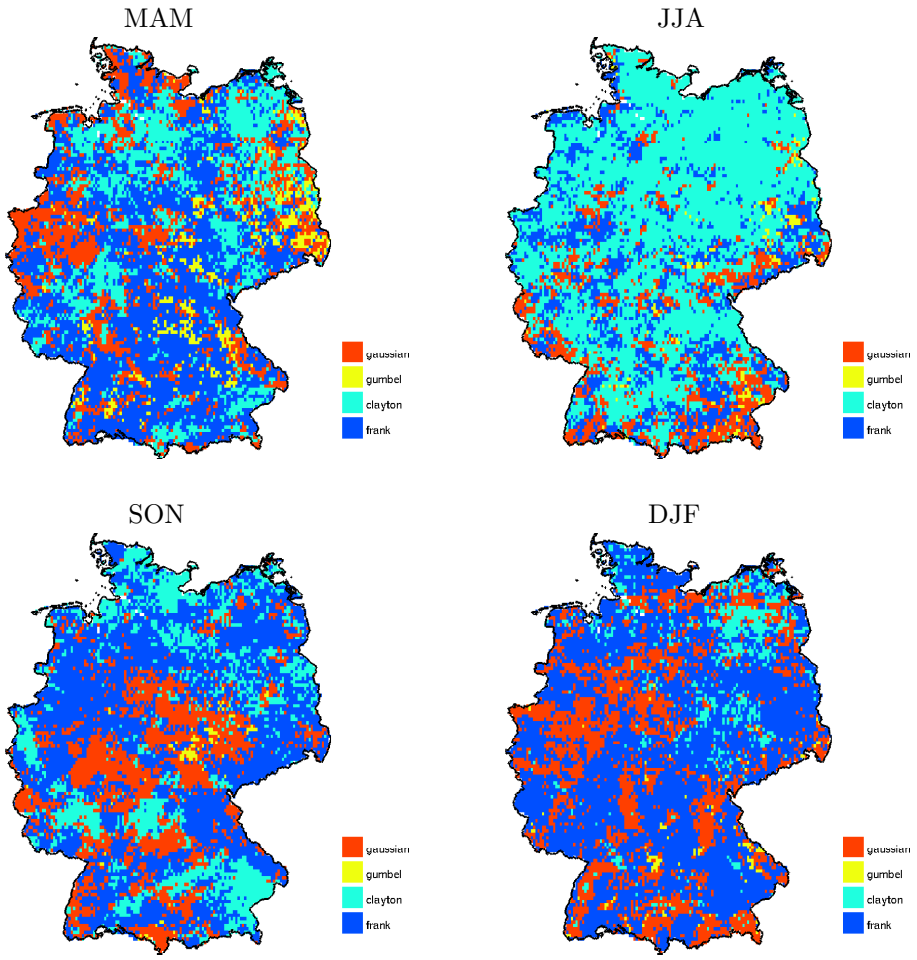

FIGURE 6.5: Fitted Copula functions between REGNIE and WRF precipitation (calibration period (1971–1985), positive pairs only). The Copulas are identified for the different seasons (spring – MAM, summer – JJA, autumn – SON, winter – DJF).

families are different in the different seasons. For spring, autumn and winter the Copulas that have no pronounced tail dependence (the Frank and Gaussian Copula) dominate (spring 49 % (Frank) + 22 % (Gaussian) = 71 %, autumn 53 % + 24 % = 77 % and winter 63 % + 28 % = 91 %), in summer the Clayton Copula provides the best fit for most of the grid cells (62 %) that indicates a lower tail dependence between REGNIE and WRF precipitation. For all seasons the Gumbel Copula is only selected for few grid cells with a maximum number of hits in spring (5 % of the grid cells). In general the differences are most prominent for winter and summer (see Fig. 6.5).

6.4 Validation of the Copula-based bias correction

The seasonal Copula model are establish while the marginal distributions and the Copula functions are estimated seasonally. Therefore the conditional distribution of REGNIE conditioned on WRF can be derived for each grid cell to generate bias-corrected WRF precipitation. As already mentioned in Sect. 4.4, the Copula-based stochastic bias correction generates the corrected precipitation through the Monte Carlo simulation with a sample size of 100. For each corrected step, a empirical predictive distribution is derived. Due to the practical reasons (e.g. spatial illustration or spatial evaluation), a "typical" value is taken as the estimator of this predictive distribution to get single corrected values. The expectation, median and mode value of the empirical predictive distribution are investigated and similar results as Fig. 5.7 are found. For the Copula-based bias correction with the seasonal Copula model approach, the expectation value of the Monte Carlo simulations has the best performance compared to other two "typical" values (i.e. the median and mode). Therefore, in the following the bias corrected value refers to the expectation value of the Monte Carlo simulations (100 realizations for each time step) only. To validate the Copula-based bias correction with the seasonal Copula model approach, the corrected WRF data are then compared to the uncorrected raw WRF data and it is shown in Fig. 6.6. Figure 6.6 (left panel) shows the original relative bias of mean daily precipitation between REGNIE and uncorrected WRF, while Figure 6.6 (right panel) shows the relative bias between REGNIE and corrected WRF by applying Copula-based method in the seasonal mode. Similar performance is found compared to the Copula-based correction in the overall mode (see Fig. 5.7). The wet biases are corrected for most of the domain and for some region small dry biases are introduced. The average of the bias for the whole study area is reduced from 10 to $-1\,\%$.

With respect to seasonal variations the correction performance is also evaluated in different seasons. The relative bias maps before and after bias correction in each season are shown in Fig. 6.7 left panel and right panel, respectively. It can be seen that the Copula-based correction in the seasonal mode efficiently removes most of the biases indicating a comparable performance for all seasons. Figure 6.7 especially for spring and winter indicates that the correction is tending to be more suitable to correct for overestimation of the rainfall. The underestimation of precipitation, that is most prominent in summer, however, is still significantly reduced. In autumn and winter the Copula-based correction in the seasonal mode reduces the rainfall amounts too much for the west of Germany, introducing a small dry bias in that region. The average bias are reduced

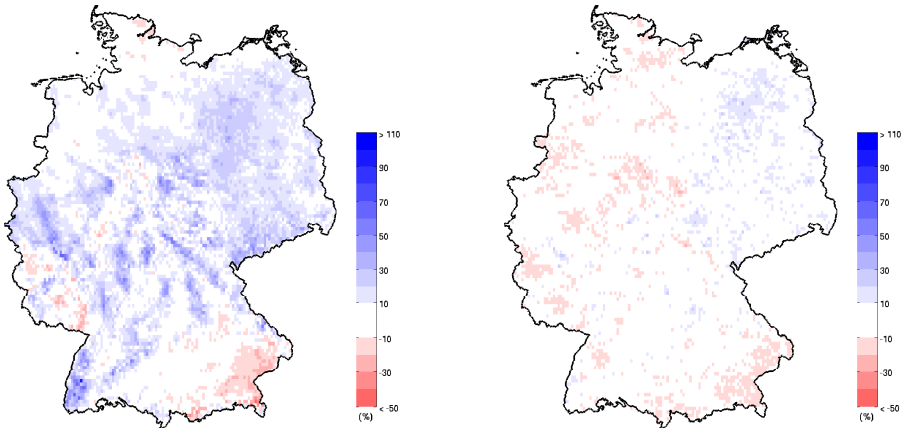

FIGURE 6.6: Relative bias map of mean daily precipitation for uncorrected precipitation field (left) and corrected WRF precipitation field by applying Copula-based method in seasonal mode. The results are based on the validation period 1986-2000.

to 16, −11, −1 and −3 % respectively for different seasons from spring to winter. By comparing to the correction in the overall mode (see Fig. 5.8), seasonal mode based correction shows a great improvement especially in summer and winter. In summer, the seasonal mode based correction is significantly improved in the entire study area. In the overall mode, the Copula-based correction shows even an increasing of dry biases compared to the original bias map. By applying the correction in the seasonal mode, the correction decreases the dry biases from −15 to −11 %. In winter, the seasonal mode is found to be able to remove the wet bias in the south and west of Germany which can not be removed by correction in the overall mode. In spring and autumn, the seasonal mode approach is also found to be improved slightly.

In order to investigate the seasonal performance in more detail, we also look at the intra-annual variability of observed and corrected model precipitation. As mentioned in chapter 5, the four specific pixels are again selected to evaluate their monthly mean precipitation. These four pixels represents four different typical situations (see Sect. 5.4). For the selected four pixels their monthly mean precipitation of REGNIE, WRF, corrected WRF in the overall mode and the corrected WRF in the seasonal mode are shown in one plot in Fig. 6.8 indicated by different colors. In each season, the Copula-based bias corrected monthly mean precipitation in the seasonal mode (the black line in Fig. 6.8) are further improved compared to the corrected monthly mean precipitation in the overall mode (blue line in Fig. 6.8). For pixel 1, the correction in the seasonal mode has improved all the year except only in December, where the monthly mean precipitation is over estimated a bit. The same performance is found for pixel 3, while the corrected monthly mean precipitation in the seasonal mode are closer to the observation

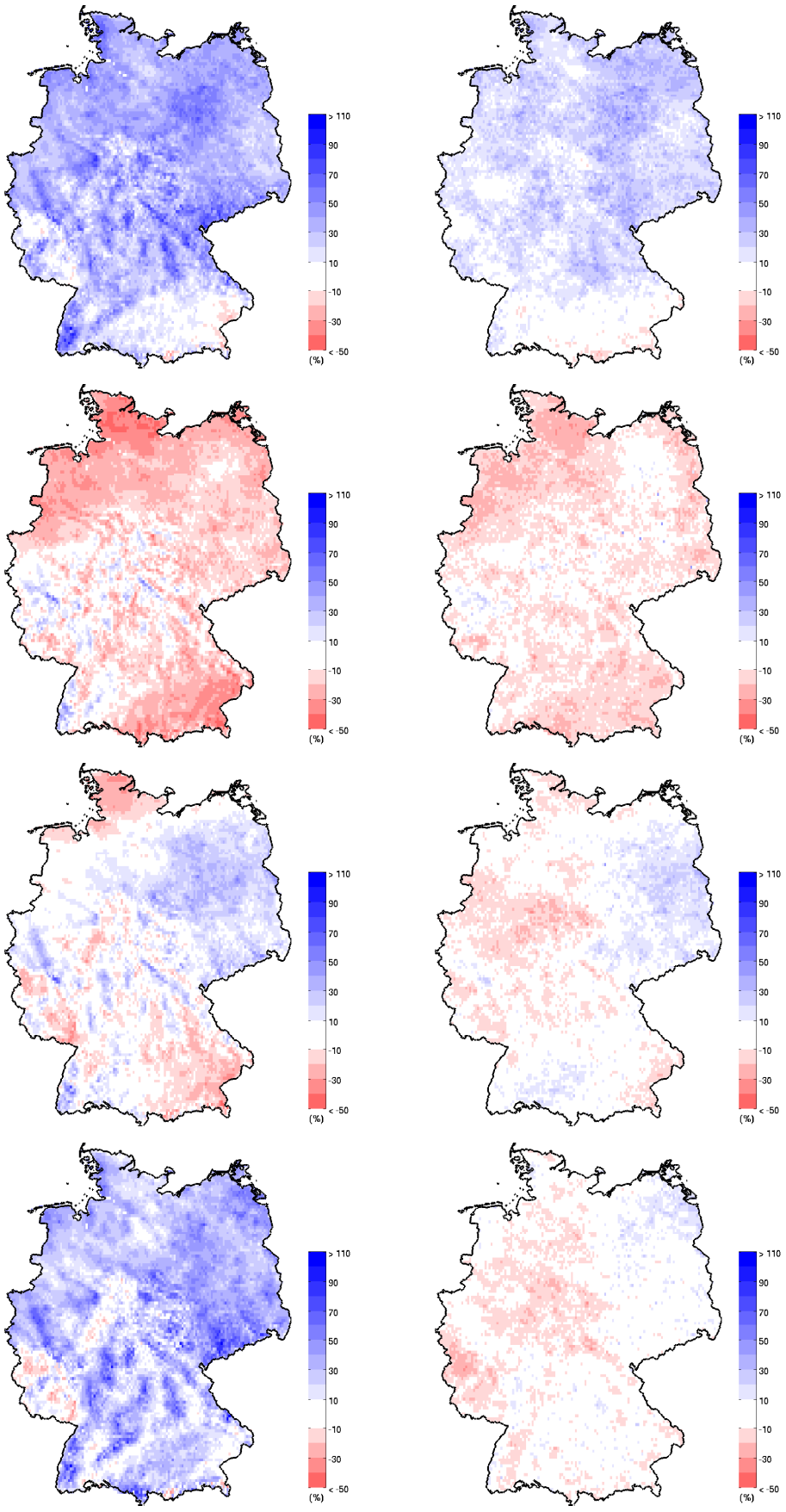

FIGURE 6.7: Relative bias between uncorrected (left) and corrected (right) WRF mean daily precipitation and the REGNIE data set in Germany for the different seasons (spring–MAM, summer–JJA, autumn–SON, winter–DJF, from top to bottom). The results are derived for the validation time period (1986-2000).

