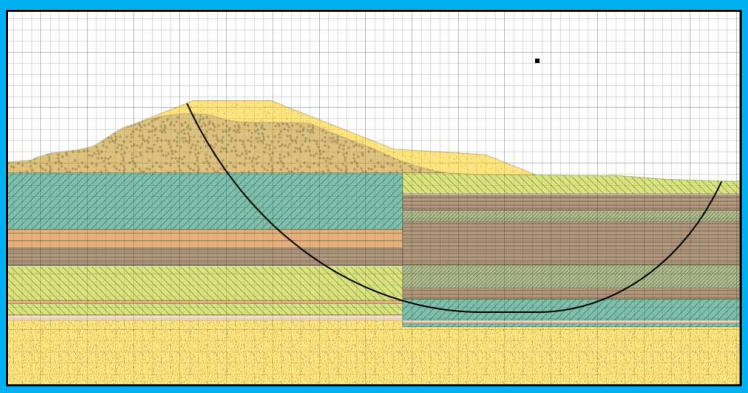
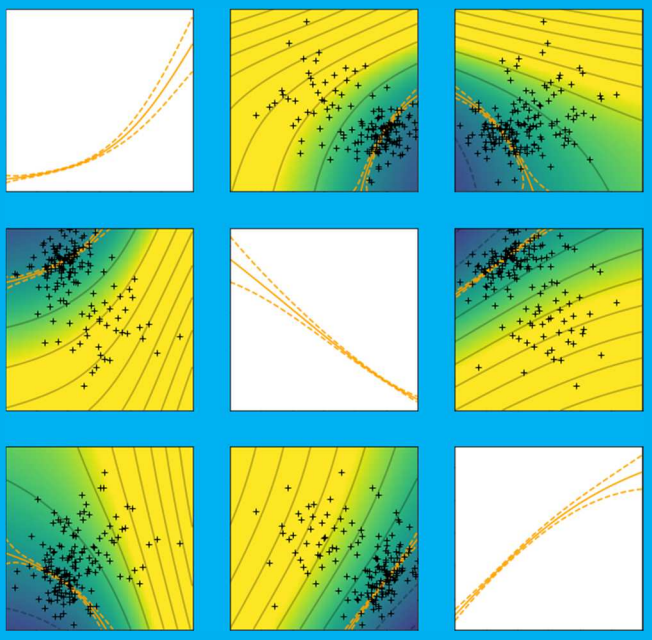
A Metamodelling Approach to Reliability Updating with Dike Construction Survival





Daniël Kentrop

MSc Thesis

Cover page: The conceptual process of metamodelling. From a real situation (Waal dike at Neerijnen, D. Kentrop 2017), to a schematization (D-Stability), to training a metamodel.

A Metamodelling Approach to Reliability Updating with Dike Construction Survival

by

Daniël Kentrop

in partial fulfillment of the requirements for the degree of

Master of Science
in
Civil Engineering

at the University of Technology, Delft
to be defended publicly on Friday 5 February 2021 at 11:00 AM

Student number:	4626125	
Project duration:	May 2020 – January 2021	
Thesis committee:	Prof. dr. ir. M. Kok (chair)	TU Delft
	Dr. ir. A.P. van den Eijnden	TU Delft
	Ir. M.G. van der Krogt (daily supervisor)	Deltares/TU Delft

Preface

Before you lies the final product of my graduation thesis "A Metamodelling Approach to Reliability Updating with Dike Construction Survival". This report is the last step in completing the master track Hydraulic Engineering at the Technical University of Delft. Reliability updating is an interesting field of research, with the potential to improve the cost-effectiveness of flood defenses. This research investigates the effect of incorporating construction survival in the reliability assessment of slopes. The methodology and findings of this research are meant to bring the application of reliability updating in practice a bit closer.

This research would not have been possible without the support of my supervisors. First, I would like to express my sincere gratitude to my daily supervisor, Mark van der Krogt, for his enthusiasm and valuable support. The meetings with Mark always provided me with new ways for approaching the issues I encountered and helped me to see the bigger picture when I got lost in a detail. His knowledge and research skills were a great help in completing this report.

I would also like to thank the members of my graduation committee, Matthijs Kok and Bram van den Eijnden, for their useful feedback and insights throughout the project. Furthermore, I would like to thank Deltares and my Deltares colleagues for the opportunity to complete my master thesis at their inspiring research institute. Finally, I would like to thank my family, friends and roommates for their support. They really made me enjoy my study time.

I hope you will enjoy reading this thesis.

Daniël Kentrop
Delft, January 2021

Summary

The safety assessment of dike slope stability often involves dealing with large uncertainty in soil properties, caused by a lack of knowledge of the exact subsoil conditions. Uncertainty has influence on the reliability estimate of slope stability. Large uncertainty may therefore lead to an overestimation of the failure probability. A way of improving the reliability assessment is by incorporating additional information such as the survival of loading conditions from the past. An example of such an observation is the survival of an extreme water level, but the survival of other loading conditions may be useful as well. A dike heightening often causes a decrease in reliability during construction. Moreover, raising a dike may lead to excess pore water pressures in the subsoil, resulting in a temporal strength decrease, lowering the stability. Observations of critical loading conditions might provide information on the strength of a dike, which can be used to improve reliability assessments. This research investigates the effect of incorporating construction survival in estimating the slope reliability of dikes.

Assessing slope reliability can be done by means of simulation. However, due to the high reliability of dikes, many realizations and thus model evaluations may be needed to achieve sufficient accuracy, leading to infeasible computation times. A possible way to reduce the calculation time is the use of an approximation model, called metamodelling. A metamodel is constructed with a concise set of model evaluations and is used for predicting the model outcome, forming a replacement of the actual model (e.g. slope stability software). In this research a method is developed for reliability updating using the metamodelling package ERRAGA (Van den Eijnden et al., 2021). The developed method involves training two metamodels, one for predicting the slope stability in the observed situation and one for the assessment situation. The reliability update is done with simulation using importance sampling. All samples are evaluated in both metamodels, after which the posterior reliability is calculated by only considering realizations that lead to survival of the observed situation.

The effect of incorporating construction survival is explored by updating the inward slope reliability of a case study dike improvement with various observations. The chosen case is Kinderdijk-Schoonhovenseveer, a Dutch dike trajectory along the river Lek. From this trajectory a dike section is chosen in which the crest was raised, and a berm was constructed on the land side. The observations comprise the survival of the original dike improvement and the hypothetical survival of alternative construction sequences (first constructing the berm and then heightening the dike, and vice versa). It was found that the reversed construction phasing leads to more effect for reliability updating then the original situation: updating the reliability index from 6.0 to 8.5. The original construction

phases were not severe enough for the construction survival to yield an effect for reliability updating.

The effectiveness of a reliability update depends on the criticality of the observation and the similarity of loading conditions in the observed and assessment situation. It was found that the updating effect is optimal when the two governing limit state functions are parallel, and the reliability in the survived situation is smaller. Then the entire failure domain of the assessment situation falls within the failure domain of the observed situation, which is the implausible domain because survival was observed. For situations with the same set of stochastic parameters with only reducible uncertainty (all parameters have full auto-correlation in time), the failure probability reduces to zero. Cases with (near) parallel limit state functions are therefore effective for updating. Further insight is obtained by comparing the design points and the importance factors of the (strength) parameters in both situations, determined with FORM analyses. It was found that limit state functions are approximately parallel when the importance factors and slip planes are similar, thus when loading conditions are alike. Cases with these characteristics are therefore good candidates for reliability updating, although parallelism is not a strict requirement. Other situations can be effective as well, though cases with limit state functions that are less alike need a more severe survived loading condition to achieve the same effect.

The case study shows that the current semi-probabilistic approach to assessing construction stability in practice leads to situations not effective for reliability updating. Though adjusting the construction phasing may lead to a significant update. The potential impact of including a reliability update in future dike improvement designs is examined by optimizing the remaining construction activities after observing the survival of a critical construction step. The design was altered such that the updated reliability is equal to the prior reliability of the original design. In the alternative construction phasing of the case study, the critical construction step used for updating took place just before construction of the berm. This allows for reduction of the berm dimensions after observing construction survival. It was found that the berm length could be shortened from 10 to 5 meter and the berm height from approximately 2 to 1 meter. A limiting factor for this approach is the flooding probability during construction, which should still comply with the statutory requirements. However, this limitation does not seem crucial, as a slope failure does not directly cause a flooding. So, besides improving the reliability, this method also has potential to reduce construction costs, limit additional space occupation and reduce other negative impacts of dike reinforcements. This makes designing the construction phasing with reliability updating in mind promising for future dike designs.

The main finding of this research is that incorporating construction survival in the reliability assessment can be very effective in terms of reliability, and in making dike improvements more efficient and less space consuming. A simplified approach is proposed for investigating the potential of optimizing a dike improvement with reliability updating, before going into the endeavor of Bayesian analysis. The developed method using metamodelling decreases the computational time from at least days to several hours compared to advanced sampling methods while not losing accuracy, which makes the practical application of Bayesian updating within reach.

Contents

List of symbols

Latin symbols

CoV	coefficient of variation (σ/μ), possibly with subscript indicating the random variable
$f_{\mathbf{X}}(\mathbf{x})$	probability density function of $\mathbf{X}$
$g(\cdot)$	limit state function
$h(\cdot)$	observation function
h	water level
$I[\cdot]$	indicator function, returns 1 if argument is true, 0 otherwise
$k(\mathrm{x}, \mathrm{x}')$	covariance (kernel) function, gives covariance between random variables at x and x'
$\mathbf{K}(\mathbf{x})$	covariance matrix of $\mathbf{x}$
$\mathrm{L}(\cdot)$	learning function
$m(\mathbf{x})$	mean vector of $\mathbf{x}$
m_d	slip plane model factor
m	strength increase exponent
M	metamodel
n	sample size
$(\cdot)$	normal distribution
p	pore pressure
$P(\cdot)$	occurrence probability of argument
$\hat{P}$	estimate of probability
P_f	probability of failure
$q_{\mathbf{X}}(\mathbf{x})$	importance sampling distribution for random variable $\mathbf{X}$
S	shear strength ratio
S_u	undrained shear strength
$\mathbf{u}_{\mathrm{IS}}$	group of importance samples in the standard normal space
$w(\mathbf{x})$	importance sampling weight of realizations $\mathbf{x}$
$\mathbf{X}$	vector of random variables
$\mathbf{x}$	realization of $\mathbf{X}$
$y(x)$	metamodel, approximates model response as a function of model parameter x
$\mathbf{y_t}$	$y(\mathbf{x_t})$ data set for training the metamodel
$\mathbf{y_p}$	$y(\mathbf{x_p})$ predicted model response

Greek symbols

β	reliability index
ε	evidence or observed event

$\sigma_{\mathbf{X}}$	standard deviation of $\mathbf{X}$
σ	total stress
σ'	effective stress
σ'_v	vertical effective stress
σ'_{vy}	yield stress
$\Phi^{-1}(\cdot)$	inverse cumulative density function of standard normal distribution
ω	contribution factor of a failure mechanism to the total failure probability

Abbreviations

AK-MCS	Active learning Kriging based Monte Carlo Simulation
DoE	Design of Experiment
FORM	First Order Reliability Method
GEV	Generalized Extreme Value distribution
GMM	Gaussian Mixture Model
HL	Head Line
IS	Importance Sampling
KIS	Kinderdijk-Schoonhovenseveer
LEM	Limit Equilibrium Method
MCS	Monte Carlo Simulation
NAP	Normaal Amsterdams Peil, the Dutch national reference height
OCR	Over-consolidation Ratio
PDF	Probability Density Function
PL	Phreatic Line
POP	Pre-overburden Pressure
SAS	Salmsteke-Schoonhoven
SF	Safety Factor
SHANSEP	Stress History and Normalized Soil Engineering Properties
SHM	Structural Health Monitoring
WBI	Wettelijk Beoordelingsinstrumentarium (Dutch statutory assessment manual)

1

Introduction

1.1 Background and motivation

Flood defenses are an essential part of infrastructure for low-lying coastal areas such as The Netherlands. For societal and economic benefits, it is crucial to minimize the risk of failure to an acceptable level. Dutch waterboards are therefore obligated to assess the safety of their system of flood defenses every 12 years. This safety assessment is to make sure that the failure probability of flood defenses meets the statutory safety requirement. The probability of failure is determined by assessing all possible failure mechanisms, of which slope stability is an important one for earthen dikes. Large uncertainty in soil properties is a typical issue encountered in the assessment of such geotechnical structures (e.g. Han et al, 2011). This uncertainty can be reduced by site investigation and lab tests, but spatial and temporal variability and the availability of resources put restraints to the reduction extent. The lack of knowledge on the exact subsoil conditions and thus soil strength affects the reliability assessment of slope stability. Large uncertainty may lead to an overestimation of the failure probability. This provokes an ongoing search for sophisticated probabilistic methods to improve the reliability estimate. A way of improving the reliability assessment is by incorporating additional information such as the survival of loading conditions from the past.

The concept of updating with past performance is meant to improve the beliefs about uncertain parameters with additional information. For instance, by improving settlement predictions (Kelly & Huang, 2015; Van der Meijs, 2015) or sheet pile deformation predictions (Papaioannou & Straub, 2012) with deformation measurements. Another example is the incorporation of a survived loading condition to improve the reliability estimate. An intuitive example for flood defenses is the survival of an observed water level, as used by Schweckendiek (2014) for updating the reliability regarding the failure mechanism piping. A less intuitive example is updating with survival of the construction of a dike, as investigated by van der Krogt et al. (2021). However, construction or improvement of a dike can cause severe loading conditions, as becomes apparent from cases in which slope instability occurred after improvement, like in the Dutch town of Streefkerk in 1984 (TAW, 1985). Such cases give reason to believe that observing dike

improvement survival could be useful for reliability updating, perhaps improving the safety assessment of dikes.

1.2 Problem definition

A way of incorporating additional information in a reliability analysis is by using Bayes theorem for conditional probabilities, known as Bayesian reliability updating (e.g. Schweckendiek, 2014; Straub & Papaioannou, 2014). Bayesian updating results in a probability conditionalized on the information. In case of slope reliability updating with past performance this would be the probability of failure given the survival of a loading condition. Determining this conditional probability can be done by means of integration over the joint probability density function. This may be feasible for simple reliability problems involving few parameters but gets difficult or impossible for more complex problems. Therefore, preceding studies have proposed alternative methods for Bayesian reliability updating. Examples are the use of fragility curves with the First Order Reliability Method (FORM) (Schweckendiek & Kanning, 2016) and approaches using the Equivalent Planes Method (Laghmouchi, 2021; Roscoe et al., 2015; van der Krogt et al., 2021). These approaches are computationally inexpensive but involve approximations, which may pose limitations for complex reliability problems with strongly nonlinear limit state functions.

A class of approaches that doesn't involve simplification of the reliability problem is simulation, like Monte Carlo Simulation (MCS). MCS involves drawing parameter realizations from the joint probability density function and evaluating the model performance (e.g. slope stability calculation). The updated reliability is obtained by only considering realizations that are consistent with the observation. However, for problems with a high reliability, many realizations may be needed to achieve sufficient accuracy of the outcome. There are methods available for enhancing the simulation efficiency, like subset simulation (Au & Beck, 2001; Straub & Papaioannou, 2014) and importance sampling (Melchers, 1989). But when failure probabilities are small and the model evaluation is computationally expensive, these methods may lead to infeasible computation times, as is the case for slope stability. Therefore, the urge arises to reduce the amount of model evaluation to a minimum. A possible way to reduce the simulation time is by the use of an approximation model, called metamodelling.

A metamodel is constructed with a concise set of model evaluations and is used for predicting the model outcome (e.g. slope stability). It can be used for replacing actual model evaluations in a simulation, thus reducing the computational time. This is beneficial in the above described case, where (repetitive) evaluation of the model is computationally expensive. Metamodelling has had various applications in slope stability analysis (Li et al., 2016), and in structural engineering for updating the reliability of bridges (Sousa et al., 2019) and sheet pile walls (Chai, 2019). A metamodelling method

called Active learning Kriging based Monte Carlo Simulation (AK-MCS) has shown to be an efficient way of approximating failure probabilities (Echard et al., 2011). It contains an iterative process in which the training set of the metamodel is expanded with extra model evaluations until a certain convergence criterion of the outcome is met. An advantage of Kriging is that besides estimating the model response, also the uncertainty of the estimate is given. This allows for assessing the accuracy of the analysis outcome. Recently, Van den Eijnden et al., (2021) took the concept further and developed a two stage AK-MCS metamodel meant for modelling noisy and incomplete models (see section 2.6). The two-stage approach allows the model to be used in the reliability updating context, in which one of the stages predicts consistency with the observation. This way the failure probability conditional on the observation can be determined. The metamodel developed by Van den Eijnden et al., (2021) may thus be applicable to reliability updating of slope stability, which is the topic of this research.

1.3 Research objective and questions

Reliability updating with dike improvement construction survival seems promising. However, a limitation for problems with a highly reliability and expensive computational models turns out to be long computation times. Therefore, the objective of this research is to update dike reliability with construction survival using a metamodel, making computation times feasible for appliance in practice, while not losing accuracy. Considering this objective, the main research question is formulated along with three sub-questions:

What is the effect of reliability updating with construction survival of a dike improvement?

- What are the characteristics of a dike improvement leading to a significant reliability update with survival of the construction?

- What is the effect of reliability updating with construction survival compared to reliability updating with an observed water level?

- How can metamodelling be used for reliability updating with past performance regarding slope stability of dikes?

1.4 Thesis overview

This thesis is structured as follows. First, in chapter 2, some essential theory regarding reliability analysis and slope stability is discussed. Furthermore, it discusses reliability updating, metamodelling and introduces the metamodel by Van den Eijnden et al., (2021). Chapter 3 proposes a method for reliability updating using metamodelling and discusses the design choices of this method. Next, the case study dike improvement Kinderdijk-Schoonhovenseveer is introduced in chapter 4, along with the prior reliability analysis of slope stability. Then, in chapter 5, the effect of reliability updating with construction survival is investigated. This effect is compared to updates with survival of the daily conditions and survival of an extreme water level. Also, the effect of incorporating a reliability update into a dike design is explored. Chapter 5 is finalized with a discussion of all the results. In chapter 6 conclusions are drawn from the case study results. It presents why certain observations are effective for reliability updating and how this can be recognized. These findings are used for answering the research questions. Hereafter, the most important points of discussion of the developed methodology are presented. Finally, several recommendations are done regarding the improvement of the methodology and for further research. Also, a simplified approach is proposed for investigating the potential of optimizing a dike improvement with reliability updating

2

Background

2.1 Flood safety

The Dutch safety requirements for flood defenses are based on optimization of economic, societal and individual risks. These risks were determined in a nationwide flood risk analysis, in which for each dike trajectory the consequences of several flood scenarios were considered (VNK2, 2014). The statutory upper limits of the flooding probabilities were derived using predetermined acceptable levels per risk category, resulting in a risk bases trajectory specific target probability.

2.1.1 Safety assessment

The failure probability of a dike trajectory consists of contributions of various failure mechanisms. A combination of these probabilities results in the failure probability of the dike trajectory. The assessment of a single mechanism is usually done by dividing the dike trajectory in sections based on similarity in properties. In each section a representative two-dimensional cross-section is chosen and used for assessment of the failure mechanism. The resulting probability on cross-sectional level then needs to be scaled to sectional level, which is be done by incorporating the length-effect (WBI, 2017). Assembly of the failure probability of all sections results in the failure probability of a single mechanism on trajectory level. Combining the failure probability of all mechanisms results in the final failure probability of the dike trajectory. This method is described in the Dutch statutory assessment manual (WBI, 2017).

2.2 Slope stability

Slope stability is an important failure mechanism for dikes. A slope instability can be described as the loss of equilibrium of (a part of) the dike slope. Its occurrence can lead to the loss of the water retaining function of a dike, therefore the probability of occurrence is assessed periodically. The stability of a slope is often expressed with the safety factor (SF), which is the resisting force (strength) divided by the driving force (load) of a slip surface. A safety factor smaller than one therefore indicates failure of the slope. Slope stability is usually assessed by considering a two-dimensional cross-section

of a dike. Software can be used to compute the safety factor. The occurrence of a slope instability is schematized in figure 2-1.

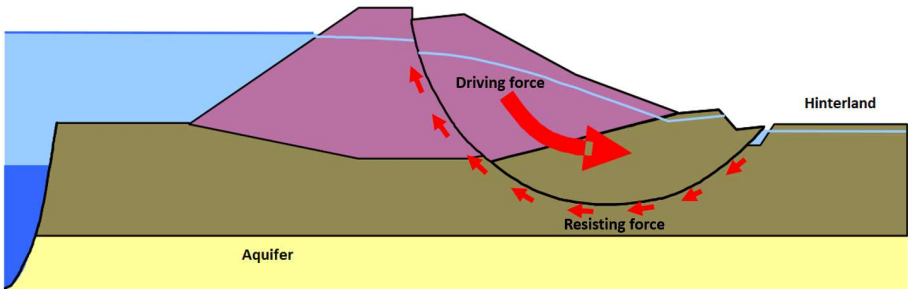


Figure 2-1: Illustration of slope instability, adapted from Deltares (2018)

2.2.1 Shear strength

The shear strength of soil is an important parameter in the slope stability assessment of dikes. The strength of soil in a slope stability calculation is determined using shear strength models. Soil types are usually divided in two categories for which different shear strength models are used, soils with drained or undrained behavior. Applying a load to a normally consolidated soil causes deformation by a decrease of pore volume. If the pores are saturated, then water needs to dissipate for the soil to deform. A soil is considered to behave drained if the dissipation process is quick relative to the load increase. This is the case for sands. The shear strength of drained soils is often modelled with the Mohr-Coulomb strength model, as is the case for the safety assessment in the Netherlands (WBI, 2019). Further details on the Mohr-Coulomb model are found in chapter 20 of (Verruijt, 2001).

Soils are considered to behave undrained if the water dissipation process is slow compared to the load increase. Applying a load can cause excess pore water pressure. For soils with such behavior, the SHANSEP (Stress History and Normalized Soil Engineering Properties) model is used. The shear strength using SHANSEP is presented in equation (1), in which S_u is the undrained shear strength, σ'_v is the in-situ vertical effective stress, S is the shear strength ratio, OCR is the overconsolidation ratio and m is the strength increase exponent. S and m are immutable soil properties, the OCR however, describes the state of a soil and depends on the stress history. The OCR is related to the pre-overburden pressure (POP) as in equation (2), with σ'_{vy} the maximum experienced effective vertical stress.

$$S_u = \sigma'_v \cdot S \cdot OCR^m \tag{1}$$

$$OCR = \frac{\sigma'_{vy}}{\sigma'_v} = \frac{\sigma'_v + POP}{\sigma'_v} \tag{2}$$

Important factors for slope stability are the hydraulic boundary conditions and the associated pore water pressures. An increase in pore water pressure causes a decrease in effective stress and thus also in shear strength. The vertical effective stress σ_v' is defined in equation (3), in which σ_v is the total vertical stress and p the pore water pressure.

$$\sigma_v' = \sigma_v - p \tag{3}$$

This explains a decrease in stability during a period of high water. A rise in water level may cause a rise in phreatic level and thus also in the weight of the soil body, which is the driving force of this mechanism. Simultaneously, the resisting force is decreased because of the decrease in effective stress. The occurrence of a high water level can therefore cause a critical loading condition in terms of inward slope stability. Another situation causing an increase in the pore water pressure is heightening a dike that is built on a soft soil foundation with undrained behavior. The load increase causes excess pore water pressures and thus a decrease in strength. The loading condition during improvement of a dike therefore has similarities with the condition during high water. The stability is often a limiting factor for the construction speed of a dike improvement. The construction phasing is determined such that between the construction steps the excess pore water pressure has time to dissipate (partially).

2.3 Reliability analysis

Failure is defined as the occurrence of an unwanted event, such as the loss of water retaining function of a dike. Whether failure occurs can be evaluated with a limit state function, here written as $g(\mathbf{x})$, with $\mathbf{x}$ the realization of $\mathbf{X}$, which is the vector of all involved random variables. The limit state function is defined such that failure occurs when $g(\mathbf{x}) < 0$. When the limit state function and joint probability density function $f_{\mathbf{X}}(\mathbf{x})$ are known, then the failure probability can be determined by evaluating the integral in equation (4). A common way to express the reliability is through the reliability index β as expressed in equation (5), in which Φ^{-1} is the inverse cumulative standard normal distribution function.

$$P_f = \int_{g(\mathbf{X})<0} f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \tag{4}$$

$$\beta = -\Phi^{-1}(P_f) \tag{5}$$

For a simple limit state function with few random variables, analytical or numerical evaluation of this integral might be possible but is difficult or impossible for a large set of dependent variables. Additionally, slope stability problems have no analytical expression for the limit state function (at least not without some effort). The use of other techniques is therefore better suited for this class of problems.

