DELFT UNIVERSITY OF TECHNOLOGY

Introducing a Load Trend to the Reliability Analysis of Hydraulic Structures

Application of Bayesian Network-supported Reliability Analysis to predict future failure of Pumping Station IJmuiden

TUDelft

This page was intentionally left blanc

Introducing a Load Trend to the Reliability Analysis of Hydraulic Structures

Application of Bayesian Network-supported Reliability Analysis to predict future failure of Pumping Station IJmuiden

By

Robin Huijmans

28-09-2018

in partial fulfilment of the requirements for the degree of

Master of Science

in Construction Management and Engineering

at the Delft University of Technology

Faculty of Civil Engineering and Geo-sciences

Chairman:	Prof. dr. ir. A.R.M. Wolfert	TU Delft, Infrastructure Design and Management
1st Supervisor:	Dr. Ir. O. Morales-Napoles	TU Delft, Hydraulic Structures and Flood Risk
2nd Supervisor:	Dr. R. Schoenmaker	TU Delft, Infrastructure Design and Management
Company Supervisor:	Ir. A. Willems	Iv-Infra

This page was intentionally left blanc

FOREWORD

This report is the result of my graduation research to complete the Master Construction Management and Engineering at the Delft University of Technology. It introduces a non-parametric Bayesian Network supported reliability analysis in order to determine the failure probability of coastal pump stations. Here I would like to emphasize that concepts of the framework can be applied to hydraulic structures in general.

The idea that initiated this research started approximately two-and-a-half years ago when the Merwede bridge closed for heavy vehicles. I wrote the idea on a small piece of paper which I took out of my closet approximately a year ago from now. From then on, I began to focus on predictive maintenance studies such as the conducted by Jan van Noortwijk. Although I understood the core, the underlying mathematics made me anxious.

Looking back on what I learn throughout the process makes me pretty proud. Moreover, I seriously got interested into mathematical concepts such as Bayesian Networks, Gamma processes and Markov chains. For example, I never would have thought that I would read an interview with Judea Pearl – the initiator of Bayesian Networks – out of interest. I reckon that a thesis subject as this is not very popular among CME-students since the predominant part sees the mathematical concepts as a huge obstacle. I think that's really a shame since it is – in my opinion – a very interesting side of infrastructure asset management. Therefore, a sub-aim of this research is to make those concepts more accessible to students with a less mathematical background.

I would like to make use of this foreword to thank the members of the graduation committee, their input, knowlegde and time which have been of great value to conduct this master thesis. Second I want to thank my colleagues at Iv Infra, especially Arno Willems, for his support during this graduation process. Due to you all I was always able to work with a laugh and work in complete silence when needed. Third, I want to thank my family and friends for stimulating me to pursue the track that has led me to this point in my life. Due you all I have gone through this process with enthusiasm and produced a report that I am proud of.

I hope that you, the reader, enjoy reading the reported and do not hesitate to contact me for any discussion or questions about the thesis.

Robin Huijmans

Delft, September 2018

This page was intentionally left blanc

EXECUTIVE SUMMARY

Research Motive

This research describes a way to determine the reliability of coastal pump stations - from which their environment is influences by climate change – in time. Last decade, the challenge within the hydraulic infrastructure network shifted the focus from expanding the infrastructure inventory towards the maintenance and management of the existing network. The reason for this tendency is that the predominant part of the Dutch infrastructure inventory is approaching its end-of-life while their functions are still required. As a consequence, the asset owner is confronted with the challenge to choose the suitable maintenance action in order to extent the assets life-cycle.

To weight the possible actions, an accurate estimation of asset failure is required. For this, many predictive failure models have been developed both deterministic as well as probabilistic, to assess the assets reliability in time. The predictive failure models are based on the fact that the structures performance deteriorates over time. The rate of deterioration is dependent on its ambient environment which is – for hydraulic structures - characterized by short- medium- and long term variabilities. Due to the environmental dependency, deterioration is fortified by long-term variabilities. In other words, mainly the long-term variabilities contribute to the complexity to estimate future asset failure. There are two underlying reasons. First, long-term variabilities often contain large uncertainties and secondly, historic data sets – on which those models are based - do not capture long-term variabilities. In this research, long-term variabilities are referred as long term trends.

Furthermore, this research states that the long term trend that impacts the hydraulic infrastructure network is climate change with its accompanied effects such as sea level rise, increased rainfall etc. To narrow the scope of this research the focus will be on coastal pump stations. Climate change affects coastal pump stations in three ways:

1. Sea level rise fortifies the deterioration rate and increases the probability that an individual pump is unavailable.
2. Sea level rise reduces the maximum pump capacity.
3. Increase of extreme rainfall events increase the number of peak-discharges per year

The practical aim is to support coastal owners with predictive failure models that incorporates the three effects of climate change in order to accurately predict future asset failure. Based on this aim, the central question is drawn:

> *'How can long-term trends be introduced to reliability analysis in order to determine the effect on the pump stations reliability on the longer term?'*

Reliability analysis for Pumping Stations

In order to answer the question, this research departed from the limit state function which can be marked as the first most important concept. Limit state functions define a condition beyond which an object or structure does no long fulfil one of its performance requirements. The limit state function for pump stations is formulated as follows:

$$Z(t) = \left(\sum_{n=1}^{N} A_{pump,n}(t) * C_{pump}(t) \right) - Q_{capacity;required}(t) \qquad (1)$$

Where $A_{pump,n}(t)$ represents the availability of the pump at a random point in time, which is a discrete value with 1 representing pump-availability and 0 pump-unavailability. $C_{pump}(t)$ is the discharge capacity of one pump (m³/s) at a random moment in time. N is the total number of pumps. The pump station is considered 'failed' when $Z(t) < 0$. Since the three variables consider uncertainties and can all be described by their (non-)parametric distributions, we are interested in the probability of failure: $P(Z(t) < 0)$. The three effects of climate change affect all three variables. Therefore, the predominant part of this research focusses on determining the distributions over time, including the three effect of climate change.

To determine those distributions in time, literature research showed that dependency modeling can be used to incorporate the long term trend, i.e. making the three limit state variables dependent on a network of affecting dependent environmental variables. Non-parametric Bayesian Networks (NPBNs) appeared to most suitable for the problem. NPBN are directed acyclic graphs (DAGs) where the nodes are represented (conditional) marginal distributions and the arcs their (conditional) copulae.

Where the environmental dependencies of $C_{pump}(t)$ and $Q_{capacity;required}(t)$ can be modelled directly via the NPBN, the effect of deterioration to $A_{pump,n}(t)$ showed to be more challenging. Therefore a framework is presented that utilized NPBNs to update the Markov chain transition matrix for each time step. Briefly, the underlying idea is to model the dependencies between the environment and the input parameters of the Markov chain transition matrix; the transition times. The outcome of this framework is the probability that a specific component is in a specific state such as 'unavailable'. Then, via the system decomposition of the pump one is able to determine the total pumps probability of unavailability at time t.

The long term trend can then be incorporated by performing inference in the NPBNs. Generally, inference updates a belief into an unobservable variable when a dependent variable is observed. In other words, the model is able to stress a belief into the discharges, when a specific future rainfall-amount is 'observed'. Inference in NPBNs only has one limitation: the future 'observations' must be captured by the data set, i.e. it must have happened before. In this research is assumed that extreme rainfall has happened before, only the probability increases as an effect of climate change. Therefore inference is suitable to model the long term trend of increase of extreme rainfall events. Unfortunately, in case of sea level rise past data does not include future values. Then the future distributions must be estimated via – for example – expert judgement. Finally, when all distributions are known for every time step t. Probability of failure can be calculated via performing simulations.

Application to the Case Study

The reliability analysis for pump stations including the long term trends as described above is applied to pumping station IJmuiden in order to validate whether the model does what was expected. Since the goal is to proof the concept of the model, simplifications are made regarding the definition of water system failure and physics of component failure.

Conclusion & Recommendations

Despite the fact that the case application shows simplifications, one can conclude that the model does what was expected. The results of the case study show a gradual increase of its failure probability, dependent on the rate of the long-term trend. Thereby, the results show that the contribution of fortified deterioration to the total probability of pump station failure is very limited under expected long-term trend circumstances. In general can be stated that the impact of climate change to the pump's reliability is very limited.

A general recommendation is given to gather environmental data in more structured- and standardized way and to gather deterioration data in accordance to the input parameters of stochastic processes such as the Markov Chain and Gamma Processes. Thereby, some more model accuracy-increasing recommendations are given such as the incorporation of Gumbel and Clayton copulae in UniNet. Also, the utilization of Dynamic Bayesian Networks (DBNs) would be an added value in user-friendly and decision-making sense.

This page was intentionally left blanc

TABLE OF CONTENT

This page was intentionally left blanc

LIST OF FIGURES AND TABLES

LIST OF ABBREVIATIONS AND NOTATIONS

ARC	Amsterdam-Rhine Channel
BN	Bayesian Network
CBM	Condition Based Maintenance
cDBN	Covariate Dynamic Bayesian Network
cdf	Cumulative Density Function
DAG	Directed Acyclic Graph
DER	Determinant of the Empirical Rank correlation matrix
DNR	Determinant of the Empirical Normal Rank correlation matrix
DBN	Determinant of the Bayesian Network rank correlation matrix
HBN	Hybrid Bayesian Network
MTTF	Mean Time to Failure
MTTR	Mean Time to Repair
NPBN	Non-Parametric Bayesian Network
NS	North Sea
NSC	North Sea Channel
pdf	Probability density function
TPM	Transition Probability Matrix

$S(t)$	Solicitation in time
$R(t)$	Resistance in time
ρ	Product moment correlation
r	Rank correlation
α	Shape-parameter, input Weibull distribution
β	Scale-parameter, input Weibull distribution
σ	Standard deviation of the variable's natural logarithm, input lognormal distribution
μ	Mean of the variable's natural logarithm, input lognormal distribution

This page was intentionally left blanc

Part I – Research Design

Merwede bridge closing

"Rijkswaterstaat stated that when the traffic flow since construction in 1961 has not increased, the bridge would easily be safe for years. Thereby the increased weight of passenger cars and trucks also contributed to the shortened technical lifetime" (ANP, 2016)

1. PROBLEM INTRODUCTION

The news article on the left can be considered as the incentives of this research. The Merwede bridge was closed for several days which caused economic damages. Although no actual collapse happened it questioned the effect of increasing loads on a structures reliability over time. Especially for hydraulic structures, the consequences of potential failure are enormous. This chapter will elaborate the urge of this research, departing from the tendency within the field of hydraulic infrastructures. Thereafter this chapter will work towards the central problem stated in this research.

1.1. DUTCH HYDRAULIC INFRASTRUCTURE INVENTORY

In the Netherlands, Rijkswaterstaat is - among others - responsible for the management and maintenance of 650 hydraulic structures in the Netherlands. Generally, hydraulic structures are part of the flood defense system (e.g. barriers), the water management system (e.g. weirs) and the navigation system (e.g. navigation locks). Many of these systems have multiple functions; a navigation lock could also be part of the flood defense system (Jonkman, Voortman, Klerk, & van Vuren, 2018). A large part of the Dutch hydraulic infrastructure has been constructed after the Second World War, see figure 1. Until the 21st century new construction governed the development of infrastructure networks (Klatter & Roebers, 2017). The last decade shows that networks have matured and further expansion of the inventory is declining.
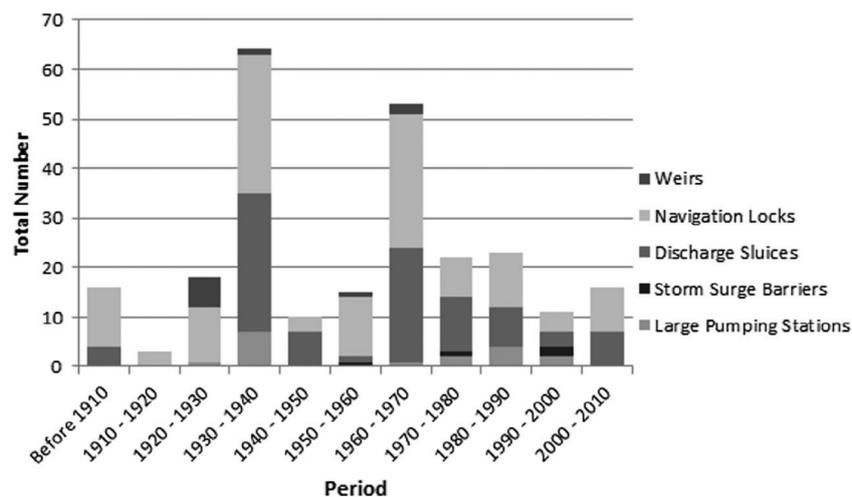
Figure 1: Construction year of Dutch hydraulic structures (Jonkman et al., 2018).

Due to ageing and more intensive use of the existing structures, many approach their end-of-life. Jonkman et al. (2018) notes that two challenges are of the main concern nowadays that arise from this tendency. First, despite the fact that the structures age, their functions might still be needed. This raises the question whether to renew, adapt or upgrade the considered structure.

The second challenge concerns the management and maintenance of existing structures. Past events have raised attention to management and maintenance aspects. In 2013, it appeared that the scour-hole protection near the Easter Scheldt barrier was eroding, whereas this was not noticed by the local management authority. Thereby, the reliability of the Maeslant barrier near Rotterdam appeared somewhat lower than expected, which resulted in the consideration of upgrades and reinforcements of the flood defenses behind the Maeslant barrier. Both events show the complexity of maintenance and management of structures that approach their end-of-life.

Overall, the combination of ageing existing structures and decline of further network expansion shifts the mindset from funding new infrastructure to the funding of renewals, adaptations, upgrades and proper management of the existing structures in order to fulfill its technical- and functional requirements. This is where infrastructure asset management comes in.

1.2. INFRASTRUCTURE ASSET MANAGEMENT

Infrastructure asset management aims at maximizing the performance of infrastructure asset systems in terms of functionality, for the energy, water and transport sector. In general, infrastructure asset management is concerned with applying technical and financial judgement to decide what infrastructure assets need to meet the performance aims over whole life, through to disposal (Hastings, 2015).

The main contributor that enhances the complexity of the decision-making process is the problem that asset performances can decline over time. This process is called deterioration. *Deterioration* is an ongoing process where the value and performance of assets reduces over time due to stressful conditions. This negatively influences length of the assets lifecycle. To counteract early lifecycle ending due to asset deterioration, van Dongen (2011) described that the asset's life cycle can be extended by performing maintenance. The type of maintenance actions – component replacements, upgrades or asset modifications etc. - are dependent on their type of ageing. One can distinguish four ageing-processes that impact the assets lifecycle: technical ageing, functional ageing, economical ageing and compliance ageing. Within infrastructure asset management, the first two ageing processes are most common. *Technical ageing* means that there is a gradual degradation of often structural parts of an asset. *Functional ageing* comprises that the asset does not meet the primary function, the services to be provided, or the products produced (van Dongen, 2011). In case of technical ageing, the lifecycle can be extended by performing critical component replacements. For functional ageing, the lifecycle can be extended by asset modifications or upgrades.

Generally, maintenance actions can be performed *correctively* or *preventively*. In this business maintenance is often preventive, since the consequences of sudden component failure are generally high (Jorissen & van Noortwijk, 1998). In the past decades, lots of preventive models have been proposed (Wang, 2002). Still they have all one thing in common; in order to make decisions whether to renew, recondition, modify or upgrade the object, the asset manager needs to have knowledge about the future moments of failure. For this reason *predictive failure models* have been developed, both deterministically and probabilistically (Jardine & Tsang, 2013). Deterministic models are those in which the timing of replacements are assumed to be known with certainty. In contrast, probabilistic models are those on which in which the timing of the failures depend on chance. Since civil assets are almost always subject to dynamic environmental conditions, failures cannot be timed with certainty. For this reason, probabilistic models are mostly used within the civil engineering domain. The adaptation of probabilistic methods forces the asset manager to think in terms of certain failure, towards thinking in terms of *probability of failure*. An important input to predictive failure models is an adequate understanding of all environmental variabilities. This typically includes the estimation of environmental conditions that can be expected over the life of an asset or asset system.

1.3. ASSET ENVIRONMENT & LOAD TRENDS

Infrastructure asset networks operate in a dynamic environment where they are exposed to short-, medium- and long-term variabilities in ambient environmental conditions (Rayner, 2010). Long-term variabilities embrace social-, economic- and climatic variabilities that affect the assets ambient environmental conditions. In most cases, the magnitude of impact on the environmental variables

posed by the long-term variables develop gradually over time. From now on, we refer to those long term variabilities as *long term trends.*

Male (2010) states that organizations - whether public, private or hybrids - that heavily rely on physical assets to function will face unprecedented challenges over the coming decades. Many of these challenges will be due to the difficulties posed by the climatic long term trends. As mentioned previously, an asset manager requires an adequate understanding of all environmental variabilities, including the climatic variabilities affected by climate change. Rayner (2010) addresses the problem that the effects of climate change often contain large uncertainties, see for example IPCC (2014). In other words, it is very hard to retrieve accurate estimates regarding the magnitude of impact on the environmental variables. Thereby, the use of historic data brings up the assumption that the use of a long enough time history of past variability will capture the future variability as well. Briefly stated, with the current models we assume that past data will be representative for the future. This assumption that is no longer valid when there is a cycle longer than the length of the measured time history (Rayner, 2010).

1.4. SYNTHESIS

Recapping, a large number of hydraulic assets will approach their end-of-life. Replacements or renovations are necessary since the function of those assets are still required. Rijkswaterstaat expects that the costs of those actions will raise up to 100 million euros per year (Rijkswaterstaat, 2014). To get an overview of the necessary actions, it is important to have the best possible estimate of the assets point of failure. For this, predictive maintenance models are available, of which the probabilistic models are preferred. Logically, the moment of asset failure is dependent on the deterioration rate which is related to the asset ambient environment.

Paragraph 1.3. described that the asset environment is actually changing as a consequence of climate change. Due to the increasing intensity of the asset environment by long term trends like increasing use, higher loads, heavier rainfall, stronger winds, deterioration is fortified (Klatter & Roebers, 2017). This means that assets deteriorate faster and will approach their end-of-life earlier than expected during construction. In probabilistic terms; the probability of asset failure increases over time and will approach its unacceptable threshold earlier than expected during construction. Researchers, private- and public- entities, support this reasoning and commonly agree that the effects like social growth, economic growth and climate change impact the lifecycle of the hydraulic infrastructure inventory (Ministery of Infrastructure and Environment, 2016; U.S. Department of Transportation, 2015). The general research problem and urge of this research can be summarized as follows:

> *Large maintenance assignments are upcomming. Hydraulic asset owners need accurate predictions of their assets lifecycle-end. Since climate change fortifies asset deterioration - and consequentially increases the failure probability over time -, long term trends as climate change must be incorporated into the predictive failure models.*

Among the hydraulic structures mentioned in figure 1, incorporating climate change into models to predict coastal pump station failure are expected to be most complex. The way how the long term trends posed by climate change affect the function of a pump station is described in the next paragraph. To limit the scope of this study we will continue this research with the focus on coastal pump stations. Still we want to emphasize that concepts can be applied to various hydraulic structures.

1.5. EFFECT CLIMATE CHANGE ON PUMP STATIONS

In this research is assumed that climate change impacts coastal pump stations via sea level rise and increase in rainfall occurrences. Here is assumed that those events affect the pump station function in three ways. For readability reasons, we already introduce those effects here.

First, sea level rise affects the operating time of the pumps. The fact that the pumps must operate more often fortifies the deterioration over time. The timeslots that the pumps operate, is dependent on the water levels on both sides of the pump station. Unfortunately, the water level on the sea side rises as a result climate change. This would mean that the time slots that pumping is required will increase over time, which enhances the wear of the pumps, see figure 2. Due to the enhanced wear, the probability that the pump is unavailable to perform its function increases.
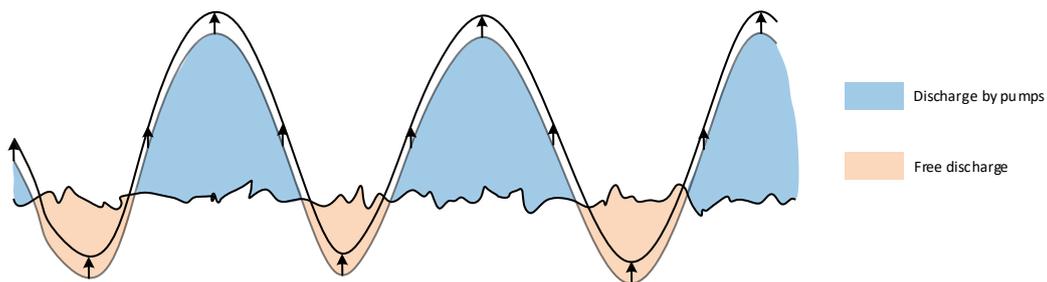
Figure 2: Effect of sea level rise on the pump- and free discharge time slots

Second, sea level rise decreases the maximum capacity of the pumps. This can best be illustrated by use of a Q-H curve, see figure 3. The Q-H curve illustrates the relation between the decreasing pump volume when the water head increases. Sea level rise causes an overall increase of the water head, which implies that the pumps have to function more often with reduced pump capacity. Since this affects all pump in the pump station simultaneously, the total required pump volume will drop dramatically.
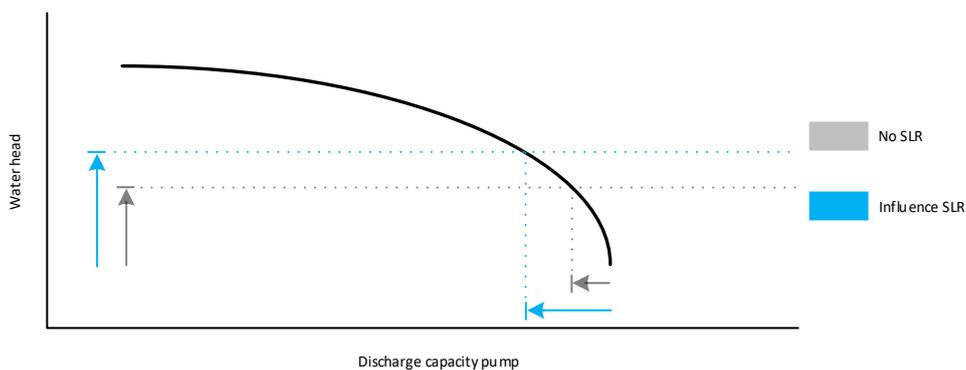
Figure 3: Example Q-H curve

Third, increase in extreme weather occurrences increases the amount and quantity of pumped discharges. Normally, the excess water in the system, of which rainfall is a large contributor, will be drained and discharged towards the main channel and thereafter discharged by the pump station. Since the amount of rainfall increases with time, the peak-dischargeable amount water will also increase, as does the number of peak-occurrences. Consequentially, the occasion that maximum pump capacity is required - the occasion that all pumps must be available - will occur more often.

2. RESEARCH DESIGN

This research aims to provide knowledge, insight and information that 'fills' the research gap discussed in paragraph 2.1. Thereafter, the research objective and research question will be briefly introduced.

2.1. RESEARCH GAP

The function of coastal pump stations is to discharge the in-land water systems excess water into the sea. Failure occurs when the total discharge capacity delivered by the pump station is smaller than the required discharge capacity. As introduced in paragraph 1.5., climate change affects both total discharge capacity of the pumping station as the required discharges. As a result, the probability that the asset cannot perform its function increases gradually.

Limited number of studies assessed the reliability of pumping stations. Those studies mainly analyzed the pumping station reliability via – for example - failure rate functions (Briggs & Hodkiewicz, 2005). In chapter 5 of this research we mention the limits of this way of assessing reliability and state that this model is insufficient to introduce long term trends. Concluding, a probabilistic way that is able to incorporate the impact of long term trends - especially climate change - on the reliability of the pumping station is not performed yet.

2.2. RESEARCH OBJECTIVE

In essence, the objective of this research is to fill the research gap, namely to conduct a reliability analysis that is able to incorporate the impact of long term trends - especially climate change. The focus of this research is therefore on introducing long term trends to reliability analysis. Based on this line of reasoning, the research objective can be formulated as:

> *'… to determine the coastal pump station reliability over time - including the effects of climate change - by introducing long term tends to reliability analysis theory'*

2.3. RESEARCH QUESTION

The research objective marks the scope of this research; the research focusses solely on introducing a long term trend to the reliability analysis of coastal pump stations that are affected by the effects of climate change. The research question is formulated as follows:

> *'How can long-term trends – such as climate change – be introduced to reliability analysis in order to determine the effect on the pump stations reliability on the longer term?'*

Based on the research question, the aimed result would be a reliability analysis that is able to introduce long term trends for coastal pump station. Ideally, a general step-by-step approach is preferred in order to be able to perform the reliability analysis to real-life cases.

3. RESEARCH METHODOLOGY

The main aim of this research is to close the research gap by introducing long term trends to reliability analysis of coastal pump stations. In order to close the gap, the research question gives a handle regarding the preferred result. This chapter generally clarifies the path to be taken in order to arrive at an answer to the formulated research question. The underlying figure shows the general.
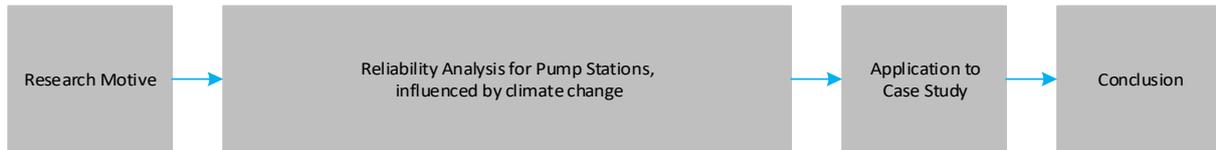
Figure 4: Overview research steps

This research is divided into four phases, corresponding to the document parts mentioned in the table of contents. At first the necessity to conduct this research will be shown under 'research motive'. The problem introduction mentioned the urge. The aim of the second phase is to find a theoretical framework to incorporate long term trends into the reliability analysis. To verify the proposed framework, the third part will reflect the proposed approach to an actual case study; Pump Station IJmuiden. Based on the findings of the case study, part four will reflect on the research objective and research question.

Research Motive

The research motive is already discussed extensively in chapter 1 – Problem introduction. Since large maintenance assignments are upcoming, the field of infrastructure asset management needs predictive failure models that are able to incorporate long term trends.

Reliability analysis for pump stations

Based on literature is aimed to find a handle to depart from in order to conduct a reliability analysis over time that is able to incorporate the three effects of climate change treated in paragraph 1.5. Literature showed that the use of a limit state function serves as the most suitable point of departure. See Appendix I. Conceptually, the limit state function can be divided into a resistance and a solicitation, both described by their (non-)parametric distributions. The main challenge is to determine the distributions for each time step, since both the resistance- as the solicitation- distributions are expected to change shape for every time step as a consequence of the three effects of climate change. Those distributions can be estimated directly, for example via expert judgement. However, in dynamic complex environments, network-based approaches are preferred. In other words, the problem asks for an approach to model the dependencies between the asset and its environment.

Application to Case Study

The framework sketched in the phase two will be applied to a case study in order to verify whether the model does what we expect it to do. The primary requirements to the choice of the case study is that its environment is subject to the effects of climate change. Secondary requirements predominantly relate to the availability of sufficient data regarding the local discharges and properties of the pump station itself. Pump Station IJmuiden appeared to be most fitting.

To show the impact of climate change and fortified deterioration, we basically acknowledge two models; *environment-independent deterioration model* and *environment-dependent deterioration*

model and expressing them in terms of resistance and solicitation. Since the load trends are inevitable, the environment-independent deterioration model shows the case that we do not consider fortified pump deterioration, but climate change only affects the increase of the solicitation, the required discharges, see the left three figures. The environment-dependent deterioration model considers both fortified deterioration as increasing discharges due to climate change, see the figures on the right. Comparing the latter two figures, the probability of pump station failure is expected to be higher in the environment-dependent deterioration model.
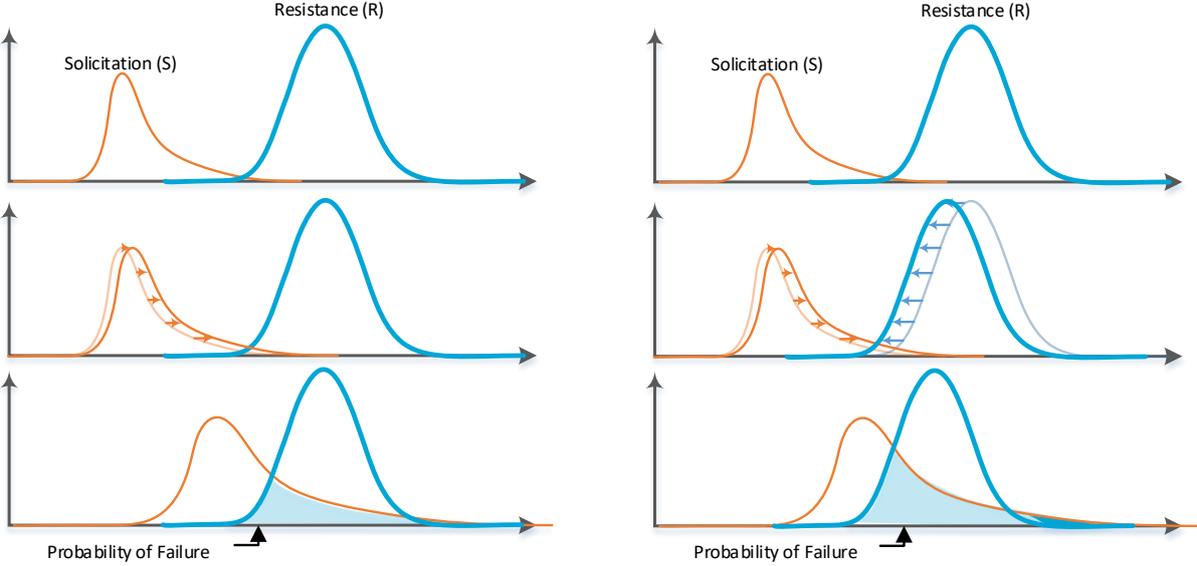
Figure 5: Schematic representation of development probability of failure.

Conclusion

The results of the case study will show whether the proposed pathway is able to determine the pump station reliability over time. Via the two models proposed before, the results show the impact of climate change to the deterioration rate. Thereafter, the framework will be evaluated, discussed and opportunities for further research will be summed up.

This page was intentionally left blanc

Part II – Reliability Analysis for Pump Stations

This page was intentionally left blanc

4. LIMIT STATE FUNCTION FOR PUMP STATIONS

This chapter considers a mixture of literature research and interpretations regarding the addressed objective. Essentially, several tools are combined in order to perform a reliability analysis for every time step. The chapter departs from defining a general limit state function, applicable for pump stations. The concepts of a reliability analysis, based on the limit state function, is elaborated in Appendix I.

Limit state functions define a condition beyond which an object or structure does no long fulfil one of its performance requirements. In order to make a limit state function suitable for the problem mentioned, the limit state function must be written in terms of discharge capacities. As mentioned, the pump station fails when the total discharge capacity is lower than the required discharge capacity. Therefore the limit state function can be written as follows:

$$Z(t) = Q_{capacity;total}(t) - Q_{capacity;required}(t) \tag{2}$$

Where $Q_{capacity;total}$ is the total discharge capacity and $Q_{capacity;required}$ is the occurring capacity that must be pumped. Both variables can be described by their distributions. In fact, the total capacity ($Q_{capacity;total}(t)$) at random moment t is related to the question whether the pump is available and what the discharge capacity is. Therefore, equation (2) can be written as follows.

$$Z(t) = \left(\sum_{n=1}^{N} A_{pump,n}(t) * C_{pump}(t) \right) - Q_{capacity;required}(t) \tag{3}$$

Where $A_{pump,n}(t)$ represents the availability of the pump at a random point in time, which is a discrete value with 1 representing pump-availability and 0 pump-unavailability. $C_{pump}(t)$ is the discharge capacity of one pump (m³/s) at a random moment in time. N is the total number of pumps. Deterioration affects the probability of failure over time, which consequentially enlarges the probability that $A_{pump,n}(t)$ returns a 0. On the other hand, sea level rise increases the probability that the individual pumps cannot pump with full pump capacity $C_{pump}(t)$. In fact, this research is all about defensibly determining the distributions mentioned in equation (3). Probably the hardest task is to model the dependency between environment and individual pump deterioration. For this reason, we depart from the basic concepts of probabilistic deterioration modelling.
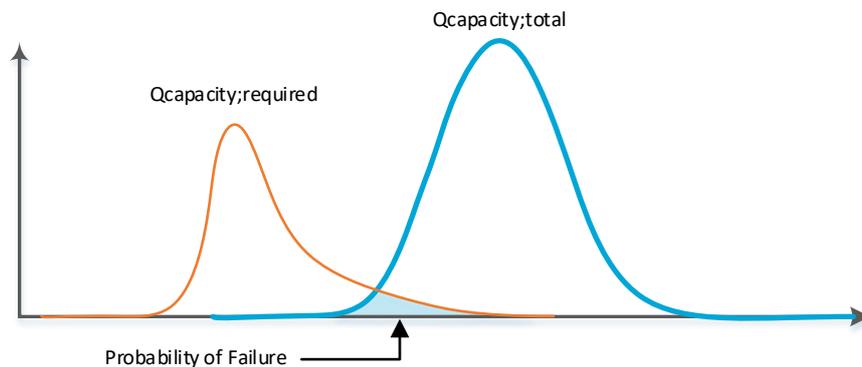
Figure 6: Definition probability of failure pump stations

5. PROBABILISTIC DETERIORATION MODELLING

By means of this paragraph we aim to find a handle to depart from. This paragraph departs from a brief explanation of stochastic-based-processes which describe the deterioration of an object, component or asset over time. Decennia after the introduction of those methods derivatives started to appear that are more suitable for the complex challenges nowadays. In this research we refer to those methods as network-based-deterioration models.

5.1. STOCHASTIC-PROCESSES-BASED DETERIORATION MODELS

Currently used probabilistic methods, such as failure rate models, are based on time-to-failure distributions which are the probabilities of failure at random time t. In failure rate models, the time-to-failure distributions are described by parametric families of distributions such as the exponential, the Gamma, the lognormal, the Pareto, the Weibull and their mixtures and multivariate analogues (Singpurwalla, 1995). Ideally, the parameters of those distributions are estimated via available data (Jardine & Tsang, 2013).

However, Singpurwalla (1995) argued that a serious disadvantage of failure rates is that they cannot be observed or measured for a particular component. Thereby, the sparse amount of failure data results in unreliable estimations of the distribution parameters. In other words, in absence of sufficient failure data, a reliability analysis solely based on time-to-failure distributions and their unobservable failure rates is unsatisfactory (van Noortwijk, 2009).

