Interdependency of Stochastic Variables Determining Normative Water Levels in the Alblasserwaard

Master Thesis
W. Roosjen, October 2019

UNIVERSITY OF TWENTE.

HydroLogic

Interdependency of Stochastic Variables Determining Normative Water Levels in the Alblasserwaard

Master Thesis Civil Engineering and Management
University of Twente
Faculty of Engineering Technology
Water Engineering and Management

Author	Wytse Roosjen BSc (s1574981)
	w.roosjen@alumnus.utwente.nl
Educational institution	University of Twente
External organisation	HydroLogic
UT supervisors	Dr.ir. M.J. Booij (Head of graduation committee)
	Ir. M.R.A. Gensen
External supervisor	Ir. B. Schnitzler
Date	31-10-2019

Preface

This master thesis marks the end of my period as a student Civil Engineering at the University of Twente, in which I learned a lot and had a great time. I would like to thank everyone involved in my time as a student and during this research project. This research project introduced me to the statistics involved in determining normative water levels in a regional water system, from which I learned a lot of new things.

Many thanks to HydroLogic for giving me the opportunity to execute this research project as an intern in Amersfoort. I have experienced this as great time in which I gained a lot of experience. Special thanks to all the supportive colleagues who created a great working atmosphere to work in. Furthermore, I would like to thank my external supervisor at HydroLogic, Bram Schnitzler, who helped me shaping the research project and provided useful feedback to improve my thesis.

Also, many thanks to my supervisors from the University of Twente, Martijn Booij and Matthijs Gensen. They helped me a lot in defining the research objective and setting up the research project. Their critical views and good discussions helped me to improve this master thesis. Also, I would like to thank Paul Torfs who helped me with this project and provided useful feedback with his experience and knowledge of the topic. Finally, many thanks to Froukje Molenkamp in supporting me during this period and also providing feedback on my writings.

Summary

The Netherlands has always been vulnerable to floods due to its low elevation. To reduce the probability of floods in regional water systems to an acceptable level of safety, norms have been introduced. These norms are based on the probability of occurrence of these floods. The regional water systems are generally assessed on the level of safety using a hydraulic model of the water system to determine the frequency of occurrence of certain water levels for each location, the so-called normative water levels. This could be used to determine the required top level of regional flood defences. There are currently two main methods which are used to determine these normative water levels, the time series method and the stochastic method. The first method uses measured time series of for example precipitation and wind as input for the model. Then, the frequencies of occurrence of water levels are determined based on the resulting water levels in the model. As these measured time series are generally not sufficient in length for the return periods of the norms (e.g. once in 100 years), extrapolation based on the available observations takes place to obtain the water levels for longer return periods. The stochastic method extrapolates on the input variables or also-called stochastic variables (e.g. precipitation and wind) to generate model events with a certain probability. An important advantage of this second method is that extreme situations are tested in the water system. However, a disadvantage compared to the time series method is that it is often assumed that these input variables are independent of each other in order to avoid (complex) multivariate probability models. This assumption potentially affects the probabilities of occurrence of certain situations (e.g. large precipitation volumes with high wind speeds and high external water levels), as interdependencies might be present in reality.

To assess whether there are any interdependencies between these input variables, and what the effects of these interdependencies on the joint probabilities and normative water levels are, a case study on the Alblasserwaard is performed (location in Figure 1). To determine the joint probabilities of the input events, the method 'copula' is used. This method is used to construct joint probability distributions based on the dependence structure between the variables and the probability distributions of each variable. Important advantages over other methods are that each probability distribution can take any form and is independently chosen of the dependence structure, while extensions to higher dimensions are also possible using the theory of vine copula. Summing up, the research goal of this study is:

"*To evaluate to what extent including interdependence of stochastic variables affects normative water levels in comparison to assuming independence between these variables, by using a copula approach in a case study of the Alblasserwaard*"

First, the relevant stochastic variables are tested on the presence of interdependencies. The strongest dependency is found between the variables wind speed and water level of the Lek (affecting the pumping station at Kinderdijk). Also, low dependency is found between precipitation and wind speed and precipitation and water level of the Lek.

Secondly, joint probability distributions are fitted in four combination studies using copula functions, three combinations of two variables and the last combination with three variables

using a vine copula. For all these combinations, appropriate probability distributions and copula functions are found to represent the observations sufficiently. In comparison to the independent situations, the dependent situations are relatively better in representing the observations, especially visible for the combination of wind speed and water level Lek. These simulations also show how the probability of joint extremes is generally higher in the dependent situation.

Lastly, the joint probability distributions are used to determine the normative water levels in the Alblasserwaard and compared to the independent reference situation. In all combination studies, differences are found. The results are highly dependent on location and return period. Figure 1 shows an example of the obtained results, which is the result of combination 4 for a return period of 100 years, involving all the relevant interdependencies. Large differences up to 10 to 12 cm are visible close to the pumping station as a result of those interdependencies. Differences up to 6 cm are visible in the east of the Alblasserwaard area as a result of the dependency between precipitation and wind.

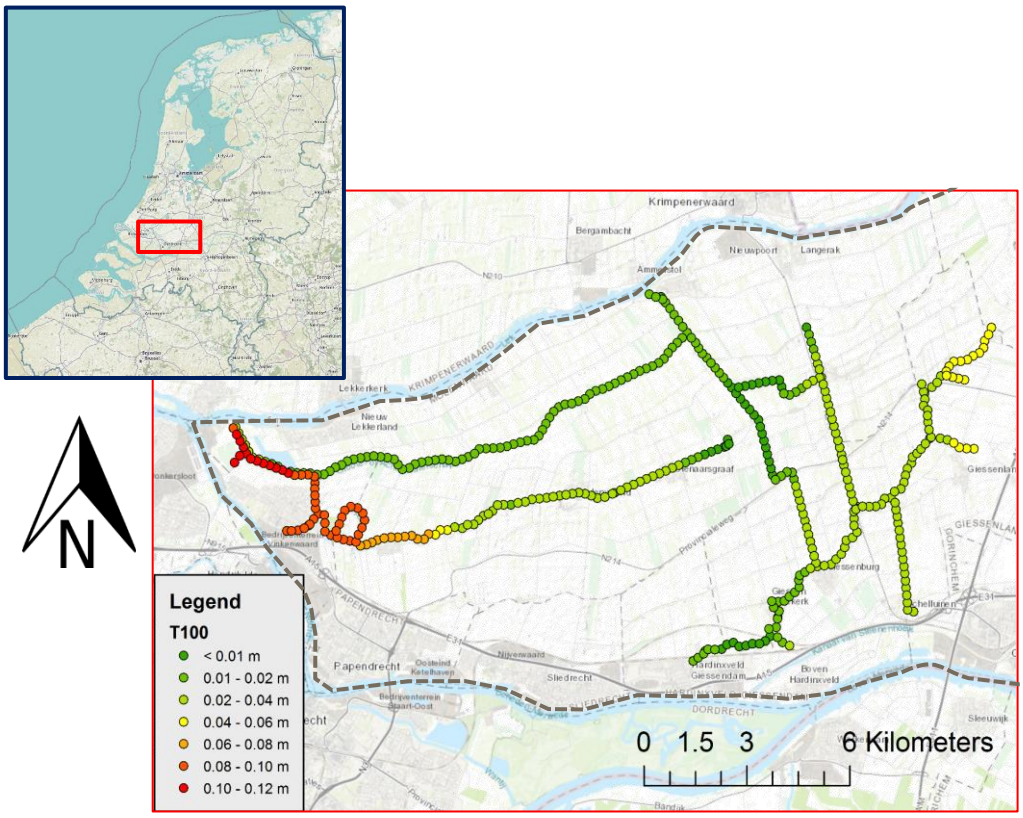


Figure 1: Differences in normative water levels between the dependent and independent situation in the Alblasserwaard area considering all the interdependencies with a return period of once in 100 years (T100). A positive difference means that the modelled dependent situation has higher normative water levels than the independent case.

This study shows that it is technically feasible to account for interdependencies determining normative water levels in a regional water system and potentially other water systems. It shows that it has an effect on the normative water levels as generally the probabilities of joint extremes are higher in the dependent situation, although the effects on the water levels are location and return period specific. When the stochastic method is applied in water management, independence between stochastic variables cannot simply be assumed but should be considered first as the impact could potentially be large as shown in this study.

Contents

1 Introduction

1.1 Background

The Netherlands is susceptible to flooding due to its low elevation. To cope with such risks in regional water systems, norms have been introduced. The norms for regional flood defences are based on the frequency that water levels exceed a certain level, with guidelines defined in 1999 (Interprovinciaal Overleg, 1999). The accepted frequencies of occurrence, ranging from once in 10 years to once in 1000 years, are based on the potential damage of a dike breach. Contrary to the primary flood defences, there is no national regulation for the assessment of regional flood defences, because this is defined by the provinces (STOWA, 2015a). Figure 2 illustrates a typical Dutch regional water system, on which this study focusses. Precipitation in the polders is drained to ditches. Water nuisance may arise in these polders when the drainage capacity is not sufficient (with specific norms in 'Nationaal Bestuursakkoord Water' (Unie van Waterschappen, 2003)). From the ditches in the polders, water is discharged to drainage canals or in Dutch 'boezems', with the help of polder pumping stations. The regional flood defences are generally located around these drainage canals. From these drainage canals, water is discharged to a primary water body (river/sea) or regional river using a pumping station.

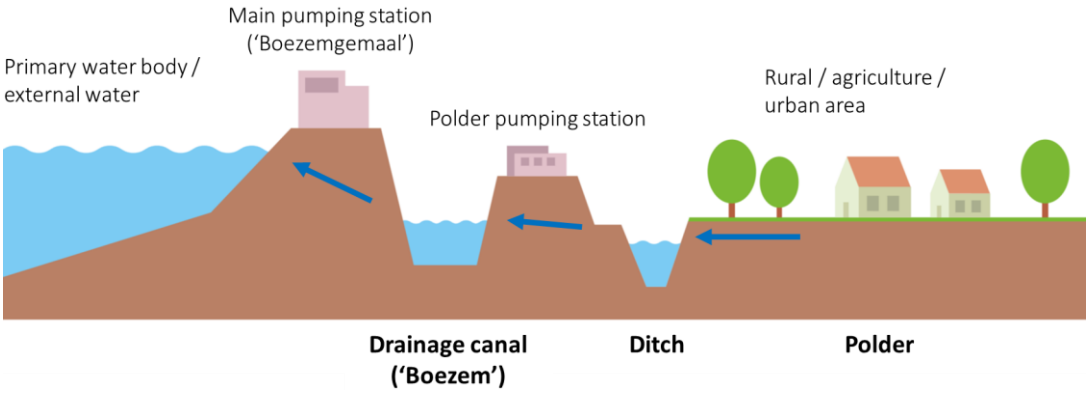


Figure 2: Typical Dutch water system from polder to primary water body (illustation based on: Unie van waterschappen, Interprovinciaal Overleg & STOWA (2016))

To assess the level of safety in regional water systems, normative water levels are determined for all relevant sections of the system to check whether flood defences need to be raised or other measures should be taken. A normative water level is a water level that has a certain frequency of occurrence, for example a water level of +1 m NAP is reached once in 100 years. Currently, there are two main methods that are used to determine normative water levels in regional water systems (STOWA, 2011; Bosch et al., 2006; De Graaff & Versteeg, 2000). The first one is the time series method, the second is the stochastic method. The time series method applies long measured time series (of input variables such as precipitation and wind) in a hydraulic/hydrological model and extrapolates the resulting water levels to longer return periods as the time series are generally not of sufficient length for the required norms (STOWA, 2011). The stochastic method extrapolates to longer return

periods by assigning probabilities to the input variables or the so-called stochastic variables (e.g. precipitation, wind). A stochastic variable is a variable whose values depend on the outcome of a random process, which has a certain probability distribution. Using the second method, the extreme situations are simulated in the water system, while the extreme water levels are extrapolated in the case of the time series method. So, the first method extrapolates on the output of the model (a posteriori), while the second method extrapolates on the input (a priori).

Another option is to measure water levels in the water system and extrapolate (if necessary) to longer return periods, which removes the model uncertainty from the assessment. However, it is then required to have long time series of measurements throughout the complete area for all sections, which is often not available. Furthermore, regional water systems and the stochastic variables are subject to of alterations over time, which invalidate the measurements. In addition, an advantage of having a model of the water system is that it is possible to test the effect of measures on the normative water levels.

This study focusses on the stochastic method for regional water systems. An important assumption that is generally used for the stochastic method as currently performed in practice is that the input variables are independent of each other, which means that the value of one variable does not affect the probability of occurrence of the other. This is generally assumed in order to avoid the need of (complex) multivariate probability models. However, this assumption is quite questionable (STOWA, 2011; Bosch et al., 2006). For example, you might expect that during an extreme precipitation event, the wind speeds are higher. Another example is that the external water level affecting the main pumping stations are affected by precipitation and/or wind, which could lead to potential problems when the regional water system needs discharge capacity from these affected pumping station. It is therefore relevant to study such dependencies as they could possibly affect the normative water levels when the stochastic method is used.

1.2 Current knowledge

To cope with dependencies between variables, several methods are found in literature. A traditional approach to measure dependency is using a coefficient such as Pearson's linear correlation coefficient or Spearman's *rho* (Hao & Singh, 2016; Genest & Favre, 2007). However, to determine the joint probabilities of dependent variables to fall in a particular range of values, with possibly complex underlying dependence structures, such a coefficient has little use. For this purpose, joint probability distributions have to be constructed, in which individual (marginal) probability distributions and their dependencies are combined. From these distributions, the frequency of occurrence of events with multiple variables to fall in a particular range of values can be determined, which can be used to determine normative water levels.

For situations with multiple variables, such as floods with volume, duration and peak discharge, an approach where the frequencies of occurrence of those variables are considered separately could lead to under- or overestimation of the risk (Hao & Singh, 2016; Gräler et al., 2013; Favre et al., 2004). Univariate frequency analyses have been used

extensively in hydrology, while multivariate frequency analyses are emerging. To construct joint probability distributions, extensions of existing parametric distributions have been made, such as the multivariate normal distribution using a linear correlation coefficient to describe its dependency. However, such distributions have several drawbacks. The same family for the distribution of each variable is required (e.g. normally distributed), the parameters for the individual distributions are also used to model the dependency and extensions to more than two variables are often not feasible (Favre et al., 2004). In addition, in the case of the multivariate normal distribution, the dependency is assumed to be linear. This limits the applicability when several hydrological and meteorological processess are considered where the individual distributions of variables are not always from the same family and the dependence structure can be complex and non-linear. A solution could be to use nonparamatric methods such as a kernel densities, where a probability density is smoothely fitted to the data. However, an important drawback of nonparametric methods is that the extrapolation beyond the historical data records may cause problems as the density is precisely fitted to the existing data (Hao & Singh, 2016). Problems with kernel densities especially occur in multivariate cases. However, it has shown potential in some studies in flood frequency analyses (e.g. Vittal et al. (2015)).

To overcome the drawbacks of the abovementioned methods, copulas have been introduced by Sklar (1959). Copulas are able to model the dependence structure independently from the individual distributions, which can take any form (Nelsen, 2006; Genest & Favre, 2007; Favre et al., 2004). In addition, extension to more than two dimensions is relatively easy using vine copulas, where bivariate copulas are used as building blocks (Aas et al., 2009). There is a wide variety of copula families available to model different kinds of dependence structures (e.g. symmetric, asymmetric, positive, negative and tail dependence). The use of copulas in hydrology is only recent with first mentions around 2004-2007 (Favre et al., 2004; Genest & Favre, 2007). Copulas have first been widely applied in the financial sector (Favre et al., 2004; Genest & Favre, 2007; Wahl et al., 2012).

A reasonable amount of literature can be found for the bivariate case, but it is quite scarce for multivariate cases (3 or more variables). For example, Zhong et al. (2013) used copulas to simulate simultaneous Rhine and Meuse discharges to estimate high water level frequencies in the Lower Rhine Delta with a 1D hydrodynamic model. Gräler et al. (2013) discussed design events and constructed a 3-D vine copula for the variables peak discharge, duration and volume and found differences in design events when this analysis is based on one, two or three variables. Ward et al. (2018) found, using copulas, that the joint exceedance probability of events in which design river discharges and design sea-levels in global deltas and estuaries are considered as dependent can be several magnitudes higher. Similar results are found in Wahl et al. (2015) for the US and Bevacqua et al. (2017) for Italy as a result of low-pressure storm systems and other meteorological processes bringing winds and precipitation simultaneously. Furthermore, copulas are used to simulate long time series, for example of temperature, precipitation and evaporation which could be used as input for a hydrological model (Pham et al., 2018). Among others, these studies show that copulas have great potential, and overcome limitations of other methods, to be used in water engineering and management but it is currently not well-established in the field.

1.3 Research gap

The stochastic method in determining normative water levels in regional water systems is commonly used in practice. An important assumption that is regularly used to apply this method, is that the stochastic variables are independent from each other (STOWA, 2011; Bosch et al., 2006). This is a convenient assumption as the probabilities of the several variables can be multiplied to determine their joint probabilities. However, this assumption might have a significant impact on the normative water levels. An analysis on the impact of this assumption on normative water levels is not found in literature. There are some examples of determining exceedance frequencies of water levels by using a joint probability approach (e.g. Zhong et al. (2013), Franken et al. (2016)), but these are not focussed on a regional water system. In addition, the comparison between normative water levels based on including and based on excluding dependencies between the variables has not been made. It is therefore not clear if this assumption has a large impact on normative water levels. This raises the question whether there is any dependency between the stochastic variables, and if so, what the effect of including dependency on the joint probabilities and normative water levels is, compared to assuming no dependency between these input variables. This also involves the technical applicability of the available theory, including the case with more than two variables (which is relatively new).

1.4 Research objective and questions

The objective of this research is based on the discussed research gap. The sensitivity of the normative water levels to the dependency of the stochastic variables is studied in a case study, which is the Alblasserwaard. This area is used as a hydrological/hydraulic model of this regional water system is available and extensively applied in practice using the stochastic method with assumed independence between the variables, to determine the normative water levels in the area. Afterwards, it is discussed if this study is generalisable and applicable to other areas based on this case study of one area. Based on the current knowledge, a copula approach is taken, as this method overcomes drawbacks of other methods. The normative water levels determined using the copula approach are compared to the normative water levels based on the assumption of independence to study the effect of this assumption. Concluding, the objective of this research is:

"To evaluate to what extent including interdependence of stochastic variables affects normative water levels in comparison to assuming independence between these variables, by using a copula approach in a case study of the Alblasserwaard"

The following research questions are answered to reach this objective:

RQ1. Are the stochastic variables used in determining normative water levels interdependent?

RQ2. What are the most suitable joint probability distributions describing each of the combinations of interdependent stochastic variables, and what is the effect of including dependency on the joint probabilities of the stochastic variables?

RQ3. What is the effect of including dependencies between the stochastic variables on the normative water levels in comparison to assuming independence between these variables and how can these differences be explained?

An overview of the approach to answer the research questions and an outline of the report is given in section 1.5.

1.5 Outline report and overview of methods

A short overview of the applied methods to answer the research questions and achieve the research objective is given here. An overview of this process is visible in Figure 4. Each of those processes are discussed in detail in the mentioned chapters.

An introduction to the study area and the availability of data is given in Chapter 2. In the third chapter, the dependencies in the observations are analysed to answer RQ1. After the exploration of dependencies in Chapter 3, the process is divided in combinations of dependent variables to stepwise study the effect of the dependencies between the variables. For example, a combination could be that the precipitation volume and water level at the Lek are considered dependent, and the other variables are considered independent. This is done to prevent missing out the individual effects of the dependencies between certain variables on the resulting normative water levels, which helps in understanding the physical basis of the results. In addition, this could be useful to state something about other areas (not in the Alblasserwaard) where only some of the variables are relevant. In Chapter 4, the methods to construct the joint probability distributions and its results are described, separately for the bivariate combinations and the final multivariate combination, which answers RQ2. In Chapter 5, the methods used to compute the normative water levels in the Alblasserwaard based on the derived probabilities from the joint probability distributions and the set-up of the model are discussed. The results in this chapter, answering RQ3, are discussed per combination of dependent variables. In each chapter, a conclusion discussing the relevant research question is given. The overarching discussion and conclusion can be found in Chapter 6 and Chapter 7 respectively.

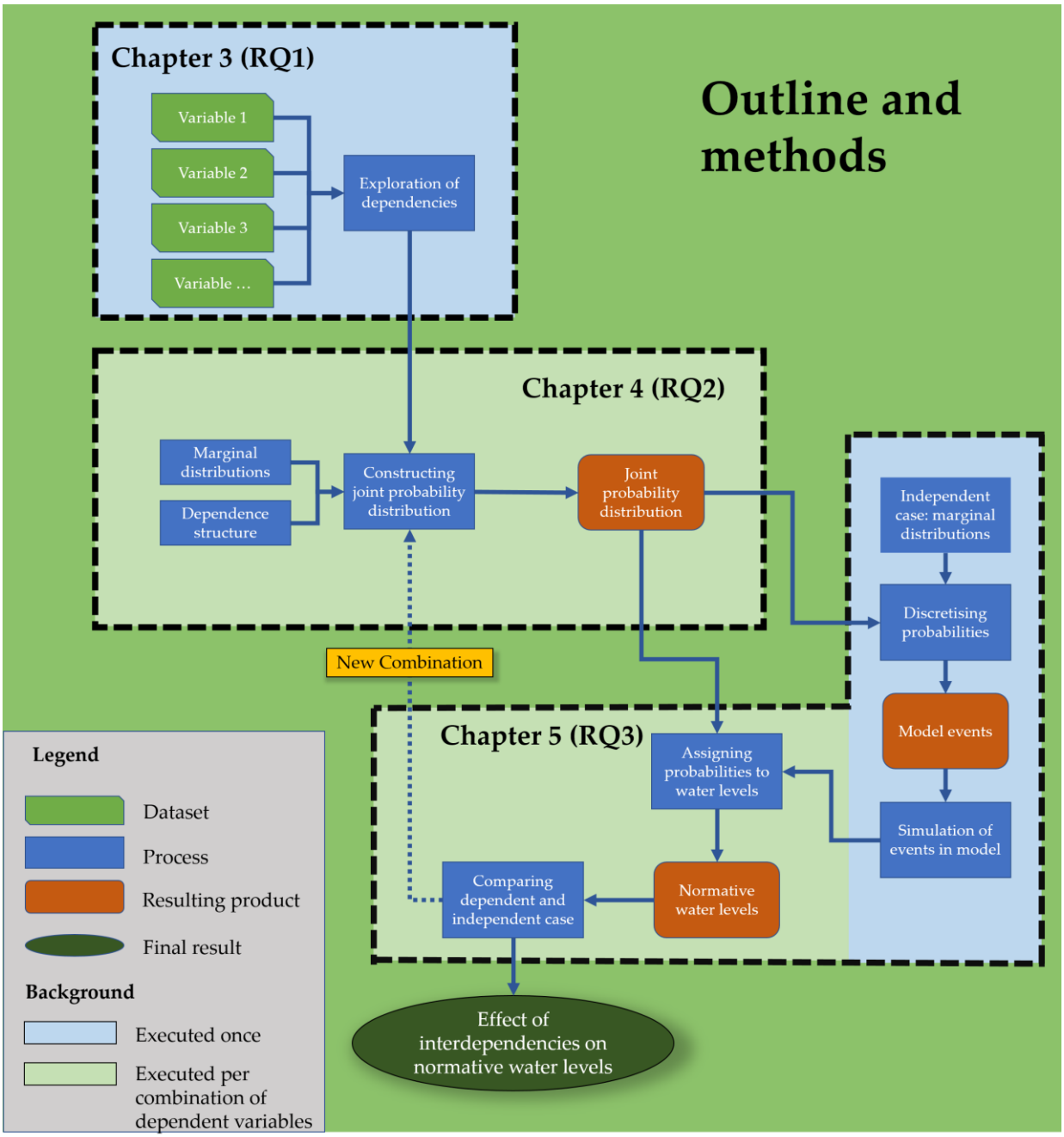


Figure 4: Overview of the methods and structure of the report

2 Study area and data

In this chapter, the case study area and the associated stochastic variables that are used in this research are discussed. Furthermore, the availability and quality of the data for this area is examined.

2.1 Case study area

The case study area, the Alblasserwaard, is visible in Figure 5. A more detailed map of the water system with fixed winter and summer water levels can be found in Appendix A.

Figure 5: Location Alblasserwaard and its two separated water systems: Overwaard and Nederwaard (source base map: OpenStreetMap (2019))

The water system consists of two separated areas, the Overwaard and the Nederwaard. The general flow of water in both the areas is in the direction of the main pumping station complex at the location of Kinderdijk (blue dot, top left in the area). The soil type in the Alblasserwaard area is mainly peat and clay on peat, which caused subsidence in the area in the past, and will continue in the future (HydroLogic, 2017). The system is divided in polder areas (NL: peilgebieden), as visible in Appendix A. Each area has a target water level, which can differ in the summer and winter periods. These levels are managed by the operation of small pumping stations (NL: poldergemalen). These pumping stations discharge the water into the larger drainage canals, which lead the water to the reservoir

areas (NL: boezem). From this reservoir area at Kinderdijk, the water from the two separated water systems is discharged to the Lek, a large river north of the Alblasserwaard.

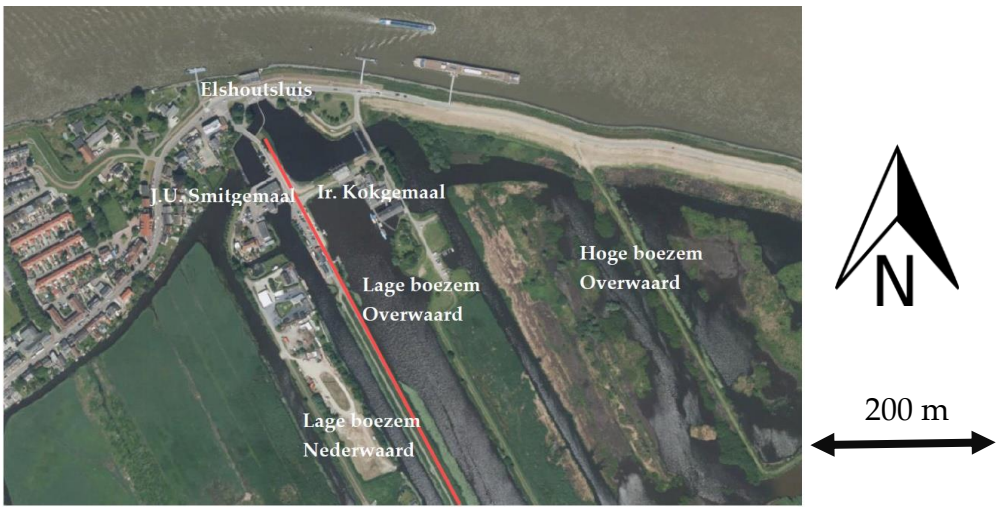


Figure 6: Overview Kinderdijk complex (HydroLogic, 2018a). The red line indicates the separation of the two water systems, the Overwaard and Nederwaard.

An overview of the area where the water is discharged into the Lek is given in Figure 6. The Nederwaard area is discharged via the J.U. Smitgemaal pumping station (max. capacity of 18.3 m^3/s) into a reservoir area with a higher elevation (NL: maalkom). From this reservoir, water is released into the Lek via the 'Elshoutsluis' sluice. This is only possible if the water level at the Lek is less than 2.25 m NAP, otherwise no water is discharged from the Nederwaard area. The Overwaard area is discharged with the Ir. Kokgemaal pumping station (max. capacity 25 m^3/s) to a reservoir area. When the water level at the Lek is higher than 1.35 m NAP, the water from the Overwaard area is released via the 'Elshoutsluis' sluice with a temporary pump (same capacity as pumping station in this study) in the section of the sluice for the Overwaard area. So, the Overwaard area is not affected by the water levels of the Lek. As you can imagine, the combination of a high-water level at the Lek and a large precipitation event in the Nederwaard area is hazardous, as the water accumulates and cannot be discharged from the area.

2.2 Data

To determine the normative water levels in this specific area, using the stochastic method and available model, data is required to assign probabilities to the variables which are relevant. The variables which are expected to have an effect on the normative water levels in the Alblasserwaard are (HydroLogic, 2018b):

- Precipitation
  - Volume
  - Distribution over time
- Wind
  - Speed
  - Direction
- Initial groundwater levels
- External water level Lek (blocking the main pumping station in the Nederwaard)

For precipitation volume, a duration of 9 days is used as normative in the Alblasserwaard (HydroLogic, 2018b). In general, the decision on the normative duration is based on the characteristics of the water system. For urban areas, it is expected that short intense precipitation events are normative, while in rural areas precipitation events of 4 to 9 days are commonly seen as normative depending on its size and characteristics (STOWA, 2011). The decision to use one duration for precipitation events as normative is questionable. However, using multiple durations of precipitation events as stochastic variable is a difficult issue when the stochastic method is applied due to the derivation of probabilities (overlap), which is discussed in Bosch et al. (2006). Therefore, in general only one duration is used to determine the frequencies of occurrence of the total volume of precipitation events. For the distribution of this precipitation volume over time, 7 'shapes' for the duration of 9 days are defined by STOWA (2004). This is an important variable to consider as different distributions of precipitation over time can lead to different effects in a water system. These 7 'shapes' that are used in this study can be found in Appendix G.

The variables related to wind (direction and speed) are important to include in the stochastic method as this affects the gradient of the water surface in the Alblasserwaard (HydroLogic, 2018b; Arcadis, 2014). From previous studies it is concluded that the Alblasserwaard is particularly susceptible for winds from northwest to southwest with velocities higher than 10 m/s, which is a result of the general flow from east to west in the area. The duration of the wind speeds could also be accounted for in a stochastic matter, but this would lead to too many variables to be handled in a manageable amount of computational time. Therefore, this variable is not handled as a stochastic variable, but typical durations are derived in Appendix I.

The initial groundwater level is another important hydrological variable, it affects the response of the water system to a precipitation event. If the groundwater levels are low, the soils can first retain a part of the precipitation volume (depending on soil type) before it is discharged. The Alblasserwaard is a polder area with mainly peat and clay on peat as soil characteristic, see Figure 7. To prevent subsidence, the polders are only drained by a certain degree with little variance over the year (HydroLogic, 2018b). Peat areas are restricted to a maximum difference of 60 cm between the surface and the groundwater level (Waterschap Rivierenland, 2010). The amount of available storage in the upper layer is therefore small. From previous studies in the Alblasserwaard is concluded that the effect of using different initial groundwater levels as input for the stochastic method on the normative water levels in the canals/reservoirs is negligible (< 2 cm) (HydroLogic, 2018b; Arcadis, 2014).

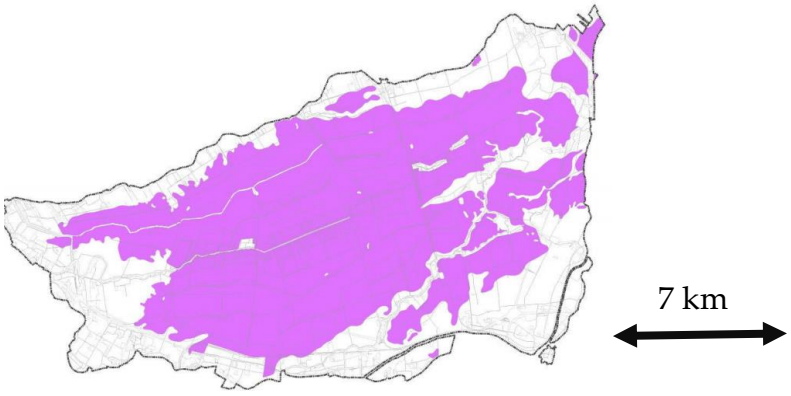


Figure 7: Purple areas represents the peat areas in the Alblasserwaard (Waterschap Rivierenland, 2010)

Table 1 shows the data sets which are used for the discussed stochastic variables of the Alblasserwaard. It is important that the measuring stations are close to the Alblasserwaard (blue outline in Figure 8) as the dependencies between the variables are studied. The end of each measuring period is set at 31-12-2018 to have year-round data and prevent seasonal effects (equal number of each month). The measuring station of Oud-Alblas only includes the daily precipitation volume, while Rotterdam airport also includes wind direction and speed. Therefore, the dataset of Rotterdam airport is used for precipitation and wind data in this research. The dataset of Oud-Alblas is used to validate whether the precipitation volume in Rotterdam is representative for the situation in the Alblasserwaard. Other weather stations close to the Alblasserwaard are present, but these are not further considered as the measured time period is shorter and/or wind is not measured. For example, in Cabauw (close to the Alblasserwaard), wind is measured since 1986.

Table 1: Data availability of stochastic variables

Stochastic variable	Location	Time period	Temporal scale	Resol-ution	Comments	Source
Precipitation volume	Rotterdam airport	1974 – 2018	hourly	0.1 mm	Value of -1 when hourly precipitation volume is < 0.05 mm.	(KNMI, 2019)
	Oud-Alblas	1951 - 2018	daily	0.1 mm	From 08:00 to 08:00	
Wind direction	Rotterdam airport	1971 - 2018	hourly	10°	Wind direction is the average over the last 10 minutes of the hour. Value of 990 when direction varies considerably.	(KNMI, 2019)
Wind speed	Rotterdam airport	1971 - 2018	hourly	1 m/s (0.5 before 1996)	Hourly average	(KNMI, 2019)
External water level (Lek)	Krimpen a/d Lek (located near pumping station Alblasserwaard)	1973 - 2018	hourly (timestep of 10 min. since 10-08-1987)	0.01 m	Longer time series available, but large changes in the water system before 1973 makes it unusable (HydroLogic, 2018b)	(Rijkswaterstaat, 2019)

Figure 8: Location of measuring stations (source map: OpenStreetMap (2019)). Blue outline shows the Alblasserwaard

The overlapping time period of all datasets is from 01-01-1974 until 31-12-2018. This time period is used in further analyses as an overlapping time period is required in the dependency analysis. The total length of the time series is 45 years, which is less than the generally higher required return periods (e.g. T100, T300, T1000), which asks for the need of extrapolation beyond the measured range. The time series of all datasets are checked for this period and no missing values were found. Wind directions with a value of 990 are removed (rotating wind direction), which are only 15 values with low wind speeds. This should not affect the analysis significantly. Furthermore, it is assumed that the observations are stationary and measured correctly. No notable outliers are found. Using the Mann-Kendall non-parametric test for trend analysis, a slight upward trend is found for precipitation volume and a slight downward trend for wind speed (5% level of significance) (Pandey et al., 2018).

The datasets of Rotterdam airport and Oud-Alblas are compared on precipitation volume to validate further usage of the Rotterdam dataset. The figures used for this analysis can be found in Appendix B (Figure 42, Figure 43 and Figure 44). The time series are aggregated over 9 days, as this duration is used in further analyses. The linear correlation coefficient between the two time series is 0.90. The mean values of 9-day periods are respectively 18.8 mm and 21.4 mm for Rotterdam airport and Oud-Alblas. The maximum values are respectively 166.5 mm and 141.2 mm. The Rotterdam dataset shows slightly more extreme values, while the average of the Oud-Alblas dataset is higher. The two time series correspond well enough to use the Rotterdam dataset in this research as the differences in frequencies are small and the Rotterdam dataset also includes wind data.

3 Exploration of dependencies

3.1 Introduction

The goal of this first step is to explore the dependencies between the variables considered in this study which answers research question 1: "*Are the stochastic variables used in determining normative water levels interdependent?*". A broad definition of dependence is used in this study, which means that any relation between two or more variables is considered, not only the linear relationship. If the occurrence of one is affected by the probability of occurrence of the other, it is considered as dependent. If the probability of occurrence of one is not affected by the occurrence of the other, it is considered as independent. This first analysis helps in understanding the interaction or dependence between the variables. Furthermore, it is used as guideline for the next steps as the strongest dependencies are investigated first, plus it provides information for the model setup. First, the methods are discussed in section 3.2, followed with the results in section 3.3. In section 3.4, the conclusion answering the first research question is given.

3.2 Methods

The methods section is divided in two parts, one (3.2.1) based on daily observations and one based on events. Initially, the dependencies between all daily observations are analysed, neglecting the serial dependency between the observations. This step prevents that dependencies could possibly be missed if only data based on events are selected. However, serial correlation (autocorrelation) can affect the distribution of the variables. For example, if the water levels at the Lek are high today, it is expected that the water levels of tomorrow are also high. This effect can be reduced by selecting relevant events from the time series by taking the serial correlation into account, the event-based approach (3.2.2).

3.2.1 Analysis based on daily aggregation

The scale of days, and not hours, is used in this section as the aggregation scale for precipitation is fixed at 9 days (discussed in section 2.2). In addition, the interest is in long-term effects and not in the scale of hours, which would be more of interest for short-term extreme summer precipitation. The maximum hourly wind speed and its corresponding direction is used as maximum wind speed on a day. The aggregation level of the hourly average is appropriate as the interest lies on sustained wind speeds which affect the water surface gradient, not on short wind gusts. The external water level at the Lek is represented by the maximum hourly average water level on a day, which is also the smallest time step of the measurements available for the total time period from 1974 to 2018. A larger aggregation level is not of interest as only one hour above 2.25 m NAP could already affect the normative water levels as it blocks the pumping stations. If the average over multiple hours would be used, this event could be missed.

For the analysis, the daily observations are plotted against each other, the histograms of each of the variables and the measures of dependence are determined to give a broad view on the mean and variance of the variables and additionally, how the variables interact (e.g. type of dependence). There are several measures of dependence such as linear and rank correlation coefficients. The Pearson correlation coefficient measures the degree of linear dependence, but is not always appropriate for hydrological/meteorological data as such variables often have nonlinear dependency and are not Gaussian (symmetric "bell shape") distributed (Hao & Singh, 2016). Rank correlation coefficients such as Spearman's *rho* and Kendall's *tau* overcome those drawbacks as they measure the association between ranks. As this first step is only exploratory and the three measures of dependency gave similar results, the Pearson's correlation coefficient is shown in the results. In addition, ties are present in the data which can affect the results of rank correlations. Ties are values that are exactly the same, which makes it impossible to determine its ranks. This is for example a result of measuring resolution (e.g. wind speed is measured with intervals of 1 m/s). The interpretation of the results is straightforward as a value of 1 represents a perfect linear correlation, 0 represents no correlation, -1 represents a perfect decreasing linear correlation. Along with the correlation coefficients, the p-values are determined to test whether the correlation is significant (95% level). For the circular variable wind direction, the linear-circular correlation coefficient as defined in Mardia et al. (2000) is used, as the scale of direction is not linear, but circular.

As the dependencies between the variables can have a certain time lag, for example the wind speed affecting the external water level of the next day, the correlation coefficient is also determined while shifting the time series with time steps of 1 day up to a maximum of 20 days. This information is used to determine in what time window the variables are coupled in the event-based approach (which is described below), and how those variables should be modelled with the appropriate time lag.

3.2.2 Analysis based on events

In this section, the event-based approach is discussed, which means that relevant events are selected from the data to use for further analysis. The event-based approach is tested as serial correlation (autocorrelation) is present in the time series of the stochastic variables (see Appendix C). In general, classical statistical modelling requires temporal independent and identically distributed (i.i.d.) data, so that the probability distribution of each draw is equal, and not affected by the previous draw (Bernardara et al., 2012; Mazas & Hamm, 2011; Vandenberghe et al., 2010). To check whether the dependencies are still present if an event-based approach is taken, independent precipitation events are selected based on a threshold for the precipitation volume and an inter-peak time to decluster the time series (Bernardara et al., 2012). The threshold and inter-peak time are chosen in such a way that the autocorrelation plots show no significant correlation anymore (95% confidence interval). In general, frequency analyses require high thresholds so that the probabilities are described by an extreme value distribution. However, as the interest in this research does not exclusively lie on extremes for each of the variables and in addition, mainly on the effects of dependencies between the variables, a low threshold is appropriate. Setting a threshold at 10 mm precipitation volume and a minimum inter-event time of 7 days, the autocorrelation plot showed no significant correlation anymore.

As with the analysis on all observations, the correlation coefficients are determined, and scatter plots are produced. Furthermore, the resulting figure is compared to the dependencies between the daily observations, to check whether the dependencies are also found in the event-based approach. The figure of the event-based approach is used as guideline for which variables the strongest dependency is found determining the order of the dependencies to study. In addition, the plots provide information about the type of dependency and give information about the behaviour of the variables.

3.3 Results

A first indication of the dependencies between the stochastic variables considered here is visible in Figure 9. In this figure, the dependencies between all daily observations are determined. Note that considering the linear correlation coefficient for wind direction is not appropriate as the scale of direction is not linear, but circular, as discussed in section 3.2. However, the scatter plots do provide information on the dependencies with the other variables. In addition, the linear-circular correlation coefficients are determined separately in Table 2.

The histograms show that the variables are not normally (gaussian) distributed, which shows the importance to use other methods (e.g. copula), which can manage non-normally distributed variables. Visible is that the strongest correlation can be found between the wind speed and the water level of the Lek. The scatter plot shows a clear positive relation between those two variables. Furthermore, we see that those higher wind speeds are mainly occurring as the wind comes from westerly directions (225° - 315°). Also visible is that the higher external water levels coincide with winds from the west. These dependencies can be mainly explained by the fact that the water levels at the Lek are affected by the North Sea (west of the Alblasserwaard), as a clear tidal effect is visible in the time series of the water levels. The surging effect of those westerly winds on the North Sea, which affect the water levels of the Lek, might explain the relationship between wind speed and water level.

Other dependencies that are visible in this figure are that the wind speeds are generally higher during precipitation events and that during precipitation events the winds are often from the west. This might be caused by meteorological circumstances. Additionally, we see that with an increasing volume of precipitation, the external water level is also increasing, which could be a result of the correlation between the precipitation volume and the wind speeds. Furthermore, the contour lines in the graphs show the complex non-linear relationships between the variables, which requires parametric models able to model such dependence structures (e.g. copula).

All the mentioned dependencies have a high level of significance ($p < 0.001$). This does not mean that there is a direct causal relationship between those variables. It is also possible that other reasons for these relationships exist, e.g. seasonality or indirect causation could partly explain dependencies. For example, more 9-day precipitation events take place during winter, while the peak water levels at the Lek are generally also higher in the winter/spring, amongst others due to less evaporation, more precipitation and meltwater in spring in the catchment area of the Rhine (Lek is a branch of the Rhine).

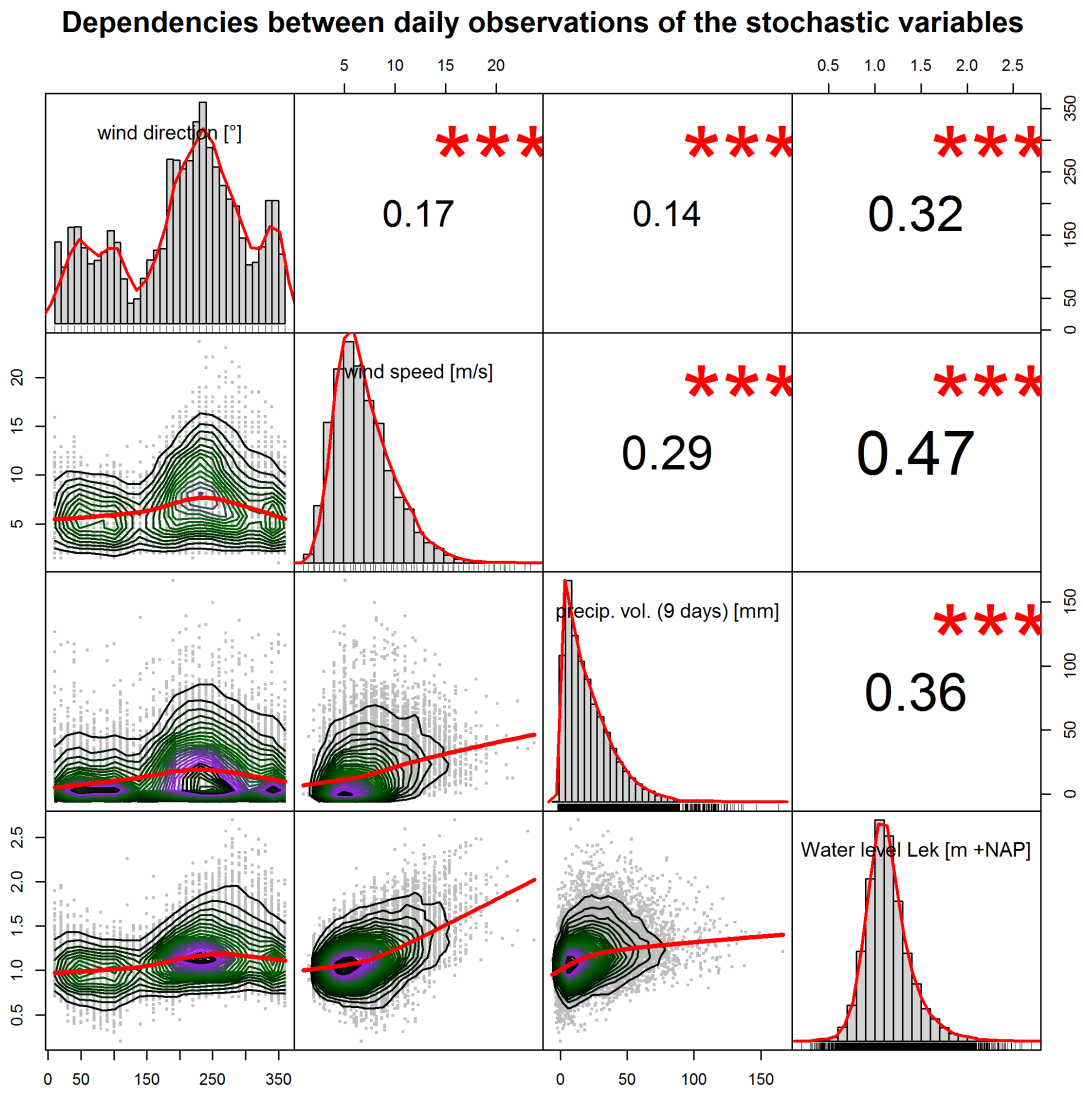


Figure 9: Diagonal represents distributions of each variable with an indicative fitted line. Lower left half represents scatter plots with indicative contour lines (lines with equal density, colours showing degree of density) with a red line representing the indicative relationship. Top right half shows correlations coefficients with p-values of *** = 0.001.

