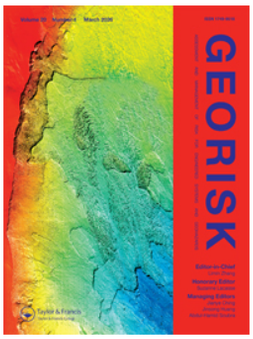


Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards

ISSN: 1749-9518 (Print) 1749-9526 (Online) Journal homepage: www.tandfonline.com/journals/ngrk20

The role of load variations in assessing credible dike failure probabilities: balancing load and strength uncertainties

Bart Strijker, Sebastiaan N. Jonkman & Matthijs Kok

To cite this article: Bart Strijker, Sebastiaan N. Jonkman & Matthijs Kok (09 Apr 2026): The role of load variations in assessing credible dike failure probabilities: balancing load and strength uncertainties, Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards, DOI: 10.1080/17499518.2026.2654177

To link to this article: https://doi.org/10.1080/17499518.2026.2654177

GEORISK
https://doi.org/10.1080/17499518.2026.2654177

The role of load variations in assessing credible dike failure probabilities: balancing load and strength uncertainties

Bart Strijker[a,b], Sebastiaan N. Jonkman[a] and Matthijs Kok[a,b]

[a]Department of Hydraulic Engineering, Delft University of Technology, Delft, The Netherlands; [b]Department of Risk and disaster management, HKV Lijn in Water, Lelystad, The Netherlands

ABSTRACT

Assessing dike safety is of key interest to societies in low-lying areas, but results can be implausible when, for example, they contradict the observed performance of the dike. To improve credibility, load monitoring data can be incorporated using reliability updating techniques. This paper investigated the role of load variations in reliability updating and assessing credible failure probabilities. It was found that the impact of reliability updating increases when load variations are small, as a large contribution to failure probabilities comes from relatively frequent load levels, of which the conditional failure probabilities are reduced most through reliability updating. Moreover, a credibility check was introduced for dikes that have been stable for decades, where load levels with return periods of up to 10 years are not expected to contribute more than 50% to the failure probability, indicating an imbalance between load variation and strength uncertainty. This imbalance occurs when the inverse gradient of the fragility curve exceeds 1.5 times the decimate height of the load. Many Dutch dikes, including canal dikes and dikes along the large lakes and delta regions, have small decimate heights. For these dikes, strength uncertainties must be sufficiently small to obtain credible failure probability estimates.

ARTICLE HISTORY
Received 25 June 2025
Accepted 29 March 2026

KEYWORDS
Dike safety; reliability updating; credible probabilities; load variations; monitoring; Bayesian approach

1. Introduction

To provide flood protection, transportation, energy and freshwater, society depends on various earthen embankments. The safety of these structures often requires periodic assessment (European Commission 2024; CEN 2004). In the Netherlands, over 10,000 kilometres of dikes provide flood protection, including both regional and primary dikes (see Figure 1). These are earthen embankments that retain water. While primary dikes protect the hinterland from floods caused by major water bodies such as the sea, lakes, and large rivers, regional dikes are located along smaller inland water bodies. In the event of a dike breach, both types of floods can lead to severe damage, while fatalities are only expected at primary dike breaches. Therefore, estimating credible dike failure probabilities is a key question for Dutch society in managing flood risk. Credible implies that the failure probability is realistic and justifiable (Phoon 2023), derived from a proper and well-informed engineering assessment using appropriate models based on available data, and not contradicting the (historical) behaviour or performance of the flood defence. Simpson, Pappin, and Croft (1981) suggested using so-called worst credible parameter values in calculations, which have a probability of being exceeded about 0.1%, as also discussed in Phoon (2016). These values reflect loads and material properties that the designer could still realistically believe might occur (Simpson, Pappin, and Croft 1981). In flood safety, however, relevant load levels can be more extreme than those with a 0.1% annual exceedance probability (Jonkman, Jongejan, and Maaskant 2011; Kind 2014). Consequently, the interpretation of what is credible, or worst credible, depends on the geotechnical structure and its observed and expected performance. The authors prefer using the term credible when linking probabilities to observed performance. For example, suppose a dike has an estimated failure probability of 1/100 per year but withstood a load level with an exceedance probability of 1/500 per year, then that failure probability estimate is not credible. Ultimately, rare events do happen, which are unlikely, but not necessarily non-credible. The pursuit of

CONTACT Bart Strijker b.strijker@tudelft.nl Department of Hydraulic Engineering, Delft University of Technology, Building 23, Stevinweg 1, Delft, 2628CN, The Netherlands

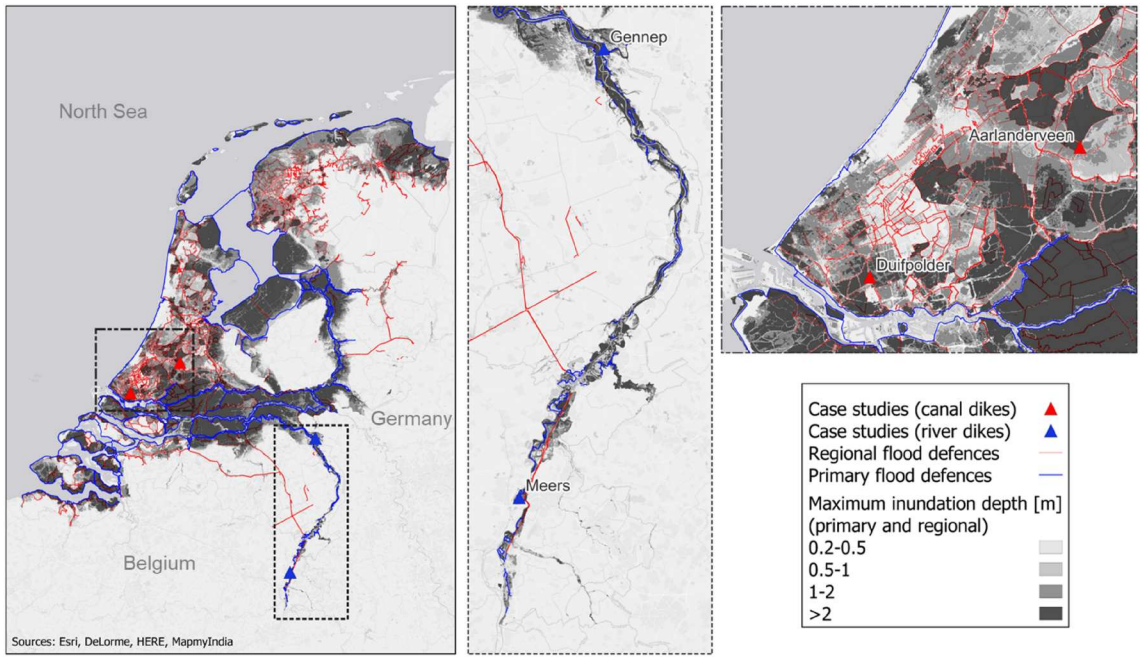


Figure 1. Maps of the Netherlands that show the maximum inundation depths (combined from many flood scenarios from primary and regional dike breaches) and the dike systems for both primary (blue) and regional (red) flood defences, along with the highlighted case study sites that are marked (▲) and labelled.

credible failure probabilities prevents dikes from being unnecessarily and unjustifiably estimated as unsafe, and supports making investment decisions based on a well-founded safety profile.

One of the failure mechanisms contributing to the overall failure probability of dikes in the Netherlands is inner-slope instability, referred to as dike instability (Jongejan and Maaskant 2015). This failure mechanism occurs when high water pressures inside the dike reduce the shear strength and induce loss of stability. It accounts for the largest number of primary dike trajectories failing to meet safety standards (Nationaal Georegister 2024), while historically it accounted for only 5% of the failures (Van Baars and Van Kempen 2009). The same holds for regional dikes, where hundreds of kilometres are considered unsafe on a national scale, which contradicts observations (Rikkert 2022). Many of the Dutch dikes are centuries old and have often been reinforced and upgraded over the years, resulting in heterogeneous dikes with varying soil compositions (Woerkom et al. 2022). This heterogeneity, combined with limited data on the subsurface buildup and the uncertain shear strength of the soil, makes the overall strength of dikes uncertain. This uncertainty complicates the assessment of whether a dike meets safety standards and rises the question: are dikes truly unsafe, or is it challenging to demonstrate adequate safety with the available data and models?

Advances in computational methods have enhanced the understanding of dike failure mechanisms and reliability assessment. Advanced numerical approaches, such as Material Point Method simulations, have demonstrated the capability to model large-deformation behaviour and dike failure (Zheng et al. 2024; Zheng et al. 2025). Furthermore, probabilistic frameworks have improved reliability assessment for embankment slopes through various approaches: Monte Carlo methods with random field modelling (Hicks and Li 2018; Li et al. 2016; Liu et al. 2020) and surrogate modelling techniques for transient seepage analysis (Wang et al. 2020). While these numerical and probabilistic methods provide valuable insights into failure mechanisms and uncertainty quantification, challenges remain in incorporating observed performance information to reduce uncertainties and refine reliability estimates. In general, assumptions and choices in failure probability modelling tend to be conservative, or safe-sided (European Commission 2024). This results in so-called hidden safety elements in safety assessments, leading to smaller observed failure rates compared to estimations made using models. This gap between observations and model outcomes can be reduced by understanding parameter sensitivity in reliability analyses (Yang et al. 2024) and the use of Bayesian inference (Huang, Liu, and Liu 2025; Yang et al. 2025) to reduce model parameter uncertainty and update the predicted reliability. The specific application of Bayesian inference is often

called reliability updating, which uses Bayesian principles to refine reliability predictions by integrating prior knowledge with new knowledge. This can be measurements, monitoring outcomes, or observations of system performance (Straub 2011). It is a well-established and frequently used technique in structural engineering to update the failure probabilities of structures (De Koker et al. 2020; Hemel et al. 2024; Straub and Papaioannou 2015). While less common for dikes, it is still applied to various failure mechanisms, like overflow, piping and slope stability (Huang, Liu, and Liu 2025; Rikkert et al. 2022; Schweckendiek et al. 2017; Schweckendiek and Vrouwenvelder 2013; Schweckendiek, Vrouwenvelder, and Calle 2014; van der Krogt et al. 2022). Piping and slope stability are failure mechanisms that often involve large epistemic strength uncertainties due to uncertain model schematizations and subsurface conditions. These uncertainties are virtually time-invariant, and as shown by Schweckendiek, Vrouwenvelder, and Calle (2014), these uncertainties can be especially reduced using reliability updating. While the role of time-invariant epistemic strength uncertainties in reliability updating is well understood, the role of time-varying inherent load uncertainties, or simpler load variations, is poorly understood.

The central question addressed in this study is: What is the role of load variations in reliability updating and assessing credible failure probabilities?

The outline of this study starts with the materials and methodology in Section 2. This section describes the characteristics of typical Dutch dikes, the methods used to calculate the probability of dike instability, along with the applied reliability updating techniques and the schematisation of the case studies. Section 3 discusses the results of the conceptual analysis and the case studies. Section 4 explores the implications of these findings for failure probability analysis and outlines the study's limitations. Finally, Section 5 concludes with final remarks.

2. Materials and methods

2.1. Some typical Dutch dikes and their key characteristics

The Netherlands has a variety of dikes that protect the low-lying land, like sea, river and lake dikes, as well as compartmentalising dikes and dikes along the canals (see Figure 1). All these dikes have different characteristics, like dimensions, subsoil conditions and boundary conditions. The case studies in this study considered canal and river dikes, of which the typical differences in dike characteristics are summarised in Figure 2 and are discussed in more detail.

The canal dikes in the Netherlands are located along drainage canals in polders in the northwestern part of the Netherlands. Water levels in these canals can be several metres above the polder levels and are maintained near target levels using pumping stations and weirs. In extreme situations, canal water levels typically rise several decimetres above the target levels. In contrast, the Dutch rivers exhibit highly dynamic water levels, often fluctuating by several metres between average annual conditions and extreme floods (Klijn, Asselman, and Mosselman 2019). This difference in water level fluctuations is a key distinction between canal and river dikes and impacts the dynamics of pore-water pressures, which are crucial for dike stability (Ridley, McGinnity,

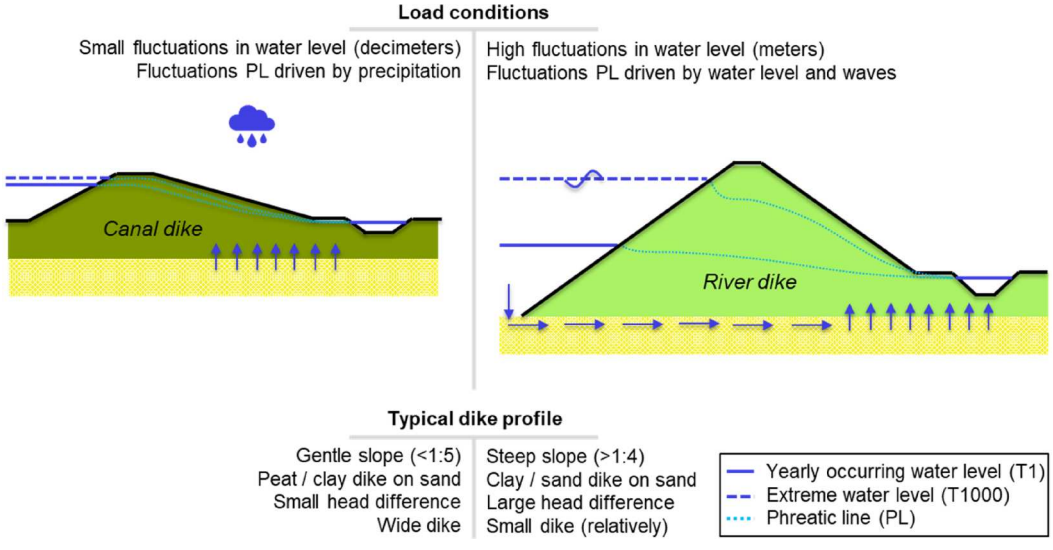


Figure 2. Schematic representation of a typical river dike and canal dike, illustrating characteristic loading conditions and dike profiles.

and Vaughan 2004; Van der Meij 2023). Pore-water pressures in the dike body are commonly expressed by the phreatic line. This phreatic line can be influenced by both the intrusion of outside water and precipitation on top of the dike. Which driver plays the biggest role varies with the dike's specific characteristics. At canal dikes, where water level fluctuations are small, precipitation is expected to be the primary factor influencing the phreatic line, as supported by previous studies (Rikkert 2022; Strijker et al. 2024). In contrast, river water levels can fluctuate significantly and for extended periods, often lasting weeks, making them the main driver of the phreatic line's position within river dikes.

