

**ORIGINAL ARTICLE**

# Multiscale Modeling of Backward Erosion Piping in Flood Protection Infrastructure

Alessandro Fascetti\*<sup>1</sup> | Caglar Oskay<sup>2</sup><sup>1</sup>School of Engineering, The University of Waikato, Hamilton, New Zealand<sup>2</sup>Department of Civil and Environmental Engineering, Vanderbilt University, Nashville (TN), USA**\*Correspondence**Alessandro Fascetti  
Gate 1 Knighton Road  
3240 Hamilton, New Zealand  
Email: alessandro.fascetti@waikato.ac.nz**Summary**

This manuscript presents a novel multiscale modeling approach to simulate the evolution of the Backward Erosion Piping (BEP) process in Flood Protection Systems (FPSs). A multiphase description of the BEP phenomenon is proposed for the numerical solution at the local scale and validated by means of full-scale experimental results available in literature. Results of the local scale simulations are used as the training set for a multilayer Machine Learning (ML) model to bridge the information between the local and system scales. Accuracy of the trained ML algorithms is demonstrated by comparing results obtained from detailed physics-based numerical models. The novelty of the proposed methodology lies in its capability of real-time predictions of the overall response at the system scale. A case study is presented where a portion of the Nashville Metro Levee System is analyzed over the span of a year, to assess the likelihood of BEP in the infrastructure. The capability of the model to accept water height data obtained from field measurements is exploited in the numerical simulations.

**KEYWORDS:**

multiscale modeling, flood protection infrastructure, erosion, backward erosion piping, machine learning

## 1 | INTRODUCTION

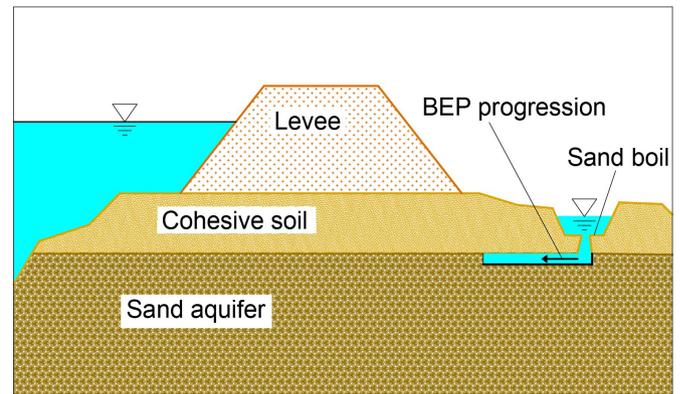
Backward erosion piping is one of the main driving factors for embankments and dam failures (Foster, Fell, & Spannagle, 2000b), and it is considered to be the cause of approximately one third of the piping failures occurred in the last 100 years (K. S. Richards & Reddy, 2007). For this reason, accurate prediction of the evolution of erosion process is critical to the maintenance and stability analysis of FPSs, particularly in view of the catastrophic consequences flooding events can have on impacted communities.

BEP refers to the progressive erosion of the fine fraction of soil from the foundation of a FPS. The process starts at the downstream side of the FPS, often caused by localized damage that exposes the permeable soils, creating high levels of local hydraulic gradient. This initial damage propagates

'backwards', towards the upstream side of the water-retaining structure, eventually leading to the formation of a full pipe (see Fig. 1 ) that undermines the stability of the whole system. BEP is a complex phenomenon, encompassing concepts of sediment transport, underground water flow, hydrodynamics and soil mechanics. In view of its catastrophic consequences, different approaches have been proposed to assess the likelihood of BEP in FPS. The classical design approaches include the Bligh's method (Bligh, 1910; Lane, 1934), Terzaghi's method (Terzaghi, Peck, & Mesri, 1996) and the blanket theory (USACE, 1956). Sellmeijer's method (J. B. Sellmeijer, 1988) offers a simplified analytical solution to distinguish between critical and non-critical conditions when a local damage is present at the downstream side of the FPS (usually referred to as "sand boils"). Schmertmann (2000) proposed a semi-empirical approach to define safety factors against BEP, by defining different correction factors based on theoretical

