MSC THESIS

Influence of measurement density in the flood risk assessment for piping

A theoretical analysis of the accuracy of strength assessments

J.M. van Stokkum

https://beeldbank.rws.nl, Rijkswaterstaat / Ron Boelens

2016, July 1st

Title:	Influence of measurement density in the flood risk assessment for piping; *A theoretical analysis of the accuracy of strength assessments*
Date:	July 1st 2016
Location:	Deventer
Author :	J.M. van Stokkum julius.van.stokkum@bzim.nl
University:	University of Twente Faculty of Engineering Technology Civil Engineering & Management/Water Engineering & Management Post box 217 7400 AE Enschede
Company:	BZ Innovatiemanagement BV Zutphenseweg 51 Post box 445 7400 AK Deventer
Graduation committee:	Dr. K.M. Wijnberg, University of Twente Dr. J.J. Warmink, University of Twente ing. W.S. Zomer MSc, BZ Innovatiemanagement BV

Preface

In this document you find the final report of my master thesis about *measurement densities in piping assessments*. It provides the results of a research trajectory of about 1 year to finalize my master Water Management & Engineering and consequently my time as student. This wonderful period started back in 2009 at the Technical University of Delft and via the University of Linkoping it ends 7 years later in Twente, back at my roots.

The last year has already been a transition from a student to working life. In September 2015 I started with my research at BZ Innovatiemanagement in Deventer. Soon the graduation process was combined with two days of work on other projects of BZ as well. This combination was a challenge and has resulted in a couple of months delay. However, it also offered me the opportunity to learn a lot about working as an advisor and the water safety sector in general. The extra lead time made it also possible to put the research results in a better perspective. All together I consider my period at BZ as very valuable and want to thank my colleagues for this. First of all, my daily supervisor Wouter Zomer. Thank you for initiating the research topic, familiarizing me with the sector, help with defining the scope and underlying problem area, feedback, discussions and daily companion at the office. Most of all I want to thank you for helping me back on track when I had a major 'graduation depression'. I also want to thank colleagues Sander Bakkenist, Caspar ter Brake and David Barmentloo. Your presence at the office was both fun and inspiring.

Furthermore, I want to thank my UT supervisors Jord Warmink and Kathelijne Wijnberg. You both helped me to make this research much better and of an academic worthy level. Besides that, I want to highlight the great attitude of you during the entire graduation process. Although critically regarding the content, it was very easy making appointments and ask for your help. Especially Jord, your help meant a grade deal during my 'graduation depression'.

Finally, I can't and won't forget to thank my girlfriend Moniek. Thank you for all the effort in reading and revising my report. I am very grateful you did that and helped me get through the tough moments.

What is left for now is to wish you pleasure reading the results of my research. I hope this research will be helpful in the challenges of protecting the Netherlands effectively against flood risks.

Summary

Piping is a failure mechanism in which a water flow below a cohesive dike body causes erosion of soil particles. This can result in failure of the dike. In periodic safety assessments, the probability of failure due to piping is calculated and compared with the legal norms. In practice the so called *detailed* assessment is often used to estimate the piping safety in a semi-probabilistic approach. A model of Sellmeijer is used to calculate the water level that is critical with respect to piping. Most important parameters in the Sellmeijer models describe soil properties. In literature is found that piping is one of the most critical failure modes of dikes, largely influenced by uncertainties in those soil properties.

Parameter values of the Sellmeijer model that are representative for a considered dike have to be estimated with use of measurements. Usually only a small number of local measurements is collected or available. Because of uncertainties about the (spatial) distribution of the parameters, a number of assumptions is made to make the strength assessment conservative. This is done to prevent unsafe situations in which the actual strength is smaller than the calculated resistance. Counterpart is that due to conservatism dikes can be rejected and reinforced while they had been meeting the norm if very detailed information was available. Hypothesis is that higher measurement densities help to improve the assessments. The goal of this research is to quantify the effect of the density of point measurements on the accuracy of the strength assessment.

A simulation is set up which enables a comparison between a reference strength and an assessed strength that is based on point measurements. A theoretical analysis is made possible by modelling soil properties as random fields. Monte Carlo is used to compare reference and assessed strength for uncertain soil conditions.

The resistance of a dike to piping is varying along the length as a consequence of varying soil properties. The cross section of a dike section with the lowest resistance is of main interest and therefore defined as the reference strength. The assessed strength follows from characteristics value calculations based on uncertain measurement data. The error in a strength assessment is defined as the difference between reference and assessment. The magnitude of the error is dependent on the soil conditions. Because the soil conditions under a dike are uncertain, the error has a range of possible values. It is shown that this range of errors exist for every measurement density but can be decreased to some extent. The density till which the range can be decreased depends on the (unknown) variation pattern of the soil and the (unknown) measurement error. It is therefore not possible to conclude about an optimal measurement density without additional sources of information.

It is also shown that the (unknown) variation pattern and measurement error determine the range of errors and the probability that an assessment underestimates or overestimates the actual strength. It is shown that for constant variation patterns the assessment is generally conservative as it intends to be. This means that it is not likely that the probability of failure is actually higher than calculated from the assessment. It is however likely that a dike is actually stronger than calculated. This can result in ineffective reinforcements. When it is expected that a dike is much safer than calculated, an increasing measurement density can favour the assessment outcome. The assessed strength becomes generally less conservative if the measurement density is increased. In some scenario's this corresponds to an increased probability of overestimation as well. If an assessment overestimates the actual strength, it is possible that the dike fails before the normative load is exceeded.

In this research is shown that the current assessment approach is not generally conservative. For some scenario's it is well possible that the actual strength is smaller than calculated. It is identified that sudden weak spots are easily overlooked by the current approach. Increasing the measurement density has not always the intended effect of preventing this danger.

In the assessment measurements are used to estimate parameter distributions. From those distributions characteristic values are determined for which it is statistically likely that actual values in the field at least result in a smaller piping risk. More measurements help to describe the distribution of a parameter more accurately. The way the representative parameter value is selected from that distribution is however static, i.e. the method for determining a representative value from a data set is not influenced by measurement density. Therefore, still an error can be present between the estimated representative parameter value and the value at the weakest cross section. The density of measurements has no influence to this type of error. This means a range of errors exist for every measurement density, even for very high measurement densities.

It is shown that if the actual strength is underestimated, this is often because unfavourable properties do not coincide while assumed they do. If the actual strength is underestimated, it is often because outliers are not captured by the assumed parameter distributions. By considering spatial variation explicitly, it is theoretically possible to get rid of inaccuracies by increasing measurement densities. It is shown that the density of measurements needs to be higher than the typical correlation length of properties to make sure the actual strength is not overestimated. Note that this only holds if measurement error is absent or very small. It is shown that errors due to measurement error can be reduced by averaging multiple measurements per location.

It is concluded the current strength assessment based on point measurements and characteristic parameter values is unreliable for every measurement density. With a couple of measurements and unknown soil variation patterns, the error in an assessment can in theory be up to almost 100%. It is also concluded that due to static assumptions in the detailed assessment, relative errors can still be up to 50% at high measurement densities. In theory measurements might contain all information to make reliable assessments when spatial distribution is made explicit. But it is concluded that the required point measurement density is not feasible in practice. Because of the possible errors in an assessment based on point measurements, it is recommended to use additional sources of information in the assessment as well. It is recommended to study the possibility of (applying) volume covering measurements as it might provide information about variation patterns and identify possible weak spots.

Where literature has been focussed on improved understanding and modelling of piping, this research has specifically highlighted the errors that can be made if model input is inadequate. It also shows that the use of point measurements inevitable results in uncertain model input. Therefore, this research supports existing initiatives in which is argued for the need of more data and information and/or alternative data and information in safety assessments. Additional data and information is needed to prevent unreliable assessments. An unreliable assessment can result in ineffective reinforcements or safety risks that are higher than accepted.

Content

List of definitions

Term/symbol	Explanation
Piping/backward erosion	The formation of empty spaces/pipes below a water retaining structure resulting as the result of a seepage flow in which sediments are transported.
d70 realisation	The spatial distribution of d70 values along the dike section. A d70 realisation is randomly acquired by input statistics mean, variance and correlation length.
k realisation	The spatial distribution of k values along the dike section. A k realisation is randomly acquired by input statistics mean, variance and correlation length.
Strength realisation/ H_c-realisation	The spatial distribution of the critical head (resistance against piping) along the dike section. The H_c realisation is acquired by combining the varying parameter realisations (d70 and k) in the model of Sellmeijer (2011).
Reference strength/ $H_{c,ref}$	The minimum value of a H_c realisation. The reference strength is the resistance to piping of a dike section as it presents the strength at the weakest cross section.
Characteristic d70 value/ p_{d70}	The theoretical 95% lower boundary of the d70 distribution, assumed to be the representative d70 value of a dike section.
Characteristic k value/ p_k	The theoretical 95% upper boundary of the k distribution, assumed to be the representative k value of a dike section.
Assessed strength/ $H_{c,assessed}$	The H_c value that is calculated with characteristic parameter values (d70 and k) using the model of Sellmeijer (2011). The assessed strength is a single critical head value that is assumed to be representative for a dike section.
Error/ Δ_x	The difference between reference strength and assessed strength as ratio of the seepage length L. The error is expressed in the deviation of critical slope and therefore has the unity m/m or dimensionless.
Accurate strength assessment/ $\Delta_x \approx 0$	The assessed strength is (almost) equal to the reference strength: the error is (close to) zero.
Underestimation $\Delta_x > 0$	Positive error: the assessed strength is lower than the reference strength.
Conservative assessment/ $\Delta_x > 0$	Positive error: probability that a dike section is rejected while it actually meets the norm
Overestimation/ $\Delta_x < 0$	Negative error: the reference strength is lower than the assessed strength.
Unsafe assessment/ $\Delta_x < 0$	Negative error: probability that a dike section is considered safe after the assessment while it actually does not meet the norm.

1. Introduction

This chapter introduces the research topic and summarizes it into a problem statement. The chapter is finalized with the goal and research questions of this research.

1.1. Background

Nowadays, the safety standards for primary flood defences have a legal status in the Netherlands, defined in the 1995 Flood Defence Act (in Dutch: Wet op de waterkering). The compliance of the defences to the safety standards needs to be verified by the water boards through periodic safety assessments (Schweckendiek, 2014). The legal procedures of the periodic assessment are described in the 'Voorschrift Toets op Veiligheid' (VTV) (in English 'Regulation Safety Assessment'). Currently the VTV2006 is still valid, but the new VTV (VTV2017) is under development.

In a safety assessment the strength of the dikes is determined in relation to the several failure mechanisms and compared to the norms. Non-complying defences are to be reinforced. Piping is one of the failure mechanism for which the safety is assessed.

1.1.1. The piping failure mechanism

The definition of piping is 'the formation of empty spaces below a water retaining structure as the result of a concentrated seepage flow in which sediment particles are transported' (Förster, van den Ham, Calle, & Kruse, 2012). A dike fails due to piping in case a head difference causes a water flow below the dike that, in case the flow is strong enough, causes the soil particles to erode. The progressing erosion ultimately results in the collapse of the dike (Kanning & Calle, 2013)

Before a dike fails due to piping two conditions need to be satisfied: burst of the covering impermeable layer (called 'uplift') and pipes creating an open connection between the outside water and the inside area of the dike (Calle, van der Meer, & Niemeijer, 1999). A schematization of a developing pipe is given in Figure 1.

Additional information about the piping mechanism is provided in Appendix 1. For an extensive description of the piping mechanism is referred to chapter 4 of the report 'Zandmeevoerende Wellen' (Förster et al., 2012).

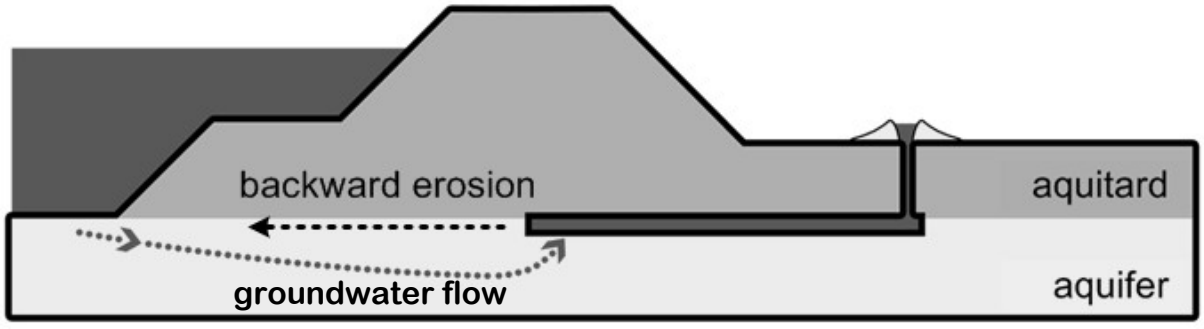


Figure 1 – illustration of the backward erosion/ piping process (Schweckendiek, 2014)

1.1.2. Assessment of piping risk

The principle of a safety assessment is as follows. Per considered dike section a limit state equation is worked out: $Z = R - L$. In this equation R is the resistance of the dike and L the load on the dike. The limit state defines the critical state for failure, i.e. the loads on the dike equal the strength (Aguilar-Lopez, Warmink, Schielen, & Hulscher, 2016).

The safety assessment consists of three types of assessments: simple, detailed and advanced. Within the safety assessment procedure, the so called 'detailed assessment' is most relevant (Förster et al., 2012). It uses a semi-probabilistic approach in comparing load and resistance. Resistance is determined using a Sellmeijer model. Sellmeijer calculates the resistance in terms of 'critical head difference' (H_c): the maximum water level difference over a dike that the dike can handle without piping to occur.

The input parameters to calculate the critical head are uncertain and can be statically described as random variables: the parameters have a range of values according to a probability density distribution. Use of random input variables results in an output (H_c) that is also a random variable. Combining this random resistance variable with a random load variable results in a probability of failure. This is referred to as probabilistic approach. The semi-probabilistic approach translates random input variables into deterministic input values. The result is a deterministic resistance value. The random load variable is also translated into a deterministic water level. Comparing the deterministic resistance with the deterministic load results in a statement about the possibility that load is higher than resistance. A partial safety factor is used in the determination of strength values to give this statement a certain reliability.

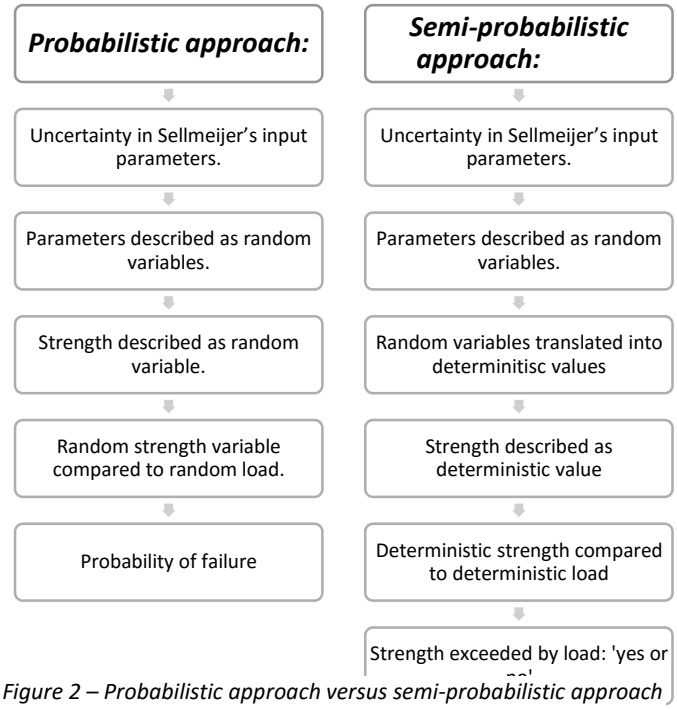


Figure 2 – Probabilistic approach versus semi-probabilistic approach in water safety assessments

The assessment specifies *how* the random variables are translated into deterministic values in a semi-probabilistic assessment. That procedure implies a certain safety margin that consist of a conservative *recipe* to estimate deterministic input parameters in combination with use of partial safety factors. The recipe for determination of input parameters is analysed in the 'problem definition' section.

More information about the procedure of the detailed assessment is given in Appendix 2. Background to the models (1989 and 2011) of Sellmeijer (including formulas) is provided in Appendix 3.

1.1.3. Soil

Piping takes place below the surface and is therefore very sensitive to ground conditions (Schweckendiek, 2014). Therefore, information about the soil is necessarily in assessing the risk of piping.

The most parameters of Sellmeijer's equations describe properties of the soil. The most important are the grainsize and permeability of the aquifer and the presence and depth of the covering clay layer (Kanning, 2012), (Schweckendiek, 2014).

Properties of naturally deposited soils generally exhibit considerable spatial fluctuations. Magnitudes of these fluctuations are often such that they may have significant effects on the design of geo-technical structures as dikes (Vrouwenvelder & Calle, 2003).

Within rather homogeneous deposits the uncertainty is called a continuous variability (see left stratification of Figure 3). Within rather homogeneous deposits minor geological details can cause variations in grainsizes and permeability. Since those are the main parameters, this would as well be very relevant for piping (Kanning, 2012).

Main deposits in the Netherlands are often intersected with anomalies such as old river gullies that later have filled with sediments (right stratification of Figure 3). High permeability or very fine grains in anomalies cause a sudden weak spot. Dike stretches may perform well for many kilometres but can fail at those weak spots (Kanning, 2012).

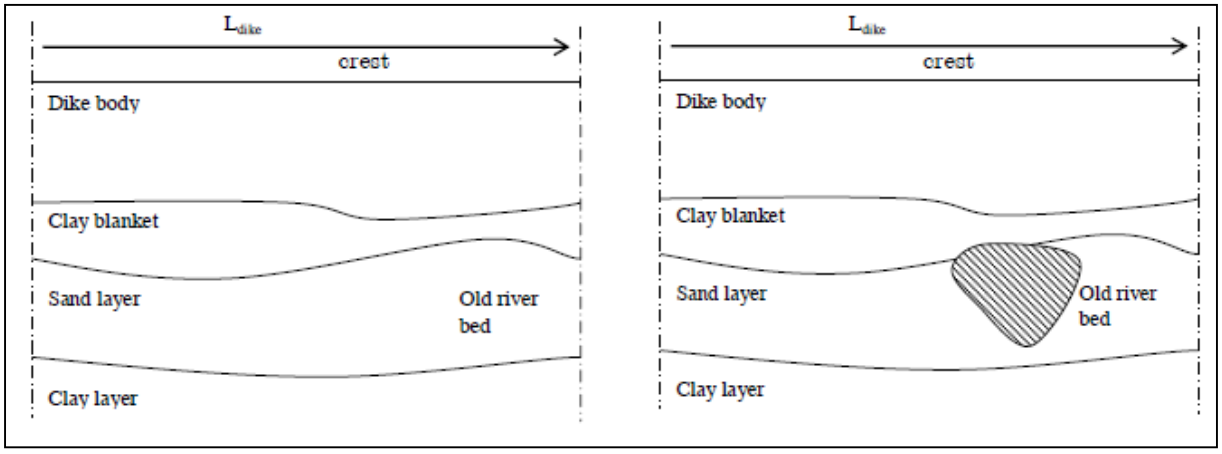


Figure 3 – Stratifications of the subsoil. Right a stratification with anomaly is presented (Kanning, 2012).

The probability that an anomaly is present in the considered dike stretch depends on the geological history. This probability can be examined with use of local data sets and geological expertise as was the approach in *Floris* (Wiersma and Vonhoegen, 2011).

1.1.4. Risk of piping in the Netherlands

Until now, the semi-probabilistic approach is used as standard in safety assessments. However, a full probabilistic approach is desirable to make flood risk analysis possible (Nicolai, Vrouwenvelder, Wojciechowska, & Steenbergen, 2013). Currently there is a transition towards the probabilistic approach, i.e. reliability analyses. The projects *Floris* (Flood Risks) and *VNK* (Veiligheid Nederland in Kaart) are examples of this and calculated the probabilities of failure of each dike ring in the Netherlands (Vergouwe, 2014).

From reliability analyses appeared that piping is one of the most critical failure modes of dikes. Piping reliability is largely influenced by uncertainties in ground conditions, especially by geo-hydrological properties such as the hydraulic conductivity of an aquifer. These properties are typically highly uncertain, which can lead to the related probabilities of failure being considerably high. (Schweckendiek & Vrouwenvelder, 2013)

According to the calculations many dikes in the Netherlands should be reinforced. This notice has increased the attention for this specific failure mechanism significantly and initiated new research and investigation towards the piping problem. In the next section the problem of uncertainties causing high risks is considered in more detail.

1.2. Problem description

The safety assessment of piping is subject to many uncertainties. The uncertainties related to resistance are of interest in this research and can be subdivided in model uncertainty and model input uncertainty.

1.2.1. Errors due to schematization and modelling

Models introduce uncertainty as they are only a simplified representation of the reality. A model can be inaccurate resulting in model bias, but a model can also be imprecise resulting in a model standard deviation. Model uncertainty can only be made explicit by comparing model predications with actual observed behaviour.

A critical head calculation with Sellmeijer has to be applied on cross section level (see Figure 4). The Sellmeijer models are developed and tested for rather homogeneous subsoils. The use of that model therefore implies uniform soil properties within the cross section. In each cross section, parameters such as grainsize can only have a single value. Because single values for sand layer depth and clay layer depth are required, a simplified soil structure as in Figure 4 is implied.

It is plausible that the uniformity assumption causes errors as the actual properties in the field are heterogeneous. Literature as well as practical experience for example indicates that heterogeneity of grainsize in the cross section, increases resistance against piping due to parallel effects. This is an example of model bias. The estimated strength with the model is consistently incorrect because based on uniformity instead of heterogeneity.

Model uncertainty is beyond the scope of this research because within the detailed assessment the model is given and not influenced by choices of the assessor such as data use. The uncertainty related to model *input* is however influenced by choices and knowledge of the soil. This is explained in the following sections.

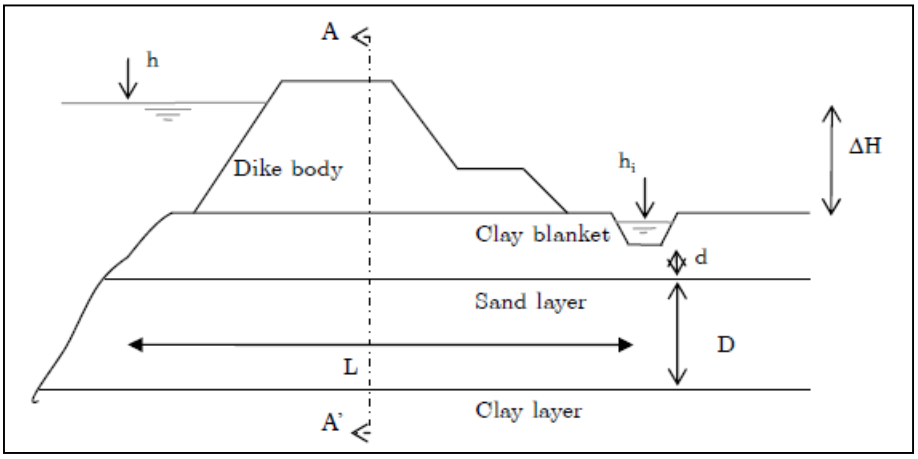


Figure 4 – Simplified cross section used in assessment, from Kanning (2012).

1.2.2. Errors due to estimation of representative parameter values

If a dike would be homogenous in length direction, it could be represented by one cross section. But because of different geometries and spatial variation of subsoil properties (see previous section), parameter values actually vary along the dike.

To be able to assess an entire dike ring with more reliable results. The dike ring, is divided into smaller sections for which at least the geometry is considered uniform. Each section is characterized with one representative cross section and the strength of this cross section is evaluated with Sellmeijer (Ministerie van Verkeer en Waterstaat, 2007). The calculated strength is assumed representative for the entire section. But obviously variations in strength are still present due to the heterogeneity of soil properties. As the dike is a series system, the representative cross section is ideally representing the weakest location of a particular dike section. However, the representative cross section is not certain but an estimate. And this estimate is mainly based on input from a couple of point measurements

(personal communication; Zomer, 2015). Point measurements provide only information about one specific position in the spatial domain and are subjected to measurement uncertainty.

A limited number of measurements introduces statistical uncertainty when estimating the probability density distribution of parameter values. A limited number of point measurements also introduces uncertainty about the spatial variation of parameter values. And furthermore measurement uncertainty is introduced as parameter values are translated from measurements.

Those three types of uncertainties leave the possibility that the properties at some location along the dike section are different then assumed in the representative cross section. An unsafe situation occurs if the assumed cross section is more favourable than the weakest location. Strength is then overestimated. As the length of sections increases, the probability that a weak location is present but not captured by the measurements and cross section calculations, increases. This is referred to as the length effect or series effect. There are two main causes of length-effect: continuous fluctuations in space due to natural variability in resistance parameters and discontinuous fluctuations due to local anomalies.

1.2.3. Limited information and dealing with uncertainties

To make a critical head calculation on a representative cross section, local information about the dike properties is needed. Information about the soil normally becomes available by taking soil measurements (sampling). The number of samples that is used to determine representative soil parameter values is usually small. The safety assessment however relies on the limited information from those samples. Therefore, the assessment uses a conservative approach to deal with the uncertainties and decrease the probability that strength is overestimated:

- Spatial distribution of parameter values is considered implicitly. This is referred to as the homogeneity assumption and a way to deal with uncertainty related to spatial variability. The most unfavourable parameter values with respect to piping are assumed to coincide at one location somewhere along the section. This approach provides no insight in the distribution of piping resistance along the dike section. Instead the strength of the entire section is assumed equal and represented by individual parameter values combined in one representative cross section.

- To quantify the unfavourable parameter values, strength parameters are described as random variables. Random variables are described by a distribution type, a mu and sigma. Most parameters are assumed to have a normal or lognormal distribution. Mu and sigma are calculated from the available samples. At least three samples are required. A *characteristic value calculation* finds the upper or lower representative parameter value from the considered probability distribution. This value is assumed conservative. The combination of conservative parameter values in the representative cross section should assure that the actual strength along the entire section is at least higher than the estimated strength.

- The estimated distribution type, mu and sigma are uncertain as long as the sample size is smaller than the population size. The characteristic value calculation therefore results in an uncertain upper or lower representative value (statistical uncertainty). To deal with statistical uncertainty a student-t factor is used. With a small number of data points, the sample mu and sigma have high probability to deviate from the population's mu and sigma. The student-t factor helps to find a representative value that is at least as conservative as when many data points where used. The student-t factor decreases if the number of used samples increases.

In appendix 4 the formulas to calculate characteristic values are provided. Notice that those classic statistical formulas are used under the assumption that measurements are as independent samples from an uncorrelated population. This means no spatial structure in measured values. However, this assumption does not fit the pattern of spatial distribution often visible in regional measurements sets (Calle, 2007)

- With each parameter assigned a conservative value, the critical head of the representative cross is calculated using a Sellmeijer model. The critical head is then divided by two safety factors. Those factors account for model and schematization uncertainties. But they also account for uncertainties in model input that are not explicitly covered by previous two bullets.

  One of those uncertainties is measurement error. It means that the estimated parameter from the measurement may not be perfectly representing the property at the measured location. Especially the estimation of permeability is often related to high uncertainties due to measurement error (personal communication; Kanning, 2015).

1.3. Problem statement

In summary the following problem is identified with respect to the design and execution of a detailed piping safety assessment:

- Previous collected data and information is often not stored or rarely used in (future) safety assessments.
- (Additional) data collection by measurements is expensive and often limited to the required minima. Technically the minimum number of measurements is three to be able to make a characteristic value calculation.
- Used measurements are characterized as point sources. They only provide information about a very small part of the subsoil.
- Consequence is that strength assessments in piping safety assessments are made based on a limited amount of information.
- To account for uncertainties arising from limited information, conservative assumptions and approaches are used to predict strength values.

Within this practice many kilometres of dike have been rejected the last years on behalf of insufficient safety with respect to piping. Experts doubt whether the calculated safety is a correct representation of the actual safety. It is already acknowledged that the assessment of piping is subjected to many uncertainties. The discussion is supported by practical cases which show that calculations and observed behaviour often mismatch. The use of a small amount of data is recognized as one possible explanation.

Usually only point measurements and cross-section calculations are available, whereas large uncertainties in the length-direction of the dike are present and possibly unaccounted for. Currently there is insufficient insight in the possible error of strength assessments as consequence of the existing uncertainties. It specifically lacks a quantitative analysis of the accuracy of assessments in relation to the quantity and quality of model input.

1.4. Goal

The goal of this research is to quantify the effect of the density of point measurements on the accuracy of the strength assessment.

The accuracy is determined by comparing strength assessments based on a number of point measurements (uncertain input) to a reference situation in which high quality sampling is available

(certain input). Sampling of high quality is then defined as an extreme dense collection of error free point measurements.

1.5. Research questions

To be able to achieve the goal of this research the following four research question were posed:

1. What is a representative model of a piping sensitive dike section?
   a) What is a representative geometry?
   b) What is are the representative (spatial) distribution of the important soil parameters?
2. How does the density of measurements influence the accuracy of strength assessments using the semi-probabilistic assessment approach?
3. How sensitive is the accuracy of strength assessments to assumed (spatial) distributions of soil parameters?
4. Is it possible to increase the accuracy of strength estimations by using information of point measurements alternatively?

Figure 5 provides a schematization on how the different research question relate to each other.

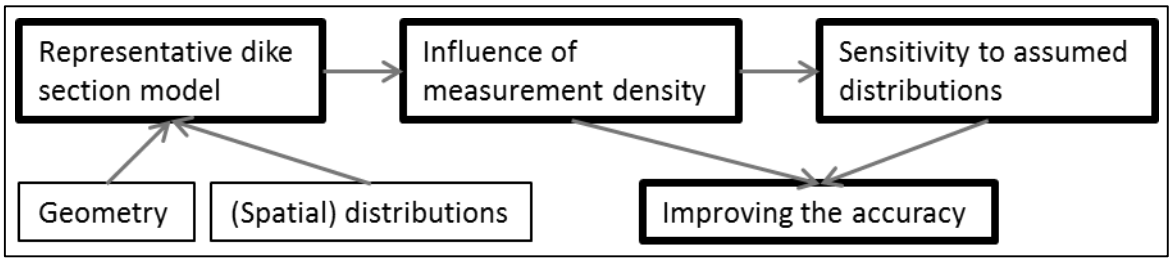


Figure 5 – Relation between research questions

2. Methods

The method used and assumptions made to be able to answer the research questions are described in this chapter. This chapter is subdivided in assumptions (section 2.1), approach (section 2.2) and scope (2.3).

A simulation is set up to study the accuracy of strength estimations in assessments. The simulation captures important aspects of the *detailed assessment* and enables a comparison between an assumed reference strength and an assessed strength. The reference strength is calculated with complete and known assessment input data. The calculation of the assessed strength is subject to the uncertainties as described in section 1.2 *problem definition.*

2.1. Assumptions of the simulation

The next subsections describe the most important principles and assumptions of the simulation. The assumptions correspond with assumptions made in actual strength assessments in the field.

2.1.1. The piping mechanism is captured by the calculation rule of Sellmeijer (2011).

As mentioned before, only the effect of model input on the accuracy of assessments is analysed. The model of Sellmeijer (2011) is used because it is the Dutch default. With Sellmeijer (2011) the strength is quantified in terms of *critical head* or *critical slope* and determined by the appearance of parameters mentioned in the formulas. The updated formulas in the Sellmeijer model (Sellmeijer et al., 2011) are used because those are, next to the original formulas (1989), already the regularly used and legally prescribed from 2017.

Each cross section is assumed to have uniform properties, which implies the research is restricted to a 1D analysis. This assumption is motivated by the use of a Sellmeijer model which requires or assumes a uniform cross section. In addition, the complexity of the simulation is limited. In the 1D analysis the dike section consists out of a large number of small consecutive subsections with homogeneous properties. To each individual subsection Sellmeijer (2011) is applicable. The subsections have certain properties. How those are distributed along the dike section is explained in section 2.2 *approach.*

2.1.2. The behaviour of individual parameters is statistically independent.

There is no cross correlation between parameters. For most parameters this assumption is logical, however the grainsize and permeability are physically connected. A layer with small grains has less cavities than a layer with bigger grains resulting in a smaller permeability. In Förster et al. (2012) is explained that only the small top layer of a water transporting layer below a cohesive soil layer is subjected to piping. Therefore, the relevant parameter is the grainsize of the layer directly below the cohesive dike body. Contrary the complete water transporting layer is affecting the piping mechanism. And consequently the average permeability of the entire sand layer is the relevant.

The assumptions on independence and uniformity in cross section, imply an aquifer layer of several sublayers. Each sublayer is uniform with respect to grainsize and permeability, but the properties per sublayer may vary (Figure 6). Then configurations are possible in which locations with small grainsizes (in the top sublayer) match with locations with high permeability's (average of entire layer). Or vice versa. Permeability is still measured as one value per cross section: the average permeability over all sublayers.

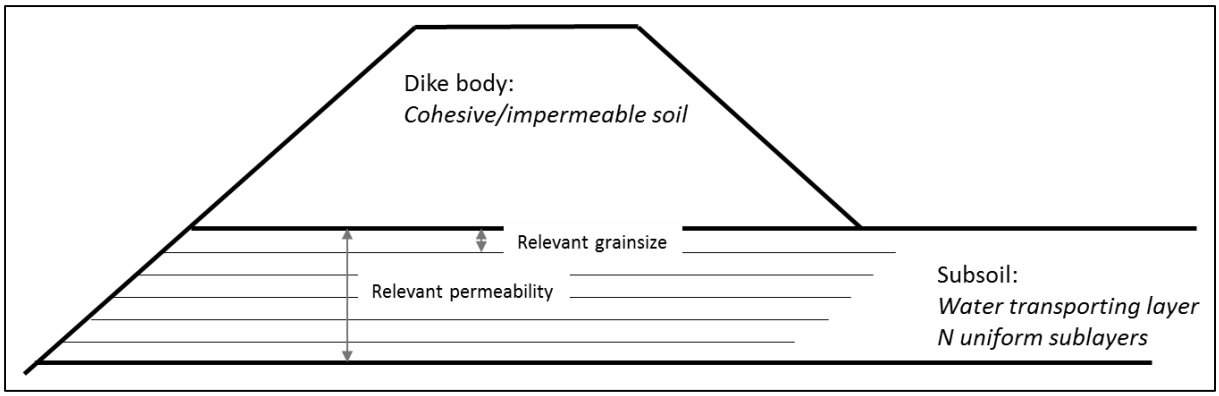


Figure 6 – Schematization of independence assumption

In piping reliability analyses, parameters are almost always modelled without cross-correlation. But Aguilar-López et al. (2016) demonstrate that cross correlation between relevant grainsize and relevant permeability is to some extend likely. The use of cross correlation decreases the probability of failure in full probabilistic approaches because there is a higher probability that small grainsizes are 'compensated' by small permeability's as well (and vice versa). In this study this would have increased the reference strength (see chapter 7 *discussion*).

In the assessment practices, it is even common to use the proposed relation between d70 and k to determine k values from grainsize measurements (Förster et al., 2012). The procedure however also mentions that to determine k, grainsize samples from the complete sand layer should be used. To determine d70 only samples from the top layer are required. This supports the assumption of this research that different layering within the sand layer introduces independence, possibly resulting in unfavourable combinations of k and d70.

2.1.3. Parameter can be estimated in-situ point measurements and can contain error

It is assumed that a measurement always measures the relevant parameter but not always the correct value belonging to that parameter on that specific location due to measurement error. The relevant parameter value at each subsection can be obtained by one measurement per parameter. Practices in which k is estimated from grainsize distributions are not covered in this research because it is in conflict with the previous assumption (section 2.1.2).

As mentioned before, the sand layer is assumed to consist out of several sub-layers. Only the d70 values of the top layer are represented in the simulation. This corresponds to a reality in which sieve samples to determine d70 are always taken from the top layer. In the simulation, the average permeability of the entire sand layer is represented. This corresponds to a reality in which the bulk permeability at one location is measured in situ with one measurement by for example pumping test. The result of such analyses is one value for the average permeability at the cross section where the test is carried out. That groundwater flow is a 3D process influenced by a 3D varying permeability is not relevant for the measured permeability at that exact location. It however is relevant for the correlation of k values between sub-sections. This is handled in chapter 3.

2.2. Approach

In this research a theoretical analysis is carried out with use of artificial data generation. Relevant soil parameters are modelled as random fields. This approach is also applied in the PhD theses of Kanning (2012) and Schweckendiek (2014). In literature it is a common way to model unknown or uncertain soil properties. Subsoil properties then show random fluctuations in space but to some extend

autocorrelation as well. Autocorrelation function (correlogram), mean and variance of soil parameters determine the behaviour in space. The correlation length indicates to which extend a parameter is fluctuating around its mean value. *The result of artificial data generation is shown in chapter 3.*

To quantify the accuracy of a strength assessment and the influence of measurement density, the following consecutive steps are performed in the simulation:

1. *Random field (RF) generation*. For each considered parameter a data realisation is made. A data realisation is a random data set with autocorrelation, according to a predefined correlogram, mean and variance. The realisation represents a possible spatial distribution of the parameter's values within a dike section. Each small subsection, as defined in section 2.1.1 is assigned a parameter value representing the property at that location.

2. *Application of the Sellmeijer (2011) model*. The different data realisations of model parameters are used as input in the Sellmeijer model. This results in a realisation of resistance values along the dike section. The strength at each location is expressed in terms of critical head (H_c).

3. *Determining the reference strength as the lowest critical head of a dike section*. The minimum value of the H_c realisation is called the reference strength. This is the H_c value that corresponds to the location of the dike where the combination of parameter values is most unfavourable with respect to piping resistance.

   Note: The reference strength is the estimated strength if certain model input would be available. The reference strength is not to be confused with actual strengths in reality. Actual strengths of dikes can normally only be determined by loading them. As mentioned this research is only a theoretical analysis.

4. *Estimating the strength with a semi-probabilistic approach*. First noise is added to the parameter realisations to represent measurement error. Then measurements are simulated by taking, for each individual parameter, a predefined number of values from the realisation with noise. With the measured values characteristic upper/lower values are calculated for each parameter. Those values are used in a representative cross section. Sellmeijer (2011) is applied on that representative cross section to obtain a representative critical head. This critical head is referred to as the *assessed strength*. An example calculation is given in *appendix 4*.

5. *The reference strength and assessed strength subtracted and divided by the seepage length*. This difference in critical slope is called the *error* (Δ_x) and is a measure for the performance/accuracy of the assessment, given the number of measurements x that is used in the assessment.

   Critical heads increase with increasing seepage lengths. Therefore, the value of L influences the calculated difference between assessed critical head and reference critical head. By defining accuracy as the difference between assessed and reference divided by *L*, the value of *L* is of less importance. There is only a small influence left, caused by non-linearity of L. By minimising the influence of dike width, the results become more generally applicable.

6. *The previous steps are applied in a Monte Carlo Simulation*. A large number of runs is made and in which randomly parameter realisations are generated. For each run the error is measured.

7. *All error's (one calculated error corresponding to one run) are statistically analysed*. The mean, standard deviation, minimum, maximum and 95% confidence interval of the errors are calculated.

8. *Previous steps are repeated for several measurement configurations.* This enables an analysis to the effect of different measurement densities.

In Figure 7 an overview of the simulation steps and the relation between them are given. In the simulation two loops are made. A Monte Carlo analysis is made to determine the error for many different soil configurations. A repetition is included to analyse the influence of measurements densities on the error. Result is a number of X error sets with each N values.

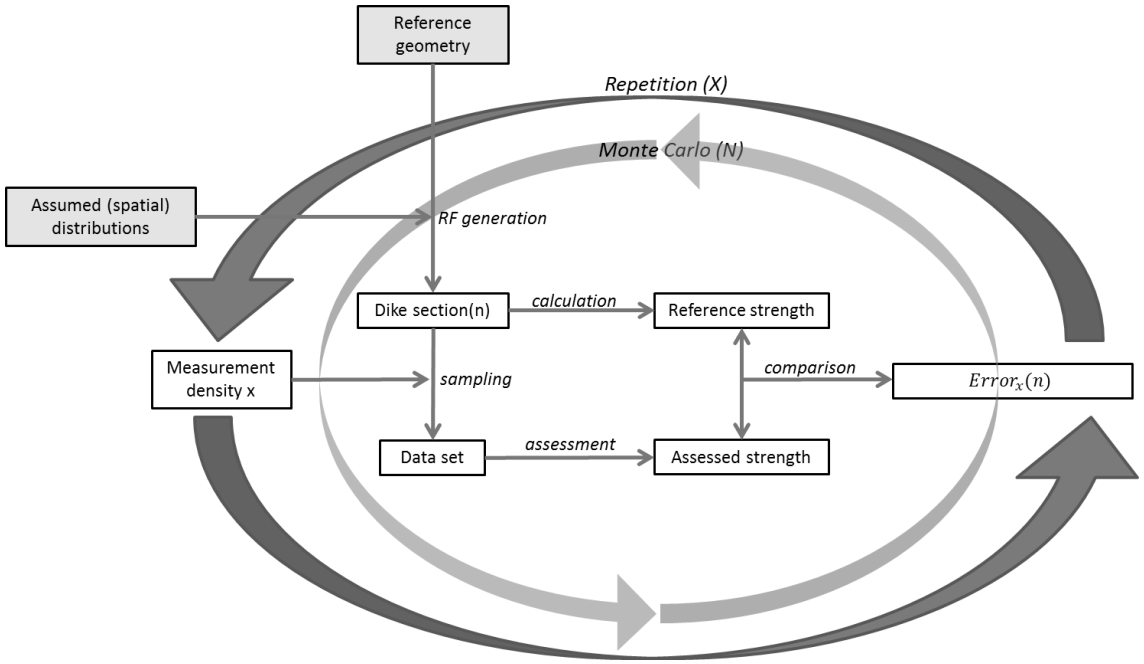


*Figure 7 – Overview of the different steps in the simulation. Start with a geometry and (spatial) distributions to obtain a dike section by RF generation. Given a measurement density a data set is obtained by sampling the dike section. From the dike section the reference strength is calculated. From the data set the assessed strength is determined. Both strengths are compared to find the error. N is the number of repetitions in the Monte Carlo analysis. X is the number of measurement densities for which the total MC analysis is repeated. In total X*N error values are obtained.*

2.3. Scope

The analysis of accuracy is subjected to several modelling choices. The following scope is used:

1. Geometry related parameters are fully correlated in length direction (uniform along the section). This uniformity is in practice quite easily reached when dividing a dike into dike sections. In this research only one dike section geometry is analysed.

2. The analysis is limited to variation of the parameters grainsize (d70) and permeability (k). Other parameters are assumed 'known'. This is done because grainsize and permeability are identified as the most important/sensitive parameters in the assessment model. Clay layer depth is normally quite sensitive as well because it influences both the parameters d and L (Calle et al., 1999). By neglecting covering clay layers, the relative importance of sand layer properties increase and it is not necessarily to make a 2D-analaysis. *In chapter 3 is explained why an analysis without covering clay layer is still valuable/important and representative for actual situations.*

3. The dike section used in the simulations is 1000 meters long. This is similar to the length used in the PhD theses of Kanning (2012) and Schweckendiek (2014) and shows good model performance in terms of calculation time. Other dike section lengths will be evaluated in the sensitivity analysis.

   Note: Typical dike stretch lengths are hundreds of meters till several kilometres. For example, dike ring 5 in the Netherlands has stretches from 340 meter to 2400 meter. Dike ring 17 in the Netherlands has stretches from 410 meter to 3050 meter.

4. The dike is subdivided into a grid of 1000 subsection of each one meter. Because the grid size is much smaller than the section length and smaller than typical correlation lengths it is assumed the grid is sufficiently fine. Connection to practice: modelling subsections of 1 meter and 1000 measurements over 1 km is very precise in perspective to what is common in practice.

5. In the simulation, see step 6 of the previous section, 20.000 repetitions are made. This number is chosen to reach stability in the results. Because with 20.000 realisations stability in results is reached, almost all possible combinations of parameter value distributions (which are random) are assumed covered.

6. The measurements densities that are analysed have respectively intervals of 500, 250, 200, 100, 50, 40, 25, 20, 10, 5, 2 and 1 meter. The number of measurements and spacing between measurements is equal for considered subsoil parameters. It means the density of measurements is equal for every parameter. In practice the number of measurements of different properties might deviate. This is neglected to limit the number of configurations that could be analysed.

Although some scope limitations are posed, the obtained results are useful because these can be extended to general concepts about measurement densities. This is showed in chapters 6, 7 and 8. Besides that the simulation captures the most important aspects of the piping assessment considering the use of measurements. In the research the effect of scope limitations and assumptions are studied. It is found out that the current scope is relevant to practice and actual situations. In the discussion (chapter 7) is reasoned that the influence of the assumptions to the general conclusions is minor.

3. Representative dike section model

This chapter is related to the first research question: *What is a representative model of a piping sensitive dike section?*

3.1. Introduction

In construction of a representative model, focus has been on the representative geometry and representative statics to model the distribution and spatial variation of the soil parameters. The choices result in a set of input parameters that are used in the simulation. The data sets generated with this input are tested to check the performance of the model.

Figure 8 provides a schematization of the considered dike, based on the assumptions and scope of the research as discussed in chapter 2.