Table 2: Linear-circular correlation with wind direction + other variables

Variables	Linear-circular correlation	P-value
wind direction + speed	0.14	0.001
wind direction + precip. volume	0.08	0.001
wind direction + water level Lek	0.17	0.001

All the mentioned dependencies are found with the observations correlated at the same day. However, as there might be a time lag between the effects of certain variables, the cross-correlation at several time lags is also determined. See Figure 10 for the resulting plots. The maximum time lag in these figures is set at 20 days. Wind direction is not considered here, as wind speed and direction are coupled at the same moment in time. The same figures based on hourly observations as opposed to the daily values can be found in Appendix D. The plots for daily values are shown here, as the relevance for time lags at the scale of hours is less as the general scale is on multi-day precipitation events, not on short intense rain storms. The results for hourly and daily observations are similar.

The correlation between precipitation volume and wind speed is strongest at a time lag of 0 days (middle of 9-day precipitation event). Furthermore, the correlation around this time lag is quite symmetrical. No significant cross-correlation is found after +- 8 days. For the variables wind speed and water level Lek, we see that the strongest correlation is found with a time lag of 0 to 1 day, so the wind speeds affect the water level of the same and next day. However, moderate correlation is found through a long range of time lags, especially when the water level is shifted to the right. The same applies to precipitation volume and the water level at the Lek, where the strongest correlation is found 0-4 days after the middle of the 9-day precipitation sum. This could be caused by the correlation between the precipitation and the wind speed, affecting the water levels at the Lek. An important note is that the precipitation and wind are measured at Rotterdam Airport, which is about 20 km downstream of where the water levels of the Lek are measured. This could also affect those results. However, on the scale of days, this effect is negligible.

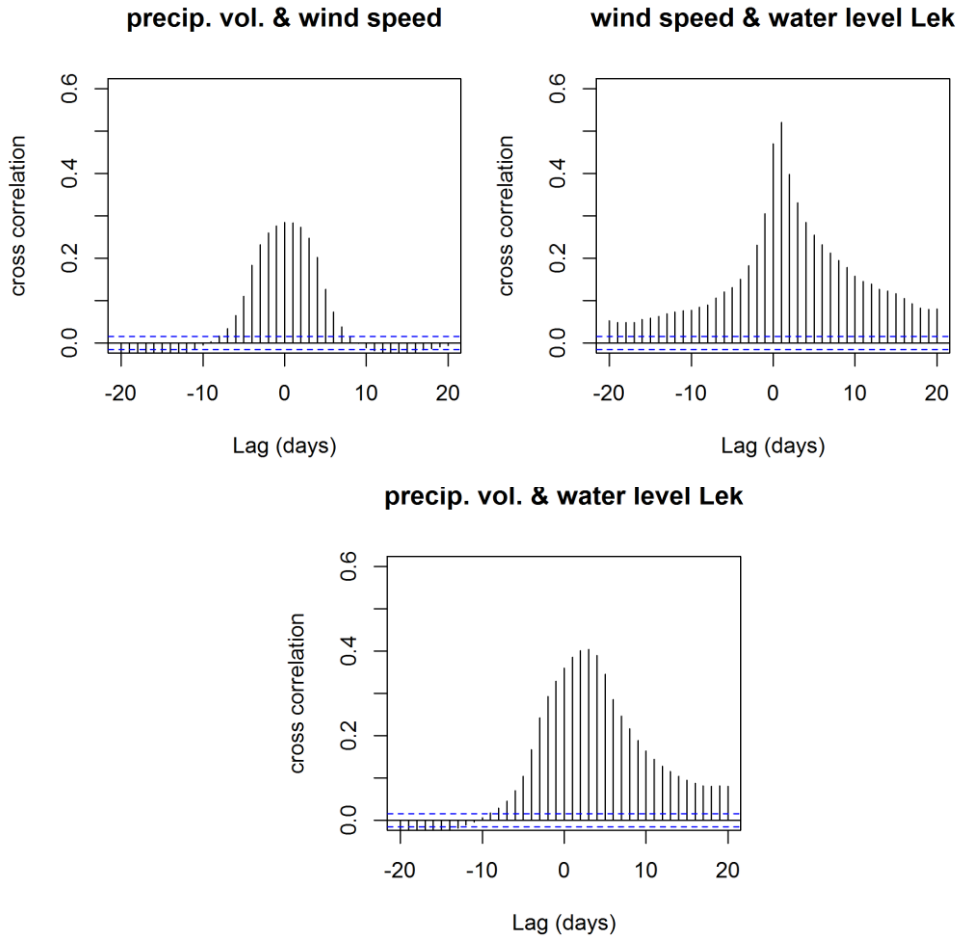


Figure 10: Cross-correlation plots where the 2nd mentioned variable in the title is shifted with a maximum of 20 days. Blue dotted lines represent the critical value for significant cross-correlation (α = 0.05).

As discussed in the methods, independent precipitation events are selected as serial dependency is present in the time series (see Appendix C). Independent events are selected based on the absence of autocorrelation. As each event is described by one value for every variable, a time window had to be determined wherein the values are coupled. This window is determined with the help of the cross-correlation plots in Figure 10.

In this case, independent events (no significant autocorrelation is found) are selected by setting a threshold at a minimum precipitation volume of 10 mm in 9 days, where the minimum time between peaks is 7 days. The accompanying wind speeds are selected with a time window of 3 days (before and after the middle of the 9 days), while the highest water level is determined by selecting the maximum within 2 days of the highest wind speed. The resulting figure for the selected events can be found in Figure 11.

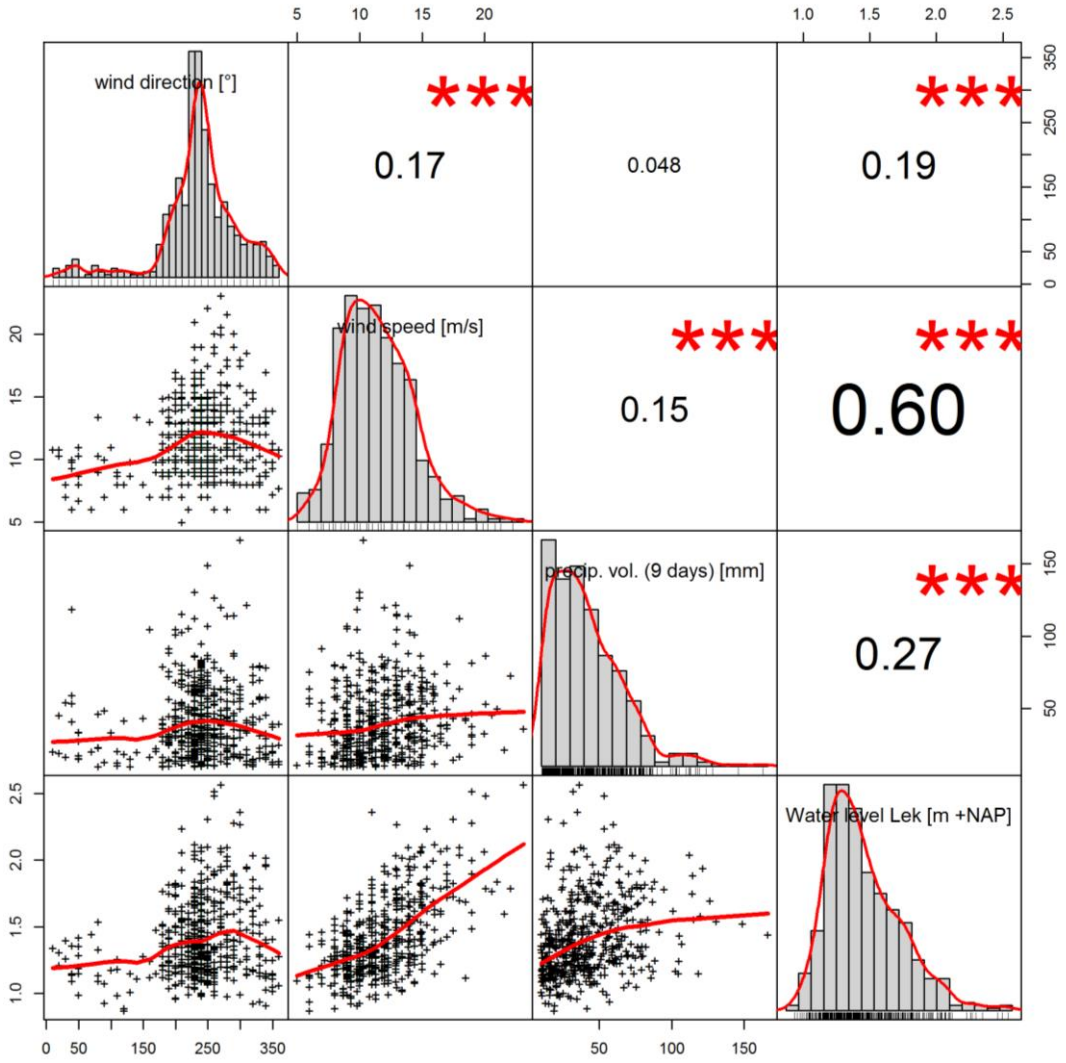


Figure 11: Dependencies between stochastic variables for the selected events. Diagonal represents distributions of each variable with an indicative fitted line. Lower left halve represents scatter plots with a fitted line representing the estimated relationship. Top right shows correlations coefficients with p-values of *** = 0.001.

Table 3: Linear-circular correlation with wind direction + other variables during selected events

Variable	Linear-circular correlation	P-value
wind direction + speed	0.11	0.001
wind direction + precip. volume	0.03	0.001
wind direction + water level Lek	0.07	0.001

Similar to the results based on all observations, the strongest dependency is found between the variables wind speed and water level Lek. For the other combinations of variables, similar or slightly lower dependencies are found. The combinations wind speed - precipitation volume and precipitation volume - water level Lek have a lower correlation

coefficient compared to the daily approach as there is dependency in the lower range of values visible in the daily approach, which is not included in the event-based approach. Notable is how often the strongest wind speeds during those precipitation events are from the southwest to northwest direction, ranging from 200° to 340°.

In Figure 12, the relationship between the variables is further explored, but now in three dimensions, which is relevant when those three variables are considered as dependent to determine the normative water levels. The precipitation events are divided in three groups based on precipitation volume, with an about equal number of events. Density plots and histograms for the variables wind speed and water level Lek are shown per precipitation category. The figure shows that with an increasing precipitation volume, the densities and probability distributions (histograms) shift to higher values of wind speed and water level Lek, which means that those variables are not independent relative to precipitation. In addition, it is visible that the water levels are higher during high wind speeds as seen in the former figures. This figure shows how increasing precipitation volumes could potentially lead to hazardous situations as the wind speeds and water levels at the Lek are also increasing, caused by the interdependencies between those variables.

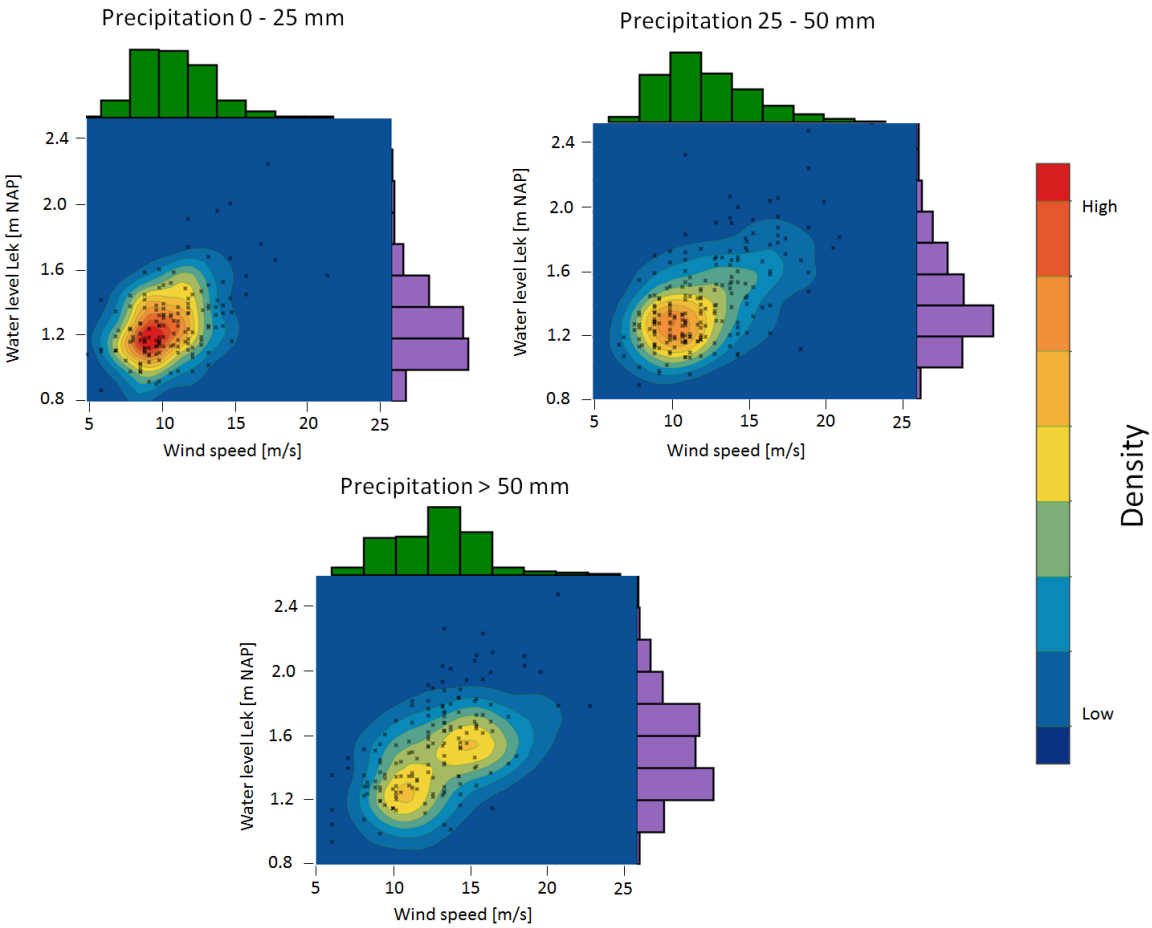


Figure 12: Density plots and histograms of variables Wind speed and Water level Lek for three different classes of precipitation volumes of the selected events

3.4 Conclusion

In this chapter, research question 1: "*Are the stochastic variables used in determining normative water levels interdependent?*" is answered. A clear overview of the present dependencies between the stochastic variables, based on all observations and based on temporally independent events is given. Similar results are found for both the event-based as the daily approach. The strongest dependency is found between the wind speed and the water level of the Lek, possibly caused by the surging effect of winds from the west/northwest on the North Sea and subsequently the Lek. In addition, it is visible that the wind speeds slightly increase with increasing precipitation volumes, which could explain the increasing water level of the Lek with increasing precipitation volume. Furthermore, it is visible that the higher wind speeds, precipitation volumes and water levels of the Lek mainly occur during winds from southwest to northwest (200° - 340°) which could all be caused by meteorological/geographical circumstances such as the presence of the North Sea, west of the Alblasserwaard. In addition, the effect of time lags on the correlation is analysed, which shows that there is a certain time lag of about 1 to 2 days of the water level Lek relative to the wind speed and precipitation volume, which is relevant to consider in the model setup.

It is clear that dependencies cannot be neglected further on in this study. As there is a positive relationship visible between the variables precipitation volume, wind speed and water level of the Lek, it is expected that the normative water levels in the Alblasserwaard are in some way affected by these dependencies and thus relevant to study.

4 Dependent joint probability distributions

4.1 Introduction

In this chapter, the second research question is discussed: "*What are the most suitable joint probability distributions describing each of the combinations of interdependent stochastic variables, and what is the effect of including dependency on the joint probabilities of the stochastic variables?*". In Chapter 3, the exploration of dependencies, (non-linear) relationships between the (non-normally distributed) variables are found. Based on the results of research question 1, the pairs of dependent variables are divided over separate combinations of variables. This prevents missing out individual effects of dependencies, which is also relevant considering the generalisation of the results to other areas. These combinations are:

1. Wind direction & speed + water level Lek (bivariate)
2. Precipitation volume + water level Lek (bivariate)
3. Wind direction & speed + precipitation volume (bivariate)
4. Wind direction & speed + precipitation volume + water level Lek (multivariate)

For each combination of variables, a joint probability distribution is constructed to analyse the joint probabilities. The described methods in section 4.2 to construct the joint probability distribution are divided in two sections, one for the bivariate case (dependency between two variables) and the other for the multivariate case (dependency between more than two variables). In addition, the validation by simulation from these joint probability distributions is discussed, which helps in visualising the effect of dependency on the probabilities and validating the appropriate choice in dependence structure and marginal distributions. Note that wind direction is handled in combination with wind speed as wind direction is a directional variable which is difficult to be handled sufficiently using parametric copulas (Carnicero et al., 2013; Soukkisian & Karathanasi, 2017). Section 4.2 could be technically/mathematically complicated, but it is not necessary (for the reader) to understand all mathematical background to understand the process and results. The results are presented per combination of dependent variables in section 4.3. The conclusion of this chapter answering the second research question can be found in section 4.4.

4.2 Methods

The methods to construct the joint probability distributions are divided in two sections. The construction based on two variables (bivariate copula) is discussed in section 4.2.1. Using more than two variables, the multivariate case, is discussed in section 4.2.2 (vine copula).

4.2.1 Bivariate copula

Before discussing the taken approach for the bivariate cases, a short introduction to the theory is given. To analyse the dependence structure and construct the joint probability distributions, the method 'copula' is used. A copula is a bivariate cumulative distribution function, which is used as a method to construct joint probability distributions. The construction is separated in the marginal distributions (probability distribution of single variable) and the dependence structure (Sklar, 1959; Nelsen, 2006; Genest & Favre, 2007). The dependence structure describes how two (or more) variables are related to each other. For example, is there a linear relationship between the variables throughout the whole range of values, is there only a relationship for low or high values, or any other kind of relation between the variables. The main advantages of the copula over other methods is that each marginal distribution can take any form and is modelled separately from the dependence structure. Using the theory of copula, the joint cumulative distribution $H(x,y)$ of any pair of continuous random variables (X,Y) with marginal cumulative distributions $F_X(x)$ and $F_Y(y)$ is given by:

$$H(x,y) = \boldsymbol{C}(F_X(x), F_Y(y); \theta) = \boldsymbol{C}(u, v; \theta) \qquad \textit{equation (1)}$$

Where **C** represents the copula function, of which the dependence structure depends on its parameter(s), represented here by θ. There are many copula functions describing all sorts of dependence structures, which will be discussed later. The variables u and v are the probabilities of non-exceedance or also-called normalised ranks between 0 and 1 (transformed from the variables X and Y), as the copula function requires uniform marginals over (0, 1). Each continuous distribution can be converted to a uniform distribution using the probability integral transform. This means that each value is represented by its probability of non-exceedance. This gives an approximate uniform distribution on (0, 1) as the number of values between for example 0.05 and 0.10 is the same as between 0.70 and 0.75.

To help understanding different copula functions and the function of its parameter(s), the probability density functions of two copula functions (Joe and Frank copula) are shown in Figure 13. The more the values are around the u/v diagonal, the stronger the dependency between the two. Visible is how the Joe copula represents strong dependency for high values of u and v, while the Frank copula represents dependency throughout the whole range of values and particularly the higher and lower values. Furthermore, the effect of the copula parameter is visible (expressed as rank correlation coefficient *tau)*, which determines the strength of the dependence between the two variables. The higher the copula parameter, the higher the strength of the dependency.

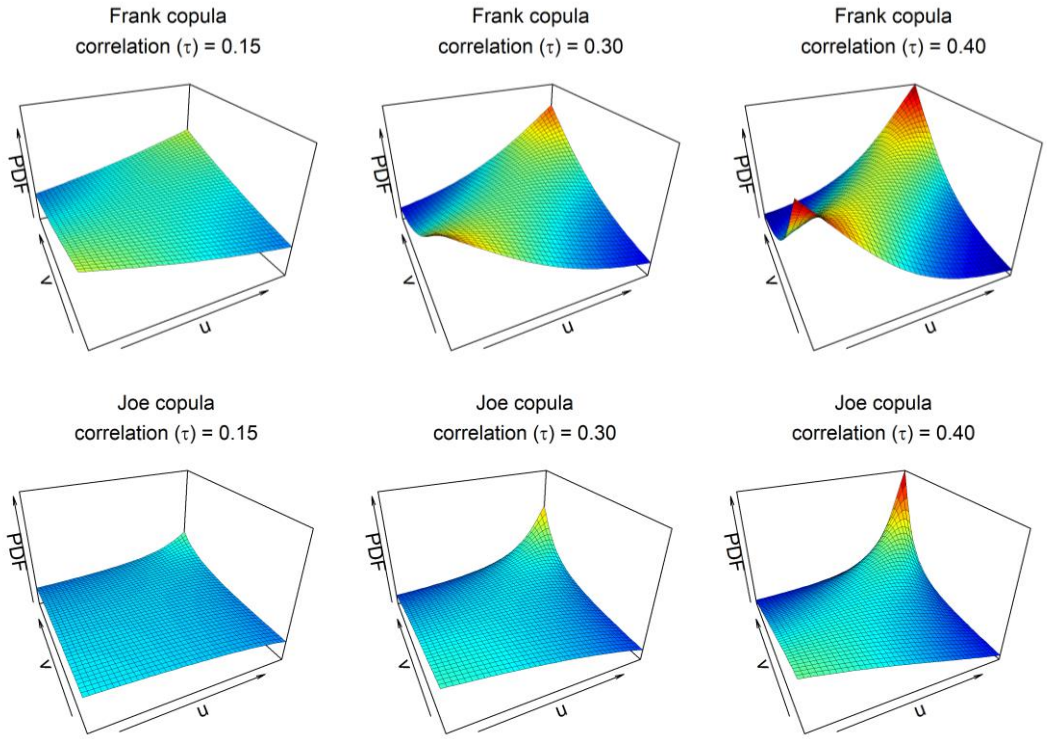


Figure 13: Probability density functions (PDF) of two copula functions (Frank and Joe) (with uniform marginals) for various values of the copula parameter, expressed here as *tau (τ)* which measures the nonlinear correlation (-1 to 1) between the variables.

To choose an appropriate copula function for the given data, the empirical copula based on the observations is used. The empirical copula is made using the probability integral transform which transforms each value to its empirical probability of non-exceedance. As an example, the normalised ranks (probability of non-exceedance) for the variables wind speed and water level at the Lek are shown in Figure 14. From this empirical copula, you can see what type of dependence is present between the two variables (e.g. left tail dependence, main body dependence). Visible is the high density of observations in the upper right corner, which means that there is a lot of right tail dependence, while in the main body of observations, the dependency is less strong. Right tail dependence means that if one of the observations is a high value, it is likely that the other one is also high.

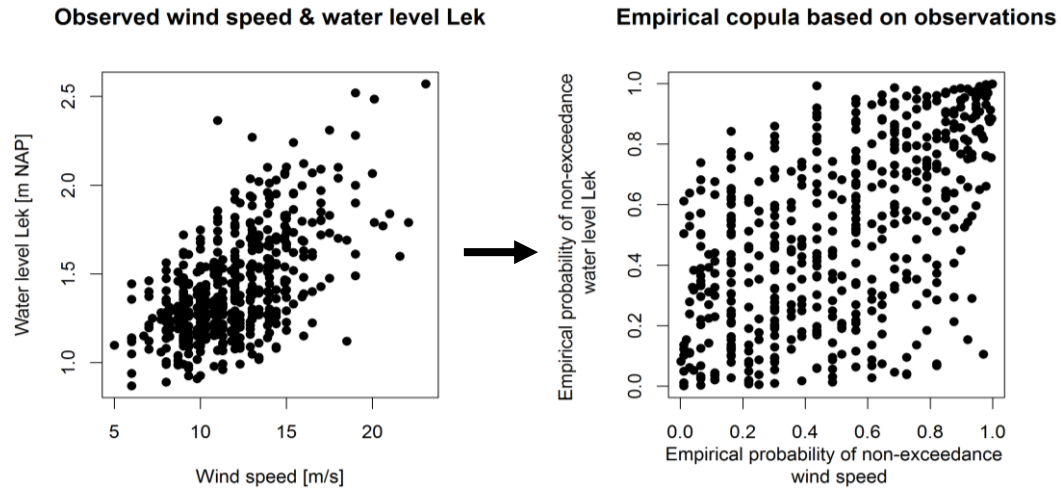


Figure 14: Example of empirical copula/normalised rank scatter plot for the variables wind speed & water level Lek using the selected events from section 3.2. Constructed using the pobs() function in the R package 'copula'.

There are several copula functions which describe different dependence structures. To illustrate how a simulation from a copula looks like in the bivariate case to compare it to the empirical copula, simulations from the Clayton (strong left tail dependence) and Joe (strong right tail dependence) copula are illustrated in Figure 15. A simulation means that a random value is picked from a continuous distribution, such as in Figure 13. Based on Figure 14 and Figure 15, you could say that the Joe copula is a potential copula function for the dependency between wind speed and water level at the Lek in this case.

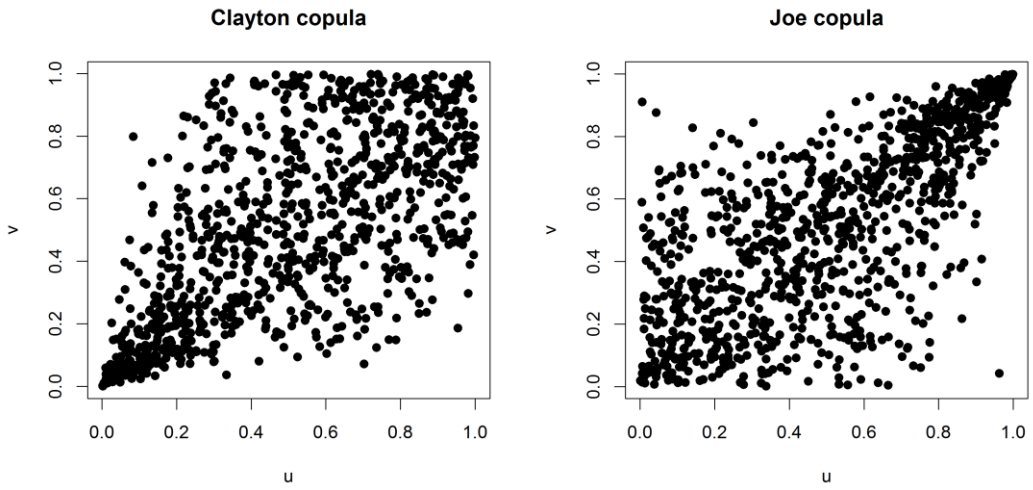


Figure 15: Simulation (N=1000) from Clayton and Joe copula with a correlation value of 0.5 (Kendall's tau) using the 'copula' package in R (Yan, 2007). Clayton copula shows strong left tail dependence, while the Joe copula shows more right tail dependence.

To summarise, a copula function describes a dependence structure on the interval of 0 to 1. This value represents the probability of non-exceedance. To construct a joint probability distribution with real values, these probabilities of non-exceedance can be tranformed back to the real values using the marginal distributions. So, the combination of the dependence structure and the marginal distributions lead to a joint probability distribution.

Constructing the joint probability distribution in the bivariate case

Now that the theory of copula is briefly explained, the stepwise approach to construct a joint probability distribution in the bivariate case is discussed. These steps are shown in the flowchart in Figure 16. The fitting of the marginal distributions and the copula functions are parallel processes as the copula functions are fitted based on the empirical probabilities of non-exceedance, so that the fitting is insensitive to the marginal distribution selection, which will be discussed later in this section.

The first step in constructing the joint probability distribution for the two variables of interest is to select the data to include in this analysis. Individual events are selected (similar to Chapter 3), as in general temporally independent observations are necessary to construct a (joint) probability distribution (Nelsen, 2006). Independent events are selected based on a threshold and minimum time between peaks. As this study focuses on the effects of dependency, a low threshold is set to preserve the dependence structure. This process is in detail discussed in Appendix F. The selected events are divided over the relevant wind directions (NW, W, SW and other) for the combinations which include wind speed.

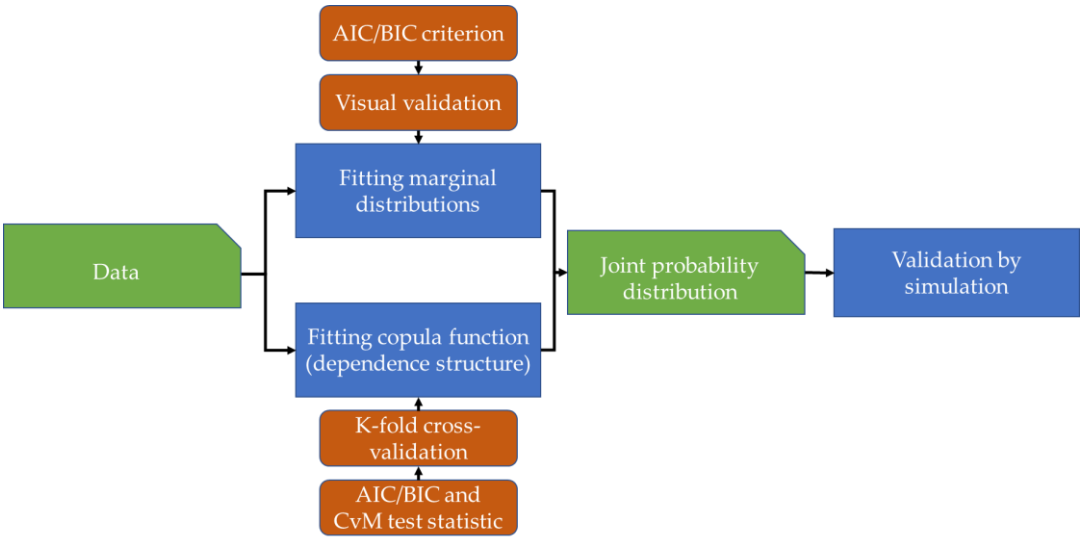


Figure 16:Flowchart constructing joint probability distribution in the bivariate case

The next step is to fit parametric marginal distributions to each of the considered variables. Typical probability functions of hydrological/meteorological variables, which describe a wide range of different distribution shapes, are tested on the data, see Table 4 (Grimaldi et al., 2011). A first selection is made based on the Akaike's information criterion (AIC) and Bayesian information criterion (BIC) (Hofert et al., 2018):

$$AIC = 2k - 2\ln(L) \qquad \textit{equation (2)}$$

$$BIC = \ln(n)\,k - 2\ln(L) \qquad \textit{equation (3)}$$

where k is the number of parameters of the distribution, n is the number of observations and L is the maximised value of the likelihood function (higher is better). This likelihood function, which depends on which distribution is tested, is a function that relates the distribution parameters with the probability of observing the sample by the parametric model. AIC and BIC are used to compare different models on how similar the observed and modelled values are, while penalising for the number of parameters to prevent overfitting (Delignette-Muller & Dutang, 2015). Both are used as in general BIC performs better for large samples, while AIC performs better for smaller samples (Shumway & Stoffer, 2017). The parameters of each distribution are estimated by maximum likelihood estimation (MLE), which in general performs well (Grimaldi et al., 2011). For the variables on which a threshold is set, also the shifted versions of the parametric distributions are tested, as values below the threshold are not possible. This is executed by setting the location parameter of the distribution at the value of the threshold. The three best scoring distributions (lowest AIC/BIC score) are then visually compared to the empirical distribution to validate whether the choice of the distribution based on the AIC/BIC scores is appropriate. This visual goodness-of-fit test is done using four classical plots for such analyses; density plot along the histogram, cumulative density function (CDF) comparison, Q-Q plot representing the empirical and theoretical quantiles and a P-P plot representing the empirical and theoretical percentiles (Delignette-Muller & Dutang, 2015). All these calculations and plots are executed using the **R** package *fitdistrplus.*

Table 4: Probability distributions which are tested on the stochastic variables in this study (Grimaldi et al., 2011). With distribution parameters $\theta_1, \theta_2, \theta_3$ and gamma function Γ.

Distribution	PDF	Range
Normal	$f_x(x) = \frac{1}{\theta_2\sqrt{2\pi}} e^{\left(\frac{1}{2}\left(\frac{x-\theta_1}{\theta_2}\right)^2\right)}$	$-\infty < x < \infty$
Lognormal	$f_x(x) = \frac{1}{x\theta_2\sqrt{2\pi}} e^{\left(\frac{1}{2}\left(\frac{\log(x)-\theta_1}{\theta_2}\right)^2\right)}$	$x > 0$
Exponential	$f_x(x) = \frac{1}{\theta_2} e^{\left(-\frac{x-\theta_1}{\theta_2}\right)}$	$x > \theta_1 \; for \; \theta_2 > 0$
Gamma	$f_x(x) = \frac{1}{\lvert\theta_1\rvert\Gamma(\theta_2)}\left(\frac{x}{\theta_1}\right) e^{\left(-\frac{x}{\theta_1}\right)}$	$x \geq 0$
Gumbel	$f_x(x) = \frac{1}{\theta_2} e^{\left(-\frac{x-\theta_1}{\theta_2} - e^{-\frac{x-\theta_1}{\theta_2}}\right)}$	$-\infty < x < \infty$
Weibull	$f_x(x) = \frac{\theta_2}{\theta_1}\left(\frac{x}{\theta_1}\right)^{\theta_2-1} e^{-\left(\frac{x}{\theta_1}\right)^{\theta_2}}$	$x > 0$
Generalised Pareto	$f_x(x) = \frac{1}{\theta_2}\left(1 - \theta_3\left(\frac{x-\theta_1}{\theta_2}\right)\right)^{\frac{1}{\theta_3}-1}$	$\theta_1 \leq x < \infty \; if \; \theta_3 > 0$

Next to the marginal distributions, an appropriate copula function is fitted to the data. These steps are performed using the **R** packages *copula* and *VineCopula*, which are widely applied packages in the research field (Yan, 2007; Hofert et al., 2018; Schepsmeier et al., 2018). Example code can be found in Appendix S. Based on the implementation in **R** and the occurrence in literature, the parametric copula functions in Table 5 are tested on the data. Also, the rotated versions (180°) are tested. Rotated versions of copulas can be visually imagined by rotating Figure 15 with 180°, so that a rotated Clayton copula shows strong right tail dependence and a rotated Gumbel copula shows strong left tail dependence. This expands the number of dependence structures which can be described.

Table 5: Copula functions which are tested on the stochastic variables (Schepsmeier et al., 2018)

Copula name	Range parameter 1	Range parameter 2	Description
Gaussian	[-1,1]	-	No tail dependency, symmetric
Student-t	[-1,1]	[2, ∞)	Lower and upper tail dependence
Clayton	(0, ∞)	-	Lower tail dependency
Gumbel	[1, ∞)	-	Upper tail dependency
Frank	(-∞ , ∞)	-	Main body dependency, no tail dependency
Joe	[1, ∞)	-	Upper tail dependency
BB1	[0, ∞)	[1, ∞)	Combination of Clayton and Gumbel copula
BB6	[1, ∞)	[1, ∞)	Combination of Joe and Gumbel copula
BB7	[1, ∞)	(0, ∞)	Combination of Joe and Clayton copula
BB8	[1, ∞)	(0, 1]	Combination of Joe and Frank copula
Tawn type 1	[1, ∞)	[0, 1]	Asymmetric, upper tail dependence
Tawn type 2	[1, ∞)	[0, 1]	Asymmetric, upper tail dependence

The first step in the fitting process is to estimate the copula parameters. Using the *BiCopEst()* function in the *VineCopula* package, copula parameters are estimated using the maximum pseudo-likelihood (MPL) (Schepsmeier et al., 2018). This method is better than other methods such as maximum likelihood (MLE) or inference functions of margins (FML) as MPL is based on the normalised ranks or pseudo-observations and is thus insensitive to the specification of the marginal distributions (Kim et al., 2007; Hofert et al., 2018). This prevents that misspecifying the marginal distributions also results in misspecifying the dependence structure. Using the maximum pseudo-likelihood estimation, it is also possible to determine its standard error. This is used to determine 95% confidence intervals (1.96*standard error), to show the effect of the uncertainty in copula parameters on the normative water levels (Genest et al., 1995).

A first selection in best fitting copula functions is based on the earlier mentioned AIC and BIC criteria (equation 2 and 3), which are commonly applied and well-performing in the context of selecting the best fitting copula (Hofert et al., 2018; Jordanger & Tjøstheim, 2014). These criteria are computationally fast and are appropriate to make a selection in best fitting copulas. The 5 best fitting copulas are then further examined with computationally costly goodness-of-fit tests.

Test results of AIC and BIC tests only state something about which parametric copula fits relatively best, and not if the copula significantly represents the empirical copula. The resulting scores (values) do not have a meaning, only the order has. Therefore, the formal goodness-of-fit test of Cramér-von Mises is used additionally. This test statistic performs better than other similar goodness-of-fit tests (Genest et al., 2009; Hofert et al., 2018). The test statistic is given by the equation:

$$S_n^{gof} = \sum_{i=1}^{n} \left(C_n(U_{i,n}) - C_{\theta_n}(U_{i,n}) \right) \qquad \textit{equation (4)}$$

Where C_n represents the empirical copula and C_{θ_n} the estimated parametric copula with $U_{i,n}$ representing the normalised observations. So, the test statistic measures the distance between the parametric and empirical copula. To validate whether the parametric copula represents the empirical copula, P-values are determined using a parametric bootstrap procedure (Genest et al., 2009; Hofert et al., 2018). This procedure produces P-values by generating N number of x observations from the parametric copula and determine the number of simulations that significantly represent the empirical copula. So, formally the following is tested:

$$\mathcal{H}_0: \; C_{emp} \in C_{par} \quad versus \quad \mathcal{H}_1: \; C_{emp} \notin C_{par}$$

If P-values are below 0.05, we can reject that the empirical copula is significantly represented by the parametric copula (95% level of significance). This test is implemented in the **R** package *copula* with the function *gofCopula()*. Based on Genest & Favre (2007), the number of simulations is set at N = 10,000. This is only performed on the 5 selected copulas as this test requires a lot of computational time. Ties in the data are handled by using a random order for these values, based on findings of Hofert et al. (2018).

The fitting and validation process are assessed on the same data to use the full length of the available time series. Otherwise, it was not possible to test the copula on extreme values, as these values do not occur frequently and are not well represented if only a part of the time series is used. Therefore, the last step in choosing the appropriate copula function is to cross-validate the potential copula functions. K-fold cross-validation is used to prevent that the choice of which part of the time series is used for training and which part is used for validation affects the results. In this case, 5-fold cross-validation is used, which means that the data is divided in 5 equally sized samples, where 4 (80%) of them are used for training and 1 (20%) is used for validation. This process is then repeated 5 times. This process is implemented in the **R** package *copula* with the function *xvCopula()*, which is based on the work of Grønneberg & Hjort (2014). The result of this test is the cross-validation copula information criterion. This test is performed on the same 5 selected copulas that are selected earlier as this test also requires a lot of computational time. This should not be a problem as the result of the AIC/BIC tests in general do not differ a lot from the result of the cross-validation copula information criterion, so the risk of missing the best fitting copula by only checking the selected 5 is small (Jordanger & Tjøstheim, 2014).

The copula function that is not rejected at a 95% level of significance and scores the highest on the cross-validation information criterion is selected as the most appropriate copula. Now the dependence structure (copula) and marginal distributions are selected, the joint distribution is constructed using the *mvdc()* function in the **R** package *copula*. This joint probability distribution is used to determine the probabilities and simulate pairs of variables.

Validation by simulation

To validate the combination of the selected marginal distributions and copula function, validation by simulation is performed. Additionally, this is used to check what the effect of including dependency on the probabilities is, compared to assuming no dependency. The distribution of the probabilities for the dependent case is based on the selected copula, the other one is based on the multiplication of the probabilities as independence is assumed (also called: independence copula).

The number of draws for this simulation is set based on the number of observations. As it is only a visual comparison and it has no effects on the results, it is set at N= 5,000 for combination 1 and N = 2,000 for the other combinations (with less observations). A lower number of draws results in a different result for every new simulation as this process is based on a random number generator. The results are checked on how well the simulated values represent the observations. The comparison with the independent case is relevant as an expectation on the effects on the normative water levels can be made, for example based on how joint extremes are represented if independence is assumed. Note that this comparison is based on the same selected observations and marginal distributions, so that the only difference is the method to calculate the probabilities. This helps in understanding the effect of dependency on the probabilities. Otherwise, it is not clear if the difference between the dependent and independent case is caused by selecting the appropriate data and marginal distributions or by the methods to determine probabilities.

4.2.2 Vine copula (multivariate)

The previous section is focussed on the bivariate copula for two variables. Constructing a higher-dimensional copula is generally recognised as a difficult problem compared to the bivariate case, as there is a vast number of bivariate copulas, while the number of higher-dimensional copulas is limited (Aas et al., 2009). In addition, the flexibility of those higher-dimensional copulas in modeling complex dependencies is limited and not often applied. To cope with these problems, Aas et al. (2009) (based on work from others (Bedford & Cooke, 2001; Bedford & Cooke, 2002; Joe, 1996)) proposed a decomposition of pair-copula to extend bivariate copulas to higher dimensions. Pair-copulas or also called vine copulas (illustrated in Figure 17) are building blocks of bivariate copulas to construct higher-dimensional copulas to model complex multivariate dependent data. By using bivariate copulas and combining them, a wide variety of dependence structures is achievable as there is a high number of bivariate copula functions.

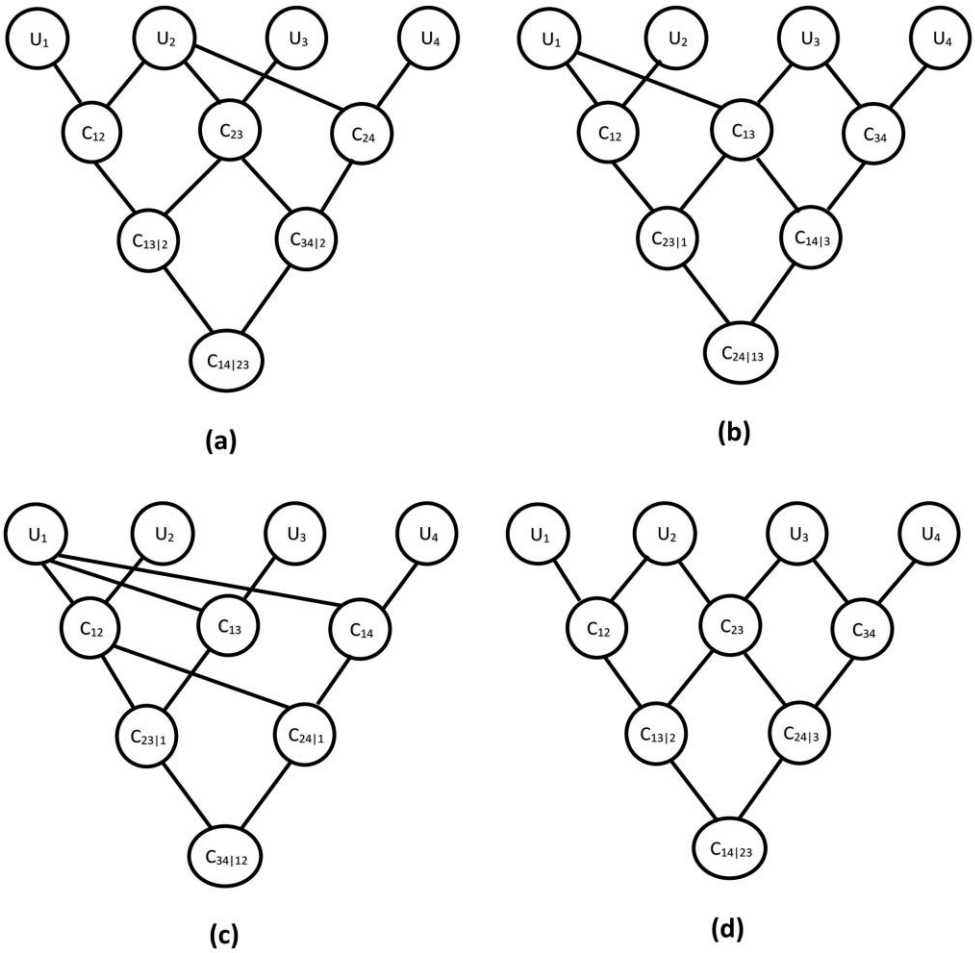


Figure 17: Illustration of R-Vine copula (a,b), C-Vine copula (c) and D-Vine copula (d) (Pham et al., 2018)

There are many possibilities to construct a vine copula, for example 240 different constructions exist in the case of 5 variables (Aas et al., 2009). Some options for four variables are shown in Figure 17, where U stands for the normalised variables uniform on (0, 1), C_{12} stands for the bivariate copula between variable 1 and 2, and $C_{13|2}$ stands for the bivariate copula between the variables 1 and 3, depending on variable 2. There are different types of vine copulas, such as regular vine copula (R-vine copula, Figure 17(a,b)), canonical vine copulas (C-vine copula, Figure 17c) and D-vine copula (Figure 17d)(Kurowicka & Cooke,

2007; Hao & Singh, 2016). C-vine and D-vine copulas are specific cases of regular vines. A C-vine copula is generally used when all the dependencies involve the same variable, a D-vine copula is used when dependencies are considered sequentially (there is no main variable that affects all dependencies) (Pham et al., 2018; Kurowicka & Cooke, 2007). Choices herein are mainly based on a physical basis to reduce the number of possible structures. If all the dependencies originate from one variable, a C-vine copula is generally used, while a D-vine is generally used when there are interdependencies between all the variables.

