Precipitation Extremes in Western Europe: a Comparative Analysis of Historical Trends

MSc Thesis

L.D. Fokke

Delft University of Technology

TU Delft

Precipitation Extremes in Western Europe: a Comparative Analysis of Historical Trends

MSc Thesis

by

L.D. Fokke

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Monday July 15, 2024 at 2:00 PM.

Student number: 4875680
Project duration: February 13, 2024 – July 8, 2024
Thesis committee: dr. ir. R. van der Ent, TU Delft, supervisor
dr. P. Mares-Nasarre, TU Delft, supervisor

Cover: Image created by AI (Microsoft Bing)

TU Delft

Preface

This master thesis is the final step for the completion of my studies at Delft University of Technology in order to obtain the degree of Master of Science in Water Resources Engineering. The topic of my research is related to changing patterns in extreme precipitation. During my study in Environmental Engineering, the topic of climate change its effect on precipitation extremes was part of several courses. I realised that in order to prepare our society and water systems for these precipitation extremes, a good understanding of them is necessary. It was a pleasure for me to explore the very complex system of precipitation behaviour. I have improved my skills in analysing large amounts of data, which will be very useful in my later career.
I hope my research will contribute to a better understanding of historical and future precipitation extremes and the effects climate change may have on them.
I would like to thank my main supervisor, Ruud van der Ent, for his guidance throughout the process of writing this master thesis. During our meetings, you asked questions about my ideas, which gave me the opportunity to improve my work each time. You learned me how to be critical to my work and how to structure this critical way of thinking in scientific research. Secondly, I would also like to thank my supervisor, Patricia, for your availability and the feedback you gave me. I really enjoyed the way you shared your knowledge related to extreme value analysis. I would also like to thank my family, friends, and boyfriend who supported me in various ways during my study and especially during this final part of writing my master thesis.
Lastly, I give thanks to the Lord, who gave me the strength to fulfill this study. I honor Him with the words of Psalm 147: *"Sing to the Lord with thanksgiving; sing praises on the harp to our God, Who covers the heavens with clouds, who prepares rain for the earth, who makes grass to grow on the mountains."*

L.D. Fokke
Delft, July 2024

Abstract

Rare precipitation extremes are projected to increase relatively more than common extremes under future climate change. However, it is unconfirmed whether these projected differences in relative change are observable in historical precipitation data. Therefore, this study investigated whether rare precipitation extremes have changed differently from more common ones during the period 1951 to 2010, in Western Europe. The differences in change between rare and common extremes ($D_{\text{rare-common}}$) and its correlation with temperature is analysed, using daily temperature and precipitation data. The $D_{\text{rare-common}}$ values are calculated based on the relative change of a specific return level estimated by three methods: the Full series, the Generalized Extreme Value (GEV) and the Metastatistical Extreme Value (MEV) distribution. A correlation with daily temperature is obtained by analysing first all day temperatures and compare these with the pattern in the $D_{\text{rare-common}}$ values. Secondly, the temperatures at the days of the annual extremes is analysed.

The results indicate that, on average for Western Europe, historical precipitation extremes have increased in magnitude by 3.0 up to 5.4%, depending on the method. Regional variations are present in Western Europe, showing negative relative changes for the most southern regions. A $D_{\text{rare-common}}$ for Western Europe equal to 1.0 and 2.3%pt is obtained, for respectively the MEV and GEV method, indicating that the rarest precipitation extremes have increased more in magnitude compared to the common extremes. However, a comparison of extremes based on the time series length using the Full series indicated no difference in relative changes for return levels of 0.1 and 10 years. A statistically significant correlation between the $D_{\text{rare-common}}$ and temperature is obtained between the results of the GEV and Tmax and MEV method and Tmin. This correlations suggests observing a relative high $D_{\text{rare-common}}$ value in a region in which the change in maximum temperature is low and the change in the minimum temperature is high, depending on the method used. Analysing the temperatures during annual extremes showed an increase in both the minimum and maximum temperature on the days of the annual extremes, despite a shift from warmer to colder months in which annual extremes occur. When dividing the annual extremes into common and rare, it was concluded that a higher increase in temperature is observed during common annual extremes compared to the rare annual extremes.

It is recommended to analyse this phenomenon in more recent years. To understand the behaviour of precipitation extremes in greater depth, other driving factors causing precipitation to become extreme could be studied. Examples of these driving factors include atmospheric water vapor and the temperature in regions where precipitation is generated.

Contents

Nomenclature

Abbreviations

Abbreviation	Definition
AMS	Annual Maxima Series
CDF	Cumulative Distribution Function
CRS	Coordinate Reference System
DOF	Degrees of Freedom
EVA	Extreme Value Analysis
Full	Full series frequency analysis
GEV	Generalized Extreme Value
MEV	Metastatistical Extreme Value
PWM	Probability Weighted Moments
std	standard deviation

Symbols

Symbol	Definition	Unit
$C_{\text{rel,Tx}}$	Relative change for a return period Tx between the two periods	%
D_{Tb-Ta}, x	Difference in $C_{rel,Tb}$ and $C_{rel,Ta}$ for method x	%pt
C_j	Weibull scale parameter	-
F	Non-exceedance probabilities (Gumbel)	-
i	ranking of observation values	-
n_j	number of wet events	-
n_{pyr}	average number of observations per year (365.25)	days
p	probability of exceedance	-
P	Precipitation	mm
T	Temperature	°C
$Tmin$	Minimum temperature	°C
$Tmax$	Maximum temperature	°C
Tx	x year return period level	mm/day
w	Weighting factor to account for area	-
w_j	Weibull shape parameter	-
$x0$	starting guess numerical solution MEV	-

Greek symbol	Definition	Unit
ζ_m	MEV-Weibull CDF	-
μ	GEV location parameter	-
ξ	GEV shape parameter	-
ρ	GEV scale parameter	-

1 Introduction

1.1. Motivation

Precipitation is the most crucial water source for sustaining life on Earth in many regions worldwide. While a certain amount of precipitation is necessary for life, excessive amounts can lead to societal issues, such as flooding and landslides (Nikolopoulos et al., 2015; Tradowsky et al., 2023). For this reason, high return levels of extreme precipitation act as design criteria for hydraulic infrastructures or the control of extreme floods. To obtain values for these high return levels, a good understanding of the behaviour of historical precipitation extremes is required. Therefore, several previous studies focused on the long-term changes in these extremes. These studies observed increases in extreme precipitation, for both the more common and rare extremes (Alexander, 2016; Groisman et al., 2005; Min et al., 2011; Sun et al., 2021). Recent climate models predict an increase in precipitation extremes in the coming centuries, due to global warming (Myhre et al., 2019). Accurate estimates of the evolution of the precipitation extremes in the future are crucial for disaster risk understanding and management. As stated before, the increase in extreme precipitation is often studied for two different types of extremes: common and rare. Common extremes can be expressed in terms of return periods, often within a year (Borodina et al., 2017; Sun et al., 2021), or in percentile-based threshold values, ranging higher than the 90th percentile up to the 99.9th (Fischer & Knutti, 2016; Mishra & Singh, 2010). These studies indicate an increase in the magnitude of common extremes. Rare extremes are defined in relation to design boundaries for infrastructure or flood management (Tradowsky et al., 2023), often spanning multidecadal time periods (Chan et al., 2018; Hodnebrog et al., 2019).
To predict future occurrences of extreme precipitation events, a good understanding of the long-term patterns of historical precipitation extremes is crucial. To analyse rare extreme precipitation events (return periods > 100 years), due to limited availability in recorded data, statistical methods are applied to infer extremes that have not been observed yet in the data. The extreme precipitation is described by the tail behaviour of the distribution, which is the upper part of a probability distribution. A heavy tail represents more frequent larger magnitude precipitation events compared to the lighter tails. A wrong choice of the tail distribution can lead to incorrect predictions of rainfall for large return periods (Papalexiou et al., 2013). As design criteria are often based on rare extreme precipitation predictions, a good understanding of the performance of extreme value analysis is crucial.

1.2. Previous studies

While many studies have indicated that precipitation extremes will increase, recent research has also demonstrated variability in the relative increase of extreme precipitation. Gründemann et al. (2022) showed that the rarest precipitation extremes are projected to experience a greater relative increase in magnitude compared to common ones. By the end of this century, daily land precipitation extremes could experience magnitude increases ranging from 10.5% to 28.2% for annual events and between 13.5% and 38.3% for centennial events, respectively for low and high emission scenarios (see Figure 1.1). This research compared two 30-year periods (1971-2000 and 2071-2100) based on historical data and predictions based on 25 CMIP6 (Coupled Model Intercomparison Project Phase 6) models.

Notably, Africa and regions around and just north of the equator are anticipated to experience disproportionately higher occurrences of the rarest rainfall extremes. Although the underlying mechanism for this phenomenon remains unclear, the findings of Gründemann et al. (2022) have significant implications for the establishment of engineering standards.

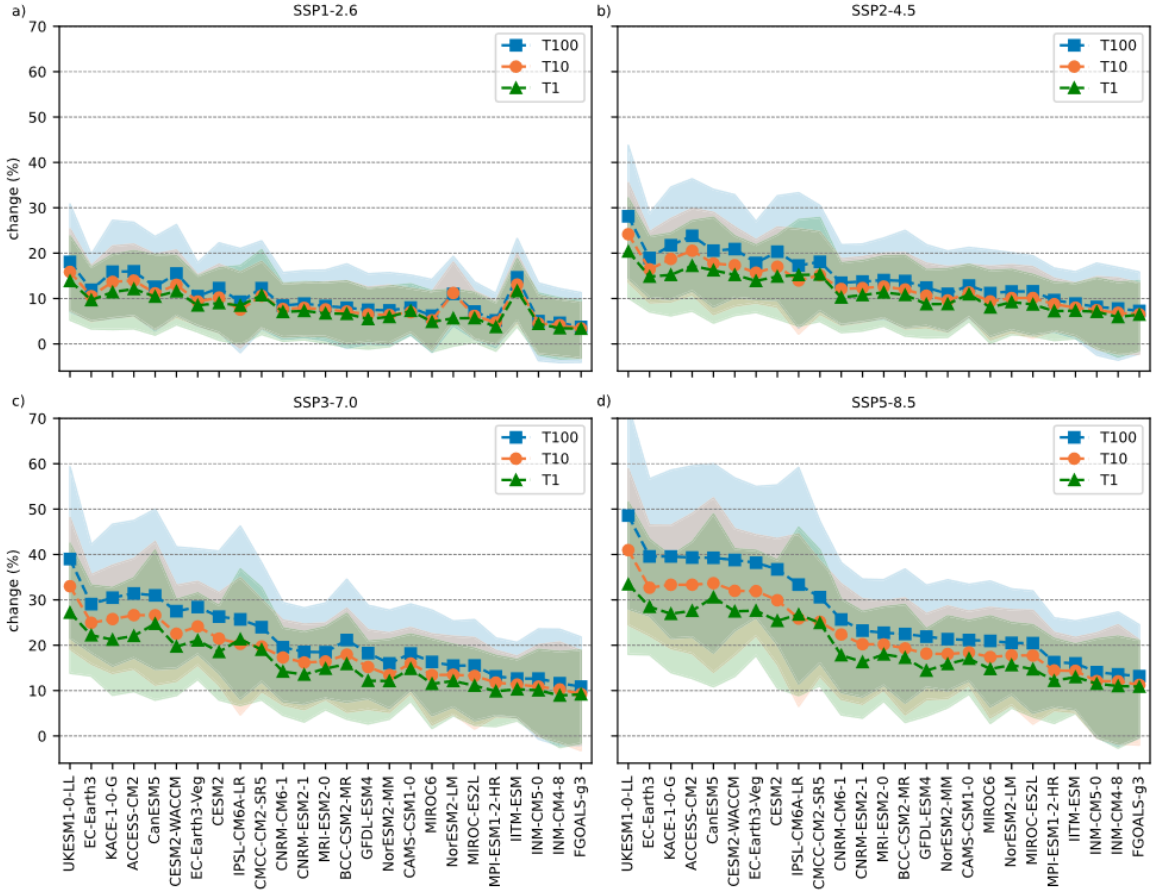


Figure 1.1: Result study by Gründemann et al.(2022) on relative increase of rare extreme rainfall events under future climate change. The figure indicates global relative change of precipitation extremes for each individual climate model. Each plot represents a different climate scenario. The blue, orange and green line indicates the 1, 10 and 100 year return levels, respectively.

The results of Gründemann et al. (2022) regarding these relative increases in extreme precipitation, based on models for future climate scenarios, show the expected behaviour of global precipitation extremes. However, as these expectations are based on climate models, it is not confirmed that these differences in relative change are also observable in historical precipitation data. Previous studies have shown the increase in both common and rare precipitation extremes. However, there are fewer studies on the differences between 'common' and 'rare' extremes for futures scenarios (Aalbers et al., 2018; Gründemann et al., 2022; Li et al., 2021; O'Gorman & Schneider, 2009; Pendergrass & Hartmann, 2014). While some of these studies combine historical data with climate-based predictions, the observed trend is not analysed based on historical data. Further knowledge about these patterns in historical data can confirm the conclusions drawn by Gründemann et al. (2022) and can indicate whether the obtained change in extreme precipitation has already started or if it is solely predicted by climate models.

1.3. Research question

This study will extend the research on future precipitation extremes by analysing historical data to confirm if rare precipitation extremes have been increasing more than the common ones. For this study, the following research question is defined:
What are the differences in the relative changes between the magnitude of common and rare precipitation extremes for historical daily observations in Western Europe over the period 1951 to 2010 and can these be associated with changes in daily temperature?

The data used in this research are precipitation and temperature observations from six countries in Western Europe: The Netherlands, Germany, France, Sweden, Norway, and Spain. Within the time frame of 1951 to 2010, two periods of 30 years are be selected for comparison. The analysed data covers a 1-day time scale.
The main question can be divided in the following sub-questions, which are also visualised in Figure 1.2:

1. *What are the relative changes in common and rare precipitation extremes in the periods 1951-1980 and 1981-2010?*
2. *How do the relative changes in precipitation extremes behave with respect to each other for both common and rare extremes in the periods 1951-1980 and 1981-2010?*
3. *What is the behaviour of daily temperatures in Western Europe within the period 1951 to 2010, and can this be associated with the difference in relative change in precipitation extremes?*

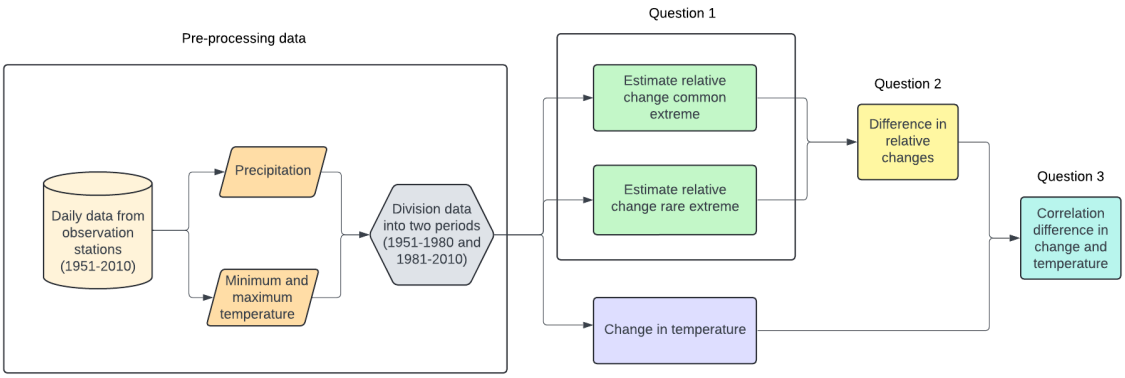


Figure 1.2: Flowchart of research questions. The three sub-questions are indicated as well as the pre-processing of the data.

1.4. Report structure

In this report, Chapter 2 describes the theoretical framework regarding precipitation and temperature data, as well as background information on extreme values analysis. The third chapter outlines the materials and methods used to address the research question. Chapter 4 presents the results of the analysis, while the discussion of these results can be found in Chapter 5. The final part of this report contains the concluding remarks by addressing the main research question in Chapter 6.

2

Theoretical Framework

2.1. Historical precipitation data

Analysing historical precipitation data requires an understanding of the quality and quantity of the collected data. While nowadays a lot of sources are available for precipitation data, such as radar data collected by satellites, the main ground-based source for historical rainfall data is the use of rain gauges. Many rain gauges are located in Europe and the US, which show the highly heterogeneous distribution of precipitation data. A substantial percentage of worldwide land is not covered by local precipitation data collection (Kidd et al., 2017). While around 150,000 rain gauges exist, they have not been operating very continuously and simultaneously (New et al., 2001). Despite the quality of the collected data by rain gauges is quite accurate at a certain locations, it does not show a high spatial coverage in many regions. The advantage of gauge measurements is that at certain locations long records are available compared to satellite measurements, which have data for a maximum of 25 years (New et al., 2001).

2.2. Development Extreme Value Analysis methods

In the last century the first methods for extreme value analysis were developed. To extrapolate observation data, statistical methods were introduced, starting with the Gumbel distribution. Compared to the Gumbel distribution, the later developed Generalized Extreme Value (GEV) was fitted to block maxima and it uses an additional parameter for more flexibility (Hosking & Wallis, 1993). This extra parameter enables the GEV to account for different tail behaviours, which is useful for accurate extreme precipitation predictions. While many other extreme value theories (EVT) are developed that only uses on a small fraction of the available observations, the recently developed Metastatistical Extreme Value distribution (MEV) accounts for the full 'ordinary' underlying events, using all observations above a certain threshold to predict the parameters. Therefore, the MEV is able to estimate higher return periods outside the period of recording with a lower uncertainty than traditional methods in EVT (Zorzetto et al., 2016). The drawback of this method is the bad estimation of low return periods with respect to the time series length.

2.3. Evolution of precipitation extremes

Precipitation extremes are increasing in more areas globally than they are decreasing (Donat et al., 2013). As stated in paragraph 2.1, there is limited availability of precipitation data for long time series globally. Despite this limiting factor, numerous studies have analysed the behaviour of extreme precipitation events (Alexander, 2016; Donat et al., 2016; Groisman et al., 2005; Min et al., 2011). While these studies differ in the dataset used, applied methods, or definition of extreme precipitation, they all conclude that there are more increases in extreme precipitation than decreases globally. The study by Min et al. (2011) concluded that approximately 65% of the globe has positive trends for annual maximum daily precipitation. The study indicate that on a global or continental scale, extreme precipitation events have been increasing in both frequency and intensity.

2.4. Trends in temperature

Global temperature patterns show consistent warming trends over the past century, particularly in the last few decades. The greatest increase in temperature is observed during the colder months (Donat et al., 2013). The most significant warming trends in temperature are those calculated based on daily minimum temperature. There is a reduction in the number of cold days and an increased frequency of warm days. This trend in temperature increase is more pronounced over the period 1951-2010 compared to 1901-2010 (Donat et al., 2013). The IPCC (2023) presented the change in global temperature over the period 1850-2020, clearly indicating an increase in the average global surface temperature (see Figure 2.1).

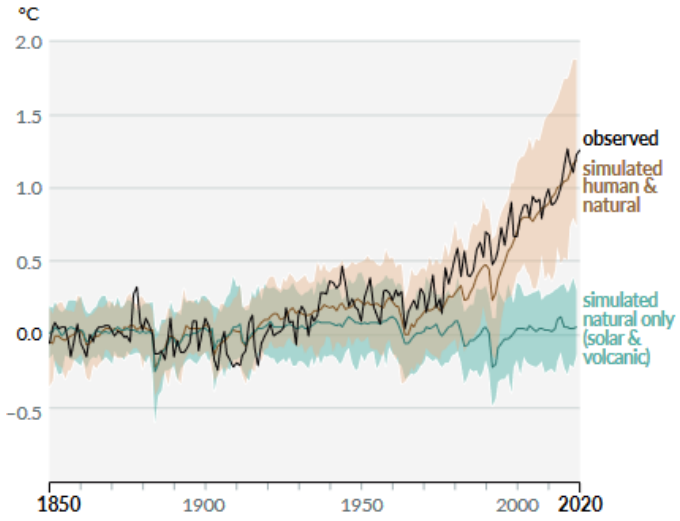


Figure 2.1: The change in global surface temperature from the IPCC. The observed (black line) and simulated (blue and orange) temperature change between 1850 and 2020 is visualised (Intergovernmental Panel On Climate Change, 2023).

As this research focuses on precipitation and temperature patterns for countries in Western Europe, it is relevant to analyse the warming trends in Europe rather than global temperature. The study by Twardosz et al. (2021) examined temperature trends based on data from 210 weather stations within the period 1951-2020, coinciding with the time frame of this research. The analysis revealed a continuous increase in air temperature, with a linear trend observed since 1985 (see Figure 2.2). During the period 1951-1985, the temperature trend remained relatively constant. The increase since 1985 is evident in all seasons except for autumn. The most significant intensification is observed in the northeast of Europe, particularly during winter and spring, while the highest increases during summer are observed in the southwest.

Several studies have examined the relationship between temperature increase and precipitation extremes. Fischer and Knutti (2015) found that with a global warming of 0.85°C, approximately 18% of daily precipitation events can be attributed to the observed temperature increase relative to the pre-industrial era. They also showed that with an expected temperature increase of 2.0°C, the fraction of precipitation extremes rises to 40%, indicating a strong relationship between extreme precipitation and temperature increase, particularly for the most rare extremes. Additionally, models project more heavy precipitation in response to increasing global temperatures (Fischer & Knutti, 2015). For a temperature increase of 2.0°C, the probability of the most extreme events increases by a factor of 1.5 to 3.0. The study by Westra et al. (2013) investigated the trends in annual maximum daily precipitation and their association with globally averaged near-surface temperature. The study concluded that there is a statistically significant association with global temperature, at a rate ranging between 5.9% and 7.7% per Kelvin, depending on the method used.

Numerous studies assume changes in extreme precipitation are constrained by the amount of precipitable water in the atmosphere, which is affected by changing water vapor content. The relationship between water vapor content and temperature changes is explained by the Clausius-Clapeyron relation, which states that an increase of 1°C leads to a 7% change in daily precipitation extremes. Previous studies also conclude that the heaviest precipitation events will occur when the maximum moisture

content is fully precipitated out. The study by Pall et al. (2007) found the greatest agreement with the Clausius-Clapeyron relation at the mid-latitudes.

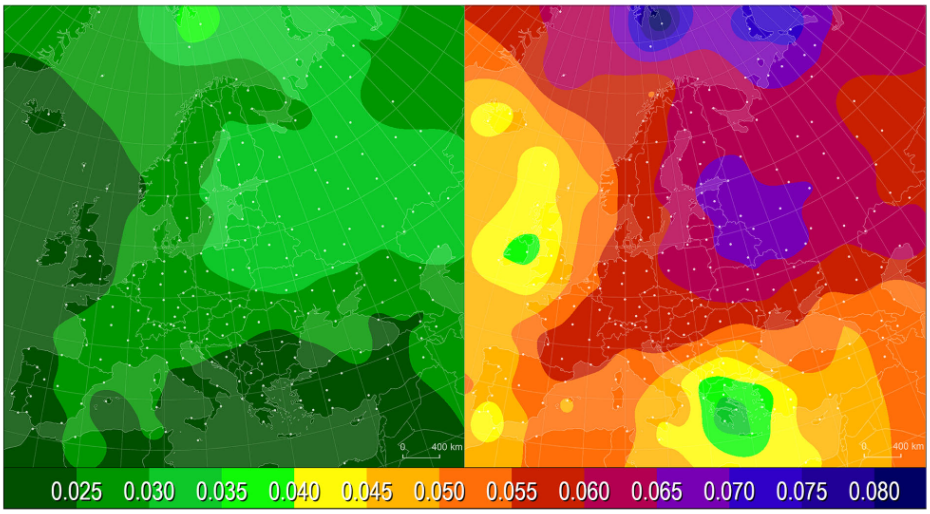


Figure 2.2: The observed annual air temperature trends (°C/year) in Europe over the time span 1951-2010 (left) and 1985-2010 (right), studied by Twardosz et al. (2021).

Gründemann et al.(2022) analysed differences in relative changes for extreme precipitation events based on four climate scenarios. The associated global temperature increases are visualised in Figure 2.3. This figure shows the relative change in global surface air temperature relative to an average of two periods for five different climate scenarios. The values presented in Figure 1.1 should be interpreted based on the corresponding global temperature increase related to each climate scenario.

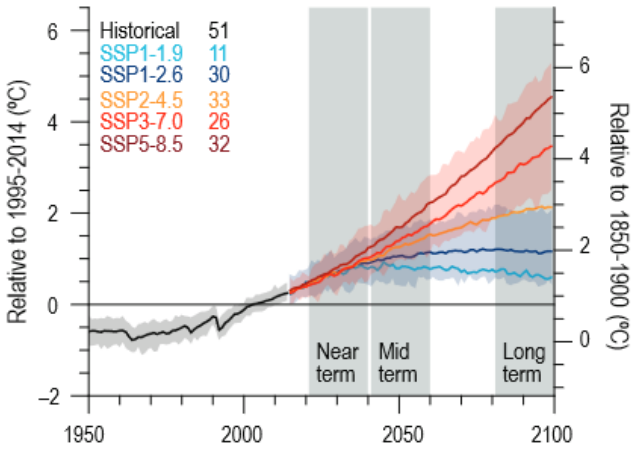


Figure 2.3: Global surface air temperature changes relative to the 1995–2014 average (left axis) and relative to the 1850–1900 average for five different climate scenarios. Black lines indicates the observed historical temperature. (Intergovernmental Panel On Climate Change (Ipcc), 2023)

3
Methodology

3.1. Data sets

To unravel the differences in relative changes in extreme precipitation in Western Europe, precipitation and temperature data is collected from six different countries: Germany, France, Sweden, Norway, Spain, and The Netherlands (Figure 3.1, dark grey countries). These countries cover the whole latitude range of Western Europe. Stations located outside the European continent are removed, as several countries have observation stations in overseas territories that are not relevant for the analysis of Western Europe.