considerations accounting for the specific features of the structure (e.g. granulometry, slope inclination, blanket thickness). Foster, Fell, & Spannagle (2000a) proposed an empirical procedure to determine the likelihood of erosion to progress in the FPS by collecting data from a total of 1462 embankments. In view of the complexity of the phenomenon, several experimental works have been carried out to quantify the dependence of BEP on the relevant, measurable soil, geometry and load parameters. Among others, Bendahmane, Marot, & Alexis (2008) designed an experimental setup comprising three triaxial cells coupled with air-water cells, in order to investigate the erodibility of soils for different levels on confinement. V. M. van Beek, Knoeff, & Sellmeijer (2011) performed experiments at different scales (i.e. small, medium and large) and compared their observations with results from Sellmeijer's model. H. Sellmeijer, de la Cruz, van Beek, & Knoeff (2011) analyzed the results obtained from different laboratory tests to perform a parametric investigation of their theoretical model. K. Richards & Reddy (2012) designed a triaxial test setup to investigate the effect of confining stress on the initiation of BEP in cohesionless soils. Fleshman & Rice (2014) tested different vertical flow conditions in order to measure the hydraulic response of the soil during erosion processes. Ke & Takahashi (2014) measured the mechanical and hydraulic characteristics of soils at the onset and during the progression of internal erosion. Sharif, Elkholy, Hanif Chaudhry, & Imran (2015) tested different mixtures of sands, silts and clays in flume experiments to investigate the effect of the compaction level on erosion resistance. The above-mentioned studies aimed at defining empirical rules to assess the likelihood of BEP progression in different types of structures, providing valuable information on the dependence of the BEP phenomenon on soil parameters and geometrical characteristics of the site (e.g. the shape and size of the exit condition, geometrical arrangement of the soil layers, presence of filter cutoffs etc.).

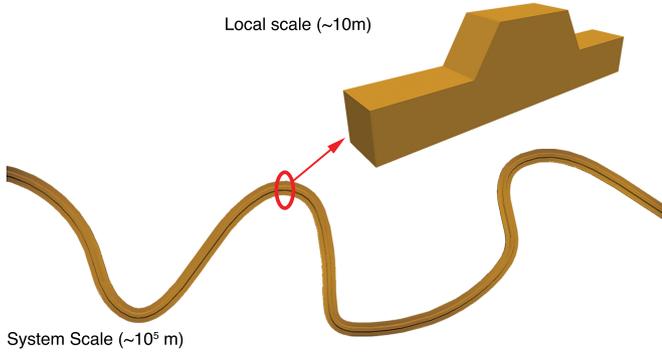
The capability of predicting the evolution of BEP in FPSs is highly desirable due to the transient nature of the phenomenon. While the experiments clarify the relevance of various quantities with respect to the evolution of the erosion process, in-situ conditions are often very different from the laboratory environment because of the time and spatial variability of the driving factors (hydraulic loads, soil properties, seasonality of the water table, presence of localized defects and damage). To this end, numerical predictions offer a valuable tool to assess the likelihood of BEP progression, helping designers and decision makers quantify the health of the systems and plan maintenance accordingly. Three different classes of numerical approaches have been proposed: 1) models in which the erosion progression is described by amplifying the soil conductivity (e.g. Fascetti & Oskay (2019); Robbins (2016);



**FIGURE 1** Schematic description of the backward erosion piping phenomenon.

Vandenboer, van Beek, & Bezuijen (2014)); 2) discrete element method (DEM) coupled with water flow description (e.g. El Shamy & Aydin (2008); Y. Wang & Ni (2013)); and 3) multiphase modeling in which the eroded particle density is described by appropriate constitutive relationships and coupled to the flow description (e.g. Fujisawa, Murakami, & Nishimura (2010); Luo (2013); D.-y. Wang, Fu, Jie, Dong, & Hu (2014); Zhang et al. (2013)). In this work, a multiphase description is used to simulate BEP progression, where a nonlinear transport model is coupled with a description of the eroded mass by means of a concentration-dependent constitutive law. The porosity evolution is then used to express the change in permeability due to the erosion process.

In this context, the present manuscript proposes a novel multiscale framework for the simulation of BEP in earthen embankments. First, a multiphase local scale model for the simulation of BEP is proposed, based on models formulated in literature for different types of erosion problems. The information obtained from a set of such numerical simulations are used as a dataset to train a multilayer Machine Learning model, which is used to track the response evolution at the system scale (see Fig. 2 ). The proposed methodology is also inherently capable of accepting data obtained from field testing. Accuracy of the system scale predictions is demonstrated by comparing results obtained from the local model. The influence of model parameters is investigated by using the ML predictions and results are reported in a compact and efficient graphical fashion using response tables. The capabilities of the proposed multiscale framework in simulating BEP progression in FPS at the system scale is demonstrated by a case study on a portion of the Nashville Metro Levee System, in which available water height recordings were used as input. To the best of the authors' knowledge, there has been no attempt yet in literature to model the flood protection infrastructure at the system scale ( $\sim 10^5$  m). Existing approaches typically focus



**FIGURE 2** Multiscale approach for the simulation of Flood Protection System infrastructure.

on simulating response at the local scale, investigating one representative cross-section of the infrastructure.

