Spatial Variability in Dike Stability Assessments

Lisa van der Burg

Delft University of Technology & HKV lijn in water

TU Delft

HKV LIJN IN WATER

Spatial Variability in Dike Stability Assessments

by

Lisa van der Burg

to obtain the degree of Master of Science
in Civil Engineering at the Delft University of Technology,
to be defended publicly on Tuesday January 30, 2024 at 14:00.

Student number: 4714377

Thesis committee:	Dr. ir. A. P. van den Eijnden,	TU Delft, committee chair
	Dr. ir. R. C. Lanzafame,	TU Delft
	Ir. J. Tigchelaar,	HKV lijn in water
	P. Kindermann Msc.	HKV lijn in water

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

TU Delft

HKV LIJN IN WATER

Preface

The Dutch people have been fighting the threat of water for generations. My master's thesis on "Spatial Variability in Dike Stability Assessments" follows in the footsteps of the work of my grandfather, an engineer who worked on the Delta Works. Similar to the years in which the Delta Works were constructed, safety remains a top priority. With the help of new technologies, we can now also strive for more efficient and economical designs of dikes by minimizing uncertainties, such as the spatial variability I studied.

This thesis is the final part of my master's in Civil Engineering at the TU Delft, specializing in Geo-Engineering. I am grateful for the opportunity from HKV lijn in water to combine my interests in geotechnical engineering and mathematics to work on an important matter, such as water safety. This issue will become even more relevant due to the growing threat of climate change.

I want to thank those who guided me through this process. Starting with Bram van den Eijnden, an assistant professor at the Geo-Engineering department and the chair of my thesis committee, for his time and our interesting discussions that constantly improved my understanding and enthusiasm for the topic. Secondly, I would like to thank Robert Lanzafame, my second supervisor from the TU Delft, I want to thank you for providing sharp input during progress meetings and offering a never-ending list of ideas to improve my research. I am thankful for the support of my HKV supervisor Jan Tigchelaar, for providing an infectious enthusiasm for everything soil-related and always reminding me of the importance of the study. And I want to especially thank Paulina Kindermann, my daily HKV supervisor, for always showing interest, positivity, and sharp observations on the subject and text.

I would also like to thank Sander Kapinga from water board Rivierenland and Dusan Bosic from water board Brabantse Delta for providing data and Wim Kanning for answering my questions. Also, thanks to the HKV colleagues for reminding me that while it is good to focus on the details, it is also important to zoom out occasionally.

Finally, I would like to thank everyone who made this experience enjoyable: my friends, family and listening ear for their support. Special acknowledgements go to the friends I often studied with over the past few months. They made it feel like a collective effort. A special thanks goes out to my "rubber duck", Marilotte, for helping me gain valuable insights. And, of course, thanks to my other friends at the Garage for the feedback on the text and for reminding me that there is more to life than writing my thesis.

I look back on the past few years with much gratitude for the academic and personal development opportunities, from doing a Computer Science minor to living in a student house and doing a board year with AIESEC and de Ondergrondse. I am excited to take all I have learned and experienced into the future and become a true mud master.

Lisa van der Burg
Delft, January 2024

Abstract

The safety of the Dutch dikes depends on various failure mechanisms. Macro stability, a geotechnical failure mechanism, is highly affected by differences in soil strength because the sliding plane propagates through areas of least resistance. The variations in soil properties in space, also known as spatial variability, are caused by geological processes and determine the locations of weaker zones in a dike. This highlights the importance of incorporating spatial variability into dike stability assessments.

The approach to incorporate spatial variability in the official Dutch assessment framework (WBI2017) relies on various assumptions. It assumes complete local variance reduction and neglects that the failure mechanism propagates through the weaker zones, leading to a mean strength reduction. Moreover, it assumes a default value for the ratio between local and regional variance $\alpha = 0.75$, which lacks empirical evidence. Another method to incorporate spatial variability in stability calculations is the Random Finite Element Method (RFEM). This probabilistic technique models strong and weak zones through random fields. However, the main drawback is its considerable computation time, making it impractical to use for the assessment of the hundreds of kilometers of dikes in the Netherlands.

To address these issues, this research answers the question: *What is an effective approach for incorporating spatial variability in soil into dike stability calculations?* The study is divided into two parts: a data analysis and the creation of an RFEM model.

The first part investigated national and regional spatial correlations using variograms. The study found that the local spatial variance cannot be analyzed with variograms based on the national dataset. This is because the variograms average the variance in the local data due to their large scale. Investigating local variability requires local data with a high enough density and accuracy in the research area.

In the second part, the inclusion of spatial variability was studied for a case study dike using RFEM, which is part of dike trajectory 34-2, located between Willemstad and Noordschans. The research highlighted two differences between assumptions made by RFEM and the WBI2017 method: (1) the inclusion of statistical uncertainty and (2) the use of different stress components in calculating the undrained shear stress. The results of the different methods can only be compared if these differences are accounted for. Furthermore, the study found that using a probabilistic calculation with $\alpha = 0.8$ better fits the results of the realistic RFEM model of the case study dike, particularly in the lower tail of the distribution of the results, compared to the default value of $\alpha = 0.75$. Therefore, it can be concluded that $\alpha = 0.8$ leads to a more realistic approximation of the probability of failure of this cross-section.

To investigate the importance of this finding, an assessment was carried out following the guidelines of WBI2017 but with $\alpha = 0.8$. This showed that the probability of failure for dike trajectory 34-2 was reduced by 39.72% but that the safety category of the dike trajectory (for macro stability) remains unchanged.

Therefore, the answer to the research question is that when considering the computational requirements of RFEM, it is more effective to keep using the WBI2017 approach of implementing spatial variability into the input parameters of dike stability calculations with $\alpha = 0.75$.

These findings are relevant as they validate the use of the current method with the available data. The study used different approaches for incorporating spatial variability in stability calculations and provides valuable insights for future research.

Contents

1 Introduction

1.1. Motivation

Throughout history, the Netherlands has faced the constant threat of water (Deltacommissie, 1961; ENW, 2016; Stijnen et al., 2014; TAW, 1998). People in the low-lying areas began building dikes centuries ago to protect their fertile lands and homes (TAW, 1998). Despite their efforts, floods have plagued this land for hundreds of years. One of the most devastating floods was the Watersnoodramp, a powerful storm surge in 1953, which flooded over $2,000$ square kilometers of land, killing $1,835$ people (Deltacommissie, 1961; Jonkman et al., 2008). To prevent such a disaster from happening in the future, the government formed the Delta Commission. In 1993 and 1995, the river dikes in the Netherlands could not handle the high river discharges of $11,000$ and $12,000$ cubic meters per second, leading to the evacuation of $250,000$ residents. (ENW, 2016). As a result, the Delta Commission expanded its efforts to improve the river dikes as well (Deltacommissie, 2008).

In the Netherlands, most of the primary dikes are earthen structures. The soil properties significantly influence the strength of these dikes (ENW, 2016). According to Clayton (2001), around 85% of the general construction risk is related to geotechnical issues. Therefore, an accurate understanding of the soil strength is crucial in reducing the risk of flooding. Unlike manufactured materials like concrete or steel, natural soil varies considerably due to geological processes (Calle et al., 2021). To accurately measure this variability in the soil, thorough soil investigations are needed. However, limited data is usually available due to time and cost limitations (Calle et al., 2021). As a result, conservative values for soil properties are often used because of a lack of knowledge about the uncertainties and their effect on the stability of dikes, potentially leading to overly cautious and uneconomic designs (Varkey, 2020). This highlights an issue: in an attempt to cut expenses on extensive soil investigations, there is a potential for incurring higher costs through uneconomic designs.

1.1.1. Background Information

In the Netherlands, dike safety is governed by the Water Law (Waterwet) and assessed through procedures of the Legal Assessment Instrument (Wettelijk BeoordelingsInstrumentarium) (WBI), with the latest version published in 2017. The Dutch dike network has been divided into trajectories, each with its own safety standard. The same threat and, to a lesser extent, the same consequences in the case of a breach are assumed over the whole trajectory (ENW, 2016). Within each trajectory, statistically homogeneous sections share the same loads, geometry, and soil layers. Evaluating representative cross-sections involves considering various failure mechanisms, as shown in the fault tree (Figure 1.1). If any section fails, flooding occurs, indicating a dike breach.

One of the failure mechanisms significantly affected by differences in soil strength, along with piping, is macro-instability. As a geotechnical failure mechanism, spatial variability is one of the primary driving factors, alongside the load and geometry of the dike (ENW, 2016). This is because the sliding plane propagates through areas of least resistance, whose locations are determined by the spatial variability of the soil.

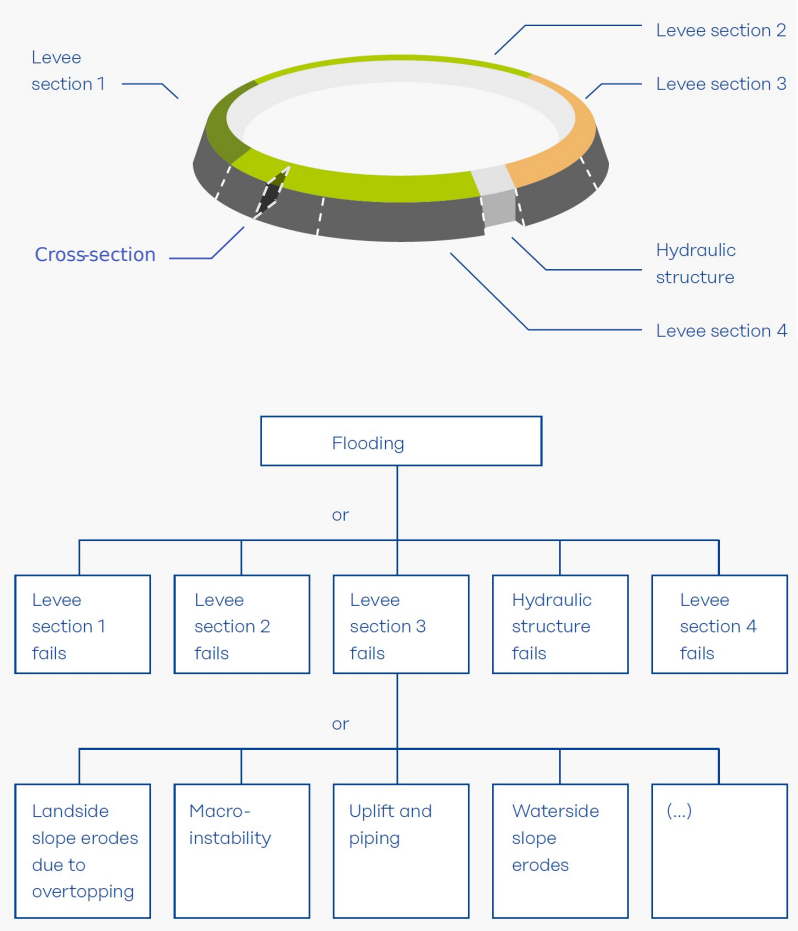


Figure 1.1: Fault tree with the different dike failure mechanisms. Adapted from ENW (2016).

In the WBI2017 assessment method, spatial variability of the soil is included in the model by Calle et al. (2021), based on Vanmarcke (1983). This model proposes complete averaging of local soil strength variability (variance reduction) due to the small scale of fluctuation compared to the layer thickness crossed by the failure surface. However, the strength of the soil fluctuates over a smaller distance in the vertical direction than in the horizontal direction due to geological deposition processes (Schweckendiek et al., 2017). Consequently, this model assumes that averaging is only relevant in the vertical direction.

Recent research has shown that spatial averaging in the Vanmarcke sense is an incomplete representation of the spatial variation in geotechnical engineering problems (Cami et al., 2020). Furthermore, studies with Random Finite Element Method (RFEM) reveal that the attraction of failure surfaces to weaker soil zones has significant implications in some cases (Calle et al., 2021), reducing mean soil properties. This is not included in the current model.

Moreover, Calle et al. (2008) and TAW (1989) found that Vanmarcke's model did not perfectly fit the dataset of available Dutch soil tests. To accommodate the wide range of means of the local test sets that made up the regional test set, the model had to be extended (Calle et al., 2021). To achieve this, Calle et al. (2008) proposed to incorporate the ratio between the local and regional variance (α) in the determination of soil strength parameters used for calculations. The default value of α, set at 0.75, was initially established by the TAW (1989) based on intuition due to the absence of adequate data, time, and techniques (Calle et al., 2021).

1.1.2. Problem Statement and Relevance

In the Netherlands, the water boards and Rijkswaterstaat (the State Department of Waterways and Public Works) maintain primary and regional dikes. They rely on (semi-)probabilistic stability calculations based on regional and local soil investigation. However, the current method of incorporating spatial variability into these calculations relies on numerous assumptions as outlined in Paragraph 1.1, which are as follows:

- The inclusion of variance reduction while neglecting the influence of reduced mean. By doing this, the method fails to account for the tendency of failure to occur in weaker zones. This neglect can lead to unconservative failure probability estimates depending on the problem.
- The default value of the α ratio, representing the relationship between regional and local data, lacks empirical evidence. It is assumed to be a conservative value (Calle, 2007, 2008). This parameter significantly influences the entire schematization, affecting the determination of the characteristic value or standard deviation.

It is widely acknowledged that spatial variability is a significant driving force behind dike instability (ENW, 2016). However, the amount of impact remains unknown due to the limited research conducted on the abovementioned assumptions. This knowledge gap shows the importance of understanding how these assumptions affect the failure probability estimation. It is necessary to bridge this gap to increase the accuracy and reliability of our stability assessments.

Although the use of RFEM provides a sophisticated approach for modeling spatial variability through random fields, its computational complexity makes it impractical for implementation in engineering practice.

1.2. Research Objective

The fact that the WBI2017 model for assessing Dutch dikes is based on assumptions (see Paragraph 1.1.2) and the impracticality of RFEM for dike maintenance institutions further highlights the need for an alternative approach. Therefore, the primary objective of this research is to develop a method that accurately and effectively accounts for spatial variability in the stability assessment of dikes. To achieve this objective, the following research question and sub-questions will be answered:

RQ *What is an effective approach for incorporating spatial variability in soil into dike stability calculations?*

SQ1 *How do different soil characteristics correlate in the spatial domain?*

SQ2 *What is the number and distribution of measurements on a local scale needed to accurately model the spatial variability in the soil?*

SQ3 *How does soil spatial variability on a regional scale influence the different steps involved in the schematization of dike stability calculations?*

SQ4 *How important is the spatial variability in the soil compared to other uncertainties?*

1.2.1. Methodology

The method used to investigate the influence of spatial variability consists of two parts. Firstly, a data analysis is conducted to investigate national and regional spatial correlation. Secondly, an RFEM model is generated to approximate reality and investigate the assumptions in the current model.

The data used in the data analysis is collected by Applied Water Management Research Foundation (Stichting Toegepast Onderzoek Waterbeheer) (STOWA) and will be filtered to handle outliers. Sample variograms will be constructed for different soil types and two different soil parameters: the volumetric weight γ_{wet} and SHANSEP parameter S. Analyzing the detail and accuracy of the sample variograms can determine whether the dataset is appropriate for answering sub-question 2. Theoretical variograms will be fitted to quantify the spatial variability per soil type and soil parameter. The data will be analyzed for different categories to gain more insight into the influences of spatial variability. This includes refining the soil classes, filtering by a specific depth range, only including the vertical variability, and investigating the influence of spatial scales. By doing this, the range over which the parameter is correlated, the

amount of total variance, and the variance due to spatial variability can be investigated. This information will answer sub-question 1 by quantifying the correlations in the spatial domain.

In the second part of the study, a realistic case of spatial variability will be created by generating a random field for a case study dike using Random Finite Element Method (RFEM). For this, a TU-Delft in-house finite element code (version Dec. 2023), built upon the finite element library by (Smith et al., 2014), as used in for example (Varkey et al., 2023), will be used. Sub-question 3 will be answered by comparing this realistic case with the current methodology for incorporating spatial variability in dike assessments as described in WBI2017. The calculations will be performed using a Finite Element Method (FEM). To calculate the probability of failure of the dike, Monte Carlo simulations will be used for the probabilistic and RFEM calculations. The distribution of the outcomes from the realizations and the shapes of the failure surfaces will be analyzed to study the assumptions of mean and variance reductions and the significance of spatial variability compared to other uncertainties (sub-question 4).

1.2.2. Scope

This research focuses exclusively on analyzing failure in primary flood defences in the Netherlands, referred to as ”dikes” hereafter. However, it should be noted that this topic is also relevant for the stability assessment in other soil bodies, like embankments.

The failure analysis will specifically consider the Ultimate Limit State (ULS) method. According to section 1, subsection 1 of the Water Act (Waterwet), the ULS is defined as follows: ”loss of flood defence capacity in a dike trajectory causing the area protected by the dike trajectory to flood in such a way that fatalities or substantial economic damage occur”. Therefore, this study will investigate the macro-stability failure of the land-side slope that directly leads to a dike breach, causing flooding. It will not address macro-stability in combination with other failure mechanisms.

For the stability of dikes, it can be assumed that the soft soils, namely peat and clay, have the most influence due to their low volumetric weight, leading to low effective stresses and undrained shear strength. Consequently, this research was focused solely on these soil types.

While the entire process shown in Figure 1.2 is essential for an effective approach to include spatial variability in dike stability assessments, this project will focus only on the initial stages. These stages involve integrating sample data into the assessment of dike stability and not the scaling process from the cross-section probability of failure to the section and trajectory level. Therefore, due to time constraints, this research will not analyze the length effect or the impact of horizontal variability parallel to the dike direction.

1.3. Structure of the Thesis

Firstly, in Chapter 2, the necessary literature on the WBI2017 method for assessing dike stability will be presented, followed by a discussion on geostatistical methods and a brief explanation of RFEM. After that, the thesis is divided into three parts, given the fact that the methodology consists of two parts, followed by a part of the discussion and conclusion.

The first part of the thesis, the data analysis, is further subdivided into three chapters. Chapter 3 provides a description of the databases and an overview of the strategy that will be used to investigate the spatial variability. Chapter 4 presents the results of the variograms, which will be discussed and concluded in Chapter 5. This chapter also includes an overview of the data from the data analysis that will be used for the model.

The second part of the thesis focuses on the model of the case study dike, with Chapter 6 introducing the case study and presenting the setup of the model. Chapter 7 presents the results of the model using different scenarios, while Chapter 8 discusses the results of the model and draws conclusions.

The third and final part of the thesis consists of two chapters, with Chapter 9 discussing the limitations of the approach of combining the data analysis with the model and the impact of the findings. Chapter 10 provides the answers to the research questions and some recommendations for future research.

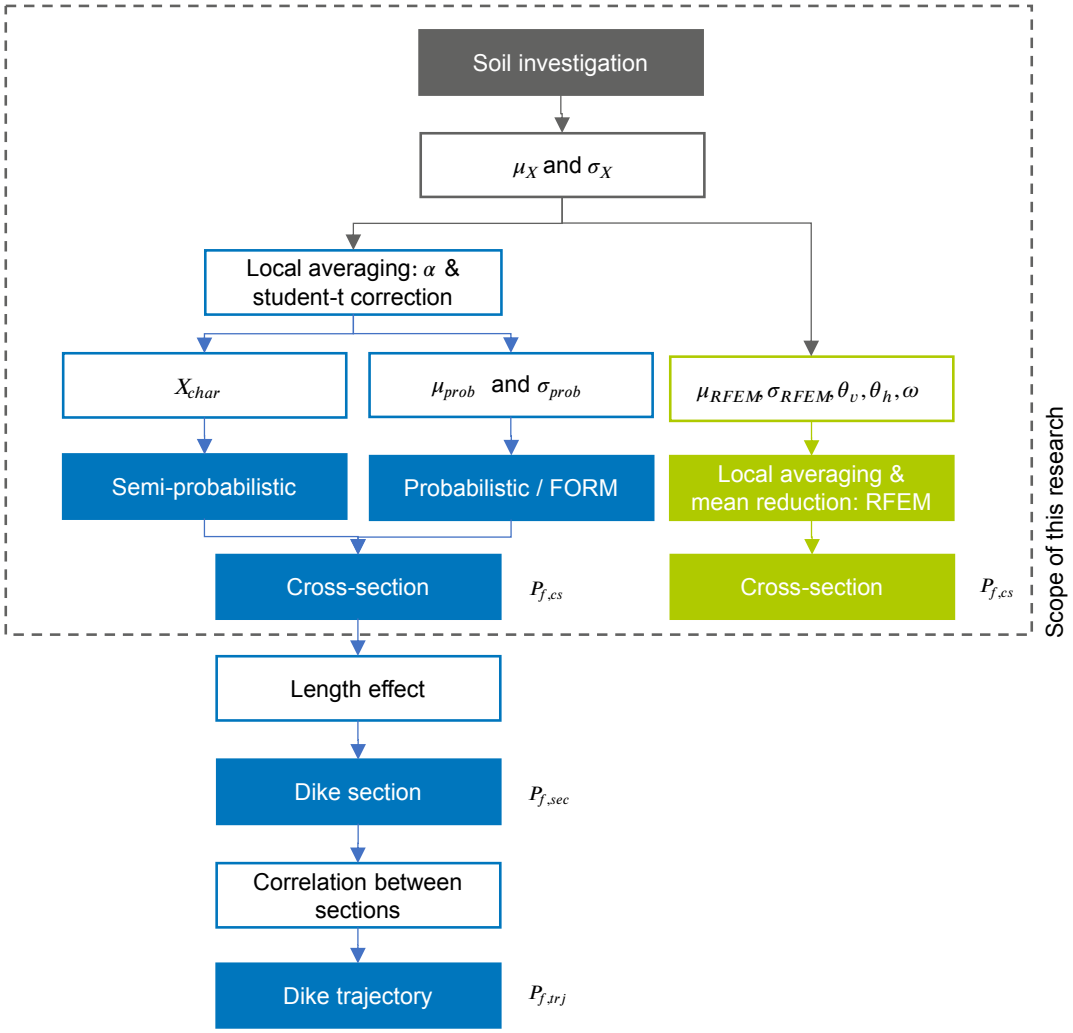


Figure 1.2: A schematization of the WBI2017 method (blue) (based on Calle et al. (2008)) and the RFEM method to include spatial variability and student-t correction of strength parameter X. Local averaging is implemented in different parts of the model. For the current method, local averaging is accounted for in the model input, while in RFEM, local averaging happens by modeling the strong and weak zones in the soil layers. In this way, the amount of local averaging can be investigated. The dashed line depicts the scope of this research.

2 Literature

2.1. WBI2017 Method

The dikes in the Netherlands are divided into separate trajectories, each with its safety norm as defined by the Water Law. To determine the failure probability of the trajectory, it is further divided into smaller sections, which are regarded as statistically homogeneous, indicating the same loads, geometry, and soil layers. For a more detailed description of the subdivision of dike trajectories, please refer to Appendix A.2. The evaluation of a representative cross-section involves considering multiple failure mechanisms. Figure 1.1 illustrates the fault tree for this analysis. If any of the dike sections fail, it leads to flooding. Consequently, the failure of a dike section occurs when a specific failure mechanism occurs within that particular section.

This research focuses on the macro-stability failure mechanism. Therefore, in the first section of this paragraph, a brief overview of the mechanism is provided. For a more extensive explanation of this mechanism, including calculation methods and information about Critical State Soil Mechanics (CSSM) and SHANSEP, please refer to Appendix A.3. Because this research investigates the current method of incorporating spatial variability into dike stability assessments, an overview of the theory behind the WBI2017 method will be given in the second section.

2.1.1. Macro-Stability

Macro-instability in dikes, as depicted in Figure 2.1, occurs when large sections of a soil mass shear along straight or curved deep shear planes. Macro-instability impacts both the dike body and the soil foundation underneath it (’t Hart et al., 2016). Together with internal erosion, the instability of a dike has been identified as one of the primary causes of past flooding events (ENW, 2016; Jonkman & Schweckendiek, 2015). A high water level will raise the water head and phreatic level by infiltration into the dike body, causing an increase in water pressures in the dike. This reduces effective stress, the primary driver of shear strength, and subsequently decreases the mobilized shear strength of the soil. This will ultimately cause the soil mass to slide and can trigger a dike breach.

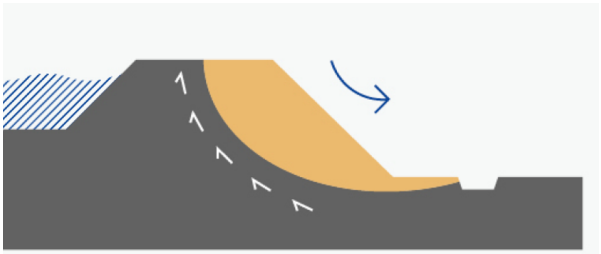


Figure 2.1: The failure mechanism macro-stability of the inner slope. Adapted from ENW (2016).

Determining the probability of dike failure due to macro-stability depends on several factors, including

the soil's strength. Since there are often significant uncertainties, probabilistic methods are frequently used to assess the probability of failure. Please refer to Appendix A.1 for the definition of the probability of failure and an outline of the various reliability analysis levels used to determine it.

The techniques used to evaluate the stability of a dike include numerical and Limit Equilibrium Methods (LEM). The LEM assumes that a dike will fail if a potential slip plane's horizontal, vertical, or moment equilibrium is disturbed and can model detailed soil profiles, seepage, and various loading scenarios (Sharp et al., 2013). Bishop (Bishop, 1955), Uplift-Van (Van, 2001), and Spencer (Spencer, 1967) are examples of LEM. On the other hand, numerical models, like the Finite Element Method (FEM), are more flexible in their choice of constitutive models. They can model groundwater flow and soil-structure interaction but are more computationally heavy and prone to input errors (Jonkman et al., 2021). More information about the second method will be provided in Chapter 2.3.1. Various factors influence the shape of the failure surface of a dike, and the choice of the method depends on the specific application.

Rijkswaterstaat (2021) has determined that the appropriate way to analyze macro-stability is through Critical State Soil Mechanics (CSSM) (Schofield & Wroth, 1968), which focuses on critical state rather than the peak shear strength. The critical state is where the soil no longer changes its volume during shearing. In CSSM, a distinction is made between drained and undrained modeling:

- *Drained:* The non-associative Mohr-Coulomb drained model is suitable for permeable layers like sand and layers above the phreatic line (Rijkswaterstaat, 2021). Excess pore pressures can dissipate, and effective stresses determine soil strength. In the CSSM framework, the dilation angle (ψ) and cohesion (c) are zero at the critical state.
- *Undrained:* The Stress History and Normalized Soil Engineering Property (SHANSEP) model (Ladd & Foott, 1974) has been used since 2017 to determine the strength of undrained layers in Dutch primary dikes. It describes the relationship between the undrained shear strength of the soil and the vertical effective stress.

2.1.2. WBI2017 Schematization of Spatial Variability

The current method for incorporating spatial variability in dike stability assessments in the Netherlands in WBI2017 is described by Rijkswaterstaat (2021) and Calle et al. (2021) and based on Calle et al. (2008). The assessment considers various factors of spatial variability, as shown in Figure 1.2. These factors include (1) the length effect, which is not within the scope of this research, (2) the complete averaging of vertical spatial variability to reduce variance, and (3) the ratio between local and regional variance.

The first point, the length effect, refers to the phenomenon where the probability of encountering a "weak spot" increases with the length of the slope, thereby raising the probability of a dike trajectory failure (Hicks & Li, 2018). Even though the length effect plays a vital role in understanding the relationship between the probability of failure of a cross-section and the probability of failure of a dike section and trajectory, this part falls outside the scope of this research. This research focuses on translating soil investigation to a probability of failure of a cross-section, as is covered in the section within the dashed line in Figure 1.2.

The last two points are incorporated in Calle et al. (2008) 's "2.5D" model, which is based on Vanmarcke (1983). These two aspects of this model will be described in this chapter.

2.1.3. Variance and Mean Reduction

The current method, described by Calle et al. (2008), assumes that the vertical scale of variation in soil properties is typically smaller than the thickness of the layer intersected by the failure plane. Consequently, the model assumes that vertical variation can be averaged out entirely, a concept referred to as variance reduction (see Figure 2.2). For a more detailed description of variance reduction, please refer to Calle et al. (2021). The current method (Calle et al., 2008) states that actual finite failure could only occur in a longitudinally unstable zone. In this zone, the average soil strength in the longitudinal direction along the failure surface is below the required stability limit. Therefore, no averaging is considered in the horizontal direction parallel to the dike, as the scale of fluctuation in that direction is often significantly larger than the width of the failure surface (Schweckendiek et al., 2017). However, due to these assumptions, this method fails to account for the tendency of failure to propagate through weaker

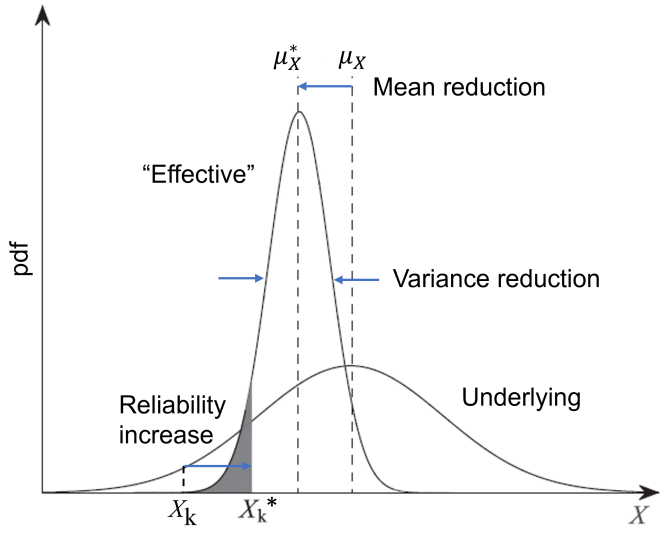


Figure 2.2: The "effective" distribution accounts for the effects of spatial correlation and the problem being studied in the derivation of characteristic property value. Modified from Hicks et al. (2019).

zones, which would reduce the mean strength. Depending on the specific problem, this exclusion may lead to overly optimistic values.

2.1.4. Ratio between Local and Regional Variance

The theory behind the current model of including spatial variability is extensively described by Calle et al. (2021). This section presents an overview of the current method for incorporating local and regional variance in three directions, as depicted in Figure 2.3.

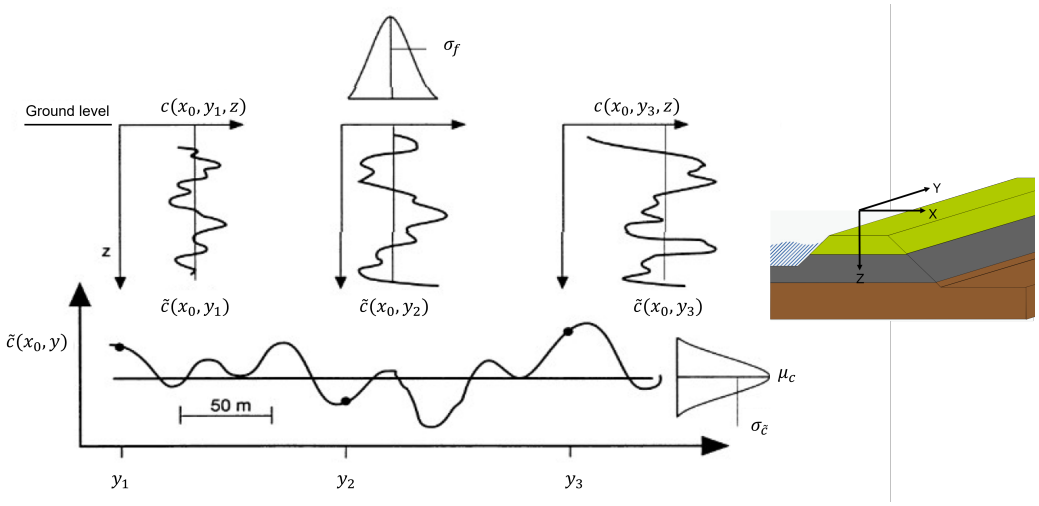


Figure 2.3: Axes along the dike and the fluctuations of a soil parameter along different directions. Adapted from Schweckendiek et al. (2017) and Calle et al. (2021).

The proposed model consists of two separate components. The first component is a three-dimensional local variation, which is represented by a function $f(x, y, z)$. The second component is a two-dimensional variation represented by another function $\tilde{c}(x, y)$. The fluctuation rate in the horizontal directions x and y are assumed to be equal. To obtain the final combined three-dimensional field $c(x, y, z)$, the two components are added together:

$$c(x,y,z) = \tilde{c}(x,y) + f(x,y,z) \tag{2.1}$$

The two-dimensional horizontal component $\tilde{c}(x,y)$ and the three-dimensional component $f(x,y,z)$ are assumed to have a normal distribution and are treated as statistically independent. The total variance of the combined field $c(x,y,z)$ can therefore be expressed as the sum of the variances of the two- and three-dimensional components:

$$\sigma_c^2 = \sigma_{\tilde{c}}^2 + \sigma_f^2 \tag{2.2}$$

Where:

σ_c^2 is the total variance.
$\sigma_{\tilde{c}}^2$ is the variance of the two-dimensional horizontal component.
σ_f^2 is the variance of the three-dimensional local component.

To incorporate these components in the current method, a factor α has been introduced in the determination of soil parameters. α is defined by Calle et al. (2021) as the ratio between the variance of the three-dimensional local component and the total variance. In other official documents, such as Rijkswaterstaat (2021), it is also referred to as the ratio between local and regional variance.

$$\alpha = \frac{\sigma_f^2}{\sigma_c^2} = \frac{\sigma_{loc}^2}{\sigma_{reg}^2} \tag{2.3}$$

Where:

σ_{loc}^2 is the local variance.
σ_{reg}^2 is the regional variance.

Because in the current method, ”local” is defined at the scale of a failure surface, the current model assumes that local variance can be completely averaged out (Calle et al., 2021). Therefore, α also determines which part of the sample variance can be averaged out. This parameter plays a role in converting sample data from soil investigation into input suitable for calculations, which are characteristic values for a semi-probabilistic analysis or mean and standard deviation for a probabilistic or FORM analysis.

The default value of α, set at 0.75, was initially established by TAW (1989) based on intuition due to the absence of adequate data, time, and techniques (Calle et al., 2021). Subsequently, Calle (2007) and Calle (2008) used a statistical approach and determined that this value is plausible yet conservative. However, it is worth noting that the suitability and the level of prudence of the α value varies depending on the regional dataset being considered (Calle et al., 2008).

2.1.5. Adjustment of Soil Parameter Input

As explained in the previous paragraph, the sample mean and standard deviation or characteristic value are adjusted for spatial variation before being used as input for (semi-)probabilistic calculations. This paragraph presents the equations used to determine the calculation inputs.

Semi-Probabilistic Analysis

Rijkswaterstaat (2021) states that the friction angle ϕ, the pre-consolidation stress σ_p' and SHANSEP parameters S and m should be modeled with a 5% lower bound value of a lognormal distribution. If the parameter is lognormal distributed, the characteristic value of the median ($X_{med,char}$) can be calculated with Equation 2.4. In addition to adjusting for spatial variation, this formulation accounts for uncertainty due to limited data points using student-t correction (statistical uncertainty).

$$X_{med,char} = \exp\left(\mu_{\ln X} \pm t_{n-1}\sigma_{\ln X}\sqrt{(1-\alpha)+\frac{1}{n}}\right) \tag{2.4}$$

Where:

$X_{med,char}$	is the characteristic value of the median lognormal distributed parameter X that accounts for local averaging.
t_{n-1}	is the student t-factor corresponding to the 5% quantile and $n-1$ degrees of freedom.
n	is the number of independent data points.

And where $\mu_{\ln X}$ and $\sigma_{\ln X}$ are the natural logarithm of the sample mean and standard deviation, respectively, and can be calculated by:

$$\mu_{\ln X} = \frac{\sum \ln X_i}{n} \tag{2.5}$$

$$\sigma_{\ln X} = \sqrt{\frac{\sum (\ln X_i - \mu_{\ln X})^2}{n-1}} \tag{2.6}$$

$$\tag{2.7}$$

When using the default value of 0.75, the characteristic value becomes lower than when using $\alpha = 1$. A lower value is a more conservative input, resulting in a more conservative estimation of the Factor of Safety (FoS).

Probabilistic and FORM Analysis

When performing a probabilistic analysis, the input required includes the mean (μ_{prob}) and standard deviation ($\sigma_{x,prob}$) of different soil parameters. If a lognormal distribution is assumed, the mean and standard deviation can be calculated by:

$$\sigma_{\ln X,prob} \approx \frac{T_{n-1}^{0.05}}{u^{0.05}}\sigma_{\ln X}\sqrt{(1-\alpha)+\frac{1}{n}} \tag{2.8}$$

$$\mu_{x,prob} = \exp\left(\mu_{\ln X} + \frac{1}{2}\sigma_{\ln X,prob}^2\right) \tag{2.9}$$

$$\sigma_{x,prob} \approx \mu_{x,prob}\left(\sqrt{\exp\left(\sigma_{\ln X,prob}^2\right)-1}\right) \tag{2.10}$$

Where:

$\sigma_{\ln X,prob}$	is the corrected standard deviation of a lognormal distributed parameter.
$\sigma_{X,loc,avg,nor}$	is the reduced standard deviation of normal distributed parameter X that accounts for local averaging.
$T_{n-1}^{0.05}$	is the student t-factor corresponding to the 5% quantile and $n-1$ degrees of freedom, where n is the number of measurements.
$u^{0.05}$	is the value that corresponds to the 5% quantile of the normal distribution. This value is equal to 1.65.
$\sigma_{\ln X}$	is the standard deviation of the log of a measurement set. It is defined in Equation 2.6
n	is the number of data points.
$\mu_{x,prob}$	is the corrected mean of a log-normally distributed parameter. This value should be used as an input for probabilistic calculations.
$\mu_{\ln X}$	is the mean of the log of a measurement set. It is defined in Equation 2.5.
$\sigma_{x,prob}$	is the corrected standard deviation of a log-normally distributed parameter. This value should be used as an input for probabilistic calculations.

When the default value of 0.75 is used, the standard deviation used as input becomes larger than when $\alpha = 1$ is used. A wider input distribution of soil parameters leads to lower values, which results in a more conservative estimation of the Factor of Safety (FoS).

Influence of α

α is a parameter that determines the part of the sample variance that can be averaged out. This affects the formulas for the standard deviation and characteristic value for calculations (see Equations 2.4 and 2.8). The impact of α on the input becomes evident when examining different values:

- *When* $\alpha = 1.0$: the regional variance equals the local variance. This means that the measured sample variance can be completely averaged out in the calculation. Suppose the number of points in your dataset becomes large enough. In that case, the input standard deviation goes to zero, and the characteristic value becomes the sample mean (transformed to a lognormal distribution).

$$\sigma_{\ln X,prob} \approx \frac{T_{n-1}^{0.05}}{u^{0.05}} \sigma_{\ln X} \sqrt{\frac{1}{n}} \xrightarrow{n\to\infty} 0 \tag{2.11}$$

- *When* $\alpha = 0$: the local variance becomes zero. Therefore, the sample variance cannot be reduced when implemented in the calculations. Suppose the number of points in your dataset becomes large enough. In that case, the input standard deviation becomes the sample standard deviation, and the characteristic value becomes the 5% quantile of the sample distribution (both transformed to a lognormal distribution).

$$\sigma_{\ln X,prob} \approx \frac{T_{n-1}^{0.05}}{u^{0.05}} \sigma_{\ln X} \sqrt{1 + \frac{1}{n}} \xrightarrow{n\to\infty} \sigma_{\ln X} \tag{2.12}$$

- *When* $\alpha = 0.75$: the local variance is 0.75 times the regional variance. So, a part of the sample variance can be averaged out. Suppose the number of points in your dataset becomes large enough. In that case, the input standard deviation becomes half of the sample standard deviation when transformed to a lognormal distribution, and the sample standard deviation only contributes half of its value to the characteristic value.

$$\sigma_{\ln X,prob} \approx \frac{T_{n-1}^{0.05}}{u^{0.05}} \sigma_{\ln X} \sqrt{(1 - 0.75) + \frac{1}{n}} \xrightarrow{n\to\infty} \frac{1}{2} \sigma_{\ln X} \tag{2.13}$$

This means that:

$$\sigma_{\ln X,prob}^{\alpha=1.0} < \sigma_{\ln X,prob}^{\alpha=0.75} < \sigma_{\ln X,prob}^{\alpha=0} \tag{2.14}$$

$$X_{med,char}^{\alpha=1.0} > X_{med,char}^{\alpha=0.75} > X_{med,char}^{\alpha=0} \tag{2.15}$$

2.2. Geostatistical Models

It is common practice in geotechnical design to consider uniform soil properties within distinct soil layers or materials. However, laboratory tests on soil samples and in situ tests conducted at different locations within these units show significant variability in the outcomes. Since the 1960s, researchers have proposed to model the variability as simple univariate statistics (Calle et al., 2021).

A soil property is often modeled as the sum of the trend function (t) and the spatial variable residual (ε). The predicted measurement of a soil property ($\tilde{X}_i$) at depth z_i can be calculated by (Ching et al., 2017):

$$\tilde{X}_i = t(z_i) + \varepsilon(z_i) \tag{2.16}$$

Spatial variability contains more than variations in soil properties at different depths; it extends throughout a three-dimensional domain. Figure 2.3 displays the definition of the axes used in this context. Typically, due to the specific depositional environment of the soil, fluctuations in the vertical direction are more pronounced compared to the horizontal direction. This behavior is depicted in Figure 2.3.

Vanmarcke (1977) modeled the residual as a stochastic field defined by a mean of zero, a standard deviation, and a correlation. The amount of correlation between variables X_1 and X_2 is expressed by the correlation coefficient ρ (Baecher & Christian, 2003):

$$\rho = \frac{\mathrm{Cov}(X_1, X_2)}{\sqrt{\mathrm{Var}(X_1)\mathrm{Var}(X_2)}} \tag{2.17}$$

An Auto Correlation Function (ACF) describes the correlation coefficient as a function of the lag (τ). The lag is a discrete distance between measurements. For example, if two points have a distance of $0.2\mathrm{m}$ between them, the correlation is defined by (τ_1), where $\tau_1 = 0.2\mathrm{m}$. The ACF defines the correlation between these points by assigning a numerical value that ranges between -1 and 1. 1 indicates a strong positive correlation, -1 indicates a strong negative correlation, and 0 indicates no correlation or independence (Baecher & Christian, 2003).

The shape of the ACF curve provides information about the spatial correlation structure. For example, a rapidly decreasing ACF curve indicates that properties at nearby locations are highly correlated, while a slower decay suggests a more extended range of correlation.

One input of the ACF is the scale of fluctuation (θ). It is defined as the distance over which a soil's properties are comparable or correlated (Cami et al., 2020). In Paragraph 2.2.1, the scale of fluctuation is discussed in more detail. Possible approaches for determining the scale of fluctuation and its relationship to theACF are provided later in this chapter.

2.2.1. Scale of Fluctuation

The distance over which a soil's properties are correlated is called the scale of fluctuation θ (Cami et al., 2020). The scale of fluctuation is given as an input of the ACF, see Table 2.3. Vanmarcke (1983) defined θ as the area under the autocorrelation function:

$$\theta = \int_{-\infty}^{\infty} \rho(\tau) d\tau = 2 \int_{0}^{\infty} \rho(\tau) d\tau \tag{2.18}$$

Table 2.1: Summary of scales of fluctuation (θ) from a literature review conducted by Cami et al. (2020) per soil types (meters). The value of θ is primarily determined by cone resistance data from CPT measurements.

	Horizontal θ				**Vertical θ**			
Soil type	**Number**	**Min**	**Max**	**Average**	**Number**	**Min**	**Max**	**Average**
Alluvial	9	1.07	49	14.2	13	0.07	1.1	0.36
Ankara clay	-	-	-	-	4	1	6.2	3.63
Chicago clay	-	-	-	-	2	0.79	1.25	0.91
Clay	9	0.14	163.8	31.9	16	0.05	3.62	1.29
Clay, sand, and silt mix	13	1.2	1,000	201.5	28	0.06	21	1.58
Hangzhou clay	2	40.4	45.4	42.9	4	0.49	0.77	0.63
Marine clay	8	8.37	66	30.9	9	0.11	6.1	1.55
Marine sand	1	15	15	15	5	0.07	7.2	1.43
Offshore soil	1	24.6	66.5	45.6	2	0.48	1.62	1.04
Overconsolidated clay	1	0.14	0.14	0.14	2	0.063	0.255	0.15
Sand	9	1.69	80	24.5	14	0.1	4	1.17
Sensitive clay	-	-	-	-	2	1.1	2.0	1.55
Silt	3	12.7	45.5	33.2	5	0.14	7.19	2.08
Silty clay	7	9.65	45.4	29.8	14	0.095	6.47	1.4
Soft clay	3	22.2	80	47.6	8	0.14	6.2	1.7
Undrained engineered soil	-	-	-	-	22	0.3	2.7	1.42

Table 2.3: Common autocorrelation models. Adapted from (Cami et al., 2020). For the Matérn function, ν is the smoothness parameter, ρ is the scale, and K_v is the Bessel function of the second kind.

Autocorrelation model	**Correlation as a function of the lag τ**
Markovian (single exponential)	$\rho(\tau) = \exp\left(-\frac{2\lvert\tau\rvert}{\theta}\right)$
Second-order Markovian	$\rho(\tau) = \left(1 + \frac{4\lvert\tau\rvert}{\theta}\right)\exp\left(-\frac{4\lvert\tau\rvert}{\theta}\right)$
Third-order Markovian	$\rho(\tau) = \left(1 + \frac{16\lvert\tau\rvert}{3\theta} + \frac{256}{27}(\frac{\lvert\tau\rvert}{\theta})^2\right)\exp\left(-\frac{16\lvert\tau\rvert}{3\theta}\right)$
Gaussian (squared exponential)	$\rho(\tau) = \exp\left(-\pi(\frac{\lvert\tau\rvert}{\theta})^2\right)$
Spherical	$\rho(\tau) = \begin{cases} \frac{4}{3} - 2\lvert\frac{\tau}{\theta}\rvert + \frac{2}{3}\lvert\frac{\tau}{\theta}\rvert^3 & \text{if } \lvert\tau\rvert \leq \theta \\ 0 & \text{otherwise} \end{cases}$
Cosine exponential	$\rho(\tau) = \exp\left(-\frac{\lvert\tau\rvert}{\theta}\cos\frac{\lvert\tau\rvert}{\theta}\right)$
Binary noise	$\rho(\tau) = \begin{cases} 1 - \frac{\lvert\tau\rvert}{\theta} & \text{if } \lvert\tau\rvert \leq \theta \\ 0 & \text{otherwise} \end{cases}$
Whittle-Mátern	$\rho(\tau) = \frac{2}{\Gamma(\nu)}\left(\frac{\sqrt{\pi}\Gamma(\nu+0.5)\lvert\tau\rvert}{\Gamma(\nu)\theta}\right)^{\nu} K_v\left(\frac{\sqrt{\pi}\Gamma(\nu+0.5)\lvert\tau\rvert}{\Gamma(\nu)\theta}\right)$

Worst-Case Scale of Fluctuation

In design and stability calculations, the ratio of θ to the size of the structure is important. If the scale of fluctuation is much larger than the structure, it can be considered practically infinite. This means that the soil volume influencing soil-structure interaction can be seen as homogeneous. However, when this is not the case, the previous method is not applicable, especially for failure mechanisms like slope failure, which has an emerging failure path. In this instance, a 'worst-case' θ should be considered. Fenton and Griffiths (2003) defined the worst-case scale of fluctuation as: "the scale of fluctuation value that results in the highest probability of failure".

An example of this is a study conducted by Ali et al. (2014), where the influence of a spatially varying hydraulic conductivity on the risk associated with rainfall-induced landslides was investigated. In the case of an infinite slope, they found that the critical $\theta_z = H$, where H is the height of the slope and θ_z is the scale of fluctuation in the vertical direction z.

In a study by De Gast et al. (2019), the impact of displacement on horizontal scales of fluctuation was examined. The research revealed that the construction process of the dike, including its loading history, significantly affects the horizontal scale of fluctuation of the undrained strength parameter s_u. Specifically, in the direction perpendicular to the dike, it was observed that θ_x varied depending on the compressibility of the soil type. The determined values of θ_x ranged from half the width of the dike to slightly larger than the width.

Identifiability Problem

Ching et al. (2017) introduced the concept of the identifiability problem, which occurs when estimating θ becomes difficult due to limited data availability. Accurate determination becomes difficult when θ is smaller than the spacing between data points. In these cases, the worst-case scale of fluctuation serves as a valuable tool for design purposes (Cami et al., 2020).

Currently, there is a lack of quantitative research regarding the scales of fluctuation for different soil characteristics. Therefore, it is unclear how scales of fluctuation relate to one another. Although the general assumption suggests that a single autocorrelation model governs all soil parameters (Cami et al., 2020), further studies are required to validate and improve our understanding of this assumption.

Estimation of the Scale of Fluctuation with Data

Several methods have been developed to characterize the fluctuation parameter scale from soil data, particularly using Cone Penetration Test (CPT) measurements. The following methods can be utilized to estimate the scale of fluctuation from soil investigation:

1. Method of Moments (MoM): This approach utilizes statistical moments to estimate the scale of

fluctuation. This includes the autocorrelation function, autocovariance function, and semivariogram.

2. Maximum-Likelihood Estimation (MLE): MLE maximizes the likelihood function to determine the most likely value for the scale of fluctuation.
3. Bayesian analysis: Bayesian analysis incorporates prior knowledge and data to estimate the scale of fluctuation. This method provides a probabilistic framework for estimating the parameter.

The method of moments (MoM) is the most used, as stated by (Cami et al., 2020). While MLE may be more precise, it is susceptible to theoretical ACF functions that fit only one parameter. The Whittle-Mátern function introduces a second parameter better suited for MLE. However, this study aims to investigate different ACF functions, and therefore, the MoM will be applied.

The following paragraph will provide a further explanation of this method.

Implementation of θ in the Current Method

In the Netherlands, the stability assessment method for dikes also includes the scale of fluctuation θ. This parameter is used in the ACF of the combined field $c(x, y, z)$, which is explained in Paragraph 2.1 and Calle et al. (2021).

The ACF is defined in three directions of the dike, as shown in Figure 2.3, and is calculated using Equation 2.19. The ACF is not directly used in determining soil parameter input for stability assessments but is part of the theoretical basis of Equations 2.4 and 2.8, as explained in Calle et al. (2021). Two assumptions are made in this method: (1) the horizontal scales of fluctuation (θ_h) in the two horizontal directions (x and y) are equal, and (2) a Gaussian ACF is used.

$$\rho_c = \exp\left(-\pi \frac{\Delta x^2 + \Delta y^2}{\theta_h^2}\right)\left((1-\alpha) + \alpha \exp\left(-\pi \frac{\Delta z^2}{\theta_v^2}\right)\right) \tag{2.19}$$

Here, the parameter α represents the ratio between the local variance (σ_f) and the total regional variance (σ_c).