Other relevant differences include the subsurface material and geometry. The canal dikes are often heightened with locally available soils, mainly a mixture of clayey and peaty material. The dikes are, in general, rather wide with gentle slopes and small head differences. The canal dike bodies are situated on Holocene deposits ranging in thickness from a few metres to over 20 metres, overlying a Pleistocene aquifer. The river dikes have a similar basis of Holocene and Pleistocene layers, but the dike bodies are typically composed of sand and clay. Although both dikes are located on the Pleistocene layers, the pore-water pressure dynamics are different. Unlike canal dikes, the underlying aquifers of river dikes are often directly connected to the outside water level. As a result, high water levels can cause a significant increase in pore-water pressure beneath the dike, which impacts dike stability. The fluctuations in pore-water pressures in the aquifer underlying canal dikes are limited and driven by seasonal variations in precipitation and evapotranspiration (Wong, Batjes, and de Jager 2007). The response is slow and damped by the overlying Holocene clay and peat layers.

2.2. Failure probability analyses and fragility curves

2.2.1. Uncertainties: epistemic and inherent randomness

The probability of dike instability results from uncertainties in both load and strength. These uncertainties can arise from inherent randomness or a lack of knowledge, referred to as inherent and epistemic uncertainties, respectively. While inherent uncertainty cannot be practically reduced through additional analysis or data collection, such as fluctuating water levels and pore water pressures, epistemic uncertainties can. Examples of epistemic uncertainties are geotechnical parameters, such as the shear strength. This uncertainty arises primarily from the limited number of soil samples tested at a given dike section, and the uncertainty in estimating this geotechnical parameter reflects spatial variability and measurement error (Phoon and Kulhawy 1999). In practice, soil samples from various locations are combined to create a regional dataset, after which factors like variance reduction from averaging spatial variability, statistical uncertainty, and measurement uncertainty are considered (Calle, Kanning, and Schweckendiek 2021; Van Der Krogt, Schweckendiek, and Kok 2018; Vanmarcke 1977). Estimates of these geotechnical parameters remain uncertain and are incorporated into the reliability analyses. These uncertainties are time-invariant and, in practice, deterministic within a year. Although soil properties can naturally vary over time due to geological processes, this is not a random process, but rather a gradual and patterned evolution. When assessing the failure probability within a specific year, the temporal variation of geotechnical parameters is expected to be small relative to the present uncertainty in estimating them from available data, and the uncertainty in soil parameters is mainly epistemic. This means that the dike strength in safety assessments is often assumed to be time-invariant and epistemically uncertain, while most loads are typically classified as time-varying and inherently uncertain (Schweckendiek et al. 2017).

2.2.2. Fragility curves

A commonly used approach to estimate the dike failure probability is the use of fragility curves (Bachmann et al. 2013; Casciati, Faravelli, and Venini 1991; Schultz et al. 2010; Schweckendiek, Vrouwenvelder, and Calle 2014). This curve describes the dike's strength by expressing the relationship between load levels and failure probabilities conditional on these loads. To calculate the failure probability of a dike, the fragility curve needs to be combined with the probability density function of the load events:

$$P(F) = \int_{S=-\infty}^{S=\infty} f_s(s)\, F_R(s)\, ds \tag{1}$$

where $f_s(s)$ is the probability density function of load levels and $F_R(s)$ is the cumulative density function of the resistance R, that gives the probability of failure conditional on the load S, also known as the fragility curve. When analysing the role of loads, it is more practical to focus on the exceedance probability, $1\text{-}F_S(s)$, rather than the probability density function. This is because extreme load events are the primary concern, and the exceedance probability effectively highlights the likelihood of these occurrences. These different components are shown graphically in Figure 3, where an example of a fragility curve (top panel), load statistics (middle panel) and resulting density

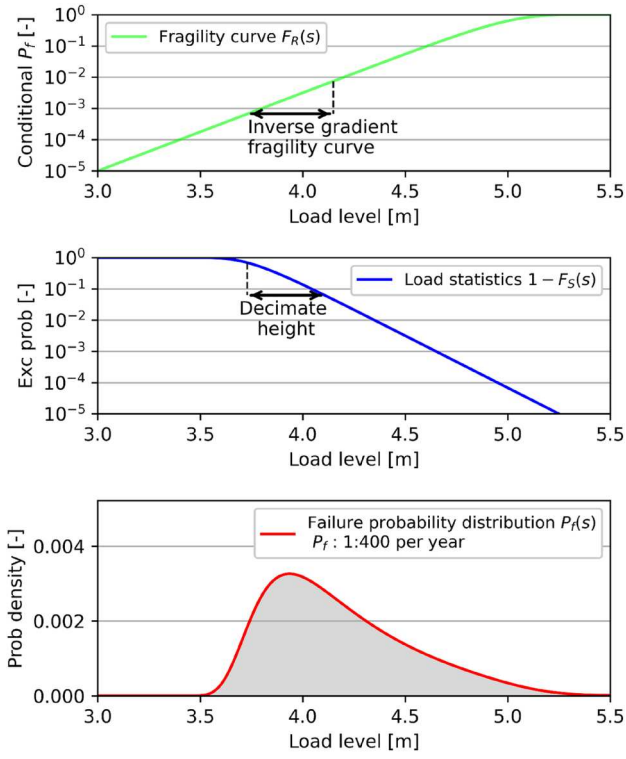


Figure 3. Graphical representation of the dike failure probability: the fragility curve (top), load statistics (middle), and the resulting failure probability distribution (bottom). The integrated failure probability density distribution gives a failure probability of 1:400 per year in this example.

function of the failure probability are shown (bottom panel). This density function is the product of the fragility curve and the probability density function of the load. It provides insight into the contribution of various load levels to the overall probability of failure, as well as the load event where failure is most likely to occur, which is at the peak of this distribution. The integral of this distribution gives the failure probability, which in this example is 1/400 per year.

The shape of the fragility curve reflects the amount of strength uncertainty. This depends not only on the stochastic input variables but also on the failure mechanism itself. In general, the curve approaches a step function for very well-understood or brittle systems, like wave-overtopping, and a more S-shaped curve corresponds to more poorly understood or elastic systems, like slope instability (Van der Meer, Ter Horst, and Van Velzen 2009). A measure to quantify this shape is the inverse gradient, which represents the increase in load level that results in a tenfold rise in the conditional failure probability, as illustrated in the upper panel of Figure 3. A higher inverse gradient corresponds to a gentler curve, indicating larger strength uncertainties. The inverse gradient can be taken at any part of the fragility curve, but in this study, the inverse gradient is considered at the load level scenario where failure is most likely to occur, so at the peak of the failure probability distribution.

A measure for the load variations is the decimate height. This is defined as the increase in load level that occurs when the exceedance probability of the load level experiences a tenfold decrease (Schweckendiek 2014; Wojciechowska 2015). Smaller decimate heights correspond to smaller load variations. This measure is illustrated in the middle panel of Figure 3.

2.2.3. Reliability updating using Bayesian analysis

Traditionally, engineers deal with uncertainties in failure probability analyses by starting with initial estimates based on available information and refining these estimates through additional field observations. This process can be made more structured by using probability theory, and specifically Bayesian probability updating, to revise prior judgments. This method can be used to update failure probability estimates based on observations, such as the survival of observed load conditions without failure indications. When survived load levels are used to update failure probabilities, Bayesian updating is applied by eliminating implausible values of uncertain strength parameters, assuming the uncertainty is epistemic. In general, this can be done in two ways that are mathematically equivalent and differ only in terms of implementation: the direct and indirect reliability updating techniques (Schweckendiek 2011; Schweckendiek 2014). Both techniques use the Bayes' Rule (Bayes 1763) that forms the basis for updating probabilities with new evidence:

$$P(F|\varepsilon) = \frac{P(\varepsilon|F)P(F)}{P(\varepsilon)} \tag{2}$$

where F is the failure event to be predicted and ε the observed event or evidence. In this study, the evidence is survival at an observed load level s_{obs}, where survival implies that the dike strength R is larger than the observed load: $R > s_{obs}$ (inequality type of information).

$$P(F|\varepsilon) = \frac{P(F \cap \varepsilon)}{P(\varepsilon)} = \frac{P(F,\, R > s_{obs})}{P(R > s_{obs}))} \tag{3}$$

where R denotes the random variable representing dike's resistance or strength, expressed in the same units as the load variable S (e.g. phreatic head or water level).

The indirect method uses Bayes' rule to update the PDFs of the underlying random variables describing the strength, while the direct method exploits the definition of the conditional probability of failure F given the evidence ε to update the failure probability directly. In this study, the direct method using fragility curves is preferred because it is simpler to implement

and aligns with the proof-of-concept focus, rather than providing detailed updates to stochastic variables. This method directly calculates the updated fragility curve and probability of failure, but does not update the underlying stochastic variables, which is beyond the scope of this study. Direct reliability updating can be applied using fragility curves and survival information, which leads to the following expression for the updated failure probability:

$$P(F|\varepsilon) = \frac{\int_{s_{obs}}^{\infty} P(s_{obs} < R < S) f_s(s) ds}{1 - F_R(s_{obs})} = \frac{\int_{s_{obs}}^{\infty} (F_R(s) - F_R(s_{obs})) f_s(s) ds}{1 - F_R(s_{obs})} \quad (4)$$

Since the load density function $f_S(s)$ multiplies the entire integral, the expression can be written more compactly as:

$$P(F|\varepsilon) = \frac{\int_{s_{obs}}^{\infty} (F_R(s) - F_R(s_{obs}))}{1 - F_R(s_{obs})} f_s(s) ds \quad (5)$$

From this, the posterior or updated fragility curve can be written as:

$$F_{R,upt}(s) = P(F|\varepsilon) / f_s(s) ds = \frac{\int_{s_{obs}}^{\infty} (F_R(s) - F_R(s_{obs}))}{1 - F_R(s_{obs})} \quad (6)$$

$$F_{R,upt}(s) = \begin{cases} \frac{F_R(S) - F_R(s_{obs})}{1 - F_R(s_{obs})}, & s \geq s_{obs} \\ 0, & s < s_{obs} \end{cases} \quad (7)$$

Figure 4 shows an example of the effect of reliability updating on the fragility curve and the probability density function of the failure probability.

Note that updating only has an effect if past and current epistemic strength uncertainties are fully correlated and time-invariant. This study assumes time-invariant uncertainties, a reasonable choice for many geotechnical properties (Schweckendiek 2014). It should be noted, however, that in practice the observed survived load level s_{obs} may itself be uncertain due to measurement errors or spatial variability in pore-water pressures. Such uncertainties effectively reduce the information content of the survival evidence, which in turn diminishes the impact of the updating. In this study, a best-estimate s_{obs} was used. This assumption, and its resulting limitations, is important to keep in mind for practical applications.

2.3. Conceptual analysis

To examine how load variations influence reliability updating, a conceptual analysis was performed on hypothetical dikes using a range of combinations of load variations and fragility curves. Both were described with two-parameter Gumbel distributions. Their spreads were varied while the prior failure probability

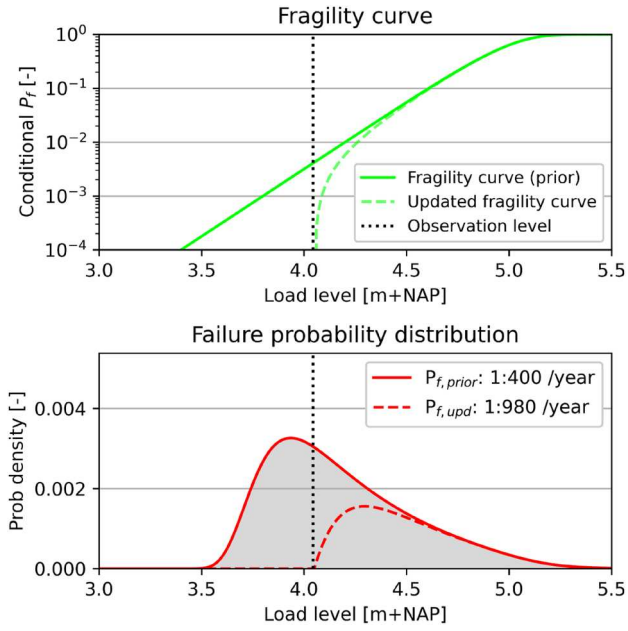


Figure 4. Example of prior and updated fragility curve (top), based on an observed and survived load level, and the resulting failure probability distribution (bottom) with the prior and updated failure probabilities. The load statistics are unchanged and correspond to those shown in Figure 3.

was fixed at 1/100 per year by adjusting the relative distance between the two distributions. In practice, fragility curves and load statistics for dikes are often derived from probabilistic analyses using numerical integration and do not necessarily follow a probability density function. As a result, the decimate height and inverse gradient of the fragility curve can be estimated by taking the tangent of the curve. For this conceptual analysis, these measures follow directly from the Gumbel distributions. The cumulative distribution function of the Gumbel distribution is:

$$F(x) = \exp\left(-\exp\left(-\frac{x-\mu}{\beta}\right)\right)$$

where μ is the location parameter and β the scale parameter. The decimate height and inverse gradient follow from the scale parameter by multiplying β with $ln(10)$.