The primary novel contributions of the present manuscript are: a) the definition of a flux-dependent constitutive relation embedded in a multiphase description of BEP at the local scale, b) a multilayer ML model composed of a combination of classification and regression tools for the solution of the system scale problem with substantial reductions in computational costs (several orders of magnitude), c) a sensitivity analysis of the response of the system with respect to the model parameters, d) a case study on a portion of the Nashville Metro Levee System infrastructure which shows the potential of the proposed approach in the definition of the likelihood of BEP to happen in FPSs of massive size using data obtained from field measurements.

## 2 | PHYSICS-BASED, "LOCAL" SCALE MODELING OF BEP

Modeling of the BEP process in the embankment is performed using a multiphase approach, in which the nonlinear diffusion equation governing the groundwater flow is coupled with the mass-balance of the fluidized particles by means of a concentration-dependent constitutive law. This section presents the system of governing equations and the solution strategy.

The groundwater flow is governed by a nonlinear diffusion equation (Hagerty & Curini, 2004; Robbins, 2016; Vandenoer et al., 2014):

$$\frac{\partial h(\mathbf{x}, t)}{\partial t} = \nabla \cdot (D(h(\mathbf{x}, t)) \nabla h(\mathbf{x}, t)) \quad \mathbf{x} \in \Omega, t \in (0, T) \quad (1)$$

where  $\Omega$  is a given domain of saturated soil,  $h$  represents the hydraulic head field,  $T$  is the total time and  $D$  is the diffusivity coefficient evaluated as the ratio between the hydraulic conductivity and the specific storage of the soil. The nonlinearity

comes from the dependence of  $D$  on the response field  $h$ . The domain is subject to the following boundary conditions:

$$\begin{aligned} h &= h_b(t) & \text{on } \Gamma_b \subset \partial\Omega \\ \mathbf{q} &\equiv -D \frac{\partial h}{\partial \mathbf{n}} = \mathbf{q}_b & \text{on } \Gamma_q \subset \partial\Omega \end{aligned} \quad (2)$$

with  $\Gamma_b \cap \Gamma_q = \emptyset$  and  $\Gamma_b \cup \Gamma_q = \partial\Omega$ .  $\mathbf{q}$  is the outward flux orthogonal to the domain boundary with normal  $\mathbf{n}$ ,  $\mathbf{q}_b$  the prescribed boundary flux and  $h_b$  is the time-dependent prescribed hydraulic head at the boundary. The description of the evolution of erosion is enriched by the mass-balance equation for the fluidized particles (Stavropoulou, Papanastasiou, & Vardoulakis, 1998):

$$\frac{\partial(c(\mathbf{x}, t)\phi(\mathbf{x}, t))}{\partial t} + \frac{\partial(c(\mathbf{x}, t)\mathbf{q}(\mathbf{x}, t))}{\partial \mathbf{x}} = \frac{\partial\phi(\mathbf{x}, t)}{\partial t} \quad (3)$$

where  $c$  represents the fluidized particle concentration defined as the ratio of the particle volume over the pore space, and  $\phi$  is the soil porosity. The right hand side of Eq. 3 represents the mass generation term, and must be defined by means of appropriate constitutive relations. Based on previous findings for different erosion problems by other authors (Sakthivadivel & Irmay, 1966; Stavropoulou et al., 1998; Vardoulakis, Papanastasiou, & Stavropoulou, 2001; Vardoulakis, Stavropoulou, & Papanastasiou, 1996), we propose the following relation for the evolution of BEP in earthen embankments:

$$\frac{\dot{m}}{\rho_s} = \frac{\partial\phi(\mathbf{x}, t)}{\partial t} = \begin{cases} 0 & \text{if } |\mathbf{q}| \leq q_{crit} \\ \lambda c \sqrt{1 - \frac{\phi^2}{\phi_{max}^2}} (|\mathbf{q}| - q_{crit}) & \text{if } |\mathbf{q}| > q_{crit} \end{cases} \quad (4)$$

where  $\dot{m}$  is the mass generation term,  $\rho_s$  the particles density,  $\phi_{max}$  the maximum allowable value of porosity (related to the fine fraction of the soil),  $q_{crit}$  the critical value of flux, and  $\lambda$  is a model parameter related to the spatial frequency of erosion starter points in the solid skeleton of the soil (Vardoulakis et al., 1996). The evaluation of  $q_{crit}$  can be performed either experimentally (Bendahmane et al., 2008; Van Beek, 2015) or numerically (Fascetti & Oskay, 2019; Robbins, 2016). In this work, the latter approach is used based on comparison with experimental evidence available in literature.