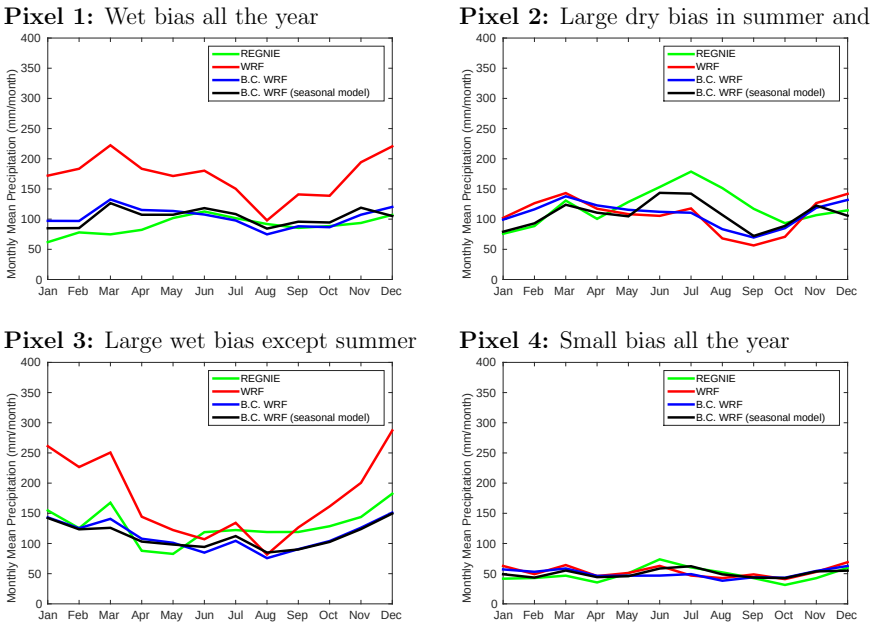

FIGURE 6.8: Comparison of bias corrected WRF mean monthly precipitation (blue: in overall mode, black: in seasonal mode) with REGNIE (green) and original WRF data (red) for the selected four pixel 1–4 in the validation period from 1986 to 2000. The number of the respective grid cell is noted in the upper left corner of each plot.

expect in March (precipitation is reduced too much). For pixel 2, the dry biases in summer is significantly reduced by the seasonal mode correction. For pixel 4 even the WRF model simulated precipitation is good enough, the correction can still improve the results by applying in the seasonal mode approach.

As described in chapter 4, the Copula-based correction is applied for each grid cell separately. Therefore, it is necessary to investigate the spatial coherence of the bias corrected precipitation fields. The sequence of three selected days (from January 9th to 11th, 1986) are exemplarily shown in Figure 6.9. The left panel from top to bottom are the observed precipitation fields for these three days. In the middle are the uncorrected (original) WRF simulated precipitation fields and the right panel indicates the bias corrected WRF precipitation fields in the seasonal mode. The results show that the WRF simulated precipitation show an overestimation in these days, and the Copula-based correction is able to reduce the precipitation amount therefore correct the wet biases. It can also be seen that the corrected fields follows the pattern of raw WRF data. Which means that while correcting the absolute precipitation values, the spatial coherence of the precipitation patterns are retained after the application of the bias correction.

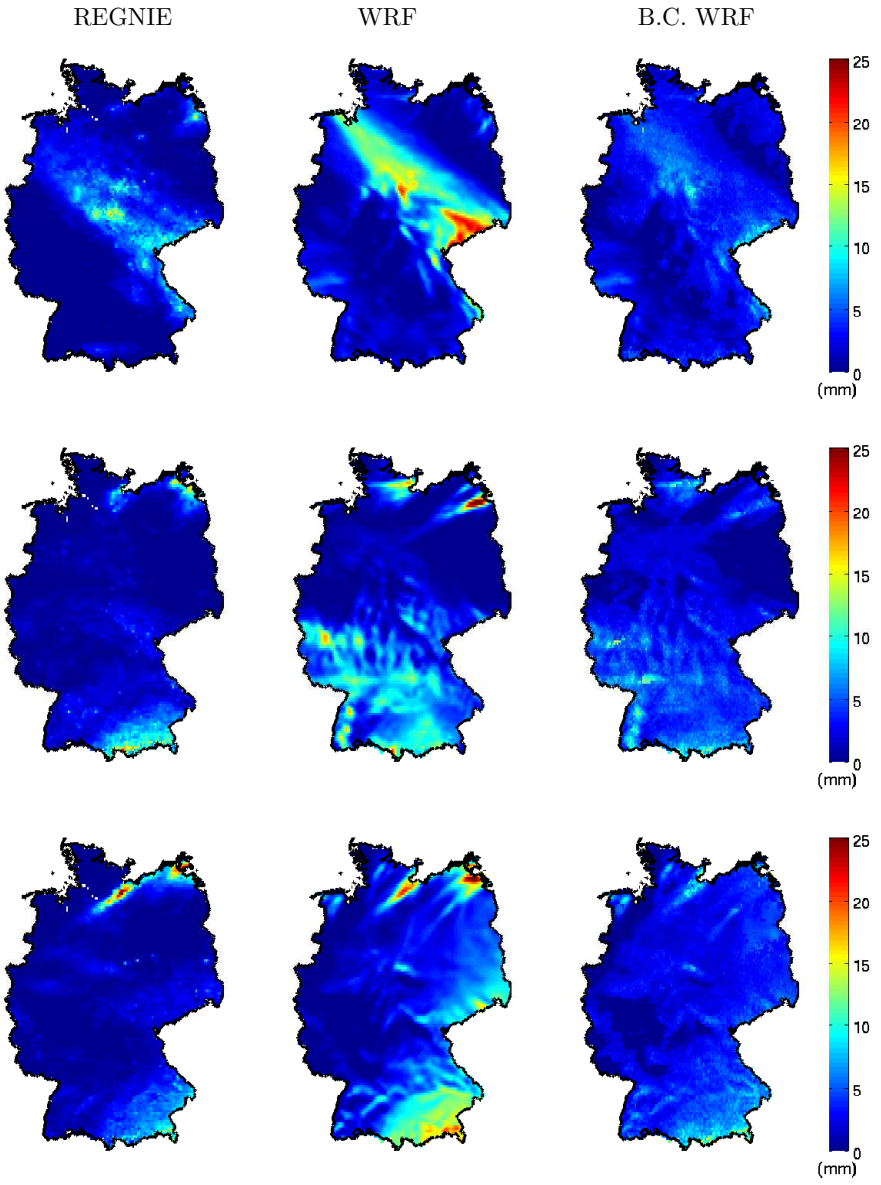

FIGURE 6.9: Daily precipitation fields over Germany for the three consecutive days from January 9 to January 11, 1986.

6.5 Summary and discussion

As already discussed in previous Chapter, the Copula-based bias correction by taking the expectation value of the Monte Carlo simulations is able to significantly reduce the relative bias of the mean daily precipitation for the fifteen years validation time. However, it has limitations to reduce the seasonal bias of the mean daily precipitation. Therefore, in this Chapter the stochastic bias correction is applied in a seasonal mode.

The marginal distributions as well as the Copula functions are estimated for each pixel in each season (spring – MAM, summer – JJA, autumn – SON, winter – DJF). The marginal distribution and Copula fitting follows the same strategy as in chapter 5. Again, only the positive pairs of REGNIE and WRF precipitation are used to construct the seasonal Copula models, therefore the models are also only applied for the (1,1) cases. The time steps that are not belong to the (1,1) cases are then kept same as the raw WRF data.

The fitted seasonal marginal family maps show that the precipitation distributions change over time and space for both REGNIE and WRF. Differences are found between fitted marginal distribution of REGNIE and WRF, which means that the WRF model has the shortcomings for simulating the precipitation distribution in seasons. The same as the annual fitted marginal distribution, in seasonal fitting the patchy patterns are also found for REGNIE precipitation. For seasonal fitted marginal distributions, the number of K–S fails are significantly reduced for both REGNIE and WRF. To evaluate the Goodness-of-fit tests, visual inspections are also applied, which give great agreement between fitted theoretical distributions and empirical distributions.

The estimated Copula functions are different in different seasons and in different area, which indicates that the dependence structures between REGNIE and WRF precipitation vary over time and space. Based on the seasonal fitted marginal distributions and Copula functions, the seasonal Copula models are established. The conditional distribution are then derived seasonally. By comparing the relative bias of mean daily precipitation, the Copula-based correction in the seasonal mode is found to be able to significantly reduce the biases in each seasons. This is due to the reason that precipitation distributions and the dependence structure between REGNIE and WRF precipitation change over seasons, with the seasonal fitted models these information can be better captured and therefore derive better corrections. The bias of monthly mean precipitation are also analyzed for four selected pixels with respect to intra-annual variability.

Results show that with seasonal fitted Copula models, the correction can be further improved.

Finally, the spatial coherence of the Copula-based correction are also investigated due to the reason that the copula models are estimated for each pixel separately. By looking at the sequence of three selected days (from January 9th to 11th, 1986), the seasonal mode corrections are found to be able to reduce the biases and also preserve the spatial structure of the WRF output. This is due to the fact that the Copula-based approach is conditioned on the WRF simulation. The method adjusts the value of the WRF precipitation according to the fitted Copula model. Even though the Copula models are estimated for each grid cell, the spatial coherence is captured by the Copula model as both the Copula families as well as the marginal distributions are also spatially clustered.

Chapter 7

Comparison of the Copula-based bias correction to the quantile mapping

7.1 Introduction

In chapters 5 and 6, the Copula-based stochastic bias correction is applied to the WRF simulated precipitation field by using both the overall Copula model approach and the seasonal Copula model approach. The results show that both Copula models are able to reduce the biases in the WRF derived precipitation field, while the seasonal Copula model approach has a better performance. In this chapter, in order to further evaluate the Copula-based bias correction, it is compared to the traditional quantile mapping correction. Like the Copula-based correction, the quantile mapping correction is also applied for all pixels in the entire study area. The corrected precipitations are then compared to the Copula-based corrections by looking at the root mean square error (RMSE), the quantile RMSE and the percentage of the corrected time steps that are closer to the observations. Due to the fact that the seasonal Copula model based correction are better than the overall Copula model, the comparison the Copula-based correction refers to the seasonal Copula model approach only.

7.2 Empirical distribution derived quantile mapping correction

The quantile mapping method is often used in bias correction of RCM derived precipitation (Dosio and Paruolo, 2011; Gudmundsson et al., 2012). This method corrects the bias by rescaling the values of the RCM so that the distribution of the RCM matches that of the observations. The quantile mapping method has already depicted in Sect. 1.2, for the sake of completeness it is briefly described here. The correction formula is followed as:

$$P^* = F_{obs}^{-1}(F_{RCM}(P)) \tag{7.1}$$

where P^* is the corrected model precipitation value, P is the original modelled precipitation value. F_{RCM} is the CDF of the RCM precipitation and accordingly F_{obs} is the CDF of the observations. F_{obs}^{-1} is the inverse function of F_{obs}, which is named quantile function.

In this study, the empirical distribution is used to model the CDF of both the REGNIE and WRF precipitation. Therefore the empirical distribution constructed transfer function(Eq. 7.1) is hence referred to as empirical quantile mapping (eQM).

To implement the empirical distribution correction method, the ranked modelled precipitation distribution is divided into a number of discrete quantiles. Following the procedure of Lafon et al. (2013), the empirical CDF is approximated using tables of empirical percentiles. Values in between the percentiles are approximated using linear interpolation. The same procedure is done for generating the empirical CDF of the observation. The inverse computation is simply using the quantile function. For each grid cell in the domain, the empirical quantile mapping correction has performed.

