Reliability updating for slope stability

Improving dike safety assessments using performance information

van der Krogt, M.G.

DOI
10.4233/uuid:2af85b15-9208-47bc-851d-9313d65cefcb

Publication date
2022

Document Version
Final published version

Citation (APA)
van der Krogt, M. G. (2022). *Reliability updating for slope stability: Improving dike safety assessments using performance information*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:2af85b15-9208-47bc-851d-9313d65cefcb

RELIABILITY UPDATING FOR SLOPE STABILITY

IMPROVING DIKE SAFETY ASSESSMENTS USING PERFORMANCE INFORMATION

RELIABILITY UPDATING FOR SLOPE STABILITY

IMPROVING DIKE SAFETY ASSESSMENTS
USING PERFORMANCE INFORMATION

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof. dr. ir. T.H.J.J. van der Hagen,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op
woensdag 11 mei 2022 om 10:00 uur

door

Marcus Gerardus VAN DER KROGT

civiel ingenieur, Technische Universiteit Delft, Nederland,
geboren te Zoeterwoude, Nederland.

Dit proefschrift is goedgekeurd door de

promotor: prof. dr. ir. M. Kok
copromotor: dr. ir. T. Schweckendiek

Samenstelling promotiecommissie:

Rector Magnificus	voorzitter
Prof. dr. ir. M. Kok	Technische Universiteit Delft, promotor
Dr. ir. T. Schweckendiek	Technische Universiteit Delft, copromotor
Onafhankelijke leden:	
Prof. dr. ir. S.N. Jonkman	Technische Universiteit Delft
Prof. dr. M.A. Hicks	Technische Universiteit Delft
Prof. dr. J.Y. Ching	National Taiwan University
Dr.-Ing. I. Papaioannou	Technische Universität München
Ir. M.M. de Visser	Rijkswaterstaat

Keywords: dikes, levees, slope stability, Bayesian reliability updating, performance information, site-specific transformation models, transformation uncertainty, proof loading, Bayesian decision analysis.

Printed by: Ipskamp Printing

Front & Back: Jessica van der Holst

ISBN 978-94-6366-531-5

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

SUMMARY

Dikes are crucial for the protection against floods. One of the ways in which dikes can fail is by the instability of the inner slope. Credible probabilities of failure for slope stability are essential for the safety assessment of existing dikes and the design of dike reinforcements. However, large uncertainties in soil properties of the dike and the subsoil often lead to high failure probabilities for slope stability of dikes (just like other geotechnical failure mechanisms). Uncertainties in soil properties mainly arise from limited data and can be reduced by obtaining more information. However, the uncertainty reduction that can be achieved with conventional methods such as soil investigation is limited by the amount of soil investigation that can reasonably be carried out. For further reduction of uncertainties and improvement of failure probability estimates, also other sources of information will have to be considered.

This dissertation focuses on improving failure probability estimates by using observed behaviour and performance of dikes. Examples of performance information are survived loads such as flood water levels or proof loads, and measurements during such loading conditions. The research uses Bayesian analysis to account for one or more performance observations or measurements in estimating failure probabilities. Using multiple case studies, this research identifies the observations and success factors leading to significantly lower failure probabilities. Furthermore, Bayesian decision analysis was used to consider the cost-effectiveness (Value of Information) of performance information, to determine which strategy of dike reinforcement and/or uncertainty reduction leads to the lowest overall cost to comply with a given safety level.

The first source of performance information considered in this research is the survival of the construction phase of dikes, the so-called 'proven strength'. Successful construction of this usually large load on a soft subsoil leads to a reduction in uncertainty of the soil properties. Depending on the situation, the probability of failure can reduce by more than a factor 10. The impact on the failure probability and a design is especially significant when the load effects during construction differ little from the future flood situation that is being assessed, such as is the case for dikes on soft subsoils.

In addition to 'proven strength', settlement measurements during the construction of a dike can be included to improve estimates of the probability of failure. In contrast to survival information, incorporating settlement measurements can also lead to higher failure probabilities, if settlements are larger than expected based on prior data. Settlement measurements lead to significant uncertainty reduction for stability especially if the uncertainty in the degree of over-consolidation of soft subsoil layers is large, which affects both settlement and stability.

Other sources of information for slope stability are pore pressure monitoring and proof loads. The cost-effectiveness of these measures depends on the expected uncertainty reduction and the costs involved in obtaining the data, such as investment costs and potential damage by proof loading. Proof loading appears to be cost-effective for

relatively expensive dike reinforcements of several million euros per kilometre of dike in combination with low risk of damage due to proof loading (e.g. as a result of contingency measures and damage control). Decision analysis enables to determine the optimal strategy for combinations of dike reinforcement and investing in performance information to comply with a given safety level. Pre-posterior decision analysis can be used to determine the optimal proof load for which the expected total costs are the lowest. Decision analysis also allows to identify the most effective method for uncertainty reduction and the optimal sequence of measures.

Incorporating multiple data of the behaviour and performance of dikes thus improves estimates of the failure probability for slope stability, leading to better safety assessments and more efficient design of dike reinforcements. The cases considered in this dissertation suggest that savings of several million euros per kilometre dike reinforcement are possible (10-35% compared to the current dike reinforcement costs), for the Dutch situation with typically relatively high cost of dike reinforcements compared to the costs and risks of obtaining performance information. The use of performance information therefore contributes to improving the efficiency of managing flood risk in the Netherlands, and in particular the dike reinforcements in the Dutch Flood Protection Programme.

This research was carried out as part of the All-Risk programme which supports the Dutch Flood Protection Programme in implementing the new safety standards for flood protection. The programme, aimed at the development of new methods aim to improve the efficiently of investments for flood protection, is largely funded by NWO, the Dutch Research Council, and supported by contributions of numerous end-users.

SAMENVATTING

Dijken zijn cruciaal voor de bescherming tegen hoogwater. Een van de manieren hoe een dijk kan falen is door het instabiel worden van het binnentalud: macro-instabiliteit. Geloofwaardige faalkansen voor macro-stabiliteit zijn essentieel voor de veiligheidsbeoordeling van bestaande dijken en het ontwerp van dijkversterkingen. Grote onzekerheden in de eigenschappen van de dijk en de ondergrond leiden echter vaak tot hoge faalkansen voor macro-stabiliteit van dijken (net zoals andere geotechnische faalmechanismen). Onzekerheden in grondeigenschappen komen voornamelijk voort uit beperkte gegevens en zijn reduceerbaar door meer informatie in te winnen. De onzekerheidsreductie die met conventionele methoden zoals grondonderzoek bereikt kan worden is echter begrensd door de redelijkerwijs uit te voeren hoeveelheid grondonderzoek. Voor verdere reductie van onzekerheden en verbetering van faalkansschattingen zal dus ook gekeken moeten worden naar andere informatiebronnen.

Dit proefschrift richt zich op het verbeteren van faalkansschattingen door gebruik te maken van geobserveerd gedrag van dijken. Voorbeelden van gedragsinformatie zijn overleefde belastingen zoals hoogwater of proefbelastingen, en metingen gedurende dergelijke belastingsituaties. In dit onderzoek wordt Bayesiaanse analyse gebruikt om faalkansen te berekenen die rekening houden met een of meerdere gedragsobservaties of metingen. Met behulp van meerdere case studies identificeert dit onderzoek de observaties en succesfactoren die leiden tot significant lagere faalkansen. Vervolgens is met behulp van Bayesiaanse beslisanalyse de kosteneffectiviteit (Value of Information) van gedragsinformatie beschouwd, om te bepalen welke strategie van dijkversterken en/of onzekerheidsreductie leidt tot de laagste totale kosten om te voldoen aan een vastgesteld veiligheidsniveau.

De eerste bron van gedragsinformatie die in dit onderzoek beschouwd is, is het overleven van de aanleg van dijken, de zogenaamde 'bewezen sterkte'. Het succesvol opbrengen van deze doorgaans flinke belasting voor de slappe ondergrond leidt tot een reductie van onzekerheid in de grondeigenschappen. Afhankelijk van de situatie kan de faalkans meer dan een factor 10 lager zijn. De impact op de faalkans en een ontwerp is vooral significant wanneer de belastingeffecten tijdens aanleg weinig verschillen van de toekomstige hoogwater situatie die beoordeeld wordt, zoals bij dijken op slappe ondergronden.

Naast 'bewezen sterkte' kunnen ook metingen van de zetting tijdens de aanleg van een dijk worden meegenomen om schattingen van de faalkans te verbeteren. In tegenstelling tot informatie van overleven, kan het meenemen van zettingsmetingen ook leiden tot hogere faalkansen, als er meer zetting optreedt dan verwacht op basis van alleen vooraf beschikbare gegevens. Zettingsmetingen leiden vooral tot significante onzekerheidsreductie voor stabiliteit als de onzekerheid in de mate van overconsolidatie van slappe grondlagen groot is, welke zowel zetting als stabiliteit beïnvloedt.

Andere bronnen van informatie voor macrostabiliteit zijn waterspanningsmonitoring en proefbelastingen. De kosteneffectiviteit van deze maatregelen hangt af van de verwachte onzekerheidsreductie en de kosten gemoeid met het inwinnen van de gegevens, zoals investeringskosten en potentiële schade bij proefbelastingen. Proefbelastingen blijken kosteneffectief voor relatief dure dijkversterkingen van meerdere miljoenen euro per kilometer dijk in combinatie met lage risicokosten van schade door proefbelasten (bijvoorbeeld als gevolg van beheersmaatregelen). Beslisanalyse maakt het mogelijk om te bepalen wat de optimale strategie is van combinaties van dijkversterken en investeren in gedragsinformatie, om een vastgesteld veiligheidsniveau te bereiken. Preposterior beslisanalyse kan gebruikt worden om de optimale proefbelasting te bepalen, waarvoor de verwachte totale kosten het laagst zijn. Beslisanalyse stelt ons ook in staat om te identificeren welke methode voor onzekerheidsreductie het meest effectief is, en om te bepalen wat de optimale volgorde van maatregelen is.

Het meenemen van meerdere gegevens over het gedrag en de prestatie van dijken leidt dus tot betere schattingen van de faalkans voor macro-stabiliteit, wat leidt tot betere veiligheidsbeoordelingen, en efficiëntere ontwerpen voor dijkversterkingen. De beschouwde casussen in dit proefschift suggereren dat besparingen van enkele miljoenen euro per kilometer mogelijk zijn (10-35% ten opzichte van de huidige dijkversterkingskosten) voor de Nederlandse situatie, met typisch relatief hoge kosten voor dijkversterkingen ten opzichte van de kosten en risico's voor het inwinnen van gedragsinformatie. Het gebruiken van gedragsinformatie draagt dus bij aan het verbeteren van de efficiëntie van de beheersing van overstromingsrisico's in Nederland, en in het bijzonder de dijkversterkingen in het Hoogwaterbeschermingsprogramma.

Dit onderzoek is uitgevoerd als onderdeel van het All-Risk programma dat het Hoogwaterbeschermingsprogramma ondersteunt bij de implementatie van de nieuwe veiligheidsnormen voor waterveiligheid. Het programma, gericht op de ontwikkeling van nieuwe methoden om investeringen voor waterveiligheid efficiënter te maken, is grotendeels gefinancierd door NWO, de Nederlandse Organisatie voor Wetenschappelijk onderzoek, en ondersteund met bijdragen van een groot aantal eindgebruikers.

ACKNOWLEDGEMENTS

I started with this research project in September 2017. At that time the four years ahead looked endless to explore everything I could imagine. I had never been a must to do a PhD, though I always liked to figure out new things, preferably beyond the existing standards and guidelines. At Deltares I had already been working for three years on such projects. For example the calibration of partial factors for inner slope stability for the statutory safety assessment (WBI), and developing a method for reliability updating for slope stability. When the possibility arose to further develop reliability updating within the research programme All-Risk, starting a PhD was a once-in-a-lifetime opportunity to not let go.

The four years have flown by, and I can agree with all the clichés about doing PhD research. The unparalleled freedom to determine what to research and how to go about it. To decide yourself on what and how to develop both personally and professionally. To work on your own topic for four years eventually getting through and completing it. What certainly helped me to not lose interest and motivation for the subject are the first years of practical experience at Deltares. Here I discovered my interest in probabilistic methods and its potential for the field. Being able to link research to practical issues has been a success factor for keeping my interest in the subject. I am thankful that Deltares has been willingly to let me go for a while, even though it meant I could only work for Deltares projects one day a week.

Doing and managing the research project is a very individual effort, yet it is not possible to do it alone. Many people have contributed implicitly and explicitly to the realisation of this work. A few of them I would like to mention specifically. First of all, I would like to thank Matthijs and Timo for being my promotor and co-promotor. Timo, as my daily supervisor you have played a major role in the successful completion of this project with your attention and care. I find it admirable that you are able to understand and interpret results so promptly and sharply. Your insights and sharp questions about my results and texts have helped me enormously in giving direction to my research. I am very grateful for your contributions to the thesis, and that you monitored the process to keep it going, even in extremely busy times. Matthijs, thank you for seeing me as a suitable candidate for this position and for allowing me the freedom to develop in this area. You have the gift of always asking questions from different perspectives, which has challenged me to place my own solutions in the social perspective. Your encouragement to always look for connections with (PhD) colleagues in other disciplines has also contributed to this.

My thanks go also to the many fellow researchers in All-Risk, especially those in Delft in room 3.58. We had a great time together! With many of you I have taken up a project or supervised students. Thank you for taking me along in your research, and for your contributions to my project. I enjoyed establishing new connections and applications this way. In particular, I would like to mention the collaborations on joint articles. Wouter Jan, the

paper on decision analysis of monitoring and proof loading. Teun, the paper on linking uncertainty of the shape of the flood wave shape to time-dependent reliability analysis. Joost, Vera and Guido, our discussions and paper about the interactions between failure mechanisms in determining the probability of *flooding*.

I would also like to thank everyone who has shown interest in my project and its progress, colleagues and friends, among others at the TU and Deltares, but also outside. Thanks also to everyone with whom I could find a bit of distraction, e.g. in the form of either a coffee, beer, game of AoE, and in establishing our home-brewery. My thanks go to my parents, who gave me every opportunity to study without any worries. You have taught me the importance of education and training from a young age. And lastly, Jessica, your boundless love, unconditional support, and all imaginable help deserve much more than just a mention in this book.

Mark van der Krogt
Zoeterwoude, March 2022

CONTENTS

1

INTRODUCTION

An instrument too often overlooked in our technical world is a human eye connected to the brain of an intelligent human being.

Ralph Peck

Dikes play a pivotal role in the protection of people and goods against flooding in flood prone countries such as the Netherlands. One of the ways how dikes fail, is through instability of the inner slope. Estimates of the probability of slope instability of dikes involve high uncertainty. Those estimates can be improved by considering additional information contained in observations of the performance of dikes. This chapter introduces the research topic and its relevance, and it provides an overview of the research questions and main contributions of this dissertation.

1

1.1. RELEVANCE OF FLOOD PROTECTION IN THE NETHERLANDS

Around 60% of the Netherlands is vulnerable to flooding. A major part of the country is below the mean sea level, and other parts of the country are liable to floods from high river discharges (ENW, 2017), see Figure 1.1. A large-scale flood has disastrous consequences for the Netherlands, as it disrupts large parts of the society and leads to considerable damage to buildings, roads, and the environment. Besides the direct damage, flooding also negatively affects economic welfare because approximately 70% of the Gross Domestic Product is generated in these areas (MinIE, 2015). Moreover, a flooding results in many casualties and loss of lives because over 9 million people live in flood prone areas in the Netherlands.

A sufficiently safe and reliable flood protection system is thus pivotal for the Netherlands. The flood defence system is assessed for compliance with the required safety level at least once in 12 years. As of 2017, new safety standards have been defined. Contrary to the former safety standards which were based on the exceedance frequency of hydraulic loads, the current safety standards are defined in terms of acceptable probabilities of flooding, based on different risk acceptance criteria, namely individual risk, economic risk, and societal risk, see (ENW, 2017; Jonkman et al., 2018).

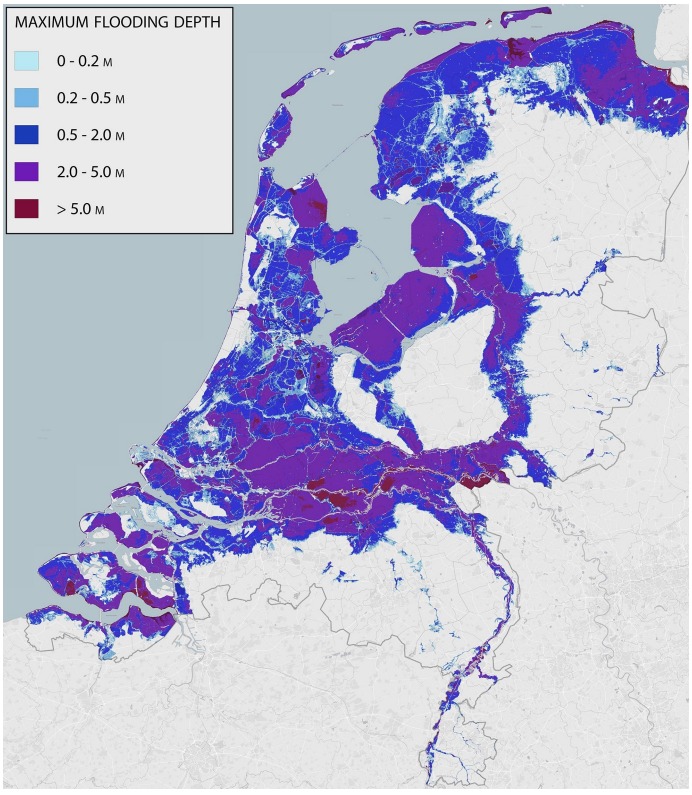


Figure 1.1: Map of the Netherlands showing the maximum flooding depth. Data source: Climate Impact Atlas.

By 2050, all flood defences must comply with the new safety standards. That means, the failure probability ought to comply with the acceptable failure probability requirement. With over 3000 km in length, earthen dikes form the largest part of the Dutch primary flood protection system. The system further consists of dunes and numerous other structures, such as flood gates and storm surge barriers.

Anno 2021, a large share of the dikes does not comply with the safety standards, see Figure 1.2. Mainly dikes sections in the river area have an estimated probability of failure that is several times higher than the acceptable probability. Therefore, more than 1600 km of dikes need reinforcement until 2050 (HWBP, 2020). The Dutch Flood Protection Program (in Dutch: Hoogwaterbeschermingsprogramma (HWBP)) coordinates these dike reinforcement projects. Especially in densely built areas, space for dike heightening and widening is limited, making dike reinforcement projects complex. Consequently, the costs are high and there is a continuous search for opportunities to make flood protection more efficient. The All-Risk research project (of which this research is part of) supports the HWBP with the implementation of the new safety standards by developing knowledge and enhanced methods which ought to contribute to a more efficient flood risk management.

Deterministic and semi-probabilistic methods can be used to estimate failure probabilities, but are typically based on conservative assumptions for the uncertain parameters, and can hence lead to conservative estimates of the safety. Full-probabilistic methods, on the other hand, provide less biased estimates because all uncertainty can be

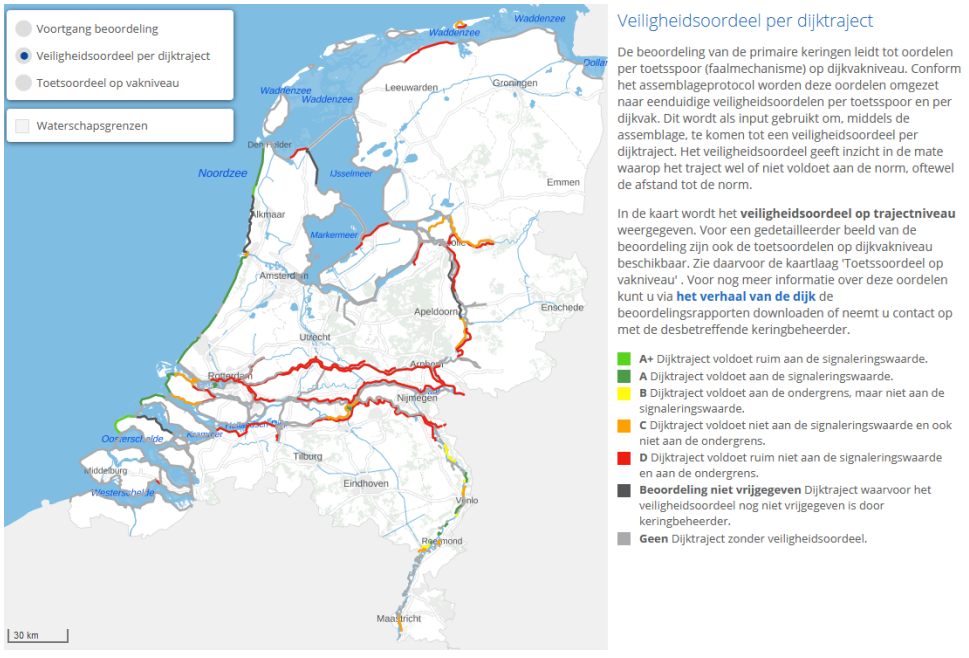


Figure 1.2: Safety assessment per dike segment, status of October 2021. Green and yellow dike segments comply with the safety standards, orange and red do not. Source: Waterveiligheidsportaal (2021)

considered explicitly. Probability estimates can be improved by including additional information using Bayesian inference. Improved probability estimates result in enhanced safety assessments of existing dikes and dike reinforcements.

One of the most important failure modes of earthen flood defences is instability of the inner slope: the mass-movement of a soil wedge along a sliding plane at the landside slope of a dike, see Figure 1.3. When water levels rise, pore water pressures in and under the dike increase, decreasing effective stresses and shear strength. The driving moment of the soil mass can cause the exceedance of the maximum shear resistance along a sliding plane, causing a large soil mass to slide off. After slope instability, dikes may be damaged to such an extent that it cannot withstand the flood water level, for example when the crest height is decreased or when the vulnerable dike core material is exposed. Under conditions of high water levels and wind waves, wave overtopping and subsequent instabilities can lead to a dike breach ('t Hart et al., 2016). All these processes must be taken into account to estimate the probability of flooding, for example using event trees, see among others Calle et al. (2003); Kanning et al. (2019); Van der Krogt et al. (2019); Van der Meer et al. (2019). As slope instability is an important initiating mechanism of dike failure, this dissertation focuses on the probability of occurrence of slope instability. Throughout this dissertation, 'failure probability' refers to the probability of occurrence of this geotechnical failure mode, and not the probability of flooding. Also, the term 'reliability' is often used in this dissertation, which is a decreasing function of the failure probability.

Figure 1.3: Instability of the inner slope of the Westfriese Zeedijk near Sint Maarten in Noord Holland on 15 September 1994 (source: Deltares Visuals)

1.2. OBJECTIVE AND RESEARCH QUESTIONS

Estimates of the reliability of slope stability are crucial for safety assessments of dikes and for the verification of reinforcement designs. For reliability analyses it is essential to employ probability theory, which requires quantification of the uncertainties involved, e.g. by means of probability distributions. For the stability of dike slopes on soft soils, the undrained shear strength of soft soil layers underneath is one of the most important input parameters. The knowledge about this parameter is usually based on prior knowledge from design codes (such as Eurocode 7 (CEN, 2004) or ISO2394 (Phoon and Retief, 2016)) or data from local soil investigation.

As slope stability involves a volume of soil, cone penetration tests (CPTs) are often used to estimate the undrained average shear strength because the measurements are virtually continuous over depth. To estimate undrained shear strength from cone resistance, a transformation model is needed, for example site-specifically calibrated with shear tests and CPTs in close proximity. In that case, various error sources (due to spatial variability, measurement errors and statistical uncertainty) propagate through the transformation model into the total uncertainty of the estimated spatial average undrained shear strength (e.g. Cao et al. (2017); Phoon and Kulhawy (1999a,b)). For adequate estimates of the total uncertainty of the spatial average, it is important to account for the different error sources appropriately.

From the Bayesian view on the concept of probability, probability reflects the degree of belief about a state of nature depending on the available information as elaborated by Baecher (2017). In the contexts of flood risk, an annual failure probability of a dike section is not the frequency of failure, but the chance that failure occurs at the considered section in a given year. The Bayesian view on probability that is adhered to in this work also implies that the estimates are no 'true' or 'fixed' values, but change with new information.

Large uncertainty can lead to high failure probabilities for slope stability (compared to the safety standards). Unlike aleatory uncertainty (actual randomness), epistemic uncertainty (due to a lack of knowledge) can be reduced with additional information (e.g., Baecher and Christian (2005); Straub and Papaioannou (2015); Papaioannou and Straub (2012). Additional soil investigation can reduce uncertainty, but the reduction in uncertainty is bounded by the amount of soil investigation that is practically feasible. Observations of historical or past performance can lead to significant uncertainty reduction and improved estimates of the failure probability. The degree to which epistemic uncertainty can reduce, depends on how informative the observation is.

Observations that relate to the performance of slope stability of dikes are the survival of loading conditions. Survival observations can be used to improve the estimates of failure probability for future conditions by incorporating the information into the reliability analysis through Bayes Theorem (Bayes, 1763). For dikes, it was shown that incorporating survival information in the reliability analysis may lead to lower failure probabilities, especially if a critical load was survived, see Zhang et al. (2011); Schweckendiek et al. (2014); Li et al. (2015). The construction phase is one such critical load condition, because raising the embankment increases the driving force, and excess pore water pressures result in lower strength. So we want to know how to incorporate 'construction survival' in the reliability estimates, and in which cases the information leads

to a significant update of the reliability.

Also other observations are available during the construction, for example settlement measurements and pore pressure monitoring. Incorporating information from multiple observations reduces uncertainty further compared to single sources of information. At the same time, the computational cost to estimate the (updated) probability of failure increases with the amount of information considered. Especially when the model runs require a few seconds or more, solving this reliability updating problem becomes very challenging, highlighting the need for efficient algorithms.

In contrast to conventional sources of information (e.g. site investigation), performance information is a direct observation of the performance (the stability of the slope), and therefore, it should be particularly valuable. Performance observations such as historically survived water levels are often readily available, or can be obtained by for example proof loading. Yet, proof loading and pore pressure monitoring are costly measures, so the cost-effectiveness needs to be evaluated. Bayesian Decision Theory (Raiffa and Schlaifer, 1961) can be used to determine the Value of Information of performance information, and to determine the optimal strategy of investing in uncertainty reduction and reinforcing dikes, leading to the lowest Total Cost.

In general, reducing uncertainty with readily available or cheap to obtain performance observations, can lead to more efficient flood risk management. Though the concepts and techniques of (pre-)posterior analysis and Bayesian decision analysis are available, the effect and cost-effectiveness of considering survival information during the construction phase and multiple sources of performance information are not examined so far. The objective of this dissertation is thus to improve probabilistic safety assessments of dike slope stability by considering performance information. The main research question and sub-questions are defined as follows:

How does performance information affect reliability estimates for dike slope stability?

1. To what extent is the uncertainty in the undrained shear strength reducible?
2. How does the performance information contained in the survival of the construction phase affect the reliability estimates for slope stability of dikes?
3. What is the effect of combining survival and settlements (multiple performance observations) during the construction of dikes?
4. How cost-effective are proof loading and pore water pressure monitoring in dike reinforcements?

1.3. OUTLINE

Each chapter of this dissertation is centred around one research question, each focusing on a specific aspect of improving reliability estimates of slope stability of dikes, see Figure 1.4.

Chapter 2 provides the relevant background about slope reliability analysis, Bayesian analysis, and decision analysis. Chapter 3 focuses on establishing the priors for the uncertainty in the undrained shear strength when estimated from CPTs and a site-specific

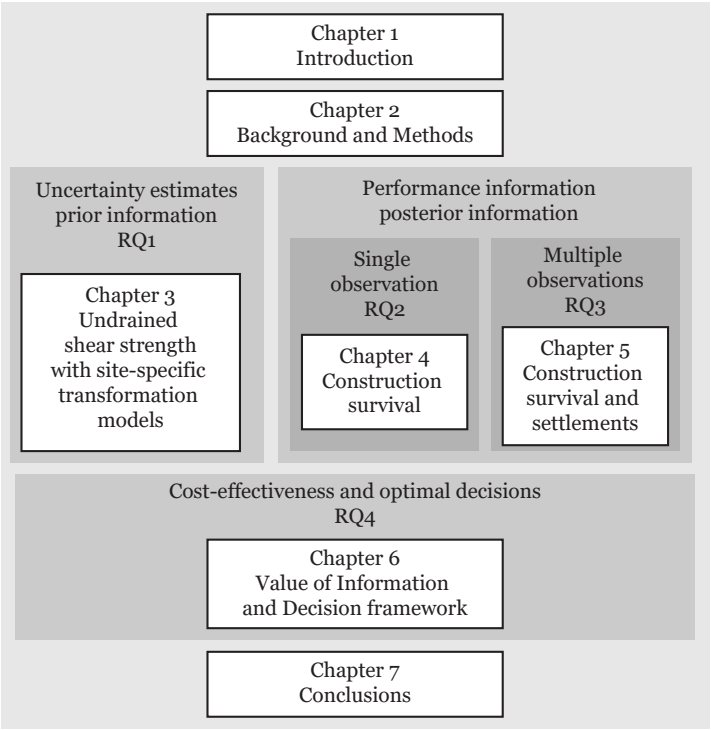


Figure 1.4: Structure of this dissertation

transformation model. The aim is to show to what degree uncertainty in undrained shear strength is epistemic, and hence, reducible. A method is proposed to estimate the uncertainty, particularly concentrating on the role of averaging of all spatially variable error components.

The two next chapters are directed towards improving reliability estimates of dikes by incorporating the information contained in performance observations. Chapter 4 specifically looks at how to use the observation of 'construction survival'. The proposed Bayesian Updating approach is demonstrated for several case studies, providing insight in the conditions influencing the degree to which the reliability changes with new information.

Chapter 5 elaborates on the effect of incorporating survival information and settlement measurements during the construction phase to explore the effect of using multiple performance observations on the reliability analysis. As the analysis is performed with computationally expensive models, different approximations are used to make the analysis computationally feasible so that it can be applied in practice too.

Finally, Chapter 6 proposes a framework to evaluate the Value of Information that is obtained by proof loading and/or pore water pressure monitoring. Using this framework, the cost-effectiveness of proof loading and monitoring for dike reinforcements is evaluated.

1.4. ORIGINALITY

The core of the contribution of this work lies in the use of performance observations for improving reliability estimates for slope stability of dikes. The main novelty is in the use of observations that have not been considered to date, such as construction survival, settlement measurements, and the combination of multiple performance observations. Furthermore, advances have been made in improving prior uncertainty estimates, and in determining whether uncertainty reduction is efficient for flood risk reduction.

The novelty of Chapter 3 is in the extensive analysis of how the total uncertainty of prior estimates is composed of various errors, when using a site-specific transformation model to estimate soil properties with indirect measurements such as CPTs. Specifically, the degree to which averaging of various random and systematic errors plays a role in estimating the total uncertainty of spatial average parameters had not been addressed conclusively in literature as now also acknowledged by Van den Eijnden and Hicks (2019) and Ching et al. (2020).

The criticality of slope stability during the construction of embankments is widely known, but the impact of considering this survival information on reliability estimates had never been demonstrated quantitatively for dike slope stability. Chapter 4 extensively investigates this source of information and the effect of reliability updating with different design flood water levels, and for different standard dike types.

The novelty of Chapter 5 is in the combination of multiple performance observations (slope stability survival and settlement) during the construction of dikes, using geotechnical models with the same level of complexity and computational expense as in practice. Though the application of nested simulation methods for solving the Bayesian inference problem is quite common in cosmology and particle physics (Feroz et al., 2009), their application in reliability analysis of civil engineering problems is new.

The application of decision theory and pre-posterior analysis in Chapter 6 has led to original insights in the role that proof loading and pore pressure monitoring can play in dike improvement projects in terms of benefits, costs and risks. Also the Value of Information analyses on using both these measures in combination, and how decisions for one measure influence the VoI of another were novel.

The various fictitious but realistic cases throughout this dissertation provide insights into the parameters and conditions determining the impact of reliability updating. The case study of the Eemdijk test dike demonstrates that the methods can also be applied to real conditions, as encountered in dike safety assessments and reinforcement projects.

1.5. ALL-RISK PROGRAMME

This research is part of the All-Risk programme which supports the Dutch Flood Protection Programme with implementing the risk-based safety standards for flood protection. The research aims to develop new methods for improving the efficiency of investments for flood protection. The All-Risk programme is carried out by five different universities in the Netherlands in cooperation with a large network of over 30 national and international partners consisting of government agencies, research institutes and companies. The programme is largely funded by NWO, the Dutch Research Council, and supported by contributions of numerous end-users.

2

BACKGROUND AND METHODS

Probability theory is nothing but common sense reduced to calculation.

Pierre-Simon Laplace

This dissertation aims at improving reliability estimates for slope stability of flood defences by reducing uncertainty with performance information. This chapter provides the relevant background of uncertainty, Bayesian reliability updating, and Bayesian decision making. Several overarching topics are introduced such as safety assessment of flood defences in the Netherlands, prior and posterior analysis, reliability methods, limit state formulations for slope instability, and the framework for Bayesian decision analysis.

Contents

2.1. SAFETY ASSESSMENT OF FLOOD DEFENCES

The Water Act (Waterwet, 2021) sets out the new flood protection standards which are based on acceptable risk (individual, economic, and societal). The acceptable flood probability offers a basic level of protection based on Local Individual Risk (in Dutch: Lokaal individueel Risico (LIR)), and a higher level of safety where economic and societal costs of flooding are relatively high. The acceptable annual probabilities of flooding are sometimes referred to as the 'requirement' or the 'target probability', denoted by P_{T}, or in terms of the reliability index β_T, see Equation 2.1, where Φ^{-1} is the inverse of the standard normal distribution.

$$\beta_{\mathrm{T}} = \Phi^{-1}(1 - P_{\mathrm{T}}). \tag{2.1}$$

Flood defences are large systems with multiple sections, and can fail due to different failure modes. The failure probability of a dike segment is therefore calculated per failure mechanism and per dike section, and afterwards combined into a system probability. The failure probability is commonly estimated for cross-sections of approximately 50m wide, and are representative for a section of roughly 100-1000 m long, see Figure 2.1.

Where flood defences in the Netherlands do not comply with the safety standards, dikes need reinforcement. It is, however, difficult and time consuming to find an optimal design given the local boundary conditions of each dike section, meeting the required reliability on a system level, though a first proof of concept is presented in Klerk et al. (2021)). Therefore, a simplified approach is to break down the maximum failure probability for the (series) system into requirements for independent elements, such as dike sections and failure modes, see e.g. Jongejan et al. (2020):

$$P_{\mathrm{T,section,mode}} = \frac{\omega \cdot P_{\mathrm{T}}}{N_{\mathrm{length}}}. \tag{2.2}$$

Here ω defines the fraction assigned to an independent failure mode, and N_{length} the number of equivalent independent dike sections to account for the length-effect, see Kanning (2012); Schweckendiek et al. (2013); Jongejan et al. (2020). N_{length} depends on

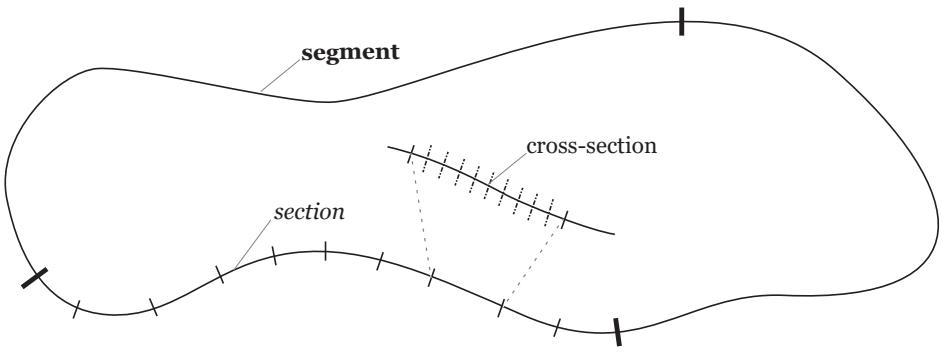


Figure 2.1: Definitions of cross-section, section, and segment in a dike system.

a, the fraction of the length of a dike segment susceptible to the failure mode, and b the length of one independent equivalent dike section, see Equation 2.3. For slope stability of dikes the default starting point is $\omega = 0.04$, $a = 1/30$, and $b = 50$ m, see TAW (1989); WBI (2013); Jongejan (2017). A target reliability for slope stability at a section level is the starting point in this dissertation.

$$N_{\text{length}} = 1 + a \cdot L_{\text{segment}} / b. \tag{2.3}$$

2.2. RELIABILITY ANALYSIS

2.2.1. PRIOR PROBABILITY OF FAILURE

In reliability analysis we estimate the probability of the event of failure F. When the function $g(\mathbf{x}) < 0$ describes the performance of the system (based on an outcome (realisation) of a stochastic input $\mathbf{x}$), the event of failure is defined as $\mathrm{F} = \{g(\mathbf{x}) < 0\}$. The failure probability P_f is calculated as follows:

$$P_f = P(\mathrm{F}) = \int_{g(\mathbf{x})<0} f_{\mathrm{X}}(\mathbf{x})\, d\mathbf{x}, \tag{2.4}$$

where f_{X} is the probability density function (pdf), $\mathbf{X}$ a vector of random variables, and $\mathbf{x}$ a realisation.

2.2.2. RELIABILITY METHODS

Equation 2.4 can be solved by numerical integration, but other methods and algorithms are more efficient to estimate $P(\mathrm{F})$. This section states common and popular methods used in this dissertation, not to give a full overview of possible methods. A comprehensive overview can be found for example in Melchers and Beck (2018); Lemaire et al. (2009).

Crude Monte Carlo (CMC) is a simulation method based on the Laplace formulation $P(\mathrm{F}) = n(\mathrm{F})/N$ (Laplace and Truscott, 1951). For $N \to \infty$, the method is exact. An implementation to estimate the failure probability with Monte Carlo using the indicator function $1[\cdot]$ is:

$$\hat{P}_f = \frac{1}{N} \sum_{i=1}^{N} 1\left[g(\mathbf{x}_i) < 0\right]. \tag{2.5}$$

The $\hat{}$ denotes an estimate for the failure probability. The relative error of $\hat{P}_f$ is estimated using Equation 2.6. It follows that the relative error in the probability is less than 5% if the number of failing realisations is larger than 400, for an example.

$$V_{\hat{P}_f} = \frac{\sqrt{\mathrm{var}\left[\hat{P}_f\right]}}{E\left[\hat{P}_f\right]} = \frac{\sqrt{\hat{P}_f\left(1-\hat{P}_f\right)/N}}{\hat{P}_f} \approx \frac{1}{\sqrt{N \cdot \hat{P}_f}}. \tag{2.6}$$

Plenty of variants on CMC are available, mostly to reduce the required sample size to estimate the probability with the same accuracy, for example Monte Carlo Importance Sampling (MCIS), Directional Sampling, or adaptive versions (Grooteman, 2011;

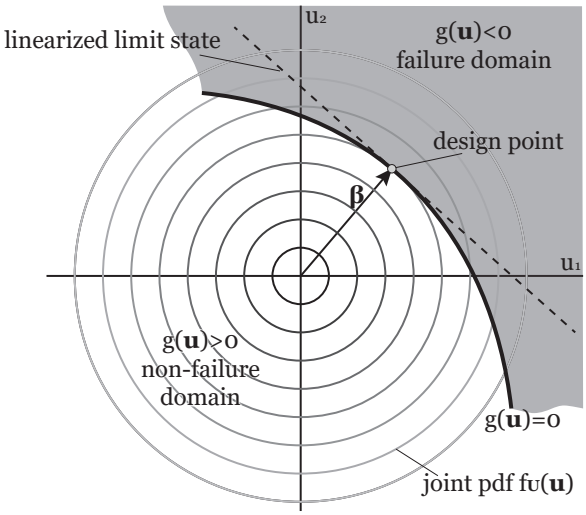


Figure 2.2: Definitions of FORM for two dimensions in standardized normal space

Waarts, 2000). MCIS is implemented similar to Monte Carlo, but using weight factors, $w_i = f(\mathbf{X})/f_*(\mathbf{X})$ to correct for a different sampling pdf f_* instead of f in Chapter 4.

Another noteworthy reliability method is subset simulation (Au and Beck, 2001), which has been demonstrated as one of the most efficient and robust reliability methods, for reliability problems with large dimensions, low failure probabilities, and Bayesian Updating, for example in Papaioannou et al. (2015); Jiang et al. (2018); Betz et al. (2018); Van den Eijnden and Hicks (2017); Ahmed and Soubra (2012); Hsu and Ching (2010); Au et al. (2007). In subset simulation, the failure probability $P(\mathrm{F})$ is calculated by the product of the conditional failure probabilities of multiple intermediate failure events (Au and Beck, 2001), see Equation 2.7.

$$P(\mathrm{F}) = P(\mathrm{F}_1)\prod_{i=2}^{m} P(\mathrm{F}_k|\mathrm{F}_{k-1}). \tag{2.7}$$

The failure event F is denoted by a sequence of intermediate events (subsets) satisfying $\mathrm{F}_1 \supset \mathrm{F}_2 \supset \ldots \supset \mathrm{F}_k$. Markov chains (using the MetropolisHastings algorithm) can be used to efficiently simulate the conditional samples in each subset (Metropolis et al., 1953; Hastings, 1970; Papaioannou et al., 2015).

The First Order Reliability Method (FORM) is a reliability method based on a first-order approximation of the limit state $g(\mathbf{X}) = 0$ in the standardized normal space, i.e. an equivalent random variable space consisting independent standard normal random variables U. According to Hasofer and Lind (1974), the reliability index β is equal to the shortest distance from the origin to the surface described by $g(\mathbf{U}) = 0$, see Figure 2.2. The point on the limit state closest to the origin is called the design point.

Fragility curves are a convenient way to calculate the slope reliability of dikes, see Schweckendiek et al. (2017). Fragility curves describe the conditional failure probability given a (load) variable, for example the water level h:

$$P(\mathrm{F}|h) = P(g(\mathbf{X}, h) < 0). \tag{2.8}$$

Here $\mathbf{X}$ is the vector of random variables except for h. The annual probability of failure is obtained by combining $P(\mathrm{F}|h)$ with the pdf of the annual maxima of the load h as follows:

$$P(\mathrm{F}) = \int P(\mathrm{F}|h) f(h) dh. \tag{2.9}$$

Fragility curves can, in principle, be made for any (load) variable. This approach is followed in Section 6.3.2.

2.2.3. POSTERIOR PROBABILITY OF FAILURE

Posterior reliability analysis is the procedure to update prior reliability estimates, based on Bayes' Rule (Bayes, 1763) and the formulation of conditional probability. Zhang et al. (2011) distinguish two methods to incorporate additional information using Bayesian Updating: the indirect method and the direct method, which are discussed below.

DIRECT METHOD FOR RELIABILITY UPDATING

The direct method is a direct application of the Bayes theorem, with F the failure event, and ε the event of the observation, see Equation 2.10.

$$P(\mathrm{F}|\varepsilon) = \frac{P(\mathrm{F} \cap \varepsilon)}{P(\varepsilon)}, \tag{2.10}$$

The observation event is expressed in terms of an observation function $h(\mathbf{x})$. Depending on the type of information, the observation is described as follows:

- inequality information; when the observation has information whether or not a limit state is exceeded, such as survival information: $h(\mathbf{x}) < 0$;
- equality information, when the observation has information about the measured performance value, such as pore pressure or settlement measurements: $h(\mathbf{x}) = 0$.