For this study, precipitation and temperature data are sourced from weather observation stations. Precipitation data is represented by the accumulated precipitation over a 24-hour period. Temperature data includes the daily minimum and maximum temperature measured in °Celsius. The data used in this study is directly measured by equipment at the observation stations; no satellite or modeled data is used.

To minimise uncertainty arising from data quantity, particularly for studying extreme precipitation events, a minimum number of observation station datasets per country in Western Europe is required. Precipitation data from a country is included if the average area covered by a station is no larger than 15,500 km² on average, as this approximates the minimum number of stations necessary to obtain an areal-averaged precipitation value (Xie & Arkin, 1995). Additionally, to account for snow, snow depth data are collected from each station, in case this is not already included in the precipitation data. Appendix A provides a description of all data per country.

3.2. Pre-processing observation data

As observation data from meteorological stations can include missing or incorrect data, a quantity and quality step is implemented. First, in the pre-processing step the precipitation data is checked. It is verified whether both rainfall and snow data are available for each station. Secondly, the availability of data for at least 60 years, starting from 1951, is confirmed, as two periods of 30 years are analysed in this study. If 60 years of data is available for a station, the dataset is examined for the amount of missing data within this time frame. To maintain statistical significance, a maximum of 10% missing data is permitted (Papalexiou & Koutsoyiannis, 2013; Zhai et al., 2005). After the quantity checks, the data undergoes quality assessment. Negative values are discarded since they are physically impossible. Subsequently, outliers are identified and removed based on a criterion of the mean plus 20 times the standard deviation, to eliminate unrealistic values.

Temperature data are not available on all locations with precipitation data (Table 3.1) in Western Europe. Nevertheless, to study the general association between difference in relative change in precipitation extremes and temperature in the same area, all temperature observation data is used from the same six Western European countries. The temperature data is checked on its availability to confirm data is available between 1951-2010, with a maximum of missing data of 10%. Missing values in the remaining data are filled by linear interpolation.

After applying all data quantity and quality checks, the number of observation stations available is shown in Figure 3.1. From the 18089 stations with precipitation data a number of 1858 is selected. For

temperature data 323 from the 13069 stations are used in this study. To provide an indication of the density, the number of stations is compared to the area of the corresponding country (see Table 3.1).

Table 3.1: Overview data availability after quantity and quality control. The second column contains the source of the data for each country. Per country the number of stations with precipitation and temperature data is given, with the density of these stations with respect to the total area of the country. The last row contains the total number of precipitation and temperature datasets.

Country	**Source**	**stations P**	P **(km²/station)**	**stations T**	T **(km²/station)**
Netherlands	KNMI	226	185	4	10463
Germany	DWD	855	418	145	2465
France	Météo-France	574	961	84	6341
Sweden	SMHI	108	4169	31	14525
Norway	MET	62	6213	27	13757
Spain	Aemet	33	15334	32	15813
Total		1858		323	

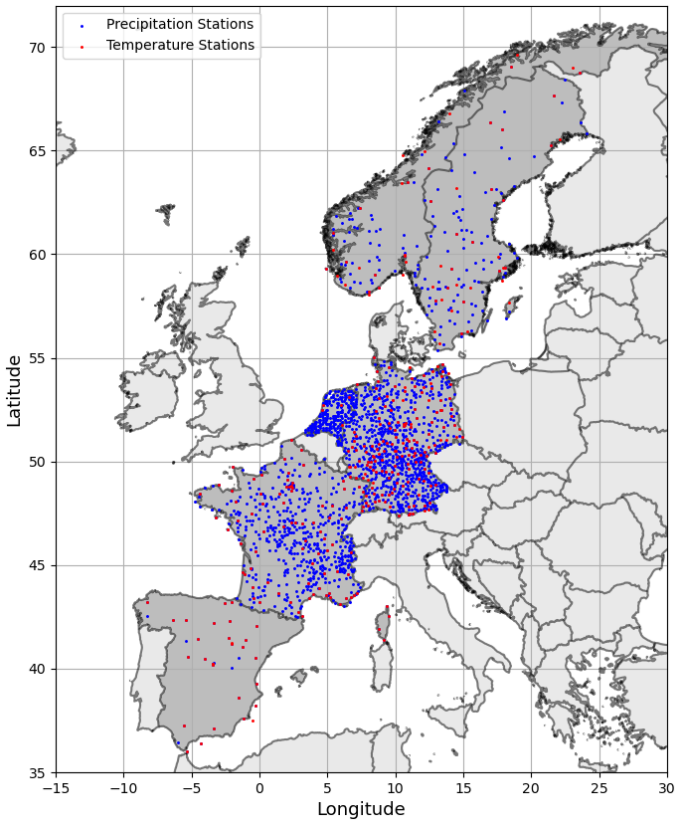


Figure 3.1: Geographical visualization of all observation stations used in this study after quantity and quality analysis. Blue dots indicate station with precipitation data. Red dots indicate an available station with temperature data. The dark grey countries are included in this study.

3.3. Data processing

Two periods of 30 years are extracted from the datasets that are chosen for comparison of relative changes in extremes (Figure 3.2). A 30-year time frame is selected, as it is the minimum standard for analysing climate changes (Aalbers et al., 2018; Arguez & Vose, 2011; Kendon et al., 2018). The period spans from 01-01-1951 to 31-12-2010, providing 60 years of data. Within this time frame, two periods are selected for relative comparison (see Figure 3.4). The first period, period 1, covers 01-01-1951 to 31-12-1980. The second period spans the next 30 years, from 01-01-1981 to 31-12-2010. To account for snow, precipitation that falls in solid form is added to the precipitation data as its melting

water equivalent. This adjustment is made only if snow data is not already included in the precipitation records. As a guideline, snow is recalculated using the rule '1 cm of snow equals 1 mm of precipitation' (Christensen, 2020).

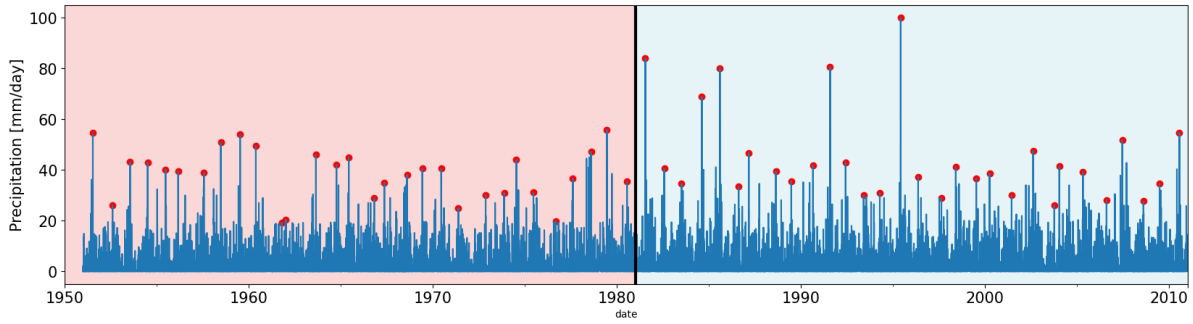


Figure 3.2: Example of time series of precipitation data from an observation station, spanning from 1951 to 2010. The blue line indicates all observed precipitation events. The time series is divided into two periods: 1951-1980 and 1981-2010, visualized by the red and blue segments, respectively. The red dots represent the annual maxima, which are the most extreme events of each calendar year.

3.3.1. Weighing factor

To account for the non-uniform data availability,to give all observation stations equal weight according to their surrounding area, the Thiessen method is applied (Han & Bray, 2006), which divides an area into parts according to defined points. Each point is assigned to an area that encompasses all other points closest to it (Figure 3.3). The weight assigned to each station is determined by the area of its corresponding Thiessen polygon, relative to the total area of all six countries. To visualise the data and to obtain the Thiessen polygons, shapefiles from Eurostat are used. These shapefiles have a scale of 1:$3 \cdot 10^6$ and are based on the most recent data from 2020, downloaded in the coordinate system CRS:4326 ('Countries - GISCO - Eurostat', n.d.). The weighted mean of a parameter is obtain as follows:

$$w_{average} = \frac{\Sigma wx}{\Sigma w} \tag{3.1}$$

where w is the fraction of the total area of the countries and x the parameter being weighted. Weighting is applied solely for averaging all stations, not per country. Therefore, by using the weighted average, the mean value of all stations is relative to the corresponding area in Western Europe.

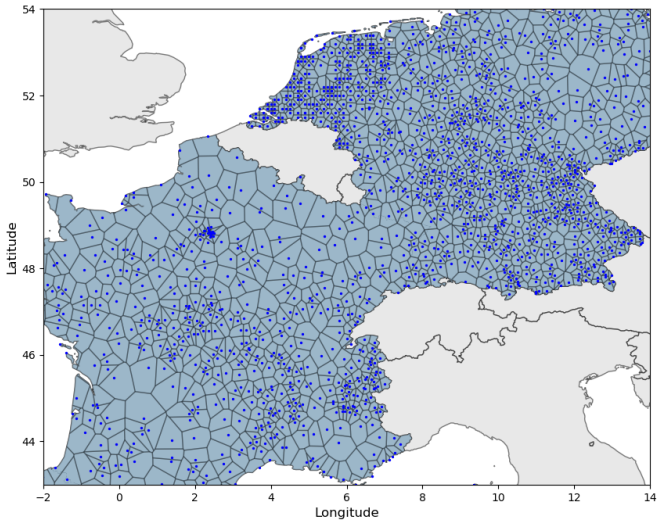


Figure 3.3: Visualization of the Thiessen polygons in central Western Europe for precipitation observation stations. Black dots represents observation stations. The weighing factor is based on the size of the polygon.

3.4. Extreme Value Analysis

In this research, the focus lies on analysing the relative change in extreme precipitation. Previous studies, as mentioned in Chapter 1, indicated that the rarest precipitation extremes are projected to increase relatively more in magnitude than common extremes by the end of the 21st century. The rare extremes in this study are calculated using the MEV and compared to the common extreme calculated using all-day percentiles. To achieve a direct comparison with this previous study, the same methodology is adopted as a starting point (see Figure 3.4, green and orange box).

To study the phenomenon of relative changes in precipitation extremes more comprehensively, different approaches are employed in this research. While the focus is on studying the relative changes using T100 and T1 values, the principle that rarer precipitation extremes increase relatively more than common ones can be applied to other return periods. Therefore the relative changes are calculated for the range of 0.1 up to 1000 return level[1], to understand to behaviour of all extremes.

In this study, three different methods are utilised. The reason for employing multiple methods is that the values obtained using extreme value analysis can be influenced by the method itself, potentially leading to under- or overfitting. By applying different methods, the impact of a specific method can be understood, which is particularly useful for future studies in this field. All methods are applied separately for each station, after which the results are averaged for Western Europe. The average difference in relative change is also visualised per latitude or grid cell. For all methods the common extremes are estimated by using the Full series frequency analysis, as this is the most accurate method to calculate return levels lower then the time series length, as extreme value distributions are designed for estimating return levels greater than the length of the time series. The rare extremes are estimated using the Full series, GEV and MEV (Figure 3.4).

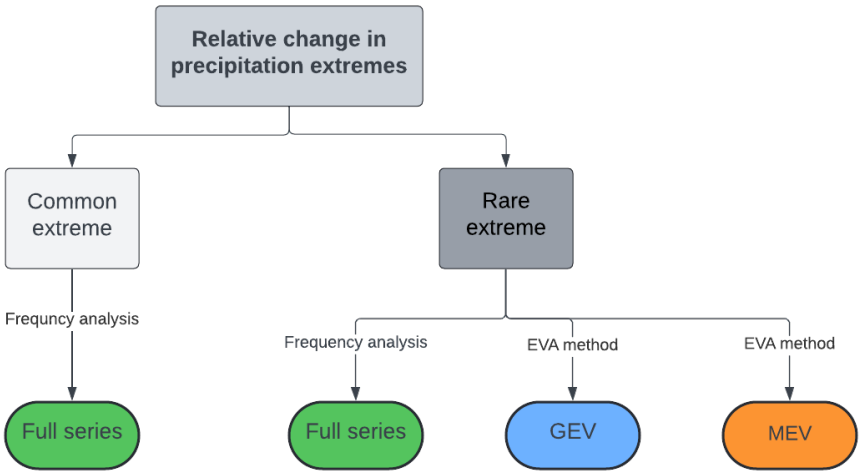


Figure 3.4: Flowchart of the different Extreme Value Analyses. The common extreme is estimated by the Full series frequency analysis. The rare extremes are estimated in three different ways: the Full series, GEV and MEV. The type of input data for each analysis is presented between the brackets

3.4.1. Full series frequency analysis

To analyse relative changes in return periods for the time series, a frequency analysis is employed (Figure 3.4, green boxes). This method relates the magnitude of precipitation extremes to probabilities of occurrence. It calculates return periods up to the length of the time series, which in this case is 30 years for both periods. The frequency analysis utilises the time series of daily observations, and as a first step, the precipitation data is ranked in descending order. The corresponding rate of exceedance is then calculated as follows:

$$p = \frac{i}{N} \tag{3.2}$$

[1]Depending on the method used. Full series ranges from 0.1 to 30 year. GEV and MEV are calculated from 0.1 to 1000 years.

where i is the ranking of the descending precipitation value and N the number of observations in the time series. The corresponding return period is calculated by:

$$T_{return} = \frac{1}{p}\frac{1}{n_{pyr}} \tag{3.3}$$

with p being the probability of exceedance and n_{pyr} the average number of observations per year (365.25) to recalculate the values from daily to yearly return periods. The Full series directly uses all precipitation data and therefore it is the most accurate method to obtain return levels up to 30 years.

3.4.2. Generalized Extreme Value (GEV) distribution

An extreme value distribution to estimate the high return levels is the widely used Generalized Extreme Value (GEV) distribution (see Figure 3.4, blue box). Extreme Value Analyses (EVA) enables the calculation of return values for return periods higher than the length of the time series. The GEV distribution can take the form of a Weibull, Frechet, or Gumbel distribution, which adds flexibility compared to using only one type of distribution (Coles, 2001). The parameters of the GEV distribution are fitted to the data using the L-moments approach. The GEV is applied to the available precipitation data, using the annual maxima of each year. The cumulative distribution function (CDF) is given by:

$$G(z) = \begin{cases} \exp\left[-\left(1+\xi\frac{z-\mu}{\sigma}\right)^{-1/\xi}\right] & \text{if } \xi \neq 0 \\ \exp\left[-\exp\left(-\frac{z-\mu}{\sigma}\right)\right] & \text{if } \xi = 0 \end{cases} \tag{3.4}$$

with μ being to the location parameter, σ the scale parameter, and ξ the shape parameter. The annual maxima, i.e. the maximum value of each year, is used to fit the parameters. The behaviour of the tail of the distribution is described by the shape parameter. A positive ξ corresponds to the *Fréchet* distribution with a heavy tail, a ξ equal to zero refers to the *Gumbel* distribution, which is characterised by a light tail and a negative value of ξ corresponds with the reversed Weibull distribution, which has an upper limit in the tail.

3.4.3. Metastatistical Extreme Value (MEV) distribution

The second method used, to calculate return levels outside the range of the time series length, is the MEV distribution (Figure 3.4, orange box). This is the same method as applied in previous studies as mentioned before. The MEV method takes into account the full distribution of the underlying 'ordinary daily' events. It has been shown to outperform the GEV distribution when estimating return periods larger than the sample size, reducing uncertainty for the rarest extremes (Schellander et al., 2019; Zorzetto et al., 2016). The MEV considers the random process of event occurrence and the potentially changing probability distribution of event magnitudes across different blocks of time. The input data for this analysis includes all events above a threshold value of 1 mm/day, which are considered wet days. These events are treated as Weibull right-tail equivalents (Wilson & Toumi, 2005). The Weibull parameters used in the MEV method are estimated for each year based on all selected wet events using the probability weighted moments method (PWM), which assigns greater weight to the tail of the distribution (Greenwood et al., 1979). The cumulative distribution function is then formulated as the MEV-Weibull distribution:

$$\zeta_m(x) = \frac{1}{M}\sum_{j=1}^{M}[1 - e^{-(\frac{x}{c_j})^{w_j}}]^{n_j} \tag{3.5}$$

where M is the total number of years and j is the year in range j = 1,2,..,M. The variable x represents the input data. The $C_j > 0$ is the Weibull scale parameter, $w_j > 0$ is the Weibull shape parameter and n_j is the number of wet events observed in the year j. The used Python script is described in the package 'mevpy' (Zorzetto, 2023).

To obtain return levels, the MEV method utilises a starting guess x_0 for the numerical solution. As this starting guess varies for each station, an algorithm is included to apply different x_0 values based on the Annual Maxima Series (AMS).

3.5. Relative changes in magnitudes of the extremes

By estimating the return levels using the different methods for the two periods, the difference in relative change can be calculated for each station. This comparison involves two main steps: comparing two periods of 30 years to obtain values for the relative change in extreme precipitation events and comparing differences in values corresponding to two return levels, i.e. rare and common (see Figure 3.4).
The functions used to describe the relative change is:

$$C_{\text{rel,t}} = \frac{Tt_{\text{period2}} - Tt_{\text{period1}}}{Tt_{\text{period1}}} \tag{3.6}$$

where $C_{\text{rel, t}}$ is the relative change for a return period T between two periods of 30 years, where period 1 equals a time frame of 30 years which is compared to period 2 which equals another 30-years period later then period 1 (Gründemann et al., 2022).
The equation to calculate the difference in relative change between common and rare extremes in percent points is described as follows:

$$D_{\text{rare-common,x}} = C_{\text{rel,rare}} - C_{\text{rel,common}} \tag{3.7}$$

For which the $C_{\text{rel,rare}}$ and $C_{\text{rel,common}}$ are the relative changes, calculated with eq. (3.6), of respectively the rare and common extreme return level in percents for an observation station location and x is the method used. The D-value indicates the difference in these relative changes for the station location. A positive D-value states that the rare precipitation events have seen a greater relative increase in magnitude compared to more common precipitation events. The mentioned analysis is illustrated in Figure 3.5.

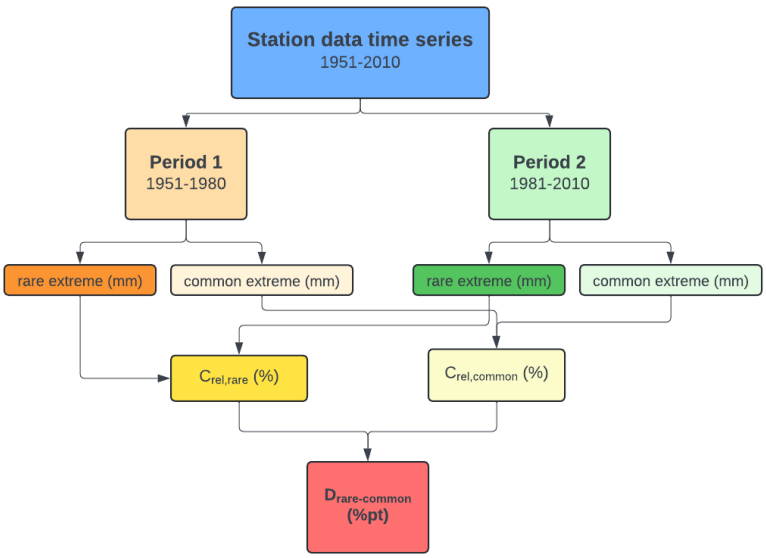


Figure 3.5: Flowchart of the method to obtain the difference in relative change between common and rare extreme precipitation return levels.

A graphical visualisation of the process for obtaining the difference in relative change, as previously explained, is provided in Figure 3.6. The graph displays two extreme value distributions: one for the period 1951-1980 (red line) and another for the period 1981-2010 (blue line). These distributions are based on parameters obtained from the input data through the extreme value method. On the x-axis, the return level is shown, ranging from 0.1 to 1000 years, while the y-axis represents the corresponding precipitation value. Both distributions estimate the values for all presented return levels, extrapolating the data beyond 30 years. The change in a specific return level, for example, the 1-year and 100-year return levels, is indicated by two black arrows. The difference in the estimation of the return values is an

indication of the relative increases $C_{\text{rel,rare}}$ and $C_{\text{rel,common}}$. If the relative change indicated by the right-side arrow is larger compared to the relative change indicated by the left-side arrow, the $D_{\text{rare-common,x}}$ is positive. The $D_{\text{rare-common,x}}$, as defined in Equation 3.7, is obtained by subtracting the relative change of the left-side arrow from the right-side arrow. This method of calculating C_{rel} and $D_{\text{rare-common,x}}$ is applied to each individual station using each of the three aforementioned methods.

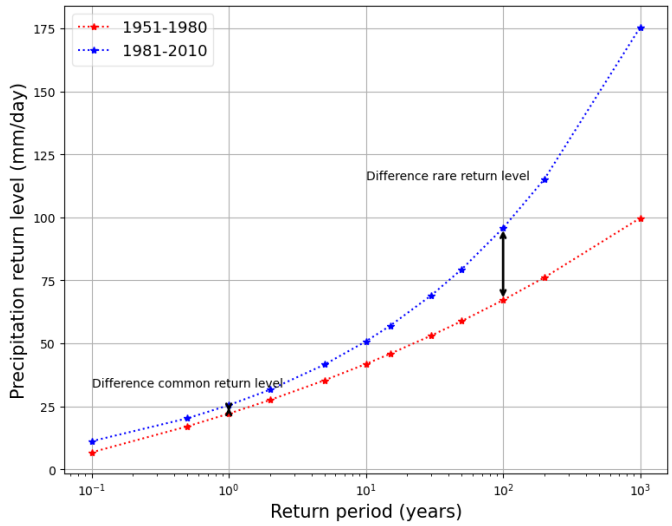


Figure 3.6: Graphical representation of the procedure for calculating the relative change $C_{\text{rel,rare}}$ and $C_{\text{rel,common}}$. The red line is the distribution fit for the first period (1951-1980) and the blue line represents the distribution fit for the second period (1981-2010). The range of return levels is from 0.1 to 1000 years. The two black arrows indicate the difference for both the common return level (T1) and the rare return level (T100). This difference, relative to the value of the first period, equals the C_{rel} values. If the relative change of the right-side arrow is larger compared to the relative change of the left-side arrow, the $D_{\text{rare-common,x}}$ is positive. The distributions shown in this example are based on the GEV.

While in this study more than one rare and common extreme is calculated to obtain a complete overview of the underlying difference between return levels, one single D-value is calculated for each station as an average for whole Western Europe. Table 3.2 provides an overview of what is defined as common and rare extremes for each type of analysis to calculate the final D-value. To obtain a comparison between rare and common extremes based on the Full series frequency analysis, two other extremes are defined as is done for the GEV and MEV, for which the common extreme is defined as T0.1 and the rare extreme as T10.

Table 3.2: Definition of common and rare extreme for each method to calculate the $D_{\text{rare-common,x}}$. The $D_{\text{rare-common,x}}$ presented in the results per method are based on these common and rare extremes.

Method	**common extreme**	**rare extreme**
Full	T0.1	T10
GEV	T1 (Full)	T100
MEV	T1 (Full)	T100

The difference in relative change is first calculated for each station separately. To further analyse the patterns in these results, the D-value is averaged on different scales, ending with a single D-value per method for Western Europe. D-values are calculated in the latitude direction, averaging the D-values of all stations for a moving window of 4 degrees latitude and per 2x2 grid cell.
To obtain a single D-value for Western Europe an weighted-average is taken of all stations. To address the sensitivity due to outliers in these values, bootstrapping is applied to calculate the D-values for each method using sampling with replacement. From the D-values for each stations random sampling is applied with sampling size equal to the degrees of freedom, which is equal to sample size minus 1. After sampling, the D-value is calculated based on the sampled values. This process is repeated 1000 times, to obtain a distribution of D-values for each method.

3.6. Statistical significance $D_{\text{rare-common}}$

3.6.1. Significance $D_{\text{rare-common}}$ for Western Europe

To test whether the difference in relative change ($D_{\text{rare-common,x}}$) for Western Europe is statistical significant, two significance tests are applied to account for both the spatial and temporal correlation.

Spatial variability

The first test accounts for the variability in the locations used to calculate the average D-value for Western Europe. First, the degrees of freedom (DOF) are determined based on the sample size. The DOF are equal to the sample size minus 1. Based on the DOF sampling is applied on both the $C_{\text{rel,rare}}$ and $C_{\text{rel,common}}$ values, with replacement. As all stations have different corresponding areas, and also contribute with a different weight to the final D-value, the w-factor of each station is used as a weight for the possibility of each value to be sampled. Based on the sampled values from the $C_{\text{rel,rare}}$ and $C_{\text{rel,common}}$, a t-test is applied to test whether the means of the two distributions are significant different from each other. This tests whether the relative change for a return level is not equal to zero, i.e. the mean is not equal to zero. The null hypothesis H0 is therefore that no relative change for a return level is present. This H0 is rejected in case the p-value is below 0.05. The equation used for this t-test is as follows:

$$t = \frac{x - \mu}{\frac{s}{\sqrt{n}}} \tag{3.8}$$

where x is the observed mean of the sample, μ the assumed mean, s the standard deviation and n is the sample size. This sampling and testing is repeated 1000 times. It is counted how often for each method the 1000 t-test results show a significant difference.