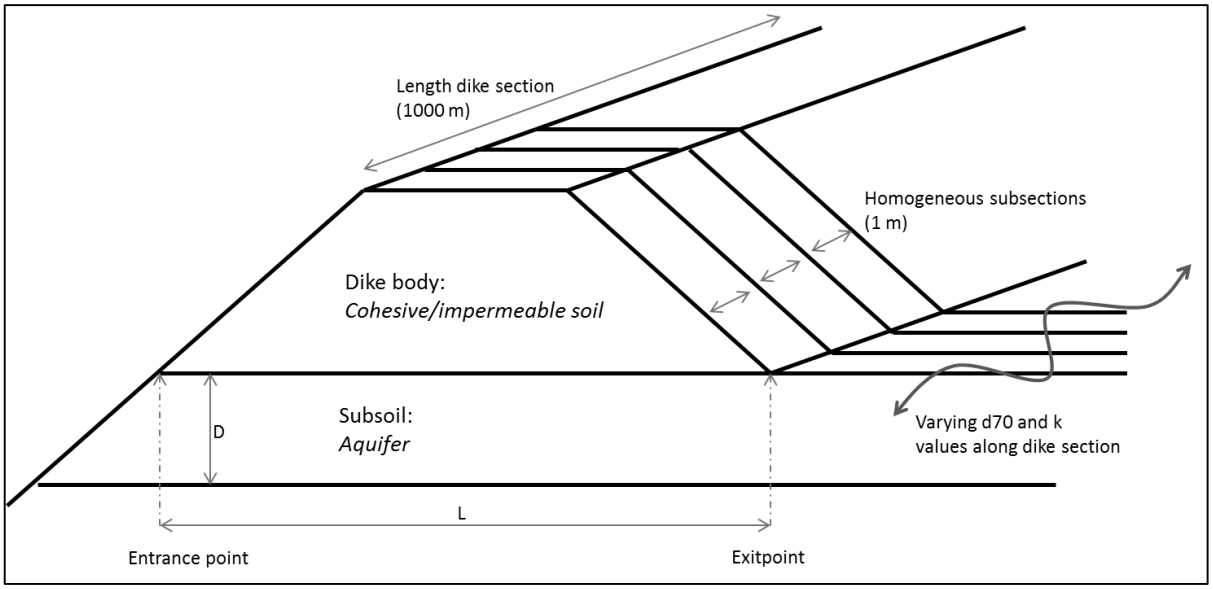


Figure 8 – Schematisation of dike section model

3.2. Representative geometry

Note that pipes develop easier for small seepage lengths; therefore, a typical piping sensitive geometry is expected to have a rather small dike body. In the Netherlands the seepage length is often increased by the presence of a consistent covering clay layer. This layer moves entrance and exit points away from the dike body. Lack of a clay layer increases the sensitivity to piping.

Dike sections where the dike body width is determining the seepage length (in Dutch also known as 'schaardijken') are in practice of main attention as the resistance is typically low (POV Piping, 2015). Especially at those places the aquifer related parameters d70 and k have major influence in determining whether piping is a threat or not.

This research focusses on subsoil parameters of the water transporting layer in a piping sensitive geometry. Only variation in grainsizes and permeability's are studied, therefore a simplified geometry is considered where,

- A clay (cohesive and impermeable) dike body on top of an aquifer which consists of sandy sublayers (Figure 8). The combination of the clay and sandy sublayers makes the dike sensitive to piping as the clay layer prevents pipes from collapsing (Förster et al., 2012).

- Covering (clay) layers are absent outside (in Dutch so called 'voorland') and inside the dike (in Dutch so called 'deklaag'). This means that the uplift mechanism can be ignored and the clay layer depth parameter *d* in the equation of Sellmeijer is zero. It represents a 'schaardijk' situation.
- The seepage length *L is* directly determined by the width of the clay dike body and is constant along the length of the dike. It must be stressed that the purpose of this research is not about absolute quantification of accuracy but about quantifications of accuracy related to measurement densities. Therefore the actual choice of L is of minor importance. A common value of 30 meters is chosen (Calle et al., 1999).

3.3. Representative (spatial) distribution

In the previous section is indicated that soil characteristics play an important role in the piping resistance. Soil characteristics cannot be observed directly and are therefore uncertain. In literature random fields based on statistical parameters are used to model uncertain soil properties (REF). Because of randomness, each generated field is different. However, on average the random fields behave according to the input statistics.

It appears very difficult to determine representative input statistics. Different geological backgrounds result in different typical fluctuation patterns in the field and little is known about the (spatial) distributions of soil properties as it lacks very fine maze soil investigations yet (Kanning, 2012), (de Visser, Kanning, Koopmans, & Niemeijer, 2015). However, some literature is available that provides statistical characteristics to describe spatial variations of parameters. *Appendix 5 gives an overview of relevant literature and statistical characteristics.*

3.3.1. Reference scenario

The statistical characteristics that are chosen to describe the spatial variations of parameters are listed in Table 1. This combination is from now on used and defined as the *reference scenario input*

Table 1 – Representative parameter variances to use as 'reference scenario'.

Soil parameter	Mean ($\boldsymbol{\mu}$)	Standard deviation ($\boldsymbol{\sigma}$)	Correlation length ($\boldsymbol{\delta_0}$)	Measurement error ($\boldsymbol{\phi}$)
d70	$2.00*10^{-4}$ [m]	$0.30*10^{-4}$ [m]	180 [m]	$0.15*10^{-4}$ [m]
k	$1.40*10^{-4}$ [m/s]	$1.40*10^{-4}$ [m/s]	600 [m]	$0.70*10^{-4}$ [m/s]

With respect to Table 1 the following considerations are of importance:

- Means, standard deviations and correlations lengths are educated guesses based on literature. The used statistical parameters are mainly based on the values used in the probabilistic analysis of piping in projects 'Floris' and project 'VNK' (PC-Ring calculations values). It is assumed that those variances are well representing the 'average' macro conditions of the aquifers as they are observed in the field in the Netherlands.
- Measurement error is modelled as white noise. The variance is estimated to be 50% of the standard deviation used to describe spatial variations. This is based on rough indications from Kanning (2012) in which it was however unclear where the indicated variances refer to model input, model output, measured nugget effect, etc. Because it lacks of more sources that indicate magnitudes of measurement error, the 50% value should be considered as a first estimate. A percentage is used to make sure that parameters with high variations show high measurement errors as well.

- It is assumed that spatial fluctuation is well approximated by a Gaussian autocorrelation function, described by:

  $$\rho(x) = e^{-\left(\frac{x}{\delta_0}\right)^2}$$

  In which $\rho(x)$ is the correlation coefficient, x is the Euclidian distance between two points and δ_0 is the 'correlation length'. The choice for a Gaussian autocorrelation structure is made following Kanning (2012) and Schweckendiek (2014).

- d70 is modelled with a normal distribution. In literature it is common to model d70 as lognormal to prevent negative values and allow for larger grainsizes. Nonetheless the variance in this case is relatively small with respect to the mean, which means negative d70 values will not appear and the difference between normal and lognormal is very small. Besides, normal distributed data is more straightforward to model.
- k is modelled with a lognormal distribution. This is necessarily to prevent negative values as the variance is in the same order as the mean. A lognormal distribution of k is very common in literature.

3.3.2. Correlation structure to model soil fluctuation patterns

There is limited literature available of scales of fluctuation of d70 and k. Especially horizontal scales of fluctuations are challenging, as sampling intervals are usually not sufficiently small (Kanning, 2012). However, the choices with respect to correlation have significant implications. Gaussian autocorrelation implies rather *smooth* fluctuation patterns. Smooth fluctuation seems to correspond to 'within deposit variations', because naturally deposits have developed very gradually. In contrary the Exponential autocorrelation function results in rather 'abrupt' fluctuations. This might be a good representation of geological details and/or interfaces between different deposits.

The correlation lengths of d70 and k, give typical ranges in which those parameter values fluctuate. A correlation length of 180 meter for d70 implies that grainsizes show almost no fluctuation on short distances (0 – 50 meters). It is questionable if this is representative for Dutch practices. Fine maze soil investigations in testing grounds show rather large fluctuations on small distances (0 – 10 meters) (de Visser, Kanning, Koopmans, & Niemeijer, 2015). However, strong fluctuations are related to fluvial deposits. There are found indications that Aeolian deposits show less fluctuations in grainsize. Especially those aeolian deposits contain the very fine grains that are sensitive to erosion.

The correlation length of k is assumed to be 600 meter. This implies a rather constant appearance of permeability in dike sections of 1000 meter. The higher correlation compared to d70 is caused by the nature of the parameter k as a layer average, meaning that small scale variations are already averaged out. A correlation length of 600 meter in combination with a smooth correlation pattern is possibly not representative for situations with anomalies such as old river gullies. Those anomalies appear rather sudden in relatively homogeneous deposits. To model the possibility of a sudden high permeability, the correlation length should equal the typical width of an old river bed.

Expert believe that the smallest correlation length of d70 is possibly zero (de Visser et al., 2015), (POV piping, Werkplaats Zwarte water, 2015). It means that d70 can behave as white noise. It is unknown whether d70 behaves as white noise due to measurement error, measurements from different (geological) layers or due to actual variation. Nevertheless, experts confirm that sand layers can be very heterogeneous, at least for some areas in the Netherlands. The smallest expected correlation length of k is equal to the dike width. This because k is influenced by a 3-dimensional flow pattern (personal communication, Kanning, 2015).

In the standard input configuration, it is chosen to model correlated soils with correlation lengths of respectively 180 meter and 600 meter. Based on literature and expert judgment a range of plausible correlation distances is tested in the sensitivity analysis (chapter 5) as well.

3.4. Generation of correlated data series

With the selected input configuration (Table 1), it is possible to provide the model with data. The generation of a correlated data set is based on normal distributed random signals and a correlation matrix which is based on an autocorrelation function (correlogram). To simulate measurement error, a white noise signal is added to the correlated data set obtaining a correlated noised signal. The following formulas are used:

$$\boldsymbol{A} = [\boldsymbol{R} * \boldsymbol{U}] + \mu + randn(0, (\phi))$$

$$\boldsymbol{R} = randn(0, \sigma)$$

$$\boldsymbol{U} = chol(\boldsymbol{C})$$

$$\boldsymbol{C} = mat[\rho(\delta_0)]$$

In which:

- $\boldsymbol{A}$ is a correlated noised signal.
- $\boldsymbol{R}$ is a 'white noise' signal with zero mean and a certain standard deviation.
- $\boldsymbol{U}$ is the output of the function *chol* that performs a Cholesky transformation to the correlation matrix $\boldsymbol{C}$.
- The function *randn* is used to generate a random signal with a normal probability density function.
- The function *mat* transforms the autocorrelation function into a correlation matrix according to the statistic δ_0 ('correlation length').
- The statistic μ (mu) is used to shift the correlated data set to the wished mean.
- The statistics σ (sigma) is used to generate a white noise signal with a certain standard deviation and zero mean.
- The statistic ϕ (phi) is used to generate a white noise signal with a certain standard deviation and zero mean.

The model input is chosen such that the output matches the representative parameter variances of Table 1. For *d70* the generation of data is straightforward as the *randn* function is used to generate normal distributed data. However, in the generation of *k* data a transformation is needed. To obtain representative means and variances without negatives, the data is transformed to a log normal distribution. This is achieved by taking the exponential of normal distributed correlated data series. This transformation results however in a difference between input parameters and output parameters. Trial and error yielded that: $[u_{in} = 0\,, \sigma_{in} = 0.85 * 10^{-4}]$ results in an output that matches the representative mean and variance of *k.* See also section 3.5 about 'model output'. Figure 9 and Figure 10 show the steps of the simulation in an example of respectively a *d70* data set and a *k* data set.

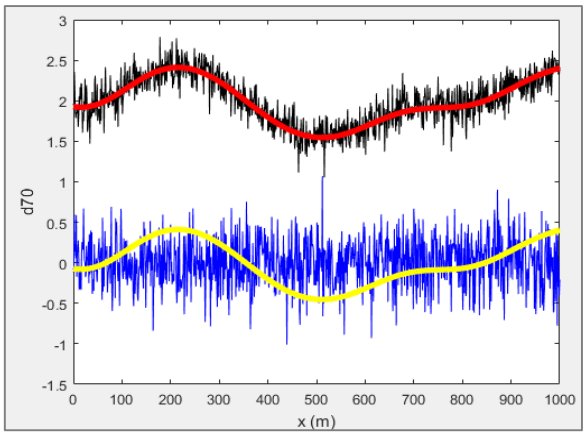


Figure 9 – Transformation of random signal (=blue) to a correlated signal (=yellow) to a correlated signal with predefined mean (=red) to a correlated noised signal (=black). Simulated realisation: standard d70 configuration with measurement error.

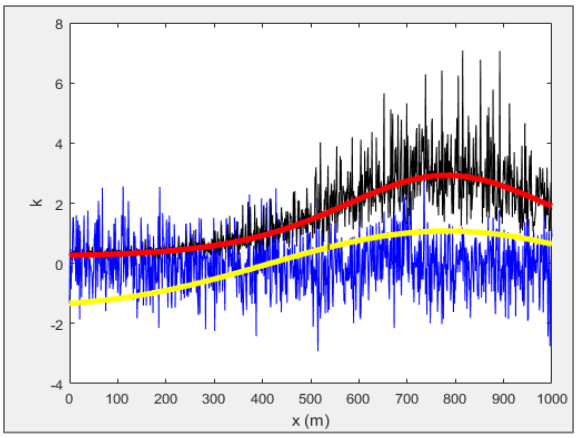


Figure 10 - Transformation of random signal (=blue) to a correlated signal (=yellow) to a strictly positive correlated signal(=red) to a strictly positive correlated noised signal (=black). Simulated realisation: standard k configuration with measurement error.

These examples show that in both cases negative values are prevented. In case of *d70* the white noise is of constant magnitude along the domain. However, in case of *k* the white noise is very small where the value of the correlated set is close to zero and much larger at locations where the value of the correlated set is higher. This is because the measurement error of *k* is the result of the same exponential transformation. This results in the logical situation that data values are in the same order of magnitude as the error.

Examples of generated data sets are given in Figure 11 until Figure 14. In these examples a domain of 3000m dike is considered to give overview of the variances and scales of fluctuation corresponding to the input statistics. Figure 12 and Figure 14 show that a large number of realisations 'cover' all kinds of possible variations. As a consequence of a random process each output is unique, although the input statistics are constant. Therefore, a simulation with an extensive number of data realisations is a method to represent uncertainty in soil conditions.

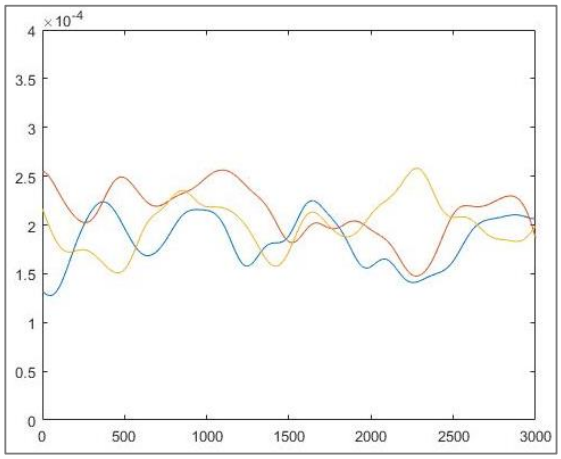


Figure 11 – 3 example data sets of d70, each as a different result of equal input statistics: in this case corresponding to the standard configuration.

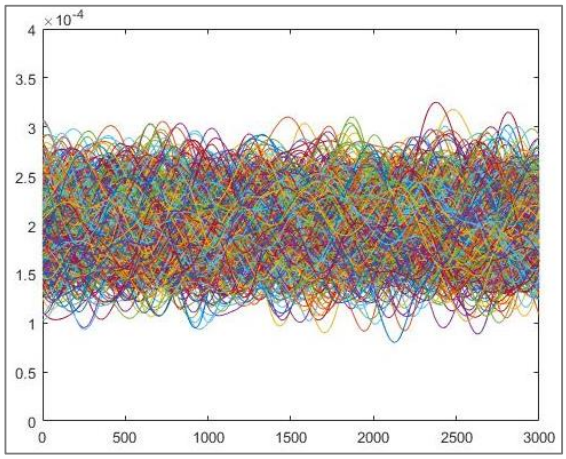


Figure 12 – result of 1000 random d70 data sets

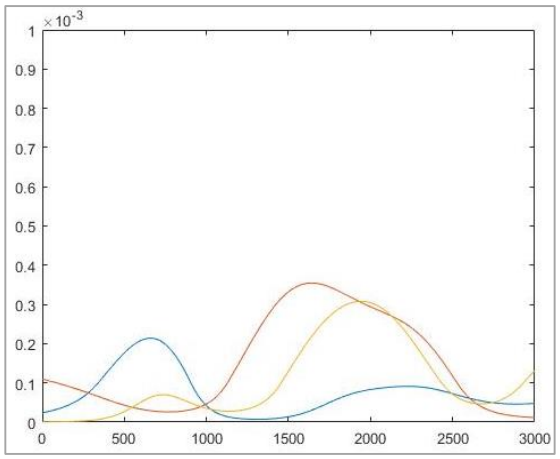


Figure 13 – 3 example data sets of k, each as a different result of equal input statistics: in this case corresponding to the standard configuration.

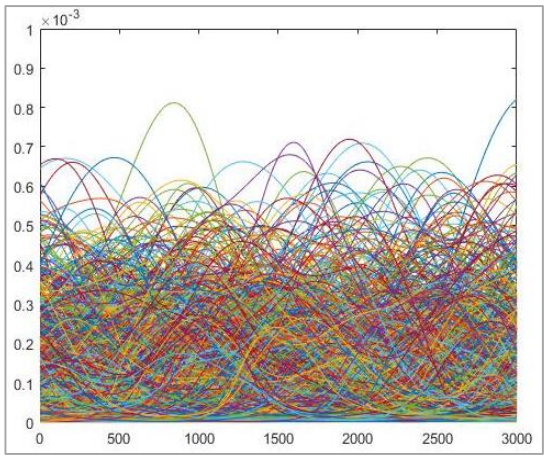


Figure 14 – result of 1000 random k data sets

3.5. Output of data generation script

Because a random data generation script is used, it is important to analyse a sufficient amount of data sets to obtain stable results. In this section the model performance of 20.000 random data sets (see chapter 2), consisting of 1001 values each, is analysed.

3.5.1. Mean and variance

Measuring over the complete data field of 2002000 data points, the *d70* field has a mean and variance equal to the input values. Notice that it can indeed be concluded that the randomness has vanished after enough realisations. The mean and variance of the *k* field differs from the input. This is caused by the translation from standard normal distributed values to lognormal distributed values. In contradiction with the *d70* field, the mean and variance of *k* are not completely stable after 20.000 realisations (randomness has not completely vanished). The fluctuation range is limited to $0.1*10^{-4}$ m/s for both the mean and variance. This is assumed acceptable with the calculation time significantly increasing and the precision of the result hardly.

Table 2 – Measured output values. Simulated field N=20000, standard input configuration, no measurement error.

	d70 [m]	**k [m/s]**
μ_N	$2.00*10^{-4}$	$[1.41 – 1.46] *10^{-4}$
σ_N	$0.30*10^{-4}$	$[1.45 – 1.52] *10^{-4}$
$\mu(\sigma_n)_N$	$0.24*10^{-4}$	$[0.55 – 0.59] *10^{-4}$

Table 3 – Measured output values. Simulated field N=20000, standard input configuration with measurement error

	d70 [m]	**k [m/s]**
μ_N	$2.00*10^{-4}$	$[1.55 – 1.59] *10^{-4}$
σ_N	$0.33*10^{-4}$	$[1.88 – 1.95] *10^{-4}$
$\mu(\sigma_n)_N$	$0.29*10^{-4}$	$[0.99 – 1.03] *10^{-4}$

Note: μ_N is the mean value of all data points, σ_N is the standard deviation of all data points, $\mu(\sigma_n)_N$ standard deviation per data set averaged over all realisations. So the average standard deviation of a data set.

Visualizations of the data in histograms (*see appendix 6*) show that indeed d70 data is normally distributed and k data is lognormally distributed. Analysis of statistics per location in the domain, show

the distribution of data is not affected by locations or boundaries. It is concluded the model functions well.

3.5.2. Correlation

Correlation is measured using the Method of Moments. To check for the overall average correlation of all realisations the unbiased correlation is calculated using the average mean and variance of the complete data set (i.e. average values over all realisations).

Estimator of the correlation coefficient using Methods of Moments is described by (Lark, 2000).

$$\rho_z(\delta) = \frac{1}{(n-\delta)s^2}\sum_{i=1}^{n-\delta}(z(x_i)-\mu_z)(z(x_{i+\delta})-\mu_z)$$

In which z is the parameter value at a certain location x_i in the domain, $(n-\delta)$ are the number of data pairs that have separation distance δ ('lag'), μ_z is the mean of the set and s^2 equals the sample variance.

Using this formula, the measured correlation is compared with the correlogram (autocorrelation function) that was used as input for the correlation matrix. The results are presented in Figure 15 and Figure 16. Those figures show that on average the output correlation is equal to the input.

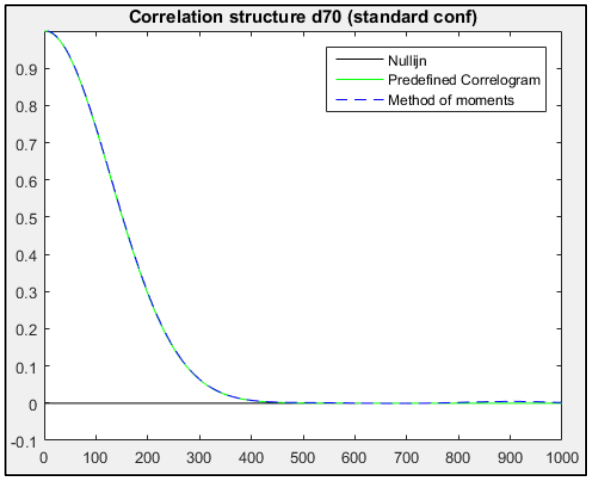


Figure 15 – Correlation structure of d70 realisations. Simulated field: N=20000, standard configuration.

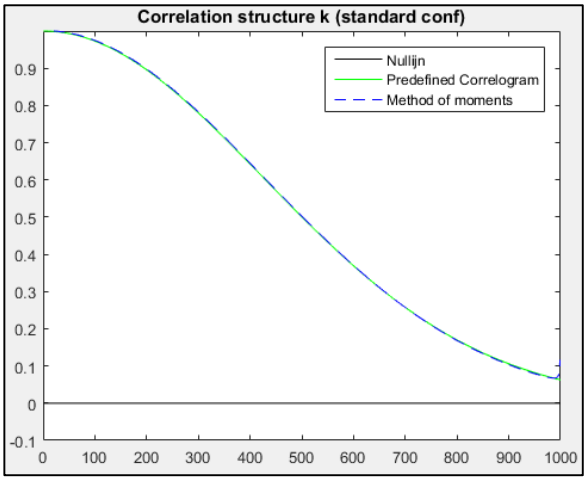


Figure 16 – Correlation structure of k realisations. Simulated field: N=20000, standard configuration.

The local variance of a data set is affected by correlation (Table 2 and Table 3). Because a single data set has a limited number of values and the values within the domain correlate with each other, there is not sufficient 'fluctuation space' to reach the input variance. Input variance is only reached when the ratio correlation length/domain length goes to zero. Local variance in a correlated set is therefore always smaller than input variance. Measuring the variance over all sets, can be seen as measuring the variance of 1 correlated set with a length of 20.000*1000 meter. Although correlation is present, the ratio correlation length/domain length then goes to zero and the measured output variance equals the input variance.

More about the model output is given in appendix 6.

3.6. Summary and conclusions

The used dike section model consists of a piping sensitive geometry and data sets which describe aquifer parameters d70 (grainsize) and k (permeability) in the length direction of the dike. It is concluded that a dike is piping sensitive in a configuration with a cohesive and impermeable dike body

on top of a permeable sand layer. A small dike body and lack of a clay layer increases the sensitivity to piping and makes the aquifer properties d70 and k extra important in the determination of piping resistance.

Aquifer properties, such as d70 and k, cannot directly be observed and are therefore considered uncertain. This uncertainty can be described by statistics: probability density distributions to indicate probable appearances and correlation lengths to indicate the variation of soil properties. The d70 and k data of the dike section model is generated randomly according to a probability density function and an autocorrelation function.

It is concluded that uncertain soil parameters in the Netherlands cannot be described with one representative distribution and autocorrelation function. Therefore, first a reference scenario is described. Other (spatial) distributions are analysed in a scenario analysis. The selected statistical parameters for the reference scenario are mainly based on PC-Ring calculations. This means d70 data is distributed normal (CoV=0.15) and shows correlation (Gaussian) over a length of 180 meters. The k data is distributed lognormal (CoV=1) and shows correlation (Gaussian) over a length of 600 meters. The choices with respect to correlation imply smooth correlation patterns and are expected to correspond to large scale *within deposit* fluctuations. The repetition of data generation according to the input statistics results in many unique possible dike sections. Appropriate functioning of the model is demonstrated as the average output statistics of soil parameters complies with the input statistics.

4. The error in an assessment of the strength and the influence of measurement density

This chapter is related to the second research question: *How does the density of measurements influence the accuracy of strength assessments using the semi-probabilistic assessment approach?*

4.1. Introduction

In the previous chapter a dike section is represented as a geometry combined with a set of parameter realisations. The parameter realisations describe the values of the individual parameter along the dike. Combining the parameter realisations in the model of Sellmeijer (2011) results in one strength realisation, which describes the strength along the dike.

In the upcoming chapter the parameter realisations of d70, k and resulting H_c realisation are used to quantify the accuracy for several measurement densities. In section 4.2. simulation results are provided. To illustrate the procedure, first the result of one arbitrary dike section and one measurement density is shown. Then it is shown that the accuracy itself is a random variable due to uncertain soil conditions. Multiple measurement densities are then compared and the result of the complete simulation is summarized in one bandwidth of errors. One complete simulation gives insight in the reliability of assessments and how this is related to measurement densities. In section 4.3 an analysis and explanation of the results is given. The chapter finalizes with a summary and preliminary conclusion.

4.2. Simulation results

4.2.1. Accuracy of a single strength assessment

Figure 17 provides an arbitrary strength realisation with the reference strength and assessed strength. In this example the characteristic values are based on a measurement density of 3 per 1000 meter.

The individual parameter realisations with calculated characteristic values that determine the assessed strength value, can be found in Appendix 7.

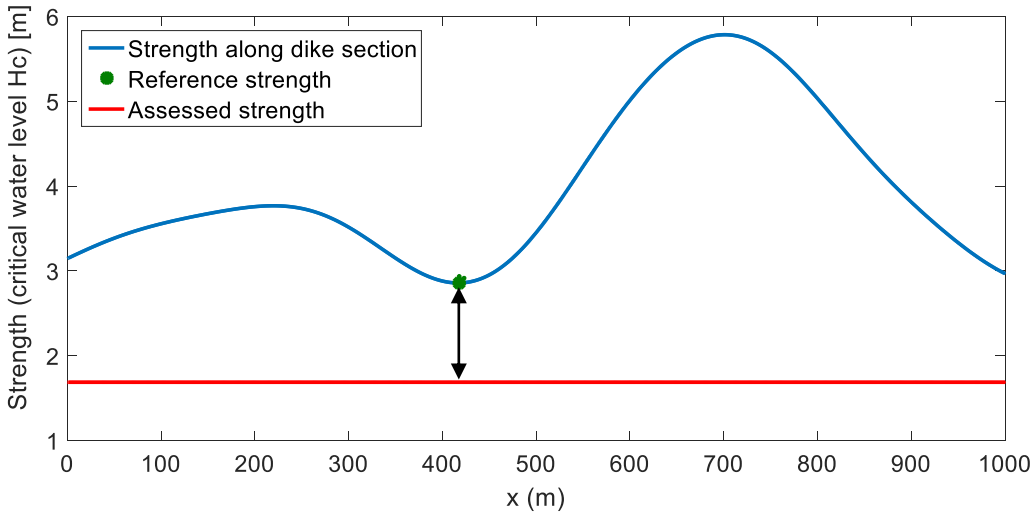


Figure 17 – Visualisation of strength realisation, reference strength and assessed strength based on 3 measurements.

The strength of a dike section is spatially varying (blue line). In the detailed assessment one strength value is assumed representative along the entire section (red line). The figure shows a difference between the assessed strength and strength realisation over the entire domain. The reference

strength, which is considered the representative strength value for this section, is higher than the assessed strength (blue arrow). For this specific section it appears that with uncertain input data (only three measurements prone to measurement uncertainty) a smaller strength is calculated than would have been calculated as the spatial distribution of properties would be completely known.

The *error* in an assessment due to uncertain input is quantified by difference in critical slope (meter/meter or dimensionless):

$$Error = [reference\ strength - \ assessed\ strength]/L$$

L is the seepage length, equal to the width of the dike body. The applicability of the results is expanded by making them independent of the width of the structure; dividing by L and expressing the error in terms of critical gradient.

The error can have three types of values:

1. $Error \approx 0$, the assessed strength is accurate.
2. $Error > 0$, the assessed strength is conservative as the reference strength is underestimated.
3. $Error < 0$, the assessed strength is unsafe as the reference strength is overestimated.

4.2.2. Accuracy as a random variable

In the previous section is illustrated how the error in an assessment is quantified. This is done for one possible strength realisation that was chosen arbitrary. In practice however, the strength and its variation along a dike section are uncertain. Monte Carlo is used to quantify the accuracy of multiple (random) strength realisations.

Figure 18 shows the distribution of the error over the interval: $-0.1 < Error < 0.1$. The error depends on the variation of the strength parameters along the dike section. Since these are uncertain, the error is uncertain as well.

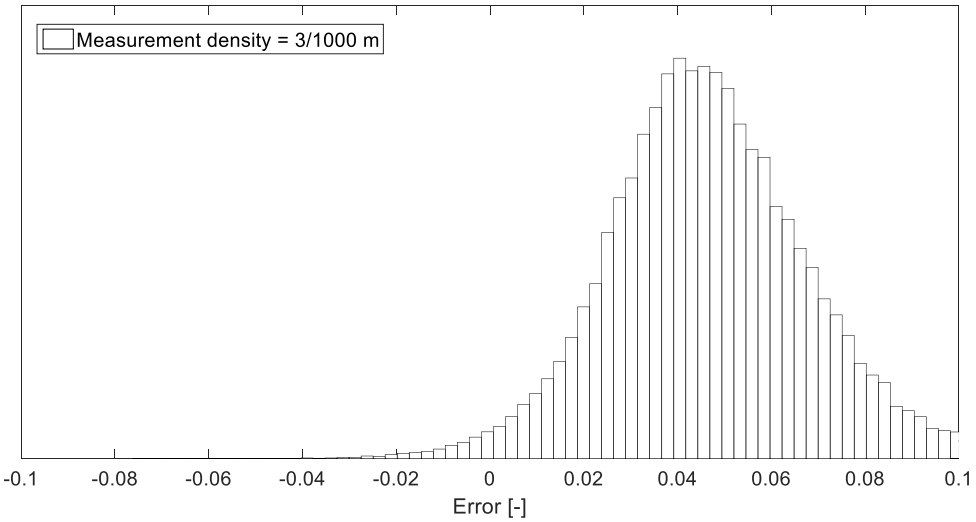


Figure 18 – histogram of errors in an assessment with 3 measurements per 1000 meter

The error of an assessment varies between the realisations. Beforehand it is unknown to which extend the assessed strength matches the actual strength, either an overestimation, accurate calculation and underestimation are plausible. Unfortunately, it is not possible to determine the exact error due to the lack of a reference strength in reality. The only available strength follows from the measurements and assessment.

The error related to a density of 3 per 1000 meter, equal to a 500-meter measurement interval (Figure 18), resembles the shape of a normal distribution. The error cannot be quantified deterministically but a simulation provides information about the mean error and cumulative distribution values which make it possible to speak in terms of probabilities. For example (using Figure 18), the probability that the error is negative is about 2%.

If the range of errors includes negative values, there is a probability that strength is overestimated: the section is actually weaker than is calculated with limited data. If the actual strength is below the normative load an unsafe situation can be the result. The section should be reinforced but there is a chance this is not noticed because from the assessed strength the conclusion is drawn the norm is met. In this case the probability of failure is higher than expected at possibly higher than allowed.

The mean error value is defined as the expected error (bias) in the assessment method. For the considered characteristic value analyses based on a measurement density of 3 per 1000 meter, the expected error is 0.04 in terms of critical gradient (m/m). Which corresponds for the considered dike geometry to an expected underestimation of 1.2 meter in terms of critical head difference: the method is on average biased towards rather conservative strength estimations. The standard deviation of errors is a measure of consistency in the error, i.e. precision in the strength assessment. A high inconsistency means a wide range of possible errors and therefore a high uncertainty in the reliability of the strength assessment outcome.

The standard deviation is good measure of consistency, but does not indicate to what extend assessments are conservative or unsafe. A 95% certainty interval solves this problem. Values above the upper boundary of the 95% interval represent errors for which one can state with 2.5% probability that the actual error is relatively more conservative. For Figure 18 this upper boundary is 0.09, corresponding to an underestimation of 2.7m in terms of critical head difference. Values below the lower boundary of the 95% interval represent errors for which one can state with 2.5% probability that the actual error is relatively more unsafe. For Figure 18 this lower boundary is 0.006, corresponding to an underestimation of 0.2m in terms of critical head difference.

4.2.3. Influence of measurement density

In the previous section is illustrated that the error is uncertain. This section elaborates on influence of the measurement density on the probability density distribution of the error. This is done by repetition of the Monte Carlo analysis for different measurement densities.

Empirical probability density distributions of four selected measurement densities (corresponding to measurement intervals of 500m, 250m, 100m, and 10m) are presented in Figure 19.

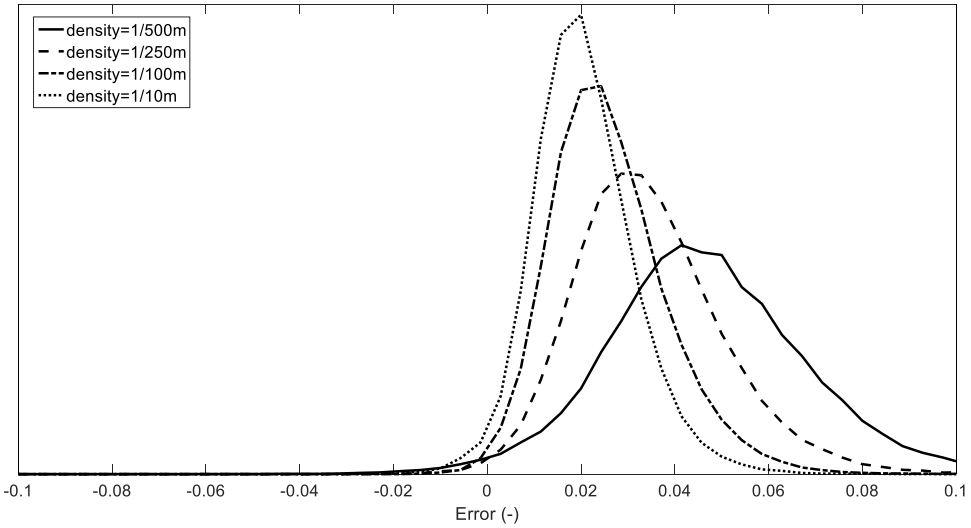


Figure 19 – Error histograms of 4 different assessments

From each of 12 considered distributions, the mean, upper boundary and lower boundary of the 95% interval are determined and plotted against the corresponding measurement density. For convenience the measurement densities are presented on a logarithmic scale. The combination of the lower and upper boundary results in a bandwidth of errors that directly visualizes the influence of measurement density to the range of the possible error.

More background about strength distributions is provided in appendix 8. It also provides a table with the average magnitudes of reference and assessed strength values (H_c).

The plot from Figure 20 summarizes the results of one complete model simulation and is referred as the *reference result*.

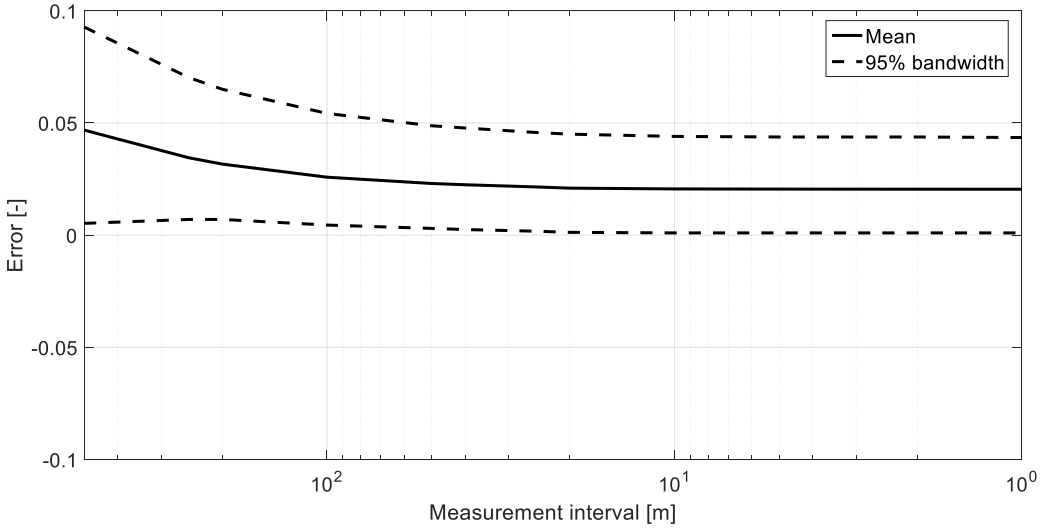


Figure 20 – Mean and 95% bandwidth of errors for measurements intervals 500m --> 1m. The Error is defined as the difference in critical slope of either the reference and assessed strength and is therefore dimensionless (m/m).

Considering the mean error (Figure 20), the following observations are relevant:

- The calculated strength (assessment) decreases as the number of measurements increases.
  - The decreasing trend is levelling off and increased measurement densities have hardly influence after approximately 1 per 50 meter

- The expected error is positive at all measurements densities. On average the assessment method underestimates the reference strength.

Regarding the 95% bandwidth of errors, the following observations are considered relevant:

- The range of errors becomes smaller up to approximately 1 per 100 meter. After that points only little changes are visible. Which means that from that point the influence of more measurements is minor.
- The calculated strength values in the upper boundary of the bandwidth decrease as the measurement density increases. The trend is equal to the trend in the mean error.
- The calculated strength values in the lower boundary of the bandwidth increase up to a measurement density of about 1 per 200 meter. At this density a trend reversal takes place.

  The result in Figure 20 is obtained from a simulation in which measured data points equally distributed along the domain. Appendix 9 provides the results of a simulation in which measurement locations are selected randomly from the domain. This increases the uncertainty for low measurement densities because the lower boundary shows a different trend. In the appendix is explained that it is favourable to place measurements at equal intervals if data is spatially correlated.

- Unsafe situations are plausible as actual strength can be overestimated by an assessment (Figure 19). Because calculated strength values are less conservative if a higher measurement densities are used, overestimation is even more likely in those cases.

The effect of measurement error is visualized in Figure 21. Considering the situations with (black lines) and without (red lines) measurement error, the following observations stand out:

- Including measurement error, calculated strength values (assessment) are even smaller than without measurement error.

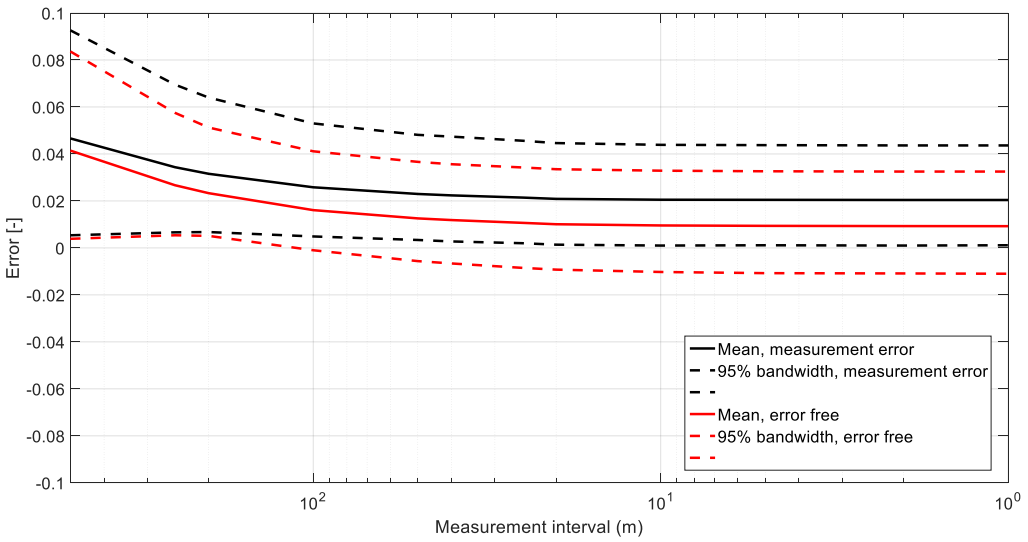


Figure 21 – Reference result in cases that 1) measurements are subjected to measurements and 2) measurements are error free

4.3. Analysis and explanation

In this section is analysed, explained and discussed how the differences between assessed and reference strength arise considering the semi-probabilistic assessment method.

4.3.1. Decreasing trend expected error

On average, measurements density influences the characteristic value calculations in mainly two ways:

- First, the decreasing trend of the expected error by increasing the measurement density is caused by an increase in the calculated characteristics values of d70 and decrease in the calculated characteristic values of k (Table 4).
- Secondly, the measured variance (Table 5) affects the calculated characteristic values. On average the measured variance depends on the measurement interval and correlation length. If the measurement interval is larger than the correlation length, an extra measurement is on average adding new information about the range of the data and increases the measured variance. For measurement intervals that are equal or smaller than the correlation length, measurements are not completely independent.

Note that because the largest measurement interval of k is already smaller than the considered correlation length, the measured variance of k decreases in the whole domain of considered measurement densities.

Table 4 - Average of characteristics values for increasing number of measurements

Number of measurements	3	5	6	11	21	26	41	51	101	201	501	1001
d70 e-4	1,24	1,42	1,46	1,54	1,57	1,58	1,59	1,60	1,60	1,60	1,60	1,60
k e-4	13,0	4,65	4,07	3,26	2,97	2,92	2,84	2,80	2,77	2,75	2,75	2,75

Table 5 – Average of measured sample variances for increasing number of measurements

Number of measurements	3	5	6	11	21	26	41	51	101	201	501	1001
$\mu(\sigma)_{d70}$ ***E-5**	2,65	2,73	2,69	2,56	2,48	2,46	2,44	2,43	2,41	2,40	2,40	2,40
$\mu(\sigma)_{k}$ ***E-5**	8,28	7,14	6,85	6,26	5,94	5,87	5,78	5,74E	5,68	5,64	5,62	5,61

4.3.2. Trend in uncertainty interval

The upper boundary of the 95% confidence range shows the same trend as the mean error. Because high measured variance results in conservative characteristics values, the upper boundary of the bandwidth represents assessments for which measured variance is relatively high. When the measured variance is high, the effect of student-t factor is relatively strong. Furthermore, it represents situations where the measured variance is accidently high although the measurement density is small, resulting in a measured variance that can only decrease for higher densities (Table 6). A decreasing variance and decreasing student-t factor result in relatively less conservative assessments, i.e. a decreasing trend of the upper boundary.

Table 6 - Measured variance in an example that is representative for a situation in which the assessment results in a small strength value (upper boundary of error range)

Number of measurements	3	5	6	11	21	26	41	51	101	201	501	1001
σ_{d70} ***E-5**	5,11	3,83	3,70	3,12	2,95	2,91	2,85	2,82	2,78	2,75	2,74	2,74

The lower boundary of the 95% confidence shows a more complicated trend. It represents situations where the measured variance is relatively small. When the variance is small, the influence of the

student-t factor in determination of the characteristic value is small as well. The trend is then dominated by the measured variance that is influenced by the measurement density. It appears that if the measured variance is accidently very small (or almost zero) for a small measurement density, the measured variance is expected to increase for higher densities as long as the measurement interval is larger than the correlation length. This effect results in a trend reversal in the lower boundary at the measurement interval that is typically equal to the smallest of correlation lengths of the different parameters. In this case an interval of 200m as this interval is closest to the correlation length of d70 (180m). The trend reversal is more clearly visible in the result of a simulation without measurement error (Figure 21) because for that case the measured variance is not influenced by noise.

4.3.3. Sources of uncertainty and influence of measurement density

Different sources of limited (use of) information contribute to inaccurate assessments. In some cases, errors add up mainly resulting in larger underestimation of the actual strength. On the other hand, uncertainties might compensate each other. This possibly results in error values close to zero. Different sources of uncertainties are already introduced in chapter 1 and the contribution to the bandwidth of errors will be discussed in the following sections.

Statistical uncertainty

The assessment estimates the probability density distribution by an assumed distribution type (for example normal or lognormal) and measured mean and variance. The number of measurements influences the measured mean and variance. Measuring each value of the population means that the uncertainty related to the mean and variance becomes zero. The reduction of this statistical uncertainty results in a smaller range of errors for high measurement densities.

Still a second source of statistical uncertainty is left because the characteristic value formula assumes that measurements can be considered as samples from an uncorrelated population (Calle E. , 2007). This means that parameter values in a data set are assumed to behave according a theoretical distribution. But the distribution in practice may deviate from that. In this research the assumed distributions in the characteristics value formula are actually equal to the data input distributions. However, if the considered dike section has a length in the same order as the present typical correlation length, the measured distribution might be significantly different then the theoretical probability density distribution. This is referred to as *boundary effect*.

With boundary effects the actual mean and variance of a smaller section still not result in the actual representative parameter value. The errors in the assessment due to the boundary effect are independent of the measurement density. A part of the range of errors (still present for high measurement densities) is therefore explained. *Further explanation and examples are given in appendix 10.*

Spatial variability

Where (part of) statistical uncertainty can be decreased with more measurements the uncertainty related to spatial variability remains unchanged. This is because measured values are not connected to measurement locations. Measured values are only used to calculate representative parameter values that are assumed constant over the entire domain (see section as well 1.2.2).