In addition, Singpurwalla (1995) suggested that a more appealing approach would be to choose a model based on the physics of failure and the characteristics of the operating environment. This brings us to *stochastic-process-based deterioration models*. Several authors have spent most of their work on those models (Cinlar, 1972a, 1972b; Frangopol, Kallen, & Noortwijk, 2004; Grall, Dieulle, Bérenguer, & Roussignol, 2002; Hastings, 2015; Jardine & Tsang, 2013; Nicolai, 2008; van Noortwijk, 2009). Nowadays, these stochastic-process-based models are supported by Rijkswaterstaat and proved their added value in the design- and the operation phase of several projects (van Noortwijk & Klatter, 1999; van Noortwijk & Peerbolte, 2000)

Generally, stochastic processes, also called random processes, model the evolution of a random system in time. The goal of stochastic-process-based deterioration models is to derive the probability of object failure in time $F(t)$ and their accompanied time-to-failure distribution $f(t)$. In mathematical terms, a stochastic process is defined as a family of random variables $X(t)$ defined on a given probability space and indexed by t belonging to a parameter set T. The set T is the time sequence of the process and it can be discrete ($T = \{0,1,2,....,t\}$) or continuous ($T = \{0, \infty\}$). Here we treat two stochastic processes that are generally accepted in maintenance optimization of infrastructural assets: (discrete-state) Markov chains and (continuous-state) Gamma processes.

Markov chains are the most elementary processes used in deterioration modeling. Markov chains enable us to calculate the probability that an object is in a certain state at time t. Markov chains have been extensively used in the context of risk, reliability, and maintenance management for civil infrastructures (Baik, Seok, Jeong, & Abraham, 2006; Edirisinghe, Setunge, & Zhang, 2015; Klutke & Sanchez-Silvia, 2016). Via estimating the transition probabilities from one state to the other, we are able to determine the probability of failure for each time step. Although challenging, those transition probabilities can be estimated directly via data (Baik et al., 2006) or expert judgement (Kosgodagan-Dalla Torre et al., 2017).

Thereafter, a lot more extended stochastic-process models are proposed based on Brownian Motion, Levy processes etc.. Yet, the Gamma process is mainly preferred since it contains solely positive increments. Abdel-Hameed (1975) was the first who proposed to use the Gamma process as a model for random deterioration in time. Despite their elegance, Gamma processes are barely applied with respect to Markov chains. One of the Gamma-process main limitations is often the availability of data on which we base the calculations the Gamma process-parameters. For a mathematical background of the two processes, we refer to appendix II.

5.2. NETWORK-BASED DETERIORATION MODELS

The previously mentioned stochastic processes address uncertainties in (mainly) environmental variables, by treating the increments as random variables. The parameters of the stochastic processes are predominantly estimated or based on available data, originating from objects that are subject to different environments. In other words, the initial assumptions regarding the input variables is our best estimate of how the object deteriorates. As time matures more information becomes available regarding environmental variables due to performing inspections. As it turns out, a powerful method based on Bayes' theorem enables us to update the initial assumptions via the obtained information. This method is called Bayesian Updating or Bayesian Inference.

Theoretically, performing Bayesian Inference to the relatively easy-observable variables influencing the deterioration process would increase the accuracy of the stochastic process and consequentially our prediction regarding the time-to-failure distribution. A model that both facilitates Bayesian Inference and is able to incorporate multiple variables that influence the stochastic process is called a Bayesian Network (BN), initiated in Pearl (1988) and explained in Appendix III. Later on, Murphy (2002) proposed a method based on Dynamic Bayesian Networks (DBNs) that is very suitable to update predictions regarding stochastic processes. DBNs are a special type of BNs. They consist of a sequence of time-slices, each of which consists of one or more BN-nodes. The slices are connected by direct links from nodes in slide t to nodes in slice $t + 1$. Straub (2009) used DBNs to model deterioration due to fatigue cracks. Via updating the inner-process dependencies, the model became able to stress a more accurate belief into the future parameters once one parameter is updated.

Kosgodagan-Dalla Torre et al. (2017) extended this theory to the so-called covariate Dynamic Bayesian Networks (cDBN). Generally, covariates are a synonym of environmental variables. Compared to Straub (2009), Kosgodagan-Dalla Torre et al. (2017) realistically made deterioration dependent on the covariates such as traffic and load. The study applied a cDBN to an interdependent steel bridge-network in the Netherlands to determine their probability of failure in time. Here, a Markov chain is utilized to model the deterioration in time, based on transition probabilities estimated via expert judgement.

Kosgodagan-Dalla Torre et al. (2017) asked experts questions regarding the time it takes to transit the steel bridge from one state to the next state. For this, Kosgodagan-Dalla Torre et al. (2017) used a four state Markov chain. Experts were requested to give their 5th, 50th and 95th quantiles. Then, Cooke's method (Cooke, 1991) is applied to weight the given quantiles by the experts, based on their calibration and information scores. The results of this expert judgement application showed the transit times of the 5th, 50th and 95th quantiles for every transition. The 5th percentile transition times, correspond to low loads on the bridge. The 50th percentiles correspond to medium load on the bridge and the 95th percentiles correspond to the high loads on the bridge. Based on the transit time-percentiles, Kosgodagan-Dalla Torre et al. (2017) was able to determine the transition probabilities that the bridge remains in the same state after one time step via the underlying formula;

$$p_{i,i} = 1 - \frac{1}{E[T_{i,i+1}]} \tag{4}$$

Here, $E[T_{i,i+1}]$ is the transition time, given a load; low, medium or high. In fact, in this way Kosgodagan-Dalla Torre et al. (2017) derived three Markov chain transition matrices corresponding to a certain load.

5.3. LIMITATIONS WITH RESPECT TO THIS RESEARCH

Sperotto, Molina, Torresan, Critto, & Marcomini (2017) reviewed the potentials of discrete BNs for climate change applications and concluded that the main limitations include amongst others: (1) that BNs have limited capacity to deal with continuous variables and (2) the growing complexity of the computational effort in case of complex systems.

We can conclude that many domains such as the engineering domain, require reasoning about the joint behavior of a mixture of discrete and continuous variables (A. Hanea, Morales Napoles, & Ababei, 2015). Focusing on the aim of this research to incorporate gradual trends, we need a model that is capable to produce small increases per time-step. Both reasons indicate that the discrete BNs are not suitable for the purpose of this research.

Thereby, Kosgodagan-Dalla Torre et al. (2017) considers the deterioration of steel bridges due to fatigue crack growth in the deck plate. In the case of bridges, functional failure and technical failure are practically the same; when the bridge is technically failed, the function - to transport vehicles safely from one side to the other - is also failed. The underlying cause is that bridge systems are series-connected; when one component fails, the asset fails. As a consequence, for the application of bridges it is sufficient to predict the point of asset failure by solely modeling the technical deterioration of the component. In contrast, for pump stations, asset components are connected in parallel; when one pump fails, the remaining pumps can also do the job. In those cases, functional failure and technical failure have two different definitions.

5.4. DIFFERENT APPROACH

In accordance with the researches mentioned in the previous chapter, the general idea is to make the deterioration dependent on the environmental variables, in literature also called *covariates*. Due to the limitations of the recently developed deterioration models we are forced to come up with a model that is able to deal with continuous variables; continuous Bayesian Networks. Domains that can handle mixtures of discrete and continuous variables are called hybrid domains and hence, BNs dealing with discrete and continuous variables are called hybrid BNs (HBNs). One method for HBNs that departs from the classical BN-approach is the so-called *non-parametric Bayesian Network (NPBN)* (A. Hanea et al., 2015).

Just as BNs, NPBN are directed acyclic graphs (DAGs) and feature the property of inference. The nodes are (conditional) joint distributions and the arcs represent (conditional) copulae. Bivariate copulae are bivariate distribution that describe the dependency structure between two marginal distribution. In that sense, they separate the effect of dependence from the effect of marginal distributions in a joint distribution. This property is precisely what is making the study of copulas the standard in modern statistics. For a more elaborate explanation on copulae we refer to appendix V. The use of copulas limits the input in NPBNs to the marginal distributions of each variable and the associated (conditional) rank correlations. The output of the NPBN is a sample-set of all variables in the NPBN. For a more detailed explanation regarding the theoretical and mathematical background of NPBNs we refer to chapter 6.

6. NON-PARAMETRIC BAYESIAN NETWORKS

This chapter mathematically describes the most important aspects of Non-Parametric Bayesian Networks. This knowledge is necessary in order to validate the model. A. Hanea (2008), devoted her doctoral thesis to the modeling statistical dependency via NPBNs. In case the reader is interested in the details of this method, we refer to the thesis. In this chapter we will elaborate only on the most relevant concepts in order to understand the characteristics of a NPBN. NPBN's combine the methods of the classical Bayesian Networks and copulae to efficiently deal with dependent continuous variables. To read this chapter pre-knowledge in BN's and copulae is advised, see appendix III and IV respectively.

6.1. JOINT PROBABILITY

NPBNs are introduced in Kurowicka & Cooke (2005) and extended in Hanea, Kurowicka, & Cooke (2007) and A. M. Hanea, Kurowicka, Cooke, & Ababei (2010). NPBNs construct the joint distributions of a set of variables represented as a Directed Acyclic Graph (DAG) by coupling the marginal distributions of all variables with the dependence structure constructed from bivariate pieces of dependence. Here, just as for the 'classical' BNs, the main aim is to calculate the joint probability of n-variables which can be determined according the underlying formula (A. Hanea et al., 2015).

$$f_{1,\dots,n}(x_1, \dots, x_2) = f_1(x_1) \prod_{i=2}^{n} f_{i|Pa(i)}(x_i|x_{Pa(i)}) \tag{5}$$

NPBNs distinguish themselves from BNs by using (conditional) one-parameter copulas as represented by the arcs of the DAG, see figure 7. Following Sklar's theorem (Appendix IV), it is possible to construct the joint distribution requiring only the marginal distributions of the variables and a copula with their copula parameters (Clemen & Reilly, 1999).
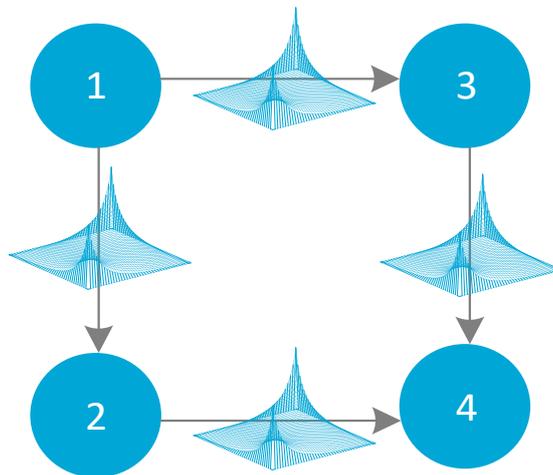
Figure 7: Schematic representation of a simplified Non-Parametric Bayesian Network

Concluding, the input to derive the joint probability of the NPBNs is limited to a number of marginal distributions equal to the number of variables, and number of (conditional) copulae that depend on dependence parameters equal to the number of arcs of the NPBN. The marginal distributions is in most cases empirical, but can also be elicited from experts. The (conditional) copulae can in principle be any copula, but only the Gaussian copula affords the advantages of rapid calculations in large and complex problems. Since the Gaussian copula is utilized, we are interested in the Gaussian copula parameter; the product moment correlation ρ. Since the Gaussian copula is essentially an bivariate normal

distribution, Pearson transformation[1] enables it to think in terms of rank correlations. Therefore dependence parameters between parent and child in the NPBN are described by (conditional) rank correlations r. The (conditional) bivariate copulae are then used as building blocks of the joint distribution.

6.2. BAYESIAN REASONING

The direction of the arrows add information regarding to the conditional (in)dependencies within the graph or more simplified: it translates the reasoning behind the variables into conditional dependencies. This encoded information is also known as d-separation or conditional independence structures. We can basically distinguish three structures:

$① → ② → ③$ The first structure states that without observing 2, observing 1 would say something about the distribution of 3, or in probabilistic language; $X_1 \not\perp X_3$. That is 1 is not marginally independent of 3. However, if 2 is known, then 1 would not add extra information to explain 3, that is 1 and 3 are conditionally independent, given 2: $X_1 \perp X_3|X_2$.

$① ← ② → ③$ The second structure is similar as the first case: $X_1 \not\perp X_3$, but $X_1 \perp X_3|X_2$.

$① → ② ← ③$ The third structure deviates from the latter cases. Here, 1 and 3 are marginally independent $X_1 \perp X_3$, but not conditionally independent when 2 is observed. That is, if we observe 1(3), without observing 2 that would say nothing about 3(1), respectively. In contrast, if we observe 2, then observing 1(3) will say something additional about the distribution of 3(1), so $X_1 \not\perp X_3|X_2$.

6.3. CONDITIONAL RANK CORRELATIONS

The encoded information in the DAG via the prementioned structures results in conditional dependencies between variables. But how are we able to incorporate conditional independencies into the NPBN?

In NPBNs the conditional dependencies can be derived by adapting the Gaussian copula dependence parameter ρ via rank correlations r. In mathematical terms, the conditional rank correlations of variables X_i and X_j given X_k and X_z is:

$$r(X_i, X_j \,|X_k, \dots, X_z) = r(\widetilde{X}_i, \widetilde{X}_j) \tag{6}$$

Where $(\widetilde{X}_i, \widetilde{X}_j)$ has the distribution of (X_i, X_j) given $(X_k = x_k, \dots, X_z = x_z)$. Then the (conditional) copulae are assigned to the arcs of the NPBN according to a protocol that depends on a ordering of the parent nodes (A. Hanea, 2008).

Briefly, this ordering implies that the correlation between the child and its first parent will be an unconditional rank correlation, and the correlations between the child and its next parents (in the ordering) will be conditioned on the values of the previous parents, etc.. Let's assume the underlying configuration of figure 8. Here node 2 and 3 both have only one parent. According to the ordering protocol, the arc will be assigned to an unconditional rank correlation. In contrast, node 4 has two parents, namely 2 and 3. Dependent on the order, one arc will be assigned to a conditional rank

---

[1] The Pearson transformation gives the relation between the product moment correlation and the rank correlation for joint normal distributions: $\rho(X, Y) = 2\sin(\frac{\pi}{6} \cdot r(X, Y))$.

correlation and the other to an unconditional rank correlation. In this case we assume node 3 as the first parent, then r_{24} will be conditionalized on node 3: $r_{24|3}$. For a more extended example and application of so-called D-vines, we refer to A. M. Hanea, Kurowicka, Cooke, & Ababei (2010).
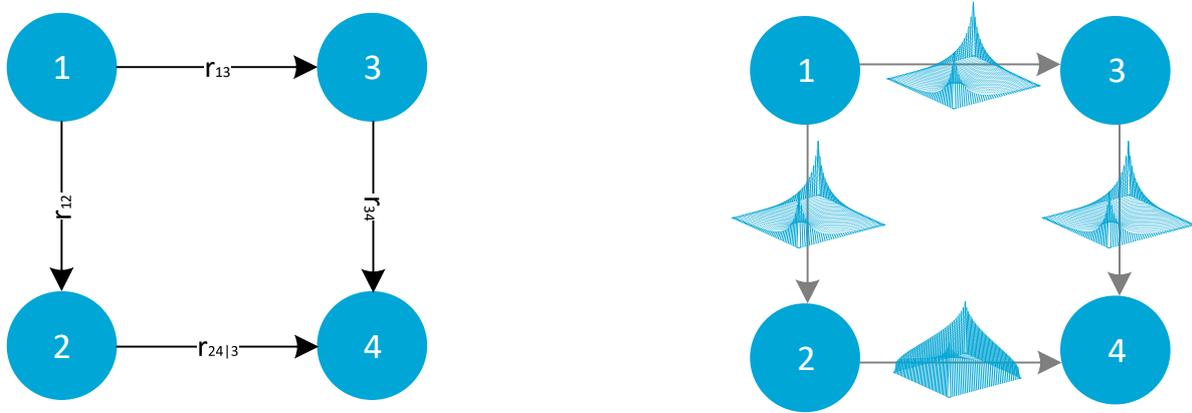
Figure 8: Schematic representation of a NPBN.

In mathematical terms we can summarize the ordering protocol as follows: for each variable X_i with m-parents $X_1 = pa_1(X_i), \ldots, X_k = pa_m(X_i)$, associate the arc $pa_j(X_i) \rightarrow X_i$ with the rank correlation:

$$r[X_i, pa_j(X_i)], j = 1$$

$$r[X_i, pa_j(X_i)|pa_1(X_i), \ldots, pa_{j-1}(X_i)], j = 2, \ldots, m$$

(7)

To calculate the conditional Gaussian copula parameters between nodes 2 and 4, we first make a transform the rank correlations to Pearson's product-moment correlations. Thereafter we utilize the underlying recursive equation (A. Hanea et al., 2007; A. M. Hanea et al., 2010).

$$\rho_{1,2|3,\ldots,n} = \frac{\rho_{1,2|4,\ldots,n} - (\rho_{1,3|4,\ldots,n})(\rho_{2,3|4,\ldots,n})}{\sqrt{(1 - \rho_{1,3|4,\ldots,n}^2)(1 - \rho_{2,3|4,\ldots,n}^2)}}$$

(8)

Via the conditional product moment correlation $\rho_{1,2|3,\ldots,n}$ one can determine the conditional Gaussian copula in the NPBN which transforms the original marginal distribution to the conditional marginal distributions.

6.4. VALIDATION METHODS

In order to check whether the NPBN represents the dependence structure of the empirical data we have several methods available, beginning with the building blocks of the NPBN; the Gaussian copula.

6.4.1. Validation Copulae

Copula validation is based on the goodness-of-fit methods such as the 'blanket'-test and semi-correlations explained in Appendix V.

6.4.2. Validation NPBN

For the validation of the total NPBN we can compare the determinants of the rank correlation matrices. A determinant equal to 1 represents independence among all variables. A determinant equal to 0 represents linear dependence between the variables. Based on the determinants we can perform two validation tests. Stepwise protocols to perform both tests can be read in A. Hanea et al. (2015).

The first test measures the suitability of the Gaussian copula to represent the empirical data dependence structure. As mentioned different copulae can be applied in order to create the dependence structure, but the Gaussian copula is the standard since it affords advantages regarding the efficiency of sampling. Due to the assumption of the Gaussian copula, the determinant of the empirical normal rank correlation matrix (DNR) will differ from the determinant of the empirical rank correlation matrix (DER) since the latter is based on the empirical copula.

Since we have the dependence parameter to construct the accompanied Gaussian copula and the marginal distributions, we are able to draw samples from the 'Gaussian' data. Then the rank correlations can be calculated with its accompanied correlation matrix. By re-sampling the normal data multiple times, we can obtain the empirical DNR and extract the 5th and 95th percentiles of the distribution. If the DER falls within the quantiles of the empirical DNR-distribution, the joint Gaussian copula represents the empirical dependence structure sufficiently.

The second test concerns the comparison of the determinant of the rank correlation matrix of a BN constructed under the assumption of the Gaussian copula (DBN) and the DNR. Since purpose of a BN is to draw the (conditional) dependencies and (conditional) the DBN > DNR, in general unless the BN is saturated. Starting adding arcs between the nodes, the DBN value increases since we increase the (conditional) dependencies.

Instead of re-sampling the DNR as in the previous test, we now re-sample the DBN to obtain the distribution of the DBN. If the DNR is within the 90% central confidence band of the DBNs distribution, then it can be concluded that enough dependencies are captured in the model. If not, then find a pair of variables such that the arc between them is not in the DAG and their rank correlation is greater than the rank correlation of any other pair not in the DAG. In this way we are able to tweak the DBNs value.

6.5.   UNINET SOFTWARE

UniNet, software initially developed by the Technical University of Delft (TU Delft), enables us to create NPBNs networks in an efficient way. UniNet is freely downloadable for academic purposes from http://www.lighttwist.net/wp/uninet.

The idea behind the use of UniNet is to model the dependence structure among all variables, under the assumption of the Gaussian copula. When the dependence structure is known, UniNet can calculate the (conditional) rank correlations for every arc and its associated (conditional) marginal distributions. Based on the known (conditional) copulae and (conditional) distributions, UniNet is able to sample from the (conditional) distributions. In this way, we are able to retrieve enormous amounts of samples with a realistic dependence structure. Although applied to different types of cases as the one proposed in this research, A. M. Hanea et al., (2010) and Morales-Nápoles & Steenbergen (2014), provide a practical example how to work and reason with UniNet and NPBNs.

7. DETERIORATION MODELLING VIA NPBNS

Chapter 5 concluded that the available models are not suitable to treat the problem covered in this research since 'classical' BNs have limited capacity to deal with continuous variables. Non-Parametric Bayesian provided a solution to deal with discrete- and continuous variables. This chapter will therefore combine the theory of NPBNs and Markov chains to propose a model similar as in Kosgodagan-Dalla Torre et al. (2017).

7.1. AVAILABILITY-MARKOV CHAIN

Paragraph 5.1 introduced two generally accepted methods to model the deterioration; the Markov chain and the Gamma Process, with input variables transition probabilities and inspection data, respectively. Although Gamma processes provide a very elegant way to model the deterioration in time, Markov chains are easier in use and more fitting to calculate the probability of unavailability. Here we propose a very simplified way to model deterioration, based on Kosgodagan-Dalla Torre et al. (2017).
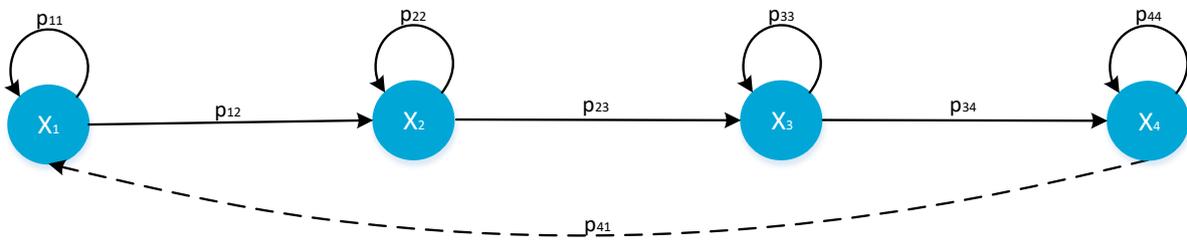
Figure 9: Availability Markov chain

Then, $p_{i,i}$ is the probability that the component is still in function after one time step. $p_{i,i+1}$ is the probability that the component switches from state X_i to X_{i+1} which is equal to its probability of failure. $p_{4,4}$ is the probability that the component remains it its unavailable state; the component is in repair. Finally, p_{41} marks the transaction probability that the component switches from the repair state to the 'as new' state. Its transition matrix can be defined as follows:

$$\boldsymbol{P} = \begin{pmatrix} p_{11} & p_{12} & 0 & 0 \\ 0 & p_{22} & p_{23} & 0 \\ 0 & 0 & p_{33} & p_{34} \\ p_{41} & 0 & 0 & p_{44} \end{pmatrix} \tag{9}$$

Note that this Markov chain has no absorbing state as the Markov chain mentioned in Appendix II. This Markov chain will converge towards a certain 'stable' probability. Estimating the transition probabilities directly is known to be challenging. Therefore, Kosgodagan-Dalla Torre et al. (2017) assumed sequential degradation and proposed a way to derive the transition probabilities via expected transition times via the underlying equation.

$$p_{i,i} = 1 - \frac{1}{E[T_{i,i+1}]} \tag{10}$$

Where $p_{i,i}$ is the probability that the system remains in a particular state after one time-step, and $E[T_{i,i+1}]$ is the expected transition time. Kosgodagan-Dalla Torre et al. (2017) performed expert judgement to derive the 5th, 50th and 95th percentiles of the transition times T_{12} , T_{23} , T_{34} ,

corresponding to *High, Medium* and *Low* loads respectively. In this way, the probability that the component would transit to the next state is higher, given a *High* load; medium, given a *Medium* load; and lower, given a *Low* load. Although this is a very elegant way to model the environment-deterioration dependency, this study strives for a way to model the dependency for continuous states. This is where the non-parametric Bayesian Network comes in.

p_{44}, marks the probability that the component remains unavailable after a specified time step. The transition time to the 'as new' state, T_{41}, can be interpreted as the Mean Time to Repair (MTTR).

7.2. DETERIORATION WITH NON-PARAMETRIC BAYESIAN NETWORK

The transition probabilities can be calculated via the expected transition times. Here we propose a framework based on NPBNs to model the dependencies between the environment and transition times, focusing on pumps. For simplicity is assumed that the transition times only depend on the operating time of the pumps per year. See the underlying figure for a stable environment.
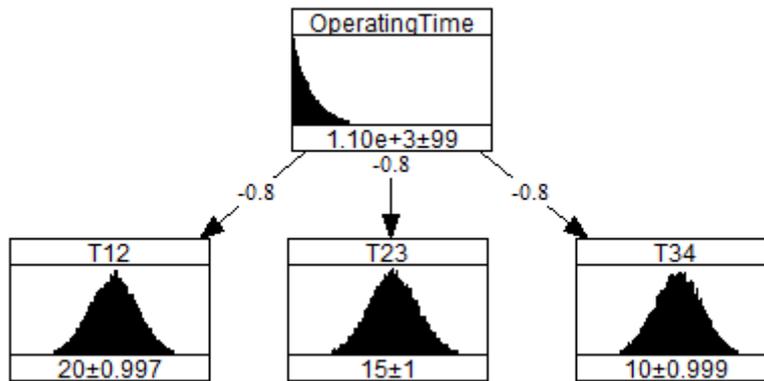
Figure 10: Unconditionalized Deterioration NPBN

The transition distributions can – just as in Kosgodagan-Dalla Torre et al. (2017) – be determined via expect judgement. Then, calculating the expected value of the transition times returns the expected transition time $E[T_{i,i+1}]$. In the above case, the accompanied transition matrix would be as follows:

$$\boldsymbol{P(0)} = \begin{pmatrix} 0.95 & 0.05 & 0 & 0 \\ 0 & 0.933 & 0.067 & 0 \\ 0 & 0 & 0.9 & 0.1 \\ p_{41} & 0 & 0 & p_{44} \end{pmatrix} \tag{11}$$

Now we introduce the dependency on the environment by performing inference. Let's assume that at time *t = 1* the pump station is in a calm environment and only assumes lower values, for example 1000 – 1200 hours per year. Then the NPBN will change as follows:
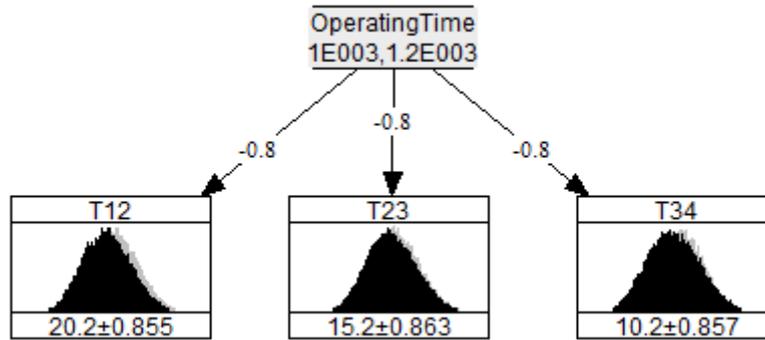
Figure 11: Conditionalized Deterioration NPBN

Note that the expected transition times increased as a result of the negative rank correlations. Now, the transition matrix would be as follows:

$$P(1) = \begin{pmatrix} 0.951 & 0.049 & 0 & 0 \\ 0 & 0.934 & 0.066 & 0 \\ 0 & 0 & 0.902 & 0.098 \\ p_{41} & 0 & 0 & p_{44} \end{pmatrix}$$

Note that the probabilities that the system remains in the same state have gone up, which is exactly what was expected. Now, based on the beginning state X_0, we are able to calculate the probabilities that the system is in a specific state at a random point in time via the underlying equation.

$$X_t = X_0 * \prod (P(1), P(2), \dots, P(t)) \tag{12}$$

7.3. LIMITATIONS

Adopting this model contains a few assumptions. First, the Markov chain has no memory. So if the component in state 1 for a long time, the probability that it transits to the next state is equal to the probability that it transits the first time. Second, every time a component fails, the component returns to its 'as new' state. In other words, the model assumes perfect maintenance.

8. NPBN-SUPPORTED RELIABILITY ANALYSIS

Chapter 6 introduced non-parametric Bayesian Networks as a tool to model the dependencies among variables. This chapter proposes two conceptual NPBNs to model (1) the total discharge capacity and its dependencies to sea level rise and (2) the required discharge capacity and its dependencies to rainfall. Thereafter this part will be closed with a road map to conduct a reliability analysis for pumping stations and include the introduction of long term trends.

8.1. MODELING ENVIRONMENTAL DEPENDENCIES WITH NPBNS

The impact on the resistance side of the limit state function – the total pump station discharge capacity – can be divided into the availability and individual pump capacity. In paragraph 1.5 is stated that both are impacted by sea level rise. For that reason, the conceptual NPBN can be divided into two branches, see figure 12; one representing the availability impacted by deterioration and one representing the individual pump capacity.

For the resistance variables, we depart from the left branch, which can be compared to the figures in chapter 7. For simplicity, here is one transit time taken into account; $Time\_to\_Failure$. Which is basically just the transit time that the component transits from available to unavailable. Here we want to emphasize the more transition time nodes could be added just as in chapter 7. Via the NPBN above we basically stress a belief into the operating time, given a certain sea level rise, and thereafter we stress a belief into the time to failure, given a certain operating time per year. Then, chapter 7 already elaborated on the use of this NPBN-output to the Markov chain in order to model deterioration. For right hand branch, the pump capacity can be deterministically derived from the water head, based on the Q-H curve described in chapter 1.5. The water head can on its turn be derived via the sea level rise and the water fluctuations influenced by storm surges and tidal variations.

The solicitation variable; the required discharge capacity is dependent on – for example - the rainfall in the area, the discharged amounts of excess water from the surrounding water systems etc. This model can contain limitless of variables that influence the required discharge capacity. Practice applications show that often sufficient data is available that can be used to define the variables marginal distributions. Therefore it is recommended to add the variables of which data is available.
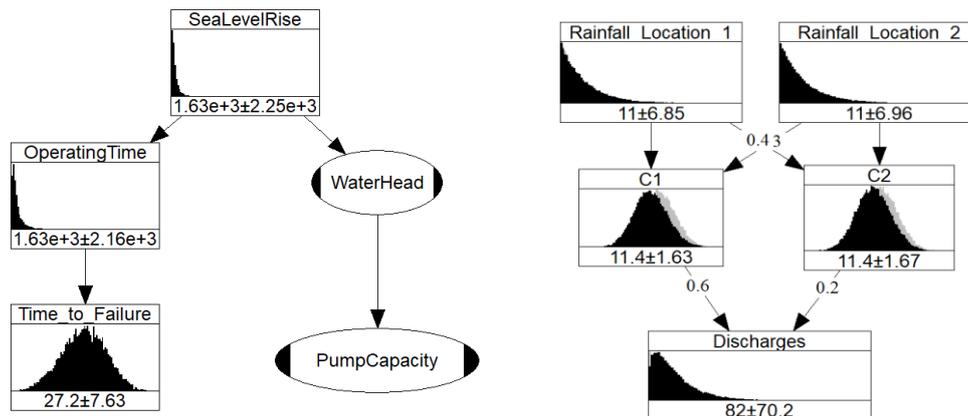
Figure 12: Conceptual NPBNs to model the three effects of sea level rise

8.2. INTRODUCING LONG TERM TRENDS

When all directly and indirectly impacting variables are gathered, we are ready to incorporate the long-term trend. Basically there are two ways to conduct this, based on the trend and data itself. First, in case the data covers the long term trend, Bayesian Inference can be used to update initial assumptions via the future obtained information. As it turns out in most cases, we already have predictions of how the long term trend would develop. This predictive information can be used to update our model before the information is actually observed. In other words, we use BN's largest beneficial element – inference – to update our beliefs into the time-to-failure distribution, see figure 13.

The best example is rainfall. Climate change will increase the number of extreme rainfall occurences. In this research is assumed that the extreme rainfall amount is already covered in the historic data. Then, by conditionalizing on the higher values of the rainfall is aimed to model the increase in discharge occurences in time. See figure 13 - top figures, notice that the mean of the time-to-failure distribution ($Time\_to\_Failure$) decreases in time, which implies a decrease of the Mean Time To Failure (MTTF) and a change in the Markov chain.