Temporal variability

The second step testing the statistical significance of the area-weighted average D-value for Western Europe accounts for the temporal variability. As two time frames are used in this study (1951-1980 and 1981-2010) the temporal effect of these periods should be tested. To account for this, random sampling from both periods is applied with replacement. For each set of sampled years, the D-value is calculated. This sampling is repeated 350 times, to obtain a distribution of D-values for each method. To test whether the mean of all 350 D-values is not equal to zero, a t-test is applied based on a normal distribution, determined by the 350 D-values, and the standard normal distribution. This standard normal distribution represents a D-value equal to zero. If the normal distribution of the 350 D-values with respect to this standard normal distribution does not show a statically significant result, it states that the D-value does not show a significantly difference from zero. Therefore, the null hypothesis that D equals zero can be rejected if the p-value is below 0.05.

3.6.2. Correlation of $C_{\text{rel,rare}}$ and $C_{\text{rel,common}}$

To check if stations with a higher increase for T1 have also a high increase for T100, the Spearman's rank correlation is applied. This is a non-parametric method to evaluate the agreement of a monotonic relationship of two ranked variables, in this case the $C_{\text{rel,rare}}$ and $C_{\text{rel,common}}$ values. To quantify the correlation the coefficient r_s is used, to state the strength and direction of a relation ship. The coefficient r_s varies between -1 and +1, where the absolute value of this coefficient is an indication of the strength of the correlation. An absolute value of 1 corresponds to a perfect monotonically correlation (Bon-Gang, 2018). In order to assess whether the observed result is significant, a hypothesis test is performed. The null hypothesis H_0 is such that there is no significant correlation between the relative change in $C_{\text{rel,common}}$ and $C_{\text{rel,rare}}$ for a specific station, so statistical significance is determined for $p < 0.05$. The definition of the coefficient r_s is given by:

$$r_s = 1 - \frac{6\Sigma(R_i - i)^2}{n(n^2 - 1)} \tag{3.9}$$

where R_i is the rank i observation X_i of the first variable, i the rank number of the corresponding second variable and n is the total number of observations.

3.7. Correlation with daily temperature

Changes in precipitation extremes are often assumed to be constrained by changes in the amount of precipitately water in the atmosphere. Because this amount is affected by the temperature, the difference in relative change between common and rare precipitation extremes are associated to changes in temperature. In this study, the daily minimum and maximum temperature[2] is analysed (description of data was provided in Chapter 3.1). As stated in Table 3.1 and Figure 3.1 (red dots), a number of 323 station data is used for the temperature analysis, with corresponding weighting factors to account for the spatial heterogeneity.

3.7.1. All day temperature development

As a starting point, the general development of the temperature in Western Europe is analysed, to state whether there is temperature increase in Western Europe or not. The change in temperature is obtained by calculating the mean daily temperature for both periods 1951-1980 and 1981-2010. For each of the 323 stations data sets the mean temperatures of all days within the 30 years is calculated. As a next step, for each station the absolute difference in temperature between both periods is calculated. This difference indicates whether a relative warming is presented at a station location. This difference is also presented as a relative change. To analyse the daily temperatures in this way an indication is obtained if there is a change in temperature between the two periods 1951-1980 and 1981-2010. In this way a correlation can be calculated with the previous calculated D-value, as this is also a comparison of estimates of return levels between both periods. To clarify if there is any pattern in the temperature change for Western Europe, the obtained mean daily temperatures per station are also averaged per 4 degrees latitude. A moving window is applied which average all values around a specific latitude, ranging from -2 to +2 degrees. As the previously calculated D-values are also averaged in the same way, the correlation between the change in temperature and the D-value per latitude can be calculated using the previous mentioned Spearman's rank test (Bon-Gang, 2018). To visualise this correlation between temperature and $D_{\text{rare-common}}$, a linear trend is plotted through the points indicating the $D_{\text{rare-common}}$ and the corresponding difference in temperature. To obtain a single value for temperature increase in Western Europe an area-weighted average value for the temperature is taken from all stations for both periods.

3.7.2. Temperature during annual precipitation extremes

As the $D_{\text{rare-common}}$ is an indication for the relative change between different precipitation extremes, it is logical to study the temperatures at the days where precipitation extremes occur, as these temperatures can be the driving force for these extremes. Therefore, the temperature at these days is extracted and the average values for each year are analysed to understand the trend in these values. The data used for this analysis can be obtained from observation stations which have both the precipitation and temperature data available. The 267 out of the 323 stations are illustrated in Appendix E.

3.7.3. Temperature during common and rare annual extremes

To understand if changes in the temperature are different for the common and rare extremes, the relationship between the magnitude of the precipitation extremes and the temperature is analysed. For both periods, 1951-1980 and 1981-2010, the average temperature during both common and rare extremes is calculated, to analyse what the relative changes of the corresponding temperatures is and if these changes are different for the common and rare extremes. The common and rare extremes for this temperature analysis are defined as the lowest five annual maxima for the common extremes and the highest five annual maxima as the rare extremes. This analysis can only be done on the the length of the time series as no temperature values can be related to extrapolated values for precipitation.

3.7.4. Supplementary analysis of dew point temperature

In addition to the comprehensive temperature analysis, a supplementary analysis is conducted on the dew point temperature data. While this type of data is not the primarily focus of this study, it is included in the results as several studies suggest a correlation exists between the local dew point temperature and extreme precipitation events (Aleshina et al., 2021; Pérez Bello et al., 2021; Zhang et al., 2019). This supplementary analysis is performed using data from three of the six countries: Sweden, The

[2]For clarification the minimum and maximum temperature are further stated as 'temperature'.

Netherlands and Germany. A total number of 53 datasets containing hourly dew point data are available. The locations of these stations are provided in Appendix E.

To analyse changing patterns in the dew point data, the average daily mean dew point temperature is first calculated by resampling the hourly data to daily mean temperatures. Secondly, all daily mean temperatures within both periods (1951-1980 and 1981-2010) are averaged to obtain a single dew point temperature for each period. To analyse regional variations, the average temperature per 4 degrees latitude is calculated. These average temperatures are then used to calculate the correlation with the previously calculated D-values per latitude. For this dew point calculation, only the D-values for the latitude degrees with available dew point data are used. The correlation coefficient is calculated using Spearman's rank coefficient, as previously explained.

Analysing the dew point temperature on days with precipitation extremes is not feasible due to limited data availability. Therefore, the supplementary analysis presents an example of one station, for which a relation is visualised between the annual precipitation extremes and the corresponding mean dew point data on those specific days. To obtain a correlation, Spearman's rank correlation is calculated between the annual precipitation maxima and the dew point temperature for both periods.

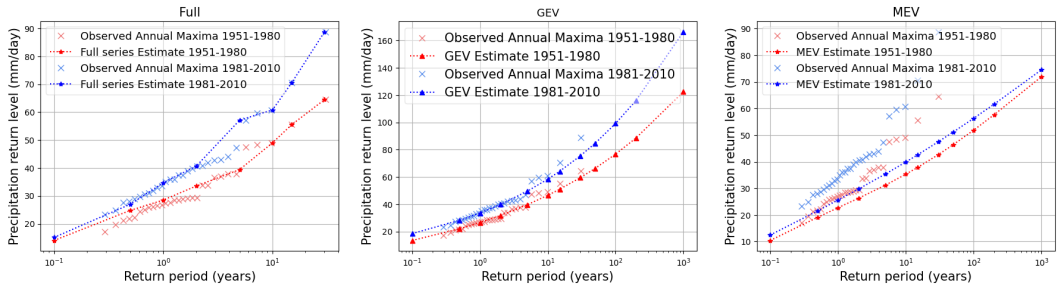
4

Results

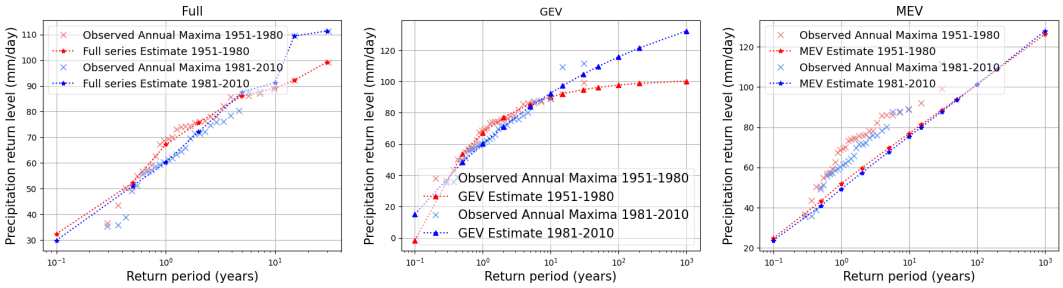
4.1. Relative changes in common and rare precipitation extremes

4.1.1. Relative changes per observation station

To compare the relative changes of return levels of precipitation extremes in Western Europe, the return levels are calculated based on the Full series, GEV, and MEV distributions for the precipitation data from each observation station. After dividing the data into two periods of 30 years (Figure 3.5), the distributions are fitted according to the methods described in Chapter 3.4. Two examples of these distributions are visualised in Figure 4.1, with additional examples provided in Appendix B. As a result of the frequency analysis (Full) and the extreme value analyses (GEV and MEV), each station yields two distributions per method (Figure 4.1, blue and red lines), which relate return periods to corresponding daily precipitation values. A precipitation amount can be linked to a particular return period. Due to the extrapolation in the GEV and MEV methods, the return values can be significantly influenced by the behaviour of the respective distributions. To understand the results from the extreme value analyses, the performance of both GEV and MEV distributions are evaluated and described in Appendix C.



(a) Observation station Netherlands

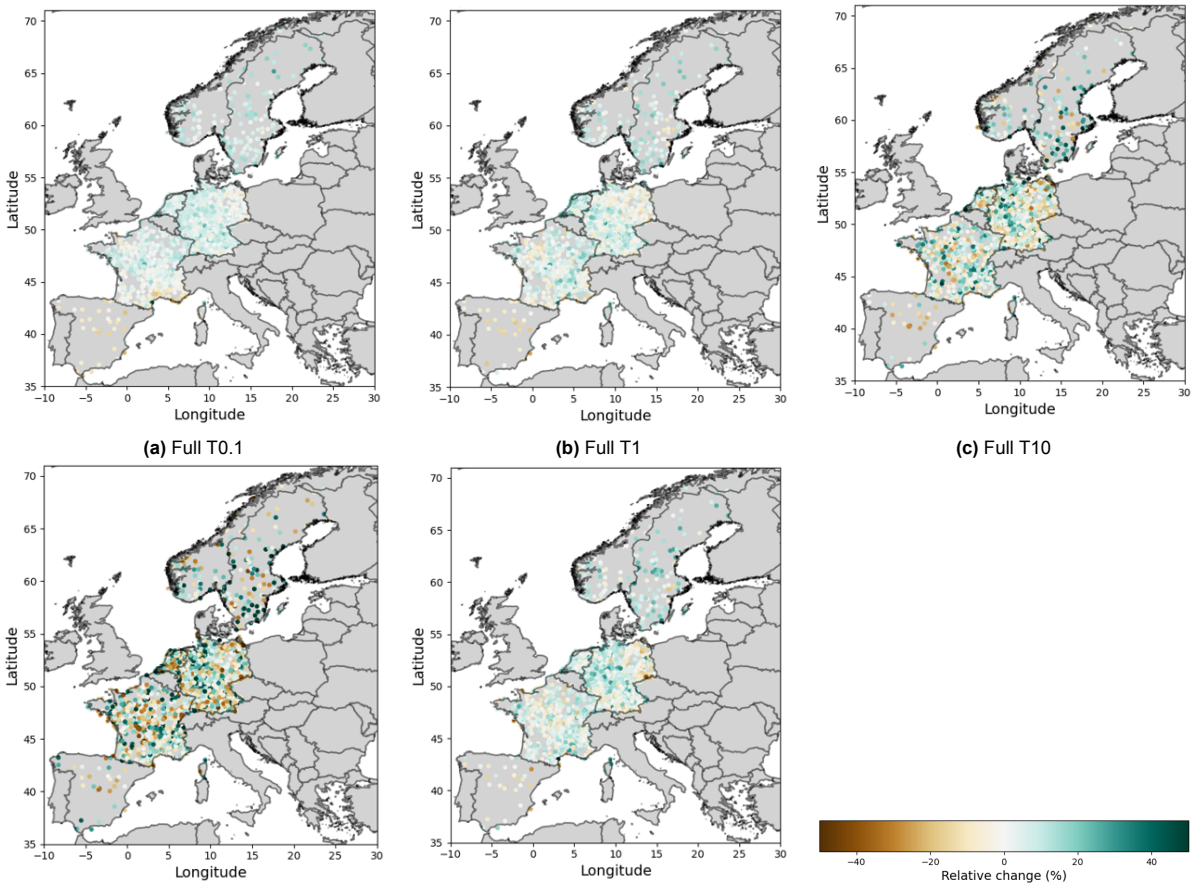


(b) Observation station Germany

Figure 4.1: Distribution fits for Full series, GEV, and MEV for a station in The Netherlands and Germany. The red lines indicate the fit for the first period (1951-1980), while the blue lines represent the fit for the second period (1981-2010). The red and blue x's denote the annual maxima for the respective periods, which are the input for the GEV analysis. The return periods (x-axis) range from T0.1 to T1000 for both the GEV and MEV. Due to the length of the time series, the Full series shows return levels up to T30.

The relative change of a return level can be obtained from the distributions by calculating the difference between the two values for a specific return period. A positive relative change means, visually, that the blue line (period 2) is above the red line (period 1) for a specific return period. In Figure 4.1a, an example of a station is shown where each blue line is above the red line for all three methods and all return periods. This indicates that, based on the results of the three methods, all return periods exhibit a positive relative change. In contrast, the example in Figure 4.1b does not show a consistent positive or negative relative change across all return periods. The Full series and GEV methods display varying positive and negative relative changes, while the MEV method shows almost no relative change for each return period. The procedure of fitting distributions is applied on all 1858 observation station data.

$D_{\text{rare-common}}$, which indicates the difference in relative change between common and rare extremes, illustrated in Figure 4.2 where the five relative changes, defined as common and rare extremes, for all stations are presented. Each dot represents the relative change (C_{rel}) for an observation station. Stations which have a brown color show a negative relative change for the given return level, while a green dot indicates a positive relative change.



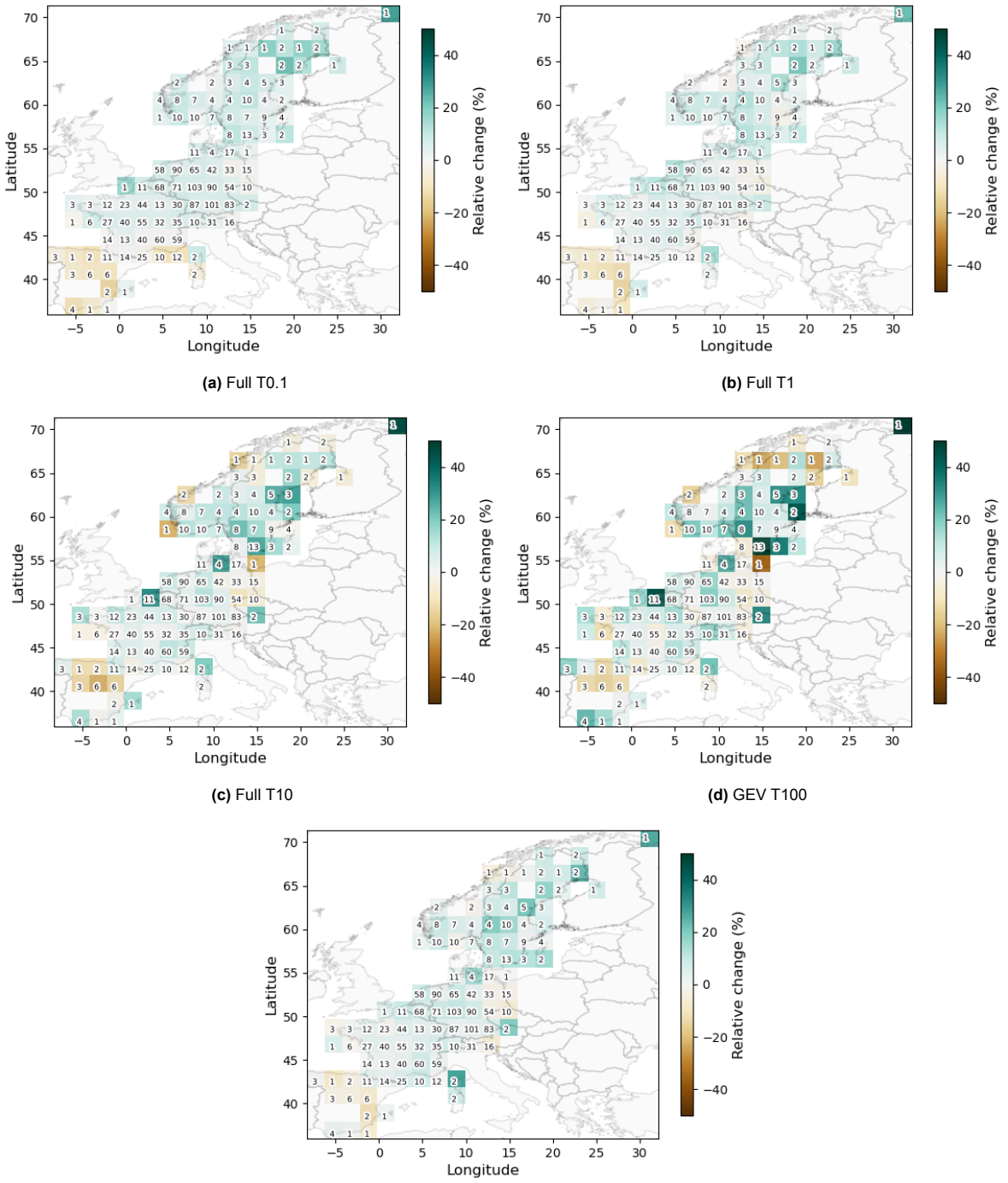
(a) Full T0.1 **(b)** Full T1 **(c)** Full T10

(d) GEV T100 **(e)** MEV T100

Figure 4.2: Geographical visualization of the relative change per observation station for **a** the Full T0.1, **b** the Full T1, **c** the Full T10, **d** the GEV T100 and **e** the MEV T100. Brown dots indicate a negative relative change; green dots represents stations with a positive relative change for the corresponding return level. The relative change is equal to the change in the precipitation corresponding to the specific return level between the periods 1951-1980 and 1981-2010.

The geographical pattern of these relative changes is further depicted in Figure 4.3. Each grid cell of 2x2 degrees represents the average value of all stations located within that region. The Full series exhibits clear patterns for T0.1 and T1: in southern Western Europe, negative relative changes are observed, indicating precipitation extremes became less extreme. The central regions of Western Europe show values close to zero, while the northern areas display clear positive relative changes. However, the pattern for T10 is less distinct, with some northern areas also showing negative values. The GEV results display negative values in the southern part, with some regions exhibiting slightly positive relative

changes. Northern areas show regions with high relative changes, but the most northern regions revert to negative values. Similarly, the MEV results for T100 show a pattern resembling that of the Full series, with negative values in southern regions, values near zero in the central areas of Western Europe, and positive relative changes in the northern regions.



(a) Full T0.1

(b) Full T1

(c) Full T10

(d) GEV T100

(e) MEV T100

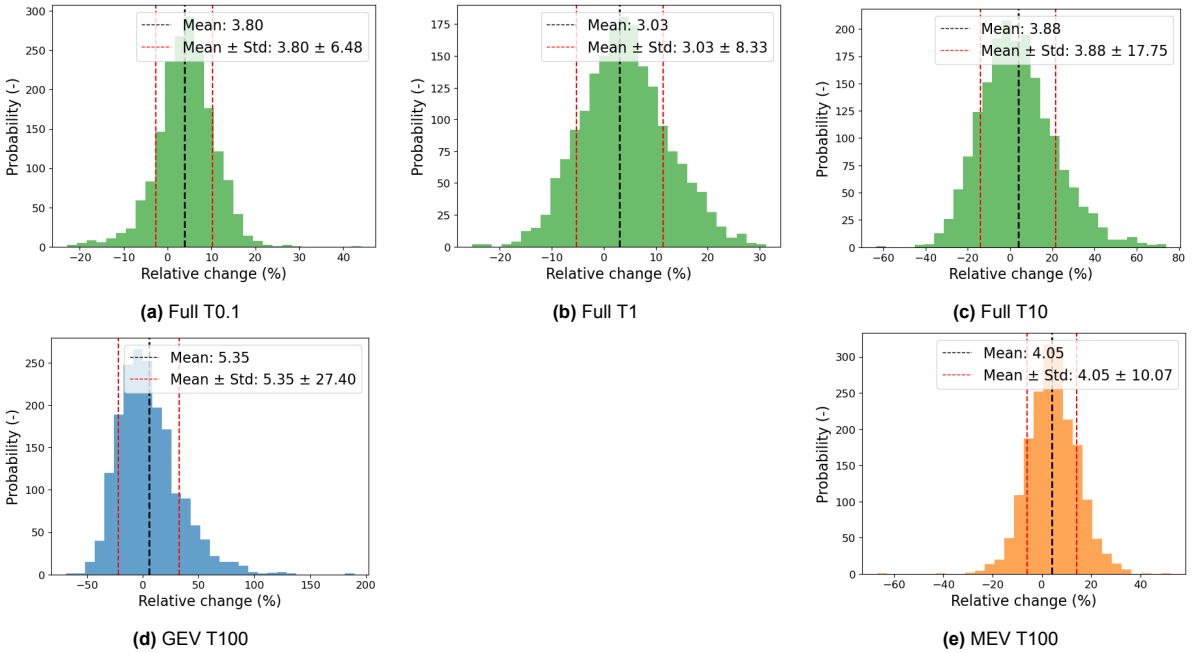
Figure 4.3: Plot of average relative change (C_{rel}) per 2x2 grid cell for **a** the Full T0.1, **b** the Full T1, **c** the Full T10, **d** the GEV T100 and **e** the MEV T100. The brown colors indicate an average relative decrease. A green color represents a relative increase. Each cell shows the mean of all stations within that region. The number in each grid is the number of stations present in this area. The five return level presented are used for the calculation of the $D_{\text{rare-common}}$ as stated in Table 3.2. The relative change is equal to the change in the precipitation corresponding to the specific return level between the periods 1951-1980 and 1981-2010.

The range of values for the relative changes for all stations, presented in Figure 4.2, are plotted in

histograms (Figure 4.4). The area-weighted average relative changes of all stations values are provided in Table 4.1. These values indicate relative changes in precipitation extremes between 3% and 5.4%. The low return level for the Full series and the MEV show lower standard deviations, whereas the Full T10 and the GEV T100 display higher standard deviations. This indicates a wide range in the values across all stations for these return levels, being either a negative or positive range.

Table 4.1: The area-weighted average relative changes for the return levels of interest. The standard deviation (std) indicates the variation in values across all observation stations. The range of this std is visualized in the histograms in Figure 4.4.

Method	Return level	Mean (%)	std (%)
Full	T0.1	3.80	6.48
Full	T1	3.03	8.33
Full	T10	3.88	17.75
GEV	T100	5.35	27.40
MEV	T100	4.05	10.07



(a) Full T0.1 **(b)** Full T1 **(c)** Full T10

(d) GEV T100 **(e)** MEV T100

Figure 4.4: Histogram showing the ranges of C_{rel}. The mean value presented in the legend represent the area-weighted C_{rel} value for Western Europe.

4.1.2. Relative changes of all return levels

To gain a comprehensive understanding of relative changes for other return levels, an overview of the return levels from common (T0.1-T1) to rare (>T100) is provided, as illustrated in the distributions in Figure 4.1. Figure 4.5 depicts the area-weighted average relative change for Western Europe, where the relative changes of individual stations are averaged based on their corresponding areas. The results of the three methods indicate a positive relative change above 2% for all return levels, suggesting that, on average, all return levels in Western Europe became more extreme during the period 1951-2010. The Full series (green line) varies between 2.5% and 5% for a return level range of T0.1-T30. The GEV reaches its lowest value of 2.5% at the return levels of T30-T50 and its maximum value of 14% at a return level of T1000. The MEV consistently shows a value of 4% for each return level.

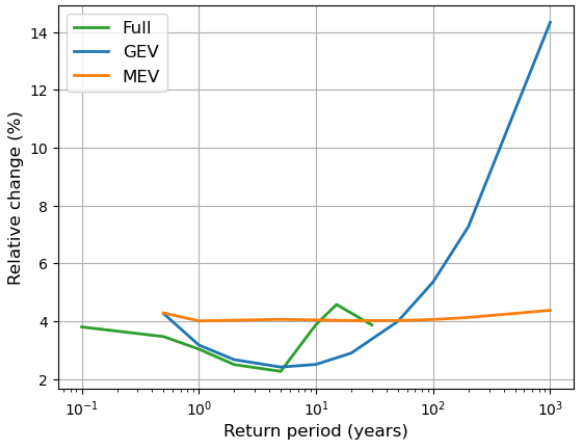


Figure 4.5: Area-weighted relative changes for Western Europe based on all observation stations. Relative changes are shown for return periods ranging from T0.1 to T1000. The green line represents the Full series, the blue line indicates the GEV estimation, and the orange line depicts the MEV distribution. The Full series shows values up to T30 due to the length of the time series.