7.3 Comparison of the Copula-based bias correction to the quantile mapping correction

The quantile mapping corrected WRF precipitations has compared to the Copula-based correction by investigating the RMSE, quantile RMSE and the percentage of the corrected time steps that are closer to the observations.

Firstly, The RMSE between the observed (REGNIE) and bias corrected modelled data is calculated for both the Copula-based correction and the quantile mapping method. The

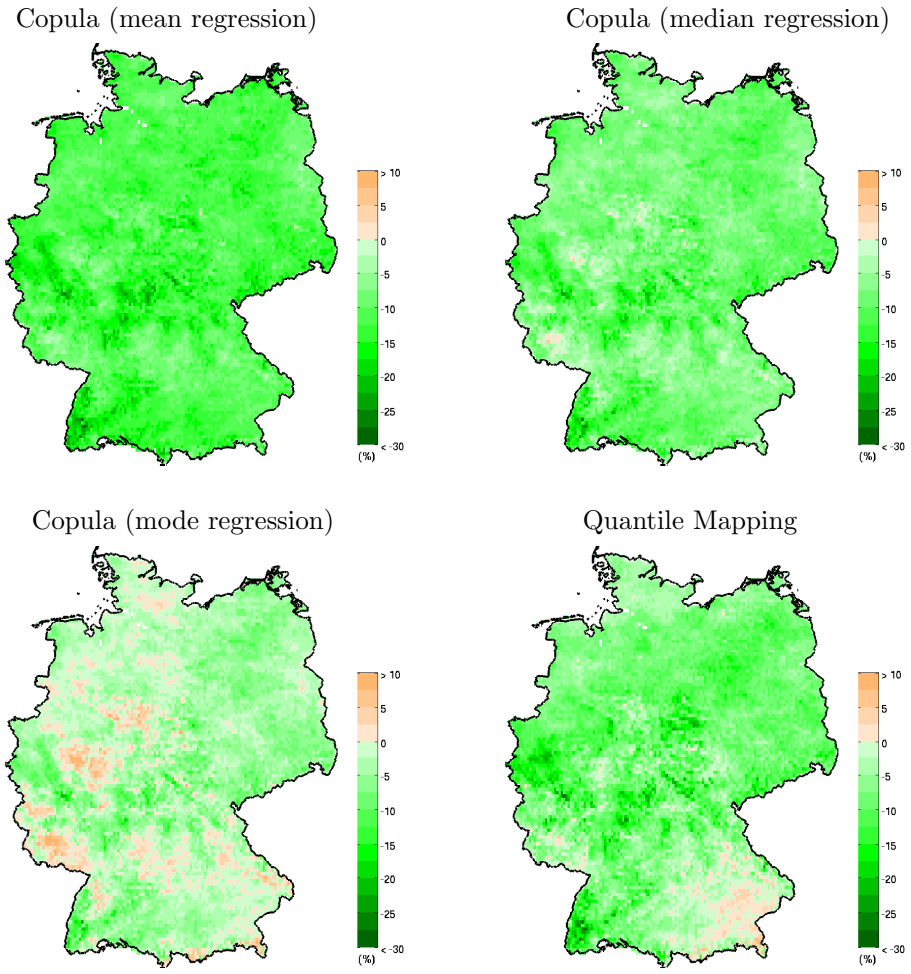

FIGURE 7.1: The changes of the RMSE in the validation period (1986-2000) by different bias correction methods. The green color indicates a decrease of the RMSE, while the ocher color implies an increase of the RMSE.

original RMSE (between REGNIE and WRF) is also computed as a reference. For the Copula-based approach, the RMSE is calculated for all the simulations with respect to the mean-, median- and mode value. The changes of the RMSE by different corrections over the study area are shown in Fig. 7.1. The Copula-based correction derived from the mean regression reduces the RMSE significantly with an average of -12% over the domain. The Copula-based correction derived from the median also reduces the RMSE, but to a less degree. The correction derived from Copula-based mode regression reduces the RMSE, but results in an increase of the RMSE in some regions. The same holds true for the quantile mapping approach.

To further assess the performance of the Copula-based method, additional performance measures are analyzed. The RMSE for different magnitudes of observed precipitation

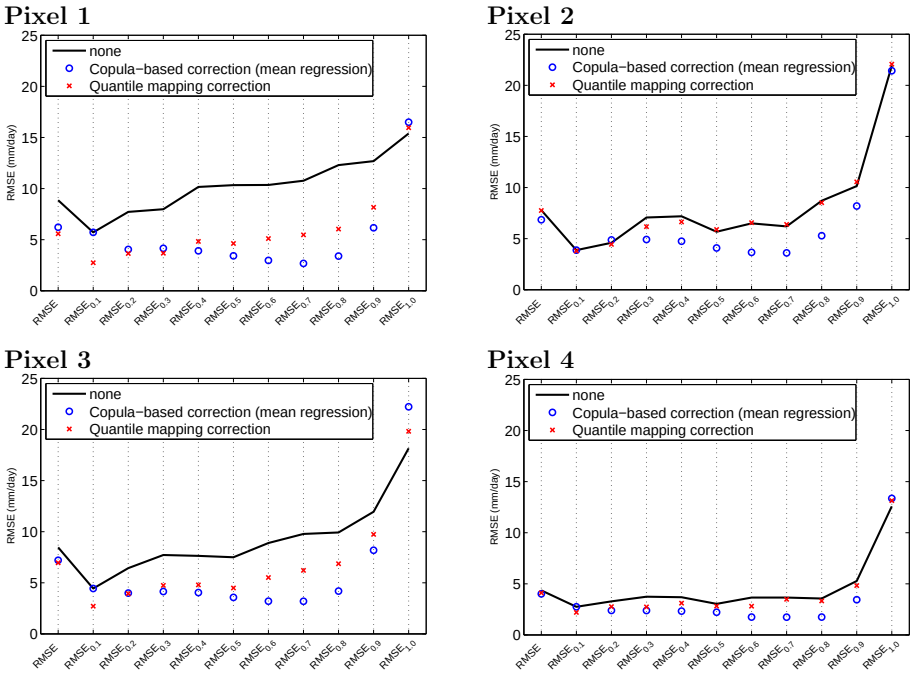

FIGURE 7.2: The root mean square errors (RMSE) and the root mean square errors for specific probability intervals ($RMSE_{0.1}$, $RMSE_{0.2}$, ..., $RMSE_{1.0}$) for different methods. The selected four pixels are the same as in Fig. 6.8. The black solid line indicates the errors without correction. The results are derived from the validation period from 1986 to 2000.

(i.e. a quantile RMSE analysis) is done for the selected four grid cells (see Fig. 2.1). The results from the validation period are shown in Fig. 7.2. The RMSE in different quantiles are represented by $RMSE_{0.1}$, $RMSE_{0.2}$, ..., $RMSE_{1.0}$, while the subscript indicates the magnitude level. $RMSE_{0.1}$ evaluates the errors in the dry part of the observation distribution, implying the (0,1) errors. From $RMSE_{0.2}$ to $RMSE_{1.0}$ the RMSE are calculated for equally spaced probability intervals of the observed empirical CDF of wet days. For example, $RMSE_{1.0}$ indicates the errors in the magnitude of the 10% highest events. As it can be seen from Fig. 7.2, the Copula-based correction performs equally or even better in terms of the RMSE in most of the quantiles. In pixel 1 and 3, both correction are able to reduce the quantile RMSE, while the Copula-based correction is more efficient. For pixel 2, the quantile mapping failed to decrease the quantile RMSE for nearly all the quantiles, while the Copula-based method reduce the errors significantly. In pixel 4, the original RMSE for each quantiles are low. By applying the quantile mapping correction the RMSE are only slightly reduced but the Copula-based method corrects the errors significantly. It is also found that the Copula-based correction is not able to reduce the RMSE in the first quantile ($RMSE_{0.1}$) and sometimes also failed in correcting RMSE in the extreme quantile ($RMSE_{1.0}$).

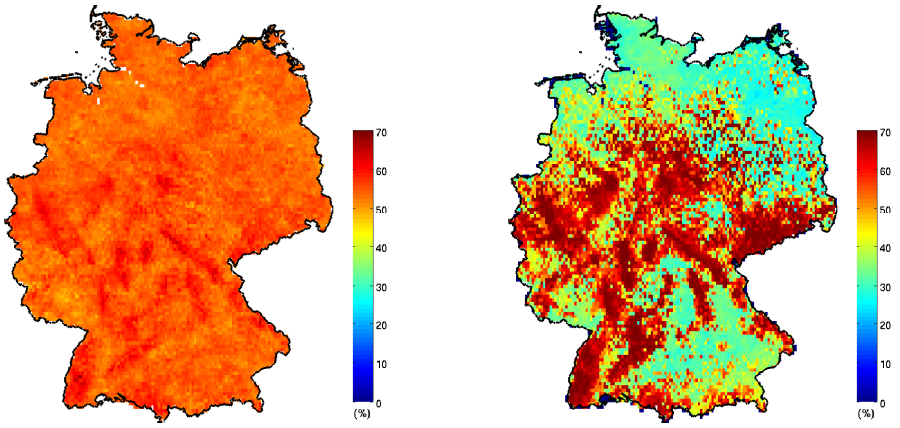

FIGURE 7.3: The percentage of the corrections that are closer to the observations. Left: Copula-based correction (mean regression); Right: quantile mapping correction. The results are derived from the validation period from 1986 to 2000.

Furthermore, the percentage of the corrected time steps that are closer to the observations compared to the quantile mapping method is investigated. The results are shown in Fig. 7.3. The values indicate the percentage of the successful corrections (i.e. closer to the observations) by the two bias correction methods. It can be seen that the results of the quantile mapping correction strongly depends on the rank correlation (see Fig. 7.4), while the Copula-based correction provides a stable correction efficiency over the entire domain. The average percentages of the successful correction are 55% for the Copula-based correction and 46% for the quantile mapping correction, respectively.

7.4 Summary and discussion

The Copula-based correction is found to be able to efficiently reduce the biases in the WRF derived precipitation. In order to further evaluate the proposed Copula-based bias correction it is compared to the most often used quantile mapping method. The empirical quantile mapping correction is applied for each pixel in the study area. As analyzed and discussed in the previous chapters (chapter 5 and 6), the seasonal Copula model based correction has a better performance. The results show that the Copula-based method has an improved performance in reducing the overall RMSE of WRF precipitation. The quantile mapping correction even increases the RMSE in some regions. The analysis of quantile RMSE for selected four pixels shows that the Copula-based correction performs equally or even better in terms of the RMSE in most of the quantiles. Finally the correction efficiency is compared between the Copula-based method and the quantile mapping correction regarding the percentage of the corrected time steps that are closer

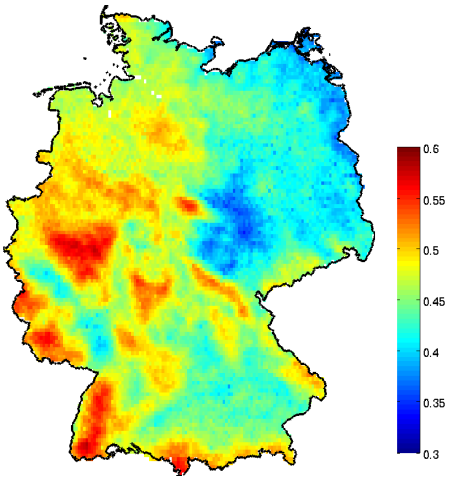

FIGURE 7.4: The rank correlations between RCM and REGNIE precipitation over the domain in the validation period from 1986 to 2000.

to the observations. Results show that the correction efficiency of the quantile mapping correction strongly depends on the rank correlation, while the Copula-based correction provides a stable correction efficiency over the entire domain and has a higher score.

It is well known that the quantile mapping method corrects all moments of the RCM precipitation distribution. However this correction is usually applied under the assumption of a perfect dependence among the ranks and this full dependence assumption is limited. In our study area, the rank correlation between the datasets varies between 0.3 and 0.6 (see Fig. 7.4). Therefore the Copula-based method are more efficient than the traditional quantile mapping correction, as it has no limitation of full dependence assumption and the Gaussian assumption.