For inequality information, we write Equation 2.10 as follows:

$$P(\mathrm{F}|\varepsilon) = \frac{P(\mathrm{F} \cap \varepsilon)}{P(\varepsilon)} = \frac{P\left(g(\mathbf{X}) < 0 \cap h(\mathbf{X}) < 0\right)}{P(h(\mathbf{X}) < 0)}. \tag{2.11}$$

For equality information, the probability that $h(\mathbf{X}) = 0$ is equal to 0, which makes it impossible to evaluate. Therefore, the equality information can be transformed into inequality information (Straub, 2011; Straub and Papaioannou, 2014), so the reliability updating problem can be evaluated as a standard reliability problem. Here the observation function is:

$$h(\mathbf{x}, u) = u - \Phi^{-1}(c \cdot L(\mathbf{x})), \tag{2.12}$$

where u is the outcome of a standard normally distributed random variable, and $L(\mathbf{x})$ the likelihood of the observation ε: $P(\varepsilon|\mathbf{x})$. The scaling constant c ensures that $0 \leq c \cdot L(\mathbf{x}) \leq 1$. The Φ^{-1} is the inverse cumulative distribution function of the standard normal distribution. An alternative form of Equation 2.12 is

$$h(\mathbf{x}, \pi) = \ln(\pi) + \hat{l} - \ln(L(\mathbf{x})), \tag{2.13}$$

in which the logarithm of the observation function is taken. In this notation u is replaced by a standard uniformly distributed random variable π, and $\ln(c^{-1})$ is denoted by $\hat{l}$. In this formulation the limit state has a more appropriate shape in evaluating $P(h(\mathbf{x}) < 0)$, leading a smoother transition of the failure domains in an implementation with subset simulation (Betz et al., 2018).

The observation event can also describe multiple independent observations: $\varepsilon = \varepsilon_1 \cap \varepsilon_2 \cap ....$ In this case, $h(\mathbf{X}) < 0$ should be read as $\bigcap^{n} h_i < 0$ with h_i the individual limit state belonging to the observations ε_i. For equality information, multiple observations can also be aggregated by multiplying the likelihoods of those observations. $L(\mathbf{x})$ should then be read as $\prod^{n} L_i(\mathbf{x})$.

INDIRECT METHOD FOR RELIABILITY UPDATING

In the indirect method, the parameter distribution $f_{\mathbf{X}}$ is first updated to $f_{\mathbf{X}|\varepsilon}$, using the likelihood of the observation ε: $L(\mathbf{x}) = P(\varepsilon|\mathbf{x})$, see Equation 2.14.

$$f_{\mathbf{X}|\varepsilon}(\mathbf{x}) = \frac{L(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x})}{\int_{\mathbf{X}} L(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}}. \tag{2.14}$$

Thereafter the probability is recalculated using the updated distribution, using Equation 2.15.

$$P(\mathrm{F}|\varepsilon) = \int_{g(\mathbf{x})<0} f_{\mathbf{X}|\varepsilon}(\mathbf{x}) d\mathbf{x}. \tag{2.15}$$

The indirect method is generally favourable when information relates to best estimates, and parameter distributions can be easily reconstructed, for example, through updated values of the distribution parameters (e.g. Jiang et al. (2020)). For survival observations (which provide information about inequalities), the likelihood function takes the value 1 if a realisation is in accordance with the observation, and else 0, as formulated in Straub and Papaioannou (2015).

2.2.4. NESTED SAMPLING

Equation 2.11 can be evaluated using standard reliability methods. A method that is specifically suited for Bayesian inference is nested sampling. Nested sampling techniques (e.g. the MultiNest algorithm) are aimed at calculating $\int_{\mathbf{X}} L(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}$ and the posterior distribution $f_{\mathbf{X}|\varepsilon}$, and promise to be efficient when the total evidence (the integral) is low, posterior distributions are multimodal, and in high dimensions.

The basis for the MultiNest algorithm (Feroz et al., 2009) is nested sampling (Skilling, 2004, 2006). With nested sampling the evidence $P(\varepsilon) = \int L(\mathbf{x}) dY$ is evaluated by integration over $dY = \pi(\mathbf{x}) d\mathbf{x}$, the element of prior mass (π) and $\mathbf{x}$ a realisation of the unknown parameters. Notice that the function π is different from the stochastic realisation π in Equation 2.13. The simplest form to estimate the evidence is to sort the likelihood values for the realisations $\mathbf{x}$ and integrate the likelihood values over Y, see Figure 2.3.

The basic routine of nested sampling starts with an initial set of realisations (so-called live points) and iteratively stores the realisation with the lowest likelihood. The sample is replaced with a new sample $Y_i < Y_{i-1}$, where $L_i > L_{i-1}$. By sampling with the prior density $\pi(\mathbf{x})$ constrained with $L_i > L_{i-1}$, the new sample automatically meets $Y_i < Y_{i-1}$, avoiding the sorting of Y. New samples therefore follow the shape of likelihood

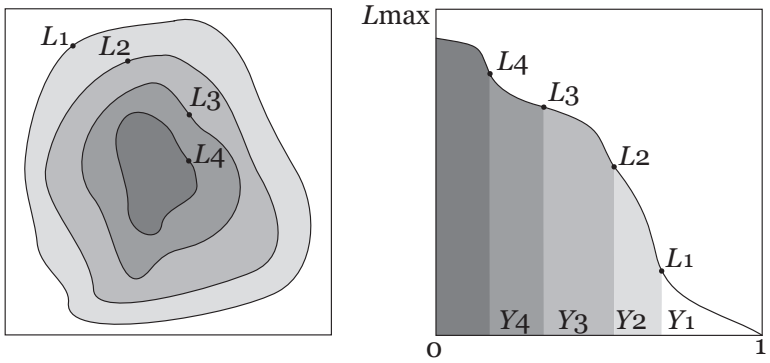


Figure 2.3: Left: Nested likelihood contours sorted to enclosed prior mass Y. Right: integration of the evidence. Source: Feroz et al. (2019).

contours, see Figure 2.3. The MultiNest algorithm improves the sampling of new points, through an ellipsoidal rejection sampling scheme. Nested Sampling and the implementation with rejection sampling is further explained in Skilling (2004, 2006), and MultiNest in Feroz et al. (2009).

One disadvantage of the nested sampling methods is that there is no rigorous stopping criterion, and that the stopping criterion does not relate to the probability of failure. The stopping criterion of MultiNest defines the maximum estimated remaining evidence after iteration i, estimated by $Z_{\text{est}} = L_{\max} \cdot Y_i$. The algorithm stops with adding points when the expected evidence contribution from the current set of live points is less than a user-defined tolerance: $\log(Z_i + Z_{\text{est}}) - \log(Z_i) < \delta_{\text{logz}}$.

2.2.5. METAMODELING

Metamodels (also known as surrogate models) predict the outcome of a model, based on only a limited number of evaluations of the actual model, reducing the computational time. Therefore, metamodels can improve the efficiency of reliability and Bayesian analysis with computationally expensive models. Examples of metamodelling approaches are (polynomial) response surface methods, Kriging-based methods, or Machine Learning-based methods. Teixeira et al. (2021) gives a recent overview of different methods, and Li et al. (2016) gives an extensive literature study of application of response surfaces for slope reliability.

An efficient metamodelling method for estimating failure probabilities is Active learning combined with Kriging and Monte Carlo Simulation (AK-MCS) (Echard et al., 2011). One of the main advantages of AK-MCS is that next added points are targeted around the limit state, based on the uncertainty estimates based on Kriging and the learning function. Eijnden et al. (2021) presented a Kriging-based metamodelling approach especially targeted at (target) reliabilities in the range of $3 \lesssim \beta \lesssim 6$, models with 10 to 20 variables, and possibly a noisy response, see Figure 2.4.

The Kriging model (Gaussian Process regression) predicts the response of the limit state function g by a mean function m and a covariance function K (the kernel), such

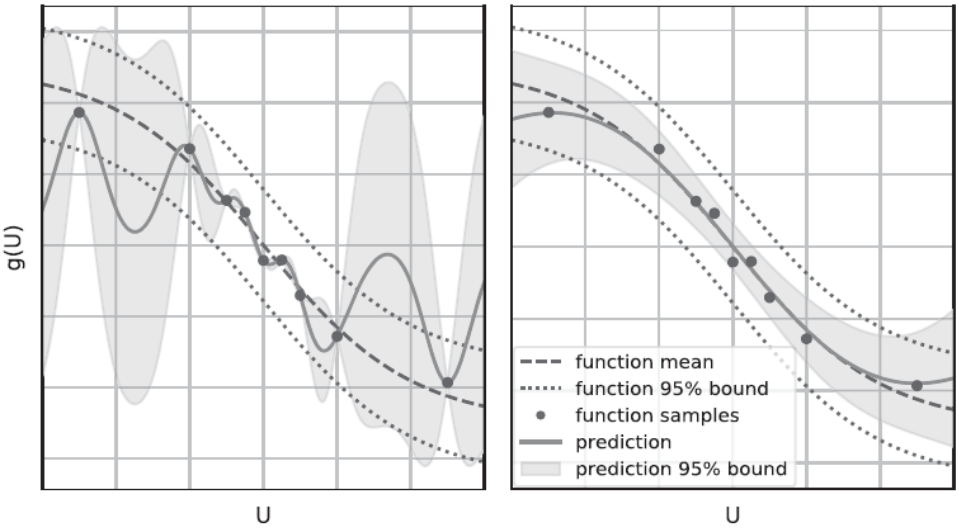


Figure 2.4: Example of a Kriging metamodel for prediction of the limit state outcome $g(U)$. The left pane shows an example of an overfitted model, the right pane shows an example where overfitting is prevented by introducing a noise term in the kernel. Source: Eijnden et al. (2021).

that the prediction $\hat{g}(\mathbf{u}) = m(\mathbf{u}) + \mathbf{K}^{1/2}\vec{\xi}$. Herein, $\mathbf{K} = k(\mathbf{u}, \mathbf{u}|\vec{\theta})$, is the kernel matrix with $\vec{\theta}$ internal shape parameters of the kernel, and $\vec{\xi}$ a standard normal multivariate.

The metamodel $\hat{g}(\mathbf{u})$ replaces the actual limit state function $g(\mathbf{x})$ in the evaluation of the failure probability using Equation 2.4. The formulation is in standard normal space with uncorrelated variables $\mathbf{U}$ by transforming the stochastic parameters $\mathbf{X}$ from (correlated) parameter space to standard normal space. Basically, one can use any simulation model to evaluate the failure probability in combination with a metamodel. The implementation of Eijnden et al. (2021) obtained efficient results using an adaptive multiple importance sampling. For the scientific background and details regarding the implementation, refer to Eijnden et al. (2021).

2.3. SLOPE RELIABILITY

2.3.1. SLOPE STABILITY MODELLING

Slope stability can be analysed using various methods such as analytical methods (e.g. Limit Equilibrium method (LEM)) or numerical methods (e.g. Finite Element Methods (FEM) (e.g. Griffiths and Lane (1999); Griffiths and Fenton (2004)), Finite Difference Methods, or Material Point Methods (Wang et al., 2016; Remmerswaal et al., 2021)). In this dissertation we use the Limit Equilibrium method because it is attractive in terms of computation time. Other advantages of the LEM method compared to more advanced models are its simplicity and robustness (Juang et al., 2018), the relatively extensive verification with case histories, and the LEM method is very common in engineering practice.

LEM analysis considers the equilibrium of a soil mass. The ratio of mobilized shear force and the available shear force along a potential slip plane (Fellenius, 1936) indicates

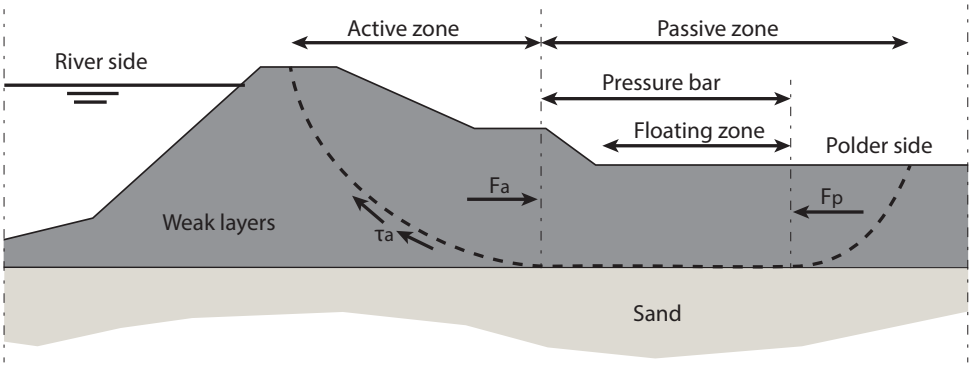


Figure 2.5: Uplift-Van slip plane (dashed) with an active and passive circular part and a horizontal pressure bar

the stability, expressed by a factor of safety (F_s). This dissertation considers the equilibrium in 2-D plain strain conditions, based on the method of slices (Morgenstern and Price, 1965). Many methods exist to analyse slope stability, see (Bishop, 1955; Spencer, 1967), and Van (2001) (Figure 2.5) among others. The methods mainly differ in the assumptions regarding the shape of the slip plane, the interslice forces, and the requirements concerning the force balances. In this dissertation, the D-Stability software with Bishop and Uplift-Van method is used for stability calculations with LEM. For the implementation, reference is made to Brinkman et al. (2018); Deltares (2019).

Soil strength can be modeled with different constitutive models, such as Mohr-Coulomb among others. When it comes to slope instability and ultimate limit states, the critical state concept suits well to a situation of continuous shear deformation along an entire slip plane (with large strains and no peak strength) according to the Critical State Soil Mechanics (CSSM) framework (Schofield and Wroth, 1968). In this concept, the shear strength for drained soils such as sand can be calculated using a critical state friction angle ϕ_{cs}. For undrained soils such as clay or peat, the undrained shear strength (s_u) can be modelled with the Stress History and Normalized Soil Engineering Properties (SHANSEP) formulation (Ladd and Foott, 1974), see Equation 2.16. The SHANSEP formulation is also well suited for stability analysis of staged construction.

$$s_u = \sigma'_v \cdot S \cdot OCR^m. \tag{2.16}$$

where, S = undrained strength ratio for normally consolidated soil, m = strength increase exponent, and OCR = over-consolidation ratio, being the ratio of in situ vertical effective stress σ'_v and preconsolidation stress $\sigma'_p = \sigma'_v + POP$, where POP = pre-overburden pressure.

2.3.2. PERFORMANCE FUNCTION

Since the factor of safety F_s indicates the ratio of resistance over load, instability occurs in a perfect model if $F_s < 1.0$. Contrary, a slope is stable if $F_s \geq 1.0$. To account for model error, F_s is multiplied with a model uncertainty factor for stability m_s, for which a realisation $m_s < 1.0$ indicates a model that over-predicts the stability. Accordingly, the

limit state function g for failure is defined as:

$$g(\mathbf{x}) = F_s \cdot m_s - 1. \tag{2.17}$$

Herein, $\mathbf{x}$ is a realisation of $\mathbf{X}$, with $\mathbf{X}$ the vector of all (random) variables. In this notation F_s is the stability model, which depends on all input parameters (i.e. soil parameters, pore pressures, etc.) of the model $\mathbf{x}$, except m_s. For survival information, generally the information is that $g(\mathbf{x}) \geq 0$, which can also be rewritten in an observation function h, for which survival information is $h(\mathbf{x}) < 0$.

$$h(\mathbf{x}) = 1 - F_s \cdot m_s. \tag{2.18}$$

2.4. UNCERTAINTY AND SPATIAL VARIABILITY

For slope stability the uncertainty is mainly in soil properties (Christian et al., 1994), which are uncertain because of spatial variability, measurement error, sparse data, and transformation error (e.g. Baecher and Christian (2005); Cao et al. (2017); Phoon and Kulhawy (1999a,b). A state of the art overview of the implications of spatial variability and uncertainty in geotechnical properties and models is given by ISSMGE-TC304 (2021). Furthermore, the flood water level and the accompanying decrease of soil shear strength due to higher pore pressures is uncertain, and also model uncertainty is significant in slope instability models (e.g. Zhang et al. (2014); Juang et al. (2018)).

Generally, uncertainty is divided in two categories: variability, and uncertainty originating from lack of knowledge (e.g. Van Gelder (2000)). Variability, often indicated as inherent uncertainty, or aleatory uncertainty, refers to randomness, like the outcome of a throw of a dice. Variability can be further split into variability in space and in time. Soil properties, for example, vary in space, but are usually constant in time. Contrary, the maximum water level in a year varies per year, but will not differ much from place to place. Another possibility is to classify uncertainty to where the uncertainty originates from, e.g. from model error, measurement error, or statistical errors, and to what degree the errors are random (scatter) or systematic (bias).

The above classifications are no fundamental properties of parameters, but a choice of how we model the uncertainty in our (reliability) problems (Kiureghian and Ditlevsen, 2009). Most parameters important for slope stability analysis are time-invariant soil properties, for which the properties are correlated between the moment of observation and the moment of assessment. This is denoted by the auto-correlation $\rho_k = 1$ for a fully correlated (time-invariant) parameter k. So, though the soil properties vary in space, the value at a specific location is mainly uncertain due to lack of knowledge, and hence it is reducible. Some stochastic properties, however, are uncorrelated in time, such as the annual maximum of the flood water level (which determine the pore pressures influencing the stability). For those time-variant properties, the auto-correlation $\rho_k = 0$ and realisations in time are stochastically independent.

Soil properties are variable in space, because of various processes during the formation process (e.g. Lumb (1966)). The impact of spatial variability on geotechnical engineering problems is widely recognised, see for example Fenton and Griffiths (2002); Griffiths et al. (2009); Hicks and Samy (2002); Cho (2007); Ahmed and Soubra (2012)).

Spatial variability can be modelled explicitly in slope stability using random fields with the statistical properties mean, variance and scale of fluctuation (Vanmarcke, 1983). For slope stability the total shear strength along a slip plane is the parameter of interest. For volumes of soil larger than the correlation distance, short-scale fluctuations will average. The spatial average can be approximated by applying variance reduction, e.g. the variance reduction factor Γ^2 (Vanmarcke, 1977, 1980).

In this dissertation the (local) average soil properties along a slip plane are approximated by modelling average values for statistically homogeneous soil layers (e.g. geological deposits) in a cross-section. The choice to approximate the average along a slip plane by a layer-average, assumes full averaging of all fluctuations, which is a reasonable assumption when the vertical scale of fluctuation is much smaller than the vertical dimension of the slip plane in the layer of interest, which is usually the case in slope stability problems (see the summary of reported scales of fluctuation in De Gast (2020)).

The above assumption ignores the effect that slip planes tend to be attracted to weaker zones, especially when the correlation distances of shear strength are in the same order as the volume of the mechanism considered (Ching et al., 2017; Zhu et al., 2019; Varkey et al., 2019). In slope stability of dikes, this effect mainly plays a role in the horizontal parts at the bottom of the slip plane, when it crosses a continuous weak zone of several meters. However, the influence of weak-zone seeking is typically relatively small when the total uncertainty is dominated by epistemic uncertainty such as transformation uncertainty, measurement uncertainty, and statistical uncertainty (Tabarroki et al., 2021), which is the case for the layer-average estimates in this dissertation.

The horizontal scale of fluctuation in the direction along a dike (or line-infrastructures) is typically tens of meters (Van Duinen, 2019), which implies that independent sections or reaches are in the order of fifty to hundred meters. The result of the a-priori reliability analyses based on local cone penetration tests or regional mean and variability of spatial averages, are thus representative for sections in the order of fifty to hundred meters. Notice that settlement information such as used in Section 5.3 for reliability updating may be representative of shorter reaches because the phenomenon is more local and less subject to spatial averaging in horizontal direction.

2.5. BAYESIAN DECISION ANALYSIS

Bayesian decision analysis concerns the logical analysis of choices and actions, under uncertain conditions about the state of the world and the outcomes of foreseen experiments and actions. Utility theory (Neumann and Morgenstern, 1947) forms the basis for optimizing decisions based on maximization of utility. In this dissertation this theory is used to optimise the costs of dike reinforcement (costs are negative utility). The experiments we consider aim to reduce uncertainties and adjust the subjective failure probabilities based on performance information. The actions aim to reinforce (strengthen) dikes, decreasing the failure probability such that a certain target reliability is met. Specifically for considering experiments and actions that affect subjective probabilities, pre-posterior analysis in combination with decision trees is used (Raiffa and Schlaifer, 1961; Benjamin and Cornell, 1970; Thöns, 2017).

The general decision problem of Raiffa and Schlaifer (1961) considers the choice an action $a \in A$, where the outcome of that action depends on the uncertain 'state of the

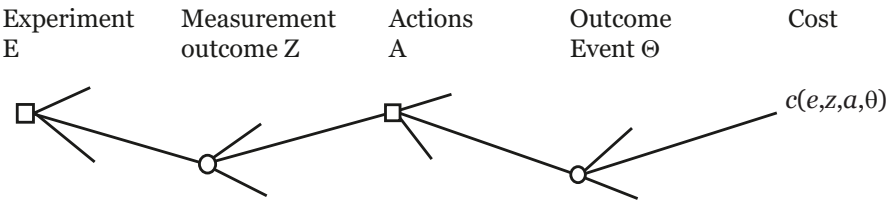


Figure 2.6: General decision tree

world' θ, and where the uncertainty about the state can be reduced by undertaking an experiment $e \in E$, at a certain cost. The outcome of the experiment (e.g. a measurement) $z \in Z$ is used to make a better decision on the action to take. The decision problem is visualized in a decision tree in Figure 2.6. The optimal strategy is defined as the set of an experiment and action (e^*, a^*) maximizing the utility, or minimizing the total expected cost:

$$C(e^*, a^*) = \min_{e,a} E\left[c\left(e, z, a, \theta\right)\right]. \tag{2.19}$$

The outcome of the experiment and actions depend on the uncertain state of the world. The total cost is therefore uncertain. When only prior information about the state is available, the expected cost is calculated using the prior probability density $f_\Theta(\theta)$:

$$C(e, a) = E\left[c\left(e, z, a, \theta\right)\right] = \int c\left(e, z, a, \theta\right) f_\Theta\left(\theta\right) d\theta. \tag{2.20}$$

When a measurement z is obtained, the prior knowledge on θ is updated, and the expected cost is calculated based on posterior probability distribution $f_\Theta(\theta|z)$, see Equation 2.21.

$$C(e, a) = \int c\left(e, z, a, \theta\right) f_\Theta\left(\theta|z\right) d\theta. \tag{2.21}$$

The formulation of Equation 2.21 applies to the situation where the outcome of the experiment z is known, which is not the case when we still have to decide whether to conduct the experiment. So, to determine the optimal strategy in advance, all possible outcomes z are accounted for in advance using pre-posterior analysis (Raiffa and Schlaifer, 1961; Benjamin and Cornell, 1970; Thöns, 2017). Here each possible outcome z is mapped into an action $a|z$ using decision rules $a = d(e, z)$. A decision rule is thus obtained by a posterior analysis, conditional on the presumed measurement. Then by integrating over all possible z, weighted with the likelihood of observing z, we obtain the expected total cost based on pre-posterior analysis. The total expected cost function to be minimized is:

$$C(e, d) = \int c\left(e, z, d(e, z), \theta\right) f_\Theta\left(\theta\right) d\theta. \tag{2.22}$$

The difference in expected cost and benefits with an experiment $(C(e, d))$ and without $(C_0(d))$ an experiment is called the Value of Information (VoI) (Raiffa and Schlaifer, 1961):

$$VoI = C_0(d) - C(e, d), \tag{2.23}$$

for which the decision analysis is aimed at maximizing the VoI.

3

UNCERTAINTY OF SPATIAL AVERAGE UNDRAINED SHEAR STRENGTH USING A SITE-SPECIFIC TRANSFORMATION MODEL

While in theory randomness is an intrinsic property, in practice, randomness is incomplete information.

Nassim Nicholas Taleb

Quantitative reliability analysis of slope stability starts with establishing prior estimates of the uncertainty of geotechnical parameters. This chapter presents a method to estimate the spatial average undrained shear strength using cone penetration tests, and the uncertainty involved. Special attention is paid to the degree to which the uncertainty is epistemic (and hence reducible) by examining the extent to which the contributing error sources are epistemic, or intrinsically variable. This chapter specifically deals with transformation models calibrated with local data, as cone penetration tests are often used to establish priors for reliability analyses of slope stability of linear infrastructures such as dikes; the proposed methods and data are also used in chapter 4 and 5.

This chapter is based on Van der Krogt, M.G., Schweckendiek, T. and Kok, M. Uncertainty in spatial average undrained shear strength with a site-specific transformation model. Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards, 1–11 (2018) and Van der Krogt, M.G. and Schweckendiek, T. (2019). Systematic Transformation Error in the Depth-Average Undrained Shear Strength. Proceedings of 7th international symposium on geotechnical safety and risk, Tapei, Taiwan.

Contents

3.1. INTRODUCTION

For the reliability based design and assessment of slope stability of dikes it is common to use cone penetration tests (CPTs) to estimate the undrained shear strength of soft soil layers because CPTs provide insight in the spatial variability of soil layers in vertical direction and the associated spatial averaging. To estimate soil properties from indirect measurements such as cone resistance, we need transformation models. A site-specific transformation model can be calibrated with laboratory tests on soil samples from borings at the site. This approach is efficient because we do not need a boring at all locations where a CPT was done, and borings are typically more labour intensive and more expensive than CPTs.

In using transformation models, we have to account for transformation uncertainty. If local data are used to calibrate an empirical transformation model, the various errors in local data also propagate into the transformation uncertainty, see Figure 3.1. In practice it remains challenging to properly quantify the transformation uncertainty. In particular it is difficult to distinguish between spatially variable components and random errors which are subject to spatial averaging, and systematic errors which are not. This chapter unravels the role of (spatial) averaging in the uncertainty of the spatially averaged undrained shear strength when using a site-specific transformation model for undrained shear strength from cone resistance to answer the sub-question:

To what extent is the uncertainty in the undrained shear strength reducible?

Section 3.2 investigates how to estimate the (de-trended) spatial average from CPTs using existing statistical approaches to quantify the uncertainty. It furthermore demonstrates those approaches for a synthetic random field example. Section 3.3 and Section 3.4 explore how the different error terms propagate into the transformation uncertainty and the estimates of the spatial average, for a synthetic random field example. Section 3.4 also proposes method to estimate the uncertainty in the spatial average at locations of interest with a CPT, accounting for the averaging of random errors and accounting for systematic errors. We also propose a method to estimate the spatial average and the uncertainty in the spatial average at other locations in the same statistically homogeneous deposit, based on the observed variability of multiple CPTs. The methods are demon-

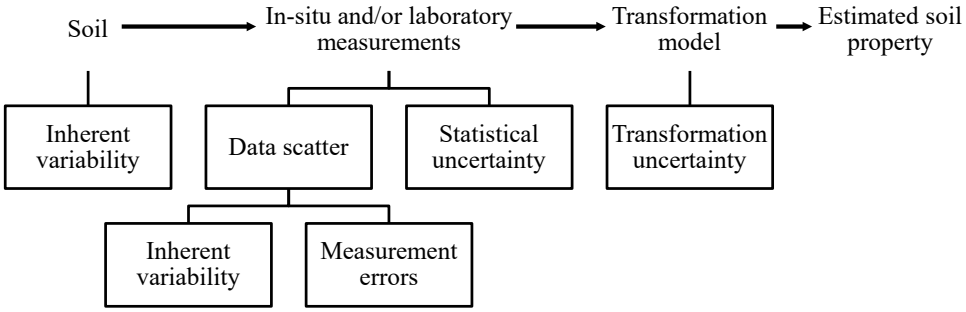


Figure 3.1: Overview of geotechnical uncertainties (Cao et al., 2017), adapted from Phoon and Kulhawy (1999a)

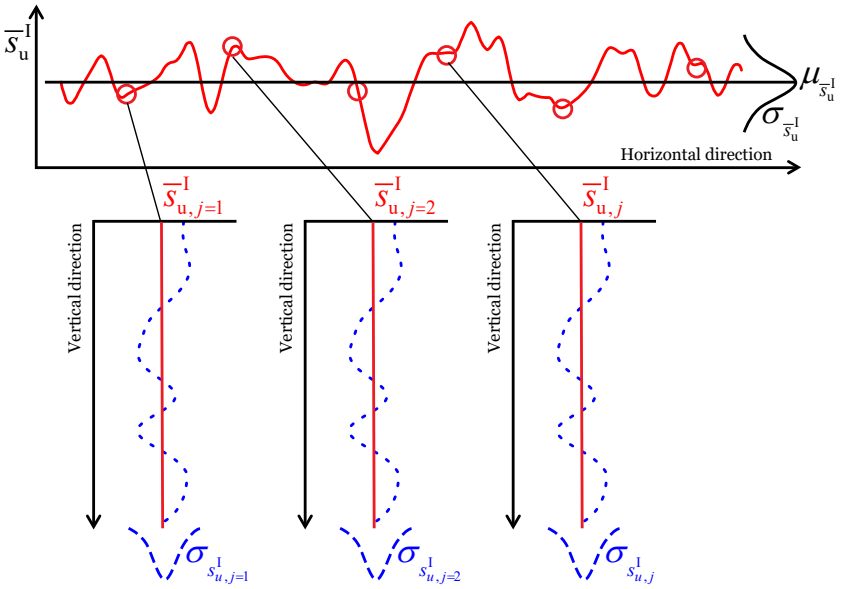


Figure 3.2: Schematic overview of a 2-D site (length and depth). Top: variability of the depth-average parameter (red) in horizontal direction. Bottom: scatter around the local depth-average over the thickness of a soil layer (blue). Figure adapted from Calle et al. (2021).

strated for case data of Eemdijk in Section 3.5, whereafter the results are summarized and the sub-question is answered.

3.2. CHARACTERIZING THE SPATIAL AVERAGE USING CPTS

In linear infrastructure projects, we often want to estimate the depth-average parameter of a soil layer at different locations across a site. Usually a soil layer is a geological deposit that is present over a site of about 1 to 10 km, at least for long linear infrastructures such as dikes, roads or railways. At each location where a CPT is available (depicted with circles in the schematic in Figure 3.2), the estimated depth-average shear strength is applicable to a cross-section, and the cross-section is representative for a horizontal section or reach shorter than the horizontal scale of fluctuation. At locations without a CPT, the depth-average can be estimated based on the statistics of the entire site, when the soil layer is statistically homogeneous.

The uncertainty of geotechnical parameters stems from various sources, see Section 2.4. According to, among others, Phoon and Kulhawy (1999a,b), the total uncertainty is a combination of the uncertainties from the various sources, see Figure 3.1. Phoon and Kulhawy (1999b) propose to apply variance reduction (as proposed by Vanmarcke (1977)) to the variance related to spatial variability ($\sigma^2_{\text{spatial}}$), and add the remaining variances to determine the total uncertainty:

$$\sigma^2 = \Gamma^2 \cdot \sigma^2_{\text{spatial}} + \sigma^2_{\text{measurement}} + \sigma^2_{\text{statistical}} + \sigma^2_{\text{transformation}}. \tag{3.1}$$

Essentially, determining the total variance as in Equation 3.1 implies the assumption that measurement uncertainty, statistical uncertainty and transformation uncertainty relate to systematic errors, which are not subject to (spatial) averaging. While this assumption is certainly conservative in the sense that it will lead to a high uncertainty estimate in engineering applications, the question is whether these error terms should be considered as entirely systematic and not subject to spatial averaging. And furthermore, what difference it would make in terms of probability distributions and characteristic values, if the assumption of the errors being systematic is relaxed. In this chapter this is demonstrated for the example of site-specific transformation models.

3.2.1. LOCATIONS WITH A CPT

The undrained shear strength at a location with a CPT can be estimated from cone resistance using a transformation model, see Equation 3.2. Here, s_u^I is the indirectly measured undrained shear strength, q_{net} the normalized cone resistance (cone tip resistance q_c corrected for pore water pressures u_2 through the cone factor a_{CPT} and normalized for the in-situ vertical stress σ_{v0}: $q_{net} = q_c + u_2 \cdot (1 - a_{CPT}) - \sigma_{v0}$), and N_{kt} the transformation model parameter.

$$s_u^I = \frac{q_{net}}{N_{kt}}. \tag{3.2}$$

The depth-average (denoted by a bar) parameter of a soil layer in CPT j, ($\bar{s}_{u,j}^I$) is estimated by the numerical mean of all CPT measurements N_j in that CPT (typically every 2 cm), see Equation 3.3 and Figure 3.2. Note that for the remainder of this chapter, also the term spatial avarage is used for the depth-average.

$$\bar{s}_{u,j}^I = \frac{1}{N_j} \sum_{i=1}^{i=N_j} s_{u,j,i}^I. \tag{3.3}$$

In line with Phoon and Kulhawy (1999b), the uncertainty in the estimated spatial average is the linear sum of measurement error (bias), transformation error, and statistical uncertainty, where the error terms are assumed to be independent, see Equation 3.4. Notice that the contribution of spatial variability is absent in the uncertainty estimate, as the spatial average is directly determined from the measurement.

$$\sigma^2_{\bar{s}_{u,j}^I} = \sigma^2_{meas.,sys.} + \sigma^2_{trans.} + \sigma^2_{stat.}. \tag{3.4}$$

The statistical uncertainty in Equation 3.4 can be estimated from the scatter around the spatial average and the number of measurements: $\sigma_{stat.} = \sigma_{s_{u,j}^I} / \sqrt{N_j}$, where $\sigma_{s_{u,j}^I}$ the observed scatter around the spatial average, and N_j is the number of measurements in one CPT. The total uncertainty Equation 3.4 can thus be written:

$$\sigma^2_{\bar{s}_{u,j}^I} = \sigma^2_{meas.,sys.} + \sigma^2_{trans.} + \frac{\sigma^2_{s_{u,j}^I}}{N_j}. \tag{3.5}$$

Theoretically, the observed data scatter around the spatial average ($\sigma_{s_{u,j}^I}$) stems from small-scale vertical spatial variability and random measurement errors (white noise).

The small-scale spatial variability can be estimated from the total point variance if we assume that the total point variance of the site $\sigma^2_{\text{spat.}}$ is the linear sum of the fluctuations around the spatial average in vertical direction σ^2_{f} and the fluctuations of that spatial average σ^2_{av}: $\sigma^2_{\text{spat.}} = \sigma^2_{\text{f}} + \sigma^2_{\text{av}}$. This concept of a 'composite random field model' (see Calle et al. (2021)) is illustrated in Figure 3.2 where blue dashed lines indicate the local fluctuations and the black line the fluctuation of the spatial average. When we substitute the definition of the variance reduction function (Vanmarcke, 1977) ($\Gamma = \sigma_{\text{av}}/\sigma_{\text{spat.}}$) into that equation, we obtain an expression for the expected variance around the spatial average in a CPT, as function of the total spatial variability: $\sigma^2_{\text{f}} = (1-\Gamma^2)\cdot\sigma^2_{\text{spat.}}$.

The random measurement error ($\varepsilon_{q_{\text{net}}}$) is an independent error source and typically modelled as multiplicative error. Hence, we write the two sources leading to the total observed scatter in terms of the coefficient of variation (CoV) V, $V = \sigma/\mu$. A theoretical estimate for the data scatter around the spatial average is thus:

$$\sigma_{s^{\text{I}}_{\text{u,j}}} = \bar{s}^{\text{I}}_{\text{u}} \cdot \sqrt{(1-\Gamma^2)\cdot V^2_{\text{spat.}} + V^2_{\varepsilon_{q_{\text{net}}}}}. \tag{3.6}$$

In practice it is challenging to distinguish the contributions of spatial variability and measurement error, and to estimate the data scatter by Equation 3.6. Therefore, we propose to simply estimate the statistical uncertainty in Equation 3.4 from the observed (total) variance in CPT j. In Section 3.2.3 (Figure 3.3) we compare this estimate with the 'true known' uncertainty of the spatial average (based on the difference between $\bar{s}^{\text{I}}_{\text{u,j}}$ and $\bar{s}_{\text{u,j}}$ for a random field example with synthetic data.

$$\sigma^2_{s^{\text{I}}_{\text{u,j}}} = \frac{1}{N_j - 1}\sum_{\text{i=1}}^{\text{i=N}_\text{j}} \left(s^{\text{I}}_{\text{u,j,i}} - \bar{s}^{\text{I}}_{\text{u,j}}\right)^2 \tag{3.7}$$

Note that random errors average, but we still have to account for systematic errors in the uncertainty estimate, see Van den Eijnden and Hicks (2019). For the remainder in this chapter, the term random error is used for white noise in the individual measurements and systematic error for a bias in a CPT or for the entire field (i.e. site).

3.2.2. LOCATIONS WITHOUT A CPT

When there is no CPT available in a section, an estimate for the shear strength could be based on the other CPTs across the site, see Figure 3.2. In a statistically homogeneous layer, our best estimate for the spatial average at a location without a CPT, is the mean value of the spatial averages from M CPTs across the site:

$$\mu_{\bar{s}^{\text{I}}_{\text{u}}} = \frac{1}{M}\sum_{\text{j=1}}^{j=M} \bar{s}^{\text{I}}_{\text{u,j}}. \tag{3.8}$$

We can estimate the uncertainty in the spatial average at a location without a CPT from the variance of the spatial averages from multiple CPTs, accounting for statistical uncertainty in the mean value (the term $\left(1+\frac{1}{M}\right)$):

$$\sigma^2_{\bar{s}^{\text{I}}_{\text{u}}} = \left(1+\frac{1}{M}\right)\cdot\frac{1}{M-1}\sum_{\text{j=1}}^{\text{j=M}} \left(\bar{s}^{\text{I}}_{\text{u},j} - \mu_{\bar{s}^{\text{I}}_{\text{u}}}\right)^2. \tag{3.9}$$

The aim is to investigate the effect of random errors in contrast with the assumption of only systematic errors. So, to have a clear comparison, we base ourselves on the same assumptions as Phoon and Kulhawy (1999a): a linear combination of spatial variability and error terms:

$$\sigma^2_{\bar{s}^{\mathrm{I}}_{\mathrm{u}}} = \Gamma^2 \cdot \sigma^2_{\mathrm{spat.}} + \sigma^2_{\mathrm{meas.,sys}} + \sigma^2_{\mathrm{trans.}} + \sigma^2_{\mathrm{stat.}}. \tag{3.10}$$

3.2.3. SYNTHETIC RANDOM FIELD EXAMPLE

The approach to estimate the uncertainty according to Equation 3.5 is evaluated with a synthetic random field example with an assumed perfect transformation model ($V_{\varepsilon_{\mathrm{t}}} = 0$). For the sake of a good comparison, we use synthetic data for which we generate the true but subsequently unknown values for the undrained shear strength s_{u} and errors, and compare the estimated uncertainty in the spatial average with the observed variability.

To this end, we generate a stationary Gaussian random field of 200 × 2000 (v×h) cells, representing a site with a 2 m thick soil layer over 2 km length, with Circulant Embedding (Kroese and Botev, 2015). The field size and resolution is expected to be large enough to sample enough independent samples and small enough to be computationally efficient. The soil property in the synthetic true field is normally distributed with mean value $\mu_{s_{\mathrm{u}}} = 20\,\mathrm{kPa}$ and standard deviation $\sigma_{s_{\mathrm{u}}} = 4\,\mathrm{kPa}$. In this example it is ignored that most geotechnical parameters are non-negative and therefore other probability distributions might be more suitable. The assumption of a statistically homogeneous random field with constant mean is justified if the field data is de-trended. A squared exponential correlation function is applied with a horizontal and vertical correlation length of respectively 25 m and 0.25 m (equivalent with $\delta_{\mathrm{h}} = 44\,\mathrm{m}$ and $\delta_{\mathrm{v}} = 0.44\,\mathrm{m}$), consistent with literature (among others, Phoon and Kulhawy (1999a); Li (2017)). From the true known field $\boldsymbol{s}_{\mathrm{u}}$, the measured field for the indirect measurement $\boldsymbol{q}_{\mathrm{net}}$ is generated, according to Equation 3.11.

$$\boldsymbol{q}_{\mathrm{net}} = \boldsymbol{s}_{\mathrm{u}} \cdot N_{\mathrm{kt}} \cdot \boldsymbol{\varepsilon}_{\boldsymbol{q}_{\mathrm{net}}}. \tag{3.11}$$

The example considers a perfect transformation model with a deterministic value of $N_{\mathrm{kt}} = 20$. The multiplicative random measurement error $\varepsilon_{q_{\mathrm{net}}}$ is modelled by a normal distribution with mean value 1 and the CoV $V_{\varepsilon_{q_{\mathrm{net}}}}$ is varied. The depth-average shear strength $\bar{s}^{\mathrm{I}}_{\mathrm{u,j}}$ is estimated according to Equation 3.3, and the uncertainty in the spatial average $\sigma_{\bar{s}^{\mathrm{I}}_{\mathrm{u,j}}}$ is estimated by Equation 3.5 (using Equation 3.7 to quantify the contribution of the statistical uncertainty). The analysis is done for a layer thickness of 0.2 , 1.0 and 2.0 m and a measurement interval of 0.02 m, such that $N_{\mathrm{j}} = [10, 50, 100]$. The estimated uncertainty is compared with the modelled uncertainty in this synthetic example: i.e. the standard deviation of the difference between the estimated and the 'true known' spatial average $\bar{s}^{\mathrm{I}}_{\mathrm{u,j}} - \bar{s}_{\mathrm{u,j}}$.

The estimated uncertainty coincides well with modelled uncertainty when there is a random error, see Figure 3.3. As the random error increases, it dominates the estimated statistical uncertainty. When there is no random measurement error, the estimate based on Equation 3.5 shows a slight difference with the true uncertainty, because it cannot be

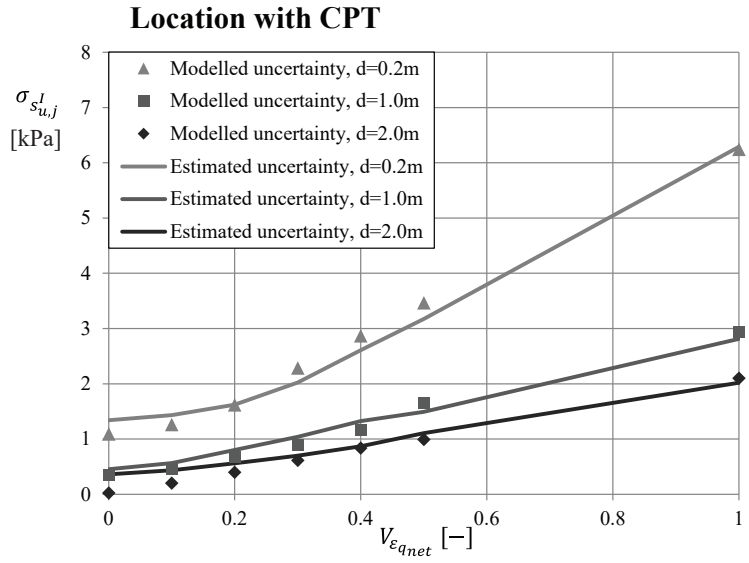


Figure 3.3: Estimated uncertainty of the indirectly estimated spatial average from a CPT compared to the actual uncertainty modelled in the synthetic example.

traced what part of the observed scatter is due to random errors, and what part is due to spatial variability.

3.3. TRANSFORMATION UNCERTAINTY OF A SITE-SPECIFIC TRANSFORMATION MODEL

The example in Section 3.2.3 assumes a perfect transformation model, which is unrealistic but serves the purpose of clarification. When we use a 'generic' transformation model from literature it is likely that the empirical model is biased for an entire site (Ching et al., 2016). This systematic transformation uncertainty should be accounted for in the estimated spatial average parameter, see Equation 3.5.

Fundamentally, transformation uncertainty is a model uncertainty. In principle, model error is meant to cover the model prediction errors for perfectly known model inputs (Ching and Phoon, 2014). In practical terms it is, however, impossible to determine model uncertainty in a clean fashion, nor transformation uncertainty for that matter, because such perfect conditions are not available. For instance, it is practically impossible to calibrate a site-specific transformation model where two paired measurements are at exactly the same location. Therefore, spatial variability causes additional error in the transformation uncertainty estimate. Moreover, there is measurement error in both CPT and laboratory measurements, which will have random and systematic components. This section analyses the propagation of these extraneous errors into the uncertainty of the transformation model parameter to explore to what degree transformation error is ultimately random or systematic.

For transformation models calibrated and used at a specific site, no systematic biases are expected for the entire site (Ching and Phoon, 2020). However, locally, the transformation model parameter may deviate from the site-average. The transformation model error is most certainly spatially variable because it is, at least to some degree, due to missing factors that are spatially variable, such as over consolidation ratio, water content and plasticity index. Since the stress state (e.g. loading history) is constant in a vertical profile, we assume that the transformation error is largely systematic per CPT, supported by Ching et al. (2016) where it was showed that the vertical scale of fluctuation of the transformation error is relatively large compared to the layer thickness. Because the horizontal distance between CPTs is usually larger than the scale of fluctuation of e.g. the stress state, we expect the transformation error is independent between CPTs. Hence we can justify the assumption that the transformation error is fully correlated in depth, and independent per CPT, at least for practical engineering purposes.