For the conceptual analyses, the load variations were varied between decimate heights of 0.05 and 0.65 m, and the strength uncertainty between inverse gradients of the fragility curve from 0.05–1.55 m. For each case, reliability updating was applied using a survived load level that occurs on average once every 10 years.

2.4. Modelling dike stability

The dike's probability of instability is generally assessed using two-dimensional limit equilibrium

methods, such as Bishop, Spencer or Uplift-Van, which calculates the factor of safety against instability considering driving forces and resisting forces acting on a slip plane (Bishop 1955; Sharp et al. 2013). These methods, in combination with search algorithms to find the slip plane with the minimum factor of safety, are implemented in D-stability, which was used in this study (Van der Meij 2023). Next to calculating the factor of safety, this software can also calculate failure probabilities using Adaptive Importance Sampling or First Order Reliability Method (FORM) analysis. While importance sampling includes a search algorithm, FORM analysis was carried out with a fixed slip plane, which was found using a calculation with design values (Van der Meij 2023). Importance sampling may not converge for very small failure probabilities. In those cases, a FORM analysis was used as an alternative approach. The resulting failure probability is always calculated for a single load scenario, representing only one point on a fragility curve. The complete fragility curve can be constructed by performing multiple calculations using a range of different load scenarios with various headlines.

2.5. Case study description

To demonstrate the role of load variations in reliability updating and assessing credible failure probabilities, four realistic case studies are examined in this section. These case studies include two canal dikes and two river dikes. This section describes the schematisation of the dikes, including geometry, subsurface materials, soil properties, load variations and the survived load levels.

2.5.1. Dike geometry and geotechnical properties

The two canal dikes are located in the western Netherlands and are named after the polder that they protect: Aarlanderveen and Duifpolder. The two river dikes are located in the southern part of the Netherlands and their names were retrieved from the nearby towns they protect; Gennep and Meers. The geographic locations of the dikes are shown in Figure 1. The canal dikes consist more of peat and organic clay, while the river dikes are made up of clay and sand, as can be seen in the cross-sectional profiles in Figure 5. All dikes are located on top of Pleistocene sand layers, which are located very deep, over 15 metres below the surface, at the Duifpolder.

The soil parameters for all layers are listed in Table 1. Shear strengths were modelled using Mohr-Coulomb failure criterion in terms of effective stress parameters, based on a regional dataset of over 5,000 shear strength tests for typical Dutch geological deposits (triaxial for clay, direct simple shear for peat) and over 13,000 classification tests (STOWA 2023). The effective stresses are calculated in D-Stability based on soil weights and the pore-water pressures. The way the pore-water pressures are schematised is described in Section 2.5.2. Strength parameters were treated as stochastic variables, while volumetric weights were considered deterministic using mean values. When the variance of the strength parameters in this dataset is derived, it includes the effects of spatial variability, statistical uncertainty, and measurement uncertainty (Calle, Kanning, and Schweckendiek 2021; Phoon and Kulhawy 1999). The strength uncertainties applied in dike-stability reliability analysis should reflect the shear strength uncertainty of a layer. Because relevant slip planes span both weaker

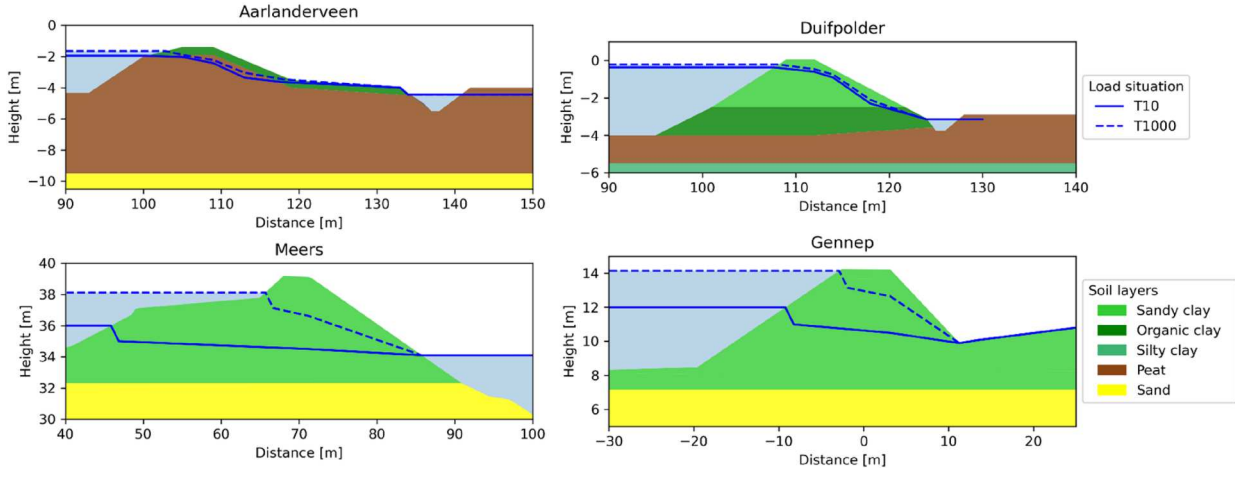


Figure 5. Cross-sectional profiles of the considered case studies showing the soil types and schematised phreatic lines for different probabilities of occurrences. Only two phreatic lines are shown here, whereas in the analyses, more load situations were considered, including pore-water pressures in the aquifer.

Table 1. Deterministic and stochastic soil parameters in the reference case. The presented standard deviations are based on the full dataset, whereas these values are reduced in the stability analyses to account for averaging spatial variability.

		Effective Cohesion (Lognormal)		Effective Friction Angle (Lognormal)	
	Vol. Weight (Deterministic) γ_{sat}/ γ_{unsat} [kN/m^3]	Mean [kPa]	Std. dev. [kPa] (COV)	Mean [°]	Std. dev. [°] (COV)
Silty clay	15.7 / 15.7	2.7	0.5 (19%)	29.9	0.8 (3%)
Organic clay	15.2 / 15.2	3.4	1.0 (29%)	26.5	1.1 (4%)
Peat	10.2 / 10.2	5.0	0.3 (7%)	13.7	0.3 (2%)
Sandy clay	16.2 / 16.2	3.7	0.6 (16%)	28.2	0.7 (2%)
Sand	20.0 / 18.0	-	-	35.0	1.8 (5%)

and stronger zones, the dataset's variance is reduced to account for spatial averaging effects. However, since the dataset consists of tests within a larger region, not all variance can be reduced. Variance reduction is done by a simplification of the basic random field model, where it is assumed that 25% of the variance in the dataset is due to regional variations of the mean, while 75% of the variance is due to variations around the local mean that spatially average out (Calle 1996; Calle, Kanning, and Schweckendiek 2021). This results in standard deviations of shear strength parameters that are approximately half those found in the dataset.

Two additional scenarios were included to assess the impact of strength uncertainty: a moderate scenario (10% variation in cohesion, 5% in friction angle) and a high strength uncertainty scenario (20% variation in cohesion, 10% in friction angle). The high strength uncertainty was based on the NEN 9997-1, and also suggested by the JRC report on the reliability background of the Eurocodes (Vrouwenvelder et al. 2024). These coefficients of variation represent the uncertainty in spatial average ground strength parameters. In both scenarios, the deterministic soil properties remained unchanged.

Note that this schematisation can differ in various aspects from those used in actual safety assessments conducted by the water authorities. For example, a less detailed soil schematisation and different soil strength parameters are used. The case studies aim to provide insights using various schematizations of dikes, rather than estimating a single best failure probability.

2.5.2. Load variations and pore-water pressures

All stability analyses employ effective stress principles, where effective stresses are calculated from total stresses and pore-water pressures schematised for each load scenario. The fragility curves are constructed conditional on load scenarios, which comprise a combination of (1) the phreatic line, including the water level acting against the dike and head levels in the dike body, and (2) the pore-water pressures in the underlying aquifer. Pore-water pressures are assumed hydrostatic up to the first cover layer above the Pleistocene sand, and linearly interpolated from there to the top of the aquifer. There are no roads on the dikes, and external traffic loads are not considered in the analysis. These load components have two-dimensional aspects. These are represented by a parameter for the height of the phreatic line, which serves as the load variable for the conditional failure probabilities. Therefore, the height of the phreatic line is defined as the average head in the dike body and is expressed in m + NAP, the Dutch vertical reference system. Appendix A.1 contains statistics on water levels and head at the inner crest used for the load scenarios. The underlying assumptions of these scenarios are discussed here.

For the load scenarios at canal dikes, it is assumed that high canal water levels and high head levels are strongly correlated and occur simultaneously, as both result from heavy rainfall (Lendering, Schweckendiek, and Kok 2018; Strijker and Kok 2023). Water level statistics were derived using log-linear interpolation between target levels and normative water levels with a 100-year return period, received from the water authorities. The decimate heights are 15 cm at Aarlanderveen and 7.5 cm at Duifpolder. The statistics of head levels from the inner crest and further down the dike profile were based on measurements and time series models (Bakker and Schaars 2019; Strijker and Kok 2025) and are shown in the right-hand graphs of Figure 6, with decimate heights under 10 cm. Heads between the canal and the inner crest are linearly interpolated.

For the load scenarios at the river dikes, the phreatic line and pore-water pressures under the dike were schematised using the official flood risk assessment toolkit for primary flood defences (Slomp et al. 2016; TAW 2004). This toolkit includes the Waternet Creator, which models water pressures based on parameters, such as the maximum water level, dike geometry, leakage lengths and subsurface conditions (Van der Meij 2023). For the river dikes, the water level statistics were based on the probabilistic tool Hydra-NL, which provides water level statistics across the Netherlands (Geerse 2011). Subsurface conditions differ between the two river dikes. The Meers dike consists of a dike body of clay overlying a sand layer, while the Gennep dike has a clay body with an underground of clay and peat layers on top of a sand layer. As a result, pore-water pressure cannot build up in the sandy aquifer at

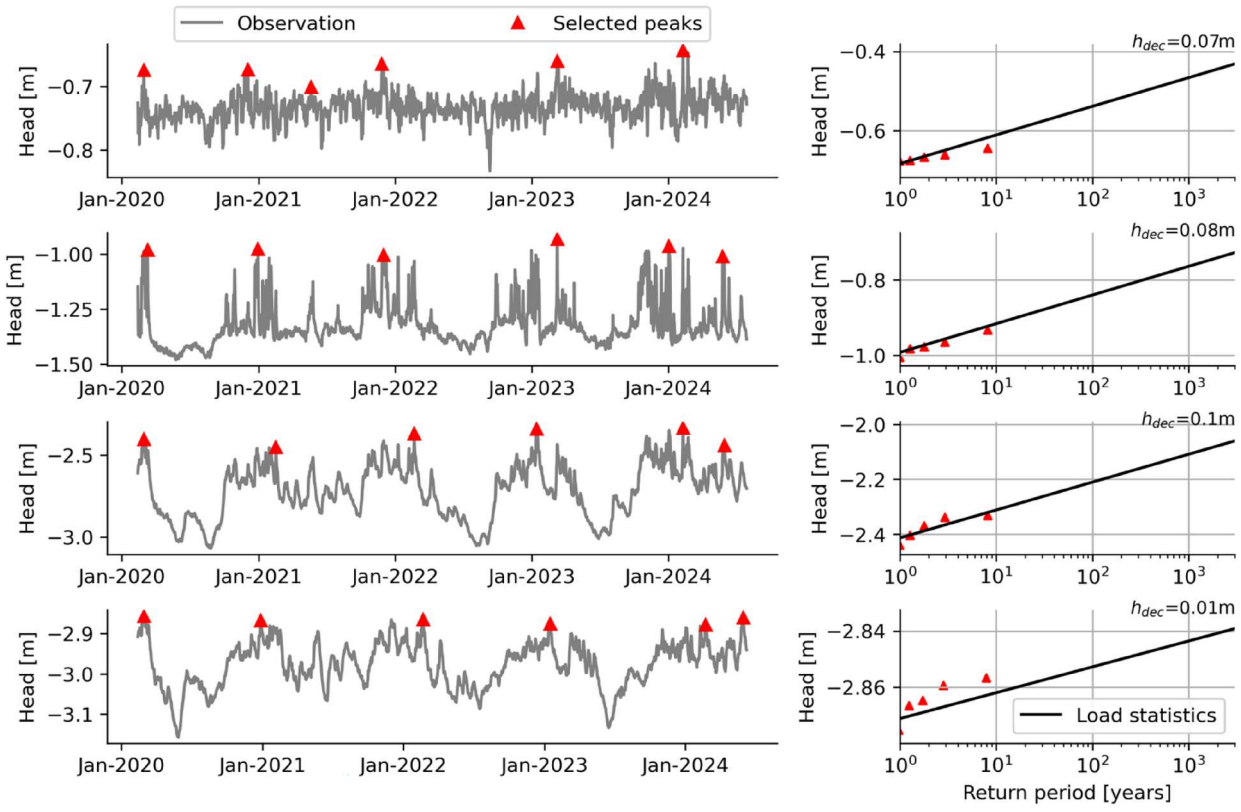


Figure 6. Time series of the observed head level at Duifpolder with selected peaks (▲) at the inner crest (top), talud zone (middle) and toe (bottom). Right: plot positions of the selected peaks with extrapolation lines. Red triangles in the right and left panels mark the same peaks, but axis scaling differs.