Constructing the joint probability distribution in the multivariate case

To apply this theory to the multivariate data of the Alblasserwaard, mainly the *R* package *VineCopula* is used (Schepsmeier et al., 2018). This package provides tools for statistical analysis of vine copula models, parameter selection and goodness-of-fit tests. The steps to construct a vine copula as used in this study (combination 4) is shown in Figure 18. Example code can be found in Appendix S.

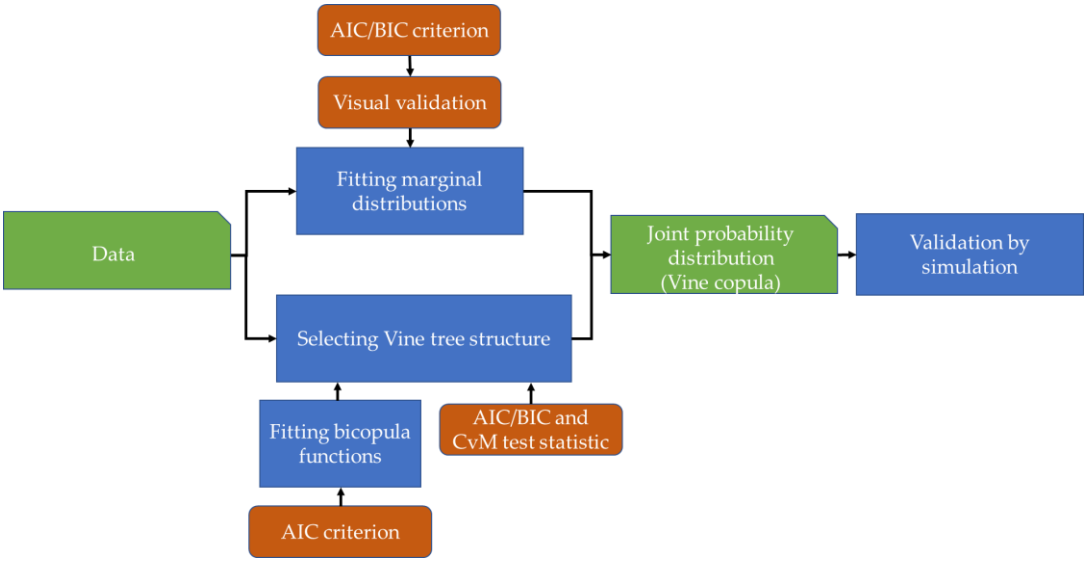


Figure 18: Flowchart constructing joint probability distribution in the multivariate case

The first step is to select temporal independent events to include in constructing the vine copula model. These events are the same events as used for combinations 2 and 3, where a threshold is set for the variable precipitation volume, which leads to 488 precipitation events (see Appendix F). To deal with wind direction, the events are divided over the relevant directions NW, W, SW and Other directions, similarly to combination 3 (precipitation and wind speed). The next steps are to fit marginal distributions to the data and select an appropriate vine copula model. Fitting the marginal distributions is similarly performed as in the bivariate case, using the AIC/BIC criterion in combination with a visual validation. Since in combination 3 the marginal distributions for the two variables (precipitation volume and wind speed) are already determined per wind direction, it is only necessary to fit the variable 'water level Lek', as in combination 2 (precipitation and water level Lek) the dataset is not divided per wind direction.

To select an appropriate vine copula model for each direction, which consist of a vine tree structure and bivariate copula functions, the following procedure is used. As there are only three variables in this case (precipitation volume, wind speed and water level Lek), a C-vine

and D-vine structure is the same (Aas et al., 2009). With three variables, there are only three different vine structures possible (see Figure 19), where only the ordering differs between the options. This prevents the need to make choices in advance and makes it possible to test all three structures within an acceptable amount of time. For each wind direction and tree structure, all bivariate copulas are selected automatically by fitting each copula using the maximum pseudo-likelihood and testing with the AIC criterion (using the *RVineCopSelect()* function). The bivariate copula functions with the lowest AIC score are selected. Advantage of the maximum pseudo-likelihood method is that standard errors can also be determined, which is used to quantify the uncertainty of the vine copula. Contrary to the bivariate combinations 1,2 and 3, goodness-of-fit tests are not applied for every bivariate copula as this would cost too much time. In addition, the number of possible bivariate copula functions (see Table 6) is reduced compared to the other combinations due to restrictions in the package *VineCopula*, which makes it necessary to re-select the copula functions. Vine tree structures with two parameter Archimedean copula functions (BB1, BB6 etc.) are not compatible with all functions in this package.

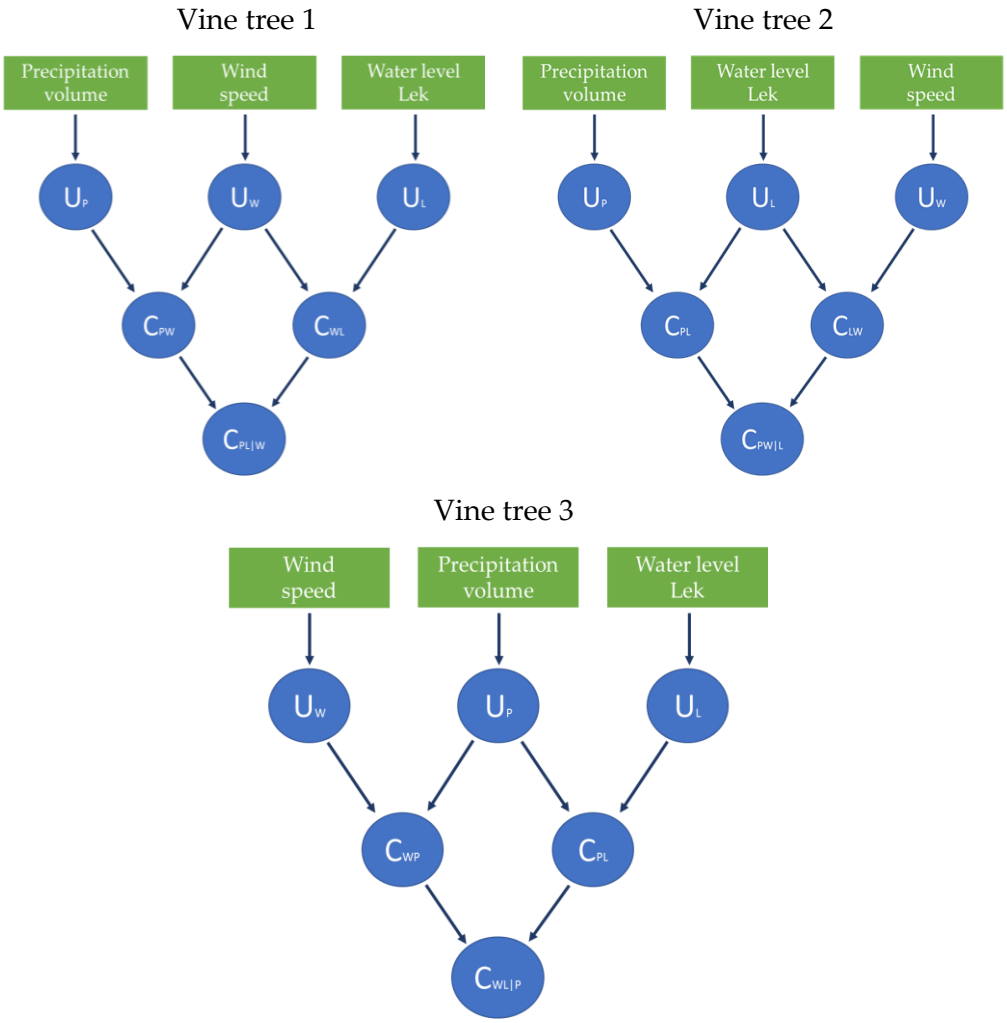


Figure 19: Three candidate vine tree structures for the three variables wind speed, precipitation volume and water level Lek

Table 6: Available bivariate copula functions for the R-vine models (Schepsmeier et al., 2018)

Copula name
Gaussian
Student-t
Clayton
Gumbel
Frank
Joe

When the three options of vine copula structures (see Figure 19) and its fitted bivariate copula functions are constructed, the selection between them is made based on the AIC/BIC scores and additionally the Cramér-von Mises test statistic. This is similar to the bivariate combinations, but now it is performed on the vine copula. The AIC/BIC score shows which of the vine copulas are relatively best able to reproduce the behaviour of the data, while it does not state anything about statistical significance. The AIC/BIC score of a vine copula model is the sum of all the individual pairwise AIC/BIC scores (Schepsmeier et al., 2018). The Cramér-von Mises test statistic (using *RVineGofTest()*) determines whether the vine copula model is significantly able to reproduce the dependence structure of the multivariate data (Schepsmeier et al., 2018; Pham et al., 2018). The implementation in *R* is based on the methods of Schepsmeier (2013). The statistically significant (CvM score higher than 0.05) vine copula with the lowest AIC/BIC score is chosen as most appropriate. This procedure is performed for the directions NW, W and SW. For the category Other directions, a bivariate copula is constructed as there is no interest in wind speed in this direction. Selection is based on the same procedure with AIC/BIC and Cramér-von Mises test statistic.

Similar to the bivariate case, a simulation (N = 2,000) from the joint probability distribution is performed to visualise how well the simulated values correspond with the observations. In addition, this is compared to a situation where independence is assumed, to analyse what the effect of dependency is, and how this compares to the bivariate cases.

4.3 Results

The resulting joint probability distributions based on the marginal distributions and copula functions are discussed here per combination of dependent variables:

1. Wind direction & speed + water level Lek (bivariate)
2. Precipitation volume + water level Lek (bivariate)
3. Wind direction & speed + precipitation volume (bivariate)
4. Wind direction & speed + precipitation volume + water level Lek (multivariate)

For each combination is discussed how the joint probability distribution is constructed and how the dependency affects the probabilities, which could be used in understanding the results in Chapter 5 on the effects on the normative water levels.

4.3.1 Combination one (Wind direction & speed + water level Lek)

Marginal distributions

The first step was to determine the best fitting marginal distributions for the selected observations. A top 3 per variable is determined based on the AIC/BIC criterion. This is visible in Table 7.

Table 7: Candidate marginal distributions and AIC/BIC scores. Bold values are best scores (lowest value).

Distribution	AIC	BIC
Wind speed (NW)		
Lognormal	**5084.7**	**5094.8**
Gumbel	5090.1	5100.2
Gamma	5098.0	5108.1
Wind speed (W)		
Lognormal	**13947.4**	**13959.3**
Gumbel	13951.3	13963.2
Gamma	13978.9	13990.8
Wind speed (SW)		
Gamma	**14082.0**	**14094.0**
Gen. Pareto	14083.8	14101.8
Lognormal	14093.4	14105.4
Water level Lek (NW)		
Lognormal	35.9	45.9
Gamma	93.9	104.0
Gumbel	**-39.8**	**-29.8**
Water level Lek (W)		
Lognormal	468.3	480.3
Gamma	602.9	614.8
Gumbel	**285.3**	**297.2**
Water level Lek (SW)		
Lognormal	-515.7	-503.7
Gamma	-433.9	-421.9
Gumbel	**-585.5**	**-573.5**
Water level Lek (O)		
Lognormal	**-974.9**	**-963.8**
Gamma	-943.4	-932.3
Gumbel	-943.1	-926.5

The visual comparisons of the top 3 distributions per variable can be found in Appendix K. Extra attention is given to the extreme values, as this could be critical in determining normative water levels. The selected marginal distributions fit the empirical distributions very well, in most cases multiple distributions could fit the data appropriately. For wind speed (W) and water level Lek (O), the best scoring (AIC/BIC) marginal distributions are not chosen, as respectively the Gamma and Gumbel distributions fitted the data better based on the visual comparison, mainly focussing on the higher values. The selected distributions and its parameters can be found in Table 8.

Table 8: Selected marginal distributions per variable with distribution parameters

Variable	Selected distribution	Parameter 1	Parameter 2
Wind speed (NW)	Lognormal	scale = 2.01 (-)	shape = 0.28 (-)
Wind speed (W)	Gamma	scale = 7.90 m/s	shape = 0.88 (-)
Wind speed (SW)	Gamma	scale = 7.80 m/s	shape = 0.95 (-)
Water level Lek (NW)	Gumbel	location = 1.13 m	shape = 0.20 (-)
Water level Lek (W)	Gumbel	location = 1.20 m	shape = 0.21 (-)
Water level Lek (SW)	Gumbel	location = 1.19 m	shape = 0.19 (-)
Water level Lek (O)	Gumbel	location = 0.99 m	shape = 0.17 (-)

Dependence structure

Parallel to the marginal distributions, the dependence structure is selected. This is performed in three steps. First, a selection among all the candidate copula functions is made based on the AIC/BIC criteria. Then p-values are determined for the top 5 copula functions using the Cramér-von Mises test statistic. To choose among the statistically significant copula functions, the 5-fold cross validation score is used. Results can be found in Table 9, Table 10 and Table 11 for the directions NW, W and SW.

Table 9: Scoring of the top 5 copula functions for the variables wind speed (NW) and water level Lek. Bold scores are best score (AIC/BIC/Cross-validation) or non-rejected copulas (CvM p-value > 0.05).

Copula name	AIC	BIC	CvM p-value	5-fold cross validation
BB8	**-336.5**	**-326.4**	**0.32**	**159.1**
rot. Clayton	-328.0	-323.0	**0.10**	152.5
rot. BB1	-326.5	-316.4	**0.06**	153.0
rot. BB7	-326.0	-315.9	0.04	153.9
Gumbel	-323.1	-318.0	**0.14**	151.0

For the northwest direction, the BB8 copula (lower and main body dependency) is clearly the best option as it has the lowest AIC/BIC scores, it is statistically significant and has the highest 5-fold cross validation score.

Table 10: Scoring of the top 5 copula functions for the variables wind speed (W) and water level Lek. Bold scores are best score (AIC/BIC/Cross-validation) or non-rejected copulas (CvM p-value > 0.05).

Copula name	AIC	BIC	CvM p-value	5-fold cross validation
Gumbel	**-1200.6**	**-1194.6**	**0.26**	566.1
BB1	-1199.5	-1187.6	**0.08**	566.9
BB6	-1198.5	-1186.6	**0.15**	**567.4**
rot. BB1	-1195.5	-1183.6	0.00	564.8
BB8	-1180.9	-1168.0	0.00	567.4

For the direction west, three statistically significant copula functions are able to describe the dependence structure of the wind speed and water level of the Lek. From these three, the BB6 copula (upper tail dependency) has the highest 5-fold cross validation score and is therefore chosen as most appropriate copula function.

Table 11: Scoring of the top 5 copula functions for the variables wind speed (SW) and water level Lek. Bold scores are best score (AIC/BIC/Cross-validation) or non-rejected copulas (CvM p-value > 0.05).

Copula name	AIC	BIC	CvM p-value	5-fold cross validation
BB1	**-906.8**	**-894.8**	**0.07**	435.4
rot. BB1	-906.3	-894.3	0.00	**436.9**
Gumbel	-901.3	-895.3	**0.05**	431.5
BB6	-899.2	-887.2	0.03	434.5
BB8	-892.5	-880.5	0.00	428.4

For the direction southwest, only two statistically significant copula functions are found. The BB1 copula (lower and upper tail dependency) is selected as it has the highest 5-fold cross validation score while the CvM p-value is higher than 0.05. It is however not the best scoring copula, which is the rotated BB1 copula, but the difference is negligible.

Visible is that for each direction, one or multiple statistically significant parametric copulas are found. So, it is possible to describe the dependence structures for the variables wind speed and water level of the Lek for each direction using a parametric copula. The eventually chosen copula functions with their parameters can be found in Table 12.

Table 12: Selected copula functions with its parameters

Variables	Selected copula	Parameter 1	Parameter 2
Wind speed (NW) - water level Lek	BB8	2.09 (-)	0.94 (-)
Wind speed (W) - water level Lek	BB6	1.00 (-)	1.59 (-)
Wind speed (SW) - water level Lek	BB1	0.09 (-)	1.43 (-)

Validation by simulation

After the joint probability distributions are defined, it is possible to sample from those distributions and compare it to the observations as discussed in the methods. For the independent case, the same selected events are used to make a clear comparison based on the differences of determining probabilities. For this visual comparison, a sample of 5,000 draws is used. The result for wind speeds from the northwest direction is visible in Figure 20

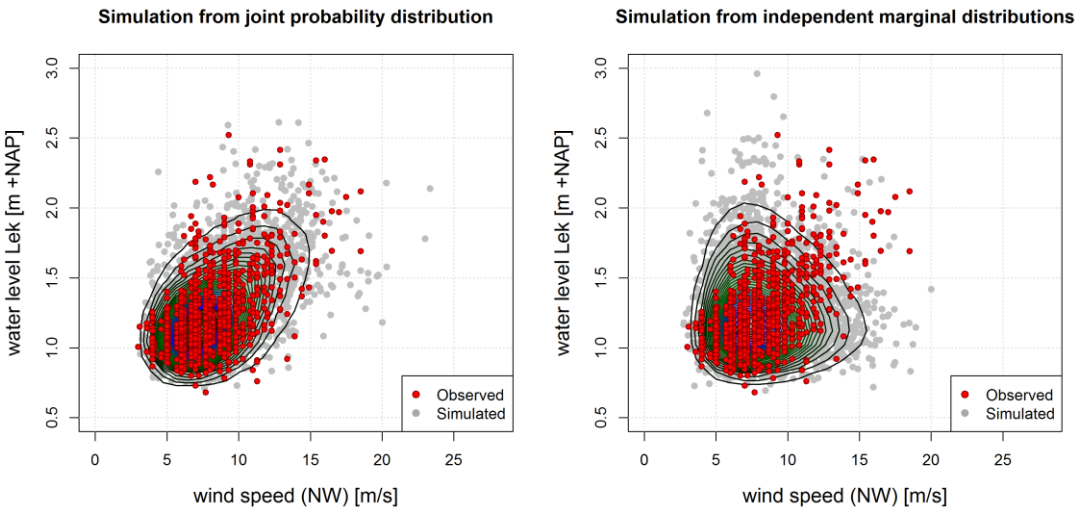


Figure 20: Simulation of stochastic variables (N=5,000) for wind speeds (NW) and water levels at the Lek, compared to the observations. Left figure is based on joint probability of those two variables. Right figure is based on assumed independence. Additionally, the contours of the densities (lines of equal density) based on simulated values are shown to visualise the behaviour in the main body of points (high to low density with colours purple to dark green).

It is clearly visible that the simulation of the stochastic variables from the joint probability distribution is closer to the observations compared to the simulation with assumed independence. It is visually validated that the selected marginal distributions and dependence structure are appropriate and represent the observations well. The comparison between the dependent and independent simulation shows that the probability of joint extremes is higher (relatively more dots) in the case of the copula approach as there is a clear dependence between the wind speed and water level at the Lek.

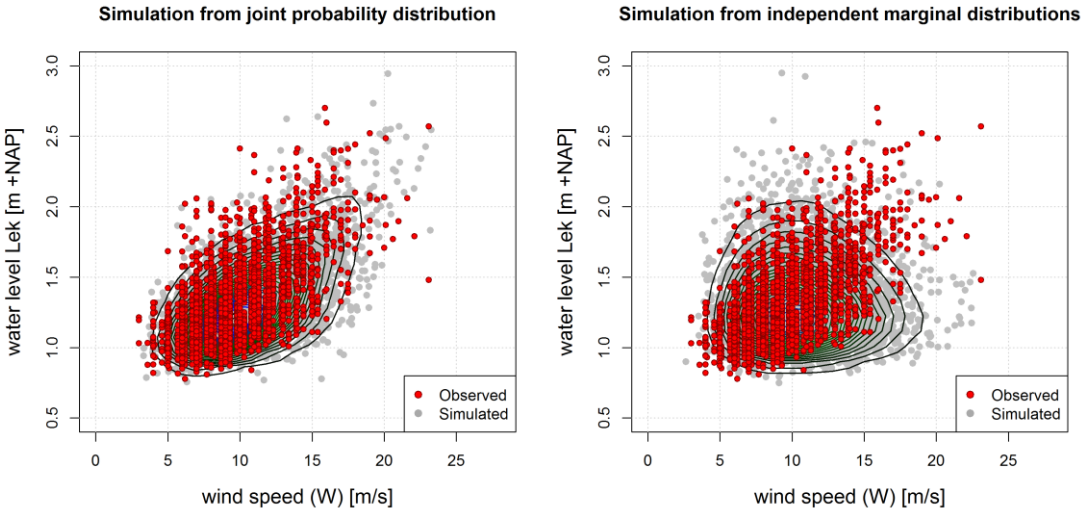


Figure 21: Simulation of stochastic variables (N=5,000) for wind speeds (W) and water levels at the Lek, compared to the observations. Left figure is based on joint probability of those two variables. Right figure is based on assumed independence. Additionally, the contours of the densities (lines of equal density) based on simulated values are shown to visualise the behaviour in the main body of points (high to low density with colours purple to dark green).

In Figure 21 and Figure 22, the comparison for wind speeds from west and southwest directions are visible. The figures show, similar to the previous figure, that the probability of joint extremes is higher in the dependent setting and that the simulation from the joint probability distributions is closer to the observations.

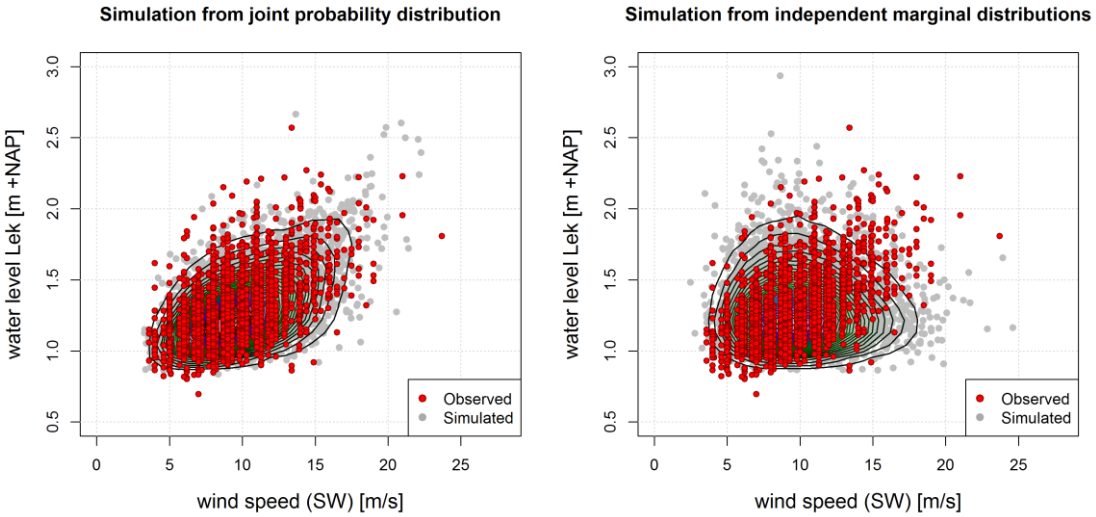


Figure 22: Simulation of stochastic variables (N=5,000) for wind speeds (SW) and water levels at the Lek, compared to the observations. Left figure is based on joint probability of those two variables. Right figure is based on assumed independence. Additionally, the contours of the densities (lines of equal density) based on simulated values are shown to visualise the behaviour in the main body of points (high to low density with colours purple to dark green).

Based on this validation by simulation, it can be said that the marginal distributions and copulas are properly able to capture the behaviour of the variables, better than in the independent setting. It shows that in general, the probability of joint extremes is higher when the dependency between the variables is accounted for as the density of dots is higher in the more extreme region (high values for both variables).

4.3.2 Combination two (Precipitation volume + water level Lek)

Marginal distributions

The first step was to determine the best fitting marginal distributions for the selected observations. A top 3 per variable is determined based on the AIC/BIC criterion. This is visible in Table 13.

Table 13: Candidate marginal distributions and resulting AIC/BIC criteria. Bold numbers are the best scores (lowest value).

Distribution	AIC	BIC
Precipitation volume		
Weibull	**4164.2**	**4176.8**
Gamma	4169.2	4181.8
Lognormal	4201.4	4214.0
Water level Lek		
Gumbel	**131.8**	**140.2**
Lognormal	149.8	158.2
Gen. Pareto	158.9	171.5

For precipitation volume, the Weibull distribution fits the data well based on the AIC/BIC scores and the visual comparison, which can be found in Appendix L. The Gamma and Lognormal distribution overestimate the probabilities of higher precipitation volumes, while the Weibull distribution is also fitting for the higher values. For water level Lek, the Gumbel distribution is the best scoring distribution. However, visible in the visual comparison is that the Gumbel distribution overestimates the probabilities of the higher values, which are critical in determining the normative water levels for this variable. Therefore, the second-best scoring distribution which fits the data well in the higher range, the Lognormal distribution is selected. The selected distributions and its parameters can be found in Table 14.

Table 14: Selected marginal distributions per variable with distribution parameters

Variable	Selected distribution	Parameter 1	Parameter 2
Precipitation volume	Weibull	scale = 27.63 mm	shape = 1.14 (-)
Water level Lek	Lognormal	scale = 0.42 (-)	shape = 0.19 (-)

Dependence structure

Next to the marginal distributions, the best fitting copula function for the observations is selected. In Table 15, the resulting scores can be found.

Table 15: Scoring of the top 5 copula functions for the variables precipitation volume and water level Lek. Bold scores are best score (AIC/BIC/Cross-validation) or non-rejected copulas (CvM p-value > 0.05)

Copula name	AIC	BIC	CvM p-value	5-fold cross validation
Frank	**-28.6**	**-24.4**	**0.29**	**13.8**
rot. BB8	-28.2	-19.8	**0.47**	11.8
Clayton	-27.5	-23.3	**0.30**	11.9
Normal/Gaussian	-27.0	-22.8	**0.54**	12.4
rot. Gumbel	-26.9	-22.7	**0.20**	12.1

In this case, it is clear that the Frank copula (main body dependency) is best scoring copula function. It has the best AIC, BIC and cross validation score, while it is also not rejected by the Cramér-von Mises test. Therefore, this copula function is selected with the following parameter, see Table 16. The Frank copula represents dependence throughout the whole range of values, so differences between the dependent and independent situation can be expected for the complete range of values.

Table 16: Selected copula function with its parameter

Combination	Selected copula	Parameter
Precipitation volume - water level Lek	Frank	1.53 (-)

Validation by simulation

From this joint probability distribution, a sample of 2000 values is drawn. The result can be found in Figure 23. The simulation from the joint probability distribution seems to fit the observations slightly better based on the contours, but the differences are small as there is not a strong dependency between the two variables such as in the case for wind speed and water level Lek. The main differences between the dependent and independent situations are found in the upper right area, which shows the higher values of both variables. The relative number of dots in this area is slightly higher in the dependent case. This subtle difference could however have an impact as the combination of higher precipitation volumes and a blockage is hazardous. It is however not clearly visible whether this is more in correspondence with the observations.

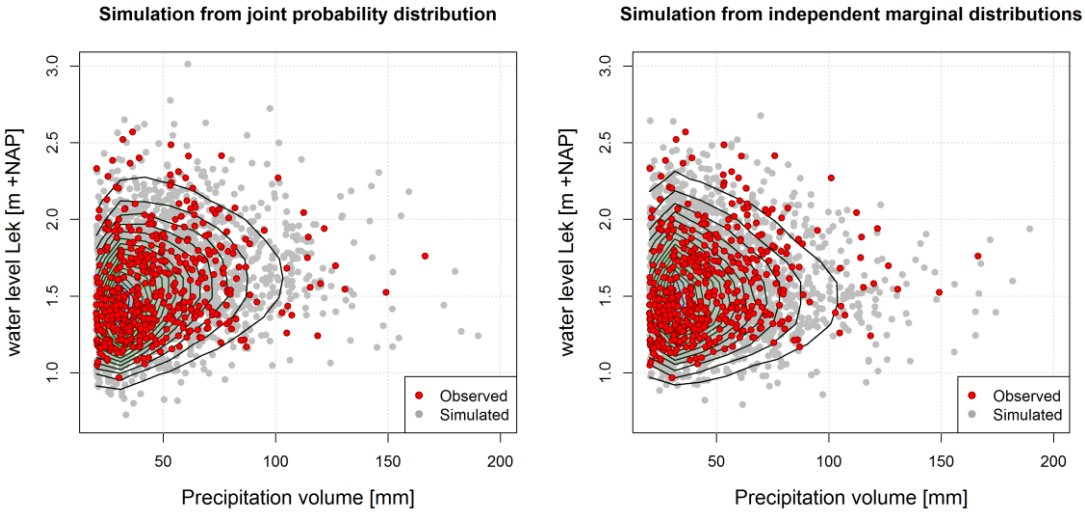


Figure 23: Simulation of stochastic variables (N=2,000) for the variables precipitation volume and water level Lek, compared to the observations. Left figure is based on joint probability of those two variables. Right figure is based on assumed independence. Additionally, the contours of the densities (lines of equal density) based on simulated values are shown to visualise the behaviour in the main body of points (high to low density with colours purple to dark green).

4.3.3 Combination three (Wind dir. & speed + precipitation volume)

Marginal distributions

The results of the AIC/BIC scores for the marginal distributions of the considered variables per wind direction can be found in Table 17. Only the three best scoring distributions are shown.

Table 17: Candidate marginal distributions and AIC/BIC scores. Bold numbers are the best scores (lowest value).

Distribution	AIC	BIC
Precipitation (NW)		
Lognormal	**322.9**	**327.8**
Gumbel	326.1	329.4
Gamma	328.6	331.9
Precipitation (W)		
Lognormal	1866.3	1876.4
Gumbel	1881.8	1888.5
Gamma	**1853.2**	**1863.3**
Precipitation (SW)		
Lognormal	1786.6	1796.6
Gumbel	1804.2	1810.8
Gamma	**1775.1**	**1785.1**
Precipitation (O)		
Lognormal	**233.4**	**237.4**
Gumbel	242.2	244.9
Weibull	237.7	241.7
Wind speed (NW)		
Lognormal	**189.7**	**193.0**
Gumbel	189.8	193.1
Gamma	190.3	193.5
Wind speed (W)		
Lognormal	1071.2	1077.9
Gumbel	1076.4	1083.2
Gamma	**1070.7**	**1077.4**
Wind speed (SW)		
Lognormal	1042.4	1049.1
Gumbel	1045.3	1051.9
Gamma	**1041.7**	**1048.4**

Additional to these scores, a visual comparison is made, which can be found in Appendix M. In some cases, the visual comparison led to another conclusion than the AIC/BIC scores. For precipitation volume, the scores and visual comparisons agree with each other. For the directions northwest and 'Other', the number of observations is low, which makes it hard to select an appropriate probability distribution. For wind speed, the Lognormal distribution is selected for each direction as it fits the data best. For the directions west and southwest, it is not the best scoring distribution, although the differences in scores are small. In general, the differences between the probability distributions are small, so the choices herein probably do not affect the results significantly.

Table 18: Selected marginal distributions per variable with distribution parameters

Variable	Selected distribution	Parameter 1	Parameter 2
Precipitation (NW)	Lognormal	scale = 3.01 (-)	shape = 0.71 (-)
Precipitation (W)	Gamma	scale = 20.66 mm	shape = 1.34 (-)
Precipitation (SW)	Gamma	scale = 20.70 mm	shape = 1.30 (-)
Precipitation (O)	Lognormal	scale = 2.42 (-)	shape = 1.24 (-)
Wind speed (NW)	Lognormal	scale = 2.32 (-)	shape = 0.27 (-)
Wind speed (W)	Lognormal	scale = 2.45 (-)	shape = 0.25 (-)
Wind speed (SW)	Lognormal	scale = 2.38 (-)	shape = 0.28 (-)

Dependence structure

In addition to the marginal distributions, the best fitting copula functions are selected for the pairs of variables per direction. In Table 19 the results for the direction NW can be found.

Table 19: Scoring of the top 5 copula functions for the variables precipitation volume and wind speed for direction NW. Bold scores are best score (AIC/BIC/Cross-validation) or non-rejected copulas (CvM p-value > 0.05).

Copula name	AIC	BIC	CvM p-value	5-fold cross validation
Frank	**-4.5**	**-2.9**	**0.86**	**2.73**
rot. Gumbel	-3.2	-1.6	**0.58**	1.71
Gaussian	-2.9	-1.2	**0.79**	1.01
Clayton	-2.7	-1.0	**0.45**	1.37
BB8	-2.4	0.8	**0.91**	-0.57

Visible is that each of the copula functions are able to significantly represent the dependence structure, which is probably a result of the low to moderate correlation between the two variables. All copula functions are quite similar when the dependence is low. The Frank copula (main body dependency) is selected as it has the lowest AIC/BIC score and highest 5-fold cross validation score. The result for direction west can be found in Table 20.

Table 20: Scoring of the top 5 copula functions for the variables precipitation volume and wind speed for direction W. Bold scores are best score (AIC/BIC/Cross-validation) or non-rejected copulas (CvM p-value > 0.05).

Copula name	AIC	BIC	CvM p-value	5-fold cross validation
Clayton	**-4.2**	**-0.8**	**0.96**	**3.10**
rot. Joe	-3.9	-0.5	**0.41**	1.42
rot. Gumbel	-3.8	-0.4	**0.76**	1.37
Gaussian	-3.1	0.3	**0.78**	1.39
Frank	-2.8	0.6	**0.84**	2.31

Again, all 5 copula functions are able to represent the dependence structure significantly. The Clayton copula (lower tail dependency) is selected as it has the lowest AIC/BIC scores and highest 5-fold cross-validation score. For direction southwest, the result can be found in Table 21.

Table 21: Scoring of the top 5 copula functions for the variables precipitation volume and wind speed for direction SW. Bold scores are best score (AIC/BIC/Cross-validation) or non-rejected copulas (CvM p-value > 0.05).

Copula name	AIC	BIC	CvM p-value	5-fold cross validation
rot. Joe	**-1.9**	**1.5**	**0.18**	0.63
rot. Gumbel	-0.5	2.8	0.03	0.77
Clayton	-0.3	3.0	**0.23**	-0.25
rot. BB8	0.1	6.8	0.02	0.93
rot. BB6	0.1	6.8	0.03	**1.54**

In this case, only two copula functions are able to significantly represent the dependence structure, the rotated Joe and rotated Gumbel copula functions. A choice between these two copula functions is made based on the 5-fold cross validation, which led to selecting the rotated Joe copula (lower tail dependency). This is also in agreement with the best AIC and BIC scores.

The selected copula functions for each direction can be found in Table 22. The selected copula functions: Frank, Clayton and rotated Joe copula represent mainly left-tail to main body dependence. This means that the interdependency is mainly found for lower to medium values and less for higher values.

Table 22: Selected copula function with its parameter

Variables	Selected copula	Parameter
Precipitation - wind speed (NW)	Frank	2.8 (-)
Precipitation - wind speed (W)	Clayton	0.2 (-)
Precipitation - wind speed (SW)	rot. Joe	1.11 (-)

Validation by simulation

In Figure 24 the result of the simulation of the stochastic variables for the direction NW is visible. The fitting process is based on a low number of observations, so there is a large uncertainty in the result. Based on these observations, the simulation from the joint probability distribution seems to fit better compared to the independent situation. Visible is that the combination of a high precipitation volume and low wind speed and vice versa is less probable in the dependent situation. Therefore, the probability that a wind speed that affects the water levels in the Alblasserwaard (>10 m/s) is reached during a substantial precipitation event, is higher in the dependent situation.

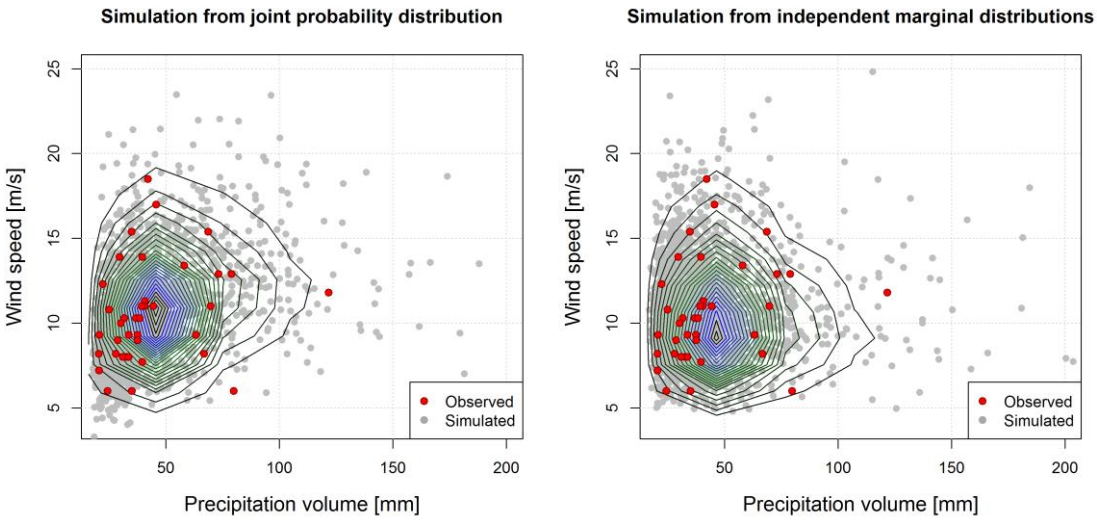


Figure 24: Simulation of stochastic variables (N=2,000) for the variables precipitation volume and wind speed for the direction NW, compared to the observations. Left figure is based on joint probability of those two variables. Right figure is based on assumed independence. Additionally, the contours of the densities (lines of equal density) based on simulated values are shown to visualise the behaviour in the main body of points (high to low density with colours blue to dark green).

For the wind direction west, the result can be found in Figure 25. The number of observations for this direction is higher, so the result is less uncertain. As the dependence between the two variables is low, it is not clearly visible whether the dependent simulation fits the observations better. However, it is visible that the wind speeds are generally slightly higher with higher precipitation volumes compared to the independent situation, so a small effect can be expected for the normative water levels.

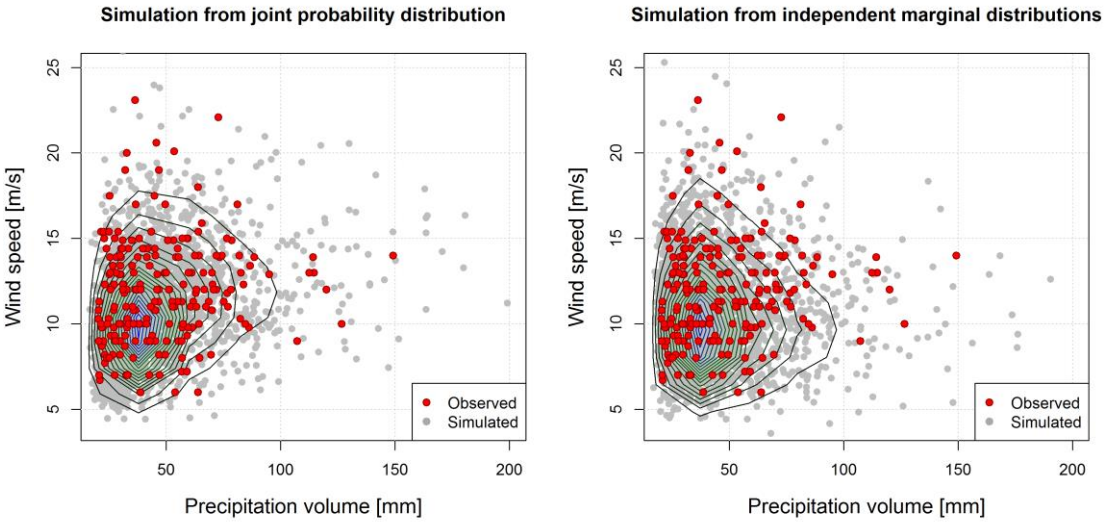


Figure 25: Simulation of stochastic variables (N=2,000) for the variables precipitation volume and wind speed for the direction W, compared to the observations. Left figure is based on joint probability of those two variables. Right figure is based on assumed independence. Additionally, the contours of the densities (lines of equal density) based on simulated values are shown to visualise the behaviour in the main body of points (high to low density with colours blue to dark green).

For the southwest direction, the result can be found in Figure 26. The conclusion is similar to the conclusion of the west direction. The dependency is small which makes the comparison less clear. The simulation from the joint probability distribution seems to fit the observations slightly better as there are less dots in the area of high precipitation volume and low wind speed. Again, the number of dots in the area of higher precipitation volumes and higher wind speeds is slightly higher in the dependent situation, so a small effect on the normative water levels is expected.

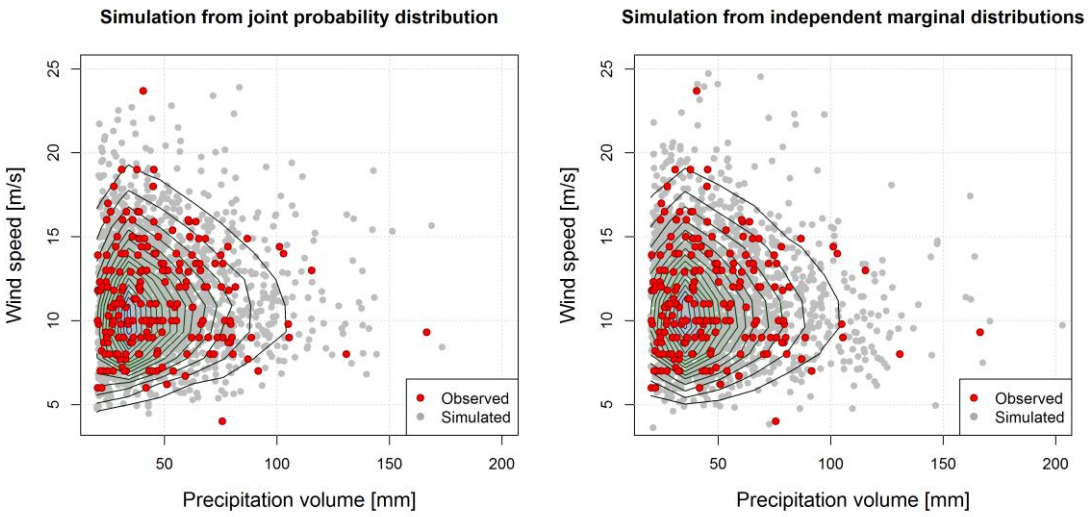


Figure 26: Simulation of stochastic variables (N=2,000) for the variables precipitation volume and wind speed for the direction SW, compared to the observations. Left figure is based on joint probability of those two variables. Right figure is based on assumed independence. Additionally, the contours of the densities (lines of equal density) based on simulated values are shown to visualise the behaviour in the main body of points (high to low density with colours blue to dark green).

So, in general, the differences between the dependent and independent situation are small as the dependence between precipitation volume and wind speed is low. In addition, the dependency is mainly found in the left tail and main body of the data.

4.3.4 Combination four (Wind dir. & speed + precipitation volume + water level Lek)

Marginal distributions

The first step to construct the vine copula is to select the appropriate marginal distributions, using the selected events. For the variables wind speed and precipitation volume, the marginals distributions per wind direction are already selected in combination three. The marginals distributions of the variable water level Lek per wind direction are selected and discussed in Appendix N. The selected marginal distributions can be found in Table 23.