If the historic dataset does not capture the future values, the future marginal distributions must be estimated via – for example – Expert Judgement, see figure 13 – lower figures. This counts for sea level rise, since the historic water levels have never assumed any values that sea level rise-models predict. Paragraph 10.5 shows a way to determine the distributions in time by solely estimating the latest marginal distribution of the considered time span.
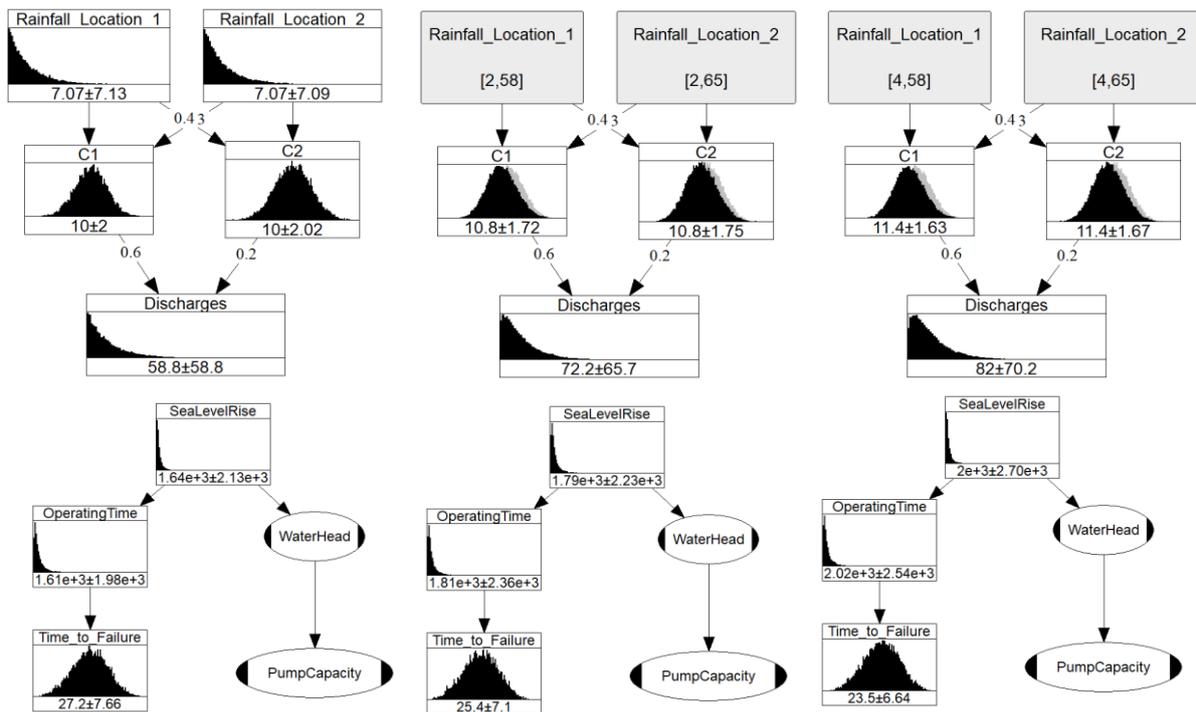
Figure 13: Incorporating long term trends

8.3. PUMP STATION RELIABILITY ANALYSIS ROADMAP

Dependency modeling via NPBNs enables it to determine the input distributions of the limit state function in time. Here, a total conceptual framework to calculate the pump station reliability in time is proposed that combines the tools given in the previous chapters, see figure 14.
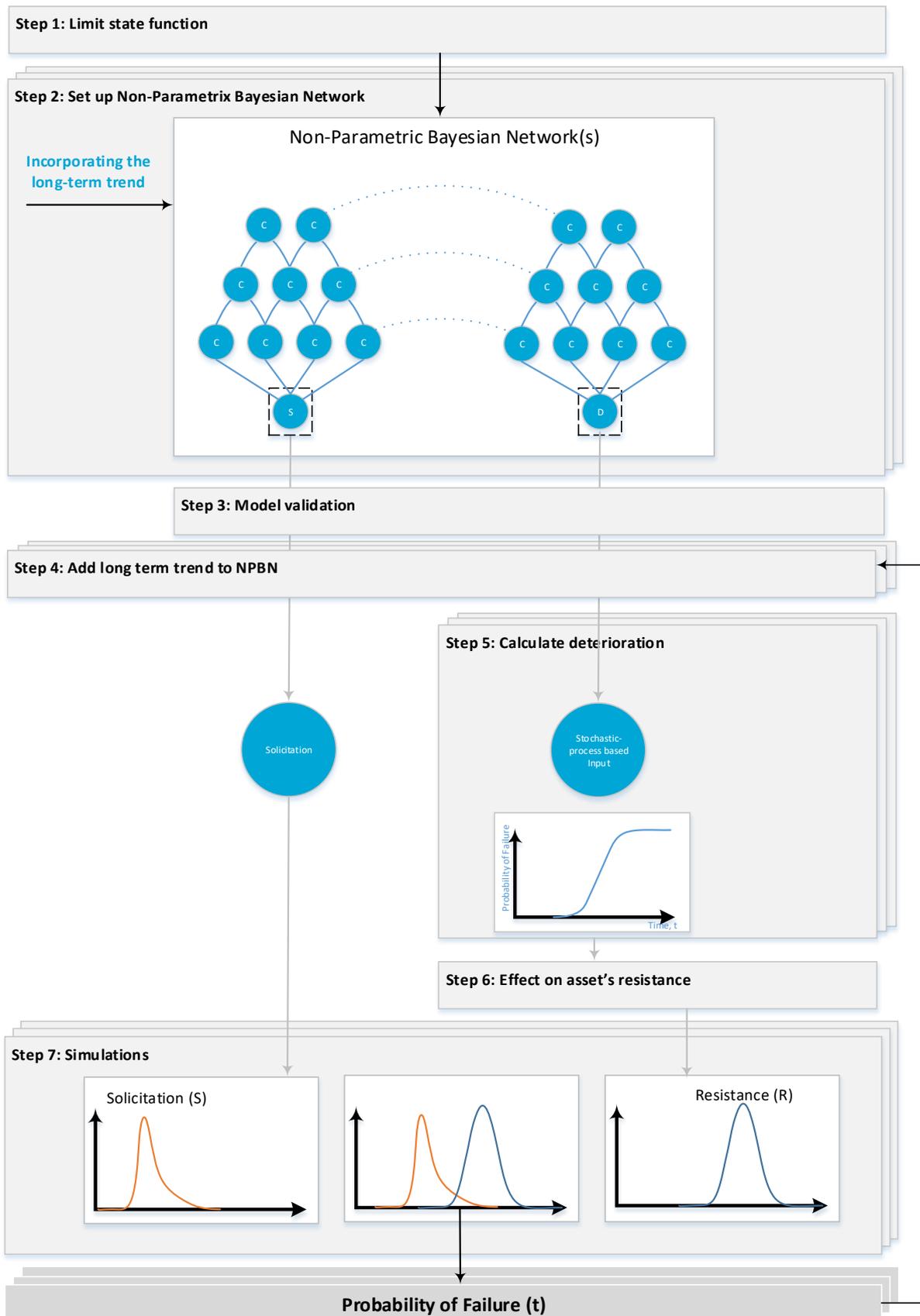
Figure 14: Methodology road map

Step 1: Limit state function

The point of departure is the limit state function, see equation (3). The limit state function must be fitted to the failure definition. The limit state function defines the variables of which we aim to describe the dependence between the covariates directly influenced by the long term trend.

Step 2: Set-up Non-Parametric Bayesian Network

This step concerns the set-up of the NPBN. Notice that the roadmap shows two NPBNs; one to model the impact of the long term trend on the resistance and one to model the impact of the long term trend on the solicitation side of limit state function. The conceptual models of chapter 8.1 can be used as a handle to gather data. Again, countless covariates could be added to the NPBN to finally arrive at an covariate that is directly influenced by the long term trend. Thereby, it could also be possible that an variable influences both the resistance side as the solicitation side. NPBNs can perfectly cover those interdependencies.

Step 3: Model Validation

When the NPBN is ready for use and able to stress beliefs in the deterioration-input variable(s) and solicitation variable(s). The NPBN must be verified whether it sufficiently represents the data and its dependency structure. Chapter 6 proposes two ways to validate the NPBN; verifying whether the copula represents the data between two variables, and whether the joint Gaussian copula is a sufficient assumption for the total NPBN.

Step 4: Add long-term tend to NPBN

For the resistance side, future distributions must be estimated. A defendable method is preferred here, for example expert judgement. For the solicitation side, inference is sufficient. See chapter 8.2.

Step 5: Calculate deterioration

The output-samples per time unit are used as input into the Markov chain, see chapter 7.

Step 6: Effect on asset's resistance

How the deterioration of one component is related to the overall resistance is highly dependent on the configuration of the component in the pump itself. Step 5 calculates the probability that the component is unavailable. Then, via the pump decomposition one is able to calculate the effect on the total pump unavailability that affects the availability variable of the limit state function.

Step 7: Simulations

Following the predefined steps one will end up with a solicitation distribution and a resistance distribution at a particular time t. The solicitation distribution directly gained by the NPBN and the resistance distribution indirectly gained via the NPBN, stochastic-process-based deterioration models and asset decomposition. With simulating the distribution samples through the system we will end up with the probability of failure.

After calculating the probability of failure for that particular time, the procedure repeats itself from step 4: the NPBN must be conditioned on the expected values for the directly impacted covariates for that particular year. Continuing this procedure will lead to the probability of asset failure in time, taking into account the deterioration and technical lifecycle enhancement actions for critical assets.

Part III - Application to Case Study

This page was intentionally left blanc

9. CASE STUDY: PUMP STATION IJMUIDEN

The research will continue with applying the theory in practice. First we will introduce the case study of discharge complex IJmuiden. Thereafter we will elaborate on the reasoning behind the modeling procedure. The roadmap of part II is used as the thread through the model building.

9.1. SYSTEM DESCRIPTION

The complex can discharge the excess water in two ways: by pumps and by sluices. The pumping station contains six pumps in total, situated below sea level. Four of the six are Stork-pumps with a maximum capacity of 40 m³/s, coming from the year 1976. The remaining two pumps are Nijhuis-pumps with a maximum capacity of 50 m³/s and were placed in 2004 in order to extend the maximum pump capacity from 160 m³/s to 260 m³/s.

In comparison to the pumping station, the sluices have a much higher capacity, namely 500 – 700 m³/s. Thereby, pumping these large amounts of water towards the sea costs high amounts of energy, while free discharge only require the opening of the sluice gates. Logically the water management prefers free discharge with respect to discharge by pumping. However it is not that simple.
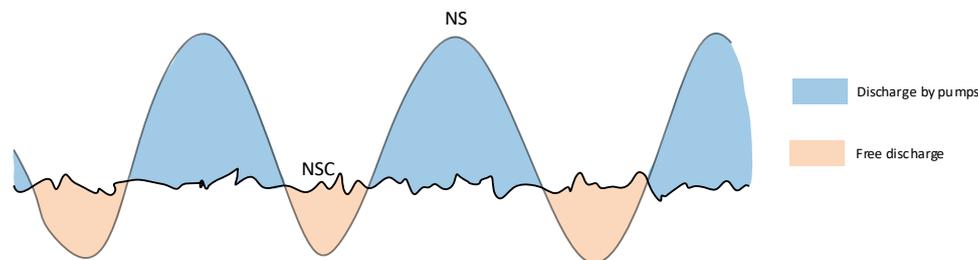
Figure 15: Time slots of discharge by pumps and free discharge

See figure 15, free discharge is only possible when the North Sea (NS) water level is below the water level of the North Sea Channel (NSC) at the considered point in time ($h_{ns} < h_{nsc}$). If the water level of the NS is higher than the water level of the NSC ($h_{ns} \geq h_{nsc}$), the pumps are switched 'on' and excess water will be pumped into the NS. Under normal conditions, the NSC/ARC-water levels fluctuate between NAP - 0,30m and NAP – 0,55m. The NS water level fluctuates between the NAP - 1,03m and the NAP – 2,20m, under normal conditions. The water level on the NS-side is subject to incoming tidal waves. So as a rule of thumb, the complex will freely discharge during two intervals a day and also intervals via the pumping station.

The function of the complex is to regulate the water level of the North Sea Channel (NSC)- and Amsterdam-Rhine Channel (ARC) water systems, see figure 16. The discharge complex of IJmuiden is for approximately 95% responsible for the total water discharge out of the water system towards the sea and is therefore an essential link in the water management of the region. But what are the factors that influence the dischargeable amount of water?

The dischargeable amount of water is determined via the water balance of the system where, among others, the rainfall, the discharge of surrounding water systems into the NSC-ARC-water system and the Lek- and Markermeer-inlets are discounted. One limitation of the NSC/ARC-water system is that it has limited storage capacity. This implies that when the complex is somehow not able to discharge the excess water into the sea, the water cannot be stored elsewhere what consequentially results to an instant increase of the ground water level. To illustrate the accompanied problems; with an water level

increase of 0,8m with respect to the NSC/ARC-target level, the cellars in Amsterdam will flood with all its consequences (van der Wiel, Persoon, & Stiksman, 2013). When the NSC-water level exceeds a certain critical water level, the surrounding water boards are forbidden to pump their excess water into the NSC/ARC water system. This consequentially results in higher water levels in these water systems.
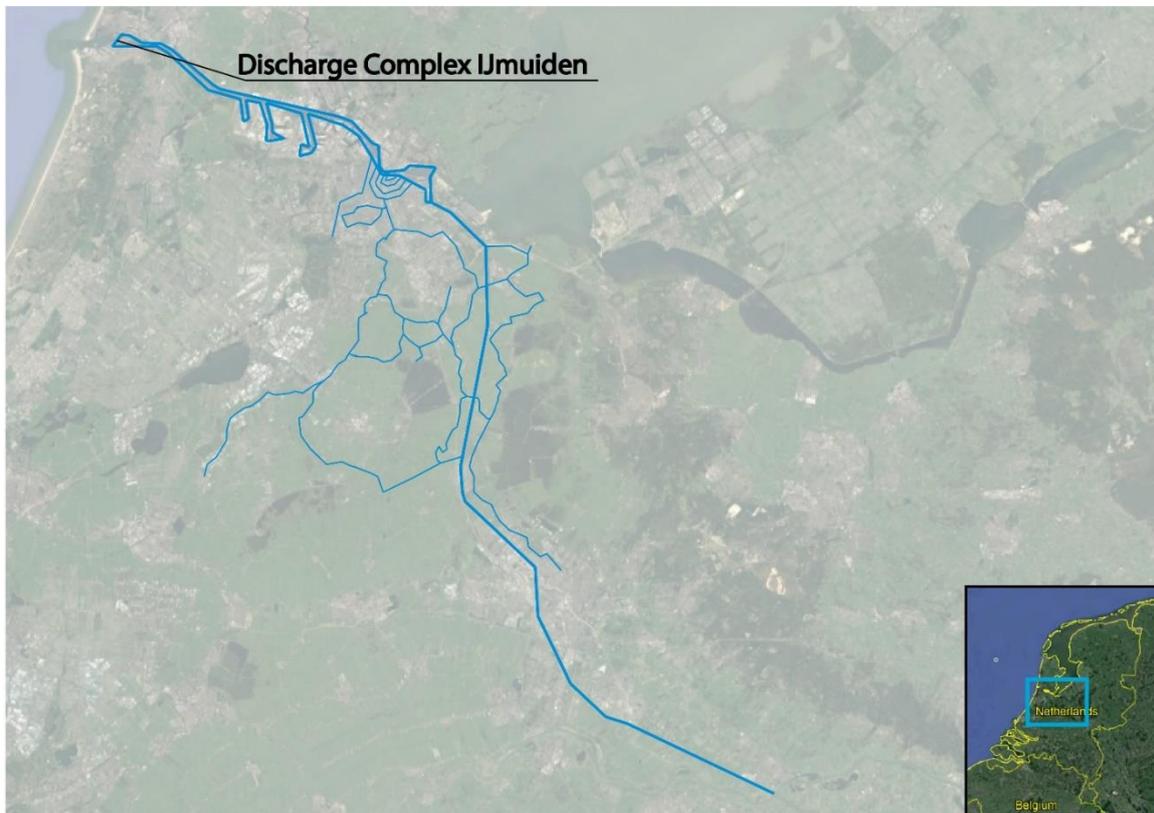
Figure 16: Overview of the NSC/ARC water system

9.2. CASE PROBLEM

Rijkswaterstaat is the execution agency of the Dutch Ministry of Infrastructure and Environment and makes sure that the Netherlands are a safe, livable and accessible country. Because a large part of the Netherlands is located below sea level, the country relies enormously on hydraulic structures such as navigation locks, bridges, dikes and storm surge barriers. Rijkswaterstaat maintains approximately 650 of these structures, of which the larger part originates from the 1950's (Jonkman et al., 2018; Rijkswaterstaat, 2014). Because the designed lifetime of these structures is around the 80 – 100 years, the larger amount will approach the end of its designed lifetime. This illustrates the problem of Rijkswaterstaat whether to renovate or replace the large amount of these important assets.

Since the pumping station of IJmuiden has been built in 1975, the structural capacity is not an issue yet. However, many doubts exist whether the complex' functional capacity is still sufficient for the longer term: can the discharge complex fulfill its functional requirements for the longer term in a reliable way? We translate this question into a failure definition.

> **Failure definition:** The asset fails when the required discharge capacity cannot be delivered when required

The main contributing environmental factors that affect the complex' reliability are the effects of climate change, specifically the increase in peak discharges and sea level rise.

9.3.   EFFECTS OF CLIMATE CHANGE TO PUMP STATION IJMUIDEN

As mentioned in paragraph 9.1. the timeslots that the pumps operate, is dependent on the NS water level and the NSC water level. Unfortunately, the sea level raises due to climate change. The Royal Dutch Royal Meteorological Institute (KNMI) states that the sea level will increase with a rate of 1-6 mm per year until 2030 (Klein Tank, Beersma, Bessembinder, Hurk, & Lenderink, 2015). This would mean that the time slots that pumping is required will increase over time, which enhances the wear of the pumps, se figure 17. But what is the magnitude of impact to the timeslots that the pumps are on?
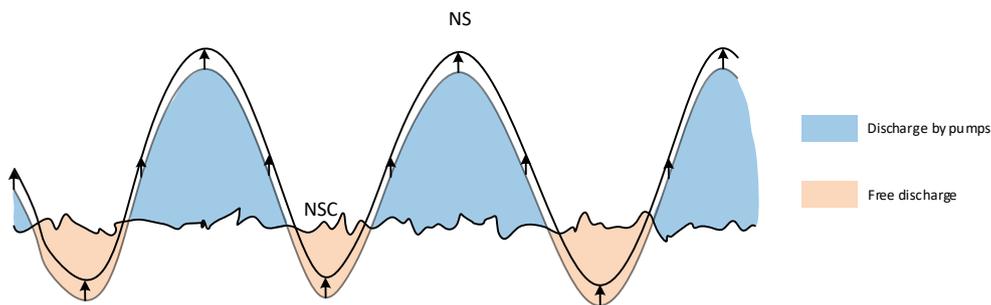
Figure 17: Effect of sea level rise on the pump- and free discharge time slots

The second effect, also introduced by sea level rise can best be illustrated by use of the Q-H curve, see figure 18. The Q-H curve illustrates the relation between the decreasing pump volume when the water head increases. Sea level rise causes an overall increase of the water head, which implies that the pumps have to function more often in the lower pump volume regime. Since this affects all pumps simultaneously, the total required pump volume will drop dramatically.
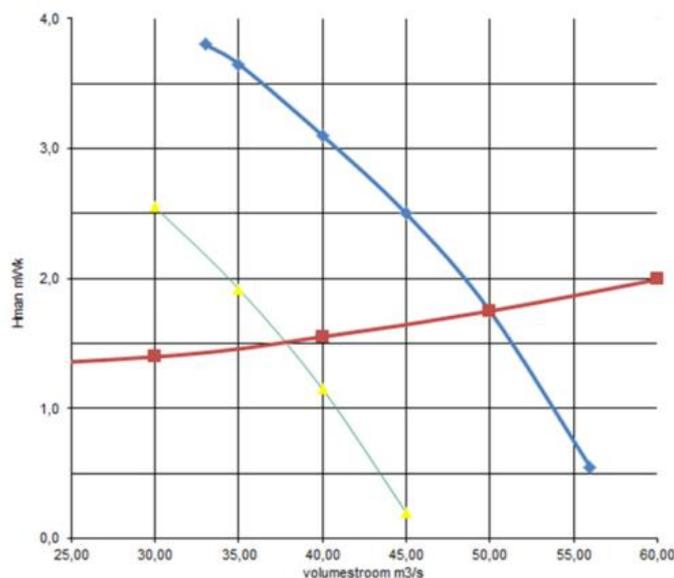
Figure 18: Pumpcurve Nijhuis (blue) and Stork (green) pumps.(van der Wiel et al., 2013)

The final effect induced by climate change is the increase of extreme weather occurrences. Normally, the excess water in the system, of which rainfall is a large contributor, will be drained and discharged towards the ARC or NSC and thereafter discharged by the discharge complex in IJmuiden. Since the

amount of rainfall increases with time, the peak-dischargeable amount water will also increase, as does the number of peak-occurrences. Consequentially, the occasion that the maximum capacity, i.e. the occasion that all pumps must be available will occur more often.

To conclude, let's assume a peak discharge of 220 m^3/s. In order to discharge this amount of water, at least 5 of the 6 pumps need to be available to provide this capacity. When two pumps fail at that specific moment, the system can be considered as failed. So combining the effects that (1) the increasing probability of pump failure due to sea level rise and (2) the increasing chance that a peak discharge occurs and maximum pump availability is required, will consequentially lead to an increasing probability of system failure.

10. MODEL BUILDING

This chapter will make reliability analysis as described in chapter 8.3. suitable to the case of pumping station IJmuiden as described in chapter 9. To simplify the problem, we make two preliminary assumptions. Thereafter we depart from the first step of the roadmap in chapter 0: determine the limit state function. On the basis of this limit state function we distinguish three models that will be treated further separately; the discharge model, the capacity model and the availability model.

10.1. PRELIMINARY ASSUMPTIONS

Before applying the methods to the case, we want to stress the main limitations that assist the interpretation of the results. For simplicity we make the following assumption:

> **Preliminary assumption 1:** As mentioned in paragraph 9.1, the ARC-NSC water system has a limited amount of storage capacity. In other words, the water level in the system will rise directly when the discharge complex fails. If the water level raises up to NAP +0.4m, the system fails. For simplicity, we assume immediate water system failure when the discharge complex cannot deliver the required capacity.

With this assumption we can state that the probability of water system-failure is equal to the probability of discharge complex-failure. In reality however, the water management of the NSC/ARC-region is able to store the excess water that cannot be pumped due to limited pump capacity. The remaining water will raise the water level in the system which again does not necessarily lead to system failure immediately.

Next to the practical assumption of immediate failure when the discharge complex cannot deliver the required capacity, another assumption can be made in order to simplify the model.

> **Preliminary assumption 2:** The second assumption concerns modeling the discharges over time. Here we assume that the probability that the system fails during lower discharges is negligibly small. The system is more likely to fail when high-discharges occur simultaneously with a particular (reduced) pump capacity.

To make this assumption clear, see the figure below. The black line represents the discharges over a time span of four years. The blue line represents the discharge capacity. The first 'drop' of the pump capacity is equal to the unavailability of one pump. However, this drop does not lead immediately to failure of the system, since the discharges at that point in time are low. In contrast, the second 'drop' of the blue line leads to system failure since the discharge is higher than the discharge capacity at that point in time. The third 'drop' is equal to the unavailability of four pumps that makes the pump capacity intersect the required discharge and consequentially leads to pump failure. The assumption states that the latter case, where four pumps fail at a random moment in time, is negligibly small compared to the failure of the second 'drop' where only one pump fails simultaneously with a high discharge.
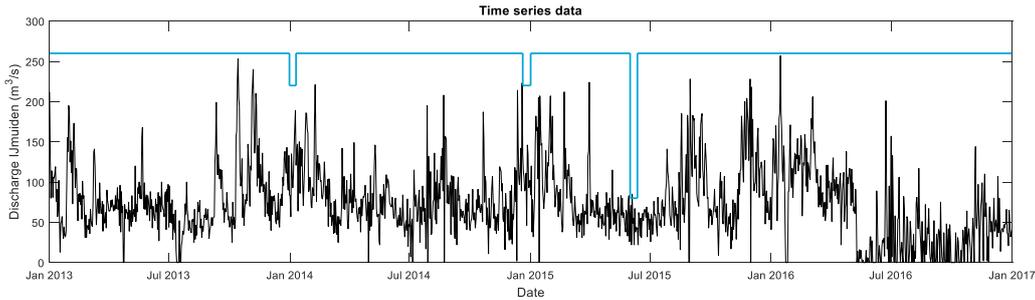
Figure 19: Data time series

Based on this assumption, we do not have to model the whole time series. Modeling the whole time series would be much more complex and the benefits would be limited. Therefore, this assumption enables us to model the peak discharges over a year instead of the whole time series. Associated with the preliminary assumption comes one of the main limitations of this research: we are only able to model per peak-discharge occurence.

10.2. LIMIT STATE FUNCTION

The starting point of the quantitative analysis is the limit state function as described in chapter 4. The function describes the boundary between the states where the system can perform according the predefined requirements and the states where those requirements are not met. Here we translate the failure definition into the language of the limit state function; resistance and solicitation. The probability that the discharge complex cannot pump the required discharge can be merged in the following limit state function:

$$Z(t) = \left(\sum_{n=1}^{6} A_{pump,n}(t) * C_{pump}(t) \right) - Q_{required}(t) \tag{13}$$

Where $A_{pump,n}(t)$ represents the availability of the pump at a random point in time, which is a discrete value with 1 representing pump-availability and 0 pump-unavailability. $C_{pump}(t)$ is the discharge capacity of one pump (m³/s) at a random moment in time and $Q_{required}(t)$ is the required discharge (m³/s) at a random moment in time, in order to pump all excess water from the water system into the NS. The capacity of the pumps represents the resistance of the system and the required discharges the solicitation. The system fails when $Z(t)$ drops below zero. Thus, the probability of failure P_f equals $P(Z < 0)$.

10.3. SET-UP NON-PARAMETRIC BAYESIAN NETWORK

This paragraph elaborates on the reasoning to derive the input for the NPBN; the marginal distribution and rank correlations.

10.3.1. Overall Model

Figure 20 shows the overall NPBN which functions as the basis of the whole model. We distinguish two separate models: the solicitation model and the resistance model. As mentioned in chapter 6, the input in NPBNs is limited to marginal distributions and conditional rank correlations. Most often there is a limited amount of data available or is not representative for the case. This also counted for this case. The solicitation model to derive the discharges of IJmuiden is solely based on datasets of the surrounding discharge stations and precipitation-amounts. The output of the solicitation-NPBN is directly used as input for the simulations, see the roadmap of paragraph 8.3.

In contrast, the resistance-model is a mixture of several manners for deriving the marginal distributions. Here, expert judgement, defendable estimations and deterministic relations are used to putt a belief in the time-to-failure distribution and the pump capacity in time, described by the Q-H curve. See figure 20, the green nodes indicate that the marginal distribution determined by data-sets, the blue nodes show that the marginal distribution are determined via expert judgement, the red nodes are marginal distributions determined by estimations and finally the white nodes are determined by deterministic functions. The time-to-failure distributions are used as input for the deterioration model, a Markov chain, conceptually elaborated in chapter 7 and fitted to the case paragraph 10.6. A more detailed figure can be found in appendix VII.
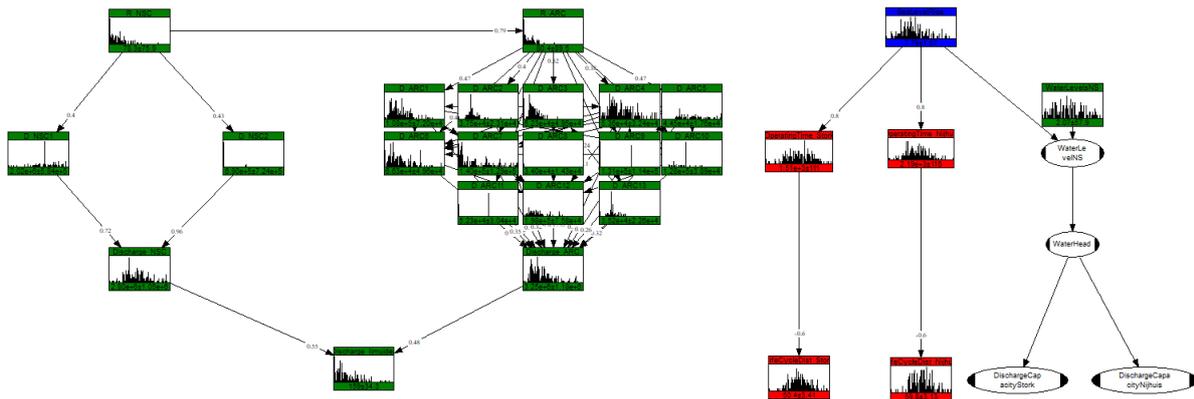
Figure 20: Overall NPBN

10.3.2. Solicitation Model

The fundamental goal for this model is to stress a belief into the discharges at IJmuiden, based on the rainfall in the area since those are expected to increase due to climate change. After raindrop, the water is expected to flow towards the discharge complex IJmuiden via the regional pump stations, Amsterdam Rhine Channel and North Sea Channel. As mentioned in paragraph 6.1., the input into NPBNs are the marginal distributions and rank correlations to determine the (Gaussian) copula. First we will sum up all variables that must be added into the NPBN. Thereafter we define the derivation of the marginal distributions and rank correlations. A more detailed figure can be found in appendix VII.
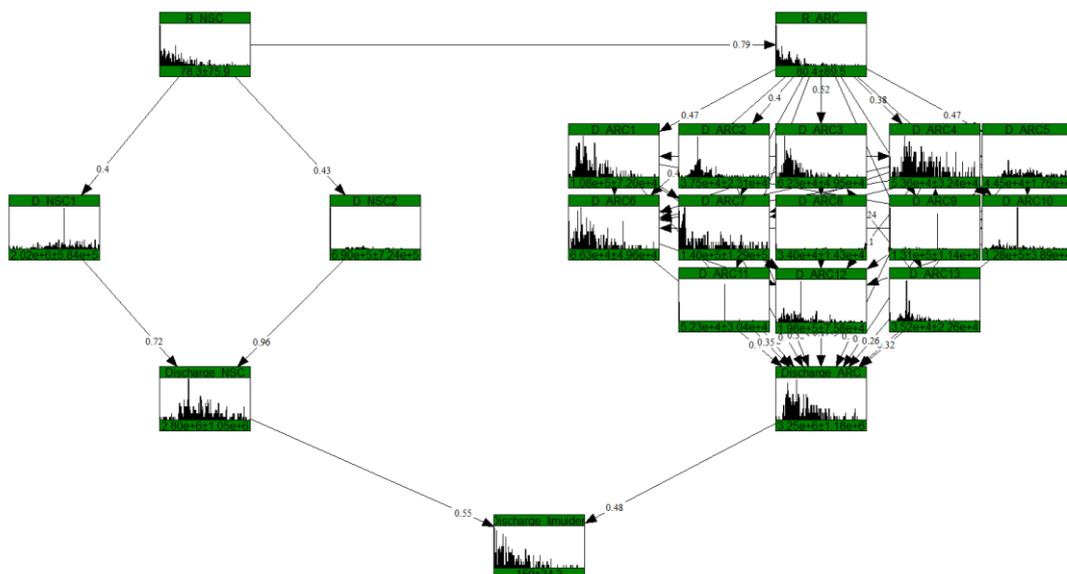
Figure 21: The solicitation model

Covariates

The water system can be considered to exist out of two regions: the NSC-region and the ARC-region, see figure 22. Since this is a large area, we chose two representative locations for the rainfall in the area: one representative for the NSC-region (R_NSC) and one for the ARC-region (R_ARC). The rainfall will be discharged into the NSC and ARC by the adjoining discharge stations, D_NSC_n and D_ARC_n respectively. Thereafter the cumulative excess water, Discharge_NSC and Discharge_ARC, that is discharged by all adjoining stations will flow towards the discharge station of IJmuiden where it will finally be discharged into the North Sea.

The ARC region consists of a large network of channels and rivers with many discharge stations with a relatively low discharge capacity. Instead of incorporating all adjoining discharge complexes, we incorporated 13 representative discharge stations that are for 53% responsible for the total discharge in the ARC between 2013 - 2016. For the names of the chosen discharge stations we refer to figure 22. The total discharges of the ARC (Discharge_ARC) represents the cumulative discharges of all adjoining discharge stations in the partial water system.

The North Sea Channel water system is a smaller system with less discharge stations, but with higher discharge capacities. The model links the rainfall in the region to two discharge stations that drain their water into the NSC. The cumulative discharge (Discharge_NSC) consists of all discharge stations south of the NSC. Unfortunately, the discharge stations north of the NSC are not discounted since the gathered data is not sufficient. Still, NPBNs enables us to retrieve a belief in the discharge at IJmuiden (D_IJmuiden). Incorporating the data north of the NSC would make the model more realistic since it can contradict or support our current belief in the discharge at IJmuiden, based on the southern discharges.

Deriving the Marginal Distributions

Now that the covariates are known, their marginal distribution must be defined. The water boards gathered the discharge data of their discharge stations accurately, which makes it relatively easy to determine their marginal distributions.

Preliminary assumption 2 enables it to only account for the peak discharges. Therefore, the first step is to set a threshold of which the discharges are equal or above this threshold can be considered as peak discharge. The only restriction related to the threshold is that enough peak discharges should be gathered to make a reliable estimation of the dependence structure between all variables in the model. A threshold of 120 m^3/s, gives us 48 peak-occurrences in 2013, 51 in 2014, 71 in 2015 and 45 peak-occurrences in 2016, which we consider as sufficient samples to use as marginal distribution.

The marginal distributions for all other nodes are based on the values that occurred at simultaneously with the considered peak discharge. The figure below illustrates how the samples are retrieved. After gathering all values associated with a certain peak-value, the marginal distributions can be generated for each node and plugged into the NPBN.