Meers, but can at Gennep. The leakage length of the sandy aquifer at Meers, expressing the distance over which the head decays to about 37%, is about 180 m. This results in high pore water pressures at the toe of the dike. These pore-water pressures can cause uplift (Van, Koelewijn, and Barends 2005), where the hydraulic head in the sand layer is maximised by the weight of the uplifted overlying layer.

For all cases, the fragility curves were set up using at least five load scenarios, ranging from yearly occurring levels to levels with a probability of 1/10,000 per year. Adjacent scenarios differ by at most a factor of ten in probability. The fragility curves were constructed using log-linear interpolation of the conditional failure probabilities between the load scenarios. Nonlinearity checks were performed, and additional scenarios were added where needed. The overall failure probability was then obtained by combining the conditional failure probabilities with the load distribution and integrating over all load levels.

2.5.3. Survived load levels in reliability updating

Extreme events during monitoring offer valuable data and much deeper insight into dike reliability than annual occurring load levels. For instance, during the 2021 Meuse flood, water levels occurred with an exceedance probability smaller than 1:100 per year (Strijker et al. 2023). However, this study focuses on how failure probability changes across dike types and loading variations. To maintain consistency and eliminate random variability, two load levels were selected for each case with fixed annual probabilities of 1/2 and 1/10 per year, respectively called T2 and T10 events. These represent typical yearly occurring and more extreme events, helping to demonstrate the added value of observing both.

3. Results

To answer the research question, the results of two analyses are discussed. First, the extent to which failure probabilities can be reduced using reliability updating for different load and strength configurations is conceptually assessed. Secondly, reliability updating is applied to the case studies involving four dikes in the Netherlands. This section finishes with a reflection on balancing strength uncertainties to load variation for assessing credible failure probabilities.

3.1. Conceptual analysis on reliability updating

To understand the role of load variations in reliability updating, a conceptual analysis was performed using hypothetical dikes with a range of combinations between load variations and strength uncertainty. The impact of reliability updating is expressed using the probability factor, defined as the ratio of prior to updated failure probability ($P_{f,prior}$ / $P_{f,upd}$). It was found that the largest reduction in failure probability occurs when the strength uncertainty is high and the load variations are low (top right corner in Figure 7). Even in cases of moderate strength uncertainties, reliability updating can lead to significantly smaller failure probabilities when the decimate height is small. This is because, under small load variations, the range of loads contributing to the failure probability is narrow and concentrated around relatively frequent load levels. This is illustrated using two cases with identical fragility curves and prior failure probabilities, but different load variations, indicated with red and blue dots in Figure 7. For the two cases shown in Figure 8, the prior failure probability is 1/100 per year, while the updated probabilities are 1/960 per year (red, small load variations) and 1/140 per year (blue, large load variations). This results in probability factors of 9.6 and 1.4, respectively. In both cases, reliability updating causes a sharp drop in the conditional failure probability near the survived load level (top row in Figure 8). When the load variations are small, this range contributes much more to the failure probability than in cases with larger load variations.

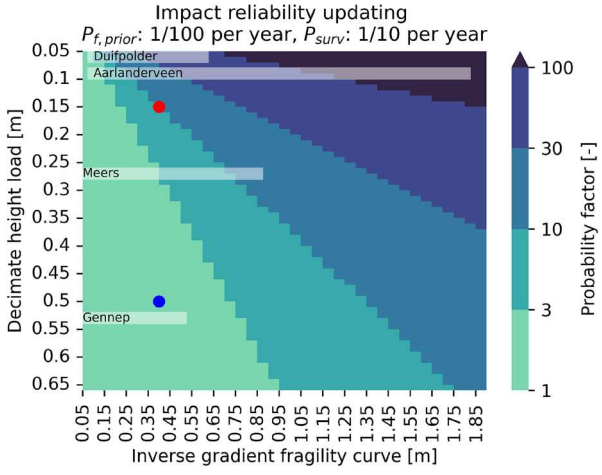


Figure 7. Impact of reliability updating across different load variations (decimate height) and strength uncertainties (inverse gradient fragility curve). Transparent blocks indicate the ranges of load variations and strength uncertainties of the case studies considered, and red and blue dots mark two cases of which the fragility curves, load statistics and resulting failure probability distribution are shown in Figure 8.

As a result, reliability updating has a larger impact on the failure probability under small load variations.

To assess when reliability updating is most effective, sensitivity analyses were performed by varying survived load levels and prior failure probabilities (see Appendix A.2 and A.3). The key finding is that the impact of reliability updating is independent of the prior failure probability. Regardless of whether the prior failure probability is high or low, the impact of updating depends primarily on the shape of the failure probability distribution, which is influenced by load variations and strength uncertainties, rather than by the absolute level or relative distance between them.

3.2. Case study results

The insights from the conceptual analysis were verified using four realistic case studies of canal and river dikes with varying strength uncertainties and load variations. Their position relative to the conceptual parameter space is illustrated in Figure 7 by the transparent blocks. The calculated probability factors for the case studies are compared with the so-called conceptual factors, defines as the expected ratio $P_{f,prior}/P_{f,upt}$ for given combinations of decimate height and fragility-curve inverse gradient, based on Gumbel distributions. The results are discussed, beginning with the canal dikes, followed by the river dikes.

3.2.1. Canal dikes

Based on the conceptual analysis, it is expected that the failure probabilities of the canal dikes can be significantly reduced through reliability updating, since the load variations are small with decimate heights of the phreatic line up to 10 cm. For each dike, three different strength uncertainty scenarios were considered. The resulting prior and updated fragility curves (T2 and T10) for Aarlanderveen are shown in Figure 9. As strength uncertainty increases, fragility curves become flatter with higher inverse gradients ranging from 0.08 m to 1.79 m, as indicated in the titles. In the middle panel, the inverse gradient for the moderate-strength case is taken at the load level where failure is most likely. This point lies at a phreatic line of roughly −0.75 m and is marked by a grey solid circle. After updating, the fragility curve is capped at the survived load level and gradually converges back to the prior curve. This offset is largest under high strength uncertainty, leading to the greatest reduction in failure probability. The updated fragility curve depends only on strength uncertainty and the survived load level, while the final impact on failure probability also depends on the load statistics.

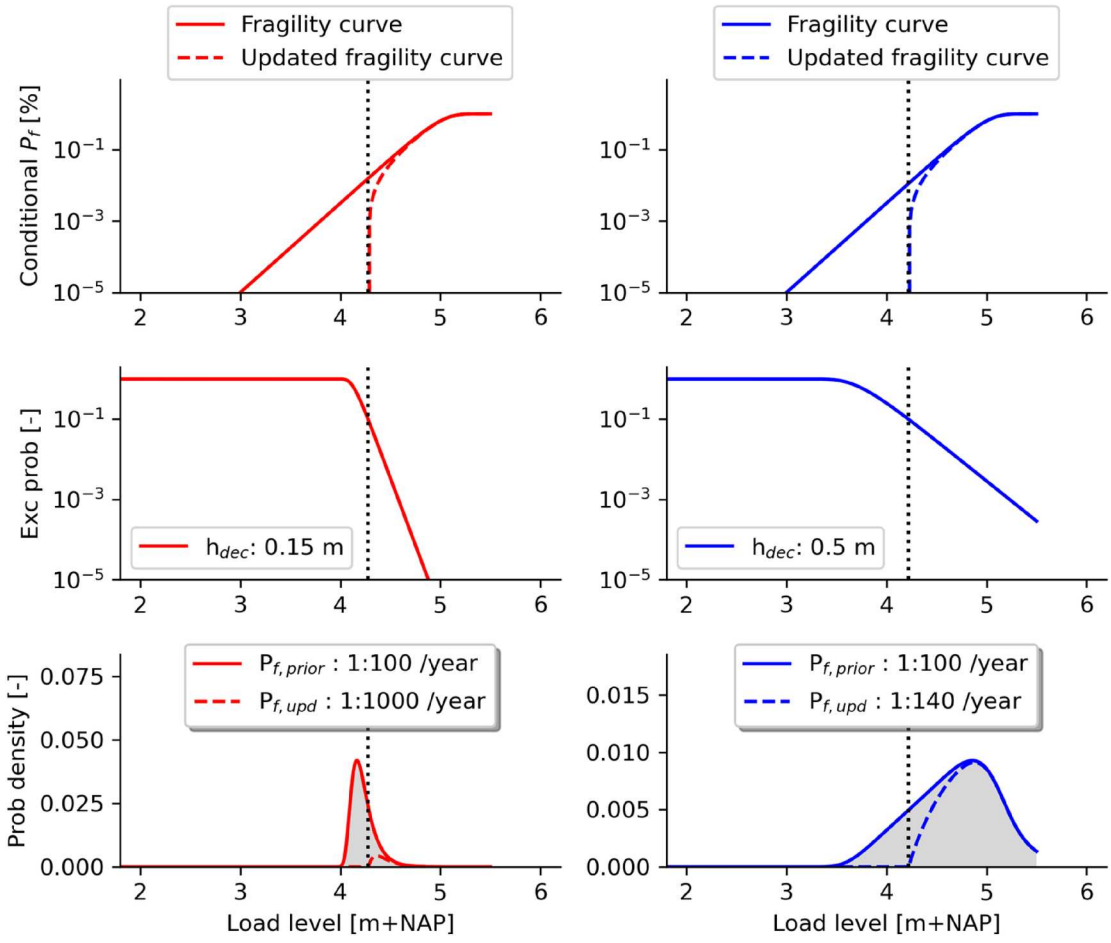


Figure 8. Two combinations of load statistics (red: small load variations, blue: large load variations) and the same prior fragility curves (inverse gradient of 0.4 m). The vertical dotted line indicates the survived 10-year load level. Both cases have the same prior failure probability, while the updated probabilities differ.

Given the small load variation of the canal dikes, how much are the failure probabilities reduced?

For the moderate strength uncertainty scenario and a T10 survival load, reliability updating reduces failure probability by a factor of 10.0 at Duifpolder and 42.4 at Aarlanderveen. Although Duifpolder has a smaller decimate height, its fragility curve is steeper (lower inverse gradient), which limits the impact of updating

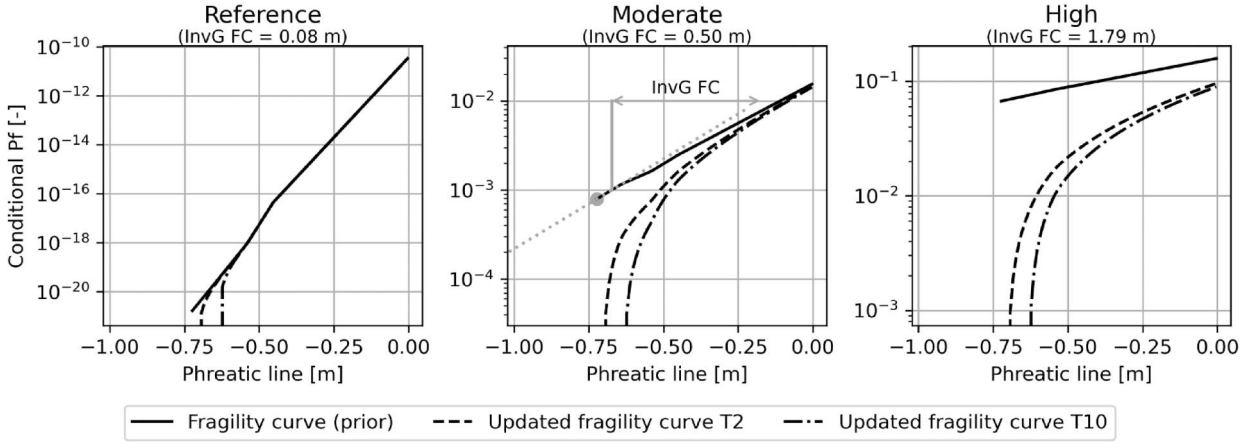


Figure 9. Fragility curves at Aarlanderveen for three different strength uncertainty scenarios (reference, moderate and high), including both the prior and two updated fragility curves, based on two different survived load levels (T2 and T10).

Table 2. Prior and updated failure probabilities for the two canal dikes, each with three different fragility curves. Reliability updates are based on T2 and T10 survival loads, with reductions expressed as probability factors and compared to conceptual factors (in italic and shown in brackets).