The system of governing equations is completed by considering Darcy's law:

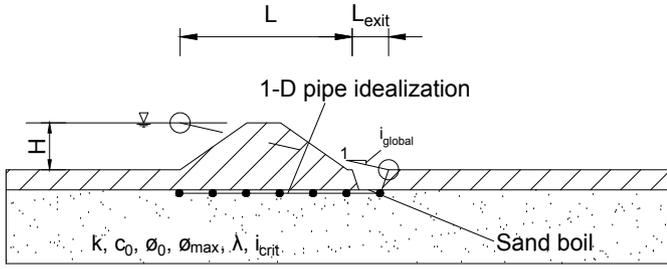
$$\mathbf{q} = -k\mathbf{i} \quad (5)$$

in which  $\mathbf{i}$  is the hydraulic gradient and  $k$  is the conductivity of the soil, related to the porosity using the Kozeny-Carman equation:

$$k = c_{kz} \frac{\phi^3}{(1 - \phi)^2} \quad (6)$$

where  $c_{kz}$  is the Kozeny-Carman constant for the given soil.

The diameter of the pipe is an important characteristic of the BEP process, as it has an influence on the shear stress



**FIGURE 3** Graphical representation of the local scale model parameters.

exerted by the fluid on the particles of the soil skeleton at the pipe tip and the internal boundary. The pipe diameter grows with the progression of the erosion mechanism due to the turbulent nature of the flow at the pipe tip (Rotunno, Callari, & Froiio, 2017). This can lead to increased water flow and erosion rates. Hole erosion tests (HETs) are generally used to characterize soil erodibility in the context of BEP (Fell & Wan, 2002). The interpretation of results is based on the Shields erosion law  $\dot{e} = c_e (< \tau_h - \tau_c >)$ . This equation describes the eroded mass rate as a function of the erosion coefficient  $c_e$ , the fluid shear stress  $\tau_h$  and its critical value  $\tau_c$ . To estimate the critical shear stress, common practice is to deduce the pipe evolution from the initial and final diameters, and subsequently the erosion rate is evaluated from the rate of growth of the pipe diameter (Lominé, Scholtes, Sibille, & Poullain, 2013) as  $\tau_h = (d_p i) / (2l_p)$ . As the pipe widens, the shear stress tends to increase, as described in the previous equation. This aspect is indirectly considered in our model through the erosion constitutive law reported in Equation 4. As the porosity increases at a given point in the domain, in fact, the rate of erosion obtained from Eq. 4b increases accordingly, scaling linearly with porosity. The same behavior was also reported in the works of Vardoulakis et al. (2001 1996) on which the constitutive law we propose is based.

The domain of the problem is schematized as 1-dimensional (the erosion path in Fig. 1). The previously described system of equations is discretized in space using the midpoint finite difference scheme, while discretization in time is performed through forward finite differences leading to an explicit solution scheme. Accuracy of the local scale model was tested numerically, by increasing the mesh density until the values of the local hydraulic gradient converged to the exact solution, for the case of highest applied global gradient in the defined intervals (see Table 3). The same mesh density was used in all the local scale simulations.

## 2.1 | Initial Conditions

In the blanket theory, the US Army Corps of Engineers (2000) defines 12 different cases and sub-cases to take into account all possible geometrical and material arrangements that can occur in river levees. These aspects are related to the 2-dimensional distribution of the flowlines induced by the different conductivity in the top and bottom soil strata in a typical levee geometry. The scope of the so-conducted seepage analysis is to evaluate the local hydraulic gradient at the toe of the levee. This value is used to perform a design check on the heave gradient at that point. In this way, it is possible to consider the effects of two-dimensional flow in the domain for the design of earthen levees.

Rather than checking the value of the hydraulic gradient at a single point, the methodology we propose considers a 1-D domain to quantify the evolution of the hydraulic head, porosity and fluidized particle concentration fields. The geometrical arrangement and the definition of the quantities used to define the local domain are now reported in a revised version of Figure 3, which shows all the parameters involved in the simulations.