Amsterdam Rhine Channel Region
North Sea Channel Region
Rainfall Location
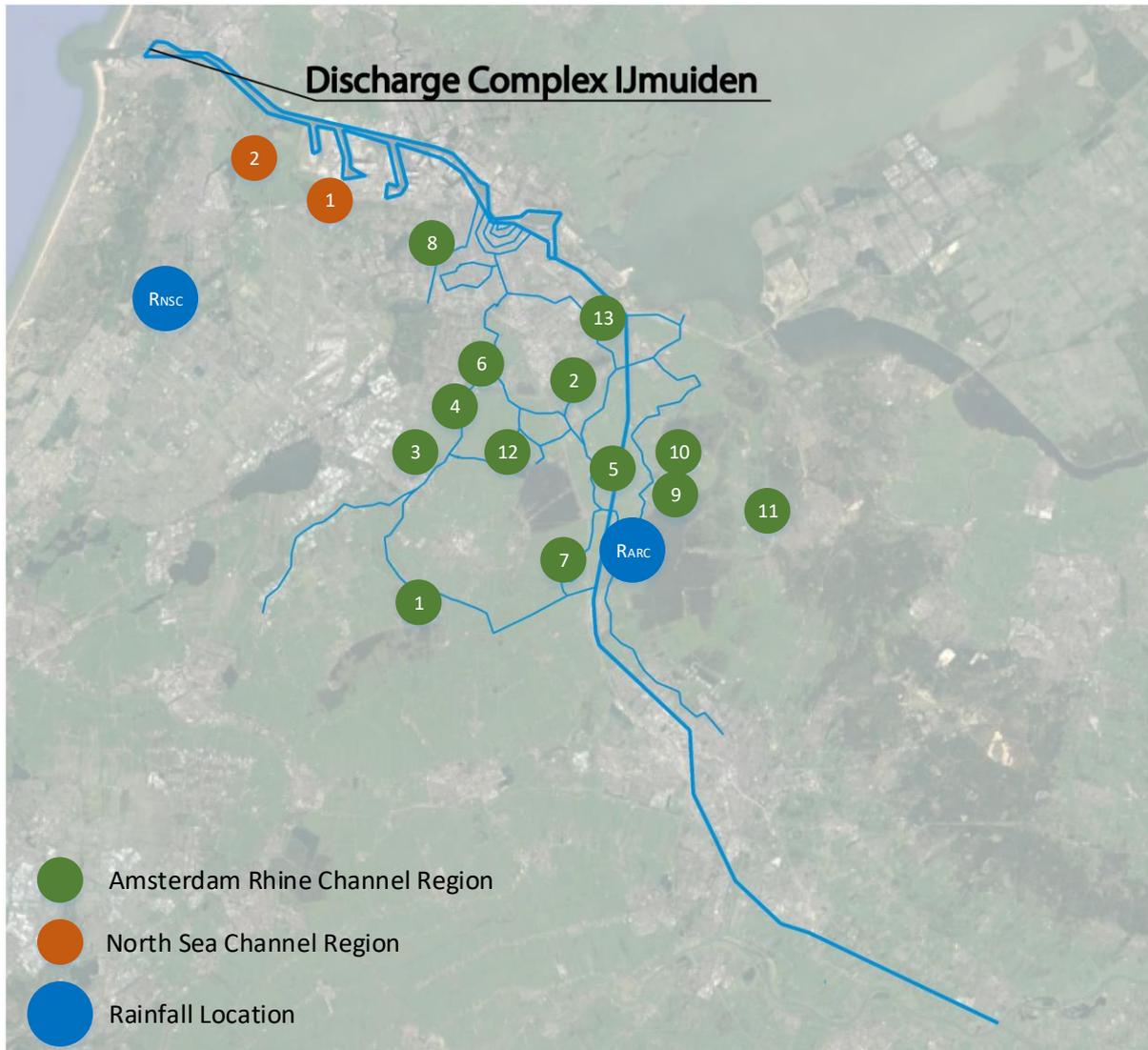
Amsterdam Rhine Channel Region		North Sea Channel Region		Rainfall Location	
1	Kromme Mijdrecht	1	Halfweg	R_{ARC}	Loenen aan de Vecht
2	Nieuwe Buillewijk	2	Spaarndam	R_{NSC}	Heemstede
3	Noorderlegmeer				
4	Ronde Hoep				
5	Baambrugge-Oostzijds				
6	Bovenkerkerpolder				
7	De Ruiter Demmerik				
8	Delflandlaan				
9	Gemaal Loosdrecht				
10	Horstermeer				
11	Kortenhoef				
12	Winkel				
13	Zandpad				

Figure 22: Representative discharge stations and rainfall locations

Date	D_IJmuiden	D_ARC1	…	R_ARC	
	[m³/s]	[m³/day]		[mm/day]	
1-1-2013	212	1.7288e+05	…	135	
2-1-2013	136	9.9637e+04	…	25	
28-1-2013	155	1.5939e+05	…	60	N₂₀₁₃
…	…	…	…	…	
3-1-2014	135	6.7902e+04		74	
…	…	…	…	…	N₂₀₁₄
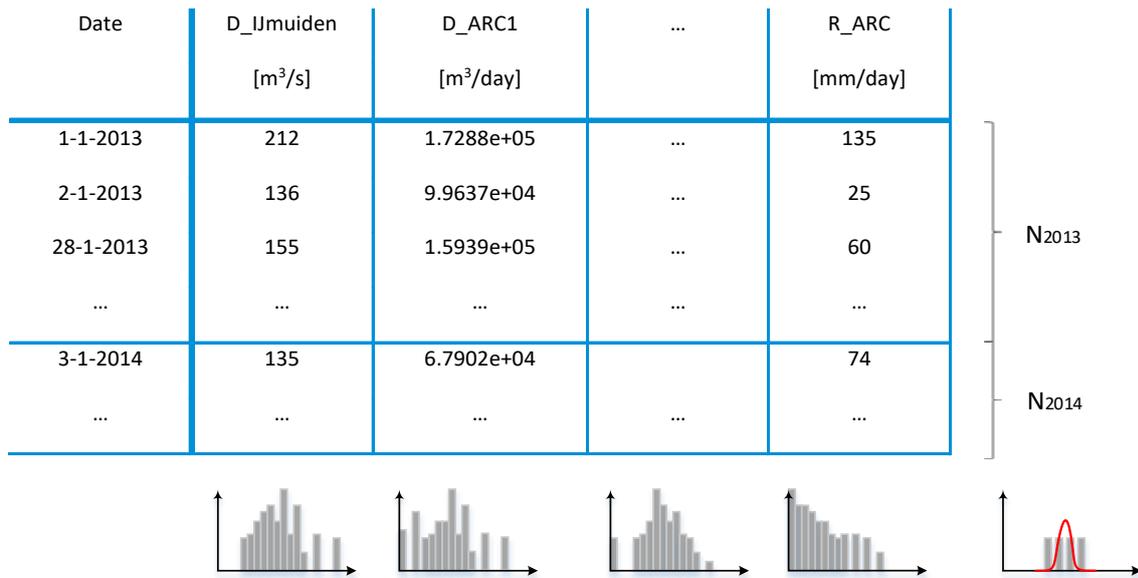
Figure 23: Process to derive the marginal distributions for each node

A very interesting parameter here is the frequency of peak discharge occurrences per year, denoted above as N_{year}. Based on the number of peak occurrences per year, we can fit a Poisson-distribution[2] to the data by estimating the only parameter λ, which represents the average peak occurrences per year. In this way we can probabilistically describe the frequency of peak occurrences under stationary circumstances. In anticipation to the creation of time-dependence, one effect of climate change, namely the expected increase in peak discharges per year, can be incorporated by tweaking this parameter λ.

Rank correlations covariates

So far, the datasets of the chosen covariates are modified on the basis of the peak-discharge-samples above 120 m³/s. The next step is to determine the dependency among all covariates, according Spearman's rank correlation, in UniNet named 'empirical rank correlation'. The correlation matrix is given in table 3. The rank correlations among the covariates are necessary to identify the covariates between which we have to draw an arc.

Since the availability of sufficient data, the rank correlations can be directly determined and is represented in Appendix VII. High rank correlations imply that there is a clear dependency between the two variables. This means that an arc must be drawn in order to represent the dependency via an copula. We drew arcs between all variables above a threshold of 0.62.

The direction of the arc adds the Bayesian reasoning component to the network. See paragraph 6.2 for the conditional independence structures. We also removed arcs that have correlations higher than the set threshold of 0.62 when correlation is assumed to be coincidence. By both adding and removing arcs, we ought to avoid sample jitter and generate reliable samples.

At the end, we are mainly interested whether the BN represents the dependency between rainfall and discharge. One can check this by comparing the empirical correlations and BN-correlations of those

---

[2] The Poisson distribution, $Poiss(\lambda)$, is a discrete distribution with probability mass function $P(X = k) = \frac{\lambda^k}{k!}e^{-\lambda}$, where $\lambda > 0$, $E[X] = \lambda$ and $Var(X) = \lambda$.

variables. According to the correlation matrix in appendix VII, the difference between the two correlations is approximately 0.1 which we consider small enough. One can conclude that the correlation is sufficiently represented in the model.

Model output

The resistance model output is the sampled distribution of the only solicitation variable: the peak discharge IJmuiden ≥ 120 m³/s . Thereby, we determined the Poisson-distribution that describes the number of peak discharges in a year. Both distribution will be used as an input into the simulations, described in paragraph 10.8.

10.3.3. Resistance model

The aim of this paragraph is to set-up a model to stress a belief into availability and pump-capacity via the NPBN illustrated below. First we will sum up all variables that must be added into the NPBN. Thereafter we define the derivation of the marginal distributions and rank correlations.
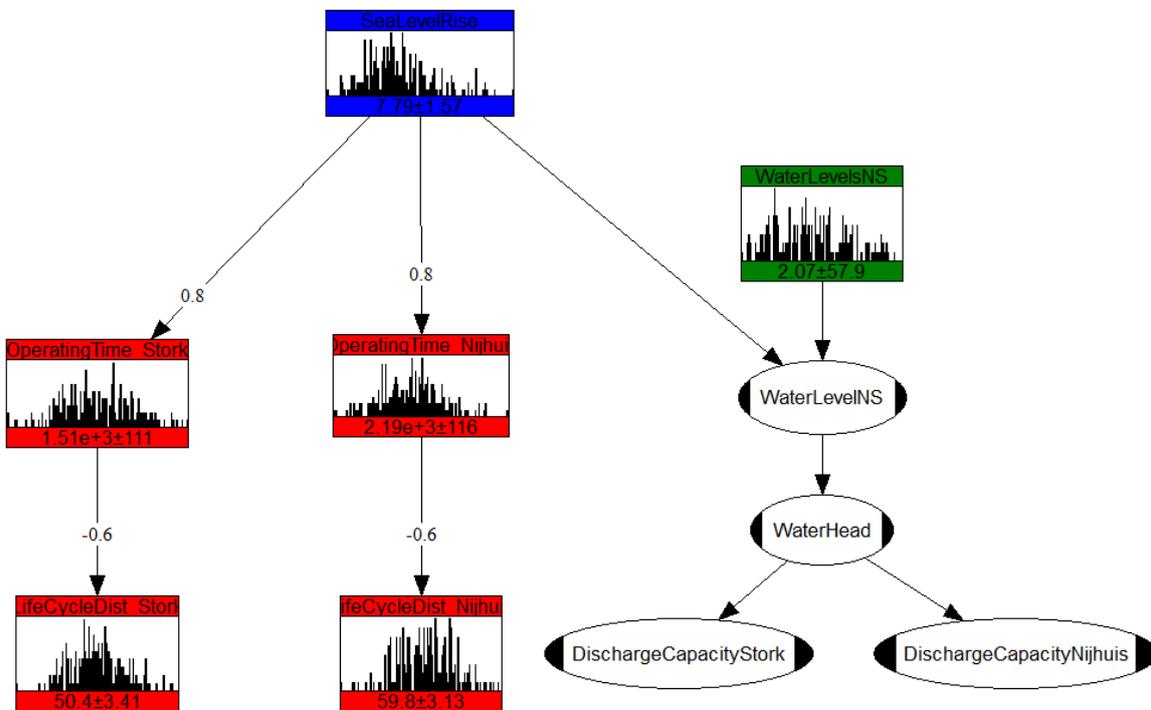
Figure 24: Resistance model

Covariates

Chapter 8 described the way sea level rise will impact the time-to-failure distribution and the pump capacity. To recap, the system will wear faster since the duration of the pumping time slots will increase (figure 17) and the Q-H curve (figure 18) shows that the maximum capacity decreases when the water head increases. For this reason we need to find covariates that link sea level rise to the time-to-failure distribution and water head.

First, we addressed the problem that the pumping slots increase. Since the pumping time slots vary in time on a yearly basis due to yearly variations in tidal fluctuations, storm surges etc., we prefer to think in terms of yearly operational hours. By deriving the marginal distributions of the time-to-failure distribution and operating time of each pump, we are able to model the dependency between sea level rise and component-time-to-failure distribution.

Secondly, we noted that the probability that the pumps not able to discharge with maximum capacity will decrease, when the average sea level rises. The Q-H curve shows a deterministic relationship with pump capacity and water head. For this reason, the water head is the node of interest. Since we assumed the water level in the NSC to be constant, the water head can be simply defined by subtracting the NS-level by the constant NSC-level (0.40m). This implies that the water level of the NS is the second node of interest. The water levels in the NS are determined by tidal fluctuations, storm surges and waves. Since the latter two contain uncertainties, we prefer to add a probabilistic node that contains the water levels at the NS in the considered period (2012-2016). The sum of sea level rise and current water level –data must represent the water levels in the NS over time.

Deriving the Marginal Distributions

In contrast to the solicitation model, there is limited data available to determine the marginal distributions of the nodes mentioned. This also counts for the parental node sea level rise. For this reason, we must switch to the other method to retrieve the distributions; expert judgement. We acknowledge that the next estimations of the marginal distributions and rank distributions contain an uncertain factor that will affect the results. Still, it is sufficient for the purpose of this research.

Bamber & Aspinall (2013) performed an expert judgement assessment of future sea level rise form the western- and eastern Antarctica-ice sheets and the Greenland ice-sheet. By estimating the contributions per sheet to sea level rise for 2100, the combination results in total sea level rise at 2100. One very important fact is that the assessed SLR appeared to be non-Gaussian with a long upper tail. In order to estimate the marginal distributions for SLR we therefore assume the SLR to be lognormal distributed. Then, assuming linear increase of the total SLR-percentiles – which are given by W. Aspinall -, we are able to fit a log-normal distribution for each year. The mathematical procedure is given in 7.5.

The next node is operating time. Estimations in van der Wiel et al. (2013) showed that the operational hours of the Stork- and Nijhuis-pumps vary. Current approximations state that the Stork pumps are approximately 1500 hrs/yr in operation and the Nijhuis pumps approximately 2200 hrs/yr. Unfortunately, data is not available to determine the marginal distributions and therefore we performed our own 'expert judgement' by estimating the 5th and 95th percentiles for 2017. As paragraph 1.5. introduced, the operation time is predominantly dependent on the tidal fluctuations. Thereby, it is also dependent on storm surges. Here we assume that the average sea level will be temporal lift up for approximately 8 days in total. Therefore we assumed the 5th and 95th percentiles to be ± 200 hours of the 50th percentiles.

Table 1: Estimations 5th, 50th and 95th percentiles of the operational time in 2017.

	Nijhuis			Stork		
	5th	50th	95th	5th	50th	95th
Operational Time 2017 (hrs)	2000	2200	2400	1300	1500	1700

The third node is the time-to-failure distribution. The failure data only shows the mean time to failure (MTTF), which we interpreted here as the 50th percentile of the distribution. Then we only have to estimate the 5th and 95th percentiles of the current distribution. Here we asked ourselves "*Given that the nth percentile of the Operational Time in 2017 is x hours, what will the nth percentile of the time-to-failure distribution be?*".

Table 2: Estimations 5th, 50th and 95th percentiles of the operational time in 2100.

	Nijhuis	Stork
Component	*'Hoofdasafdichting'*	*'Hoofdas'*

	5th	50th	95th	5th	50th	95th
Time-to-failure distribution 2017 (years)	55	60	70	45	50	57

On the pump capacity side, the nodes are mainly derived via a function, except the *WaterLevelsNS*-node. This node is simply derived from Rijkswaterstaat data and need no further explanation. Since the functions are an alternative for a copula, we will treat the functions hereafter.

Rank correlations covariates

Due to the lack of data, we estimated the marginal distributions of SLR, operational time and lifetime distributions. Now, we also need to estimate the copula in order to represent the dependency. First the SLR-Operation Time copula. Essentially, this is a monotone function which indicates full dependency. However, since the pumps operate randomly when discharge is required, there is a small uncertainty in it. For this reason we estimated the rank correlation to be high; 0.8. The Operation Time – time-to-failure distribution copula is more uncertain and negative; high operating time goes accompanied with a low time-to-failure distribution. For this reason we estimated the rank correlation on -0.6.

Now, we focus on the functional arc between the water head and the pump capacities for both the Stork- and Nijhuis pumps. The Q-H curve (figure 18) shows a deterministic relationship between water head and discharge capacity. This deterministic relationship is represented by fitting a curve on the known points in the figure. Those functions, both for the Stork- and Nijhuis-pump are represented in the NPBN do derive the discharge capacity-distributions.

Model output

The resistance model output are the sampled distributions of the four variables: the time-to-failure distribution(s) of the Stork component(s), the time-to-failure distribution(s) of the Nijhuis component(s), the discharge capacity of the Stork pumps and the discharge capacity of the Nijhuis pumps. The time-to-failure distributions are used as input into the deterioration model, described in chapter 7. The sampled distribution of the discharge capacities will directly be used as an input into the simulations, described in paragraph 7.8.

10.4. MODEL VALIDATION

The third step of the road map illustrated in paragraph 8.3, is the model validation. Since we used data we are interested whether the dependence structure is sufficiently represented in the NPBN. First we have to justify the utilization of the Gaussian copula. For this we use the validation-methods described in Appendix IV. The NPBN validation is described in paragraph 6.4.2. The Matlab-scribs can be found on https://github.com/RHuijmans.

10.4.1. Copula Justification

The justification of the Gaussian copula is based on the study of semi-correlations and a 'Blanket Test', extensively discussed in Appendix IV. The results are shown in Appendix VII. The results can be interpreted as follows, relatively small differences in semi-correlations and low values of the 'Blanket Test' statistic indicate a Gaussian copula, relatively high upper-tail correlations and low values of the 'Blanket test' statistic indicate a Gumbel copula, and relatively high lower-tail correlations and low values of the 'Blanket Test' statistic indicate a Clayton copula.

Analysis of the results indicate that Gaussian copula – utilized in UniNet – is only a good representation for a few bivariate pairs of variable. For 42 out of the 61 bivariate pairs, the Gumbel copula shows to

be the best fit. Only for 18 out of the 61 pairs, the Gaussian copula shows the best fit and for 1 out of the 61 pairs, the Clayton copula shows to be the best choice.

The dominance of the Gumbel copula can be a consequence of the preliminary assumption to only incorporate the peak discharges. Despite this dominance, we maintain the assumption of the Gaussian copula due to the computational advantages via UniNet. Preserving this assumption would not have any very dramatic consequences for the sampling procedure; since the Gumbel copula illustrates upper-tail dependency, the lower-tail dependency is smaller than the Gaussian copula. The effect on the higher discharges is limited, which is the main spectrum of interest.

10.4.2. NPBN Validation

The first test measures the suitability of the Gaussian copula to represent the empirical data dependence structure. The determinant of the empirical normal rank correlation matrix (DNR) will differ from the determinant of the empirical rank correlation matrix (DER) since the latter is based on the empirical copula. If the DER is within the 90% confidence bound of the DNR, the joint Gaussian copula is a reasonable assumption. In case of our BN, the DER remained within the 90% bound of the DNR if no more than ca. 350 samples were drawn.

The second test concerned the comparison of the determinant of the rank correlation matrix of a BN constructed under the assumption of the Gaussian copula (DBN) and the DNR. In case of our BN, the DER remained within the 90% bound of the DNR if no more than ca. 200 samples were drawn. Both sample-values are relatively small, indicating that the joint normal copula may not be the best assumption to utilize in the model. Still we stick to the utilization of the Gaussian copula.

10.5. ADD LONG-TERM TREND

The fourth step concerns the addition of the long term trend. As illustrated in paragraph 8.2, the most attractive way to do this, is to conditionalize the NPBN on the forecasted values. One requirement is necessary to perform this; the forecasted values have been occurred in the past. In this research we assume that this requirement is met for precipitation; i.e. we say that extreme rainfall has occurred simultaneously with an particular peak discharge and is captured in the historic data set. Unlike future precipitation, the future sea level is not captured in past data. This requires a different, more complex approach to model the long-term trend.

10.5.1. Precipitation

Klein Tank et al. (2015) state in 'KNMI'14-klimaatscenario's' that the probability of extreme precipitation events increase under influence of climate change. Here, we assume that an extreme precipitation event has occurred in the past and is captured in the dataset. This enables us to add the long term trend via inference.

In our case, we perform inference by conditioning the rainfall on the upper tail-percentiles, for example the 5th – 100th percentile. Although we assume then that a peak discharge will always occur together with rainfall, it puts higher probabilities in the other rainfall amounts, including the extremes. This is our best estimation to represent the future rainfall distribution. The effect on the discharge distribution is illustrated in the figure below. Repeating inference for every time step enables it to represent the effect of climate change on the discharges in IJmuiden for the longer term.
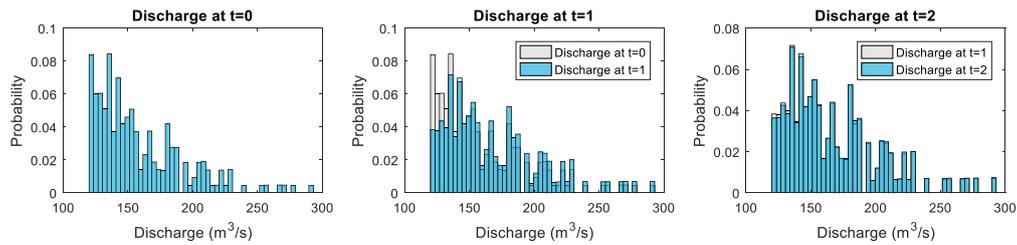
Figure 25: Effect inference on discharges IJmuiden

10.5.2. Sea Level rise

In contrast to the precipitation, we are not able to perform inference for the sea level rise case. The main reason is that NS-water level data does not capture any values that can occur when the sea level rises with one meter, which is realistic according (Bamber & Aspinall, 2013). For this reason, we must propose a more extended way to model the long-term trend of sea level rise and its impact on the child-nodes.

The method we propose here can be compared to the Dynamic Bayesian Network described in paragraph 5.2.; for every time-slice we create a BN as illustrated in figure 20. The parental SLR node's marginal distribution is described by the SLR-distribution at that particular time. This also implies that the (conditional) distributions of the operating times and time-to-failure distributions must be conditionalized on the sea level rise. In other words, the marginal distributions of the probabilistic child-nodes also vary for every time slot. This procedure implies that the we need to define the SLR-distribution, operating time-distributions and lifecycle distributions for every time slot. One way to derive this is explained hereafter.

In case of the SLR-distribution, the point of departure is the paper of Bamber & Aspinall (2013) that performed expert judgement to estimate the SLR-contributions of three ice-sheets in 2100. Based on the retrieved 5th, 50th and 95th percentiles they fitted a lognormal distribution. By sampling those distributions and sum the samples they retrieved the total SLR-rate in 2100. The sampling procedure included the fact that the ice-sheets may be correlated. In their article, they assumed the correlations; EAIS – WAIS = -0.2, EAIS – GrIS = -0.2 and WAIS – GRIS = +0.7. For the purpose of this research, author W. Aspinall provided us with perfect positively correlated samples. Perfect positive correlation implies that all possible SLR-rates are covered, minimum and maximum. Consequentially, the uncertainty bandwidth of possible SLR-rates increases.

As done in the B&A-paper, we converted the rates of SLR into cumulative values by assuming a linear increase from the experts' estimate for the past decade (0.9 mm/year) to the 5th, 10th, 25th, 50th, 75th, 90th and 95th- percentiles of the 2100-total SLR-rate distribution. Then, integrating the range values from 2010 to the year of interest gives us the SLR for that particular time. For the record, we used the findings of the 2012 elicit. Figure 26 illustrates the SLR over time, based on the given values in Bamber & Aspinall (2013).
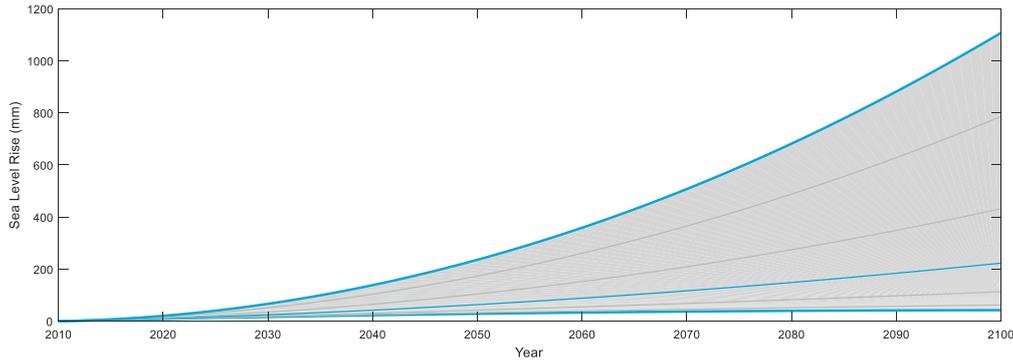
Figure 26: Sea level rise over time, based on the values of Bamber & Aspinall (2013)

Then, we are able to determine the 'emperical' 5th, 10th, 25th, 50th, 75th, 90th and 95th percentiles for every year. The calculated values are given in Appendix VI. Since it is already mentioned that the SLR is non-Gaussian, we fitted a lognormal distribution to the calculated percentiles via the underlying formula:

$$\min S(\mu, \sigma) = \sum_{i=n} \left(F_{lognormal}(x_n, \mu, \sigma) - F_{emperical}(x_n)^2 \right) \tag{14}$$

Where $n = \{5, 10, 25, 50, 75, 90, 95\}$. This formula calculates the lognormal distribution's parameters (μ, σ) that minimizes the differences between the fitted lognormal distribution and empirical distributions. Here we used the cumulative distribution, but the density function can also be used. The results for years 2017 and 2100 are given in the table below. All results for the years in between are given in Appendix VI.

Table 3: Estimations 5th, 10th, 25th, 50th, 75th, 90th and 95th percentiles of the sea level rise in 2017 and 2100

	5th	10th	25th	50th	75th	90th	95th
Sea Level Rise 2017 (mm)	6.10	6.20	6.47	7.04	8.14	10.00	11.68
Sea Level Rise 2100 (mm)	42.17	62.21	113.22	222.99	431.92	785.93	1107.50
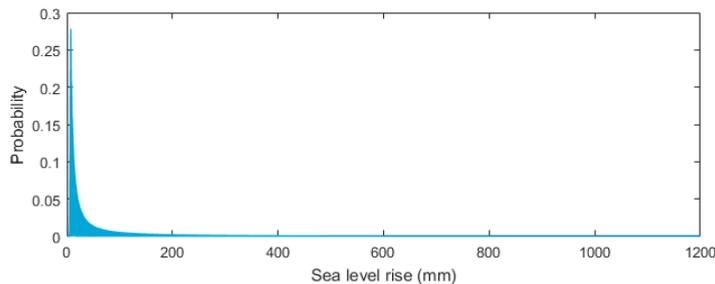
Figure 27: Development distribution sea level rise from 2017 to 2100.

This trick is also used for the child nodes. First the Operating time. In order to retrieve the percentiles, we asked ourselves the question "*Given that the nth percentile of the Sea Level Rise in 2100 is x mm, what will the nth percentile of the time-to-failure distribution be?". S*ee the underlying table for the outcome. Then as done for the SLR we assumed a linear increase over time to withdraw the distributions for each year, the results are given in Appendix VI.

Table 4: Estimations 5th, 50th and 95th percentiles of the operational times in 2017 and 2100.

	Nijhuis			Stork		
	5th	50th	95th	5th	50th	95th
Operational Time 2017 (hrs)	2200	2400	2600	1500	1700	1900
Operational Time 2100 (hrs)	2300	3000	4700	1900	2100	3400
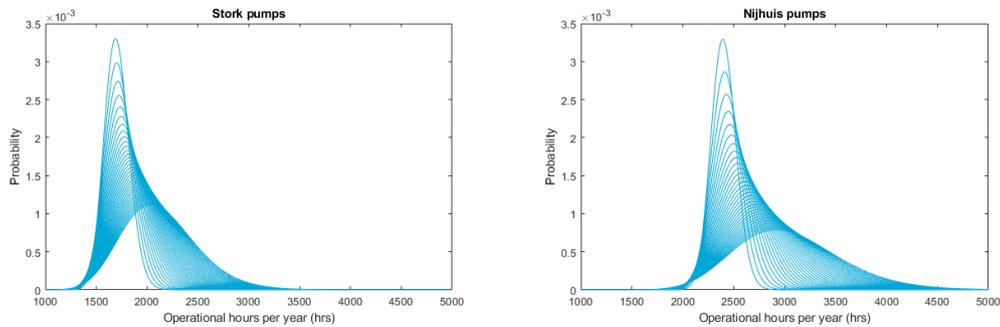
Figure 28: Development Opderational hours distribution from 2017 to 2050.

Secondly, the time-to-failure distributions are purely based on estimation. In order to retrieve the percentiles, we asked ourselves the question "*Given that the nth percentile of the Operational Time in 2100 is x hours, what will the nth percentile of the time-to-failure distribution be?".* The results are shown in the underlying table. The stochastic based-processes of appendix II both show that the time-to-failure distribution has a Weibull distribution-shape. We therefore fit a Weibull distribution to the calculated percentiles. The results are given in Appendix VI.

Table 5: Estimations 5th, 50th and 95th percentiles of the time-to-failure distributions in 2017 and 2100.

	Nijhuis			Stork		
Component	*'Hoofdasafdichting'*			*'Hoofdas'*		
	5th	50th	95th	5th	50th	95th
Time-to-failure distribution 2017 (years)	56	58	60	46	48	50
Time-to-failure distribution 2100 (years)	48	55	59	38	44	48
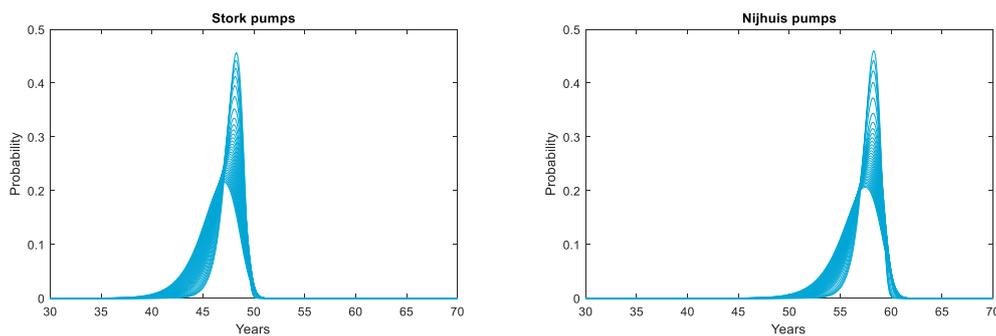
Figure 29: Development time-to-failure distribution from 2017 to 2050.

By plugging the determined distribution at time t into the BN of time-slice t, we are able to include sea level rise. In comparison to the previously mentioned DBN, this model contains no dependencies between the time-slices. In other words, drawing a sea level rise-sample at time t is independent on the sea level rise-sample draw at time $t-1$. In reality their actually is dependency within this process. This dependency is not captured in this model.

10.6. DETERIORATION MODEL

NPBNs enable us to make to model the dependencies between a continuous-state deterioration and continuous covariates. In this research, we create dependency the Markov chain input – the time-to-failure distribution and the covariates. For the case purposes, we need to take one component per pump that is critical, and deteriorates due to an increase operational time. Based on the (partial) system decomposition in Appendix V, we choose the component *'Hoofdas'* for the Stork-pump and *'Hoofdasafdichting'* for the Nijhuis pump.

In this research, an oversimplified two-state Markov Chain is utilized due to the lack of sufficient data to define multiple states as described in chapter 7. For an visualization of the used Markov chain, see figure 30.
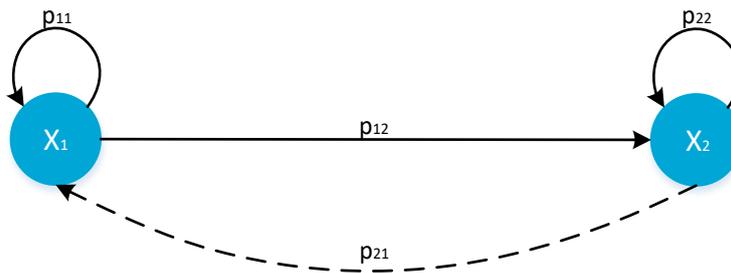
Figure 30: Availability Markov chain

The accompanied transition matrix is defined in equation (15). The MTTR of both components can be expressed in hours. Since this research models on a peak discharge occurrence time-scale, we assume that the component is brought back to its 'as new' state before the following peak-discharge occurs. Then, p_{21} can be considered to be 1.

$$P(t) = \begin{pmatrix} 1 - \dfrac{1}{MTTF(t)} & \dfrac{1}{MTTF(t)} \\ 1 & 0 \end{pmatrix} \tag{15}$$

The utilization of this oversimplified Markov chain contains one main limitations. The use of two states limits the deterioration model to integrate the physics of failure as wear out. For example, once the component is already corroded for 10%, the transition time to 20% is less than the time it took the component to arrive at 10% departing from no corrosion. Now, the two-state Markov chain essentially assumes a constant failure rate: the chance that failure occurs at *t = 0* is equal to the chance that the component is operational for 30 years.

10.7. EFFECT ON THE ASSET'S RESISTANCE

Step six of the roadmap concerns the translation of the deterioration in terms of assets resistance. Based on the failure definition, we concluded that pump function unavailability is the point of interest. In the previous paragraph 10.6. we choose the two components to model the deterioration; one for the Stork-pump and one for the Nijhuis-pump. However, the pump's unavailaibility is not determined by those components only. The pumps unavailability is determined by a whole process from one side to the other; the pump process. All components in the pump process have a specified time-to-failure distribution, MTTR and unavailability. In order to estimate the combined effect of each component to the availability of one pump, we made a NPBN-shaped fault tree where the nodes are the (discrete) probabilities of component availability and unavailability, see figure 31. By sampling the fault tree NPBNs we are able to calculate the total unavailability of the Stork- and Nijhuis pumps. Since the

previous chapter states that the probability of component unavailability changes in time, we have to adapt the fault tree NPBNs for each time slice.
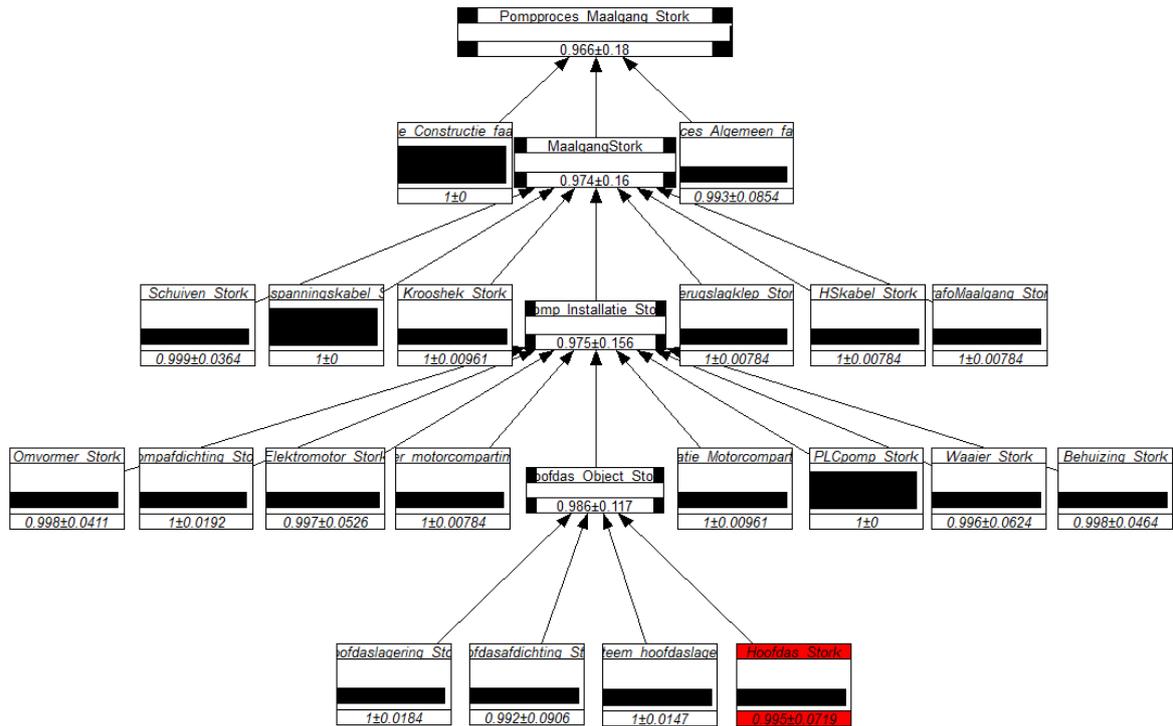
Figure 31: Fault tree NPBN of the Stork pump process after sampling
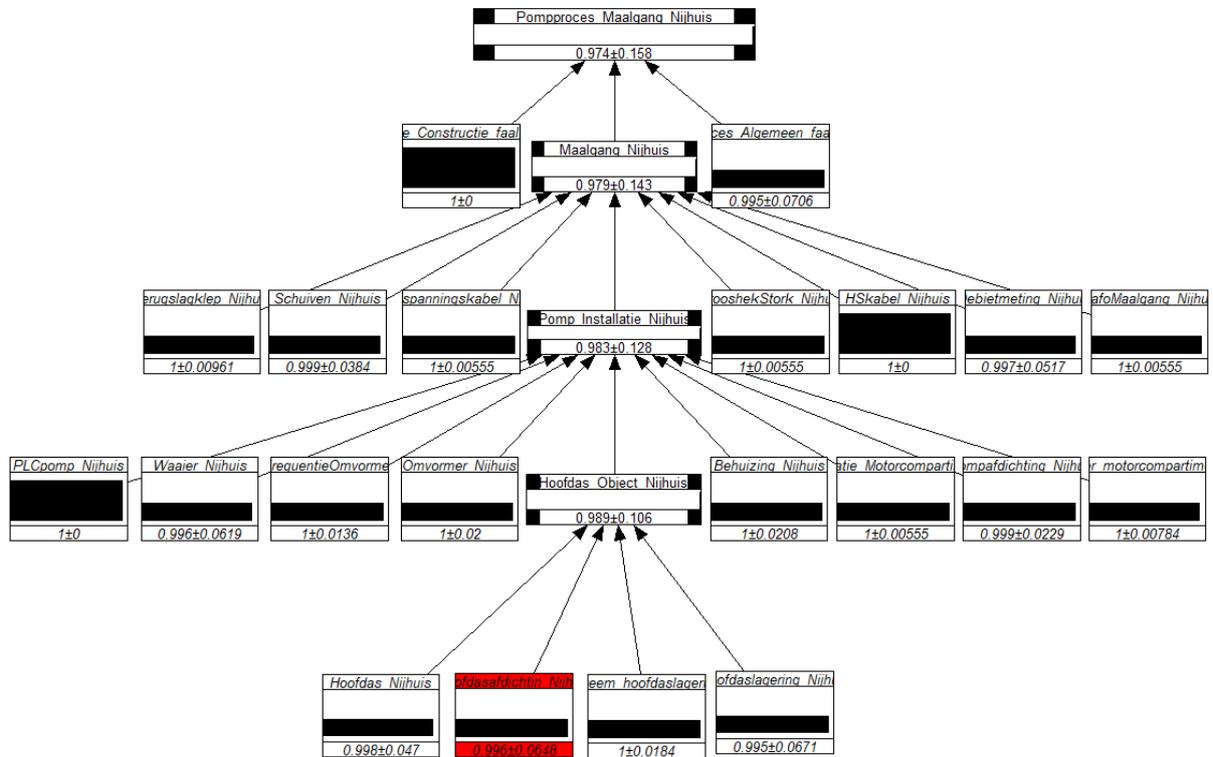
Figure 32: Fault tree NPBN of the Nijhuis pump process after sampling

Model output

The model output are the sampled distributions of the Stork- and Nijhuis pumps unavailability, denoted in the upper nodes. The sampled distribution will be used as an input into the simulations, described in paragraph 10.8.