The spatial variability explains a large part of the bandwidth of errors. Multiple varying parameters make it possible that unfavourable values of one parameter are compensated. Due to compensation it is likely that the reference strength is underestimated in an assessment. Namely in the assessment is assumed that the strength is given by the combination of unfavourable parameter values.

In the detailed assessment the statistical upper or lower 95% characteristic values are used. This means in theory a probability of 5% that there is actually a more unfavourable value present. If unfavourable parameter values coincide at one location, it is possible that the calculated characteristics values are allow overestimation of the actual strength (negative error values).

In the simulation no safety factor is applied. However, in practice a safety factor is prescribed in the detailed assessment to prevent that the assessed strength is higher than the actual strength. A safety factor accounts for the length effect. *In Appendix 11 the effect of applying a safety factor is analysed. Furthermore, is shown that in general the same effect is obtained when smaller characteristics boundaries are used. For example, the statistical 99% lower boundary instead of 95% lower boundary.*

4.4. Summary and conclusions

The reference strength is defined as the minimum resistance of a section. The resistance is varying along the dike section as a consequence of varying d70 and k parameters. The assessed strength follows from characteristics value calculations that are based on measurement. Measurements are taken from noised d70 and k sets to simulate measurement error. Per random dike section an error is found as the difference between reference and assessment. Because the soil conditions of a dike section are uncertain; the error is uncertain as well. For each measurement interval a range of errors is calculated which is combined in a bandwidth of possible errors. This bandwidth visualizes the accuracy of the assessment in relation to the measurement density.

With respect to the influence of measurements, the main conclusion is found to be that the strength assessment with use of characteristic values is unreliable for every measurement density. This conclusion relates to the wide range of possible errors. Depending on the spatial appearances of soil properties, the assessment of the strength can be either accurate or inaccurate. In this chapter the accuracy of the reference scenario is considered. It is concluded that with respect to this reference scenario the characteristic value analysis is a conservative method because in most cases the assessed strength is smaller than the reference strength. This is also what the method intends to be to prevent that the actual probability of failure is higher than expected from uncertain data (overestimation). Because of this conservatism it is likely that actual safety is often underestimated in a safety assessment. It is shown that the conservatism (on average) decreases when the measurement density is increased up to a density of about 1 per 50 meter. At the same time this increases the probability of overestimation as well. It is therefore concluded that a higher measurement density does not always decrease the probability of failure.

Because the strength assessment based on characteristic values is not accurate for the reference scenario, it is concluded that two type of unwanted situations are possible. When the load and resistance are close to each other there is a probability that the norm is not met but the dike is not reinforced because the actual strength is overestimated. When the load is smaller than the resistance there is a probability that the dike is still reinforced because the actual strength is underestimated.

5. Scenario analysis – influence of assumed (spatial) distributions

This chapter is related to the third research question: *How is the accuracy of strength assessments influenced by the representative variances of the soil parameters?*

5.1. Introduction

The result in the previous chapter is obtained from a simulation with model input corresponding to the reference scenario. To be able to draw more general conclusions on the accuracy of strength assessments, a scenario- and sensitivity analysis is executed. Only the input related to measurement error, correlation structure and distribution (range of parameter values) is varied. Input is equal to the reference scenario if not mentioned differently.

This analysis identifies how the results and conclusions of chapter 4 are if the most important model input is (chosen) differently.

5.2. Measurement error

It is plausible that the measurements of an assessment are subjected to measurement error. An assumption has to be made on the type and magnitude of the error because specific numbers about measurement error are absent (in literature). In chapter 2 a first estimate is made in which the measurement error is assumed to be a white noise signal equal to a fraction (phi) of the spatial variability of the parameter. Figure 22 and Figure 23 give example realisations for d70 and k data respectively with both three values of phi simulating none (black lines), minor (red lines) and major (blue lines) measurement errors. A higher value of phi means more noise.

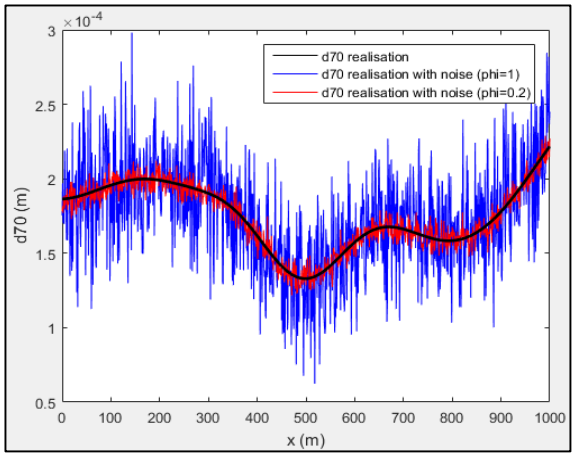


Figure 22 – Example of d70 realisation that is noised with $\phi = 1$ respectively $\phi = 0.2$

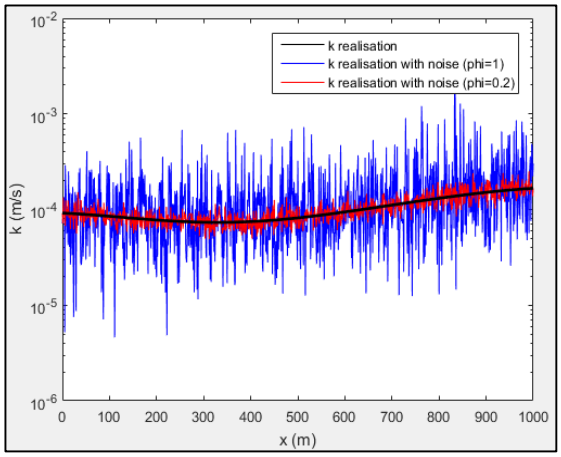


Figure 23 – Example of k realisation that is noised with $\phi = 1$ respectively $\phi = 0.2$

Sensitivity analysis *(see appendix 14)* showed that the accuracy of assessments is sensitive to both measurement error in d70 and k data. For both parameters holds that an increasing measurement error results in more conservative strength. This is explained by the increasing variance in measured data. Because the variance in the actual data is less than measured, the characteristic value analysis underestimates actual representative parameter values.

In Figure 24 two bandwidths are show to highlight two scenarios with respect to measurement error:

- Measurement error is absent, phi = 0.
- Measurement error is present, phi = 1 (with respect to phi is 0.5 in the reference input configuration), meaning that the noise variance is equal to the spatial variance.

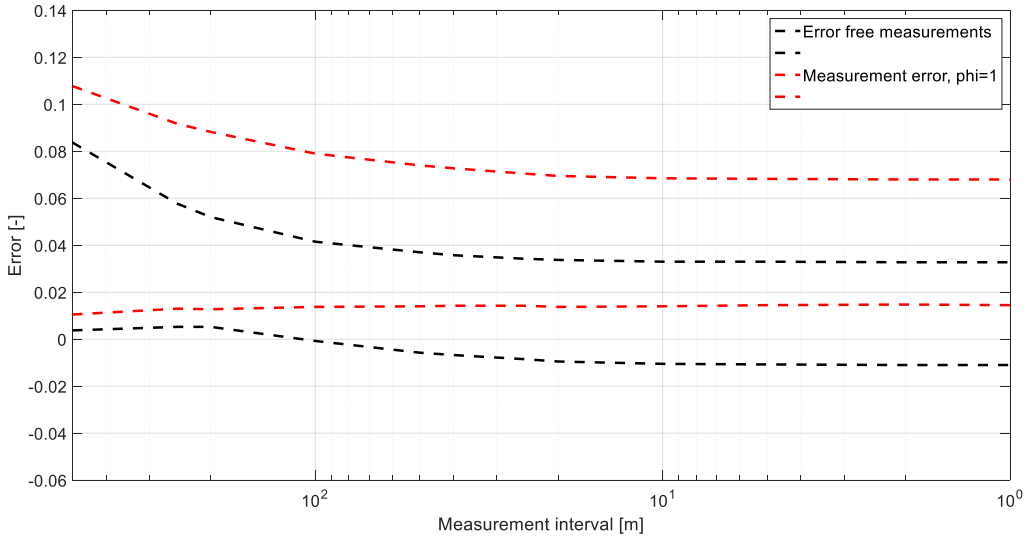


Figure 24 – Comparison of simulation result with and without measurement error

5.3. Correlation structure

In this section the influence of the correlation structure of d70 and k is evaluated. The correlation of data is determined by the shape of the correlogram and the correlation length.

5.3.1. Correlogram

A Gaussian correlation structure is assumed to be representative to model spatial variations in d70 and k data (chapter 3). Alternatively, an exponential correlation structure can be used to generate data sets. Example correlogram of both Gaussian and Exponential autocorrelation functions are presented in Figure 25. The scale of fluctuation 'δ_u' is defined as the area below the function: $\delta_u = \int_{-\infty}^{\infty} \rho_{x,x+\delta}(\delta) d\delta$ (Vanmarcke, 1983). The scales of fluctuation in Figure 25 are equal and therefore Gaussian and Exponential correlation structures can be compared.

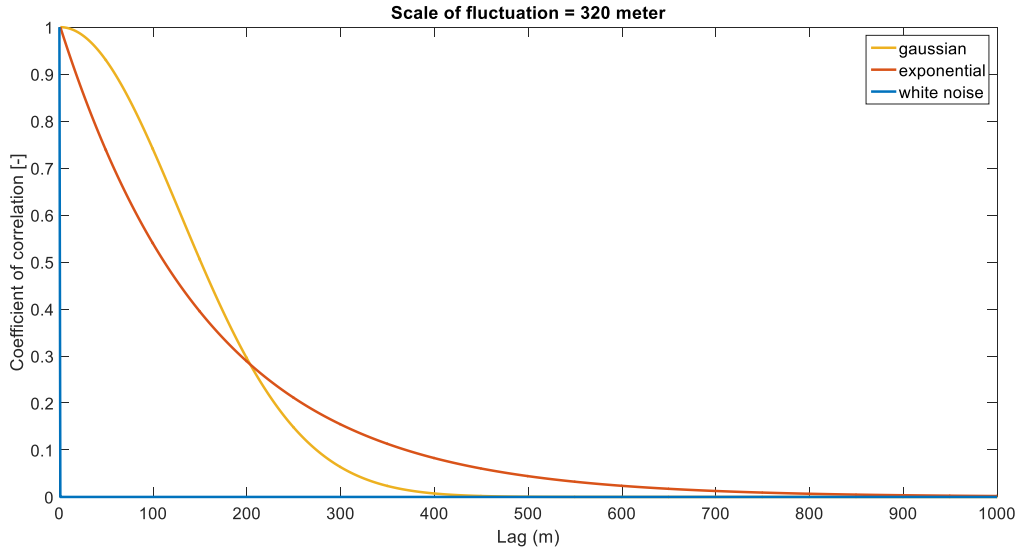


Figure 25 – Gaussian and Exponential correlogram with a comparable scale of fluctuation. The correlation structure of white noise (uncorrelated) is plotted as reference.

The exponential correlogram is steeper for small lags causing fluctuations to be less smooth than for Gaussian structures (white noise effects). Small scale and sudden variations in soil properties are therefore better described with exponential correlation. Figure 26 provides a Gaussian as well as

Exponential correlated data set. Both are based on the same random signal (the white noise in the same figure).

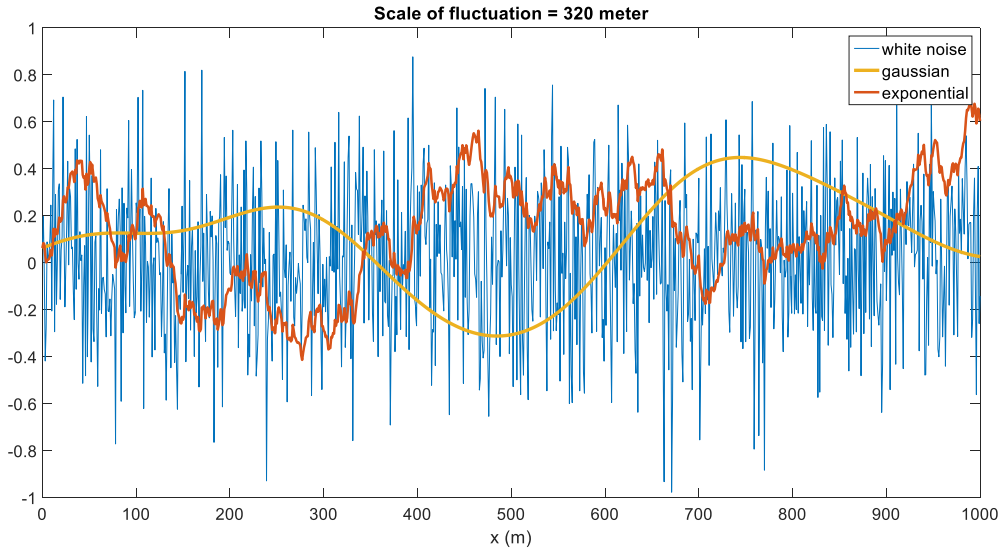


Figure 26 –Gaussian and Exponential correlated data. Both based on the same data signal and equal scales of fluctuation

Several data sets show that Gaussian correlated data needs more space to fluctuate and therefore shows on average less variance per data set (Table 7). This effect increases with increasing correlation lengths.

Table 7 – Effect of correlation structure on average variance in data sets (1000 data points) with scales of fluctuation of 320m (d70) respectively 1040m (k). The variance of white noise (uncorrelated) is given as reference.

	$\mu(\sigma_n)_{d70}$	$\mu(\sigma_n)_k$
Gaussian	2.4E-5	5.7E-5
Exponential	2.5E-5	7.6E-5
White Noise	3.0E-5	1.5E-4

The influence of correlation structure to the accuracy is illustrated in Figure 27. For now, measurement error is neglected to make the influence of correlation better visible. For exponentially correlated data the assessment is relatively less conservative. Exponential correlation results in more and faster fluctuation pattern. Actually the net correlation is less than for Gaussian correlation which is much smoother and needs more fluctuation space to reach the potential variation. This makes the probability of an actual weak parameter value to be present in the domain or coincide with a weak spot of another parameter higher in case of exponential correlation. It is noticed here that the probability of overestimation increases if soils fluctuate spatially.

The effect of measurement density is more or less equal in both correlation structures with a decreasing uncertainty range towards a measurement interval that is equal to the typical correlation length. Because correlation is less dominant in exponential correlation, the trend reversal in the lower boundary is less striking than in case of Gaussian correlation.

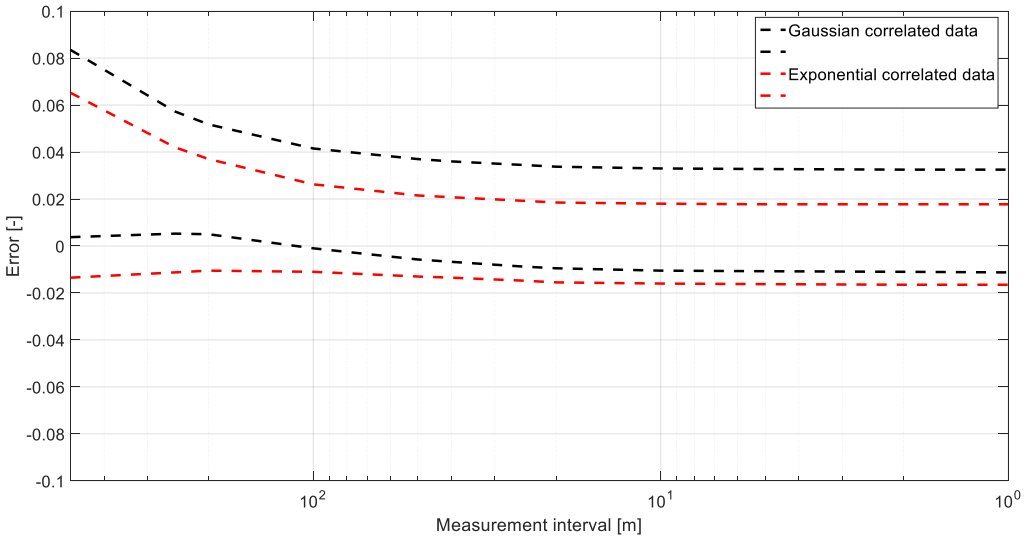


Figure 27 – Influence of correlation structure to the bandwidth of errors. Measurement error is neglected in both simulations.

5.3.2. Correlation length

The correlation length determines to what distance data values influence each other within one data set. Figure 28 provides four data sets with different correlation lengths in case of Gaussian correlation. All four sets are based on the same input signal and a Gaussian structure to be able to make a comparison. Increasing the correlation length (in a limited domain length) means less fluctuation and a decreasing output variance compared to the input signal. The data sets in Figure 28 show that data with little correlation has typically a wider range than data with high correlation. If correlation is absent or small, is resembles white noise. The probability of extremes is then higher.

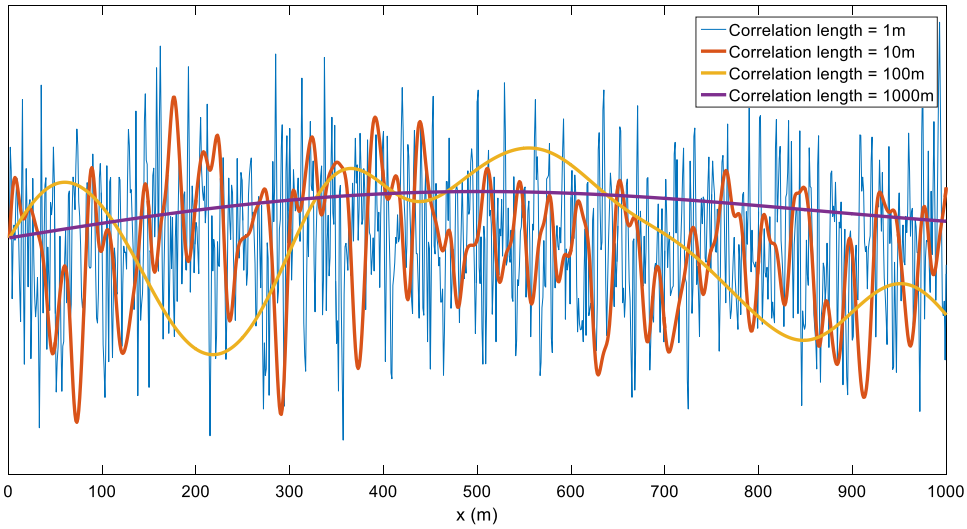


Figure 28 – An example signal, prone to 4 different correlations lengths.

The influence of (lack of) correlation to the error is illustrated in Figure 29. The figure provides the result of 2 different simulations both representing a different correlation scenario (measurement error is again neglected for this case). In the reference scenario standard correlation lengths of 180 (d70) and 600 (k) meter are used. The *uncorrelated* scenario represents a situation with the least correlation

considered plausible. From chapter 3 is known that this is 1m (d70) and 30m (k). Note a correlation length of 30m for k is assumed minimal as the dike width in this case is 30m.

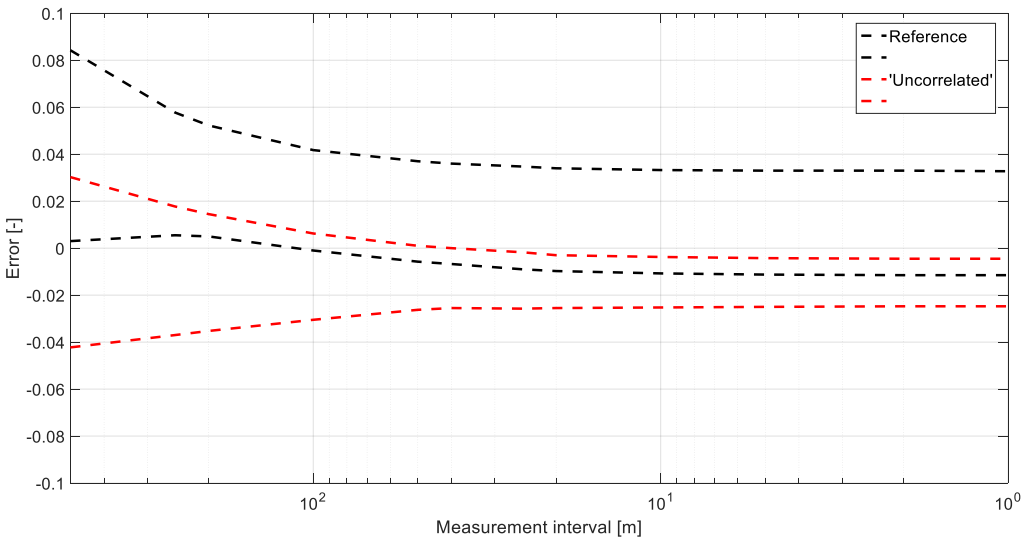


Figure 29 – Influence of correlation length to the bandwidth of errors. Measurement error is neglected in both simulations.

The two bandwidths presented in Figure 29 show that lack of correlation results in relatively less safe assessments compared to the reference case. Analogue to the previous section this is explained by the increased probability that an outlier is present in the data or that unfavourable parameter values coincide at one location. Referred to as 'length effect'. If properties have little spatial correlation, the 95% characteristic lower or upper boundary are often not safe enough.

High measurement densities have more effect in the uncorrelated case as the range of errors becomes smaller. With more measurements the precision of the assessment is increased significantly as the absolute range of the error can be decreased with about 70%. Errors due to the boundary effect are mainly absent in uncorrelated cases. Which means that it is more likely that correct estimation of mean and variance result in the actual 95% lower or upper characteristics (see section 4.3.3.). Application of a safety factor in an uncorrelated scenario can compensate the bias towards unsafe strength assessments.

Dike section length

The length of a dike section has the same influence to the accuracy as the correlation length. Long dike sections allow for more fluctuations and variance in the data, which increases the probability that unfavourable parameter values of combinations of parameter values appear (length effect). In fact, not the correlation length or dike section length is of importance, but the ratio between them. This ratio determines whether boundary effects take place that transform data distributions at limited domains. *In appendix 13 simulation results for several dike section lengths are provided.*

5.4. Distributions of parameter values (range)

In this section the sensitivity of the accuracy to the range of d70 and k values, determined by the model input mean and variance, is evaluated.

5.4.1. Average grainsize

In the standard input configuration an average d70 of 2.0E-4 meter and average k of 1.4E-4 meter per second is used. According to Table 8 this could be indicated as *very fine* or *medium fine* sand. Due to variation, the range of sands covers *extremely fine* to *medium coarse* sands as well. Piping sensitivity

is often related to fine sands as the resistance to erosion is small. However, dikes can be sensitive to piping as well as coarse sands are present. In de Visser et al. (2015) and Aguilar-Lopez et al. (2016) examples are given.

Table 8 – Grainsizes, sand medians and permeability's (meter/day) of different sand classes (Bot, 2011)

Grainsize	Sand median micro meter	Without sludge	Weak sludge containing	Heavy sludge containing
Extreme fine	63-105	**3**	**2**	**0.5**
Very fine	105-150	**6**	**4**	**1**
Medium fine	150-210	**15**	**10**	**3**
Medium coarse	210-300	**30**	**20**	**5**
Very coarse	300-420	**55**	**35**	**10**
Extreme coarse	420-2000	**250**	**150**	**50**

Sensitivity analysis (Appendix 16) shows that the influence of measurement density on the accuracy of an assessment is dependent of the input mean of d70. The error ranges show similar trends in simulations that represent finer or coarser grainsize situations. The absolute magnitude of errors is however affected. This is because equal absolute deviations at different magnitudes of d70 and/or k values have a different impact on the calculated strengths in absolute terms (non-linearity in Sellmeijer model).

The bandwidths of different sand class scenarios are illustrated in Figure 30. A coarser sand scenario (Table 9), based on *de Visser et al. (2015)* and *Aguilar-Lopez et al. (2016),* is compared to the reference result (measurement error is again neglected). Only the means of d70 and k are increased while the relative variation (Coefficient of Variation) is kept constant. In Table 9 the average reference strength in terms of critical head is provided as well. This shows that the larger grainsizes result in increased piping resistance, although the increase in permeability, as a consequence of larger grains (Table 8).

Table 9 – Statistics of two sand class scenarios, of which simulation results are provided in Figure 30

	Reference		**'Coarse sand'**	
	mean	**CoV**	**mean**	**CoV**
d70 (m)	2.0E-4	0.15	3.5E-4	0.15
k (m/s)	1.4E-4	1	3.0E-4	1
Average reference strength – H_c (m)	2.3	0.3	3.8	0.3

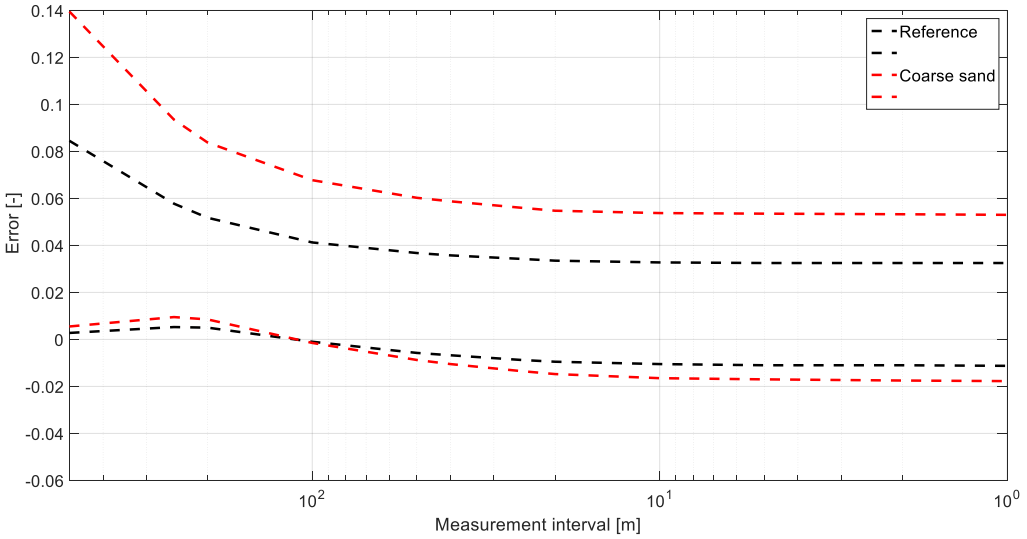


Figure 30 – Influence of average grainsize to the bandwidth of errors. Measurement error is neglected in both simulations.

From Figure 30 is learned that existing inaccuracies are expanded. Positive errors get larger positive values, negative errors get larger negative values and accurate values are unaffected. Expansion of errors is because the same absolute difference between an actual representative parameter values do not result in the same absolute errors. A deviation at a dike with large grains and therefore more resistance to piping results in a higher error between reference strength and assessed strength than if that same deviation is present at a dike with small grains and a small resistance to piping. This is illustrated with a numeric example:

	'Fine sands'	**'Coarse sands'**
$\mathbf{d70_{representative}}$	2.0E-4	4.0E-4
$\mathbf{d70_{characteristic}}$	1.5E-4	3.5E-4
$\mathbf{Deviation_{d70}}$	0.5E-4	0.5E-4
Reference strength	2.92	7.71
Assessed strength	1.95	6.39
Δ (=Reference – Assessment)	0.97	1.42
Error (=reference-assessment/L)	0.032	0.047
$\Delta_{\mathbf{relative}}$ (=Δ/reference)	0.33	0.18

Note that the absolute error is higher for coarse sands. However, the relative error is smaller. Those effects are explained by the non-linear dependency of the strength to d70.

Stronger dike sections due to larger grains or lower permeability's have generally a wider range of possible errors. This accounts for all other parameters that determine the strength to piping (Sellmeijer's model). For example, the seepage length L is considered. An error in the determination of the representative d70 value has more effect when the seepage length is longer and therefore the resistance to piping than if that same error is made at a dike with small seepage length and therefore a lower resistance. This is again illustrated with a numeric example:

	‘Small dike’	‘Wide dike’
$\mathbf{d70_{representative}}$	2.0E-4	2.0E-4
$\mathbf{d70_{characteristic}}$	1.9E-4	1.9E-4
$\mathbf{Error_{d70}}$	0.1E-4	0.1E-4
L	30	60
Reference strength	2.92	5.34
Assessed strength	2.72	4.97
Δ (=Reference – Assessment)	0.20	0.37
Error (=reference-assessment/L)	0.0066	0.0062
$\Delta_{relative}$ (=Δ/reference)	0.068	0.069

Note that the absolute difference increases but the relative error stays in this case more or less equal. This shows that the dependency of the piping strength to L is almost linear. This is not surprising as Sellmeijer (2011) actually calculates a critical gradient. Multiplying this gradient with L results in the critical head. However, the ratio D/L has also a small dependency in the calculation of the critical gradient.

5.4.2. Field variance

In the previous section the sensitivity to the average grainsize is analysed with a constant coefficient of variation. In this section the input variance is varied with a constant mean to analyse the sensitivity to the coefficient of variation (Figure 31).

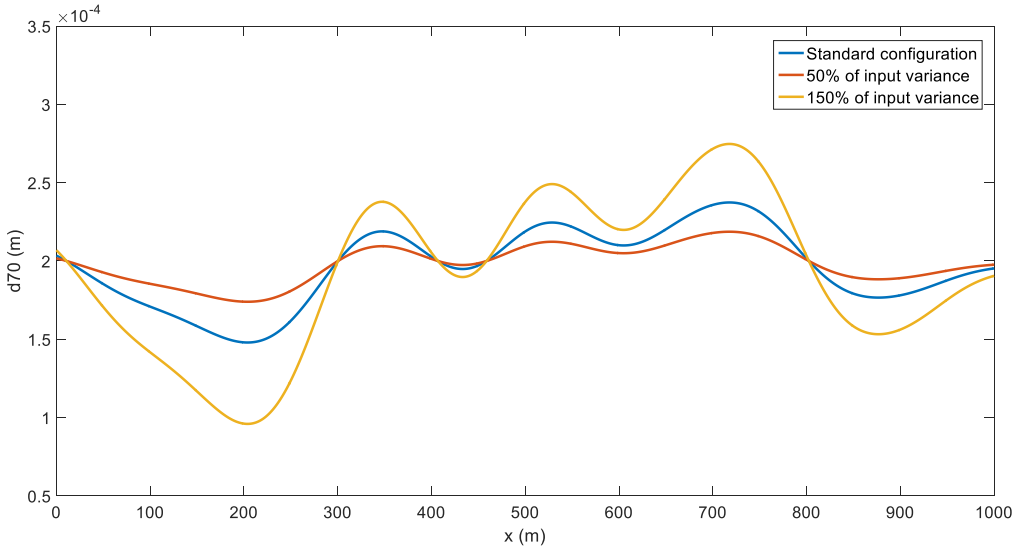


Figure 31 - Example realisations of d70 data generated with respectively 50% and 150% of the input variance in de standard configuration.

Several simulations (appendix Xx) show that the bandwidth of errors is generally wider if the variation in data is relatively high. Small coefficients of variation result in a relatively small bandwidth.

The effect of variation is illustrated with a reference and two example scenarios from which the input is based on the measured mean and variance in the d70 testing grounds of *Veessen* and *IJzendoorn* (de Visser et al., 2015). In all simulations the d70 data is assumed to have a lognormal distribution and for simplicity k is assumed constant at a value that is in accordance Bot (Table 8). Parameter statistics of

the three scenarios are provided in Table 10 and the bandwidths related to the different scenarios are presented in Figure 32.

Table 10 – Statistics of 4 field variance scenarios

	'Veessen'			'IJzendoorn'			Reference		
	mean	**CoV**	**Minimum value**	**mean**	**CoV**	**Minimum value**	**mean**	**CoV**	**Minimum value**
d70 (m)	4.2E-4	0.5	3.8E-5	4.2E-4	0.25	1.3E-4	4.2E-4	0.15	2.1E-4
k (m/s)	3.5E-4	0	-	3.5E-4	0	-	3.5E-4	0	-

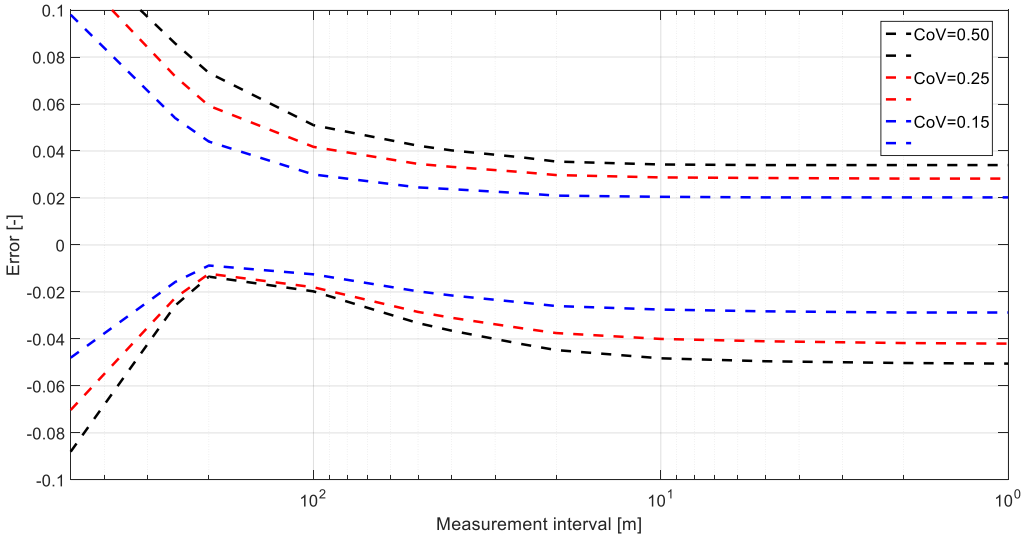
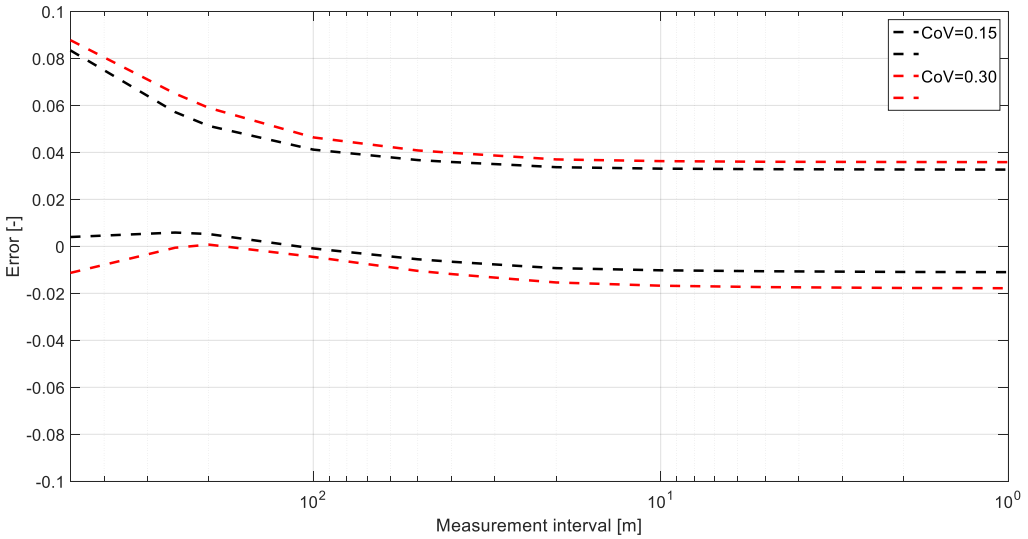


Figure 32 – Influence of field variance to the bandwidth of errors. Measurement error is neglected in both simulations.

The increased bandwidth of errors is explained by the possible deviations between actual representative values and measured values. If the variance in data is high, the possible deviation of an extreme value estimation can be high as well.



5.5. Relative error

All the previous results are presented with absolute errors. Result is that the upper and lower boundaries of the uncertainty range depend on the reference strength and how this reference results from the (spatial) distribution of parameters. At dike sections with high reference strength values,

underestimation of strength (high positive error) is possible. In data sets with a relatively low reference strength, overestimation of strength (high negative error) is possible. It appears that if the measurement density is low, the relative error can be up to 100%. A relative error close to 100% means:

- A dike section has a high reference strength of for example 5m in terms of critical head. With a measurement interval of 500m there is a probability that, due to statistical uncertainty, measurement error and spatial variability, an assessed strength in the order of 0.1m is found. So the error, difference between reference and assessment, is then in the same order as the reference strength. It means that if a strength value close to zero is found it cannot be stated that the section is actually sensitive to piping. In fact, it is possible that the resistance to piping is up to 100 times the assessed strength. But it is also possible that the assessment is accurate.
- A dike section has a low reference strength, in the order of 0.1m critical head. With a measurement interval of 500m there is a probability that an assessed strength in the order of for example 5m is found. The error is then in the same order as the assessed strength. Which means that a dike which is considered insensitive to piping, can be in fact very sensitive to piping.

Note that a reference strength of 0.1m is possible due to the random nature of the simulation. Dikes that are that sensitive to piping are most probably in reality already failed. However, this theoretical analysis does show that not only the absolute error can be high, but the relative error as well. Which is opposed to the intuitive idea that at weak sections a small error is made and only at strong section a high error is made.

The high relative errors follow from the possible high variability in strength values within one dike section. With high measurement densities the upper boundary of relative errors can be brought back to about 50% due to a reduction of statistical uncertainty.

5.6. Summary and conclusions

Uncertain soil conditions are modelled with a set of statistical characteristics. Where a specific error depends on the soil conditions per dike section, the bandwidth of errors depends on the average assumed (spatial) distributions of the soil properties. From the sensitivity- and scenario analyses is therefore concluded that no general or absolute statements can be made about the expected bandwidth of errors.

If in the detailed assessment, measurements are the only source of information, the average characteristics (distributions) are unknown. Therefore, the bandwidth of errors that needs to be considered is even wider than the single bandwidths calculated per individual scenario. It is concluded that the detailed assessment method is even more unreliable than was expected based on the results of the reference scenario. The total uncertainty is given by the combination of error bandwidths from a whole range of plausible soil condition scenarios.

It is concluded that trends related to the influence of measurement density appear to be general, independent of the way soil parameters are(spatially) distributed. Next, the sensitivity analysis shows important trends in the influence of spatial autocorrelation. With decreasing correlation or increasing section length, the reference strength is generally smaller. This is mainly due to the increased probability that an outlier is present. Outliers are not always covered by the conservative assumptions in the strength assessment. Due to outliers it is well possible that the probability of failure is actually higher than expected. Therefore, it is concluded that the ratio correlation length/section length is important to consider in the assessment. Furthermore, the range of errors is generally smallest close to the measurement density that is equal to the typical correlation length. If correlation lengths of typical soils are known, this can give rough indications to what extend more measurements can be valuable in uncertainty reduction.

6. Usefulness of point measurements in increasing the accuracy

Is it possible to increase the accuracy of strength estimations by using information of point measurements alternatively?

6.1. Introduction

In chapter 4 the influence of measurements on the accuracy of strength assessments is evaluated using the characteristic value analysis prescribed in the detailed level of the safety assessment. However, the safety assessment includes also an advanced assessment for which no specific criteria are given (Ministerie van Verkeer en Waterstaat, 2007). This advanced assessment provides the option to use data from point measurements alternatively.

The advantage of the characteristic value approach is the ability to deal with small number of measurements. The disadvantage of a small number of measurements is the likely possibility of errors (chapters 4 and 5). Instinctively the most logical solution to improve the accuracy is to increase the amount of data. The analysis of chapter 4 however shows that with use of statistical descriptions to estimate representative parameters, only statistical uncertainty can be reduced. The other option to improve estimations is to use the point measurements alternatively.

Chapter 6 is devoted to investigate on the potential of point measurements to reduce the range of errors and make estimations generally more reliable.

6.2. Translation of measurements to representative calculation values

In this section different approaches for using point measurements to estimate representative parameter values are evaluated. The following four approaches are tested:

1. *Characteristic value analysis*
   Each data set is translated into a parameter specific characteristic value. The characteristic values are used to estimate one representative strength. This approach equals the standard detailed assessment as used up till now.
2. *Conservative value analysis*
   Each data set is translated into a parameter specific characteristic value. But if one of the measured values is more unfavourable than the characteristic value, the more conservative measured value is used in the Sellmeijer model.
3. *Measured critical value analysis*
   From each set the most critical value is derived (smallest measured in case of d70 and highest measured in case of k). The most critical of measured values is used in the Sellmeijer equation.
4. *Measured strength analysis*
   Per location, a uniform subsection of 1-meter length, a set of one d70 and one k value is obtained by measurements. This set of parameter values is used in the Sellmeijer model to calculate the strength for that specific location/subsection. The measurement density determines how many locations along the dike section are evaluated. The location from which the combination of a d70 and k measurement the smallest strength is calculated, is assumed representative for the considered dike section. This means that parameters are not considered individually, as in the other 3 approaches, but coherent to other parameters and specific locations. In this approach is looked how many subsections should be assessed to find the subsection that is representative for the entire section. In this measured strength analysis, the method to assess the strength is the same method as the reference strength is determined in this study.

In the following subsections the results of the four approaches are given. The analysis is made for a reference scenario (see chapter 3) with error free measurements. This means for now only statistical uncertainty and spatial variability are considered. In section 6.3 measurement error is explicitly considered.

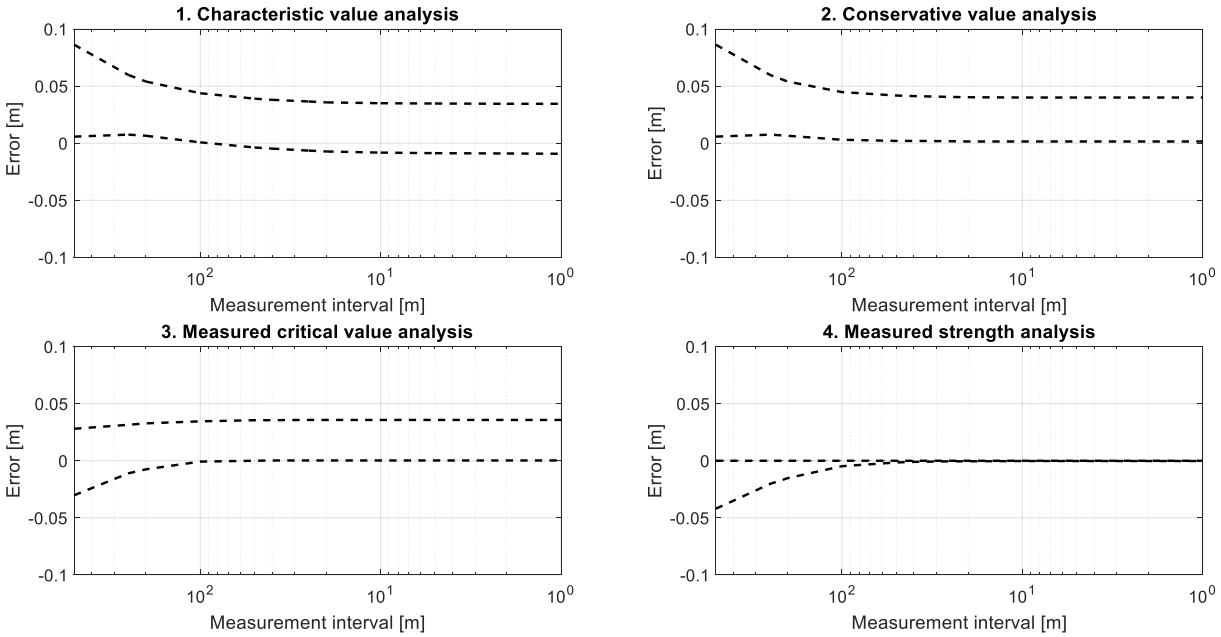


Figure 33 – bandwidth of errors for 4 different approaches to translate measurement into representative parameter values

6.2.1. Characteristic value analysis

The result of the approach based on characteristic value analysis is presented in plot 1 of Figure 33. The result is already extensively analysed in chapter 4. Here it functions only as a reference to the other approaches.

6.2.2. Conservative value analysis

The result of the approach based on conservative value analysis is presented in plot 2 of Figure 33. In this approach the individual measured values are considered explicitly. With increasing number of measurements, the probability increases that the most unfavourable value (or a value close to that) is measured. This method uses measurements explicitly to replaces optimistic characteristic values with more pessimistic values if those are actually measured. It is shown that this method prevents overestimation as the lower boundary of the bandwidth reaches the value zero.

Drawback of this method is the increased underestimation for high measurement densities with respect to characteristic value analysis. Because of spatial variability, the most critical values of individual parameters can underestimate the actual strength. In a characteristic value analysis this is less the case because only a 95% boundary is used. This effect is visible when comparing the upper boundaries of plot 1 and 2 of Figure 33.

6.2.3. Measured critical value analysis

The result of the approach based on measured critical value analysis is presented in plot 3 of Figure 33. In this method no characteristic values are used. From each parameter set the most critical (unfavourable) value is selected. For a small measurement density, the probability that an unsafe value is selected is high as the available measurements are not a representative sample. For a high

measurement density, the probability of overestimation becomes (almost) zero. With the distribution of values completely described (instead of estimated by a shape, mean and variance) the actual critical values can be estimated with great accuracy. The statistical uncertainty, including boundary effect, vanishes for high measurement densities.

In the detailed assessment the implicit assumption is made that critical parameter values appear at the same location. The measured critical value analysis gives insight in the range of errors as result of that assumption. At high measurement density there is no statistical uncertainty because from both d70 and k the most unfavourable values are known. However, that values do not have to coincide at one location. But because it is assumed they do, strength is underestimated in many assessments. The bandwidth of positive error values (underestimation) show that the assumption of coinciding unfavourable values is conservative.

6.2.4. Measured strength analysis

The result of the approach based on measured strength analysis is presented in plot 4 of Figure 33. In this method the strength is related to locations. In this way it is possible to banish out all uncertainties with increasing measurement density.