4.2. Differences in relative change between common and rare extremes

To determine whether rare precipitation extremes have different relative changes compared to more common extremes in Western Europe, the entire range of return levels is first analysed. Subsequently, the exact $D_{\text{rare-common}}$ values are presented based on the extremes defined in Table 3.2.

4.2.1. Area-weighted average Western European $D_{\text{rare-common}}$ for all return levels

The values presented in Figure 4.5 represent the relative changes for all return levels, ranging from 0.1 to 1000 years. Figure 4.6 shows the differences in relative change for three three methods: Full series, GEV and MEV. The green line shows the difference with respect to the Full series 0.1 year return level. The blue and orange line represent the difference in relative change for, respectively the GEV and MEV, with respect to the Full series T1 value (as described in Table 3.2). The Full series shows negative D-values up to a return period of 10 years, after which the D-value increases up to 1.0%pt for the T10 return level, af which the D-values decreases again. The GEV starts at T2 at a negative difference in the change. The differences in relative changes became positive at a return period of T20, after which it continues to increase. The MEV shows a much more stable result, with the differences in relative change, with respect to T1, approximately equal to 1.0%pt for each return period. After T100 there is a slight increase up to 1.5%pt for a return period of T1000.

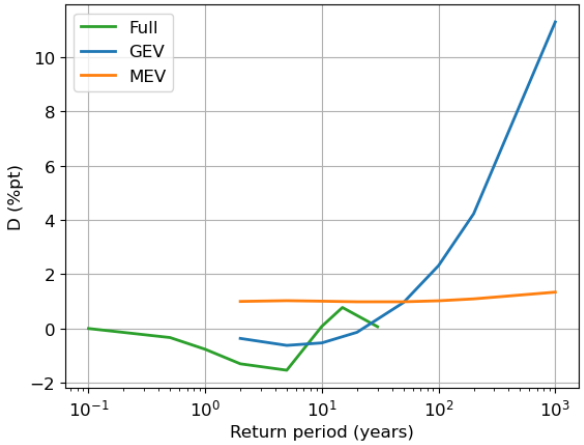


Figure 4.6: Difference in relative change with respect to the common extreme as defined in Table 3.2. The Full series is shown with respect to the T0.1. The GEV and MEV with respect to the Full series T1 value. The blue line indicates the Full series, the orange line the GEV estimation and the green line the MEV distribution. The D-values are area-weighted average values for Western Europe.

4.2.2. Area-weighted average Western European $D_{\text{rare-common}}$

Based on the previously values for $D_{\text{rare-common}}$ (Figure 4.6) the mean difference between common and rare extremes can be calculated. For each of the three methods, the D-value represents the difference in relative changes shown in Table 4.2. The difference in relative change between common and rare precipitation extremes, as calculated in previous studies (Gründemann et al., 2022), $D_{\text{rare-common,MEV}}$ equals 1.02 %pt. This value states that, on average for Western Europe, the rarest precipitation extremes have increased more in magnitude relative to the common extremes. The same conclusion can be stated if the GEV distribution is applied instead of the MEV, which gives a $D_{\text{rare-common,GEV}}$ equal to 2.32%pt. The Full series shows a value close to zero, which states that, based on all-day percentiles, there is no difference in the relative change between the common and rare extremes.

Table 4.2: Differences in the change of common and rare extremes for Western Europe (as defined in Table 3.2).

Method	$D_{\text{rare-common}}$ **(%pt)**
Full	0.08
GEV	2.32
MEV	1.02

As the $D_{\text{rare-common}}$ are averages of a range of all 1858 stations, the sensitivity of the average $D_{\text{rare-common}}$ due to its outliers is visualised in boxplots (see Figure 4.7). The GEV shows the highest variability, indicating that the GEV distribution is the most sensitive to outliers. This sensitivity is due to the fact that the GEV extrapolates annual extremes and is therefore sensitive to only a few observed precipitation events. Conversely, the MEV shows the lowest range in $D_{\text{rare-common}}$, suggesting that it is less affected by outliers compared to the other methods. A deeper clarification of the outliers in the GEV is provided in Appendix C.1.

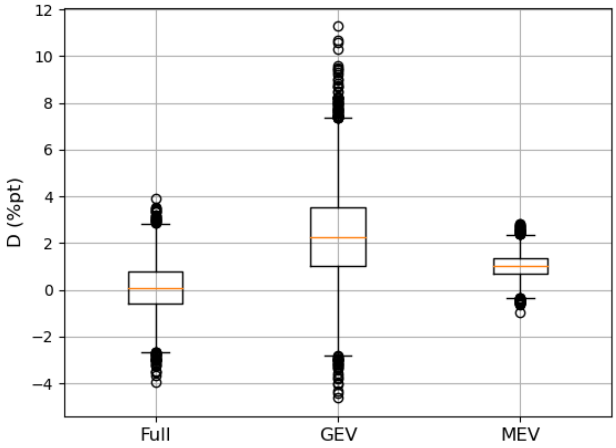


Figure 4.7: Boxplots illustrating the area-weighted average $D_{\text{rare-common}}$ obtained by bootstrapping for each method. The boxplots demonstrate the impact of outliers on $D_{\text{rare-common}}$ for Western Europe using random sampling from all stations.

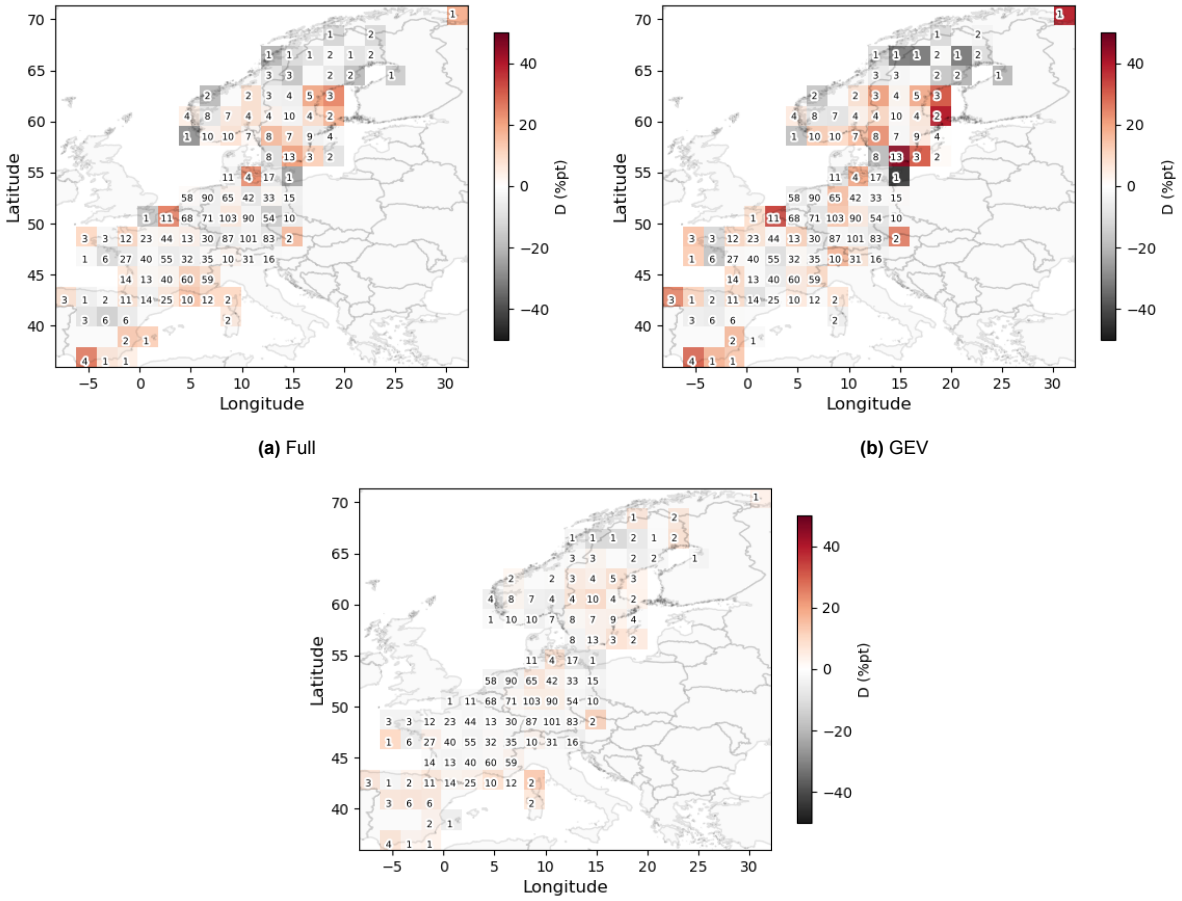
4.2.3. Regional variations in the $D_{\text{rare-common}}$

Hidden variations may exist in the area-weighted average $D_{\text{rare-common}}$ for Western Europe, as not all regions within Western Europe have a uniform behaviour regarding changes in precipitation extremes. To unravel these variations, the percentage of the total area of the six Western European countries which have positive or negative values for the $D_{\text{rare-common}}$ is examined for each method. This percentage is based on the corresponding area of each station within Western Europe as described in Chapter 3.3.1. The percentage values represent the fraction of the total area which demonstrates positive or negative D-values. The results indicate that, based on the Full series and GEV, more areas in Western Europe exhibit decreasing $D_{\text{rare-common}}$ values, whereas the MEV shows that 58% of the area demonstrates a positive $D_{\text{rare-common}}$. Despite minor variations in the percentage for each method, the proportion of areas with a positive $D_{\text{rare-common}}$ is around 50% (Table 4.3). This suggests a high level of agreement that, based on the total area of Western Europe, the rarest precipitation events will not experience the greatest relative increase in magnitude.

Table 4.3: Percentage of the total area of all six countries showing increasing and decreasing $D_{\text{rare-common}}$ per method during the periods 1951-1980 and 1981-2010.

Method	% increase	% decrease
Full	46.4	53.6
GEV	48.5	51.5
MEV	57.7	42.3

For a geographical analysis of $D_{\text{rare-common}}$ for each method, the average $D_{\text{rare-common}}$ per 2x2 grid cell is illustrated in Figure 4.8. Once again, these values generally have agreement, while there is not clear pattern of exclusively positive or negative values relative to latitude. In the northern regions of Sweden and Norway, the majority of grid cells show notably negative D-values, whereas the southern parts of both countries display a mixture of positive and negative values. Sweden shows predominantly positive D-values, while the south of Norway consistently shows negative D-values. Similar, high values and some negative values are observed in the southern part of Western Europe. The central part, comprising Germany, France, and The Netherlands, exhibits a mixture of negative and positive values, with the average values hovering around zero. This suggests a balance between the differences in relative changes in rare and common extremes in these regions.



(a) Full

(b) GEV

(c) MEV

Figure 4.8: Average $D_{\text{rare-common}}$ per 2x2 grid cell for **a** the Full series, **b** the GEV and **c** the MEV. Grey colors indicate an average negative D-value, while red colors represent positive D-values. The number in each grid cell denotes the number of observation stations used to calculate the average $D_{\text{rare-common}}$ for the corresponding cell.

The average $D_{\text{rare-common}}$ per latitude is visualised in Figure 4.9, using a moving window in the latitude direction with a width of 4 degrees. Positive $D_{\text{rare-common}}$ value are observed in northern countries Sweden and Norway for each method. In the central part of Western Europe, values close to zero are observed for each method, while the southern regions of France and Spain display positive values again, indicating the rarest precipitation extremes increase relatively more than the common ones in these areas. This pattern becomes more apparent when considering the average of all three methods (see Figure 4.9, represented by the black dashed line).

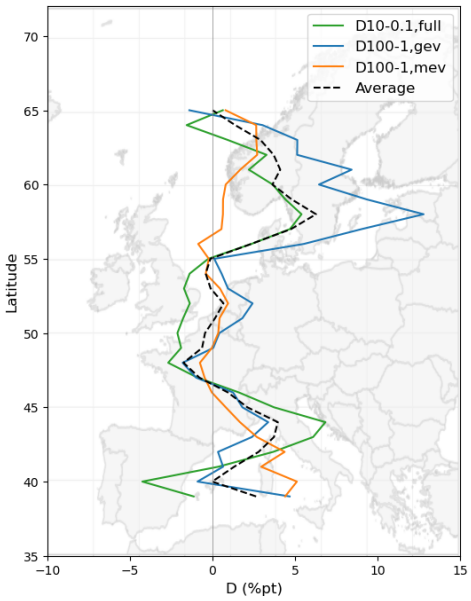


Figure 4.9: Average $D_{\text{rare-common}}$ per latitude, with a window size of 4 degrees. Map shows the average D-value (%pt) for each method. The latitude ranges from 37 to 69 degrees. The black dashed line represents the average of all three methods.

4.3. Statistical Significance of the $D_{\text{rare-common}}$ for Western Europe

To test the statistical significance of the result that the average $D_{\text{rare-common}}$ for Western Europe (presented in Table 4.2) is positive, various tests are applied to account for spatial and temporal variability.

4.3.1. Spatial variability

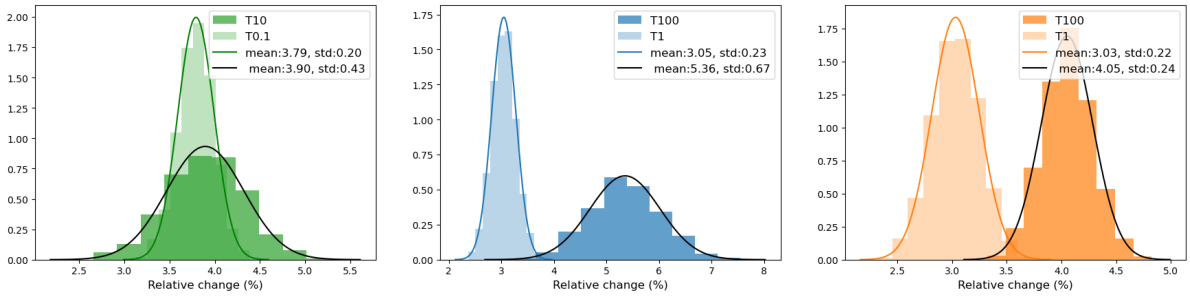
The first test accounts for the spatial variability of the 1858 observation stations for precipitation used in this study. For each of the three methods, a weighted sampling with replacement is applied as described in Section 3.7.2. A number equal to the degree of freedom is drawn from the $C_{\text{rel,rare}}$ and $C_{\text{rel,common}}$ distributions, resulting in 1857 values for both relative changes. A t-test was conducted to test whether the means of these two distributions are significantly different from each other. This random sampling is repeated 1000 times. Table 4.4 shows the frequency with which these samples show a significant difference, defined as a p-value of the t-test being lower than 0.05. Given that the obtained $D_{\text{rare-common}}$ for the Full series is very close to zero (Table 4.2), the t-tests indicate that only 5% of all samples shows a significant difference of the means of the two distributions. Both the GEV and MEV show a percentage of significant difference of 90%, indicating that the results for these methods are statistically significant in most cases by randomly selecting the station locations. Based on these results, the test demonstrates that the null hypothesis of no difference in the means of the $C_{\text{rel,rare}}$ and $C_{\text{rel,common}}$ can be rejected in 90% of all samples for both the GEV and MEV, while this is only 5% for the Full series.

Table 4.4: Output t-tests for spatial variability for area-weighted $D_{\text{rare-common}}$ for Western Europe (Table 4.2). Percentage of the 1000 samples accounting for spatial variability where $D_{\text{rare-common}}$ shows a significant difference between $C_{\text{rel,rare}}$ and $C_{\text{rel,common}}$. A significant difference occurs when the p-value of the t-test is below 0.05.

Method	Occurrence of significant difference (%)
Full	4.9
GEV	91.8
MEV	89.0

For each of the 1000 samples, the area-weighted mean of the $C_{\text{rel,rare}}$ and $C_{\text{rel,common}}$ is calculated and plotted in a histogram (Figure 4.10). These histograms illustrates the distribution of the 1000 means for both the $C_{\text{rel,rare}}$ and $C_{\text{rel,common}}$. The Full series shows overlapping distributions, where the $C_{\text{rel,T10}}$ has a higher standard deviation compared to the $C_{\text{rel,T0.1}}$. Both the GEV and MEV show distributions

which do not overlap. However, the $C_{\text{rel,T100}}$ for the GEV has a three times higher standard deviation compared to the $C_{\text{rel,T1}}$. The MEV distributions have equal standard deviations for both the $C_{\text{rel,T100}}$ and $C_{\text{rel,T1}}$.



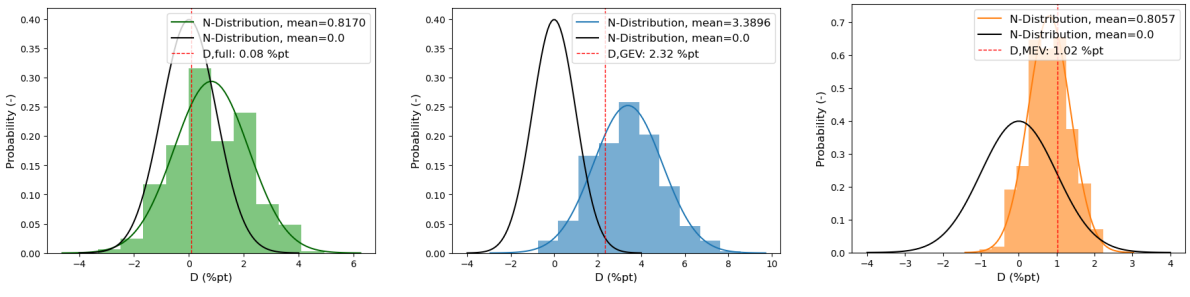
(a) Full **(b)** GEV **(c)** MEV

Figure 4.10: Plot shows the results of all area-weighted 1000 samples of the significance test for spatial variability for each method: **a** the Full series, **b** the GEV and **c** the MEV. For each sample, the mean of all selected $C_{\text{rel,rare}}$ and $C_{\text{rel,common}}$ is calculated and plotted. The previously shown t-test result is based on the two means of each sample.

4.3.2. Temporal variability

To account for the temporal variability, a second test is performed by sampling the years from both periods. From both periods, 1951-1980 and 1981-2010, random years were selected with replacement, equal to a number of 30 years. For each randomly selected pair of years, the area-weighted average $D_{\text{rare-common}}$ for Western Europe is calculated. This sampling is repeated 350 times. All 350 D-values for each method are plotted in a histogram (Figure 4.11). Each plot contains a red dashed line, which corresponds with the area-weighted D-value for Western Europe presented in Table 4.2.

Based on the result of the t-test, it was assessed whether the distribution of D-values for each method differs from the standard normal distribution. The D-value is assumed to be equal to zero for a method if the distribution of all sampled values equal the standard normal distribution, with a mean of 0. The standard normal distribution is therefore assumed to represent a variability in the D-value equal to zero mean and standard deviation 1. The mean of all 350 sampled values for each method varies from the values presented in Table 4.2. Sampling of the years for the Full series results in an average D-value of 0.82%pt. The average D-value for the GEV increases to 3.39%pt, while the value for the MEV decreases to 0.81%pt. As shown in Table 4.5, again the GEV and MEV have a statistical significant, with a p-value lower than 0.05. This indicates that the null hypothesis (H0), that the mean of the D-value distribution is equal to mean of the standard normal distribution, can be rejected. However, the p-value for the Full series is above 0.05, which suggesting that there is no significant difference between means of the standard normal distribution and the D-values.



(a) Full **(b)** GEV **(c)** MEV

Figure 4.11: Histograms illustrates the results of all 350 samples of the significance test for temporal variability for each method: **a** the Full series, **b** the GEV and **c** the MEV. The black line indicates the standard normal distribution. The histograms for each method shows the distribution of all 350 area-weighted $D_{\text{rare-common}}$ obtained by sampling of the years. Red dashed lines indicate the $D_{\textbf{rare-common}}$ obtained for each method.

Table 4.5: t-test output for temporal variability of the area-weighted average $D_{\text{rare-common}}$ for Western Europe. A p-value lower than 0.05 indicates that the null hypothesis that there is a standard normal distribution of D-values, can be rejected.

Method	t-value	p-value
Full	1.87	0.0624
GEV	2.74	0.0068
MEV	-3.42	0.0008

4.3.3. Correlation of $C_{\text{rel,rare}}$ and $C_{\text{rel,common}}$

To examine the correlation between increases in the relative change of rare and common extremes, the Spearman's rank correlation coefficient is applied (see Table 4.6). This test determines whether there is a correlation between the relative change for $C_{\text{rel,rare}}$ and the relative change for $C_{\text{rel,common}}$. A correlation value close to 1 indicates that a high relative change in rare extremes corresponds with a high relative change in more common extremes for a specific station location. The results from the MEV distribution show a correlation of around 0.6. However, the other methods show almost no correlation. The relatively higher correlation observed in the MEV may be influenced by the input data used in the analysis. Since the MEV method utilises all precipitation data instead of just annual maxima, the input data may exhibit less variance, potentially affecting the correlation results.

Table 4.6: Spearman's rank correlation (r_s) between $C_{\text{rel,rare}}$ and $C_{\text{rel,common}}$. The correlation parameter indicates whether a relative change in $C_{\text{rel,rare}}$ corresponds to a similar change in $C_{\text{rel,common}}$. The p-value states whether the result is significant.

Method	r_s	p-value
Full	0.15	<0.001
GEV	0.12	<0.001
MEV	0.59	<0.001

4.4. Correlation $D_{\text{rare-common}}$ and daily temperature

The minimum and maximum daily temperatures are analysed in the same regions as the precipitation data is examined. Initially, the average temperature for both in Western Europe is presented. Next, the temperatures on the days with annual maxima are analysed. Finally, the temperature during common and rare extremes is compared to see if there is a pattern which can be associated with the obtained $D_{\text{rare-common}}$.

4.4.1. All day temperature development Western Europe

For each observation station with temperature data (Figure 3.1 red dots) the average temperature is calculated in both periods (1951-1980 and 1981-2010). This average temperature for each of the two period represents the mean of all day temperature within the time frame of 30 years. Figure 4.12a, b, d and e represents these mean daily temperatures per station. The mean daily temperatures show a similar pattern for both periods 1951-1980 and 1981-2010. As expected, the highest latitudes experience both for the minimum and maximum temperature the lowest temperatures, while the most south regions show the highest temperatures. To analyse the development of the temperature of the observation stations, the difference between the mean daily temperatures are visualized in Figure 4.12c and e, respectively for the minimum and maximum temperature. Stations marked with a brown color indicate locations where the mean daily temperature has decreased, while green dots represent stations where the mean daily temperature has increased. The difference in minimum temperature ranges from -0.29°C to 2.01°°C, whereas the range for the maximum temperature is -1.68°C to 1.98°C. The percentage of stations with a positive difference for both minimum and maximum temperatures is 97.5%.