10.8. SIMULATIONS

The final step of the roadmap are the (Monte Carlo-) simulations of the retrieved sample distributions. Here, we elaborate on the sampling procedure that is programmed in Matlab®. First, we recap the distributions and their sample space, they are given in the table below.

Table 6: Overview simulation variables

λ	Peak discharge occurences per year	Discrete
D	Discharge at IJmuiden	Continuous
A_{Stork}	(Un)availability Stork pump	Discrete
$A_{Nijhuis}$	(Un)availability Nijhuis pump	Discrete
C_{Stork}	Capacity Stork Pump	Continuous
$C_{Nijhuis}$	Capacity Nijhuis Pump	Contninous

The sampling procedure is as follows, see figure 33. For every year y, we perform n number of simulations. For every simulation n we sample a the number of peak discharge occurences $\lambda_{y;n}$ for the particular sample in that particular year. Thereafter we sample λ-times from the remaining five distributions which represent all variables in the limit state function, equation (3). In other words, we are able to calculate the probability of failure by dividing the number of failures by the number of simulations for that particular year $N = (n_y * \sum_{i=1}^{n} \lambda_{y;i})$. Repeating this procedure for every time slice, we are able to calculate the probability of failure in time. The Matlab-scribs can be found on https://github.com/RHuijmans.
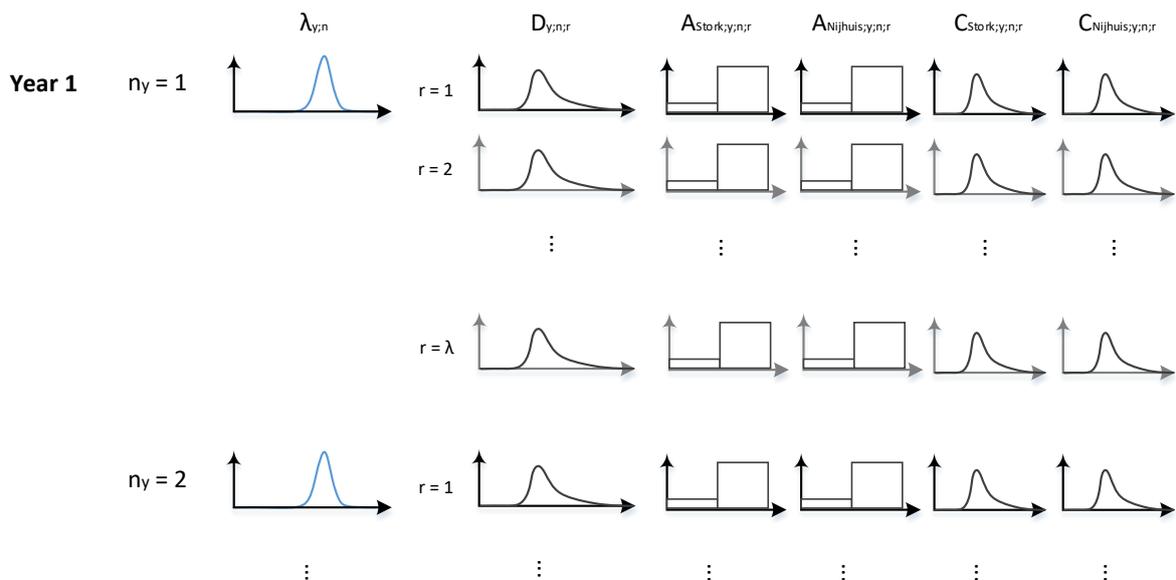
Figure 33: Sampling Procedure for one year

11. CASE STUDY RESULTS

This chapter includes the results of the built model of chapter 10. The aim of this chapter is to verify that the model does what we expect it to do. The output results of the NPBN and deterioration model will be given in paragraph 11.2 and 11.3. Thereafter we introduce the hypothetical case that the current predictions of sea level rise are underestimated and happen twice as fast. The distributions of The Matlab-scribs to calculate the results can be found on https://github.com/RHuijmans.

11.1. SCENARIO'S

As introduced in chapter 3, to show the effect of making the environment dependent on the deterioration in order to model the fortified deterioration, we divide the model into two sub-models; the environment-independent deterioration model and the environment-independent deterioration model.

Again, the *environment-independent deterioration model* calculates the probability of failure of discharge station IJmuiden assuming an increasing environment but a constant time-to-failure distribution and MTTF. The *environment-dependent deterioration model* calculates the probability of failure assuming dependency between the increasing environment and decreasing time-to-failure distribution. As mentioned in paragraph 8.2, we are able to add the long term trend on rainfall by conditionalizing on intervals. The underlying tables indicates on which values we conditionalize the covariates to represent the three scenarios. Here we would like to emphasize that all values are just estimations.

Besides the independent- and dependent-deterioration models two extreme scenarios are developed in indicate the minimum and maximum impact of the long term trend. that different scenarios have been developed by conditionalizing on different percentiles representing extreme- and medium cases.

Table 7: Conditional inference intervals to illustrate the minimum impact of the long term trend

Years	2017 - 2019	2020 - 2029	2030 - 2039	2040 - 2050
Rainfall occurences	53.75	55	57	59
Rainfall	$0^{th} - 100^{th}$	$7^{th} - 100^{th}$	$10^{th} - 100^{th}$	$12^{th} - 100^{th}$
SeaLevelRise	$0^{th} - 5^{th}$			

Table 8: Conditional inference intervals to illustrate the average impact of the long term trend

Years	2017 - 2019	2020 - 2029	2030 - 2039	2040 - 2050
Rainfall occurences	53.75	57	62	67
Rainfall	$0^{th} - 100^{th}$	$9^{th} - 100^{th}$	$11^{th} - 100^{th}$	$13^{th} - 100^{th}$
SeaLevelRise	$5^{th} - 95^{th}$			

Table 9: Conditional inference intervals to illustrate the maximum impact of long term trend

Years	2017 - 2019	2020 - 2029	2030 - 2039	2040 - 2050
Rainfall occurences	53.75	59	67	72
Rainfall	$0^{th} - 100^{th}$	$10^{th} - 100^{th}$	$12^{th} - 100^{th}$	$14^{th} - 100^{th}$
SeaLevelRise	$95^{th} - 100^{th}$			

11.2. EFFECTS OF CLIMATE CHANGE TO THE PUMP STATION

This chapter will reflect the results to the three effects to pump stations posed by climate change as introduced in paragraph 1.5:

1. **Increased pump deterioration.** Sea level rise fortifies the deterioration rate and increases the probability that an individual pump is unavailable.
2. **Reduction of pump capacity.** Sea level rise reduces the maximum pump capacity that can be retrieved from the Q-H curve.
3. **Increasing peak-discharges.** Increase of extreme rainfall evens increase the number of peak-discharges per year

Although numerous results can be shown, the results elaborated here will be limited to the medium case solely. One of the main reasons is that the individual differences among the scenarios are hard to note on the eye; incorporating them would be nothing-saying. Only the distributions of the 2017-, 2030- and 2050- time-slices will be shown here. We will come back on the scenarios when we are defining the final probability of asset failure.

11.2.1. Increased Pump Deterioration

As elaborated in previous chapters, sea level rise indirectly affects the time-to-failure of the component. The figure below shows the impact of sea level rise on the two considered deteriorating components; the Hoofdas of the Stork-pump and the 'hoofdasafdichting' of the Nijhuis pump. Here the shift to the lower lifetimes can be noticed. This would impact the input into the deterioration-model since the MTTF decreases. To get an understanding of the magnitude, the MTTF of the Nijhuis-pump Hoofdafdichting in 2017 is determined on 57.84 years, the MTTF in 2030 and 2050 are calculated on 56.94 and 56.54 years, respectively.
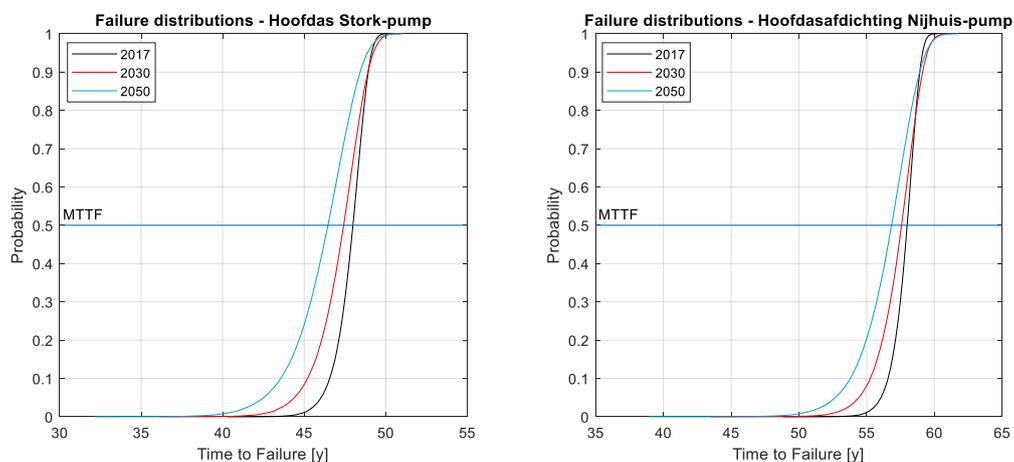
Figure 34: CDFs of the Time-to-Failure distributions over time

Now the MTTF's for each year can be are retrieved from the conditional time-to-failure distributions - dependent on operation time and future sea level rise - the model is able to calculate the availability deterioration of both components. The figure below shows the unavailability over time of both components for the deterioration-independency model as well as for the deterioration-dependency. The deterioration-independency model shows a stable unavailability rate, which is as expected since Markov chains always search for a stable state. In contrast, the proposed dependent Markov chain, part of the deterioration-dependency model, clearly shows a gradual increase of the components' probability of unavailability. For clarification, the first drop from 2017 to 2018 is a consequence of the Markov chain finding its stable value.
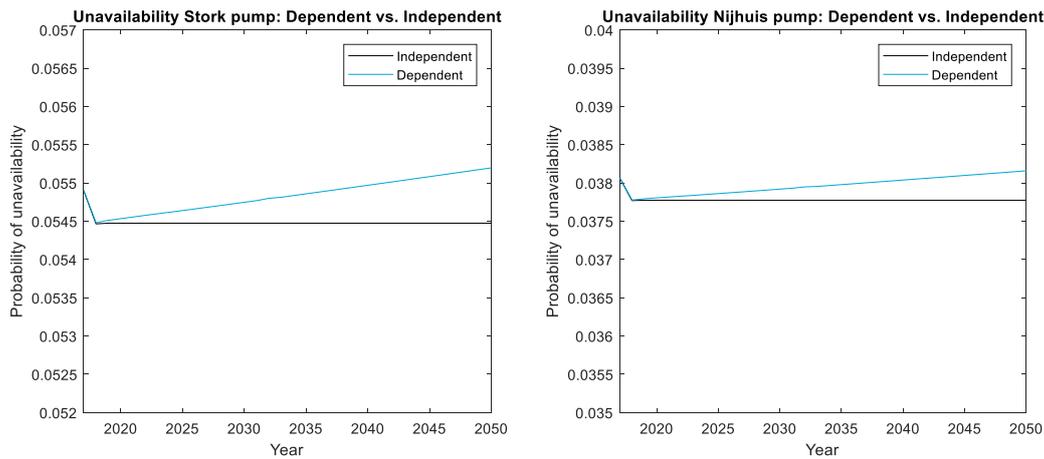
Figure 35: Availability deterioration of both pump's components

11.2.2. Reduction of Pump Capacity

The second effect of sea level rise concerns the reduction of pump capacity due to the increasing water head. The figure below shows the occurring pump capacities per pump type. Those can be compared with the Q-H curve given by figure 36. Note that - when time matures – the discharge capacites shift to the lower discharges.
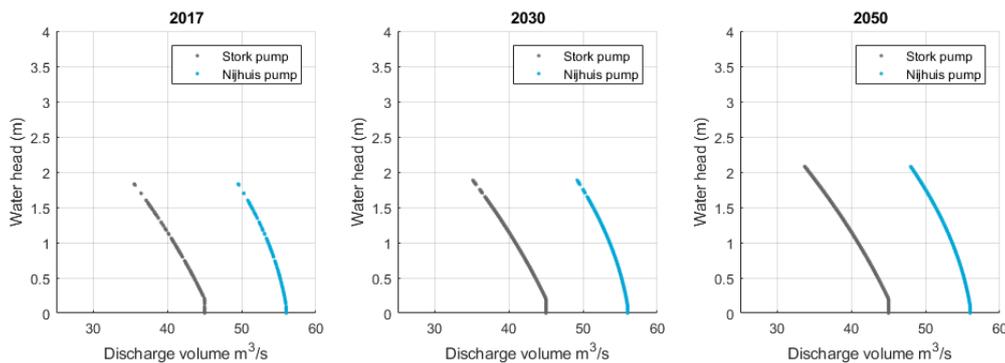
Figure 36: Occurring discharge volumes pump per year

The combined effect of figure 35 and 36 - deterioration of the pump availability and the shift to the lower discharge capacities, respectively – shows the effect of sea level rise on the total pump capacity, see equation (3). The cumulative distribution of the total discharge capacity is shown in figure 37. The distribution is retrieved after the simulations. To recap, the long term trend is here represented by deriving a specific marginal distribution for each time step. According to Bamber & Aspinall (2013), the sea level has risen 40 - 200 mm. The cumulative distribution can be interpreted as the probability that the discharges capacity is lower than the given discharge on the x-axis: $P(Q_{capacity;total} < q)$.

The figure shows a small decrease of the total pump discharges when time matures: the probability that the total pump capacity is smaller than q is higher in 2050 than in 2017 and 2030. Although the model does wat is expected, the impact of sea level rise to the total pump capacity is relatively small. Also the differences between the environment-independent deterioration model and the environment-dependent deterioration model apprear to be very small.
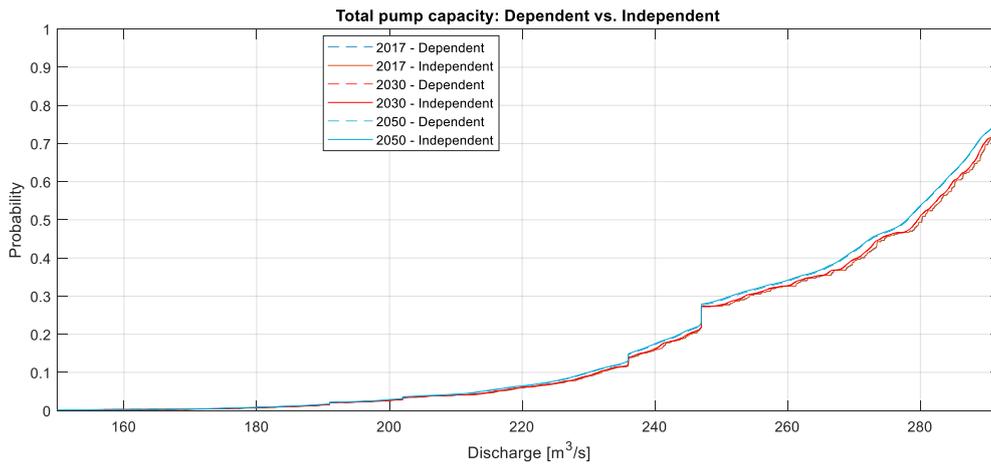
Figure 37: CDF of the total Discharge Capacity of Pump Station IJmuiden

11.2.3. Increasing peak-discharges

By sampling the NPBN, without any conditionalization, we are able to withdraw the distributions of 2017, the starting point of our belief in future discharges. The solicitation part of the limit state function is represented by the discharges of IJmuiden. The figure below shows the effect on the discharges of Ijmuiden when NSC- and ARC- rainfall is 'observed'. One can note that the lower discharge probabilities slowly shift towards the higher-end values. Sampling the 2050 distribution will result in more extreme discharges which require full pump capacity.
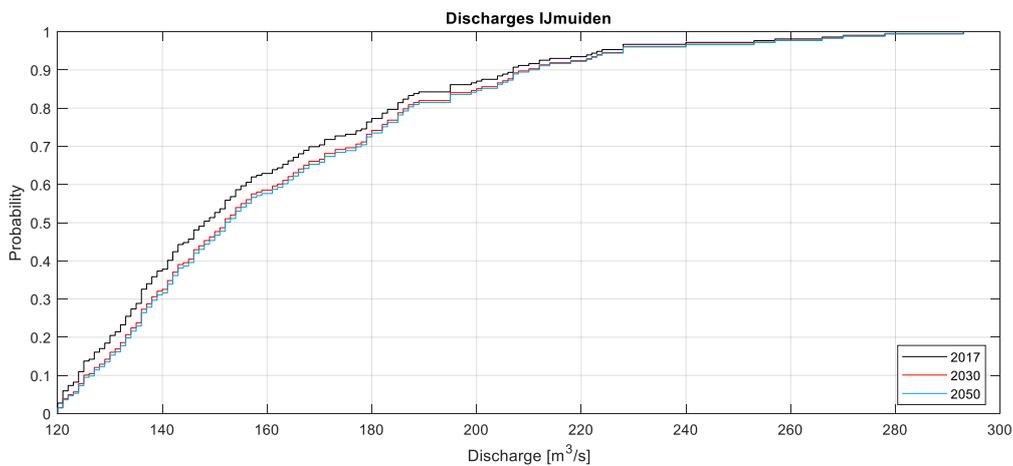
Figure 38: CDF of the Required Discharges over time

11.3. PROBABILITY OF PUMP STATION FAILURE IN TIME

Sampling and simulating the before-mentioned distributions - derived via conditionalizing the NPBN - returns the underlying probability of asset failure. The probability of failure can be interpreted as 'the probability that the system cannot deliver the required capacity, while required'. Figure 39 is a result of the simulations, conditionalized on the values mentioned in paragraph 8.1 which represent the different scenario-packages. The *High* and *Low* values can be interpreted as the extreme probabilities of failure.
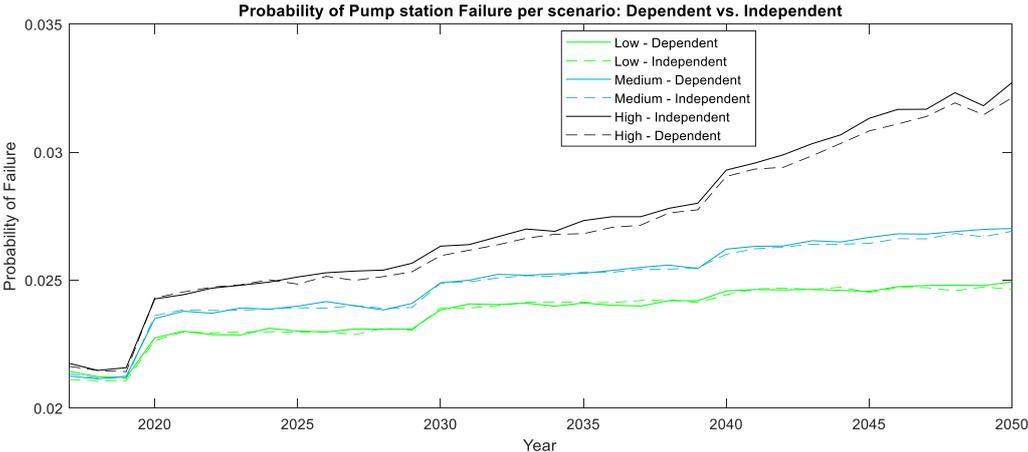
Figure 39: Probability of asset failure in time per scenario

Note that these probabilities of failure are relatively high. The main contributor is the fact that the marginal distribution of the discharges in IJmuiden in the NPBN is an approximated dataset, see appendix V.2.. The original dataset covered all discharged data – by pumps as well as by free discharge. Still this is our best approximation, which shows to contain higher discharges than the cumulative density function of the pumped discharges in van der Wiel et al. (2013). Consequentially, the probabilities will rise significantly. Although the results of the simulations can be doubtable since they rely on many assumption, three main conclusions can be drawn.

First note that the probability of pump station failure increases over time. The overall increase of the curves can be explained by the fact that the probabilities of the extreme discharges and maximum total discharge capacities – see figures 37 and 38 - a slowly converge.

Secondly, note that the slopes of mainly the lines are very small. The small slope indicates that climate change impacts the failure probability of pumps to very small proportions. In other words, the ARC/NSC-area does not have to fear for the effects of climate change on the short term. However, for the asset manager a small slope illustrates danger since a 'hard' moment of failure cannot be designated.

Third, the contribution of the environment-dependent model is limited. The transition probabilities – dependent on the time-to-failure and sea level rise – show a very small changes per time step. This can also be attributed to the very gradual effect of climate change.

Overall one can say that the currently estimated effects of climate change are very limited. However, this statement immediately raises the question whether the effects are also limited when the current predictions are wrong.

11.4. HYPOTHETICAL CASE: UNEXPECTED SEA LEVEL RISE

This paragraph considers the case that the current predictions are incorrectly determined and sea level rise occurs twice as fast. The sea level rise increments are also based on Bamber & Aspinall (2013). Here the probability of failure will be briefly examined and reflected to the previous findings, discussed in the previous chapter. The results only consider the Medium case.

Figure 40 shows the probability of failure of pump station failure influences by sea level rise that develops twice as fast as currently predicted. For comparison, the medium case of the previous paragraph is illustrated. Note that both lines are diverging over time. Thereby, note that the

environment-dependent deterioration model shows larger differences with the environment-independent deterioration model than in the previous case.
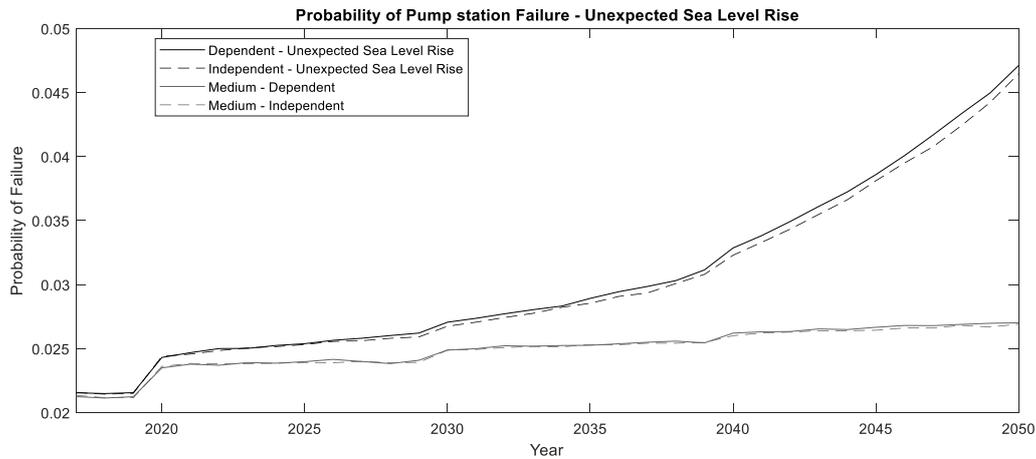
Figure 40: Probability of asset failure based on 'new' sea level rise predictions

The exponential increase of probabily of failure can be directly related to the Q-H curve. The figure below shows the discharge capacity occurrences per time unit. Note that – due to water head increase – the minimum pump capacities shift to the lower values as expected in paragraph 1.5. However, due to the enormous sea level rise – maximally 900mm – the probability that the pump cannot pump the discharges increases. Note for example that the pump already hits its minimum capacity in 2050. Although it sounds very logical, the model that once the water levels exceed the pump's discharge range, the probability of failure increases exponentially and system failure is upcoming.
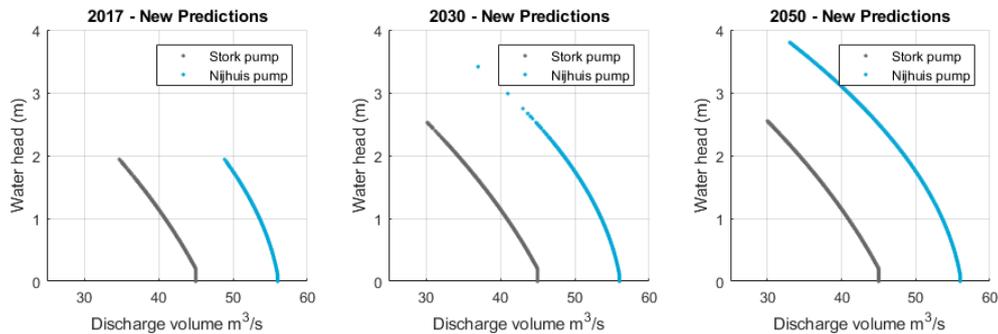
Figure 41: CDF of the Required Discharges over time

Part IV – Conclusions and Recommendations

This page was intentionally left blanc

12.   CONCLUSION

This chapter concludes on the main research question which evolved from the research problem and research question. To recap, this research stated that since large maintenance assignments are upcomming, hydraulic asset owners require accurate predictions of their assets lifecycle-end. Since climate change fortifies asset deterioration - and consequentially increases the failure probability over time -, long term trends as climate change must be incorporated into the predictive failure models. To narrow the scope of the research, the focus is on coasta pump stations. Following from this statement, we derived the following research question:

> *'How can long-term trends – such as climate change – be introduced to reliability analysis in order to determine the effect on the pump stations reliability on the longer term?'*

This research provided a reliability analysis of coastal pump station that is supported by non-parametric Bayesian Networks and Markov chains to quantify the effect of the long term trend on the assets probability of failure. Part II: Reliability Analysis for pump stations described the path from NPBN set-up, through theoretical validation, to the performance of simulations in order to require the probability of failure in time.

This formation of the model departed from Kosgodagan-Dalla Torre et al., (2017), which is the first who proposed a method that made a stochastic-process-based deterioration model dependent on its environment. In his study, Bayesian Networks are utilized to model the dependencies between covariates and the input of the deterioration model, represented by a discrete Markov chain. Unfortunately, the use of the discrete Bayesian Network makes the model unsuitable to incorporate long-term trends which requires the support continuous distributions. For this reason, Non-Parametric Bayesian Networks – that are able to support discrete- and continuous marginal distributions – are adopted to make the deterioration input dependent on the covariates. The NPBN's input is limited by its marginal distributions and rank correlations. Ideally those are derived from data, but most often – particularly in failure modelling – the marginal distributions are not known. In the latter case, expert judgement can be applied to determine the marginals.

Based on these tools, the reliability analysis by handling the following seven steps. **First,** the limit state function should be derived. In this research a general limit state function for pump stations is presented:

$$Z(t) = \left(\sum_{n=1}^{N} A_{pump,n}(t) * C_{pump}(t) \right) - Q_{capacity;required}(t)$$

Secondly, the NPBN must be set up. Here, the dependencies must be drawn between environmental variables such as rainfall or sea level, and the variables of interest. **Third,** if the marginal distributions are determined by data, the model must be verified whether it sufficiently represents the dependence structure. This research presented and applied three varication tests. **Fourth,** dependent on the effect of climate change, the marginal distributions must be determined per time slice. This can be performed via inference or distribution estimations methods such as expert judgement. In case of rainfall, Klein Tank et al. (2015) expect the extremes to occur more often. Assuming that the extremes are covered in the marginal distributions, the marginal distributions do not have to be modified for every time slice. Now, Bayesian inference enables it to stress a belief into the deterioration input by conditionalizing on the variables directly impacted by the long term trend. In case of sea level rise, Bamber & Aspinall

(2013) showed that the magnitudes of future sea levels are not incorporated in the current datasets. Then the marginals should be adapted for each time slice in order to represent all possible sea levels at time t. **Fifth,** component deterioration must be added in order to determine the pump availability in time. This research proposes a combined approach of NPBNs and Markov chains that is based on Kosgodagan-Dalla Torre et al., (2017). The underlying idea is to update the transition probabilities, based on the environment. The outcome of is the pump-component's unavailability in time. **Six.** Based on the system decompositions and the gained component unavailability per time unit, one is able to calculate the probability that an individual pump is unavailable. The **seventh -** and final - step is to determine the total asset's probability of failure in time by conduction simulations.

In order to verify whether the model does what it is supposed to do, the model has been applied to the case study of pumping station IJmuiden. The pumping station's performance is affected by long term trends as increase in extreme rainfall occurrences and sea level rise. Both effects are incorporated into the model. After running the simulations one can say that the currently estimated effects of climate change to the pump stations reliability are very limited for the 2050 estimations. However, the model shows that when the sea level rises to such extent that it is out of pump-curve range, failure behaves exponentially. We can state that the NPBN-supported reliability analysis sufficiently incorporates long-term trends in order to determine the probability of failure.

13. DISCUSSION

This chapter discusses the journey throughout this research to propose a method and fit the proposed methods to the case study. Here, the fundamental choices are critically discussed. This chapter discusses the impact of those assumptions and simplifications which serves as points of reference for further research and summons the link to the recommendations in the next chapter.

13.1. METHOD RELATED

This research proposed a melting pot of probabilistic methods to describe the relationship between deterioration and environment. Probabilistic methods showed to be commonly accepted in infrastructure asset management. Therefore this research builds on probabilistic concepts such as stochastic-based deterioration models. Although the method proved to be valid, a lot of time, effort and statistical background knowledge is required to set up a model like this. Therefore one can question whether probabilistic methods are the best way to tackle the problem.

The opposite of probabilistic methods are deterministic methods. One deterministic method that is able to represent the dependencies between variables is *system dynamics*. System dynamics elementary builds on differential equations, which directly marks the lack of adaptability once a variable is observed, see appendix III – Bayesian Reasoning. Thereby, system dynamics assumes a perfect correlation between variables, while this is absolutely not true. In other words, once one is interested in the development of a certain variable over time and its dependent variables can be observed, the probabilistic method can be assumed to be the most suitable method. Despite the fact that it requires lots of afford to set up, the benefits are great.

13.2. CASE RELATED

The theoretical method is reflected to the case study of Pumping station IJmuiden. To prove that the model actually worked as we expected, several assumptions are made that impact the case related results. Here, the case-related assumptions and simplification will be criticized.

- **Use of the real pump discharges**
  The marginal distribution of the discharges in IJmuiden in the NPBN is determined by the dataset that covers all discharged data – by pumps as well as by free discharge. In the data analysis we voided the discharges > 260 m^3/s since they cannot be pumped by the pumping station. What remains is a data set of discharge samples 120 – 260 m^3/s of which we do not know which are pumped by the pumping station or freely discharged. Still this was the best approximation regarding the pumped discharges at IJmuiden.

- **Unjustified use of the Gaussian copula**
  The model validation in paragraph 10.4. showed that the utilization of the Gaussian copula might not be the best assumption. The goodness-of-fit tests showed that the utilization of the Gumbel copula would be more justified. Due to the benefits of Gaussian copula utilization – rapid calculations and UniNet-support – we retained the Gaussian copula assumptions. The consequences of the assumption of the Gaussian copula are hard to estimate qualitatively; it dependents on the dependence structure of the whole network.

- **Estimations of the unknown marginal distributions**
  Then the unknown marginal distributions – related to the input-parameter of the stochastic-process-based deterioration model - are assumed via estimating by the 5th, 50th and 95th

percentiles, given its parental node. Then, a parametric distribution is fitted to the percentiles. Where the results show that the probability of failure is highly sensitive to the outcomes of the Markov chain, more accurate estimations of the unknown marginal distributions is required. The same counts for the estimations of the rank correlations.

- **Use of an oversimplified Markov chain**
  One of the main discussion points is the use of an oversimplified deterioration model is used; a two-state Markov chain. The transition probabilities are determined by their MTTF and MTTR. The MTTF is determined by the mean of the drawn time-to-failure distribution samples. Mainly in the low and high scenario showed to the MTTF to fluctuate over time. This is because the MTTF is calculated based on 2500 samples of a (conditionalized) Weibull distribution. Since the increase is very small per time step, this small sample amount can cause shifts in the MTTF.

- **Number of simulations**
  Although the probability of asset failure-curve shows an increasing trend, it is pretty 'bumpy' where we expect a smooth line. One of the main reasons is that the involved distributions increase in uncertainty over time. This is why the line gets more 'bumpy' when time matures. In this research, we performed 50.000 simulations times the number of peak-discharge occurrence per year, which is already an enormous amount since the simulations are performed over 34 years.

Out of the above mentioned discussion-points we can conclude that mainly the application of the proposed framework contains some doubtable points that require further research. The next chapter summons some ideas regarding future research to make proposed framework more accurate.

14. RECOMMENDATIONS

This study introduced a combination of reliability analysis theory, Non-Parametric Bayesian Networks and stochastic-process-based deterioration models to model the dependency between component-deterioration and a dynamic environment, influenced by a long term trend such as climate change. Finally, the proposed model is able to calculate the probability of system failure. During the development of the model ideas came up that would improve the model or even extend it. The proposed ideas to improve the model mainly concern improvements regarding the accuracy of the model outcome. Before those ideas will be elaborated, probably the most important recommendation will be discussed first in order to emphasize its importance.

