Hindcasting of levee failures

Deterministic and probabilistic methods

Kool, J.J.

DOI
10.4233/uuid:6d76d0e3-18ea-4fad-821f-2267262172ac

Publication date
2022

Document Version
Final published version

Citation (APA)
Kool, J. J. (2022). *Hindcasting of levee failures: Deterministic and probabilistic methods*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:6d76d0e3-18ea-4fad-821f-2267262172ac

Hindcasting of levee failures
Deterministic and probabilistic methods

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology
by the authority of the Rector Magnificus Prof.dr.ir. T.H.J.J. van der Hagen,
Chair of the Board for Doctorates
to be defended publicly on
Wednesday 7, December 2022 at 17:30 o'clock

by

Job Jurriaan KOOL

Master of Science in Civil Engineering,
Delft University of Technology, The Netherlands,
born in Amsterdam, The Netherlands

This dissertation has been approved by the promotors.

Composition of the doctoral committee:

Rector Magnificus	chairperson
Prof.dr.ir. S.N. Jonkman	Delft University of Technology, *promotor*
Prof.dr. ir. C. Jommi	Delft University of Technology, *promotor*
Dr. ir. W. Kanning	Delft University of Technology, *copromotor*

Independent members:

Prof. dr. ir. M. Kok	Delft University of Technology
Prof. dr. K.G. Gavin	Delft University of Technology
Univ.-Prof. Dr.-ing. J. Stamm	Technical University of Dresden, Germany
Dr. ir. T. Schweckendiek	Delft University of Technology
Prof. dr. ir. P.H.A.J.M. van Gelder	Delft University of Technology, reserve member

Keywords: hindcasting, forensic engineering, river levee, slope instability, levee failures, Bayesian techniques, failure rate, likelihood ratio
Printed by: IPSKAMP printing
Cover image: Weichel, T. (2013). Failure of the Breitenhagen levee (2013). LHW Sachsen-Anhalt

ISBN 978-94-6469-161-0

An electronic version of this thesis is available at http://repository.tudelft.nl/.

TU Delft

Contents

Summary

In the last century, approximately 100,000 people lost their lives during a flood event and over 1.4 billion people were affected. As the population and economic activities grow in flood-prone areas and the frequencies and intensities of flood events increase due to climate change, damage due to flooding is expected to increase. To limit and control the potentially increased risks, in many locations flood defences, such as levees, are built and existing flood defences are reinforced. It is thus important to be able to properly estimate the reliability of these levees and to understand their potential failure mechanisms. In particular, there are significant uncertainties associated with the occurrence of geotechnical failure mechanisms such as slope instability and piping.

The hindcasting of levee failures can provide valuable information about the factors and uncertainties that dominate levee performance and reliability. Systematic forensic engineering approaches to evaluate failed structures and methods of hindcasting have been developed in the field of structural engineering, but these are not well applicable to failed levees. This is mostly due to the scarcity of relevant information prior to, during or after the levee failure, which leaves multiple scenarios and alternative model choices possible to characterize the event.

This thesis proposes and demonstrates methods for systematic analysis of levee failures at the individual section and system level. These methods of hindcasting are expected to contribute to the overall quality, repeatability, transferability, transparency and recognisability of the analysis of levee failures. In this thesis, existing approaches for evaluating structural failures have been adapted to analyse levee failures using both deterministic and probabilistic techniques. This thesis focuses on the levee failure mechanism of slope instability of the inner slope.

Firstly a deterministic method is proposed which is applied to a slope failure. In this method, the uncertainties in possible causes and computational models are modelled by defining possible scenarios explaining the failure based on all the information available. The influence of the identified scenarios and possible alternatives in model choices are analysed through a sensitivity analysis. Results of the computations are confirmed or refuted by observation information of the failure such as the shape of the failure surface. To illustrate the method, it is applied to the levee failure near Breitenhagen (2013), in Germany in Chapter 2 of this dissertation. The levee near Breitenhagen is located at the intersection of the Saale and the Elbe and it failed due to instability of the slope at the polder side of the levee. Unexpected saturation of the levee, steep slope of the levee, and the influence of the tree roots were identified to cause of the levee failure by previous reports. However, in the present study, an old breach was found to be there (the first proxy was a pond likely caused by this old breach next to the levee; the old breach was later confirmed with archive research). This old breach and pond resulted in a scenario with low strength and high water pressures in both levee and the aquifer and was identified to be the most likely scenario explaining the failure. The results indicate that locally low values of shear strength (low values of pre-overburden pressure or cohesion) explain the failure. Other scenarios that were evaluated resulted in a situation that was not likely to fail or, resulted in a slip surface that differs from the observed failure surface.

The deterministic method does not quantify uncertainties explicitly. That makes it difficult to uniquely identify the most likely scenario to explain the failure. Therefore the deterministic method is advanced by making it probabilistic and by including Bayesian techniques in Chapter 3. Thereby a better insight is provided into the relative likelihoods of the various scenarios explaining the failure. Failure observations (water level at failure, the shape of the slip surface, etc.) and a-priori levee information (soil layering, shear strength etc) are systematically taken into account to quantitatively identify the most likely scenario explaining the failure and the most representative model choices to most accurately characterise the failure. The Bayesian techniques are also used for updating the scenario and possible alternatives in model choices using the observations of the actual failure (if present) such as the shape of the slip surface. To illustrate the method, it is also applied to the levee failure near Breitenhagen (2013) in Germany. Similar to the deterministic method, the old breach resulting in a scenario with locally weak soil and aquifer connection is found to be the most likely scenario. Further, the Limit equilibrium using Spencer's approach and undrained soil response is identified to be the most representative model choices. The shear strength ratio is identified as the

most dominant contributor to the failure. Compared to the “deterministic method” introduced in chapter 2, the probabilistic method adds the possibility to quantitatively substantiate the identification of the most likely scenario explaining the failure as well as the most representative model choices.

Both methods of hindcasting have had little application and validation. Therefore both methods have been applied to a large-scale levee failure experiment. The levee of the Leendert de Boerpolder, in the Netherlands, was brought to failure under controlled circumstances. As a result, very detailed information is available. The levee was brought to failure by gradually lowering the water level in an excavated ditch at the polder side of the levee. Since the water level drawdown is known at the time of failure, this information is used to validate the outcome of both methods of hindcasting. The available levee information was used in two steps. In the first instance, only basic information was used in the hindcasting. In the second step, the geometry of the observed slip surface is also included. The probabilistic method using Bayesian techniques required some adjustment, to account for the survival of previous load phases during a stepwise increase of the load. Both methods of hindcasting identified the same water level drawdown at the moment of failure, but different model choices. In addition, the identified water level drawdown is confirmed by the observed water level drawdown at the time of the failure, i.e. 1.6 m.

Finally, this thesis introduces a method to quantify the influence of deviating conditions on the failure rate of a levee by looking at failures on a system level. The annual failure rate of a levee section is assessed based on information from historical floods. The return period of past events is also taken into account. The presence of deviating conditions at failed and survived levee sections is analysed based on satellite observations. Bayesian techniques and likelihood ratios are used to update the average failure rate as a function of the presence of a deviation. The river system of Sachsen-Anhalt, Germany, is used as a case study. It experienced severe floods with many levee failures in the years 2002 and 2013 resulting in the failure of 41 levee sections due to internal erosion, instability or overflow. It is found that the presence of geological deviations has a significant influence on the observed failure rate and that the failure rate increases with the magnitude of the hydraulic loading. The results show that in the case of the occurrence of a visually identifiable geological deviation in the subsurface, the updated failure rate of a section is about 14 times high than when there is no visually identifiable deviation. The presence of other deviations, such as bushes or trees, or permanent water near the levee also results in a somewhat higher failure rate (20–30% higher) than the calculated average annual failure rate. It is also discussed how the expected number of failures in a system during a high water event with a certain magnitude can be estimated. The results of this research can be used to further optimize soil investigations, calibrate the results of more advanced reliability analyses, and complement risk assessments. The method offers opportunities in particular in environments where little data is available.

Overall, the methods and insights developed in this thesis can contribute to a better understanding of the performance and reliability of flood defence systems.

Samenvatting

In de vorige eeuw verloren ongeveer 100.000 mensen het leven tijdens overstromingen en werden meer dan 1.4 miljard mensen hierdoor getroffen. Naarmate de bevolking en de economische activiteiten groeien in overstromingsgevoelige gebieden en de frequentie van overstromingen toenemen als gevolg van klimaatverandering, zal de schade als gevolg van overstromingen naar verwachting toenemen. Om de mogelijk verhoogde risico's te beperken en te beheersen, worden waterkeringen, zoals dijken, gebouwd en worden bestaande waterkeringen versterkt. Het is dus belangrijk om de betrouwbaarheid van deze dijken goed in te kunnen schatten en hun faalmechanismen te begrijpen. Er zijn met name grote onzekerheden verbonden aan het optreden van geotechnische faalmechanismen zoals instabiliteit en piping.

Modelmatige reconstructie (hindcasting) van gefaalde dijken kunnen waardevolle informatie opleveren over de factoren en onzekerheden die de prestaties en betrouwbaarheid van de dijken domineren. Systematische forensische technieken voor het analyseren van gefaalde constructies zijn ontwikkeld in de utiliteitsbouw, maar deze zijn niet goed toepasbaar op gefaalde dijken. Dit is vooral te wijten aan de schaarste van relevante informatie voor, tijdens of na de dijkdoorbraak. Hierdoor zijn er meerdere scenario's en alternatieve modelkeuzes mogelijk om de gebeurtenis te karakteriseren.

In dit proefschrift worden methodes voor het systematisch analyseren van dijkfalen op dijksectie- en systeemniveau voorgesteld en gedemonstreerd. Deze methoden van modelmatige reconstructie zullen naar verwachting bijdragen aan de algehele kwaliteit, repliceerbaarheid, overdraagbaarheid, transparantie en herkenbaarheid van de analyse van dijkdoorbraak. In dit proefschrift zijn bestaande benaderingen voor het evalueren van constructief falen aangepast om dijkdoorbraak te analyseren met behulp van zowel deterministische als probabilistische technieken. Dit proefschrift richt zich voor een groot deel op het faalmechanisme van instabiliteit van het binnentalud.

Eerst wordt een deterministische methode van analyseren voorgesteld en toegepast op een taludinstabiliteit. In deze methode worden de onzekerheden in mogelijke oorzaken en rekenmodellen gemodelleerd door mogelijke scenario's te definiëren die het falen verklaren. Deze scenario's zijn gebaseerd op alle beschikbare informatie. De invloed van geïdentificeerde scenario's en mogelijke alternatieven in modelkeuzes worden geanalyseerd door middel van een gevoeligheidsanalyse. Resultaten van de berekeningen worden bevestigd of weerlegd door observatie-informatie van de gefaalde dijk. Om de methode te illustreren, wordt deze toegepast op de dijkdoorbraak bij Breitenhagen (2013), in Duitsland in hoofdstuk 2 van dit proefschrift. De dijk bij Breitenhagen ligt op de kruising van, de rivieren, de Saale en de Elbe en is gefaald door instabiliteit van het talud aan de landzijde van de dijk. Onverwachte verzadiging van de dijk, het steile talud van de dijk en de invloed van de boomwortels werden door eerdere rapporten geïdentificeerd als oorzaak van het falen. In het huidige onderzoek werd echter een oude bres aangewezen (de eerste proxy was een vijver die waarschijnlijk werd veroorzaakt door deze oude bres naast de dijk; de aanwezigheid van de oude bres werd later bevestigd met archiefonderzoek). Deze oude bres en vijver resulteerde in een scenario met lage sterkte en hoge waterdrukken in zowel de dijk als de aquifer. Dit scenario werd geïdentificeerd als het meest waarschijnlijke scenario dat het falen verklaart. De resultaten geven aan dat lokaal lage waarden van afschuifsterkte (lage waarden van pre-overburden pressure of cohesie) het falen verklaren. Andere scenario's die werden geëvalueerd resulteerden in een situatie die waarschijnlijk niet tot falen zou leiden of resulteerden in een glijvlak dat afwijkt van het waargenomen glijvlak.

De deterministische methode kwantificeert onzekerheden niet expliciet. Dat maakt het moeilijk om een unieke en meest waarschijnlijke scenario te identificeren om de dijkdoorbraak te verklaren. Daarom wordt de deterministische methode verder ontwikkeld in hoofdstuk 3 door deze probabilistisch te maken en Bayesiaanse technieken op te nemen. Hierdoor ontstaat er een beter inzicht in de relatieve waarschijnlijkheid van de verschillende scenario's die het falen verklaren. Observatie die gedaan zijn tijdens het falen van de dijk (waterniveau bij bezwijken, de vorm van het glijvlak, enz.) en a-priori dijkinformatie (bodemlagen, schuifsterkte, enz.) worden systematisch in de

analyse verwerkt. Hierdoor is het mogelijk om kwantitatief het meest waarschijnlijke scenario te identificeren dat het falen verklaart, evenals de meest representatieve modelkeuzes. De Bayesiaanse technieken worden ook gebruikt de waarschijnlijkheid op een scenario en mogelijke alternatieven in de modelkeuzes aan te scherpen met behulp van de waarnemingen van het daadwerkelijke falen (indien aanwezig), zoals de vorm van het glijvlak. Om de methode te illustreren is deze ook toegepast op de dijkdoorbraak bij Breitenhagen (2013) in Duitsland. Net als bij de deterministische methode blijkt de oude bres die resulteert in een scenario met lokaal zwakke bodem- en aquiferverbinding met behulp van Spencer's benadering en ongedraineerde grondrespons het meest representatief te zijn. De shear strengt ratio wordt geïdentificeerd als de meest dominante bijdrage aan het falen. Vergeleken met de "deterministische methode", biedt de probabilistische methode zowel het meest waarschijnlijke scenario te identificeren als de modelkeuzes die het falen het beste karakteriseren. Ook wordt deze identificatie kwantitatief onderbouwd.

Beide methoden van modelmatige reconstructie hebben relatief weinig toepassing en validatie gehad, daarom zijn beide methoden toegepast op een grootschalig dijkdoorbraakexperiment. De dijk van de Leendert de Boerpolder, in Nederland, is onder gecontroleerde omstandigheden tot bezwijken gebracht. Hierdoor is er zeer gedetailleerde informatie beschikbaar. De dijk is tot bezwijken gebracht door het waterpeil in een uitgegraven sloot aan de polderzijde van de dijk stapsgewijs te laten zakken. Aangezien de waterstandsdaling bekend was op het moment van falen, wordt deze informatie gebruikt om de uitkomsten van beide methoden van modelmatige reconstructie te valideren. De beschikbare informatie over de dijk is in twee stappen gebruikt. In eerste instantie wordt er alleen basisinformatie over dijk gebruikt in de modelmatige reconstructie en in tweede instantie wordt ook de geometrie van het waargenomen glijvlak meegenomen. De probabilistische methode met behulp van Bayesiaanse technieken vereiste enige aanpassing. Deze houdt nu rekening met het overleven van eerdere belastingfasen tijdens een stapsgewijze toename van de belasting. Beide methoden van modelmatige reconstructie identificeerden hetzelfde waterpeildaling op het moment van falen, maar verschillende modelkeuzes. Het geïdentificeerde waterpeildaling op het moment van de falen wordt bevestigd door de waargenomen waterpeildaling van 1.6m.

Tenslotte introduceert dit proefschrift een methode om de invloed van afwijkende omstandigheden op de faalkans van een dijksectie te kwantificeren. Dit is gedaan door analyse van waarnemingen over dijkdoorbraken in een dijksysteem. De jaarlijkse faalkans van een dijksectie wordt beoordeeld op basis van informatie uit historische overstromingen. Er wordt ook rekening gehouden met de terugkeerperiode van hoogwatergolven uit het verleden. Aan de hand van satellietwaarnemingen wordt de aanwezigheid van afwijkende omstandigheden op gefaalde en overlevende dijksecties geanalyseerd. Bayesiaanse technieken en waarschijnlijkheidsratio's worden gebruikt om de gemiddelde faalkans aan te scherpen als een functie van de aanwezigheid van een afwijkende conditie. Het riviersysteem van Saksen-Anhalt, Duitsland, wordt als case study gebruikt. Dit riviersysteem heeft in de jaren 2002 en 2013 te maken gehad met zware overstromingen met veel dijkdoorbraken, waardoor 41 dijksecties door interne erosie, instabiliteit of overstroming bezweken. Het blijkt dat de aanwezigheid van visuele herkenbare geologische afwijkingen in de ondergrond een significante invloed heeft op de waargenomen faalkans en dat de faalkans toeneemt met de grootte van de hydraulische belasting. De resultaten laten zien dat bij het optreden van een visueel herkenbare geologische afwijking in de ondergrond, de aangescherpte faalkans van een dijksectie ongeveer 14 keer zo hoog is als wanneer er geen afwijkende condities aanwezig zijn. Ook de aanwezigheid van andere afwijkende condities, zoals struiken of bomen, of permanent water bij de dijk, resulteert in een wat hogere faalkans (20-30% hoger) dan het berekende gemiddelde jaarlijkse faalkans. Ook is behandeld hoe het verwachte aantal dijkdoorbraken in een systeem tijdens een hoogwatergolf met een bepaalde omvang kan worden geschat. De resultaten van dit onderzoek kunnen worden gebruikt om bodemonderzoek verder te optimaliseren, de resultaten van meer geavanceerde betrouwbaarheidsanalyses te kalibreren en risicoanalyses aan te vullen. De methode biedt vooral kansen in een omgeving waar weinig data beschikbaar is.

De methoden en inzichten die in dit proefschrift zijn ontwikkeld, bijdragen aan een beter begrip van de prestaties en betrouwbaarheid van waterkeringen.

1 Introduction

1.1 Context of this research

All over the world, many areas in deltas or along rivers are prone to flooding. These flood-prone areas in deltas or along rivers are often protected by levees that retain the water in a high-water situation and thus prevent flooding of the hinterland. However, this will often lead to enormous damages and loss of life, if such a levee fails during a high water situation and the hinterland is nevertheless flooded. Levees are defined here as water retaining structure consisting of soil with a sufficient elevation and strength to be able to retain water under extreme circumstances (Jonkman et al., 2017). In the last century, approximately 100.000 people lost their lives during a flood event and over 1.4 billion people were affected (Jonkman, 2005). Without further flood control measures such as levees, damages due to floods are expected to grow, as the population and economic activities grow in flood-prone areas and it is expected that the frequency and intensity of flood events are increasing due to changes in extreme weather (Winsemius et al., 2016).

Therefore, it is very important to know the performance and reliability of these levees and to understand the failure mechanisms (Jonkman & Schweckendiek, 2015; Vorogushyn et al., 2009). Over the past years, several studies have developed methods for reliability analysis for the Netherlands (Jongejan & Maaskant, 2015) and other countries, for example in the UK (Hall et al., 2003) and the USA (IPET, 2009) and China (Jiabi et al., 2013). Despite many efforts in advancing our understanding of the performance of levees, levee failure happens quite often. Geotechnical failure mechanisms, such as instability of slopes, generally occur before the water level reaches or exceeds the crest of the levee. Since this is typically unforeseen, this results in significant consequences. An inventory of over 1500 failure and performance cases of levee showed that over 90% were due to either external erosion, internal erosion or instability and a small fraction were due to overflow/overtopping or other causes (I.E. Ozer et al., 2020). In particular, there are considerable uncertainties in the occurrence of these geotechnical failure mechanisms (Vorogushyn et al., 2009). It is thus important to get a better understanding of the performance and reliability of levees, and to take into account this knowledge in their design and engineering.

1.2 Knowledge gaps

This thesis addresses the following knowledge gaps. Hindcasting of levee failures (i.e. analysis after failure has occurred) can provide valuable information about the factors and uncertainties that have determined the levees' performance and reliability. Generally, geotechnical failure causations are related to judgement errors of soil behaviour and missing out on specific boundary conditions during a strength assessment or design of a levee. The information that is gained from the hindcasting of levee failures can be used to support and advance the field of research, design, and safety assessment of levees. Past studies have used failure information to reduce uncertainties on failure mechanisms (Schweckendiek et al., 2014), field observations and monitoring (Kanning et al., 2008; Seed et al., 2008), and derivation of design criteria (Bligh, 1915; Lane, 1935; van Baars, 2005). Over the years much research has been done on the subject of forensic engineering[1] and hindcasting[2] to

[1] Forensic Engineering is “the investigation of failures - ranging from serviceability to catastrophic - which may lead to legal activity, including both civil and criminal” (Neale, 1999). However, Specter (1987) says, “Forensic Engineering is the art and science of professional practice of those qualified to serve as engineering experts in matters before courts of law or in arbitration proceedings”. However, since this thesis focusses on the hindcasting part; the litigation part is not included. In other work, the term “accident reconstruction” is used when the litigation is not part of the analysis (Carper, 2000).

[2] Hindcasting: back analysis using structural and geotechnical analyses in investigations to determine the causes of failure.

improve our understanding of geotechnical failure and the design of levees significantly, e.g. for failures in New Orleans, USA (Seed et al., 2006), along the Yabe river, Japan (Honjo et al., 2015), Breitenhagen, Germany (Grubert, 2013a), and several other cases (Bligh, 1915; Heyer, 2011; Heyer & Stamm, 2013; Kanning et al., 2008; Kortenhaus, 2003; Lane, 1935; Schweckendiek, 2014b; van Baars, 2005). Each of these studies uses their own approach or method. But methods for systematically hindcasting individual levee failures are not available.

In other fields of engineering, systematic approaches have been developed to evaluate failures – often referred to as forensic engineering. Within this discipline, there are roughly three activities to distinguish (Borsje et al., 2014; Carper, 2000; Neale, 1999; Rao, 2016; Specter, 1987; Terwel et al., 2018): collecting evidence, performing a hindcast to determine the root cause of failure and reporting. The current practice of forensic engineering suggests a systematic investigation of structural incidents that provides the use of all information during analysis. This approach is designed to make the input and modelling uncertainty explicit, in order to characterize the actual event as realistically as possible.

The current practice of forensic engineering analysis uses the collected data in two steps (Borsje et al., 2014). The first part results in an overview of all possible scenarios that might explain the failure, based on the situation before the failure. The history of loading and performance of a structure is investigated up to the failure. The first sign of failure is identified. Then all possible scenarios are verified or eliminated, leaving the most likely scenarios explaining the failure for further consideration. Detailed structural model computations are conducted to examine the most likely scenario explaining the failure, and to reproduce the event of failure (hindcasting). Data is used to verify the model outcome. The conclusion is based on rational deductions and every decision related to uncertainties is rationally argued during the process.

The uncertainties in resistance properties of levees (soil structures) are generally larger than for structures consisting of concrete or steel. Also, it is generally more complicated to build a failure model that characterizes and describes the failure event accurately for analysis purposes of levee failures, due to the scarcity of relevant and documented information. Documentation of the complete history and performance of a levee is often limited or incomplete. Even more, photo and video footage or other documentation of the levee during the failure are even more scarce. The scarcity of this information results in uncertainties about the conditions of the levee at the time of the failure. Examples of these uncertain parameters are (CIRIA, 2013; Deen & Duinen, 2016):

- Stratigraphy of the soil
- Properties of the soil layers, e.g. values of the geotechnical parameters and drainage rate of the soil
- Hydraulic loads and pore pressures

Possible loadings and subsoil conditions related to hydraulic loads, pore pressures and stratigraphy that possibly explain the failure are referred to as scenarios. For example, when such a scenario is implemented with alternatives of Limit Equilibrium Methods (LEM), soil reaction behaviour and soil parameters a slope stability model can be built to characterize the failure. Such a slope stability model can be used to evaluate which scenario explains the slope instability and which parameters are dominant. However, even when the data of relevant parameters are available from local site investigations, they do not necessarily represent the site-specific data in great detail (Schweckendiek & Vrouwenvelder, 2014; van Baars, 2005). Limited observation generally is not representative of the soil on-site. Even more, substantiation of the model choices is difficult as during the failure all physical processes are out of sight since they are located in the cross-section of the levee and after the failure, all data is swept away. Also, different failure mechanisms may overlap or influence each other (Kanning et al., 2008). Due to the combination of these uncertainties, it is difficult to characterize the conditions of the levee at the moment the load exceeded the resistance of the levee. It is complicated to identify the most likely scenario that explains the failure using geotechnical models. Past studies show that:

- Multiple scenarios can explain the failure and cannot be excluded from further investigation a-priori (see e.g. Seed et al. (2006) for the failure of the London Avenue canal in New Orleans).
- The collected evidence is typically very limited, especially in rural areas as the breach washes away the evidence.
- It appears to be difficult to substantiate the different choices in the models with collected evidence

Significant deviations are observed between computed probabilities of failure (per reference period) of individual levee sections using geotechnical models and the failure rates that are observed on macro-scale system failure (Rikkert & Kok, 2019). This is unexpected, as both assessments evaluated the same failure events and associated failure rates. The failure rate is the conditional probability of failure at time t, i.e. the probability of failure in an infinitesimal unit interval of time, given that the unit has survived until then. Typically, failure rates are associated with a characteristic life cycle failure function of an object (Bathtub curve) in three phases, i.e. infant mortality where construction or engineering errors display, the usage period that is marked with a near constant failure rate, and the wearing period where the failure rate increases due to deterioration of the object (Finkelstein, 2008). Literature on failure rates has proven to be very useful and provides insights into the capacity of the evaluated structures for dams, pipeline and other infrastructure (Baecher et al., 1980; Dawotola et al., 2011; DeNeale et al., 2019; Hatem, 1985; Hutchison et al., 2009). Only a few studies have been performed on observed levee failure frequencies (Foster et al., 2000; Major, 2019; Rikkert & Kok, 2019) but did not consider local deviating conditions that may affect the strength and reliability of the levee (Baecher, 2017; Baecher & Christian, 2015). Moreover, they do not consider how many breaches occur at certain high water levels (Bernitt & Lynett, 2010).

1.3 Aim of this dissertation

This research aims to develop and advance methods for the hindcasting of levee failures, at the individual and system levels. This will help the identification of the conditions that will affect the performance and reliability of levees, and will develop knowledge to make levees more reliable in the future. This research addresses the following general research question:

- How can methods for the hindcasting of levee failures be improved?

The following four research (sub) questions (RQ) will be addressed:

1. How can the most likely cause of levee failures be determined?
2. How can the most likely scenario explaining the failure, and the most representative model choices be quantitatively substantiated using Bayesian techniques?
3. How do the newly developed methods of hindcasting apply to another case (i.e. the Leendert de Boerspolder failure experiment, in the Netherlands)?
4. How can the failure rate of levees within a levee system be quantitatively assessed and what is the influence of deviating conditions?

1.4 Research approach

Different methods of levee failure hindcasting are developed in this thesis to provide insights into failure modelling and to identify possible conditions that might undermine the intended performance of a levee. The word method is used in this thesis to refer to a systematic procedure to the new problem of hindcasting of levees. We thereby include existing (scientific) methods such as Bayesian approaches and geotechnical fundamentals. We follow the definition of methods from the Cambridge

dictionary (“a particular way of doing something”) and Nickles (1987) who sees a method as “a detailed, content-specific (and therefore usually quite local) problem-solving procedures“. This is done by enhancing and adapting current applications of forensic analysis and hindcasting to an individual levee failure in chapters 2, 3. and 4. In these chapters, the methods are applied to a case study for an actual levee failure which is used for demonstration and validation purposes. Second, a method is introduced to quantitatively assess the failure rate of a levee section within a system based on historical information, as well as to determine what is the influence of deviating conditions on these failure rates, in chapter 5.

In chapter 2, the most likely cause of levee failures is identified in roughly three steps (Borsje et al., 2014) even though limited data are available, i.e.:

1) Collecting information for exploring possible scenarios that explain the failure,
2) Evaluating these scenarios with computational models to identify the most likely scenario explaining the failure (hindcasting) and
3) reporting the conclusion.

Based on the information prior to, during and after the failure, a computational model is developed to evaluate the failure event deterministically, which is called hindcasting. The uncertainty of the model is included in the analysis by defining possible scenarios explaining the failure. The influence of the identified scenarios and alternatives in model choices is investigated using a sensitivity analysis. This is done to identify the combination of the most likely scenario explaining the failure and the most representative model choices to characterize the levee failure most accurately. The developed method is applied to the levee failure near Breitenhagen (2013), at the junction where the Saale meets the Elbe, in Germany, which failed during the floods in the year 2013 due to slope instability of the inner slope.

In chapter 3, the procedure of hindcasting of a levee failure is enhanced by using Bayesian techniques. Failure observations and a-priori levee information are systematically taken into account to quantitatively identify the most likely scenario explaining the failure and the most representative model choices to most accurately characterise the failure. Bayesian techniques are used to incorporate observations of the actual failure for updating the failure probabilities for various scenarios. Again, the developed method is applied to the levee failure near Breitenhagen (2013), Germany. For this particular case, the failure observation concerns the observed geometry of the slip surface.

In chapter 4, both methods of hindcasting are applied to the levee failure experiment at the Leendert de Boerspolder. This provides validation and demonstration of how both newly developed methods of hindcasting can be used. The levee of the Leendert de Boerspolder was brought to failure (de Gast, 2020) under controlled conditions which means that very detailed information is available, prior to and during the slope failure. This information is used in this thesis to demonstrate the hindcasting method. As this was a controlled experiment with gradually increasing load conditions, the probabilistic method had to be extended to include the survival of previous load steps. As the conditions of the levee were observed at the time of the failure (i.e. the water level at failure, the geometry of the slip surface), the results of the hindcasting can be verified as a validation of the method.

Chapter 5 focuses on larger levee systems consisting of multiple sections. The failure rate of levees within a levee system and the influence of deviating conditions (e.g. anomalies and irregularities) are quantitatively assessed, using Bayesian techniques, satellite images and historical data. The collection of historical data is used to estimate failure rates and a prior probability of occurrence of specific failure mechanisms, i.e. macro-stability and internal erosion, and overflow, as a function of outside water levels. Second, both the failed and survived levee sections are analysed for deviating conditions. These results are used to quantify the relations between observed deviations and the occurrence of failure. Third, Bayesian techniques are used to update the earlier estimated failure probabilities (as done in the first step) on the information on the occurrence of these

specific deviations. The developed method is applied to the river system in the state of Sachsen-Anhalt, Germany; which is the same river system where the levee failure near Breitenhagen (2013) analysed in the earlier chapters occurred.

This dissertation's novel contributions are:

- The development of both a deterministic and a probabilistic method for systematic hindcasting of levee failures. These are characterized by the specific challenges of limited information and many possible failure modes. It is shown how Bayesian techniques can help systematically identify the most likely scenario explaining a levee failure, by including available information and observations. Both methods can be used for systematic hindcasting as part of forensic analysis after a failure has occurred
- The validation of both developed methods of levee failure analysis by applying them to another case study (Leendert de Boerspolder) where the actual failure conditions were mostly known. The probabilistic hindcasting method has been extended to incorporate the effect of gradually increasing load conditions. The case allowed validating the method using increasing levels of information.
- The development of a method to systematically determine failure rates based on the historical observed failure in a levee system, and to assess the effects of deviating conditions on failure rates. These methods will be particularly valuable for levee systems and regions with limited detailed data available.

1.5 Dissertation overview

The topics in this dissertation can be categorized as hindcasting on individual levee failures and hindcasting on levee failures as part of a levee system. Chapters 2, 3, and 4 focus on the hindcasting of individual levee failures while the research in chapter 5 focuses on the hindcasting of levee failures as part of the levee system. An overview of the chapters is presented in Figure 1.1.

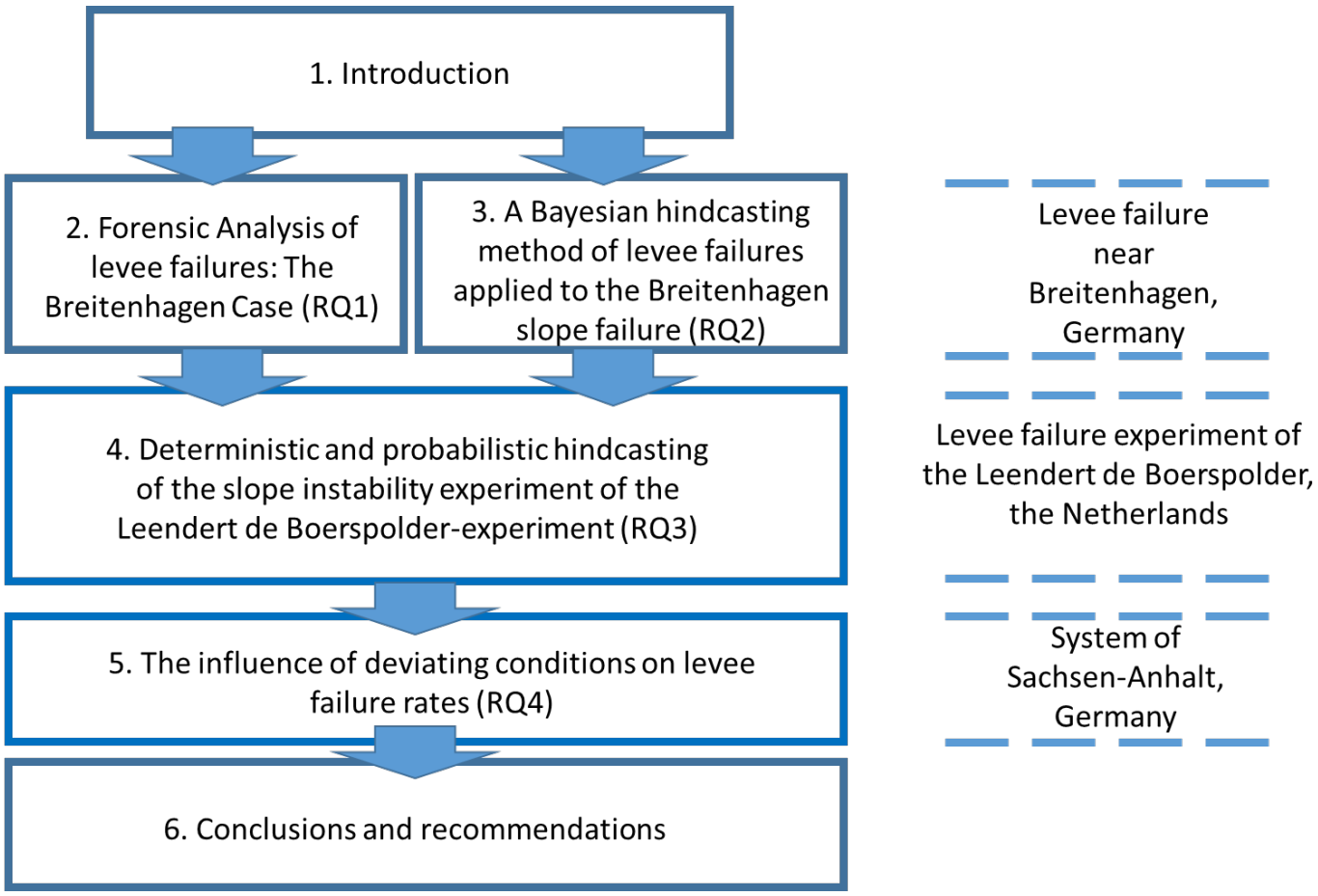


Figure 1.1 Overview of this dissertation on the left side of the figure. Each panel corresponds with a chapter in this dissertation and the associated research question is indicated. On the right side of the figure, the case studies are indicated.

Chapter 2 introduces a generic deterministic framework of forensic analysis for levee failures and a method of hindcasting to identify the most likely scenario and most likely cause of triggering the levee failure that occurred near Breitenhagen (Sachsen-Anhalt in Germany, 2013). This method of hindcasting is improved by using probabilistic methods and Bayesian techniques to substantiate the identification of the most likely scenario explaining the failure quantitatively in chapter 3. In chapter 4, both the deterministic and probabilistic methods of hindcasting are validated by applying them to the Leendert de Boerspolder experiment (Noord-Holland in the Netherlands). Chapter 5 uses a pool of historical levee failure and survival data of the overall levee system of Sachsen-Anhalt, in Germany (of which the Breitenhagen failure was one of the failures). The applicability of the research is discussed in each chapter itself. The conclusions and recommendations are reported in chapter 6. chapters 2, 3, and 4 are based on previously published articles in journals, which have been included in their original form. Therefore there may be some smaller overlaps between chapters (e.g. in the case study description) and some minor differences in the terminology. To support the readers during reading of this dissertation there is a glossary included in **Error! Reference source not found.** which highlights some of the key terms and discusses how some of these terms have been used differently in the different chapters.

As mentioned, chapters (2, 3 and 5) are based on previously published journal papers with the author of this thesis as a first author. Although the journal papers would not have achieved their quality without the contributions of co-authors, the first author has done the bulk of the concept development, all analysis, and most of the writing.

1.6 SAFElevee project

This research was one of the subprojects of the SAFElevee project that focused on improving the understanding of failure mechanisms and breaching of flood defence systems and providing systematically collected datasets for future scientific research. Other research projects within SAFElevee have focussed on and development of advanced levee breaching models (van Damme, 2020), levee deformation and monitoring (Ozer et al., 2019). As part of the SAFElevee project, the International Levee Performance Database has been developed which includes information on more than 1500 failure and performance cases (I.E. Ozer et al., 2020), some of which are described in more detail in this thesis. The SAFElevee project was supported by STW (now NWO – the Dutch Research Council) under project number 13861.

2 Forensic analysis of levee failures: The Breitenhagen case[3]

2.1 Introduction

It is of the utmost importance that the understanding of levee failures is increased continuously since the consequences of flood defence failures during extreme events are very large, including damages and possible loss of lives (Jonkman, 2005). The hindcasting of levee failures can provide valuable information about general levee performance, the quality of the strength models and the dominant factors that contribute to levee failure. New insights that are gained from the hindcasting of breaches can be used to support and introduce developments in the field of research, design and engineering of flood defences. Examples are the utilisation of failure data (Schweckendiek, 2014b), field observations and monitoring (Kanning et al., 2008; Seed et al., 2008; Zang et al., 2013), or even new design criteria (Bligh, 1915; Lane, 1935; van Baars, 2005). Over the years, much research has been done on the subject of forensic analysis (strategies) of historical (near-) failures to increase the understanding of levee performance behaviour under extreme circumstances and to develop calculation models to predict the performance of the levees and to make their uncertainties explicit (Jonkman & Schweckendiek, 2015; Rao, 2016; Vorogushyn et al., 2009). Nevertheless, the rivers in Germany experienced a large discharge in the years 2003 and 2013. This resulted in floodings in 12 of the 16 federal states in Germany causing enormous damage (A.H. Thieken et al., 2016). The levee in Breitenhagen, Germany failed in 2013 during a high river level and is considered to be a failure due to the instability of the landside slope of the levee (Grubert, 2013a). The first explorative calculations of the strength and performance are based on best estimates of the situation and do not indicate that this particular levee should have failed. This is indicating a further need for a systematic investigation of the failure.

The discipline of forensic analysis roughly distinguishes three activities (Bell, 2001): collecting data and reviewing them, executing a hindcast analysis with calculative models to determine and report the cause to learn from the findings. Literature shows that forensic engineering and hindcasting can be used on different professional levels (Terwel et al., 2012). Current applications of forensic engineering to levee failure are dominated by large parameter and model uncertainties leading to seemingly arbitrary decisions that do not account for the uncertainties, leaving the outcome of the analysis open for discussion (Grubert, 2013a; Kanning et al., 2008; Vorogushyn et al., 2009).

This chapter aims to identify the most likely causes of the Breitenhagen levee failure with the help of a generic systematic approach to forensic analysis. The original elements in this chapter are the analysis of the Breitenhagen levee failure and the development of a generic approach for identifying the most likely causes of geotechnical failures. Attention is paid to a well-argued deduction, how to account for all considerations and modelling choices that were made during the process of forensic analysis. The analysis considers the uncertainties introduced by common geotechnical analytical simple models in combination with the lack of typical vital information which is common when a levee accidentally fails. At the same time the developed approach of forensic analysis is expected to contribute to the overall quality, repeatability, transferability, transparency

[3] This chapter has been published as: Kool, J., Kanning, W., Heyer, T., Jommi, C., & Jonkman, S. N. (2019). Forensic analysis of levee failures: The Breitenhagen case. *International Journal of Geoengineering Case Histories, 5*(2), 70-92. https://doi.org/10.4417/IJGCH-05-02-02

and recognisability of the analysis of levee failures. It can thereby contribute to insights into the strength and performance of levees.

First, the relevant data are presented for the case (section 2.2). The method of the analyses is explained in section 2.3. This is followed by the assumptions regarding the sensitivity calculations in section 2.3 as well. Subsequently, the results are presented in section 2.4. Finally, the approach of the analysis in relation to the results is discussed in section 2.5, followed by the conclusion (section 2.6) and the recommendations (section 2.6).

2.2 The Breitenhagen levee failure

During the floods of the Saale River in June 2013, the east embankment breached near Breitenhagen (Saxony, Germany) (Figure 2.1 and Figure 2.2) (Grubert, 2013a). The breach of the levee caused a lot of economic damage that was financially covered by the state of Saxony-Anhalt (Drews, 2017b). The failure appears to be the result of slope instability due to the long-lasting high water wave in both the Elbe and the Saale (Drews, 2017b; Grubert, 2013a). Historical documents indicate that the levee was built in the second half of the 19th century (Sixdorf, 2016). The overview photo shows a grassy levee with a paved road on top of the levee which can be used by heavy-weight traffic (Figure 2.1). Trees grow along the length of the levee in the floodplain of the Saale. The overview photo also shows a pond near the location of the breach at the river side of the levee (Figure 2.2).

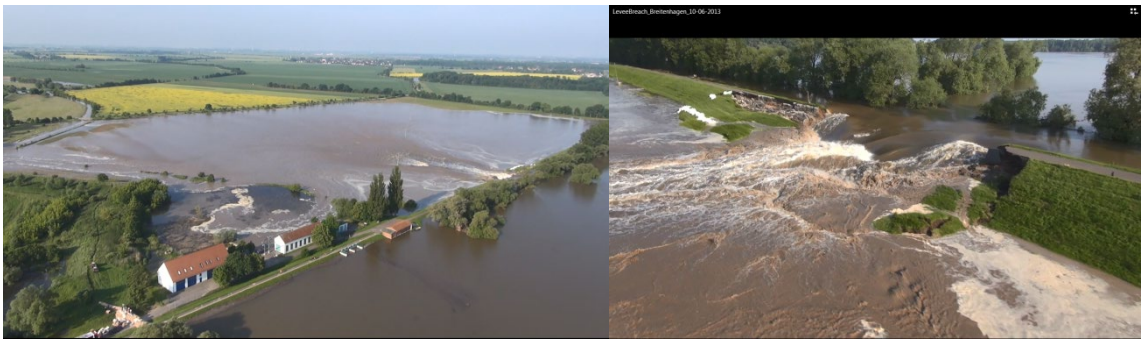


Figure 2.1 Left: overview photo of the landside while the flooding evolves, Right: overview photo of the breach on June 10th 2013 (Weichel, 2013).

Figure 2.2. Overview of the location of the breach (GeoBasis-DE/BKG, 2009).

An analysis of the case is carried out by using the provided and relevant information. The collected information is analysed on factors that will influence the performance behaviour being input parameters (possible causes of failure e.g. pore water pressures and shear strength of the soil) and model choices (e.g. LEM and (un)drained soil behaviour). These factors are used as bases of the computational models that are used to execute the hindcasting to determine the most likely cause

of failure. This section presents an overview of relevant information that is available. This includes an introduction of the history of the construction, failure and repairs, the water levels, an analysis of the local conditions, the external features, the known geotechnical aspects and the findings from a previous analysis.

Table 2.1 Overview of collected documentation (Grubert, 2013a; Sixdorf, 2016; Weichel, 2013)

Document	Information
Design of the upgrade 1846 (Sixdorf, 2016)	Design drawings of the cross sections. The design was not realized.
Photo reportage 2003 (Sixdorf, 2016)	Photos of the installation of sheet piles to prevent seepage near the pumping station
Photo reportage 2004 (Sixdorf, 2016)	Photos of the construction of the road on top of the crest
Video footage, 2013 (Weichel, 2013)	Video footage during the breach by drone flight
Saaledeich bei Breitenhagen, geotechniscche untersuchungen der Bruchstelle Empfehlungen zut Sanierung, dr.-Ing. P. Grubert, 2013 (Grubert, 2013a)	Photo reportage (during and after the breach) Analysis based on calculations Location overview Levee profile (measurements) Levee profile km 0+590 (incl. borings) Soil mechanic laboratory tests Water content determinations Sieve curves Geotechnical stability calculations (Bishop and Janbu) variety of scenarios Underground hydraulic analysis (stationary and transient pressures) Earth stance, assessment of uplift
Photo reportage and paper clippings of the repair 2013 (Sixdorf, 2016)	Photo footage of the repairing of the levee Clippings of the plans in the area Photo footage right after the repair

2.2.1 Water levels during Elbe and Saale flood of 2013

In the summer of 2013 both the Elbe and the Saale experienced a high water wave due to heavy rainfall upstream. The peaks of the high water waves met each other at the intersection of the Saale and the Elbe near Breitenhagen. The actual water height was not measured at the location of the breach but the water levels were measured at the nearest upstream and downstream measurement station and the levels at Breitenhagen were interpolated (Figure 2.3) (Drews, 2017b). The elevation of the (water) level is expressed as a standard reference level called the NormalHohenNull (NHN) which is approximately equal to the mean sea level. The breach occurred just at the moment that the water levels in both rivers reached their peak on June 8th, 2013. It appeared that the water levels of 2013 were the highest water levels ever measured in both the Saale and the Elbe (Drews, 2017b). Both water levels have a return period of 100 to 200 years (Drews, 2017b).

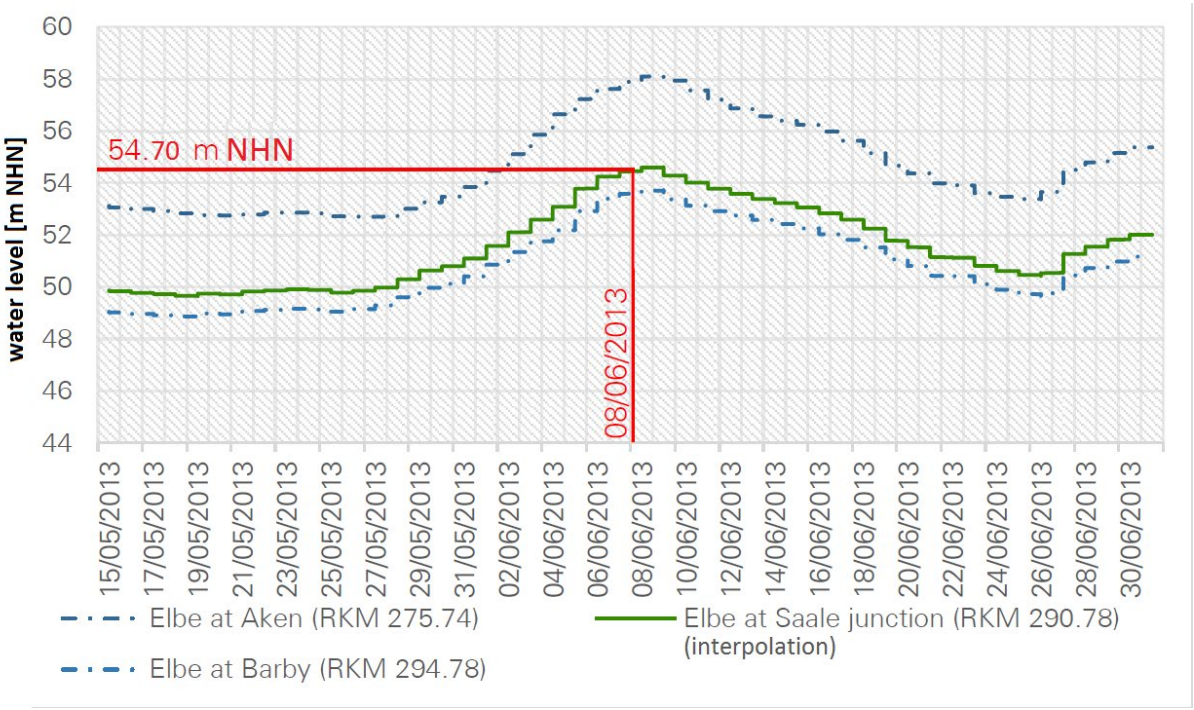


Figure 2.3. Graphs with water levels over time of the Elbe and Saale near Aken and Barby. Waterlevel at Elbe and Saale junction, RKM 290.8, is NHN+54.7 m (Drews, 2017b; Grubert, 2013a)

2.2.2 Cross-section and stratigraphy

The stratigraphy is reconstructed using the borings that were done over five cross-sections, including samples at the toe and from the crest of the levee (Grubert, 2013a). The borings were taken near the location of the breach (length marker +600m). In total twelve borings were completed to a depth of approximately 3 m to 4 m. The borings show a little variation in width over the cross-section (Figure 2.4).

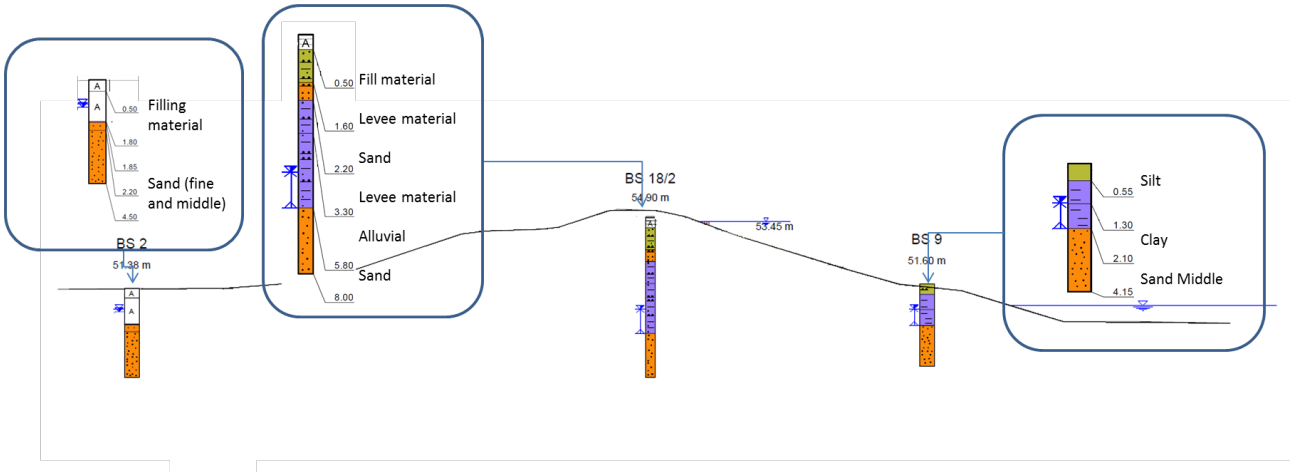


Figure 2.4 Cross-section of the levee (right is riverside) near the location of the breach (length marker +590m), with the water levels, groundwater levels and the results of the borings (titled BS2, BS18/2, BS9 taken about 10m in distance of the start of the breach (Grubert, 2013a). The borings were taken both before and after the breach of 2013.

The cross-section generally consists of two layers (Figure 2.4 and Table 2.2) (Grubert, 2013). The top layer of the levee consists of largely homogeneous cohesive material (clay-like). A few inclusions of locally mixed-grained soils are identified. The bottom layer consists of a sand layer that can be found throughout the cross-section at an approximate elevation of NHN+49.5 m about 5.5 m deep from the surface (Table 2.2). This layer likely functions as an aquifer. The identified soil types

are characterised by the typical characteristic parameters listed in Table 2.3 (Grubert, 2013; Schneider & Albert , 2014).

Table 2.2. Stratigraphy at the location of the breach (*this boring was done in 2007) (Grubert, 2013a). The description of each layer per boring is presented. The position of each soil layer is related to the NHN, with matching soil layer numbers in the calculative models used in Grubert (2013a), see Figure 2.5.

Toe (landside) code: BS2			Crest code: BS18/2*			Toe (waterside) code: BS9		
Soil description	Layer Nr.	m NHN	Soil description	Layer Nr.	m NHN	Soil description	Layer Nr.	m NHN
fill up material	Nr.1 and Nr.3	51.4-49.6	fill up material	Nr.1	54.9-54.4	Silt	Nr.1	51.6-51.0
Sand (fine and middle)	Nr.4	49.6-46.9	Levee material	Nr.2	54.4-53.3	Clay (fine)	Nr.2	51.0-50.3
-		-	Sand	Nr.2	53.3-52.7	Clay	Nr.2 and Nr.3	50.3-49.5
-		-	Levee material	Nr.2	52.7-51.6	Sand (middle)	Nr.4	49.5-47.5
-		-	Alluvial	Nr.3	51.6-49.1	-		-
-		-	Sand	Nr.4	49.1-46.9	-		-

The dimensions of the cross-section of the levee are mapped at the location of the breach (Figure 2.5). The measurements show that the levee is about 3.50 m high with a crest width of 3.00 m. The landside slope is 1:2.1 and the waterside slope is 1:3.1. The distance between the waterside toe and the pond is about 3.00 m.

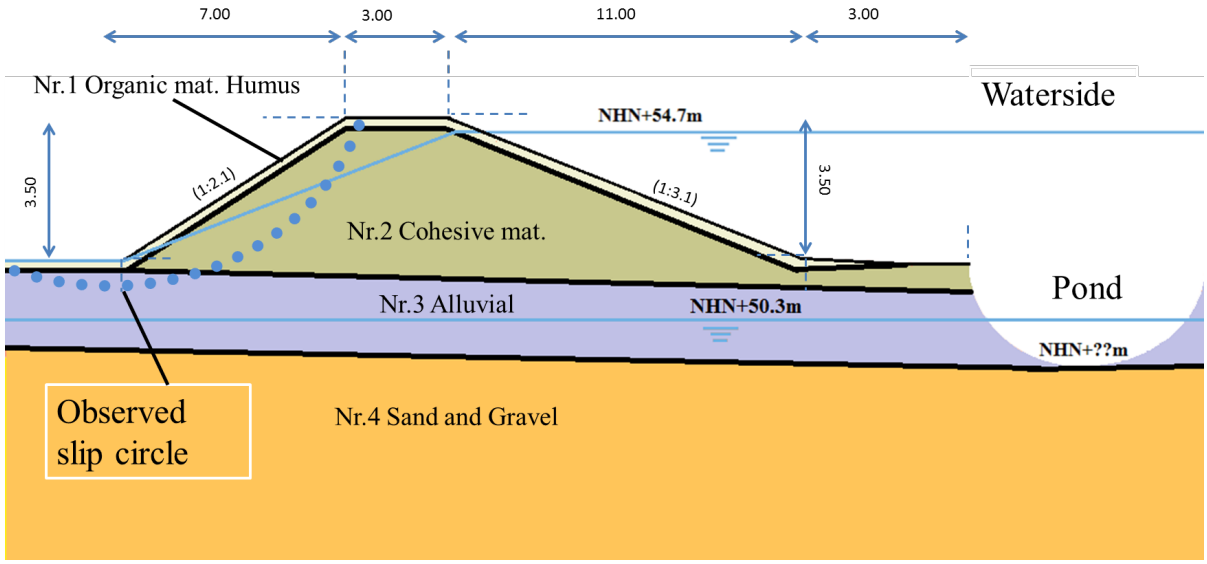


Figure 2.5 Dimensions of the Breitenhagen levee including simplified soil layering flood level and assumed aquifer level (pre-failure, left side is landside, right is the waterside). The 2013 high water level is NHN+54.7m, the measured groundwater level is NHN+50.3m. The soil modelling is based on the forensic engineering report (Grubert, 2013a), incorporated with the maximum occurred flood level, and soil types (Grubert, 2013a).