3.3.1. CALIBRATION OF THE TRANSFORMATION MODEL

To calibrate an empirical transformation model for undrained shear strength s_u^I from normalized cone tip resistance (Equation 3.2), we pair the measured cone resistance with direct (laboratory) measurements from (nearly) the same location ($q_{net,i}; s_{u,i}^D$). We can obtain the transformation model parameter N_{kt} from a linear regression analysis on n pairs from different locations within the same site (and deposit). The residuals represent the variability in s_u^I for a given value of q_{net}. Under the assumption of a constant CoV, we can write $V_{N_{kt}} = V_{s_u^I}$. Note that because of soil mechanical considerations, the regression line is forced through the origin.

Two regression methods are compared: minimizing the standard deviation (SD) and minimizing the coefficient of variation (CoV), see Equation 3.12 and 3.13, respectively. If the variability around the regression line is constant, minimizing the SD should be the correct regression method; if the scatter around the regression line increases with the mean, minimizing the CoV should be the better option.

$$\sigma_{s_u^I} = \sqrt{\frac{1}{n-1} \cdot \sum_{i=1}^{i=n} \left(s_{u,i}^D - q_{net,i} / \hat{N}_{kt} \right)^2}. \tag{3.12}$$

$$V_{s_u^I} = \sqrt{\frac{1}{n-1} \cdot \sum_{i=1}^{i=n} \left(\frac{s_{u,i}^D - q_{net,i} / \hat{N}_{kt}}{q_{net,i} / \hat{N}_{kt}} \right)^2}. \tag{3.13}$$

3.3.2. ERROR PROPAGATION INTO THE TRANSFORMATION MODEL

Since there will always be a non-zero distance between direct and indirect measurements, spatial variability will propagate in the transformation model (parameter). Also measurement error and transformation error propagate into the transformation model. Assuming that these errors are independent, the uncertainty in the transformation model parameter N_{kt} can be written as linear sum of the independent contributions:

$$V_{N_{kt}}^2 = V_{spat.}^2 + V_{\varepsilon_{s_u^D}}^2 + V_{\varepsilon_{q_{net}}}^2 + V_{\varepsilon_t}^2. \tag{3.14}$$

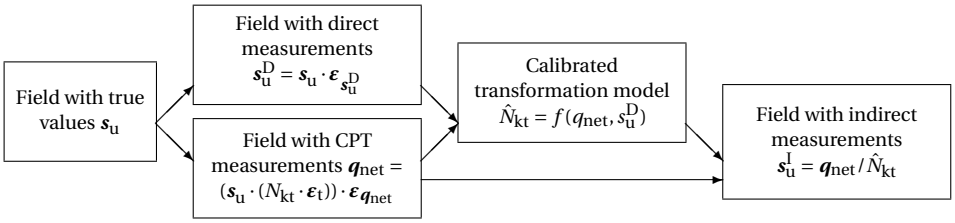


Figure 3.4: Schematic representation of the simulated random fields and transformations.

3

The contribution of the spatial variability to the transformation uncertainty can be estimated using the semi-variogram:

$$V_{\text{spat.}} = \sqrt{2 \cdot V_{s_u}^2 \cdot \left(1 - \rho\left(\Delta x, \Delta y\right)\right)}. \tag{3.15}$$

For the measurement and transformation errors, we expect that only random errors and spatially variable errors lead to variability of the indirectly measured undrained shear strength. Therefore, $V_{s_u}^2$, $V_{\varepsilon_{q_{net}}}^2$, and $V_{\varepsilon_t}^2$ in Equation 3.14 relate to the random and spatially variable errors. For errors in the independent variable (cone resistance), we also expect a bias due to the nature of the regression analysis, see e.g. Greene (2002).

Systematic measurement errors in the CPT measurements (such as a bias in the pressure transducer or incorrect zero point calibration) do not add to the variability in the transformation model, but will lead to a higher or lower value of the transformation model parameter. However, the measurements are still correlated to the correct direct measurements and therefore, systematic measurement errors in the CPT measurements cancel out if we use equally biased measurements with a biased transformation model. Systematic measurement errors in the direct measurement however, are problematic, because those lead to a non-quantifiable bias in the transformation model.

3.3.3. RESULTS SYNTHETIC RANDOM FIELD EXAMPLE

The calibration of the site-specific transformation model is demonstrated for the synthetic random field example from Section 3.2.3. We sample random fields for the direct measurement s_u^D and the indirect measurement q_{net} from the random field with true known values s_u. Random measurement errors (e.g. sample disturbance) and model errors are added to the samples according to Equation 3.16 and 3.17. Then we select n independent locations with sufficient inter-distance where we pair the measured shear strength with the measured cone resistance, and perform a regression analysis to obtain the estimate for the transformation model parameter, see Figure 3.5.

$$s_u^D = s_u \cdot \varepsilon_{s_u^D}. \tag{3.16}$$

$$q_{net} = (s_u \cdot (N_{kt} \cdot \varepsilon_t)) \cdot \varepsilon_{q_{net}}. \tag{3.17}$$

First, we consider an ideal transformation model with a deterministic value of $N_{kt} = 20$ and perfect measurements ($V_{\varepsilon_{q_{net}}} = V_{\varepsilon_{s_u^D}} = V_{\varepsilon_t} = 0$). The transformation model for

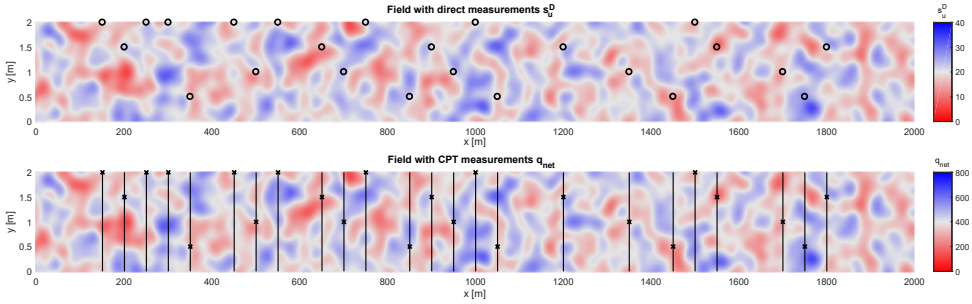


Figure 3.5: Direct and indirect measurements (circles and crosses, respectively) from synthetic random fields.

the site is calibrated with 25 CPTs (minimum spacing 50 m) and 25 laboratory test at arbitrary depths (see Figure 3.4). We estimate the transformation model parameter from the slope of the regression (Figure 3.6a): $\hat{N}_{kt} = 1/0.05 = 20$. The scatter in the regression is zero due to the absence of measurement and transformation errors.

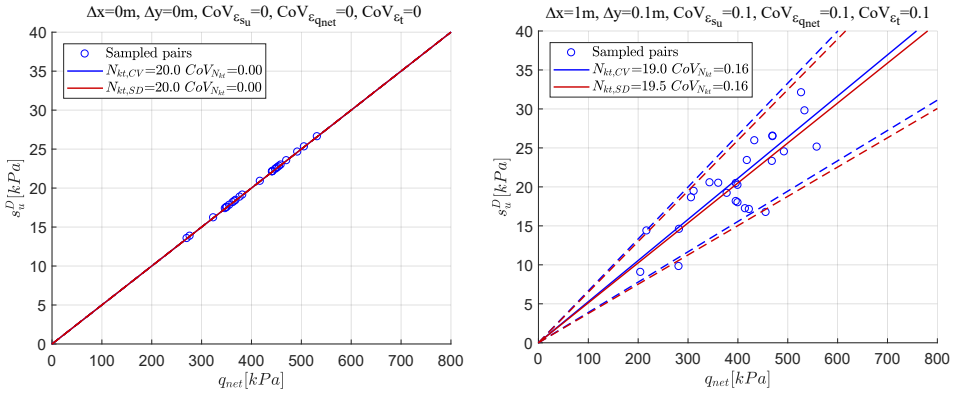
Furthermore, we investigate the effect of the above mentioned errors on the calibrated transformation model parameter $\hat{N}_{kt}$ and the uncertainty. We compare the variability obtained from the regression $V_{s_u^I}$ with the estimate based on Equation 3.14. To that end, we assume a horizontal distance between direct and indirect measurements of 1.0 m and vary the vertical distance between 0 and 0.20 m. For the other errors we assume CoV between 0 and 0.2 under the following assumptions:

- random measurement error in the direct measurement: $\varepsilon_{s_u^D}$ normally distributed with mean value 1 and CoV $V_{\varepsilon_{s_u^D}}$.
- random measurement error in the indirect measurement: $\varepsilon_{q_{net}}$ normally distributed with mean value 1 and CoV $V_{\varepsilon_{q_{net}}}$
- spatially variable transformation error: ε_t normally distributed with mean value 1 and CoV V_{ε_t}

For the case with $V_{\varepsilon_{s_u^D}} = 0.1$, $V_{\varepsilon_{q_{net}}} = 0.1$, and $V_{\varepsilon_t} = 0.1$, the results of a simulation are shown in Figure 3.6b. Notice that the total scatter is relatively large, compared to the spatial variability in s_u^I itself. It is quite common that the epistemic uncertainty dominates the spatial variability in geotechnical engineering.

Figure 3.7 shows the results for $\hat{N}_{kt}$ and $V_{N_{kt}}$ for 1000 random fields. On average, the transformation model parameter is biased for both regression methods. This bias is caused by the scatter due to spatial variability, random measurement error in the cone resistance and transformation error. The uncertainty in the transformation model parameter (and variability in the indirectly measured undrained shear strength) is estimated to be: $V_{N_{kt}} = \sqrt{0.11^2 + 0.1^2 + 0.1^2 + 0.1^2} = 0.21$ and is in accordance with the observed uncertainty in Figure 3.7.

Notice that random measurement errors in the shear strength from laboratory tests lead to scatter, but do not contribute to the bias, contrary to random measurement errors in the cone resistance. Compared to the statistical uncertainty, regression with minimizing SD is virtually unbiased for a Gaussian (normally distributed values) field, see Figure



(a) Without uncertainty

(b) With extraneous uncertainty

Figure 3.6: Calibration of the transformation model parameter. Dashed lines indicate the 90% confidence bounds.

Table 3.1: Average results for 1000 times repeated calibration of the transformation model parameter and uncertainty, for different combinations of errors and using two different regression methods.

	Modelled extraneous errors				**Regression method**				**Expected**
	Error term CoV			$\Delta x = 1$ m	Minimizing CoV		Minimizing SD		**uncertainty**
	ε_{s_u}	$\varepsilon_{q_{net}}$	ε_t	Δy	N_{kt}	$V_{N_{kt}}$	N_{kt}	$V_{N_{kt}}$	$V_{N_{kt}}$
Case 1	0	0	0	0	20	0.01	20	0.01	0.01
Case 2	0.05	0.05	0.05	0.05	19.6	0.11	20.1	0.11	0.10
Case 3	0.10	0.10	0.10	0.1	18.7	0.21	20.6	0.23	0.21
Case 4	0.15	0.15	0.15	0.15	17.1	0.31	21.2	0.39	0.30
Case 5	0.20	0.20	0.20	0.2	15.3	0.40	22.1	0.59	0.40

3.6b. The results for $\hat{N}_{kt}$ and CoV with different values of the CoV of the error terms are shown in Table 3.1.

It is found that the variability in indirectly measured undrained shear strength (or uncertainty in the transformation model parameter) obtained by minimizing SD, is on average slightly higher than what was expected based on the modeled errors. The difference increases with increasing variability and can likely be attributed to additional model error due to the regression method.

The statistical uncertainty in $\hat{N}_{kt}$ is a systematic error and depends only on the number of independent measurement pairs. For this example, the uncertainty is $V_{\text{trans.,stat.}} = V_{N_{kt}}/\sqrt{n}$. Note, that multiple measurement pairs in one CPT can be not fully independent, because they can have a correlated error. If the soil property in the random field is assumed to be log-normal distributed, it is found that both methods are equally biased. In this case there is no preference for one of the two regression methods.

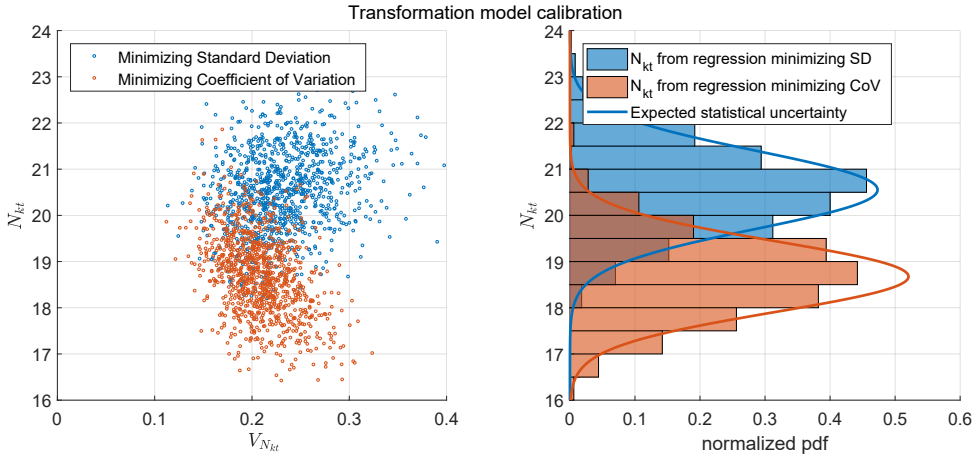


Figure 3.7: Results of the 1000 times repeated calibration of the transformation model parameter, using two different regression methods.

3.4. ESTIMATING THE UNCERTAINTY IN THE SPATIAL AVERAGE

For most geotechnical problems we are interested in spatial averages, for example, along a shear plane. The approaches proposed in literature in this respect consider only the averaging of the true spatial variability (Phoon and Kulhawy, 1999a,b; Vanmarcke, 1977), yet also other random (i.e. non-systematic) errors are also subject to averaging, at least when multiple measurements are available. Applying Equation 3.1 straightforwardly leads to an overestimation of the total uncertainty. Arguably, the systematic error component of the transformation uncertainty, is the uncertainty we are actually facing when executing a geotechnical analysis based on indirect measurements.

We propose two methods to estimate this systematic uncertainty in the transformation model parameter. The first is more theoretically founded, and is based on the estimated contribution of random and systematic components in the total uncertainty. The second is more empirically, and is based on the observed data from the field. Subsequently we investigate the appropriateness of the uncertainty estimates by analysing the difference between the indirectly estimated spatial average and the true known value Special attention is paid to the difference between random, systematic and spatially variable errors. In estimating the spatial average, it is differentiated between locations where a CPT is present and locations without.

3.4.1. PROPOSED METHOD

As argued above, only the systematic part of the transformation model parameter constitutes the uncertainty in the spatial average. To estimate this systematic part, we assume that the total variance in the transformation model parameter consists of a random and a systematic part: $V^2_{N_{\text{kt}}} = V^2_{N_{\text{kt}},\text{sys.}} + V^2_{N_{\text{kt}},\text{rand.}}$. We introduce the ratio of random variability and total point variability: $r = V^2_{N_{\text{kt}},\text{rand.}} / V^2_{N_{\text{kt}}}$, such that the systematic component in the transformation uncertainty can be estimated as follows: $V^2_{N_{\text{kt}},\text{sys.}} = (1-r) \cdot V^2_{N_{\text{kt}}}$. The

random and total part in the variability follow from Equation 3.14:

$$r = \frac{V^2_{\text{spat.}} + V^2_{\varepsilon_{s_u}} + V^2_{\varepsilon_{q\text{net}}}}{V^2_{N_{kt}}}. \tag{3.18}$$

The systematic part of the transformation uncertainty in the indirectly estimated spatial average can be estimated using Equation 3.19. Here the term $1/n$ accounts for the statistical uncertainty in the estimate value of N_{kt}, due to the limited number of data points. The statistical uncertainty is also a systematic uncertainty.

$$\sigma_{\text{trans.}} = \mu_{\bar{s}^{\text{I}}_u} \cdot \sqrt{V^2_{N_{kt},\text{sys.}} + V^2_{N_{kt},\text{stat.}}} = \mu_{\bar{s}^{\text{I}}_u} \cdot \sqrt{(1-r) + \frac{1}{n}} \cdot V_{N_{kt}}. \tag{3.19}$$

In practice, we often do not have quantitative information on the random (and systematic) error. and thus values of r will have to be based on expert judgement. To get a rough estimate of r we propose an alternative approach to estimate the value of r based on the observed variability of the indirectly measured undrained shear strength. We can write the total point variability of the indirectly measured undrained shear strength of the entire site as the summation of random and a systematic variance, i.e. fluctuations around the spatial average in CPT j and fluctuations of the spatial average: $\sigma^2_{s^{\text{I}}_u} = \sigma^2_{s^{\text{I}}_{u,j}} + \sigma^2_{\bar{s}^{\text{I}}_u}$, see Figure 3.2. Then, the ratio r can then be rewritten as follows:

$$r \approx \frac{\sigma^2_{s^{\text{I}}_{u,j}}}{\sigma^2_{s^{\text{I}}_u}}. \tag{3.20}$$

This estimate for the share of random and systematic uncertainty in the transformation model parameter is dependent on the spatial variability, through the total uncertainty (Equation 3.6 and Equation 3.5). Hence, the estimated value according to Equation 3.20 may deviate from the definition in Equation 3.18. However, when epistemic uncertainty is dominant in the total uncertainty (which is often the case in geotechnical engineering, see e.g. Nadim (2015)), then the difference is small.

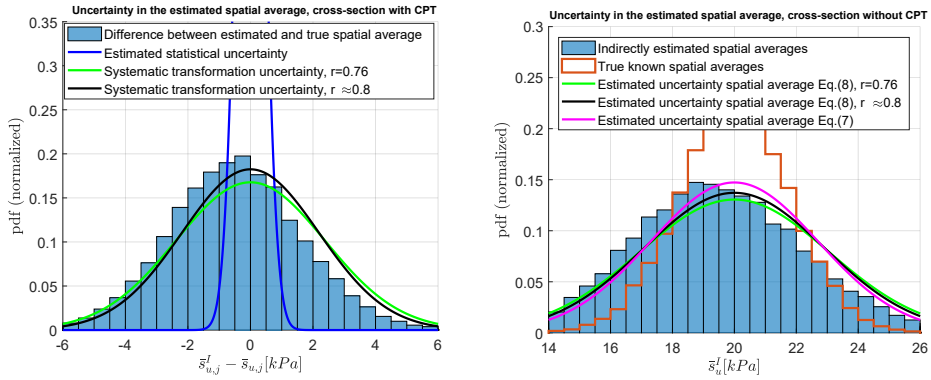
Including the estimated systematic transformation uncertainty in the uncertainty estimate of the spatial average for a location with a CPT (Equation 3.5) leads to:

$$\sigma^2_{\bar{s}^{\text{I}}_{u,}} = \frac{1}{N_j} \cdot \sigma^2_{s^{\text{I}}_{u,j}} + \sigma^2_{\text{meas.,sys.}} + \left(\mu_{\bar{s}^{\text{I}}_u} \cdot \sqrt{(1-r) + \frac{1}{n}} \cdot V_{N_{kt}} \right)^2. \tag{3.21}$$

The uncertainty in the spatial average at locations without a CPT is estimated by Equation 3.9. Since this uncertainty is based on the spatial average (of indirect measurements), it includes the averaging of random errors and true spatial variability already.

3.4.2. RESULTS SYNTHETIC RANDOM FIELD EXAMPLE

The proposed approach is demonstrated for the synthetic random field in Section 3.3.3. We use the calibrated transformation model factor from the 1000 random fields to estimate the spatial average undrained shear strength at locations with and without a CPT. For locations with a CPT, the uncertainty in the spatial average ($\bar{s}^{\text{I}}_{u,j} - \bar{s}_{u,j}$) is shown by the



(a) Uncertainty of the spatial average

(b) Variability of the spatial average

Figure 3.8: Uncertainty and variability of the spatial average estimated using indirect measurements.

histogram in 3.8a. Taken across all sites, the transformation model is virtually unbiased, reflected in the histogram centred around 0. However, for estimates at specific locations, the uncertainty is larger than only the expected statistical uncertainty (the blue line) due to the spatially variable transformation model error, which is systematic per location.

The estimate for the systematic part of the transformation uncertainty based on the imposed random errors (values in Table 3.1) using Equation 3.14: $r = (0.11^2+0.1^2+0.1^2)/(0.11^2+0.1^2+0.1^2+0.1^2) = 0.76$, shown by the green line. The factor r based on the variability of the indirect measurements: $r \approx \sigma^2_{s^I_{u,j}} / \sigma^2_{s^I_u} = 0.80$, shown by the black line. We find that the estimate of r with the second method is higher than the value obtained by Equation 3.18, because the ratio of local point variance and total point variance comprises, besides averaging of random errors, also averaging of true spatial variability. Therefore, the systematic part in the transformation uncertainty is underestimated by Equation 3.20. In this numerical example where epistemic uncertainty is dominant, the difference between the two methods is negligible (green versus black lines in Figure 3.8a).

For estimating the spatial average undrained shear strength at locations without a CPT, we estimate the variability of the indirectly estimated spatial averages for 1000 random fields. The found variability is shown by the histogram in Figure 3.8b. Besides the spatial (average) variability, it also contains systematic uncertainty because estimates at individual locations are biased (see Figure 3.13a). The green and black line depict the uncertainty estimates using Equation 3.10, with the estimated values for r using Equations 3.18 and 3.20. The magenta line shows the estimated uncertainty according to Equation 3.9: the estimate based on the observed variability. The results are in line with each other for this example, which substantiates the appropriateness of the proposed approach.

3.4.3. PRACTICAL IMPLICATIONS

The presented method appropriately accounts for averaging of both spatial variability and random errors in the uncertainty estimate of the spatial average from indirect CPT measurements. The present example has contemplated one end of the spectrum in the

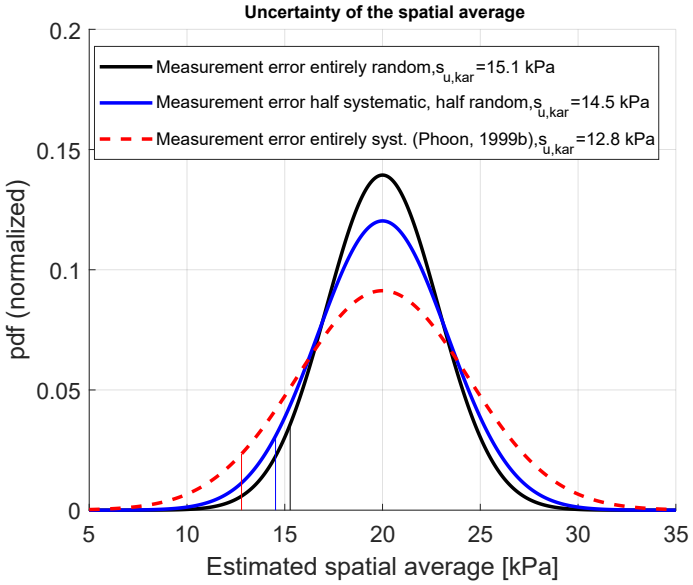


Figure 3.9: Comparison presented method with established method.

sense that measurement errors have been assumed entirely random (i.e. white noise). On the other hand, assuming these errors entirely systematic as done in Phoon and Kulhawy (1999b) is conservative. The comparison in Figure 3.9, in terms of 5%-quantile characteristic values, $s_{\text{u,kar.}}$, shows that there is considerable margin between these two assumptions. Even with half of the measurement error being systematic and half random, it is still likely that a substantial part of the transformation uncertainty is random, because of spatial variability in the transformation model.

3.5. CASE STUDY: EEMDIJK

The results so far only considered synthetic data - in this section we investigate the implications for a case study of a real site. Here a site-specific transformation model was used to infer the undrained shear strength from CPTs. The case study demonstrates what the impact is of using a more differentiated approach of separating random and systematic uncertainty. The following proposed step-wise approach is followed:

1. Calibrate a site-specific transformation model based on n pairs of cone resistance and laboratory tests. Obtain the estimate of N_{kt} and the uncertainty in this parameter $V_{N_{\text{kt}}}$ using Equation 3.13.
2. Estimate the share of random error in $V_{N_{\text{kt}}}$, preferably using Equation 3.18. A rough estimate can be made using Equation 3.20 when the random errors are dominant.
3. Determine the systematic part of the transformation uncertainty by accounting for averaging of random errors, and by accounting for statistical uncertainty in the site-specific transformation model parameter, using Equation 3.19.

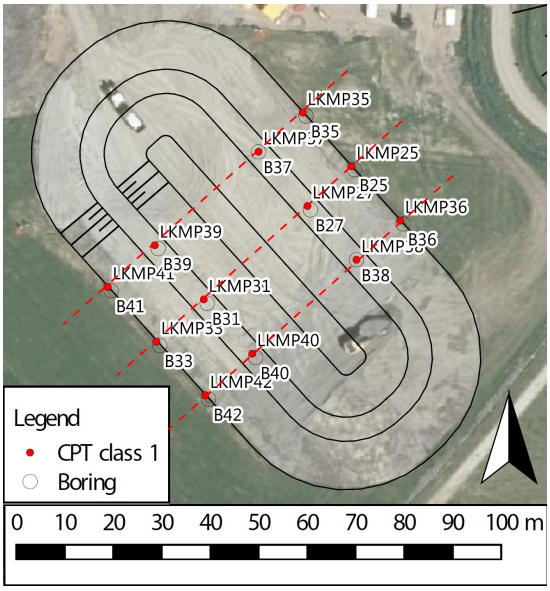


Figure 3.10: Satellite image of the test site during site preparation. The black lines indicate the intended ring dike. The red dots indicate the location of the CPTs, the open circles the borings.

4. Determine the total uncertainty in the depth-average undrained shear strength for locations with a CPT using Equation 3.21, and for locations without a CPT using Equation 3.9.

3.5.1. CASE DESCRIPTION

The case study concerns a full-scale test of two dikes, one conventional dike and one dike reinforced with a sheet pile wall. Both dikes were loaded until failure, with the main objective of validating models for slope failure of dikes with and without sheet piles. For more information on the 'Eemdijkproef', see Lengkeek et al. (2019).

The test dike was built in several construction stages to avoid disturbances in the subsurface by slope instability. Therefore, the stability was analysed during the construction phase, and the undrained shear strength was estimated across the site of approximately 100×100 m, see Figure 3.10.

The site is located in Eemdijk (The Netherlands). The surface level is around the Dutch vertical datum Amsterdam Ordnance Datum (in Dutch: Normaal Amsterdams Peil (NAP)): at −0.1 m NAP. The subsoil consists of a 1.5 to 2.0 m thick (organic) clay layer, an approximately 2.5 m thick peat layer and a Pleistocene sand layer of 6 m, see Figure 3.11.

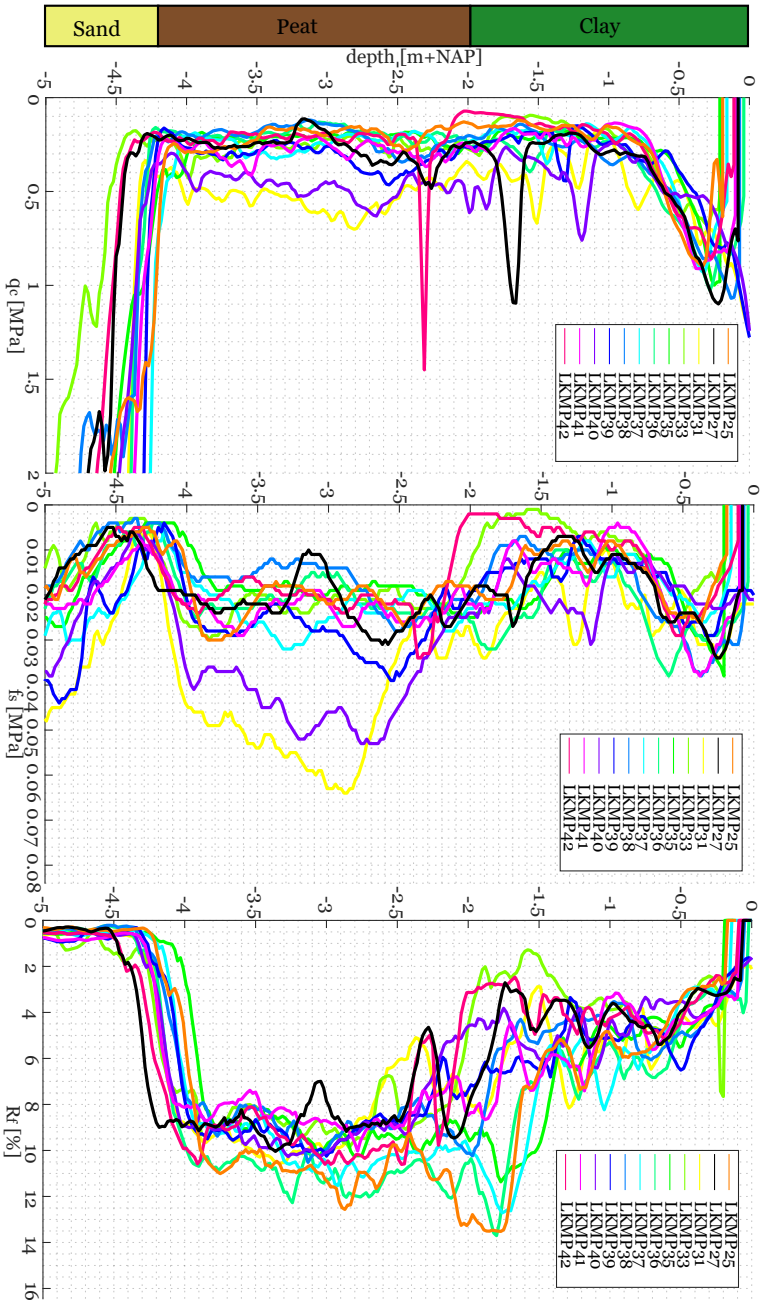


Figure 3.11: Cone resistance qc [MPa], sleeve friction fs [MPa] and friction ratio Rf [%] of 12 CPTs at the test site.

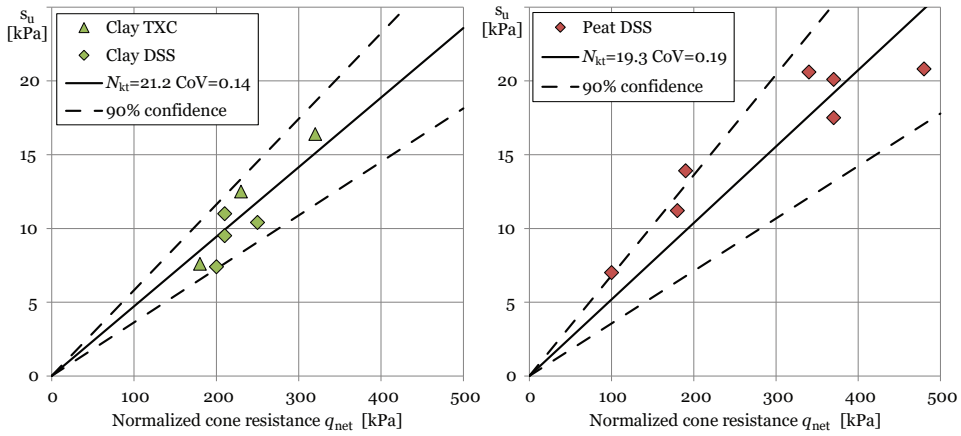


Figure 3.12: Cone resistance qc [MPa], sleeve friction fs [MPa] and friction ratio Rf [%] of 12 CPTs at the test site.

3.5.2. SITE-SPECIFIC TRANSFORMATION MODEL FOR UNDRAINED SHEAR STRENGTH

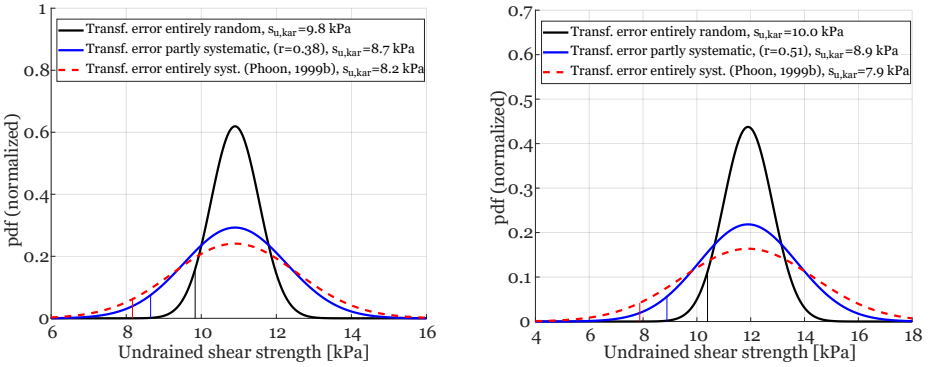
At the site, 12 pairs of mechanical borings and Class 1 CPTs with u_2 water pressure measurements were carried out (see locations in Figure 3.10) and used to derive the vertical soil profile, see Figure 3.11. The ground water table is around 0.75 m NAP, therefore the top part of the clay layer is unsaturated, which is reflected in a higher cone resistance.

Of seven clay and seven peat samples from the borings, the in-situ undrained (critical state) shear strength is measured in the laboratory with Unconsolidated Undrained triaxial compression (TXC) (TXC-UU) and direct simple shear (DSS) tests. These direct measurements of the shear strength (s_u^D) are paired with the normalized cone resistance q_{net} of the CPT at nearly the same location and depth as the direct measurement, see Figure 3.12. The transformation model parameter is derived using linear regression, as shown in Equation 3.12. The result is shown in Figure 3.12 and the uncertainty is depicted by the 90% confidence interval. To avoid large statistical uncertainty due to a limited number of data points, both the TXC and DSS tests are used for the transformation model for clay.

3.5.3. UNCERTAINTY IN THE DEPTH-AVERAGE UNDRAINED SHEAR STRENGTH

Using the empirical transformation model, the depth-average undrained shear strength in the clay and peat layer is estimated. Figure 3.13a and 3.13b show the uncertainty calculated using Equation 3.5 and 3.19 for CPT LKMP27. The black and red lines exemplify the assumptions of an entirely random ($r = 1$) or entirely systematic ($r = 0$) transformation uncertainty, respectively. The choice for either of both assumptions significantly impacts the 5% characteristic value.

A more realistic estimate of the systematic transformation uncertainty is made using Equation 3.18. Based on the assumptions that the random measurement error in



(a) Clay layer

(b) Peat Layer

Figure 3.13: Uncertainty in the depth-average undrained shear strength in clay and peat layer at CPT LKMP27.

cone resistance and laboratory tests is respectively 1% and 5%, and the spatial variability contributes 5% random error in the transformation model, this yields a value of $r = \sqrt{0.05^2 + 0.01^2 + 0.05^2}/0.19 = 0.38$ for the transformation model for peat and $r = 0.51$ for clay. The corresponding estimated uncertainty is shown by the blue line in Figure 3.13a and Figure 3.13b. For the considered CPTs, this more realistic assumption of the systematic transformation uncertainty leads to approximately 10% higher characteristic values compared to a conservative assumption of the transformation uncertainty being entirely systematic.

If r is approximated based on the ratio of local variance versus total point variance (Equation 3.20), we find a ratio of 0.64 for both the clay and peat layer (coincidentally). The approximation underestimates the systematic transformation uncertainty because the approximation falsely includes spatial averaging (just like the example in Section 3.4.2), and does not account for random errors in direct measurements. However, the difference between the two approaches is not very large, but it underlines that a larger part of the uncertainty may be random.

The CPTs are used to determine the shear strength in the (NE-SW oriented) cross-sections. Each cross-section is representative for a section of approximately the distance between the CPTs. For the sections at the head sides of the ring dike (without a CPT), the depth mean are estimated using Equation 3.9. Based on the results of 10 CPTs, the average shear strength is 9.4 kPa and 10 kPa for the clay and peat layer, respectively. The estimated uncertainty is 1.3 kPa and 1.8 kPa, respectively. Note that two CPTs are excluded from the statistical analysis, because these locations have a different stress history, because these locations were pre-loaded by an old embankment.

3.6. SUMMARY AND RECOMMENDATIONS

This chapter shows that we can use direct measurements (e.g. laboratory tests) and indirect measurements (e.g. CPTs) from a site to calibrate a site-specific transformation model. This site-specific transformation model can be used to estimate the spatial average of the soil parameter of interest using indirect measurements, which are often

cheaper and provide better spatial coverage. A virtually unbiased transformation model for a site can be obtained by linear regression, contrary to generic transformation models, which are biased for the entire site.

We demonstrated that the uncertainty in the transformation model parameter contains random errors, which are subject to averaging in estimating the spatial average. Therefore, we should not only account for spatial averaging of the actual soil heterogeneity, but also for averaging of random measurement errors. The remaining component in the uncertainty of the indirectly estimated spatial average is the statistical uncertainty of the transformation model, and the locally unknown bias in the transformation model.

This systematic component of the uncertainty in the site-specific transformation model can be estimated using information on the contributions from random and systematic errors involved (Equation 3.18), or based on the ratio local versus total (site) point variance Equation 3.20). For the CPTs considered in the Eemdijk case study, such a realistic assumption of the systematic transformation uncertainty leads to approximately 10% higher characteristic values compared to conservative assumptions of the transformation uncertainty being entirely systematic. Therefore, we should not only focus our site investigation on estimating the heterogeneity of the subsoil, but also on differentiating between systematic and random errors, e.g. by repetitive laboratory measurements or analysing the spatial variability of the transformation error.

The considerations and results imply that there are several possibilities to reduce the uncertainty in the indirect estimate of the undrained shear strength (or any other parameters obtained in a similar manner). One option is to minimize the distance between a direct and indirect measurements, as spatial variability propagates into transformation uncertainty. Because the transformation model error is largely systematic in a vertical, it is recommended to add direct measurements at different CPTs/boreholes, rather than at different depths in the same vertical. Reducing measurement error helps too, but particularly the bias in direct measurements is to be avoided.

3.7. CONCLUDING REMARKS

In addition to adequately establishing the uncertainty of geotechnical parameters using CPTs, for Bayesian analysis and reliability updating it is important to know to what extent the uncertainty reducible is. This chapter unravelled the role of (spatial) averaging in the uncertainty of undrained shear strength obtained with site-specific transformation models to answer the question:

To what extent is the uncertainty in the undrained shear strength reducible?

The uncertainty in the spatially averaged undrained shear strength arises from several components. Spatial variability, measurement error, statistical error, and transformation error all contribute to the total uncertainty. Although spatial variability does lead to variation of soil properties over space, the properties at a specific location are constant. The estimates, however, exhibit a considerable epistemic uncertainty because all error sources contributing to the total uncertainty of the spatial average are due to a lack of knowledge, and thus the uncertainty is in principle reducible.

4

IMPROVING DIKE RELIABILITY ESTIMATES BY INCORPORATING CONSTRUCTION SURVIVAL

When utilizing past experience in the design of a new structure we proceed by analogy and no conclusion by analogy can be considered valid unless all the vital factors involved in the cases subject to comparison are practically identical. (...) Hence our practical experience can be very misleading unless it combines with it a fairly accurate conception of the mechanics of the phenomena under consideration.

Karl von Terzaghi, 1939

The construction of dikes on soft soils is one of the possibly critical loading conditions to which dikes and the subsoil are subjected. The observation of survival of the construction contains information about the minimum strength along the slip plane, and can be used to reduce uncertainty, and update the failure probability. This chapter investigates how this performance information affects the reliability estimates for slope stability of dikes, using Bayesian updating to incorporate the information into the reliability analyses.

This chapter is based on Van der Krogt, M.G., Schweckendiek, T. and Kok, M. (2021). Improving dike reliability estimates by incorporating construction survival, *Engineering Geology*, 280, 105937.

Contents

4.1. INTRODUCTION

Incorporating survival information in probabilistic analyses can significantly reduce uncertainty and hence significantly improve reliability estimates for slope stability (Zhang et al., 2011; Li et al., 2015; Schweckendiek, 2010; Schweckendiek et al., 2014). The degree to which the new information leads to significant uncertainty reduction (the *informativeness* of an observation) depends, broadly speaking, on the plausibility of that observation, in relation to the hypothesis. Based on earlier findings (Zhang et al., 2011; Schweckendiek, 2010; Schweckendiek et al., 2014), a significant reliability update is mainly to be expected when extreme loading conditions are survived. Observations of survived loading conditions such as extreme flood water levels are rare, and hence, not always available. Instead, we may consider another potentially critical and more widely available loading condition for dikes: the construction. During the construction of embankments on soft soils in general, and dikes in particular, the stability typically reaches critical levels. The main cause is the resulting excess pore water pressure in the foundation as the embankment is raised. Baecher and Ladd (1997) showed that performance information of survived construction stages can be used to update the slope stability predictions in later construction stages. That case study, however, only considered the stability during the construction, not flood loading.

The objective of this chapter is to examine how the survival of the construction stage can be used to improve the reliability estimates of dike slopes in flood conditions. The main question is formulated as:

> *How does the performance information contained in the survival of the construction phase affect the reliability estimates for slope stability of dikes?*

To answer the research question, this chapter is structured as follows. First, we compare the factors of safety at the end of construction with the factors of safety in flood conditions for a range of hypothetical cases of typical dikes in Section 4.2. This indicates for which cases the construction of a dike is a critical loading condition, and hence if survival of that situation is valuable information to consider in the reliability analysis. Section 4.3 proposes a practical approach to incorporate construction survival in the probabilistic analysis of a dike in flood conditions using Bayesian updating, to investigate the effect for different cases, under various hypothetical survived conditions in Section 4.4. Furthermore, Section 4.5 presents the results for the recently constructed Eemdijk test dike from Section 3.5, demonstrating the practical applicability of the proposed approach. The chapter concludes with a summary and a discussion on how the approach can be applied in practice, which further developments are desirable, and how the performance information can be further used to optimize flood risk management.

4.2. DIKE CONSTRUCTION AS CRITICAL LOADING CONDITION

The construction of a dike is one of the loading conditions to which a slope and subsoil is subjected, when regarding the safety with respect to slope stability (U.S. Army Corps of Engineers (USACE), 2003). Raising an embankment leads to excess pore water pressures in soft soil foundation layers (see Figure 4.1), resulting in low effective stresses. Due to feasibility and economic reasons, critical levels of stability (low factors of safety) are

often accepted during the construction of dikes, all the more because potential damage in terms of loss of life and injuries during construction is usually low, compared to the potential flooding damage at design conditions.

If the construction of a dike is a critical loading condition, then survival of this loading condition provides additional information about the shear strength properties involved. The corresponding information of construction survival is that the factor of safety must have been greater than 1.0 at the time of the observation. Field observations to substantiate observations of construction survival can be the absence of cracks or excessive deformations, or other monitoring indicating that a rotational shear failure was not initiated under de observed loading conditions (Tavenas et al. (1979) describes some of those).

The question is, however, what information does construction survival provide about the reliability under the design loading conditions of a dike, namely flood loading. Although flood loading and the construction seem to be two different loading conditions, they are in fact quite similar. First, because the main load effect in terms of increased pore water pressures is, in principle, comparable. Second, because the failure mode is similar, in terms of potential slip planes intersecting the dike body and mostly the same subsoil layers.

