

Master Thesis

Development of a statistical daily
precipitation model and its
application to precipitation records



submitted by

Philipp von Bomhard

to achieve the academic degree

– Master of Science –

1st Supervisor: Prof. Dr. Yaping Shao

2nd Supervisor: Prof. Dr. Joachim Krug

Köln, 08.09.2017

Abstract

In recent years, numerous record-breaking precipitation events have caused several deaths and high economic losses. However, current precipitation models do not adequately capture exceptionally high precipitation events. To address this problem, a new statistical daily precipitation model including several new aspects was developed. The model distinguishes between stratiform and convective precipitation, whereby the disaggregation of these two types of precipitation is based on SYNOP reports of 300 German weather stations. By combining a Weibull distribution with power law, a new probability distribution was derived and later implemented in the amount process. This four-parameter distribution improves the modelling of extreme precipitation amounts enormously and still delivers good results for small amounts. To take account of variability, a confined random walk was additionally implemented in the occurrence process of the model. The model was developed in a manner that allows universal use. By incorporating the elevation of a station and the time of the year as input parameters, the model was made applicable at any time and for any location in Germany. In a second step the model was applied to investigate daily precipitation records. For that, linear changes were implemented into the model. As result, a previously found decrease of records in the summer season can be explained by changes in the stratiform precipitation distribution. However, the decrease of precipitation records in the summer season is too low to rule out random processes as cause of this decrease. In contrast, the increase in the mean number of precipitation records in winter season cannot be reproduced with the developed model. A possible explanation for that is the neglect of spatial correlations in the amount process. An appropriate method for taking spatial correlations into account, could be a Copula approach.

Table of Contents

1. Introduction	5
1.1. Precipitation models	8
1.1.1. The occurrence process	8
1.1.2. The amount process	11
1.2. Records	14
1.3. Aim	18
2. Data and data processing	19
2.1. Dataset DWD _{all}	19
2.2. Dataset DWD _{con}	21
2.3. SYNOP-Dataset	21
2.4. Disaggregating convective and stratiform precipitation	23
3. Development of a precipitation model	27
3.1. Development of the occurrence process	28
3.2. Development of the amount process	33
3.2.1. Development of a distribution for precipitation	33
3.2.2. Time and altitude dependence	40
3.3. Taking account of variability in the occurrence process	42
3.4. Model results vs. observation	44
3.4.1. Precipitation occurrence	44
3.4.2. Precipitation amounts	46
4. Application to precipitation records	48
4.1. Observed changes in the past	48
4.1.1. Observed changes in the occurrence process	49
4.1.2. Observed changes in the amount process	50
4.2. Implementation of a drift in the precipitation model	52
4.3. Modelled precipitation records	55
5. Conclusion and Outlook	58

Appendix	63
Bibliography	68
Acknowledgements	75




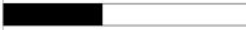




1. Introduction

Precipitation is essential for life. It provides water for the continents, an important prerequisite for life on earth. When there are reduced amounts or even a lack of rain over a longer period of time, there are widespread consequences: increased risk of fires, shortage in water- and power supply as well as crop shortfall and resulting increase in food prices. A well-known example is the heatwave in Central Europe in 2003: a long lasting dry spell in combination with extremely high temperatures lead to 70,000 fatalities and a financial loss of US\$ 13.8bn [Robine et al. (2008); Munich RE (2017)]. Another example was the dry spell in November 2011 in which a mean precipitation of 3mm lead to the driest month ever recorded in Germany since the beginning of comprehensive weather records in 1881 [Müller-Westermeier (2012)].

On the other hand, high amounts of precipitation over a long period of time also have negative consequences, especially for agriculture. Crops cannot be harvested and are often spoiled if it is too wet. The most fatal situations often occur when extremely high amounts of precipitation are limited to a very short period of time leading to flooding and high economic losses. The highest daily precipitation ever recorded in Germany was 312mm, measured in Zinnwald-Georgenfeld on August 12th 2002 [Rudolf and Rapp (2003)]. In the following weeks, the area around the river Elbe suffered from devastating flooding. There was an economic damage of US\$ 11.6bn [Munich RE (2017)] in Germany. In May 2013, there was another extreme flooding in south and eastern Germany: Even if the highest daily precipitation was far below previously reported records in 2002, the month May was the second wettest since 1881. Several local records were recorded in May and June [DWD (2013)] and the overall losses were estimated to US\$ 10.4bn [Munich RE (2017)].

In the summer months of previous years heavy precipitation was mainly observed very locally in Germany. This summer official 197mm of rain were recorded within several hours in Berlin by the German Weather Service while daily rain totals of a private weather station even exceeded 260mm [Gebauer et al. (2017)]. Even more rain in a shorter period of time was recorded in Münster in July 2014: Within 7

Table 1.1.: List of the eight costliest hydrological events from 1980–2016, sorted by the convective percentage of the total rainfall amount in the given periods. The convective percentages were calculated by using SYNOP reports to disaggregate types of precipitation. Data source: Munich RE (2017) & SYNOP reports (see sec. 2).

Period	Event	Overall losses	Fatalities	Convective rain
28.07. – 29.07.2014	Flash flood	US\$ 600m	2	
27.05. – 30.05.2016	Flash flood	US\$ 830m	4	
31.05. – 01.06.2016	Flash flood	US\$ 2,000m	7	
16.07. – 04.08.1997	Flood	US\$ 370m	–	
25.05. – 15.06.2013	Flood	US\$ 10,400m	8	
06.08. – 09.08.2010	Flood	US\$ 1,100m	4	
04.08. – 13.08.2002	Flood	US\$ 11,600m	21	
17.12. – 27.12.1993	Flood	US\$ 600m	5	

hours 292mm and within just 2 hours 245mm of rain was measured [Becker et al. (2014); Axer et al. (2015)]. The German Weather Service declared this 2-hour-value as a new German record [Becker et al. (2014)]. According to Munich RE (2017) this was the most expensive flash flood since 1980 with an overall damage of US\$ 510m whereof US\$ 230m was insured. However, in 2016 this record was even beaten twice: In the small town of Braunsbach (Baden-Württemberg) damage of US\$ 830m (US\$ 500m insured [Munich RE (2017)]) was caused by a flash flood on May 29th while only three days later a damage of US\$ 2bn (US\$ 830m insured [Munich RE (2017)]) in Simbach am Inn was caused by another flash flood. In Gundesheim, a town 50 km West of Braunsbach, a daily precipitation record of 122mm was recorded for the day of the flooding. The weather station close to Simbach am Inn did not record any values at the critical hours of the event but the total daily precipitation sum is expected to be similar of these in Gundelsheim [Piper et al. (2016)]. The discrepancy between the high overall damage and the relatively low measured precipitation amounts can be explained by the topographic position of Braunsbach and Simbach am Inn [Axer et al. (2017)].

The main cause of costly and destructive events can be identified by looking at the previously described examples: On the one hand heavy continuous rainfall can

Table 1.2.: Overview of typical characteristics of stratiform and convective precipitation.

	stratiform	convective
scale	large-scale	local
duration	long	short
intensity	less intensive	more intensive
typical example	large-scale frontal rain, drizzle	thunderstorms, showers

lead to supra-regional flooding (e.g. flooding in 2002 and 2013), on the other hand extreme local rainfalls can swamp the sewage water system leading to flash floods (e.g. flash flood events 2014 and 2016). In fact these two event types result from two major processes of precipitation generation: *stratiform* and *convective* precipitation (see Tab. 1.1). These two precipitation types differ in many aspects, such as the cloud type of which the precipitation falls out or the fall velocity of the rain drops in relation to the vertical air motion [Houze Jr (2014)]. Stratiform precipitation is characterized by long-lasting large-scale precipitation (e.g. drizzle) while convective precipitation is typically very local and more intensive but of short duration (e.g. thunderstorms) (see Tab. 1.2 for a summary of the main differences). Furthermore, it has a highly seasonal dependency with much more convective events occurring in summer compared to winter season.

The extreme precipitation events mentioned above are often associated with climate change. According to the Clausius–Clapeyron rate, the water holding capacity of the atmosphere increases by around 7% per degree warming. Since precipitation releases latent heat, the global total rainfall amount is expected to increase by just 1% – 3% K^{-1} with warming [Solomon (2007); Stephens and Ellis (2008)]. However, according to the latest studies [Berg et al. (2013); Zhang et al. (2017)] the convective precipitation – which is already very intense – increases the most, even more than the Clausius-Clapeyron rate. This is probably due to local feedbacks related to the convective activity [Lenderink et al. (2017)]. The increase in the intensity of strong rainfall is expected to be associated with a decrease in light and moderate rains and/or a decline in the frequency of precipitation events [Trenberth et al. (2003)].

These findings lead to the assumption that an increase in precipitation records is due to global warming. This is mainly true for the summer months in which more

convective precipitation events occur. The latest extreme precipitation events observed in Germany seem to validate this hypothesis. Surprisingly, an investigation of daily precipitation records in Germany [von Bomhard (2014)] led to an opposite result: in the summer months fewer records were observed than theoretically expected in a stationary climate. However, in the winter months around 16% more records were observed than expected. The reason of the increased (winter) or decreased (summer) precipitation records will be investigated in this work by using a newly developed statistical precipitation model (see also sec. 1.3).

The following subsections are intended to give a comprehensive overview of statistical precipitation models (sec. 1.1) and record statistics (sec. 1.2).

1.1. Precipitation models

Generally stochastic precipitation models have two submodules: an occurrence process and an amount process. The occurrence process has to classify whether a certain day is a wet or a dry day. For dry days the precipitation amount is equal to zero but for a day to be classified a wet day the amount process has to determine a nonzero precipitation amount.

1.1.1. The occurrence process

Since the first computer models were used for generating daily weather variables a lot of different possibilities for implementing the occurrence process were tested. A recent work from Ng and Panu (2010) compares four different models based on the short-term temporal-dependency, dry- and wet-spell length and goodness-of-fit. Using these three assessment criteria, a two state, second order Markov chain showed the best performance. In addition to the Markov chain the authors also investigated an alternating renewal process (ARP) and introduced the Dictionary approach (originally developed in the field of genome pattern) and a probabilistic word matrix model referring to a list of “words” comprised of precipitation “letters”. The ARP spell length model and the Markov chain model are well performing models which are frequently used. These models will be discussed below in more detail. For more details about additional approaches in the context of precipitation occurrence processes see Ng and Panu (2010) and references therein.

Markov chain

In general, a Markov chain model is characterized by its order n a number of states m . In the context of rainfall, the Markov chain is usually implemented with two states (a dry day and a wet day). In addition to a wet state and a dry state some studies included a certain threshold for separating a state with little rainfall amounts from a state with higher rainfall amounts. For example, Stern and Coe (1984) defined a third “trace” state for days with precipitation amounts less than 2.5mm.

The order of the Markov chain can be imagined as the number of days the chain remembers (therefore also called memory depth). With an increasing order of the Markov chain process the number of required parameters increases exponentially. For a two state, k -th order Markov chain 2^k parameters are required. To determine the optimum order of a Markov chain model for a given set of data, usually maximum likelihood based information criteria are used. These information criteria detect the goodness of fit using a maximum likelihood but also include a penalty which increases with adding more parameters and therefore with a higher order of the process. Usually the Akaike Information Criterion (AIC) [Akaike (1974)] or the Bayes Information Criterion (BIC) [Schwarz et al. (1978)] are used which only differ by the used penalty. A more recent study about modelling precipitation [Lennartsson et al. (2008)] uses a Generalized Maximum Fluctuation Criterion (GMFC). In contrast to the AIC and BIC the GMFC is not based on maximum likelihood. Such a maximal fluctuation method was first developed by Peres and Shields (2005) and further established by Dalevi et al. (2006) to the more general GMFC estimator. In comparison to four other estimators the authors identified the GMFC to be superior in several respects, while the BIC underestimated the optimum order for a moderate data sample noticeable [Dalevi et al. (2006)].

The first order two state Markov chain model is the simplest and most commonly used precipitation occurrence process [e.g. Gabriel and Neumann (1962); Katz (1977); Wilks (1989, 1999)]. As it has a memory depth of one, it follows a Geometric distribution, given as

$$Pr(X = x) = p_{r/d} (1 - p_{r/d})^{x-1} \quad (1.1)$$

where X is the length of a wet/dry spell. For this reason, the probability for gener-

ating a long interval of x consecutive dry (with $p = p_d$) or wet (with $p = p_r$) days is relatively small. Some studies argue that long dry spells are modelled too infrequently by this approach [e.g. Buishand (1978); Racsco et al. (1991)]. To handle this problem while keeping the number of parameters as small as needed, a hybrid-order Markov chain model was established [Stern and Coe (1984); Wilks (1999)]. This means that a first order Markov chain is used for modelling the wet state but higher orders for the dry state are allowed. For this hybrid-order Markov chain model the number of parameters only increases with $k + 1$ rather than 2^k for a k -th hybrid-order Markov chain [Wilks (1999)].

Spell length models

Another approach for generating the right occurrence rate of (long) dry and wet spells is given by the spell length model (also called alternating renewal process – ARP-model). As mentioned in the previous section the first order two state Markov chain model can be rewritten into a spell length model following a Geometric distribution (see eq. 1.1). In general, a spell length model generates the length of either a dry or a wet spell by a given distribution, in the following it generates the spell length of the opposite type and so forth [Wilks and Wilby (1999)]. In the literature also the Negative Binomial distribution [e.g. Wilby et al. (1998)], a modified truncated Negative Binomial distribution [e.g. Woolhiser and Roldan (1982)] or a superposition of two distributions – for example a mixture of two Geometric distributions [Racsco et al. (1991)] – are used instead of the geometric distribution (first order two state Markov chain) for implementing a spell length model.

Comparison of occurrence processes

As presented above (see chapter 1.1.1), for stations in Canada a second order two state Markov chain was superior to three other occurrence processes including a spell length model [Ng and Panu (2010)]. In addition, a lot more investigations about the optimum occurrence process for different locations were made. Stern and Coe (1984) compared different realizations of the Markov chain model (up to three states, five orders and including hybrid models) for different stations around the world. They concluded that different stations need different realizations of the Markov chain. Furthermore, the authors pointed out that for most parts of the world an assumption of stationarity throughout the year is not appropriate even for periods as short as one month. A similar result was found by Lennartsson et al. (2008). By comparing the

best orders of Markov chains found for 20 stations in Sweden they concluded that the optimum order of the Markov chain varies between the stations as well as during the year. These authors are taking the time dependency of the Markov chain in account by determining the optimal Markov chain order for each month. However, some other studies established a time response model based on Fourier series to describe the seasonal variability [e.g. Woolhiser and Pegram (1979); Woolhiser and Roldan (1982); Stern and Coe (1984)]. One of these studies compared a first-order Markov chain and a spell length model with a geometric distribution for the wet days and a truncated negative binomial distribution for dry days [Woolhiser and Roldan (1982)]. Both models were nonstationary by allowing daily variation of the parameters of both models (described by a Fourier series). Using the AIC estimator, the Markov chain model was superior to the spell length model for five tested stations in the U.S. [Woolhiser and Roldan (1982)]. Consistent with this study a comparison of different realizations of the Markov chain and different spell length models for 30 stations in the U.S. showed that a first-order Markov chain model is superior to higher order Markov chain models as well as spell length models according to the BIC estimator [Wilks (1999)]. In this comparison a mixed geometric spell length model was found to be worst, but a negative binomial spell length model performed best for the west stations after dividing the stations according to their geographic location [Wilks (1999)].

1.1.2. The amount process

As soon as a day is declared a wet day by the occurrence process, an additional process has to determine the amount of rainfall. To do so, probability distributions are compared to find the best fit with observed precipitation amounts. Subsequently, the often used Gamma distribution and the Weibull distribution will be presented in more detail. The latter is of special interest for extreme precipitation amounts. Additionally, further distributions have been tested such as the lognormal [Shoji and Kitaura (2006)] or the generalized Pareto distribution [Lennartsson et al. (2008)] which will not be described in more detail.

The Gamma distribution

The Gamma distribution is the most popular choice for simulating daily precipitation amounts [e.g. Thom (1958); Katz (1977)]. The probability density function (PDF) of the Gamma distribution is given by [Wilks (2011)]

$$f_{Gamma}(x) = \left(\frac{x}{\beta}\right)^{\alpha-1} \frac{e^{-\frac{x}{\beta}}}{\beta\Gamma(\alpha)}, x, \alpha, \beta > 0. \quad (1.2)$$

With α and β being the shape and the scale parameter respectively and in the context of precipitation, x is the daily amount of rainfall. The most common method to determine α and β is to use maximum likelihood estimators [Wilks (2011)]. For $\alpha = 1$ the Gamma distribution can be limited to a one-parameter distribution which is called exponential distribution. This simplified form of the Gamma distribution was also used in the past for generating the daily precipitation amount [e.g. Woolhiser and Roldan (1982)].

While the amounts are generally modelled as being independent and identically distributed (i.i.d.) some approaches used distributions in which the amounts depend on previous day(s). For example, Katz (1977) used two Gamma distributions depending on whether the previous day was wet or dry (later referred to as chain-dependent). In a slightly different approach three Gamma distributions with a fixed shape parameter were used by Wilby et al. (1998) as well as by Wilks (1999) depending on the position in a wet spell (later referred to as position-dependent).

A very interesting investigation about the validity of the Gamma distribution for 90 stations in Europe pointed out that the Gamma distribution is probably not as valid as commonly believed [Vlček and Huth (2009)]. The authors state that a Kolmogorov-Smirnov (KS) test is often used to assess the goodness-of-fit. They argue that this is incorrect in the sense that the shape and scale parameter of the Gamma distribution are estimated from the data sample and therefore the KS modification of Lilliefors [Lilliefors (1967)] has to be used (for more details see Wilks (2011)). When using the Lilliefors instead of the KS test the Gamma distribution is more often rejected. For example for modelling the winter season 42% are rejected instead of 14% [Vlček and Huth (2009)].

Due to this result it should be considered to test other probability distributions such as the Weibull distribution which is described in more detail below.

The Weibull distribution

Like the Gamma distribution the Weibull distribution is a two parameter distribution. It is given by [Wilks (2011)]

$$f_{Weibull}(x) = \left(\frac{x}{\beta}\right)^{\alpha-1} \left(\frac{\alpha}{\beta}\right) \exp\left[-\left(\frac{x}{\beta}\right)^{\alpha}\right], \quad x, \alpha, \beta > 0, \quad (1.3)$$

with α , β and x being again the shape and the scale parameter and the daily precipitation amount respectively. Like in the case of the Gamma distribution the Weibull distribution follows the exponential distribution when the shape parameter is equal to one. An important difference between the two distributions is the shape of their tail. While the tail of the Gamma distribution is exponential for all shape parameters, the tail of the Weibull distribution gets heavier with a decreasing α and lighter with an increasing α . The tail is called “heavy tail” for $\alpha < 1$ and “light tail” for $\alpha > 1$. Because distributions with a heavy tail have higher probabilities for generating high precipitation amounts, a heavy tailed Weibull distribution might be a better choice than the Gamma distribution for an implementation of the amount process in the application of records.

Comparison of amount processes

A comparison of an Exponential, Gamma, and Mixed Exponential distribution with respect to chain-dependency and independency was published by Woolhiser and Roldan (1982). Here a mixed exponential distribution means a mixture of two Exponential distributions in the same manner as it was already discussed for the case of the spell length models. Using the AIC estimator, the independent distributions were superior to their chain-dependent companions which is consistent with findings from Katz (1977) for the case of a Gamma distribution. The independent Mixed Exponential distribution was found to be the best choice of all compared distributions [Woolhiser and Roldan (1982)]. A similar study for 30 U.S. stations was published by Wilks (1999). He compared an independent Gamma, a position-dependent Gamma and also an independent Mixed Exponential distribution. Interestingly the use of three Gamma distributions dependent on the position in a wet spell was superior to the i.i.d. Gamma distribution according to the BIC estimator. Nevertheless, the Mixed Exponential distribution even led to a further enhancement. In addition to the BIC estimator, Wilks (1999) also tested the interannual variability as a goodness criteria to evaluate models. For this purpose he summed 30 consecutive daily precipitation amounts and counted the number of rainy days in this time period. Based on this, he calculated a variance overdispersion which is given by the relation of the observed to the modelled variance. With this method Wilks (1999) pointed out that the combination of a Mixed Geometric spell-length model with a Mixed Exponen-

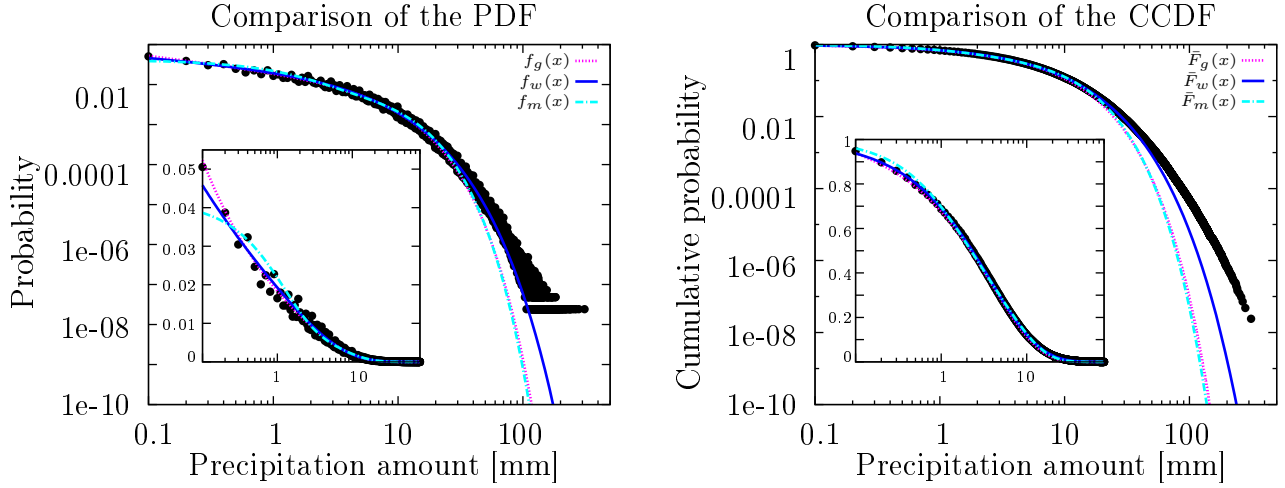


Figure 1.1.: Comparison of fits of the Gamma distribution (purple), Weibull distribution (blue) and Mixed Exponential distribution (light blue) of observations in Germany (dataset DWD_{all}). Both for the probability density function (left) and the complementary cumulative distribution function (right).

tial distribution was superior compared to all other combinations of occurrence and amount processes tested. This is a surprising result as the Geometric distribution was actually inferior compared to all other tested occurrence processes (see section 1.1.1). An additional investigation refers to extreme precipitation amounts: By comparing the largest precipitation amounts modelled with the observed ones, all tested amount processes turned out to be unsuitable to generate very high precipitation amounts as frequently as they are observed in reality. Especially for precipitation amounts larger than 100mm the models reach their limit [Wilks (1999)]. As this is an important feature referring to the investigation of records the use of extreme-value distributions should be considered. Unfortunately, extreme-value distributions such as the Weibull distribution are very rarely used for simulating daily precipitation amounts. Nevertheless, a Weibull distribution with a shape parameter of $\alpha = 0.75$ was found to be superior compared to an Exponential and a Beta-P distribution for 33 stations east of the Rocky Mountains [Selker and Haith (1990)]. Not surprisingly the best improvement was found for the largest precipitation amounts.

