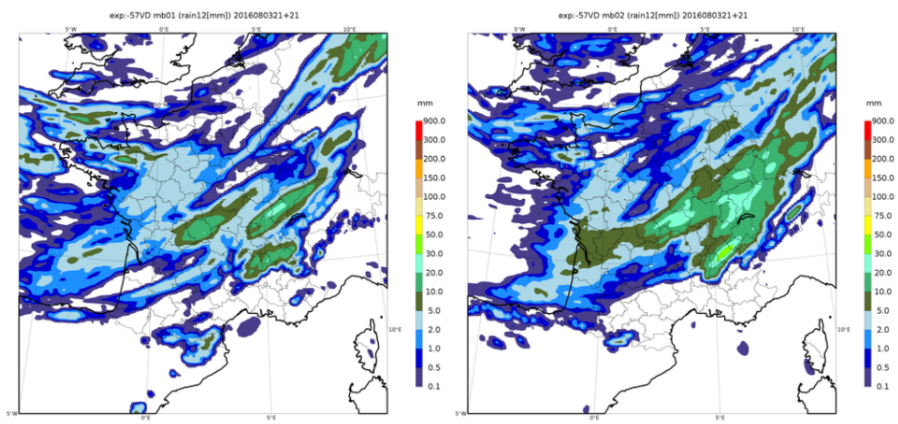
Trend analysis of extreme precipitation events in the Meuse catchment, obtained with re-forecasts from the ECMWF

A bachelor thesis report

Author: **Berend-Jan te Booij**
External Supervisor: **Frederiek Sperna Weiland**
Internal Supervisor: **Jaap Kwadijk**



Source: Encyclopedia of the environment

Institution: University of Twente
Host Company: Deltares
Version: Final Version
Date: 31st May 2022

1 Preface

This final thesis describes the investigations, analyses and conclusions in the context of my Bachelor Civil Engineering at the University of Twente, Enschede, the Netherlands. This assignment, "Trend analysis of extreme precipitation events in the Meuse catchment, obtained with re-forecasts from ECMWF", investigates if the trend is significant enough to be provable with statistical trend tests. The research is conducted at Deltares and aims to contribute to a better understanding on our future climate and further research.

During the project I was given lots of freedom to develop my own ideas, because the use of ensemble re-forecasts was very new. I want to thank my external supervisor Frederiek Sperna Weiland, who gave me helpful advice on how data should be processed. Also a huge thank you goes to Jaap Kwadijk who supported me during the entire project and critically looked at my work. It pushed me to go deeper into the resources and provided me relevant feedback, observing my work from a different perspective.

Finally, I would like to thank my parents who supported me during the whole project, even when sometimes the motivation and pressure was not 100% on point. Their patience and support helped me, making this paper understandable for people who have less knowledge in area of this study.

I hope this research is interesting for the readers and will be used for further research.

2 Abstract

A fast-changing climate results in extreme events, that become more extreme as time proceeds, in the Meuse catchment. Therefore, extreme weather events from the past are not that relevant for the present. This reduces the amount extreme events that are suitable for trend analysis. Due to the shorter observation period that is relevant, the technique of ensemble re-forecasts is used to 'increase' the number of extreme events of the most recent years (1996-2015). By applying a trend analysis, it is possible to see if those extreme events indeed become more extreme over time. This is beneficial for water management and shows how drastic the change of extreme events is. This research studies, if there is a positive trend in re-forecasts of the European Centre for Medium-Range Weather Forecasts (ECMWF) in the Meuse catchment.

The trend analysis is applied to different scenarios, to see whether there is a positive trend in short or long precipitation events. The type of events are one-day, three-day and five-day events. Additionally, an analysis of annual and seasonal events (summer and winter) was done, to determine if there is a difference between summer and winter. A total of four statistical tests have been used for the trend analysis, including a power of detection analysis and extension of the data.

The results showed a rather low detection power for all statistical tests. However, a positive trend was present for all types of events annually. Only the one-day event received a p-value below the critical 5%, making it very significant. Especially the summer events have a large influence on the annual trend analysis. The winter events show a negative trend, proved by the very low p-values of the statistical tests and the visual analysis.

Contents

3 Introduction

During the past decades, global warming has become an increasingly important topic worldwide. Due to the accumulation of greenhouse gas emissions from energy production and fossil fuel use, long-lasting environmental changes can be seen [16]. This also has an effect on weather events around the world, such as increased extreme droughts, heat waves, and cyclones [7] [14] [5].

It is widely acknowledged scientifically, that due to climate change, extreme weather events increase in magnitude and frequency [9] [18] [13] [26], especially an increase in single day events [1]. In water management practice, professionals are preparing for extreme events which are, like the climate, gradually changing [25] [22] [23]. The preparation will be much more complicated if frequency and magnitude of those events would change in a disruptive way rather than gradually. This could mean that old observations are less relevant for trend analysis and decreasing the effective time span of the observations that are usable for trend analysis. This makes a rare extreme event even rarer. Therefore classical approaches, for example observations, are no longer reliable for estimating the trend in extreme weather events. This will have severe consequences for the engineering designs and the way they work, as well as the the social and economical consequences [10].

These observations are not new to the scientific world. However, the events in the last decade may be signals, making scientists and decision-makers aware, that alternative approaches should be considered rapidly. Using observational records to prove or disapprove that extremes do not change gradually but disruptively requires decades of data that are not yet available due to the short observational period [2]. This would mean that for a long period the hazard of extreme weather events is much larger than the prediction model assumes. An alternative approach for assessing historic trends on extreme weather events is the use of ensemble weather re-forecasts. These are forecasts that are generated by present models for past dates. Those re-forecasts are beneficial because they consist of several ensemble members with each member representing a realistic forecast of the weather for that specific re-forecasted date. Instead of using a single observation, it allows several outcomes to be used. It results in a larger set of available data and can reduce the required time frame. However, this in addition with longer records, the estimates of the extreme weather events can become more robust.

The Meuse catchment is an area where information about precipitation is of high importance. Extreme events that have happened in the past years have resulted in new records of water discharges and water levels [17] [21]. To help water management of the Meuse catchment, information about the trend of rainfall is therefore critical in order to evaluate the measures taken so far.

3.1 Problem statement

The Meuse basin is an area where extreme precipitation events occur more and more frequently [8]. One of those events was the flooding in July 2021, which was called a unique event [21] and resulted in record-breaking peak discharges at some locations. This leads to the question of whether those type of events are still rare events or are part of the present world. A trend analysis can give more insight, however is hampered by the limited length of the observational records. This, because old observations are less relevant for the climate where we live in. This could mean that the hazard of extreme weather in the river basin of the Meuse may be much larger than is assumed currently. This Bachelor thesis aims to test whether this ca be overcome by using ensemble weather forecast data.

Research objective The objective of this research is to find out whether a positive trend in the precipitation data of the ensemble re-forecasts can be identified, focusing on the extreme precipitation events in the river Meuse catchment area.

3.2 Ensemble re-forecasts

In this research the ensemble re-forecasts produced by the ECMWF (European Centre for Medium-Range Weather Forecasts) are used. These re-forecasts are forecasts for the whole world of past dates, that are produced with a present model making them more accurate [24]. Ensemble re-forecasts of the ECMWF are produced with a numerical weather prediction model which is a type of model that is capable of dealing with uncertainties. One source of uncertainties are the initial conditions that are used for the forecasting model and can lead to large forecast errors. A second source of increasing uncertainty is the model itself: for example the simplification of the dynamics and physics of the atmosphere in numerical algorithms. Both errors interact with each other. To account for the uncertainties, ensembles are used, which help predicting, quantatively, the probability density of the state of the atmosphere in the future [12]. The ensemble comprises a series of members, each being a single weather forecast, created with a unique set of initial conditions [24]. Although the outcome is different, each ensemble member is a possible weather prediction and therefore gives us additional information about the possible behaviour of the weather at that point in time. It enables the assortment of all different weather re-forecast predictions and to filter them on specific properties such as a minimum or maximum precipitation or temperature. Especially for extreme events, this can be helpful because of their low occurrence.

To illustrate such an ensemble, an example of a prediction of the route of Hurricane Sandy in the United States of America is used. Figure 1 shows the outcome: Different routes, representing the different ensemble members, were possible at the time of the prediction. The different colours represent the different time periods, which show the spread of the different routes: the farther the prediction is out, the larger the spread. Every modelled path is a possible trajectory [4]. The objective of this thesis is to see if there is a possibility of finding a trend in the full set of ensemble forecasts combined over a period of several years.

The data used for this project originate from a numerical weather prediction model operated by ECMWF which produces the re-forecasts. The available time period from March 1996 to March 2016, a total of 20 years of data, is made available by Deltares. The data can be found in the Meteorological Archival and Retrieval System (MARS). A ensemble is generated every Monday and Thursday, each with a total of ten ensemble members, each for a period of 15 days into the

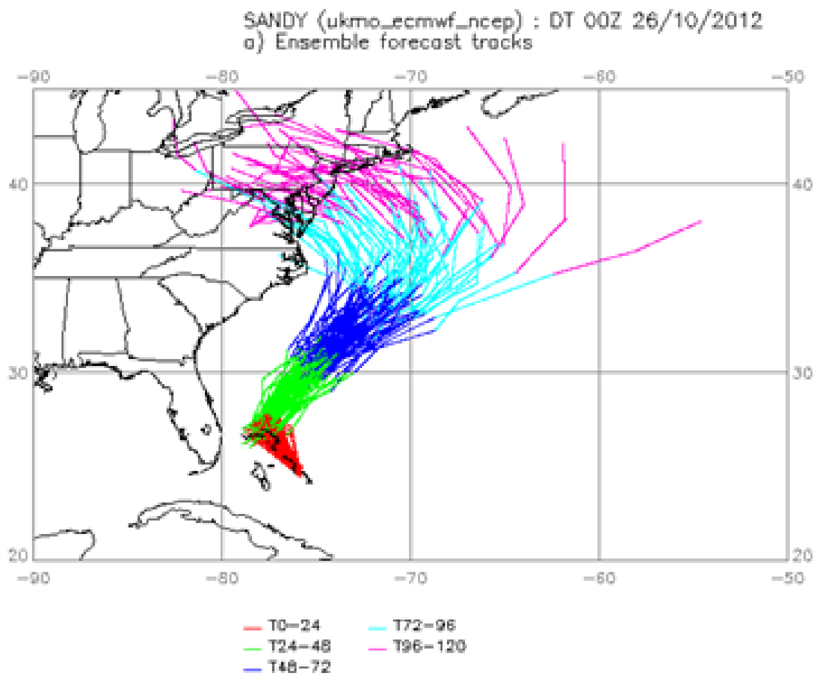


Figure 1: The different possible trajectories are shown at the time of the prediction. The time periods are shown by changes in colour.[4]

future with time steps of six hours. This means that between 40 and 50 different predictions are available for each day during the time period. This can be calculated by dividing the length of the forecast by the frequency of the forecast.

3.3 Studies conducted in the past

Previous research has used ensemble re-forecasts to either increase the observation period or the sample size in order to obtain a more reliable result. One re-forecast exists of ten ensemble members each predicting a possible and realistic weather forecast. To increase the total observation period those forecasts can be combined to produce an extended time series [6]. Another application is the increase of the sample size of extreme events and therefore improve the representation of extreme events [27].

In 2021 master student Daniel Eduardo Villarreal Jaime aimed at deriving extreme return periods more accurately for the river Vecht [6]. This was done by using hydrological ensemble forecasts to extend the discharge records of the project area. The project provided a satisfactory result by reducing the uncertainty from 1.5 to 3.5 times, compared to the estimation from extrapolation of observed data. These uncertainties depended on the length and independence of observed records, the number of ensemble members, and the ability of the model to produce the extreme events and lead time. However, with a special case where Daniel picked out two extreme precipitation events, the peak discharges obtained were not the maximum peak discharges in the records, which indicated that the size of the discharge also depends on the initial conditions in the catchment area, next to the precipitation. The methodology used could help identifying very extreme events with low probability of occurrence to prepare for flooding and idealize the initial conditions to reduce the peak discharge.[6]

In an other report, called "Using ensemble reforecasts to generate flood thresholds for improved global flood forecasting", the authors aimed to explore the potential benefits of using ensemble river flow reforecasts to generate flood thresholds that can deliver improved reliability and experience, increasing confidence in forecasts [27]. A flood threshold describes the maximum discharge needed for floods to occur on land. These are often determined by an extreme value distribution on a set of annual maxima. By using ensemble reforecasts the sample size was aimed to be increased due to the increase of the data in a certain time period. It allows to find extreme events that may no occur in the typical 30-50-year-long sample of traditional observation.With the use of ensemble-reforecast-based-thresholds, the global average forecast reliability and quality improved over the reanalysis-based threshold for up to the evaluated day 30 lead time. This is possible because of a better representation of the extreme events in the forecast climatology by using ensemble reforecasts to increase the sample size.