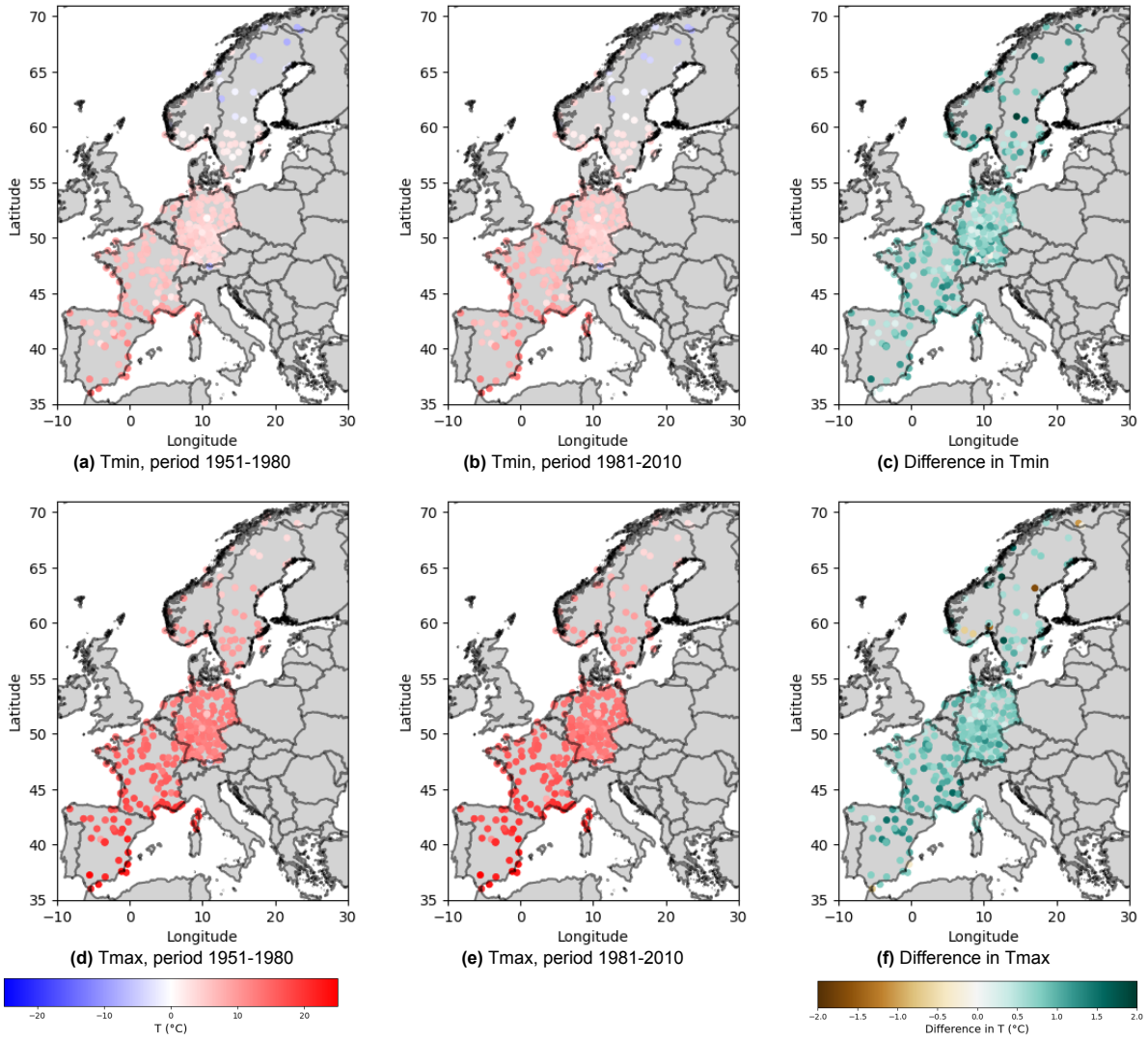
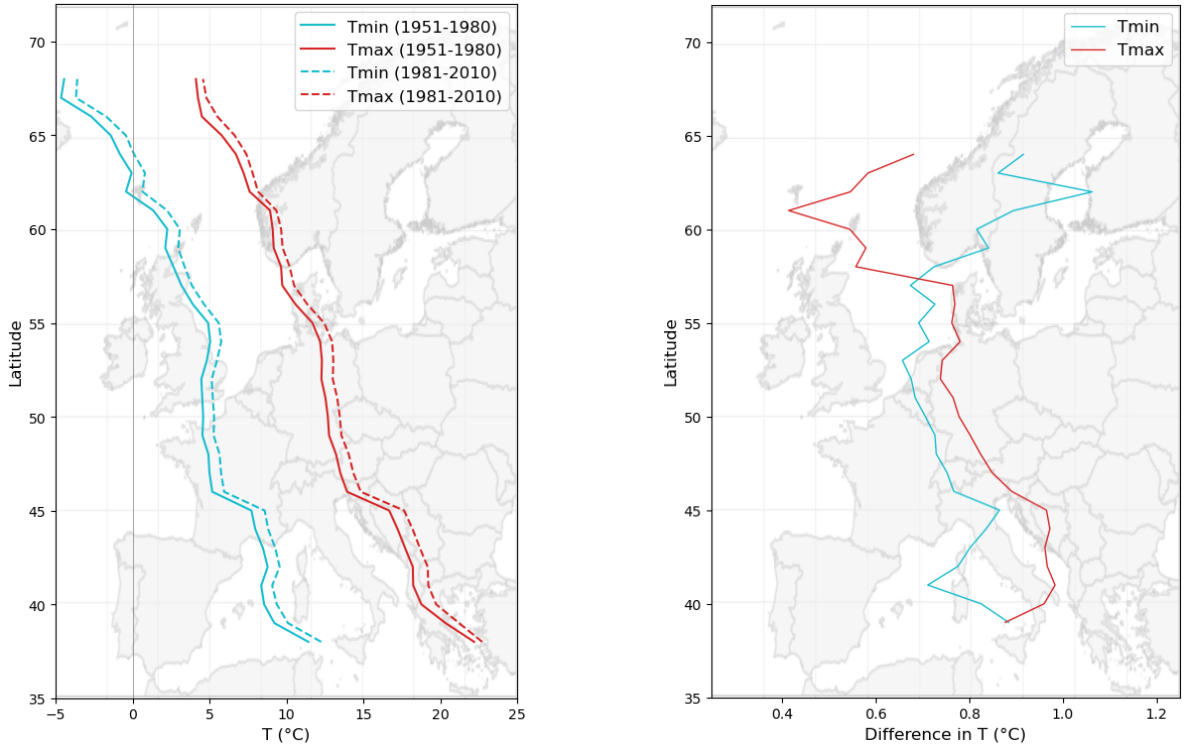


Figure 4.12: Geographical plots showing the mean daily temperature (°C) over the specified periods for each observation station. **a, b, d and e** represents the average temperature for both the Tmin and Tmax for the two periods. Temperatures below zero are indicated with blue, while the positive average temperatures are presented in red. **c and f** shows the difference in temperature between both periods. Brown colors indicate a stations which has a negative difference in T between the two periods (1951-1980 and 1981-2010), while the green colors represents a positive difference in temperature.

The mean daily temperature in both periods (1951-1980 and 1981-2010) for Tmin and Tmax per latitude is visualised in Figure 4.13a, similar to the plot for $D_{\text{rare-common}}$ in Figure 4.9. The dashed line indicates a higher value for all latitudes for both the minimum and maximum temperature, stating a positive difference in temperature on average for each latitude. Figure 4.12b presents these average temperature difference per latitude. The minimum temperature (blue line) shows the highest values in the northern region, approximately equal to 1 °C. The central part of Western Europe is characterised by lower values, around 0.6-0.7 °C, which slightly increases further to the south. The maximum temperature (red line) shows the lowest values for the northern region, while the temperature difference increases in the central part of Western Europe and continues to rise further to the south. Observing both patterns of the difference in minimum and maximum temperature, it is evident that the temperature difference in the northern region is highest for the minimum temperature, whereas the central and southern part show the highest difference for the maximum temperature. Based on this, it can be concluded that the maximum temperature has increased more in Western Europe compared to the minimum temperature, except for the northern region.



(a) Average mean daily temperature per latitude

(b) Average difference in T per latitude

Figure 4.13: Temperature presented per latitude. **a** shows for both the Tmin (blue lines) and Tmax (red lines) the average temperature over the specified period. **b** shows the average difference in both temperatures. The size of the moving window equals 4 degrees.

To obtain a single value for the average change in temperature for Western Europe, the mean daily temperatures of all stations are averaged and weighted according to their corresponding area (Table 4.7). The absolute and relative difference is presented in Table 4.8. The average change in the minimum and maximum temperature in Western Europe equals, respectively, 0.83 °C and 0.70 °C. While the absolute change is approximately the same, the relative change for the minimum temperature is five times larger due to the lower mean daily minimum temperature. This indicates that the minimum temperature has increased relatively more compared to the maximum temperature.

Table 4.7: Average area-weighted Tmin and Tmax for Western Europe for both periods (Figure 3.1 red dots). The temperatures are based on the daily mean Tmin and Tmax value per station and averaged for Western Europe according to the station area.

Period	Tmin (°C)	Tmax (°C)
1951-1980	3.15	11.92
1981-2010	3.98	12.62

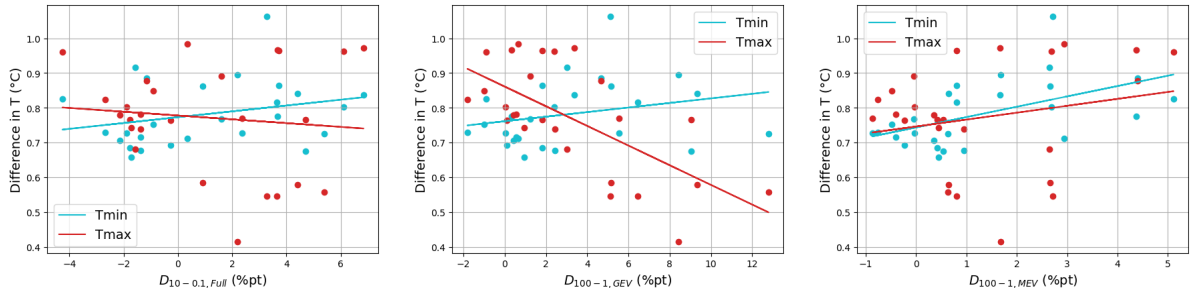
Table 4.8: Average area-weighted difference in Tmin and Tmax for Western Europe for both periods (Figure 3.1 red dots). The differences are based on values shown in Table 4.7

Temperature	Difference in T (°C)	Relative change T (%)
Tmin	0.83	26.4
Tmax	0.70	5.9

4.4.2. Correlation $D_{\text{rare-common}}$ and all day temperature

Based on the area-weighted temperature change stated in Table 4.8, the expected change in precipitation extremes according to the Clausius-Clapeyron relationship should be 7% for each degree of temperature increase. The difference in both minimum and maximum mean daily temperature equals

+/- 0.75°Celsius, a change in precipitation extremes close to 5.3% should be expected. However, as can be seen in Table 4.1, the average values obtained from the observation stations are slightly lower, except for the GEV. The change in temperature extremes, as an average for Western Europe, can therefore be explained by the Clausius-Clapeyron relationship, depending on the used method for estimating the precipitation extremes.
To further quantify the correlation between the $D_{\text{rare-common}}$ for precipitation and the change in temperature for each method, the obtained $D_{\text{rare-common}}$ values per latitude, presented in Figure 4.9, are correlated to the average mean in the minimum and maximum temperature presented in Figure 4.13b. The correlation is calculated using the Spearman's rank correlation coefficient. The highest correlation is obtained between the change in Tmax and the $D_{\text{rare-common}}$ based on the GEV, which provides a correlation of -0.58, which is statistically significant (Table 4.9). This correlation suggests that if the change in Tmax is relative high in a region, the $D_{\text{rare-common, GEV}}$ in this area is relatively low, compared to other regions. Based on the p-values the $D_{\text{rare-common, MEV}}$ shows also a statistical significant correlation with the change in the minimum temperature, with a correlation coefficient equal to 0.53. This correlation indicates that the output of the MEV distribution shows a relatively high $D_{\text{rare-common, MEV}}$ in the regions for which the change in the minimum temperature is relative high.



(a) Full **(b)** GEV **(c)** MEV

Figure 4.14: Correlation of the temperature and the $D_{\text{rare-common}}$ for the methods **a** Full series, **b** the GEV and **c** the MEV. The correlation is calculated for the average values per latitude presented in the Figures 4.9 and 4.13. Blue dots and lines indicate the minimum temperature; the red dots and lines the maximum temperature.

Table 4.9: Output of Spearman's rank test correlation $D_{\text{rare-common}}$ and temperature change per latitude (Figure 4.14). The r_s close to +1 or -1 indicates very strong correlation. A p-value below 0.05 states that the null hypothesis (H0) that there is no correlation can be rejected.

Method	**Tmin**		**Tmax**	
	r_s	**p-value**	r_s	**p-value**
Full	0.27	0.190	-0.03	0.891
GEV	0.27	0.177	-0.58	0.002
MEV	0.53	0.005	0.115	0.575

4.4.3. Daily temperature during annual precipitation extremes

As this study focuses on changing patterns in extreme precipitation events, the minimum and maximum temperature on the days of annual precipitation extremes are analysed. The temperature for each year during the annual extremes are plotted as an area-weighted average of all stations which have both the temperature and precipitation data (Figure 4.15). The average Tmin increases with 0.019°C/year, the Tmax increases with 0.022°C/year, both being statistically significant trends, based on the Mann-Kendall test. Based on this, it can be stated that for Western Europe, both the average Tmin and Tmax increases during the period 1951-2010 at the days of the annual extremes.

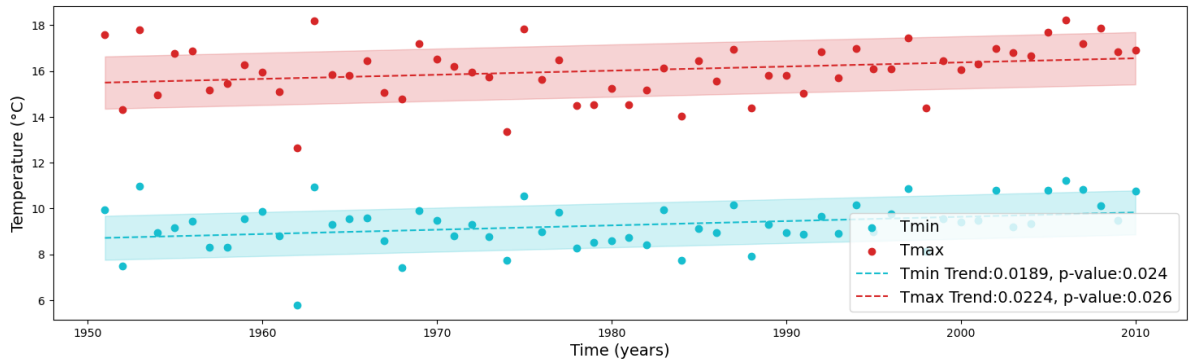


Figure 4.15: Area-weighted average Tmin and Tmax of all temperature stations during annual extremes for each year. A trend line is plotted through the points, indicating the statistical significant (p<0.05) slope of these values. Shading shows the mean +/- standard deviation.

This change in temperature during annual extremes can be caused by the general increase of the minimum and maximum temperature (Table 4.7), but can also be affected by a shift from colder to warmer months as only one day of each year is analysed in this case. Therefore, the months in which the annual extremes occur during both periods 1 and 2 for all precipitation stations are presented in Figure 4.16. In the winter months (Oct-Mar) the second period has 9.5% more annual maxima, compared to the first period 1951-1980, while the summer period (Apr-Sep) shows a decrease of 4.5%. Despite the fact that more annual extremes are present in the summer period for both periods, there is a shift visible towards more precipitation extremes in the winter period, i.e., towards the colder months. The increase in both Tmin and Tmax is therefore not affected by a shift from colder to warmer months.

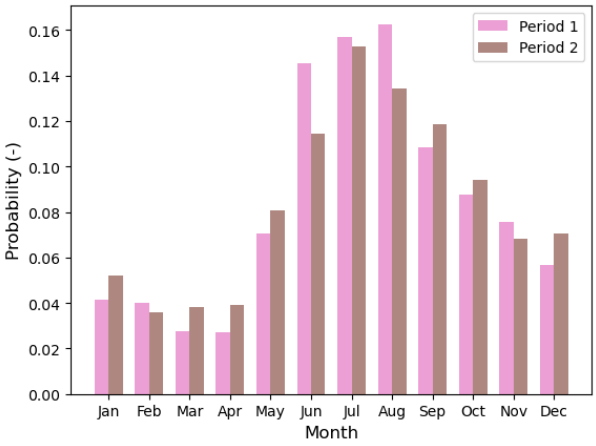


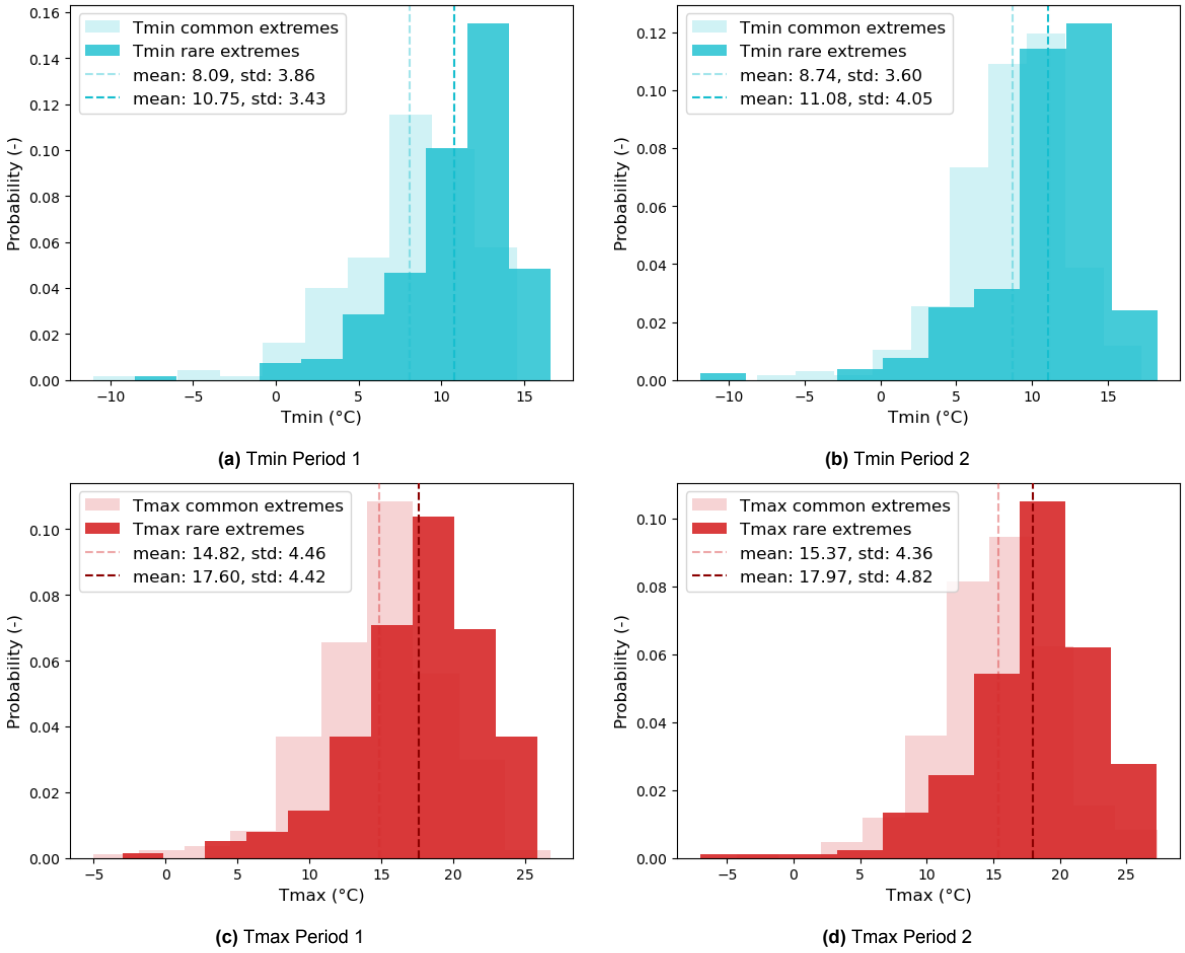
Figure 4.16: Months in which annual precipitation maxima occurs. The bars represent the sum of months in which the annual maxima occur for all precipitation observation stations. Pink bars indicate the months in period 1951-1980, while brown bars represent the months for the period 1981-2010.

4.4.4. Daily temperature during common and rare annual extremes

To assess whether a change in daily temperature can be associated with the positive change between common and rare extremes (Table 4.2), the relation between minimum and maximum temperature and the common and rare extremes is examined. In literature, it is often suggested that an increase in temperature is a driving force for more extreme precipitation.

The results of the temperature analysis during both common and rare extremes are visualised in Figure 4.17. For each station with both temperature and precipitation data, the minimum and maximum temperature during the lowest five annual extremes are averaged, as well as for the five highest annual extremes. This procedure is applied for both period 1 (1951-1980) and period 2 (1981-2010). Figure 4.17 presents for each period, both temperature distributions for the common and rare extremes for all stations. Upon visual inspection of the histograms for each temperature and period, it is observed

that the average temperature at the days with rare precipitation extremes is higher compared to the temperature at the days with common precipitation extremes.



(a) Tmin Period 1 **(b)** Tmin Period 2

(c) Tmax Period 1 **(d)** Tmax Period 2

Figure 4.17: Histograms illustrates the **a** minimum temperature during period 1, **b** minimum temperature during period 2, **c** maximum temperature during period 1 and **d** maximum temperature during period 2, all observed during annual maxima. The temperatures during common extremes represent the average temperature during the five lowest annual precipitation extremes, while the temperature during rare extremes denotes the average temperature during the highest annual precipitation extremes.

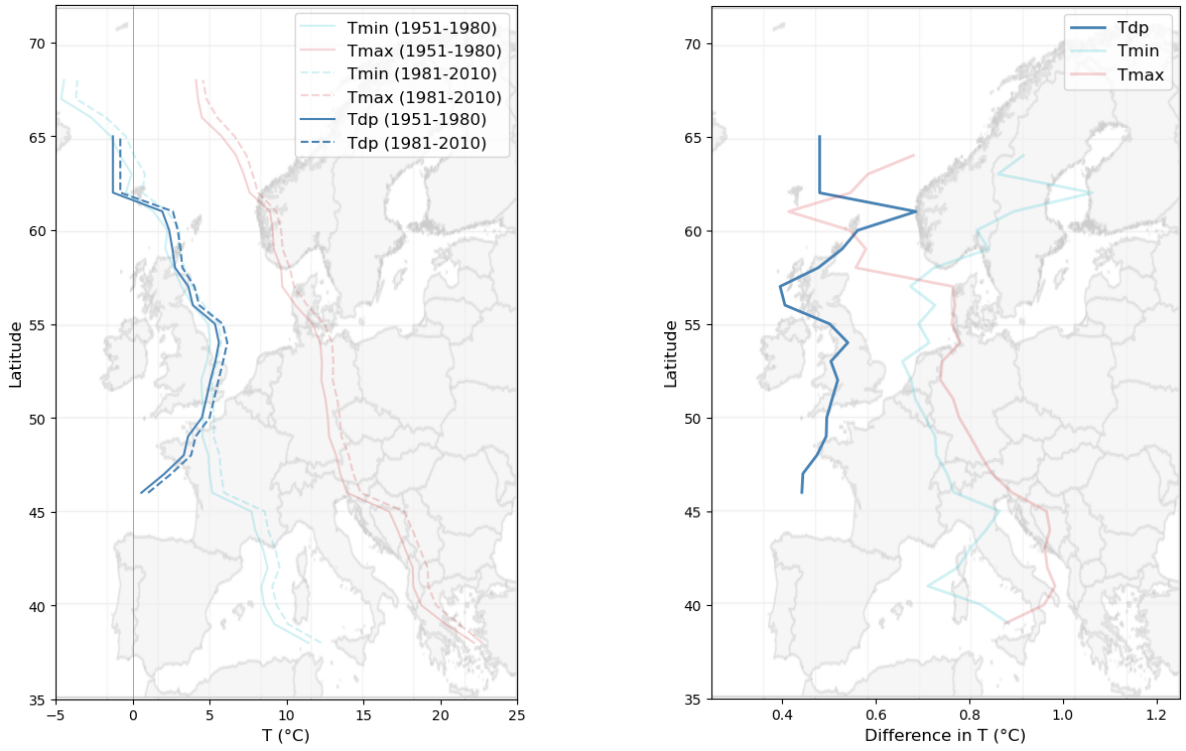
The average minimum temperature demonstrates a greater increase (Table 4.10) during both common and rare extreme precipitation events (8.05 > 3.58 and 3.08 > 2.17). The temperature rise during more common extremes is notably higher compared to that during rare extremes (8.05 > 3.08 and 3.58 > 2.17). This suggests that a greater temperature increase is observed during common precipitation extremes, contradicting the findings in Table 4.2, which indicate a slightly higher increase in rare extremes compared to common ones. These results imply that, based on the datasets used, a relatively higher change in rare extremes is not necessarily associated with a greater temperature increase during those rarer events.

Table 4.10: Relative change in average daily temperature during common and rare extremes for the periods 1951-1980 and 1981-2010.

Temperature	C_{rel} **for common extreme (%)**	C_{rel} **for rare extreme (%)**	**D (%pt)**
Tmin	8.05	3.08	-4.97
Tmax	3.58	2.17	-1.41

4.4.5. Supplementary analysis of dew point temperature

In addition to the comprehensive temperature analysis, a supplementary analysis is conducted on the dew point temperature data for three countries: Germany, The Netherlands and Sweden. An overview of the station location is provided in Appendix E. Although this is not the primary focus of this study, it shows valuable insights into the atmospheric conditions influencing precipitation extremes. This analysis aims to explore the relationship between dew point variations and extreme precipitation events, further enhancing the understanding of the factors driving these extremes. This part does not contain a full analysis, but show the relevance of understanding the behavior of dew point temperature in relation with extreme precipitation patterns.



(a) Average mean daily temperature per latitude

(b) Average difference in T per latitude

Figure 4.18: Average daily dew point temperature (dark blue) presented per latitude. The figure show in **a** shows also both the Tmin (blue lines) and Tmax (red lines) the average temperature over the specified period. **b** shows the average difference in the dew point temperature. The size of the moving window equals 4 degrees. Dew point data is collected from 53 stations in Norway, The Netherlands and Germany.

Averaging the daily dew point temperatures from all stations yields a value of 3.99°C for the period 1951-1980 and 4.49°C for the period 1981-2010 (Table 4.11). On average, the dew point temperature has increased by 0.5°C, representing a relative increase of 12.53%.

Table 4.11: Average dew point temperature for both periods. The temperatures are based on the daily mean dew point temperature value per station.

Period	Tdp (°C)
1951-1980	3.99
1981-2010	4.49

Similar to the procedure used for the minimum and maximum temperatures, the average dew point temperature for both periods, 1951-1980 and 1981-2010, was calculated. Figure **??**a shows the average dew point temperature per latitude, with the minimum and maximum temperatures also plotted for comparison. The pattern of the average dew point temperature is similar to that of the minimum temperature (light blue line). The differences between the average temperatures for both periods are

calculated and presented in Figure 4.18b. The difference in dew point temperature exhibits a pattern similar to that of the minimum temperature, though it is approximately half as large. Nonetheless, the difference remains positive for each region, indicating that the dew point temperature has increased over the period from 1951 to 2010.
Based on the Spearman's rank coefficient, no statistical significant correlation is found between the D-value and the difference in dew point temperature between both periods, as each of the p-values is above the 0.05 (Figure 4.19).

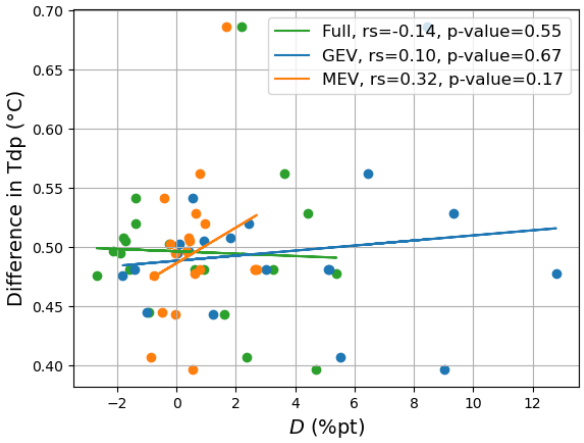


Figure 4.19: Correlation of the dew point temperature and the $D_{\text{rare-common}}$ for the methods Full (green), GEV (blue), and MEV (orange). The Spearman's rank correlation coefficient (r_s) and the corresponding p-value are stated in the legend.