The findings from the borings were translated into a 2D-geotechnical model to simulate the resistance of the levee (Figure 2.4) (Grubert, 2013a). The cross-section in this model consists of mostly cohesive material in the top layers and levee itself, with a subsoil of sandy material (Table 2.3 and Figure 2.5) which is considered to be very conductive.

2.2.3 Geotechnical parameters

The results of the borings as shown in Table 2.2 are related to the values of parameters with similar soil descriptions (Table 2.3). Best estimates for relevant soil parameters from the standard best estimate values provided by the Deutsches Institut fur Normung (DIN) or table 2.7 from “Bautabellen für Ingenieure” (Grubert, 2013a; Schneider & Albert, 2014).

Table 2.3 Best estimate of relevant parameters used for the different soil type (bulk unit soil weight:γ, friction angle: φ ', cohesion: c') (Grubert, 2013a; Schneider & Albert, 2014).

Layer Nr.	Name soil layer (Grubert, 2013a)	Description of the layers of soil (Schneider & Albert, 2014)	γ kN/m^3	φ ' degr	c' kN/m^2
Nr.1.	S 0a-Mutterboden	Organic mat. Humus	18.0	26.0	7.0
Nr.2.	S1a-Deichkorper, bindig	Cohesive mat.	20.0	25.0	10.0
Nr.3.	S 2-Aueablagerungen	Alluvial	20.0	25.0	10.0
Nr.4.	S 3-Sand und Kiese	Sand and gravel	18.0	33.0	0.0

2.2.4 History of the Breitenhagen levee: construction, failure and repair

In this section, the past performances of the levee are summarized in terms of construction, upgrades, repairs and breaching thus providing insight into the past performance of the levee (Sixdorf, 2016):

- 1845: the first documentation of the levee protection with a crest height of 0.50 m;
- 1862: the levee is breached on purpose to drain the polder which was flooded. Followed by a repair of the levee;
- 1955: the installation of an electric pump near the levee. For this purpose, a cable was installed from Breitenhagen to the pumping station inside the levee cross-section;
- 2003: a high water wave took place;
- 2003: the installation of sheet piles into the levee body, just up to the location of the pumping station as a measure against seepage;
- 2004: the construction of the crest road;
- 2013: the levee breached;
- After the flood, installation of more sheet piles as a prevention of seepage and the installation of a stability berm to prevent instability of the slope at the landside.

2.2.5 Findings previous analysis of the Breitenhagen failure

The Breitenhage levee failure was studied by Grubert (2013a). The purpose of this study was to investigate the cause of the breach and the measures required for permanent repairs. At some distance from the breach, cracks were found in the dyke body. The investigations were therefore extended to the adjacent area. The analysis in this study on the Breitenhagen case notes that the roots of the trees at the riverside of the levee grow inside the levee (Grubert, 2013a). The contribution of the tree roots inside the levee is considerable and is indicated as a probable cause of the breach. The analysis is based on geotechnical investigations and stability analyses that are included in the report. The results show that a part of the cause is probably uplift of the top layer related due to the little coverage of the cohesive top layer. This is considered a secondary cause, since the observed sliding mode does not meet the typical dimensions of such a mechanism. The most important causes of the failure and adjacent cracks are a combination of (Grubert, 2013a):

- The relatively steep angle of the slope at the landside;
- An unusual vertical crack occurred in the body of the levee due to the drying out of the clay;

- Horizontal conductive layer due to the growing of the tree roots on the riverside in a horizontal plane;
- The levee is made out of plastic clay which is low in water content.

Two different scenarios of water pressure development in the cross-sections were analysed in the stability models (i.e. with tree roots and without tree roots growing in the cross-section). The results are expressed as a ratio between the resistance (R) and the loads (S) which is called the factor of safety (FoS=R/S). The scenario with tree roots (FoS =0.92) and without tree roots (FoS=2.04) were both analysed with the help of the Janbu stability model. The Bishop model of the levee shows that the levee without the trees is calculative stable (FoS=1.45). Both analyses use the characteristic values of soil of Table 2.3, as a conservative estimate. The findings of Grubert (2013a) are incorporated in this study.

2.2.6 Observations before, during and after the breach in 2013

With the help of the collected data (mostly photographs), an exploration of visual indicators is performed. The analysis of the data also provides possible input of the model (factors of influence) such as the geometry of the levee and the geometry of the slip surface, stratigraphy, phreatic lines and transient groundwater flow in possible high conductive layers. ‘’The data indexed as prior, during and after the breach.

Observations before the breach

The collected data which were taken prior to the breach indicates the condition of the levee prior to the breach. The overview photo shows a pond in front of the levee possibly caused by an earlier breach (Figure 2.2). This indicates the possibility of a direct connection between the outside water level and the aquifer below the levee. The photos show a row of trees at the riverside toe of the levee which might influence the conductivity of the levee due to the intrusion of the roots inside the levee, or might indicate the presence of a more conductive layer inside the levee (Figure 2.2). The slope of the levee is very steep (about 1:2.1) with a gradient that surpasses the friction angle of most soil types. The slope at the landside is prone to instability (Figure 2.4).

Observations during the breach

Photo data were collected during the breaching process as well, showing recognizable indications of the levee failure. The indicators were used to identify the deformations and the geometry of the slip surface (Figure 2.6). Also, ponding due to possible seepage of water at the landside is shown. Possible high water pressures in the aquifer may cause seepage and saturation of the soil at the landside of the levee.

Figure 2.6. Photo series over time (upper left: 08.06.2013-10:37, initial break at the edge of the slope, upper right: 08.06.2013–14:10, horizontal shifts, bulge of the toe, lower left: 08.06.2013–20:51 continuing failure, lower right, 08.06.2013–20:55, continuing failure) with the visual predictors (Grubert, 2013a). The riverside of the levee is on the right for all photos.

On the left side of the picture an electricity pole is recognized which seems to tilt over time, until it is almost completely horizontal (Figure 2.6). Also, the slip surface is expanding until the road crumbled indicating an increasing deformation over time. The initial failure is considered to have weakened the levee to such a level that the levee lost all its retaining capacity.

Observations after the breach

The photos taken after the breach show the presence of roots in the cross-section even after the breach in Figure 2.7. The roots are numerous and the length of the roots indicate that they have grown deep into the levee. The presence of the tree roots might indicate a highly conductive layer which could have a direct effect on the local pore pressures inside the levee. The soil surface that is exposed by the breach, seems to consist of a cohesive material.

Figure 2.7. Photo of the cross-section of the breached levee (Grubert, 2013a)

2.3 Framework for forensic engineering analysis and its application to the Breitenhagen levee failure

In this section, a framework for forensic analysis of levee failures is developed which is based on a generic approach of forensic engineering originally developed by the Dutch Organization for Applied Scientific Research (TNO) to analyze structural failures (Borsje et al., 2014). However, the TNO forensic engineering approach is not sufficiently applicable to analyse levee failures that are associated with different and larger uncertainties and different failure modes. In the following section, the proposed “generic” approach for general forensic engineering by TNO will be discussed first. In the second section, a specific approach to forensic engineering of levee failures is introduced. The third and fourth sections describe in more detail the approach and considered scenarios for the Breitenhagen failure respectively.

2.3.1 Forensic engineering

TNO has developed a generic approach for a systematic investigation of (steel) structural incidents that propagating a clear insight in the use of all information during the analysis (Borsje et al., 2014). The approach accounts for all decisions as some typical and vital information might be uncertain. This approach is designed to make the input and modelling uncertainties explicit in order to simulate the actual event as realistic as possible without the usage of probabilistic techniques. The generic approach of forensic engineering analysis of TNO uses the collected data in two steps (Borsje et al., 2014). The first part results in an overview of all possible scenarios, based on the situation before the failure. The history of loading and performing of a structure is investigated up to the failure. The first sign of failure is located. Then all possible scenarios are verified or falsified, leaving the most likely causes for further consideration. Detailed structural calculations are used to examine the most likely scenario of combined causes of the failure, in order to reproduce what happened (hindcasting). The conclusion is based on rational deductions and every decision related to uncertainties is rationally argued during the process.

Compared to steel structures, which were the object of investigation of TNO, geotechnical failures introduce an additional complexity regarding the input and modelling uncertainties. Documentation of the complete history and performance of a geotechnical structure (in this case a levee) is often limited or incomplete. Even when the data of relevant parameters are distributed from local site investigations, they do not necessarily represent the site-specific data in great detail (Schweckendiek et al., 2014; van Baars, 2005). In the analysis of the levee using geotechnical models, multiple model choices can be used to determine the strength performance behavior of the levee. No matter how advanced these models are, they will always suffer and reflect uncertainties in the model outcome. Typical input parameters from standard laboratory data are used, which in general do not account for time-dependent behavior and, thus cannot be representative of the natural heterogeneity of the soil on site. Validation of the geotechnical models is complex as during the failure all detailed processes within the cross-section are withdrawn from observation, and after the failure usually all evidence is washed away. Also, different failure mechanisms may overlap or influence each other (Kanning et al., 2008). The sum of uncertainties result in many of combinations and possibilities that could lead to failure. The suggested approach of forensic engineering for our case of levee failure focuses on several individual factors that are typically points of special attention when performing a simulation of the strength performance of a levee. The approach analyses their influence one by one.

2.3.2 Proposed approach for forensic analysis of levee failures

To identify the most likely cause of a levee failure a systematic approach for forensic analysis of levee failures is suggested. In order to make the generic forensic analysis approach more applicable to levee failures, the standard three steps of forensic engineering as proposed by TNO (see section 3.1) are adapted into three more specific step for levees (A, B and C, see Figure 2.8).

In stept A., the collected data that are typically relevant for levee breaches, are sorted in three ways, i.e. data collected prior, during and after the failure (Figure 2.8). This provides a clear overview

of which data are relevant in terms of factors that are of influence the performance behaviour and a first impression of the performance behaviour of the levee during extreme circumstances. A distinction is made between input parameters (which are considered as possible causes e.g. pore water pressures and shear strength of the soil) and model choices (e.g. LEM and (un)drained soil behaviour). With the data during and after the failure, possible failure scenarios are determined and the behavior of a levee is studied.

In step B., a sensitivity analysis is used to validate or falsify possible causes contributing to the levee failure (hindcasting) (Figure 2.8). Using the data and additional literature, the best estimate scenario and expected values for the input parameter are modelled (further referred to as “base model”). During the process of developing a “base model”, all the individual uncertainties are identified and introduced as scenarios (in terms of input parameters and modelling choices). During the process of the sensitivity analysis, the expected values of the still uncertain input parameter are replaced one by one by realistic upper and lower limit boundary values.

Exploration of all possible combinations of individual scenarios can be a very time-consuming exercise. The focus of the scenarios and sensitivity analysis is mainly on the influence of the different individual uncertain factors of influence, except for a number of combinations of causes which are suspected to cause failure. The results of the sensitivity analysis validates or falsifies the different scenarios of individual possible causes of the failure and is intended to identify the most likely causes in step C. The most likely causes can be further validated by performing additional field tests, research or more detailed calculations resulting in the final conclusion assuming the most likely individual cause of the levee failure.

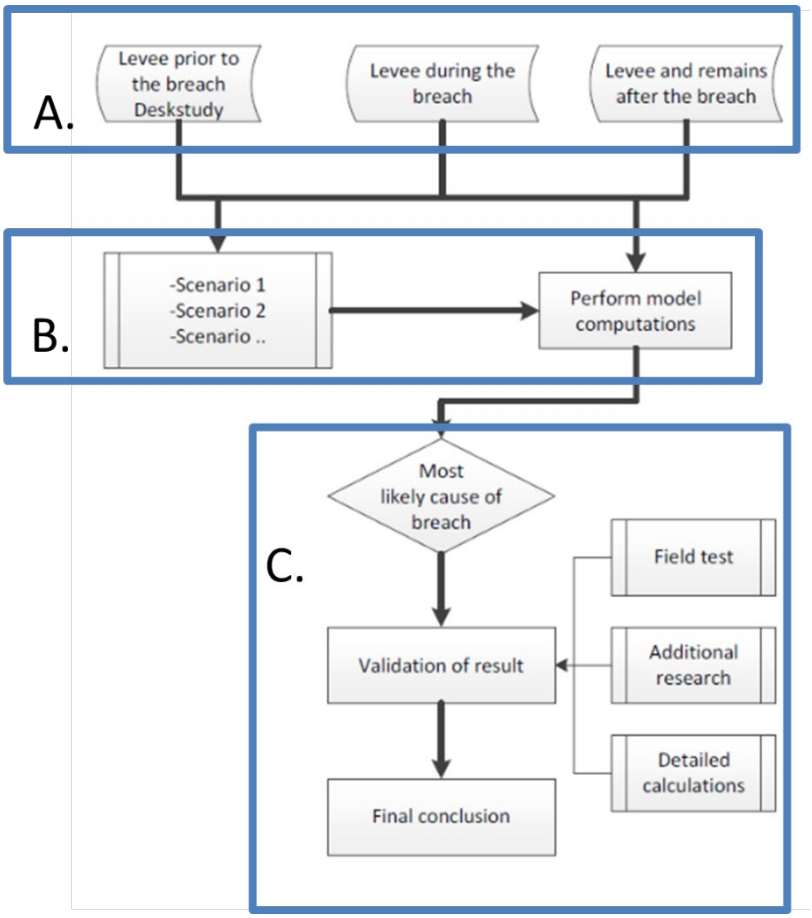


Figure 2.8. Generic approach for levee forensic analysis based on (Borsje, Renier, & Bruggraaf, 2014) Step A.: Collecting information for exploring possible scenarios, Step B.: Evaluate with a calculative model, Step C.: When results do not identify the cause, additional information can be gathered.

2.3.3 Forensic analysis of the Breitenhagen failure

The generic approach of forensic analysis for levee failure as presented in the previous section is applied to the Breitenhagen levee failure. This paragraph elaborates on the identification of all possible scenarios and causes of failure that are distilled from the collected data (step A) which are used for the sensitivity analysis in step B as proposed in the previous section.

Based on the known evidence (section 2.2), the Breitenhagen case is considered to be a failure due to the instability of the slope at the landside. When performing a levee stability analysis the following factors are of influence on the performance behaviour and are categorized in terms of input parameters and model choices (CIRIA, 2013; Deen & Duinen, 2016):

- The geometry of the levee;
- Stratigraphy of the soil;
- Water levels (inside and outside);
- The water pressure (development) inside the levee and hydraulic head in the subsoil;
- The values of the geotechnical parameters of the soil;
- The drainage rate of the soil;
- The specific Limit Equilibrium Method (LEM).

In step A. the collected data are thoroughly researched for data that is related to the factors of influence in order to identify all uncertainties. The geometry, stratigraphy of the soil, and water level could be identified utilizing the collected evidence. The hydraulic head in the subsoil, and the soil strength in terms of geotechnical parameters of the soil are still uncertain and are considered possible causes of the levee failure. Also, the model choices in terms of the drainage rate of the soil of the LEM are still uncertain.

To determine the possible contribution to the failure of the uncertain input parameters and model choices, all possible scenarios and causes of failure are identified. The collected data are used to construct a stability model in order to simulate the best estimate scenario together with the expected values of input parameters (representing the best estimate conditions). This model is called the “base model” as shown in Figure 2.9. In the case of the Breitenhagen levee failure, the focus is on the uncertainties in water pressure development (level 1), the drainage rate of the soil (level 2) (whether the response of the soil is better represented by drained or undrained schemes) and the related values of shear strength parameters (level 3). Moreover different LEM schemes (Bishop, Uplift Van and Spencer) are used to analyse the calculation model uncertainty (level 4).

The next step is to identify the individual influence of the uncertain factors (level 1 to level 4) and verify whether the calculated slip surface naturally matches the actual observed slip surface (Figure 2.11) and doesn’t intersect with the sand layer. For the input parameters, the lower and upper limit boundary values are introduced, one by one, into the model of “base model” (Figure 2.9). The actual input values and factors are introduced and discussed in the next section (Figure 2.10, Table 2.4 and Table 2.5). The results that are expressed as factors of safety (FoS) show a bandwidth of possible outcomes, related to each individual factor of influence. Additional analyses on combinations of possible causes which are suspected to have occurred are executed. Thus an overview with one or more dominant contributors is developed. When the related FoS is below a value of one, the situation is considered to be unstable and the scenario is identified as a possible cause of failure.

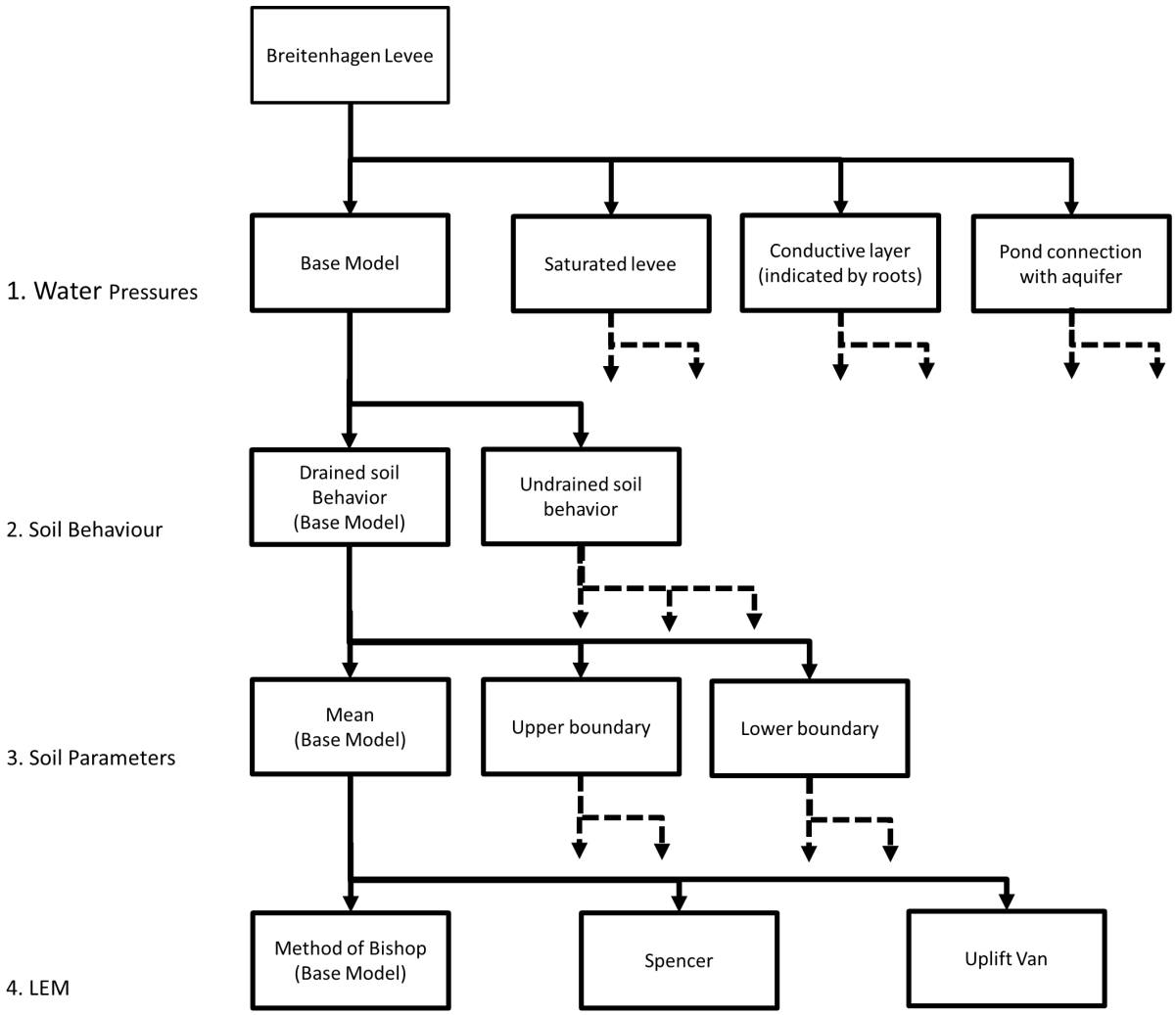


Figure 2.9. Overview of all factors of influence to vary in the sensitivity analysis one by one. Level 1.: Water pressure scenarios. Level 2.: Soil reaction behaviour scenarios. Level 3.: Soil parameter scenarios. Level 4.: LEM scenarios.

2.3.4 Scenarios for analysing the performance behaviour of the Breitenhagen levee

In this section, the scenarios of water pressure and hydraulic head in the subsoil, soil strength in terms of soil parameter, soil reaction behaviour scenarios, or LEM as shown in Figure 2.9, are further elaborated and quantified.

Pore water pressures

The intrusion of water at high water levels inside the levee causes the strength of the levee to decrease. The process is best described by the two sub-processes i.e. infiltration and groundwater flow, which initiate local high water pressures inside the levee. High water pressures lead to low effective stresses and result in local shear strength reduction. The water pressure in the base model (A) is modelled with a phreatic line in the levee body and water pressures in the aquifer (Figure 2.10). The analysis of the collected data shows that there are three possible scenarios next to the base model (A): increased rising of the phreatic line (scenario B.:" Saturated levee", modelled as 'high' phreatic line), a conductive layer as a result of tree roots growth (scenario C.:"Conductive layer, indicated by roots"), and high water pressure under the levee due to a connection of the outside water level and the aquifer (scenario D.:"Pond connection with the aquifer") (Figure 2.10). In the

sensitivity analysis, the input uncertainty is taken into account for scenario B. and D. by introducing the extreme upper, best estimate and extreme lower boundary value of water pressure development (Figure 2.10). The four water pressure scenarios are further elaborated below the figure.

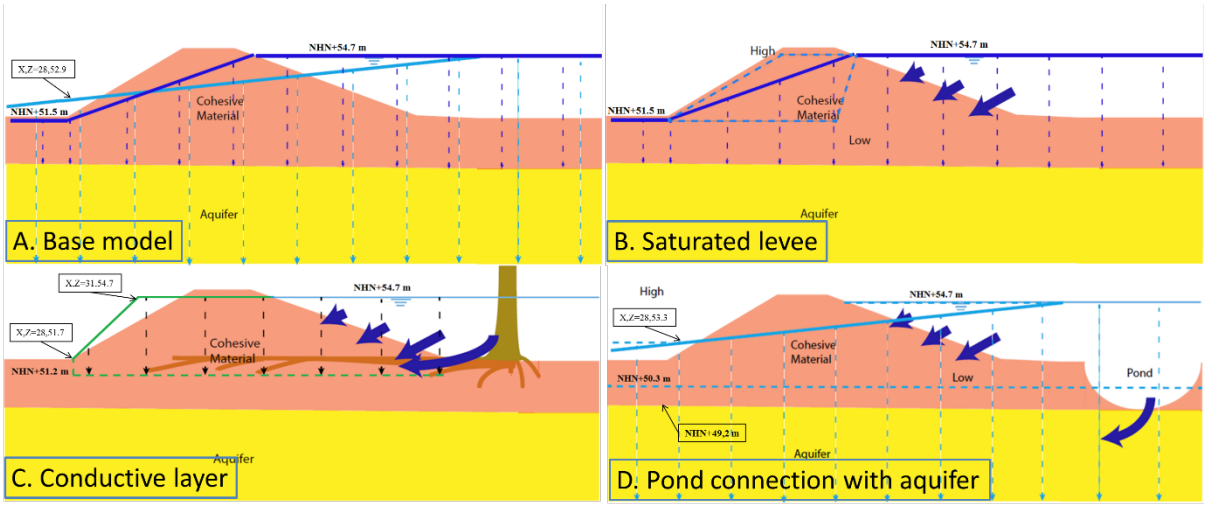


Figure 2.10. View of the four possible scenarios that are derived from the data analysis. Scenario A.: Best estimate of the phreatic line inside the levee (solid dark blue) and the best estimate hydraulic head in the aquifer (solid light blue), called "Base model". The river head (NHN+54.7m) and locally heightened polder head (NHN+51.5m). Scenario B: Infiltration of water through the levee slope which is called "Saturated levee" (dashed light blue lines); Scenario C.: Infiltration through a conductive layer related to the tree roots which is called "Conductive layer, indicated by the tree roots" (situated at NHN+51.2m)(the hydraulic head in green is the head at the location of the roots, the phreatic line is the same as the Base model. Inside the levee, water pressures in the levee are interpolated between the phreatic line and the green line); and Scenario D.: Infiltration through the connection between the outside water level and the aquifer which is called "Pond connection with aquifer" (polder head of NHN+51.5m) with an upper and lower head (dashed light blue lines);.

The analyses and the assumptions of the Grubert report (2013) are used to model the expected phreatic line and the hydraulic head inside the aquifer. Observations indicate that the phreatic line equals the surface level at the landside at the time of failure. The phreatic line is interpolated linearly between the river level and the polder head. At the same time, the hydraulic head in the aquifer beneath the layer might increase due to a connection between the river level and the aquifer. This causes a local increase in water pressure in the aquifer compared to the water pressure corresponding to the hydrostatic phreatic line.

The water potential inside the aquifer is limited by the weight of the blanket layer, which might otherwise cause uplift. The hydraulic head of scenario A., i.e. the "Base model" (best estimate) is fitted to the weight of the covering soil, at the toe of the levee. The water pressure inside the levee is interpolated between the phreatic line and the hydraulic head at the interface between sand and clay.

In the scenario of the "Saturated levee" (scenario B.), the only phenomenon that is taken into account is the development of the phreatic line in the levee body (Figure 2.10). The development of the phreatic line in the levee itself is caused by the infiltration of the high water wave from the river in the levee body. Due to the low permeability of the cohesive material the infiltration of the water in the levee is a relative slow process. The intrusion length of the high water wave is limited by the width of the levee. The extreme low and high boundaries of the phreatic line are based on standard phreatic models to represent infiltration (upper extreme boundary) and transient (lower extreme boundary) flow are suggested by literature (TAW, 2004). The models incorporate the influence of the dominant factors that influence the phreatic lines, i.e. infiltration and groundwater flow in order to simulate the transient flow and the different possibilities of steady-state flow.

In the scenario of the "Conductive soil" (scenario C.), the roots inside the levee indicate a more conductive layer due to a relatively large permeability (Figure 2.10). Therefore the intrusion

length of the water wave is increased in this layer. This causes a local increase in the water pressure. Moreover, the velocity of the high water pressure wave in the conductive layer is relatively fast compared to the levee body. In this area, the water potential equals the outside water level. At the inside toe of the levee, the potential inside the rooted area is limited by the covering soil. The analysis of Scenario C. is executed as a direct relation to Scenario B. to show its influence on the upper, mean and lower bound of the phreatic line. The water pressure inside the levee is interpolated between the phreatic line and the hydraulic head inside the conductive soil layer.

In the scenario of the “Pond connection with aquifer” (scenario D.), the pond in front of the levee might introduce a connection between the aquifer layer and the outside water level (Figure 2.10). This causes a local increase in water pressure in the aquifer compared to the water pressure corresponding to the phreatic line. Because of a relatively large leakage length of the outside water level, the water pressures inside the aquifer might cause a shear strength reduction at the landside. The weight of the blanket layer limits the hydraulic head in the aquifer. The water pressure inside the levee and the blanket layer is interpolated between the phreatic line and the (limited) hydraulic head at the interface between sand and clay.

Geotechnical shear strength

The shear strength of the soil is introduced for both drained and undrained response behaviour is used to determine the amount of shear stress that the soil can resist. Besides the drainage rate of the soil, the shear strength of the soil depends on the values of the geotechnical input parameter.

The drained soil response behaviour is introduced by the Mohr-Coulomb model and depends on three parameters, i.e. cohesion (c’ in kPa), soil friction angle (φ in deg) and the effective vertical stress (σ’ in kPa). The undrained soil response behaviour for low permeability materials is introduced with help of the SHANSEP implementation (Ladd, 1991) of the Critical State Soil Mechanics (Schofield & Wroth, 1968) (2-1):

$$s_u = \sigma'_{v,i} SOCR^m \text{ with } OCR = \sigma'_{vy}/\sigma'_{v,i} \text{ and } \sigma'_{vy} = \sigma'_{v,i} POP \quad (2\text{-}1)$$

Where s_u is the undrained shear strength ratio in kPa, $\sigma'_{v,i}$ is the in-situ effective vertical stress in kPa, S is the undrained shear strength ratio, OCR the Over Consolidation Ratio, m the strength increase exponent, σ'_{vy} the vertical yield stress in kPa, and POP the pre-overburden pressure in kPa.

The shear strength of the soil is influenced not only by the pore water pressures but also by the values of the geotechnical input parameters. The parameters are uncertain (Figure 2.5 and Table 2.2). Therefore values of the parameters are based on values coming from Dutch literature (Deen & Duinen, 2016; Normcommissie, 2011; RWS, 2016b). In this sensitivity analysis the input uncertainty is taken into account by applying the extreme upper, mean and extreme lower boundary value for the soil parameters related to the identified soil type, for both drained and undrained soil behavior (

Table 2.4 and Table 2.5). The coarse-grained soil (sand and gravel) is expected to behave drained and is implemented as such, also in the undrained analyses. Since the behaviour of the fine-grained soils is unknown, these are modelled as both drained and undrained. The sensitivity analysis does not consider individual layers of soil. The introduction of the upper, mean or low boundary values affects all soil layers at the same time.

Table 2.4. Assumed lower and upper bound of soil properties. Mohr-Coulomb parameters are used for the sensitivity analyses (extreme value for the soil weight: $\gamma_{up/low}$, expected values: γ_{mean}, extreme value soil friction angle: $\varphi_{up/low}$, expected values: φ_{mean}, extreme value cohesion: $c'_{up/low}$, expected values: c'_{mean}) (Grubert, 2013a; Normcommissie, 2011).

Layer Nr. and Soil type (Grubert, 2013)	Description of soil layer (Normcommissie, 2011)	γ_{mean} kN/m^3	$\gamma_{up/low}$ kN/m^3	φ_{mean} deg	$\varphi_{up/low}$ (deg)	c'_{mean} (kPa)	$c'_{up/low}$ (kPa)
Nr. 1. Organic mat Humus	Clay, clean and moderately stiff	17	20/14	21.25	25/17.5	7.5	15/0
Nr.2. Cohesive mat.	Clay, little bit of sand, stiff	18	21/15	25	27.5/22.5	7.5	15/0
Nr. 3. Alluvial	Clay, little bit of sand, stiff	18	21/15	25	27.5/22.5	7.5	15/0
Nr.4. Sand and gravel	Gravel, bit silty, clean	18,5	22/19	36.25	40/32.5	0	0/0

Table 2.5. Typical parameter values of undrained soil behaviour (Critical State Soil Mechanics) that are used for the sensitivity are very common in the Netherlands with similar soil description (where $S_{low/up}$ are the extreme boundary values of undrained shear strength ratio, S_{mean}, are the expected values, $m_{low/up}$ are the extreme boundary values of strength increase exponent, m_{mean} are expected values, $POP_{low/up}$ are extreme boundary values of pre-overburden pressure, POP_{mean} are expected values) (Deen & Duinen, 2016; Ladd, 1991; RWS, 2016b).

Layer Nr. And Soil type (Grubert, 2013)	Description of soil layer	$S_{low/up}$	S_{mean}	$m_{low/up}$	m_{mean}	$POP_{low/up}$ kPa	POP_{mean} kPa
Nr. 1. Organic mat Humus	Sandy and Silty Clay	0.22/0.50	0.30	0.5/1.00	0.9	0/75	22
Nr.2. Cohesive mat.	Levee material	0.23/0.50	0.31	0.5/1.00	0.9	0/150	30
Nr. 3. Alluvial	Levee material	0.23/0.50	0.31	0.5/1.00	0.9	0/150	30

2.4 Breitenhagen levee failure: sensitivity calculations

With the chosen scenarios introduced in the previous section a sensitivity analysis in order to find the most likely causes of the failure (Figure 2.9). In the following paragraphs, the results related to the scenarios of water pressures are presented, followed by the influences of the soil response behavior (drained and undrained), the influence of the different shear strength parameters and the influence of the different LEM. In order to highlight the potential of further research, the results of the more prominent combinations of possible causes are presented as well. In total 32 scenarios and different combinations are taken into account. The model of the best estimated situation represents the base of the sensitivity analysis (referred to as “base model) and is presented in the first paragraph. In the last paragraph, an overview of all results is given.

2.4.1 Base model

The “base model” (scenario A) represents the best estimate scenario together with the expected values of input parameters. The “base model” functions as the basis of the sensitivity analyses (Figure 2.9). The LEM calculation is executed using the method of Bishop and incorporating the drained soil behaviour. The method of Bishop is used as the best estimate since this method reflects the circular observed slip surface relatively well. The results of the simulation of the best estimated

situation show a value of FoS=1.61, which is considered stable (FoS>1.0). The shape of the slip surface shows realistic similarities with the actual slip surface (Figure 2.11). However, the slip surface starts at the riverside of the top of the levee, instead of the inside. Also, the slip surface intersects with the sand layer, which is not considered realistic.

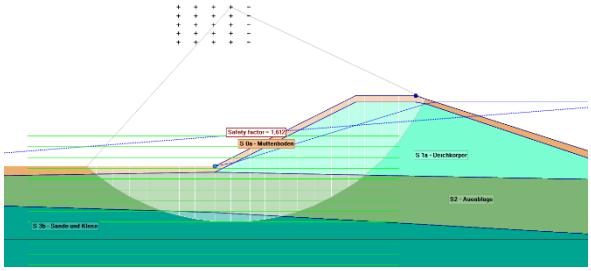


Figure 2.11 View of the LEM calculation using the Bishop stability model with drained soil behaviour (FoS=1.61).

2.4.2 Water pressure scenarios

The development of the water pressure inside the levee is hard to determine but has a significant influence on the strength of the levee (Figure 2.9 at “level 1”). There are three different scenarios of pore pressure development incorporated into the sensitivity analysis, i.e. scenario B.: “Saturated levee”, scenario C.: “Conductive layer (related to the roots)” and scenario D.: “Pond connection with aquifer” (Figure 2.10). The different scenarios are introduced to simulate the transient flow and the different possibilities in steady state flow at level 1 of the flow chart (Figure 2.9 “level 1”). All scenarios are assessed with the help of LEM using Bishop’s method and a drained analysis.

Scenario B.:”Saturated levee” and Scenario C.:“Conductive layer”

The analysis of the “Saturated levee” uses 3 different possible scenarios of water pressures to incorporate the uncertainty in transient flow and the contributors in the phreatic line (rain, groundwater, high water wave, etc.) e.g. convex, straight and concave (Figure 2.10). The analyses with the straight phreatic line is similar to the base model that is introduced as the base case (Figure 2.9). The analysis of the “Conductive layer (indicated by roots)” is executed as a direct relation to the saturation of the levee as described in section 2.3.4.

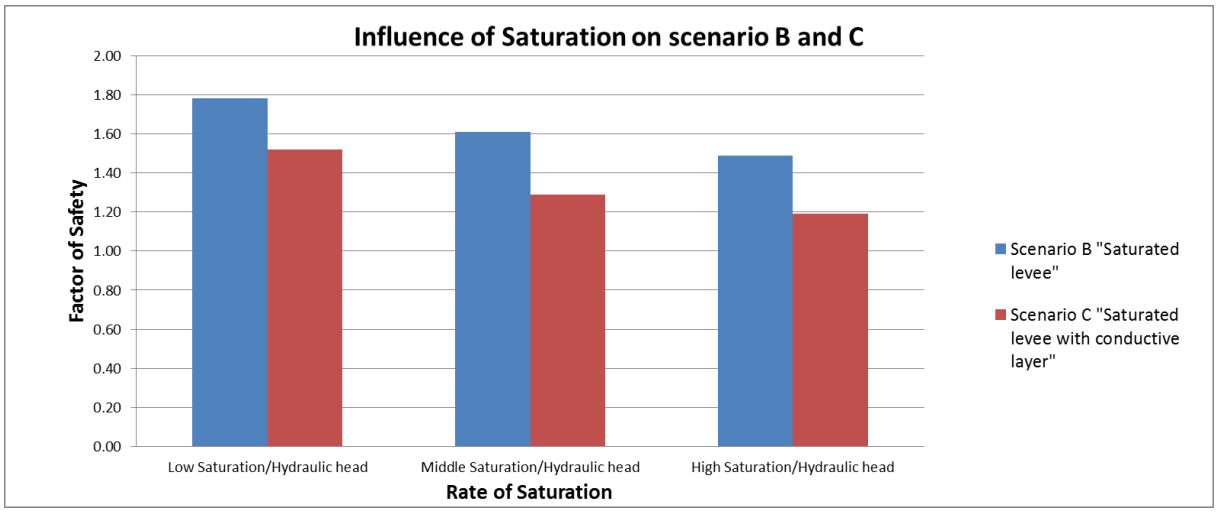


Figure 2.12. Blue: Influence of the saturation of the levee on the FoS. Red: Influence of the saturation when the local high water pressures in the conductive layer is introduced.

The FoS is negatively correlated with an increase in water pressures because the effective stresses are reduced, which in turn causes the shear strength to decrease. Due to the rising phreatic line, the range of FoS is between 1.78 and 1.49 (Figure 2.12). The analysis of the increased potential head in the conductive layer around the tree roots results in FoS of 1.52, 1.29 and 1.19. The lowest

factor of safety is found for the highly saturated cross-section in combination with the conductive layer. All models, in any case, yield safe values for the FoS.

Scenario D.: "Pond connection with aquifer"

The stratigraphy shows an aquifer below the levee which might be in direct contact with the water level at the riverside by means of a pond in front of the levee (Figure 2.10). Similar to the phreatic line, the different water pressures in the aquifer influence the shear strength as well. For the analysis of the possible contribution of the aquifer to the cause of failure, the model uses three scenarios for the water pressure as described in section 2.3.4. The results are shown in Figure 2.13.

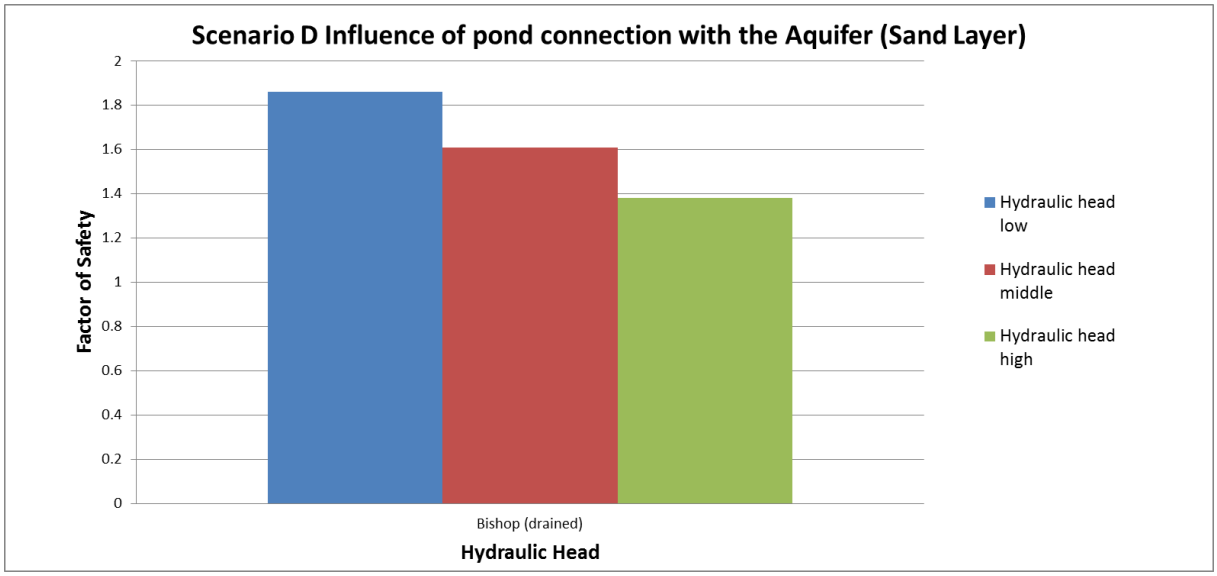


Figure 2.13. Influence of the rising of the hydraulic head in the aquifer with 3 different levels in hydraulic head in scenario D.

Due to the rising of the hydraulic head in the aquifer the range of the FoS overall is between 1.86 and 1.38. The shape of the slip surface shows realistic similarities with the actual slip surface. However, the slip surface still intersects the levee crest instead of starting from the landside of the levee.

Introducing undrained soil behaviour for the clay, scenario D., shows results of FoS between 1.66 and 1.28. The slip surface of the actual failure (Figure 2.6) and the calculated slip surface (for example with the use of Bishop with undrained soil behaviour and Spencer with drained soil behaviour) differ in geometry. The slip surface of the Spencer method seems to show a more realistic geometry of the slip surface than the Uplift Van method and Bishop's method in case of low effective soil pressures at the interface between sand and clay (Figure 2.14).

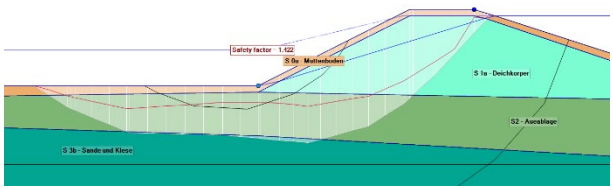


Figure 2.14 Screenshot of scenario D.: "Pond connection with aquifer" where the high hydraulic head is included and the Spencer model is used (with drained soil behaviour results in FoS=1.42).

The joint occurrence of a saturated cross-section (high phreatic line) and a high water pressure in the aquifer are considered with help of Bishop's method and both drained and undrained

soil response. Assuming drained soil behaviour results in a FoS of 1.28 and assuming undrained soil behaviour results in a FoS of 1.20 with both realistic and similar slip surface shapes in comparison with the actual slip surface (except for the intersection of the sand layer with the slip surface). In addition to earlier calculations, a combination of scenario 1,2 and 3 is taken as extreme and combined into one scenario (with the Bishops' method). Assuming drained soil behaviour results in a FoS of 1.23 and assuming undrained soil behaviour results in a FoS of 1.21. Both calculations show realistic slip surfaces (except for the intersection of the sandlayer with the slip surface).

2.4.3 Influence of the soil model

Since the cross-section of the levee mostly consists of clay, the influence of undrained soil response behaviour of the cohesive soils (Verruijt, 2010) is assessed relative to the "base model". The calculations are executed using the method of Bishop and incorporate the undrained shear strength with the expected values of the input parameters, see section 2.3.4. The result of the simulation shows a value of FoS of 1.50, where a FoS=1.61 is found when the "base model" incorporates drained soil behaviour. The shape of the slip surface shows realistic similarities with the actual slip surface similar to the "base model" (Figure 2.11). However the slip surface starts at the riverside of the top of the levee, instead of on the landside of the levee and shows an intersection of the sandlayer with the slip surface.

2.4.4 Influence of the soil parameters

The analysis of the extremes of the soil parameters to establish the influence of each individual input parameter and the influence on the performance behaviour is compared to the "base model" (Figure 2.9 at level 3). The contribution of the weight (for both drained and undrained soil model) and the internal friction angle (for drained soil model) are very limited, implying that they do not have a dominant contribution in the failure and therefore are excluded from further analysis. The sensitivity analysis is performed using Bishop's LEM.

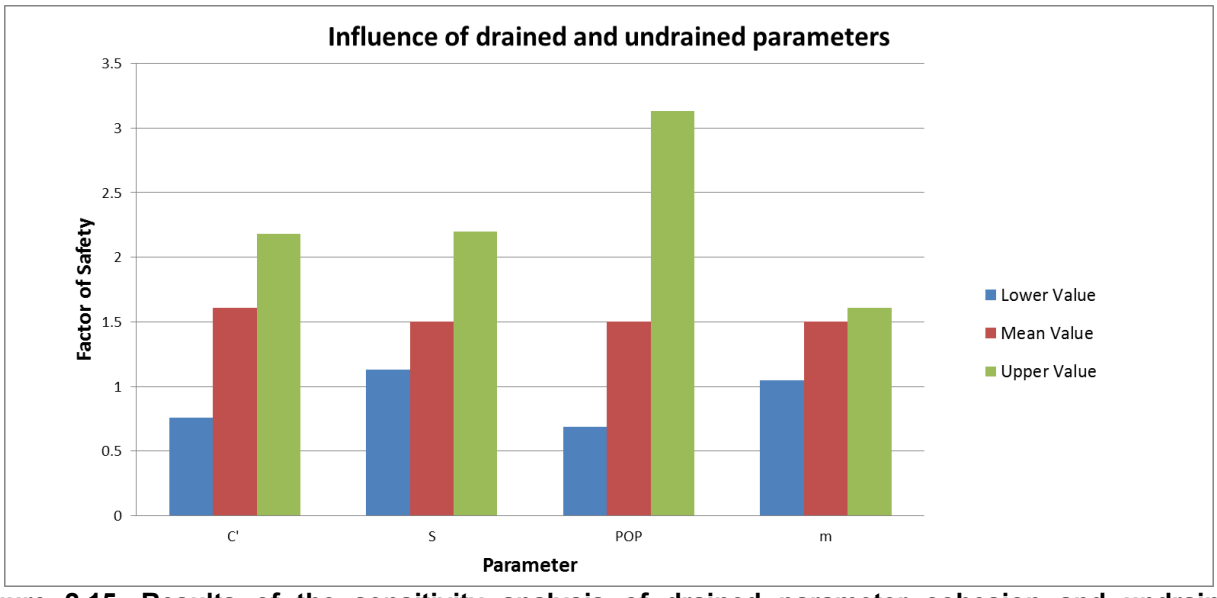


Figure 2.15. Results of the sensitivity analysis of drained parameter cohesion and undrained parameters.

Varying the cohesion, which is the dominant parameter of the drained soil model, the FoS covers a range between 0.76 and 2.18 (Figure 2.15). The undrained shear strength has high variability, due to the uncertainty of the previous stress history and related overconsolidation state (the POP, pre-overburden pressure, was used to describe this uncertainty). The results of the calculations greatly reflect this uncertainty, with the calculated FoS ranging from 3.13 and 0.69. The

parameters S and m, used to model undrained shear strength, give large variation in FoS but do not result in values below 1, which would represent the failure of the levee. The calculation using the low cohesion values and the Bishop model shows similarities with the actual slip surface and do not reach the sand layer underneath the cohesive layers (Figure 2.16) .

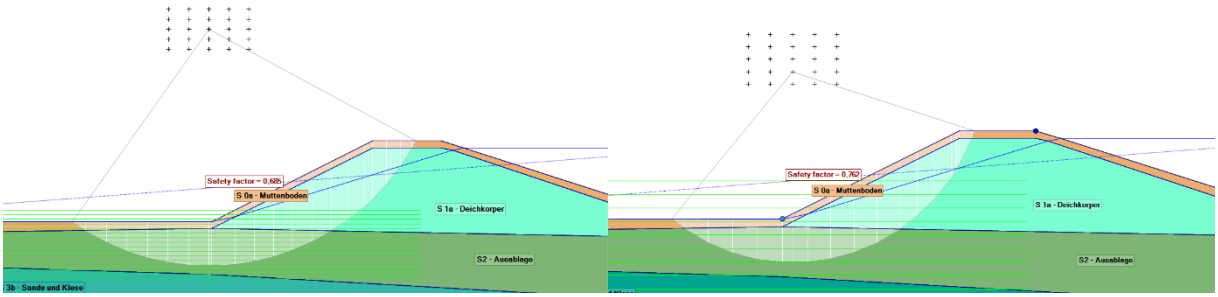


Figure 2.16 Left: Screenshot of scenario A. "Base model" using the Bishop slip surface and including low values of POP and shows FoS=0.69 (undrained soil behaviour). Right: Scenario A. "Base model" using the Bishop model and including low values of cohesion and shows FoS= 0.76 (drained soil behaviour).

The additional calculation run with a combination of low values of cohesion with high water pressures in the aquifer results in a FoS of 0.55, with Bishop's LEM. The resulting slip surface seems shallow compared to the actual observed failure surface.

2.4.5 Limit equilibrium method

Three different LEM are used in the assessment on the levee strength, i.e. Bishop, Uplift Van and Spencer (Figure 2.9 level 4) in the D-Geo Stability software (Deltares, 2016). All three LEMs are applied to the "base model". The sensitivity analysis is used to establish the influence of every individual LEM. The FoS determined with the help of the Bishop, Uplift Van and Spencer model covers a range between 1.55 and 1.61. All models result in approximately the same slip surface (Figure 2.11, Figure 2.17). The Spencer method shows the most realistic geometry of the slip surface (considering the slip surface doesn't intersect the sand layer).

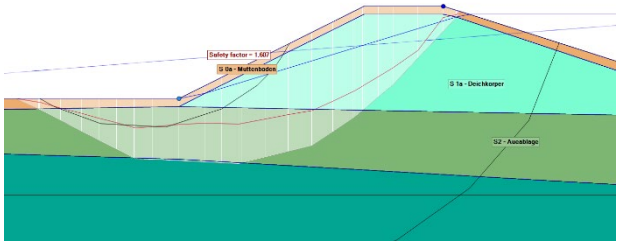


Figure 2.17 Screenshot of scenario A. "Base model" using the Spencermodel presenting the calculated critical slip surface (FoS=1.61).

2.4.6 Results of the sensitivity calculations

The FoS of the individual causes and additional analyses with combinations of scenarios on the stability of the levee are presented in Table 2.6. The results are presented in terms of upper, best estimated and lower boundary values of FoS (Table 2.6). Four types of uncertainties are considered e.g. water pressures, soil reaction behaviour, soil parameters and LEM (Figure 2.9).

As expected, the upper bounds of the water pressures in the scenarios of "saturated levee" and "pond connection with the aquifer" are related to lower values in FoS. All water pressure-related

scenarios of failure seem unable to create an unstable situation as an individual cause but do have a significant influence.

Lower values of cohesion and POP lead to FoS values that are significantly lower than 1. The calculations suggest that the individual cause of low shear strength can result in calculative slope instability. Moreover, the shape of the slip surface (Bishop) associated with the low value of cohesion shows large similarities with the shape of the actual slip surface. For the scenario with the low value of POP a slip surface with less resemblance is obtained.

The use of difference LEMs does not lead to significantly different results. Although the failure plane of the LiftVan model intersects the interface of the sand and clay layer and therefore shows non-realistic results. This is probably due to the skewed interface between the sand and clay layer and the high values of the hydraulic head.

The analyses focus on combinations of individual possible causes that show low results of FoS and result in calculative instability (FoS<1). The results of the combinations of causes show that the combination where low values of POP or cohesion are included can result in calculative instability. Especially when the low values of cohesion are combined with the presence of the conductive layer associated with the tree roots and the pond connection with the aquifer, predicts an instability situation. When the dominant individual factors are combined the FoS drops to a value of 0.16. However, the shape of the slip surface, associated with the low cohesion and the high water pressure values in the aquifer is very shallow and does not show similarities with that of the actual slip surface.

Table 2.6. Overall summary of results over the upper limit and the lower limit boundary in FoS per factor (uncertainty). The best estimated situation is calculated with the help of the Bishop method with drained soil behaviour (Figure 2.9). *Calculated slip surface matches the shape of the actual slip surface.shows a slip surface that intersects the sand layer**

	Uncertainty	Upper FoS	Best estimate situation	Lower FoS
Base model	Scenario A.: Best estimate situation (drained model, Bishop)	-	1.61**	-
Water pressures (Level 1.)	Scenario B.: Saturated levee (drained model, Bishop)	1.78**	1.61**	1.49**
	Scenario C.: Increased pore by conductive layer (drained model, Bishop)	1.52	1.29	1.19
	Scenario D.: Aquifer connection with pond (drained model, Bishop)	1.86	1.61**	1.38**
Soil behaviour (Level 2.)	Soil model Un-drained (Bishop)	-	1.50**	-
Soil parameters (Level 3.)	Cohesion (c') (Bishop)	2.18**	1.61**	**0.76***
	Undrained Shear strength ratio (S) (Bishop)	2.20**	1.50**	1.13
	Strength increase exponent (m) (Bishop)	1.61**	1.50**	1.05
	Pre-overburden pressure (POP) (Bishop)	3.13**	1.50**	**0.69**
LEM (Level 4.)	Spencer	-	1.61	-
	Uplift Van	-	1.55**	-
Combinations	High phreatic line and high aquifer connection (drained, Bishop)	-	-	**1.28****
	High phreatic line and high aquifer connection (undrained, Bishop)	-	-	**1.20****
	Increased pore by conductive layer (undrained, Bishop)	-	-	1.33
	High phreatic line, conductive layer and aquifer connection (drained, Bishop)	-	-	**1.23****
	High phreatic line, conductive layer and aquifer connection (undrained, Bishop)	-	-	**1.21****
	Aquifer connection (drained model, Spencer)	-	-	**1.42**
	Aquifer connection (drained model, Uplift Van)	1.85	1.55**	**1.34****
	Aquifer connection (undrained model, Bishop)	1.66	1.50**	**1.28****
	Low value in cohesion and high aquifer connection (drained, Bishop)	-	-	**0.55**
	Low value in cohesion, high phreatic line, conductive layer and aquifer connection (Bishop)	-	-	**0.16**

2.5 Discussion

In the next sections, the newly developed generic approach of forensic analysis of levee failures and the results of the Breitenhagen case are separately discussed.

2.5.1 Proposed generic framework for forensic analysis of levee failures

The proposed framework intends to provide a generic approach to forensic analysis of failed levees, to explicitly and transparently account for the most relevant uncertainties and modelling decisions. The systematic approach enhances the process of logical deduction in forensic analysis of levee failures dealing with significant uncertainties since the information prior to the breach is typically limited. In addition, obtaining more information after the breach is also limited since part of the information is washed away. Part of the forensic analysis is the simulation of the event, called hindcasting. The models that are used to simulate the event introduce input parameters- and model uncertainties. The proposed approach is successfully applied to the Breitenhagen levee failure; however, it might need adaptations for future use, once multiple breaches have been assessed using this approach.

The arrangement of the information in chronological order (prior, during and after the failure) provides a start for a structured overview of the evidence. Moreover, it gives good insights into the levee history and, the behaviour during the failure event. By isolating and indicating the typical relevant data, an assessment of available and missing data is achieved. Also, the overview of evidence demonstrates the consistency between the identified most likely possible causes, the earlier collected evidence and the remaining uncertainties of input parameters and model uncertainties that are involved.

Hindcasting of geotechnical structures is typically dominated by several input and model uncertainties. The introduction of all possible scenarios of individual causes and combinations of possible causes, enables hindcasting that simulates the actual situation, and identifies the most likely causes. The proposed approach makes all uncertainties explicit but doesn't quantify the uncertainties that reflect on the likelihood of occurrence. The likelihood of occurrence has to be estimated by the forensic engineer to come to a final answer. Therefore the actual situation is not necessarily objectively reflected in the outcome, since it is possible to have multiple outcomes in terms of likely causes.

Analysing all possible combinations of lower and upper bound of parameters results in many scenarios and is very time-consuming. Limiting the analysis to the individual possible causes and suspected combinations of causes as possible scenarios gives important insights into the individual contribution of the performance behaviour and is less time-consuming. However, it does not reflect possible correlations between contributing factors when taking only the individual possible causes into account.

As suggested in the approach itself (step C, Figure 2.8), the next step would be to weigh the outcome based on the likelihood. However, as no probabilities of scenarios are estimated, this is not possible within this approach. To take more information into account, more advanced probabilistic approaches are recommended. This would make it possible to introduce the visual observations into the analysis as well, further limiting the input parameter uncertainties and model uncertainties.

2.5.2 Application to the Breitenhagen levee failure

The forensic analysis approach was developed in conjunction with the Breitenhagen analysis. Hence, challenges encountered during the approach such as the sensitivity of the results to shear strength parameters are incorporated into the results. That said, the approach might need small adaptations in case of different available datasets, while the general framework is expected to be sufficiently robust.

The chosen limits of soil parameters are used to incorporate the uncertainty in the soil parameters. However, the chosen extreme values of soil parameters dominate the outcome, especially since all layers (levee body and blanket) are assumed weak simultaneously. This can put too much emphasis on locally weak soils and individual weak soil layers should be considered as well. Since most shear strength is generated by the levee body, considering individual layers is not expected to influence the outcome of the analysis much. The values of parameters of the drained soil behaviour are based on Dutch literature, the chosen values might be very conservative and justify further research. German literature suggests larger values of cohesion (Schneider & Albert, 2014) and therefore shear strength is expected to be less dominant for the Breitenhagen case. However, very low values of cohesion (c'~0 kPa) result in low values of FoS and indicate instability. These results seem obvious since the applied slope is steeper than the applied friction angle of the soil. The angle of the slope of the levee was also mentioned in the earlier forensic engineering report of Grubert (2013a) as a point of attention. More data on local soil characteristics would reduce the uncertainty of the input parameters.

Which soil model is most appropriate to use in a stability analysis depends on many different factors. Whether the response of the soil was undrained or drained cannot be observed, since it is hidden inside the levee. However, the observed relatively slow process (more likely for drained behaviour) in combination with low permeability materials (more likely for undrained analysis), indicates a partially drained behaviour. Partially undrained soil behaviour is not taken into account in the type of computational models that are used in this section.

The hydraulic head in the conductive layer around the tree roots and the influence of these on the surrounding soil are assumed conservatively. For instance, it is very unlikely that the roots have sprouted through the levee and reached the other side of the levee, but it is expected that they are present in the levee as this is shown in Figure 2.7. The measurement of the actual development of the hydraulic head and the range of the influence of the conductive layer could limit the influence of pore water pressures resulting in a more accurate simulation of the influence on the levee behavior.

The models that are used to predict the performance behaviour are developed under controlled conditions and are little validated to predict the performance behaviour of a levee under extreme conditions or when confronted with variability and heterogeneity of the soil. Using the scenarios approach gives insights into which model simulates the most accurate and limits the uncertainties as much as possible. More advanced Finite Element Method (FEM) based models would possibly be able to describe the behaviour better and to reduce the uncertainty, but advanced models are very time-consuming and introduce other uncertainties. Furthermore, the used LEM models reflect the failure mode sufficiently well.

The proposed approach is working well for the Breitenhagen levee failure partly because it concerns an unexpected instability (FoS>1.0 for the best estimate conditions). This implies that possible failure cases can be found by searching for combinations of parameters that result in a FoS<1.0, which results in a limited number of combinations for this case study. In situations, where the best estimate situation results in a FoS closer to, or below 1.0, differentiating between the possible causes (using the criteria of FoS<1.0) becomes more complex since many possible causes will result in FoS<1.0. Moreover, the cause resulting in the lowest FoS is not necessarily the most likely cause. In this case, probabilistic methods would be a useful addition.

2.5.3 Likely causes of the Breitenhagen levee failure

The sensitivity analysis identifies two most likely causes of instability of the slope of the levee which are low values of POP and cohesion, both related to weak soil. The terms of low values of POP and cohesion result in low values of FoS (FoS<1.0). The high resemblance between the calculated shape of the slip surface when analysing the influence of the cohesion and the actual slip surface reinforces these findings. The influence of the hydraulic head in the aquifer especially in combination with high

saturation of the levee (tree roots) is significant. The collected data shows that at this particular location the levee near Breitenhagen has most likely been breached before (Sixdorf, 2016). As a result of the repairs of the breach, the soil conditions might deviate from the other soil conditions of the levee and are suggested to be weak. It seems that the pond in front of the levee is a leftover of this former breach. The relation between the breach, repairs, weak soil and the high hydraulic head in the aquifer would explain the occurrence of the breach at this location in 2013. Although the breach and the repairs do not support the contribution of the conductive layer around the tree roots, they do not exclude the conductive layer around the roots as a possible contributor to failure.

Earlier forensic engineering analysis by (Grubert, 2013a) identified a combination of causes of failure which are similar to the findings of this analysis. The conclusion of the earlier forensic engineering analysis suggests that the cause of the breach relates to the high saturation of the levee, the relative steep angle of the levee and the high conductive soil layer that is associated with the tree roots. The conclusion does not explicitly address the pond connection with the aquifer or weak soil conditions, although the calculations by Grubert (2013a) seems to acknowledge the connection of the outside water level and the aquifer and assume a direct connection with the outside water. Here, special attention is paid to the possibility of uplift, but this is indicated to be secondary to the influence of the tree roots. The present study supports the findings of Grubert (2013a) and added additional insights in different scenarios of possible causes of failure and in how these relate to the history and the collected evidence. Which findings are more likely, can not be concluded from this analysis.

2.6 Conclusion and recommendations

The developed approach of forensic analysis in this section has been applied to the levee failure near Breitenhagen to identify the most likely cause of failure. The collected evidence is assessed, the results of the sensitivity analysis are discussed. Conclusions and recommendations are presented below.

2.6.1 Conclusions

This article suggests a newly developed generic approach to forensic analysis of levee failures. The approach includes the most relevant uncertainties and modelling decisions explicitly and transparently. The approach enables the engineer to systematically generate possible realizations of reality. Also, the approach provides insight into the uncertainty of input parameters and model uncertainties explicitly by introducing all possible causes as scenarios and by validating or falsifying each scenario. The forensic analysis approach requires to analyse all possible scenarios including all combinations of possible causes, thus making this approach very time consuming. The forensic analysis approach does not quantify the uncertainties, as in a probabilistic manner, and does not prioritize the most likely causes based on the probability of occurrence.

In this case study, the results indicate that locally low strength associated with low values of POP or cohesion justify the failure. Other scenarios analysed resulted in either a FoS that does not justify failure or, show a slip surface that differs from the observed failure surface. High hydraulic head in the aquifer and unexpected saturation of the dyke body are considered to have contributed significantly in bringing the levee to instability. Both the presence of a pond and locally weaker soil may be justified by an old levee breach (section 2.5) previously occurred at the location of the breach in 2013 (Sixdorf, 2016). Historical data can not confirm or rule out the contribution of unexpected high saturation, reducing the resistance of the levee, that is related to the presence of tree roots.

2.6.2 Recommendations

The generic approach is developed for the purpose of forensic analysis of levee failures in general. It is advised to apply the approach to another type of levee failure mechanism for the purpose of further development of the forensic engineering approach to geotechnical failure cases.

In future forensic engineering analyses of levee failures, it is recommended to use a probabilistic analysis which is capable to quantify the input parameter and model uncertainties and make them more explicit by incorporating the likelihood of each scenario. This approach takes the correlation of all contributors into account and enables to explore of all possible combinations of probable causes. This will increase insight into the collection and the actual weight of every contributor and support an unambiguous outcome.

Furthermore, there are several refinements possible for the Breitenhagen analysis. The uncertainty regarding shear strength parameters, and the resulting high used band-width, may be reduced by performing local field test (e.g. CPTs) or collecting samples for lab tests. This cannot be done on the failure soil, since that has washed away, but local soils should give a better approximation and lower band-width of the parameters than the current used default values for Germany and the Netherlands Also, the results show that it is essential to include observational information in the analysis to identify the most likely scenario explaining the failure. It is advised to explore whether it is possible to have more information using Bayesian techniques in the hindcasting, e.g. past performance information (Schweckendiek, van der Krogt, Teixeira, et al., 2017). Discerning drained from undrained response seems not feasible in most breach cases. Investigating the influence of a time-dependent, partly drained/undrained, response could give better insight into the “true” failure conditions.

Applying parameters that are more representative for the actual location would limit the influence of the input parameter uncertainty on the result. This can be done by doing more field tests. But also consultation with local authorities and companies would lead to better assumptions for input. Especially when better insights are available on the possible values of cohesion and POP and values of the potential head in the aquifer and the phreatic line. However, this is not expected to change the overall conclusions of this case much.

Acknowledgements

We thank Dr.-Ing. Thilo Weichel of the Landesbetrieb für Hochwasserschutz und Wasserwirtschaft Sachsen-Anhalt for providing us with his drone movie of the breach of the levee near Breitenhagen and other relevant data.

This research was performed as part of the NWO TTW project SAFElevee (project number 13861)

3 A Bayesian hindcasting method of levee failures applied to the Breitenhagen slope failure[4]

3.1 Introduction

Forensic analysis provides a systematic procedure of analysis for the investigation of failures. The procedure of forensic analysis roughly consists of three stages: 1) collecting and reviewing of evidence, 2) utilizing calculative models for a back analysis (hindcasting) to identify the cause, and 3) reporting the findings (Carper, 2000). In previous work, it appeared that hindcasting using a deterministic sensitivity analysis identifies several possible causes of failure (Kool et al., 2019; Zhang et al., 2010). However, the deterministic sensitivity approach of hindcasting did not provide an explicit insight into the likelihood of various failure scenarios.

In order to enhance levee hindcasting, probabilistic Bayesian techniques are used in this section to identify the most likely scenario of failure, the most representative model choices to characterize the failure, and the most dominant parameters triggering the failure. The prior and posterior probabilities reflect the probability that a model best describes the reality and that a scenario of failure was present given the evidence. The combination of the most likely scenario of failure and most representative model choices is referred to as "MLC" in the remainder of this section. In the past, Bayesian techniques have been used to characterize the lifetime reliability of geotechnical structures (Baecher, 2017; Schweckendiek, 2014b; Zhang et al., 2010), and have been applied to the hindcasting of different types of geotechnical related structures (Gilbert, 2016; Gilbert et al., 1998; Luckman et al., 1987; Zhang et al., 2010). However, to our knowledge, Bayesian techniques have not been applied to hindcasting of specific levee failures. Another novel element in the proposed approach in this section is that Bayesian updating is carried out using specific information on the geometry of the observed failure, in this case, the slip surface. This approach allows us to account for the scarcity of evidence, as the available information after failures is generally limited to a small number of observations, photos and videos.

The article is structured as follows: The background of the method for the hindcasting of levee failures using Bayesian techniques is presented in section 3.2. In section 3.3 the case study of the levee failure near Breitenhagen (2013) is analysed using the Bayesian hindcasting method. The results are discussed in section 3.4, and the conclusion and recommendations in section 3.5.

[4] This chapter has been published as: Kool, J., Kanning , W., Jommi, C., & Jonkman, S. N. (2020). A Bayesian hindcasting method of levee failures applied to the Breitenhagen slope failure. *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*. https://doi.org/10.1080/17499518.2020.1815213

3.2 Probabilistic hindcasting of slope instability using Bayesian updating

3.2.1 Method background for Bayesian levee hindcasting

The proposed approach for hindcasting levee failures consists of defining scenarios that could explain the failure and assigning prior probabilities of occurrence to these scenarios. Subsequently, the probability of failure given a scenario is calculated (which is used as likelihood) and Bayesian updating is used to update the prior probabilities with the likelihood and failure observation (see e.g., (Gilbert et al., 1998; Schweckendiek, 2014b)). The objective is to find the most likely scenario (S_i) that has resulted in the failure event (F).

Prior probabilities

The first step of the proposed hindcast approach is to use the collected information for the identification of possible scenarios (S_i) that could have resulted in the failure of the levee. A scenario typically represents the loading conditions possibly triggering the failure as well as the subsoil conditions (e.g. pore water pressures, stratigraphy). Prior scenario probabilities (i.e. $P(S_i)$) are assigned in such a way that they total a sum of 1 (e.g. (Schweckendiek, van der Krogt, Rijneveld, et al., 2017)).

Likelihood function based on failure and observed evidence

The Bayes' theorem makes it possible to incorporate evidence in the analyses, such as observed failure information (Schweckendiek, 2014a). To apply Bayesian updating of the previously assigned prior scenario probabilities, two pieces of information are incorporated using a likelihood function in the next steps, i.e. (A) the actual occurrence of failure and (B) the geometry of the observed slip surface (when available).

Hence, the likelihood function in step (A, referred to as likelihood A) is calculated by determining the probability of failure (F) per scenario $P(F|S_i)$ according to:

$$P(F|S_i) = P(Z(X) < 0|S_i) \tag{3-1}$$

In this equation, the limit state function Z is used to describe when the failure occurs. This is evaluated in this section by using a slope stability model. The joint probability density function f(x) of the random variables (X) is used to describe Z. Negative values of the limit state function (Z) describe which combinations of random variables (X) combination results in failure of a scenario (S_i). Hence, the likelihood function A shows how likely it was that, according to a model, failure would have occurred given a certain scenario. And thus how likely it was that failure was observed given a certain scenario. This likelihood is expressed as a probability of failure given the scenario. This probability of failure is the results of all the uncertain input parameters (X) in the model.

The First Order Reliability Method (FORM) is used in this section to estimate the probability of failure, Equation (3-1). Subsequently, Equation (3-2) is used to find the design point at Z(X)=0 which is the combination of parameter values that provide the highest probability density (see (Rackwitz, 2001). Hence, this is the most likely combination of parameters values triggering a failure. When this design point is found, the probability of failure is calculated, using $P_f=1-\phi(\beta)$. The design point is found using a process of iterations in evaluating:

$$Z = \beta - \sum_{m=1}^{n} \alpha_m u_m \tag{3-2}$$

Where β is the reliability index and α_m is the influence coefficient, or the FORM sensitivity coefficient (for which $\Sigma\alpha_m^2=1$) of variable X_m; u_m is the standard normally distributed variable representing a normalized stochastic variable. The value of the influence coefficient of a variable provides a measure of its contribution to the reliability. The highest value of the influence coefficients (max α_m) identifies the dominant basic variables. The process of iterations in the used FORM analysis using the Probabilistic Toolkit is described in (Brinkman, 2015).

Additionally, for slope instability analyses, the observed shape of the slip surface of slope failure (h') can be combined with the likelihood function (A) and, thereby as a next step, can be used for updating the scenario probabilities. This combination of likelihoods is further referred to as the likelihood (B). Evidence is expressed by an observation function h' and limit state function (Z), as in Equation (3-3). Negative values of the h' conditioned limit state function (Z) describe which random variable (X) combination results in both potential failure of a scenario (S_i) and observed slip surface shape (h'). The following likelihood function combines the influence of failure and slip surface (h') (into likelihood function (B)):

$$P(F \cap h' | S_i) = P(Z(X) < 0 | S_i) \tag{3-3}$$

In which Z(X) is the limit state function that includes the observed slip surface in the slope stability computation.

Posterior distribution of scenarios probabilities with observed failure

Subsequently, the earlier assigned prior probabilities of scenario S_i are updated using both the information that the slope has failed (A) and the geometry of the slip surface that is observed (B). This posterior distribution is shown in Equation (3-4):

$$P(S_i | F \cap h') = \frac{P(F \cap h' | S_i) P(S_i)}{P(F \cap h')} \tag{3-4}$$

3.2.2 Bayesian hindcasting of slope instability

Steps of Bayesian hindcasting of slope instability

A Bayesian method for the hindcasting of slope instability is proposed in six steps, based on the method by Kool et al. (2019) and described in Figure 3.1.

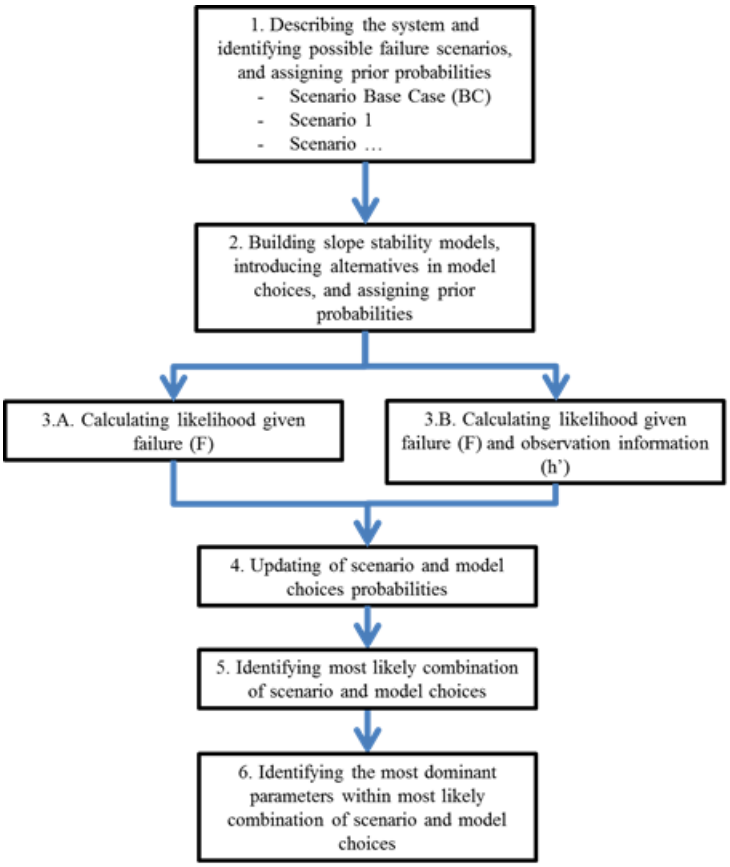


Figure 3.1. Method of hindcasting of levee failures using Bayesian techniques.

The presented Bayesian theory and the model for slope instability analysis (introduced in the following sections) are used to identify the combination of the most likely scenario triggering the failure and the most representative model choice (MLC), step by step:

1. System description and identification of possible failure scenarios (e.g. stratigraphy and water pressures) with the help of the collected evidence. Prior probabilities are assigned to all scenarios (S_i), and a uniform prior can be chosen in case no prior knowledge is available to discern the scenarios:

$$P(S_i) \tag{3-5}$$

2. Introduction of alternatives for model choices (M_j) and parameter choices (X) into the scenarios (S_i) and building slope stability models (noted as ($S_i \cap M_j$)). Prior probabilities are assigned to all model choices:

$$P(M_j|S_i) \tag{3-6}$$

3. System description Evaluation of the limit state function using the generated slope stability models ($S_i \cap M_j$), to calculate the likelihood, one by one:
   a. The conditional probability of event F (likelihood (A)):

$$P(F|S_i \cap M_j) \tag{3-7}$$

b. The conditional probability of F (likelihood (B)) including the field observation information when available (observed slip surface information (h')):

$$P\left(F \cap h' | S_i \cap M_j\right) \tag{3-8}$$

4. Update of prior scenario and model choices probabilities into posterior probabilities using likelihood (A) given failure (F) and (B) slip surface information (h'):

$$\mathrm{P}(\mathrm{S_i}|\mathrm{F} \cap \mathrm{h}') \tag{3-9}$$
$$\mathrm{P}\left(\mathrm{M_j}|\mathrm{F} \cap \mathrm{h}'\right) \tag{3-10}$$

5. Identification of the combination of most likely scenario and model choices: the scenario and model choices resulting in the highest posterior probabilities, given failure (F) and slip surface information (h'), are considered the most likely to characterize the failure most accurately (MLC):

$$\mathrm{maxP}(\mathrm{S_i}|\mathrm{F} \cap \mathrm{h}') \tag{3-11}$$
$$\mathrm{maxP}(\mathrm{M_j}|\mathrm{F} \cap \mathrm{h}') \tag{3-12}$$

6. Identification of the most dominant parameter contributing to failure: the most dominant variables are found by assessing the influence factors of each basic variable for the combination of most likely scenario and most representative model choices (MLC):

$$\max {\alpha_{\mathrm{m}}}^2 \tag{3-13}$$

Please, note that the prior probabilities refer to the probability that a model (M) best describes reality and that a scenario (S) was present given the information. In the next paragraphs step 1 to step 6 will be explained in more detail.

3.2.3 Elaboration of the hindcasting steps

System description and inputs (Step 1. and Step 2.)

A general model of slope stability analysis of a levee is used to analyse the failure of the levee. The analysis calculates the Factor of Safety (FoS) which is a function of the resisting moment (M_R) and the driving moment (M_S). Higher pore water pressures increase M_S and decrease the available shear strength of the soil and lead to a reduction of M_R. The limit state of the slope stability model is used with the purpose to determine the conditional failure of the levee and to incorporate the influence of the observation information (Step 3.A. and 3.B).

The performance of a levee depends on various typical variables (Deen & Duinen, 2016) (Table 3.1). Collected evidence is thoroughly studied on the typical variables influencing the performance of the levee to build a slope stability model. Further, the collected evidence is used to identify scenarios which might support conditions for triggering the levee to fail. Generally accepted literature is used to collect possible alternatives to typical variables supplementing the required set of variables in building the slope stability model.

Table 3.1. Overview of uncertainties and related typical variables of influence on slope instability performance. Variables are introduced as either a discrete or continuous variables. *chosen to exemplify the method on the Breitenhagen case, discussed in the next section (see Figure 3.6).

Uncertainties slope instability (Baecher & Christian, 2003)	Variables (Deen & Duinen, 2016)	Variable type	Specific Variables/choices related to Breitenhagen levee
Site characterization	Geometry of the levee	Discrete	*
	Stratigraphy of the soil layers and the levee body		*
	Water level (landside and waterside)		*
Model	Limit Equilibrium Method (LEM)		Bishop
			Spencer
			Uplift-Van
	Soil Shear Strength Model		Drained
			Undrained
Parameters	Soil parameters	Continuous	Cohesion
			Friction angle
			Shear strength ratio
			Strength increase exponent
			Pre-Overburden Pressure
	Hydraulic pressure of relevant layers		Simplified model (steady-state flow)

Each discrete variable (Table 3.1) contributes a finite set of possible choices of model and parameter choices. All possible combinations of choices are used to generate slope stability models covering the possible characteristics of the failure.

A uniform prior can be chosen in case no prior knowledge is available to discern the scenarios of failure and alternatives of typical discrete variables. For example; the evidence is used to construct four scenarios possibly triggering the failure. Each of them is assigned a probability of 0.25 (Equation (3-5)).

Likelihood given observed failure (Step 3.)

The generated slope stability models (step 1 and step 2) are used in step 3A to determine the conditional probability of failure, which is taken as likelihood given failure (step 3A). The probability of slope instability is calculated using the limit state function Z(X), see Equation (3-14). This is repeated for each generated slope stability model and its corresponding random FoS(X); with each model based on a combination failure scenario, model and parameter choices:

$$Z(X) = FoS(X) - 1 \tag{3-14}$$

Equation (3-15) is used to calculate the conditional probability of failure (step 3A.).

$$P\left(F \middle| S_i \cap M_j\right) = P\left(Z(X) < 0 \middle| S_i \cap M_j\right) \tag{3-15}$$

The collected evidence is subsequently used in step 3B. to estimate the geometry of the observed slip surface (Figure 3.2) or a zone in which the slip surface might have occurred. The slope stability models that produce a slip surface that correspond to the observed information (shaded area in Figure 3.3 and denoted with h') are taken into account of the analysis and expressed by Equation (3-16) (step 3B.).

$$P\left(F \cap h' \middle| S_i \cap M_j\right) = P(Z(X) < 0 \cap h' | \, S_i \, \cap \, M_j) \tag{3-16}$$

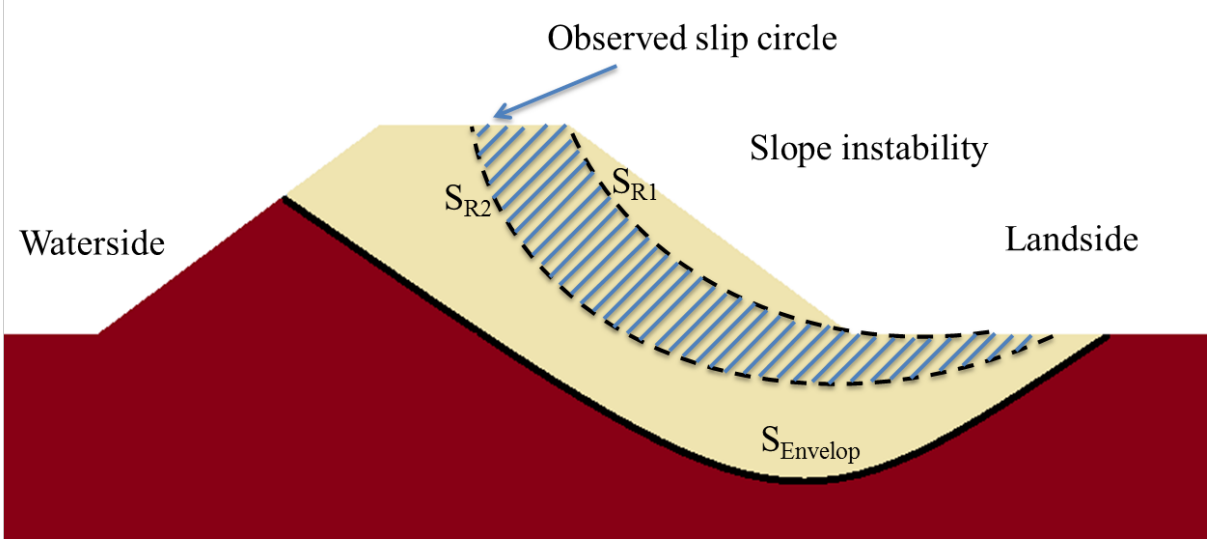


Figure 3.2. Yellow ($S_{Envelop}$): Yellow area marks the envelope where the critical slip surface can be found regardless whether this corresponds to the observed information. Shaded area marks the possible zone of the slip surface (h')

In practice, the observations of a slip surface are not exact. Therefore h' represents an estimated area with an upper boundary (S_{R1}) and lower boundary (S_{R2}), see Figure 3.2. The yellow area shows the envelope of all possible slip surfaces that is the result of step 3A. The shaded area shows the boundaries that are introduced to incorporate the observations, as done in step 3B.

Posterior probabilities identifying the most likely scenario and model choices (Step 4. and Step 5.)

Posterior probabilities are used to identify the MLC for hindcasting purposes. Equation (3-4) is calculated by applying a tree diagram in which all combinations of scenario and model choices are shown. For each combination of scenario and model choice, the likelihood is calculated in a slope stability computation by computing the probability that failure would have occurred. Combining these likelihoods with the prior probabilities of the scenarios finally gives the posterior distribution. For explanatory purposes, a simplified example is shown, in this section.

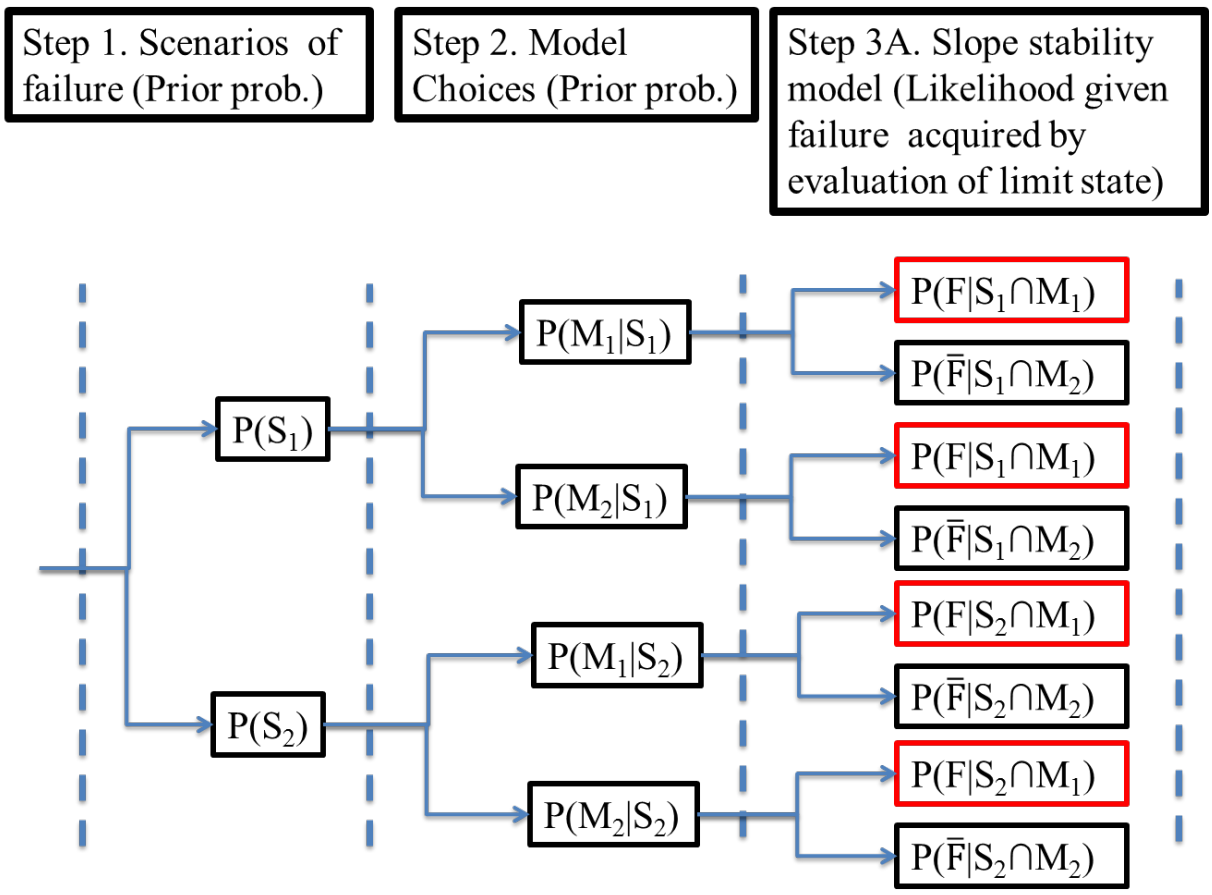
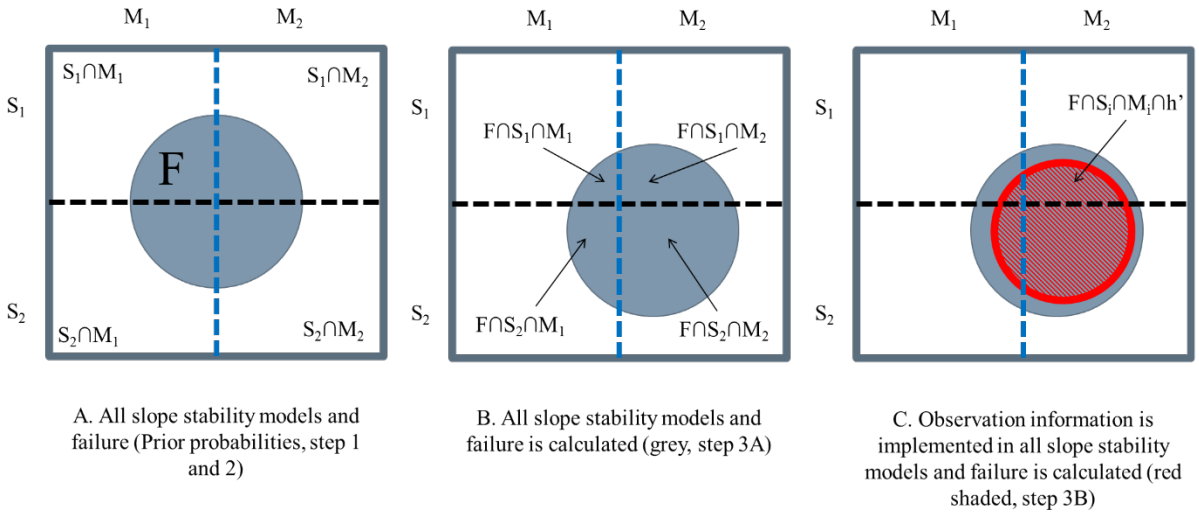


Figure 3.3. Event tree with the incorporation of the scenarios of failure ("S_1", "S_2"), alternatives in model choices ("M_1", "M_2"). The slope stability models are put to purpose to determine the related likelihood given failure ("$F|S_i \cap M_j$") and given non-failure ("$\bar{F}|S_i \cap M_j$").

The simplified example takes only two discrete choices (n=2) into account per discrete variable, that is two scenarios of failures ("S_1","S_2",) and two model choices ("M_1","M_2"). An overview of the resulting four possible slope stability models (noted as ($S_i \cap M_j$)) is shown both by an event tree (Figure 3.3) and an Edwards-Venn diagram (Figure 3.4). The slope stability models are used to determine $P(F|S_i \cap M_j)$. Prior probabilities are assumed $P(S_1)=P(S_2)=0.5$ and $P(M_1)=P(M_2)=0.5$; since $P(S_i)$ is independent of $P(M_j)$ result in $P(S_1 \cap M_1)=P(S_1) \cdot P(M_1)=0.25$.



A. All slope stability models and failure (Prior probabilities, step 1 and 2)

B. All slope stability models and failure is calculated (grey, step 3A)

C. Observation information is implemented in all slope stability models and failure is calculated (red shaded, step 3B)

Figure 3.4. (A.) Edwards-Venn-based diagram to illustrate the relation between the scenarios of failure (S_1, S_2), the model choices (M_1, M_2) and all the four resulting slope stability models that are built ($S_i \cap M_j$)

to calculate the probability of failure in the next step (noted as F and marked as the grey area). Prior probabilities of scenarios and model choices are included. (B.) Probability of failure per slope stability model is calculated and implemented in the diagram ($P(F \cap S_i \cap M_j)$ marked as grey area). This illustrates the intersection of the calculated F and the slope stability models. (C.) The red-shaded area is a subset of F and incorporates the observed slip surface information (h'). The most likely scenario (S_2) and model choices (M_2) are identified.

The four slope stability models ($S_i \cap M_j$) are assumed mutually exclusive, and the total probability of failure is the union in event F (Figure 3.4). This total probability indicates whether the failure was likely (could be expected) or not (a surprise); the total probability of failure is calculated as:

$$\begin{aligned} P(F) = {} & P(F|M_1 \cap S_1)P(M_1|S_1)P(S_1) + \\ & P(F|M_2 \cap S_1)P(M_2|S_1)P(S_1) + \\ & P(F|M_1 \cap S_2)P(M_1|S_2)P(S_2) + P(F|M_2 \cap S_2)P(M_2|S_2)P(S_2) \end{aligned} \tag{3-17}$$

Bayes rule is used to update the prior model and scenario probabilities by the information that failure occurred (step 3A). When slip surface information is available, the set of likelihoods given failure and observation information (step 3B) replaces the set of likelihoods given failure (step 3A, Figure 3.3). The posterior probabilities of the scenarios of failure and model choices are obtained by alternately evaluating Equation (3-16) and Equation (3-18):

$$P(S_2|\,F) = \frac{\sum_{j=1}^{n} P\left(F|M_j \cap S_2\right)P\left(M_j|S_2\right)P(S_2)}{P(F)} \tag{3-18}$$

$$P(M_2|\,F) = \frac{\sum_{i=1}^{n} P(F|S_i \cap M_2)P(S_i|M_2)P(M_2)}{P(F)} \tag{3-19}$$

This way, the most likely scenario leading to failure and model choice to represent the levee failure most accurate is obtained by evaluating Equation (3-11) and Equation (3-12).

Dominant influence factor (Step 6.)

The MLC is used to determine the most dominant variable leading to the failure of the levee. The influence factor is a by-product from the evaluation of the limit state function using FORM, as is done in step 3A and step 3B, and provides a measure of the contribution of each variable ($\Sigma_m \alpha^2 = 1$) to failure. The most dominant variable is identified utilizing Equation (3-13).

3.3 Case study of levee failure near Breitenhagen, Germany in 2013

3.3.1 The Breitenhagen case: levee failure and input data

General failure information

In 2013, the levee near Breitenhagen (Germany) failed and caused considerable economic damages to the area, see e.g. Figure 3.5 (Weichel, 2013). Previous analyses of the levee failure show that the levee would be calculatively stable (FoS>1) under conditions of best estimates of levee characteristics (Grubert, 2013b; Kool et al., 2019). This indicates that possibly an anomaly (e.g. old breach or conductive layer) dominated the outcome (Grubert, 2013b; Kool et al., 2019). Details on the interpretation of the collected data and the identification of anomalies, which are introduced in the analyses as possible scenarios leading to failure, are discussed by Kool et al. (2019). The collected data are summarized in Table 3.2. The data that support the findings of this study are available in the International Levee Performance Database (ILPD) – a public database with information on levee failure and performance cases (at https://leveefailures.tudelft.nl/, failure-id 121001: Breitenhagen, 2013).

Figure 3.5. Left: overview photo of the inland while the flooding evolves, Right: overview photo of the breach at 10-jun-2013 (Grubert, 2013b; Weichel, 2013)

Table 3.2. Collected data on the levee near Breitenhagen that failed, The data that support the findings of this study are available in the International Levee Performance Database (ILPD) (at https://leveefailures.tudelft.nl/, failure-id 121001: Breitenhagen, 2013).

Document	Information
Design of upgrade 1846 (Sixdorf, 2016)	Design drawings of the cross-sections. The design was not realized.
Photo report 2003 (Sixdorf, 2016)	Photos of the installation of sheet piles to prevent seepage near the pumping station
Photo report 2004 (Sixdorf, 2016)	Photos of the construction of the road on top of the crest
Video footage, 2013 (Weichel, 2013)	Video footage during the breach by a drone flight
Saaledeich bei Breitenhagen, (Grubert, 2013b)	Photo report (during and after the breach) Hindcasting based on calculations Location overview Levee profile (measurements) Levee profile km 0+590 (incl. borings) Data based on laboratory test of local samples: water content w in %, yield point wL in %, Roll-out limit wP in %, Plasticity number IP in %, Consistency number IC) Moist sample [gr], dry sample [gr], pore water [gr], water content [%] Sieve curves Geotechnical stability calculations in a variety of scenarios Underground hydraulic analysis (stationary and transient pressures) Uplift assessment
Photo report and paper clippings of the repair 2013 (Sixdorf, 2016)	Photo footage of the repairing of the levee Clippings of the plans in the area Photo footage right after the repair

Step 1: System description, identification of possible failure scenarios

General. The collected evidence is examined to identify possible failure scenarios, i.e. loading and subsoil conditions that possibly explain the failure. The Breitenhagen levee has a typical cross-section of a levee on top of an aquitard on both land- and riverside, consisting of layers of cohesive and alluvial soil, and a high permeable aquifer consisting of sand and gravel (Figure 3.6, based on (Grubert, 2013b)). At the location of the breach, the levee was about 3.50 m high and had a crest width of 3.00 m. A possible failure scenario is caused by the trees growing at the section between the levee and the pond near the toe of the levee at the waterside. Furthermore, this particular section of the levee has likely been breached in the past, resulting in a pond in front of the levee (Sixdorf, 2016).

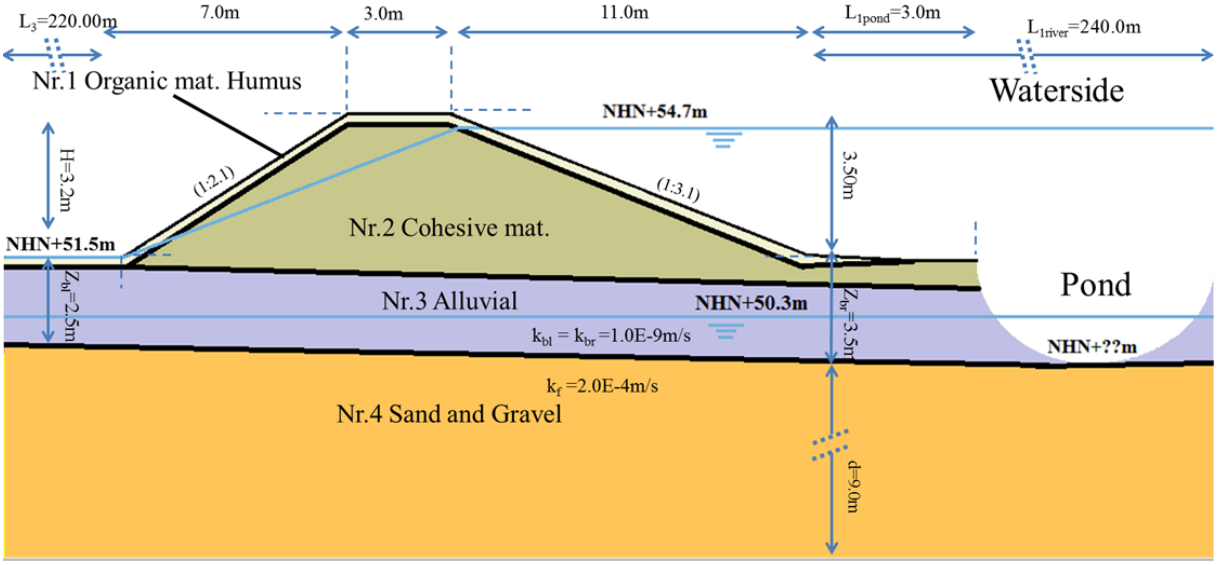


Figure 3.6. Simplified soil layering of the failed levee near Breitenhagen, including dimensions (Drews, 2017a; Grubert, 2013b). Coordinates are expressed in a vertical datum: Normalhöhennull (NHN), which is roughly sea level. The observed high water level NHN+54.7m, measured groundwater of NHN+50.3m. Aquitard extends up to L_3=220m on the land side and up to L_{1river}=240m on the riverside. Furthermore, k_{bl}, k_{br} and k_f indicate the hydraulic conductivity of respectively the landside blanket, waterside blanket and aquifer; while Z_{bl} and Z_{br} indicate the landside and waterside blanket thickness.

Scenarios of failure. Seepage of water from the riverside causes local high water pressures inside the levee to reduce the shear strength and possibly trigger slope instability. The best-estimated situation is referred to as the "Base case" (BC). Subsequently, three possible scenarios of water pressures are identified as shown in Figure 3.7. To emphasize the scenario-specific dominant water pressures, the influence of each scenario-specific water pressure is introduced in relative to the BC. Kool et al. (2019) further elaborate and motivate the different scenarios:

- Scenario "Base case" (BC): the best estimate of the base case is based on normal levee investigation data (borings, etc.) and best estimates of water pressures,
- Scenario "Saturated levee" (S1): higher phreatic as a result of a more permeable levee than expected, based on evidence from photos,
- Scenario "Conductive layer" (S2): higher water pressures inside the levee due to a conductive layer inside the levee as a result of tree roots, based on evidence from photos,
- Scenario "Pond connection with aquifer" (S3): higher water pressures in the aquifer under the levee due to a close connection of the outside water level with the aquifer, based on evidence from photos.

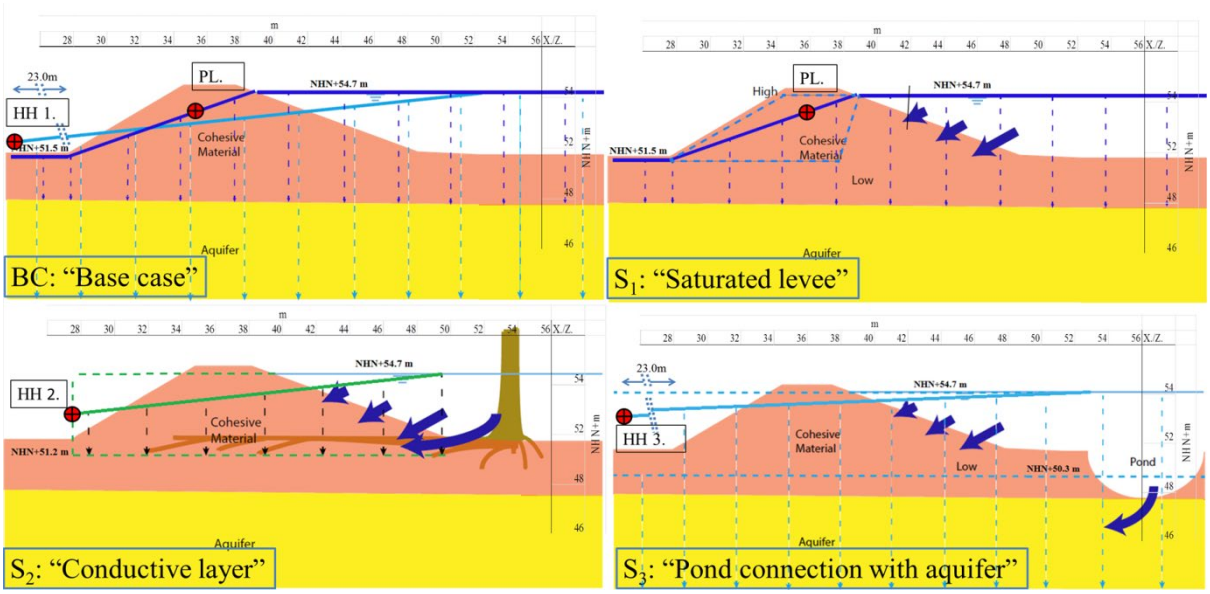


Figure 3.7. Modelling of the four scenarios of water pressure. Horizontal coordinates are project-specific, and vertical coordinates are relative to NHN. The dashed lines are (high and low) boundaries of the piezometric lines. The red node is the point of manipulation in the X-axis and Z-axis to incorporate the uncertainty in the phreatic line (Table 3).

Water pressures in failure scenarios. This analysis adopts a pragmatic implementation of water pressures using simplified models of quasi-steady-state flow to incorporate the transient flow (Figure 3.7; (TAW, 2004)). Three aspects are considered: A. the phreatic line (PL), B. the aquifer head (HH1) and C. the interpolation between A. and B. (this is done linearly and not further discussed). Whether steady-state conditions are reached depends on the duration of the water level and the hydraulic conductivity of the soil. In this case, the higher water event lasted over a month, of which 12 days the water was relatively close to the maximum (see (Kool et al., 2019)). Analysis by Drews (2017b) shows that the aquifer conditions mostly reach steady conditions for the various scenarios, which is incorporated in the modelling in this study. For the phreatic line, this is highly dependent on the hydraulic conductivity of the levee material (Drews, 2017b). For the reported low conductivity, steady-state conditions are not met. However, since photos show wet soils during, the slope failure; indicating a high phreatic line. Hence, a phreatic line that represents (close to) steady-state conditions is used in the base case.

It should be noted that using steady-state conditions is usually pragmatic and conservative for design considerations; for hindcasting, this may lead to the identification of the wrong dominant scenario in case there are significant transient effects. A sensitivity analysis shows that using transient water pressures based on a low conductivity dike body does not influence the outcomes much in this case since the aquifer pressures are much more important, see 'Discussion'.

Modelling water pressures. The water pressures are modelled in each scenario using probability distributions to reflect uncertainty in actually occurred pressures. This is shown in Figure 3.7 and summarized in Table 3.3. The red points in Figure 3.7 are assigned the probability distribution, linear interpolation is used to connect this location with waterside and landside boundaries. The following choices are made:

- Base Case (BC): the mean of the PL (dark blue) and aquifer (light blue) are shown in Figure 3.7 and are based on quasi-steady-state conditions. For the PL, there is a small uncertainty modelled (0.3 m; see (Rozing 2015)). Also, a truncated distribution is used, similar to the dashed line of S1, to reflect the physical boundaries as the water level cannot exit the levee body. The PL can be higher than steady-state conditions though due to e.g. rainfall. For the aquifer, blanket theory (USACE, 2000) is used to compute the water pressures (HH1). In this case, a semi-pervious top stratum to calculate the head is shown in Figure 3.6. The uncertainty is 0.3 m in both x-, and z-direction. Again a truncated distribution is used with

upper bound equal to the maximum river level and lower bound equal to the pre-flood ground water table.

- Scenario “Saturated levee” (S1): The only difference with BC is the increased standard deviation of the PL (of 1.1 m) to reflect the higher potential for full saturation.
- Scenario “Conductive layer” (S2): The only difference with BC is the introduction of an extra piezometric head (HH2) at NHN+51.2 m with increased head. There is a linear interpolation between HH2 and both the PL and HH1 which have the same properties as the BC (but are not shown in the figure for clarity)
- Scenario “Pond connection with aquifer” (S3): The only difference with BC is the increased mean of the aquifer head (HH3) due to the pond connection, resulting in a lower L1 in the blanket equations, see Figure 3.6.

Table 3.3. Uncertainties in water pressures. The hydraulic head distribution is modelled in the red nodes (Figure 3.7) and modelled as a truncated normal distribution. The coordinates correspond with Z and X in Figure 3.7. *S1, S2, and S3 have the same distribution for PL and HH as the base case, except for what is shown in the “water pressure” column.

Scenario of failure	Line	Axis	μ m	σ m	Truncating Bounds m – m
BC: “Base Case”	Phreatic line in levee (PL)	Z	54	0.3	51-54.7
BC: ”Base Case”	Phreatic line in levee (PL)	X	36.5	0.3	35-38
BC: ”Base Case”	Hydraulic head in aquifer (HH1)	Z	52.8	0.3	49-54.7
S1 “Saturated levee”*	Phreatic line in levee (PL)	Z	54	1.1	51 -54.7
S1 “Saturated levee”*	Phreatic line in levee (PL)	X	36.5	1.1	35 – 54.7
S2: “Conductive layer”	Extra hydraulic head rooted conductive layer (HH2)	Z	52.8	1.11	51.5-54.7
S3: “Pond connection with aquifer”*	Hydraulic head in aquifer (HH3)	Z	54	1.11	49-54.7

Step 2: Alternatives for model choices and parameter choices

The possible model choice and parameter choices are derived that, together with the failure scenarios of step 1, will characterize the levee failure most accurately. The different scenarios are implemented with alternatives of Limit Equilibrium Methods (LEM), soil reaction behaviour and properties of soil parameters, to generate a set of slope stability models that include all possible characterizations of the failure (Table 3.1).

The LEM of Bishop (M_{Bishop}), Uplift Van (M_{Uplift}) and Spencer ($M_{Spencer}$) are applied as possible analytical slope stability methods (Bishop, 1955; CIRIA, 2013; Spencer, 1967). LEMs are chosen over more complicated methods such as Finite Elements because of their short running time and sufficient accuracy for the not-complex slip surface as observed in this failure. The different LEMs mainly differ in the various shapes of slip surfaces that they take into account. Moreover, the LEM of Bishop is considered less accurate than the LEM of Uplift Van and Spencer but is incorporated as a control method. The uncertain typical variables (step 1) are the subject of this study as presented in Table 3.1. Therefore, the overall imperfection of the simulation is accepted and compensatory measures, such as an overall model uncertainty factor, are left out of the assessment.

How soils react to loading depends on the rate of loading and the conductivity of the soil. Often, the shear strength of the soil is modelled as drained or undrained, while in reality the soil likely reacts as partially drained. Initially, drained (SB_{dr}) and undrained (SB_{und}) (Ladd, 1991; Schofield &

Wroth, 1968) are used in this section as these are the most commonly used. Partially drained behaviours should be considered in case this is expected to have a large influence on the results; which is not expected for this case. Undrained soil behaviour takes the possibly generated water pressures by deformations into account, in contrast to drained soil behaviour (Deen & Duinen, 2016).

The undrained soil response for low permeability materials is implemented using the SHANSEP implementation (Ladd, 1991) of the Critical State Soil Mechanics (Schofield & Wroth, 1968):

$$s_u = \sigma'_{v,i} S\ OCR^m\ with\ OCR = \frac{\sigma'_{v,y}}{\sigma'_{v,i}} and\ \sigma'_{v,y} = \sigma'_{v,i} POP \quad (3\text{-}20)$$

Where s_u is the undrained shear strength ratio in kPa, $\sigma'_{v,i}$ is the in-situ vertical effective stress in kPa, S the undrained shear strength ratio, OCR the over-consolidation ratio, m the strength increase exponent, σ'_{vy} the vertical yield stress in kPa and POP the pre-overburden pressure in kPa.

Table 3.4. The probability distribution of soil properties (parameters of drained soil behaviour), soil weight (γ), wet soil weight (γ_{wet}), soil friction angle (φ), cohesion (c'). *(Grubert, 2013b) used local soil investigations to identify the soil types, **(Normcommissie, 2011) provides typical mean values of strength parameters which are common in the Netherlands with similar soil description, ***Baker and Calle (2006) provide default typical values of variation coefficients with similar soil description from point to point. The probability distributions are log-normal with a mean (μ) and a standard deviation (σ) (Schweckendiek, van der Krogt, Rijneveld, et al., 2017)

Soil type *	Description of a soil layer **	μ_γ kN/m^3 **	σ_γ kN/m^3 v=0.05 ***	$\mu_{\gamma,wet}$ kN/m^3 **	$\sigma_{\gamma,wet}$ kN/m^3 v=0.05 ***	μ_φ deg **	σ_φ deg v=0.2 ***	$\mu_{c'}$ kPa **	$\sigma_{c'}$ kPa v=0.5 ***
Nr. 1. Organic mat Humus	Clay, clean and moderately stiff	17	0.85	17	0.85	21.25	4.25	7.5	3.75
Nr.2. Cohesive mat.	Clay, little bit of sand, stiff	18	0.90	18	0.90	25	5	7.5	3.75
Nr. 3. Alluvial	Clay, little bit of sand, stiff	18	0.90	18	0.90	25	5	7.5	3.75
Nr.4. Sand and gravel	Gravel, bit silty, clean	18,5	0.93	20.5	1.03	36.25	7.25	0	0

Table 3.5. Probability distributions of soil properties (undrained soil behaviour), shear strength ratio (S), strength increase exponent (m), Pre-overburden Pressure (POP). *(Grubert, 2013b) used local soil investigations to identify the soil types, **(RWS, 2016a) provides typical values of strength parameters which are common in the Netherlands with similar soil descriptions. The probability distributions are log-normal with a mean (μ) and a standard deviation (σ) (Schweckendiek, van der Krogt, Rijneveld, et al., 2017)

Soil type *	Description of soil layer **	μ_S **	σ_S **	μ_m **	σ_m **	μ_{POP} kPa **	σ_{POP} kPa **
Nr. 1. Organic mat Humus	Sandy and Silty Clay	0.30	0.03	0.9	0.03	22	6.6
Nr.2. Cohesive mat.	Levee material	0.31	0.06	0.9	0.03	30	9
Nr. 3. Alluvial	Levee material	0.31	0.06	0.9	0.03	30	9

The report of Grubert (2013) describes the local soil types and characteristics but does not report the probability distributions of the strength parameters that represent the diversity of shear strength in space (Calle 2008; TAW, 2001). Based on the descriptions of the soil characteristics by Grubert (2013) and (Normcommissie, 2011), the corresponding values of soil strength are collected

for both drained and undrained soil behaviour (Table 3.4 and Table 3.5). The probability distributions of the parameter input are adopted from (Baker & Calle 2006).

Conform steps 1 and 2, failure scenarios are identified: a finite set of interchangeable alternatives of model choices and parameter choices are systematically collected, which results in 24 combinations of failure scenarios and model choices ($S_i \cap M_j \cap SB_n$). Each combination is used to build a slope stability model (see Figure 3.8), which is evaluated for the probability of failure given the a-priori conditions. When observed slip surface information is available (see step 3 below). The observed slip surface information is included in all 24 slope stability models by conditioning the search area of the critical slip surface to the observed slip surface. Both the soil parameters and hydraulic pressures of relevant soil layers are implemented as continuous probability distributions (Table 3.3, Table 3.4, and Table 3.5). An overview of all combinations is presented using an event tree in Figure 3.8. and in 0, which presents an elaborate overview of all combinations and calculation results, and a calculation example that demonstrates how the posterior probability of S_3 is calculated.

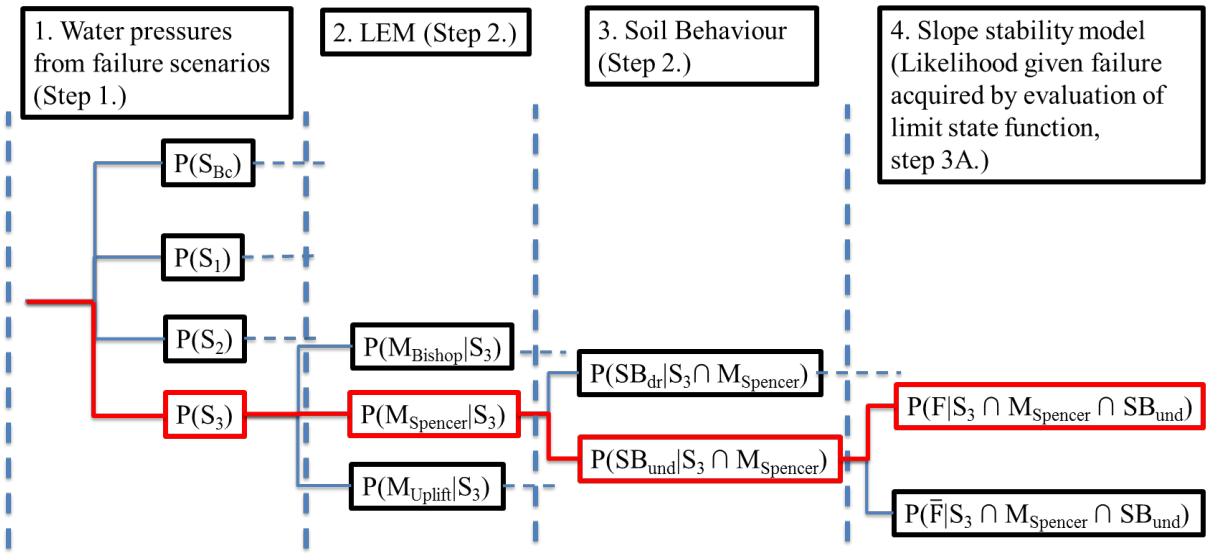


Figure 3.8. Presents an overview of the different combinations of scenarios and model choices. Level 1: Water pressure from failure scenarios (i.e. S_{Bc}: "Base case", S_1: "Saturated levee", S_2: "Conductive layer", and S_3: "Pond connection with aquifer"). Level 2: LEM (i.e. M_{Bishop}, $M_{Spencer}$ and M_{Uplift}). Level 3: Soil reaction behaviour (i.e. SB_{dr}: drained, and SB_{und}: undrained). Level 4: Likelihood given failure (F) and non-failure ($\bar{F}\bar{F}$). Red branch contains MLC, as an example.

Step 3: Likelihood given failure and field observational information

The assembled 24 slope stability models are used to evaluate the accompanied limit state functions with FORM. This results in the conditional probability of failure of each slope stability model, which is adopted as likelihood in this analysis. The likelihood is determined for two situations:

A. Field observation information is not available and the search area of the critical slip surface is not defined in the slope stability models, resulting in $P(F|S_i \cap M_j \cap SB_m)$ (step 3A.),
B. Field observation information is available and the search area of critical slip surface is conditioned to the observed slip surface information in the slope stability models, resulting in $P(F \cap h'|S_i \cap M_j \cap SB_m)$ (step 3B.).

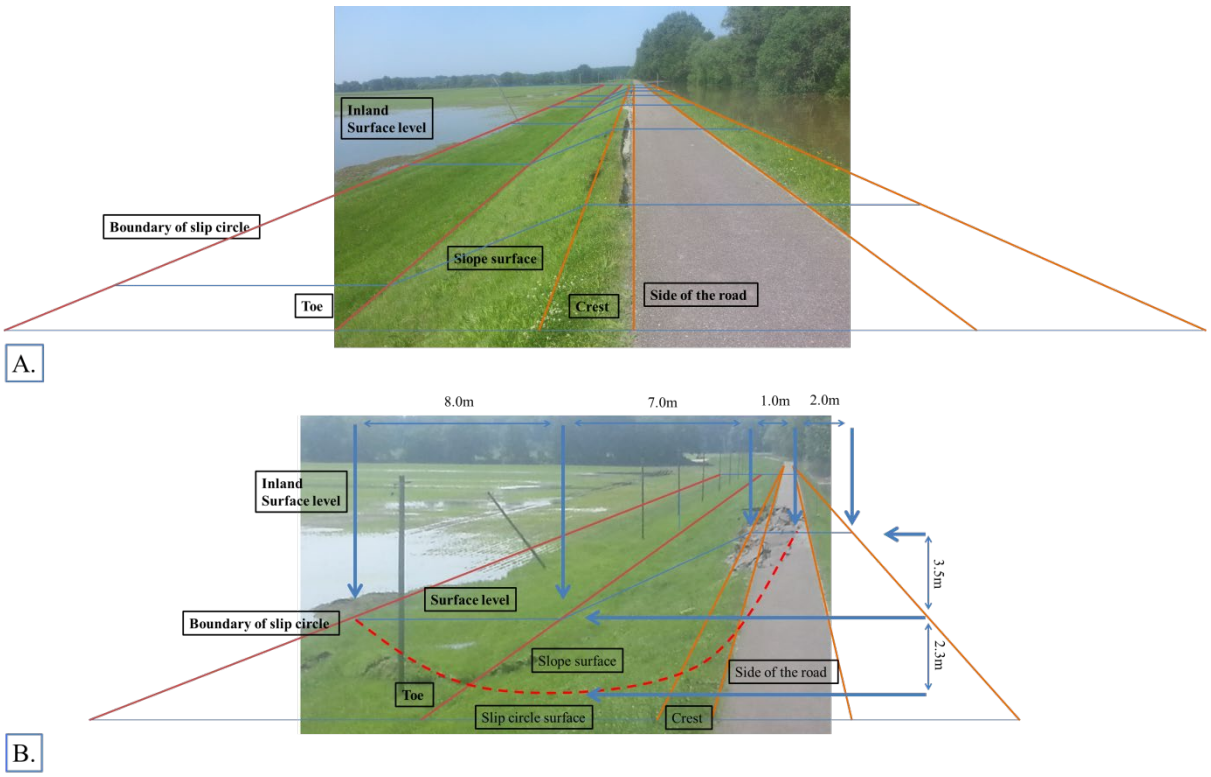


Figure 3.9. Geometric analysis of the failed levee. Longitudinal guides to the levee (in Orange, Red), guide across to the levee (Blue). (A.) First distortions are apparent, (B.) Larger distortions are apparent and used to estimate the geometry of the slip surface (Red dotted)

Figure 3.9 shows the field observation information that is available. The slope stability models are conditioned (h') according to the observed shape of the initial slip surface as part of step 3B. (Equation (3-8)). The exact geometry of the slip surface is difficult to estimate, and therefore estimated by an upper and lower boundary that are 2m apart, as suggested in Figure 3.2.

Steps 4, 5, and 6: Systematic analysis of uncertainties

Both sets of the 24 likelihoods (steps 3A. and 3B.) and the prior probabilities (steps 1 and 2) are implemented alternately in Bayes rule and used to determine the posterior probabilities for each scenario of failure and model choice (step 4). The resulting posterior probabilities substantiate the identification of the MLC (step 5.) which is used to identify the parameters that contributed to the failure the most.

Evaluating the limit state with FORM enables to estimate the conditional probability of failure, but also determines the contribution of the individual parameters to failure ($\Sigma\alpha_i^2=1$). Because FORM might introduce approximation errors on the probability of failure, the results have been verified with Monte Carlo which gave very similar results for the considered cases.

3.3.2 Results

Updated scenario probabilities

The two sets of the 24 likelihoods (steps 3A. and 3B.) are used to update the prior probabilities of earlier identified scenarios that possibly trigger a failure (step 1.). The prior and the resulting posterior probabilities are shown in Figure 3.10, step by step.

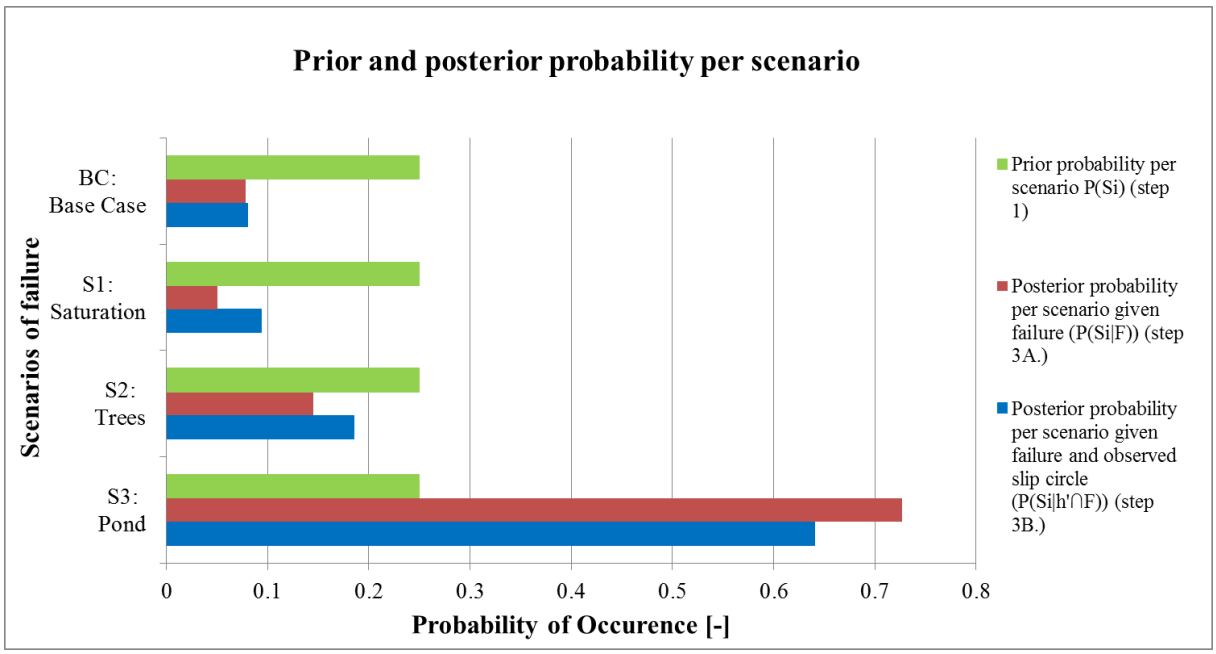


Figure 3.10. Prior probability per scenario BC, S_1, S_2, and S_3 (equally distributed probabilities (green)), the posterior probabilities per scenario given failure (red) and observed geometry of slip surface (blue). These are conditional probabilities that sum over all scenarios (i.e. BC, S1, S2 and S3) to one.

Updating the prior probabilities of the scenario BC, S_1, S_2, and S_3, based on the information of the actual failure (step 3A.) emphasizes S_3 as the most likely scenario ($P(S_3)=0.25$ to $P(S_3|F)=0.72$). When the observed slip surface information is included in the update (step 3B), the posterior probability of S_3 reduces slightly (from $P(S_3|F)=0.72$ to $P(S_3|h'\cap F)=0.64$). The results identify S_3 as the most likely scenario of failure.

Systematic analysis of uncertainties and most dominant parameters

Additionally, the set of the 24 likelihoods (steps 3A. and 3B.) is used to update the prior probabilities of alternatives in the model choices (step 2), in the process of identifying the MLC. Both the prior and the resulting posterior probabilities are shown in Figure 3.11 and Figure 3.12. Then, the MLC is deployed to determine the most dominant contributor to the failure (step 6).

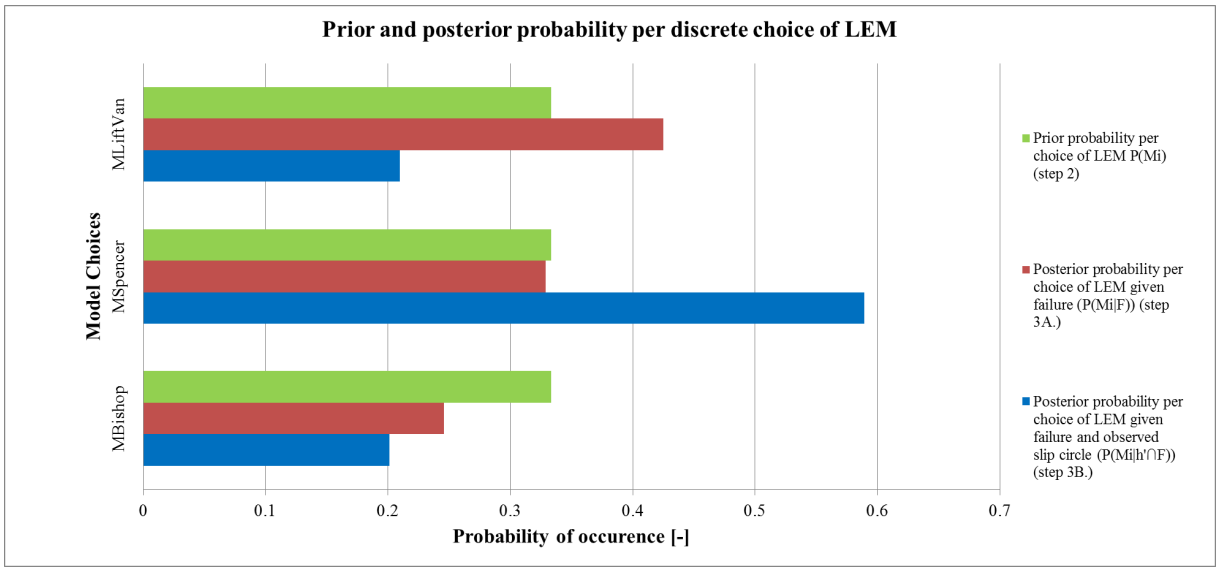


Figure 3.11. Prior probability per choice of LEM Uplift Van, Spencer, and Bishop and equally distributed probabilities (green), the posterior probabilities per choice of LEM given failure (red) and observed geometry of slip surface (blue). These are conditional probabilities that sum over all alternative choices of LEM to one.

Figure 3.11 shows that updating of the prior probabilities on the information that the actual failure happened (step 3A.), emphasizes the LEM of Uplift Van as the most likely choice of LEM ($P(M_{Uplift})=0.33$ to $P(M_{Uplift}|F)=0.43$). However, when the observed slip surface information is included in the update (step 3B.), the resulting posterior probability emphasizes the LEM of Spencer as a more likely choice than the LEM of Uplift Van, i.e. $P(M_{Spencer}|F)=0.33$ to $P(M_{Spencer}|h'\cap F)=0.59$. The posterior probabilities identify the LEM of Spencer to characterize the failure most likely most accurately. Figure 3.12 shows the most likely resulting slip surface (in the design point) when S3: Pond and undrained soil behaviour is assessed using LEM of Uplift Van, Spencer and Bishop without observational information (step 3A., upper row) and with observational information (step 3B., lower row) included.

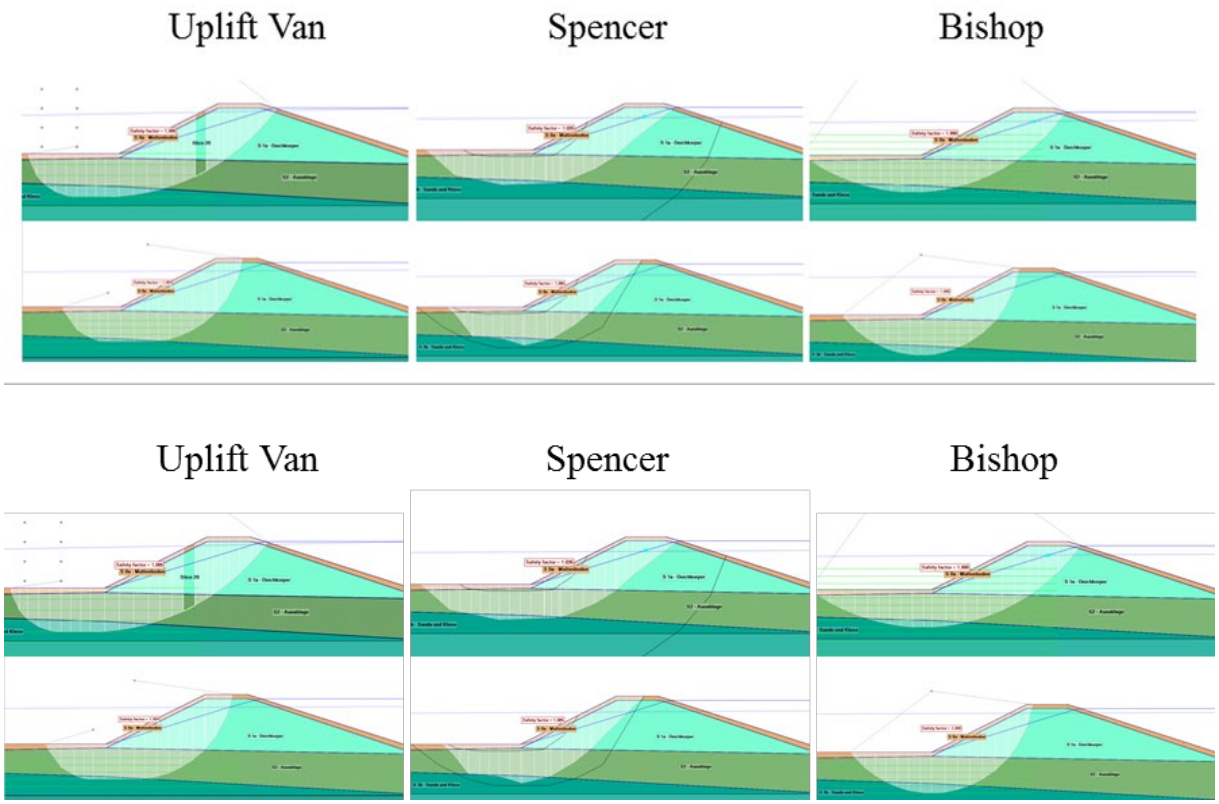


Figure 3.12 Slip surface in the design point of the most likely scenario (S3: Pond) and soil behaviour (undrained soil behaviour) LEM Uplift Van (left), Spencer (middle), and Bishop (right). Upper figures: search area of the critical slip surface is not defined in the slope stability models (step 3A). Lower figures: search area of the critical slip surface is conditioned to the observed slip surface information in the slope stability models (Step 3B.).

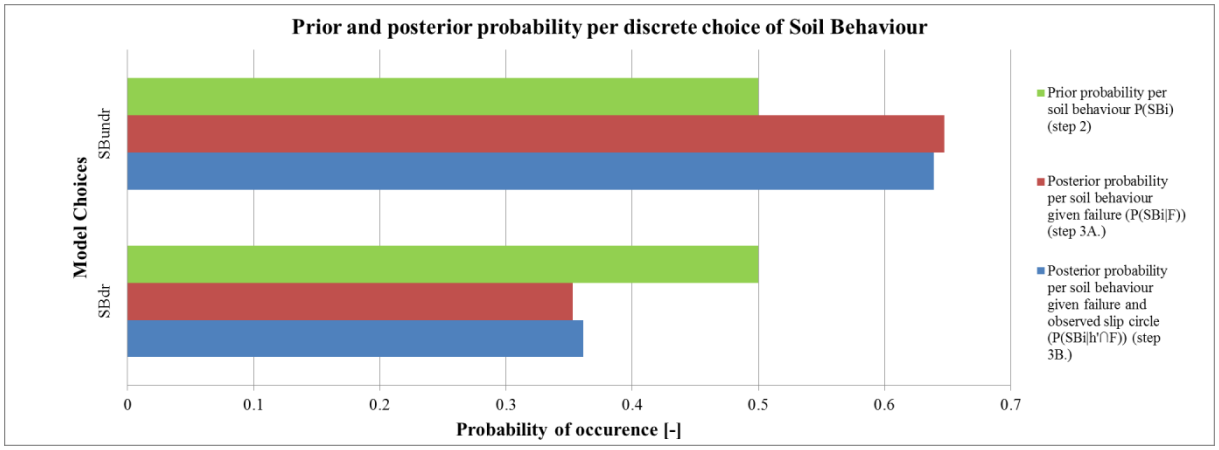


Figure 3.13. Prior probability per choice of soil behaviour both drained (SB_{dr}) and undrained (SB_{undr}) and equally distributed probabilities (green), the posterior probabilities per the choice of soil behaviour given failure (red) and observed geometry of slip surface (blue). These are conditional probabilities that sum over both shear strength models to one.

Figure 3.13 shows that updating prior probabilities of soil response behaviour on the information that the levee failed (step 3A.) emphasizes the posterior probabilities of the undrained soil response behaviour as the most likely soil behaviour, i.e. $P(SB_{undr})=0.50$ to $P(SB_{undr}|F)=0.65$. Including the observed slip surface information in the update (step 3B.) has little effect on the posterior probabilities. The posterior probabilities identify the undrained soil response behaviour to characterize the failure most likely most accurately; however, the differences are small and no firm conclusion can be drawn.

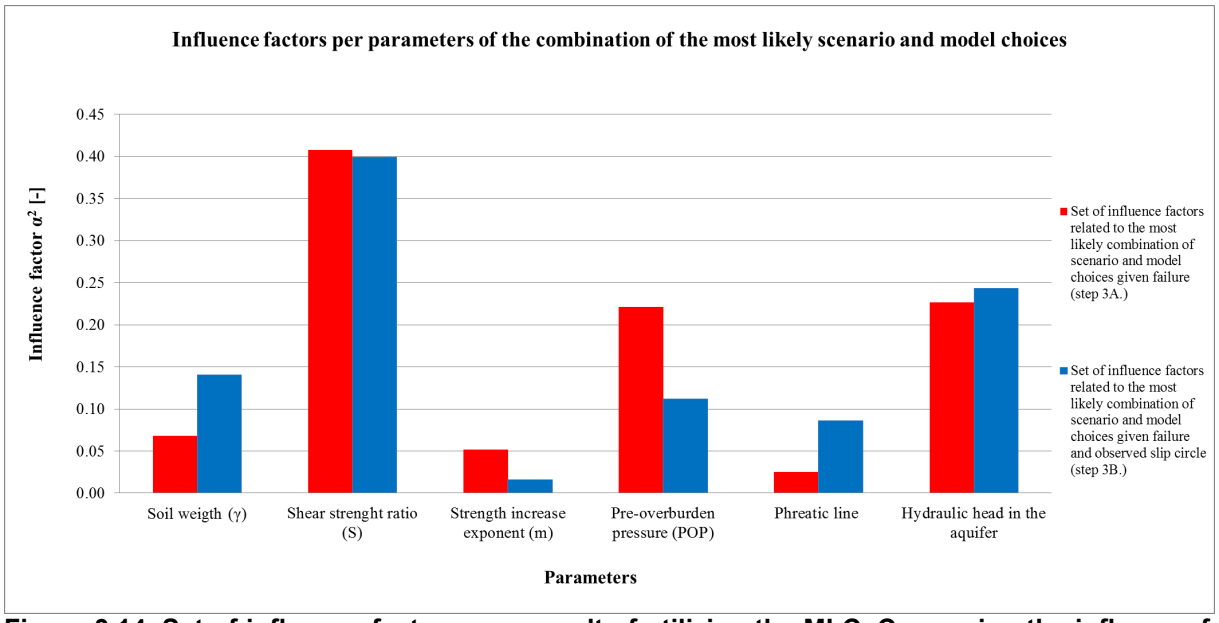


Figure 3.14. Set of influence factors as a result of utilizing the MLC. Comparing the influence factors of the individual parameters without (red) or with (blue) incorporation of observational information ($\Sigma\alpha^2=1$)

The slope stability model implemented with the combination of scenario pond connection with the aquifer (S_3), method of Spencer ($M_{Spencer}$) and undrained response soil behaviour (S_{udr}) characterizes the failure most likely most accurately. This slope stability model is used to evaluate the limit state and to estimate the contribution of the individual variables to the failure. Figure 3.14 shows that overall the shear strength ratio and the hydraulic head in the aquifer are considered the most dominant parameters contributing to the failure. The influence of the Pre-overburden pressure seems to decrease as a result of the incorporation of the observed slip surface information.

3.4 Discussion

3.4.1 Framework for probabilistic hindcasting of levee failures

The six steps of hindcasting are based on a framework which explicitly and transparently accounts for significant uncertainties and model choices by quantitative means that reflect on the likelihood of occurrence. Although, consideration of all possible combinations of scenarios of failure and model choices is time-consuming; it does provide a thorough analysis of the failure possibilities.

Even more, the suggested Bayes technique based approach enables the implementation of observational information in the hindcasting. The approach provides insights on how each piece of evidence influences the scenario and model choice-related posterior probabilities, building up to the identification of the slope stability model best representing the failure, piece by piece. The proposed slope stability model facilitates the identification of the most dominant parameter contributing to failure. The information that the event of failure happened is decisive for the identification of the most likely failure scenario. The information on the shape of the slip surface is decisive for the identification of the most representative model choices and the most dominant parameter contributing to failure.

The First Order Reliability Method (FORM) is used in this section to estimate the probability of failure for every generated slope stability model. However, FORM gives an approximation of the probability of failure meaning that the exact value of the probability of failure is unknown. FORM has the benefit of providing a quick answer and insight into the influence factors. It also has (potential) drawbacks. For instance, the found local design point might correspond to a non-representative failure probability. In such cases, the calculated influence factors are non-representative as well. Also, the iterative process might not converge to one answer due to numerical problems. However, the software that was used (Probabilistic Toolkit of (Brinkman, 2015)) provides the option of detecting numerical problems and this was not an issue for this assessment. Also, several results of FORM calculations have been verified with parallel Monte Carlo simulation, providing very similar results.

Scenarios of failure and alternatives of model choices are assumed equally probable as a first estimate, despite that some of the scenarios and the model choices are considered less likely, such as the introduction of Bishop's method. Experts consider the Bishops' method to be less accurate than both Spencer's and Uplift Van's method. Moreover, slope stability computations are the bases of the likelihoods and result in posterior probabilities to determine the most likely causes. Field observations can be implemented into the analyses by including these in the slope stability computations, such as the geometry of the slip surface or the actual slope failure. The absence of clear evidence impedes the estimation of prior probabilities, but appears less important for results.

Furthermore, by including the likelihood of occurrence in the hindcasting, the outcome is considered to reflect the actual situation more objectively than a hindcasting based on a deterministic sensitivity analysis. With a deterministic sensitivity analysis, the results are highly determined by the chosen input parameter bounds. Also, despite the equally assigned prior probabilities over the discrete choices as a first estimate, which is considered a very rough estimate, experts in the field might be able to suggest more appropriate prior probabilities. Altogether, it is still possible that the actual cause was not identified as part of the considered scenarios.

3.4.2 Application of probabilistic hindcasting to the Breitenhagen case

The six steps of hindcasting are applied to the levee failure near Breitenhagen. Several elements in the analysis are based on local data such as the scenarios of water pressure and part of the soil parameters. Other elements rely on generic inputs such as some shear strength parameters; see Table 3.4 and Table 3.5. The more local data are available, the more accurate the hindcast. In this case, it was not feasible to collect more local data and we believe there was sufficient data available to conclude. But more local data might yield in even more conclusive conclusions.

The connection between the river and the aquifer, as presented in "Pond connection with aquifer" Scenario (S_3), results in high water pressures building up underneath and in the toe area of the levee. High water pressures in these areas reduce the local shear strength. These conditions

influence the toe area (passive area) of the levee the most, resulting in a non-linear and stretched slip surface when slope instability occurs. Observations, illustrated in Figure 3.9, confirm that the passive area of the slip surface stretches up to the electrical pole at the toe of the levee and tilts the pole. Thus, the observed stretched passive part of the slip surface indicates high pore pressures underneath the levee. Other scenarios of water pressures concentrate on the active area of the slip surface, leaving the passive area relatively intact and resulting in a slip surface which is less stretched. Even though the other scenarios are not very probable, they cannot completely be excluded. Also, the method of Spencer is most robust and accurate when calculating critical slip surfaces that are not typically circular-shaped. This explains why the posterior probability favours the method of Spencer when incorporating the observed slip surface information (h') (Figure 3.11).

Whether the soil is best characterized by drained or undrained behaviour is, in reality, less binary than the models suggest. The soil behaviour is probably best characterized as partially drained which explains the relatively small difference between the resulting posterior probabilities (Figure 3.13).

The forensic analysis of Kool (2019) using a deterministic sensitivity analysis-based hindcasting, reports a collection of identified dominant parameters contributing to failure: that deviating low shear strength associated with low values of POP or cohesion justify the failure. This is in contrast to the findings of the current analysis, which identifies the low shear strength associated with low values of the Shear strength ratio. The difference is explained by: 1.) Undrained soil behaviour is identified as the most representative model choice in the analysis and, therefore, automatically excludes the cohesion as the most dominant parameter, 2.) The applied range of the values of POP (by Kool, 2019) covers situations that are less relevant for this specific situation and results in FoS<<1. This low value of FoS explains why the POP is identified as a dominant contribution to failure when a deterministic sensitivity analysis of hindcasting is used. However, when parameter values are related to their likelihood of occurrence, the Shear strength ratio values resulting in failure are more probable to occur, which explains the high influence factor.

In this study, the water pressures are mostly based on steady-state assumptions for the phreatic line and hydraulic head in the aquifer. While a pragmatic and conservative choice for engineering purposes, this assumption can lead to the wrong conclusions for hindcasting. In general, transient effects should be incorporated in a hindcast. In this study, steady-state conditions were likely reached for uplift due to the long duration of the high water and the high conductivity of the aquifer. For the phreatic line, steady-state may be more questionable because of the low conductivity of the levee material (though the pictures do show a mostly saturated levee). A sensitivity analysis with a lower phreatic line (reflecting transient effects) does not yield a different conclusion. This is because the stability is mainly determined by the aquifer head. For cases where the levee body is of more importance for the stability, transient effects also become more important.

LEMs are used in this section to model slope stability, where a Finite Element Model (FEM) would be better in capturing complex geometries and soil behaviour. However, in this case, a sensitivity analysis shows that the scenarios of water pressures are more important for failure than the soil behaviour. Furthermore, the observed slip surface does not follow a complex shape. Therefore, in combination with its computational efficiency, we chose a LEM as this is deemed sufficiently accurate. FEM might, however, be more appropriate in case of more complex geometries and soil behaviour, and in case water pressure scenarios are less dominant. In order to use the benefits of FEM, more local information would be needed as well (hindcasting this is difficult to collect more data as this mostly vanished), especially if spatial variability is to be incorporated.

The findings of the forensic engineering report by Grubert (2013) identified a collection of causes: unexpected saturation of the levee, steep slope of the levee and, foremost, the influence of the tree roots. However, this present study shows that a pond in front of the levee is the most likely cause of failure. The findings of Grubert (2013) are possible as well, as the computed scenario probabilities are not negligible. Both visual and historical data indicate that the pond in front of the levee is likely to be leftover from a former breach (Sixdorf, 2016). Due to the reparations of this breach, the conditions of the soil might deviate from other stretches of the levee. The most likely

scenario of a local high hydraulic head in the aquifer and the most dominant parameter identified by the present study can directly be connected to the former breach and the reparations, which would explain the actual location of the breach in 2013.

3.5 Conclusions and recommendations

3.5.1 Conclusions

This chapter demonstrates the application of probabilistic Bayesian techniques to the hindcasting of levee failures due to slope instability in six steps. The method provides insights into the most relevant uncertainties of hindcasting by evaluating all possible scenarios of failure and model choices on the likelihood of occurrence. This results in the identification of the most likely scenario and most dominant parameters triggering a failure. The suggested steps of analysis provide a thorough, workable and transparent method of analysis. Furthermore, this method is an improvement of existing deterministic hindcasting methods (Kool et al., 2019) in the sense that it provides a better insight into the relative likelihoods of the various possible causes of failure.

The developed approach of hindcasting is applied to the levee failure near Breitenhagen, with the identified most likely scenario of failure, the most representative model choices characterizing the failure, and the most dominant parameters triggering the failure as a result, despite the scarcity of evidence. The levee failure near Breitenhagen in Germany in 2013 is most likely triggered by locally weak soil conditions and unexpected high water pressures due to a connection between a pond on the riverside of the levee and the aquifer underneath the levee. The slope stability model that characterizes this failure most accurately is implemented with LEM of Spencer and undrained shear strength soil behaviour. Within this combination of a failure scenario and model choices, the shear strength ratio is identified as the most dominant contributor to the failure. The contribution of the high hydraulic head due to the pond connection is identified to be the second dominant. Based on the available evidence, an old levee breach explains both the presence of a pond and the locally weak soil (Sixdorf, 2016) and would thus explain the specific location of the breach.

3.5.2 Recommendations

Even though the probabilistic method of hindcasting is successfully applied to the levee failure near Breitenhagen as part of the forensic analysis, the method can be improved by analysing more levee failures. Historical cases such as New Orleans (2005), would make an interesting object of study, and the method can also be further developed for other failure mechanisms, such as piping. Overall, it is expected that the developed approach can support a more systematic analysis of other levee failures.

In this chapter, we have used FORM for probabilistic calculations. In some cases, this method could lead to inaccurate estimates of the design point and probability of failure. In future assessments, it is recommended to investigate the (parallel) use of other methods, such as Monte Carlo, especially in case on non-linear limit state functions and complex failures.

This method shows that including observational information in hindcasting is vital to the identification of the most dominant contributing variable. Therefore, it is recommended to explore whether it is possible to include more evidence in the hindcasting with the help of Bayesian techniques, e.g. past performance information (Schweckendiek, van der Krogt, Teixeira, et al., 2017).

In order to characterize failure, LEMs are used in this chapter as the slope stability is mainly determined by water pressure scenarios and the slip surface relatively simple. Finite Elements Method (FEM) analyses can lead to a more accurate characterization of slope instability, and should especially be considered for complex geometries and soil behaviour (see e.g. (Varkey et al., 2017). However, this comes at the cost of a computational burden.

Acknowledgements

We thank both Dr.-Ing. Thilo Weichel of the Landbetrieb fur Hochwasserschutz und Wasserwirtschaft Sachsen-Anhalt and Dr.-Ing. Torsten Heyer of Dresden University of Technology for providing us with relevant data of the levee failure near Breitenhagen.

This research was performed as part of the NWO TTW project SAFElevee (project number 13861). We acknowledge the valuable inputs of the anonymous reviewers.

4 Deterministic and probabilistic hindcasting of the slope instability of the Leendert de Boerspolder- experiment

4.1 Introduction

To identify the cause of a failure, a forensic engineer would like to study and evaluate the situation of the failure as accurately as possible. In this chapter, we look into the slope failure of a levee. Using a slope stability analysis could help the forensic engineer explain why the driving moment (M_S) of the levee exceeded the resisting moment (M_R). However, typically it is complicated to generate a slope stability model that characterizes and models the failure accurately for analysis purposes, due to the scarcity of relevant and documented information. The lack of this information results in uncertainties about the actual present circumstance explaining the failure. Examples of these uncertain parameters are (CIRIA, 2013; Deen & Duinen, 2016):

- Stratigraphy of the soil
- Layer properties, e.g. values of the geotechnical parameters and drainage rate of the soil
- Hydraulic loads and pore pressures

Possible loadings and subsoil conditions related to hydraulic loads, pore pressures and stratigraphy that possibly explain the failure are referred to as scenarios. For example, when such a scenario is implemented with alternatives of Limit Equilibrium Methods (LEM), soil reaction behaviour and soil parameters a slope stability model can be built to characterize the failure. Such a slope stability model can be used to evaluate which scenario explains the slope instability and which parameters are dominant. However, it is difficult to find the most likely scenario using models because of e.g.:

- Multiple scenarios of failure that explain the instability are possible and cannot be excluded from the analysis a-priori (see e.g. Seed et al. (2006) for the failure of the London Avenue canal in New Orleans).
- It is not possible to substantiate the different choices in the models with collected evidence.

To deal with these difficulties, two methods of hindcasting a levee failure systematically were developed in the previous chapters of this thesis, whereby uncertainties can be incorporated into the analysis: a deterministic method (Kool et al., 2019) and a probabilistic method using Bayesian techniques (Kool et al., 2020). The latter method is further referred to as the Bayesian method. However, the validation of these two methods so far is limited and mostly based on the Breitenhagen slope failure during the 2013 river floods in Germany.

The objective of this chapter is to validate and show how these newly developed methods of hindcasting can be used in other situations, based on another failure. The Leendert de Boerspolder experiment is used where a levee was brought to failure under controlled circumstances, of which very detailed information is available. This detailed information is used to accurately determine the most likely scenario explaining the failure and the most accurate slope stability model. Roughly explained, this levee of the Leendert de Boerspolder was brought to failure by drawing down the water level in an excavated ditch at the polder side of the levee in steps of 0.5 m. In doing so the resisting moment gets lowered with every step of the water level drawdown until the driving moment finally exceeds the resisting moment. In this chapter, the failure scenarios cover all steps of water level drawdown which all possibly have triggered the failure. The failure scenarios are combined with

alternatives of model choices, i.e. soil response (whether the soil responds with undrained shear strength or drained shear strength due to changing loadings) and method to determine the slip surfaces (LEM of Bishop, Spencer and Uplift Van). All other parameters are assumed to be known. To validate the outcome of the analysis, the recorded water level drawdown observed during failure (approximately 1.6 m) is used. But the analysis is performed as if the water level drawdown at the moment of failure was not recorded, as this is generally not known during actual unexpected failures. Hence, the Leendert de Boerspolder failure allows an accurate comparison between the observations and the most likely scenario predicted by the Bayesian and deterministic methods.

The two methods of hindcasting were earlier applied to the Breitenhagen slope failure. In the case of the Breitenhagen slope failure, multiple scenarios of local pore pressures inside the levee were possible while the water level in the polder was known. In the case of the Leendert de Boerspolder-experiment, multiple scenarios of water level drawdown in the ditch behind the levee are possible and the pore pressures inside the levee are known as they were observed and the levee was kept saturated fully. Even more, we focus, as much as possible, on the use of results from existing computations as documented in de Gast (2020).

Two levels of information are considered. The first step in applying the two hindcasting methods is to find the most likely water level drawdown using basic information (soil layering etc.). In the second step, more detailed failure information is included in the analysis; the geometry of the observed slip surface. This enables a better identification of the most likely scenario of water level drawdown explaining the failure. At the same time, the applicability of both methods is demonstrated.

This chapter is structured as follows: the case study of the Leendert de Boerspolder experiment is presented in summary in chapter 4.2. In chapter 4.3, both methods are applied to the case study and the results are shown. The results are discussed in chapter 4.4, and the conclusion and recommendations are in chapter 4.5.

4.2 Case study: the Leendert de Boerspolder levee failure experiment

We use the Leendert de Boerspolder experiment as a case study (de Gast, 2020). Extensive site investigation and laboratory testing programs provide very detailed information, which has provided many insights into the failure behaviour of the levee. This levee protects a regional area, is exposed to a constant water level, and is covered by grass. The levee was brought to failure by excavating a ditch (initially filled with water) at the polder side of the levee and drawing down the water level in the ditch. This levee dates from before the year 1610 and has been maintained and upgraded over time. The levee is mostly made of cohesive material, however, it contains also some heterogeneous recycled material used over centuries to strengthen it.

This section presents an overview of the geometry of the levee, the setting of the experiment, and the interpretation of the site investigations and laboratory results as input to calculate material model parameters, only. See de Gast (2020) for further details. Information on maintenance, upgrades, and repairs due to failure or others over time is unknown.

4.2.1 Cross-section and stratigraphy

The experiment site covers only a small section of the Leendert de Boerspolder's levee which is brought to failure. A small triangular area in the south of the polder is designed as an experiment location. The experiment site parts into three sections to ensure enough detailed information, which is called the North, Centre and South row (Figure 4.1). From this section, the Centre section is normative and is included in this analysis. The altitude coordinates are expressed relative to the Normal Amsterdam Level (NAP). A NAP+0 m is approximately equal to the mean sea level of the North Sea. The stratigraphy is reconstructed using the five Cone Penetration Tests (CPT) borings in the crest in two rows (de Gast, 2020). The borings show little variation over the different cross-sections.

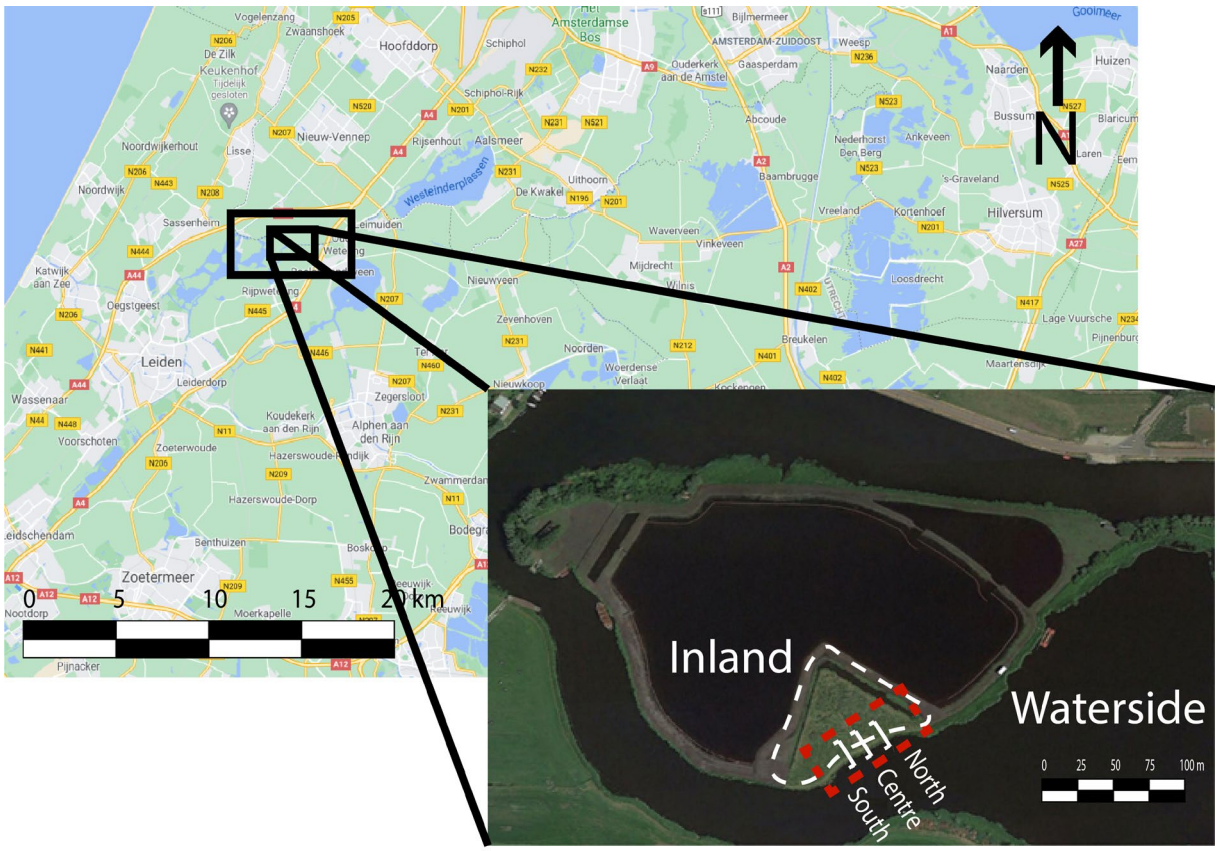


Figure 4.1 Overview of Leendert de Boerspolder, province of Noord-Holland, the Netherlands, taken before the experiment (adapted from de Gast (2020)). The inland is a small triangular area in the south of the polder. The location of the experiment and failure is marked by the red box and includes the three cross-sections of the levee that are evaluated, i.e. North, Centre, and South. For this analysis, the Centre cross-section is identified as the normative cross-section. The inland is protected by a levee, which is marked by the white dotted line, from the outside water.

The dimensions of the cross-section of the levee are mapped at the location of the breach, as shown in Figure 4.2. The measurements show that the levee is about 1.5 m high with a crest width of 2.5 m. The landside slope is 5.5 m in length. The (outside) water level is continuously present near the crest at a level of NAP-0.6 m. During the experiment, a ditch is excavated at the toe of the levee on the polder side of the levee. The water in the ditch at the polder side has a reference level of NAP-2.0 m, which is almost equal to the polder level (NAP-1.9 m).

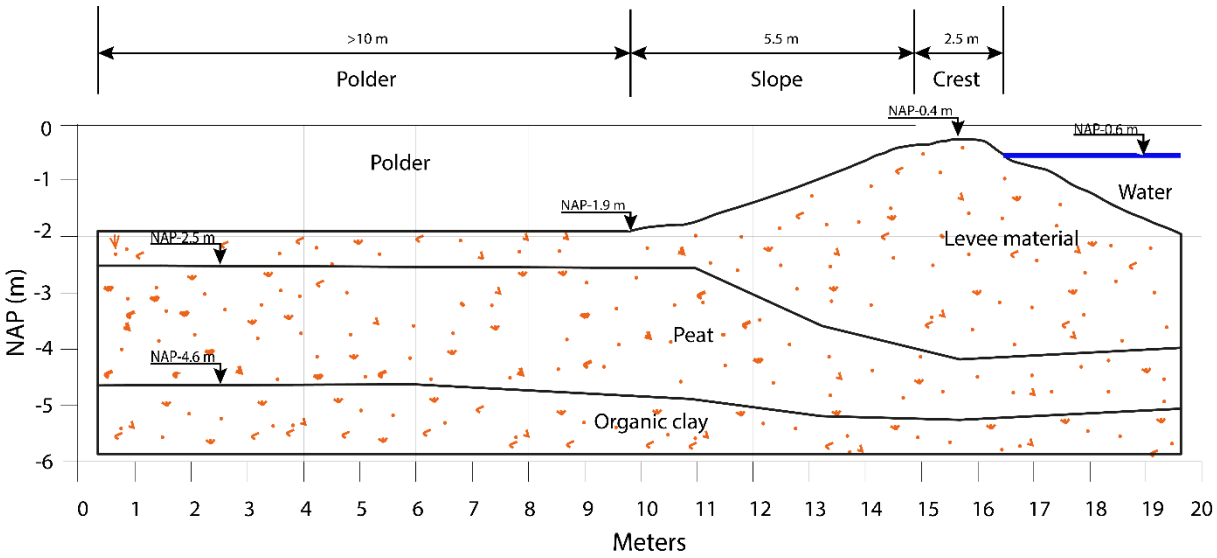


Figure 4.2 Geometry of centre cross-section and the stratigraphy of the top layers (adapted from de Gast (2020)), i.e. levee material, peat, and organic clay viewed from the southwest. The levee material consists of clay, silt, sand and rubble. Dimensions of the levee are given by approximation.

The stratigraphy of the cross-sections generally consists of four layers and is described by de Gast (2020). The top layer of the levee consists of levee material, which is placed over time. It's periodically added for maintenance purposes. The levee material consists of sand, silt, clay, and rubble. The second layer consists of peat and is influenced by the dewatering, in the polder, and by the overlaying levee material. The third layer consists of organic clay, which organic content is more at the top than at the bottom. The fourth layers consist of silty clay with an increasing silt content over the depth (bottom at NAP -16 m).

4.2.2 Failure experiment design

The experiment was designed to induce a slip surface that occurs through the peat layer in seven stages (Figure 4.3, Figure 4.4 and Table 4.1). The design, execution, monitoring and analyses of the experiment were done by de Gast (2020) as part of his dissertation. Failure behaviour has been monitored and the retrieved data from de Gast (2020) is the basis for this analysis.

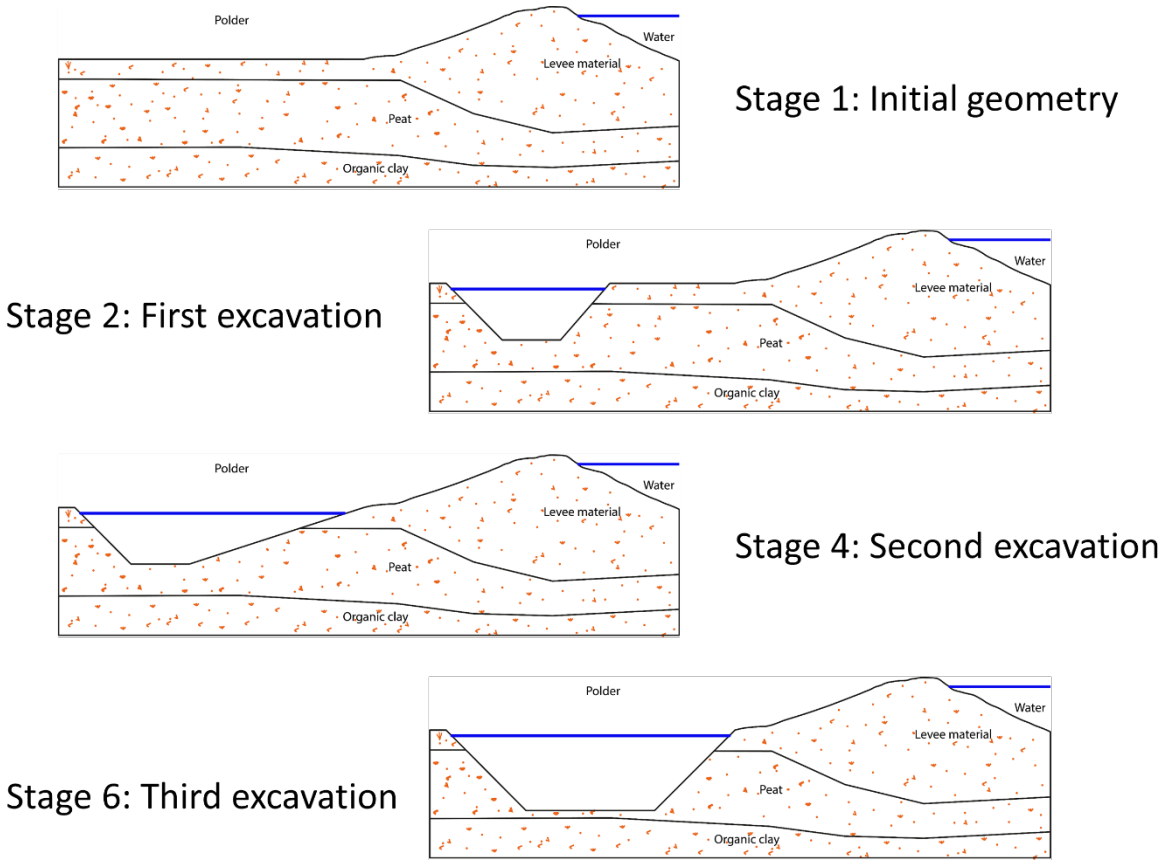


Figure 4.3 Overview of the different stages of excavation during the experiment (adapted from de Gast (2020)), i.e. stage 1, stage 2, stage 4, and stage 6 as introduced in Table 4.1. Dimensions of the levee are given by approximation. In stage 2 the ground at the toe of the levee was excavated forming a ditch with a depth of 2.0 m and a width of 5.5 m. In stage 4 the ditch was widened to a width of 9.0 m. In stage 6, the ditch was deepened to a depth of 2.5 m, as shown in Figure 4.4. Note: the excavation is 2.5 m in depth while the levee is about 1.5 m high.

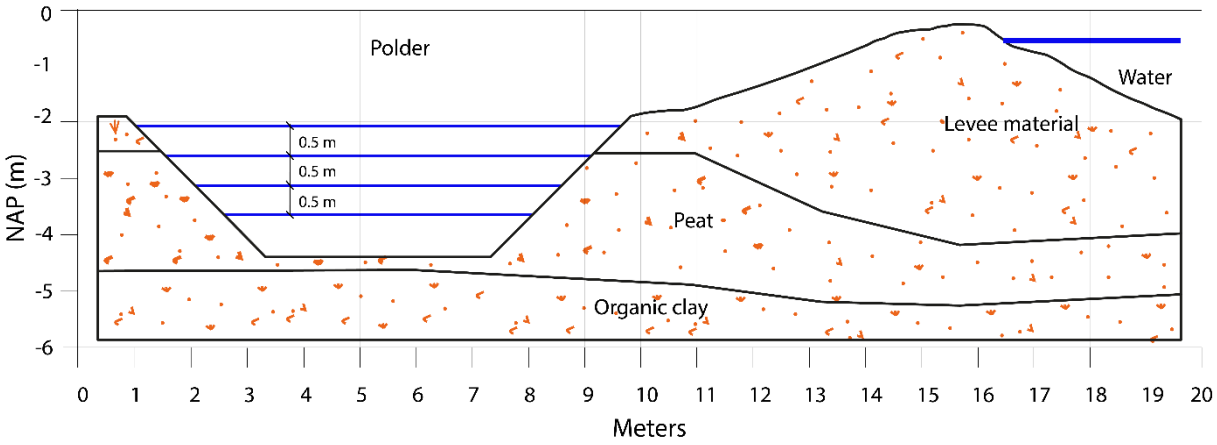


Figure 4.4 Drawing down the water level in steps of 0.5 m during the final stage (stage 7) (adapted from de Gast (2020)). Dimensions of the levee are given by approximation. The levee failed at a water level drawdown of 1.6 m by approximation.

Table 4.1 The seven stages during the experiment towards inducing failure of the levee, the starting date and which actions are affiliated with this stage.

Stage	Starting date	Actions
1	18-09-2015	Data acquisition and wetting of the levee
2	28-09-2015	The first excavation of the ditch
3	30-09-2015	Pumping of the ditch and refilling
4	05-10-2015	Second excavation of the ditch
5	07-10-2015	Pumping of the ditch and refilling
6	12-10-2015	Third excavation of the ditch
7	14-10-2015	Drawing down the water level in the ditch and inducing failure

The levee was brought to failure in seven stages (as illustrated in Figure 4.3, Figure 4.4 and Table 4.1).. By studying the strength behaviour of the levee during these seven stages, it is easier to determine the stress and strain behaviour during the actual failure. In stage 1, artificial rain was applied to the levee saturating (equilibrating pore pressures) the levee for completely over a week. This means uncertainties related to unsaturated soil response are reduced. In stage 2, the first excavation is executed which is filled with water. In stage 3, the water level in the ditch is lowered, from NAP-2.0 m, by 1.0 m and refilled to its original water level. In stage 4, a second excavation is executed. During stage 5, the water level is lowered by 1.0 m and refilled again. In this stage, the levee showed large deformation and the levee was suspected to fail accidentally over a shallow slip surface. Mitigation measures were taken to prohibit this accidental failure. During stage 6 the excavation is widened. In the final stage (stage 7), the water level in the ditch is lowered in steps of 0.5 m. Finally, failure occurred between a water level drawdown of 1.5 m and 1.75 m. The polder water level at failure is estimated at a 1.6 m water level drawdown.

4.2.3 Geotechnical parameters

Expected values and standard deviations for relevant soil parameters are derived from laboratory tests (de Gast, 2020) as shown in Table 4.2. Depending on the drainage rate of the soil, the soil shear strength is characterized by either a Mohr-Coulomb drained shear strength model or an undrained shear strength model (Robertson, 2009). Results from the laboratory tests are summarised in terms of average values for a Mohr-Coulomb drained shear strength model (c', φ') in Table 4.2. The undrained shear strength model, s_u, (Equation (4-1)) was used to characterise the undrained shear strength of the saturated soil and values were computed from the CPT test and its resulting cone resistance over depth, utilizing N_{kt} correlation (Robertson, 2009):

$$s_u = \frac{q_t - \sigma_v}{N_{kt}} \tag{4-1}$$

In which q_t represents the cone resistance (kPa), σ_v represents the total vertical stress (kPa), and N_{kt} represents an empirical correction factor. Moreover, the median of the measured shear strength samples approximates the generic shear strength of a levee best in case soil responds with undrained shear strength due to changing loadings. In his thesis, de Gast (2020) provides more elaborate information on the derivation of the undrained shear strength model parameters.

Table 4.2 Soil parameters derived from triaxial laboratory data at 5 % axial strain and N_{kt} correlation (Robertson, 2009), γ_{dry}: mean dry unit weight, γ_{sat}: mean saturated unit weight, c': mean effective cohesion, ϕ': mean effective friction angle, s_u using N_{kt} correlation: is expressed in two values which characterize the median shear strength both at the top of a layer the bottom of the layer is linearly interpolated over the layer.

Description Soil layer	γ_{dry}/ γ_{sat} kN/m^3	φ' degr	c' kN/m^2	s_u median, top-bottom layer kN/m^2
Dyke material	13.0/18.0	33.0	5.0	6.9-22.9
Peat	9.0/10.0	28.8	2.5	9.7-11.7
Organic clay	14.5/15.0	29.5	4.4	6.2-9.2
Silty clay	14.0/17.0	30.0	1.9	18.0

4.2.4 Observations before, during and after the induced failure

Both methods of analysis use factors of influence that provide information on the failure behaviour of the structure coming from data that were taken before, during and after the induced failure. These factors of influence can substantiate the conclusion greatly (Kool et al., 2019). In this section, the data collected before, during and after the failure is analysed.

Observations before the failure

The experiment showed that parts of the peat layer were more permeable than anticipated. During the experiment, a significant amount of water was drained from the peat layer (of the levee body) as a result of the water level drawdown during stage 3 (Figure 4.5). This was such a large amount of water that the drainage showed in the water pore pressure measurements. After refilling the ditch, the pore pressure inside the levee was restored to the original water level. During the whole experiment, the levee was under saturated conditions and the outside water level has maintained at the same water level.

Figure 4.5 Significant amount of water draining from the peat layer as a result of the water level drawdown at the polder side of the levee running for a couple of hours.

During stage 5 (Figure 4.3 and Table 4.1.), the levee experienced large deformations suggesting the first signs of failure, after drawing down the water level in the ditch. This was not according to the experiment design. In reaction, the water level was raised again, preventing the failure of the levee.

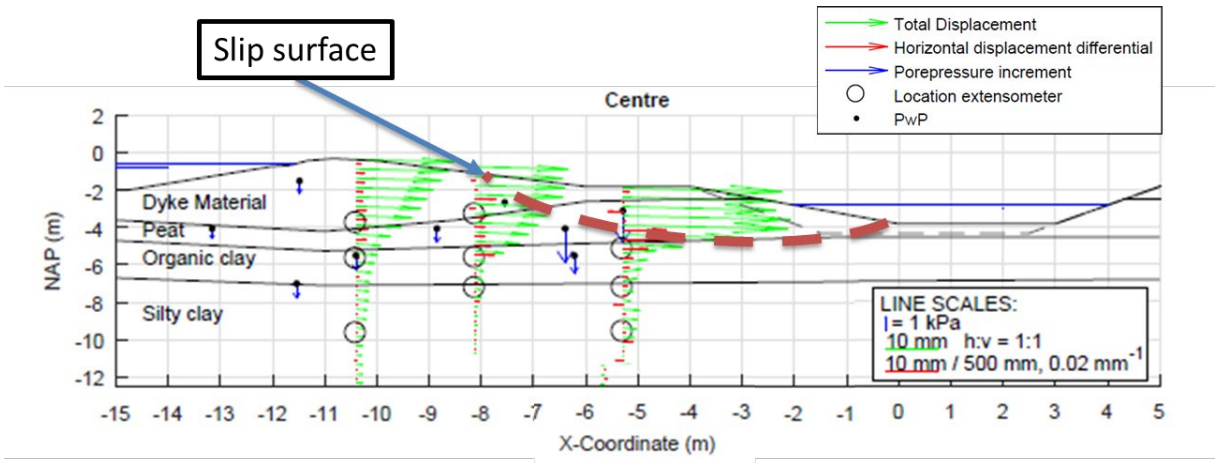


Figure 4.6 shows the centre cross-sections of the levee (adapted from de Gast (2020)). On the left side, is the waterside with a water level of NAP-0.6 m and the polder on the right. Displacement vectors during stage 5 after drawing down the water level at the polder side during the time interval of October 7 and October 8 (de Gast, 2020). Vertical displacements are measured using an extensometer and the location of the extensometers is noted with a circle in the figure. The registered negative pressure is considered proof that local reorganisation of grain occurs. We estimate the geometry of the resulting slip surface supported by the deformation measurements and water pressure measurements (de Gast, 2020).

Both pore pressure sensors and displacement sensors were installed inside the levee to monitor the levee's behaviour during steps of excavation and drawing down the water level (Figure 4.6). These have provided information on the total horizontal deformation (green vector) and the relative and measured horizontal deformation between two points successive inclinometer points as a function of time (red vector) during stage 5 (Table 4.1). The pore pressure sensors measure the maximum pore pressure differences during each stage (in bleu).

The deformation sensors do not provide enough information to estimate the geometry of the slip surface (Figure 4.6). However, measurements registered by the water pressure sensors complement the needed information to estimate the geometry of the slip surface. The development of sudden pressure loss is registered, during stage 5's pumping and refilling. We believe that this is the result of the drawing down of the water level that initiates large deformations and a smaller amount of water drainage as shown in Figure 4.5. Local deformations within a soil body, that exists of interlocking grains, will initiate a lever motion between neighbouring grains, which increases the void space between the grains and creates negative pressure. Hence, this creates a dilatant soil response. The registered negative pressure is, therefore, proof that local deformations do occur and that soil is loaded locally to its maximum capacity. Moreover, an increase in relative deformation is measured between inclinometer points as it would have happened if a clear slip surface had been developed. The measured relative deformation between inclinometer points is, therefore, proof that local deformations do occur and that soil is loaded locally to its maximum capacity. In both measurements, a number of points indicate where the slip surface has developed and these points are enough to estimate how the geometry of the slip surface has developed by the eye. Introduced mitigation measures avoided that the soil exceeded its loading capacity and a full slip surface could not develop. All information is used to reconstruct the expected geometry of the beginning slip surface (dashed brown line as shown in Figure 4.6).

Observations during the failure

In the final stage (stage 7), the levee was brought to complete failure, which was registered using measurement sensors (Figure 4.7) and video footage. On the morning of October 14_{th}, 4.30 am local time, the deformation measurement equipment registered deformations of the levee and two hours later, at 6.30 am, the levee finally failed. The failure slip surface has left a breach of 27 m wide by approximation (STOWA, 2015). Information on the geometry of the slip surface, that occurred, has provided valuable information related to the probability of failure occurrence, i.e. updating purposes. In this section, we have attempted to reconstruct the geometry of the slip surface.

Both pore pressure sensors and displacement sensors were installed inside the levee to monitor the levee's behaviour during the steps of excavation and drawing down the water level (Figure 4.7). These have provided information on the total horizontal deformation (green vector) and the relative and measured horizontal deformation between two points successive inclinometer points as a function of time (red vector) during stage 7. The pore pressure sensors have measured the maximum pore pressure differences over time (in bleu).

Again, solely the deformation measurements have not provided enough information to reconstruct the geometry of the slip surface sufficient, as explained in the previous section. Therefore, information taken from the water pressure sensors has complemented the information that is needed to reconstruct the geometry of the final slip surface (dashed brown line as shown in Figure 4.7). The local deformation triggers the dilatant response of the soil as the voids between the grains have been increased and have resulted in negative pore pressures locally. Moreover, an increase in relative deformation is measured between inclinometer points as it would have happened if a clear slip surface had been developed. The relative deformation between inclinometer points is, therefore, proof that local deformations do occur and that soil is loaded locally to its maximum capacity. Again, this information is used to estimate the geometry of the slip surface by the eye.

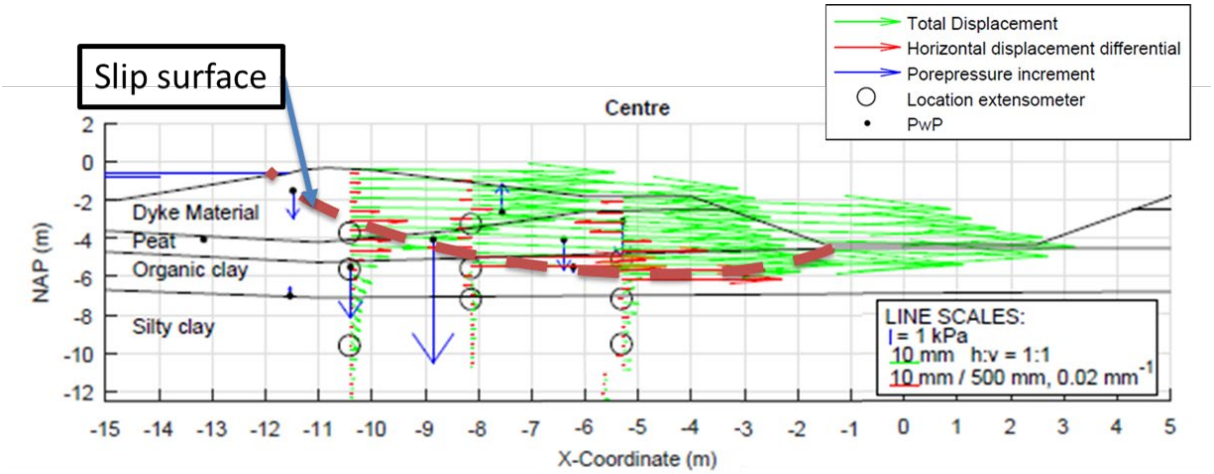


Figure 4.7 shows the centre cross-sections of the levee (adapted from de Gast (2020)). On the left side, is the waterside with a water level of NAP-0.6 m and the polder on the right. As part of stage 7, displacement vectors of the centre row, after drawing down the water level at the polder side, during a time interval from September 19 to October 14 (de Gast, 2020). Vertical displacements are measured using an extensometer and the location of the extensometers is noted with a circle in the figure. The registered negative pressure is considered proof that local reorganisation of grain occurs and we estimate the geometry of the resulting sliding plane supported by the deformation measurements and water pressure measurements (de Gast, 2020).

Observations after the failure

No observation information of the failure gathered from the site after the induced failure is available.

4.3 Hindcasting of slope instability of Leendert de Boerspolder

4.3.1 General approach

The Leendert de Boerspolder experiment is analysed as part of the hindcasting analysis and validation of both the deterministic and the Bayesian hindcasting methods as described in chapter 2 (Kool et al., 2019) and chapter 3 (Kool et al., 2020). In turn, the hindcasting analysis is part of forensic analysis. The process of forensic analysis roughly consists of three stages (Carper, 2000): Collecting and reviewing evidence, using calculative models for a back analysis to identify the cause (hindcasting), and coming to a conclusion and reporting the findings (Carper, 2000). This analysis focuses on the hindcasting part, where a calculative model is built that characterizes the levee failure and the performance of a levee at the time of the failure most accurately and is enabling a forensic

engineer to explore the failure space (e.g. all combinations of possible stratigraphy of the soil, layer properties and hydraulic loads and pore pressures that could result in failure).

The performance of a levee can be determined by analysing the ratio between the driving moment (M_S) and the resisting moment (M_R). This ratio is expressed as the factor of safety (FoS). The FoS is used in this chapter with the purpose to determine the conditions that most likely explain the failure. Values of the FoS that are just under one (FoS<1) describe a condition that the driving moment just exceeds the resisting moment. Thus, this is the moment that failure occurs. A levee stability analysis is performed by building a calculative model to evaluate the performance. The following variables are influencing the levee performance (CIRIA, 2013; Deen & Duinen, 2016):

- The geometry of the levee;
- Stratigraphy of the soil;
- Water levels (polder side and waterside);
- Water pressure (development) inside the levee;
- Values of the geotechnical parameters of the soil;
- Drainage rate of the soil. Hence, whether the soil shear strength is best characterized by a drained or undrained shear strength model;
- Most appropriate Limit Equilibrium Method (LEM).

In the process of building a slope stability model, there are multiple scenarios of water level drawdown possibly explaining the failure and model choices to characterize the failure due to the lack of information about the actual present circumstance explaining the failure. The most accurate characterization is reproducing a failure that resembles the observation information the most. How this most accurate characterization of the levee failure is identified differs between both the methods of hindcasting that are used (i.e. deterministic method and Bayesian method). But when this set is found, the water level drawdown is found which most likely explains the failure.

The collected data are analysed to identify all variables that influence the levee's performance, as shown in Table 4.3. The geometry of the levee includes the geometry of the ditch (as a result of the excavation), the stratigraphy of the soil, the water level at the waterside and polder side, water pressures inside the levee, and possible values of the geotechnical parameters of the soil are identified. Also, the geometry of the slip surface is known which is added in the analysis as a second step to update introducing a second level of information into the hindcasting. The water levels at the polder side at the moment of failure are used for validation purposes, although the water level at the time of the failure is known. Therefore, the water level variable is for this purpose considered as if it is not known. The drainage rate of the soil and the most appropriate LEM are still uncertain and will be identified according to the two methods of hindcasting, as part of the validation.

Table 4.3 Overview of typical variables of influence on slope instability performance and uncertainties that are either identified using the data and factors of influence or are unknown. Variables are taken either as discrete or continuous variables.

Identified	Unknown
• The geometry of the levee (including the excavation); • Stratigraphy of the soil; • Water levels (waterside and polder side); • Water pressure (development) inside the levee; • Values of the geotechnical parameters of the soil. • Added as a second step: the geometry of the observed slip surface	• Hydraulic head during failure • Drainage rate of the soil; • Most appropriate Limit Equilibrium Method.

In section 4.3.2, the deterministic method is explained and followed by the section where the results of the deterministic method are presented. In section 4.3.4, the Bayesian method is presented and followed by the results in section 4.3.5.

4.3.2 Deterministic method

The deterministic method is used to identify the most likely scenario of water level drawdown and model choices to characterize the failure in two steps and consider two levels of information. The levee was made to failure by drawing down the water level in an excavated ditch at the polder side. In the first step, different scenarios of water level drawdown explaining the failure and model choices are systematically analysed on their influence on the FoS. In the second step, the observational information is included in the analyses as well, i.e. the resulting critical slip surfaces from the first step are compared to the observed geometry of the slip surface. The water level explaining the failure at the polder side was measured but is kept as an unknown factor for this analysis. The drainage rate of the soil (drained or undrained response of the soil) and the most appropriate LEM are also assumed uncertain (though known). In this section, the general method is explained and how it is applied to the Leendert de Boerspolder-experiment in two steps. In the next section, the results are presented.

In the first step of the deterministic method, special attention is paid to how to identify the critical water level drawdown, because the failure is explained by successive water level drawdowns. Within the full range of potential successive water level drawdowns that aim to initiate a levee into failure, the driving moment (M_s) increases when the water level drawdown increases. The different scenarios of water level drawdown are analysed as a sequence, as they are also in practice. The first set of water level drawdown and model choices in this sequence that results in a driving moment that exceeds the resisting moment is called the most likely water level drawdown explaining the failure (FoS<1). This is the water level drawdown that is most likely triggering the failure first in the sequence of water-level drawdowns.

In this analysis, the cross-section and the failure surface are considered two-dimensional, while the actual failure surface has three dimensions (van Duinen, 2015). In practice, only the critical cross-section is checked for exceeding the computed driving moment on the compute resisting moment under heavy loading. While in reality, the slope of the levee fails when the total driving moment over the length of the wedge exceeds the total resisting moment over the length of the wedge. De Gast (2020) has proposed to correct the calculated factor of safety in the critical cross-section with a correction factor of 1.13 to correct for three-dimensional effects to mark the moment where the driving moment exceeds the resisting moment ($FoS_{3D}=1.13 \cdot FoS<1$).

Following Figure 4.8 and as the first step of the analysis, the following practical computations are performed in systematically investigating the influence of the different alternatives in water levels of water level drawdown and model choices on the FoS_{3D}. One base slope stability model is built which is based on the information (identified variables of influence) that is known as presented in the previous section. This base slope stability model consists of a water level drawdown of 1.5 m, a LEM of Bishop and a drained shear strength model. Failure of the levee is initiated by drawing down the water level gradually on the polder side. Thus, each water level is considered as a possible scenario for explaining the failure while taking into account that these different water levels are occurring in a logical sequential order. Therefore, the base model is varied with different water level drawdowns: When the scheme of Figure 4.8 is taken as an example, first the water level drawdown level of 1.0 m is implemented in the base model than a water level drawdown of 2.0 m. Second, the levee failed due to slope instability after developing a slip surface that seems to have a relatively simple-shaped slip surface (Figure 4.6 and Figure 4.7). Therefore, both alternative LEMs are implemented in the base model, one by one, i.e. UpliftVan and Spencer (Figure 4.8). As of last, due to the water level drawdown at the polder side, the soil inside the levee has experienced a change in shear stresses, that has to be resisted for the levee to maintain its integrity. How much shear stress the soil can resist depends on the shear strength the soil can generate and on the drainage rate of the soil. However, whether the soil responded drained or undrained is unknown. The drained soil shear strength model is implemented in the base model and the undrained soil shear strength model as an alternative (Figure 4.8) using the parameter values of Table 4.2. By doing so, the influence of these variables on the factor of safety (FoS_{3D}) is explored systematically.

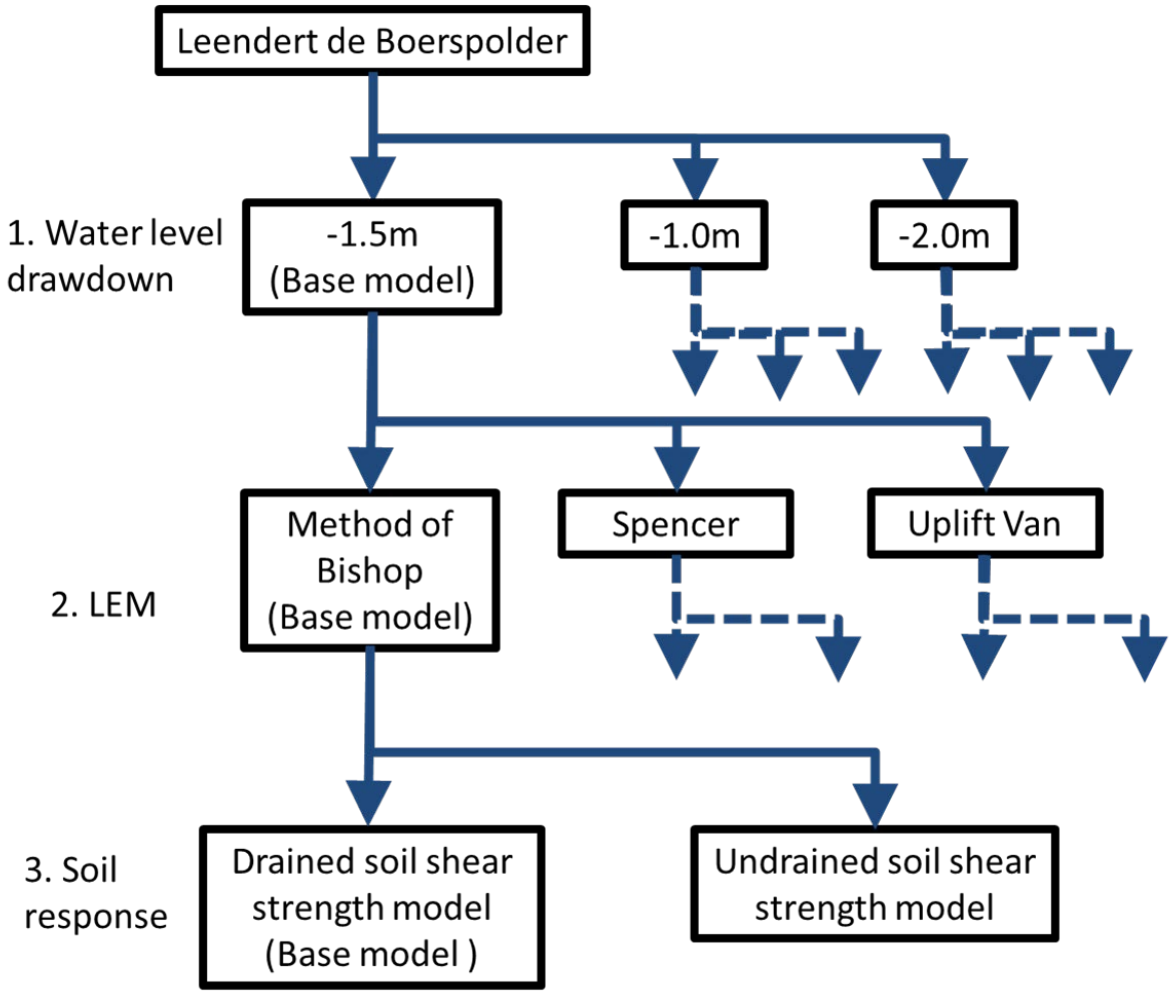


Figure 4.8 Overview of all scenarios of water level drawdown and model choices to vary in the sensitivity analysis one by one, resulting in 6 slope stability models in total. Level 1: Water level drawdown, Level 2: LEM, and Level 3: Soil shear strength model. The base model is implemented with a water level drawdown of 1.5 m (observed water level drawdown during failure), LEM of Bishop and drained soil shear strength model. This results in a total of one base model and five alternative models.

As a second step of the deterministic method, observational information is included in the analysis as an additional level of information. Following the original deterministic method of hindcasting, in this analysis, it was chosen to assess the geometry of the computed critical slip surface on its natural resemblance with the observed slip surface using a free search grid. Hence, finally, the FoS_{3D} is taken into account in the analysis given that the critical slip surface naturally resembled the observed slip surface.

4.3.3 Results of the deterministic analysis

Following the previously described method, the resulting FoS_{3D} values are presented for 6 slope stability models (model numbers 1 to 6 in Table 4.4). These slope stability models consist of sets of a water level drawdown scenario and model choices according to Figure 4.8 and three additional for elaboration purposes (model numbers 7 to 9 in Table 4.4). In short, only three scenarios of water level drawdown are evaluated. The three additional stability models are used to get more specific insights into how the different model choices influence the FoS_{3D}. As a first step, the most likely set of water level drawdowns explaining the failure and model choices are identified based on the resulting values of FoS_{3D}. This is done by looking at increasing water level drawdowns. The most likely water level drawdown explaining the failure is the first water level drawdown with model choices that result in $FoS_{3D}<1$. As a second step, the identified a most likely set of the scenario of water level drawdown explaining the failure and model choices are valued visually on the natural resemblances between the resulting computed critical slip surface and the observed slip surface during failure.

Table 4.4 Summary of results over six slope stability models in total and one slope stability model for elaboration purposes. The FoS is given per water level drawdown in combination with different alternatives of model choices conform to Figure 4.8 and is also corrected for three-dimensional effects with the suggested factor of 1.13 by de Gast (2020). This is noted as FoS_{3D} instead of FoS. The base slope stability model is implemented with a water level drawdown of 1.5 m (water level drawdown observed during failure), the Bishop method and the drained shear strength model and highlighted in bold. *computed slip surface matches the geometry of the observed slip surface.

Model number	Scenario of water level drawdown and model choices	FoS_{3D}
1	Water level drawdown 1.0m-Bishop-drained	1.04
2	**Water level drawdown 1.5m-Bishop-drained (base slope stability model)**	**0.89**
3	Water level drawdown 2.0m-Bishop-drained	0.81
4	Water level drawdown 1.5m-Uplift-drained	0.88
5	Water level drawdown 1.5m-Spencer-drained	1.06
6	Water level drawdown 1.5m-Bishop-undrained	1.02*
7	Water level drawdown 2.0m-Bishop-undrained	0.96*
8	Water level drawdown 1.5m-Spencer-undrained	1.04*
9	Water level drawdown 2.0m-Spencer-undrained	0.98*

In step 1: The set of a water level drawdown of 2.0 m, LEM of Bishop, and drained shear strength model results in the smallest value of $FoS_{3D}=0.81$. However, the situation when the expected driving moment exceeds the expected resisting moment for the first time ($FoS_{3D}<1$) is expected to explain the failure and affect the levee most, i.e. taking into account the staged excavation of the ditch. This is the case for the LEM of Bishop and drained shear strength model with a water level drawdown of 1.5 m (resulting in a $FoS_{3D}=0.89$) and suggests a water level drawdown between 1.0 m (where FoS_{3D} is just above 1.0) and 1.5 m is explaining failure best.

In step 2: The geometry of the observed slip surface is included, as shown in Figure 4.9. The slope stability models that are implemented with an undrained shear strength model are producing computed critical slip surfaces that are showing the most geometry similarities with the observed slip surface. The slope stability model that is implemented with a set of a water level drawdown of 1.5 m, LEM of Bishop, and a drained shear strength model is resulting in a critical slip surface with a FoS_{3D} =0.89. But the geometry of the computed critical slip surface shows very few similarities with

the geometry of the observed slip surface (Figure 4.9). The geometry of the critical slip surface that was computed using a slope stability model implemented with a water level drawdown of 2.0 m, LEM of Spencer, and undrained shear strength model (FoS_{3D} =0.98) shows more similarities with the geometry of the observed slip surface and still predicts computational failure (FoS_{3D}<1) (Figure 4.9). The slope stability model that was implemented with the LEM of Spencer and the undrained shear strength model resulted in FoS_{3D}=1.04 for a water level drawdown of 1.5 m, and FoS_{3D}=0.98 for a water level drawdown of 2 m. This suggests that the failure occurred between these steps.

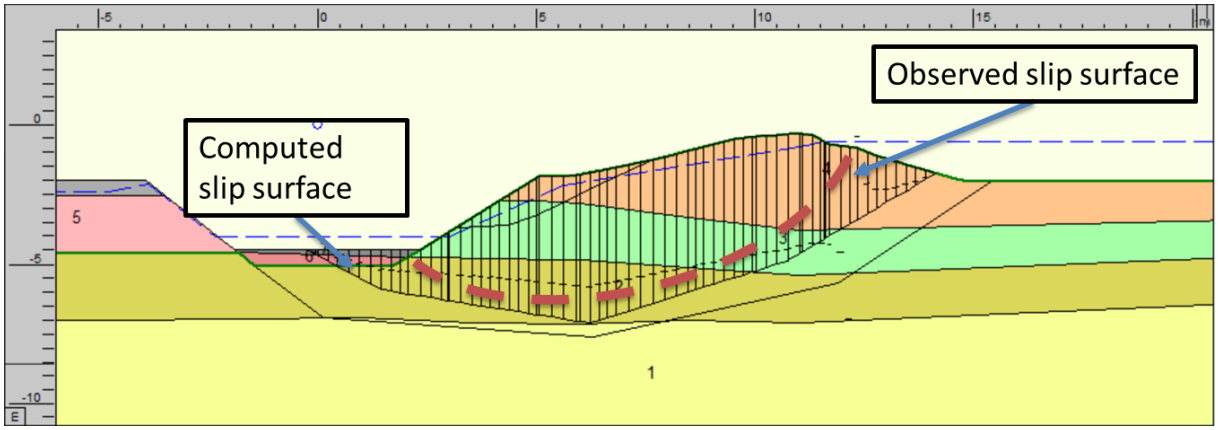


Figure 4.9 Geometry of a computed critical slip surface when undrained shear strength model, LEM of Spencer, and water level drawdown of 2.0 m are implemented in a slope stability model. This results in a FoS_{3D} of 0.98. The observed slip surface is marked with a dashed brown line.

4.3.4 Probabilistic method using Bayesian techniques

The Bayesian method enables the exploration of sets of a wide range of water level drawdown and model choices as well as prior estimates of the likelihood of these. This is done to identify the most likely set of water level drawdown and model choices that most accurately characterizes the failure in six steps. As these can be used in a slope stability model, which is used for hindcasting purposes. Since the Bayesian method is more complex than the deterministic method, the Bayesian method is more elaborately explained. Contrary to what the title of this chapter suggests, no actual probability computations were done. Instead, the FoS was computed deterministically, and the probability of failure is determined using a generic relation between the FoS and the probability of failure.

In this section, the general method is explained first, followed by the application of the method to the Leendert de Boerspolder-experiment and the results are shown in section 4.3.5.

Steps of Bayesian hindcasting of slope instability

The Bayesian method of hindcasting is proposed in six steps following (Kool et al., 2020) with one deviating step to account for the stepwise increase of water level drawdown. To account for this a correction for surviving the previous water level drawdown has been included in the probability calculation (in step 4):

1. The system is described and possible failure scenarios are identified, i.e. different water level levels due to water level drawdown in the ditch at the toe of the levee in the case of the Leendert de Boerspolder-experiment. Prior probabilities are assigned to all water level drawdowns (D_i). Since no prior knowledge is obtainable to distinguish between the scenarios of water level drawdown, a uniform prior is chosen:

$$P(D_i) \tag{4-2}$$

2. Alternatives for model choices and parameter choices (X) are introduced into the scenarios of different water level drawdowns (D_i) and slope stability models are built. These are noted as ($M_j \cap D_i$). Each model choice is assigned a prior probability:

$$P(M_j|D_i) \tag{4-3}$$

3. The limit state function is evaluated to determine the likelihood using the built slope stability models ($D_i \cap M_j$), one by one:
   A. The conditional probability of event F (likelihood (A)):

$$P(F|M_j \cap D_i) \tag{4-4}$$

   B. The conditional probability of F (likelihood (B)) including the observed slip surface information (h):

$$P(F \cap h|M_j \cap D_i) \tag{4-5}$$

4. The computed likelihood of failure (step 3a or step 3b) is corrected for the fact that the levee survived the previous water level drawdown. This probability shows the likelihood that the levee survived the previous loading step and fails at the considered one:

$$P(F_{corr} \cap h|D_i) = (\bar{F} \cap h|D_{i-1})^{i-1}(F \cap h|D_i) \tag{4-6}$$

5. Prior probabilities of water level drawdown and model choices are updated into posterior probabilities using likelihood (A) given failure and (B) slip surface information (h):

$$P(D_i|F_{corr} \cap h) \tag{4-7}$$

$$P(M_j|F_{corr} \cap h) \tag{4-8}$$

6. The set of the most likely water level drawdown and model choices are identified by the resulting highest posterior probabilities:

$$max\ P(D_i|F_{corr} \cap h) \tag{4-9}$$

$$max\ P(M_j|F_{corr} \cap h) \tag{4-10}$$

System description and inputs (Step 1 and Step 2)

As said in the deterministic method, the performance of a levee depends on various typical variables which influence the driving moment (M_S) and the resisting moment (M_R). The Bayesian method is based on the probability that the driving moment exceeds the resisting moment. A simplified limit state function (Z) is used to describe when the driving moment exceeds the resisting moment.

$$\mathrm{Z} = \mathrm{FoS} - 1 = \frac{M_R}{M_S} - 1 \tag{4-11}$$

In which the FoS is the factor of safety and the ratio between the driving moment (M_S) and the resisting moment (M_R). Combinations of driving moments (M_S) and resisting moments (M_R) that result in negative values of Z and FoS smaller than 1 produce slope instability. As discussed in

section 4.3.1, the FoS is influenced by various variables e.g. water level drawdown and alternative model choices, therefore the limit state is influenced as well. The data as presented in section 4.2 is analysed for possible scenarios of water level drawdown explaining the failure ($P(D_i)$). Alternatives for model choices and parameter choices are introduced to characterize the levee's physical response to the different scenarios of water level drawdown ($P(M_j)$). Alternatives of failure scenarios of water level drawdown and model choices are used to build slope stability models to characterize different possible characterizations of the failure ($(D_i \cap M_j)$). A uniform prior is chosen since no knowledge is available to differentiate between the different water level drawdowns and alternative model choices.

Likelihood given observed failure (step 3)

The built slope stability model (as a result of steps 1 and 2) is used to compute a FoS, which is used to compute the conditional probability of failure. The computed conditional probability of failure is taken as the likelihood given failure (3A). This is done for each built slope stability model and its corresponding FoS, which result from steps 1 and 2. The resulting factors of safety (FoS) are connected with the probability of failure using the relation as suggested (Appendix C which is based on results provided by Kanning (2017)). Originally, this relation was used to identify a minimum FoS to comply with a levee to meet a target reliability as support for a reliability assessment of a levee. To this end, an upper bound is used and the relation is fitted to the 20% quantiles of the reliability indexes. However, for hindcasting purposes, the relation is fitted to 50% of the reliability indexes. The probability of failure given the built slope stability models $(D_i \cap M_j)$ is computed as a reliability index (β_H).

$$\beta_H = g(FoS) = (FoS - 0.96)/0.13 \tag{4-12}$$

In which, FoS is the ratio between the mean driving moment (M_S) and the mean resisting moment (M_R) where the material factor $\gamma_m = 1$ is applied. The probability of failure given a slope stability model ($P(F|D_i \cap M_j)$) can easily be computed once a reliability index is known:

$$P(F|M_j \cap D_i) = \Phi(-\beta_H) \tag{4-13}$$

In which Φ is the cumulative standard normal distribution and β_H the computed reliability index.

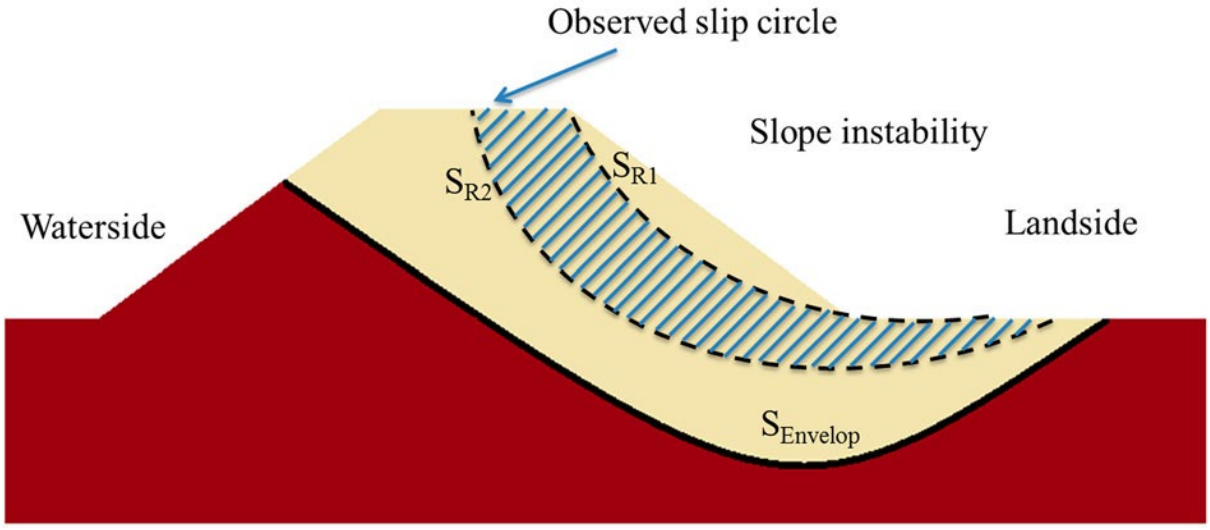


Figure 4.10 The yellow area is where the critical slip surface can be located. The shaded area marks the region of the observed slip surface (h') (Kool et al., 2020)

The collected slip surface information is subsequently used in step 3B. to estimate a region in which the slip surface might have occurred (Figure 4.10). This means that the slope stability

models that naturally produce a critical slip surface that correspond to the observed information (h) are taken into account in the analysis and other slip surfaces are excluded:

$$P(F \cap h | M_j \cap D_i) \tag{4-14}$$

Since the observation of a slip surface is not exact, h is indicating an estimated region with an upper boundary and a lower boundary (Figure 4.10).

Correcting the likelihood of failure and identifying the most likely water level drawdown and model choices by posterior probabilities (step 4 - Step 6)

The most likely water level drawdown and model choices characterizing the levee failure most accurately are identified using the posterior probabilities (in step 6). Whereas, the posterior probabilities (step 5) result from the earlier identified prior probabilities of water level drawdown and model choices (steps 1 and 2), the computed conditional probability of failure (step 3), and using these in a tree diagram. All are demonstrated using a simple case as an example. In this simple case, steps 1 to 6 are applied. To explain steps 4 to 6 effectively, steps 1 to 3 are included in this section as well. Moreover, in this section, it is shown how the conditional probabilities of failure (step 3) are corrected for surviving previous steps of water level drawdown (step 4).

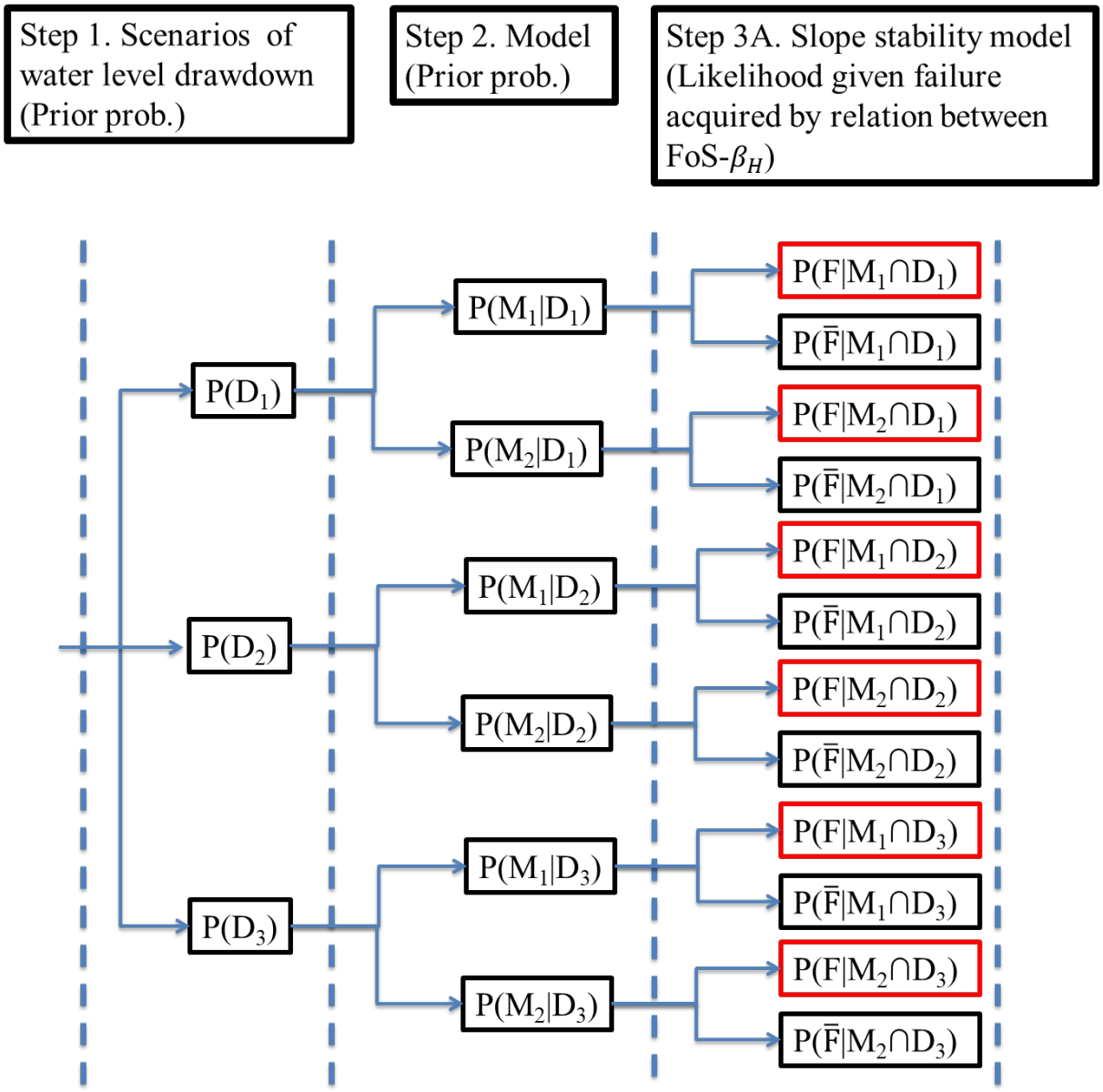


Figure 4.11 Event tree including the scenarios of water level drawdown ("D_1", "D_2", and "D_3") and the alternative models ("M_1" and "M_2"). The resulting slope stability models are used to compute the related likelihood given failure ("$F|D \cap M$") and non-failure ("$\bar{F}|D \cap M$")

The posterior probability is computed using a tree diagram in which all sets of water level drawdown and model choices are included (Figure 4.11).). In this simple case as an example, the failure is assessed using 2 models ("M_1" and "M_2") and the water level in the polder is lowered in 3 steps to induce the levee to failure ($D_1 = 0.0$ m, $D_2 = 1.0$ m, and $D_3 = 2.0$ m) (steps 1 and 2). From the stem of the tree diagram, the first level of branch nodes contains all scenarios of water level drawdown (D_1, D_2, D_3) while the second branch nodes contain the alternatives of model choices (M_1, M_2 as shown in Figure 4.11). All branch nodes are assigned prior probabilities, which are discussed later in this section.

Since no prior knowledge is available to distinguish the scenarios of water level drawdown and model choices, a uniform prior can be chosen. Prior probabilities are $P(D_1) = P(D_2) = P(D_3) = 0.33$ and $P(M_1) = P(M_2) = 0.5$. The six slope stability models ($M \cap D$) are assumed mutually exclusive, while the variables D_i M_j are assumed independent. The sum of the computed probabilities of failure is unioned into the total probability of failure (P(F)), which shows whether the failure was likely to happen or not before the event (i.e. a-priori):

$$P(F) = \sum P(F|M_j \cap D_i) P(M_j|D_i) P(D_i) \tag{4-15}$$

Together, all sets of water level drawdown and model choices are cumulating in six slope stability models ($M \cap D$) and therefore result in $P(M_j \cap D_i) = 0.165$. As part of step 3, each slope stability model is used to evaluate the ratio between driving moments and resisting moments which is expressed in a FoS. As shown in Figure 4.12, the resulting FoS will decrease as the water level is drawn down. This FoS is computed into a corresponding probability of failure ($P(F(FoS)|M \cap D)) = P(F|M \cap D)$) as suggested under step 3 presented in the previous section. These probabilities of failure are included in the leaf nodes of the tree diagram as shown in Figure 4.11 and taken as a likelihood. The tree diagram (Figure 4.11) and Edwards-Venn-based diagrams (Figure 4.13), which are later discussed in this section, provide an overview of the resulting 6 slope models.

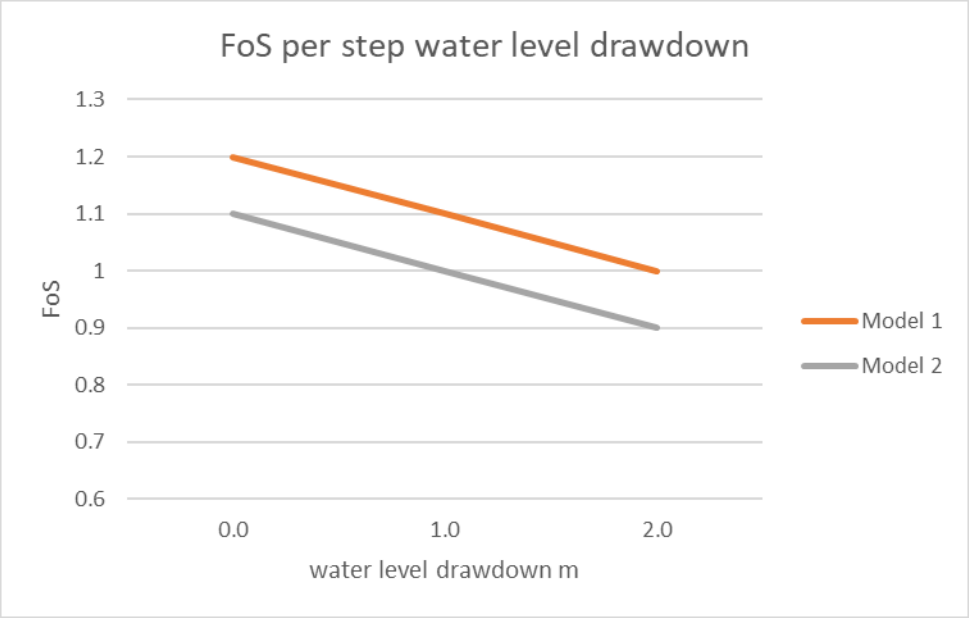
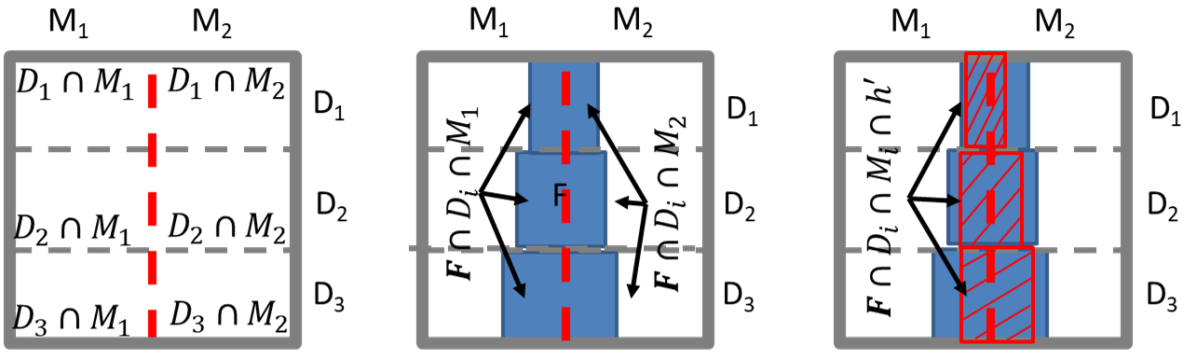


Figure 4.12 Overview of FoS per step of water level drawdown and model choices which are used to compute the probability of failure per step of water level drawdown (P(F|D)).

Since in this analysis the slip surface information is available (h), the set of likelihoods given failure (step 3A, Figure 4.13) is replaced by a set of likelihoods given failure and observation information (step 3B, Figure 4.13). This means that the slope stability models that produce a critical slip surface that naturally correspond to the observed slip surface information (h) are taken into account in the analysis and other slip surfaces are excluded.



A. Possible outcome space of the models and scenarios of water level drawdown (prior probabilities, step 1 and 2). Edwards-Venn diagram with a surface of 1 (Ω=1)

B. Failure space (step 3A, blue area)

C. Failure space that is updated on the observation information (step 3B, red shaded area)

Figure 4.13 A. Edwards-Venn based diagram presents the relation between the different steps of water level drawdown (D_1, D_2, D_3) and the model choices (M_1, M_2) that result in six slope stability models that are built ($D_i \cap M_j$) for calculating the probability of failure which is displayed under step B of this figure. The event of failure is noted as F and marked as a blue area. Events of steps of water level drawdown and the model choices include prior probabilities. B. Each slope stability model computes a probability of failure ($F \cap D_i \cap M_j$) which is marked as a blue area in the Edwards-Venn based diagram (step 3A). C. The red-shaded area incorporates the observed slip surface information (h') as a subset of F (step 3B). The most likely scenario of water level drawdown (D_i) and model choices (M_j) are found.

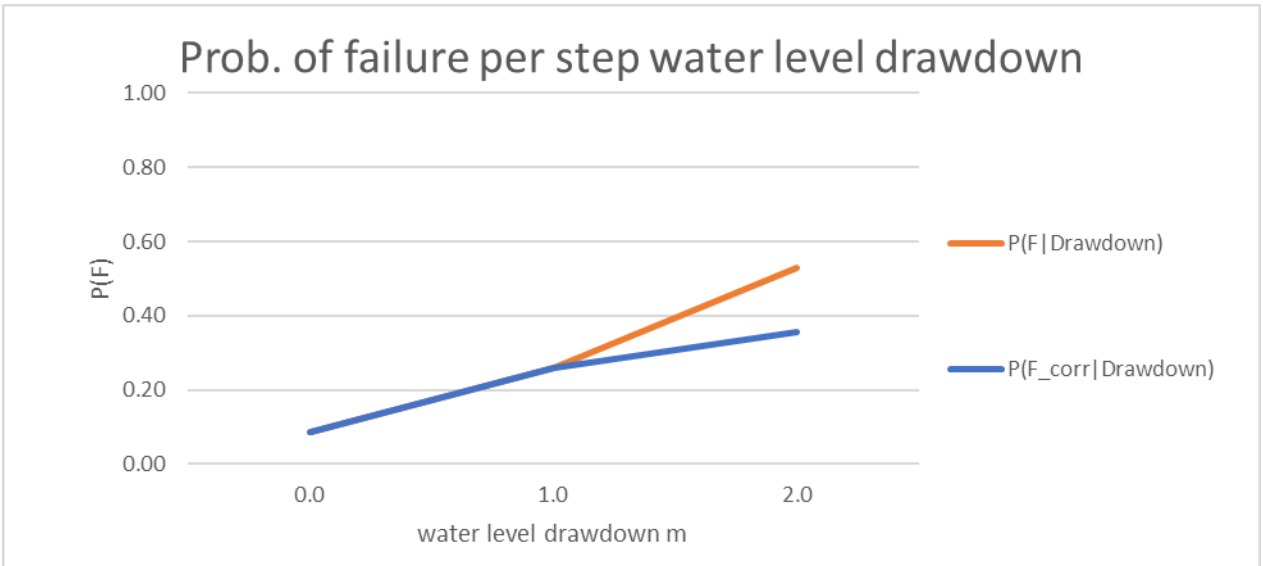


Figure 4.14 Overview of probabilities per step of water level drawdown, i.e. 0m, 1m and 2m. The probabilities of failure $P(F|D)$, coming from step 3A and illustrated in the orange line, is used to calculate the probability of occurrence of failure and non-failure of the previous water level drawdown $P(F_{corr}|D)$, as step 4. Hence, that the levee survived the previous step of the water level drawdown. The calculated failure probability is corrected into the probability that this step of water level drawdown is the first step of water level drawdown that explains the failure ($P(F_{corr}|D_i)$) (step 4) and is illustrated in the blue line. The probability that a water level drawdown is the first water level to trigger the failure per slope stability model ($P(F_{corr}|D_i \cap M_j)$) is calculated and implemented in the tree diagram.

These computed probabilities given failure and observation information (step 3B) per water level drawdown (i) are corrected for the fact that the levee survived the previous levels of water level drawdown (i-1). Although the boundary conditions are not fully met, this principle is based on the geometrical distribution (Jonkman et al., 2016). As shown in Figure 4.14, assessing the water level drawdown as separate events will lead to an increased probability of failure per increasing water level drawdown. Even so, the assessment is performed to identify the first water level in the sequence of successive water level drawdowns to trigger the failure. In the example, there are three steps of water level drawdown ($P(D_1) = P(D_2) = P(D_3)$) and two model choices ($P(M_1) = P(M_2)$). First, the union of the probability of failure given a water level drawdown ($P(F|D_i)$) are computed using the rule of total probability. Where the probability of failure per water level drawdown and model choices are summed over the model choices:

$$P(F|D_i) = \sum_{j=1}^{2} P(F|M_j \cap D_i) \tag{4-16}$$

Then each probability of failure per water level drawdown is computed including the probabilities of survival of the previous water level drawdown (Figure 4.14). This is done for the third level of water level drawdown (D_3), as an illustration:

$$P(F_{corr}|D_3) = (F|D_3)(\bar{F}|D_2)(\bar{F}|D_1) \tag{4-17}$$

In this the probability of survival per level of water level drawdown is $P(\bar{F}|D_i) = 1 - P(F|D_i)$. After correction, the calculated probability of failure per water level drawdown ($P(F_{corr}|D_i)$) is distributed back over the probability of failure given the model choices per level of water level drawdown $P(F_{corr}|M_j \cap D_i)$. The corrected probability of failure given the model choices per level of water level drawdown is computed using their relative contribution to the probability of failure per water level drawdown. This is done for the third level of water level drawdown (D_3), as an illustration:

$$P(F_{corr}|M_1 \cap D_3) = P(F_{corr}|D_3)P(F|M_1 \cap D_3)/P(F|D_3) \tag{4-18}$$

$$P(F_{corr}|M_2 \cap D_3) = P(F_{corr}|D_3)P(F|M_2 \cap D_3)/P(F|D_3) \tag{4-19}$$

The set of likelihoods given the failure in the tree diagram is replaced by the set of corrected probabilities given failure which is taken as likelihoods again, as illustrated in Figure 4.11.

Similar to the method in Kool et al. (2020), the posterior probabilities of the water level drawdown and model choice are evaluated by:

$$P(D_1|F_{corr}) = \frac{\sum_{j=1}^{2} P(F_{corr}|M_j \cap D_1)P(M_j|D_1)P(D_1)}{P(F_{corr})} \tag{4-20}$$

$$P(M_1|F_{corr}) = \frac{\sum_{i=1}^{3} P(F_{corr}|D_i \cap M_1)P(D_i|M_1)P(M_1)}{P(F_{corr})} \tag{4-21}$$

The most likely level of water level drawdown leading to failure and model choice to characterize the levee most accurately is obtained by evaluating the equation above for all alternatives of water level drawdown and model choices.

Water level Drawdown and model choices to characterize the Leendert de Boerspolder experiment

In total, 36 slope stability models are generated to substantiate the identification of the most likely water level drawdown and model choices to characterize the failure using the Bayesian method. An overview of the 36 slope stability models is presented in Figure 4.15, which follows from the 6 scenarios of water level drawdown (D), three LEM (M) and two possible soil shear strength models (SS) that are possible (i.e. $6\ drawdown\ levels\ (D) \times 3\ LEMs(M) \times 2\ soil\ shear\ strength\ models\ (SS) = 36\ slope\ stability\ models\ (SS \cap M \cap D)$). Again as done in the deterministic method of analysis, all scenarios of water level drawdown and model choices are based on computations by de Gast (2019). Because the analyses of de Gast (2019) and this analysis use the same boundary conditions, the results of both analyses are comparable. In this section, an overview of all the different sets of scenarios of water level drawdown and model choices are presented along with their prior probabilities.

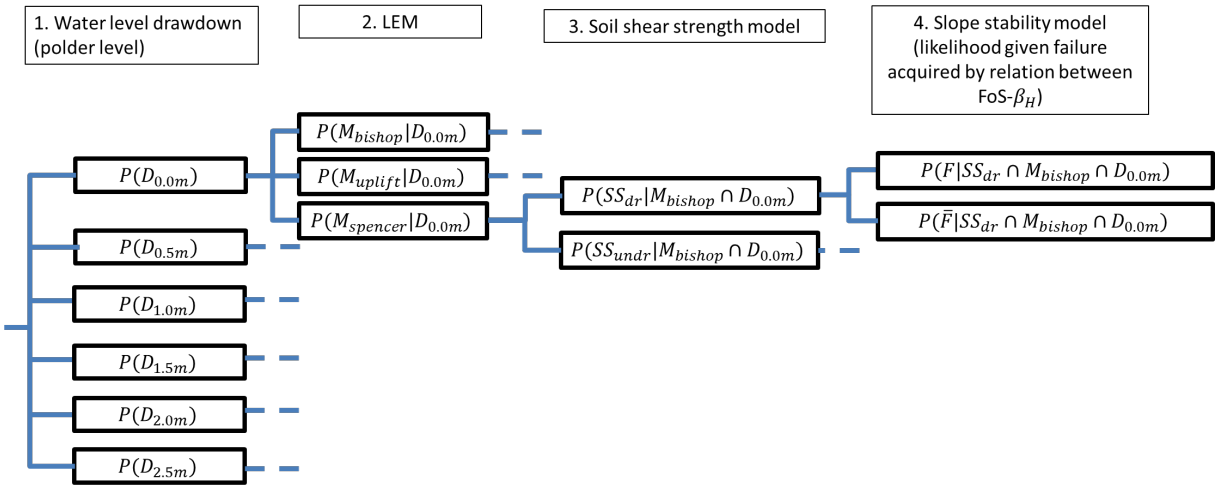


Figure 4.15 presents an overview of the different sets of scenarios of water level drawdown and model choices. Level 1: Scenarios of water level drawdown (i.e. $D_{0.0m}$, $D_{0.5m}$, $D_{1.0m}$, $D_{1.5m}$, $D_{2.0m}$, and $D_{2.5m}$) Level 2: LEM (i.e. M_{bishop} ,$M_{spencer}$, and M_{uplift}). Level 3: soil shear strength model (i.e. SS_{dr}: drained, and SS_{undr}: undrained). Level 4: likelihood given failure (F) and non-failure ($\bar{F}$), which follows from evaluating the slope stability models, resulting in FoS and using the relation of FoS-β_H.

The water level at the polder side is drawn down gradually to induce failure (Figure 4.7). The water level in the polder initially is NAP-2.0 m which initiates a head difference of 1.4 m (NAP-0.6 m-NAP-2.0 m=1.4 m). A water-level drawdown from 0 m to 2.5 m in steps of 0.5 m is introduced in the tree diagram as possible scenarios of water level drawdown explaining the failure, i.e. from 0.0 m ($D_{0.0m}$) to 2.5 m ($D_{2.5m}$) (Figure 4.15). Hence, there are six scenarios of water level drawdown in total. The prior probability of each water level drawdown is $P(D_i) = \mathbf{0.167}$

The levee failed due to slope instability after a simple shaped slip surface developed (Figure 4.7). Both LEMs of Bishop (M_{Bishop}), Spencer ($M_{Spencer}$), and Uplift Van ($M_{liftvan}$) are implemented as alternatives to LEM, which are used to characterize the levee's physical response to the gradual drawing down of the water level. Hence, there are used three LEM in total, which means that the prior probability of each model is $P(M_j) = 0.333$.

How much shear strength the soils have generated as a response to the loadings depends on the rate of the loading and the conductivity of the soil, as mentioned earlier. Often, the shear strength of the soil is modelled as a completely drained soil response or undrained soil response, while in reality the soil likely responds as partially drained. Initially, drained and undrained are used in this section as these are most commonly used ($P(SS_{dr})$ and $P(SS_{und})$). An undrained shear strength model has taken the possibly generated water pressures by deformations into account, in contrast to a drained shear strength model. Hence, there are used two types of soil shear strength models to characterize the shear strength in total, which means that the prior probability of each soil response is $P(\boldsymbol{SS}_k) = 0.5$.

After combining all scenarios of water level drawdown and model choices, 36 slope stability models ($SS \cap M \cap D$) are generated (steps 1 and 2) to compute the likelihood given failure ($P(\mathrm{F}|SS \cap M \cap D$), step 3A). As an additional step, the observational information on the slip surface is implemented in the stability models ($P(\mathrm{F} \cap h|SS \cap M \cap D$), step 3B). This is done by restricting the search area for the most critical slip surface in the slope stability models. These generated slope stability models have provided for a calculated and deterministic ratio between the driving moment and the resisting moment on which the further steps of the Bayesian method (steps 4 to step 6) elaborate.

4.3.5 Results Bayesian method

Factor of safety and probability of failure as a function of water level drawdown

The FoS has resulted from the earlier built 36 slope stability models (step 1 and step 2) and are used to compute the conditional probabilities of failure ($P(\mathrm{F}|SS \cap M \cap D$) in step 3A and $P(\mathrm{F} \cap h|SS \cap M \cap D$) in step 3B). Each slope stability model contains a unique set of a scenario of water level drawdown (D), a LEM (M) and a soil shear strength model (SS) to characterize the failure. As the resisting moment of the slope of the levee degreases with every step of the water-level drawdown, the value of the different computed FoS is strongly influenced by the water-level drawdown. Therefore, the computed conditional probabilities of failure have to be corrected for surviving the previous steps of the water level drawdown (step 4). In this section, it is shown how both the resulting FoS and the probability of failure of the levee are influenced by the water level drawdown and model choices.

In Figure 4.16, the top panels show the FoS per water level drawdown, whereby the observed slip surface is not incorporated in the slope stability model (on the left the slope stability model is implemented with a drained shear strength model and on the right with undrained shear strength model). In the lower panels, the FoS is presented per water level drawdown where the observed slip surface information is incorporated into the slope stability model. The FoS values from the deterministic analysis can be found in the top panels. Generically, the results of FoS resulting from the slope stability model are increased when the observed slip surface information is incorporated in the slope stability model because other slip surfaces than observed are excluded.

The slope stability models result in approximately the same FoS, where the observed slip surface information is not incorporated, implemented with the drained shear strength model, and either the LEM of Bishop or Uplift. This is also to be expected, given no uplift conditions are present. Slope stability models have returned the same FoS, that are implemented with the undrained shear strength model, and with LEM of Bishop or LEM of Spencer. While slope stability models have estimated the capacity of the levee optimistically, they are implemented with the undrained shear strength model and LEM of Uplift. The slope stability models cluster when the observation information is incorporated into the slope stability model and are implemented with the drained shear strength model and the LEM of Bishop or LEM of Uplift. While the slope stability model has estimated the capacity of the levee conservatively, that is implemented with the drained shear strength model and LEM of Spencer.

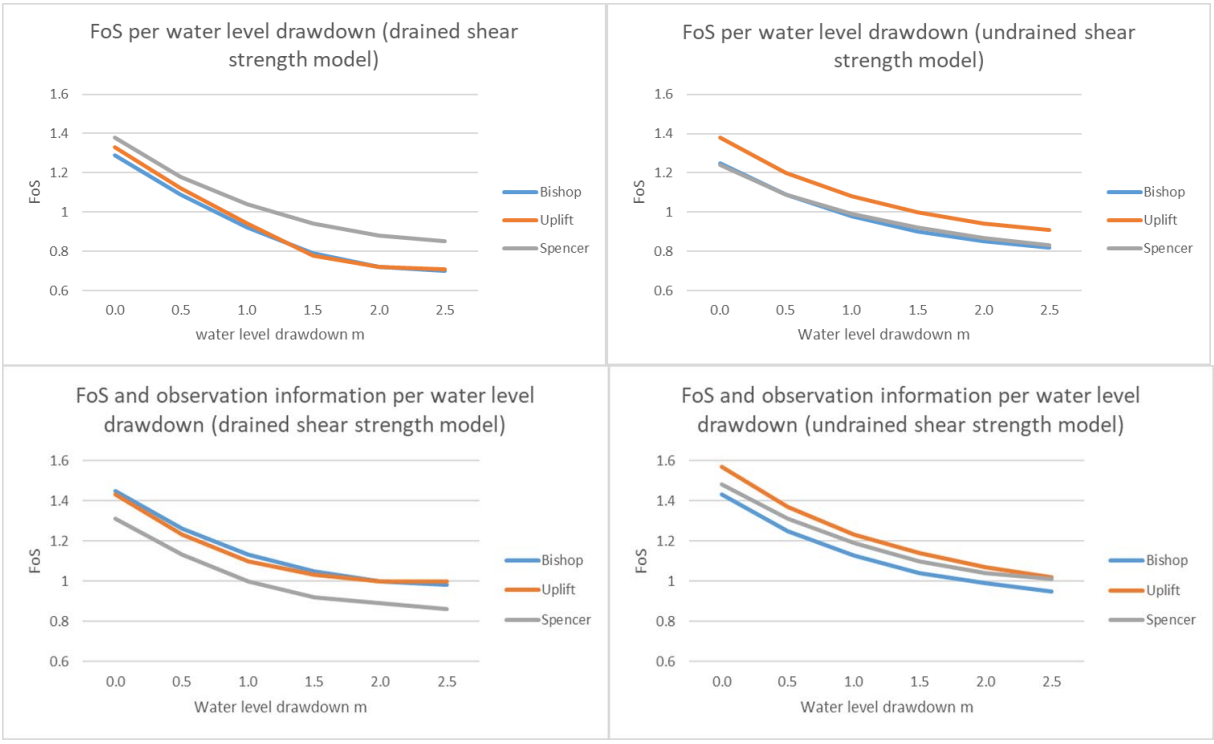


Figure 4.16 The FoS per water level drawdown is used as input to compute the probability of failure. The top-left plane shows the FoS per water level drawdown and different LEMs given a drained shear strength model. The top-right plane shows the FoS per water level drawdown and different LEMs given an undrained shear strength model. The lower-left plane shows the FoS per water level drawdown and different LEMs given a drained shear strength model and observed geometrical information of the slip surface. The lower-right plane shows the FoS per water level drawdown and different LEMs given an undrained shear strength model and observed geometrical information of the slip surface.

The conditional probabilities of failure given the branch nodes (presented in Figure 4.15 in the previous section) are included in the leaf nodes of the tree diagram that is used in the Bayesian method, as computed under steps 3A and 3B. In Figure 4.17, the failure probabilities over the different LEM per water level drawdown have been combined and computed in the total probability that the levee will fail at this specific water level drawdown (following the completed event tree in Figure 4.15). Since the different steps of water level drawdown are sequential, the computed probabilities of failure are corrected for the probabilities that the levee survived the earlier steps of water level drawdown, as shown in Figure 4.17 and stated earlier. The results show that the largest water level drawdown is less probable to be the first step of the water level drawdown explaining the failure when the probability of failure is corrected. When the observational information is included in the analysis, the probability of failure becomes generally smaller per step of the water level drawdown. In the corrected probability of failure, the highest probability of failure is assigned to a water level drawdown between 1.5 m and 2.0 m. These calculated probabilities of failure are further used as likelihoods of failure (step 4 of the Bayesian method).

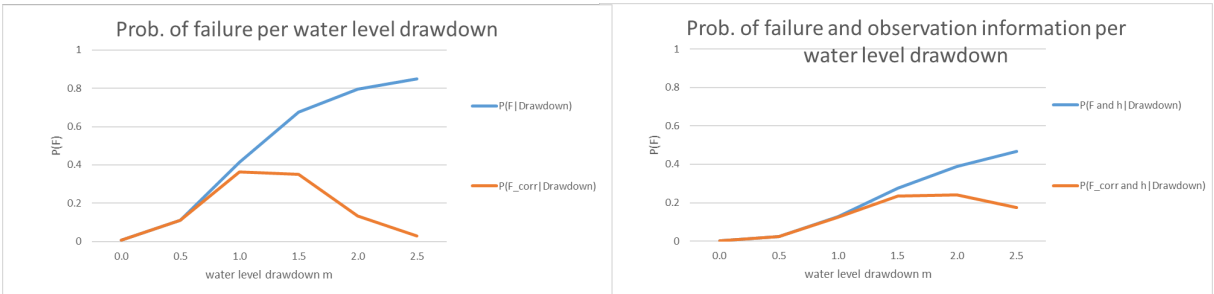


Figure 4.17 The probability of failure per water level drawdown is in blue. The corrected probability of failure per water level drawdown is in orange (step 4). The left panel shows the probability of failure as a result of slope stability models where the observed slip surface information is not incorporated in the slope stability model, as done under step 3A. The right panel shows the probability of failure as a result of slope stability models where the observed slip surface information (h) is incorporated, as done under step 3B.

Updated water level drawdown probabilities

The posterior probabilities of the water level drawdown given failure ($P(D_i|F)$) are used to identify the most likely water level drawdown explaining the failure (step 6). These posterior probabilities have resulted from combining prior probabilities of water level drawdown with likelihoods that result from the earlier generated 36 slope stability models. Whereas, the value of these likelihoods is strongly influenced by the water level drawdown as shown if not corrected in this section. In this section the most likely water level drawdown explaining the failure is identified, based on likelihoods:

- That is not corrected for surviving the previous water level drawdown as a first step and that results in a slope stability model where the observed slip surface was not incorporated (step 3A, Figure 4.18).
- That is not corrected for surviving the previous water level drawdown as a first step and that results in a slope stability model where the observed slip surface was incorporated (step 3B, Figure 4.18).
- That is corrected for the survival of the previous water level drawdown as a second step and that results in a slope stability model where the observed slip surface was not incorporated (step 3A, Figure 4.19).
- That is corrected for the survival of the previous water level drawdown as a second step and that results in a slope stability model where the observed slip surface was incorporated (steps 3B, Figure 4.19)

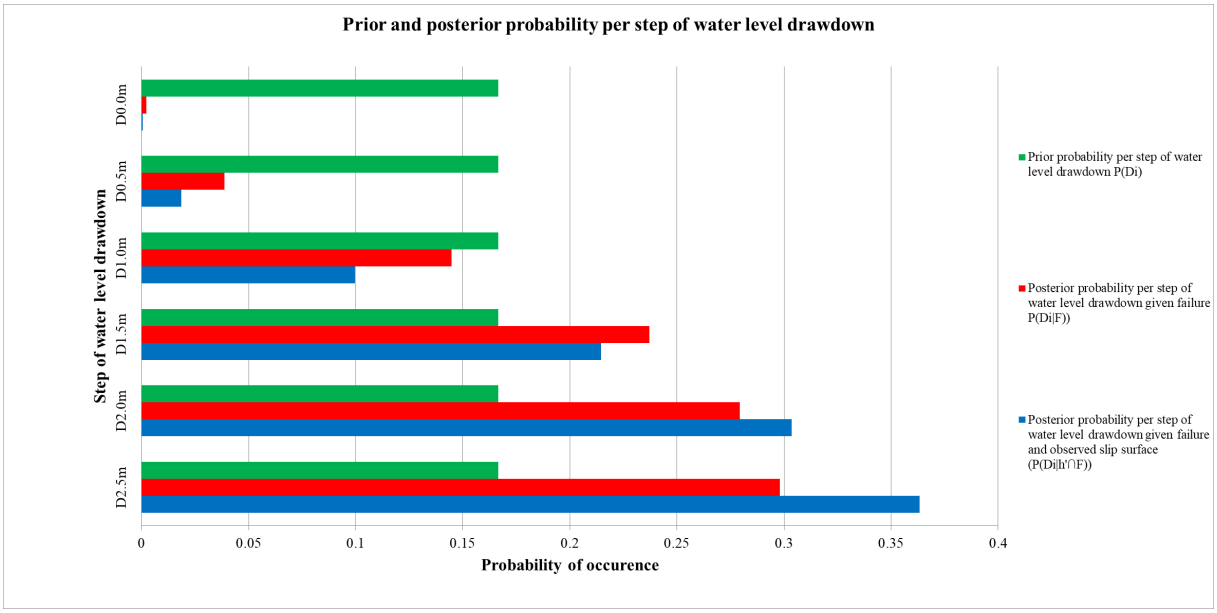


Figure 4.18 Prior probability per water level drawdown (D) at failure regime and equally distributed probabilities (green). The posterior probabilities per water level drawdown given failure (red) and observed geometry of slip surface (blue). These are conditional probabilities that sum over all drawdown levels (i.e. $D_{0.0m}$, $D_{0.5m}$, $D_{1.0m}$, $D_{1.5m}$, $D_{2.0m}$, and $D_{2.5m}$) to one.

Based on the likelihoods that are not corrected for the probability of surviving previous levels of water level drawdown as shown in Figure 4.19, the posterior probability ($P(D_i|F)$) is increasing when the water level drawdown is increasing. This is probably due to the decreasing of the resisting moment with every increasing step of the water level drawdown. The prior probability of the steps of water level drawdown is updated on the information that failure occurred (3A in red) and the observation information (3B in blue). These posterior probabilities per water level drawdown given failure (red) and observed geometry of slip surface (blue) are not based on likelihoods that are corrected for the survival of previous steps of water level drawdown. As shown in Figure 4.18, these posterior probabilities will always identify the highest water level drawdown as the most likely, as the conditional failure probabilities ($P(F|D_i)$) increase with each step of water level drawdown.

Based on the likelihood given failure (red) that is corrected for surviving the previous steps of water level drawdown, both a water level drawdown of 1.0 m and 1.5 m is most likely to result in the most accurate characterisation of the failure (Figure 4.19). Moreover, the posterior probability given corrected failure and observed slip surface information (blue) is concentrated around a water level drawdown of 1.5 m and 2.0 m (($P(D_{1.5m}|F_{corr} \cap h)$) and ($P(D_{2.0m}|F_{corr} \cap h)$, Figure 4.19). The slip surfaces for these water level drawdowns match the observational information better. A water-level drawdown of 1.6-1.7 m at the time of failure was observed. The observation information is thus essential in estimating the most accurate characterisation of the failure. The analysis is insufficiently differentiated (in terms of water level drawdown steps) to make a clear distinction between these two most likely scenarios of water level drawdown.

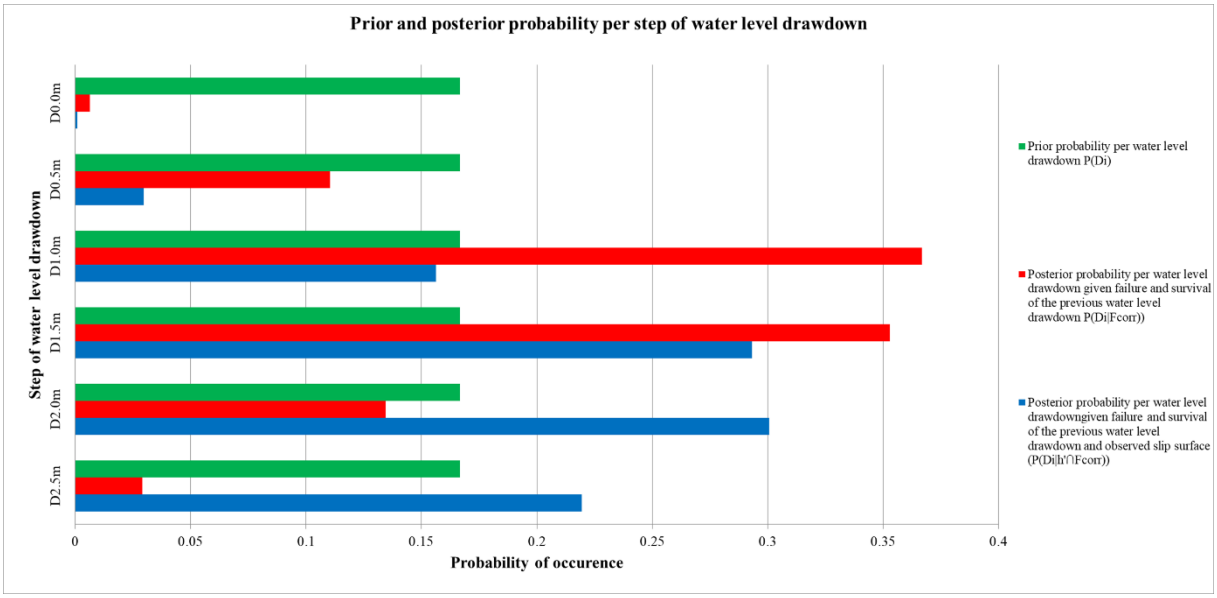


Figure 4.19 Prior probability per water level drawdown (D) at failure step and equally distributed probabilities (green). The posterior probabilities per water level drawdown given failure (red) and observed geometry of slip surface (blue) are corrected for the survival of previous steps of water level drawdown. These are conditional probabilities that sum over all drawdown levels (i.e. $D_{0.0m}$, $D_{0.5m}$, $D_{1.0m}$, $D_{1.5m}$, $D_{2.0m}$, and $D_{2.5m}$) to one.

Updated model choice probabilities

The posterior probabilities of the LEM and the shear strength model of the soil are used to identify the most likely LEM and shear strength model to characterize the failure most accurately (step 6). These posterior probabilities have resulted from combining prior probabilities of model choices with likelihoods that result from the earlier generated 36 slope stability models. In this section the most likely set of LEM and shear strength models to characterize the failure most accurately are identified, based on likelihoods:

- That is corrected for surviving the previous steps of water level drawdown and that results in a slope stability model where the observed slip surface was not incorporated (step 3A, Figure 4.20)
- That is corrected for surviving the previous steps of water level drawdown and that results in a slope stability model where the observed slip surface was incorporated (steps 3B, Figure 4.21).

In both Figure 4.20 and Figure 4.21, both the prior and posterior probabilities are shown.

As shown in Figure 4.20, Spencer's LEM appears the most likely model choice for LEM considering the posterior probability given failure and the observed slip surface (blue), based on the corrected likelihoods ($P(M_i|h \cap F_{corr})$). Again, the incorporation of the geometry of the observed slip surface into the slope stability model influences the posterior probabilities in identifying the most accurate characterisation of the failure. The difference between the posterior of the LEM of Uplift and Bishop is not significant. This is probably because of the simple geometry of the observed slip surface. However, the geometry of the observed slip surface was not completely circular and somewhat superficial. The LEM of Spencer has the property to evaluate somewhat superficial slip surfaces. Therefore the LEM of Spencer finds the geometry of the slip surface that has occurred most accurately. The observation of the slip surface is essential in estimating the most accurate characterisation of the LEM.

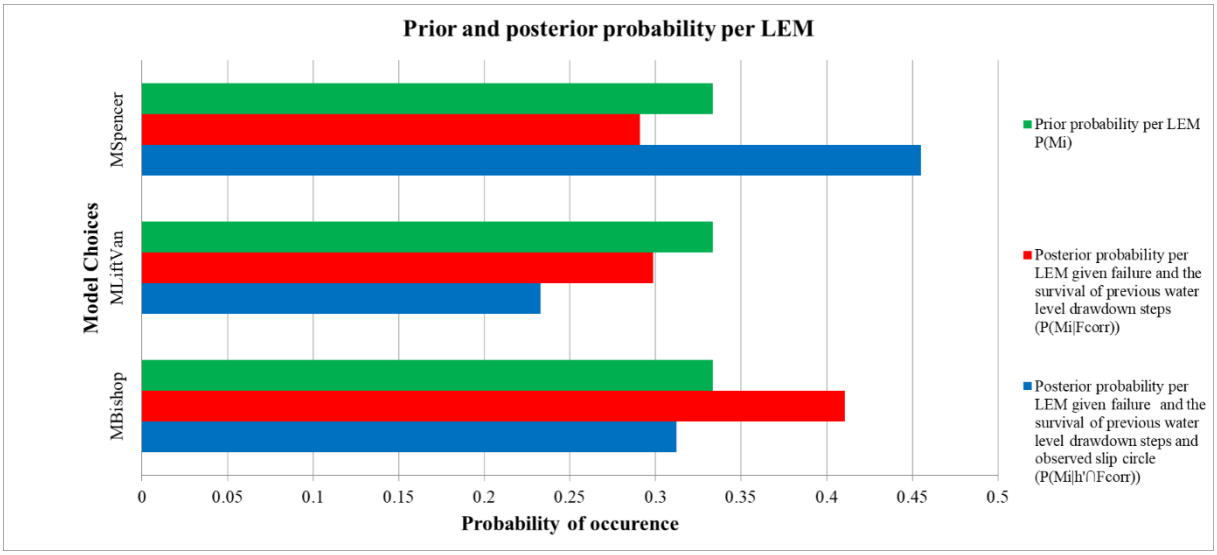


Figure 4.20 Prior probability per the choice of LEM of Uplift Van, Spencer, and Bishop and equally distributed probabilities (green). The posterior probabilities per the choice of LEM given failure (red) and observed geometry of slip surface (blue) are corrected for survival. These are conditional probabilities that sum over both model choices to one.

Conform Figure 4.21, the drained shear strength model provides the best characterization of the soil's shear strength at the time of the failure ($P(SS_i|h \cap F_{corr})$). The posterior probabilities per type of shear strength model given failure are evenly distributed. Again, the incorporation of the geometry of the observed slip surface into the slope stability model influences is essential in estimating the most accurate characterisation of the soil shear strength.

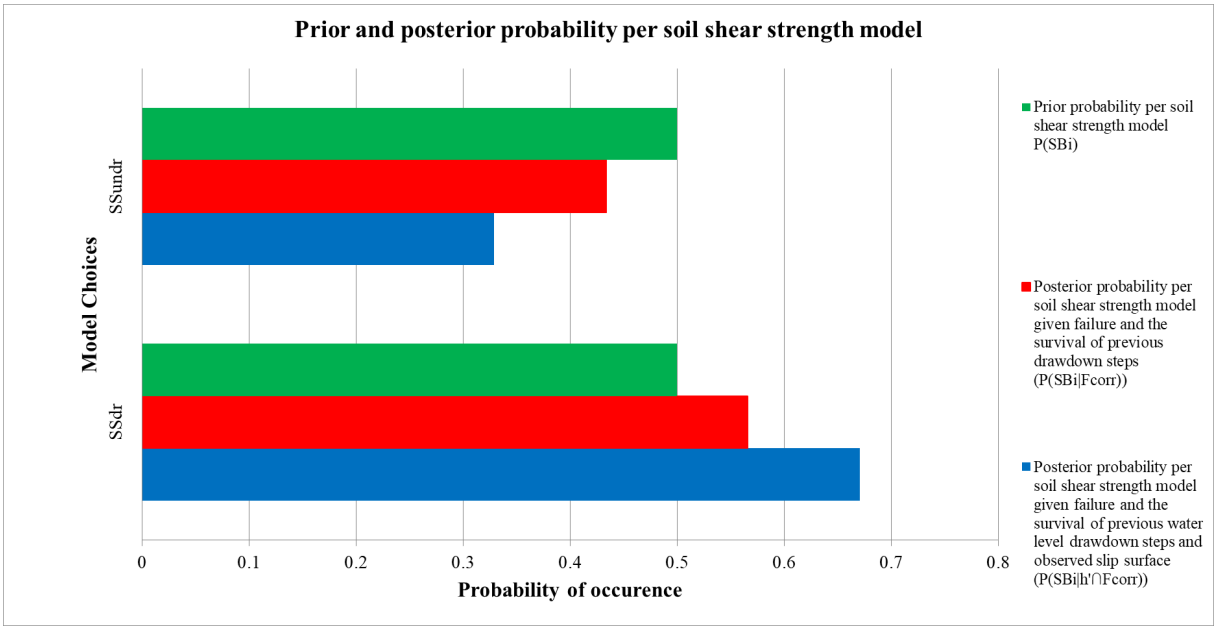


Figure 4.21 Prior probability per the choice of shear strength model both drained (SS_{dr}) and undrained (SS_{undr}) and equally distributed probabilities (green). The posterior probabilities per the choice of shear strength model given failure (red) and observed geometry of slip surface (blue) are corrected for the survival of previous steps of water level drawdown. These are conditional probabilities that sum over both model choices to one.

In summary, a set of a water level drawdown between 1.5 m and 2.0 m, the LEM of Spencer and the drained shear strength model provide the best characterization of failure based on the

posterior probability given failure, the observed slip surface, and the survival of previous water level drawdown steps.

4.4 Discussion

4.4.1 Discussion of results

Both the deterministic and Bayesian methods are expected to identify the most likely set of model choices to characterize the initiated levee failure of the Leendert de Boerspolder. More specifically, both analyses aim to identify the most likely level of water level drawdown explaining the failure, while this is known from the observation. This is 1.6 m of water level drawdown by approximation. An overview of the identified most likely water level drawdown and model choices to characterise the failure are presented in Table 4.5 per method of hindcasting. The table shows the outcomes for a situation, whether the slope stability model was or was not incorporated with the observed slip surface information, and the limitations of both methods. The content of this Table 4.5 will be discussed in the next sections.

Table 4.5 Summary of results over the different methods for the slope stability models that were or were not incorporated with the observed slip surface. Both methods result in the most likely water level drawdown and model choices to characterise the failure

	Deterministic method		Bayesian method	
	Without observed slip surface information	With observed slip surface information	Without observed slip surface information	With observed slip surface information
Most likely water level drawdown	1.0 m to 1.5 m	1.5 m to 2.0 m	1.0 m to 1.5 m	1.5 m to 2.0 m
LEM	Bishop	Spencer	Bishop	Spencer
Shear strength model	Drained	Undrained	Drained	Drained
Remarks (issues/limitations)	• The identification of the water level drawdown and model choices are not based on the likelihood of occurrence. • Small differences in LEM • No computations with fixed grids to allow better incorporation of the observed slip surface		• The cases that are used to derive the $FoS - \beta_H$-relation is not fully representative of this levee. • The process of applying the Bayesian method is time-consuming since all sets of water level drawdown and model choices have to be computed	

4.4.2 Findings based on the deterministic method

The hindcasting process of geotechnical structures is typically dominated by several input and model uncertainties, due to the lack of information about the actual present circumstances explaining the failure. The proposed approach makes all uncertainties explicit in a systematic fashion. Also, the proposed method provides more insights into how the uncertainties from different levels influence the capacity of the levee to some extent, i.e. the water level drawdown, LEM, and shear strength model. This is done by analysing the influence of different scenarios of water level drawdown and model choices on the FoS_{3D}. However, it takes still some effort to interpret the results and identify an unambiguous most likely water level drawdown explaining the failure. This is because the method does not quantify the likelihood of the occurrence of the outcome or the relative accuracy of the different model choices, as noted earlier by Kool et al. (2019). The substantiation of explaining the failure will benefit when it is possible to compare the different steps of water level drawdown and model choices quantitatively to one another as done in the Bayesian method.

Even more in the case of the Leendert de Boerspolder experiment, the difference in FoS_{3D} related to the influence of the LEM of Bishop and Uplift is very small. This is probably because there

are no uplift conditions to which the LEM of Uplift is very sensitive. It is therefore difficult to identify a clear most likely choice of LEM. However, the deterministic method does offer the opportunity to incorporate the order of the water level drawdown such that all loads occur gradually as a result of the drawdown steps. The method allowed the identification of a water level drawdown between 1.5 m and 2.0 m explaining the failure. In this situation, the driving moment exceeded the resisting moment of the slope (FoS_{3D}) for the first time and it is thus considered the most likely water level drawdown explaining the failure. This identified water level drawdown is confirmed as the observed water level drawdown was approximately 1.6-1.7 m.

The results do not distinguish well between the water level drawdown of 1.5 m and 2.0 m as the most likely water level drawdown at the time of failure. Possibly, because the chosen water level drawdown step distance is too coarse. The observed water level drawdown was 1.6-1.7 m and is therefore a value between the chosen water level drawdown. In the absence of this value, the closest water level drawdown is identified as the most likely water level drawdown which are 1.5m and 2.0 m. For future research, a more refined step grid should be used, so more insight is gained into the accuracy of the method.

4.4.3 Findings based on the Bayesian method

The set of a water level drawdown of 1.0 m and 1.5 m, the LEM of Bishop, and the drained shear strength model are identified to characterize the failure most accurately, before including the observed slip surface information. Moreover, the water level drawdown of 1.5 m and 2.0 m and the drained shear strength model are identified to characterize the failure most accurately when the survival of the previous water level and the visual information of the observed slip surface are taken into account. However, the identification of the LEM to characterize the failure most accurately does not follow clearly from the probabilistic calculations. Nevertheless, the LEM of Spencer is identified as LEM to characterize the failure most accurately. The water level drawdown at the moment of the failure was between 1.5 m and 2.0 m, which corresponds to the observed water level drawdown was approximately 1.6-1.7 m. As stated earlier, observational information is crucial for identifying the most accurate characterization.

Similar to the comment in the previous section, the results do not distinguish well between the water level drawdown of 1.5 m and 2.0 m as the most likely water level drawdown at the time of failure. Possibly, because the chosen water level drawdown step distance is too coarse. The observed water level drawdown was 1.6-1.7 m and is therefore a value between the chosen water level drawdown. In the absence of this value, the closest water level drawdowns are identified as the most likely water level drawdown which are 1.5 m and 2.0 m. For future research, a more refined step grid should be used so more insight is gained into the accuracy of the method.

4.4.4 Discussion of deterministic method and implications

Several specific elements of the deterministic method could be improved. The computed critical slip surfaces resulting from the identified sets of water level drawdown and model choices are judged visually on their natural resemblance with the observed slip surface during failure. Due to the reasonable simple geometry of the observed slip surface, several elements in the geometry of the computed critical slip surface can objectively be judged on their resemblance with the observed slip surface, e.g. entry points or the general geometry. However, the geometry of the observed slip surface is not fully incorporated in the analysis which means that the conclusion will always be based on engineering judgement to some extent.

The sets of water level drawdown and model choices that result in critical slip surfaces that show natural resemblances to the observed slip surface result in higher FoS. Hence, hindcasting models are likely to underestimate the capacity of the levee when the observation information is not included because other slip surfaces than observed are excluded. Even more, during the assessment of the reliability of an existing levee, the capacity is likely to be underestimated since the slip surface can be freely assumed (ENW, 2009; Kanning et al., 2016; Schweckendiek, van der Krogt, Rijneveld, et al., 2017).

Generally, it is assumed that the undrained shear strength model parameters should be specified relative to the slope, crest, and polder (3 zones) to provide a more accurate characterisation of the soil's shear strength (RWS, 2019). This is because of 2 reasons. First, the properties of the soil shear strength model in these zones can change significantly due to different load histories along the slip surface. Second, the direction of the slip surface relative to the main stresses in the soil determine the shear strength at any specific point in the slip surface. In the current approach, a single set of undrained shear strength model parameters is. Based on the result of de Gast (2020), the FoS calculated using Bishop increases by 4% (FoS_{3zone}/ FoS_{1zone} =1.28/1.23=1.04), Uplift Van decreases by 37% (FoS_{3zone}/ FoS_{1zone} =1.14/1.79=0.63), and Spencer increases with 10% (FoS_{3zone}/ FoS_{1zone} =1.23/1.12=1.10) as the result of specifying the soil properties relative to these 3 zones. When these ratios are used to correct the results of Table 4.4, a set of water level drawdowns of 0.0 m, Uplift Van and undrained shear strength model are identified to characterize the failure most accurately. These results are based solely on the FoS_{3D} and not on the resemblance of the observed slip surface. A most likely water level drawdown of 0.0 m at the time of failure suggests a situation without any water level drawdown and this does not match with the water level drawdown that was observed at the time of failure. Especially the LEM of Uplift Van is affected by the introduction of the shear strength model parameters that are specified relative to the slope, crest, and polder. It appears that specifying the undrained shear strength model parameters relative to the slope, crest, and polder might lead to a severe underestimation of the water level drawdown at the time of the levee failure when using the LEM of Uplift Van.

4.4.5 Discussion of the Bayesian method and its implications

Several specific elements of the Bayesian method could be improved. The established relationship between the reliability index and the factor of safety (the $FoS - \beta_H$ relationship), as used in this analysis, is based on 48 cases of primary flood defences coming from 27 separate locations in total. Kanning et al. (2016) computed both the reliability index (β_H) and FoS of each case based on the mean values of strength parameters and mean values of loadings. This implies that cases with different safety standards are used in this relation. However, the levee of the Leendert de Boerspolder is regional. This means that it is constructed from different materials than most primary flood defences are and is subjected to different conditions than the primary flood defences as used by Kanning et al. (2016). Even more, the established $FoS - \beta_H$ relation had included both uncertainties of load (water level) and strength, while in the Leendert de Boerspolder experiment, in which a known water level is taken into account. Moreover, the relation was derived so it could be used to identify the FoS a levee should at least have to comply with the target reliability index. However, in this analysis, the relation is used to determine the reliability index a levee should at least have to comply with the computed FoS. And hence, the annual failure probabilities in the cases from (Kanning et al., 2016) are used in this chapter as conditional failure probabilities given a certain load, which is highly questionable. The $FoS - \beta_H$ relation should be further developed for regional flood defences, or other more suitable techniques can be used to make an estimation of the probability of failure, e.g. Monte Carlo simulation or FORM.

The resisting moment of the Leendert de Boerspolder levee has decreased step by step with a stepwise increasing water level drawdown. First, it has been assumed that the water level drawdown steps are independent events. However, as there could be a weakening of the levee due to previous water level drawdown steps, the resisting moment in the various steps will not be fully independent in reality. How well the random variable of the driving moment and the resisting moment are independent of each water level drawdown step is for future studies (Klerk et al., 2018).

Despite this simplification, the method identified the water level drawdown that corresponds to the water level drawdown at the time of the failure. Many failure experiments use the principle of increasing the load or decreasing the resistance in steps as long as the structure fails. The method described in this section can help significantly in estimating the number of steps until failure will occur or estimating the probability of failure at each step of failure (van der Krogt et al., 2020). This method may also be used in experiments with other failure mechanisms such as piping (Pol et al., 2020).

Apart from an experiment, the Bayesian method could help with a hindcasting analysis of a failed levee or identify the water level at the time of failure, e.g. the Breitenhagen levee failure in Germany (Kool et al., 2019; Kool et al., 2020). The Leendert de Boerspolder experiment was set up to cancel out the uncertainty caused by the water pore development inside the levee due to the water level drawdown e.g. phreatic line development. How this uncertainty affects the outcome of hindcasting of a real levee failure is for future studies (Kanning & van der Krogt, 2016; Schwiersch, 2021).

Within the Netherlands, levees are designed to maintain a certain reliability level in a given year. This study underlines that when accurately estimating the reliability of an existing levee in any given year, the information should be taken into account that an existing levee has already survived a certain load or period (Hall, 1987; Klerk et al., 2018; Schweckendiek, 2014b; Schweckendiek, van der Krogt, Teixeira, et al., 2017).

Additional calculations have been made using the mean, instead of the median, of the measured undrained shear strength, but were not presented in this analysis. These additional resulting FoS were all very high, i.e. none of the models showed a driving load that exceeded the resisting moment regardless of the water level drawdown. In their dissertations, Muraro (2019) and de Gast (2020) states that the median of the measured shear strength approximates the generic strength of a levee best in case soil responds with undrained shear strength due to changing loadings. Finally, the median of the measured shear strength was used in this analysis.

Similarly to the point raised in the discussion considering the deterministic method, generally it is assumed that the undrained shear strength model parameters should be specified relative to the slope, crest, and polder (3 zones) to provide a more accurate characterisation of the soil's shear strength. This is because of 2 reasons. First, the properties of shear strength in these zones can change significantly due to different load histories along the slip surface. Second, the direction of the slip surface relative to the main stresses in the soil determine the shear strength at any specific point in the slip surface. The results as presented by Gast (2020) are used to derive correction ratios. These correction ratios are presented in the discussion section of the deterministic method. In the current approach, a single set of undrained shear strength model parameters is used. A set of a water level drawdown of 1.0 m, Uplift Van and the undrained shear strength model characterize the failure most accurately, when the presented findings in Figure 4.19, Figure 4.20, and Figure 4.21 are corrected accordingly. This identified water level drawdown of 1.0 m differs from the 1.6 m observed at the time of the failure. Especially Uplift Van is affected by the application of the shear strength model parameters that are specified relative to the slope, crest, and polder. These results show that the estimated performance of a levee during a reliability assessment using a LEM and SHANSEP is significantly influenced by the level of detail of the characterization of the soil properties, i.e. whether the soil properties are specified relative to the slope, crest, and polder or not. Hence, correcting for the model uncertainty using a single variable that is independent of the level of detail of characterization of the shear strength model may under or overestimate the performance of a levee during a reliability assessment (RWS, 2019). This is in contrast to what is generally assumed that the undrained shear strength model parameters should be specified relative to the slope, crest, and polder (3 zones) to provide a more accurate characterisation of the soil's shear strength. Future research might explain this observation.

4.5 Conclusions and recommendations

This chapter has demonstrated the application of both the deterministic method and the Bayesian method to the hindcasting of the induced failure of the Leendert de Boerspolder levee. It was found that the original Bayesian method (chapter 3) needed some extension for analysing the induced levee failure of the Leendert de Boerspolder experiment. Eventually, both methods have provided insights in:

- The most relevant uncertainties
- The most likely set of water level drawdown and model choices to characterize the failure.

The deterministic method has the advantage that it results in a relatively little number of slope stability models and is, therefore, less time-consuming. However, in the Bayesian method, the geometrical information of the slip surface is fully incorporated in the analysis while in the deterministic method this is not the case. In the deterministic method, the resulting critical slip surface is visually assessed for its resemblance to the observed geometry as is done in the analysis of the levee failure near Breitenhagen. Interpreting the results from the deterministic method, takes some more effort since steps of water level drawdown and model choices are not individually valued for their likelihood of occurrence nor their relevance for characterizing the failure. The actual situation or model choices are not necessarily objectively mirrored in the results using the deterministic method. The Bayesian method makes it possible to quantify how much more likely it is to fail at the identified water level drawdown instead of the water level drawdown a bit higher or lower. The same goes for alternatives in model choices. This makes it easier to interpret the results and identify the most accurate characterization. The Bayesian method appears to have more advantages over the deterministic method.

The observation information was essential for the identification of the most likely water level drawdown in both methods. When taking into account the observed geometry of the slip surface, both methods identify the same water level drawdown at the moment of failure, but with different model choices (Table 4.5). The deterministic method supports the LEM of Spencer and the undrained shear strength model, with a failure for a water level drawdown between 1.5 m and 2.0 m to characterize the failure most accurately. The Bayesian method supports the LEM of Spencer and the drained shear strength model with a water level drawdown of 1.5 m and 2.0 m to characterize the failure most accurately. The difference between both methods is best explained by the difference in how the observed slip surface information is incorporated into the hindcasting.

In this chapter, the Bayesian method has been enhanced to include the gradual increase of the hydraulic head and the survival of previous load phases. Originally, the Bayesian method was applied to the Breitenhagen levee failure. In the case of the Breitenhagen levee failure, the water level during failure was known while the scenarios of subsoil composition were uncertain. In the Leendert de Boerspolder, the subsoil composition is well known while the water level during a failure is not. Although the Bayesian method is time-consuming and further development is possibly needed, it provides a robust method to hindcast and it explains the failure based on likelihoods. During real failures, water levels will gradually rise, thus increasing the likelihood of failure over time. For future work, the method of Bayesian hindcasting should be further developed so it can handle loads that gradually change over time better. Also in this analysis, the slope instability is characterized using a LEM as the slope surface has a relatively simple geometry. Finite Element Methods (FEM) analyses can lead to a more accurate characterization of slope instability, and should especially be considered for complex geometries and soil response (see e.g. (Varkey et al., 2017)).

5 The influence of deviating conditions on levee failure rates[5]

5.1 Introduction

The reliability of a levee system is generally decomposed into levee sections and every section is assessed on different failure mechanisms. For every failure mechanism, both the load and strength terms and model uncertainties are incorporated in the evaluation of the limit states using probabilistic methods (Jongejan & Maaskant, 2015; Schweckendiek et al., 2014; Steenbergen et al., 2004; Vrijling & Gelder, 2002; Vrouwenvelder, 2006; Vrouwenvelder & Steenbergen, 1999; Zhang et al., 2011). Stochastic distributions are used to characterize loads and strength properties, as well as model uncertainties. When the load exceeds the strength, the limit state is exceeded and failure occurs – often leading to breaching and consequential damages.

However, the resulting failure probabilities from such reliability analysis could significantly deviate from actual (observed) failure rates of levees (see definition below), despite both methods targeting the same phenomenon. Sometimes there can be good reasons for the differences between both approaches, such as levee upgrades or the occurrence of multiple high water level events within a short period. However, in other cases, the difference between calculated and observed levee failure probabilities (per unit time) is more difficult to explain. As an illustration, in recent reliability analyses of levee systems in the Netherlands (using advanced methods and input data), relatively high failure probabilities were found for multiple levee systems in the order of ¼ to 1/10 per year. Results were considered less credible as no failures have been observed for many decades in the Netherlands and levees are generally designed for very low probabilities of failure. Also, Rikkert (2019) shows that regional levees in the Netherlands are designed at an annual failure probability of 1/50 to 1/500 per year, however, he observed average failure rates are at least 1/600 per year. In the context of levee reliability analysis and flood risk management, historical information on past failure and failure rates could provide important additional information. Therefore, a more “frequentist” approach is developed and implemented in this chapter, which is based on observed failure rates of levees within a system.

In this chapter, the failure rate is used to describe a first estimate of the probability of failure within an interval of time (Finkelstein, 2008). In the past, failure rates have been used to estimate the annual probability of failure of pipelines (Dawotola et al., 2011), power plants (Hutchison et al., 2009), or embankments dams in the United States (Baecher et al., 1980; DeNeale et al., 2019; Hatem, 1985; Major, 2019; Von Thun, 1985). There have only been a few studies on observed failure frequencies of levees (Foster et al., 2000; Major, 2019; Rikkert & Kok, 2019), but these did not consider the effect of the return periods of loads and local deviating conditions that may affect strength and reliability.

This study presents a method to quantitatively assess the failure rates of levees based on historical information of failure and the return period of the considered events, as well as the influence of deviating conditions on these failure rates. The following general approach is used: Failure rates

[5] This chapter has been published as: Kool, J., Kanning , W., & Jonkman, S. N. (2022, 8 December 2021). The influence of deviating conditions on levee failure rates. *Journal of Flood Risk Management*. https://doi.org/10.1111/jfr3.12784

are derived based on observations for past flood events, identifying which sections failed or survived flood events with certain return periods. Also, the expected amount of failures at a system level is considered.

The occurrence of failures is expected to be related to local load and strength characteristics. During a single flood event groups of levees in a system are often loaded by a similar load level. Particular differences in strength properties will affect whether a local failure occurs or not. Examples are 1) changes in the geometry of the cross-section of the levee, 2) bushes or trees near the levee, 3) permanent surface water directly next to the levee, 4) the presence of old geological features, such as old river meanders, or 5) animal burrows (Sharp et al., 2013). Such circumstances are referred to as “deviating conditions” in the context of this study, where it is noted that conditions 1-4 have been included in the analysis. Satellite images are used for the identification of these four types of deviating conditions. Bayesian techniques are used to determine the effects of deviating conditions on the failure rate of an individual levee section. The river system in the state of Sachsen-Anhalt, Germany is used as a case study. This system experienced severe flooding in 2002 and 2013 and relatively good information is available on the levee failures (Heyer & Stamm, 2013; I.E. Ozer et al., 2020).

The chapter is structured as follows. Section 5.2 introduces the method adopted in this section in more detail. The third section introduces general information about the Sachsen-Anhalt case study. Section 5.4 presents the results in terms of failure rates at the levee section level, the effect of deviating conditions, and the expected number of failures for a flood event. The section is finalized with a discussion (section 5.5), and conclusions and recommendations (section 5.6).

5.2 Method for the assessment of failure rates and the effects of deviating conditions

5.2.1 General approach

This section introduces the method for assessing the observed levee failure rate of a levee section, the influence of the deviating conditions on observed levee failure rates, and the number of expected failures during a single event. To illustrate the approach, a simplified example is presented. A schematic view of the analysed levee system is shown in Figure 5.1, which shows similarities to the case described in section 5.3.

A continuous levee system ensures the hinterland's safety, by separating the hinterland from the floodplain. Historical data shows that the levee system failed on two occasions when it was exposed to high water levels (h_T) with different return periods (T). This information can be used to determine the conditional failure rate for a certain load level, and the weighted annual failure rate (see details below). There are various deviations in the system that could influence the failure rate of specific sections. In this section, Bayesian inference is used to assess likelihood ratios that show the increase or decrease of the levee failure rate due to such deviating conditions. We assume that the groups of sections that are characterized by similar loading and certain deviating conditions (e.g. vegetation) are statistically homogeneous. This is substantiated by the observation that all levees have been designed according to safety standards, implying relatively homogeneous strength. An important cause of inhomogeneity would be the load, this is included by differentiating between loading levels. We consider the failure rate of levee sections that were exposed to water levels of different return periods. We also distinguish so-called deviating conditions (e.g. vegetation – see Figure 5.1) to account for potential differences in resistance. We thus assume that levees in a certain group of similar load and resistance conditions can be considered as a statistically homogenous group to derive typical failure rates.

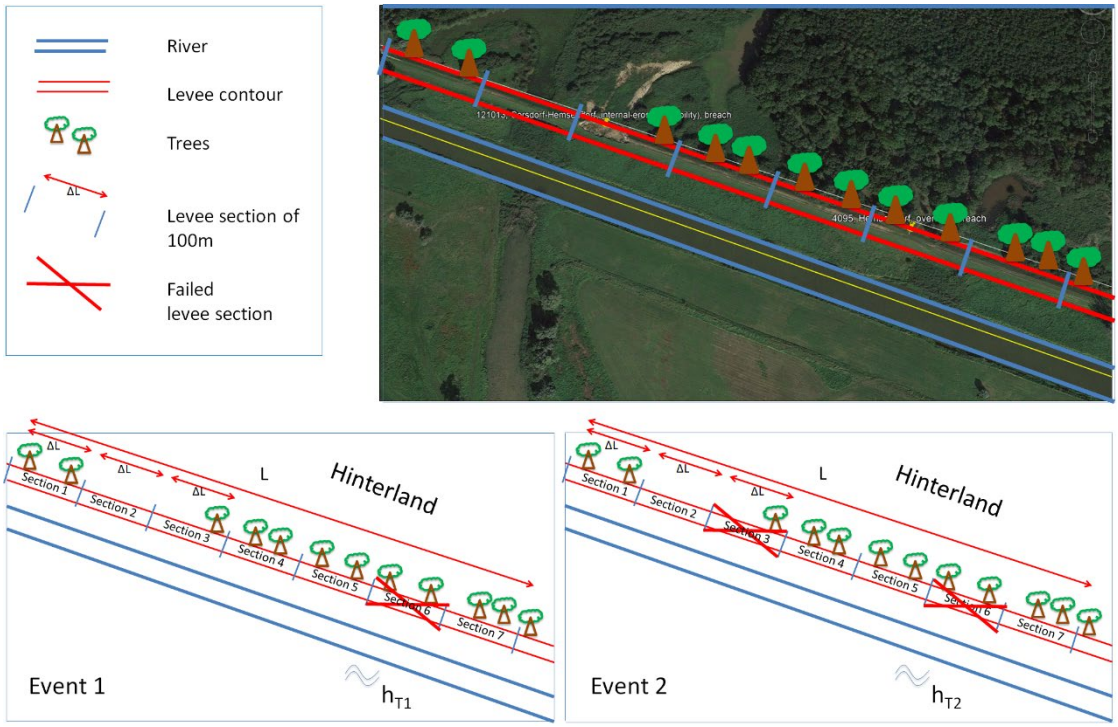


Figure 5.1 Schematic example of the performance of a part of the levee system during two high water events. The system is divided into seven sections, with trees present at most of the locations. Lower left panel: during event 1, the system is exposed to a high water event (h_{T1}) with a return period T1, which results in one breach. Lower right panel, during event 2: the system is exposed to a high water event (h_{T2}) that exceeds the return period of T2, which results in two breaches

Following this concept, a stepwise method is introduced below to calculate the failure rates and the influence of deviating conditions. The steps of the method are further elaborated in the following sections:

1. Divide the levee system into statistically independent levee sections (N_{exp}),
2. Determine the conditional failure rate $P(F|h_T)$ for all known events,
3. Estimate the annual failure rate of a section,
4. Analyse the failed levee sections for deviating conditions,
5. Analyse the survived levee sections for deviating conditions,
6. Determine the likelihood ratio (LR) of deviating conditions,
7. Calculate the annual failure rate ($P(F)$) given deviating conditions (D) and no deviating conditions ($\overline{D}$) and compare them,
8. As an optional step: Estimate the expected amount of levee section failures.

The method can roughly be divided into 3 parts: analysing failure rates based on historical events (steps 1-3), the influence of deviating conditions on observed levee failure rates (steps 4-7), and the expected amount of failures for flood events (step 8). The various steps are further described in the following sections. The following output is the results of the steps:

- An annual failure rate that shows the annual probability of failure of a random levee section without further information about e.g. local subsoil conditions (step 3)
- Likelihood ratios that show how much the annual failure probability changes due to deviating levee conditions e.g. subsoil conditions (step 6.)
- A posterior failure rate that shows the annual failure probability given deviating conditions (step 7)
- An estimate of the expected amount of levee failures in a system (step 8).

5.2.2 Step 1: Division of the levee system into independent levee sections

In the first step, the levee system is broken down into statistically independent levee sections. This implies that the system can be schematised as a series system of uncorrelated elements, under the assumption that all sections are equally strong and exposed to the same load. Hence, we are looking at how many independent levee sections (N_{exp}) there are as a function of the levee system's length (L), using Equation (5-1). And thus, we are looking for how long the sections (ΔL) should be for the sections to be independent (Vrouwenvelder & Vrijlling, 1982).

$$N = \frac{L}{\Delta L} \tag{5-1}$$

In this section, when referring to an independent section we refer to their strengths being independent since the loads are typically highly correlated between levee sections. This is illustrated in Figure 5.2, where the probability distribution and a realization of strength (resistance) and load are shown. The strength and load (and their underlying variables) usually exhibit spatial dependence which can be expressed by an (auto-)correlation function with correlation lengths θ_R and θ_S (see e.g. Vanmarcke (1977)). As the load typically has much higher correlation lengths than the strength, and we are looking for failure rates conditional to a certain water level, we are thus looking for the length of ΔL for the strengths to be independent. In other words, ΔL has to be large enough for the correlation between sections to be negligible.

We do not have detailed information on spatial variability of strength parameters available and it is, therefore, difficult to determine exact correlation lengths. It is also noted that different failure mechanisms have different typical correlation lengths, i.e. 50m for stability and 300m for piping for conditions in the Netherlands (Jonkman et al., 2017). However, Vrouwenvelder (2006) suggests that a connection between the correlation parameters and observable quantities like the length of a failure mechanism is possible. A section length of 100m corresponds roughly to the observed failure widths as well as the lengths of the deviating conditions and is therefore chosen as a value for ΔL.

It should be noted that Figure 5.2 and the text above assume that R can be modelled with a continuous probability distribution. This is not always the case as there can be discontinuous, deviating anomalies such as old gullies. The width of these deviating conditions should thus also be considered in determining ΔL.

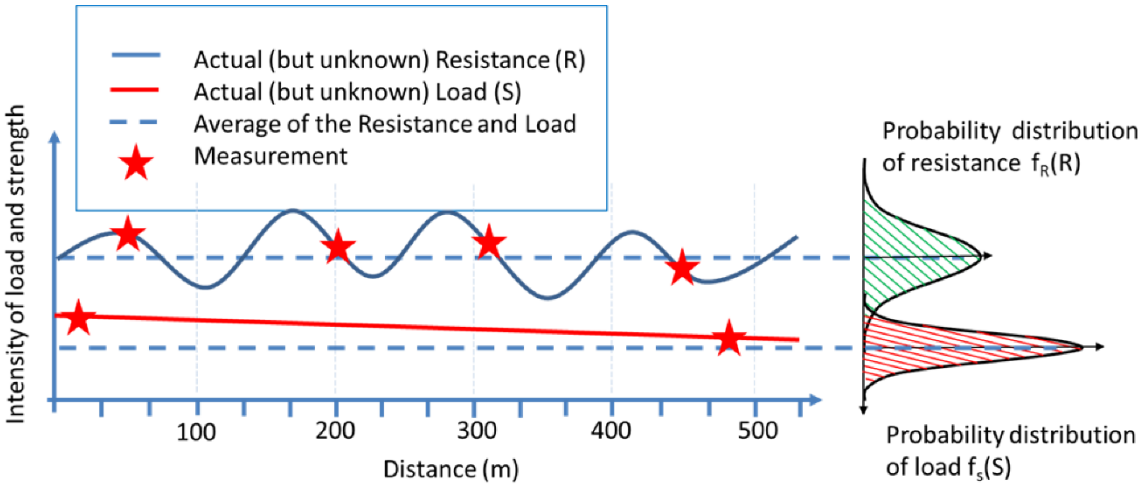


Figure 5.2 Statistical representation of the distribution of strength and load intensity. The system consists of sections 100 m in length. The strength and load are only known at the locations of the measurements but are uncertain in the other locations. The measurements can be used to derive the probability distribution of strength $f_R(r)$ and load $f_S(s)$ throughout the levee system

5.2.3 Step 2: Determination of the conditional failure rate

In this section, we determine an observed failure rate, further referred to as the annual failure rate, which empirically shows the annual probability of failure of a section or system, without assessing the underlying base variables. This is a different approach than the computation of the annual failure probability by evaluating the limit state function, which takes into account the underlying distributions of load and strength variables (see e.g. Jongejan and Maaskant (2015)). To find the annual failure rate, we have to determine the conditional failure rate for various extreme water levels (h_T) with return level T. Conditioning the failure rate by the return level T reinforces the assumption that the different sections are equal in strength and load. Hence the system is discretized in sections that were equally loaded and are statistically sufficiently homogeneous. A section's conditional failure rate is expressed as the ratio between the sections that failed (N_{fail,h_T}) and the total amount of sections that were exposed (N_{exp,h_T}), both to the same water level with a specific return period (h_T):

$$\boldsymbol{P(F|h_T) = \frac{N_{fail,h_T}}{N_{exp,h_T}}} \tag{5-2}$$

This information can be used to determine a fragility curve based on observations and the annual failure rate.

5.2.4 Step 3: Estimation of the annual failure rate using a fragility curve based on observations

If there is information on multiple loading events it is also possible to construct a fragility curve based on observations, which shows the conditional failure rate $\left(P(F|h_T)\right)$ as a function of the occurred load (water level, h_T), as shown in Figure 5.3. In this empirical fragility curve (i.e. as based on observations), the load can be expressed as a water level or return period (T). Furthermore, the probability density function (PDF) of the water level should be constructed ($f_S(h_T)$). This PDF is shown in the lower part of Figure 5.3. In the case of this section, this PDF is implicitly used, as the observed return periods are directly used to determine the probability mass contributions to the water level.

This fragility curve based on observations in combination with the PDF of the water level can be used to compute the annual failure rate by defining intervals. The number and size of intervals are dependent on the resolution of input data and the events that occurred. The red dots are representative points of the probability mass. In reality, the distribution is not that angular as shown in Figure 5.3. Therefore, the points between the boundaries represent the mass that is used in the integration. The annual failure rate is computed by integrating the points on the fragility curve $P(F|h_T)$ with the probability of occurrence of the water level $P(h_T)$ as defined by the probability mass:

$$\boldsymbol{P(F) = \sum P(F|h_T)P(h_T)} \tag{5-3}$$

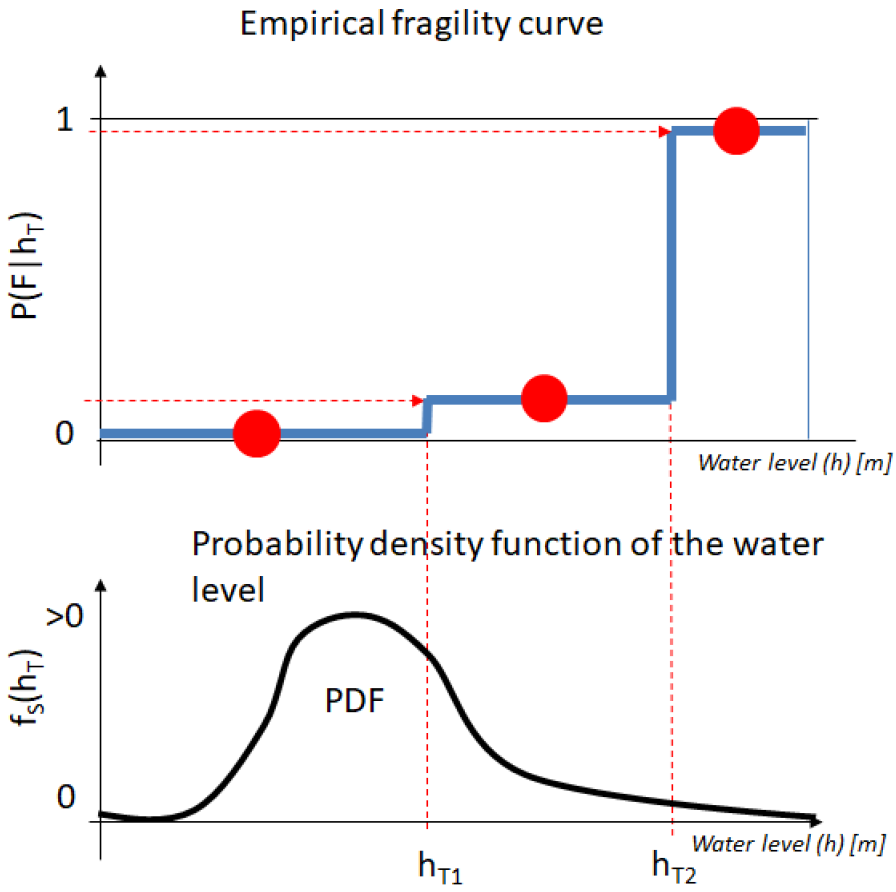


Figure 5.3 Upper: in blue, the fragility curve based on observations that relate the conditional failure rate (P(FjhT)) and water levels (h_{T1}, h_{T2}) as shown in the adopted example from Figure 5.1. The blue line is the average conditional failure rate considering all water levels in the zone: (1.) smaller than h_{T1}, (2.) between h_{T1} and h_{T2}, and (3.) larger than h_{T2}. Lower: probability density function (PDF) of the water levels. The probability of exceedance of the water level derives from the cumulative distribution function of the water levels and the return periods of the water level

The $P(h_T)$ is derived from the exceedance frequency of the events. Figure 5.3 shows an example of three intervals:

- Water levels lower than h_{T1}, with $P(F|h_T) = 0$ and $P(h_T)$ equal to all the probability mass left of h_{T1}
- Water levels between h_{T1} and h_{T2}, based on the fragility curve and the $P(h_T)$ equal to the probability mass between h_{T1} and h_{T2}.
- Water levels higher than h_{T2}. $P(h_T)$ based on the fragility curve and the $P(h_T)$ equal to all the probability mass right of h_{T2}

Hence, the method assumes that the conditional failure rate for intervals of water levels is known, as well as the return periods corresponding to these intervals.

When Equation (5-3), is applied to the example of two events in Figure 5.1, the annual failure rate $P(F)$ follows by:

$$\boldsymbol{P(F) = \mathrm{P}(\mathrm{F_0}|\mathrm{h}_{T_0} < \mathrm{h} < \mathrm{h}_{T_1})P(h < \mathrm{h}_{T_1}) + \mathrm{P}(\mathrm{F_1}|\mathrm{h}_{T_1} < h < \mathrm{h}_{T_2})P(\mathrm{h}_{T_1} < h < \mathrm{h}_{T_2}) + \mathrm{P}(F_2|\mathrm{h}_{T2} < \mathrm{h})P(\mathrm{h}_{T_2} < h)} \quad (5\text{-}4)$$

For this schematic case $\mathrm{P}\left(\mathrm{F}_0|\mathrm{h}_{T_0} < \mathrm{h} < \mathrm{h}_{T_1}\right)$ is considered negligible and therefore this term, from Equation (5-4), will not be considered separately.

As an alternative metric, a generic (annual) failure rate is provided by Major (2019) to give a first estimate of the failure rate of a section. It can be estimated by dividing the number of failures by the number of sections-years or system years of operation:

$$\lambda = \frac{N_{fail}}{t\, N_{exp}} \tag{5-5}$$

Where λ is the failure rate, N_{fail} is the number of registered failures over a time t (in years) and N_{exp} is the number of sections or systems that are evaluated in time t.

5.2.5 Steps 4-7: Determination of the likelihood ratio and update the annual failure rate

As a next step, the failed and survived levee sections are screened for the presence of deviation conditions (steps 4 and 5) and the gathered information is merged in likelihood ratios (steps 5 and 6) (Baecher & Christian, 2015). The effects of deviating conditions on the annual failure rate of an individual levee section are incorporated using two separate Bayesian techniques, the likelihood ratio and Bayesian inference.

The strength of evidence contained in the observations of deviating conditions (D) given failure (F) and survival ($\bar{F}$) can be expressed by the likelihood ratio (LR) which is determined in step 6:

$$\boldsymbol{LR = \frac{P(D|F)}{P(D|\bar{F})}} \tag{5-6}$$

A classification of the strength of the evidence is given by Jeffreys (1998) (Table 5.1). In steps 4 and 5, satellite imagery of both the considered failed and survived sections are analysed for the presence of deviating conditions. The obtained information is used to compute $P(D|F)$ and $P(D|\bar{F})$ which enable the calculation of the likelihood ratio (LR). This likelihood ratio provides a measure of the strength of the evidence.

Table 5.1 Likelihood ratios and strength of evidence according to (Jeffreys, 1998)

Likelihood Ratio (LR)	Strength of evidence
<1	Negative
1-3	Barely worth mentioning
3-10	Substantial
10-30	Strong
30-100	Very strong
>100	Decisive

Bayesian inference is subsequently used (step 7) to update the annual failure rate of a section (P(F)) into a posterior failure rate that uses information about the failure and survival of the sections with and without deviating conditions. Thereby, we can compare the annual failure rate of a section, for which the same information is used as for the likelihood ratio (Equation (5-6)) that has deviating conditions to the annual failure rate of a section that has no deviating conditions. This is done by using the following equation based on Bayesian inference:

$$\boldsymbol{P(F|D) = \frac{P(F)P(D|F)}{\sum P(F_i)P(D|F_i)}} \quad (5\text{-}7)$$

Here, P(F|D) is the annual failure rate given that a deviating condition (D) is observed, P(F) is the annual failure rate before an analysis is done on deviating conditions, see Equiation (5-4); and P(D|F) is how probable observing deviating conditions is given the failure of a section. The normalizing constant for this equation is the probability of observing (D) overall sections, regardless of (F).

5.2.6 Step 8: Expectation of the amount of failures in a levee system

Once the failure rate of individual sections is known, it is still an open question of what the expected number of failures given a certain loading event will be. The expected amount of failures depends on the correlation between the section failures and the corresponding probability distribution. We consider a levee system that consists of independent sections in terms of strength, that are all exposed to the same deterministic load (h_T), as introduced in section 5.2.2 (step 1). In such a situation it is possible to describe the probabilistic properties of this number of failures in a system using a Binomial distribution (Jonkman, 2007) to estimate the expected amount of failures and their standard deviation.

The probabilistic properties using a Binomial distribution, i.e. the probability mass function in Equation (5-8), the expected amount of failures for a specific water level in Equation (5-9) and the standard deviation in Equation (5-10), follow from:

$$P\left(N_{fail} = n|h_T\right) = \frac{N_{exp,h_T!}}{n!\left(N_{exp,h_T} - n\right)!} {F_{h_T}}^n (1 - F_{h_T})^{N_{exp,h_T} - n} \quad (5\text{-}8)$$

$$E\left(N_{fail}|h_T\right) = F_{h_T} N_{exp,h_T} \quad (5\text{-}9)$$

$$\sigma\left(N_{fail}|h_T\right) = \sqrt{N_{exp,h_T} F_{h_T}(1 - F_{h_T})} \quad (5\text{-}10)$$

Where N_{exp,h_T} is the number of sections exposed to the high water (h_T); n is the number of section failures; F_{h_T} is the conditional failure rate $\left(P(F|h_T)\right)$ of Equation (5-2).

5.3 Case study: The Sachsen-Anhalt floods in 2002 and 2013

The case study adopted in this section concerns the major river system in the state of Sachsen-Anhalt in Germany, i.e. the Elbe river system and its tributaries Saale, Mulde, Schwarze Elster, and Weisse Elster, as shown in Figure 5.4. High discharges originate from the precipitation that falls in these catchments. The studied river system covers a length of about 581 km with a levee system adjacent that currently covers a total length of 673,2 km (Weichel, 2020). The focus of this analysis will be on the two major high water events in the year 2002 and 2013, which have caused the flood defences along the river system to fail at many locations (Schwandt & Hubner, 2019) resulting in damages in the range of 11.6 billion euro in 2002 and 6 to 8 billion euros in 2013 (A.H. Thieken et al., 2016).

For 23 years (from 1995 to 2018), the major river system of interest was confronted with high water events with levee failures on several occasions, i.e. in the years 2002, 2003, 2010, and 2013 (Schwandt & Hubner, 2019). The 2003 and 2010 high water events concerned individual failures and floods were associated with lower return periods which are not taken into account in the analysis here: The levee failure in 2003 concerned a summer levee parallel to the Elbe, with no real damage

as a result. The levee failure in 2010 concerned a levee that needed rebuilding (DLR, 2010; Siebert, 2020)

The two flood events, in the years 2002 and 2013, have resulted in the failure (i.e. not necessarily breaching) of 41 levee sections. The failures are split into 3 categories, the levees that failed due to either “internal erosion or instability”, due to “overflow” and “other causes” (Table 5.2). Levees fail due to overflow when water levels exceed the height of the levee, leading to consequent erosion. Levees fail due to internal erosion or instability simply because they could not retain the water and often fail before water levels overflow the crest. This analysis only considers the total amount of levee failures (Table 5.2). It is noted that two levee failures upstream of the Schwarze Elster have not been included in the analysis (Figure 5.5). These were not directly the result of extremely high water, but because of construction activities that were not secured before the water had reached the 10-year event water level (Siebert, 2020). An overview of the locations of all failures, within the borders of Sachsen-Anhalt, is presented in Figure 5.4 (Google, 2020c) and further details are presented in I.E. Ozer et al. (2020). Both full failures (i.e. breaches) and partial failures (occurrence of the failure mechanism, without full breaching, e.g. a partial slide of the levee body) are included in this analysis.

Table 5.2 Number of levee failures (41 in total) during the high water events (the years 2002 and 2013) categorized as failures due to internal erosion or instability, overflow, and other types of failures.

year	Internal erosion or instability	Overflow	Other types of failures	Total levee failures
2002	5	20	4	29
2013	8	4	-	12

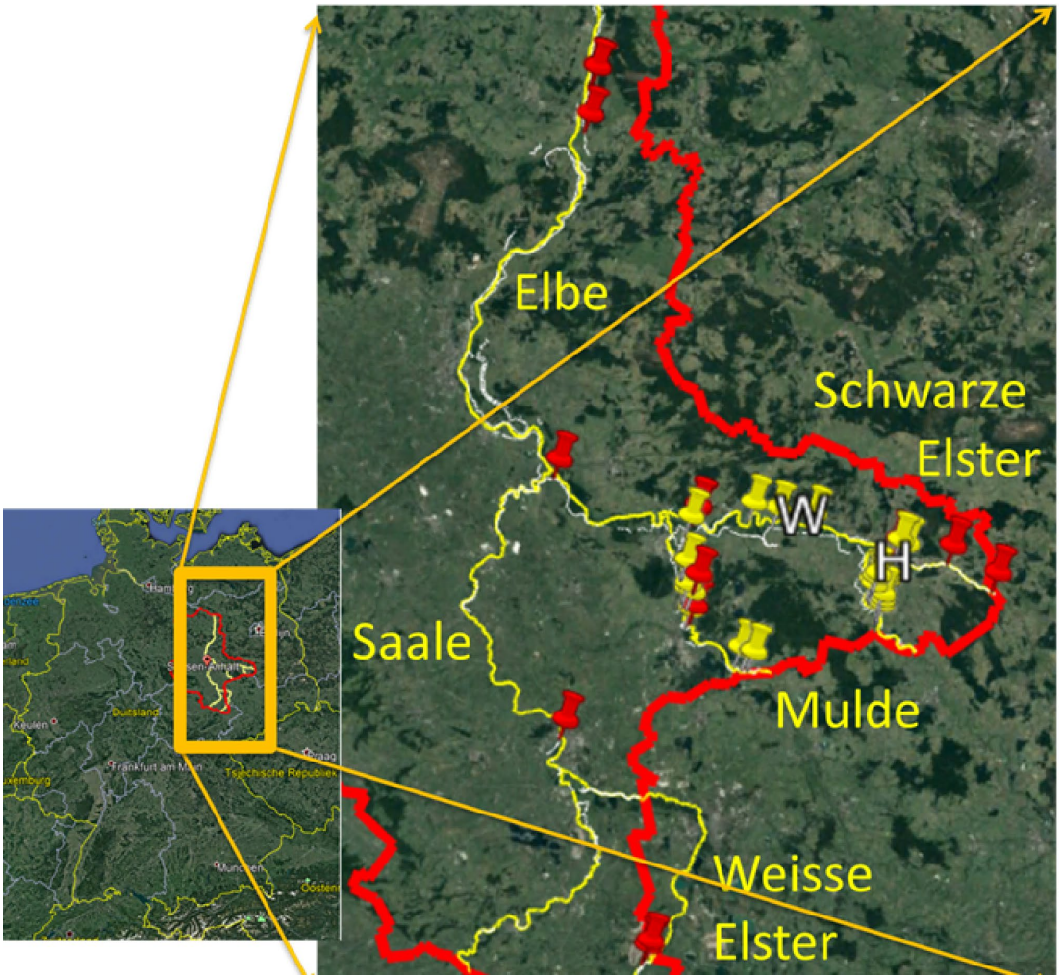


Figure 5.4 Left: river system of Sachsen-Anhalt in Germany. Right: all 43 failures projected in the river system (during the high water events of 2002 (yellow) and 2013 (red)). The white “W” marks Wittenberg and the white “H” marks Hemsendorf (Google, 2020c)

Originally, the design heights of the levees along the Elbe focused on a flood event from the early ’80s and the water levels that occurred during that event were used as absolute “values” plus freeboard (Weichel, 2021). This design height originates from the time that Sachsen-Anhalt was part of the former GDR. In the run-up to the first high water event in 2002, the local government had the intention of designing the entire levee system for a water level with a return period of 100 years plus freeboard. However, at the time of the first high water wave in 2002, approximately 5% of the levee system of Sachsen-Anhalt was rehabilitated to conform to that intention. In the aftermath of the 2002 flood event, Gocht (2004) investigated the design elevation of prominent locations in the levee system that were exposed to high water in 2002. The design height of these investigated levee sections corresponded to a 100-year high water level event and an additional 1m freeboard. Approximately an additional 45% of the levees were rehabilitated within Sachsen-Anhalt between the period 2002 and 2012, before the 2013 flood event (Weichel, 2021).

The hydraulic loads and the associated return periods for both events (2002 and 2013) are estimated, per river section (Schroter et al., 2015). The gauge data were made available by the Federal Institute for Hydrology and environmental state offices of the federal states. These give insight into the return period of flood peak discharges and associated water levels using an interval of return periods (upper and lower bound) per river section. The results are found using extreme value statistics and based on a 50-year reference period (Schroter et al., 2015), which are shown in Figure 5.5. For several locations, more detailed information was needed to connect the local hydraulic loads and the associated return periods to the levee failures, e.g. at the Schwarze Elster near Hemsendorf, and the Elbe near Wittenberg (Figure 5.5) (Gocht, 2004; IKSE, 2004; LHW, 2014; Siebert, 2020). This information has been for analysing failure rates and the influence of deviating conditions.

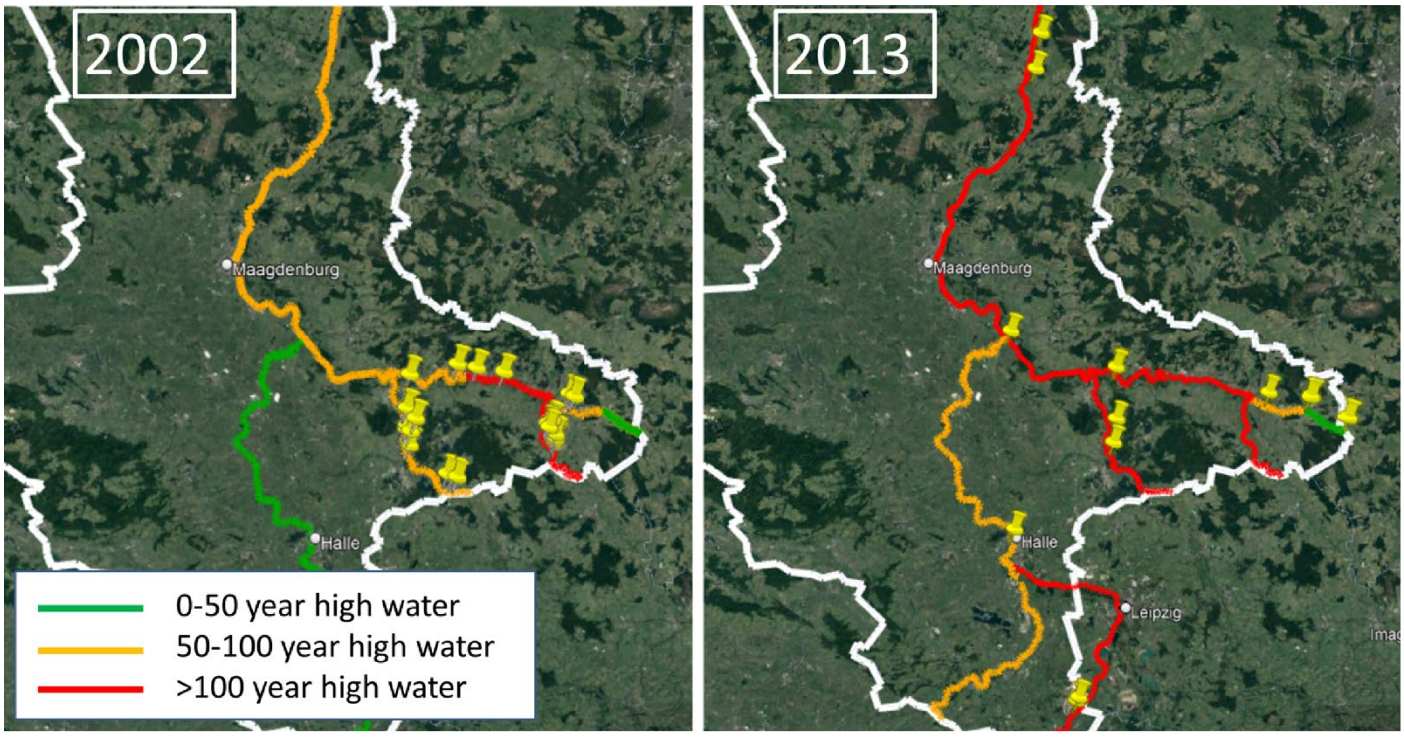


Figure 5.5 Return periods of highwater discharges and associated water levels present per event (left: the data of the 2002 flood, and right: the 2013 flood (Gocht, 2004; IKSE, 2004; LHW, 2014; Schroter et al., 2015; Siebert, 2020). The marked locations (yellow pins) show the location of the failures. In the right panel: two pins on most right are locations along the Schwarze Elster and have not been included in the analysis

5.4 Results: Failure rates and the influence of deviating conditions

Based on the method introduced in section 5.2, this section presents the results of the analyses. First, the failure rates are presented (section 5.4.1), subsequently the influence of deviating conditions (section 5.4.2) and finally the expected number of failures per event (section 5.4.3).

5.4.1 Failure rates (steps 1-3)

Step 1: Schematization of the Sachsen-Anhalt levee system.

The continuous levee system of Sachsen-Anhalt is divided into statistically independent sections to approach the levee system as a series system. The length of the section (ΔL) influences the number of sections within the system and is, therefore, an important choice in the assessment of the failure rate of a section.

Dutch procedures suggest independent equivalent lengths of 50m in the evaluation of slope instability and 300m in the evaluation of piping/internal erosion (helpdesk, 2015). For the sake of simplicity and uniformity, we adopt a section length of 100m (ΔL) for all failure mechanisms in this study. A section length of 100m corresponds roughly to the observed failure widths as well as the lengths of the deviating conditions that are the object of this study (Figure 5.6)(Google, 2020b). This results in 6732 sections within the system.

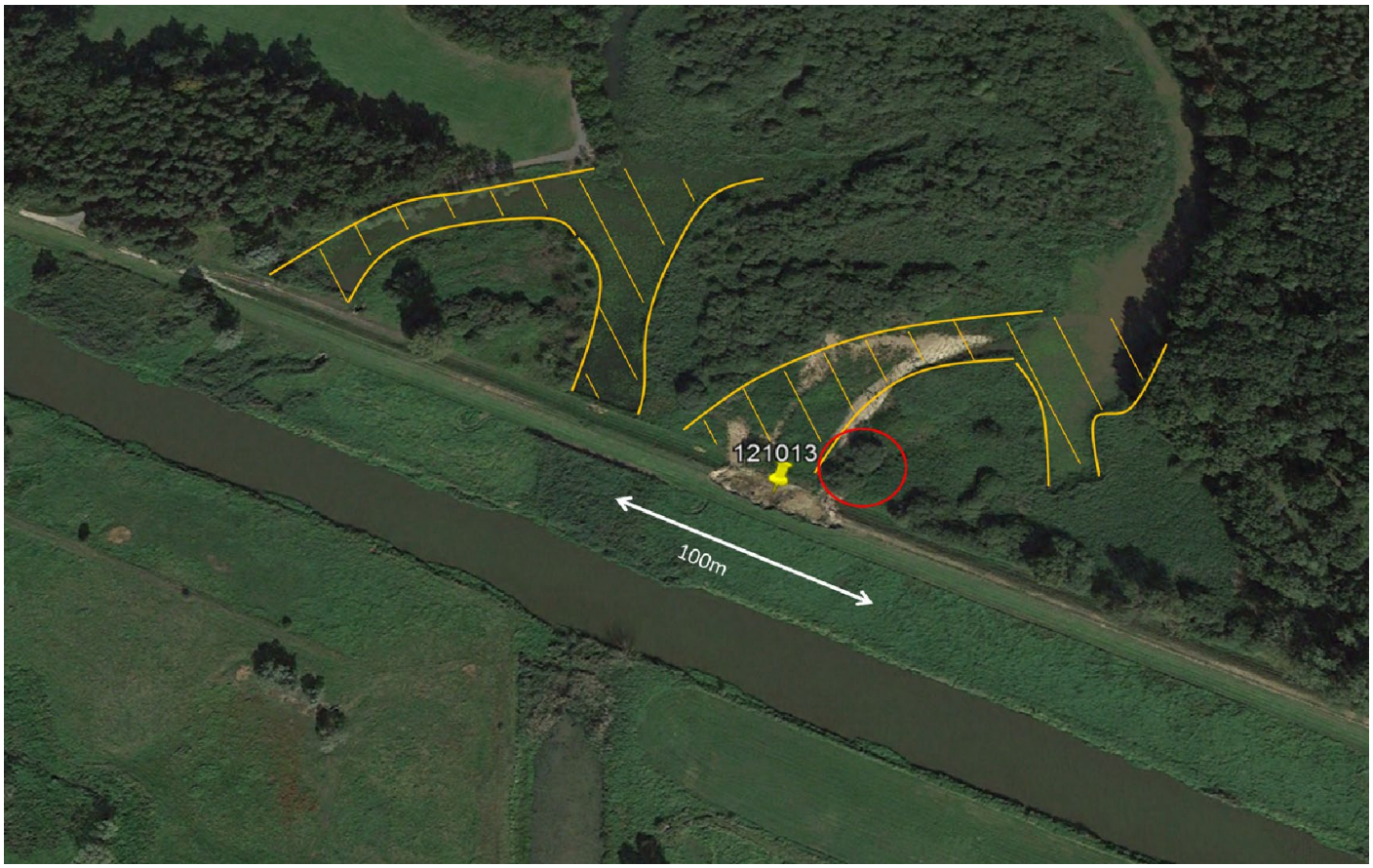


Figure 5.6 Overview of the breach, no. 121013 (ILPD) in the aftermath of the flood in the year 2013 (Google, 2020b). It occurred along the river called Schwarze Elster near Hemsendorf (located at the yellow marker), the figure shows the width of the breach (white arrow) initiated by internal erosion as well as bushes or trees next to the levee (marked with a red circle), and observable geological deviation in the subsurface (marked with orange shadings) for example, old river belts

Step 2: Derivation of conditional failure rates

Historical high-water information shows the high water levels varied between the stretches of the rivers during the flood events of 2002 and 2013. Therefore, as explained in section 5.2.4, the failure rates are assessed conditional to the return periods of the high water level events. It has been determined how many sections were exposed to a water level corresponding to a certain return period (N_{exp,h_T}) per high water event or over a certain period. This enables building a fragility curve. The exact water levels are unknown and are expressed in terms of ranges of return periods (T) in previous studies. Three categories of water levels were distinguished (see Table 5.3):

- Lower than a 50-year event ($h_{T0}<h_T<h_{T50}$),
- Between a 50-year event (h_{T50}) and 100-year event ($h_{T50}<h_T<h_{T100}$),
- Higher than a 100-year event ($h_{T100}<h_T$).

The numbers of failures have been assessed for the individual high water events (2002 and 2013) and the summed number of failures over both events (see Table 5.3). Most of the failures occurred in 2002 due to water levels exceeding the 100 years return period, but also a substantial number of failures occurred for lower water levels. In 2013 almost all sections failed due to exposure to high-water levels associated with a 100-year return period or higher ($h_{T100}<h_T$). The conditional failure rates have been calculated using Equation (5-2) and are presented in Table 5.4. These present the conditional failure rate per high water event (2002 and 2013) and the total failure rate of a levee section exposed to a water level with a certain load level. The total failure rates are found by taking the total number of sections that were exposed and failed over both the high water events (2002 and 2013) and using these as input for Equation (5-2). These are used for further analysis.

Table 5.3 The number of failed and exposed sections as a function of the return periods of high water levels that these were exposed to for the high-water events of 2002 and 2013.

return periods of high-water levels (h_T, T in years)	number of sections that failed (N_{fail,h_T}) per event and in total			number of sections that were exposed to h (N_{exp,h_T}) per event and in total		
	2002	2013	Total of both events	2002	2013	Total of both events
($h_{T0}<h_T<h_{T50}$)	0	0	0	2812	112	2924
($h_{T50}<h_T<h_{T100}$)	13	1	14	3252	2812	6063
($h_{T100}<h_T$)	16	11	27	668	3808	4476

Table 5.4 The conditional failure rates are based on the information in Table 5.3.

return periods of high-water levels (h_T, T in years)	Conditional failure rate $P(F\|h_T)$		
	2002	2013	Total over both events
($h_{T0}<h_T<h_{T50}$)	0	0	0
($h_{T50}<h_T<h_{T100}$)	$4.00 \cdot 10^{-3}$	$3.56 \cdot 10^{-4}$	$2.31 \cdot 10^{-3}$
($h_{T100}<h_T$)	$2.40 \cdot 10^{-2}$	$2.89 \cdot 10^{-3}$	$6.03 \cdot 10^{-3}$

Step 3: Construction of the fragility curve based on observations and estimation of the annual failure rate

Based on the data of Table 5.4, a fragility curve can be compiled to relate the conditional failure rates ($P(F|h_T)$) to the return period of the water level (h_T). Figure 5.7 shows the results. For both events, the conditional failure rate increases with the load level. This is as expected, as higher loads will likely lead to overloading and failure of more sections. Also, the conditional failure rates of 2002 were higher than in 2013. A likely cause could be the reinforcements that took place after the 2002 event (Gocht, 2004).

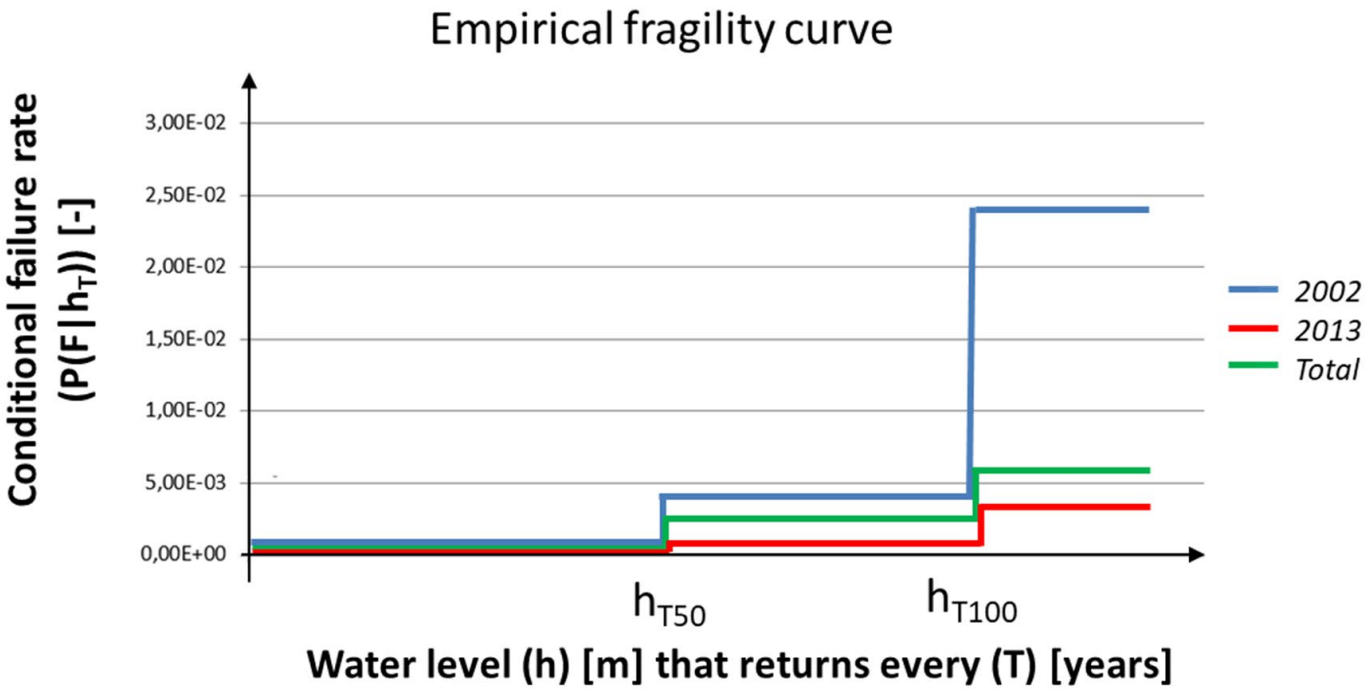


Figure 5.7 The fragility curve shows the relation between conditional failure rates of a section and the exposure of water levels while only the water levels that have a return period of 50 and 100 years are known

As a next step, the annual probability of failure for a section is computed for the combination of the two events according to Equation (5-3) and Equation (5-3):

$$\begin{gathered} \boldsymbol{P_{f,section} = P(F|h_{T0} < h_T < h_{T50})P(h_{T0} < h_T < h_{T50}) +} \\ \boldsymbol{P(F|h_{T50} < h_T < h_{T100})P(h_{T50} < h_T < h_{T100}) +} \\ \boldsymbol{P(F|h_{T100} < h_T)P(h_{T100} < h_T) =} \\ \boldsymbol{0 \cdot \left(1 - \frac{1}{50}\right) + 2.31 \cdot 10^{-3} \cdot \left(\frac{1}{50} - \frac{1}{100}\right) + 6.03 \cdot 10^{-3} \cdot \frac{1}{100}} \\ \boldsymbol{= 8.34 \cdot 10^{-5} [\frac{failure}{year}]} \end{gathered} \quad (5\text{-}11)$$

We have compared the results with the often-used generic failure rate using the method suggested by (Major, 2019). It is defined as the predicted number of times an item will fail in a specified period. Since the number of levee failures between the years 1995 and 2018 is known, we have calculated this generic failure rate as well. The earlier presented information shows that of the 6732 sections in total, 41 sections failed over a period between 1995 and 2018, during two events of overloading (in the years 2002 and 2013). The failure rate of a random section over 23 years:

$$\lambda_{section} = \frac{number\ of\ failed\ sections}{years * number\ of\ sections} = \frac{41}{23 * 6732} = 2.65 \cdot 10^{-4} [\frac{failure}{year}] \quad (5\text{-}12)$$

It is noted that this generic failure rate is higher than the annual failure rate computed in Equation (5-11). The latter takes into account that two rare events have occurred within a relatively short period. The observed probability of system failure per unit time of the levee system is found by dividing the two events with failure (in 2002 and 2013) over 23 years of operation:

$$\lambda_{system} = \frac{number\ of\ system\ failures}{years} = \frac{2}{23} = 8.70 \cdot 10^{-2}[\frac{failure}{year}] \quad (5\text{-}13)$$

Considering these failure rates (Equation (5-11), Equation (5-12), and Equation (5-13)), the failure rate of the system is about 300 to 1000 times higher than on the section level.

5.4.2 The influence of deviating conditions on the annual failure rate (steps 4-7)

Based on satellite imagery processing we have analysed the statistical influence of a deviating condition on the annual failure rate of a levee section. Satellite imagery of the river system enables analysis of both the failed levee sections and survived sections for the presence of deviating conditions over a section of 100m in length (steps 4 and 5) using imagery from (Google, 2020a). The obtained information enables the calculation of the likelihood ratio (LR) of Jeffreys (1998) to classify the influence of the deviating condition (step 6). Bayes' rule is used to determine the annual failure rate (posterior) of a section given that there is a certain deviation present in the section and these results are used to determine the influence of a deviating condition (step 7, Equation (5-7)).

The following types of deviating conditions have been considered, as illustrated in Figure 5.8:

1. Bushes or trees next to the levee,
2. Permanent water directly next to the levee,
3. Geometric changes of the cross-section of the levee,
4. Geological deviations in the subsurface are observable from images.

Figure 5.8 Illustration of deviating conditions (samples were taken from the Sachsen-Anhalt river system (Google, 2020a), 1: bushes or trees next to the levee, 2: permanent water directly next to the levee, 3: geometric changes of the cross-section of the levee (the road in orange on the levee in red branches off with several turns toward the river), 4: geological deviation observable in the subsurface, for example, old river belts

In total, all 41 failed sections and a sample of 164 survived sections, of the total of 6732 sections, have been analysed. To choose the location of the samples of surviving sections, markers were chosen-distributed evenly with a constant distance between them and manually over the entire system in the centre of the river (yellow line in Figure 5.9). This ensures a certain degree of randomness in the collected samples. Subsequently, the samples were placed alternately and perpendicular to the river in a north-easterly direction (label-E) or in a south-western direction (label-

W) of the initial sample, as illustrated in Figure 5.9 (with the red arrow). There are two options where the mark can be placed:

- Levees parallel to the rivers as marked by the chart of Sachsen-Anhalt (LVermGeo, 2019), which resulted in 164 samples
- Other sections, where no levee parallel to the river were present, for example, because the height of the landscape ensures that no flood defences are needed to protect the hinterland. These marks were excluded for further analysis.

The resulting likelihood ratios and conditional failure rates are summarized in Table 5.5.

Figure 5.9 Left: satellite imagery showing the markers that were chosen-distributed evenly with a constant distance between them and manually over the whole river system. Lower right panel: close up Satelite imagery (dated from 1-6-2001) of samples 37_E, 38_W, 39_E, and 40_W (E: north-easterly side, W: southwestern side), the red arrows demonstrate how the samples are placed, perpendicular from the centre of the river alternately in north-easterly (label-E) or south-western direction (label-W), top right panel: close up of sample 37_E: there is no presence of bushes or trees and geometric changes. Permanent water is present next to the slope and there are observable geological deviations (Google, 2020a)

Table 5.5 The likelihood, likelihood ratios and conditional failure rate for the various deviating conditions.

	No deviation (D_{non})	Bushes or trees (D_{bt})	Permanent water (<25m) (D_{pw})	Geometric changes in cross-section (D_{gc})	Geological deviations (D_{gd})
Likelihood in case of failure $P(D_i\|F_i)$	0.05	0.84	0.28	0.33	0.70
Likelihood in case of survival $P(D_i\|\overline{F_i})$	0.20	0.64	0.23	0.32	0.21
Likelihood ratio (LR)	0.3	1.3	1.2	1.0	3.3
Annual failure rate given a deviating condition $P(F_i\|D_i)$	$1.91 \cdot 10^{-5}$	$1.09 \cdot 10^{-4}$	$1.03 \cdot 10^{-4}$	$8.51 \cdot 10^{-5}$	$2.79 \cdot 10^{-4}$

The results in Table 5.5 show that only the geological deviations have a significant influence and a substantial amount of evidence of influencing the levee failure rate with a value of the likelihood ratio of LR=3.3. The analysis of satellite images showed that geological deviations were present for 70% of the failures, but only for 21% of the survived sections. This is in line with previous research (Buijs et al., 2013; Kool et al., 2019; Kool et al., 2020; Zwanenburg et al., 2018) that pointed out that local geological deviation in the subsoil could be an important trigger of levee failure. A section has an annual failure rate of $P(F_i|D_{gd})=2.79 \cdot 10^{-4}$ in case of the presence of geological deviations, which is about 14 times higher than a section that has no deviation (Table 5.5). Other deviations have a very limited influence on the annual failure rate. For example, bushes or trees are located at 84% of the failed sections, but also for 64% of the sections that survived the high loading.

The previously calculated prior annual failure rate is $8.34 \cdot 10^{-5}$. Here it appeared that sections with no deviation have an annual failure rate of $P(F_i|D_{non})=1.91 \cdot 10^{-5}$ (the posterior), which is about 4 times smaller. Hence, the information that there are no deviations observable influences the annual failure rate.

5.4.3 The expected amount of levee failures (step 8)

Once the conditional failure rates of individual sections are known, the expected number of failures given a certain loading event can be determined following the method introduced in section 5.2. The binomial distribution is used, using the conditional failure rate ($F_{h_T} = P(F|h_T)$) given the high-water levels (h_T) and the number of sections that were exposed (N_{exp,h_T}), as inputs. These specific inputs were obtained for the combined 2002 and 2013 events from Table 5.2 and Table 5.3 respectively, distinguishing the amount of failures and exposed sections that were exposed to water levels in the 50-100 year return period range, and those with a 100 year return period or higher. Table 5.4 and Figure 5.7 show the results of the analysis of the total dataset of the 2002 and 2013 events. As expected, the number of observed failures is reproduced, and – as an additional parameter – the standard deviation of the Binomial distribution is determined (Figure 5.10 and Table 5.6).

To show the potential application of the method to estimate the number of failures for other and future flood events, two hypothetical floods are assumed. One moderate flood during which all levee sections are exposed to flood levels with return periods between 50 and 100 years, and one more extreme flood with water levels that exceed the 100 year return period. It is assumed that the future levees in the system have the same properties as those in the past, so the earlier derived failure rates and assumptions are adopted. Failure rates have been derived from Table 5.4. This leads to the expected number of failures and the standard deviation as shown in the last two columns

of Table 5.6, clearly indicating that the expected number of failures increases with increasing loads. This illustrates how the failure rates can be used in flood risk assessments at a system level.

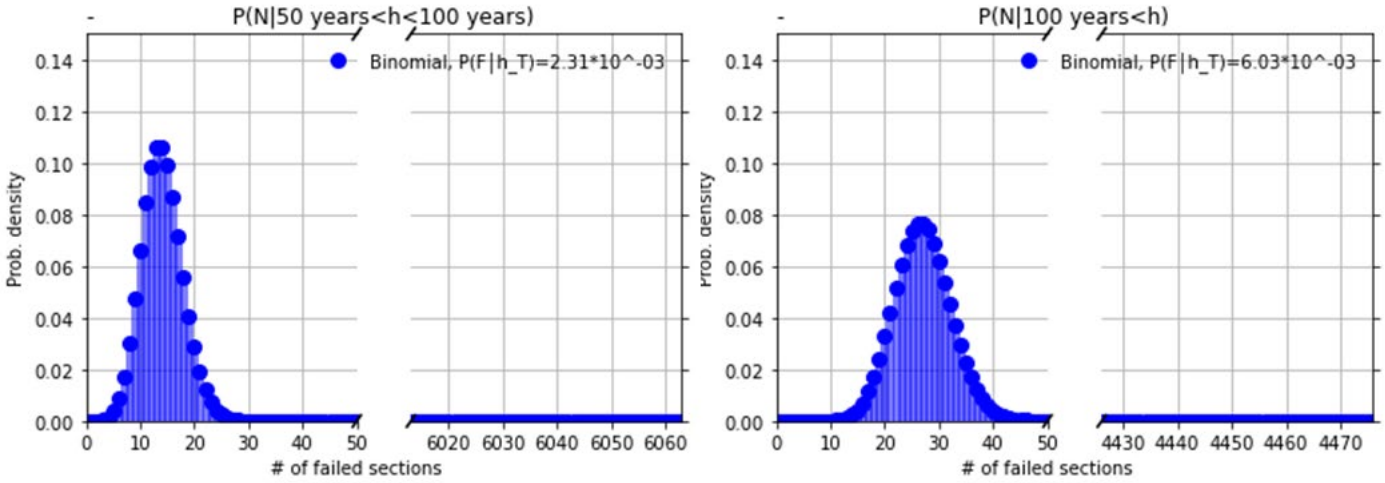


Figure 5.10 Probability mass distribution of the number of failures: Left: for the 6063 sections (see Table 5.3) exposed to a water level with return periods between once in 50 and 100 years (hT50 < hT < hT100). Right: for the 4476 sections exposed to water levels with return periods 100 years or higher (hT100 < hT)

Table 5.6 Observed and expected number of failed sections for various cases.

	2002 and 2013 flood events (combined)		Hypothetical floods	
	Return periods of high water levels ($h_{T50}<h_T< h_{T100}$)	Return periods of high water levels above 100 years ($h_T > h_{T100}$)	All sections exposed to ($h_{T50}<h_T< h_{T100}$)	All sections exposed to ($h_T > h_{T100}$)
The observed number of failures (Table 5.3)	14	27	-	-
Binomial distribution				
Exposed sections: N_{exp,h_T}	6063	4476	6732	6732
Failure rate: $P(F\|h_T)$	$2.31 \cdot 10^{-3}$	$6.03 \cdot 10^{-3}$	$2.31 \cdot 10^{-3}$	$6.03 \cdot 10^{-3}$
Expected number of failures $E(N_{fail}\|h_T)$	14	27	16	41
Standard deviation of failures $\sigma(N_{fail}\|h_T)$	3.7	5.2	3,9	6.4

5.5 Discussion

The presented analysis for the levee system of Sachsen-Anhalt focuses on the high-water events of 2002 and 2013 and the data collected for these events. The presented results for these events can be obtained under the assumption of homogeneity and stationarity. However, within the considered

period of analysis, there could be changes to the levee system over time, such as degradation processes or reinforcements of levee sections or –over a longer period- possible changes in trends in discharges and the associated frequencies (Faulkner et al., 2018; Hall et al., 2014; Merz et al., 2018). These changing conditions make it less likely that the results based on past observations can be used to characterize the future performance of the levee system directly. However, there are many levee systems with limited data availability, e.g. lack of soil investigations. For such situations, an estimate of the failure rate of a levee section can still be made using simplified methods based on observed historical failure rates, as proposed by Foster et al. (2000), Major (2019) and Rikkert and Kok (2019). However, these latter methods do not take into account specific characterizations of strength and return periods of water levels, which are factors that have been included in the present section. Information on observed failure rates could also be used as a calibration for reliability estimates from more detailed studies. The failure rates could also be coupled with (simplified) risk assessments and optimization, e.g. as proposed by Hui et al. (2016). In this approach, the levee fragility is parameterized with levee height, crown width and condition, based on synthetic levee performance curves.

Several improvements in specific elements of the proposed method could also be made. The calculated conditional failure rate (step 2) is based on the information on frequencies of high water levels for the 2002 and 2013 events, expressed using intervals of return periods. The current section utilizes a relatively coarse discretization in three intervals of load return periods, as this is the data that is available in terms of hydraulic loads (Schroter et al., 2015). In future work, we suggest differentiating more load classes based on more detailed hydraulic loading data for future cases. Also, it would be good to validate the assumption of statistical homogeneity if there is more detailed subsoil data available to check the homogeneity of the resistance properties or further divide the system into more subgroups. It is noted that particular geotechnical failures are unlikely to occur exactly at the moment that the highest water level is reached (Hui et al., 2016), and variables such as load duration could be included as an additional parameter. Furthermore, a section of a length of 100m has been used. It would be useful to assess whether this 100m could be validated or differentiated to failure mode and used in other systems, e.g. using a sensitivity analysis under various section lengths.

Moreover, other high-water events occurred in the system, such as in the years 2006, 2010 and 2011, although they did not lead to substantial failures. Including these data will provide a more accurate estimation of a section's failure rate. Also, the analysis focused on the sections that failed in general but focussed less on the individual failure mechanisms. Although this generalization supports the demonstration of the method, it ignores the specific characteristics related to different failure mechanisms and spatial uncertainties, e.g. the load duration, soil properties and levee geometry. In future work, more specific information related to failure mechanisms could be included in the analysis, and it could even be considered to vary the length of the sections depending on which failure mechanisms would be assessed.

The occurrence of deviations was assessed in this section through a visual analysis of a limited amount of manually selected samples of satellite imagery. This was a labour-intensive process. In future work, more advanced methods and other data sources could be utilized to include more samples and more information in the assessment (Aanstoos et al., 2012; I.E. Ozer et al., 2020). This will increase the robustness of the results and may also help to identify additional factors that could influence the reliability of a levee section. Further coupling with soil datasets is also a promising direction to get better insights into local weaknesses.

Vegetation (bushes or trees) was found to be a frequently present deviating condition: for the case study 84% of the sections that failed and for 64% of the sections that survived the high water events. This implies that bushes or trees are frequently present, but does not have a major influence on the failures. Geological deviations were observed for 70% of the sections that failed and in 21% of the sections that survived, therefore they are more likely to be a marker of failures. This section does not elaborate on how the (geological) deviations physically affect the strength and reliability of a levee section. However, previous work has shown that such deviations (e.g. old river meanders and old levee failures) could have a major influence on levee safety (Buijs et al., 2013; Kool et al.,

2019; Kool et al., 2020; Zwanenburg et al., 2018). Further soil investigation and modelling of the effects of such conditions on the factor of safety and failure probability of typical levees in the river system is recommended.

Failure scenarios with multiple breaches get limited attention in risk assessments, although real events show that multiple breaches can be expected when extreme high water levels occur (Jonkman et al., 2008). Even more, Apel et al. (2004) suggest that the occurrence of upstream levee failures can significantly reduce the failure probabilities of downstream levees. Note, some approaches for deriving statistics of hydraulic loads already include upstream events and interactions, such as dam and reservoir operation, e.g. Ciullo et al. (2019). In estimating the expected number of levee failures (step 8 in the framework) independence between sections has been adopted, assuming that uncertainties in strength are greater than those in loads. In the future, further formal verification, also of the role of spatial variability in hydraulic loads including interactions, and analysis of correlation would be useful.

5.6 Conclusions and recommendations

This study introduces a method for assessing the annual failure rate of levees based on information from historical floods, while also considering the return period of past events. Also, an approach has been developed to quantify the influence of deviating conditions on these failure rates. Satellite images have been processed to assess the presence of deviating conditions for failed and survived sections. Bayesian techniques are used to update the failure rate as a function of the presence of deviations.

The river system of Sachsen-Anhalt, Germany (2002 and 2013) was used as a case study. It experienced severe floods with many levee failures in the years 2002 and 2013. In assessing the failure rate an fragility curve based on observations was constructed and it was used to calculate an annual failure rate, which is $8.34 \cdot 10^{-5}$ per year per section for the case study. This differs from the base failure frequency of $1.30 \cdot 10^{-1}$ per year per section that would be found according to the method of Major (2019) which does not take into account the likelihood of the experienced flood events. Both analyses are based on historical data and satellite imagery processing of sections that failed or survived high water events. But the failure rate determined in the proposed approach in this section takes into account that two rare flood events have occurred within a relatively short period.

The results show that the occurrence of a visually identifiable geological deviation in the subsurface has a significant influence on the calculated failure rate of a levee section. The updated failure rate of a section is about 3.3 times higher than the initially calculated failure rate and about 14 times higher than when there is no visually identifiable deviation. The fact that no deviations are observed results in a likelihood ratio of 0.3; hence a much lower annual failure probability. The presence of other deviations, such as bushes or trees, or permanent water near the levee has a more limited influence on the failure rate. Nevertheless, results show that levee sections that have either bushes or trees, or permanent water near the levee have a somewhat higher failure rate (20-30% higher) than the calculated annual failure rate.

Also, the results for the Sachsen-Anhalt river system, show that rare events of high water levels are expected to result in multiple levee failures. It has been shown how the conditional failure rates and information on the return periods of loading conditions can be used to estimate the distribution of the number of failures. The inclusion of more detailed data on loads and resistance properties could lead to more detailed assessments of vulnerable groups of levees and corresponding failure rates. Also, after the failure of a dike section, reinforcements and repairs of dike sections have been performed over time implying that the associated frequencies change. For this reason, it is recommended to distinguish between reinforced and non-reinforced dike sections using the likelihood ratios in future work.

Based on this study, we recommend that the strength of levee sections with either geological deviations that are visually observable on the surface, bushes or trees, or permanent water near the levee are investigated more closely. We recommend using satellite imagery processing to analyse

the subsurface near the levee for geological deviations, such as old river bends. Based on this, further soil data acquisition could be optimized and targeted to inform the assessment of strength and reliability. This could contribute to a more targeted assessment and more robust flood defences. For future studies, we also recommend assessing other deviating conditions as a possible markers for failure and include data from other levee systems to identify general features of interest. It is also recommended to investigate how observed failure rates can be used to calibrate and improve reliability assessments. It is also expected that the presented methodology could play a role in risk and reliability assessments in data-poor environments. In such circumstances, there are often limited insights into soil data and levee profiles, but some information on historical failures is generally available, thus allowing quantification of failure rates.

Also, we recommend that future risk assessments consider failure scenarios with multiple breaches in case of high water events. Further formal verification and analysis of the correlation between failures would be useful. Overall, it is expected that the approaches that are introduced in this section can complement and improve levee management and future flood risk assessments.

Acknowledgements

We thank both Dr.-Ing. Thilo Weichel and Ms. Antje Siebert of the Landbetrieb fur Hochwasserschutz und Wasserwirtschaft Sachsen-Anhalt for providing us with relevant data. This research was performed as part of the NWO TTW project SAFElevee (project number 13861).

6 Conclusions and recommendations

6.1 Conclusions

This research aims to advance the methods for the hindcasting of levee failures, at the individual and system level. The main research question and the four supporting sub-research questions are answered below.

The main research question is; How can methods for the hindcasting of levee failures be improved?

This research proposes and demonstrates methods for systematic analysis of levee failures at the individual section and system level. The methods for hindcasting a levee failure for an individual section provide better insight into the causes and dominant uncertainties for failures. The methods developed at the system level provide more insight into the failure rates of levee sections and their dependence on loading levels and the presence of local deviations.

For the individual levee sections, approaches for forensic analysis from structural engineering have been extended to be applicable to levee failures (Borsje et al., 2014). Through deterministic and probabilistic methods of hindcasting developed in this thesis, a systematic analysis of the causes of individual levee failures can be made. Uncertainties are included in the analysis by defining possible loading and subsoil conditions related to hydraulic loads, pore pressures and stratigraphy which are referred to as scenarios explaining the failure. For example, a slope stability model can be built to characterize the failure when such a scenario explaining the failure is implemented with alternatives of model choices i.e. alternatives of LEM, soil reaction behaviour and soil parameters. This slope stability model can be used to evaluate which scenario explains slope instability and which parameters are dominant. Also, observational information can be included, i.e. the geometry of the slip surface. The deterministic method uses a sensitivity analysis, while the probabilistic method quantifies the likelihood of occurrence of each scenario and model choice, and enables updating of the likelihoods using the observational information.

The methods developed for hindcasting levee failures at the system level provide more insight into the annual failure rates of levee sections. Moreover, these methods enable quantification of the influence of local deviating conditions, loading levels and their return periods on the annual failure rates of levees.

Both developed hindcasting methods are expected to contribute to the overall quality, repeatability, transferability, transparency and recognisability of the analysis of levee failures. Therefore in the future, applying the methods can contribute to better insight into the performance and reliability of levees.

RQ 1. How can the most likely cause of levee failures be determined?

Approaches for forensic analysis have been developed for structural failures, but not yet for levee failures. Since the evidence after a failure is often very sparse, it is difficult to build a model that characterizes the failure accurately as part of the hindcasting process.

Chapter 2 offers a newly developed method to systematically analyse a levee failure based on the three steps of forensic analysis developed in structural engineering (Borsje et al., 2014). Utilizing the collected information prior to, during and after the levee failure, a computational model is developed to evaluate the failure event. A deterministic approach is followed in this chapter. The uncertainties in possible causes and computational models are included by defining possible scenarios explaining the failure. The influence of the identified scenarios and possible alternatives in model choices are analysed through a sensitivity analysis. Results of the computations are confirmed or refuted by observation information of the failure. However, due to a large number of combinations of scenarios explaining the failure and model choices to evaluate, this method is relatively time-consuming. Also, it is impossible to say something about the likelihood of different scenarios, leaving multiple scenarios possible as a failure cause.

To illustrate the method, it is applied to the levee failure near Breitenhagen (2013), in Germany in Chapter 2 of this dissertation. The levee near Breitenhagen is located at the intersection of the Saale and the Elbe and it failed due to instability of the slope at the polder side of the levee. Unexpected saturation of the levee, steep slope of the levee, and the influence of the tree roots were identified to cause of the levee failure by the forensic engineering report by Grubert (2013a). However, in the study in this section, an old breach was found to be there (the first proxy was a pond next to the levee and later confirmed with archive research). This led to scenarios with low strength and high water pressures, which was identified to be the most likely scenario explaining the failure. For this particular case, the failure observation concerns the observed geometry of the slip surface. The results indicate that locally low values of strength associated with low values of pre-overburden pressure or cohesion explain the failure. Other scenarios that were evaluated resulted in a situation that was not likely to fail or, resulted in a slip surface that differs from the observed failure surface. Both the pond and the locally weaker soil are likely to facilitate the high hydraulic head in the aquifer and unexpected saturation of the levee body that has contributed significantly to bringing the levee to instability (Grubert, 2013a). Historical data cannot confirm or exclude the contribution of the present tree roots to unexpected high saturation which reduced the resistance of the levee.

RQ 2. How can the most likely scenario explaining the failure, and the most representative model choices be quantitatively substantiated using Bayesian techniques?

The previously developed deterministic method does not quantify uncertainties explicitly. That makes it difficult to uniquely identify the most likely scenario to explain the failure.

Therefore the deterministic procedure of hindcasting of a levee failure as proposed in Chapter 2 is advanced by including Bayesian techniques in Chapter 3. Thereby a better insight is provided into the relative likelihoods of the various scenarios explaining the failure and possible root causes of failure. Failure observations and a-priori levee information are systematically taken into account to quantitatively identify the most likely scenario explaining the failure and the most representative model choices to most accurately characterise the failure. The Bayesian techniques are also used for updating the failure probabilities on the observations of the actual failure (if present) such as the shape of the slip surface. The suggested steps of analysis provide a thorough, workable and transparent method of analysis.

To illustrate the method, it is also applied to the levee failure near Breitenhagen (2013) in Germany in Chapter 3 of this dissertation. The levee failure near Breitenhagen is most likely triggered due to there being an old breach (Sixdorf, 2016), most likely resulting in an aquifer connection and locally weak soils. The LEM using Spencers' approach and undrained shear strength soil response are identified to be the most representative model choices. Within this combination of a scenario explaining the failure and model choices, the shear strength ratio is identified as the most dominant contributor to the failure. The contribution of the high hydraulic head due to the pond connection is identified to be the second dominant factor. The old breach would explain, the pond, locally weak soil, and the specific location of the breach as stated earlier in this section. Compared to the

"deterministic method" introduced in chapter 2, the probabilistic method adds the possibility to quantitatively substantiate the identification of the most likely scenario explaining the failure and the model choices most accurately characterizing the failure, but also the relative likelihood of the conditions at the time of failure.

RQ 3. How do the newly developed methods of hindcasting apply to other cases (i.e. the Leendert de Boerspolder failure experiment, the Netherlands)?

To validate and test new methods using other cases (then the Breitenhagen case) the methods are applied to the Leendert de Boerspolder levee in the Netherlands (Chapter 4). In this experiment, the hydraulic head gradient over an instrumented levee was gradually increased, to observe and measure slope instability (de Gast, 2020). These load steps were introduced as scenarios that possibly explain the failure in this analysis

The probabilistic method using Bayesian techniques required some adaptation. This was because of the design of the experiment, where the levee was brought to failure by a stepwise increase of the hydraulic head gradient, by drawing down the water level in an excavated ditch at the polder side of the levee. This lowered the resisting moment until the driving moment finally exceeded the resisting moment. The Bayesian method has been extended to include the gradual increase of the hydraulic head gradient and the survival probability of previous load phases. This makes the method applicable for the analysis of failure cases where the actual water level at the time of the failure is unknown.

Both the deterministic and the probabilistic methods using Bayesian techniques result in the identification of the most likely drawdown level as observed at the time of the failure. The Bayesian method provides more insights into how the geometry of the slip surface influences the computed likelihood of occurrence of the water level drawdown more objectively. This is because the geometrical information of the slip surface is fully incorporated in the Bayesian method, but not in the deterministic method. However, the deterministic method is less time-consuming.

The deterministic method indicates the LEM using Spencer approach and an undrained shear strength model, with a failure for a water level drawdown between 1.5 m and 2.0 m to characterize the failure most accurately. The Bayesian method suggests the same LEM with a drained shear strength model with a water level drawdown of 1.5 m and 2.0 m to characterize the failure most accurately. Both the deterministic and probabilistic methods identified the most likely water level drawdown of 1.6 m by approximation at the time of the failure. Both methods identify different model choices as most representative.

RQ 4. How can the failure rate of levees within a levee system be quantitatively assessed and what is the influence of deviating conditions?

The failure rate of levees within a levee system and the influence of deviating conditions are quantified, using Bayesian techniques, satellite images and historical data, while also considering the return period of past events. This is demonstrated in Chapter 5. Data from historical floods is used to estimate failure rates and a prior probability of occurrence of specific failure mechanisms, i.e. macro-stability and internal erosion, and overflow. Second, satellite images have been processed to analyse both the failed and survived levee sections in the presence of deviating conditions. These results are used to quantify the relations between observed deviations and the occurrence of failure. Third, Bayesian techniques are used to update the earlier computed average failure rate (as done in the first step) on the information on the occurrence of these specific deviations.

The developed method of quantitatively assessing the failure rate of a levee and the influence of deviating conditions is applied to the river system of Sachsen-Anhalt, Germany. This river system experienced many levee failures due to the high water events in 2002 and 2013. In assessing the failure rate a fragility curve based on observations was constructed and it was used to calculate an average annual failure rate. The average annual failure rate is $8.34 \cdot 10^{-5}$ per year per section for the considered system considering all failure mechanisms.

The influence on the failure rate of 1) changes in the geometry of the cross-section of the levee, 2) bushes or trees near the levee, 3) permanent surface water directly next to the levee, and 4) the presence of old geological features, such as old river meander, is assessed. The results show that the updated failure rate of a section is about 3.3 times higher in the case of the occurrence of a visually identifiable geological deviation in the subsurface. This updated failure rate is about 14 times higher than when there is no visually identifiable deviation. The presence of other deviations, such as bushes or trees, or permanent water near the levee results in a somewhat higher failure rate (20–30% higher) than the calculated average annual failure rate. The results show that rare events of high water levels are expected to result in multiple levee failures in the river system of Sachsen-Anhalt. As it is also discussed how the expected number of failures in a system during a flood event with a certain magnitude can be estimated.

The developed method can be used to complement more advanced reliability analyses. The presented methodology could play a role in risk and reliability assessments in data-poor environments. In such circumstances, there is often limited insight into soil data and levee profiles, but some information on historical failures is generally available, thus allowing quantification of failure rates.

6.2 Recommendations

6.2.1 Recommendations for scientific research

In this research, several advancements are presented in the hindcasting of levee failure on both individual and system levels, which are applied for demonstration purposes as well. In this section, various suggestions are given for further research on how to hindcasting levee failures.

Deterministic and probabilistic hindcasting of individual levees

The newly developed method of hindcasting, which is presented in Chapter 2, is developed for purpose of forensic analysis of levee failures in general. This research has been applied to a slope instability of a levee. To test its robustness and possibly advance the method, it is advised to apply the approach to another type of levee failure mechanism such as piping.

Furthermore, there are several refinements possible for the Breitenhagen analysis. The uncertainty regarding shear strength parameters, and the resulting high variations in input, may be reduced by performing local field tests (e.g. CPTs) or collecting samples for lab tests. But also consultation with local authorities and companies would lead to better assumptions for inputs, e.g. cohesion, POP, and values of the potential head in the aquifer and phreatic line. In the model, the shear strength is characterized as fully drained and undrained. This seems a bit crude since, in reality, the actual strength will be in between these two extremes. In both the case of the levee failure near Breitenhagen and the Leendert de Boerspolder-experiment, the difference between drained soil response and undrained soil response is not very significant, but this should not be considered a conclusion for every case. Research into the influence of a time-dependent, partly drained/undrained shear strength response could provide a better characterization of the soil shear strength.

Probabilistic analysis can help to estimate the likelihood of each scenario explaining the failure. This approach takes the correlation of all contributors into account and explores all possible combinations of possible scenarios explaining the failure and most model choices to identify the most accurate characterization of the failure. In future work, more attention can be paid to the correlations between the input parameter and how these correlations will influence the identification of the most likely scenario explaining the failure.

In the hindcasting of the levee of the Leendert de Boerspolder, the established relationship between the reliability index and the factor of safety is used to compute the probability of failure given the resulting factor of safety of each built slope stability model. However, this relationship is based on data coming from levees that do not represent the conditions of the levee of the Leendert de Boerspolder well, which means that the computed probabilities of failure are not sufficiently accurate. For future assessments, it is suggested to use relationships between factors of safety and reliability index derived based on data from comparable cases as for geometry materials and loads. Applying such a relationship will provide a more accurate probabilities of failure.

This study shows that observational information is vital to the identification of the most dominant contributing variable. Therefore it is recommended to explore other possibilities to include more evidence in the hindcasting with the help of Bayesian techniques. Examples are the inclusion of past performance information such as observed water seepage, unexpected relative deformations, or survival of high water levels.

Since the slip surface has a relatively simple geometry in the case of the Leendert de Boerspolder-experiment, the slope instability has also been characterized in this analysis using a LEM. Finite Element Methods (FEM) analyses may provide a more accurate characterization of slope instability. For future work and especially for complex geometries and shear strength, FEM should be considered (see e.g. (Varkey et al., 2017)) as part of hindcasting. However, this comes at the expense of computational time.

Influence of deviating conditions on the failure rates

As observed failure rates and the computed reliability of individual levee sections often differ significantly, it is recommended to investigate how observed failure rates can be used to calibrate and improve reliability assessments. Possibly, the levee failure near Breitenhagen (Kool et al., 2020) and the river system of Sachsen-Anhalt (Kool et al., 2022) can be used as a case study.

It is recommended to add data on other levee systems in the failure rate analysis that might provide more insights into the typical performance of a levee section to develop generic failure rates. For instance, the failure rates could also be coupled with (simplified) risk assessments and optimisation or other deviating conditions, then included in this research. As an example, in the work of Hui et al. (2016), the levee fragility is parameterized with levee height, crown width and conditions, based on synthetic levee performance curves.

6.2.2 Recommendations for engineering practice, design and policy

The findings of this study could contribute to advancing our understanding of the failure mechanisms of levees, to improve the estimate of the performance of levees. In this section, various suggestions are given for engineering practice, design and policy regarding levees.

Deterministic and probabilistic hindcasting

It is recommended to introduce a standardized approach for analysing levee failures. The methods developed in this thesis can be used to identify which information needs to be collected. Information

can also be stored and shared in public datasets such as the ILPD (leveefailures.tudelft.nl). This contributes to the overall quality, repeatability, transferability, transparency and recognisability of the analysis of failures. This will hopefully lead to more insights into the performance of levees.

The levee failure near Breitenhagen is explained by the occurrence of unexpected low shear strength, possibly related to the old breach. More attention should be paid to the identification of the values of strength parameters during the engineering of levees and safety assessment.

The influence of deviating conditions on the failure rates

Research in this dissertation shows that the unexpected soil scenarios dominated the failure of the levee failure near Breitenhagen, Sachsen-Anhalt in Germany (2013) and the updated failure rate of a levee section within the river system of Sachsen-Anhalt, Germany (2002 and 2013) is significantly affected by visually identifiable deviations. Therefore, it is advised to pay more attention to the identification of anomalies, and effects of unexpected vegetation which possibly affect the performance of the levee locally. During engineering or safety assessments of a levee, different scenarios of soil scenarios should be identified and included in the engineering or assessments. This is possibly done by analysing historical data on structural modifications or other issues, satellite imagery, and geological information. This information could be used to tailor soil investigations. Formalisation of these analyses could be done by setting up a database in which these types of data are managed and supplemented in sufficient detail, structurally and systematically.

Future risk assessments should incorporate the possibility of multiple breaches during high water events. Further analysis of the correlation between the breaches would be valuable for emergency response management.

It is also expected that the presented methodology could play a role in risk and reliability assessments in data-poor environments. In such circumstances, there are often limited insights into soil data and levee profiles, but some information on historical failures and satellite imagery is generally available thus allowing quantification of failure rates. It is recommended to set up one or more pilot cases to further test and develop the approach for systematic analysis of historical failure rates. Also, the failure rates could be coupled with (simplified) risk assessments and optimization, e.g. as proposed by Hui et al. (2016). In this approach, the levee fragility is parameterized with levee height, crown width and condition, based on synthetic levee performance curves.

Bibliography

Aanstoos, J., Hasan, K., O'Hara, C., Dabbiru, L., Mahrooghy, M., Nobrega, R., & Lee, M. (2012). Detection of slump slides on earthen levees using polarimetric SAR imagery. https://doi.org/10.1109/AIPR.2012.6528207

Apel, H., Thieken, A. H., Merz, B., & Bloschl, G. (2004). Flood risk assessment and associated uncertainty. *Natural Hazards and Earth System Sciences, 4*, 295-308.

Baecher, G. (2017). Bayesian Thinking in Geotechnics. Proceeding of GeoRisk2017, Denver.

Baecher, G., Paté, M., & de Neufville, R. (1980). Risk of Dam Failure in Benefit-Cost Analysis. *Water Resources Research, 16*(3), 449-456. https://doi.org/10.1029/WR016i003p00449

Baecher, G. B., & Christian, J. T. (2003). *Reliability and Statistics in Geotechnical Engineering*. John Wiley and Sons Ltd.

Baecher, G. B., & Christian, J. T. (2015). Bayes Factors and the Observational Method. *Geotechnical Safety and Risk V*. https://doi.org/10.3233/978-1-61499-580-7-939

Baker , J., & Calle , E. (2006). *JCSS Probabilistic Model Code, Section 3.7: SOIL PROPERTIES (revised version)*.

Bell, G. R. (2001). Civil Engineering Investigation. In K. L. Carper (Ed.), *Forensic Engineering, .* CRC Press.

Bernitt, L., & Lynett, P. (2010). Breaching of sea dikes *Coastal engineering*

Bishop, A. W. (1955). The use of the slip cicle in the stability analysis of slopes. *Géotechnique, 5*, 7-17.

[Record #495 is using a reference type undefined in this output style.]

Borsje, H., Renier, B., & Bruggraaf, H. (2014). Collapse of the roof of a football stadium. 37th IABSE, Madrid.

Brinkman, R. (2015). *Probabilistic Toolkit.* In (Version 1.8.12.629) Deltares. https://www.deltares.nl/en/software/probabilistic-toolkit-ptk/

Buijs, M., Wiersma, A., & Balen, A. v. (2013). *Applying a piping model on field data and historical data of dike failure events* [Master thesis, Vrije Universiteit].

Calle , E. (2008). Statistiek bij regionale proevenverzamelingen. *GEOtechniek 1*, 40-44.

Carper, K. L. (2000). *Forensic engineering*. CRC Press LLC.

CIRIA. (2013). *The International Levee Handbook*.

Ciullo, A., De Bruijn, K. M., Kwakkel, J. H., & Klijn, F. (2019). Systemic FLood Risk Management: The Challenge of Accounting for Hydraulic Interactions. *Water, 11*. https://doi.org/10.3390/w11122530

Dawotola, A., Gelder, F. H. A. J. M. v., Charima, J. J., & Vrijling, J. K. (2011). Estimation of failure rates of crude product pipelines. *Applications of Statistics and Probability in Civil Engineering*, 1741-1747.

de Gast, T. (2020). *Dykes and Embankments: a Geostatistical Analysis of Soft Terrain* (Publication Number ISBN 978-94-028-1915-1) [Dissertation, Delft University of Technology]. Delft.

Deen, J. v., & Duinen, A. v. (2016). *Schematiseringshandleiding macrostabiliteit*.

Deltares. (2016). *D-Geo Stability.* In

DeNeale, S., Baecher, G., Steward, K., Smith, E., & Watson, D. (2019). *Current State-of-Practice in Dam Safety Risk Assessment*. R. US Nuclear Regulatory Commission, MD, US DEPARTMENT OF ENERGY.

DLR. (2010). *Floods on the river Schwarze Elster (Germany) 2010*. Center for Satellite Based Crisis Information (ZKI). Retrieved 20 Nov. from https://activations.zki.dlr.de/en/activations/items/ACT088.html#:~:text=On%20September%2030%2C%202010%2C%20heavy,other%20dykes%20in%20the%20region

Drews, N. (2017a). *Investigation of the dike failure near Breitenhagen (Saale River) during the flood in June 2013* TU Dresden]. Dresden

Drews, N. (2017b). *Investigation of the dike failure near Breitenhagen (Saale River) during the flood in June 2013* TU Dresden]. Dresden.

ENW. (2009). *Technisch Rapport Actuele sterkte van dijken*

Faulkner, D., Warren, S., Spencer, P., & Sharkey, P. (2018). Can we still predict the future from the past? Implementing non-stationary flood frequency analysis in the UK. *Journal of Flood Risk Management*. https://doi.org/10.1111/jfr3.12582

Finkelstein, M. (2008). *Failure Rate Modelling for Reliability and Risk* (A. D. S. Rees, Springer, Ed.) https://doi.org/10.1007978-1-84800-986-8

Foster, M., Fell, R., & Spannagle, M. (2000). The statistics of embankment dam failures and accidents. *Canadian Geotechnical Journal, 37*, 1000-1024.

GeoBasis-DE/BKG. (2009). *Overview photo of location of levee failure near Breitenhagen*. Retrieved May 5 from

Gilbert, R. B. (2016). Important Role of Uncertainty in Forensic Geotechnical Engineering. In *Forensic Geotechnical Engineering* (pp. 493-504). Springer

Gilbert , R. B., Wright, S. G., & Liedtke, E. (1998). Uncertainty in back analysis of slopes: Kettleman Hills case history. *J. Geotech. Geoenviron. Eng.*, 1167-1176.

Gocht, M. (2004). *Deichbruche und Deichuberstromungen an Elbe und Mulde im August 2002*.

Google, E. P. (2020a). *(1-6-2001, 31-12-2001) Herz, Sachsen-Anhalt, Germany ,52022'02.86" N 11055'54.00" O, Eye alt 727m* Version 7.3.2.5776).

Google, E. P. (2020b). *(30-9-2013). Sachsen-Anhalt, Germany. 51° 48' 01.22"N, 12° 52' 56.44"O, Eye alt 446 m.* http://www.earth.google.com Version 7.3.2.5776).

Google, E. P. (2020c). *(2020). Sachsen-Anhalt, Germany. 51° 52' 24.18"N, 15° 08'18.25"O, Eye alt 311.03 km. Landsat/Copernicus, http://www.earth.google.com [august 27 2021]* Version 7.3.3.7786).

Grubert, P. (2013a). *Saaledeich bei Breitenhagen, Geotechnische Untersuchungen der Bruchstelle Empfehlungen zur Sanierung*.

Grubert, P. (2013b). *Saaledeich bei Breitenhagen, Geotechnische Untersuchungen der Bruchstelle Empfehlungen zur Sanierung*

Hall, J., Arheimer, B., Borga, M., Brazdil, R., Claps, P., Kiss, A., Kjeldsen, T. R., Kruayciuniene, J., Kundzewicz, Z. W., Lang, M., Llasat, M. C., Macdonald, N., McIntyre, N., Mediero, L., Merz, B., Merz, R., Molnar, P., Montanari, A., Neuhold, C., Parajka, J., Perdigao, R. A. P., Plavcova, L., Rogger, M., Salinas, J. L., Sauquet, E., Schar, C., Szolgay, J., Viglione, A., & Bloschl, G. (2014). Understanding flood regime changes in Europe: a state-of-the-art assessment. *Hydrology and Earth System Sciences, 18*, 2735-2772. https://doi.org/10.5194

Hall, J. W., Dawson, R. J., Sayers, P. B., Rosu, C., Chatterton, J. B., & Deakin, R. (2003). A methodology for national-scale flood risk assessment. *Proceedings of the Institution of Civil Engineers - Water and Maritime Engineering, 156*(3), 235-247. https://doi.org/10.1680/wame.2003.156.3.235

Hall, W. B. (1987). Reliability of Service-Proven Structures. *Journal of sSTructural Engineering 114*(3), 608-624.

Hatem, G. A. (1985). *Development of a Data Base on Dam Failures in the United States: Preliminary Results* Stanford University].

helpdesk. (2015). *Veiligheidsfactoren en belastingen bij nieuwe overtromingskansnormen* (Handreiking ontwerpen met overstromingskansen, Issue. V. e. L. Rijkswaterstaat Water.

Heyer, T. (2011). *Zuverlässigkeitsbewertung von Flussdeichen nach dem Verfahren der Logistischen Regression*. Technische Universität Dresden.

Heyer, T., & Stamm, J. (2013). Levee reliability analysis using logistic regression models - abilities, limitations and practical considerations. *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards, 7*(2), 77-87. https://doi.org/10.1080/17499518.2013.790734

Honjo, Y., Mori, H., Ishihara, M., & Otake, Y. (2015). On The Inspection of River Levee Safety in Japan by MLIT. *Geotechnical Safety and Risk V*, 873-878.

Hui, R., Jachens, E., & Lund, J. (2016). Risk-based planning analysis for a single levee. *Water Resources Research, 52*(4), 2513-2528. https://doi.org/10.1002/2014WR016478

Hutchison, K., Quigley, J. L., Raza, M., & Walls, L. (2009). *Empirical Bayes methodology for estimating equipment failure rates with application to power generation plants* Industrial Engineering and Engineering Management,

IKSE. (2004). *Dokumentation des Hochwassers vom August 2002 im Einzugsgebiet der Elbe*.

IPET. (2009). *Final Report of the Interagency Performance Evaluation Task Force* (Engineering and operational Risk and Reliability Analysis, Issue. USACE.

Jeffreys, H. (1998). *Theory of probability*. Oxford University Press.

Jiabi, X., Sayers, P. B., Dongya, S., & Hanghui, Z. (2013). Broad-scale reliability analysis of the flood defence infrastructure within the Taihu Basin, China. *Journal of Flood Risk Management, 6*, 42-56. https://doi.org/10.1111/jfr3.12034

Jongejan, R. B., & Maaskant, B. (2015). Quantifying Flood Risks in the Netherlands. *Risk Analysis, 35*, 252-264}

Jonkman, S. N. (2005). Global perspectives of loss of human life caused by floods. *Natural Hazards 34*, 151-175.

Jonkman, S. N. (2007). *Loss of life estimation in flood risk assessment* [Docteral thesis, Technical University of Delft]. Delft.

Jonkman, S. N., Jorissen, R. E., Schweckendiek, T., & Bos, J. P. v. d. (2017). *Flood Defences, Lecture notes CIE5314*. Delft University of Technology.

Jonkman, S. N., Kok, M., & Vrijling, J. K. (2008). Flood Risk Assessment in the Netherlands: A Case Study for Dike Ring South Holland. *Risk Analysis, 28*(5). https://doi.org/10.1111/j.1539-6924.2008.01103.x

Jonkman, S. N., & Schweckendiek, T. (2015). Briefing: Lessons learned from failure of flood defences. *Forensic Engineering, 168*, 85-88.

Jonkman, S. N., Steenbergen, R. D. J. M., Morales-Napoles, O., Vrouwenvelder, A. C. W. M., & Vrijling, J. K. (2016). *Probabilistic Design: Risk and Reliability Analysis in Civil Engineering, Lecture notes CIE 4130*.

Kanning, W., Teixeira, A., Krogt , v. d. M., & Rippi, K. (2016). *Derivation of the semiprobabilistic safety assessment rule for inner slope stability`*. Deltares.

Kanning , W., van Baars, S., & Vrijling, J. K. (2008). The stability of flood defences on permeable soils- the London Avenue Canal failure in New Orleans. Sixth International Conference on Case Histories in Geotechnical Engineering,

Kanning, W., & van der Krogt, M. G. (2016). *Memo: Pore water pressure uncertainties for slope stability*.

Klerk, W. J., Kanning , W., & Kok, M. (2018). *Time-dependent reliability in flood protection decision making in Netherlands* Safety and Reliability - Safe Societies in a Changing World,

Kool, J., Kanning, W., Heyer, T., Jommi, C., & Jonkman, S. N. (2019). Forensic analysis of levee failures: The Breitenhagen case. *International Journal of Geoengineering Case Histories, 5*(2), 70-92. https://doi.org/10.4417/IJGCH-05-02-02

Kool, J., Kanning , W., Jommi, C., & Jonkman, S. N. (2020). A Bayesian hindcasting method of levee failures applied to the Breitenhagen slope failure. *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*. https://doi.org/10.1080/17499518.2020.1815213

Kool, J., Kanning , W., & Jonkman, S. N. (2022, 8 December 2021). The influence of deviating conditions on levee failure rates. *Journal of Flood Risk Management*. https://doi.org/10.1111/jfr3.12784

Kortenhaus, A. (2003). *Probabilistische Methoden für Nordseedeiche* [Ph.D. Thesis]. Technische Universität Carolo-Wilhelmina zu Braunschweig.

Ladd, C. C. (1991). Stability evaluation during staged construction (22nd Terzaghi Lecture). *J. of Geotech. Eng.*, 540-615.

Lane, E. (1935). Security from under-seepage masonry dams on earth foundations. *Transactions American Society of Civil Engineers*(100), 1235-1351.

LHW. (2014). *Bericht uber das Hochwasser im Juni 2013 in Sachsen-Anhalt*. L. f. H. u. W. Sachsen-Anhalt. https://lhw.sachsen-anhalt.de/fileadmin/Bibliothek/Politik_und_Verwaltung/Landesbetriebe/LHW/neu_PDF/4.0/SB_Hochwasserschutz/Hochwasserbericht_2013.pdf

Luckman, P. G., Der Kiureghian, A., & Sitar, N. (1987). Use of stochastic stability analysis for Bayesian back calculation of pore pressures acting in a cut at failure. 5th Int. Conf. on Application of Statistics and Probability in Soil and Struct. Engr., Vancouver, Canada.

LVermGeo. (2019). *Kostenfreie Digitale Topographische Karte 1 : 100 000 (DTK100) mehrfarbig* [maps of Sachsen-Anhalt in tiff]. https://www.lvermgeo.sachsen-anhalt.de/de/dtk100/kostenfreie-dtk100-mehrfarbig.html

Major, J. (2019). *Best Practices in Dam and Levee Safety Risk Analysis* https://www.usbr.gov/ssle/damsafety/risk/methodology.html

Merz, B., Nguyen, V. D., Apel, H., Gerlitz, L., Schroter, K., Steirou, E., & Vorogushyn, S. (2018). Spatial coherence of flood-rich and flood-poor periods across Germany. *Journal of Hydrology, 559*, 813-826.

Muraro, S. (2019). *The deviatoric behaviour of peat: a route between past empiricism and future perspectives* [PhD thesis]. Delft University of Technology.

Neale, B. S. (1999). *Forensic engineering: A professional approach to investigation*. Thomas Telford Ltd

Nickles, T. (1987). Methodology, heuristics, and rationality. In J. C. Pitt (Ed.), *Rational changes in science: Essays on Scientific Reasoning* (pp. 103–132). Springer.

Normcommissie. (2011). *NEN9997-1 Geotechnisch ontwerp van constructies - deel 1: Algemene regels*.

Ozer, I. E., Damme, M. v., & Jonkman, S. N. (2020). Towards and International Levee Performance Database and Its Use for Macro-Scale Analysis of Levee Breaches and Failures. *Water, 12*(1), 119.

Ozer, I. E., Damme, v. M., & Jonkman, S. N. (2020). Towards and International Levee Performance Database and Its Use for Macro-Scale Analysis of Levee Breaches and Failures. *Water, 12*(1), 119.

Ozer, I. E., Rikkert, S., van Leijen, F., Jonkman, S. N., & Hanssen, R. F. (2019). Sub-seasonal Levee Deformation observed Using Satellite Radar Interferometry to Enhance Flood Protection. *Scientific Reports* https://doi.org/10.1038/s41598-019-39474-x

Pol, J., Kanning , W., & Jonkman, S. N. (2020). Timeporal Development of Backward Erosion Piping in a Large-scale Experiment. *Journal of Geotechnical and Geoenvironmental Engineering*. https://doi.org/10.1061

Rackwitz, R. (2001). Reliability analysis - a review and some perspectives. *Structural Safety, 23*(4), 365-395.

Rao, V. V. S. (2016). *Forensic Geotechnical Engineering* (V. V. S. Rao, Ed.). Springer. https://doi.org/10.1007/978-81-322-2377-1

Rikkert, S., & Kok, M. (2019). FAALKANSEN VAN BOEZEMKADEN VANUIT EEN STATISTISCH PERSPECTIEF. *H2O*.

Robertson, P. K. (2009). Interpretation of cone penetration tests - a unified approach. *Canadian Geotechnical Journal 46*(11), 1337-1355.

Rozing , A. (2015). *Onzekerheden Waterspanningen in WTI 2017.*

RWS. (2016a). 05_shmacrostabiliteit_31jan2017. *Helpdesk Water*. Retrieved december 1, from https://www.helpdeskwater.nl/publish/pages/132667/05_shmacrostabiliteit_31jan2017.pdf

RWS. (2016b). 05_shmacrostabiliteit_31jan2017. *Helpdesk Water*. Retrieved december 1, from https://www.helpdeskwater.nl/publish/pages/132667/05_shmacrostabiliteit_31jan2017.pdf

[Record #487 is using a reference type undefined in this output style.]

Schneider, K. J., & Albert, A. (2014). *Bautabellen fur Ingeriuere*.

Schofield, A. N., & Wroth, C. P. (1968). *Critical state Soil mechanics*. McGraw-Hill.

Schroter, K., Kunz, M., Elmer , F., & Muhr , B. (2015). What made the June 2013 flood in Germany an exceptional event? A hydrometeorological evaluation. *Hydrology and Earth System Sciences*, 309-327.

Schwandt, D., & Hubner, G. (2019). *Extremereignisse im Elbegebiet: Hochwasser, Niedrigwasser*. http://undine.bafg.de/elbe/extremereignisse/elbe_extremereignisse.html

Schweckendiek, T. (2014a). *On Reducing Piping Uncertainties*. T.Schweckendiek.

Schweckendiek, T. (2014b). *On Reducing Piping Uncertainties - A Bayesian Decision Approach*. Delft University of Technology.

Schweckendiek, T., Kanning, W., & Jonkman, S. N. (2014). Advances in reliability analysis of the piping failure mechanism of flood defences in the Netherlands *Heron 59* ((2/3)), 101-127.

Schweckendiek, T., van der Krogt, M., Rijneveld, B., & Teixeira, A. M. (2017). *Handreiking Faalkansanalsye Mactrostabiliteit*.

Schweckendiek, T., van der Krogt, M., Teixeira, A., Kanning , W., Brinkman, R., & Rippi, K. (2017). Reliability Updating with Survival Information for Dike Slope Stability Using Fragitlity Curves *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards, 7*, 494-503.

Schweckendiek, T., & Vrouwenvelder, A. C. W. M. (2014). Updating piping reliability with field performance observations. *Structural Safety, 47*, 13-23.

Schwiersch, N. D., B.O.R.I.S.

Stamm, J. . (2021). Probabilistic Determination of the Phreatic Line in River Levees Under STeady-State Conditions and its Effect on the Stability Statement. European Safety and Reliability Conference Singapore.

Seed, R. B., Bea, R. G., Abdelmalak, R. I., Athanasopoulos, A. G., Boutwell, G. P., Bray, J. D., Briaud, J. L., Cheung, C., Cobos-Roa, D., Cohen-Waeber, J., Collins, B. D., Ehrensing, L., Farber, D., Hanemann, M., Harder, L. F., Inkabi, K. S., Kammerer, A. M., Karadeniz, D., Kayen, R. E., Moss, R. E. S., Nicks, J., Nimmala, S., Pestana, J. M., Porter, J., Rhee, K., Riemer, M. F., Roberts, K., Rogers, J. D., Storesund, R., Govindasamy, A. V., Vera-Grunauer, X., Wartman, J. E., Watkins, C. M., Wenk Jr., E., & Yim, S. C. (2006). *Investigation of the Performance of the New Orleans Flood Protection Systems in Hurricane Katrina on August 29, 2005* [Final report]. U. o. California.

Seed, R. B., Bea, R. G., Athanasopoulos, A. G., Boutwell, G. P., Bray, J. D., Cheung , C., & Wartman, J. E. (2008). The 17th Street Drainage Canal. In New Orleans and Hurrican Katrina. *Geotech. Geoenviron. Eng*(740-761).

Sharp, M. K., Wallis, M., Deniaud, F., Hersch-Burdick, R., Tourment, R., Matheu, E., Seda-Sanabria, Y., Wersching, S., Veylon, G., Durand, E., Smith, P., Forbis, J., Spliethoff, C., van Hemert, H., Igigabel, M., Pohl, R., Royet, P., & Simm, J. (2013). *The International Levee Handbook*. CIRIA.

Siebert, A. (2020, 19 november 2020). *Level information Schwarze Elster* [Interview].

[Record #116 is using a reference type undefined in this output style.]

Specter, M. M. (1987). National Academy of Forensic Engineers. *Journal of Performance of Constructed Facilities (ASCE), 3*(1).

Spencer, E. (1967). A METHOD OF ANALYSIS OF THE STABILITY OF EMBANKMENTS ASSUMING PARALLEL INTER-SLICE FORCES. *Géotechnique*, 11-26.

Steenbergen, H. M. G. M., Lassing, B. L., Vrouwenvelder, A. C. W. M., & Waarts, P. H. (2004). Reliability analysis of flood defence systems. *HEron, 49*(1), 51-73.

STOWA. (2015). *Stress test made ancient clay-on-peat levee finally to fail, the Netherlands*. NL Netherlands,
STOWA,
PARTNERS FOR WATER. Retrieved 2022-1-27 from https://www.dutchwatersector.com/news/stress-test-made-ancient-clay-on-peat-levee-finally-to-fail-the-netherlands

TAW. (2001). *Technisch Rapport Waterkerende Grondconstructies (TRWG)*.

TAW. (2004). Technisch Rapport Waterspanningen bij dijken.

Terwel, K., Loeve, A., & Schuurman, M. (2018). Improving reliability in forensic engineering: the Delft approach. *Proceedings of the Institution of Civil Engineers – Forensic Engineering, 171*(3), 99-106.

Terwel, K., Ratay, R., Bruhwiler, E., Wood, J., Godart, B., & Dalmy, D. (2012). *An Initial Survey of Forensic Engineering Practices in Some European Countries and the USA.* Int. Symp ASCE 6th Congress of Forensic Engineering, San Francisco.

Thieken, A. H., Kienzler, S., Kreibich, H., Kuhlicke, C., Kunz, M., Mühr, B., Müller, M., Otto, A., Petrow, T., Pisi, S., & Schröter, K. (2016). Review of the flood risk management system in Germany after the major flood in 2013. *Ecology and Society, 21*(2), Article 51. https://doi.org/http://dx.doi.org/10.5751/ES-08547-210251

Thieken, A. H., Kienzler, S., Kreibich, H., Kuhlicke, C., Kunz, M., Mühr, B., Müller, M., Otto, A., Petrow, T., Pisi, S., & Schröter, K. (2016). Review of the flood risk management system in Germany after the major flood in 2013. *Ecology and Society 21*(2), Article 51. https://doi.org/http://dx.doi.org/10.5751/ES-08547-210251

USACE. (2000). *Design Guidance for Levee Under-Seepage*.

van Baars, S. (2005). The horizontal failure mechanism of the Wilnis peat dyke. *Géotechnique, 55*.

van Damme, M. (2020). An analytical process-based approach to predicting breach width in levees constructed from dilatant soils. *Natural Hazards*. https://doi.org/10.1007/s11069-020-03862-8

van der Krogt, M., Schweckendiek, T., & Kok, M. (2020). Improving dike reliability estimates by incorporating construction survival *Engineering Geology*

van Duinen, A. (2015). *Modelonzekerheidsfactoren Spencer-Van de Meij model en ongedraineerde schuifsterkte* (1207808-001-GEO-0006). (Programma WTI 2017, cluster Stabiliteit, Issue. Deltares.

Vanmarcke, E. (1977). Probabilistic Modeling of Soil Profiles *Journal of Geotechnical and Geoenvironmental Engineering*. https://doi.org/10.1016/0148-9062(78)90012-8

Varkey, D., Hicks, M., & Vardon, P. (2017). *Influence of Spatial Variability of Shear Strength Parameters on 3D Slope Reliability and Comparison of Analysis Methods*. https://doi.org/10.1061/9780784480717.038

Von Thun, J. L. (1985). *Application of Statistical Data From Dam Failures and Accidents to Risk-Based Decision Analysis on Existing Dams*.

Vorogushyn, S., Merz, B., & Apel, H. (2009). Development of dike fragility curves for piping and micro-instability breach mechanisms. *Natural Hazards and Earth System Sciences, 9*, 1883-1401.

Vrijling, J. K., & Gelder, P. H. A. J. M. v. (2002). *PROBABILISTIC DESIGN IN HYDRAULIC ENGINEERING*. Delft University of Technology.

Vrouwenvelder, A. (2006, April 26-29). Spatial effects in reliability analysis of flood protection systems. *International Forum on Engineering Decision Making, Second IFED Forum*.

Vrouwenvelder, A. C. W. M., & Steenbergen, H. M. G. M. S., K.A.H. (1999). *THEORIEHANDLEIDING PC-RING, Deel B: Statisctische modellen* (98-CON-R1431).

Vrouwenvelder, A. C. W. M., & Vrijlling, J. K. (1982). *Probabilistisch Ontwerpen CTow4130*. Technische Universiteit Delft.

Weichel, T. (2013). *Failure of the Breitenhagen levee (2013)*. Landesbetrieb für Hochwasserschutz und Wasserwirtschaft Sachsen-Anhalt.

Weichel, T. (2020, 20 November 2020). *Length of earthen levees along the big rivers of Sachsen-Anhalt* [Interview]. Kool, J.

Weichel, T. (2021, 27-8-2021). *AW: length of earthen levees along the big rivers of Sachsen-Anhalt* [Interview].

Winsemius, H. C., Aerts, J. C. J. H., van Beek, L. P. H., Bierkens, M. F. P., Bouwman, A., Jongman, B., Kwadijk, J. C. J., Ligtvoet, W., Lucas, P. L., van Vuuren, D. P., & Ward, P. J. (2016, april 2016). Global drivers of future river flood risk. *nature climate change, 6*. https://doi.org/10.1038/NCLIMATE2893

Zang, L. M., Xu, Y., & Peng, M. (2013). Assessment of flood risks in Pearl River Delta due to levee breaching. *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards, 7*(2), 122-133.

Zhang, J., Tang, W. H., & Zhang, L. M. (2010). Efficient Probabilistic Back-Analysis of Slope Stability Model Parameters. *Journal of Geotechnical and Geoenvironmental Engineering*, 99-109.

Zhang, J., Zhang, L. M., & Tang, W. H. (2011). Slope Reliability Analysis Considering Site-Specific Performance Information. *Journal of Geotechnical and Geoenvironmental Engineering, 137*, 227-238.

Zwanenburg, C., López-Acosta, N. P., Tourment, R., Tarantino, A., Pozzato, A., & Pinto, A. (2018). Lessons Learned from Dike Failures in Recent Decades. *International Journal of Geoengineering Case Histories, 4*(2), 203-229. https://doi.org/10.4417/IJGCH-04-03-04

Appendix A. Glossary

Chapters 2, 3 and 4 are based on previously published articles in journals, which are reproduced in their original form. Therefore, there are some minor differences in terminology across the different chapters. This glossary highlights some of the key terms and discusses how some of these terms have been used differently in the different chapters.

Forensic engineering:
Forensic Engineering is “the investigation of failures - ranging from serviceability to catastrophic - which may lead to legal activity, including both civil and criminal” (Neale, 1999). However, since this thesis focuses on the hindcasting part; the litigation part is not included. In other work, the term “accident reconstruction” is used when the litigation is not part of the analysis (Carper, 2000).

Hindcasting:
Building a calculative model that characterizes the (levee) failure at the time of the failure most accurately and is enabling a forensic engineer to determine the most likely cause(s) of failure.

Characterization of the failure:
The combination of a scenario explaining the failure and the most representative model choices that are included in a built, calculative, model to evaluate the failure

Scenario:
The term “scenario” has been given a different meaning in Chapter 2 than in Chapters 3 and 4.
In Chapter 2, the term “scenario” refers to the combination of the loading and subsoil scenario possibly explaining the failure in combination with the model choices, which is similar to characterization.
In Chapters 3 and 4, the term “scenario” refers to loading and subsoil conditions related to hydraulic loads, pore pressures and stratigraphy that possibly explain the failure.

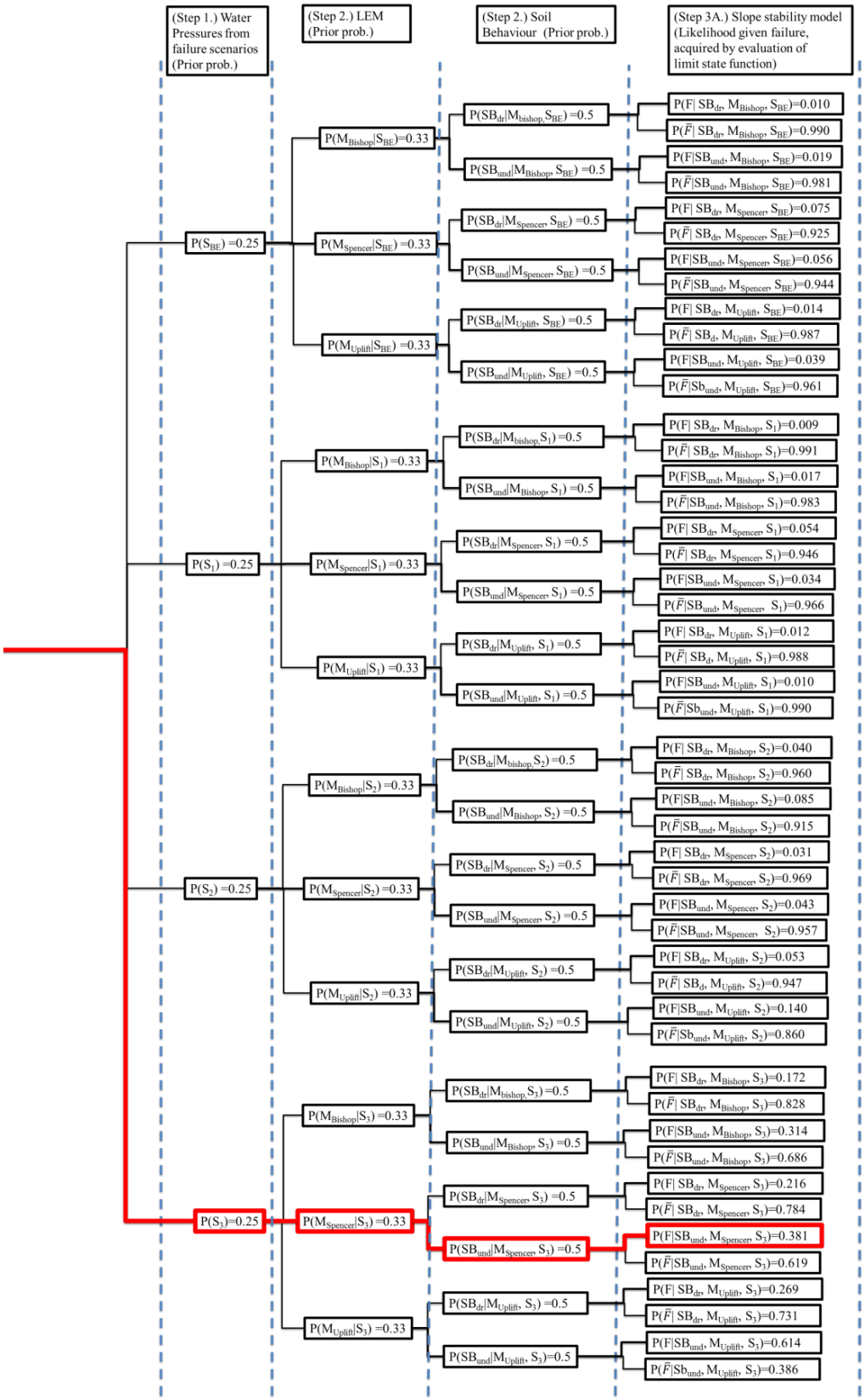
Failure observations:
The term failure observation concerns information that is observed from the structure at time the failure occurred, i.e. in the case of the levee failure near Breitenhagen and the levee failure of Leendert de Boerspolder this refers to the geometry of the slip surface

The most likely cause triggering the failure:
The dominant uncertaint parameter that contributed to the failure given the most likely scenario as well as the most representative model choices.

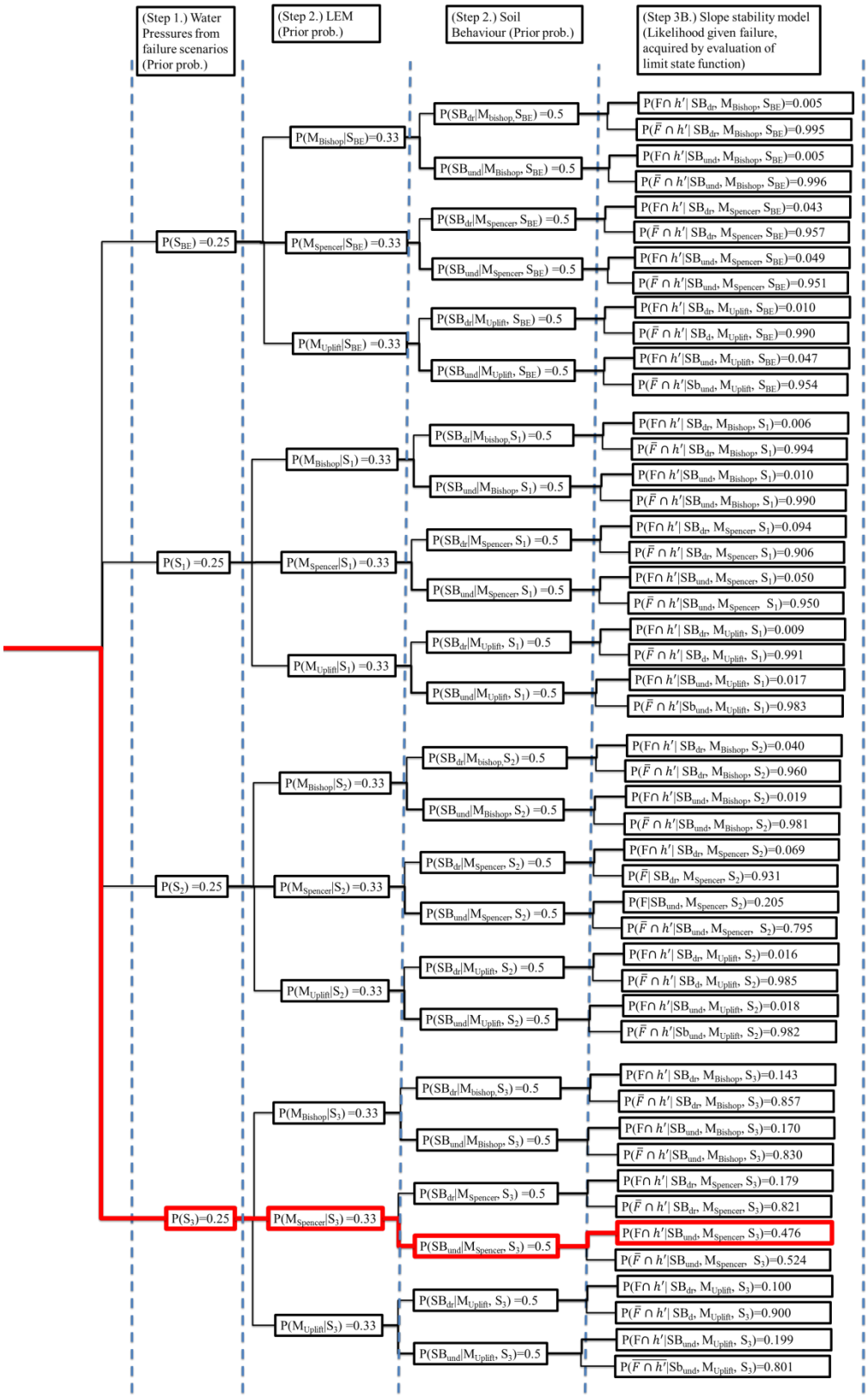
Failure rate:
The failure rate is the probability of failure per unit of time, or more formally the conditional probability of failure at time t, i.e. the probability of failure in an infinitesimal unit interval of time, given that the unit has survived until then. Typically, failure rates are associated with a characteristic life cycle failure function of an object (Bathtub curve) in three phases, i.e. infant mortality where construction or engineering errors display, the usage period that is marked with a near constant failure rate, and the wearing period where the failure rate increases due to deterioration of the object (Finkelstein, 2008).

Appendix B. Complete event tree for Bayesian hindcating of the Breitenhagen case

Event tree with scenarios of failure, alternative in model choices, and related likelihood given failure and survival. The red line indicates the most likely scenario of failure and the most representative model choices to characterize the failure.



Event tree with scenarios of failure, alternative in model choices, and related likelihood given failure and survival, including the field observation information. The red line indicates the most likely scenario of failure and the most representative model choices to characterize the failure, which results in the highest conditional probability of failure.



Example calculation of determining the posterior probability of S_3 "Pond connection with aquifer" given failure and field observation information. This is applied on the event tree with scenarios of failure, alternative in model choices, and related likelihood given failure and survival, including the field observation information.

Appendix C. FoS-β_H relation based on mean values

The data provided by Kanning et al. (2017) connect the resulting factors of safety (FoS) with the probability of failure in 44 cases. This data makes it possible to establish a general relationship between the computed FoS, based on mean values, and the minimum reliability index. The FoS's of the different cases are based on mean values. Originally, this relation was used to identify a minimum FoS to comply with a levee to meet a target reliability. This relation supports for a reliability assessment of a levee (Figure C.1). For hindcasting purposes, the relation is fitted to 50% of the FoS. The probability of failure given the built slope stability models $(D_i \cap M_j)$ is computed as a reliability index (β_H).

$$\beta_H = g(FoS) = (FoS - 0.96)/0.13$$

In which, FoS is the ratio between the mean driving moment (M_S) and the mean resisting moment (M_R) where the material factor $\gamma_m = 1$ is applied.

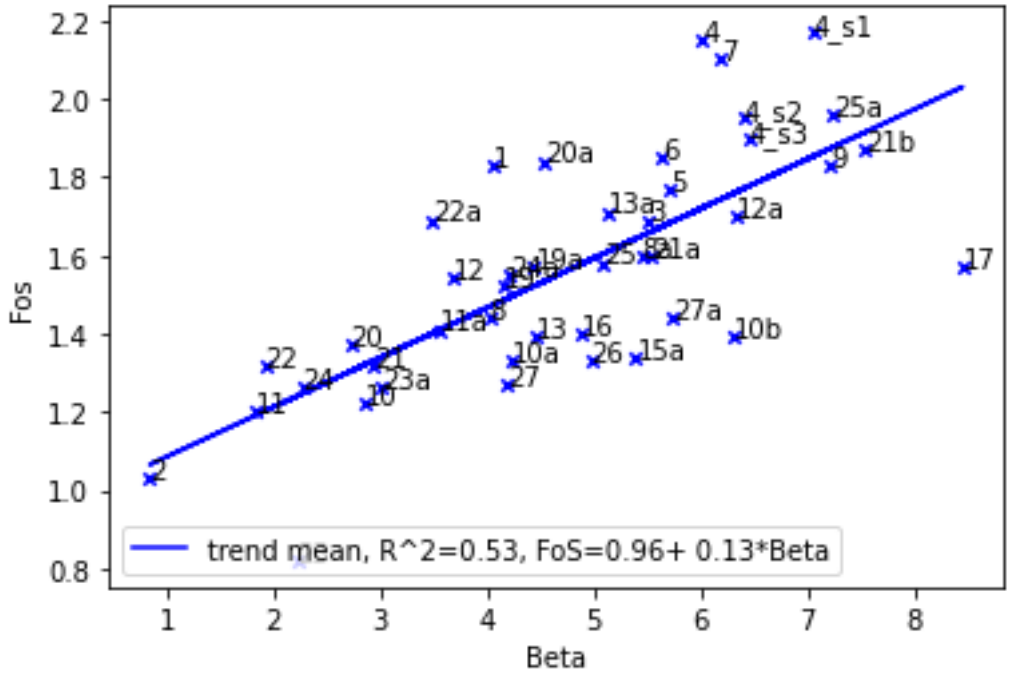


Figure C.1 Data of 44 cases and the linear relation between the computed FoS and the minimum reliability index

Acknowledgements

The process of researching and writing that resulted in this thesis has been very instructive for me and I have experienced it as a great adventure that I would not have wanted to miss. Although my name is on this thesis, it could never have achieved such quality without the support and advice I received from many people. I would therefore like to take this opportunity to thank a few people in particular, with which I realize that I can never be complete.

I would like to thank my supervisors. My (co-) promotors Bas Jonkman, Cristina Jommi and Wim Kanning, who have given me the opportunity and support. I want to thank you for the many discussions we have had and the many texts you have provided with comments and suggestions to improve them.

I would also like to thank my colleagues, who have always welcomed me warmly, and whom I could always go to for advice or moral support. In particular, I would like to thank my roommates Ece, Ermano, Erik, Orson and Stephan, with whom I started this adventure and who proved to be great travel companions. But also Jarit de Gijt, whose enthusiasm, creativity, enormous knowledge and beautiful stories have been a great source of inspiration over the years. I would also like to thank my colleagues from Royal HaskoningDHV for their flexibility, which enabled me to combine a position in their company and the realization of this thesis.

Finally, I would like to thank my family and friends for their encouragement and for having patience when I wanted to talk about my work in analyzing failed dikes again. In particular, Marielle, Vera and Hanna, who joined me during the realization of this thesis and made this adventure extra interesting.