1.2. Records

The basic theory of records on independent and identically distributed (i.i.d.) random variables (RV's) was developed over six decades ago. One of the pioneers of

the mathematical theory of records was Chandler. He published one of the first papers on this issue in 1952 [Chandler (1952)]. Since then the theory of records has continuously been developed to a own research area in probability theory. A recent summary of the theory of record statistics including the state of research of a few application examples can be found in the review of Wergen (2013). The following remarks in this section refer to this review.

Let X_1, X_2, \dots, X_n be a time series of RV's. Then entry n is a new upper record, if it exceeds all previous entries:

$$X_n > \max \{X_1, X_2, \dots, X_{n-1}\}. \quad (1.4)$$

Analogously, the n th event is a record low, if it is below all previous entries:

$$X_n < \min \{X_1, X_2, \dots, X_{n-1}\}. \quad (1.5)$$

Of high interest is the probability, that entry n is a record – also known as the record rate. For an upper record these probability is given by

$$P_n = \text{Prob} [X_i > \max \{X_1, X_2, \dots, X_{n-1}\}]. \quad (1.6)$$

The first entry is by definition a record: $P_1 = 1$. The second entry is a record, if it exceeds the first one. Its probability is $P_2 = \frac{1}{2}$ for i.i.d. RV's. Analogously, $P_3 = \frac{1}{3}$, $P_4 = \frac{1}{4}$, \dots . This leads to

$$P_n = \frac{1}{n}, \quad (1.7)$$

for the probability of a record at time n .

Another commonly used quantity in record statistics is the mean record number R_n , which is the number of records that occurred in the time series up to time n . It is simply given by a harmonic series:

$$R_n = 1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{n} = \sum_{k=1}^n P_k = \sum_{k=1}^n \frac{1}{k}. \quad (1.8)$$

In the context of daily precipitation it has to be taken into account that a record can only occur on days with rain. Let p be the probability for getting a rainy day (= rain probability) and m ($m \in \{1, 2, 3, \dots, n\}$) be the number of rainy days up to day n , then the probability of getting a precipitation record is given by [von Bomhard (2014)]:

$$P_n = \sum_{m=1}^n \frac{1}{m} \binom{n-1}{m-1} p^m (1-p)^{n-m}. \quad (1.9)$$

Evaluating the sum and using the binomial theorem gives:

$$P_n = \frac{1}{n} (1 - (1-p)^n). \quad (1.10)$$

With that, the mean record number of daily precipitation is given by:

$$R_n = \sum_{k=1}^n \frac{1}{k} (1 - (1-p)^k). \quad (1.11)$$

This expression is only valid under the assumption that the rain probability p is constant. So, just in case that every year the same number of rainy days occur. Of course, this is not the case. In reality p is very variable. For example in the German summer of 1983 it was raining at less than 30% of all days, while in the summer of 1987 at more than 60% of all days precipitation was recorded (compare Fig. 4.5). To minimize the confounding influence of the rain probability p , k in eq. 1.11 can be raised. E.g. for Germany it was found, that the confounding influence can be negligible for $k = 20$ [von Bomhard (2014)]:

$$\tilde{R}_{n,20} = \sum_{k=20}^n \frac{1}{k} (1 - (1-p)^k) \approx \sum_{k=20}^n \frac{1}{k}. \quad (1.12)$$

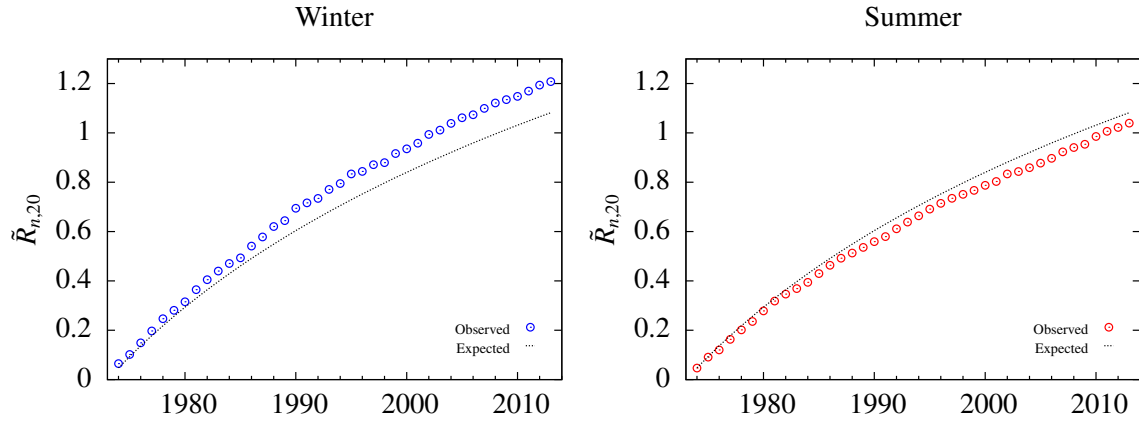


Figure 1.2.: Additional mean records $\tilde{R}_{n,20}$ since 1974 based on time series for the years 1954–2013 for the winter (left) and summer (right) seasons. Dotted lines show the prediction for a stationary climate and circles show the observations.

This modified mean record number $\tilde{R}_{n,20}$ is the number of additional records from entry $k = 20$ to the n th step in a time series of length n . For example, the number of additional records (modified mean record number $\tilde{R}_{n,20}$) in a stationary climate (i.i.d. case) is expected to be 1.08 for the last 40 years of a time series of 60 years. Interestingly, von Bomhard (2014) counted 1.25 additional daily precipitation records since 1974 in the winter seasons from 1954 to 2013 and in the summer seasons just 1.02 additional records were observed (see Fig. 1.2).

So far, there are not many more studies in the context of precipitation records. One very recent study used both a statistical- as well as a dynamical model to show that there is a high chance for new record highs of rainfall totals in winter months in the UK under current climate conditions [Thompson et al. (2017)]. Another recent paper used gridded data of monthly 1-day precipitation amounts to relate an increase in record-breaking rainfall events of 12% over 1981–2010 to global warming [Lehmann et al. (2015)]. However, no abnormalities in daily precipitation records of stations in Scandinavia were found by Benestad (2003). Interestingly, Benestad (2003) also used the same stations for an investigation of daily temperature records. Though he found an increase in daily temperature records, he was unable to prove a significant trend. Indeed, other studies also failed in a proof of increasing temperature records [e.g. Redner and Petersen (2006)] although a global warming of 0.85 °C is observed since 1880 [Stocker et al. (2013)]. Only since 2009 the first studies demonstrated a significant trend in temperature records [Wergen and Krug (2010); Meehl et al. (2009)]. Wergen and Krug (2010) observed a significant increase of

upper temperature records. For the year 2005 they found an increase of about 40% in temperature records registered at European stations compared to the period of 1976 to 2005. By using a linear drift model (LDM) they concluded that this increase is due to climate change.

The first study of a LDM was published by Ballerini and Resnick [Ballerini and Resnick (1985, 1987)]. They considered a model with i.i.d. RV's X_n being exposed to a linear drift of the form cn :

$$Y_n = X_n + cn, \quad n \geq 1. \quad (1.13)$$

Where c is a positive constant ($c > 0$). The RV's Y_n on the left hand side of eq. 1.13 are no longer identically distributed. The LDM, therefor, depends on the distributions of the RV's. A detailed discussion of the LDM for the three different classes of extreme value statistics – *Weibull*, *Gumble* and *Fréchet* class – is given in Franke et al. (2010).

1.3. Aim

The aim of this work is to develop a statistical precipitation model for Germany which can be used to investigate the observed differences (to theory expectations) in the mean record number (see Fig. 1.2). In the context of precipitation records it seems highly important to simulate high precipitation amounts as close to reality as possible. Since this aspect is a major weakness of commonly used amount processes (sec. 1.1) a different probability distribution will be developed in this work and implemented in the model. Furthermore, the occurrence process should also be able to discriminate between the two types of precipitation (*convective* and *stratiform*). In addition the model should not be specific to one station, as it is the case for previously developed models but valid throughout Germany. For that, the utilized precipitation data (sec. 2) is analyzed according to their dependence on topographic height (above sea level) of the station and the time of the year (sec. 3). The variable *height* of station and *time of the year* will then be used as input parameters in the model. Finally, observed changes in the precipitation pattern will be implemented in the model by linear drifts (see LDM in sec. 1.2) and the influence on the mean record number will be investigated (chapter 4).

2. Data and data processing

In this work three different sets of data were used for analysis: Two of them are rain gauge data of the German weather service (DWD) and the third one includes SYNOP (synoptic observation data) reports. The latter one was used for disaggregating convective and stratiform precipitation.

The data of the DWD datasets are provided via a ftp-server¹. For each station there is a separate file including the daily precipitation in mm in which 0.1mm is the minimal documented precipitation amount. The recording of precipitation was first started at Hohenpeißenberg in 1781, which is the oldest meteorological mountain station worldwide [Strauch (2011)].

In the following chapter the three sets of data will be described in more detail and subsequently the processing of the data will be explained.

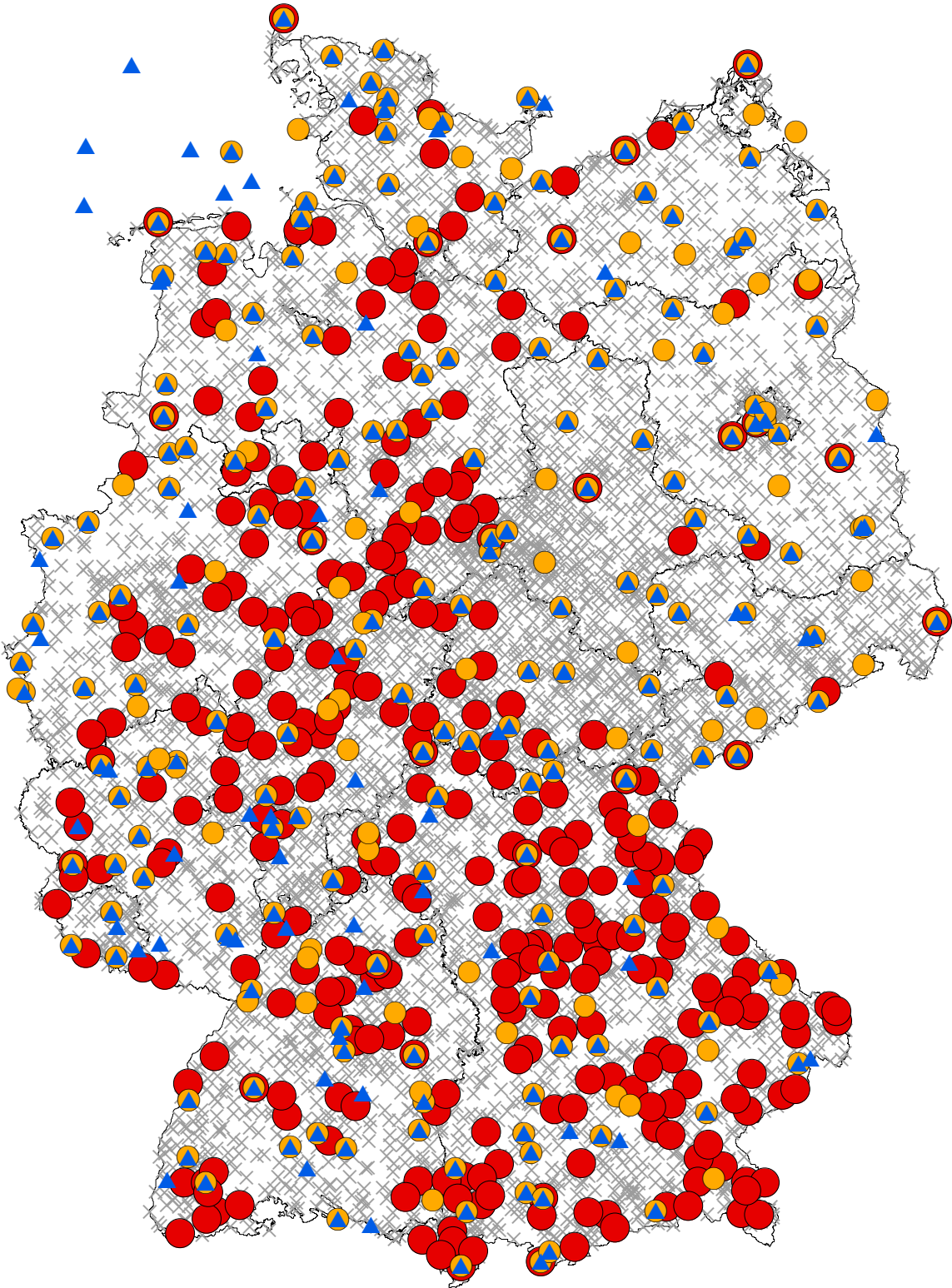
Table 2.1.: Overview of the three different datasets used in this study.

Dataset	Time range	Rain gauges	Rainy days	Source
DWD _{all}	1950–2016	5,400	45,000,000	DWD
DWD _{con}	1954–2013	320	3,500,000	DWD
SYNOP	1950–2016	300	1,400,000	NOAA

2.1. Dataset DWD_{all}

The dataset DWD_{all} includes all precipitation data provided by the DWD since 1950. For each station there are metadata available including the history of the rain gauge. It includes information about the elevation (above sea level) as well as the geographic longitude and latitude in decimal degrees. For the year 1901 already more than 1,400 stations are available and the precipitation network reached its peak with 4,500 stations in the 1980s [Kaspar et al. (2013)]. In total the ftp-server

¹Available online: `ftp-cdc.dwd.de`.



▲ SYNOP (NOAA) ● SYNOP (OGIMET) ● DWD_{con} × DWD_{all}

Figure 2.1.: Spatial distribution of the used stations throughout Germany. Symbols show the rain gauges of dataset DWD_{all} (gray crosses), of dataset DWD_{con} (red circles) as well as the synoptic stations provided by OGIMET (orange circles) and NOAA (blue triangles).

provides information of nearly 5,400 rain gauges with data on approximately 45 million rainy days available since 1950 (see Tab. 2.1). But great deficiencies exist in digitally available precipitation data. More than 30% of all the available precipitation records are not yet digitized. The greatest lack of information exists in eastern Germany, where only about 50% of the rain gauges are digitally available before 1969 [Kaspar et al. (2013, 2015)]. The observed data passes through several quality checks. However, the DWD estimates that still around 0.1–1% of the values in its provided climate (not only precipitation) data are doubtful [Kaspar et al. (2013)].

2.2. Dataset DWD_{con}

The dataset DWD_{con} is the same as the one used by von Bomhard (2014). It consists only of rain gauges which fulfill the following criteria:

- Data of the respective station has to be complete and consistent for the period of investigation. This means that either a gap of data or a measuring error (marked by -999) leads to the exclusion of the entire station.
- If a rain gauge has been moved in the period of investigation, the maximal distance of movement should not exceed 1 km and 50 m of altitude. These values were set arbitrary but should ensure that the data remains comparable after the movement.

After applying these criteria to all rain gauges in the period of 1954–2013, only 320 stations out of 5,400 (complete dataset DWD_{all}) remain in this consecutive dataset. In particular, just very few stations are included in the former GDR due to the lack of information in eastern Germany (see Fig. 2.1). However, in this work a dataset that is consistent and consecutive is needed for the investigation of records (sec. 4).

2.3. SYNOP-Dataset

The dataset SYNOP was used to investigate stratiform and convective daily precipitation separately. For this, SYNOP reports from the Integrated Surface Database² (ISD) for January 1950 to September 2016 provided by the National Oceanic and Atmospheric Administration (NOAA) were used. Since – by unknown reasons – the archiving of synoptic data of ISD decreases since the beginning of 1990, additional

²Available online: <https://www.ncdc.noaa.gov/isd>

Table 2.2.: Example of a SYNOP report from June 1st 2016 at Cologne/Bonn airport. Colored numbers show the information which were used in this study: present weather code (red numbers), cloud type (orange numbers) and precipitation totals in mm (blue numbers) for a given period of time (green numbers; 1 = 6 hours, 2 = 12 hours).

SYNOPS from 10513, Cologne / Bonn (Germany)		
01/06/2016 18:00->	AAXX 01181	10513 01130 82603 10175 20167 39985 40101 56016 60312 78186 879// 333 10175 20148 32015 55300 20017 30017 60167 84803 87905 91007 90760 91108 90760 91206 96482 96463==
01/06/2016 17:00->	AAXX 01171	10513 41128 82705 10167 20165 39989 40106 57018 78188 873// 333 55300 20076 30076 86803 87904 91007 90710 91108 90710 91206==
01/06/2016 16:00->	AAXX 01161	10513 41250 82704 10167 20164 39995 40112 58016 78186 869// 333 55300 20181 30181 85704 85907 91007 90710 91107 90710 91205==
01/06/2016 15:00->	AAXX 01151	10513 21220 82704 10166 20163 30001 40118 58014 78162 869// 333 55300 20176 30176 60037 85707 86904 91006 90730 91106 90730 91204==
01/06/2016 14:00->	AAXX 01141	10513 41217 82803 10165 20162 30007 40124 58011 76166 877// 333 55300 20187 30187 87704 91005 90710 91106 90710 91204==
01/06/2016 13:00->	AAXX 01131	10513 41230 82803 10164 20161 30011 40128 58008 76362 877// 333 55300 20193 30193 87704 91005 90710 91105 90710 91203==
01/06/2016 12:00->	AAXX 01121	10513 01230 82702 10162 20158 30015 40132 58004 6013 72086 877// 333 31/// 55300 20364 30364 60037 87706 91005 90760 91105 90760 91203==

SYNOPSIS reports of NOAA were used that were provided by the internet platform OGIMET³ since 1999. In total, data of 800,000 rainy days provided by ISD and further 600,000 by OGIMET (see Fig. 2.2, top) were available for the data analysis.

The synoptic data are hourly observation data which are available in encoded form (see Tab. 2.2 for an example of a SYNOPSIS report). These data could be either entered manually by an employee of the weather station (manned station) or could be generated automatically (automatic station). 14% of the data used in this work were generated automatically while the majority of data were recorded manually. The SYNOPSIS reports contain information about the current weather situation: wind speed, temperature, dewpoint, pressure and many more. For this study the information about precipitation (blue in Tab. 2.2), cloud type (orange in Tab. 2.2) and the present weather (red in Tab. 2.2) was used. For full details on the SYNOPSIS reports including decode tables see the Federal Meteorological Handbook (FMH) number 2 [US Dept. of Commerce (1979)]. The following section describes the used method of data processing to obtain a daily dataset with information about the amounts of precipitation and the type of precipitation.

³<http://ogimet.com/index.phtml.en>

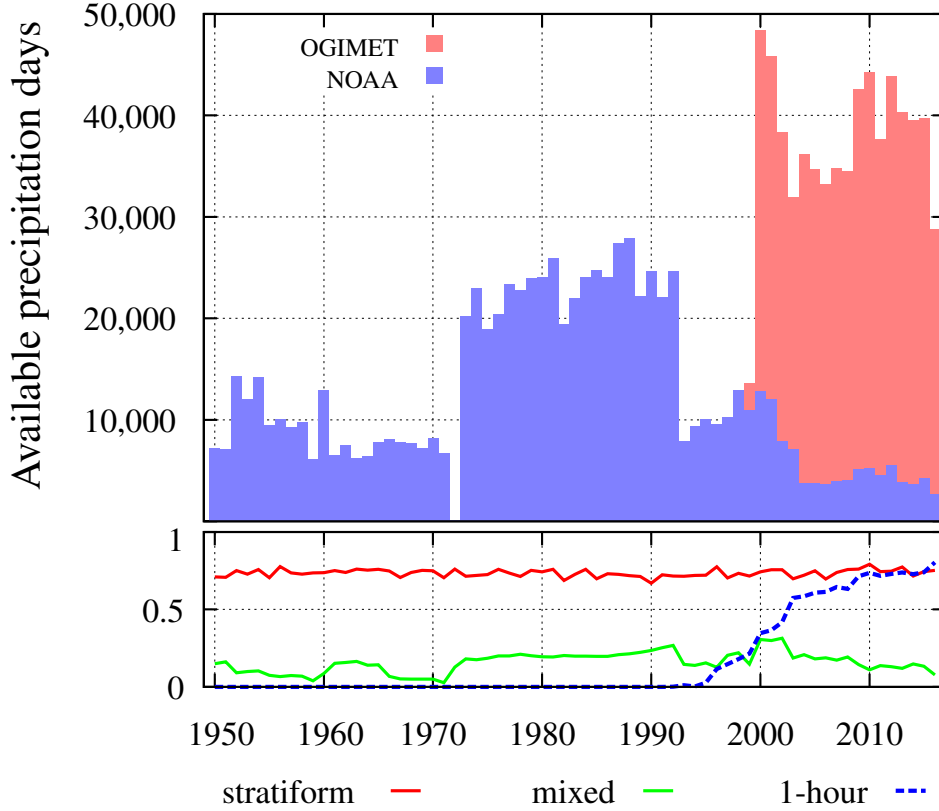


Figure 2.2.: Overview of the used synoptic data. Top: Availability of days giving information on precipitation from SYNOP reports, provided by OGIMET (red bars) and NOAA (blue bars). Bottom: Percentages of days with no identified precipitation type (green line), of stratiform precipitation of remaining identified days (red line) and proportion of disaggregation based on hourly precipitation data (blue dashed line) for each year from 1950–2016.