Chapter 8

Outlook: Framework for the application of the Copula-based correction in climate projections

8.1 Introduction

In the previous chapters, a Copula-based stochastic bias correction method was introduced and applied for the correction of WRF derived precipitation. The focus was to demonstrate the feasibility of this method to correct precipitation time series for the past. If applied to future climate projections, this method has to be extended further. The proposed Copula-based method is applied for positive pairs of REGNIE and WRF data, i.e. the (1,1) cases. When dealing with future climate information, e.g. from climate projections (where no observations are available), one cannot identify if a time step belongs to (1,1) cases or not. In this case, the methodological framework as ooutlined in the previous sections has to be extended.

In this chapter, the Copula-based bias correction framework is further elaborated to allow its use for climate projections by applying a Markov-based precipitation cases identification model. This model allows the identification of (0,0), (0,1), (1,0) and (1,1) cases for every time step in the future period. The methodology is briefly described and applied for selected four pixels in a validation mode, i.e. it is assumed that no observations are available for the period under consideration.

8.2 Precipitation cases identification model

The proposed precipitation cases identification method consists of a hidden Markov model and a Viterbi process. A first-order hidden Markov model (HMM) is used to model the precipitation cases, i.e., (0,0), (0,1), (1,0) and (1,1) cases, in precipitation pairs between REGNIE and WRF. Four Markov states are set up to describe these four precipitation cases: S1 indicates the (0,0) case, S2 implies the (0,1) case, S3 represents the (1,0) case and S4 stands for the (1,1) case. The state diagram is shown in Fig. 8.1. Each state has a probability to move to other states or stay in the same state. In this model, the emissions refer to the WRF derived precipitation and are only configured with two levels, i.e. $E = \{d, w\}$, where d indicates a dry day and w stands for a wet day. Therefore, the emission probability is known as follows: S1 has 100 % chance to generate a dry day; S2 has 100 % chance to generate a wet day; S3 has 100 % chance to generate a dry day; S4 has 100 % chance to generate a wet day. Note again that the emissions refer to the WRF derived precipitations, since the purpose of this method is to predict the four precipitation cases in the future where only WRF precipitations are available.

The first-order HMM is firstly estimated based on the training datasets (i.e. the past observations and WRF simulations). The future precipitation cases can then be predicted through the Viterbi algorithm by using the estimated HMM and the precipitation projections. The details of the hidden Markov model and the Viterbi algorithm are described in Appendix A and Appendix B. To illustrate the precipitation cases identification method, a small example is also shown in Appendix C by taking the data from the pixel 1 in Fig. 2.1.

8.3 Application of the Copula-based bias correction combined with the precipitation cases identification

To evaluate the precipitation cases identification method, four pixels are selected to apply with it and the results are analyzed. The selected four pixels are marked in Fig. 2.1. The 30-year time series is split into a calibration (1971–1985) and a validation (1986–2000) period of equal length. For each pixel, the hidden Markov model is firstly estimated in calibration period. The precipitation cases are then predicted in validation period by using the Viterbi algorithm.

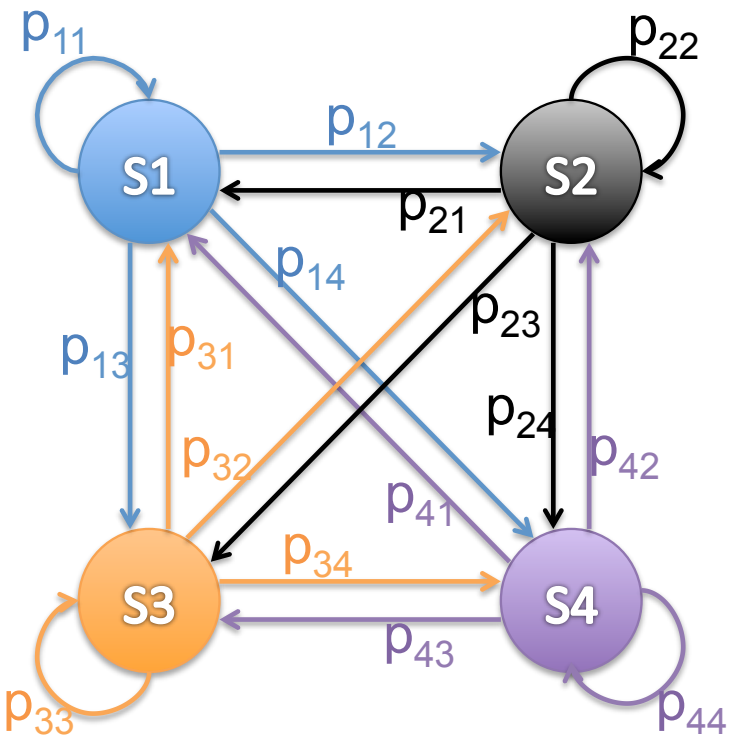

FIGURE 8.1: State diagram for the hidden Markov model with four states from S1 to S4. These four states indicates four different cases in precipitation pairs: S1 indicates the (0,0) case, S2 implies the (0,1) case, S3 represents (1,0) the case and S4 stands for the (1,1) case.

The predicted precipitation cases are compared to the actual precipitation cases in the validation period via confusion matrix and is shown in Table 8.1. Each row of the matrix represents the numbers of actual precipitation cases, while each column represents that of the predicted precipitation cases (in %). The major diagonal shows the fraction of concurring precipitation cases. By summing up the major diagonal, it can be seen that the accuracy of the precipitation cases identification for these four pixels are 72 %, 71.5 %, 69 %, 69 %, respectively.

In the prediction model, it is also found that (1,0) cases are quite often predicted as (0,0) and the (0,1) cases are easy to be predicted as (1,1) cases. E. g. for pixel 1, there are 9.2 % (8.6 % + 0.8 %) of days belong to the (1,0) cases in actual situation, however 93 % (8.6 % ÷ 9.2 %) of them are predicted as (0,0) and only 7 % (0.8 % ÷ 9.2 %) are predicted correctly. There are also 62 % (10.92 % ÷ (6.54 % + 10.92 %)) of the (0,1) cases are predicted as (1,1) cases and only 38 % (6.54 % ÷ (6.54 % + 10.92 %)) of them are predicted as such. The prediction accuracy for (0,0) and (1,1) cases are high, especially for the (0,0) case, which is 96 % (29.68 % ÷ (29.68 % + 1.24 %)). For the (1,1) cases, the

TABLE 8.1: Confusion matrix between the actual precipitation cases and the predicted precipitation cases for selected four pixels.

Pixel 1		Predicted cases			
		(0,0)	(0,1)	(1,0)	(1,1)
Actual cases	(0,0)	29.68%	0%	1.24%	0%
	(0,1)	0%	6.54%	0%	10.92%
	(1,0)	8.4%	0%	0.8%	0%
	(1,1)	0%	7.41%	0%	35.01%

Pixel 2		Predicted cases			
		(0,0)	(0,1)	(1,0)	(1,1)
Actual cases	(0,0)	20.6%	0%	4.9%	0%
	(0,1)	0%	1.63%	0%	13.8%
	(1,0)	7.8%	0%	4.9%	0%
	(1,1)	0%	2%	0%	44.3%

Pixel 3		Predicted cases			
		(0,0)	(0,1)	(1,0)	(1,1)
Actual cases	(0,0)	26.6%	0%	3.31%	0%
	(0,1)	0%	2.37%	0%	14.13%
	(1,0)	9.55%	0%	3.81%	0%
	(1,1)	0%	4.03%	0%	36.21%

Pixel 4		Predicted cases			
		(0,0)	(0,1)	(1,0)	(1,1)
Actual cases	(0,0)	26.01%	0%	3.6%	0%
	(0,1)	0%	6.35%	0%	14.44%
	(1,0)	9.8%	0%	3.19%	0%
	(1,1)	0%	3%	0%	33.6%

prediction accuracy is 83 % (35.01 % ÷ (7.41 % + 35.01 %)). Similar results are found for other three pixles.

The precipitation cases identification model allows the prediction of the precipitation cases in the future, where only RCM projections are available. Therefore it can be

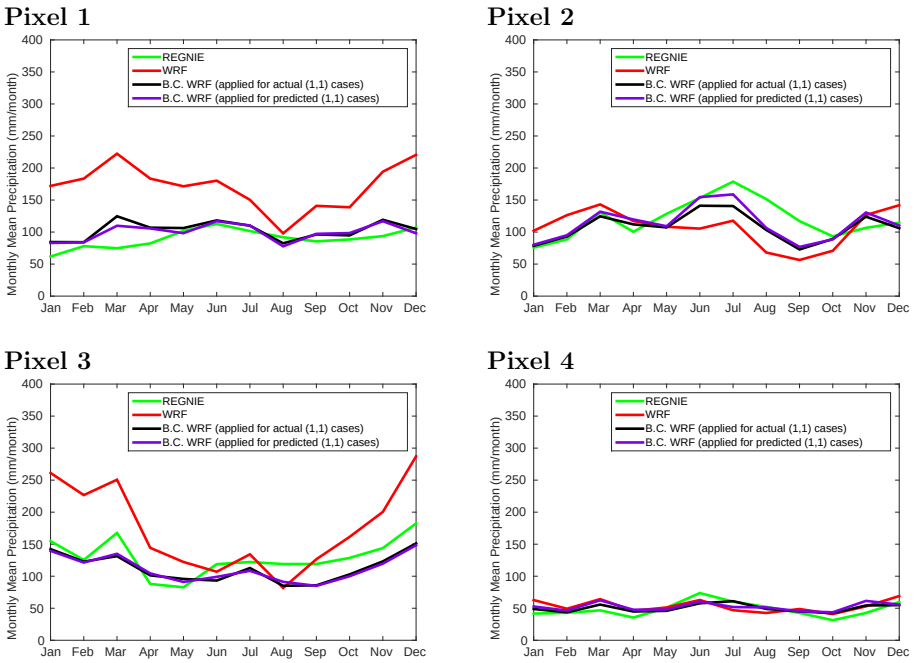

FIGURE 8.2: Comparison of bias corrected WRF mean monthly precipitation (black: applied for actual (1,1) cases, purple: applied for predicted (1,1) cases) with REGNIE (green) and original WRF data (red) for the selected four pixels (marked in Fig. 2.1) in the validation period from 1986 to 2000.

applied together with Copula-based bias correction to reduce the errors in RCM projections. We assume the validation time to be the future. By applying the Copula-based correction for these predicted (1,1) cases. The bias corrected mean monthly precipitation are shown in Fig. 8.2. It can be seen that with the predicted (1,1) cases, the errors are also significantly reduced. It has nearly the same performance as the application of the bias correction for the actual (1,1) cases. For some pixels, the performance of the application for predicted (1,1) cases are even better than that for the actual cases in some months. Furthermore, the RMSE and the quantile RMSE are analysed and the results are shown in Fig. 8.3. For all the selected pixels, the bias correction applied for predicted (1,1) cases also has close performane compared to that for the actual (1,1) cases. The quantile RMSE are only slightly higher due to the predict accuracy. It is also necessary to note that the first quantile RMSE ($RMSE_{0.1}$) is reduced with the predicted (1,1) cases. This is due to the reason that some of the (0,1) cases are predicted as (1,1) cases (see Table 8.1), therefore these time steps are then corrected by the Copula-based approach.