Table 23: Selected marginal distributions per variable with its distribution parameters

Variable	Selected distribution	Parameter 1	Parameter 2
Precipitation (NW)	Lognormal	scale = 3.01 (-)	shape = 0.71 (-)
Precipitation (W)	Gamma	scale = 20.66 mm	shape = 1.34 (-)
Precipitation (SW)	Gamma	scale = 20.70 mm	shape = 1.30 (-)
Precipitation (O)	Lognormal	scale = 2.42 (-)	shape = 1.24 (-)
Wind speed (NW)	Lognormal	scale = 2.32 (-)	shape = 0.27 (-)
Wind speed (W)	Lognormal	scale = 2.45 (-)	shape = 0.25 (-)
Wind speed (SW)	Lognormal	scale = 2.38 (-)	shape = 0.28 (-)
Water level Lek (NW)	Lognormal	scale = 0.41 (-)	shape = 0.19 (-)
Water level Lek (W)	Lognormal	scale = 0.44 (-)	shape = 0.19 (-)
Water level Lek (SW)	Gumbel	location = 1.41 m	scale = 0.23 (-)
Water level Lek (O)	Lognormal	scale = 0.30 (-)	shape = 0.13 (-)

Vine tree structure

The next parallel step is to select the vine tree structure which consists of the tree structure and its bivariate copula functions. All three optional tree structures including its fitted bivariate copula functions are tested with AIC and BIC. In addition, its statistical significance at a 5% level is tested using the Cramér-von Mises goodness-of-fit test. The resulting scores for the direction NW can be found in Table 24.

Table 24: Selecting the vine tree structure for the direction NW. Best AIC/BIC scores and non-rejected (CvM > 0.05) vine tree structures are in bold.

Vine copula (NW)	AIC	BIC	CvM
Vine tree 1	-19.9	-15.0	**0.86**
Vine tree 2	-18.3	-13.4	**0.44**
Vine tree 3	**-22.5**	**-17.5**	**0.87**

Visible is that each tree structure is able to significantly capture the dependency in the data. The selected tree structure (tree 3) is based on the lowest AIC/BIC score. Between the best and second-best vine tree AIC/BIC score is quite a large gap, so vine tree 3 is the best option by quite a margin. The result for direction W can be found in Table 25.

Table 25: Selecting the vine tree structure for the direction W. Best AIC/BIC scores and non-rejected (CvM > 0.05) vine tree structures are in bold.

Vine copula (W)	AIC	BIC	CvM
Vine tree 1	-125.6	-115.5	**0.21**
Vine tree 2	-126.5	-116.4	**0.76**
Vine tree 3	**-126.8**	**-116.7**	**0.06**

Again, all three tree structures are not rejected and able to capture the interdependency in the data. The differences between the three in terms of AIC/BIC scores is very small, so choices herein do not have a large impact on the outcome (tested in visual validation). Vine tree 3 is selected based on the lowest score. In Table 26, the result for direction SW can be found.

Table 26: Selecting the vine tree structure for the direction SW. Best AIC/BIC scores and non-rejected (CvM > 0.05) vine tree structures are in bold.

Vine copula (SW)	AIC	BIC	CvM
Vine tree 1	**-74.4**	**-64.4**	**0.08**
Vine tree 2	-70.3	-60.3	0.02
Vine tree 3	-71.2	-62.0	**0.88**

In this case, vine tree 2 is rejected, while the other two are accepted. Out of these two options, vine tree 1 has the best AIC/BIC score by a relatively large margin. Therefore, this tree structure is selected for the direction SW. Notable is that the direction NW and W have the same tree structure, while the direction SW has another tree structure as best option. For the direction "Other", a bivariate copula between precipitation volume and water level Lek is fitted as there is no interest in wind speed from this direction (not relevant, only affecting the water levels in the Alblasserwaard with wind speeds > 21 m/s occurring once in 3543 years (Arcadis, 2014)). This resulted in a Frank copula as best option. The resulting selected copula functions per wind direction, its parameters and tree structures can be found in Table 27.

Table 27: Selected copula functions for the four wind directions and its copula parameters.

Direction	Tree structure and copula functions (parameter)
Northwest	C_{WP}: Frank (par = 2.8) - C_{PL}: Clayton (par = 0.75) $C_{WL\|P}$: Frank (par = 4.2)
West	C_{WP}: Clayton (par = 0.25) - C_{PL}: Clayton (par = 0.12) $C_{WL\|P}$: Gaussian (par = 0.61)
Southwest	C_{PW}: rot. Joe (par = 1.04) - C_{WL}: Gaussian (par = 0.48) $C_{PL\|W}$: Frank (par = 1.16)
Other	C_{PL}: Frank (par = -0.52)

Validation by simulation

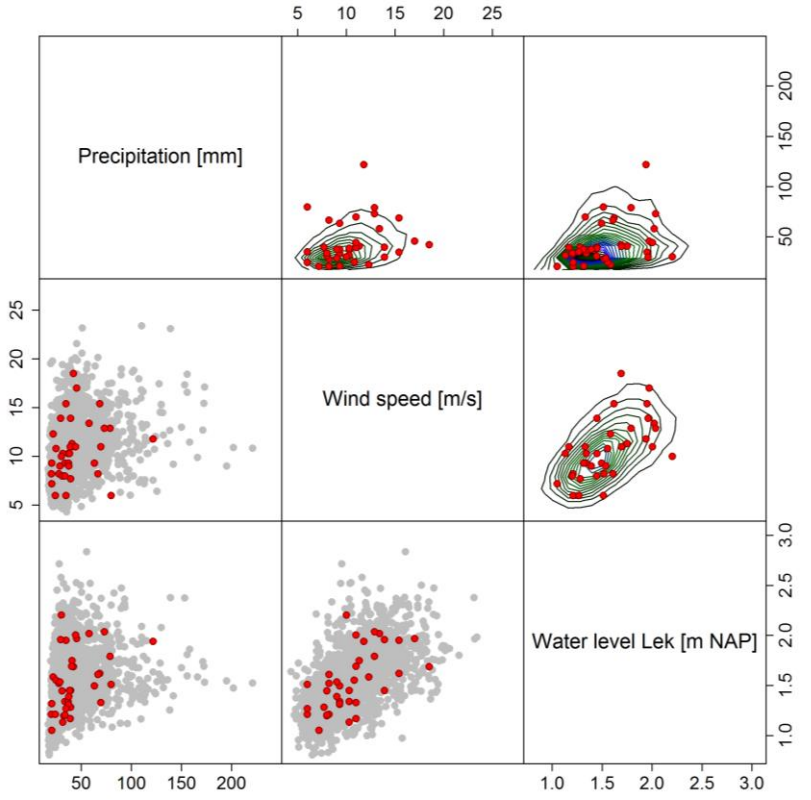
To check whether the combination of marginal distributions and vine copula structures are selected properly, a simulation from the vine copula is performed. Simulation (N =2,000) from the vine copula is done using the *RVineSim()* function (Schepsmeier et al., 2018). In addition, a simulation is performed where independence between the variables is assumed, to check what the effect of dependency on the simulation is. In Figure 27 the result for the direction northwest can be found. The results for directions west and southwest can be found in Appendix O.

Visible for the direction northwest is that the simulation from the vine copula is much closer to the observations than the independent simulation. For all three combinations of variables is quite a large difference visible between the dependent and independent situation. Furthermore, it is visible that in the upper right corner of the graphs, relatively more dots are present, which means that the probability of joint extremes is higher (based on this simulation) in the dependent situation. The results are similar to the bivariate combination studies.

For the directions west and southwest, similar results are found. Although, the interdependencies between precipitation and wind speed and precipitation and water level Lek are low for these directions. The clearest differences are found for the combination of wind speed and water level Lek for both directions. These results are also similar to the bivariate combinations.

In general, the vine tree structures and the marginal distributions seem to capture the behaviour of the data very well, which makes it appropriate to use to determine the joint probabilities. As the density of the dots is in most cases relatively higher in the upper right corners of the graphs, the probabilities of joint extremes seem to be increased relative to the independent situation based on these simulations. Therefore, it is reasonable to expect that the normative water levels will be in some way affected, considering the differences between the dependent and independent situations. The results are similar to the bivariate combinations, which is an additional confirmation that the vine copula structures are able to capture the dependencies well.

Dependent



Independent

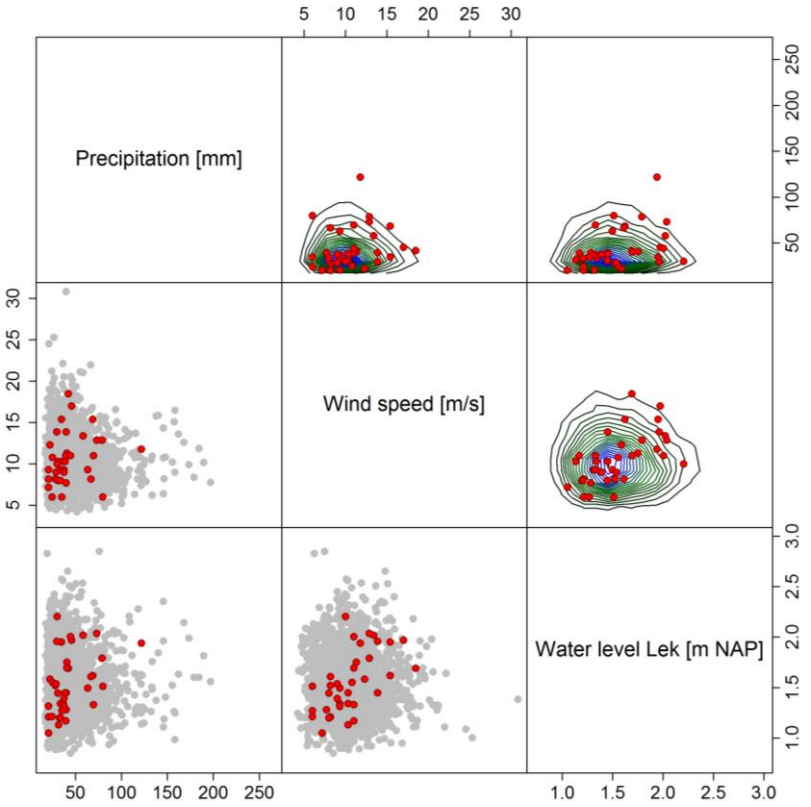


Figure 27: Comparison dependent (above) and assumed independent (below) simulation (N = 2,000) compared to observed values for the direction NW. Lower left half of each figure is the simulation with points, upper right half are the density contours based on simulated values. Note that each half is mirrored to match the axes.

4.4 Conclusion

In this chapter, research question 2: "*What are the most suitable joint probability distributions describing each of the combinations of interdependent stochastic variables, and what is the effect of including dependency on the joint probabilities of the stochastic variables?"* is answered. For each combination, a joint probability distribution is constructed using a bivariate or vine copula. In all cases, it was possible to construct a statistically significant dependence structure and find fitting marginal distributions. In addition, it is concluded that it is feasible to construct these joint probability distributions with the available *R software* packages and an acceptable amount of time and basic knowledge of copulas. So, the used available dependence structures and marginal distributions are sufficient. In addition, from each joint probability distribution, a simulation is performed and compared to the observations. These analyses show that the joint probability distributions were able to represent the observations well, and better compared to the independent simulations.

The clearest differences are found for the combinations of wind speed and water level Lek, as the dependency between those variables is the strongest. The other combinations also showed some differences, but less clear. The general conclusions for all the combinations are that the dependent simulation is relatively closer to the observations than the independent simulation and that the probability of combined extremes is generally higher in the dependent situation relative to the independent situation. Whether this conclusion affects the normative water levels, is analysed in Chapter 5.

5 Normative water levels

5.1 Introduction

In this chapter, RQ3 is discussed: *"What is the effect of including dependencies between the stochastic variables on the normative water levels in comparison to assuming independence between these variables and how can these differences be explained?"*. First, the methods on how the normative water levels are derived are discussed in section 5.2. Secondly, the results are discussed per combination of dependent variables in section 5.3. The results of each combination are compared to a reference case, in which the probabilities are determined based on assumed independence. This is done to show the effect of including dependencies on the normative water levels and to answer the research question. It is also discussed how these differences in normative water levels can be explained. In section 5.4, an overall conclusion on the results in this chapter is given.

5.2 Methods

The methods are divided in four sections. First an introduction to the model and input variables is given in section 5.2.1. Then, the derivation of the probabilities for the reference case and the dependent cases is discussed in respectively section 5.2.2 and 5.2.3. The processing of the model results is discussed in section 5.2.4.

5.2.1 Introduction to the model and input variables

To derive normative water levels for the Alblasserwaard, a model of the regional water system is used. The model is a one-dimensional (1D) SOBEK model with modules 1DFLOW (Rural) and RR (Rainfall-Runoff), to simulate hydrodynamic and rainfall-runoff processes, respectively, the functioning of weirs and pumps, paved and unpaved runoff areas, cross-sections/bed friction numbers of rivers/canals and other objects (HydroLogic, 2010). Water flow is computed by solving the Saint Venant equations for unsteady flow in one dimension (Deltares, 2019). Rainfall is distributed uniformly over the entire area. For each rainfall-runoff node in the model, the total area, land use, surface level, soil type and initial groundwater level is defined. The drainage from unpaved/rural area towards open water is modelled using the Ernst drainage formulation. More details about the functioning of the SOBEK model can be found in the User Manual of SOBEK (Deltares, 2019). The SOBEK model has been used in previous studies and is continuously developed based on modifications in the area (HydroLogic, 2010; Arcadis, 2015). Calibration and verification steps have been conducted on the 2010 version of the model, mainly focussing on calibrating the effects of wind by testing different roughness heights and comparing it to observations (HydroLogic, 2010). The version of the model used in this study is the reference model of Infram, Strootman & HydroLogic (2017), in which the effects of potential measures for 2050 are compared to the reference situation. The only implemented difference is the upgrade of

the temporary pump in the Overwaard area to 25 m^3/s (an ongoing development), which makes the Overwaard area independent of the water level of the Lek.

The stochastic method is used to determine the normative water levels. This method is based on probability distributions assigned to each of the variables that affect the normative water levels. The variables are discretised in classes and then all combinations (events with known probability) of these discretised classes are used as input in the model. The derivation of the resulting return periods of water levels is further described in section 5.1.3. The stochastic variables used in this study are:

- Precipitation
    - Volume
    - Distribution over time
- Wind
    - Speed
    - Direction
- External water level Lek
- Initial groundwater levels (fixed)

Following the results of the first research question in Chapter 3, the variables of which the probabilities are determined in a jointly matter are divided over four combinations. For all these combinations, the same number of events and the same discretisation of the variables are used, so that the only difference between them is the method to determine the probabilities for the events based on which variables are considered as interdependent. Otherwise, it would not be clear if the computed normative water levels are different as a result of the discretisation and number of events or the derived probabilities. In addition, this allows that the SOBEK model is only run once, as the probabilities are assigned to the events afterwards. The results of each combination are compared to the independent reference case, where the probabilities are derived by multiplication of the individual probabilities as independence is assumed. This independent case also uses the same discretisation. The derivation of the probabilities of the reference case and the dependent cases are shortly discussed in sections 5.2.2 and 5.2.3.

The timing of the wind speed compared to blockage of the pumping station in the model is important as it affects the water levels. This timing is determined based on results of the exploration of dependencies in which the cross-correlations for different lags are determined. Herein the strongest correlation for the wind speed is found at the middle of a 9-day precipitation event, while the strongest correlation between the water level of the Lek relative to the precipitation and wind speed was found about one day later. Therefore, the peak of the wind speed is modelled in the middle of the 9-day period, while the blockage is modelled one day after this peak. The derivation of the wind event shape based on the maximum wind speed can be found in Appendix I. Initial groundwater levels are set at the average groundwater level (GG) as Arcadis (2014) showed that using the initial groundwater levels as stochastic variable has a negligible effect on the normative water levels. As the water management of the Alblasserwaard is based on polder pumping stations that are activated based on a water level above the local threshold, which pump into the reservoir canals with a year-round fixed water level, it is not necessary to divide the

probabilities into summer and winter settings. The water system functions similarly in winter and summer. Again, the same choices are made in both the dependent case and the independent case to prevent differences due to choices herein.

To determine the duration of each event in the model, some test simulations are performed. In the most extreme case (210 mm with 2 peaks (pattern G)), the maximum water level in some branches is reached just after 9 days. Therefore, it is decided that each event simulation has a duration of 10 days to be certain that the maximum water level as a result of the precipitation event is included.

The discretisation of the stochastic variables is based on a range of approximately T0.5 to T1000, as once in 1000 years is the most strict frequency of occurrence for regional flood defences (STOWA, 2015b). For the distribution of the precipitation over time, the 7 shapes as defined in STOWA (2004) are used, see Appendix G. The uniform and 1-peak 12.5% shapes are combined as one shape, as they are almost identical and have the same effect on the water levels (HydroLogic, 2018b). In Table 28, each precipitation shape is described with the letters A to G, referring to Appendix G. Wind direction and speed is discretised in such a way that only the relevant directions and speeds (affecting both the Lek and the water levels in the Alblasserwaard) are used (Arcadis, 2014; HydroLogic, 2018b). This means that only wind speeds >10 m/s from northwest to southwest are discretised, while lower wind speeds from these directions and all wind speeds from other directions are combined in one class. Winds from the direction east only impact the water levels in the Alblasserwaard at higher return periods that are not relevant as higher wind speeds are uncommon from this direction (Arcadis, 2014). The discretisation for the water level Lek scenarios is derived in Appendix J, which resulted in three scenarios; no blockage, 2 hours blockage and 4 hours blockage. This is based on the relationship between the maximum water level and the duration of the blockage. The derivation of the probabilities of each event is shortly described per combination in sections 5.2.2 and 5.2.3. The probabilities for precipitation shapes can be found in Table 31 in Appendix G, which are equal in both situations, dependent and independent. As visible in Table 28, the total number of events is 12 x 6 x 10 x 3 = 2160.

Table 28: Discretisation of the stochastic variables

Precipitation volume (12 classes)	Precipitation 'shape' (6 classes)	Wind direction + speed (10 classes)	Scenario external water level (3 classes)
65 - 85 mm	A/B	Northwest 12 m/s	No blockage
85 - 105 mm	C	Northwest 17 m/s	2h blockage
105 - 125 mm	D	Northwest 22 m/s	4h blockage
125 - 135 mm	E	West 12 m/s	
135 - 145 mm	F	West 17 m/s	
145 - 155 mm	G	West 22 m/s	
155 - 165 mm		Southwest 12 m/s	
165 - 175 mm		Southwest 17 m/s	
175 - 185 mm		Southwest 22 m/s	
185 - 195 mm		Others (no wind)	
195 - 205 mm			
205 - 215 mm			

5.2.2 Independent/reference case

The focus of this study is on determining joint probabilities, but to compare it to the reference situation, also the independent probabilities are determined. The probabilities for the independent case are determined by multiplying the probabilities of each stochastic variable for all events:

$$P(i) = P_p(i) * P_s(i) * P_w(i) * P_l(i) \qquad \textit{equation (5)}$$

where P_p is the probability of precipitation, P_s the probability of precipitation shape, P_w the probability of wind direction and speed and P_l the probability of the water level Lek scenario. These probabilities are all based on a 9-day period. You could argue that for wind and blockage of the pumping station, not all 9 days are relevant, but this depends on location and the distribution of the precipitation over time (e.g. uniform, one-peak or two-peak). Therefore, the conservative approach of using the probability of occurrence of wind and blockage somewhere in the 9 days is taken. To derive the probability in a year, which is relevant to determine the return period, the probability in 9 days is converted to the probability in a year. With the assumption of independence between consecutive 9-day periods, the probability in a year can be determined by subtracting the probability of no event from 1 and multiplying it by the number of 9-day periods in a year (Rumsey, 2006):

$$P_{year}(i) = 1 - \ (1 - \ P_{9\ days}(i))^N \qquad \textit{equation (6)}$$

where N stands for the average number of 9-day periods in a year (N = 365.25/9). The probabilities for precipitation shapes can found in Table 31. The univariate probabilities of precipitation volumes are derived in Appendix E using a Peak over Threshold approach. The probabilities of maximum wind speed and accompanying direction in 9 days are derived in Appendix H. The probabilities for the Lek scenarios (No blockage, 2h blockage, 4h blockage) are derived in Appendix J. These probabilities are also (partly used) for the other combinations where not all probabilities of every stochastic variable are determined in a jointly matter.

5.2.3 Dependent cases

In the dependent cases, the probabilities of different combinations of stochastic variables are determined in a jointly matter. These probabilities are derived from the joint probability distributions based on the copula functions. To derive the probabilities within a certain interval from a copula function, the following general equation is used (Nelsen, n.d.):

$$P(a \le U \le b, c \le V \le d) = \big(C(b,d) - C(a,d)\big) - \big(C(b,c) - C(a,c)\big) \qquad \textit{equation (7)}$$

where a, b, c and d are the boundaries of the interval, and C is the copula function. In *R*, the probabilities are derived using the *pMvdc()* function. This probability is then multiplied with the other assumed independent probabilities to determine the probability of an event in 9 days. Similarly to the reference case, all probabilities are determined as the probability of occurrence in 9 days and then converted to probability in a year.

For the first combination, the probabilities of wind speed and water level Lek are derived in a jointly matter per wind direction. The probability of an event is determined using the univariate probabilities for precipitation volume and shape, and the joint probabilities of wind speed, direction and water level of the Lek. The probabilities of wind speed and Lek scenario are divided per wind direction (estimated with fraction of selected events in that direction) and converted to probability of occurrence in 9 days:

$$P_{wl}(i) = 1 - (1 - P(a \leq w \leq b, c \leq l \leq d))^N = 1 - (1 - (C(b,d) - C(a,d) - C(b,c) + C(a,c)))^N \quad \textit{equation (8)}$$

where P_{wl} is the joint probability of wind (w) and Lek scenario (l), C stands for the copula function and N stands for a factor including the adjustment for wind direction (fraction of events in that direction) and the occurrence in 9 days. For the wind category 'Others' in combination with a blockage, the marginal distribution of water level Lek (O) and the copula functions for the northwest to southwest directions with wind speeds lower than 10 m/s are used to determine the probabilities. For the combinations two and three, the same approach is taken (equation 8) using the joint probability distributions of the relevant variables. However, extrapolation to 9 days are for these combinations not required as the events in the data are already coupled within an appropriate time frame of the precipitation event.

For combination four, the same approach is taken, but there is no function available in *VineCopula* to determine the probabilities of the values to be in a certain range. This is solved by performing a simulation from the vine copula with N = $1*10^8$ and transforming these values between 0 and 1 to their real values using the marginal distributions. Then, probabilities are derived by determining the fraction of the draws to be within a certain range. This method is tested on the bivariate combinations and gave similar results to the *pMvdc()* function as used for the other combinations. In addition, the number of draws is sufficient as multiple simulations (with other random number seeds) gave the same results.

5.2.4 Processing model results

After the model simulation of all events, the maximum water levels of all reaches in the complete area per event are known. From this data, the frequency curves for all the reaches can be determined as the probability of every event is known. Using the stochastic method, it is assumed that the sum of the probabilities (of the input events) above a water level is equal to the probability of that water level (Bosch et al., 2006). This process is illustrated in Figure 28. This involves the following steps. First, all events above a certain water level are selected. Secondly, the sum of the probabilities of all events above this water level is taken. Lastly, this process is iterated over a range of water levels (minimum to maximum) to derive the probabilities for all water levels in that section. This is then repeated for all the sections. The return period in years is then determined by the inverse of the probability in a year. A return period of for example 100 years or a T100 event, means that that specific event takes on average place once in 100 years.

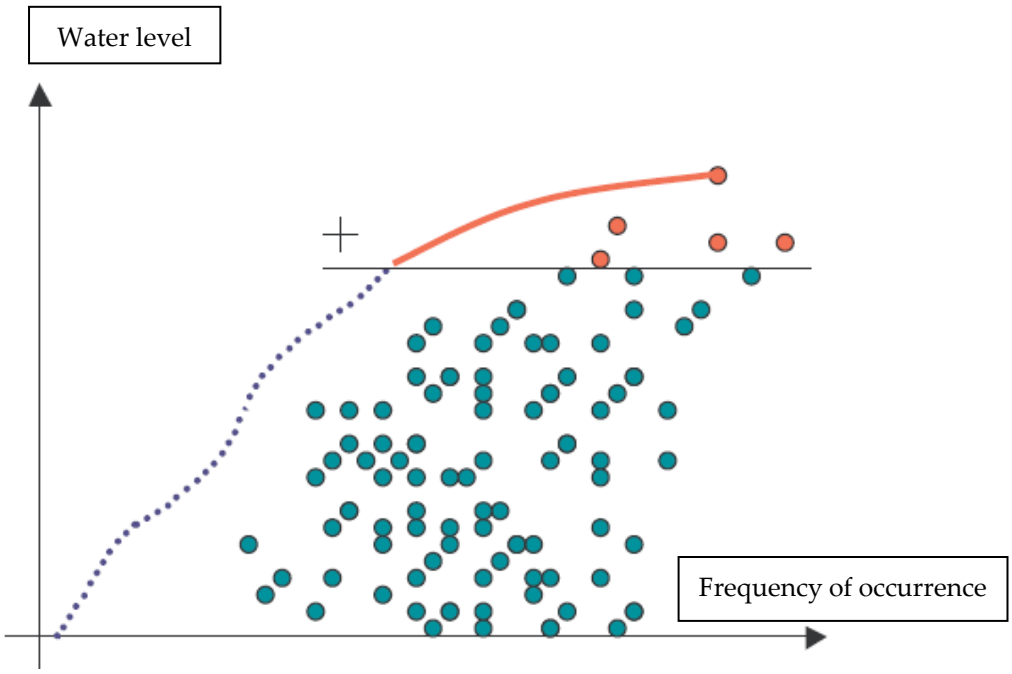


Figure 28: Illustration of determining frequency of occurrence using the stochastic method (Bossenbroek, 2004). For each water level, the sum of the probabilities of the events above this level are taken, which is assumed to represent the probability of that water level.

The resulting normative water levels are then analysed in two ways. First, the differences in normative water levels between the dependent and independent case are plotted over a map of the Alblasserwaard. Then, for specific locations which show a notable result, for example the highest difference, the frequency curves are presented. This is used to analyse whether the differences in water levels are equal for all return periods or differ over the range of return periods. The most relevant return periods in the Alblasserwaard are T100 and T300, as these are the current norms for regional flood defences in the area, see Appendix T. The frequency curves include the effect of the uncertainty introduced by the copula with confidence intervals at the 95% significance level. Only the uncertainty in the dependency is quantified as the focus of this study is on the effect of interdependencies, not on quantifying all the uncertainties involved in determining normative water levels.

5.3 Results

The results, answering research question 3, are discussed per combination of dependent variables in the following sections.

5.3.1 Combination one (Wind direction & speed + water level Lek)

In the first combination, the variables wind speed (per direction) and water level of the Lek are considered dependent. High water levels at the Lek cause that the water from the Nederwaard area (southwestern part) cannot be discharged. Wind speeds from westerly directions affect both the Lek and the water levels in the Alblasserwaard. The resulting differences in normative water levels relative to the independent case for a return period of 300 years are visible in Figure 29. This return period is chosen as for lower return periods, there was not a significant difference visible.

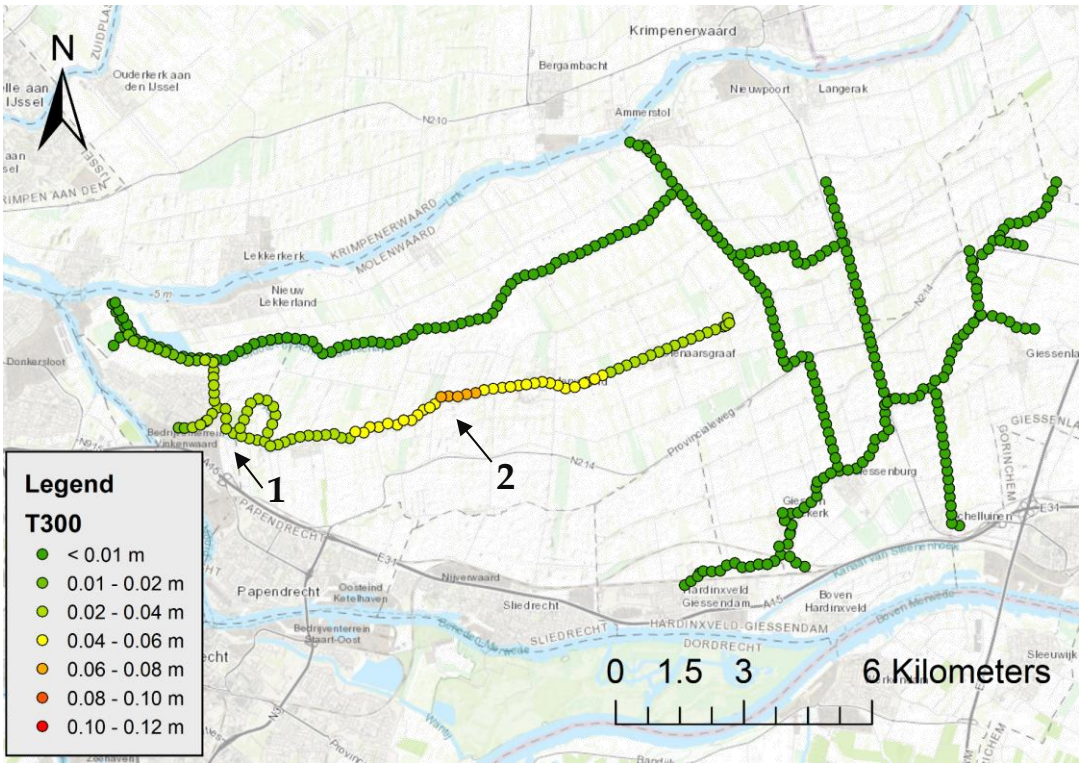


Figure 29: Normative water levels relative to the independent case at T300 (occurring on average once in 300 years). Location 1 = WL_36_2 and location 2 = WL_40_5. A positive difference means that in the dependent situation the normative water levels are higher compared to the reference case.

Visible is that especially in the middle section of the Nederwaard area (around location 2), the normative water levels are up to about 6 cm higher relative to the independent case. The Overwaard area is not affected, as this area is discharged using a temporary pump during high water levels on the Lek. Due to the higher joint probability of winds from westerly directions and a blockage of the pumping station, the situation where the water is displaced in eastward direction while no water is discharged occurs more frequently. To explain this result, the effects of wind and blockages on the Nederwaard area are visualised and discussed in Appendix P. Concluded is that the main effect of a 4-hour blockage without wind reaches no further than about location 2, while only wind and no blockage mainly affects east of location 2. The spatial distribution of the result in Figure 29 is a consequence of the situation that the middle section of the Nederwaard (around location 2) is the most

affected by the combination of wind and a blockage compared to only wind or only a blockage. So, only in the case of winds from the west during a blockage, which has a higher joint probability in the dependent case, the effect of the blockage reaches location 2. As the wind has almost no impact close to the main pumping station, there are no differences in normative water levels visible there.

At locations 1 and 2, the frequency curves are presented in Figure 30 and Figure 31 respectively. Visible at location 1 is that the normative water level is significantly different around T50 and above T200. The effects of a blockage of 2 hours and 4 hours is clearly visible in the graph. The maximum difference at a return period of 1000 years is about 3 cm.

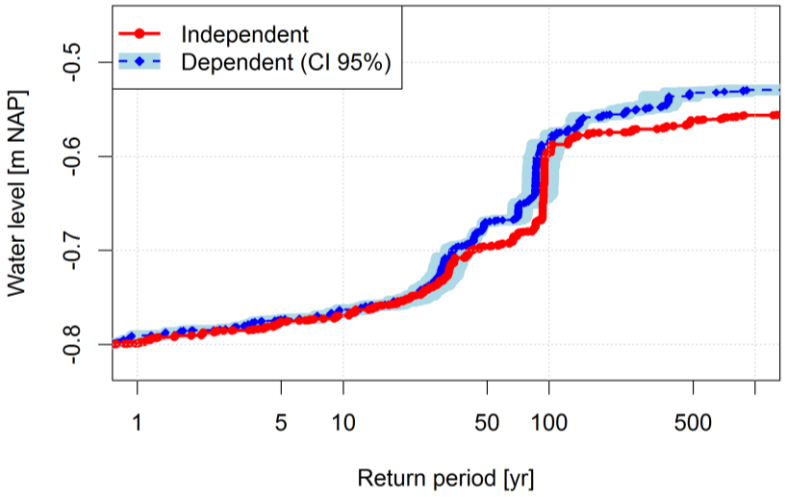


Figure 30: Water level frequency curve at location 1 for the dependent and independent case. The dependent curve includes the uncertainty introduced by the copula.

Location 2 shows the largest difference between the dependent and independent case, as this area is sensitive to both the blockage and wind speeds from westerly directions. Clearly visible is that the effect of the blockage scenarios is less pronounced at this location compared to location 1 as the distance from the pumping station is larger. Significant differences between the dependent and independent case start at around T150, with a maximum difference of about 7 cm around T500.

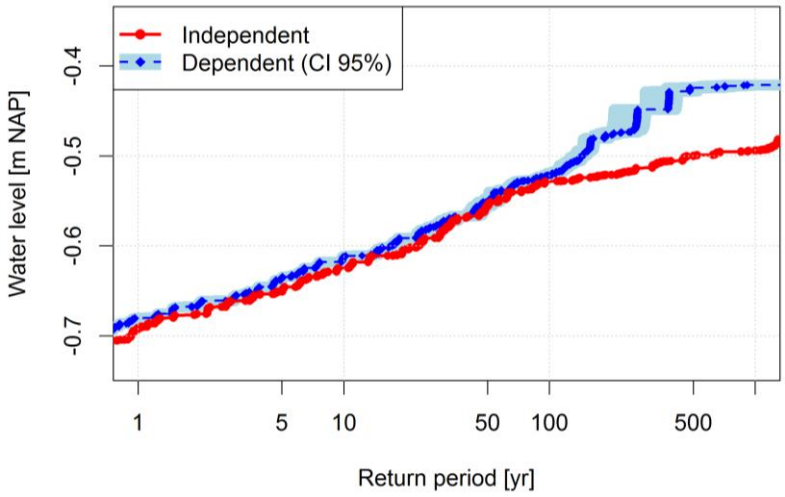


Figure 31: Water level frequency curve at location 2 for the dependent and independent case. The dependent curve includes the uncertainty introduced by the copula.

In conclusion, the dependency between wind and the water level of the Lek has a significant effect up to about 7 centimetres at higher return periods at locations where the wind and blockage both affect the water levels significantly. Even considering the confidence intervals, the difference is clear for those locations in the Nederwaard area at higher return periods. For lower return periods (< T100), the effect is negligible (maximum about 2 centimetres).

5.3.2 Combination two (Precipitation volume + water level Lek)

In the second combination, the variables precipitation volume and water level Lek are considered as dependent. In Figure 32 the relative difference between the dependent and independent situation for a return period of 25 years is shown. At lower return periods, there is no significant effect as a blockage in combination with a precipitation event does not occur frequently. Visible is that locally, close to the pumping station, the effect in the Nederwaard area is large (> 10 cm). However, close to location 4 and eastward, the effect is negligible. The Overwaard area is not affected by a blockage of the pumping station.

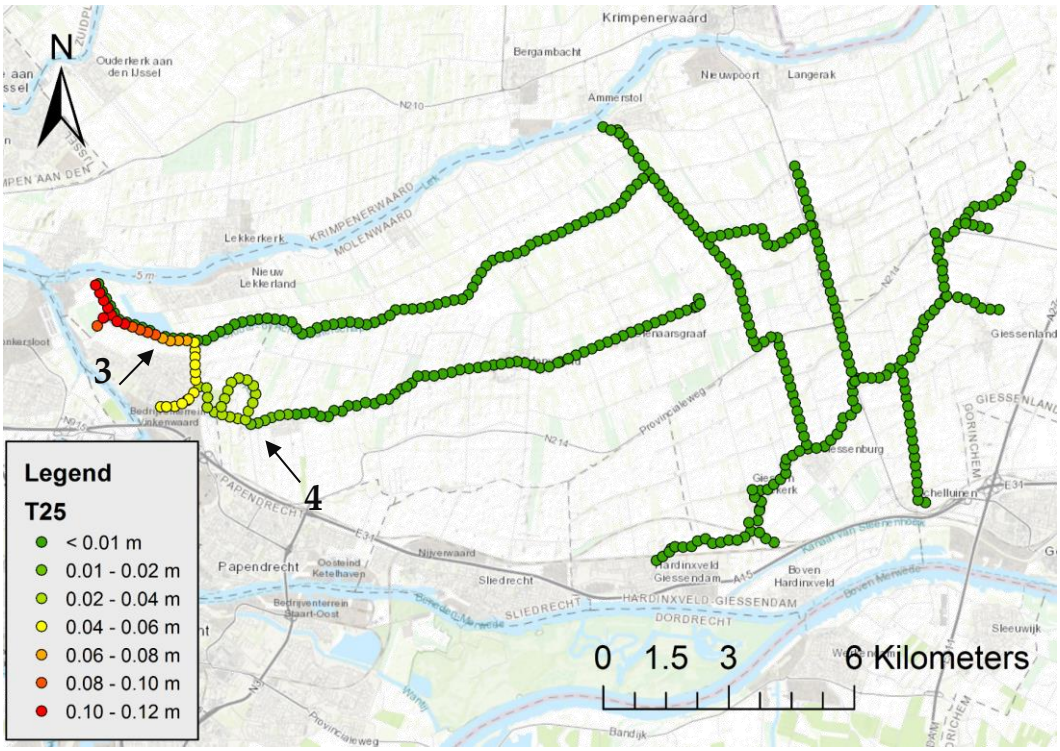


Figure 32: Normative water levels relative to independent case at T25. Location 3 = WL_33_6 and location 4 = 74_2. A positive difference means that in the dependent situation, the normative water levels are higher compared to the reference case.

This spatial distribution is not equal for every return period. In Figure 33 the same comparison is made for a return period of 100 years. Visible is how the effect of the dependency is extended relative to the pumping station in comparison to the return period of 25 years. However, the difference in the section close to the pumping station is not increased. To explain this result, the effect of the combination of precipitation and a blockage is analysed and discussed in Appendix Q. Concluded is how an increasing precipitation volume does not have a large impact close to the pumping station during a blockage, as these water levels are mainly determined by the duration of the blockage, which is contrary to the water levels in the eastern part of the Nederwaard. For higher return periods with higher precipitation volumes, the impact of a blockage shifts farther away from the pumping station.

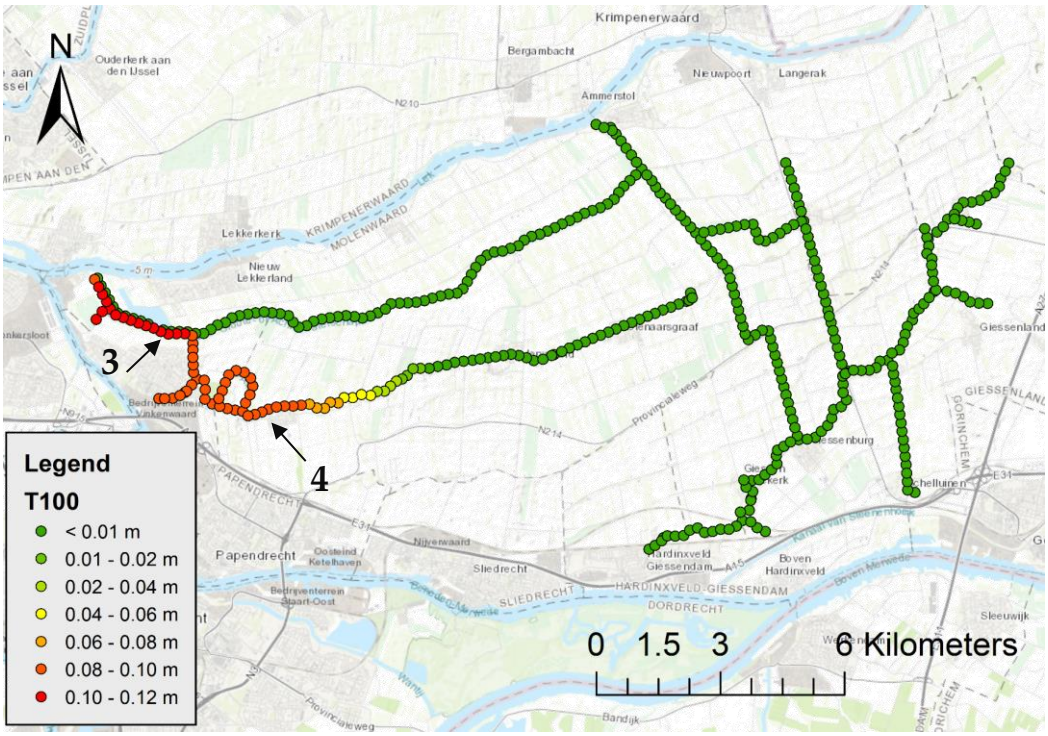


Figure 33: Normative water levels relative to independent case at T100. Location 1 = WL_33_6 and location 2 = 74_2. A positive difference means that in the dependent situation, the normative water levels are higher compared to the reference case.

At locations 3 and 4, the frequency curves are presented in Figure 34. This confirms the conclusion that at location 3, the impact of dependency is visible at smaller return periods than at location 4. Considering the confidence intervals, it is visible that at location 4 the significant difference is small. The difference at location 3 is clearer as this section is more impacted by a blockage of the pumping station.

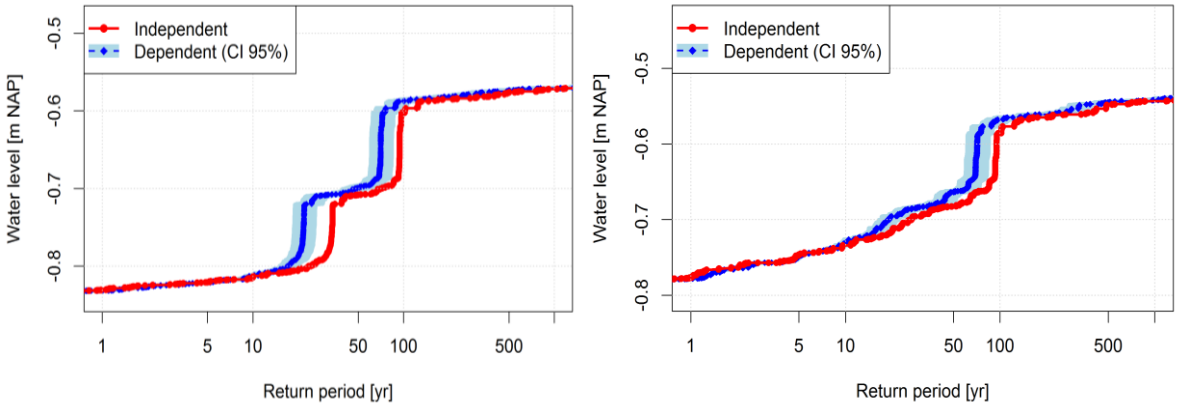


Figure 34: Water level frequency curved at location 3 (left) and location 4 (right) for the dependent and independent case. The dependent curve includes the uncertainty introduced by the copula.

In conclusion, the dependency between precipitation volume and water level Lek does impact the normative water levels. The effect decreases with a larger distance from the pumping station. The difference is however only significant close to the pumping station, where the normative water levels are mainly determined by the probability of a blockage. As the dependency between the two variables is only moderate, the normative water levels of the independent case fall most of the times within the confidence interval of the dependent situation.

5.3.3 Combination three (Wind dir. & speed + precipitation volume)

In the third combination, the effect of the interdependency between precipitation and wind is analysed. In Figure 35 the result is visible for a return period of 100 years. For lower return periods, the spatial distribution of the differences is similar, but the absolute differences are smaller. The figure shows that mainly the Overwaard area is affected by the dependency between wind and precipitation. In most of the area, the difference is about 2 cm, which is considering the uncertainties, not significant. In the most eastern part of the Overwaard area, around location 6, the difference is about 5 cm, as this area is considerably sensitive to wind from westerly directions.