2.4. Disaggregating convective and stratiform precipitation

Because of the differences in the characteristics of the two precipitation types (as mentioned earlier in sec. 1) many different techniques were developed to distinguish stratiform from convective rainfall events. In 2003 a comparison of six different methods was presented by Lang et al. (2003) but also hereafter new techniques were established. In this study an algorithm developed by Rulfová and Kyselý (2013) (see Fig. B.1 in the appendix) was used for disaggregating *convective* and *stratiform* precipitation from station weather data. This algorithm uses the precipitation amounts, cloud type and present weather information of SYNOP reports. The present weather observations are numerically coded from 0 to 99 (red numbers in Tab. 2.2). The

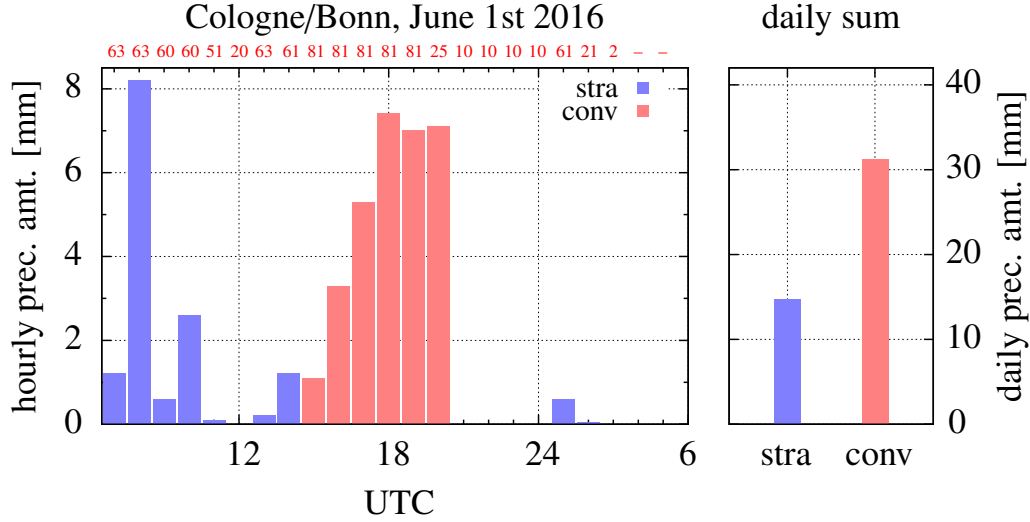


Figure 2.3.: Example of the used disaggregation process for Cologne/Bonn airport on June 1st 2016. Left: Hourly precipitation amounts are identified as stratiform (blue bars) or convective (red bars) driven by hourly present weather codes (red numbers). Right: All collected stratiform as well as convective precipitation of one day are summed up to figure out which type of precipitation amount is predominant. In this example, the convective precipitation dominates, giving a convective day with a rain total of 46mm by the algorithm.

decoding of the present weather observations with the used disaggregation of Rulfová and Kyselý (2013) is given in appendix A.1 (stratiform) and A.2 (convective). Since measurements of precipitation are only recorded in intervals of 6 hours in the SYNOP reports (blue numbers in Tab. 2.2), it is possible that a precipitation value can not be clearly allocated to one type of precipitation. One example for that is the SYNOP report from June 1st 2016 at Cologne/Bonn airport (Tab. 2.2). Since 1995 the German Weather Service provides hourly precipitation values for many stations. Deviating from the algorithm of Rulfová and Kyselý (2013), in this work the SYNOP reports were supplemented with these hourly data whenever possible to gain a clear classification of the types of precipitation (see Fig. 2.5). The availability of hourly precipitation values has been increased over the years, so that in 2016 more than 80% of all SYNOP reports could be complemented with these values (see blue dotted curve in Fig. 2.2 bottom). For all other stations where this was not possible it was preceded as defined in the algorithm of Rulfová and Kyselý (2013): in the case of a non-distinctive allocation based on the present weather situation, information about low level clouds (orange numbers in Tab. 2.2) was used as secondary criterion (see appendix B.1). Findings of Langer and Reimer (2007) were used to classify

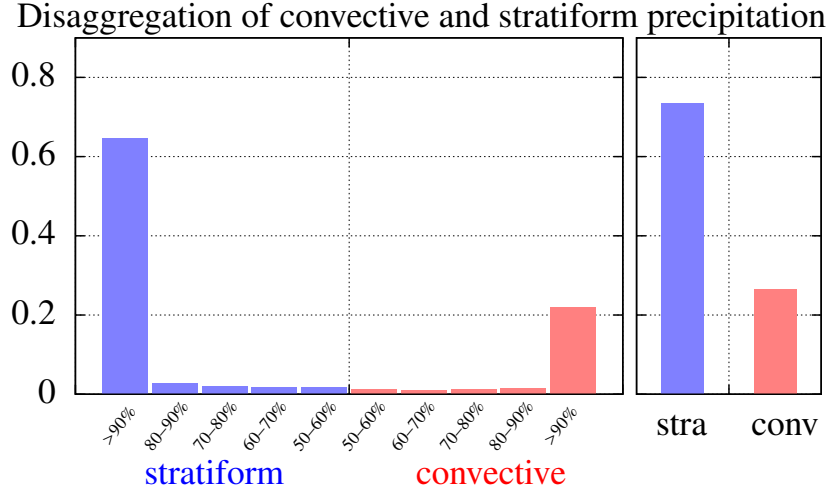


Figure 2.4.: Overview of the total disaggregation of convective (red) and stratiform (blue) precipitation. The intervals specify the predominant percentage of stratiform and convective precipitation, respectively. The graphic to the right shows the total identified stratiform (stra) and convective (conv) percentages.

the numerical codes to *convective* and *stratiform* cloud types (see Tab. A.3 in the appendix).

Since the investigations of this work refer to daily precipitation, each day had to be declared as convective or stratiform based on the SYNOP reports. Whenever a thunderstorm was recorded (present weather codes 90-99) this day was automatically recorded as convective day. In all other cases the convective and stratiform precipitation amounts of each day were summed up and the dominating type of precipitation for the respective day was adopted. For example, on June 1st 2016 46mm precipitation was measured at the rain gauge Cologne/Bonn. According to the present weather codes, 31mm (68%) of this were convective precipitation and the remaining 15mm (32%) were classified as stratiform precipitation (see Fig. 2.3). Because more convective than stratiform precipitation was recorded, this day was identified as a convective day with a total amount of 46mm.

The classification into stratiform and convective rain was by 56% based on hourly, by 24% on 6-hour, by 14% on 12-hour and by 6% on 24-hour precipitation data. In most cases the classification into convective and stratiform days was easy to determine: In more than 80% of the cases the precipitation of one day was to 90% of only one type of precipitation (Fig. 2.4). In 18% of the days it was not possible to clearly define the type of precipitation (mixed, green line in Fig. 5

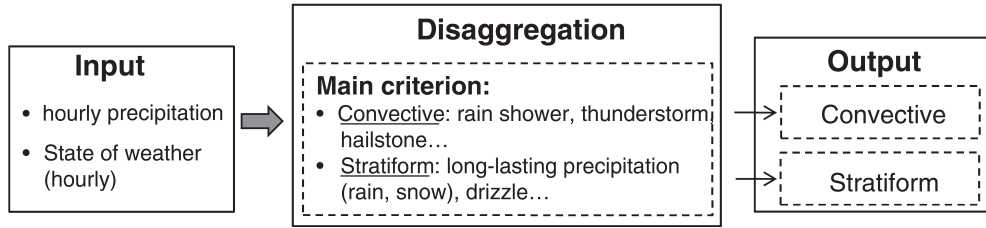


Figure 2.5.: Schematic illustration of the primary used algorithm.

bottom). This was mainly due to missing SYNOP-data for several hours of the day. For the successfully classified days it was found that 73.6% of the days had primary stratiform precipitation while on the remaining 27.4% the precipitation was convective driven. This frequency of little more than 70% stratiform days per year remains approximately constant: no fluctuation or trend can be observed throughout the time period of 67 years (red curve in Fig. 2.2 bottom).

3. Development of a precipitation model

This chapter describes the development of a statistical precipitation model for Germany. Differing from previously developed statistical precipitation models (described in sec. 1.1) this model should not be based on just one station. Instead it should be able to realistically generate the occurrence and amounts of precipitation at any time and at any location in Germany. This should be implemented by distinguish between different types of precipitation, namely stratiform and convective precipitation. The main difference between stratiform and convective precipitation is driven by its seasonal dependency (see also sec. 1). However, there is also another mechanism that has an enhancing effect. It is called the orographic effect and causes the precipitation recorded in the German mountain regions (altitude $> 500\text{m}$, annual average: 1246mm) to be in average 60% higher than elsewhere (altitude $< 500\text{m}$, annual average: 787mm ; compare Fig. 3.1).

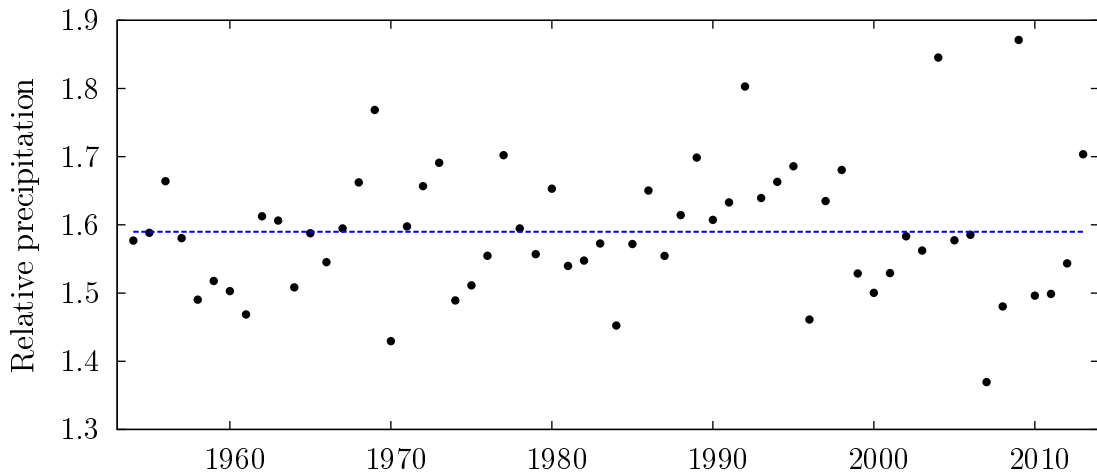


Figure 3.1.: Relative annual precipitation of German mountain stations (altitude $> 500\text{m}$) compared to gauges in flatter terrain (altitude $< 500\text{m}$). Dots show the annual values and dotted line indicates the average.

The orographic precipitation is especially important in the mountains: Orographic precipitation happens when an air mass is forced towards a mountain range and is rapidly lifted upwards. This causes moisture to cool down and create precipitation in the form of rain or snow. Thus, orographic precipitation is mainly dependent on the altitude of a rain gauge. In fact, it also depends on whether precipitation occurs on a windward or a leeward side of a mountain [Smith (1979)] but since collected data do not provide information about the mountain profile these influences are not taken into account. Hence, in this model the orographic effect is considered by the elevation of a station.

This model utilizes the two input parameters t (= day of the year) and h (= altitude of the station) to represent the main characteristics of the precipitation anywhere and at any season in Germany. In the following chapters the t - and h -dependence of both the precipitation occurrence and the precipitation amount will be investigated to develop the occurrence and the amount process respectively.

3.1. Development of the occurrence process

To discriminate between stratiform and convective precipitation the separation method described in sec. 2.4 is applied to the dataset SYNOP. Thus our model has three different states: either no precipitation (dry state), stratiform precipitation (stratiform state) or convective precipitation (convective state). Previous studies showed that a first order Markov chain is mostly superior to Markov chains with higher orders (see section 1.1.1). Therefore a simple first order Markov chain with three states is used.

Figure 3.2 shows a schematic of the state diagram of the Markov chain used. To estimate the state transitions a similar approach as the one of Woolhiser and Pegram (1979) [Woolhiser and Pegram (1979); Woolhiser and Roldan (1982); Stern and Coe (1984)] is used: The set of data is separated into groups of twelve consecutive days and for each group the state transitions are calculated. In a next step a Fourier series is fitted to the 30 groups of data ($360/12 = 30$) to get a time response model (seasonality of the model). The Fourier series is given by

$$U_{Fourier}(x) = A_0 + \sum_{l=1}^k A_l \cos \frac{2\pi lx}{365} + B_l \sin \frac{2\pi lx}{365} \quad (3.1)$$

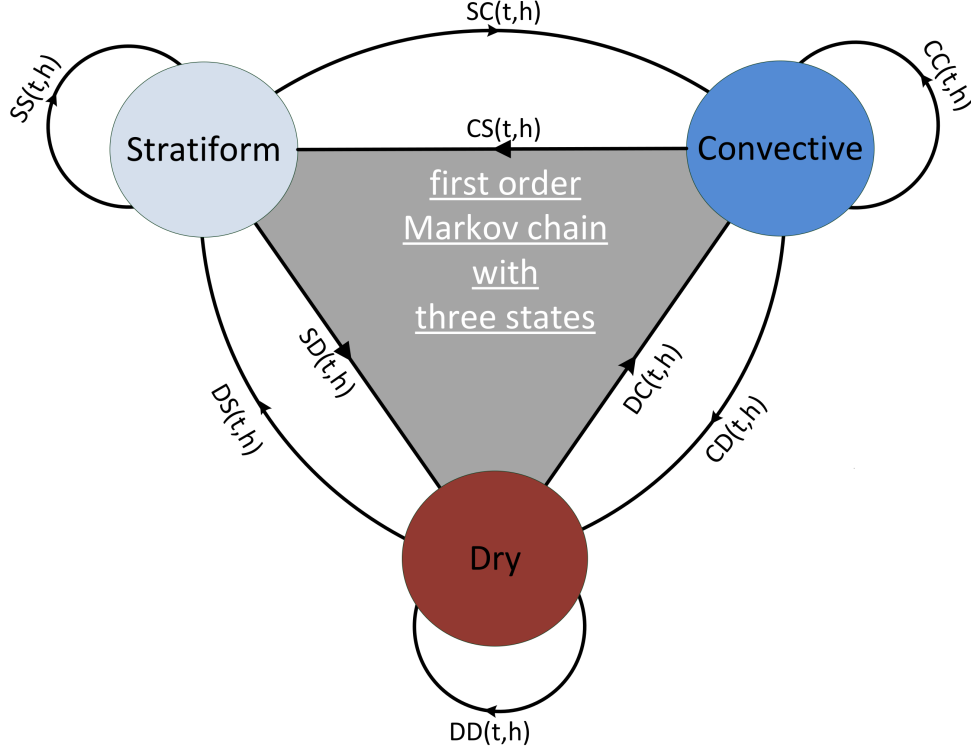


Figure 3.2.: Schematic illustration of the used first order Markov chain with the three states stratiform (S, light blue), convective (C, blue) and dry (D, red). Arrows show the state transitions.

where k is the number of harmonics and A_0 , A_l and B_l are the Fourier coefficients. For most state transitions one harmonic ($k = 1$) was enough to get good fit results and B_1 was close to zero. For these cases a cosines-function was fitted to the group of data which is given by

$$U_{cos}(t, h) = m_0(h) + m_1(h) \cos \left(\frac{2\pi t}{365} + m_2 \right). \quad (3.2)$$

For the probabilities to remain in a dry state (state transitions DD) and to change from a dry state into a convective state (state transition DC) two harmonics ($k = 2$) were needed. For these two cases a Fourier series of the following form was used:

$$U_{fr}(t, h) = n_0(h) + n_1(h) \cos \frac{2\pi t}{365} + n_2(h) \sin \left(\frac{2\pi t}{365} \right) + n_3 \cos \left(\frac{4\pi t}{365} \right) + n_4 \sin \left(\frac{4\pi t}{365} \right). \quad (3.3)$$

The coefficients of the cosine- ($m_0(h)$, $m_1(h)$ and $m_2(h)$) as well as those of the Fourier-series ($n_0(h)$, $n_1(h)$, $n_2(h)$, $n_3(h)$ and $n_4(h)$) in eq. 3.2 and eq. 3.3 respec-

Table 3.1.: Overview of the used altitude ranges and the corresponding amounts of data within these classes.

Altitude range	Data	Altitude class
0-200m	45%	lowland
200-500m	29%	hill area
500-1000m	22%	lower mountains
1000-1500m	3%	pre-alps
1500-3000m	1%	high mountains

tively are dependent on h to include a dependency of the altitude. To get information about the altitude dependency the data is additionally separated into five altitude classes: 0-200m (lowland), 200-500m (hill area), 500-1000m (lower mountains), 1000-1500m (pre-alps) and 1500-3000m (high mountains) (see Tab. 3.1). The dependencies of the altitude were estimated by fitting the cosines- (eq. 3.2) or the Fourier-series (eq. 3.3) respectively to each altitude class of a state transition. Regression analysis was then used to get approximations of the altitude dependency of each coefficient.

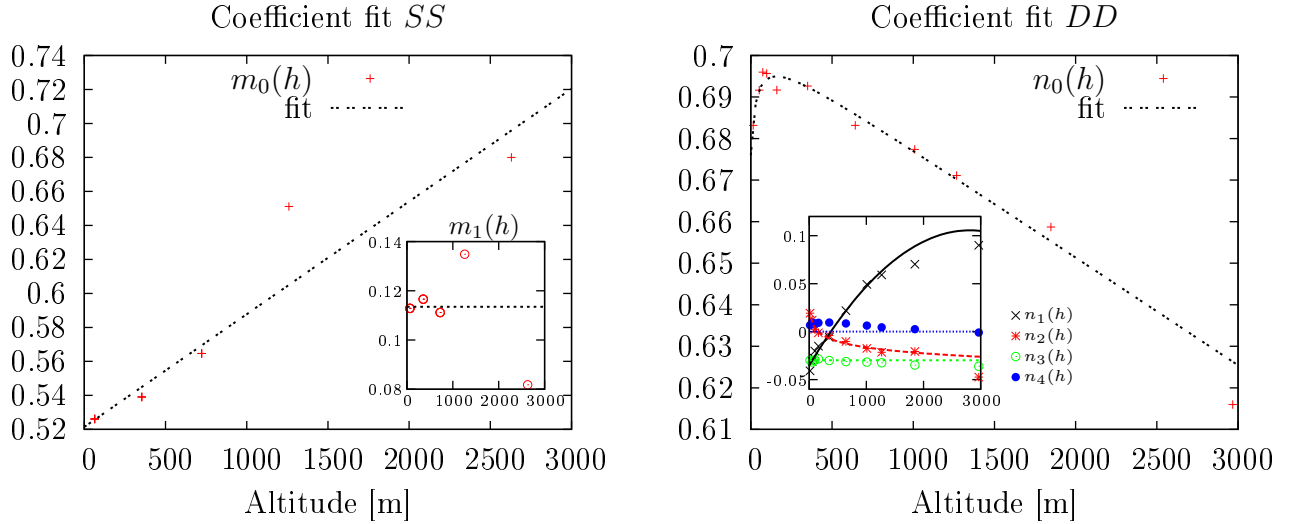


Figure 3.3.: Examples of the coefficient calculations for the state transition *stratiform-stratiform* (SS, left) and the state transition *dry-dry* (DD, right). Symbols show the estimated coefficients for given altitude classes based on the dataset SYNOP (state transition SS) and dataset DWD_{all} (state transition DD), respectively. Lines show the used fits of the coefficients. For the state transition DD in total eleven instead of five altitude classes were used, as DWD_{all} provides a much larger set of data.

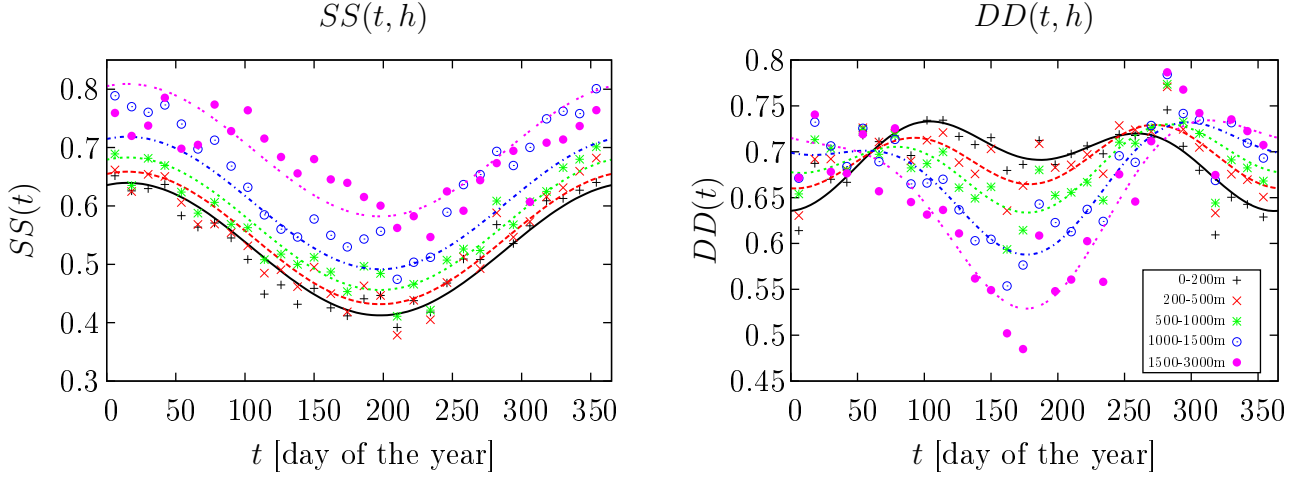


Figure 3.4.: Used altitude and time dependency of the transitions SS (left) and DD (right). Symbols show the five altitude classes used and lines their best fits, respectively.

Fig. 3.3 shows the regression analysis for the state transitions of a stratiform day followed by another stratiform day (SS , left) and for the state transitions of a dry day followed by another dry day (DD , right). The coefficient m_0 of the state transition SS is well approximated by a linear regression while for the coefficient m_1 a constant is assumed (see Fig. 3.3 left). It should be noted that each coefficient was weighted by the amount of data available on the respective altitude range (see Tab. 3.1). Thus a good fit result for the altitude classes *lowland*, *hill area* and *lower mountains* is of major importance while a good fit result for the altitude classes *pre alps* and *high mountains* is less important. This is why the coefficient m_0 of the *pre-alps*-altitude class is relatively far away from the regression line (see Fig. 3.3 left). Therefore, it seems like the probability to remain in a stratiform state in the wintertime is underestimated with the cosine-function of the *pre-alps*-altitude-class (see Fig. 3.4 left). Furthermore the coefficient $m_2(h)$ was set to be independent of the altitude and equal to $m_2 = -0.26$. This is because m_2 is a phase shift in eq. 3.2 and a phase shift of $m_2 = -0.26$ is equal to a shift of 15 days ($\frac{2\pi \cdot 15}{365} = 0.26$). With this shift of 15 days the cosine-function has its maximum on the 15th of January and its minimum on the 16th of July which is in the middle of the meteorological winter or summer season respectively.

For the state transition DD a linear regression worked also fine for the coefficient $n_0(h)$ (see Fig. 3.3 right). Note that the dataset DWD_{all} was used for the studies of the altitude and time dependency of the state transition DD as it provides a

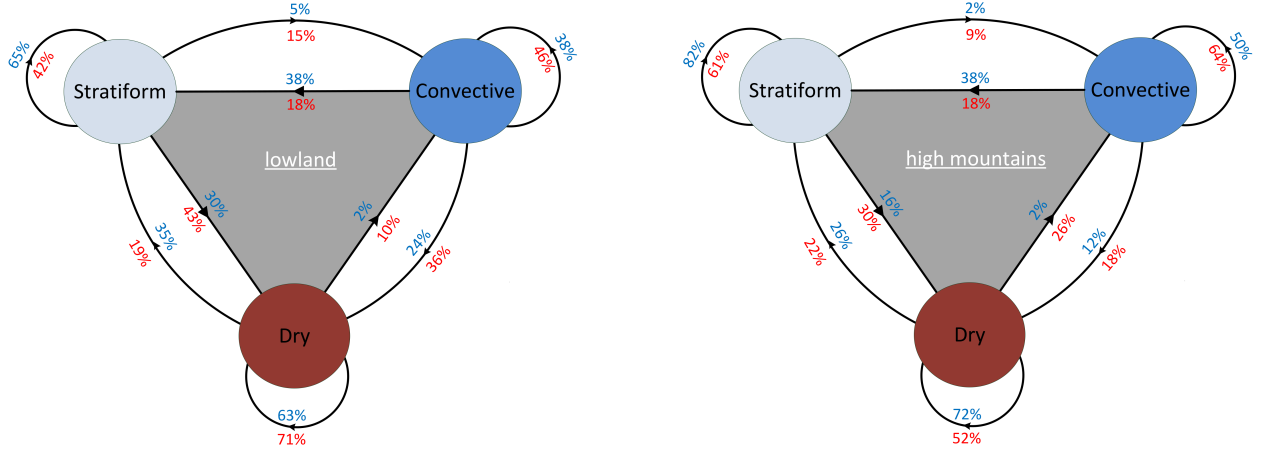


Figure 3.5.: Same as Fig. 3.2, but with given transition probabilities for a lowland station (elevation: 50m, left) and a station in the high mountain altitude class (elevation: 2,000m). The numbers show the transition probabilities for January 15th (blue) and July 16th (red).

much larger set of data (compare Tab. 2.1). This huge set of data was separated into eleven instead of five classes of altitudes and more complicated non-linear regressions were used. The coefficient $n_0(h)$ for example is actually a combination of two linear regressions and $n_2(h)$ appears to have a logarithmic shape. Moreover the coefficient $n_4(h)$ was set to $n_4 = 0$ as its values were very small. A summary of the time- and altitude dependencies for all state transitions used in the occurrence process of the model is given in the appendix (Tab. C.1).