At small measurement densities there is uncertainty about the distribution of strength values resulting in assessed strength overestimating the reference strength (unsafe). Because of the small size of the data set, it is likely that the taken sample is not representative for the entire section. With increasing number of measurements, the set becomes a more representative sample. The most unfavourable (minimum) strength value from a large set of strength values is an accurate estimation of the actual representative strength.

Note that the actual representative strength is called the reference strength in this text and is actually defined as the minimum strength value of a dike section. So if the assessment finds this minimum value, the error compared to the reference strength is indeed zero.

The d70 and k parameters are used to directly calculate the strength at each measurement location. It means only one relevant property (H_c) is left to consider. Measuring H_c values therefore only leaves uncertainty about the distribution of that single property. That uncertainty can be reduced by increasing the number of measurements. Therefore, the range of errors becomes zero for high measurement densities.

The measured strength analysis might suggest to use that method in the assessment instead of the characteristic value method. Plot 4 of Figure 33 suggests that measurements are able to reduce the range of errors to zero from a measurement interval of approximately 50m. However, some drawbacks exist. In practice a few hundred meters is common (de Visser et al., 2015), therefore this interval is already quite small. Notice that in this case the measurements interval of 50m is smaller than the correlation lengths of the data. The hypothesis is that much smaller measurement intervals are needed to provide accuracy in situations where parameters show less or no spatial correlation.

Secondly this method requires that the strength parameter is assigned to each specific location. As a consequence, each measurement location requires an accurate measurement of every relevant property. Only error free measurements result in error free strength values at each measurement location. Therefore, it is necessarily to prevent or reduce measurement error.

6.3. Influence of measurement error to the accuracy

This section studies the effect of measurement error in detail. First is analysed how measurement error influences the accuracy of assessments for the four different approaches. Secondly the possibility to use more point measurements to reduce uncertainty due to measurement error is evaluated.

6.3.1. Effect of measurement error

Figure 34 shows the influence of measurement error in the four different data use approaches. From the analysis is learned that measurement error results in increased inaccuracy, as well bias to underestimation and increased imprecision.

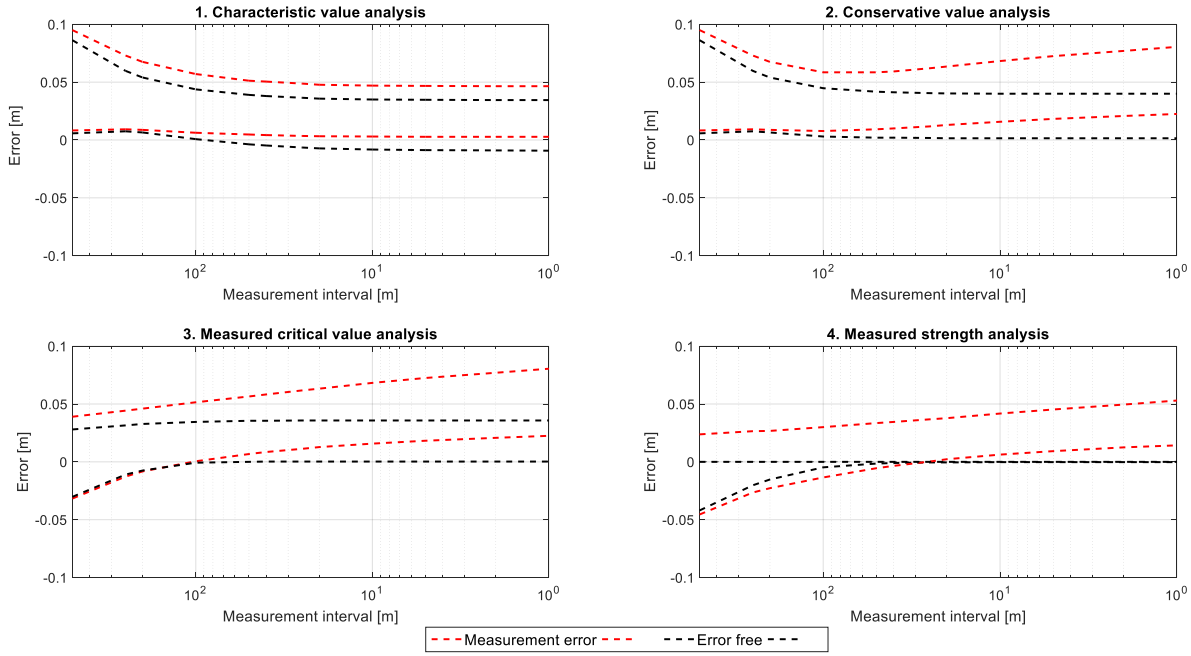


Figure 34 – Influence of measurement error in four data use approaches

Especially in an approach where measured values are explicitly used (approaches 2 and 3), the measurement error can result in the use of more pessimistic values than are actually present. In the measured strength analysis (approach 4), the measurement error causes inaccuracy and increasing conservatism at increasing densities because the probability that a too conservative strength value is measured increases with the number of measurements.

Based on the previous section, the use of measured strength analysis argues for the importance of high measurement densities because then accurate assessments are possible. But when considering measurement error, it appears that even for very high measurement densities, assessments are not accurate (plot 4 of Figure 34).

6.3.2. Reduction of measurement error

The magnitude of measurement error in piping analysis is uncertain, especially with respect to k. Therefore, it is tried to deal with measurement error in a general applicable approach. Increasing the accuracy of measurement equipment is one possibility. However, the accuracy of equipment is beyond the scope of this study. Averaging out the random error by increasing the number of measurements is applicable and is further analysed.

Averaging out error in in sections with uniform properties

First assessments of completely uniform dike sections are analysed. In Figure 35 the result of this analysis is presented for two different data use approaches and two different magnitudes of

measurement uncertainty. In the upper plot the measured data points are simply averaged, resulting in a decreasing range of errors for increasing number of measurements. The slight average underestimation is caused by the logarithmic character of the k parameter. In the lower plot the measured data is used to calculate characteristic values. This plot shows that if the soil has uniform properties, measurement error causes structural underestimation when calculating characteristic values from measured data. This means that in a scenario where properties are highly uniform, measurement error has major influence. From the measurements the idea can rise that the soil is heterogenic. With characteristic value calculation then safe lower or upper boundaries are calculated. These result in much smaller strength values than actually present. The probability of failure of the structure is then smaller than concluded from the assessment. Therefore, it is very important to question whether variance in measured data is because of spatial variation or measurement errors.

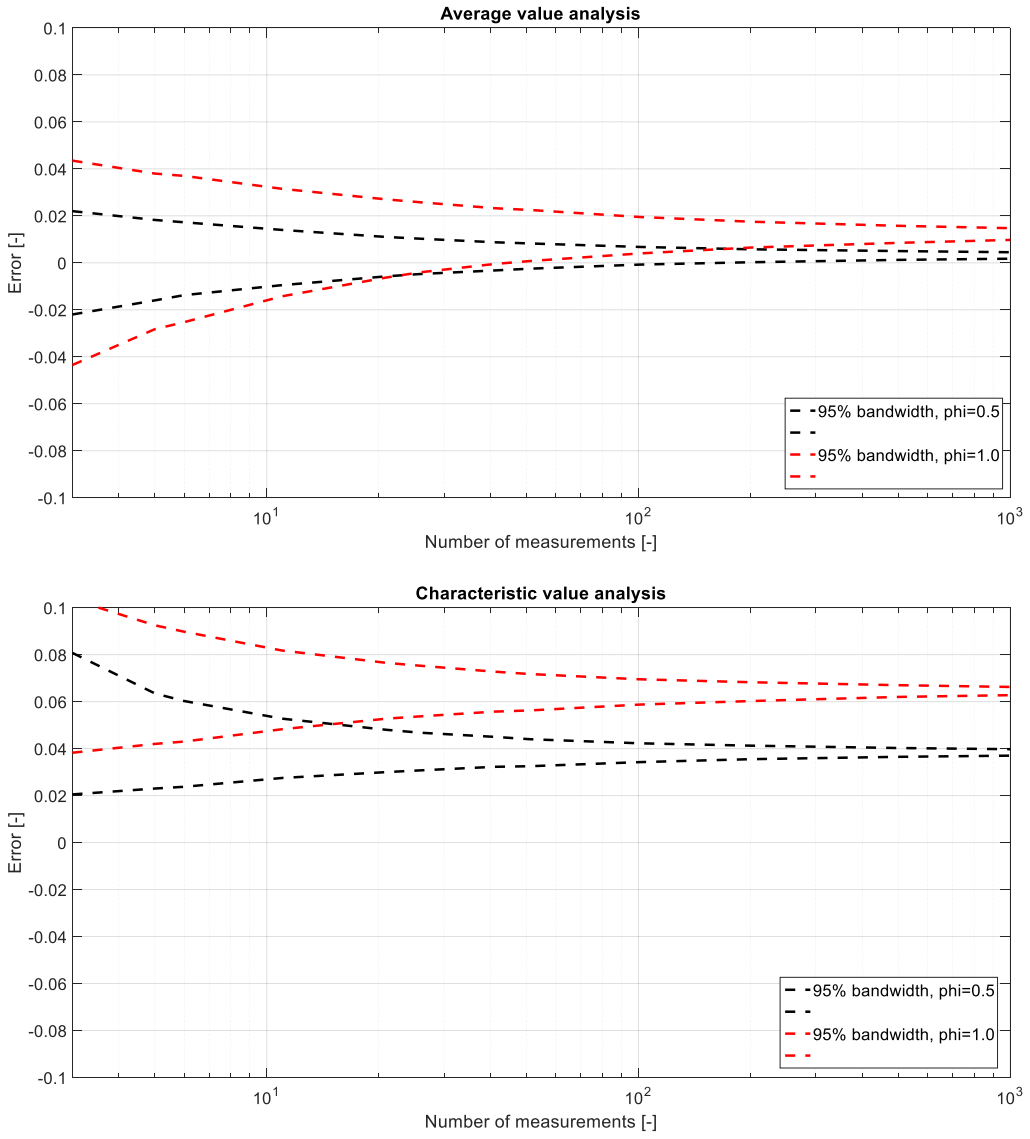


Figure 35 – Effect of number of measurement in case of measuring at a homogeneous dike section with measurement error.

Notice that a lot of measurements are required to average errors and make precise assessments. Besides it is only useful to average measurements if measurement error is randomly and not biased.

Averaging out error in spatial variable soils

From Figure 34 is concluded that measured strength analysis results in accurate assessment for high measurement densities. This accuracy vanishes when measurement error is included/considered. In this section is investigated whether error can be averaged out by taking multiple measurements per measurements location.

Figure 36 shows that accuracy is generally improved when multiple measurements per location per parameter are taken. The errors become smaller and consistency higher. Because the k data is exponentially distributed the original standard configuration without measurement error cannot be reached completely. Also with high numbers of measurements per location the average of exponential noise is higher than the original data realisation.

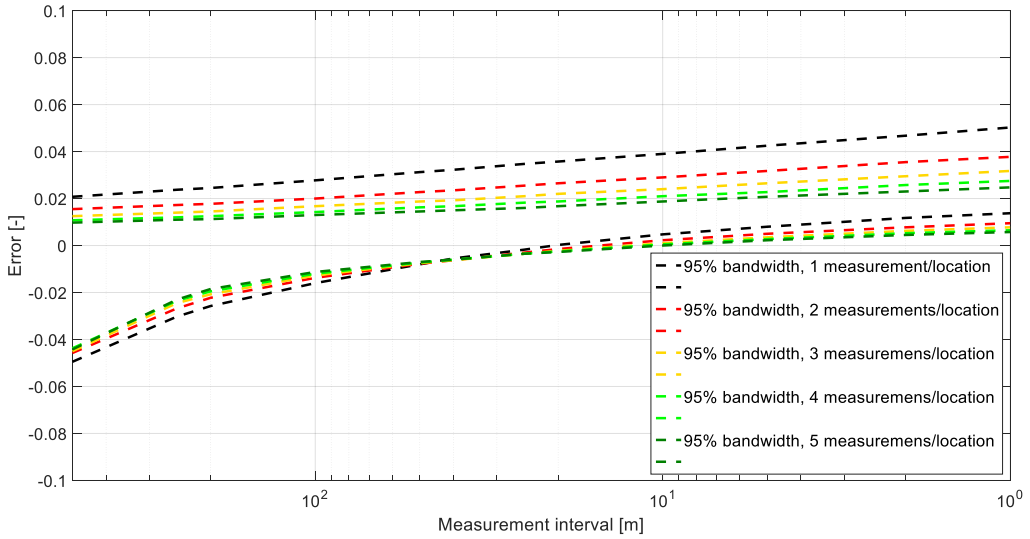


Figure 36 – Effect of multiple measurements per measurement location to average out measurement error (measured strength approach).

6.4. Use of spatial correlation to increase the accuracy

The probability of geo-technical failure depends for a significant part on random system effects, i.e. length effects, parallel system and series system effects. The structure of spatial correlation of soil properties is decisive for these effects (Vrouwenvelder & Calle, 2003). The spatial correlation in parameter values is not explicitly used in the current assessment. However, the safety factor partly accounts for the length effect. This section evaluates the possibility of using correlation lengths implicitly to increase the reliability of the strength assessment

6.4.1. Influence of correlation

The scenario analysis (chapter 5) already showed that the correlation in the d70 and k data affects the bandwidth of errors. Therefore, it might be valuable to know on beforehand of an assessment whether the soil properties are expected to be more or less correlated. If the correlation length is in the order of the domain length (or higher), data is hardly fluctuating but it has a constant descending or ascending trend. Peaks are therefore often at the boundary of the considered domain (illustrated in Figure 37). Measurements at the boundary of the domain will most probably give a correct estimate of the range of the data. If the peak is not at the boundary of the domain, the error is at least small. With strong correlated properties, a measured critical value or measured critical strength analysis can therefore be preferred above a characteristics value analysis.

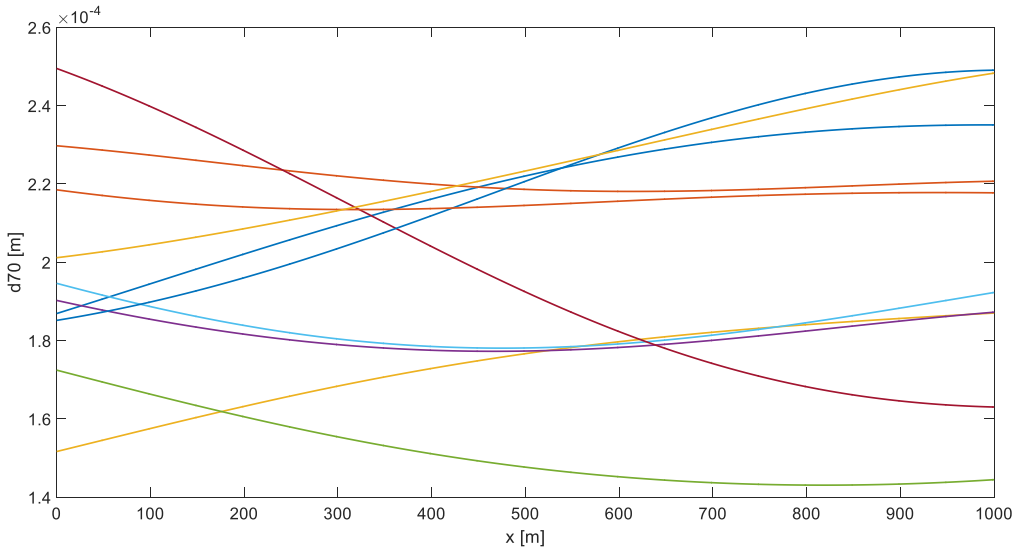


Figure 37 – 10 example data sets with a correlation length in same order as domain length (1000m)

If the correlation length is small with respect to the considered domain, data is fluctuating heavily within that domain. It would be a coincidence if a couple of measurements, for example at the boundaries of the domain, resulted in an accurate estimate of the range. Then it fits better to collect measurements at more or less random locations and translate those into a characteristic value. A safety factor can then be applied to compensate the probability that the characteristic value is underestimating the actual critical value. This probability mainly depends on the fluctuation space which depends on the length of the considered domain. With the standard characteristics value analysis, the theoretical probability of exceeding the 95% characteristic value is 5% per fluctuation. A parameter with a correlation length of 1m, makes about 1000 fluctuations in a domain of 1000m and it is therefore almost certain that the critical value will exceed the characteristic value. The random uncorrelated data set of Figure 38 illustrates this.

Besides the correlation length, the domain length influences the probability of exceeding the characteristic (lower) boundary. If the domain length is increased with a factor 10 (Figure 39), the probability of exceeding the characteristic boundary is increases ten times as well. Therefore, the ratio correlation length/domain length determines the probability of exceeding a boundary value.

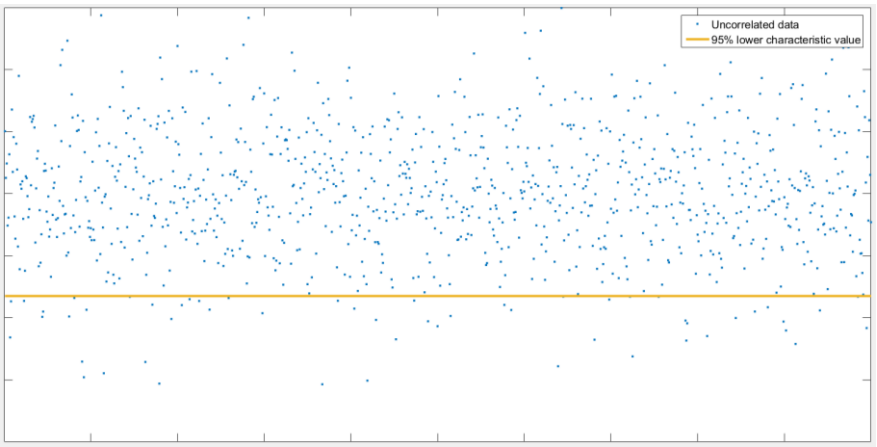


Figure 38 – Random uncorrelated data with 54 out of 1000 data points exceeding the 95% lower characteristics value

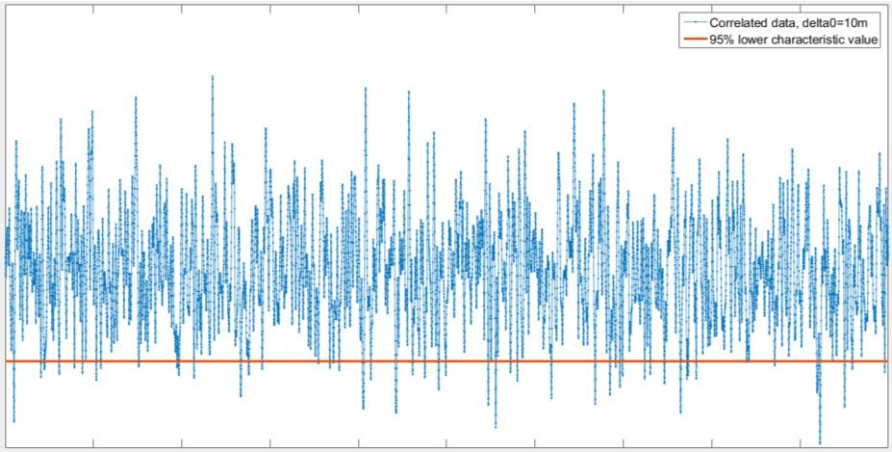


Figure 39 - Random data applied to a correlation length of 10m, with 50 out of 10000 data points exceeding the 95% lower characteristics value

Knowledge of the correlation scale can be beneficial for the decisions made during the assessment as more information is available. However, it is not the key to accurate assessments. It only helps to estimate the expected bandwidth of error and to anticipate on that *(see section 5.3 as well).*

6.4.2. Usefulness of measured correlation

Although much is known about the effect of correlation, very little is known about existing correlation lengths in practice (Vrouwenvelder & Calle, 2003). Correlation of two data points can be expressed with a correlation coefficient which indicates about the similarity of two data values with a certain lag. This is easily determined between a couple of measurements. However, of interest is whether a certain correlation pattern is present in a larger domain, such as a dike section. The question rises whether it is useful to estimate this pattern with point measurements within an assessment.

First of all, it is only possible to estimate the correlation length if the measurement interval is smaller than the correlation length. Furthermore, the measurement domain length has to be much larger than the present correlation length to be able to recognize that pattern. Finally, many measuring points are needed be able to average out possible measurement error. In conclusion: to make reliable estimations of correlation patterns, a very large and dense data set is required. And if this set is available, the statistic *correlation length* loses its value because representative values can also be observed directly from the measurement set. As with all statistical descriptions, they are only useful to make estimation if a small amount of data is available.

However, the quantification of correlation lengths can be useful if a relatively large area is expected to show equal variation patterns. For example, a specific geomorphological deposit or geological unit. With a fine measurement grid a relatively small part of the area is then investigated and the presence or absence of correlation of the whole deposits is characterized by a representative correlation length.

If correlation lengths can be estimated from other sources, such as geological maps, it makes sense to use this information the development of a measurement and data translation strategy. It is however not useful to estimate correlation lengths from measured data that is also meant to assess the strength.

6.4.3. Quantification of correlation

If a large data set is available, correlation within that set can be evaluated with the use of a correlogram. A correlogram provides the *average* correlation coefficients between all pairs of a data set for number of lags. From a correlogram the correlation length in a data set can be estimated.

In chapter 3, unbiased estimates of correlation coefficients are made to check the functioning of data generation. In an unbiased estimate the mean and variance are known and equal to the average or input mean and variance that accounts for a data set with infinite length. However, in practice only a biased estimate can be made. In a biased estimate the local measured mean and variance are used to estimate correlation. In that case the ratio of correlation length and domain length affects the measured correlogram. The measured correlation per section (method of moments) becomes on average smaller with respect to the input correlation predefined correlogram if the ratio correlation length/domain increases (0→ 1). This effect is illustrated in Figure 40 for a ratio of 0.1.

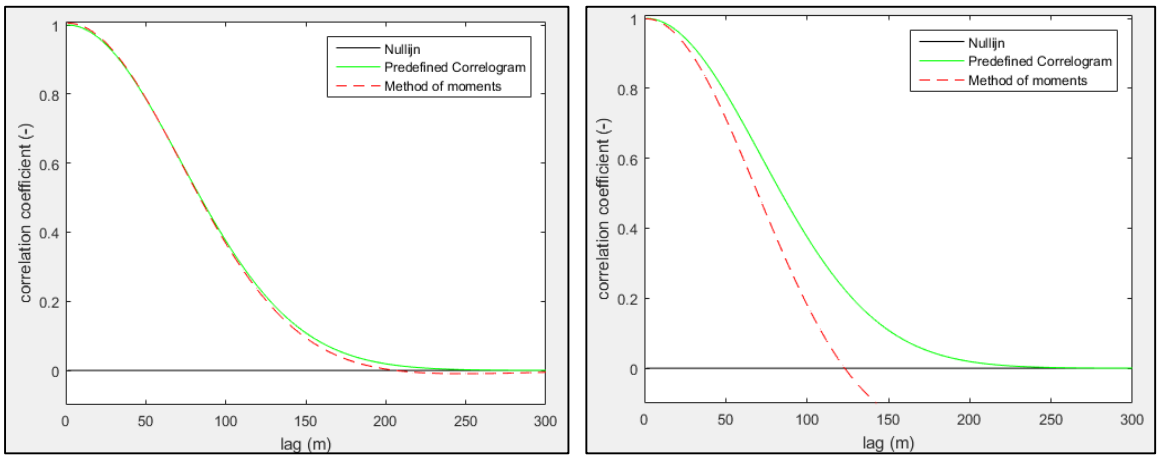


Figure 40 – Left: Average correlogram using the unbiased estimate with Simulated field: N=2000, m=0, s=1 and δ_0=100 m. Right: Average correlogram using the biased estimate. Simulated field: N=2000, m=0, s=1 and δ_0=100 m

Correlation lengts can be estimated as a fraction of the surface below a correlogram. Another way to estimate correlation lengths is to measure the correlation coefficient at one specific lag and estimate the autocorrelation function. For example the Gaussian autocorrelation function can be rewritten to derive the correlation length from a measured correlation coefficient:

$$\rho(x) = e^{-\left(\frac{x}{\delta_0}\right)^2} \rightarrow \ln(\rho_x) = -\left(\frac{x}{\delta_0}\right)^2 \rightarrow \sqrt{-\ln(\rho_x)} = \frac{x}{\delta_0} \rightarrow \delta_0 = \frac{x}{\sqrt{-\ln(\rho_x)}}$$

One should be careful using the information of a correlogram as it presents average correlation coefficients. Sudden variabilities or different correlation regimes within a larger measurement domain are averaged and therefore invisible in one representative correlation length. For example: A sudden anomaly, such as an old river bed, crosses a dike section with further strongly correlated properties in length direction. The soil is on average quantified as correlated and therefore a relatively high measurement density can be used to find representative parameter values. In that case the probability that the actual representative anomaly is missed, is very large. If due to the anomaly the section was quantified as uncorrelated, a characteristic value analysis with safety factor could have been logical. However, because of correlation in the majority of the section, the measured variance would be low, probably resulting in a too optimistic characteristics value. Therefore, average correlation lengths are only useful in areas with constant correlation patterns.

6.5. Summary and conclusions

In this chapter the accuracy of four different methods that translate measurements into a strength value are evaluated. It is concluded that the potential of a very dense measurement grid is insufficiently used as long as the measurements are only used to estimate probability density distributions of parameters individually. An alternative is to calculate strength values at small subsection level and consider spatial variation explicitly. The section representative strength is then defined as the weakest

of all measured subsections. In this approach an increasing measurement density has the following two advantages:

- Ongoing increase of accuracy with increasing number of measurements.
- Location specific insight in strength along the dike length that allows for customized analysis.

It is concluded that these two advantages mainly/only count when measurement error is small or absent and the actual values of parameters can be measured accurately at each location. It is shown that by averaging multiple measurements per location, the inaccuracies caused by measurement error are reduced. Furthermore, it is concluded that the density of measurements needs to be higher than the typical correlation length to make accurate assessments. In cases of strong fluctuating soil properties, the measurement density should be in the order of meters to be sure that the most critical location is actually measured. It is concluded that the required point measurement density to prevent possible overestimation, is infeasible in practice. It is concluded that in general the probability of failure is smaller if limited measurements are translated in a strength estimate by use of the characteristic value analysis (standard detailed assessment).

Knowledge of correlation patterns of soil properties can be useful to narrow down/decrease the range of errors. It is concluded that very large data set is required to make an unbiased estimation of the correlation pattern. It is concluded that it is not useful to estimate correlation length of a section while assessing the same section because the predictive value becomes irrelevant as many data is already available. However, correlation lengths become relevant if they can be assigned/linked to geologic units. The correlation length statistic has then predictive value which can be used to increase the accuracy of the assessment.

7. Discussion

In this chapter the value the meaning of this research is discussed.

7.1. Interpretation of the results

The error that is quantified in this research is about the difference in what we think that is the strength if the data quality is optimal and what we think that the strength is if the data quality is limited. Data quality is in this case about quantity of measurements and noise of measurements. Insight in the error tells us something about how our view of safety depends on the quality of data. In this matter we assume that the view of safety is in better correspondence with reality as the quality of data is high. This because the view of reality is more detailed. The definition of the reference strength as the weakest cross section is in line with this assumption: the reference strength only tells what the representative strength of a longer section would be if every meter dike would have been assessed individually with error free parameter estimates. Therefore, the quantified error is an indication of possible differences between the assessment outcome and the reality. But notice that the actual differences can only be quantified when the dike is loaded. However, with the results of this study it is still possible to get insight in how the perception of safety is influenced by the way we assess our dikes with point measurements and conservative assumptions. Two interesting consequences of errors are considered in an example:

The reference strength of a dike section is equal to the norm. For example, the critical water level is NAP +3m, corresponding to a probability of failure of 1/1000. With an assessment a critical water level of NAP +2m is found. For certain scenario's this is a likely possibility according to the results of this study. The assessor thinks that the probability of failure is higher than allowed and decides to reinforce the dike with a berm of for example 20 meter. After reinforcement the actual resistance increases to a critical water level of about NAP +4m and the probability of failure of that section becomes much smaller than the norm. Although it was not necessarily to reinforce based on the norms, it can be argued that at least the investment resulted in a very safe dike section.

Unfortunately, a dike consists out of multiple sections and because a dike is a series system the weakest link determines the flood risk. Imagine that this dike consists out of two sections. So for the neighbouring dike section the same assessment is made. This section has a reference strength of NAP +2,8m, corresponding to a probability of failure of 1/800. From the assessment follows that the critical water level is NAP +3m as well. The results of this study show that it is a possibility that the reference strength is overestimated. The assessor considers this section as safe as from the assessment follows that the probability of failure is equal to the norm. Consequently, the probability of failure of the entire dike is 1/800, dominated by the weaker section. This not only means that the norm is not met, but also that the reinforcement of the first section is completely ineffective because it does not increase the safety level of the hinterland.

This example shows a possible consequence of an unreliable assessment. The difficulty is that the errors can be in both overestimation (negative errors) as underestimation (positive errors) depending on the soil scenario. Positive errors mean that it is possible that a dike is reinforced unnecessarily. The range of positive errors determines to what extend it is likely that a difference between load and assessed strength can be caused by an error instead of an actual lack of strength. Consider an example in which the load is 1 meter higher than the assessed strength. If the range of errors lies between 0 and 2 meter, the probability is 50% that the actual strength is at least as high as the load and that the reinforcement wat not needed to meet the norm. The relative influence of increased measurement density to the probability of unnecessarily reinforcement is illustrated in Figure 41.

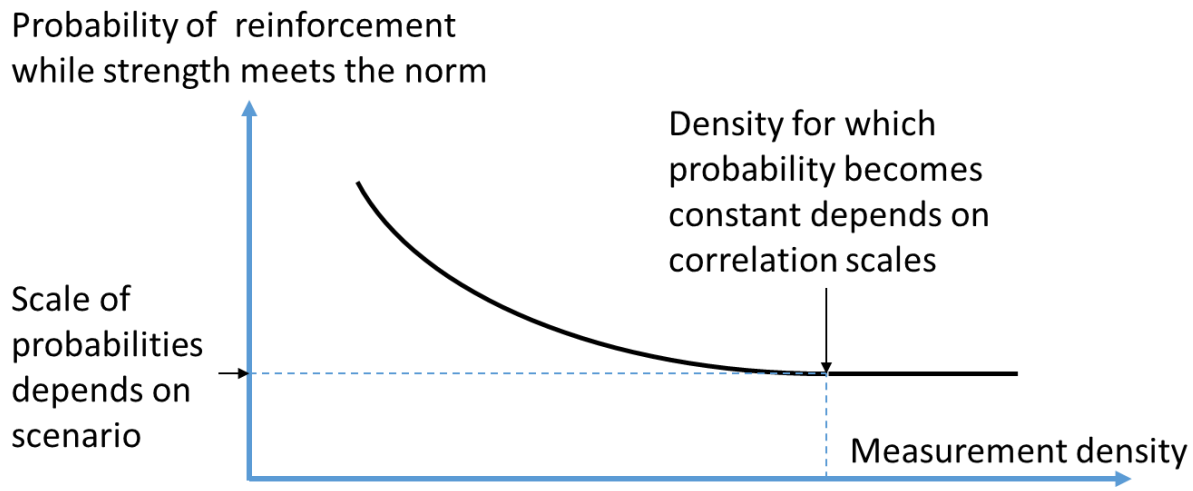


Figure 41 – relative influence of increased measurement on the probability of (unnecessarily) reinforcement

It shows that the probability that a section is unnecessarily reinforced decreases if a higher density is used. But it appears that the probability at a high measurement density is still in the same order as for a low density. Furthermore, the order is different for different scenarios and dependent on the ratio between strength and load. Therefore, it is not possible to generally quantify the probability of underestimating strength nor the probability of unnecessarily reinforcement.

Within a detailed assessment more measurements can reduce the reinforcement task if the assessed strength is smaller than the load. In general, the assessed strength becomes less conservative when using more measurements.

Negative errors mean that the reference strength is overestimated and the probability of failure is actually higher than calculated. Then it is possible that a section that should be reinforced, is not reinforced because it is considered safe. The range of negative errors determines to what extend it is likely that a difference between load and reference strength will not result in a reinforcement. Consider an example in which the load is 1 meter higher than the reference strength. If the range of errors lies between 0 and -2 meter, the probability is 50% that the assessed strength is equal or higher than the load. This means 50% chance that the dike is will fail before the normative load. The relative influence of increased measurement density to the probability of such an unsafe situation is illustrated in Figure 42.

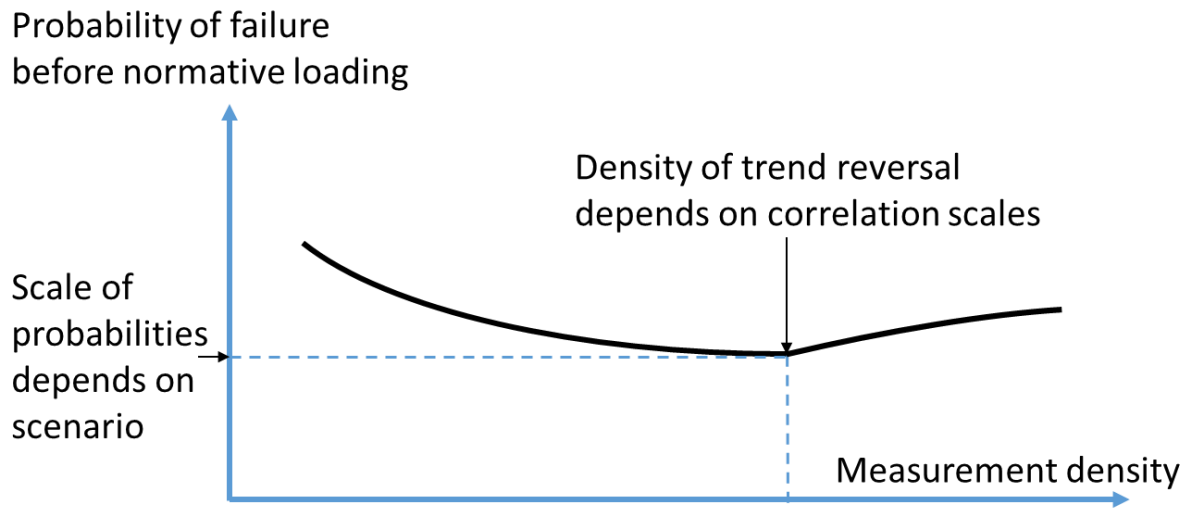


Figure 42 - relative influence of increased measurement on the probability of failure before the normative load is exceeded.

Where the probability of unnecessarily reinforcement decreases or is constant (Figure 41), the probability of failure before normative load is minimum at a certain density that depends on the correlation scale. This insight shows that if the correlation scale is unknown, it is not per definition useful to/you should not blindly increase the measurement density in order to decrease the probability of an unsafe situation.

The order of probability that an unsafe situation occurs is different for different scenarios and dependent on the ratio between strength and load. Therefore, it is not possible to generally quantify the probability of overestimation or probability of an unsafe situation.

In summary, both the probability of unnecessarily reinforcement as the probability of failure before normative load are influenced by the density of measurements. The density with the smallest total range of errors results in the smallest combined probability of one of the two unwanted situations. Whether the probability of unnecessarily reinforcement or probability of failure before normative load dominates, depends on the soil scenario:

In the reference results is shown that the assessments tend to underestimate the actual representative strength. So often dikes are actually stronger than expected. The validly of this result is supported by practices in which dike sections are clearly rejected in the assessment but not considered weak based on experiences with the pas (van Putten, 2013). So, the difference between assessment outcome and expected safety based on experience can possibly be explained by uncertainty in model input. Note that model uncertainty and uncertainty in loads/inexperience with extreme loads can be debit to the discrepancy as well.

Opposite situations in which sections are determined safe by the assessment but considered weak by practical experience are apparent as well. Cases exist in which sand boils are observed at relatively low water levels. This is a sign of increased piping risk, but from the assessment a safety factor of above 1 is found. This might be explained by high fluctuation scales in soil parameters, because then the assessment tends to overestimate strength. However, when considering parallel effects due to heterogeneity, significant overestimation is not very likely (see section 7.3 of this chapter). An explanation is found in the presence of anomalies. Anomalies are not considered explicitly in this research. But it can be reasoned that in a trajectory with a constant variation pattern, a sudden weak spot in undetected by a characteristics value approach. If the weak spot is for example an old giver gully that is completely intersecting the dike width, a dangerous situation occurs. At that spot sand

boils might develop while the measured variance from the rest of the section might be too low to calculate a sufficiently conservative d70 or k value. This research mainly shows the likeliness of underestimation and unnecessarily reinforcements in soils with constant variation patterns. Insight in the characteristic value analysis and nature of point measurements argues that overestimation and increased probability of failure is likely in case anomalies are present. Note that the importance of the measurement density is actually minor: for every density a range of errors exist and it is unknown of the estimated strength overestimates or underestimates and to what extent. Tco estimate the piping risk adequately it is much more important to identify the variation pattern of the important soil parameters and identify the presence of anomalies.

7.2. Expansion of the research scope

In this research the scope is limited to the analysis of uncertainties in parameters d70 and k. Within a simple geometry without covering clay layers ('schaardijk' concept) one other parameter, the sand layer depth D, is considered spatially variable in the assessment. This means that the value of D contains the same type of uncertainties as d70 and k. The parameter D is however neglected because it has a rather small influence to the sensitivity of a dike to piping. The effect of incorporating this extra uncertain parameter in relation to d70 and k is showed in Figure 43.

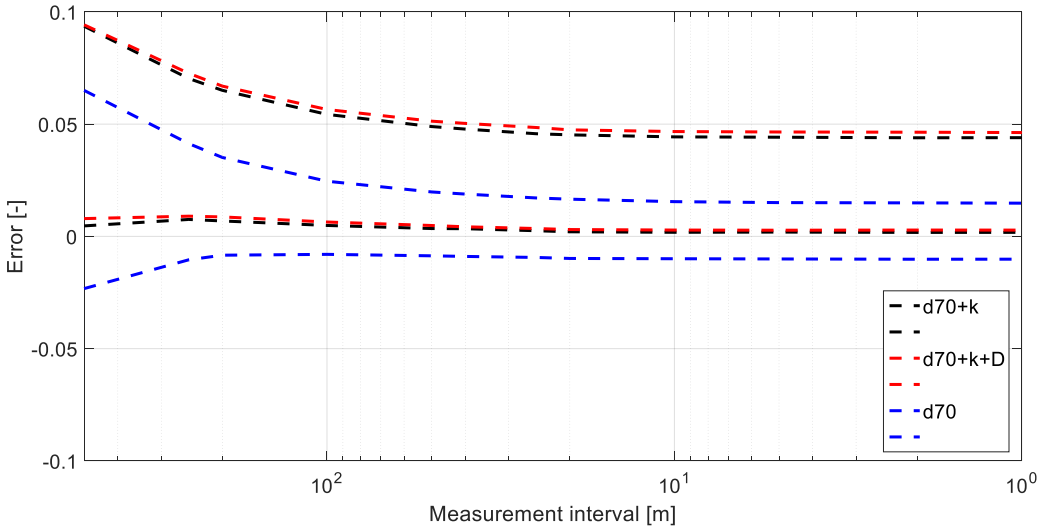


Figure 43 – Contribution of d70, k and D to the bandwidth of errors in a standard input configuration. Input of D: mean=15m, standard deviation=3m, correlation=200m (according to PC-Ring calculations)

Most obvious is that the influence of D to the error is very small with respect to the influence of k and d70 which again confirms the relative importance of that two parameters in the piping assessment. From the figure also follows that the bandwidth of errors increases and the assessment is on average more conservative when more parameters are considered. This is because the probability that unfavourable conditions coincide then decreases.

In practice, at many potentially piping sensitive dikes a covering clay layer is (partly) present. The clay layer is of influence to the Sellmeijer model parameters d and L. Especially the determination of a representative L value is often subject of discussion (POV piping; werkplaats Zwarte Water, 2015). Because the presence, depth and continuity of a clay layer are uncertain, a conservative attitude is common to the contribution of covering clay to the resistance. But if covering clay is actually present it can result in significant increase of the seepage length. It is expected that uncertainty related to clay covers can cause significant deviations between estimated strength and actual strength.

Incorporation of more uncertain parameters results in less precision and more bias to conservative assessments. It is expected that this also accounts for parameters L and d as those are as well determined with measurements and characteristic value analyses. However, quantifying the effects of clay cover is relatively complicated because a 2D analysis is necessarily. Next, the determination of a reference strength is not straightforward as a representative seepage length is not per definition captured in one cross section anymore.

7.3. Validly of the reference strength

In this research is focused is on errors due to uncertain model input. Other uncertainties such as model and schematization uncertainties are neglected. In this section is discussed how the definition of the reference strength can be of influence to the error in the assessment.

The reference strength is constantly defined as the most unfavourable combination of independent parameters within a limited domain. Parameters are assumed to vary only in the length direction implying uniformity within cross sections. How the combination of independent parameters results in a strength value is captured by the model of Sellmeijer. The use of a model introduces model uncertainty: reality is possibly different than the model representation. The accuracy of the original Sellmeijer model (1989) has been studied past years, which resulted in improved versions of the original model. However, the only way to quantify model uncertainty is by comparing model calculations with actual performance of dikes (when loaded). In conclusion, there is no proven better alternative to calculate a critical head value then by Sellmeijer.

However, there is a major discussion point concerning the use of a Sellmeijer model in this research. The models are developed and calibrated in small scale laboratory tests and the relation between parameters and piping resistance is determined empirically for rather homogeneous sands. Consequently, the use of the model implies more or less uniform cross sections. The resistance of a dike section is then indeed given by the smallest value of d70. The uniform cross section assumption is permanently made in this research and seems reasonable when d70 is correlated in space (see Figure 48). Problem is that cross sections are not uniform when grainsizes are highly variable in space. So in scenarios with small correlation length, the uniformity assumption is not feasible (see Figure 45). If a cross section is not uniform, the resistance to piping is not automatically determined by the smallest value of d70.

Kanning and Calle (2013) described a theory in which the resistance is determined by the largest d70 in the erosion path. The strength of a section is then defined as the weakest link: the path in which erosion takes place easiest. The representative d70 to use in the Sellmeijer model is defined as the minimum of the largest d70 of every possible erosion path. This would mean that next to a length effect also a parallel effect is present. Which means that high variability in d70 has also a positive contribution to the strength. The concept of parallel effects due to heterogeneity is more and more recognized in the field of piping risks (personal communication with van Beek & Koelewijn, 2015), (personal communication with Blinde, 2015).

In a 1D analysis, correlation scale has clear influence to the bias in assessments. When a property is quickly fluctuating, the probability of an unfavourable parameter value increases. This results in general in smaller reference strengths. However, if those quick fluctuations also appear in the width of the dike, the probability increases that somewhere in the representative erosion path larger grains are present. Result is that reference strengths are possibly higher than expected. This is of influence to the bandwidth of errors and the influence of correlation length to this bandwidth (see the sensitivity of correlation length in chapter 5).

A 1D analysis suggests that it depends on the soil conditions whether the characteristics value analysis tends to overestimate or underestimate the strength. To capture parallel effects a relatively simple 2D simulation is set up in which the concept of erosion paths is incorporated in determination of the reference strength (analogue the weakest link analysis of Kanning, 2012). In a 2D simulation the variations both in the length direction as the width of the dike are explicitly considered. The used 2D simulation is in principle equal to the standard simulation (see chapter 2) except for the determination of the reference strength. For simplicity k is taken constant and the strength is given by the weakest erosion path of a 2D d70 field, see Figure 45 until Figure 47 for 2D fields. The 2D simulation results are given for correlation lengths of 180 meter and 1 meter are given in Figure 48 and Figure 49. Correlation lengths are equal in both length as width direction. In the same figures the results in case of uniform cross sections are given for comparison.

The results of Figure 48 and Figure 49 provide an indication of the expected errors if heterogeneity is contributing to the strength. It results in the hypothesis that the correlation in d70 is of inverse influence to the bias in the assessments then was expected based on the 1D analysis. More variability results in increased reference strength and therefore decreased probability of failure. Based on chapter 5 it was expected that lack of correlation would result in overestimation of strength by an assessment. This could be dangerous because the probability of failure is then higher than expected. However, a simple 2D analyses suggests that the opposite might be the case. Lack of correlation results in underestimation of strength which means that the probability of failure is much smaller than was expected based on the assessment.