2.3.1 Monte Carlo Simulation

Monte Carlo simulation (MCS) is a sampling technique in which the integral of equation (4) is approximated. Random samples are drawn from the join distribution function and are evaluated in the limit state function. The estimate of the probability of failure $\hat{P}_f$ is presented in equation (6), in which the sample size is n and $I[\cdot]$ is the indicator function, returning 1 if the argument is true and 0 otherwise.

$$\hat{P}_f = \frac{1}{n}\sum_{i=1}^{n} I[g(\mathbf{x_i}) < 0] \tag{6}$$

The uncertainty of the estimate is expressed through the coefficient of variation. It can be determined when realizing that the number of failures in an MCS is binomial distributed and that its variance is $n \cdot P_f(1 - P_f)$. After some math the coefficient of variation can be written as in equation (7) (Jonkman et al., 2017):

$$CoV_{P_f} = \sqrt{\frac{1 - P_f}{n \cdot P_f}} \tag{7}$$

The probability of failure of a flood defense is typically very low. For a probability of failure $P_f = 10^{-5}$ and a required CoV of 0.05 already $4 \cdot 10^7$ model evaluations are needed. This clearly demonstrates the limitation of applying MCS to low probability problems like flood defenses. Various methods are developed to enhance the efficiency of MCS, like Importance Sampling (Melchers, 1989) and Subset Simulation (Au & Beck, 2001).

2.3.2 Importance sampling

Importance sampling is a technique of simulation in which samples are generated from a different distribution than the original one, called the importance distribution. This is a useful technique when the region of interest is far away from the mean, for instance the failure domain in a reliability analysis. The efficiency of the simulation can be enhanced by shifting some of the density towards this region such that failure is more frequently observed. Though, the importance distribution must be chosen carefully, as a poor choice may result in an erroneous result. The probability of failure using importance sampling is determined with equation (8) and its variance with equation (9). The symbols are defined as in the previous paragraph, with additionally the importance sampling weight w, defined in equation (10). It corrects the for the difference in density between the original distribution $f_{\mathbf{X}}$ and the importance distribution $q_{\mathbf{X}}$.

$$\hat{P}_f = \frac{1}{n}\sum_{i=1}^{n} I[g(\mathbf{x}_i) < 0] \cdot w(\mathbf{x}_i) \tag{8}$$

$$\sigma^2_{\hat{P}_f} = \frac{1}{n}\left(\frac{1}{n}\sum_{i=1}^{n} w(\mathbf{x}_i)^2 \;\; I[g(\mathbf{x}_i) < 0] - \hat{P}_f^2\right) \tag{9}$$

$$w(\mathbf{x}) = \frac{f_{\mathbf{X}}(\mathbf{x})}{q_{\mathbf{X}}(\mathbf{x})} \tag{10}$$

2.4 Reliability updating

The reliability obtained from a reliability assessment is an estimate. Its quality depends on the assessment method and the knowledge about the problem under consideration. A typical issue for slope reliability analysis is the lack of knowledge on the exact subsoil conditions. The subsoil conditions are determined through site investigation and lab tests, but uncertainty always plays a role due to the spatial variability of soil and the availability of resources. The lack of knowledge on the subsoil conditions and thus soil strength affects the reliability assessment of slope stability. Large uncertainties herein may lead to an overestimation of the failure probability. Including additional information like the survival of a critical loading condition may improve the reliability estimate. This section goes into inclusion of additional information with Bayesian updating, based on the method used by Schweckendiek, (2014).

2.4.1 Bayes' Theorem

Bayesian reliability updating is based on Bayes' Theorem for calculating conditional probabilities, presented in equation (11). With this theorem the occurrence probability of some event E conditionalized on the occurrence of another event $\boldsymbol{\varepsilon}$ can be calculated. In the case of reliability with past performance, the event E is the occurrence of failure and $\boldsymbol{\varepsilon}$ is the observation of past performance, also called the evidence. Calculation of the conditional probability involves evaluation of two reliability problems, as is seen from the right part of equation (11). The joint probability of seeing event E together with the evidence $\boldsymbol{\varepsilon}$, and the occurrence probability $\boldsymbol{\varepsilon}$ itself.

$$P(E|\varepsilon) = \frac{P(\varepsilon|E)P(E)}{P(\varepsilon)} = \frac{P(E \cap \varepsilon)}{P(\varepsilon)} \tag{11}$$

2.4.2 Type of information

The implementation of equation (11) in the reliability analysis depends on what information is contained by the evidence $\boldsymbol{\varepsilon}$. Generally, the information in the evidence can be of two types, equality or inequality information. A measurement of a system

characteristic or performance is of the equality type. For example, measured settlements in a settlement prediction analysis. So, the performance of the model (e.g. settlement), which is a function of the involved set of random variables $\mathbf{x}$, should be equal to the measured value (or close, when including measurement uncertainty). This is expressed in equation (12) using the observation function $h(\mathbf{x})$. The evidence is defined as the observation function being zero.

$$\varepsilon = \{h(\mathbf{x}) = 0\} \tag{12}$$

For inequality information no exact model quantity is known. It is only known that some model quantity is larger or smaller to a certain value. This is the case for a dike that survived a loading condition. This observation does not provide a safety factor, survival only shows that the limit state function was not exceeded under the observed conditions. Hence the stability factor must have been equal to or larger than one. The evidence with inequality information is expressed in equation (13). The observation function is defined such that it results in zero or positive values for parameter combinations leading to the observation being true. This definition deviates from the one defined by Schweckendiek (2014) (which is $\varepsilon = h(\mathbf{x}) < 0$), but is a more convenient definition for considering dike slope survival information. It leads to similar definitions of the observation function and limit state function, presented in equation (24) and (25).

$$\varepsilon = \{h(\mathbf{x}) \geq 0\} \tag{13}$$

In this research only inequality observation is considered. Substitution of equation (13) in equation (11) and taking slope failure as the event under consideration results in the failure probability conditional on the evidence with inequality information:

$$P(F|\varepsilon) = \frac{P(\{g(\mathbf{X}) < 0\} \cap \{h(\mathbf{X}) \geq 0\})}{P(h(\mathbf{X}) \geq 0)} \tag{14}$$

Equation (14) can be interpreted as only taking failure into account over the domain in the parameter space which results in consistency with the observation, so $h(\mathbf{x}) \geq 0$ must be true.

2.4.3 Type of uncertainty

The purpose of reliability updating by including additional information is reducing uncertainty, but not all uncertainty can be reduced. Generally, there are two types of uncertainty: reducible (epistemic) uncertainty and unreducible (aleatory) uncertainty. When a random variable has only epistemic uncertainty, then a true value exists for this variable, only it is unknow. Adding information can theoretically reduce all uncertainty. Another way to see this is that the random variable is fully autocorrelated in time,

learning the value in the past results in knowing the value in the future. An example is the shear strength ratio of soil. For variables with only aleatory uncertainty, a random process over time is involved. Information on this variable from the past does not say anything about the variable in the future, there is no (or limited) correlation in time. An example is the water level in a river. Though its distribution can be inferred from data, the uncertainty cannot be reduced by knowing the water level at a point in time. A random variable can also have a mix of aleatory and epistemic uncertainty. Then the variable has some autocorrelation in time, due to the epistemic uncertainty. Accounting for these different types of uncertainty in the reliability analysis is done by modelling two correlated sets of random variables, one for the observation and one for the assessment. A method for computing the autocorrelation is given by Roscoe et al. (2015). The type of uncertainty must be carefully considered when updating the reliability or parameter distributions. Reducing aleatory uncertainty could lead to an overestimation of the updated reliability.

2.4.4 Implementation

The updated probability presented in equation (14) consists of two reliability problems. Both can be determined using any reliability method, but an approach using importance sampling is chosen here as discussed in 2.3. Applying equation (8) for calculating both probabilities in equation (14) results in the following equation, in which the sample size n drops out.

$$\hat{P}(F|\varepsilon) = \frac{\sum_{i=1}^{n} I[g(\mathbf{x_i}) < 0 \ \cap \ h(\mathbf{x_i}) \geq 0] \cdot w(\mathbf{x_i})}{\sum_{i=1}^{n} I[h(\mathbf{x}_i) \geq 0] \cdot w(\mathbf{x_i})} \tag{15}$$

Note that equation (15) uses the same set of random variables in both limit state functions. This assumes that there is only reducible uncertainty present. For problems with also unreducible uncertainty, two correlated sets of random variables should be used, as explained in the previous paragraph.

2.4.5 Past applications

A lot of research has been done in the field of reliability updating. Some focusing on the methodology and others on finding new useful applications. A selection of research is discussed here.

Applications in flood defenses

A useful application of reliability updating in the Dutch flood control context was found by Schweckendiek (2014). The failure mechanism piping is one of the main contributors to the failure probability of Dutch river dikes, mainly because of the uncertainty in subsoil conditions. It was found that reliability updating with field observations has a

significant effect, while also pointing out the most probable cause of the change by updating the parameter distributions. Another practical example is the appliance of Bayesian updating to settlement predictions by Van der Meijs (2015). During a dike improvement, settlements and pore water pressures are monitored to ensure stability, but also for predicting long term settlements. These measurements were used for updating the beliefs on long term settlement.

Schweckendiek & Kanning (2016) developed a practical reliability updating framework for updating slope reliability with past performance. They used FORM to construct so called fragility curves. Fragility curves give the failure probability conditional on a dominant load variable and are constructed by performing reliability analyses (e.g. FORM) at discrete points of that load variable. A fragility curve can be seen as the cumulative density function of the overall strength, Schweckendiek & Kanning (2016) therefore used it to sample from in an MCS. The updated reliability is then obtained by filtering out the realizations that are inconsistent with the observation. The case study report accompanying the research (Schweckendiek et al., 2016) showed that in one of their two cases the failure probability decreased orders of magnitude due to updating, showing the potential in practice.

Structural Health Monitoring

Reliability updating is also applied in structural engineering, but mainly with measurements instead of survival information. A lot of infrastructure was built after the second world war and is now approaching the end of the design lifetime. Collective replacement is expensive so a good estimation of the state of infrastructural works could mean an extension of their lifetime. A management strategy called Structural Health Monitoring (SHM) uses deformation and settlement measurements to update the reliability of structures. Sousa et al., (2019) established a SHM approach for updating the parameter distribution with damage scenarios and demonstrated it in a case study. Updating the parameters was done by means of sampling and with a finite element method combined with a metamodel. Another SHM example is the work of Chai (2019), who used strain measurements to update the reliability of sheet pile walls. A finite element model was combined with a metamodel, as structural problems typically have a low failure probability and computations are expensive.

2.5 Metamodelling

A metamodel, or surrogate model, is a function $y(x)$ that approximates an unknown function describing the response of a system. It is based on a set of points in the parameter space where the response is known. A regression method is used to approximate the system response at points where it is unknown. This comes in handy when the system response is obtained by an expensive computational model. Perhaps the simplest interpolation methods thinkable are linear interpolation and linear

regression for a one-parameter system, presented in figure 2-2 for illustrative purposes. The blue line is the metamodel and approximates the system response. Of course, more sophisticated methods are needed for multi-parameter systems with a non-linear model response like slope stability problems. A method capable of grasping the response of such systems is Kriging, explained in the following paragraph.

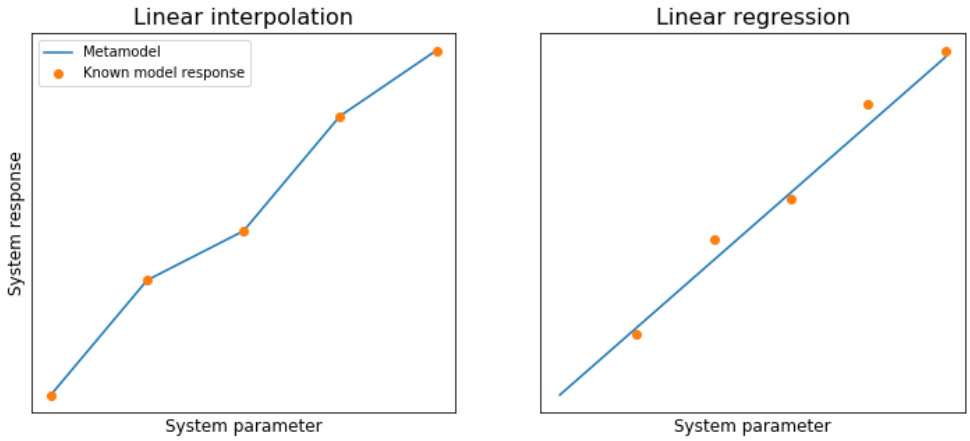


Figure 2-2: Illustration of simple metamodels for a one-parameter system

2.5.1 Kriging

Kriging is an interpolation method in which the approximate system response function $y(x)$ is modelled as a Gaussian process. This means that $y(x)$ is normally distributed at a certain x and that any finite collection of model responses has a multivariate normal distribution. The big advantage of Kriging is that it estimates the model response but also shows the uncertainty. This is a convenient property because it allows for optimal expansion of the training set in an iterative learning process. This paragraph goes into the mathematics of simple Kriging. It assumes that the a-priori global mean is zero, which is different from ordinary Kriging where the global mean is determined. In all the presented equations a one-parameter model is considered with x the model parameter, $y(x_i)$ or y_i the model response and $\mathbf{x}$ a collection of model parameters with the corresponding model response vector $\mathbf{y}$. A collection of model responses has a multivariate normal distribution as presented below:

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \mathbf{y} \sim \quad (\mathbf{0}, \mathbf{K}(\mathbf{x})) \tag{16}$$

With

$$\mathbf{K}(\mathbf{x}) = \begin{pmatrix} k(x_1, x_1) & \dots & k(x_1, x_n) \\ \vdots & \ddots & \vdots \\ k(x_n, x_1) & \cdots & k(x_n, x_n) \end{pmatrix} \tag{17}$$

In which **K(x)** is the covariance matrix. The dependence structure of the joint probability density function enclosed in the covariance matrix is expressed by the covariance function, or kernel function $k(x, x')$. It gives the covariance of two random variables separated a distance $x - x'$. Various kernel functions are available in literature, a choice must be made on which fits best. The vector $\mathbf{y}$ can be partitioned in two parts, a known part $\mathbf{y_t}$ ($y(\mathbf{x_t})$) used for training the metamodel, and a part to be predicted $\mathbf{y_p}$. Following Murphy (2012), the partitioning is written as:

$$\begin{bmatrix} \mathbf{y}_p \\ \mathbf{y}_t \end{bmatrix} \sim \left(\mathbf{0}, \begin{bmatrix} \mathbf{K}_{pp} & \mathbf{K}_{pt} \\ \mathbf{K}_{pt}^T & \mathbf{K}_{tt} \end{bmatrix} \right) \tag{18}$$

This partitioning allows solving for $\mathbf{y_p}$. The mean vector and covariance matrix of $\mathbf{y_p}$ are given by equation (19) and (20), using the proof of Murphy (2012)[1] for conditionalizing Gaussian distributions:

$$m(\mathbf{y}_p \mid \mathbf{y}_t) = \mathbf{K}_{pt}\mathbf{K}_{tt}^{-1}\mathbf{y}_t \tag{19}$$

$$\mathbf{K}(\mathbf{x}_p \mid \mathbf{x}_t) = \mathbf{K}_{pp} - \mathbf{K}_{pt}\mathbf{K}_{tt}^{-1}\mathbf{K}_{pt}^T \tag{20}$$

Taking the diagonal of the covariance matrix $\mathbf{K(x_p|x_t)}$ gives the variance for each point in $\mathbf{y_p}$ (given $\mathbf{y_t}$). So, besides a prediction of the model response, also the uncertainty in the prediction is known. An example of Kriging is shown in figure 2-3, in which the solid line is the estimated model response, the red dots the known data, the gray area is the 95% confidence interval and the dashed lines are some realizations. It shows that the uncertainty increases as the distance from the known points increases.

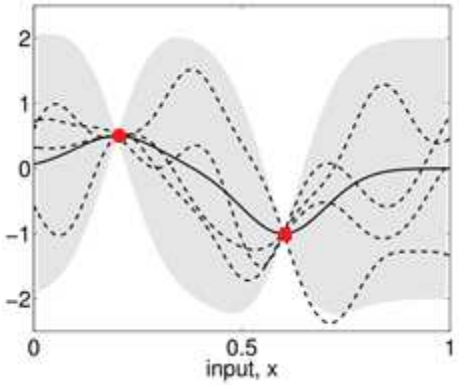


Figure 2-3: Example of Kriging (Rasmussen & Williams, 2006)
(solid line is the mean, gray area is the 95% certainty interval, dashed lines are realizations)

[1] Chapter 4.3 equation 4.69 of Murphy (2012)

2.5.2 Active learning Kriging based Monte Carlo simulation

Kriging is a convenient regression method, as explained above. A metamodelling method that uses Kriging is Active learning Kriging based Monte Carlo Simulation (AK-MCS) and has shown to be an efficient way of approximating failure probabilities (Echard et al., 2011). Active learning means that an iterative process is used in which the training set is expanded based on the uncertainty of the model, especially in the region of interest (i.e. region near the limit state). The process is repeated until a chosen convergence criterion is met. The metamodel is used together with MCS for determining the system response of interest, as the method name indicates. Recently, Van den Eijnden et al., (2021) developed a two-stage AK-MCS metamodel for geotechnical reliability meant for modelling noisy and incomplete models. The model is explained in the next section.

2.6 ERRAGA

The AK-MCS based metamodel by Van den Eijnden et al., (2021) is named ERRAGA (Efficient and Robust Reliability Analysis for Geotechnical Applications) and is meant for modelling noisy and incomplete models. A model is incomplete if it is not able to return an answer for some combinations of input parameters. That could be the case if a part of the parameter space causes a physically impossible situation in the model, called the infeasible domain. One of the two stages in ERRAGA is a classification model. It is trained to predict feasibility of a parameter set. The other stage is a regression model for predicting the model outcome. Both models can be used during simulation to estimate the probability of failure given feasibility. In this work, ERRAGA is used for predicting the model response of slope stability in a reliability updating framework presented in the following chapter. Therefore, this section elaborates on some of the characteristics of ERRAGA and the training of the regression model, the classification model is not discussed. For further explanation the reader is referred to the work by Van den Eijnden et al., (2021).

2.6.1 Learning procedure regression model

The metamodel approximates the real model response, based on a concise set of model evaluation. This set of model evaluations is called the design of experiment (DoE). The prediction of the metamodel is most accurate around the model evaluations. The accuracy decreases with increasing distance from those points in the parameter space. For a reliability analysis, it is important that failure or survival of the system is predicted accurately. The metamodel should therefore be most accurate around the limit state function, at the edge of failure. This reduced the probability of misclassification of failure. However, the location of the limit state function is often not known a-priori. Therefore, an iterative procedure is implemented in ERRAGA to allow for optimal expansion of the DoE during training.

The training starts with an initial set of model evaluations with parameters randomly drawn for the joint parameter distribution. This initial DoE is used for fitting the first metamodel. Then the failure probability is determined by drawing a group of parameter sets $\mathbf{u}_{\mathbf{IS}}$ from the initial importance sampling distribution, elaborated further in the next paragraph. A sample is selected from this group based on a learning function, which is meant to choose the most optimal point for learning. The selected sample is used as input for a model evaluation. The result is added to the DoE, the metamodel is fitted and the reliability is determined with the same $\mathbf{u}_{\mathbf{IS}}$. The process of selecting the optimal learning sample from the sample group is repeated until two convergence criteria are met, one for the certainty of the metamodel prediction, called the stopping criterion and one for the certainty of the calculated probability of failure. The latter is explained in the next paragraph. Multiple stopping criteria are available in ERRAGA. The stopping criterion used in this work is "BetaStop" and is discussed here. This criterion is based on the 1-α certainty bounds of the predicted limit state outcome for $\mathbf{u}_{\mathbf{IS}}$. The reliability is determined for the predicted limit state outcome and the lower and upper bound, denoted by β, β^- and β^+. These reliabilities are used for the stopping criterion presented in equation (21), in which k is the amount of standard deviations from the mean to either one of the certainty bounds (so $k = \Phi^{-1}(\alpha/2)$) .

$$\frac{\beta^+ - \beta^-}{\beta} < 2 \cdot k \cdot \alpha \tag{21}$$

For the learning functions multiple options are available as well. The used learning function in this work is the UNIS learning function. The function predicts the expected reduce in the probability of misclassification regarding failure for the sample under consideration. The function is applied to every sample in $\mathbf{u}_{\mathbf{IS}}$, after which the sample with the highest reduction is selected for addition to the DoE. Details of this learning function are given by Van den Eijnden et al. (2021).

2.6.2 Adaptive importance sampling

The importance sampling methods implemented in ERRAGA make use of a Gaussian Mixture Model (GMM). This model is a combination of multiple Gaussian distributions with a certain mean and standard deviation. The failure probability using simulation in combination with an importance sampling distribution was presented in equation (8) and its variance in (9).

The importance distribution at the start of training begins as a regular Gaussian distribution with a variance such that some cases of failure are probably observed in the initial sampling pool $\mathbf{u}_{\mathbf{IS}}$. After initializing the importance distribution, the learning procedure as described in the previous paragraph commences. After each addition to the DoE, the stopping criterion and the criterion for the probability of failure are checked.

The latter criterion is given in equation (22). If the stopping criterion is met, but the convergence of the failure probability is not, then the importance sampling distribution is adapted. All parameter sets used for expanding the DoE are added to the GMM, resulting in an importance distribution with extra weight around the limit state line. The importance distribution is also adapted if a user-specified maximum of model evaluations is reached. ERRAGA allows for three cycles of adapting the distribution. If both criteria are met, then the learning is finished. The total learning procedure of ERRAGA is presented in appendix B.

$$CoV_{P_f} = \frac{\sigma_{P_f}}{\hat{P}_f} \leq 0.05 \tag{22}$$

3 Methodology

This chapter provides a framework for updating the reliability of slope stability, using dike improvement as the survived loading condition. It discusses the combined use of ERRAGA and slope stability software to obtain a metamodel for slope stability, as well as how to deal with the observed loading condition. Finally, estimation of the updated probability of failure is discussed. The methodology established in this chapter is applied in a case study in chapter 4.

3.1 Model setup

In chapter 2 it was discussed that the failure probability of a dike due to slope instability is assessed periodically. In this assessment the actual dike situation is used, including any stochastic boundary conditions like the water level. This is referred to as the assessment situation. The situation used for reliability updating is referred to as the observed situation, which in this research is the survival of a dike improvement. The intention is to use the observation to reduce the uncertainty in the assessment situation, hereby improving the reliability estimate. Incorporating the observation in the reliability analysis is done with equation (23), as presented in section 2.4.

$$P(F|\varepsilon) = \frac{P(\{g(\mathbf{X}) < 0\} \cap \{h(\mathbf{X}) \geq 0\})}{P(h(\mathbf{X}) \geq 0)} \tag{23}$$

The probability of failure F given the observation ε is the updated probability, it only accounts for failure in the part of the possibility space were the observation is also true. In a probabilistic method using sampling, this is equivalent to only accepting samples that are consistent with the observation. Thus, for the case of updating with construction survival, if a realization survives the observed conditions. Only those realizations are used for computing the updated reliability, details are given in paragraph 3.4.1. So, in order to assess construction survival, also the stability in the observed situation must be evaluated. Hence, two slope stability problems need to be evaluated, for both the observed and the assessment situation. This method uses the metamodelling package ERRAGA to create metamodels for predicting slope stability (Van den Eijnden et al., 2021). The limit equilibrium software D-Stability 20.2.1 by Deltares is used for training the metamodels (Van der Meij, 2019).

ERRAGA is suited to only accept realizations that are consistent with the observation. The default implementation is using a binary classification model, as was done by Van den Eijnden et al. (2021). This binary feature is especially useful when the computational model does not provide a model response for certain parameter sets. That could be the case if those sets cause a physically impossible situation in the model. The parameter combinations without a model response are then used to train the classification model. However, this research uses a limit equilibrium method (LEM) to compute the safety factor, which nearly always returns a model response. Training the classification model with factors of safety means only utilizing the binary component (failure/survival), and not the information on how far the realization is from the point of failure.

Therefore, it is proposed to use a second regression model instead of the classification model. It uses all information and is therefore expected to be more efficient during training of the metamodel. It also provides more insight into the stability behavior in the observed situation, as it returns the limit state response instead of just a classification.

3.2 Model assumptions

This section explains two assumptions regarding this method for reliability updating. Both assumptions are substantiated in the following two paragraphs.

3.2.1 Slip plane consistency

The common way to schematize the subsoil in a slope stability analysis is by subdividing it in layers with homogeneous properties. Geotechnical site investigation such as CPT's and borehole data is used for subdividing the subsoil in layers based on the soil type. Each layer is assigned a set of soil parameters which represent the spatial averaged properties. The result is a subsoil schematization with deterministic layer boundaries, with one set of soil parameters per layer. This approach is used for the evaluation of slope stability in this research. This is however a simplification of reality, as soils are spatially variable. Soil properties can vary from place to place within a soil layer. Therefore, two slip planes passing through a soil layer at different locations may experience different soil properties. One could pass a weaker spot, while the other could go through a stronger area. This phenomenon, which is not represented in the schematization, should be considered when updating the reliability. Updating the reliability is done by only considering parameter value combinations which lead to survival in the observed situation. But for a schematization without spatial variability, gaining knowledge on the strength of one sliding plane may not be enough to update the beliefs of the strength along other slip planes.

This is clarified by means of the example situation in figure 3-1. In the example the spatial variability of a strength parameter of the actual situated is represented by weaker

and stronger areas. This spatial variability is not modelled in the slope stability analysis, as a strength parameter has a single value per soil type. For this realization, a smaller slip plane is critical in the observed situation, while a deeper one is critical in the assessment situation. Both critical slip planes result in failure. In reality, the small slip plane crosses a stronger area, while more averaging takes place along the larger plane. The slip plane averaged strength along the two slip planes is thus not equal. The strength found by observing survival of the smaller plane is therefore not representative for the strength along the deeper plane. In this example, this could lead to overestimation of the strength of the deeper layer. Reliability updating considering different slip planes could therefore lead to an overestimation of the updated reliability.