Fig. 3.5 shows the results of the occurrence model for the winter season (15th of January, blue numbers) and the summer season (16th of July, red numbers) as well as for a station on the lowland area (at 50m altitude, left) and a station on a high mountain (at 2000m altitude, right). There is a significant difference in the occurrence of precipitation between these different altitudes and seasons. For example in the winter season the probability of a day with stratiform precipitation followed by a dry day is 30% at a rain gauge at 50m altitude whereas it is just 16% at a rain gauge at 2000m altitude. In the summer season the probability is 43% and 30% respectively. Moreover it is more likely to get two consecutive dry days in the winter season at high mountains than on the lowland area (72% vs. 63%) and vice versa in the summer season (52% vs. 71%).

3.2. Development of the amount process

This chapter describes the development of another distribution for the generation of precipitation amounts. In particular the focus was on good realizations of the highest precipitation amounts. For the development of the distribution the complementary cumulative distribution function (CCDF, also called tail distribution) was used which is defined as

$$\bar{F}(x) = P(X > x), \quad (3.4)$$

where x is the amount of precipitation and X is the observed value. In contrast to the classical cumulative distribution function (CDF), which is defined as

$$F(x) = P(X \leq x). \quad (3.5)$$

The CCDF gives the probability of a random variable being *above* instead of being *under* a particular level. Thus the CCDF and the CDF are connected by

$$\bar{F}(x) = 1 - F(x). \quad (3.6)$$

For the fitting of the tail Clauset et al. (2009) recommend to work with the CCDF as its visual form is more robust than its PDF. The probability density function (PDF) and the CCDF are connected by

$$f(x) = \frac{d}{dx}F(x) = \frac{d}{dx} [1 - \bar{F}(x)] . \quad (3.7)$$

3.2.1. Development of a distribution for precipitation

The generation of very high precipitation amounts has been the main problem of statistical precipitation models in the past. As discussed in chapter 1.1 the frequency of extreme precipitation is much lower in statistical precipitation models than observed in nature. Since a realistic occurrence of very high precipitation amounts is essential for an application to records, the initial focus was on the tail of the distribution.

Fitting of the tail

On a log-log plot of the CCDF the tail becomes roughly straight for all sets of data. This could be an indication for the tail of the precipitation distribution to be well described by a power law. Indeed, a recently published paper by Yalcin et al. (2016) suggested that daily precipitation can be described by a power law approach.

In general, the PDF of a power law is given by [Clauset et al. (2009)]

$$f_{pl}(x) = \frac{\alpha - 1}{x_{min}} \left(\frac{x}{x_{min}} \right)^{-\alpha}, \quad x_{min} \geq x \quad (3.8)$$

and the CCDF is given by [Clauset et al. (2009)]

$$\bar{F}_{pl}(x) = \int_x^\infty f_{pl}(x') dx' = \left(\frac{x}{x_{min}} \right)^{(-\alpha+1)} \quad (3.9)$$

where x_{min} is the minimum possible precipitation amount and for this study can be interpreted as the beginning of the tail (beginning of the validity of the power law) whereas α is the scaling exponent.

To estimate the scaling parameter α the *method of maximum likelihood* is used. The maximum likelihood estimator (MLE) for $\hat{\alpha}$ is given by Clauset et al. (2009)

$$\hat{\alpha} = 1 + n \left[\sum_{i=1}^n \ln \frac{x_i}{x_{min}} \right]^{-1} \quad (3.10)$$

where x_i ($i = 1, 2, \dots, n$) are the observed precipitation amounts with $x_i \geq x_{min}$. For this equation the value x_{min} , which is still unknown, is needed. One method to estimate \hat{x}_{min} as well as $\hat{\alpha}$ is to use eq. 3.10 for all possible \hat{x}_{min} and to plot the estimated $\hat{\alpha}$ as a function of \hat{x}_{min} . The values of \hat{x}_{min} and $\hat{\alpha}$ can now be visually estimated at the point where the value of $\hat{\alpha}$ appears to be stable.

In addition to the visual estimation of \hat{x}_{min} , \hat{x}_{min} can be calculated with an approach proposed by Clauset et al. (2007) which uses the KS statistic to quantify the distance between the CCDF of the observations $\bar{S}(x)$ and the CCDF of the best fit of the data $\bar{F}_{pl}(x)$ for values of x greater than x_{min} :

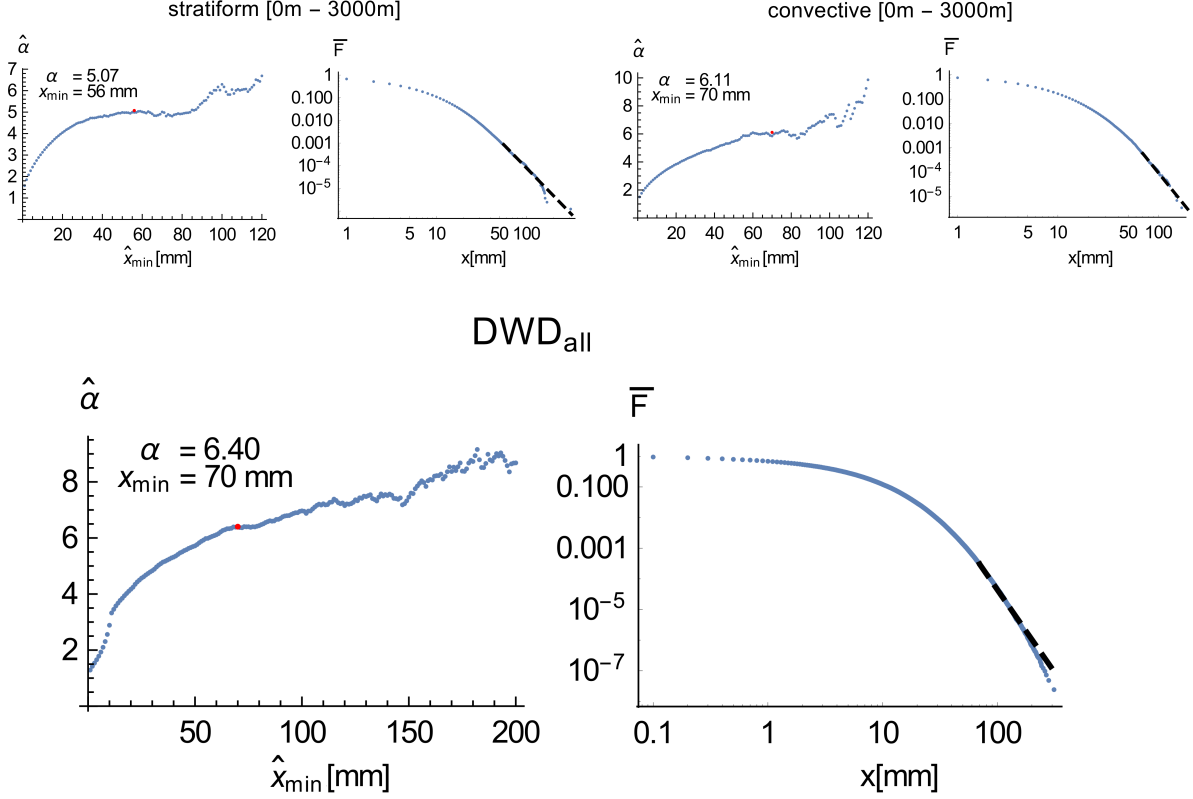


Figure 3.6.: Estimated $\hat{\alpha}$ as a function of \hat{x}_{min} (left charts) and CCDF (right charts) for both the stratiform (top left) and convective (top right) precipitation of dataset SYNOP as well as for dataset DWD_{all} (bottom). Dashed line in the CCDF shows the fit with the calculated values of eq. 3.11. Values are provided in the charts and visualized as red dots in the α -vs.- x_{min} plot.

$$D = \max_{x \geq x_{min}} |\bar{S}(x) - \bar{F}_{pl}(x)|. \quad (3.11)$$

The value of x_{min} that minimizes D is the estimated \hat{x}_{min} . In doing so, the values of \hat{x}_{min} are $\hat{x}_{min} = 70$ mm for the dataset DWD_{all} as well as for the convective precipitation of the dataset SYNOP and $\hat{x}_{min} = 56$ mm for the stratiform part of the dataset SYNOP. The estimated value of $\hat{\alpha}$ lies between $\hat{\alpha} = 5.1$ and $\hat{\alpha} = 6.4$ (see red dot in Fig. 3.6). Visually the highest precipitation amounts seem to follow approximately a power law. However, a closer look at the results of dataset DWD_{all} (bottom plot in Fig. 3.6) suggests that a power law is probably not the true distribution from which the highest precipitation amounts were drawn (the tail is not exactly straight on the log-log-plot). But even if a more detailed study (not yet done) would lead to a rejection of the power law a power-law-like distribution should be implemented

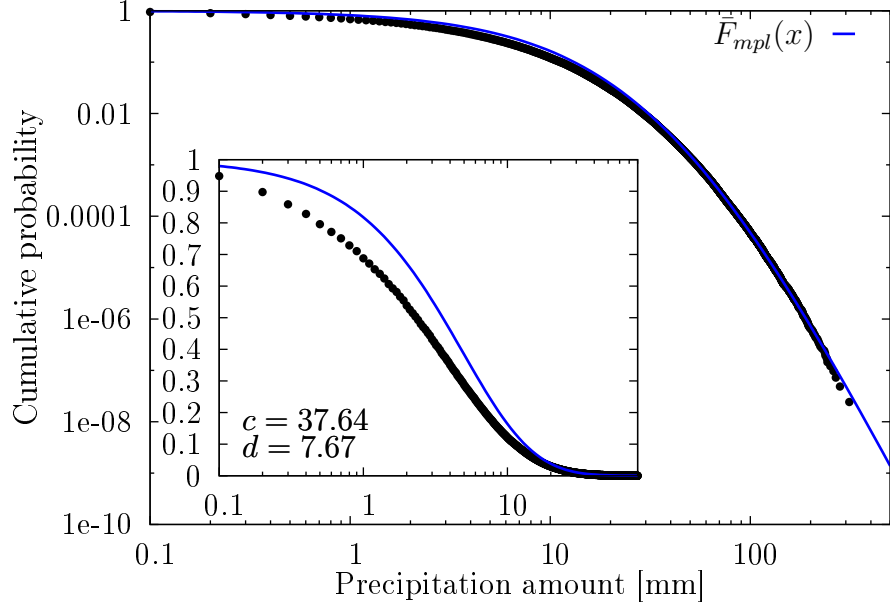


Figure 3.7.: Complementary cumulative distribution function of dataset DWD_{all} (black dots) and its best fit of $\bar{F}_{mpl}(x)$ (blue curve). Numbers in the chart indicate the parameter values of the fit.

in the amount process as it seems to improve the generation of high precipitation amounts significantly.

The precipitation model should also be able to simulate precipitation amounts smaller than x_{min} . For that the CCDF \bar{F}_{pl} (eq. 3.9) is modified so that \bar{F}_{pl} is also valid for values smaller than x_{min} . Let \bar{F}_{mpl} be the CCDF of the modified power law. Because of the definition of the CCDF it is known that $\bar{F}_{mpl}(x = 0) = 1$. A possible form of \bar{F}_{mpl} could then be

$$\bar{F}_{mpl} = \left(1 + \frac{x}{c}\right)^{-d}. \quad (3.12)$$

Note that $c \neq x_{min}$ and $d \neq \alpha$.

The best fit of eq. 3.12 to the precipitation data leads to good results for the tail (see log-log-plot of Fig. 3.7) but still overestimates amounts of $x \leq \hat{x}_{min}$, especially for values close to zero (see log-plot of Fig. 3.7).

Combining Weibull distribution and power law

So far a distribution which generates good results for the tail but non-satisfying results for small precipitation amounts was developed. In contrast common precipitation models use distributions like the Gamma distribution or, less common, the Weibull distribution which fit well for small precipitation amounts but underestimate the tail of the distribution (see section 1.1). A distribution which follows a Gamma- or Weibull distribution for small values and merges with a power law for higher values might be a distribution which describes the whole range of precipitation amounts. To realize this idea it can be written:

$$\bar{F}_{com}(x) = \gamma(x)\bar{F}_A(x) + (1 - \gamma(x))\bar{F}_B \quad (3.13)$$

where \bar{F}_{com} is the CCDF of a combination of the two distributions \bar{F}_A and \bar{F}_B and $\gamma(x)$ is a x -dependent weighting of \bar{F}_A . For small precipitation amounts (small values of x) \bar{F}_{com} should mainly be described by \bar{F}_A and for high precipitation amounts (high values of x) \bar{F}_{com} should mainly be described by \bar{F}_B . Therefore

$$\bar{F}_{com}(x = 0) \stackrel{!}{=} \bar{F}_A(x) \quad (3.14)$$

and

$$\bar{F}_{com}(x = \infty) \stackrel{!}{=} \bar{F}_B(x) \quad (3.15)$$

hence it follows

$$\gamma(x = 0) \stackrel{!}{=} 1 \quad (3.16)$$

and

$$\gamma(x = \infty) \stackrel{!}{=} 0. \quad (3.17)$$

This condition is satisfied by

$$\gamma(x) = \frac{1}{1 + \gamma'(x)}. \quad (3.18)$$

Next it is assumed

$$\gamma'(x) = \frac{x}{\epsilon} \quad (3.19)$$

which means that for $x = \epsilon$ the weighting for both \bar{F}_A and \bar{F}_B is the same ($\bar{F}_A = \bar{F}_B = \frac{1}{2}$). In the following, let $\bar{F}_A = \bar{F}_w$ be the CCDF of the Weibull distribution and $\bar{F}_B = \bar{F}_{mpl}$ be the CCDF of the modified power law in the previous section (see eq. 3.12):

$$\begin{aligned} \bar{F}_{w+pl}(x) &= \frac{1}{1 + \frac{x}{\epsilon}} \bar{F}_w(x) + \left(1 - \frac{1}{1 + \frac{x}{\epsilon}}\right) \bar{F}_{mpl} \\ &= \frac{1}{1 + \frac{x}{\epsilon}} e^{-\left(\frac{x}{b}\right)^a} + \left(1 - \frac{1}{1 + \frac{x}{\epsilon}}\right) \left(1 + \frac{x}{c}\right)^{-d} \end{aligned} \quad (3.20)$$

In principle ϵ and c do have a similar physical meaning: both are an indicator of the beginning of the tail of the distribution. It seems to be fairly reasonable to set $\epsilon = c$. Which reduces eq. 3.20 to

$$\begin{aligned} \bar{F}_{w+pl}(x) &= \frac{1}{1 + \frac{x}{c}} e^{-\left(\frac{x}{b}\right)^a} + \left(1 - \frac{1}{1 + \frac{x}{c}}\right) \left(1 + \frac{x}{c}\right)^{-d} \\ &= \frac{c}{c+x} e^{-\left(\frac{x}{b}\right)^a} + \left(1 - \frac{c}{c+x}\right) \left(\frac{c+x}{c}\right)^{-d} \\ &= \frac{ce^{-\left(\frac{x}{b}\right)^a} + x \left(\frac{c+x}{c}\right)^{-d}}{c+x}. \end{aligned} \quad (3.21)$$

This expression is the final CCDF for the precipitation model used. The CDF is simply given by:

$$\begin{aligned} F_{w+pl}(x) &= 1 - \bar{F}_{w+pl}(x) \\ &= 1 - \frac{ce^{-\left(\frac{x}{b}\right)^a} + x \left(\frac{c+x}{c}\right)^{-d}}{c+x} \end{aligned} \quad (3.22)$$

and the PDF is given by the first derivative of F_{w+pl} (eq. 3.22):

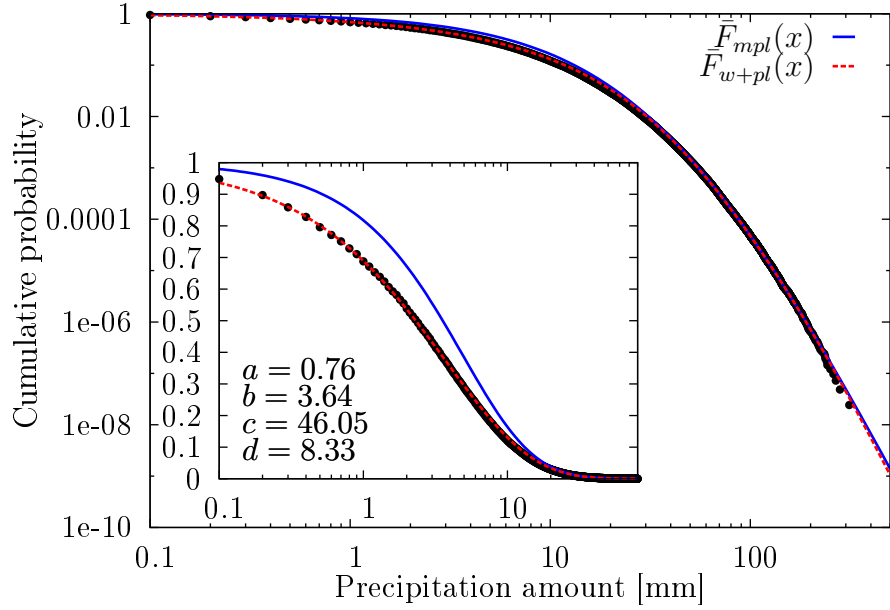


Figure 3.8.: Same as Fig. 3.7, but additionally with \bar{F}_{w+pl} (red dashed line). Numbers in the chart indicate the parameter values of the fit of \bar{F}_{w+pl} .

$$\begin{aligned}
f_{w+pl}(x) &= \frac{d}{dx} F_{w+pl}(x) \\
&= \frac{d}{dx} \left[1 - \frac{-ce^{-\left(\frac{x}{b}\right)^a} - x \left(\frac{c+x}{c}\right)^{-d}}{c+x} \right] \\
&= \frac{(c+x) \frac{d}{dx} \left[-ce^{-\left(\frac{x}{b}\right)^a} - x \left(\frac{c+x}{c}\right)^{-d} \right] - \frac{d}{dx} [x+c] \left(-ce^{-\left(\frac{x}{b}\right)^a} - x \left(\frac{c+x}{c}\right)^{-d} \right)}{(c+x)^2} \\
&= \frac{(c+x) \left(\frac{ac}{b} \left(\frac{x}{b}\right)^{a-1} e^{-\left(\frac{x}{b}\right)^a} - \left(\frac{c+x}{c}\right)^{-d} + \frac{dx}{c} \left(\frac{c+x}{c}\right)^{-d-1} \right) - \left(-ce^{-\left(\frac{x}{b}\right)^a} - x \left(\frac{c+x}{c}\right)^{-d} \right)}{(c+x)^2} \\
&= \frac{(c+x) \left(\frac{ac}{x} \left(\frac{x}{b}\right)^a e^{-\left(\frac{x}{b}\right)^a} + \left(\frac{dx}{c+x} - 1\right) \left(\frac{c+x}{c}\right)^{-d} \right) - \left(-ce^{-\left(\frac{x}{b}\right)^a} - x \left(\frac{c+x}{c}\right)^{-d} \right)}{(c+x)^2} \\
&= \frac{\left(\frac{ac(c+x)}{x} \left(\frac{x}{b}\right)^a e^{-\left(\frac{x}{b}\right)^a} + (dx - (c+x)) \left(\frac{c+x}{c}\right)^{-d} \right) - \left(-ce^{-\left(\frac{x}{b}\right)^a} - x \left(\frac{c+x}{c}\right)^{-d} \right)}{(c+x)^2} \\
&= \frac{\left(\frac{a(c+x)}{x} \left(\frac{x}{b}\right)^a + 1 \right) ce^{-\left(\frac{x}{b}\right)^a} + (dx - (c+x) + x) \left(\frac{c+x}{c}\right)^{-d}}{(c+x)^2} \\
&= \frac{(a(c+x) \left(\frac{x}{b}\right)^a + x) ce^{-\left(\frac{x}{b}\right)^a} + x(dx - c) \left(\frac{c+x}{c}\right)^{-d}}{x(c+x)^2}.
\end{aligned} \tag{3.23}$$

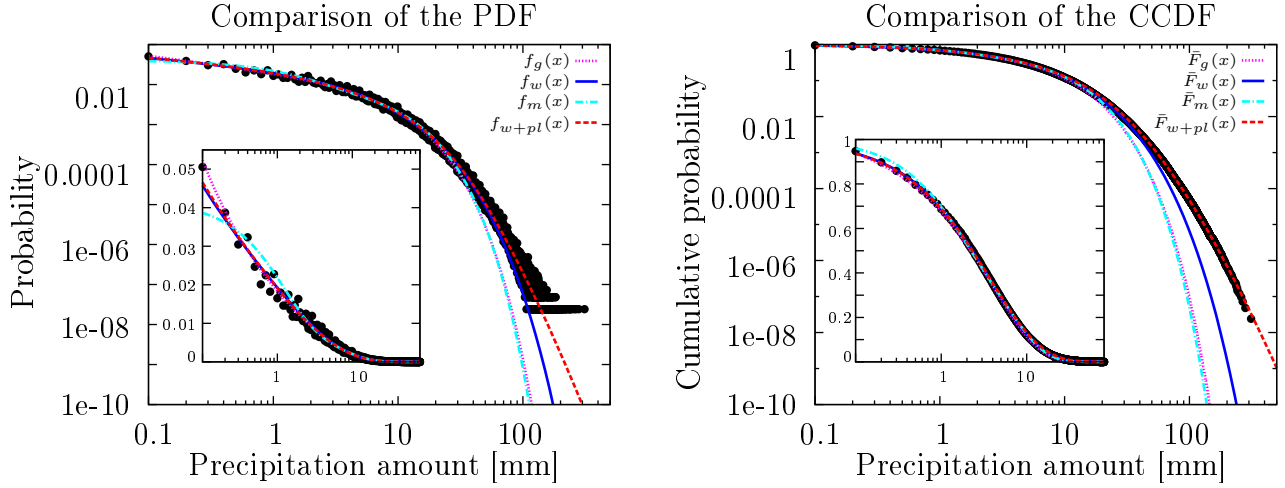


Figure 3.9.: Same as Fig. 1.1 but with $f_{w+pl}(x)$ (left, dashed red line) and \bar{F}_{w+pl} (right, red dashed line).

Fig. 3.8 shows the best fit of $\bar{F}_{w+pl}(x)$ (eq. 3.21, dotted line) and $\bar{F}_{mpl}(x)$ (eq. 3.12, solid line) for the dataset DWD_{all} . $\bar{F}_{w+pl}(x)$ shows an almost perfect fit for the smallest precipitation amounts and still a very good fit in the tail.

In comparison to previously developed amount processes $\bar{F}_{w+pl}(x)$ improves the generation of high precipitation amounts enormously (see Fig. 3.9).

3.2.2. Time and altitude dependence

Like the occurrence process (see sec. 3.1) also the amount process should be dependent on the time of the year t and the altitude h . After the disaggregation of the stratiform and convective precipitation (method described in sec. 2.4) every parameter was estimated for both types of precipitation at each station and plotted against the altitude of the stations (Fig. 3.10). For the parameters a and b an altitude dependency is clear noticeable whereas the altitude dependencies of the parameters c and d are not obvious. However, for all four parameters linear regressions were used to represent the altitude dependency.