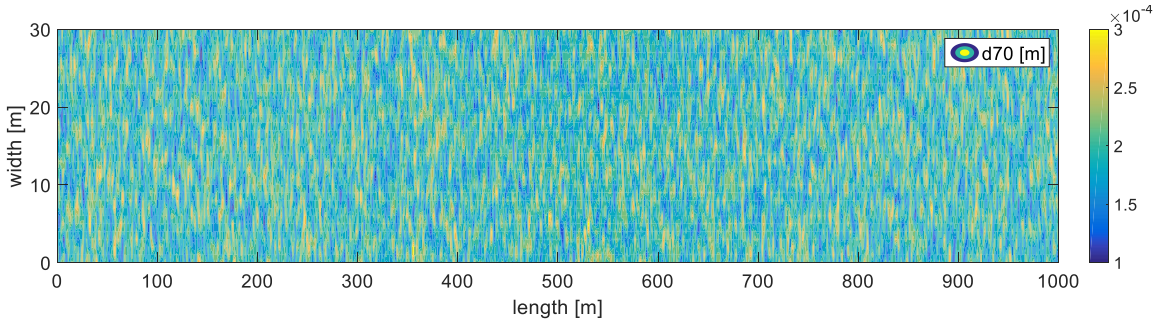


Figure 44 – Example d70 field (2D) with correlation length=1m (in both x and y)

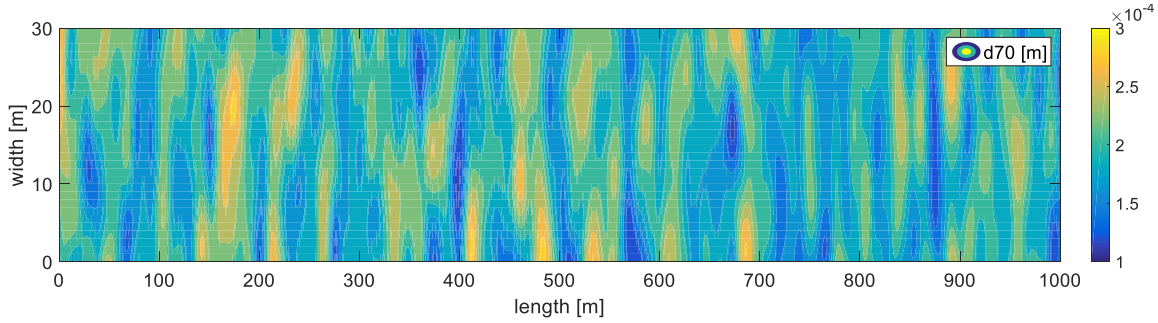


Figure 45 – Example d70 field (2D) with correlation length=10m (in both x and y)

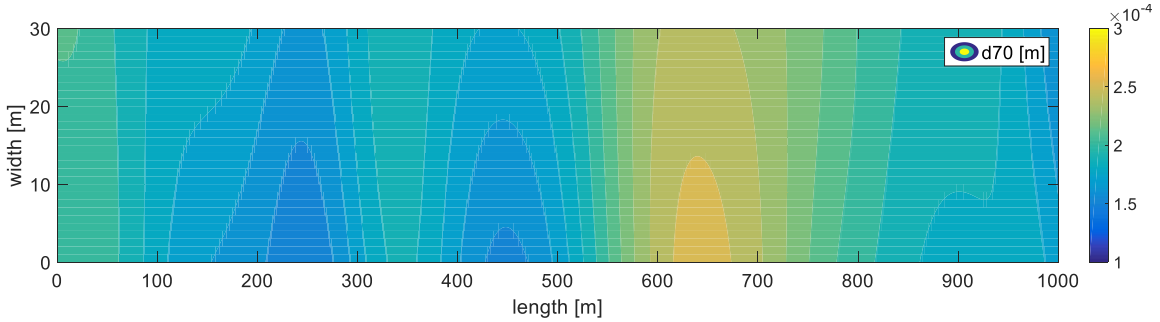


Figure 46 – Example d70 field (2D) with correlation length=100m (in both x and y)

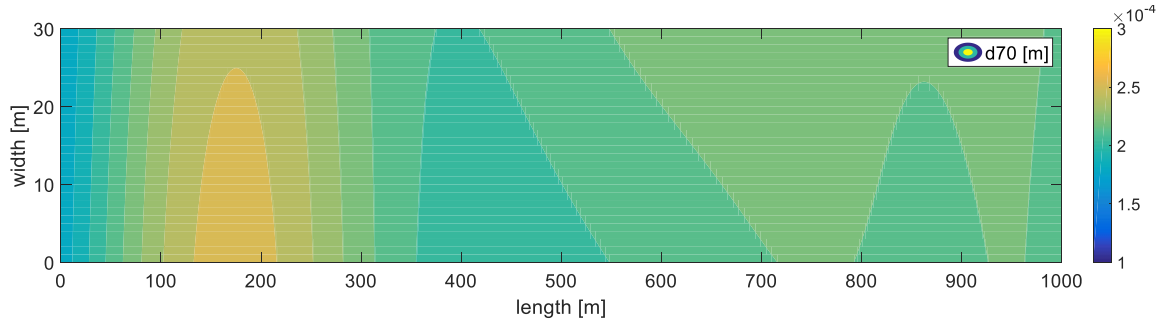


Figure 47 – Example d70 field (2D) with correlation length=200m (in both x and y)

Appendix 17 shows the influence of 2D modelling to the distribution of reference strength and assessed strength (measurement density is 1 per 1 meter) for several correlation lengths. Appendix 18 shows the influence of 2D modelling to the bandwidth of errors for several correlation lengths.

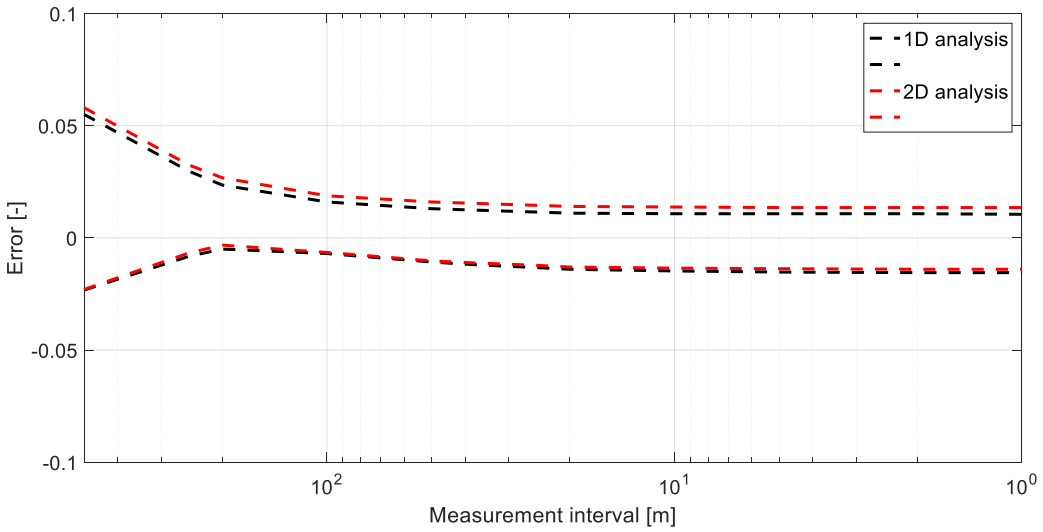


Figure 48 – Bandwidths of errors in case d70 is varying with a correlation length of 180m in both x and y direction. Results of both the standard 1D simulation and developed 2D analysis show the influence of parallel effects in correlated soils.

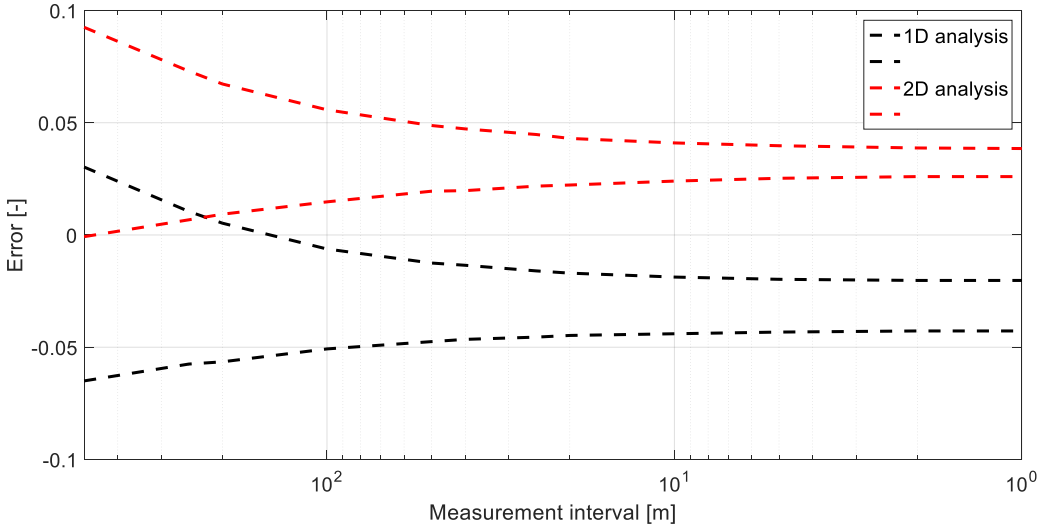


Figure 49 – Bandwidths of errors in case d70 is varying with a correlation length of 1m in both x and y. *Results of both the standard 1D simulation and developed 2D analysis show the influence of parallel effects in uncorrelated soils*

Because k is representing a layer average permeability, heterogeneity within a cross section will not directly influence the character of this parameter. However, calculating strengths with single k values does not completely justify that groundwater flow leading to erosion is actually a 3D process (Vandenboer, van Beek, & Bezuijen, 2014).

Finally, the independency of properties is discussed. In determination of the reference strength no cross correlation between d70 and k is modelled. It is questionable how realistic this is since small grains usually cause low permeability. But to what manner does permeability in the lower sand layers influence the force on grains in the top layer?

Aguilar-Lopez et al. (2016) discusses cross correlation. They show cross correlation is likely to be present and expected to decrease the probability of failure. In this research cross correlation would imply increase underestimation of actual strength by the assessment. Which means a smaller probability of failure is present than calculated. Note that the assessment assumes complete independence of properties. Especially when properties are uncorrelated the reference strengths will increase strongly as the probability that unfavourable parameter values correspond will decrease.

Remark: some cross correlation is implied by the fixed and coupled input means of d70 and k. This prevents to some extend that very large grains occur at locations with very low permeability.

7.4. Simplified assessment

In this research a simplified semi-probabilistic assessment is simulated. The limitations in representing an actual detailed assessment are discussed here.

Measurement intervals are assumed constant and equal for both k and d70 measurements. In practice the number of measurements per parameter can deviate (see appendix 20) and the positioning of measurements along the dike is not per definition with equal intervals between measurements (see appendix 9).

To schematize a representative cross section and estimate piping resistance, multiple sources of information can/should be used (ENW, 2012). In the simulation only one source of information is used, namely parameter values based on measurements. So actually only one degree of freedom, the density of measurements, is analysed. Furthermore, the assessment is not static as considered in this research

but dynamic as the understanding of the dike and subsoil increases while assessing it. In reality measurements are reviewed and not blindly entered to a formula. In addition to static formulas, also static approaches are analysed in this research. In chapter 4 only characteristic value analysis is used, also in case of high measurement density. In practice it is more likely that the approach of data use depends on the collected information and measured parameter values.

More design choices mean more uncertainties which might increase the probability of an error between the actual strength and assessed strength. On the other hand, multiple errors can omit each other. In addition, it is reasoned that if more information is available upfront to an assessment, measurements can be used more effectively and possibly more accurate assessments can be made.

Within the assessment procedure, also measuring is simplified. The measuring of k values is less straightforward as presented in this research. It is presented as point measurements that represents one k value. In reality the k values are often estimated from samples (d60 and d10) out of the entire sand layer. In that method some correlation between k and d70 is assumed. It is not straightforward to measure k values in situ. In this research the complexity of measuring accurate k values is neglected and all uncertainties are discounted in the measurement error. The measurement error is arbitrarily chosen.

In an actual assessment correlation between d70 and k is ruled out for sake of conservatism. The lower characteristic value of d70 has to be combined with the upper characteristic value of k. But when k is estimated from sand fractions (d10 and d60 measurements) some correlation is implied. The selection of k samples from all over the sand layer is not always guaranteed. Whether this implicit correlation is representative for reality or not is not known. This possible error is however not analysed in this research but could have influenced the error range. For example, if fine sands are on top of coarse sands, the use of d70 samples from the top layer to estimate k can result in significant error as a too optimistic permeability is considered.

7.5. Contribution to literature

In existing literature is mainly focussed on improved understanding and modelling of piping, amongst others (van Beek, 2015), (Robbins, Sharp, & Corcoran, 2015), (Vandenboer, van Beek, & Bezuijen, 2014), (Kanning, 2012), (Förster et al., 2012). In literature is also focussed on the improvement of reliability analysis of piping, for example (Aguilar-Lopez et al., 2016), (Schweckendiek, 2014). Furthermore, attention is paid to the influence of spatial correlation and determination of correlation scales (de Visser, Kanning, Koopmans, & Niemeijer, 2015), (Vrouwenvelder & Calle, 2003). So research has been done to describe mechanisms that influence the probability that piping occurs. The ability of models to describe actual processes as realistic as possible has been of primary focus.

As consequence of the legal safety assessment, a broad discussion has been ongoing about the actual safety risks with respect to piping. Many kilometres of dikes have been rejected based on the existing piping models and assessment procedures. There has been discussion about the correctness of assessment outcomes in representing actual piping risks (Vrijling, et al., 2009), supported by cases in which practical knowledge about the safety is in contradiction with the official assessment (POV piping; werkplaats Zwarte Water, 2015.

Additional research towards the piping problem was needed in order to reduce the gap between actual risks and estimated risks. It has amongst others resulted in full probabilistic reliability analyses, 3D erosion modelling, the weakest path theory and an updated/improved Sellmeijer model. Furthermore, there is consensus in literature that high calculated piping risks might be (partly) due to high uncertainties in model input and corresponding conservative assumptions. Although identified, this is

not jet supported by academic research. With this study a first step is made in the quantification of errors between estimated risks and actual risks. Specifically, to the errors that can be made if model input is inadequate. The results of this study show that the probability of failure of a dike due to piping might be underestimated in some cases. In those cases, the legal norm might not be met. However, in other cases the probability of failure is overestimated resulting in reinforcements while the dike section actually meets the norm.

This research also shows that the use of point measurements inevitable results in uncertain model input. Therefore, this research contributes to initiatives, such as coordinated by *Stichting FloodControl IJkdijk*, in which is argued for the need of more data and information and/or alternative data and information in piping risk analyses.

8. Conclusions and recommendations

In this chapter the answer to each research question is provided. Based on the drawn conclusions recommendations are given with respect to data use and further research.

8.1. Research questions

1. What is a representative geometry and soil characteristics of a piping sensitive dike section?

It is concluded that a dike is piping sensitive in a configuration with a cohesive and impermeable dike body on top of a permeable sand layer that is in connection with outside water. Pipes develop easier for small seepage lengths; therefore, a typical piping sensitive geometry has a rather small dike body. The seepage length is increased by the presence of a consistent covering clay layer moving entrance and exit points away from the dike body. Lack of a (consistent) clay layer increases the sensitivity to piping.

Uncertain soil characteristics and spatial variation can be described by probability density distributions and scales of fluctuation. In this research the statistical variables used in PC-Ring are followed to describe a reference scenario. The result of other representative distributions and correlation patterns are analysed in a scenario analysis to determine the influence of modelling choices.

It is not straightforward to describe generally representative soil characteristics because of different geological backgrounds and morphological units. In addition, little is known about soil variations on small scale because extensive data is missing. Recent extensive piping reliability analyses have calculated piping risks with correlation lengths that are estimated by experts and match with theory regarding large scale variations within rather homogeneous deposits. Some testing grounds with fine grids of d70 measurements, show no significant spatial correlation in grainsizes. Experts confirm that sand layers can be very heterogeneous, at least for some areas in the Netherlands. In conclusions, there is disagreement about the spatial correlation in soil characteristics that should be considered to analyse piping risks.

2. How does the density of measurements influence the accuracy of strength assessments using the approach described in the detailed assessment level?

Given the reference scenario, it is concluded that the assessment biases to conservative assessments and measurements decrease this bias to some extent. With a minimal measurement density of 3 per 1000 meter an expected error in the assessed strength of 1.4m underestimation of critical head is found. At a measurement density of 1 per 10 meter the expected error decreases to 0.6m underestimation of critical head. A decrease in the order of 50%. Next, it is concluded that the assessed strength value is not always close to the actual representative strength because a range of possible errors is found for every measurement density. Increase of measurement density decreases the range of errors to some extend due to reduction of statistical uncertainty. But at dense grids, density of 1 per 10 meter and denser, the range of errors is constant between accurate and 1.3m underestimation of critical head. For comparison, the average strength of the considered dike sections is 2.3m critical head.

The characteristics value approach is able to translate limited data into low (conservative) strength assessments that tend to underestimate the actual representative strength. The approach is not meant to translate high measurement densities into accurate assessments (error range of zero, close to zero). Although statistical uncertainty is reduced when using higher measurement densities, the assessment will not find representative parameters values consistently. Due to the use of static assumptions in the assessment method, the increased information about the (spatial) soil distributions stays unused.

3. *How sensitive is the accuracy of strength assessments to assumed (spatial) distributions of soil parameters?*

It is concluded that the bandwidth of errors depends on the average statistical characteristics of the soil parameters. When the average statistical characteristics are uncertain, the actual bandwidth of errors is much wider. When point measurements are the only source of information in the assessment, the expected bandwidth of errors is given by the combination of bandwidths from a whole range of plausible scenarios.

It is concluded that no general quantitative statements can be made about the bias in the assessments and range of possible errors. Due to the combination of statistical uncertainty, boundary effect, measurement error and spatial variability, relative errors up to almost 100% are possible. The scenario analysis shows that the range of possible errors depends on the actual representative strength, determined by the actual (unknown) spatial distribution of properties.

It is concluded that the influence of measurement density is independent of the average characteristics of soil parameters. Assessments generally become less conservative due to extra measurements and the error range to some extend smaller due to reduction of statistical uncertainty. By reducing statistical uncertainty, the range of possible relative errors can be decreased to about 50%.

Finally, it is concluded that due to the length effect assessments tend to be relatively unsafe at dike sections with small correlation lengths. This notice is mainly of importance with respect to k. With respect to d70 it is likely that parallel effects are present as counterforce to the length effect. Furthermore, it is concluded that the range of errors is generally smallest for the measurement density that is somewhat higher than typical correlation length. If correlation lengths of typical soils are known, this can give rough indications to what extend increase of measurement density can be valuable in uncertainty reduction.

4. *Is it possible to increase the accuracy of strength estimations by using information of point measurements alternatively?*

With location specific measurements it is possible to decrease errors due to spatial variability and statistical uncertainty. Random measurements errors can also be reduced by averaging multiple measurements per location. It is concluded that in theory it is possible to make accurate strength assessments with use of point measurements by making the density of measurements high in relation to the correlation length. It is concluded that to be certain that the error in the assessed strength is small, measurement densities are needed that are not feasible in practice. The use of point measurements does not justify the spatial variability of soil properties and related uncertainty about piping risks. Furthermore, the use of point measurements does not justify inaccuracies as result of measurement error.

To conclude about the potential of point measurements, different data use methods are evaluated. In both characteristic value analyses as measured critical value analyses, parameters are considered individually. This means that only uncertainty about the probability density distribution of parameters is reduced with increasing measurement densities. Measuring strength per location makes it possible to reduce errors due to spatial variability as well.

With high measurement densities it is possible to estimate statistical descriptions of parameters accurately. But at the same time, it is then possible to directly allocate strength values to many locations. But even with a large data set it is not straightforward to predict the strength accurately. Pipes develop on microscale and a point measurement is in principle only giving information about one specific location in the spatial domain. So the mechanism that is looked for is in principle as small as the size of the used observations. Spatial correlation of relevant properties determines how dense measurement grids should be to find reliable strength estimations.

The correlation length is a statistic that is not explicitly used in the detailed assessment. With high densities of point measurements, it is possible to estimate correlation lengths, but only if correlation length is much smaller than the measurement domain. It is concluded that quantification of correlation length is only useful to characterize larger deposits or geologic units. With knowledge of the correlation patterns upfront to the assessment, the range of possible errors can be narrowed.

8.2. Recommendations

First it is recommended to acknowledge that estimation of piping risks based on point measurements is unreliable because of a large range of possible errors in strength assessments. With characteristic value analysis (detailed assessment), errors in the same range as the actual strength or assessed strength are possible. Therefore, it is recommended not to rely solely on information from point measurements.

It is shown that the error in a strength assessment is location specific, i.e. depends on the local soil variations. Therefore, it is recommended to use customized above standardized assessments. It is recommended to use all available information (dynamically) to divide trajectories into sections and to develop a measurement plan. Information about variation scales, correlation patterns and possible anomalies should always be considered.

Assessing piping risks is mainly about identification of weak locations in the length direction of the dike. It should be noticed that the nature of point measurements does not justify existing spatial variations and the possibility of anomalies. It is recommended to use alternative techniques and additional information as well to locate possible weak spots and use traditional point measurement to quantify soil properties at those locations.

With point measurements accurate assessments can be made if the density is high enough and measurement error is ruled out. The number of measurements that would be needed is not considered plausible in practice. Instead it is recommended to consider surface covering or volume covering measurements because of its potential to provide insight in the variation of a certain property at high density. Besides that, possible measurement error is relative, which increases the accuracy in identification of actual weak spots.

It is recommended to do further research into the possibilities of (using) surface covering or volume covering measurements. With respect to the k parameter it might be valuable instead of permeability measurements (monitoring well, pumping test) at an interval that is smaller than the width of possible gullies (Schweckendiek, 2014). With respect to the d70 parameter it might be valuable to characterize heterogeneity (de Visser et al., 2015) and identify/rule out anomalies with fine sand fractions that cross the dike width in full length.

Experiments showed that dikes fail later than expected based on the current theoretical knowledge. In practice this is visible at water boards that in theory have to reject many kilometres of dike while practice shows many of these dikes are fine (van Putten, 2013). Because discrepancy between assessed strength and actual representative strength is likely (showed in this research) it is recommended to do additional research in case of doubt. Monitoring and use of sensors is a possibility/shows to be promising (van Putten, 2013), (Koelewijn, Pals, Sas, & Zomer, 2010), (Sluis, Sirks, Koelewijn, & Veenstra, 2016). *For extensive information about monitoring is referred to Kennisplatform Dijkmonitoring (www.dijkmonitoring.nl).*

It has been reasoned in literature that behaviour of soil is determined by mechanisms on microscale. Therefore, it can be questioned whether it is possible to measure the (all) relevant/representative parameters. Because soil bodies are so heterogeneous and complex it is hardly possible to estimate the

behaviour with models and model input completely accurately. To increase insight in behaviour it could be useful to consider the monitoring of output parameters as well. In case of piping for example measuring of increased seepage during high water. In case of macro stability for example measuring movement of the dike body during high water.

This research has been a first step in the quantification of possible errors when assessing the risk for piping with limited data and information. From an academic perspective it would be interesting to extend this research to a more sophisticated 2D/finite elements analysis. It is recommended to focus on a more realistic determination of the reference strength, in which parallel effects (d70), 3D groundwater flow (k) and covering clay layers (d and L) are incorporated as well. It would be of interest to compare this reference strength with assessed strengths from both a random finite element analysis as with the characteristic value analysis and Sellmeijer model. Both assessment methods subjected to the limitations of point measurements model input.

In this research is shown that soil conditions have impact on the safety of dikes with respect to piping. Mainly about the correlation scales of soil properties and presence of anomalies much uncertainties/unknowns exist. It is recommended to increase the insight in the spatial variations and anomalies with respect to geographic locations in the Netherlands.

Bibliography

Abdul Rahman Shibli, S. (2003, February). Geostatistics FAQ. Retrieved from https://wiki.52north.org/pub/AI_GEOSTATS/AI_GEOSTATSFAQ/FAQ_Geostatistics_01.pdf

Aguilar-Lopez, J., Warmink, J. J., Schielen, R. M., & Hulscher, S. J. (2016). Soil stochastic parameter correlation impact in the piping erosion failure estimation of riverine flood defences. *Structural Safety, Volume 60*, 117-129.

Bot, A. P. (2011). *Grondwaterzakboekje.* Atelier Rijksbouwmeester.

Calle, E. (2007). Statistiek bij regionale proevenverzamelingen:Het ruimtelijk model. *Geotechniek 11 (3)*, 40-44.

Calle, E., van der Meer, M. T., & Niemeijer, J. (1999). *Technische rapport Zandmeevoerende wellen.* Technische Adviescommissie voor de Waterkeringen. Retrieved from http://repository.tudelft.nl/assets/uuid:d5354144-9b49-48b2-a6cc-0e13d0e38691/TR_15_Technisch_Rapport_Zandmeevoerende_wellen.pdf.

de Smidt, J. T., Verruijt, A., Barends, F. B., Dekker, J., Epema, W. G., de Haan, W., . . . Vrijling, J. K. (1994). *Handreiking constructief ontwerpen.* Bennekom: Drukkerij Modern. Retrieved from http://repository.tudelft.nl/assets/uuid:93574db5-ea9b-49a1-9e1c-ecb2b20d36e9/L9_-HandreikingConstructiefontwerpen.pdf.

de Visser, M. M., Kanning, W., Koopmans, R., & Niemeijer, J. (2015). Determination of Spatial Variability in d70 Grain Size Values Using High density Site Measurements. *Geotechnical Safety and Risk V* (pp. 213-218). Amsterdam: IOS Press.

Drenth, P. (2015). *Effect van meetkwantiteit op toetsingsresultaat (BSc thesis).* Deventer: Universiteit Twente/BZ Innovatiemanagement.

Engerink, F., Calle, E., Kruse, G., Selmeijer, J., & Blinde, J. (2007). *Module MPinping-VNK voor piping-analyse in PC-Ring.* Delft: GeoDelft.

Expertisenetwerk waterveiligheid. (2012). *Technisch Rapport Grondmechanisch Schematiseren bij dijken.* Rijkswaterstaat Water, Verkeer en Leefomgeving. Retrieved from http://repository.tudelft.nl/islandora/object/uuid:bd8f4337-6519-474d-abb6-236c8574f19f?collection=research

Fenton, G. A., & Vanmarcke, E. H. (1991). Proceedings of the ASCE Geotechnical Engineering Congress. *Spatial variation in liquefaction risk asessment*, (pp. 594-607). Boulder, Colorado.

Förster, U., van den Ham, G., Calle, E., & Kruse, G. (2012). *Onderzoeksrapport zandmeevoerende wellen.* Deltares.

Förster, U., van den Ham, G., Calle, E., & Kruse, G. (2012). *Onderzoeksrapport Zandmeevoerende Wellen.* Deltares. Retrieved from http://library.wur.nl/WebQuery/hydrotheek/2040046.

Geodelft. (1991). *Verificatie Piping Model - Proeven in de Deltagoot - Evaluatierapport. Geodelft report CO-317710/9.* Retrieved from http://repository.tudelft.nl/islandora/object/uuid:83be44f0-1340-4f3b-bccd-c9e2623b9188?collection=research

Inspectie Leefomgeving en Transport. (2013). *Verlengde derde toets primaire waterkeringen.* Utrecht: Ministerie van Infrastructuur en Milieu. Retrieved from

https://www.rijksoverheid.nl/binaries/rijksoverheid/documenten/rapporten/2014/02/28/landelijke-rapportage-2012-2013-verlengde-derde-toetsing-primaire-waterkeringen/landelijke-rapportage-2012-2013-verlengde-derde-toetsing-primaire-waterkeringen.pdf.

Jongejan, R., Stefess, H., Roode, N., ter Horst, W., & Maaskant, B. (2011). 5th International Conference on Flood Managament. *The VNK2 Project: a detailed, large scail quantitative flood risk analysis for the Netherlands.* Tokyo-Japan.

Kanning, W. (2012). *PhD thesis: The weakest link - Spatial Variability in the Piping Failure Mechanism of Dikes.* Delft: Technische Universiteit Delft.

Kanning, W. (2012). *The weakest link - Spatial Variability in the Piping Failure Mechanism of Dikes.* Delft: Technische Universiteit Delft.

Kanning, W. (2012). *The weakest link - Spatial Variability in the Piping Failure Mechanism of Dikes (PhD Thesis).* Delft: Technische Universiteit Delft. Retrieved from http://repository.tudelft.nl/view/ir/uuid:5fb7b121-dc00-48aa-bda2-b163f10513bf/

Kanning, W., & Calle, E. (2013). Derivation of a representative piping resistance parameter based on random field modelling of erosion paths. *Georisk, Vol 7, Iss 2*, 99-109.

Koelewijn, A., Pals, N., Sas, M., & Zomer, M. (2010). *IJkdijk pipingexperiment; validatie van sensor- en meettechnologie voor detectie van optreden van piping in waterkeringen.* Stichting IJkdijk.

Lark, L. M. (2000). Estimating variograms of soil properties by the method-of-moments and maximum likelihood. *European Journal of Soil Science 51*, 717-728.

Ministerie van Verkeer en Waterstaat. (2007). *Voorschrift toetsen op veiligheid primaire waterkeringen.* Retrieved from https://www.rijksoverheid.nl/documenten/rapporten/2007/08/01/voorschrift-toetsen-op-veiligheid-primaire-waterkeringen-voor-de-derde-toetsronde-2006-2011

Ministerie van Verkeer en Waterstaat. (2007). *Voorschrift toetsen op veiligheid primaire waterkeringen.*

Nicolai, R., Vrouwenvelder, T., Wojciechowska, K., & Steenbergen, H. (2013, Oktober). Omgaan met onzekerheden in het waterveiligheidsbeleid. *STator*, pp. 20-25.

Oldhoff, R. (2013). *Meetkwantiteit versus toetsingskwaliteit (BSc thesis).* Deventer: Universiteit Twente/BZ Innovatiemanagement.

Rijkswaterstaat. (2015). *Handreiking ontwerpen met overstromingskansen - Veiligheidsfactoren en belastingen bij nieuwe overstromingskansnormen.* Rijkswaterstaat Water, Verkeer en Leefomgeving. Retrieved from http://repository.tudelft.nl/assets/uuid:92dda204-f729-47b1-be4c-66dc61af64c9/Handreiking-Ontwerpen-met-overstromingskans-v2.5-20150610.pdf.

Robbins, B. A., Sharp, M. K., & Corcoran, M. K. (2015). Geotechnical Safety and Risk V. *Laboratory Tests for Backwards Piping Erosion* (pp. 849-854). http://repository.tudelft.nl/islandora/object/uuid%3A927db8d5-ea11-4cf0-a7e4-0ffd26195bcd?collection=research: IOS Press.

Royal Haskoning. (2010). *Veiligheidstoetsing primaire waterkering.* waterschap Hunze en Aa's.

Royal Haskoning. (2010). *Veiligheidstoetsing Waddenzeedijk - primaire waterkeringen dijkringgebied 6.* waterschap Noorderzijlvest.

Schweckendiek, T. (2014). *On Reducing Piping Uncertainties: A Bayesian Decision Approach (PhD thesis).* Delft: TU Delft/Deltares. Retrieved from http://repository.tudelft.nl/view/ir/uuid:f9be2f7e-7009-4c73-afe5-8b4bb16e956f

Schweckendiek, T. (2014). Using head monitoring for reliability updating of levees. *Geotechnical Safety and Risk IV - Proceedings of the 4th International Symposium on Geotechnical Safety and Risk, ISGSR 2013*, (pp. 365-370). Hong Kong.

Schweckendiek, T., & Vrouwenvelder, A. (2013). Reliability updating and decision analysis for head monitoring of levees. *Georisk 7 (2)*, 110-121.

Sellmeijer, H., de la Cruz, J., van Beek, V., & Knoeff, H. (2010). Fine-tuning of the backward erosion piping model through small-scale, medium-scale and IJkdijk experiments. *European Journal of Environmental and Civil Engineering 15 (8)*, 1139-1154.

Sellmeijer, J., López de la Cruz, J., van Beek, V., & Knoeff, H. (2011). Fine-tuning of the backward erosion piping model through smaal-scale, medium-scale and IJkdijk experiments. *European Journal of Environmental and Civil Engineering 15/8*, 1139-1154.

Sluis, J., Sirks, E., Koelewijn, A., & Veenstra, J. (2016). Resultaten monitoringsproef op Ameland. *Geotechniek Januari*, 8-12.

Technische Adviescommissie voor de Waterkeringen. (1999). *Technische rapport Zandmeevoerende Wellen.*

Technische Adviescommissie voor de Waterkeringen. (2000). *van Overschrijdingskans naar Overstromingskans.* Retrieved from http://repository.tudelft.nl/assets/uuid:b30132ee-95af-4ed8-96be-9580a9bdb14a/P_00_0420Van20Overschrijdingskans20naar20Overstromingskans20Hoofdrapport.pdf.

van Beek, V. M. (2015). *Backward erosion piping: Initiation and progression (Phd thesis).* http://repository.tudelft.nl/islandora/object/uuid%3A4b3ff166-b487-4f55-a710-2a2e00307311?collection=research: TU Delft.

van Noortwijk, J. M., Vrouwenvelder, A. C., Calle, E., & Slijkhuis, K. (n.d.). *Probability of dike failure due to uplifting and piping.* Retrieved July 2015, from EWI: http://www.ewi.tudelft.nl/fileadmin/Faculteit/EWI/Over_de_faculteit/Afdelingen/Applied_Mathematics/Risico_en_Beslissings_Analyse/Papers/noopod1.pdf

van Putten, R. (2013). Dijken optimaliseren met sensoren. *Geotechniek special*, 56-58.

van Zwieteren, T. J. (2013). *Spatial variation in dike safety management for uplift (MSc thesis).* Delft: TU Delft/Fugro. Retrieved from http://repository.tudelft.nl/islandora/object/uuid%3A06906a57-07d0-4a75-8d3b-1bb8882c91c6?collection=education

Vandenboer, K., van Beek, V., & Bezuijen, A. (2014). 3D finite element method (FEM) simulation of groundwater flow during backward erosion piping. *Frontiers of Structural and Civil Engineering, Vol 8, Iss 2*, 160-166.

Vergouwe, R. (2014). *De veiligheid van Nederland in kaart.* Rijkswaterstaat Projectbureau VNK. Retrieved from https://www.helpdeskwater.nl/publish/pages/33875/eindrapport_vnk.pdf

Vienken, T., & Dietrich, P. (2011). Field evaluation of methods for determining hydraulic conductivity from grain size data. *Journal of Hydrology 400*, 58-71.

Vrijling, J., Kok, M., Calle, E., Epema, W., van der Meer, M., van den Berg, P., & Schweckendiek, T. (2009). *Piping - Realiteit of Rekenfout.* Expertisenetwerk Waterveiligheid. Retrieved from http://repository.tudelft.nl/islandora/object/uuid%3Af5b79879-f4d2-4fce-9b11-38547db4509f?collection=research

Vrouwenvelder, A. C., & Steenbergen, H. M. (2003B). *Theoriehandleiding PC-RING deel B: Statistische modellen. .* TNO report 2003-CI-R0021.

Vrouwenvelder, A., Steenbergen, H., & Slijkhuis, K. (1999). *Theoriegandleiding PC-Ring - Deel B: Statistische modellen.* Delft: TNO.

Vrouwenvelder, T. (2006). Spatial effects in reliability analysis of flood protection systems. Lake Louise, Canada: International Forum on Engineering Decision Making.

Vrouwenvelder, T., & Calle, E. (2003). Measuring spatial correlation of soil properties. *HERON*, Vol 48, No. 4.

Appendices

1. The piping failure mechanism

The occurrence of piping is related to several factors (Ministerie van Verkeer en Waterstaat, 2007):

- The head difference over the dike. The head difference causes pressure difference resulting in high water pressures in the sand layer below the dike (if this sand layer is in connection with the outside water).
- The thickness and permeability of the water transporting layer (sand layer) under the dike body. These parameters determine the magnitude of flow through the sand layer as a result of head difference. Higher flows increase the risk of sand particles to move.
- The seepage length: distance between entry point and exit point. A longer seepage length means a lower pressure gradient and increased resistance preventing sand particles to move.
- The thickness and volumetric weight of the covering (clay) layer above the permeable sand layer. Such a layer prevents water to flow towards the surface and therefore increases the seepage length. If the water pressure in the sand layer is high enough this layer can burst (uplift) causing water to flow from under the dike to the surface. If the flow is strong it can start to transport sand particles from below the dike towards the surface.
- Grainsize distribution. Especially of importance at the exit point. Finer materials have lower resistance against internal erosion as the flow is better able to move these particles.
- The presence and dimensions of a ditch. A ditch is possibly a weak spot in the covering clay layer as the layer is thinner there (uplift). The ditch functions often as the exit point of the seepage flow.

Usually only the cross section is used to visualize piping (see example in Figure 50). In reality after the occurrence of uplift a water boil is created causing a three dimensional flow pattern towards the boil. If the flow is strong enough erosion starts. The channel develops from inner side to outer side but also in the length direction of the dike depending on the grain distributions below the blanket layer. The channel can also develop in depth, but collapses cause it to continue just below the blanket layer which is usually cohesive. The sand layer thickness D below the blanket layer is not important for the resistance (because only in the top layer channels can develop), but is important for the development of flow patterns and flow velocity (Kanning, 2012).

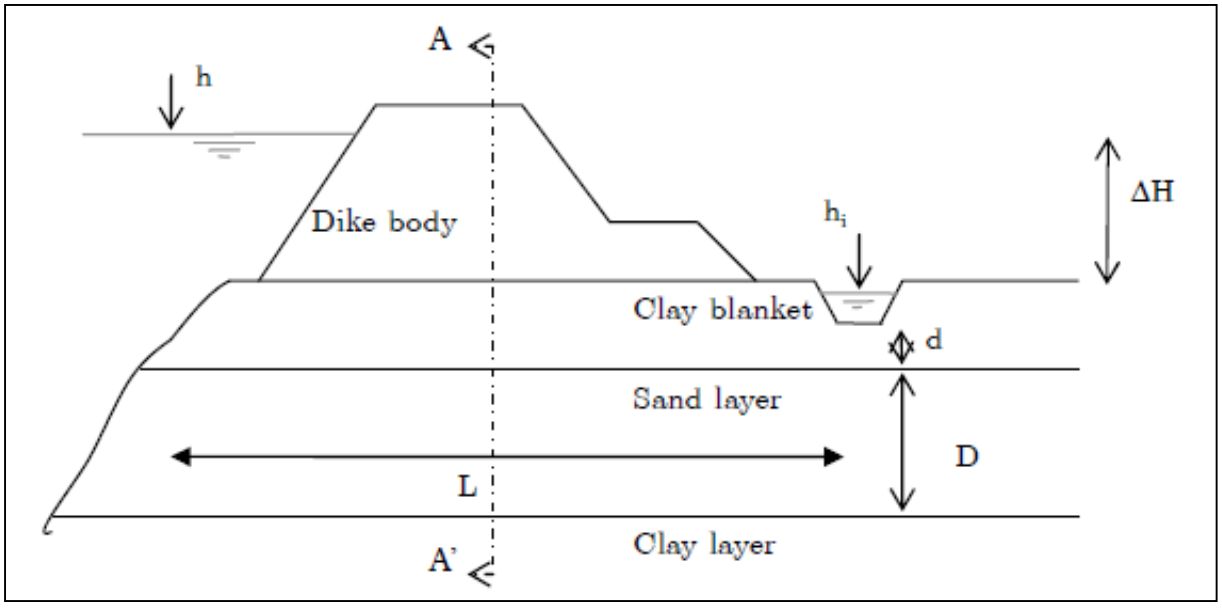


Figure 50 – typical cross-section of a piping sensitive dike (Kanning, 2012).

2. The legal assessment of piping risks

Three levels of assessment are possible to determine the safety of a dike against the piping failure mechanism:

- A simple global check in which mainly the geometry of the dike is considered.
- A detailed check in which the safety is checked according to the model of Sellmeijer (Sellmeijer, 1989).
- An advanced assessment in which the local (sub) soil is modelled in detail or in which a probabilistic analysis is carried out. No explicit procedures exist when performing an advanced assessment, but it has to sufficiently convincing in proving the safety of the assessed dike (Ministerie van Verkeer en Waterstaat, 2007).

If from the global check it appears the dike might be unsafe the detailed check is carried out. If this check is does not lead to the conclusion the dike is safe, it is possible to do an advanced assessment. For every step, from global to advanced, the number of input parameters and needed amount of data increases (Ministerie van Verkeer en Waterstaat, 2007).

In the safety assessment the detailed check is most used. For this purpose, a semi-probabilistic approach is used, in which stochastic parameters related to strength are translated to characteristic calculation values (Förster, van den Ham, Calle, & Kruse, Onderzoeksrapport zandmeevoerende wellen, 2012). The procedure is as follows:

- First a dike ring is split into several small dike section based on (sub) soil characteristics
- From every stretch a representative cross section is made. This cross section holds the information to calculate the resistance of that dike section: for piping the critical head difference or critical slope is calculated as strength indicator.
- To make a representative cross section, (local) information about the dike geometry and subsoil is needed. The subsoil information becomes available by doing (additional) measurements. Because of spatial variability of soil characteristics within the assessed dike stretch, the measured strength parameters differ from place to place.
- Strength parameters (for example grain size, seepage length, sand layer thickness and clay layer thickness) are therefore translated into random variables (based on at least 3 sample points). Most parameters are assumed to have a normal or lognormal distribution. Some parameters can hardly be measured; in that case nominal values are assumed (safe conservative values).
- From the random distributions, characteristic values are calculated based on statistical analysis. This means a value is chosen such that with 95% certainty the actual values in the field are more favourable with respect to piping.
- With each parameter assigned a characteristic (assumed safe) value, the critical head is calculated using the calculation rule of Sellmeijer.
- The critical head is divided by partial safety factors to account for (model) uncertainties and the translation of dike cross section safety to dike ring safety.
- The ‘corrected’ critical head is compared to the head difference (load) as determined from the norm. The norm provides a water level that is exceeded with a certain frequency.
- If the strength is less than the load, the dike section is determined ‘unsafe’. Each section needs to be determined ‘safe’ in order to conclude a dike ring to be ‘safe’.

3. Piping model of Sellmeijer

To determine the possibility of piping several models can be used, varying from simple calculation rules to complex 3D finite element methods. A very simple and often used model to do a first check is the empirical calculation rule of Bligh (based on a critical average gradient). A more detailed physical model is developed by Sellmeijer. This calculation rule of Sellmeijer is most often used in detailed piping assessments. To have a safe situation the critical head calculated with the models should be higher than the effective head difference over a dike (Förster, van den Ham, Calle, & Kruse, Onderzoeksrapport zandmeevoerende wellen, 2012).

The model of Sellmeijer is the more 'sophisticated model for the determination of the critical water level. The Sellmeijer model is prescribed in the detailed assessment of dike safety towards piping. The model is an equilibrium model that checks if a critical situation is developed. The idea behind the model, validated with observations, is that the piping channel can reach an equilibrium if it not passes half the leakage length. Sellmeijer considers the occurrence of sand boils still as a possible safe situation as the piping channel can reach an equilibrium and stops developing. The safety of a structure when observing a sand boil can therefore not exactly be indicated. The critical head according to Sellmeijer (1989) is calculated by:

$$\Delta H_c = \alpha c L \left(\frac{\gamma_p}{\gamma_w} - 1\right)(0.68 - 0.1\ln(c))tan\theta_R > 0$$

$$\alpha = \left(\frac{D}{L}\right)^{\left(\frac{0.28}{\left(\frac{D}{L}\right)^{2.8}-1}\right)}$$

$$c = \eta * d_{70}\left(\frac{1}{\kappa L}\right)^{\frac{1}{3}}$$

$$\kappa = \frac{\upsilon}{g}k$$

In which:
L is leakage/seepage length
$D$is sand layer thickness
α included limited thickness of the sand layer
c incorporates the erosion resistance of the sand layer
γ_p is wet soil weight
γ_p is water weight
θ_R is the rolling friction angle
η is the White's constant
d_{70} is the 0.70 grain size fractile of the sand
κ is the intrinsic permeability
υ is the kinematic viscosity
g is the gravitational acceleration
k is the permeability

Adjusted model

In the context of the research program *Sterkte en Belastingen van waterkeringen,* a new comprehensive research towards piping and heave is carried out in order to identify the uncertainties within the current assessment procedures. The results are described in the report *Zandmeevoerende wellen* (Förster et al., 2012) and have suggested a renewed and improved calculation rule of Sellmeijer.

They also state that the rule of Bligh is no longer applicable as it can overestimate the safety. The adapted Sellmeijer model calculates the critical head according to (Sellmeijer et al., 2011):

$$\Delta H_c = L * F_{resistance} * F_{scale} * F_{geometry}$$

$$F_1 = F_{resistance} = \frac{\gamma_p}{\gamma_w}\{\eta \tan(\theta)\}$$

$$F_2 = F_{scale} = \frac{d_{70m}}{\sqrt[3]{\kappa L}}\left(\frac{d_{70}}{d_{70m}}\right)^{0.6}$$

$$F_3 = F_{geometry} \underset{=}{MSeep} F(G) \underset{=}{standard\ dike} 0.91 * \left(\frac{D}{L}\right)^{\frac{0.28}{\left(\frac{D}{L}\right)^{2.8}-1}+0.04}$$

In which:

d_{70m} is the average d_{70} of the in small scale trial applied types of sand for which this formula is fitted: $d_{70m} = 2.08 * 10^{-4}$ [m].

This adjusted rule is supposed to be used for safety assessments in the Netherlands from now on.

Parameter (estimations)

In relation to the uplift and piping problem the following categories of parameters are present in the piping model:

- Hydraulic boundary conditions: water levels with a certain exceedance frequency that have to be retained safely.
- Construction dimensions (geometry).
  Often construction dimensions can be distracted from (old) design specifications.
- Structure of dike body and subsoil and geo-hydrologic system (are water transporting layers in contact with outside water and the head ('stijghoogtes') in these layers).
  Information about this often follows from a first global soil investigation.
- Material characteristics.
- Geo-hydrologic characteristics.

The models of Sellmeijer hold more information than the average gradient models and thus require more data. When using the calculation model of Sellmeijer some explicit estimates are required (Technische Adviescommissie voor de Waterkeringen, 1999):

- Permeability of the sand layer
- 70-percentilevalue of the grain size distribution (d70)
- thickness of sand layer and the development of it below and beside the dike
- Specific parameter indications: 'the White's constant', 'the rolling friction angle'. These are very hard to determine with measurements; therefore, nominal prescribed values are used (partly based on laboratory tests).

Grain size distribution

The grain size distribution influences the stability of grains in the potential pipe. Therefore, it is important to know specifically the grain size distributions just below the impermeable layer close to

the exit point. Samples can best be taken close to the dike toe (inside of the dike). In de piping formula a conservative estimation (low representative value) of d70 should be used.