Therefore, it is assumed that the reliability can only be updated if the two considered slip planes are equal. The chosen strategy in this methodology is to first calculate the slope stability in the assessment situation freely with an algorithm that finds the critical slip plane. Then the same slip plane is used in the stability analysis of the observed situation. Considering equal slip planes prevents overestimation of the updated reliability. This is however a conservative assumption because the observation may also carry information when the slip planes are not exactly equal.

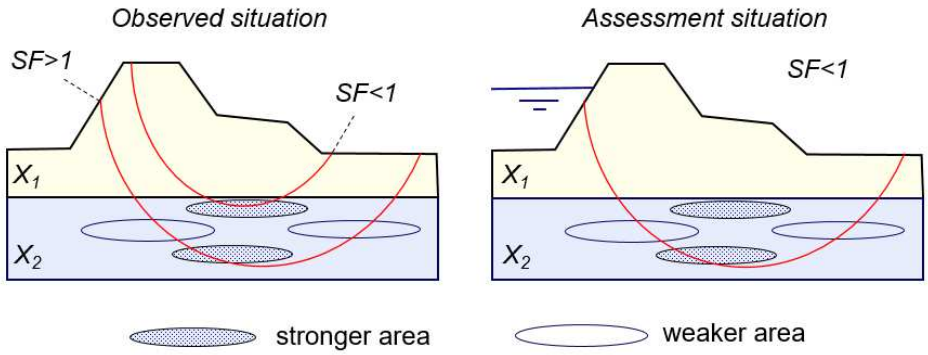


Figure 3-1: Example situation leading to over estimation of the reliability

3.2.2 Reducible uncertainty

There are two types of uncertainty, reducible and non-reducible uncertainty, explained in paragraph 2.4.3. In this method the shear strength ratio and strength increase exponent are assumed to be time invariant, and hence, they are modelled as fully correlated between the observation and assessment situation. The values do not change; the uncertainty is reducible. Also, the POP is assumed to be time invariant, although the analysis accounts for development of stress history due to construction activities and change of hydraulic conditions. The uncertainty of the water level is modelled as non-reducible. The observation of a water level does not provide information for future water levels. For the model factor, the classification of reducibility is less straightforward. The model factor takes account for the possible error between the slip plane model and the slope stability situation in reality. The uncertainty of a model factor in general typically consists of reducible and non-reducible uncertainty (Schweckendiek, 2014). In some situations, the model will perform better than in others, resulting in different model factors from case to case. If the model factor does not change in time, then in theory all its uncertainty can be reduced by considering some observation in the past. However, if there are factors changing the model performance over time, then perhaps the uncertainty can only be reduced partially. For the problems considered by this method it holds that the subsoil stratification is the same for both situations. Also, the considered slip planes are equal. The factors that do change are the pore water pressure and a part of the geometry, due to the improvement. Due to the similarity of the situations it is assumed that the non-reducible part of the model factor is small. The uncertainty of the model factor is assumed to be fully reducible.

3.3 Model training

Section 2.6 explains the training procedure of a single metamodel. In an iterative loop the training set is expanded with extra realizations until the convergence criteria are met. This is true for the approach with two metamodels as well, but the models are not trained separately due to the requirement of equal slip planes explained in the previous section. The learning algorithm implemented for training the model for reliability updating is presented in figure 3-2.

In the learning algorithm, the subscripts A and O denote the assessment situation and observed situation respectively. The design of experiment (DoE) is the set of model evaluations on which the metamodel is trained. M denotes the metamodel and g is the vector of the predicted limit state outcome for the sample group $\mathbf{u_{IS}}$, drawn for the importance sampling distribution. The first step is creating the initial DoE. The stability in both situations is evaluated for a group of samples randomly drawn from the probability distribution. Then the metamodel of the assessment situation is trained. The additional model evaluations of slope stability are based on optimal learning of this situation. In a model evaluation, both stability situations are evaluated with the same slip planes. So, during training of the metamodel in the assessment situation, also the DoE of the observed situation is expanded. After training the metamodel for the assessment situation, training of the observed situation follows. It starts with the importance distribution developed during training of the assessment model. Multiple options are available in ERRAGA for the importance distribution, learning function and stopping criterion. The used settings are presented in table 3-1, which are discussed in section 2.6. The learning algorithm for training the original two staged metamodel is presented in appendix B. Many steps of the algorithms are similar because those steps are implemented in ERRAGA and are used in this method. The updated reliability is determined after training both models, for which the method is given in the next section.

Parameter	**Value**
Importance distribution	Gaussian Mixture Model based on DoE
Learning function	UNIS learning function
Stopping criterion for convergence meatamodel	BetaStop (on the reliability index β)
Stopping criterion value	0.01
Convergence criterion probability of failure	0.05

Table 3-1: ERRAGA settings

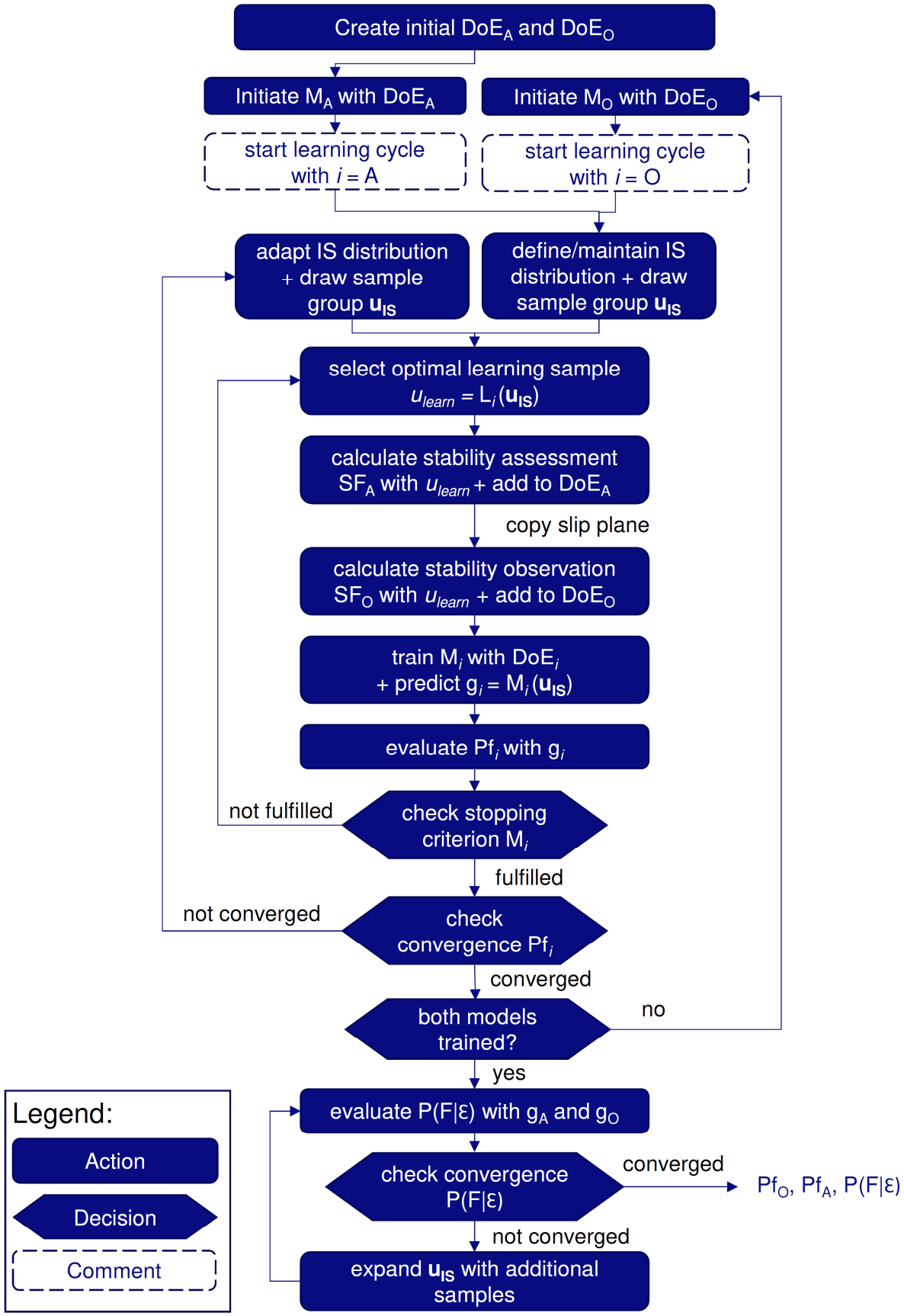


Figure 3-2: Model learning algorithm

3.4 Reliability analysis

In reliability analyses, the failure or survival of a system is usually expressed by the limit state function. It is defined such that negative values represent system failure. For the slope stability analyses in this research, a safety factor below one is defined as failure. This leads to the definition of the limit state functions in the assessment and observed situation in equation (24) and (25) respectively, including the model factor m_d for the chosen slip plane model. These equations are used for identifying slope failure in the reliability analysis. $\mathbf{x}$ represents a single realization of the involved parameters and $\mathbf{x}^*$ the same set but without the model factor. The safety factor is obtained by performing a slope stability software. The reliability of a slope stability problem can be determined by simulation, as elaborated in paragraph 2.3.1.

$$g(\mathbf{x}) = \frac{SF_A(\mathbf{x}^*)}{m_d} - 1 \tag{24}$$

$$h(\mathbf{x}) = \frac{SF_O(\mathbf{x}^*)}{m_d} - 1 \tag{25}$$

3.4.1 Reliability update

The updated probability of failure is determined by only considering realizations which result in survival of the observed situation. The expression in equation (8) for determining probabilities with importance sampling is combined with Bayes theorem from equation (23) to obtain an expression for the updated failure probability, presented in equation (26). In the equation n represents the sample size, $\mathbf{x_i}$ is a single realization of parameters, and $I[\cdot]$ is the indicator function, which is one if the argument is true. The importance sampling weight is $w(\mathbf{x_i})$ and is defined in equation (10).

$$\hat{P}(F \mid \varepsilon) = \frac{\sum_{i=1}^{n} I[g(\mathbf{x}_i) < 0 \cap h(\mathbf{x}_i) \geq 0] \cdot w(\mathbf{x}_i)}{\sum_{i=1}^{n} I[h(\mathbf{x}_i) \geq 0] \cdot w(\mathbf{x}_i)} \tag{26}$$

The convergence criterion for this probability is a maximum coefficient of variation of 0.05, as given in equation (27).

$$CoV_{P(F|\varepsilon)} = \frac{\sigma_{P(F|\varepsilon)}}{\hat{P}(F|\varepsilon)} \leq 0.05 \tag{27}$$

An expression for estimating the failure probability is available in equation (26), but also an expression for its uncertainty is needed to check the convergence criterion. The expression for the conditional probability is a quotient of two probabilities, as seen in

Bayes theorem in formula (23). These two probabilities are random variables with a binomial distribution, as was discussed in paragraph 2.3.1. However, the normal distribution is a reasonable approximation for the binomial distribution for certain conditions for the probability of occurrence and sample size, given by Box et al. (1978). The condition is given in equation (28), in which n is the sample size and p is the probability of occurrence of the event. This probability is not equal to the probability of failure when using importance sampling. In that case, p is the probability of seeing failure in the simulation. An efficient importance distribution should attribute a significant portion of the weight to the region of failure. The probability p should therefore be larger than the failure probability if a reasonable importance distribution is used. For $p = 0.01$, the condition is fulfilled for a sample size of approximately 10^5. In the case study presented in chapter 4, the lowest value found for p is in the order of 10^{-5}, which results in a sample size in the order of 10^6. Such sample sizes are not a restriction from the computational point of view.

$$\frac{1}{\sqrt{n}} \cdot \left| \sqrt{\frac{1-p}{p}} - \sqrt{\frac{p}{1-p}} \right| < 0.3 \tag{28}$$

Approximating the binomial distributions as normal allows the expression in equation (26) to be seen as a quotient of normally distributed random variables. Díaz-Francés & Rubio (2013) proved that the distribution of the ratio of two independent normally distributed random variables with positive means $Z = X/Y$ can to a certain extent be approximated as normal under certain conditions. From Díaz-Francés & Rubio (2013) it follows that for the coefficients of variation $\sigma_Y/\mu_Y < 0.1$ and $\sigma_X/\mu_X < 0.5$ the cumulative density function of the actual distribution and the approximation have a maximum difference of 0.03 within two standard deviations of the mean of Z. In case of the expression in equation (26), the criterion of the denominator is fulfilled, as a coefficient of variation of smaller than 0.05 is required for convergence. Only a check on the coefficient of variation of the joint probability in the numerator of equation (26) remains needed when updating the reliability. However, a variation coefficient larger than 0.5 will probably not result in convergence of the updated probability. It is therefore expected that this criterion does not require the sample set to be expanded.

The only criterion that probably cannot be fulfilled is the requirement of independence. The updated probability is a quotient of probabilities, both involving the same event, construction survival. Therefore, if there is any correlation between the probabilities, then it would probably be positive. However, a positive correlation reduced the spread of a quotient of random variables. For the purpose of checking the convergence criterion, assuming independence is thus a conservative assumption. On this matter, it results in a stricter convergence criterion because the uncertainty is overestimated. The variance

of the quotient is given in Díaz-Francés & Rubio (2013) and presented in equation (29). For applying it to updating the reliability, X represents the estimate of $P(F \cap \varepsilon)$, Y the estimate of $P(\varepsilon)$ and Z the estimate of $P(F|\varepsilon)$.

So, in order to use a normal approximation for $P(F|\varepsilon)$, the criterion in equation (28) must be met and the coefficient of variation of $P(F \cap \varepsilon)$ may not be greater than 0.5. Then the convergence criterion of the updated probability in equation (27) can be evaluated using the expression of the variance in equation (29).

$$\sigma_Z^2 = \frac{\mu_X^2}{\mu_Y^2} \cdot \left(\frac{\sigma_X^2}{\mu_X^2} + \frac{\sigma_Y^2}{\mu_Y^2} \right) \tag{29}$$

The expression for the variance of the failure probability for a single metamodel is used from Van den Eijnden et al. (2021) and presented in equation (30).

$$\sigma_{\hat{P}_f}^2 = \frac{1}{n} \left(\frac{1}{n} \sum_{i=1}^{n} w(\mathbf{x}_i)^2 \; I[g(\mathbf{x}_i) < 0] - \hat{P}_f{}^2 \right) \tag{30}$$

3.4.2 Combined importance sampling

Dike failure mechanisms generally have a low probability of occurrence. Estimating a low probability with simulation requires a large sample size, which may cause impractical computational times. This explains the need for importance sampling methods (paragraph 2.3.1). The importance sampling methods implemented in ERRAGA make use of a Gaussian Mixture Model (GMM), described in paragraph 2.6.2. During the training of a metamodel this GMM is adapted for efficiently estimating the reliability of the system under consideration. This is achieved by allocating extra probability density to the region around the limit state function. In the training procedure presented in figure 3-2, first the metamodel in the assessment is trained, together with a GMM. Hereafter, the metamodel for the observed situation is trained. The training does not start with a new GMM but continues with the GMM from the assessment situation. The resulting GMM is thus a distribution with extra weight around both limit state functions. This is an efficient approach when the limit state functions of the observed and assessment situation are somewhat similar, as is the case for the slope stability problems considered in this work. Though a different approach may be needed for problems in which the limit state functions are very different. In that case the current approach may result in an inefficient learning procedure. The combined GMM is used for determining the updated probability. This distribution is not suggested to be the most efficient, but covers the important domain of the parameter space and was found to be effective in this work.

4

Case study: Kinderdijk – Schoonhovenseveer

In this chapter the case study dike improvement Kinderdijk-Schoonhovenseveer is introduced. This case is used for investigating the effect of incorporating dike improvement construction survival in the reliability assessment of slope stability. First the case is introduced, and the improvement data is presented in section 4.1 and 4.2. Then the hydraulic and geotechnical boundary conditions are discussed in 4.3. Finally, the prior reliability is determined and discussed in 4.4. In chapter 5, the reliability of this case study is updated with various loading conditions.

4.1 Case introduction

The dike between Kinderdijk and Schoonhovenseveer (KIS) is part of the primary flood defense trajectories 16-2 and 16-3 and is located east of Rotterdam. The dike is maintained by Waterboard Rivierenland and was reinforced in 2016-2018. For design purposes, the dike was subdivided into sections. The reliability updating methodology is applied to section L1, where the dike was heightened, and a berm was constructed. The situation of the dike trajectory and section L1 is presented in figure 4-1.

Figure 4-1: Location KIS and section L1

4.2 Construction data

During the dike reinforcement, section L1 was monitored with four water pressure gauges and various settlement plates. Together with construction verification reports provided by waterboard Rivierenland, this gives a clear insight into how the reinforcement was phased and how the pore water pressure developed. The location and names of the settlement plates and water pressure measurements are shown in figure 4-2. Two of the pressure gauges were positioned underneath the crest and two in the hinterland beneath the new berm, all at different depths. The measured heads are presented in figure 4-3. The head measurements are corrected for settlements, for which the nearest settlement plate is used. The amount of excess pore water head at the maximum level is presented in table 4-1 and the position of the pressure gauges in the dike profile is presented in figure 4-4.

Other data sources data sources that are used for the schematization of the stability calculations are listed below. The remainder of this chapter uses this list for referencing:

[1] Hydraulic boundary conditions database WBI2017_Benedenrijn_16-2_v04, Rijkswaterstaat, Helpdesk water, www.helpdeskwater.nl
[2] TNO Hydraulic head map, https://www.grondwatertools.nl
[3] Algemene Hoogtekaart Nederland AHN3, https://www.pdok.nl

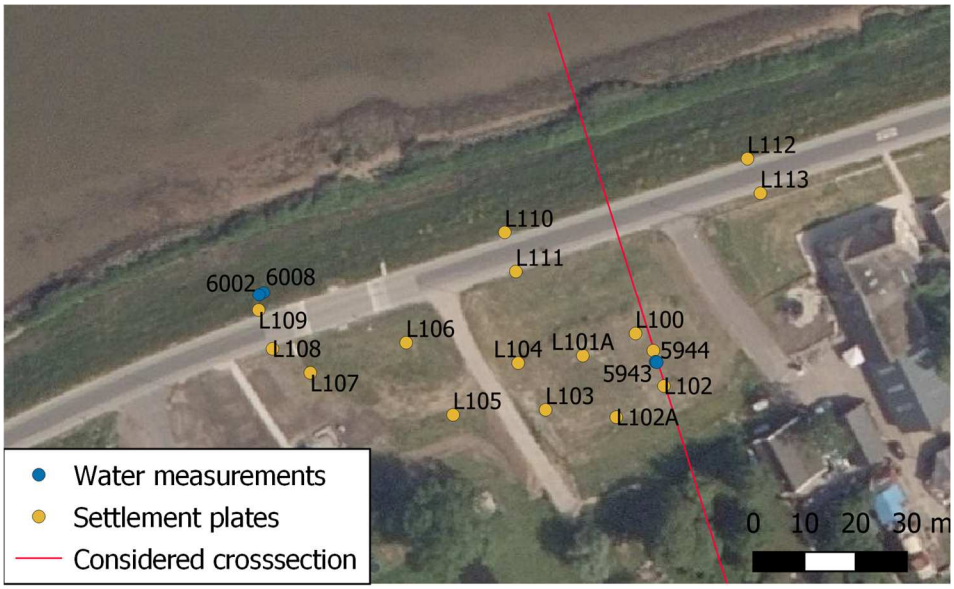


Figure 4-2: Measurement points dike section L1

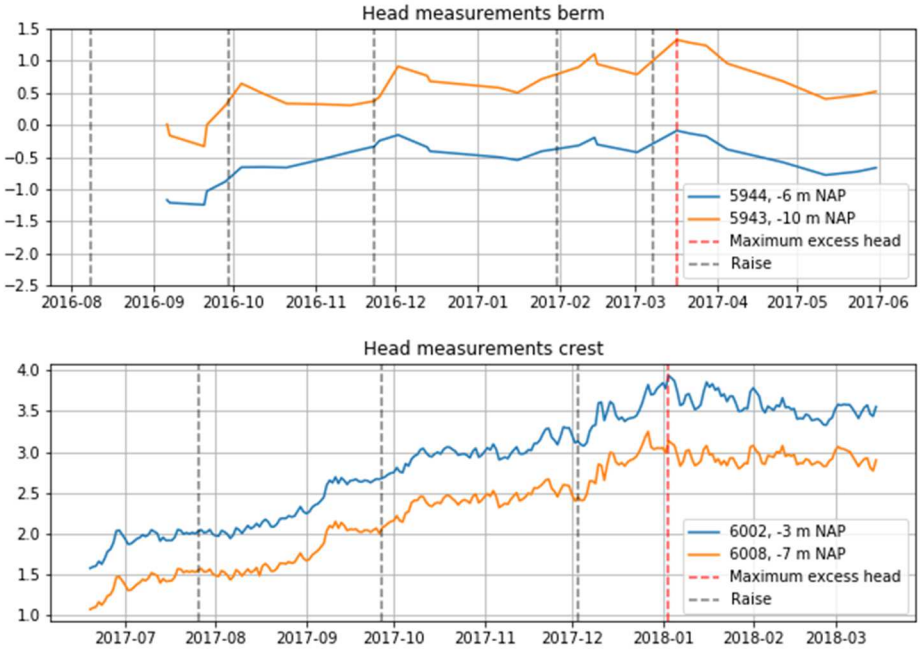


Figure 4-3: Pore water pressure measurements

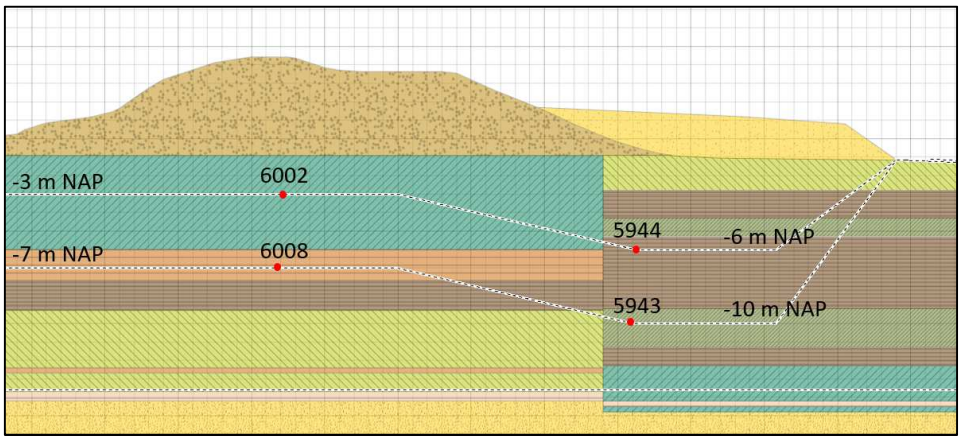


Figure 4-4: Location pore water pressure measurements (red dots indicate position)

	Δh_{crest} [m]		Δh_{berm} [m]	
Construction phase	**-3 m NAP (6002)**	**-7 m NAP (6008)**	**-6 m NAP (5944)**	**-10 m NAP (5943)**
Construction berm	0.00	0.00	2.10	2.20
Raise of crest	1.80	2.70	0.00	0.00

Table 4-1: Excess pore water head per phase

4.3 Boundary conditions

The boundary conditions are determined such that the analyses performed here are as close to reality as possible. Therefore, several conditions are adopted from the dike improvement design. However, the dike improvement was designed according to the old safety standards, which used exceedance instead of failure probabilities and only used drained shear strength parameters for soils. For these parameters new values are determined, elaborated in the coming paragraphs.

4.3.1 Hydraulic boundary conditions

Extreme river water level

In the dike improvement design only one design water level was determined. However, in a probabilistic analysis a probability distribution is needed. With use of the software Hydra-NL, various water levels are determined and used for fitting an extreme value distribution. The water levels are determined at a location 1000 meter upstream of section L1 and are presented in table 4-2. The following holds for the derived the water levels:

- The current distribution is considered, no climate change is included
- Model and statistical uncertainty is included (built-in in Hydra-NL)
- Hydraulic boundary condition database "WBI2017_Benedenrijn_16-2_v04" is used [1]
- Hydra-NL version 2.7.1 (November 2019)

Exceedance probability [1/year]	**Water level [m + NAP]**
1/10	2.89
1/30	3.07
1/50	3.14
1/100	3.23
1/300	3.35
1/1000	3.48
1/3000	3.58
1/10000	3.71
1/30000	3.90
1/100000	4.14

Table 4-2: Water levels from Hydra-NL

The water levels above are shown as red dots in the exceedance probability plot in figure 4-5. A Generelized Extreme Value distribution (GEV) is fitted to the data, which is a combination of three extreme value distributions, Gumbel, Fréchet and Weibull. The CDF of a GEV distribution is shown in equation (31), in which ξ, σ and μ are the shape, scale and location parameter respectively. Fitting a single distribution to the data did

not provide a good fit, hence a second GEV distribution was used to model the behavior of the tail. The fitted distribution parameters are shown in table 4-3.

$$F(x) = \exp\left(-\left(1+\xi\left(\frac{x-\mu}{\sigma}\right)\right)^{-1/\xi}\right) \; for \; \xi \neq 0 \tag{31}$$

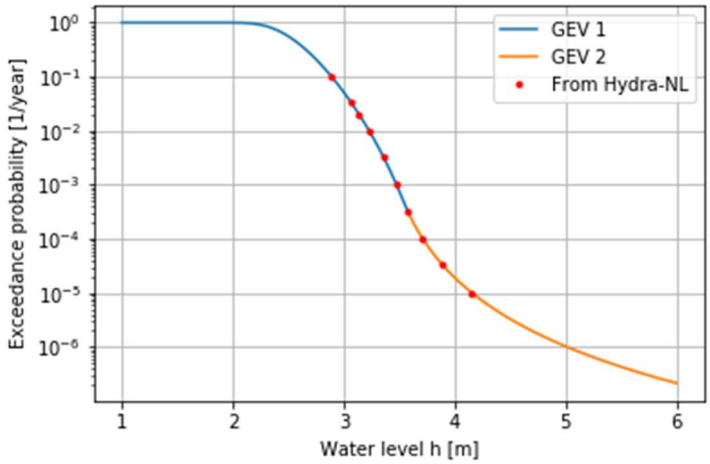


Figure 4-5: Exceedance probability curve

Distribution	Domain	Shape ξ	Scale σ	Location μ
GEV 1	h <= 3.52 m	0.127	2.457	0.222
GEV 2	h > 3.52 m	-0.287	3.272	0.010

Table 4-3: GEV distribution parameters

Daily river water level

The average water level in the river is 0.40 m + NAP, adopted from the dike improvement design.