		Aarlanderveen (h_{dec} = 0.09 m)			Duifpolder (h_{dec} = 0.06 m)		
Strength uncertainty scenario		Reference	Moderate	High	Reference	Moderate	High
Inverse gradient fragility curve [m]		*0.08*	*0.50*	*1.79*	*0.08*	*0.10*	*0.61*
Prior failure probability		<1/1,000,000	1/1,300	1/18	1/47,000	1/120	1/8
T2	Updated failure prob	<1/1,000,000	1/12,000	1/580	1/47,000	1/370	1/130
	Probability factor *(conceptual factor)*	**1.0** *(1.0)*	**9.3** *(7.7)*	**31.8** *(30.7)*	**1.0** *(1.0)*	**3.0** *(1.9)*	**16.3** *(14.8)*
T10	Updated failure prob	<1/1,000,000	1/57,000	1/2,700	1/47,000	1/1,200	1/440
	Probability factor *(conceptual factor)*	**1.0** *(1.0)*	**42.4** *(31.8)*	**149.1** *(162.5)*	**1.0** *(1.0)*	**10.0** *(4.1)*	**54.3** *(70.7)*

compared to Aarlanderveen. These outcomes align with the conceptual analysis, as shown in Table 2, where probability factors are summarised for all strength scenarios and survival loads for both the conceptual analyses and case studies. In the reference strength-uncertainty scenario, the inverse gradient of the fragility curve is small relative to the load decimate height, so reliability updating does not reduce the updated failure probability. While the case study results closely match the conceptual factors, small differences arise. These differences can be explained by the fact that the fragility curves and load statistics of the case studies do not follow statistical distributions, as used in the conceptual analysis. In practice, both the decimate height and the inverse gradient of the fragility curve vary across the range of phreatic line levels.

3.2.2. River dikes

For the river dikes, the phreatic lines exhibit larger variations than in the canal dikes, with decimate heights around 30 and 50 cm. This leads to a total range of about two metres, covering load scenarios that occur on average once per year up to once every 10,000 years. As a result, the fragility curve span a much wider range of load levels, as shown in Figure 10, compared to about 70 cm at Aarlanderveen. Despite this, the inverse gradients of the fragility curves are within a similar range as those of the canal dikes. The updated fragility curve reduces the conditional failure probability for several decimetres above the survived load level, although this effect diminishes at higher phreatic levels.

For the river dikes, the impact of reliability updating is minimal and updated failure probabilities are nearly identical to the prior values, see Table 3. Even for the Gennep dike, which has a relatively high prior failure probability of 1/32 per year under the high strength uncertainty scenario, a survived load level with a 10-year return period provides little new information due to the large load variation. Even more extreme events, such as the 2021 Limburg floods with a return period of 100 years (Strijker et al. 2023), add little value when updating failure probabilities. For example, the Gennep dike's failure probability under T100 conditions is only reduced by a factor of three to seven, depending on the strength uncertainty scenario. This shows that surviving extreme loads doesn't guarantee a significant reduction in failure probabilities, as the impact depends on how the load varies.

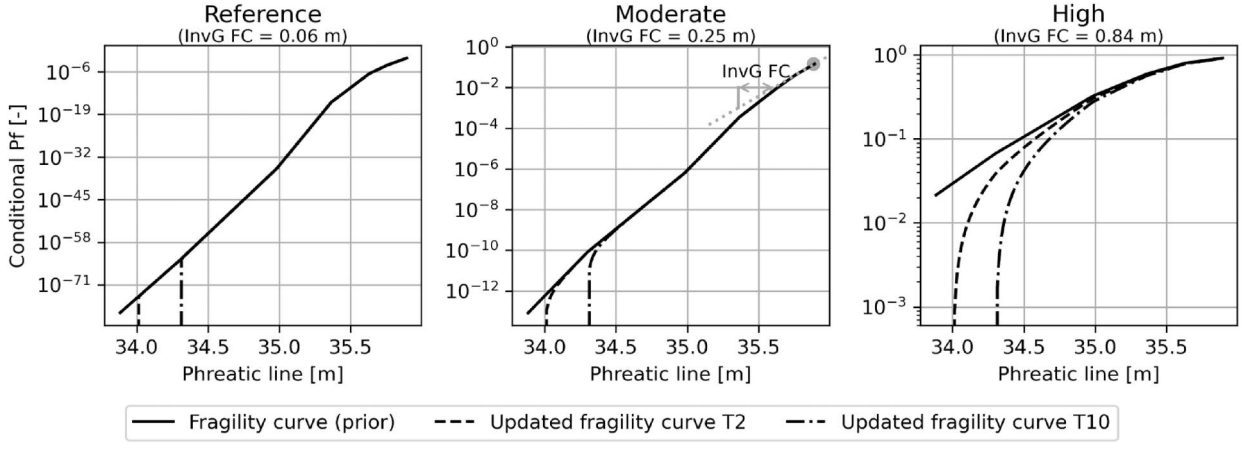


Figure 10. Fragility curves at Meers for three different strength uncertainty scenarios (reference, moderate and high), including both the prior and two updated fragility curves, based on two different survived load levels (T2 and T10).

Table 3. Prior and updated failure probabilities for the two river dikes, each with three different fragility curves. Reliability updates are based on T2 and T10 survival loads, with reductions expressed as probability factors and compared to conceptual factors (in italic and shown in brackets).

		Meers (h_{dec} = 0.27 m)			Gennep (h_{dec} = 0.53 m)		
Strength uncertainty scenario		Reference	Moderate	High	Reference	Moderate	High
Inverse gradient fragility curve [m]		0.06	0.25	0.84	0.49	0.06	0.27
Prior failure probability		<1/1,000,000	1/240,000	1/24	1/49	1/49	1/32
T2	Upd. failure prob	<1/1,000,000	1/240,000	1/59	1/49	1/49	1/32
	Probability factor	**1.0**	**1.0**	**2.5**	**1.0**	**1.0**	**1.0**
	(conceptual factor)	*(1.0)*	*(1.1)*	*(2.4)*	*(1.2)*	*(1.0)*	*(1.3)*
T10	Upd. failure prob	<1/1,000,000	1/240,000	1/110	1/49	1/49	1/35
	Probability factor *(conceptual factor)*	**1.0** *(1.0)*	**1.0** *(1.6)*	**4.7** *(5.7)*	**1.0** *(1.0)*	**1.0** *(1.0)*	**1.1** *(1.8)*

In conclusion, the case studies confirm the findings from the conceptual analysis. The calculated probability factors in the case studies lie close to the ones calculated by the conceptual analysis. Therefore, the results of the conceptual analysis can be seen as a first approximation for estimating the impact of reliability updating under different load and strength characteristics.

3.3. Balancing strength uncertainties to load variations

To prevent non-credible failure probability estimates, a credibility check is proposed. This check helps identify situations where the strength uncertainty is disproportionately high compared to the load variations. Specifically, it targets situations where the failure probability of a dike is mainly influenced by load levels that occur every 10 years or more frequently, and these frequent loads contribute more than 50% to the total failure probability. According to conceptual analysis results, such situations typically arise when the inverse gradient of the fragility curve is at least about 1.5 times greater than the decimate height of the load, as shown in Figure 11. These situations may appear unlikely in dike safety, particularly when the dike has demonstrated stability over many decades. In such cases, the load variations and strength uncertainty are out of balance, leading to unrealistic strength uncertainty compared to load variations, and consequently, non-credible failure probabilities. Reducing the uncertainties related to the dike's strength will lower the inverse gradient of the fragility curve, resulting in more credible failure probability estimates. This can be done by, for example, monitoring load levels in combination with reliability updating methods, or gathering more detailed data on soil properties to obtain a clearer and less uncertain assessment of the dike's strength.

The proposed criterion (inverse gradient $\geq 1.5 \times$ decimate height) assumes that epistemic uncertainties in strength parameters and load distributions remain quasi-stationary over time. The criterion should be applied with caution when:

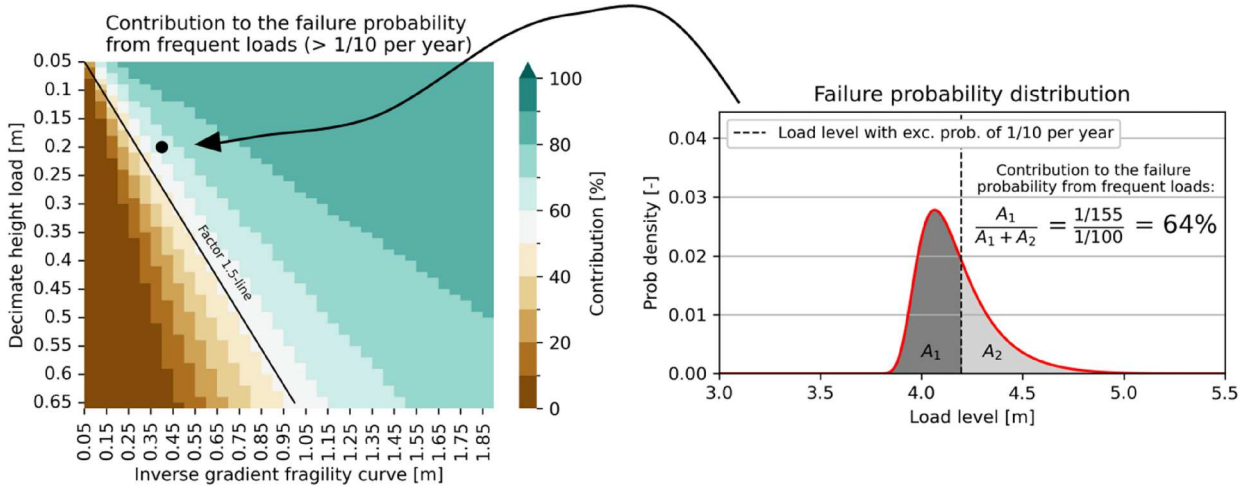


Figure 11. Left: heatmap showing the contribution of load levels that occur frequently (>1/10 per year) to the failure probability across different load variations and strength uncertainties. The factor 1.5-line shows where the inverse gradient is 1.5 times larger than the decimate height. Right: an example of a hypothetical dike with a decimate height of the load of 20 cm and a fragility curve with an inverse gradient of 40 cm, where frequent loads contribute 64% to the overall failure probability.

- Significant strength degradation is expected, for example, due to wet-dry cycles (Azizi, Musso, and Jommi 2020; Tang et al. 2010), soil strain softening (Skempton 1964), or other time-dependent strength loss mechanisms
- Major reinforcement or reconstruction alters the dike cross-section
- Climate-driven shifts substantially change the load distributions

In such cases, the balance between load variation and strength uncertainty may evolve, requiring reassessment of the criterion's applicability.

4. Discussion

4.1. Implications

This section reflects on the way the findings of this study can help estimate probabilities of dike instability in practice. It starts with identifying regions in the Netherlands where dike failure probability estimates can potentially be improved using reliability updating, based on information about the load variations. It finishes with a call to consider load variations as stochastic variables, rather than design values, even when the variations are small.

4.1.1. Regions in the Netherlands with small load variations

One of the key characteristics of the over 10,000 km of Dutch canal dikes is small load variations in both the water level and the phreatic line (Strijker and Kok 2025). Furthermore, there are many primary flood defences across the Netherlands with a large variety in load variations. The variety in water level statistics is available throughout the Netherlands and is calculated using a probabilistic tool called Hydra-NL, which combines physical and statistical models (Geerse 2011). As can be seen in Figure 12, the decimate heights of water levels along the primary dikes in the lake and delta areas are the smallest. These are areas where water levels are regulated by weirs, pumping stations, and storm surge barriers, which reduce variations in water levels. Therefore, these controlled systems with small water level variations offer advantages: primary dikes are exposed to lower absolute loads than in uncontrolled systems, and the limited variation helps to assess actual dike's strength and performance, since extreme levels are close to those occurring annually.

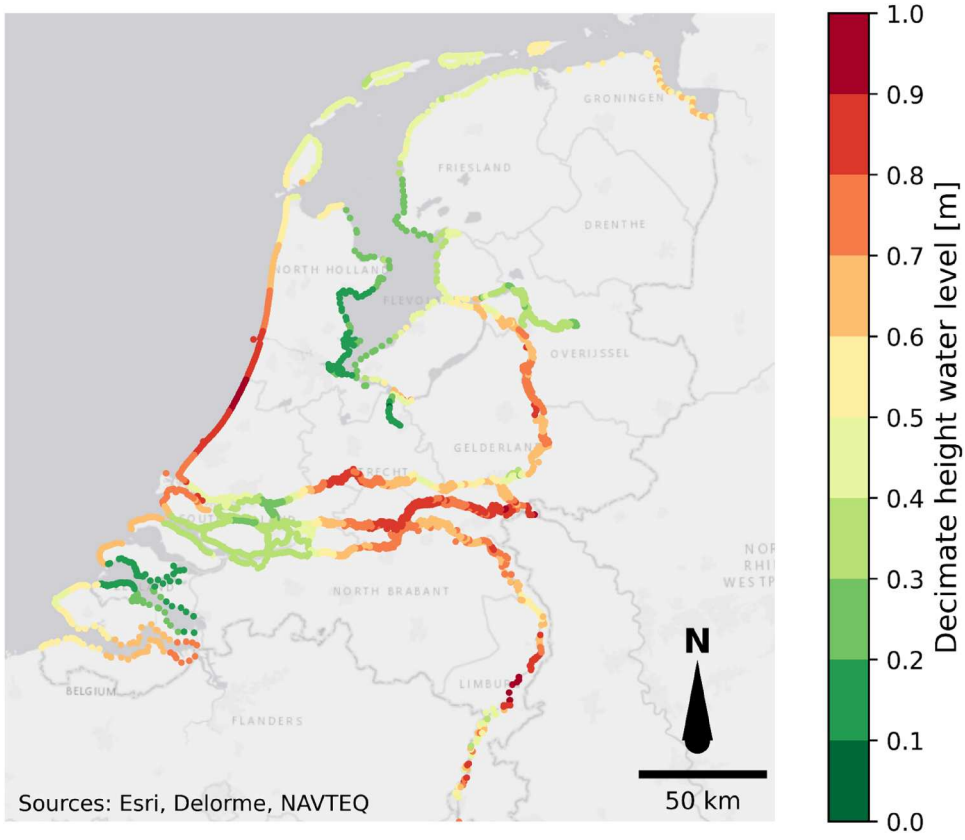


Figure 12. The decimate height of the water levels along the primary dikes in the Netherlands, based on data from TMR2006 (Bouw 2008).