Permeability

The permeability of the sand layer is a very sensitive parameter in the piping mechanism. But acquiring reliable estimation of the permeability is a problem, in which the effect of heterogeneity in the sand layer plays a difficult role. In the piping formula a conservative estimation (high representative value of the layer's average) of the permeability should be used. In the assessment practice usually estimations based on TNO-ground water maps or estimations based on grain size distributions are used. The ground water maps are probably conservative (deeper sand layers, less fine grains, more permeable) while the grain size relation is probably optimistic (from sand samples in the top of the layer, which are usually fine and less permeable) (Förster, van den Ham, Calle, & Kruse, Onderzoeksrapport zandmeevoerende wellen, 2012).

The permeability can be estimated based on grain size distribution samples. In TAW (1994) a procedure is described to calculate the permeability based on samples of the sand layer. In the Netherlands almost all water transporting sand layers consist out of sand with a D60/D10 ratio smaller than 10. The Dutch sand is relatively fine-grained and uniform, varying between 150 and 350 micro meter. The uniformity (D60/D10) lies between de 1.5 and 3 (Förster, van den Ham, Calle, & Kruse, Onderzoeksrapport zandmeevoerende wellen, 2012).

Furthermore, it is possible to obtain estimates of the permeability with in situ tests as the 'pumping test', the 'falling head test' and the use of monitoring wells. In the piping mechanism in situ test are preferred over estimates from grain size distribution samples. This is because the local permeability is not so important compared to the 'bulk' permeability of the entire sand layer (TAW, 1994).

Effective head

The calculated critical head is compared with the effective head in order to determine the possibility of piping to occur. This effective head is the difference between the outside water level and the inside (ground) water level. If a ditch is present the water level in this ditch is used, otherwise the surface level at the inside of the dike is considered (Ministerie van Verkeer en Waterstaat, 2007).

In case of a covering impermeable layer (blanket layer) at the inside of the dike the piping mechanism is affected. This is modelled as a reduction of the effective head difference: $\Delta H_e = \Delta H - 0.3 * d$. In which d is the thickness of the impermeable layer. The thickness of the blanket layer is difficult to estimate due to variability and local excursions as ditches and gravel pits (Kanning W. , The weakest link - Spatial Variability in the Piping Failure Mechanism of Dikes, 2012).

4. Characteristic value analysis

Sampling

Samples are taken from noised correlated realisations to simulate measurements. Characteristics value calculations are made for the following measurement configurations:

Configuration:	1	2	3	4	5	6	7	8	9	10	11	12
No. measurements	3	5	6	11	21	26	41	51	101	201	501	1001
Measurement density [1/m]	500	250	200	100	50	40	25	20	10	5	2	1

The example realisations in Figure 51 show measurement configurations 1, 4 and 9 for example realisations of d70 and k.

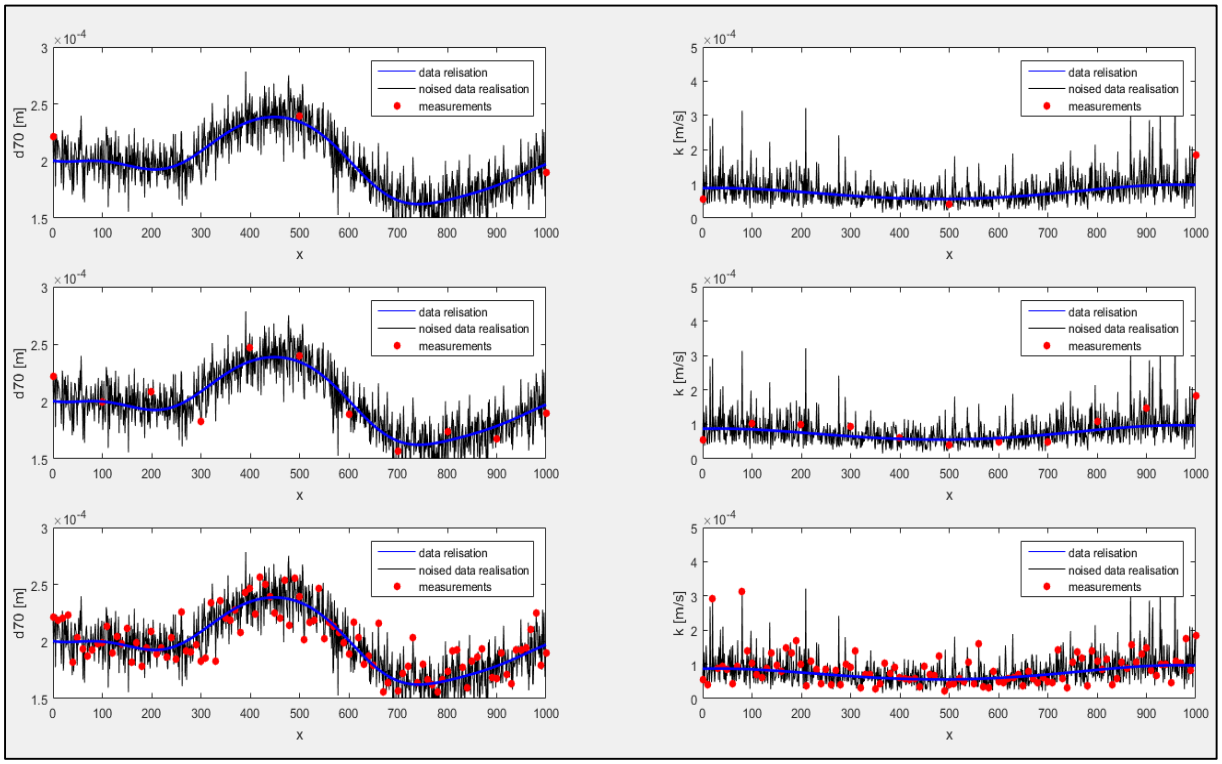


Figure 51 – Simulation of d70 and k measurements according to 3 different measurement configurations (No. of measurements). Simulation based on standard input configuration with measurement error

Formulas for characteristic value calculations

In this section the used formulas for characteristic value calculations are given. Parameter related samples are represented by the symbol p.

Normal distributed data

Because the variance of *d70* data is relatively small, it is assumed that d70 is normally distributed. In case of relative small variations or if a normal distribution of the property is assumed. For d70 the lower boundary of the data set is representative for piping. The related characteristic value is calculated according to:

$$p_{char,d70} = mean(p) - t_{95} * std(p) \qquad \textit{Equation 1}$$

Lognormal distributed data

Permeability data sets show relative high variances. TAW (1999) and Deltares (2002) advise to consider high variable data to be lognormal distributed. In this study the k values are in fact lognormal distributed and show relatively high variances. For the k parameter the upper boundary of the data set is representative for piping. The characteristic upper value is then calculated according:

$$p_{char} = \exp[mean(\ln p) + t_{95} * std(\ln p)] \qquad \textit{Equation 2}$$

Instead of the mean and variance of the sample values, the mean and variance of the natural logarithmic of the sample values are used

Parameters

The characteristic value of a sample is calculated with the mean and standard deviation (std) of the sample set. The mean is a measure of the expected value of the parameter. The standard deviation is a measure for the variation of the parameter values around the mean. The third parameter is the student-t factor. Which is determined by the number of used measurements *N* and the required certainty interval. In case of piping the 95% interval is always used (Technische Adviescommissie voor de Waterkeringen, 1999). The student-t factors are given in the following table:

N-1	***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***10***
$\boldsymbol{t_{N-1}^{0.95}}$	6.314	2.920	2.353	2.132	2.015	1.943	1.895	1.860	1.833	1.812
N-1	*11*	*12*	*13*	*14*	*15*	*16*	*17*	*18*	*19*	*20*
$\boldsymbol{t_{N-1}^{0.95}}$	1.796	1.782	1.771	1.761	1.753	1.746	1.740	1.734	1.729	1.725
N-1	*21*	*22*	*23*	*24*	*25*	*26*	*27*	*28*	*29*	∞
$\boldsymbol{t_{N-1}^{0.95}}$	1.721	1.717	1.714	1.711	1.708	1.706	1.703	1.701	1.699	1.645

The student-t factor is a statistical scalar to find a certain value from a probability density distribution. In case of piping, the value is searched for which the probability that the actual value is more favourable is at least 95%.

Example realisation with characteristic value calculation

In this section the characteristic values and resulting strength assessment is visualized with an example run. Figure 52 shows an example of a d70 realisation as the result of the standard input configuration. Next to the data realisation the minimum present d70 value is highlighted. Furthermore, the characteristic value as calculated based on three measurements is visualized. This is done for situations with measurement error (noised) and without measurement error (error free) This gives, for this typical example, insight in the influence of measurement error on assessments/calculations of representative parameters: in this case the characteristic d70 value.

Figure 53 shows the same but now for an example realisation of k. In this case the maximum value is searched for so the characteristic value is the theoretical upper 95% boundary based on the (three) measurements. For this figure the logarithmic scale is used. Because of the marginal sensitivity of the strength to D variations, no figure is related to the D parameter. However, the characteristic value calculation is in the same analogy as for d70.

Figure 54 gives the resulting strength realisation. The reference strength and assessed strength are visualized. The assessed strength (error free) is based on characteristic value inputs that are calculated using measurements without measurement error. The assessed strength (noised) is based on characteristic value inputs that are calculated with measurements subjected to measurement error.

Because measurements of d70, k and D are taken at the same locations, the actual strength at the three locations is measured. Because three locations are not representing the whole dike section, characteristic value calculations are normally used. Because for each parameter in principle a 'conservative' calculations value is used (95% upper/lower boundaries), the combination of conservative values leads to a theoretical 'weak spot' where all properties are possibly unfavourable. Figure 54 shows that it is possible that the predicted 'weak spot' is not present in the considered section. It is possible that weak locations in terms of d70 are for example counteracted by strong locations in terms of k. However, in this specific case the underestimation of strength is also caused by too conservative characteristic values of both k and d70 individually. As can be seen in Figure 52 and Figure 53. The numbers and calculated values related to this example runs are presented in *Table 11* and *Table 12*. From the figures and tables is learned that measurement error can also improve the accuracy of a representative value assessment (see k realisation). Furthermore, this example shows that the measured strengths as combination of parameter measurements per location can give valuable insight. The minimum measured strength is in this case (almost) equal to the reference strength. The assessed strength based on individual characteristic parameter values however, is underestimating the reference strength by almost one meter.

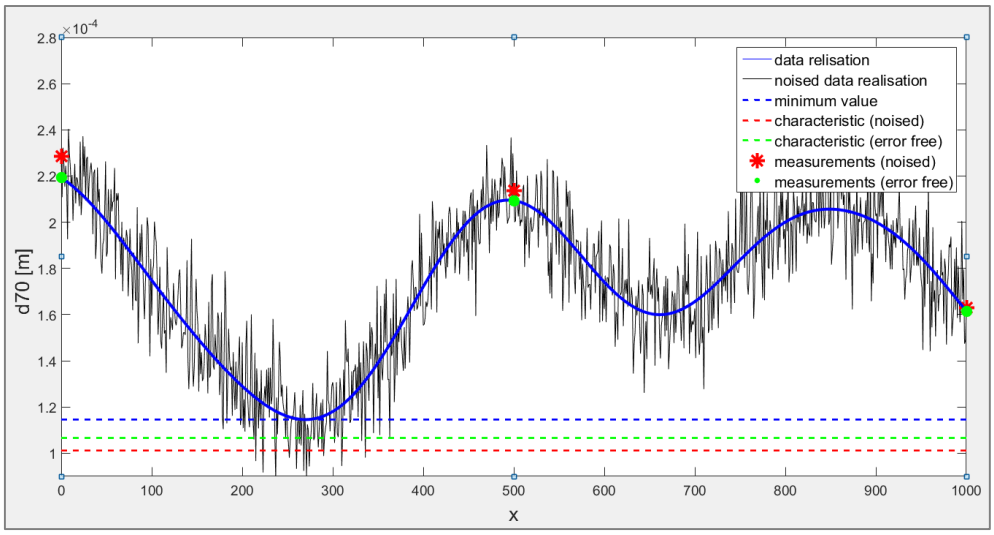


Figure 52 – Schematisation of a characteristic d70 value calculation based on 3 measurements; and the influence of measurement error on the calculation outcome. Example realisation based on standard input configuration.

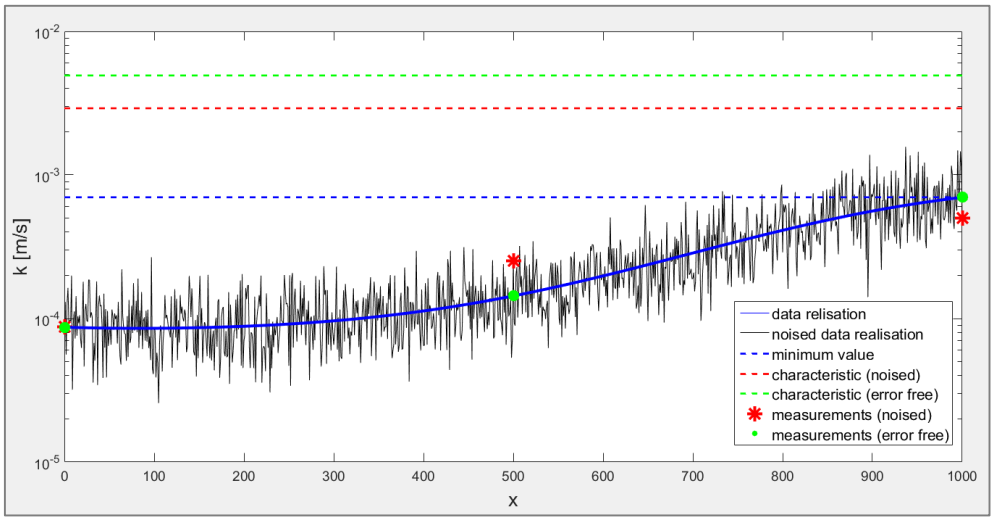


Figure 53 – Schematisation of a characteristic k value calculation based on 3 measurements; and the influence of measurement error on the calculation outcome. Example realisation based on standard input configuration.

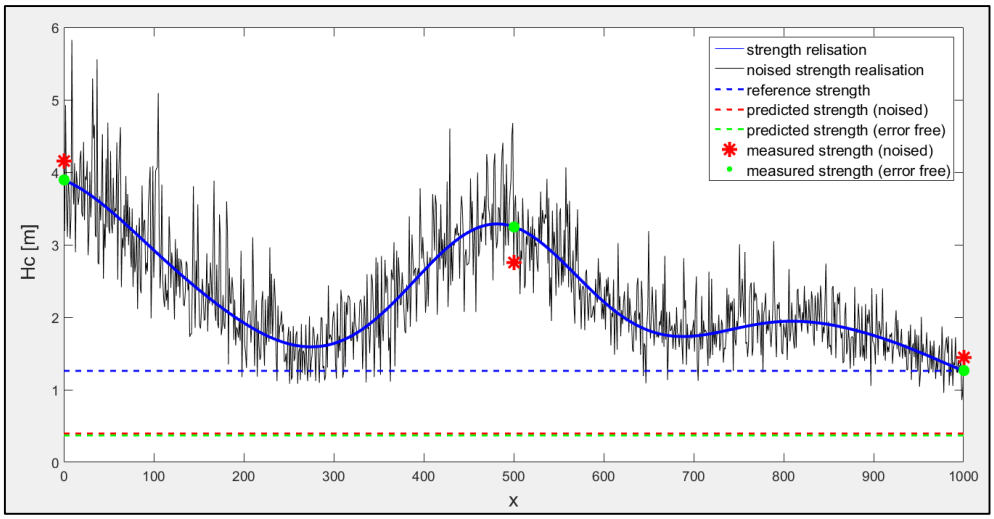


Figure 54 – Example of a strength (in terms of Hc) realisation with schematised the influence of measurement error on the assessed strength. Underlying data realisations have standard input configuration.

Table 11 – Numbers/calculation related to the example realisations in Figure 52 and Figure 53

	d70			**k**		
Measurements (noised)	2,29E-04	2,14E-04	1,63E-04	8,75E-05	2,52E-04	5,03E-04
Measurements (error free)	2,19E-04	2,09E-04	1,62E-04	8,69E-05	1,44E-04	6,99E-04
Mean measurements (noised)	m=2,02E-04			m=2,81E-04		
Mean measurements (error free)	m=1,97E-04			m=3,10E-04		
Std measurements (noised)	s=3,44E-05			s=2,09E-04		
Std measurements (error free)	s=3,09E-05			s=3,38E-04		
Student-t factor	t=2.92			t=2.92		
formula	p=m-(t*s)			p=exp(m(ln met)-(t*s(ln met)))		
Characteristic value (noised)	p=1,01E-04			p=2,91E-03		
Characteristic value (error free)	p=1,07E-04			p=4,93E-03		

Table 12 - Numbers/calculation related to the example strength realisation in Figure 54.

	Hc (m)		
Measured strength (noised)	4,15	2,75	1,44
Measured strength (error free)	3,89	3,25	1,26
Minimum measured strength (noised)	1,44		
Minimum measured strength (error free)	1,26		
Reference strength	1,26		
Assessed strength (noised)	0,40		
Assessed strength (error free)	0,37		

5. Representative variances

Overview of literature mentioning statistical parameters to represent distributions of soil properties d70 and k:

	Correlation d70	Correlation k	Mean d70	Mean k	Variance d70	Variance k
PC-Ring default values, table 3.1, (Engerink, Calle, Kruse, Selmeijer, & Blinde, 2007)	-	-	Nom 2E-4	Nom	CoV=0.15	CoV=1
Statistische invoergegevens PC Ring, table 3.4 (Vrouwenvelder, Steenbergen, & Slijkhuis, 1999)	180 m	-	Nom	-	CoV=0.15	
Representatieve parameterkeuze in formule van Sellmeijer, table 4.2 (TAW, 1999)	-	-	-	-	CoV=0.25	-
Overview of Sellmeijer parameters for standard dike, table 4.9 (Kanning W. , 2012)/ (TAW,1999)	180 m	600 m	2E-4	1.39E-4	CoV=0.182	CoV=1
Theoriehandleiding PC-Ring deel B: statistische modellen (Vrouwenvelder & Steenbergen, 2003B)	180 m	600 m	-	-	CoV=0.15	CoV=1
Spatial variation in liquefaction (Fenton & Vanmarcke, 1991)	-	~40 m	-	-	-	-
Variability in d70 of Vianen dataset, figure 4.24 (Kanning, 2012)	-	-	2.9E-4	-	CoV=0.344	-
Expert judgement from Betuwe Water Board (Geodelft, 1991)	-	-	3.6E-4	-	CoV=0.15	-
k values measured by slug tests, table 7 (Vienken & Dietrich, 2011)	-	-	-	4.61E-4	-	σ=5.48E-4
Input data for stochastic failure estimation (Aguilar-Lopez, Warmink, Schielen, & Hulscher, 2016)	-	-	3.33E-4	3.02E-4	CoV=0.15	CoV=1
d70 proeftuin bij Veessen (de Visser, Kanning, Koopmans, & Niemeijer, 2015)	-	-	4.16E-4	-	σ=2.15E-4	-
d70 proeftuin bij Veecaten (de Visser, Kanning, Koopmans, & Niemeijer, 2015)	-	--	2.05E-4	-	σ=0.25E-4	-
d70 proeftuin bij IJzendoorn (de Visser, Kanning, Koopmans, & Niemeijer, 2015)	-		4.20E-4	-	σ=1.08E-4	-
Input of random ground water flow (Kanning, 2012)/(Terzaghi, 1929)		10-60 m				

6. Model output

Because a random data generation script is used, it is important to analyse a sufficient amount of data sets to obtain stable results. In this section the model performance of 20.000 random data sets, each as a set of 1001 values, is analysed.

Mean and variance

Measuring over the complete data field of 2002000 data points, the *d70* field has a mean and variance equal to the input values: randomness is averaged out. The mean and variance of the *k* field differs from the input. This is caused by the translation from standard normal distributed values to lognormal distributed values. In contradiction with the *d70* field, the mean and variance of *k* are not completely stable after 20.000 realisations (randomness is not completely averaged out). The fluctuation range is limited to $0.1*10^{-4}$ m/s for both the mean and variance. With the calculation time significantly increasing using more data sets, this range is considered acceptable.

Table 13 – Measured output values. Simulated field N=20000, standard input configuration – no measurement error.

	d70 [m]	*k [m/s]*
μ_N	$2.00*10^{-4}$	[1.41 – 1.46] $*10^{-4}$
σ_N	$0.30*10^{-4}$	[1.45 – 1.52] $*10^{-4}$
$\mu(\sigma_n)_N$	$0.24*10^{-4}$	[0.55 – 0.59] $*10^{-4}$

Table 14 – Measured output values. Simulated field N=20000, standard input configuration + measurement error

	d70 [m]	**k [m/s]**
μ_N	$2.00*10^{-4}$	[1.55 – 1.59] $*10^{-4}$
σ_N	$0.33*10^{-4}$	[1.88 – 1.95] $*10^{-4}$
$\mu(\sigma_n)_N$	$0.29*10^{-4}$	[0.99 – 1.03] $*10^{-4}$

Note: μ_N is the mean value of all data points, σ_N is the standard deviation of all data points, $\mu(\sigma_n)_N$ standard deviation per data set averaged over all realisation. So the average standard deviation of a data set.

d70 data

In Figure 55 histograms of the simulated *d70* field are presented. The left figure is a histogram of the complete field, comprising 20.000*1001 values. The middle and right figures are histograms at locations x=0 and x=1000, comprising both 20.000 values.

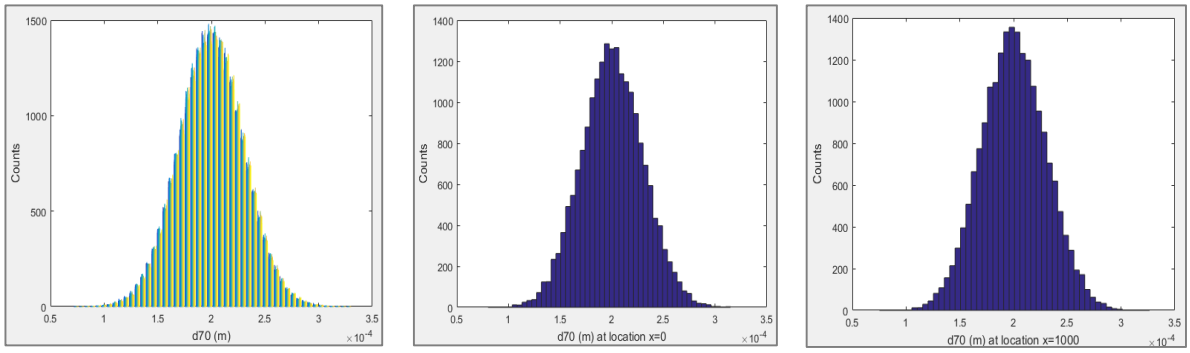


Figure 55 – Histograms of 20.000 d70 realizations. Simulated realisations according to standard input configuration

These figures show that the *d70* realizations are normally distributed as was planned in the model set-up. It also shows that the values at the boundaries of the model are normally distributed and that no negative *d70* values appear.

Figure 56 and Figure 57 present a typical example of the average mean and variance of *d70* values over all realizations for each location. Because this mean and variance have a very small deviation range along the domain it is concluded the model is not affected by location. This represents a theoretical equal probability of a value to occur for each location within the domain.

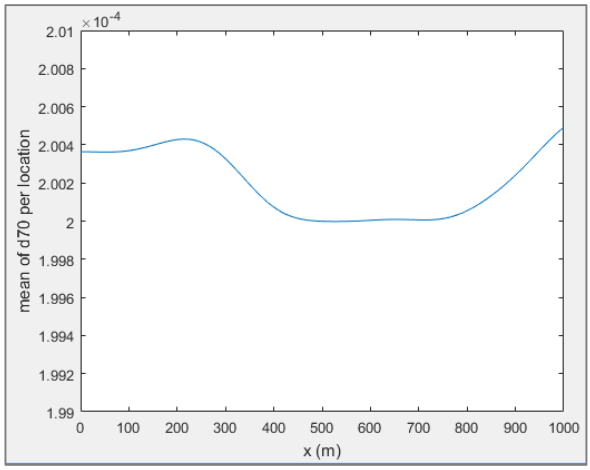


Figure 56 – Mean of d70 over 20.000 realizations for each location in the domain. Simulated field: standard configuration.

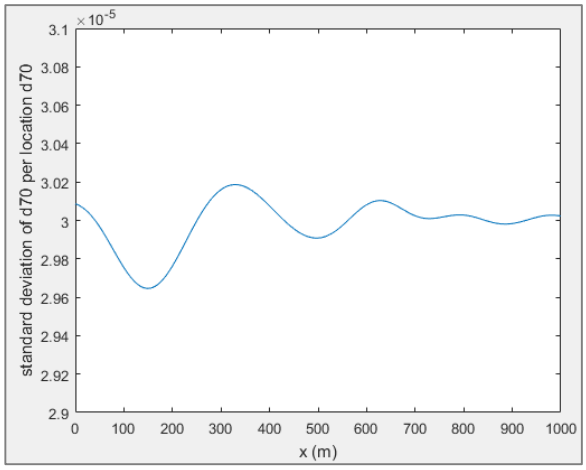


Figure 57 – Standard deviation of d70 over 20.000 realizations for each location in the domain. Simulated field: standard configuration.

k data

In Figure 58 histograms of the simulated *k* field are presented. The left figure is a histogram of the complete field, comprising 20.000*1001 *k* values. The middle and right figures are histograms at locations x=0 and x=1000, comprising both 20.000 values.

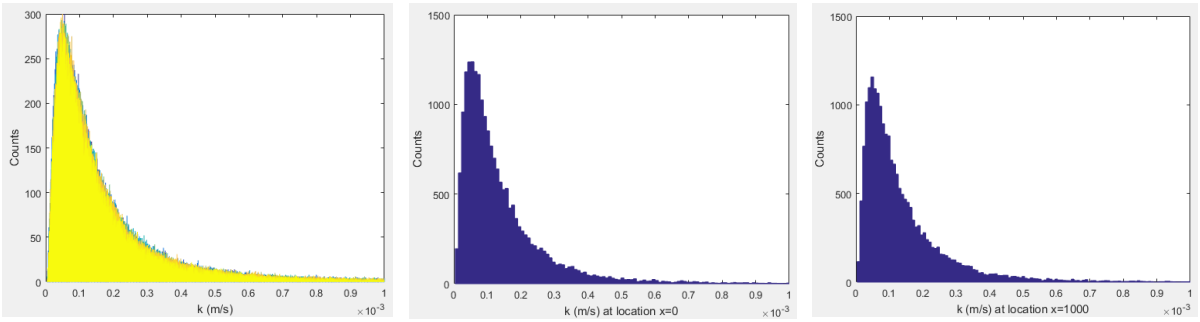


Figure 58 – Histograms of 20.000 k realizations of 1001 vales. Simulated field: standard configuration.

These figures show that the *k* realizations are log normally distributed as was planned in the model set-up. And that this is the case at the boundaries of the model as well. It shows that negative do indeed not occur.

Figure 59 and Figure 60 present a typical example of the mean and variance of *d70* values over all realizations for each location. Because this mean and variance have a very small deviation range along the domain it is concluded the model is not affected by location. This represents a theoretical equal probability of a value to occur for each location within the domain.

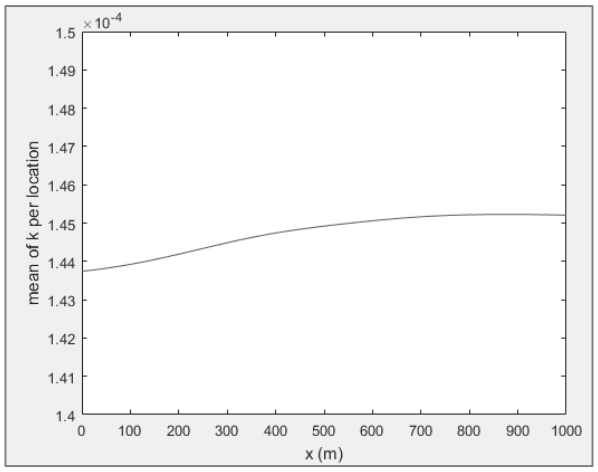


Figure 59 – Mean of k over 20.000 realizations for each location in the domain. Simulated field: standard configuration.

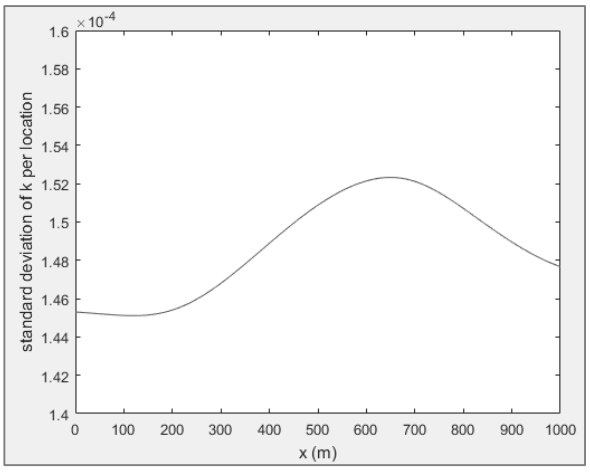


Figure 60 – Standard deviation of k over 20.000 realizations for each location in the domain. Simulated field: standard configuration.

Correlation structure

Correlation is measured using the Method of Moments. To check for the overall average correlation of all realizations the unbiased correlation is calculated using the average mean and variance of the complete data set (i.e. average values over all realisations).

Estimator of the correlation coefficient using Methods of Moments is described by:

$$\rho_z(\delta) = \frac{1}{(n-\delta)s^2}\sum_{i=1}^{n-\delta}(z(x_i)-\mu_z)(z(x_{i+\delta})-\mu_z)$$

In which $(n-\delta)$ are the number of data pairs that have separation distance δ ('lag'), μ_z is the mean of the set and s^2 equals the sample variance.

Using this formula, the measured correlation is compared with the correlogram (autocorrelation function) that was used as input for the correlation matrix. The results are presented in Figure 15 and Figure 16.

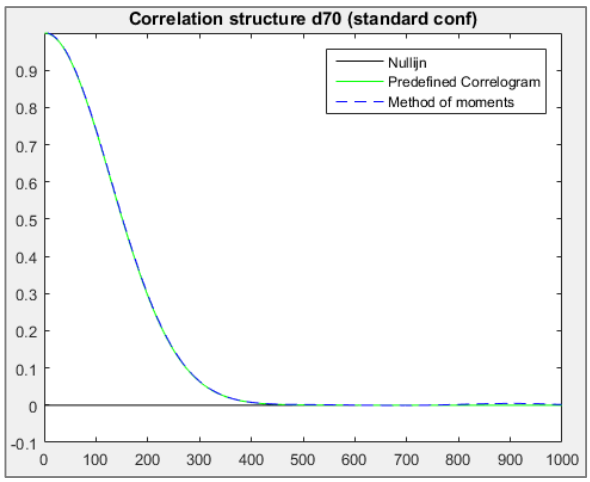


Figure 61 – Correlation structure of d70 realizations. Simulated field: N=20000, standard configuration.

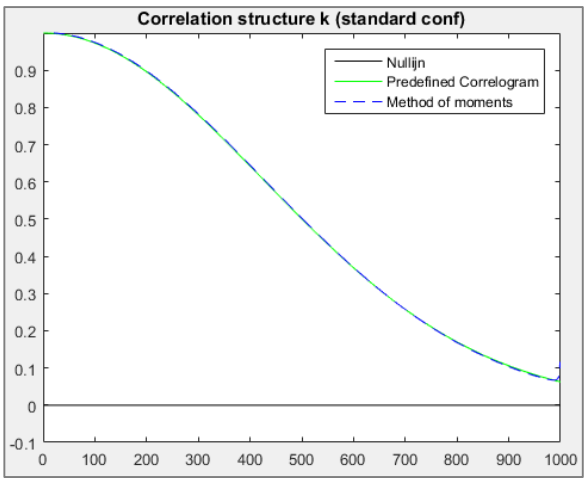


Figure 62 – Correlation structure of k realizations. Simulated field: N=20000, standard configuration.

Influence of model boundaries

The ratio correlation length/section length (from now on referred to as δ_0/x') influences the model output. If the correlation length is not much smaller than the model length, the measured variance within a realisation is affected.

First, because of correlation the highest or lowest value of a realization is mostly found at the boundaries of the domain. This is independent on how the domain is chosen. If for example a domain of 1000 meter is randomly selected from a much larger domain, still de maximum/minimum values of the smaller domain within the larger domain are most of the cases at the boundaries.

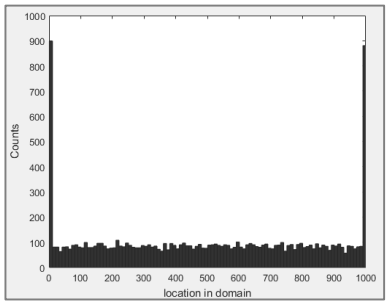


Figure 63 – Histogram of locations of observed d70 minima over all realizations. Simulated field: N=20000, standard configuration.

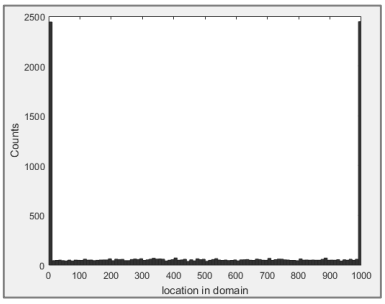


Figure 64 – Histogram of locations of observed k maxima over all realizations. Simulated field: N=20000, standard configuration.

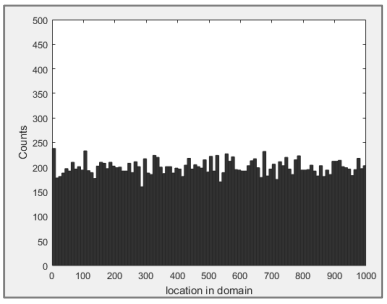


Figure 65 – Histogram of test run. Simulated field: N=20000, $\boldsymbol{\mu}$=0, $\boldsymbol{\sigma}$=1, $\boldsymbol{\delta_0}$=5 and $\boldsymbol{x}'$=3000

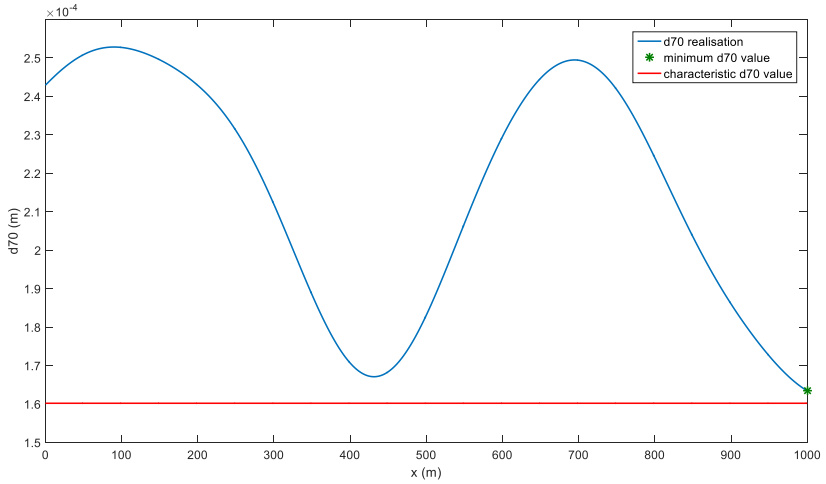
A test run, see Figure 65, shows that when the ratio $\frac{\delta_0}{x'} \to 0$ the minima/maxima of realizations are uniformly distributed along the domain.

The variance, represented with the standard deviation, can be measured in 2 ways. First as the deviation of values from the 'global average' and second as the deviation of values from the 'local average'. The global average is defined as the mean of the entire set including 1001 values multiplied with 20.000 realizations. However, each realization has also an own mean which is defined as local mean.

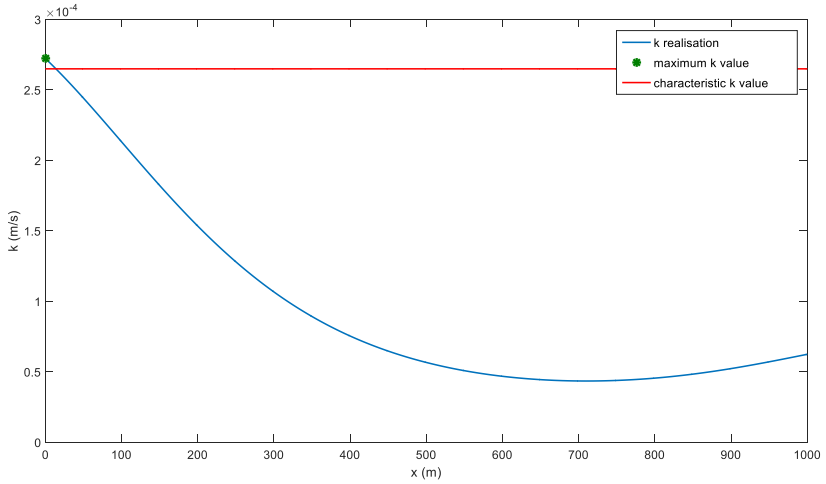
Within one realizations the mean differs from the average mean over all realizations. Considering the mean of one realizations the local variance is the standard deviation from one realization. This standard deviation can be averaged with all other local variances. Within one realization the variance is affected by correlation. Because a realization has a limited number of values and the values within the domain correlate with each other, there is not sufficient 'fluctuation space' to reach the underlying mean and variance. Complete variance can only occur if the ratio $c\delta_0/x'$ goes to zero. Therefore the local variance in a correlated realization is always smaller than the global variance. Measuring the variance over all realization can be seen as measuring the variance of 1 correlated series with 'infinite' length (all individual lengths are added up). Then, although the correlation is present, the ratio correlation length/domain length goes to zero and the measured variance equals the underlying variance.

7. Example realisation

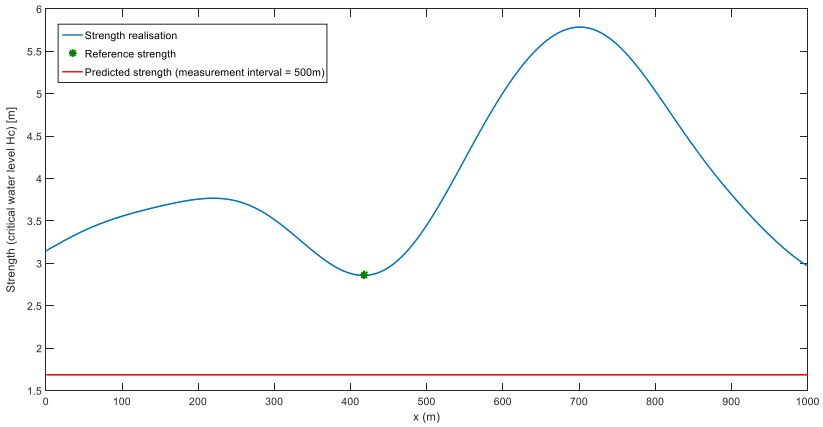
The combination of a d70 and k realisation result (by applying the model of Sellmeijer) in one strength realisation. The combination of actual representative parameter values results in the reference strength: the minimal strength of a realisation. The actual representative values are the parameter values at the weakest location, i.e. the location where the reference strength is determined. The combination of calculated characteristic values results in an assessed strength, assumed constant for one dike section.



+



=



8. Strength distributions

In Table 15 an overview of some main strength parameters is provided. From 20.000 runs the average the reference strength and several assessed strengths are calculated. On average (over all measurement configurations) the reference strength is underestimated by 0.5 meter in the standard configuration and 0.8 meter in de standard configuration with measurement error.

Table 15 – Main strength parameters calculated of a simulation with standard input

Property	Explanation	measurement error neglected	with measurement error
$\boldsymbol{\mu_{Hc}}$ $\boldsymbol{\mu_{i_c}}$	Field mean of all Hc values consisting out of 20.000 runs of each 1001 values	$H_c = 3.4\ m$ $i_c = 0.113$	$H_c = 3.4\ m$ $i_c = 0.113$
$\boldsymbol{\mu_{ref}}$	Mean of 20.000 reference strengths	$H_c = 2.3\ m$ $i_c = 0.077$	$H_c = 2.3\ m$ $i_c = 0.077$
	Mean of 20.000 assessed strengths per measurement interval		
	500 meter	$H_c = 1.0\ m$ $i_c = 0.033$	$H_c = 0.8\ m$ $i_c = 0.027$
	250 meter	$H_c = 1.4\ m$ $i_c = 0.047$	$H_c = 1.2\ m$ $i_c = 0.040$
	100 meter	$H_c = 1.7\ m$ $i_c = 0.057$	$H_c = 1.4\ m$ $i_c = 0.047$
	50 meter	$H_c = 1.9\ m$ $i_c = 0.063$	$H_c = 1.5\ m$ $i_c = 0.050$
	1 meter	$H_c = 2.0\ m$ $i_c = 0.067$	$H_c = 1.6\ m$ $i_c = 0.053$
$\boldsymbol{Hc(\mu_{output})}$ $\boldsymbol{i_c(\mu_{output})}$	Hc as calculated with the output mean of d70 (=2.0e-4), k(=1.4e-4) and D(=15)	$H_c = 2.9\ m$ $i_c = 0.097$	$H_c = 2.9\ m$ $i_c = 0.097$

To provide more insight in the influence of measurement intervals, probability density functions (histograms) of the reference strength and assessed strengths related to 5 measurement intervals are made. The results are presented in Figure 66 for standard input.

The figures show that the total of runs provide a range of reference strengths and assessed strengths. With decreasing measurement intervals, the probability density functions shift towards the reference strength. But the difference between a measurement interval of 50 meter or 1 meter is limited. Furthermore, the difference between reference strengths and assessed strengths, even if the measurement interval is 1 meter, stays significant. This result is even more obvious if measurement error is simulated as well (Figure 66). In the case with measurement error the distributions of assessments are even further off from the distributions of the reference strengths.

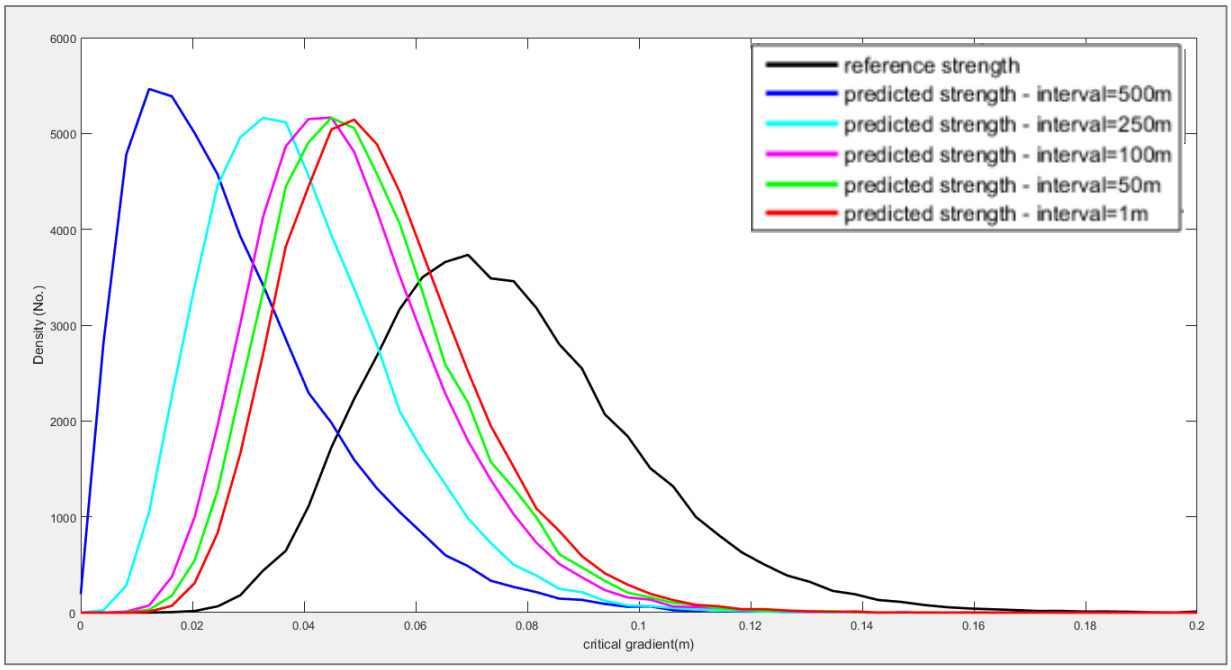


Figure 66 – Result of Monte Carlo analysis: overview of density distribution of the reference strength and assessed strengths as result of five different measurement configurations. Simulation based on standard input configuration.

Where Figure 66 gives an overview of the reference strengths and assessed strengths individually, Figure 67 provide probability density distributions of the difference between reference and assessed strengths (as calculated for each run apart). Those figures show clearly the effect of measurements on the accuracy of assessments.