Polder water level

In the summer the polder level is maintained at -2.11 m NAP and in the winter at -2.21 m NAP. In the daily condition the winter polder level is used. In the design condition the water level in the polder ditches may be increased by seepage and overtopping water, the polder water level is therefore schematized close to the polder surface level.

Daily head aquifer

The hydraulic head in the aquifer is -1.00 m NAP, adopted from the dike improvement design. The value of the head in the aquifer is verified using the TNO hydraulic head map [2].

Pore water pressure

The schematization of the pore water pressure is adopted from the dike improvement design. Three head lines are used, the phreatic line (HL1), the aquifer head line (HL3),

and the head under daily circumstances (HL2), used to schematize the intrusion length during high water conditions. The phreatic pressure is assigned to the subsoil from the surface level to two meters below the surface level of the hinterland in all conditions. The aquifer head is assigned to the top of the aquifer. Under daily conditions, linear interpolation takes place between the former two heads. Under high water conditions, the phreatic and aquifer head interpolate towards HL2, which is assigned three meters above the aquifer, the intrusion length. The schematization of the head lines is illustrated by means of figure 4-6 and table 4-4, in which h_d and h_h are the daily and high river water level, aq is the daily head in the aquifer, PL is the polder level and SFL is the surface level. r is the response factor and equals 0.35, as in the improvement design (W+B, 2013).

Situation	**Head line**	**a**	**b**	**c**	**d**	**e**	**f**
Daily conditions	HL1	h_d	h_d+1.0	h_d+1.0	PL	-	PL
	HL3	aq	-	-	-	-	aq
High water	HL1	h_h	h_h -1.0	h_h -1.5	SFL - 0.1	SFL	SFL
	HL2	h_d	-	-	-	PL	PL
	HL3	h_h	-	-	h_d+r(h_h-h_d)	-	h_h

Table 4-4: Head line values at reference points

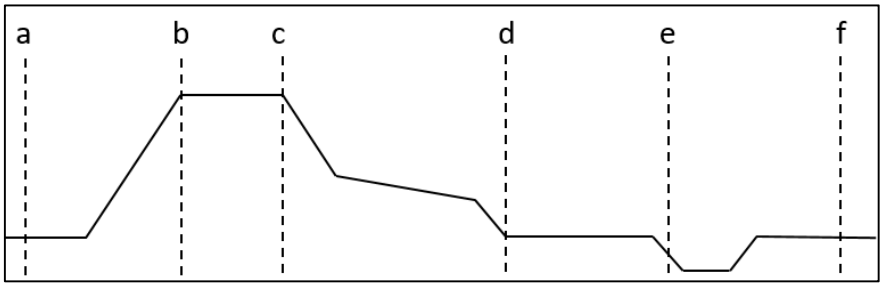


Figure 4-6: Reference points for head line schematization

4.3.2 Geometry

The cross-section line in figure 4-1 was projected to the Dutch national height map AHN3 [3] to obtain the cross-sectional geometry. It was recorded in 2015 and thus represents the situation before improvement. The geometries of the various construction phases are derived using the settlement plate recordings. For the situation after settlement the dike improvement design is used (W+B, 2013). The geometry of four construction stages is presented in figure 4-7. These are the stages that are used for updating the reliability in the next chapter.

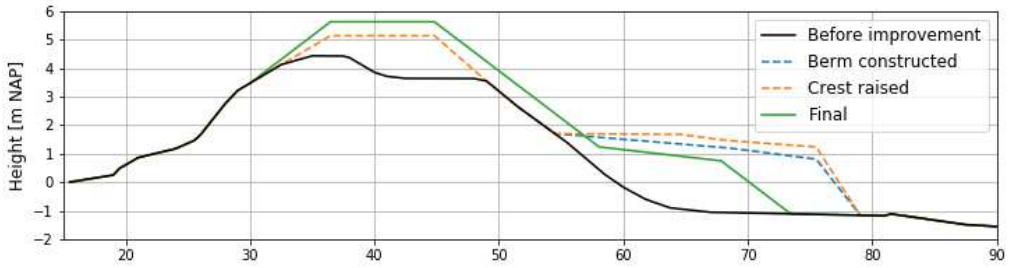


Figure 4-7: Geometry of construction stages

4.3.3 Soil parameters

During the design of dike improvement KIS, it was common practice for Dutch engineers to use a drained shear strength model for soils with undrained soil behavior. However, new insights have led to the use of the SHANSEP model for modelling undrained shear strength, elaborated on in paragraph 2.2.1. As a result, the collection of shear strength parameters used for the dike improvement design is not suited for the analyses in this research. Instead, the collection of shear strength parameters of the region "Neder-Betuwe" is used, obtained from Deltares, involved in reinforcing dike trajectories in that region. The Neder-Betuwe is located around the city of Tiel, some 60 kilometers upstream of KIS. Seen the scale of geological depositions, the location of KIS and the resemblance of the regions, it is expected that the use of this alternative parameter set will not significantly alter the reliability of the dike trajectory. This regional collection however, does not contain state parameters (as they should be site specific). Therefore, compression tests of dike improvement Salmsteke – Schoonhoven (SAS), a dike trajectory just upstream of KIS, are used for deriving distributions for the POP. The compression test data is presented in appendix C.

The SAS collection of 42 tests is divided in four groups for which separate distributions for the POP are derived. A distinction is made between peat and light clay (<17.5 kN/m^3), and between soils under, and besides the dike. The latter distinction is made because of the difference in stress history. A lognormal distribution is fitted to each of the groups, representing the POP at a certain point in a soil layer. In appendix C it can be seen that for soil samples taken from the same layer in the same borehole at different depths, the resulting POP values vary, indicating spatial variation. So, it is expected that some spatial averaging takes place along a potential sliding plane. The method from the TAW (2001) report on water retaining earthen structures is used to adjust the distributions to account for spatial averaging. It uses the variance reduction factor Γ^2 to account for spatial variation along a sliding plane. Although the value of Γ^2 is uncertain, 0.25 seems to be a reasonable choice (TAW, 2001). The spatial averaged standard deviation is shown in equation (32). The method is described for normally distributed variables, therefore the standard deviation of the logarithm of the POP is used. Thereafter a sample size correction is done to the standard deviation using equation (33), in which $t_{n-1}^{0.95}$ is the 95th-percentile of the student-t distribution with n-1 degrees of

freedom, and 1.645 the 95th-percentile of the standard normal distribution. The student-t distribution is often used to account for statistical uncertainty when determining characteristic values of soil properties, equation (33) should be seen as applying this method to a whole distribution, in which the 5th percentile serves as a calibration point.

$$\sigma_{\ln(POP),av} = \sigma_{\ln(POP),point}\sqrt{\Gamma^2 + \frac{1}{n}} \tag{32}$$

$$\sigma_{\ln(POP),av,corr} = \sigma_{\ln(POP),av} \cdot \frac{t_{n-1}^{0.95}}{1.645} \tag{33}$$

All the soil parameters are presented in table 4-5. The weight of the soils is adopted from the dike improvement, which is based on site specific weight tests. The uncertainty of the parameters is expressed through the coefficient of variation (CoV), which is σ/μ. The presented CoV values for the POP include spatial averaging as presented above. The other soil parameters are not corrected for spatial variability. All soil strength parameters are assumed to be lognormally distributed.

	Saturated weight	**Dry weight**	**Friction angle**		**S**		**m**		**POP**			
Soil type	[kN/m3]	[kN/m3]	μ [°]	CoV [-]	μ [-]	CoV [-]	μ [-]	CoV [-]	μ below [kPa]	CoV below [-]	μ besides [kPa]	CoV besides [-]
Dike material (clay)	18.54	-	35.0	0.07	-	-	-	-	-	-	-	-
Peat	10.63	-	-	-	0.41	0.06	0.75	0.08	35.7	0.29	33.9	0.29
Peat clayey	11.69	-	-	-	0.33	0.03	0.95	0.04	35.7	0.29	33.9	0.29
Clay strong organic	13.08	-	-	-	0.29	0.05	0.93	0.10	40.7	0.40	28.4	0.32
Clay silty organic	15.16	-	-	-	0.30	0.04	0.77	0.10	40.7	0.40	28.4	0.32
Clay silty	16.82	-	-	-	0.32	0.05	0.83	0.07	40.7	0.40	28.4	0.32
Sand	20.00	18.00	34.0	0.05	-	-	-	-	-	-	-	-
Raise material sand	19.00	17.00	32.6	0.05	-	-	-	-	-	-	-	-

Table 4-5: Soil property parameters

4.3.4 Soil stratification

The soil stratification of section L1 is adopted from the dike improvement design and presented in table 4-6 and figure 4-8. The site investigation used for the dike design (W+B, 2013) shows a significant difference between the stratification under the dike and in the hinterland. Hence different soil stratifications are used, separated at the inner dike toe.

Dike		**Hinterland**	
Soil type	**Top layer [m NAP]**	**Soil type**	**Top layer [m NAP]**
Dike material clay	-	Clay silt organic	-
Clay silt	0.9	Peat clayey	-2.8
Peat	-6.0	Clay organic	-4.3
Peat clayey	-7.7	Peat clayey	-5.3
Clay silt organic	-9.3	Clay organic	-9.2
Peat	-12.4	Peat clayey	-11.3
Clay silt organic	-12.7	Clay silty	-12.3
Heavy peat	-13.7	Heavy peat	-14.2
Sand	-14.2	Clay silty	-14.5
		Sand	-14.8

Table 4-6: Soil stratification L1

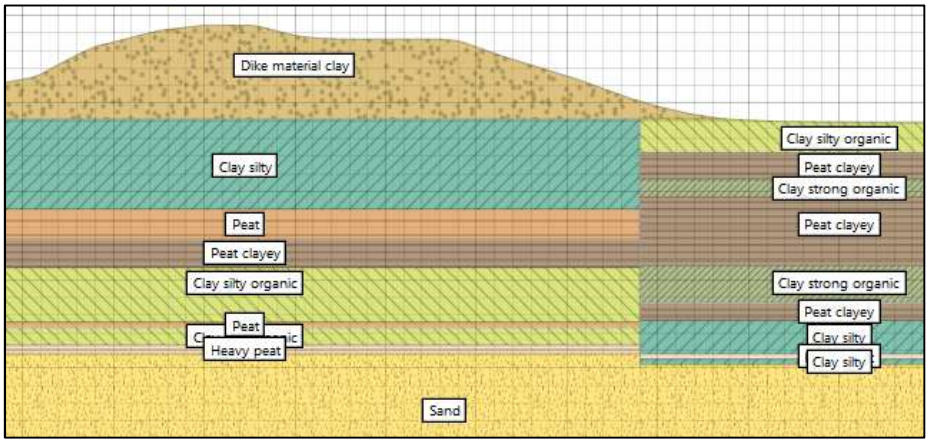


Figure 4-8: Visualization of stratification section L1 (D-Stability)

4.4 Prior reliability

This section elaborates on the estimation of the prior reliability and forms a basis for the reliability update in the next section. The reliability is estimated using ERRAGA, as described in paragraph 2.6.1. The prior reliability is determined for five points in time. The initial condition, the situation after improvement for both the daily conditions, and with inclusion of the extreme water level distribution, and two points in time during construction of the dike reinforcement. The reinforcement was executed in many steps, with time in between to allow for consolidation. The two points in time during construction are chosen based on the pore water pressure measurements. The stability is expected to be critical when the excess pressure reaches a maximum. Maxima can be found after construction of the berm, and after partially raising the crest, presented in figure 4-3 and table 4-1. The situation after these construction steps are used for updating the reliability in chapter 5.

4.4.1 Stochastic parameters

The parameters listed below are treated as random variables in the reliability analyses, adding up to a total of 14. All soil strength parameters are assumed to be lognormally distributed with the parameters from table 4-5. The extreme value distribution of the water level was presented in paragraph 4.3.1. The model factor quantifies discrepancies between a real slip plane and the slip plane model used for schematization. It has a normal distribution for which the parameters are given in table 4-7. This work uses the slip plane model Uplift-Van.

- Water level (h)
- Shear strength ratio (S)
- Strength increase exponent (m)
- Pre-overburden pressure (POP)
- Model factor (m_d)

Model	$\boldsymbol{\mu_{m_d}}$ [−]	$\boldsymbol{\sigma_{m_d}}$ [−]
Bishop	1.025	0.050
Uplift-Van	1.005	0.033
Spencer	1.008	0.035

Table 4-7: Distribution parameters model factor (Schweckendiek, van der Krogt, et al., 2017)

4.4.2 Parameter dependency

The POP of a soil layer has two origins, it develops over time but is also dependent on the stress history of the soil layer. A load applied to the surface level has effect over a certain depth, hence dependency in the vertical direction is expected over that depth. Due to this effect it is assumed that the values for the POP of the soil layers below the crest are fully dependent, as are those of the layers in the hinterland. Assuming full

correlation seems contradictory to considering spatial averaging in paragraph 4.3.3. However, from the analysis in paragraph 4.3.3 it followed that various values for the POP were found in the same soil layer, hence the true correlation differs from one. Still, considering spatial averaging is a conservative choice from the reliability updating perspective. It results in distributions less spread around the mean and hence a higher prior reliability estimate, leaving less to be updated. The approach of spatial averaging combined with correlated POP values is considered to be reasonable for analyzing the effectiveness of reliability updating in this research. All values for S and m are assumed to be independent per soil type.

4.4.3 Phasing

State parameters like the POP are parameters that depend on the circumstances. When the circumstances change (e.g. increase in load), the POP also changes. It is therefore crucial to consider all loading conditions that affect the POP, from determining the POP, up to the point in time of interest. Otherwise, some development or loss of POP might not be considered. The stages listed underneath are considered for achieving this. In D-Stability all stages can be schematized in one calculation. This ensures that the appropriate values for the POP are used in each phase.

- Initial condition (daily water level)
- Construction of the berm
- Raising the crest
- Final geometry (consolidated, daily water level)
- Final geometry (extreme water level distribution)

4.4.4 ERRAGA parameter settings

The ERRAGA settings used for calculating the prior reliability are shown in table 4-8. The function of the ERRAGA parameters are specified in section 2.6. The chosen convergence criterion value of 0.01 results in a 99-percent certainty range of 0.3 on the reliability index for a reliability index around 6. This range decreases for decreasing reliability indices.

Parameter	**Value**
Initial training set size	50
Maximum training set size	450
Learning function	UNIS
Convergence criterion	BetaStop (on the reliability index β)
Convergence criterion value	0.01
Minimum amount of white noise	0.001 (on the limit state outcome)
Number of kernel parameter optimizations	20

Table 4-8: ERRAGA parameter settings

4.4.5 Results

A reliability analysis with ERRAGA was performed for every phase of the dike improvement, for which the results are presented in table 4-9. The reliability index is given for each phase, accompanied with the safety factor resulting from a 5^{th}-percentile semi-probabilistic calculation (model factor not included). The extreme water level distribution is only included in the assessment situation. In the other stages the water level is fixed to a daily level determined in paragraph 4.3.1.

FORM analyses were done on the trained metamodels to obtain some insight into the importance of the variables involved. For this Python package OpenTURNS (Baudin et al., 2015) is used. The results are presented in table 4-10. The design points in standard normal space (U-space) are given, as well as the squared values of the importance factors α^2. Also, the reliability index resulting from the FORM analyses are presented. In a FORM analysis the design point is defined as the point on the limit state line with the highest probability density.

From the analyses follow that all construction phases have a failure mode contributing most to the probability of failure, here called the main sliding mode. Going through the slope stability calculations in the metamodel training sets gives an idea of which slip planes are important. Performing a slope stability analysis with the design point values results in the most probable sliding plane. The slip planes resulting from a design point slope stability analysis of the berm phase and assessment situation are given in figure 4-9. The main mode of the crest phase and daily conditions are very similar to the main mode of the assessment situation and thus not presented. Looking at the main sliding modes turns out to be useful for reliability updating, further explained in chapter 5.

The metamodels trained in the analysis are hard to visualize as a whole. Presenting the metamodel as an analytical expression wouldn't provide much insight, as it is a function of all the (hundreds of) model evaluations in the training set. Plotting the metamodel is possible but the dimensionality of 14 prevents the metamodel to be plotted as a function of all the variables. Therefore, in figure 4-10, the metamodel prediction of the limit state function in the assessment situation is presented in a matrix plot. The diagonal gives the limit state outcome as a function of one parameter, and the non-diagonal as a function of two. The remaining parameters are fixed in the design point so that the most relevant part of the model is visible. For readability, only the four most important parameters are plotted. The difference in importance factors is big compared to the other variables (table 4-10), so these four parameters are decisive and represent the important part of the metamodel. The limit state plot over the whole parameter space is presented in appendix D.

Situation	Initial conditions	Berm phase	Crest phase	Daily conditions	Assessment situation
Safety factor	1.11	1.14	1.09	1.05	1.03
Reliability index β	5.8	6.7	6.9	6.1	6.0

Table 4-9: Result prior reliability analyses with ERRAGA

Phase	Initial conditions		Berm phase		Crest phase		Daily conditions		Assessment situation	
FORM reliability	β = 5.7		β = 6.6		β = 6.8		β = 6.1		β = 5.9	
Parameter	$\mathbf{U_{design}}$	$\boldsymbol{\alpha^2}$	$\mathbf{U_{design}}$	$\boldsymbol{\alpha^2}$	$\mathbf{U_{design}}$	$\boldsymbol{\alpha^2}$	$\mathbf{U_{design}}$	$\boldsymbol{\alpha^2}$	$\mathbf{U_{design}}$	$\boldsymbol{\alpha^2}$
Water level	-0.01	0.00	0.01	0.00	0.01	0.00	-0.03	0.00	0.39	0.00
POP below dike	-2.76	0.23	-2.53	0.15	-2.48	0.13	-1.65	0.07	-1.78	0.09
POP hinterland	-2.71	0.23	-3.44	0.27	-2.67	0.15	-2.35	0.15	-2.15	0.13
Model factor	2.93	0.26	3.91	0.35	4.22	0.38	4.03	0.44	3.85	0.43
S silty clay	-2.73	0.23	-1.66	0.06	-3.67	0.29	-3.06	0.25	-2.94	0.25
m silty clay	-0.61	0.01	-0.46	0.01	-0.34	0.00	-0.70	0.01	-0.92	0.02
S peat clayey	-0.25	0.00	-0.87	0.02	-0.74	0.01	-0.80	0.02	-0.66	0.01
m peat clayey	-0.12	0.00	-0.26	0.00	-0.29	0.00	-0.27	0.00	-0.31	0.00
S clay silty organic	-0.16	0.00	-0.06	0.00	-0.99	0.02	-1.03	0.03	-0.98	0.03
m clay silty organic	-0.20	0.00	-0.09	0.00	-0.18	0.00	-0.14	0.00	-0.18	0.00
S organic clay	-0.92	0.03	-2.47	0.14	-0.26	0.00	-0.23	0.00	-0.21	0.00
m organic clay	-0.43	0.01	-0.36	0.00	-0.18	0.00	-0.17	0.00	-0.01	0.00
S peat	0.03	0.00	-0.75	0.01	-0.80	0.01	-0.90	0.02	-0.82	0.02
m peat	-0.02	0.00	-0.17	0.00	-0.21	0.00	-0.01	0.00	0.03	0.00

Table 4-10: Results FORM analyses
(the squared importance factors may not add up to one due to rounding)

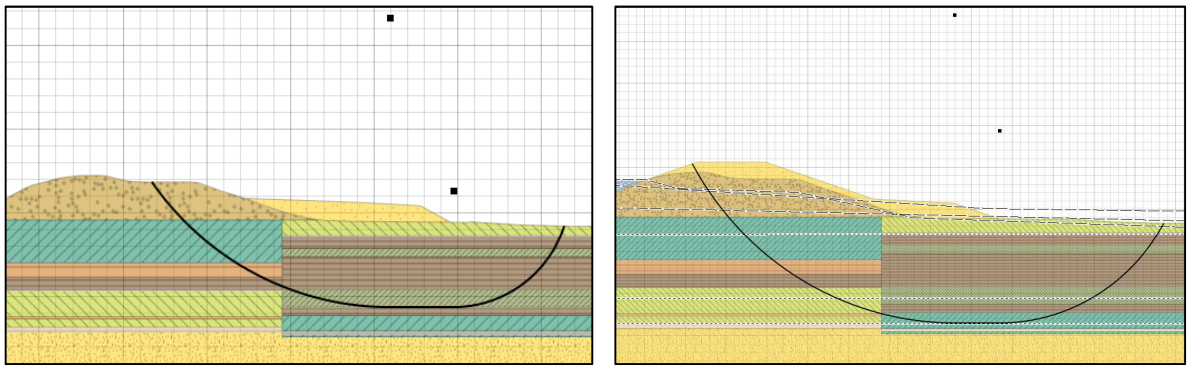


Figure 4-9: Main sliding modes of the berm phase (left) and the assessment situation (right)

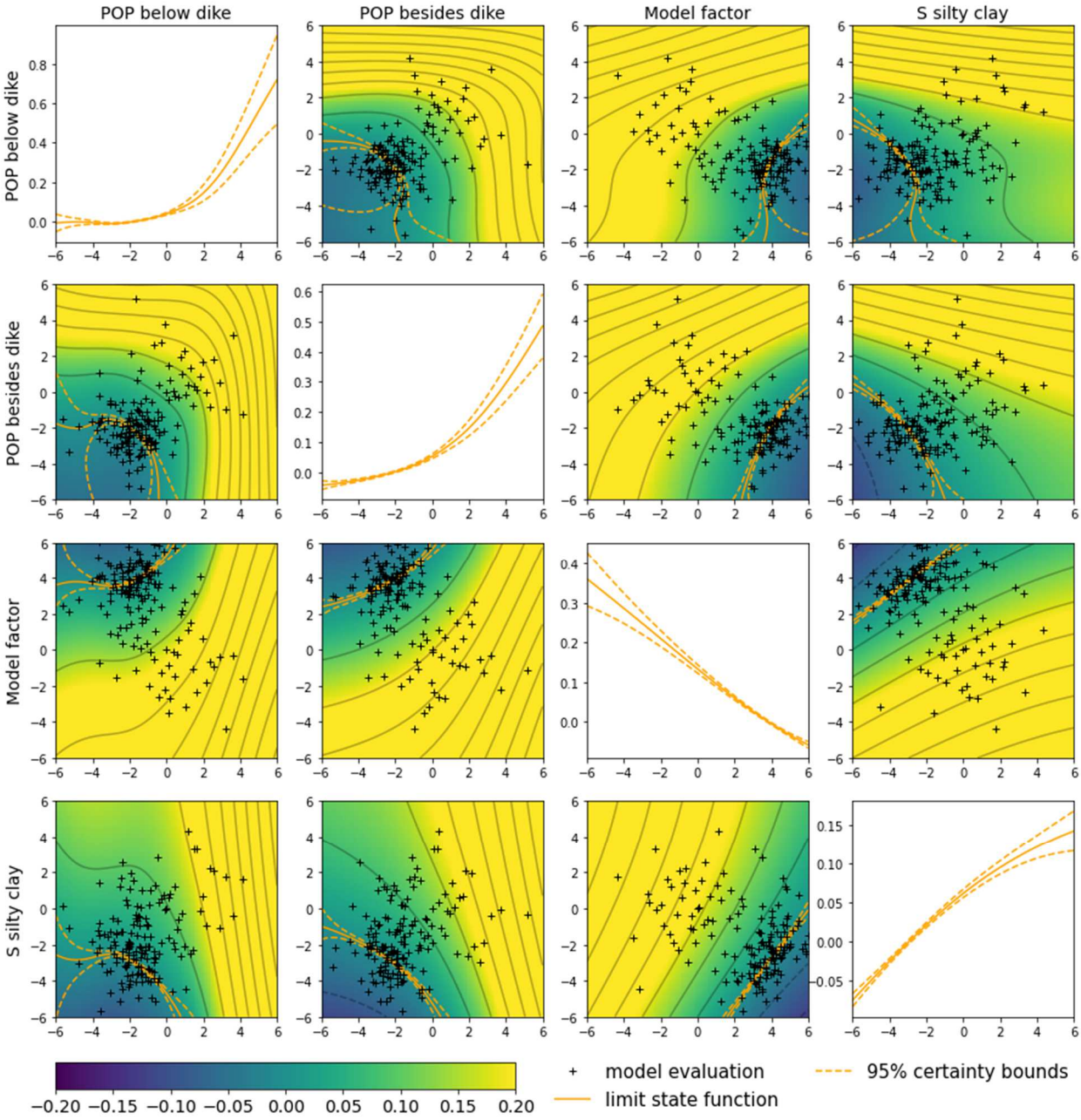


Figure 4-10: Limit state outcome of the four most important parameters in U-space obtained by ERRAGA

(Diagonal: limit state outcome as a function of one variable. Non-diagonal: colors indicate the outcome of the limit state function, orange line is the limit state line)

4.4.6 Discussion

The safety factors from the semi-probabilistic slope stability analyses show that the assessment situation is the most critical loading condition. Constructing the berm has a stabilizing effect. This allows the remaining construction activities to be executed without the safety factor being worse than in the assessment situation. The reliability indices largely confirm the former findings, with the noteworthy difference that the reliability is worst in the initial condition. It is concluded that in this case, the construction of the dike improvement is a less severe loading condition than the assessment situation. Note that the safety factor and reliability during daily conditions is lower than after the crest raise. This is because another raise takes place after the crest raising phase considered here, causing the safety factor and reliability to decrease.