To study an additional t -dependency every month and four altitude classes (0-200m, 200m-500m, 500m-1000m and 1000-3000m, same as in sec. 3.1) were merged. At first the t -dependency was only detectable for the parameter b (see Fig. 3.11). Again, for the fit the data points were weighted by the respective data availability. Especially for the winter months the data availability of convective precipitation is

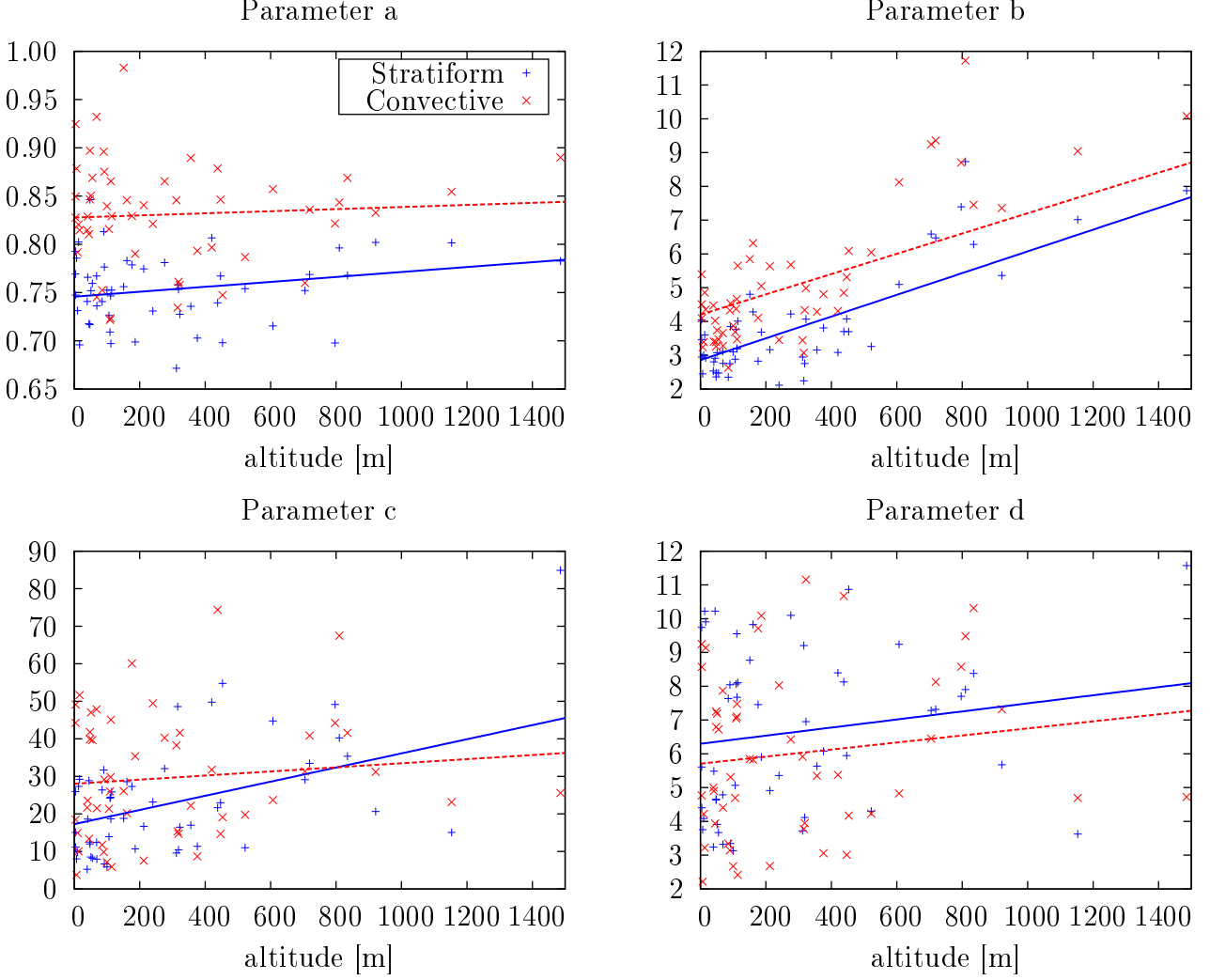


Figure 3.10.: Altitude dependencies of the parameters a , b , c and d . Crosses show the estimated values of convective (red) and stratiform (blue) precipitation for each station of dataset SYNOP. Red and blue lines show the best linear fits, respectively. For display purposes, the x-axis of the charts have been cut at 1,500m (Zugspitze, 2964m, not shown).

very low (Fig. 3.11 right). However, first runs of the model showed higher values in the winter season than observed. To generate less high amounts in the winter season it seemed to be necessary to give the parameter d also a t -dependency (because d is the scaling parameter of the tail). This was realized by a cosine function giving d around 50% higher values in winter time than during summer months (see Tab. C.2 in the appendix).

A summary of the time- and altitude dependencies of all four parameters of the used probability distribution is given in the appendix (Tab. C.2).

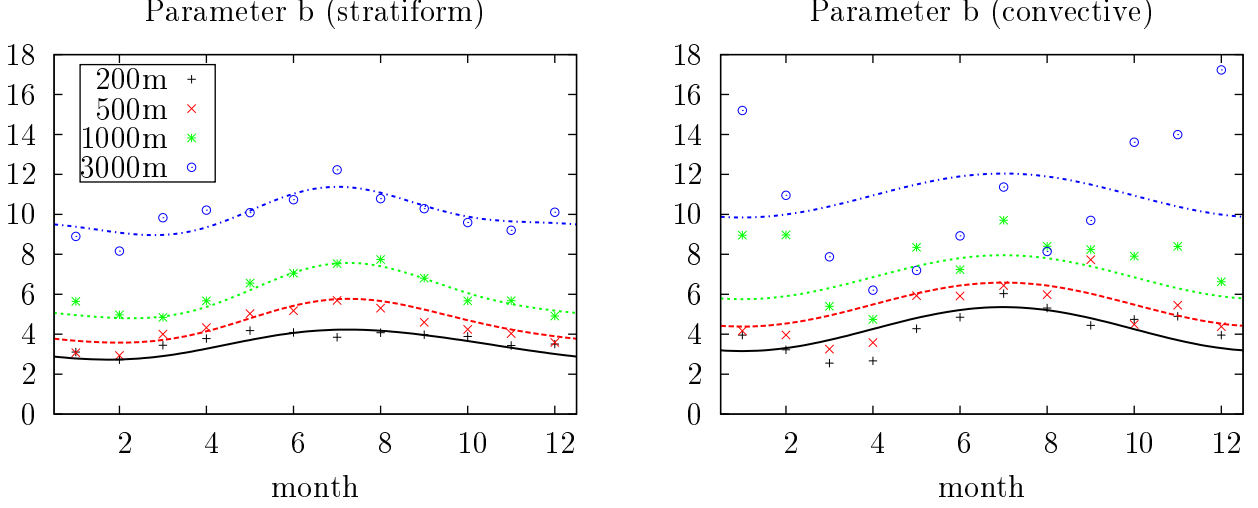


Figure 3.11.: Time- and altitude dependencies of parameter b for both stratiform (left) and convective (right) precipitation. Symbols show the altitude classes for each month. Lines show weighted fits of the data, respectively.

3.3. Taking account of variability in the occurrence process

The occurrence process described above is still too static – especially regarding the modelling of extreme dry and wet spells. Therefore, the model still requires a modification which represents the variability of the general weather situation. In the model this is implemented by using a confined Gaussian random walk. This is a random walk which generates the next step by using a normal distribution with the standard deviation σ . Additionally, the range of values of the random walk is limited by two fixed boundaries. These boundaries have been chosen in a manner that they meet the condition of all state transitions to remain between 0 and 1. As a result, the random walk determines a shift Δ which modifies the transition probabilities as follows:

$$\widetilde{XX} = XX + \Delta_{wet/dry} \quad (3.24)$$

for the probabilities of remaining in a state and

$$\widetilde{XY} = XY - \Delta_{wet/dry} \frac{XY}{XY + XZ} \quad (3.25)$$

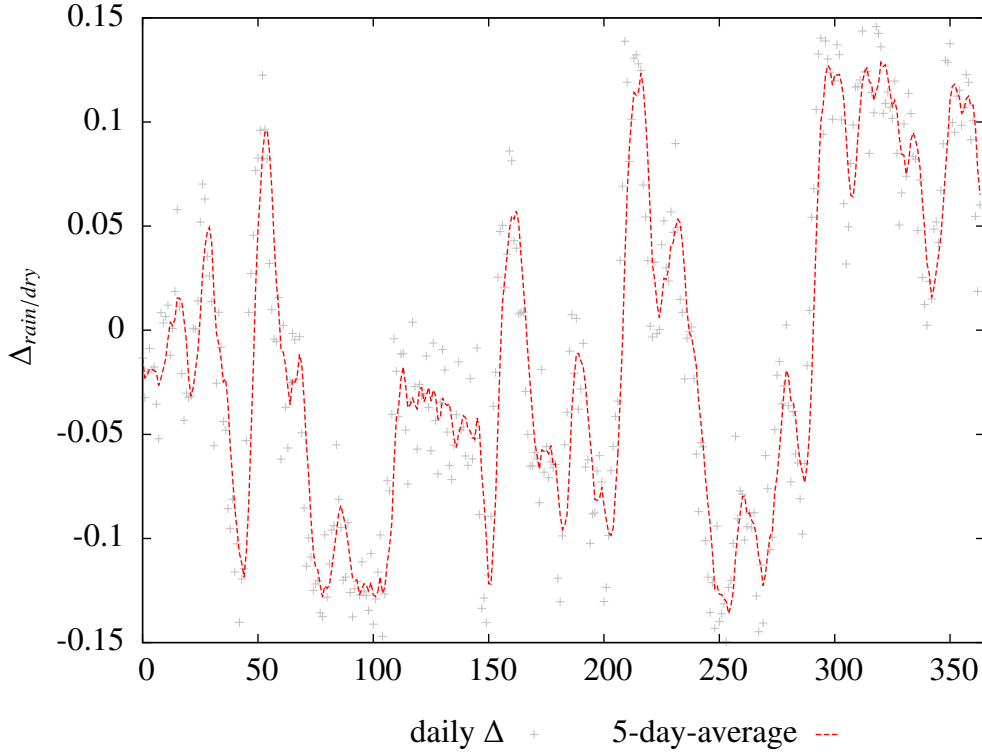


Figure 3.12.: Example of a confined random walk for an arbitrary year. Gray crosses show the daily values of the shift Δ and dashed red line its 5-day-average.

for all transition probabilities with a changing state.

Thereby, X, Y and Z each represent one of the three conditions stratiform (*S*), convective (*C*) or dry (*D*). Δ_{wet} was used for the initial states *stratiform* and *convective* as well as a Δ_{dry} for the initial state *dry*.

The thresholds and the standard deviation were varied and identified based on validations of the most extreme wet and dry spells. Thereby the thresholds ± 0.15 for both Δ and the standard deviation $\sigma = 0.03$ were identified. An example of such a random walk for a time interval of one year can be seen in Fig. 3.3: assuming that the random walk shown represents the value Δ_{dry} , this would mean that for a few days around day 100, the probability that a dry day would be followed by another day without precipitation is over 10% lower (unstable weather situation). Thereby, for a few days around day 300, the probability that it will remain dry is over 10% higher (stable weather situation). Overall, the effects balance each other out so that the mean value of all Δ_{dry} and Δ_{wet} values always results in zero.

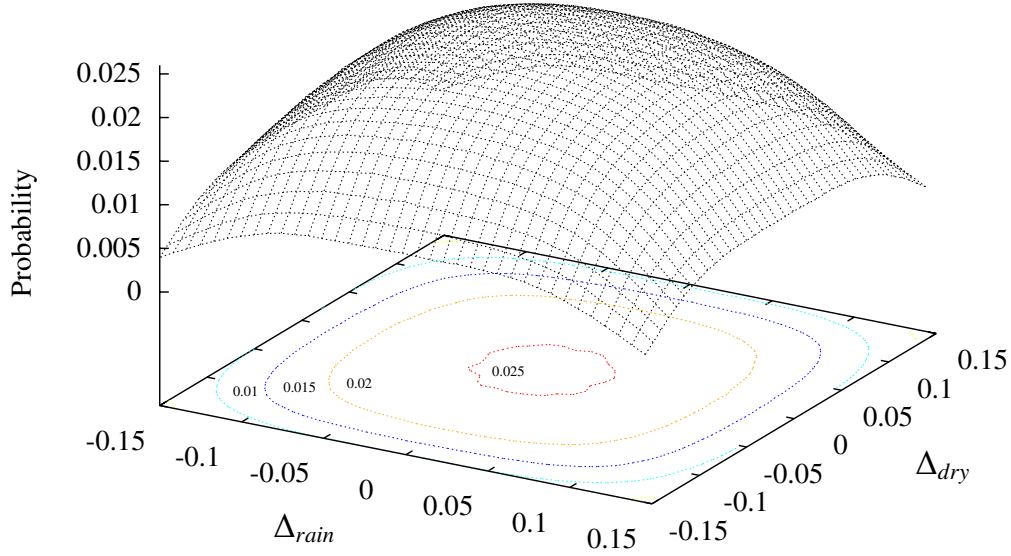


Figure 3.13.: 3-D plot of the probability distribution of the bivariate shifts Δ_{rain} and Δ_{dry} .

The random walk was used in the model in such a way that Δ_{dry} and Δ_{wet} were identified for each day and were valid for all stations. With this approach, occurrence process correlations between the stations were considered. A simulation for 100 million days demonstrates the effect of this modification of the occurrence process (see Fig. 3.13): On most days the random walk does not influence the occurrence process particularly strong. Nevertheless, there are also phases in which the probability of remaining in a *wet*- and/or *dry* state is significantly heightened or reduced (boundaries in the 3-D-plot in Fig. 3.13). Thereby phases of stable and unstable weather conditions are represented as they occur in reality.

3.4. Model results vs. observation

In this chapter, the results from the model will be compared with the actual observations. The topographic height of a station and the number of measurement days available for each station were used as input parameters to comparably reproduce the observations of the SYNOP and DWD_{all} datasets. On the one hand, a realistic occurrence of precipitation events (sec. 3.4.1), and on the other hand, a realistic simulation of high precipitation values was tested (sec. 3.4.2).

3.4.1. Precipitation occurrence

A comparison of modelled and observed dry and wet spell length is frequently carried out to test whether the occurrence process developed in the model describes reality

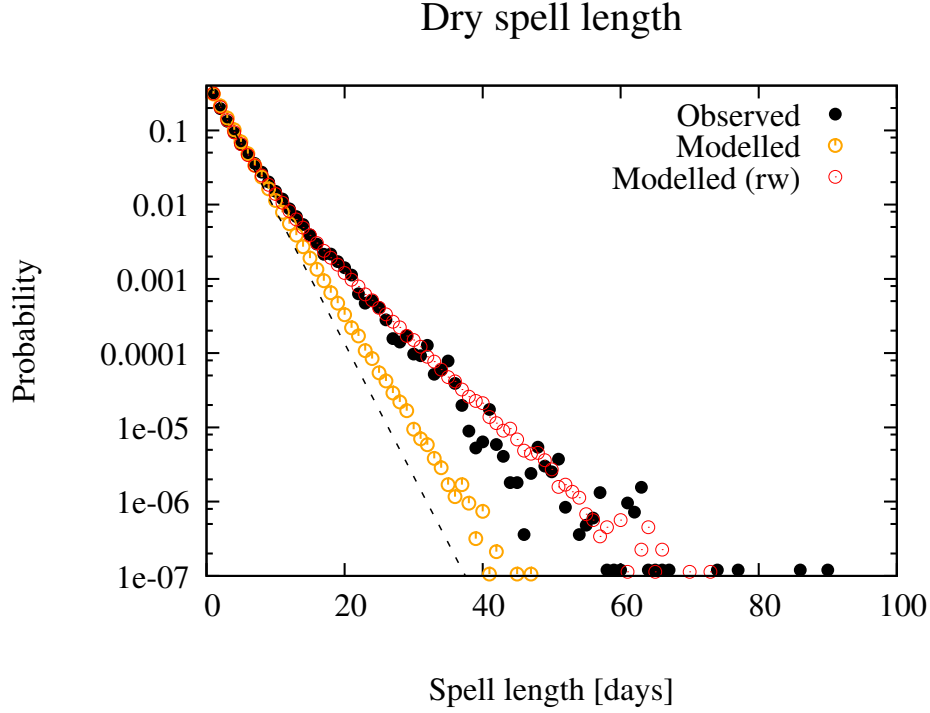


Figure 3.14.: Comparison of the modelled and observed dry spell length. Dashed line show the best fit of a geometric distribution. Dots show dry spell length like observed- (black), modelled- (orange) as well as modelled with confined random walk (red).

correctly [e.g. Wilks (1999)]. Simple (first order) Markov chains follow a geometric distribution in the simulation of spell length (see sec. 1.1). As previously noted, big under-estimates of the dry spell length occurred in previous studies using a Markov chain [e.g. Buishand (1978); Racsco et al. (1991)]. With the first order Markov chain used here, higher probabilities for very long dry spells were considered, than through a geometric distribution (see Fig. 3.14). This will mainly be due to the averaging of more than 300 stations and the consideration of the *time of year*. Nevertheless, the likelihood for a very long spell length to occur (e.g. 60 days), for dry as well as wet spells, continues to be underestimated. Therefore, the changes to the occurrence process described in the previous section were made. By taking into consideration this atmospheric variation, the dry spell lengths generated in the model are a good match for the observations (red circles in Fig. 3.14). The modelled wet spell length, however, still slightly underestimates the observations (purple circles in Fig. 3.15 left). However, by defining a wet day as a day on which more than 0.1mm (instead of $>0.0\text{mm}$) are observed, the modelled wet spell length fits the observations (purple circles in Fig. 3.15 right). Therefore, the slight underestimation on the wet spell length can be seen as negligible.

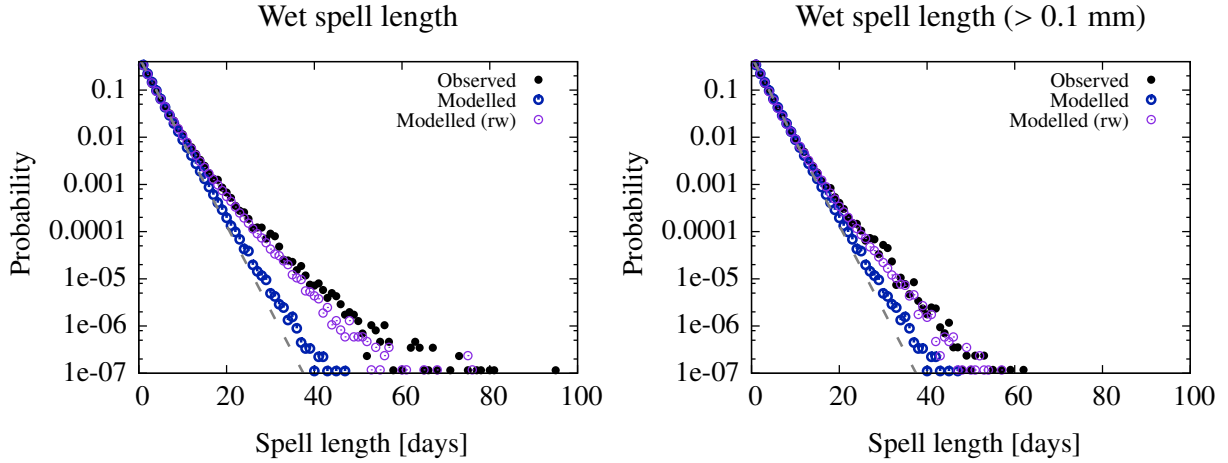


Figure 3.15.: Comparison of the modelled and observed wet spell length. Once for a spell length defined as $>0\text{mm}$ (left) and once for $>0.1\text{mm}$ (right). Dashed lines show the best fit of a geometric distribution. Dots show wet spell length like observed- (black), modelled- (blue) as well as modelled with confined random walk (purple).

3.4.2. Precipitation amounts

For the investigation of precipitation records, it appears especially important that very high precipitation is simulated close to reality. To test this, 30 simulations were created for each of the SYNOP and DWD_{all} datasets. As a next step, the highest modelled precipitation amounts were compared with the highest observed precipitation amounts (Fig. 3.16).

It is easy to see that the precipitation model uses the data from the SYNOP data set: The modelled precipitation amounts are well distributed about the origin to give an exact model of the observations (Fig. 3.16 left). If instead of the SYNOP dataset the very large DWD_{all} is used, an overestimation by the model of about 150mm is apparent. A possible cause could be the disregard of spatial correlations in the amount process of the model. A consideration of correlations appears to be necessary especially with a very high station density such as in the DWD_{all} dataset. However, for the purpose of use in this thesis the model is considered as well usable.

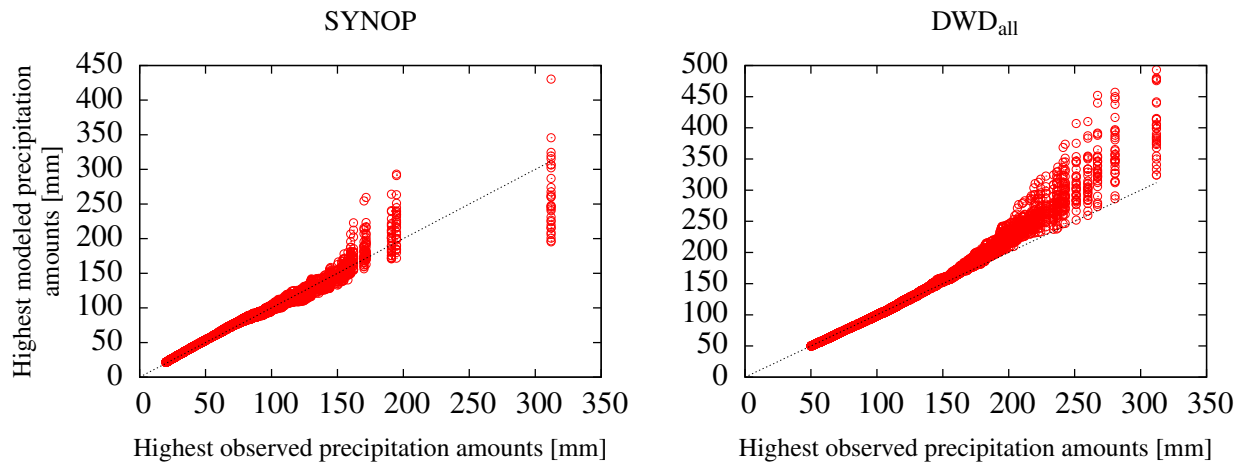


Figure 3.16.: Comparison of the highest observed precipitation amounts of dataset SYNOP (left) as well as dataset DWD_{all} (right) to the highest modelled precipitation amounts using 30 simulations, respectively.

4. Application to precipitation records

In this chapter, the precipitation model developed in section 3 will be used to prove whether the higher than expected mean record number $\tilde{R}_{n,20}$ (eq. 1.12) in the winter season, or respectively, slightly lower $\tilde{R}_{n,20}$ in the summer season (see Fig. 1.2) is due to a change in the occurrence process and/or in the amount process. To do so, a linear drift will be implemented into the model. The same strategy was used by Wergen and Krug (2010) in the context of temperature records. In the context of precipitation, a change in the precipitation pattern can be caused by a change in the occurrence and/or in the amount process. Therefore, first of all, changes in the precipitation pattern in Germany since 1954 will be investigated with dataset DWD_{con} (sec. 4.1). The implementation of the drifts is described in subsection 4.2 and subsection 4.3 compares the results of the linear drift model with the observations.