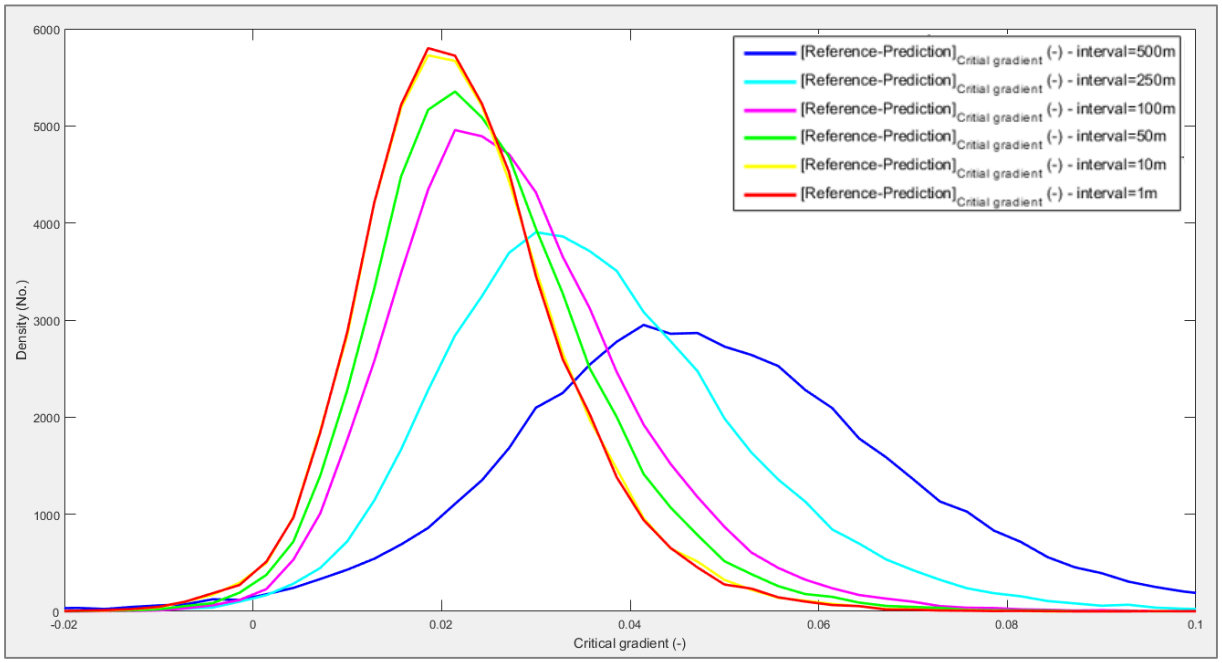


Figure 67 – Result of Monte Carlo analysis: overview of density distributions of the difference between reference strength and assessed strengths as result of six different measurement configurations. Simulation based on standard input configuration with measurement error.

9. Specified vs random measurement locations

In the standard case the analysis is made with fixed distances between the measurements. These distances are varying from 500 meters (3 measurements in case of a 1000-meter section length) to 1 meter (1000 measurements in case of a 1000-meter section length).

Next to equal measurement locations and density configurations for each realization, it is possible to analyse the case where measurements are taken randomly along the domain. In Figure 68 (standard configuration) and Figure 69 (standard configuration with measurement error) the result of this analysis is plotted. To make a comparison possible the case of specified measurement locations (standard case) is plotted as well in the same figure (black is standard case, red is deviating case).

With the numbers of measurements increasing the cases merge. This is because also in the approach of random measurements, eventually all locations are measured (locations can only me measured once in this set up). Only for low number, the randomness appears in increased bandwidths. This is caused by an increase in the number of runs where the strength is overestimated.

In the case with measurement error this effect is less. That is explained by the random character that measurement error already implies.

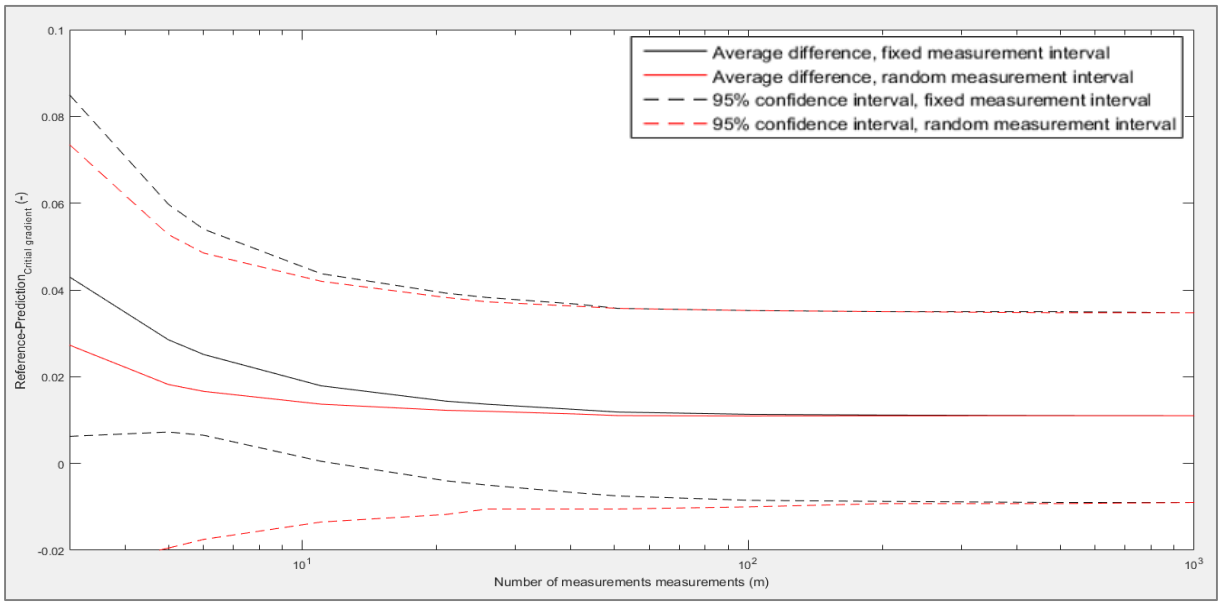


Figure 68 – Comparison between an approach in which fixed and equal measurement intervals are used and an approach in which measurement locations are random. Realisations based on standard input configuration, no measurement error

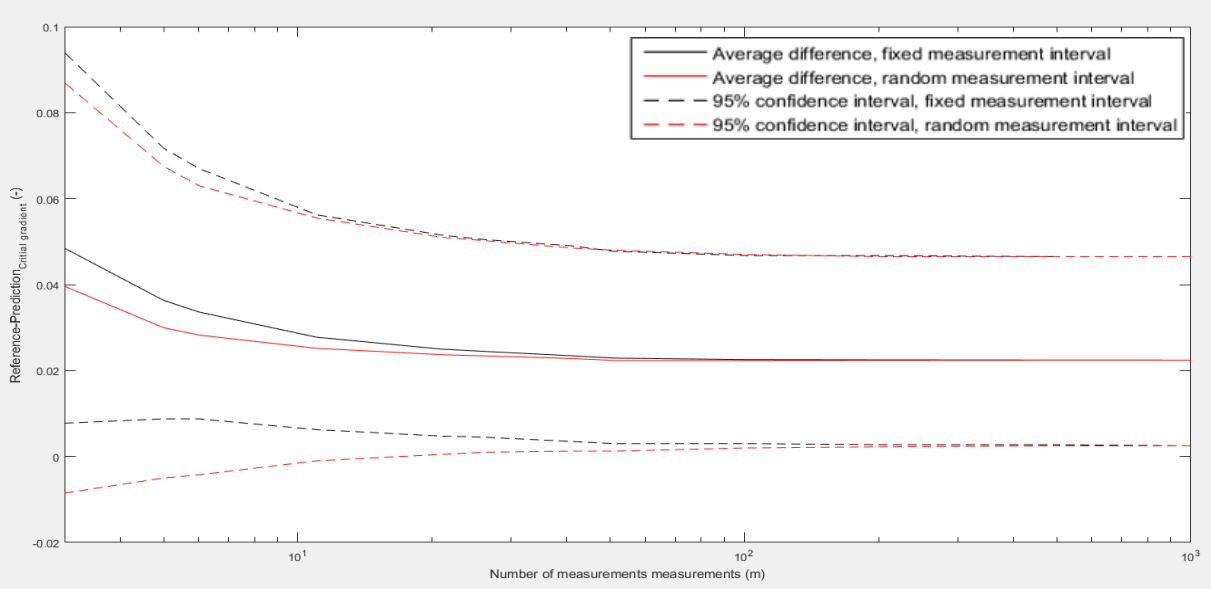


Figure 69 – Comparison between an approach in which fixed and equal measurement intervals are used and an approach in which measurement locations are random. Realisations based on standard input configuration with measurement error

The effect that specified measurement locations and intervals result in smaller error bandwidths for small measurement densities is explained by correlation in the d70 and k data. Due to correlation in a relatively small domain, extreme values within a set are often found at the boundaries of the domain. Therefore, a relatively large measurement interval results in a high measured variance and conservative assessment. While with random measurement locations it is possible that measurement result in in very small range of values: measured values can be accidently almost the same. This is especially the case if measurement locations are close to each other in a correlated soil. The variance of that measurement set is then very small, possibly resulting in a relatively unsafe assessment.

10. Boundary effect

In case all values of a data set (realisation) are known, the calculated mean and variance are the actual mean and variance of that set. Using a data set that has a perfect normal density function, the characteristic value calculation would give the actual 95% boundaries.

Considering the *d70* data sets it is expected these are indeed normally distributed because a normal distributed data generator is used. But because of correlation and a random generator in combination with limited domain length it is possible that the distribution is deviating from the distribution of input values. The exact underlying input properties are not reached at smaller domains because of boundary influence. In Figure 70 an example realisation is presented. The data is correlated and standard normally distributed. In the figure the average of the complete set (5000 values) and averages of subsets (of 1000 values) are pictured. It is noticed that even if the correlation length/domain length is relatively small ($\frac{\delta_0}{x'} \ll 1$), the output mean is not necessarily equal to the input mean. This deviation is for local averages sometimes even higher.

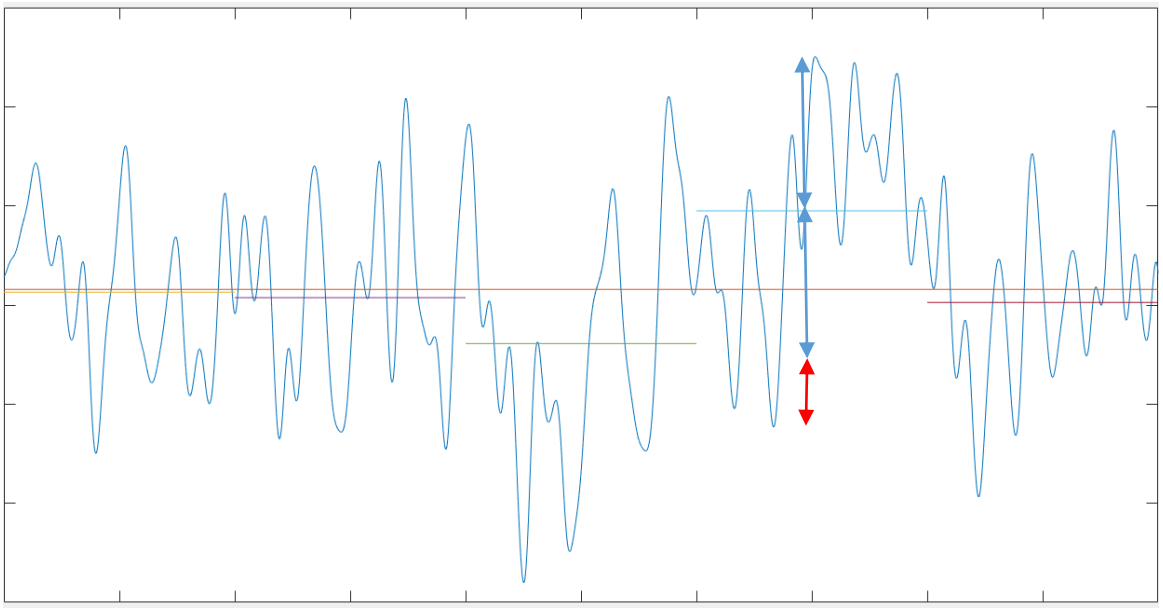


Figure 70 – Difference between global and local averages. Used input configuration: $\mu = 0$, $\sigma = 1$, $\delta_0 = 50$

With respect to characteristic value calculation, next to the mean also the variance of the data is of importance. Of special interest here is the way peaks in the data set contribute to the calculated variance of a data set. In a data set with limited length (compared to the correlation length), it is possible that the observed minimum and maximum are not evenly far off from the mean. This is illustrated for the fourth subsection in Figure 70. This means that calculated variance might not be a good representation of a local peak, either the maximum or the minimum.

In case of *d70* data, a high maximum increases the variance of the total set and therefore will decrease the characteristic 95% lower boundary. If the minimum value is then less far off from the average than the maximum, the result is an overestimation of the minimum present *d70* value. The opposite may also occur if the variance is too little to account for local minima. This results in underestimation of the actual present d70 value. Both cases are illustrated in Figure 71 with an example realization and characteristic value calculation where influence of measurement error is ignored for the moment.

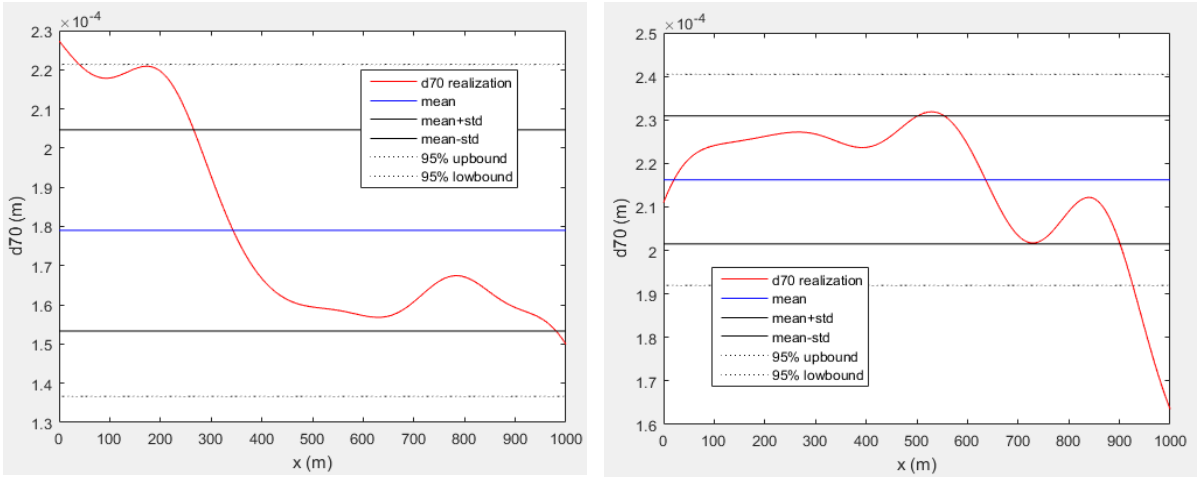


Figure 71 - 2 example realizations of d70 to show cases of under- and overestimation. In both examples the same calculation procedure is followed.. Realisations based on standard input configuration, no measurement error.

Permeability data sets show relative high variances. TAW (1999) and Deltares (2002) advise to consider high variable data to be lognormal distributed. In this study the k values are in fact lognormal distributed and show relatively high variances. Instead of the mean and variance of the sample values, the mean and variance of the natural logarithmic of the sample values are used. In this approach boundary effect are also present when data is correlated. Analogous to the analysis of d70, two cases in which k is overestimated/underestimated are presented in Figure 72.

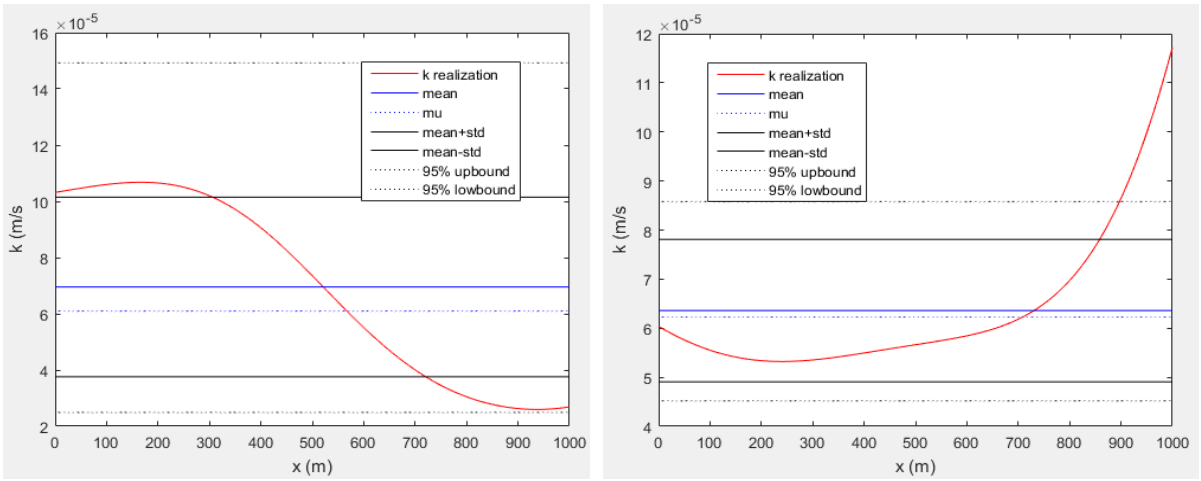


Figure 72 – 2 example realizations of k to show cases of under- and overestimation. In both examples the same calculation procedure is followed. Realisations based on standard input configuration, measurement error is neglected.

To show what characteristic value calculations actually do, an example realization of an uncorrelated *d70* data set is presented in Figure 73. About 5% of the values exceed the 95% upper boundary and 5% of the values drop below the 95% lower boundary. This means that in uncorrelated data many peaks along the domain are present, while in correlated data the number of peaks is smaller. Uncorrelated soil parameters result in more possible weak spots within the domain beyond the 95% boundary that is assumed to be the conservative boundary.

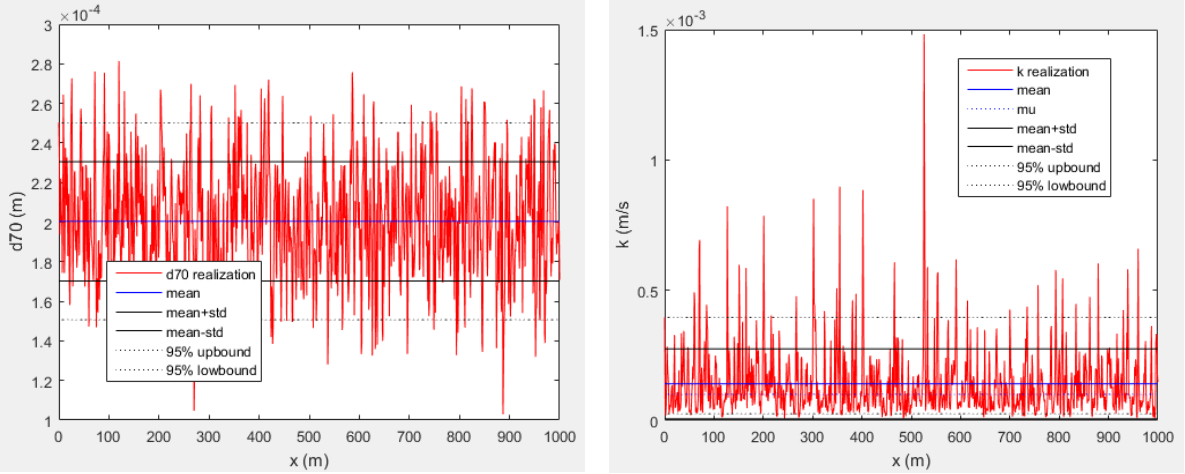


Figure 73 – Example realization of uncorrelated d70 data set (right) and uncorrelated k data set (left)

11. Safety level

Safety factor

In the detailed assessment the critical strength is corrected with a partial safety factor:

- Safety criterion follows: $\Delta H * F_s \leq H_c \quad , \ [1.2 \leq F_S \leq 1.6]$. This means that the situation is safe if the load is smaller than the calculated resistance divided by the safety factor. In the analysis this is represented by dividing the assessed strengths by the safety factor before those are subtracted from the reference strengths.
- Average difference and confidence interval shifts upwards. Unsafe assessments, for as far as they were present, are banned out.
- Simulations with safety factors of 1, 1.2 and 1.6 are presented in Figure 74 for the standard input configuration.

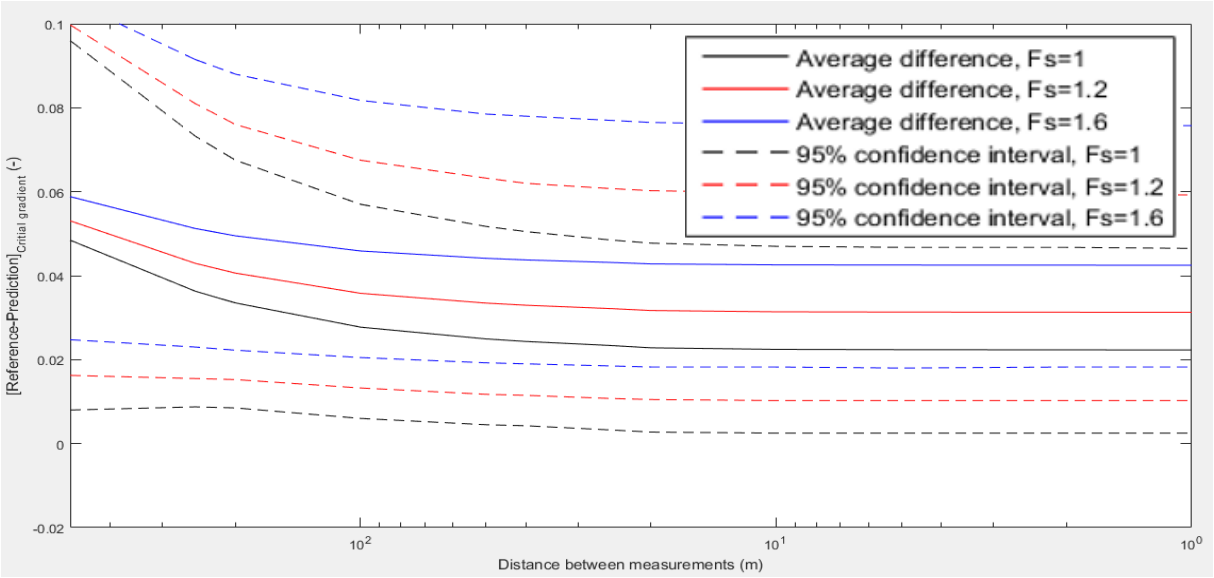


Figure 74 – influence of safety factors on average difference between and 95% confidence interval of reference and assessed strength. Simulation with standard input configuration and measurement error.

Xx% characteristic upper/lower boundary

In Figure 75 is illustrated what the effect of a certain characteristic boundary is. In the assessment usually a 95% lower boundary is used to estimate the characteristic value of d70. This is compared with the bandwidth of errors when a 99% lower boundary is used.

Figure 75 presents the bandwidths of errors for which standard input is used to generate data sets. *Figure 81* presents the results in case d70 is uncorrelated.

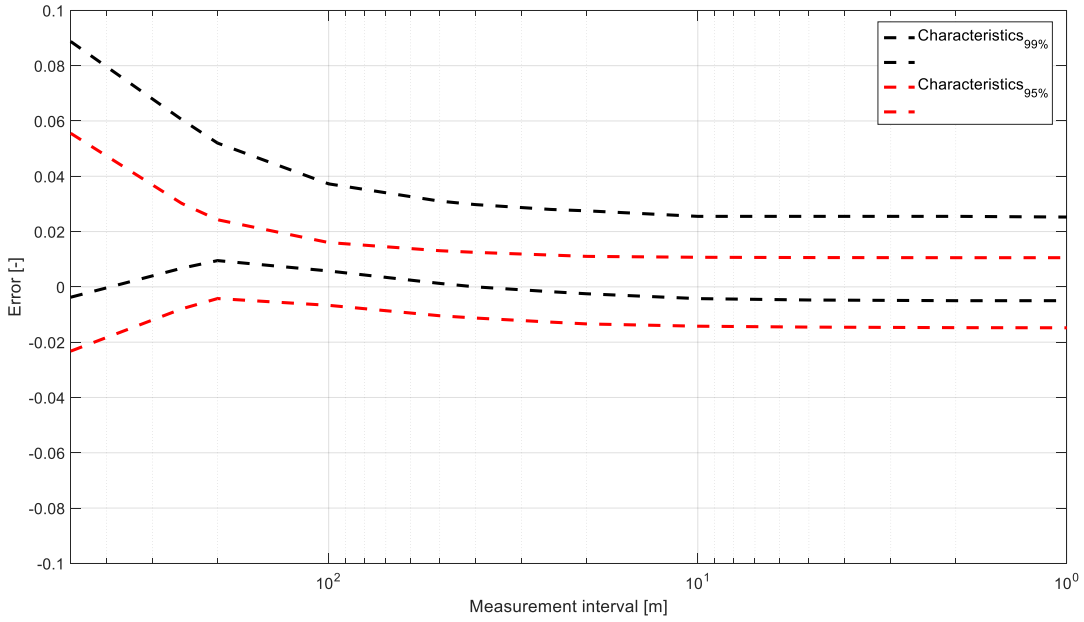


Figure 75 – influence of characteristic lower/upper boundaries: comparison between characteristic 95% lower d70 boundary and characteristic 99% lower d70 boundary. Simulation: only variations in d70, standard input, measurement error neglected.

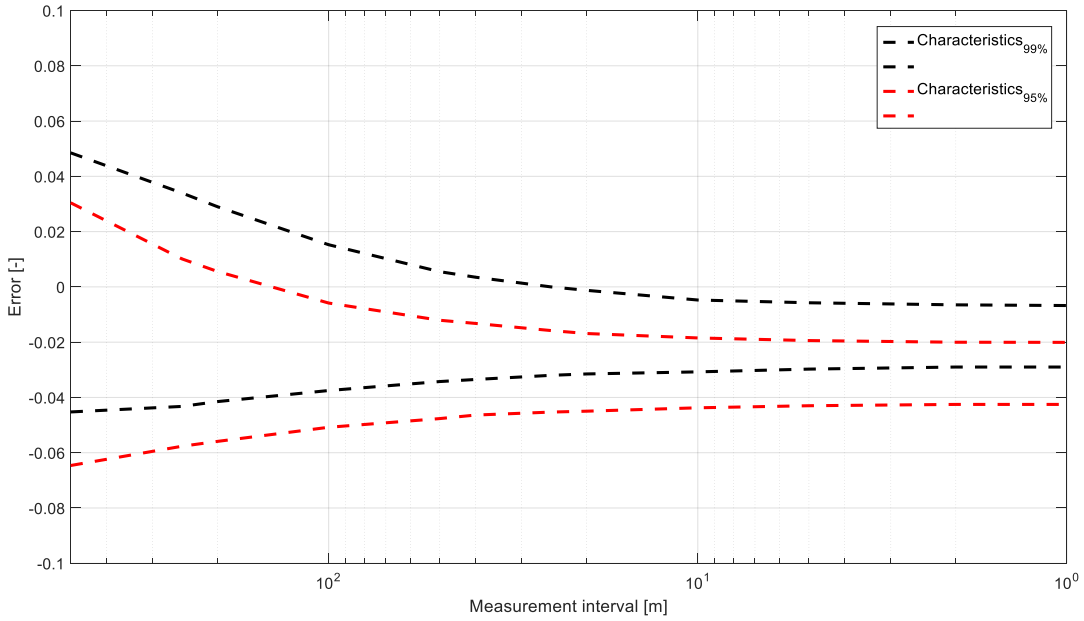


Figure 76 – influence of characteristic lower/upper boundaries: comparison between characteristic 95% lower d70 boundary and characteristic 99% lower d70 boundary. Simulation: only variations in d70, uncorrelated, measurement error neglected.

12. Influence of seepage length

The error is defined as the difference between reference and assessed strength relative to the seepage length. Therefore, the bandwidth of errors is hardly influenced by the chosen seepage length (Figure 77). But in terms of critical water level, the bandwidth of errors is much more influenced. Soil and measurement conditions that result in an error, result in an increased error if the seepage length is increased (Figure 78). This is because a higher actual strength allows for higher absolute deviations,

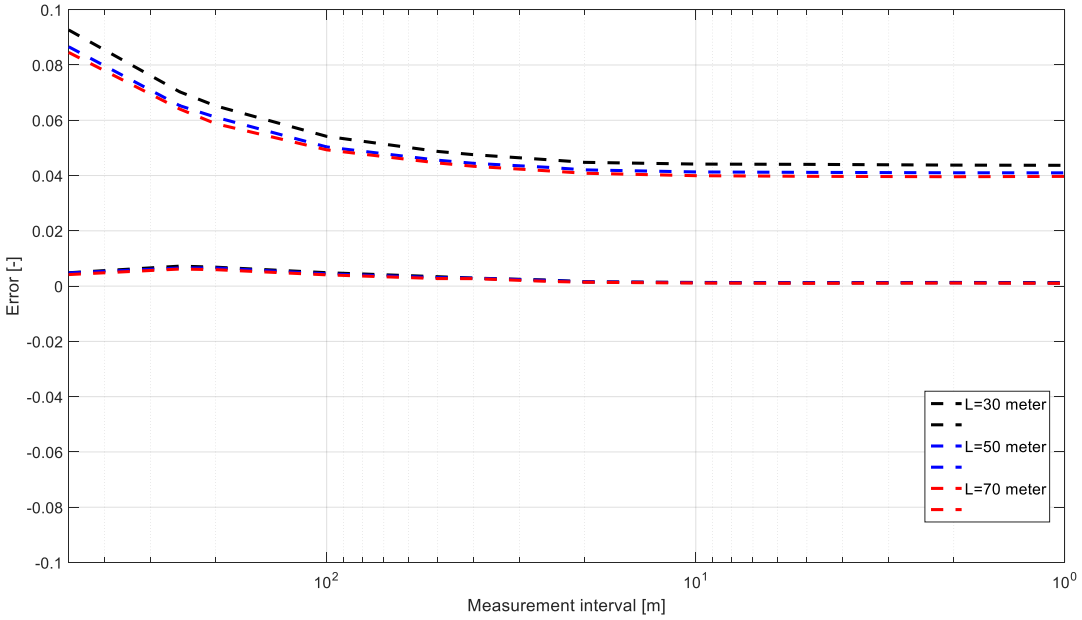


Figure 77 – influence of seepage length L to the bandwidth of errors

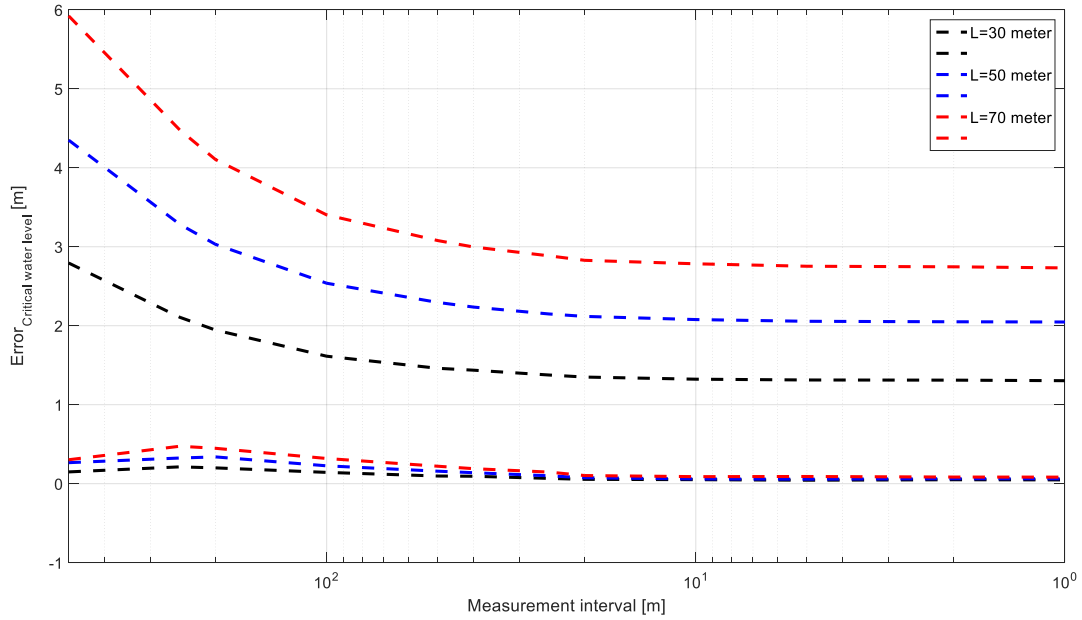


Figure 78 – influence of seepage length to the absolute deviation of reference and assessed strength in terms of critical water level (m).

13. Influence of dike section length

The bandwidth of errors shifts downwards in the error domain if the section length is increased. This is explained by the length effect. The length effect states that a probability of an unfavourable combination of soil properties increases if more fluctuation are present. The number of possible fluctuations is given by the correlation length and domain length. When the domain length is increases, the probability of overestimation increases. Unfavourable combinations are more likely to be present and are undetected in the assessments.

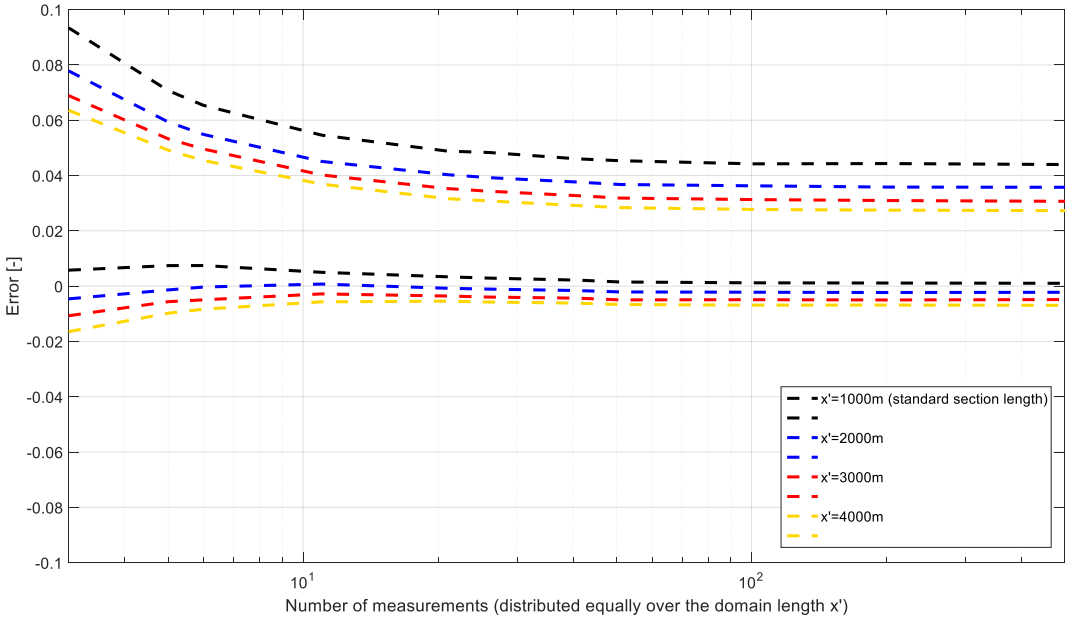


Figure 79 – Influence of section length if d70 (180m) and k (600m) are strongly correlated.

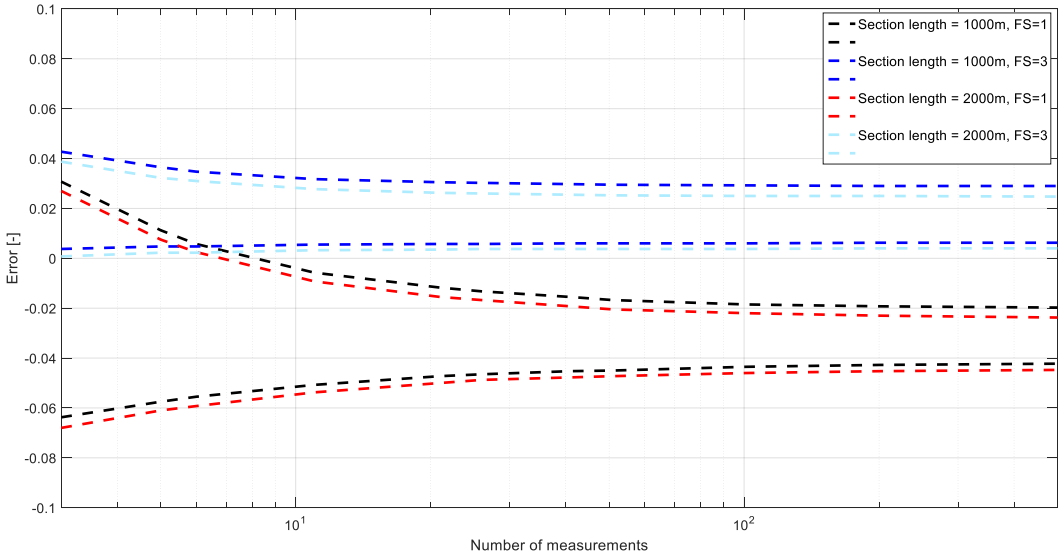


Figure 80 – Influence of section length if d70 (1m) and k (30m) are hardly correlated. Cases in which a safety factor of 1 respectively 3 is applied to the assessed strengths before compared with the reference strength.

14. Influence of measurement error

In Figure 81 the result of the sensitivity analysis towards ϕ is presented. Four subfigures are presented. In 3 subfigures the sensitivity of the model outcome to variations in phi related to only respectively d70, k or D data. For example, in case of the sensitivity to measurement error in d70 data, the magnitude of measurement error of d70 data is varied while the error in k and D data is set zero. In the lower right subfigure, the result is presented of an analysis in which measurement error of all three parameters is varying combined.

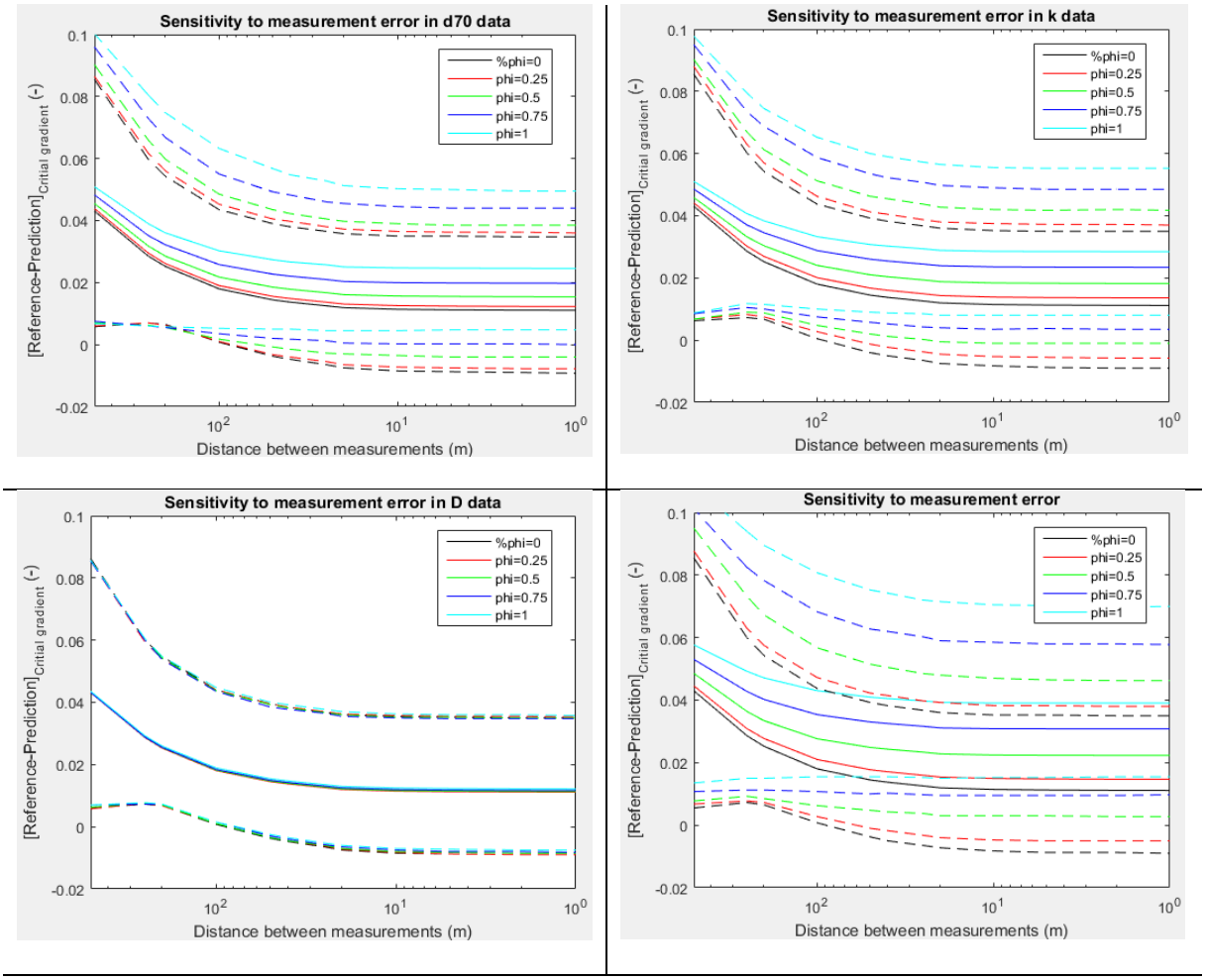


Figure 81 – Overview of model sensitivity with respect to ϕ: representing the magnitude of measurement error as fraction of the spatial variability.

15. Influence of correlation length

Varying d70 correlation length

First the sensitivity of the model outcome to the correlation length used to simulate d70 data is evaluated. The correlation lengths of k is kept constant (standard input)

The sensitivity is tested for 5%, 10%, 15%, 20%, 25%, 50%, 75%, 100%, 125% and 150% of the correlation length of d70 in the standard configuration:

Measurement error neglected

The result of the analysis for the simulations without noise, so no measurement error, is presented in Figure 82. The measured output statistics are presented in Table 16.

Table 16 – Measured d70 output with respect to varying d70 correlation length input. Simulation based on standard input configuration, measurement error neglected.

d70 correlation	**5%**	**10%**	**15%**	**20%**	**25%**	**50%**	**75%**	**100%**	**125%**	**150%**
$\boldsymbol{\delta_{0,input}}$ **[m]**	9	18	27	36	45	90	135	180	225	270
$\boldsymbol{\mu_N\ [10^{-4}\ m]}$	2,00	2,00	2,00	2,00	2,00	2,00	2,00	2,00	2,00	2,00E
$\boldsymbol{\sigma_N\ [10^{-5}\ m]}$	3,00	2,99	3,00	3,00	3,00	2,99	3,01	3,01	3,0	3,00
$\boldsymbol{\mu(\sigma_n)_N\ [10^{-5}\ m]}$	2,97	2,93	2,91	2,88	2,84	2,68	2,54	2,39	2,26	2,14

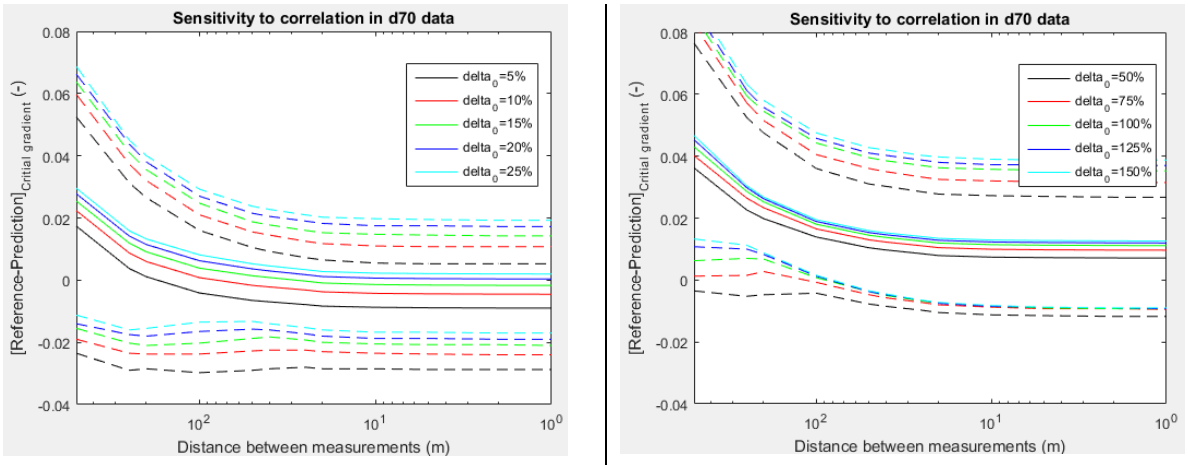


Figure 82 – Sensitivity of model to correlation in d70 data. Standard input configuration, no measurement error, $\boldsymbol{\delta_0(d70)}$ *is varying between 5% and 150% of the standard input value (180m).*

The green line in the left plot is equal to the base case as the correlation length for that simulation is equal to 100% of the standard configuration. It can be noticed that the shape and magnitude of the 95% confidence interval is hardly changing with varying correlation scales. However, the bandwidth shifts downwards meaning in more runs strength is overestimated instead of underestimated. For a correlation length of for example 9 meter, the average difference (reference-assessment) is even negative for high measurement densities.

Measurement error considered

The result of the analysis for the simulations with noise, so with measurement error, is presented in Figure 84. The measured output statistics are presented in Table 17.