From the FORM analyses follow that the POP, the shear strength ratio of silty clay and the model factor are by far the most influential parameters. The sum of squared influence factors of these four variables adds up to at least 0.80 in all phases. The importance of these variables thus remains consistent throughout the phases, although there is a redistribution between the four. The squared importance factors of both POP variables approximately decrease by a factor two from the first to the last phase. This is explained by the load increase due to the dike raise and the loss of stress history inherent to it (decrease of the POP). When the POP in the initial condition is smaller than the addition of effective stress due to improvement, then all stress history is lost. This causes the lower tail of the distribution of the POP to become less important, as all POP is lost up to a certain value. This can be recognized in the top left graph of figure 4-10, where the limit state outcome is plotted as a function of the POP below the dike. It shows the invariant behavior of the limit state outcome up to a value of -2 in U-space. The water level is in this case barely of influence. This is typical for a situation with a thick soft soil blanket and an aquifer that is not very sensitive to an external water level rise (response factor is 0.35).

The difference in main sliding mode is also a feature that can be found in the importance factors. The sliding mode in the berm phase differs from the other phases. It is smaller and shallower and does not pass the deep silty clay layer in the hinterland (blue layer in figure 5-3). This results in a lower importance factor for the shear strength ratio of silty clay, while the shear strength ratio of organic clay has a much higher value than in the other phases.

5 Case study: Posterior Analysis

This chapter investigates the effect of reliability updating with observations of survived loading conditions for the case study Kinderdijk-Schoonhovenseveer, introduced in chapter 4. Also, the potential impact of including a reliability update in the dike design is explored. This chapter concludes with a discussion of all the results.

5.1 Observations

The effect of reliability updating is explored by updating the slope reliability of the case study with various observed loading conditions. A reliability update is done with the observed survival of the dike improvement construction phases. Also, two hypothetical construction sequences are used for updating the reliability, referred to as alternative A and B. The actual dike improvement construction sequence is referred to as the original case. Besides updating with construction survival, also an update is done with survival of the daily conditions and with a hypothetical observed water level. The updating results are presented individually, after which an overview is given. Finally, the results are analyzed in the discussion.

In each reliability update four values for the reliability β are presented. For the assessment situation the prior and posterior reliabilities are given, β_{prior} and $\beta_{posterior}$. For the observation two reliabilities are given as well. The actual reliability in the observation β_{obs}, determined with a free slip plane search, and the reliability in the observed situation obtained by using the critical slip plane from the assessment situation, $\beta_{obs,fixed}$. The latter shows how critical the observed loading condition was for the slip plane area relevant to the assessment situation. The process of training the metamodel to determine this value is presented in figure 3-2 (where $\beta_{obs,fixed}$ relates to $P_{f,O}$). To obtain insight in the importance of the parameters, a FORM analysis is done on each metamodel. The resulting importance factors are used for comparing loading conditions.

5.1.1 Construction sequences

The original dike improvement sequence and the two alternatives are presented in figure 5-1. In the original case the berm was constructed first, then the crest was raised partially, and finally the rest of the crest was constructed and the berm was reduced to its final dimensions. In alternative A, the partial crest raise takes place before construction of the berm. This is expected to create a more severe loading condition than in the original

improvement, which should affect the reliability update. In alternative B, the whole crest is constructed before constructing the berm.

A point of attention is how much of excess pore water pressure should be used in the alternative construction sequences, as the pressure measurements represent the original phased construction. The crest raise took place after dissipation of the excess pressure caused by the berm construction. The excess pressure found in the measurements after the crest raise were thus only caused by the crest raise, as the soil under the berm already consolidated. Therefore, it is assumed that the pressure measurements after berm and crest construction can be treated as independent. Thus, the actual measurements after berm construction and raising the crest can be used in alternative A. In alternative B the whole crest is raised before constructing the berm. This has the potential to create more excess pore water pressure than the partial crest raise, depending on the construction speed. Yet, to create a not too extreme loading condition, the actual measurements are used for alternative B as well.

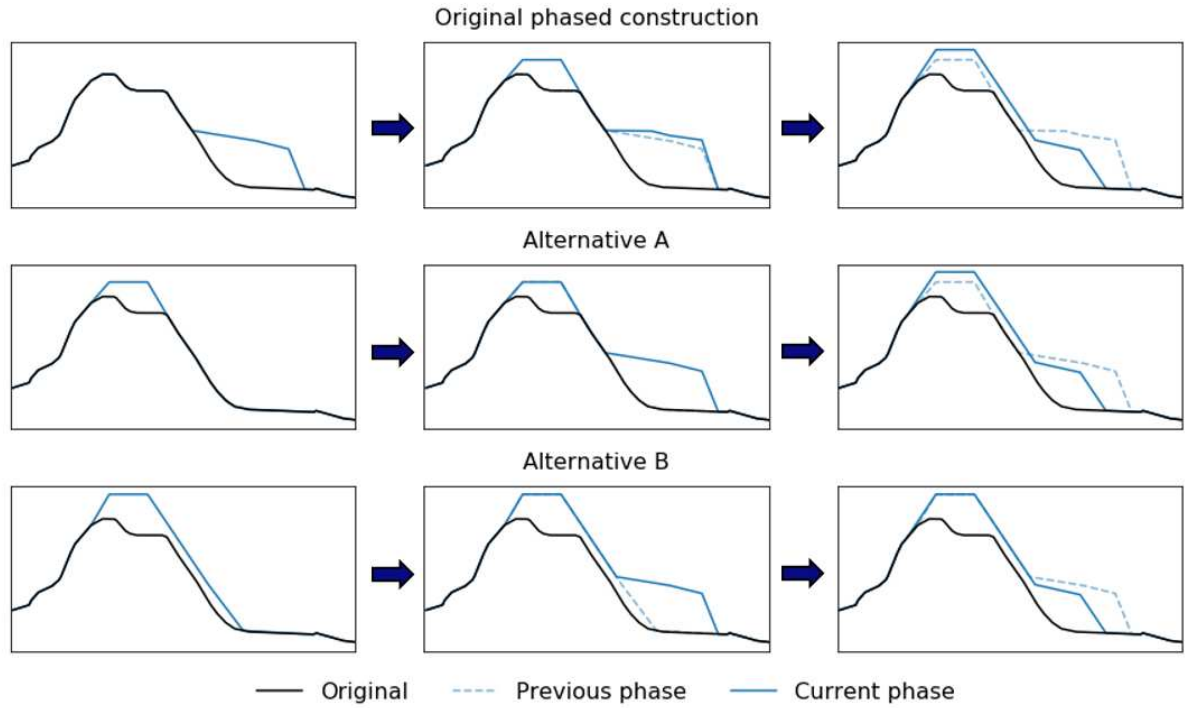


Figure 5-1: Different construction sequences (construction order is from left to right)

5.1.2 Water level

The effect of reliability updating with the survival of an observed water level is investigated in section 5.5. The slope reliability is updated with an extreme water level with a return period of 100 years. It is determined with the extreme water level distribution given in paragraph 4.3.1 and is 3.23 m NAP.

5.2 Reliability update original case

The reliabilities that result from updating with the original construction phases are presented in table 5-1, as are the semi-probabilistic safety factors. SF_{obs} is the safety factor of the critical slip plane in the observation, and $SF_{obs,fixed}$ is the safety factor of the observation using the critical slip plane from the assessment situation. The reliabilities and design points resulting from FORM analyses are presented in table 5-2. Note that these are different from the FORM analyses results shown in table 4-10, which used metamodels trained with a free slip plane search. The results shown here follow from FORM analyses performed on the metamodels trained with the critical slip planes from the assessment situation, as elaborated in section 3.3. The reliability update using construction survival turns out to have no effect in the original case. The least reliable observed situation is the daily situation. The limit state function in the daily and assessment situation are plotted in figure 5-2 for comparison. The limit state outcome of the assessment situation is plotted in the background as a reference.

Survived load condition	$\boldsymbol{\beta_{obs}}$	$\boldsymbol{\beta_{obs,fixed}}$	$\boldsymbol{\beta_{posterior}}$	$\boldsymbol{SF_{obs}}$	$\boldsymbol{SF_{obs,fixed}}$
berm construction	6.7	8.2	6.0	1.14	1.21
crest construction	6.9	7.6	6.0	1.09	1.12
daily conditions	6.1	6.1	6.0	1.05	1.05

Table 5-1: Reliability updating results original case with ERRAGA ($\beta_{prior} = 6.0$)

Phase	Berm phase		Crest phase		Daily conditions		Assessment situation	
FORM reliability	$\beta_{FORM} = 8.1$		$\beta_{FORM} = 7.5$		$\beta_{FORM} = 6.1$		$\beta_{FORM} = 5.9$	
Parameter	$\mathbf{U}_{\text{design}}$	$\boldsymbol{\alpha^2}$	$\mathbf{U}_{\text{design}}$	$\boldsymbol{\alpha^2}$	$\mathbf{U}_{\text{design}}$	$\boldsymbol{\alpha^2}$	$\mathbf{U}_{\text{design}}$	$\boldsymbol{\alpha^2}$
Water level	0.09	0.00	0.18	0.00	0.02	0.00	0.39	0.00
POP below dike	-3.32	0.17	-2.81	0.14	-1.77	0.09	-1.78	0.09
POP hinterland	-3.78	0.22	-3.03	0.16	-2.31	0.15	-2.15	0.13
Model factor	4.37	0.29	4.48	0.36	4.04	0.45	3.85	0.43
S silty clay	-1.34	0.03	-3.99	0.28	-2.96	0.24	-2.94	0.25
m silty clay	-0.19	0.00	-0.12	0.00	-0.73	0.02	-0.92	0.02
S peat clayey	-0.88	0.01	-0.95	0.02	-0.75	0.02	-0.66	0.01
m peat clayey	-0.20	0.00	-0.34	0.00	-0.25	0.00	-0.31	0.00
S clay silty organic	0.09	0.00	-1.30	0.03	-1.09	0.03	-0.98	0.03
m clay silty organic	-0.31	0.00	-0.13	0.00	-0.22	0.00	-0.18	0.00
S organic clay	-4.01	0.25	-0.09	0.00	-0.18	0.00	-0.21	0.00
m organic clay	-0.85	0.01	-0.11	0.00	-0.19	0.00	-0.01	0.00
S peat	-1.28	0.03	-0.86	0.01	-0.84	0.02	-0.82	0.02
m peat	-0.22	0.00	-0.14	0.00	-0.04	0.00	0.03	0.00

Table 5-2: Results FORM analyses original case

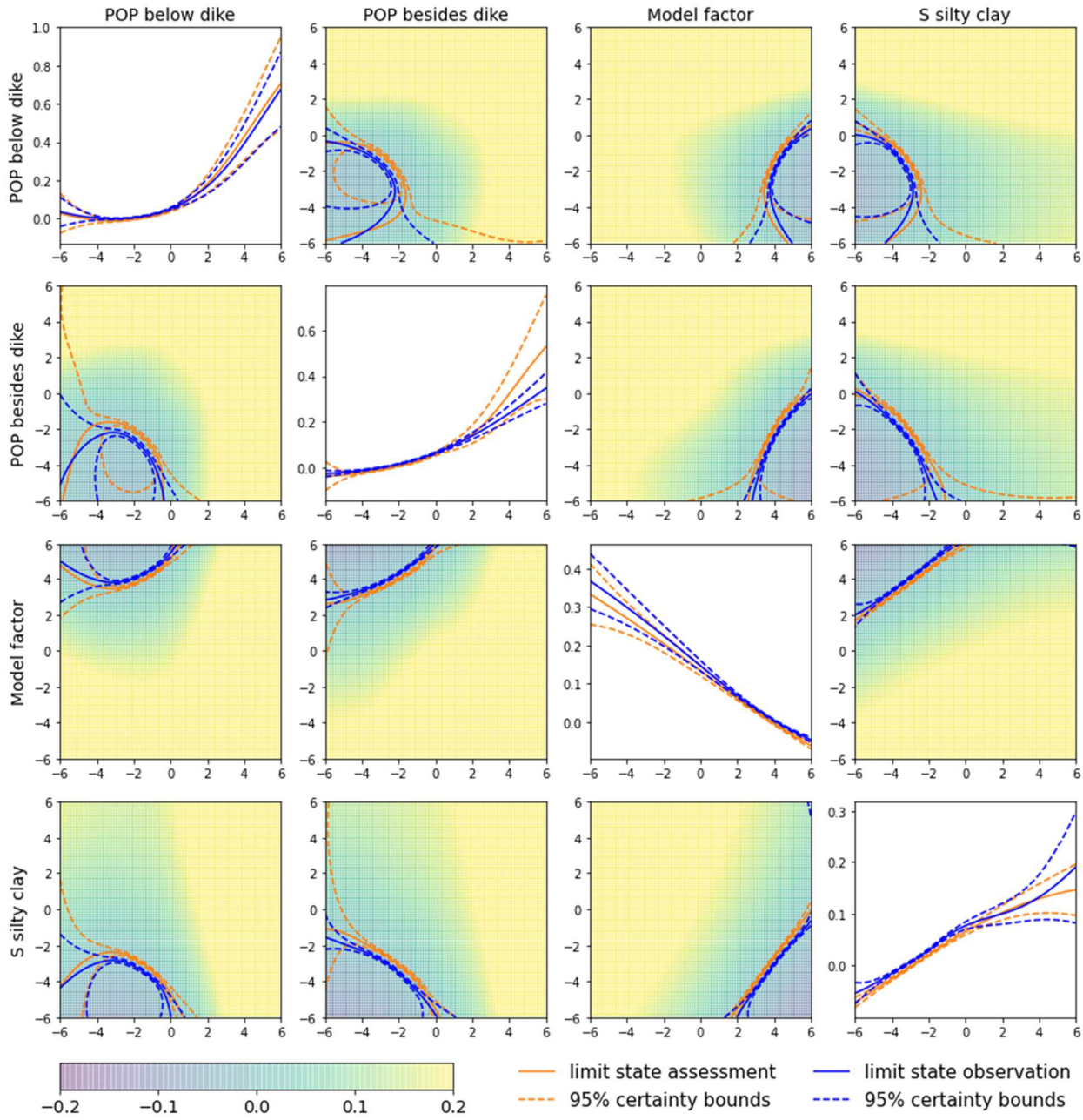


Figure 5-2: Comparison of limit state functions in the daily conditions and assessment situation for the original case

5.3 Reliability update alternative A

The reliability updating results using survival of the crest construction in alternative A are presented in table 5-3. The results from the FORM analyses using the metamodels are shown in table 5-4 and the comparison of the metamodels in the observed and assessment situation is presented in figure 5-4. The main sliding modes of the observed and assessment situation are shown in figure 5-3.

Survived load condition	β_{prior}	β_{obs}	$\beta_{obs,fixed}$	$\beta_{posterior}$	SF_{obs}	$SF_{obs,fixed}$
crest construction	6.0	4.3	5.6	6.1	0.98	1.04

Table 5-3: Reliability updating results alternative A with ERRAGA

Phase	Crest phase		Assessment situation	
FORM reliability	β_{FORM} = 5.5		β_{FORM} = 5.9	
Parameter	$\mathbf{U}_{\text{design}}$	$\boldsymbol{\alpha^2}$	$\mathbf{U}_{\text{design}}$	$\boldsymbol{\alpha^2}$
Water level	0.01	0.00	0.39	0.00
POP below dike	-2.58	0.22	-1.78	0.09
POP hinterland	-2.98	0.29	-2.15	0.13
Model factor	3.26	0.35	3.85	0.43
S silty clay	-1.16	0.04	-2.94	0.25
m silty clay	-0.14	0.00	-0.92	0.02
S peat clayey	-0.61	0.01	-0.66	0.01
m peat clayey	-0.25	0.00	-0.31	0.00
S clay silty organic	-0.92	0.03	-0.98	0.03
m clay silty organic	-0.38	0.01	-0.18	0.00
S organic clay	-0.85	0.02	-0.21	0.00
m organic clay	-0.61	0.01	-0.01	0.00
S peat	-0.76	0.02	-0.82	0.02
m peat	-0.24	0.00	0.03	0.00

Table 5-4: Results FORM analyses alternative A

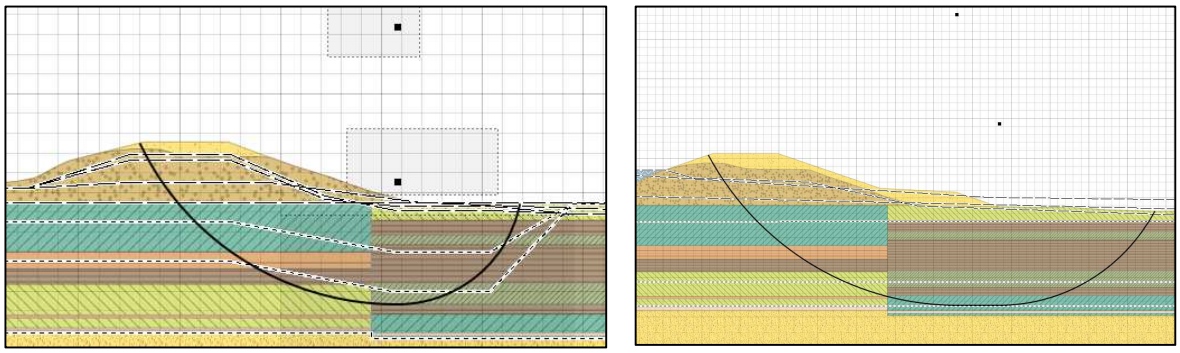


Figure 5-3: Main sliding modes of the crest phase (left) and the assessment situation (right)

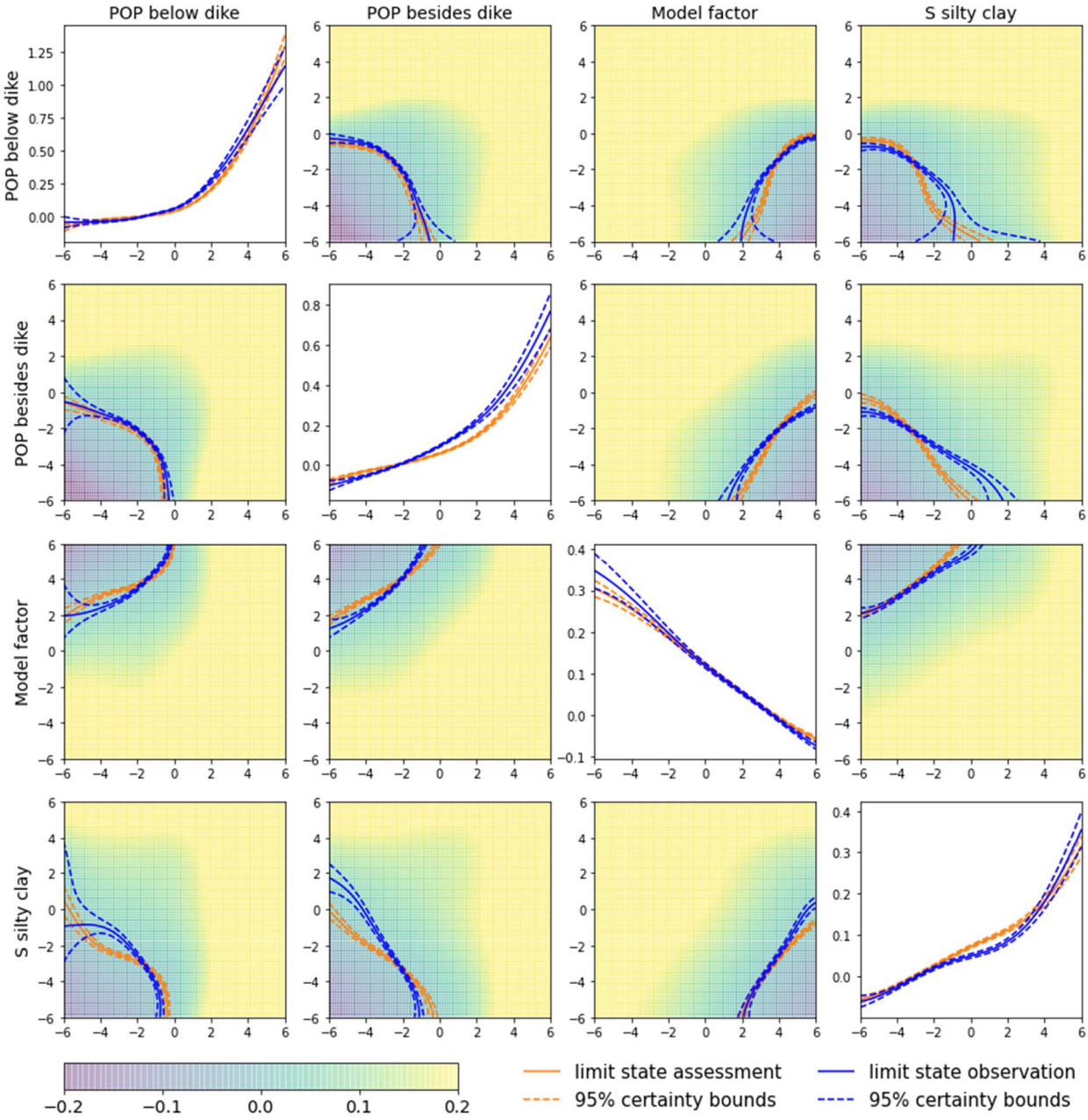


Figure 5-4: Comparison of limit state functions in the observation and assessment situation for alternative A

5.4 Reliability update alternative B

The reliability updating results using survival of the crest construction in alternative B are presented in table 5-5. Updating the reliability with construction survival of the crest has significantly improved the reliability. To see the influence of the excess pore water pressure, an update was also done with a consolidated variant of the crest construction. The consolidated case still causes a significant update. Table 5-8 shows the results from the FORM analyses using the fitted metamodels and a comparison of the metamodels in the observed and assessment situation is presented in figure 5-6, both for the unconsolidated case.

Survived load condition	β_{obs}	$\beta_{obs,fixed}$	$\beta_{posterior}$	SF_{obs}	$SF_{obs,fixed}$
crest construction	3.1	4.0	8.5	0.92	0.96
crest construction consolidated	3.3	4.7	7.5	0.95	0.99

Table 5-5: Reliability updating results alternative B with ERRAGA ($\beta_{prior} = 6.0$)

To find out more about the reliability update, several analyses were done with the metamodels. First, the posterior parameter distributions of the four most important parameters are presented together prior distributions in figure 5-5. The posterior distributions are obtained by taking a sample set and creating histograms of only the data consistent with the observation, thus construction survival. Then, for insight in the implausible part of the parameter space, the correlation matrix of samples causing failure in the observation are shown in table 5-8. Finally, a scatter plot in figure 5-7 is presented to obtain insight in the prior and posterior failure region of the assessment situation. Determining the reliabilities is done with a Gaussian mixture model as the importance distribution. However, using sampling data from the mixture model for this figure would result in a messy plot, providing little information on the failure domain and how it developed after updating. Therefore, a multivariate Gaussian distribution was fitted to the prior and posterior samples causing failure in the assessment situation. This distribution was used for creating the sample sets presented in figure 5-7. So, the figure gives a representation of where in the parameter space failure occurs, but the actual shape of the failure domain is different.

The mean value of a failure region is an estimate for the center of gravity of the failure distribution, the point in the parameter space around which the probability density of failure is evenly distributed. This point is quite like a design point from a FORM analysis, which is the point on the limit state line with the highest probability density. The prior FORM design points and prior and posterior center of gravity design points are presented in table 5-7. It shows that the prior FORM and center of gravity design points are very alike. It also shows that after the update, previously unimportant parameters start to play an important role. A scatter plot of the prior and posterior domains of failure of all the 14 variables is given in appendix E.