Analysing the dew point temperature on days with precipitation extremes is not feasible due to limited data availability in the three countries. However, an example from one station is presented in Figure 4.20 to illustrate the potential change in the relationship between precipitation extremes and the dew point temperature on those days. The red dots represent data from the period 1951-1980, and the blue dots represent data from the period 1981-2010. For both periods, the average dew point temperature on the days of the annual extremes is presented by a horizontal line. An increase in the average dew point between the two periods is observed, suggesting that the dew point is higher in the second period, on average. To determine if there is a relationship between precipitation and the daily mean dew point temperature, Spearman's rank correlation is applied. This analysis shows no significant correlation for this specific station, as both p-values are above 0.05.

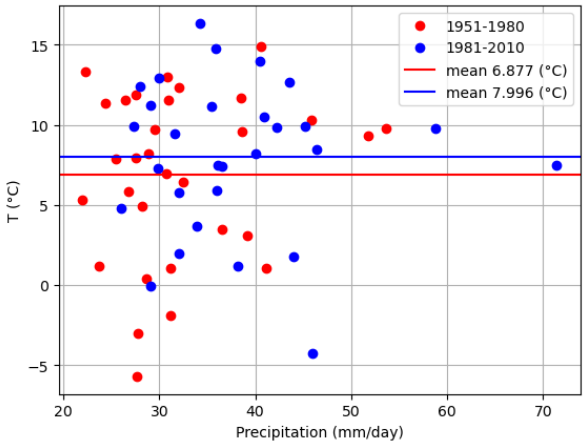


Figure 4.20: Example of a station for which the annual precipitation maxima is related to the dew point temperature on the same day. The data from the period 1951-1980 is presented in red, and the data from the period 1981-2010 is presented in blue.

Table 4.12: Output of Spearman's rank test correlation between the annual precipitation maxima and the dew point temperature on the same day, calculated for each of the two periods, 1951-1980 and 1981-2010.

Period	r_s	**p-value**
1951-1980	0.08	0.68
1981-2010	-0.02	0.91

5 Discussion

5.1. Interpreting results

Given that design criteria related to precipitation extremes are often based on the expected frequencies of rare extreme precipitation events, it is necessary to understand whether the predicted patterns in precipitation extremes are reliable. If rare precipitation extremes are expected to become more extreme compared to the more common ones, and this expectation fails, considerable investments may be made to improve engineering design, which may ultimately prove unnecessary. As it is recognised by Schiermeier (2010) that predictions of precipitation changes represent one of the most complex aspects of climate science, this study highlights the necessity of exercising discretion when deriving conclusions from climate scenarios. It underscores the possibility that in the future, extremes may not occur to the extent initially estimated. Therefore, the results of this study contribute to understanding the quality of the predictions derived from climate models and provide additional evidence to support the finding whether rare extremes will see the greatest relative increase in magnitude.

5.1.1. Relative changes in precipitation extremes

The relative changes in precipitation extremes across Western Europe are obtained by averaging the individual changes over the 1857 stations. As illustrated in Figure 4.4 and Table 4.1, standard deviations up to 27.4% are observed, particularly for the GEV distribution. These standard deviations indicate the range of values across all stations. Figure 4.4d presents relative changes of individual stations up to 100% for the GEV. As explained in Appendix C the GEV distribution can be significantly affected by some 'extreme' events in one of the two periods. An example of such a time series is shown in Figure D.3, which contains a few rare extremes in the second period. These events affect the distribution fit for this second period, causing the high relative change. This explains that obtaining values up to 100% can be driven only by a few extreme events, but is not unrealistic, as it indicates the change in occurrences of these precipitation extremes. Averaging the results of all 1857 station will decrease the effect of individual events at a specific location and provide an indication of the average pattern in Western Europe.

For interpreting the average relative change in return levels for Western Europe, it is important to realise that the study spans thirty degrees of latitude, encompassing various climates that can potentially influence the observed changes in precipitation extremes. Climates are often characterised by different precipitation patterns, which can affect the behavior of precipitation events. Additionally, temperatures vary significantly across the different climates in Western Europe. Therefore, in addition to the average values for the entire region, this study focused on average values per latitude to analyse patterns in the different climates. It is suggested that future research should examine the specific characteristics of each climate in Western Europe, alongside the analysed temperature data.

5.1.2. Difference in relative changes

While this study presented the a large range of return levels for both the C_{rel} and $D_{\text{rare-common}}$ (Figure 4.5), the final $D_{\text{rare-common}}$ values are based on two specific return levels. To interpret the obtained $D_{\text{rare-common}}$ values, each of the three methods is discussed separately.

Full series

In the first method, this study obtained the D-value based on the Full series frequency analysis. The D-values are directly estimated from the time series, making this method the most stable for obtaining return values up to length of the time series, in this case 30 years. The D-value calculated using the Full series method is based on the return levels T0.1 and T10. According to the definition of extreme precipitation, the 0.1-year return level is expected to be exceeded on average every month, while T10 represents a precipitation event expected to be exceeded once every 10 years. Although these return levels differ from those used in the GEV and MEV distributions, the difference in relative changes still indicates whether more common precipitation extremes behave differently compared to more rare extremes. The Full series method yields a D-value close to zero, suggesting no significant difference in relative change between common and rare precipitation extremes. Considering the full range of return levels, as shown in Figure 4.6 for the Full series, the D-value varies depending on how common and rare return levels are defined. Therefore, analysing the behavior of both common and rare extreme should be done using different definitions of what is common and rare to prevent observing a phenomenon that may only be valid for two specific return levels.

GEV

In the second method, the $D_{\text{rare-common}}$ is obtained through the GEV distribution, yielding a value of 2.32%pt, the most significant result among all three methods.
Analysing the entire range of D-values up to a return level of 30 years, the difference in relative change for the GEV is negative. However, for return levels greater than 30 years, the D-value increases sharply, reaching up to 14.0%pt for $D_{1000\text{-}1}$. This pattern indicates that when the rare extreme is defined as exceeding 30-year return level, the GEV suggests that the rare precipitation extremes have increased more in magnitude relative to the more common extremes. Since the GEV uses the annual precipitation extremes as input for defining the parameters, the tail behavior represents the pattern in these extremes, making the obtained D-value representative of only the extremes' behavior. Compared to the Full series, the GEV can estimate return levels exceeding the time series length.

MEV

The $D_{\text{rare-common}}$ based on the MEV distribution shows a value close to 1.0%pt. As depicted in Figure 4.6, which displays the D-values across the range of return levels, this value of 1.0%pt is consistent for each return period. The relative change for each return is therefore almost identical. As the MEV shows an equal D-value for different definitions of the rare extreme, the MEV distribution itself does not show any difference in relative change on average for Western Europe. In comparison to the GEV, the MEV performs better for return levels exceeding the time series length of 30 years, as highlighted in previous comparisons (Zorzetto et al., 2016). One of the reasons for the stable result is the input data used to obtain the MEV distribution, which includes all precipitation events above a threshold of 1 mm/day. Consequently, the distribution is based on all precipitation events, not solely extreme ones. Additionally, the statistical significance test in Chapter 4.3.3 reveals a correlation coefficient between $C_{\text{rel,rare}}$ and $C_{\text{rel,common}}$ of 0.59. This correlation suggests that a high relative increase in the rare extremes corresponds with a high relative change in more common extremes, contributing to the stable result observed for each return level on average across Western Europe.

5.1.3. Correlation with temperature

The results of the temperature analysis showed an average increase in both minimum and maximum temperature over the period 1951-2010 in Western Europe. The temperature analysis by latitude underscores the importance of examining the correlation between differences in relative change and temperature, as the temperature pattern varies across latitudes. This highlights the existence of different climates within Western Europe, suggesting that while it is useful to analyse the overall trend for Western Europe, it is also essential to consider regional variations. The correlation between the average D-value per latitude and the change in temperature is not significant for most of the results. However, a significant result is obtained for the $D_{\text{rare-common,GEV}}$ and the change in maximum temperature, as well as for $D_{\text{rare-common,MEV}}$ and the change in the minimum temperature. These findings indicate a weak overall correlation between changes in temperature and the obtained $D_{\text{rare-common}}$ values. Nonetheless, the two statistically significant results suggest that a relationship may exist. This highlights the complexity of factors influencing $D_{\text{rare-common}}$. The presence of some significant correlations indicates that

temperature changes could play a role, but other factors are also likely to affect $D_{\text{rare-common}}$. Therefore, further analysis is needed to understand what drives the $D_{\text{rare-common}}$ value in specific areas, considering factors beyond just temperature changes.
Analysing temperatures on the days of annual extremes reveals an increase over time, indicating that temperatures on these specific days have been rising from 1951 to 2010. Although this trend is based on a single day each year, the observed increase is of the same order as the overall temperature change.
According to the Clausius-Clapeyron relation, which states that water vapor pressure increases by approximately 7% per degree Celsius, higher temperatures should lead to more water being stored in the atmosphere, resulting in more extreme precipitation events. However, the analysis of temperatures during common and rare precipitation extremes shows a greater temperature increase during common extremes compared to rare extremes. If the Clausius-Clapeyron relation were directly applicable, we would expect greater temperature increases to correspond to more extreme precipitation events. Contrary to this expectation, the results of this study indicate that rare precipitation extremes have not increased proportionately to the higher temperature increases. This discrepancy suggests that other factors may be influencing the intensification of rare precipitation extremes. These factors could include the water vapor content at the time of the precipitation extreme or the temperatures at locations where water enters the atmosphere. Additionally, it is possible that temperature increases have a non-linear effect on water vapor content, with higher temperatures potentially having a more significant impact on water vapor content compared to lower temperatures.

5.2. Comparison with previous study

As mentioned in the introduction, the previous study by Gründemann et al. (2022) indicated that rarest precipitation extremes will see the greatest relative increase in magnitude compared to more common extremes, based on climate scenarios. To verify if the results of this research align with this previous conclusions, it is essential to consider both the length of the time scale and the associated temperature increase should be taken into account. The previous study analysed changes in extreme precipitation over a period of 130 years, while this research focuses on changes over 60 years (1951-2010). To make a meaningful comparison between the two studies, these different time scales must be accounted for. In this study, D-values varying between 0.08 and 2.32%pt were found for Western Europe. In contrast, the study of Gründemann et al. (2022) reported global (land) D-values ranging from 3.06 up to 10.65%pt, depending on the climate scenario (Table 5.1). Specific $D_{\text{rare-common}}$ values for Western and Central Europe were lower, ranging from 2.55 to 8.06%pt. Considering the longer time frame of Gründemann's study which can cause larger difference in relative change, the results of this research are comparable to those corresponding to the lowest climate scenario for Western Europe, excluding the results of the Full series. Despite the results of this study are lower compared to the study of Gründemann et al.(2022), both results show positive D-value on a large scale.

Table 5.1: $D_{\text{rare-common}}$ results previous study (Gründemann et al., 2022) in relation with temperature increase (global values). Temperature change is given for the period 1971-2100 based on Figure 2.3.

Climate scenario	Period	Global $D_{\text{rare-common}}$ (%pt)	Increase global T (°C)
SSP1-2.6	1971-2100	3.06	1.7
SSP5-8.5	1971-2100	10.65	5.1

5.3. Limitations of the method

The methodology used in this study to calculate the $D_{\text{rare-common}}$ and its correlation with daily temperature, contains several limitations in terms of the used data and method.

5.3.1. Data availability and quality

In the analysis of precipitation and temperature data from observation stations, it is crucial to consider the heterogeneity in the available data. Globally, many areas do not have a dense network of meteorological data stations. Moreover, when analysing long time series, limitations arise from the fact that numerous observation stations in many countries did not start data collection more than 60 years ago. For instance Norway, with 2264 observation stations with recorded precipitation data, yet only 62 sta-

tions have maintained continuous precipitation measurements throughout the entire time series from 1951 to 2010. Overall, in this study, 1858 out of 18089 observation stations could be used for analysis, due to shorter time series or missing data. Particularly in other regions worldwide for which the density of observation stations is even lower compared to Western Europe, the analysis of observation data time series may consequently be subject to high uncertainties.
The examination of observation station distributions across the six countries (Germany, France, the Netherlands, Spain, Norway, and Sweden) revealed a non-uniformity in data availability (Figure 3.1). Notably, Germany, France, and the Netherlands exhibited a dense network of observation stations, characterised by extensive coverage. Conversely, Spain, Norway, and Sweden displayed significantly fewer availability of observation stations scattered across their regions. This variation in data underscores a challenge in achieving spatial coverage and highlights potential limitations in data-driven analyses, particularly in regions with sparse observation networks. To mitigate the impact of non-uniform spatial coverage of observation stations, in this study area-weighting techniques were implemented to ensure equal contribution across the geographic regions. However, its crucial to note that this weighting method, while it helps addressing the spatial disparities, does not account for the coverage of the available station, which can cause incorrect assumptions of precipitation and temperature patterns.
The precipitation data for this study is sourced from 1857 stations across Western Europe. Ideally, daily temperature data for all these stations would be available, but this is not the case. Instead, temperature data is available for only 323 locations. Although this number is about one-sixth of the number of precipitation stations, the area coverage is approximately equal, as each region with precipitation data also has a station with temperature data available. When critically evaluating the correlation between precipitation and temperature, the disparity in the number of stations must be considered. The general temperature change presented in this study is based on the average values from the 323 locations. To address the non-uniformity in the data, area-weighting techniques were applied to ensure equal contribution from all regions. Analysing temperature trends on specific days can only be performed at observation stations where both temperature and precipitation data are available. As detailed in Appendix E, 267 of the 323 temperature locations also have corresponding precipitation data. A noticeable reduction in data availability is particularly evident in Norway and Sweden.
A second limitation in data collection is the accessibility of long time series. Data retrieval from meteorological institutes became a time-consuming process due to the complexity of accessing data, especially if the language is not standardised to English. Several institutes in Western Europe improved recently the availability of large datasets, for example via API's, but lots of countries do not have the opportunity to obtain their long term data. For example, for this study, the data from Spain is available until 2012. To get access to the more recent data, payment is required, which limits the use of this data.

5.3.2. Limitations in estimates of precipitation extremes

In this study, three different methods were applied to obtain return levels. Each method has distinct advantages and limitations, impacting their reliability and applicability in different scenarios.
The Full series method is considered the most reliable as it solely uses the available data to calculate return levels. This direct approach ensures that no extrapolation beyond the observed data is necessary, making it robust for shorter return levels. However, its significant limitation is the inability to estimate return levels beyond the length of the time series. This is particularly relevant because most meteorological observation time series do not span at least 100 years, making it difficult to estimate high return levels using this method alone.

The GEV distribution is widely used for extrapolating time series to estimate return levels beyond the observed data range. It offers more flexibility than simpler methods like the Gumbel distribution. However, a key limitation of the GEV method is that it uses only a limited subset of all precipitation events, typically annual maxima, as input for calculating the distribution. This approach significantly reduces the amount of data available for analysis. For shorter time series, this can result in only a few observations being used to fit the distribution. As observed in the fitting of the distribution for several stations in this study (Appendix B), the tail of the distribution is primarily influenced by the highest extremes, which determines the direction of the tail. This can lead to substantial overfitting or underfitting, particularly when comparing high return levels with lower return levels, which can significantly affect the results.

The MEV distribution outperforms the GEV in estimating return levels larger than the sample size (Zorzetto et al., 2016). To obtain the final $D_{\text{rare-common}}$, the MEV is used to obtain a return level of

100 years for a time series of 30 years, considering the MEV performs better for these return levels compared to the GEV. This study also calculates other return levels, ranging from 2 to 1000 years. It would be valuable to investigate for which return levels, relative to the sample size, the MEV results can be reliably used in comparative analyses as conducted in this study. A limitation of the MEV is the lower accuracy of estimates for low return periods, necessitating the use of another method, in this case, the Full series frequency analysis, to estimate low return levels.

While this study investigated the difference in change between common and rare precipitation extremes using the three aforementioned methods. While these three methods differ from each other in terms of data used and calculating return values, several other methods exist to perform extreme value analysis. Some of these other methods are very useful in the field of estimating precipitation extremes as done in this study. An example is the Peaks over Threshold (POT) approach, which analyses precipitation events exceeding a specific threshold. The benefit of this method compared to the use of annual maxima in the GEV is that the GEV excludes the second highest precipitation events, potentially excluding significant extreme precipitation events which can be quite useful for the analysis. Especially in a year which has a very high annual extreme and a relative high second highest extremes, which are not correlated. The POT will in this case include both events in the analysis. For using the POT approach, the effect of the threshold choice should be studied. This POT approach tends to converge to a Generalized Pareto Distribution (GPD) (Beguería & Vicente-Serrano, 2006). This distribution, described by two parameters (a shape and a scale parameter) can be used similarly to the GEV distribution to extract return levels higher than the time series length.

5.3.3. Temperature analysis

Analysing changes in temperature can be approached in several ways. The method used in this study for temperature analysis consists of several steps. First, the average daily temperature for the periods 1951-1980 and 1981-2010 is calculated. While this value provides a direct comparison of temperatures between these periods, it does not show the temperature increase over the entire period from 1951 to 2010 or within each period. Secondly, averaging all daily values over a 30-year period does not account for seasonal variations. Analyzing the behavior of daily temperatures by season can reveal in which season the temperature changes the most.

To obtain a correlation coefficient for the relationship between temperature and the D-value, the average temperature per latitude is used. For visualising this relationship, a linear trend line is plotted in the figures (Figure 4.14). While this trend line indicates the type of correlation, it should be noted that the correlation does not necessarily need to be linear.

The next step in the temperature analysis considers the temperatures on the days of annual extremes. To identify a trend in these temperatures, the values per day are averaged for all stations together, and then the annual average is calculated. Again, a linear trend is assumed to analyze this pattern. However, assuming a linear trend does not account for fluctuations within the values themselves. Figure 4.15 shows that a seasonal trend is visible, which is neglected in this research by using only the linear trend.

The final step involves an analysis of the temperature development during common and rare extreme. As no extreme value analysis is applied to the temperature, the temperature during common extremes is defined as the average temperature during the five lowest annual extremes, while rare extremes are defined as the five highest annual extremes. While this definition indicates a difference in precipitation extremes, it differs from the way common and rare precipitation extremes are defined in previous analyses. Future studies could benefit from analysing the temperatures directly at the events defined for a specific return level in, for example, the GEV, to ensure that the temperatures are more closely aligned with the return levels obtained in the extreme value analysis.

5.4. Recommendations for further research

This study presents a comparative analysis of precipitation and temperature data from observation stations in six Western European countries over the period 1951-2010. While this analysis utilises all available precipitation observation data within this time frame, future research should consider expanding the scope to include data from additional European countries and extending the time series. Particularly, the past 14 years have seen notable extreme weather patterns, both in terms of precipitation and temperature. Consequently, it would be valuable to investigate the $D_{\text{rare-common}}$ for this more

recent period as well.

5.4.1. Driving factors of extreme precipitation

In this study, only a limited subset of the driving factors behind precipitation extremes was analysed, focusing on the correlation with general temperature patterns in Western Europe and local temperatures on the days of annual extremes. Further research should delve into additional precipitation-generating mechanisms. Precipitation extremes are influenced by numerous factors beyond local temperature at the precipitation site, making it worthwhile to investigate atmospheric water vapor, which determines the precipitable amount of water (Kunkel et al., 2020). For this, the dew point temperature and the humidity can be used. Other key factors include the temperature development in regions where precipitation is generated, such as the Atlantic Ocean, and the atmospheric circulation patterns, which can affect the behaviour of cloud systems.

This study includes a supplementary analysis of dew point temperature data. It is recommended to extend this analysis by including additional countries and conducting a more in-depth examination of dew point temperature trends. Specifically, it would be valuable to study hourly data for both precipitation and temperature. This approach would allow for a more precise correlation of local dew point temperature at the moment of extreme precipitation events.

5.4.2. Other data sources

Given the limited availability of observation data due to short time series or the absence of precipitation records numerous regions worldwide, alternative methods for estimating return levels become essential. One such method involved utilising satellite-derived data, which provides high spatial estimations globally. However, a drawback of this approach is the relatively short time series length associated with satellite data, which makes this type of data still unsuitable for the type of analysis conducted in this study. Nevertheless, using other data types can be studied to explore the possibilities for research in this field. An example of another data source is the Multi-Source Weighted-Ensemble Precipitation (MSWEP) dataset. This product merges gauge, satellite and reanalysis data to obtain high-quality of precipitation estimates at a 0.1 °resolution, with a global coverage (Beck et al., 2019).

6

Conclusion

The aim of this study was to examine historical daily precipitation data in Western Europe to determine whether there is a difference in changes between common and rare precipitation extremes. Previous studies, based on climate models, have predicted that under future climate change, the rare precipitation extremes will see the greatest relative increase in magnitude compared to the more common extremes (Gründemann et al., 2022).

This study aimed to investigate whether this finding is evident in historical data from Western Europe and whether these observations can be associated with changes in daily temperature. In this research, rare extremes are defined as events with a return level of at least 100 years, whereas more common extremes are defined as a one-year return level. The difference in relative change for these return levels are calculated based on the relative change of a specific return level estimated by three methods: the Full series, the Generalized Extreme Value (GEV) and the Metastatistical Extreme Value (MEV) distribution.

The results of this study indicate that, on average, historical precipitation extremes in Western Europe have increased in magnitude between the periods 1951-1980 and 1981-2010. However, in the southern regions of Western Europe, negative relative changes were observed, suggesting that precipitation extremes became less extreme. Conversely, the highest relative increases were found in the northern regions of Western Europe. Although the three different methods used in this study show some local differences, they agree on the general pattern for Western Europe. The study obtained an area-weighted average relative change in daily common precipitation extremes, ranging from 3.03% and 3.80%, based on the Full series method. The daily rare precipitation extremes have increased by 5.35% and 4.05%, based on, respectively, the GEV and MEV distribution. The estimated change in rare extremes based on exclusively the time series length using the Full series method, presents a change in the rare extreme equal to 3.88%.

The rarest precipitation extremes increased more relative to the common extremes, according to the estimations of both the GEV and MEV method. The observed difference in relative change between common and rare precipitation extremes equals 2.32%pt and 1.02%pt, respectively, for the GEV and MEV. These values are area-weighted averages for Western Europe. This behaviour of the return levels is statistically significant both for spatial and temporal variability. A comparison of relative changes based on the time series length indicates no differences in relative change for return levels 0.1 and 10 years. Analysing regional variations reveals a varying pattern across different latitudes in Western Europe, with the highest values in the northern regions, while the central part of Western Europe is characterised, on average, by values close to 0%pt.

The mean daily temperature has increased for all regions in Western Europe between the two periods. The highest increase in the minimum temperature is observed in the northern region, while the highest increase in maximum temperature is present in the most south regions. Regions in Western Europe with a relatively high change in the minimum temperature show a greater relative change for the rarest precipitation extremes compared to the more common ones, based on the MEV. According to the GEV a small change in the maximum temperatures occurs in regions in which high values of $D_{\text{rare-common}}$ are obtained. Both the minimum and maximum temperature on the days with annual ex-

tremes have increased with an average rate of 0.019 and 0.022°C/year, respectively, despite a shift from warmer to colder months in which annual extremes occur. It is observed that for both the minimum and maximum temperature, a higher relative increase in temperature is observed during common annual extremes compared to rare annual extremes, while the average temperature during these rare extremes is higher.

To conclude, the differences in the relative changes between the magnitude of common and rare extremes in Western Europe varies at a rate between 0.08 and 2.32%pt, depending on the method used. A significant correlation is obtained for the results of the GEV and MEV distribution and the change in both the minimum and maximum temperature per region in Western Europe. The temperature at the days of annual extremes has increased over the period 1951-2010, while this increase is higher during common extremes compared with the temperature during rare extremes.