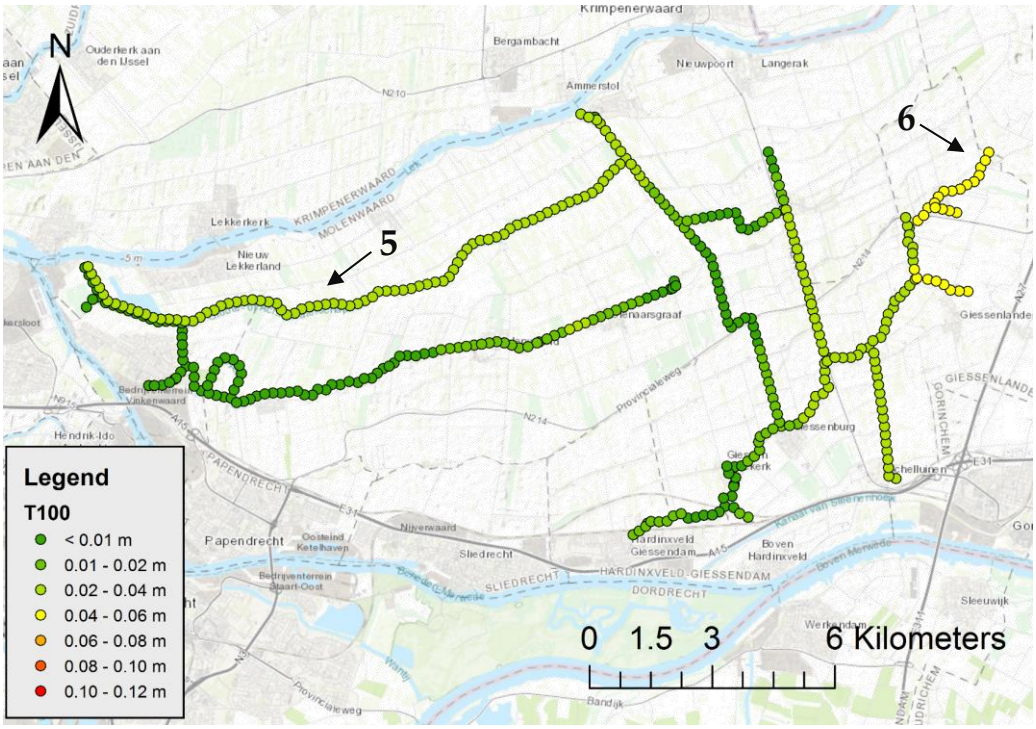


Figure 35: Normative water level relative to independent case at T100. Location 5 = WL_30_14 and location 6 = 923. A positive difference means that in the dependent situation, the normative water levels are higher compared to the reference case.

The sensitivity of the water levels to wind is analysed and discussed in Appendix R. Concluded is that especially winds from the west and northwest affect the water levels considerably in the east part of both the Overwaard and Nederwaard area. The precipitation volume does not clearly affect the effect of wind on the water levels, so there is no extra effect of simultaneous occurrence.

For locations 5 and 6, the water level frequency curves are presented in Figure 36. Visible is that for both locations, differences in water levels start at return periods of about 5 years. Furthermore, the difference at location is 5 is nearly equal throughout the range of return period, while this alternates a lot at location 6. When the confidence intervals are considered, the difference between the dependent and independent situation is in most cases no more than 2 cm.

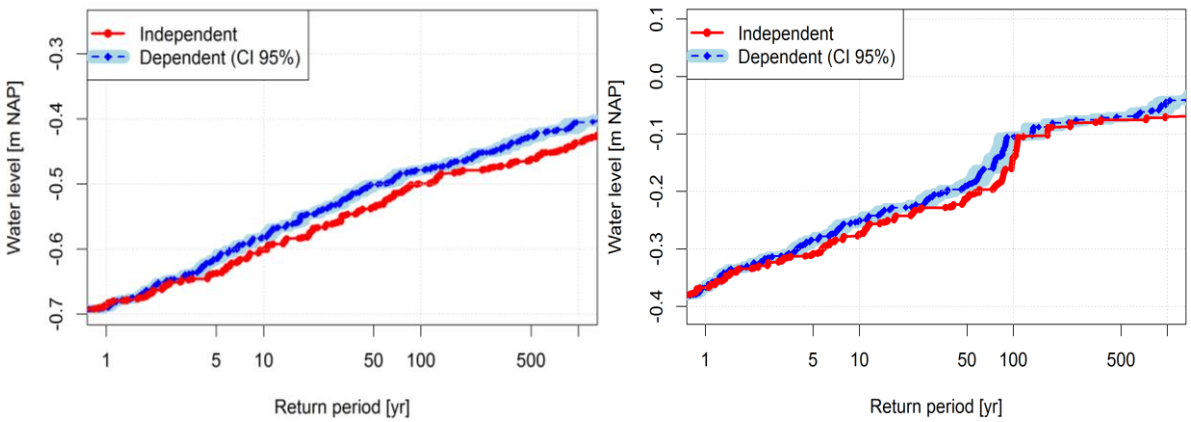


Figure 36: Water level frequency curved at location 5 (left) and location 6 (right) for the dependent and independent case. The dependent curve includes the uncertainty introduced by the copula.

To conclude, the interdependency between precipitation volume and wind speed does affect the normative water levels in the Alblasserwaard, especially the locations which are sensitive to wind. However, due to the low dependence between the two variables, the differences are small, especially considering the uncertainties in the process. At some locations, the differences are equal throughout the range of return periods, while at other locations this difference is only visible at some return periods.

5.3.4 Combination four (Wind dir. & speed + precipitation volume + water level Lek)

In the fourth combination, the effect of the interdependencies between precipitation volume, wind and water level Lek is analysed. In Figure 37 the result is visible for a return period of 25 years. This is the lowest return period with clear differences between the dependent and independent case visible. The figure shows that mainly the Nederwaard area, especially close to the pumping station, is affected by the dependencies (up to 12 cm). In the rest of the Alblasserwaard area, the differences are in the range of about 1-4 cm. This result is expected, looking at the results of combinations 1 (wind and water level Lek), 2 (precipitation and water level Lek) and 3 (precipitation and wind), where the first mainly affects the middle section of the Nederwaard, the second mainly the water levels close to the pumping station, while the third affects a large part of the Overwaard area. As they each affect other parts of the Alblasserwaard area, there is not an 'extra' effect of this fourth combination leading to extra considerable water level differences between the dependent and independent situation.

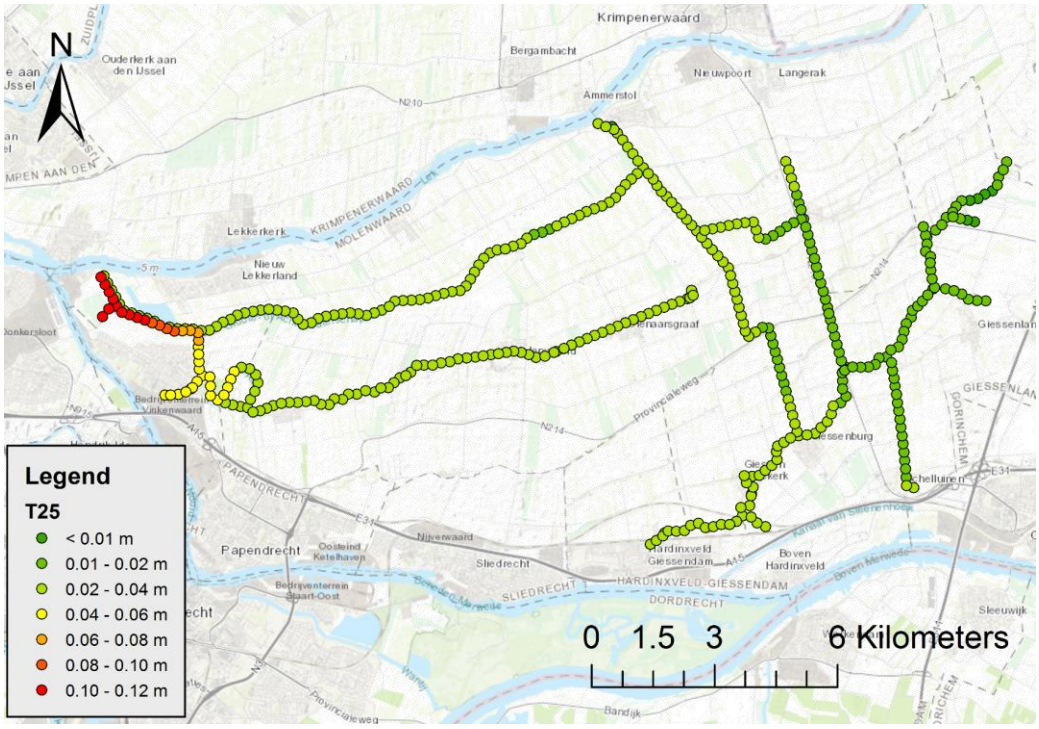


Figure 37: Normative water level relative to independent case at T25. A positive difference means that in the dependent situation, the normative water levels are higher compared to the independent case.

The same figure, but then with a return period of 100 years, can be found in Figure 38. Again, a substantial effect is visible in the Nederwaard area, close to the pumping station. In addition, the effect of the interdependencies with wind is visible in the eastern part of the Alblasserwaard, up to about 4-6 cm. This result is expected, looking at the results of the bivariate combinations. For locations 2 and 4 in Figure 38, the frequency curves are presented in Figure 39. These locations are both in the Nederwaard area as the Overwaard area is only affected by the interdependency between precipitation and wind, which has similar results as the third combination.

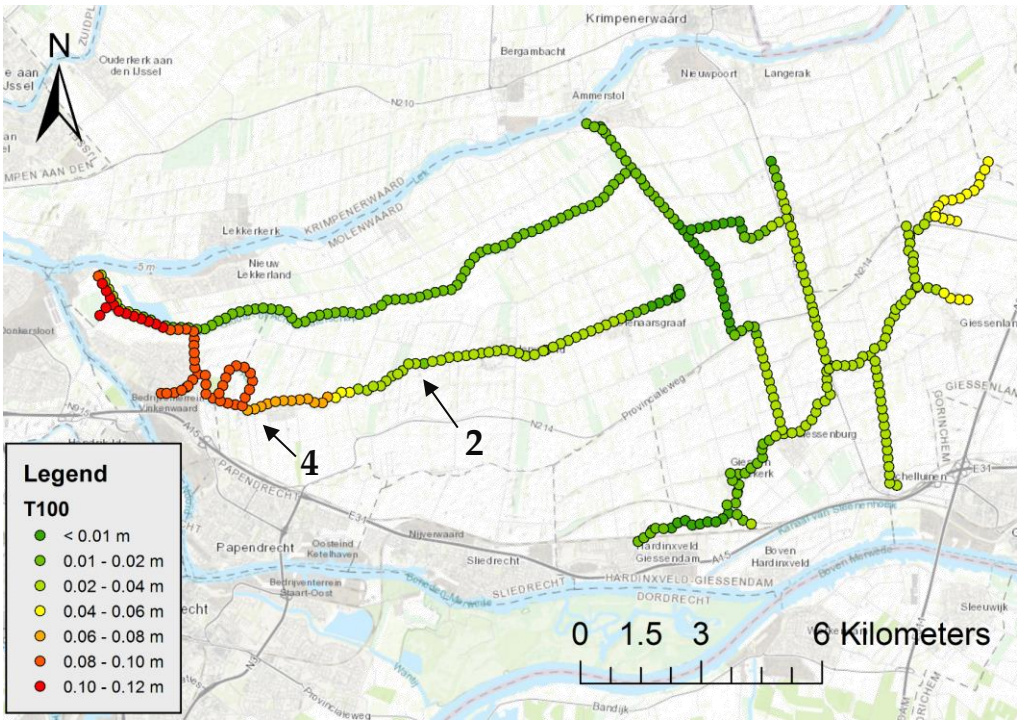


Figure 38: Normative water level relative to independent case at T100. Location 4 = 74_2 and location 2 = WL_40_5. A positive difference means that in the dependent situation, the normative water levels are higher compared to the independent case.

For both locations, it is visible that from a return period of 25 years, there is a significant difference between the dependent and independent situation. Location 4 is mainly determined by the interdependencies with the water level Lek, while location 2 is affected by all the interdependencies of all three the variables. Comparing the result at location 4 with the result of combination 2, it is visible that the shape of the curve is considerably changed as a result of the interaction with the other dependencies. The differences are generally slightly increased and start at a lower return period. These differences are also visible for location 2, looking at the result of combination 1. As a result of those multiple dependencies having an effect, the uncertainty as a result of the vine copula has also increased relative to combination 1.

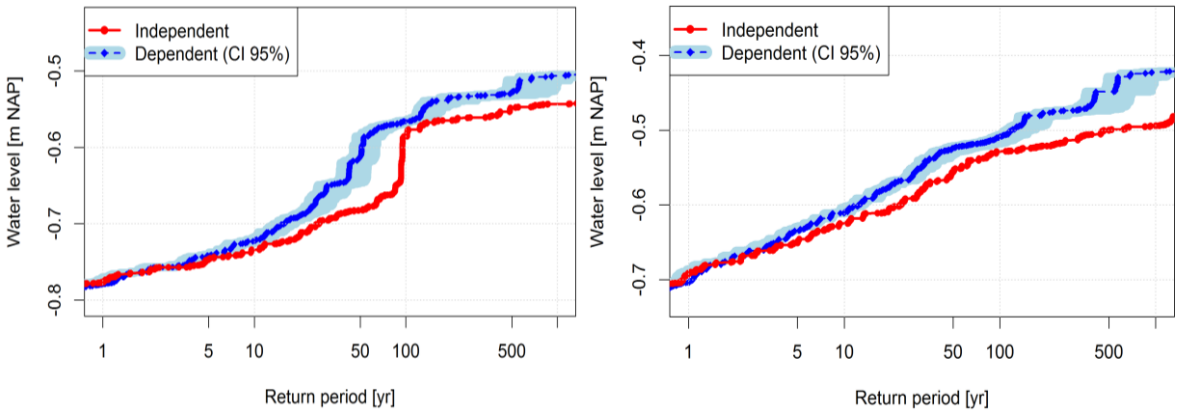


Figure 39: Water level frequency curve at location 4 (left) and location 2 (right) for the dependent and independent case. The dependent curve includes the uncertainty introduced by the vine copula.

To conclude, using a vine copula, it is possible to capture the dependencies between the three variables. Compared to the independent case, there is a large difference in normative water levels of up to 10 cm in some areas (close to the pumping station in the Nederwaard).

In the rest of the Alblasserwaard area, the difference is relatively small to moderate (1-4 cm) with some locations in the east up to 6 cm. All these differences can be derived from the bivariate combinations, as the differences per combination generally affect other parts of the Alblasserwaard. There is not a substantial extra difference in normative water levels due to the interdependencies between all three variables.

5.4 Conclusion

In Chapter 5, the third research question is answered: *"What is the effect of including dependencies between the stochastic variables on the normative water levels in comparison to assuming independence between these variables and how can these differences be explained?"*. This is answered for each of the four combinations. The first combination (wind and water level Lek) shows mainly a significant effect (up to 7 cm) at higher return periods (T300+) at the middle section of the Nederwaard area, which is affected by the combination of wind and a blockage of the pumping station. Combination two (precipitation and water level Lek) shows a significant effect close to the pumping station of the Nederwaard. This effect is visible from a return period of 25 years and the differences in normative water levels can be up to 10 cm. With a higher return period, the effect reaches further from the pumping station. Combination three shows a significant effect (in the range of 1-4 cm) over a large part of the Alblasserwaard, where a small effect can already be seen for lower return periods. In the east part of the Overwaard area, an effect of up to 6 cm is found, as this area is significantly affected by westerly winds. In combination four, the effect of the combination of all the dependencies is tested. This results in similar conclusions as the other combinations, as each combination generally affects other parts of the Alblasserwaard. This is visible in Table 29 for T100, as the sum of the effects of the first three combinations is about equal as the result of combination 4. On the basis of analyses of various types of single event results (e.g. high wind speed/no blockage, no wind speed/blockage and high wind speed/blockage etc.), the effects are found at locations that are expected.

Figure 40: Location numbers referring to table 29

Table 29: Differences between the dependent and independent situation in normative water levels at T100 for several locations in the Nederwaard area (affected in all 4 combinations). Location ID is given with distance to pumping station and reference number (see Figure 40).

Combination → Location ↓	1 (wind-blockage)	2 (precip.-blockage)	3 (precip.-wind)	4 (combined)
(1) J.U. Smitgemaal	0.00 m	0.11 m	0.00 m	0.11 m
(2) WL_36_2 (+5 km)	0.01 m	0.09 m	0.00 m	0.09 m
(3) WL_40_5 (+10 km)	0.01 m	0.00 m	0.01 m	0.02 m
(4) WL_42_8 (+15 km)	0.01 m	0.00 m	0.02 m	0.02 m

So, in general, including the interdependencies between the variables leads to an increase in normative water levels. However, this depends a lot on the location. Some locations are only affected by one variable (precipitation volume), so there will be no effect of including dependencies. However, when locations are affected by the external water level and/or wind, there is a substantial effect visible as a result of the interdependences. Especially the effect of the water level of the Lek is important, leading to large differences (up to 11 cm) close to the pumping station of the Nederwaard, even at lower return periods of 25 years. Another conclusion is that the strongest dependency does not directly cause the largest differences in normative water levels, as the dependency between wind speed and water level Lek only led to large differences in a specific area at higher return periods (T300+).

The results in this chapter are specific to the Alblasserwaard, a polder water system in the Netherlands. Conclusions from this chapter could potentially be generalised to other areas by analysing its characteristics. For example, if the water levels in the water system and the external water level (affecting the pumping stations) are both affected by wind or precipitation, it can be expected that the normative water levels in the water system increase relative to the independent case. Also, the higher joint probabilities of higher wind speeds and precipitation can potentially affect a lot of other water systems. The effects on the normative water levels depend on the characteristics of a water system, for example the direction of flow in the rivers and canals. So, the results in this chapter can be used as a reason to look further into a certain water system or to conclude that no further investigation is needed based on a quick analysis of its characteristics.

6 Discussion

In this chapter, first the potential, relevance and implications of this study are discussed in section 6.1. Secondly, the assumptions, simplifications and uncertainties involved in the research process will be reflected on in section 6.2. Lastly, the generalisability of the results of this case study is discussed in section 6.3.

6.1 Potential

This research is the first (to the best of my knowledge) to study the effects of interdependencies on the normative water levels in a regional water system in the Netherlands. It showed that it is feasible to include interdependencies in determining normative water levels using available packages in the software *R* and additionally, it is possible to reproduce the results as objective statistical tests available in these packages are applied. Dependencies between multiple stochastic variables in the Alblasserwaard are found. Using the copula theory and SOBEK model, it is found that the effects of these interdependencies on the normative water levels depend a lot on location and considered return period. Some locations are not affected by interdependencies while some locations are substantially (up to about 10 cm) impacted, especially locations close to the pumping station in the Nederwaard area. So, this study shows that it is possible to obtain such results, with methods which are reproducible and applicable to be applied in other studies.

This study is relevant as the assumption of independence between the stochastic variables led, in this case, to an underestimation of the normative water levels and thus an underestimation of the risk. This could potentially lead to under-dimensioning of dikes and/or pumping stations, also in other types of water systems. If the stochastic method is applied in regional water systems, it is recommended to study the presence of dependencies and its potential effects on the normative water levels. Whether it makes a significant difference to include dependencies relative to assuming independence depends on the characteristics of a water system and the range of other uncertainties (e.g. model uncertainty) involved in the process of determining normative water levels.

The results of this study give a similar direction to the effects of interdependencies on joint probabilities as other studies in the field of water engineering and management. Generally, positive dependencies between variables lead to an increase of the joint probability of extreme events. For example, Ward et al. (2018), Wahl et al. (2015) and Bevacqua et al. (2017) found that the joint probability of events in which design river discharges and design sea-levels in deltas and estuaries are considered as dependent, can be several magnitudes higher. This is a result of low-pressure storm systems and other meteorological processes bringing winds and precipitation simultaneously. This study adds another case study, a regional water system in the Netherlands, to the existing literature on the potential effects of dependencies on the joint probabilities of combined extreme events.

Compared to other methods such as the multivariate normal distribution, the copula method has shown its potential in this study and other studies to construct joint probability distributions as it shows great flexibility (Hao & Singh, 2016; Genest & Favre, 2007). Other dependence methods that are used in the Netherlands for modelling the dependencies in the primary water systems, such as the Homoscedastic model (HOS), Heteroscedastic model (HES) and De Haan method (HN), also have some limitations as opposed to the copula method. For example, the assumptions about the marginal distributions, limitations in types of dependence structures (e.g. only positive, homoscedastic or asymptotical) and no possibilities to extend it to more than two variables (Diermanse & Geerse, 2010; Deltares, 2016). The copula method is more flexible as all sorts of marginal distributions and dependence structures are possible and is not limited to only two dimensions.

6.2 Limitations

The results are achieved by making some assumptions and simplifications to keep the execution of the study feasible, while the results also have some limitations. Topics that are discussed here are the limitations of the available data, event selection, number of probability distributions, number of copula functions and simplifications, assumptions and uncertainties in applying the stochastic method. Also discussed is the impact of those limitations on the results.

Stationarity of the data and correct measurements without errors are assumed for the used datasets. A slight upward trend for precipitation volume and a slight downward trend for wind speed is noted. No suspicious outliers are found in the datasets for the relevant time period. In addition, anthropogenic processes could affect the water level of the Lek over time, such as dredging or the functioning of hydraulic structures (e.g. Maeslant barrier and Hagestein weir). However, as the focus of this study is on the differences in normative water levels between the dependent and independent situation, where for both situations the same datasets are used, this will not substantially impact the results.

In the process of selecting events from the datasets, choices are made which could affect the results. Temporally independent events are selected as this is a general requirement for fitting probability distributions and copula functions. In most cases, this requirement is met, but in some cases a small but significant autocorrelation is found. The thresholds based on a minimum value and interevent time are kept low as with high thresholds, a lower number of events are selected and/or smaller dependencies are found between some variables. This is also discussed in Paprotny et al. (2018), who studied compound flood potential and found more realistic return periods of combined high river discharges and high sea levels by using all observations instead of only using the upper 5%. In addition, a high threshold leads to a low number of events, increasing the uncertainty in the selected probability distributions and copula functions. A disadvantage of a low threshold is that the probability distribution is relatively better fitted to the lower values and not specifically to the extremes, this is however mitigated by focussing on the higher values in the visual selection of probability distributions. Therefore, it is expected that the choices herein do no impact the results substantially.

The number of copula functions tested on the data is limited by the availability in the software *R*. To handle several different dependence structures and marginal probability distributions, the available number of parametric copula functions and a wide variety of marginal probability distributions are tested on the data. However, some variables and its dependence structure with other variables are hard to capture with parametric functions, such as the variable wind direction. This could be further explored with nonparametric copula functions and/or probability distributions, which showed potential in for example Vittal et al. (2015). However, based on the statistical tests and visual validation simulations, the used number and variety of functions seems sufficient (excluding wind direction) as the simulations and visual comparisons are relatively close to the observations. This also applies to the vine copula. The use of vine copulas is relatively new and the amount of available literature in the field of water engineering and management is low. However, the results in this study using a vine copula seem appropriate based on the results of the bivariate combinations. Its usage is not yet extensively analysed in literature, but this study and others show the potential due to its flexibility in combining all sorts of marginal distributions and bivariate copula functions to model dependencies.

In the process of determining normative water levels in the model study, simplifications and assumptions are made. One normative duration for the precipitation events is chosen (9 days), standard shapes for precipitation and wind are used, choices in discretisation of variables, timing of the blockage of the pumping station and a uniform distribution of precipitation over the complete area is assumed. In addition, the model to determine water levels in the Alblasserwaard comes with a certain degree of uncertainty. All these simplifications and assumptions affect the normative water levels, therefore the uncertainty and the effects of choices on the normative water levels should be further explored. In this study only the uncertainty caused by the dependency is shown as the uncertainty range in the normative water levels as this is the focus of this study. The other uncertainties affect both the dependent and independent cases, which means that the differences between these two should largely remain equal. However, it cannot definitively be said that the found differences between the dependent and independent situation are significantly different as the total uncertainty band is not known.

6.3 Generalisation

In this study, the methods are only applied to one area, the Alblasserwaard. Some aspects are very specific for this area, such as the functioning of the pumping stations and the dependency between the wind and the water level of the Lek. This is not always generalisable to other areas. In addition, in other areas there might be other stochastic variables that are relevant to consider which are not used in this study, e.g. initial groundwater levels or failure of structures. Still, some expectations could be made based on this study on how interdependencies affect normative water levels based on e.g. layout of the area and which variables affect the normative water levels in that specific area. For example, the sensitivity of the pumping station(s) to the external water level, which could be affected by wind and precipitation or the direction of the flow (east to west in this case study) in the water system relative to the wind field. This study showed that it is possible to account for interdependencies and that it can impact normative water levels, especially for

the combinations in which the external water level of the Lek is involved. Therefore, it is relevant to consider these methods in studies where dependencies are potentially relevant. The potential of further analysis can be assessed based on the results of this study. So, it can also be used to conclude that no further analysis is necessary based on the characteristics of a water system. Further research could be done on other types of water systems (e.g. free flowing water systems, river-sea interaction and river-lake interaction) in the Netherlands and elsewhere. These studies could also bring up other issues such as the availability of data and the effects of other variables. In addition, this could help to define whether some dependencies such as the combination of precipitation and wind are similar in other areas and if this relation could be generalised over larger areas.

7 Conclusion and recommendations

7.1 Conclusion

The goal of this study is *to evaluate to what extent including interdependence of stochastic variables affects normative water levels in comparison to assuming independence between these variables*. This is studied in a case study of the Alblasserwaard by using a copula approach.

The first research question '*Are the stochastic variables used in determining normative water levels interdependent?'* is answered by analysing the relationships between the variables visually and with a correlation coefficient. The clearest dependency is found between wind speed and water level Lek, while other dependencies are also present but less defined. Precipitation events mostly take place with winds from southwest to northwest direction and with an increasing precipitation volume, the probabilities of higher wind speeds and higher water levels increase. It is clear that no dependencies can be neglected in the analysis.

The results of the second research question '*What are the most suitable joint probability distributions describing each of the combinations of interdependent stochastic variables, and what is the effect of including dependency on the joint probabilities of the stochastic variables?'* showed that by including interdependencies, the probability distributions are closer to the observations than in the case of assumed independence. Furthermore, it is clear that the joint probabilities of combined extremes are higher in the case of dependence relative to independence. This is especially visible in the case of wind speed and water level Lek.

For the third and last research question '*What is the effect of including dependencies between the stochastic variables on the normative water levels in comparison to assuming independence between these variables and how can these differences be explained?'*, the joint probability distributions are used to determine normative water levels with a hydrological/hydraulic model of the Alblasserwaard and compared to the independent reference case. The resulting differences are location specific and vary over the return periods. The largest differences are found close to the pumping station (up to 10-12 cm at T100), mainly as a result of the dependency between precipitation and water level of the Lek. In the east part of the Alblasserwaard, some smaller differences (about 5 cm at T100) are found as a result of the dependency of precipitation with wind. In other areas, the differences are smaller than 2-3 cm. These results are explainable based on the bivariate combinations and the analysis of model results of specific events.

To conclude, there are interdependencies between the stochastic variables that affect the normative water levels. These results are location and return period specific. Furthermore, it is applicable and feasible to use the copula method to produce these results.

7.2 Recommendations

This study shows that interdependencies between stochastic variables are present, affect the joint probability distributions and normative water levels in a regional water system. Therefore, it is recommended to study the potential presence of interdependencies between the considered stochastic variables in future regional water system analyses. A first indication of potential effects on the normative water levels could be made based on the characteristics of the water system. For example, whether the pumping station is affected by the external water level which in turn is affected by precipitation and/or wind. Another example resulting from this study is the flow from east to west in the Alblasserwaard, where winds from the west direction hinder the flow to the main pumping station. When such characteristics and dependencies are present and the normative water levels need to be accurately determined, an in-depth assessment considering the dependencies as in this study could be made.

It is recommended to apply similar methods as used in this study to other case studies with different characteristics, where the effects of dependencies are potentially different and where other variables affect the normative water levels. In this study, a typical polder water system is used which mainly depends on pumping stations. In other types of water systems (e.g. free flowing water system, river-lake/sea interactions), other effects might be found. In addition, the generalisability of the results of this study could be discussed when more case studies are available. These case studies could help in setting up guidelines on which characteristics of water systems could potentially lead to differences in normative water levels due to dependencies. Also, the availability of data is a point of attention for the generalisability. Especially data on wind speed, which is not widely available or only for a relatively short amount of time. If the joint probabilities of, for example, wind speed and precipitation are found similar in multiple areas, it is possible to use these for other areas with less data availability (e.g. based on characteristics such as distance to the sea).

As described in the discussion, there are a lot of uncertainties involved in the process of determining normative water levels in a regional water system using the stochastic method. The choices and assumptions herein could potentially have more impact than the interdependencies on the normative water levels. Therefore, it is recommended to study the sensitivity of the normative water levels to several assumptions, uncertainties and choices. This could for example include event selection, uncertainties in marginal distributions, using other measuring stations, using other normative durations of precipitation events, discretisation of marginal distributions and the timing of wind and a blockage. When the range of effects of these uncertainties and assumptions are better known, the impact of the dependencies on the normative water levels can be estimated more accurately.

It is recommended to the field of water engineering and management to further study and establish the use of copulas and vine copulas for dependence modelling. It has shown its potential in this study and other studies and could be used for all sorts of research where dependencies are considered, not only for the specific goal of determining normative water levels in a regional water system. For example, studies on failure mechanisms of flood defences, water quality with interactions between different elements or flood and drought predictions.

References

Aas, K., Czado, C., Frigessi, A. & Bakken, H. (2009). Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics*, (44) 182-198.

Arcadis. (2014). *Memo Integreren wind- en hoogwatersyatistiek.* Apeldoorn: Arcadis.

Arcadis. (2015). *Integraal verbeterprogramma Alblasserwaard en Vijfheerenlanden.* Apeldoorn: Arcadis.

Bedford, T. & Cooke, R. (2001). Probability Density Decomposition for Conditionally Dependent Random Variables Modeled by Vines. *Annals of Mathematics and Artificial Intelligence*, (32) 245-268.

Bedford, T. & Cooke, R. (2002). VINES - A new graphical model for dependent random variables. *The Annals of Statistics*, (4) 1031-1068.

Bernardara, P., Mazas, F., Weiss, J., Andreewsky, M., Kergadallan, X., Benoît, M. & Hamm, L. (2012). On the two step threshold selection for over-threshold modelling. *Coastal Engineering*, 1-6.

Bevacqua, E., Maraun, D., Haff, I., Widmann, M. & Vrac, M. (2017). Multivariate statistical modelling of compound events via pair-copula constructions: analysis of floods in Ravenna (Italy). *Hydrology and Earth System Sciences*, (21) 2701-2723.

Bosch, S., Hakvoort, H., Diermanse, F. & Verhoeve, C. (2006). Verantwoord omgaan met de nieuwe neerslagstatistiek. *Stromingen*, (1) 13-24.

Bossenbroek, J. (2004). *Statistiek vóóraf of statistiek achteraf?* Delft: TU-Delft.

Buishand, T., Jilderda, R. & Wijngaard, J. (2009). *Regionale verschillen in extreme neerslag.* De Bilt: KNMI.

Carnicero, J., Ausin, M. & Wiper, M. (2013). Non-parametric copulas for circular–linear and circular–circular data: an application to wind directions. *Stochastic Environmental Research and Risk Assessment*, (27) 1991-2002.

Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values.* Bristol: Springer Series in Statistics.

Daneskhah, A., Remesan, R., Chatrabgoun, O. & Holman, I. (2016). Probabilistic modeling of flood characterizations with parametric and minimum information pair-copula model. *Journal of Hydrology*, (540) 469-487.

De Graaff, B. & Versteeg, R. (2000). Wateroverlast, zo goed als zeker. *H2O*, 28-30.

Delignette-Muller, M. & Dutang, C. (2015). fitdistrplus: An R Package for Fitting Distributions. *Journal of Statistical Software*, (64) 1-34.

Deltares. (2016). *Hydra-Ring.* Delft: Deltares.

Deltares. (2019). *SOBEK User Manual.* Delft: Deltares.

Diermanse, F. & Geerse, C. (2010). Correlation models in flood risk analysis. *Relliability, Risk and Safety*, 432-439.

Far, S. & Khairi Abd. Wahab, A. (2016). Evaluation of Peaks-Over-Threshold Method. *Ocean Sciences Discussions*, 1-25.

Favre, A., El Adlouni, S., Perreault, L., Thiémonge, N. & Bobée, B. (2004). Multivariate hydrological frequency analysis using copula. *Water resources research*, (40) 1-12.

Franken, T., Blacnkaert, J. & Leyssen, G. (2016). Synthetic events for flood risk calculation by using a nested Copula. *FLOODrisk 2016 - 3rd European Conference on Flood Risk Management.* E3S Web of Conferences.

Genest, C. & Favre, A. (2007). Everything You Always Wanted to Know about Copula Modeling but Were Afraid to Ask. *Journal of hydrologic engineering*, (12) 347-368.

Genest, C., Ghoudi, K. & Rivest, L.-P. (1995). A semiparametrie estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, (3) 543-552.

Genest, C., Rémillard, B. & Beaudoin, D. (2009). Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics*, (44) 199-213.

Gräler, B., van den Berg, M., Vandenberghe, S., Petroselli, A., Grimaldi, S., De Baets, B. & Verhoest, N. (2013). Multivariate return periods in hydrology: a critical and practical review focusing on synthetic design hydrograph estimation. *Hydrology and Earth System Sciences*, (17) 1281-1296.

Grimaldi, S., Kao, S.-C., Castellarin, A., Papalexiou, S.-M., Viglione, A., Laio, F., . . . Gedikli, A. (2011). *Statistical Hydrology.* Oxford: Academic Press.

Grønneberg, S. & Hjort, N. (2014). The Copula Information Criteria. *Scandinavian Journal of Statistics*, (41) 436-459.

Hao, Z. & Singh, V. (2016). Review of dependence modeling in hydrology and water resources. *Progress in Physical Geography*, (40) 549-578.

Hofert, M., Kojadinovic, I., Mächler, M. & Yan, J. (2018). *Elements of Copula Modeling with R.* Cham, Switzerland: Springer.

Hu, H. (2008). *Poisson distribution and application.* Knoxville, Tennessee, USA: Department of Physics and Astronomy,University of Tennessee.

HydroLogic. (2010). *Maatregelenstudie regionaal watersysteem Nederwaard en Overwaard.* Amersfoort: HydroLogic.

HydroLogic. (2017). *Feiten over werking van het watersysteem in de Alblasserwaard en achtergronden bij mogelijke maatregelen.* Amersfoort: HydroLogic.

HydroLogic. (2018a). *Overdrachtsdocument Sobek modellering Alblasserwaard.* Amersfoort: HydroLogic.

HydroLogic. (2018b). *Stochastenmethode berekeningen complex Kinderdijk.* Amersfoort: HydroLogic.

Infram, Strootman & HydroLogic. (2017). *Een nieuw begin voor een iconische polder: Visie voor 2050 op het watersysteem in de Alblasserwaard – Een toekomstbestendige inrichting van het watersys-teem, door een adaptieve uitvoeringsstrategie.* Tiel: Waterschap Rivierenland.

Interprovinciaal Overleg. (1999). *IPO-richtlijn ter bepaling van het veiligheidsniveau van boezemkaden.* The Hague: Interprovinciaal Overleg (IPO).

Joe, H. (1996). Families of m-Variate Distributions With Given Margins and m(m - 1)/2 Bivariate Dependence Parameters. In *Distributions with Fixed Marginals and Related Topics* (pp. 120-141). Vancouver, Canada: Department of Statistics, University of British Columbia.

Jordanger, L. & Tjøstheim, D. (2014). Model selection of copulas: AIC versus a cross validation copula information criterion copula information criterion. *Statistics and Probability Letters*, (92) 249-255.

Kamaruzaman, I., Zin, W. & Ariff, N. (2018). A Generalized Bivariate Copula for Flood Analysis in Peninsular Malaysia.

Kew, S., Selten, F., Lenderink, G. & Hazeleger, W. (2013). The simultaneous occurrence of surge and discharge extremes for the Rhine delta. *Natural Hazards and Earth System Sciences*, (13) 2017-2029.

Kim, G., Silvapulle, M. & Silvapulle, P. (2007). Comparison of semiparametric and parametric methods for estimating copulas. *Computational Statistics & Data Analysis*, (51) 2836 - 2850.

KNMI. (2019). *Uurgegevens van het weer in Nederland*. Retrieved 15 03, 2019, from KNMI: https://projects.knmi.nl/klimatologie/uurgegevens/selectie.cgi

Kurowicka, D. & Cooke, R. (2007). Sampling algorithms for generating joint uniform distributions using the vine-copula method. *Computational Statistics & Data Analysis*, (51) 2889-2906.

Li, F., van Gelder, H., Ranasignhe, R., Callaghan, D. & Jongejan, R. (2014). Probabilistic modelling of extreme storms along the Dutch coast. *Coastal Engineering*, (86) 1-13.

Mardia, K. & Jupp, P. (2000). *Directional statistics.* Chicester: John Wiley & Sons.

Masseran, N., Razali, A., Ibrahim, K., Zaharim, A. & Sopian, K. (2013). The Probability Distribution Model of Wind Speed over East Malaysia. *Research Journal of Applied Sciences, Engineering and Technology 6*, (10) 1774-1779.

Mazas, F. & Hamm, L. (2011). A multi-distribution approach to POT methods for determining extreme wave heights. *Coastal Engineering*, (58) 385 - 394.

Nelsen, R. (2006). *An Introduction to Copulas - Second edition.* New York: Springer.

Nelsen, R. (n.d.). *Properties and applications of copulas: A brief survey.* Department of Mathematical Sciences, Lewis & Clark College.

OpenStreetMap. (2019). *OpenStreetMap*. Retrieved 05 03, 2019, from OpenStreetMap: https://www.openstreetmap.org/

Pandey, P., Das, L., Jhajharia, D. & Pandey, V. (2018). Modelling of interdependence between rainfall and temperature using copula. *Modeling Earth Systems and Environment*, (4) 867-879.

Paprotny, D., Vousdoukas, M., Morales Napoles, O., Jonkman, S. & Feyen, L. (2018). Compound flood potential in Europe. *Hydrology and Earth System Sciences*.

Pham, M., Vernieuwe, H., De Baets, B. & Verhoest, N. (2018). A coupled stochastic rainfall–evapotranspiration model for hydrological impact analysis. *Hydrology and Earth System Sciences*, (22) 1263-1283.

Provincie Zuid-Holland. (2019). *Waterveiligheid regionale keringen*. Retrieved 20 08, 2019, from Provincie Zuid-Holland: https://staatvan.zuid-holland.nl/portfolio_page/waterveiligheid-regionale-keringen/

R Core Team. (2018). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. Retrieved from URL https://www.R-project.org/

Ribatet, M. (2011). *A User's Guide to the POT Package (Version 1.4).* Montpellier, France: University of Montpellier II, Department of Mathematics.

Rijkswaterstaat. (2019). *Waterinfo*. Retrieved 05 03, 2019, from Rijkswaterstaat: https://www.waterinfo.rws.nl

Rumsey, D. (2006). *Probability For Dummies.* Hoboken, NJ: Wiley Publishing Inc.

Schepsmeier, U. (2013). A goodness-of-fit test for regular vine copula models. *Econometric Reviews*.

Schepsmeier, U., Stoeber, J., Brechmann, E., Graeler, B., Nagler, T. & Erhardt, T. (2018). *Package 'VineCopula'.*

Shafaei, M., Fakheri-Fard, A., Dinpashoh, Y., Mirabbasi, R. & De Michele, C. (2017). Modeling flood event characteristics using D-vine structures. *Theoretical and Applied Climatology*, (130) 713-724.

Shumway, R. & Stoffer, D. (2017). *Time Series Analysis and Its Applications.* Cham, Switzerland: Springer International Publishing.

Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Stat. Univ. Paris*, 229-231.

Sohoni, V., Gupta, S. & Nema, R. (2016). A comparative analysis of wind speed probability distributions for wind power assessment of four sites. *Turkish Journal of Electrical Engineering & Computer Sciences*, (24) 4724-4735.

Soukkisian, T. & Karathanasi, F. (2017). On the selection of bivariate parametric models for wind data. *Applied Energy*, (188) 280-304.

STOWA. (2004). *Statistiek van extreme neerslag in Nederland.* Utrecht: STOWA.

STOWA. (2011). *Standaard werkwijze voor de toetsing van watersystemen aan de normen voor Regionale Wateroverlast.* Amersfoort: STOWA.

STOWA. (2015a). *Flood Risk of Regional Flood Defences.* Amersfoort: STOWA.

STOWA. (2015b). *Leidraad toetsen op veiligheid regionale waterkeringen.* Amersfoort: STOWA.

Unie van Waterschappen. (2003). *Het Nationaal Bestuursakkoord Water.* Unie van Waterschappen.

Unie van waterschappen, Interprovinciaal Overleg & STOWA. (2016). *Visie op de regionale waterkeringen 2016.* Amersfoort: STOWA.

Vandenberghe, S., Verhoest, N. & De Baets, B. (2010). Fitting bivariate copulas to the dependence structure between storm characteristics: A detailed analysis based on 105 year 10 min rainfall. *Water Resources Research*, (46) 1-17.

Vittal, H., Singh, J., Kumar, P. & Karmakar, S. (2015). A framework for multivariate data-based at-site flood frequencyanalysis: Essentiality of the conjugal application of parametricand nonparametric approaches. *Journal of Hydrology*, (525) 658 - 675.

Wahl, T., Jain, S., Bender, J., Meyers, S. & Luther, M. (2015). Increasing risk of compound flooding from storm surge and rainfall for major US cities. *Nature Climate Change*, (5) 1093-1098.

Wahl, T., Mudersbach, C. & Jensen, J. (2012). Assessing the hydrodynamic boundary conditions for risk analyses in coastal areas: a multivariate statistical approach based on Copula functions. *Natural Hazards and Earth Sustem Sciences*, (12) 495-510.

Ward, P., Couasnon, A., Eilander, D., Haigh, I., Hendry, A., Muis, S., . . . Wahl, T. (2018). Dependence between high sea-level and high river discharge increases flood hazard in global deltas and estuaries. *Environmental Research Letters*, (13).

Waterschap Rivierenland. (2010). *GGOR en peilbesluit Alblasserwaard .* Tiel: Waterschap Rivierenland.

Yan, J. (2007). Enjoy the Joy of Copulas: With a Package copula. *Journal of Statistical Software*.

Zhong, H., van Overloop, P. & van Gelder, P. (2013). A joint probability approach using a 1-D hydrodynamic model forestimating high water level frequencies in the Lower Rhine Delta. *Natural Hazards and Earth System Sciences*, (13) 1841-1852.

Appendices

A. Detailed map of the Alblasserwaard

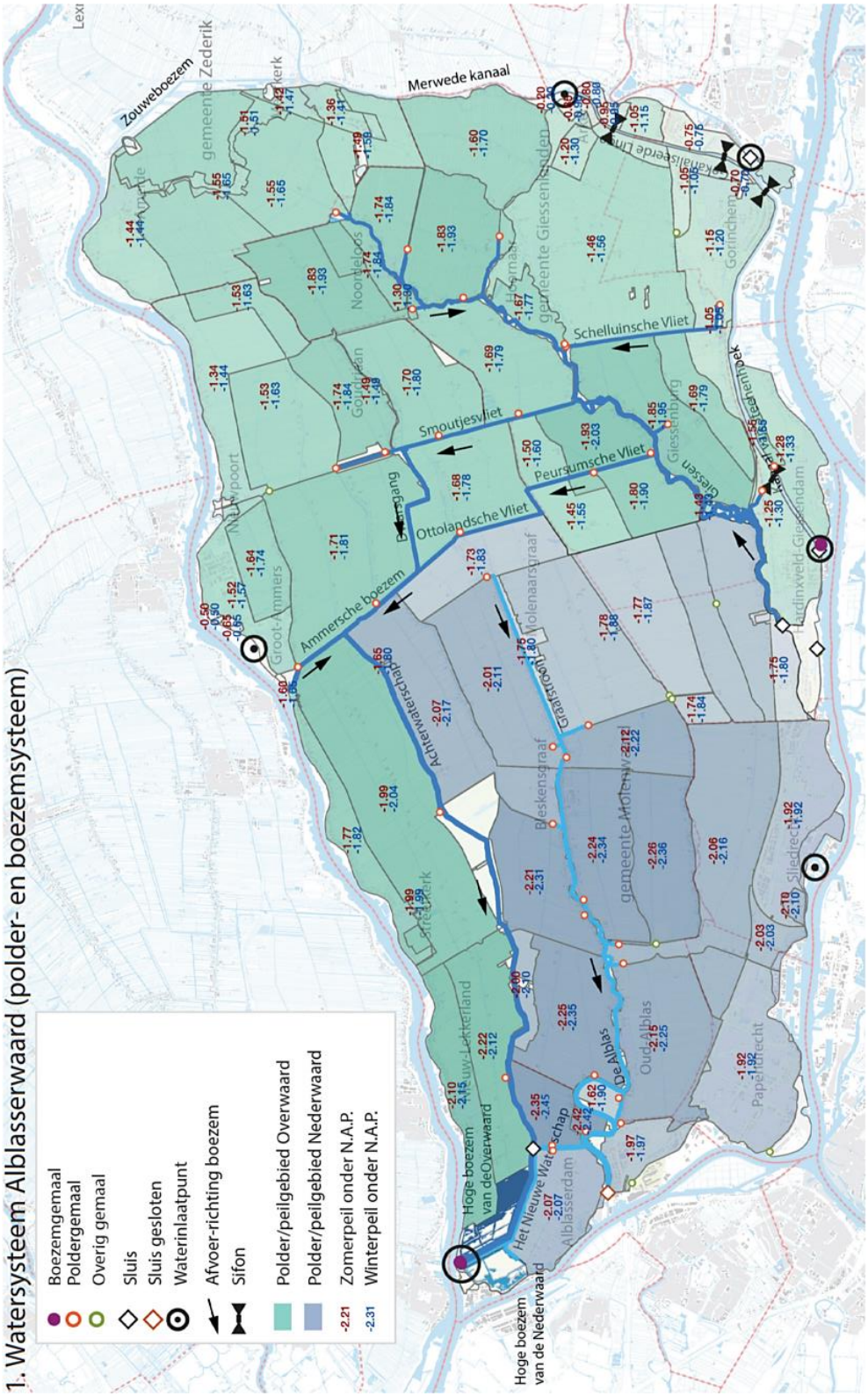


Figure 41: Overview water system Alblasserwaard (HydroLogic, 2017). Arrows represent general direction of flow. Red numbers are water levels (m NAP) in summer, blue numbers are water levels in winter (m NAP). Purple dot in the top left is the main pumping station complex which discharges the water from the Alblasserwaard to the Lek.