Survived load condition	Crest phase		Assessment situation	
FORM reliability	β_{FORM} = 4.0		β_{FORM} = 5.9	
Parameter	$\mathbf{U_{design}}$	$\boldsymbol{\alpha^2}$	$\mathbf{U_{design}}$	$\boldsymbol{\alpha^2}$
Water level	0.01	0.00	0.39	0.00
POP below dike	-1.85	0.21	-1.78	0.09
POP hinterland	-2.27	0.32	-2.15	0.13
Model factor	2.27	0.32	3.85	0.43
S silty clay	-0.94	0.06	-2.94	0.25
m silty clay	-0.08	0.00	-0.92	0.02
S peat clayey	-0.48	0.01	-0.66	0.01
m peat clayey	-0.24	0.00	-0.31	0.00
S clay silty organic	-0.44	0.01	-0.98	0.03
m clay silty organic	-0.11	0.00	-0.18	0.00
S organic clay	-0.75	0.04	-0.21	0.00
m organic clay	-0.32	0.01	-0.01	0.00
S peat	-0.55	0.02	-0.82	0.02
m peat	-0.03	0.00	0.03	0.00

Table 5-6: Results FORM analyses alternative B

Design point type	FORM	Centre of gravity	Centre of gravity
Parameter	**Prior $\mathbf{U_{desing}}$**	**Prior $\mathbf{U_{desing}}$**	**Posterior $\mathbf{U_{desing}}$**
Water level	0,39	0,38	0,76
POP below dike	-1,78	-1,80	-0,37
POP hinterland	-2,15	-2,29	0,19
Model factor	3,85	4,08	5,66
S silty clay	-2,94	-2,85	-3,48
m silty clay	-0,92	-0,86	-0,51
S peat clayey	-0,66	-0,75	-0,23
m peat clayey	-0,31	-0,41	-1,16
S clay silty organic	-0,98	-0,95	0,49
m clay silty organic	-0,18	-0,23	-1,31
S organic clay	-0,21	-0,40	-3,19
m organic clay	-0,01	-0,31	-3,20
S peat	-0,82	-0,83	-1,42
m peat	0,03	-0,02	-0,38

Table 5-7: Prior and posterior design points of the assessment situation of alternative B

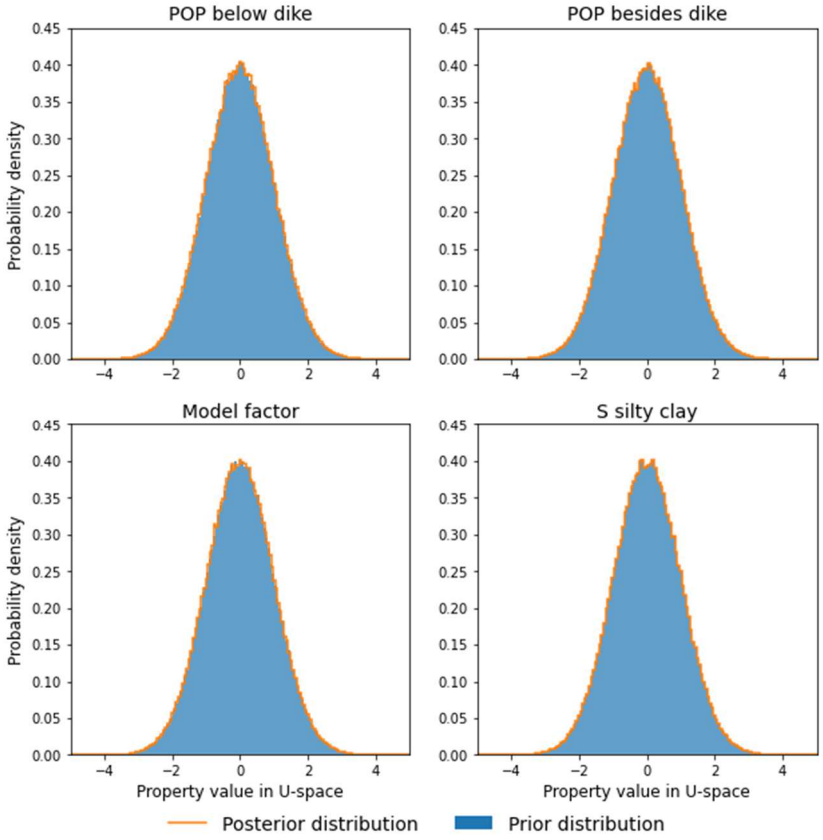


Figure 5-5: Prior and posterior distributions of the four important parameters

	Water	POP below	POP besides	Model factor	S silty clay	m silty clay	S peat clayey	m peat clayey	S clay silty organic	m clay silty organic	S organic clay	m organic clay	S peat	m peat
Water	1	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
POP below		1	-0,16	0,31	-0,14	-0,07	-0,07	-0,02	-0,05	-0,02	-0,07	-0,02	-0,04	0,00
POP besides			1	0,47	-0,11	-0,04	-0,09	-0,07	-0,08	-0,03	-0,08	-0,07	-0,10	0,00
Model factor				1	0,20	-0,01	0,07	0,03	0,11	0,01	0,11	0,05	0,12	0,01
S silty clay					1	0,06	-0,06	-0,03	0,01	-0,01	-0,25	-0,07	-0,05	0,00
m silty clay						1	0,01	0,00	0,02	0,00	-0,06	-0,01	0,00	0,00
S peat clayey							1	0,00	-0,01	0,00	-0,01	-0,01	-0,02	0,00
m peat clayey								1	0,00	0,00	-0,01	0,00	0,00	0,00
S clay silty organic									1	0,00	-0,06	-0,02	-0,02	0,00
m clay silty organic										1	-0,01	0,00	0,00	0,00
S organic clay											1	0,04	0,00	0,00
m organic clay												1	-0,01	0,00
S peat													1	0,00
m peat														1

Table 5-8: Correlation matrix disregarded samples alternative B (only top half is given for readability)

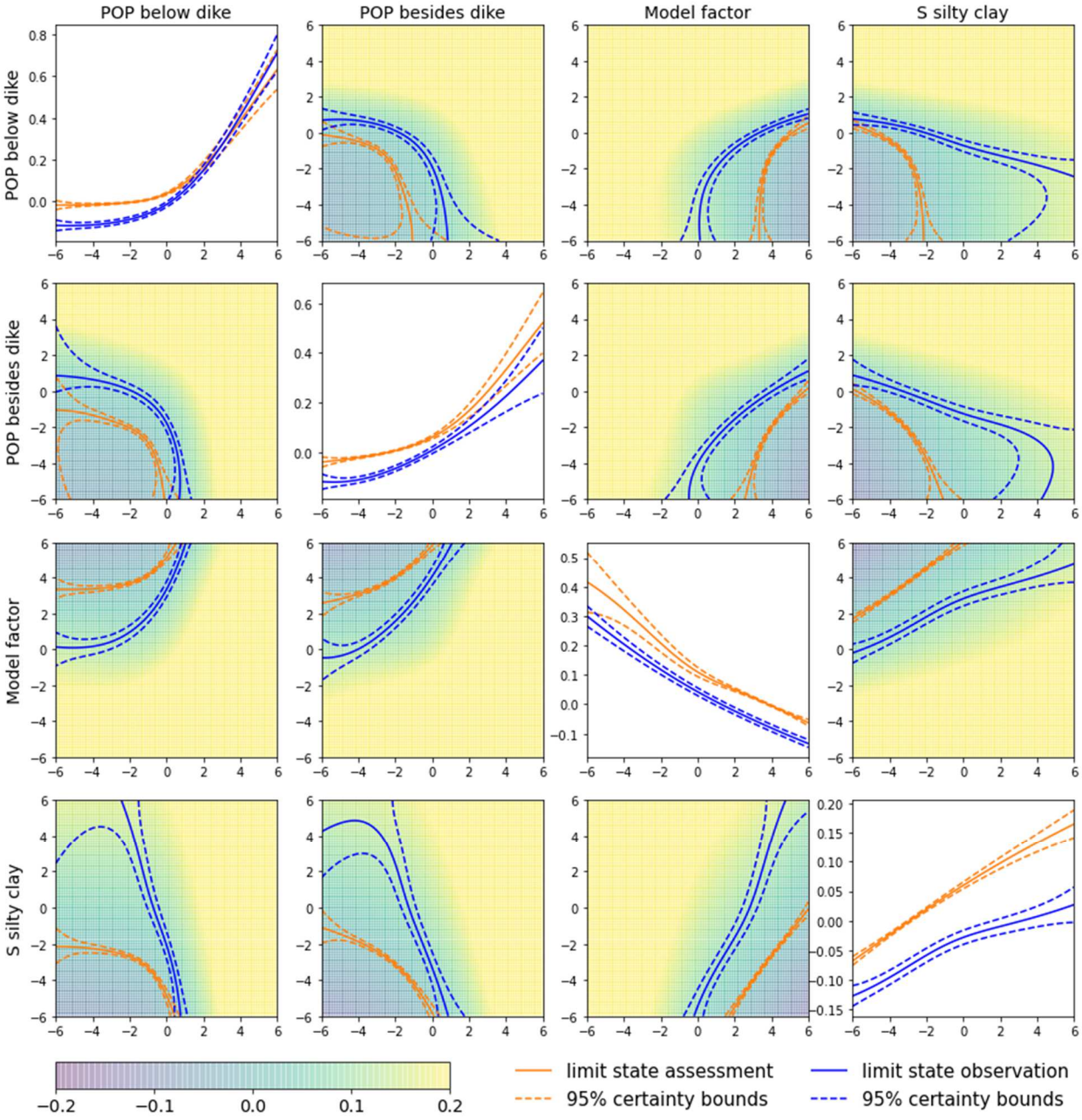


Figure 5-6: Comparison of limit state functions in the observation and assessment situation for alternative B

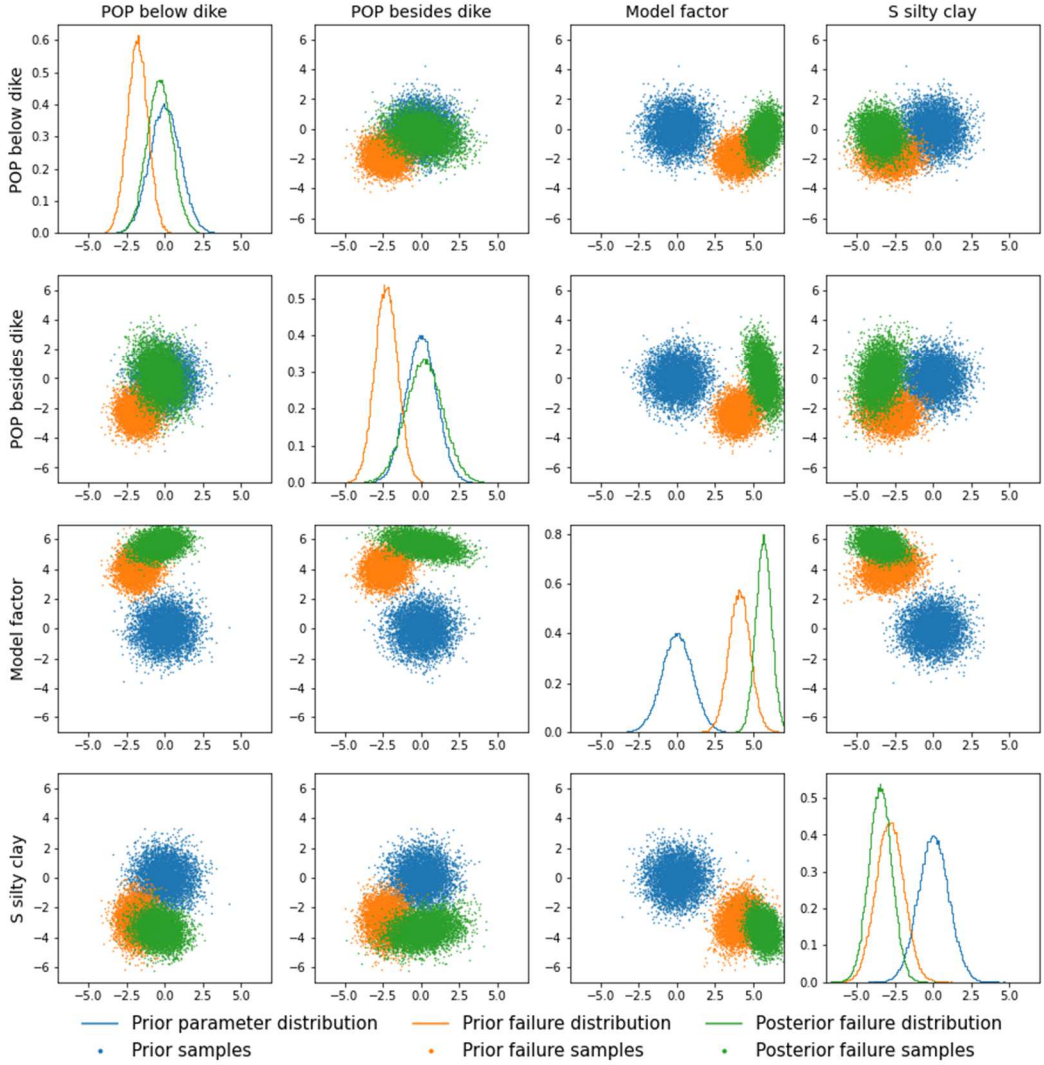


Figure 5-7: Prior parameter distributions and prior and posterior domains of failure

5.5 Reliability update observed water level

Information about the survival of an observed water level could be useful for reliability updating, see for example Schweckendiek, Kanning, et al., (2017). The update with daily circumstances in paragraph 5.2 actually already was an update with an observed water level, which is the daily water level of 0.4 m NAP. Whether observing a more extreme water level is useful for this case study is demonstrated by a reliability update using a water level with a return period of 100 years. It is determined with the extreme water level distribution given in paragraph 4.3.1 and is 3.23 m NAP. The result is presented in table 5-9.

Survived load condition	β_{prior}	β_{obs}	$\beta_{obs,fixed}$	$\beta_{posterior}$	SF_{obs}	$SF_{obs,fixed}$
crest construction	6.0	5.8	5.8	6.3	1.03	1.03

Table 5-9: Result for updating with an observed water level using ERRAGA

5.6 Overview updating results

All updating results are collected in table 5-10. Table 5-11 gives an overview of the FORM analyses. The presented FORM results are for the metamodels trained with the critical slip plane from the assessment situation.

Survived load condition	β_{obs}	$\beta_{obs,fixed}$	$\beta_{posterior}$	SF_{obs}	$SF_{obs,fixed}$
original case: berm phase	6.7	8.2	6.0	1.14	1.21
original case: crest phase	6.9	7.6	6.0	1.09	1.12
original case: daily conditions	6.1	6.1	6.0	1.05	1.05
original case: 1/100-year water level	5.5	5.5	6.3	1.03	1.03
alternative A: crest phase	4.3	5.6	6.1	0.98	1.04
alternative B: crest phase	3.1	4.0	8.5	0.92	0.96
alternative B: crest phase, consolidated	3.3	4.7	7.5	0.95	0.99

Table 5-10: Overview of reliability updating results ($\beta_{prior} = 6.0$)

Case	Original			Alternative A	Alternative B	All
Load condition	Berm phase	Crest phase	Daily conditions	Crest phase	Crest phase	Assessment situation
FORM reliability	8.1	7.5	6.1	5.5	4.0	5.9
Parameter	α^2	α^2	α^2	α^2	α^2	α^2
Water level	0.00	0.00	0.00	0.00	0.00	0.00
POP below dike	0.17	0.14	0.09	0.22	0.21	0.09
POP hinterland	0.22	0.16	0.15	0.29	0.32	0.13
Model factor	0.29	0.36	0.45	0.35	0.32	0.43
S silty clay	0.03	0.28	0.24	0.04	0.06	0.25
m silty clay	0.00	0.00	0.02	0.00	0.00	0.02
S peat clayey	0.01	0.02	0.02	0.01	0.01	0.01
m peat clayey	0.00	0.00	0.00	0.00	0.00	0.00
S clay silty organic	0.00	0.03	0.03	0.03	0.01	0.03
m clay silty organic	0.00	0.00	0.00	0.01	0.00	0.00
S organic clay	0.25	0.00	0.00	0.02	0.04	0.00
m organic clay	0.01	0.00	0.00	0.01	0.01	0.00
S peat	0.03	0.01	0.02	0.02	0.02	0.02
m peat	0.00	0.00	0.00	0.00	0.00	0.00

Table 5-11: Overview FORM analyses

5.7 Optimized dike design

The reliability update with the crest construction phase of alternative B turned out to be very effective. It resulted in a large reliability update not accounted for in the design. A higher reliability seems nice, but a dike reinforcement always has negative aspects such as high costs, nuisance for residents nearby and demolishing houses for creating space. From this perspective, it is interesting to see if a less radical dike improvement would be possible with inclusion of the reliability update.

Therefore, three updates are done with a berm of reduced cross-sectional area, with reductions of 50%, 75% and without a berm. The reduction of 50% is applied to the length of the berm, which decreases from 10 to 5 meters. The additional reduction needed for the 75% update was applied to the height of the berm, decreasing from approximately 2 to 1 meter. In the reinforcement, the berm is initially constructed larger than the design, probably to enhance settlement. This has effect on the stress history of the subsoil. Therefore, the cross-sectional area of initial over-dimensioned berm is reduced with the same percentage in each case. The results are presented in table 5-12. The safety factor of the observation using a free slip plane search is 0.92 for each update, as the crest construction phase takes place before constructing the berm. From the updates follow that a berm with just 25% of the cross-sectional area of the original berm provides a similar reliability after updating as the original dike design. The reduced berm design is presented in figure 5-8.

The observation takes place before constructing the berm, which means that reducing the berm dimensions would have been possible during the dike reinforcement. So, the combination of updating with construction survival and adaptation of the dike design has potential to save costs and reduce other negative effects in future dike reinforcements.

Survived load condition	**Berm reduction**	β_{prior}	$\beta_{obs,fixed}$	$\beta_{posterior}$	$SF_{obs,fixed}$	SF
alternative B: crest phase	0%	6.0	4.0	8.5	0.96	1.03
alternative B: crest phase	50%	5.0	3.5	7.4	0.95	1.00
alternative B: crest phase	75%	4.3	3.3	6.1	0.93	0.97
alternative B: crest phase	100%	3.6	3.2	5.0	0.92	0.94

Table 5-12: Reliability updating results for a reduced dike improvement design

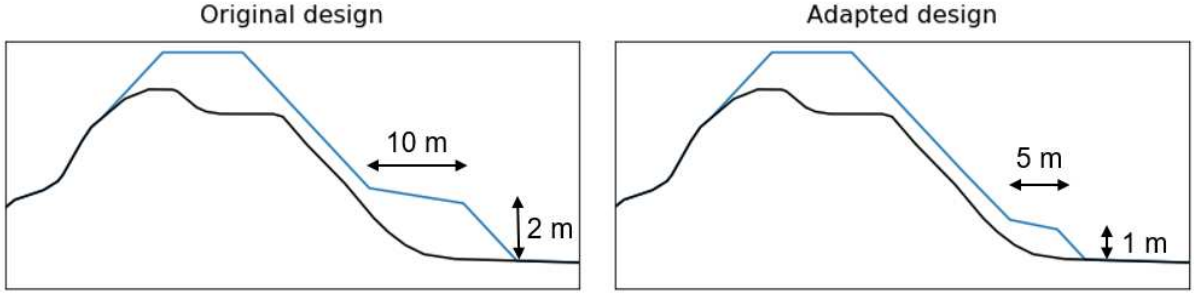


Figure 5-8: Original and adapted dike design

5.8 Discussion

Update original case

The prior reliability analysis already showed that the reliability during the construction phases was greater than in the assessment situation. That the survival of these relatively reliable loading conditions is not effective for reliability updating becomes apparent from the update with the berm and crest construction phase, which showed no effect. Besides the higher reliability of the observation, also the similarity of sliding planes in the observed and assessment situation plays a role. This is due to the assumption that uncertainty can only be updated if the considered slip planes are equal. Therefore, it was chosen to train the observation metamodels with the critical slip planes from the assessment situation (see paragraph 3.2.1 for details). The resulting reliability, $\beta_{obs,fixed}$, shows how critical the observed loading condition was for the slip plane area relevant to the assessment situation. From the prior analysis followed that the main sliding modes of the berm phase and assessment situation are different (figure 4-9). The difference between $\beta_{obs,fixed}$ and β_{obs} of the berm phase shown in table 5-10 now also demonstrates that the loading conditions in the observation are unimportant for the slip plane area of the assessment situation. This dissimilarity of slip planes for the crest phase is a less but still present, as is seen by comparing $\beta_{obs,fixed}$ and β_{obs}.

The situation during daily loading conditions on the other hand, has a reliability like the assessment situation and equal values for $\beta_{obs,fixed}$ and β_{obs}, indicating similarity of the sliding modes. Also, the importance factors are almost identical (table 5-11). For two linear limit state functions in U-space it holds that they are parallel if the importance factors are equal. In a FORM analysis the limit state function is linearized in the design point, meaning that two limit state functions with equal importance factors are at least parallel in the design points. That the similarity of importance factors is an indication for parallelism of limit state functions can be seen in figure 5-2, where the limit state functions of the daily conditions and assessment situation is presented. The lines are quite parallel, with the line for the assessment situation closest to the origin. This is explained by the smaller reliability in the assessment situation. In the two parameter graphs (non-diagonal), the area beyond the limit state line of the observation seen from the origin is the implausible area. The parameter combinations in that area would lead to failure in both the observation and assessment situation and are disregarded when updating the reliability. The area between the limit state lines is thus the area contribution to the updated probability of failure, where only failure in the assessment situation is predicted. Although it seems small, apparently the bulk of probability density is between those lines, as the update is not effective. Note that the limitation of presenting only two of the fourteen probably plays a role here. If the limit state line of the observation would be on the other side of the assessment limit state line, then this would be a very strong case for reliability updating. Then most of the parameter sets

leading to failure in the assessment situation would also lead to failure in the observed situation, resulting in a large update.

Update alternative A

It is concluded that the loading conditions during construction and in the daily conditions are not severe enough to lead to an update in the original case. Therefore, two alternative construction sequences are used to see if a more severe loading condition during construction is effective for updating the reliability. In alternative A, the construction sequence is altered such that a partial crest raise takes place before constructing the berm. The partial crest raise is used as the observed situation. The reliability of the observed situation β_{obs} of alternative A is indeed much lower than in the assessment situation. However, the slip plane region relevant for the assessment situation is not affected severely, as demonstrated by the higher reliability $\beta_{obs,fixed}$, similar to the assessment reliability. This is explained by comparing the dominant sliding modes of the observation and assessment situation in figure 5-3. The sliding part is smaller and shallower than the main mode in the assessment situation. This dissimilarity in important sliding modes causes the update to be insignificant. The update shows a small effect of 0.1 in reliability, which is not big with respect to the uncertainty interval of the metamodels (0.3 at 99% certainty for a reliability around 6). So, this loading condition is not more critical than the assessment situation for the slip plane area relevant to the assessment situation. Using the observed construction survival for updating does not have a significant effect.

The update with the daily conditions in the original case showed that similar importance factors are an indication for parallel limit state functions. The importance factors of the update of alternative A are not very similar, as can be seen in table 5-11. The importance of the POP values is smaller in the assessment situation, while the shear strength ratio of silty clay is much more important. The difference in importance of the parameters is partially explained by the slip plane dissimilarity, but also due to the loss of stress history caused by raising the dike, as discussed in the prior reliability analysis (paragraph 4.4.6). The dissimilarity of importance factors can be seen in the plot of the limit state lines in figure 5-4. The limit state lines as a function of both the POP variables are still parallel, but the remaining parameter plots are not and have crossing points.

Update alternative B

In alternative B the construction phasing is altered such that the total crest is raised before constructing the berm. Updating with the crest construction phase shows to be very effective, with the reliability increasing from 6.0 to 8.5. Not only β_{obs}, but also $\beta_{obs,fixed}$ is significantly lower than the prior reliability, which was not the case in alternative A. The higher crest in the observation adds extra driving force to the failure mechanism. The smaller sliding mode is still most important for the observation, but the extra driving force causes the deeper sliding mode to become relevant too, as is seen from a lower $\beta_{obs,fixed}$.

Analyzing updated parameter distributions could provide some insight into the origin of the update. However, the updated distributions shown in figure 5-5 cannot point out one or more parameters that were of great influence solely. Only four distributions are shown but the other distributions are not visually altered as well. Possibly, the update is mainly caused by the exclusion of combinations of unfavorable parameter values. Therefore, the correlation matrix of disregarded samples is presented table 5-8. It shows that combinations of low values of the POP, shear strength ratio of silty clay and high values of the model factor are correlated, with absolute correlations of 0.11 to 0.47. It is not a surprise that the updating mainly takes place through these four variables, as the design point from the prior reliability analysis showed these variables to be most important (see table 5-11). Although, they are not the only contributors. The rest of the parameters also show some correlation with the important four, indicating that it really is a combination of unfavorable parameters.

Updating the reliability is reducing uncertainty. And, apparently in this case, uncertainty about the occurrence of parameter combinations is reduced. It results in a change in the joint probability distribution, which might affect the influence of parameters on the slope reliability. To see a possible change of influence, the prior and posterior center of gravity design points are determined and presented in table 5-7. It shows that the design point changed significantly after updating. The values for the POP clearly become less important, while the model factor and the shear strength ratio of silty clay grow in importance. Also, the shear strength parameters of organic clay and peat significantly grow in importance. A visual representation of the shift in design point is given in figure 5-7, in which two phenomena are observed. For the POP variables, the failure distributions shift towards the prior mean and become almost the same shape. The correlation matrix of implausible parameter sets in table 5-8 shows correlation for unfavorable combinations of the POP with other variables. From the shift in failure distribution is concluded that this causes the POP to lose its dominant role in causing slope instability. Where a-priori failure was most likely for low percentiles of the POP, now failure at a certain POP value is almost as likely as the occurrence of that value, showing the indifference of model outcome to the POP. So, the first phenomenon is a

shift of the design point towards the parameter prior mean due to the reduction of uncertainty. The second phenomenon is a shift of the design point away from the mean due to the increased relative influence of certain parameter sets. This is the case for the model factor and the shear strength ratio of silty clay in figure 5-7. This part of the parameter space contributed little to the prior probability of failure because of its low probability of occurrence. But, due to updating, this part becomes relevant.