Please note that the Equation 2.19 provided by Calle et al. (2021) uses a correlation length d. For a Gaussian ACF, $\theta = d\sqrt{\pi}$.

2.2.2. Method of Moments

Vanmarcke (1977) introduced the method of moments to estimate θ. With this method, the scale of fluctuation can directly be approximated from the characteristics of the variance function or ACF (Christodoulou et al., 2021). The ACF described in Equation 2.17 is often hard to determine because of limited data (Phoon et al., 2006). Therefore, an approximation $\hat{\rho}$ can be used. The first equation is suitable for equally spaced data, while the second formulation can be used for non-uniformly distributed spaced data:

$$\hat{\rho}(\tau_j) = \frac{\sum_{i=1}^{k-j}(X_i - \mu_X)(X_{i+j} - \mu_X)}{\sigma_X^2 k} \qquad j = 0, \ldots, k-1 \tag{2.20}$$

$$\hat{\rho}(\tau_j) = \frac{\sum_{i=1}^{t}(X_i - \mu_X)(X_{i+\Delta i} - \mu_X)}{\sigma_X^2 (t-1)} \tag{2.21}$$

Where:

$\hat{\rho}(\tau_j)$	is the experimental or sample correlation coefficient between two points that are τ_j apart.
k	is the number of sample points [-].
i	is the index of the point pair at lag distance τ_j.
t	is the number of point pairs at lag distance τ_j.
Δi	is the index of the distance between the points of a point pair for unequally distributed data.

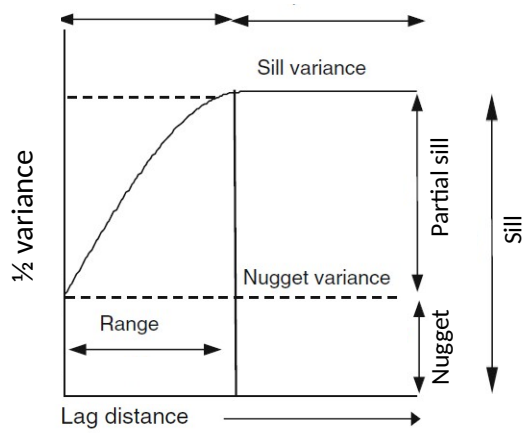


Figure 2.4: Semivariogram and its parameters. Adapted from Oliver and Webster (2015).

With the Method of Moments, θ can be found by minimizing the error between the experimental moments $\rho(\tau)$ and the theoretical moments (Cami et al., 2020) as described in Table 2.3. ”moments” refers to the statistical quantities used to match a theoretical model’s properties to the observed data.

This technique may be divided into two categories: semivariogram or variogram fitting and autocorrelation or autocovariance function fitting (Cami et al., 2020). In this report, only the semivariogram and autocorrelation will be investigated.

Semivariogram Fitting

The semivariogram was first introduced by Krige (1951) to determine the spatial correlations for gold mines in South Africa. Semivariograms are often used to define the spatial correlation in geology but are not often used in geotechnical engineering (Cami et al., 2020). The semivariogram $\gamma(\tau)$ expresses the average dissimilarity of a random variable X value between sample points at various distances (u_i):

$$\gamma(\tau) = \frac{1}{2}\mathrm{Var}\left[X(u_i) - X(u_i + \tau)\right] \tag{2.22}$$

Where $X(u_i)$ is the value of parameter X at location u_i. Half of the average squared difference in X as a function of the spatial displacement is quantified by the (semi)variogram. Because it is half of the squared difference, it approximates the variance and, therefore, goes by ”semivariance”. ”Semivariogram” will be abbreviated to ”variogram”.

Some important parameters can be defined from the semivariogram: the sill, nugget, and range, see Figure 2.4. They are defined as follows:

- *The Sill (s)*: The semivariogram eventually reaches an upper bound known as the sill. Depending on the variogram, the sill can remain constant or act as an asymptote. It represents the maximum variance of the process. The partial sill is the difference between the nugget and the sill and quantifies the variance due to spatial variability.
- *The Range (r)*: When a semivariogram reaches its sill at a finite lag distance, it has a range. The range defines the limit of spatial correlation, beyond which the autocorrelation becomes 0. Locations that are farther apart than the range are considered spatially uncorrelated or independent ($\mathrm{Cov}(\tau) = 0$). For semivariograms that approach their sills asymptotically, strict ranges do not exist. However, in practice, an effective range is determined at the lag distances where the variograms reach 95% of their sills (Oliver & Webster, 2015). The effective range (r') is defined to approximate the range. How this relates to the range depends on the theoretical function (see Table 2.4).
- *The Nugget (n)*: The variogram often intersects the y-axis at a positive value. The term ”nugget” originated from gold mining, where gold nuggets were found to occur randomly and indepen-

dently of each other. In the context of the semivariogram, the nugget represents a discontinuity in variation, signifying an uncorrelated component. This occurs when the variation at a location is unrelated to the variation at neighboring sites.

These three components, the sill, range of spatial correlation, and nugget provide insights into the behavior and characteristics of the semivariogram. It can help to understand spatial variability and correlations in the soil. The experimental or sample semivariogram is defined as (Cami et al., 2020):

$$\hat{\gamma}(\tau) = \frac{1}{2k}\sum_{i=1}^{k}[X(u_i) - X(u_i + \tau)]^2 \tag{2.23}$$

To define the parameters as mentioned earlier, a theoretical function (see Table 2.4) is fitted to the experimental points calculated by Equation 2.23. The theoretical functions of the ACF in Table 2.3 can also be used as variogram functions when subtracted from the variance.

Table 2.4: Examples of theoretical semi variance models as a function of the lag τ. Where p is the partial sill (p=sill-nugget), r is the range, r' is the effective range, n is the nugget. h is the shape parameter of the Stable function, which determines if the shape of the function looks more like a Gaussian or Exponential and Spherical shape. For the Matérn function, ν is the smoothness parameter, ρ is the scale, and K_v is the Bessel function of the second kind. All the functions are fitted to p, s and n. The Stable function is also fitted to h, and Matérn is also fitted to ν.

Name semivariogram model	Semivariogram function	Range
Gaussian (Allard, 2013)	$\gamma(\tau) = p\left(1 - \exp\left(-\frac{\tau^2}{(r/2)^2}\right)\right) + n$	$r \approx \frac{r'}{2}$
Exponential (Allard, 2013)	$\gamma(\tau) = p\left(1 - \exp\left(-\frac{\tau^2}{r'/3}\right)\right) + n$	$r \approx \frac{r'}{3}$
Spherical (Burgess & Webster, 1980)	$\gamma(\tau) = \begin{cases} p\left(\frac{3d}{2r} - \frac{\tau^3}{2r^3}\right) + n & \text{if } \tau \leq r \\ p + n & \text{otherwise} \end{cases}$	
Stable	$\gamma(\tau) = n + p\left[1 - \exp\left(-\frac{\tau}{\frac{r'}{3^{1/h}}}\right)^h\right]$	$r \approx \frac{r'}{3^{1/h}}$
Matérn	$\gamma(\tau) = n + p\left[1 - \frac{1}{2^{\nu-1}\Gamma(\nu)}(\frac{\tau}{r'/2})^\nu K_v(\frac{\tau}{r'/2})\right]$	$r \approx \frac{r'}{2}$
Pure nugget	$\gamma(\tau) = n$	

A variogram can be created when the data exhibits intrinsic stationarity (Marchant & Lark, 2004). This concept is closely tied to second-order stationarity. When the data is second-order stationary, the observations at different locations can be treated as sampled from the same stationary process. In this context, a ”process” refers to a collection of random variables organized in time or space. The variance of the difference between the values of a variable at two different places is influenced solely by the lag vector between these locations, not their absolute positions. This assumption rests on the idea that statistical characteristics like mean, variance, and autocorrelation remain constant or unvarying with location. If the increments of $X(u)$ are second-order stationary, the process can be defined as intrinsically stationary.

To create a variogram and determine its parameters, the following steps should be performed:

1. Detrending;
2. Binning the variogram cloud;
3. Selecting an appropriate bin width;
4. Select an appropriate cutoff value;
5. Fitting theoretical functions.

For a more detailed description of the steps, please refer to Appendix A.4.

Autocorrelation Function Fitting

The fitting described above can also be done for the ACF instead of the semivariogram. In this case, the error between the experimental ACF and a theoretical ACF (Table 2.3) is defined in Equation 2.24.

$$\text{Error} = \sum_{i=1}^{k} [\hat{\rho}(\tau_i) - \rho(\tau_i)]^2 \tag{2.24}$$

Autocorrelation and autocovariance are limited to second-order stationary processes, while variograms apply to intrinsically stationary processes. This implies that variograms are more widely applicable (Baecher & Christian, 2003).

2.3. RFEM

The Finite Element Method (FEM) is a numerical model that evaluates the response of the dike by solving the partial differential equations within the predefined elements that make up the dike's geometry (the mesh). When combined with the technique of generating random fields of soil parameters, it is known as the Random Finite Element Method (RFEM) (Hicks & Li, 2018). The random field represents the spatial variability of soil by modeling stronger and weaker zones in a soil layer, as illustrated in Figure 2.5. The primary advantage of RFEM over traditional Limit Equilibrium Method (LEM) and FEM is that the failure surface can "seek out" the weaker zones and propagate through them (Fenton & Griffiths, 2008). This makes the analysis more realistic and accurate. By generating a large number of realizations of random fields using a Monte Carlo analysis, the failure probability of a dike can be estimated while considering the spatial variability.

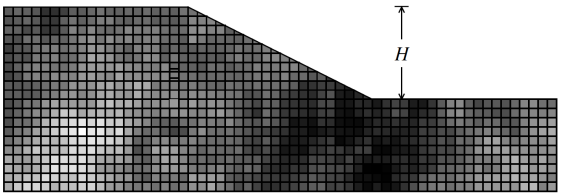


Figure 2.5: Example of a realization of a RFEM random field. The strong zones are depicted in black, while the weak zones are white. Adapted from Fenton and Griffiths (2008).

This chapter explains the general concept of an RFEM analysis by following the flowchart shown in Figure 2.6. The process involves four main steps:

1. Firstly, soil investigation is conducted, and the dataset of soil parameters is analyzed to determine the marginal distribution for one or multiple soil parameters X. An appropriate spatial correlation structure should also be defined, as this is crucial input for the RFEM analysis. More information about geostatistical methods can be found in Chapter 2.2.
2. Based on the point statistics and the spatial correlation structure, a random field can be generated, as will be explained in Paragraph 2.3.2.
3. A realization of a random field is created and mapped onto a FEM mesh. A numerical analysis is performed to calculate the system response to this realization, and the outcome of this analysis is a single Factor of Safety (FoS). This step is repeated multiple times.
4. Finally, the distribution of the calculated factors of safety is analyzed, and by dividing the number of failed realizations (FoS < 1), the failure probability of the dike can be calculated.

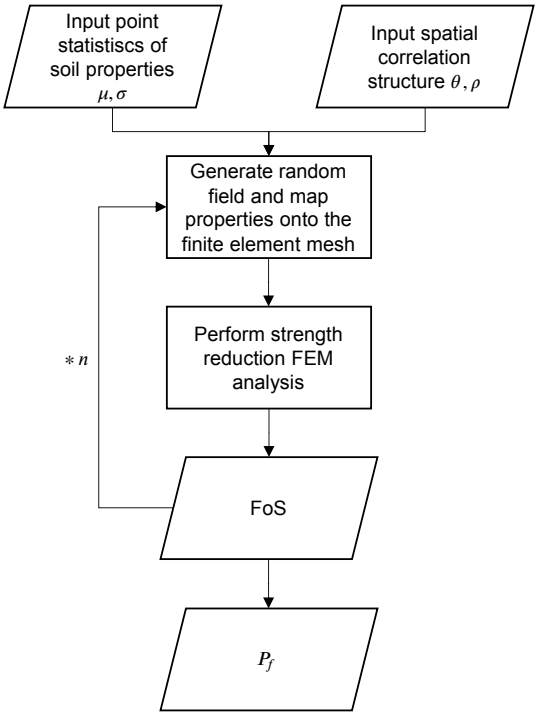


Figure 2.6: General flowchart RFEM.

2.3.1. FEM

The Finite Element Method (FEM) was developed in the 1960s to approximate solutions that could not be calculated with classical analytical solutions. To solve complicated physical processes, like slope stability under gravitational loading, FEM uses numerical techniques. Geometries are split into discrete linked areas called elements, which form the mesh. The response of the entire domain to a particular set of boundary conditions is then determined by solving a group of partial differential equations at the element level. The differential equation that is solved for is known as the governing equation and can be formulated as:

$$\left.\begin{aligned} \frac{\delta\sigma'_x}{\delta_x} + \frac{\delta\tau_{xy}}{\delta_y} + F_x = 0 \\ \frac{\delta\tau_{xy}}{\delta_x} + \frac{\delta\sigma'_y}{\delta_y} + F_y = 0 \end{aligned}\right\} \rightarrow \left[\mathbf{A}\right]^T \left\{\boldsymbol{\sigma}\right\} = -\left\{\boldsymbol{F}\right\} \tag{2.25}$$

Where:

σ_x and σ_y	are the effective stresses in the x- and y-direction respectively [kPa].
τ_{xy}	is the shear stress [kPa].
F_x and F_y	are the body "forces" in the direction respectively [$\mathrm{force/m^3}$]

An additional advantage of using FEM, next to the free shape of the failure surface, is the ability to model complex material behavior by employing advanced constitutive models (Hicks & Spencer, 2010).

This research uses a non-linear Finite Element Method (FEM) that uses quadrilateral elements with eight nodes and four integration points. The algorithm used is based on the process outlined in the book written by Smith et al. (2014). According to Griffiths and Lane (1999), the FEM algorithm involves the following general steps:

1. The calculation of the normal and shear stresses at each integration point in the mesh, assuming that the soil behaves elastically.

2. The comparison of these stresses with the Mohr-Coulomb failure criteria at each integration point in the mesh. If a point falls inside the failure criteria, it will behave elastically.
3. Yielding occurs at the points that are at the failure envelope. A visco-plastic algorithm, as described by Zienkiewicz and Cormeau (1974), redistributes the yielding stresses to neighboring integration points and calculates the plastic strains.
4. When enough integration points have yielded, a failure surface can develop, resulting in overall shear failure.

Slope Stability Analysis

The stability of a dike is assessed using a strength reduction algorithm in FEM, as described by Matsui and San (1992). This thesis uses a strength reduction algorithm developed by TU Delft. This algorithm combines strength reduction with Mohr-Coulomb and SHANSEP. Smith et al. (2014) presents this algorithm in program 6.4. Varkey et al. (2023) describes the extension of this algorithm with SHANSEP. The first three steps described in the previous section calculate the strains in the integration points.

The loading stages are also described by Varkey et al. (2023). The application of gravity loading generates the initial stress state. Subsequently, two stages of loading are applied to consider the normal daily water level condition in the first stage, followed by a high water level condition.

The user can choose between drained or undrained analysis per layer. The drained shear strength calculation uses Mohr-Coulomb, defined in Equation A.12 in Appendix A.3. The undrained shear strength is calculated using SHANSEP, which is defined for the FEM code in Equation 2.26.

$$s_u = \sigma'_1 * S * OCR^m \tag{2.26}$$

$$OCR = \frac{\sigma'_{1,max}}{\sigma'_1} \tag{2.27}$$

The ratio between the strength and the loading of the dike is the Factor of Safety (FoS). In FEM, the FoS is determined by reducing the strength of the structure by a factor known as the Strength Reduction Factor (SRF). The SRF is gradually increased until the point where the numerical model fails to converge, which means that the global equilibrium cannot be reached within a user-defined maximum number of iterations (Griffiths & Lane, 1999).

To find the minimum SRF that leads to failure, the algorithm first evaluates the SRF of 0.5, and if it does not lead to failure, it increases the SRF by values of 0.1 until it does. Once the first SRF that leads to failure is found, the algorithm iteratively searches for the lowest value of SRF that results in failure using a bisection method. The FoS of the dike is then defined by this value (Smith et al., 2014).

In a drained analysis, the SRF is applied to the tangent of the friction angle (ϕ) and the cohesion (c), as presented in Equations 2.28 and 2.29 respectively. In an undrained SHANSEP analysis, the shear strength (s_u) is calculated using Equation 2.26, which is then reduced following Equation 2.30. It is important to note that the same SRF is applied for all the terms.

$$\phi_f = \arctan\left(\frac{\tan\phi}{SRF}\right) \tag{2.28}$$

$$c_f = \frac{c}{SRF} \tag{2.29}$$

$$s_{u,f} = \frac{s_u}{SRF} \tag{2.30}$$

2.3.2. Random Fields

The random process of $X(u)$, such as the cone resistance of a CPT, can be approximated by a Gaussian random field that is second-order stationary, according to Fenton and Griffiths (2008). An example of a random field realization can be seen in the upper plot in Figure 2.7. This type of random field is widely

applicable (Varkey, 2020) and is also used in this study to create the spatial variability in Random Finite Element Method (RFEM). In this research, the random field and a FEM mesh are combined by allocating the cell average of the underlying random field to the integration point (Van Den Eijnden & Hicks, 2017). According to Fenton and Griffiths (2008), to create a random field, the following is needed:

1. The mean μ_X;
2. The variance σ_X^2;
3. The rate at which the random field changes as we move through space.

The covariance function can define the last two. Specifically, the covariance of the values at two points $u = u_i$ and $u = u_j$ can be defined as:

$$\mathrm{Cov}\left[X(u_i), X(u_j)\right] = \mathrm{E}\left[(X(u_i) - \mu(u_i)) * (X(u_j) - \mu(u_j)\right] \tag{2.31}$$

The correlation function, as defined in Equation 2.17, can be rewritten using $\mathrm{E}\left[X(u_i)\right] = \mu(u_i)$ and $\mathrm{Var}\left[X(u_i)\right] = \sigma(u_i)$:

$$\rho\left(X(u_i), X(u_j)\right) = \frac{\mathrm{Cov}\left[X(u_i), X(u_j)\right]}{\sigma(u_i)\sigma(u_j)} \tag{2.32}$$

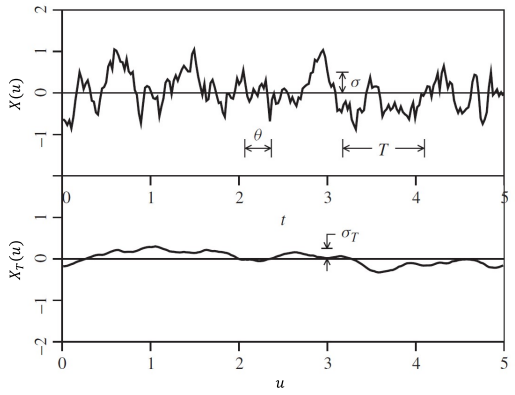


Figure 2.7: The impact of variation on local averaging: The lower plot is obtained by averaging the plot throughout a changing window length denoted by T. Figure from Fenton and Griffiths (2008).

As the random field is second-order stationary, the covariance is only determined by the distance between the points, referred to as the lag (τ), and not the absolute location within the random field. This allows for expressing Equation 2.32 in a generalized form for any given lag (τ):

$$\rho\left(\tau\right) = \frac{C\left(\tau\right)}{\sigma_X^2} \tag{2.33}$$

Where $C(\tau)$ represents the covariance between two points separated by lag τ, and σ_X^2 represents the variance of the random field. The autocorrelation is described using an adapted version of the Markov ACF, which is shown in Table 2.3 and is based on van den Eijnden and Hicks (2019):

$$\rho(\tau) = \omega + (1 - \omega)\exp\left(-\frac{2|\tau|}{\theta}\right) \tag{2.34}$$

Where:

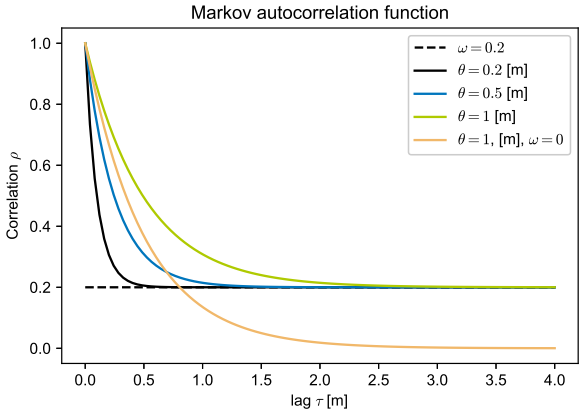


Figure 2.8: Markov autocorrelation function with different values of ω and θ

$\rho(\tau)$ is the correlation between two points with separating lay τ.
ω is residual correlation factor.
θ is the scale of fluctuation [m].

In Chapter 2.2.1, the scale of fluctuation θ is defined as the distance over which a soil's properties are similar or related (Cami et al., 2020). When the scale of fluctuation decreases, the correlation between points at the same lag τ also decreases. Thus, the correlation function ρ decays faster with a smaller θ. In slope stability analyses, Equation 2.34 is defined in all directions, meaning each direction has its own θ.

For the purpose of this study, only two directions are modeled: the vertical and horizontal directions of a cross-section. Even though the ratio between the out-of-plane horizontal scale of fluctuation and the domain size is crucial for the probability of failure of a dike section, as researched by Hicks and Spencer (2010), the out-of-plane spatial correlation is not considered and is assumed to be greater than the width of the failure surface.

When the spatial variability is on a smaller scale than the system domain, averaging occurs. This process is illustrated in Figure 2.7. For example, suppose the variability in a soil layer is smaller than the layer thickness intersected by the failure plane. In that case, there will be a reduction in the variance of the soil properties. This results in a σ_T less than σ, as shown in the figure.

Averaging the variance is only permitted when the scale of fluctuation is sufficiently small relative to the problem size. The measured variance of the samples is defined by local and regional variance. Local variance can be averaged out if the scale of fluctuation is small enough. In contrast, regional variance should not be averaged out since it represents spatial variability on a larger scale than the failure surface domain. Moreover, statistical uncertainty resulting from the limited size of a dataset is an inherent component of variance that cannot be effectively averaged (van den Eijnden & Hicks, 2019).

Figure 2.8 shows the Markov autocorrelation function. To prevent total averaging of the variance, a residual correlation factor called ω is introduced.

The mean and the covariance function define the random field. The covariance function is the product of the correlation function $\rho(\tau)$ and the standard deviation of the random field σ_X^2 (see Equation 2.33). If the scale of fluctuation is small enough compared to the thickness of the layer, the variance of the random field will be averaged out for very large lags when $\rho \to 0$, which is not desirable. Therefore, residual correlation is introduced as ω to prevent the correlation function from going to zero.

Covariance Matrix Decomposition

Random fields are often generated by transforming standard normal random fields (Van Den Eijnden & Hicks, 2017). In this thesis, standard normal random fields are generated by covariance matrix decomposition using Cholesky decomposition. This approach is described in more detail by Van Den Eijnden

and Hicks (2017). To obtain the standard normal random field vector ($\boldsymbol{Z}$), a vector of uncorrelated standard normal random numbers ($\boldsymbol{U}$) is multiplied with the decomposed covariance matrix ($\mathbf{A}$) (Varkey, 2020):

$$\boldsymbol{Z} = \mathbf{A}\boldsymbol{U} \tag{2.35}$$

The covariance of a random field is usually a positive definite covariance matrix $\mathbf{C}$. The elements of this matrix are represented by $C_{ij} = C(\tau_{ij})$, where τ_{ij} is the lag between the points u_i and u_j (as shown in Equation 2.31) (Fenton & Griffiths, 2008). The matrix that is decomposed by Cholesky decomposition ($\mathbf{A}$) can be used to determine $\mathbf{C}$:

$$\mathrm{E}\left[\boldsymbol{Z}\boldsymbol{Z}^{\boldsymbol{T}}\right] = \mathbf{A}\mathbf{A}^{\mathbf{T}} = \mathbf{C} \tag{2.36}$$

According to Van Den Eijnden and Hicks (2017), this method is more exact and poses fewer limitations on the mesh when compared to other methods, such as Local Average Subdivision (Fenton & Vanmarcke, 1990).

Part I

Data Analysis

3

Database

One of the objectives of this research is to explore the spatial variability of soil properties in the Netherlands. This chapter will investigate two soil parameters: the wet volumetric weight γ_{wet} and the SHANSEP parameter S. This will help to verify whether the WBI2017 method of incorporating spatial variation is accurate. Data from two databases are utilized for this purpose. One is on a national scale to investigate any national spatial correlations within the parameters, while the second database is sampled on a more regional scale to compare the spatial variance for the national and regional scales.

This chapter introduces the sources of the data and explains how the data is prepared for use. Moreover, it provides a brief overview of the investigated soil parameters.

For this analysis, two different databases are utilized: (1) the STOWA database (STOWA, n.d.) and (2) the Rivierenland database. The geographical distribution of the data in these databases is illustrated in Figure 3.1.

The first database is a collection of samples assembled by Kindermann and Tigchelaar (2022) on behalf of the Applied Water Management Research Foundation (Stichting Toegepast Onderzoek Waterbeheer) (STOWA). It contains data collected by multiple water boards, including HH Delfland, WS Rivierenland, Wêtterskip Fryslan, HHNK, WDOD, HDSR, HH Rijnland, WS Brabantse Delta, and WS Rijn en IJssel. This database provides geotechnical information both from regional and primary Dutch dikes. This dataset includes data from the following laboratory and in situ tests:

- Classification tests
- Grain size analysis
- Cone Penetration Test (CPT)
- Constain Rate of Strain (CRS) tests
- Oedometer tests
- Direct Simple Shear (DSS) tests
- Triaxial tests

These tests can be used to obtain an understanding of the soil properties and are used in the design and stability calculations of dikes.

The data is consolidated in an Excel sheet, including all the necessary information extracted from these tests. The unfiltered data consists of $30,413$ sample points, each with a distinct (x, y, z) location.

Kindermann and Tigchelaar (2022), on behalf of STOWA, researched the distribution of some soil strength parameters by performing a data analysis on clay and peat. Apart from some conclusions about the mean value, standard deviation and characteristic value of these parameters, some interesting points on the national spatial distribution of the parameters were presented. They concluded that it may be possible to use only one dataset for the whole of the Netherlands since the national

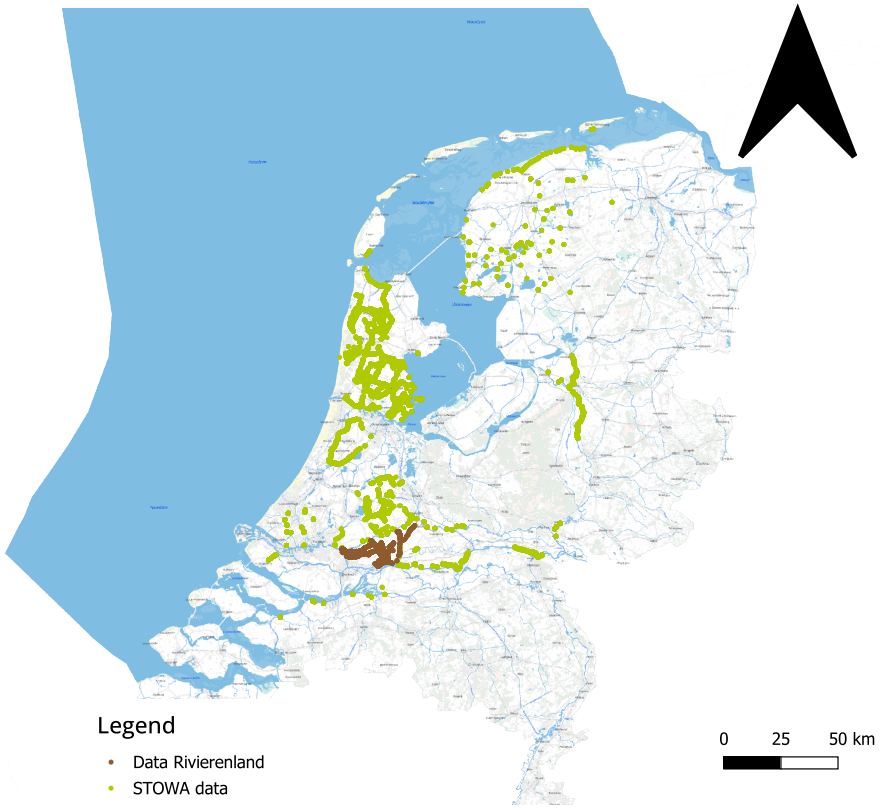


Figure 3.1: Locations of the data points. Green dots illustrate the STOWA database, and brown dots illustrate the Rivierenland database.

spatial distribution of the strength parameters was found to be minimal (see Figure 3.2). However, they recommended further investigation of this hypothesis through a statistical test.

The Rivierenland (RL) database, collected by water board Rivierenland, consists of 2655 datapoints primarily located in the Alblasserwaard. This is a predominantly agricultural area in the south-east of the Dutch province Zuid-Holland. Some data points are located further east, just over the border in the province of Utrecht, along the Merwede channel.

In line with the recommendations of Kindermann and Tigchelaar (2022), a statistical test is performed to further investigate the distribution of some soil strength parameters across the Netherlands. To determine the spatial variability on the national scale, semivariograms from the STOWA dataset were constructed for different soil types (for more information on variograms, please refer to Chapter 2.2.2). The two datasets were kept separate to compare the results on both national and local scales.

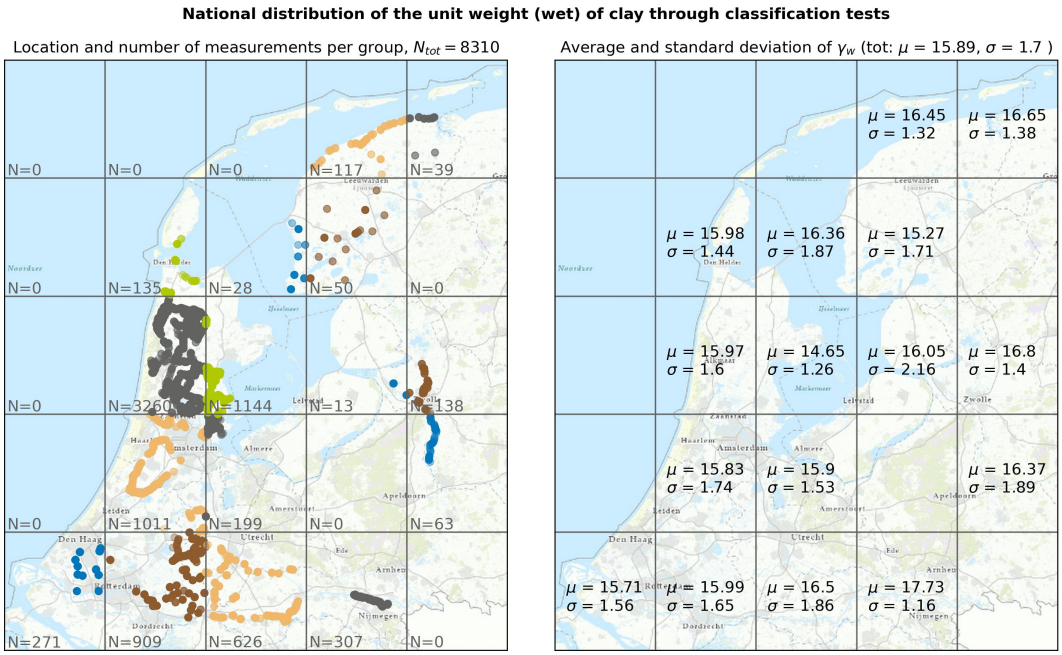


Figure 3.2: National distribution of the wet unit weight of clay. Translated and modified from Kindermann and Tigchelaar (2022). The authors concluded: "It can be seen that the dispersion is generally not very large."

3.1. Interpretation Strategy

In Chapter 2.2.2, the use of the variogram of the Method of Moments (MoM) is discussed. MoM is the most standard method in statistics (Cami et al., 2020) and the variogram is more widely applicable than autocorrelation function fitting (Baecher & Christian, 2003). For a more detailed description of creating variograms, please refer to Chapter 2.2.2 and Appendix A.4.

To analyze the data, the following steps were taken to investigate the spatial variability:

1. Data subtraction:
    - For SHANSEP parameter S:
        (a) The data was subdivided based on the soil type as defined in Paragraph 3.2. The single-stage undrained triaxial tests were used for clays, while for peats, DSS were used. Both were evaluated at large strains, and only the normally consolidated and anisotropically loaded tests were considered when calculating S.
        (b) To handle outliers, the three filters described in 3.4 are used.
    - For the wet volumetric weight γ_{wet}:
        (a) The data was subdivided based on the soil type.
        (b) To handle outliers, the first two filters described in 3.4 are used.
2. Variogram Cloud Construction: A variogram cloud was constructed to visualize the spatial variability.
3. Sample Variogram Construction: An appropriate bin width was selected to create the sample variogram. Subsequently, an appropriate cutoff lag value was chosen.
4. Theoretical Variogram Fitting: The theoretical variogram was fitted to the sample variogram through the following steps:
    (a) An initial guess for the range, sill, and nugget parameters was made based on the results of the sample variogram.

(b) Initially, only the Spherical function was fitted to the variogram using non-linear least squares optimization. The optimized parameters of this function were then used as a refined initial guess for other theoretical functions.

(c) Based on this refined initial guess, the Spherical, Exponential, Stable, and Matérn theoretical functions (see Table 2.4) were fitted to the sample variogram using weighted non-linear least squares optimization. The Gaussian function was not employed because it does not account for the expected discontinuity in the variogram.

(d) The optimal theoretical function is selected based on the WNLS score.

3.2. Soil Classification

The data analysis involved different soil types. However, before 2022, the original soil types in the database did not meet the standard soil description of NEN5104. The soil descriptions were interpreted and adjusted during the preparation of the study conducted by Kindermann and Tigchelaar (2022). Although this standard is no longer in use, it was used for simplicity. This allowed for the direct correlation of inherent parameters, such as S, instead of wet volumetric weight (γ_{wet}), a usual starting parameter for regional datasets.

Clay was further subdivided into silty clay, organic clay and sandy clay. Each subclass was further divided into three sub-classes: weakly (1), medium (2), strongly (3), silty, organic or sandy. While the organic clay can also be subdivided based on its sandy or silty additives, Kindermann and Tigchelaar (2022) found that this subdivision gives a smaller difference in volumetric weight than subdivision with weak, medium and strongly organic. Therefore, the latter subdivision was chosen for performing the analysis.

3.3. Soil Parameters

The data analysis focuses on the volumetric weight (γ_{wet}) and the SHANSEP parameter S. γ_{wet} is selected because a significant amount of data is available from the STOWA and Rivierenland (RL) databases. The dataset from STOWA includes a larger number of triaxial and DSS tests compared to the number of CRS and Oedometer tests, which are necessary for determining the SHANSEP parameter m. Therefore, it was decided to focus on the analysis of SHANSEP parameter S in this research.

3.3.1. Volumetric Weight

For both the Rivierenland and STOWA database, the wet volumetric weight (γ_{wet}) for each soil type was determined by examining the classification tests of the dataset. Here, the volumetric weight was evaluated and recorded for different soil samples. The data was subsequently filtered, with the applied filters described in Paragraph 3.4.

3.3.2. SHANSEP S

The SHANSEP parameter S is an important soil strength parameter. This parameter, defined as the ratio of undrained shear strength (s_u) to the consolidation stress in a normally consolidated state (σ_{vc}), is calculated as $S = \frac{s_u}{\sigma'_{vc}}$.

When assessing Dutch dikes, S must be determined at the critical state, characterized by shearing at "large" strains. For clay and sands, "large" corresponds to 25% axial strain in a single stage triaxial test, while for peat, it corresponds to 40% shear strain during a DSS test (van Duinen, 2012). More detailed information about the SHANSEP parameter can be found in Appendix A.3.3.

In the t, s-space, where the DSS or triaxial test results can be analyzed, S can be computed. Here, the undrained shear strength is equal to half the deviator stress ($s_u = t$), and the consolidation stress can be calculated by adding the mean stress at the end of consolidation to half of the deviator stress at the end of consolidation ($\sigma'_{vc} = s'_c + t_c$). This results in:

$$S = \frac{s_u}{\sigma'_{vc}} = \frac{t}{s'_c + t_c} \tag{3.1}$$

The STOWA dataset provides these values for both single-stage triaxial tests and DSS tests at large strains:

- For clays:
  - t: half the deviator stress at the end of a triaxial test.
  - s'_c: average effective stress at the end of consolidation in a triaxial test.
  - t_c: half the deviator stress at the end of consolidation in a triaxial test.
- For peats:
  - $s_u = \tau_v$: the shear stress at the end of the DSS test.
  - σ'_{vc}: the effective vertical stress during the maximum consolidation stress that occurs during consolidation.

Because S should be determined at a normally consolidated state, only the normally consolidated samples were used in this determination. Further explanation about the filters on the data are given in the next paragraph.

3.4. Filtering the Data

During the study of Kindermann and Tigchelaar (2022), some inconsistencies in the data were found. To ensure sufficient data quality, three specific filters were applied. This section describes these filters in detail because they were also applied in both datasets of this research.

1. Kindermann and Tigchelaar (2022) observed classification errors in soil types, particularly visible in the volumetric weight analysis of clay. Some clay types had unrealistically high volumetric weights, which exceeded physically possible values. Because sandy and silty clays cannot have a water content higher than 150%, these data points with higher values and those with values lower than 10% were deleted.
2. To effectively deal with outliers in the volumetric weight, data points in the 5% percentile and outside the 95% are excluded.
3. For the analysis of S, four additional filters are applied. It is determined that the determination of S requires a single-stage, anisotropically loaded, normally consolidated test conducted at large strains (van Duinen, 2012). Therefore, these conditions are specifically selected. Furthermore, a triaxial test on clay behaves differently when a higher sand content is present in the soil. Consequently, sandy clays were filtered out and data points with a volumetric weight exceeding $17\mathrm{kN/m^3}$ were removed.

4

Results Semivariogram Analysis

In this chapter, the results of the variogram analyses are presented for both the S parameter and γ_{wet}. To provide a clearer picture, an example of the construction of a variogram is given in Chapter 4.1.1. Furthermore, some comparisons between results are made:

1. Soil type versus their sub-classes;
2. With and without filtering for a certain depth range;
3. Horizontal versus vertical variograms;
4. Scale differences (Rivierenland versus STOWA).

4.1. Volumetric Weight

Variograms were constructed for different soil types to investigate spatial correlation for wet volumetric weight γ_{wet}. First, an instructional example for creating a variogram for the clay soil type will be given.

4.1.1. Variogram of clay

To illustrate the variogram analysis process, this example involves the 'k' soil group (clay) within the STOWA database. The filters described in Paragraph 3.4 are applied to obtain these data points.

To create an initial understanding of the distribution of the data, the values of γ_{wet} are visualized on the map of the Netherlands (see Figure 4.1a). There are no clear trends or spatial correlations visible in this visualization. The values for γ_{wet} do not show a distinctive difference when further apart. To illustrate, there is no noticeable difference between the Dutch regions Noord- and Zuid-Holland.

However, it is noteworthy that some values tend to cluster together. Additionally, it is important to acknowledge that the eastern and eastern-southern part (Limburg) of the Netherlands has a limited number of data points. This is consistent across all datasets in this study. Consequently, this region is not well-represented in the analysis.

As no clear correlation can be seen, the spatial correlation is further investigated by constructing a variogram, following the steps described in Chapter 2.2.2.

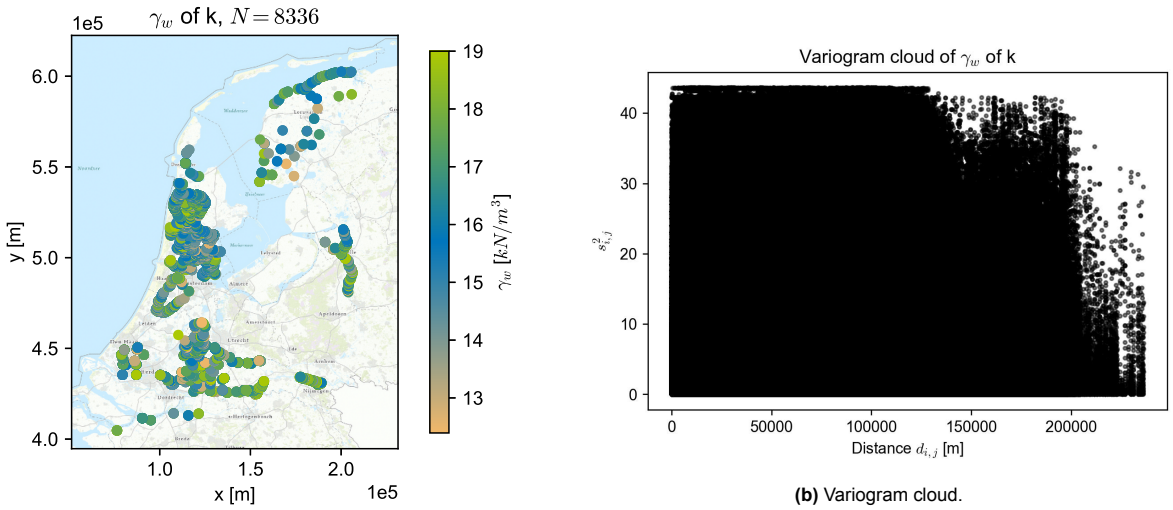
The first step requires creating a variogram cloud by plotting the squared differences in values for all point pairs in the dataset against the distances separating these points. The variogram cloud is displayed in Figure 4.1b). Several observations can be made from this cloud:

- The maximum distance between data points is approximately $236\mathrm{km}$.
- The maximum squared difference in values is around $43.7(\mathrm{kN/m^3})^2$.
- It is clear that the number of data pairs decreases as the distance between them increases.

It should be noted that in this example, only the distances in the horizontal plane are considered. Given that volumetric weight is assessed per soil sample, and each borehole (with x and y coordinates) may

contain multiple soil samples at varying depths (z coordinates), many locations exhibit a distance of 0m. However, the number of point pairs with a 0-meter distance is only 0.08% of the total point pairs.

Due to the large amount of data points in the variogram cloud (around 34.7 million points), no meaningful conclusions can be drawn about spatial correlation. Therefore, it is necessary to bin and average the data to create a better overview.



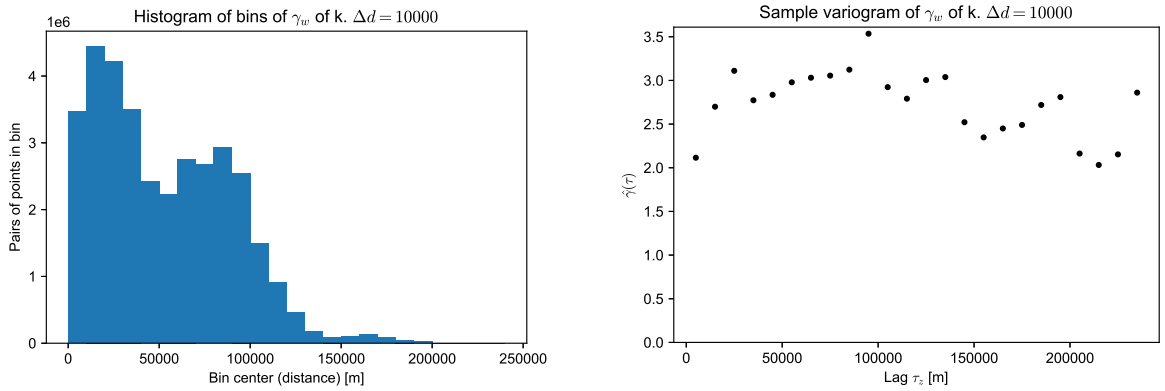
(a) Data points on the map of the Netherlands.

(b) Variogram cloud.

Figure 4.1: Data of γ_{wet} of subset 'k' (clay).

To construct the variogram, the data should be binned based on a specific bin width. Determining the bin width is a process of iterative evaluation, where different bin widths are tested to identify the most appropriate value. Starting with a bin width of 10km, the results are depicted in Figure 4.2. The histogram (Figure 4.2a) shows that for distances up to until 100km, each bin contains more than 1.5 million pairs of data points. This indicates that further refinement is possible by applying a narrower bin width.

It can be observed that the variogram in Figure 4.2b lacks detail. In the first 30km, there is an upward trend in the sample variogram, which gradually flattens beyond this range. Beyond 100km, the points in the sample variogram show weaker alignment with the straight line, and there is more noise. This corresponds with the reduced number of data point pairs in the histogram bins. A smaller bin width is needed to gain more detailed information about the spatial correlation.



(a) Histogram of binned data.

(b) Variogram.

Figure 4.2: Histogram and variogram of γ_{wet} of subset "k" (clay) with a bin width of 10 km.

A higher level of detail is needed to estimate the variogram parameters. This requires selecting a smaller bin width of 900m (see Figure 4.3). The histogram (Figure 4.3a) shows that every bin contains more than $100,000$ point pairs up to a distance of 100km. The sample variogram (Figure 4.3b) shows significant scatter beyond approximately 120km. An appropriate cutoff value is applied to further investigate the beginning of the variogram, which is crucial for parameter determination. The choice of this number depends on how clearly the variogram's initial segment can be seen. In this case, a value of 35km is chosen (see Figure 4.3c). This variogram shows a continuous and increasing trend that appears to approach an asymptote. Using this sample variogram, an initial estimation of the variogram parameters can be made:

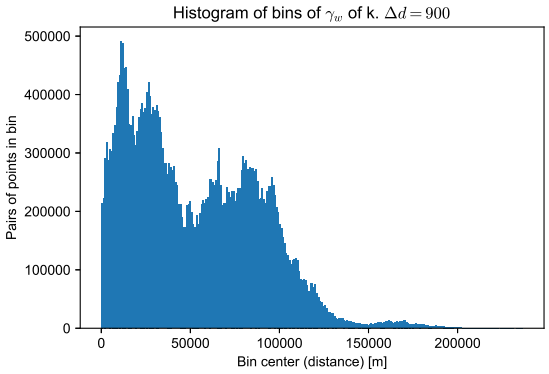
- Range (r) $\approx 18\text{km}$. This means that points within a distance of approximately 18km show spatial correlation. Beyond this range, points are considered spatially uncorrelated or independent.
- Sill (s) ≈ 3.2. The sill describes the maximum variability observed within the spatial domain. It represents the total variance in the dataset, incorporating both spatial and non-spatial sources of variability.
- Nugget (n) ≈ 1.9. The nugget is the intersection with the y-axis, and it represents the non-spatial variance in the measurements. It might indicate that the measurements could not capture correlations at small spatial scales. It is worth noting that the nugget accounts for more than half of the sill, suggesting that the contribution of spatial variability to the overall variance is relatively small.

This study uses Equation A.24 for the weighting scheme. Although Equation A.23 is a stricter weighting scheme, it is only more useful when the number of point pairs decreases rapidly with increasing distance. However, in this study, multiple clusters are illustrated in Figure 4.1a, which means that a more flexible weighting scheme like Equation A.24 would be more appropriate.

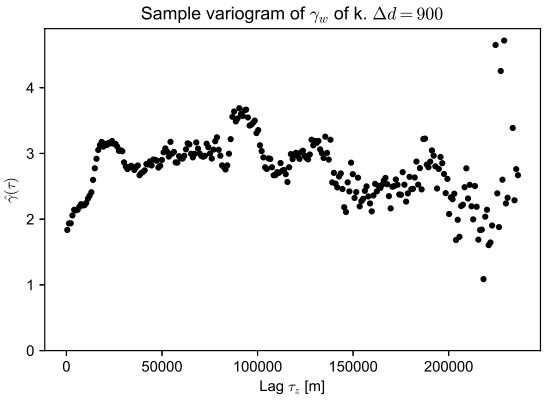
As shown in Figure 4.3b, the variogram has a wave-like shape. Typically, the variogram increases as the distance between points increases since the values become less alike. However, in this case, the variogram decreases after approximately 30km and then increases again at around 80km before decreasing again at 90km. This phenomenon is likely caused by the presence of clusters, as illustrated in Figure 4.1a. The wave-like shape could be due to the fact that the data does not meet the requirement of intrinsic stationarity. For more information on this requirement, please refer to Section 2.2.2. This phenomenon will be discussed in more detail in Chapter 5.3.3.

The next step in creating the variogram involves fitting the theoretical functions to the sample variogram. Optimization of the parameters of these theoretical functions is accomplished through the Weighted Non-Linear Least Squares (WNLS). To determine the best theoretical variogram fit, the WNLS scores are compared among the different theoretical functions. The fit with the smallest WNLS score is selected, and its parameters are saved. From the partial sill p and effective range r', the sill s and range r were calculated:

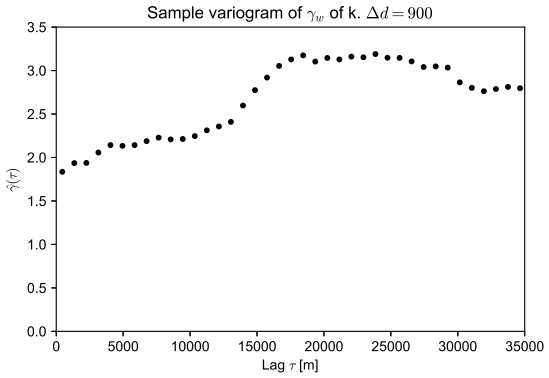
- $r' = 17,635\text{m}$. $r = 14,248\text{m}$.
- $p = 0.94$. $s = 3.00$.
- $n = 2.07$.



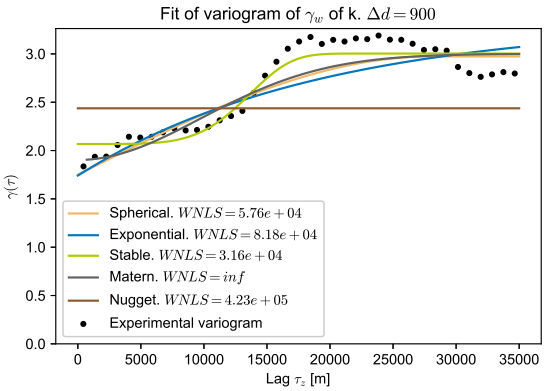
(a) Histogram of binned data.



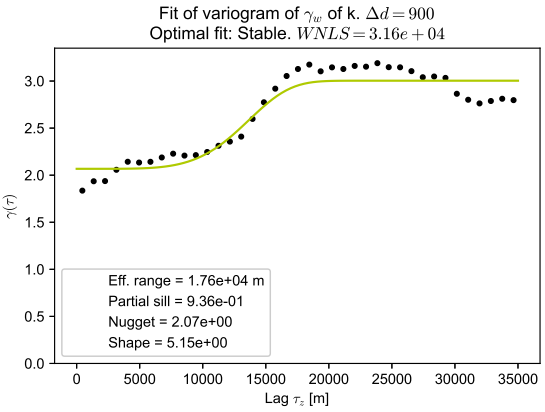
(b) Variogram.



(c) Variogram with a cut off at 35 km.



(d) Variogram with a cut off at 35 km and the fitted theoretical functions.



(e) Variogram with a cut off at 35 km and the optimal fitted theoretical function with its parameters.

Figure 4.3: Histogram and variogram of γ_{wet} of subset 'k' (clay) with a bin width of 900 m.