3.4 Research questions

Based upon the research objective, a number of main research questions are defined. These research questions concern [1] the determination of the extreme precipitation events and [2] the analysis of a possible trend therein.

- **How can extreme precipitation events be defined, using the ensemble re-forecast methodology?**

As mentioned in section 3.2, Ensemble re-forecasts, the low amount of precipitation data is a problem for finding trends, and therefore ensemble re-forecasts are used to extend the data. The words "joint together" in the first sub question refers to the extension of the dataset.

  - How can the data of the ensembles be joint together to increase the data?
  - How can extreme precipitation events be filtered out of the daily precipitation?

Extreme precipitation events are needed before any trend analysis can be performed. This requires the formulation of a methodology to obtain a usable dataset.

- **Is a trend detectable in the obtained extreme precipitation events?**

For this main research question no sub-questions are needed. The focus of this research is towards a positive trend and therefore one-tailed statistical tests are applied on the data set.

3.5 Methodology

In order to answer the research questions, a methodology was formulated that shows the steps taken in this thesis. Figure 2 shows the steps that are carried out. The bold-printed steps are targeted to answer the two main research questions.

It begins by defining the type of events that will analysed in this thesis, which is described in section 4. In that section, the raw data used is explained and how the data is process in a MATLAB model. It is used for extracting the daily precipitation and calculating the extreme events. This will result in a list that contains all extreme events. To receive a table with the ten most extreme events for each year a filter is applied that assures the independence of the events, where the determined time gap is part of.

In section 5 the statistical tests are described, which includes the setup of a null-hypothesis. Additional to that, the applied sensitivity tests are explained that will give more insight on the behaviour of the applied statistical tests.

The results of the statistical tests and the data selection is presented in section 6 for the different extreme precipitation events. The outcome of the sensitivity tests is also explained. The discussion about the results can be seen in section 7, which also includes a recommendation for future research projects.

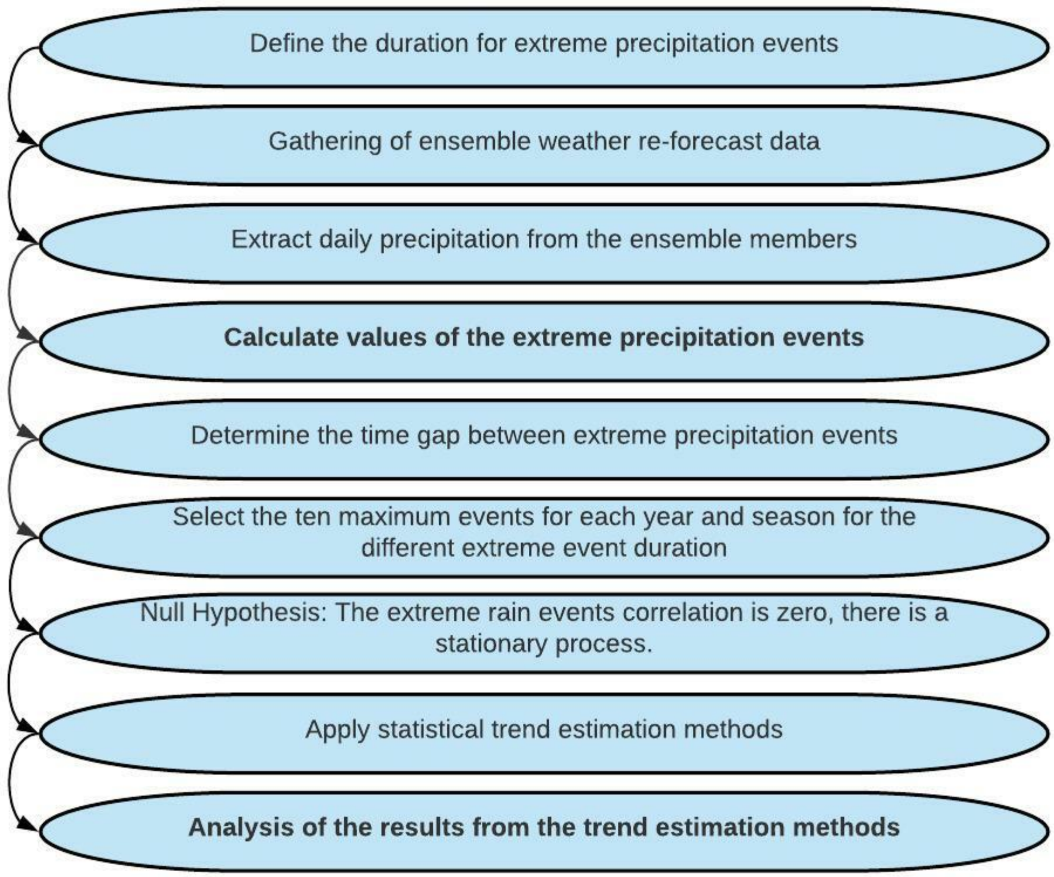


Figure 2: The steps used in this thesis are described above. The bold printed sentences are steps with results to answer the research questions.

4 Approach

The following section explains the time series and the raw data used for this research. Additionally, the boundary conditions are summarised, as well as the different scenarios in which the statistical tests are applied. The raw data functions as input for a MATLAB model that extracts the extreme precipitation events, explained afterwards.

4.1 Time series

The data used in this research are re-forecasts from March 1996 until March 2016 which are made available by Deltares (see Section 3.2 for more information). The data set consists of re-forecasts which are made twice a week with either three or four days between the release of the re-forecast. Each re-forecast consists of ten perturbed forecasts and one unperturbed forecast. In this research the focus lies on the ten perturbed forecasts which are individual ensemble members that predict the weather for 15 days with a time step every six hours resulting in 60 time steps. In figure 3, a scheme of the data is shown which is an extraction of the year 1996. Due to the frequency of twice a week and the forecast horizon of 15 days, for each day 4-5 forecasts are available (see bottom row in figure 3). This results in a total of 40 or 50 individual re-forecasts, each predicting a different but realistic weather scenario.

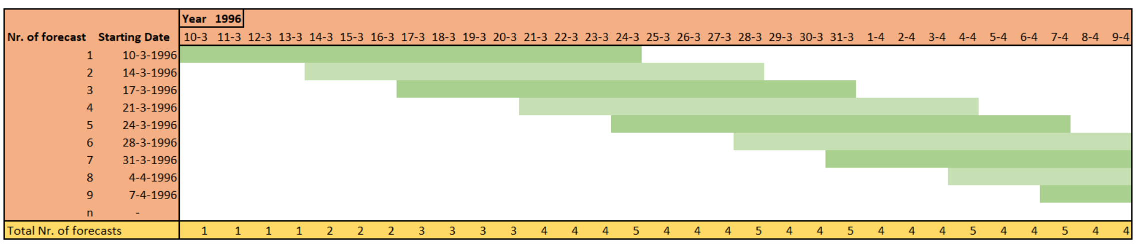


Figure 3: This image shows the overlap of the individual re-forecasts and how many data is available for one day.

The data files received from Deltares consist of a 5-D matrix. The five parameters are listed in the table below. By filling in the first four parameters the fifth, which is the precipitation, is obtained.

1. Longitude
2. Latitude
3. Ensemble member
4. Time
5. Precipitation

The Meuse basin is the study area to which this research method is applied. In figure 4, on page 11, the Meuse basin is indicated by the white and blue area, where the thinner lines display the tributaries and the thick blue line the main river Meuse. The red square is the area, which is used as a simpler replication of the Meuse catchment area.An extreme rain event can last for different periods, from several hours up to a week or even longer. The choice in this research was put on one-day, three-day and five-day events. This because they have different effects on the Meuse catchment, where one-day extreme events put a high strain on tributaries and five-day extreme events on the main river Meuse. To allow a better comparison between two points, a point in between is chosen which is a three-day extreme event in this case.

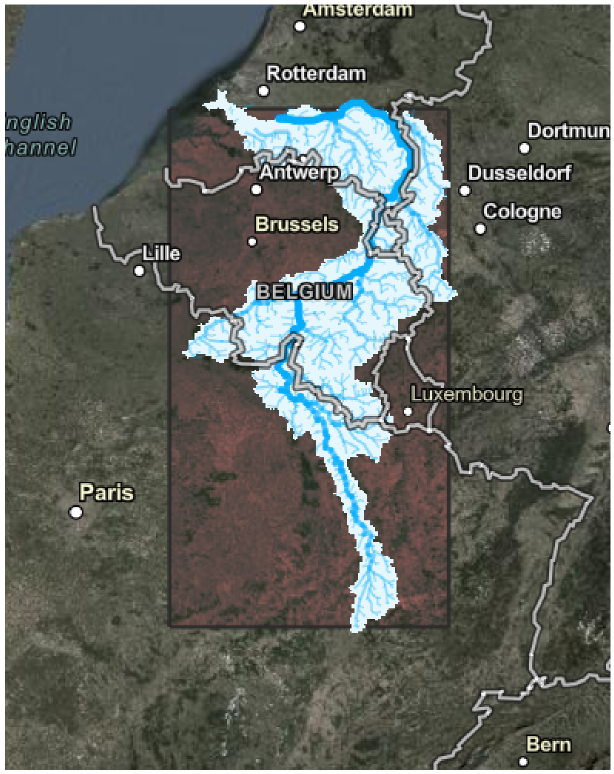


Figure 4: The blue area shows the Meuse catchment and the red square displays the simplified area of the Meuse catchment used in this research

Some reports show that the intensity and amount changes differently between summer and winter [3], which is the reason for analysing the data on trends annually (calendar year) and seasonally (summer and winter). Short one-day extreme events are typical for the summer and events for several days for the winter. The exact dates for the winter and summer are displayed below in table 1.

As mentioned before one re-forecasts consists of ten ensemble members. From each ensemble one event is selected, resulting in a total of 1040 events annually. This is because instead of 365 days of data a total of 15600 days are available. A short explanation of how that number is accomplished is shown below.

$$52 \text{ weeks} * 2 \text{ re-forecasts} * 10 \text{ ensemble members} * 15 \text{ days} = 15600 \text{ days}$$

For further analysis only the highest ten events per period (year or season) are selected to be used. This means that in total 200 events are used which is a reasonable amount for applying statistics and much more than when performing statistics on observations.

Table 1: The time period of the summer and the winter used in the model

summer	**1 April - 31 September**
winter	**1 October - 31 March**

When choosing the ten most extreme events for each period some rules are applied to prevent that events from the same day are used or belong to the same precipitation event.

1. The extreme precipitation events do not have the same date
2. Between the extreme precipitation events are at least two days
3. The day of issue determines to which period (annually, summer or winter) the events belongs

The reason the dry period is two days is because the ensemble members can shift the precipitation, i.e. there is a forecast uncertainty in the timing of the events. A re-forecasts on Monday 10-3-1996 has one ensemble member predicting the one-day extreme on 13-3-1996 and another ensemble member predicts the one-day extreme on 14-3-1996. Only the most extreme of the two is chosen, even if both belong to the ten most extreme events in 1996.

4.2 MATLAB model

To process the raw data from ECMWF a MATLAB model was developed to calculate the extreme precipitation events and sort out the ten most extreme events for each period. Below in figure 5 a flow diagram is shown of how the extreme events are selected from the data. How the precipitation retained in a grid of the whole area in time steps of six hours, became an average precipitation over the whole area per day. A total of 2000 re-forecasts make up the total raw data. The process below is used for each year separately resulting in 100 re-forecasts, which are a total of 1000 ensemble members, per loop (one run through the process).