4.1.2. The value of considering the load probabilistically

In many geotechnical problems, the uncertainty in the strength dominates the reliability estimates, while the influence of loads is relatively low (European Commission 2024). Therefore, it can be practical to treat the load side of the problem semi-probabilistically using design values. This approach is also applied in the current safety assessment of Dutch canal dikes, where load variations are relatively small. However, when assessing failure probabilities, loads can provide insight into the schematisation of the strength. Considering the load probabilistically helps to realistically estimate stochastic strength variables through reliability updating. It also provides insight into failure processes, such as the sensitivity to load variations that can trigger failures. Importantly, dike instabilities occur due to variables that change over time, even when their fluctuations are small. In safety assessment, time-variant and time-invariant, inherently or epistemically uncertain, are often treated similarly probabilistically. However, in reality, the time-variant variables trigger failure, and we must keep these variables in mind in safety assessments. Therefore, we argue that time-variant loads should always be considered as probabilistic variables.

4.2. Limitations and future research

4.2.1. Model uncertainties

The approach in this study did not include any model uncertainties. Model uncertainties were excluded to avoid introducing additional complexity and to allow for a clearer interpretation. However, there are model uncertainties associated with, for example, modelling the stability and estimating the pore-water pressures. In practice, these model uncertainties tend to be compensated using conservative and safe assumptions in safety assessments, like a safe assumption for the phreatic line in the river dikes or strength modelling (European Commission 2024). Nevertheless, adding these overlooked model uncertainties is certain to increase the prior failure probability, but their influence on the updated failure probability is less straightforward, and a qualitative reasoning is given. Adding additional uncertainties in the strength, such as model uncertainty related to stability modelling, results in a fragility curve that shifts upward and becomes flatter. This, in turn, is likely to increase the impact of reliability updating, resulting in larger probability factors. Adding model uncertainty to the schematisation of the load does not necessarily change the decimate height; it is more likely to affect only the absolute level. This does not affect the impact of reliability updating in terms of the probability factor, but it does increase the prior and updated failure probability. Furthermore, this study assumed a clear survived load level for reliability updating, which can be in practice rather uncertain, because of measurement errors and the translation from point measurements to spatial variables, such as the phreatic line. Accounting for these uncertainties, rather than using a best estimate, will reduce the impact of reliability updating.

4.2.2. Multivariate load variations

Failure probabilities were estimated using fragility curves, which express the relation between load and conditional failure probability. In the case studies, we only considered one load variable, namely the pore-water pressures in and under the dike. However, the load can also be a combination of pore-water pressures and traffic load. In these cases, a fragility plane can be used where the failure probability is conditional on two variables (Nofal, van de Lindt, and Do 2020). This can make failure probability analysis more challenging, as well as reliability updating. This traffic load is time-variant and, therefore, inherently uncertain (Lendering, Schweckendiek, and Kok 2018). The way this load variable is considered within failure probability analysis, deterministic or probabilistic, affects the potential of reliability updating. In future extensions, attention should be given to the dependency between loads, such as time-varying pore-water pressures and traffic load, as well as to the incorporation of multiple monitoring indicators.

4.2.3. Modelling shear strength and the value of historic data

When it comes to the reliability of dikes, the behaviour of dike material is never constant over the long term. Soils are subjected to weather and other forces of the environment, making soil characteristics vary over time. For example, strength parameters can evolve due to factors such as soil aging, dry-wet cycling, and human interventions like maintenance or reinforcement efforts. Additionally, load characteristics can change, because of climate change with increasing extreme precipitation and more frequent high water levels (Van Dorland et al. 2023). This dynamic nature of load and strength means that failure probabilities and dike safety change over time. Whether this has a positive or negative impact on dike stability depends on site-specific dike characteristics. Ultimately, the challenge to reliability updating lies in finding the right balance between understanding evolving strength and load parameters and leveraging historical measurements. Regular monitoring, combined with probabilistic

methods that incorporate past data, helps create a clearer picture of the dike's true performance.

5. Conclusions

This study aims to better understand the role of load variations in reliability updating and in assessing credible failure probabilities. First, this was conceptually examined using hypothetical dikes with various load and strength configurations described by Gumbel distributions. It was found that the impact of reliability updating increases when load variations are smaller, regardless of the prior failure probabilities. The reason is that only a relatively small load range contributes to the calculated failure probability, which, in absolute terms, is close to the load levels that occur every year. By updating the fragility curve, the conditional failure probabilities closest to the survived load level are reduced the most. Therefore, reliability updating has a larger impact on the failure probability for situations with smaller load variations. Second, these findings were confirmed in the case studies of two canal dikes and two river dikes in the Netherlands, which varied in terms of fragility curves and load variations. The two canal dikes with decimate heights smaller than 10 cm benefited the most from reliability updating, while the updated failure probabilities of the river dikes changed only slightly. This is because the load variations of the two river dikes are larger with decimate heights of the phreatic line of approximately 30 and 60 cm, while the inverse gradients of the fragility curve were similar. In the moderate-strength scenario for the canal dikes, reliability updating with a load that occurs once every two years reduces the failure probabilities by a factor of three to nine, relative to the prior estimates. However, for the river dikes, even extreme load levels with a 1/100 probability per year, as occurred during the 2021 Limburg floods, contribute little when it comes to improving failure probabilities through reliability updating.

Based on the conceptual analysis, a credibility check for safety assessments was proposed, identifying situations where the failure probability is dominated by load levels that occur every 10 years or more frequently. The contribution of frequent load levels to the total failure probability exceeds 50% when the inverse gradient of the fragility curve is about 1.5 times greater than the decimate height of the load. In these situations, the load variations and strength uncertainty are out of balance, making it challenging to estimate credible failure probabilities, as this can contradict the stability of the dikes over many decades. Situations of small load variation occur widely in the Netherlands, for example, along much of the over 10,000 kilometres of canal dikes and in primary flood defences in lake and delta regions. In these situations, reducing the uncertainties related to the dike's strength will lower the inverse gradient of the fragility curve, resulting in more credible failure probability estimates. This can be achieved by, for example, monitoring load levels in combination with reliability updating methods or gathering more detailed data on soil strength.

Acknowledgments

We gratefully acknowledge the support and valuable insights provided by the water authorities Hoogheemraadschap Schieland & de Krimpenerwaard (HHSK), Hoogheemraadschap Delfland (DL), and Hoogheemraadschap Rijnland (RL) regarding practical challenges in the field of dike safety.

Author contributions

B. Strijker: Writing – review & editing, Writing – original draft, Visualization, Methodology, Formal analysis, Data curation, Conceptualization; M. Kok: Writing – review & editing, Supervision, Resources, Project administration, Methodology, Conceptualization; S.N. Jonkman: Writing – review & editing, Supervision, Conceptualization

Disclosure statement

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Bart Strijker is employed by HKV Consultants. Matthijs Kok is also employed by HKV Consultants.

Funding

This work was funded by STOWA and Rijkswaterstaat, The Netherlands.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the author(s) used ChatGPT (4o) to support the editing process and improve clarity and logical flow. The author(s) subsequently reviewed and revised the content and take full responsibility for the final publication.

Code availability statement

The script used to produce example cases, key figures illustrating the methodology, and the main results of the conceptual analysis are available from 4TU. ResearchData at: https://doi.org/10.4121/acd1f26c-15c4-47f6-b01a-0120d104031a

It includes the construction of fragility curves and the direct reliability updating procedure described in the manuscript.

References

Azizi, Arya, Guido Musso, and Cristina Jommi. 2020. "Effects of Repeated Hydraulic Loads on Microstructure and Hydraulic Behaviour of a Compacted Clayey Silt." *Canadian Geotechnical Journal* 57 (1): 100–114. https://doi.org/10.1139/cgj-2018-0505.

Bachmann, D., N. P. Huber, G. Johann, and H. Schüttrumpf. 2013. "Fragility Curves in Operational Dike Reliability Assessment." *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards* 7 (1): 49–60. https://doi.org/10.1080/17499518.2013.767664.

Bakker, M., and F. Schaars. 2019. "Solving Groundwater Flow Problems with Time Series Analysis: You May Not Even Need Another Model." *Groundwater* 57 (6): 826–833. https://doi.org/10.1111/gwat.12927.

Bayes, T. 1763. "LII. An Essay towards Solving a Problem in the Doctrine of Chances. By the Late Rev. Mr. Bayes, FRS Communicated by Mr. Price, in a Letter to John Canton, AMFR S." *Philosophical Transactions of the Royal Society of London* 53:370–418. https://doi.org/10.1098/rstl.1763.0053.

Bishop, Alan W. 1955. "The Use of the Slip Circle in the Stability Analysis of Slopes." *Géotechnique* 5 (1): 7–17. https://doi.org/10.1680/geot.1955.5.1.7.

Bouw, R. 2008. "Decimate Heights TMR2006" (Report No. 373295) [in Dutch]. Witteveen+Bos." 2008. https://open.rijkswaterstaat.nl/open-overheid/onderzoeksrapporten/@26517/decimeringshoogten-tmr2006/.

Calle, Ed. 1996. *Characteristic Values of Geotechnical Parameters*. Lecture notes from the PAO Seminar on Soft Soil Engineering, June 10–14, Noordwijk, The Netherlands. Delft: Foundation Post Academic Education.

Calle, E. O. F., W. Kanning, and T. Schweckendiek. 2021. "Characteristic Values of Soil Properties in Dutch Codes of Practice-Theoretical Backgrounds and Assumptions." *Deltares 11206883-014-GEO-0001*, Delft.

Casciati, F., L. Faravelli, and P. Venini. 1991. "Stochastic Equivalent Linearization of Nonlinear Complex Structures." In *Computational Stochastic Mechanics*, edited by D. Spanos and C. A. Brebbia, 349–358. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-011-3692-1_30.

CEN. 2004. "Eurocode 7: Geotechnical Design. Part 1: General Rules, EN1997-1".

De Koker, N., C. Viljoen, R. Lenner, and S. W. Jacobsz. 2020. "Updating Structural Reliability Efficiently Using Load Measurement." *Structural Safety* 84:101939. https://doi.org/10.1016/j.strusafe.2020.101939.

European Commission. 2024. "Joint Research Centre, van Den Eijnden, B., Knuuti, M., Lesny, K., Löfman, M., Mavritsakis, A., Roubos, A., Schweckendiek, T., Sciarretta, F., Ebener, A., Escher, K., Spross, J., Commend, S., Hehenkamp, M., Arnold, P., Wilhelm, S., Ene, A., Rimoldi." 2024. https://data.europa.eu/doi/10.27601342542.

Geerse, C. P. M. 2011. "Hydra-Zoet for the Fresh Water Systems in the Netherlands-Probabilistic Model for the Assessment of Dike Heights." *PR2168 HKV Rapport Voor Rijkswaterstaat, Waterdienst*.

Georegister, Nationaal. 2024. "Toets- En Veiligheidsoordelen Primaire Waterkeringen (WBI)." *Metadata Date* 0–2024:4–09.

Hemel, M. J., D. J. Peters, T. Schweckendiek, and S. N. Jonkman. 2024. "Reliability Updating for Lateral Failure of Historic Quay Walls." *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards* 18 (4): 882–903. https://doi.org/10.1080/17499518.2024.2302141.

Hicks, Michael A., and Yajun Li. 2018. "Influence of Length Effect on Embankment Slope Reliability in 3D." *International Journal for Numerical and Analytical Methods in Geomechanics* 42 (6): 891–915. https://doi.org/10.1002/nag.2766.

Huang, Wei-Che, Wen-Cheng Liu, and Hong-Ming Liu. 2025. "Uncertainty Analysis of Overflow Due to Sea Dike Failure during Typhoon Events." *Journal of Marine Science and Engineering* 13 (3): 573. https://doi.org/10.3390/jmse13030573.

Jongejan, R. B., and B. Maaskant. 2015. "Quantifying Flood Risks in the Netherlands." *Risk Analysis* 35 (2): 252–264. https://doi.org/10.1111/risa.12285.

Jonkman, S. N., R. Jongejan, and B. Maaskant. 2011. "The Use of Individual and Societal Risk Criteria within the Dutch Flood Safety Policy – Nationwide Estimates of Societal Risk and Policy Applications." *Risk Analysis: An International Journal* 31 (2): 282–300. https://doi.org/10.1111/j.1539-6924.2010.01502.x.