It can be observed from the results that the spatial correlation of γ_{wet} of clay extends up to 14km, with a maximum total variance of 3.01. The additional variance caused by spatial variation is at most 0.94, whereas the variance due to, for example, measurement uncertainty accounts for more than $2/3$ of the total variance, which is 2.07. Consequently, it can be concluded that the spatial variability is less significant than the sum of other uncertainties. However, it should not be disregarded, and national γ_{wet} measurements of clay should only be used in calculations after accounting for an increased variance due to spatial variability.

This process was repeated for all the remaining soil types. The results can be found in Table 4.1.

Table 4.1: Variogram parameters of the different soil types in the STOWA database for the volumetric weight γ_{wet}.

Soil	Number of points N	Bin width Δd	Effective range r' [m]	Range r [m]	Partial sill p	Sill s	Nugget n	Ratio p/s
k	8336	900	17635.17	14248.06	0.94	3.00	2.07	0.31
ks	1182	2000	20825.21	15876.61	1.08	2.74	1.66	0.39
ks1	142	8000	17492.09	16096.20	1.36	2.65	1.28	0.52
ks2	564	3000	36052.33	18026.16	1.00	2.34	1.34	0.43
ks3	378	2000	4063.35	3557.47	0.52	2.04	1.52	0.26
kh	5680	1000	19280.04	15000.47	1.14	3.21	2.07	0.35
kh1	3349	1000	25323.63	17258.26	1.49	2.76	1.27	0.54
kh2	1561	1500	21110.48	17471.34	0.60	1.58	0.98	0.38
kh3	688	1400	15572.56	12225.11	0.27	1.49	1.22	0.18
kz	596	450	2624.68	1105.83	0.68	1.66	0.97	0.41
kz1	306	10000	19125.34	15803.18	1.08	2.38	1.30	0.45
kz2	118	8000	15154.69	12378.87	0.80	2.23	1.43	0.36
kz3	123	1000	2510.88	2510.88	0.27	0.71	0.45	0.37
v	3625	4000	18002.41	17822.71	0.04	0.41	0.37	0.11
vk	753	2000	17777.29	17777.29	0.09	0.40	0.31	0.22
vk1	404	3000	33447.78	11149.26	0.08	0.32	0.24	0.26
vk2	0	-	-	-	-	-	-	-
vk3	329	1500	2740.45	2176.79	0.09	0.40	0.31	0.23
vm	2563	5000	170907.29	170907.29	0.03	0.21	0.18	0.16

4.1.2. Variogram comparisons

A detailed comparison of variograms will be given in this paragraph to investigate the influence of spatial variability on γ_{wet}. Specifically, the influence of the soil type, depth, and spatial scale on γ_{wet} will be examined. An example comparison using the variograms for clay will be given first, followed by a summary of observations for the other soil types.

In order to compare variograms effectively, a new parameter has been introduced - the ratio between the partial sill (p) and the sill (s). This parameter is used to describe the ratio between the variance caused by spatial variability (p) and the total variance (s):

$$R_{ps} = \frac{p}{s} \tag{4.1}$$

Where:

- R_{ps} is the ratio between the variance caused by spatial variability and the total variance of the soil parameter.
- p is the partial sill (variance caused by spatial variability).
- s is the sill (total variance).

Soil Types

To accurately compare the variance for different soil types, the variograms can be normalized by dividing the semi-variance by the squared average γ_{wet} (μ^2) of the soil type. These variograms are constructed with the same bin width (Δd) for consistency. This approach accounts for the significant variation in mean value for different soil types, ensuring that differences in variance can be compared accurately. The result can be seen in Figure 4.4.

The graphs demonstrate a higher relative variance for the clay types than the peat soils. Moreover, no

clear range is visible for the peat soils. When considering the sub-classes of the clay, the organic clay types ('kh') show a smaller variance. The p-s ratio (R_{ps}) is generally smaller for the peat soil types.

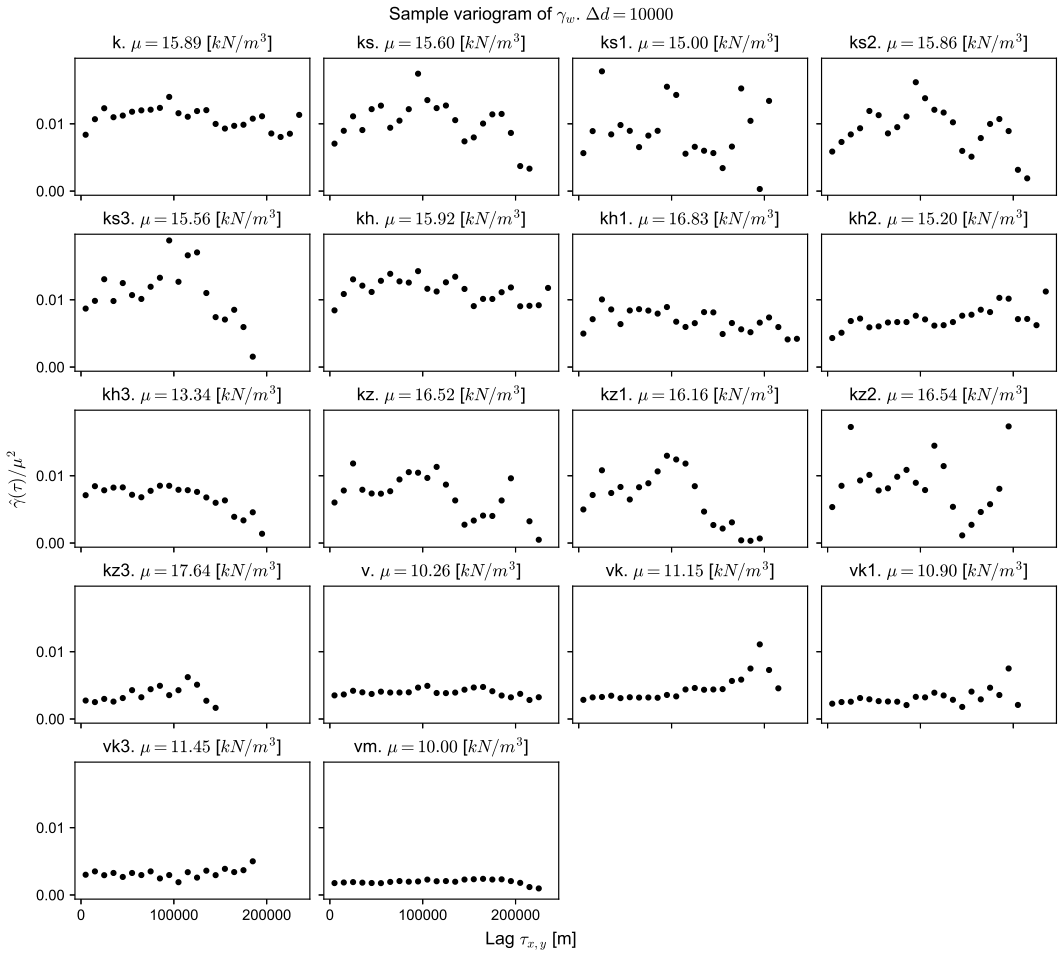


Figure 4.4: Normalized sample variograms with the same $\Delta d = 10\text{km}$ of γ_{wet} in the STOWA database.

Sub-Classes of Soil Types

Figure 4.3 illustrates the variograms for the data set of 'k', including all the clay types as outlined in 3.2. The broader clay category is expected to show less spatial correlation compared to more specific soil types. To test this, a comparison is made in Figures 4.5, 4.6 and 4.7 between the variograms of clay ('k'), silty clay('ks') and its subset, moderately silty clay ('ks').

Figure 4.5 shows a significantly larger number of data points available for clay compared to silty and moderately silty clay, which aligns with expectations since they are subsets of each other. Figure 4.6 shows that a greater level of refinement is possible in the 'k' dataset, which can be explained by the larger number of data points in that set. Note that the wave-like from, similar to the variogram of clay in Figure 4.3b, is also evident in these variograms, which will be discussed in Section 5.3.3.

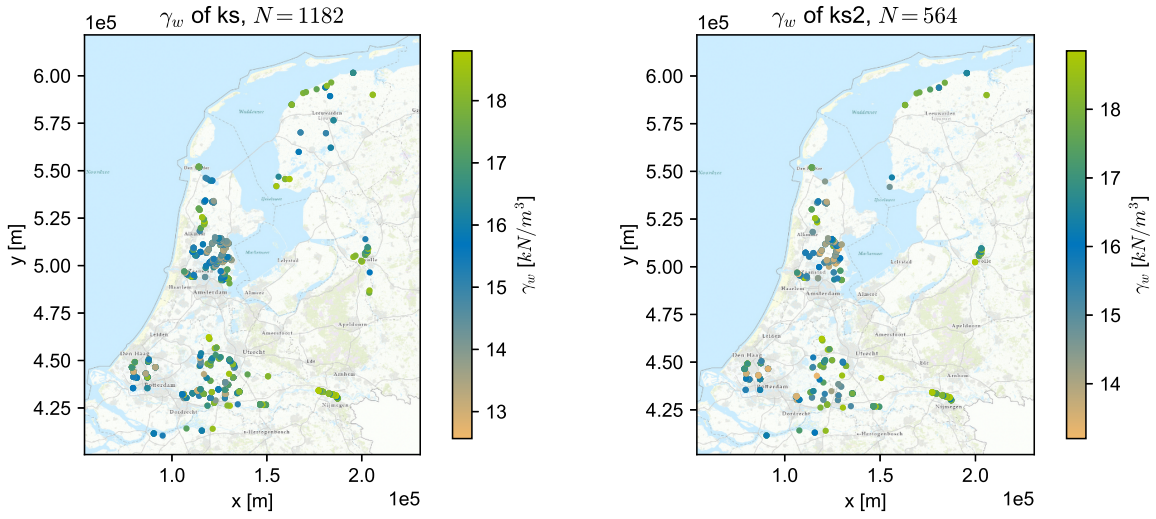
The optimal theoretical fits to the sample variogram also reveal some differences. The variogram of clay presents a more continuous line than its subsets, implying a higher noise level in the subsets. However, the nugget decreases with each sub-sampling step, and the effective range widens with increased sub-sampling. From the effective range, the ranges can be calculated (see Table 4.1). It is important to note that while the WNLS score of the silty clay variogram is smaller, it does not necessarily indicate a variogram with reduced uncertainty.

Based on the results for 'k' and 'ks', it seems that sub-sampling for clay results in wider ranges, indicating spatial correlation over longer distances. Additionally, increasing sub-sampling levels can reduce

the general variance between data points as the nugget decreases. However, the partial sill increases, indicating an increase in variance due to spatial variability. For all 'ks' sub-types, except for 'ks3', the p-s ratio (R_{ps}) increases, suggesting that a larger part of the total variance is due to spatial variability.

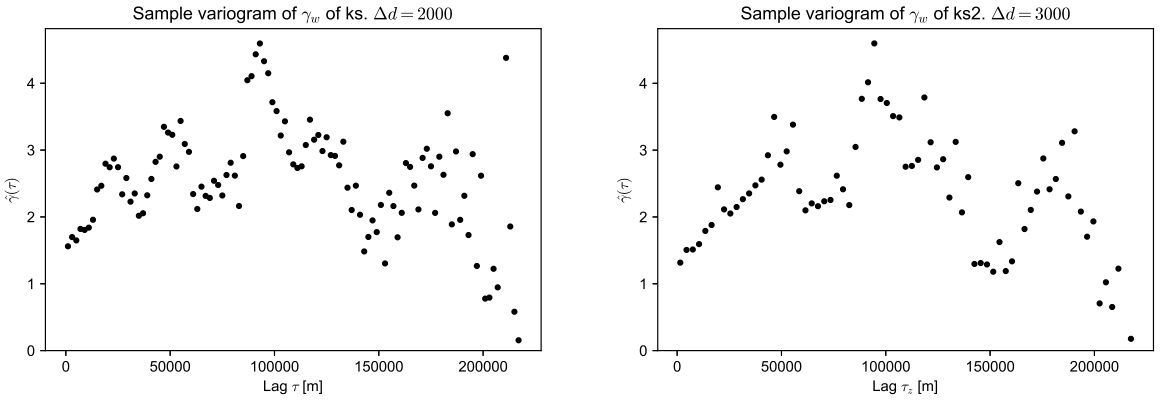
The following observations can be made for the other subsets:

- *Organic clay ('kh')*: 'kh1' has a larger range and partial sill but a smaller nugget than the other subsets. The p-s ratio is also larger. 'kh2' has a larger range but smaller partial sill and nugget and a larger p-s ratio. On the other hand, 'kh3' has a smaller range, partial sill, nugget, and p-s ratio.
- *Sandy clay ('kz')*: The range of 'kz1' and 'kz2' is more than ten times larger than their parent soil type. This indicates a strong spatial correlation over a much larger domain. However, this could also be due to the smaller bin width (Δd) for 'kz'. More detail is visible with a smaller Δd, especially in the first section of the variogram. This could indicate that the range calculated with a smaller Δd is more reliable. It should also be noted that the number of points in the sub-classes is much smaller, especially for 'kz2' and 'kz3', indicating less reliability for the results. The partial sill is larger for 'kz1' and 'kz2' but smaller for 'kz3'. Only the nugget for 'kz3' is smaller, and the p-s ratio is larger only for 'kz1'.
- *Clayey peat ('vk')*: No variogram could be created for 'vk2' since there were no points. 'vk1' has a smaller range, a slightly smaller partial sill and nugget than 'vk'. Conversely, 'vk3' has a larger range, while the difference between the partial sill and the nugget and the parent soil type 'vk' is negligible. For 'vk1', the non-spatial variability (nugget) is smaller, while the spatial variability (partial sill) is almost the same when sub-sampled. This causes the ratio between spatial and total variance (p-s ratio) to be larger for 'vk1'.



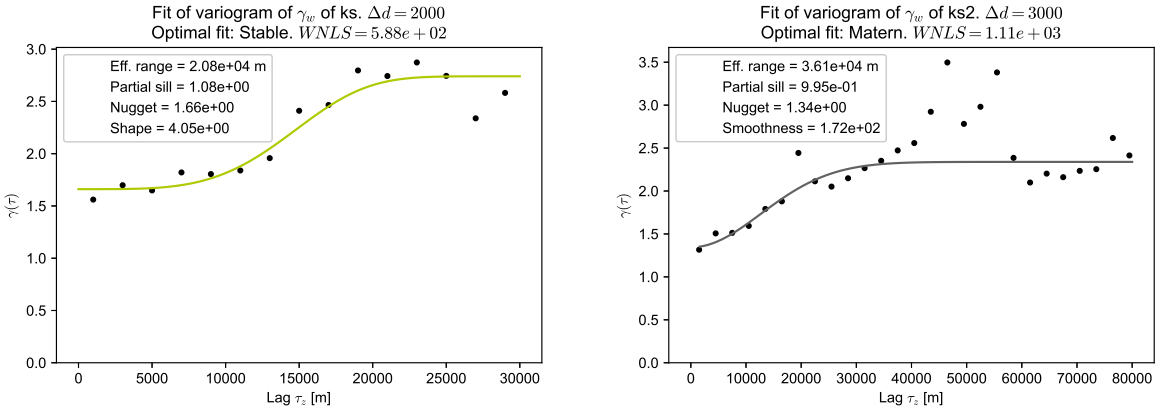
(a) 'ks' (silty clay). **(b)** 'ks2' (clay moderately silty)

Figure 4.5: Distribution of data for one subset: silty clay.



(a) 'ks' (silty clay). **(b)** 'ks2' (clay moderately silty)

Figure 4.6: Full variogram for one subset: silty clay.



(a) 'ks' (silty clay). **(b)** 'ks2' (clay moderately silty)

Figure 4.7: Full variogram for one subset: silty clay.

Depth Range

A comparison is made between various depths for selected soil types to investigate the impact of different vertical positions within soil samples. Only soil types with the largest data points were chosen, ensuring the representation of different soil sub-types. The objective is to reduce the effects caused by stress history and vertical spatial correlation.

Table C.1 in Appendix C displays the results, while Figure 4.9 illustrates the results for clay. One of the key observations is that the number of data points decreases with depth. This is due to the higher costs of performing a deeper borehole or the lesser importance of the properties of the deeper layers for the intended purpose. In the case of the STOWA database, the intended purpose is to study macro-stability and piping. The reliability of calculated variograms is affected by the low number of samples at greater depths since the semi-variance in a bin is averaged over fewer points.

Additionally, less refinement is possible, resulting in a larger value for Δd. For specific depth ranges (16 - 20m), the data points become so sparse that fitting is impossible. The sample variogram for the depths between 12 and 16m shows a large amount of noise, making it difficult to draw any clear conclusions from the theoretical functions fitted in Figure 4.9. This was also observed for other soil types.

For clay ('k') in Figure 4.9, the variograms for the first few depth ranges share similar shapes, indicating consistency in spatial correlation across different depths. The nugget and sill values remain relatively constant across different depths. The majority of partial sills (p) are larger than the partial sill for the reference clay variogram (Table 4.1). The same applies to the range values (r). However, the nuggets

(n) for the separate depths are mostly smaller. This pattern is also observed in the average values. In cases where a good fit can be made for a specific depth, the average p, r and n values are calculated, and the results are presented in Table 4.3. It can be seen that filtering for depth causes an increase in partial sill and range values while the nugget value decreases. Furthermore, the p-s ratio (R_{ps}) is higher for the averaged values.

For the remaining soil types, the following observations can be made:

- *Silty clay ('ks'), moderately silty clay ('ks2') and organic clay ('kh')*: Compared to the unfiltered dataset, the average range, average partial sill, and p-s ratio of these soil types increase. However, the average nugget decreases. The sample variograms for the first few depth ranges have a similar shape.
- *Organic clay ('kh1')*: Compared to the unfiltered dataset of 'kh1', the average range of this soil type increases. However, the average nugget, partial sill, and p-s ratio decrease. The sample variograms for the first few depth ranges have a similar shape.
- *Peat ('v'), clayey peat ('vk') and weakly clayey peat ('vk1')*: Compared to the unfiltered dataset, the average partial sill and p-s ratio of these soil types increase. However, the average range and nugget decrease. The sample variograms for the first few depth ranges have a similar shape.

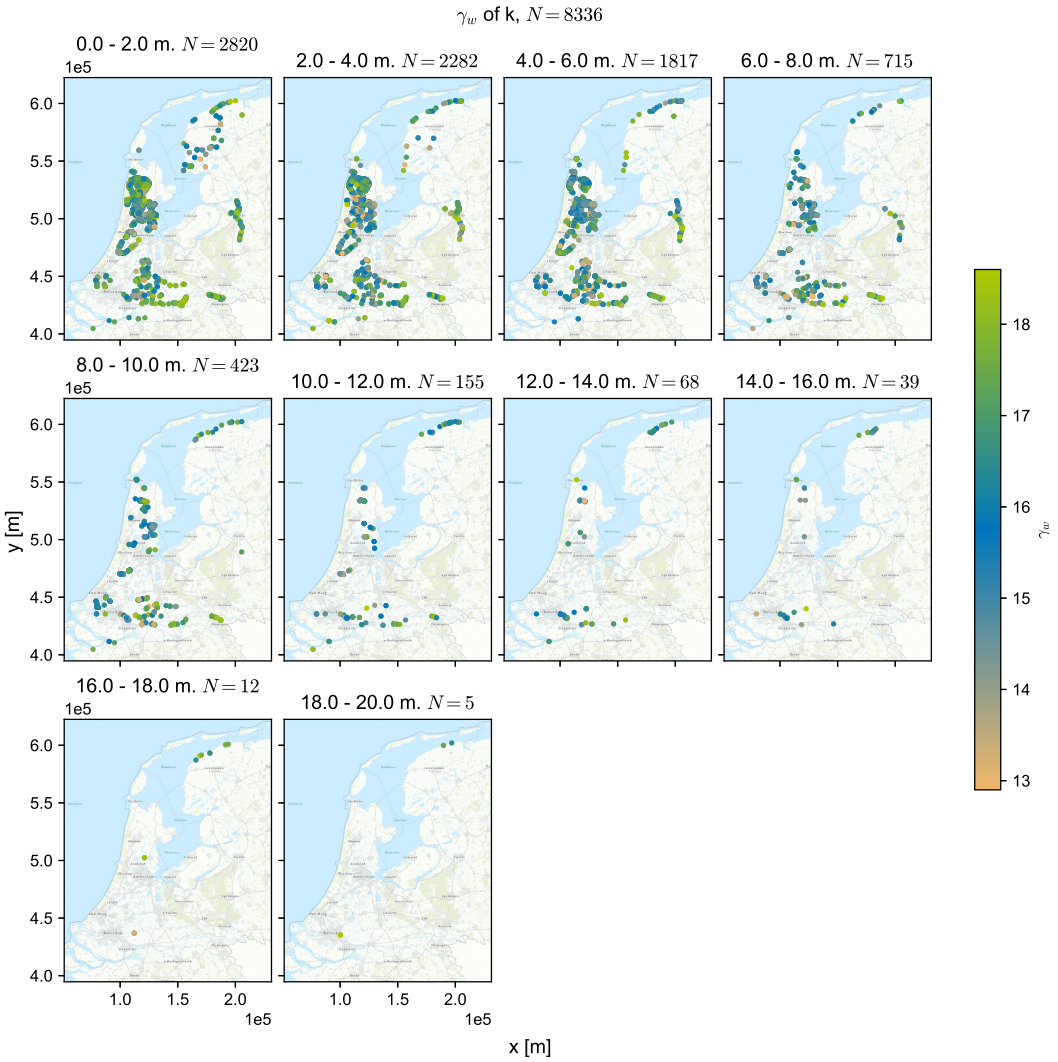


Figure 4.8: Distribution of data of γ_{wet} for different depth ranges within clay.

Table 4.3: Average variogram parameters over the different depth ranges in the STOWA database for the volumetric weight γ_{wet}. The depth ranges were selected on a 2m interval from 0 to 20m below ground level, as illustrated in Figure 4.8.

Soil	Average effective range r' [m]	Average range r [m]	Average partial sill p	Average sill s	Average nugget n	Ratio p_{avg}/s_{avg}
k	30662.68	27894.32	1.08	2.79	1.72	0.39
ks	15769.27	11730.47	1.00	2.33	1.33	0.43
ks2	45895.75	22642.07	1.41	2.18	0.77	0.65
kh	21508.59	16103.40	1.19	2.69	1.50	0.44
kh1	37844.58	35004.41	1.00	2.01	1.01	0.50
v	15045.73	14508.23	0.11	0.37	0.26	0.30
vk	9164.71	7665.89	0.16	0.41	0.24	0.40
vk1	9963.07	6466.42	0.15	0.29	0.13	0.53

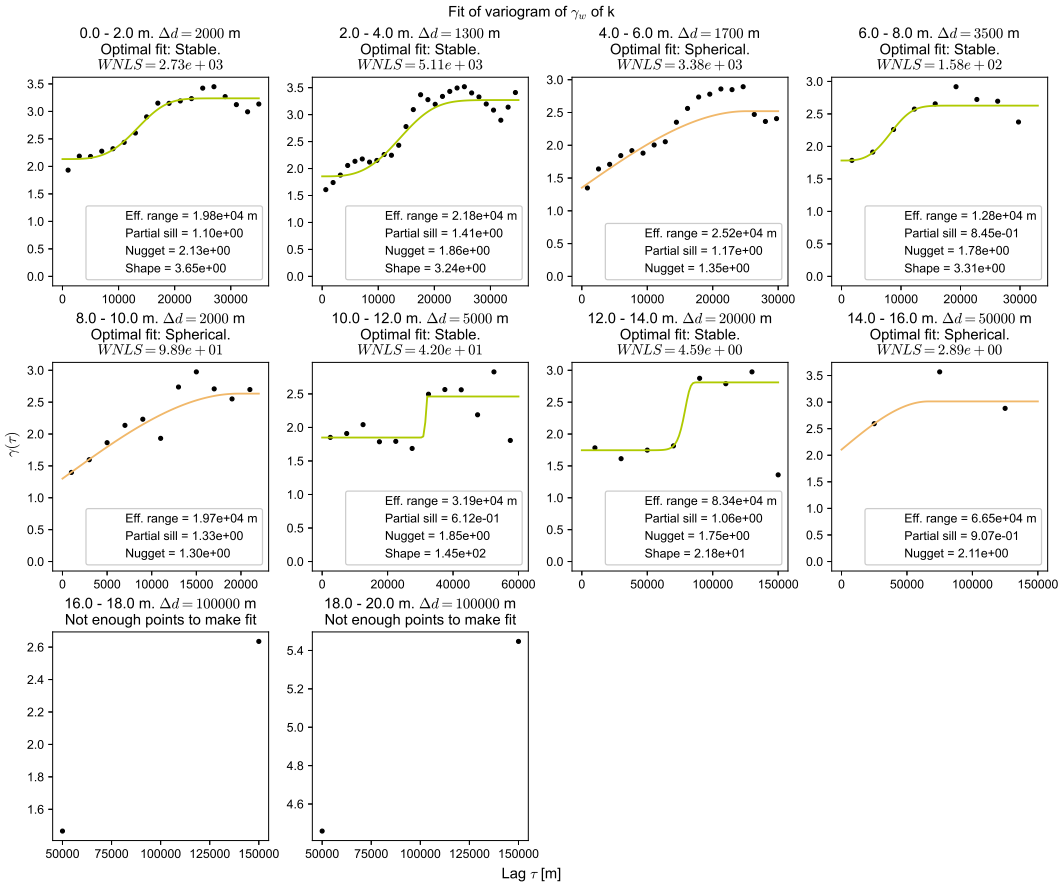


Figure 4.9: Variograms for different depth ranges for γ_{wet} of clay.

Vertical Variograms

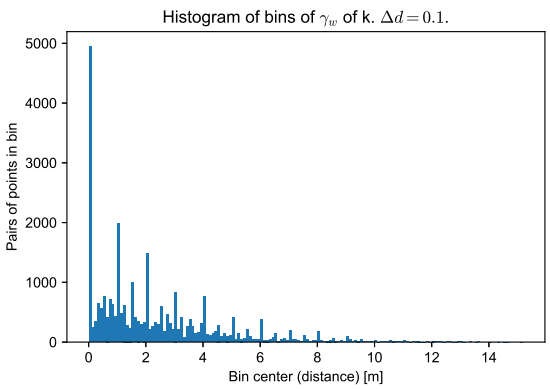
Based on literature, it can be assumed that the range of spatial variation in the vertical direction is much less than in the horizontal direction, mainly because θ_z is often greater than θ_h (see Table 2.1). This is due to the depositional environment of the soils. To investigate this, variograms have been created in the vertical direction, using only the points with the same x and y coordinates. The results of this analysis can be found in Table 4.5.

Since the maximum distance between two points in the vertical direction is considerably smaller than

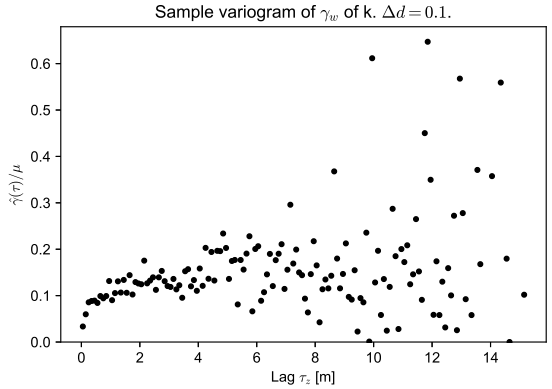
that in the horizontal direction, comparing the ranges is inappropriate due to the different scales. The ranges obtained for the clays are approximately equal to the average values presented in Table 2.1, which validates the variogram results.

When comparing Figure 4.3 in the horizontal direction with Figure 4.10 in the vertical direction, it can be observed that the partial sill is larger and the nugget is smaller in the vertical direction. However, it can be seen that the p-s ratio (R_{ps}) increases, which indicates that a larger proportion of the total variance is due to the spatial variation between the samples.

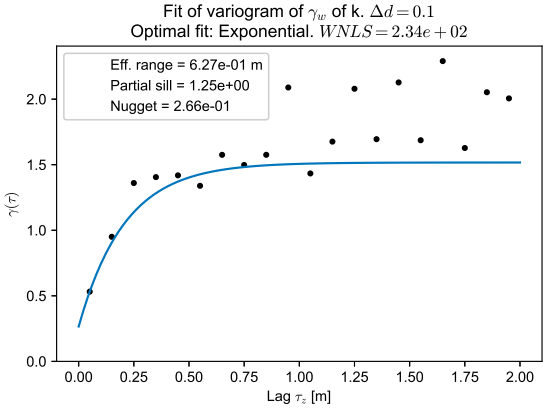
For the other soil types (refer to Table 4.1 and 4.5), a similar trend is visible. Only for 'kh1' is the partial sill smaller in the vertical direction. All p-s ratios are larger for the vertical direction.



(a) Histogram of the binned data with $\Delta d = 0.1\text{m}$



(b) Variogram with bin width $\Delta d = 0.1\text{m}$.



(c) Optimal variogram fit.

Figure 4.10: Variogram of data in the vertical direction of γ_{wet} of clay ('k') in the STOWA database.

Table 4.5: Variogram parameters for the vertical direction in the STOWA database for the volumetric weight γ_{wet}.

Soil	Number of points N	Bin width Δd	Effective range r' [m]	Range r [m]	Partial sill p	Sill s	Nugget n	Ratio p/s
k	8336	0.10	0.63	0.21	1.25	1.52	0.27	0.82
ks	3000	0.30	1.63	0.82	1.43	1.57	0.15	0.91
ks2	564	0.50	2.09	1.05	1.19	1.37	0.18	0.87
kh	5680	0.70	1.96	0.98	1.69	2.03	0.35	0.83
kh1	3349	0.60	1.18	0.70	0.81	0.94	0.13	0.86
v	3625	0.30	0.89	0.66	0.31	0.39	0.08	0.80
vk	753	0.8	No fit					
vk1	404	1	No fit					

Scale Differences

To compare the spatial variability between a national and regional scale, the variograms of two datasets are compared: the STOWA and the Rivierenland (RL) dataset. The area of the RL dataset is shown in Figure 4.11a, and it covers only about 3.3% of the total area covered by the STOWA dataset.

As shown in Figure 4.11, the range of the RL dataset is $756\mathrm{m}$, while the range of the STOWA dataset is much larger, with a range of $14248\mathrm{m}$. This means that the range of the STOWA dataset is approximately 19 times greater than that of the RL dataset. In terms of maximum distance between points, the STOWA dataset has a maximum distance of approximately $236\mathrm{km}$, while the RL dataset has a maximum distance of only $33\mathrm{km}$. By comparing the relative ranges (see Equation 4.2) of the two datasets, it can be seen that the range of the STOWA dataset is significantly larger than that of the RL dataset.

$$r_{rel} = \frac{r}{d_{max}} \tag{4.2}$$

Where:

r_{rel} is the relative range [m].
r is the range [m].
d_{max} is the maximum distance between two points in the dataset [m].

For 'k', $r_{rel,STOWA} = 14,248/236,000 = 0.06$ and $r_{rel,RL} = 756/33,000 = 0.02$. Due to the smaller relative range of the RL dataset, it suggests that the data is spatially correlated over a shorter distance, which is unexpected.

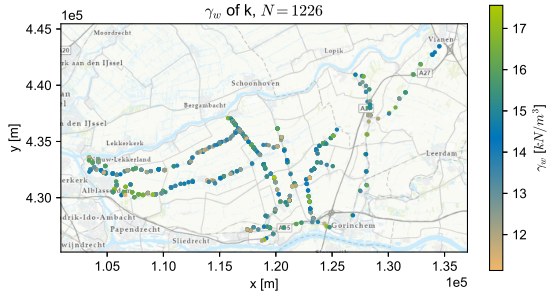
It is also unexpected that the nugget value for 'k' is found to be higher for the RL dataset ($n_{RL} = 2.29$) in comparison to the STOWA dataset ($n_{STOWA}2.07$). However, it is noteworthy that the additional variance beyond the nugget value is lower for the regional Rivierenland dataset when compared to the STOWA dataset. This can be observed by comparing the partial sill (sill-nugget) of the Rivierenland dataset ($p_{RL} = 0.38$) with the STOWA dataset ($p_{STOWA} = 0.94$), indicating that the variance due to spatial variability is more pronounced in the case of the larger region. Although the nugget value is higher for RL, the part of the total variance due to spatial variation (p-s ratio) is lower for RL.

The characteristics of the other soil types are:

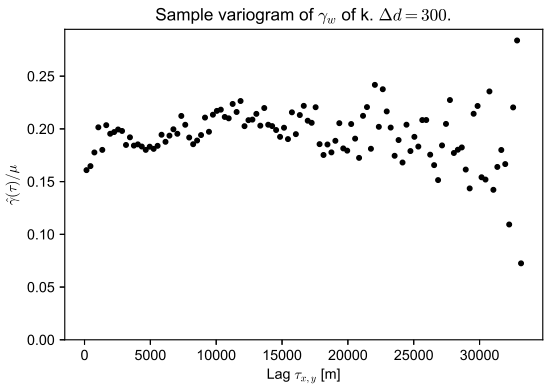
- *Silty clay ('ks') and moderately silty clay ('ks2')*: The relative range is smaller, along with a smaller nugget and partial sill.
- *Organic clay ('kh') and clayey peat ('vk')*: No appropriate fit could be made.
- *Weakly organic clay ('kh1')*: The relative range is bigger while the nugget is larger and the partial sill is smaller.

- *Peat ('v')*: The relative range is bigger while the nugget is smaller, and the partial still remains the same.
- *Weakly clayey peat ('vk1')*: The relative range is bigger, the nugget is smaller, and the partial sill is equal to zero, making it smaller than the other soil types.

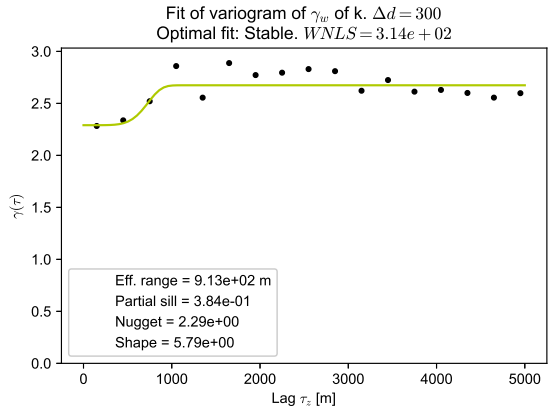
The Rivierenland dataset was also evaluated for the vertical direction. When comparing the vertical variograms of the STOWA and RL dataset, a few observations can be made for all soil types: (1) the RL dataset has consistently larger ranges and nuggets, and (2) smaller partial sills and p-s ratios for all soil types. These observations for the nugget, partial sill, and p-s ratio are consistent with the horizontal comparisons, meaning that the spatial variation is smaller for the smaller scale, even in the vertical direction. In this direction, the range is larger for Rivierenland than STOWA (see Figures 4.10c and 4.11d), which is inconsistent with the findings in the horizontal direction. It should be noted that the selected bin width (Δd) in the vertical direction of the RL dataset is almost always larger than the range in the vertical STOWA variograms. This could be the reason for the significant difference between the two datasets. Furthermore, when comparing the p-s ratio between the horizontal and vertical directions in the Rivierenland dataset, an increase can be observed across all soil types, similar to the STOWA dataset.



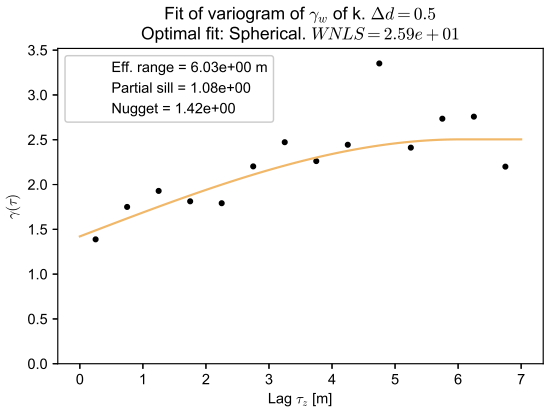
(a) Distribution of data.



(b) Horizontal variogram with bin width $\Delta d = 300\text{m}$.



(c) Optimal variogram fit in the horizontal direction.



(d) Optimal variogram fit in the horizontal direction with bin width $\Delta d = 0.5\text{m}$.

Figure 4.11: Horizontal and vertical variograms of data in the Rivierenland data set of γ_{wet} of clay ('k').

Table 4.7: Variogram parameters in the Rivierenland (RL) database for the volumetric weight γ_{wet}.

Soil	Number of points N	Bin width Δd	Effective range r' [m]	Range r [m]	Partial sill p	Sill s	Nugget n	Ratio p/s
k	1226	300	913.39	755.63	0.38	2.67	2.29	0.14
ks	584	500	1074.30	963.10	0.24	1.62	1.37	0.15
ks2	285	500	1816.47	1299.58	0.34	1.34	1.00	0.25
kh	484	600	No fit					
kh1	260	2000	13207.06	9836.34	0.54	2.24	1.70	0.24
v	989	1500	21707.97	14576.50	0.04	0.22	0.18	0.17
vk	240	400	No fit					
vk1	198	700	3648.19	3648.19	0.00	0.21	0.20	0.01

4.2. SHANSEP S

In Figure 4.13, a highly discontinuous sample variogram can be observed. Moreover, there appears to be much more scatter for the other soil types than the variograms for γ_{wet}. One possible explanation could be that less data is available for S. This lack of data makes it challenging to obtain accurate fits, especially when implementing filters to analyze soil sub-classes, depth ranges, and the vertical direction. Additionally, this specific Rivierenland dataset lacks triaxial or DSS tests, preventing comparisons at a smaller scale. However, the water board Rivierenland has incorporated these tests into the STOWA dataset. Unfortunately, these do not have the required density for constructing variograms.

When comparing the variograms, the following observations can be made:

- S *and* γ_{wet}: Generally, the range is larger for S than γ_{wet}. The partial sill and nugget are both smaller for S. However, this is mainly because the value of S is much smaller. The p-s ratio (R_{ps}) is also larger for almost all soil types.
- *Different Soil Types*: The nuggets and partial sills of S are generally smaller for peat than for clay. However, the p-s ratios are not significantly different for the different soil types.
- *Sub-Classes of Soil Types*: When comparing 'ks' and 'ks2', a larger range, partial sill, and p-s ratio can be observed. The nugget is smaller. However, 'kh' and 'kz' do not have an appropriate fit, making it impossible to compare their sub-classes.
- *Filtering for a Depth Range*: No clear patterns or trends are visible for all the variogram parameters and p-s ratios.
- *Vertical*: The theoretical functions are all horizontal lines, indicating no vertical spatial variability.

It should be noted that a proper variogram could not be constructed for many soil types. Only discontinuous and sample variograms with high scatter were found, for which no theoretical function could be fitted properly.

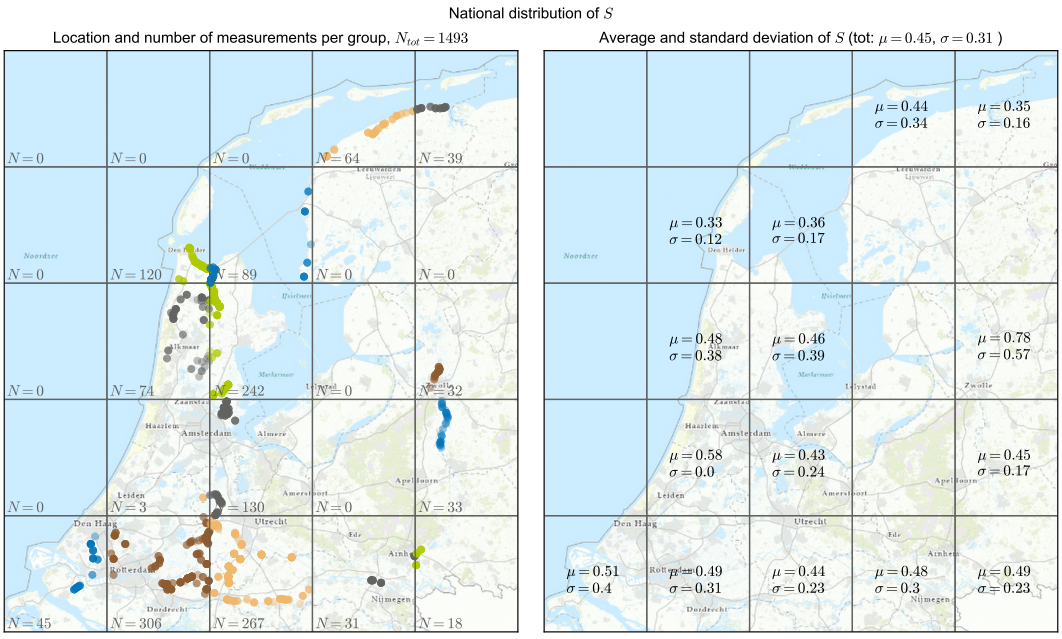
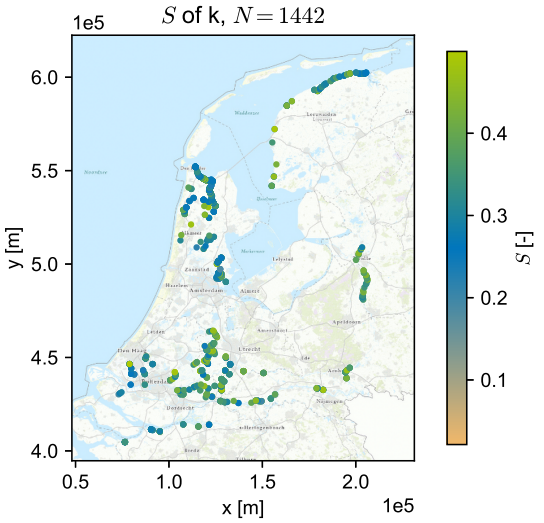
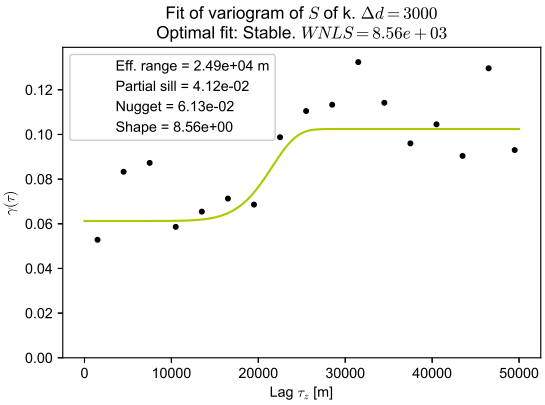


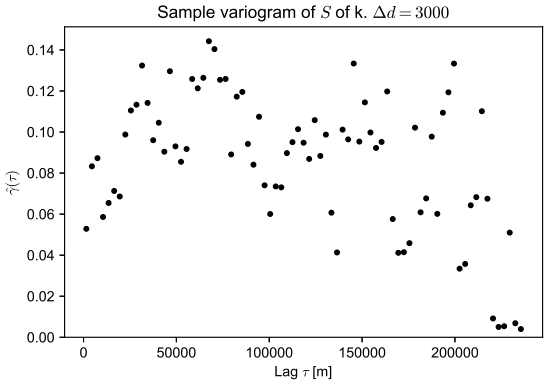
Figure 4.12: National distribution of the SHANSEP parameter S of clay. Translated and modified from Kindermann and Tigchelaar (2022).



(a) Distribution of data.



(b) Optimal variogram fit.



(c) Variogram with bin width $\Delta d = 3000\text{m}$.

Figure 4.13: Variogram of data for S in the STOWA data set for clay ('k').

Table 4.9: Variogram parameters in the Applied Water Management Research Foundation (Stichting Toegepast Onderzoek Waterbeheer) (STOWA) database for the SHANSEP parameter S.

Soil	Number of points N	Bin width Δd	Effective range r' [m]	Range r [m]	Partial sill p	Sill s	Nugget n	Ratio p/s
k	1442	3000	24893.22	21893.90	0.04	0.10	0.06	0.40
ks	377	5000	16817.43	14235.63	0.07	0.09	0.02	0.76
ks1	50	30000	No fit					
ks2	154	3000	7979.49	7714.12	0.05	0.08	0.03	0.66
ks3	155	10000	No fit					
kh	997	500	No fit					
kh1	607	4000	29054.44	29054.44	0.05	0.12	0.07	0.44
kh2	264	600	No fit					
kh3	138	2500	19245.97	6415.32	0.08	0.09	0.01	0.85
kz	17	10000	No fit					
kz1	10	10000	No fit					
v	340	20000	28646.53	14323.27	0.01	0.02	0.01	0.60
vm	314	1000	7091.25	6602.54	0.02	0.02	0.00	0.87

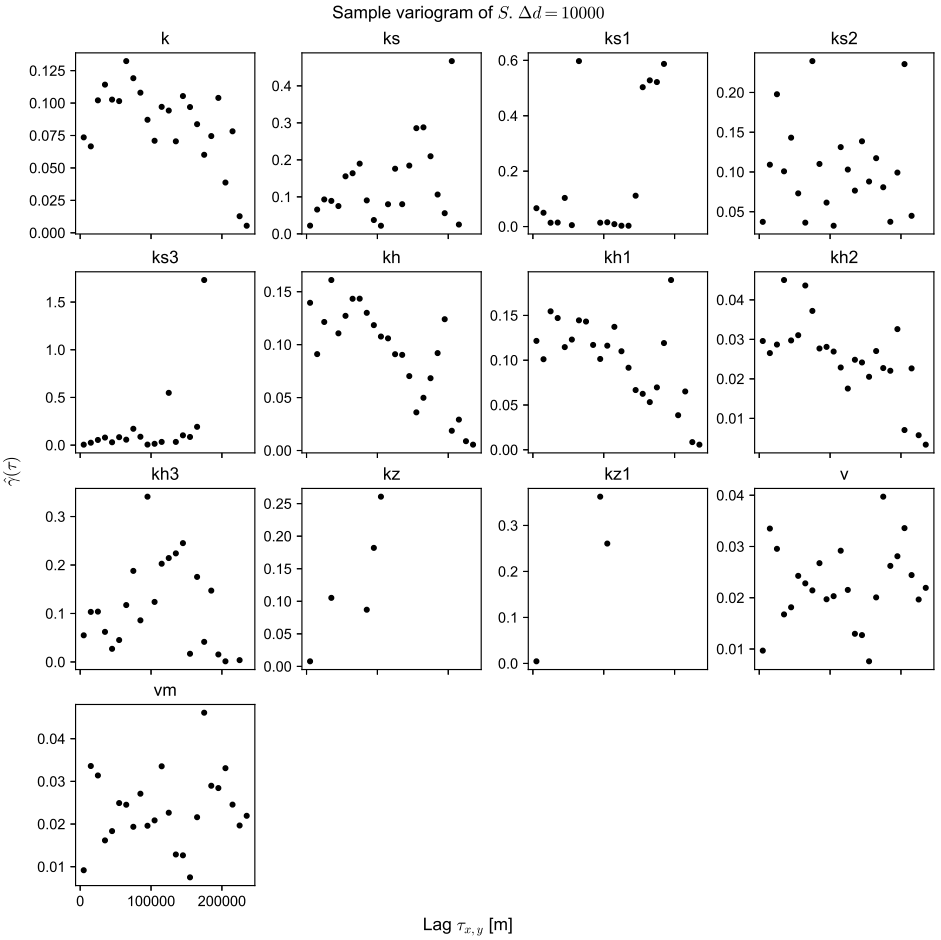


Figure 4.14: Variograms of S for different soil types in the Applied Water Management Research Foundation (Stichting Toegepast Onderzoek Waterbeheer) (STOWA) dataset. All variograms were constructed with $\Delta d = 10,000$m.

4.3. Local and Regional Variance

To further analyze the spatial variation in the soil parameters γ_{wet} and SHANSEP S, the relationship between the variogram parameters the sill (s) and nugget (n) will be investigated. In Chapter 2.2.2, the nugget is introduced as the point where the variogram intersects the y-axis. According to Clark (2010), the nugget can be associated with random noise, measurement error, and short-scale variability. The sill represents the maximum variability of the process and is an indicator of regional scale variability. Assuming that the nugget is solely caused by short-scale variability, the local divided by the regional variance can be calculated using the nugget-sill ratio. This concept is illustrated in Figure 4.15 for γ_{wet} and in Figure 4.16 for SHANSEP S.

This section refers to the WBI2017 method used for accounting for spatial variability in Dutch dike assessments, as explained in Chapter 2.1. The formulation of α is provided in Equation 2.3, which is defined in Calle et al. (2021). This formulation is used by Rijkswaterstaat (2021) to define the standard deviation that should be used in probabilistic calculations and the formulation of the characteristic value for semi-probabilistic calculations.

When assuming the nugget to be fully defined by the local variance, α can be approximated by the nugget-sill ratio:

$$\alpha = \frac{\sigma^2_{loc}}{\sigma^2_{reg}} = \frac{n}{s} \tag{4.3}$$

Where:

σ^2_{loc}	is the variance of the local dataset.
σ^2_{reg}	is the variance of the regional dataset.
n	is the nugget.
s	is the sill.

Equations 2.8 and 2.10 show that a larger value of α results in a smaller standard deviation compared to the standard value of 0.75.

4.3.1. Volumetric Weight

First, the volumetric weight of the national scale in the STOWA dataset is analyzed, followed by the regional scale analysis of the Rivierenland dataset.

National Scale

The trend of the sill and nugget follows a linear trend, with most of the data points falling within the 95% confidence interval:

$$n = \alpha * s = 0.58 * s \tag{4.4}$$

Because the STOWA database is on a national scale, α is in this case defined as:

$$\alpha = \frac{\sigma^2_{loc}}{\sigma^2_{nat}} \tag{4.5}$$

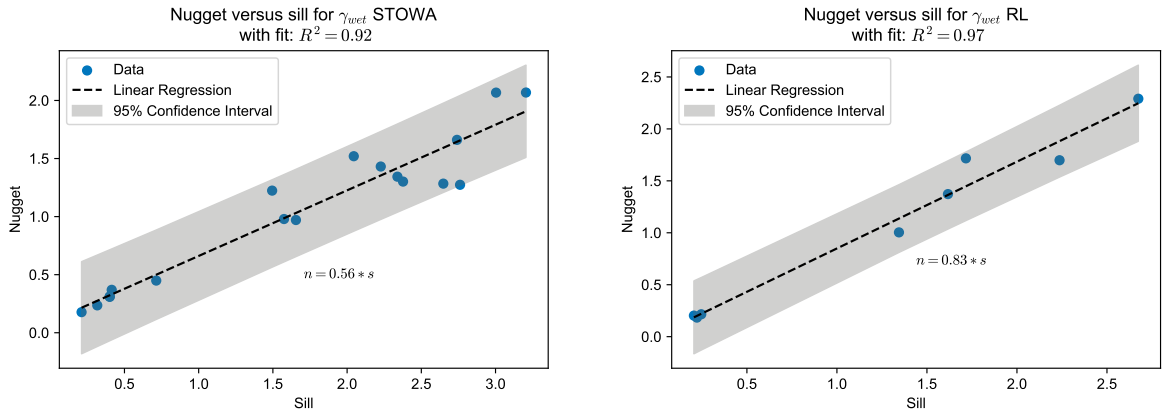
On a national scale, the variance is indicated by σ^2_{nat}. In the STOWA database, the value of α is 0.58, which is lower than the recommended value of 0.75 in Rijkswaterstaat (2021).