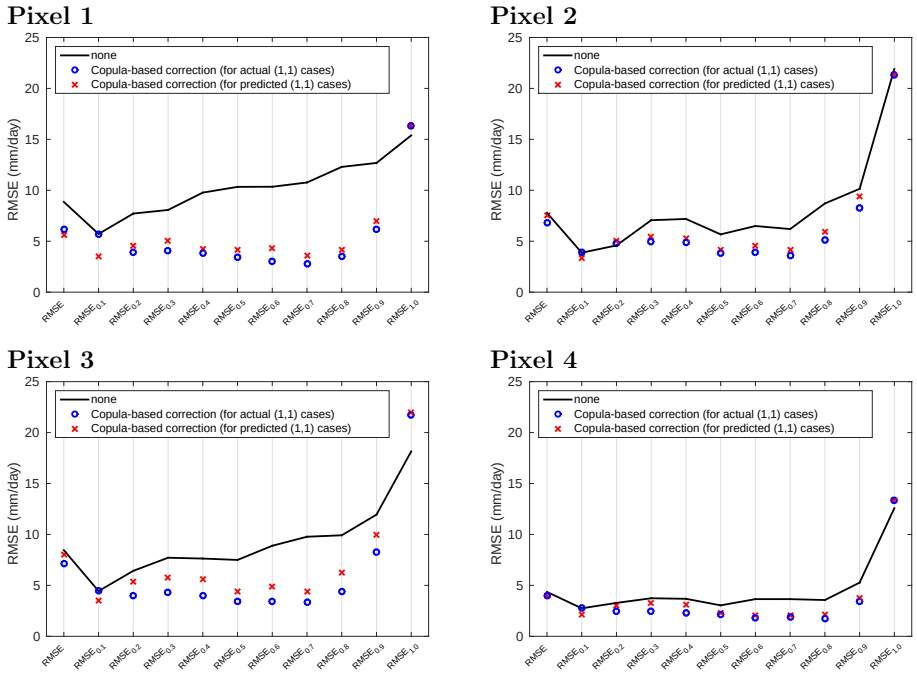

FIGURE 8.3: The root mean square errors (RMSE) and the root mean square errors for specific probability intervals ($\mathrm{RMSE}_{0.1}$, $\mathrm{RMSE}_{0.2}$, ..., $\mathrm{RMSE}_{1.0}$) for different methods. The selected four pixels the same as in Fig. 8.2. The black solid line indicates the errors without correction. The results are derived from the validation period from 1986 to 2000.

8.4 Summary

In this chapter the Copula-based correction is further extended to be capable of applyig in the climate projections. The approach is not elaborated in full detail in the Ph.D. dissertation thesis. It is rather focused on the validation instead of the application for the entire domain. Results show that the method is able to predict the precipitation cases with a high overall accuracy (around 70 %). For the (1,1) cases, the prediction accuracy is even higher. By applying the Copula-based stochastic bias correction for the predicted (1,1) cases, it is found that the monthly mean errors, RMSE and the quantile RMSE in the WRF precipitation time series can be successfully reduced. It is also found that the correction for the predictied (1,1) cases has close performance compared to the correction for the actual (1,1) cases and only the quantile RMSEs are slightly higher.

The precipitation cases identification model is configured with a first-order HMM, the emissions are only set to two levels, i.e. dry day and wet day. It can be further extended by increasing the emission levels, e.g. to four levels: dry ($< 0.1\,mm$), lower intensity ($0.1\,mm - 5\,mm$), moderate intensity ($5\,mm - 50\,mm$) and extremes ($> 50\,mm$).

AIasseur et al. (2004) used 42 emission levels in his study to simulate the rain events time series with a Markov model. The method can also be extened by increasing the order of the hidden Markov model. Which may result a higher prediction accuracy for the precipitation cases.

The precipitation cases identification method introduced here is an extension of the Copula-based stochastic bias correction. It allows the bias correction also for the climate projections. Furthermore, the method predicts not only the (1,1) cases for the precipitation prejections but also the other three cases, i.e. (0,0) cases, (0,1) cases and (1,0) cases. Therefore, it potentially provides the possibility to correct also the (0,1) and (1,0) errors separately, which may lead to a better correction of the biases.

Chapter 9

Conclusions

In this study, a Copula-based stochastic bias correction technique for RCM-output is introduced. Different to triditional transfer function-based statistical corrections, The strategy of this method is the identification and description of underlying dependence structures between RCM and observed precipitation and its application for bias correction. A bivariate Copula model which consists of two marginal distributions and a Copula function, forms the basis of this approach. The marginal distributions describe the statistical properties of the variates (here: RCM and observed precipitation) and the Copula function captures the dependence structure between them. The WRF precipitation is then corrected based on the conditional distribution which is derived from the estimated Copula model. It is important to note that the proposed method is only applied for the positive pairs of RCM and observed precipitations, i.e. the (1,1) cases. In order to generate a complete bias corrected time series of WRF output, the events that are not covered by the (1,1) case are left unchanged (see Sect. 4.6).

The advantages of this approach are: 1) It is able to capture the non-linear dependencies between variables including a reliable description of the dependence structure in the tails of the joint distribution. This is not possible e.g. by using a Gaussian approach or methods based on the Pearson's correlation coefficient; 2) The univariate marginal distributions can be modeled independently from the dependence function, i.e. the Copula function. This provides more flexibility to construct a correction model by combining different marginal distributions and Copula functions, as many parametric univariate distribution and theoretical Copulas are available; 3) It provides the possibility to access all the possible outcomes of the corrected value and additionally gives the

information of a PDF for each corrected time step. 4) It has no limitations like Gaussian assumption or full dependence assumption which is hold by the quantile mapping correction.

The Copula-based correction is applied to correct a 30-year dynamical downscaled preipitation field (WRF-ERA40). As observation, the REGNIE data from the German Weather Service is used. The 30-year time series is split into a calibration (19711985) and a validation (19862000) period of equal length. For the application of bias correction, both the overall mode approach and the seasonal mode approach are used to investigate the intra-annual variability. In the overall mode approach, for each grid cell an overall Copula is estimated. While in the seasonal mode approach, the Copula models are fitted for every seasons (MAM, JJA, SON, DJF) separately in each grid cell. When estimating the Copula models, five theoretical distributions (Generalized Pareto distribution; Gamma distribution; Exponential distribution; Weibull distribution and Normal distribution) and four different types of Copulas (Clayton, Frank, Gumbel and Gaussian) are investigated.

The results of marginal fitting indicate discrepancies between the fitted marginal distributions of REGNIE and WRF-EAR40 data. The estimated marginal distributions for WRF show distinct spatial (strongly related to the orography of the domain) and seasonal patterns (clear differences between summer and winter, similar patterns for spring and fall season). The distributions are more scattered for the REGNIE data. The fitted Copula families imply that the dependence structure between REGNIE and WRF precipitation vary both in space and time (seasonally). The fact that different dependence structures exist for the different seasons indicates that the method corrects for different dominating precipitation types, i.e. convective and stratiform precipitation.

The assumption of this approach is that the dependence structure between observed and modelled precipitation is stationary over the period of interest. The corrected WRF precipitations are analyzed for both the overall mode approach and the seasonal mode approach. The Copula-based stochastic correction provides a full ensemble of corrected WRF precipitation for each time step through the Monte Carlo simulations. For the investigation of the spatial performance, the mean value of the Monte Carlo simulated realizations is applied after the comparison with other two statistical values, i.e. the median and mode value. Results show that the proposed approach in the both overall mode approach and the seasonal mode approach successfully corrected the errors in RCM derived precipitation. The seasonal mode based correction are found to be more efficient than the overall mode correction. It is also found that the correction method

in seasonal mode performs better for overestimations rather than for underestimations. By investigating the spatial coherence, the proposed method is found to be able to preserve the spatial structure of the WRF model output. This is due to the reason that the Copula-based approach is conditioned on the WRF simulation. The method is adjusting the value of the WRF precipitation according to the fitted Copula model. Even though the Copula models are estimated for each grid cell, the spatial coherence is captured by the Copula model as both the Copula families as well as the marginal distributions are also spatially clustered.

Finally the Copula-based statistical bias correction is compared to the quantile mapping method. It is found that the Copula-based method in seasonal mode has an improved performance in reducing the RMSE and quantile RMSE. It is also found that the Copula-based method allows for a better correction with respect to the percentage of the time steps that are closer to the observations after the correction. The Copula-based method is able to provide a stable correction efficiency over the entire domain, even if the rank correlations between the RCM- and observed precipitation are low.

Finally, the Copula-based stochastic bias correction is extended to be applicable for climate projections by combining with a precipitation cases identification model. This model is based on a first-order HMM with two emission levels and four Markov states which captures the four precipitation cases, i.e. (0,0) cases, (0,1) cases, (1,0) cases and (1,1) cases. It predicts the precipitation cases for the precipitation projections, where no observations are available, via the Viterbi algorithm. So far, this study focused on the introduction of the method, not a final application for the entire domain. To demonstrate this approach, it is applied for selected four pixels and the results are analyzed. Firstly, the predicted precipitation cases are compared to the actual precipitation cases. It is found that the method is able to predict the precipitation cases with a high overall accuracy (around 70 %). The Copula-based correction is then applied for the predicted (1,1) cases. Results show that the monthly mean errors, RMSE and the quantile RMSE in the WRF precipitation time series for the selected pixel can be successfully reduced. It is also found that the correction for the predictied (1,1) cases has close performance compared to the correction for the actual (1,1) cases and only the quantile RMSEs are slightly higher. Currently, this precipitation model is configured with a first-order HMM and only emission levels of two. It can be further extended by increasing the emission levels and the order of the HMM. This may result a higher prediction accuracy for the precipitation cases. Furthermore, this method potentially provides the possibility to correct also the (0,1) and (1,0) errors separately, since it predicts not only the (1,1)

cases for the precipitation prejections but also the other three cases, i.e. (0,0) cases, (0,1) cases and (1,0) cases.

Based on the analysis carried out in this study, the results discussed in the previous chapters, and the scope of this research, the following recommendations for future research are suggested:

It is well known that the daily precipitation distributions are typically heavily skewed towards low-intensity values. When fitting a set of theoretical distributions, the distribution parameters will be dictated by the most frequently occurring values, but may then not accurately represent the extremes. Using the combination of two or more distributions to can better characterize the precipitation property and therefore improve the correction.

The Copula-based method provides the information of the full PDF for each individual time step. For practical reasons, only single corrected values are required for each time step. In this study, only the mean, median and the mode value of the Monte Carlo simulation are investigated. It is worthful to investigate also other statistical values or make use of the complete conditional Copula CDF.

Appendix A

The hidden Markov model

A Markov chain (Rabiner, 1989) can be described as follows: Assume one has a set of states, $S = \{s_1, s_2, \cdots, s_n\}$. The process starts in one of these states and moves successively from one state to another. Each move is called a step. If the chain is currently in state s_i, then it moves to state s_j at the next step with a probability denoted by p_{ij}. In a first-order Markov process, the probability to have a particular state s_j at time t depends solely on the condition of the previous state s_i at time $(t-1)$, i.e.:

$$p_{ij} = Pr(q_t = s_j | q_{t-1} = s_i) \tag{A.1}$$

The probabilities p_{ij} are called transition probabilities. The process can remain in the state it is in, and this occurs with probability p_{ii}. The matrix which contains all of the transition probabilities is called the transition matrix.

$$T = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix} \tag{A.2}$$

The sum of each column or row in the transition matrix T is equal to 1. The state diagram for a Markov model with two states is shown in Fig. A.1. The transition

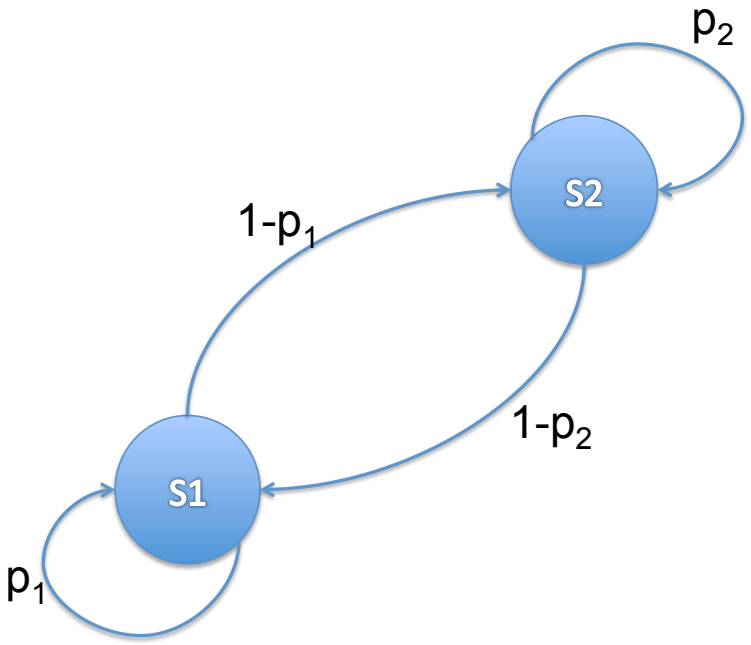

FIGURE A.1: State diagram for the Markov model with two states S1 and S2.

matrix for this model is the expressed as $\begin{bmatrix} p_1 & 1-p_1 \\ 1-p_2 & p_2 \end{bmatrix}$. In this model, the Markov process has a probability of p_1 to stay in state S1 and has a probability of $1-p_1$ to move from S1 to S2. The probability of stay in S2 is p_2 and the probability is $1-p_2$ if it moves from S2 to S1. In a Markov process, the initial state probabilities can be expressed as $\lambda_i = Pr(q_1 = s_i)$.