References

Aalbers, E. E., Lenderink, G., van Meijgaard, E., & van den Hurk, B. J. J. M. (2018). Local-scale changes in mean and heavy precipitation in western europe, climate change or internal variability? *Climate Dynamics*, *50*(11), 4745–4766. https://doi.org/10.1007/s00382-017-3901-9

AEMET. (2012). *AEMET OpenData*. Retrieved February 29, 2024, from https://opendata.aemet.es/centrodedescargas/inicio

Aleshina, M. A., Semenov, V. A., & Chernokulsky, A. V. (2021). A link between surface air temperature and extreme precipitation over russia from station and reanalysis data [Publisher: IOP Publishing]. *Environmental Research Letters*, *16*(10), 105004. https://doi.org/10.1088/1748-9326/ac1cba

Alexander, L. V. (2016). Global observed long-term changes in temperature and precipitation extremes: A review of progress and limitations in IPCC assessments and beyond. *Weather and Climate Extremes*, *11*, 4–16. https://doi.org/10.1016/j.wace.2015.10.007

Arguez, A., & Vose, R. S. (2011). The definition of the standard WMO climate normal: The key to deriving alternative climate normals [Publisher: American Meteorological Society Section: Bulletin of the American Meteorological Society]. *Bulletin of the American Meteorological Society*, *92*(6), 699–704. https://doi.org/10.1175/2010BAMS2955.1

Beck, H. E., Wood, E. F., Pan, M., Fisher, C. K., Miralles, D. G., Dijk, A. I. J. M. v., McVicar, T. R., & Adler, R. F. (2019). MSWEP v2 global 3-hourly 0.1° precipitation: Methodology and quantitative assessment [Publisher: American Meteorological Society Section: Bulletin of the American Meteorological Society]. *Bulletin of the American Meteorological Society*, *100*(3), 473–500. https://doi.org/10.1175/BAMS-D-17-0138.1

Beguería, S., & Vicente-Serrano, S. M. (2006). Mapping the hazard of extreme rainfall by peaks over threshold extreme value analysis and spatial regression techniques [Publisher: American Meteorological Society Section: Journal of Applied Meteorology and Climatology]. *Journal of Applied Meteorology and Climatology*, *45*(1), 108–124. https://doi.org/10.1175/JAM2324.1

Bon-Gang, H. (2018, January 1). Chapter 3 - methodology. In H. Bon-Gang (Ed.), *Performance and improvement of green construction projects* (pp. 15–22). Butterworth-Heinemann. https://doi.org/10.1016/B978-0-12-815483-0.00003-X

Borodina, A., Fischer, E. M., & Knutti, R. (2017). Models are likely to underestimate increase in heavy rainfall in the extratropical regions with high rainfall intensity. *Geophysical Research Letters*, *44*(14), 7401–7409. https://doi.org/10.1002/2017GL074530

Chan, S., Kahana, R., Kendon, E., & Fowler, H. (2018). Projected changes in extreme precipitation over scotland and northern england using a high-resolution regional climate model. *Climate Dynamics*, *51*. https://doi.org/10.1007/s00382-018-4096-4

Christensen, J. (2020). Snow water equivalent (SWE) - its importance in the northwest. https://www.climatehubs.usda.gov/hubs/northwest/topic/snow-water-equivalent-swe-its-importance-northwest

Coles, S. (2001). *An introduction to statistical modeling of extreme values*. Springer. https://doi.org/10.1007/978-1-4471-3675-0

Countries - GISCO - eurostat. (n.d.). Retrieved April 8, 2024, from https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administrative-units-statistical-units/countries

Donat, M. G., Alexander, L. V., Yang, H., Durre, I., Vose, R., Dunn, R. J. H., Willett, K. M., Aguilar, E., Brunet, M., Caesar, J., Hewitson, B., Jack, C., Klein Tank, A. M. G., Kruger, A. C., Marengo, J.,

Peterson, T. C., Renom, M., Oria Rojas, C., Rusticucci, M., … Kitching, S. (2013). Updated analyses of temperature and precipitation extreme indices since the beginning of the twentieth century: The HadEX2 dataset [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jgrd.50150]. *Journal of Geophysical Research: Atmospheres*, *118*(5), 2098–2118. https://doi.org/10.1002/jgrd.50150

Donat, M. G., Alexander, L. V., Herold, N., & Dittus, A. J. (2016). Temperature and precipitation extremes in century-long gridded observations, reanalyses, and atmospheric model simulations [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/2016JD025480]. *Journal of Geophysical Research: Atmospheres*, *121*(19), 11, 174–11, 189. https://doi.org/10.1002/2016JD025480

DWD. (2023). *Climate data center*. Retrieved February 29, 2024, from https://cdc.dwd.de/portal/

Fischer, E. M., & Knutti, R. (2015). Anthropogenic contribution to global occurrence of heavy-precipitation and high-temperature extremes [Number: 6 Publisher: Nature Publishing Group]. *Nature Climate Change*, *5*(6), 560–564. https://doi.org/10.1038/nclimate2617

Fischer, E. M., & Knutti, R. (2016). Observed heavy precipitation increase confirms theory and early models [Number: 11 Publisher: Nature Publishing Group]. *Nature Climate Change*, *6*(11), 986–991. https://doi.org/10.1038/nclimate3110

Greenwood, J., Landwehr, J., Matalas, N., & Wallis, J. (1979). Probability weighted moments: Definition and relation to parameters of several distributions expressable in inverse form. *Water Resources Research*, 1049–1054. https://doi.org/10.1029/WR015i005p01049

Groisman, P. Y., Knight, R. W., Easterling, D. R., Karl, T. R., Hegerl, G. C., & Razuvaev, V. N. (2005). Trends in intense precipitation in the climate record [Publisher: American Meteorological Society Section: Journal of Climate]. *Journal of Climate*, *18*(9), 1326–1350. https://doi.org/10.1175/JCLI3339.1

Gründemann, G. J., van de Giesen, N., Brunner, L., & van der Ent, R. (2022). Rarest rainfall events will see the greatest relative increase in magnitude under future climate change [Number: 1 Publisher: Nature Publishing Group]. *Communications Earth & Environment*, *3*(1), 1–9. https://doi.org/10.1038/s43247-022-00558-8

Han, D., & Bray, M. (2006). Automated thiessen polygon generation. *Water Resources Research*, *42*(11). https://doi.org/10.1029/2005WR004365

Hodnebrog, Ø., Marelle, L., Alterskjær, K., Wood, R. R., Ludwig, R., Fischer, E. M., Richardson, T. B., Forster, P. M., Sillmann, J., & Myhre, G. (2019). Intensification of summer precipitation with shorter time-scales in europe [Publisher: IOP Publishing]. *Environmental Research Letters*, *14*(12), 124050. https://doi.org/10.1088/1748-9326/ab549c

Hosking, J. R. M., & Wallis, J. R. (1993). Some statistics useful in regional frequency analysis [ADS Bibcode: 1993WRR....29..271H]. *Water Resources Research*, *29*, 271–281. https://doi.org/10.1029/92WR01980

Intergovernmental Panel On Climate Change. (2023, July 6). *Climate change 2021 – the physical science basis: Working group i contribution to the sixth assessment report of the intergovernmental panel on climate change* (1st ed.). Cambridge University Press. https://doi.org/10.1017/9781009157896

Intergovernmental Panel On Climate Change (Ipcc). (2023, July 6). *Climate change 2021 – the physical science basis: Working group i contribution to the sixth assessment report of the intergovernmental panel on climate change* (1st ed.). Cambridge University Press. https://doi.org/10.1017/9781009157896

Kendon, E. J., Blenkinsop, S., & Fowler, H. J. (2018). When will we detect changes in short-duration precipitation extremes? [Publisher: American Meteorological Society Section: Journal of Climate]. *Journal of Climate*, *31*(7), 2945–2964. https://doi.org/10.1175/JCLI-D-17-0435.1

Kidd, C., Becker, A., Huffman, G. J., Muller, C. L., Joe, P., Skofronick-Jackson, G., & Kirschbaum, D. B. (2017). So, how much of the earth's surface is covered by rain gauges? [Publisher: American

Meteorological Society Section: Bulletin of the American Meteorological Society]. *Bulletin of the American Meteorological Society*, *98*(1), 69–78. https://doi.org/10.1175/BAMS-D-14-00283.1

KNMI. (2023). *KNMI - dagwaarden neerslagstations*. Retrieved December 20, 2023, from https://www.knmi.nl/nederland-nu/klimatologie/monv/reeksen

Kunkel, K. E., Stevens, S. E., Stevens, L. E., & Karl, T. R. (2020). Observed climatological relationships of extreme daily precipitation events with precipitable water and vertical velocity in the contiguous united states [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2019GL086721]. *Geophysical Research Letters*, *47*(12), e2019GL086721. https://doi.org/10.1029/2019GL086721

Li, C., Zwiers, F., Zhang, X., Li, G., Sun, Y., & Wehner, M. (2021). Changes in annual extremes of daily temperature and precipitation in CMIP6 models [Publisher: American Meteorological Society Section: Journal of Climate]. *Journal of Climate*, *34*(9), 3441–3460. https://doi.org/10.1175/JCLI-D-19-1013.1

Meteo France. (2023). *Meteo.data.gouv.fr*. Retrieved February 29, 2024, from https://meteo.data.gouv.fr/datasets?topic=6571f26dc009674feb726be9

Min, S.-K., Zhang, X., Zwiers, F. W., & Hegerl, G. C. (2011). Human contribution to more-intense precipitation extremes [Publisher: Nature Publishing Group]. *Nature*, *470*(7334), 378–381. https://doi.org/10.1038/nature09763

Mishra, A. K., & Singh, V. P. (2010). Changes in extreme precipitation in texas. *Journal of Geophysical Research: Atmospheres*, *115*. https://doi.org/10.1029/2009JD013398

Myhre, G., Alterskjær, K., Stjern, C. W., Hodnebrog, Ø., Marelle, L., Samset, B. H., Sillmann, J., Schaller, N., Fischer, E., Schulz, M., & Stohl, A. (2019). Frequency of extreme precipitation increases extensively with event rareness under global warming [Number: 1 Publisher: Nature Publishing Group]. *Scientific Reports*, *9*(1), 16063. https://doi.org/10.1038/s41598-019-52277-4

New, M., Todd, M., Hulme, M., & Jones, P. (2001). Precipitation measurements and trends in the twentieth century. *International Journal of Climatology*, *21*, 1899–+. https://doi.org/10.1002/joc.680.abs

Nikolopoulos, E. I., Borga, M., Marra, F., Crema, S., & Marchi, L. (2015). Debris flows in the eastern italian alps: Seasonality and atmospheric circulation patterns. *Natural Hazards and Earth System Sciences*, *15*(3), 647–656. https://doi.org/10.5194/nhess-15-647-2015

O'Gorman, P. A., & Schneider, T. (2009). The physical basis for increases in precipitation extremes in simulations of 21st-century climate change [Publisher: Proceedings of the National Academy of Sciences]. *Proceedings of the National Academy of Sciences*, *106*(35), 14773–14777. https://doi.org/10.1073/pnas.0907610106

Pall, P., Allen, M. R., & Stone, D. A. (2007). Testing the clausius–clapeyron constraint on changes in extreme precipitation under CO2 warming. *Climate Dynamics*, *28*(4), 351–363. https://doi.org/10.1007/s00382-006-0180-2

Papalexiou, S. M., Koutsoyiannis, D., & Makropoulos, C. (2013). How extreme is extreme? an assessment of daily rainfall distribution tails [Publisher: Copernicus GmbH]. *Hydrology and Earth System Sciences*, *17*(2), 851–862. https://doi.org/10.5194/hess-17-851-2013

Papalexiou, S. M., & Koutsoyiannis, D. (2013). Battle of extreme value distributions: A global survey on extreme daily rainfall [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2012WR012557]. *Water Resources Research*, *49*(1), 187–201. https://doi.org/10.1029/2012WR012557

Pendergrass, A. G., & Hartmann, D. L. (2014). Changes in the distribution of rain frequency and intensity in response to global warming [Publisher: American Meteorological Society Section: Journal of Climate]. *Journal of Climate*, *27*(22), 8372–8383. https://doi.org/10.1175/JCLI-D-14-00183.1

Pérez Bello, A., Mailhot, A., & Paquin, D. (2021). The response of daily and sub-daily extreme precipitations to changes in surface and dew-point temperatures. *Journal of Geophysical Research: Atmospheres*, *126*(16), e2021JD034972. https://doi.org/10.1029/2021JD034972

Schellander, H., Lieb, A., & Hell, T. (2019). Error structure of metastatistical and generalized extreme value distributions for modeling extreme rainfall in austria. *Earth and Space Science*, *6*(9), 1616–1632. https://doi.org/10.1029/2019EA000557

Schiermeier, Q. (2010). The real holes in climate science [Publisher: Nature Publishing Group]. *Nature*, *463*(7279), 284–287. https://doi.org/10.1038/463284a

SHMI. (2023). *Ladda ner meteorologiska observationer | SMHI*. Retrieved February 29, 2024, from https://www.smhi.se/data/meteorologi/ladda-ner-meteorologiska-observationer#param=precipitation24HourSum,stations=active

Sun, Q., Zhang, X., Zwiers, F., Westra, S., & Alexander, L. V. (2021). A global, continental, and regional analysis of changes in extreme precipitation [Publisher: American Meteorological Society Section: Journal of Climate]. *Journal of Climate*, *34*(1), 243–258. https://doi.org/10.1175/JCLI-D-19-0892.1

Tradowsky, J. S., Philip, S. Y., Kreienkamp, F., Kew, S. F., Lorenz, P., Arrighi, J., Bettmann, T., Caluwaerts, S., Chan, S. C., De Cruz, L., de Vries, H., Demuth, N., Ferrone, A., Fischer, E. M., Fowler, H. J., Goergen, K., Heinrich, D., Henrichs, Y., Kaspar, F., … Wanders, N. (2023). Attribution of the heavy rainfall events leading to severe flooding in western europe during july 2021. *Climatic Change*, *176*(7), 90. https://doi.org/10.1007/s10584-023-03502-7

Twardosz, R., Walanus, A., & Guzik, I. (2021). Warming in europe: Recent trends in annual and seasonal temperatures. *Pure and Applied Geophysics*, *178*(10), 4021–4032. https://doi.org/10.1007/s00024-021-02860-6

Westra, S., Alexander, L. V., & Zwiers, F. W. (2013). Global increasing trends in annual maximum daily precipitation [Publisher: American Meteorological Society Section: Journal of Climate]. *Journal of Climate*, *26*(11), 3904–3918. https://doi.org/10.1175/JCLI-D-12-00502.1

Wilson, P. S., & Toumi, R. (2005). A fundamental probability distribution for heavy rainfall [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2005GL022465]. *Geophysical Research Letters*, *32*(14). https://doi.org/10.1029/2005GL022465

Xie, P., & Arkin, P. A. (1995). An intercomparison of gauge observations and satellite estimates of monthly precipitation [Publisher: American Meteorological Society Section: Journal of Applied Meteorology and Climatology]. *Journal of Applied Meteorology and Climatology*, *34*(5), 1143–1160. https://doi.org/10.1175/1520-0450(1995)034<1143:AIOGOA>2.0.CO;2

Zhai, P., Zhang, X., Wan, H., & Pan, X. (2005). Trends in total precipitation and frequency of daily precipitation extremes over china [Publisher: American Meteorological Society Section: Journal of Climate]. *Journal of Climate*, *18*(7), 1096–1108. https://doi.org/10.1175/JCLI-3318.1

Zhang, W., Villarini, G., & Wehner, M. (2019). Contrasting the responses of extreme precipitation to changes in surface air and dew point temperatures. *Climatic Change*, *154*(1), 257–271. https://doi.org/10.1007/s10584-019-02415-8

Zorzetto, E., Botter, G., & Marani, M. (2016). On the emergence of rainfall extremes from ordinary events [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/2016GL069445]. *Geophysical Research Letters*, *43*(15), 8076–8082. https://doi.org/10.1002/2016GL069445

Zorzetto, E. (2023, May 11). *EnricoZorzetto/mevpy* [original-date: 2017-12-15T01:29:18Z]. Retrieved January 26, 2024, from https://github.com/EnricoZorzetto/mevpy

A

Overview raw data

This appendix provides an overview of the data used for the analysis. There is some variation in how meteorological institutes in West-Europe store their data, which makes it necessary to have a comprehensive understanding of the data and the available parameters. The time frame for which data is obtained spans from January 1, 1951 to December 31, 2010.

A.1. The Netherlands

Meteorological data in The Netherlands is available from the Royal Netherlands Meteorological Institute (KNMI). The KNMI maintains a total of 469 weather stations with data for the specified time frame and 50 stations with temperature data. The data is provided in text file format(.txt) for each station separately. The precipitation data includes daily precipitation an snow cover observations (KNMI, 2023). The data files contains the following parameters:

- STN: station ID
- YYYYMMDD: date (year, month, day)
- RD: daily precipitation amount in 0.1 mm over the period 08:00 preceding day - 08:00 UTC present day (0.1 mm)
- SX: code for snow cover at 08:00 UTC

The snow cover (SX) indicates the type of snow as follows:

- 1 - 996: snow cover in cm
- 997: broken snow cover < 1 cm
- 998: broken snow cover > 1 cm
- 999: snow dunes

To recalculate snow to liquid precipitation, the snow water equivalent (SWE) is an important parameter. The KNMI does not provide measurements relate to SWE. As an indication, the snow can be recalculated using the rule '1 cm snow equals 1 mm precipitation'. If data for a given day is not measured, the column contains five spaces. The data was validated by the KNMI every 10 days and updated when the validated values became available. A comment by KNMI indicates that the temperature time series are inhomogeneous due to station relocations and changes in observation techniques. Consequently, these series are not suitable for trend analysis. However, a total of four weather stations contain homogenized series, namely de Kooy, Eelde, de Bilt and Vlissingen.

The temperature data contains the following parameters:

- TG = daily mean temperature in (0.1°Celsius)
- TN = daily minimum temperature in (0.1°Celsius)
- TX = daily maximum temperature in (0.1°Celsius)

The coordinates of the observation stations are available in a separate file. The KNMI has chosen to provide the coordinates with one decimal, which may not give the exact location of the stations.

A.2. Germany

The meteorological institute in Germany is the Deutsche Wetterdienst (DWD), which provides time series data via the Climate Data Center (CDC). The DWD offers precipitation and snow fall data for Germany for 5537 weather stations and temperature data from 1058 stations. Several files are available for each location (DWD, 2023). The precipitation time series data is delivered in files, named as follows: *'produkt nieder tag 19510101 20061231 00001.txt'* for station number 1. A separate file, 'Parameter nieder tag', provides a description of all the parameters. It is stated that the time frame used until 1990 is 07:00 - 07:00 FT. GZ, and after 1990 the time frame is 07:30 - 07:30 FT. GZ. The following relevant parameters are available:

- STATIONS ID: station ID
- MESS DATUM: date in (YYYYMMDD)
- RS: daily precipitation (tgl. Niederschlagshoehe Messnetz 6) in (mm)
- SH TAG: daily snowcover in cm (Schneehoehe Tageswert)
- NSH TAG: daily new snowcover in cm (measured at 07:00-07:30 AM)

Another file is available with data labeled as incorrect, 'Metadaten Fehldaten'. This file contains an overview of periods in which the data is incorrect or unavailable. The parameter is indicated along with the station ID. The time frame of the 'fail data' is specified as 'Von Datum' until 'Bis Datum'. Any part the time series is removed that appears in this list is removed. In the data processing phase, the snow data is added to the rainfall data. This achieved by using the NSH TAG parameter, which contains the new daily snow cover. To account for the addition of snow to liquid precipitation, the difference between the snow cover of consecutive days is used. If this value is negative, it indicates that the snow cover is melting and contributes to liquid precipitation.
The temperature data is downloaded separately from the precipitation data and is available in .csv format. For each station, a separate file is provided for the minimum and maximum temperature. The files contain the following parameters:

- Produkt code: indicates the type of parameter (Tmax: OBS DEU P1D T2M X, Tmin: OBS DEU P1D T2M N)
- Zeitstempel: date in (YYYY-MM-DD)
- Wert: value of the parameter in (°Celsius)

A.3. France

Météo France is the French Institute for Meteorology. The precipitation and temperature data is available in .csv files for the 101 departments in France. In total, 10060 observation stations are available (Meteo France, 2023). The data is provided under the name 'Données climatologiques de base - quotidiennes'. It is assumed that snow is included in the RR parameter, as for days with daily maximum temperature below -2°Celsius do not always have an RR value of 0. It is assumed that the snowfall is recalculated to its water equivalent.

The following relevant parameters are available:

- NUM POSTE: station id
- LAT: latitude of the station
- LON: longitude of the station
- AAAAMMJJ: date in (YYYYMMDD)
- RR: Total daily precipitation sum (mm)
- TN: Minimum temperature in (°Celsius)
- HTN: Time of day of TN (HH:MM)
- TX: Maximum temperature in (°Celsius)
- HTX: Time of day of TX (HH:MM)
- TM: Daily average temperature in (°Celsius)
- QQR: Quality indication (0: Final quality check. Check by climatologist, 1: Automatic validated data, 2: Data labeled as unreliable by automatic system, 9: Filtered data)

A.4. Sweden

The SHMI (Sweden Hydrological and Meteorological Institute) provides historical data from its observation stations (SHMI, 2023). Stations can be filtered based on non-active/active stations. Only active stations are utilized in this study. A total of 554 active stations are available for data retrieval. According to the description on SHMI website, the precipitation data includes rainfall, snow and hail (SHMI, 2023). The following relevant parameters are available for the precipitation data:

- Från Datum Tid (UTC): start date and time (DD-MM-YYYY HH:MM)
- Till Datum Tid (UTC): end date and time (DD-MM-YYYY HH:MM)
- Representativt dygn: Representative day (DD-MM-YYYY)
- Nederbördsmängd: Total daily precipitation sum (mm)
- Kvalitet: quality indication (G: checked and approved values, Y: Suspicious or aggregated values. Roughly controlled archive data and uncontrolled real-time data (last 2 hours))

For the temperature the same parameters are available for the general information. The temperature is stated as follows:

- Lufttemperatur min: Minimum daily temperature (°Celsius)
- Lufttemperatur max: Maximum daily temperature (°Celsius)

These temperature data is available in separate .csv files for a total number of 224 stations.

A.5. Spain

Precipitation en temperature data from Spain is available from the AEMET (Agencia Estatal de Meteorologica). The data used in this research dates from 2012, as data has not been freely available since that year. The data comprises separate .csv files for each of the 114 observation stations (AEMET, 2012). Snow data is not available. The date is stated in three separate column:s Ano (year), Mes (month) and Dia (day). Missing data is stated as 'lp' or 'Acum'.

The following relevant parameters are available:

- Indicativo: station ID
- Prec: Total daily precipitation sum (mm)
- T.max: Maximum Temperature (°C)
- T.min: Minimum Temperature (°C)
- T.med: Average Temperature (°C)

The precipitation and temperature data is available in one file.

A.6. Norway

The data is sourced from the Norwegian Meteorological Institute (MET). Approximately 2000 stations are accessible, with 1364 located within Norway. Some stations downloaded are located outside of Norway. The data compromises precipitation values in mm along with a corresponding quality code. The quality code is equal to 0, 1 or 2, where 0 signifies high-quality data, 1 indicates minor issues, and a value of 2 denotes significant errors or missing data.
The following relevant parameters are available:

- elementId: parameter
- value: corresponding value for the parameter
- unit: unit of the given value
- timeOffset: PT6H or PT18H, value for the different timeframes (PT6H: time frame starts at 6:00 AM)
- referenceTime: date and time (YYYY-MM-DD HH:MM:SS)
- sourceId: station ID

The precipitation and temperature data is downloaded separately.