Regional Scale

The data sampled from Alblasserwaard, a smaller region (Figure 4.11a), can be considered at a regional scale. The nugget-sill trend is shown in Figure 4.15b.

$$n = \alpha * s = 0.83 * s \tag{4.6}$$

As expected, the ratio of the local variance to the regional variance is greater than that of the local and national variance. In this scenario, the value of α is larger than the standard value of 0.75. By using a larger value of α in Equations 2.8 and 2.10, a smaller standard deviation can be applied in the probabilistic analysis and a larger characteristic value in the semi-probabilistic analysis.



(a) STOWA dataset.

(b) Rivierenland dataset.

Figure 4.15: Sill versus nugget for γ_{wet} and a fitted trend for the different datasets.

When comparing the differences between national and regional scales, it can be observed that the total variance is larger for the national database. However, it should be noted that the regional dataset is not a part of the national dataset. Therefore, part of the variance on the national scale is not caused by the variance on the regional scale. In the industry, data for stability assessments of a cross-section is often collected on a waterboard level, which is approximately the same scale as the RL dataset. Therefore, the analysis with the RL dataset better represents use cases in the industry.

4.3.2. SHANSEP S

Figure 4.16 shows the relationship between sill and nugget for the national scale using the STOWA database. It can be observed that the values for different soil types are much more varied for S. This could be due to various factors that influence the variogram's quality, as discussed in Chapter 5.2.

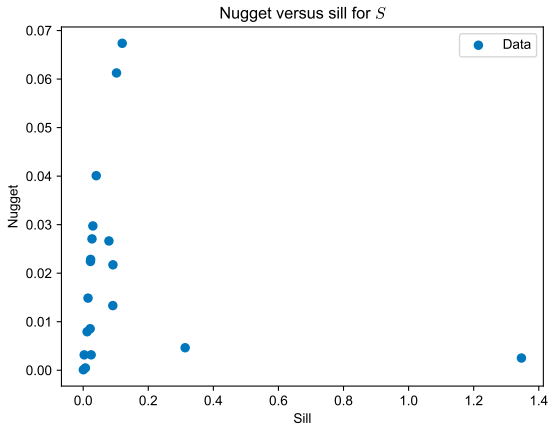


Figure 4.16: Sill versus nugget for S.

5

Discussion and Conclusion Data Analysis

A geostatistical analysis was implemented to explore the spatial variability in two parameters: γ_{wet} and the SHANSEP parameter S. Variograms were constructed for both parameters, and a series of filters were applied to further investigate the spatial variability. Additionally, vertical variograms were constructed. For the γ_{wet} parameter, variograms were created for the national and regional scale. The results of these analyses will be discussed in this chapter, and the conclusions will be presented. Finally, the applicability of using variograms for determining local spatial variability will be discussed and concluded.

5.1. Volumetric weight

Based on the constructed variograms, it can be seen that the spatial variability varies considerably with soil type, scale and direction. This paragraph will discuss the main observations, followed by a comment on the vertical variability. The analysis will conclude with an overview of the key findings.

5.1.1. Discussion Results

After comparing the variograms of γ_{wet} for different soil types, it could be seen that the variance due to spatial variability and total variance is more significant for clay compared to peat. However, when more soil sub-types are defined, the total variability increases only for silty clay and clayey peat soil types. This highlights the importance of differentiating between subtypes when incorporating spatial variability in assessments for these soil types.

Filtering clay for specific depth ranges leads to a more significant variance because of spatial variation. As a result, the uncertainty of measurements decreases, and the p-s ratio for the averaged values is also larger. For other soil types, filtering by depth ranges usually increases the variance due to spatial variation while the general uncertainty decreases. However, the reliability of these parameters decreases with increasing depth when filtering because of the decreasing number of points.

A comparison was made between the variograms of two datasets, one at a national scale (from the STOWA database) and the other at a regional scale (Rivierenland (RL) database). Inconsistent results for the range can be observed across different soil types, making it difficult to draw any conclusions about the area over which the measurements are correlated. However, the partial sill, nugget, and p-s ratio are smaller for the RL dataset, indicating a decrease in uncertainty due to spatial variability and general uncertainty when looking at a smaller area, which is expected.

Most of the time, the ratio of the variance due to spatial variability (p-s ratio) is less than 0.5. This indicates that other sources of uncertainties contributing to the variance are more significant than the spatial variability. However, spatial variability still plays an important role in most cases and accounts for a part of the total variance.

5.1.2. Vertical Variability in the Horizontal Variograms

It is important to note that the variograms calculated in the horizontal direction ignore the depth location (z). As a result, samples from the same borehole are assigned a distance of $0\mathrm{m}$ to each other and fall into the first bin. Even when the horizontal direction is considered, the distances in the z-direction are so small compared to the x, y-direction that all the distances in the vertical direction are binned together in the first bin. Therefore, the first bin in the horizontal direction includes the spatial variance present in the vertical direction to some extent. Due to the large scale of the horizontal variograms, the small-scale variance could not be adequately captured, and the estimated nugget may be larger than the actual value present in the data.

Tables 4.1 and 4.5 show the comparison of γ_{wet} results in the horizontal and vertical directions in the STOWA database. An interesting observation from these tables is that the sill value in the vertical variogram is smaller for nearly all soil types than the nugget value in the horizontal variogram. However, this could not be observed for the variograms of the Rivierenland database. Therefore, it is unclear how much vertical variance is incorporated into the horizontal variogram.

5.1.3. Conclusions

From the results and discussion of the variograms for γ_{wet}, the following conclusions can be made:

- There is a certain degree of spatial variability in the value of γ_{wet}, particularly for clay samples and when considering only their vertical locations.
- Spatial variability decreases at a regional scale compared to a national scale as the ratio between spatial variability and total variance decreases.
- It is unclear how much vertical variance is incorporated into the horizontal variogram.

5.2. SHANSEP S

Most variograms for S could not correctly be constructed because they did not show a clear spatial correlation pattern. The possible reasons behind the low quality are discussed in this paragraph. Nevertheless, the variograms that could be constructed show larger p-s ratios for S, indicating that S exhibits higher spatial variability than γ_{wet}. In general, the variance for peat is smaller than for clay. However, filtering for sub-soil types or depth ranges does not provide a clearer view of spatial correlation, and no spatial variability is visible in the vertical direction.

5.2.1. Quality of the Variograms

Most of the constructed variograms for S were either highly discontinuous or could not be computed. Attempts to filter the data based on depth or subsoil types did not cause significant improvements. This could be attributed to three possible reasons: (1) there may be no spatial correlation for S; (2) the quantity of data available for S may not be sufficient; or (3) the quality of the data for S may not be good enough.

The SHANSEP parameters are related to the soil's mineralogy (Ajmera et al., 2016). It can be hypothesized that the spatial correlation will not impact the variance. However, the sample variograms constructed for different soil types ('k', 'ks', 'kh', 'v' and 'vm') in Figure 4.14 show an increasing trend in the first part, which suggests the presence of spatial variability. Even though it is not always a clear continuous line like the variogram in Figure 4.3b, it can be concluded that in some cases, spatial correlation may play a role in the total variance.

Around 150 data points and 100 point pairs in each bin are needed to construct a variogram (Webster & Oliver, 1992). The results presented in Table 4.9 show that most soil types have more than 150 data points available. Furthermore, variograms with more than 100 data point pairs per bin could be constructed for most soil types. However, the variogram is still scattered and discontinuous even for soil type 'k', which has 1442 data points. This would suggest that the amount of data available is not the primary issue.

The quality of the variogram can be affected by several components, including spatial density and patterns, outliers, and epistemic uncertainties. These components and some methods to eliminate their influence will be discussed shortly. It is recommended to investigate the cause of poor variogram

quality for S using these methods in further research.

Spatial Density and Patterns

The variogram's accuracy is influenced by data availability, and regions with sparse data are less reliable than those with high data density. This results in irregularities in the variogram. The use of the appropriate sampling scheme can help resolve this issue. Geotechnical designs often use regular spaced sampling schemes, as recommended by, for example, Eurocode 7 (CEN, 2004). However, the distances recommended by EC7 (25 - $75\mathrm{m}$ for dams) are not sufficient for studying spatial variability in soil parameters, as at least 5 points within the range are needed to construct an appropriate variogram (Oliver & Webster, 2015).

Recent research has been focusing on Bayesian approaches (Wang et al., 2019) and declustering techniques such as centroidal Voronoi tessellation (Shi et al., 2023) to address the ongoing challenge of obtaining an effective sampling scheme in three dimensions.

An appropriate sampling scheme can be tested by implementing, for example, a dummy dataset.

Outliers

Filtering outliers is important for stabilizing variograms (Oliver & Webster, 2015). Even though outliers were removed in the data preparation stage, there were still some outliers visible in the variogram plot for S compared to γ_{wet} (as shown in Figure 5.1). This could be due to possible errors in soil type classification or in the DSS or triaxial tests used to calculate the values. These outliers can significantly affect the calculated sample variogram and cause scatter.

It should be noted that since this investigation aims to assess dike stability, the low and high values are of interest, and removing outliers could be counter-intuitive. However, it is important to remember that this analysis focuses on investigating spatial variation rather than identifying an appropriate distribution for soil parameters.

A stricter filtering scheme could eliminate outliers. Cressie and Hawkins (1980), Dowd (1984) and Genton (1998) also proposed robust variograms against outliers.

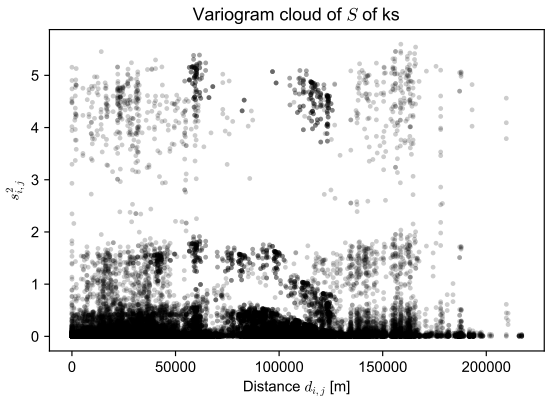


Figure 5.1: Variogram cloud of silty clay ('ks') of S. It can be seen that there are some outliers above a squared difference of 3.

Epistemic Uncertainties

Schweiger and Peschl (2005) define epistemic uncertainties as: "the type of uncertainty which results from a lack of knowledge about a parameter (state of knowledge, subjective uncertainty or ignorance)". This differs from aleatoric uncertainty, which originates from inherent variability caused by heterogeneity or the random nature of the natural processes (Rohmer & Verdel, 2014). While aleatoric uncertainty can be described using traditional probabilistic theory like a Probability Density Function, epistemic uncertainty often relies on expert opinions and cannot be properly described using these methods. The uncertainty related to the definition of S and the inability to identify and quantify a representative soil strength parameter can cause epistemic uncertainties. The discontinuity of the sample variograms of S might be explained by the variogram's inability to model these uncertainties.

Using Bayesian approaches can help investigate epistemic uncertainties (Loquin & Dubois, 2010). Handcock and Stein (2022) presents a method that uses this approach, allowing for expert opinions or previous studies to be incorporated into the a priori distribution. This can help in the evaluation of the fitted theoretical variogram.

5.2.2. Conclusions

The following conclusions can be drawn from the results of the variograms for S:

- Most variograms for S could not be properly constructed because of their poor quality.
- The variograms show that the variance for clay is higher than for peat.
- No spatial variability is visible in the vertical direction.
- The exact cause of the poor quality of the variograms is unclear, but it is likely due to the poor quality of the data. The spatial density and patterns found in the data, outliers, and incorrect inclusion of the epistemic uncertainties can affect the quality of the variogram. Additional research on this matter is recommended.
- From these results, the hypothesis that S is not spatially correlated cannot be ruled out.

5.3. Applicability of a Variogram

As part of the process of constructing variograms, it became clear that there are certain constraints to using variograms to determine spatial relationships for soil parameters. This paragraph will discuss these limitations and give a conclusion on the applicability of using national variograms to determine local spatial variance.

5.3.1. The Number of Data Points

The construction of the variogram was not possible for some datasets due to a small number of points, which limited the comparison material, especially for the S analysis.

The accuracy of the variogram is highly dependent on the number of data points in a dataset (Oliver & Webster, 2015). Here are some guidelines related to the number of points:

- Around 150 points are required to construct a variogram, but more points are preferable when accurately estimating the sill (Webster & Oliver, 1992) (Webster & Lark, 2013) (Oliver & Webster, 2014).
- Bins in the variogram must have at least 100 pair points. If the histogram shows a decrease below 100 points per bin, the variogram should be cut off at that point.

More data points result in a smaller bin width (Δd), which captures more details and leads to better estimation of variogram parameters, especially the nugget and range. A smaller bin width will also lead to a better approximation of the spatial variability on a smaller scale.

5.3.2. Analysis of the Range

The ranges for S and γ_{wet} were found to be inconclusive. The values fluctuated, indicating that a variogram on a national scale is unreliable for determining the horizontal scale of fluctuation θ_h for a stability calculation. Although the vertical variograms of clay have similar ranges to those found in literature (see Table 2.1), the ranges in the horizontal variogram did not match these values.

5.3.3. Wave Shape variograms

In some variograms for γ_{wet}, a wave-like pattern can be observed, as shown in Figures 4.3b, 4.5a, and 4.5b. These figures show that the variance initially increases, then decreases after reaching a certain point, and sometimes, this pattern is repeated several times. However, this pattern is unusual as it is unexpected that the difference between points decreases when the distance between them increases.

It is worth noting that the decreasing number of points in a bin is not the reason for the decrease in variance, as most variograms still have a large number of points in the bin at the lag where the variance decreases again. This is particularly evident when comparing the sample variogram in Figure 4.3b with the histogram in Figure 4.3a.

The likely cause of the issue seems to be the presence of clusters. As can be seen from Figure 4.1, the measurements are distributed across the country in clusters. When considering the distance between these clusters, it becomes clear that the points within a cluster are compared to one another within a limited radius. The variance between these points increases with distance. However, beyond a certain maximum distance, the points between two different clusters are compared, resulting in a decrease in variance. This suggests that the dataset may not be intrinsically stationary, which is necessary for creating the variogram.

Intrinsic stationarity refers to the requirement that the increments of a parameter, denoted by $X(u)$, are second-order stationary. This means that the increments should have the same statistical characteristics such as mean, variance, autocorrelation, etc. For a more detailed description, please refer to Chapter 2.2.2 and, for example, in Marchant and Lark (2004).

Figure 3.2 shows the mean and standard deviation for different parts of the Netherlands. Based on their analysis, Kindermann and Tigchelaar (2022) concluded that the dispersion between the parts of the Netherlands was not too large, and therefore, it was assumed that the dataset is intrinsically stationary.

The statement that the dispersion is not large is subjective and can be questioned based on the wave shape found in the variograms. If the data is not intrinsically stationary, a variogram may not be the appropriate way to display spatial variation. The wave-like form makes it difficult to fit the theoretical variogram and read the variogram parameters, which can also affect the accuracy of the calculated variogram parameters of γ_{wet}.

Based on Figure 4.12, it can be observed that the mean and standard deviation of S differ for different parts of the Netherlands, indicating non-intrinsic stationarity. However, no clear waveform can be seen for S, likely due to the poor quality of the variograms.

5.3.4. Local and Regional Variance Analysis

When analyzing the value of α in relation to the nugget and sill, it is assumed that the nugget is entirely caused by small-scale variations that cannot be observed. In this case, the nugget can be defined as the local variance, denoted as σ^2_{loc}. However, the nugget can be caused by various factors, such as random noise and measurement errors (Clark, 2010). Assuming that small-scale variations cause the nugget entirely implies that the variogram created for the smaller, local scale would show zero variance at zero lag. However, due to sample disturbance, mineralogy, and errors during the testing process, it is highly unlikely that samples taken at the exact location would have precisely the same value for S or γ_{wet}.

When considering this in the nugget-sill analysis of γ_{wet}, the portion of the nugget caused by local variance decreases, resulting in a decrease in α. In the case of the national dataset, the difference between the standard value would increase, leading to an increase in the probabilistic standard deviation and a decrease in the semi-probabilistic characteristic value, thus creating more unsafe designs when incorporating $\alpha = 0.75$ compared to this decreased value of α. In the regional dataset, the value of α would be more similar to the standard value of 0.75, increasing the probabilistic standard deviation and decreasing the semi-probabilistic characteristic value.

It is worth noting that the current Dutch dike assessments often include the stochastic nature of S but not for γ_{wet}. Instead, the mean of local samples is used in both probabilistic and semi-probabilistic analyses (Rijkswaterstaat, 2021). When modeling the SHANSEP parameter S, it is important to consider the spatial variability in either the characteristic value or the standard deviation. Unfortunately, no trend was identified in the nugget-sill analysis for S. Therefore, the results of γ_{wet} provide only a rough indication of the ratio between local and regional variance. The observed values for α cannot be immediately applied in the assessments.

5.3.5. Scale Differences

In Table 4.7, it is indicated that the smallest bin width (Δd) for the RL database is 300m and 450m. Therefore, any point pairs that are closer than 300m apart are considered part of the local scale in this analysis. For these points, the variance is averaged and represented by the nugget. It is important to note that the definition of the ”local” scale used in this variogram analysis is much larger than that of

the ”local” scale used in WBI2017, which is on the scale of a failure surface. According to Calle et al. (2021), the width of a failure surface is 30 - $70\mathrm{m}$ parallel to the dike and 10 - $20\mathrm{m}$ perpendicular to the dike. The same applies for the STOWA database, with the minimum Δd being $450\mathrm{m}$. This highlights the difference in scale between the variograms and the definition of local in the dike assessments.

5.3.6. Conclusions

Based on the discussion of the limitations, the following conclusions can be made:

- Some variograms could not properly be constructed due to the limited number of data points.
- The horizontal scale of fluctuation could not be calculated from the range over which parameters are correlated in the variograms.
- The calculated values for α in the local and regional variance analysis cannot be used in the stability assessment of dikes.
- Due to some limitations of the use of national variograms, like the requirement of intrinsic stationarity of the data and the large difference in scale, it cannot be used for determining the local-scale spatial variability.

5.4. Use of Data in Model

In the following chapter, a model will be set up using some data from the STOWA database. Because the case study is located in the maintenance area of waterboard Brabanetse Delta, the STOWA data is filtered and used accordingly. The locations of the data used for this purpose are shown in Figure 5.2. Most data points are located along the dikes on the north side of the maintenance area. In total, 135 data points are used, some of which are from the same borehole, so they have the same x and y coordinates. Tables D.1 and D.2 in Appendix D show the data that is used as input for clay and peat respectively.

From the data analysis, it is concluded that the variograms on a national scale are not suitable for determining the range and scale of fluctuation of the soil. Therefore, these parameters cannot be used as input for the model.

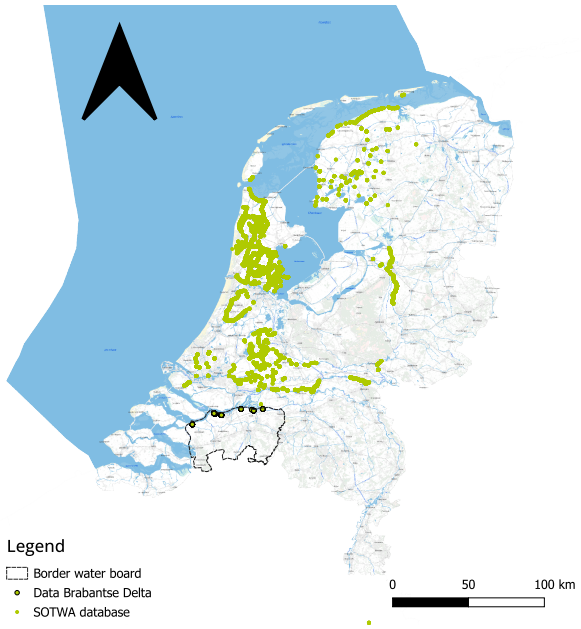


Figure 5.2: Map with the STOWA database where the data from the maintenance area of water board Brabantse Delta is highlighted.

Part II

Modeling

6

Model Setup

This study aims to create a realistic model of the spatial variability in Dutch dikes to investigate its impact on macro-stability. To achieve this, a Random Finite Element Model (RFEM) is created for a case study dike. To compare the RFEM model with the WBI2017 method of assessing dike stability in the Netherlands, the RFEM model will be adapted to a reference D-Stability model. This reference model was developed for this specific location by Tigchelaar and van Haaren (2022) and is based on the current requirements of the Dutch guidelines of Rijkswaterstaat (2021). Figure 6.1 provides an overview of the results of the D-Stability model.

This chapter presents the model's setup. First, the case study will be introduced. Then, the model's inputs will be described in detail. Random Finite Element Method (RFEM) requires three main inputs: soil parameters, hydraulic conditions, and a mesh. A detailed description of each of these parts will be given in this chapter.

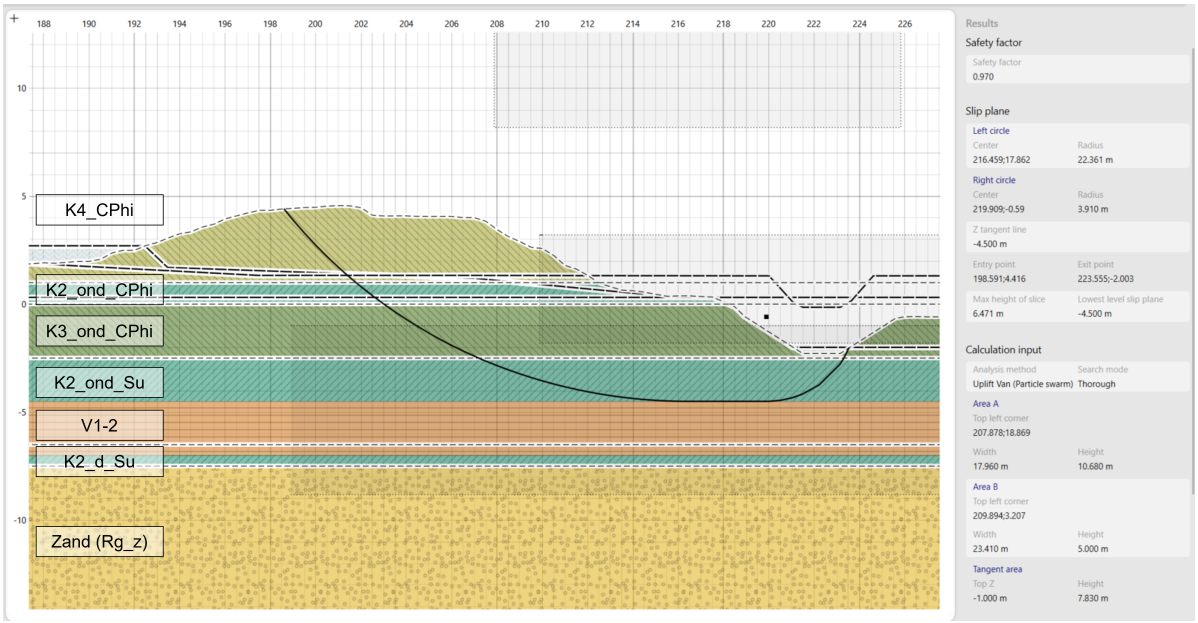


Figure 6.1: Results of the D-Stability model of the case study dike. Adapted from Tigchelaar and van Haaren (2022).

6.1. Case Study Location

The cross-section to be investigated is situated in the maintenance area of the water board Brabantse Delta. Specifically, this case study will be focused on dike section 20, which is part of trajectory 34-2 (see Figure 6.2).

This chapter will describe the starting points of the dike model, including the general geological and geotechnical layers and the hydraulic conditions.

The Waterlaw has established a probability of failure (signal) norm of $1/1,000$ per year for this dike trajectory. When considering the length effect, this can be scaled down to a norm of $1/404,880$ per year for the dike cross-section (Tigchelaar, 2022). When performing a semi-probabilistic calculation, the norm can be translated into a normative Factor of Safety (FoS), taking into account the model factor of 1.06 for Uplift-Van, which is the standard model used for Dutch dike assessments (Rijkswaterstaat, 2021). This results in a minimum FoS_{min} of 1.16 (Tigchelaar, 2022). The probabilistic calculation will be assessed for the cross-section norm.

Trajectory 34-2 experienced several breaches during the watersnoodramp in 1953, which have since been repaired. Moreover, the trajectory has been reinforced multiple times, including in the 70s and 2001 (Tigchelaar, 2022).

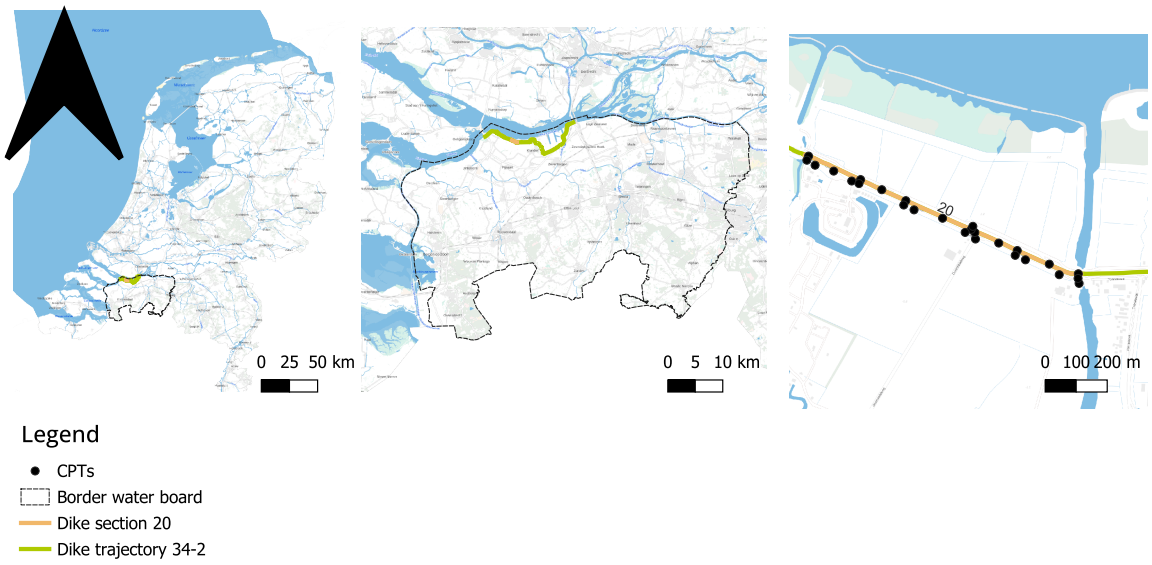


Figure 6.2: Map of the water board maintenance area with the location of the dike section. The illustrated CPTs are the ones that are used to determine the scale of fluctuation.

6.1.1. Geotechnical Conditions

The geological and geotechnical conditions of the dike section were studied by Tigchelaar (2018). Tigchelaar used various resources, including the VNK2 (Veiligheid in Nederland in Kaart) project (VNK2, n.d.-a) (VNK2, n.d.-b), the stochastic soil model (WTI-SOS) (Hijma & Sun Lam, 2015), and maps from TNO-NITG (GeoTop) (TNO-NITG, n.d.), and found the following structures:

- The Pleistocene sand layer is located at approximately -8m NAP.
- The Holocene layer consists of formations of Naaldwijk and Nieuwkoop.
- The layers between the Pleistocene and Holocene include formations of Boxtel and Kreftenheye and the formation of Peize-Waalre.

The geotechnical soil layers for the dike reinforcement project have been established by Tigchelaar

(2018) and Tigchelaar and van Haaren (2022) through soil investigations and data from previous dike reinforcement projects. The dike trajectory consists of both clay and sand dikes, and a peat layer can be found throughout the trajectory at around $-4\mathrm{m}$ NAP, with varying thickness from 1 to $2.5\mathrm{m}$. In section 20, the dike is made of clay, and the peat layer is located between -4.5 and $-7\mathrm{m}$ NAP. According to the Dutch dike assessment regulations (Rijkswaterstaat, 2021), the impermeable layers beneath the phreatic surface will be modeled undrained using SHANSEP. Table 6.1 and Figure 6.1 provide an overview of the layers. The volumetric weight of soil is determined by its relation to a WTI-SOS soil class.

Table 6.1: Soil layers found in the case study dike section. Based on Tigchelaar (2018) and (Tigchelaar & van Haaren, 2022). Note that soil type P_Ova_sd will not be modeled due to its significant depth. The volumetric weight of soil is determined by its relation to a WTI-SOS soil class (Hijma & Sun Lam, 2015).

Nr.	Name	Soil type	Top [m NAP]	Bottom [m NAP]	Model	γ_{wet} [$\mathrm{kN/m^3}$]
1	K4_Cphi	Clay sandy	Ground level	1	Mohr-Coulomb	18.8
2	K2_ond_Cphi	Clay weakly silty	1	0	Mohr-Coulomb	15.9
3	K3_ond_CPhi	Clay moderately to strongly silty	0	-2.5	Mohr-Coulomb	17.3
4	K2_ond_Su	Clay weakly silty	-2.5	-4.5	SHANSEP	15.9
5	V1-2	Peat low mineral content	-4.5	-7	SHANSEP	10.3
6	K2_d_Su	Clay weakly silty	-7	-7.5	SHANSEP	15.9
7	Zand (Rg_z)	Sand	-7.5	-15	Mohr-Coulomb	17
8	P_Ova_sd	Sand	-15		Mohr-Coulomb	17

Hydraulic Conditions

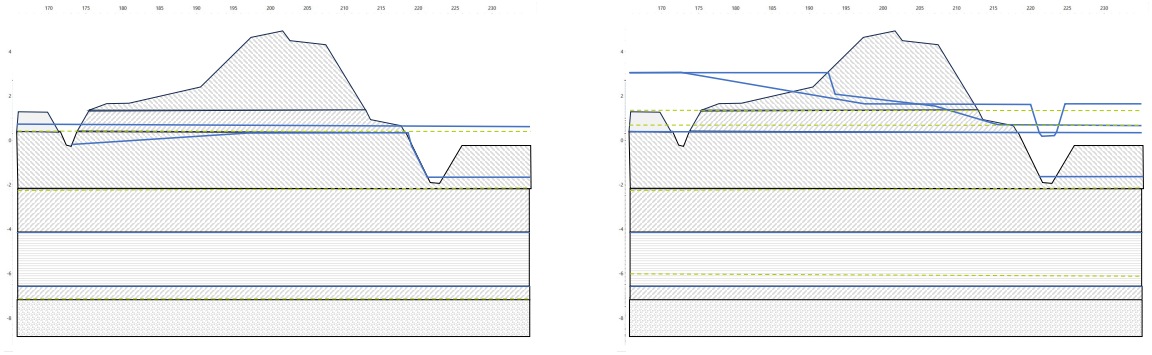
The hydraulic conditions were also analyzed by Tigchelaar and van Haaren (2022) and Tigchelaar (2022). They have been based on the guidelines from Rijkswaterstaat (2021) and data from piezometers (TAUW, 2021).

Under normal conditions, the phreatic level is located at $-0.68\mathrm{m}$ NAP in front of the dike, while behind it, in the polder, the phreatic level is regulated at $-2\mathrm{m}$ NAP. The phreatic level in the area between these points bulges to $0\mathrm{m}$ NAP in the dike. The aquifer's hydraulic head is $0.3\mathrm{m}$ above NAP. Figure 6.3a illustrates the water levels during normal conditions.

During extreme weather conditions, the water level outside the dike can rise to $2.7\mathrm{m}$ above NAP (Tigchelaar, 2022). This level is called the Waterstand Bij Norm (WBN) in Dutch, and this is the water level that needs to be evaluated in a semi-probabilistic analysis according to Rijkswaterstaat (2021). A jump is modeled in the phreatic line through the revetment layer to simulate the transition between the water level outside and inside the dike. The hydraulic head progresses through the dike and reaches the water level inside the polder ditch at a depth of $-2\mathrm{m}$ NAP. The hydraulic head in the aquifer also reacts to the increase in water level outside the dike. The hydraulic head in front of the dike is at the same level as the phreatic surface and decreases to $1.3\mathrm{m}$ in the polder.

This increased water pressure in the aquifer causes water to seep through the clay and peat layer on top of the sand. The intrusion length, also known as "indringingslengte" in Dutch, is the vertical distance at the bottom of the poorly permeable aquitard over which the pore pressure changes with hydraulic head variations in the aquifer layer, as defined by Rijkswaterstaat (2021). At the top of this

intrusion zone, the hydraulic head is $0.3\mathrm{m}$ NAP, equal to the hydraulic head in the aquifer during normal conditions. Figure 6.3b illustrates the water levels in the extreme conditions.



(a) Normal conditions.

(b) Extreme conditions.

Figure 6.3: Normal and extreme hydraulic conditions modeled in D-Stability. The blue lines are the water lines and the reference lines are depicted in green.

6.2. Soil Parameters

The soil parameters are an important input for the model. These parameters are based on a regional dataset of the maintenance area of the water board Brabantse Delta and can be found in Appendix D. The dataset was created by Tigchelaar (2018). Please refer to his research for a detailed description of the field and laboratory tests that were conducted. This dataset was also included in the STOWA dataset that was investigated in the data analysis in part 1 of this research. For this study, the data was further analyzed to determine the soil parameters.

In this chapter, the process of determining these parameters will be presented. Four types of analyses will be conducted in this study: deterministic, semi-probabilistic, probabilistic, and Random Finite Element Method (RFEM) analyses. The point statistics of the dataset will be presented first, followed by an explanation of how the stochastic soil parameters will be determined in the semi-probabilistic and probabilistic analyses. After that, the method for the RFEM analysis will be introduced. Finally, the other parameters will be discussed, where no spatial variability should be considered according to Rijkswaterstaat (2021). To meet the requirements of Rijkswaterstaat (2021), all parameters will be modeled using a log-normal distribution.

According to Rijkswaterstaat (2021), it is necessary to consider spatial variability when determining the SHANSEP parameter S and m and the friction angle ϕ. However, as it is assumed that the FoS is not significantly impacted by m, no spatial variability is considered when modeling m. After performing the data analysis, it is concluded that it is challenging to determine spatial correlation for S. Further research on this assumption is therefore recommended. However, to compare the model results with the current method, spatial variability is accounted for in S.

These values for S and ϕ are calculated using Equations 2.8, 2.9, 2.10, and 2.4. First, the dataset in Appendix D was analyzed to determine the mean (μ_X) and standard deviation (σ_X). These values were then transformed to fit a log-normal distribution using Equations 2.5 and 2.6. Table 6.3 presents the resulting values for different soil types.

It is assumed that the sand layer does not affect the dike's stability, so it is not modeled stochastically.

Table 6.3: Summary data from dataset Tigchelaar (2018).

Parameter	Soil dataset	Soil model	n	μ_X	σ_X	$\mu_{\ln X}$	$\sigma_{\ln X}$
S [-]	K2	K2_onder_Su and K2_d_Su	42	0.323	0.045	-1.14	0.14
	V2	V1-2	22	0.458	0.053	-0.787	0.115
ϕ [°]	K2	K2_ond_CPhi	42	36.67	2.05	3.6	0.06
	K3	K3_ond_CPhi	13	35.28	3.58	3.56	0.1
	K4	K4_CPhi	34	31.94	4.24	3.46	0.13
	Zand	Zand	5	34.57	-	-	-

For the semi-probabilistic and probabilistic analyses, the soil will be modeled using different values of α: $\alpha = 1$, $\alpha = 0.8$, the default $\alpha = 0.75$, and $\alpha = 0.5$. The student-t factor is the same for the different values of α and is calculated based on the number of data points as presented in Table 6.3. For each α value, the mean, standard deviation, and characteristic value for S and ϕ are calculated. The results are presented in Table 6.5.

When $\alpha = 0.5$, only the characteristic value is calculated as it is exclusively analyzed in a semi-probabilistic analysis. It can be observed from Table 6.5 that $\sigma_{X,prob}$ decreases for all soil types as α increases, whereas X_{char} increases and becomes more similar to $\mu_{X,prob}$ as α increases. According to Equation 2.9, it can be seen that the value of $\mu_{X,prob}$ is also affected by the value of α through the value of $\sigma_{\ln X,prob}$. However, the values in the Table show that this influence is minimal and can be considered negligible in most cases.

Table 6.5: Calculated soil parameters for the probabilistic and semi-probabilistic analyses for different values of α based on Equations 2.8, 2.9, 2.10, and 2.4. In the table, p and c in the subscript equals $prob$ $char$ in the equations respectively.

		$\alpha = 1$				$\alpha = 0.8$			
Param	**Soil**	$\sigma_{\ln X,p}$	$\mu_{X,p}$	$\sigma_{X,p}$	X_c	$\sigma_{\ln X,p}$	$\mu_{X,p}$	$\sigma_{X,p}$	X_c
S	K2	0.022	0.32	0.007	0.308	0.067	0.321	0.022	0.286
	V1-2	0.026	0.455	0.012	0.436	0.060	0.456	0.027	0.413
ϕ	K2	0.009	36.62	0.321	36.09	0.027	36.63	0.986	35.03
	K3	0.03	35.12	1.066	33.39	0.058	35.16	2.025	31.92
	K4	0.023	31.67	0.737	30.47	0.065	31.73	2.063	28.44
		$\alpha = 0.75$				$\alpha = 0.5$			
S	K2	0.075	0.32	0.024	0.283				0.27
	V1-2	0.065	0.456	0.03	0.409				0.393
ϕ	K2	0.03	36.63	1.091	34.86				33.3
	K3	0.063	35.17	2.202	31.66				29.11
	K4	0.072	31.74	2.279	28.13				25.23

One of the purposes of the RFEM analyses is to study the effect of spatial averaging, as defined in Chapter 2.3.2 and Figure 2.7, on the failure probability of a dike. As a result, the soil parameter distributions should not account for spatial averaging and variance reduction at this stage. Unlike the (semi-)probabilistic calculations, spatial variability will not be considered when determining the input mean and standard deviation. However, to account for the limited number of data points, Equations 2.8, 2.9, and 2.10 will be used with $\alpha = 0$.

When $\alpha = 0$ in Equation 2.3, the local variance is zero since the regional variance cannot be. In the WBI2017 method, only the local variance can be averaged out. Hence, no averaging or variance reduction should be considered in this case. This implies that only the RFEM model itself can account for any variance reduction based on the configuration of stronger and weaker zones in the model and the propagation of the failure surface through them.

The input data required for the RFEM model is presented in Table 6.7.

Table 6.7: Calculated soil parameters for the RFEM analyses based on Equations 2.8, 2.9, and 2.10 with $\alpha = 0$

		$\alpha = 0$			
Param	**Soil**	$\sigma_{\ln X,RFEM}$	$\mu_{X,RFEM}$	$\sigma_{X,RFEM}$	ω
S [-]	K2	0.144	0.323	0.047	0.023
	V1-2	0.109	0.518	0.057	0.043
ϕ [°]	K2	0.058	36.68	2.111	0.023
	K3	0.114	35.33	4.023	0.071
	K4	0.138	31.96	4.419	0.029

Some uncertainty should not be averaged out, which is the uncertainty due to the limited number of samples in a dataset and the regional variance. Therefore, a residual correlation factor (ω) is introduced in the Auto Correlation Function (ACF) in Equation 6.1. The purpose of this factor is to prevent the covariance in the random field from going to zero, as seen in Figure 6.4, thereby maintaining a degree of variance in the random field. For more information about random fields, please refer to Chapter 2.3.

$$\rho(\tau) = \omega + (1 - \omega) \exp\left(-\frac{2|\tau|}{\theta}\right) \tag{6.1}$$

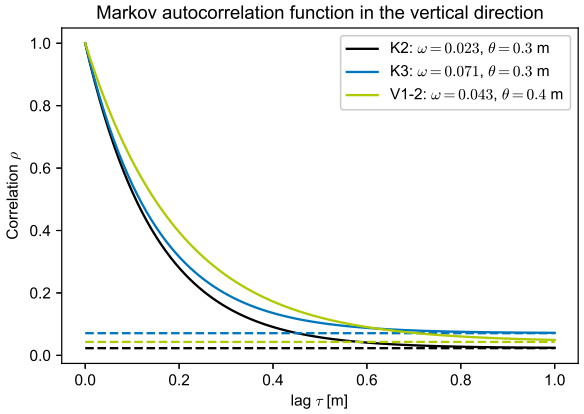


Figure 6.4: The used Markov ACF in the vertical direction for the different soil types in the model. The dashed line is the value for ω for that soil type, which is the asymptote for the ACF.

To determine the value of ω, the ratio between the variance that cannot be averaged out ($\sigma^2_{no\ ave}$) and the total variance (σ^2_{tot}) is calculated. This can be seen in Equation 6.2.

The WBI2017 method assumes that the local variance can be completely averaged out, while the regional variance cannot. Therefore, there is no regional variance when using a local dataset, and $\alpha = 1$. In this case, only the statistical variance remains, which cannot be averaged out. Conversely, when α is set to 0, the local variance is zero, and only the regional variance remains in combination with the statistical variance, which cannot be averaged out in accordance with WBI2017.

The data analysis in part I of this study has shown low partial sill-sill ratios. This indicates that only a small portion of the total variance is defined by regional variance. As a result, in this research, ω is defined as the ratio of the measured sample variance when all variance is averaged ($\sigma_{prob,\alpha=1}$) to the measured sample variance when no variance is averaged ($\sigma_{prob,\alpha=0}$). Both include the statistical variance, so this is the only variance that cannot be averaged in RFEM. This definition allows the possibility of also averaging the regional variance in the RFEM analysis. The calculated values of ω are presented in Table 6.7.

$$\omega = \frac{\sigma^2_{no\ ave}}{\sigma^2_{tot}} \tag{6.2}$$

$$\omega = \frac{\sigma^2_{prob,\alpha=1}}{\sigma^2_{prob,\alpha=0}} \tag{6.3}$$

The values for the SHANSEP parameter m are based on recommendations from Rijkswaterstaat (2021). Also, according to their guidelines, the Critical State Soil Mechanics (CSSM) framework is used to determine the values for the dilation angle and cohesion. For a more detailed description of CSSM, please refer to Chapter A.3.3. The drained CSSM analysis uses the non-associative Mohr-Coulomb model, where the dilation angle (ψ) and cohesion (c) are utilized. No plastic volume changes occur at the critical state. Therefore $\psi = 0°$. The CSSM defines cohesion as a form of over-consolidation, which does not play a role in the critical state. Therefore, the cohesion can be defined as $c = 0\text{kPa}$. Young's modulus (E) and Poisson's ratio (ν) have been determined based on the values used in the research of Varkey (2020). Note that the absolute value of E does not affect the ultimate strength, but it is nonetheless a required input for the model to calculate the strains. Table 6.9 summarizes these parameters.

It is important to note that only the friction angle (ϕ) and the SHANSEP parameters S and m have been modeled stochastically, in line with the recommendations of Rijkswaterstaat (2021).

Table 6.9: Additional geotechnical parameters, independent of the soil type, used as input for the model.

Parameter	μ_m	σ_m	m_{char}	ψ [°]	E [MPa]	ν [-]	c [kPa]
Value	0.9	0.03	0.86	0	0.1	0.3	0.01

6.2.1. Scales of Fluctuation

Other important model inputs are the scales of fluctuation (θ) for each soil type in horizontal and vertical directions. However, due to the lack of laboratory data on a small scale (see the data analysis in part 1), Cone Penetration Tests (CPTs) are used to determine the scale of fluctuation in the vertical direction (θ_v). In the horizontal direction, the average distance between the CPTs is 45m, which is larger than the average θ_h found in literature (refer to Table 2.1). Therefore, it is impossible to determine θ_h from the available data. To investigate its impact, a sensitivity analysis will be conducted for different ratios between θ_h and θ_v, including 10, 15, 20, and ∞.

De Gast et al. (2019) outlined three steps to determine the value of θ from CPTs

1. Eliminate the trend from the CPTs.
2. Using the CPT data, compute the experimental Auto Correlation Function (ACF). The experimental ACF is the mean of the individual experimental ACFs in case of multiple CPTs.
3. Determine θ by methodically evaluating a theoretical ACF that has been produced using various θ's. The θ that results in a theoretical function with the smallest error compared to the experimental ACF should be selected.

The research follows these steps for the different soil layers defined in Table 6.1. The data is analyzed individually by subtracting the parts of the CPTs that fall within the depth ranges of the layers defined in the table. The CPTs used for this analysis are shown on the map of the dike in Figure 6.2. Figure 6.5 displays the cone resistance of the CPTs plotted over the depth. The x-axis shows the different x-coordinates of the CPTs. The different soil layers are labeled as defined in Table 6.1.

Removing the Trend

One of the requirements for using an ACF is second-order stationarity. The depth trend should be removed from the CPT data to achieve this. The trend is defined by:

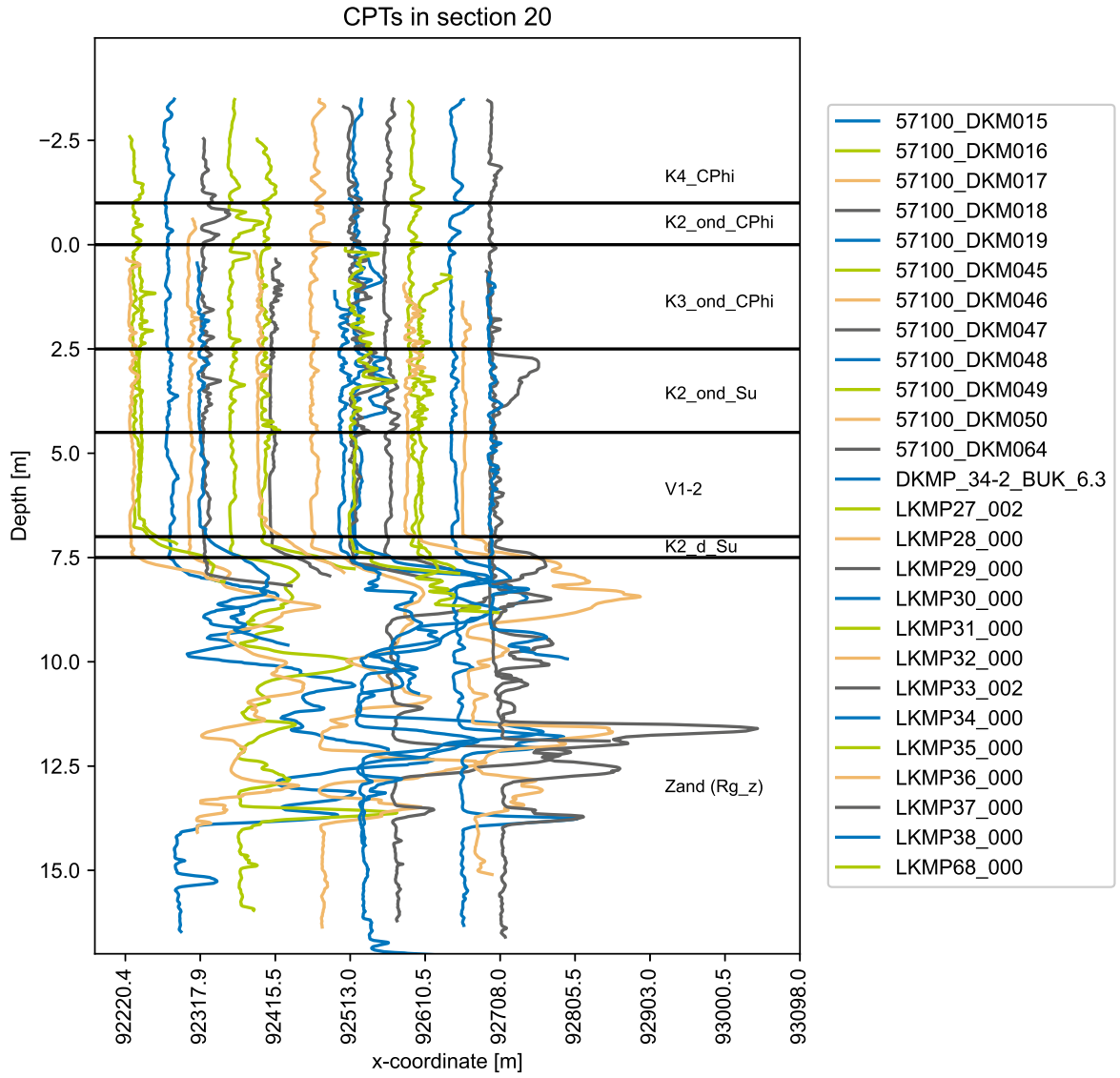


Figure 6.5: The cone resistance and x-coordinates of the CPTs that were used to determine the scale of fluctuation (θ)

$$q_c = (b + az) + \varepsilon(z) \tag{6.4}$$

Where:

q_c	is the cone resistance of a CPT [MPa].
b	is the intersect of the depth trend [MPa].
a	is the slope of the depth trend [MPa/m].
z	is the depth [$mathrmm$].
$\varepsilon(z)$	the spatial variability around the trend as a function of the depth z [MPa].

The spatial variability, denoted by $\varepsilon(z)$, should be normalized before use:

$$U(z) = \frac{\varepsilon - \mu_\varepsilon}{\sigma_\varepsilon} \tag{6.5}$$

Where:

$U(z)$ is the detrended and normalized spatial variability in q_c as a function of the depth z [-].
μ_ε is the mean of ε [MPa].
σ_ε. is the standard deviation of ε [MPa].

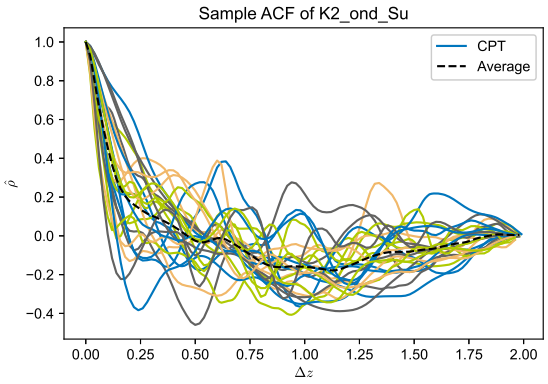
Autocorrelation Functions

In the second step of the process, an experimental autocorrelation function, also called the sample ACF as defined in Equation 2.17, is created. Then, multiple autocorrelation functions are averaged to form one sample ACF. Next, the average ACF is fitted with a Markovian ACF, as suggested by De Gast et al. (2019) and defined in Table 2.3. The theoretical ACF, a function of θ, is then fitted by minimizing the error between the theoretical and the average experimental functions. Figure 6.6 shows an example of this process for soil type K2_ond_Su.