A hidden Markov model is a model in which the state sequence (which is a Markov chain) is not directly observable. It generates an emission from each states. The emission is a probabilistic function of an underlying Markov state sequence. At each time step, an emission is drawn from the (discrete or continuous) probability distribution associated with the current state. The emission probability for the observable a from state s_i is $b_i(a) = Pr(E_t = a|q_t = s_i)$ (where E_t is the emission at time t). The hidden Markov model is a doubly embedded stochastic process with an underlying stochastic process that is not observable (it is hidden), but can only be observed through another set of stochastic processes that produce the sequence of emissions. The set of parameters $H = \{p_{ij}, b_i(a), \lambda_i\}$ gives a full probabilistic description of the aforementioned HMM model.

Appendix B

The Viterbi algorithm

Given a sequence of emissions $E = \{e_1, e_2, \cdots, e_T\}$, and an HMM $H = \{p_{ij}, b_i(a), \lambda_i\}$. There are several paths through the hidden states that lead to the given sequence, but they do not have the same probability. The maximum probability state path $Q = \{q_1, q_2, \cdots, q_T\}$ can then be estimated recursively using the Viterbi algorithm, which is a dynamical programming algorithm. The estimated maximum probability state path is also called Viterbi path.

Let $v_i(t)$ be the probability of the most likely path ending in state s_i at time t, i.e.,

$$v_i(t) = \max_{q_1, q_2, \cdots, q_{t-1}} Pr(q_1 q_2 \cdots q_{t-1}, q_t = s_i, e_1 e_2 \cdots e_t | H),$$

and let λ_i be the initial probabilities of the states s_i at time $t = 1$. Then $v_j(t)$ can be calculated recursively using

$$v_j(t) = \max_{1 \le i \le N} [v_i(t-1) p_{ij}] b_j(e_t)$$

together with initialization

$$v_i(1) = \lambda_i b_i(e_1) \qquad 1 \le i \le N$$

and termination

$$P^* = \max_{1 \le i \le N} [v_i(T)].$$

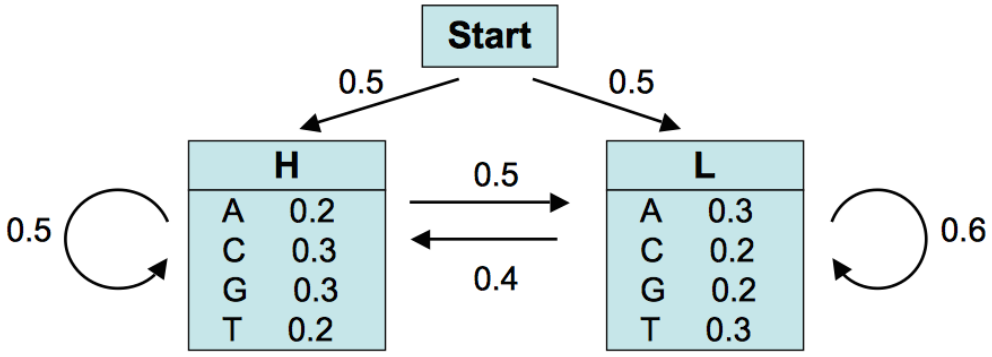

FIGURE B.1: Illustration of a simple hidden Markov model with two states and four emission levels for each state.

The method chooses the highest probability endpoint, and then backtrack from there to find the highest probability path.

To illustrate this algorithm, a small example is shown in the following (Borodovsky and Ekisheva, 2006): Let's consider a simple HMM which is described in Fig. B.1. This model is composed of 2 states, H (high GC content) and L (low GC content). We can for example consider that state H characterizes coding DNA while L characterizes non-coding DNA. For each state there are four different emission levels, i.e. A, C, G and T, with different emission probabilities.

Assume a sequance as $S = \mathbf{GGCA}$, there are several paths through the hidden states (H and L) that lead to this sequence, e.g. $P = LLHH$. The probability of the HMM to produce sequence S through the path $P = LLHH$ can be calculated as:

$$\begin{aligned} v &= \lambda_L * b_L(G) * p_{LL} * b_L(G) * p_{LH} * b_H(C) * p_{HH} * b_H(A) \\ &= 0.5 * 0.2 * 0.6 * 0.2 * 0.4 * 0.3 * 0.5 * 0.2 \\ &= 0.000144 \end{aligned}$$

However, this path may not be the most probable path. To search for the most probable path for this sequence $S = \mathbf{GGCA}$, the viterbi algorithm is depicted step by step in following, i.e. backtracking from the endpoint to the start point.

$P^* = \max(v_H(4), v_L(4))$

$v_H(4) = b_H(A)\max(v_H(3) * p_{HH}, v_L(3) * p_{LH}) = 0.2 * \max(v_H(3) * 0.5, v_L(3) * 0.4)$
$v_L(4) = b_L(A)\max(v_H(3) * p_{HL}, v_L(3) * p_{LL}) = 0.3 * \max(v_H(3) * 0.5, v_L(3) * 0.6)$

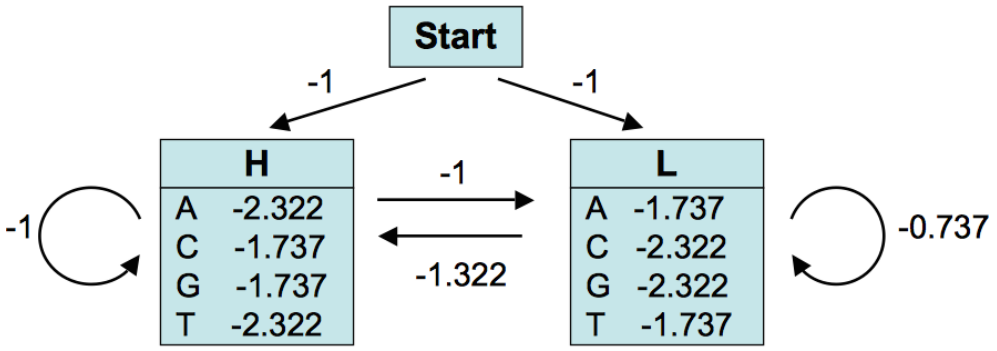

FIGURE B.2: Illustration of a simple hidden Markov model with two states and four emission levels for each state. The probabilities are transformed by a log operation ($\log_2$)

$$v_H(3) = b_H(C)\max(v_H(2) * p_{HH}, v_L(2) * p_{LH}) = 0.3 * \max(v_H(2) * 0.5, v_L(2) * 0.4)$$
$$v_L(3) = b_L(C)\max(v_H(2) * p_{HL}, v_L(2) * p_{LL}) = 0.2 * \max(v_H(2) * 0.5, v_L(2) * 0.6)$$

$$v_H(2) = b_H(G)\max(v_H(1) * p_{HH}, v_L(1) * p_{LH}) = 0.3 * \max(v_H(1) * 0.5, v_L(1) * 0.4)$$
$$v_L(2) = b_L(G)\max(v_H(1) * p_{HL}, v_L(1) * p_{LL}) = 0.2 * \max(v_H(1) * 0.5, v_L(1) * 0.6)$$

$$v_H(1) = b_H(G) * \lambda_H = 0.3 * 0.5$$
$$v_L(1) = b_L(G) * \lambda_L = 0.2 * 0.5$$

where v is the probability of the most likely path ending in a certain state at a certain step, e.g. $v_H(4)$ is the probability of the most likely path ending in state H at the $4th$ step. p is the transition probability that the Markov chain moves from one state to another state, e.g. p_{LH} is the transition probability that the Markov chain moves from state L to state H. For the calculations, it is convenient to use the log of the probabilities (rather than the probabilities themselves). This allows to compute *sums* instead of *products*, which is more efficient and accurate. In this example, $\log_2$ is used and the probalibites in Fig. B.1 are then transformed to that in Fig. B.2. Therefore the probabilities of the path that end in a certain state at each time step are computed as follows. It is noted that the probabilites are obtained after a log transformation ($\log_2(v)$).

$$\log_2(v_H(1)) = -1 - 1.737 = -2.737$$
$$\log_2(v_L(1)) = -1 - 2.322 = -3.322$$

$$\log_2(v_H(2)) = -1.737 + \max(-2.737 - 1, -3.322 - 1.322) = -5.474$$
$$\log_2(v_L(2)) = -2.322 + \max(-2.737 - 1, -3.322 - 0.737) = -6.059$$

$\log_2(v_H(3)) = -1.737 + \max(-5.474 - 1, -6.059 - 1.322) = -8.211$
$\log_2(v_L(3)) = -2.322 + \max(-5.474 - 1, -6.059 - 0.737) = -8.796$

$\log_2(v_H(4)) = -2.322 + \max(-8.211 - 1, -8.796 - 1.322) = -11.533$
$\log_2(v_L(4)) = -1.737 + \max(-8.211 - 1, -8.796 - 0.737) = -10.948$

$\log_2(P^*) = \max(-11.533, -10.948) = -10.948$
$P^* = 2^{-10.948} = 5.06E - 4$

Finally, the probability of the Viterbi path is calculated, i.e. $P^* = 5.06E - 4$, which is given by the path ending in state L at the $4th$ step. By backtracking from the endpoint to the start point (highlighted as red during the calculation), the path which corresponds to the highest probability is found, i.e. **HHHL**.

Appendix C

Illustration of the precipitation cases identification

For illustration of the precipitation cases identification approach, a step-by-step example is shown here by taking the data from the pixel 1 in Fig. 2.1. Using the same strategy as before, the 30-year time series is split into two 15-year periods. The first 15-year (1971–1985) is set for the calibration period and the last 15-year (1986–2000) is set for the validation period. We assume the validation time to be the future. A threshold of rainfall amount of 0.1 mm per day was used to identify a wet day with respect to the usual precision of rain gauges (Dieterichs, 1956; Moon et al., 1994). The data are shown in Table C.1, where X indicates the REGNIE data and Y represents the WRF data. The complete time series in calibration period includes 5479 days and only the first 15 days are shown in the Table for the sake of intuition.

As described in Sect. 8.2, the hidden Markov model is configured with two emission levels (i.e. d/w) and four Markov states (i.e. S1 indicates the (0,0) case, S2 implies the (0,1) case, S3 represents the (1,0) case and S4 stands for the (1,1) case). For the selected dataset the emissions and the states can be simply derived in Table C.2.

TABLE C.1: Training data set for the precipitation cases identification model (mm/day)

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	...
X_i	0	0	0	0	0	0	0	0	0	0	0	0	1.5	4.6	0.6	...
Y_i	0.83	1.38	12.26	0.97	0	0	0	0	0	0	0	0	0	22.38	10.83	...

TABLE C.2: The emissions and the states that derived from the training data set

t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	...
$E(t)$	w	w	w	w	d	d	d	d	d	d	d	d	d	w	w	...
$S(t)$	S2	S2	S2	S2	S1	S1	S1	S1	S1	S1	S1	S1	S3	S4	S4	...

TABLE C.3: The emissions in the validation period

t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	...
$E(t)$	w	w	w	w	w	w	w	w	w	w	w	w	d	d	w	...

With the emissions and the states in the calibration period, the first-order HMM is then estimated which consists of the initial state probabilities $\lambda_i = \{S1 : 0.2774, \quad S2 : 0.2361, \quad S3 : 0.0807, \quad S4 : 0.4058\}$, the transition matrix

$$T = \begin{bmatrix} 0.60 & 0.15 & 0.11 & 0.14 \\ 0.27 & 0.40 & 0.04 & 0.29 \\ 0.23 & 0.11 & 0.24 & 0.42 \\ 0.07 & 0.22 & 0.06 & 0.65 \end{bmatrix},$$

and the emission matrix

$$EM = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

The transition matrix contains the probabilities that a state jump to another state or stay in the same state. E. g. in this model, it has a probability of 0.6 that the state $S1$ stays as the same and has a probability of 0.15 that the state $S1$ jump to $S2$. The probability that the state $S2$ move to $S1$ is 0.27. The emission matrix contains the probabilites that a state generates a certain emission. The column indicates the emission levels, i.e. d/w, and the row implies the states. E. g in this model, state $S1$ has 100 % chance to generate a dry day (d) and has no chance to generate a wet day (w). State $S2$ has 0 chance to generate a dry day (d) and has 100 % chance to generate a wet day (w). It is noted that the emissions are only refer to the WRF projections.