4.1. Observed changes in the past

In this subsection, a change in the precipitation pattern in Germany is investigated for the period of 1954–2013. On the one hand, changes in the state transitions of the occurrence process are investigated. On the other hand, changes in the probability distribution (amount process) are studied for the winter (DJF) and summer (JJA) seasons. To keep it very simple, the time series of dataset DWD_{con} is split into two halves (1954–1983 and 1984–2013), which is occasionally used in climate science [e.g. Malitz et al. (2011); Thompson et al. (2017)]. For both the Fourier- respectively Cosine-coefficients (occurrence process) as well as the parameters a , b , c and d of the probability distribution (amount process) are determined. For this study, the complete, continuous dataset DWD_{con} was used since for both time periods the same number of stations and days were recorded. This dataset, however, has no information about the type of precipitation namely stratiform or convective. Therefore, for the amount process study, also the dataset SYNOP was used. Data

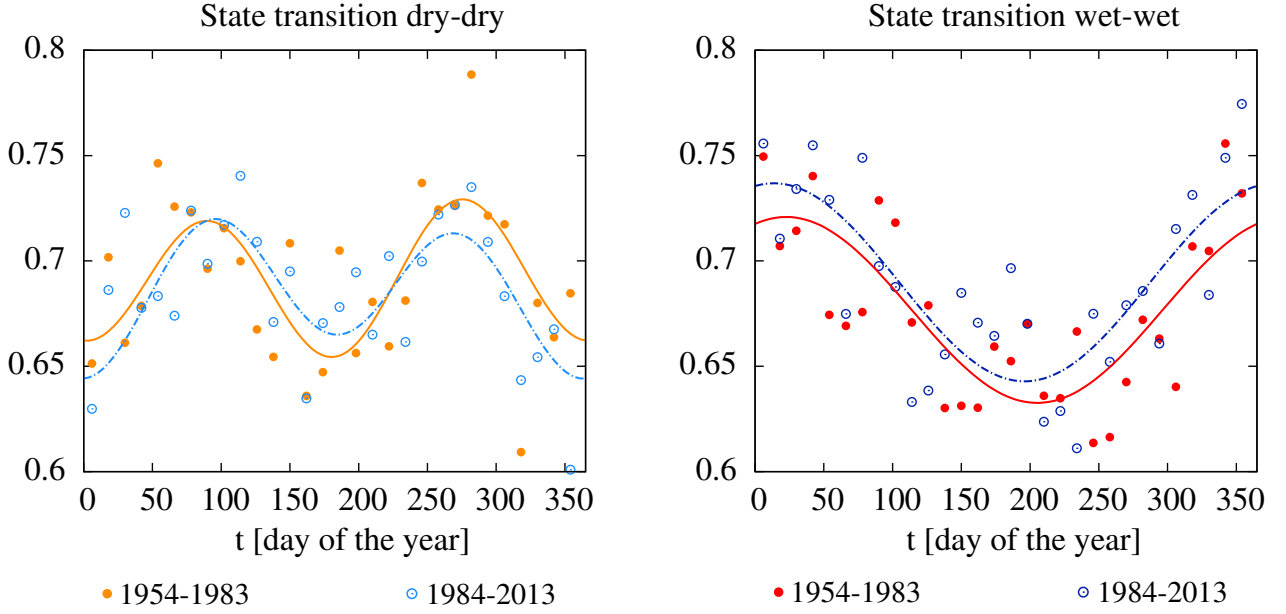


Figure 4.1.: Changes in the state transitions *dry-dry* (left) and *wet-wet* (right) by splitting dataset DWD_{con} into two halves.

from this dataset must be used with care because both periods are based on different stations. For example, between 1954 and 1983, it is weighted by containing more mountain stations (average station altitude: 345m) than in the period 1984–2013 (average station altitude: 306m).

4.1.1. Observed changes in the occurrence process

Due to the lack of information concerning the type of precipitation in dataset DWD_{con} , just the state transitions *dry-dry* and *wet-wet* could be investigated. By comparing the period 1954–1983 (orange dots) and the period 1984–2013 (light blue dots), a shift of the Fourier-series towards a slightly higher probability in the spring season (\sim day 60–150) and part of the summer season (\sim day 150–210) was observed for the state transition *dry-dry* (Fig. 4.1 left). For the rest of the year, the Fourier-series shifted towards lower probabilities for getting a dry day after a previous dry day. The shift was most significant in the winter season (\sim day 335–60) with a gap of up to two percent. This was also the case for the state transition *wet-wet* (Fig. 4.1 right) but with a shift to higher probabilities of up to two percent for the winter season. Differing from the state transition *dry-dry*, the Cosine-function of the state transition *wet-wet* shifted to higher probabilities all year-long. However, the shift is much weaker in the spring and part of the summer season.

In general, a higher probability of getting a rainy day in the winter season was ob-

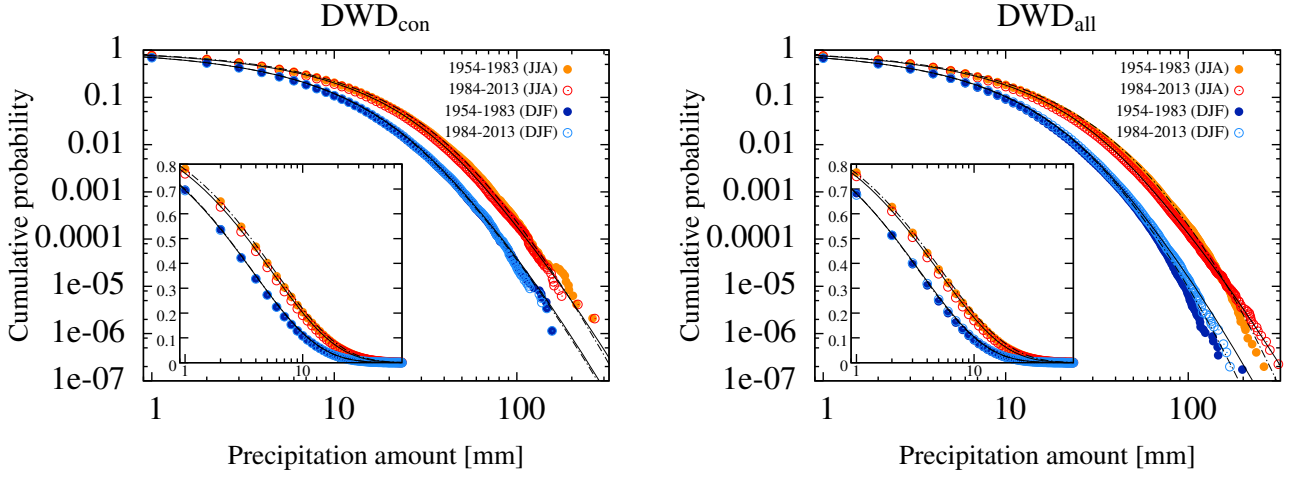


Figure 4.2.: Changes in the probability distributions of both the winter (dark blue 1954–1983; light blue 1984–2013) and summer (orange 1954–1983; red 1984–2013) season for dataset DWD_{con} (left) as well as DWD_{all} (right), respectively.

served, whereas in the summer season a more stable weather situation was observed: The probabilities increased for getting consecutive dry days as well as consecutive wet days.

4.1.2. Observed changes in the amount process

In the study of the amount process, a change between the period of 1954–1983 and the period of 1984–2013 is scarcely noticeable (Fig. 4.2 left). On closer study, there seems to be a slight reduction in high precipitation in summer, which is shown by orange (1954–1983) and red circles (1984–2013). This observation was confirmed through an identical evaluation of the markedly larger DWD_{all} dataset (Fig. 4.2 right). Therefore, a reduction in heavy precipitation (especially for precipitation amount $> 30\text{mm}$) seems more significant. Interestingly, this observation is reversed for extreme precipitation: the probability of daily precipitation, in excess of 200mm , was higher in the period of 1984–2013 than in 1954–1983.

In contrast to the dataset DWD_{con} , an increase in very heavy winter precipitation was observed for the dataset DWD_{all} (especially for precipitation amounts $> 100\text{mm}$) shown by dark (1954–1983) and light (1984–2013) blue dots (Fig. 4.2 right).

An investigation of the dataset SYNOP, including a separation of stratiform and convective rainfall, suggests that the decrease in high daily summer precipitation is

driven by a change in the stratiform precipitation (Fig. 4.3c). This could also be partly due to the lower average elevation of the stations in the period of 1984–2013 (314m in the summer season) compared to the period of 1954–1983 (360m in the summer season). In contrast, for convective rain, a slight increase in precipitation of less than about 50mm was observed (Fig. 4.3d).

Hardly any change was noticed in stratiform rain in winter (Fig. 4.3a). Nevertheless, the chance of convective precipitation of less than 25mm has, in contrast, increased (Fig. 4.3b).

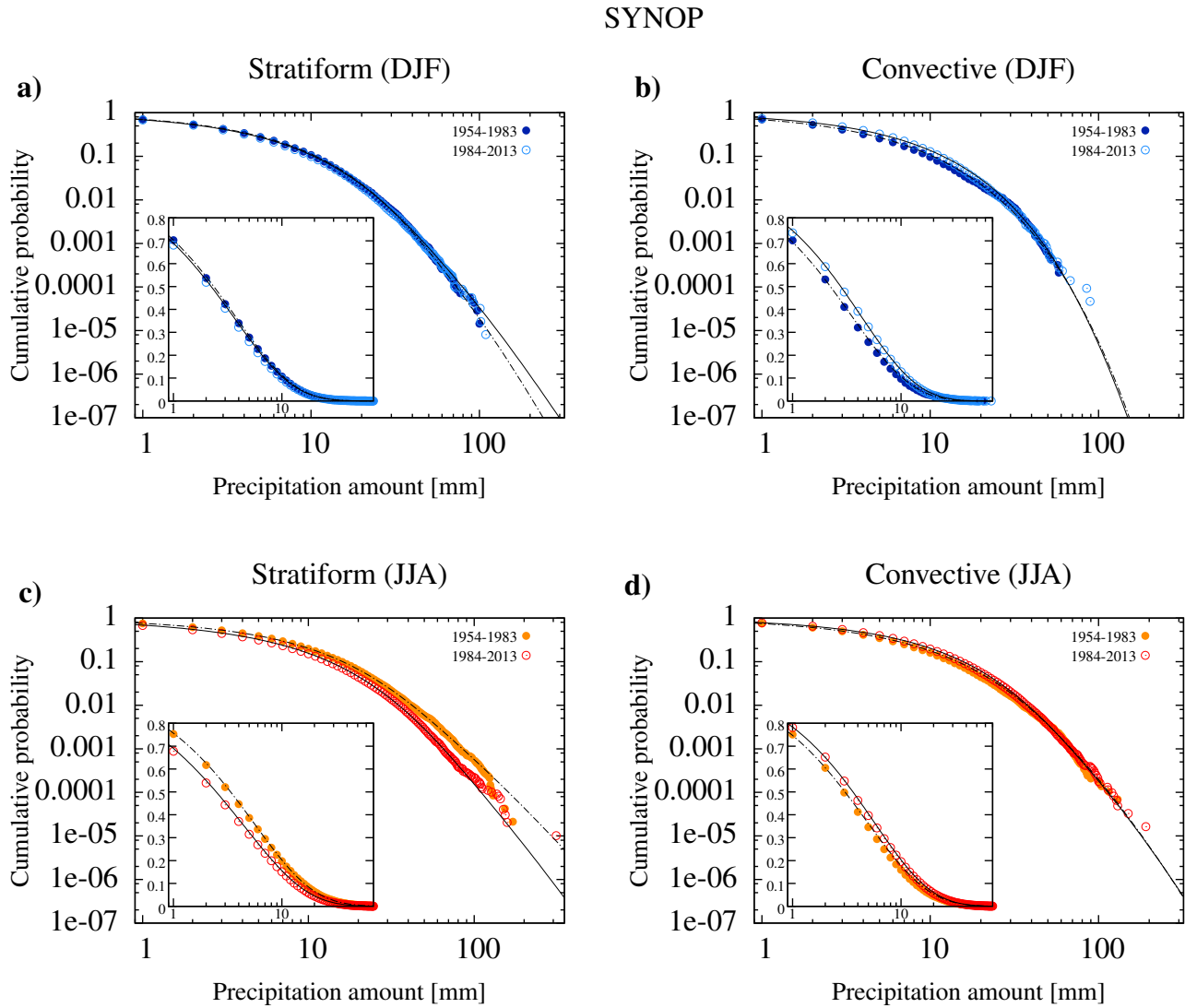


Figure 4.3.: Changes in the probability distributions of both stratiform (a) and convective (b) precipitation in the winter season as well as stratiform (c) and convective (d) precipitation in the summer season of dataset SYNOP.

4.2. Implementation of a drift in the precipitation model

In this chapter, the modification of the precipitation model by an implementation of the observed changes (see sec. 4.1) is described. The implementation of the observed changes was realized by using linear drifts in the occurrence and amount processes.

In the occurrence process, the linear drifts arise from the change in the coefficients of the Fourier series and Cosine-function. With that implementation the models state transitions *dry-dry* and *stra-stra* shift as shown in Fig. 4.1. In the models state transition *conv-conv* no linear drift is implemented, because convective rain is more likely to occur in unstable weather situations. Thus, it is assumed that the observed change in the state transition *wet-wet* is mainly driven by a shift in the state transition *stra-stra*.

The changes in the amount process were introduced into the model by using a linear drift of the parameters of the precipitation spread. Thereby, a slight linear drift of the convective precipitation was used in the winter season (as shown in Fig. 4.3b), while for the stratiform amount process no linear drift was used. For the summer, the changes in stratiform and convective distribution were implemented as shown in Fig. 4.3c and Fig. 4.3d, respectively.

To study the effects of these changes on the mean record number, the changes to the model were divided into a total of five model runs (see Tab. 4.1). Model run I represents a stationary precipitation model without linear drift and correlations. In Model run II and III the stationary model is modified by adding a linear drift in the occurrence and amount process, respectively. In model run IV, the linear drifts in the occurrence and amount process is implemented, including correlated fluctuations in the occurrence process. The correlations in the occurrence process are realized as

Table 4.1.: Overview with descriptions of the model runs used.

Model run I	Stationary model (sec. 3.1 and 3.2)
Model run II	Model run I + linear drift in the occurrence process (sec. 4.1.1)
Model run III	Model run I + linear drift in the amount process (sec. 4.1.2)
Model run IV	Model run II + III + correlation in the occurrence process (sec. 3.3)
Model run V	Model run I but with empiric distribution of the occurrence process

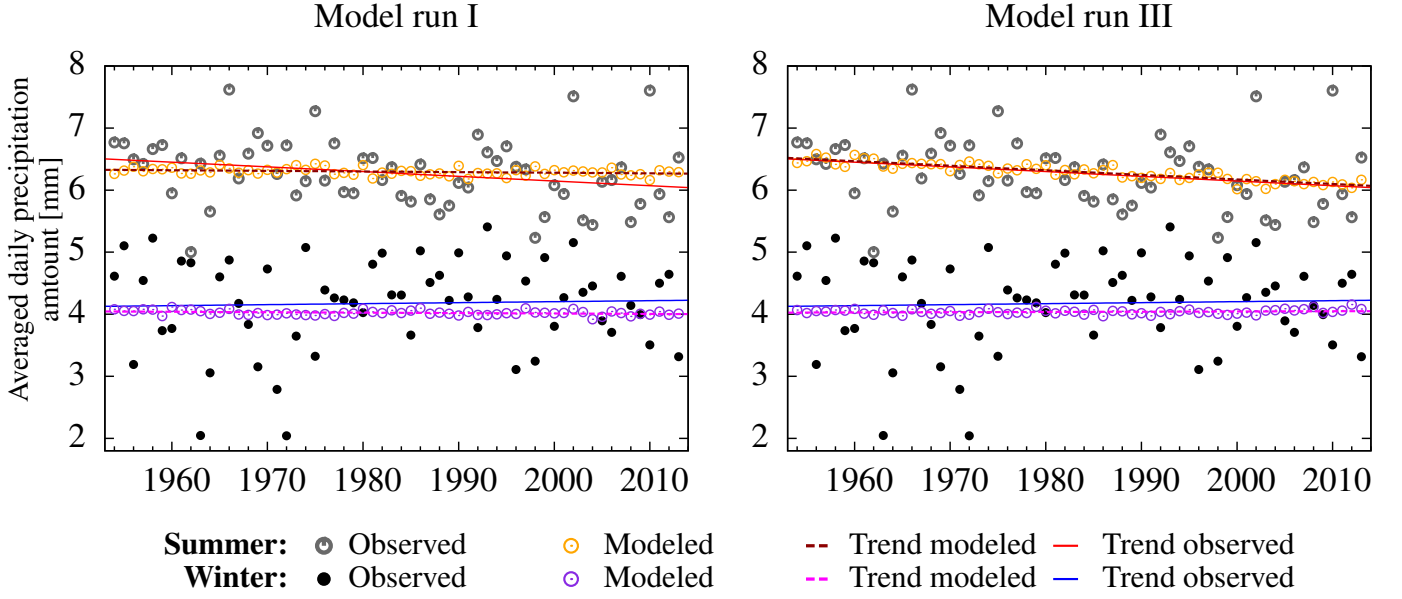


Figure 4.4.: Comparison of observed daily precipitation averages to those of model run I (left) and model run III (right) for given years.

described in chapter 3.3. Finally, model run V uses the empiric distribution of the occurrence process.

To validate this model, the individual model runs were compared to the observations for rain probability (Fig. 4.5) and to the average daily precipitation quantities (Fig. 4.4). As it was expected from the results from Chapter 4.1.1, an increase in the probability of precipitation during winter can be found in the observations (blue lines in Fig. 4.5 left).

In summer, the changes noticed in the occurrence process nearly cancel each other out and only a small increase in the probability of precipitation can be seen in the observations (red lines in Fig. 4.5 right). The mean changes in the probability of precipitation could be realistically reproduced with the help of the linear drifts in the occurrence process (model run II) (Fig. 4.5c and Fig. 4.5d). Through the addition of spatial correlations in the occurrence process, the model outputs and the observations show very similar results (Fig. 4.5e and Fig. 4.5f). However, the average probability of rain generated by the model is slightly lower than the observations (dotted lines in Fig. 4.5). The difference can be explained by the fact that the precipitation model is based on the SYNOP data, while the observed changes were studied with the dataset DWD_{con} . However, for the investigation of the mean record number, especially the changes over time are important.

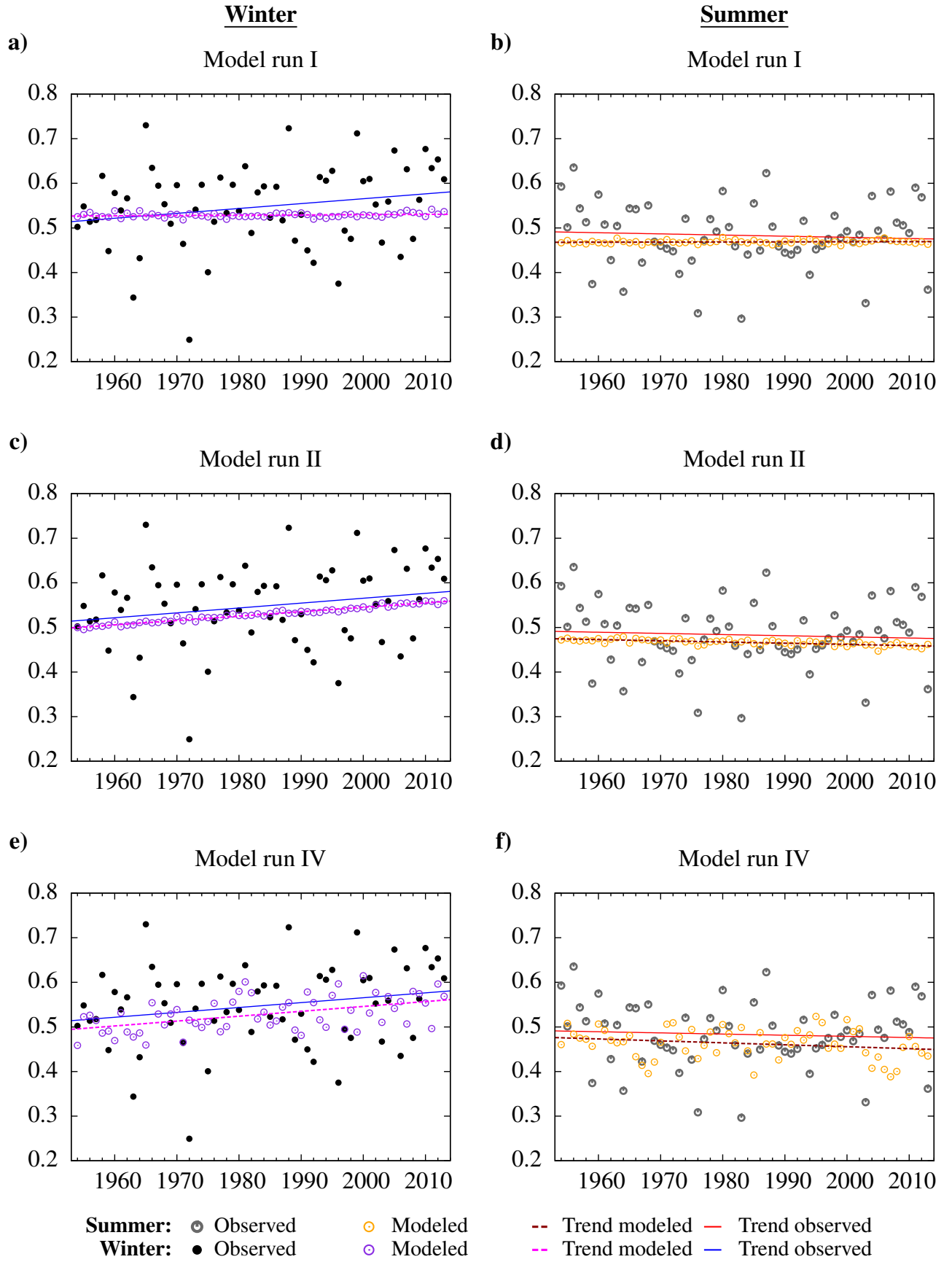


Figure 4.5.: Comparison of the observed rain probabilities to simulations of model run I in winter (a), model run I in summer (b), model run II in winter (c), model run II in summer (d), model run IV in winter (e) and model run IV in summer (f).

For the trend, a comparison of the daily average precipitation also shows a realistic modelling of the observations using the linear drift in the amount process – especially for the summer season (yellow circles in Fig. 4.4 right). In winter, a slight linear drift in the convective amount process has no effect on the average amount of precipitation (purple circles in Fig. 4.4 right). This can be explained by the very low probability of convective precipitation in winter.

However, it is easy to see that annual variability is not at all captured by the model (Fig. 4.4). This is an indication on missing correlations in the occurrence process that should be considered for further research (see also sec. 5).

4.3. Modelled precipitation records

In this chapter, the effects of the changes described in the previous chapter (sec. 4.2) on the mean record number are discussed. The evaluation is based on 20 simulations for each model run.

The modelled results for the stationary example (model run I), for both summer and winter, are very close to the theoretical expectations (eq. 1.12 in sec. 1.2) from i.i.d. random variables (dotted lines in Fig. 4.6a and Fig. 4.6b). By the implementation of linear drifts in the occurrence process, an increase in the mean record number in winter was observed (Fig. 4.6c). In comparison to that, no obvious changes were found for the summer season (Fig. 4.6d). When a linear drift in the amount process is considered, the result is exactly reversed: In winter, there is no change in the mean record number, while in summer the result is a decrease in the mean record number of the magnitude actually observed (Fig. 4.7a and Fig. 4.7b).

If additional spatial correlations in the occurrence process are considered, the variation in the mean record number increases noticeably (shaded area in Fig. 4.7c and Fig. 4.7d). Interestingly, despite taking the correlations into account, none of the 20 simulations produced a result on the scale of the actual mean record numbers observed in winter.