There are, of course, differences, which make it impossible to use the survival information in a deterministic approach, while it can be done in Bayesian analysis. For example, the pore water pressures during flood loading are induced by seepage and mainly affect the dike body. The pore water pressures in the soft foundation soil are typically less affected, depending mainly on the flood duration. In contrast, the main increase in pore pressure during construction occurs in the soft soil foundation below the dike. Differences can be treated in Bayesian analysis, for example, by using the information of construction survival for updating the time-invariant soil parameters, instead of relating the information to the stress-dependent shear strength. Moreover, slightly lower effective stresses do not necessarily lead to significantly lower undrained strength of overconsolidated soil when using Critical State Soil Mechanics (CSSM) and Stress History and Normalized Soil Engineering Properties (SHANSEP), as pointed out by Shewbridge

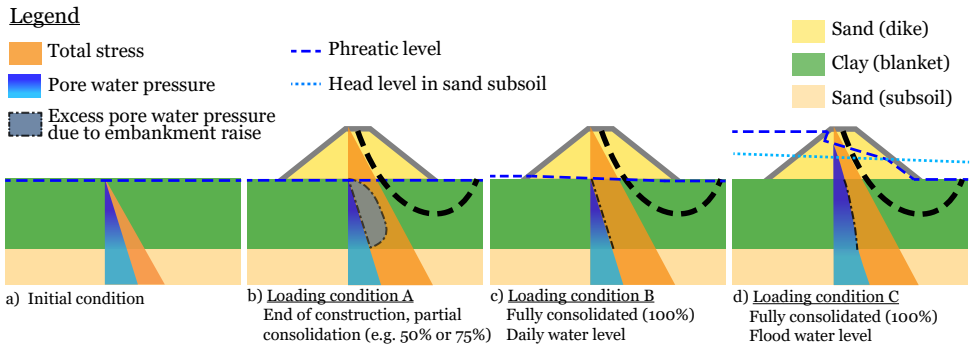


Figure 4.1: Overview of different loading conditions of a sand dike (with clay cover) on a soft soil blanket layer and sand subsoil. The typical failure mode (slip plane) is indicated with the dashed line.

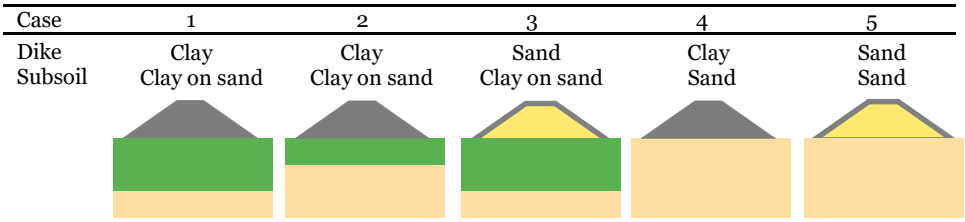


Figure 4.2: Overview of characteristic dike profiles. Grey and green indicate clay soil (for the dike core and subsoil, respectively). Yellow and orange indicate sand soil (for the dike core and subsoil, respectively).

and Schaefer (2013). Thus, the undrained shear strength is also approximately the same in both loading conditions.

4.2.1. DETERMINISTIC CASE STUDY

In order to sort out in which situations the construction phase is indeed a critical loading condition, we compare the factors of safety for five hypothetical, typical dike cross sections (outlined in Figure 4.2). These cases cover different combinations of subsoil (thick clay layer on sandy subsoil, thin clay layer on sandy subsoil, and only sand subsoil) and dike material (clay or sand with a clay cover). For the cases with a clay layer, we distinguish cases with normally consolidated (NC) versus over-consolidated (OC) soil. All cases involve a dike height of 4.0 m and a 1:4 slope (v:h). The five cases certainly do not cover all possible cases, but the cases are typical for the Dutch situation, and also characteristic for other locations, especially in deltaic areas.

The soil shear strength is modelled according to the CSSM framework (Schofield and Wroth, 1968) because the critical state concept suits best to a situation of continuous shear deformation along an entire slip plane (with large strains and no peak strength). The shear strength of the sand foundation layer and the unsaturated part of the dike body are modelled with a critical state friction angle ϕ_{cs}. The undrained shear strength (s_u) of saturated clay is modelled with the SHANSEP formulation (Ladd and Foott, 1974), see Equation 2.16.

The probability distributions of the soil properties (see Table 4.1) are chosen such that they represent realistic probability distributions. Spatial variability is implicitly modelled in the parameter distributions, by modelling spatial average soil properties. The probability distributions resemble values which are typically encountered in the field, in line with values reported in various case studies: e.g., Berre and Bjerrum (1975); Ochiai (1980); Ladd (1991); Baecher and Ladd (1997); Watabe et al. (2002); Stuedlein et al. (2012); De Koning et al. (2019).

The factor of safety F_s has been examined for three consecutive loading conditions, see Figure 4.1b-d) and listed below:

- (A) end of construction, partially consolidated,
- (B) fully consolidated, daily water levels,
- (C) fully consolidated, flood water levels.

Table 4.1: Probability distributions of soil properties used in the case studies.

Soil	**Volum. weight** $\gamma_{sat.}/\gamma_{unsat.}$ [kN m^{-3}]	**Normally consolidated undrained shear strength ratio** S [−]	**Strength increase exponent** m [−][a]	**Pre-overburden pressure** POP[b] [kPa]	**Critical state friction angle** ϕ_{cs}[°]
Dike Clay[c]	17/17	Log-normal μ 0.30, σ 0.03	Deterministic 0.80		Log-normal μ 32.0, σ 3.2
Dike Sand	17/19				Log-normal μ 32.0, σ 3.2
Clay NC	14/14	Log-normal μ 0.30, σ 0.03	Deterministic 0.80	Log-normal μ 10.0, σ 2.0	
Clay OC	14/14	Log-normal μ 0.30, σ 0.03	Deterministic 0.80	Log-normal μ 30.0, σ 6.0	
Sand	18/20				Log-normal μ 35.0, σ 1.75

[a] Realisations of S and m might be negatively correlated, however, we neglect this effect because the influence of m is typically small (for this reason m is modelled deterministically).
[b] The pre-overburden pressure refers to a the initial situation, before construction of the dike. A NC POP =10 kPa, $OCR \approx 1-3$ and OC POP =30 kPa, $OCR \approx 3-5$ situation is distinguished.
[c] Saturated Dike Clay is modelled using SHANSEP parameters S, m and POP (or OCR), unsaturated Dike Clay is modelled using a critical state friction angle ϕ_{cs}.

Loading condition A is immediately after finishing the construction of the dike. As a result of the staged construction process, excess pore pressures are still present in the clay foundation layers. The degree of consolidation (i.e. the dissipated excess pore water pressures in the foundation) are typically based on actual pore water pressure data, for real-life situations. For the hypothetical cases we assumed typical average degrees of consolidation of 50% and 75% to study the sensitivity to this parameter. Although the actual excess pore pressure varies with depth, we model a constant excess pore water pressure distribution over the soil layer. This represents the average degree of consolidation along potential slip planes. The phreatic level at the end of construction is assumed to be at surface level, and the dike body unsaturated. The construction is assumed to be with a new material with no significant unsaturated contribution to the shear strength (e.g. no suction forces) leading to the assumption that the unsaturated contribution in the construction phase is neglected. A sensitivity analysis in Appendix A shows that the assumption may have small impact but it does not lead to different conclusion.

Loading condition B is defined as the situation in which the subsoil has been fully consolidated. In this situation all excess pore pressures have dissipated (i.e. the degree of consolidation is 100%). The phreatic level is at surface level, identical to loading condition A. Loading condition C considers the dike with a fully consolidated subsoil in flood conditions, with steady-state seepage conditions (i.e. largely saturated dike body and increased head in the foundation layers, schematically indicated in Figure 4.1d).

The slope stability has been calculated with the LEM Bishop (Bishop, 1955), using the D-Stability software (Brinkman et al., 2018). Advantages of the LEM method compared to more advanced models are its simplicity, robustness, and the favourable computation

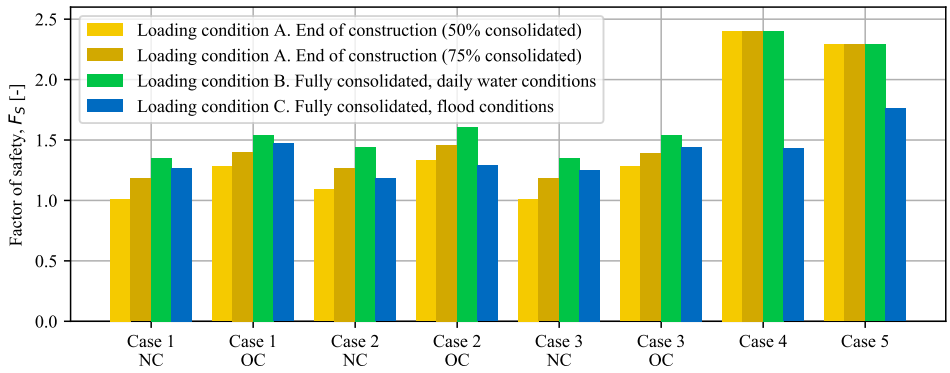


Figure 4.3: Factor of safety for slope stability considering different loading conditions for various typical dike profiles with NC and OC subsoil. For dikes on clay subsoil (cases 1-3), the stability at the end of construction is much lower than during flood conditions.

time. Although advanced methods might predict deformations of a slope better, they usually require more computation time and are less robust (e.g. Juang et al. (2018)). Moreover, the deformations are not of interest in this chapter, since the main focus is on predicting the slope stability. The factors of safety F_s presented throughout this chapter are calculated with the expected values of all random variables. Excess pore water pressures during construction are taken into account in the calculation of the effective stress per slice. Seepage is considered through a gradient over the slices in the LEM model.

Figure 4.3 confirms that the slope stability increases with increasing consolidation from loading condition A to B, for dikes on clay (cases 1-3). For a sand foundation layer, we consider no significant excess pore pressures. Hence, there is no difference in F_s between loading conditions A and B in cases 4 and 5.

The most important observation in reliability updating context is that the stability at the end of construction is critical (F_s close to 1.0), or in any case more critical than during flood conditions. For dikes on a clay foundation (cases 1-3), the F_s in loading condition A is lower than the F_s in loading condition C, in almost all cases. The reason is that for dikes on a clay foundation the subsoil is not fully consolidated at the end of the construction, unlike for sand. Hence, cases 4 and 5 do not show a lower stability at the end of construction than during flood conditions.

For dikes on a thin soft soil foundation (case 2), the difference in F_s between the end of construction and flooding conditions is less than in case 1 and 3. The reason is that water pressures in the foundation sand layer during floods have a greater influence on the stability in case 2. For dikes on sufficiently thick soft soil foundation layers, we may expect that incorporating construction survival will yield an increased reliability estimate. Vice versa, for the dikes on sand subsoil, we do not expect a significant effect on the reliability, and therefore cases 4 and 5 are not considered in the further analysis.

4.3. RELIABILITY UPDATING USING CONSTRUCTION SURVIVAL

To incorporate construction survival in the probabilistic analysis of a dike in flood conditions, a practical approach is proposed based on probabilistic slope stability calculations using FORM and Bayesian updating (Section 2.2).

For the envisaged application, we define the stochastic variables $\mathbf{X}_a$ for a dike in the assessed flood conditions, and $\mathbf{X}_s$ for the survived construction. Notice that $\mathbf{X}_a$ and $\mathbf{X}_s$ contain the same variables, but they represent the states of the variables at different moments in time, namely the assessed flood conditions and survived construction. Many of them are time-invariant soil properties which are modelled as fully correlated between the assessed flood conditions and the survived construction. Other variables (such as pore pressures) are independent/uncorrelated between the assessed and survived situation.

Let F be the event of slope failure in the assessed loading condition B or C (i.e., daily or flood water level, respectively), where $g(\mathbf{x}_a)$ the domain of failure in the assessed loading condition, and $\mathbf{x}_a$ a realisation of $\mathbf{X}_a$. The probability of failure (further referred to as prior probability) is given by Equation 2.4, with $f_{\mathbf{X}_a}$ the probability density of $\mathbf{X}_a$. Similarly, let ε be the event of construction survival (the evidence), thus where $h(\mathbf{x}_s) < 0$ is the domain of survival with the survived loading conditions, and $\mathbf{x}_s$ a realisation of $\mathbf{X}_s$. The probability of survival is given by by Equation 2.4, with $f_{\mathbf{X}_s}$ the probability density of $\mathbf{X}_s$.

As the main attention is to the reliability update in this chapter, the direct method of Bayesian updating is used to incorporate construction survival in the probabilistic analysis (see Section 2.2.3). The posterior failure probability is thus calculated according to Equation 2.10 and Equation 2.11.

4.3.1. IMPLEMENTATION WITH CRUDE MONTE CARLO

The posterior probability can be estimated using CMC by counting the number of realisations in the failure domain, where we only take into account the realisations for which the construction has been survived. Thus if $g(\mathbf{x}_{a,i}) < 0$, where $g(\mathbf{x}_{s,i}) \geq 0$ for realisations $\mathbf{x}_{a,i}$ and $\mathbf{x}_{s,i}$. Using the indicator function $1[\cdot]$, we estimate the posterior probability as follows:

$$P(\mathrm{F}|\varepsilon) \approx \frac{\frac{1}{n}\sum_{i=1}^{i=n} 1\left[g(\mathbf{x}_{a,i}) < 0 \cap h(\mathbf{x}_{s,i}) < 0\right]}{\frac{1}{n}\sum_{i=1}^{i=n} 1\left[h(\mathbf{x}_{s,i}) < 0\right]}. \tag{4.1}$$

Although a single slope stability calculation to evaluate the limit state function g or h does not require much computation time, evaluating a large sample becomes computationally expensive – especially for low probabilities. Therefore, we propose an approximation to evaluate the outcome of $g(\mathbf{x}_{a,i}) < 0$ and $h(\mathbf{x}_{s,i}) < 0$. To that end, we first replace $g(\mathbf{x}_{a,i}) < 0$ and $h(\mathbf{x}_{s,i}) < 0$ by indicator functions $\chi_{F,i}$ and $\chi_{\varepsilon,i}$, respectively. Since indicator functions only take values 1 or 0, the AND-gate ($\cap$) can be replaced by the logical operation of a parallel system: a multiplication. The $1/n$ terms drop from the equation, and the posterior probability of failure is written as follows:

$$P(\mathrm{F}|\varepsilon) \approx \frac{\sum_{i=1}^{i=n} \chi_{\mathrm{F},i} \cdot \chi_{\varepsilon,i}}{\sum_{i=1}^{i=n} \chi_{\varepsilon,i}}. \tag{4.2}$$

We approximate the outcome of the indicators $\chi_{\mathrm{F},i}$ and $\chi_{\varepsilon,i}$ based on the prior probabilities for failure and survival. This is done by comparing realisations of the standard uniform variables ν_1 and ν_2 with the previously calculated probabilities $P(\mathrm{F})$ and $P(\varepsilon)$:

$$\chi_{\mathrm{F},i} = 1[\nu_{1,i} < P(\mathrm{F})], \text{and} \tag{4.3}$$

$$\chi_{\varepsilon,i} = 1[\nu_{2,i} < P(\varepsilon)]. \tag{4.4}$$

This principle is similar to generating random realisations using the inverse cumulative distribution function. The main advantage of this method is that we only have to carry out stability analyses to estimate $P(\mathrm{F})$ and $P(\varepsilon)$, and not $P(\mathrm{F}|\varepsilon)$. The $P(\mathrm{F})$ and $P(\varepsilon)$ may be evaluated for example by methods that require fewer model evaluations, such as FORM (Hasofer and Lind, 1974).

Since $\mathbf{X}_\mathrm{a}$ and $\mathbf{X}_\mathrm{s}$ are correlated due to the subset of time-invariant variables they have in common, the realisations of the limit state functions $g(\mathbf{x}_\mathrm{a})$ and $h(\mathbf{x}_\mathrm{s})$ are correlated. This correlation is taken into account by simulating equally correlated values of $\nu_{1,i}$ and $\nu_{2,i}$. In fact, this is the application of the EPM (Roscoe et al., 2015) for calculating system probabilities from the linearised limit state in the design point. The correlation ρ_{ν_1,ν_2} can be estimated from the influence coefficients α_k of the FORM calculation of $P(\mathrm{F})$ and $P(\varepsilon)$ for all parameters k, considering the autocorrelation of those parameters ρ_k, see Equation 4.5. The autocorrelation depends, among others, on which parameters are variable in time, as discussed in the next section. The effectiveness of the practical approach using indicators and the approximation of the correlation is discussed in Section 4.4.4.

$$\rho_{\nu_i,\nu_2} \approx \sum_k \alpha_k^a \cdot \alpha_k^s \cdot \rho_k \tag{4.5}$$

4.3.2. REDUCIBLE UNCERTAINTY

Although the survived and assessed situations are largely comparable, mainly because the soil properties are time-invariant, there are differences between the two situations. Therefore, we cannot guarantee with certainty that any flood situation will be survived when the construction was survived, even though the calculated F_s would suggest a less critical loading under flood conditions. The extent to which uncertainty in all stochastic variables in $\mathbf{X}$can be reduced (and reliability can increase) depends on whether the uncertainty in the parameters is epistemic, and hence, reducible, and whether the parameters are time-invariant, or not. To this end, we distinguish between random variables addressing predominantly epistemic uncertainty (due to a lack of knowledge) and random variables modelling aleatory uncertainty (i.e. actual randomness in time).

The uncertainty in soil properties is arguably of predominantly epistemic nature because most uncertainty relates to a lack of knowledge, for example due to limited soil investigation, see Chapter 3. Soil properties are also assumed to be time-invariant, i.e.

identical in the survived and assessed situation. For other stochastic parameters the values can differ between loading conditions, for example the phreatic level, the shear strength of unsaturated versus saturated clay, and the model uncertainty factor. Hence, information on the survived conditions does not give information about those parameters in flood conditions. The autocorrelation between realisations of variables in flood conditions and survived conditions is taken into account by the autocorrelation ρ_k in Equation 4.5 in Section 2.4.

4.4. CASE STUDY: CHARACTERISTIC DIKES

In the next sections, the proposed approach is applied to several case studies to analyse the impact of incorporating construction survival in the reliability estimate. This section examines the impact of incorporating construction survival in the reliability estimate for daily conditions and in flood conditions (loading condition B and C) for the characteristic dike profiles. To that end, the reliability (conditionally to daily and flood water levels) is estimated with and without construction survival. First the case input is described, then the reliability results are presented. Next, a sensitivity study regarding different survived conditions is presented to address the influence of OC and NC soil on the reliability update. Then the effectiveness of the proposed approach is discussed. The last section summarizes the influences of the specific dike profiles on the reliability updating.

4.4.1. CASE DESCRIPTION

The prior probability of slope instability (conditional on the considered loading condition) is estimated using FORM because it can reach accurate results with limited computational time, (e.g., De Koker et al. (2019); Huber et al. (2016); Kanning et al. (2017)). The prior probability distributions are shown in Table 4.1. In this study we use a log-normal distribution with $\mu = 1.0$ and $\sigma = 0.05$ for the model uncertainty factor for stability m_s. We assumed an unbiased model, since most of the model error contributions (3D-effects, appropriateness of strength data, strain-compatibility, etc.) are likely to cancel each other out (Ladd and Degroot, 2003). The standard deviation (SD) represents a central value in the range from 0.02 to 0.10 found in literature (e.g., Azzouz et al. (1983); Christian et al. (1994); Ladd and Degroot (2003)).

The soil properties S, m, POP, and ϕ_{cs} are assumed to be time-invariant parameters, for which full autocorrelation in time is assumed ($\rho_k = 1$). The phreatic level (and pore water pressures) are assumed to be time-variant, so these were independently modelled in the assessed and survived situation ($\rho_k = 0$). The model uncertainty is assumed to be uncorrelated in time ($\rho_k = 0$). Although the model errors in the LEM method and software will be correlated over time to some extent, the model uncertainty may also cover differences between the observed and future situation which are not correlated in time. Therefore, a realistic auto-correlation will not be 0 nor 1. We choose here the conservative side, to not overestimate the effect of updating in this regard. Along the same line, we assumed the drained strength parameter ϕ_{cs} and undrained strength parameters (S, m, POP) of the same soil are uncorrelated, although the drained and undrained strength could exhibit a correlation to some extent.

4.4.2. RELIABILITY RESULTS

Figure 4.4 shows the reliability index for the considered cases, with and without incorporating construction survival. The results are presented for the survived conditions where 50% of the pore pressures are dissipated at the end of the construction. Incorporating construction survival leads to an increase in reliability in all cases, but the reliability updating effect differs from case to case. For the daily water level, the posterior probability of failure is a factor 10 to 1000 lower than the prior probability of failure (0.45 to 1.7 in terms of the reliability index). For the flood water level, the reduction in failure probability is a factor 2 to 70 (0.04 to 1.3 in terms of the reliability index).

The impact on the reliability at flood conditions is lower than conditional on the daily water level. This is mainly a result of a lower correlation between the design situation and the survived situation, because the load effects (specifically the pore water pressures) are different between the situations. For example, in case of thin blanket layers, such as in case 2, pore water pressures in the sand foundation have a large effect on the shear strength, which makes flood loading conditions quite different from the construction conditions. Consequently, the survival of the construction provides little extra information about the stability in flood conditions, and therefore, the impact on the reliability is low.

The results, however, are presented conditional to the water level, whereas the total probability of failure is the probability-weighted sum of the conditional failure probabilities for all water levels in the relevant range; two of which are considered by loading condition B and C. To provide a rough estimate of the total annual failure probability for the characteristic dike cases, we combined the conditional failure probabilities $P(\mathrm{F})|h$ with the pdf of the water level f_h, using Equation 2.9. We use a Gumbel-distributed probability distribution of the water level based on the assumption that the flood level of loading condition C corresponds to an exceedance probability of 1/100 per year. Due to the larger effect of the survival information on lower water levels (which have the highest probability density), the annual reliability increases significantly, see Table 4.2. This demonstrates that the increase in the total reliability can be still significant, even though the effect of reliability updating on the conditional reliability for flood conditions is limited.

4.4.3. INFLUENCE OF OC AND NC SOIL ON THE RELIABILITY UPDATE

It is striking that the reliability update in all cases with NC soil is much higher than in the cases with OC soil. This is explained by the relatively high probability of failure for the dikes on weak NC soil considered in this example, compared to dikes on OC soil, see Table 4.2. This is in accordance with the F_s close to 1.0 during construction, see Figure 4.3. It confirms earlier findings that the reliability update depends among others on the criticality of the survived situation.

However, remarkably, it turns out that the update generally is higher in the cases with OC soil, at least when looking at roughly equally critical survived situations. This follows from a sensitivity study to the effect of different degrees of consolidation at the end of construction on the reliability update. Figure 4.5 shows the results of this study with first the probability of failure at the end of construction (a), and subsequently the updated reliability for daily (b) and flood conditions (c) when that survived degree of

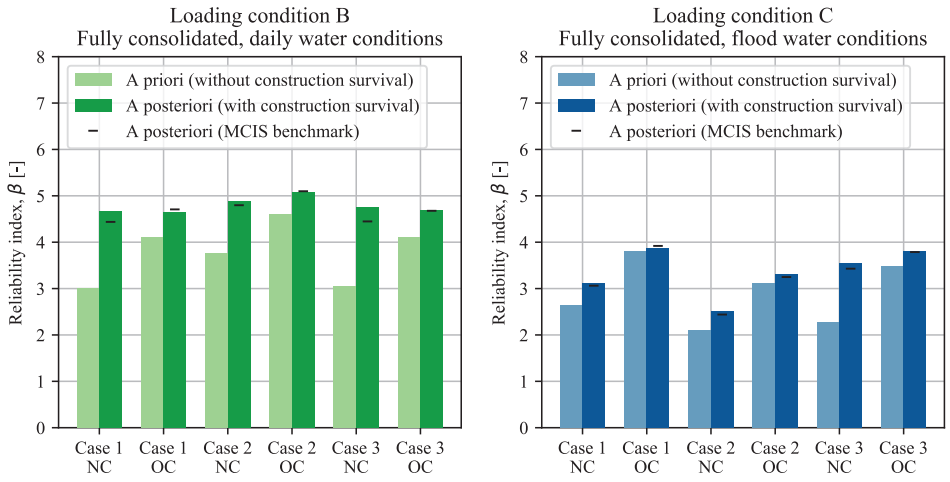


Figure 4.4: Prior and posterior reliability index of slope stability (with and without construction survival with 50% consolidation at the end of the construction, respectively). Results are shown conditional to daily (left) and flood water levels (right). The bars show the results obtained with the proposed simplified method in this chapter; the horizontal black lines show benchmark results obtained with MCIS.

Table 4.2: Failure probability at the end of construction, and prior and posterior annual reliability estimates of slope stability for the characteristic dike profiles considered.

	Failure probability P_f during construction	**Annual reliability index β^a of the assessed situation**		**Ratio prior vs. posterior failure probability**
	Loading condition A End of construction	A priori (without construction survival)	A posteriori (with construction survival)	
Case 1, NC	0.49	2.97	4.26	146
Case 1, OC	0.02	4.04	4.52	9
Case 2, NC	0.21	3.37	3.96	10
Case 2, OC	0.01	4.23	4.40	2
Case 3, NC	0.49	2.93	4.49	476
Case 3, OC	0.02	4.01	4.57	12

[a] probability-weighted sum of the conditional failure probabilities for all water levels

consolidation is incorporated.

From Figure 4.5a, it can be read at what degree of consolidation the probability of failure is more or less equal, and thus at which conditions the situations are equally critical. A failure probability of 0.01–0.05 is found at a degree of consolidation of 80% for the NC cases, and 50% for the OC cases in Figure 4.5a. When regarding these survived conditions specifically, the posterior failure probabilities for dikes on NC soil are approximately a factor 4 lower than the prior failure probabilities, compared to roughly a factor 10 for dikes on OC soil, see Figure 4.5b and c. The main reason for the larger update is

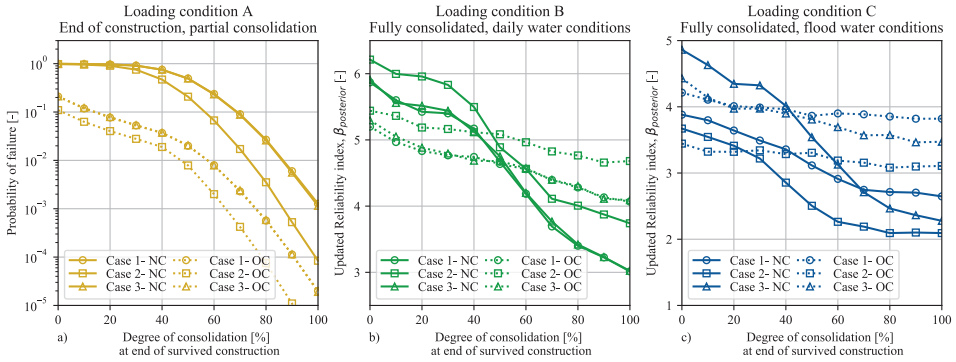


Figure 4.5: Influence of the survived degree of consolidation at the end of construction on the updated reliability for slope stability of different cases. The left graph (a) shows the probability of failure after construction, the middle graph (b) the updated reliability index conditional to daily water level, the right graph (c) conditional to flood water levels.

that the total uncertainty is larger for dikes on OC soil, which is discussed in more detail in the next section. The analysis further shows that significant reliability updates are possible, even when the probability of failure during construction is lower than 0.05; to our judgement a reasonable and generally accepted safety level during construction.

4.4.4. ACCURACY OF THE SIMPLIFIED METHOD

To demonstrate the accuracy of the proposed approach based on the indicator function (using Equation 4.2), the results were benchmarked with MCIS. The calculation of the failure probability is similar to Equation 4.1, but with importance weighting factors correcting for the shifted sampling distributions (shifted to the FORM design point of the prior analysis). The results of the proposed method are in reasonable accordance with the MCIS results, depicted by black horizontal lines in Figure 4.4.

From the MCIS it also follows that the correlation between realisations of the limit state functions $g(\mathbf{x}_a)$ and $h(\mathbf{x}_s)$ is very similar to the estimate based on Equation 4.5. For example for case 1 NC, the Pearson correlation coefficient between loading condition B and A based on MCIS is 0.72, The estimate using Equation 4.5 is 0.73. For loading condition C and A, the correlation coefficient is 0.56 based on MCIS, and 0.58 based on Equation 4.5. Notice that the correlation coefficient in the considered cases mainly depends on the choice about the autocorrelation of parameters (ρ_k). The contribution of a different importance of parameters in both calculation is minor. Similar results are found for the other cases. The proposed approach is therefore a convenient, practical alternative for estimating low posterior conditional probabilities of failure.

The Monte Carlo analysis further provides insights in how the information contained in the survival of construction leads to uncertainty reduction. Figure 4.6 shows the MCIS realisations for case 1 NC and OC, for the two parameters which have (a priori) the largest influence on the failure probability: the *POP* and the *S*-ratio of the clay layer. The figure highlights the samples of the MCIS analysis which are not consistent with the observation, for which $h(\mathbf{x}_{s,i}) > 0$. From the figure, two things are be concluded.

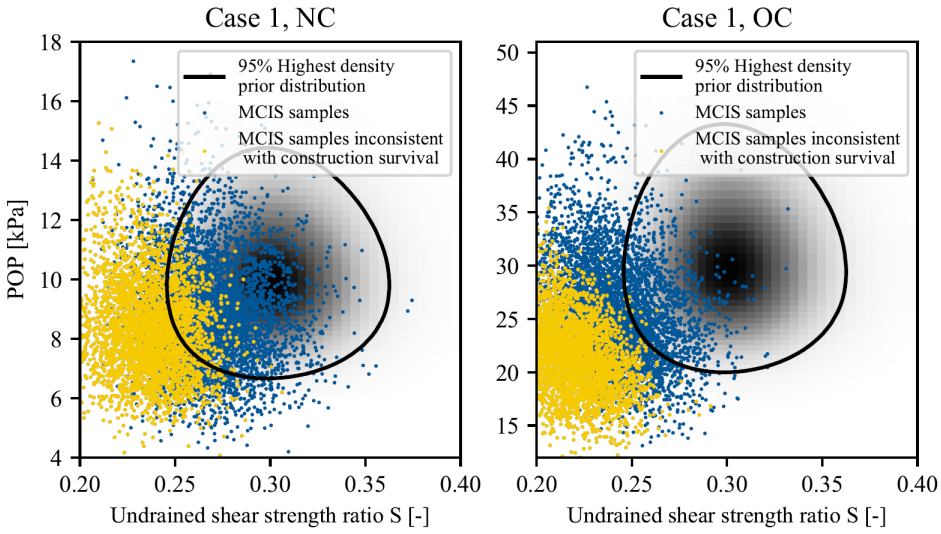


Figure 4.6: Joint probability density (grey shading) for undrained shear strength ratio (S) and POP values. Blue dots show all MCIS samples, yellow dots show only realisations that are not consistent with the survived construction (i.e. for which $h(\mathbf{x}_{s,i}) > 0$.

First, the survival information relates to realisations at the low end of the joint parameter distribution, reflected in a lower density of the tail of the joint parameter distributions. Specifically for the latter, it is not straightforward to reconstruct the posterior probability distributions including the survival information, without losing information about the induced partial cross correlation due to the exclusion of samples in the joint tail.

Second, there is a difference in uncertainty reduction between the NC and OC case. For the OC case the uncertainty reduction is mainly in the tail of the S and POP of the clay layer. For the NC case, the uncertainty reduction in POP is less, and the update relates mainly to the S parameter. The main reason is that the uncertainty in preconsolidation is large for OC soils, whereas preconsolidation does not play a role for NC soil as the preconsolidation is almost directly exceeded by the embankment raise. For the NC case, the shear strength is therefore only determined by the relatively less uncertain undrained shear strength ratio (S).

4.4.5. SUMMARY OF CASE STUDY RESULTS

Based on the presented results, the following conclusions are drawn regarding the influence of the considered characteristic dike cases on the reliability update:

- For dikes on a sand foundation (case 4 and 5), the construction is not a critical load because excess pore water pressures dissipate almost instantaneously during construction. Hence, no reliability update is expected from incorporating construction survival for these cases.

- In case 2, the presence of a relatively thin clay blanket layer leads to considerably different pore water pressures in the subsoil at flood loading conditions than during the survived conditions. This limited similarity between loading conditions results in a lower reliability update effect.
- For the specific cases considered, the effect of incorporating construction survival turned out to be generally larger for dikes on OC than on NC soil, for the same criticality of stability during construction. The main reason is the larger uncertainty in properties of OC soil, allowing for (and leading to) more uncertainty reduction. In the cases with dikes on NC soil, the uncertainty was smaller, thus less uncertainty could be reduced.
- No firm conclusions can be drawn about the influence of different dike materials. Although differences in the response to flood levels and difference in probability distributions are expected to have a different impact on the reliability update, the considered cases were too limited to draw conclusions. Moreover, for the considered cases, the main uncertainty is not in the dike material, but in the shear strength of the soft soil blanket. Hence, the reliability update in both cases is mainly due to uncertainty reduction in the soft soil parameters, not the dike material parameters.

4.5. CASE STUDY: EEMDIJK

This section demonstrates the application of the proposed approach to the Eemdijk test dike. The case study is already introduced in Section 3.5. In contrast to the previous sections, this case study uses the actual data from site investigation and pore water pressure monitoring.

4.5.1. CASE DESCRIPTION

The dike, built of clean sand, 5.3 m high and slope 1:2 (v:h), was constructed in multiple stages (see Figure 4.8) with no excessive deformations or cracks being observed. During construction, pore water pressures were continuously measured in the clay and peat layers at multiple locations. The undrained shear strength was estimated by CPTs using a site-specific transformation model, see Section 3.5. The other soil properties were obtained from laboratory tests (Deltares, 2018), prior to the construction of the test dike. The stochastic soil properties used for this case study are shown in Table 4.3.

4.5.2. RELIABILITY RESULTS

Following the proposed approach, we calculate the prior reliability (without construction survival) and posterior reliability (with construction survival) of the test dike. The design conditions comprise different flood water levels and fully consolidated subsoil, as if the dike had been built as flood defence. The survived conditions are based on the actual measurement data acquired during construction Lengkeek et al. (2019); Van der Krogt (2018). This involves data such as the geometry, phreatic level and the excess pore water pressures over the soil layers. The degree of consolidation below the dike was approximately 60%. The Uplift-Van model (Van, 2001) was used because of the possibility of uplift at the toe.

Figure 4.7: Aerial photo of the Eemdijk ring-shaped test dike, with slope instability of the conventional (ground) dike (left). The test dike with sheet pile, is located on the right side. Photo courtesy of Eric Feijten / NOS (Feijten , Photographer)

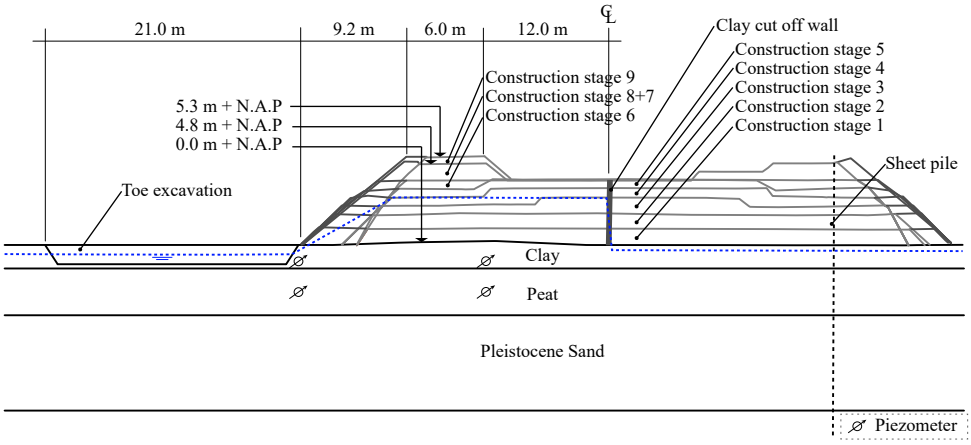


Figure 4.8: Schematic overview of cross-section A-A', see Figure 4.7. The blue dotted line indicates the phreatic level during the failure test. Heights relative to NAP.

Figure 4.9 shows that the failure probability for the daily phreatic level reduces significantly (factor 13 lower) by considering the survival of the last two construction stages. The effect of incorporating the survival information on the failure probability is negligible for design conditions with fully consolidated subsoil and a very high phreatic level, due to the low similarity with the survived conditions. This is in line with the determin-

Table 4.3: Soil properties of the Eemdijk test site Deltares (2018)

Soil	Property	Symbol	Distribution	Mean	Deviation	Type	Auto-correlation
Dike Sand	Friction angle	ϕ_{cs}	Log-normal	35.0	1.5	SD	1
Clay	NC shear stress ratio	S	Log-normal	0.38	0.05	SD	1
Clay	Strength increase exponent	m	Log-normal	0.91	0.02	SD	1
Clay below crest[a]	Undrained shear stress (CPT)	s_u	Log-normal	13.2	0.14[b]	CoV	1
Clay below toe[a]	Undrained shear stress (CPT)	s_u	Log-normal	6.2	0.14[b]	CoV	1
Peat	NC shear stress ratio	S	Log-normal	0.50	0.04	SD	1
Peat	Strength increase exponent	m	Log-normal	0.87	0.03	SD	1
Peat below crest[a]	Undrained shear stress (CPT)	s_u	Log-normal	24.6	0.19[b]	CoV	1
Peat below toe[a]	Undrained shear stress (CPT)	s_u	Log-normal	12.0	0.19[b]	CoV	1
Sand Aquifer	Friction angle	ϕ_{cs}	Log-normal	35.0	1.5	SD	1
Model uncertainty factor Uplift-van (Van Duinen, 2015)		m_s	Log-normal	0.995	0.033	SD	0

[a]The mean undrained shear strength under the dike crest differs from the toe as the area below the crest had been preloaded by an old, meanwhile removed dike.
[b]The uncertainty in the indirectly measured undrained shear stress (using CPT) before construction at a specific location is largely epistemic, since it includes a large transformation error (Van der Krogt et al., 2018; Van der Krogt and Schweckendiek, 2019).

Table 4.4: Factor of safety (F_s) in different loading stages and different phreatic levels. Mean values of the soil parameters are used.

	Phreatic level [m NAP]			
Loading stages	+0.25	+1.0	+2.0	+3.0
Construction, 4.8 m bank height	1.03			
Construction, 5.3 m bank height	1.02			
Failure test	1.02	1.00	0.95	0.87
Failure test + toe excavation	0.92	0.90	0.85	0.76
Fully consolidated state (hypothetical)	1.17	1.13	1.05	0.94

istic analyses reported in Table 4.4 as the factor of safety in the design conditions itself, is lower than the stability during construction.

In terms of the criticality of the survived situation (F_s close to 1.0), and the lower similarity of the extreme situation with respect to the survived situation, the results of the Eemdijk case study are most similar to Case 2 (thin blanket layer) of the characteristic dikes considered in Section 4.4. The effect of the reliability update for the Eemdijk case study is a little less than for the characteristic case study, presumably because the prior uncertainty is lower due to relatively extensive site investigation. Nonetheless, the transformation uncertainty is still present and reduces if we incorporate the information of the survived construction.

These results, however, are conditional failure probabilities. Therefore, we combine the conditional reliability indices in Figure 4.9 with a pdf of the phreatic level (the phreatic level virtually being the response to flood loading). A Gumbel distribution with annual exceedance probabilities of 1/2 and 1/100 for phreatic levels of +1.0 m NAP and +2.0 m NAP, respectively, results in a total prior annual reliability index of 1.59, and a total posterior annual reliability index of 2.51. This is a reduction in failure probability of almost a factor 10.

Similar to the cases of characteristic dikes in Section 4.4, the reliability estimate

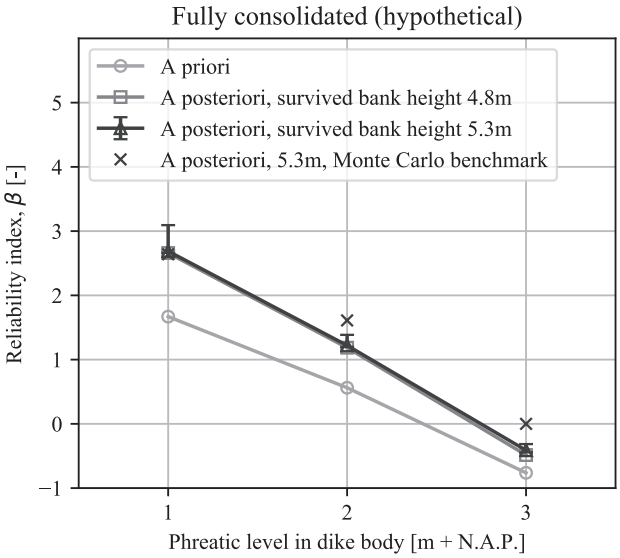


Figure 4.9: Prior and posterior reliability index conditional to the phreatic level of a hypothetical design situation with a fully consolidated subsoil. Error bars show the influence of uncertainty in the measured pore water pressures (one standard deviation lower or higher) on the calculated posterior reliability.

weighted with pdf of the water level improves significantly, despite an insignificant reduction of the conditional failure probability at high water levels (i.e. design conditions). Again, the overall effect is due to the reduced conditional probability of failure for lower water levels, which are more similar and less critical than the observed loading conditions.

The effect of uncertainty in the measured excess pore water pressures at the survived loading conditions is investigated through a sensitivity analysis. From the data, the variation of the excess pore water pressure across the site is estimated 3 kPa. The influence on the posterior reliability of a 3 kPa higher or lower pore water pressure appears to be very limited, see the error bars in Figure 4.9. In addition, the results obtained with CMC using Equation 4.1, are in good accordance with the proposed approximation method.

4.5.3. IMPROVED PREDICTION FOR FAILURE TEST

Instead of the dike being an actual flood defence (on fully consolidated soil as considered above), the failure test was executed directly after the last construction stage, without full consolidation. The slope failure of the test dike was induced by increasing the phreatic level in the dike body, and additionally by excavating a trench adjacent to the dike toe, see Figure 4.8. Consequently, the stability in terms of F_s during the failure test was lower than during construction, see Table 4.4. Nevertheless, the dike was critically loaded during construction (F_s was close to 1.0), thus considering construction survival should also improve the predicted performance during the failure test.

The reliability update for the failure test is depicted by the fragility curves (empirical

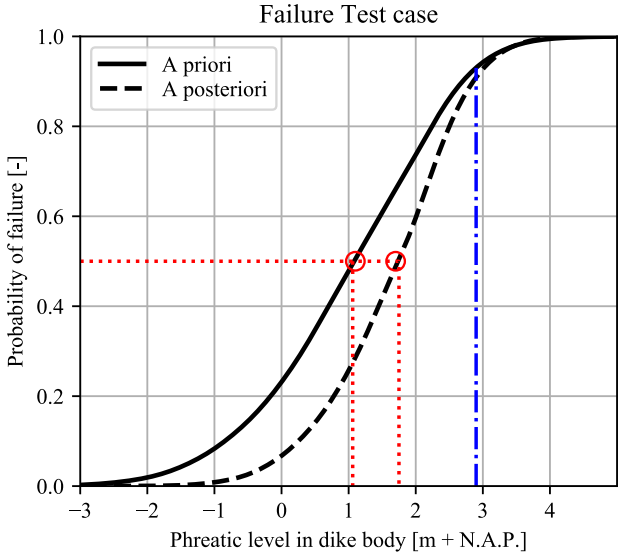


Figure 4.10: Prior and posterior fragility curves of the failure test. Red lines indicate the prior and posterior estimate for the phreatic level triggering failure (at which $P(\mathrm{F}) = 0.5$). The blue line indicates the phreatic level at which the test dike actually failed.

cdfs) in Figure 4.10. The empirical cdfs does not show a sharp cut-off at the survived phreatic level during construction, but a gradual reduction of probability density at low phreatic levels, and a redistribution of density mass over the higher phreatic levels, due to the differences in loading conditions between the failure test and the survived construction. Although the effect of considering construction survival is limited for the overall reliability (reduction by factor 2), the graph clearly shows the reduction of uncertainty by incorporating the construction survival. For example, it can be derived that the phreatic level at which $P(\mathrm{F}) = 0.5$ (i.e. approximating the most likely loading condition triggering failure) increased from +1.0 m NAP (a priori) to +1.7 m NAP after updating.

In the test, the slope eventually failed at a phreatic level of +2.9 m NAP. Hence, we may conclude that the prediction of failure could be improved by using the information from the construction phase, even though we remain with an under-estimation of the slope resistance after updating.

4.6. CONCLUSION

As survival information promises to lead to reliability updates, and specifically the information contained in the survival of the often critical construction phase, we strived to answer the following sub-question:

> *How does the performance information contained in the survival of the construction phase affect the reliability estimates for slope stability of dikes?*

This chapter developed a practical method to incorporate the observation of construction survival in the reliability analysis. Depending on the subsoil and loading conditions, incorporating this survival information can lead to a significantly higher posterior reliability, especially for dikes on undrained subsoil. The main reason is that the construction of dikes on soft soils is a critical loading condition, and the stability directly after construction is often lower than during the design flood conditions.