TABLE C.4: Predicted states in the validation period

t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	...
$S(t)$	S4	S4	S4	S4	S4	S4	S4	S4	S4	S4	S4	S2	S1	S1	S4	...

With the estimated HMM and the future WRF precipitation (i.e. the WRF time series in validation period in this application), the precipitation cases can then be predicted. The emissions in the validation time are shown in Table C.3 and the predicted states are shown in the Table C.4.

Bibliography

AIasseur, C., Husson, L., and Perez-Fontan, F.: Simulation of rain events time series with Markov model, 2004 IEEE 15th International Symposium on Personal, Indoor and Mobile Radio Communications (IEEE Cat. No.04TH8754), 4, 2801–2805, doi: 10.1109/PIMRC.2004.1368831, 2004.

Arnell, N. W.: Climate change scenarios from a regional climate model: Estimating change in runoff in southern Africa, Journal of Geophysical Research, 108, 4519, doi: 10.1029/2002JD002782, 2003.

Bachner, S., Kapala, A., and Simmer, C.: Evaluation of daily precipitation characteristics in the CLM and their sensitivity to parameterizations, Meteorologische Zeitschrift, 17, 407–419, doi:10.1127/0941-2948/2008/0300, 2008.

Baigorria, G. A., Jones, J. W., Shin, D.-W., Mishra, A., and OBrien, J. J.: Assessing uncertainties in crop model simulations using daily bias-corrected Regional Circulation Model outputs, Climate Research, 34, 211–222, doi:10.3354/cr00703, 2007.

Bárdossy, A.: Copula-based geostatistical models for groundwater quality parameters, Water Resources Research, 42, W11 416, doi:10.1029/2005WR004754, 2006.

Bárdossy, A. and Pegram, G.: Multiscale spatial recorrelation of RCM precipitation to produce unbiased climate change scenarios over large areas and small, Water Resources Research, 48, 1–13, doi:10.1029/2011WR011524, 2012.

Bárdossy, A. and Pegram, G. G. S.: Copula based multisite model for daily precipitation simulation., Hydrology and Earth System Sciences, 13, 2299–2314, doi: 10.5194/hess-13-2299-2009, 2009.

Berg, P., Wagner, S., Kunstmann, H., and Schädler, G.: High resolution regional climate model simulations for Germany: part Ivalidation, Climate Dynamics, 40, 401–414, doi:10.1007/s00382-012-1508-8, 2013.

Boé, J., Terray, L., Habets, F., and Martin, E.: Statistical and dynamical downscaling of the Seine basin climate for hydro-meteorological studies, International Journal of Climatology, 27, 1643–1655, doi:10.1002/joc.1602, 2007.

Borodovsky, M. and Ekisheva, S.: Problems and solutions in biological sequence analysis, Cambridge University Press, Cambridge, U.K., 2006.

Chen, F. and Dudhia, J.: Coupling an Advanced Land SurfaceHydrology Model with the Penn StateNCAR MM5 Modeling System. Part I: Model Implementation and Sensitivity, Monthly Weather Review, 129, 569–585, doi:10.1175/1520-0493(2001)129⟨0587:CAALSH⟩2.0.CO;2, 2001.

Deheuvels, P.: La fonction de dépendance empirique et ses propriétés: Un test non paramétrique dindépendance, Académie Royale de Belgique, Bulletin de la Classe des Sciences, 65, 274–292, 1979.

Dettinger, M. D., Cayan, D. R., Meyer, M. K., and Jeton, A. E.: Simulated hydrologic responses to climate variations and change in the Merced, Carson, and American River basins, Sierra Nevada, California, 1900-2099 *, Climatic Change, 62, 283–317, doi:10.1023/B:CLIM.0000013683.13346.4f, 2004.

Diaz-Nieto, J. and Wilby, R. L.: A comparison of statistical downscaling and climate change factor methods: Impacts on low flows in the River Thames, United Kingdom, Climatic Change, 69, 245–268, doi:10.1007/s10584-005-1157-6, 2005.

Dieterichs, H.: Frequency of dry and wet spells in san salvador, Geofisica pura e applicata, 33, 267–272, 1956.

Dobler, A. and Ahrens, B.: Precipitation by a regional climate model and bias correction in Europe and South Asia, Meteorologische Zeitschrift, 17, 499–509, doi:10.1127/0941-2948/2008/0306, 2008.

Dosio, A. and Paruolo, P.: Bias correction of the ENSEMBLES high-resolution climate change projections for use by impact models: Evaluation on the present climate, Journal of Geophysical Research, 116, D16 106, doi:10.1029/2011JD015934, 2011.

Dudhia, J.: Numerical Study of Convection Observed during the Winter Monsoon Experiment Using a Mesoscale Two-Dimensional Model, doi:10.1175/1520-0469(1989)046⟨3077:NSOCOD⟩2.0.CO;2, 1989.

Dupuis, D. J.: Using copulas in hydrology: Benefits, cautions, and issues, Journal of Hydrologic Engineering, 12, 381–393, doi:10.1061/(ASCE)1084-0699(2007)12:4(381), 2007.

DWD: Abteilung Hydrometeorologie: REGNIE (REGionalisierte NIEderschlage): Verfahrensbeschreibung und Nutzeranleitung, interner Bericht im DWD, Offenbach, 2011.

Embrechts, P., McNeil, A., and Straumann, D.: Correlation and dependence in risk management: properties and pitfalls, in: Risk Management: Value at Risk and Beyond, edited by Dempster, M. A. H., pp. 176–223, Cambridge University Press, Cambridge, U.K., 2002.

Embrechts, P., Lindskog, F., and McNeil, A.: Modelling dependence with copulas and applications to risk management, in: Handbook of Heavy Tailed Distributions in Finance, edited by Rachev, S. T., chap. 8, pp. 329–384, Elsevier/North Holland, 2003.

Fisher, N. I. and Sen, P. K.: The Collected Works of Wassily Hoeffding, New York : Springer-Verlag, New York, 1994.

Fowler, H. J. and Ekstr, M.: Multimodel ensemble estimates of climate change impacts on UK seasonal precipitation extremes, International Journal of Climatology, 29, 385–416, doi:10.1002/joc.1827, 2009.

Fowler, H. J., Blenkinsop, S., and Tebaldi, C.: Linking climate change modelling to impacts studies : recent advances in downscaling techniques for hydrological modelling, International Journal of Climatology, 27, 1547–1578, doi:10.1002/joc.1556, 2007.

Fréchet, M.: Sur les tableaux de corrélation dont les marges sont données, Annales de l'Université de Lyon. Section A: Sciences mathématiques et astronomie, pp. 53–77, 1951.

Fredricks, G. A. and Nelsen, R. B.: On the relationship between Spearman's rho and Kendall's tau for pairs of continuous random variables, Journal of Statistical Planning and Inference, 137, 2143–2150, doi:10.1016/j.jspi.2006.06.045, 2007.

Gao, H., Wood, E. F., Drusch, M., and McCabe, M. F.: Copula-derived observation operators for assimilating TMI and AMSR-E retrieved soil moisture into land surface models, Journal of Hydrometeorology, 8, 413–429, doi:10.1175/JHM570.1, 2007.

Genest, C. and Favre, A.-C.: Everything you always wanted to know about copula modeling but were afraid to ask, Journal of Hydrologic Engineering, 12, 347–368, doi:10.1061/(ASCE)1084-0699(2007)12:4(347), 2007.

Genest, C., Rémillard, B., and Beaudoin, D.: Goodness-of-fit tests for copulas: A review and a power study, Insurance: Mathematics and Economics, 44, 199–213, doi:10.1016/j.insmatheco.2007.10.005, 2009.

Ghosh, S. and Mujumdar, P. P.: Climate change impact assessment: Uncertainty modeling with imprecise probability, Journal of Geophysical Research, 114, D18 113, doi:10.1029/2008JD011648, 2009.

Grégoire, V., Genest, C., and Gendron, M.: Using copulas to model price dependence in energy markets, Energy Risk, 5, 62–68, 2008.

Gudmundsson, L., Bremnes, J. B., Haugen, J. E., and Engen-Skaugen, T.: Technical Note: Downscaling RCM precipitation to the station scale using statistical transformations a comparison of methods, Hydrology and Earth System Sciences, 16, 3383–3390, doi:10.5194/hess-16-3383-2012, 2012.

Gutjahr, O. and Heinemann, G.: Comparing precipitation bias correction methods for high-resolution regional climate simulations using COSMO-CLM, Theoretical and Applied Climatology, 114, 511–529, doi:10.1007/s00704-013-0834-z, 2013.

Hagemann, S., Chen, C., Haerter, J. O., Heinke, J., Gerten, D., and Piani, C.: Impact of a Statistical Bias Correction on the Projected Hydrological Changes Obtained from Three GCMs and Two Hydrology Models, Journal of Hydrometeorology, 12, 556–578, doi:10.1175/2011JHM1336.1, 2011.

Hay, L. E., Wilby, R. L., and Leavesley, G. H.: A Comparison of Delta Change and Downscaled GCM Scenarios for Three Mountainous Basins in the United States, Journal of the American Water Resources Association, 36, 387–397, doi:10.1111/j.1752-1688.2000.tb04276.x, 2000.

Hertig, E. and Jacobeit, J.: A novel approach to statistical downscaling considering nonstationarities: Application to daily precipitation in the mediterranean area, Journal of Geophysical Research: Atmospheres, 118, 520–533, doi:10.1002/jgrd.50112, 2013.

Hohenegger, C., Brockhaus, P., and Schär, C.: Towards climate simulations at cloud-resolving scales, Meteorologische Zeitschrift, 17, 383–394, doi:10.1127/0941-2948/2008/0303, 2008.

Hong, S.-Y. and Lim, J.-O. J.: The WRF single-moment 6-class microphysics scheme (WSM6), URL http://www.mmm.ucar.edu/wrf/users/docs/WSM6-hong_and_lim_JKMS.pdf$\delimiter"026E30F$nhttp://search.koreanstudies.net/journal/thesis_name.asp?tname=kiss2002&key=2525908, 2006.

Hong, S.-Y., Dudhia, J., and Chen, S.-H.: A Revised Approach to Ice Microphysical Processes for the Bulk Parameterization of Clouds and Precipitation, Monthly Weather Review, 132, 103–120, doi:10.1175/1520-0493(2004)132⟨0103:ARATIM⟩2.0.CO;2, 2004.

Hong, S.-Y., Noh, Y., and Dudhia, J.: A New Vertical Diffusion Package with an Explicit Treatment of Entrainment Processes, Monthly Weather Review, 134, 2318–2341, doi:10.1175/MWR3199.1, 2006.

Horton, P., Schaefli, B., Mezghani, A., Hingray, B., and Musy, A.: Assessment of climate-change impacts on alpine discharge regimes with climate model uncertainty, Hydrological Processes, 20, 2091–2109, doi:10.1002/hyp.6197, 2006.

Ines, A. V. M. and Hansen, J. W.: Bias correction of daily GCM rainfall for crop simulation studies, Agricultural and Forest Meteorology, 138, 44–53, doi:10.1016/j.agrformet.2006.03.009, 2006.

Joe, H.: Multivariate Models and Dependence Concepts, Chapman and Hall/CRC, London, 1997.

Kain, J. S.: The KainFritsch Convective Parameterization: An Update, Journal of Applied Meteorology, 43, 170–181, 2004.

Kunstmann, H., Schneider, K., Forkel, R., and Knoche, R.: Impact analysis of climate change for an Alpine catchment using high resolution dynamic downscaling of ECHAM4 time slices, Hydrology and Earth System Sciences, 8, 1031–1045, doi:10.5194/hess-8-1031-2004, 2004.

Lafon, T., Dadson, S., Buys, G., and Prudhomme, C.: Bias correction of daily precipitation simulated by a regional climate model: a comparison of methods, International Journal of Climatology, 33, 1367–1381, doi:10.1002/joc.3518, 2013.