Finally, the days on which it rained were set empirically rather than being modelled by the occurrence process (model run V, Fig. 4.7e and Fig. 4.7f). In winter, this led to similar results for the mean record number than obtained through modelling with a linear drift in the occurrence process (model run II, Fig. 4.6c). In

summer, a lower mean record number was found in comparison to the stationary model. This is particularly remarkable as the effect does not occur (or was not noticed) when the linear drift in the occurrence process is considered (Fig. 4.6d).

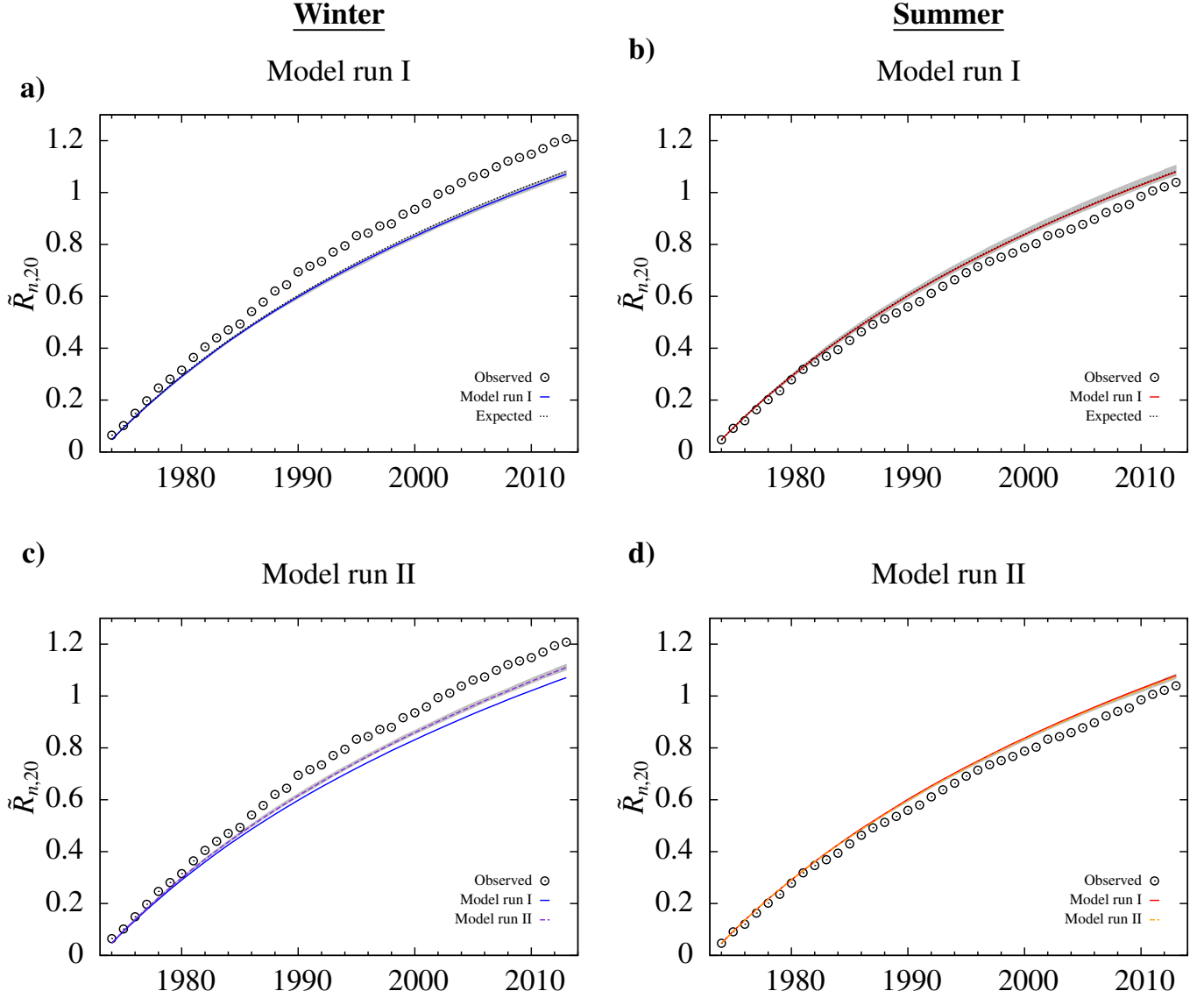


Figure 4.6.: Modified mean record number (eq. 1.12) of dataset DWD_{con} (circles) compared to 20 simulations of model run I (a-b) and model run II (c-d) for both winter (left) and summer (right) season. Model run I is also compared to the prediction for a stationary process (dotted lines in (a) and (b)), while model run II is additionally compared to model run I (solid lines in (c) and (d)). Shaded areas show the ranges of variation of the simulations.

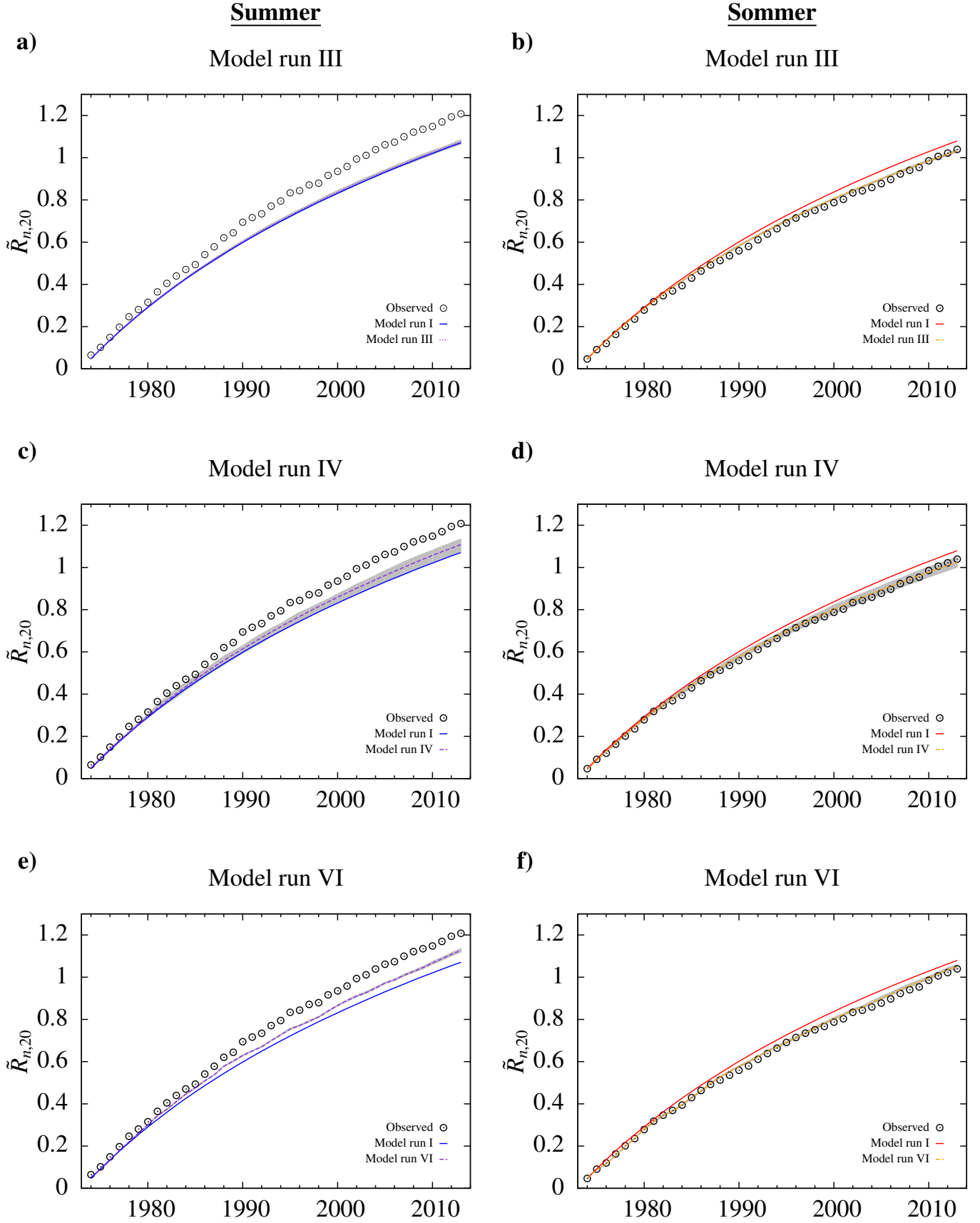


Figure 4.7.: Same as Fig. 4.6, but for model run III (a-b), model run IV (c-d) and model run V (e-f).

5. Conclusion and Outlook

In this study a new statistical precipitation model was developed and has been applied to investigate the results of previous work in the context of precipitation records [von Bomhard (2014)]. To this model several new aspects were included: On the one hand it is the first statistical daily precipitation model which distinguishes between convective and stratiform precipitation. On the other hand a new probability distribution was developed for the models amount process. Furthermore the model was developed for universal use, as it uses dependencies on the elevation of a station and on the time of the year as input parameters.

By differentiating convective and stratiform precipitation the model was supposed to reflect the different characteristics of the two precipitation types. In the occurrence process this differentiation of the types of precipitation leads to state transitions as expected – namely higher probabilities for remaining in a stratiform-state than for remaining in a convective-state. In the probability distribution of the amount process the differences of convective and stratiform precipitation is more significant in the parameters a and b than in the parameters c and d (see Fig. 3.10). Nevertheless, the daily precipitation probability distributions for these both types of rain are much closer than it would be expected due to their characteristics. The reason is, that the more intensive convective precipitation is typically of much shorter duration than the less intensive stratiform precipitation. On a time scale of one day (24 hours) this can lead to precipitation totals of similar magnitude (see Fig. 5.1).

By comparing the results of the model with the observations, it was found that the model was still to static. In particular the occurrence of very long dry- as well as wet-spells were underestimated. Hence, a confined random walk was implemented in the occurrence process for taking variability into account. Because the determined shift Δ of the random walk is valid for all stations, this leads to correlations in the occurrence process. In the amount process no correlations were taken into account, as a comparison of the highest modelled versus the highest observed precipitation amounts showed good agreement for the dataset SYNOP (see Fig. 4.2 left).

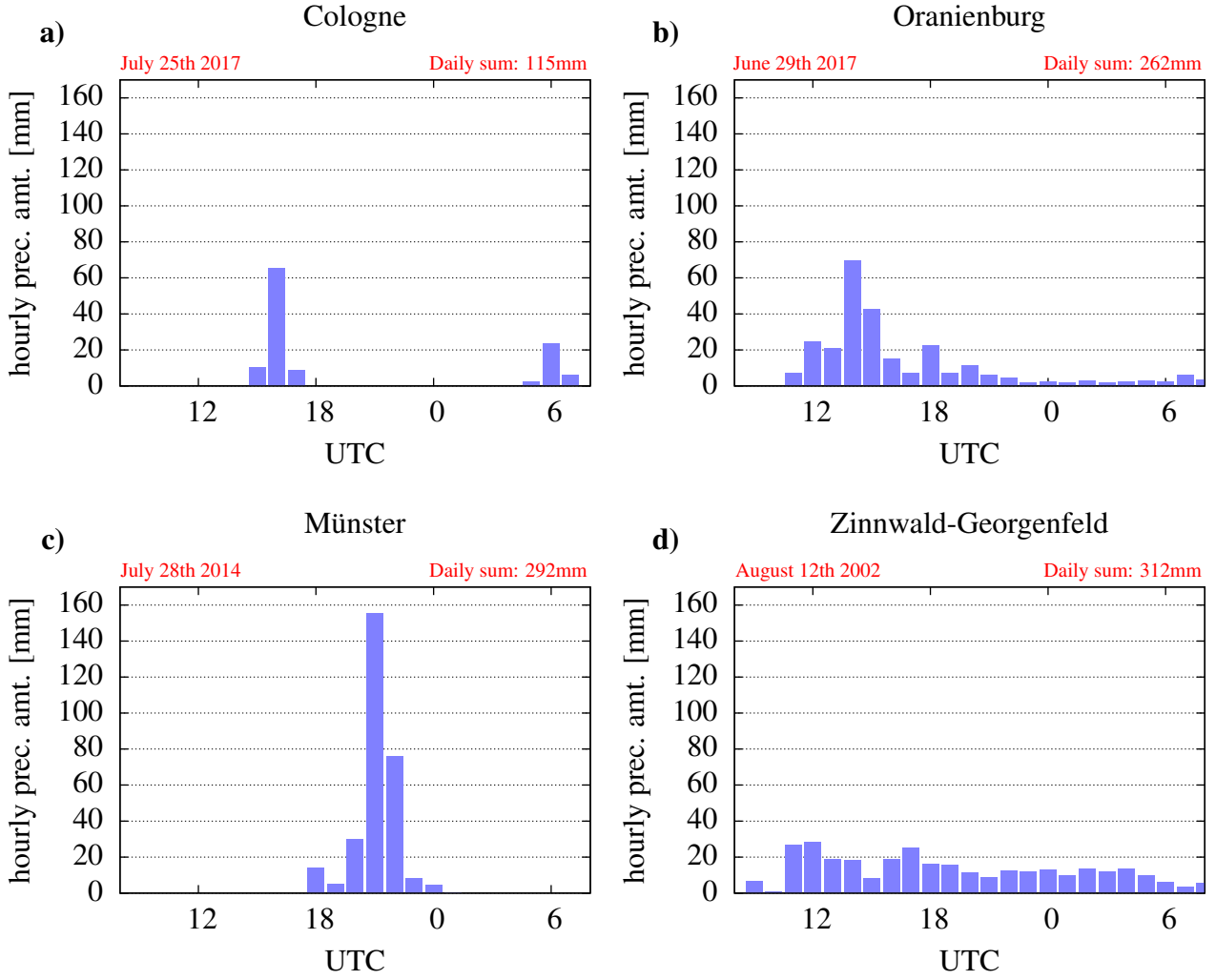


Figure 5.1.: Examples of extreme daily precipitation events on hourly resolution. For the convective events in Münster (a) and Cologne (c) as well as the mixed (convective followed by stratiform precipitation) event of Berlin-Oranienburg (b) and the stratiform event of Zinnwald-Georgenfeld (d).

In a next step the model was used for investigating precipitation records. For that, five different simulation runs with incremental changes in the model were utilized. The modifications of the model include implementations of linear drifts. These linear drifts were simply realized by changes in the parameters, which were estimated by a comparison of two 30-year partial periods (period 1954–1983 and period 1984–2013) of the dataset DWD_{con} .

As result, for the winter season, a significantly increased probability of rainy days but no changes in the amount process was found. Comparable results were found by Malitz et al. (2011), who – based on 83 German weather stations in the period 1901–2000 – found a much more significant increase in the frequency of heavy pre-

cipitation occurrence, than in its intensity. However, an increase in very heavy and weak precipitation events at the expense of medium rainfall as presented by some studies [e.g. Hänsel et al. (2005); Hattermann et al. (2013)] was not found by the comparison of the two partial periods with dataset DWD_{con} .

In summer, however, a decrease in medium and heavy precipitation events and hardly no change in the occurrence process was found, which is in consistence with previous studies (see Kunz et al. (2017) and references therein).

After these linear drifts were implemented to the model, a lower mean record number than expected for stationary conditions was found for the summer season. It was of similar magnitude as observed in reality (for 2013: $\tilde{R}_{60,20} = 1.03 \pm 0.01$ vs. $\tilde{R}_{60,20} = 1.04$ observed). An important change in the amount process is mainly the decrease of medium and strong stratiform precipitation. However, also changes in the occurrence of precipitation events in the summer months may have contributed to less records than expected: An implementation of the empirical occurrence of precipitation led to a mean record number of $\tilde{R}_{60,20} = 1.05 (\pm 0.01)$ for the year 2013. However, this is not necessarily due to significant changes in the occurrence process, but can also be due to random processes as it was shown by simulations considering correlations. To sum this up, one can conclude that a too low mean record number can result from changes in the amount process or/and coincidences in the occurrence process.

In the winter season, the increase in daily precipitation records can only be partially explained by an increase in the rain probability (for 2013: $\tilde{R}_{60,20} = 1.11 \pm 0.01$ vs. $\tilde{R}_{60,20} = 1.21$ observed). Taking correlations in the occurrence process into account it has the potential to slightly increase the mean record number further (for 2013: $\tilde{R}_{60,20} = 1.11 \pm 0.03$). Finally the simulations using the empirical occurrence of precipitation gives a mean record number of $\tilde{R}_{60,20} = 1.13 (\pm 0.01)$ for the year 2013. In conclusion it remains an open question, which additional enhancing effects have contributed to the observed mean record number of $\tilde{R}_{60,20} = 1.21$. One effect could be that actually also a linear drift in the winters amount process must be considered. The assumption for not taking linear drifts into account, was due to results of splitting the dataset DWD_{con} into two 30-year periods. As this method is very simple, the assumption might be wrong. That a linear drift should be implemented is supported by previous studies which found, as mentioned earlier, an increase in heavy precipitation events for German winters (see Kunz et al. (2017) and references

therein). Another reason could be the neglect of spatial correlations in the amount process. Spatial correlations in the amount process are expected to be especially important for stratiform precipitation. While precipitation in winter time is almost always of stratiform character, the neglect of correlations in the amount process can have significant effects on the mean record number of the winter season.

Spatial correlations could be, for instance, implemented to the model by calculating cross correlations [e.g. Brommundt and Bárdossy (2007)]. To keep the model as universal as developed in this study, another idea is to generate correlated random numbers for each day which are independent of the marginal distributions. This can be realized, for example, by using a copula approach. Copula are often used in the fields of risk management, insurance and especially in finance. However, in the last decade copulas become also popular on the issue of precipitation models [e.g. Lennartsson et al. (2008); Bárdossy and Pegram (2009); Nguyen-Huy et al. (2017)]. A good introduction of using copula in meteorology and climate research including an application to daily rainfall can be found in Schoelzel and Friederichs (2008).

However, on a daily scale, the variability of the weather seems to cause that the identification of significant trends in records is more difficult than for longer time scales. This aspect was also discussed by Rahmstorf and Coumou (2011) and Wergen et al. (2014) in the context of temperature records: Rahmstorf and Coumou (2011) argued that already a noticeable increase in records can be found for a slight increase in annual temperatures on a global scale due to its annual and spatial averaging. Also Wergen et al. (2014) justify a stronger increase in monthly record highs compared to daily record highs with a smaller standard deviation in monthly values. Therefore, it might be interesting for further study to use the daily precipitation model (by merging 30-day intervals) to investigate monthly precipitation records or to use the maximum daily precipitation amount of each month (similar to Lehmann et al. (2015)).

Nevertheless, a higher temporal resolution of the model should be also of great interest. As previously mentioned, the different characteristics of stratiform and convective precipitation should be more sharply distinguishable on a temporal resolution of 1-hour. For example, results of Langer et al. (2008) suggest, that the probability distribution of 1-hourly convective and stratiform precipitation amounts is of much higher difference than in the findings of this study for daily values. With an even higher resolution of 5 minutes the difference seems to be even more apparent

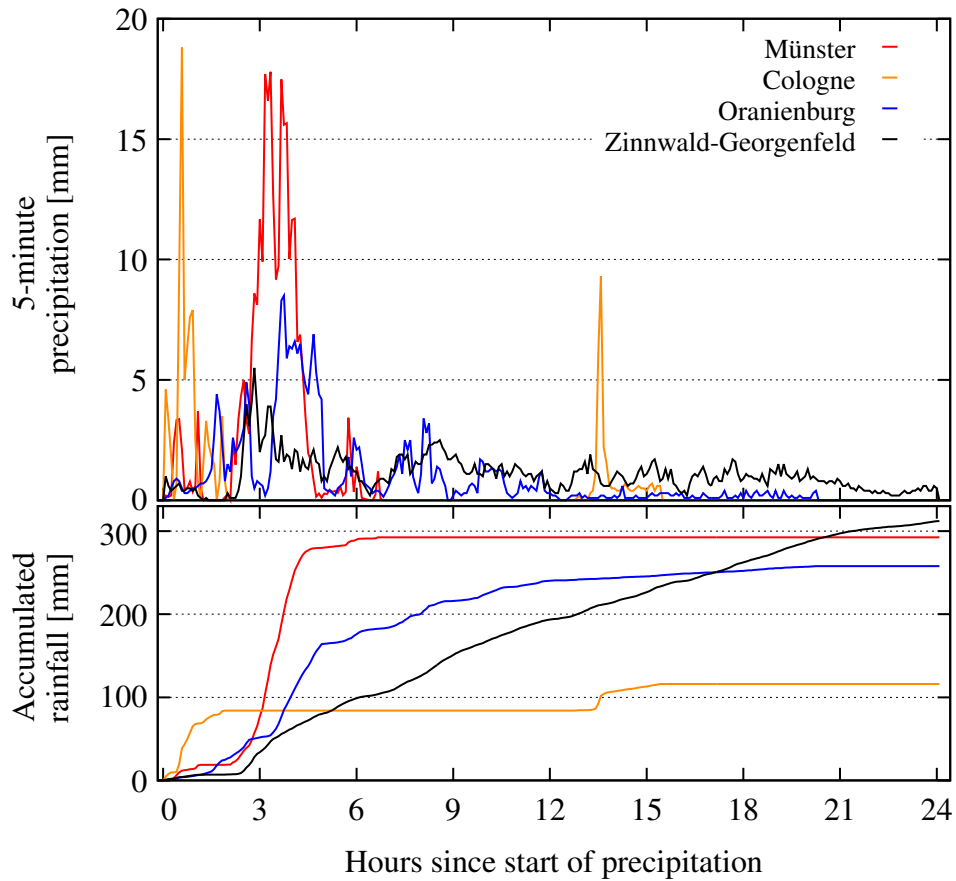


Figure 5.2.: Daily precipitation totals with temporal resolution of 5-minutes (top) for the same events as in Fig. 5.1. In addition accumulated rainfall of these events is shown (bottom).

(see Fig. 5.2). It is interesting to note in this context that very recently a new law entered into force¹, stating that the DWD has to provide all collected data, including 1-minute precipitation measurements, for free.

Using this high temporal resolution data and implementing spatial correlations by using a copula approach should further improve the model. All in all, it is expected that such a universal and well working precipitation model has the potential for applications in many more research areas, such as implementation in a statistical weather generator or modelling damages of flash floods.

¹see the press release from July, 25th 2017: http://www.dwd.de/EN/press/press_release/EN/2017/20170725_amendment_to_the_DeutscherWetterdienst.pdf?__blob=publicationFile&v=3

Appendix A.

Table A.1.: SYNOP-codes of weather state characterizing stratiform precipitation.

Table taken from Rulfová and Kyselý (2013).

1. Precipitation within past hour but not at observation time 20 drizzle 20 drizzle 21 rain 22 snow 23 rain and snow 24 freezing rain	2. Drizzle 50 intermittent light snow 51 continuous light drizzle 52 intermittent moderate drizzle 53 continuous moderate drizzle 54 intermittent heavy drizzle 55 continuous heavy drizzle 56 light freezing drizzle 57 moderate to heavy freezing drizzle 58 light drizzle and rain 59 moderate to heavy drizzle and rain
3. Rain (not in the form of showers) 60 intermittent light rain 61 continuous light rain 62 intermittent moderate rain 63 continuous moderate rain 64 intermittent heavy rain 65 continuous heavy rain 66 light freezing rain 67 moderate to heavy freezing rain 68 light rain and snow 69 moderate to heavy rain and snow	4. Snow (not in the form of showers) 70 intermittent light snow 71 continuous light snow 72 intermittent moderate snow 73 continuous moderate snow 74 intermittent heavy snow 75 continuous heavy snow 76 diamond dust 77 snow grains 78 snow crystals 79 ice pellets

Table A.2.: SYNOP-codes of weather state characterizing convective precipitation.