The model picks out the first re-forecast and then chooses the first ensemble member and calculates the average precipitation over the whole area per time step (every 6 hours). This means that the precipitation of all grid cells is summed and divided by the total number of cells. Having done that, the time steps of 6 hours are summed to obtain the daily precipitation. Depending on the event length(one, three or five days), a moving sum is applied to the data. The date with the highest value is selected and saved in a table.In that table, the following attributes are gathered from the event:

- year of the re-forecast
- month of the re-forecast
- day of the re-forecast
- days since 0-0-0000 of the event
- ensemble member
- day in the ensemble number (between 1-15)
- precipitation in m

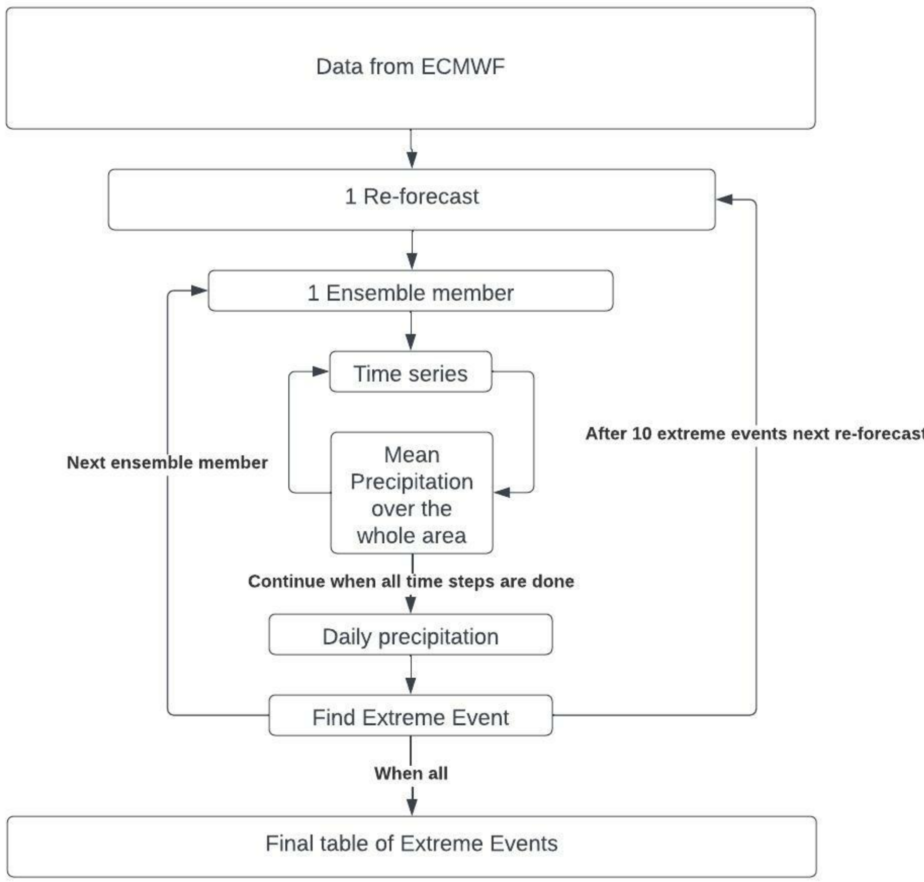


Figure 5: Process description of how the model calculates the extreme events from the raw data

With the date of the re-forecast and the day in the ensemble member, the exact day is calculated on which the extreme event occurs, which is expressed in days since 0-0-0000. This allows for the comparison of dates, even if there is a change in months or years, because it is a continuous value. A detailed description of the selection of the events can be found in the appendix in section 9.1. Having finalised a table with all extreme events in a year, a sorting and filtering process is used to pick out the ten most extreme events in a specific period (annually or seasonally), avoiding events that overlap. That process is visualised in figure 6 below.

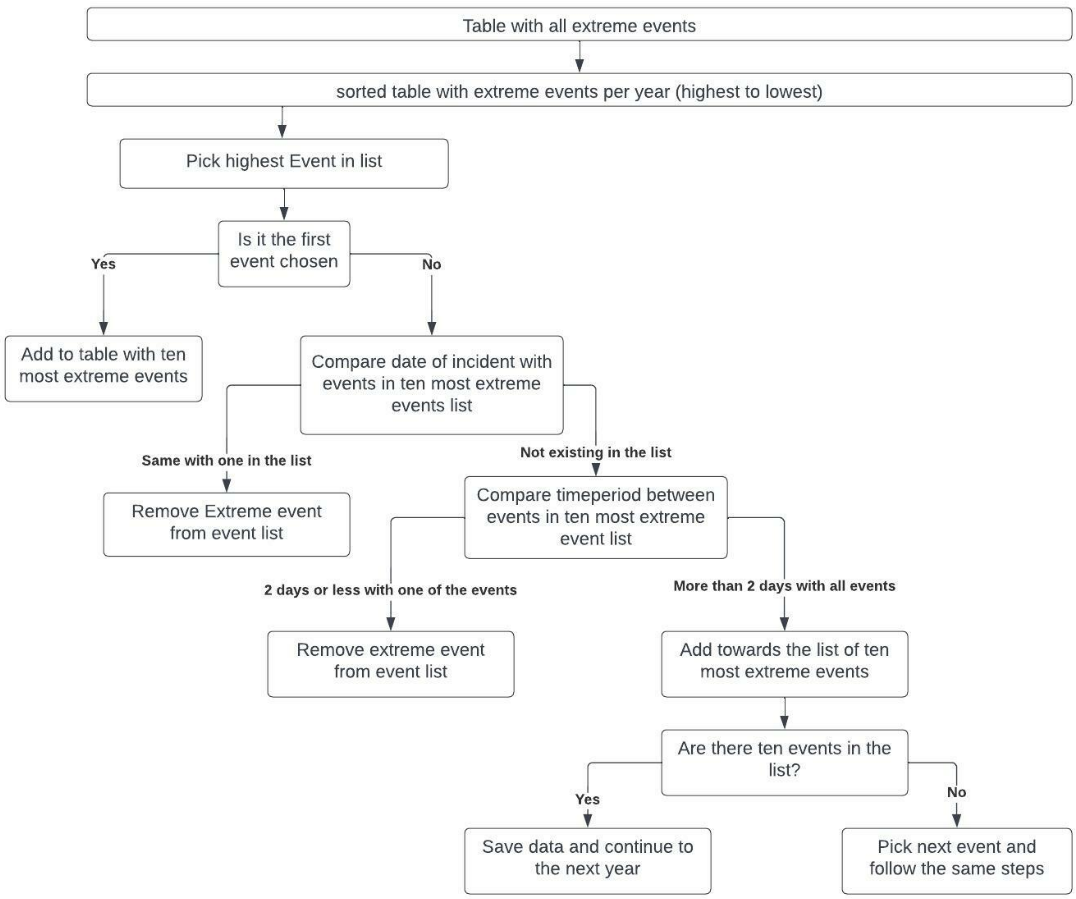


Figure 6: The flow diagram above shows the process of obtaining one list with the ten most extreme events obeying the rules of the date of incident and the time gap between events.

To begin with, the table of extreme precipitation events is sorted by precipitation from highest to lowest (descending order). The highest value is picked and forms the beginning of the list with the ten extreme events. From then on, the next highest value is picked and is compared with the date of the incident from the events already in the list. If there is no similar day the event is kept, otherwise the event is removed. The second comparison is the time gap between the events. That time gap is always two days between the end of one event and the start of the next event. This is again compared to all events in the list of the ten extreme events. If there is no coincidence, the extreme event is added to the final list of the ten extreme events of that specific year. When the list consists of ten events, the process is stopped, and the next year is processed. The same is done with the data for the seasons winter and summer; however, the beginning dates of the extreme events determine if the event belongs towards the winter or the summer. A detailed description of the filtering process is described in the appendix in section 9.2.

5 Data Analysis

The data obtained using the MATLAB model described in the previous section is at first inspected visually, see figure 7. This gives a first impression, if a trend exists in the data. Additionally, the data is tested by statistical trend tests to show the significance of the possible trend. The values obtained of the statistical tests show the relation between the values of the extreme precipitation events and their position on the time line. This either supports or disproves the visual findings, if present. For applying the statistical trend tests, a statistical hypothesis test is done for which a null hypothesis and a alternative hypothesis are postulated. The objective is to reject the null hypothesis and therefore accept the alternative hypothesis as true.

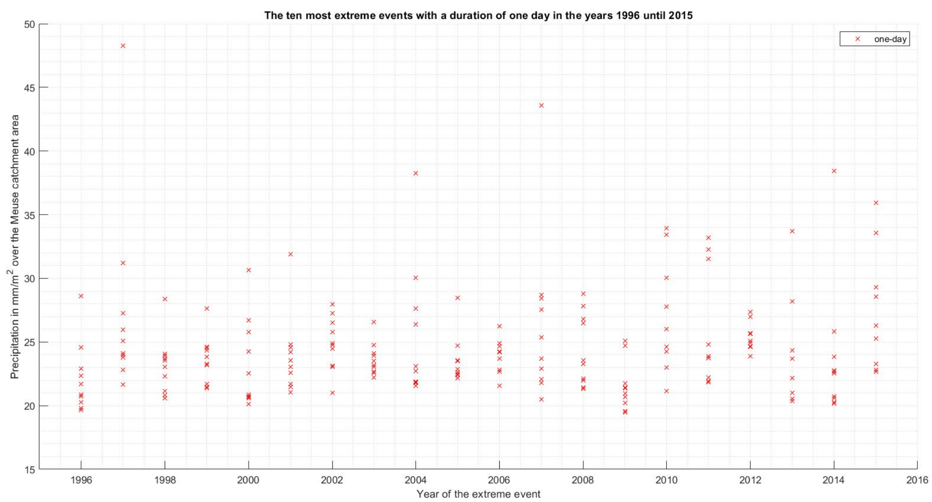


Figure 7: Caption

To begin with, a null and alternative hypothesis need to be formulated. The test conducted will be a one-tailed test because the research only aims to find positive trends. In the following, the null hypothesis (H_0) and the alternative hypothesis (H_1) are shown.

H_0 : *The correlation between extreme rain events is zero (the process is stationary).*

H_1 : *The correlation coefficient between extreme rain events is greater than zero, a positive correlation could exist.*

With the use of the statistical trend test the procedure is to reject the null hypothesis with sufficient evidence in the form of statistical tests. In case of rejection, the alternative hypothesis is favoured instead, would indicate a trend in the extreme rain events over the Meuse catchment, hence there would be no stationary process. A p-value is used to find out if the null hypothesis should be rejected or not. The value can be between zero and one, where a lower value means stronger evidence for rejecting the null hypothesis. A p-value of 5 percent is statistically significant and therefore used in this experiment. A value below 5% shows that the null hypothesis is false with more than 95% probability. In the case of a p-value above the significance level, the question is at what point the results become significant. For this, the data will be manipulated in two different ways, in order to verify if it is due to insufficient observations or if there is no trend at all.

5.1 Statistical tests

Four different statistical trend tests are applied to the observed data. Hereafter follows a short description of the test.