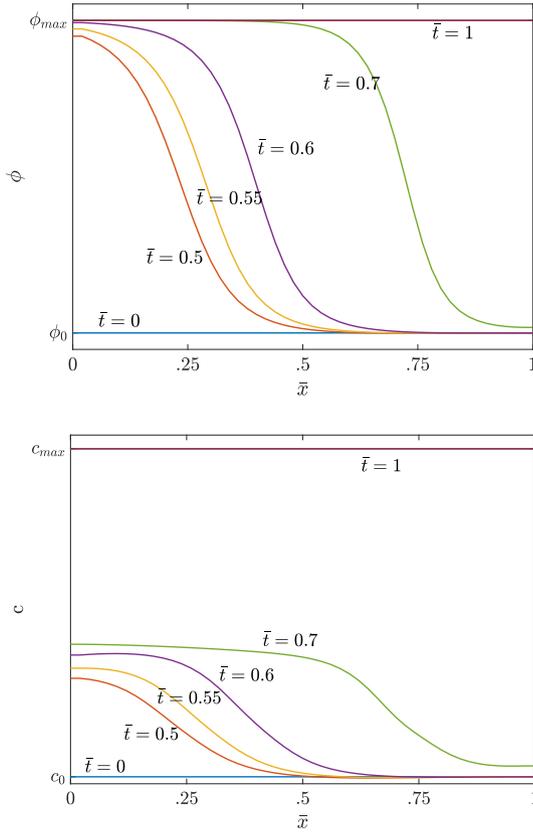
Even though the local scale model we propose uses a 1-dimensional discretization to describe pipe progression, it is possible to consider the effect of the 2-D groundwater flow by adopting an initial distribution for the hydraulic head field derived from a 2-D seepage analysis such as the blanket theory. This allows, similarly to the provisions given in US Army Corps of Engineers (2000), to use values of the hydraulic head field compatible with the local geometry of the levee in the simulations.

BEP is triggered by the presence of localized damage in the impermeable soil layers on the downstream side of the structure (see Fig. 1). In the model, the dimension of such damage is considered a known value and introduced as the initial condition for the hydraulic head  $h(\mathbf{x}, t)$ . The effect of the damage is a local decrease of the hydraulic head in the initiation region, as shown in Fig. 3. This choice is consistent with experimental and numerical observations available in literature (De Wit, Sellmeijer, & Penning, 1981). The head drop is considered to happen instantaneously in the foundation soil, increasing the local hydraulic gradient. The BEP simulation starts at the moment when the head drop is imposed.

## 2.2 | Solution Strategy

The steps involved in the solution of the problem are the following:

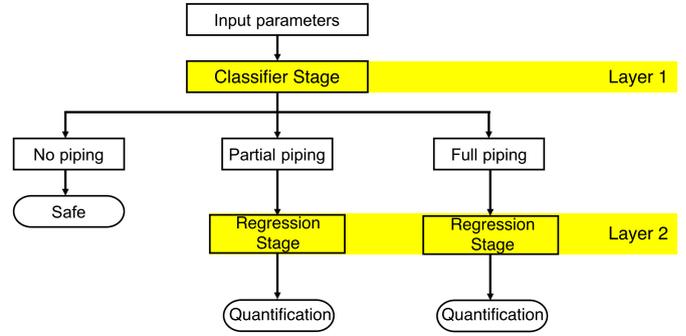
1. Define domain, mesh size, boundary and initial conditions;
2. Solve continuity equation for  $h(\mathbf{x}, t)$ ;



**FIGURE 4** Typical porosity (top) and particle concentration (bottom) evolution obtained from the proposed model.  $\bar{x}$  and  $\bar{t}$  denote normalized position and time, respectively.

3. Calculate  $\mathbf{i}$  from the so obtained hydraulic head field and  $\mathbf{q}$  from Darcy's law (Eq. 5);
4. Evaluate the porosity evolution term  $\frac{\partial \phi(\mathbf{x}, t)}{\partial t}$  from the constitutive law (Eq. 4);
5. Calculate concentration  $c(\mathbf{x}, t)$  from the mass-balance equation (Eq. 3);
6. Calculate new values of conductivity  $k$  using the Kozeny-Carman equation;
7. Update diffusivity  $D(\mathbf{x}, t)$ ;
8. Update time step;
9. Iterate steps 2-8 until total time is met.

Fig. 4 shows the typical evolution of the porosity and particle concentration in the domain, as a function of normalized space and time. The length measure was normalized to the total length of the pipe, while the time was normalized to the total time to complete evolution of the piping process. The erosion phenomenon evolves from the downstream to the



**FIGURE 5** Multilayer Machine Learning modeling approach.

upstream side of the domain as typically observed in the experiments (Bendahmane et al., 2008; H. Sellmeijer et al., 2011; V. M. van Beek et al., 2011).

### 3 | DATA-BASED "SYSTEM" SCALE MODELING OF BEP

Machine learning algorithms are receiving growing attention in different applications for engineering problems (Adeli, 2001; Bogdanor, Mahadevan, & Oskay, 2013; Bogdanor, Oskay, & Clay, 2015; Lin, Nie, & Ma, 2017; Nabian & Meidani, 2018; Rafiei & Adeli, 2018; Reich, 1997; Zhang & Oskay, 2018). The growing capabilities and possibilities to obtain big databases to use as training sets for ML models make it a convenient and effective tool for the solution of problems that are in general too expensive, or when the time required for the solution is not acceptable with respect to the need they are trying to address. In this work, the information from the physics-based local scale simulations is bridged to the system scale (i.e., the FPS) using the multilayer prediction strategy graphically described in Fig. 5. The employed multilayer prediction strategy encompasses a set of classifier and regression models that are trained by means of a database of results of the local scale problem. The classifier model is used to assess whether piping will occur or not, while the regression models furnish quantitative information of the extent of piping for all the cases for which BEP is predicted.

#### 3.1 | Classifier Learner

Classifier algorithms are employed when, given an input data set  $\mathbf{x}$ , it is necessary to assign each entry to one of a set of discrete classes  $C_k$ . The goal is to divide the input hyperspace into a set of decision regions by means of decision boundaries or surfaces. In this study, Support Vector Machines (SVMs) are employed to classify the possibilities of "full piping", "partial piping" or "no piping" in the FPS system at a specified time  $T$ .

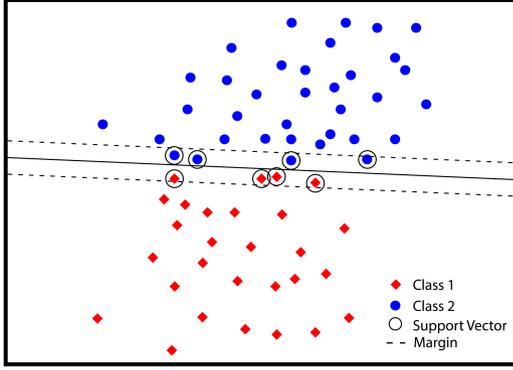


FIGURE 6 Two-class SVM and support vectors.

The three states refer to the occurrence of erosion mechanisms in the whole domain, in a portion of it, or to a condition where of no erosion, respectively.

SVM models are a class of kernel-based algorithms with sparse solutions (Vapnik, 1998). Rather than discarding the dataset after the training phase, SVM use the training set (or a subset of it) in the prediction phase as well. Predictions are made by means of a linear combination of a set of functions defined at the data points. These functions are called kernels, and for models comprising a fixed feature space mapping  $\mathbf{N}$ , they are defined symmetric:

$$k(\mathbf{x}, \mathbf{x}') = \mathbf{N}(\mathbf{x})^T \mathbf{N}(\mathbf{x}') \quad (7)$$

The idea of using kernel functions was first presented by (Aizerman, Braverman, & Rozonoer, 1964) and then applied to ML classifiers by (Boser, Guyon, & Vapnik, 1992), which originated the SVM models. The main advantage of SVM is that the kernel-based algorithm has a sparse solution, meaning that the prediction phase only relies on a subset of the training dataset, allowing for both faster training and predictions.

SVMs separate the samples space through the concept of margin, defined as the smallest distance between the decision boundary and the closest data point. The decision boundary is set to be the one that maximizes the margin. By doing so, the location of the boundary can be determined by using a subset of the data points, known as Support Vectors. SVM models are a decision machine and therefore they do not provide posterior probabilities.

Given that the margin can be calculated as the distance of the border to the closest dataset point  $\mathbf{x}_n$ , the maximization of the margin is equivalent to solving the optimization problem:

$$\max_{\mathbf{w}, b} \left\{ \frac{1}{\|\mathbf{w}\|} \min_n [t_n (\mathbf{w}^T \mathbf{N}(\mathbf{x}_n) + b)] \right\} \quad (8)$$

where  $\mathbf{w}$  and  $b$  represent the model parameters,  $t_n$  the target vector for the input  $\mathbf{x}_n$ . The sign of the quantity  $t_n (\mathbf{w}^T \mathbf{N}(\mathbf{x}_n) + b)$  is positive for correctly classified items and negative otherwise.

For the point that is closest to the surface, one can rescale  $\mathbf{w}$  and  $b$  so that (Bishop, 2006):

$$t_n (\mathbf{w}^T \mathbf{N}(\mathbf{x}_n) + b) = 1 \quad (9)$$

By doing so, it is guaranteed that all other points will satisfy:

$$t_n (\mathbf{w}^T \mathbf{N}(\mathbf{x}_n) + b) \geq 1 \quad (10)$$

which is the canonical representation of the decision hyperplane. By introducing a set of Lagrange multipliers  $a_n \geq 0$  (one for each of the constraint given by Eq. 10), the Lagrangian function is obtained:

$$L(\mathbf{w}, b, \mathbf{a}) = \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{n=1}^N a_n \{t_n (\mathbf{w}^T \mathbf{N}(\mathbf{x}_n) + b) - 1\} \quad (11)$$

By setting  $\partial L / \partial \mathbf{x}$  and  $\partial L / \partial b$  equal to zero, two relations are obtained:

$$\begin{aligned} \mathbf{w} &= \sum_{n=1}^N a_n t_n \mathbf{N}(\mathbf{x}_n) \\ 0 &= \sum_{n=1}^N a_n t_n \end{aligned} \quad (12)$$

that can be used to eliminate  $\mathbf{w}$  and  $b$  from Eq. 11, obtaining the so called dual representation of the model:

$$\tilde{L}(\mathbf{a}) = \sum_{n=1}^N a_n - \frac{1}{2} \sum_{n=1}^N \sum_{m=1}^N a_n a_m t_n t_m k(\mathbf{x}_n, \mathbf{x}_m) \quad (13)$$

where the kernel function definition provided in Eq. 7 has been used. The constraints of the dual problem are:

$$\begin{aligned} a_n &\geq 0, \quad n = 1, \dots, N \\ \sum_{n=1}^N a_n t_n &= 0 \end{aligned} \quad (14)$$

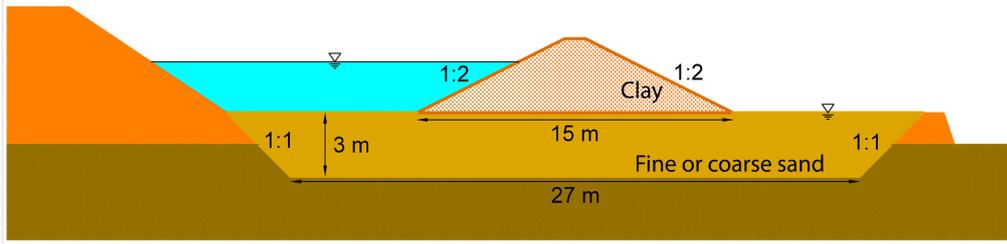
In order to classify a new data point, the sign of the function:

$$y(\mathbf{x}) = \mathbf{w}^T \mathbf{N}(\mathbf{x}) + b = \sum_{n=1}^N a_n t_n k(\mathbf{x}, \mathbf{x}_n) + b \quad (15)$$

is evaluated. Equation 15 explains how the choice of support vectors is made. Every point for which  $a_n = 0$  has no role in the prediction phase, and can therefore be discarded. Every other point with  $a_n > 0$  is retained as a support vector. Figure 6 shows a graphical interpretation of the choice of support vectors for a two-class SVM.

### 3.2 | Regression Learner

Regression algorithms are used to predict the value of a set of continuous target variables  $\mathbf{t}$  given the value of a  $D$ -dimensional vector  $\mathbf{x}$  defining the input. The simplest example of this procedure is the interpolation of available data to obtain values of a defined output quantity inside or outside of the original range in which the data was collected. In this study,



**FIGURE 7** Experimental setup used in the Ijkdijk full-scale experiments.

Gaussian Process (GP) models are used to predict the time to failure or length of pipe progression in the FPS under the given loading conditions in the cases of "partial" and "full" piping, respectively.