B

Visualization analysis fits

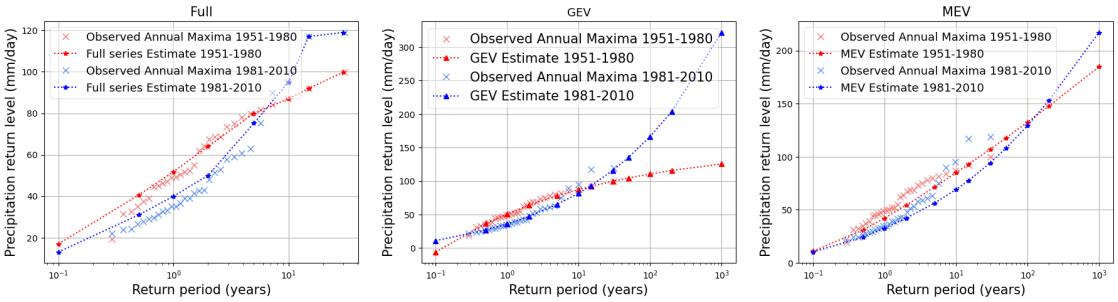


Figure B.1: France station 13028001

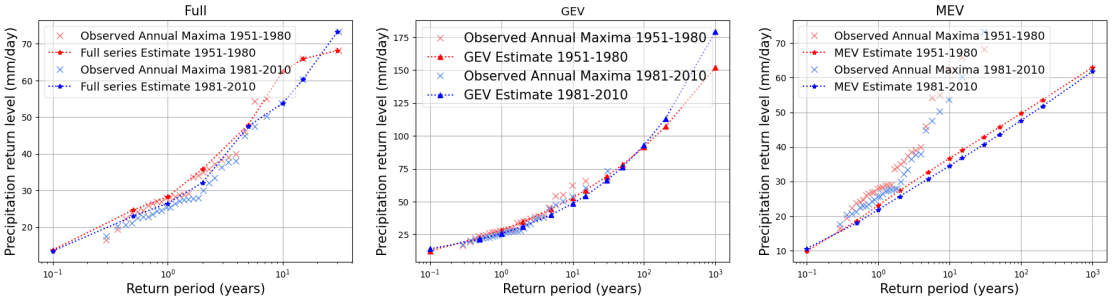


Figure B.2: Germany station 5382

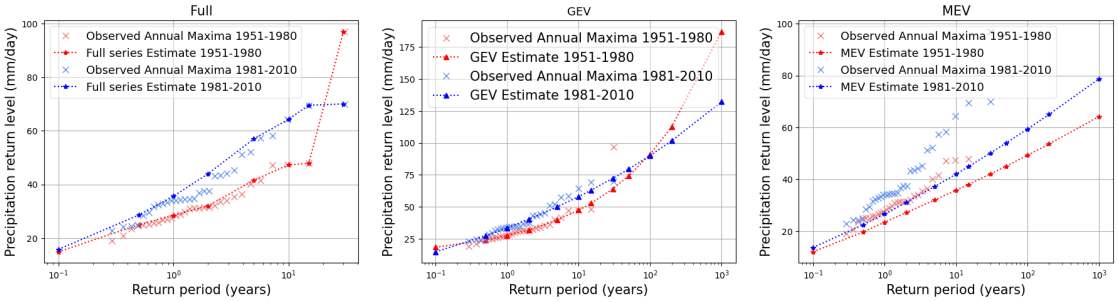


Figure B.3: Netherlands station 234

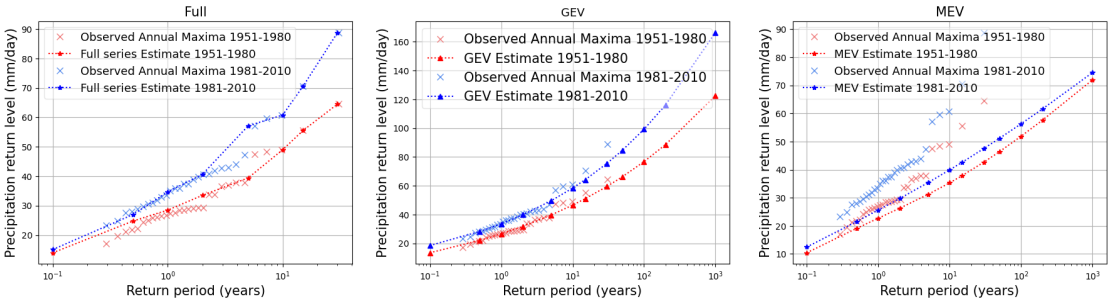


Figure B.4: Netherlands station 740

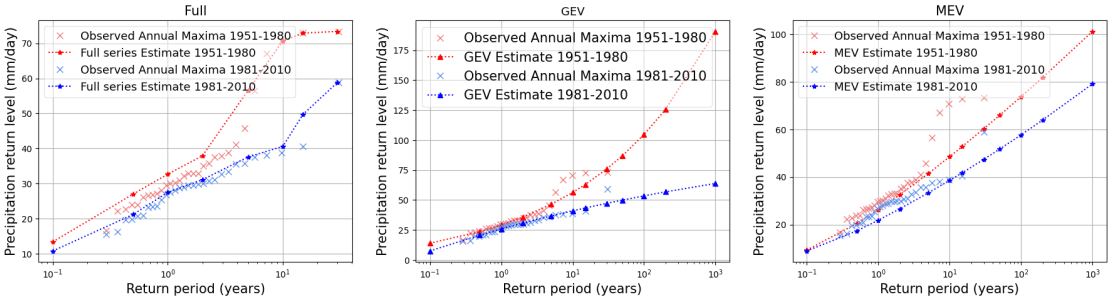


Figure B.5: Spain station 3129

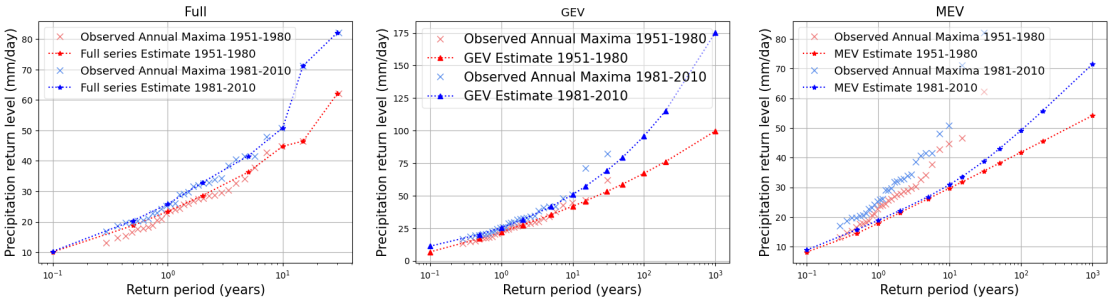


Figure B.6: Sweden station 64020

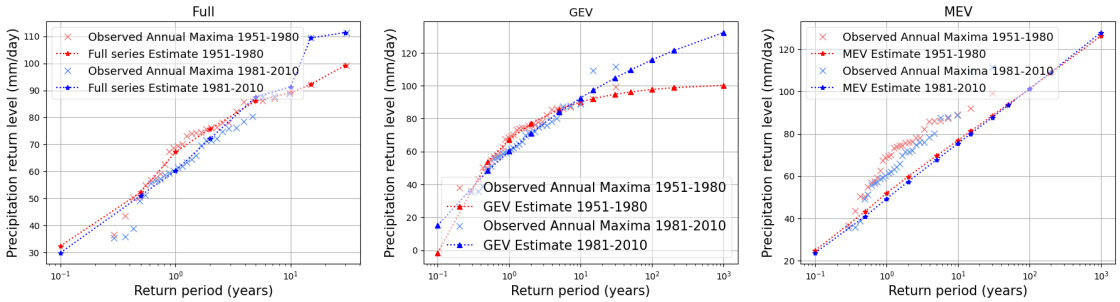


Figure B.7: Germany station 1346

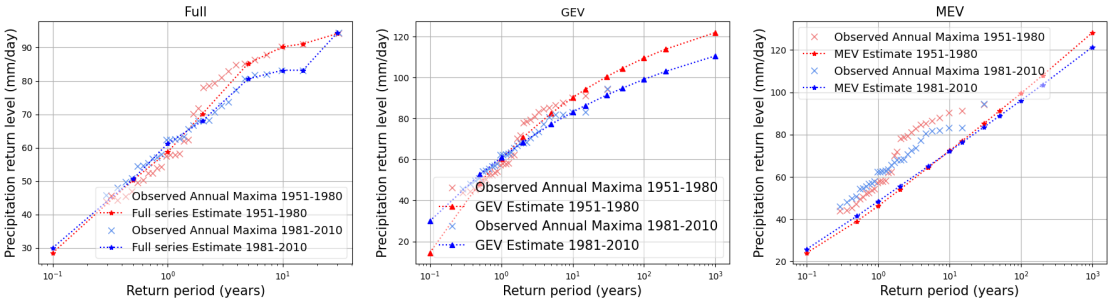


Figure B.8: Norway station SN75100

C

Performance Extreme Value Analyses

The results obtained from the GEV and MEV distributions can strongly be influenced by the characteristics and behaviours of these methods. Therefore, it is essential to analyse and describe the performance of both methods to comprehend the final outcomes. Both the GEV and MEV methods extrapolate the available data to estimate values for higher return periods. To visualise the fitting process and its implications, distributions are plotted for eight stations across different countries (see Appendix B). An example of such a plot is illustrated in Figure C.1, which depicts the relative behavior of both fits for each method. The plots demonstrate the relative behaviour of both fits for each method. The GEV shows a behaviour which corresponds with the fact that rarer extremes became more extreme compared to common extremes, as the distance between the fits becomes bigger for an increasing return period. The fits of the MEV explain why the difference in relative change for this method are not deviating a lot from 0. The lines show that the precipitation events became extremer in the second period (blue line), but the distance between both lines does not change, which indicate a D value close to zero. For other stations this pattern is approximately equal (see Appendix B).

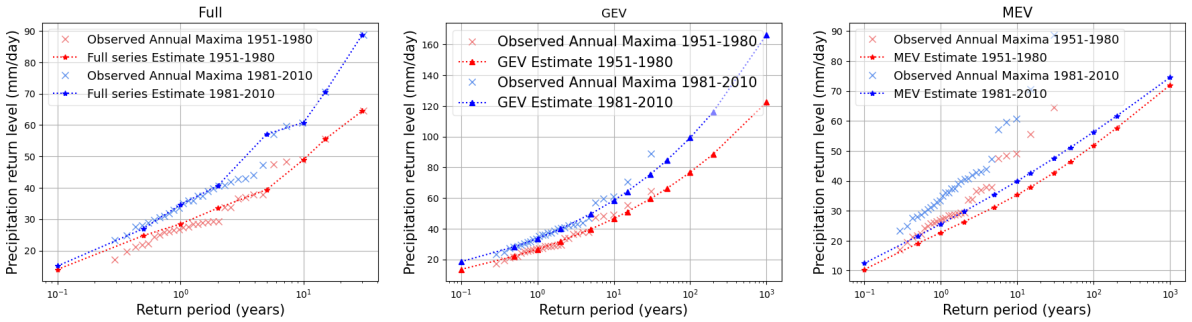


Figure C.1: Visualisation of the fits for the Full series, GEV and MEV for the station 740 in The Netherlands. Red lines indicates the fit for the first period 1951-1980. The blue line is the fit for the second period 1981-2010. The red and blue x's are the annual maxima for the corresponding periods. These annual maxima are the input for the GEV distribution.

C.1. GEV analysis

In the statistical significance test, it was evident that the GEV distribution has the highest standard deviation and the most extreme outliers compared to the other methods (see Figure 4.7). Therefore, it is worthwhile to check the behaviour of the data associated with these highest relative changes. Specifically, an analysis was conducted on the 10 stations with the highest relative change for the return level of 100 years. To assess the impact of outliers in the data from these specific stations, bootstrapping was employed. This technique involved randomly sampling from the 30 annual precipitation extremes, for which the GEV distribution was calculated, and repeating this process 1000 times. Figure C.2 illustrates the boxplots resulting from this bootstrapping. It is notable that the boundaries of the boxplots, particularly for the stations 'Sweden 92530', 'Sweden 97520', and 'Germany 5320', are extensive. However, upon analysing the full time series of these stations (detailed in Appendix D.2), it was observed that

the data does not exhibit any unrealistic characteristics. Consequently, these stations were deemed suitable for inclusion in the analysis. Further examination of the time series from these three stations revealed that the second period (1981-2010) encompassed a few extreme precipitation events with values reaching 80-100 mm/day. Given that the GEV fit for this second period is heavily influenced by these 'extreme' events, it becomes apparent why the extreme D value was observed.

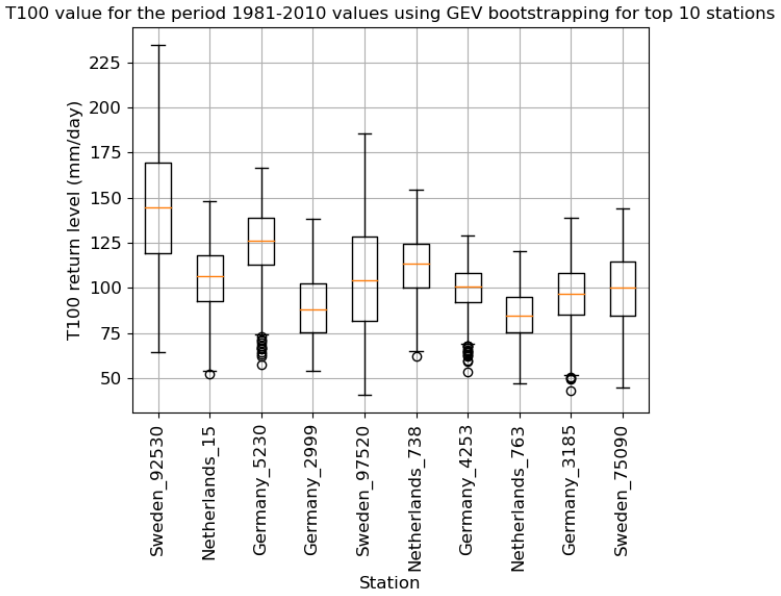


Figure C.2: Boxplots of bootstrapping the 10 highest outliers in the $C_{\text{rel},100}$ for the periods 1951-1980 and 1981-2010

The effect of overfitting due to only a few numbers of extremes, is visualised in Figure B.5 for a station in Spain. For this location, period 1 (1951-1980) contains several high extremes compared to period 2 (1981-2010). Consequently, the red line representing the GEV fit for period 1 appears to overfit significantly when contrasted with the fit for period 2. This discrepancy leads to substantial relative changes for the higher return periods. Although this phenomenon is observed for just one station, it underscores the influence of the GEV distribution on estimating relative changes for high return levels.

C.2. MEV analysis

The performance of the MEV method was evaluated by examining outliers and the parameters derived to achieve the fit. Initially, the MEV results revealed an extreme outlier that significantly impacted the histogram. To evaluate the behaviour and ascertain the cause of this outlier, the time series of the associated station was plotted (Figure D.1). This visualisation revealed unrealistic patterns in the data. Consequently, the data from this station was excluded from the analysis to prevent its influence on the MEV analysis.

To gain further insight in the performance of the MEV distribution, the underlying parameters are visualised. The relative change for a specific return level can be influenced by variations in parameters such as the number of wet days (N), the scale parameter (C), or the shape parameters (W). As expected, the parameters for period 1 and period 2 are approximately the equal (Figure C.3). The number of wet days (N) are slightly higher for period 2 compared to period 1, while the scale parameter (C) is only increased by 0.2. In previous studies (Gründemann et al., 2022) it was suggested that the observed trend of rare extremes increasing relatively more than common extremes could be attributed to a decrease in the number of wet days coupled with an increase in the shape parameter (C).However, based on the data analyzed in this study, the number of wet days is actually increasing, while the shape parameter (C) remains relatively stable. Additionally, another behavior that could explain a positive D value is a decrease in the shape parameter (W). However, the decrease observed in this parameter (0.8%) is insufficient to account for the D value. The fact that the MEV distribution results in a stable D value for all return levels (Figure 4.6, orange line), is related to the fact that for both periods 1 and 2 the fitted line behaves the same. can be attributed to the fact that the fitted lines for both periods 1 and

2 exhibit similar behavior. Moreover, for all return levels, the values of approximately 4% for $C_{\text{rel,MEV}}$ indicate a consistent increase in extreme precipitation. However, this increase is uniform across all return levels, suggesting that the method does not differentiate between common and rare extremes.

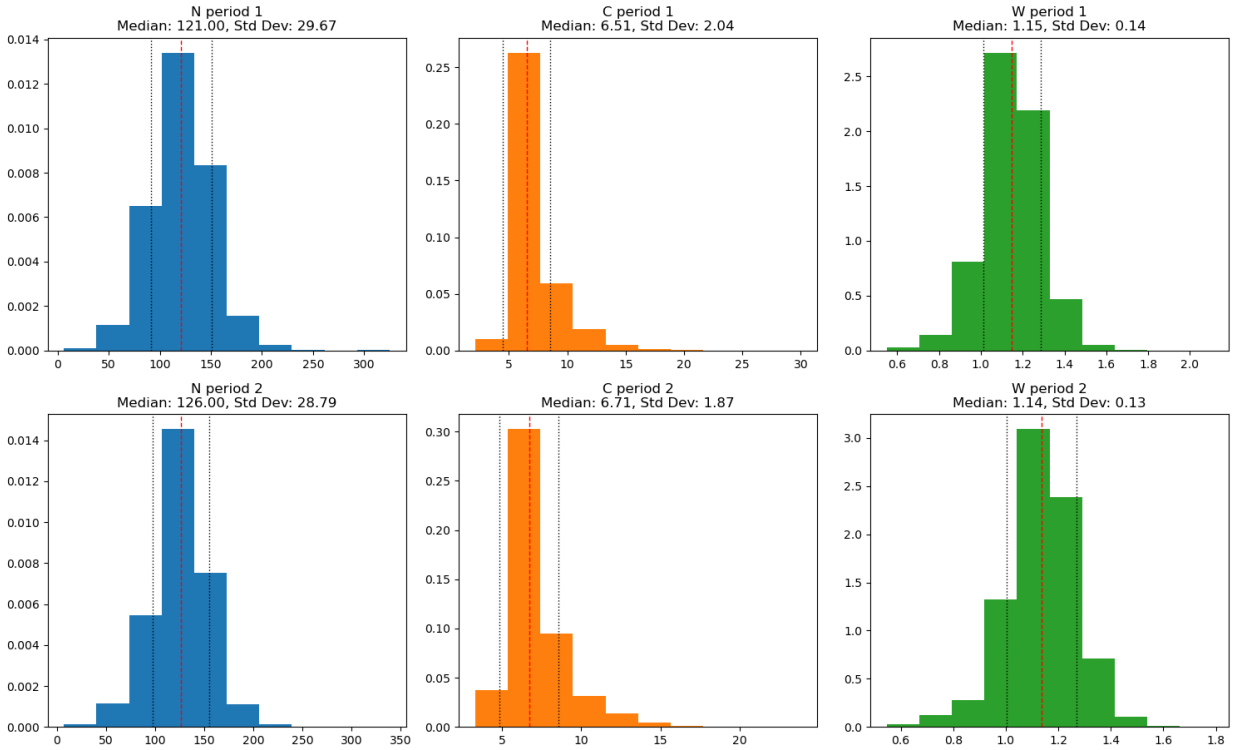


Figure C.3: Histogram of MEV parameters for period 1 and period 2.

D

Overview quality check data

D.1. Outliers MEV

The analysis of the MEV showed high outliers, which effect the final result. Checking the data of this station it is seen that the standard quantity and quality check did not account for this type of wrong data.

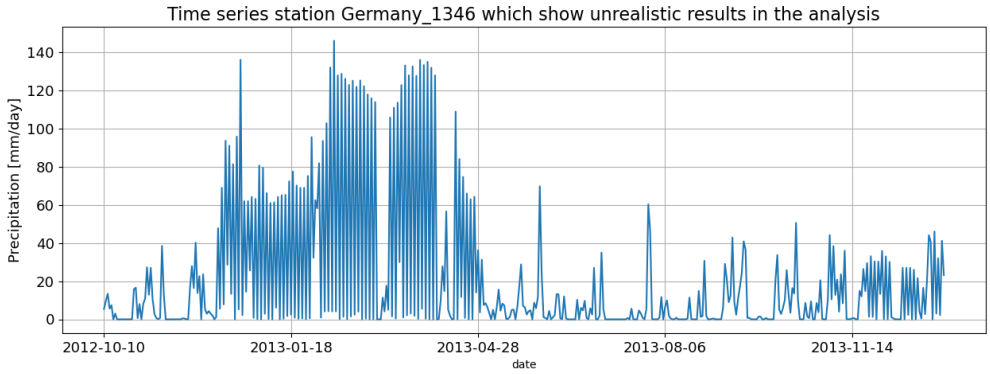


Figure D.1: Time series of station in Germany which show outlier in the MEV analysis. Plotting the data shows a unreliable time series in the showed time frame. For months the daily precipitation is around 60 mm, which is not realistic, especially compared with stations close to this location.

D.2. Outliers GEV

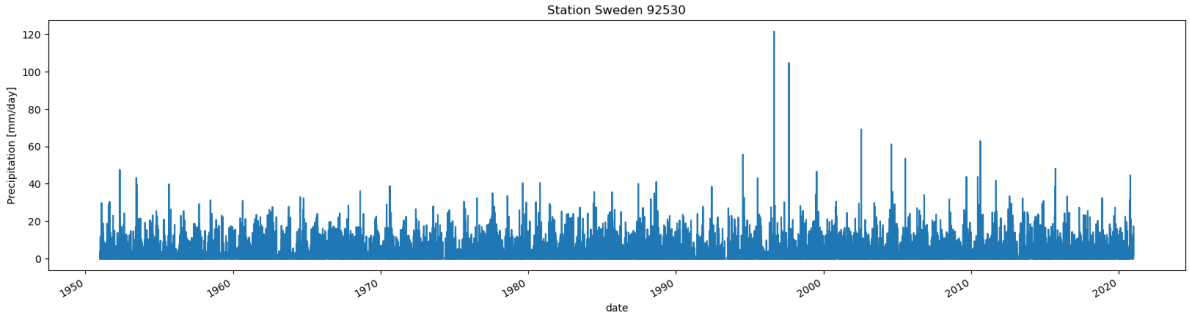


Figure D.2: Time series of station in Sweden which show outlier in the GEV analysis.

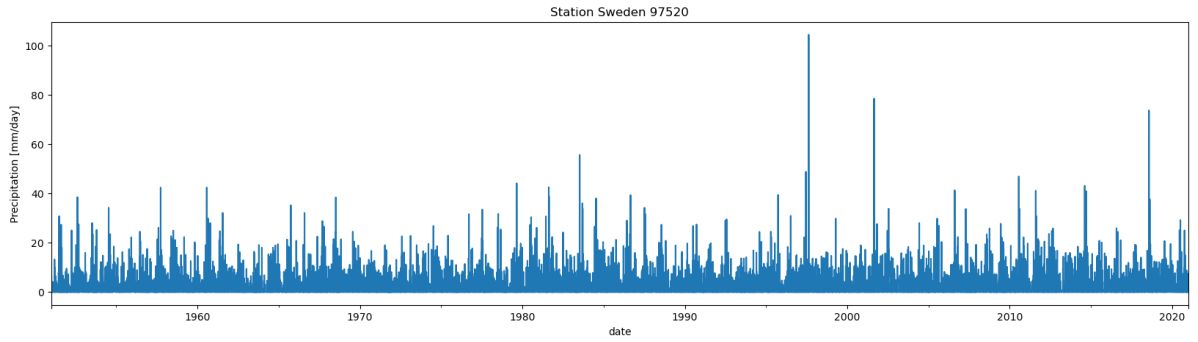


Figure D.3: Time series of station in Sweden which show outlier in the GEV analysis.

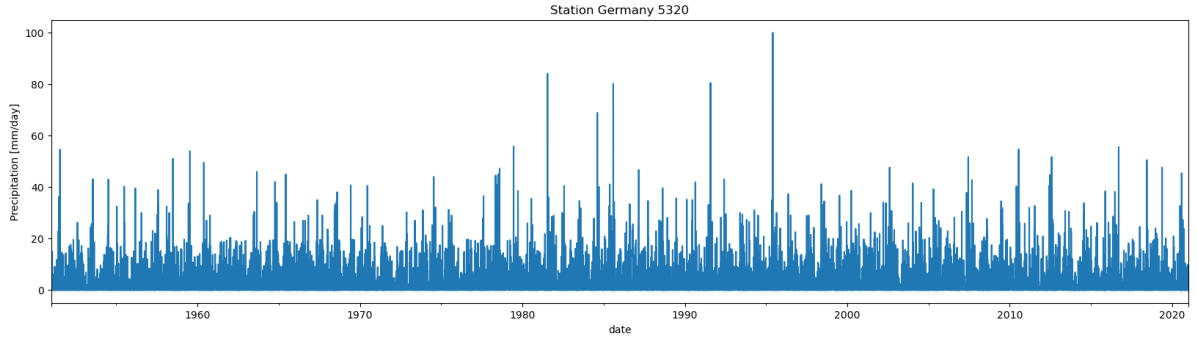


Figure D.4: Time series of station in Germany which show outlier in the GEV analysis.

E

Extra overview of station locations

E.0.1. Stations with both precipitation and temperature data

Table E.1: Overview data availability after quantity and quality control. The second column contains the source of the data for each country. Per country the number of stations with precipitation and temperature data is given, with the density of these stations with respect to the total area of the country. The last row contains the total number of precipitation and temperature datasets.

Country	**stations P and T**	P **(km²/station)**
Netherlands	5	8370
Germany	122	2931
France	84	6568
Sweden	16	28143
Norway	13	29631
Spain	27	18741
Total	267	

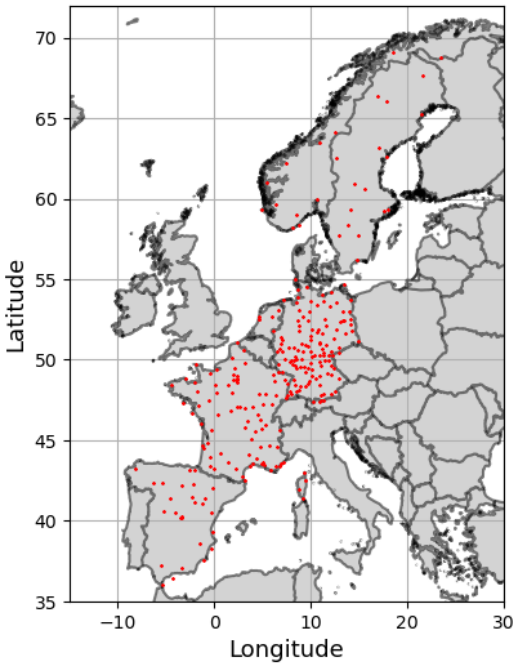


Figure E.1: Geographical visualization of all observation stations used in the temperature analysis for temperature during annual extremes and the common and rare extremes.

E.0.2. Stations dew point temperature

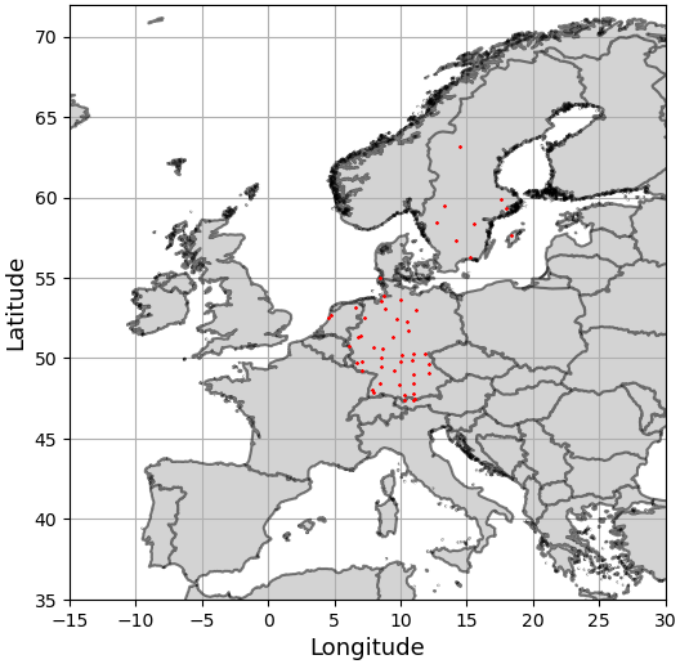


Figure E.2: Geographical visualization of all observation stations used in the dew point temperature analysis. Data is collected from Sweden, The Netherlands and Germany.