5.1.1 Pearson t-test

The Pearson t-test is the only parametric test used in this research. It measures the strength of a linear correlation between the measured extreme rain events and their date of happening. The Pearson correlation coefficient does not represent the slope of the line of best fit but how the variation of the data towards the line of best fit is. For a valid result the data should comply with several assumptions as following a normal distribution.[20]

5.1.2 Spearman rank test

This test is a non-parametric test where the correlation coefficient measures the strength and direction between extreme weather events and their date of happening by ranking. By using ranked values no assumption of the distribution function is needed. Instead a monotonic relation is preferred which means that when one variable increases, so does the value of the other variable. The correlation value is calculated by the difference in ranking of the x value and y value of all data points.[19]

5.1.3 Wilcoxon-Mann-Whitney test

The third test applied in this research is called the Wilcoxon-Mann-Whitney test which compares the difference between two independent groups. It allows to split up the data set into two independent groups and compare the years at the beginning with the years available at the end of the observation time period. The two independent groups are combined and ranked, and the sum of the ranking of the two separate groups are compared to each other. In this research, the first five years will be compared with the last five years. In research 5 years is a rather short period, however this is compensated by the use of the ensemble members.[11]

5.1.4 Anderson-Darling

The last test used for this research is the Anderson-Darling two-sample test, which investigates the distribution of two different data sets and whether they are the same or different. Instead of only using the ranks of the data as the Wilcoxon-Mann-Whitney, Anderson-Darling also takes the difference between the data points into account. Same as the Wilcoxon-Mann-Whitney test, it uses the first five years and the last five years from the data set.[15]

5.2 Sensitivity analysis

The result from the statistical tests can lead to two different outcomes: either there is a trend or not. A sensitivity analysis will give insight in the strength of the tests and at what point in time there may be a chance of finding a trend. These help to understand the results in case no trend is found. If there is indeed a trend, these additional tests can improve the strengths of the outcome. In the following, two methods are explained that are used in this research:

- Analysis of the detection power of the statistical tests with the use of a generated series derived from observed data
- Extending the data set by repeating the most recent years

The method applied to the data is described in the section below.

5.2.1 Detection power of the statistical tests

For the detection power of the statistical tests, a generated series of extreme precipitation events is used. For these series, a normal distribution, obtained from results of the annual one-day extreme events from the re-forecast data, is used with the following mean and standard deviation:

$$\text{Mean:} = 0.024519$$

$$\text{Standard deviation:} = 0.004063$$

A total of 200 series was generated with the values mentioned above. For the detection power, two different methods are used to see at which point a trend is possible to be detected in the data. The first one is using a fixed time period of 20 years as the original data in this research. A forced linear trend will then be applied to the data which has different strengths as 5% and 10%. This will be done by multiplying the precipitation values by a factor that increases linearly over time, resulting in increases of either 5% or 10% at the end of the period. Using statistical tests, the probability of detecting a trend was evaluated. Therefore, the outcome is displayed as percentage, indicating for how many series it was possible to detect a trend.

The second method uses a fixed linear trend of 5% over a period of 20 years. This time, the time period is the variable and changes between 10 and 30 years. This means that the trend of 5% over 20 years is continuous up to 30 years, resulting in a total increase of 7.5%. Again the data will be multiplied by a factor that increases linearly in time. The outcome will also be displayed as a percentage, indicating how many series a trend is detected. It allows us to find out at which time period a higher chance is to find a trend.

5.2.2 Extension of the data set

To study the behaviour of a possible trend in the future years where no trend is found yet, recent years are used to extend the data-set in two different ways. This is done to see, if the existing trend continuous with the same speed over the next few years, the significance changes below the critical level. Events in the future are likely to be similar to those of the most recent years.

The first one is by adding the most recent events (last year 2015). The last year (2015) will be added to the observation period, resulting in 21, 22, and so on, up to 25 years of observation.

The reason for not going beyond 25 years is because the Wilcoxon-Mann-Whitney and Anderson-Darling tests compare the first five years and the last five years of the time period. Adding a sixth year does not influence the data, and therefore results in the same result.

The second method, is using several recent years, which are the years 2011 until 2015. For one additional year only year 2015 is used. When the data-set is extended for two additional years, the years 2015 and 2014 are used. This continuous until five years are added, which means that year 2011 until 2015 is added a second time to the time series of 1996 until 2015. This results in a total of 25 years, equal to the first method. Having the same number of years in both methods allows a better comparison of the results.

The results obtained will be a p-value. Additionally, if the p-value goes below the critical value, the year in which that was achieved is written next to it. This means that when the critical p-value is achieved when the most recent year is added three times, a value of Y23 will be written down. It means that when the time period is 23 years, the critical p-value is achieved and a trend is detected statistically by extending the data series. This gives a quick overview at what time period length a significant value is exceeded.

6 Results

In the following section, the results of the statistical tests are shown and interpreted. These will be done separately for the annual extreme events and seasonal extreme events. At the end, the outcome from the sensitivity analysis and extension of the data are described.

6.1 Annual Extreme Events

The primary focus of this research was on the change in annual extreme events. In figure 8 the annual extreme events are shown grouped by their year of occurrence. The plotted linear lines show an increase in extreme events for all three durations, especially for one-day extreme events. The five-day extreme events show some positive trend, but surprisingly there is rather no change in the three-day extreme events. Astonishingly some one-day extreme events of some years are larger than some three-day or even five-day events in other years. Between three-day and five-day events, this is even more present.

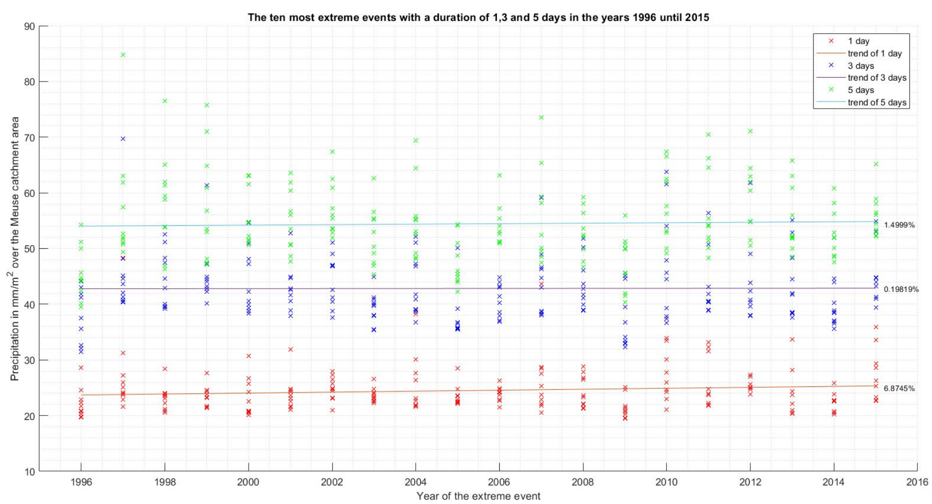


Figure 8: The graph shows the ten annual extreme events between 1996 and 2015 for the one-day, three-day and five-day events inclusive the trend-line determined by least square linear regression. A larger representation can be found in the appendix, in section 10

In table 2 the results of the four statistical tests that were applied to test the significance of the trend are shown. Based on these results, the null hypothesis can only be rejected for the one-day extreme event with the Pearson t-test, Spearman and Mann-Whitey-Wilcoxon results which have an outcome below the defined critical level of 5% (see the green values in table 2). It can be said that with (more than) 95% certainty, there is a positive trend for one-day extreme events. For the other events this cannot be stated, as none of the results are below the critical level of 5%. The five-day extreme event is closer to rejection of the null-hypothesis than the three day-extreme event. The results of the statistical tests reflect the trend line obtained by the least squares linear regression in figure 8.

Table 2: Outcome of the statistical tests applied to annually extreme events

Statistical test	**one day**	**three day**	**five day**
Pearson t-test	3.9%	46.7%	29.2%
Spearman	2.9%	67.9%	13.4%
Mann-Whitney-Wilcoxon	2.4%	74.2%	12.2%
Anderson Darling	5.9%	48.0%	8.1%

6.2 Seasonal Extreme Events

After knowing the results of the annual extreme events, the question is if there is a difference between the events in the summer and winter. Below in figure 9 the summer events and in figure 10 the winter events for one day, three days and five days are shown. Additional to the events a trend line is added which is determined with the least square linear regression. The graphs show some surprising results:for example, the one-day and five-day extreme events only show a positive trend in the summer and not in the winter. That negative trend is noticeable for three-day events in both summer and winter. Comparing the extreme events of the summer and winter, the one-day events have similar values, however the five-day events differ. In the winter more extreme events occurred and a clearer division between one-day, three-day and five-day events exists. In the summer the different events overlap each other, where many three-day events are similar to five-day events.

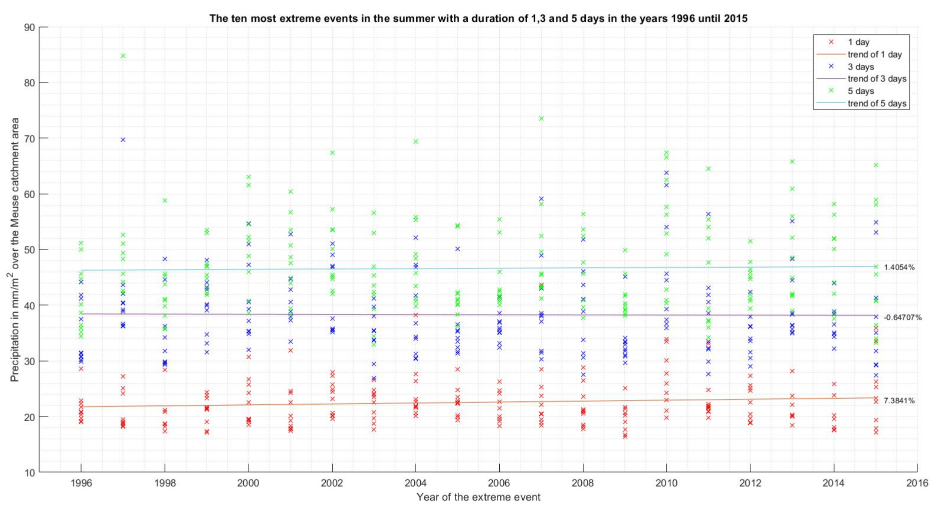


Figure 9: The graph shows the ten annual extreme events between 1996 and 2015 for the one day, three day and five day events inclusive the trend-line determined by least square linear regression for the summer season.

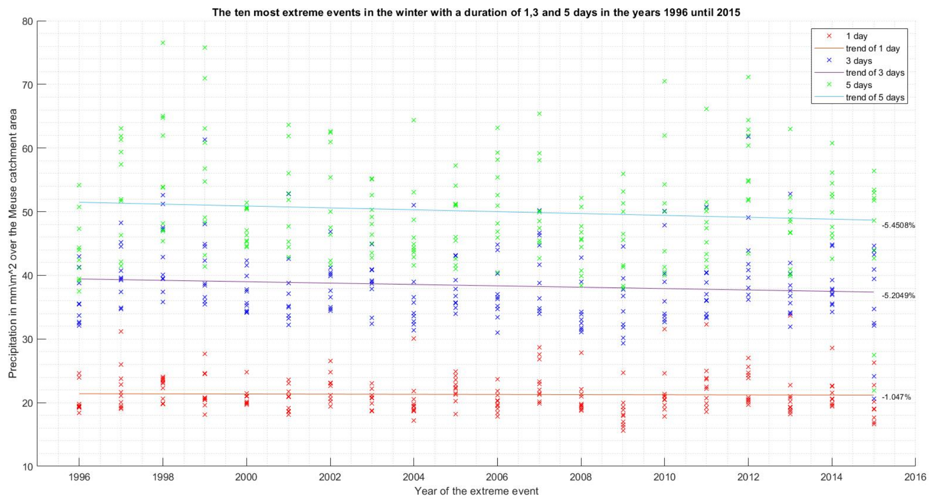


Figure 10: The graph shows the ten annual extreme events between 1996 and 2015 for the one day, three day and five day events inclusive the trend-line determined by least square linear regression for the winter season.

In Table 3 the results of the four statistical tests are shown that are applied to test the significance of the trend. Only the outcome of the Spearman and Mann-Whitney-Wilcoxon for the one-day extreme event in the summer are below the defined critical level of 5%, resulting in the rejection of the null hypothesis. For all other scenarios, the results mostly do not come close to the defined critical level, and therefore the null hypothesis cannot be rejected. These results support the data in the graphs in Figures 9 and 10, where the negative trend lines are supported by high significance levels in the table.

Table 3: The statistical result of the tests applied to extreme precipitation events in the summer and winter season.

Statistical test	**one day**		**three day**		**five day**	
	summer	winter	summer	winter	summer	winter
Pearson t-test	6.9%	63.0%	54.7%	94.2%	35.9%	93.3%
Spearman	3.9%	84.1%	67.7%	92.9%	46.8%	71.9%
Mann-Whitney-Wicoxon	3.8%	68.8%	73.5%	84.9%	49.2%	56.6%
Anderson Darling	10.2%	45.3%	48.4%	48.0%	66.1%	79.5%

6.3 Detection power of statistical tests

The first sensitivity analysis is used to see at what point a trend is possible to be detected with the observation period that is available (20 years). Below in table 5 the results from the statistical tests are shown of the data which has been modified with a fake trend of five and ten percent.

With a trend of 5%, the statistical tests detect only a trend between 15% and 30% of the series, where a series is a time period over 20 years with 10 extreme events per year obtained from a normal distribution, . When there is a trend of at least 10%, the trend is significant enough to be detected by more than half the series. Pearson and Spearman showed the same results with the modified data. The results of varying the duration of the observation period show that, when there is a trend of 5% from 30 years onward, more than 50% of the series are significant detected by statistical tests. When increasing the length the Mann-Kendall-Wilcoxon and Anderson-Darling show similar results, where as by increasing the trend different results are obtained. Overall the sensitivity outcome of the statistical test show a rather low chance of detecting a trend. It shows that the chance of detecting a trend of 10%, which is a very strong increase, is zero for 1/3 of the series with the Pearson, Spearman and Anderson-Darling.

Table 4: The results (in percentages) of applying a fake trend of 0%, 5% and 10% to the generated data by a normal distribution.

statistical test	**0%**	**5%**	**10%**
Pearson t-test	6	31.5	71
Spearman	4	31	71
Mann-Kendall-Wilcoxon	6.5	24	100
Anderson-Darling	6.5	17	51

Table 5: The results (in percentages) of a fixed trend in the generated data of a normal distribution, but the time period used remained variable between 10 and 30 years.

statistical test	**10**	**20**	**30**
Pearson t-test	8	31.5	76.5
Spearman	9	31	71.5
Mann-Kendall-Wilcoxon	9.5	24	59
Anderson-Darling	6	17	45.5

6.4 Extension of the data

The extension of the data-set is tested to see how the behaviour of the trend is for all scenarios when increasing the observation period. The aim is to see if a trend is detectable in the future, assuming the existing trend continuous. This can be seen in the p-value and if that is becoming lower or higher. A lower becoming p-value will indicate a trend that becomes stronger and maybe below the critical p-value resulting in a positive trend. The tables below show the results of the two different methods. For each scenario, the year is shown where the p-value is below the critical level of 5% with the p-value. If the critical level of the p-value is not reached, the final p-value and the year (25) is listed in the table.

Table 6 shows the summary of the results of the first method where the year 2015 is used for extending the data-set. In many scenarios the critical significant level is accepted, as, for example, with annual 3 day events a trend is detectable. A different result is achieved for seasonal events with a length of 5 days, where a trend is rather more difficult to be verifiable the longer the observation period is. All p-values from the extension with the year 2015 can be found in the appendix, section 11. Additional to the outcome of the statistical test a table is added that shows how wet 2015 was compared to the other year. The lower the value the wetter the year is.

Table 6: Outcome of the first extension method with the statistical tests applied to annual and seasonal scenarios showing in which year the p-value has reached the significant value. If a p-value of 5% or less is not reached, the final p-value is listed of the 25th year.

First extension method (Year 2015)								
statistical test	Pearson t-test		Spearman		Mann-Kendall-Wilcoxon		Anderson-Darling	
Annual 1 day	**20Y**	*4%*	**20Y**	*3%*	**20Y**	*2%*	**21Y**	*2%*
Annual 3 day	**23Y**	*5%*	**24Y**	*2%*	**24Y**	*1%*	**24Y**	*2%*
Annual 5 day	**25Y**	*8%*	**22Y**	*3%*	**24Y**	*2%*	**21Y**	*4%*
Summer 1 day	**21Y**	*3%*	**20Y**	*4%*	**20Y**	*4%*	**22Y**	*5%*
Winter 1 day	**25Y**	*99%*	**25Y**	*100%*	**25Y**	*100%*	**25Y**	*100%*
Summer 3 day	**25Y**	*28%*	**25Y**	*69%*	**25Y**	*69%*	**23Y**	*5%*
Winter 3 day	**25Y**	*100%*	**25Y**	*97%*	**25Y**	*98%*	**25Y**	*100%*
Summer 5 day	**25Y**	*52%*	**25Y**	*74%*	**25Y**	*80%*	**24Y**	*2%*
Winter 5 day	**25Y**	*100%*	**25Y**	*92%*	**25Y**	*94%*	**25Y**	*100%*

Table 7: The rank of 2015 compared to all years from 1996 until 2015. The lower the rank, the higher the extreme events were in that year. A one means that the year was the wettest year of all years.

Scenario	*Rank*
one-day annual	1
three-day annual	2
five-day annual	8
one-day summer	2
one-day winter	19
three-day summer	9
three-day winter	20
five-day summer	11
five-day winter	20

The results of the second method are shown in table 8. For this the years 2011-2015 were used, depending on the length by what the original data is extended. The results of the annual scenarios show less significant p-values, especially for the three-day extreme events. When looking at the seasonal results, similar values can be seen comparing it with the first method. However the results are less extreme than the first method, where all p-value get close to 100% in the winter. The second method shows similar p-values in summer and winter. All p-values are shown in the appendix, in section 12. It shows that the p-value sometimes moves up or down, depending on how wet the added year has been.

Table 8: Outcome of the second extension method with the statistical tests applied to annual and seasonal scenarios showing in which year the p-value has reached the significant value. If a p-value of 5% or less is not reached, the final p-value is listed of the 25th year.

Second extension method (Year 2011-2015)								
Statistical test	Pearson t-test		Spearman		Mann-Kendall-Wilcoxon		Anderson-Darling	
Annual 1 day	**20Y**	*4%*	**20Y**	*3%*	**20Y**	*2%*	**21Y**	*2%*
Annual 3 day	**25Y**	*36%*	**25Y**	*56%*	**25Y**	*74%*	**25Y**	*48%*
Annual 5 day	**25Y**	*9%*	**24Y**	*5%*	**25Y**	*12%*	**21Y**	*4%*
Summer 1 day	**21Y**	*3%*	**20Y**	*4%*	**20Y**	*4%*	**25Y**	*10%*
Winter 1 day	**25Y**	*34%*	**25Y**	*64%*	**25Y**	*69%*	**25Y**	*55%*
Summer 3 day	**25Y**	*65%*	**25Y**	*73%*	**25Y**	*73%*	**25Y**	*48%*
Winter 3 day	**25Y**	*80%*	**25Y**	*73%*	**25Y**	*85%*	**25Y**	*48%*
Summer 5 day	**25Y**	*44%*	**25Y**	*53%*	**25Y**	*49%*	**25Y**	*66%*

7 Discussion

In this section, the results shown will be critically discussed and reflected on, followed by a recommendation.

7.1 Result Discussion

Both the visual analysis (figure 8) and the trend analysis (table 2) of the annual events show similar results. A trend exist for all three scenarios. Only one-day extreme events are 95% significant for the Pearson, Spearman and Wilcoxon-Mann-Whitney test. This corresponds with literature that tells that there exists especially an increase in short precipitation events. When taking a closer look at figure 8, the most extreme five-day events are not very different in heaviness between 1996 en 2015, however for one-day extreme an increase is noticeable. This can indicate that the peak precipitation values of an event over several days is much higher as often the one-day extreme events are part of the five-day extreme events, instead of having it distributed over the affected days. A different development is noticeable for three days, where a very low trend is shown in the graph that is different to one-day and five-day extremes. An explanation could be that two one-day extremes that are closely forecasted after each other may form one five-day extreme. In a three-day event, this is never possible due to the time gap of 2 days. A possible option to avoid that, is applying a larger time gap for the shorter extreme events, however this maybe results in a larger spread of the most extreme events, which may has a negative impact on the trend analysis.

A surprising difference in significant p-value is obtained between annual and seasonal extreme precipitation events when comparing table 2 and table 3. Only one-day extreme events in the summer receive a p-value that is below the critical value that is set at the beginning of the research, which indicates the increase of short duration extreme events in the summer for the future. The summer extreme events are the events that have the most effect on the positive trend in all three scenarios annually. The winter extreme events rather show a negative trend, that can be seen in figure 10 and the low statistical results in table 3 compared to the results of the summer extreme events. However, part of the precipitation in the winter can fall as snow and therefor is not a problem on the day of incident, but could become a problem for the discharge when it is melting.

The p-values for most seasonal scenarios have turned out to be very high which means that a trend is rather difficult to be detected. A reason for that is the low detecting power of the four statistical test. This can be seen with the forced trend of 10% on the series, where three of the four tests detected no trend for 1/3 of the series. A trend of 10% is very high and having a chance of 33% failing detecting the trend, is rather high. A reason for this is the large spread of extreme event values, where the difference between the highest value and lowest value is quit high. This also means that a possible trend can exist in the data set, however is not detected by the applied statistical tests.

The amount of extreme events chosen for each year also affects the seasonal outcomes. The number of extreme events remained the same for the annual analysis (12 months) and the seasonal analysis (6 months). Choosing, for example, 10 extreme events with a duration of five days, including the dry period (time gap) of two days between the events, a total of 70 days are covered. This means that more than two months are included which are annually about 16%, on the other hand more than 30% for the seasonal time period. Applying a larger time gap increases the time period which makes it difficult to implement. Covering such large time period means that much lower extreme precipitation events are chosen to finally have ten extreme events for each period that have a strong effect on the outcome. Lowering the total extreme events for each period will give more extreme events, however it could results in to few data points which make the tests less decisive. This is the difficult part of balancing the amount of data points for the statistical tests and the total amount time period covered by the extreme events, including their independence depending on the time gap between the events. A possibility is also to let the amount of events chosen, be depending on the total time period covered by the events including their time gap. This will results in less events for the seasonal scenarios, than the annual scenarios.

The results of the extension of the data showed rather less impressive results. For method one, year 2015 was often either very wet or very dry which had a high influence in detecting a trend or not. A different results can be seen when adding the years 2011-2015 where the p-value lowers just slightly but not a lot, therefore becoming slightly more significantly. Therefore the outcome of the sensitivity test give low evidence on upcoming significant trends in the different scenarios. It depends a lot on how wet the most recent years have been.

7.2 Recommendations

As mentioned in the discussion, the number of the most extreme events chosen for the statistical tests may have a negative influence on the trend. A possible option for further research is the analysis on reducing the amount of events without loosing the strength of the statistical test by having less data points. Another factor which can be analyzed further, is the time gap applied to assure the independence of the different extreme events. When increasing the total period of the data until 2020 or further, a deduction of the total extreme events used for the statistical tests is possible.

The model currently uses a very rough replication of the Meuse basin to calculate daily precipitation. Due to the fact that an average over the whole replication is used, values outside the original Meuse catchment can have a negative or positive effect on the outcome. This can be improved by selecting that area much more accurately and even more realistic results can be achieved.

Interesting for water management is also the discharge that is caused by the precipitation obtained from the re-forecasts and what the impact is at Borgharen, the location where the Meuse enters the Netherlands. The question is whether the discharge and precipitation values follow the same trend, or if different results would be obtained due to the characteristics of the Meuse catchment with its tributaries. Of course the position of the precipitation plays an important role for discharges in the river Meuse, however this data can give a first impression of the impact on discharges. The next step then would be taking the movement of the precipitation event into account, as for example when it is moving north, a larger discharge can be expected in the lower areas when discharge from the upper areas collides with precipitation in lower areas.

8 Conclusion

The objective of this thesis is to answer the question, whether a positive trend can be identified in precipitation data in the ensemble re-forecasts of the ECMWF in the Meuse catchment.

The research showed that a positive trend exists in one-day extreme events, annually, according to the Pearson, Spearman and Wilcoxon-Mann-Whitney statistical tests. All three tests yielded a p-value below the critical value. This is mainly caused by the events in the summer period which, compared to the winter period, showed a strong trend. The p-values in the summer period were more significant than the results of the winter period. Analysis of the one-day events during the summer period resulted in a p-value below the critical level. For the three-day and five-day event a trend can be observed as well, but it is weaker and the results are not found to be statistically significant.

The power of detection was low for all statistical tests, which means that having detected the trend for one-day events, clearly shows that there indeed is a trend. However, there is a probability that for the other scenarios, a trend exists but was not detected. Taking into account that the data set used covers the period 1996-2015, plus the fact that the present is 2022, there is the option to expand the data set with at least an additional 5 years, allowing for a much better trend analysis. After extension of the data set, the results shown were significant, but they were depending on the amount of precipitation in the latest year (2015). Using a larger extension period resulted into a slightly lower p-value over time, in the majority of cases. However, also that was associated with a large dependency on the precipitation levels in the latest years (2011-2015). However, in general, extension of the time period of the data set is expected to increase the reliability and validity of the statistical tests.

A rather positive result was the low p-value for the winter events, indicating a negative trend. This implies that an increase of extreme precipitation events in the winter is not expected. For all scenarios based upon the winter period, the statistical significance reduced, in spite of extending the time period. It is recommended that further research focuses on the short extreme events during the summer period.

References

[1] United States Environmental Protection Agency. Climate change indicators: Heavy precipitation. *United States Environmental Protection Agency*, 2021.

[2] F. Diermanse, Jaap Kwadijk, J. Beckers, and J. Crebas. Statistical trend analysis of river rhine and meuse discharges. *ResearchGate*, 2013.

[3] Pavel Ya. Groisman, Thomas R. Karl, David R. Easterling, Richard W. Knight, Paul F. Jamason, Kevin J. Hennessy, Ramasamy Suppiah, Cher M. Page, Joanna Wibig, Krzysztof Fortuniak, Vyacheslav N. Razuvaev, Arthur Douglas, Eirik Førland, and Pan mao Zhai. Changes in the probability of heavy precipitation: important indicators of climatic change. *Climatic Change*, 42:243–283, 1999.

[4] Group on Earth Observations. Weather: Improving weather information, forecasting, and warning. *Thorpex*, 2012.

[5] Stephanie C. Herring. Observing mother nature, a look at severe weather events in a changing climate. *Climate Risk:Emerging Risks*, 2018.

[6] Daniel Eduardo Villarreal Jaime. Ensemble hydrological forecasts to derive extreme return periods: Case study of the overijsselse vecht river using the wflow_sbm model. *Master Thesis*, 2021.

[7] Thomas R. Karl, Richard W. Knight, David R. Easterling, and Robert G. Quayle. Indices of climate change for the united states. *Bulletin of the American Meteorological Society*, 77:279–292, 1996.

[8] Frank Kreienkamp, Sjoukje Y. Philip, Jordin S. Tradowsky, Sarah F. Kew, Philip Lorenz, Julie Arrighi, Alexandre Belleflamme, Thomas Bettman, Steven Caluwaerts, Steven C. Chan, Andrew Ciavarella, Lesley De Cruz, Hylke de Vries, Norbert Demuth, Andrew Ferrone, Erich M. Fischer, Hayley J. Fowler, Klaus Georgen, Dorothy Heinrich, Yvonne Henrichs, Geert Lendering, Frank Kaspar, Enno Nilson, Frederike E L Otto, Francesco Ragone, Sonia I. Seneviratne, Roop K. Singh, Amalie Skålevåg, Piet Termonia, Lisa Thalheimer, Maarten van Aalst, Joris van den Bergh, Hans van de Vyver, Stéphane Vannitsem, Geert Jan van Oldenborgh, Bert van Schaeybroeck, Robert Vautard, Demi Vonk, and Niko WandersSkålevåg. Rapid attribution of heavy rainfall events leading to the severe flooding in western europe during july 2021. *World Weather attribution*, 2021.

[9] Zbigniew W. Kundzewicz. Extreme weather events and their consequences. *Papers on Global IGBP*, 23:59–69, 2016.

[10] Zbigniew W. Kundzewicz and Pawel Licznard. Climate change adjustments in engineering design standards: European perspective. *Water Policy, Official Journal of the World Water Council*, 23, 2021.

[11] Wayne W. LaMorte. Mann whitney u test (wilcoxon rank sum test). *Nonparametric Tests*, 2017.

[12] M. Leubecher and T.N. Palmer. Ensemble forecasting. *Journal of Computational Physics*, 227:3515–3539, 2008.

[13] V. Masson-Delmotte, P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Watterfield, O. Yelekçi, R. Yu, and B. Zhou. Summary for policymakers. in:climate change 2021: The physical sciencse basis. contribution of working group i to the sixth assessment report of the intergovernmental panel on climate change. *IPCC*, pages 3–32, 2021.

[14] John F.B Mitchell, Jason Lowe, Richard A Wood, and Michael Vellinga. Extreme events due to human-induced climate change. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 364:2117–2133, 2006.

[15] National Institute of Standards and Technology. Anderson-darling test. *Engineering Statistics Handbook*, 2022.

[16] Andrew Revkin. Climate change first became news 30 years ago. why haven't we fixed it? *National Geographic*, 2018.

[17] Rijskwaterstaat. Extreem hoge waterstanden op maas en rijn: uniek voor de zomer. *Rijkswaterstaat*, 2021.

[18] Sonia I. Seneviratne, Neville Nicholls, David Easterling, Clare M. Goodess, Shinijro Kanae, James Kossin, Yali Luo, Jose Marengo, Kathleen McInnes, Mohammed Rahimi, Markus Reichstein, Asgeir Sorteberg, Carolina Vera, and Zuebin Zhang. Changes in climate extremes and their impacts on the natural physical environment. *Managing the Risks of Extreme Events and Disasters o Advance Climate Change Adaption:Special Report of the Intergovernmental Panel on Climate Change*, 9781107025:109–230, 2012.

[19] Statistics Solutions. Conduct and interpret a spearman rank correlation. *Complete Dissertation*, 2022.

[20] Statistics Solutions. Pearson correlation assumptions. *Complete Dissertation*, 2022.

[21] Task Force Fact-finding hoogwater 2021. Hoogwater 2021, feiten en duiding. *expertisenetwerk waterveiligheid*, 2, 2021.

[22] Jan Maurits van Linge, Johan de Putter, Dorine Epping, and Miranda Koffijberg. Samen maken we het aa-dal klimaatrobuust. *Waterschap Aa en Maas*, 2020.

[23] E. Verheijen, A. van Kruchten, A. van Hal, R. Potter, A.H. te Linde, and S.M.G. Arts. Beheerplan waterkeringen 2017-2022. *Waterschap Limburg*, 2:0–49, 2017.

[24] Frédéric Vitart, Gianpaolo Balsamo, Jean-Raymond Bidlot, Simon Lang, Ivan Tsonevsky, David Richardson, and Magdalena Balmaseda. Use of era5 reanalysis to initialise re-forecasts proves beneficiel. *Number 161 - Autumn 2019*, 2019.

[25] Werkgroep Klimaatadaptie Gemeente Deventer. Aanpassen aan klimaatverandering deventer ambitie en aanpak. *Werkgroep Klimaatadaptie Gemeente Deventer*, pages 1–50, 2016.

[26] Joel Zeder and Erich M. Fischer. Observed extreme precipitation trends and scaling in central europe. *Weather and Climate Extremes*, 29, 2020.

[27] Ervin Zsoter, Christel Prudhomme, Elisabeth Stephens, Florian Pappenberger, and Cloke Hannah. Using ensemble reforecasts to generate flood thresholds for improved global flood forecasting. *WILEY and CIWEM*, 2020.

9 Appendix A

9.1 Selecting the extreme events

The steps below describe how the 'extreme event' was selected from a single ensemble member from all re-forecasts.

- Data is obtained from the ECMWF data base. These files are a 5-D matrix which contain Longitude, Latitude, Ensemble member number, Time and the Precipitation.
- The data is a re-forecast over the whole world. However, only the data for the Meuse basin are interesting; therefore, the grid variables are only in the range of the simplified Meuse basin (3.4N-6.6N and 51.8W-48.0W), reducing the overall data and file size.
- The files are grouped in years from the date of creation of the file. This means that it starts with the group 1996, 1997, 1998, and so on.
- Starting with the group of re-forecasts in 1996, the first file (re-forecast) is picked out.
- Next step is to pick-out the first ensemble member from the re-forecast.
- Having one ensemble member, the first step is to calculate the average precipitation over the whole simplified Meuse catchment by taking the mean over all the single grid cells in the re-forecast file. This will result in a vector file that gives the precipitation throughout the area per time step
- The time step is set at 6 hours in the data files. This is changed by summing up four values which results in precipitation for one day. The new vector gives a precipitation value for each day in the ensemble member. This vector is called daily precipitation.
- To pick out the extreme event from the ensemble member, a moving sum is used to calculate the possible extreme events in the vector of daily precipitation.
  - For one-day events, this step is skipped because the values already contain daily precipitation.
  - For three-day events, a three-day moving sum is used on the daily precipitation vector, which calculates the sum of the three consecutive values. This means that the last two values can be neglected because they are the same as or lower than the third last day in the precipitation vector.
  - For five-day events, a moving sum of five is used which calculates the sum of the following five values. This means that the last four values can be neglected because they are the same as or lower than the fifth last day in the daily precipitation vector.
- With the result of the moving sum, a vector is obtained with all possible extreme events. From these events, the maximum value is chosen. About that extreme event useful data for later processing is stored. The different parameters of interest are listed below:
  - year of the re-forecast (between 1996-2015)
  - month of the re-forecast (between 1-12)
  - day of the re-forecast (between 1-31)

- days since 0-0-0000 of the event
  - ensemble member (between 1-10)
  - day in the ensemble number (between 1-15)
  - precipitation in m
- Now the next ensemble member is chosen from the re-forecast and the same steps from... are followed again. When the last ensemble member (number 10) is done the next re-forecast is chosen from the group and the steps from ... are followed again.
- When an extreme event is selected from each ensemble member in the group of that certain year, a single matrix for each extreme event duration (1,3 and 5 days) is the result. The total number of extreme events in those matrices is the number of re-forecasts times ten. The next step is the selection and filtering of the data to finally end with the ten most extreme events of that group (year/season).

9.2 Filtering the extreme events

After selecting the extreme event, the large matrix is sorted and filtered to finally have 10 extreme events for each year/season at the end. The steps of this process are described below. The process is divided into annual extreme events and seasonal extreme events. First, the annual process is described.

- As input for this process, the matrix with the extreme events of one group (one year) is used. Then, this matrix is sorted by the amount of precipitation in descending order, with the highest value being the top row.
- When the table of the ten most extreme events is empty, the highest value of the sorted table with all extreme events is selected. That event will form the start of the list of the ten most extreme events in that group(year).
- The next highest extreme event is picked.
- The first step is the comparison of the date of the incident of the chosen event and the listed events in the list of the ten most extreme events. This is done with the use of the unit of the days since 0-0-0000. If there is a match the chosen event is removed and step ... is repeated. If not continue to the next step.
- The second step is to see if there is at least a 2 day time gap between the chosen event and the existing events in the list of the ten most extreme events. The value that must be between the date of the incident is the duration of the extreme event plus two. That means three days for one-day extremes, five days for three-day extremes, and seven days for five-day extremes. Using that value, the unit of days since 0-0-0000 for the date of the incident can be used. If the difference between events is within the time gap the event is removed and step ... is repeated. If not, the chosen event complies with the rules and becomes the next event in the list of the ten most extreme events. If this list consists of ten events, this process is stopped. If not, step... is repeated.

When the process is stopped, it means that ten events are selected. These ten events are saved in the final list of extreme events between 1996 and 2015. Now the next group (year) is selected, and the first process of selecting the extreme events is started over again. These processes are repeated for all groups(years), resulting in a matrix with ten extremes of all years (1996-2015), which are in total 200 events. However, an aspect of this research is also to find out if there is a difference in summer and winter. For that a different sorting and filtering process is applied.

- As input for this process, the matrix with the extreme events of one group (one year) is used. However, before this is sorting out, the group is divided into three groups. The three different groups are shown below.
  - Winter second half (1.January - 31. March)
  - Summer (1. April - 31. September)
  - Winter first half (1. October - 31. December)
- They are grouped by dates. After they are grouped, they are sorted from highest precipitation to lowest precipitation.
- For the first year (1996), the winter second half is removed, because the data starts in march 1996 which means that there is almost no data. Therefore, the first group that is filtered are the summer events. The filtering functions the same as in the filtering process for the annual year until the ten most extreme events are found, therefor that process can be followed. The only difference is the final matrix, which is called extreme events from 1996 until the 2015 summer.
- After the summer 1996, exactly the same is done with the events in the first half of winter 1996. The outcome of these events are, as described later, compared to the events in the group winter second half from the upcoming year.
- From the second year on (1997-2015) the data is grouped by dates (groups described in step...). Additionally, the ten most extreme events from the first half in the previous year are called back to be added to the group for winter second-half events and sorted from highest precipitation to lowest (descending order). The same filter process used for annual events is used to produce a table with the ten most extreme events in winter. They get an extra column that receives the value of the year where the winter started. The data is stored in the list of extreme events from 1996 until the 2015 winter.
- The summer and winter first half groups use the same steps as the first group (1996) described in step....
- When all the years are complete, there will be a matrix for the winter events and a matrix for the summer events. Both will have 200 extreme events for the time period 1996-2015.

10 Appendix B

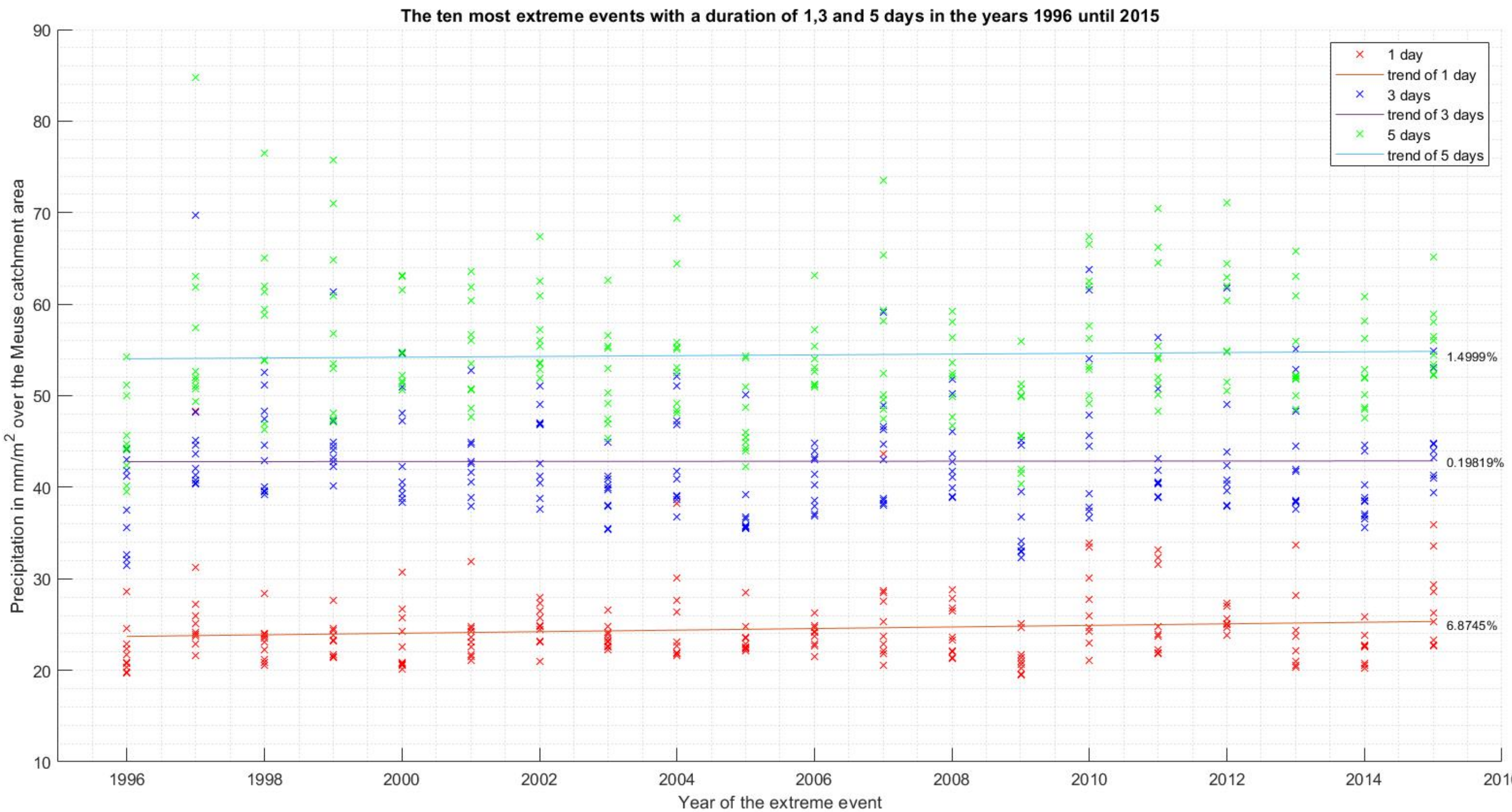


Figure 11: The graph shows a larger representation of the ten annual extreme events between 1996 and 2015 for the one-day, three-day and five-day events inclusive the trend-line determined by least square linear regression.

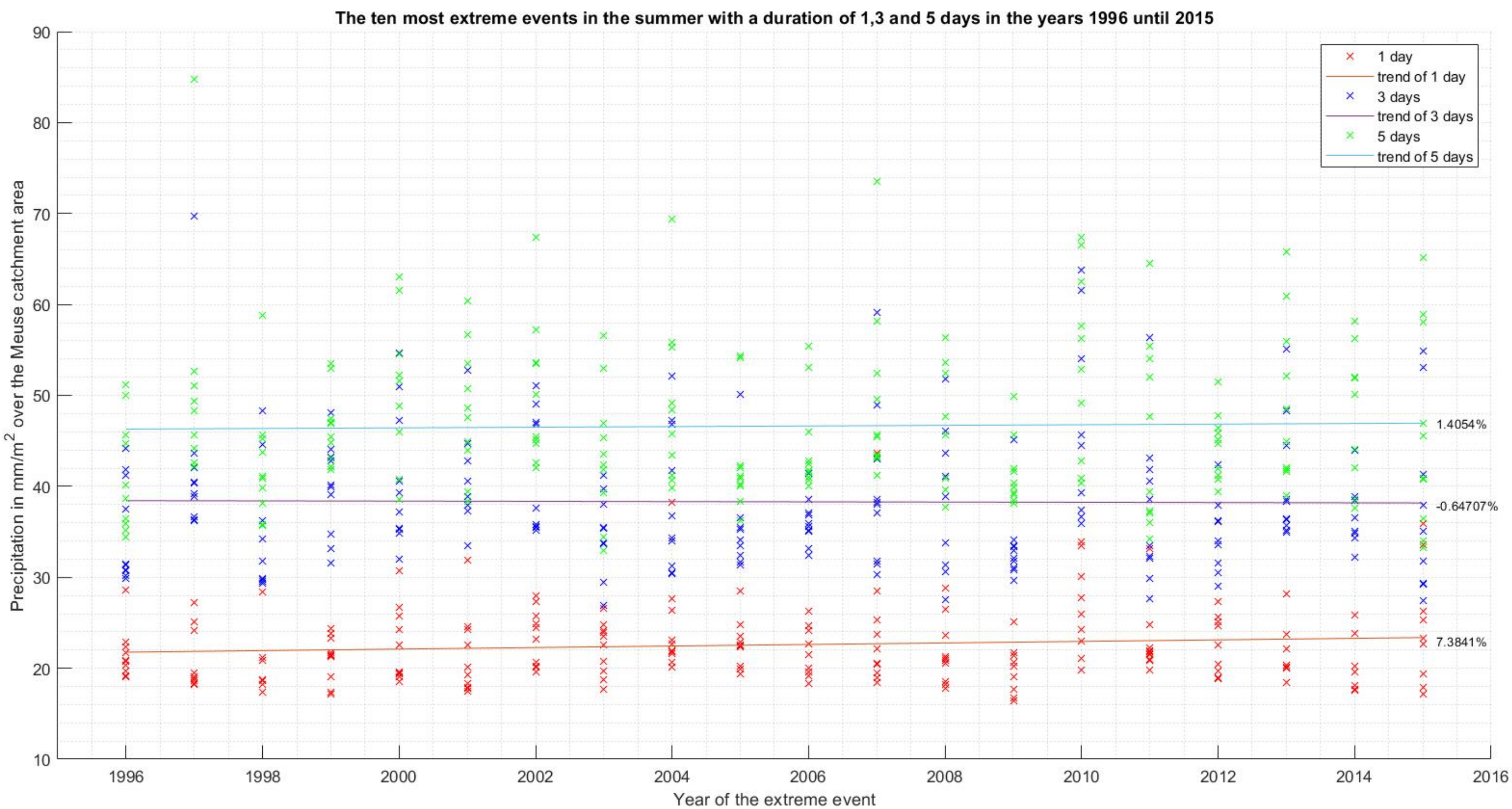


Figure 12: The graph shows a larger representation of the ten annual extreme events between 1996 and 2015 for the one day, three day and five day events inclusive the trend-line determined by least square linear regression for the summer season.

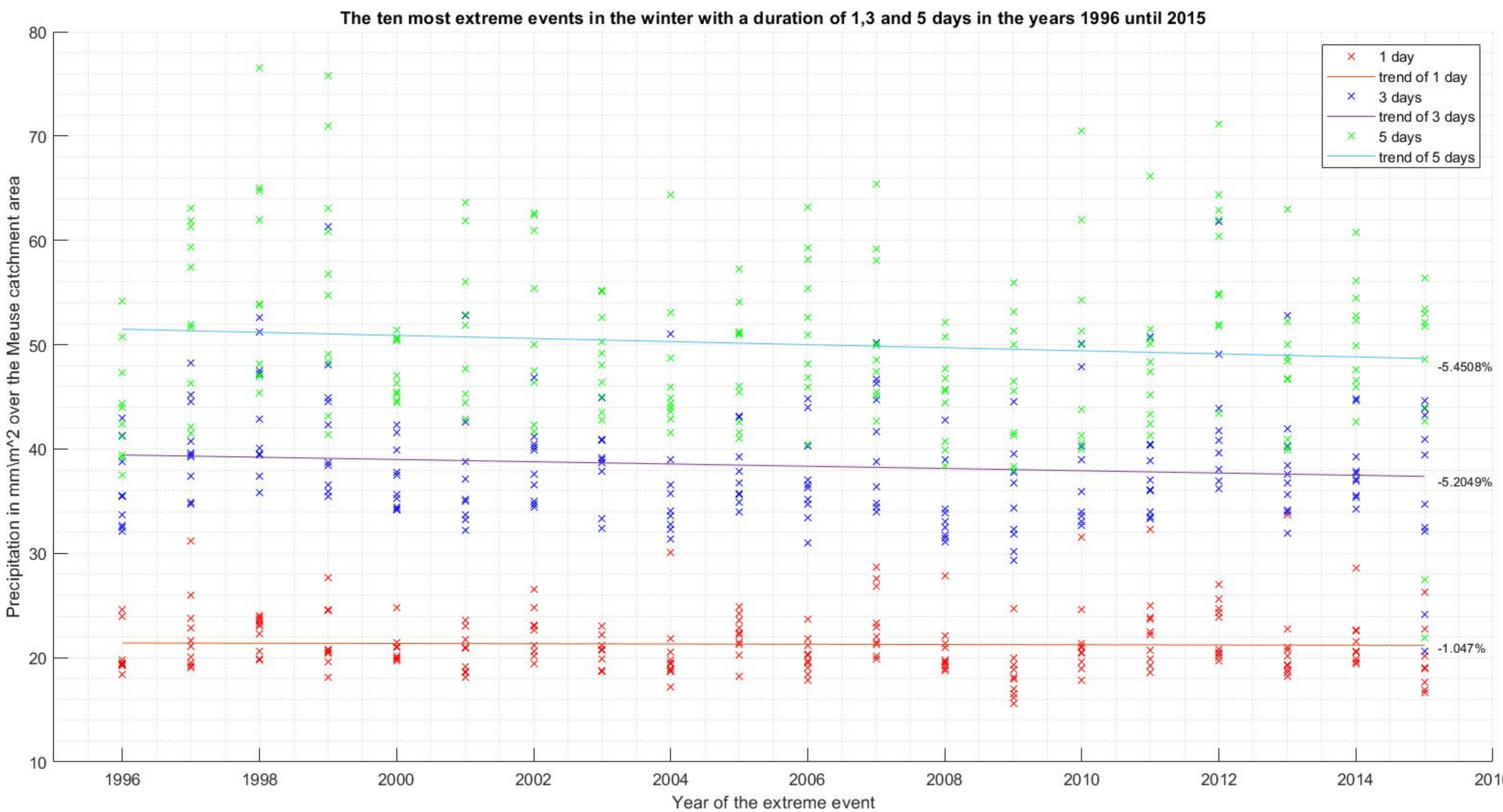


Figure 13: The graph shows a larger representation of the ten annual extreme events between 1996 and 2015 for the one day, three day and five day events inclusive the trend-line determined by least square linear regression for the winter season.

11 Appendix C

In the following section the outcome of the extension of the data is shown of the first method. The first method is extending the data set by only using the year 2015. The year 2015 is added up to five times to the existing data set. The tables below show the p-value obtained by the four statistical tests.

Table 9: The results of the one-day extremes annually with the first method. The values are the p-values obtained by the statistical tests.

Annual one-day event				
Time period	Pearson t-test	Spearman	Mann-Kendall-Wilcoxon	Anderson-Darling
20 years	3.9%	2.9%	2.4%	5.9%
21 years	0.8%	0.4%	0.7%	1.8%
22 years	0.2%	0.1%	0.9%	2.0%
23 years	0.1%	0.1%	0.1%	0.1%
24 years	0.1%	0.1%	0.1%	0.1%
25 years	0.1%	0.1%	0.1%	0.1%

Table 10: The results of the three-day extremes annually with the first method. The values are the p-values obtained by the statistical tests.

Annual three-day event				
Time period	Pearson t-test	Spearman	Mann-Kendall-Wilcoxon	Anderson-Darling
20 years	46.6%	67.9%	74.1%	48.0%
21 years	25.3%	38.4%	56.0%	66.9%
22 years	12.2%	17.1%	30.5%	57.1%
23 years	5.4%	6.4%	14.1%	35.9%
24 years	2.3%	2.1%	0.7%	1.8%
25 years	1.0%	0.7%	0.7%	1.8%

Table 11: The results of the five-day extremes annually with the first method. The values are the p-values obtained by the statistical tests.

Annual five-day event				
Time period	Pearson t-test	Spearman	Mann-Kendall-Wilcoxon	Anderson-Darling
20 years	29.2%	13.4%	12.2%	8.1%
21 years	22.2%	6.7%	10.4%	3.7%
22 years	16.8%	3.3%	12.5%	1.5%
23 years	12.8%	1.6%	8.2%	0.5%
24 years	9.8%	0.8%	2.3%	0.1%
25 years	7.5%	0.4%	2.3%	0.1%

Table 12: The results of the one-day extremes in the summer period with the first method. The values are the p-values obtained by the statistical tests.

Summer one-day event				
Time period	Pearson t-test	Spearman	Mann-Kendall-Wilcoxon	Anderson-Darling
20 years	6.9%	3.9%	3.8%	10.2%
21 years	2.6%	2.0%	4.2%	8.2%
22 years	1.0%	1.0%	4.1%	4.7%
23 years	0.4%	0.5%	2.3%	1.2%
24 years	0.2%	0.3%	0.3%	0.2%
25 years	0.1%	0.2%	0.3%	0.2%

Table 13: The results of the one-day extremes in the winter period with the first method. The values are the p-values obtained by the statistical tests.

Winter one-day event				
Time period	Pearson t-test	Spearman	Mann-Kendall-Wilcoxon	Anderson-Darling
20 years	63.0%	84.1%	68.8%	54.7%
21 years	81.3%	95.4%	94.5%	94.4%
22 years	91.5%	98.9%	99.8%	99.9%
23 years	96.4%	99.7%	99.9%	100%
24 years	98.5%	99.9%	100%	100%
25 years	99.4%	100%	100%	100%

Table 14: The results of the three-day extremes in the summer period with the first method. The values are the p-values obtained by the statistical tests.

Summer three-day event				
Time period	Pearson t-test	Spearman	Mann-Kendall-Wilcoxon	Anderson-Darling
20 years	54.7%	67.7%	73.5%	48.4%
21 years	47.1%	68.1%	71.6%	32.4%
22 years	40.8%	68.5%	58.5%	17.0%
23 years	35.7	68.9%	72.1%	5.0%
24 years	31.5%	69.2%	68.5%	0.8%
25 years	28.2%	69.5%	68.5%	0.8%

Table 15: The results of the three-day extremes in the winter period with the first method. The values are the p-values obtained by the statistical tests.

Winter three-day event				
Time period	Pearson t-test	Spearman	Mann-Kendall-Wilcoxon	Anderson-Darling
20 years	94.2%	92.9%	84.9%	48.0%
21 years	97.5%	94.5%	85.9%	25.8%
22 years	98.9%	95.6%	97.0%	23.2%
23 years	99.5%	96.4%	97.4%	0.6%
24 years	99.7%	97.0%	98.2%	0.1%
25 years	99.8%	97.4%	98.2%	0.1%

Table 16: The results of the five-day extremes in the summer period with the first method. The values are the p-values obtained by the statistical tests.

Summer five-day event				
Time period	Pearson t-test	Spearman	Mann-Kendall-Wilcoxon	Anderson-Darling
20 years	35.9%	46.8%	49.2%	66.1%
21 years	40.2%	54.4%	53.0%	65.3%
22 years	44.0%	60.8%	53.0%	18.7%
23 years	47.2%	66.1%	70.7%	6.8%
24 years	50.0%	70.6%	79.6%	2.4%
25 years	52.5%	74.2%	79.6%	2.4%

Table 17: The results of the five-day extremes in the winter period with the first method. The values are the p-values obtained by the statistical tests.

Winter five-day event				
Time period	Pearson t-test	Spearman	Mann-Kendall-Wilcoxon	Anderson-Darling
20 years	93.3%	77.7%	56.6%	20.5%
21 years	97.9%	82.4%	57.1%	59.3%
22 years	99.3%	85.9%	89.5%	95.2%
23 years	99.8%	88.5%	88.1%	98.0%
24 years	99.9%	90.4%	94.4%	99.5%
25 years	100%	91.9%	94.4%	99.5%

12 Appendix D

In the following section the outcome of the extension of the data is shown of the second method. The second method is extending the data set by adding the period 2011-2015. When the data set is extended by one year only year 2015 is used. When the data set is extended by two years, the year 2014 and 2015 are added. This pattern is followed until five year are added, which is the period 2011-2015. The values in the table below are the p-values obtained by the four statistical tests. Because the added years have different precipitation values, the p-value can go up and down.

Table 18: The results of the one-day extremes annually with the second method. The values are the p-values obtained by the statistical tests.

Annual one-day event				
Time period	Pearson t-test	Spearman	Mann-Kendall-Wilcoxon	Anderson-Darling
20 years	3.9%	2.9%	2.4%	5.9%
21 years	0.8%	0.4%	0.7%	1.8%
22 years	2.4%	3.5%	13.1%	19.3%
23 years	4.1%	7.7%	13.1%	19.3%
24 years	3.2%	1.6%	0.7%	1.8%
25 years	2.1%	1.3%	2.4%	5.9%

Table 19: The results of the three-day extremes annually with the second method. The values are the p-values obtained by the statistical tests.

Annual three-day event				
Time period	Pearson t-test	Spearman	Mann-Kendall-Wilcoxon	Anderson-Darling
20 years	46.7%	67.9%	74.2%	48.0%
21 years	25.3%	38.4%	56.0%	66.9%
22 years	53.4%	72.3%	81.5%	24.7%
23 years	47.3%	67.4%	81.5%	24.7%
24 years	43.9%	65.9%	56.0%	66.9%
25 years	36.3%	56.3%	74.2%	48.0%

Table 20: The results of the five-day extremes annually with the second method. The values are the p-values obtained by the statistical tests.

Annual five-day event				
Time period	Pearson t-test	Spearman	Mann-Kendall-Wilcoxon	Anderson-Darling
20 years	29.2%	13.4%	12.2%	8.1%
21 years	22.2%	6.7%	10.4%	3.7%
22 years	33.4%	13.3%	29.3%	29.3%
23 years	29.7%	12.1%	29.3%	4.5%
24 years	13.6%	4.9%	10.4%	3.7%
25 years	9.1%	3.5%	12.2%	8.1%

Table 21: The results of the one-day extremes in the summer period with the second method. The values are the p-values obtained by the statistical tests.

Summer one-day event				
Time period	Pearson t-test	Spearman	Mann-Kendall-Wilcoxon	Anderson-Darling
20 years	6.9%	3.9%	3.8%	10.2%
21 years	2.6%	2.0%	4.2%	8.2%
22 years	5.5%	8.3%	16.4%	16.1%
23 years	8.1%	9.9%	16.4%	16.1%
24 years	8.6%	7.1%	4.2%	8.2%
25 years	9.0%	5.2%	3.8%	10.2%

Table 22: The results of the one-day extremes in the winter period with the second method. The values are the p-values obtained by the statistical tests.

Winter one-day event				
Time period	Pearson t-test	Spearman	Mann-Kendall-Wilcoxon	Anderson-Darling
20 years	63.0%	84.1%	68.8%	54.7%
21 years	81.3%	95.4%	94.5%	94.4%
22 years	76.8%	92.1%	98.5%	98.2%
23 years	74.6%	95.3%	98.5%	98.2%
24 years	54.3%	80.9%	94.5%	94.4%
25 years	34.1%	64.0%	68.8%	54.7%

Table 23: The results of the three-day extremes in the summer period with the second method. The values are the p-values obtained by the statistical tests.

Summer three-day event				
Time period	Pearson t-test	Spearman	Mann-Kendall-Wilcoxon	Anderson-Darling
20 years	54.7%	67.7%	73.5%	48.4%
21 years	47.1%	68.1%	71.6%	32.4%
22 years	56.5%	70.6%	62.4%	28.2%
23 years	44.3%	54.0%	62.4%	28.2%
24 years	62.5%	69.6%	71.6%	32.4%
25 years	65.2%	72.6%	73.5%	48.4%

Table 24: The results of the three-day extremes in the winter period with the second method. The values are the p-values obtained by the statistical tests.

Winter three-day event				
Time period	Pearson t-test	Spearman	Mann-Kendall-Wilcoxon	Anderson-Darling
20 years	94.2%	92.9%	84.9%	48.0%
21 years	97.5%	94.5%	85.9%	25.8%
22 years	96.7%	91.5%	95.8%	7.4%
23 years	95.7%	91.0%	95.8%	7.4%
24 years	78.8%	71.0%	85.9%	25.8%
25 years	79.7%	73.5%	84.9%	48.0%

Table 25: The results of the five-day extremes in the summer period with the second method. The values are the p-values obtained by the statistical tests.

Summer five-day event				
Time period	Pearson t-test	Spearman	Mann-Kendall-Wilcoxon	Anderson-Darling
20 years	35.9%	46.8%	49.2%	66.1%
21 years	40.2%	54.4%	53.0%	65.3%
22 years	39.0%	50.8%	41.8%	32.1%
23 years	27.2%	38.3%	41.8%	32.1%
24 years	38.6%	45.6%	53.0%	65.3%
25 years	43.6%	53.5%	49.2%	66.1%

Table 26: The results of the five-day extremes in the winter period with the second method. The values are the p-values obtained by the statistical tests.

Winter five-day event				
Time period	Pearson t-test	Spearman	Mann-Kendall-Wilcoxon	Anderson-Darling
20 years	93.3%	77.7%	56.6%	79.5%
21 years	97.9%	82.4%	57.1%	40.7%
22 years	96.2%	73.0	80.2%	11.0%
23 years	97.5%	81.8%	80.2%	11.0%
24 years	78.4%	44.9%	57.1%	40.7%
25 years	82.5%	55.2%	56.6%	79.5%