For several characteristic dike profiles, the (conditional) probability of failure reduced by a factor of 10 to 1000 for relatively low water levels. For high water levels representing design flood conditions, the impact was less significant with a reduction by a factor of 2 to 10. Primarily as a result of lower correlation or similarity between the survived and the assessed conditions. Nevertheless, the total dike reliability estimate (e.g., annual) can improve significantly because it considers the entire range of potential flood levels. Herein, extreme flood stages have a low probability of occurrence, and hence, a lower weight in the total reliability estimate.

The results obtained from the proposed approximation approach based on FORM calculations agree well with results obtained from Monte Carlo simulations, for conditional probabilities for the cases presented in this chapter. Therefore, the proposed approach is a convenient, practical alternative for estimating low probabilities. The use of surrogate-models for slope stability calculations (e.g. Jiang et al. (2015); Li et al. (2016)) may allow evaluation of Bayes rule (Equation 2.11) directly, improving computational efficiency while using less approximations.

The case studies provide insight into the factors determining a significant reliability update. We highlight two of those. The first prerequisite to be met is that the assessed and survived situations need to be a high degree of similarity, in the simplified method depicted by a high correlation. A second condition is that sufficient knowledge is available about the survived conditions, because large uncertainty about the survived conditions (due to limited availability of data) reduces the impact of the survival observation. An example is the contribution of unsaturated conditions to the shear strength, which was neglected in the examples. The influence of unsaturated conditions on the reliability update is investigated by a sensitivity study, demonstrating that additional unsaturated strength reduces the effect of reliability updating, but not completely. To assess whether incorporating construction survival can improve the reliability estimate in practical projects, we recommend analysing first how critical were the loading conditions during construction, compared to the design loading conditions, in terms of the respective factors of safety. And thereafter, to consider whether the two loading conditions are sufficiently similar, in terms of the critical slip circle and the influence of the parameters involved.

The results indicate that dikes on soft soil blankets that survived the construction will generally have a higher reliability than the reliability of the design based on prior knowledge. For existing dikes, incorporating the information of construction survival in the reliability analysis can therefore significantly improve safety assessments. As the survival information is typically available over distances longer than a slip plane, the results (e.g. posterior reliability or updated estimates of the local average properties) are representative for horizontal sections that can be considered as homogeneous in terms of the schematized properties (typically in the order of hundreds of meters). To update

the reliability of larger sections e.g. when survival information is available along longer stretches, Bayesian analyses must be performed for multiple cross-sections.

For new dikes and dike reinforcement projects, designs can be optimized by anticipating the survival of the construction stage, using pre-posterior analysis, see Appendix A. The design, the criticality of the loading condition during construction, the required monitoring (e.g. pore water pressures), and contingency actions (e.g. stopping criterion) become then elements in the optimization. Bayesian decision theory (Raiffa and Schlaifer, 1961) and the Observational Method (Peck, 1969; Spross and Johansson, 2017) provides an appropriate risk-based framework to consider whether the expected benefits in the design outweigh the additional risk of failure during construction. Chapter 6 further elaborates on this concept.

5

MULTI-SOURCE PERFORMANCE INFORMATION FOR SLOPE RELIABILITY UPDATING

Science never solves a problem without creating ten more.

George Bernard Shaw

By considering observations of the performance of dikes, epistemic uncertainty reduces and estimates of the failure probability improve. Uncertainty can be further reduced by considering multiple performance observations, from different sources. This chapter investigates the effect of settlement measurements and survival during the construction of dikes on the reliability estimates by using Bayesian analysis. We also explore workable approximation methods Bayesian analysis with geotechnical models used in practice, for which evaluation of a large number of realisations is computationally challenging.

Contents

5.1. INTRODUCTION

In Chapter 4 it was shown that incorporating construction survival information using Bayesian Updating increases reliability estimates because survival of critical loads typically contains information about the minimum strength of the soil and affects the tail of the joint probability density function (pdf) of the parameters. However, still considerable uncertainty in the parameters remains, reflected by the posterior (updated) probability distributions.

Uncertainty can be further reduced by incorporating additional performance information that is available during the survived construction. Another source of information of the embankment construction are settlements during and after the raise of the embankment. The settlement is among others influenced by the preconsolidation of the soil, see Figure 5.1. In reliability analysis for slope stability where s_u is calculated within the LEM model using the SHANSEP formulation (Equation 2.16), preconsolidation is one of the main uncertain parameters. Settlement measurements over time can reduce uncertainty in the preconsolidation stress, and hence update the reliability estimate for slope stability.

In addition, combining the observations of settlement and survival should reduce uncertainty further because they are from different sources. Furthermore, settlements are 'equality information' which should lead to uncertainty reduction and updated values for best estimates (shifted mean), contrary to survival information which is 'inequality information' and for which the update mainly relates to the tail of the joint probability distributions. Settlement and survival information are thus complementary sources of information, and the combination of equality and inequality information may be specifically useful for uncertainty reduction.

Some examples of incorporating multiple measurements in the application of geotechnical engineering are given by Vardon et al. (2016); Li et al. (2018); Peng et al. (2014); Zheng et al. (2018); Sun et al. (2019). None of the examples, however, considers the combination of equality and inequality performance information. Moreover, the examples are not focused on improving failure probability estimates. This chapter examines how multiple performance observations such as survival and settlement during

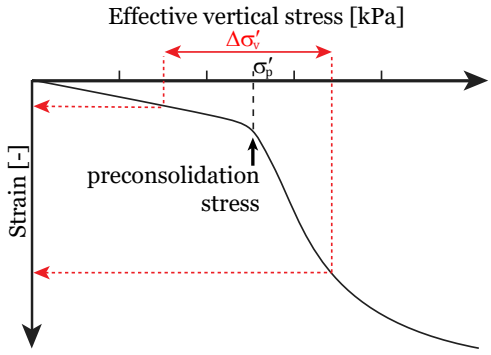


Figure 5.1: Typical settlement development of soft soil (depicted by the strain of a soil layer) due to effective stress increase $\Delta\sigma'$. The preconsolidation stress is indicated with σ'_p.

the construction phase can improve reliability estimates, using Bayesian updating. The research question is formulated as follows:

What is the effect of combining survival and settlements (multiple performance observations) during the construction of dikes?

To answer this question, we estimate the failure probability with single and multiple performance observations, using Bayesian updating. Straightforward reliability methods are ineffective for estimating low probabilities of failure and multiple observations due to the computation time of limit equilibrium state models for slope stability analysis and the models for settlement prediction (simple geotechnical models with computation times of at least several seconds), contrary to closed form solution as used in e.g. Spross and Larsson (2021). In order to achieve practically tractable computation times we use approximations for estimating the posterior probability of failure conditional to all observations: $P(\mathrm{F}|\varepsilon_1,...,\varepsilon_n)$. Here F indicates failure (instability), and $\varepsilon_i,...,\varepsilon_n$ the observations of survival and settlements.

Among the various reliability methods (of which some are introduced in Section 2.2.2), there is no obvious candidate to solve the reliability updating problem with multiple observations of different sources and different types. While not the main goal of this chapter to make an extensive performance comparison of methods for estimating the posterior probability, we investigate the most interesting possibilities to examine the effect of incorporating multiple performance observations, from multiple sources, and being of the equality and inequality type. In addition, we consider methods that, in addition to be feasible for academic research, are also practically workable in engineering practice, i.e. that the analysis is performed within one working day on a desktop computer, and is sufficiently accurate to compare with reliability requirements.

We consider subset simulation as one of the most effective simulation methods for estimating small probabilities, also dealing with Bayesian Updating (Straub (2011); Straub and Papaioannou (2014); Betz et al. (2018)). Secondly we consider the MultiNest algorithm (Feroz et al., 2009) as particularly promising method to efficiently estimate the total evidence, with low likelihoods. Finally, we follow an approach using a kriging-based metamodelling approach for reliability analysis, Efficient and Robust Reliability Analysis for Geotechnical Applications (ERRAGA), which drastically decreases the number of model evaluations (Eijnden et al., 2021). The comparison with CMC serves the primary goal of investigating the impact of considering the combination of survival and settlement information. In addition, the CMC serves a secondary goal to give recommendations about the tested reliability methods, for application of Bayesian Updating in geotechnical engineering practice.

This chapter is structured as follows: Section 5.2 outlines the different methods and implementations to calculate the posterior probability of failure. The envisaged methods are applied to a case study in Section 5.3, to examine the effect of incorporating survival and settlement observations for the case study of the Eemdijk test dike. Section 5.4 contains a further interpretation of the results and findings, and 5.5 finally answers the research question and gives recommendations and directions for further developments.

5.2. INCORPORATING SURVIVAL AND SETTLEMENT OBSERVATIONS

To examine the effect of considering survival and settlement observations, we recalculate the probability of failure conditional to both observations. As in Chapter 4, F is the event of slope stability failure, ε_1 the event of construction survival, and ε_2 the observation of settlement during construction. To estimate the posterior probability of failure $P(\mathrm{F}|\varepsilon)$, where $\varepsilon = \varepsilon_1 \cap \varepsilon_2$, three limit state equations are relevant: $g(\mathbf{x})$ for failure (Equation 2.18), $h_1(\mathbf{x})$ for construction survival (Equation 2.18), and $h_2(\mathbf{x})$ for settlement observations during construction, derived from Equation 2.12:

$$h_2(\mathbf{x}) = u - \Phi^{-1}\left(c \cdot L\left(\mathbf{x}, s_{\text{obs}}\right)\right). \tag{5.1}$$

Herein, is $L(\mathbf{x})$ the likelihood of observing a settlement:

$$L(\mathbf{x}) = \varphi\left(\frac{s(\mathbf{x}) - s_{obs}}{\epsilon_{\text{sett}}}\right), \tag{5.2}$$

with $s(\mathbf{x})$ the calculated settlement for a realisation $\mathbf{x}$ and s_{obs} the observed settlement. The function φ is the pdf of the standard normal distribution, and ϵ_{sett} is the standard deviation of the measurement error. This implies a normal distributed, unbiased measurement error, which is independent from measurement to measurement. The constant c in Equation 5.1 is a scaling factor, defined by the reciprocal of the maximum likelihood, thus $c = \epsilon_{\text{sett}}/\varphi(0)$. In case multiple independent measurements are available, the total likelihood is calculated by the multiplication of the individual likelihoods $\prod_i^n L(\mathbf{x}, s_{\text{obs},i})$.

According to Equation 2.11, the posterior probability is evaluated by:

$$P(\mathrm{F}|\varepsilon_1 \cap \varepsilon_2) = \frac{P(g(\mathbf{x}) < 0 \cap h_1(\mathbf{x}) < 0 \cap h_2(\mathbf{x}) < 0))}{P(h_1(\mathbf{x}) < 0 \cap h_2(\mathbf{x}) < 0))}. \tag{5.3}$$

To evaluate Equation 5.3 with simulation methods (Monte Carlo and subset simulation), we use the max of the limit state realisations to evaluate the $\cap$-gate.

$$P(\mathrm{F}|\varepsilon_1 \cap \varepsilon_2) = \frac{P(\max\left(g(\mathbf{x}), h_1(\mathbf{x}), h_2(\mathbf{x})\right) < 0)}{P(\max\left(h_1(\mathbf{x}), h_2(\mathbf{x})\right) < 0)}. \tag{5.4}$$

For the implementation with Crude Monte Carlo, the probability is calculated using Equation 5.5, with the indicator function $1\,[\cdot]$.

$$P(\mathrm{F}|\varepsilon) \approx \frac{\frac{1}{n}\sum_{i=1}^{i=n} 1\left[\max\left(g(\mathbf{x}_\mathrm{i}), h_1(\mathbf{x}_\mathrm{i}), h_2(\mathbf{x}_\mathrm{i})\right) < 0\right]}{\frac{1}{n}\sum_{i=1}^{i=n} 1\left[\max\left(h_1(\mathbf{x}_\mathrm{i}), h_2(\mathbf{x}_\mathrm{i})\right) < 0\right]}. \tag{5.5}$$

In the implementation with subset simulation, the numerator and denominator in Equation 5.4 are calculated separately and divided afterwards. Within the implementation using a metamodel, the limit state functions are replaced by surrogates $\hat{g}(\mathbf{x})$, $\hat{h}_1(\mathbf{x})$, and $\hat{h}_2(\mathbf{x})$, using a Kriging-based metamodel, see Section 2.2.5.

The implementation to calculate the failure probability using MultiNest (Feroz et al., 2009) is slightly different because the MultiNest algorithm is aimed at calculating the total evidence. Since the realisations of the nested sampling comply with the posterior distribution, and the weights w_i are already normalized with the total evidence ($\sum w_i = 1$),

the posterior failure probability is calculated by summing the weights of the realisations in the failure domain:

$$P(\mathrm{F}|\varepsilon) = \sum_{i=1}^{i=n} w_i \cdot 1\left[g(\mathbf{x_i}) < 0\right] \tag{5.6}$$

5.3. CASE STUDY: EEMDIJK

This section describes the results of incorporating settlement information and survival information during the staged construction of the Eemdijk test dike. The case study considers the stability at two locations across the test site (location 38 and 39, see Figure 3.10). At both locations the dike survived the construction and settlements are measured. The case study is already introduced in Section 3.5 and 4.5.

5.3.1. CASE DESCRIPTION

Prior information of the geotechnical parameters of the entire site, and the information of the staged construction phasing is taken from Deltares (2018); Van der Krogt (2018). The vector **X** denotes the 18 stochastic time-variant and time-invariant variables (modelled by the autocorrelation) for soil strength and stiffness in the three soil layers, see Table 5.1 and Table 4.3.

Contrary to 4.3, the strength-increase exponent m and properties of the sand aquifer are taken as deterministic value (mean) because the influence is negligible. The model uncertainty of the 2D plane strain settlement model is taken by a Log-normal distribution with mean 1.0 and standard deviation 0.05, in accordance with the results of Muhammed et al. (2020), in absence of values for soft soil in ISSMGE-TC304 (2021).

The stiffness parameters are derived from (K0-)constant rate of strain tests. A cross-correlation was found between the data points of the laboratory test results for the isotache model parameters a, b, and c, which are modelled accordingly in **X**, see Table 5.2. The consolidation is calculated using Darcy's consolidation model using a strain

Table 5.1: Stiffness properties of the Eemdijk test site Deltares (2018)

Soil	Property	Symbol	Distribution	Mean	Deviation	Type	Auto-correlation
Clay	Pre-overburden pressure[a]	POP	Log-normal	17.1	5.6	SD	1
	Isotache swelling index	a	Log-normal	0.014	0.004	SD	1
	Isotache compression index	b	Log-normal	0.155	0.025	SD	1
	Isotache creep constant	c	Log-normal	0.011	0.004	SD	1
	Initial permeability	k_{e_0}	Log-normal	$1.0 \cdot 10^{-9}$	0.5	CoV	1
	Permeability strain factor	c_k	Log-normal	0.3	0.3	CoV	1
Peat	Pre-overburden pressure[a]	POP	Log-normal	8.7	5.2	SD	1
	Isotache swelling index	a	Log-normal	0.045	0.013	SD	1
	Isotache compression index	b	Log-normal	0.332	0.048	SD	1
	Isotache creep constant	c	Log-normal	0.035	0.007	SD	1
	Initial permeability	k_{e_0}	Log-normal	$1.0 \cdot 10^{-8}$	0.5	CoV	1
	Permeability strain factor	c_k	Log-normal	0.26	0.3	CoV	1

[a]POP prior to the construction of the dike.

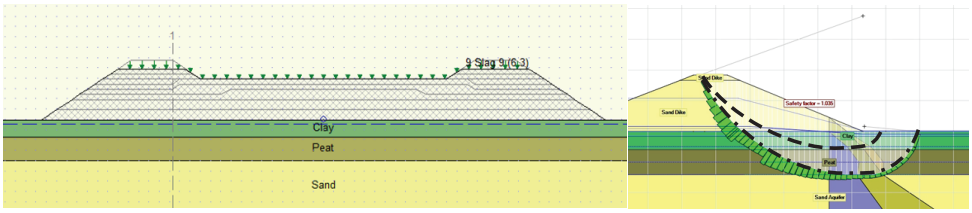


Figure 5.2: Staged construction settlement model and stability model. Dashed and dash-dotted lines indicate the shallow and deep failure mode for the slope stability. Clay layer is indicated by a green color, the peat layer is indicated by an olive color.

dependent permeability that decreases with the void ratio.

Table 5.2: Correlation between isotache parameters for clay (left) and peat (right), found between the data points of the laboratory test results.

Clay	a	b	c
a	1	0.21	0.47
b	0.21	1	0.85
c	0.47	0.85	1

Peat	a	b	c
a	1	0.53	0
b	0.53	1	0
c	0	0	1

The probability density functions of the soil parameters describe the uncertainty of the locally unknown depth-average value. As the uncertainty was estimated from laboratory tests from the entire site, it includes spatial variability. However, the local average is constant over time so the uncertainty in the local average is reducible. The uncertainty furthermore includes statistical uncertainty due to a limited number of measurements.

For the survival observation, we consider only the most critical construction stage at t=147 days, with the survived conditions as described in Section 4.5.2. The stability is assessed at t=224 days; the day the dike was tested, with the phreatic level at 1.0 m NAP. We evaluate the effect of survival and settlements for a fixed slip plane through the dike and clay layer underneath, see Figure 5.2. The impact of other potential slip planes is discussed in Section 5.4.1. Notice that the prior reliability estimate differs from the results in Section 4.5.2 because the POP in this study is based on laboratory tests only, and not based on local CPTs.

The settlement is calculated in 20 time steps from the start of the construction (t=0 days) until the successful construction of the full embankment (t=147 days), taking into account all construction stages. For the settlement observation we consider the settlement at both locations, at $t = 94$ days and $t = 147$ as independent measurements. The measurement error of the settlements (ϵ_{sett}) is taken by a CoV of 0.02, based on the reading accuracy of the measurement and deviation (root-mean-square error) of the individual measurements compared to the calculated settlement curve with the best fit in Figure 5.3.

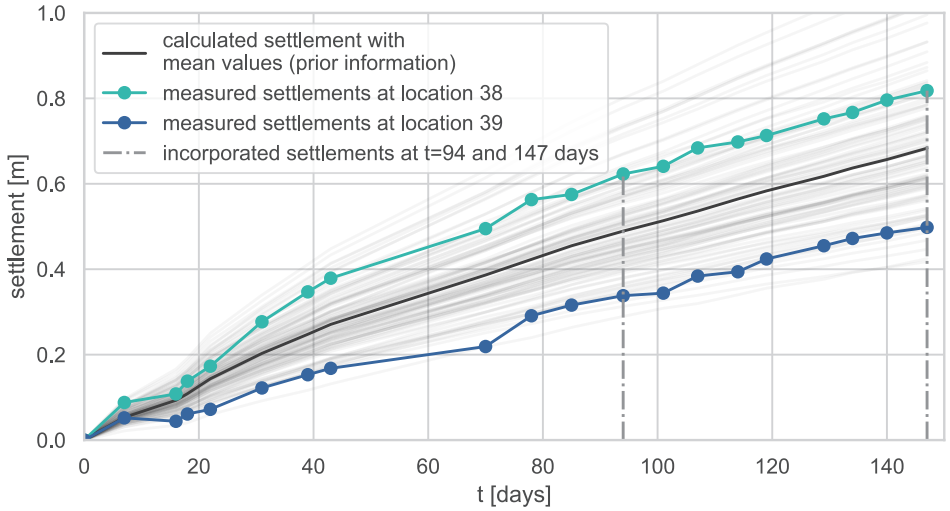


Figure 5.3: Calculated settlements for 100 realisations based on prior information (grey), calculated settlements based on the expected values (black), and the actual measurements of two locations at the site. For the Bayesian Updating, only the measurements at t=94 and t=147 days are used, indicated by dash-dotted lines.

5.3.2. RESULTS

Following the implementation described in Section 5.2, the reliability for slope stability was estimated for two different locations, incorporating the survival and settlement information at each location. The subset size for subset simulation and the number of live points for MultiNest were set to 1000 , such that a reasonable computation time was reached on normal workstation (maximum 1 day calculation time). For the metamodel a 60 initial realisations were selected, and 60 additional realisations. For Crude Monte Carlo 800 000 realisations were taken to obtain a sufficiently reliable benchmark (3 weeks calculation time).

PRIOR STABILITY AND SETTLEMENT ANALYSIS

The prior reliability index β for a slip circle through the clay layer was found 1.93 using Monte Carlo. The subset simulation, the analysis with the MultiNest algorithm, and the analyses with the metamodel were repeated 4 times to evaluate the variability and potential bias related to the chosen methods. The repeated results show some scatter, but on average the results agree well with the Monte Carlo. The results obtained with the metamodel show the least scatter among the different runs, see Figure 5.4. The calculated settlements for 100 realisations based on the prior information are shown grey in Figure 5.3. The black line indicates the calculated settlement for the expected values of the stochastic parameters.

POSTERIOR STABILITY ANALYSIS

To update the failure probabilities for stability at the two locations across the site, we considered the construction survival and the settlement measurements at the two locations.

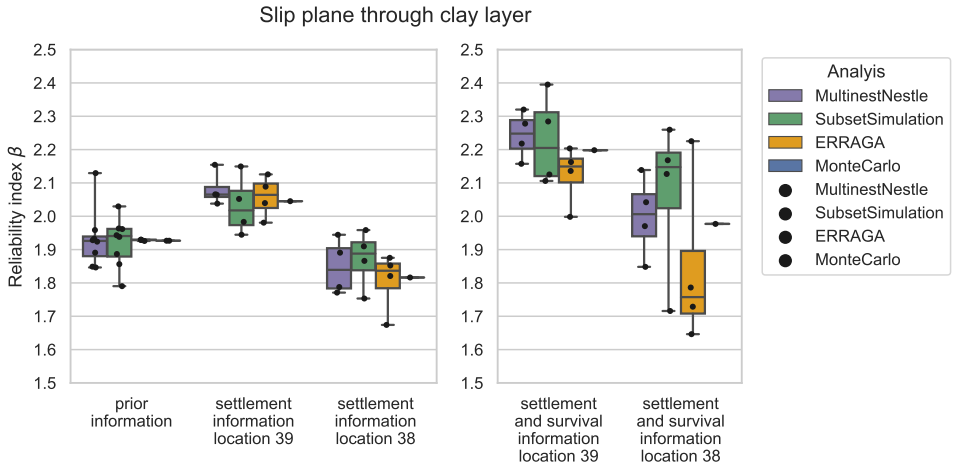


Figure 5.4: Reliability result for slope stability based on prior information, and posterior information, incorporating settlement measurements (left) and settlement and survival information (right). Results of 4 independent runs are shown by dots, and in boxplots. Results are conditional probabilities (daily water level).

5

The measured settlements are shown in Figure 5.3, one location representing an above average, and one representing a below-average settlement compared to the expected value.

Incorporating the settlement measurements of location 39 (less settlement than expected) increases the reliability for stability at that location (the cross-section indicated in Figure 3.10), while incorporating the settlement measurements of location 38 (more settlement than expected) decreases the reliability estimate of the local stability, see the left panel in Figure 5.4. This is expected because stability and settlement are typically negatively correlated through the influence of the preconsolidation.

Due to the Bayesian updating, many realisations were rejected from the Monte Carlo reliability analysis. From the only 20 000 samples which were not rejected, approximately 400 samples failed, so the numerical accuracy is approximately $V_{\hat{P}_f} \approx 0.05$. Each result of the 4 repeated analyses with subset simulation, MultiNest, and the metamodel is plotted with a black dot in Figure 5.4, and for better interpretation the results are also shown in boxplots. The middle lines of the boxplots describe the median values of the estimated reliability index.

The impact of the settlement measurements on the reliability estimate for stability is limited for this case. The increase and decrease of the failure probability is approximately 25% (from 0.027 to 0.035 and 0.020, respectively). The decrease in failure probability is relatively low due to the low correlation between stability and settlement in the considered case (Pearson correlation coefficient −0.10). Though the results of subset simulation, MultiNest, and ERRAGA show some scatter among different runs, the median value seems unbiased compared to the Monte Carlo benchmark. The scatter of subset simulation is consistent with the results reported in Straub et al. (2016).

Combining the information of settlement and survival of the most critical construc-

tion stage, leads to further uncertainty reduction. For the stability at locations 38 and 39, the posterior reliability considering both settlement and survival information is higher than the posterior with settlement information only, see the right panel in Figure 5.4. For location 38 the survival information compensates for the decrease in reliability due to settlement information, compared to the situation where only only prior information is used. Hence, combining different sources of performance information still leads to lower posterior failure probabilities. The results obtained with MultiNest and Subset Simulation show some scatter but seem unbiased, the results obtained with ERRAGA seem to under-predict the reliability slightly.

UNCERTAINTY REDUCTION

The CMC analysis provides insight into the uncertainty reduction and the resulting reliability increase for stability. The settlement measurements mainly lead to a small shift of the cumulative density function, to the lower side in case of high settlements (location 38), and to to the higher side in case of low settlements (location 39), see the left panel in Figure 5.5. The right panel in Figure 5.5 visualizes the effect of survival: where settlement information tends to shift the density function, survival information tends to reduce the density in the tail of the distribution, see the dark grey line in Figure 5.5. The dark blue line indicates the combined effect of settlement measurements and survival information, showing that both the equality and inequality data contribute to the uncertainty reduction at the lower tail of the cdf of the factor of safety F_s.

The little influence of the settlement observation on the stability in this case study, is also reflected in the posterior probability density functions of the parameters after incorporating settlement measurements, see Figure 5.6. The largest uncertainty reduction appears in parameters which have no influence on the slope stability, such as the initial porosity (e_0) and the permeability (k_{e_0}). These parameters are updated more than the preconsolidation (POP), mainly due to the high prior uncertainty.

Note that settlements measurements can fluctuate over shorter length scales. This limits the representativeness a measurement at a location and hence the result of the Bayesian analysis to distances within the auto-correlation length of the settlements. For use of settlement measurements for reliability updating of longer reaches of dike, multiple settlement measurements are thus necessary over a shorter distance.

5.4. DISCUSSION

5.4.1. OTHER SLIP PLANES

The results in Section 5.3.2 are, conditional to a fixed slip circle through the clay layer. An instability can, however, take any form, so ideally we evaluate the stability of all slip planes. To check whether other slip planes have a considerable contribution to the failure probability, we analysed a second potential slip plane: a deep failure mode intersecting the clay and peat layer of the subsoil, see the dash-dotted line in Figure 5.2. The prior reliability conditional to this slip plane is 2.70. The contribution of this slip plane to the total failure probability is thus negligible.

The impact of considering settlements for this deep slip plane is similar to the results for the shallow slip plane, see Figure 5.7. As the stability analysis now benefits also from uncertainty reduction of the POP of the peat layer (see Figure 5.6), the reliability update

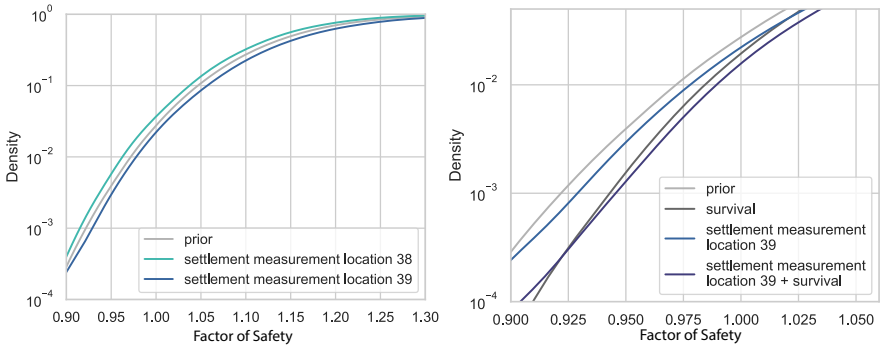


Figure 5.5: Fitted cumulative density (cdf) of the factor of safety with and without settlement information (left), and the combined effect of low settlement observation and survival information. Shallow failure mode through clay layer.

is a little larger. The increase and decrease of the failure probability is approximately 25-50% (from $3.5 \cdot 10^{-3}$ to $5.5 \cdot 10^{-3}$ and $2.7 \cdot 10^{-3}$ for location 38 and 39 respectively).

It strikes that the MultiNest analyses (purple in Figure 5.7 show much more scatter than the results with subset simulation and ERRAGA, for the considered case. This results in inaccurate reliability estimates. The reason is that the reliability estimate using MultiNest is heavily dependent on the number of failed samples in the initial set of live points (1000). As the (prior) reliability for the deep failure mode is already much higher than the shallow failure mode, there is a very low number of failed samples in the first set of live points (only 2, see the left panel in Figure 5.8). Because the algorithm samples only new realisations with a higher likelihood, hardly any new samples in the failure domain are simulated, contrary to subset simulation (see the right panel in Figure 5.8).

5.4.2. PERFORMANCE OF RELIABILITY METHODS

Even with the relatively simple geotechnical models we used in this chapter (which require approximately 1 s to run), the total computation time of the analyses is long but acceptable, see Figure 5.9. In engineering practice, however, more advanced geotechnical models are often used (e.g. with Finite Element Methods (FEM)), requiring more computation time. Computation times in the order of several days for a subset simulation or analysis MultiNest may be acceptable for academic research, but are not desirable in engineering practice. Basically, only implementations with metamodels seem suitable for applying Bayesian analysis in those situations. Actually, when a manifold of observations are to be considered in the analysis, implementations with metamodels seem the only realistic option. The next subsections give specific recommendations per method.

SUBSET SIMULATION

We chose an implementation in which we combine the limit states for failure, settlement measurements, and survival into one limit state function, and where we calculate the numerator and denominator of the Bayesian equation separately. This implementation requires only two subset simulations to estimate the updated probability. In this implementation it is important to normalise the individual limit state functions to prevent bias

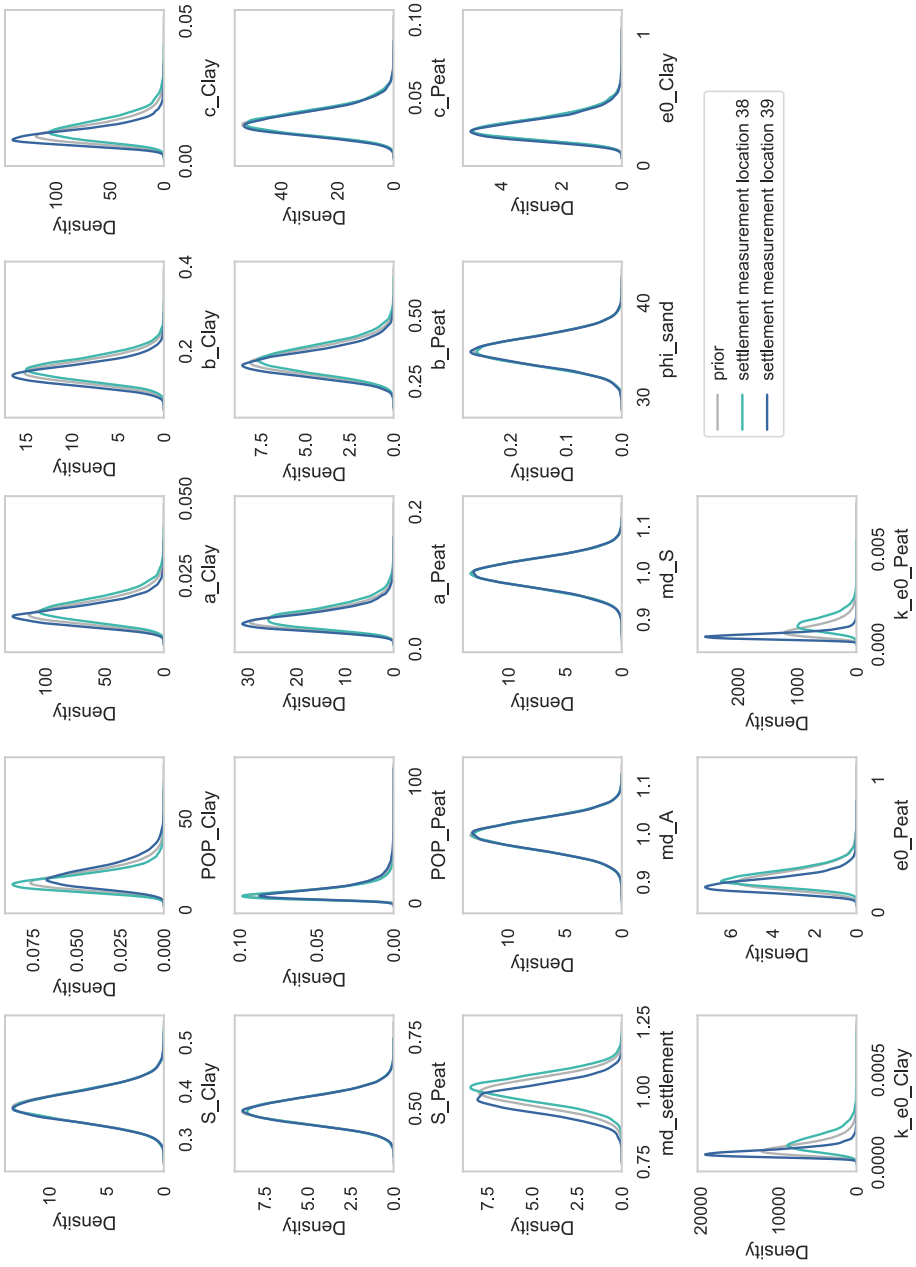


Figure 5.6: Local posterior pdfs of the settlement parameters after incorporating settlement measurements.

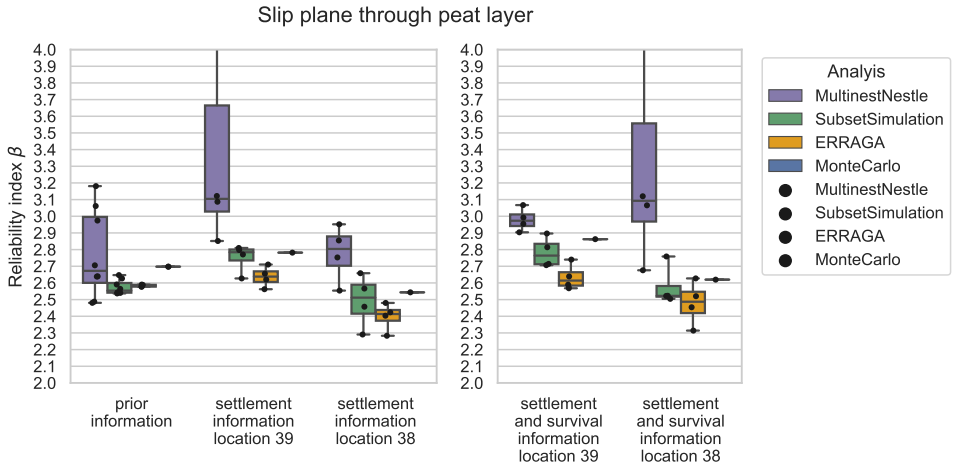


Figure 5.7: Reliability result for slope stability. Fixed deep slip circle through clay and peat layer. Results are conditional probabilities (daily water level).

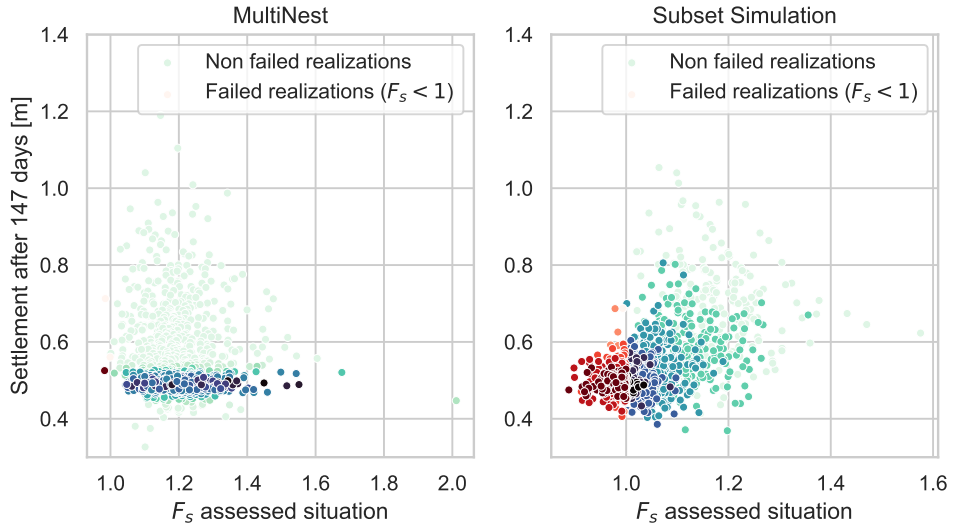


Figure 5.8: Realizations in MultiNest and subset simulation for the deep failure mode. Darker colors indicate subsequent live points in the MultiNest analysis (left) and subsequent subsets in the subset simulation (right).

in the results, because the selection of new seeds for subsequent subset runs depends on the limit state value. This normalisation would not be relevant if we followed a sequential Bayesian Updating framework (Cao et al., 2016; Straub et al., 2016), where $P(\mathrm{F}|\varepsilon)$ is calculated from the conditional samples in the last step of the calculation of $P(\varepsilon)$. The latter may involve more subset simulations, and hence more calculation time.

We chose a subset size of 1000 , leading to considerable scatter in the results among

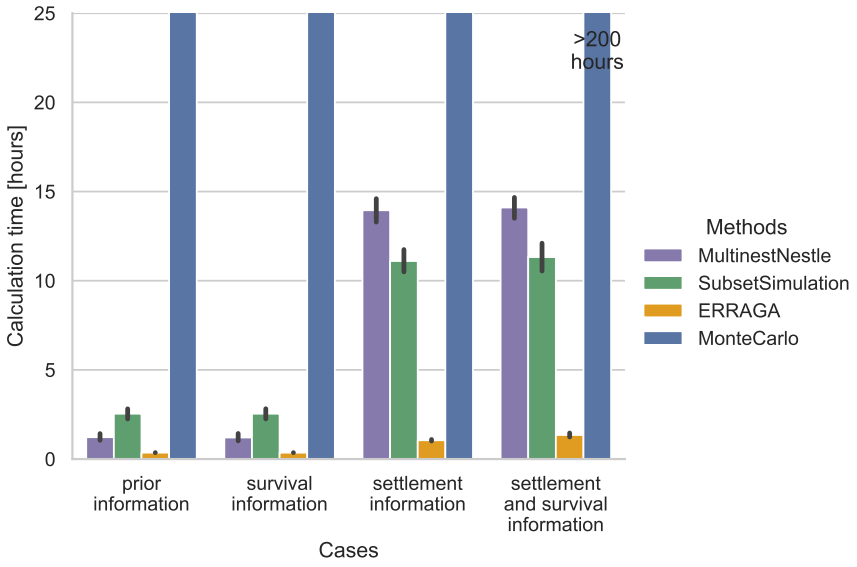


Figure 5.9: Calculation time of the Bayesian reliability analysis with different methods.

different runs. A larger subset size might work better with a combination of multiple limit states. A sensitivity analysis with a subset size of 3000 , shows that the scatter over different runs decreases, so the bias of a single analysis decreases too.

MULTINEST

From Section 5.4.1 follows that the results obtained with MultiNest seem to be (on average) unbiased but are heavily dependent on the number of failed samples initially sampled. To use MultiNest as a method for solving reliability problems with low prior failure probabilities, adaptions are needed to increase the number of samples in the failure domain, for example using Importance nested sampling Feroz et al. (2019). Here, $w_i = w_{i,\mathrm{MN}} \cdot w_{i,\mathrm{IS}}$ with $w_{i,\mathrm{MN}}$ the MultiNest weight factor, and $w_{i,\mathrm{IS}} = \pi(\mathbf{x}_i)/\pi'(\mathbf{x}_i)$ the importance weight factor with π and π' the actual and proposal pdf, respectively.

An initial trial was made using a variance increase of a factor 1.5 in standard normal space on all parameters, resulting in more failure samples. The accuracy of the failure probability was not substantially better than the calculation without variance increase. We recommend to explore adaptive importance sampling methods that can be better tailored at shifting or inflating particular probability distributions.

ERRAGA

The implementation with ERRAGA is clearly the best in terms of total calculation time and the number of model evaluations, see Figure 5.9. Overfitting of the metamodel was prevented by adding a noise term in the kernel and the UNIS learning function (see Eijnden et al. (2021)) was found to be effective in selecting new samples that contribute most to reducing the prediction uncertainty for the limit states for failure and survival.

Yet, the reliability results show some scatter and seem to under-predict the posterior reliability, possibly due to large uncertainty in the metamodel for the settlement observation ($\hat{h}_2$). Therefore improvements are needed to improve the fit of the limit state function for settlement, and to reduce the uncertainty on prediction of the metamodel, for example through the settings of the learning function, with another learning function, or with a different kernel. Generally, increasing the number of model evaluations also reduces uncertainty in the prediction, however, based on the uncertainty estimates of the metamodel, it does not seem likely that adding model evaluations will improve the results for this case study.

5.5. CONCLUDING REMARKS

This chapter showed that reliability estimates for slope stability improve further by combining the information of survival and settlement during the construction of dikes. Survival information and settlement information are complementary because the both sources of information are from different sources (independent data). Moreover, survival information is inequality data, for which the uncertainty reduction happens in the tail of the joint probability distribution, and settlement information is equality data, which updates the best estimates of the settlement and stability prediction. Thus, the uncertainty reduction affects the probability of failure in two ways.

Though the correlation between settlement and stability is not very large, the estimated probability of failure in the considered case study was between 25 and 50% higher or lower (factor 1.2-1.5) depending on the settlement measurement. Settlement measurements may lead to more uncertainty reduction for stability when the correlation between settlement and stability is larger. For example if the uncertainty and the relative influence of the degree of over-consolidation of soft subsoil layers is larger in both models, or when there are more 'shared' parameters affecting both settlement and stability.

There are practical limits to adding more information from additional observations and from multiple sources, namely the computational feasibility of such Bayesian reliability analysis in practice. Because, for each piece of information that is added, the number of required realisations increases drastically. With geotechnical models that require only a few seconds to calculate, most standard reliability methods become computationally intractable, and implementation of the Bayesian analysis is only practically feasible using metamodels. This is particularly the case for the safety assessment of flood defences and other geotechnical constructions in the Netherlands, with high target reliabilities (low probabilities of failure). A point of attention for using settlement measurements to update reliability is that the result of the Bayesian analysis is typically limited to the auto-correlation distance of settlements. In order for the reliability updating result to be representative for longer reaches of dike, settlement measurements of multiple locations are thus required.

6

COST-EFFECTIVENESS OF PERFORMANCE INFORMATION FOR DIKE REINFORCEMENTS

Today's posterior is tomorrow's prior.

Dennis Lindley

If performance information is available, it generally leads to more credible estimates of the failure probability. However, sometimes performance information is not yet available, and it requires an investment to obtain performance information. For example, setting up a monitoring campaign or carrying out proof loading involves investment cost, and proof loading comes with the additional risk of damage if a test fails. Whether it is cost-effective to invest in obtaining performance information depends on the investment costs and the benefits in terms of lower construction reduction costs. In this chapter the cost-effectiveness is evaluated of uncertainty reduction by proof loading and pore pressure monitoring in dike designs which take into account performance information.

This chapter is based on Van der Krogt, M.G., Klerk, W. J., Kanning, W., Schweckendiek, T. and Kok, M. (2020). Value of Information of Combinations of Proof Loading and Pore Pressure Monitoring for Flood Defences. *Structure and Infrastructure Engineering*, 1–16.
The research underlying this chapter was conducted as a joint project of Wouter Jan Klerk and Mark van der Krogt. Both researchers contributed equally to the research and the journal article.

Contents

6.1. INTRODUCTION

Insufficient safety against slope instability is traditionally remedied by increasing the resisting weight at the passive side of the slip plane, for example by decreasing the slope angle or by constructing a stability berm at the inner toe of the dike toe. The optimal reliability targets (see Section 2.1) are the starting point for the required safety of a reinforcement design. When space is available, reinforcing dikes is relatively cheap as the construction and material costs of soil are low. However, when space is scarce (e.g. in densely built areas like the Netherlands), reinforcement can become highly expensive. For example because adjacent home owners have to be moved and compensated, or because other design options are applied such as expensive sheet pile walls and diaphragm walls. In such cases, methods for reducing uncertainty might result in lower reinforcement costs because the required reliability can be achieved at much lower cost. For example, a less costly reinforcement method becomes feasible, or the reinforcement of certain dike sections becomes unnecessary.

As performance information can reduce uncertainty and improves reliability estimates of dikes, it is an interesting alternative or addition to traditional ways of strengthening dikes to resolve insufficient safety for slope stability. However, performance information is not always yet available. For example when the construction of a dike reinforcement has not started, or when we consider to impose a proof load (an example of a proof load of dikes is shown in Figure 6.1). The question arises when obtaining performance information is cost-effective, because the potential savings must outweigh the additional costs of, and relating to, obtaining the performance information. For example the costs of carrying out proof loading, and the failure costs if the load is not survived, must be balanced by the potential savings in terms of lower construction cost and reduced flood

Figure 6.1: Example of proof loading of dikes. Photo courtesy of Michiel van der Ruyt / Deltares

risk. For proof loading this balance is particularly interesting because both the amount of information and the risk of failure increase with an increasing (proof) load. This chapter explores the cost-effectiveness of performance information obtained with proof loading and monitoring, in optimizing the strategy for reinforcing dikes (leading to the lowest Total Cost (TC) involving all costs and benefits). The sub-question to be answered is:

How cost-effective are proof loading and pore water pressure monitoring in dike reinforcements?

First, this chapter presents the decision framework to evaluate the benefits of proof loading and pore pressure monitoring for dike reinforcements in Section 6.2. Thereafter a case study is presented in which a dike is subjected to proof loading by raising the phreatic level, and/or monitoring of pore pressures in Section 6.3. A simplified version of the decision framework is also used in an additional case study in Appendix A where it is explored whether a faster construction phasing is cost-effective.

6.2. FRAMEWORK TO SUPPORT DECISION-MAKING

To evaluate the benefits of pore pressure monitoring and proof loading Bayesian pre-posterior analysis is used, see Section 2.5. Decision trees are the most common approach to visualise and structure pre-posterior decision analysis (Raiffa and Schlaifer, 1961; Spross and Johansson, 2017; Thöns, 2017). A decision tree shows a sequence of decision (choice) nodes and outcomes (chance). Decision nodes are typically choices made by a decision maker as part of some (optimized) strategy, such as the decision to do proof loading. Chance nodes are outcomes of choices and depend on the action and prior information on the state of the system, for instance failure after a proof load test.

A disadvantage of a decision tree is that it can become cumbersome to visualise and solve if many sequential decisions are considered, in such cases other approaches such as influence diagrams (i.e., an extension of Bayesian networks) are more adequate (Luque and Straub, 2019), possibly combined with heuristic decision rules. This study considers three decision options (proof loading, monitoring and dike reinforcement), hence a decision tree is well suited. Figure 6.2 presents the decision tree for the sequential decision strategy of proof loading, pore pressure monitoring and dike reinforcement, denoted with p, m, and a, respectively. Note that a specific sequence for proof loading and pore pressure monitoring is assumed, the effect of reversing the sequence is discussed in Section 6.3.4.

The first step is the decision whether to execute a proof load test of a certain magnitude $p \in \mathbf{P}$. The outcome z_p (a survived or failed proof load test) depends on the magnitude of the proof load and the prior belief $f_{\mathbf{X}}(\mathbf{x})$ on the random variables in $\mathbf{X}$. The higher the magnitude of the survived test load (i.e. the artificially induced phreatic line), the more uncertainty is reduced, and the higher the updated reliability. On the other hand, the higher the magnitude of the survived test load, the higher the probability that the test is not survived. In that case the dike is damaged and needs to be reinforced immediately and the part of the section that was proof loaded has to be repaired such that extra costs are incurred.

After deciding whether to do a proof load test (and on the magnitude of the test load),

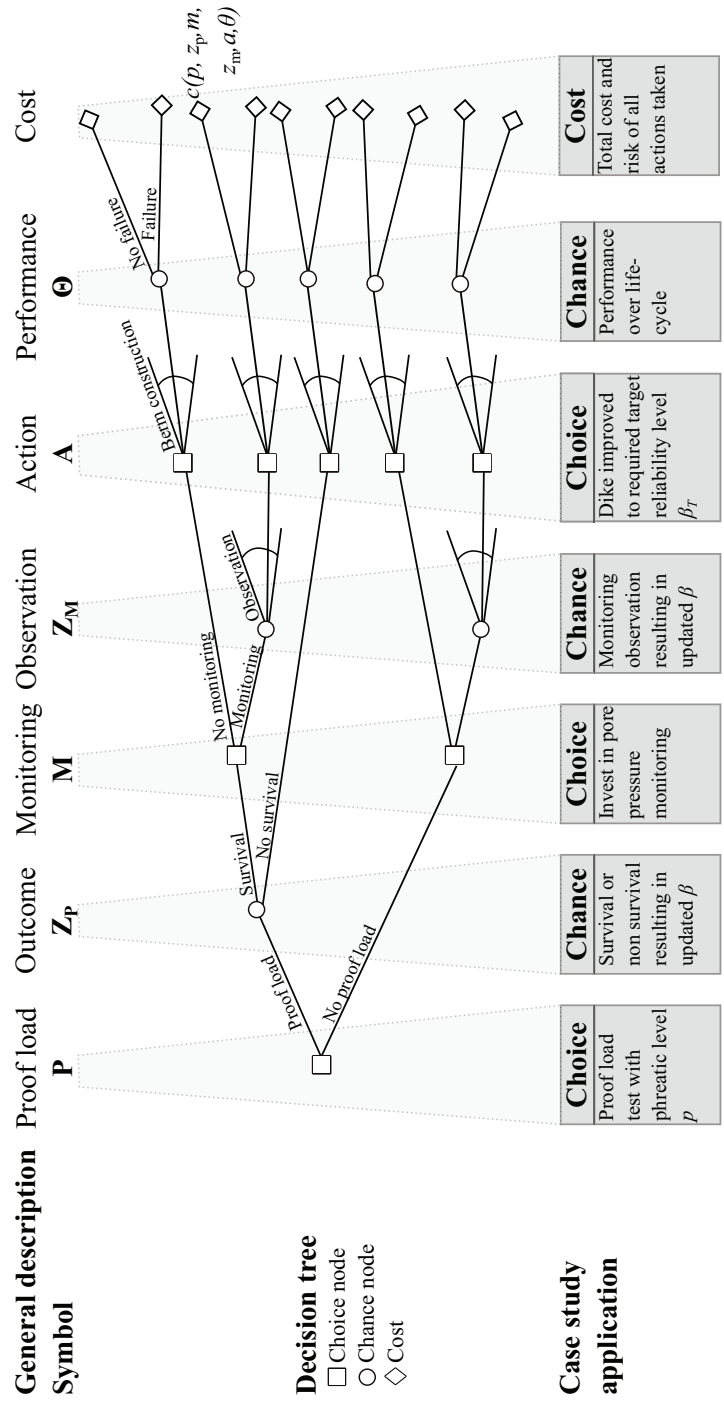


Figure 6.2: Decision tree for a sequential decision on proof loading, monitoring, and reinforcement of a dike section. The decision tree is a graphical presentation of the choices $p \in \mathbf{P}$, $m \in \mathbf{M}$, and $a \in \mathbf{A}$, and chances $z_p \in \mathbf{Z}_p$, and $z_m \in \mathbf{Z}_m$.

it can be decided to invest in pore pressure monitoring ($m \in \mathbf{M}$) in order to reduce uncertainty on the response of the phreatic line to outside water levels. Again, two outcomes are possible: either an observation is made or not. The observation z_{m} depends on the belief after proof loading $f_{\mathbf{X}|\mathbf{z}_{\mathrm{m}}}(\mathbf{x})$. Whether an observation is made in the considered time period depends on whether the water level exceeds a certain threshold required to obtain useful measurements (Frangopol et al., 2008; Klerk et al., 2019). Note that the probability of making and observation is time-dependent: the longer the monitoring period, the higher the probability of a useful observation, opposite to a proof load test which is time-independent.

After the outcome of the monitoring, the dike is improved to the required target reliability level using decision rules that translate combinations of outcomes of proof loading and monitoring to actions: $\mathbf{d}(\mathbf{Z}_{\mathrm{p}}, \mathbf{Z}_{\mathrm{m}}) = \mathbf{A}$, where for an individual decision rule d it holds that $d \in \mathbf{d}(\mathbf{Z}_{\mathrm{p}}, \mathbf{Z}_{\mathrm{m}}) = \mathbf{A}$. Note that through an action $a \in \mathbf{A}$ also some design variables in $\mathbf{X}$ can be adapted (e.g., the length of the stability berm).

The cost of a branch in the decision tree is determined by the costs of every individual step and the expected damage given the performance θ (failure/no failure): $c(p, z_{\mathrm{p}}, m, z_{\mathrm{m}}, d, \theta)$. The cost of the optimal strategy $c(p^*, m^*, d^*)$ can be computed by combining the cost of different branches over the possible outcomes:

$$c(p^*, m^*, d^*) = \min_{p \in \mathbf{P}, m \in \mathbf{M}, d} E_{\Theta|\mathbf{X}}\left[E_{z_{\mathrm{m}}|\mathbf{X}}\left[E_{z_{\mathrm{p}}|\mathbf{X}}\left[c\left(p, z_{\mathrm{p}}, m, z_{\mathrm{m}}, d, \theta\right)\right]\right]\right]. \tag{6.1}$$

Specifically in this case study, the cost of a strategy $c(p, m, d)$ is defined by the sum of costs of each step in the decision tree (decision and outcome):

$$\begin{aligned} c(p, m, d) = {} & I_{\mathrm{p}} \cdot C_{\text{proofload}} + I_{\mathrm{m}} \cdot C_{\text{monitoring}} + \\ & \int \left(P(\overline{S}|p \cap \mathbf{x}) \cdot C_{\text{repair}} + C_{\text{reinforcement}}(d, \mathbf{x}) + C_{\text{failure}}(d, \mathbf{x})\right) f_{\mathbf{X}}(\mathbf{x}) \mathrm{d}\mathbf{x}, \end{aligned} \tag{6.2}$$

where I_{p} and I_{m} are indicator random variables (value 0 or 1) that indicate whether proof loading or monitoring is done. C parameters indicate different cost components. The cost components of proof loading and monitoring are independent of the prior belief $f_{\mathbf{X}}(\mathbf{x})$. There are three cost components that depend on the prior belief: the cost of failure after a failed proof load test, where $P(\overline{S}|p \cap \mathbf{x})$ is the probability of not surviving a proof load with magnitude p, and C_{repair} are the repair costs. The costs of reinforcement $C_{\text{reinforcement}}$ depend on decision d and the realisation $\mathbf{x}$. The annual failure probability is assumed to be constant in time. Thus, for the Present Value of the failure costs $C_{\text{failure}}(d, \mathbf{x})$ an infinite time horizon can be considered, such that:

$$C_{\text{failure}}(d, \mathbf{x}) = \frac{P(\mathrm{F}|d, \mathbf{x}) \cdot D}{r}, \tag{6.3}$$

where $C_{\text{failure}}(d, \mathbf{x})$ is the cost of failure in € for an infinite time horizon, D is the expected damage in case of a flood (in €), r is the annual discount rate, and $P(F|d, \mathbf{x})$ is the annual failure probability given an action following from decision rule d and a realisation of the set of random variables $\mathbf{x}$. A reference period of 1 year is assumed, in line with common practice for flood defence structures. It should be noted that in

some cases for geotechnical structures use of other reference periods might be more adequate (Roubos et al., 2018), and for instance the time factors provided in Diamantidis et al. (2019) may be applied. The cost of the reference strategy without monitoring and proof loading is defined as $c_0(d)$, the cost of a strategy of proof loading, monitoring, and reinforcing is defined as $c(p,m,d)$, so the Value of Information (VoI) of proof loading and/or monitoring can be computed by:

$$VoI = c_0(d) - c(p,m,d) \tag{6.4}$$

The next subsections go into the choices the decision maker is faced with in more detail.

STEP 1: DOING A PROOF LOAD TEST (P)

Proof loading involves imposing a representative design load on the dike body, for example a high phreatic line (see Figure 6.5). If such a proof load is survived, it proves that there is a minimum resistance along a slip plane. Conversely, when the dike fails under the conditions of the proof load test, it reveals that the structure was not safe enough. Note that a higher proof load yields more information, but also results in a higher risk of failure during the test. The outcome of the proof load test is used to update the failure probability based on the outcome z_p of the proof load test.

The updated failure probability is directly calculated by applying Bayes rule (Equation 2.11):

$$P(\mathrm{F}|z_\mathrm{p}) = \frac{P(\mathrm{F}\cap z_\mathrm{p})}{P(z_\mathrm{p})}, \tag{6.5}$$

where the observation of no instability at a proof load level p is noted by z_p. The performance function for stability at a proof load level p is denoted here by $h(\mathbf{X},p) < 0$. It is noted that a proof load test does not update all parameters, for example for those related to response of the phreatic line to an extreme flood water level (e.g., the head level in the aquifer below the soft soil blanket or pore pressures in the dike body in flood conditions) no additional information is obtained.

STEP 2: SETTING UP A PORE PRESSURE MONITORING CAMPAIGN (M)

After or instead of proof loading, uncertainty can be reduced by setting up a pore pressure monitoring campaign. Pore pressure monitoring aims to reduce uncertainty about the response of the phreatic line in the dike. The parameters characterizing this response are part of the belief $f_{\mathbf{X}|\mathbf{z}_\mathrm{p}}(\mathbf{x})$, where conditioning on z_p is not needed if proof loading was not done beforehand. If a pore pressure monitoring campaign yields an observation z_m, $f_{\mathbf{X}|\mathbf{z}_\mathrm{p}}(\mathbf{x})$ can be updated to a posterior estimate including z_m, according to Equation 2.14:

$$f_{\mathbf{X}|z_\mathrm{p}\cap z_\mathrm{m}}(\mathbf{x}) = \frac{P(\mathbf{x}\cap z_\mathrm{m})\,f_{\mathbf{X}|z_\mathrm{p}}(\mathbf{x})}{P(z_\mathrm{m})} = \frac{L(\mathbf{x})\,f_{\mathbf{X}|z_\mathrm{p}}(\mathbf{x})}{\int_\mathbf{X} L(\mathbf{x})\,f_{\mathbf{X}|z_\mathrm{p}}(\mathbf{x})\,d\mathbf{x}}. \tag{6.6}$$

Here $L(\mathbf{x})$ is the likelihood $P(z_\mathrm{m}|\mathbf{x})$ calculated with the updated probability distribution $f_{\mathbf{X}|\mathbf{z}_\mathrm{p}}(\mathbf{x})$. The posterior probability of failure with monitoring and proof load is calculated using Equation 2.15, with the posterior pdf $f_{\mathbf{X}|z_\mathrm{p}\cap z_\mathrm{m}}$. Note that the parameters in $\mathbf{X}$ related to the response of the phreatic line are now directly updated, as there are direct observations of input parameters, contrary to proof loading.

As was indicated by Klerk et al. (2019), an important parameter for pore pressure monitoring is the probability that a useful observation is obtained. Often discontinuities in a dike body (e.g., an older clay dike), or different permeability values in general can result in different responses of the phreatic line for different outside water levels, and therefore, an observation z_{m} is to give more useful information if measurement conditions are closer to design conditions. To incorporate this, it is assumed that a valuable measurement (i.e. uncertainty reduction) is only obtained if the annual maximum water level h exceeds a predefined threshold water level h_{thresh} during monitoring. Thus the probability of obtaining a valuable measurement z_{m} can be computed using the following formula:

$$P(\text{valuable observation } z_{\mathrm{m}}) = 1 - F_h(h_{\mathrm{thresh}})^t \tag{6.7}$$

where F_h is the cumulative density function of the annual maximum outside water level, and t the duration of monitoring in years.

STEP 3: DIKE REINFORCEMENT (A)

In practice, numerous reinforcement measures are available to increase the stability of dikes, for example: stability berms, sheet pile or diaphragm walls, or soil anchoring techniques. Here, only the most common (and often cheapest) method of stability berm construction is considered. Adding a stability berm at the inner toe of the dike increases the weight on the passive side of the slip plane and increases the resisting shear stress.

The target reliability that has to be satisfied after a dike reinforcement is often predetermined, and typically based on an optimization of various risk indicators and costs of reinforcement (e.g., Eijgenraam et al. (2017); Voortman (2003); Vrijling (2001)). If the reliability of dikes is changing significantly in time, one also has to consider reinvestments. However, due to the dependence of slope stability reliability on time-independent ground-related uncertainty, slope stability of dikes (and other geotechnical structures, e.g. Roubos et al. (2018)) is typically rather time independent. Therefore, in an economic optimization one can estimate the annual target reliability by considering an infinite time horizon, such that the optimal level of protection β_{T}, follows from the following minimization:

$$\beta_{\mathrm{T}} = \arg\min_{\beta} \left(C(\beta) + \frac{\Phi^{-1}(\beta) \cdot D}{r} \right) \tag{6.8}$$

where D is the annual expected damage in case of flooding, r is the annual discount rate and $C(\beta)$ is the cost of achieving a certain reliability index. Notice that this formulation assumes a situation where the reliability target is prescribed by a standard. A situation where the reliability target is case-specifically optimized using a cost-benefit analysis is further addressed in the sensitivity analysis in Section 6.3.3.

6.3. CASE STUDY

This case study considers pore pressure monitoring of the position of the phreatic line and proof loading by raising the phreatic line in the dike, as the phreatic line is one of the main factors causing slope instability. Pore pressure monitoring is aimed at reducing

uncertainty on the position of the phreatic line (Koelewijn et al., 2013), representing the response of pore water pressures in the dike body to hydraulic loads. Such responses are typically dependent on the hydraulic conductivity of the dike material, which is often heterogeneous hence uncertain due to the limited amount of measurements that are available.

An important aspect of pore pressure monitoring is that the information obtained (resulting in uncertainty reduction) is dependent on the water levels observed during the monitoring period (Klerk et al., 2019; Schweckendiek et al., 2014). In some cases, for instance at locations with a large tidal range, frequently occurring situations are similar to design conditions, resulting in significant uncertainty reduction. At other locations, such as the river dikes regarded in this case study, conditions leading to large uncertainty reduction occur less often. Consequently, the longer the monitoring period, the higher the probability of obtaining useful information that can be used to reduce uncertainty, as was shown by Frangopol et al. (2008) and Klerk et al. (2019).

Proof loading in this case study consists of a controlled experiment to artificially raise the phreatic line. We assume a successful test in the sense that it always succeeds in increasing the water pressures to the desired level, throughout the dike body. Note that proof loading only reduces uncertainty in the (variables relating to the) overall resistance, conditional to the imposed proof load. It does not lead to additional knowledge about the actual response of the phreatic line to flood conditions. Thus, pore pressure monitoring and proof loading are complementary.

6.3.1. CASE DESCRIPTION

The reference case is a dike section of 1 km in length, inspired by an actual dike section currently being reinforced. It is slightly simplified such that it contemplates a typical dike section in the Dutch riverine area. The dike cross section, displayed in Figure 6.3, consists of a traditional clay dike which has been reinforced with sand in the past. It is assumed that the dike is scheduled for reinforcement in 5 years as it currently does not meet the safety standard. Until that time there is opportunity to do a proof load test and pore pressure monitoring to reduce uncertainty on the resistance parameters and the position of the phreatic line in the dike body, respectively. The goal is to determine the optimal course of action for the coming 5 years.

The dike consists partly of clay and partly of sand, has a crest level at +14.0 m ref. (reference level), a landside elevation of +6.0 m ref., an inner slope of 1:3 (v:h) and is situated on (Holocene) clay layers on top of a (Pleistocene) aquifer. A cross-section of the considered dike is shown in Figure 6.3. The strength of the soil is modelled according to the CSSM (Schofield and Wroth, 1968) with a critical state friction angle (φ_{cs}) or undrained shear strength (s_u) calculated using the SHANSEP formulation (Ladd and Foott, 1974), see Equation 2.16. Note that a high phreatic line leads to higher pore pressures, thus lower s_u, and lower stability (F_s). Additionally, the stability decreases because a higher phreatic line corresponds to a higher weight of the dike body. Table 6.1 lists the input probability distributions for parameters in the reference case. The probability distributions for these spatially averaged soil parameters are derived from regional data for typical geological deposits of the Dutch situation, see RWS (2019).

[1] Note that μ is the mean value, not the lognormal distribution parameter, CoV is the coefficient of variation.

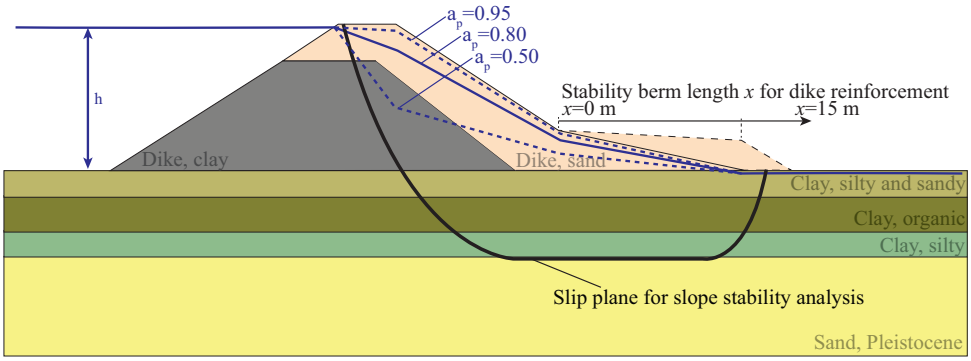


Figure 6.3: Cross section of the considered case study. Blue lines indicate the simplified schematization of the phreatic line for different response factors a_p (at an extreme water level).

Table 6.1: Random variables in the reference case.

Property	Symbol	Unit	Soil type	Distribution[1]
Normally consolidated undrained shear strength ratio	S	-	Clay, silt and sand	Lognormal(μ=0.36, CoV=0.15)
			Clay, organic	Lognormal(μ=0.29, CoV=0.15)
			Clay, silt	Lognormal(μ=0.32, CoV=0.25)
Strength increase exponent	m	-	Clay, silt and sand	Lognormal(μ=0.84, CoV=0.05)
			Clay, organic	Lognormal(μ=0.93, CoV=0.05)
			Clay, silt	Lognormal(μ=0.83, CoV=0.05)
Pre-overburden Pressure at daily stress conditions (no flood)	POP	kPa	Clay, silt and sand	Lognormal(μ=27.0, CoV=0.45)
			Clay, organic	Lognormal(μ=27.0, CoV=0.45)
			Clay, silt	Lognormal(μ=27.0, CoV=0.45)
Critical state friction angle	ϕ_{cs}	°	Dike, sand	Lognormal(μ=32.6, CoV=0.05)
			Dike, clay	Lognormal(μ=35.0, CoV=0.05)
			Clay, silt and sand	Lognormal(μ=32.3, CoV=0.05)
			Sand, Pleistocene	Lognormal(μ=35.0, CoV=0.05)
Model factor stability model	m_s	-		Lognormal(μ=0.995, CoV=0.033)
Parameter for phreatic line	a_p	-		Uniform(a=0.5, b=0.95)
Water level	h	m ref.		Gumbel(loc=11.9, scale=0.2)

In the case study, only monitoring of the phreatic line in the dike body is considered, not of the pore water pressures in other soil layers. The position of the phreatic line in the dike at flood conditions typically depends on the permeability of the dike material which is often heterogeneous and uncertain. Especially when a dike has a long history of reinforcements with various materials, the phreatic line is uncertain. For example the considered case study of a traditional clay dike reinforced with sand. Therefore, the position of the phreatic line in steady state seepage conditions is parametrised, using an uncertain response factor (a_p). The response factor represents the degree of saturation

of the dike body at the inner crest line, in response to an extreme water level. Values can range between $a_p = 0$ (phreatic level at the landside elevation level) and $a_p = 1$ (phreatic level equal to the outside water level). For intermediate values of a_p, the phreatic line is interpolated accordingly, see Figure 6.3. Because the dike body will always saturate to some degree, and in case of a fully saturated dike ($a_p = 1$) other mechanisms such as micro-instability become dominant, the value of a_p is limited between 0.5 and 0.95. The bounds represent realistic values based on physical considerations. Furthermore, the lower bound has a limited influence on the reliability, indicated by the results in the next paragraph.

6.3.2. IMPLEMENTATION OF RISK REDUCTION STRATEGIES

To facilitate the probability updating outlined in Section 6.2, fragility surfaces are derived, where the failure probability is conditional to response factor a_p and the water level h, analogous to Equation 2.8 for fragility curves with one parameter. The fragility surfaces are derived both for the prior situation, and the situation posterior to surviving a certain proof load level p. Figure 6.4 shows this fragility surface, plotted in terms of reliability index for convenience. The reliability is calculated at discrete intervals of h and a_p, and linearly interpolated to obtain intermediate values. The fragility surface directly shows the influence of the response factor a_p (mainly at high water levels), and clearly illustrates the potential benefit of reducing uncertainty herein. Separate fragility surfaces $\beta(h, a_p)$ are derived for berm lengths of 5, 10, 15, and 20 m. For other values fragility surfaces are interpolated or extrapolated. Integration of the prior fragility surface with the prior probability distribution of a_p and h along the lines of Equation 2.9, results in a prior failure probability of 2.7×10^{-4} per year ($\beta = 3.46$).

PROOF LOADING

Proof loading is done by artificially raising the phreatic line in the dike by infiltrating water into the dike from the crest (similar to Van Hoven and Noordam (2018)), see Figure 6.5. Survival of the situation with an imposed phreatic level leads to a higher reliability because of an implicit update of the probability density of soil parameters involved (which are a subset of $\mathbf{X}$). The higher the phreatic level, the larger the uncertainty reduction, and hence, the larger the reliability update; but also the higher the probability the proof load is not survived.

Contrary to the phreatic level in flood conditions (dependent on amongst others the flood water level, the duration of the flood wave, and permeability of the outer slope cover layer), the phreatic level during proof loading is induced/imposed by infiltrating water into the dike using e.g. infiltration wells or an irrigation system (see van Van Hoven and Noordam (2018) for pictures). Therefore, the outcome of the proof load test (and hence the updated reliability) is independent of the response factor a_p, and the posterior reliability β conditional to a_p: $\beta|a_p = -\Phi^{-1}(P(\mathrm{F}|z_p, a_p))$.

Figure 6.4 shows that a significantly updated reliability for water levels lower than the survived proof load of +12.5 m ref. is to be expected. The reliability update is relatively larger for lower values of a_p. This is in line with expectations because the survived proof load becomes more valuable if a high phreatic line is less likely. Note that the failure probability for water levels lower than the survived proof load level is not reduced to 0

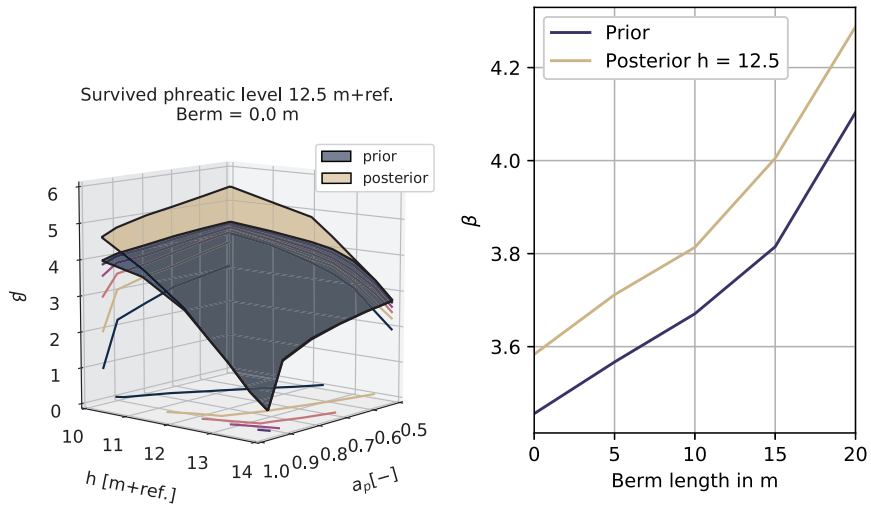


Figure 6.4: (left) Prior and posterior fragility surface (in terms of reliability index β) of the water level h and response factor a_p, for the considered case study without berm. The overall reliability index (integrated with the prior probability density of a_p and h) is 3.46. (right) Relationship between berm length and overall reliability index β for the prior situation and posterior after a proof load level of +12.5 m ref.

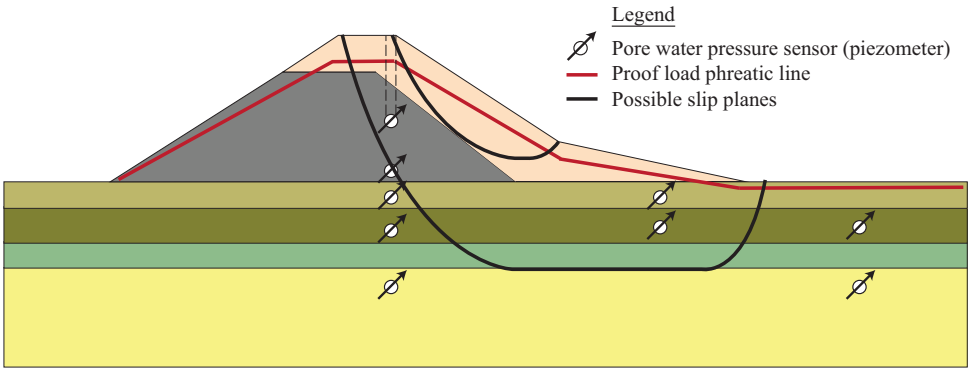


Figure 6.5: Overview of the positioning of sensors installed for pore water pressure monitoring, and the imposed phreatic level during a proof load test. The larger black line indicates the slip plane relevant for flooding, the smaller slip plane is relevant for failure of the proof load test but does not cause flooding.

(infinite beta) because of irreducible uncertainty.

It is assumed that the proof load is applied over a stretch of 100 m length. This is considered representative for the 1 km dike section because of a limited variation in the dike body in longitudinal direction as a result of the quite recent reconstruction with sand, see Figure 6.3. The total cost of a proof load test is assumed to be 500 000 € consisting of costs for equipment required for infiltration, monitoring during the test, emergency measures to mitigate slope failures induced by the test and analysis of the test results. It is assumed that the test is carried out in a period where a potential failure

does not cause flooding. Therefore, the costs of not surviving a proof load only consist of repairing the damaged slope. These costs are estimated to be 2 000 000 €, based on the costs of full reconstruction of the existing dike over a length of 100 m. Additional costs such as follow-up damage to buildings, transportation infrastructure, agricultural areas etc. are disregarded in this case study. For damage during a proof load test (C_{repair}) occurrence of each slip circle (also very shallow) is considered as failure, contrary to flooding. For flooding damage (D) only larger slip circles which will lead to flooding of the hinterland are considered, as is depicted in Figure 6.5. After a proof load test failure, no pore pressure monitoring is done.

PORE PRESSURE MONITORING

Pore pressure monitoring is carried out by measuring the phreatic line in the dike body (see Figure 6.5 for location of sensors). The measurement will lead to an update of the probability distribution of a_{p} ($a_{\text{p}} \in \mathbf{X}$). Because of the chosen limits of the prior distribution of a_{p} it is assumed that the posterior distribution of a_{p} is a truncated normal distribution with μ the observed value (i.e., based on possible state), standard deviation $\sigma = 0.05$ and upper and lower bound equal to the upper and lower bound of the uniform prior. The value of σ accounts for measurement errors and transformation errors, and corresponds with a standard deviation of 0.3 m in the position of the phreatic line. This value is in accordance with commonly found values in the Dutch practice (Kanning and Van der Krogt, 2016).

Due to the old clay dike located in the cross section the sensors will only yield relevant results if the water level is somewhat above the crest of the old clay dike (see Figure 6.3, it is assumed that this threshold is +12.2 m ref. (0.2 m above the top of the clay). With the local probability distribution for water levels, and 5 years of monitoring the probability that a relevant observation is obtained is 67% (using Equation 6.7).

While not explicitly modelled, the costs are based on plans for measuring the entire section including redundancy in measurements and multiple cross-sections with sensors. The cost of pore pressure monitoring is estimated at 100 000 € for 5 years and include cost for installation, maintenance, decommissioning and analysis of the obtained data, based on the number of sensors in Figure 6.5, installed at two cross-sections.

DIKE REINFORCEMENT

The reliability requirement for the dike section is determined based on the level of protection with minimal total cost (see Equation 6.8). This value is derived based on the prior $f_{\mathbf{X}}(\mathbf{x})$. The costs for reinforcement are shown in Figure 6.6, both for the reference case and some alternatives that will be used in a sensitivity analysis. Except for alternative 2, these curves have been derived using *KOSWAT*, a software program used for cost calculations for dike reinforcements in the Netherlands (Deltares, 2014). Only reinforcement through a stability berm is considered (see Figure 6.3 for dimensions). The costs are calculated using Equation 6.2. Note that the risk in the 5 years before reinforcement is not considered, as this is the same for each strategy (and thus does not lead to differences in VoI).

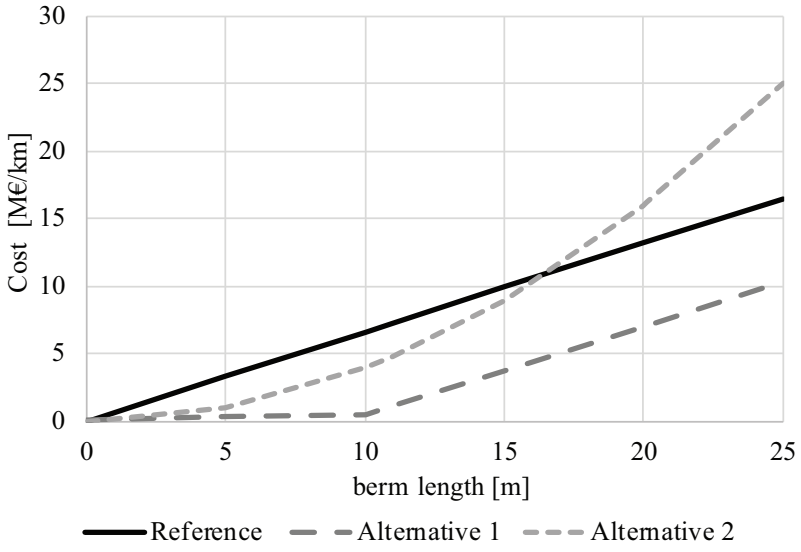


Figure 6.6: Different cost functions for dike reinforcements.

Table 6.2: Cost/Benefit parameters.

Parameter	Description	Unit	Value
r	Annual discount rate	-	0.035
D	Damage in case of flooding	million €	5000
C_{repair}	Cost of repair after failed proof load test	million €	2.0
$C_{\text{monitoring}}$	Cost of 5 years of pore pressure monitoring	million €	0.1
$C_{\text{proofload}}$	Cost of proof load test	million €	0.5
σ	Uncertainty in observation of a_{p}	-	0.05
h_{thresh}	Minimum water level for a useful observation	m +ref	12.2

6.3.3. RESULTS

We compare the total life cycle cost of the dike for three cases with and without proof loading and monitoring, with a conventional strategy of only reinforcing, in order to explore the conditions for which to invest in uncertainty reduction for dikes.

REFERENCE CASE

First, it is evaluated whether proof loading and/or pore pressure monitoring reduces overall total cost for a reference case. Here, a proof load test where the phreatic line is artificially increased to +12.5 m ref. is considered. For 5 years of monitoring the probability of having a useful observation is 67%. The parameters used for the cost benefit analysis are shown in Table 6.2.

Figure 6.7 shows total cost and VoI (see Equation 6.4) for all combinations of proof loading and monitoring, compared to a conventional strategy without monitoring and

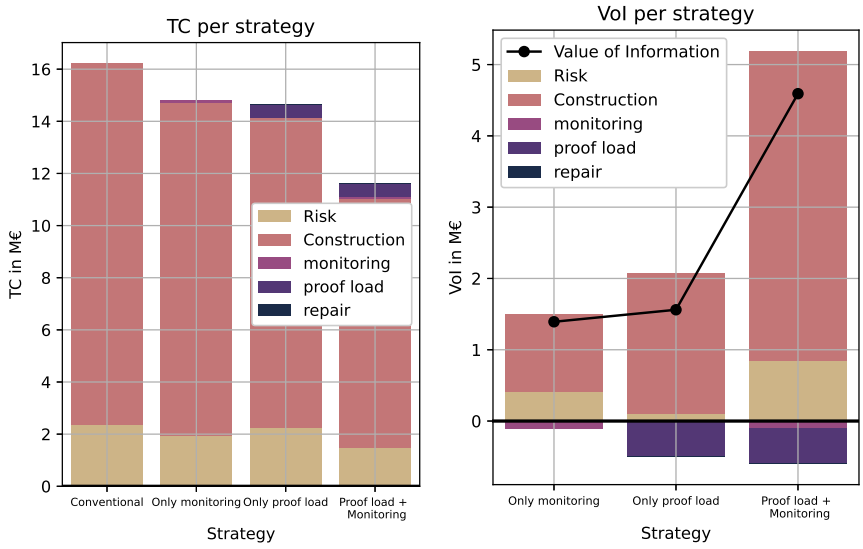


Figure 6.7: Total Cost (TC) (left) and Value of Information (VoI) (right) per strategy for the reference case. Colors indicate what the contribution is of different components to the TC (left) and VoI (right). The VoI for each strategy (the sum of the components) is calculated relative to the conventional strategy.

proof loading. Both monitoring and proof loading reduce total cost, with the optimal strategy being a combination of proof loading and monitoring (VoI = 4.6 M€). For the optimal strategy the reduction in total cost is 31% compared to a conventional reinforcement, strategies with only proof loading or monitoring have a lower but also positive VoI. The most important component for the VoI is the reduction in construction cost, which significantly outweighs the costs of monitoring and proof loading.

OPTIMIZATION OF PROOF LOAD LEVEL

Although Figure 6.7 clearly shows that a combination of monitoring and proof loading is an effective approach to reduce total cost, another important choice is the phreatic level that is to be tested. While lower levels will result in a smaller reduction of uncertainty, higher levels have higher uncertainty reduction but also the added risk that the dike section fails during the test and has to be repaired. Figure 6.8 shows the relation between phreatic level in the proof load test and the VoI. The red line indicates the VoI for different combinations of proof loading and monitoring, for which the optimum is at a proof load phreatic level +13.0 m ref.. If no monitoring is done, the optimal proof load level is +13.5 m ref. (see yellow bars). However, combined with monitoring, the VoI is highest with a lower proof load level. For proof load levels above +13.5 m ref., the VoI becomes negative because of the high risk of failure during the test (i.e., there is a critical proof load level where the VoI = 0).

Another interesting observation is that in this case the VoI of monitoring after a proof load test (purple bars) is higher than the VoI without a preceding proof load test (left purple bar). Thus, the monitoring becomes more valuable after reducing uncertainty

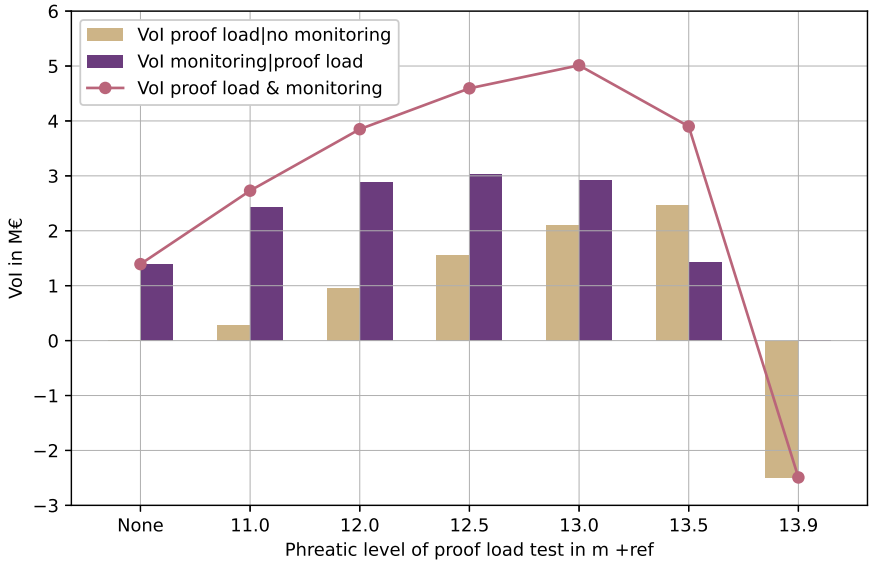


Figure 6.8: Value of Information (VoI) related to the level of the proof load test. Red line indicates the VoI for a combination of proof loading and monitoring, yellow bars indicate the value of a proof load test without monitoring. Purple bars indicate the added value of monitoring after a proof load test.

through proof loading. Obviously, this can differ per case, and it is also dependent on for instance the shape of the relationship between construction cost and berm length.

SENSITIVITY ANALYSIS

Dike sections that are part of longer dike segments can differ significantly. This section discusses several of these differences encountered in practical situations, and their influence on the VoI, namely:

- Influence of the reliability requirement: in many practical cases reliability requirements are not based on an economic optimization, such that the VoI might be different.
- Influence of different soil parameters: different locations can have significantly different mean values and variance of soil parameters, such that the benefits of different types of uncertainty reduction might shift.
- Influence of different cost functions: due to local circumstances (e.g., density of adjacent buildings) costs of reinforcement can vary, which can influence the VoI.

A proof load level of +13.0 m ref. is assumed in all cases of the sensitivity analysis, which is (close to) optimal in all cases and strategies (see also Figure 6.8).

Influence of the reliability requirement

In the reference case an optimal target reliability level is determined based on a total

cost minimization using prior information. In reality, the section studied is part of a larger flood defence system where other safety requirements (e.g., loss-of-life) might be dominant, or requirements are based on general codes. It would therefore be unlikely that the safety standard is exactly economically optimal for this specific dike section, with its specific characteristics. Figure 6.9a and 6.9b show a comparison of total cost and Value of Information for 4 cases: the reference case with optimized target reliability based on the prior information ($\beta_T = 4.13$), a case with 10 times higher requirement ($\beta_T = 4.63$), 10 times lower requirement ($\beta_T = 3.56$), and a case where the optimal target reliability is determined based on the posterior information after a proof loading and/or monitoring.

Without monitoring, the cases with lower and higher reliability requirements are significantly more expensive in terms of total cost. For the case with a higher requirement this is mainly caused by higher reinforcement costs, whereas for the case with a lower requirement this is due to higher risk costs. As reinforcement costs for the case with a higher requirement are still high after monitoring, the VoI is limited for this case. For the case with a lower requirement, the VoI of a combination of a proof loading and pore pressure monitoring is very high. The reason is that in case of very unfavourable values of a_p (and therefore high risk costs), observations are very valuable. In addition, it is prevented that an insufficiently safe dike is constructed as a result of an already too low reliability requirement.

The most efficient strategy in terms of total cost is if proof loading and monitoring are combined with a posterior optimal reliability requirement. Concretely, the optimal target reliability to be met after the dike reinforcement is determined based on the posterior information after monitoring and/or the proof load test (using Equation 6.8), rather than the prior information. Consequently, the optimized target reliability depends on the obtained information z_p and z_m, and the determination of β_T becomes a part of the decision rules $\mathbf{d}(\mathbf{Z}_p, \mathbf{Z}_m)$ in the decision tree. Hence, each branch in the decision tree can have a different β_T, dependent on the observations. This is slightly more efficient than having a requirement based on prior information, especially in case of a very favourable or unfavourable outcome after monitoring, because the change in expected reinforcement cost can be adjusted in the posterior optimization of the requirement. It has to be noted that the differences with the reference case with (prior) optimized β_T are limited, but it demonstrates that using a suboptimal target reliability has a large influence on the results of a Value of Information analysis.

Influence of different soil parameters

The reference dike section is characterised by relatively large uncertainties in soil parameters, and therefore the VoI of both proof loading and monitoring is found to be relatively large. However, not all dikes might have such large uncertainties, and therefore the VoI is assessed for two other cases: dike section A with lower uncertainty in soil parameters and a prior reliability index of 3.99, with a target reliability of 4.07. Hence, there is only a small reliability deficit, that would in practice likely be accepted as is. Section B also has relatively low uncertainty in soil parameters but lower mean values, so the prior reliability index is 3.61, with a target reliability of 4.02. Figure 6.9c and 6.9d show the Total Cost and VoI for each dike section for 4 different strategies.