Table taken from Rulfová and Kyselý (2013).

1. Non-precipitation events	2. Precipitation within past hour but not at observation time
17 thunderstorm but no precipitation falling at station	25 snow showers
18 squalls within sight but no precipitation falling at station	26 snow showers
19 funnel clouds within sight	27 hail showers
	29 thunderstorms
3. Showers	4. Thunderstorms
80 light rain showers	91 thunderstorm in past hour, currently only light rain
81 moderate to heavy rain showers	92 thunderstorm in past hour, currently only moderate to heavy rain
82 violent rain showers	93 thunderstorm in past hour, currently only light snow or rain/snow mix
83 light rain and snow showers	94 thunderstorm in past hour, currently only moderate to heavy snow or rain/snow mix
84 moderate to heavy rain and snow showers	95 light to moderate thunderstorm
85 light snow showers	96 light to moderate thunderstorm with hail
86 moderate to heavy snow showers	97 heavy thunderstorm
87 light snow/ice pellet showers	98 heavy thunderstorm with duststorm
88 moderate to heavy snow/ice pellet showers	99 heavy thunderstorm with hail
89 light hail showers	
90 moderate to heavy hail showers	

Table A.3.: SYNOP-codes of low level clouds and its decoding. Given types of precipitation are based on findings of Langer and Reimer (2007).

Code	Cloud type	Type of prec.
0	no low clouds	dry
1	cumulus (<i>Cu</i>) humulis or fractus (no vertical development)	dry
2	cumulus (<i>Cu</i>) mediocris or congestus (moderate vertical development)	convective
3	cumulonimbus (<i>Cb</i>) calvus (no outlines nor anvil)	convective
4	stratocumulus (<i>Sc</i>) cumulogenitus (formed by spreading of cumulus)	stratiform
5	stratocumulus (<i>Sc</i>) (not formed by spreading cumulus)no low clouds	stratiform
6	stratus (<i>St</i>) nebulosus (continuous sheet)	stratiform
7	stratus (<i>St</i>) or cumulus (<i>Cu</i>) fractus of bad weather	stratiform
8	cumulus (<i>Cu</i>) with stratocumulus (<i>St</i>), with differing bases	mixed
9	cumulonimbus (<i>Cb</i>) with anvil	convective
/	low clouds unobserved due to darkness or obscuration	—

Appendix B.

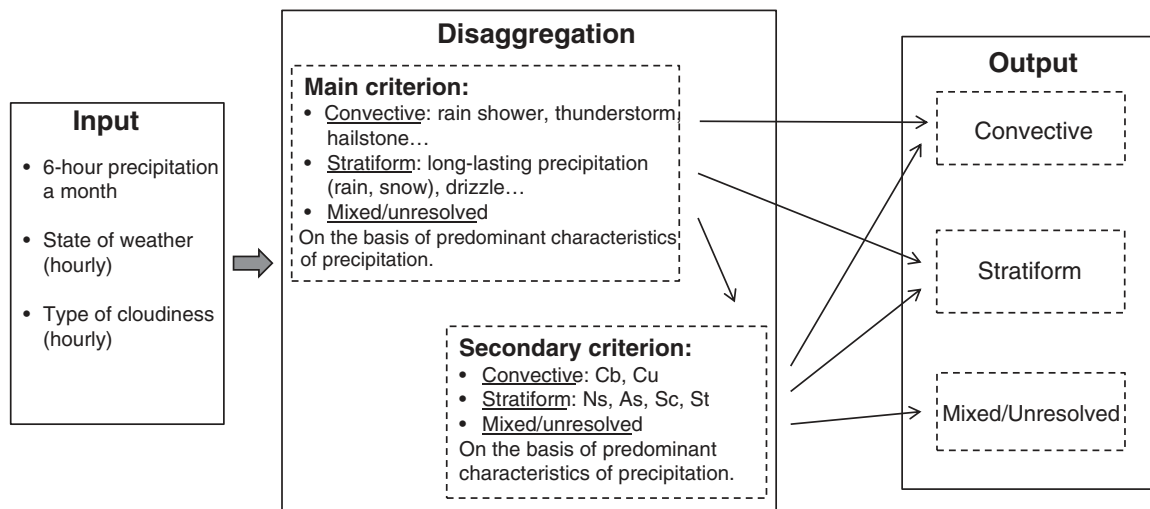


Figure B.1.: Schematic illustration disaggregation of the algorithm used Rulfová and Kyselý (2013).

Appendix C.

Table C.1.: Overview of the time- (t) and altitude- (h) dependent state transitions used in the model. $U_{cos}(t, h)$ and $U_{fr}(t, h)$ refer to eq. 3.2 and 3.3, respectively.

	Stratiform	Convective	Dry
Stratiform	$SS(t, h) = U_{cos}(t, h)$ $m_0(h) = 7 \times 10^{-5}h + 0.52$ $m_1 = 0.11$ $m_2 = -0.26$	$SC(t, h) = U_{cos}(t, h)$ $m_0(h) = -2 \times 10^{-5}h + 0.1$ $m_1(h) = 3 \times 10^{-6}h - 0.05$ $m_2 = -0.26$	$SD(t, h) = 1 - SS - SC$
Convective	$CS(t, h) = U_{cos}(t, h)$ $m_0 = 0.28$ $m_1 = 0.1$ $m_2 = -0.26$	$CC(t, h) = 1 - CS - CD$	$CD(t, h) = U_{cos}(t, h)$ $m_0(h) = -6 \times 10^{-5}h + 0.3$ $m_1 = -0.06$ $m_2 = -0.26$
Dry	$DS(t, h) = 1 - DC - DD$	$DC(t, h) = U_{fr}(t, h)$ $n_0(h) = 2 \times 10^{-5}h + 0.05$ $n_1(h) = -3 \times 10^{-5}h - 0.04$ $n_2 = 0$ $n_3(h) = 8 \times 10^{-6}h + 0.01$	$DD(t, h) = U_{fr}(t, h)$ $n_0(h) = \gamma(h)\nu_1(h) + (1 - \gamma(h))\nu_2(h)$ $\gamma(h) = \left(1 + \frac{h}{30.85}\right)^{-2}$ $\nu_1(h) = -1 \times 10^{-3}h + 0.68$ $\nu_2(h) = -3 \times 10^{-5}h + 0.7$ $n_1(h) = 1 \times 10^{-4}h - 0.04 - \frac{h^2}{5.63 \times 10^7}$ $n_2(h) = -0.01 \log(0.76h) + 0.04$ $n_3 = -0.03$

Table C.2.: Overview of the used parameters of f_{w+pl} (see eq. 3.23) with dependencies on the topographic height of a station (h) and the time of the year (t). $U_{cos}(t, h)$ and $U_{fr}(t, h)$ refer to eq. 3.2 and 3.3, respectively.

	Stratiform	Convective
a	$a_s(h) = 3 \times 10^{-5}h + 0.75$	$a_c(h) = 1 \times 10^{-5}h + 0.81$
b	$b_s(t, h) = U_{fr}(t, h)$ $n_0(h) = 4 \times 10^{-3}h + 3.2$ $n_1(h) = 5 \times 10^{-7}(h - 1000)^2 - 1$ $n_2(h) = n_1(h)$ $n_3(h) = 1 \times 10^{-4}h + 0.04$ $n_4(h) = -2 \times 10^{-7}(h - 1300)^2 + 0.3$	$b_c(t, h) = U_{cos}(t, h)$ $n_0(h) = 4 \times 10^{-3}h + 4.0$ $n_1 = 1.1$ $n_2 = 2.25$
c	$c_s(h) = 0.02h + 17.28$	$c_c(h) = 5 \times 10^{-3}h + 28.03$
d	$d_s(t, h) = U_{cos}(t, h)$ $n_0(h) = 1 \times 10^{-3}h + 6.94$ $n_1(h) = -1 \times 10^{-3}h - 0.54$ $n_2 = \pi$	$d_c(t, h) = U_{cos}(t, h)$ $n_0(h) = 9 \times 10^{-4}h + 6.43$ $n_1(h) = 4 \times 10^{-4}h + 1.31$ $n_2 = -0.75$

Bibliography

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6):716–723.
- Axer, T., Bistry, T., Klawe, M., Müller, M., and Süßer, M. (2017). Sturmdokumentation 2016 Deutschland. *Deutsche Rück, Düsseldorf*. Available online at https://www.deutscherueck.de/fileadmin/files/Sturmdokumentation_2016_.pdf; visited on August 5th 2017.
- Axer, T., Bistry, T., Klawe, M., Müller, M., Süßer, M., and von Bomhard, P. (2015). Sturmdokumentation 2014 Deutschland. *Deutsche Rück, Düsseldorf*. Available online at http://www.deutscherueck.de/fileadmin/user_upload/Sturmdoku_2014_WEB.pdf; visited on August 5th 2017.
- Ballerini, R. and Resnick, S. (1985). Records from improving populations. *Journal of Applied probability*, 22(3):487–502.
- Ballerini, R. and Resnick, S. I. (1987). Records in the presence of a linear trend. *Advances in Applied Probability*, 19(4):801–828.
- Bárdossy, A. and Pegram, G. (2009). Copula based multisite model for daily precipitation simulation. *Hydrology and Earth System Sciences*, 13(12):2299.
- Becker, A., Weigl, E., Winterrath, T., Junghänel, T., Schmitt, A., Hafer, M., Sterker, C., Müller, H.-J., Tracksdorf, P., and Helmert, K. (2014). Das Münsterereignis: 292 l/qm binnen 7 Stunden – Klimatologische Einordnung solcher Ereignisse beim DWD derzeit und in Zukunft. 9. *Extremwetterkongress, Hamburg*.
- Benestad, R. (2003). How often can we expect a record event? *Climate research*, 25(1):3–13.
- Berg, P., Moseley, C., and Haerter, J. O. (2013). Strong increase in convective precipitation in response to higher temperatures. *Nature Geoscience*, 6(3):181–185.

- Brommundt, J. and Bárdossy, A. (2007). Spatial correlation of radar and gauge precipitation data in high temporal resolution. *Advances in Geosciences*, 10:103–109.
- Buishand, T. (1978). Some remarks on the use of daily rainfall models. *Journal of Hydrology*, 36(3-4):295–308.
- Chandler, K. (1952). The distribution and frequency of record values. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 220–228.
- Clauset, A., Shalizi, C. R., and Newman, M. E. (2009). Power-law distributions in empirical data. *SIAM review*, 51(4):661–703.
- Clauset, A., Young, M., and Gleditsch, K. S. (2007). On the frequency of severe terrorist events. *Journal of Conflict Resolution*, 51(1):58–87.
- Dalevi, D., Dubhashi, D., and Hermansson, M. (2006). Bayesian classifiers for detecting HGT using fixed and variable order markov models of genomic signatures. *Bioinformatics*, 22(5):517–522.
- DWD (2013). Deutschlandwetter im Mai 2013. *Deutscher Wetterdienst Pressestelle*. Available online at http://www.dwd.de/DE/presse/pressemitteilungen/DE/2013/20130529_DeutschlandwetterimMai.html; visited on August 5th 2017.
- Franke, J., Wergen, G., and Krug, J. (2010). Records and sequences of records from random variables with a linear trend. *Journal of Statistical Mechanics: Theory and Experiment*, 2010(10):P10013.
- Gabriel, K. and Neumann, J. (1962). A Markov chain model for daily rainfall occurrence at Tel Aviv. *Quarterly Journal of the Royal Meteorological Society*, 88(375):90–95.
- Gebauer, P., Myrcik, G., and Schenk, F. (2017). Berlin unter Wasser. *Berliner Wetterkarte*, 40(17):1–12.
- Hänsel, S., Küchler, W., and Matschullat, J. (2005). Regionaler Klimawandel Sachsen. *Umweltwissenschaften und Schadstoff-Forschung*, 17(3):159–165.
- Hattermann, F. F., Kundzewicz, Z. W., Huang, S., Vetter, T., Gerstengarbe, F.-W., and Werner, P. (2013). Climatological drivers of changes in flood hazard in Germany. *Acta Geophysica*, 61(2):463–477.

- Houze Jr, R. A. (2014). *Cloud dynamics. Chapter 6 – Nimbostratus and the Separation of Convective and Stratiform Precipitation*, volume 104. International Geophysics, Academic press.
- Kaspar, F., Müller-Westermeier, G., Penda, E., Mächel, H., Zimmermann, K., Kaiser-Weiss, A., and Deutschländer, T. (2013). Monitoring of climate change in Germany—data, products and services of Germany’s National Climate Data Centre. *Advances in Science and Research*, 10(1):99–106.
- Kaspar, F., Tinz, B., Mächel, H., and Gates, L. (2015). Data rescue of national and international meteorological observations at Deutscher Wetterdienst. *Advances in Science and Research*, 12(1):57–61.
- Katz, R. W. (1977). Precipitation as a chain-dependent process. *Journal of Applied Meteorology*, 16(7):671–676.
- Kunz, M., Mohr, S., and Werner, P. (2017). Niederschlag. In *Klimawandel in Deutschland*, pages 57–66. Springer.
- Lang, S., Tao, W., Simpson, J., and Ferrier, B. (2003). Modeling of convective–stratiform precipitation processes: Sensitivity to partitioning methods. *Journal of Applied Meteorology*, 42(4):505–527.
- Langer, I. and Reimer, E. (2007). Separation of convective and stratiform precipitation for a precipitation analysis of the local model of the German Weather Service. *Advances in Geosciences*, 10:159–165.
- Langer, I., Reimer, E., and Oestreich, A. (2008). Relation of rain clouds from satellite cloud classification to conventional precipitation surface data for Central Europe. *Meteorologische Zeitschrift*, 17(1):29–37.
- Lehmann, J., Coumou, D., and Frieler, K. (2015). Increased record-breaking precipitation events under global warming. *Climatic Change*, 132(4):501–515.
- Lenderink, G., Barbero, R., Loriaux, J., and Fowler, H. (2017). Super Clausius-Clapeyron scaling of extreme hourly convective precipitation and its relation to large-scale atmospheric conditions. *Journal of Climate*, (2017).
- Lennartsson, J., Baxevani, A., and Chen, D. (2008). Modelling precipitation in Sweden using multiple step Markov chains and a composite model. *Journal of Hydrology*, 363(1):42–59.

- Lilliefors, H. W. (1967). On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of the American statistical Association*, 62(318):399–402.
- Malitz, G., Beck, C., and Grieser, J. (2011). Veränderung der Starkniederschläge in Deutschland (Tageswerte der Niederschlagshöhe im 20. Jahrhundert). In *Warnsignal Klima: Genug Wasser für alle?, 3. Aufl. Universitätsverlag, Hamburg*, pages 311–316. Lozán JL, Graßl H, Hupfer P, Karbe L, Schönwiese CD (Hrsg).
- Meehl, G. A., Tebaldi, C., Walton, G., Easterling, D., and McDaniel, L. (2009). Relative increase of record high maximum temperatures compared to record low minimum temperatures in the US. *Geophysical Research Letters*, 36(23).
- Müller-Westermeier, G. (2012). Neueste Zahlen zum Jahr 2011 und zu den ersten Monaten des Jahres 2012. *Klima-Presskonferenz des Deutschen Wetterdienstes, Berlin*.
- Munich RE (2017). NatCatSERVICE database. Loss Events Worldwide 1980–2016. *Munich Reinsurance, Munich, Germany*. Available online at <http://natcatservice.munichre.com>; visited on August 5th 2017.
- Ng, W. and Panu, U. (2010). Comparisons of traditional and novel stochastic models for the generation of daily precipitation occurrences. *Journal of hydrology*, 380(1):222–236.
- Nguyen-Huy, T., Deo, R. C., An-Vo, D.-A., Mushtaq, S., and Khan, S. (2017). Copula-statistical precipitation forecasting model in Australia’s agro-ecological zones. *Agricultural Water Management*, 191:153–172.
- Peres, Y. and Shields, P. (2005). Two new Markov order estimators. *arXiv preprint math/0506080*.
- Piper, D., Kunz, M., Ehmele, F., Mohr, S., Mühr, B., Kron, A., and Daniell, J. (2016). Exceptional sequence of severe thunderstorms and related flash floods in May and June 2016 in Germany-Part 1: Meteorological background. *Natural Hazards and Earth System Sciences*, 16(12):2835.
- Racsko, P., Szeidl, L., and Semenov, M. (1991). A serial approach to local stochastic weather models. *Ecological modelling*, 57(1-2):27–41.
- Rahmstorf, S. and Coumou, D. (2011). Increase of extreme events in a warming world. *Proceedings of the National Academy of Sciences*, 108(44):17905–17909.

- Redner, S. and Petersen, M. R. (2006). Role of global warming on the statistics of record-breaking temperatures. *Physical Review E*, 74(6):061114.
- Robine, J.-M., Cheung, S. L. K., Le Roy, S., Van Oyen, H., Griffiths, C., Michel, J.-P., and Herrmann, F. R. (2008). Death toll exceeded 70,000 in Europe during the summer of 2003. *Comptes rendus biologies*, 331(2):171–178.
- Rudolf, B. and Rapp, J. (2003). The century flood of the River Elbe in August 2002: Synoptic weather development and climatological aspects. *Quart. Rep. German NWP-System Deutscher Wetterdienst*, 2:8–23.
- Rulfová, Z. and Kyselý, J. (2013). Disaggregating convective and stratiform precipitation from station weather data. *Atmospheric research*, 134:100–115.
- Schoelzel, C. and Friederichs, P. (2008). Multivariate non-normally distributed random variables in climate research—introduction to the copula approach. *Nonlin. Processes Geophys.*, 15(5):761–772.
- Schwarz, G. et al. (1978). Estimating the dimension of a model. *The annals of statistics*, 6(2):461–464.
- Selker, J. S. and Haith, D. A. (1990). Development and testing of single-parameter precipitation distributions.
- Shoji, T. and Kitaura, H. (2006). Statistical and geostatistical analysis of rainfall in central Japan. *Computers & Geosciences*, 32(8):1007–1024.
- Smith, R. B. (1979). The influence of mountains on the atmosphere. *Advances in geophysics*, 21:87–230.
- Solomon, S. (2007). *Climate change 2007-the physical science basis: Working group I contribution to the fourth assessment report of the IPCC*, volume 4. Cambridge University Press.
- Stephens, G. L. and Ellis, T. D. (2008). Controls of global-mean precipitation increases in global warming GCM experiments. *Journal of Climate*, 21(23):6141–6155.
- Stern, R. and Coe, R. (1984). A model fitting analysis of daily rainfall data. *Journal of the Royal Statistical Society. Series A (General)*, pages 1–34.
- Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, B., and Midgley, B. (2013). IPCC, 2013: climate change

- 2013: the physical science basis. Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change.
- Strauch, N. (2011). Das Meteorologische Observatorium Hohenpeißenberg. Brochure. Available online at https://www.dwd.de/SharedDocs/broschueren/DE/presse/standorte/obs_hpb_g_pdf?__blob=publicationFile&v=3; visited on August 5th 2017.
- Thom, H. C. (1958). A note on the gamma distribution. *Monthly Weather Review*, 86(4):117–122.
- Thompson, V., Dunstone, N. J., Scaife, A. A., Smith, D. M., Slingo, J. M., Brown, S., and Belcher, S. E. (2017). High risk of unprecedented UK rainfall in the current climate. *Nature Communications*, 8.
- Trenberth, K. E., Dai, A., Rasmussen, R. M., and Parsons, D. B. (2003). The changing character of precipitation. *Bulletin of the American Meteorological Society*, 84(9):1205–1217.
- US Dept. of Commerce (1979). Federal Meteorological Handbook No. 2. Synoptic Code. US Department of Commerce, US Department of Defense, US Department of Transportation. The instructions for encoding the land station surface synoptic code is available online at <http://atmo.tamu.edu/class/atmo251/LandSynopticCode.pdf>; visited on August 5th 2017.
- Vlček, O. and Huth, R. (2009). Is daily precipitation Gamma-distributed?: Adverse effects of an incorrect use of the Kolmogorov–Smirnov test. *Atmospheric Research*, 93(4):759–766.
- von Bomhard, P. (2014). Bachelorarbeit: Rekordstatistik von Niederschlägen. *unpublished*.
- Wergen, G. (2013). Records in stochastic processes—theory and applications. *Journal of Physics A: Mathematical and Theoretical*, 46(22):223001.
- Wergen, G., Hense, A., and Krug, J. (2014). Record occurrence and record values in daily and monthly temperatures. *Climate dynamics*, 42(5-6):1275–1289.
- Wergen, G. and Krug, J. (2010). Record-breaking temperatures reveal a warming climate. *EPL (Europhysics Letters)*, 92(3):30008.

- Wilby, R. L., Wigley, T., Conway, D., Jones, P., Hewitson, B., Main, J., and Wilks, D. (1998). Statistical downscaling of general circulation model output: a comparison of methods. *Water resources research*, 34(11):2995–3008.
- Wilks, D. S. (1989). Conditioning stochastic daily precipitation models on total monthly precipitation. *Water Resources Research*, 25(6):1429–1439.
- Wilks, D. S. (1999). Interannual variability and extreme-value characteristics of several stochastic daily precipitation models. *Agricultural and Forest Meteorology*, 93(3):153–169.
- Wilks, D. S. (2011). *Statistical methods in the atmospheric sciences*, volume 100. Academic press.
- Wilks, D. S. and Wilby, R. L. (1999). The weather generation game: a review of stochastic weather models. *Progress in physical geography*, 23(3):329–357.
- Woolhiser, D. A. and Pegram, G. (1979). Maximum likelihood estimation of Fourier coefficients to describe seasonal variations of parameters in stochastic daily precipitation models. *Journal of Applied Meteorology*, 18(1):34–42.
- Woolhiser, D. A. and Roldan, J. (1982). Stochastic daily precipitation models: 2. A comparison of distributions of amounts. *Water resources research*, 18(5):1461–1468.
- Yalcin, G. C., Rabassa, P., and Beck, C. (2016). Extreme event statistics of daily rainfall: dynamical systems approach. *Journal of Physics A: Mathematical and Theoretical*, 49(15):154001.
- Zhang, X., Zwiers, F. W., Li, G., Wan, H., and Cannon, A. J. (2017). Complexity in estimating past and future extreme short-duration rainfall. *Nature Geoscience*, 10(4):255–259.

Acknowledgements

First of all I would like to thank my supervisor Prof. Dr. Joachim Krug for his constant support throughout the entire time of my master project. Thank your for the helpful discussions and for reading this manuscript.

I would like to give special thanks to Prof. Dr. Yaping Shao for the opportunity to perform this interesting master thesis and his immediate agreement to be my supervisor.

Additionally I want to thank my supervisors as well as my colleagues at Deutsche Rück AG for their support and understanding of doing my job and my master thesis in parallel.

Thanks a lot to my friend Rika Langen for reading this manuscript and her useful remarks.

At the end I am most grateful for my friends and family, especially my parents who constantly supported me during my whole studies. Most importantly I would like to thank my wife Inga who supported me in every respect throughout my master thesis.

Versicherung an Eides statt

Hiermit versichere ich an Eides statt, dass ich die vorliegende Arbeit selbstständig und ohne die Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten und nicht veröffentlichten Schriften entnommen wurden, sind als solche kenntlich gemacht. Die Arbeit ist in gleicher oder ähnlicher Form oder auszugsweise im Rahmen einer anderen Prüfung noch nicht vorgelegt worden. Ich versichere, dass die eingereichte elektronische Fassung der eingereichten Druckfassung vollständig entspricht.

Ort, Datum

Unterschrift