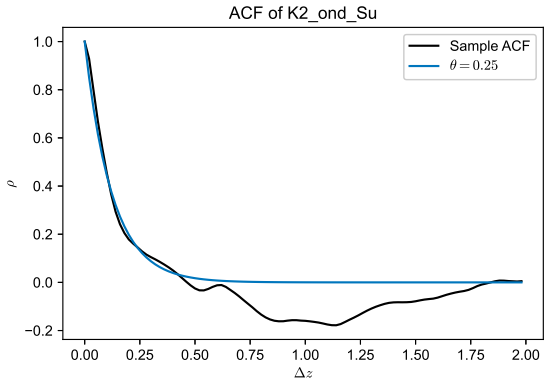
In Figure 6.6, it can be seen that the experimental ACF can have negative values. In contrast, the theoretical Markov ACF, defined in Table 2.3, cannot have negative values. The θ value plays a role in determining how fast the ACF approaches zero (see Figure 2.8). Therefore, to determine θ, it is more important that the theoretical function matches the first part of the experimental ACF, which in this case is up to a lag of approximately 0.65m.

The fitted value of θ determines the scale of fluctuation for a particular type of soil. The calculated values of θ_v for the clay layers are small compared to the values that could be found in literature. For the peat layers, θ_v corresponds to the results of de Gast (2020). A possible reason for the low clay θ_v values could be the thickness of the layers being too small to capture the entire θ_v.

Table 6.11 shows the results for different soil types. The value of θ_v is adjusted to match the mesh size of Finite Element Method. A mesh size of 0.1m would be needed to model the scales of fluctuation accurately, but this would significantly increase computational time. Therefore, a larger mesh size is selected. Choosing a larger value of θ_v is considered conservative, as it would result in larger weaker zones that would have a larger impact on the FoS than smaller stronger zones. In other words, a larger θ_v would lead to less averaging and a large mean reduction of the soil strength parameters over the failure surface, resulting in a more conservative approach. Lastly, it is worth noting that sand (zand) is not modeled stochastically and thus does not have a θ implemented in the model.



(a) Experimental autocorrelation functions for the different CPTs and the average of these experimental autocorrelation functions.



(b) Theoretical autocorrelation function, fitted to the average experimental autocorrelation function.

Figure 6.6: Experimental and theoretical Auto Correlation Function (ACF) for soil type K2_ond_Su.

Table 6.11: The calculated horizontal and vertical scales of fluctuation (θ) in meters. The value of θ_v has been determined by fitting ACFs to CPT data, which have then been adjusted to fit the mesh size of the model. However, θ_h could not be determined based on CPTs. Therefore, a sensitivity analysis will be conducted with different ratios between θ_h and θ_v. It is worth noting that sand (Zand) will not be modeled stochastically.

Soil	θ_v	θ_v model	$\theta_h = 10 * \theta_v$	$\theta_h = 15 * \theta_v$	$\theta_h = 20 * \theta_v$
K2_ond_Su	0.25	0.3	3	4.5	6
K2_d_Su	0.11	0.3	1	1.5	2
V1-2	0.4	0.4	4	6	8
Zand	0.17	-	-	-	-
K2_ond_CPhi	0.12	0.3	1	1.5	2
K3_ond_CPhi	0.17	0.3	2	3	4
K4_CPhi	0.21	0.3	2	3	4

6.3. Hydraulic Conditions

The RFEM software calculates the pore pressures in the dike by performing a flow calculation. To carry out this calculation, the permeability is assumed to be equal in each soil type. The input for this flow calculation is water lines, which are assigned to the bottom and top of each layer to determine the water pressure. If the water line at the top of the layer differs from the bottom, a hydraulic head difference occurs, and a flow calculation is performed. When there is no difference in hydraulic head over the layer, the water pressures are assumed to be hydrostatic. The boundary conditions of the flow calculations are the hydrostatic layers.

Two water lines are modeled: one for the phreatic surface and one for the hydraulic head in the aquifer. The phreatic line can be copied from the D-Stability model, but the hydraulic head cannot be immediately duplicated.

The RFEM code calculates the pore pressures differently from D-Stability. RFEM performs a water flow calculation, while D-Stability assumes stationary flow. In D-Stability, one can also assume a reference line, the location on which the assigned hydraulic head is applied, at any location within a soil layer. However, in RFEM, it is only possible to assign a water line at the border of a layer. This is a problem for implementing the intrusion length in the RFEM model. An additional layer can be added with the border on the same location of the intrusion length in D-Stability. However, this causes problems when generating the random fields in RFEM, which are defined in a specific layer. Therefore, a simplification has been added to the hydraulic conditions of the RFEM model. The intrusion zone is assumed to equal the clay layer (K2_d_Su) on top of the aquifer.

To model the intrusion, two flow calculations are needed to model a hydraulic head difference: (1) the hydraulic head difference between the phreatic line and the hydraulic head in the layers K2_ond_Su and V1-2 and (2) the hydraulic head difference between the intrusion zone and aquifer. However, there is no hydrostatic layer in between these flow calculations. And because the flow calculations need hydrostatic layers as boundary conditions, these two flow calculations cannot be performed. Therefore, a second simplification is added. The hydraulic head at the top of the deepest clay layer (K2_d_Su) will serve as the second water line. This line is simplified and represents the hydraulic head at the top of the intrusion zone instead of the hydraulic head in the aquifer. The flow boundary condition was dealt with by adopting a hydrostatic gradient in the clay and sand layers.

The difference in pore pressure is assumed to be minimal because the difference in intrusion length is small, causing only a small decrease in pore water pressure in the undrained layers. Furthermore, it is assumed that the deepest clay and sand layers do not contribute to the stability of the dike. Therefore, it does not matter if the pore pressures are changed to hydrostatic conditions. These assumptions were verified by altering the D-Stability model and checking the difference in FoS.

6.4. Mesh

The mesh in the RFEM model is based on the D-Stability cross-section. The used elements are 0.3m wide, except for the layers that will not be modeled stochastically, with an element width of 0.9m to limit computational time. However, due to the algorithm used to create the mesh, some elements may be smaller than 0.3m. This can cause issues when small elements are located along the failure surface, as they can experience much larger strain rates than larger elements. As a result, the FEM program may converge incorrectly to failure. The distance between the bottom of the ditch and the layer boundary is 0.2m, which causes the algorithm to create very small elements in this specific location. To avoid this issue, the bottom of the ditch has been lowered so that it is directly located on the layer boundary between K3_ond_CPhi and K2_ond_Su.

When comparing the cross-section and mesh used in the RFEM model to the D-Stability model, it is clear that extra soil layers have been added on the polder-side slope. This is because this research aims to investigate the effect on the macro-stability of a dike, and only macro-stability should be modeled. Micro-instability can cause a breach only as part of a progressive failure mechanism, so it will not be considered. The failure surface that will cause a breach of the dike has been identified by Tigchelaar and van Haaren (2022) in D-Stability and can be seen in Figure 6.1. To simulate this failure surface in RFEM, stronger zones were added to the slope on the dike's polder side. This slope is prone to micro-stability failures due to its steepness. By adding stronger zones, the possibility of micro-instability failures is eliminated. These stronger zones were not modeled stochastically and had a consistent value of $\phi = 35°$ and $c = 30\text{kPa}$, but with the same volumetric weight as the original soil types.

These modifications made to the model, along with the changes in hydraulic conditions, were tested using D-Stability. The adapted model showed a FoS of 0.98, while the original model had a FoS of 0.97. This indicates a minimal difference of 0.01 in the FoS, which is considered insignificant.

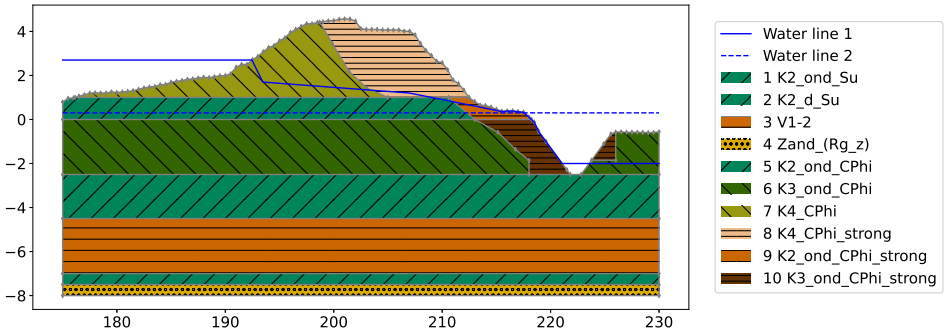


Figure 6.7: Cross section as modeled in RFEM.

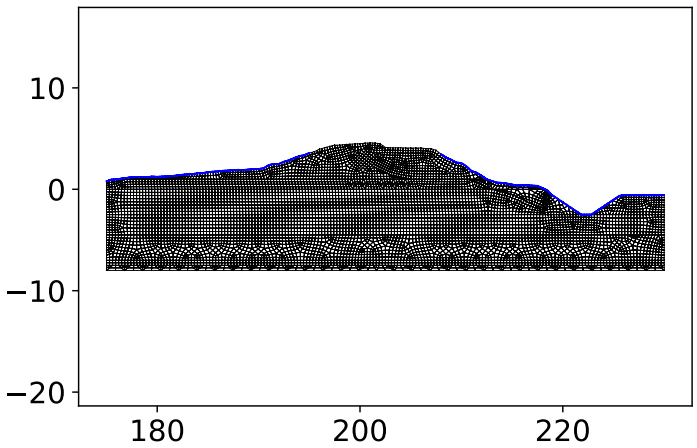


Figure 6.8: Generated mesh for RFEM. The elements of the mesh are 0.3m for the layers that will be modeled stochastically and 0.9m for the other layers.

7

Model Results

In this chapter, the results of the models will be discussed. Firstly, the various scenarios for which the models have been created will be introduced. Following that, the outcomes of deterministic and semi-probabilistic scenarios will be presented. Finally, the results of the probabilistic and Random Finite Element Method (RFEM) scenarios will be shown.

The RFEM code used for this purpose is a TU Delft in-house finite element code (version Dec. 2023), built upon the finite element library by (Smith et al., 2014), as used in, for example, (Varkey et al., 2023).

7.1. Scenarios

Different cases will be explored to examine the impact of spatial variability on macro-stability. These scenarios will be calculated using the Finite Element Method (FEM) code of TU Delft, as described above. The various scenarios are:

0. *Deterministic*: It is important to note that while a deterministic analysis cannot determine the probability of flooding (ENW, 2016), it is still performed for comparison purposes. In a deterministic analysis, only the mean values of the point statistics for S, m, and ϕ are used. This means that spatial variability and statistical uncertainties regarding the number of samples in the dataset are not considered. As a result, no random fields need to be created, and only one schematization is needed, which involves homogeneous layers with values equal to the mean of the dataset.
1. *Semi-probabilistic*: According to Rijkswaterstaat (2021), it is important to consider the spatial variability and statistical uncertainty when determining specific soil parameters. In the case of a semi-probabilistic analysis, characteristic values of S, m, and ϕ should be used for this purpose. These values should be determined to ensure a calculated reliability of 95% for the dike (Hicks, 2012) and can be calculated using Equation 2.4. To investigate the impact of α, various scenarios will be examined:
   (a) $\alpha = 1$
   (b) $\alpha = 0.8$
   (c) $\alpha = 0.75$
   (d) $\alpha = 0.5$

   The characteristic values are presented in Table 6.5.

   A semi-probabilistic analysis differs from a fully probabilistic analysis as it does not require a probability distribution for the soil parameters. Instead of multiple realizations, a single schematization with homogeneous layers is calculated per scenario using characteristic values. The characteristic values take into account the spatial variability of the soil.
2. *Probabilistic*: The probabilistic analyses are performed using Monte Carlo simulation based on soil parameter distributions, including S, m, and ϕ, with input mean and standard deviation. The

standard deviations and means of the point statistics are modified to incorporate the additional uncertainty of spatial variability and the limited amount of samples in the dataset. The influence of α is investigated by considering three cases:

(a) $\alpha = 1$

(b) $\alpha = 0.8$

(c) $\alpha = 0.75$

The spatial variability is modeled by modifying the input standard deviation and mean (see Equations 2.10 and 2.9 respectively), which are presented in Table 6.5. No random fields are used to model the spatial variability. The horizontal and vertical scales of fluctuation, θ_h and θ_v, are assigned a very large value, larger than the problem domain. As a result, each realization consists of homogeneous layers modeled by a stochastic value of the soil parameters. The Monte Carlo analysis is carried out with 200 realizations with the RFEM code.

3. *RFEM*: The RFEM scenario combines the FEM code with random fields. To model the distribution of the structural response, 200 realizations are calculated. For more information on the random fields, please refer to Chapter 2.3.

   To create the random fields, horizontal and vertical scales of fluctuation (θ_h and θ_v) are implemented. Two cases of θ_h and θ_v will be investigated:

   (a) $\theta_h = 20 * \theta_v$ and θ_v based on the values as presented in Table 6.11.

   (b) $\theta_h = \infty$ and θ_v based on the values as presented in Table 6.11. The purpose of this scenario is to investigate the influence of θ_h on the calculation.

   When creating an RFEM model to analyze the influence of variance and mean reduction, it is important that the input parameters do not already account for this, like in probabilistic analyses. To achieve this, an α value of 0 is used to determine the model input, presented in 6.7.

   It is important to consider statistical uncertainty to make a meaningful comparison with probabilistic results, it is important to consider statistical uncertainty. This is done by incorporating a residual correlation factor, denoted as ω, in the Auto Correlation Function (ACF), which was explained in Chapter 6.2.

See Table 7.1 for a summary of the cases.

Table 7.1: Different Random Finite Element Method (RFEM) cases that will be investigated.

Level	Case	α	θ	Number of realizations
0: Deterministic	0	-	-	1
1: Semi-Probabilistic	1a	1	-	1
	1b	0.8	-	1
	1c	0.75	-	1
	1d	0.5	-	1
2: Probabilistic	2a	1	$\theta_v = \infty$ $\theta_h = \infty$	200
	2b	0.8	$\theta_v = \infty$ $\theta_h = \infty$	200
	2c	0.75	$\theta_v = \infty$ $\theta_h = \infty$	200
3: RFEM	3a	0	$\theta_v = \theta_v$ $\theta_h = 20\theta_v$	200
	3b	0	$\theta_v = \theta_v$ $\theta_h = \infty$	200

For each Monte Carlo analysis, 200 realizations are performed. This decision was made based on the

number of realizations in previous studies, such as Arel and Mert (2021), Chen et al. (2020), Pan et al. (2018), and Zhu and Zhang (2015), and to ensure that computation time remains within practical limits.

7.2. Deterministic and Semi-Probabilistic Analysis

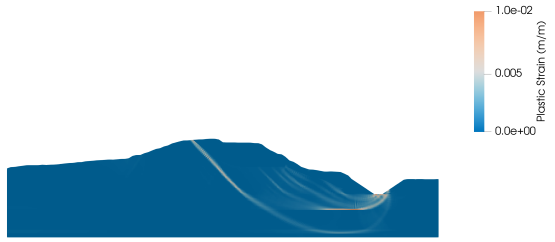
For the deterministic and semi-probabilistic cases, the calculated Factor of Safety (FoS) and failure surface are presented in Table 7.3 and Figure 7.1. The deterministic case has an FoS greater than 1, while the semi-probabilistic safety factors are below 1. If only the deterministic case had been considered, it would have resulted in a false sense of safety. This emphasizes the importance of accounting for uncertainties. Moreover, the FoS is higher for cases with a higher value of α. This is as expected because the characteristic value decreases with an increasing α (see Equation 2.4).

The probability of failure of a cross-section ($P_{f,cs}$) can be calculated using Equation 7.1 (Kanning et al., 2017), as explained in Paragraph A.2.1. The model factor for FEM should be $\gamma_d = 1.06$, according to the WBI2017 (POVM, 2020). Table 7.3 presents the computed failure probabilities. In accordance with the results of the FoS, it can be seen that the P_f decreases with an increasing value of α.

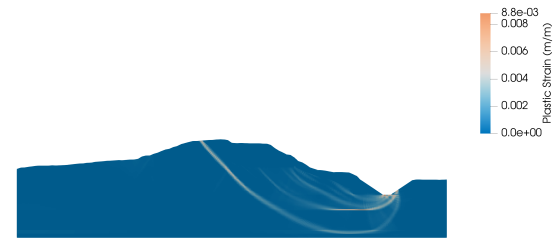
$$\beta_{cs} = \frac{\frac{FoS}{\gamma_d} - 0.41}{0.15} \tag{7.1}$$

Table 7.3: The calculated FoS FEM results of the deterministic and semi-probabilistic analyses.

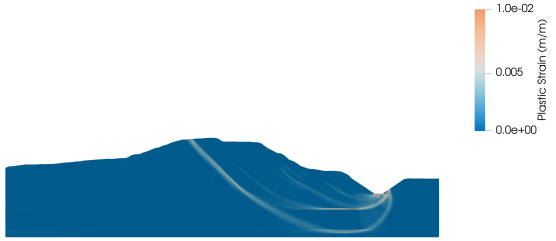
Level	Case	FoS	$P_{f,cs}$ with Equation 7.1
0: Deterministic	0	1.025	$1.023 * 10^{-5}$
1: Semi-Probabilistic	1a ($\alpha = 1.0$)	0.997	$2.022 * 10^{-5}$
	1b ($\alpha = 0.8$)	0.984	$2.748 * 10^{-5}$
	1c ($\alpha = 0.75$)	0.980	$3.016 * 10^{-5}$
	1d ($\alpha = 0.5$)	0.961	$4.653 * 10^{-5}$



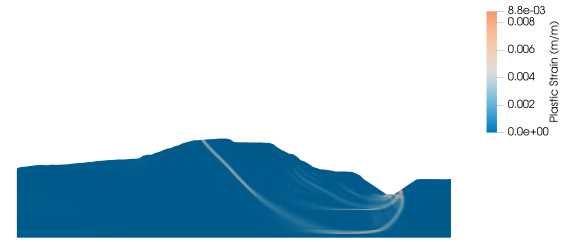
(a) Deterministic analysis. FoS $= 1.025$



(b) Semi-probabilistic case 1a. FoS $= 0.997$

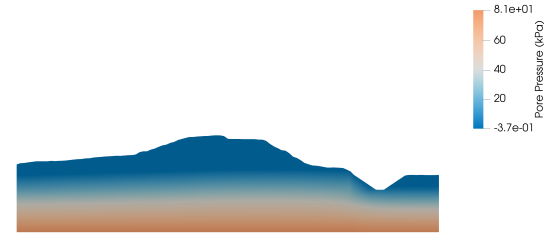


(c) Semi-probabilistic case 1b. FoS $= 0.980$

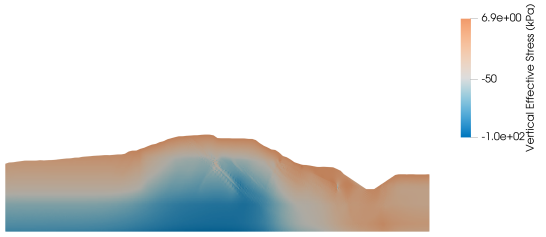


(d) Semi-probabilistic case 1c. FoS $= 0.961$

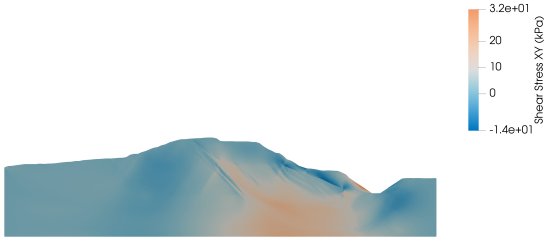
Figure 7.1: Calculated plastic strain for the deterministic case and semi-probabilistic cases.



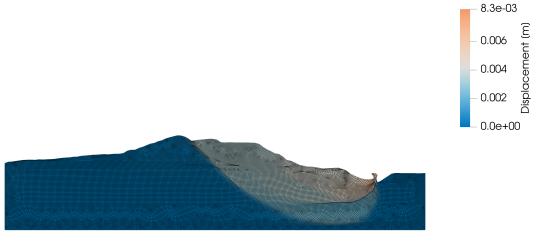
(a) Pore pressure.



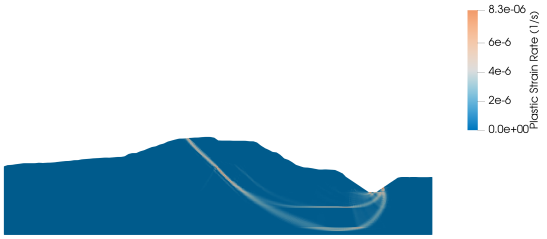
(b) Vertical effective stress.



(c) Shear stress.



(d) Displacements (color) and displaced mesh (scaled by a factor of 300).



(e) Strain rate at failure.

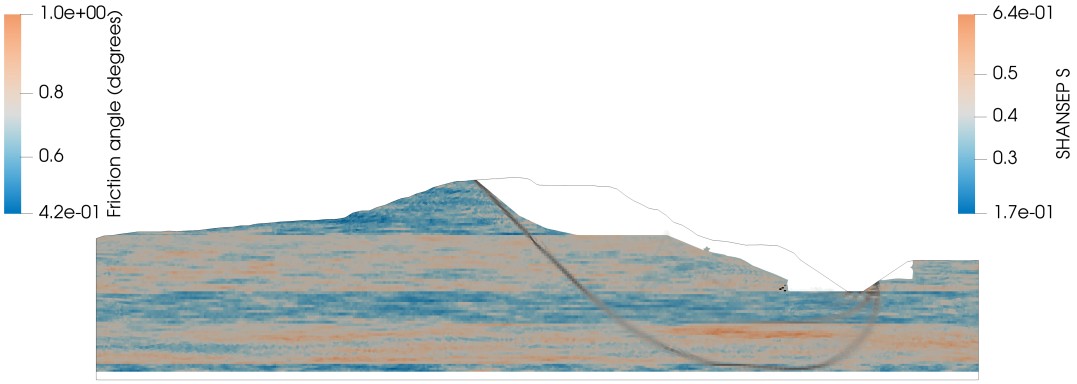
Figure 7.2: Results of case 1b. FoS $= 0.98$

7.3. Probabilistic and RFEM Analyses

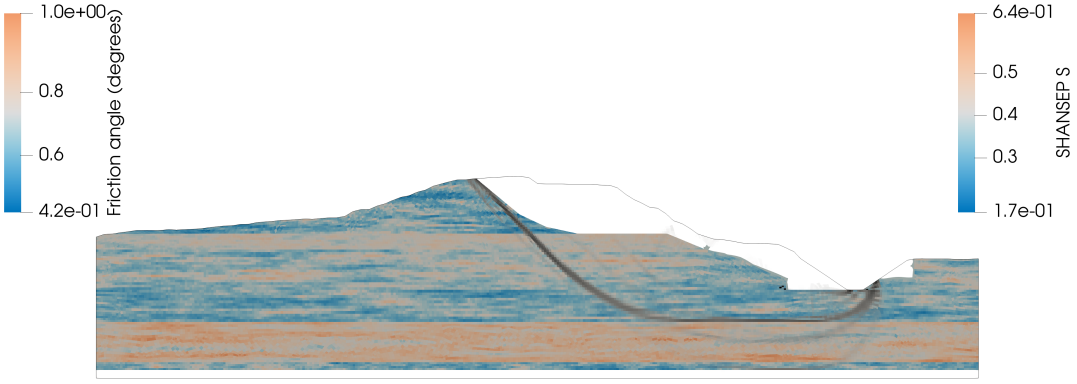
To perform probabilistic and RFEM analyses, 200 realizations were generated. Each realization was given a distinct group of soil parameters for the homogeneous soil layers in the probabilistic analyses. While for the RFEM cases, each realization modeled both strong and weak zones. As a result, each realization in both cases led to different failure surfaces and calculated factors of safety. The soil parameter S, m and ϕ were modeled with a probability distribution for the respective soil parameter. This is not the case for θ, which determines the variation of strong and weak zones in the random fields and remains constant across all realizations. Note that m is not modeled with a random field in RFEM because it is assumed that the value of m does not significantly affect the FoS.

Some results of the RFEM analysis scenario 3a are displayed in Figure 7.3. Various observations can be made for the different realizations:

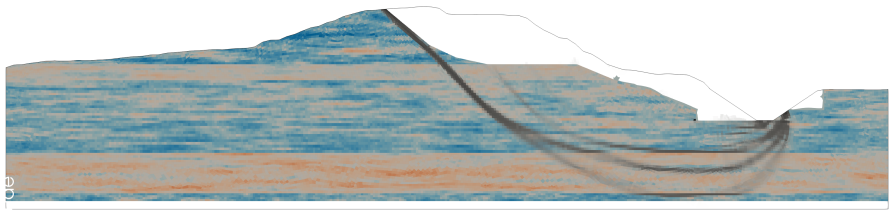
- *Realization a*: In this case, the deep failure surface propagates through the peat layer to the lower clay layer. Consequently, it results in a low FoS of 0.9937. One thing to note is that a part of the active side of the failure surface is located in a relatively weak zone in the peat layer.
- *Realization b:* This realization has the highest FoS of all the examples with FoS $= 1.0406$. The failure surface is shallow and does not propagate through the peat layer.
- *Realization c:* The failure surface in this example consists of three failure surfaces, two of which propagate through the peat layer. The middle one is located in a weaker zone in the peat layer. The FoS of this example is 1.0234.
- *Realization d:* This realization has a deep failure surface with a part of the active side located in a weaker zone of the peat layer, which results in the lowest FoS of all the examples with a value of 0.9906.



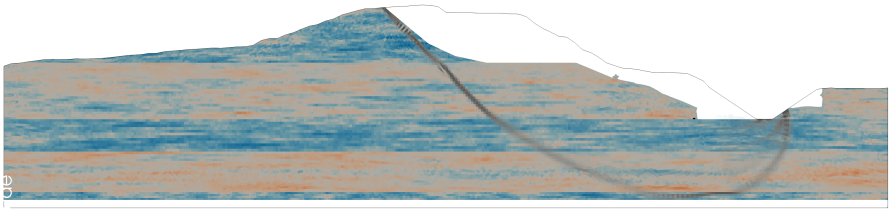
(a) FoS $= 0.9937$



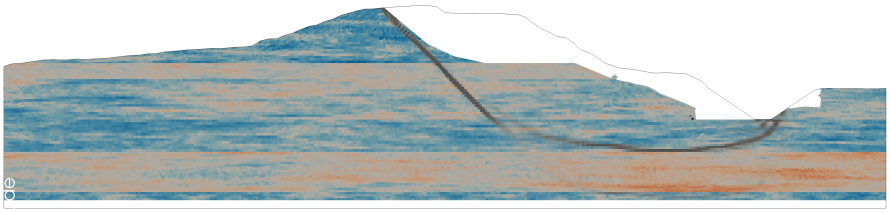
(b) FoS $= 1.0406$



(c) FoS $= 1.0234$



(d) FoS $= 0.9906$



(e) FoS $= 1.0375$

Figure 7.3: Results of different realizations of case 3a. Drained layers are represented by a value for ϕ, while undrained layers are colored by their value of S. The plastic strain rate is depicted in black.

- *Realization e:* Some strong zones are located in the peat layers through where the failure surface cannot propagate. As a result, the failure surface is shallow with a FoS of 1.0375.

It can be observed from Figure 7.3 that the failure surfaces propagate through the dike at a steep angle in the first 7 meters, crossing both strong and weak zones in this area. This part of the failure surface appears consistent across all realizations and is primarily influenced by the locations of strong and weak zones for either ϕ or S. The peat layer seems to have the most significant impact, particularly since it has larger scales of fluctuation in both the horizontal and vertical directions (θ_v and θ_h).

It is worth noting that among the different examples, those with the lowest Factor of Safety (FoS) (a and d) both have a deep failure surface, where the active side passes through a weaker zone in the peat layer. However, it is clear from the other 195 realizations that a deeper failure surface does not always lead to a lower FoS, and a lower FoS is not always associated with a part of the active side of the failure surface passing through a weak zone in the peat layer. Yet, most shallow failure surfaces propagate precisely along the boundary between the clay and peat layer, highlighting the importance of the location of the peat layer. Furthermore, throughout all the realizations, the first 7 meters of the failure surface are all at the same location. This part does not ”seek out” a weak zone in the soil layer.

The calculated FoS for each calculation can be used to generate a histogram, from which a Probability Density Function (PDF) and Cumulative Density Function (CDF) can be computed. For scenario 2c, the results are presented in Figure 7.4. The black dots represent the calculated factors of safety and their corresponding cumulative density value. Please note that the displayed number of points does not equal the number of realizations. This is due to the limited precision of the strength reduction algorithm in FEM. Appendix B describes this in more detail. As only 200 realizations were created, a theoretical CDF must be fitted to analyze the full CDF. In this case, the normal and log-normal CDFs are fitted to see which function best fits the lower tail. The lower tail represents the small probabilities of failure (P_f), which is interesting for civil engineers as they want to design for a small P_f. The right part of the plot zooms in on the data in the lower tail. The log-normal function fits these data points somewhat better, so this function is selected to represent the distribution. The difference between the normal and log-normal fits is insignificant due to the small standard deviation of the CDF. This process is repeated for the other scenarios. Figure 7.5 displays the best fit of the CDF for each scenario.

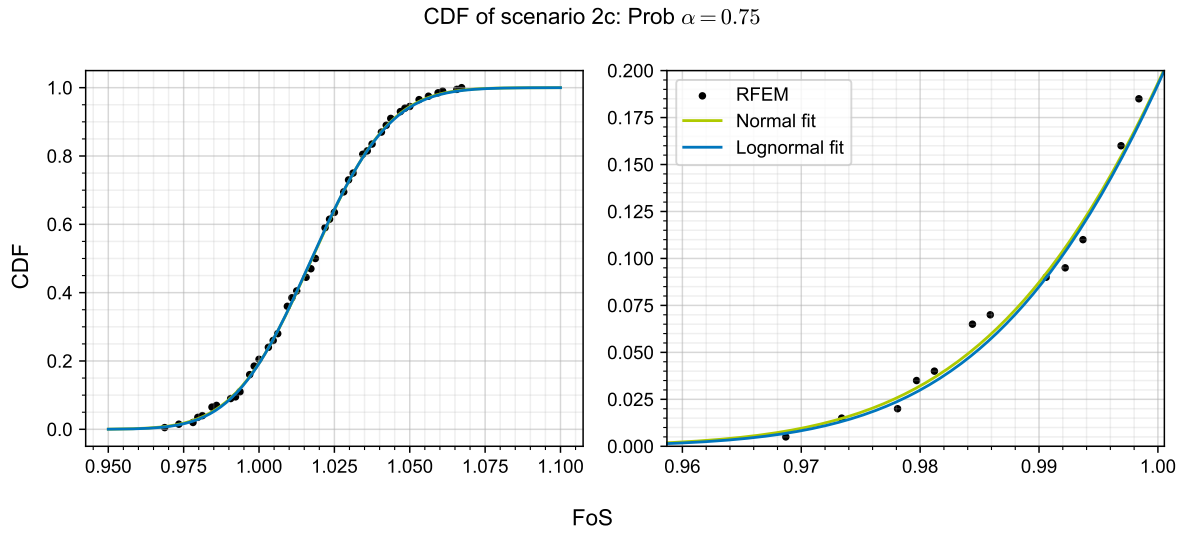


Figure 7.4: Cumulative Density Function (CDF) of the data and the fitted normal and log-normal functions for the probabilistic scenario 2c. Please note that the displayed number of points does not equal the number of realizations. This is due to the limited precision of the strength reduction algorithm in FEM. Appendix B describes this in more detail.

It is important to note that the probabilities of failure determined from the probabilistic and RFEM results are conditional on the chosen water level. As a result, these probabilities do not represent the full probability of failure of the cross-section due to macro-stability. To obtain this full probability of failure, different water levels and their probability of occurrence should be considered. Therefore, the results

presented in Table 7.5 cannot be compared to the norm for this particular cross-section or the outcomes of the semi-probabilistic analysis. For simplicity, the conditional probabilities of failure, calculated with the probabilistic and RFEM scenarios, are referred to as the probability of failure.

Table 7.5: The calculated (conditional) probability of failure P_f and standard deviation of the fitted CDF (σ_{CDF}) of the FEM results of the probabilistic and RFEM analyses.

Level	Case	P_f	σ_{CDF}
2: Probabilistic	2a: $\alpha = 1.0$	0.024	0.011
	2b: $\alpha = 0.8$	0.117	0.017
	2c: $\alpha = 0.75$	0.192	0.020
3: RFEM	3a: $\theta_h = 20 * \theta_v$	0.126	0.015
	3b: $\theta_h = \infty$	0.216	0.019

The full CDF for all the probabilistic and RFEM scenarios is displayed in Figure 7.5. The standard deviations of the fits are presented in Table 7.5. Scenario 2a with $\alpha = 1.0$ is the lower limit for the width of the CDF, as it has the lowest input values for σ, which considers the full averaging of the local variability. The widest CDF corresponds to scenario 2c, which would result in the safest designs, as it considers the least amount of local averaging. The CDF of the second probabilistic scenario lies between the CDFs of scenario 2a and 2c, as the input standard deviations are also between those input standard deviations.

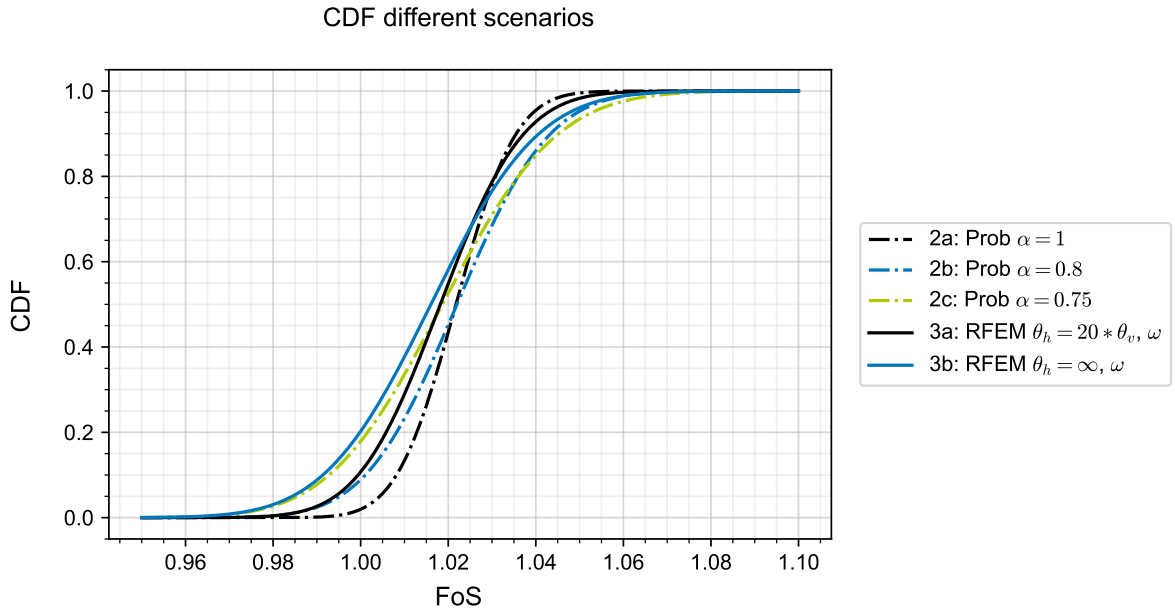


Figure 7.5: Cumulative Density Function (CDF) of the fitted functions for the different scenarios.

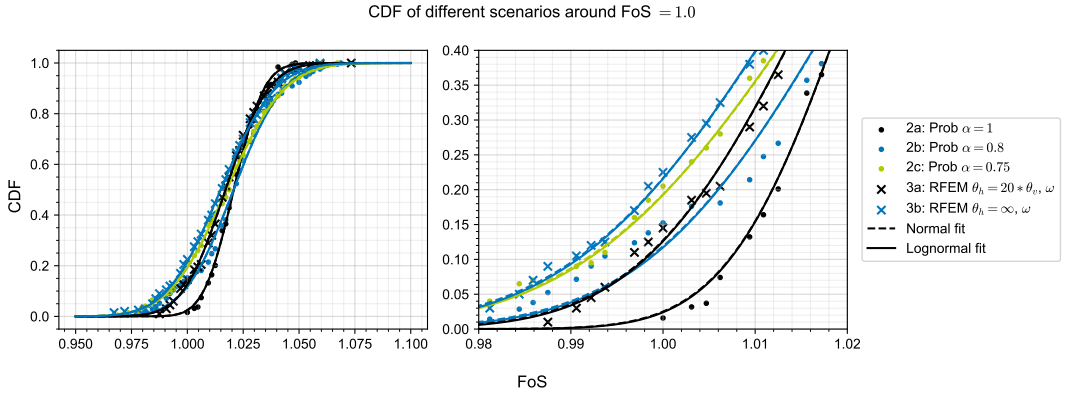
The RFEM scenario does not consider local averaging in the input distribution, making the input σ larger than those for the probabilistic cases. This model experiences local averaging by modeling stronger and weaker zones in the soil layers (see Figure 7.3). Therefore, the CDF should be wider than the CDF of scenario 2a, where full averaging is assumed. This is true and can also be observed in Figure 7.5 and Table 7.5, where the RFEM CDF of scenario 3a lies between the CDF of scenario 2a and 2b. In the lower tail, scenario 2c seems to approximate the RFEM result of scenario 3a, while in the upper tail, scenario 2a is closer.

In scenario 3b, θ_h is larger than the width of the cross-section ($\theta_h = \infty$), which is an extreme situation. This leads to a significant difference between the CDF fits of scenarios 3a and 3b. As a result, the CDF

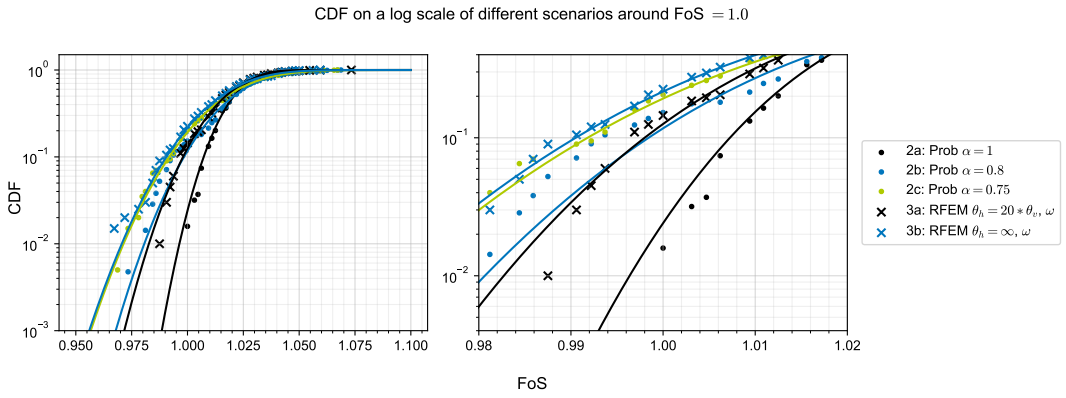
becomes wider with a lower mean, as it tends to ”seek out” the weaker zones more. This situation is best described by probabilistic scenario 2c with $\alpha = 0.75$ in the lower tail, while it approaches probabilistic scenario 2b with $\alpha = 0.8$ in the upper tail.

It is observed that the probabilistic scenarios do not perfectly match the scenarios of RFEM in both tails. This difference is due to the mean reduction of the RFEM cases. As shown in Figure 7.5, the means of the CDFs are shifted to the left compared to the probabilistic cases. This mean reduction occurs because the failure surfaces in the RFEM model ”seek out” the weak zones. Even when the underlying input distribution is identical, the failure surface’s ability to propagate through the weaker zones causes a decrease in the response’s mean (mean reduction) (Hicks et al., 2019). Scenario 3b shows a lower mean, indicating that a higher θ_v leads to a more significant mean reduction. The current model for probabilistic calculations does not account for this mean reduction. Therefore, with the current method, no fit that completely aligns with the results of the RFEM scenario could be found.

To ensure the reliability of a dike, it is important to have a low probability of failure. To evaluate this probability, the lower tails of the Monte Carlo simulations are depicted in Figure 7.6a. A close-up view around the value of FoS $= 1.0$ is provided in the figure to investigate differences between scenarios at the location where the (conditional) probability of failure (P_f) is calculated. The fits of scenarios 2c and 3a are similar in the lower tail around FoS $= 1.0$. However, scenario 2c would lead to a slightly safer design as its data points lie higher than the points of scenario 3a, which can be clearly seen in Figure 7.6b, where the CDF is plotted on a log-scale. Scenario 2b is the scenario that most closely resembles scenario 3b in the lower tail and is almost equal at FoS $= 1.0$. However, when the FoS increases, this scenario becomes less conservative compared to scenario 3b.



(a) CDF



(b) CDF on a logarithmic scale.

Figure 7.6: Cumulative Density Function (CDF) and their lower tail of the fitted functions for the different scenarios. Please note that the displayed number of points does not equal the number of realizations. This is due to the limited precision of the strength reduction algorithm in FEM in Appendix B describes this in more detail.

7.4. Impact of Changing α to 0.8

The results of the probabilistic and RFEM analysis show that a probabilistic analysis with $\alpha = 0.8$ (scenario 2b) provides a better fit to RFEM scenario 3a. The alteration of the value for α will influence the probability of failure of the cross-section ($P_{f,cs}$) and, therefore, also the probability of failure of a dike section ($P_{f,sec}$) and trajectory ($P_{f,traj}$). This influence will be further investigated in this paragraph.

The dike trajectory's signal norm is described in Chapter 6.1 and is $1/1000$, while the lower limit norm is $1/300$. Following WBI2017, the failure probability budget (faalkansbegroting in Dutch) for macro stability is 0.04. Using this budget, the norm for the failure mechanism macro stability can be calculated by multiplying it with the signal and lower limit norms. Therefore, only these norms will be considered as this research investigates the effect on the failure mechanism macro stability.

First, the influence of $P_{f,cs}$ is calculated by performing a semi-probabilistic and FORM analysis in D-Stability with Uplift-Van, where the model factor is also considered. The results are presented in Table 7.7. It is observed that the $P_{f,cs}$ decreases 40% for the semi-probabilistic analysis, while the conditional $P_{f,cs}$ decreases by 17.2% for the probabilistic analysis.

The values for $P_{f,cs}$ differ significantly between the semi-probabilistic and FORM analyses. This is because the results for the probabilistic calculation represent the conditional probabilities of failure to the outside water level. To obtain the full probability of failure for the cross-section with a probabilistic analysis, probabilities of failure for multiple water levels should be calculated, combined with the probability of occurrence of this water level. However, due to time constraints, this calculation could not be performed. The full $P_{f,cs}$ can be computed from the semi-probabilistic calculation through the calibration formula. Therefore, the change in $P_{f,cs}$ by 40% will be used to investigate the effect of different α values.

Table 7.7: An overview of the calculated probabilities of failure of the case study cross-section (P_f) and factors of safety (FoS) with the current method of the WBI2017, performed in D-Stability with Uplift-Van. *The P_f has been obtained via a different method for the semi-probabilistic analysis compared to the FORM analysis. For the first level, P_f is calculated via the calibration formula as defined in the WBI2017. For the FORM analysis, P_f is determined by finding the design point. Both methods consider the model factor.

Level	α	**FoS**	P_f*	β
Semi-probabilistic	0.8	0.936	$2.28 * 10^{-4}$	3.51
	0.75	0.915	$3.80 * 10^{-4}$	3.37
FORM	0.8	-	$4.90 * 10^{-2}$	1.66
	0.75	-	$5.91 * 10^{-2}$	1.56

To determine the impact of the change in $P_{f,cs}$ on the $P_{f,traj}$, three different tests are conducted and compared with the original case ($\alpha = 0.75$, without reduction of $P_{f,cs}$). When assembling the dike sections for $P_{f,traj}$, data from Tigchelaar and van Haaren (2022) and older WBI calculations are used. The first test involves reducing only $P_{f,cs}$ for section 20 by 40%. In the second test, only the dike sections with the same order of magnitude $P_{f,cs} > 1 * 10^{-5}$ are reduced by 40%. Finally, in the third test, the $P_{f,cs}$ for all dike sections are reduced by 40%.

The $P_{f,cs}$ provides a starting point for calculating the probabilities of failure for the cross-section ($P_{f,sec}$) and the trajectory ($P_{f,traj}$) of the dike. The $P_{f,sec}$ can be calculated from $P_{f,cs}$ by factoring in the length effect in the dike section. Each section can be assigned a category, defining the safety level of that section. The $P_{f,traj}$ can be calculated by assuming that all dike sections are independent, multiplying the probabilities of failure of all dike sections. To determine the safety level of the dike trajectory, the $P_{f,traj}$ value is compared to the signal value and the lower limit norm value of the dike trajectory for macro stability. This process is described in more detail by Diermanse (2021). The results for this case study are presented in Table 7.9. Please note that there is a slight difference between the original $P_{f,cs}$ in Table 7.9 and the semi-probabilistic calculation with $\alpha = 0.75$ in Table 7.7. This is due to a small difference in the model of the dike, which is not relevant to the purpose of these tests.

Table 7.9: An overview of the influence of adopting another value of α on the probability of failure of dike section 20 in trajectory 34-2 for different tests. Data used for this calculation is based on Tigchelaar and van Haaren (2022).

Case	$P_{f,cs}$	$P_{f,sec}$	**Category section**	$P_{f,traj}$	**Category trajectory**
Original	$3.02*10^{-4}$	$5.4*10^{-4}$	$\mathrm{V_v}$	$9.87*10^{-3}$	D
Test 1 (section 20)	$1.81*10^{-4}$	$3.2*10^{-4}$	$\mathrm{V_v}$	$9.66*10^{-3}$	D
Test 2 (same order of magnitude)	$1.81*10^{-4}$	$3.2*10^{-4}$	$\mathrm{V_v}$	$5.94*10^{-3}$	D
Test 3 (all sections)	$1.81*10^{-4}$	$3.2*10^{-4}$	$\mathrm{V_v}$	$5.93*10^{-3}$	D

From the results of these tests, a few observations can be made regarding the stability assessment of the dike:

- It was found that the probability of failure of dike section 20 decreased by 40.74% across all tests compared to the original case. However, it is worth noting that the safety category for this section remained the same as in the original case.
- The probability of failure of the dike trajectory also decreased, with a reduction of 2.13% in the first test, and 39.65% and 39.72% in the second and third tests, respectively. Again, the safety category for the trajectory did not change from the original case.
- The small difference between the probability of failure in the second and third tests highlights the impact of dike sections with a higher probability of failure on the overall probability of failure of a dike trajectory.

8

Discussion and Conclusion Modeling

The purpose of using Random Finite Element Method (RFEM) for modeling the spatial variability is to create a model that closely approximates reality by simulating the spatial variability of soil using random fields. The aim is to compare these results with the results of (semi-)probabilistic analyses based on the Legal Assessment Instrument (Wettelijk BeoordelingsInstrumentarium) (WBI) 2017 guidelines in the Netherlands. This primarily focuses on investigating the influence of variance and mean reduction.

The study conducted four different analyses to compare the results. These analyses included a deterministic analysis, four semi-probabilistic scenarios, three probabilistic scenarios, and two RFEM scenarios. The (semi-)probabilistic scenarios differed based on different input values that were determined by considering different ratios between local and regional variance (α). The current method assumes that local variance can be fully averaged out (Calle et al., 2021), and the parameter α determines which part of the sample variance can be averaged out. A higher value of α results in a lower input standard deviation for the probabilistic analysis and a characteristic value that is closer to the mean in the semi-probabilistic analysis.

This chapter discusses the results of the different types of analyses to draw conclusions about modeling spatial variability.

8.1. Deterministic and Semi-Probabilistic Analyses

In order to arrive at a conclusion regarding the deterministic and semi-probabilistic analyses, a comparison of various scenarios is conducted. Additionally, three factors that have an impact on the results produced by the analyses are investigated and discussed in this paragraph. Based on this, at the end of the paragraph. a conclusion can be drawn regarding the influence of spatial variability in semi-probabilistic analyses.

8.1.1. Discussion Results

From the calculated FoS, it can be observed from Table 7.3 that the FoS decreases with decreasing values of the ratio between local and regional variance (α). To quantify the influence of incorporating spatial variability, the relative changes in FoS and P_f compared to the standard calculation (scenario 1c with $\alpha = 0.75$) can be calculated. The maximum relative change can be found for the scenario 1d where a decrease of 1.94% and an increase of 54.28% can be calculated for the FoS and P_f, respectively.

According to a study by Teixeira and Wojciechowska (2023), a significant change in the failure probability ranges from a factor 10 to $1,000$. As a result, it can be argued that the change in the probability of failure due to a different value of α may not be significant. In Chapter 7.4, the impact of changing α from 0.75 to 0.8 on the probability of failure of a dike trajectory due to macro-stability was analyzed. The results showed that the safety category of both the dike section and trajectory, based on the comparison to the norm, did not change. This further demonstrates that the influence is not significant. It is important to note that although it was concluded that the impact does not alter the safety category of the dike for macro stability, this study was only conducted for one dike trajectory. Therefore, for other

sections of the dike, with varying datasets, soil types, geometries, and hydraulic conditions, there is a possibility that the impact may be more significant.

8.1.2. Comparison FEM and D-Stability

To be able to compare the RFEM results with the WBI2017 method that uses Limit Equilibrium Method (LEM), it is important to validate the RFEM outcome. For this purpose, a comparison is made between the semi-probabilistic results obtained with LEM and Finite Element Method (FEM), with the aim of investigating the differences.

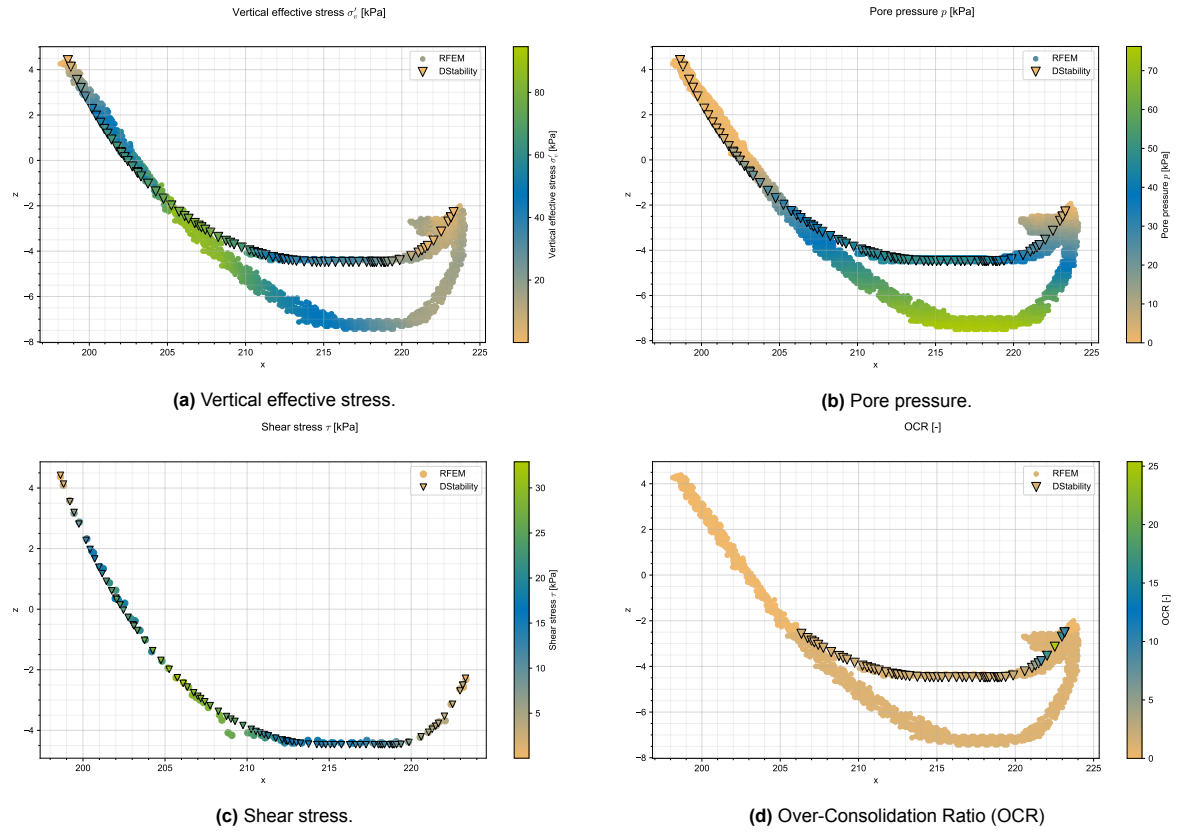
The Factor of Safety (FoS) for scenario 1b is the same as the result obtained from the reference D-Stability model (see Chapter 6.4). Figure 8.1 shows the failure surfaces and the stresses calculated along the failure surface to further investigate any differences. The calculated Factor of Safety (FoS) is directly related to the shear strength of the material along the failure surface as shown in Figure 8.1c. The shear strength calculation is performed by rotating the vector so that it is parallel to the failure surface at all points depicted in the figure. The shear strength value obtained through FEM corresponds to the D-Stability shear strength value, resulting in an identical FoS value for both methods.