Table 17 – Measured d70 output with respect to varying d70 correlation length input. Simulation based on standard input configuration and measurement error.

d70 correlation	5%	10%	15%	20%	25%	50%	75%	100%	125%	150%
$\delta_{0,input}$ [m]	9	18	27	36	45	90	135	180	225	270
$\mu_N\ [10^{-4}\ m]$	2,00	2,00	2,00	2,00	2,00	2,00E	2,00	2,00	2,00	2,00
$\sigma_N\ [10^{-5}\ m]$	3,00	3,00	3,00	3,00	3,00	3,35	3,35	3,36	3,35	3,36
$\mu(\sigma_n)_N\ [10^{-5}\ m]$	3,32	3,30	3,27	3,24	3,21	3,09	2,97	2,86	2,76	2,66

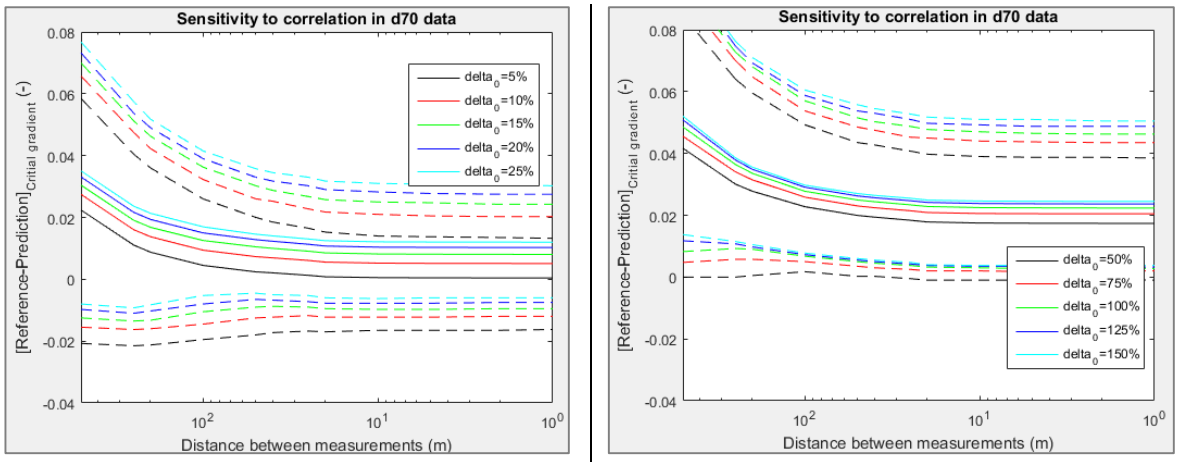


Figure 83 – Sensitivity of model to correlation in d70 data. Standard input configuration with measurement error, $\delta_0(d70)$ is varying between 5% and 150% of the standard input value (180m).

For the simulation with noise data mainly the same effects are observed. The simulation with noise results in less overestimation.

Varying k correlation length

Secondly the sensitivity of the model outcome to the correlation length used to simulate k data is evaluated. The correlation lengths of d70 is kept constant (standard input).

The sensitivity is tested for 5%, 10%, 15%, 20%, 25%, 50%, 75%, 100%, 125% and 150% of the correlation length ok k in the standard configuration.

Measurement error neglected

The result of the analysis for the simulations without noise, so no measurement error, is presented in Figure 85. The measured output statistics are presented in Table 18.

Table 18 – Measured k output with respect to varying k correlation length input. Simulation based on standard input configuration.

k correlation	5%	10%	15%	20%	25%	50%	75%	100%	125%	150%
$\delta_{0,input}$ [m]	30	60	90	120	150	300	450	600	750	900
$\mu_N\ [10^{-4}\ m/s]$	1,43	1,43	1,43	1,44	1,44	1,43	1,44	1,43	1,43	1,45
$\sigma_N\ [10^{-4}\ m/s]$	1,48	1,47	1,48	1,47	1,48	1,48	1,46	1,46	1,48	1,47
$\mu(\sigma_n)_N\ [10^{-4}\ m/s]$	1,33	1,24	1,17	1,10	1,05	0,84	0,68	0,56	0,48	0,42

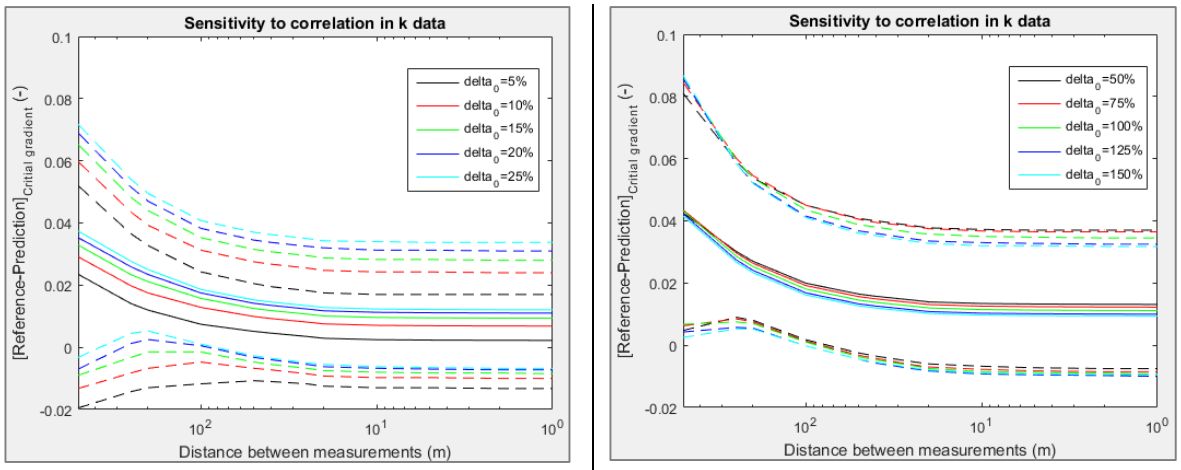


Figure 84 – Sensitivity of model to correlation in k data. Standard input configuration, no measurement error, $\boldsymbol{\delta_0(k)}$ is varying between 5% and 150% of the standard input value (600m).

The same trend is visible as for d70. Small correlation lengths shift the bandwidth downwards to more overestimation and less underestimation of strength. It can furthermore be noticed that the model is less sensitive to changes in case of high correlation lengths (right plot) than to changes in case of small correlation lengths (left plot).

Measurement error considered

The result of the analysis for the simulations with noise, so with measurement error, is presented in Figure 86. The measured output statistics are presented in Table 19.

Table 19 – Measured k output with respect to varying k correlation length input. Simulation based on standard input configuration and measurement error.

k correlation -with noise	**5%**	**10%**	**15%**	**20%**	**25%**	**50%**	**75%**	**100%**	**125%**	**150%**
$\boldsymbol{\delta_{0,input}}$ **[m]**	30	60	90	120	150	300	450	600	750	900
$\boldsymbol{\mu_N\ [10^{-4}\ m/s]}$	1,57	1,57	1,57	1,56	1,56	1,57	1,58	1,57	1,57	1,59
$\boldsymbol{\sigma_N\ [10^{-4}\ m/s]}$	1,89	1,90	1,89	1,89	1,89	1,88	1,93	1,89	1,92	1,91
$\boldsymbol{\mu(\sigma_n)_N\ [10^{-4}\ m/s]}$	1,74	1,64	1,57	1,49	1,44	1,23E	1,1	1,00	0,94	0,89

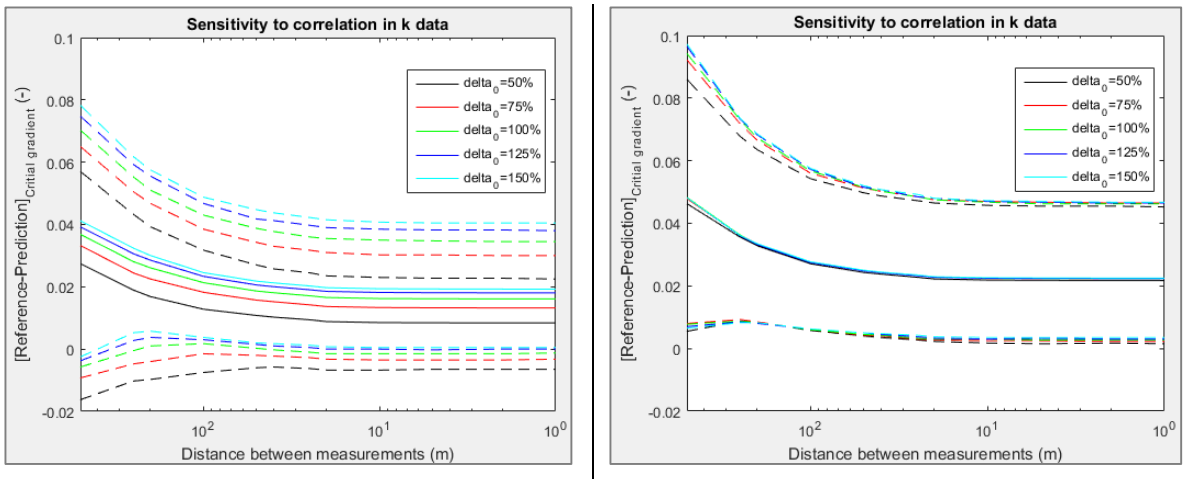


Figure 85 – Sensitivity of model to correlation in k data. Standard input configuration with measurement error, $\boldsymbol{\delta_0(k)}$ is varying between 5% and 150% of the standard input value (600m).

For the simulation with noise data the same effects are observed. As is the case for the standard configuration, the simulation with noise results in less negative differences. For small correlation lengths this effect is less striking than for d70 data.

Combining variation of d70 and k correlation lengths

Finally, the sensitivity of the model outcome to simultaneously changing correlation lengths is evaluated. The sensitivity is tested for 5%, 10%, 15%, 20%, 25%, 50%, 75%, 100%, 125% and 150% of the correlation lengths in the standard configuration.

Measurement error neglected

The result of the analysis for the simulations without noise is presented in Figure 86. The y-axis is shifted downwards 0.02 to be able to capture the negative differences better in the plot.

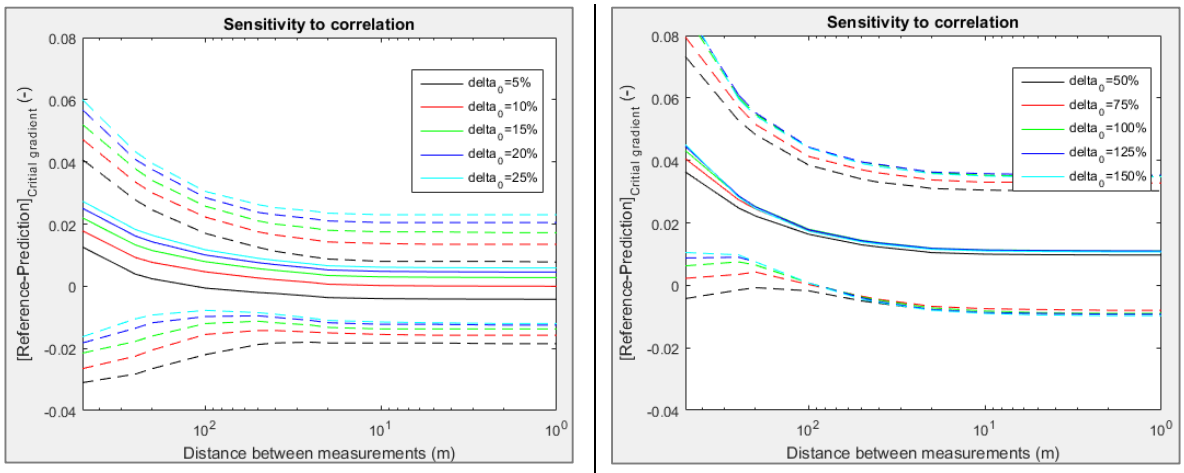


Figure 86 – Sensitivity of model to correlation lengths. Standard configuration, no measurement error, $\boldsymbol{\delta_0(d70)}, \boldsymbol{\delta_0(k)}$ are varying between 5% and 150% of the standard input values (d70=180, k=600).

Measurement error considered

The result of the analysis for the simulations with noise is presented in Figure 87.

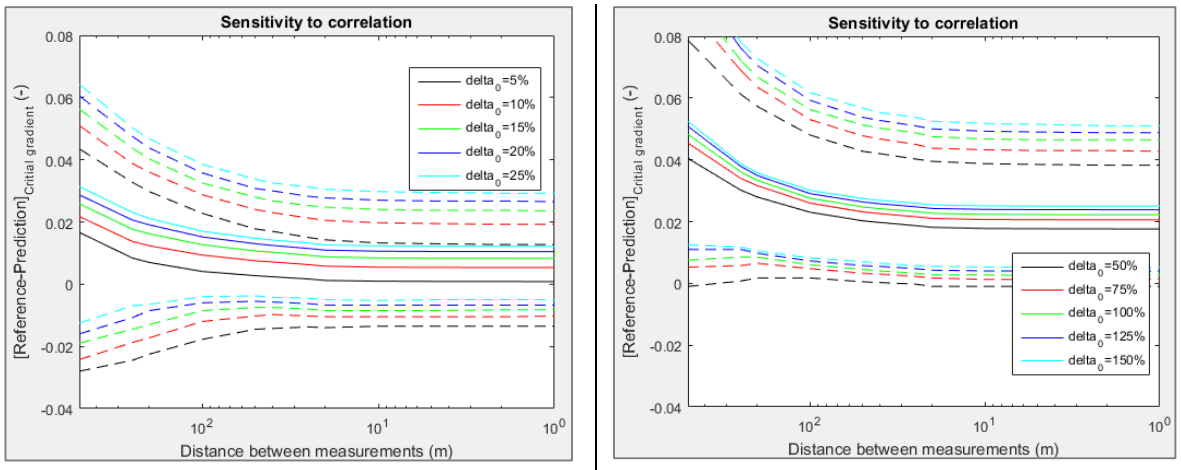


Figure 87 – Sensitivity of model to correlation lengths. Standard configuration with measurement error, $\boldsymbol{\delta_0(d70)}, \boldsymbol{\delta_0(k)}$ are varying between 5% and 150% of the standard input values (d70=180, k=600).

Also in the combined analysis more or less equal effects as in the previous figures can be observed. The model outcome is less sensitive close to the standard configuration than for correlation lengths much smaller than the standard configuration. Furthermore, it should be noticed that the magnitude of the 95% confidence interval becomes slightly smaller with decreasing correlation lengths. This mostly account for smaller measurement intervals. Associated with this is the shape of the lower 95% boundary. For small correlation lengths the advantage of extra measurements (smaller measurement intervals) is present for all measurement strategies. While for high correlation lengths extra measurements result in equal or even decreased precision (see low measurement intervals at the right plot of Figure 86 for example).

16. Influence of data distributions (range)

The sensitivity is tested for 50%, 75%, 100%, 125% and 150% of the standard deviation in the standard configuration. Mention that the magnitude of measurement error increases along with increasing parameter variance. This is because the value of noise input variance is 0.5 of the data variance input value.

Input mean of d70

Table 20 – Measured d70 output with respect to varying d70 mean parameter input. Simulations based on standard input configuration and standard configuration with measurement error.

d70 mean	**No noise**					**Noised ($\phi_{input} = 0.5 * \sigma_{input}$)**				
	50%	75%	100%	125%	150%	50%	75%	100%	125%	150%
$\mu_{input,d70}\ [10^{-4}\ m]$	1,00	1,50	2,00	2,50	3,00	1,00	1,50	2,00	2,50	3,00
$\mu_N\ [10^{-4}\ m]$	1,02	1,50	2,00	2,50	3,00	1,02	1,50	2,00	2,50	3,00
$\sigma_N\ [10^{-5}\ m]$	2,71	2,97	3,00	2,99	3,01	3,01	3,33	3,35	3,36	3,36
$\mu(\sigma_n)_N\ [10^{-5}\ m]$	2,18	2,38	2,39	2,39	2,40	2,57	2,85	2,86	2,86	2,87

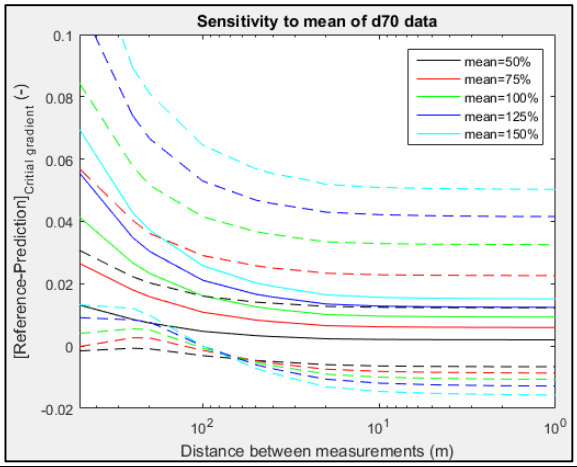


Figure 88 - Sensitivity of model to mean of d70 data. Standard input configuration, measurement error neglected, $\mu(d70)$ is varying between 50% and 150% of the standard input value (2e-4).

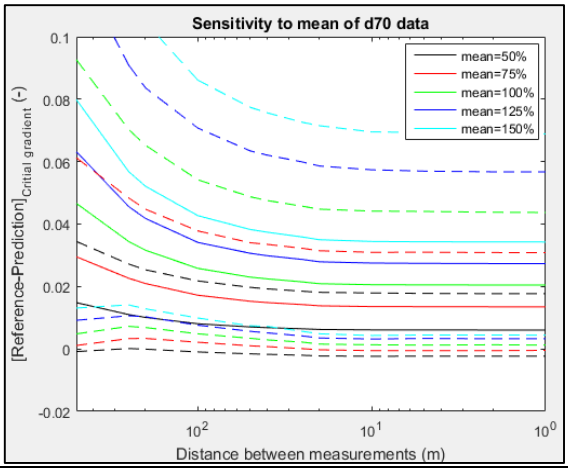


Figure 89 - Sensitivity of model to variance in d70 data. Standard input configuration with measurement error, $\mu\ (d70)$ is varying between 50% and 150% of the standard input value (2e-5).

Input variance of d70

Table 21 – Measured d70 output with respect to varying d70 variance parameter input. Simulations based on standard input configuration and standard configuration with measurement error.

d70 variance	**No noise**					**Noised ($\phi_{input} = 0.5 * \sigma_{input}$)**				
	50%	75%	100%	125%	150%	50%	75%	100%	125%	150%
$\sigma_{input,d70}\ [10^{-5}\ m]$	1.50	2.25	3.00	3.75	4.50	1.50	2.25	3.00	3.75	4.50
$\phi_{input,d70}\ [10^{-5}\ m]$	0	0	0	0	0	0.75	1.13	1.50	1.88	2.25
$\mu_N\ [10^{-4}\ m]$	2,00	2,00	2,00	2,00	2,00	2,00	2,00	2,00	2,00	2,00
$\sigma_N\ [10^{-5}\ m]$	1,50	2,25	3,00	3,75	4,48	1,68	2,51	3,36	4,19	5,00
$\mu(\sigma_n)_N\ [10^{-5}\ m]$	1,20	1,80	2,40	2,99	3,57	1,44	2,15	2,86	3,57	4,26

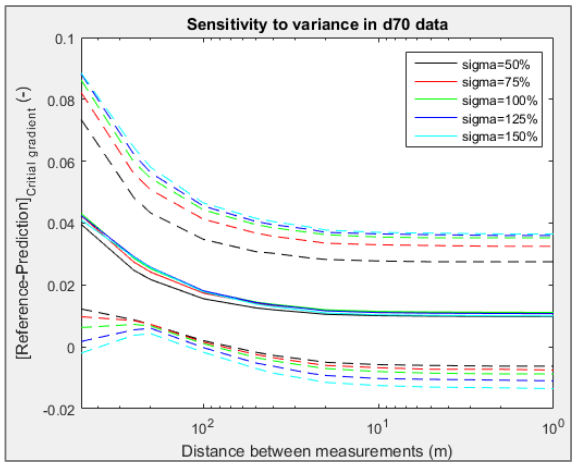


Figure 90 - Sensitivity of model to variance in d70 data. Standard input configuration, measurement error neglected, $\sigma(d70)$ is varying between 50% and 150% of the standard input value (3e-5).

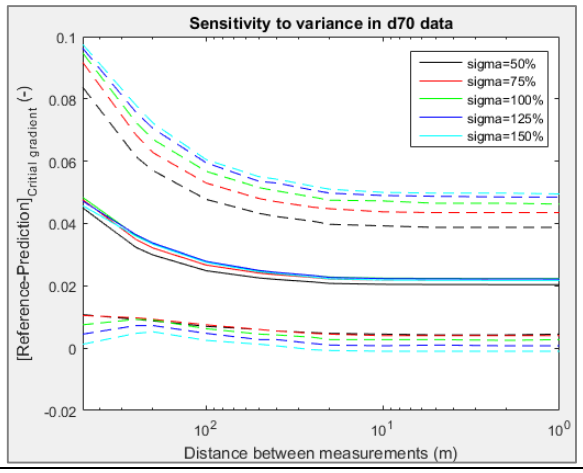


Figure 91 - Sensitivity of model to variance in d70 data. Standard input configuration with measurement error, $\sigma(d70)$ is varying between 50% and 150% of the standard input value (3e-5).

Input variance of k (and mean correspondingly)

Table 22 - Measured k output with respect to varying k variance parameter input. Simulations based on standard input configuration and standard configuration with measurement error.

k variance	No noise					Noised ($\phi_{input} = 0.5 * \sigma_{input}$)				
	50%	75%	100%	125%	150%	50%	75%	100%	125%	150%
$\sigma_{input,k}$ $[10^{-4}\ m]$	0,43	0,64	0,85	1.07	1.28	0,43	0,64	0,85	1.07	1.28
$\phi_{input,k}$ $[10^{-4}\ m]$	0	0	0	0	0	0,22	0,32	0,43	0,54	0,64
μ_N $[10^{-4}\ m]$	1,10	1,23	1,44	1,76	2,26	1,12	1,29	1,58	2,03	2,77
σ_N $[10^{-4}\ m]$	0,49	0,88	1,51	2,56	4,33	0,56	1,06	1,94	3,58	6,80
$\mu(\sigma_n)_N$ $[10^{-4}\ m]$	0,22	0,37	0,57	0,88	1,32	0,34	0,61	1,01	1,69	2,85
σ_N/μ	0,45	0,72	1,05	1,45	1,92	0,50	0,82	1,23	1,76	2,45
$\mu(\sigma_n)_N/\mu$	0,20	0,30	0,40	0,50	0,58	0,30	0,47	0,64	0,83	1,03

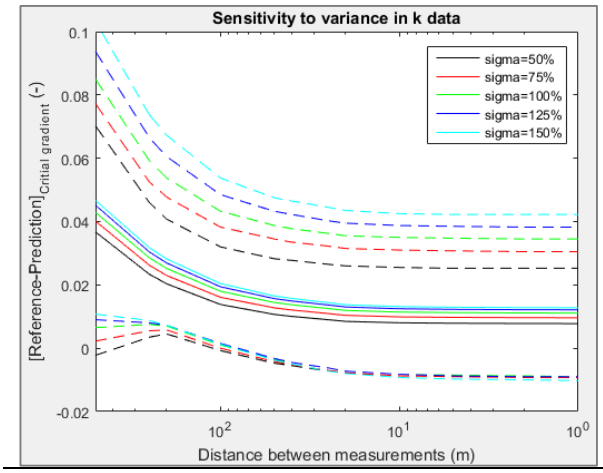


Figure 92 - Sensitivity of model to variance in k data. Standard input configuration, no measurement error, $\sigma(k)$ is varying between 50% and 150% of the standard input value (8.5e-5).

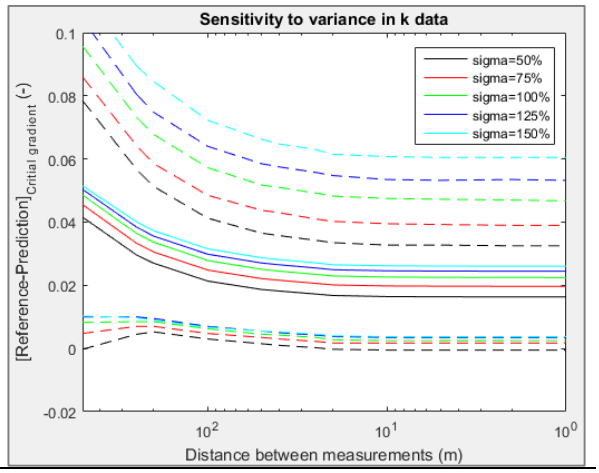


Figure 93 - Sensitivity of model to variance in k data. Standard input configuration with measurement error, $\sigma(k)$ is varying between 50% and 150% of the standard input value (8.5e-5).

Input mean of k (and variance correspondingly)

Table 23 - Measured k output with respect to varying k mean parameter input. Simulations based on standard input configuration and standard configuration with measurement error.

k variance	No noise					Noised ($\phi_{input} = 0.5 * \sigma_{input}$)				
	50%	75%	100%	125%	150%	50%	75%	100%	125%	150%
$\mu_{input,k}$ $[10^{-4}\ m]$	-0,42	-0,21	0	0,21	0,42	-0,42	-0,21	0	0,21	0,42
$\sigma_{input,k}$ $[10^{-4}\ m]$	0,85	0,85	0,85	0,85	0,85	0,85	0,85	0,85	0,85	0,85
$\mu_{N,k}$ $[10^{-4}\ m]$	0,94	1,16	1,44	1,78	2,20	1,03	1,27	1,58	1,95	2,41
$\sigma_{N,k}$ $[10^{-4}\ m]$	0,97	1,19	1,51	1,84	2,27	1,25	1,53	1,94	2,37	2,93
$\mu(\sigma_n)_{N,k}$ $[10^{-4}\ m]$	0,37	0,46	0,57	0,70	0,87	0,66	0,81	1,01	1,24	1,55

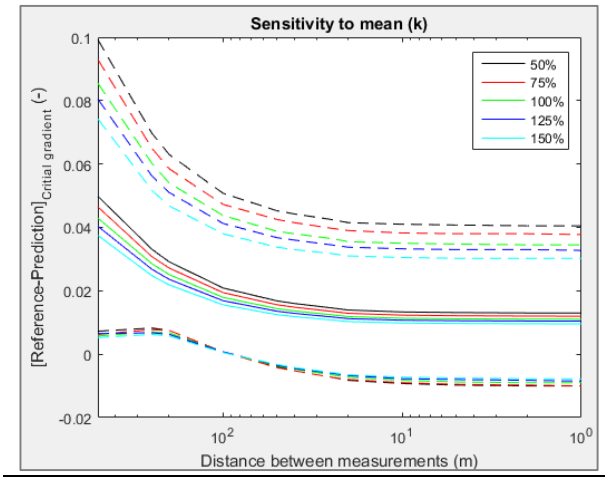


Figure 94 - Sensitivity of model to mean in k data. Standard input configuration, no measurement error, $\sigma(k)$ is varying between -0,42 and 0,42 of the standard input value (0).

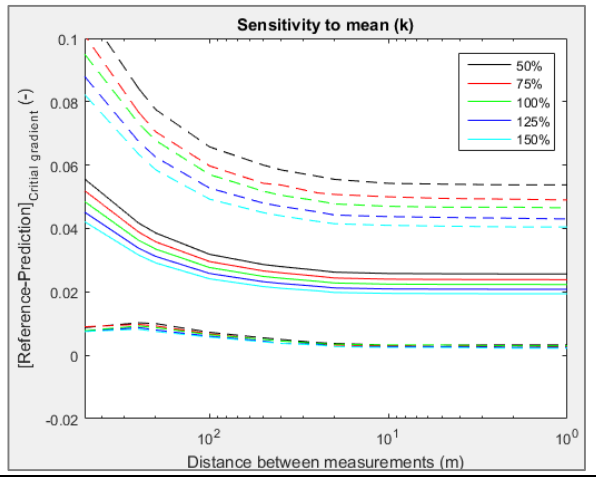


Figure 95 - Sensitivity of model to mean in k data. Standard input configuration with measurement error, $\sigma(k)$ is varying between -0,42 and 0,42 of the standard input value (0).

Combined: varying input variance of d70 and k

The sensitivity is tested for 50%, 75%, 100%, 125% and 150% of the variance in the standard configuration.

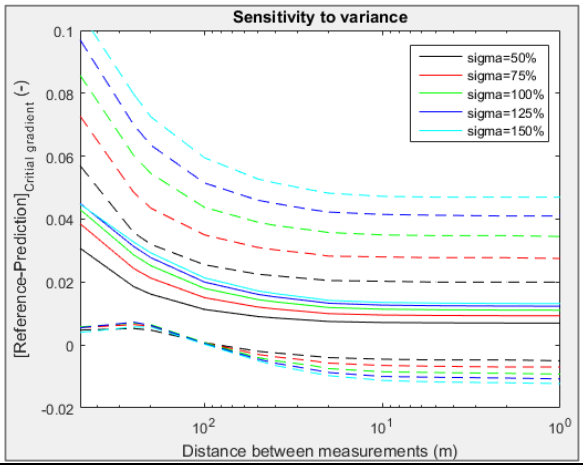


Figure 96 – Sensitivity of model to variance in data. Standard configuration, no measurement error, $\sigma(d70), \sigma(k), \sigma(D)$ are varying between 50% and 150% of the standard input values (d70=3e-5, k=8.5e-5, D=2).

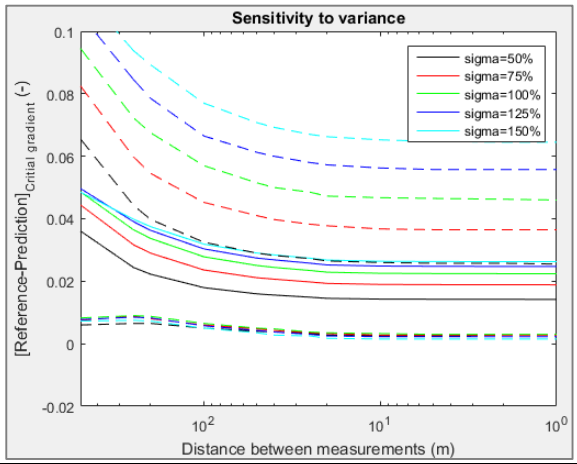


Figure 97 – Sensitivity of model to variance in data. Standard configuration with measurement error, $\sigma(d70), \sigma(k), \sigma(D)$ are varying between 50% and 150% of the standard input values (d70=3e-5, k=8.5e-5, D=2).

17. Influence of 2D d70 modelling to distribution of reference and assessed strength

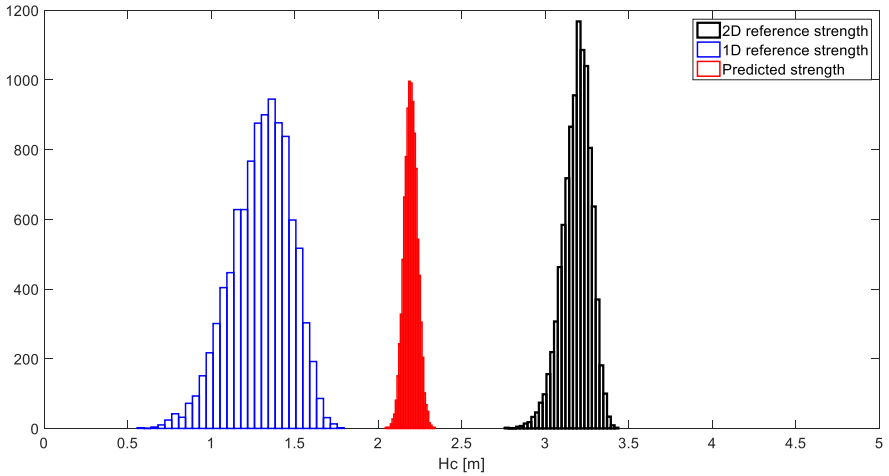


Figure 98 – Influence of parallel effects: Only d70 varying, correlation length of 1m.

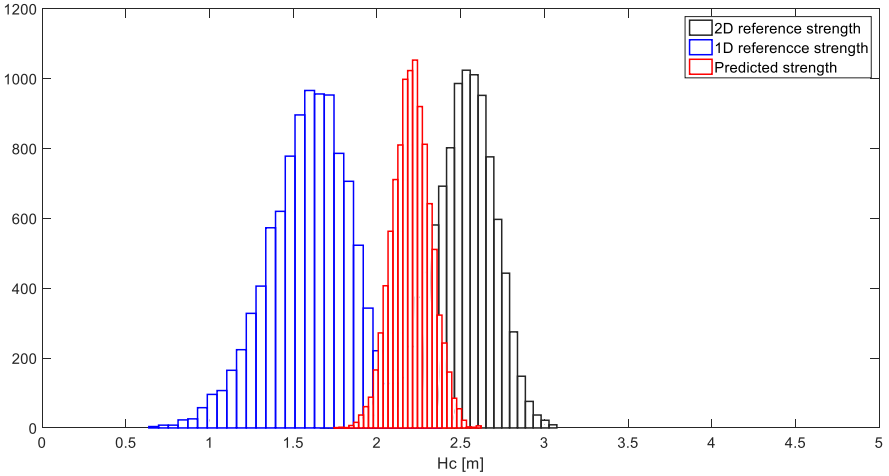


Figure 99 – Influence of parallel effects: Only d70 varying, correlation length of 10m.

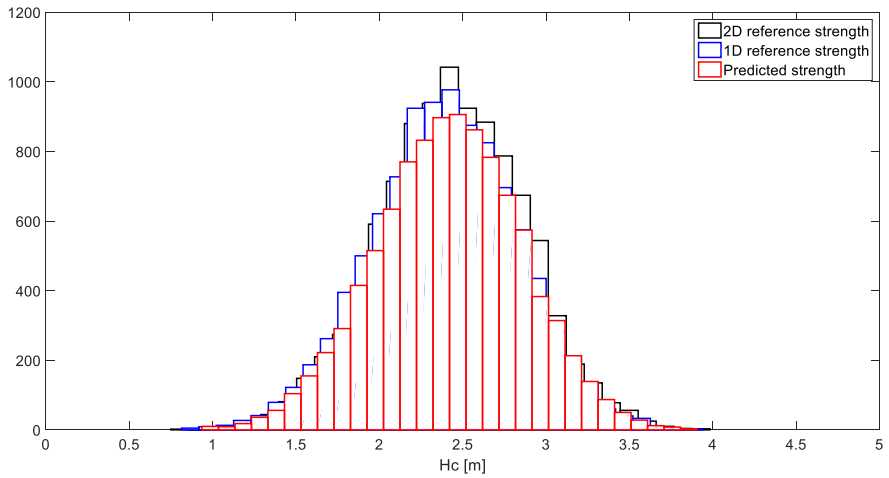


Figure 100 – Influence of parallel effects: Only d70 varying, correlation length of 180m.

18. Influence of 2D modelling of d70 to the bandwidth of errors

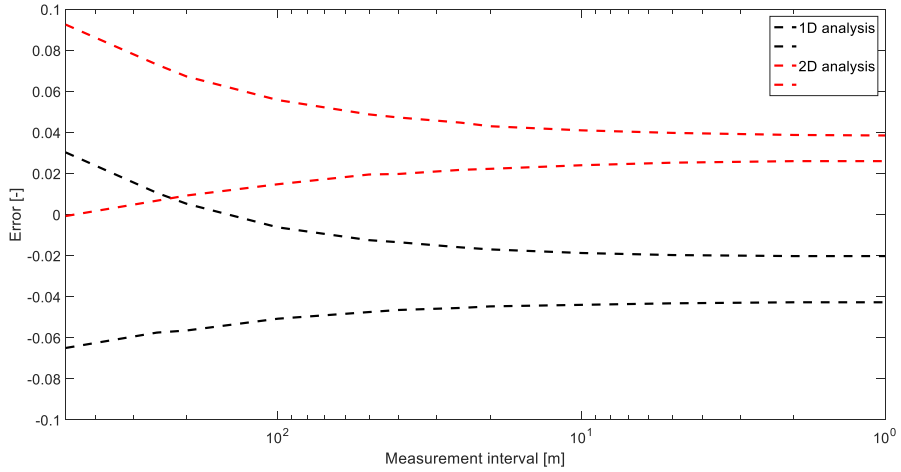


Figure 101 – Influence of parallel effects: Only d70 varying, correlation length of 1m.

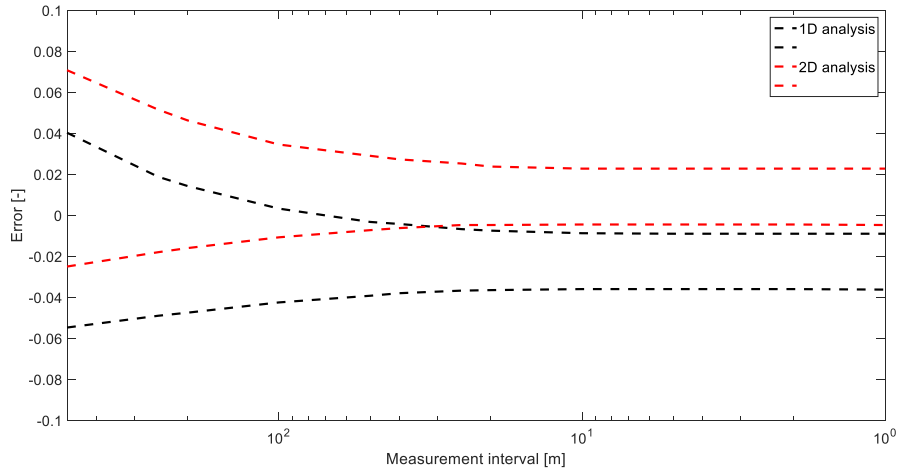


Figure 102 – Influence of parallel effects: Only d70 varying, correlation length of 10m.

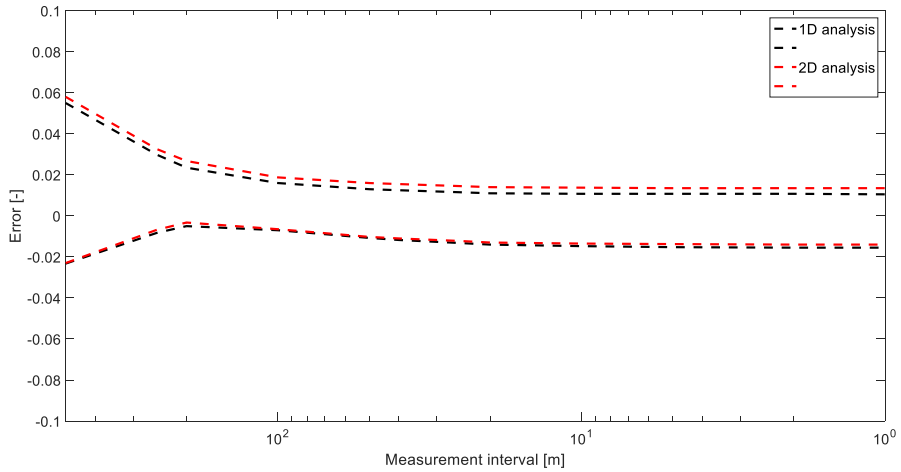


Figure 103 – Influence of parallel effects: Only d70 varying, correlation length of 180m.

19. Deviating measurement numbers for d70 and k.

In the standard model set-up, the number of measurements is equal for d70, k and D in each measurement configuration. However, in practice it might be well possible that the number of measurements per parameter deviate. Furthermore, it is interesting to know whatever the influence of more measurements to individual parameters is. Therefore 19*19 measurement configurations are tested in which the number of d70 and k measurements deviate from each other. The result of this analysis is presented in Figure 104 and Figure 105 for respectively the standard input configuration and standard input configuration with measurement error. For convenience D is not considered here. As the influence of D already appeared minor this will hardly affect the results.

For this analysis the magnitude of 95% confidence boundary is used as a measure of accuracy. From the figures it can be seen that only increasing either the number of d70 or k measurements without increasing the number of measurements of the other parameter has less or no effect. For example, using 3 d70 measurements and 20 k measurements results in the same accuracy as only 3 d70 and k measurements. In case of measurement error extra k measurements have some effect as possible errors are better averaged out. Vice versa the effect is more or less the same. An increasing number of d70 measurements has way more effect as the number of k measurements 5 or 6 instead of 3. Striking from this analysis is the inaccuracy that stays with higher number of measurements. For example, the accuracy that is reached with 21 d70 and 21 k measurements is, in terms of 95% confidence interval, more or less equal 10 d70 and 10 k measurements (for the case without measurements error even less measurements are needed to obtain the same accuracy). Investigation of higher measurements numbers reveals that the accuracy will not increase significantly.

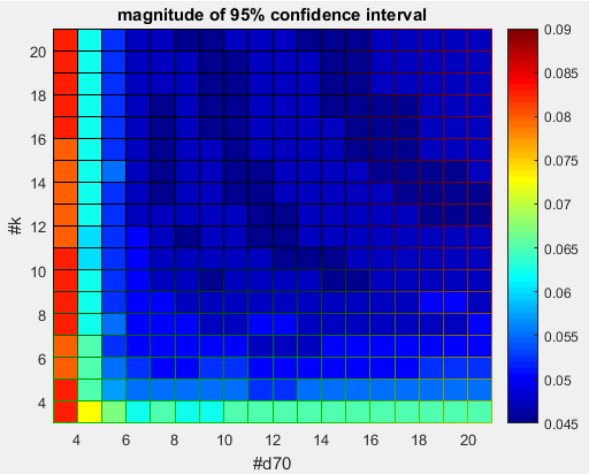


Figure 104 – effect of certain combination of d70 and k measurements on magnitude of 95% certainty interval. Simulated field: N=5000, standard configuration, no measurement error

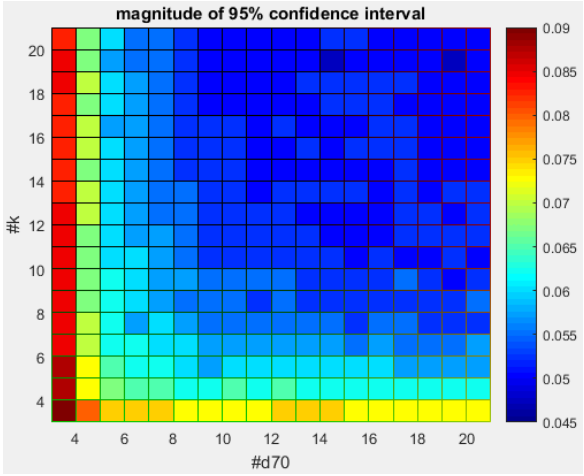


Figure 105 – effect of certain combination of d70 and k measurements on magnitude of 95% certainty interval. Simulated field: N=5000, standard configuration with measurement error

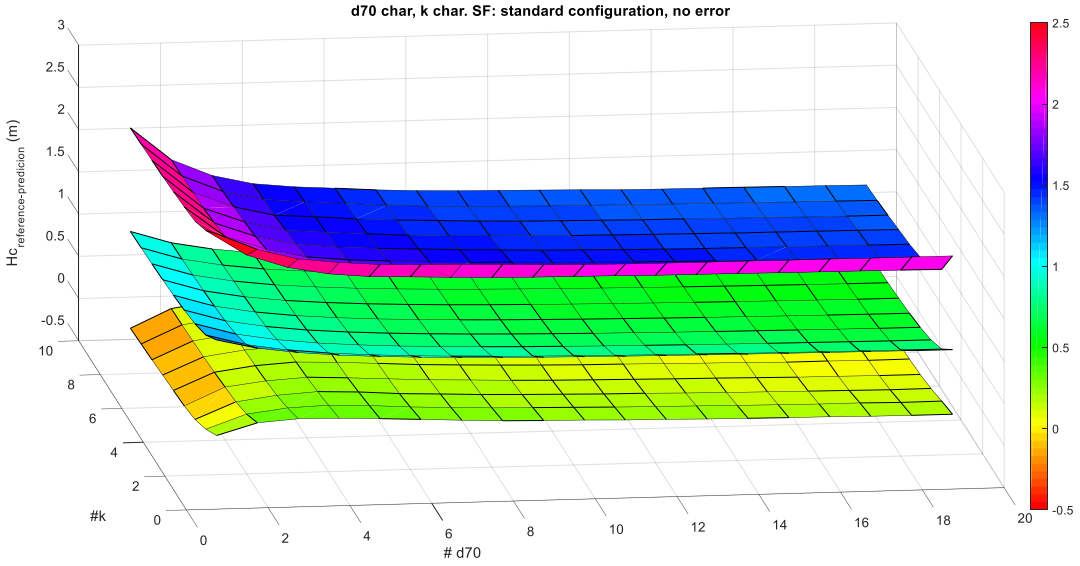


Figure 106 - 3D plot of mean error and 95% bandwidth of errors for deviating combinations of measurement numbers. Reference scenario, no error

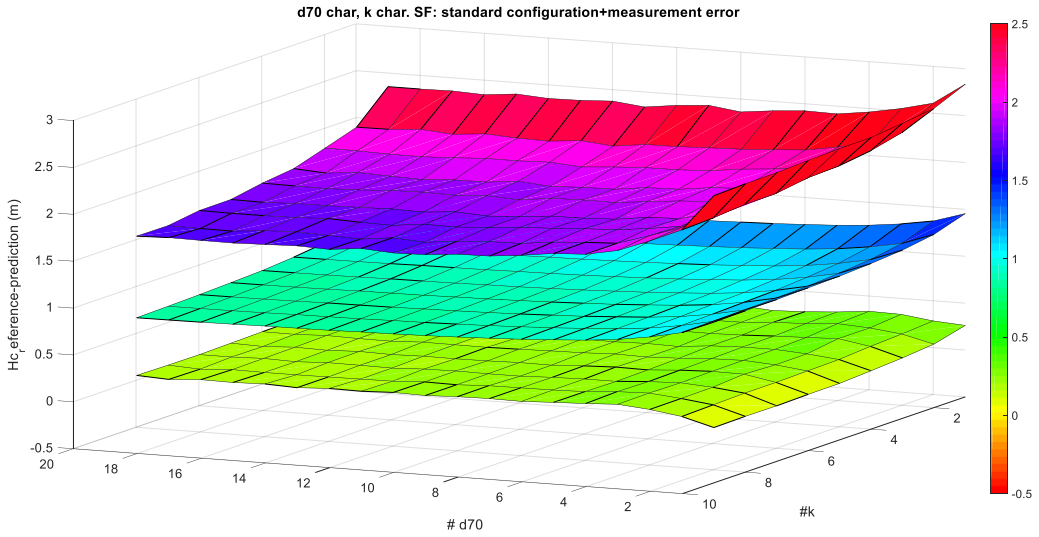


Figure 107 - 3D plot of mean error and 95% bandwidth of errors for deviating combinations of measurement numbers. Reference scenario, including measurement error