Update with a water level

The reliability update with a 1/100-year water level resulted in an update of 0.3 in terms of the reliability index. So, there is an effect, but not a very big one considering that a 1/100-year water level is quite a unique observation. From the prior reliability analysis followed a squared importance factor close to zero for the water level, indicating its limited influence. A large updating effect was therefore not expected.

The role of pore water pressure

At forehand, the excess pore water pressure was assumed to be a critical factor for updating the reliability. To see the sensitivity of the update to the excess pressure, an update was done with the crest construction phase of alternative B, without the excess pressure. It turns out that the reliability index still increased from 6.0 to 7.5. It shows that besides the excess pressure, the increased driving load is of major importance for updating the reliability. Although the combination with excess pore water pressure is most effective, updating the reliability to 8.5. This however does show that exact knowledge of the extreme excess pressure situation is not a strict requirement. In this case study the excess pore water pressure was treated as a deterministic variable. However, uncertainty always plays a role due to scarcity of data or measurement inaccuracy. Doing a cautious estimate of the excess pressure or including uncertainty in the schematization would however still cause a significant update, with a reliability between 7.5 and 8.5.

FORM reliability

For each loading condition the reliability was also determined with a FORM analysis, applied to the fitted metamodels. Comparing the reliabilities in table 5-10 with those in table 5-11 shows that the difference in reliability is 0.1 at most. That makes FORM a useful estimator of the reliability. Furthermore, it can serve as an indication for the effectiveness of a reliability update, as the reliability in the observation $\beta_{obs,fixed}$ compared to the reliability in the assessment situation turned out to be one of the key indicators. The use of FORM for estimating whether an update would have effect is useful because it is a widely used well know reliability method, while metamodelling can be omitted initially.

Factors of safety

FORM showed to be good for indication if some observation would be useful for reliability. An even more straightforward indication is given by the factors of safety from semi-probabilistic analyses. Table 5-10 shows that the loading conditions with a lower safety factor than in the assessment situation are effective for reliability updating, using the critical slip plane from the assessment in both situations ($SF_{obs,fixed}$). The safety factor in the assessment situation is 1.03. Table 5-10 shows that the load cases with safety factors 1.04 and 1.05 are not effective for reliability updating, while the load case with a safety factor of 1.03 shows an update of 0.3 in the reliability index. So, the safety factor in the observed situation with the critical slip plane from the assessment, is in this case a very good indication for the effectivity of a reliability update.

Failure probability during construction

In this case study, the construction sequence was altered to create a more severe observed situation in terms of stability. Updating with this alternative construction sequence showed to be very effective. The reliability of the observed situation is in the order of 3 in reliability index, or 1/1000 in terms of failure probability, which makes survival of this loading condition plausible. However, there are two considerations for creating a severe loading condition for sake of a reliability update, which are the flooding probability during construction and the cost efficiency.

Deliberately creating a loading condition with a decreased reliability conflicts with the purpose of the dike reinforcement and the dike itself. So, it needs to be considered whether such a loading condition is acceptable from the perspective of flood risk. The computed reliability of 3 however concerns the occurrence of a slope instability, and not the occurrence of a flooding event. The flooding reliability can be determined by considering the event that the water level exceeds the residual water retaining height after an instability, within the period until the dike is repaired. This reliability may already be significantly different from the slope reliability. The flooding reliability can be increased by constructing the critical step in the summer season, when high water levels are less likely. If additionally, preparations are done for emergency repairs in case of an instability to reduce the flooding opportunity window, then the flooding reliability during construction can be increased further. This way, creating a more severe loading condition may be acceptable from the flood risk point of view.

The second consideration in creating a severe loading condition is the cost efficiency. Incorporating construction survival in a dike improvement design has a financial benefit, as shown in section 5.7. However, it also has a financial downside. Increasing the failure probability during construction increases the probability that a repair is needed, thus increasing the costs expectation. Of course, the expected benefits should exceed the increase in expected costs. The cost efficiency can be assessed and optimized by

considering multiple loading conditions in a decision tree framework for cost benefit analysis, considering all costs like failure costs or additional monitoring efforts (van der Krogt et al., 2020).

ERRAGA

ERRAGA has shown to be suited for modelling the slope stability behavior of a dike using LEM. In this case study with a typical reliability index of six, fourteen variables were treated as stochastic. Training a metamodel for this problem with the chosen convergence criterion takes about two to four hours with approximately 300 model evaluations, using a computer with 16GB of installed memory (RAM) and a 3.5GHz processor. For performing a reliability update, two models must be trained, taking about six to ten hours. More than double the time is needed compared to training a single model because of the growing size of the training set. Initially, the slope stability calculations take up most of the time, with fifteen seconds each. But the time needed for fitting the metamodel grows when expanding the training set, taking about two min for a training set of 300 model evaluations, including the slope stability calculation. Addressing this problem with Monte Carlo simulation would require $4 \cdot 10^{11}$ realizations (equation (7) with $CoV = 0.05$), resulting in a computation time of thousands of years. This can of course be enhanced by using more advanced simulation techniques and parallel processing, but clearly demonstrates the added value of metamodelling for reliability problems like these. An overview of the mentioned numbers on training the metamodels is presented in table 5-13. Some experiences regarding the use of ERRAGA and the influence of ERRAGA's settings is presented in appendix A.

Characteristic	**Value**
Number of model evaluations single model (on average)	300
Single model training time (on average)	2-4 hours
Double model training time (on average)	6-10 hours
Duration single slope stability analysis	15 seconds

Table 5-13: Model training statistics

6

Conclusion

In this thesis, an approach to reliability updating using construction survival is presented and applied to a case study dike improvement. Based on the reliability update, an optimized dike design is presented. This chapter presents the main findings of this research with which the research questions are answered. Also, some points of discussion are presented, as well as several recommendations for improving the developed methodology.

6.1 Findings

The objective of this thesis is to update dike reliability with construction survival using metamodelling. With this objective, the main research question was formulated in chapter 1, along with three sub-questions. This section answers the sub-questions, after which an answer to the main question is presented.

What are the characteristics of a dike improvement leading to a significant reliability update with survival of the construction?

One of the two most important characteristics for an effective reliability update is the similarity of probable slope failure modes in the observation and after improvement. The total probability of failure after improvement generally consists of contributions from various sliding modes. If the stability of an important sliding mode reaches a critical level during improvement and survival is observed, then this observation might carry information on the strength involved in that sliding mode. Updating the reliability with this loading condition could therefore reduce the failure probability. The second important characteristic is the level of severity of the observed loading condition. The severity of a loading condition is usually expressed by the reliability. However, only sliding modes significantly contributing to the failure probability after improvement are relevant for the effectivity of the update. If was found that the reliability of the observed situation, determined by only considering the important sliding modes after improvement, is a good indication for how severe the observation was for the sliding modes of interest. If this reliability is lower than the reliability after improvement, then the loading condition is considered to be severe and an update is expected to have effect.

The difference in influence of the stochastic parameters on the stability in the observed and assessment situation also plays a role in the effectivity of the update, although this characteristic is less apparent. It was found that the similarity of FORM importance factors in the observed and assessment situation indicates parallel limit state lines. And moreover, that a reliability update is very effective if the limit state lines are parallel and the reliability in the observed situation is lower. When the limit state functions are parallel, but the observed situation is more reliable, then no significant effect is expected from updating. In the case that the limit state lines are not parallel, and reliabilities are similar, then no decisive indication can be given. Although, the lower the reliability in the observation, the more likely an effective update is. So, using the FORM reliability and the importance factors as a proxy for parallelism, an indication can be given about the effectivity of a reliability update.

Yet, a more straightforward indication for a significant reliability update using construction survival is obtained by looking at the semi-probabilistic factors of safety. The case study showed that if in both the observed and assessment situation the critical sliding plane from the assessment is used, then reliability updating is effective if the safety factor of the observed situation is lower.

What is the effect of reliability updating with construction survival compared to reliability updating with an observed water level?

The effect of updating with a water level depends on how important the water level is for stability and how extreme the observed water level was. In the case study, it was shown that for a situation in which the water level is of minor influence, updating with an extreme water level has a limited effect. The observed extreme water level did not cause a severe loading condition in terms of stability. For the same situation, a significant effect could be achieved by updating with construction survival. So, in this situation the effect of updating with construction survival is much bigger. Part of the explanation is the severity of the observed situation, as discussed in the previous sub-question. But additionally, the amount of unreducible uncertainty also plays a role. The uncertainty in the water level cannot be reduced, as it is not time-invariant. If the failure probability is mainly governed by the uncertainty of the water level, then updating is not expected to be effective. This can only be the case if the water level is important for the slope stability situation. So, for situations in which the water level is less important, possibly the failure probability is mainly determined by parameters with reducible uncertainty, like soil parameters. Situations with a limited influence of the water level are thus good candidates for reliability updating with construction survival.

However, situations in which an extreme water level creates a severe loading condition, and in which a large part of the failure probability is governed by reducible uncertainty, updating with a water level might be more effective. Also, situations in which only minor changes are done to the dike profile might not cause a severe loading condition during construction, resulting in no effect for reliability updating. The relative effect of reliability updating with a water level versus construction survival will therefore differ from case to case, depending on the influence of the water level, how extreme the observed water level was in terms of stability and how critical the stability situation during improvement was.

How can metamodelling be used for reliability updating with past performance regarding slope stability of dikes?

The case study in this thesis shows that ERRAGA can model a full probabilistic slope stability problem. Furthermore, the metamodels in the case study show logical behavior and uncertainty regarding the model outcome can be reduced to an acceptable level. The developed methodology with two regression metamodels, for the observed situation and final situation, has shown to be successful. Metamodelling makes reliability updating of slope stability problems feasible, reducing the analysis time from at least days to approximately 10 hours, compared to an importance sampling method.

The reduction in computational time is large because using a sampling method would involve many realizations, each requiring a (relatively) time-consuming model evaluations (i.e. slope stability calculations). However, expensive model evaluations leading to large computational times is not unique to slope stability problems. Examples of problems in which this also plays a role are the reliability assessment of quay walls and sheet piles and probabilistic settlement predictions. Metamodelling is therefore interesting for other fields as well.

What is the effect of reliability updating with construction survival of a dike improvement?

From answering the sub-questions, it became clear that the effect of reliability updating using construction survival is dependent on a couple conditions. If those conditions are right, then a reliability update can decrease the estimate for the probability of failure orders of magnitude, as shown by the case study of this research. However, the case study also shows that the approach to assessing construction stability in practice leads to situations not effective for reliability updating. Though adjusting the sequence of construction activities led to a significant update. This makes designing the construction phasing with reliability updating in mind promising for future dike reinforcements. In the case study, it was found that adapting the dike improvement design after survival

of the critical construction step resulted in a reduction from 10 to 5 meters in length and from 2 to 1 meter in height of the berm to be constructed, while achieving the same reliability as a priori. So, reliability updating can improve the reliability in hindsight. But integrating an update in the dike design and adapting the design during construction can have effect beyond improving the reliability only. This method for adapting the dike design has the potential to reduce construction costs, limit additional space occupation and reduce other negative impacts of a dike reinforcement.

6.2 Discussion

This section goes into some points of discussion regarding the methodology and findings from the case study. Recommendations are formed from these points and presented in the next section, along with several other recommendations.

In the case study, the construction sequence was altered to create a more severe observed situation in terms of stability. The hypothetical survival of this alternative construction sequence showed to be very effective for reliability updating. However, deliberately creating a loading condition with a decreased reliability seems to conflict with the purpose of the dike reinforcement and the dike itself. However, it was reasoned that a slope instability does not necessarily causes a flooding event. For a flooding to occur the water level must exceed the residual water retaining height after the instability, within the period until the dike is repaired. Measures like constructing the critical step in the summer and including the seasonality of the water level will limit the flooding probability. Additionally, preparing an emergency repair strategy to reduce the repair time and thus the period of reduced strength will limit the flooding probability further. This way, creating a more severe loading condition may be acceptable from the flood risk point of view.

Another point of discussion is the excess pore water pressure. At forehand, the excess pressure was assumed to be a critical parameter for updating the reliability. However, in the case study it was found that a loading condition with an increased driving load can solely result in a significant reliability update. Therefore, in this case exact knowledge of the pore water pressure situation is not strictly necessary, as a cautious estimate or inclusion of uncertainty would still cause a significant update. This is beneficial for future adaptive dike designs, as a dense network of pressure gauges for monitoring the head may be a limitation for construction activities. However, it is noted that in the case study the water level was barely of influence on the stability. The effect of excess pressures on reliability updating may be different for situations in which the water level is important. A loading condition with excess pore pressure is to some extent similar to a high water situation. In both conditions the shear strength is decreased due to an increase in pressure. So, for situations in which the water level is important, the amount of excess pressure may be important too. The influence of excess pore water pressures may thus differ from case to case and should be examined thoroughly when updating the reliability with construction survival.

6.3 Recommendations

In this section several recommendations are done regarding the improvement of the methodology and further research. Also, an approach is presented for indicating the effectiveness of a reliability update with construction survival, without metamodelling.

This research has shown that reliability updating with construction survival is promising for future dike reinforcement. Integration of reliability updating in a dike design can reduce the final dimensions of the improvement needed for achieving the reliability requirement. From answering the first sub-quest it followed that results from FORM analyses and semi-probabilistic analyses provided an indication for whether reliability updating with construction survival would be effective. Furthermore, it was found that FORM is quite an accurate estimator of the reliability compared to estimates using metamodelling. This allows for exploring the potential of incorporating construction survival and for making preliminary dike designs, without using metamodelling. This is very useful for engineers in practice, as FORM is a well-known and widely used reliability method.

Therefore, the approach below is proposed for creating a loading condition during construction that is effective for reliability updating. It uses the findings from answering the first sub-question. If the analysis concerns a past observation, then step 2 is left out. In step 5 the results from the FORM analysis are compared. If the importance factors are similar and the reliability in the observed situation is lower, then updating is expected to be effective. A larger difference indicates a larger update. When the importance factors are similar, but the observed situation is more reliable, then no effect is expected. In the case that the importance factors are not similar, and the reliabilities are similar, then no decisive indication can be given. Although, the lower the reliability in the observation, the more likely an effective update is. In a design process, step 2 – 5 can be iterated through to explore all possibilities.

1. Determine the critical slip plane of the assessment situation with a semi-probabilistic slope stability analysis
2. Create a loading condition with a lower safety factor than in the assessment situation, considering the same slip plane
3. Perform a FORM analysis on the critical slip plane in the assessment situation
4. Perform a FORM analysis in the observed situation, considering the same slip plane
5. Compare the importance factors and reliabilities.

Another recommendation is done regarding an important assumption of the methodology developed in this thesis. It was assumed that the reliability can only be updated if the two considered slip planes in the observed and assessment situation are equal. The need for this assumption originates from the variable nature of the subsoil, which is not modelled in the slope stability analyses. Considering two different slip planes can lead to unjustly disregarding parameter combinations and hence an overestimation of the updated reliability. On the other hand, the observation may also carry information when the slip planes are not exactly equal. This causes the assumption to be conservative, possibly leading to an underestimation of the updated reliability. To improve the reliability updating results, it is therefore recommended to incorporate spatial variability into the slope stability analysis. A possibility could be the use of random fields for soil parameters (see e.g. Griffiths et al., 2007).

This recommendation follows from the discussion of flood risk in the previous section. In is reasoned that creating a critical loading condition, with a relatively low slope reliability, may be acceptable in terms of flood risk. This qualitative reasoning however does not provide any guidance for what level of slope reliabilities is acceptable. For future applications of reliability updating with construction survival, it is recommended to quantify the probability of flooding due to a slope instability during construction. To this end, it needs to be determined what the water retaining height is after the slope failure and what time is needed for repairing the damage. The flooding probability is then the simultaneous occurrence of a slope failure and a water level exceeding the residual water retaining height. Based on this flooding probability and what probability is allowed, a criterion can be determined regarding the slope reliability during construction. This would allow for risk assessment and optimization of the effect of reliability updating.

Currently, the training of the metamodels is based on their individual uncertainty regarding the reliability. This causes the models to be most accurate around the design points, where failure is most probable. However, the design point changes by updating the reliability. For smaller updates, it is expected that the accuracy of the updated reliability is not affected much, as the shift in design point is small. But for larger updates causing a significant shift in design point, an improvement in the reliability estimate is probably possible. In order to increase the prediction of the updated reliability, it is recommended to research a joint learning function and convergence criterion for training the two metamodels needed for updating, based on the accuracy of the updated reliability. Defining a convergence criterion regarding the updated reliability would be useful for assessing the certainty of the outcome. As an improvement involving minor adjustments, it is suggested to add a joint learning cycle after convergence of the individual metamodels. This can be done by creating an importance sampling distribution around the new design point and using the existing learning functions and

convergence criterion only on the plausible samples. Yet, a more efficient approach would be simultaneous training of the models involving a joint learning function and a convergence criterion for the updated reliability from the start of the training procedure.

In the case study of this research, the uncertainty of all parameters was modelled as epistemic (reducible), except for the water level. This is a solid assumption for soil parameters, but arguable for the uncertainty of the model factor regarding the slip plane model. It was reasoned that the unreducible part of the uncertainty is small due to the great similarity of the observed and assessment situation. Hence, the uncertainty was assumed epistemic. However, in the case study the model turned out to be a significant contributor to the reliability update. It is therefore recommended to investigate the sensitivity of the update to the amount of unreducible uncertainty and to research how much unreducible uncertainty should be accounted for in the model factor.

A final recommendation is done regarding the applicability of reliability updating with construction survival. In the case study presented in this thesis the water level is of little influence. However, in large parts of the Netherlands and other deltas the water level is of great influence, like situations with a thin blanket layer sensitive for uplift. Therefore, further research into cases where the water level does play an important role is recommended to explore the applicability of reliability updating with construction survival.

References

Au, S.-K., & Beck, J. L. (2001). Estimation of small failure probabilities in high dimensions by subset simulation. *Probabilistic Engineering Mechanics*, *16*(4), 263–277. https://doi.org/https://doi.org/10.1016/S0266-8920(01)00019-4

Baudin, M., Dutfoy, A., Iooss, B., & Popelin, A.-L. (2015). *OpenTURNS: An industrial software for uncertainty quantification in simulation*.

Box, G. P., Hunter, W., & Hunter, J. (1978). *Statistics for experimenters : an introduction to design, data analysis, and model building*.

Chai, X. (2019). *Probabilistic system identification and reliability updating for hydraulic structures – Application to sheet pile walls*. TU Delft.

Deltares. (2018). *Fenomenologische beschrijving - Faalmechanismen WBI*. https://doi.org/document 11200574-007

Díaz-Francés, E., & Rubio, F. J. (2013). On the existence of a normal approximation to the distribution of the ratio of two independent normal random variables. *Statistical Papers*, *54*(2), 309–323. https://doi.org/10.1007/s00362-012-0429-2

Echard, B., Gayton, N., & Lemaire, M. (2011). AK-MCS: An active learning reliability method combining Kriging and Monte Carlo Simulation. *Structural Safety*, *33*(2), 145–154. https://doi.org/https://doi.org/10.1016/j.strusafe.2011.01.002

Griffiths, D. V., Fenton, G. A., & Denavit, M. D. (2007). Traditional and Advanced Probabilistic Slope Stability Analysis. In *Probabilistic Applications in Geotechnical Engineering* (pp. 1–10). https://doi.org/doi:10.1061/40914(233)19

Han, B., Liao, H.-J., Li, H., & Xiao, Z. (2011). Effects of Soil Properties Uncertainty on Reliability Index of Slope. *Advanced Materials Research*, *250–253*, 1947–1950. https://doi.org/10.4028/www.scientific.net/AMR.250-253.1947

Jonkman, S. N., Steenbergen, R. D. J. M., Morales-Nápoles, O., Vrouwenvelder, A. C. W. M., & Vrijling, J. K. (2017). *Probabilistic Design: Risk and Reliability Analysis in Civil Engineering* (4th ed.). TU Delft.

Kelly, R., & Huang, J. (2015). Bayesian updating for one-dimensional consolidation measurements. *Canadian Geotechnical Journal*, *52*(9), 1318–1330. https://doi.org/10.1139/cgj-2014-0338

Laghmouchi, K. (2021). *Reliability updating of existing quay walls based on the effects of past performance* [Delft University of Technology]. http://resolver.tudelft.nl/uuid:1c263993-09d1-4de1-8373-883a5a366504

Li, D.-Q., Zheng, D., Cao, Z.-J., Tang, X.-S., & Phoon, K.-K. (2016). Response surface methods for slope reliability analysis: Review and comparison. *Engineering Geology*, *203*, 3–14. https://doi.org/https://doi.org/10.1016/j.enggeo.2015.09.003

Melchers, R. E. (1989). Importance sampling in structural systems. *Structural Safety*, *6*(1), 3–10. https://doi.org/https://doi.org/10.1016/0167-4730(89)90003-9

Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. The MIT Press.

Papaioannou, I., & Straub, D. (2012). Reliability updating in geotechnical engineering including spatial variability of soil. *Computers and Geotechnics*, *42*, 44–51. https://doi.org/https://doi.org/10.1016/j.compgeo.2011.12.004

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Machine Learning in Python. Journal of Machine Learning Research*, *12*, 2825–2830.

Rasmussen, C. E., & Williams, C. K. I. (2006). *Gaussian Processes for Machine Learning*. MIT Press. http://www.gaussianprocess.org/gpml

Roscoe, K., Diermanse, F., & Vrouwenvelder, T. (2015). System reliability with correlated components: Accuracy of the Equivalent Planes method. *Structural Safety*, *57*, 53–64. https://doi.org/https://doi.org/10.1016/j.strusafe.2015.07.006

Schweckendiek, T. (2014). *On reducing piping uncertainties - A Bayesian Decision Approach*. TU Delft.

Schweckendiek, T., & Kanning, W. (2016). *Reliability updating using past performance - Background report*.

Schweckendiek, T., Kanning, W., Brinkman, R., Klerk, W., van der Krogt, M., Rippi, K., & Teixeira, A. (2017). *Reliability updating for slope stability of dikes*.

Schweckendiek, T., Teixeira, A., van der Krogt, M., & Kanning, W. (2016). *Reliability updating for slope stability of dikes (test cases report)*.

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

Sousa, H., Rózsás, Á., Slobbe, A. T., & Courage, W. M. G. (2019). A novel pro-active approach towards SHM-based bridge management supported by FE analysis and Bayesian methods. *Structure and Infrastructure Engineering*. https://doi.org/10.1080/15732479.2019.1649287

Straub, D., & Papaioannou, I. (2014). Bayesian Updating with Structural Reliability Methods. *Journal of Engineering Mechanics*, *141*(3), 4014134. https://doi.org/10.1061/(ASCE)EM.1943-7889.0000839

TAW. (1985). *Dijkversterking Alblasserwaard, traject Streefkerk-Midden - Onderzoek naar oorzaken afschuiving nabij d.p. 211 in oktober 1984*. http://resolver.tudelft.nl/uuid:26039668-9383-4be8-b92f-0d8ce8d81971

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

Van den Eijnden, A. P., Schweckendiek, T., & Hicks, M. A. (2021). Metamodelling for geotechnical reliability analysis with noisy and incomplete models. *Under Review*.

van der Krogt, M. G., Klerk, W. J., Kanning, W., Schweckendiek, T., & Kok, M. (2020). Value of information of combinations of proof loading and pore pressure monitoring for flood defences. *Structure and Infrastructure Engineering*, 1–16. https://doi.org/10.1080/15732479.2020.1857794

van der Krogt, M. G., Schweckendiek, T., & Kok, M. (2021). Improving dike reliability estimates by incorporating construction survival. *Engineering Geology*, *280*, 105937.

https://doi.org/https://doi.org/10.1016/j.enggeo.2020.105937

Van der Meij, R. (2019). *D-Stability - User manual* (20.1). Deltares. https://www.deltares.nl/nl/software/d-stability-nl/

Van der Meijs, R. C. J. (2015). *Increasing the Reliability of Settlement Predictions - A Bayesian Approach* [TU Delft]. http://resolver.tudelft.nl/uuid:e977cc50-39d7-4ab7-a8c5-c7501dfdc03c

Verruijt, A. (2001). *Soil Mechanics*. Delft University of Technology. https://ocw.tudelft.nl/wp-content/uploads/SoilMechBook.pdf

VNK2. (2014). *De veiligheid van Nederland in kaart Eindrapport*. https://www.helpdeskwater.nl/onderwerpen/waterveiligheid/programma-projecten/veiligheid-nederland/publicaties/

W+B. (2013). *Geotechnische rapportage dijkversterking Kinderdijk-Schoonhovenseveer*.

WBI. (2017). *Regeling veiligheid primaire waterkeringen 2017 - Bijlage III Sterkte en Veiligheid*.

WBI. (2019). *Schematiseringshandleiding macrostabiliteit*. https://www.helpdeskwater.nl/onderwerpen/waterveiligheid/primaire/beoordelen/ontwikkelingen-wbi/release-notes-28-november-2019/aanpassingen-per-product/aanpassingen-per-product/schematiseringshandleiding-macrostabiliteit/

Appendices

A

ERRAGA user experience

This appendix describes a couple of ERRAGA's settings and gives advice regarding the use of these settings. It is meant to make future users of ERRAGA aware of the choices regarding these settings and perhaps kickstart their analysis.