The shear strength in the undrained layers is calculated using the SHANSEP framework (see Appendix A.3.3) with Equation A.14. The vertical effective stress (σ'_v) and the Over-Consolidation Ratio (OCR) are essential in this shear strength formulation. Therefore, they are also shown in Figure 8.1a and 8.1d, respectively. It can be observed that both the σ'_v and the OCR differ from coordinate $(220, -4)$ to $(223, -2)$, corresponding to the passive part of the failure surface. In D-Stability, σ'_v has a smaller value than in FEM, while the OCR is much larger. Since the strength increase component m is less than 1 (0.86), the more significant increase in OCR is not directly translated to an equal increase in τ. This effect is then canceled due to the D-Stability model's decrease in σ'_v.

The effective stress is calculated using the total and pore pressure (p). The pore pressure is illustrated in Figure 8.1b and does not show any differences in the passive zone of the failure surfaces. Therefore, the differences in the calculation procedure of the stresses are the most probable reason for the difference in σ'_v.

In D-Stability, the user defines the OCR or another form of pre-consolidation input in the form of a state point. These can be defined anywhere in the geometry to define the loading history. In this layer, Tigchelaar and van Haaren (2022) defined a Pre-Overburden Pressure (POP) of 14.77kPa. The OCR is then calculated by this input and taking into account the daily conditions of the dike. In contrast, in FEM, the state of the soil is only defined by the daily conditions and then calculated using Equation 2.27. It can be seen that FEM calculates $OCR = 1$ in all elements in the cross-section. This means that the maximum effective stress equals the calculated effective stress in the element. This difference in formulation causes the difference in values.

It is clear from Figure 8.1 that FEM calculates a second sliding surface that is deeper than the shallow sliding surface calculated by D-Stability. Unlike D-Stability, FEM can generate multiple failure surfaces. To investigate the impact of the secondary sliding surface on the FoS, the two deepest soil layers in FEM are made very strong so that the failure surface cannot propagate through these layers. The results of this calculation are shown in Figure 8.2, where only a shallow failure surface is calculated. The resulting FoS value is 0.988 with the strong layers, which is only 0.008 larger than the original result for scenario 1b (refer to Table 7.3). Therefore, it shows that the difference is minimal, and the deeper failure surface does not significantly affect the stability of the dike.



(a) Vertical effective stress.

(b) Pore pressure.

(c) Shear stress.

(d) Over-Consolidation Ratio (OCR)

Figure 8.1: Comparison of the results of scenario 1b ($\alpha = 0.75$) for the FEM and D-Stability calculations.

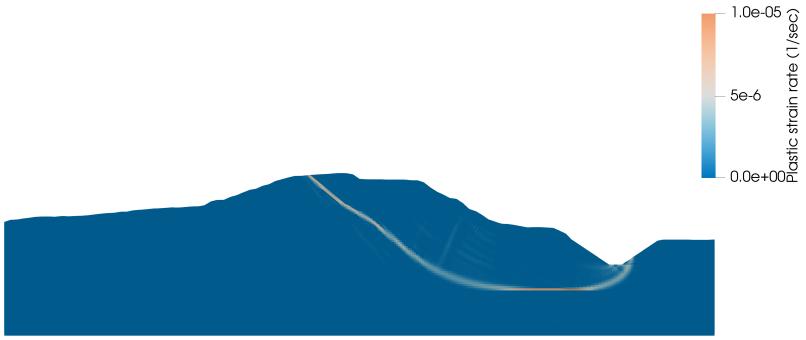


Figure 8.2: Calculated strain rate for scenario 1b, where the deepest layers are modeled very strongly to investigate the effect of the deeper failure surface. The calculated FoS $= 0.988$.

8.1.3. Difference of using Vertical or Principal Stresses in FEM

In FEM, the undrained shear strength is typically calculated with the major principal effective stress, following Equations 2.26 and 2.27 in Chapter 2.3. Appendix A.3.4 discusses how the major and minor principal stresses rotate along the failure surface. From this, it can be seen that the direction of the major principal stress does not remain vertical along the failure surface.

D-Stability (van der Meij, 2023) follows the original formulation of Ladd and Foott (1974), where the undrained shear strength is calculated via Equation A.14. In this formulation, the vertical effective stresses define the shear strength. However, because of the rotation of major principal stress along the failure surface, it is not equal to the vertical effective stress.

The influence of this definition on the FoS is tested by changing the formulation in the FEM code. The result can be seen in Figure 8.3. The calculated FoS is 1.072, which is approximately a 9% increase

compared to the reference D-Stability calculation of FoS 0.98.

Since this research aims to compare the current practice to a realistic case, and the current practice is to use D-Stability, the formulation as defined in D-Stability (van der Meij, 2023) in Equation A.14 is used in FEM. However, because of the calculated difference in this research, this effect should be further investigated in other cross-sections.

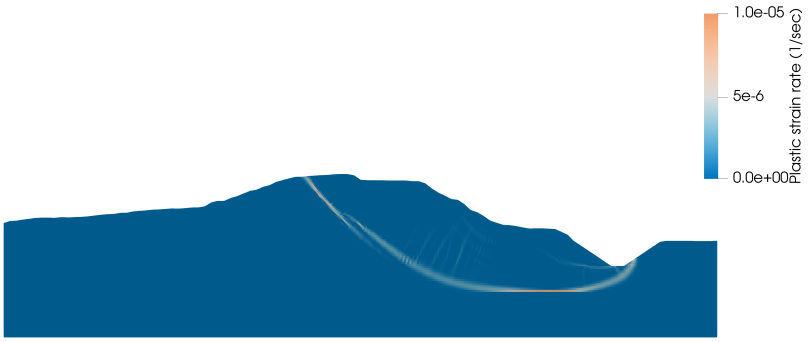


Figure 8.3: Calculated strain rate for scenario 1b, where the shear strength is calculated using the major principal effective stress instead of the vertical effective stress. The calculated FoS $= 1.072$.

8.1.4. Influence of SHANSEP m

When creating the input for the models, it was assumed that the strength increase component of the SHANSEP model (m) does not have a significant influence on the calculated FoS or P_f. Therefore, the characteristic value was not further reduced, and the input standard deviation was not increased to consider additional statistic uncertainty and spatial variability in this parameter during (semi-)probabilistic calculations. In the RFEM calculations, ω was set to 1.0, which prevented the creation of random fields for m. To validate this assumption, a sensitivity analysis for m was performed to investigate its influence on the calculated FoS in semi-probabilistic analyses.

According to Rijkswaterstaat (2021), m ranges between 0.5 and 1.0. Therefore, 15 values in this range were implemented in both FEM and D-Stability calculations. Just as in the original model, m is the same for all undrained soil types (see Table 6.1). The results can be seen in Figure 8.4.

For the FEM model, all calculated FoS values are equal, as the calculated OCR is 1, causing m not to influence the FoS. For the D-Stability model, the FoS increases as m increases. The FoS values with m between 0.714 and 0.857 are lower than the other results due to a consistently different shape of the failure surface. Here, the failure surface has a shorter horizontal part, and the active side of the failure surface is larger, causing a decrease in strength and, therefore, a decrease in FoS.

For the whole range of m, it can be observed that the FoS has a variation of 0.151 which ranges from 0.901 to 1.052. Based on a mean value of $\mu_m = 0.9$ and a standard deviation of $\sigma_m = 0.03$ (Rijkswaterstaat, 2021), the 95% confidence interval for m is between 0.841 and 0.960. If only the values within this interval are considered, instead of the entire range, the FoS ranges from 0.98 to 1.04.

It should be noted that the conducted sensitivity analysis did not consider a large number of values for the parameter m. Therefore, it is recommended to perform this analysis with a larger number of data points in future studies, especially to investigate the changes in the shape of the failure surface and the sudden jump in the values of the FoS.

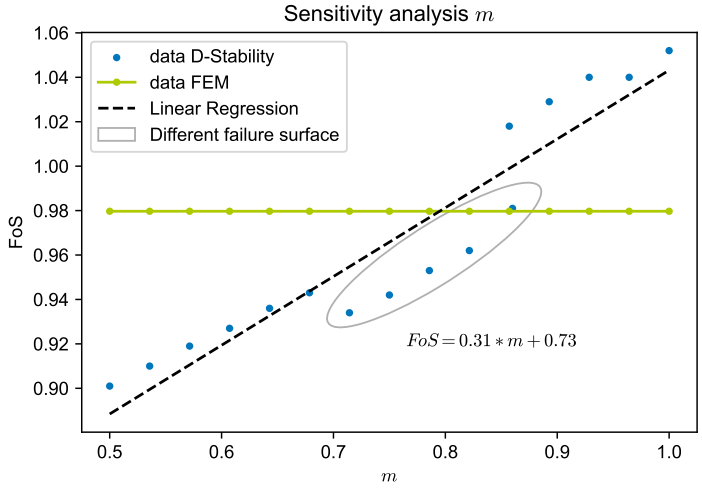
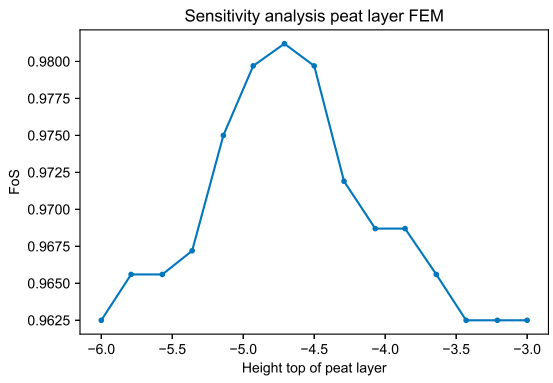


Figure 8.4: The results of a sensitivity analysis for SHANSEP m.

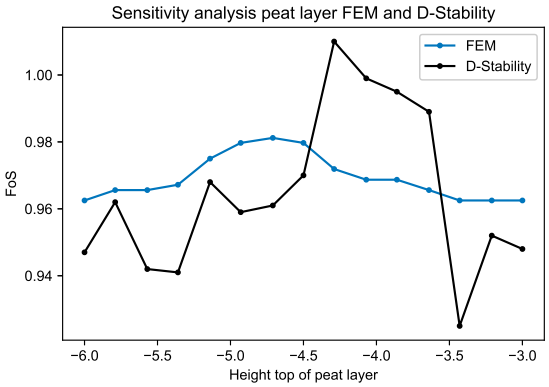
8.1.5. Influence of the Location of the Peat Layer

In Paragraph 7.3, it was noted that for the RFEM scenarios, the shallow failure surfaces often seek out the boundary between the clay layer and the top of the peat layer. To investigate the impact of the location of the peat layer, a sensitivity analysis is performed. The CPTs for the dike section indicated that the top of the peat layer could be situated somewhere between -6 and -3m NAP. Therefore, 15 values within this range are tested in semi-probabilistic calculations using FEM and D-Stability. The parameter input is based on scenario 1b. The results of this analysis are presented in Figure 8.5.

The FEM results show that when the peat layer starts at the most shallow depth (-3m NAP), the FoS is low. However, it increases when the peat layer becomes deeper, reaching a maximum when the top of the layer is located around -4.71m NAP and then decreases again until it reaches the same minimum when the top is at -6m NAP. All shallow failure surfaces follow the top of the peat layer, as depicted in Figure 8.6. In cases where the top of the peat layer is located between -3 and -4.29m NAP, it is observed that deep failure surfaces that propagate horizontally through the deepest clay layer are dominant, as seen in Figure 8.6a. This type of failure is often combined with a secondary shallow failure surface that follows the top of the peat layer. The more shallow the peat layer, the lower the calculated FoS. From the point where the top of the peat layer is located at -4.71m NAP and onwards, shallow failure surfaces become more dominant, as shown in Figures 8.6b and 8.6c. In this case, the FoS decreases as the peat layer is located deeper in the subsurface.

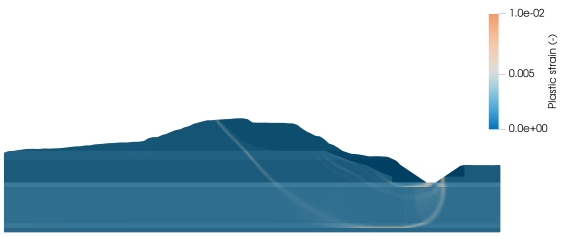


(a) Only the FEM results.

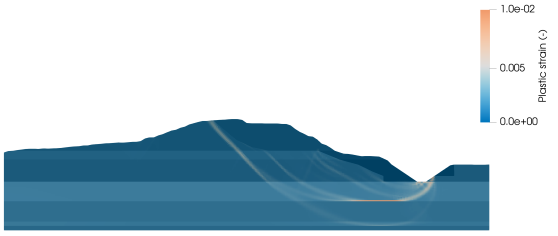


(b) FEM and D-Stability results.

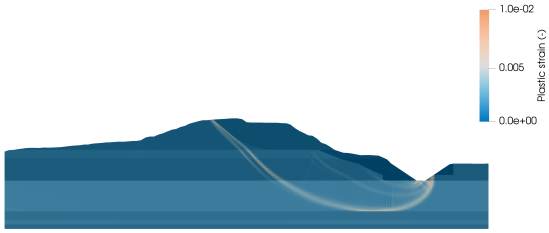
Figure 8.5: Calculated FoS for different locations of the top of the peat layer in FEM and D-Stability semi-probabilistic calculations.



(a) Top peat layer at -3m NAP.



(b) Top peat layer at -4.71m NAP.



(c) Top peat layer at -6m NAP.

Figure 8.6: Failure surfaces of FEM calculations with different depths of the top of the peat layer.

On the other hand, the results of the D-Stability analysis, as shown in Figure 8.5b, are inconclusive. No trend or pattern can be observed in this case. When examining the calculated failure surfaces, it can be seen that they often propagate through the middle of the peat layer. The comparison between FEM, where the failure surface follows the boundary of the peat layer, indicates that because of the constrained shape of the failure surfaces in D-Stability with the Uplift-Van model, the model cannot properly follow the top of the peat layer. The Spencer model in D-Stability allows for a more flexible failure surface shape compared to other models. This could potentially allow it to more accurately follow the peat layer boundary. However, this was not investigated in this study. Therefore, it is recommended to further investigate this behavior, together with a sensitivity analysis with more data points.

A comparison between sensitivity analysis for the parameter m and the location of peat can be used to examine their impact on the FoS. In this case study, changing m within its 95% percentile results in a FoS difference of 0.118, while changing the location of the peat layer by 3m results in a FoS difference of 0.019. These findings suggest that, for this particular case study, the effect of m is more significant than the location of the peat layer. However, because this case study considers a relatively strong peat layer it is recommended to investigate the impact of the location of a weaker peat layer on FoS to gain a more comprehensive understanding of its influence.

8.1.6. Conclusions

The following conclusions can be drawn from the semi-probabilistic analyses:

- By increasing the α value, the FoS increases, and P_f decreases. The maximum change in P_f is 54.28%.
- When α changes from 0.75 to 0.8, the probability of failure of a dike cross-section changes decreases with 40% when calculated with a semi-probabilistic calculation in D-Stability. The safety category of a dike trajectory due to macro-stability does not change. When this is translated into the probability of failure of a dike section and trajectory, it can be seen that the safety category does not change. Therefore, it can be concluded that the change in α does not have a significant impact on the safety level of dike trajectory 34-2 and its dike section 20.
- A comparison between FEM and D-Stability shows that FEM can be used for evaluating the stability of Dutch dikes when performing a semi-probabilistic analysis. However, it was found that the vertical effective stress must be adopted when calculating the shear strength in layers modeled by SHANSEP.
- Considering the influence of m and the location of the peat layer on the FoS, it is observed that changing m results in a higher difference in FoS than changing the location of the peat layer.

Therefore, for this case study, it can be concluded that the influence of m is more significant than the location of the peat layer.

- The sensitivity analysis of the location of the peat layer indicates that the Uplift-Van model in D-Stability does not follow the top of the peat layer because of its constrained shape.

8.2. Probabilistic and RFEM Analyses

In this paragraph, the outcomes of both the probabilistic analysis and RFEM analyses will be discussed. The impact of incorporating spatial variability on the resulting distribution will be investigated, along with the influence of the horizontal scale of fluctuation and statistical uncertainty on the results. Finally, the conclusions on the influence of spatial variability on the calculated P_f in probabilistic and RFEM calculations will be presented.

8.2.1. Discussion Results

First, the results of the probabilistic analyses will be discussed followed by the results of the RFEM analyses.

Probabilistic Analyses

Similar to the semi-probabilistic analyses, it can be observed that the P_f decreases with decreasing values of the ratio between local and regional variance (α). The results in Table 7.5 show that the P_f for $\alpha = 1.0$ compared to $\alpha = 0.75$ decreases with 87.5%. Please note that these P_f are conditional probabilities of failure based on one water level and can therefore not be compared to the P_f calculated a semi-probabilistic calculation.

When comparing the first two moments (μ and σ) of the fitted CDF, the amount of mean and variance reduction can be quantified. From the relative change in σ_{CDF}, it can be seen that variance reduction increases with an increase in α, due to the decreasing standard deviation of the soil parameter input.

RFEM Results

The results of the RFEM analysis, compared to the probabilistic scenario 2c ($\alpha = 0.75$), show that the variance reduction decreases with an increase in θ_h (a decrease of 25.0% in σ_{CDF} for scenario 3a compared to 5% for 3b), whereas the mean reduction increases with an increase in θ_h (a decrease of 0.05% in μ_{CDF} for scenario 3a compared to 0.26% for 3b). This results in an increase in the probability of failure by 12.5% for $\theta_h = \infty$ while it decreases by 34.38% with $\theta_h = 20 * \theta_h$. These observations highlight the influence of the horizontal scale of fluctuation, which will be explored in more detail in section 8.2.3.

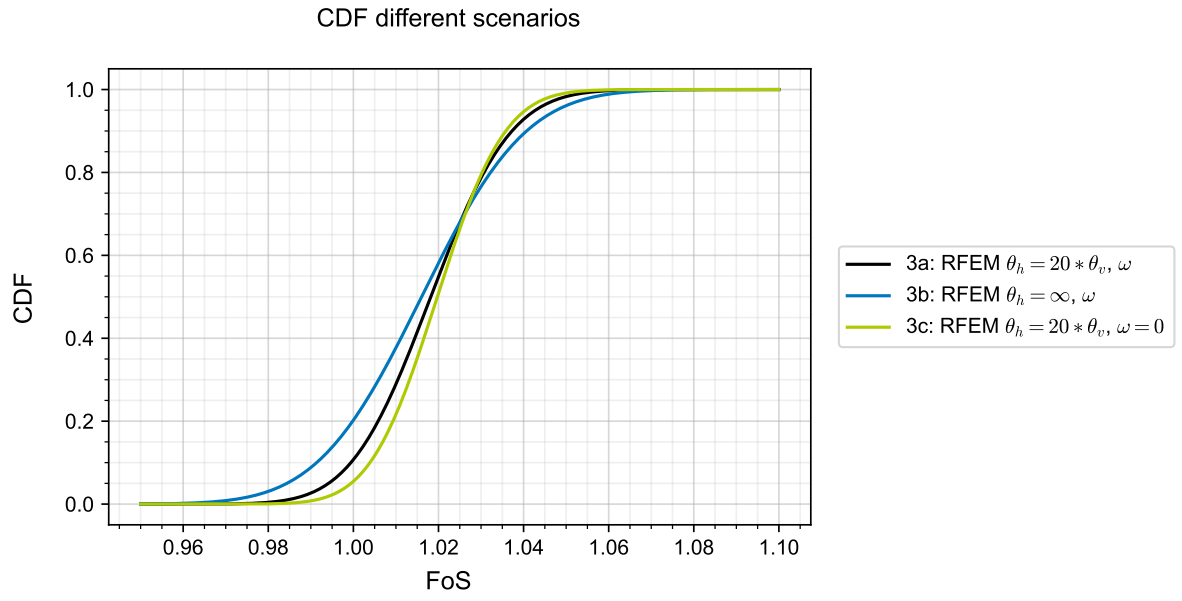
The current model assumes that the local variance, defined as the variance at the scale of a failure surface, is fully averaged out. However, based on the results presented in Figure 7.5 and Table 7.5, it is clear that the fit of the CDF in scenario 3a has a larger standard deviation as compared to scenario 2a. In scenario 2a, the value of α is 1.0, which implies that the input for the standard deviation in this analysis is only the result of statistical uncertainty, as explained in Chapter 2.1.5. Therefore, it should be considered as the minimum width of the distribution of the results. On the other hand, scenario 3a includes the reduction of local variance through the failure surface, having to pass both strong and weak zones in the random field. A larger standard deviation of the results of scenario 3a indicates that with the scales of fluctuation used in this case study, not all local variance can be averaged out.

As dikes are designed for a low probability of failure, the lower tail of the CDF is important. Out of all the possible scenarios with different values of α, the scenario where $\alpha = 0.8$ is closest to the first RFEM scenario in the lower tail, while using α equal to 0.75 is over-conservative with the given values of θ. However, the probabilistic scenario with α equal to 0.75 approximates the extreme RFEM scenario with $\theta_h = \infty$ most closely.

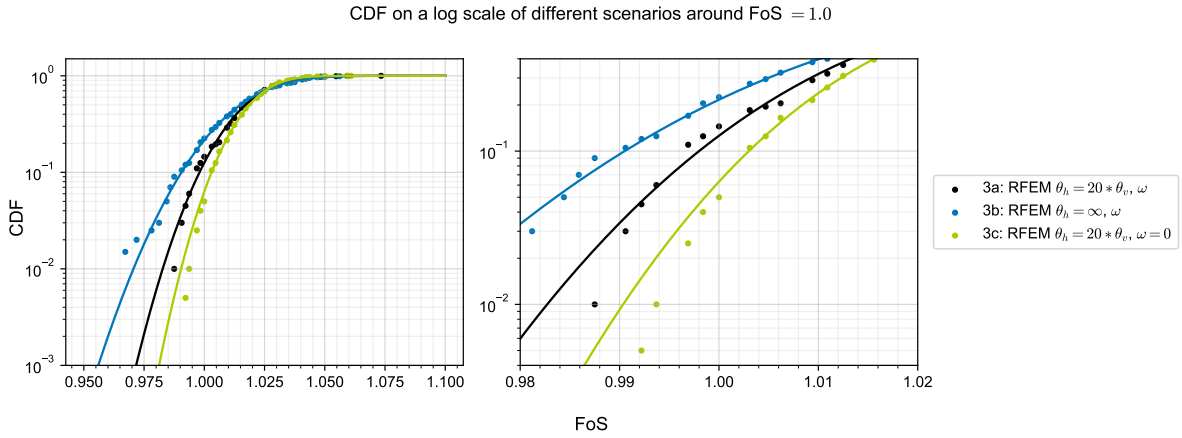
Additionally, it can be observed that in the lower tail, the mean reduction does not significantly impact the fitted probabilistic scenarios. However, it should be noted that entirely fitting the probabilistic function with only adjusting α is not possible due to the mean reduction.

8.2.2. Influence ω

As discussed in Chapters 2.3.2 and 6.2, an ω is introduced in the correlation function in RFEM when generating random fields. This is done to consider the statistical uncertainty described by van den Eijnden and Hicks (2019). When excluding this type of uncertainty, all of the variance can be averaged out when a small enough scale of fluctuation compared to the problem size is selected. van den Eijnden and Hicks (2019) found that even small amounts of systematic uncertainty can significantly impact the estimated probability of failure. This research tests this influence by adding a third RFEM scenario with $\omega = 0$ and comparing the results with the other RFEM results.



(a) CDF



(b) CDF on a logarithmic scale with the lower tail.

Figure 8.7: CDFs of the RFEM scenarios to investigate the influence of ω on the results.

Based on the results in Figure 8.7, it can be seen that applying $\omega = 0$ results in a narrower distribution of results, with $\sigma_{CDF} = 0.012$. This is due to the fact that a larger portion of the variance can be averaged out, leading to a larger amount of variance reduction. When $\omega = 0$ is adopted, the value of $P_f = 0.063$, which is 50% smaller compared to the P_f when ω is used. This highlights the importance of incorporating statistical uncertainty through ω in RFEM calculations.

In order to determine the value of ω for scenarios 3a and 3b, the ratio between the variance, determined with $\alpha = 1$ and $\alpha = 0$, is used, as shown in Equation 6.3. This determination of ω was based on the assumption that the regional variance is small compared to the total variance, as could be observed

from the results of the data analysis. However, it should be considered that the regional variance might not be negligible. This would lead to an increase in the value of ω and a wider CDF. Consequently, the calculated values of P_f with the current values of ω would be less conservative. To further investigate the impact of incorporating regional variance into the determination of ω on the distribution of the results, it is recommended to perform such an analysis as a part of future research.

8.2.3. Influence Horizontal Scale of Fluctuation

In Chapter 6.2.1, it was discussed that the horizontal scale of fluctuation (θ_h) could not be determined with the available data. To further investigate the influence of the horizontal scale of fluctuation, a sensitivity analysis was performed for three different ratios between θ_h and θ_v: 10, 15, and 20. The results are presented in Figure 8.8.

The generated CDFs do not differ significantly. This, combined with a Monte Carlo of only 200 realizations, makes fitting a theoretical function difficult, providing no further insights into the differences between the results. Therefore, it is not included in the figure. However, it can be observed that the standard deviation of the CDF increases a bit with an increasing θ_h-θ_v ratio. However, this is not significant. It was observed that a higher value of θ_h results in an small increase of the probability of failure, based on a slight difference in distribution.

As could be observed from Figure 7.5, more significant differences between the values of θ_h have a more substantial effect, as seen in Figure 7.5, where the $\theta_h = \infty$ results are modeled.

It is possible that the minor difference in distribution can be explained by θ_h values in the sensitivity analysis falling below a certain threshold value of θ. From this threshold value onwards, θ_h may have a greater impact on stability, similar to the influence of scenario 3b. This concept is related to "worst-case θ," as explained in Chapter 2.2.1 and defined by Fenton and Griffiths (2003). While Ali et al. (2014) has researched the "worst-case" value of θ_v in a spatially variable hydraulic conductivity in a dike, no study has yet been conducted to determine the "worst-case θ_h" for soil strength parameters. Therefore, it is recommended to investigate the value of θ_h at which the probability of failure would start to be more significantly affected.

It should be noted that these results have been generated with a smaller σ_{RFEM} than the other RFEM scenarios 3a and 3b. Therefore, the CDFs of the sensitivity analysis cannot be compared with the other RFEM CDFs.

Furthermore, the horizontal scales of fluctuation calculated for Dutch soil by de Gast (2020) were smaller than found in the literature overview in Table 6.11. He concluded that this was caused by the loading and deformation of the soil underneath the dike caused by its weight. Furthermore, he found that θ_x (direction perpendicular to the dike) is smaller than θ_y (direction parallel to the dike). The first of these two scales of fluctuation has been considered (θ_x) in this research.

In de Gast (2020), a ratio between θ_h and θ_v of 1 to 8 could be found for Dutch soils. This is lower than the assumed $\theta_h = 20 * \theta_v$ in this study. This means that the actual θ_h in this case study, and in general in the Netherlands, is probably closer to $\theta_h = 20 * \theta_v$ than $\theta_h = \infty$ for this case study.

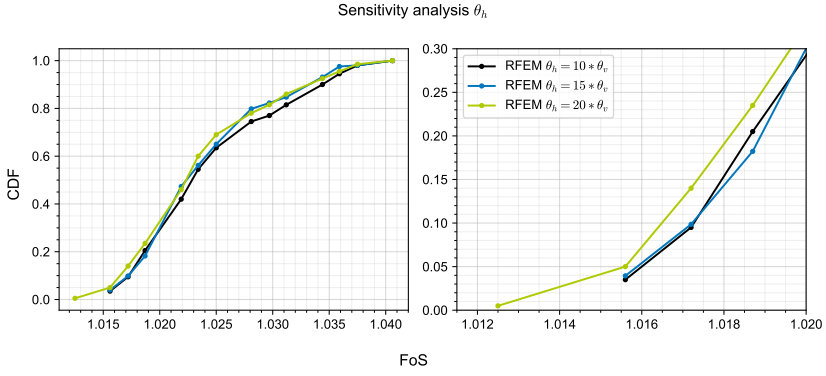


Figure 8.8: The results of a sensitivity analysis for θ_h: CDFs of different θ_v and θ_h.

From Figure 8.7, it can be observed that the differences between the distributions for scenarios 3a and 3c are greater than those for 3a and 3b. The difference between the distributions of 3a and 3b highlights the effect of θ_h on the P_f, which increases with 71.43%. The P_f decreases with 50% when the $\omega = 0$, demonstrating the impact of incorporating statistical uncertainty. From the relative changes in P_f, it can be seen that the influence of θ_h on the P_f is greater than the influence of ω.

8.2.4. Approximation of the Probability of Failure

As a Monte Carlo analysis only approximates the actual probability of failure, it is not 100% precise. The approximations are presented in Table 8.1. The more realizations are performed, the higher the accuracy of the approximation of P_f. Furthermore, the lower the calculated P_f, the larger the uncertainty about the value of P_f. This can be quantified by defining a standard deviation for P_f (Fenton & Griffiths, 2008):

$$\sigma_{P_f} = \sqrt{\frac{P_f(1-P_f)}{n}} \tag{8.1}$$

This can be scaled relative to the mean of P_f to form the coefficient of variation of P_f:

$$CoV_{P_f} = \sqrt{\frac{1-P_f}{nP_f}} \tag{8.2}$$

Table 8.1: The calculated probability of failure P_f and its CoV_{P_f} of the FEM results of the probabilistic and RFEM analyses.

Level	Case	P_f	CoV_{P_f}
2: Probabilistic	2a ($\alpha = 1.0$)	0.024	0.450
	2b ($\alpha = 0.8$)	0.117	0.194
	2c ($\alpha = 0.75$)	0.192	0.145
3: RFEM	3a ($\theta_h = 20 * \theta_v$, ω)	0.126	0.186
	3b ($\theta_h = \infty$, ω)	0.216	0.135
	3c ($\theta_h = 20 * \theta_v$, $\omega = 0$)	0.063	0.273

The calculated values of the CoV_{P_f} are presented in Table 7.5. As expected, the calculated CoV_{P_f} is higher for a lower P_f. For case 2a, σ_{P_f} is almost half of the calculated P_f, which is substantial. The accuracy of these calculated failure probabilities can, therefore, be doubted.

A CoV_{P_f} of maximally 0.1 is often preferred. The maximum P_f that has a $CoV_{P_f} < 1.0$ with 200 realizations can be calculated through Equation 8.2. This results in a value of 0.33 and is illustrated in Figure 8.9. This highlights the inaccuracy of the calculated P_f in combination with 200 realizations.

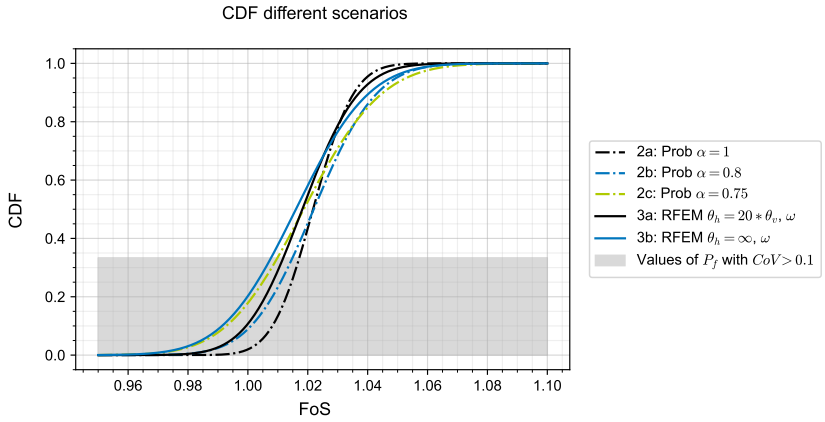


Figure 8.9: The distributions of probabilistic and RFEM Monte Carlo results. In grey, the probabilities of failure, which cannot be calculated with a $CoV_{P_f} < 0.1$, are highlighted.

To compare the results obtained from the probabilistic and RFEM calculations, a D-Stability FORM analysis that employs the same input as in scenario 2c can be used. By doing this, a (conditional) probability of failure of 0.0292 has been determined. This value is 84.79% lower than the probabilistic scenario 2c and lower than all other failure probabilities for both probabilistic and RFEM scenarios. It is important to note that the difference between the calculated probabilities of failure could be caused by the limited number of realizations or difference in the calculation methods, which is LEM in combination with FORM versus FEM in combination with Monte Carlo. Therefore, it is recommended to investigate this in further research to determine the applicability of probabilistic FEM can be used to assess Dutch dikes. Furthermore, it should be noted that no model factor is considered in the D-Stability analysis, to be able to compare the results.

The results of the probabilistic analysis cannot be compared to the results of other dike stability assessments in the Netherlands because, in this calculation, only the conditional probability of failure for one outside water level is computed. This gives an incomplete picture of the full P_f of a dike. Further research is recommended to provide a comprehensive understanding of the influence of spatial variability in probabilistic calculations.

8.2.5. Conclusions

The following conclusions were drawn from the probabilistic and RFEM analyses:

- An increase in α results in a decrease in P_f for probabilistic analyses. The difference in P_f compared to the current method with $\alpha = 0.75$ is 85.5%.
- Increasing α leads to an increase in variance reduction of up to 45% compared to the current method with $\alpha = 0.75$ for the probabilistic calculations.
- A probabilistic analysis with $\alpha = 0.8$ is a better fit to reality than with $\alpha = 0.75$, as concluded from the comparison with the probabilistic and RFEM results.
- With the used scales of fluctuation in RFEM, it can be concluded that with the used values of θ, not all local variance averages out because the realistic RFEM CDF (scenario 3a) has a width that is larger than the probabilistic scenario with $\alpha = 1.0$, which considers full local averaging.
- The probabilistic method does not consider mean reduction, whereas the RFEM results conclude that mean reduction does happen when strong and weak zones are modeled. The mean is reduced by 25% when comparing the current method (scenario 2c) with RFEM scenario 3a. However, in this case study, excluding the effect of mean reduction does not lead to a lower P_f.
- The statistical uncertainty should be accounted for in RFEM by incorporating an ω factor in the autocorrelation function. Excluding this could result in less conservative values of P_f.
- The used value of ω

- A small change in the value of θ_h in RFEM analyses does not significantly influence the results. However, when θ_h approaches infinity, the P_f increases. Underestimating the horizontal scale of fluctuation could, therefore, result in less conservative results of P_f.
- Based on a comparison with values from literature, it can be concluded that a ratio between θ_h and θ_v of 20 is more realistic for Dutch soils than $\theta_h = \infty$.
- In this case study, the influence of the horizontal scale of fluctuation on the P_f is larger than the incorporation of statistical uncertainty.
- When performing a Monte Carlo analysis to assess P_f, it is typically preferred to have a maximum coefficient of variation (CoV) of 0.1. However, if the analysis only includes 200 realizations, the lowest value of P_f that can be accurately determined with this maximum CoV is 0.33, which is much larger than most of the required failure probabilities in the Netherlands.
- It could be seen that the probabilistic scenario 2c results in a 84.79% higher (conditional) P_f than when using the same input for a FORM analysis in D-Stability. Further research is recommended to investigate this difference.
- The computed probabilities of failure in the probabilistic and RFEM analyses are conditional probabilities based on one outside water level only. They can, therefore, not be compared with the norm or the P_f calculated from a semi-probabilistic analysis.

Part III

Discussion and Conclusion

9 Discussion

This study on the impact of spatial variability was divided into two parts. Firstly, a data analysis was conducted to investigate national and regional spatial correlation. The analysis used data collected by STOWA, and variograms were constructed to quantify the spatial variability per soil type and soil parameter.

The second part involved conducting a Monte Carlo simulation in combination with random fields to investigate the response of a case study dike in a probabilistic manner (conditional to the outer water level). The assumptions of mean and variance reduction within the WBI2017 methodology were investigated by comparing this with the Random Finite Element Method (RFEM) model (which approximates reality better).

The findings of the individual parts were discussed separately in Chapters 5 and 8. As a result, this chapter will focus on discussing the limitations of the methodology and a discussion on the generalization of the results.

9.1. Limitations

It can be observed that the results of the data analysis do not highly influence the setup of the RFEM model. Some of the reasons for this are already discussed in Chapter 5. To highlight their influence on the RFEM input, a brief overview will be given:

- The wave shapes in the variograms, as discussed in Chapter 5.3.3, suggest that the data is not intrinsically stationary. As this is a variogram requirement, it is not appropriate to use a variogram to visualize spatial variation in this case. The wave-like shape makes it difficult to fit the theoretical variogram, affecting the accuracy of the conclusions about the spatial variability from the data analysis. This could have caused the other limitations of the data analysis that will be presented in the next points.
- As discussed in Chapter 5.3.2, the variograms could not accurately calculate the range, and hence no values for the horizontal and vertical scale of function for the RFEM model could be calculated. Therefore, assumptions on the horizontal scale of fluctuation needed to be made in the setup of the model. For the vertical scale of fluctuation (θ_v), CPTs had to be used to determine this value instead of laboratory tests.
- No useful approximation could be made in Chapter 5.3.4 for the value of α from the nugget-sill analysis. Therefore, the influence of an altered value of α could not be tested in RFEM.
- As discussed in Chapter 5.2, the variograms for the SHANSEP parameter S were highly discontinuous and with a lot of scatter. Therefore, no conclusions about the spatial correlation could be made from the data analysis. These conclusions, which could have been tested in the RFEM analysis, did, therefore, not give any additional insights.

Additionally, due to a limited amount of data and time constraints, only the spatial correlations of S

and γ_{wet} were investigated. If this analysis was also performed for the friction angle ϕ and SHANSEP parameter m, the difference in α for different parameters could have been investigated and modeled. In that case, the assumption of equal α values for different soil parameters could have been investigated. The findings in Chapter 8.1.4 suggest that m has an influence on the stability of a dike. It can be assumed that the incorporation of random fields for m could lead to larger mean and variance reduction due to the modeling of stronger and weaker zones for both m and S. This assumption could have been tested in combination with the insights from the data analysis into the spatial correlation structure. However, it should be noted that in this case study, the FEM code calculates an $OCR = 1$ throughout the entire dike cross-section. For m to have an influence in RFEM, it is recommended to test this in a cross-section where the OCR has a larger influence on the undrained shear strength.

Furthermore, the SHANSEP parameters are related to the soil's mineralogy. It can be hypothesized that spatial correlation will not impact the variance of S because the mineralogy is mainly related to the soil type. Yet, spatial correlation was observed in the constructed variograms. However, these variograms were highly discontinuous, and therefore the quality of these results can be questioned. If no correlation could be assumed for S, the correlation function in RFEM would be 0 for all distances, corresponding to full spatial averaging along the failure surface of the sample variance, reducing the width of the CDF of the results and creating the possibility for more economical designs. Therefore, it is recommended to further investigate the spatial correlation in S because there is room for optimization if this hypothesis turns out to be correct.

A limitation of using 2D RFEM is that it does not account for the influence of the horizontal spatial variability in the direction parallel to the dike. The current method used by WBI2017 accounts for this by following the model explained by Calle et al. (2021). The model assumes that true finite failure can only occur in a longitudinally unstable zone, where the average soil strength in the longitudinal direction along the failure surface is below the required stability limit. As a result, no averaging is considered in the horizontal direction since the scale of fluctuation in that direction is often larger than the width of the failure surface. However, Hicks and Spencer (2010) found that the ratio between θ_y and the slope length affects the probability of dike failure, which contradicts the assumption of the current Dutch assessment method. Therefore, it is recommended to investigate the influence of θ_h in the direction parallel to the dike on the probability of dike failure, specifically focusing on its relation to the current assessment method in the Netherlands.

9.2. Generalization

It is important to note that the study examined a single case study dike. However, there are two main points that need to be taken into consideration before generalizing the findings. Firstly, although the calculated shear stresses were equal in both the FEM and D-Stability calculations, the calculated vertical effective stresses and OCR were observed to be different. The differences cancel each other out in this particular case. However, it is possible that in other dike cross-sections the results could be different. Therefore, it is recommended to conduct further research about differences between D-Stability and FEM calculations before generalizing the findings of this study to other dike sections. Secondly, the found probabilistic fit with $\alpha = 0.8$ is only valid for cases where θ_h is less than or equal to 20 times θ_v. Even though it can be argued that the scenario where $\theta_h = 20 * \theta_v$ is a more realistic approximation of θ_h in the Netherlands, the results in the RFEM analysis for $\theta_h = \infty$ highlight the influence of θ_h on the probability of failure of a dike. Therefore, caution should be applied when doing this, and it is recommended that the value for θ is investigated for the specific dike section to investigate dissimilarity to the values used for this case study.

10

Conclusion

Spatial variability in soil has a significant impact on the failure mechanism macro-stability. However, the current approach for incorporating spatial variability into dike stability calculations has limitations. It assumes that variance reduction along the failure surface should be accounted for while it neglects the additional effect that the failure surface ”seeks out” the weak zones, leading to mean reduction. Additionally, a default value of the ratio between local and regional variance (α) is set to 0.75, which lacks empirical evidence. The impact of these assumptions on failure probability estimation is unknown, which highlights the need for a better understanding of the issue. Although the sophisticated approach called the Random Finite Element Method (RFEM) can address this issue, its computational complexity makes general implementation in dike stability assessments impractical. Therefore, the objective of this research is to create a methodology that accounts for the spatial variability of soil properties in dike stability assessments. This will be done by answering the following research question and sub-questions:

RQ What is an effective approach for incorporating spatial variability in soil into dike stability calculations?

- *SQ1 How do different soil characteristics correlate in the spatial domain?*
- *SQ2 What is the number and distribution of measurements on a local scale needed to accurately model the spatial variability in the soil?*
- *SQ3 How does soil spatial variability on a regional scale influence the different steps involved in the schematization of dike stability calculations?*
- *SQ4 How important is the spatial variability in the soil compared to other uncertainties?*

The investigation of spatial variability consisted of analyzing data and creating an RFEM model. The data analysis involved constructing variograms to quantify spatial variability per soil type and parameter. Answers to the first and second sub-research questions can be found in the results for the first part. The second part can answer the third and fourth sub-questions, where an RFEM model is created for a case study dike and compared to the current methodology to investigate the assumptions made. Monte Carlo simulations are used to calculate the probability of failure.

This chapter will answer the sub-research questions based on the conclusions drawn from the data analysis and RFEM modeling. The answer to the main research question will follow this. Finally, the chapter will end with a summary of recommendations.

10.1. Data Analysis

In this section, a summary of the results obtained through a geostatistical analysis focused on investigating the spatial variability is presented for two important parameters: γ_{wet} and SHANSEP 's S. This is done to provide answers to the first and second sub-research questions. To achieve this goal, variograms were created at a national level with data from the STOWA dataset, and various filters were

applied to explore the structures of spatial variability. Additionally, regional-scale variograms were constructed for γ_{wet} from a Rivierenland (RL) dataset to gain further insights. For an overview of all the conclusions that could be drawn from the data analysis, please refer to chapter 5.

SQ1: How do different soil characteristics correlate in the spatial domain?

The resulting variograms indicated that γ_{wet} had less spatial variability than S, with peat having a smaller variance than clay. Additionally, the variograms for S were highly discontinuous and scattered, possibly due to poor data quality or the parameter not being spatially correlated.

Furthermore, the study compared the ratio between variance due to spatial variability and the total variance for variograms carried out on both national and regional data. The findings showed that spatial variability decreases at a regional scale compared to a national scale. However, due to uninvestigated assumptions and the poor quality of the variograms for S, an appropriate value for α for the soil parameters could not be found through the construction of the variograms.

It was observed from the variograms of γ_{wet} that the ratio between local and regional variance is likely to be higher than 0.75, mainly if data is collected at a smaller scale than the maintenance area of a water board.

Lastly, the study showed that the calculated ranges of correlation in the horizontal direction were inaccurate, making the variogram on a national scale unreliable for determining the scale of fluctuation on a smaller scale.

SQ2: What is the number and distribution of measurements on a local scale needed to accurately model the spatial variability in the soil?

According to the data analysis results, it is impossible to use the STOWA dataset to determine local-scale spatial variability through the construction of a variogram. Although this study did not further investigate the reasons for this limitation, some potential causes were identified based on existing literature. Considering these factors when designing a soil investigation plan to study local spatial variability is also important. Therefore, this section gives an outline of the key points.

Small-scale variance cannot be analyzed with variograms based on the national STOWA dataset because spatial variability falls in the first bin and is averaged with other uncertainties. Additionally, the first bin in all the constructed variograms was much larger than the definition of ”local”, as described in WBI2017, which is at the level of a failure mechanism. Investigating local variability requires local data with a high enough density and accuracy in the research area.

The accuracy of the variogram also depends on the number, distribution and quality of data points:

- Around 150 data points are required to create a reliable variogram.
- Each bin in the variogram should contain at least 100 point pairs. If the histogram shows a decrease below 100 points per bin, the variogram could be cut off at that point.
- The reliability of a variogram also depends on the density and distribution patterns of data. Areas with fewer data points are less reliable. At least five points are required within the variogram range to study the spatial variability in soil parameters.
- Outliers should be filtered before performing geostatistical analyses to estimate the spatial variability.
- When using variograms to define spatial variability, it is important to consider that only intrinsically stationary data should be used. This means that the statistical properties of the data, such as its mean and variance, remain constant in space.

10.2. Modeling

The soil parameter input for part two can be calculated from the national STOWA dataset used in the data analysis. Part two involves using Random Finite Element Method (RFEM) for a case study dike located within the maintenance area of water board Brabantse Delta. The RFEM model simulates the soil’s spatial variability by modeling strong and weak zones in the soil through Monte Carlo simulations

with random fields. The results obtained from the RFEM model are then compared with the results of probabilistic analyses with FEM to answer the main research question.

The study conducted multiple analyses to investigate the impact of variance and mean reduction. Various (semi-)probabilistic scenarios were created based on different input values for α. The first two RFEM scenarios generated random fields with two different values for the horizontal scale of fluctuation (θ_h). The third RFEM scenario was created to evaluate the results without accounting for statistical uncertainty. A summary of the factors of safety (FoS) and (conditional) probabilities of failure (P_f) is presented in Table 10.1.

Table 10.1: An overview of the calculated probabilities of failure (P_f) and factors of safety (FoS) from the model results. *The P_f has been obtained via a different method for deterministic and semi-probabilistic analyses compared to the probabilistic and RFEM analyses. For the first two levels, P_f is calculated via the calibration formula defined in the WBI2017. For the probabilistic and RFEM analyses, the conditional P_f is determined at FoS $= 1.0$ at the CDF from the results. The first method also considers the model factor, while the second does not. Because of the differences in the methods and the fact that for the last two methods, P_f is conditional, the absolute values of P_f should not be compared.

Level	Scenario	FoS	P_f*
Deterministic	0	1.025	$1.023 * 10^{-5}$
Semi-probabilistic	1a ($\alpha = 1.0$)	0.997	$2.022 * 10^{-5}$
	1b ($\alpha = 0.8$)	0.984	$2.748 * 10^{-5}$
	1c ($\alpha = 0.75$)	0.980	$3.010 * 10^{-5}$
	1d ($\alpha = 0.5$)	0.961	$4.653 * 10^{-5}$
Probabilistic	2a ($\alpha = 1.0$)	-	0.024
	2b ($\alpha = 0.8$)	-	0.117
	2c ($\alpha = 0.75$)	-	0.192
RFEM	3a ($\theta_h = 20 * \theta_v$, ω)	-	0.126
	3b ($\theta_h = \infty$, ω)	-	0.216
	3c ($\theta_h = 20 * \theta_v$, $\omega = 0$)	-	0.063

SQ3: How does soil spatial variability on a regional scale influence the different steps involved in the schematization of dike stability calculations?

The current WBI2017 method considers spatial variability by only including the regional variance and averaging the local variance entirely (Calle et al., 2021). Increasing the value of the ratio between local and regional variance (α) allows a larger portion of the sample variance to be averaged out before using it as input for the calculation. As the dike assessments in the Netherlands can be performed through two reliability levels, the conclusions will be presented separately.

This study found that this results in a higher Factor of Safety (FoS) in semi-probabilistic calculations. The probability of failure (P_f) can be calculated from the Factor of Safety (FoS) via the calibration formula. It was found that an increase in α from 0.75 to 0.8 resulted in a decrease of 32.96%. From a comparison with a reference D-Stability calculation, it can be concluded that Finite Element Method (FEM) can only be used to assess the stability of a Dutch dike when adopting the vertical effective stress in the calculation of the undrained shear strength via SHANSEP. Because FEM often uses the major principal stress in the calculation, which is different than the definition of SHANSEP in WBI2017. This difference resulted in 9% higher FoS when using the major principal effective stress in this case study.

Monte Carlo simulations were used with homogeneous soil layers in FEM to calculate the conditional probability of failure to one outside water level for probabilistic calculations. This results in a Cumulative Density Function (CDF) of calculated factors of safety. The comparison of the different CDFs shows that a higher α gives a smaller Cumulative Density Function (CDF) of the results without reducing the mean. For this case study, implementing a higher value of α results in a maximum reduction of 87.5% in conditional P_f compared to the current value of 0.75. The probabilistic results should only be compared with the other FEM results because the calculated probability of failure differs significantly from a probabilistic calculation in D-Stability and because it is a conditional probability, not a total probability

of failure of a cross-section.