Kind, J. M. 2014. "Economically Efficient Flood Protection Standards for the N Etherlands." *Journal of Flood Risk Management* 7 (2): 103–117. https://doi.org/10.1111/jfr3.12026.

Klijn, F., N. Asselman, and E. Mosselman. 2019. "Robust River Systems: On Assessing the Sensitivity of Embanked Rivers to Discharge Uncertainties, Exemplified for the Netherlands' Main Rivers." *Journal of Flood Risk Management* 12 (S2): e12511. https://doi.org/10.1111/jfr3.12511.

Krogt, M. G. van der, W. J. Klerk, W. Kanning, T. Schweckendiek, and M. Kok. 2022. "Value of Information of Combinations of Proof Loading and Pore Pressure Monitoring for Flood Defences." *Structure and Infrastructure Engineering* 18 (4): 505–520. https://doi.org/10.1080/15732479.2020.1857794.

Lendering, K., T. Schweckendiek, and M. Kok. 2018. "Quantifying the Failure Probability of a Canal Levee." *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards* 12 (3): 203–217. https://doi.org/10.1080/17499518.2018.1426865.

Li, Dian-Qing, Shui-Hua Jiang, Zi-Jun Cao, Wei Zhou, Chuang-Bing Zhou, and Limin Zhang. 2016. "A Multiple Response-Surface Method for Slope Reliability Analysis considering Spatial Variability of Soil Properties." *Engineering Geology* 187:60–72. https://doi.org/10.1016/j.enggeo.2015.12.003.

Liu, Xin, Dian-Qing Li, Zi-Jun Cao, and Yu Wang. 2020. "Adaptive Monte Carlo Simulation Method for System Reliability Analysis of Slope Stability Based on Limit Equilibrium Methods." *Engineering Geology* 264:105384. https://doi.org/10.1016/j.enggeo.2019.1053.

Nofal, O. M., J. W. van de Lindt, and T. Q. Do. 2020. "Multi-variate and Single-Variable Flood Fragility and Loss Approaches for Buildings." *Reliability Engineering & System Safety* 202:106971.

Phoon, K. K. 2023. ""What Geotechnical Engineers Want to Know about Reliability." *ASCE-ASME Journal of Risk and*

Uncertainty in Engineering Systems, Part A: Civil Engineering 9 (2): 03123001. https://doi.org/10.1061/ajrua6.rueng-1002.

Phoon, Kok-Kwang.. 2016. "Chapter 1 Reliability as a Basis for Geotechnical Design." In *Reliability of Geotechnical Structures in Iso2394*, edited by K. K. Phoon and J. V. Retief, 1–32. Leiden: CRC Press. https://doi.org/10.1201/9781315364179-2.

Phoon, Kok-Kwang, and Fred H. Kulhawy. 1999. "Characterization of Geotechnical Variability." *Canadian Geotechnical Journal* 36 (4): 612–624. https://doi.org/10.1139/t99-038.

Ridley, A., B. McGinnity, and P. Vaughan. 2004. "Role of Pore Water Pressures in Embankment Stability." *Proceedings of the Institution of Civil Engineers-Geotechnical Engineering* 157 (4): 193–198. https://doi.org/10.1680/geng.2004.157.4.193.

Rikkert, S. J. H. 2022. "A System Perspective on Flood Risk in Polder Drainage Canal Systems." PhD Thesis, Delft University of Technology.

Rikkert, S. J. H., M. Kok, K. Lendering, and R. Jongejan. 2022. "A Pragmatic, Performance-Based Approach to Levee Safety Assessments." *Journal of Flood Risk Management* 15 (4): e12836. https://doi.org/10.1111/jfr3.12836.

Schultz, Martin, Ben Gouldby, Jonathan Simm, and J. L. Wibowo. 2010. *Beyond the Factor of Safety: Developing Fragility Curves to Characterize System Reliability*. Vicksburg, MS: Water Resources Infrastructure Program. https://doi.org/10.21236/ada525580.

Schweckendiek, T. 2014. "On Reducing Piping Uncertainties: A Bayesian Decision Approach." PhD Thesis, Delft University of Technology.

Schweckendiek, T., M. G. Van Der Krogt, A. Teixeira, W. Kanning, R. Brinkman, and K. Rippi. 2017. "Reliability Updating with Survival Information for Dike Slope Stability Using Fragility Curves." *Geo-Risk* 2017:494–503.

Schweckendiek, T., and A. C. W. M. Vrouwenvelder. 2013. "Reliability Updating and Decision Analysis for Head Monitoring of Levees." *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards* 7 (2): 110–121. https://doi.org/10.1080/17499518.2013.791034.

Schweckendiek, T., A. C. W. M. Vrouwenvelder, and E. O. F. Calle. 2014. "Updating Piping Reliability with Field Performance Observations." *Structural Safety* 47:13–23. https://doi.org/10.1016/j.strusafe.2013.10.002.

Schweckendiek, Timo. 2011. "Reassessing Reliability Based on Survived Loads." *Coastal Engineering Proceedings* 32 (January): 23. https://doi.org/10.9753/icce.v32.structures.23.

Sharp, M., M. Wallis, F. Deniaud, R. Hersch-Burdick, R. Tourment, E. Matheu, Y. Seda-Sanabria, et al. 2013. "The International Levee Handbook".

Simpson, B., J. W. Pappin, and D. D. Croft. 1981. "An Approach to Limit State Calculations in Geotechnics." *Ground Engineering* 14 (6): 21–28.

Skempton, A. W. 1964. "Long-Term Stability of Clay Slopes." *Géotechnique* 14 (2): 77–102. https://doi.org/10.1680/geot.1964.14.2.77.

Slomp, R., H. Knoeff, A. Bizzarri, M. Bottema, and W. de Vries. 2016. "Probabilistic Flood Defence Assessment Tools." In *E3S Web of Conferences (Vol.* p. 03015). EDP Sciences: 7.

STOWA. 2023. "Soil Parameters for the Assessment and Design of Flood Defences (STOWA 2023-37). Foundation for Applied Water Research (STOWA)".

Straub, D. 2011. "Reliability Updating with Equality Information." *Probabilistic Engineering Mechanics* 26 (2): 254–258. https://doi.org/10.1016/j.probengmech.2010.08.003.

Straub, D., and I. Papaioannou. 2015. "Bayesian Updating with Structural Reliability Methods." *Journal of Engineering Mechanics* 141 (3): 04014134. https://doi.org/10.1061/(asce)em.1943-7889.0000839.

Strijker, B., N. Asselman, J. de Jong, and H. Barneveld. 2023. "The 2021 Flood Event in the Dutch Meuse and Tributaries from a Hydraulic and Morphological Perspective." *Journal of Coastal and Riverine Flood Risk* 2:6. https://doi.org/10.59490/jcrfr.2023.0006.

Strijker, B., T. J. Heimovaara, S. N. Jonkman, and M. Kok. 2024. "Exploring Subsurface Water Conditions in Dutch Canal Dikes during Drought Periods: Insights from Multiyear Monitoring." *Water Resources Research* 60 (9): e2023WR036046. https://doi.org/10.1029/2023wr036046

Strijker, B., and M. Kok. 2025. "The Dynamics of Peak Head Responses at Dutch Canal Dikes and the Impact of Climate Change." *EGUsphere* 2024:1–36. https://doi.org/10.5194/egusphere-2024-1495.

Strijker, Bart, and Matthijs Kok. 2023. "The Joint Impact of Rainfall Events on Water- and Dike Systems in Dutch Polders." May. https://doi.org/10.5194/egusphere-egu23-9048.

Tang, Chao-Sheng, Yu-Jun Cui, Anh Minh Tang, and Bin Shi. 2010. "Experimental Evidence on the Temperature Dependence of Desiccation Cracking Behavior of Clayey Soils." *Engineering Geology* 114 (3-4): 261–266. https://doi.org/10.1016/j.enggeo.2010.05.003.

TAW. 2004. "Technisch Rapport Waterspanningen Bij Dijken Waterkeringen (TAW)".

Van, M. A., A. R. Koelewijn, and F. B. Barends. 2005. "Uplift Phenomenon: Model, Validation, and Design." *International Journal of Geomechanics* 5 (2): 98–106. https://doi.org/10.1061/(asce)1532-3641(2005)5:2(98).

Van Baars, S., and I. M. Van Kempen. 2009. "The Causes and Mechanisms of Historical Dike Failures in the Netherlands." *E-Water Official Publication of the European Water Association* : 1–14. https://doi.org/10.2175/193864702784164505.

Van Der Krogt, M. G., T. Schweckendiek, and M. Kok. 2018. "Uncertainty in Spatial Average Undrained Shear Strength with a Site-specific Transformation Model." *Georisk Assessment and Management of Risk for Engineered Systems and Geohazards* 13 (3): 226–236. https://doi.org/10.1080/17499518.2018.1554820.

Van der Meer, J. W., W. L. A. Ter Horst, and E. H. Van Velzen. 2009. "Calculation of Fragility Curves for Flood Defence Assets." In *Flood Risk Management*, edited by P. Samuels, S. Huntington, W. Allsop, and J. Harrop, 567–573. London: Taylor & Francis Group.

Van der Meij, R. 2023. "D-Stability User Manual (Version 2024.01, December 2023)." 2023. https://www.deltares.nl.

Van Dorland, R., J. Beersma, J. Bessembinder, N. Bloemendaal, H. van Den Brink, M. Brotons Blanes, S. Drijfhout, et al. 2023. *Knmi National Climate Scenarios 2023 for The Netherlands*. De Bilt, The Netherlands: KNMI.

Vanmarcke, Erik H. 1977. "Reliability of Earth Slopes." *Journal of the Geotechnical Engineering Division* 103 (11): 1247–1265. https://doi.org/10.1061/ajgeb6.0000518.

Vrouwenvelder, Ton, Silvia Dimova, Luisa Sousa, Jana Marková, Giuseppe Mancini, Ulrike Kuhlmann, Andreas Taras, et al. 2024. *Reliability Background of the Eurocodes: Support to the Implementation, Harmonization and Further Development of the Eurocodes*. Luxembourg: Publications Office of the European Union. https://doi.org/10.2760/9482837.

Wang, Lu, Changzhi Wu, Xing Gu, Honglei Liu, Guifu Mei, and Wengang Zhang. 2020. "Probabilistic Stability Analysis of Earth dam Slope under Transient Seepage Using Multivariate Adaptive Regression Splines." *Bulletin of Engineering Geology and the Environment* 79 (6): 2763–2775. https://doi.org/10.1007/s10064-020-01730-0.

Woerkom, T. van, R. van Beek, H. Middelkoop, and M. F. Bierkens. 2022. "Assessing Lithological Uncertainty in Dikes: Simulating Construction History and Its Implications for Flood Safety Assessment." *Journal of Flood Risk Management* 15 (4): e12848. https://doi.org/10.1111/jfr3.12848.

Wojciechowska, K. A. 2015. "Advances in Operational Flood Risk Management in the Netherlands." PhD Thesis, Dissertation (TU Delft), Delft University of Technology. K. A. Wojciechowska. https://doi.org/10.4233/uuid:5d719beb-bbbf-4fde-8e10-526785315fd1.

Wong, T. E., D. A. J. Batjes, and J. de Jager. 2007. *Geology of the Netherlands*. Amsterdam: Royal Netherlands Academy of Arts and Sciences.

Yang, Zhiyong, Xueyou Li, Chao Xu, and Guangming Yu. 2025. "Bayesian Updating of Soil-Water Characteristic Curves for Slope Reliability Analysis." *Engineering Geology* 353:108148. https://doi.org/10.1016/j.enggeo.2025.108148.

Yang, Zhiyong, Chengchuan Yin, Xueyou Li, Shuihua Jiang, and Dianqing Li. 2024. "Efficient Slope Reliability and Sensitivity Analysis Using Quantile-Based First-Order Second-Moment Method." *Journal of Rock Mechanics and Geotechnical Engineering* 16 (11): 4192–4203. https://doi.org/10.1016/j.jrmge.2024.04.007.

Zheng, Yanhao, Jinhui Li, Xianshuo Zheng, Ning Guo, and Guanghua Yang. 2024. "Pre- and Post-failure Behaviour of a Dike after Rapid Drawdown of River Level Based on Material Point Method." *Computers and Geotechnics* 170:106269. https://doi.org/10.1016/j.compgeo.2024.106269.

Zheng, Yanhao, Junru Li, Jinhui Li, and Yintang Wang. 2025. "Quantitative Risk Assessment of Road Posed by Levee Slope Failure: A Novel Framework Integrating Monte Carlo Simulation and Material Point Method." *Engineering Geology* 353:108148. https://doi.org/10.1016/j.enggeo.2024.108148.

Appendix

A1. Load scenarios in case studies

For every case, at least five load situations are considered, ranging from yearly occurring levels to a situation with a probability of occurrence of 1/10,000 per year, discretized with load levels that differ by a maximum probability factor of 10.

The water level statistics of the various case studies are as follows.