14.1. GENERAL RECOMMENDATIONS

The general recommendations concern the overall recommendations that are not entirely the aim of this research, but are worth to be mentioned.

- **Standardize the data-gathering procedure**
  Within this research it showed to be challenging to gather sufficient- and suitable data. While the world is slowly shifting towards data-based applications, such as Artificial Intelligence, the input into these new concepts is lacking behind. Consequentially, the motivation to initiate those new concepts into the business is very low since it currently takes much effort to retrieve suitable input. The first main reason noticed throughout the research process is that the entities responsible for the asset's management are not critical enough regarding their datasets. Although there are exceptions, mainly the water board's data sets contain enormous amounts of gaps while this can easily be solved by reliable measurements. The second reason is that the data-sets are all gathered on different intervals. In this research, the discharges passing by were gathered every day, every ten minutes or every 20 minutes. In order to compare the data, the researcher must first transform the datasets to higher time-intervals; one day or even one year. Transformation limits the possibilities of the model; for example, to model the maintenance over time one actually needs an hour time-scale.

  Ideally, ideally continuous measurements should solve this but are known to be expensive since continuous data extends the required storage capacity enormously. Therefore we argue mainly governmental bodies – in this case Rijkswaterstaat and the water boards – to contribute to a standardized data gathering procedure in order to derive data with the same time-interval and invest in reliable measurements to collect the data without any gaps.

  Thereby, asset managers should gather their data in more structured in accordance to the input parameters of stochastic processes such as the Markov Chain and Gamma Processes. Admitting that it might be relatively hard to define more than 2 observable states, it would be an added value to do more accurate estimations regarding future failure and base it on the physics of failure. For the Markov Chain this would be expected transition time-distributions for each state, such as sketched in chapter 7. For the Gamma Process this would be a distributions of the cumulative deterioration on predefined inspection intervals.

14.2. MODEL RELATED IMPROVEMENTS

This paragraph sums up the general model improvements to the reliability analysis presented to determine the probability of coastal pump failure in time that is influenced by a long term trend. The prosed recommendations are presented by importance.

- **Including the Clayton and Gumbel copula in UniNet**

  Although the Gaussian copula contains huge benefits and is applicable to the predominant part of applications, this research showed to be an exception. Here, the Gumbel copula is the best fit. Unfortunately, to our knowledge no software has yet been developed that is able to represent the NPBN-arcs with Gumbel or Clayton copulae. A software package that supports the Gaussian, Gumbel and Clayton copula would be a solution to optimize the dependency structure.

- **Towards a Dynamic Non-Parametric Bayesian Network**

  In this research are the separate time-slices treated as independent. However, in reality they actually are dependent. For example; given a high SLR this year, then the SLR is also likely to be high the next year. In order to model dependencies between time slices, Kosgodagan-Dalla Torre et al. (2017) and Straub (2009) showed that Dynamic Bayesian Networks are perfectly capable to stress a belief into future variable distributions based on observations in the present, see paragraph 5.1.2.. Another advantages of DBNs is that modelling the yearly dependencies would also make the reliability analysis more suitable for future decision-making. For example, the construction of a lock in 2030 leads to higher leakages. Then, one is able to stress a future belief regarding the pump capacities of surrounding pump stations, given higher discharges in 2030.

- **Utilization of the Gamma Process**

  In contrast to the previous recommendation for future research, one can also use the Gamma process as stochastic-based deterioration method. Appendix II elaborated the Gamma process. Compared to the Markov chain, the Gamma process is a more elegant process to derive the time-to-failure distribution. Adopting the Gamma process instead of the Markov chain would mean the following for the model; instead of making the environment dependent on the time-to-failure distribution, adaptation of the Gamma process indicated that the environment must be made dependent on the data describing the cumulative deterioration at predefined points in time.

- **Detailed applications**

  This research determined the probability of failure on a yearly-time scale basis including renewals in case of failure. Although this time-scale shows to be suitable for the problem mentioned, other cases ask for a more detailed approach by modeling on a daily-time scale basis. Note that the models complexity would then exponentially increase since the number of time-slices – and with that nodes and arcs – increase. Thereby, modeling on a daily time-scale would also include seasonal changes. In other words, modeling on smaller time scales would be a research itself. Here must be noted that a more detailed approach would increase the computation effort enormously.

14.3. CASE RELATED IMPROVEMENTS

This paragraph sums up the recommendations for future research that were encountered through the process. The proposed ideas to improve the model mainly concern extensions that would increase the accuracy of the model outcome. The recommendations are order by their importance, beginning with the right inclusion of the stochastic-based process, since the used two-state Markov chain shows severe limitations to reliably assess the probability of asset failure.

- **Add pump policies**

  In this research is assumed that the NSC/ARC-channel behaves as a river; the upcoming discharges must be pumped with certainty. However, in reality, when the discharges are too high, the surrounding water boards stop pumping their excess water into the NSC/ARC-channel to reduce the required pump discharges at IJmuiden. This would give more reliable quantities for the probability of pump station failure.

- **Increase number of states Markov chain process**

  A n-state Markov chain contains benefits in modeling failure and maintenance. Including more than two states enables the deterioration process to include the rates of deterioration. For example, the probability that the component transits from state 1 to 2 can be lower than the probability that the component transits from state 2 to 3. This more detailed approach would contribute to the accuracy of component deterioration. Concerning maintenance modelling, the adopted two-state Markov chain assumes renewals to the 'as new' state. A benefit of an n-state Markov chain, maintenance can be modeled as being 'imperfect' by defining the transition probabilities that the component goes to state n at time t after maintenance.

- **Utilization of Expert Judgement for marginal distributions and rank correlations**

  In chapter 6 is stated that the NPBN's input is limited to its marginal distributions and rank correlations. In the ideal case, the both are defined by the associated data sets of two random variables. However, in many cases data is not available or not representative. In this research the unknown marginal distributions are simply determined by estimating their 5th, 50th and 95th percentiles, given the parental nodes. Thereby, the unknown rank correlations are estimated by logical reasoning. Although not applied in this research, there are methods to defensibly quantify both inputs. Cooke (1991) presents a method to determine the unknown marginal distributions based on expert elicitations. Thereby, Morales, Kurowicka, & Roelen (2008) proposed a model to defensibly elicit the (un)conditional rank correlations.

This page was intentionally left blanc

15. BIBLIOGRAPHY

Abdel-Hameed, M. (1975). A Gamma Wear Process. *IEEE Transactions on Reliability*, *R-24*(2), 152–153. https://doi.org/10.1109/TR.1975.5215123

ANP. (2016). "Bouwers Merwedebrug hadden zo veel verkeer niet voorzien." *NOS*. Retrieved from https://nos.nl/artikel/2139101-bouwers-merwedebrug-hadden-zo-veel-verkeer-niet-voorzien.html

Baik, H., Seok, H., Jeong, D., & Abraham, D. M. (2006). Estimating Transition Probabilities in Markov Chain-Based Deterioration Models for Management of Wastewater Systems. *Journal of Water Resources Planning and Management*, *132*(February), 15–24. https://doi.org/10.1061/(ASCE)0733-9496(2006)132:1(15)

Bamber, J. L., & Aspinall, W. P. (2013). An expert judgement assessment of future sea level rise from the ice sheets. *Nature Climate Change*, *3*(4), 424–427. https://doi.org/10.1038/nclimate1778

Briggs, E. K., & Hodkiewicz, M. (2005). Maintenance and Reliability of a Water Pumping Station. In *9th International Conference of Maintenance Societies*.

Cinlar, E. (1972a). Markov Additive Processes {I}. *Z. Wahrscheinl. Verw. Geb.*, *24*, 95–121. https://doi.org/10.1007/BF00532536

Cinlar, E. (1972b). Markov Additive Processes {II}. *Z. Wahrscheinl. Verw. Geb.*, *24*, 85–93. https://doi.org/10.1007/BF00532536

Clemen, R. T., & Reilly, T. (1999). Correlations and Copulas for Decision and Risk Analysis. *Management Science*, *45*(2), 208–224. https://doi.org/10.1287/mnsc.45.2.208

Cooke, R. M. (1991). *Experts in uncertainty: opinion and subjective probability in science*. New York: Oxford University Press.

Edirisinghe, R., Setunge, S., & Zhang, G. (2015). Markov Model — Based Building Deterioration Prediction and ISO Factor Analysis for Building Management. *Journal of Management in Engineering*, *31*(6), 1–9. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000359.

Frangopol, D. M., Kallen, M.-J., & Noortwijk, J. M. van. (2004). Probabilistic models for life-cycle performance of deteriorating structures: review and future directions. *Progress in Structural Engineering and Materials*, *6*(4), 197–212. https://doi.org/10.1002/pse.180

Genest, C., Rémillard, B., & Beaudoin, D. (2009). Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics*, *44*(2), 199–213. https://doi.org/10.1016/j.insmatheco.2007.10.005

Grall, A., Dieulle, L., Bérenguer, C., & Roussignol, M. (2002). Continuous-time predictive-maintenance scheduling for a deteriorating system. *IEEE Transactions on Reliability*, *51*(2), 141–150. https://doi.org/10.1109/TR.2002.1011518

Hanea, A. (2008). *Algorithms for non - parametric bayesian belief nets*.

Hanea, A., Kurowicka, D., & Cooke, R. M. (2007). Hybrid Method for Quantifying and Analyzing Bayesian Belief Nets. *Quality and Reliability Engineering International*, *22*(2006), 709–729. https://doi.org/10.1002/qre.808

Hanea, A. M., Kurowicka, D., Cooke, R. M., & Ababei, D. A. (2010). Mining and visualising ordinal data with non-parametric continuous BBNs. *Computational Statistics and Data Analysis*, *54*(3), 668–687. https://doi.org/10.1016/j.csda.2008.09.032

Hanea, A., Morales Napoles, O., & Ababei, D. (2015). Non-parametric Bayesian networks: Improving theory and reviewing applications. *Reliability Engineering and System Safety*, *144*, 265–284. https://doi.org/10.1016/j.ress.2015.07.027

Hartnett, K. (2018). To Build Truly Intelligent Machines, Teach Them Cause and Effect. *Quantamagazine*, 1–12.

Hastings, N. A. J. (2015). *Physical Asset Management Handbook*. https://doi.org/10.1007/978-3-319-14777-2

Hong, H. P. (1999). Inspection and maintenance planning of pipeline under external corrosion considering generation of new defects. *Structural Safety*, *21*(3), 203–222. https://doi.org/10.1016/S0167-4730(99)00016-8

IPCC. (2014). *Summary for Policymakers. Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. https://doi.org/10.1017/CBO9781107415324

Jardine, A., & Tsang, A. (2013). *Maintenance, replacement and reability: Theory and Applications*.

https://doi.org/10.1017/CBO9781107415324.004

Joe, H. (2014). *Dependence Modeling with Copulas*. Boca Raton, FL: Taylor & Francis.

Jonkman, S. N., Steenbergen, R. D. J. M., Morales-Nápoles, O., Vrouwenvelder, A. C. W. M., & Vrijling, J. K. (2015). *Probabilistic Design: Risk and Reliability Analysis in Civil Engineering*. (M. H. G. Baas, Ed.).

Jonkman, S. N., Voortman, H. G., Klerk, W. J., & van Vuren, S. (2018). Developments in the management of flood defences and hydraulic infrastructure in the Netherlands. *Structure and Infrastructure Engineering*, 1–16. https://doi.org/10.1080/15732479.2018.1441317

Jorissen, R. E., & van Noortwijk, J. M. (1998). Instrumenten voor optimaal beheer van waterkeringen. *Het Waterschap*, (83), 6–12 & 42–47.

Klatter, L., & Roebers, H. (2017). Assessment of Need for Renewal on a Multi-Network Level. In J. Bakker, D. M. Frangopol, & K. van Breugel (Eds.), *Life-Cycle of Engineering Systems: Emphasis on Sustainable Civil Infrastructure* (pp. 1643–1649). London: Taylor & Francis Group.

Klein Tank, A., Beersma, J., Bessembinder, J., Hurk, B. van den, & Lenderink, G. (2015). KNMI'14-klimaatscenario's voor Nederland; Leidraad voor professionals in klimaatadaptatie. *KNMI Klimaatscenario's '14*, 34.

Klutke, G., & Sanchez-Silvia, M. (2016). *Reliability and Life-Cycle Analysis of Detoriorating Systems*.

Kosgodagan-Dalla Torre, A., Yeung, T. G., Morales-Nápoles, O., Castanier, B., Maljaars, J., & Courage, W. (2017). A Two-Dimension Dynamic Bayesian Network for Large-Scale Degradation Modeling with an Application to a Bridges Network. *Computer-Aided Civil and Infrastructure Engineering*, *32*(8), 641–656. https://doi.org/10.1111/mice.12286

Kurowicka, D., & Cooke, R. M. (2005). Distribution-Free continuous bayesian belief nets. *Modern Statistical and Mathematical Methods in Reliability*, 309–323.

Male, S. (2010). The challanges facing public sector asset management. In *Asset management : whole-life management of physical assets* (pp. 50–73). ICE Publishing.

Ministery of Infrastructure and Environment. (2016). *Onderhoud Strategische Bruggen Hoofdwegennet*. Den Haag. https://doi.org/10.1007/s00244-010-9510-9.-

Morales-Nápoles, O., & Steenbergen, R. D. J. M. (2014). Analysis of axle and vehicle load properties through Bayesian Networks based on Weigh-in-Motion data. *Reliability Engineering and System Safety*, *125*, 153–164. https://doi.org/10.1016/j.ress.2014.01.018

Morales, O., Kurowicka, D., & Roelen, A. (2008). Eliciting conditional and unconditional rank correlations from conditional probabilities. *Reliability Engineering and System Safety*, *93*(5), 699–710. https://doi.org/10.1016/j.ress.2007.03.020

Murphy, K. P. (2002). *Dynamic Bayesian Networks: Representation, Inference and Learning*. University of California. https://doi.org/10.1.1.129.7714

Nicolai, R. P. (2008). Maintenance Models for Systems subject to Measurable Deterioration, 198.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems*. San Francisco, CA.: Morgan Kauffmann Publishers.

Rayner, R. (2010). Incorporating climate change within asset management. In *Asset management : whole-life management of physical assets* (pp. 161–180). ICE Publishing. Retrieved from https://app.knovel.com/web/toc.v/cid:kpAMWLMPAF/viewerType:toc/root_slug:asset-management-whole

Rijkswaterstaat. Vervangingsopgave Natte Kunstwerken; Meerwaarde van de Gevoeligheidstest Natte Kunstwerken voor Adaptief Deltamanagement (2014).

Ross, S. M. (2014). *Introduction to Probability Models* (11th Editi). Elsevier.

Singpurwalla, N. D. (1995). Survival in Dynamic Environments. *Statistical Science*, *10*(1), 86–103. https://doi.org/10.1111/j.l558-5646.2010.01056.ADAPTATION

Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, *8*, 229–231.

Sperotto, A., Molina, J. L., Torresan, S., Critto, A., & Marcomini, A. (2017). Reviewing Bayesian Networks potentials for climate change impacts assessment and management: A multi-risk perspective. *Journal of Environmental Management*, *202*, 320–331. https://doi.org/10.1016/j.jenvman.2017.07.044

Straub, D. (2009). Stochastic Modeling of Deterioration Processes through Dynamic Bayesian Networks. *Journal of*

Engineering Mechanics, *135*(10), 1089–1099. https://doi.org/10.1061/(ASCE)EM.1943-7889.0000024

U.S. Department of Transportation. (2015). *Climate Change Adaptation Guide for Transportation Systems Management, Operations, and Maintenance*. Retrieved from file:///C:/Users/Kyle P. Kwiatkowski/Documents/Mendeley Desktop/Asam et al. - 2015 - Climate Change Adaptation Guide for Transportation Systems Management, Operations, and Maintenance.pdf

van der Wiel, W. D., Persoon, E., & Stiksman, G. LCC-optimalisatie Gemaal IJmuiden - Korte en Lange Termijn (2013).

van Dongen. (2011). *Referentiemodel voor levensduurverlenging van technische assets*.

van Noortwijk, J. M. (2009). A survey of the application of gamma processes in maintenance. *Reliability Engineering and System Safety*, *94*(1), 2–21. https://doi.org/10.1016/j.ress.2007.03.019

van Noortwijk, J. M., & Klatter, H. E. (1999). Optimal inspection decisions for the block mats of the Eastern-Scheldt barrier. *Reliability Engineering and System Safety*, *65*(3), 203–211. https://doi.org/10.1016/S0951-8320(98)00097-0

van Noortwijk, J. M., & Peerbolte, E. B. (2000). Optimal Sand Nourishment Decisions. *Journal of Waterway, Port, Coastal, and Ocean Engineering*, *126*(January/Februari), 30–38.

Verma, A. K., Ajit, S., & Karanki, D. R. (2016). *Mechanical Reliability*. *Reliability and Safety Engineering*. https://doi.org/10.1007/978-1-4471-6269-8_7

Wang, H. (2002). A survey of maintenance policies of deteriorating systems. *European Journal of Operational Research*, *139*(3), 469–489. https://doi.org/10.1016/S0377-2217(01)00197-7

This page was intentionally left blanc

Appendices

This page was intentionally left blanc

APPENDIX I – RELIABILITY ANALYSIS

Reliability can be defined as the ability of an asset to perform a required function under state conditions for a specified period of time (Hastings, 2015). Pump stations does not perform its function when the discharge capacity is lower than the required capacity. This chapter elaborates on the central concept to introduce the long term trend.

In order to treat the problem described in chapter 1 in a probabilistic way, the point of departure is the reliability analysis. In terms of structural reliability the probability that the system functions at a specific time, can be represented as follows:

$$P[System\ is\ reliable] = P[R \geq S] \tag{16}$$

Where R is the resistance and S the solicitation, both described by their distributions. Logically, but still worth to mention, those distributions can be parametric or non-parametric in case of data. Instead of thinking in terms of system reliability, literature prefers to reason in terms of system unreliability. The unreliability of a system is also denoted as the probability of failure (Verma, Ajit, & Karanki, 2016). In other words, the system reliability is simply the complement of the probability of failure. In its simplest format the system is failed when the resistance (R) is smaller than the solicitation (S), $R < S$.
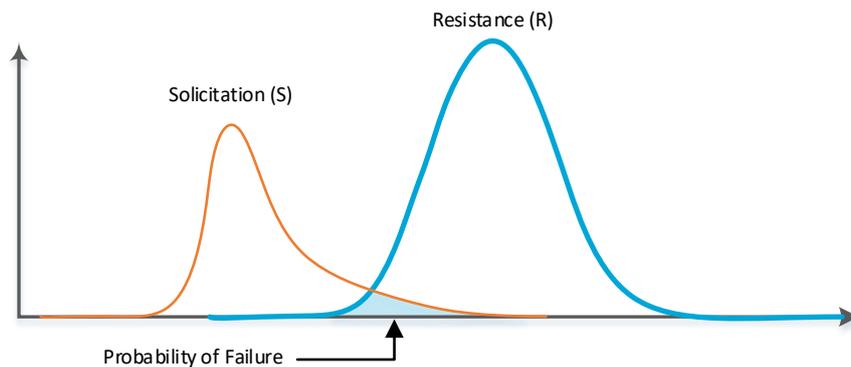
Figure 42: Definition probability of failure

When the distributions of the resistance (R) and solicitation (S) are known, the probability of failure P_f can be calculated as the probability that the solicitation (S) is higher than the resistance (R), see figure 42. The same problem can be formulated by means of a limit state. This is an equation beyond which the system or part of the structure does no longer fulfil one of its performance requirements, technically of functionally. The limit state (Z) can be assessed by considering the resistance (R) and the solicitation (S):

$$Z = R - S$$
$$P_f = P[Z < 0]. \tag{17}$$

The equation above is called the limit state with accompanied definition of its probability of failure. Depending on the type of infrastructure under consideration, the limit state of various functions can be defined. For example, breaching is the crucial limit state for flood defenses. For other structures the key requirement can also refer to another function, such as the available discharge capacity of a discharge sluice or the availability of a navigation lock (Jonkman et al., 2018).

APPENDIX II - STOCHASTIC-PROCESS-BASED MODELS

The aim of this appendix is (1) to explain stochastic-process-based methods which Singpurwalla (1995) prefers above the currently used failure rate models and (2) to quantify the independent $X(t)$ mentioned in the previous appendix. Stochastic-process-based models are also called deterioration-models since they describe the decay of a certain object. First the reason to use stochastic processes to describe the decay of an object will be explained in more detail. Thereafter two stochastic processes that are commonly accepted, Markov- and Gamma processes, will be elaborated.

II.1. STOCHASTIC PROCESSES

Generally, stochastic processes, also called random processes, model the evolution of a random system in time. Mathematically, a stochastic process is defined as a family of random variables $X(t)$ defined on a given probability space and indexed by t belonging to a parameter set T. The set T is the time sequence of the process and it can be discrete ($T = \{0,1,2,\dots.,t\}$) or continuous ($T = \{0,\infty\}$). See figure 43, the range of the random variable $X(t)$ reduces a state space S, which can be discrete or continuous. Here we treat two stochastic processes that are generally accepted in maintenance optimization of infrastructural assets: (discrete-state) Markov chains and (continuous-state) Gamma processes.
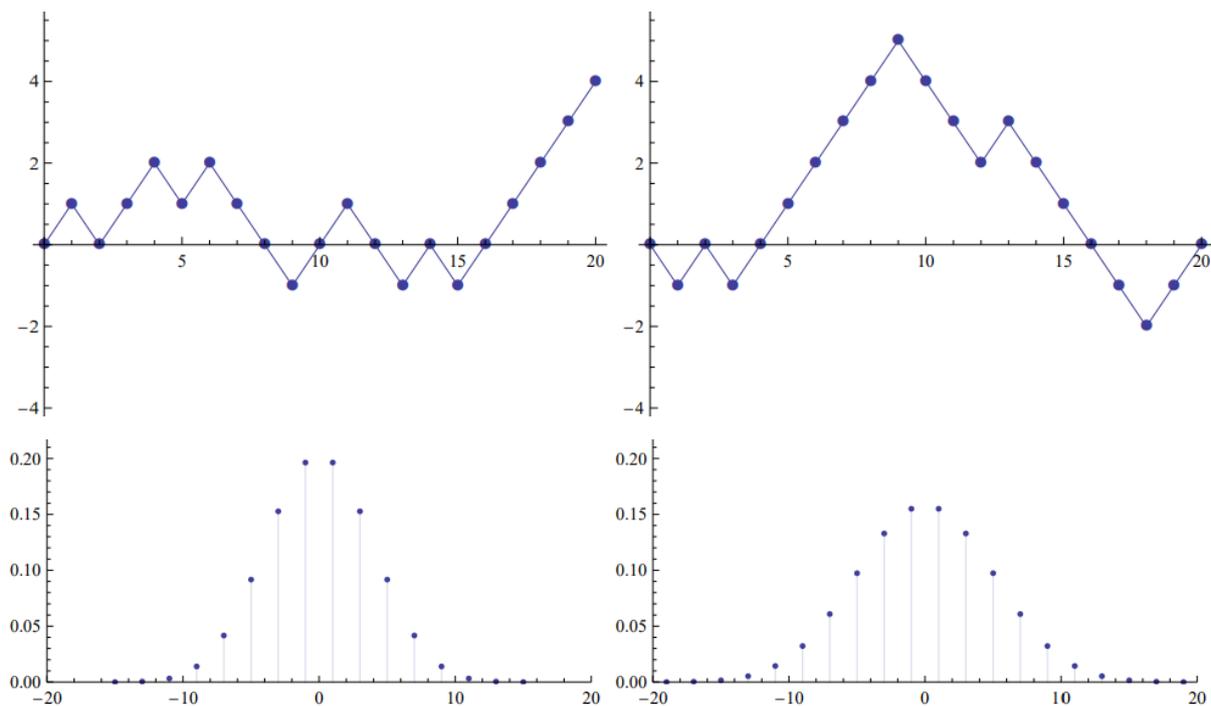
Figure 43: Two different trajectories of random walks

A stochastic process is named a Markov process when it contains the so-called Markov property. The Markov property states that the future state of $X(t)$ is only dependent on the present state, i.e. the past deterioration has no influence on the future deterioration. Another stochastic process is the Gamma process, initiated by Jan van Noortwijk. Published examples of stochastic-process applications to deterioration are erosion (van Noortwijk, 2009) or corrosion (Hong, 1999). The former is modeled by a Gamma process and the latter by a (discrete-state) Markov chain.

II.2. Markov chains

Markov chains have been extensively used in the context of risk, reliability, and maintenance management for civil infrastructures (Baik et al., 2006; Edirisinghe et al., 2015; Klutke & Sanchez-Silvia, 2016). Let's assume that random deterioration quantity $X(t)$ has a discrete state space {1,2,3,4}, which implies that $X(t)$ can take 4 states, for example {$As\ new, good, bad, failed$}, see figure 44.

II.2.1. Mathematics

In order to determine the probability that the component reaches the fourth state "failed" over time, we first have to determine the probabilities that the component goes from state i to state j for one time step Δt, see the underlying figure.
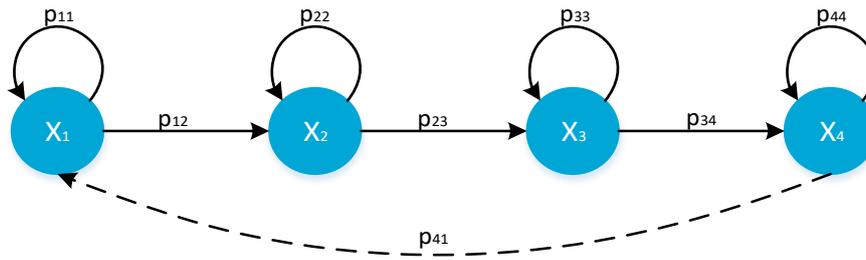
Figure 44: Schematic representation Markov chain, including renewals

This probability is called the transition probability, denoted as p_{ij}. Logically, p_{ii} represents the probability that the system remains in the same state as the previous state. According the Markov property, which states that the future state is only dependent on the present state, p_{ij} can mathematically be defined as follows for the first time step (Ross, 2014):

$$p_{ij} = P(X_1 = good|X_0 = As\ new) \tag{18}$$

p_{ij} is often also named as the one-step transition probability from state i to state j, with $p_{ij} \geq 0$, $i,j \geq 0$ and $\sum_{j=0}^{\infty} p_{ij} = 1, i = 1,2,\dots,\Omega$. Let $\boldsymbol{P_{ij}}$ denote the matrix of one-step transition probabilities p_{ij}, so that:

$$\boldsymbol{P_{ij}} = \begin{pmatrix} p_{1,1} & \cdots & p_{1,\Omega} \\ \vdots & \ddots & \vdots \\ p_{\Omega,1} & \cdots & p_{\Omega,\Omega} \end{pmatrix} \tag{19}$$

Now we have to define the t-step transition probability matrix $\boldsymbol{P_{ij}^t}$, which represents the transition probabilities p_{ij}^t for time t:

$$p_{ij}^t = P(X_{t+k} = j|X_k = i) \tag{20}$$

Then, via the Chapman-Kolmogorov equations is shown that the t-step transition matrix $\boldsymbol{P_{ij}^t}$ may be obtained by multiplying the matrix $\boldsymbol{P}$ by itself t times (Ross, 2014):

$$\boldsymbol{P_{ij}^t} = \boldsymbol{P^t} \tag{21}$$

Finally, we define the state probability vector at time n, $\boldsymbol{p^t}$, as a row vector. Then, given the initial state probability $\boldsymbol{p^0}$ and the one-step transition probability matrix $\boldsymbol{P}$, we can easily determine $\boldsymbol{p^t}$ by (Klutke & Sanchez-Silvia, 2016):

$$\boldsymbol{p^t} = \boldsymbol{p^t P} = \boldsymbol{p^0 P^t} \tag{22}$$

Since $\boldsymbol{p^0}$ is a $1 \times m$-matrix and $\boldsymbol{P^t}$ a $m \times m$-matrix, $\boldsymbol{p^t}$ is a $1 \times m$-matrix. Then, the mth-entry of $\boldsymbol{p^t}$ represents the probability that the component is in the failed-state at time t, i.e. the probability of component failure at time t. After applying the Markov process, the underlying figures show representative pdf and cdf.
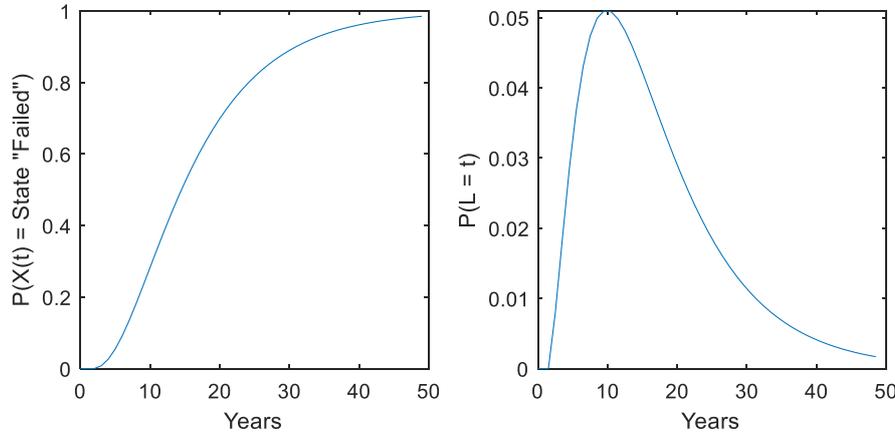
Figure 45: Output Markov chain

II.2.2. Input parameters

To conclude, via estimating the transition probabilities from one state to the other, we are able to determine the probability of failure for each time step. Those probabilities can be estimated by data or via expert judgement (Baik et al., 2006; Kosgodagan-Dalla Torre et al., 2017).

II.3. GAMMA PROCESSES

In comparison with the Markov chain, the Gamma process is a more complex deterioration process. Abdel-Hameed (1975) was the first who proposed to use the Gamma process as a model for random deterioration in time. Since the initial paper, the same author used the Gamma wear process as a building block for developing mathematical models for optimizing time-based maintenance and condition-based maintenance. Thereafter, many other researchers adopted the initial model of Abdel-Hameed as a base for other applications and extensions.

II.3.1. Mathematics

The Gamma processes is illustrated in the underlying figure and can be defined as follows. A random deterioration quantity $X(t)$ has a Gamma distribution with shape parameter $\alpha > 0$ and scale parameter $\beta > 0$ if its probability density function is given by[3]:

$$Ga(x|\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp\{-\beta x\} \qquad (23)$$

Where $\Gamma(\alpha)= \int_{z=0}^{\infty} z^{\alpha-1} e^{-z} dz$ is the Gamma function for $a > 0$. As drawn in figure 46, we make a Gamma distribution for every time step Δt, which creates a series of Gamma distributions: a Gamma process. Time-increasing uncertainty and the behavior of deterioration can be incorporated in the shape parameter v. Then, let $\alpha(t)$ be a non-decreasing, right-continuous, real-valued function for $t \geq$

---

[3] In literature, the Gamma distribution is also frequently denoted as follows:

$$Ga(x|k,\theta) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} \exp\left(-\frac{x}{\theta}\right)$$

The notation of the shape- and scale parameters k, θ, respectively, deviates from the paramters mentioned in equation (5): $k = \alpha$ and $\theta = 1/\beta$. MATLAB® uses the notation above.

0,with $\alpha(0) \equiv 0$. The Gamma process with shape function $\alpha(t) > 0$ and scale parameter $\beta > 0$ is a continuous-time stochastic process with the following properties (van Noortwijk, 2009):

(1) $X(0) = 0$ with probability one:
(2) $X(\tau) - X(t) \sim Ga(v(\tau) - v(t), u)$ for all $\tau > t \geq 0$;
(3) $X(t)$ has independent increments.

The shape parameter $\alpha(t)$ is defined as the expected maximum deterioration at time t. In most cases, this function behaves according the following power law:

$$\alpha(t) = at^b \tag{24}$$

With physical constants $a > 0$ and $0 < b < 1$. Thereby, we also incorporate the uncertainty parameter ψ. This parameter represents the uncertainty in the deterioration process: the larger the ψ, the more uncertain the deterioration process. Now, denote $X(t)$ as the maximum deterioration at time t, and substitute the parameter θ and equation (24) into equation (23), the probability density function of $X(t)$ be given via the underlying equation:

$$p_{X(t)}(x) = Ga(x|[at^b]/\psi, 1/\psi) \tag{25}$$
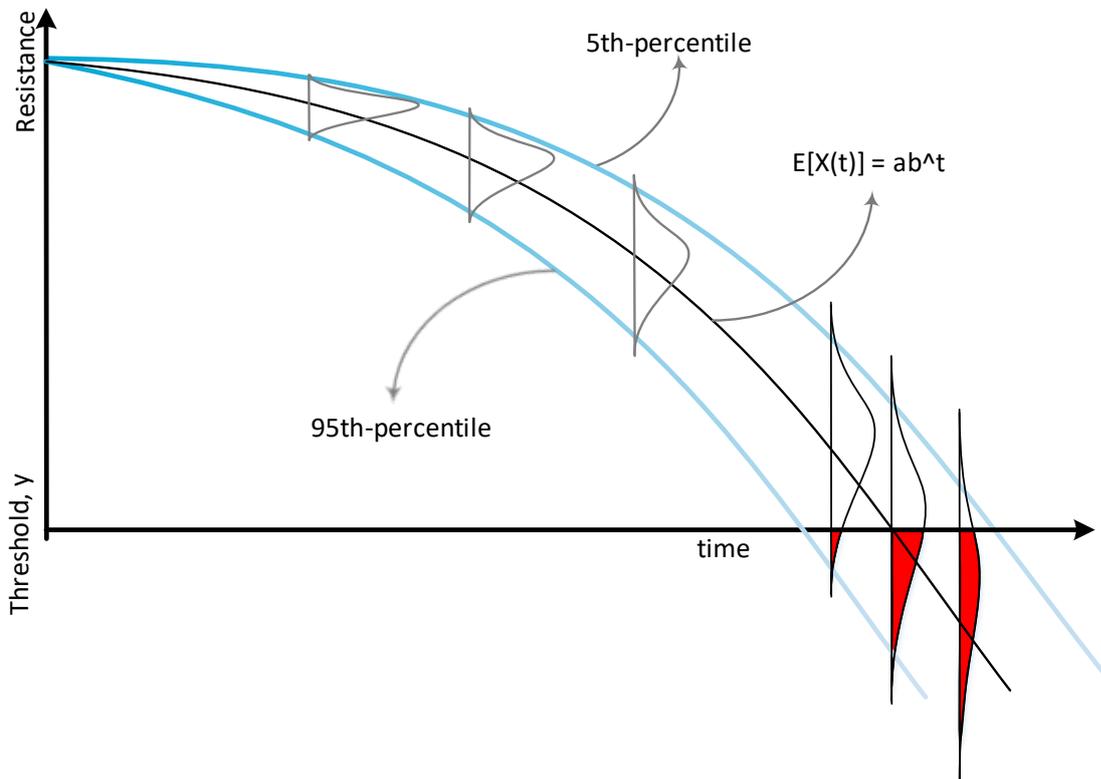
Figure 46: Gamma process

As illustrated in figure 46, the component fails when it crosses as certain threshold, say y. Let the time at which failure occurs, i.e. at which the failure level is crossed, be denoted as the life time L. Owing the Gamma distributed maximum deterioration of equation (25), the (cumulative) time-to-failure distribution can be written as:

$$F(t) = P(L \leq t) = P(X(t) \geq y) = \int_{x=y}^{\infty} p_{X(t)}(x)dx \tag{26}$$

$$= \frac{\Gamma([at^b]/\theta, y/\theta)}{\Gamma([at^b]/\theta)}$$

Where $\Gamma(\alpha, x) = \int_{t=x}^{\infty} t^{\alpha-1} e^t dt$ is the incomplete Gamma function for $x \geq 0$ and $a > 0$. The probability density function can logically be calculated by determining the derivative of equation (8). However, a more appealing approach is to determine the probability that $X(t) = y$ for each time step:

$$P(X(t) = y) = Ga(x|[at^b]/\psi, y/\psi) \tag{27}$$

After all, the underlying figures show representative pdf and cdf after the application of the Gamma process.
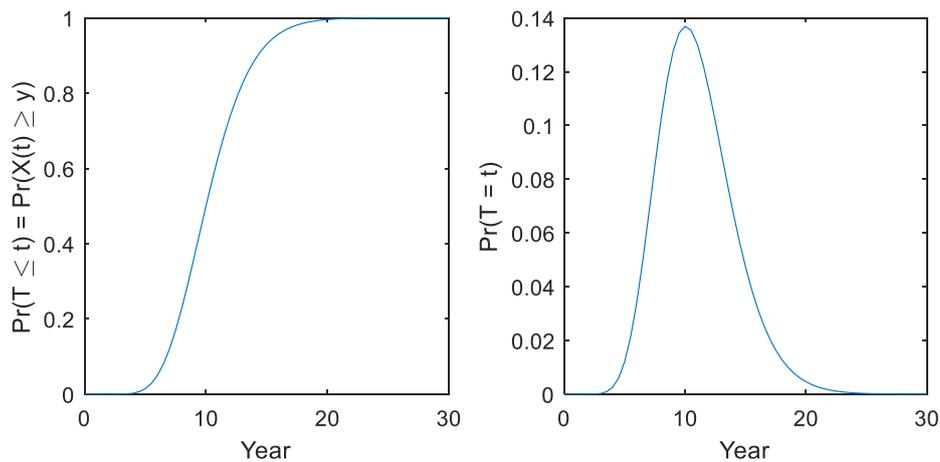
Figure 47: Output Gamma process

II.3.2. Input parameters

The challenge to model the Gamma process is to determine the input parameters a, b and ψ. van Noortwijk (2009) proposes three models based on data which should represent the cumulative deterioration at several inspection-moments; method of maximum likelihood, method of moments, method of Bayesian statistics. In absence of representative data, expert judgement can be used. For more in-depth information we refer to the mentioned source.

APPENDIX III - 'CLASSIC' BAYESIAN NETWORKS

III.1. INTRODUCTION

This appendix elaborates on the concept of Bayesian Networks. A Bayesian Network (BN), also called Bayesian Belief Networks, is a model that provides an elegant way of expressing joint distributions of a large number of interrelated variables. According to Pearl (1988), the initiator of BNs, a BN consists of two parts: a qualitative and a quantitative part. The qualitative part is the graphical part of the BN which consists of a set of nodes and a set of arcs, also known as an directed acyclic graph (DAG). The quantitative part of the BN contains the marginal distributions per node, conditioned on the parental nodes. The marginal distributions can be taken from data directly or that may be elicited from experts, for example via Cooke's method (Cooke, 1991). Applications of Pearl's theory, predominantly in the medical domain, demonstrate one of the main strength of Bayesian Networks; their ability to cope with large numbers of variables.

III.2. BAYESIAN REASONING

Three decades ago, a prime challenge in artificial intelligence research was to program machines to associate a potential cause to a set of observable conditions. Pearl figured out how to do it using Bayesian networks. In Hartnett (2018), Judea Pearl state the following:

> 'The language of algebra is symmetric: If X tells us about Y, then Y tells us about X. There's no way to write in mathematics a simple fact – for example, that the upcoming storm causes the barometer to go down, and not the other way around. Mathematics has not developed the asymmetric language required to capture our understanding that if X causes Y that does not mean that Y causes X.

The combination of the directed acyclic graph (DAG) and conditional independence statements, or semantics, provide the Bayesian network of this asymmetric property.

III.2.1. Directed Acyclic Graphs

Briefly, the qualitative part of BNs consist of a set of variables (nodes) and a set of directed edges (arcs) between variables. The only constraint on the arcs allowed in a BN is that there must not be any directed cycles. The figure below shows a BN model for five random variables $X = \{X_1, \ldots, X_5\}$. Note that random variable X_4 is attached to X_1 and X_2 via two arcs. According BN terminology, X_4 is called a 'child' node of 'parent'-nodes X_1 and X_2.
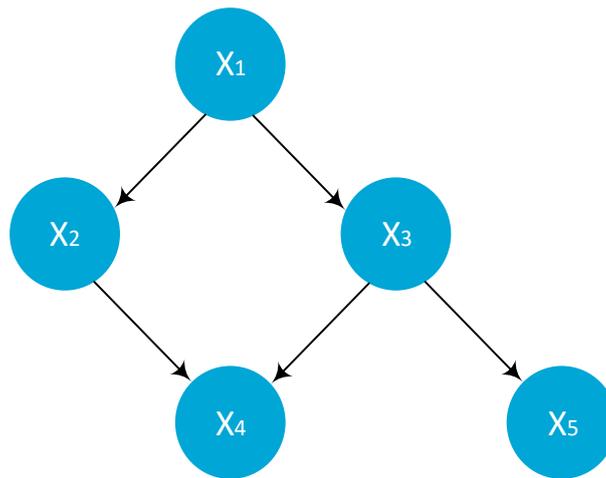
Figure 48: An example of a Directed Acyclic Graph (DAG)

III.2.2. Semantics

The direction of the arrows add information regarding to the conditional (in)dependencies within the graph or more simplified: it translates the reasoning behind the variables into conditional dependencies. This encoded information is also known as d-separation or conditional independence structures. We can basically distinguish three structures:

① → ② → ③   The first structure states that without observing 2, observing 1 would say something about the distribution of 3, or in probabilistic language; $X_1 \not\perp X_3$. That is 1 is not marginally independent of 3. However, if 2 is known, then 1 would not add extra information to explain 3, that is 1 and 3 are conditionally independent, given 2: $X_1 \perp X_3|X_2$.

① ← ② → ③   The second structure is similar as the first case: $X_1 \not\perp X_3$, but $X_1 \perp X_3|X_2$.

① → ② ← ③   The third structure deviates from the latter cases. Here, 1 and 3 are marginally independent $X_1 \perp X_3$, but not conditionally independent when 2 is observed. That is, if we observe 1(3), without observing 2 that would say nothing about 3(1), respectively. In contrast, if we observe 2, then observing 1(3) will say something additional about the distribution of 3(1), so $X_1 \not\perp X_3|X_2$.

III.3. CONDITIONAL PROBABILITY FUNCTIONS

As mentioned in the introduction of this appendix, the quantitative part consists of the marginal distributions per node, conditioned on the parental nodes. Obviously, these conditional probabilities are built Bayes's theorem which describes the probability of an event, based on prior knowledge of conditions that are related to the event:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Here, $P(A|B)$ is the conditional probability: the likelihood that event A occurs, given that B occurs. $P(B)$ is the marginal probability of observing parental event B. Note that if events $P(B)$ and $P(A)$ are independent, i.e. $P(A \cap B) = P(A) \times P(B)$, then $P(A|B) = P(A)$. If we want to know the joint probability density function of multiple dependent variables, one uses the following notation that is based on conditional probabilities.

$$P(A_1, A_2, A_3, \dots, A_n) = \prod_{i=1}^{n} P(A_i | A_{i+1}, \dots, A_n)$$

Now, let's combine the conditional independence statements and the joint conditional probability function above. Then, a fundamental conclusion is that every variable is independent of its ancestors, given its parents. Therefore if every variable is associated with a conditional probability function of the variable given its parents $P(A_i | Pa(A_i))$ then the joint probability may be written as:

$$P(A_1, A_2, A_3, \dots, A_n) = \prod_{i=1}^{n} P(A_i | Pa(A_i)) \tag{28}$$

If a node has no parents, then $P(A_i | Pa(A_i)) = P(A_i)$. For example, In figure 48, node $\{X_1\}$ have no parents, these nodes are solely defined by their marginal distributions. The other 'child'-nodes' marginal distributions, $\{X_2, X_3, X_4, X_5\}$ must be conditioned on the accompanied parental nodes.
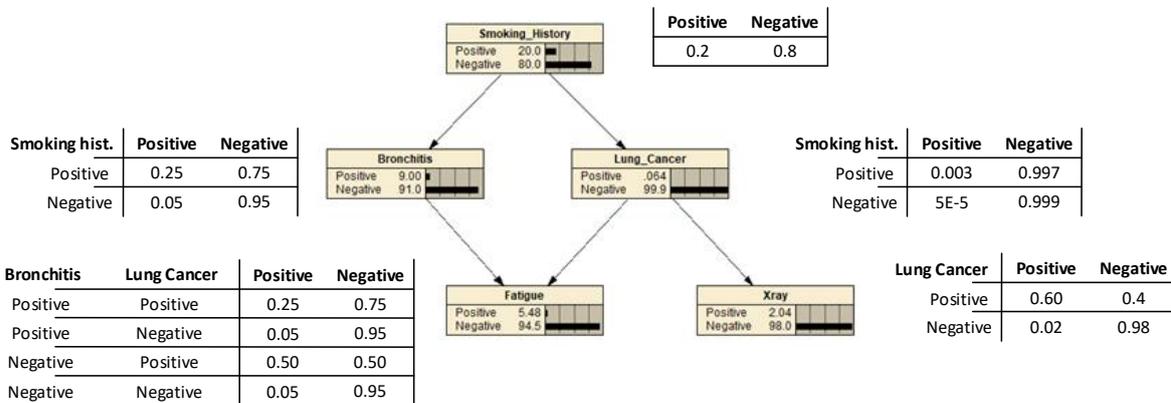
Figure 49: An example of an Bayesian Network with probability tables

Let's assume the BN above, similar as figure 48, with discrete probabilities. According equation (28), the joint probability can be defined as follows:

$$P(H, B, L, F, X) = P(H)P(B|H)P(L|H)P(F|B,L)P(X|L)$$

III.4. INFERENCE

The probabilities are then given by the conditional probabilities in the associated tables. Based on the joint probability distribution we are able to calculate the conditional probabilities given that a certain variable is observed via, again, the conditional probability function. This is called inference. If one is interested in the probability that the patient has bronchitis, given that the X-ray is negative and the patient has a smoking history:

$$P(B|H^+, X^-) = \frac{\sum_{L,F} P(H^+, B, L, F, X^-)}{\sum_{B,L,F} P(H^+, B, L, F, X^-)}$$

APPENDIX IV - BIVARIATE COPULAE

In reliability analysis theory, one often utilizes the bivariate distributions of two random variables, for example the resistance R and solicitation S. As mentioned, in some cases those variables can be treated as independent which simplifies calculations. When variables are dependent, not enough information is provided by the separate one-dimensional marginal distributions to shape the two-dimensional probability density function. Therefore, modern statistics shift more and more to the study of copulas. Their importance has clearly grown in the last ten years. Despite their growth in popularity, copulas are not too familiar year. Here we will briefly sketch the general concept of copulae.

IV.1. SKLAR'S THEOREM

Roughly speaking, copulas separate the effect of dependence from the effect of marginal distributions in a joint distribution (Jonkman, Steenbergen, Morales-Nápoles, Vrouwenvelder, & Vrijling, 2015). Following the copula approach, it is possible to construct the joint distribution requiring only the marginal distributions of the variables and measures of their dependence (Clemen & Reilly, 1999).The concept of copula is based on Sklar's theorem which states that any multivariate joint distribution can be written in terms of the univariate marginal distribution functions $\{F_x(x), F_y(y)\}$ and a copula C which describes the dependence between the random variables (Sklar, 1959):

$$F_{xy}(x,y) = C\{F_x(x), F_y(y)\} \quad x, y \in \mathbb{R} \tag{29}$$

IV.2. COPULA FAMILIES

There is a large variety of copula types of which the Gaussian-, Clayton- and Gumbel copula are the most familiar one-parametric copulae. First, the Gaussian copula is given by:

$$C_\rho(u,v) = \phi_\rho\big(\phi^{-1}(u), \phi^{-1}(v)\big)$$

Where ϕ_ρ denotes the bivariate standard normal cumulative distribution function with product moment correlation ρ and ϕ^{-1} is the inverse of the univariate standard normal distribution function. The Clayton and Gumbel copulas are two of the most used one-parameter Archimedean copulas. The Clayton copula, parameterized by θ, is given by:

$$C_\theta(u,v) = \exp\left\{-\big([-\log(u)]^\theta + [-\log(v)]^\theta\big)^{1/\theta}\right\}, \quad \theta \geq 1$$

And the Gumbel copula, parametrized by β, is defined as:

$$C_\beta(u,v) = \big(u^{-\beta} + v^{-\beta} - 1\big)^{-\frac{1}{\beta}}, \quad \beta \in [-1, \infty)$$

Those three copula types present an important aspect of joint distributions, known as tail dependence. Joe (2014) introduced an upper tail dependence coefficient λ_U that measures the magnitude of tail dependence. A value $\lambda_U > 0$ indicates that it is likely to observe values of U greater than u given that V is greater than u for u arbitrarily close to 1. Via this coefficient, Joe (2014) showed that the Gaussian copula presents no tail dependence ($\lambda_U = 0$), the Clayton copula presents lower tail dependence ($\lambda_U = 2^{-\frac{1}{\theta}}$) and the Gumbel copula presents upper tail dependence ($\lambda_U = 2 - 2^{\frac{1}{\beta}}$).

In the figure below, you can find the separate copula-types, with its accompanied randomly sampled scatter plots. In the scatter plots one sees the (tail)dependence structure between the two marginal U and V.
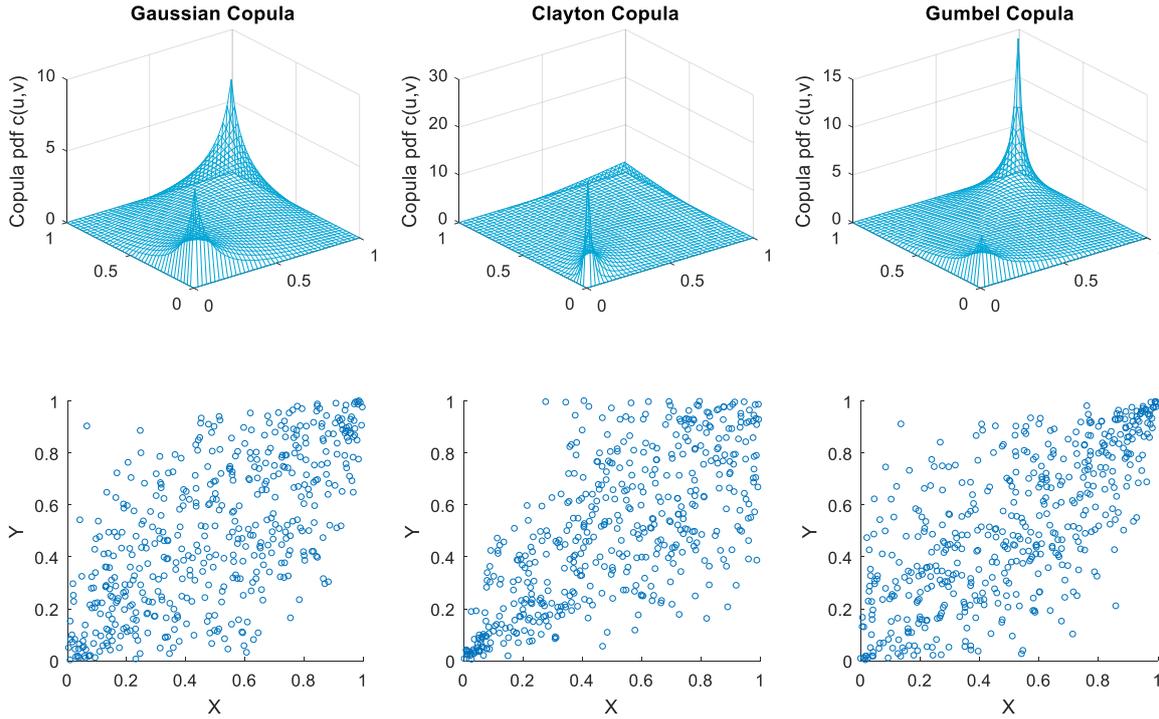
Figure 50: Most frequently used copulas and its accompanied scatterplots.

IV.3. COPULA VALIDATION

Since the benefits and theory of around the study of copulas are familiar now, the focus shifts to the practical side now, namely fitting a copula to data. Many proposals have been made for goodness-of-fit testing of copula models. Joe (2014) suggested an approach based on semi-correlations and Genest, Rémillard, & Beaudoin (2009) propose a set of statistical 'blanket tests' of which the Cramèr-von Mises statistic test will be utilized here.

IV.3.1. Semi-Correlations

According to Joe (2014), semi-correlations can be defined as the Pearson's product moment correlation coefficients computed in the upper and lower quadrants of the normal transforms of the original variables. First, the original variables (X_1, X_2) need to be transformed to standard normals (Z_1, Z_2), via:

$$Z(X_{i,j}) = \phi^{-1}(F(X_{i,j})) \tag{30}$$

Where i is the number of the considered variable and $j = 0, 1, \dots, n$, with n is the number of observations. For clarification, first one should calculate the ranks per value via the empirical cumulative distribution function: $F(X_1)$ and $F(X_2)$. Then every observation $X_{i,j}$ has an associated rank-value $F(X_{i,j}), [0,1]$. The ranks per variable $U \sim F(X_i)$ is uniformly distributed. Plugging the uniformly distributed rank values of X_i into the inverse standard normal distribution will give the transformed standard normal values $Z_{i,j}$. For positive correlation, semi-correlations in the upper right (ρ_{NE}) and lower left (ρ_{SW}) quadrants are:

$$\rho_{NE} = \rho(Z_1, Z_2 | Z_1 > 0, Z_2 > 0) \tag{31}$$

$$\rho_{SW} = \rho(Z_1, Z_2 | Z_1 < 0, Z_2 < 0) \tag{32}$$

For negative correlations, semi-correlations in the upper left (ρ_{NW}) and lower right (ρ_{SE}) are defined similarly. In general, larger absolute values of the semi-correlations than the 'overall' correlations indicate tail dependence.
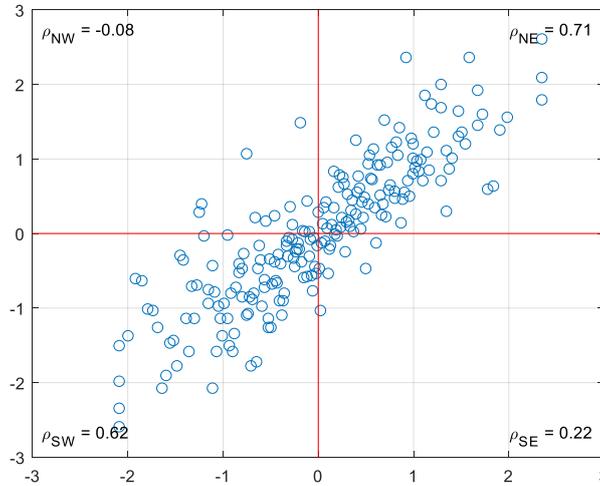
Figure 51: Semi-correlations per quadrant

IV.3.2. Blanket test

The 'blanket' test is discussed in Genest et al. (2009) and is based on Cramèr-von Mises statistics. Cramèr-von Mises statistics is used for judging goodness-of-fit of a cumulative distribution function compared to a given empirical distribution function. Translating the this application to the copula case: the idea behind this test is to measure the difference between the cumulative empirical copula and the cumulative theoretical copula. In essence the smaller the calculated differences are, the better the theoretical copula fits. The difference between the copula is measured according the 'sum of the squared differences'-principle, stated below:

$$M_n(\boldsymbol{u}) = \sum_{|\boldsymbol{u}|}\left\{C_{\hat{\theta}_n}(\boldsymbol{u}) - C_n(\boldsymbol{u})\right\}^2, \boldsymbol{u} \in [0,1]^2 \tag{33}$$

Where, $C_{\hat{\theta}_n}(\boldsymbol{u})$ is the theoretical copula and $C_n(\boldsymbol{u})$ is the empirical copula. The empirical copula is based on pseudo-observations $\boldsymbol{U}_1, \dots, \boldsymbol{U}_n$. Pseudo-observations can simply be deduced from the ranks, $U_{ij} = R_{ij}/(n+1) = nF(X_{i,j})/(n+1)$ and can be interpreted as a sample of the underlying copula. Then, the empirical copula is simply the bivariate cumulative distribution of pseudo-observations $(\boldsymbol{U}_1, \boldsymbol{U}_2)$:

$$C_n(\boldsymbol{u}) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{1}(U_{i1} \leq u_1, \dots, U_{ij} \leq u_j) \tag{34}$$

$$\boldsymbol{u} = (u_1, \dots, u_j) \in [0,1]^2$$

In this research we will only treat the one-parameter theoretical copulas. Figure 52 shows an example of the empirical copula and the cdf's of the three treated theoretical copulae. Ostensibly, the differences between the theoretical copulae are very small, but large enough to designate the best fit. Where figure 51 also shows a magnitude of tail dependence with $\rho_{NE} = 0.71$, the best-fit Gumbel copula also designate upper tail dependence.
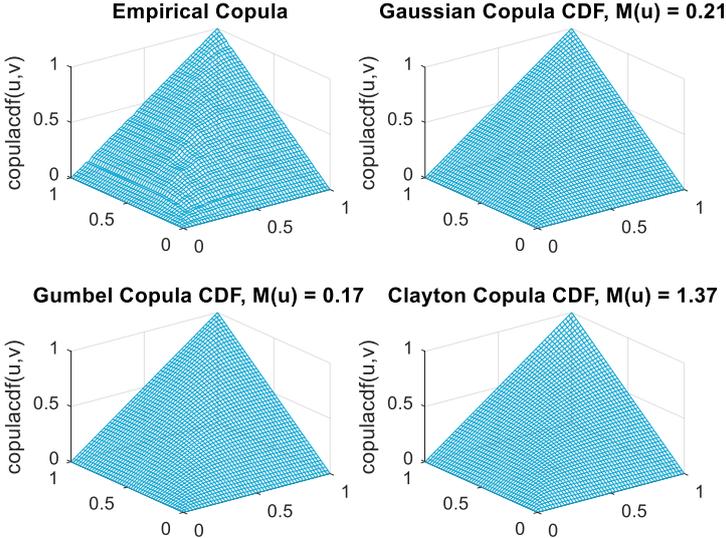
Figure 52: Example Blanket test

APPENDIX V – SYSTEM DECOMPOSITION & DATA

This appendix elaborates on essential case characteristics such as the system decomposition and the background of the retrieved data.

V.1. PUMP DECOMPOSITION

The underlying figures show the simplified the pump-decomposition. The green nodes contain sub-elements but for simplicity reasons those are the omitted.

V.1.1. Stork Pump
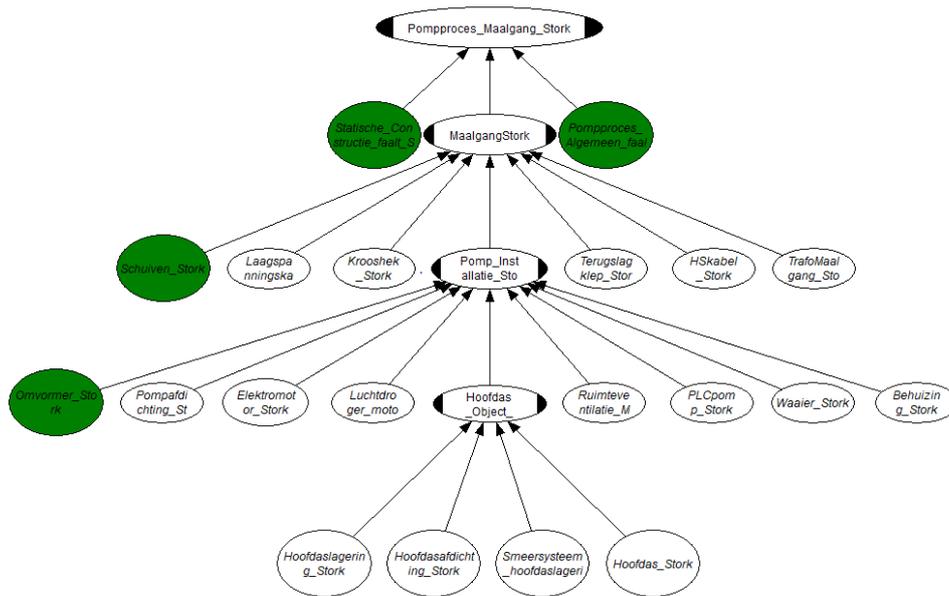
Figure 53: Pump decomposition Stork-pump

V.1.2. Nijhuis Pump
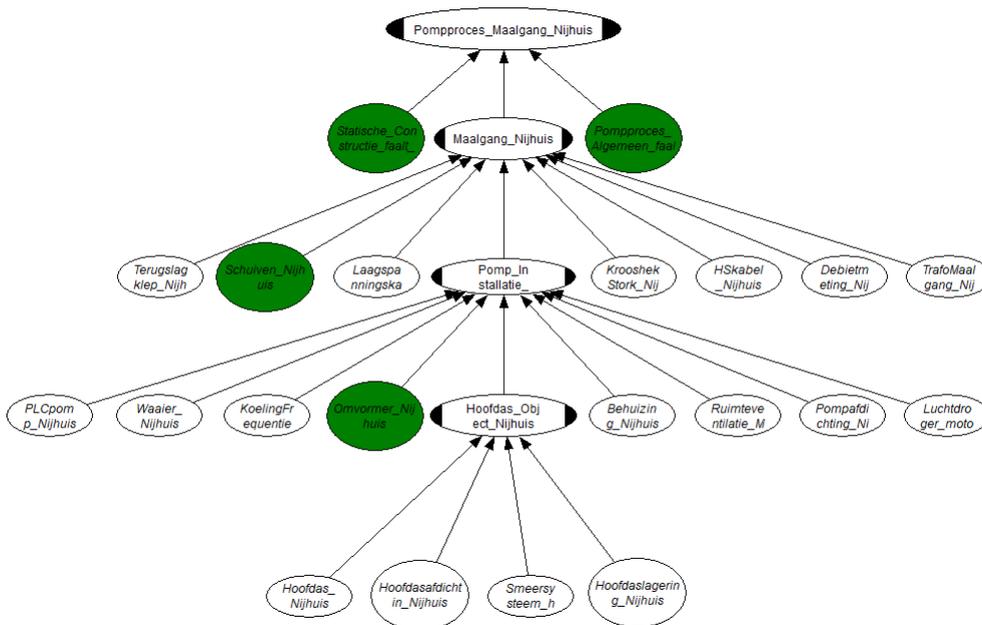
Figure 54: Pump decomposition Nijhuis pump

V.2. BACKGROUND DATA

Since the NPBN builds on the input of marginal distributions and rank correlation coefficients, see chapter 6, the first step to build the NPBN is to analyze the environmental data. Every node in the network is attached to a certain data set to retrieve the marginal distribution. Unfortunately, we are dedicated to four-year dataset; from 2013 to 2016. The underlying reasons can be read in the following paragraph's where the individual datasets are described.

V.2.1. Data Discharge Complex IJmuiden

The discharge-data of the IJmuiden complex is retrieved from the Rijkswaterstaat data-base on http://live.waterbase.nl. Here, we can retrieve historical discharge data until 21-7-2017. The retrieved data considers discharge measurements with an interval of 24h, measured at 08:00 in m^3/s-units.

In order to make the data suitable for the NPBN, we limited the data set until 31-12-2016 since we are interested in full year data. This date sets the upper limit for all following data. Thereby the lower limit is set on 1-1-2013. The dataset captures the free discharges as well as the discharge by pumps. Thereby, since the maximum pump capacity of the complex is 260 m^3/s, all values above can be assumed to be freely discharged. These values are cancelled out of the dataset.

V.2.2. Data Discharges Amsterdam Rhine Channel

The original discharge-data of the ARC-water system is retrieved from Waternet which is the governmental agency that regulates the drinking water supply, sewage systems and water management within the water board of 'Amstel, Gooi en Vecht'. For this research we received data from 17-5-2012 until 9-2-2018. The retrieved data considers discharge measurements with an interval of 1h, measured in m^3/h-units.

In order to make the data suitable for the NPBN, we limited the data set until 31-12-2016. The lower limit is set on 1-1-2013, since we are interested in full years. This date is the under limit of all following data sets. Thereby, it will take time to transport the discharged water that is discharged by the considered pump station to get actually pumped by the discharge complex of IJmuiden. In order to catch this delay, we transformed the data from hourly discharges to cumulative discharges per day. By this, we capture the time it takes that the water will actually be pumped away by the discharge complex of IJmuiden, assuming that it will take one day.

V.2.3. Data Discharges North Sea Channel

The entities that are responsible for the discharges into the North Sea Channel are the water boards of Rijnland and Hollands Noorderkwartier. The water board of Rijnland is responsible for the discharges on the south side of the NSC and the water board of Hollands Noorderkwartier is responsible for the discharges north of the NSC. The data for the discharges is gather via these entities. Again, the data of the Hollands Noorderkwartier contains lots of unreliable measurements and is not considered in the NPBN. The discharge data of Rjjnland also contains errors for three of the five discharge complexes, only the complexes of Halfweg and Spaarndam are reliable. This enables us to calculate the cumulative discharges of all discharge complexes and transform these values, together with the two discharge complexes that contain no errors, to daily values. In this way we capture the time it takes that the water will actually be pumped away by the discharge complex of IJmuiden, assuming that it will take less than one day.

V.3.4. Data Rainfall

The daily cumulative rainfall-data has been registered by the Royal Dutch Meteorological Institute (KNMI) since 1951. Throughout the Netherlands, the KNMI has 325 precipitation stations that measure

the daily rainfall between 08:00 and 08:00 the next day. For each station, the rainfall can be retrieved from https://www.knmi.nl/nederland-nu/klimatologie/monv/reeksen.

For the purpose of this research, the rainfall data in the water system is divided into two locations; Loenen aan de Vecht and Heemstede. The first station must represents the ARC-region and the latter is considered representative for the NSC southern-region since we only work with the discharge data on the south-side of the NSC.

APPENDIX VI – ESTIMATION OF UNKNOWN VARIABLES

VI.1. DISTRIBUTION FIT SEA LEVEL RISE

Year	5th	10th	25th	50th	75th	90th	95th	μ	σ
2017	6.10	6.20	6.47	7.04	8.14	10.00	11.68	2.039	0.190
2018	6.93	7.07	7.43	8.19	9.65	12.13	14.38	2.201	0.217
2019	7.75	7.93	8.39	9.38	11.25	14.44	17.33	2.346	0.242
2020	8.56	8.79	9.36	10.60	12.94	16.92	20.53	2.477	0.266
2021	9.37	9.64	10.34	11.85	14.72	19.58	24.00	2.597	0.290
2022	10.16	10.49	11.33	13.14	16.58	22.42	27.72	2.708	0.312
2023	10.94	11.33	12.33	14.47	18.53	25.43	31.69	2.812	0.334
2024	11.72	12.17	13.33	15.83	20.57	28.62	35.92	2.908	0.355
2025	12.48	13.01	14.34	17.22	22.70	31.98	40.41	2.999	0.375
2026	13.24	13.84	15.37	18.65	24.91	35.52	45.16	3.085	0.394
2027	13.98	14.66	16.39	20.12	27.22	39.24	50.16	3.166	0.413
2028	14.72	15.48	17.43	21.62	29.61	43.13	55.41	3.242	0.431
2029	15.44	16.30	18.48	23.16	32.08	47.20	60.93	3.316	0.448
2030	16.16	17.11	19.53	24.74	34.65	51.44	66.70	3.385	0.465
2031	16.86	17.91	20.59	26.35	37.30	55.86	72.72	3.452	0.481
2032	17.56	18.72	21.66	27.99	40.04	60.46	79.01	3.516	0.497
2033	18.25	19.51	22.74	29.67	42.87	65.23	85.54	3.578	0.513
2034	18.92	20.30	23.82	31.39	45.78	70.18	92.34	3.637	0.528
2035	19.59	21.09	24.91	33.14	48.79	75.30	99.39	3.694	0.542
2036	20.25	21.87	26.01	34.92	51.88	80.60	106.70	3.749	0.556
2037	20.90	22.65	27.12	36.74	55.05	86.08	114.26	3.802	0.570
2038	21.53	23.43	28.24	38.60	58.32	91.73	122.08	3.854	0.584
2039	22.16	24.19	29.37	40.49	61.67	97.56	130.16	3.903	0.597
2040	22.78	24.96	30.50	42.42	65.11	103.56	138.49	3.951	0.610
2041	23.39	25.72	31.64	44.39	68.64	109.75	147.08	3.998	0.622
2042	23.99	26.47	32.79	46.39	72.26	116.10	155.93	4.044	0.635
2043	24.58	27.22	33.95	48.42	75.96	122.63	165.03	4.088	0.647
2044	25.16	27.97	35.11	50.49	79.75	129.34	174.39	4.131	0.659
2045	25.73	28.71	36.29	52.60	83.63	136.23	184.00	4.173	0.670
2046	26.29	29.44	37.47	54.74	87.60	143.29	193.87	4.213	0.681
2047	26.84	30.17	38.66	56.91	91.65	150.52	204.00	4.253	0.692
2048	27.38	30.90	39.86	59.12	95.80	157.94	214.38	4.292	0.703
2049	27.91	31.62	41.06	61.37	100.03	165.52	225.02	4.330	0.714
2050	28.44	32.34	42.28	63.65	104.34	173.29	235.92	4.367	0.724
2051	28.95	33.05	43.50	65.97	108.75	181.23	247.07	4.403	0.735
2052	29.45	33.76	44.73	68.33	113.24	189.35	258.48	4.438	0.745
2053	29.94	34.46	45.97	70.71	117.82	197.64	270.14	4.472	0.755
2054	30.43	35.16	47.21	73.14	122.49	206.11	282.06	4.506	0.764
2055	30.90	35.85	48.47	75.60	127.24	214.75	294.24	4.539	0.774
2056	31.36	36.54	49.73	78.09	132.09	223.57	306.67	4.572	0.783
2057	31.82	37.23	51.00	80.63	137.02	232.57	319.36	4.603	0.793

2058	32.26	37.91	52.28	83.19	142.03	241.74	332.31	4.635	0.802
2059	32.70	38.58	53.56	85.79	147.14	251.09	345.51	4.665	0.811
2060	33.12	39.25	54.86	88.43	152.33	260.61	358.97	4.695	0.819
2061	33.54	39.92	56.16	91.10	157.61	270.31	372.69	4.724	0.828
2062	33.94	40.58	57.47	93.81	162.98	280.19	386.66	4.753	0.837
2063	34.34	41.23	58.79	96.56	168.44	290.24	400.89	4.782	0.845
2064	34.72	41.88	60.11	99.33	173.98	300.47	415.37	4.810	0.853
2065	35.10	42.53	61.45	102.15	179.62	310.88	430.11	4.837	0.862
2066	35.47	43.17	62.79	105.00	185.33	321.46	445.11	4.864	0.870
2067	35.82	43.81	64.14	107.88	191.14	332.21	460.36	4.890	0.878
2068	36.17	44.44	65.50	110.81	197.04	343.15	475.87	4.916	0.886
2069	36.51	45.07	66.87	113.76	203.02	354.26	491.64	4.942	0.893
2070	36.84	45.69	68.24	116.75	209.09	365.54	507.66	4.967	0.901
2071	37.16	46.31	69.62	119.78	215.24	377.00	523.94	4.992	0.908
2072	37.46	46.93	71.02	122.84	221.49	388.64	540.47	5.016	0.916
2073	37.76	47.54	72.41	125.94	227.82	400.45	557.26	5.040	0.923
2074	38.05	48.14	73.82	129.08	234.24	412.44	574.31	5.064	0.931
2075	38.33	48.74	75.24	132.24	240.75	424.60	591.61	5.087	0.938
2076	38.60	49.33	76.66	135.45	247.34	436.94	609.17	5.110	0.945
2077	38.86	49.92	78.09	138.69	254.03	449.46	626.99	5.133	0.952
2078	39.11	50.51	79.53	141.96	260.80	462.15	645.06	5.156	0.959
2079	39.35	51.09	80.98	145.28	267.66	475.02	663.39	5.178	0.966
2080	39.58	51.67	82.43	148.62	274.60	488.07	681.97	5.199	0.972
2081	39.80	52.24	83.89	152.00	281.63	501.29	700.81	5.221	0.979
2082	40.02	52.81	85.37	155.42	288.76	514.69	719.91	5.242	0.986
2083	40.22	53.37	86.85	158.87	295.96	528.26	739.27	5.263	0.992
2084	40.41	53.93	88.33	162.36	303.26	542.01	758.88	5.284	0.999
2085	40.59	54.48	89.83	165.88	310.64	555.93	778.74	5.304	1.005
2086	40.76	55.03	91.33	169.44	318.12	570.03	798.86	5.324	1.011
2087	40.93	55.57	92.84	173.04	325.67	584.31	819.24	5.344	1.018
2088	41.08	56.11	94.36	176.67	333.32	598.76	839.88	5.364	1.024
2089	41.22	56.64	95.89	180.33	341.06	613.39	860.77	5.383	1.030
2090	41.36	57.17	97.43	184.03	348.88	628.20	881.92	5.402	1.036
2091	41.48	57.70	98.97	187.77	356.79	643.18	903.32	5.421	1.042
2092	41.60	58.22	100.52	191.54	364.78	658.33	924.98	5.440	1.048
2093	41.70	58.73	102.08	195.35	372.87	673.67	946.90	5.458	1.054
2094	41.80	59.24	103.65	199.19	381.04	689.18	969.07	5.477	1.060
2095	41.88	59.75	105.22	203.07	389.30	704.86	991.50	5.495	1.065
2096	41.96	60.25	106.81	206.98	397.65	720.72	1014.19	5.513	1.071
2097	42.03	60.74	108.40	210.93	406.08	736.76	1037.13	5.530	1.077
2098	42.08	61.24	110.00	214.92	414.61	752.97	1060.33	5.548	1.082
2099	42.13	61.72	111.61	218.94	423.22	769.36	1083.78	5.565	1.088
2100	42.17	62.21	113.22	222.99	431.92	785.93	1107.50	5.582	1.093

VI.2. DISTRIBUTION FIT OPERATION TIME & TIME-TO-FAILURE DISTRIBUTIONS

	Stork				Nijhuis			
	Operational Time		Time-to-failure distribution		Operational time		Time-to-failure distribution	
Year	μ	σ	α	β	μ	σ	α	β
2017	7.438	0.071	48.296	59.935	7.783	0.051	58.294	72.920
2018	7.445	0.078	48.258	58.028	7.791	0.058	58.270	70.052
2019	7.453	0.085	48.221	56.076	7.799	0.064	58.249	67.020
2020	7.461	0.090	48.186	54.023	7.807	0.069	58.234	63.552
2021	7.468	0.095	48.155	51.755	7.815	0.074	58.231	59.000
2022	7.475	0.100	48.130	49.066	7.822	0.079	58.238	54.427
2023	7.482	0.104	48.114	46.053	7.829	0.083	58.230	51.764
2024	7.489	0.107	48.094	43.757	7.837	0.087	58.213	49.920
2025	7.496	0.111	48.066	42.147	7.844	0.090	58.192	48.449
2026	7.503	0.114	48.034	40.894	7.851	0.094	58.169	47.191
2027	7.510	0.117	47.999	39.837	7.858	0.097	58.145	46.074
2028	7.516	0.119	47.963	38.906	7.865	0.100	58.119	45.066
2029	7.523	0.122	47.926	38.067	7.872	0.104	58.093	44.138
2030	7.529	0.124	47.888	37.296	7.879	0.107	58.066	43.275
2031	7.535	0.126	47.850	36.570	7.885	0.110	58.039	42.465
2032	7.542	0.129	47.811	35.901	7.892	0.113	58.012	41.699
2033	7.548	0.131	47.771	35.276	7.899	0.116	57.984	40.973
2034	7.554	0.134	47.731	34.675	7.906	0.120	57.955	40.290
2035	7.561	0.136	47.691	34.102	7.912	0.123	57.927	39.634
2036	7.567	0.138	47.651	33.556	7.919	0.126	57.899	39.006
2037	7.573	0.141	47.611	33.030	7.926	0.129	57.870	38.400
2038	7.579	0.143	47.570	32.529	7.932	0.132	57.841	37.826
2039	7.585	0.145	47.530	32.046	7.939	0.136	57.812	37.268
2040	7.592	0.148	47.489	31.580	7.945	0.139	57.783	36.723
2041	7.598	0.150	47.448	31.127	7.952	0.142	57.753	36.214
2042	7.604	0.153	47.407	30.694	7.958	0.145	57.724	35.714
2043	7.610	0.155	47.366	30.273	7.964	0.148	57.695	35.230
2044	7.616	0.157	47.324	29.862	7.971	0.152	57.665	34.761
2045	7.622	0.160	47.283	29.469	7.977	0.155	57.636	34.306
2046	7.627	0.162	47.242	29.083	7.983	0.158	57.606	33.866
2047	7.633	0.165	47.200	28.711	7.990	0.161	57.577	33.435
2048	7.639	0.167	47.159	28.348	7.996	0.164	57.547	33.022
2049	7.645	0.169	47.117	27.994	8.002	0.167	57.517	32.619
2050	7.651	0.172	47.076	27.648	8.008	0.171	57.487	32.226

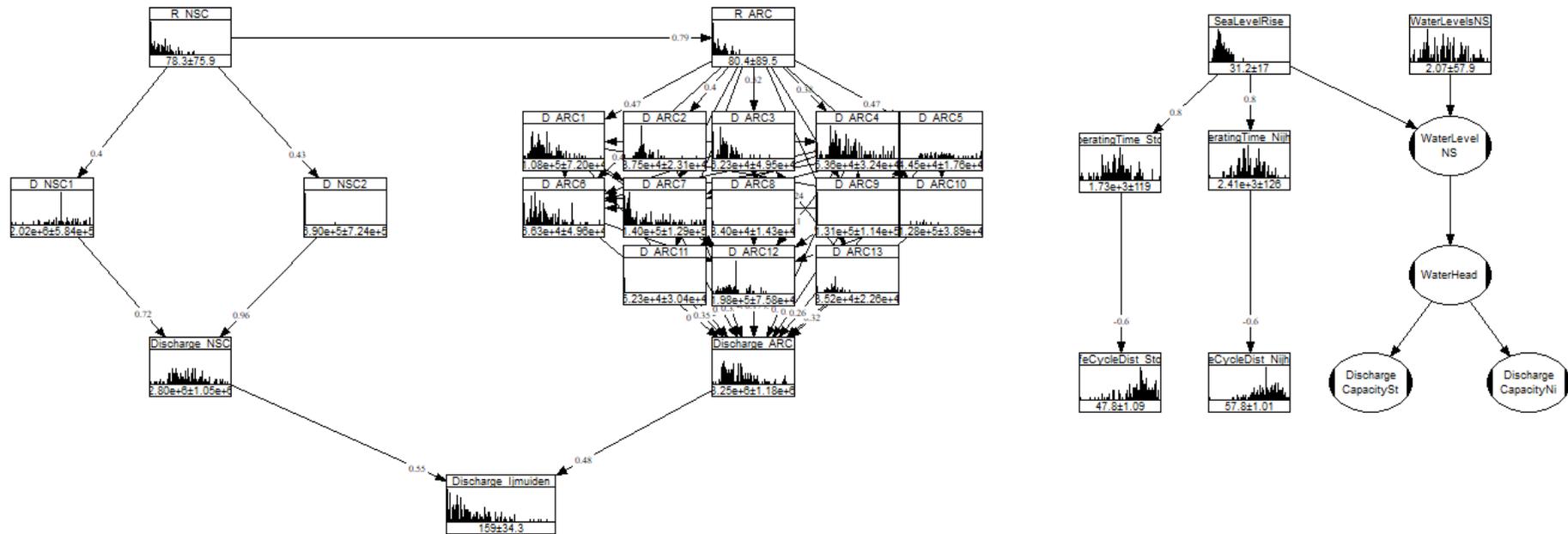# APPENDIX VII – NPBN FOR PUMPING STATION IJMUIDEN

VII.1. TOTAL NPBN

Figure 55: Total non-parametric Bayesian Network of pump station IJmuiden

VII.1. CORRELATION MATRICES

	Discharge_IJmuiden	Discharge_NSC	Discharge_ARC	R_NSC	R_ARC	D_NSC1	D_NSC2	D_ARC1	D_ARC2	D_ARC3	D_ARC4	D_ARC5	D_ARC6	D_ARC7	D_ARC8	D_ARC9	D_ARC10	D_ARC11	D_ARC12	D_ARC13
Discharge_IJmuiden	1.00	0.54	0.69	0.52	0.49	0.42	0.40	0.66	0.52	0.61	0.57	0.62	0.60	0.63	0.54	-0.02	0.43	0.56	0.38	0.47
Discharge_NSC	0.54	1.00	0.58	0.55	0.51	0.75	0.80	0.56	0.51	0.51	0.57	0.49	0.47	0.50	0.50	-0.12	0.39	0.52	0.25	0.49
Discharge_ARC	0.69	0.58	1.00	0.47	0.55	0.49	0.42	0.94	0.65	0.69	0.87	0.87	0.86	0.81	0.86	-0.10	0.66	0.80	0.51	0.55
R_NSC	0.52	0.55	0.47	1.00	0.82	0.43	0.42	0.43	0.45	0.57	0.37	0.45	0.43	0.42	0.33	-0.08	0.11	0.42	0.22	0.54
R_ARC	0.49	0.51	0.55	0.82	1.00	0.39	0.41	0.50	0.44	0.54	0.40	0.48	0.50	0.44	0.42	-0.11	0.22	0.53	0.26	0.54
D_NSC1	0.42	0.75	0.49	0.43	0.39	1.00	0.25	0.48	0.44	0.38	0.52	0.39	0.41	0.43	0.38	-0.12	0.35	0.42	0.11	0.35
D_NSC2	0.40	0.80	0.42	0.42	0.41	0.25	1.00	0.39	0.36	0.41	0.37	0.37	0.30	0.35	0.39	-0.05	0.26	0.37	0.25	0.41
D_ARC1	0.66	0.56	0.94	0.43	0.50	0.48	0.39	1.00	0.59	0.65	0.85	0.85	0.81	0.76	0.77	-0.07	0.59	0.76	0.44	0.45
D_ARC2	0.52	0.51	0.65	0.45	0.44	0.44	0.36	0.59	1.00	0.56	0.57	0.52	0.58	0.54	0.49	-0.06	0.38	0.51	0.32	0.55
D_ARC3	0.61	0.51	0.69	0.57	0.54	0.38	0.41	0.65	0.56	1.00	0.51	0.62	0.62	0.61	0.54	-0.03	0.29	0.53	0.26	0.51
D_ARC4	0.57	0.57	0.87	0.37	0.40	0.52	0.37	0.85	0.57	0.51	1.00	0.78	0.75	0.72	0.75	-0.13	0.62	0.68	0.40	0.45
D_ARC5	0.60	0.47	0.86	0.43	0.50	0.41	0.30	0.81	0.58	0.62	0.75	1.00	0.76	0.67	0.75	-0.11	0.56	0.73	0.45	0.43
D_ARC6	0.63	0.50	0.81	0.42	0.44	0.43	0.35	0.76	0.54	0.61	0.72	0.77	1.00	0.67	0.73	-0.02	0.45	0.55	0.39	0.50
D_ARC7	0.54	0.50	0.86	0.33	0.42	0.38	0.39	0.77	0.49	0.54	0.75	0.78	0.75	1.00	0.73	-0.09	0.53	0.69	0.51	0.46
D_ARC8	-0.02	-0.12	-0.10	-0.08	-0.11	-0.12	-0.05	-0.07	-0.06	-0.03	-0.13	-0.06	-0.11	-0.02	1.00	-0.09	-0.18	-0.20	-0.10	-0.13
D_ARC9	0.43	0.39	0.66	0.11	0.22	0.35	0.26	0.59	0.38	0.29	0.62	0.48	0.56	0.45	0.53	1.00	-0.18	0.51	0.38	0.22
D_ARC10	0.56	0.52	0.80	0.42	0.53	0.42	0.37	0.76	0.51	0.53	0.68	0.70	0.73	0.55	0.69	-0.20	1.00	0.51	0.35	0.42
D_ARC11	0.38	0.25	0.51	0.22	0.26	0.11	0.25	0.44	0.32	0.26	0.40	0.42	0.45	0.39	0.51	-0.10	0.38	1.00	0.35	0.20
D_ARC12	0.62	0.49	0.87	0.45	0.48	0.39	0.37	0.85	0.52	0.62	0.78	1.00	0.76	0.77	0.78	-0.06	0.48	0.70	0.42	0.47
D_ARC13	0.47	0.49	0.55	0.54	0.54	0.35	0.41	0.45	0.55	0.51	0.45	0.47	0.43	0.50	0.46	-0.13	0.22	0.42	0.20	1.00

	Discharge_IJmuiden	Discharge_NSC	Discharge_ARC	R_NSC	R_ARC	D_NSC1	D_NSC2	D_ARC1	D_ARC2	D_ARC3	D_ARC4	D_ARC5	D_ARC6	D_ARC7	D_ARC8	D_ARC9	D_ARC10	D_ARC11	D_ARC12	D_ARC13
Discharge_IJmuiden	1.00	0.55	0.66	0.50	0.48	0.42	0.42	0.62	0.48	0.56	0.52	0.56	0.61	0.50	0.00	0.42	0.54	0.38	0.59	0.43
Discharge_NSC	0.55	1.00	0.61	0.50	0.51	0.72	0.78	0.58	0.52	0.50	0.57	0.46	0.51	0.53	-0.10	0.40	0.53	0.25	0.50	0.49
Discharge_ARC	0.66	0.61	1.00	0.83	0.63	0.47	0.45	0.92	0.55	0.63	0.84	0.81	0.80	0.46	-0.06	0.65	0.78	0.51	0.85	0.52
R_NSC	0.50	0.50	0.83	1.00	0.47	0.34	0.42	0.73	0.41	0.49	0.71	0.69	0.70	0.31	-0.07	0.53	0.67	0.49	0.74	0.42
R_ARC	0.48	0.51	0.63	0.47	1.00	0.44	0.35	0.57	0.40	0.51	0.56	0.53	0.53	0.41	-0.02	0.37	0.50	0.30	0.49	0.49
D_NSC1	0.42	0.72	0.47	0.34	0.44	1.00	0.21	0.46	0.37	0.33	0.50	0.41	0.41	0.40	-0.10	0.35	0.40	0.10	0.35	0.33
D_NSC2	0.42	0.78	0.45	0.42	0.35	0.21	1.00	0.42	0.43	0.43	0.37	0.29	0.38	0.43	-0.06	0.28	0.40	0.26	0.40	0.41
D_ARC1	0.62	0.58	0.92	0.73	0.57	0.46	0.42	1.00	0.47	0.59	0.83	0.74	0.75	0.39	-0.05	0.59	0.73	0.41	0.81	0.42
D_ARC2	0.48	0.52	0.55	0.41	0.40	0.37	0.43	0.47	1.00	0.52	0.38	0.47	0.45	0.79	-0.09	0.22	0.53	0.28	0.47	0.55
D_ARC3	0.56	0.50	0.63	0.49	0.51	0.33	0.43	0.59	0.52	1.00	0.44	0.55	0.57	0.54	0.00	0.26	0.49	0.24	0.58	0.46
D_ARC4	0.52	0.57	0.84	0.71	0.56	0.50	0.37	0.83	0.38	0.44	1.00	0.70	0.68	0.33	-0.11	0.61	0.65	0.37	0.74	0.42
D_ARC5	0.56	0.46	0.81	0.69	0.53	0.41	0.29	0.74	0.47	0.55	0.70	1.00	0.63	0.39	-0.09	0.53	0.68	0.44	0.70	0.39
D_ARC6	0.61	0.51	0.80	0.70	0.53	0.41	0.38	0.75	0.45	0.57	0.68	0.63	1.00	0.42	0.01	0.46	0.54	0.40	0.75	0.47
D_ARC7	0.50	0.53	0.46	0.31	0.41	0.40	0.43	0.39	0.79	0.54	0.33	0.39	0.42	1.00	-0.07	0.11	0.39	0.23	0.41	0.50
D_ARC8	0.00	-0.10	-0.06	-0.07	-0.02	-0.10	-0.06	-0.05	-0.09	0.00	-0.11	-0.09	0.01	-0.07	1.00	-0.18	-0.18	-0.10	-0.05	-0.11
D_ARC9	0.42	0.40	0.65	0.53	0.37	0.35	0.28	0.59	0.22	0.26	0.61	0.53	0.46	0.11	-0.18	1.00	0.51	0.39	0.49	0.21
D_ARC10	0.54	0.53	0.78	0.67	0.50	0.40	0.40	0.73	0.53	0.49	0.65	0.68	0.54	0.39	-0.18	0.51	1.00	0.33	0.69	0.40
D_ARC11	0.38	0.25	0.51	0.49	0.30	0.10	0.26	0.41	0.28	0.24	0.37	0.44	0.40	0.23	-0.10	0.39	0.33	1.00	0.41	0.19
D_ARC12	0.59	0.50	0.85	0.74	0.49	0.35	0.40	0.81	0.47	0.58	0.74	0.70	0.75	0.41	-0.05	0.49	0.69	0.41	1.00	0.43
D_ARC13	0.43	0.49	0.52	0.42	0.49	0.33	0.41	0.42	0.55	0.46	0.42	0.39	0.47	0.50	-0.11	0.21	0.40	0.19	0.43	1.00

	Discharge_IJmuiden	Discharge_NSC	Discharge_ARC	R_NSC	R_ARC	D_NSC1	D_NSC2	D_ARC1	D_ARC2	D_ARC3	D_ARC4	D_ARC5	D_ARC6	D_ARC7	D_ARC8	D_ARC9	D_ARC10	D_ARC11	D_ARC12	D_ARC13
Discharge_IJmuiden	1.00	0.55	0.54	0.43	0.43	0.38	0.42	0.49	0.29	0.30	0.44	0.45	0.44	0.44	-0.04	0.32	0.45	0.12	0.46	0.26
Discharge_NSC	0.55	1.00	0.27	0.54	0.44	0.70	0.78	0.23	0.19	0.24	0.19	0.23	0.22	0.21	-0.04	0.12	0.26	0.13	0.23	0.25
Discharge_ARC	0.54	0.27	1.00	0.44	0.55	0.18	0.19	0.92	0.48	0.45	0.84	0.81	0.80	0.83	-0.05	0.62	0.79	0.16	0.85	0.35
R_NSC	0.43	0.54	0.44	1.00	0.79	0.40	0.43	0.38	0.32	0.42	0.30	0.38	0.36	0.33	-0.07	0.18	0.43	0.22	0.38	0.44
R_ARC	0.43	0.44	0.55	0.79	1.00	0.32	0.34	0.47	0.40	0.52	0.38	0.47	0.45	0.41	-0.09	0.22	0.53	0.28	0.47	0.55
D_NSC1	0.38	0.70	0.18	0.40	0.32	1.00	0.18	0.16	0.13	0.17	0.12	0.16	0.15	0.14	-0.03	0.07	0.18	0.09	0.16	0.18
D_NSC2	0.42	0.78	0.19	0.43	0.34	0.18	1.00	0.17	0.14	0.18	0.13	0.17	0.16	0.15	-0.03	0.08	0.19	0.10	0.17	0.19
D_ARC1	0.49	0.23	0.92	0.38	0.47	0.16	0.17	1.00	0.43	0.34	0.83	0.74	0.75	0.73	-0.05	0.59	0.73	0.13	0.81	0.28
D_ARC2	0.29	0.19	0.48	0.32	0.40	0.13	0.14	0.43	1.00	0.51	0.40	0.40	0.40	0.36	-0.04	0.09	0.38	0.11	0.42	0.30
D_ARC3	0.30	0.24	0.45	0.42	0.52	0.17	0.18	0.34	0.51	1.00	0.30	0.40	0.41	0.32	-0.05	0.12	0.35	0.15	0.40	0.46
D_ARC4	0.44	0.19	0.84	0.30	0.38	0.12	0.13	0.83	0.40	0.30	1.00	0.70	0.68	0.71	-0.04	0.60	0.65	0.11	0.74	0.24
D_ARC5	0.45	0.23	0.81	0.38	0.47	0.16	0.17	0.74	0.40	0.40	0.70	1.00	0.64	0.69	-0.05	0.45	0.68	0.13	0.70	0.30
D_ARC6	0.44	0.22	0.80	0.36	0.45	0.15	0.16	0.75	0.40	0.41	0.68	0.64	1.00	0.70	-0.04	0.43	0.61	0.13	0.75	0.29
D_ARC7	0.44	0.21	0.83	0.33	0.41	0.14	0.15	0.73	0.36	0.32	0.71	0.69	0.70	1.00	-0.04	0.46	0.67	0.12	0.74	0.25
D_ARC8	-0.04	-0.04	-0.05	-0.07	-0.09	-0.03	-0.03	-0.05	-0.04	-0.05	-0.04	-0.05	-0.04	-0.04	1.00	-0.02	-0.05	-0.03	-0.05	-0.05
D_ARC9	0.32	0.12	0.62	0.18	0.22	0.07	0.08	0.59	0.09	0.12	0.60	0.45	0.43	0.46	-0.02	1.00	0.44	0.06	0.49	0.12
D_ARC10	0.45	0.26	0.79	0.43	0.53	0.18	0.19	0.73	0.38	0.35	0.65	0.68	0.61	0.67	-0.05	0.44	1.00	0.15	0.71	0.31
D_ARC11	0.12	0.13	0.16	0.22	0.28	0.09	0.10	0.13	0.11	0.15	0.11	0.13	0.13	0.12	-0.03	0.06	0.15	1.00	0.13	0.15
D_ARC12	0.46	0.23	0.85	0.38	0.47	0.16	0.17	0.81	0.42	0.40	0.74	0.70	0.75	0.74	-0.05	0.49	0.71	0.13	1.00	0.30
D_ARC13	0.26	0.25	0.35	0.44	0.55	0.18	0.19	0.28	0.30	0.46	0.24	0.30	0.29	0.25	-0.05	0.12	0.31	0.15	0.30	1.00

VII.2. COPULA JUSTIFICATION

The underlying table show the results of the two validation tests. The bolt numbers indicate the best copula that shows the best fit. Thereby, graphs for a few selected cases are given in figure 56.

X	Y	ρ	ρ_{NE}	ρ_{SW}	ρ_{NW}	ρ_{SE}	M Gumbel	M Gaussian	M Clayton
D_ARC1	D_ARC10	0.70	0.66	0.18	0.36	0.19	2.21	**2.12**	3.19
D_ARC1	D_ARC12	0.41	0.13	0.10	-0.01	-0.07	**11.34**	13.94	13.05
D_ARC1	D_ARC4	0.84	0.63	0.62	0.46	0.04	0.21	**0.20**	1.34
D_ARC1	D_ARC5	0.77	0.61	0.31	0.28	0.17	2.59	**0.86**	2.87
D_ARC1	D_ARC6	0.75	0.61	0.32	0.28	0.27	**0.17**	0.28	1.82
D_ARC1	D_ARC7	0.73	0.50	0.21	-0.34	0.46	**0.39**	0.43	2.30
D_ARC1	Discharge_ARC	0.92	**0.89**	0.66	0.29	0.51	**0.10**	0.16	1.39
D_ARC10	D_ARC12	0.33	0.16	-0.18	-0.05	-0.12	**13.06**	16.89	15.06
D_ARC10	Discharge_ARC	0.76	0.72	0.23	0.29	0.24	2.23	**2.21**	2.97
D_ARC11	Discharge_ARC	0.52	0.23	0.25	0.04	0.11	**11.63**	14.55	13.48
D_ARC12	Discharge_ARC	0.84	**0.73**	0.42	-0.27	-0.25	**0.50**	0.53	1.71
D_ARC13	D_ARC3	0.49	**0.57**	-0.07	-0.12	-0.18	**1.02**	1.37	2.68
D_ARC13	Discharge_ARC	0.54	0.49	0.17	0.18	0.49	**0.51**	0.80	2.01
D_ARC2	D_ARC1	0.58	0.49	0.39	-0.01	0.19	**0.14**	0.19	0.87
D_ARC2	D_ARC4	0.56	0.37	0.29	-0.20	-0.06	**0.21**	0.27	0.88
D_ARC2	Discharge_ARC	0.64	0.52	0.44	-0.08	-0.17	**0.19**	0.28	1.05
D_ARC3	D_ARC1	0.62	**0.61**	0.24	-0.61	0.14	**0.71**	1.17	3.07
D_ARC3	D_ARC12	0.27	0.27	0.21	-0.23	-0.19	**12.49**	14.06	15.55
D_ARC3	D_ARC5	0.57	0.31	0.31	-0.42	-0.16	2.03	**1.03**	1.93
D_ARC3	D_ARC6	0.60	**0.60**	0.05	-0.42	0.02	**0.66**	1.01	2.63
D_ARC3	Discharge_ARC	0.67	**0.68**	0.34	-0.40	-0.06	**0.74**	1.17	3.16
D_ARC4	D_ARC10	0.65	0.42	0.23	0.22	0.31	**1.83**	1.86	2.27
D_ARC4	D_ARC12	0.38	0.17	-0.08	0.18	0.26	**11.26**	14.00	13.10
D_ARC4	D_ARC5	0.72	0.47	0.37	-0.05	0.30	2.05	**0.67**	2.00
D_ARC4	D_ARC6	0.69	0.59	0.19	0.44	0.38	**0.44**	0.56	2.26
D_ARC4	D_ARC7	0.71	0.45	0.29	-0.02	0.12	0.46	**0.45**	1.91
D_ARC4	Discharge_ARC	0.85	**0.71**	0.62	-0.08	0.22	**0.17**	0.21	1.37
D_ARC5	D_ARC10	0.69	0.48	0.16	0.22	0.30	2.68	**1.93**	2.68
D_ARC5	D_ARC12	0.48	0.12	0.00	0.04	-0.08	**12.45**	14.59	14.31
D_ARC5	D_ARC6	0.65	0.32	0.23	0.41	0.08	1.69	**0.42**	1.48
D_ARC5	D_ARC7	0.70	0.50	0.22	0.19	-0.08	2.01	**0.82**	2.49
D_ARC5	Discharge_ARC	0.83	0.64	0.55	0.06	-0.36	2.19	**0.66**	1.90
D_ARC6	D_ARC12	0.41	0.18	0.16	0.39	-0.30	**12.06**	15.26	14.73
D_ARC6	Discharge_ARC	0.80	0.67	0.38	0.18	0.29	**0.12**	0.19	1.53
D_ARC7	D_ARC10	0.65	0.49	0.12	0.35	0.07	2.19	**2.16**	2.41
D_ARC7	D_ARC12	0.50	0.25	-0.05	0.15	0.00	**10.94**	14.19	13.12
D_ARC7	D_ARC6	0.70	0.40	0.18	0.45	0.31	0.34	**0.27**	1.42
D_ARC7	Discharge_ARC	0.82	**0.62**	0.33	0.52	0.03	**0.32**	0.35	2.15
D_ARC8	Discharge_ARC	-0.08	-0.02	0.11	0.03	0.09	22.66	**21.41**	22.92
D_ARC9	D_ARC4	0.57	0.27	0.29	0.12	0.20	**14.31**	16.77	15.84
D_ARC9	Discharge_ARC	0.62	0.31	0.38	-0.07	0.25	**15.66**	18.13	16.40

D_NSC1	Discharge_NSC	0.75	0.36	0.55	-0.06	0.79	1.05	**0.77**	0.84
D_NSC2	Discharge_NSC	0.79	0.82	0.06	0.69	0.18	29.08	30.09	**28.91**
Discharge_ARC	Discharge_IJmuiden	0.67	0.45	0.15	0.42	0.26	**0.26**	0.45	1.84
Discharge_NSC	Discharge_IJmuiden	0.55	0.41	-0.01	0.32	0.08	**0.20**	0.44	1.27
R_ARC	D_ARC1	0.48	0.38	0.06	0.07	-0.11	**0.52**	0.72	1.59
R_ARC	D_ARC10	0.51	0.60	0.26	0.19	0.00	**3.06**	3.10	2.90
R_ARC	D_ARC11	0.27	0.21	0.12	0.12	-0.17	**12.87**	15.50	16.63
R_ARC	D_ARC12	0.27	0.21	0.12	0.12	-0.17	**12.87**	15.50	16.63
R_ARC	D_ARC13	0.56	0.54	0.09	-0.03	0.09	**0.77**	0.95	1.76
R_ARC	D_ARC2	0.39	0.35	0.09	-0.01	-0.32	**0.46**	0.56	0.98
R_ARC	D_ARC3	0.54	0.53	0.00	0.31	-0.26	**0.91**	1.09	2.10
R_ARC	D_ARC4	0.40	0.26	0.07	-0.16	-0.15	**0.42**	0.52	0.95
R_ARC	D_ARC5	0.50	0.34	0.20	0.09	-0.34	1.50	**0.64**	0.93
R_ARC	D_ARC6	0.46	0.38	0.23	0.11	-0.10	**0.47**	0.60	1.10
R_ARC	D_ARC7	0.40	0.36	0.02	-0.18	-0.42	**0.46**	0.59	1.09
R_ARC	D_ARC8	-0.11	0.00	0.09	0.25	-0.09	21.58	**19.80**	21.58
R_ARC	D_ARC9	0.19	0.28	0.19	0.14	-0.24	**13.09**	14.66	16.43
R_NSC	D_NSC1	0.43	0.30	0.22	0.23	0.27	0.66	**0.66**	0.92
R_NSC	D_NSC2	0.47	0.26	0.01	0.03	0.11	**22.39**	22.90	23.86
R_NSC	R_ARC	0.81	0.67	0.70	-0.80	-0.53	**0.57**	0.65	0.93

Table 10: Semi-correlations and 'Blanket Test' statistic for all pairs of variables used in the NPBN to model the peak discharges at pumping station IJmuiden.
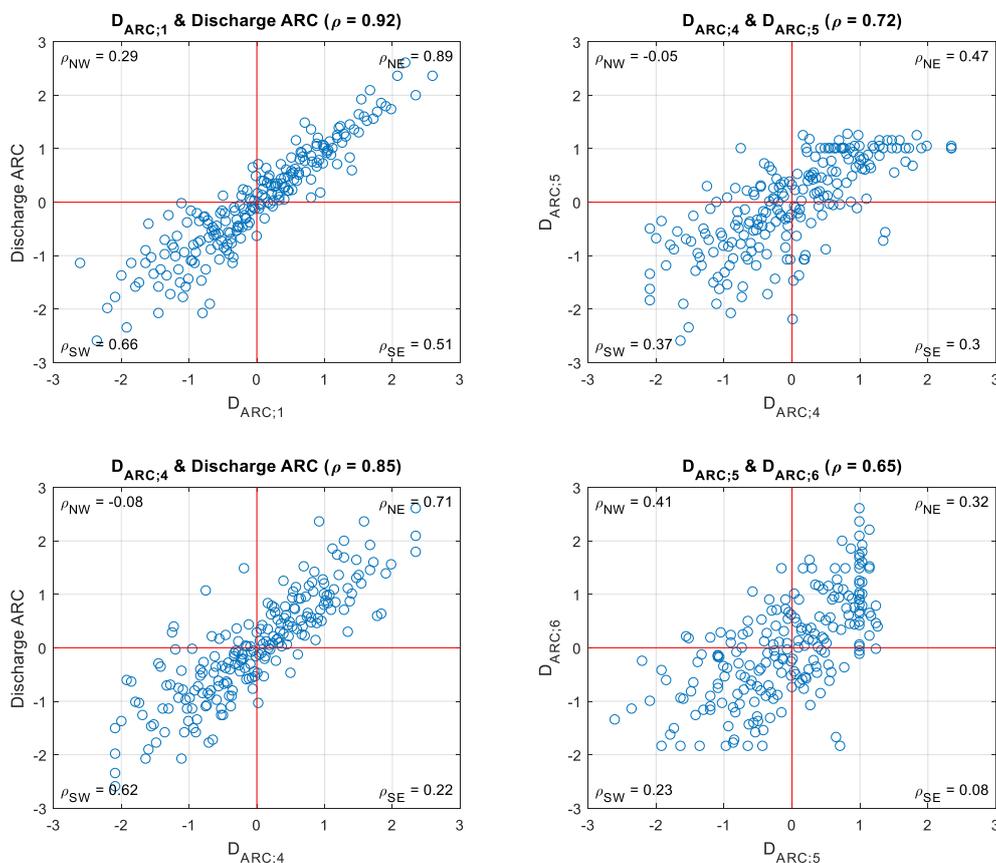

Figure 56: Graphs of selected pairs of variables of the NPBN.