Choice of learning function and initial training set size
Metamodels are trained in ERRAGA with an iterative learning procedure. Additional model evaluations are added to the Design of Experiment (DoE) until the metamodel fulfills the convergence criterion (see section 2.6). The learning procedure starts by drawing a sample pool from the (importance) distribution. In every learning iteration a sample is chosen from the sampling pool which is expected to be most informative for training the model. The selection of the most informative sample is done by a learning function, of which various types are implemented in ERRAGA. This research has experimented with using the U-learning and UNIS-learning function, the findings are shared here.

The U-learning function selects the sample from the sampling pool for which the probability of misclassification is largest (e.g. predicting failure instead of survival and vice versa). This is done by considering the absolute ratio of predicted limit state outcome and prediction uncertainty. The sample with the smallest ratio is chosen. The function was however meant for a sample pool sampled with crude Monte Carlo, in which all samples have equal weight. Using the U-learning function together with importance sampling causes a suboptimal learning process because sampling weights are not included. It leads to preferred learning around the limit state line. This seems desirable, but also samples around the limit state line that have a very low sampling weight are evaluated, which may not add much information to the training set. However, something more important is the general convergence behavior of the metamodel. It was found that using U-learning with importance sampling can lead to an insufficient amount of realizations around the mean, leading to inaccurate predictions in this region of the parameter space, affecting convergence of the model. In the case study it was found that for initial DoE of 50 model evaluations, the metamodels did not converge after 800 model evaluations. Choosing a larger set of 200 model evaluations did result in convergence.

So, the U-learning function is an efficient learning approach for crude Monte Carlo sampling in which samples have equal weight, but this is not the case when importance sampling is used. Therefore, Van den Eijnden et al. (2021) developed the UNIS-learning function in which includes the sampling weight and considers noise in the model. As recommended in the article of Van den Eijnden et al. (2021), the UNIS-learn function is more suited for combination with adaptive importance sampling. Still, initiating a metamodel with an initial DoE of limited size could affect the convergence behavior. In the case study it was found that an initial DoE with 50 model evaluations always resulted in convergence of the metamodels.

Minimum amount of white noise

In some computational models a bit of (numerical) noise is present such that performing the same calculation twice will result in slightly different answers. This could be problematic for training a regression model if it does not consider possible ambiguity due to model noise. This is solved in ERRAGA by the possibility to deploy a white noise component in the kernel. ERRAGA tries to find a best fit for this white noise component, but this can be steered by setting upper and lower bounds the white noise.

In some cases, ERRAGA finds an optimal solution for a regression model with a noise term that is lower than the actual noise present in evaluating the model responses, possibly overfitting the model. To investigate this behavior a four parameter metamodel for slope stability was trained with ERRAGA, using a fixed value for the white noise standard deviation σ_{wn}. The results are presented in figure A.1, in which the limit state outcome is plotted as a function of two of the model parameters. In this example, a white noise term with a standard deviation of 0.002 on the limit state outcome is enough to obtain a good fit. For the model used in the case study, containing 14 variables, 0.001 was already enough. Although giving a minimum noise term in the 14-parameter case was not necessary most of the time. Based on the findings described above, it is recommended to always check the limit state functions for logical behavior and overfitting. If overfitting is suspected, then rerunning the analysis with a (increased) minimum white noise may yield better results.

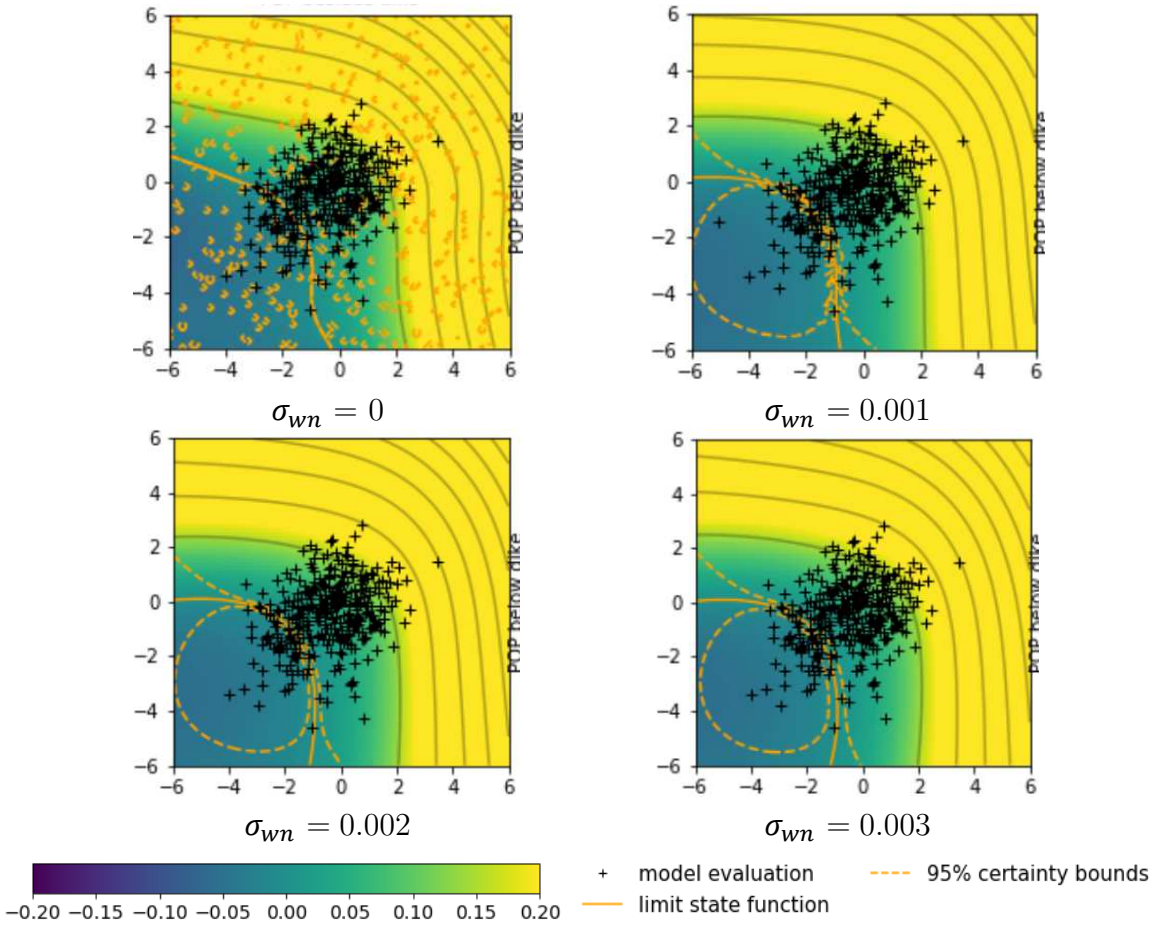


Figure A.1: Plot of limit state function for different amounts of white noise

Amount of kernel parameter optimizations

The kernel parameters of the metamodel are (re)fitted each time the DoE is expanded with an extra model evaluation. The fit is based on the maximum likelihood of the kernel parameters. A numerical method is used for finding the maxima, as computing the likelihood over the total parameters space would be computationally expensive. Using a numerical method could however result in finding a local instead of the global maximum. To reduce the probability that this occurs, an optimization procedure is implemented[2] in which the maximum likelihood determination is repeated a couple of times, each time with a random starting point. The amount of repetitions can be set in ERRAGA with the attribute "OptIter" of ERRAGA's Learn class.

For the fourteen-parameter model used in the case study, it was found that sometimes the optimal set of kernel parameters varied strongly by adding a model realization to the DoE. This behavior is presented in the left plot of figure A.2, in which the failure probability estimate is plotted as a function of the amount of model realizations (made with ERRAGA's ConvPlot.py). This behavior occurs because the most likely set of

[2] Implemented in Scikit-learn, the Python Machine Learning model used in ERRAGA (Pedregosa et al., 2011)

kernel parameters is not found every time that a model evaluation is added to the DoE. By default, the number of optimizations is 10. For the cases treated in this research, increasing it to 20 solves the behavior observed in the left plot and results in the behavior presented in the right plot. It is noted that the figures below are made with different models, which explains the difference in failure probability.

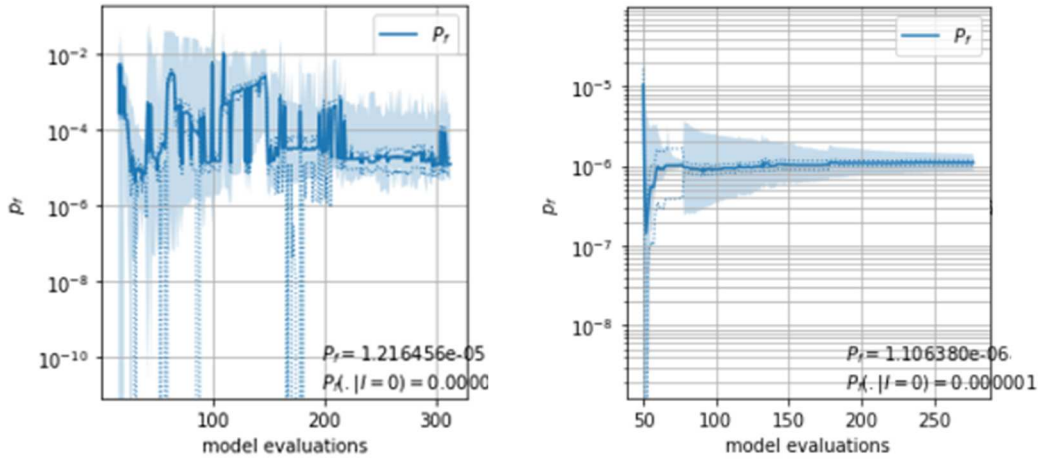


Figure A.2: Convergece plot with with 10 kernel parameter optimizations (left) and 20 kernel parameter optimizations (right)

Determining marginal parameter distributions

A last remark is made on the use of ERRAGA's CrossScatter.py for constructing the marginal (updated) parameter distributions together with importance sampling. CrossScatter can be used to visualize the samples drawn in the reliability analysis and gives insight in which samples lead to failure, as presented in the figure A.2. It was found that when using importance sampling, the marginal distributions might not turn out as expected and can have large spikes, shown on the diagonal of figure A.2. This behavior is caused by under sampling of the region of failure. A couple of samples with a high importance sampling weight dominate the histograms as shown. In this case the default amount of 100 000 samples in ERRAGA is not enough for constructing the marginal distributions. Increasing the sample size solves the problem.

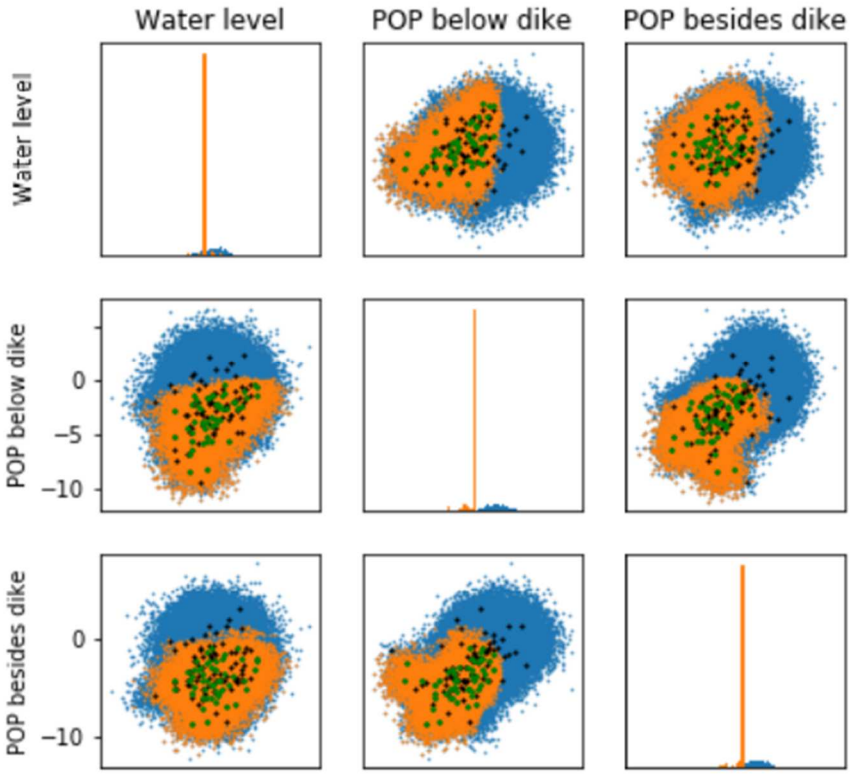
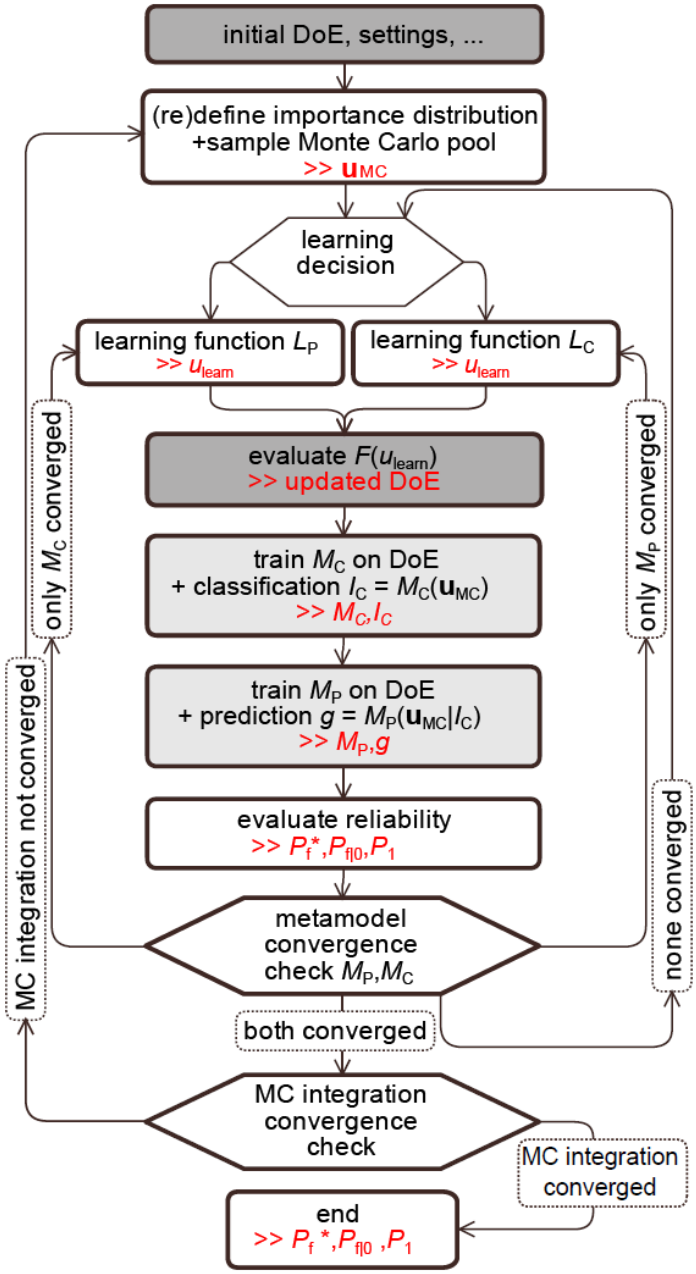


Figure A.3: Samples plotted in U-space (blue: all samples, orange: samples leading to failure)

B

ERRAGA's learning algorithm



Original ERRAGA learning algorithm (Van den Eijnden et al., 2021)

C

Compression tests Salmsteke – Schoonhoven

Clay below dike

Sample name	location	surface level [m+NAP]	top sample [m+NAP]	bottom sample [m+NAP]	in-situ effective stress [kPa]	yield stress [kPa]	POP [kPa]	OCR [-]
163_DP111+003_B_BERM_23a1	berm	2.25	-6.55	-6.65	74.4	101.7	27.3	1.4
413_DP168+051_B_BERM_031-a	berm	1.88	-10.12	-10.48	118.2	133.7	15.5	1.1
430_DP168+050_B_KR_011-a	crest	5.83	0.83	0.56	72.3	134.0	61.7	1.9
438_DP168+050_B_KR_017-a	crest	5.83	-1.57	-1.85	88.3	130.6	42.3	1.5
441_DP168+050_B_KR_019-a	crest	5.83	-2.37	-2.66	94.0	109.7	15.7	1.2
526_DP186+000_B_BERM_011-a	berm	3.03	-0.97	-1.29	56.2	110.6	54.4	2.0
554_DP186+000_B_BERM_033-a	berm	3.03	-9.77	-10.06	102.1	153.8	51.7	1.5
577_DP185+100_B_KR_015-a	crest	5.47	-1.33	-1.62	85.3	155.8	70.5	1.8
581_DP185+100_B_KR_018-a	crest	5.47	-2.53	-2.82	94.5	140.1	45.6	1.5

Clay besides dike

Sample name	location	surface level [m+NAP]	top sample [m+NAP]	bottom sample [m+NAP]	in-situ effective stress [kPa]	yield stress [kPa]	POP [kPa]	OCR [-]
79_DP111+000_B_BUT_09a2	outer toe	2.19	-1.21	-1.34	24.4	67.7	43.3	2.8
93_DP111+000_B_BUT_20a1	outer toe	2.19	-5.41	-5.79	35.6	49.3	13.7	1.4
112_DP111+003_B_AL_07a1	hinterland	0.66	-1.74	-2.01	43.0	65.9	22.9	1.5
367_DP168+049_B_AL_022-a	hinterland	-0.86	-9.26	-9.58	52.4	68.4	16.0	1.3
369_DP168+049_B_AL_023-a	hinterland	-0.86	-9.66	-10.02	55.3	83.0	27.7	1.5
461_DP168+055_B_BUT_007-a	outer toe	2.09	-0.31	-0.67	37.1	66.3	29.2	1.8
511_DP185+099_B_AL_021-a	hinterland	-0.62	-8.62	-8.95	43.6	92.6	49.0	2.1
514_DP185+099_B_AL_023-a	hinterland	-0.62	-9.42	-9.77	48.5	81.8	33.3	1.7

Peat below dike

Sample name	location	surface level [m+NAP]	top sample [m+NAP]	bottom sample [m+NAP]	in-situ effective stress [kPa]	yield stress [kPa]	POP [kPa]	OCR [-]
66_DP110+099_B_KR_24a1	crest	6.56	-4.24	-4.39	142.1	188.8	46.7	1.3
147_DP111+003_B_BERM_11a1	berm	2.25	-1.75	-1.85	47.6	71.7	24.1	1.5
152_DP111+003_B_BERM_14a1	berm	2.25	-2.95	-3.15	51.1	71.5	20.4	1.4
167_DP111+003_B_BERM_26b1	berm	2.25	-7.95	-8.05	79.8	115.7	35.9	1.4
393_DP168+051_B_BERM_016-a	berm	1.88	-4.12	-4.47	74.1	107.5	33.4	1.5
396_DP168+051_B_BERM_018-a	berm	1.88	-4.92	-5.27	76.2	101.5	25.3	1.3
398_DP168+051_B_BERM_020-a	berm	1.88	-5.72	-6.07	76.9	108.4	31.5	1.4
453_DP168+050_B_KR_027-a	crest	5.83	-5.57	-5.96	109.8	125.5	15.7	1.1
538_DP186+000_B_BERM_020-a	berm	3.03	-4.57	-4.95	77.6	123.8	46.2	1.6
540_DP186+000_B_BERM_022-a	berm	3.03	-5.37	-5.68	78.3	117.2	38.9	1.5
544_DP186+000_B_BERM_024-a	berm	3.03	-6.17	-6.48	79.2	144.8	65.6	1.8
591_DP185+100_B_KR_026-c	crest	5.47	-5.85	-6.00	108.2	175.1	66.9	1.6

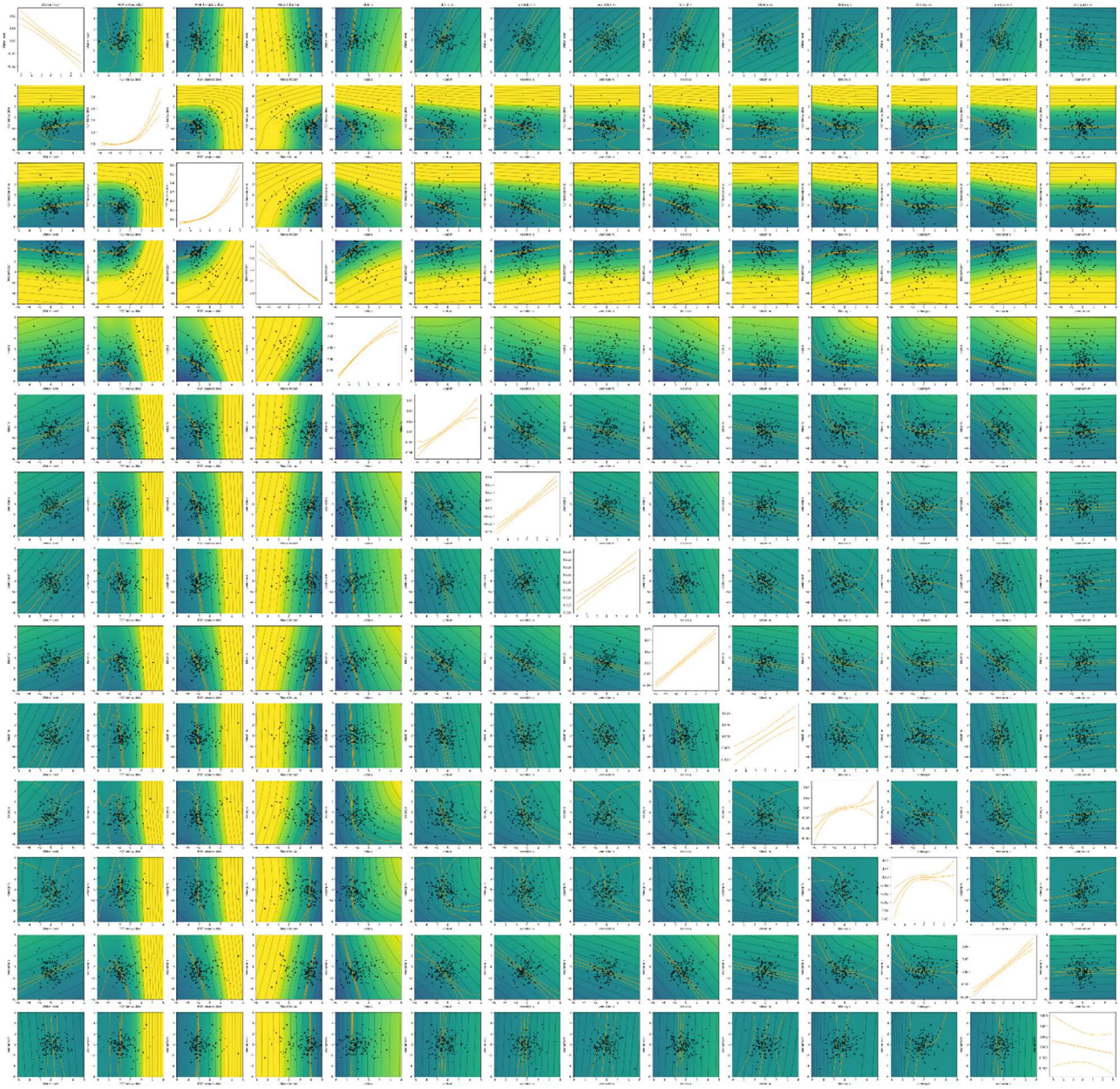
Peat besides dike

Sample name	location	surface level [m+NAP]	top sample [m+NAP]	bottom sample [m+NAP]	in-situ effective stress [kPa]	yield stress [kPa]	POP [kPa]	OCR [-]
86_DP111+000_B_BUT_14a2	outer toe	2.19	-3.21	-3.36	28.4	90.0	61.6	3.2
89_DP111+000_B_BUT_16a1	outer toe	2.19	-3.81	-4.20	29.4	78.4	49.0	2.7
127_DP111+003_B_AL_20a2	hinterland	0.66	-7.09	-7.23	75.2	98.4	23.2	1.3
128_DP111+003_B_AL_21a1	hinterland	0.66	-7.34	-7.73	76.2	111.3	35.1	1.5
130_DP111+003_B_AL_23a1	hinterland	0.66	-8.14	-8.48	77.3	98.1	20.8	1.3
345_DP168+049_B_AL_006-a	hinterland	-0.86	-2.86	-3.06	18.3	37.0	18.7	2.0
348_DP168+049_B_AL_008-a	hinterland	-0.86	-3.66	-4.04	19.0	39.0	20.0	2.1
353_DP168+049_B_AL_012-a	hinterland	-0.86	-5.26	-5.41	20.0	60.8	40.8	3.0
473_DP168+055_B_BUT_019-a	outer toe	2.09	-5.11	-5.41	69.2	121.1	51.9	1.7
475_DP168+055_B_BUT_020-a	outer toe	2.09	-5.51	-5.73	70.6	107.7	37.1	1.5
491_DP185+099_B_AL_007-a	hinterland	-0.62	-3.02	-3.39	23.0	58.5	35.5	2.5
494_DP185+099_B_AL_009-a	hinterland	-0.62	-3.82	-4.20	24.0	59.7	35.7	2.5
497_DP185+099_B_AL_011-c	hinterland	-0.62	-4.93	-4.99	23.8	44.8	21.0	1.9

D

Limit state plot

Assessment situation base case:



E

Prior and posterior domain of failure

Alternative B, prior parameter distributions and prior and posterior failure domains: