On the reliability assessment of a controlled dyke failure

de Gast, T.; Hicks, M.A.; van den Eijnden, A.P.; Vardon, P.J.

DOI
10.1680/jgeot.19.SiP.003

Publication date
2020

Document Version
Final published version

Published in
Geotechnique: international journal of soil mechanics

On the reliability assessment of a controlled dyke failure

TOM DE GAST*, MICHAEL A. HICKS*, ABRAHAM P. VAN DEN EIJNDEN* and PHILIP J. VARDON*

A reliability-based analysis framework, accounting for uncertainty arising from the spatial variability of soil properties, has been validated for the controlled, well-instrumented slope failure of an historic dyke in the Netherlands. Using soil property statistics derived from the results of laboratory and cone penetration test (CPT) data for the different soil layers at the site, the dyke was analysed for the initial (i.e. operating) conditions, as well as for the later stage of the test leading up to failure. The computed probabilities of failure and back-figured factors of safety were consistent with the point at which failure occurred in the test, as was the range of possible failure mechanisms. The uncertainty in the stability assessment was reduced by considering the spatial nature of the soil variability, and by conditioning analyses to CPT measurement data. It is shown that the reliability-based approach enables more informed stability assessments that could make the difference between a dyke being assessed as safe or requiring costly improvement.

KEYWORDS: embankments; failure; finite-element modelling; full-scale tests; statistical analysis

INTRODUCTION

The spatial variability of soils at different scales influences material behaviour and the response of geotechnical structures, as well as causing uncertainty in stability assessments and geotechnical design (Hicks, 2005, 2006). There has therefore been much research into the measurement and quantification of spatial variability (e.g. Vanmarcke, 1977; Campanella *et al.*, 1987; Wickremesinghe & Campanella, 1993; Phoon & Kulhawy, 1999; Lloret-Cabot *et al.*, 2014; Fenton *et al.*, 2018; de Gast *et al.*, 2019), and into probabilistic methods of analysis for propagating the effects of uncertainty from the material level to the geotechnical structure level. These methods have included semi-analytical methods, such as the point estimate method (Rosenblueth, 1975), first-order reliability method (Ang & Tang, 1984) and first-order second moment method, as well as computational methods, such as those linking random fields with various limit equilibrium methods (Cho, 2007; Jiang *et al.*, 2014; Javankhoshdel *et al.*, 2017), sometimes referred to as the random limit equilibrium method (RLEM), and the random finite-element method (RFEM) (Griffiths & Fenton, 1993; Fenton & Griffiths, 2008). There now exists a wide body of literature investigating the influence of spatial variability (in so-called uniform layers of soil) on the performance of geotechnical structures, although the practical application of probabilistic methods remains low, especially for those methods that are computationally intensive.

Examples of the use of probabilistic methods in geotechnical slope stability case histories involving the influence of spatial variability include: Alonso (1976), who used the method of slices to study the Green Creek slide in sensitive clay; El-Ramly *et al.* (2002, 2003, 2005), who used limit equilibrium methods to analyse a series of case histories; and Cho (2007) and Cami *et al.* (2018), who used RLEM to analyse the stability of the Sugar Creek embankment using one-dimensional (1D) and two-dimensional (2D) random fields, respectively. Hicks & Onisiphorou (2005) used RFEM to investigate the influence of spatial variability of the state parameter (Been & Jefferies, 1985) on the liquefaction potential of the Nerlerk underwater berm (Hicks & Boughrarou, 1998), demonstrating that it was possible for a predominantly dilative fill to liquefy due to failure through deposition-induced, semi-continuous weak zones. Recently, Hicks *et al.* (2019) used RFEM to assess the stability of an existing dyke in Starnmeer, North Holland, that was not meeting design safety requirements according to existing assessment methods. It was demonstrated that a consideration of the spatial nature of the variability not only led to a more realistic and less pessimistic safety assessment of a dyke that had remained stable for hundreds of years, but also to more economical and environmentally less intrusive mitigation measures.

This paper reports a unique opportunity to validate advanced probabilistic techniques by analysing a case history involving the controlled failure of an historic dyke in the Netherlands (de Gast, 2020). It includes a description of the site investigation prior to the failure test, brief details of the test itself, the evaluation of material property statistics including spatial correlation scales, and RFEM analyses of the dyke before the test and in the final stage of the test leading up to failure. The results of the analyses are consistent with the field observations and recorded measurements. It is demonstrated that the quantification of uncertainty associated with spatial variability is beneficial to an objective approach to slope stability assessment and design. Other sources of uncertainty (e.g. measurement, statistical, model) have not been studied in this paper, but may be included in a more general framework characterising total uncertainty (van den Eijnden & Hicks, 2019).

BACKGROUND TO THE CASE HISTORY

There is a continuous need to maintain and improve the geotechnical safety of dykes in delta regions around the world. In the Netherlands, around €1 billion per year are required to maintain and upgrade the dyke network, which protects around 40% of the Netherlands from inundation. To reduce the risk of flooding, 18 000 km of dykes are assessed at regular intervals, of which 14 000 km are classified as

Manuscript received 20 March 2019; revised manuscript accepted 8 June 2020.
Discussion on this paper is welcomed by the editor.
* Geo-Engineering Section, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, Netherlands.

regional dykes and do not protect against flooding from major rivers or the sea. The methods of assessing (and maintaining) regional dykes are strongly intertwined with the methods of assessing (and research related to) primary dykes; however, they are a very different type of structure, due to the lower risk levels and significantly shorter lengths. Both types of dyke are often founded on soft soils, but, whereas primary dykes are mainly engineered of selected sand or clay, regional dykes may be constructed of locally found clays, peats, debris and occasionally sand. Moreover, the hydraulic boundary conditions are different: primary dykes are designed to withstand tidal, storm, high-water and wave loadings, whereas regional dykes are characterised by artificially controlled high water tables, with, in general, only 10–20 cm change in external water level and only small wind waves due to the limited extent of open water areas.

The current assessment criteria for primary dykes in Dutch norms/guidelines were initially developed after the devastating storm surge of 1953 and subsequently evolved, starting with the probabilistic assessment of water heights and global factors of safety for slope stability, through to the adoption of statistical methods that enable the use of partial factors in dyke assessments. Partial factors allow for the inclusion of improved and more detailed knowledge, which limits the uncertainties and the level of acceptable risk in the calculation. The start of regional assessment of dykes started after a regional dyke failed in 1960; an overview of how the required factors of safety for regional dykes has changed over the past 50 years was given by de Gast *et al.* (2015), who highlighted the continued debate on the best approach. In particular, Hicks *et al.* (2019) emphasise that the partial factor approach used in the Netherlands includes conservative estimates of the strength of the material and has led to the calculation of very low factors of safety (as low as 0·5) for existing embankments, some of which have remained standing for hundreds of years.

In 2014, the Netherlands Organisation for Scientific Research (NWO) approved funding for the project 'Reliable dykes', with a view to developing new and improved geomechanical assessment tools for regional dykes. A central part of this project, which was also supported by the Foundation for Applied Water Research (STOWA), Dutch water boards and Dutch provinces, among others, was a full-scale field test for validation purposes.

LEENDERT DE BOERSPOLDER FIELD TEST

Leendert de Boerspolder was a small polder in South Holland, located south of one of the more economically important polders of the Netherlands, the Haarlemmermeerpolder, where Schiphol airport is situated. The Rijnland water board decided to flood Leendert de Boerspolder in 2015, in order to comply with the European water framework directive, which stipulates that when water storage capacity is removed (in this instance, by the extension and maintenance of flood defence structures at other locations), water storage capacity has to be added elsewhere. This provided the rare opportunity to carry out the controlled failure of an historic dyke. The design, implementation and evaluation of the failure test is described by de Gast (2020).

Figure 1 shows a cross-section through the dyke at the location of the failure test. The crest of the embankment was at Normaal Amsterdams Peil (NAP) −0·4 m and the depth of the polder varied between NAP −1·9 m and NAP −2·1 m, in which NAP is the Dutch national reference level. The body of water south of the polder (the Hanepoel) has a fixed level of NAP −0·6 m, and over the course of the year this level varies ±0·05 m. In the polder, the water level was maintained at NAP −2·45 m in a series of ditches (not present at the failure test location).

Site investigation

Two site investigations were conducted at Leendert de Boerspolder as part of the Reliable dykes project: at one location, 100 cone penetration tests with pore pressure measurement (CPTus) and six borings were undertaken as part of a detailed study on soil heterogeneity not reported in this paper (de Gast *et al.*, 2017, 2019, 2020); the second series of tests was conducted prior to, and at the location of, the failure test and the tests were used in the analyses presented herein. Fig. 2 shows a plan view of the second site investigation, which included 17 CPTus, meeting the specifications of NEN-ISO 22476-1 (Class 1), and one ball CPT. These penetrated as far as a deep underlying sand layer at NAP −16 m (not included in any of the cross-sections analysed) and were arranged in three rows of either five or six CPTs aligned perpendicularly to the dyke at 20 m intervals; they are denoted as the north, centre and south rows in the figure. In each row, from the polder to the dyke crest the spacing between the CPTs was (in general) 2·5 m, and the last CPT was located in the canal and undertaken from a pontoon 10 m from the crest. Equidistant between the rows of CPTs were two rows of three borings, made using a 100 mm dia. piston sampler and providing semi-continuous samples in 70 cm sections. Inclinometers were installed to measure the horizontal displacement in line with the centre row CPTs and close to the borings equidistant between the rows of CPTs.

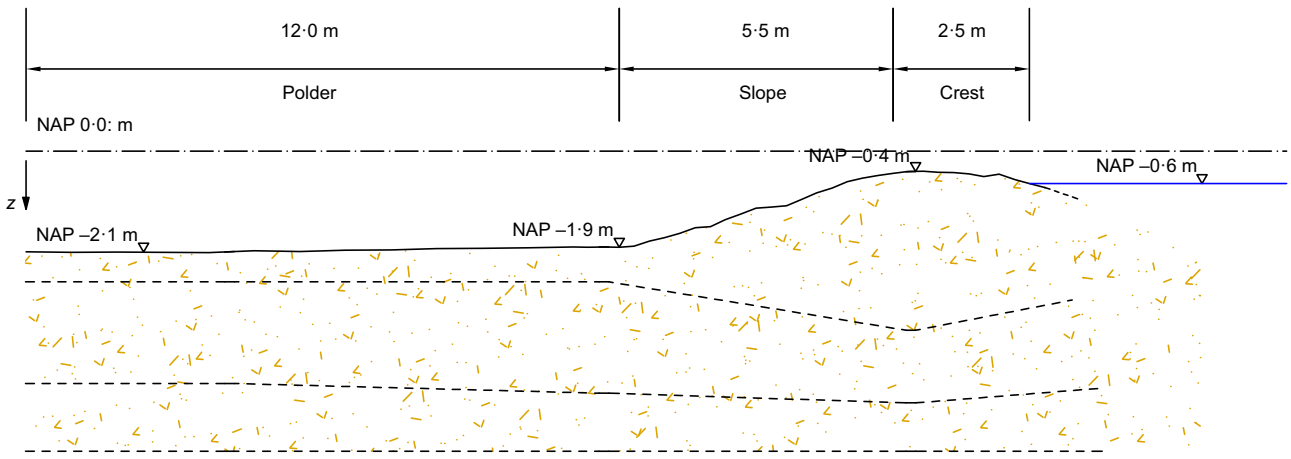

Fig. 1. Geometry of Leendert de Boerspolder (centre cross-section)

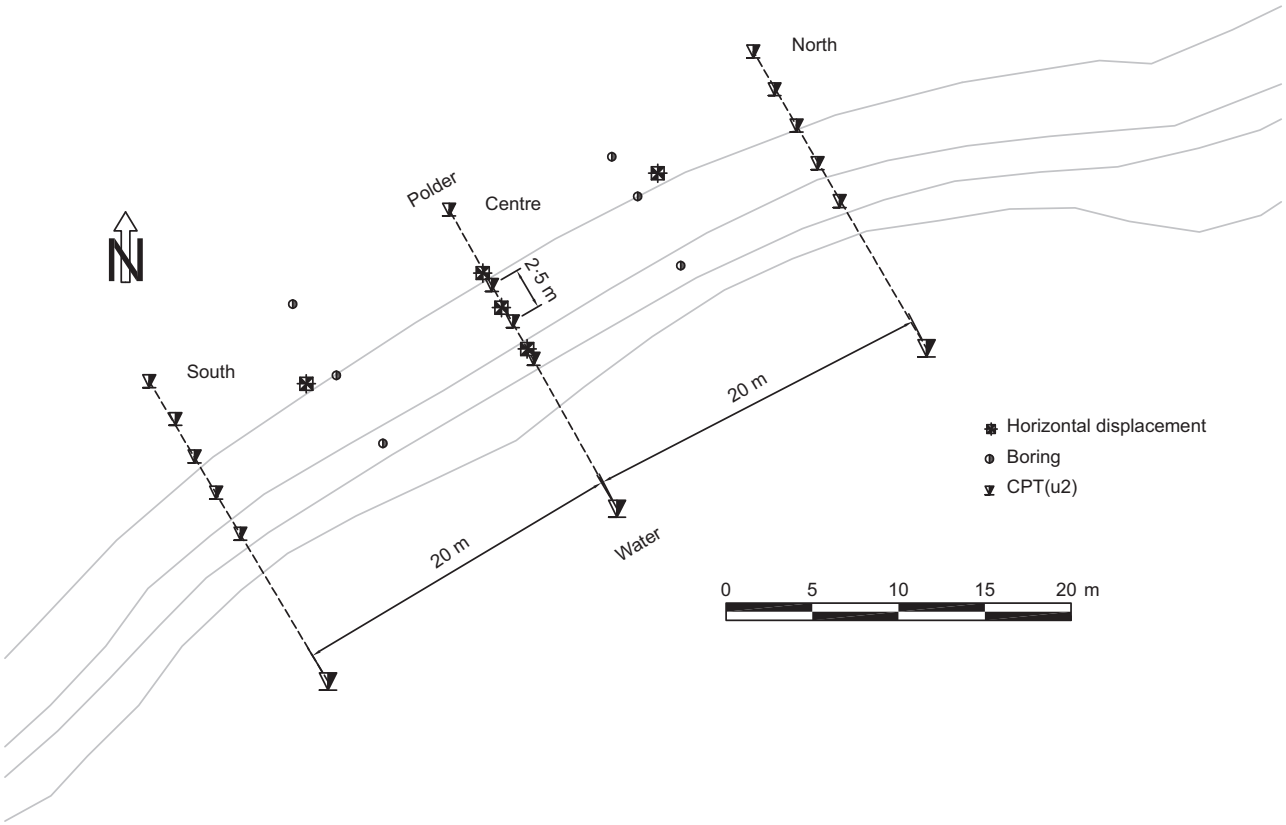

Fig. 2. Plan view of site investigation: three lines, each with four or five closely spaced CPTs on land and one CPT in the water, and two lines of three semi-continuous borings in the crest, slope and polder. From left to right, these lines are south, centre and north

Based on the CPT measurement data (i.e. cone tip resistance, sleeve friction and water pressure behind the cone), and the borehole data, the following four major material layers were identified, starting from the ground surface.

(*a*) Dyke material: this had been placed over time, since ~1600 AD, initially for the original construction and subsequently by adding to the dyke periodically for maintenance.
(*b*) Peat: this layer has been affected by the overlying dyke material and, in the polder, by constant dewatering. The thickness of the layer varies from 1·0–1·8 m under the dyke to 1·8–2·2 m in the polder.
(*c*) Organic clay: the layer starts with a high organic fraction at the top, decreasing with depth. Its thickness is 1·6–1·8 m under the dyke to 2·4–2·5 m in the polder.
(*d*) Silty clay: as the organic content decreases, the silt fraction increases and the clay layer continues until ~NAP −16 m, beyond which sand is found.

A cross-section through the stratigraphy detected by each of the three CPT rows is shown in Fig. 3, as well as a single

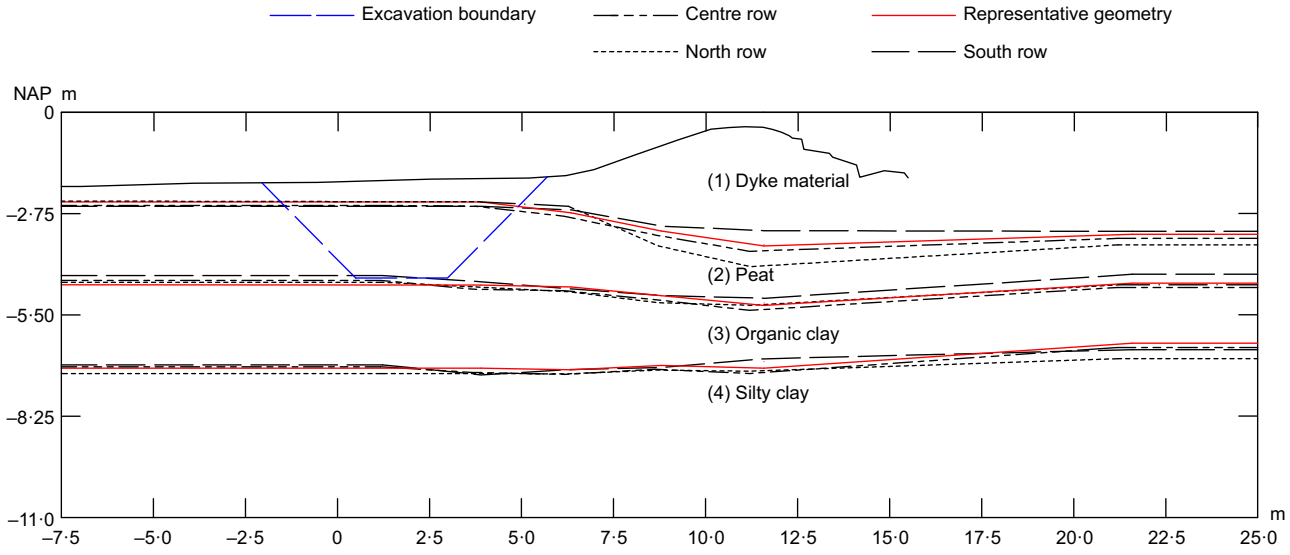

Fig. 3. Stratigraphy interpretation from the three rows of CPTs

Table 1. Derived point statistics of undrained shear strength (in kPa)

CPTs used:	All		North row		Centre row		South row	
	$s_{u;\mu}$	$s_{u;\sigma}$	$s_{u;\mu}$	$s_{u;\sigma}$	$s_{u;\mu}$	$s_{u;\sigma}$	$s_{u;\mu}$	$s_{u;\sigma}$
Dyke material	17·69	12·98	18·89	11·63	14·78	10·05	19·46	16·36
Peat	4·88 −1·90z	5·07	8·76 −1·17z	5·09	3·39 −2·46z	3·57	4·94 −1·45z	5·61
Organic clay	−8·32 −4·43z	5·22	3·21 −2·64z	4·93	−17·26 −6·25z	6·27	−6·95 −3·78z	2·97
Silty clay	−5·68 −3·62z	4·02	−3·10 −3·61z	4·72	−7·11 −3·80z	2·63	−6·08 −3·37z	2·25

Note: z is the location relative to NAP.

representative stratigraphy. All four stratigraphies have been considered in the numerical analyses.

Failure test

The dyke was saturated with water by using a water sprinkler system that was left on continuously for a period of 1 month, and then the soil in front of the toe was excavated in stages and replaced by water, effectively increasing the height of the dyke (de Gast, 2020). The excavation was performed by a global positioning system (GPS) guided excavator, ensuring centimetre-level precision. Eventually, the ditch was excavated to the bottom of the peat layer (i.e. NAP −4·4 m), at 2·5 m below the ground surface (NAP −1·9 m). Then, in the final stage of the experiment, the water in the excavation was removed and the dyke failed under its own weight, with the main failure occurring just south of the centre of the excavation. Large differential displacements were measured in the toe, and in the organic clay layer just below the boundary between the peat and organic clay. The failure occurred at a drawdown somewhere between 1·5 m and 2·0 m (i.e. between NAP −3·4 m and NAP −3·9 m), and was estimated to be at 1·6 m (i.e. NAP −3·5 m).

MATERIAL PROPERTIES

The point and spatial statistics of undrained shear strength were determined using the 17 CPTs (Fig. 2), as well as data from laboratory tests on the borehole samples.

The undrained shear strength profiles for the soil layers were determined from the CPT data using the relation (Robertson, 2009)

$$s_u = \frac{q_t - \sigma_v}{N_{kt}} \tag{1}$$

where q_t is the total cone resistance; σ_v is the total vertical stress; and N_{kt} is an empirical correction factor. Robertson (2009) suggested values for N_{kt} in the range 10–20. The procedure used here to estimate the N_{kt} values was as follows (de Gast, 2020): (*a*) the laboratory data, comprising 20 consolidated triaxial compression tests and four direct simple shear tests (Ponzoni, 2017; Muraro, 2019), were used to determine the maximum shear strength as a function of depth; (*b*) the mean and standard deviation of the q_t data were determined for each 0·25 m depth interval; (*c*) equation (1) was used to calculate an N_{kt} value for each material, so that the s_u determined from the CPT data best fitted the laboratory data. The values of N_{kt} determined for the four soil layers were: dyke material, $N_{kt} = 20$; peat, $N_{kt} = 15$; organic clay, $N_{kt} = 10$; and silty clay, $N_{kt} = 10$.

Using the derived values of N_{kt}, the CPT data were transformed to s_u profiles, and the mean (or mean trend) and standard deviation of s_u for each soil layer were then found. Table 1 summarises the point statistics based on all CPT profiles, as well as those based on the three CPT rows individually. The mean trend was computed for the peat and both clay layers, but no clear trend was identified for the dyke material. The standard deviation was calculated relative to the depth-dependent mean for the peat and clay layers. Fig. 4 shows the CPT profiles for s_u, which are plotted with respect to local horizontal axes (with the scale shown representing 0–10 kPa). The figure shows how the shear strength varies with respect to depth, for each CPT row and for each material, especially for the clay layers. When approximating the distribution of s_u data for each soil layer, a normal distribution relative to the depth-dependent mean was adopted for the peat and clay layers, whereas a lognormal probability density function was found to be the best fit for the dyke material. Note that a small number of extreme data values, for example, as are apparent for the centre row CPTs in Fig. 4(b), were considered anomalous and not included in the derived statistical values listed in Table 1.

Table 2 summarises the computed vertical and horizontal scales of fluctuation for the four soil layers. These were obtained by minimising, in both the vertical and horizontal directions, the squared difference ($E(\theta)$) between the experimental autocorrelation function ($\hat{\rho}(\tau)$) obtained from the CPT data and a theoretical (Markov) autocorrelation function ($\rho(\tau) = \exp(-2|\tau|/\theta)$), as shown in Figs 5(a)–5(b) (de Gast, 2020). Specifically, the error to be minimised is

$$E(\theta) = \sum_i [\rho(\tau_i) - \hat{\rho}(\tau_i)]^2 \tag{2}$$

where τ_i are all available lag distances for which data are available and $\hat{\rho}(\tau)$ is given by

$$\hat{\rho}(\tau) = \frac{\hat{\gamma}(\tau)}{\hat{\gamma}(0)} \tag{3}$$

where $\hat{\gamma}(\tau)$ is the experimental covariance function and $\hat{\gamma}(0)$ is the experimental covariance function when $\tau = 0$ (i.e. the point variance). For unequally spaced data, this is given by (Vanmarcke, 1983)

$$\hat{\gamma}(\tau) = \frac{1}{t-1} \sum_{j=1}^{t} \left(y_j - \hat{\mu}\right)\left(y_j' - \hat{\mu}\right) \tag{4}$$

where y_j and y_j' are two data at locations separated by lag distance τ; $\hat{\mu}$ is the estimated mean (or trend) of the dataset; and $j = 1,2,\ldots,t$ is a counter representing the number of pairs of data at lag distance τ. In this paper, for the purpose of deriving the experimental autocorrelation functions, and thereby θ_v and θ_h, the CPT data were first de-trended (i.e. with respect to the actual depth-dependent mean). Table 2 shows the scales of fluctuation to be small (0·3–0·8 m) in the vertical direction, as has generally been found by other researchers for other soils, and as found for the same soil layers at the adjacent test site at Leendert de Boerspolder (de Gast, 2020; de Gast *et al.*, 2020). The scales of fluctuation

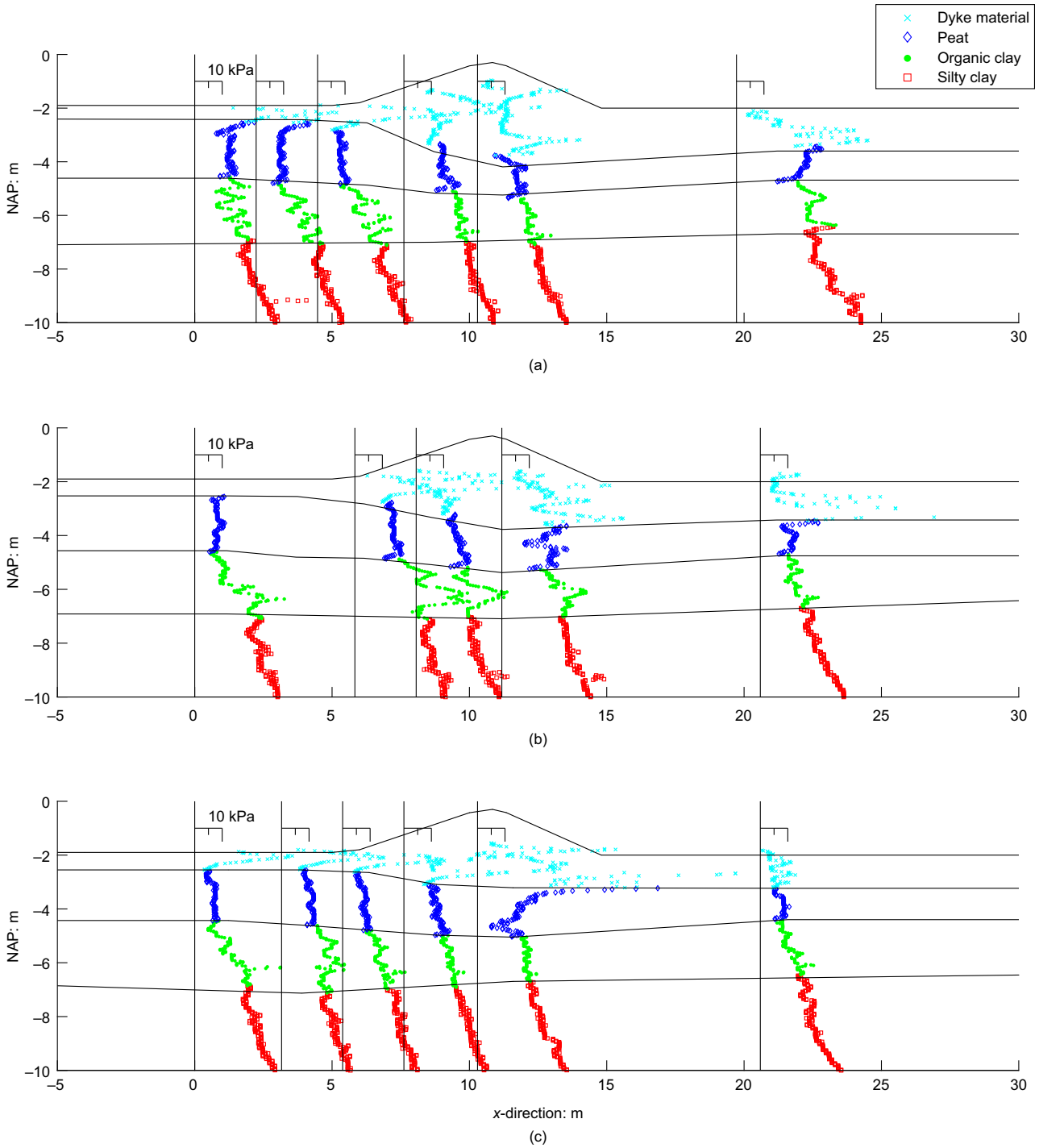

Fig. 4. Input data for conditioning random fields. Three rows of CPTs with the derived s_u values per (coloured) soil layer: (a) north row; (b) centre row; (c) south row

Table 2. Derived scales of fluctuation

Soil type	θ_v: m	θ_h: m
Dyke material	0·40	2·13
Peat	0·76	2·84
Organic clay	0·76	2·84
Silty clay	0·26	2·10

in the horizontal direction were found to be 4–8 times larger than the vertical scales of fluctuation. However, it should be noted that θ_h was determined along the CPT rows perpendicular to the dyke (Fig. 2), with the CPTs in each row positioned at 2·5 m centres on average. As the derived values of θ_h were similar in magnitude to the CPT spacing, there is some doubt as to their accuracy, although they are generally consistent with values obtained at the adjacent test site, which used CPT spacings of only 1·25 m (de Gast *et al.*, 2017, 2019, 2020). Moreover, it seems reasonable that θ_h would be at least several times greater than θ_v, due to the natural process of deposition, and there is confidence in the derived values of θ_v owing to the large number of closely spaced data available to construct the experimental autocorrelation functions in the vertical direction.

The estimation of the spatial statistics for the organic clay layer using equation (4) is illustrated in Figs 5(a) and 5(b). The experimental autocorrelation function in the vertical direction (Fig. 5(a)) is equivalent to the average of 17 experimental functions, as obtained from each of the 17 CPTs. In contrast, the experimental autocorrelation function in the horizontal direction (Fig. 5(b)) is equivalent to the average of a larger number of experimental functions, with each one corresponding to a different depth interval but based on the number of CPTs in the row. Fig. 5 illustrates how the experimental autocorrelation function is less reliable at larger lag lengths, owing to the fewer pairs of data then available for input into equation (4).

The unit weights for the four soil layers are (Ponzoni, 2017): dyke material, $\gamma_{sat} = 18$ kN/m^3; peat, $\gamma_{sat} = 10$ kN/m^3; organic clay, $\gamma_{sat} = 15$ kN/m^3; and silty clay, $\gamma_{sat} = 17$ kN/m^3. No information on the variability of the unit weights was determined, although previous researchers have generally found the coefficient of variation to be low and typically in the range 0·05–0·1 (Phoon & Kulhawy, 1999). In any case, any variability in the unit weights should have a minimal impact on the numerical analyses presented for two reasons: (*a*) the analyses are total stress, so that the soil resistance is independent of the unit weight; and (*b*) the scales of fluctuation are small relative to the geometry of the sliding mass, so there will be significant spatial averaging of unit weights and thereby minimal influence on the overturning moment.

NUMERICAL MODELLING

The dyke has been analysed using RFEM (Griffiths & Fenton, 1993; Fenton & Griffiths, 2008). In this method, multiple realisations of the boundary value problem are analysed by the finite-element method, with the spatial variability of material properties being modelled by different random fields in each realisation. These fields are based on an assumed probability density function and the point statistics (mean μ and standard deviation σ) of each spatially varying soil property, and by an assumed covariance function and the spatial correlation statistics (i.e. the scales of fluctuation in the vertical and horizontal directions, θ_v and θ_h, respectively, which represent the spatial scales over which the soil property values are significantly correlated). For a given set of point and spatial soil property statistics, RFEM computes an ensemble of responses for the boundary value problem (based on the different random fields), leading to a reliability-based safety assessment (as opposed to a single factor of safety).

One advantage of using the finite-element method to analyse the boundary value problem is that no prior

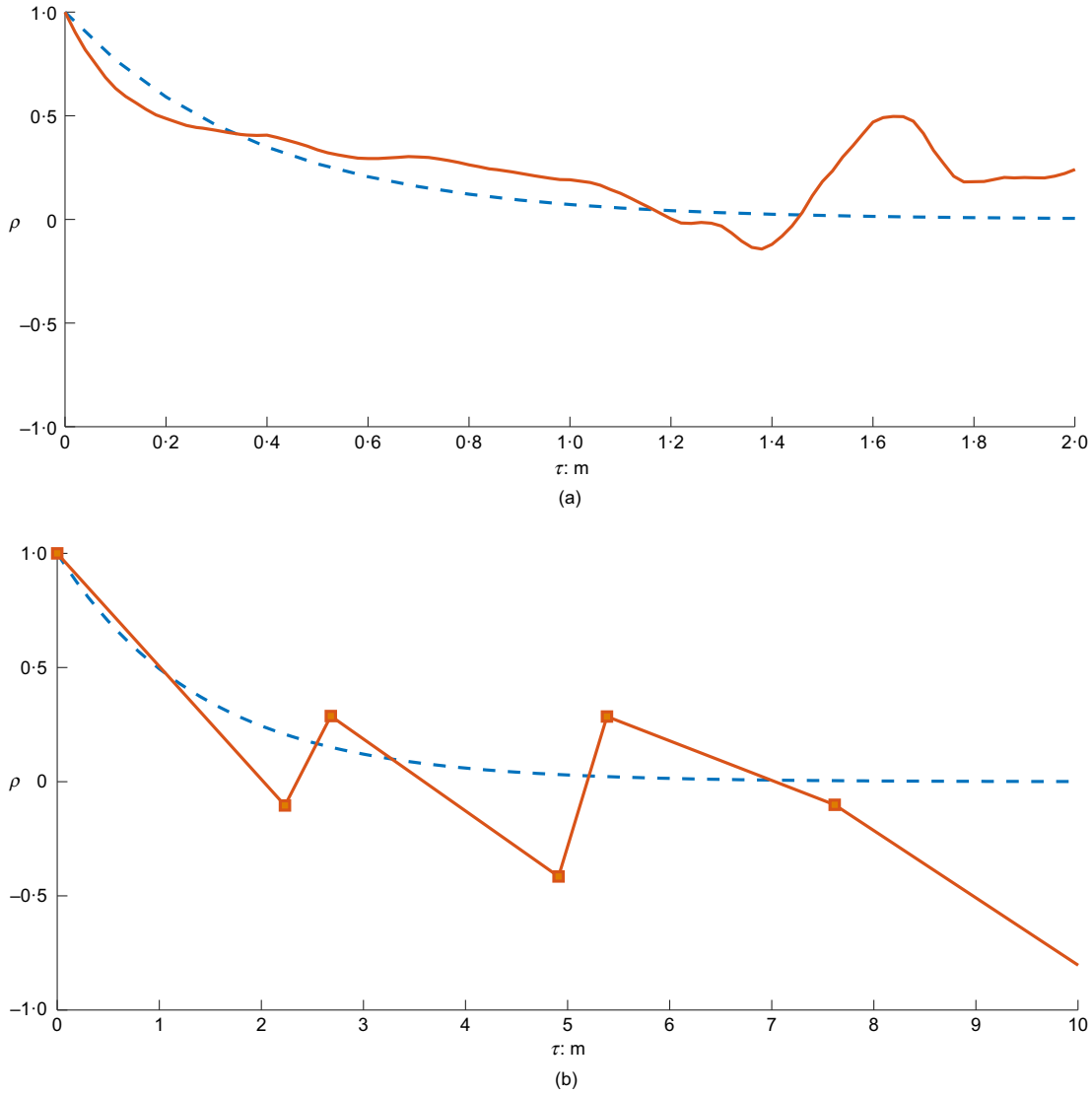

Fig. 5. Derivation of point and spatial statistics for organic clay layer: (a) vertical autocorrelation functions (experimental, solid line; theoretical, broken line); (b) horizontal autocorrelation functions (experimental, solid line; theoretical, broken line)

assumptions are needed regarding the failure mechanism geometry or its location (Hicks & Boughrarou, 1998), in contrast to the simpler limit equilibrium methods, such as those based on the traditional method of slices. This is particularly relevant when accounting for the effects of soil spatial variability, as failure mechanisms can take on many forms when attracted to paths of least resistance (Hicks & Samy, 2002; Hicks & Spencer, 2010; Hicks *et al.*, 2014). Although more advanced limit equilibrium methods can analyse more general mechanisms (Cho, 2007; Javankhoshdel *et al.*, 2017), finite elements are also able to account for more realistic soil constitutive behaviour (Hicks & Onisiphorou, 2005). An obvious disadvantage of RFEM, especially when considering problems in three dimensions, is that it is computationally expensive, although recent advances in computer power (e.g. through grid and cloud computing) are making such analyses more accessible (Li *et al.*, 2015; Hicks & Li, 2018).

In this paper, the spatial variation of material properties has been modelled by random fields generated using covariance matrix decomposition with local averaging for unstructured meshes (van den Eijnden & Hicks, 2017a). The method starts by generating a random field with values at locations matching the Gauss points of the finite-element spatial discretisation, based on a standard normal distribution and a spatial correlation function incorporating the scales of fluctuation. The standard normal field is then transformed to the appropriate distribution using the marginal distribution of the property being modelled. In any realisation, each soil layer is modelled by a separate random field characterised by the soil property statistics for that layer. Figs 6(a)–6(d) show typical random field

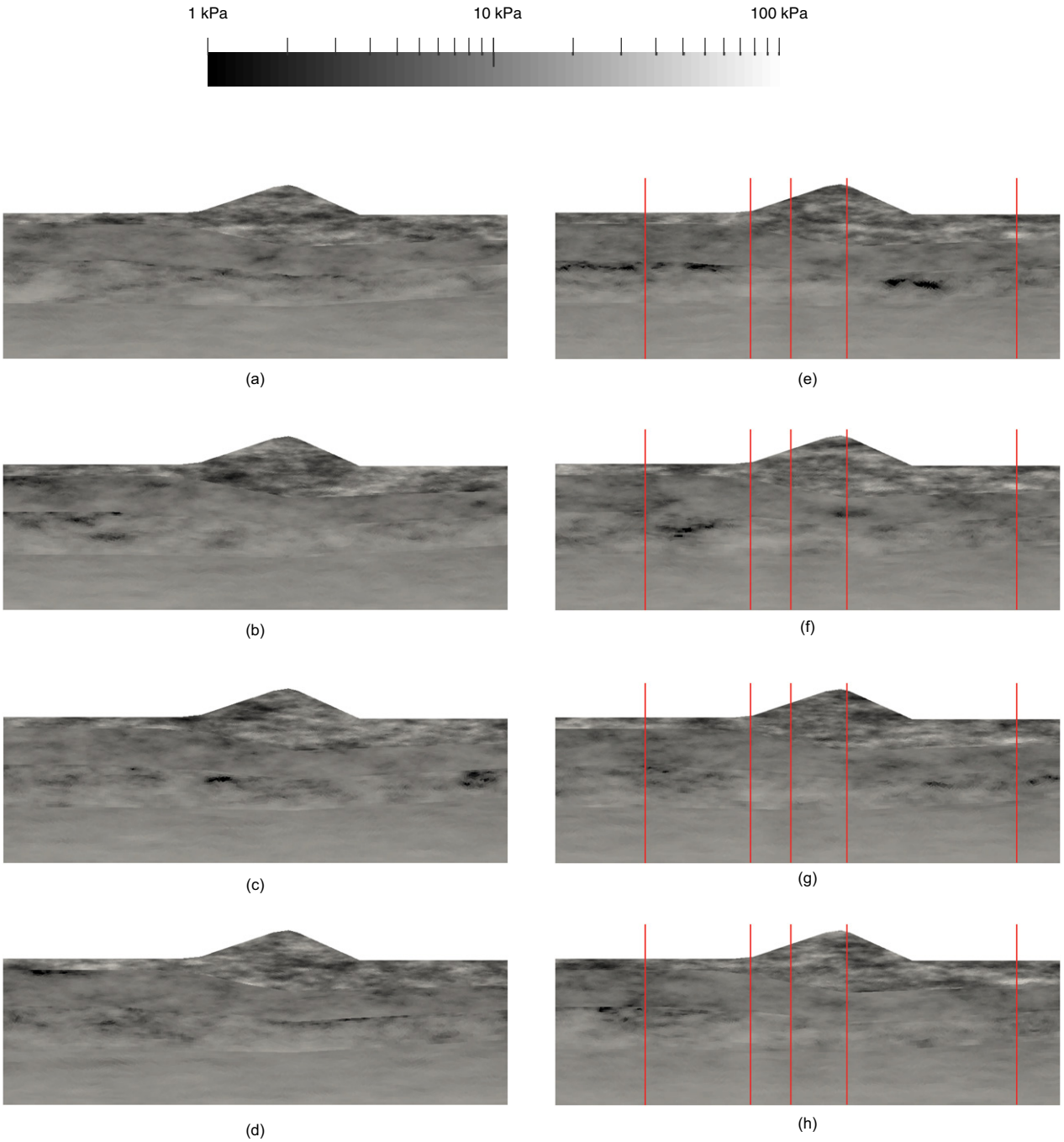

Fig. 6. Typical random fields of undrained shear strength for centre row (in kPa on a logarithmic scale): (a)–(d) unconditional random fields; (e)–(h) conditional random fields, conditioned to centre row CPTs at locations indicated by vertical lines

realisations of shear strength for the dyke (before excavation of the ditch), in which a log scale has been adopted to improve the clarity of the visualisations. For any given soil layer, the spatial distributions of shear strength are statistically similar (but spatially different) across the realisations, because they are based on the same point and spatial statistics in each realisation. Figs 6(e)–6(h) show typical realisations in which the random fields are conditioned to measurement data (i.e. so that the random field values match the CPT data (van den Eijnden & Hicks, 2017b)) at the centre row CPT locations (Fig. 2); hence, because the properties at the CPT locations are the same in each realisation, there is less uncertainty in the spatial variability (Lloret-Cabot *et al.*, 2012; Li *et al.*, 2016).

Various cross-sections through the dyke failure test have been analysed in plane strain. Fig. 7(a) shows the finite-element mesh and problem geometry based on the north row stratigraphy in Fig. 3, in which the boundaries of the domain are located far enough from the dyke so as not to influence the results. The bottom boundary is fixed, whereas the vertical boundaries allow only vertical displacement, and, as only the stability of the dyke is of interest, only one side of the excavated ditch has been modelled. On each side of the dyke an external hydrostatic load was applied representing the water load: on the canal side the water level was at NAP −0·6 m, whereas on the polder side the water level in the excavation depended on the particular stage of the test being modelled. The finite-element mesh comprised 4388 eight-node quadrilateral elements, with each element using 2×2 Gaussian integration. Similar finite-element meshes were generated for analyses involving the stratigraphy encountered at the centre and south rows, as well as for analyses involving the representative stratigraphy at the site (Fig. 3).

In this paper, each RFEM analysis has involved 1000 realisations, and, in each realisation of the RFEM analysis, the factor of safety of the dyke has been computed using the strength reduction method. Specifically, the dyke has been repeatedly analysed, by generating the in situ stresses due to gravity loading and external water loading, and by sequentially scaling down the soil shear strength, with the factor of

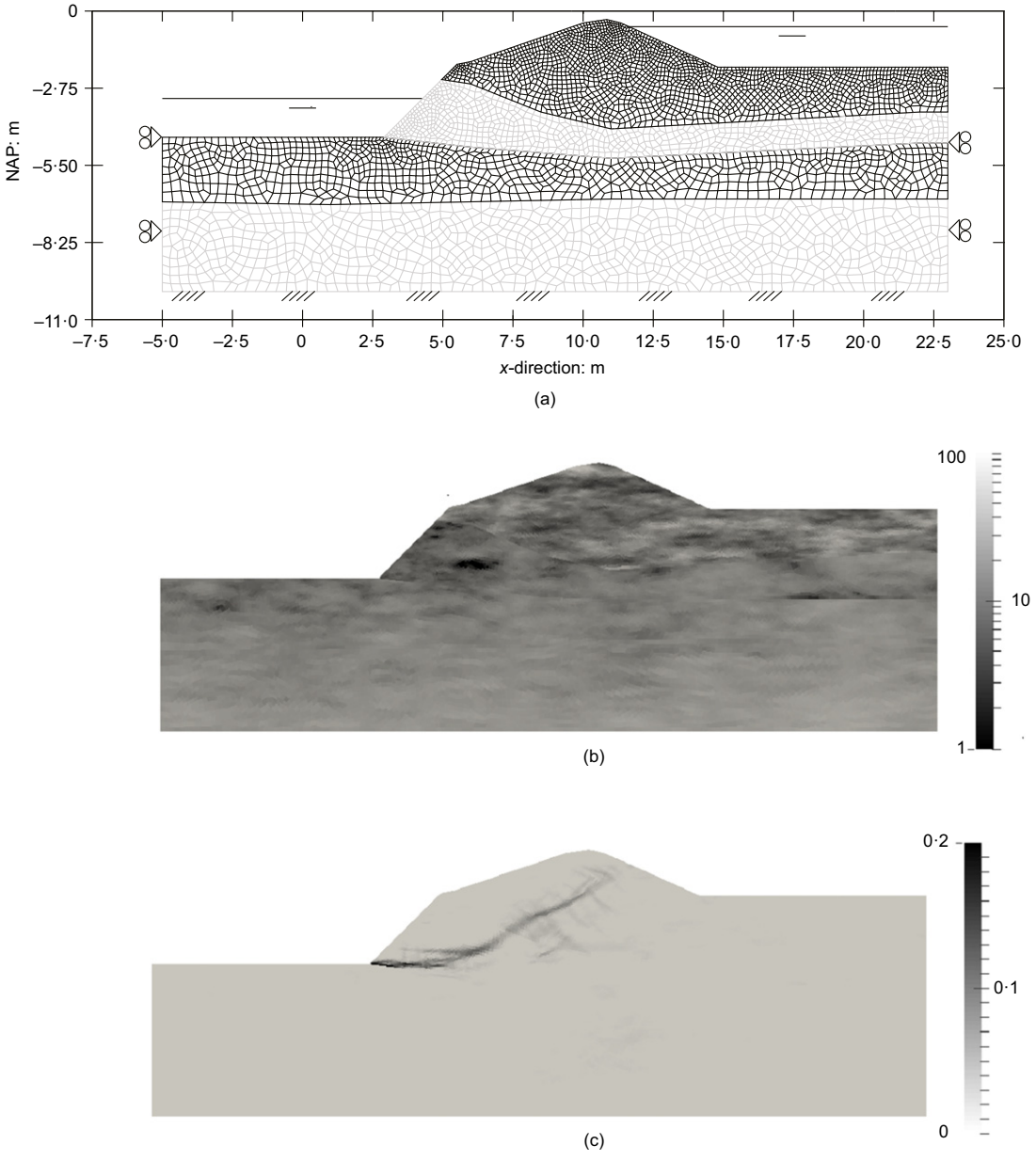

Fig. 7. Analysis of embankment after ditch excavation: (a) finite-element mesh for north row stratigraphy; (b) example random field of undrained shear strength (in kPa on a logarithmic scale); (c) contours of deviatoric strain at slope failure

safety being the smallest scaling factor required to bring the dyke to failure under its own weight. Each realisation typically took 15 min on a standard personal computer (PC); however, grid computing was used to enable many realisations to be undertaken at the same time. Figs 7(b) and 7(c) show a typical undrained shear strength distribution before strength reduction and the resulting failure mechanism after strength reduction, respectively, for a single RFEM realisation using the finite-element mesh in Fig. 7(a). A total stress analysis has been performed, using a linear elastic, perfectly plastic Tresca soil model and material properties based on the data presented by de Gast (2020). However, only the shear strength has been taken as spatially random, while other parameters have been assumed constant for each layer. The shear strength profile for each layer has been defined by either a lognormal or normal probability density function (characterised by a mean $s_{u,\mu}(z)$ and standard deviation $s_{u,\sigma}$, see Table 1, where z is the location relative to NAP), and by a Markov covariance function (characterised by θ_v and θ_h, see Table 2).

Table 3 summarises the analyses carried out, in which two stages of the test have been investigated in detail: the initial condition, before any excavation, and the final stage leading up to the dyke failure, in which the water level in the excavated ditch was gradually lowered until failure at a water level in the range of NAP from −3·4 m to NAP −3·9 m. For each stage and/or water level considered, up to four RFEM analyses have been performed: one based on unconditional random fields, and three based on random fields conditioned against one of the three CPT rows shown in Fig. 2 (north, centre or south). In addition, deterministic analyses based on single 'representative' strengths (i.e. mean, median and five-percentile s_u values), for the individual soil layers, have been carried out for comparative purposes for the initial condition.

ANALYSES

In each RFEM analysis, 1000 realisations have been performed to obtain a distribution of possible factors of safety (F). For a typical RFEM analysis, Fig. 8 shows how the mean and standard deviation of F evolve as more realisations are analysed. In this figure, the mean and standard deviation of F have been normalised by their respective values after 1000 realisations. It is seen that 1000 realisations are more than enough to achieve sufficient convergence of the output statistics for interpreting the results.

Reliability at initial conditions

Figure 9 presents the results of RFEM and deterministic analyses carried out for the initial condition; that is, before excavation of the ditch had commenced. This was analysed using a comparable finite-element mesh to Fig. 7(a), but with a horizontal ground surface in place of the ditch. For each RFEM analysis, the results are presented as (*a*) a cumulative

Table 3. Overview of RFEM and deterministic analyses

Geometry	Water level: m to NAP	RFEM analysis conditioned by CPT row?				Deterministic
		Unconditional	North	Centre	South	
Initial	—	✓(All*), ✓(Centre*)	—	✓	—	✓(Mean, med., 5%*)
Ditch	−1·9	✓(All*)	✓	✓	✓	—
Ditch	−2·4	✓(All*)	✓	✓	✓	—
Ditch	−2·9	✓(All*)	✓	✓	✓	—
Ditch	−3·4	✓(All*)	✓	✓	✓	—
Ditch	−3·9	✓(All*)	✓	✓	✓	—

*Data used in the unconditional analysis.

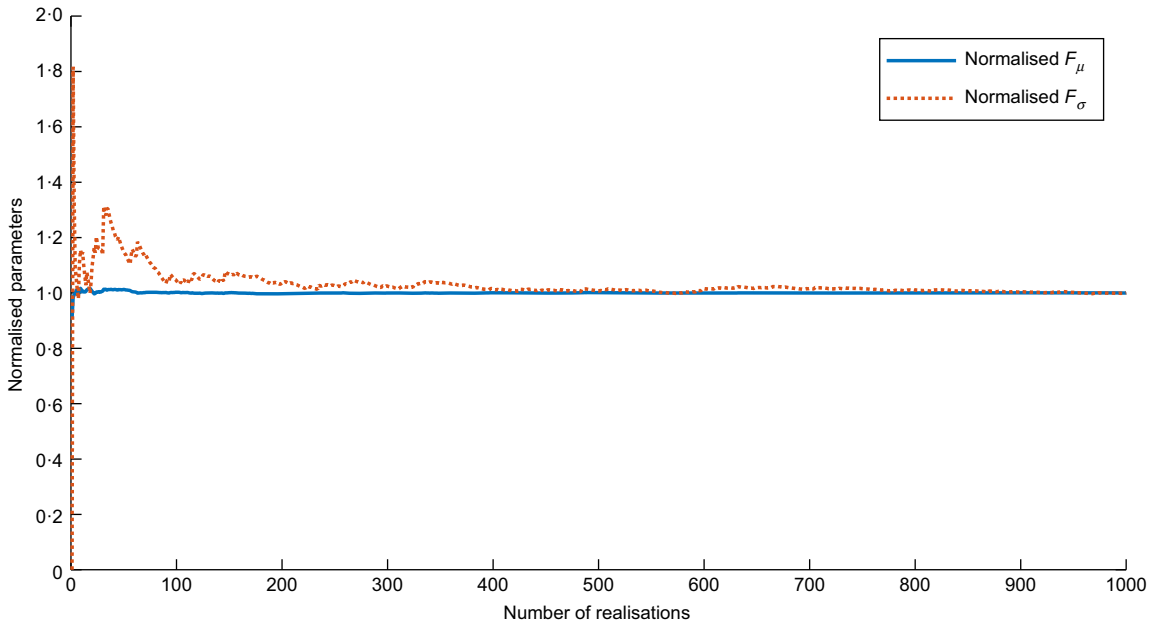

Fig. 8. Illustration of convergence of mean and standard deviation of F in typical RFEM analysis

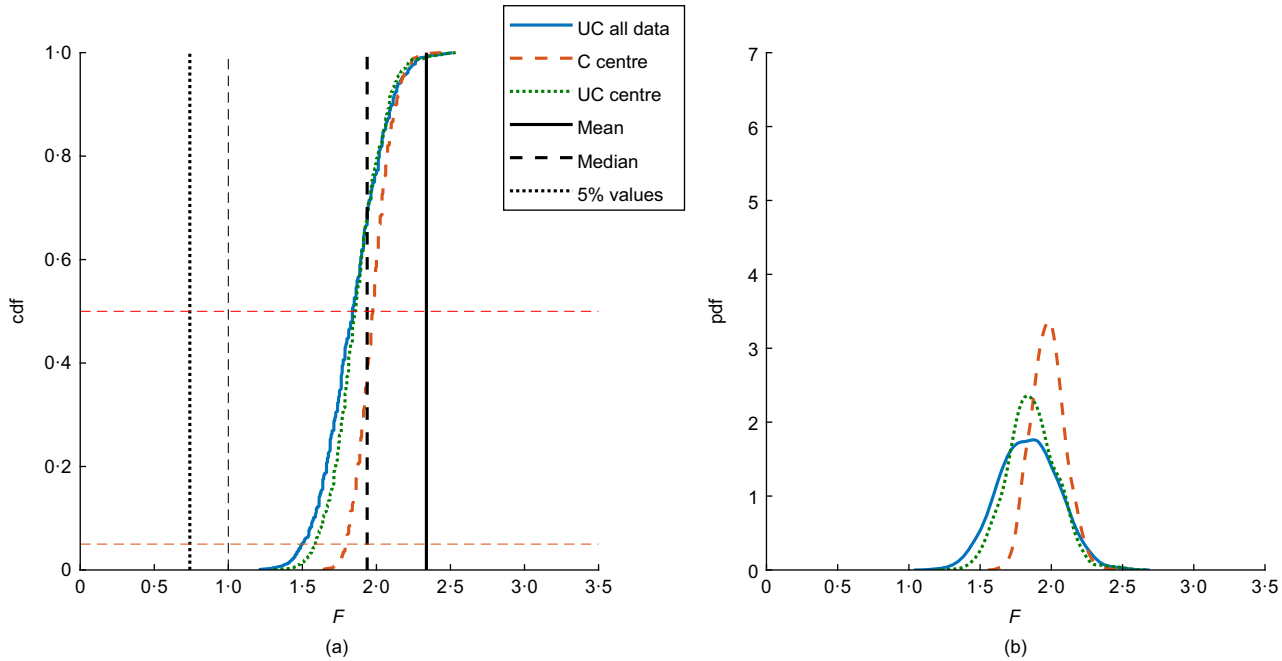

Fig. 9. Comparison of deterministic and RFEM analyses for initial condition of dyke (prior to excavation of the ditch): (a) cdf; (b) pdf

distribution function (cdf) and (*b*) a probability density function (pdf). A pdf defines the range of possible factors of safety and their relative likelihood, whereas a cdf is the cumulative area under the pdf. In Fig. 9, the probability of failure predicted by each RFEM analysis is the area under the pdf for $F < 1$, and the reliability is the area under the pdf for $F \geq 1$. Alternatively, the probability of failure is equal to the cdf when $F = 1$, whereas the reliability is equal to (1 − cdf) at $F = 1$. As one would expect for a slope that has remained standing for centuries, the computed probability of failure is almost 0% for the initial condition of the test.

Figure 9 compares the results of three RFEM simulations: one based on the statistics obtained from all 17 CPTs but without conditioning of the random fields to the CPT data ('UC all data'); and two further simulations based only on the statistics obtained from the centre row CPTs, one using random fields conditioned to the centre row CPTs ('C centre') and the other using unconditional random fields ('UC centre'). The figure shows that there is a relatively small difference between the unconditional distributions of factor of safety obtained using statistics of the material parameters derived from only the centre row CPTs and those derived from all CPTs, and that, for $R = 95\%$ (given by cdf = 0·05), the respective values of F are 1·59 and 1·50. Moreover, $F = 1{\cdot}50$ may be compared with factors of safety of 0·74, 1·94 and 2·34 from deterministic analyses based on the five-percentiles, medians and means of the distributions of undrained shear strength for each soil layer, respectively. It is seen that, in almost all cases, the RFEM realisations return a much lower factor of safety than that computed based on mean strengths, due to failure being attracted to weaker zones and avoiding (where possible) the stronger zones; indeed, Fig. 9 shows the deterministic solution based on mean strengths to be the upper-bound solution in this instance. Conversely, the RFEM realisations generally return a much higher factor of safety than that computed based on the five-percentile strengths, due to the averaging of property values along potential failure planes.

Figure 9 also shows that the uncertainty in the dyke response is reduced when conditioning the random field against the centre row CPT measurements, as indicated by the narrower range for the cdf and pdf. Specifically, for $R = 95\%$, based on the statistics derived from the centre row CPTs, F increases from 1·59 to 1·79 if the random fields are conditioned to the CPT data. Overall, it is clear from Fig. 9 that a consideration of the spatial correlation of soil properties has a significant impact on the stability assessment. Whereas assessments based on mean strengths may be unconservative, those based on five-percentiles may be significantly overconservative and lead to costly re-designs and mitigation measures.

Reliability at the final stage of failure test

Figure 10 summarises results obtained for various drawdown levels during the final stage of the field test, as obtained by RFEM analyses based on unconditional random fields and a finite-element mesh similar to that in Fig. 7(a). By the start of the final stage, the ditch had been excavated to its maximum depth of 2·5 m, with 1 in 1 side slopes, and filled with water to the brim of the excavation at a level of NAP −1·9 m. Five drawdown levels (in 0·5 m steps) have been analysed and are shown in Fig. 10, starting with NAP −1·9 m, and ending with NAP −3·9 m, corresponding to the lowest possible drawdown level at which the dyke failure occurred. Once again, the results for each analysis are shown as (*a*) a cdf and (*b*) a pdf. Considering dyke failure to be represented by a factor of safety less than 1·0, the probability of failure increases from 3·25% before the start of drawdown, to 10·50%, 25·00%, 38·75% and finally to 50·00% for drawdowns of 0·5 m, 1·0 m, 1·5 m and 2·0 m, respectively. Table 4 lists the same results for the drawdown levels relative to NAP, and compares them with the results of RFEM analyses using random fields conditioned on the individual CPT rows, using finite-element meshes consistent with the stratigraphy at each cross-section being analysed. These analyses are illustrated in Fig. 11.

Comparing the probabilities of failure for the different analyses shows a clear difference between the unconditional and conditional RFEM analyses. By conditioning the random fields based on known data at specific locations (i.e. at the CPTs), the range of possible spatial distributions of

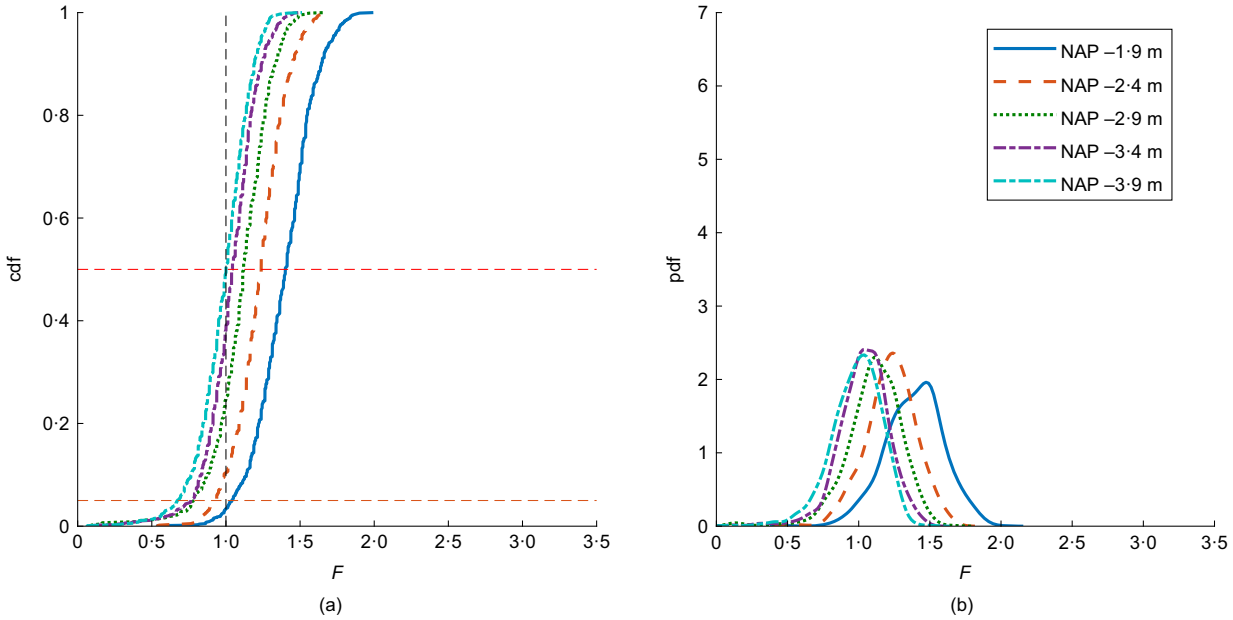

Fig. 10. Unconditional RFEM analyses for different levels of drawdown: (a) cdf; (b) pdf

Table 4. Evolution of failure probabilities (in %) as a function of drawdown level

	Drawdown level: NAP				
	−1·9 m	−2·4 m	−2·9 m	−3·4 m	−3·9 m
Unconditional	3·25	10·50	25·00	38·75	50·00
North row	<0·25	<0·25	<0·25	3·50	11·25
Centre row	<0·25	<0·25	0·50	0·60	3·60
South row	<0·25	<0·25	<0·25	0·75	2·25

undrained shear strength for the dyke cross-section reduces. This in turn leads to a smaller range of computed factors of safety when the random fields are used in the RFEM simulations, as is apparent by the factor of safety distributions being narrower in Fig. 11 than in Fig. 10. Note, however, that the narrower (conditional) distributions are completely contained within the respective wider (unconditional) distributions.

Figure 12 shows the computed failure mechanisms from the weak tail of each pdf (i.e. for those realisations for which $F < 1$), for the three conditional analyses based on the maximum possible drawdown of 2·0 m. The largest number of potential mechanisms is seen for the north cross-section (Fig. 12(a)) and the lowest number for the south cross-section (Fig. 12(e)), reflecting the relative differences in failure probability shown in Table 4. However, the mechanisms generally follow a similar trend; that is, failure tends to initiate at the bottom of the peat layer, or just inside the clay layer; it then continues just above or just below the boundary between the two layers, before turning upwards and exiting the ground surface on the outward sloping face of the dyke. This mechanism is consistent with the deterministic solution based on mean layer strengths (factored down to trigger failure), as shown by the thicker lines drawn on three cross-sections. It is also consistent with measurement data from the test itself, as recorded using inclinometers positioned at the locations indicated in Fig. 2. This included three inclinometers in-line with the centre row CPTs, plus single inclinometers between the north and centre rows, and between the centre and south rows. Fig. 12(c) includes the horizontal displacements recorded at the centre cross-section at three times: just before failure t_0 and at two times (separated by around 1 min) during the failure, t_1 and t_2. It is seen that failure occurs suddenly and involves large horizontal displacements. Moreover, the measurements indicate a similar failure surface location and geometry to the numerical analyses. Figs 12(b) and 12(d) show the displacements recorded by the inclinometers located between rows, which for convenience have been plotted relative to the stratigraphies at the north and south rows, respectively. The measurements indicate that failure was confined to the centre and southern half of the failure test, and visual observations, including a filmed recording, of the test itself confirmed that failure started between the centre and south rows. This is at variance with the RFEM results, which suggest a greater tendency to fail at the northern end of the test, although the failure probabilities in Table 4 and Fig. 11 also indicate that the three rows are in similar states as failure of the dyke is approached. A possible cause of the difference between the general tendency predicted by the RFEM results and the actual location of the failure is the spatial variability of soil properties along the length of the dyke, coupled with the narrow probability distributions being sensitive to moderately small changes in soil strength. Although some account has been taken of the spatial variability in the third dimension through the three rows of CPTs, these rows were at a spacing of 20 m, which is much greater than the horizontal scales of fluctuation listed in Table 2 and determined at the adjacent test site (de Gast *et al.*, 2017, 2019, 2020). Hence, it seems likely that, while the whole slope was near failure at the maximum drawdown, the actual location of the failure was determined by the spatial variation of soil properties along the dyke and, in

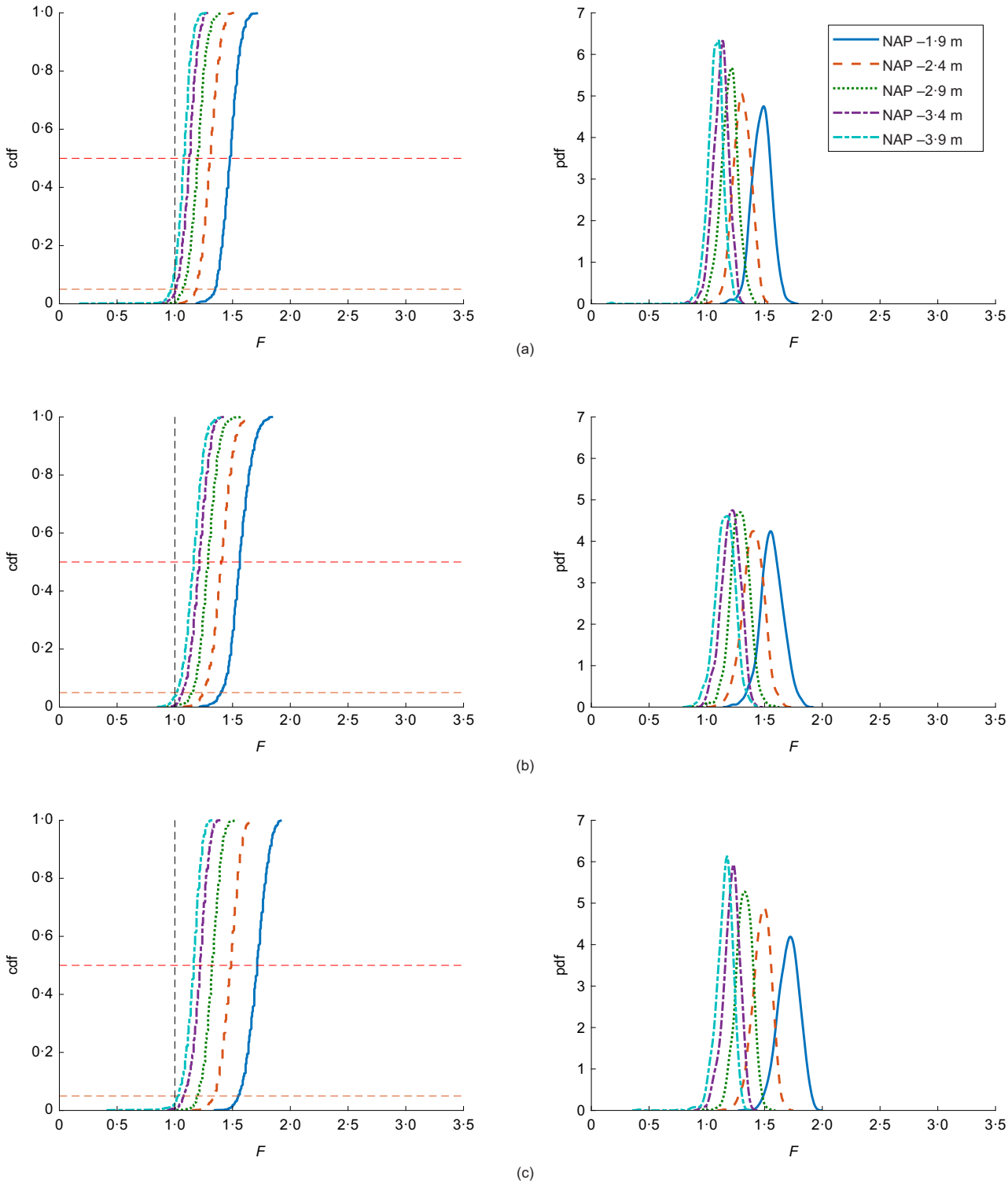

Fig. 11. Conditional RFEM analyses for different levels of drawdown: (a) north row; (b) centre row; (c) south row

particular, by the spatially variable soil profile at a location for which little prior information was available.

Discussion

The failure probabilities approaching failure, listed in Table 4, are supplemented in Table 5 by (*a*) the fifth percentile, (*b*) the 50th percentile (i.e. median) and (*c*) the mean factors of safety back-figured from the probability distributions in Figs 10 and 11. This table shows that, for all analyses, there is very little (or no) difference between the median and mean values of F, thereby reinforcing the impression given in Figs 10 and 11 that the distributions of F are approximately symmetrical. The table also shows that the mean values of F obtained from the conditional simulations are around 10–15% greater than the mean values for the corresponding unconditional simulations, even though both sets of simulations are based on the same statistics that have been derived from the same CPT data. While it is clear that conditioning the random fields will result in a reduction in the standard deviation of F, as is apparent by comparing the widths of the distributions in Figs 10 and 11, it might

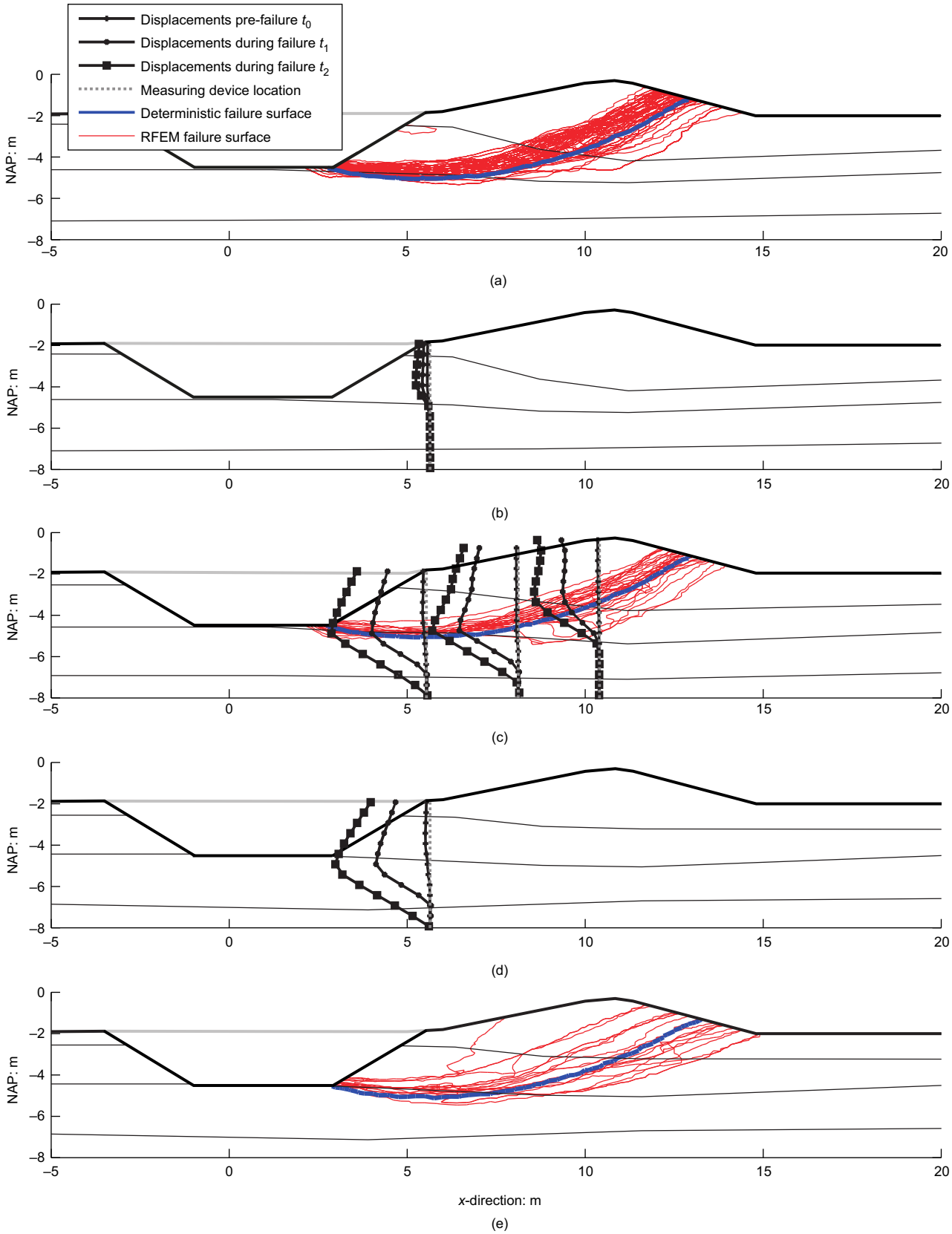

Fig. 12. Computed failure mechanisms for 2·0 m drawdown, and comparison with deterministic mechanisms and measured horizontal displacements (to scale): (a) north row; (b) between north and centre rows; (c) centre row; (d) between centre and south rows; (e) south row

seem counterintuitive that all three cross-sections return mean values of F that are greater than those for the corresponding unconditional simulations. However, from the CPT data in Fig. 4 it may be observed that the s_u profiles have a tendency to be stronger under the dyke (most likely due to the extra load imposed and consequential consolidation of the soil layers). As these profiles are in the immediate vicinity of the failure mechanism and therefore more influential, this is the likely explanation for the increase in mean factors of safety.

Table 5. Evolution of safety factors back-figured from Figs 10 and 11 as a function of drawdown level: (a) based on fifth percentile; (b) based on 50th percentile; (c) based on mean

		Drawdown level: NAP				
		−1·9 m	−2·4 m	−2·9 m	−3·4 m	−3·9 m
(a) Fifth percentile	Unconditional	1·04	0·93	0·79	0·76	0·67
	North row	1·35	1·19	1·07	1·01	0·96
	Centre row	1·40	1·24	1·15	1·06	1·02
	South row	1·55	1·34	1·20	1·08	1·03
(b) 50th percentile	Unconditional	1·40	1·24	1·12	1·04	1·00
	North row	1·48	1·30	1·19	1·12	1·08
	Centre row	1·56	1·40	1·29	1·21	1·16
	South row	1·71	1·49	1·32	1·22	1·16
(c) Mean	Unconditional	1·39	1·23	1·11	1·04	0·98
	North row	1·48	1·30	1·20	1·12	1·07
	Centre row	1·56	1·40	1·29	1·21	1·16
	South row	1·71	1·48	1·32	1·21	1·15

Because, for any stage in the failure test, the conditional distribution of F is narrower than in the unconditional simulation (and has a higher mean), the probability of failure is smaller in the conditional simulations. Table 4 shows that, whereas the probability of failure increases steadily in the unconditional analyses as the drawdown level lowers, it remains very low in the conditional analyses until the drawdown moves between NAP −3·4 m and NAP −3·9 m, whereupon there is a rapid increase in the failure probability. Between these two drawdown levels, corresponding to the approximate situation at failure in the test, the computed probability of failure in the unconditional analysis rises from 38·75% to 50·00%, giving a strong indication of failure. Conversely, at NAP −3·9 m, the computed probabilities of failure in the conditional analyses vary between 2·25% and 11·25%, which, although an indication that failure may be imminent (or has been reached), is not conclusive. This may in part be a reflection of the failure itself initiating between the centre and south rows, rather than at the cross-sections that have been analysed. However, it may also be a sign that, although conditioning the random fields significantly reduces the uncertainty in the influence of the spatial variability on the dyke performance, there are other uncertainties (e.g. measurement, statistical, transformation, material model) that have not been accounted for which may be influential (van den Eijnden & Hicks, 2019). The narrower distributions also mean that computed probabilities of failure may be more sensitive to perturbations in load and soil strength than for unconditional analyses or analyses based on point statistics.

Overall, for the two larger drawdowns considered of 1·5 m and 2·0 m, the failure probabilities in Table 4 and factors of safety in Table 5 are reasonably consistent with the outcome of the failure test. However, no account has been taken of the case history's three-dimensionality, other than to consider three cross-sections along the length of the failure test. De Gast (2020) conducted simple comparative deterministic analyses in two dimensions and three dimensions, and demonstrated an increase in F of around 15% (due to end effects) when analysing the problem in three dimensions. However, no account was taken of spatial variability, and it is clear from the field test that three-dimensional (3D) variability has a significant impact on the mechanism location and extent. Previous parametric studies by Hicks & Spencer (2010), Hicks *et al.* (2014) and Hicks & Li (2018), investigating the influence of spatial variability in 3D analyses of embankment slope stability, have shown that the reliability may be lower in three dimensions than in two dimensions due to the tendency for discrete failures to initiate in weaker zones in the direction of the embankment length.

CONCLUSIONS

A reliability-based analysis framework, accounting for uncertainty arising from the spatial variability of soil properties, has been validated for the controlled, well-instrumented slope failure of an historic dyke in the Netherlands. Using soil property statistics derived from the results of laboratory and CPT data for the different soil layers at the site, the dyke was analysed for the initial (i.e. operating) conditions, as well as for the later stage of the test leading up to failure. The computed probabilities of failure and back-figured factors of safety were broadly consistent with the point at which failure occurred in the test, as was the range of possible failure mechanisms. However, it was not possible to predict the correct location of the failure mechanism in the third dimension, suggesting that the consideration of 3D variability and 3D analysis may be important in such assessments.

It has been shown that, by considering the spatial nature of soil variability, there is a significant reduction in the calculated uncertainty of the structure response, particularly with respect to deterministic and stochastic assessments based only on the point statistics (i.e. the mean and standard deviation) for which there is a much larger range of possible solutions. For example, for the initial configuration of the dyke, a deterministic assessment based on mean soil properties gave an upper-bound (unsafe) solution for the factor of safety, due to failure mechanisms in reality seeking out the weakest path. Conversely, a deterministic assessment based on five-percentile strengths gave a lower-bound (and highly overconservative) solution, because it failed to account for the averaging of properties along potential failure planes.

The results have shown how the uncertainty in dyke assessments may be further reduced, by conditioning analyses to actual site (e.g. CPT) data, and that, for practical (i.e. high) levels of slope reliability, the probability of failure generally reduces as more data become available. However, the results have also highlighted that conditional results can be not sufficiently cautious without a proper consideration of sources of uncertainty other than those associated only with spatial variability – for example, 3D effects – and that they may be

important and should be considered. The reliability-based approach provides a rational framework for quantifying the effects of uncertainties, and enables more informed stability assessments that could make the difference between a dyke being assessed as safe or requiring costly improvement.

ACKNOWLEDGEMENTS

This work is part of the research programme 'Reliable dykes' with project number 13864, which is financed by the Netherlands Organisation for Scientific Research (NWO). The analyses were carried out on the Dutch National e-infrastructure with the support of SURF Foundation.

NOTATION

$E(\theta)$	error function
F	factor of safety
j	counter
N_{kt}	empirical correction factor
q_t	total cone resistance
R	reliability
s_u	undrained shear strength
t	number of pairs of data at lag distance τ
y_j	data
z	location relative to NAP
γ_{sat}	saturated unit weight
$\hat{\gamma}(\tau)$	experimental covariance function
θ_v, θ_h	vertical and horizontal scales of fluctuation
μ	mean (or trend)
$\hat{\mu}$	estimated mean (or trend)
$\rho(\tau)$	theoretical autocorrelation function
$\hat{\rho}(\tau)$	experimental autocorrelation function
σ	standard deviation
σ_v	total vertical stress
τ	lag distance

REFERENCES

Alonso, E. E. (1976). Risk analysis of slopes and its application to slopes in Canadian sensitive clays. *Géotechnique* **26**, No. 3, 453–472, https://doi.org/10.1680/geot.1976.26.3.453.

Ang, A. H. S. & Tang, W. H. (1984). *Probability concepts in engineering planning and design*. New York, NY, USA: Wiley.

Been, K. & Jefferies, M. G. (1985). A state parameter for sands. *Géotechnique* **35**, No. 2, 99–112, https://doi.org/10.1680/geot.1985.35.2.99.

Cami, B., Javankhoshdel, S., Yacoub, T. & Bathurst, R. J. (2018). 2D spatial variability analysis of Sugar Creek embankment: comparative study. In *Proceedings of the 2nd GeoMEast international congress and exhibition on sustainable civil infrastructures: advances in numerical methods in geotechnical engineering* (eds H. Shehata and C. S. Desai), pp. 118–125. Cham, Switzerland: Springer.

Campanella, R. G., Wickremesinghe, D. S. & Robertson, P. K. (1987). Statistical treatment of cone penetrometer test data. In *Proceedings of the 5th international conference on applications of statistics and probability in soil and structural engineering* (ed. N. C. Lind), vol. 2, pp. 1011–1019. Waterloo, Canada: Institute for Risk Research, University of Waterloo.

Cho, S. E. (2007). Effects of spatial variability of soil properties on slope stability. *Engng Geol.* **92**, No. 3–4, 97–109.

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

de Gast, T., Vardon, P. J., Jommi, C. & Hicks, M. A. (2015). The history of safety factors for Dutch regional dykes. In *Geotechnical safety and risk V, proceedings of the 5th international symposium on geotechnical safety and risk (ISGSR15)* (eds T. Schweckendiek, A. F. van Tol, D. Pereboom, M. T. van Staveren and P. M. C. B. M. Cools), pp. 364–370. Amsterdam, the Netherlands: IOS Press.

de Gast, T., Vardon, P. J. & Hicks, M. A. (2017). Estimating spatial correlations under man-made structures on soft soils. In *Geo-risk 2017: impact of spatial variability, probabilistic site characterization, and geohazards* (eds J. Huang, G. A. Fenton, L. Zhang and D. V. Griffiths), Geotechnical Special Publication (GSP) 284, pp. 382–389. Reston, VA, USA: American Society of Civil Engineers.

de Gast, T., Vardon, P. J. & Hicks, M. A. (2019). Observations and considerations regarding estimating horizontal scales of fluctuation around linear infrastructure. In *Proceedings of the 7th international symposium on geotechnical safety and risk (ISGSR 2019)* (eds J. Ching, D. Q. Li and J. Zhang), pp. 340–345. Singapore: Research Publishing.

de Gast, T., Vardon, P. J. & Hicks, M. A. (2020). Assessment of soil spatial variability for linear infrastructure using cone penetration tests. *Géotechnique* (in press).

El-Ramly, H., Morgenstern, N. R. & Cruden, D. M. (2002). Probabilistic slope stability analysis for practice. *Can. Geotech. J.* **39**, No. 3, 665–683.

El-Ramly, H., Morgenstern, N. R. & Cruden, D. M. (2003). Probabilistic stability analysis of a tailings dyke on pre-sheared clay-shale. *Can. Geotech. J.* **40**, No. 1, 192–208.

El-Ramly, H., Morgenstern, N. R. & Cruden, D. M. (2005). Probabilistic assessment of stability of a cut slope in residual soil. *Géotechnique* **55**, No. 1, 77–84, https://doi.org/10.1680/geot.2005.55.1.77.

Fenton, G. A. & Griffiths, D. V. (2008). *Risk assessment in geotechnical engineering*. New York, NY, USA: John Wiley and Sons.

Fenton, G. A., Naghibi, F. & Hicks, M. A. (2018). Effect of sampling plan and trend removal on residual uncertainty. *Georisk: Assessment Manage. Risk Eng. Syst. Geohazards* **12**, No. 4, 253–264.

Griffiths, D. V. & Fenton, G. A. (1993). Seepage beneath water retaining structures founded on spatially random soil. *Géotechnique* **43**, No. 4, 577–587, https://doi.org/10.1680/geot.1993.43.4.577.

Hicks, M. A. (2005). Risk and variability in geotechnical engineering. *Géotechnique* **55**, No. 1, 1–2, https://doi.org/10.1680/geot.2005.55.1.1.

Hicks, M. A. (2006). *Risk and variability in geotechnical engineering*. London, UK: The Institution of Civil Engineers.

Hicks, M. A. & Boughrarou, R. (1998). Finite element analysis of the Nerlerk underwater berm failures. *Géotechnique* **48**, No. 2, 169–185, https://doi.org/10.1680/geot.1998.48.2.169.

Hicks, M. A. & Li, Y. (2018). Influence of length effect on embankment slope reliability in 3D. *Int. J. Numer. Analyt. Methods Geomech.* **42**, No. 7, 891–915.

Hicks, M. A. & Onisiphorou, C. (2005). Stochastic evaluation of static liquefaction in a predominantly dilative sand fill. *Géotechnique* **55**, No. 2, 123–133, https://doi.org/10.1680/geot.2005.55.2.123.

Hicks, M. A. & Samy, K. (2002). Influence of heterogeneity on undrained clay slope stability. *Q. J. Engng Geol. Hydrogeol.* **35**, No. 1, 41–49.

Hicks, M. A. & Spencer, W. A. (2010). Influence of heterogeneity on the reliability and failure of a long 3D slope. *Comput. Geotech.* **37**, No. 7–8, 948–955.

Hicks, M. A., Nuttall, J. D. & Chen, J. (2014). Influence of heterogeneity on 3D slope reliability and failure consequence. *Comput. Geotech.* **61**, 198–208.

Hicks, M. A., Varkey, D., van den Eijnden, A. P., de Gast, T. & Vardon, P. J. (2019). On characteristic values and the reliability-based assessment of dykes. *Georisk: Assessment Manage. Risk Eng. Syst. Geohazards* **13**, No. 4, 331–319.

Javankhoshdel, S., Luo, N. & Bathurst, R. J. (2017). Probabilistic analysis of simple slopes with cohesive soil strength using RLEM and RFEM. *Georisk: Assessment Manage. Risk Eng. Syst. Geohazards* **11**, No. 3, 231–246.

Jiang, S. H., Li, D. Q., Cao, Z. J., Zhou, C. B. & Phoon, K. K. (2014). Efficient system reliability analysis of slope stability in spatially variable soils using Monte Carlo simulation. *J. Geotech. Geoenviron. Engng* **141**, No. 2, 04014096.

Li, Y. J., Hicks, M. A. & Vardon, P. J. (2015). High performance computing strategies for nonlinear finite element analysis of long heterogeneous soil slopes. In *Proceedings of the 23rd UK conference of the Association for Computational Mechanics in Engineering* (eds A. J. Gil and R. Sevilla), pp. 427–430.

Swansea, UK: Association of Computational Mechanics in Engineering – UK.

Li, Y. J., Hicks, M. A. & Vardon, P. J. (2016). Uncertainty reduction and sampling efficiency in slope designs using 3D conditional random fields. *Comput. Geotech.* **79**, 159–172.

Lloret-Cabot, M., Hicks, M. A. & van den Eijnden, A. P. (2012). Investigation of the reduction in uncertainty due to soil variability when conditioning a random field using Kriging. *Géotechnique Lett.* **2**, No. 3, 123–127, https://doi.org/10.1680/geolett.12.00022.

Lloret-Cabot, M., Fenton, G. A. & Hicks, M. A. (2014). On the estimation of scale of fluctuation in geostatistics. *Georisk: Assessment Manage. Risk Eng. Syst. Geohazards* **8**, No. 2, 129–140.

Muraro, S. (2019). *The deviatoric behaviour of peat: a route between past empiricism and future perspectives*. PhD thesis, Delft University of Technology, Delft, the Netherlands.

Phoon, K. K. & Kulhawy, F. H. (1999). Characterization of geotechnical variability. *Can. Geotech. J.* **36**, No. 4, 612–624.

Ponzoni, E. (2017). *Historical constructions on natural silty soils accounting for the interaction with the atmosphere*. PhD thesis, Università degli Studi di Brescia, Brescia, Italy.

Robertson, P. K. (2009). Interpretation of cone penetration tests – a unified approach. *Can. Geotech. J.* **46**, No. 11, 1337–1355.

Rosenblueth, E. (1975). Point estimates for probability moments. *Proc. Natl Acad. Sci. USA* **72**, No. 10, 3812–3814.

van den Eijnden, A. P. & Hicks, M. A. (2017a). Efficient subset simulation for evaluating the modes of improbable failure. *Comput. Geotech.* **88**, 267–280.

van den Eijnden, A. P. & Hicks, M. A. (2017b). Investigating the influence of conditional simulation on small-probability failure events using subset simulation. In *Geo-risk 2017: reliability-based design and code development* (eds J. Huang, G. A. Fenton, L. Zhang and D. V. Griffiths), Geotechnical Special Publication (GSP) 283, pp. 130–139. Reston, VA, USA: American Society of Civil Engineers.

van den Eijnden, A. P. & Hicks, M. A. (2019). On the importance of a complete characterization of site investigation data uncertainty: a computational example. In *Proceedings of the 7th international symposium on geotechnical safety and risk (ISGSR 2019)* (eds J. Ching, D. Q. Li and J. Zhang), pp. 237–242. Singapore: Research Publishing.

Vanmarcke, E. H. (1977). Probabilistic modeling of soil profiles. *J. Geotech. Engng Div., ASCE* **103**, No. 11, 1227–1246.

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

Wickremesinghe, D. & Campanella, R. G. (1993). Scale of fluctuation as a descriptor of soil variability. In *Probabilistic methods in geotechnical engineering* (eds K. S. Li and S. C. R. Lo), pp. 233–239. Rotterdam, the Netherlands: Balkema.