Return period [years]	Aarlanderveen	Duifpolder	Meers	Gennep
1	−2.10	−0.45	33.88	9.23
10	−1.95	−0.375	36.01	11.99
100	−1.80	−0.3	37.37	13.28
1,000	−1.65	−0.225	38.13	14.14
10,000	−1.50	−0.15	38.66	14.82

The head level at the inner-crest in the various case studies is as follows:

Return period [years]	Aarlanderveen	Duifpolder	Meers	Gennep
1	−2.53	−0.68	33.88	9.8
10	−2.43	−0.61	34.51	10.49
100	−2.33	−0.54	35.87	11.78
1,000	−2.23	−0.47	36.63	12.64
10,000	−1.40	−0.15	37.16	13.92

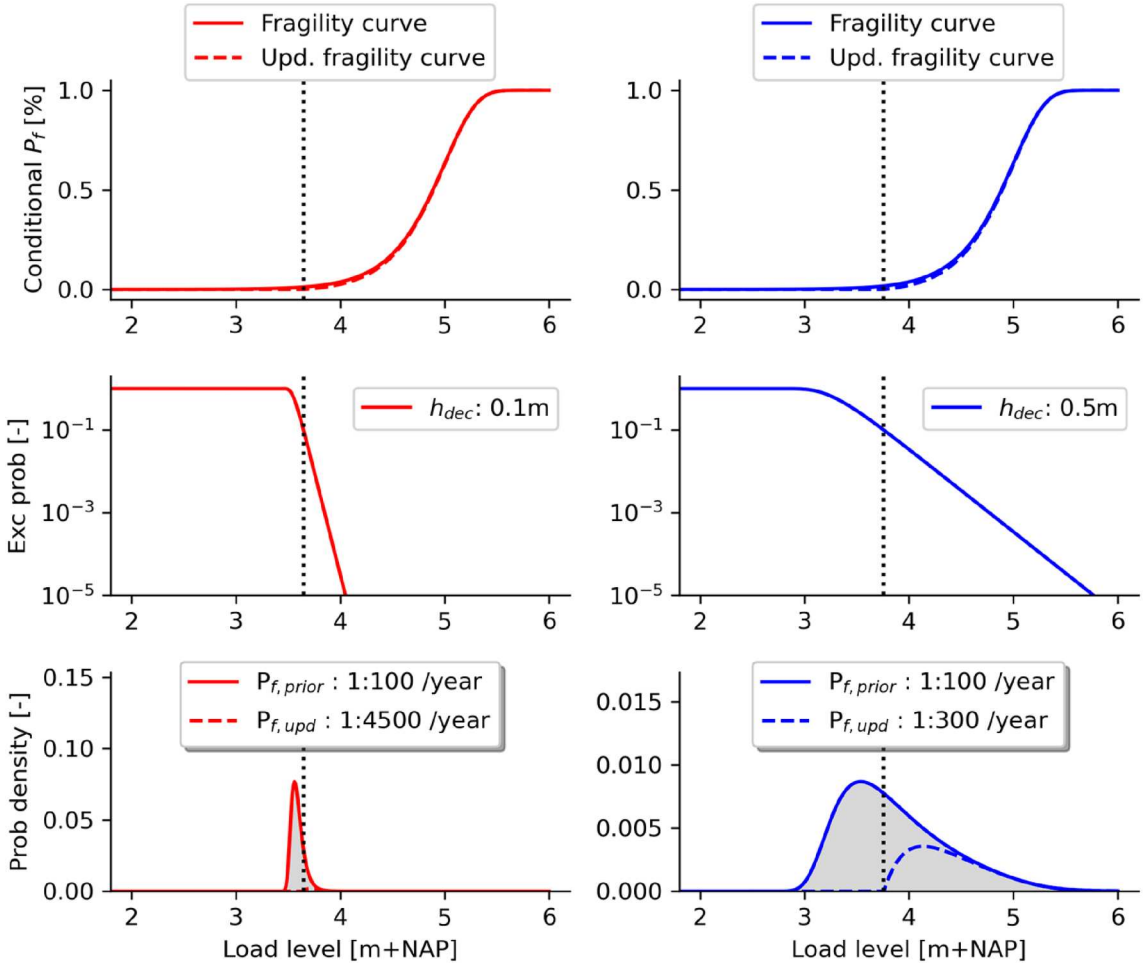
In three of the four cases, the contribution of the load scenarios with a return period of 10,000 years is small, except for the dike at Meers. Therefore, two additional scenarios were added with average return periods of 30,000 and 100,000 years.

Furthermore, for the Gennep case, two additional load scenarios with return periods of 30 and 300 years were added, because uplift occurred there. This caused a bending fragility curve that was better captured by including these extra scenarios.

A2. Illustrative example of two dikes

Let's look at two dikes with the same fragility curve and an initial failure probability of 1/100 per year. The key difference lies in the load statistics: one dike faces small load variations (left), while the other faces large load variations (right). Both dikes have survived a load level corresponding to a 10-year return period. After updating, the failure probability is lower for the dike with smaller load variations.

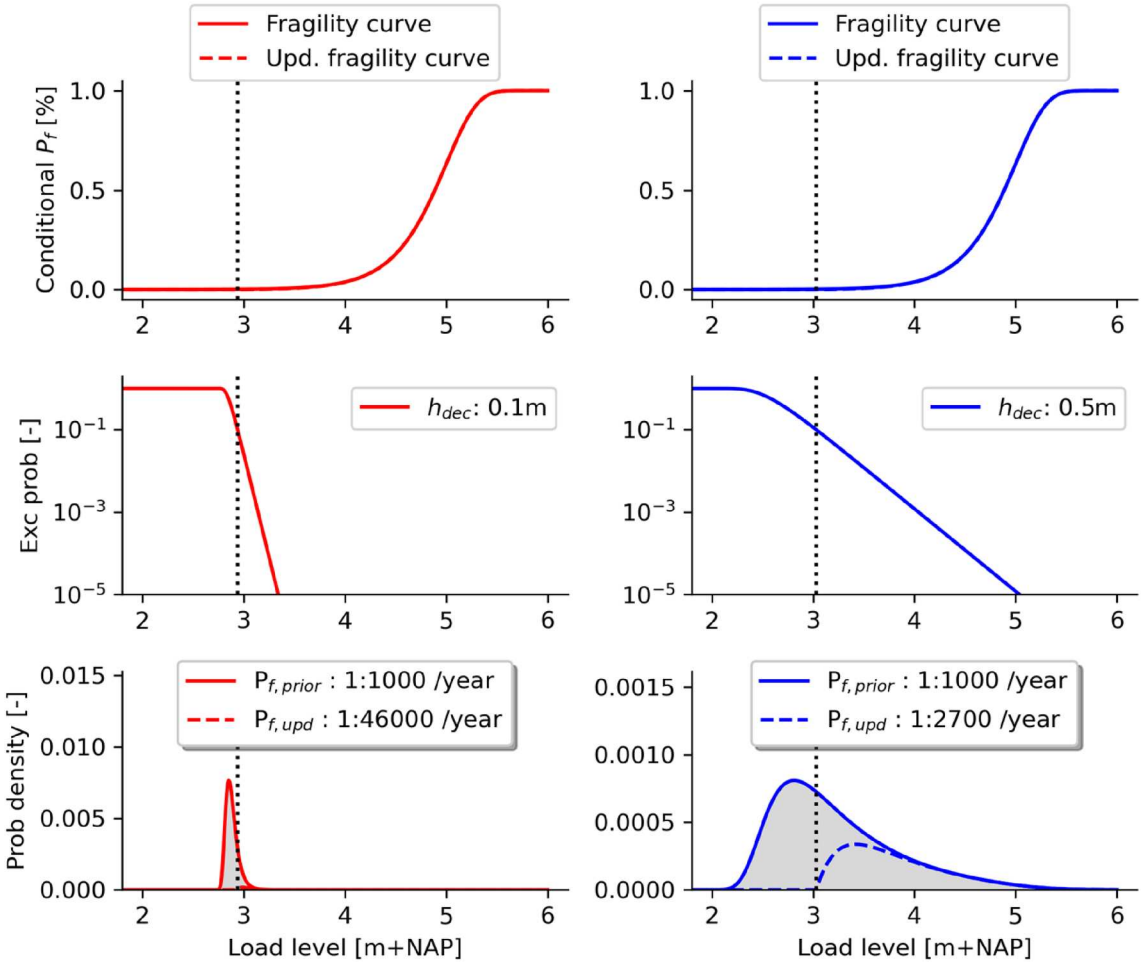
Now, consider two new dikes with similar load statistics and inverse fragility curve gradients, but with a different mean strength: the fragility curve is shifted to the right. The initial failure probabilities are now 1/1,000 per year – ten times smaller than in the first example. Again, they have survived a load level corresponding to a 10-year return period. The failure probability distributions have the same shape, but the scale of the y-axis decreased by a factor of ten. The effect of updating is similar in terms of the ratio between the prior and the updated failure probabilities.



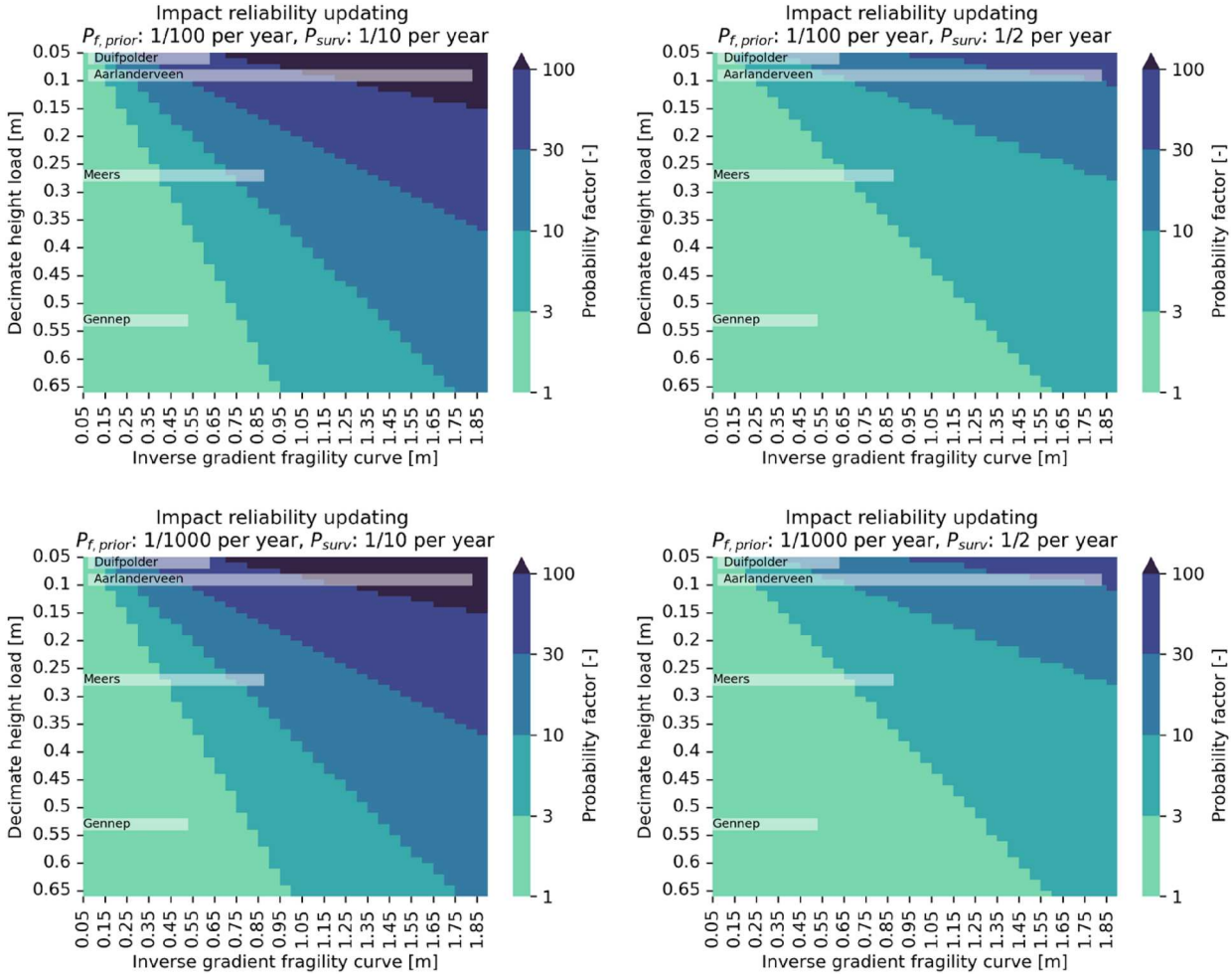
Two combinations of load statistics (red: small load variations, blue: large load variations) and the same prior fragility curves (inverse gradient of 0.4m). The vertical dotted line indicates the survived 10-year load level. Both cases have the same prior failure probability of 1/100 per year, while the updated probabilities differ.

A3. Additional figures for conceptual analysis

The figures below show the impact of reliability updating under different prior failure probabilities (1/100 and 1/1,000 per year) and survival load levels (levels with exceedance probabilities of 1/2 and 1/10 per year).



Two combinations of load statistics (red: small load variations, blue: large load variations) and the same prior fragility curves (inverse gradient of 0.4m). The vertical dotted line indicates the survived 10-year load level. Both cases have the same prior failure probability of 1/1,000 per year, while the updated probabilities differ.



Four heatmaps showing the probability factor after reliability updating as a function of the inverse gradient of the fragility curve and the decimate height load, for two prior failure probabilities (1:100 and 1:1000 per year) and two survived load levels (exceedance probabilities of 1:10 and 1:2 per year). The four case study sites are indicated by horizontal markers.