SQ4: How important is the spatial variability in the soil compared to other uncertainties?
The research investigated the impact of four uncertainties on the resistance against dike instability:

1. A sensitivity analysis was conducted to test the assumption that the SHANSEP parameter m would not significantly impact the model.
2. The location of the boundary of the peat layer was studied as the failure surface often propagated along it in the RFEM results.
3. Statistical uncertainty was incorporated in both the current WBI2017 method and RFEM to compare their outcomes. Afterwards, its influence on the RFEM results is investigated.
4. The horizontal scale of fluctuation could not be determined from the data analysis. Therefore, different values were assumed for the model setup. A sensitivity analysis was performed to evaluate the impact of this assumption on the P_f.

Because the first two factors are investigated through semi-probabilistic calculations, while the last two are investigated through RFEM calculations, they will be presented separately.

Value of m and the Location of the Peat Layer

Through separate sensitivity analyses for different values of m and different locations for the bottom of the peat layer, it can be concluded that both influence the FoS. When altering the value of m within the 95% confidence interval of its distribution, the difference between FoS is 0.118. When changing the location of the top of the peat layer with $3\mathrm{m}$, the difference between FoS is 0.019. Therefore, it can be concluded that for this case study, the value of m has a larger impact on the strength of the dike than the location of the peat layer.

Statistical Uncertainty and the Horizontal Scale of Fluctuation

In the RFEM analyses in this research, statistical uncertainty is included through a residual correlation factor (ω). The purpose of defining ω is to consider part of the soil parameter variance, which is the variance due to statistical uncertainty, that is not susceptible to averaging. This results in a wider CDF with a reduced mean, leading to a 50% increase in conditional P_f. The sensitivity analysis on the horizontal scale of fluctuation (θ_h) showed that a small variation of this value has a minor impact on the resulting distribution. However, when θ_h is increased to infinity, indicating that the soil strength only fluctuates vertically, it leads to a larger variance and mean reduction, which causes an increase of the conditional P_f by 71.43%.

Therefore, it can be concluded that including statistical uncertainty and accurately determining the horizontal scale of fluctuation are critical when conducting RFEM analyses. Excluding these factors can result in less safe outcomes. When comparing the impact of these two factors, it can be concluded that the vertical scale of fluctuation significantly influences the conditional probability of failure in this case study.

10.3. Answer to the Research Question

The impact of the assumptions in the WBI2017 model and the answer to the research question can be found by comparing the probabilistic analysis and RFEM analysis results. They can only be compared if the RFEM calculation uses the vertical effective stress in calculating the undrained shear strength and considers the statistical uncertainty through incorporating a residual correlation factor in the creation of the random fields.

The failure surfaces in RFEM did not show a clear pattern, but the mean reduction caused by the propagation of the failure surface through weak zones was evident in the resulting CDF. The horizontal scale of fluctuation significantly influences the amount of mean reduction. Investigating its influence is important because it is difficult to determine this value in practice, as seen in this study's data analysis. Furthermore, the RFEM scenario does not show full spatial averaging of local variance as the CDF is wider than a probabilistic calculation with $\alpha = 1.0$, where full spatial averaging is accounted for in the input parameters. This leads to a higher probability of failure by 88.9% than a probabilistic analysis

with $\alpha = 1.0$. Therefore, assuming complete averaging of local spatial variance without considering regional variance underestimates the probability of failure.

In a comparison between the RFEM and probabilistic calculations, it was found that using a value of $\alpha = 0.8$ in the probabilistic calculations closely matches the RFEM scenario (where $\theta_h = 20 * \theta_v$) in the lower tail. Therefore, it is concluded that probabilistic calculation with $\alpha = 0.8$ is a better fit to reality than $\alpha = 0.75$. Moreover, it was observed that the mean reduction does not significantly impact the fitted probabilistic scenarios for determining the probability of failure. The results of the probabilistic scenarios showed that increasing the α value to 0.8 leads to a decrease in the probability of failure (P_f) by 39%.

To investigate the importance of this finding, the effect of incorporating $\alpha = 0.8$ in the WBI2017 method of assessing dike stability was studied. The outcome of the study revealed that the probability of failure for dike trajectory 34-2 was reduced by 39.72% (see Table 10.3). However, the (macro-stability) safety category of the dike trajectory remains unchanged.

Table 10.3: An overview of the influence of adopting another value of α on the probability of failure of dike section 20 in trajectory 34-2 when applying a decrease in $P_{f,cs}$ for all cross-sections in the trajectory.

Case	$P_{f,cs}$	$P_{f,sec}$	Category section	$P_{f,traj}$	Category trajectory
Original	$3.02 * 10^{-4}$	$5.4 * 10^{-4}$	$\mathrm{V_v}$	$9.87 * 10^{-3}$	D
Decrease of $P_{f,cs}$ in all sections	$1.81 * 10^{-4}$	$3.2 * 10^{-4}$	$\mathrm{V_v}$	$5.95 * 10^{-3}$	D

To conclude, after investigating the influence of variance and mean reduction by approximating the reality with RFEM, this research has shown that, for the case study, a probabilistic calculation with a value for the ratio between local and regional variance (α) of 0.8 better approximates reality, particularly in the lower tail of the distribution of the results, compared to the default value of $\alpha = 0.75$. Since dikes are designed for a low probability of failure, a correct fit in the lower tail is important. When implementing a value of $\alpha = 0.8$ instead of 0.75, following the WBI2017 guidelines, the probability of failure for this cross-section and dike trajectory will decrease by approximately 40%. Because the (macro-stability) safety category of the dike trajectory remains unchanged, it can be concluded that adopting $\alpha = 0.8$ in Dutch dike stability assessments does not significantly influence the safety of a dike trajectory. The answer to the question ***What is a practical approach for incorporating spatial variability in soil into dike stability calculations?*** is, therefore, that when considering the computational requirements of RFEM, it is more effective to keep using the WBI2017 approach of implementing spatial variability into the input parameters of (semi-)probabilistic calculations with $\alpha = 0.75$.

10.4. Recommendations

In this research, two procedures for integrating spatial variability into dike stability calculations have been compared. The first method is currently being used for dike stability assessments in the Netherlands, according to the guidelines of WBI2017. The second method, RFEM, models both strong and weak zones and will be used as a model for reality. This study has shown that both methods involve multiple assumptions. To further investigate these assumptions, some recommendations will be presented for future research. As the previous chapters have discussed most of these recommendations, this paragraph will present only a brief overview. Finally, one additional recommendation is highlighted.

The following topics are recommended for investigation in future research:

- The parameter S of SHANSEP exhibits highly discontinuous and scattered variograms, making it challenging to identify spatial correlations. Therefore, it is important to investigate the root cause of this issue. Although no spatial correlations were observed nationally, it is advisable to analyze the spatial correlations locally with high-density and quality data. This approach will help to gain a better understanding of the underlying factors contributing to the observed behavior of the parameter S. Furthermore, it is hypothesized that S has no spatial correlation because it is

only related to the mineralogy of the soil. Investigating this hypothesis could potentially lead to more cost-effective designs.

- Since different soil parameters have different spatial correlation patterns, it is advisable to investigate whether different values of α can be used for each parameter.
- It is not recommended to use the national STOWA database to determine local-scale spatial variability. Instead, a (multi)fractal analysis could be more useful to determine local variability from larger-scale databases. Further research on this is recommended.
- The calculation of undrained shear strength is different for D-Stability and FEM because they use a different stress component, resulting in a 9% difference in FoS. The effect of this difference should be studied for other cross-sections as well.
- A sensitivity analysis of the SHANSEP m parameter showed a sudden increase in the calculated FoS due to a change in the failure surface shape. An additional sensitivity analysis with more values for m is recommended to better quantify the impact of m on the FoS. Furthermore, it is recommended to investigate the influence of random fields for m in an RFEM analysis.
- The sensitivity analysis of the location of the peat layer revealed that the calculated failure surfaces in FEM follow the top boundary of the peat layer, which could not be modeled in D-Stability. It is assumed that this is caused by the limited shape of the Uplift-Van model. This assumption should be further investigated, and the results should be compared with models that allow more freedom in determining the failure surface, such as the Spencer model.
- The influence of a weaker layer of peat should also be studied since it is expected to significantly impact the FoS.
- Possible methods for incorporating mean reduction in the current probabilistic method model should be investigated since the RFEM model showed a mean reduction in the CDF of the results.
- The difference between the calculated probability of failure from the D-Stability FORM calculations and the Monte Carlo FEM analysis is quite significant. Further investigation is recommended to determine the cause of these differences.
- The impact of the change in α from 0.75 to 0.8 was tested in semi-probabilistic and FORM calculations in D-Stability. The FORM calculation considered only a single water level. This resulted in an incomplete representation of the probabilistic P_f. Therefore, conducting a full probabilistic calculation is recommended to thoroughly assess the impact of changing α. This will also help with determining if Monte Carlo in combination with FEM is suitable for dike stability assessments in the Netherlands.
- The RFEM models used values of vertical scales of fluctuation that were higher than the scales of fluctuation found in the CPT analysis due to computational constraints. An RFEM analysis with the true values should be performed to quantify its influence on the probability of failure.
- The current research incorporates only statistical variance in determining the value of the residual correlation factor (ω). This is based on the assumption that regional variance is relatively small compared to total variance. However, the WBI2017 method also considers regional variance, which cannot be averaged out. Therefore, exploring the possibility of including regional variance in determining ω in future research is important.
- The sensitivity analysis of the horizontal scale of fluctuation (θ_h) showed that only a large change in this value affects the probability of failure. Further investigation is necessary to determine the value of θ_h at which the probability of failure would start to be significantly affected. Additionally, spatial variability in 3D should be further investigated, including the direction parallel to the dike.
- As the Monte Carlo analyses were performed with a small number of realizations (200), the accuracy of approximating the conditional P_f from these results is limited. Therefore, further investigation is needed to determine how the results would change when the number of realizations increases to achieve a CoV of P_f smaller than 0.1.
- It is important to note that the study is carried out on a single dike section focused on one water level and specific values for the scales of fluctuations. However, to obtain a comprehensive

understanding of the impact of spatial variability in dike assessments, it is highly recommended to extend the research to multiple dike sections across the Netherlands.

Furthermore, the term ”local” has been widely used in this research. In the context of WBI2017, it refers to the scale of a failure surface. However, the official documents such as Rijkswaterstaat (2021) and the theoretical background in Calle et al. (2021) do not clearly define this term. As a result, ”local” can be interpreted differently, leading to a possible misunderstanding that it refers to, for example, the scale of a dike section. Such misinterpretation may result in assuming that all variance at this scale can be averaged out, which is incorrect in many dike sections. To avoid such misunderstandings, it is recommended to clearly state the definition of ”local” in the guidelines.

References

Ajmera, B., Tiwari, B., & Pantha, K. (2016). Mineralogical effect on idealized, normalized strength curves for over-consolidated clays [Publisher: American Society of Civil Engineers], 1532–1539. https://doi.org/10.1061/9780784479742.128

Ali, A., Huang, J., Lyamin, A. V., Sloan, S. W., Griffiths, D. V., Cassidy, M. J., & Li, J. H. (2014). Simplified quantitative risk assessment of rainfall-induced landslides modelled by infinite slopes [ISBN: 0013-7952 Publisher: Elsevier]. *Engineering Geology*, *179*, 102–116.

Allard, D. (2013). J.-p. chilès, p. delfiner: Geostatistics: Modeling spatial uncertainty: 2nd edition. wiley, 2012. *Mathematical Geosciences*, *45*(3), 377–380. https://doi.org/10.1007/s11004-012-9429-y

Arel, E., & Mert, A. C. (2021). Field simulation of settlement analysis for shallow foundation using cone penetration data. *Probabilistic Engineering Mechanics*, *66*, 103169. https://doi.org/10.1016/j.probengmech.2021.103169

Baecher, G., & Christian, J. (2003). *Reliability and statistics in geotechnical engineering*. J. Wiley.

Bishop, A. W. (1955). The use of the slip circle in the stability analysis of slopes [Publisher: ICE Publishing]. *Géotechnique*, *5*(1), 7–17. https://doi.org/10.1680/geot.1955.5.1.7

BOI. (2023, July). *Geotechnische instabiliteit dijken/dammen deel 1 - handleiding overstromingskansanalyse*. Rijkswaterstaat.

Burgess, T. M., & Webster, R. (1980). OPTIMAL INTERPOLATION AND ISARITHMIC MAPPING OF SOIL PROPERTIES: I THE SEMI-VARIOGRAM AND PUNCTUAL KRIGING. *Journal of Soil Science*, *31*(2), 315–331. https://doi.org/10.1111/j.1365-2389.1980.tb02084.x

Calle, E. (2007). Statistiek bij regionale proevenverzamelingen: Het ruimtelijk statistische model. *Geotechniek*, *11*(3), 40–45.

Calle, E. (2008). Statistiek bij regionale proevenverzamelingen. *Geotechniek*, *12*(1), 40–44.

Calle, E., Kanning, W., & Schweckendiek, T. (2021, June 29). Characteristic values of soil properties in dutch codes of practice theoretical backgrounds and assumptions.

Calle, E., Vrouwenvelder, A., Lindenberg, J., Hannink, G., & Bruijn, E. (2008). Representatieve waarden voor grondparameters in de geotechniek. *Geotechniek*, *12*(2), 24–29.

Cami, B., Javankhoshdel, S., Phoon, K.-K., & Ching, J. (2020). Scale of fluctuation for spatially varying soils: Estimation methods and values. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*, *6*(4), 03120002. https://doi.org/10.1061/AJRUA6.0001083

CEN. (2004). Eurocode 7 geotechnical design. part 2 : Ground investigation and testing, EN 1997-2.

Chen, X. Y., Zhang, L. L., Zhang, L. M., Yang, H. Q., Liu, Z. Q., Lacasse, S., Li, J. H., & Cao, Z. J. (2020). Investigation of impact of submarine landslide on pipelines with large deformation analysis considering spatially varied soil. *Ocean Engineering*, *216*, 107684. https://doi.org/10.1016/j.oceaneng.2020.107684

Ching, J., Phoon, K., Beck, J. L., & Huang, Y. (2017). Identifiability of geotechnical site-specific trend functions [ISBN: 2376-7642 Publisher: American Society of Civil Engineers]. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*, *3*(4), 04017021.

Christodoulou, P., Pantelidis, L., & Gravanis, E. (2021). A comparative assessment of the methods-of-moments for estimating the correlation length of one-dimensional random fields. *Archives of Computational Methods in Engineering*, *28*(3), 1163–1181. https://doi.org/10.1007/s11831-020-09408-2

Clark, I. (2010). Statistics or geostatistics? sampling error or nugget effect? *Fourth World Conference on Sampling and Blending*, *110*, 13–18.

Clayton, C. R. I. (2001). *Managing geotechnical risk: Improving productivity in UK building and construction*. Institution of Civil Engineers; Thomas Telford.

Cressie, N., & Hawkins, D. M. (1980). Robust estimation of the variogram: I. *Journal of the International Association for Mathematical Geology*, *12*(2), 115–125. https://doi.org/10.1007/BF01035243

Das, B. M. (2019). *Advanced soil mechanics*. CRC press.

De Gast, T., Vardon, P., & Hicks, M. (2019). Observations and considerations regarding estimating horizontal scales of fluctuation around linear infrastructure, 340–345. https://doi.org/10.3850/978-981-11-2725-0IS12-3-cd

de Gast, T. (2020, January). *Dykes and embankments: A geostatistical analysis of soft terrain* (Doctoral dissertation). Delft University of Technology. https://doi.org/10.4233/UUID:4CE3B4EC-0A6A-4886-9A82-5945A1F9EA50

Deltacommissie. (1961). Rapport deltacommissie. deel 1. eindverslag en interimadviezen.

Deltacommissie. (2008). *Samen werken met water: Een land dat leeft, bouwt aan ziin toekomst* (September 2008).

Diermanse, F. (2021, May 11). Assemblageprotocol WBI2017 nadere uitwerking van het beoogde assemblageprotocol voor het wettelijke beoordelingsinstrumentarium.

Dowd, P. A. (1984). The variogram and kriging: Robust and resistant estimators. In G. Verly, M. David, A. G. Journel, & A. Marechal (Eds.), *Geostatistics for natural resources characterization: Part 1* (pp. 91–106). Springer Netherlands. https://doi.org/10.1007/978-94-009-3699-7_6

ENW. (2016). *Grondslagen voor hoogwaterbescherming* (2nd ed.) [OCLC: 1023024002]. ministerie van Infrastructuur en Milieu.

Fenton, G., & Griffiths, D. (2003). Bearing-capacity prediction of spatially random c φ soils [ISBN: 0008-3674 Publisher: NRC Research Press Ottawa, Canada]. *Canadian geotechnical journal*, *40*(1), 54–65.

Fenton, G., & Griffiths, D. (2008, August 19). *Risk assessment in geotechnical engineering* (1st ed.). Wiley. https://doi.org/10.1002/9780470284704

Fenton, G., & Vanmarcke, E. (1990). Simulation of random fields via local average subdivision [Publisher: American Society of Civil Engineers]. *Journal of Engineering Mechanics*, *116*(8), 1733–1749. https://doi.org/10.1061/(ASCE)0733-9399(1990)116:8(1733)

Genton, M. G. (1998). Highly robust variogram estimation. *Mathematical Geology*, *30*(2), 213–221. https://doi.org/10.1023/A:1021728614555

Griffiths, D. V., & Lane, P. A. (1999). Slope stability analysis by finite elements [Publisher: Thomas Telford Ltd]. *Géotechnique*. https://doi.org/10.1680/geot.1999.49.3.387

Handcock, M. S., & Stein, M. L. (2022). A bayesian analysis of kriging. *35*(4).

Hicks, M. A., & Spencer, W. A. (2010). Influence of heterogeneity on the reliability and failure of a long 3d slope. *Computers and Geotechnics*, *37*(7), 948–955. https://doi.org/10.1016/j.compgeo.2010.08.001

Hicks, M. (2012). An explanation of characteristic values of soil properties in eurocode 7. In *Modern geotechnical design codes of practice: Implementation, application and development* (pp. 36–45). IOS.

Hicks, M., & Li, Y. (2018). Influence of length effect on embankment slope reliability in 3d (D. Felix, Ed.). *International Journal for Numerical and Analytical Methods in Geomechanics*, *42*(7), 891–915. https://doi.org/10.1002/nag.2766

Hicks, M., Varkey, D., Van Den Eijnden, A., De Gast, T., & Vardon, P. (2019). On characteristic values and the reliability-based assessment of dykes. *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, *13*(4), 313–319. https://doi.org/10.1080/17499518.2019.1652918

Hijma, M., & Sun Lam, K. (2015, September). *Globale stochastische ondergrondschematisatie (WTI-SOS) voor de primaire waterkeringen*. Deltares.

Jongejan, R. (2017, June 24). *WBI2017 code calibration*. Rijkswaterstaat.

Jonkman, S. N., Kok, M., & Vrijling, J. (2008). Flood risk assessment in the netherlands: A case study for dike ring south holland. *Risk Analysis*, *28*(5), 1357–1374. https://doi.org/10.1111/j.1539-6924.2008.01103.x

Jonkman, S. N., & Schweckendiek, T. (2015). Briefing: Lessons learned from failures of flood defences. *Proceedings of the Institution of Civil Engineers - Forensic Engineering*, *168*(2), 85–88. https://doi.org/10.1680/feng.15.00004

Jonkman, S., Jorissen, R., Schweckendiek, T., & van den Bos, J. (2021, April). *Flood defences lecture notes CIE5314* (4th ed.). TU Delft.

Jonkman, S., Steenbergen, R., Morales-Nápoles, O., Vrouwenvelder, A., & Vrijling, J. (2015, November). *Probabilistic design: Risk and reliability analysis in civil engineering: Lecture notes CIE4130*. TU Delft.

Kanning, W., Teixeira, A., Van Der Krogt, M. G., & Rippi, K. (2017, April). Derivation of the semiprobabilistic safety assessment rule for inner slope stability.

Kindermann, P., & Tigchelaar, J. (2022, November). Landelijke grondparameterset. afleiden van kentallen voor grondsterkte.

Krige, D. G. (1951). A statistical approach to some basic mine valuation problems on the witwatersrand [ISBN: 0038-223X Publisher: Southern African Institute of Mining and Metallurgy]. *Journal of the Southern African Institute of Mining and Metallurgy*, *52*(6), 119–139.

Ladd, C. C., & Foott, R. (1974). New design procedure for stability of soft clays [Publisher: American Society of Civil Engineers]. *Journal of the Geotechnical Engineering Division*, *100*(7), 763–786. https://doi.org/10.1061/AJGEB6.0000066

Loquin, K., & Dubois, D. (2010). Kriging and epistemic uncertainty: A critical discussion. In R. Jeansoulin, O. Papini, H. Prade, & S. Schockaert (Eds.), *Methods for handling imperfect spatial information* (pp. 269–305). Springer. https://doi.org/10.1007/978-3-642-14755-5_11

Marchant, B. P., & Lark, R. M. (2004). Estimating variogram uncertainty. *Mathematical Geology*, *36*(8), 867–898. https://doi.org/10.1023/B:MATG.0000048797.08986.a7

Matsui, T., & San, K.-C. (1992). Finite element slope stability analysis by shear strength reduction technique. *Soils and Foundations*, *32*(1), 59–70. https://doi.org/10.3208/sandf1972.32.59

Oliver, M. A., & Webster, R. (2014). A tutorial guide to geostatistics: Computing and modelling variograms and kriging. *CATENA*, *113*, 56–69. https://doi.org/10.1016/j.catena.2013.09.006

Oliver, M. A., & Webster, R. (2015). *Basic steps in geostatistics: The variogram and kriging*. Springer.

Pan, Y., Liu, Y., Xiao, H., Lee, F. H., & Phoon, K. K. (2018). Effect of spatial variability on short- and long-term behaviour of axially-loaded cement-admixed marine clay column. *Computers and Geotechnics*, *94*, 150–168. https://doi.org/10.1016/j.compgeo.2017.09.006

Phoon, K., Nadim, F., Uzielli, M., & Lacasse, S. (2006, November 7). Soil variability analysis for geotechnical practice. In K. Phoon, D. Hight, S. Leroueil, & T. Tan (Eds.), *Characterisation and engineering properties of natural soils*. Taylor & Francis. https://doi.org/10.1201/NOE0415426916.ch3

POVM. (2020, March). POVM eindige-elementenmethode.

Rijkswaterstaat. (2021, May 28). Schematiseringshandleiding macrostabiliteit WBI 2017.

Rohmer, J., & Verdel, T. (2014). Joint exploration of regional importance of possibilistic and probabilistic uncertainty in stability analysis. *Computers and Geotechnics*, *61*, 308–315. https://doi.org/10.1016/j.compgeo.2014.05.015

Sadrekarimi, A. (2016). Static liquefaction analysis considering principal stress directions and anisotropy. *Geotechnical and Geological Engineering*, *34*(4), 1135–1154. https://doi.org/10.1007/s10706-016-0033-7

Schofield, A., & Wroth, P. (1968). Critical state soil mechanics. Retrieved November 29, 2023, from https://www.semanticscholar.org/paper/Critical-State-Soil-Mechanics-Schofield-Wroth/dd4f92e451984b65e71dd83bbf87e8e9c1fa1c0b

Schweckendiek, T., van der Krogt, M., Rijneveld, B., & Martins Teixeira, A. (2017, October 3). Handreiking faalkansanalyse macrostabiliteit.

Schweiger, H. F., & Peschl, G. M. (2005). Reliability analysis in geotechnics with the random set finite element method. *Computers and Geotechnics*, *32*(6), 422–435. https://doi.org/10.1016/j.compgeo.2005.07.002

Sharp, M., Wallis, M., Deniaud, F., Hersch-Burdick, R., Tourment, R., Matheu, E., Seda-Sanabria, Y., Wersching, S., Veylon, G., Durand, E., Smith, P., Forbis, J., Spliethoff, C., Hemert, H., Igigabel, M., Pohl, R., Royet, P., & Simm, J. (2013, January 1). *The international levee handbook*. CIRIA.

Shi, C., Wang, Y., & Kamchoom, V. (2023). Data-driven multi-stage sampling strategy for a three-dimensional geological domain using weighted centroidal voronoi tessellation and IC-XGBoost3d. *Engineering Geology*, *325*, 107301. https://doi.org/10.1016/j.enggeo.2023.107301

Smith, I. M., Griffi, D. V., & Margetts, L. (2014). *Programming the finite element method* [Pages: 1 online resource]. Wiley,

Spencer, E. (1967). A method of analysis of the stability of embankments assuming parallel inter-slice forces [Publisher: ICE Publishing]. *Géotechnique*, *17*(1), 11–26. https://doi.org/10.1680/geot.1967.17.1.11

Stijnen, J., Kanning, W., Jonkman, S., & Kok, M. (2014). The technical and financial sustainability of the dutch polder approach: Technical and financial sustainability of the dutch polder approach. *Journal of Flood Risk Management*, *7*(1), 3–15. https://doi.org/10.1111/jfr3.12022

STOWA. (n.d.). STOWA database. https://www.helpdeskwater.nl/onderwerpen/waterveiligheid/primaire/beoordelen/vragen/macrostabiliteit/macrostabiliteit/landelijke-database/

't Hart, R., De Bruijn, H., & de Vries, G. (2016, February). Fenomenologische beschrijving faalmechanismen WTI.

TAUW. (2021, November 19). *Analyse monitoringsplan willemstad noordschans (voorverkenning)* (R001-1281337XRV-V05).

TAW (Ed.). (1989). *Leidraad voor het ontwerpen van rivierdijken. d. 2: Benedenrivierengebied*. TAW.

TAW. (1998, January). *Grondslagen voor waterkeren* [OCLC: 68126170]. Technische Adviescommissie voor de Waterkeringen.

Teixeira, A., & Wojciechowska, K. (2023, February 21). Vervolgprocessen na eerste afschuiving.

Tigchelaar, J. (2018, October). *Proevenverzameling - sterktegegevens ondergrond* (PR3478.10). HKV lijn in water.

Tigchelaar, J. (2022, July). *Voorverkenning dijktraject 34-2 - (probabilistische) analyse STBI vak 27* (PR4495.10). HKV lijn in water.

Tigchelaar, J., & van Haaren, D. (2022, May 25). *Memo - macrostabiliteit WiNo – fase 1* (PR4495.10). HKV lijn in water.

TNO-NITG. (n.d.). GeoTop.

Van, M. A. (2001). New approach for uplift induced slope failure. *International Conference on soil mechanics and geotechnical engineering*, 2285–2288.

Van Den Eijnden, A., & Hicks, M. (2017). Efficient subset simulation for evaluating the modes of improbable slope failure. *Computers and Geotechnics*, *88*, 267–280. https://doi.org/10.1016/j.compgeo.2017.03.010

van den Eijnden, B., & Hicks, M. (2019). On the importance of a complete characterization of site investigation data uncertainty: A computational example. https://doi.org/10.3850/978-981-11-2725-0_IS9-3-cd

van der Meij, R. (2023, January). *D-stability - user manual*. Deltares. https://www.deltares.nl/software-en-data/producten/d-stability

van Duinen, A. (2012, February 10). Toelichting bij het protocol voor het uitvoeren van laboratoriumproeven.

van Duinen, A. (2013, March 18). *Memo back analyses of dikes that withstand a high water level*.

van Duinen, A. (2014, December 18). *Handreiking voor het bepalen van schuifsterkte parameters*.

van Duinen, A., Förster, U., van Hoven, A., Teunissen, H., & Zwanenburg, C. (2009, April 16). *SBW project werkelijke sterkte - projectplan 2009*. Deltares.

Vanmarcke, E. (1977). Probabilistic modeling of soil profiles. *Journal of the Geotechnical Engineering Division*, *103*(11), 1227–1246. https://doi.org/10.1061/AJGEB6.0000517

Vanmarcke, E. (1983). *Random fields: Analysis and synthesis* [OCLC: 875492250]. MIT Press.

Varkey, D. (2020). *Geotechnical uncertainties and reliability-based assessments of dykes* (Doctoral dissertation). Delft University of Technology. https://doi.org/10.4233/UUID:48460103-30A0-4338-8C4A-EBC80C73C2D3

Varkey, D., Hicks, M., & Van Den Eijnden, A. (2023). Reliability-based partial factors considering spatial variability of strength parameters [Series Title: Lecture Notes in Civil Engineering]. In *Challenges and innovations in geomechanics* (pp. 299–304). Springer International Publishing. https://doi.org/10.1007/978-3-031-12851-6_36

Verruijt, A. (1983). Grondmechanica.

VNK2. (n.d.-a). *1206930-000-BGS-0013-v4-m-geologische beschrijving dijkring 34a*.

VNK2. (n.d.-b). *1206930-000-BGS-0052-v4-m-geologische beschrijving dijkring 35*.

Wang, Y., Guan, Z., & Zhao, T. (2019). Sample size determination in geotechnical site investigation considering spatial variation and correlation [Publisher: NRC Research Press]. *Canadian Geotechnical Journal*, *56*(7), 992–1002. https://doi.org/10.1139/cgj-2018-0474

Webster, R., & Lark, M. (2013). Field sampling for environmental science and management. *Precision Agriculture*, *14*(1), 133–134. https://doi.org/10.1007/s11119-012-9293-2

Webster, R., & Oliver, M. A. (1992). Sample adequately to estimate variograms of soil properties. *Journal of Soil Science*, *43*(1), 177–192. https://doi.org/10.1111/j.1365-2389.1992.tb00128.x

Zhu, H., & Zhang, L. M. (2015). Evaluating suction profile in a vegetated slope considering uncertainty in transpiration. *Computers and Geotechnics*, *63*, 112–120. https://doi.org/10.1016/j.compgeo.2014.09.003

Zienkiewicz, O. C., & Cormeau, I. C. (1974). Visco-plasticity—plasticity and creep in elastic solids—a unified numerical solution approach. *International Journal for Numerical Methods in Engineering*, *8*(4), 821–845. https://doi.org/10.1002/nme.1620080411

Acronyms

L	Load	105–107
R	Resistance	105–107
Z	Limit State Function	106
ACF	Auto Correlation Function	12, 14, 16, 20, 59, 60, 62, 63, 66
CDF	Cumulative Density Function	70–72, 81–83, 85, 89, 92, 93, 95, 119, 120
CPT	Cone Penetration Test	12, 13, 24, 79, 88, 95
CRS	Constain Rate of Strain	24, 27, 115
CSSM	Critical State Soil Mechanics	6, 7, 60, 112–115
DSS	Direct Simple Shear	24, 26–28, 42, 49, 115, 116
FEM	Finite Element Method	4, 7, 17–20, 62, 64–67, 70–72, 76–80, 84, 85, 89, 92, 95, 108, 110, 119
FORM	First Order Reliability Method	9, 73, 85, 86, 95, 106, 108
FoS	Factor of Safety	10, 11, 17, 19, 55, 57, 62–64, 67–70, 72, 73, 75–80, 92, 93, 95, 108, 110, 111, 119
LAS	Local Average Subdivision	22
LEM	Limit Equilibrium Method	7, 17, 76, 85, 110
MLE	Maximum-Likelihood Estimation	14
MoM	Method of Moments	13, 14, 26
NC	Normally Consolidated	113
OC	Over Consolidated	113
OCR	Over-Consolidation Ratio	76–78, 89, 113, 115
PDF	Probability Density Function	49, 70, 105
POP	Pre-Overburden Pressure	76, 115

RFEM	Random Finite Element Method	ii, 2–5, 17, 18, 20, 54, 57–59, 63–66, 68, 70–73, 75, 76, 78, 79, 81–86, 88–95, 119, 120
RL	Rivierenland	25, 27, 40–42, 45–48, 51, 91, 122
SHANSEP	Stress History and Normalized Soil Engineering Property	3, 6, 7, 9, 19, 24, 26, 27, 43–45, 47, 48, 51, 56, 57, 60, 76, 78–80, 88–90, 92–95, 114, 115, 122, 123
SRF	Strength Reduction Factor	19, 119
STOWA	Applied Water Management Research Foundation (Stichting Toegepast Onderzoek Waterbeheer)	3, 24, 25, 27, 28, 33, 34, 36, 38–48, 52, 57, 88, 90, 91, 95, 121–124, 126
ULS	Ultimate Limit State	4
WBI	Legal Assessment Instrument (Wettelijk BeoordelingsInstrumentarium)	1, 4, 75, 89
WNLS	Weighted Non-Linear Least Squares	27, 31, 34, 118

Symbols

Sign	Description	Unit
H	Height of the slope	m
P_f	Failure probability	$1/\mathrm{year}$
$P_{f,cs}$	Probability of failure of a dike section	$1/\mathrm{year}$
R_{ps}	p-s ratio	
S	Undrained shear strength ratio	
$X_{med,char}$	Characteristic value of the sample median	
X	Random soil strength parameter X	
Δd	Bin width	m
α	Ratio between local and regional variance	
γ_{wet}	Wet volumetric weight	$\mathrm{kN/m^3}$
γ	Semivariogram	
$\hat{\gamma}$	Semivariance	
$\hat{\rho}$	Experimental or sample autocorrelation function	
$\mu_{\ln X}$	The natural logarithm of the sample mean	
ω	Residual correlation factor	
ϕ'_{cs}	Critical state friction angle	$^\circ$
ϕ	Friction angle	$^\circ$
ρ	Correlation coefficient	
σ'_p	Pre-consolidation stress	$\mathrm{kN/m^2}$
$\sigma_{X,loc,avg,nor}$	Adjusted standardard deviation of normal distributed parameter X	
$\sigma_{\ln X}$	The natural logarithm of the sample standard deviation	
τ	Lag	
θ_x	Scale of fluctuation perpendicular to the dike	m
θ_z	Vertical scale of fluctuation	m
θ	Scale of fluctuation	m
$\tilde{X}_i$	Predicted measurement of a soil property	
ε	Residual of a soil property X	
m	Exponent for additional strength in overconsolidated soils	
n	Nugget	
p	Partial sill	
r'	Effective range	m
r	Range	m
s_u	Undrained shear strength	kPa
s_{ij}^2	Squared difference	
s	Sill	
t	Trend function of a soil property X	
u_i	Location of sample point	

A

Supplementary Literature

A.1. Reliability Analysis Theory

Reliability refers to the ability of structures and elements to function without failing to maintain acceptable risk levels. Risk is determined by the probability of occurrence multiplied by its potential consequences. Reliability assessment ensures that the design meets acceptable risk levels and can withstand various loads and conditions (Fenton & Griffiths, 2008).

The limit state function specifies the relationship between the Load (L) acting on the structure and its Resistance (R):

$$Z = R - L \tag{A.1}$$

When the limit state function, also known as the Z-function, has a negative value, it indicates structural failure. The limit state function (also $Z(\boldsymbol{X})$) includes all variables and factors contributing to the structure's strength and the load it experiences, which can be combined in the vector $\boldsymbol{X}$ (ENW, 2016). Note that in geotechnical engineering, the loading and resisting components are not distinctly separable, and it is difficult to separate them as independent variables.

The definition of the probability of failure (P_f) is the probability that the limit state function is lower than zero (Schweckendiek et al., 2017).

$$P_f = P(R < S) = P(Z(\boldsymbol{X}) < 0) \tag{A.2}$$

Parameters defining the strength and load of a dike ($\boldsymbol{X}$) are represented by their Probability Density Function (PDF). These PDFs describe the parameters' uncertainty or (spatial) variability. The joint probability density function ($f_{\boldsymbol{x}}$) can be obtained by combining all the individual PDFs together, as shown in Figure A.1. This joint probability density function can be used to calculate the probability of failure (P_f):

$$P_f = \int_{Z(\boldsymbol{X})<0} f_{\boldsymbol{x}}(\boldsymbol{x}) d\boldsymbol{x} \tag{A.3}$$

The reliability index defines the reliability of a dike and is a function of P_f.

$$\beta = \Phi^{-1}(1 - P_f) \tag{A.4}$$

$$P_f = \Phi(-\beta) \tag{A.5}$$

Where Φ is the cumulative distribution function of a standard normal distribution.

A.1.1. Reliability Analysis Levels

There are four levels used to assess the reliability and failure probability of a dike cross-section:

0. Deterministic method;
1. Semi-probabilistic method;
2. Probabilistic approximations;
3. Probabilistic method.

In probabilistic calculations, the joint probability density function is determined through analytical formulations, numerical integration, a First Order Reliability Method (FORM), or a Monte Carlo analysis (Jonkman et al., 2015).

The First Order Reliability Method is commonly used as an approximation of probabilistic methods. It involves linearizing the failure probability function (Limit State Function Z) using a Taylor expansion at the design point. The design point represents the most probable failure location, with the highest probability density along the limit state ($Z = 0$). The precise location of the design point is determined through an iterative process in the FORM analysis.

In a semi-probabilistic calculation, design values are used in failure mechanism models. These design values combine a characteristic value of a parameter with a safety factor (ENW, 2016) (see Figure A.2). The characteristic value should be determined so that the calculated reliability of the geotechnical structure is 95% (Hicks, 2012).

The deterministic method is unsuitable for determining the probability of flooding (ENW, 2016). It relies on single values for strength, resistance, and a global (empirical) safety factor to quantify safety (Jonkman et al., 2015).

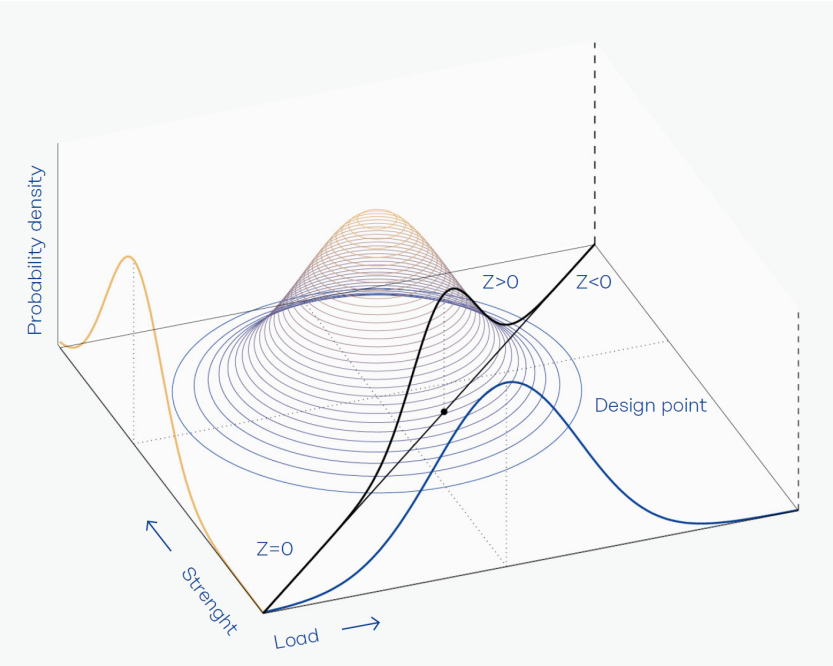


Figure A.1: The joint probability density function of the Resistance (R) and Load (L), with the Limit State Function (Z) (ENW, 2016).

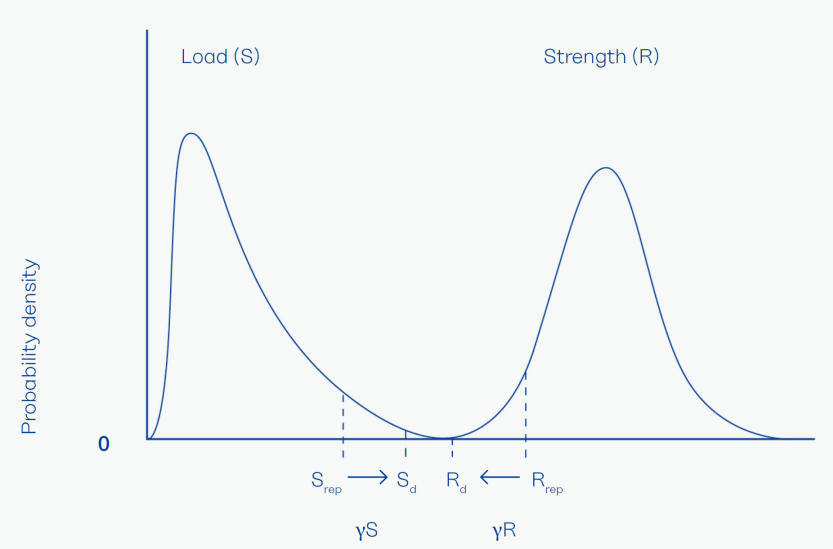


Figure A.2: The probability density function of the Resistance (R) and Load (L) of a dike. Here, the determination of the design values for a semi-probabilistic approach is illustrated. (ENW, 2016)

A.2. Current Method: Dike Trajectory Subdivision

Rijkswaterstaat (2021) has established a specific procedure to determine the subdivision of a dike trajectory into dike sections, which must be evaluated separately for each failure mechanism:

1. The first step involves creating a dike trajectory using a global SOS segment. The Stochastische Ondergrondschematisatie (SOS) is a stochastic subsurface schematization developed by Hijma and Sun Lam (2015). It defines different scenarios, each with its probability of occurrence, to provide an initial indication of the various geotechnical conditions that must be assessed in different dike sections.
2. As more localized data on soil layering becomes available, such as CPTs and boreholes, the trajectory can be divided into initial sections, and scenarios can be adjusted if, for example, the thickness of layers varies significantly.
3. The next step involves investigating the geotechnical conditions of the layers to refine the subdivision further. If there are significant changes in the geotechnical conditions, separate sections can be defined.
4. The final step involves considering other factors when defining dike section boundaries. Rijkswaterstaat (2021) identifies several factors:
    - The boundary between different types of flood defenses (e.g., dike and dune);
    - A significant change in hydraulic conditions;
    - Differences in geometry such as height, width, and slope angle of the dike or the presence of a berm;
    - The identification of a weak zone during site investigation;
    - Changes in dike body material;
    - The presence of constructional elements like sheet piles.

It is important to note that the process of defining dike sections is iterative and may involve defining multiple representative cross-sections for each dike section. These cross-sections should then be evaluated separately.

A.2.1. Calculation of P_f

The Water Law (Waterwet) states the failure probability requirement ($P_{f,req}$) for each primary dike trajectory in the Netherlands. When considering the length effect, this can be translated to a requirement

on a cross-sectional level. Only the failure probability of a cross-section due to macro-stability ($P_{f,cs}$) will be investigated as this research does not discuss the length effect.

The conditional probability of failure (P_{cs}) is the outcome of a probabilistic or FORM analysis, while for a semi-probabilistic analysis, it is a Factor of Safety (FoS). The P_{cs} can be calculated from the FoS using a calibration formula. This formula relates the reliability index (see Appendix A.1) with a target reliability-dependent safety factor (γ_n).

A study by Kanning et al. (2017) determined the current calibration formula for macro-stability, used in the WBI2017 Rijkswaterstaat (2021), based on 34 cases from different locations in the Netherlands. In this study, the $P_{f,cs}$ was calculated using a FORM analysis and compared with the calculated FoS. This resulted in the calibration formula in Equation A.6.

$$\gamma_n = 0.150 * \beta_{req,cs} + 0.410 \tag{A.6}$$

This calibration formula can be used to obtain the reliability index of a cross-section. For example, Equation A.6 can be re-formulated as:

$$\beta_{cs} = \frac{\frac{FoS}{\gamma_d} - 0.41}{0.15} \tag{A.7}$$

Where γ_d is the model factor, which is 1.06 in the case of Uplift-Van and FEM (POVM, 2020; Rijkswaterstaat, 2021). By using the reliability index, the probability of failure of a cross-section can be determined using Equation A.5.

A.3. Macro-Stability

Spatial variability is an important driving factor in failure due to macro-stability next to the load and geometry (ENW, 2016). This paragraph will describe the failure mechanism and its calculation techniques.

A.3.1. Mechanism

Macro-stability impacts both the dike body and the soil foundation underneath it. As defined by 't Hart et al. (2016) (translated from Dutch): "Macro-instability occurs when large sections of a soil mass shear along straight or curved deep shear planes. Shallow shear (shear plane up to about 1 m below ground level) is addressed under dikes and/or micro-instability." Together with internal erosion, instability of a dike has been identified by Jonkman and Schweckendiek (2015) and ENW (2016) as one of the primary causes of past flooding events.

Instability within a dike occurs when the equilibrium cannot be maintained. The weight of the dike body is a critical factor in both driving and resisting an instability where the active side acts as a load, while the passive side serves as resistance to sliding. Shear forces along the shear plane play an important role in the resistance against a slip plane. Figure A.3 shows a graphical representation of this failure mechanism.

A high water level will raise the water head and phreatic level by infiltration into the dike body, causing an increase in water pressures in the dike. This reduces effective stress, the primary driver of shear strength, and subsequently decreases the mobilized shear strength of the soil. A failure surface will propagate through the plastic zone in the dike where the shear strength is exceeded. This will ultimately cause the soil mass to slide and can trigger a dike breach (see Figure A.4). The failure is mainly triggered by the decrease in resisting forces rather than an increase in overturning forces due to the higher load from the rising water level, as is often the case with other failure mechanisms ('t Hart et al., 2016).

Several factors that will influence the (rate of the) rise in water head and phreatic level:

- The dike material and the presence of impermeable layers between the aquifer influence the rate of adaption of water head and phreatic level inside and under the dike. For example, when faced

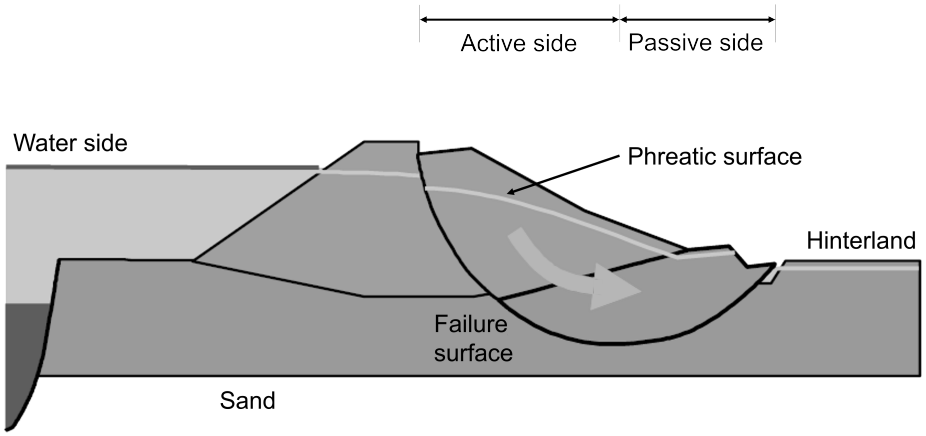


Figure A.3: Macro-instability of the inner slope of a dike. Adapted from 't Hart et al. (2016).

with high outside water levels, a more permeable, sandy dike will experience a faster increase in pore pressures inside the dike compared to a dike with a clay cover.

- The distance between the outside water level and the dike also plays a role. The greater the distance, the slower water will penetrate through the soil and affect the water pressures inside the dike.
- While large amounts of precipitation can increase the phreatic level, the infiltration rate is considerably lower than in the case of wave overtopping ('t Hart et al., 2016).

The high hydraulic head in the aquifer can cause the impermeable layer on top of the aquifer to be uplifted, which decreases the shear strength further (BOI, 2023). When the pressure gets too high, water can seep through the impermeable layers, causing a blowout and reducing the shear stresses to zero.

After the initial sliding of the slope, subsequent mechanisms can contribute to breaching the dike. The causes and consequences of instability are depicted in Figure A.4.

A complete breach of the dike occurs only when there is displacement over the entire width of the crest, allowing water to flow freely to the other side. It takes time, ranging from an hour to several days, for a new equilibrium to be established ('t Hart et al., 2016). The soil maintains some residual strength during sliding, leading to a new equilibrium between resisting and overturning forces. In many cases, the sheared zone displaces only 1 or 2 meters under the original crest ('t Hart et al., 2016). If part of the crest remains intact after slope sliding, a breach occurs only after subsequent slides when the entire crest is demolished.

One observable sign of a developing failure surface is the creation of large cracks parallel to the dike's crest. In the plastic zone, these cracks can develop when the mobilized shear strength just counteracts the development of the failure surface. The presence of a sandy top layer on the dike results in smaller deformation until failure, but the time between crack creation and failure is shorter compared to a top layer of clay, which is more prone to creep ('t Hart et al., 2016). Cracks do not necessarily lead to failure if appropriate measures are taken for repair. They can also indicate dehydration of a clay dike, which may trigger horizontal sliding ('t Hart et al., 2016).

As explained above, determining the probability of dike failure due to macro-stability depends on several factors, including the soil's strength. Since there are often significant uncertainties in this regard, probabilistic methods are frequently used to assess the probability of failure. Please refer to Appendix A.1 for the definition of the probability of failure and an outline of the various reliability analysis levels used to determine it.

A.3.2. Calculation Techniques

Various techniques exist to evaluate the stability of a dike. Each approach varies in its assumptions, for example, about the shape of the failure surface: (multi)planar, circular, or non-circular. However, in

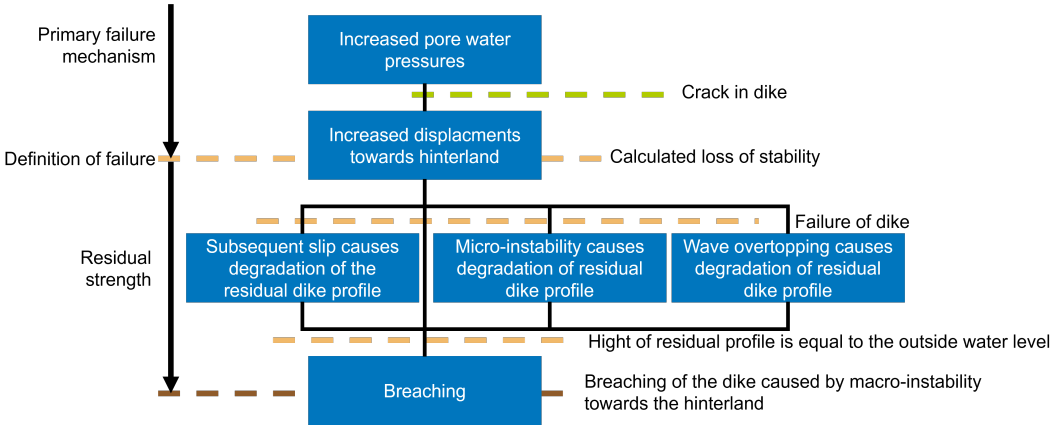


Figure A.4: Events that cause a breach of a dike after an instability. Adapted from 't Hart et al. (2016).

reality, the failure surface varies widely in form and is predominantly influenced by factors such as the geometric model, hydro-geological conditions, and the geological environment (Sharp et al., 2013).

Two main categories of dike assessments will be discussed: numerical and Limit Equilibrium Method (LEM). Within LEM, three specific methods exist: Bishop (Bishop, 1955), Uplift-Van (Van, 2001), and Spencer (Spencer, 1967). These methods are the most commonly used methods in Dutch practice and can be modeled using the widely used software D-Stability (van der Meij, 2023).