B. Comparison weather stations Rotterdam Airport & Oud-Alblas

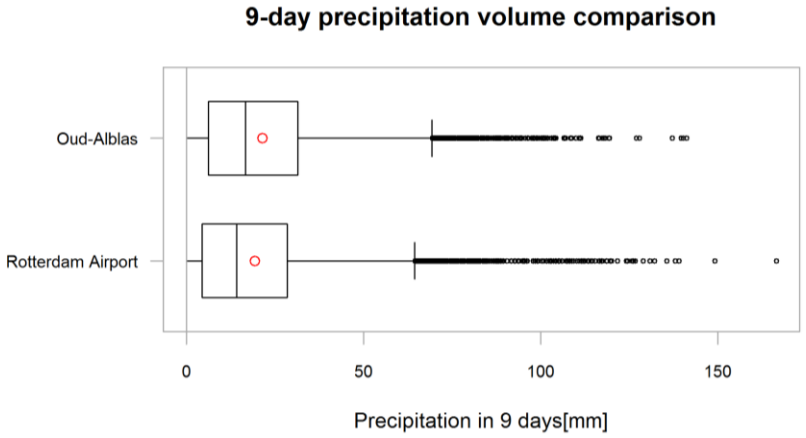


Figure 42: Box plot comparison Oud-Alblas and Rotterdam airport. Red dot represents mean value. Box represent 2nd to 3rd quartile with a line representing the median. The extreme values (outliers) are plotted separately.

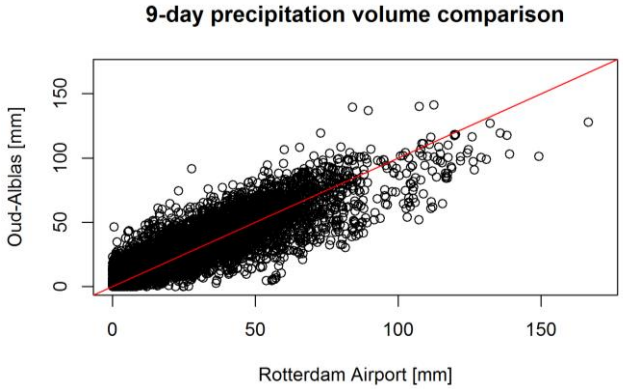


Figure 43: Comparison 9-day precipitation volume (moving sum) Oud-Alblas and Rotterdam. Red line represents diagonal y=x.

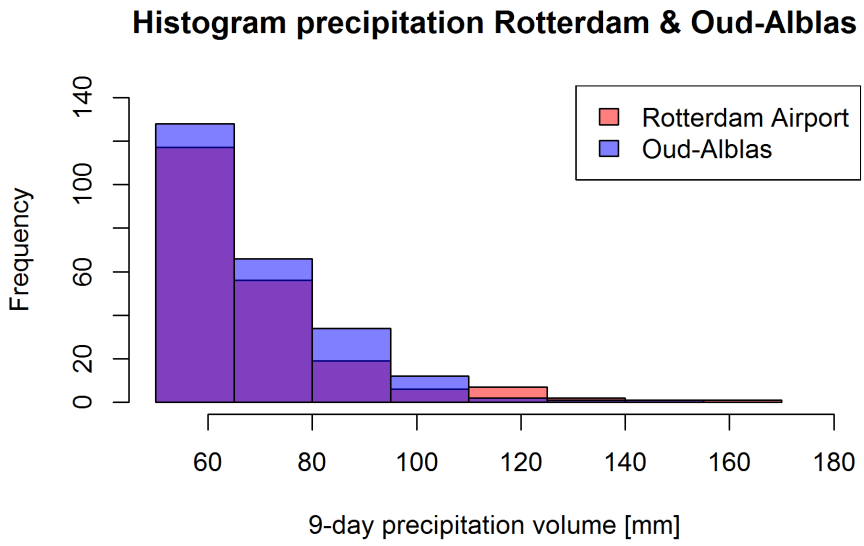


Figure 44: Comparison histogram precipitation volume Rotterdam Airport & Oud-Alblas of precipitation events >50 mm. Note that histograms are overlapping with different colouring

C. Autocorrelation of stochastic variables

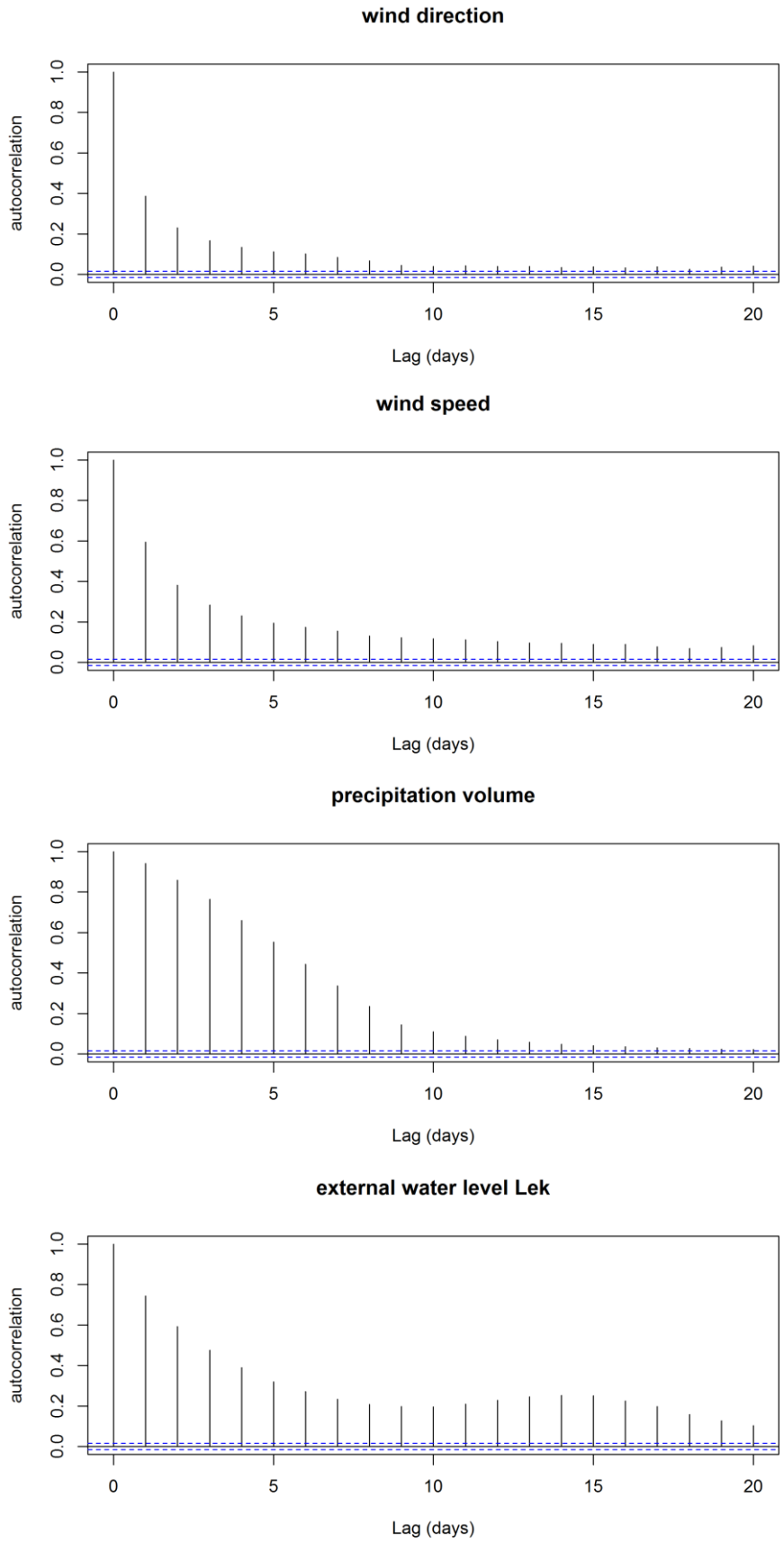


Figure 45: Autocorrelation plots of the stochastic variables based on daily values. Blue dotted line represents significant autocorrelation (α=0.05)

D. Cross-correlation plots based on hourly observations

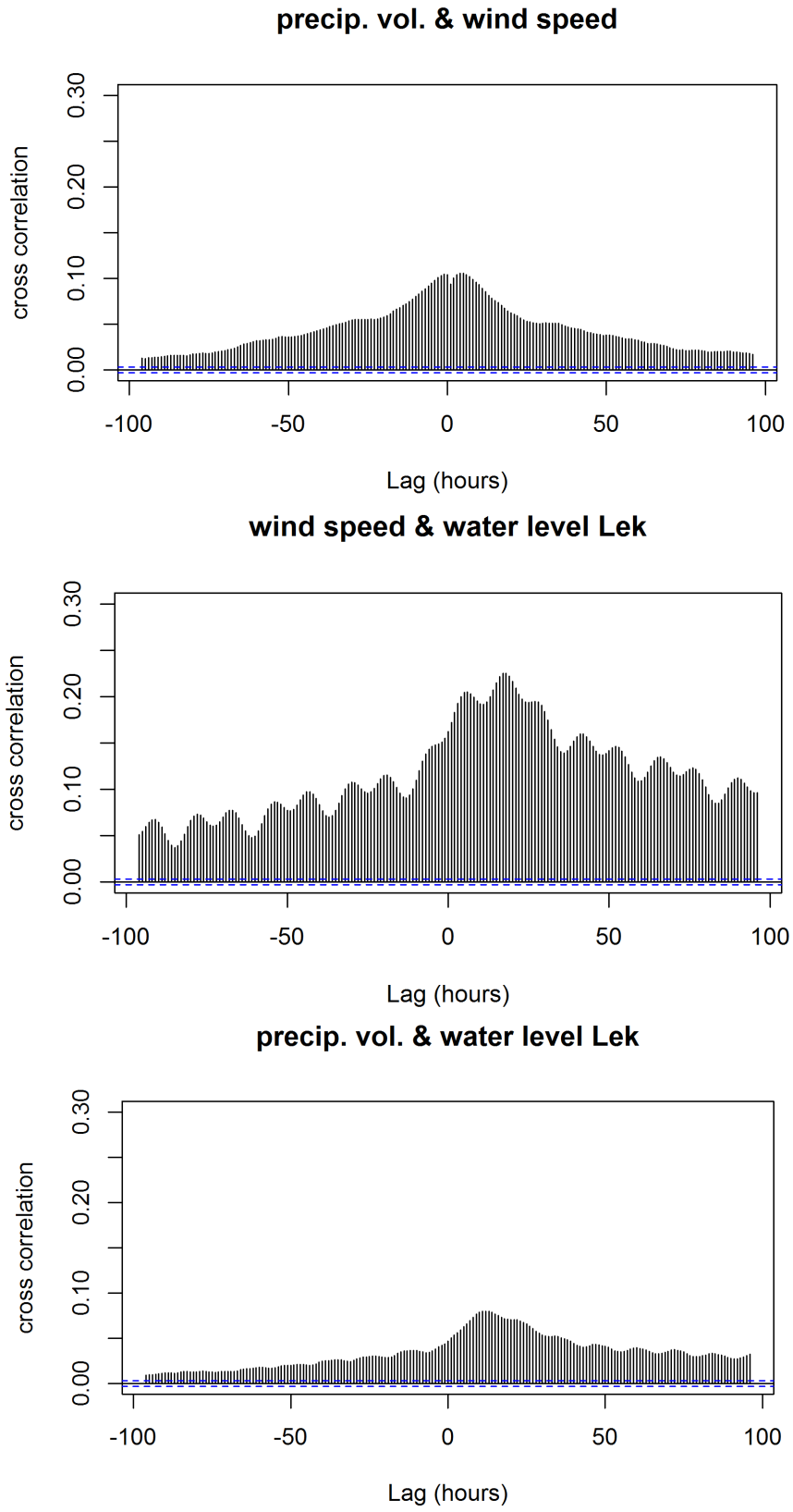


Figure 46: Cross-correlation plots where the 2nd mentioned variable in the title is shifted with a maximum of 100 hours. For all observations, the hourly values are used. Blue dotted lines represent the critical value for significant cross-correlation (α = 0.05).

E. Univariate precipitation volume frequency analysis

To derive the univariate frequency of the variable '9-day precipitation volume', extreme value statistics are applied as the interest lies on return periods up to 1000 years. As this analysis is only about one variable, it is not necessary to deal with issues such as choosing low thresholds to gain information about the dependence structure. There are two main methods to derive extreme frequencies, the Block Maxima and Peak Over Threshold (POT) techniques (Coles, 2001)

The Block Maxima approach uses the maximum value in a certain period, e.g. the annual maxima. The Peak Over Threshold (POT) approach uses all observations above a certain threshold. The POT approach has some advantages over the block maxima approach, as it is less wasteful than the Block Maxima approach (Coles, 2001; Far & Khairi Abd. Wahab, 2016; Ribatet, 2011). In addition, it is possible that the second largest peak in a year is larger than the highest peak in another year, which would be neglected in the Block Maxima approach. Therefore, the Peak Over Threshold approach is used here.

The first step is to select a threshold, such that the behaviour is described by the Generalised Pareto Distribution (GPD). A too low threshold results in a bias, while a too high threshold results in high variance for extreme estimates. The threshold selection approach as described and recommended in Coles (2001) is used here. The approach is fitting the GPD using multiple thresholds and look at the stability of the modified scale and shape parameter of the distribution against the threshold. This fitting process is based on Maximum Likelihood estimates as is recommended in Coles (2001). The result is visible in Figure 47. All calculations and plots here, are executed with the help of the **R** package *POT* (Ribatet, 2011).

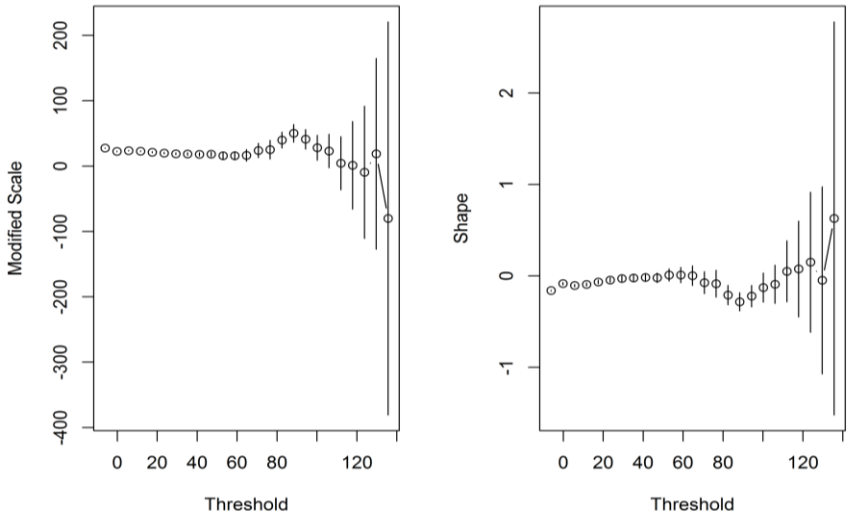


Figure 47: Threshold selection by analysing the stability of the modified scale and shape parameters of the Generalised Pareto Distribution.

Until about a threshold of 60 mm, the distribution parameters are stable. Therefore, this threshold is proposed to use for further analyses. This threshold gives about 3 events per year, which is a typical number of observations in a Peak Over Threshold setting. Using this threshold and plotting its results against the empirical probabilities, see Figure 48, shows that the use of the Generalised Pareto distribution, the selected threshold and its estimated

parameters is appropriate. In addition, trying different thresholds did not affect the results significantly.

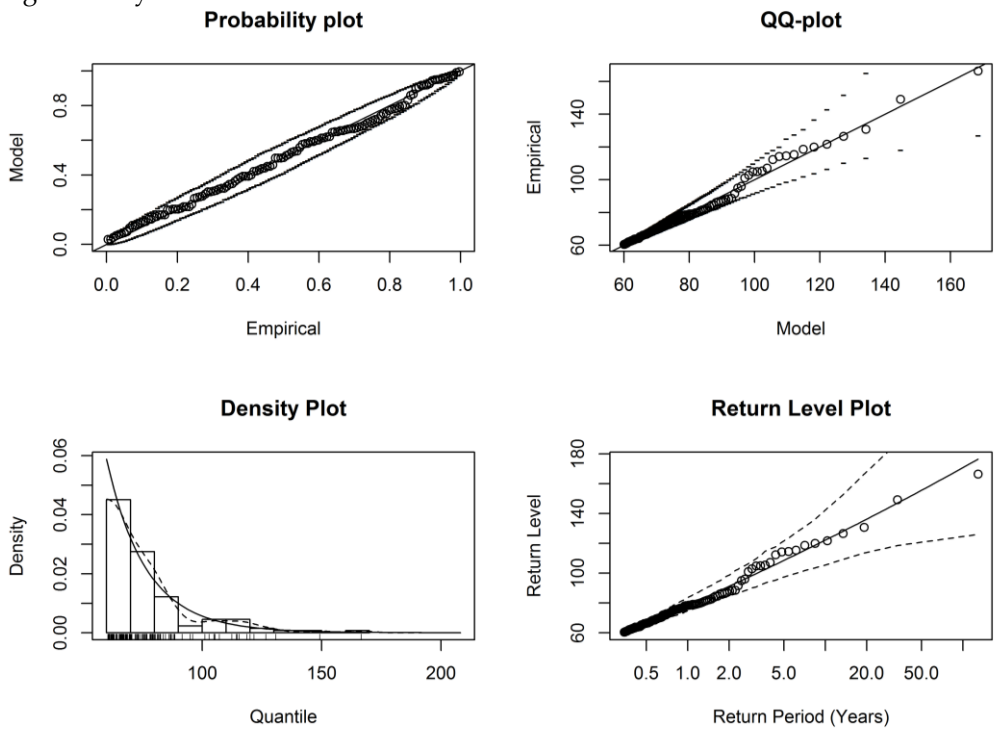


Figure 48: Univariate precipitation volume frequency analysis showing the comparison of the probabilities, quantiles (QQ-plot), density plot and return level plot (dotted line showing 95% confidence interval)

From this distribution, the frequencies of occurrence of precipitation volumes within a year are derived to compare it to previous studies. In addition, the probabilities in a 9-day period are derived. The result is visible in Table 30. Following (HydroLogic, 2018b), a range up to a return period of once in 1000 years is derived. Compared to previous analyses, two new classes (185-195 & 195-205 mm) are added as the derived frequencies are different. This could be the result of using a different measuring station for these statistics, as previous studies used the derived probabilities for de Bilt, while the dataset of Rotterdam airport is used here. As shown in Buishand et al. (2009), the Rotterdam area shows higher multiple-day precipitation volume extremes than de Bilt, which is consistent with this analysis.

Table 30: Exceedance frequency of precipitation volume

Precipitation volume [mm]	Value in model [mm]	Frequency of occurrence [jr^{-1}]	Probability in 9 days [-]
65 - 85	75	1.4703	0.049260
85 - 105	95	0.4620	0.012334
105 - 125	115	0.1539	0.003893
125 - 135	130	0.0338	0.000837
135 - 145	140	0.0203	0.000502
145 - 155	150	0.0124	0.000305
155 - 165	160	0.0076	0.000187
165 - 175	170	0.0047	0.000116
175 - 185	180	0.0029	0.000073
185 - 195	190	0.0019	0.000046
195 - 205	200	0.0012	0.000029
205 - 215	210	0.0008	0.000019

F. Event selection from time series

This appendix refers to the event selection process in Chapter 4. The first step in constructing the joint probability distribution for the two or more variables of interest is to select the data to include in this analysis. Individual events are selected, as in general independent observations are necessary to construct a (joint) probability distribution (Nelsen, 2006). Independent events are selected based on a threshold and minimum time between peaks. As this study focuses on the effects of dependency, a low threshold is set to preserve the dependence structure. Independence was checked by checking the autocorrelation plots. In some cases, autocorrelation plots still showed some significant correlation. This is partly caused by seasonality. This serial dependency could be further reduced by dividing the events per season or per month, but this is not done in this study. As the focus in this study is on the dependence structure, the requirement of independent events is not strictly taken. This is also discussed in Paprotny et al. (2018), who studied compound flood potential in Europe. They compared the dependence structure using all observations of river discharges and storm surges with only using the upper 5% of observations and found that the dependence structure is much better represented using all observations, which gave more realistic return periods of compound floods. In this study, a compromise is used by setting a low threshold and minimum inter-event time to preserve the dependence structure while the autocorrelations plots show a substantial decrease in serial correlation.

To find all relevant events for the first combination of wind speed and water level Lek (see Chapter 4), a threshold is set for each of the variables, to prevent that events with the combination of a high value for one of the variables and a low value for the other are missed (Mazas & Hamm, 2011; Li et al., 2014). Wind events are selected based on a threshold of 6 m/s (moderate wind speed, 4 on the scale of Beaufort) and a minimum inter-event time of 24 hours. Wind direction is selected based on the timing of the peak value of the wind speed. The maximum coupled water level is selected within a time frame of 72 hours based on Figure 10. Water level events are selected based on a threshold of 0.8 m NAP and a minimum inter-event time of 24 hours. Accompanying wind direction and speed are also coupled within the same time window of 72 hours. Autocorrelation is still found, mainly for the water levels at the Lek (weak correlation of 0.3). It is however reduced compared to using all observations. In the study of Li et al. (2014), a minimum inter-event time for storms surges along the Dutch coast is set at 6 h, but this would not result in a reduced serial correlation in this study. The risk of using a longer inter-event time is that events, where the pumping station at Kinderdijk is impeded, are missed. In addition, as the events are divided by wind direction, the goal was to select a high number of events. In order to take wind direction in account, the dataset is divided in the 4 directions as used for the univariate wind frequency analysis (NW, W, SW and Other directions, see Appendix H). The total number of events for this is 8820 with 1148, 2827, 2972 and 1873 in respectively NW, W, SW and Other directions. For the 'Other' directions, it is not required to fit a copula as there is no interest in the joint probability of wind speeds from these directions and the water level at the Lek as wind from these directions do not affect the water levels in the Alblasserwaard (Arcadis, 2014; HydroLogic, 2018b). Therefore, also the marginal distribution for wind speeds from this direction is not relevant. In addition, there is only a weak correlation of 0.09 found for wind speeds from this direction and the water levels at the Lek. In the other combinations (2, 3 and 4), where precipitation volume is considered as dependent with other variables,

selecting events based on two thresholds is not needed as there is no interest in events where there is no precipitation. Therefore, the same dataset with the same events is used for these combinations. For the third and fourth combination (precipitation and wind speed), the dataset is again divided in the same four directions (NW, W, SW and Other). This is not necessary for the second combination (precipitation and water level Lek). A threshold on precipitation is set at 20 mm and a minimum inter-peak time of 7 days to ensure no serial correlation for this variable. Wind speed is coupled with a time window of 5 days based on the cross-correlation. Water level Lek is coupled with a time window of 9 days as the cross-correlation is preserved over a longer time period than wind speed. Only the autocorrelation plots of water level Lek show a significant correlation (correlation of 0.2) at the first lag. However, when the events are divided over the four wind directions, the serial correlation is not significant anymore. The selection procedure resulted in 488 precipitation events with respectively 38, 215, 207 and 28 events in the directions NW, W, SW and Other. This shows that precipitation events mainly occur with winds from the directions west and southwest. This could impact the uncertainty for the events from the other directions as the number of observations is low.

G. Precipitation distribution over time

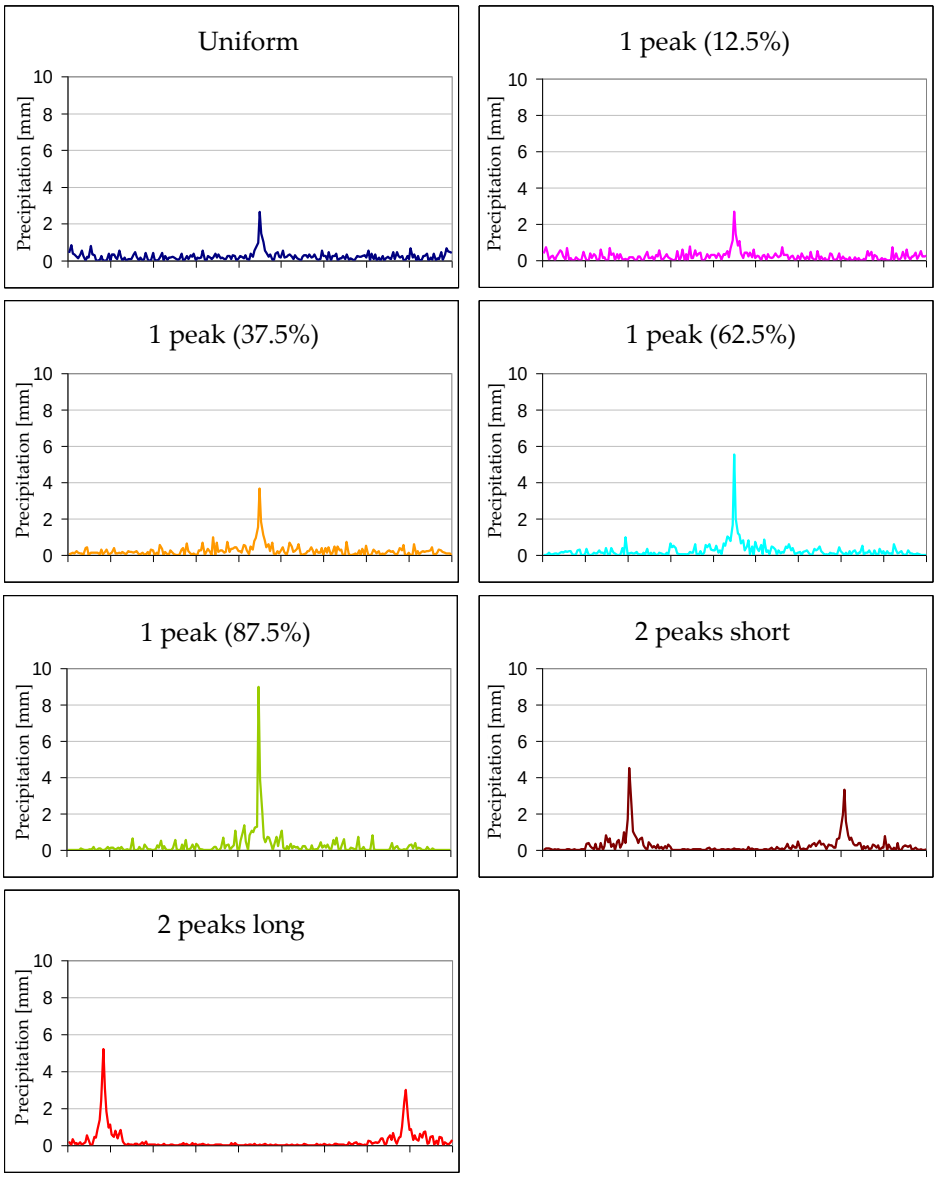


Figure 49: 7 types of precipitation distributions over time (STOWA, 2004; HydroLogic, 2018b). Percentages between the brackets is percentage of volume in peak.

Table 31: Precipitation patterns and its probability considering a 9-day precipitation event (HydroLogic, 2018b; STOWA, 2004)

Precipitation pattern	Probability
A: uniform	0.14
B: 1 peak – 12.5%	0.14
C: 1 peak – 37.5%	0.14
D: 1 peak – 62.5%	0.14
E: 1 peak – 87.7 %	0.14
F: 2 peak-short	0.18
G: 2 peak-long	0.12

H. Univariate wind frequency analysis

The water levels in the Alblasserwaard are affected by winds from the southwest to northwest (Arcadis, 2014). Winds from other directions can also affect the gradient of the water surface, but only in very extreme cases (return period > 3000 years). Therefore, only the wind speeds from northwest to southwest directions are of interest. For these directions, only the wind speeds above 10 m/s affect the maximum water levels significantly (> 2 cm) (Arcadis, 2014).

To derive the probability of the maximum wind speed in a 9-day period, the same approach as discussed in Chapter 4, about the marginal distributions is used. First, the data is aggregated over 9 days. Then, a range of potential probability distributions are tested using the AIC & BIC criteria and additionally a visual comparison took place. Potential distributions for wind speeds are the same as selected in Chapter 4, as these are also accepted to use for wind speeds (Sohoni et al., 2016; Masseran et al., 2013). The probabilities are derived for three directions (NW, W and SW), following the recommendations in Arcadis (2014). The resulting AIC/BIC scores of the four best scoring distributions can be found in Table 32. The visual comparisons between the empirical and parametric distributions can be found in Figure 50 - Figure 52.

Table 32: potential probability distributions and resulting AIC/BIC scores

Distribution	AIC	BIC
Wind speed (NW) (304°-11°)		
Lognormal	956.7	963.4
Gamma	962.1	968.8
Gumbel	**954.7**	**961.5**
Weibull	1001.8	1008.5
Wind speed (W) (236°-304°)		
Lognormal	**3632.0**	**3641.2**
Gamma	3635.1	3644.2
Gumbel	3640.6	3649.7
Weibull	3720.1	3729.2
Wind speed (SW) (169° - 236°)		
Lognormal	**2864.4**	**2873.2**
Gamma	2867.4	2876.1
Gumbel	2873.7	2882.5
Weibull	2943.6	2952.3

Based on the scores and visual comparisons, the following distributions are selected for the three wind directions, see Table 33. The selected distributions show a good fit, with extra attention given to the higher values (>10 m/s).

Table 33: Selected distributions and its parameters

Variable	Selected distribution	Par 1	Par 2
Wind speed (NW)	Lognormal	2.28 (-)	0.23 (-)
Wind speed (W)	Lognormal	2.39 (-)	0.27 (-)
Water level Lek (SW)	Lognormal	2.37 (-)	0.25 (-)

This study is based on the data of Rotterdam Airport, and not Schiphol Airport as in Arcadis (2014), which results in differences in the probabilities of wind speeds. In Rotterdam, the probabilities on higher wind speeds are higher for northwest directions, while at Schiphol the probabilities on wind speeds are generally higher for the west and southwest directions. The Overwaard area in the Alblasserwaard is in general more susceptible for winds from southwest directions, while the Nederwaard area is also susceptible for winds from the northwest. Therefore, these derived statistics probably have an impact on the derived normative water levels compared to previous studies, which used the derived statistics for Schiphol Airport.

In the study of Arcadis (2014), the wind speed is discretised in 4 classes per direction (10-15 15-20, 20-25, 25+) based on the differences in effect on the water levels. However, as the probabilities of the class of 25+ m/s are too low in this case (Rotterdam Airport data) to impact the normative water levels in the range of T0.5 - T1000, the wind speeds are discretised in 3 classes (10-15 15-20, 20+). The resulting probabilities can be found in Table 34. The derivation of the duration/shape of the wind event to use in the model is performed in Appendix I.

Table 34: resulting probabilities of mean hourly maximum wind speed on a day per direction

Wind direction (model value)	Hourly wind speed [m/s] (model value)	Probability of maximum hourly wind speed in a 9-day period
Northwest (337.5°)	12	0.0503
	17	0.0036
	22	0.0001
West (270°)	12	0.2034
	17	0.0424
	22	0.0048
Southwest (202.5°)	12	0.1684
	17	0.0258
	22	0.0019
Other directions/low speed	-	0.4994

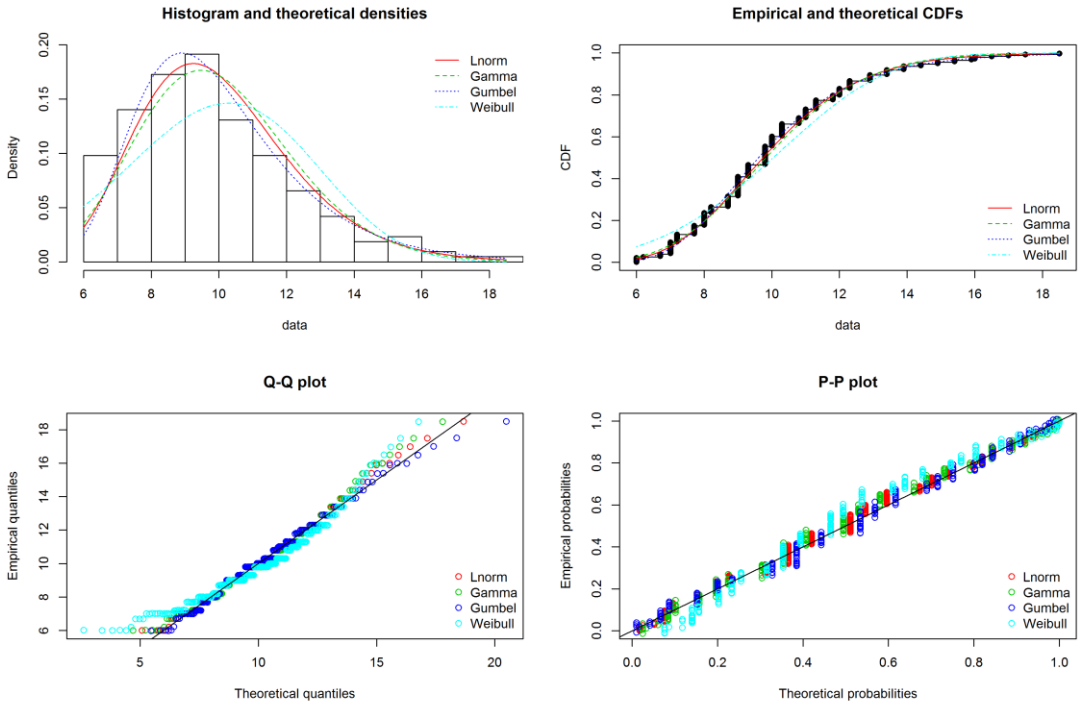


Figure 50: Comparison empirical and parametric probability distributions for wind direction NW

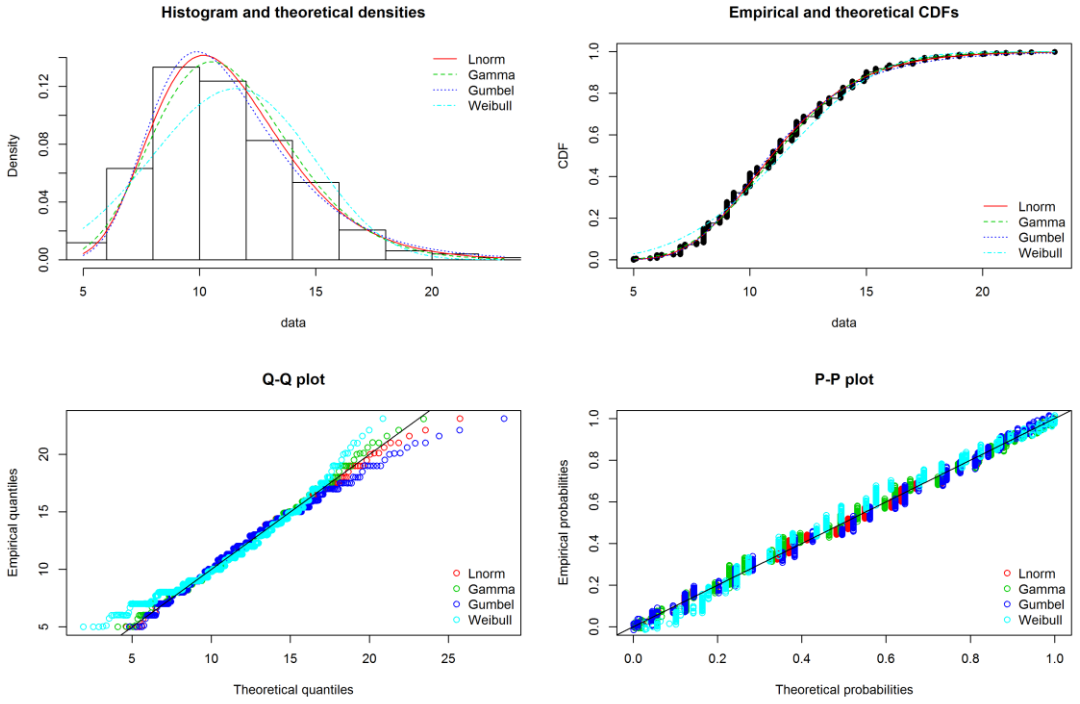


Figure 51: Comparison empirical and parametric probability distributions for wind direction W

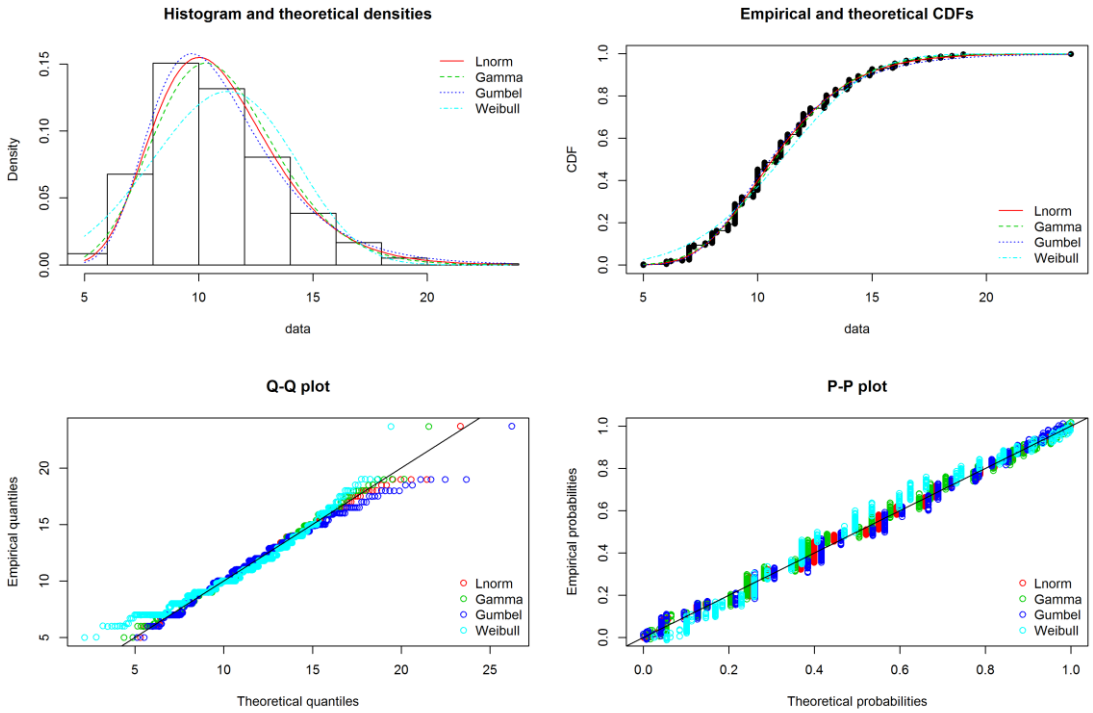


Figure 52: Comparison empirical and parametric probability distributions for wind direction SW

I. Duration of wind analysis

As the duration of a wind event is not seen as a stochastic variable in this study, a typical duration is selected. In previous studies (Arcadis, 2014; HydroLogic, 2018b), the shape of a wind event is defined as a trepeziod with 4 hours top wind speed and 24 hours for the increase to the top wind speed and 24 hours for the decrease to 0 m/s. This shape is independent of its top wind speed. In this study, this shape is refined and adapted to the data that is used in this study.

The duration is linked with the top hourly wind speed, as there is a certain relation between those two variables. Wind events are selected from the historical records based on a threshold and an inter-event time of 8 hours to ensure independency. Three threshold are defined: 10 m/s, 15 m/s and 20 m/s, similar to the categories of which the frequency of wind speeds are derived. The results can be found in Figure 53.

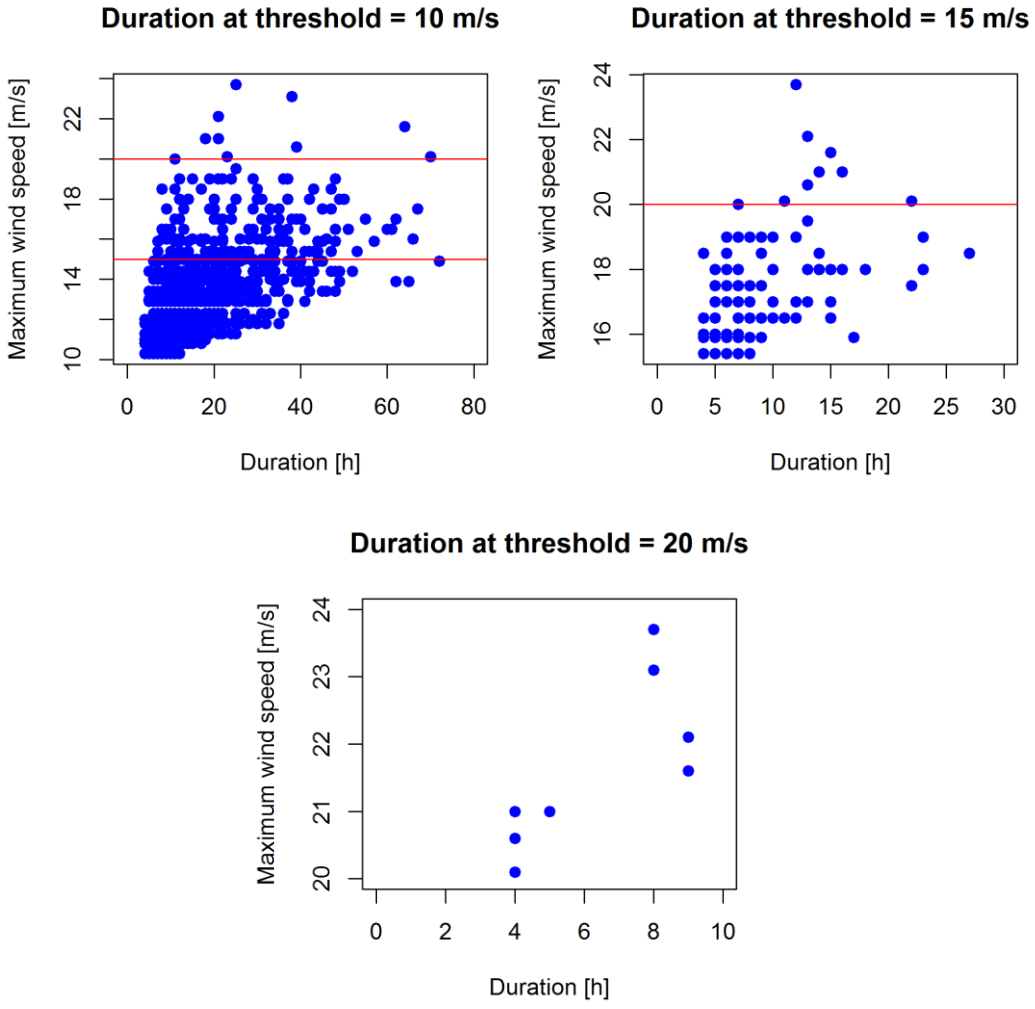


Figure 53: Analysis of duration of wind events with maximum value and duration over the named threshold

Based on these results, typical shapes for the wind events are determined. In Table 35 is visible, what the rounded average duration in that interval is, which is used to derive wind shapes.

Table 35: Duration over threshold based on maximum hourly wind speed

Interval	Representative value	Duration above 10 m/s [h]	Duration above 15 m/s [h]	Duration above 20 m/s [h]
10 - 15 m/s	12 m/s	20	-	-
15 - 20 m/s	17 m/s	35	10	-
20+ m/s	22 m/s	45	15	7

This results in the following wind events to use in the model based on a trapezoid shape, see Figure 54. This is placed in the middle of a 9-day precipitation events as a result of the strongest cross-correlation at lag 0 (see Figure 10).

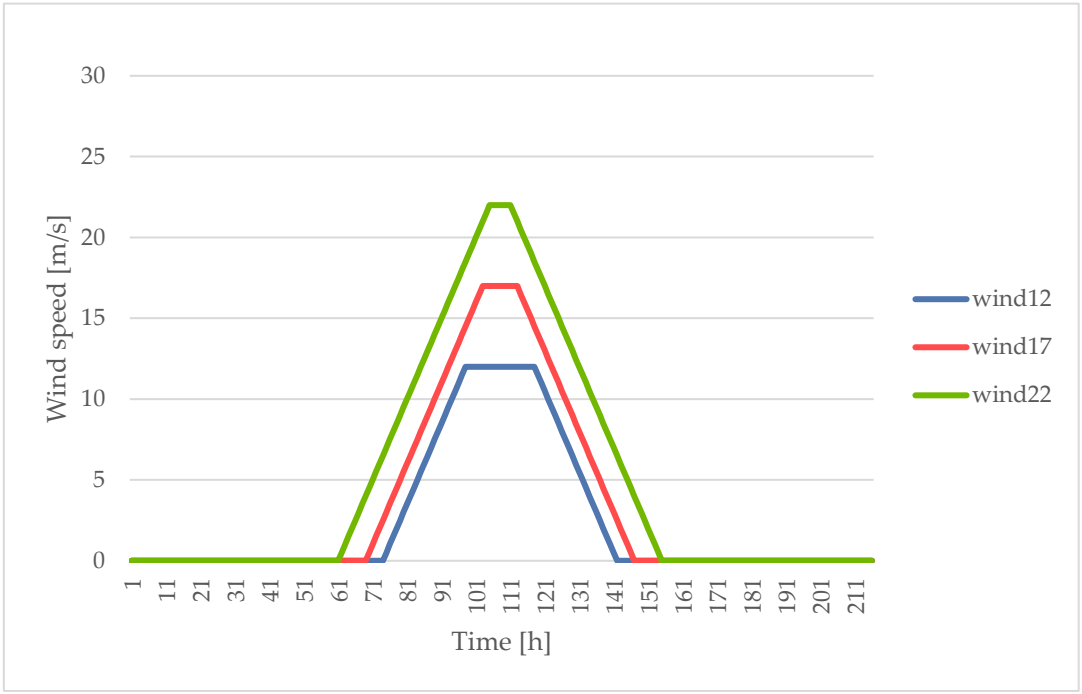


Figure 54: Shapes wind events based on top wind speed (12, 17 and 22 m/s)

J. Duration of blockage analysis

To determine the probabilities of the duration of a blockage of the pumping stations, historical events over the threshold of 2.25 m NAP on the Lek have been analysed. The minimum inter-event time has been set at 8 hours. As the threshold is high (only 28 events selected), the parameter of minimum inter-event time does not affect the number of selected events significantly. In Figure 55 the maximum water level and the duration over the threshold of the events is visible. Visible is that most events have a duration of 1-3 hours over the threshold. Only 2 events of 4 hours have been measured.

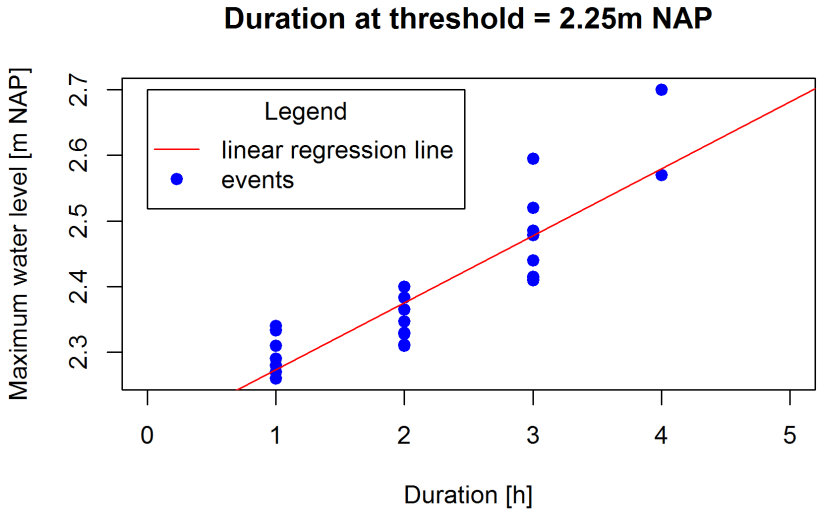


Figure 55: Analysis duration water level Lek with maximum water level above 2.25 m NAP. Red line is the linear regression (Maximum water level ~ Duration over threshold)

As there are only 28 events with 4 different values, it is not appropriate to try to fit a probability distribution of the duration over threshold. Therefore, in the case of independence, the empirical probability of an event in a 9-day period is simplified by counting the number of events and use the Poisson distribution to determine the probability of occurrence in 9 days. The Poisson distribution is used to determine the probability of a number of rare independent events given an expected frequency of occurrence (Hu, 2008). The Poisson distribution is given by the equation:

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

With k is the number of events and λ is the expected value, which is determined by counting the number of events and divide it by the length of the measured period. The probability of one or more events in 9 days is then equal to 1-(P(0 events))$^{9\ \text{days}}$.

Table 36: Analysis historical blockage events

Interval	Nr. of events	Probability in a 9-day period [-]	Model value [h]
No blockage	-	0.9847	0
1 - 2.5 h	18	$P = 1-(e^{-\left(\frac{18}{16436}\right)})^9 = 0.0098$	2
2.5+ h	10	$P = 1-(e^{-\left(\frac{10}{16436}\right)})^9 = 0.0055$	4

In the case of joint probabilities, this cannot be used as only the probability on a certain water level is available. Therefore, a simplified relation between the maximum water level and the duration of the event is derived. A linear regression based on the method of least squares is applied with the **R** function *lm()*. The resulting fitted linear line is visible in Figure 55.Using this regression line, it is possible to predict the duration of an event based on the maximum water level. Events with water levels between 2.25 and 2.45 m NAP are given the duration of 2 hours, while water level above 2.45 m NAP are set to have a duration of 4 hours.

K. Marginal distributions combination one

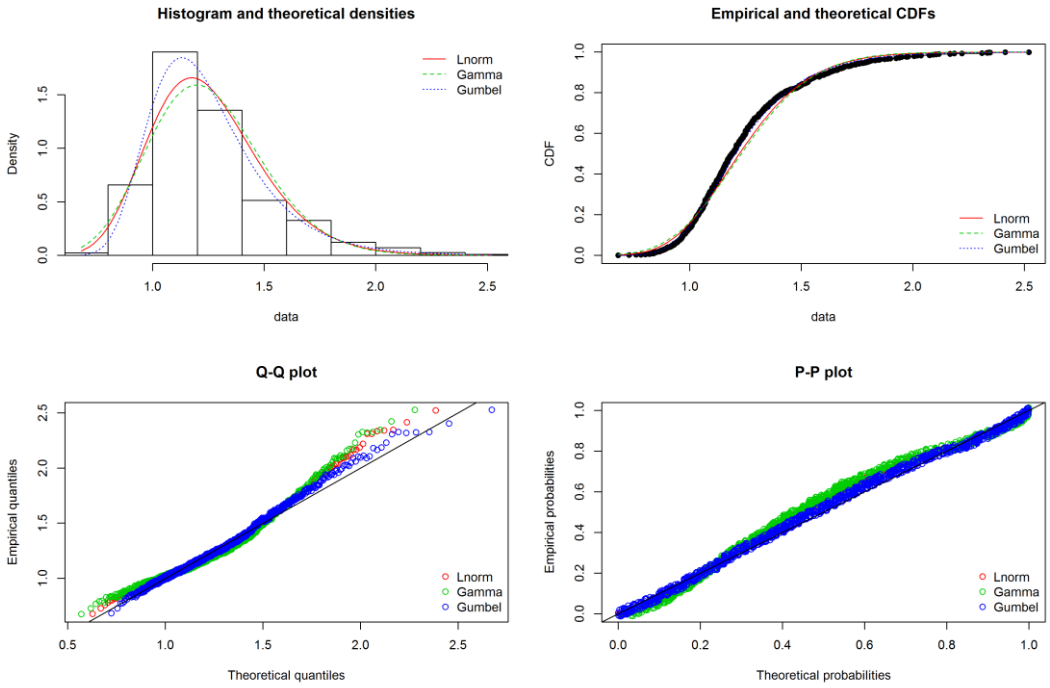


Figure 56: Marginal distributions water level Lek for direction NW

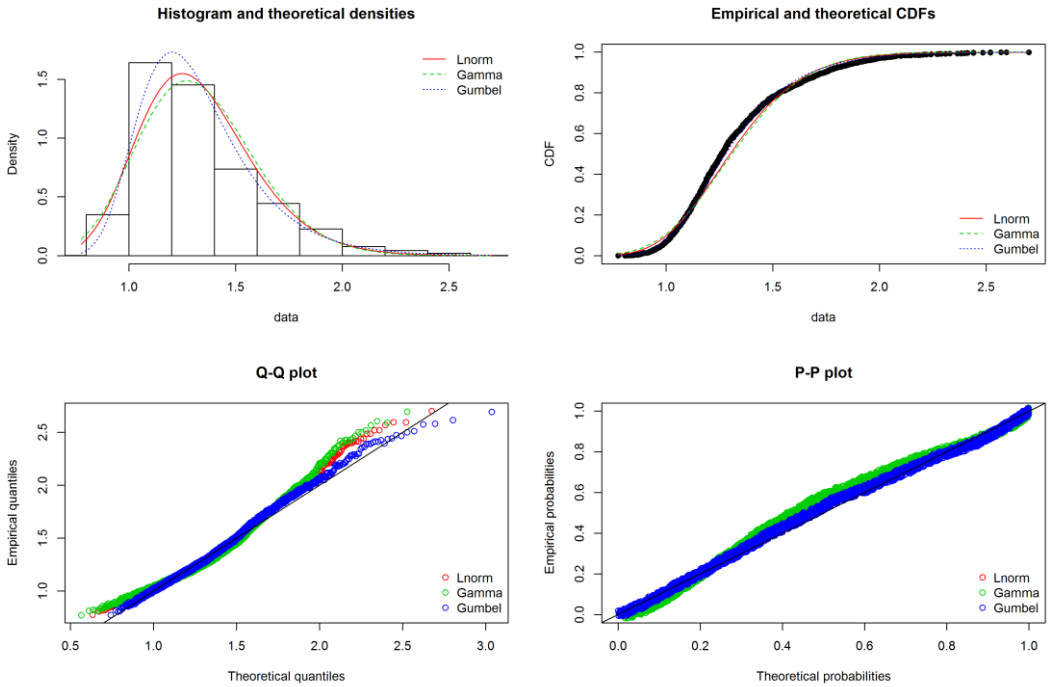


Figure 57: Marginal distributions water level Lek for direction W

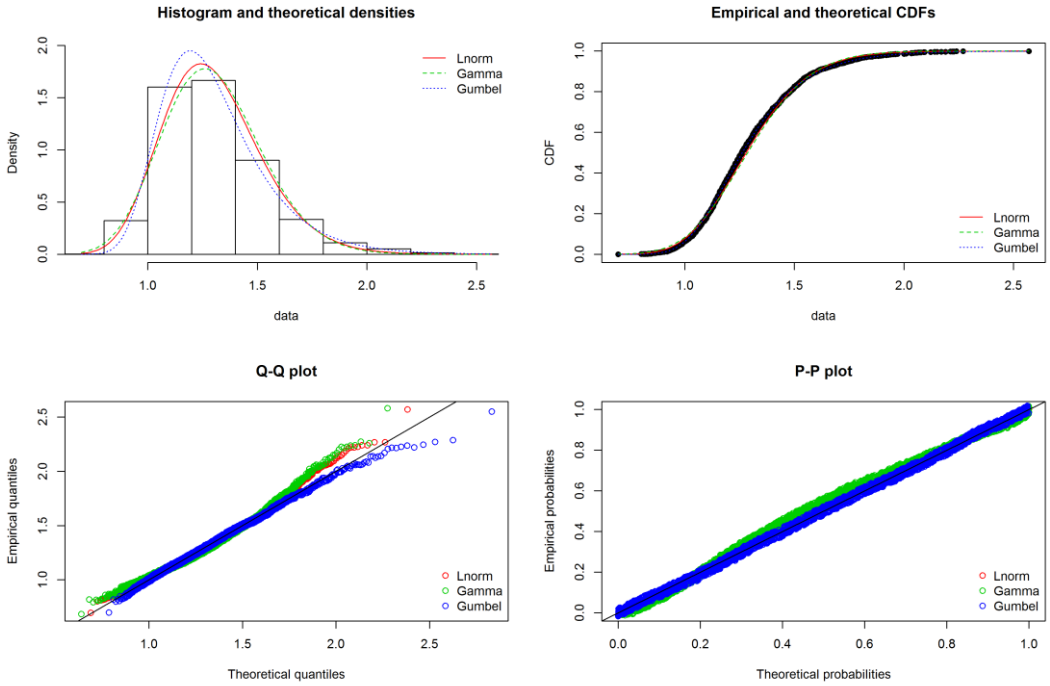


Figure 58: Marginal distributions water level Lek for direction SW

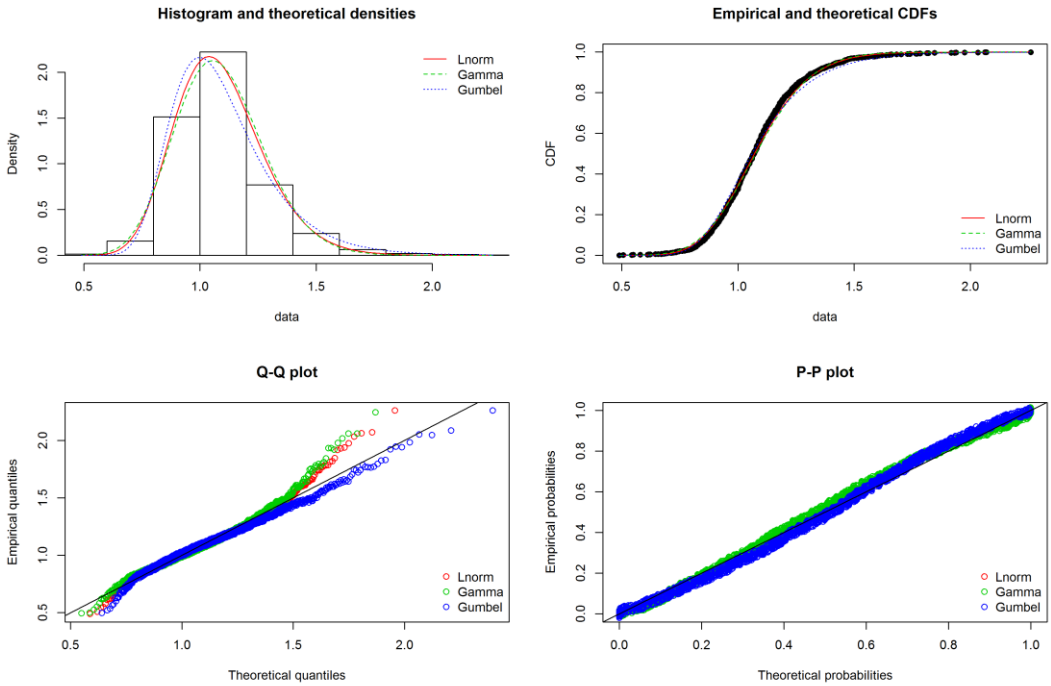


Figure 59: Marginal distributions water level Lek for direction Other

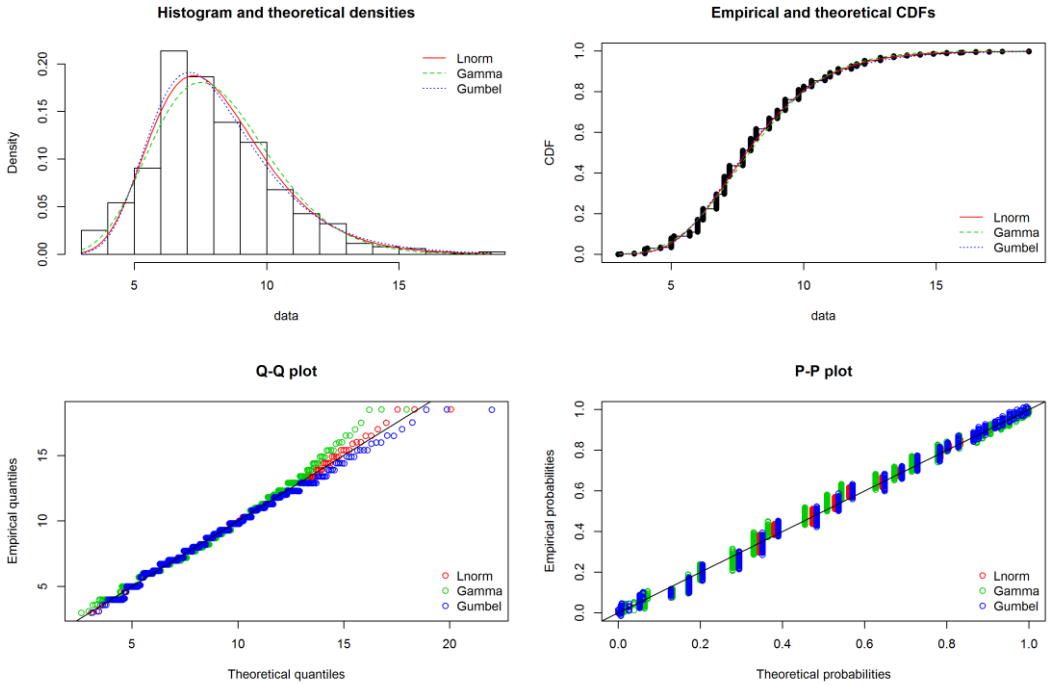


Figure 60: Marginal distributions wind speed for direction NW

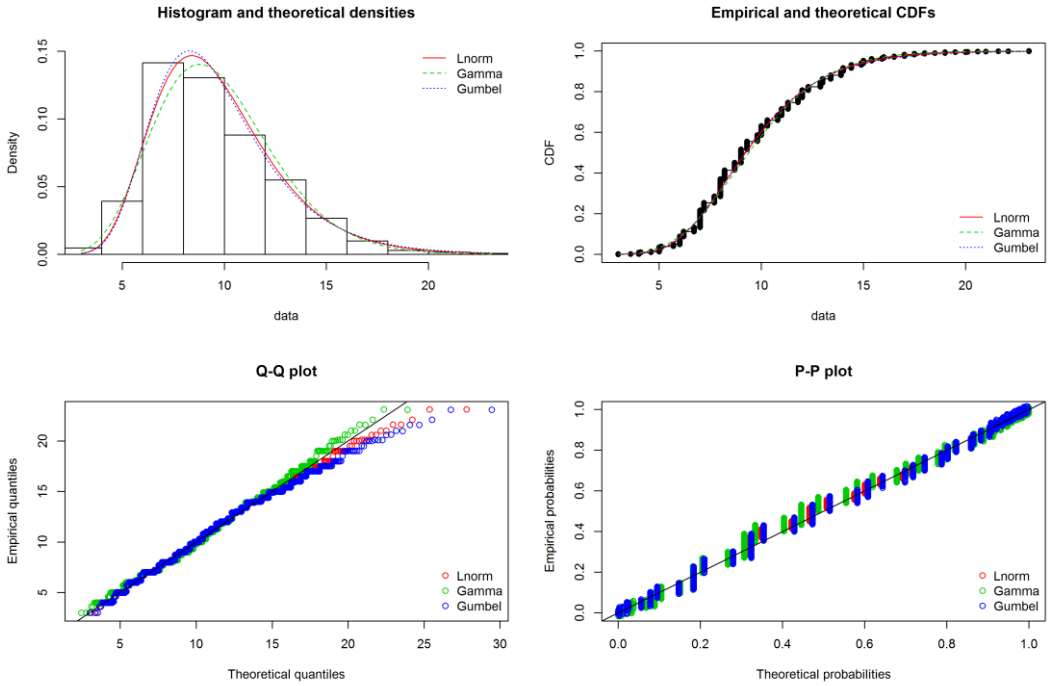


Figure 61: Marginal distributions wind speed for direction W

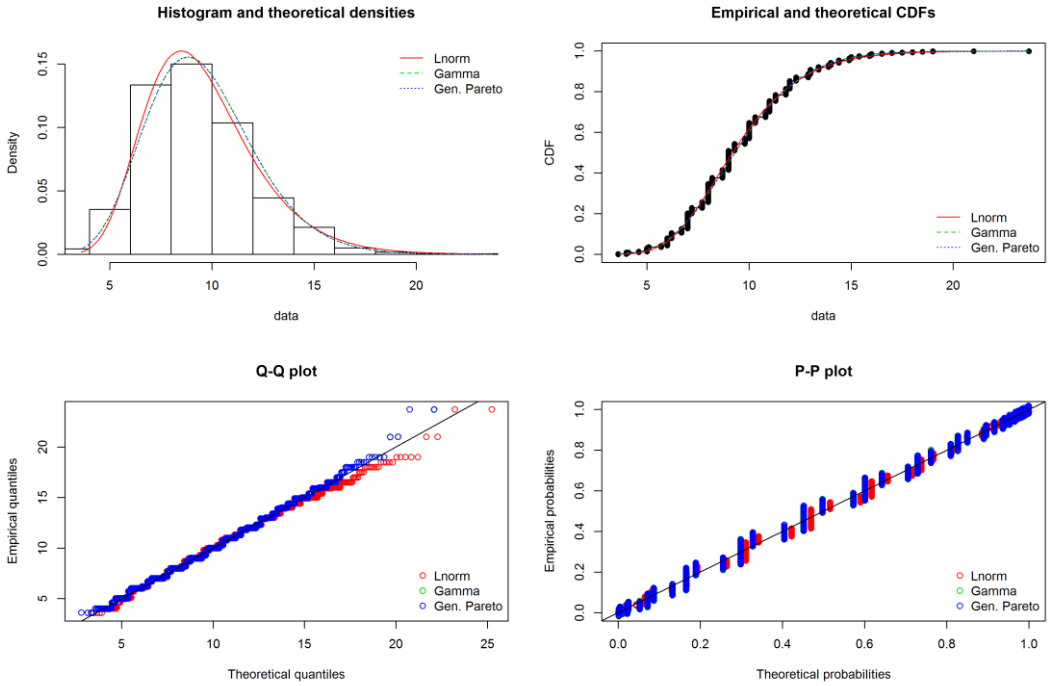


Figure 62: Marginal distributions wind speed for direction SW

L. Marginal distributions combination two

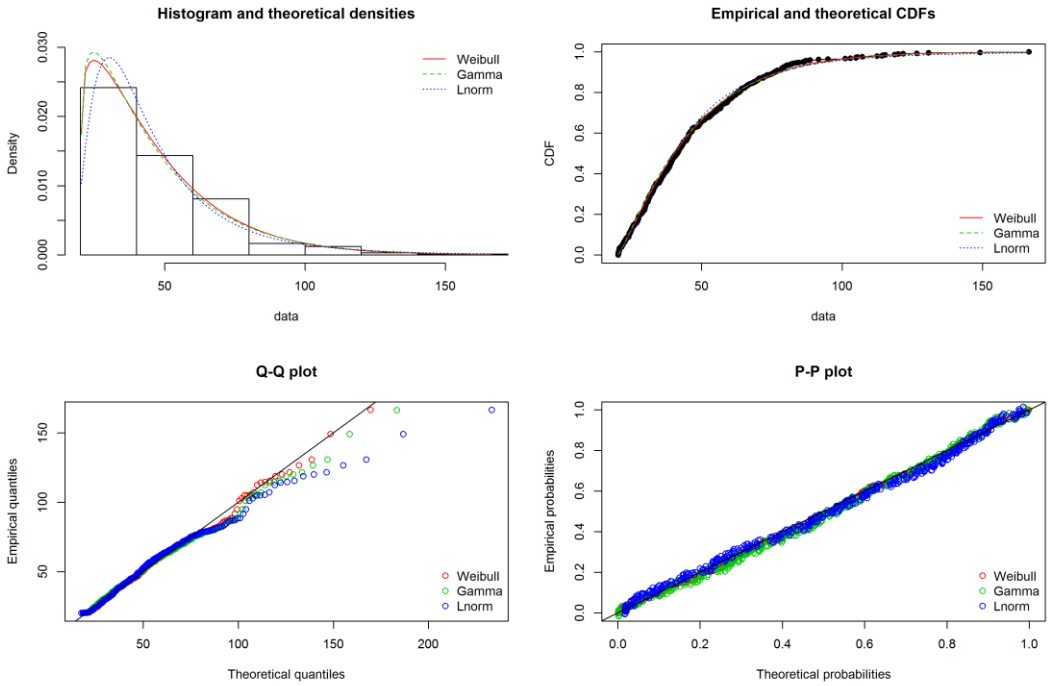


Figure 63: Marginal distributions precipitation volume

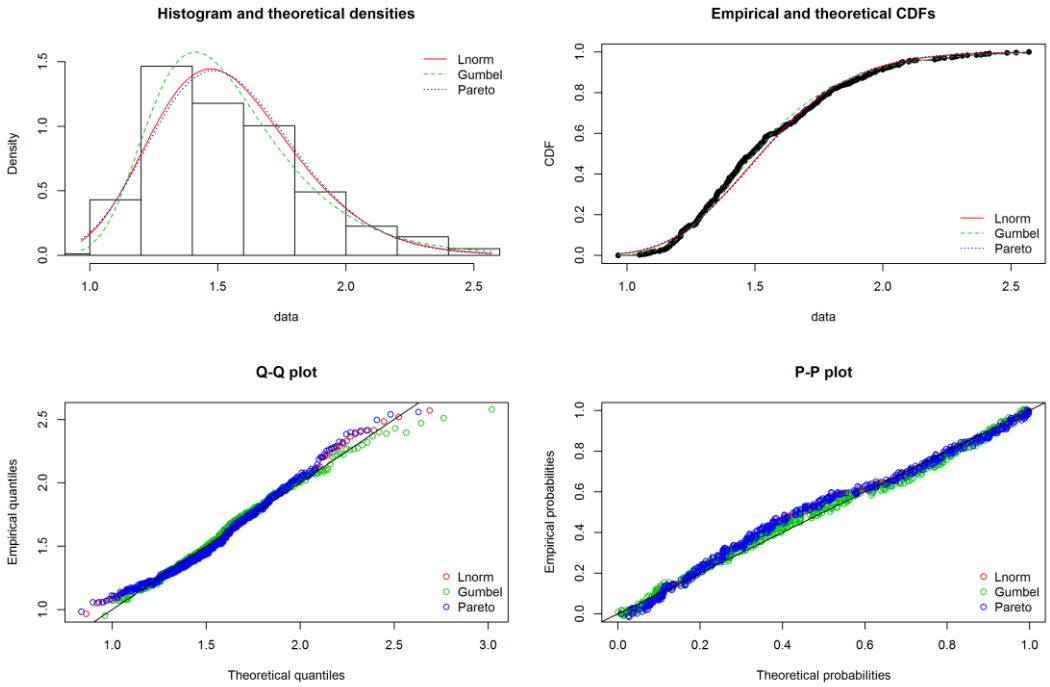


Figure 64: Marginal distributions water level Lek

M. Marginal distributions combination three

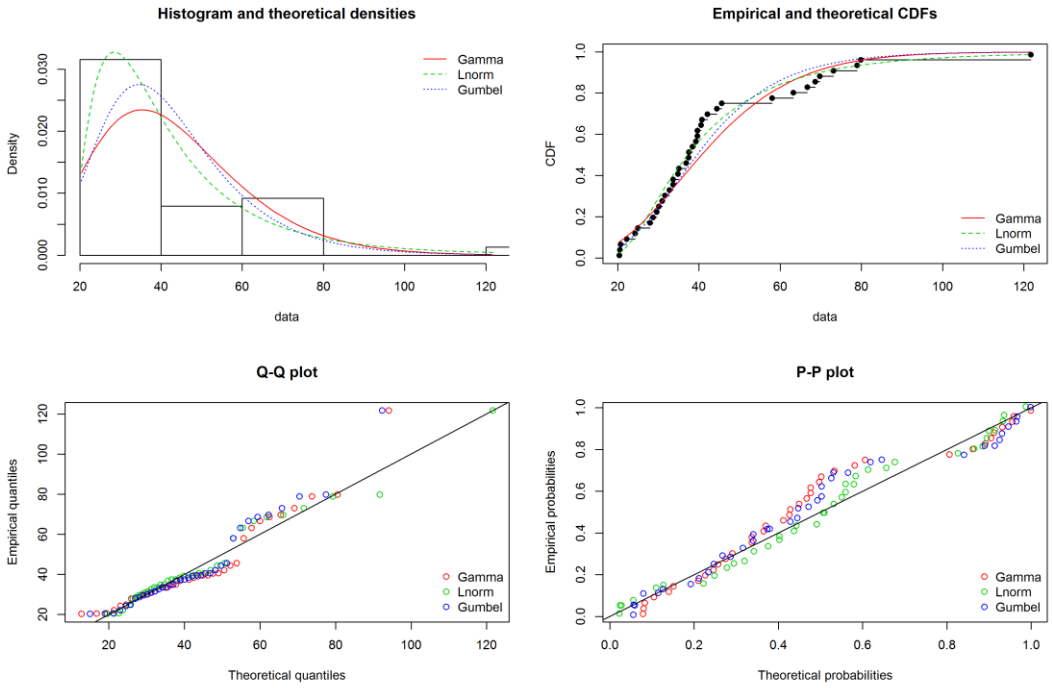


Figure 65: Marginals distributions precipitation (NW)

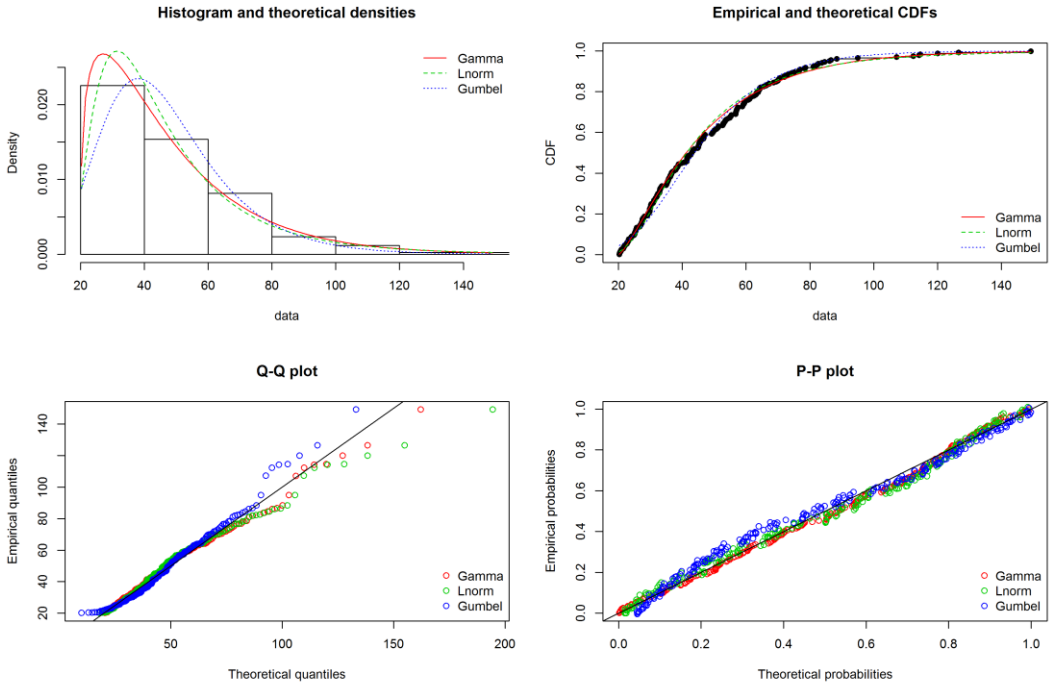


Figure 66: Marginals distributions precipitation (W)

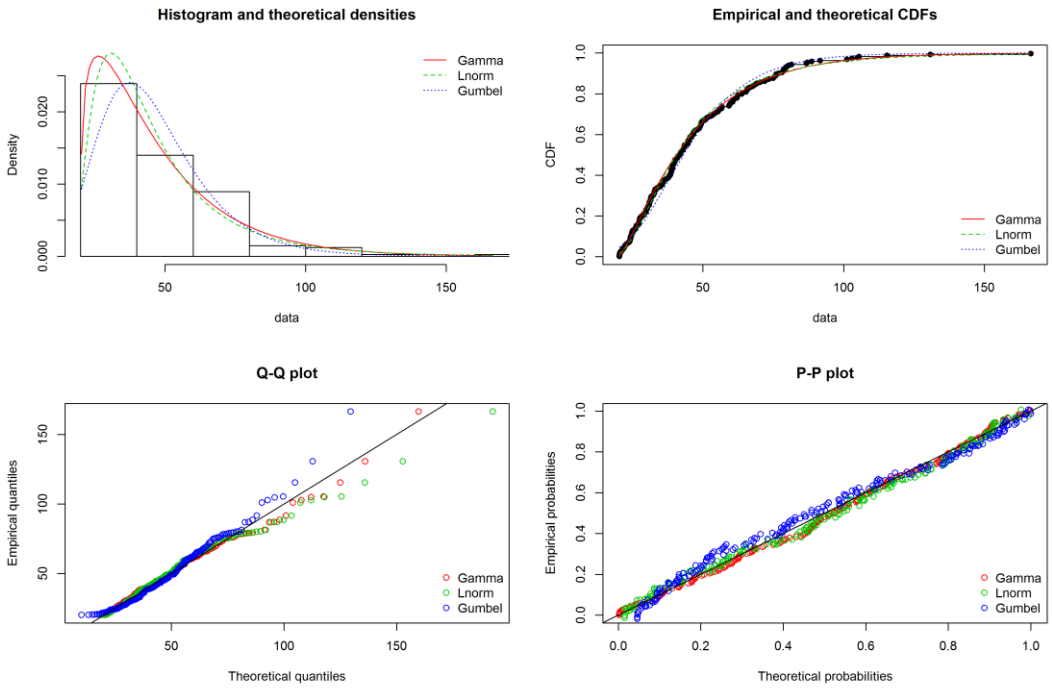


Figure 67: Marginals distributions precipitation (SW)

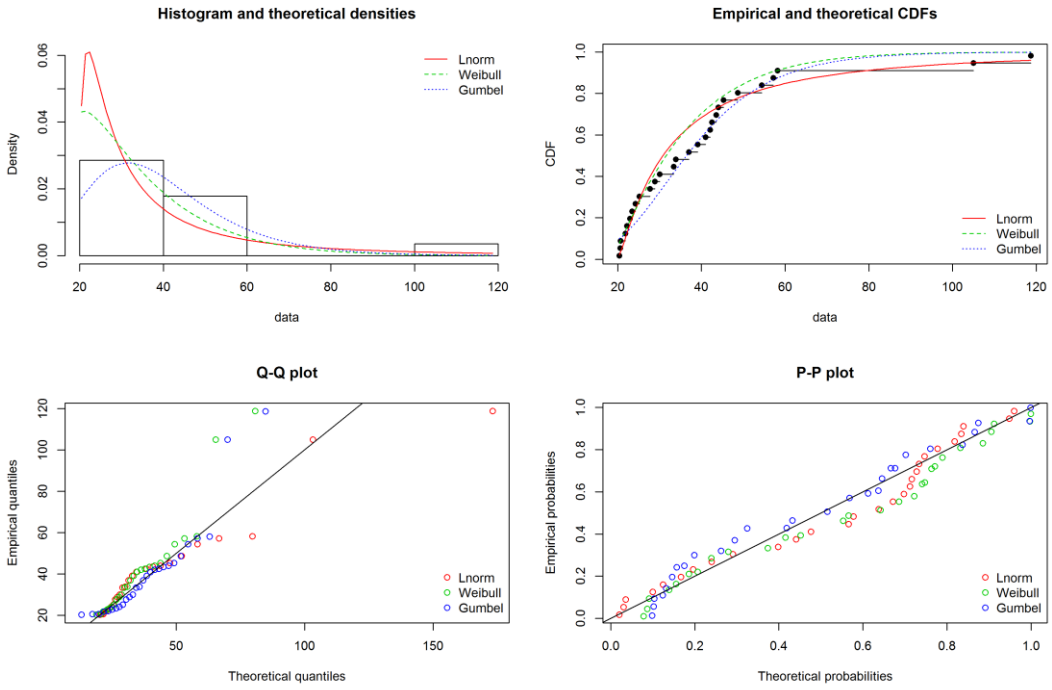


Figure 68: Marginals distributions precipitation (O)

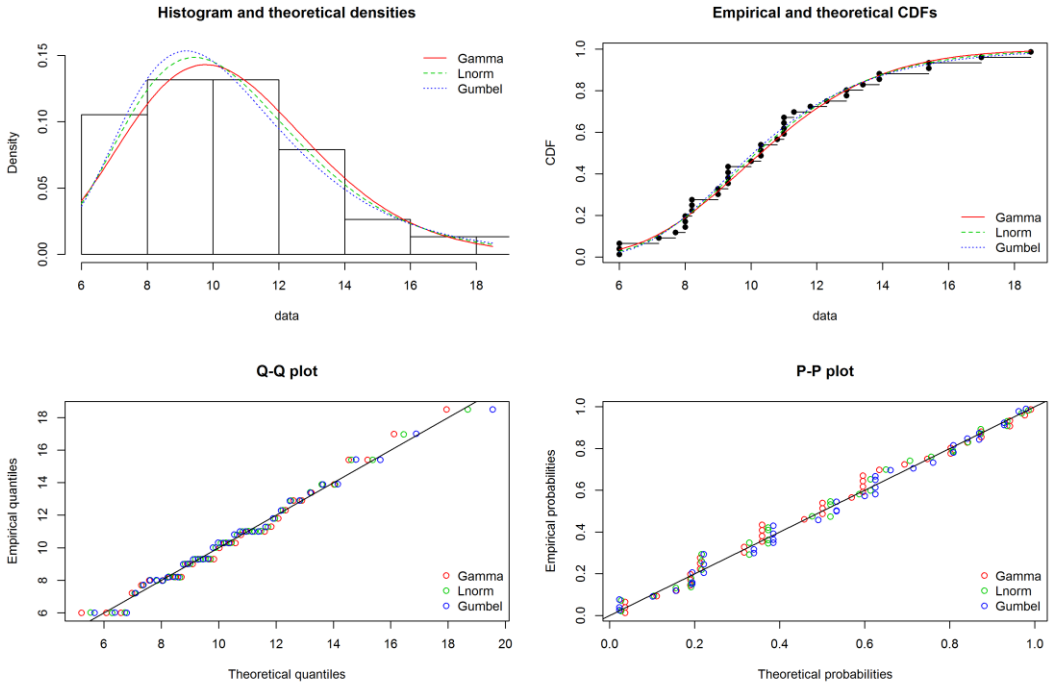


Figure 69: Marginal distributions wind speed (NW)

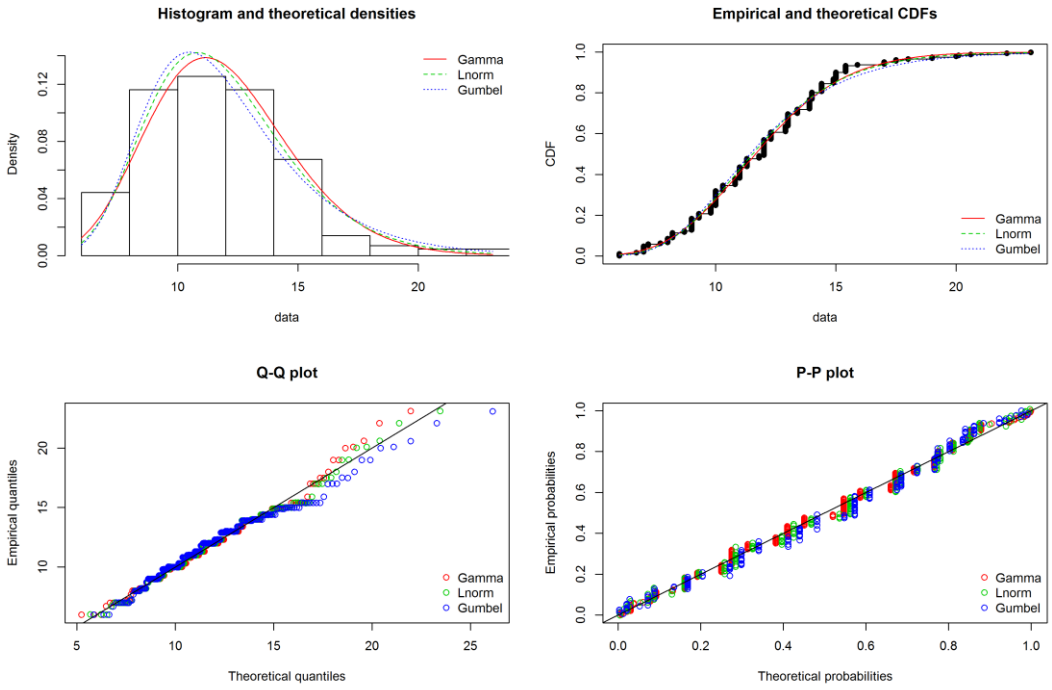


Figure 70: Marginal distributions wind speed (W)

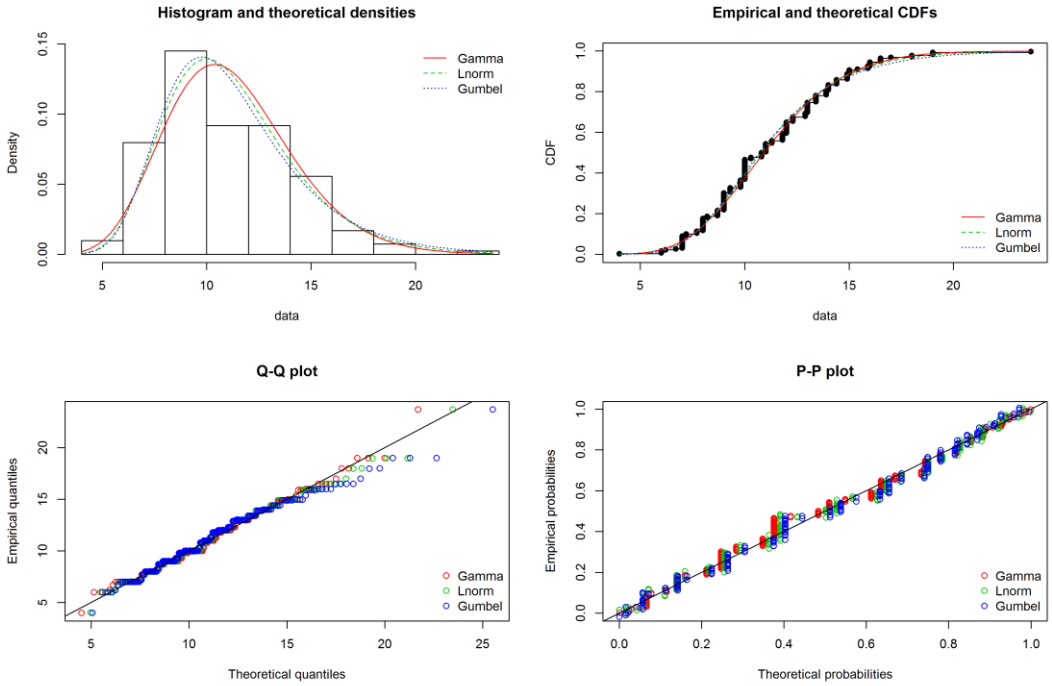


Figure 71: Marginal distributions wind speed (SW)

N. Marginal distributions combination four

To select the appropriate marginal distribution for the variable water level Lek for each wind direction, first the AIC/BIC scores are determined. Then, the top 3 scoring distributions are visually compared to the observations to check whether the distributions fit the data well, and if the visual comparison agrees with the AIC/BIC scores. The scores are visible in Table 37.

Table 37: Candidate marginal distributions water level Lek and AIC/BIC scores. Bold numbers are the best scores (lowest value).

Distribution	AIC	BIC
Water level Lek (NW)		
Lognormal	17.5	20.8
Gumbel	**17.1**	**20.4**
Gamma	18.1	21.4
Water level Lek (W)		
Lognormal	83.9	90.7
Gumbel	**77.4**	**84.2**
Gamma	89.1	95.8
Water level Lek (SW)		
Lognormal	53.1	59.7
Gumbel	**44.9**	**51.5**
Gamma	59.4	66.1
Water level Lek (O)		
Lognormal	-12.9	-10.2
Gumbel	**-13.2**	**-10.6**
Gamma	-12.6	-9.9

The visual comparison of the marginal distributions can be found in Figure 72 - Figure 75. For the direction NW, the Gumbel distribution scores as best, but for higher values, the Gumbel distribution seems to overestimate probabilities. Therefore, the second-best scoring distribution, Lognormal, is selected as most appropriate. The same applies for the wind directions W and Others. For the direction SW, the best scoring distribution (Gumbel) is selected as the other distribution seem to underestimate the range of 2.0 - 2.4 m NAP. The resulting distributions and its parameters are shown in Table 38.

Table 38: Selected marginal distribution per wind direction with distribution parameters

Variable	Selected distribution	Parameter 1	Parameter 2
Water level Lek (NW)	Lognormal	0.41 (-)	0.19 (-)
Water level Lek (W)	Lognormal	0.44 (-)	0.19 (-)
Water level Lek (SW)	Gumbel	1.41 m	0.23 (-)
Water level Lek (O)	Lognormal	0.30 (-)	0.13 (-)

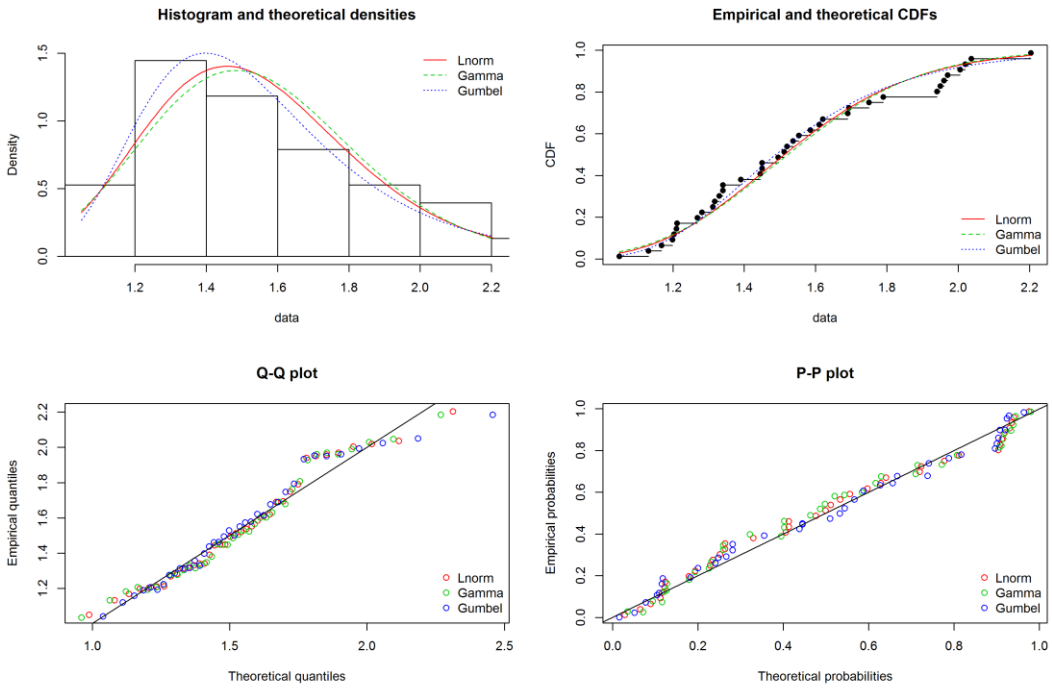


Figure 72: Marginals distributions water level Lek (NW)

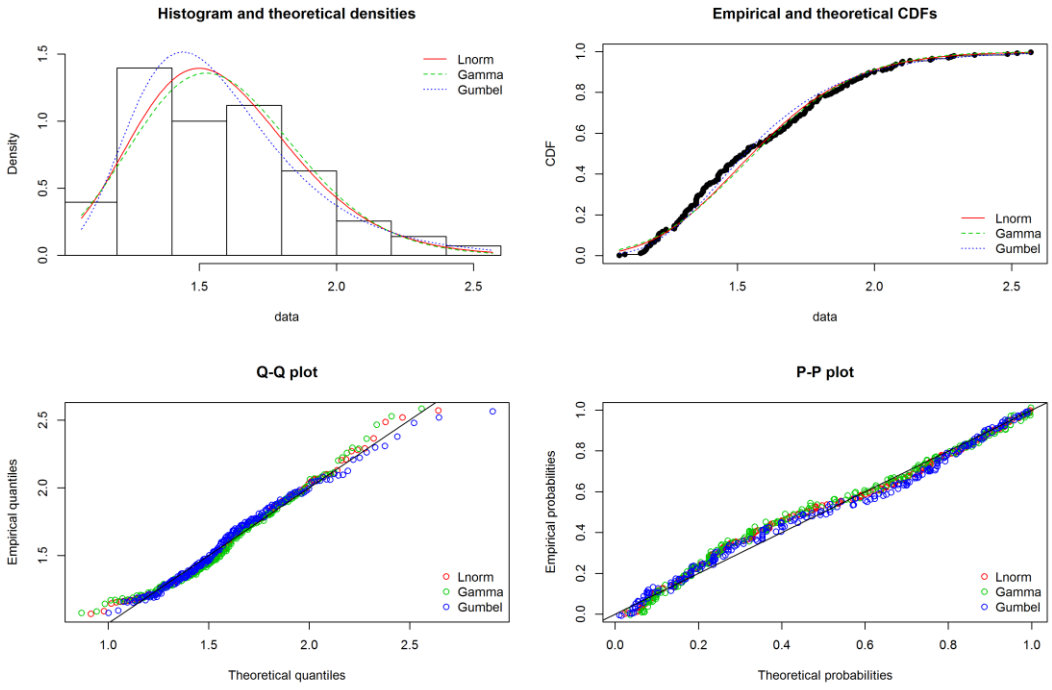


Figure 73: Marginals distributions water level Lek (W)

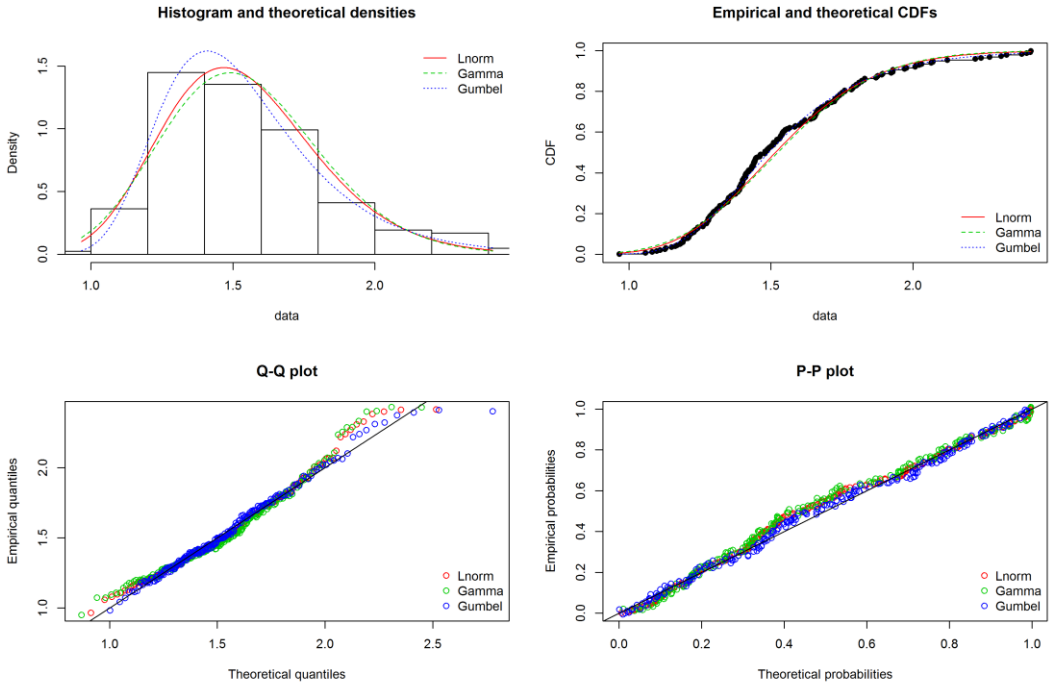


Figure 74: Marginals distributions water level Lek (SW)

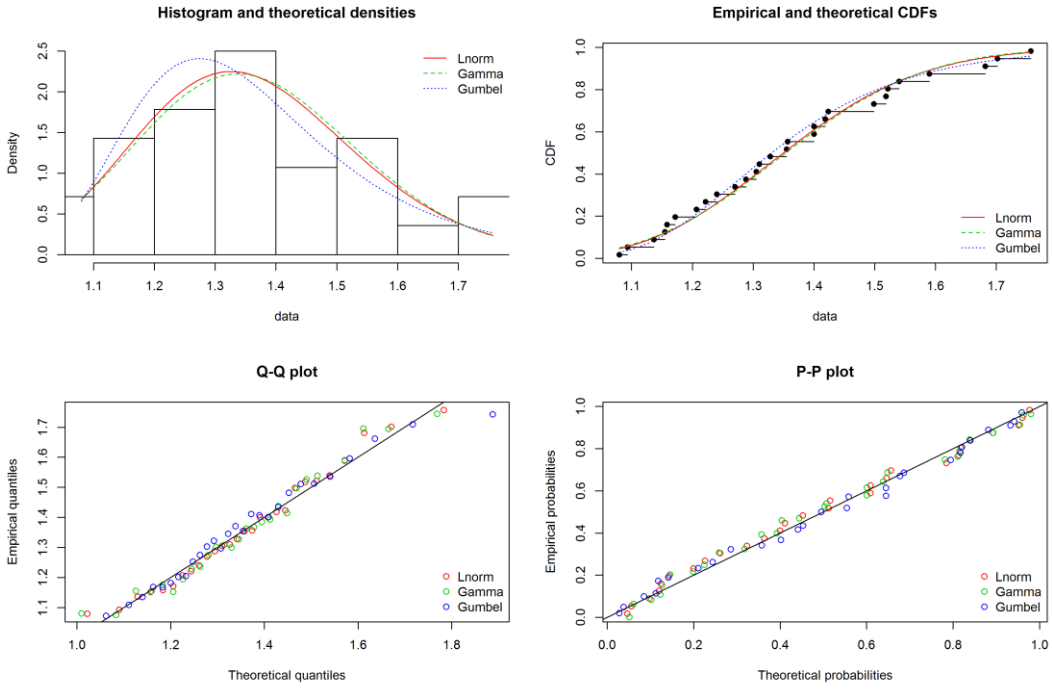


Figure 75: Marginals distributions water level Lek (O)

O. Validation by simulation combination four (vine copula)

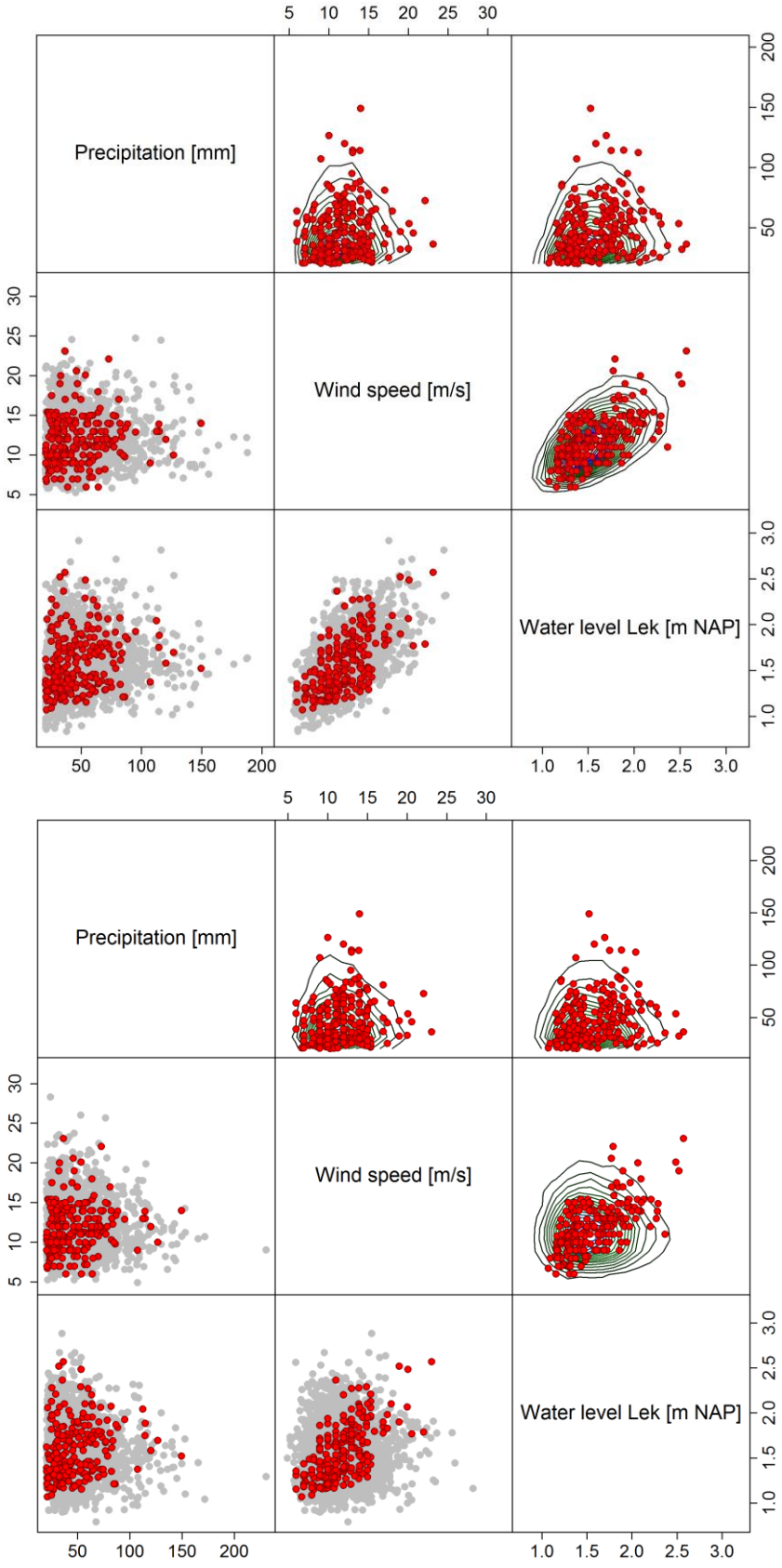


Figure 76: Comparison dependent (above) and assumed independent (below) simulation (N = 2,000) with observed values for the direction W. Lower left half of each figure is the simulation with points, upper right half are the density contours based on the simulation. Note that each half is mirrored to match the axes.

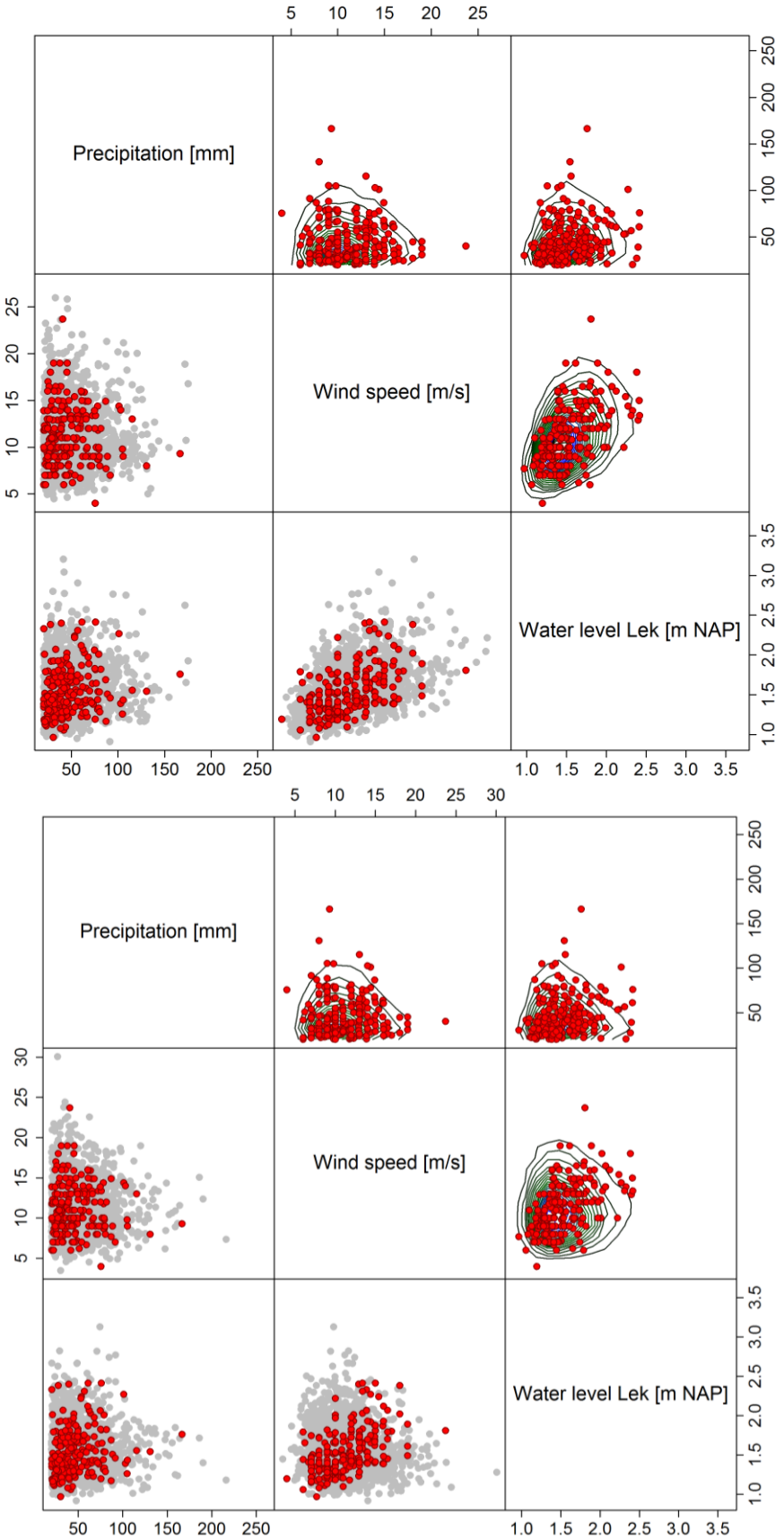


Figure 77: Comparison dependent (above) and assumed independent (below) simulation with observed values for the direction SW. Lower left half of each figure is the simulation with points, upper right half are the density contours based on the simulation. Note that each half is mirrored to match the axes.

P. Nederwaard wind & blockage effect analysis

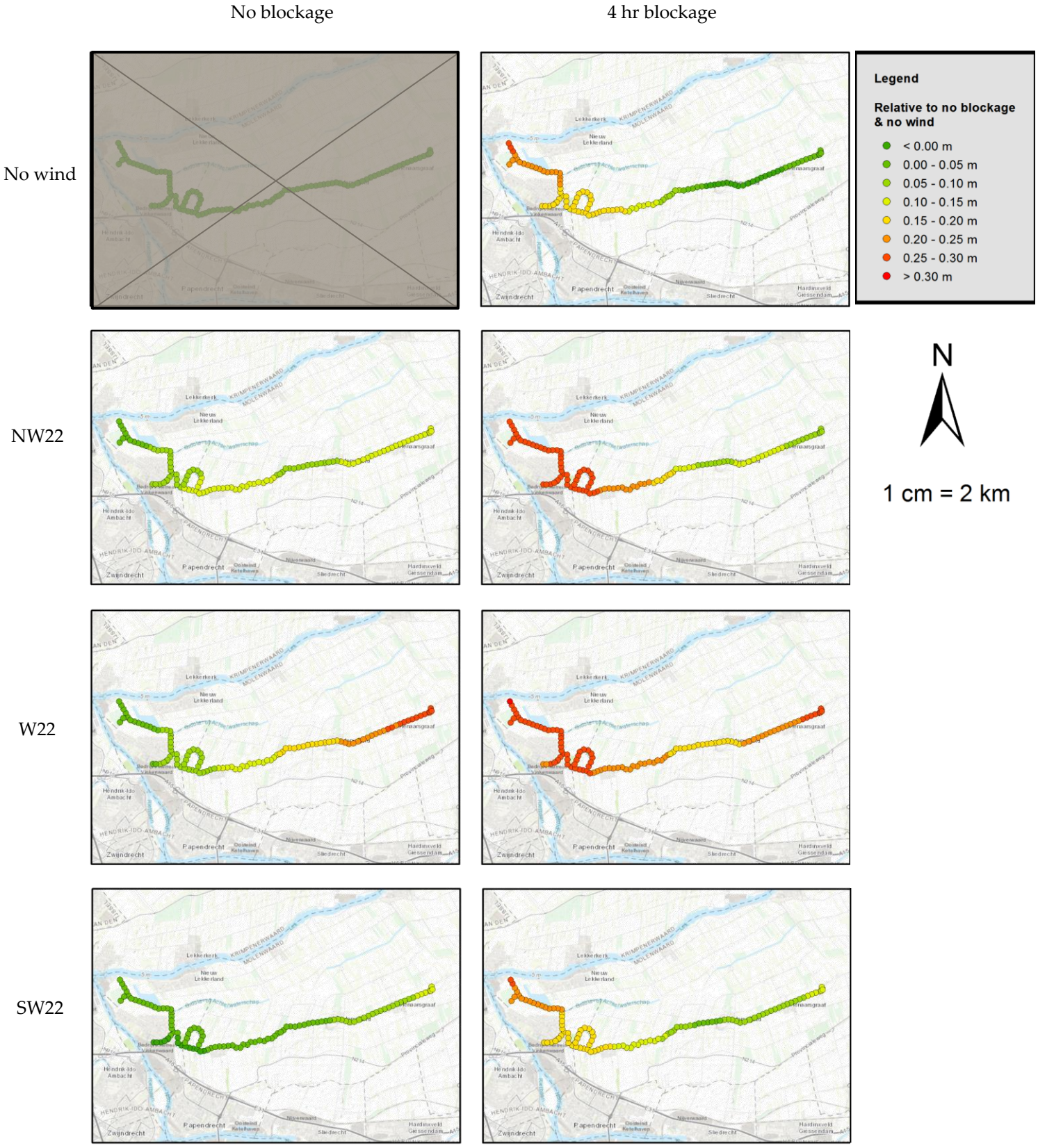


Figure 78: Effect of wind and blockage on the Nederwaard area with a precipitation event of 210 mm with one peak (shape E). Values are water depths relative to situation with no blockage and no wind (upper left frame). All frames on the left are without blockage, all frames on the right are with a blockage of 4 hours. First row is no wind, second row is wind with direction northwest and speed 22 m/s etc.

To understand and support the result of the first combination, the effect of wind and a blockage on the Nederwaard area is visualised in Figure 78, using the results of single model events. Visible is that the main effect of a 4-hour blockage without wind reaches no further than about the middle section of the Nederwaard area (location 2 in section 5.3.1). Furthermore, it is visible that the effect of wind (without blockage) is mainly in the eastern part of the Nederwaard area, with especially a strong effect for wind direction west (22 m/s) of up to about 30cm. When a wind event takes place simultaneously with a blockage, it is visible that the effect of the blockage is amplified and reaches farther from the pumping station relative to no wind. This explains the spatial distribution of the result of the first combination, discussed in section 5.3.1, as the middle section of the Nederwaard is the most affected by the combination of wind and a blockage compared to only wind or only a blockage.

Q. Nederwaard precipitation & blockage effect analysis

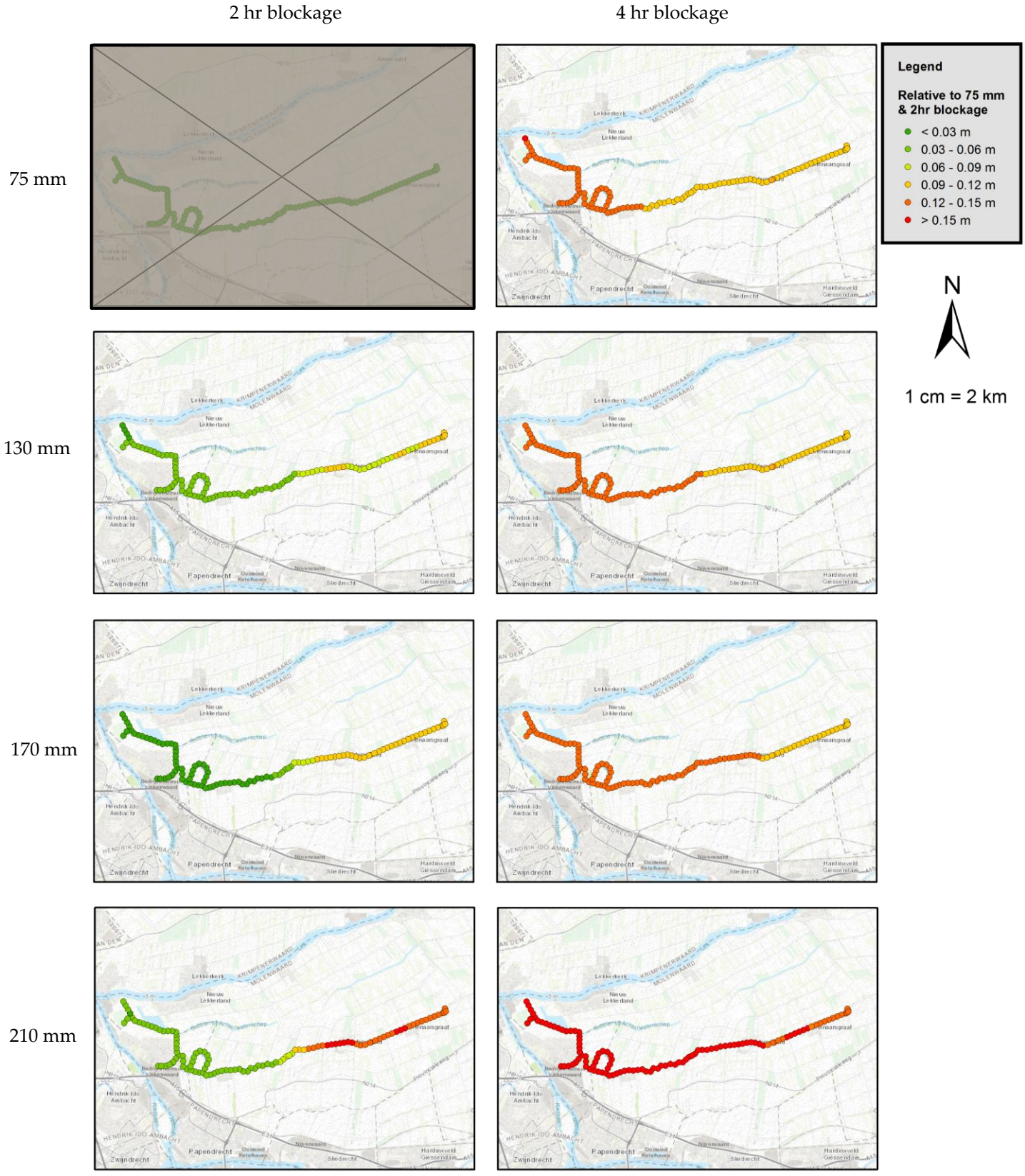


Figure 79: Effect of precipitation and blockage on the Nederwaard area with precipitation events with one peak (shape E) and no wind. Values are water depths relative to situation with 2hr blockage and 75 mm rain (upper left frame). All frames on the left are with a blockage of 2 hours, all frames on the right are with a blockage of 4 hours.

To support the result of combination 2 in section 5.3.2, the effect of the combination of precipitation and a blockage is analysed in Figure 79, using the results of single model events. Visible is how an increasing precipitation volume does not have a large impact close to the pumping station during a blockage, as these water levels are mainly determined by the duration of the blockage. However, as the precipitation volume increases, the water levels in the eastern part of the Nederwaard (farther away from the pumping station) are considerably affected compared to less precipitation. So, for higher return periods with higher precipitation volumes, the impact of a blockage shifts farther away from the pumping station, while the water levels close to the pumping station do not change a lot.

R. Alblasserwaard precipitation & wind effect analysis

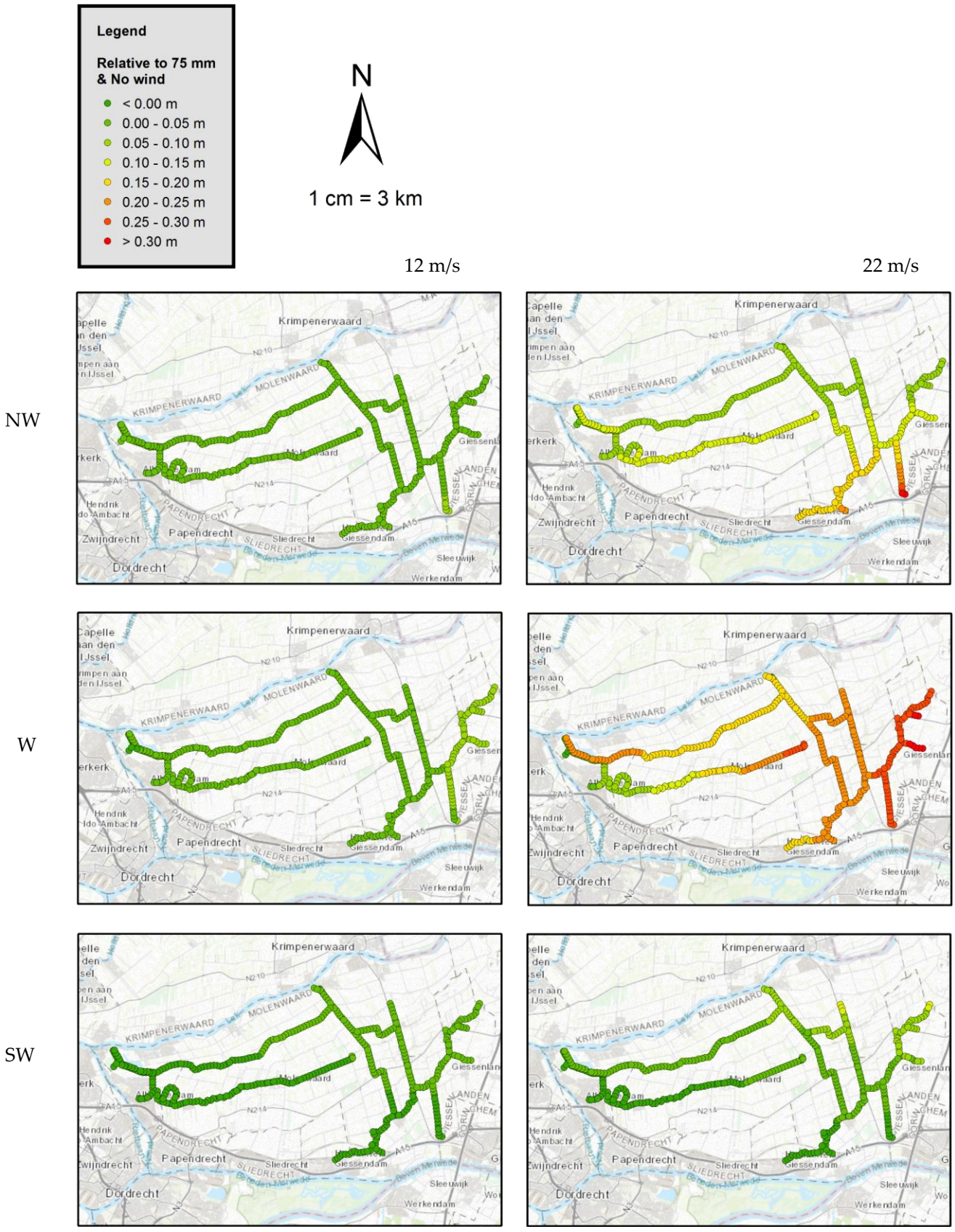


Figure 80: Effect of wind on the Alblasserwaard for the direction NW, W and SW with a precipitation event with one peak (shape E) and 75 mm. Values are water depths relative to situation with no wind. All frames on the left are with a wind speed of 12 m/s, frames on the right are with a wind speed of 22 m/s.

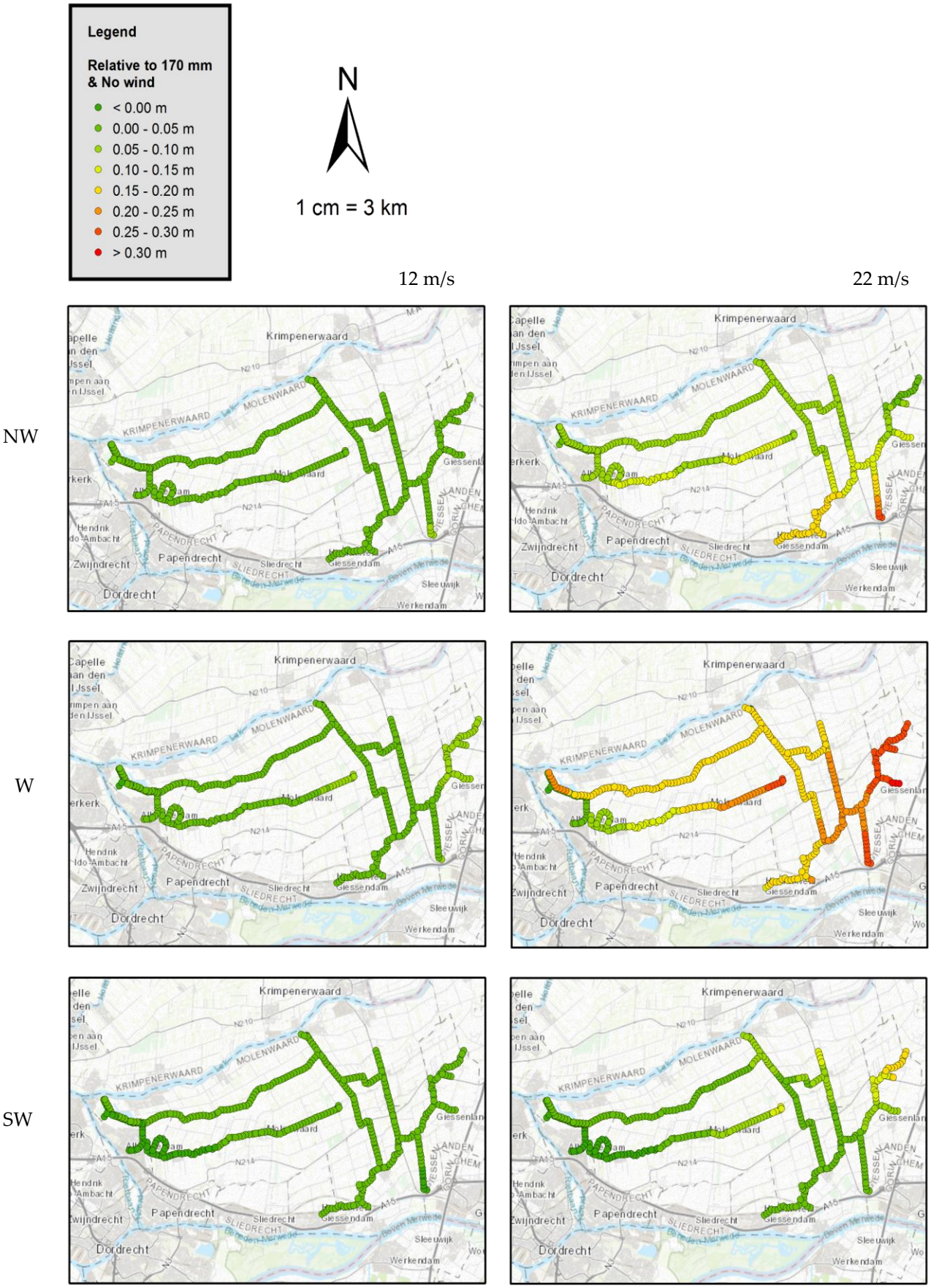


Figure 81: Effect of wind on the Alblasserwaard for the direction NW, W and SW with a precipitation event with one peak (shape E) and 170 mm. Values are water depths relative to situation with no wind. All frames on the left are with a wind speed of 12 m/s, frames on the right are with a wind speed of 22 m/s.

The sensitivity of the water levels to wind is analysed in Figure 80 and Figure 81, to support the results of the third combination in Section 5.3.3. These figures are based on single model results (no probabilities involved). Visible in both figures is that especially winds from the west and northwest affect the water levels considerably, up to 30 cm in the east part of both the Overwaard and Nederwaard area when the wind speed is 22 m/s. For wind speeds of 12 m/s, the maximum differences are about 5 - 10 cm. This analysis is performed with a precipitation event of 75 mm and an event of 175 mm. The precipitation volume does not affect the effect of wind on the water levels considerably. For the wind directions northwest and west, the effect on some locations is slightly decreased with more precipitation, while the effect is slightly increased for the wind direction southwest.

S. Example code

```
##Example code to construct the joint probability distribution using a copula
##between the variables Precipitation volume and water level of the Lek.
##Code is simplified to only show relevant functions

#Import libraries
library(readr)
library(xts)
library(zoo)
library(copula)
library(VineCopula)
library(POT)
library(fitdistrplus)


#Import data

#setting work directory
setwd("D:/work/...")

#import waterlevel data from csv
waterleveldata <-readr::read_delim("Datasets/RWSLEK.csv", ";",
                                   escape_double = FALSE, trim_ws = TRUE)

#import precipitation data from txt
precipitationdata <- readr::read_delim("Datasets/KNMI_Rotterdam.txt",
                                   ",", na = "NA", comment = "#")

#convert data to xts to easily manipulate time series
precipitation.xts <- xts(precipitationdata, format = "%Y%m%d %H")

#subset of timeseries for relevant period
precipitation19742018 <- precipitation.xts["1974/2018"]


#Extract relevant events from data

#determine the maximum water level in a 9 day period
waterleveldata9 <- zoo::rollapply(waterleveldata, width = 9, FUN = max)

#extract precipitation events over 20 mm out of data
threshold <- 20
events <- POT::clust(precipitation19742018, threshold, tim.cond = 9)

#build dataset with the relevant events
dataevent <- data.frame(precipitation19742018[events,],
                        waterleveldata9[events,])
#check correlation
cor(dataevent, method = 'kendall')
```

```
#Determine marginal distributions

#fit marginal distribution to precipitation with method mle
fit.precipitation <- fitdist(dataevent$precipitation, distr = "lnorm",
                             start = list(...), method = "mle")
#fit marginal distribution to water levels with method mle
fit.waterlevel <- fitdist(dataevent$waterlevel, distr = "gumbel",
                             start = list(...), method = "mle")
#aic score of fit
fit.waterlevel$aic

#bic score of fit
fit.waterlevel$bic

#visual inspection of fit
plot(fit.waterlevel)

#Determine copula function
#plot empirical copula for visual inspection
plot(copula::pobs(dataevent[,1]), copula::pobs(dataevent[,2]))

#Determine AIC/BIC scores for all available copula functions
BiCopEstList(u1 = pobs(dataevent[,1]), u2 = pobs(dataevent[,2]),
             rotations = TRUE)

#Determine GOF scores (CvM)(n = 10000) for potential copula
BiCopGofTest(u1 = pobs(dataevent[,1]), u2 = pobs(dataevent[,2]),
             family = BiCopName("Clayton"), method = 'kendall', B = 10000)

#Determine 5-fold cross-correlation score for potential copula
xvCopula(claytonCopula(), k = 5, x = dataevent)

#Visual application to compare the copula functions to the observations
#as extra check
BiCopCompare(u1 = pobs(dataevent[,1]), u2 = pobs(dataevent[,2]))

#Construct joint probability distribution
#joint probability distribution with copula function and margins
jpd <- mvdc(claytonCopula(param = c(1.509)), margins = c("lnorm",
"gumbel"),list(list(par1 = ..., par2 = ...),list(par1 = ..., par2 = ...)))

#validation by simulation
#simulate 3000 values from joint probability distribution
simvalues <- rMvdc(3000, jpd)

#plot observations and simulated values
plot(dataevent)
points(simvalues)

#Deriving probabilities from joint probability distribution
#Determine boundaries of interval to determine its probability
int.precipitation <- c(50,60)
int.waterlevel <- c(2.25, 2.50)

prob <- (pMvdc(c(int.precipitation[2], int.waterlevel[2]), mvdc = jpd)
         -pMvdc(c(int.precipitation[1], int.waterlevel[2]), mvdc = jpd))
         -(pMvdc(c(int.precipitation[2], int.waterlevel[1]), mvdc = jpd)
           -pMvdc(c(int.precipitation[1], int.waterlevel[1]), mvdc = jpd))
```

```
##Vine Copula example

#construct dataset
dataeventvine <- data.frame(precipitation, windspeed, waterlevel)

#define structure and test the copula (only shown for vine tree 1)
Matrix1 <- c(3, 1, 2,
             0, 2, 1,
             0, 0, 1)
Matrix1 <- matrix(Matrix1, 3, 3)

Vine1 <- VineCopula::RVineCopSelect(pobs(dataeventvine,
                        Matrix = Matrix1, familyset =
                        c(1,3,4,5,6,13,14,16,23,24,26,33,34,36), se =
                        TRUE))
summary(Vine1) #summary of constructed vine copula
contour(Vine1) #contour plot of constructed vine copula

Vine1$AIC
Vine1$BIC
Vine1$GOF <- RVineGofTest(pobs(dataeventvine, Vine1, method = "White",
statistic = "CvM", B = 10000))
Vine1$se #standard error

##Simulation from vine copula
Sim <- data.frame(VineCopula::RVineSim(1000, Vine1))

#transform values between 0 and 1 to their real values using the marginal
distributions
#only one shown as example
sapply(Sim[,1], function(x){qlnorm(x, 0.41,0.19)})
```

T. Safety level regional flood defences Alblasserwaard

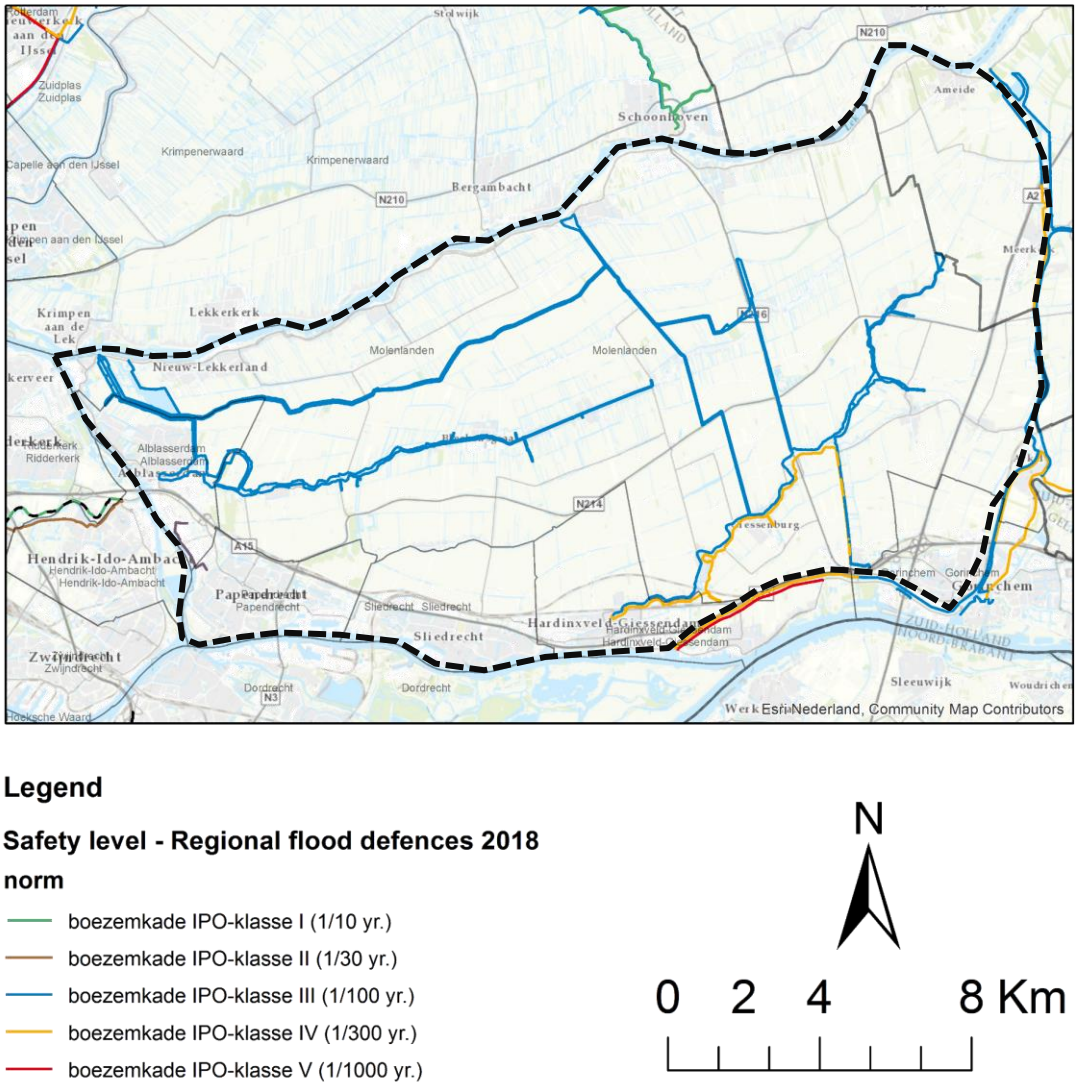


Figure 82:Safety norms of the regional flood defences in the Alblasserwaard (dark blue dashed outline) (Provincie Zuid-Holland, 2019)