Laux, P., Vogl, S., Qiu, W., Knoche, H. R., and Kunstmann, H.: Copula-based statistical refinement of precipitation in RCM simulations over complex terrain, Hydrology and Earth System Sciences, 15, 2401–2419, doi:10.5194/hess-15-2401-2011, 2011.

Leander, R. and Buishand, T. A.: Resampling of regional climate model output for the simulation of extreme river flows, Journal of Hydrology, 332, 487–496, doi:10.1016/j.jhydrol.2006.08.006, 2007.

Leander, R., Buishand, T. A., van den Hurk, B. J., and de Wit, M. J.: Estimated changes in flood quantiles of the river Meuse from resampling of regional climate model output, Journal of Hydrology, 351, 331–343, doi:10.1016/j.jhydrol.2007.12.020, 2008.

Lenderink, G., Buishand, A., and van Deursen, W.: Estimates of future discharges of the river Rhine using two scenario methodologies: direct versus delta approach, Hydrology and Earth System Sciences, 11, 1145–1159, doi:10.5194/hess-11-1145-2007, 2007.

Li, H., Sheffield, J., and Wood, E. F.: Bias correction of monthly precipitation and temperature fields from Intergovernmental Panel on Climate Change AR4 models using equidistant quantile matching, Journal of Geophysical Research: Atmospheres, 115, D10 101, doi:10.1029/2009JD012882, 2010.

Mao, G., Vogl, S., Laux, P., Wagner, S., and Kunstmann, H.: Stochastic bias correction of dynamically downscaled precipitation fields for Germany through Copula-based integration of gridded observation data, Hydrology and Earth System Sciences, 19, 1787–1806, doi:10.5194/hess-19-1787-2015, 2015.

Maraun, D., Wetterhall, F., Ireson, A. M., Chandler, R. E., Kendon, E. J., Widmann, M., Brienen, S., Rust, H. W., Sauter, T., Themeßl, M., Venema, V. K. C., Chun, K. P., Goodess, C. M., Jones, R. G., Onof, C., Vrac, M., and Thiele-Eich, I.: Precipitation downscaling under climate change: recent developments to bridge the gap between dynamical models and the end user, Reviews of Geophysics, 48, 1–34, doi:10.1029/2009RG000314, 2010.

Massey, F. J.: The Kolmogorov-Smirnov test for goodness of fit, Journal of the American statistical Association, 46, 68–78, doi:10.2307/2280095, 1951.

Melchiori, M. R.: Which Archimedean Copula is the right one?, YieldCurve, pp. 1–20, doi:10.2139/ssrn.1123135, 2003.

Mlawer, E. J., Taubman, S. J., Brown, P. D., Iacono, M. J., and Clough, S. A.: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave, Journal of Geophysical Research, 102, 16 663–16 682, doi:10.1029/97JD00237, 1997.

Moon, S.-E., Ryoo, S.-B., and Kwon, J.-G.: A Markov chain model for daily precipitation occurrence in South Korea, International Journal of Climatology, 14, 1009–1016, doi: 10.1002/joc.3370140906, 1994.

Myung, I. J.: Tutorial on maximum likelihood estimation, Journal of Mathematical Psychology, 47, 90–100, doi:10.1016/S0022-2496(02)00028-7, 2003.

Nelsen, R. B.: On measures of association of positive dependence, Statistics & Probability Letters, 14, 269–274, 1992.

Nelsen, R. B.: An Introduction to Copulas, Springer, New York, 1999.

Oakes, D.: On the preservation of copula structure under truncation, Canadian Journal of Statistics, 33, 465–468, 2005.

Ott, I., Duethmann, D., Liebert, J., Berg, P., Feldmann, H., Ihringer, J., Kunstmann, H., Merz, B., Schaedler, G., and Wagner, S.: High resolution climate change impact analysis on medium sized river catchments in Germany: An ensemble assessment, Journal of Hydrometeorology, 14, 1175–1193, doi:10.1175/JHM-D-12-091.1, 2013.

Panofsky, H. A. and Brier, G. W.: Some applications of statistics to meteorology, University Park : Penn. State University, College of Earth and Mineral Sciences, 1968.

Paxian, a., Hertig, E., Seubert, S., Vogt, G., Jacobeit, J., and Paeth, H.: Present-day and future mediterranean precipitation extremes assessed by different statistical approaches, Climate Dynamics, 44, 845–860, doi:10.1007/s00382-014-2428-6, 2014.

Piani, C., Haerter, J. O., and Coppola, E.: Statistical bias correction for daily precipitation in regional climate models over Europe, Theoretical and Applied Climatology, 99, 187–192, doi:10.1007/s00704-009-0134-9, 2010a.

Piani, C., Weedon, G. P., Best, M., Gomes, S. M., Viterbo, P., Hagemann, S., and Haerter, J. O.: Statistical bias correction of global simulated daily precipitation and temperature for the application of hydrological models, Journal of Hydrology, 395, 199–215, doi:10.1016/j.jhydrol.2010.10.024, 2010b.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P.: Numerical Recipes in FORTRAN; The Art of Scientific Computing, Cambridge University Press, Cambridge, U.K., 1993.

Rabiner, L. R.: A tutorial on hidden Markov models and selected applications in speech recognition, Proceedings of the IEEE, 77, 257–286, doi:10.1109/5.18626, 1989.

Roeckner, E., Brokopf, R., Esch, M., Giorgetta, M. a., Hagemann, S., Kornblueh, L., Manzini, E., Schlese, U., and Schulzweida, U.: Sensitivity of simulated climate to horizontal and vertical resolution in the ECHAM5 atmosphere model, Journal of Climate, 19, 3771–3791, doi:10.1175/JCLI3824.1, 2006.

Rojas, R., Feyen, L., Dosio, A., and Bavera, D.: Improving pan-European hydrological simulation of extreme events through statistical bias correction of RCM-driven climate simulations, Hydrology and Earth System Sciences, 15, 2599–2620, doi:10.5194/hess-15-2599-2011, 2011.

Salvadori, G. and Michele, C. D.: On the use of copulas in hydrology: theory and practice, Journal of Hydrologic Engineering, 12, 369–380, doi:10.1061/(ASCE)1084-0699(2007)12:4(369), 2007.

Salvadori, G., Michele, C. D., Kottegoda, N. T., and Rosso, R.: Extremes in Nature: An Approach Using Copulas, Springer, New York, 2007.

Schmidli, J., Frei, C., and Vidale, P. L.: Downscaling from GCM precipitation: a benchmark for dynamical and statistical downscaling methods, International Journal of Climatology, 26, 679–689, doi:10.1002/joc.1287, 2006.

Schmidt, T.: Coping with copulas, in: Copulas - From Theory to Application in Finance, pp. 3–34, Risk Books, London, 2007.

Schweizer, B. and Wolff, E. F.: On nonparametric measures of dependence for random variables, The annals of statistics, 1981.

Sennikovs, J. and Bethers, U.: Statistical downscaling method of regional climate model results for hydrological modelling, in: 18th World IMACS/MODSIM Congress, pp. 3962–3968, Cairns, Australia, 2009.

Serinaldi, F.: Copula-based mixed models for bivariate rainfall data: an empirical study in regression perspective, Stochastic Environmental Research and Risk Assessment, 23, 677–693, doi:10.1007/s00477-008-0249-z, 2008.

Shabalova, M. V., Deursen, W. P. A. V., and Buishand, T. A.: Assessing future discharge of the river Rhine using regional climate model integrations and a hydrological model, Climate Research, 23, 233–246, 2003.

Skamarock, W., Klemp, J., Dudhia, J., Gill, D., Barker, D., Duda, M., Huang, X., Wang, W., and Powers, J.: A description of the advanced research WRF version 3, Technical Report 459, NCAR, Boulder, Colorado, USA, 2008.

Sklar, A.: Fonctions de répartition à n dimensions et leurs marges, Publications de l'Institut Statistique de l'Université de Paris, 8, 229–231, 1959.

Smiatek, G., Kunstmann, H., Knoche, R., and Marx, A.: Precipitation and temperature statistics in high-resolution regional climate models: Evaluation for the European Alps, Journal of Geophysical Research, 114, D19 107, doi:10.1029/2008JD011353, 2009.

Stehman, S. V.: Selecting and interpreting measures of thematic classification accuracy, Remote Sensing of Environment, 62, 77–89, doi:10.1016/S0034-4257(97)00083-7, 1997.

Teng, J., Potter, N. J., Chiew, F. H. S., Zhang, L., Wang, B., Vaze, J., and Evans, J. P.: How does bias correction of regional climate model precipitation affect modelled runoff?, Hydrology and Earth System Sciences, 19, 711–728, doi:10.5194/hess-19-711-2015, 2015.

Terink, W., Hurkmans, R. T. W. L., Torfs, P. J. J. F., and Uijlenhoet, R.: Evaluation of a bias correction method applied to downscaled precipitation and temperature reanalysis data for the Rhine basin, Hydrology and Earth System Sciences, 14, 687–703, doi:10.5194/hess-14-687-2010, 2010.

Teutschbein, C. and Seibert, J.: Regional climate models for hydrological impact studies at the catchment scale: a review of recent modeling strategies, Geography Compass, 4, 834–860, doi:10.1111/j.1749-8198.2010.00357.x, 2010.

Teutschbein, C. and Seibert, J.: Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods, Journal of Hydrology, 456-457, 12–29, doi:10.1016/j.jhydrol.2012.05.052, 2012.

Themeßl, M. J., Gobiet, A., and Leuprecht, A.: Empirical-statistical downscaling and error correction of daily precipitation from regional climate models, International Journal of Climatology, 31, 1530–1544, doi:10.1002/joc.2168, 2010.

Themeßl, M. J., Gobiet, A., and Heinrich, G.: Empirical-statistical downscaling and error correction of regional climate models and its impact on the climate change signal, Climatic Change, 112, 449–468, doi:10.1007/s10584-011-0224-4, 2011.

Uppala, S., Kå llberg, P., Simmons, A., Andrae, U., da Costa Bechtold, V., Fiorino, M., Gibson, J., Haseler, J., Hernandez, A., Kelly, G., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R., Andersson, E., Arpe, K., Balmaseda, M., Beljaars, A., van de Berg, L., Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac,

M., Fisher, M., Fuentes, M., Hagemann, S., Hólm, E., Hoskins, B., Isaksen, L., Janssen, P., Jenne, R., McNally, A., Mahfouf, J.-F., Morcrette, J.-J., Rayner, N., Saunders, R., Simon, P., Sterl, A., Trenberth, K., Untch, A., Vasiljevic, D., Viterbo, P., and Woollen, J.: The ERA-40 re-analysis, Quart. J. R. Meteorol. Soc., 131, 2961–3012, doi:10.1256/qj.04.176, 2005.

van den Berg, M. J., Vandenberghe, S., De Baets, B., and Verhoest, N. E. C.: Copula-based downscaling of spatial rainfall: a proof of concept, Hydrology and Earth System Sciences, 15, 1445–1457, doi:10.5194/hess-15-1445-2011, 2011.

Venter, G. G.: Tails of copulas, Proceedings of the Casualty Actuarial Society, LXXXIX, 68–113, 2002.

Vlček, O. and Huth, R.: Is daily precipitation Gamma-distributed?. Adverse effects of an incorrect use of the Kolmogorov-Smirnov test, Atmospheric Research, 93, 759–766, doi:10.1016/j.atmosres.2009.03.005, 2009.

Vogl, S., Laux, P., Qiu, W., Mao, G., and Kunstmann, H.: Copula-based assimilation of radar and gauge information to derive bias-corrected precipitation fields, Hydrology and Earth System Sciences, 16, 2311–2328, doi:10.5194/hess-16-2311-2012, 2012.

Weakliem, D. L.: A Critique of the Bayesian Information Criterion for Model Selection, Sociological Methods & Research, 27, 359–397, doi:10.1177/0049124199027003002, 1999.

Wood, a. W., Leung, L. R., Sridhar, V., and Lettenmaier, D. P.: Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs, Climatic Change, 62, 189–216, doi:10.1023/B:CLIM.0000013685.99609.9e, 2004.

Yang, W., Andréasson, J., Phil Graham, L., Olsson, J., Rosberg, J., and Wetterhall, F.: Distribution-based scaling to improve usability of regional climate model projections for hydrological climate change impacts studies, Hydrology Research, 41, 211, doi: 10.2166/nh.2010.004, 2010.