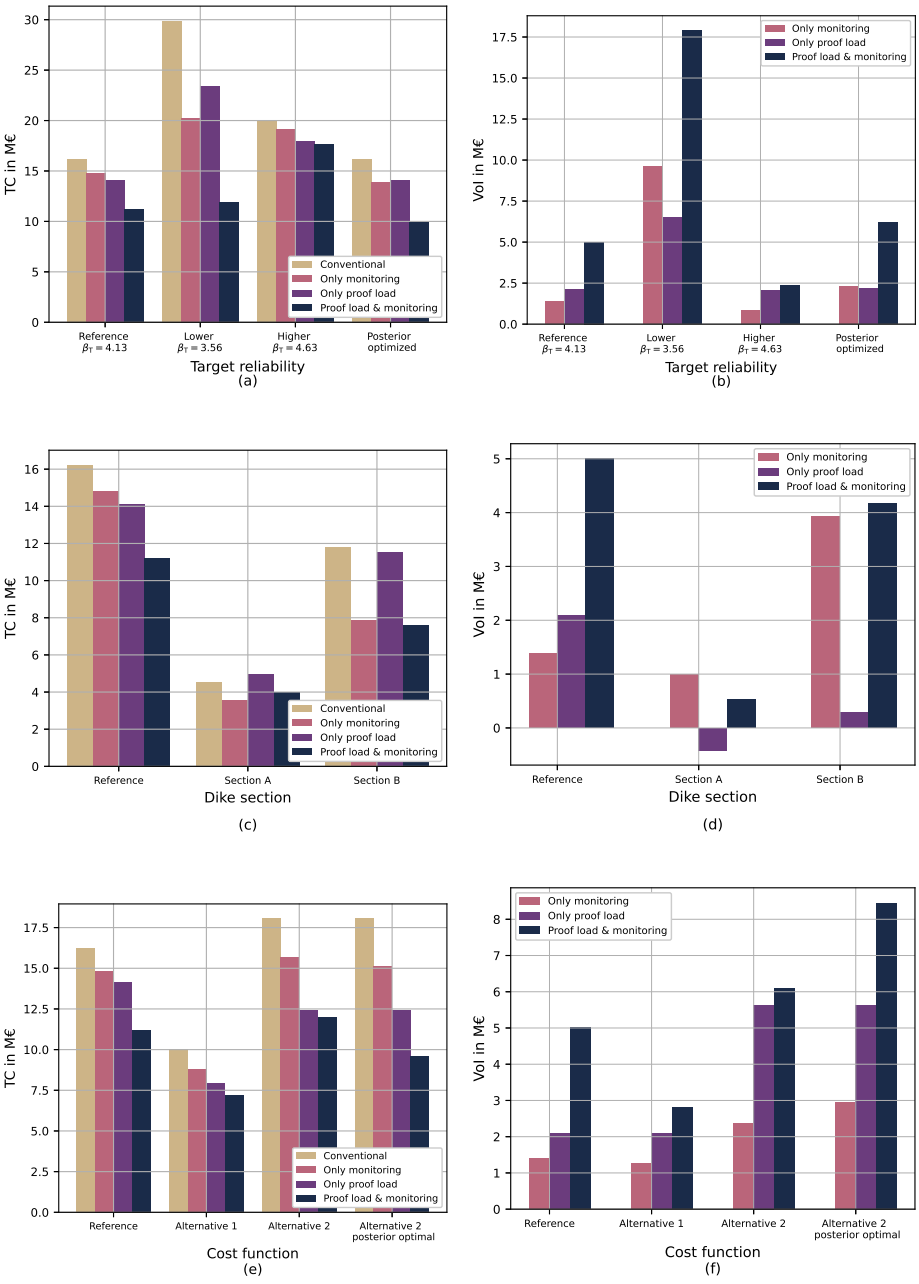


Figure 6.9: Total Cost (TC) and Value of Information (Vol) for different target reliability values (panes a and b), for different dike sections (panes c and d) and for different reinforcement cost functions (panes e and f). Proof load test level for all strategies is +13.0 m ref.. Conventional strategy has no proof load test and no monitoring.

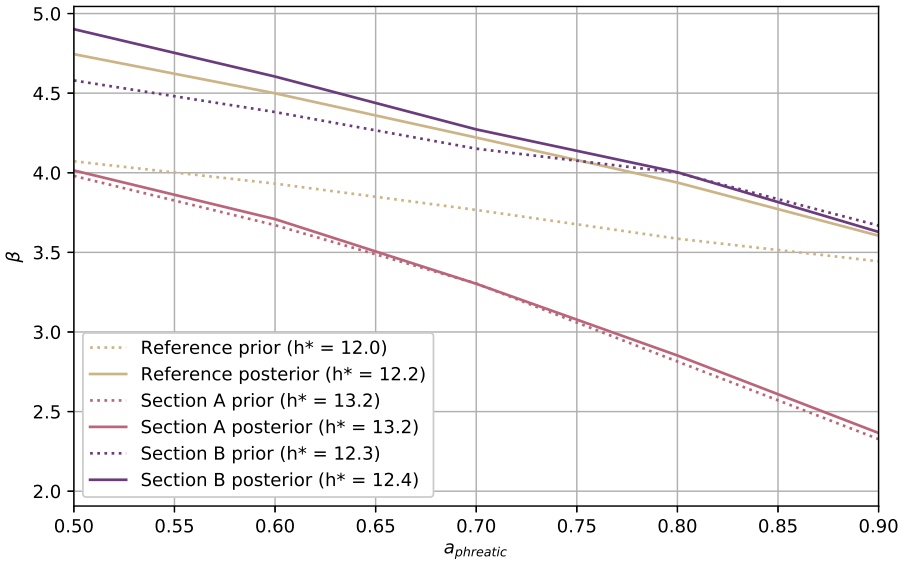


Figure 6.10: Fragility curves at the design point water level, showing an increasing dependency for the response factor $a_{\mathrm{p(hreatic)}}$ after proof loading (steeper curve).

Compared to the reference case, section A has considerably lower total cost as it is much closer to the target reliability (so the construction costs are much lower). At the same time, the VoI for proof loading is negative, which is due to the fact that the initial reliability is relatively high and the influence of soil parameter uncertainty is limited. Therefore a high proof load has to be applied to learn anything, which results in a higher probability of failure during the test. Thus, for this section proof loading adds very limited value. Although the uncertainty of soil parameters for section B is similar to that of section A, the fact that the initial reliability is lower results in a small but positive VoI for proof loading.

For pore pressure monitoring the VoI is positive in all cases. While the absolute VoI for section A is quite low compared to the other cases, relatively, monitoring reduces total cost by 22%. One thing that is quite apparent for the reference case is that the relevance of monitoring increases significantly once it is combined with a proof loading, which is not the case for the other cases with lower uncertainty in soil parameters. This can be explained as follows: a priori, the reliability in the reference case is hardly influenced by the response factor a_{p}, whereas, a posteriori, the reliability is dependent on the response factor. This is shown by a less steep fragility curve for the reference case in Figure 6.10. The curves are plotted conditional to the design point (i.e. most probable failure point) of the water level such that it best illustrates the contribution to the failure probability. So the results show that when geotechnical uncertainty is the dominant uncertainty in the prior failure probability (as it is only in the reference case), pore pressure monitoring is less effective than proof loading. After proof loading, geotechnical uncertainty is reduced, and pore pressure monitoring becomes much more effective.

Influence of different cost functions

Local differences in density of buildings, land prices, and available space for reconstruction, can significantly influence the costs for reinforcing dikes using stability berms. The reference dike section is considered for three different cost functions (see Figure 6.6). Figure 6.9e and 6.9f show total cost and Value of Information for the three different functions. Alternative 2 has relatively large benefits for proof loading, compared to the reference case (relative to total cost). This is caused by the lower marginal cost of the berm in €/m after proof loading, due to the fact that part of the cost function is less steep than the reference case. However, for alternative 2 the benefits of monitoring are much larger if the reliability requirement is optimized based on the posterior information after monitoring, rather than the prior information. Note that the same holds for alternative 1, but results are not shown. The reason is that the marginal costs of reinforcement differ per berm length. Henceforth, if the posterior reliability estimate differs significantly from the prior estimate, the marginal costs of reinforcement might change significantly as well. As a consequence, the initial reliability requirement might be suboptimal given the obtained information. Thus, especially if a cost function is highly non-linear, such a posterior optimization of the target reliability might yield significant benefits.

6.3.4. TOWARDS PRACTICAL IMPLEMENTATION OF THE DECISION FRAMEWORK

This case study demonstrates that in most cases pore pressure monitoring and proof loading yield a positive VoI for dikes that are sensitive to slope stability failures. However, proof loading is not economically efficient in all cases, and in some cases also pore pressure monitoring has very limited benefits. In a practical situation, decision makers therefore have to carefully consider what are the uncertainties that dominate the reliability, and from that determine measures to reduce these uncertainties (if available). For example, the design point (the approach used in this case study, see Figure 6.10) can provide indications to estimate the relative influence of different uncertain parameters, after which the proposed decision tree framework can be used to structure sequential decisions.

This case study only considered slope stability failures, whereas in practical situations there are often multiple failure modes that can be of relevance. Considering multiple failure modes will change the VoI for reduction of uncertainties in slope stability reliability, for instance if an increase in crest level is also required to mitigate risks from overtopping failure. However, the presented framework facilitates such a straightforward extension.

The failure probabilities in different years are assumed to be uncorrelated in this case study. While this is in line with common practice in flood defence reliability analysis, knowledge uncertainties on soil parameters are typically correlated in time. Consequently the future failure probability might be overestimated in cases with large knowledge uncertainty, most notably the case without uncertainty reduction. However, as the failure probabilities are relatively small, the overall effect is expected to be small as well (Klerk et al., 2018; Roubos et al., 2018).

In practice, reliability requirements are often prescribed by law, and are not necessarily derived solely on a local optimization of total cost, for instance requirements to loss

of life can also determine the target reliability. The sensitivity analysis in Section 6.3.3 shows that this can have a large influence both on total cost and on Value of Information. Aside from different target reliability levels that are optimal for prior information, a case where the target reliability level is optimized based on the posterior information after reducing uncertainty is also presented. It is found that this increases the VoI, in particular if the marginal cost of a dike reinforcement varies for different dimensions of the reinforcement (i.e., different increases in berm length). Specifically for cases with highly non-linear cost functions or jumps in cost functions, a local optimization based on posterior information after uncertainty reduction efforts can increase the effectiveness of uncertainty reduction, and flood risk management in general. The cases in this case study do not explicitly consider a fixed cost component, which could slightly lower the marginal costs. However, if 3 million € starting costs are added to the reference case the influence on VoI is still minor. Analysing different cost functions is straightforward within the presented framework.

While the influence of several important influential factors is explored, some are not. First of all, repair costs and other costs involved with proof loading can differ significantly depending on the design of the test. For example, damage and repair costs can be much larger than solely costs for fixing the dike itself, for instance if buildings are close by. In addition, there could also be immaterial consequences of a failing proof load test.

Secondly, it is assumed that proof loading is executed first, and after that pore pressure monitoring. However, in practice it might also make sense to alter the sequence of testing, for instance if it is expected that the outcome of pore pressure monitoring is already sufficient to ensure that the target reliability is met. Such strategies can be incorporated in the presented framework as well.

A third point concerns the inclusion of other methods for uncertainty reduction, most notably carrying out additional site investigation. This case study, in comparison to other decision analysis on reduction of geotechnical uncertainty (e.g., Schweckendiek and Vrouwenvelder (2013); Spross and Johansson (2017)), does combine multiple sequential uncertainty reduction effects. However, in a practical consideration also other approaches for uncertainty reduction such as additional soil investigation should not be overlooked and included in the analysis if they are found to be relevant based on an analysis of the most influential uncertainties. Next to that, it has to be noted that this chapter does not include potential uncertainty reduction on dike body permeability through a proof load test. This is an assumption that might influence the VoI estimates for the proof load test and in reality at least some information on this permeability might be obtained.

Spatial variability of the dike body in longitudinal direction might hinder the extrapolation of proof load test results from cross-section to dike section. In the present case study, it is assumed that the tested section is representative for the full dike section because of the relatively recent reconstruction of the inner slope. However, in other practical situations, additional site investigation might be required to substantiate the representativeness of the tested section, or else it remains uncertain how to translate the test results to other parts in longitudinal direction. Such site investigation efforts could then also be considered as a step in the presented framework.

Practical applications of pore pressure monitoring might or might not concern cross-

sections with a threshold, such as the old clay dike in the cross section considered in this study. If there is not such a clear threshold, including monitoring can be done in a similar manner, although more monitoring outcomes have to be considered than merely (no) observation. Another point of attention is that in this case a useful observation is obtained at a water level that occurs approximately once per 5 years. There might be situations where useful observations are less (or more) frequent, which obviously has an influence on the VoI of pore pressure monitoring. These considerations have been elaborated further in Klerk et al. (2019). The presented framework using a decision tree approach does facilitate adding additional outcomes or changing the threshold level.

6.3.5. SUMMARY AND CONCLUSION

This case study demonstrates the applicability of a decision tree framework in a sequential application of methods for reduction of geotechnical uncertainty. This framework can answer the question under what conditions to invest in different measures to reduce uncertainty for a dike section. The considered uncertainty reduction measures are a proof loading, which consists of artificially infiltrating the dike body with water and thus increasing the phreatic level in order to reduce uncertainty in soil properties, and pore pressure monitoring to reduce uncertainty in the response of the phreatic level to extreme hydraulic loads.

It is found that a strategy consisting of a proof load test and/or pore pressure monitoring has a positive VoI. The effectiveness of both methods depends greatly on the specific case. The relative reduction in total cost for the cases considered in this chapter ranges between 11% and 60% (on average 35%), of which the main contribution is a reduction in construction costs. However, the optimal strategy is not the same in all cases. Proof loading is most beneficial for cases where the uncertainty in soil properties is dominant and where the initial reliability is relatively low. Obviously the potential benefit must outweigh the additional risk of a failing proof load test and its costs. Pore pressure monitoring is most beneficial for cases where the uncertainty in the phreatic response is dominant.

Additionally, the influence of several factors is considered through a sensitivity analysis. The main findings are enlisted in Table 6.3, together with practical advice and remarks for implementation. For example, it is found that the choice of the target reliability requirement has a large influence on the estimate of the VoI. Therefore it is important that reliability requirements are adequately chosen, either by economic optimization or by other (optimized) requirements (e.g., Individual Risk). Only then the value of measures to reduce uncertainty can be quantified properly. Typically target reliability requirements are determined upfront (i.e., before monitoring and or proof load testing), but in this case study it was also considered whether optimizing reliability requirements after obtaining additional information improves decisions. It is found that this is typically the case, which is in line with the findings that a suboptimal choice of reliability requirements can obscure the results of the Value of Information analysis.

Decision makers can determine which measure might be worthwhile to consider in a VoI analysis by first identifying the dominant uncertainties determining the probability of failure. For example, plotting the conditional failure probability in fragility surfaces (as demonstrated in this case study) is found to be an effective and practical approach

Table 6.3: Overview of influential factors for decisions on proof loading and/or pore pressure monitoring. Some of these factors are influenceable by the decision maker (e.g., the proof load level), others are autonomous (e.g., amount of geotechnical uncertainty). For each factor a positive impact is named and potential remarks for practical implementation are given.

Influential factors for decision	**Positive impact**	**Remark**
Proof load level	Higher proof load, more uncertainty reduction.	The increased risk of failure does not always outweigh the potential benefits, especially if consequential damage is high.
Optimization of target reliability before uncertainty reduction	Can lead to significant reduction of total cost	In practice only possible if economic risk is the governing risk indicator rather than e.g., individual risk.
Optimization of target reliability after obtaining information	Reduction of total cost through inclusion of obtained information in target reliability	If target reliability was already optimized this will only be beneficial in very specific cases where information results in a posterior that strongly differs from the prior.
Larger geotechnical uncertainty	Proof loading is more effective	Pore pressure monitoring might become attractive only after reducing geotechnical uncertainty. It is recommended to determine the sequence of measures based on their relative uncertainty contribution and consider other methods (e.g., site investigation).
Higher construction cost of stability berms	Uncertainty reduction methods are more attractive as the benefits are larger.	Other methods for reinforcement might be more effective.

in identifying whether the soil properties or pore pressure are the dominant uncertainty; and thus whether to invest in proof loading or pore pressure monitoring. It was also shown that, in cases with large geotechnical uncertainty, the value of monitoring increases after a proof load, which demonstrates the relevance of considering multiple methods for uncertainty reduction in a single decision tree. In case other failure modes also have a significant contribution to flood risk, it is recommended to extend the approach to include these failure modes in the analysis.

Overall, this work puts in evidence to decision makers the criticality of carefully considering how and which uncertainties can be reduced, which is essential in achieving efficient flood defence asset management.

6.4. CONCLUDING REMARKS

This chapter aimed at answering the following research question:

> *How cost-effective are proof loading and pore water pressure monitoring in dike reinforcements?*

Bayesian decision analysis and pre-posterior analysis were used to provide insight in the factors that determine the cost-effectiveness of these uncertainty reduction methods. The main conclusions are:

- Proof loading and pore water pressure monitoring are typically cost-effective for dike sections in the Netherlands because the reduction of reinforcement cost is significant: up to 5 million € per km in the considered case study. Also, the investment costs are relatively low.
- Deliberately taking more risk through proof loading is worthwhile when the risk of failure during proof loading is not too large or when the risk is mitigated, and when the costs of traditional dike reinforcements are relatively high. Therefore, proof loading is especially interesting in dike reinforcements with space constraints, where expensive structural solutions would otherwise be required.
- Monitoring and proof loading are aimed at different uncertainties to reduce, and hence they are complementary. The results imply that different measures should be targeted at reducing different uncertainties. When time constraints limit the number of measures that can be implemented, the measure should be aimed at reducing the dominant uncertainty (i.e. the uncertainty that dominates the failure probability).

7

Conclusion

We need to carry out a vast amount of observational work, but what we do should be done for a purpose and done well.

Ralph Peck

7.1. FINDINGS

The aim of this dissertation was to improve reliability estimates for dike slope stability through considering performance information. The main question was "How does performance information affect reliability estimates for dike slope stability?".

In summary, by incorporating the information of survived (proof) loads and monitoring during the construction of dikes, reliability estimates become more credible, safety assessments improve, and design of dike reinforcements can be made more efficient. The cases considered in this dissertation suggest that savings of several million euros per kilometre dike reinforcement are possible (10-35% compared to the current dike reinforcement costs), for the Dutch situation with typically relatively high cost of dike reinforcements compared to the costs and risks of obtaining performance information. Even when it takes money or risk to obtain the performance information, a strategy with obtaining performance information can be cost-effective, and thus improve the efficiency of flood risk management. Four sub-questions were defined each centred around a specific aspect. The following paragraphs state the main findings for each research question.

To what extent is the uncertainty in the undrained shear strength reducible?

Quantitative reliability analysis of slope stability requires establishing prior estimates of the uncertainty of geotechnical parameters. For updating reliability estimates with additional information, uncertainty needs to be reducible, i.e. epistemic. The main uncertainties that affect the failure probability of slope instability are soil properties such as the undrained shear strength. The undrained shear strength is often derived from indirect measurements like cone penetration tests (CPTs) using transformation models.

The uncertainty in the spatially averaged undrained shear strength, when using transformation models calibrated with local data arises from several components. Spatial variability, measurement error, statistical error, and transformation error all contribute to the total uncertainty. Although spatial variability does lead to variation of soil properties over space, the property at a specific location is constant (but uncertain). The estimates, however, exhibit a considerable epistemic uncertainty because all error sources contributing to the total uncertainty of the spatial average are due to a lack of knowledge, and thus the uncertainty is in principle reducible.

The components contributing to the uncertainty of the transformation model are partially random and partially systematic. The random errors (e.g. measurement error, and due to spatial variability) are subject to averaging in estimating the spatial average. The remaining components in the uncertainty of the indirectly estimated spatial average for locations with a CPT, are the local bias in the transformation model and the statistical uncertainty.

The findings imply that there are several possibilities to reduce the (prior) uncertainty in the indirect estimate of the undrained shear strength (or any other parameters obtained in a similar manner). One option is to minimize the distance between a direct and indirect measurements, as spatial variability propagates into transformation uncertainty. Another option is to add direct measurements at different CPTs/boreholes, rather than at different depths in the same vertical, because the transformation model error is presumably systematic in one vertical (due to the same stress history).

How does the performance information contained in the survival of the construction phase affect the reliability estimates for slope stability of dikes?

Survived loads are a source of performance information that is informative for the stability of dikes. For dikes on soft soils, the stability during the construction stage is often critical. The survival of that loading condition with high excess pore pressures induced by raising the embankment contains information to update reliability estimates. Since the information regards survived conditions, the reliability always increases. The estimated failure probability can be more than a factor 10 lower if the observed survival during construction is incorporated (i.e. the non occurrence of instability). The degree to which epistemic uncertainty reduces (and thus the reliability increases), depends on the degree of criticality of the survived conditions and the degree of similarity between the survived and assessed situation.

The criticality relates to the extent to which stability in the survived situation is close to failure. For dikes on soft soils such as clay or organic soils, this is typically the case, due to excess pore pressures by incomplete consolidation during the raise of the embankment. For dikes constructed on highly permeable soils such as sand, the pore pressures during construction dissipate right away, and the stability during the construction is not a critical load. The degree of similarity refers to the failure mode, the loading condition, and the relative influence of uncertain parameters between the situations being assessed and the situation which is survived. The reliability update is significant when the uncertain parameters are largely epistemic, and when the failure mode in both conditions is similar, such as for dikes on soft subsoils.

The results indicate that no noticeable effect on the reliability is expected in situations where, for example, the soil blanket layer uplifts due to high pore pressures below the soft soil layer blanket, or when the design situation concerns extremely high phreatic levels due to overtopping waves. In such situations the construction survival is less informative. When considering a Bayesian analysis, a good indicator to examine the degree of similarity between the two situations is the correlation between the two limit states based on the FORM design points.

The findings also suggest that a higher raise of the dike than necessary for the design height can be considered to test the strength of the dike for stability, as long as the risk of failure outweighs the benefits in terms of an optimized design. The investigated examples indicate that significant reliability updates are possible, even when the probability of failure during construction is lower than 5%; a reasonable and generally accepted safety level during construction. Information about what conditions have been survived is necessary; practically this entails measurements of the (excess) pore pressures and the imposed load of the embankment raise.

In conclusion, incorporating the information of construction survival can significantly improve safety assessments of existing dikes. As the survival information is typically available over distances longer than a slip plane, the results (e.g. posterior reliability or updated estimates of the local average properties) are representative for horizontal sections that can be considered as homogeneous in terms of the schematized properties (typically in the order of hundreds of meters). The results also imply that in practice, many dikes have a considerably higher reliability than what they were designed for.

What is the effect of combining survival and settlements (multiple performance observations) during the construction of dikes?

Reliability estimates for slope stability improve further by combining the information of survival and settlement during the construction of dikes. Survival information and settlement information are complementary because the both sources of information are from different sources (independent data). Moreover, survival information is inequality data, for which the uncertainty reduction affects the tail of the joint probability distribution, and settlement information is equality data, which updates the best estimates of the settlement and stability prediction. Thus, the uncertainty reduction affects the probability of failure in two ways.

Though the correlation between settlement and stability is not very large, the estimated probability of failure in the considered case study was between 25 and 50% higher or lower (factor 1.2-1.5) depending on the settlement measurement. Settlement measurements may lead to more uncertainty reduction for stability when the correlation between settlement and stability is larger. For example if the uncertainty and the relative influence of the degree of over-consolidation of soft subsoil layers is larger in both models, or when there are more 'shared' parameters affecting both settlement and stability.

There are practical limits to adding more information from additional observations and from multiple sources, namely the computational feasibility of such Bayesian reliability analysis in practice. Because, for each piece of information that is added, the number of required model evaluations increases drastically. With geotechnical models that require only a few seconds to calculate, most standard reliability methods become computationally intractable, and implementation of the Bayesian analysis seems only practically feasible using metamodels. This is particularly the case for the safety assessment of flood defences and other geotechnical constructions in the Netherlands, with high target reliabilities (low probabilities of failure). A point of attention for using settlement measurements to update reliability is that the result of the Bayesian analysis is typically limited to the auto-correlation distance of settlements. In order for the reliability updating result to be representative for longer reaches of dike, settlement measurements of multiple locations are thus required.

How cost-effective are proof loading and pore water pressure monitoring in dike reinforcements?

Proof loading and monitoring reduce uncertainty, and thus improve reliability estimates. If dikes are strengthened to meet the required reliability (as defined by safety standards), dike designs that account for performance information will be - on average - smaller and less costly than conventional dike designs. However, whether the investment in proof loading or monitoring is worthwhile depends on the additional costs, for example the risk of a failing proof load. Because both the benefits (lower construction cost) and the costs (executing a proof load test and the risk of a failing test) increase with higher imposed proof loads, the cost-effectiveness is not straightforward.

The cost-effectiveness of proof loading and monitoring was evaluated using Bayesian Decision analysis. Proof loading and pore water pressure monitoring are typically cost-

effective for dike sections in the Netherlands because the reduction of reinforcement cost is significant and the investment costs are relatively low. For example in the considered case study, the Total Cost reduces with 1.5–5 million € per km for strategies with monitoring and/or proof loading, which is a 10-35% reduction compared to current dike reinforcement costs. Furthermore, the sensitivity analyses suggest savings ranging from 11% to 60% (on average 35%), mainly due to a reduction in dike reinforcement costs.

Deliberately taking more risk through proof loading is worthwhile when the risk of failure during proof loading is not too large or when the risk is mitigated, and when the costs of traditional dike reinforcements are relatively high. Therefore, proof loading is especially interesting in dike reinforcements with space constraints, where expensive structural solutions would otherwise be required. The results demonstrate that monitoring is more beneficial in combination with proof loading than without because the relative influence of pore pressure uncertainty increases after proof loading. This implies that different measures should be targeted at reducing different uncertainties.

The results of the sensitivity analyses show that the choice of the target reliability strongly influences the Total Cost. If the chosen target is higher than the economic optimum (for example because it is determined by other risk factors such as individual risk or societal risk), the Total Cost is higher through higher construction costs. If the chosen target is lower than the economic optimum, flood risk damage has a higher contribution to the Total Cost. In such cases the Total Cost can significantly be reduced by choosing higher reliability targets. The Total Cost is lowest if proof loading and monitoring is combined with economically optimized target reliabilities based on posterior information. The sensitivity analysis with respect to target reliability (factor 10 higher/lower) shows that while the Total Cost and Value of Information do change significantly, the Value of Information remains positive, indicating the combination of measures remains worthwhile.

7.2. RECOMMENDATIONS AND FURTHER RESEARCH

Based on the results and findings, this section formulates recommendations for improvements and further research.

- To improve prior estimates of uncertainty, (site) investigation should be aimed at reducing the systematic uncertainty (since random errors average), for example by repeating tests, and verifying results with different tests. For site-specific transformation models, we recommend to further investigate the extent to which transformation models are systematically or locally biased.

- This dissertation used performance information to update local spatial average soil properties, for sections that can be considered as homogeneous in terms of the schematized properties (typically in the order of hundreds of meters). It is recommended to investigate the effect of reliability updating when heterogeneity is explicitly modeled with auto-correlations instead of implicitly with local average properties. Specific attention should there be given to the difference in the auto-correlation of spatial variable components and epistemic uncertainties, and for what distance the performance information is representative.

- Additionally to investigating the effect of performance information with explicit modelling of heterogeneity, it is recommended to investigate how performance information can improve estimates of the usually very uncertain auto-correlation distance. For example the survival information during construction or historic floods is typically available over longer distances, which might improve estimates for the auto-correlation distance of the failure mechanism (length-effect). Settlement measurement with high spatial resolution might be of use to update correlation lengths of individual parameters such as the overconsolidation.

- Model uncertainty strongly affects reliability estimates. It is therefore recommended to investigate the extent to which the model uncertainty varies randomly or is systematically biased, and thus to what extent the model uncertainty is reducible.

- In order to optimally use the information contained in observations of the construction phase, it is essential to choose settlement and stability models such that the factors influencing both processes are appropriately captured in both, in order to fully reflect the degree of mutual correlation. The shared influence factors or parameters mostly concern the degree of over-consolidation and the permeability (and consolidation coefficients) governing the pore pressure response to external loading.

- The approximations for reliability updating using FORM and metamodels worked well with straightforward limit states such as only survival information, but not so well irregular limit states and combinations thereof such as settlement and surival. For Bayesian analysis in broader application in geotechnical engineering it is recommended to further explore the performance with more complex models, computationally demanding models, and combinations of multiple observations and limit states.

- The MultiNest algorithm is suitable for Bayesian updating, however less suitable for reliability analysis because the algorithm is not aimed at finding samples in the failure domain. Combining Multinest with Importance Sampling, and defining a convergence criterion on the probability of failure (in stead of the remaining evidence) can be an improvement. Another option to consider is ellipsoidal nested sampling to improve the selection of new samples in subset simulation.

7.3. RECOMMENDATIONS FOR VALORISATION IN PRACTICE

This dissertation is part of the All-Risk research programme which supports the Dutch Flood Protection programme (in Dutch: Hoogwaterbeschermingsprogramma, HWBP) with the implementation of the new safety standards. The primary goal of the HWBP is to ensure that all primary flood defences comply with the new safety standards before the year 2050. For that, many dike reinforcements are planned in the coming years. The HWBP is working towards this goal under the motto 'slim, sober en doelmatig' (English: smart, frugal, and efficient), of which 'frugal and efficient' relates to the cost of dike

reinforcement, mentioned in the Water Act. 'Smart' was added recently[1] to give room for ambitions to also improve spatial quality and improve the integration in the environment of such large and complex infrastructural projects.

The transition to safety standards defined in terms of acceptable probabilities of flooding offers great opportunities to improve flood risk management. In essence, this dissertation contributes to improved efficiency of flood risk management in the Netherlands in three ways. First, the probabilistic analyses of case studies of typical dike sections in the Netherlands provide insights in the dominant uncertainties to be reduced to obtain more credible probability estimates. Usually these are the local soil properties and the load effects. Secondly, this dissertation shows that performance information reduces uncertainty, and improves probability estimates. So, information during the construction of dikes, monitoring data, and proof loading can improve safety assessments. And lastly, this dissertation provides a framework to assess the cost-effectiveness of performance information in combination with dike reinforcement, in order to decide whether or not it is worthwhile to invest in performance information. The following paragraphs go further into detail how to utilize the concepts and knowledge developed in this dissertation in practice.

- Cone penetration tests are often used in dike reinforcement designs to estimate the undrained shear strength of soft soil layers. The transformation models needed to estimate soil properties based on cone resistance are often calibrated with local data, and involve large uncertainty. Chapter 3 presents a method to appropriately account for averaging of random errors and spatial variability. An alternative simplified method to roughly estimate the uncertainty is also presented. Recommendations for the use of local transformation models dike assessments and dike reinforcement projects are: 1) minimize the distance between a CPT and a borehole to prevent spatial variability to propagates into the transformation uncertainty, and 2) calibrate transformation models for statistically homogeneous geological deposits (on a regional scale) with sufficient independent measurements. Concretely, it is better to use multiple boreholes sufficiently distanced, rather than multiple measurements within one borehole. At least 10 but preferably 15-20 boreholes are recommended to decrease statistical uncertainty.

- The use of performance information during construction of dike reinforcements (such as the survival of the construction phase in Chapter 4 and measurements of settlement after raising dikes in Chapter 5) is important information for reducing uncertainties. Information about the construction phasing such as the embankment raise over time and excess pore pressures should be collected and used to optimize the design during the dike reinforcement, as it generally leads to less required space and less costly dike reinforcements. Settlements measurements can also be used to estimate the survived excess pore water pressures from the consolidation. The case studies in this dissertation and the HWBP are focused on primary flood defences in the Netherlands, however, the findings are also of particular importance to secondary/regional flood defences because the extreme

[1] Beleidsreactie advies College van Rijksadviseurs over ruimtelijke kwaliteit in het hoogwaterbeschermingsprogramma

situation might have a higher degree of similarity with the survived situation for those cases.

- Dike reinforcements that have already started (and where the surviving construction phase and other performance information during construction are not taken into account) can benefit from performance information in future safety assessments. This may result an extension of the lifetime, which is particularly interesting in the light of accelerated climate change, probably leading to changing boundary conditions such as additional sea level rise and more extreme flood events. A requirement is the availability of soil investigations, as-built information, and monitoring of pore water pressures, settlements and displacements during the construction phase. Such data is crucial to be able to continuously improve the understanding of dikes and to keep enhancing safety assessments. It is strongly recommended to continue monitoring pore pressures and settlements even after the dike reinforcement, as this performance data is invaluable for calibrating models and increasing knowledge about dikes on the long term in line with the recommendations of ENW[2] concerning long-term monitoring.

- Adopting a more critical staged loading schemes can be attractive to obtain a larger reliability update, and thus a less costly design. Such an approach is particularly interesting at locations with low consequential damage (no direct flooding, no damage to pipes, houses, etc.) and where otherwise expensive design solutions are needed. The variable size of a reinforcement design is expected to fall within the usual 'bandwidth' for the required space of a dike reinforcement. For practical implementation it seems reasonable that both the benefits and any additional risk are for the account of the financing institution. It is desirable that the uncertainty reduction from construction survival can be taken into account in easier ways (for example in practical guidelines), so the approach is widely accepted and also attractive to contractors (as dike reinforcements may involve less work after all).

- The decision tree framework presented in Chapter 6 seems very suitable for assisting dike managers and technical managers of dike reinforcements with deciding on what methods for uncertainty reduction are expected to be cost-effective. In particular, proof loading is efficient when geotechnical uncertainty is dominant in the failure probability estimates and where the risks of proof loading are low or can be mitigated. In dike reinforcements with space constraints, where expensive structural solutions would otherwise be required, it is typically worthwhile to obtain performance information by proof loading. Though applying a proof load to every dike is not feasible, proof loading has great potential as tailored solution for small dike sections where otherwise unrealistic, technically challenging, and expensive solutions would be required.

- Notice that implementation of proof loading and adopting a more critical staged loading scheme to optimize dike designs should always be conducted in conjunction with monitoring and a contingency plan. The Observational Method (Spross

[2] ENW advies 17-19: Beter Leren Keren door veldmetingen en monitoring

and Johansson, 2017; Nicholson et al., 1999; Peck, 1969) provides a theoretical framework for this and is highly recommended.

- Finally, this dissertation stresses the importance of performance information like survival observations and monitoring for later use. Future flood risk management will be directed towards continuously updating the remaining life time of the flood defence. That requires the possibility to retrieve and reuse historic information at later stages, thus requires rigorous data storage and management.

BIBLIOGRAPHY

Ahmed, A. and Soubra, A.-H. (2012). Probabilistic analysis of strip footings resting on a spatially random soil using subset simulation approach. *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, 6(3):188–201.

Au, S.-K. and Beck, J. L. (2001). Estimation of small failure probabilities in high dimensions by subset simulation. *Probabilistic Engineering Mechanics*, 16(4):263–277.

Au, S. K., Ching, J., and Beck, J. L. (2007). Application of subset simulation methods to reliability benchmark problems. *Structural Safety*, 29(3):183–193.

Azzouz, A. S., Baligh, M. M., and Ladd, C. C. (1983). Corrected field vane strength for embankment design. *Journal of Geotechnical Engineering*, 109(5):730–734.

Baecher, G. B. (2017). Bayesian thinking in geotechnics. In *Geo-Risk 2017*. American Society of Civil Engineers.

Baecher, G. B. and Christian, J. T. (2005). *Reliability and statistics in geotechnical engineering*. John Wiley & Sons.

Baecher, G. B. and Ladd, C. C. (1997). Formal Observational Approach to staged loading. *Transportation Research Record: Journal of the Transportation Research Board*, 1582(1):49–52.

Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London*, 53:370–418.

Benjamin, J. R. and Cornell, C. A. (1970). *Probability, statistics, and decision for civil engineers*. McGraw-Hill, New York.

Berre, T. and Bjerrum, L. (1975). Shear strength of normally consolidated clays. In *Proc. 8th International Conference on Soil Mechanics and Foundation Engineering*, volume 1, pages 39–49.

Betz, W., Papaioannou, I., Beck, J. L., and Straub, D. (2018). Bayesian Inference with Subset Simulation: Strategies and improvements. *Computer Methods in Applied Mechanics and Engineering*, 331:72–93.

Bishop, A. W. (1955). The use of the slip circle in the stability analysis of slopes. *Géotechnique*, 5(1):7–17.

Brinkman, R., Nuttall, J. D., Noordam, A. F., and Teixeira, A. C. M. (2018). Project software macro stability. User manual 2018, Part 3: Scientific Background. Technical report, Deltares report 11201523-005.

Calle, E. O. F., Knoeff, J. G., and Verheij, H. J. (2003). Residual strength after initial failure by overflow/overtopping. Technical report, Delft Cluster.

Calle, E. O. F., Schweckendiek, T., and Kanning, W. (2021). Characteristic values of soil properties in Dutch codes of practice. theoretical backgrounds and assumptions. Technical report, Deltares report.

Cao, Z., Wang, Y., and Li, D. (2017). *Probabilistic approaches for geotechnical site characterization and slope stability analysis*. Springer Berlin Heidelberg.

Cao, Z.-J., Wang, Y., and Li, D.-Q. (2016). Site-specific characterization of soil properties using multiple measurements from different test procedures at different locations A Bayesian sequential updating approach. *Engineering Geology*, 211:150–161.

CEN (2004). Eurocode 7: Geotechnical Design. Part 1: General Rules, EN1997-1. Technical report, European Committee for Standardisation.

Ching, J. and Phoon, K.-K. (2014). Transformations and correlations among some parameters of clays the global database. *Canadian Geotechnical Journal*, 51(6):663–685.

Ching, J. and Phoon, K.-K. (2020). Constructing a site-specific multivariate probability distribution using sparse, incomplete, and spatially variable (MUSIC-X) data. *Journal of Engineering Mechanics*, 146(7):04020061.

Ching, J., Phoon, K.-K., Chen, K.-F., Orr, T. L. L., and Schneider, H. R. (2020). Statistical determination of multivariate characteristic values for Eurocode 7. *Structural Safety*, 82:101893.

Ching, J., Phoon, K.-K., and Sung, S.-P. (2017). Worst case scale of fluctuation in basal heave analysis involving spatially variable clays. *Structural Safety*, 68:28–42.

Ching, J., Phoon, K.-K., and Wu, T.-J. (2016). Spatial correlation for transformation uncertainty and its applications. *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, 10(4):294–311.

Cho, S. E. (2007). Effects of spatial variability of soil properties on slope stability. *Engineering Geology*, 92(3-4):97–109.

Christian, J., Ladd, C., and Baecher, G. (1994). Reliability applied to slope stability analysis. *Journal of Geotechnical Engineering*, 120(12):5–417.

De Gast, T. (2020). *Dykes and embankments: a geostatistical analysis of soft terrain*. PhD thesis, Delft University of Technology.

De Koker, N., Day, P., and Zwiers, A. (2019). Assessment of reliability-based design of stable slopes. *Canadian Geotechnical Journal*, 56(4):495–504.

De Koning, M., Simanjuntak, T. D. Y. F., Goeman, D. G., Bakker, H. L., Haasnoot, J. K., and Bisschop, C. (2019). Determination of SHANSEP parameters by laboratory tests and CPTu for probabilistic model-based safety analyses. *Proceedings of the XVII European Conference on Soil Mechanics and Geotechnical Engineering*, pages 334–342.

Deltares (2014). KOSten voor versterking WATerkeringen. KOSWAT, systeemdocumentatie. Technical report, Deltares report.

Deltares (2018). POVM Eemdijkproef. Geotechnisch Basisrapport. Product F. Technical report, Deltares report 11200956-002-GEO-0015 (in Dutch).

Deltares (2019). D-Stability. user manual. Technical report, Deltares.

Diamantidis, D., Sykora, M., and Sousa, H. (2019). Quantifying the Value of Structural Health Information for decision support: guide for practising engineers. In *COST Action TU1402 Guidelines*. COST Action TU1402.

Echard, B., Gayton, N., and Lemaire, M. (2011). AK-MCS: An active learning reliability method combining Kriging and Monte Carlo Simulation. *Structural Safety*, 33(2):145–154.

Eijgenraam, C., Brekelmans, R., den Hertog, D., and Roos, K. (2017). Optimal strategies for flood prevention. *Management Science*, 63(5):1644–1656.

Eijnden, A. P., Schweckendiek, T., and Hicks, M. A. (2021). Metamodelling for geotechnical reliability analysis with noisy and incomplete models. *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, pages 1–18.

ENW (2017). *Fundamentals of flood protection*. Ministry of Infrastructure and the Environment and the Expertise Network for Flood Protection.

Feijten (Photographer), E. (2018). De ringdijk bij Eemnes vanuit de lucht.

Fellenius, W. (1936). Calculation of the stability of earth dams. In *Trans. 2nd Int. Cong. Large Dams (Washington)*, pages 445–459.

Fenton, G. A. and Griffiths, D. V. (2002). Probabilistic foundation settlement on spatially random soil. *Journal of Geotechnical and Geoenvironmental Engineering*, 128(5):381–390.

Feroz, F., Hobson, M. P., and Bridges, M. (2009). MultiNest: an efficient and robust Bayesian inference tool for cosmology and particle physics. *Monthly Notices of the Royal Astronomical Society*, 398(4):1601–1614.

Feroz, F., Hobson, M. P., Cameron, E., and Pettitt, A. N. (2019). Importance nested sampling and the MultiNest algorithm. *The Open Journal of Astrophysics*, 2(1).

Frangopol, D. M., Strauss, A., and Kim, S. (2008). Use of monitoring extreme data for the performance prediction of structures: General approach. *Engineering Structures*, 30(12):3644–3653.

Greene, W. (2002). *Econometric analysis*. Pearson Education, Upper Saddle River, NJ.

Griffiths, D. V. and Fenton, G. A. (2004). Probabilistic slope stability analysis by finite elements. *Journal of geotechnical and geoenvironmental engineering*, 130(5):507–518.

Griffiths, D. V., Huang, J., and Fenton, G. A. (2009). Influence of spatial variability on slope reliability using 2-D random fields. *Journal of Geotechnical and Geoenvironmental Engineering*, 135(10):1367–1378.

Griffiths, D. V. and Lane, P. A. (1999). Slope stability analysis by finite elements. *Geotechnique*, 49(3):387–403.

Grooteman, F. (2011). An adaptive directional importance sampling method for structural reliability. *Probabilistic Engineering Mechanics*, 26(2):134–141.

Hasofer, A. and Lind, N. (1974). Exact and invariant second-moment code format. *Journal of the Engineering Mechanics division*, 100(1):111–121.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109.

Hicks, M. A. and Samy, K. (2002). Influence of heterogeneity on undrained clay slope stability. *Quarterly Journal of Engineering Geology and Hydrogeology*, 35(1):41–49.

Hsu, W.-C. and Ching, J. (2010). Evaluating small failure probabilities of multiple limit states by parallel subset simulation. *Probabilistic Engineering Mechanics*, 25(3):291–304.

Huber, M., van der Krogt, M., and Kanning, W. (2016). Probabilistic slope stability analysis using approximative FORM. In *14th International Probabilistic Workshop*, pages 299–316. Springer International Publishing.

HWBP (2020). *Projectenboek 2021*. Hoogwaterbeschermingsprogramma.

ISSMGE-TC304 (2021). *State-of-the-art review of inherent variability and uncertainty in geotechnical properties and models*. International Society for Soil Mechanics and Geotechnical Engineering (ISSMGE) - Technical Committee TC304 'Engineering Practice of Risk Assessment and Management'.

Jiang, S.-H., Huang, J., Qi, X.-H., and Zhou, C.-B. (2020). Efficient probabilistic back analysis of spatially varying soil parameters for slope reliability assessment. *Engineering Geology*, 271:105597.

Jiang, S.-H., Li, D.-Q., Cao, Z.-J., Zhou, C.-B., and Phoon, K.-K. (2015). Efficient system reliability analysis of slope stability in spatially variable soils using Monte Carlo simulation. *Journal of Geotechnical and Geoenvironmental Engineering*, 141(2):04014096.