LEM is a widely used technique to assess the stability of dikes under static conditions. It assumes that a dike will fail if a potential slip plane's horizontal, vertical, or moment equilibrium is disturbed. LEM is a simple method, yet it can model detailed soil profiles, seepage, and various loading scenarios (Sharp et al., 2013). It is also limited as it can not model deformation-dependent soil behavior (Jonkman et al., 2021).

When using numerical models, also known as the Finite Element Method (FEM), the choice for the constitutive model is highly flexible. Additionally, it is also capable of modeling groundwater flow and soil-structure interaction. However, due to the complexity of the FEM, input errors are more easily made, and the calculations are more computationally intensive and time-consuming than the LEM (Jonkman et al., 2021). A more detailed explanation of the theory behind this method will be provided in Chapter 2.3.1.

Bishop

Bishop (1955) assumes a circular slip surface divided into slices. The method calculates the equilibrium of forces on each slice, as shown in Figure A.5, and the Factor of Safety (FoS) can be calculated by:

$$FoS = \frac{M_r}{M_s} \tag{A.8}$$

The resisting moment, denoted by M_r, is determined by the sum of the moments caused by the shear stress along the slip surface and the weight of the soil on the passive side of the failure surface. On the other hand, the driving moment M_s is determined by the weight of the soil and any additional loads on the active side (Jonkman et al., 2021).

Uplift-Van

The method proposed by Van (2001) is an extension of Bishop's method (Bishop, 1955) that incorporates a horizontal and a second circular part to account for uplift effects. When high water pressures occur on one side of a dike, water can flow through the aquifer to the other side. An impermeable blanket layer can limit the water from flowing to the surface. However, pore pressures may increase under this blanket layer, significantly reducing the effective stress and shear strength (Jonkman et al., 2021). Since Dutch dikes often feature a blanket layer over an aquifer (Jongejan, 2017), this method is the default for Dutch dike assessments (Rijkswaterstaat, 2021). By applying the horizontal forces

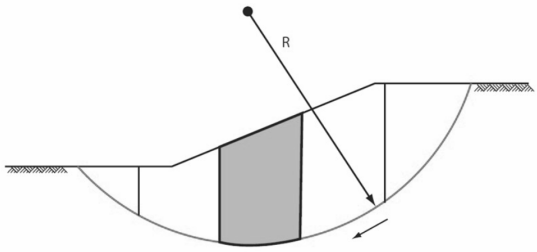


Figure A.5: Schematic representation of a slip circle as modeled with methods of slices such as Fellenius or Bishop (Jonkman et al., 2021).

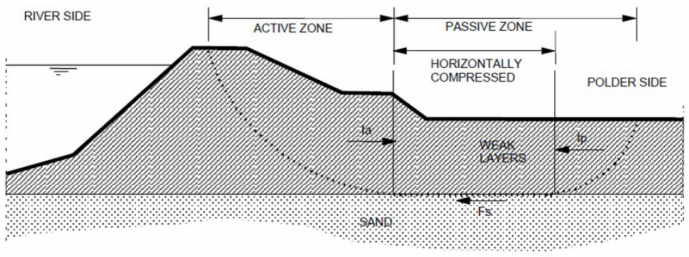


Figure A.6: Schematic representation of the Uplft-Van model with a horizontal element in between two circular segments (one on the active side, one on the passive side) (Jonkman et al., 2021).

equilibrium on the horizontal part, the factor of safety (FoS) can be calculated. See Figure A.6 for an illustration of the method.

D-Stability identifies the normative failure surface by iteratively testing surfaces within a user-defined search area. The surface with the lowest FoS is the normative one.

Spencer

The Uplift-Van and Bishop method is limited as it assumes the shape of the failure surface. However, the Spencer method (Spencer, 1967) overcomes this limitation by defining the failure surface with an entrance and exit point and some points in between. These points are connected by linear lines, as shown in Figure A.7.

Previously, determining the failure surface was a manual task that required engineering experience (Jonkman et al., 2021). However, as implemented in D-Stability, the Spencer-van der Meij method uses a genetic algorithm to automate this process. This method is based on the concept of genetic survival of the fittest. van der Meij (2023) describes the algorithm as: ”In a genetic algorithm, a population of candidate solutions (individuals) to an optimization problem is evolved toward better solutions. Each candidate solution has a set of properties. In the case of this kernel, a set of properties that describe the slip plane. These properties evolve over several generations towards an optimal solution.”

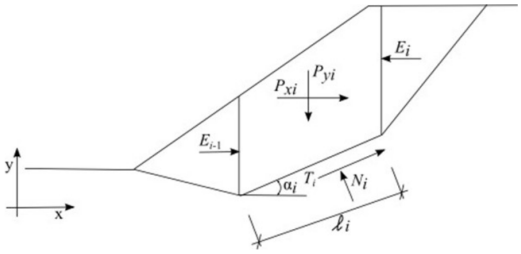


Figure A.7: Schematic representation of the Spencer model with piece-wise linear elements (Jonkman et al., 2021).

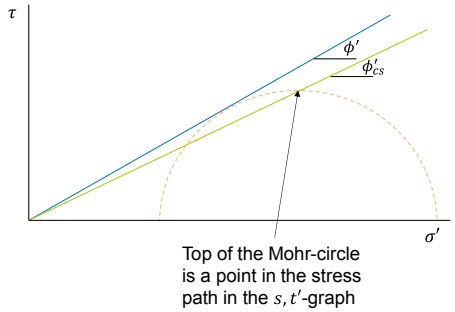


Figure A.8: The red line crossing the top of the Mohr circle, which represents the critical state line, can be transformed into the blue line that touches the Mohr circle at the point of contact, known as the failure envelope. Adapted from van Duinen (2014).

A.3.3. Critical State Soil Mechanics

Large strains are needed to achieve macro-instability that mobilizes the strength over the entire failure surface (BOI, 2023). Rijkswaterstaat (2021) has determined that the appropriate way to analyze macro-stability is through the concept of Critical State Soil Mechanics (CSSM) (Schofield & Wroth, 1968). This approach focuses on critical state shear strength rather than peak shear strength. The critical state is where the soil no longer changes its volume during shearing. All soils ultimately reach the critical state line, regardless of their initial conditions. As shown in Figure A.8, the top of the Mohr circle represents a point on the critical state line.

The critical state line can be better visualized in Figure A.9 by plotting the half of the deviator stress t against the average effective principal stress s'.

$$t = \frac{1}{2}(\sigma'_1 - \sigma'_3) \tag{A.9}$$

$$s' = \frac{1}{2}(\sigma'_1 + \sigma'_3) \tag{A.10}$$

Where σ'_1 and σ'_3 are the major and minor principal stresses, for more information about the principal stresses, please refer to Paragraph A.3.4. The failure envelope in Figure A.8 is positioned at an angle known as the critical state friction angle (ϕ'_{cs}). In the s', t-space, this angle can be defined as follows:

$$t_{max} = s' \sin(\phi'_{cs}) \tag{A.11}$$

Where t_{max} is the maximum mobilized shear strength [$\mathrm{kN/m^3}$] at 25% axial strain. van Duinen (2012) recommends determining ϕ'_{cs} from multiple anisotropically loaded undrained triaxial tests on normally consolidated soils.

Stress Paths in the CSSM Framework

The stress paths in the CSSM framework are shown in Figure A.9. These stress paths indicate the shear strength t that can be mobilized based on the stress and drainage conditions at the beginning of the stress paths. The initial stress conditions are determined by the effective stress s' and its relative position to the pre-consolidation stress (σ'_p). The stress paths in triaxial tests are divided into four categories, illustrated in the figure.

1. *Drained versus undrained shearing*: When soil shearing occurs under undrained conditions, excess loads are transferred to the pore water because the water cannot flow away, generating excess pore water pressure. This usually happens in soils when the applied load is faster than the consolidation rate. In dikes, rapidly occurring shear is typical for macro-stability (Rijkswaterstaat, 2021). The Dutch guidelines require modeling undrained layers for impermeable soils like

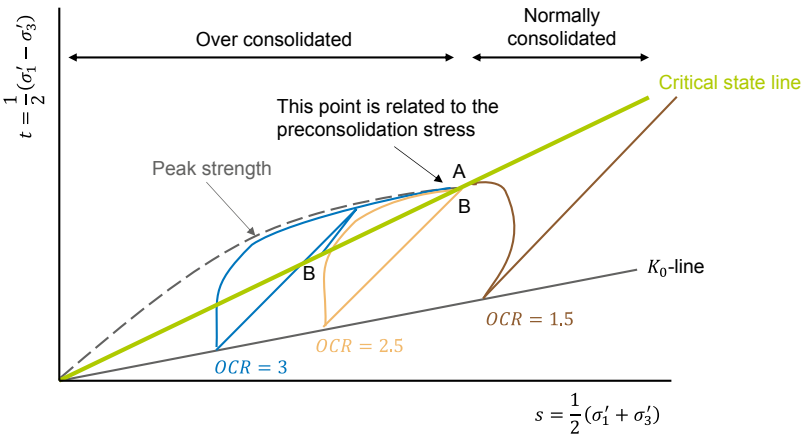


Figure A.9: In the Critical State Soil Mechanics (CSSM) framework, the behavior of soil can be analyzed in terms of drained (dashed lines) and undrained behavior. Three lines describe the behavior of different types of soils. The yellow line represents Normally Consolidated (NC) or lightly Over Consolidated (OC) soils with an Over-Consolidation Ratio (OCR) range of 1 to 2. For these soils, the soil will contract upon undrained shearing. The green line represents OC soils with an OCR range of 2 to 3, which will not contract or dilate upon undrained shearing. Finally, the blue line represents OC soils with an OCR greater than 3, which will contract upon undrained shearing. Adapted from Rijkswaterstaat (2021).

peat and clay under the phreatic line. In the next paragraphs, additional information will be given about the drained and undrained modeling of layers in the CSSM framework.

In Figure A.9, the difference between the drained (dashed lines) and undrained (continuous lines) stress paths is the generated excess pore water pressure.

2. *Normally consolidated or lightly over-consolidated soils*: When the current effective stress is lower than the pre-consolidation stress, the soil will contract upon shearing (see Figure A.10). This happens when the soil is normally consolidated or lightly over-consolidated ($1 \leq$ OCR ≤ 2). This contraction will cause excess pore water pressure to be generated. At the critical state during undrained shearing, the stress path will reach the critical state line at a smaller shear strength t than in the drained case.

   In Figure A.9, this case is illustrated by the yellow lines.

3. *Medium over-consolidated soils*: When the OCR is between 2 and 3, the soil neither contracts nor dilates, and the drained stress path will end up at the same shear strength t at the critical state line.

   In Figure A.9, this case is illustrated by the green lines.

4. *Highly over-consolidated soils*: When the soil is highly over-consolidated (OCR > 3), it will dilate upon shearing (see Figure A.10). This dilation generates under-pressures in the pores, increasing the soil's effective stresses. Therefore, in the critical state, the undrained soil has a higher shear strength t than the undrained case, as shown by the blue lines in Figure A.9.

All samples start on the K_0 line due to anisotropic consolidation with no radial strain. In over-consolidated soils, the undrained stress path approaches the peak strength before ending up at the critical state line. Figure A.10 also illustrates this extra mobilized peak strength.

The amount of over-consolidation is important because it determines whether the soil has extra strength during undrained shearing, as explained in the fourth case. According to Rijkswaterstaat (2021), the Holocene layers in the Netherlands play a significant role in macro-stability assessments. These layers were formed after the ice ages and have not been subjected to heavy ice loads. Their over-consolidation, which typically ranges with an OCR from 1 to 5, is mainly caused by factors such as creep, variations in water levels in the polder, weather conditions, biochemical activity, etc. (Rijkswaterstaat, 2021) and can therefore vary significantly.

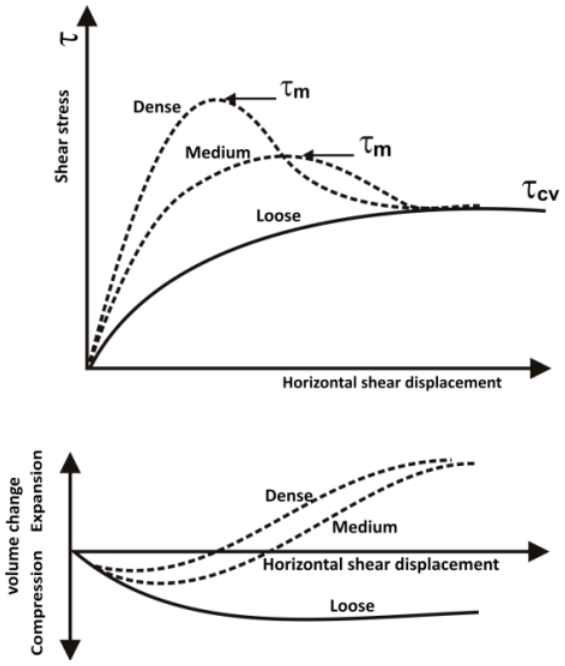


Figure A.10: Shear stress versus shear displacement for loose, medium, and dense soils (Das, 2019).

Drained Modeling in the CSSM Framework

The non-associative Mohr-Coulomb drained model should be used to model permeable layers such as sand and the layers above the phreatic line, as required by Rijkswaterstaat (2021). Since water can flow freely, any generated excess pore pressures can dissipate, and the effective stresses determine the shear strength of the soil:

$$\tau = \frac{(c + \cos\phi + \sigma' \sin phi) \cos\psi}{1 - \sin\psi \sin\phi} \tag{A.12}$$

Where:

- τ is the drained shear strength [kPa].
- c is the cohesion [kPa].
- ϕ is the friction angle [°].
- σ' is the normal effective stress [kPa].
- ψ is the dilation angle [°].

The dilation angle (ψ) is a measure of the increase in volume of a compact material during shear deformation. In the context of CSSM, there is no change in volume at the critical state, so $\psi = 0°$. Additionally, in the CSSM framework, cohesion is defined as added strength resulting from over-consolidation. However, at the critical state, initial conditions such as over-consolidation no longer have any effect, so $c = 0\text{kPa}$. This simplifies Equation A.12 to:

$$\tau = \sigma' \sin\phi \tag{A.13}$$

Undrained Modeling in the CSSM Framework: SHANSEP

Two studies, conducted by van Duinen (2013) and van Duinen et al. (2009), examined failed and stable dikes during high water conditions in the Netherlands. The studies found that the shear strength was significantly overestimated when undrained layers were modeled with a drained Mohr-Coulomb model. They concluded that a drained model would be more suitable for the soil's behavior. As a result, the Stress History and Normalized Soil Engineering Property (SHANSEP) model, developed by Ladd and

Foott (1974), has been used since 2017 to determine the strength of undrained layers in the Dutch primary dikes. This model is related to the CSSM framework as it can describe the added strength during undrained shearing in the case of highly over-consolidated soils. Ladd and Foott (1974) found that cohesive soils with equal OCR display similar stress-strain behavior when normalized over the vertical effective stress. They defined a relationship between the undrained shear strength of the soil and the vertical effective stress σ'_v:

$$s_u = S * \sigma'_v * OCR^m \tag{A.14}$$

Where:

s_u	is the undrained shear strength [kPa].
S	is the normally consolidated undrained shear strength ratio [-].
σ'_v	is the effective vertical stress [kPa].
OCR	is the overconsolidation ratio of the sample [-].
m	is the strength increase exponent [-].

The OCR can be determined by:

$$OCR = \frac{\sigma'_p}{\sigma'_v} \tag{A.15}$$

Where σ'_p is the pre-consolidation stress [kPa].

SHANSEP S
The normally consolidated undrained shear strength ratio (S) is a measure of the undrained shear strength s_u of a sample regarding its consolidation stress. This ratio considers the pore pressures and frictional forces generated during undrained shearing (Rijkswaterstaat, 2021).

To determine S, it is important to test the sample at "large" strains. This corresponds to a 25% axial strain in a triaxial test for clays and sands. A 40% shear strain should be used for peat during a Direct Simple Shear test. Both tests should be consolidated anisotropically, performed in a single stage and at a constant height (van Duinen, 2012).

SHANSEP m
The undrained shear strength of soil is affected by changes in the effective stress due to variations in the pore pressure. This sensitivity is defined by the SHANSEP parameter m (Rijkswaterstaat, 2021). m should preferably be determined based on Oedometer and Constain Rate of Strain (CRS) tests.

To explain the influence of m, two cases are highlighted:

- When m equals 1: a decrease in effective stress caused by increased pore pressure does not affect the undrained shear strength. This is because the decrease in σ'_v is compensated by an increase in the OCR, also defined by σ'_v. Since m equals 1, the increase in the OCR is fully incorporated into the undrained shear strength, keeping it constant.
- When m equals 0.5: a decrease in σ'_v will cause the same increase in OCR, but this time the OCR is not fully incorporated into the undrained shear strength. Instead, it is incorporated with an exponent of 0.5. Therefore, a decrease in σ'_v will not be canceled out by the increase in OCR, causing the undrained shear strength to decrease.

Preconsolidation Stress
As explained in the description of the different stress paths, the level of over-consolidation is a crucial factor in the CSSM framework as it determines the soil's additional strength during undrained shearing. To determine the OCR, the ratio between the current effective stress and the pre-consolidation stress should be analyzed.

The pre-consolidation level in Holocene layers is caused by various processes and is, therefore, highly variable. Next to the OCR and σ'_p, it can also be defined by the Pre-Overburden Pressure (POP) [kN/m^3]:

$$POP = \sigma'_v(OCR - 1) \tag{A.16}$$

Oedometer tests can be conducted to determine the pre-consolidation stress. However, it is important to ensure that the sample is not disturbed.

A.3.4. Principal Stresses along the Failure Surface

Consider the stresses acting on a soil element as depicted in Figure A.11. The principal stresses are the maximum normal stresses that can act on the component. The soil element can be rotated at a certain angle (known as the principal angle, θ_p) so that no shear stress is exerted on the sides of the element. The normal stresses acting on this principal plane are the principal stresses. These principal stresses can be defined in three dimensions: σ_1, σ_2, and σ_3, respectively called the major, intermediate, and minor principal stress.

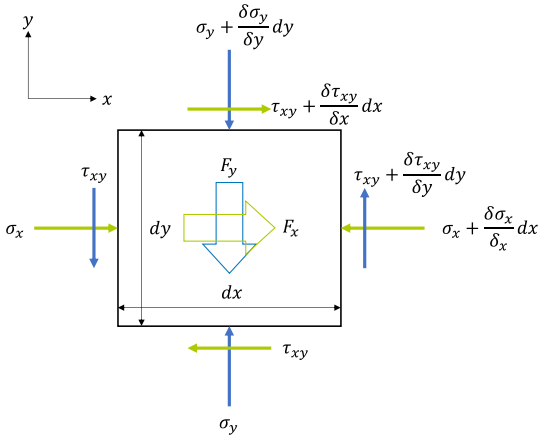


Figure A.11: Stresses on a soil element. The vertical stresses are depicted in blue, and the horizontal stresses in green.

In the case of a dike, the intermediate principal stress is somewhere between σ_1 and σ_3, and it always acts in the out-of-plane direction. Suppose the major and minor principal stresses, along with the principal angle, are known. In that case, the horizontal and vertical stresses can be calculated (and vice versa) by using the following formula (Verruijt, 1983) (only major and minor principal stress):

$$\sigma_x = \sigma_1 \sin^2 \theta_p + \sigma_3 \cos^2 \theta_p \tag{A.17}$$

$$\sigma_y = \sigma_1 \cos^2 \theta_p{}^2 + \sigma_3 \sin^2 \theta_p \tag{A.18}$$

$$\tau_{xy} = \sigma_1 \sin \theta_p \cos \theta_p - \sigma_3 \cos \theta_p \sin \theta_p \tag{A.19}$$

Where the shear stress in the x, y-plane is denoted by τ_{xy}. These equations are based on the Mohr circle, a graphical method for presenting stress transformations.

In a triaxial compression test, the major principal stress (σ_1) acts in the vertical direction, while the minor principal stress (σ_3) acts horizontally. In this case, the intermediate principal stress (σ_2) equals σ_3. In a triaxial extension test, the principal planes are flipped compared to the compression test: the major principal stress acts horizontally.

Figure A.12 illustrates the rotation of the principal stresses along the failure surface. At the entry point of the failure surface, the major principal stresses act vertically, corresponding to the relatively large vertical compression stresses that occur in this location of the slope. This is similar to the conditions in a triaxial compression test. As the exit point is approached, on the passive side of the failure surface, the vertical stresses will be smaller than the horizontal compressive stresses, and the major principal stress direction will be horizontal, corresponding to triaxial extension. Between these zones, the governing stress will be the shear stress, corresponding to a state similar to a Direct Simple Shear (DSS) test. Here, the principal stresses rotate with the angle of the failure surface.

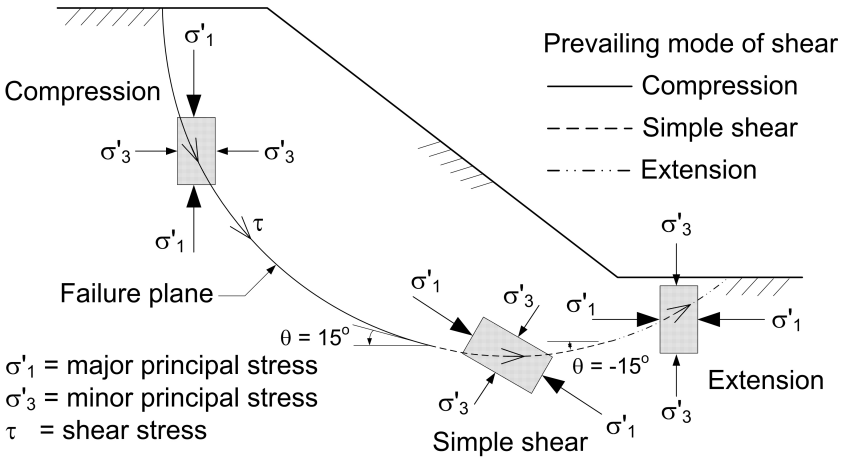


Figure A.12: Rotation of the principal stresses along the failure (Sadrekarimi, 2016).

A.4. Variogram Construction

To create a variogram and determine its parameters, the following steps should be performed:

1. *Detrending*: This step is optional since a variogram can be developed when the data is intrinsically stationary. However, trends make it harder to interpret the variogram parameters. A parameter with a trend is defined as the sum of the trend, residual, and noise. Where the residual is a zero-mean, second-order process and the noise a zero-mean, uncorrelated noise process. Detrending is often performed numerically by polynomial fitting or linear regression. However, since these functions involve estimating the trend, they introduce additional errors. Therefore, detrending should only be applied when the trend can be reliably estimated; otherwise, estimation errors can become significant. After detrending, the remaining data is second-order stationary, characterized by the sum of the residual and the noise.
2. *Creating a Variogram Cloud*: To create a variogram cloud, plot the squared differences versus the distance between all pairs of data points. The squared difference s_{ij}^2 between the value of data point i and data point j is defined as:
$$s_i j^2 = \{x(u_i) - x(u_j)\}^2 \tag{A.20}$$
3. *Binning the Variogram Cloud*: Group the distances between data pairs into bins with a bin width, Δd. Calculate the average squared differences of data pairs within each bin, representing that bin's semivariance. The sample variogram is constructed by plotting the semivariance of every bin against the bin's center. The semivariance $\hat{\gamma}$ of bin B_i can be calculated as:
$$\hat{\gamma}(d_i) = \frac{1}{2N_i} \sum_{||u_i - u_j|| \in B_i} s_i j^2 \tag{A.21}$$
Where:
N_i is the number of data pairs in bin B_i.
d_i is the center of bin B_i [m].
4. *Selecting an Appropriate Bin Width*: The bin width is crucial as it influences the scale over which squared differences are averaged, impacting the variogram's accuracy and resolution. A wide bin width averages out more measurement uncertainty but may lose information about small-scale variability. A narrow bin provides more information but retains more measurement uncertainty. An optimal bin width should be selected to balance uncertainty and detail (Oliver & Webster, 2015). In literature, guidelines about the number of pairs in each bin are given: 100 - 200. The number of points in each bin can be evaluated by using a histogram.

5. *Select Appropriate Cutoff Value*: To simplify variogram interpretation and limit uncertainty, a cutoff value for the lag axis should be chosen. Typically, the number of data pairs decreases as the distance between pairs increases. Therefore, a maximum distance exists beyond which the calculated semivariance becomes less reliable. The cutoff should be set in the variogram at the point in the histogram where the number of pairs in each bin begins to decrease. Set the cutoff at the point in the histogram where the number of pairs in each bin begins to decrease.
6. *Fitting Theoretical Functions*: A range of theoretical functions (see Table 2.4) can be fitted to the sample variogram ($\hat{\gamma}$) using Weighted Non-Linear Least Squares (WNLS). WNLS optimizes variogram parameters using non-linear regression by minimizing the sum of squared differences between sample variogram points and theoretical variogram fits. Additionally, it applies weights to the data points in the sample variogram with more reliability. It is defined as:

$$\sum_{i=1}^{n} w_i \{\hat{\gamma}(d_i) - \gamma(d_i|\Theta)\}^2 = min \tag{A.22}$$

   Where:

   Θ is the set of all model parameters $\Theta = (n, s, r)$.
   γ is the variogram.
   w_i is the weight for bin B_i.

   For the weight, which determines the reliability of each bin, various options can be used:

$$w_i = \frac{n_i}{d_i^2} \tag{A.23}$$

$$w_i = \frac{n_i}{\hat{\gamma}(d_i)^2} \tag{A.24}$$

$$w_i = n_i \tag{A.25}$$

$$w_i = \mathrm{Var}[\hat{\gamma}(d_i)]^{-2} \tag{A.26}$$

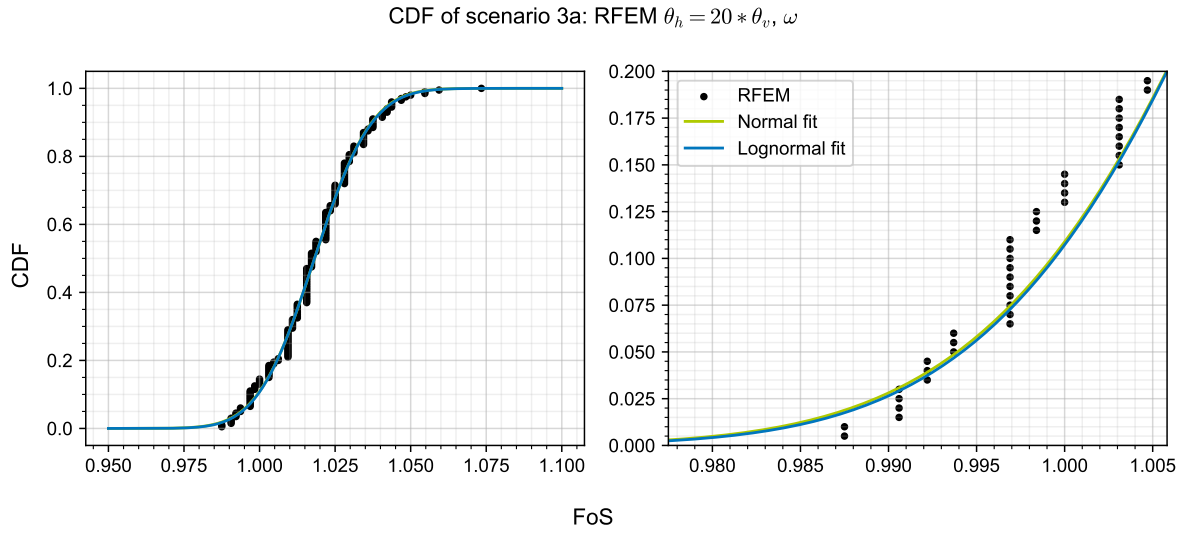
   Where n_i is the number of point pairs in bin B_i.
7. *Selecting a Theoretical Function*: To determine the semivariogram parameters, the optimal theoretical function should be selected. This can be based on criteria such as the lowest WNLS value. Furthermore, it should be noted that Gaussian variograms are only suitable for modeling extremely smooth, continuous data.

B

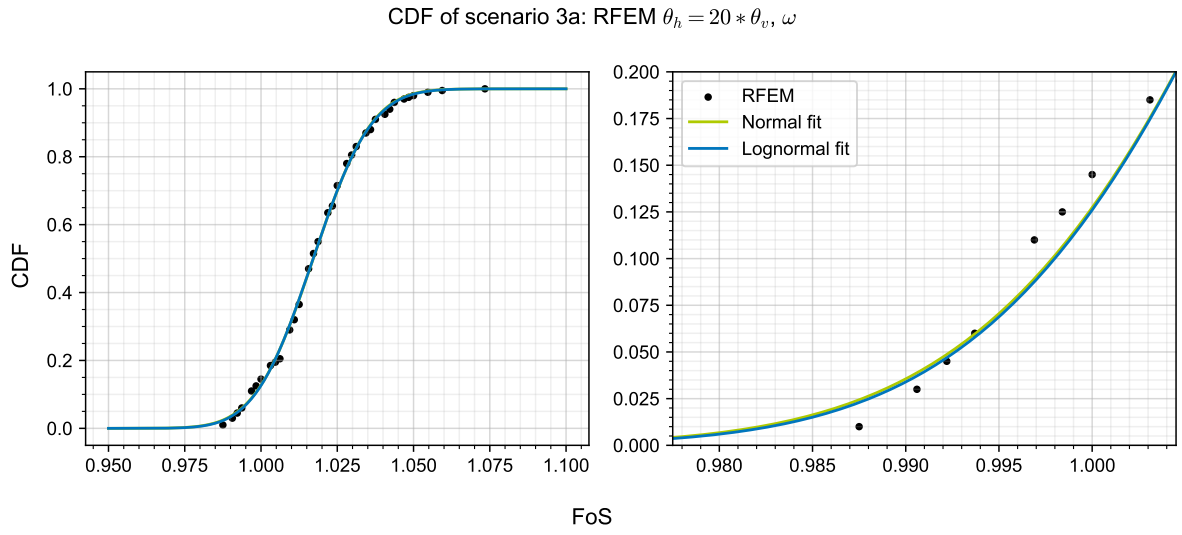
Precision RFEM

The study conducted Monte Carlo analyses for the probabilistic or Random Finite Element Method (RFEM) scenarios, where a total of 200 realizations were performed. The empirical Cumulative Density Function (CDF) of the calculated factors of safety (FoS) was constructed (see Equation ...), resulting in 200 data points. However, in scenario 3a, multiple points were found to be stacked on top of each other, indicating that the precision of the determination of the FoS in the FEM code is not high enough. This is primarily due to the precision of the strength reduction algorithm, which first reduces the Strength Reduction Factor (SRF) with steps of $\Delta SRF = 0.1$ until failure is achieved. After this, the algorithm refines the value of SRF iteratively by halving ΔSRF: $\Delta SRF_i = \Delta SRF_{i-1}/2$ and $SRF_i = SRFi - 1 \pm \Delta SRF$ where i is the iteration step in the algorithm. The minimum value of ΔSRF is 0.0015, which, combined with a narrow distribution of the factors of safety, results in the calculation of the same FoS multiple times, as illustrated in Figure B.1a. Note that, although not tested in this research, the precision of the strength reduction algorithm can be adjusted. However, this will result in a longer computation time.

To accurately model the probability of the occurrence of a particular value of FoS, only the top of the stacked points at a specific FoS should be used in the CDF, as this results in the determination of $P(X < FoS)$. This results in the CDF as depicted in Figure B.1b.



(a) CDF with 200 points.



(b) CDF with only the relevant points.

Figure B.1: CDFs for RFEM scenario 3a, plotted with 200 points or only the relevant points.

C

Extra Results Data Analysis

C.1. Volumetric Weight

C.1.1. Results Depth Range Analysis

Table C.1: Variogram parameters for the different depth ranges in the Applied Water Management Research Foundation (Stichting Toegepast Onderzoek Waterbeheer) (STOWA) database for the volumetric weight.

Soil	Depth range [m]	Nr. of points N	Bin width Δd	Eff. range r' [m]	Range r [m]	P. sill p	Sill s	Nugget n	Ratio p/s
k	0.0-2.0	18	2000	19835.63	14679.15	1.10	3.24	2.13	0.34
	2.0-4.0	27	1300	21774.47	15518.40	1.41	3.27	1.86	0.43
	4.0-6.0	18	1700	25179.77	25179.77	1.17	2.52	1.35	0.46
	6.0-8.0	9	3500	12840.27	9210.38	0.84	2.63	1.78	0.32
	8.0-10.0	11	2000	19715.73	19715.73	1.33	2.63	1.30	0.51
	10.0-12.0	12	5000	31924.50	31683.95	0.61	2.46	1.85	0.25
	12.0-14.0	8	20000	83368.41	79272.85	1.06	2.81	1.75	0.38
	14.0-16.0	3	50000	No fit					
	16.0-18.0	2	100000	No fit					
	18.0-20.0	0	100000	No fit					
ks	0.0-2.0	10	5000	No fit					
	2.0-4.0	9	3500	20050.06	12328.64	1.34	2.76	1.42	0.48
	4.0-6.0	15	2000	No fit					
	6.0-8.0	10	3000	11488.49	11132.30	0.66	1.90	1.23	0.35
	8.0-10.0	5	25000	No fit					
	10.0-12.0	3	100000	No fit					
	12.0-14.0	2	100000	No fit					
	14.0-16.0	0	100000	No fit					
ks2	0.0-2.0	13	4000	No fit					
	2.0-4.0	13	3000	7339.21	2446.40	1.32	1.92	0.60	0.69
	4.0-6.0	6	25000	154787.08	77393.54	3.11	4.47	1.36	0.70
	6.0-8.0	9	8000	21456.71	10728.35	1.22	1.45	0.23	0.84
	8.0-10.0	3	50000	No fit					
	10.0-12.0	3	100000	0	0	0	0.89	0.89	0
	12.0-14.0	2	100000	No fit					
	14.0-16.0	0	100000	No fit					
kh	0.0-2.0	18	2000	31551.88	15775.94	1.62	3.71	2.09	0.44
	2.0-4.0	18	2000	21175.96	16001.40	1.52	3.29	1.78	0.46
	4.0-6.0	12	600	3748.69	2571.10	0.73	1.97	1.24	0.37
	6.0-8.0	5	4000	15897.64	15897.64	2.14	2.78	0.64	0.77
	8.0-10.0	10	6000	11410.50	9390.95	0.83	2.34	1.51	0.36
	10.0-12.0	4	25000	45266.85	36983.38	0.30	2.05	1.75	0.15
	12.0-14.0	3	80000	No fit					
	14.0-16.0	2	100000	No fit					
	16.0-18.0	0	100000	No fit					
	18.0-20.0	2	100000	No fit					
kh1	0.0-2.0	20	1500	No fit					
	2.0-4.0	27	1500	24827.71	17677.55	1.28	2.40	1.12	0.53
	4.0-6.0	10	600	4705.61	4705.61	0.97	1.48	0.51	0.65
	6.0-8.0	6	5000	17641.22	10590.54	1.17	1.92	0.74	0.61
	8.0-10.0	9	22000	63376.18	63376.18	0.63	1.95	1.31	0.32
	10.0-12.0	3	40000	78672.18	78672.18	0.93	2.31	1.38	0.40
	12.0-14.0	4	50000	No fit					
	14.0-16.0	2	100000	No fit					
	16.0-18.0	0	100000	No fit					
	0.0-2.0	10	5000	19043.70	16616.45	0.08	0.44	0.36	0.19
	2.0-4.0	33	3000	4167.31	4167.31	0.03	0.42	0.39	0.08

Continued on next page

Table C.1 – continued from previous page

Soil	Depth range [m]	Nr. of points N	Bin width Δd	Eff. range r' [m]	Range r [m]	P. sill p	Sill s	Nugget n	Ratio p/s
	4.0-6.0	17	3000	13942.10	13681.84	0.07	0.33	0.26	0.21
	6.0-8.0	10	6000	8386.69	8386.69	0.18	0.34	0.17	0.51
	8.0-10.0	10	20000	29688.84	29688.84	0.19	0.33	0.14	0.58
	10.0-12.0	4	50000	No fit					
	12.0-14.0	3	100000	No fit					
	14.0-16.0	2	100000	No fit					
	16.0-18.0	0	100000	No fit					
	0.0-2.0	6	800	3997.38	3997.38	0.27	0.53	0.26	0.51
	2.0-4.0	4	500	899.20	899.20	0.18	0.32	0.14	0.57
	4.0-6.0	7	2400	9111.23	6964.58	0.09	0.34	0.25	0.26
vk	6.0-8.0	4	12000	22651.03	18802.39	0.12	0.45	0.33	0.27
	8.0-10.0	3	40000	No fit					
	10.0-12.0	1	100000	No fit					
	12.0-14.0	0	100000	No fit					
	0.0-2.0	29	1700	No fit					
	2.0-4.0	5	1000	2190.08	1749.69	0.14	0.26	0.12	0.54
	4.0-6.0	6	3500	15202.29	7601.14	0.08	0.13	0.05	0.62
vk1	6.0-8.0	3	12000	12496.83	10048.42	0.24	0.48	0.24	0.51
	8.0-10.0	1	100000	No fit					
	10.0-12.0	0	100000	No fit					
	12.0-14.0	1	100000	No fit					

C.1.2. Vertical direction Rivierenland Dataset

Table C.2: Variogram parameters for vertical direction in the Rivierenland (RL) database for the volumetric weight.

Soil	Number of points N	Bin width Δd	Effective range r' [m]	Range r [m]	Partial sill p	Sill s	Nugget n	Ratio p/s
k	1226	0.50	6.03	6.03	1.08	2.50	1.42	0.43
ks	584	1	5.24	2.62	0.68	1.50	0.82	0.45
ks2	285	2	5.15	2.58	0.85	1.33	0.48	0.64
kh	484	2	5.32	3.82	0.76	1.84	1.08	0.41
kh1	206	2	0	0	0	0.95	0.95	0
v	989	0.50	1.27	1.27	0.02	0.18	0.16	0.11
vk	240	5	No fit					
vk1	198	5	No fit					

C.2. SHANSEP Parameter S

C.2.1. Results Depth Range Analysis

Table C.4: Average variogram parameters over the different depth ranges in the STOWA database for the SHANSEP parameter S.

Soil	Average effective range r' [m]	Average range r [m]	Average partial sill p	Average sill s	Average nugget n	Ratio p_{avg}/s_{avg}
k	28695.29	26996.39	0.02	0.09	0.07	0.18
kh	9471.26	7938.10	0.00	0.08	0.08	0.05
kh1	1935.16	1659.72	0.01	0.02	0.01	0.40
v	6931.79	6594.66	0.03	0.03	0.00	0.92

Table C.6: Variogram parameters in the Applied Water Management Research Foundation (Stichting Toegepast Onderzoek Waterbeheer) (STOWA) database for the SHANSEP parameter S in the vertical direction and for specific depth ranges.

Soil	Depth range [m]	Nr. of points N	Bin width Δd	Eff. range r' [m]	Range r [m]	P. sill p	Sill s	Nugget n	Ratio p/s
k	0.0-2.0	10	6000	30282.01	30282.01	0.05	0.18	0.13	0.27
	2.0-4.0	6	17000	0.58	0.58	0.00	0.08	0.08	0.00
	4.0-6.0	5	20000	58139.28	58139.28	0.01	0.07	0.06	0.11
	6.0-8.0	10	10000	26359.27	19563.68	0.01	0.01	0.01	0.52
	8.0-10.0	7	30000	No fit					
	10.0-12.0	5	30000	No fit					
	12.0-14.0	2	200000	No fit					
	14.0-16.0	0	100000	No fit					
	16.0-18.0	2	100000	No fit					
kh	0.0-4.0	7	6000	0	0	0	0.15	0.15	0
	4.0-8.0	6	17000	No fit					
	8.0-12.0	5	17000	18942.53	15876.19	0.01	0.01	0.00	0.73
	12.0-16.0	2	200000	No fit					
	16.0-20.0	0	200000	No fit					
kh1	0.0-4.0	10	1000	No fit					
	4.0-8.0	4	1000	1935.16	1659.72	0.01	0.02	0.01	0.40
	8.0-12.0	4	60000	No fit					
	12.0-16.0	1	200000	No fit					
v	0.0-4.0	16	1000	6931.79	6594.66	0.03	0.03	0.00	0.92
	4.0-8.0	3	50000	No fit					
	8.0-12.0	3	100000	No fit					
	12.0-16.0	2	200000	No fit					

C.2.2. Vertical direction STOWA Dataset

Table C.7: Variogram parameters for the vertical direction in the Applied Water Management Research Foundation (Stichting Toegepast Onderzoek Waterbeheer) (STOWA) database for the SHANSEP parameter S.

Soil	Number of points N	Bin width Δd	Effective range r' [m]	Range r [m]	Partial sill p	Sill s	Nugget n	Ratio p/s
k	1142	1	0	0	0	0.03	0.03	0
ks	377	1	No fit					
ks2	154	1	0	0	0	0.02	0.02	0
kh	997	1	0	0	0	0.04	0.04	0
kh1	607	1	No fit					
v	340	1	0	0	0	0.00	0.00	0
vk	28	1	No fit					
vk1	15	3	No fit					

D

Data Case Study

Table D.1: Clay data from the STOWA database for water board Brabantse Delta which is used for the case study.

Boring ID	γ_{wet} [$\mathrm{kN/m^3}$]	S [-]	ϕ [°]	x-coordinate	y-coordinate
B101_St7	17.8	0.480	30.44	76464.3	404735.8
B101_St14	18.1	0.331	34.32	76464.3	404735.8
B101_St15	18.3	0.384	37.95	76464.3	404735.8
B101_St15-2	20.1	0.672	34.57	76464.3	404735.8
B101_St26	19.5	0.792	34.34	76464.3	404735.8
B101_St27	18.5	0.342	33.96	76464.3	404735.8
B102_St15-2	19.6	0.771	34.49	76457.8	404734.2
B102_St15-1	18.7	0.634	32.68	76457.8	404734.2
B102_St30	17	0.342	34.49	76457.8	404734.2
B102_St31	14.5	0.232	29.88	76457.8	404734.2
B103_St11	17.9	0.281	34.39	76440.5	404716
B103_St13-1	17	0.303	34.23	76440.5	404716
B103_St13-2	16.3	0.327	39.99	76440.5	404716
B103_St14	15.5	0.232	36.04	76440.5	404716
B103_St15	15.9	0.262	37.77	76440.5	404716
B103_St16	16.6	0.296	36.27	76440.5	404716
B103_St20	19.9	0.247	21.54	76440.5	404716
B103_St21	20	0.371	30.79	76440.5	404716
B104_St13-1	17.4	0.338	36.39	76443.7	404714.3
B104_St13-2	16.8	0.251	32.46	76443.7	404714.3
B104_St14	17	0.225	35.31	76443.7	404714.3
B104_St16	16.7	0.384	38.07	76443.7	404714.3
B104_St21	19.1	0.352	29.59	76443.7	404714.3
B104_St22	20.1	0.249	21.82	76443.7	404714.3
B2_St13	14.1	0.279	46.69	91240.5	411132.6
B2_St14	14.5	0.235	33.83	91240.5	411132.6
B2_St15	15	0.314	38.56	91240.5	411132.6
B201_St7	17.9	0.318	29.02	90445.5	411545.8
B201_St9	17.5	0.286	29.39	90445.5	411545.8
B201_St14	16.3	0.313	34.24	90445.5	411545.8
B201_St15	15.7	0.249	34.13	90445.5	411545.8
B201_St24	16.4	0.305	36.11	90445.5	411545.8
B201_St30	16	0.296	36.76	90445.5	411545.8
B201_St32	17.2	0.280	35.89	90445.5	411545.8
B202_St9-2	18.8	0.405	31.79	90448.8	411543.8

Continued on next page

Table D.1 – continued from previous page

Boring ID	γ_{wet} [$\mathrm{kN/m^3}$]	S [-]	ϕ [°]	x-coordinate	y-coordinate
B202_St9	18	0.340	31.56	90448.8	411543.8
B202_St14	15.7	0.327	39.73	90448.8	411543.8
B202_St17-2	16.1	0.295	34.85	90448.8	411543.8
B202_St17	16.2	0.333	38.10	90448.8	411543.8
B202_St23	16.2	0.292	34.54	90448.8	411543.8
B202_St24	16.7	0.374	35.81	90448.8	411543.8
B203_St7-1	16	0.343	36.63	90426.4	411511
B203_St7-2	15.8	0.291	35.40	90426.4	411511
B203_St8-2	16.5	0.390	37.10	90426.4	411511
B203_St8-1	15.7	0.312	37.51	90426.4	411511
B203_St15-2	14.3	0.283	37.82	90426.4	411511
B203_St15	14.3	0.288	39.77	90426.4	411511
B203_St22	15.4	0.301	36.76	90426.4	411511
B204_St6-1	17.3	0.513	40.83	90426.4	411511
B204_St6-2	15.5	0.309	34.18	90426.4	411511
B204_St8	15.3	0.324	36.74	90426.4	411511
B204_St15	13.9	0.289	40.96	90426.4	411511
B301_St6	18.3	0.386	31.65	95103.9	410439.4
B301_St11	16.3	0.289	34.00	95103.9	410439.4
B302_St13-2	18.2	0.426	34.11	95107.9	41039.5
B302_St13	17.9	0.436	34.88	95107.9	41039.5
B302_St15	16.8	0.318	34.76	95107.9	41039.5
B302_St16-2	16.6	0.313	34.80	95107.9	41039.5
B302_St18-2	18.9	0.691	33.46	95107.9	41039.5
B302_St19	18	0.369	33.80	95107.9	41039.5
B302_St20	17.3	0.340	37.92	95107.9	41039.5
B302_St23	15.4	0.286	33.94	95107.9	41039.5
B303_St5-1	16.1	0.300	36.64	95104.9	410406
B303_St5-2	16.5	0.414	38.21	95104.9	410406
B303_St6	17.5	0.441	42.05	95104.9	410406
B304_St4	16.6	0.370	36.60	95108.6	410406.2
B304_St5	15.5	0.382	37.44	95108.6	410406.2
B401_St19	15.6	0.362	38.91	107658.9	414174.1
B401_St20-1	15.8	0.349	36.42	107658.9	414174.1
B401_St20-2	15.8	0.335	37.41	107658.9	414174.1
B401_St21-2	14.6	0.321	37.26	107658.9	414174.1
B402_St17-2	16.2	0.372	37.81	107661.3	414177.3
B402_St17	15.5	0.318	38.46	107661.3	414177.3
B601_St6	18	0.262	25.72	121474	414020.3
B601_St8-2	20	0.470	34.31	121474	414020.3
B601_St8	19.7	0.502	34.38	121474	414020.3
B601_St10	19.8	0.551	33.90	121474	414020.3
B601_St11	19.8	0.528	33.13	121474	414020.3
B602_St5A	18.3	0.285	25.05	121477.4	414017.9
B602_St5	18.2	0.204	20.35	121477.4	414017.9
B602_St8	19.8	0.619	33.18	121477.4	414017.9
B602_St8A	19	0.604	34.82	121477.4	414017.9
B602_St13-2	18.4	0.360	36.09	121477.4	414017.9
B602_St13-1	18.4	0.357	33.16	121477.4	414017.9
B701_St8	18	0.370	32.93	115758.7	413046
B701_St9	18.8	0.596	34.62	115758.7	413046
B701_St13-2	16.4	0.347	38.79	115758.7	413046
B701_St13-1	16.7	0.361	37.85	115758.7	413046

Continued on next page

Table D.1 – continued from previous page

Boring ID	γ_{wet} [$\mathrm{kN/m^3}$]	S [-]	ϕ [°]	x-coordinate	y-coordinate
B702_St5	16.6	0.416	37.53	115766.7	413045.7
B702_St9-2	18.5	0.414	33.64	115766.7	413045.7
B702_St9	17.7	0.429	34.53	115766.7	413045.7
B702_St13-2	15.7	0.339	38.46	115766.7	413045.7
B702_St13-1	14.6	0.357	40.73	115766.7	413045.7

Table D.2: Peat data from the STOWA database for water board Brabantse Delta which is used for the case study.

Boring ID	γ_{wet} [$\mathrm{kN/m^3}$]	S [-]	ϕ [°]	x-coordinate	y-coordinate
B102_St24	11.0	0.473	38.50	76457.8	404734.2
B103_St9-2	10.2	0.537	53.54	76440.5	404716
B103_St9-3	10.1	0.519	50.10	76440.5	404716
B104_St10-2	10.1	0.492	45.17	76443.7	404714.3
B104_St10-1	10.2	0.505	55.95	76443.7	404714.3
B201_St26-1	10.4	0.423	34.03	90445.5	411545.8
B201_St26-2	10.4	0.446	35.37	90445.5	411545.8
B201_St26-3	10.4	0.425	34.45	90445.5	411545.8
B201_St26-4	10.3	0.470	36.37	90445.5	411545.8
B202_St26	10.6	0.403	34.02	90448.8	411543.8
B202_St27	10.3	0.414	34.24	90448.8	411543.8
B203_St10-1	10.1	0.633	58.72	90426.4	411511
B203_St10-2	10.4	0.536	52.06	90426.4	411511
B203_St25	10.4	0.479	46.43	90426.4	411511
B204_St10-1	10.2	0.489	42.97	90426.4	411511
B204_St10-2	10.0	0.434	35.40	90426.4	411511
B204_St25	10.4	0.469	52.83	90426.4	411511
B301_St27-1	10.6	0.460	41.82	95103.9	410439.4
B301_St27-2	10.6	0.455	38.87	95103.9	410439.4
B302_St27-1	10.9	0.448	43.78	95107.9	41039.5
B302_St27-2	10.7	0.512	46.93	95107.9	41039.5
B302_St30-1	11.4	0.357	39.39	95107.9	41039.5
B302_St30-2	12.1	0.347	38.73	95107.9	41039.5
B303_St9-1	10.3	0.583	59.31	95104.9	410406
B303_St9-2	10.2	0.483	43.65	95104.9	410406
B303_St9-3	10.0	0.546	42.81	95104.9	410406
B303_St10	10.0	0.389	32.02	95104.9	410406
B303_St11	10.5	0.448	41.28	95104.9	410406
B304_St10	10.2	0.479	35.25	95108.6	410406.2
B304_St11-3	10.1	0.563	60.70	95108.6	410406.2
B401_St22	9.8	0.520	39.50	107658.9	414174.1
B401_St23-1	9.9	0.489	39.67	107658.9	414174.1
B401_St23-2	9.9	0.478	39.50	107658.9	414174.1
B402_St20	9.6	0.497	41.04	107661.3	414177.3
B402_St21-1	10.3	0.465	39.47	107661.3	414177.3
B402_St21-2	10.3	0.501	41.75	107661.3	414177.3
B403_St7-1	10.0	0.505	45.14	107695.8	414150.1
B403_St7-2	10.0	0.582	51.16	107695.8	414150.1
B403_St8-2	9.8	0.561	52.58	107695.8	414150.1
B404_St8-2	9.7	0.546	53.82	107698.1	414153.3
B404_St8-3	10.1	0.535	47.45	107698.1	414153.3
B404_St8-1	10.1	0.545	47.83	107698.1	414153.3