Jiang, S.-H., Papaioannou, I., and Straub, D. (2018). Bayesian updating of slope reliability in spatially variable soils with in-situ measurements. *Engineering Geology*, 239:310–320.

Jongejan, R. B. (2017). WBI2017 Code calibration. reliability-based code calibration and semi-probabilistic assessmentrules for the WBI2017. Technical report, Rijkswaterstaat.

Jongejan, R. B., Diermanse, F., Kanning, W., and Bottema, M. (2020). Reliability-based partial factors for flood defenses. *Reliability Engineering & System Safety*, 193:106589.

Jonkman, S. N., Voortman, H. G., Klerk, W. J., and van Vuren, S. (2018). Developments in the management of flood defences and hydraulic infrastructure in the Netherlands. *Structure and Infrastructure Engineering*, 14(7):895–910.

Juang, C. H., Gong, W., Martin, J. R., and Chen, Q. (2018). Model selection in geological and geotechnical engineering in the face of uncertainty - Does a complex model always outperform a simple model? *Engineering Geology*, 242:184–196.

Kanning, W. (2012). *The weakest link. Spatial variability in the piping failure mechanism of dikes*. PhD thesis, Delft University of Technology.

Kanning, W., Teixeira, A., van der Krogt, M., Rippi, K., Schweckendiek, T., and Hardeman, B. (2017). Calibration of factors of safety for slope stability of dikes. In *Geo-Risk 2017*. American Society of Civil Engineers.

Kanning, W., ter Horst, W., Pol, J., Kok, M., Knoeff, H., Hüsken, L., and de Visser, M. (2019). Faalpaden. Mogelijkheden faalpaden voor Nederlandse dijkveiligheidsanalyses. Technical report, Deltares report 11202560-004-GEO-0006. In Dutch.

Kanning, W. and Van der Krogt, M. G. (2016). Pore water pressure uncertainties for slope stability. Technical report, Deltares report 1230090-034.

Kentrop, D. (2021). A metamodelling approach to reliability updating with dike construction survival. Master's thesis, Delft University of Technology.

Kiureghian, A. D. and Ditlevsen, O. (2009). Aleatory or epistemic? Does it matter? *Structural Safety*, 31(2):105–112.

Klerk, W., Kanning, W., and Kok, M. (2018). Time-dependent reliability in flood protection decision making in the Netherlands. In *Safety and Reliability – Safe Societies in a Changing World*, pages 3167–3174. CRC Press.

Klerk, W. J., Kanning, W., Kok, M., and Wolfert, R. (2021). Optimal planning of flood defence system reinforcements using a greedy search algorithm. *Reliability Engineering & System Safety*, 207:107344.

Klerk, W. J., Schweckendiek, T., den Heijer, F., and Kok, M. (2019). Value of information of structural health monitoring in asset management of flood defences. *Infrastructures*, 4(3):56.

Koelewijn, A., de Vries, G., van Lottum, H., Förster, U., van Beek, V., and Bezuijen, A. (2013). Full-scale testing of piping prevention measures. In *ICPMG2014 – Physical Modelling in Geotechnics*, pages 891–897. CRC Press.

Kroese, D. P. and Botev, Z. I. (2015). Spatial process simulation. In *Stochastic geometry, spatial statistics and random fields*, pages 369–404. Springer.

Ladd, C. (1991). Stability evaluation during staged construction. *Journal of Geotechnical Engineering*, 117(4):540–615.

Ladd, C. and Degroot, D. (2003). Recommended practice for soft ground site characterization. In *Proc., 12th Panamerican Conference on Soil Mechanics and Geotechnical Engineering*, pages 1–55, Cambridge, United States of America.

Ladd, C. C. and Foott, R. (1974). New design procedure for stability of soft clays. *Journal of the geotechnical engineering division*, 100(7):763–786.

Laplace, P. S. and Truscott, F. W. (1951). *A philosophical essay on probabilities*, volume 166. Dover Publications New York.

Lemaire, M., Chateauneuf, A., and Mitteau, J.-C. (2009). *Structural Reliability*. ISTE and John Wiley & Sons.

Lengkeek, H., Post, M., Breedeveld, J., and Naves, T. (2019). Eemdijk full-scale field test programme: ground dyke and sheet pile dyke failure test. In *Proc., XVII European Conference on Soil Mechanics and Geotechnical Engineering*, Iceland, Reykjavik.

Li, D.-Q., Zhang, F.-P., Cao, Z.-J., Zhou, W., Phoon, K.-K., and Zhou, C.-B. (2015). Efficient reliability updating of slope stability by reweighting failure samples generated by Monte Carlo simulation. *Computers and Geotechnics*, 69:588–600.

Li, D.-Q., Zheng, D., Cao, Z.-J., Tang, X.-S., and Phoon, K.-K. (2016). Response surface methods for slope reliability analysis: Review and comparison. *Engineering Geology*, 203:3–14.

Li, X., Zhang, L., and Zhang, S. (2018). Efficient Bayesian networks for slope safety evaluation with large quantity monitoring information. *Geoscience Frontiers*, 9(6):1679–1687.

Li, Y. (2017). *Reliability of long heterogeneous slopes in 3D: Model performance and conditional simulation*. PhD thesis, Delft University of Technology.

Lumb, P. (1966). The variability of natural soils. *Canadian Geotechnical Journal*, 3(2):74–97.

Luque, J. and Straub, D. (2019). Risk-based optimal inspection strategies for structural systems using dynamic Bayesian Networks. *Structural Safety*, 76:68–80.

Melchers, R. E. and Beck, A. T. (2018). *Structural Reliability Analysis and Prediction*. John Wiley & Sons Ltd, Hoboken, NJ.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092.

MinIE (2015). Our water in the Netherlands. New National Water Plan 2016-2021. Ministery of Infrastructure and the Environment.

Morgenstern, N. R. and Price, V. E. (1965). The analysis of the stability of general slip surfaces. *Geotechnique*, 15(1):79–93.

Muhammed, J. J., Jayawickrama, P. W., and Ekwaro-Osire, S. (2020). Uncertainty analysis in prediction of settlements for spatial prefabricated vertical drains improved soft soil sites. *Geosciences*, 10(2):42.

Nadim, F. (2015). Accounting for uncertainty and variability in geotechnical characterization of offshore sites. In *Proc. 5th Int. Symp. Geotechnical Safety and Risk*, pages 23–35.

Neumann, J. v. and Morgenstern, O. (1947). *Theory of games and economic behavior*. Princeton, NJ. Princeton University Press, 2nd edition.

Nicholson, D. P., Tse, C., and Penny, C. (1999). The Observational Method in ground engineering - Principles and applications. Technical report, CIRIA report 185, London, UK.

Ochiai, H. (1980). Undrained strength of normally consolidated clay measured in the simple shear test. Technical report, Nagasaki University Research Report.

Papaioannou, I., Betz, W., Zwirglmaier, K., and Straub, D. (2015). MCMC algorithms for subset simulation. *Probabilistic Engineering Mechanics*, 41:89–103.

Papaioannou, I. and Straub, D. (2012). Reliability updating in geotechnical engineering including spatial variability of soil. *Computers and Geotechnics*, 42:44–51.

Peck, R. B. (1969). Advantages and limitations of the Observational Method in applied soil mechanics. *Géotechnique*, 19(2):171–187.

Peng, M., Li, X. Y., Li, D. Q., Jiang, S. H., and Zhang, L. M. (2014). Slope safety evaluation by integrating multi-source monitoring information. *Structural Safety*, 49:65–74.

Phoon, K.-K. and Kulhawy, F. H. (1999a). Characterization of geotechnical variability. *Canadian Geotechnical Journal*, 36(4):612–624.

Phoon, K.-K. and Kulhawy, F. H. (1999b). Evaluation of geotechnical property variability. *Canadian Geotechnical Journal*, 36(4):625–639.

Phoon, K. K. and Retief, J. V. (2016). *Reliability of Geotechnical Structures in ISO2394*. CRC Press.

Raiffa, H. and Schlaifer, R. (1961). *Applied statistical decision theory*. Division of Research, Harvard Business School, Boston, MA., Boston, MA.

Remmerswaal, G., Vardon, P. J., and Hicks, M. A. (2021). Evaluating residual dyke resistance using the Random Material Moint Method. *Computers and Geotechnics*, 133:104034.

Roscoe, K., Diermanse, F., and Vrouwenvelder, A. C. W. M. (2015). System reliability with correlated components: Accuracy of the Equivalent Planes Method. *Structural Safety*, 57:53–64.

Roubos, A. A., Steenbergen, R. D. J. M., Schweckendiek, T., and Jonkman, S. N. (2018). Risk-based target reliability indices for quay walls. *Structural Safety*, 75:89–109.

RWS (2019). Schematiseringshandleiding macrostabiliteit WBI-2017 version 3.0 (in Dutch). Technical report, Rijkswaterstaat. Ministry of Infrastructure and Water Management.

Schofield, A. and Wroth, P. (1968). *Critical State Soil Mechanics*. McGraw-Hill.

Schweckendiek, T. (2010). Reassessing reliability based on survived loads. In *Proc. International Conference of Coastal Engineering*, pages 1–14, Shanghai.

Schweckendiek, T., van der Krogt, M. G., Teixeira, A., Kanning, W., Brinkman, R., and Rippi, K. (2017). Reliability updating with survival information for dike slope stability using fragility curves. In *Geo-Risk 2017*. American Society of Civil Engineers.

Schweckendiek, T., Vrouwenvelder, A. C. M., Calle, E. O. F., Kanning, W., and Jongejan, R. B. (2013). Target reliabilities and partial factors for flood defenses in The Netherlands. *Advances in Soil Mechanics and Geotechnical Engineering*, 1:311–328.

Schweckendiek, T. and Vrouwenvelder, A. C. W. M. (2013). Reliability updating and decision analysis for head monitoring of levees. *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, 7(2):110–121.

Schweckendiek, T., Vrouwenvelder, A. C. W. M., and Calle, E. O. F. (2014). Updating piping reliability with field performance observations. *Structural Safety*, 47:13–23.

Shewbridge, S. and Schaefer, J. (2013). Some unexpected modern complications in seepage and slope stability analysis: Modeling pore pressures and strengths for flood-loaded structures. In *Proc. Annual Conference, Association of State Dam Safety Officials*, pages 694–714.

Skilling, J. (2004). Nested sampling. In *AIP Conference Proceedings*, volume 735, pages 395–405. American Institute of Physics.

Skilling, J. (2006). Nested sampling for general Bayesian computation. *Bayesian Analysis*, 1(4).

Spencer, E. (1967). A method of analysis of the stability of embankments assuming parallel inter-slice forces. *Geotechnique*, 17(1):11–26.

Spross, J. and Johansson, F. (2017). When is the observational method in geotechnical engineering favourable? *Structural Safety*, 66:17–26.

Spross, J. and Larsson, S. (2021). Probabilistic observational method for design of surcharges on vertical drains. *Géotechnique*, pages 1–13.

Straub, D. (2011). Reliability updating with equality information. *Probabilistic Engineering Mechanics*, 26(2):254–258.

Straub, D. and Papaioannou, I. (2014). Bayesian updating with structural reliability methods. *Journal of Engineering Mechanics*, 141(3):04014134.

Straub, D. and Papaioannou, I. (2015). *Risk and Reliability in Geotechnical Engineering*, chapter Bayesian analysis for learning and updating geotechnical parameters and models with measurements, pages 221–264. CRC Press Boca Raton, FL.

Straub, D., Papaioannou, I., and Betz, W. (2016). Bayesian analysis of rare events. *Journal of Computational Physics*, 314:538–556.

Stuedlein, A. W., Kramer, S. L., Arduino, P., and Holtz, R. D. (2012). Geotechnical characterization and random field modeling of desiccated clay. *Journal of Geotechnical and Geoenvironmental Engineering*, 138(11):1301–1313.

Sun, Y., Huang, J., Jin, W., Sloan, S. W., and Jiang, Q. (2019). Bayesian updating for progressive excavation of high rock slopes using multi-type monitoring data. *Engineering Geology*, 252:1–13.

't Hart, R., de Bruijn, H., and de Vries, G. (2016). Fenomenologische beschrijving faalmechanismen WTI. Technical report, Deltares report 1220078-000-GEO-001.

Tabarroki, M., Ching, J., Phoon, K.-K., and Chen, Y.-Z. (2021). Mobilisation-based characteristic value of shear strength for ultimate limit states. *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, pages 1–22.

Tavenas, F., Mieussens, C., and Bourges, F. (1979). Lateral displacements in clay foundations under embankments. *Canadian Geotechnical Journal*, 16(3):532–550.

TAW (1989). Leidraad voor het ontwerpen van rivierdijken. Deel 2 benedenrivierengebied (in Dutch). Technical report, Technische Adviescommissie voor de Waterkeringen.

Teixeira, R., Nogal, M., and O'Connor, A. (2021). Adaptive approaches in metamodel-based reliability analysis: A review. *Structural Safety*, 89:102019.

Thöns, S. (2017). On the Value of monitoring Information for the structural integrity and risk management. *Computer-Aided Civil and Infrastructure Engineering*, 33(1):79–94.

U.S. Army Corps of Engineers (USACE) (2003). EM 11102-1902, Engineering Manual, Engineering and Design, Slope Stability. Technical report, USACE.

Van, M. (2001). New approach for uplift induced slope failure. In *Proc. XVth International Conference on Soil Mechanics and Geotechnical Engineering*, pages 2285–2288, Istanbul.

Van den Eijnden, A. P. and Hicks, M. A. (2017). Efficient subset simulation for evaluating the modes of improbable slope failure. *Computers and Geotechnics*, 88:267–280.

Van den Eijnden, A. P. and Hicks, M. A. (2019). On the importance of a complete characterization of site investigation data uncertainty: a computational example. In *Proceedings of the 7th international symposium on geotechnical safety and risk (ISGSR 2019)*, pages 237–242.

Van der Krogt, M. (2018). POVM eemdijkproeven. factual report monitoring aanleg full-scaleproeven. Technical report, Deltares report 11200956-000-GEO-0003 (in Dutch).

Van der Krogt, M. G. and Schweckendiek, T. (2019). Systematic transformation error in the depth-average undrained shear strength. In *Proceedings of the 7th International Symposium on Geotechnical Safety and Risk (ISGSR 2019)*. Research Publishing Services.

Van der Krogt, M. G., Schweckendiek, T., and Kok, M. (2018). Uncertainty in spatial average undrained shear strength with a site-specific transformation model. *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, 13(3):226–236.

Van der Krogt, M. G., Schweckendiek, T., and Kok, M. (2019). Do all dike instabilities cause flooding? In *13th International Conference on Applications of Statistics and Probability in Civil Engineering (ICASP13), Seoul, South Korea, May 26-30, 2019*.

Van der Meer, A. W., Schweckendiek, T., Kruse, H., and Schelfhout, H. (2019). Integral failure analysis of pipelines in flood defenses. In Ching, J., Li, D.-Q., and Zhang, . J., editors, *Proceedings of the 7th International Symposium on Geotechnical Safety and Risk*, pages 810–815.

Van Duinen, A. (2015). Modelonzekerheidsfactoren Spencer-Van der Meij model en ongedraineerde schuifsterkte. Technical report, Deltares report 1207808-001 (in Dutch).

Van Duinen, T. A. (2019). Variability and accuracy of shear strength measurements in soft soils. mathesis, Utrecht University.

Van Gelder, P. H. A. J. M. (2000). *Statistical methods for the risk-based design of civil structures*. PhD thesis, Delft University of Technology.

Van Hoven, A. and Noordam, A. (2018). POVM Infiltratieproef II - Analyse infiltratieproef IJsseldijk (in dutch). Technical report, Deltares report 11202663-002.

Vanmarcke, E. (1983). *Random fields: analysis and synthesis*. MIT Press, Cambridge, Massachusetts.

Vanmarcke, E. H. (1977). Probabilistic modeling of soil profiles. *Journal of the Geotechnical Engineering Division*, 103(11):1227–1246.

Vanmarcke, E. H. (1980). Probabilistic stability analysis of earth slopes. *Engineering Geology*, 16:90005–90013.

Vardon, P. J., Liu, K., and Hicks, M. A. (2016). Reduction of slope stability uncertainty based on hydraulic measurement via inverse analysis. *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, 10(3):223–240.

Varkey, D., Hicks, M. A., and Vardon, P. J. (2019). An improved semi-analytical method for 3D slope reliability assessments. *Computers and Geotechnics*, 111:181–190.

Voortman, H. G. (2003). *Risk-based design of large-scale flood defence systems*. PhD thesis, Delft University of Technology.

Vrijling, J. K. (2001). Probabilistic design of water defense systems in The Netherlands. *Reliability Engineering & System Safety*, 74(3):337–344.

Waarts, P. H. (2000). *Structural reliability using finite element methods. An appraisal for DARS: Directional Adaptive Response surface Sampling.* PhD thesis, Delft University of Technology.

Wang, B., Hicks, M. A., and Vardon, P. J. (2016). Slope failure analysis using the random material point method. *Géotechnique Letters*, 6(2):113–118.

Watabe, Y., Tsuchida, T., and Adachi, K. (2002). Undrained shear strength of pleistocene clay in Osaka Bay. *Journal of Geotechnical and Geoenvironmental Engineering*, 128(3):216–226.

Waterveiligheidsportaal (2021). Veiligheidsoordeel per dijktraject. URL: https://waterveiligheidsportaal.nl/. Accessed 6 Oct 2021.

Waterwet (2021). URL: https://wetten.overheid.nl/BWBR0025458.

WBI (2013). Vaststellen uitgangspunten definitieve kalibratie. Technical report, WBI 2017, cluster Raamwerk. Deltares report 1207803-003-GEO-0003.

Zhang, J., Huang, H. W., Juang, C. H., and Su, W. W. (2014). Geotechnical reliability analysis with limited data: Consideration of model selection uncertainty. *Engineering Geology*, 181:27–37.

Zhang, J., Zhang, L. M., and Tang, W. H. (2011). Slope reliability analysis considering site-specific performance information. *Journal of Geotechnical and Geoenvironmental Engineering*, 137(3):227–238.

Zheng, D., Huang, J., Li, D.-Q., Kelly, R., and Sloan, S. W. (2018). Embankment prediction using testing data and monitored behaviour: A Bayesian updating approach. *Computers and Geotechnics*, 93:150–162.

Zhu, D., Griffiths, D. V., and Fenton, G. A. (2019). Worst-case spatial correlation length in probabilistic slope stability analysis. *Géotechnique*, 69(1):85–88.

APPENDICES

A

PRE-POSTERIOR ANALYSIS: DIKE REINFORCEMENT WITH CONSTRUCTION SURVIVAL

INTRODUCTION

Chapter 4 concluded that dikes that are successfully built, generally have a higher posterior reliability than what they were designed for. Thus, the reliability is higher than the target reliability, and therefore, not optimal. This case study demonstrates how the information of the survival of the construction phase can be used to optimize dike designs, and whether it is cost-effective to adopt a more risky construction phasing as it results in more uncertainty reduction, but also in higher chances of failure during the construction.

The decision problem is formulated as follows. The case study considers a typical situation of a dike that needs to be raised to prevent overtopping, and needs to be strengthened with respect to the slope stability. Traditionally, dikes are strengthened before the dikes are raised, as this is the safest building sequence. In such situations, the survived situation is not a critical loading situation and incorporating the survival information barely effects the posterior reliability. Therefore a dike reinforcement based on posterior information will not be notably different from a design based on prior information.

In this case study we investigate an alternative construction sequence of first raising the dike, and then strengthen it, see Figure A.1. This reversed construction phasing is more risky because the stability will be lower during construction. However, if the construction is survived, the reliability can significant increase, see Kentrop (2021). The higher reliability requires a less wide stability berm to meet a fixed reliability target. The question is, however, if the reversed construction phasing is worthwhile the additional risk.

A

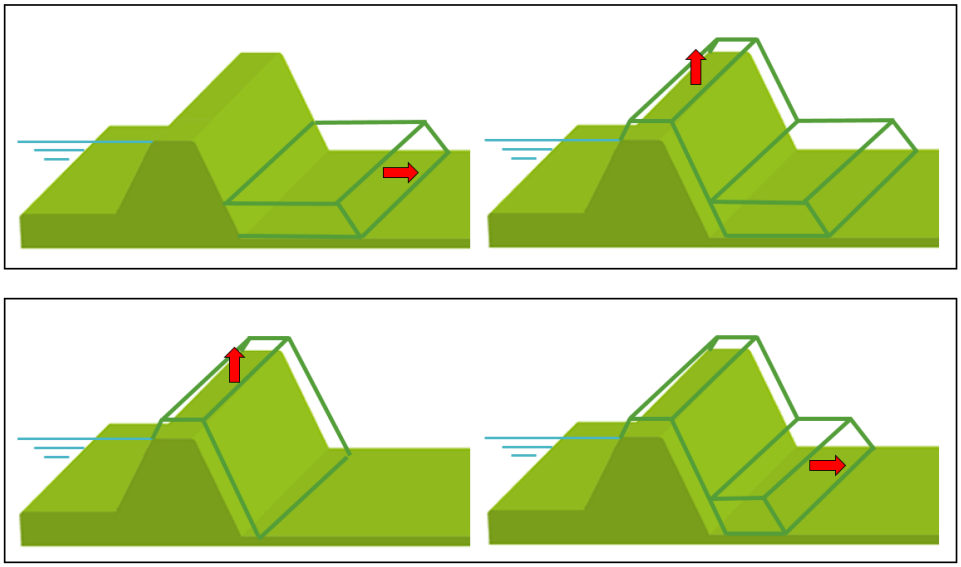


Figure A.1: Traditional building sequence of first constructing stability berm and then raising the dike (top), and more risky building sequence of first raising the dike and then constructing a stability berm (bottom).

METHOD

Whether the reversed construction sequence is cost-effective depends on different factors, among others, the probability of failure during construction, the costs of damage related to failure, and the savings of a cheaper dike design. To investigate the cost-effectiveness of such a pre-posterior design of a dike reinforcement compared to a conventional strategy based on prior information, we adopt a simplified version of the decision framework introduced in Section 6.2.

To investigate the sensitivities and circumstances that lead to whether pre-posterior dike reinforcements are cost efficient, we compare the costs of a design based on prior information with a traditional construction phasing (s_1), and a pre-posterior design incorporating the information of survived construction with a reversed construction phasing (s_2). Similar to Equation 6.2, the expected direct cost of a strategy is calculated as follows:

$$E[c(s_i)] = C_{\text{reinforcement}}(d(s_i)) + C_{\text{damage}} \cdot P(\overline{S}|s_i) \tag{A.1}$$

With d the decision rule determining the reinforcement action a needed to meet the reliability target after incorporating the information of the survived construction, see Figure A.3. In this case: the required width of the stability berm. C_{damage} is the cost of damage and repairs related to instability during the construction (including the construction cost of a new dike in case the proof load fails). Note that we keep the contribution of flood risk out of the equation because it does not differences between the direct cost of a strategy, since in both situations we reinforce the dike to the same target reliability. We also assume that reversing the construction phasing does not lead

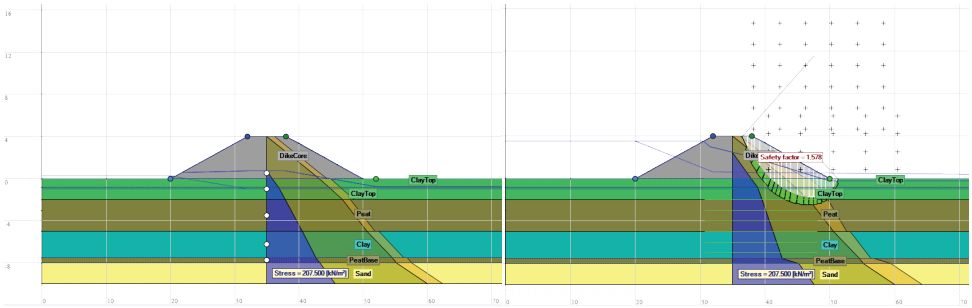


Figure A.2: Cross section of the case study in daily (left) and flood conditions (right).

to additional costs for example due to differences in total duration (for consolidation and settlements).

To optimize the dike design, we account for the survival of the construction stage in advance. In fact, we want to determine the design for which the (pre-)posterior reliability meets the target reliability. The same Bayesian updating method is used as in Chapter 4, but in stead of looking at newly built dikes, we look at existing dikes on soft soil foundations. Though this may lead to less critical stability situations than new dikes, it leads to a larger degree similarity between the survived and future situation.

Case description

The case study is a fictitious existing dike of clay/silt with a slope 1:3 (v:h) and a initial height of +4.0 m ref.. The subsoil is a typical Dutch composition of different soft soil layers: 2 m clay, 3 m peat, 2 m clay, and 0.5 m basal peat, on a Pleistocene sand layer. In daily circumstances, the phreatic level is around +0.75 m ref.and the head in the soft soil and Pleistocene sand layer −0.75 m ref.and −1.0 m ref. respectively, see Figure A.2. The soil parameters are listed in Table A.1.

We regard a situation where the total dike raise is deterministic 0.6 m, and the design for the slope stability must meet an prescribed target reliability assumed β_T=5.04. While ideally the dike height is also part of the risk optimization, we focus only on the comparison between prior and pre-posterior design for slope stability. To reduce the degrees of freedom, we only vary the width of the stability berm as measure to solve the stability deficit, and do not consider other solutions (that may be more efficient in terms of total cost).

Results

The prior reliability analysis for the slip circle indicated in Figure A.2 results in a reliability estimate of β=4.6 before the dike is reinforced. Note that this is just slightly below the target reliability, but the reliability will decrease when the dike is raised. After raising the dike, the reliability estimate drops to β=3.1. From Figure A.3 it follows that a berm of 4.7 m width is required to reach target reliability after full consolidation.

Incorporating the survival information of the survived construction, following the

A

Table A.1: Probability distributions of soil properties used in the case study.

Soil	**Volumetric weight** $\gamma_{sat.}/\gamma_{unsat.}$ [kN m^{-3}]	**Normally consolidated undrained shear strength ratio** S[−]	**Strength increase exponent** m [−]	**Pre-overburden pressure** POP[a] [kPa]	**Critical state friction angle** ϕ_{cs}[°]
Sand	20/18				Log-normal μ 35.0, σ 1.5
Peat	10.5/10.5	Log-normal μ 0.50, σ 0.05	Log-normal μ 0.85, σ 0.05	Log-normal μ 10.0, σ 3.0	
Basal Peat	11/11	Log-normal μ 0.50, σ 0.05	Log-normal μ 0.85, σ 0.05	Log-normal μ 20.0, σ 6.0	
Clay	13/13	Log-normal μ 0.40, σ 0.04	Log-normal μ 0.85, σ 0.05	Log-normal μ 15.0, σ 4.5	
Clay toplayer[b]	15/15	Log-normal μ 0.35, σ 0.04	Log-normal μ 0.85, σ 0.05	Log-normal μ 15.0, σ 4.5	Log-normal μ 32.0, σ 2.0
Dike Core[b]	17/17	Log-normal μ 0.32, σ 0.07	Log-normal μ 0.85, σ 0.05	Log-normal μ 20.0, σ 6.0	Log-normal μ 32.0, σ 2.0

[a] The pre-overburden pressure refers to a the initial situation, before construction of the dike.
[b] Saturated clay is modelled using Stress History and Normalized Soil Engineering Properties (SHANSEP) parameters S, m and POP (or OCR), unsaturated clay is modelled using a critical state friction angle ϕ_{cs}.

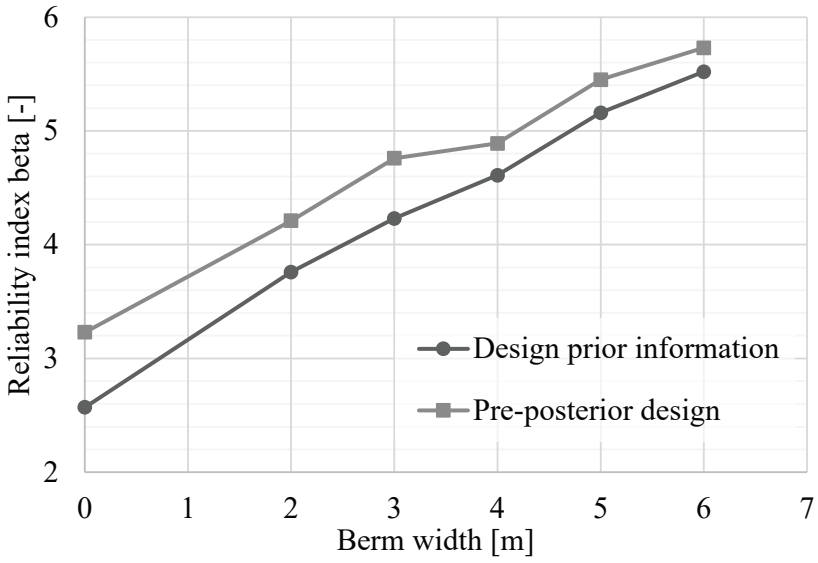


Figure A.3: Decision rules for the berm design as measure to meet the target reliabliity.

approach of Chapter 4, leads to a significant increase of the reliability estimate, depending on the modelling choices. If a situation with 50% consolidation is survived (i.e. 50%

Table A.2: Reliability results posterior analysis. Survived situation with 50% excess pore pressure

autocorrelation model uncertainty	**with survived construction**	**with additional strength in unsaturated zone**	
		only during construction	during construction and flood
uncorrelated $\rho = 0$	4.40	3.89	4.60
correlated $\rho = 1$	3.75	3.44	4.17

excess pore water pressures still present in the subsoil), the failure probability decreases a factor 15–250 depending on the assumption of the autocorrelation of the model uncertainty, see Table A.2.

The effect of additional strength in the unsaturated zone of the soil on the reliability update is investigated by a sensitivity analysis. Unsaturated strength has a positive effect on the stability during construction, making the construction a less critical situation, and hence less informative. Accounting for the additional unsaturated strength lowers the reliability updating effect, but does not eliminates it completely: the posterior failure probability is still a factor 5 lower.

Arguably, if such an additional strength is present during construction, the effect will also be present to some (and likely lesser) extent during flood. If we take the effects of unsaturated soil into account in both cases, the effect of reliability updating will be slightly lower, but de posterior reliability is much higher than in case we do not take it into account. Therefore, neglecting the influence will lead to conservative estimates of the reliability.

The decision whether or not to reverse the construction phasing depends on the one end the savings of the reinforcement, and on the other hand the additional risk of reversing. In Figure A.4 the total expected direct costs of the strategies are shown for different reinforcement costs and different damage costs of failure, to analyse the sensitivity. Equation A.1 was used to calculated the cost. The failure probability during construction with the 'safe' construction phasing (strategy 1) is 0.004, with a more risky construction phasing of strategy 2, the failure probability is 0.11.

To analyse in what situations a more risky construction phasing is typically cost-effective, we consider four typical situations: high and low damage and repair cost, and high and low reinforcement cost. High and low damage and repair costs resemble situations with and without consequential damage to e.g. cables, pipelines, infrastructure, and nearby buildings. High and low reinforcement costs per meter berm per kilometer typically resemble situations where space is scarce or not. Figure A.4 shows each of the four situations, where in each panel the damage/repair cost and reinforcement cost is fixed to the low and high values.

With increasing cost for the dike strengthening, the benefits of a pre-posterior design increase. However, Figure A.4 shows that there is a tipping point where it is not worth to take the additional risk. For example, when the consequential damage is high. The tipping point is also influenced by the sensitivity to different cost functions of the dike reinforcement. For 'cheap' dike reinforcements (e.g. when the construction of a berm is cheap), the tipping point is already at damage costs 0.5 M€ in this case study. For

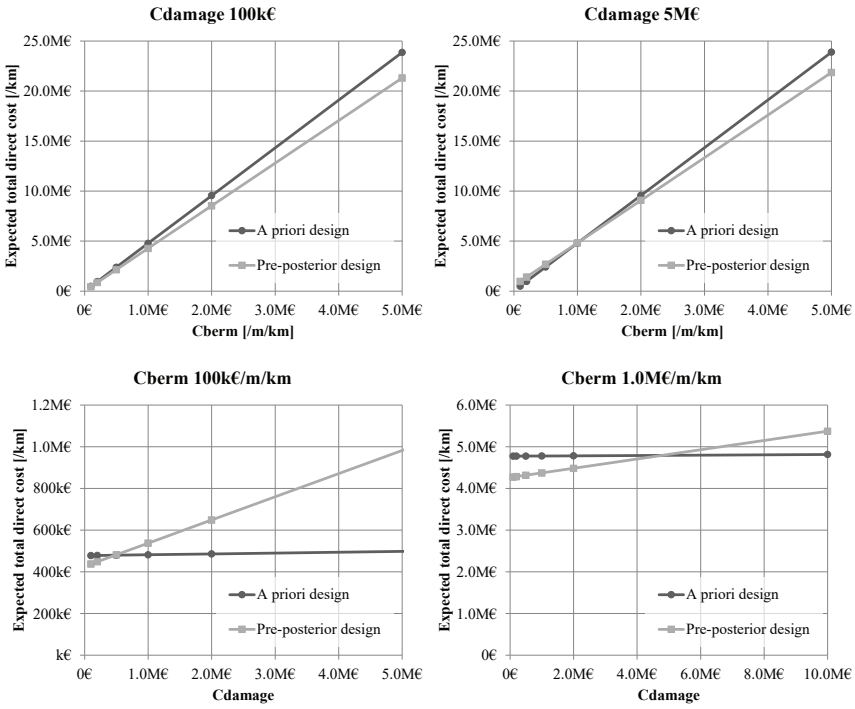


Figure A.4: Sensitivity analysis of total direct costs of a prior and pre-posterior design strategy.

expensive dike reinforcements (e.g. when constructing a berm is expensive because adjacent home owners need to be compensated), the tipping point is at 5.0 M€.

Figure A.5 illustrates another typical example of a situation with a discrete cost function. For example, when a very expensive sheet pile or diaphragm wall (20.0 M€) is required if the stability berm width is larger than for example 4.5 m. In that case, a pre-posterior design is practically always worthwhile.

CONCLUSION

From the sensitivity analysis it follows that when optimizing the total cost, it is worth accepting a lower stability during the construction, when:

- The damage and repair (risk) of a possible instability is not high, relative to the reinforcement costs, for example, with a green dike without cables and pipes and relatively high costs per meter of berm
- There is a jump in the cost function, for example when a hard construction is needed.

The additional risk can also be reduced by mitigation measures such as monitoring etc. A pre-posterior design for dike reinforcements should therefore be applied within the

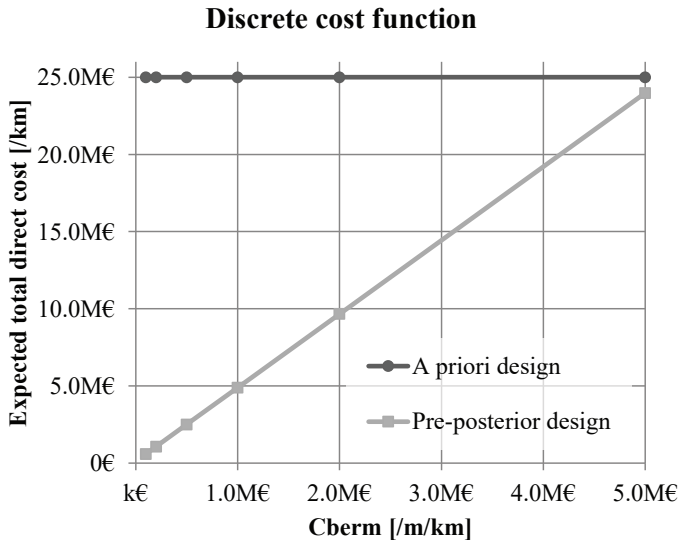


Figure A.5: Expected Total Cost of a prior and pre-posterior design strategy with a discrete cost function.

framework of the Observational Method in Eurocode 7 (CEN, 2004). Another option to take advantage of the survived construction is to extend the lifetime, i.e. to extend the time before a new dike reinforcement has to be carried out. So we can accept a temporarily too high reliability during the lifetime, to be prepared for more extreme climate scenarios with increasing loads.

List of Acronyms

cdf	cumulative density function.
CMC	Crude Monte Carlo.
CoV	coefficient of variation.
CPT	cone penetration test.
CSSM	Critical State Soil Mechanics.
DSS	direct simple shear.
EPM	Equivalent Planes Method.
ERRAGA	Efficient and Robust Reliability Analysis for Geotechnical Applications.
FEM	Finite Element Methods.
FORM	First Order Reliability Method.
HWBP	Hoogwaterbeschermingsprogramma.
LEM	Limit Equilibrium method.
LIR	Lokaal individueel Risico.
MCIS	Monte Carlo Importance Sampling.
NAP	Normaal Amsterdams Peil.
NC	normally consolidated.
OC	over-consolidated.
pdf	probability density function.
ref.	reference level.
SD	standard deviation.
SHANSEP	Stress History and Normalized Soil Engineering Properties.
TC	Total Cost.
TXC	triaxial compression.
VoI	Value of Information.

NOTATION

LATIN

A	set of dike reinforcement actions
a	decision on dike reinforcement design action (berm length)
a_{CPT}	CPT cone factor
a_{p}	response factor of the phreatic level to flood water levels
C	cost components (proof load, monitoring, repair, reinforcement, failure)
c	scaling constant for indirect reliability updating
c_0	total cost of the reference case (without proof loading and monitoring)
c	cost for each step in the decision tree (decision and outcome)
D	annual expected damage flooding
d	decision rule
F	failure event
F_s	factor of safety
h	water level
h_{thresh}	threshold water level from where valuable measurements are obtained
i	numerator for the number of measurements in a CPT, or numerator for realisation
I_m	indicator whether monitoring is done
I_p	indicator whether proof loading is done
j	numerator for CPTs, or numerator for realisation
$\hat{l}$	maximum log-likelihood value
M	number of CPTs
m	strength increase exponent SHANSEP
m	decision whether or not to invest in pore pressure monitoring
m_{s}	model uncertainty factor
N	number of measurements in a CPT
n	total number of CPTs in the calibration of the transformation model, or total number of realisations
N_{kt}	transformation model parameter
$\hat{N}_{\text{kt}}$	estimated (calibrated) transformation model parameter
OCR	over-consolidation ratio
p	decision on whether or not to execute a proof load test of a certain magnitude
P_f	probability of failure
POP	pre-overburden pressure
q_c	CPT cone tip resistance
$\boldsymbol{q}_{\text{net}}$	random field of CPT measurements
$q_{\text{net,i}}$	point measurement cone resistance from CPT
r	annual discount rate

r	ratio between random variability and total point variability
S	normally consolidated undrained shear strength ratio
S	event of survival of a proof load
$\boldsymbol{s}_{\mathrm{u}}$	random field of undrained shear strength (true known, directly measured, or indirectly measured)
$\bar{s}^{\mathrm{D}}_{\mathrm{u,j}}$	depth average undrained shear strength, directly measured using laboratory tests
$\bar{s}^{\mathrm{I}}_{\mathrm{u,j}}$	depth average undrained shear strength, indirectly measured using CPTs
s_{u}	undrained shear strength
$s^{\mathrm{D}}_{\mathrm{u,i}}$	direct point measurement undrained shear strength (laboratory test)
$s^{\mathrm{I}}_{\mathrm{u,i}}$	indirect point measurement undrained shear strength
t	duration of monitoring
u	standard normal stochastic parameter
u_2	CPT pore water pressure measurement
V	coefficient of variation (CoV)
$v_{1,i}$	random uniform number in the i^{th} Monte Carlo realization
$v_{2,i}$	random uniform number in the i^{th} Monte Carlo realization
w_i	importance weight factor for realisation i
$\mathbf{X}$	stochastic variables
$\mathbf{x}$	realization of random variables
$\mathbf{Z}_{\mathrm{m}}$	set of all possible outcomes of pore water pressure monitoring
z_{m}	outcome of pore water pressure monitoring (observation of phreatic level reaction to flood)
$\mathbf{Z}_{\mathrm{p}}$	set of all possible outcomes of a proof load test
z_{p}	observation of survival of a proof load test (at the imposed phreatic level)

GREEK

α_k	FORM influence factor for parameter k
β	reliability index
β_T	optimal target reliability index
$\chi_{\mathrm{F},i}$	indicator value of failure in the i^{th} Monte Carlo realizatio
$\chi_{\varepsilon,i}$	indicator value of evidence in the i^{th} Monte Carlo realizatio
ε	evidence/survived event
$\boldsymbol{\varepsilon}$	random field of error term
$\varepsilon_{q_{\mathrm{net}}}$	measurement error q_{net}
$\varepsilon_{s_{\mathrm{u}}}$	measurement error s_{u}
ε_{t}	transformation error
Γ^2	variance reduction factor
γ	volumetric weight
μ	mean value
π	standard uniform stochastic parameter
Φ^{-1}	inverse standard normal cumulative distribution function
φ_{cs}	critical state friction angle

ρ	correlation coefficient
σ	standard deviation
σ^2	variance
σ'_v	effective vertical stress
σ_p	pore pressure
σ'_p	preconsolidation stress
σ_v	total vertical soil stress
θ	performance (failure/no failure)

MATHEMATICAL OPERATORS

$1[\cdot]$	indicator function
$f(\cdot)$	probability density function
$g(\cdot)$	performance function failure
$h(\cdot)$	performance function observation
$F(\cdot)$	cumulative distribution function
$L(\cdot)$	likelihood function
$P(\cdot)$	probability operator
$\varphi(\cdot)$	standard normal probability density function

List of Figures

List of Tables

List of Publications

Peer-reviewed journals

Van der Krogt, M.G., Schweckendiek, T. and Kok, M. (2021). Improving dike reliability estimates by incorporating construction survival, *Engineering Geology*, 280, 105937.

Van der Krogt, M.G., Klerk, W. J., Kanning, W., Schweckendiek, T. and Kok, M. (2020). Value of Information of Combinations of Proof Loading and Pore Pressure Monitoring for Flood Defences. *Structure and Infrastructure Engineering*, 1–16.

Van der Krogt, M.G., Schweckendiek, T. and Kok, M. (2019). Uncertainty in spatial average undrained shear strength with a site-specific transformation model. *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, 1–11.

Conferences

Van der Krogt, M.G., Schweckendiek, T. and Kok, M. (2019). Do all Dike Instabilities Cause Flooding? Proceedings of ICASP13, 461:1–8. 13th International Conference on Applications of Statistics and Probability in Civil Engineering, Seoul, South Korea.

Van der Krogt, M.G. and Schweckendiek, T. (2019). Systematic Transformation Error in the Depth-Average Undrained Shear Strength. Proceedings of 7th International Symposium on Geotechnical Safety and Risk, Tapei, Taiwan.

Schweckendiek, T., **van der Krogt, M. G.**, Teixeira, A., Kanning, W., Brinkman, R. and Rippi, K. (2017). Reliability updating with survival information for dike slope stability using fragility curves. Geo-Risk 2017, American Society of Civil Engineers, 2017.

Kanning, W., Teixeira, A., **van der Krogt, M.G.**, Rippi, K., Schweckendiek, T. and Hardeman, B. (2017). Calibration of Factors of Safety for Slope Stability of Dikes. Geo-Risk 2017, American Society of Civil Engineers, 2017.

Huber, M., **van der Krogt, M. G.**, Kanning, W. (2016). Probabilistic slope stability analysis using approximative FORM. *14th International Probabilistic Workshop, 299–316. Springer International Publishing*.

van der Krogt, M. G., van den Ham, G. A. and Kok, M. (2015) Safety Assessment Method of Flood Defences for Flow Sliding. Geotechnical Risk and Safety V; 5th International Symposium on Geotechnical Safety and Risk; Rotterdam (The Netherlands), 13-16 Oct. 2015, IOS Press, 522–527.

CURRICULUM VITÆ

Mark VAN DER KROGT

06-05-1991 Born in Zoeterwoude, the Netherlands.

EDUCATION

2017–2021 PhD Candidate
Delft University of Technology, Delft
Dissertation: Reliability updating for slope stability. Improving dike safety assessments using performance information

2012–2015 MSc Civil Engineering, specialization in Hydraulic Structures & Flood Risk
Delft University of Technology, Delft
Thesis: Safety assessment method of flood defences for flow sliding

2009–2012 BSc Civil Engineering
Delft University of Technology, Delft

2003–2009 Secondary School (VWO)
Da Vinci College, Leiden

EXPERIENCE

2015 – present Researcher / Advisor
Deltares, unit Geo-Engineering