GP approaches are used in several fields, such as geostatistics (kriging models), 3-d navigation (Kalman filters) or financial science (autoregressive-moving-average models). A GP is defined as a collection of random variables representing the value of a function  $f(\mathbf{x})$  at the input locations  $\mathbf{x}$ . The values of  $y(\mathbf{x})$  evaluated at any arbitrary set of input points have a joint Gaussian distribution (Rasmussen, 2004). Each observation  $y$  is related to the function  $f(\mathbf{x})$  through a Gaussian model:

$$y(\mathbf{x}) = f(\mathbf{x}) + \mathcal{N}(0, \sigma_n^2) \quad (16)$$

The process  $f(\mathbf{x})$  is therefore completely defined by the mean  $m(\mathbf{x})$  and covariance  $k(\mathbf{x}, \mathbf{x}')$  functions, defined as:

$$\begin{aligned} m(\mathbf{x}) &= \mathbb{E}[f(\mathbf{x})], \\ k(\mathbf{x}, \mathbf{x}') &= \mathbb{E}[(f(\mathbf{x}) - m(\mathbf{x}))(f(\mathbf{x}') - m(\mathbf{x}'))] \end{aligned} \quad (17)$$

The accuracy of the predictions depends on the chosen covariance function. In this work, a squared exponential function is used:

$$k(\mathbf{x}, \mathbf{x}') = \sigma_f^2 \exp\left(-\frac{1}{2l^2} |\mathbf{x} - \mathbf{x}'|^2\right) + \sigma_n^2 I \quad (18)$$

where  $l$  is a length-scale,  $\sigma_f$  is the signal variance and  $\sigma_n$  the noise variance. The previous equation defines the covariance between pairs of random variables. This particular choice has the advantage of yielding values of the covariance close to unity between variables with similar inputs, and decreases as the distance in the input is increased.

In this framework, the joint distribution of the observed target values and the function values at test location is:

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{f}_* \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} K(X, X) + \sigma_n^2 I & K(X, X_*) \\ K(X_*, X) & K(X_*, X_*) \end{bmatrix}\right) \quad (19)$$

where  $X$  represents a training subset of the input points,  $X_*$  a subset of test points, and  $K_*$  represents the  $n \times n_*$  covariance matrix evaluated at every training-test pair. The conditional

probability  $p(\mathbf{f}_* | \mathbf{y})$  indicates the likelihood of the prediction for  $\mathbf{f}_*$  given the input data:

$$\mathbf{f}_* | \mathbf{y} \sim \mathcal{N}(K_* K^{-1} \mathbf{y}, \text{cov}(\mathbf{f}_*)) \quad (20)$$

The accuracy of the predictions performed by means of the previous set of equations depends on the covariance function. The set of hyperparameters  $\boldsymbol{\theta}$  that gives the maximum a posteriori estimate is the one for which  $p(\boldsymbol{\theta} | \mathbf{x}, \mathbf{y})$  is maximized, which is equivalent to maximizing the marginal likelihood:

$$\begin{aligned} \log p(\mathbf{y} | \mathbf{X}) &= -\frac{1}{2} \mathbf{y}^T (K + \sigma_n^2 I)^{-1} \mathbf{y} - \frac{1}{2} \log |K + \sigma_n^2 I| + \\ &\quad - \frac{n}{2} \log(2\pi) \end{aligned} \quad (21)$$

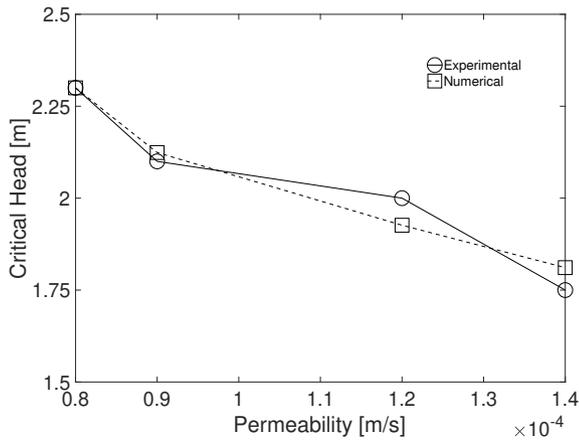
The local and system scale models defined above are embedded in a multiscale framework to obtain real-time predictions of the response of the FPS. The multiscale strategy is based on the use of Machine Learning models trained by means of a database of results obtained from a large set of local, physics-based models.

## 4 | NUMERICAL RESULTS

This section will present the numerical results obtained from the previously described models. In what follows, "local scale" model refers to the physics-based model described in Section 2. The "system scale" model refers to the multilayer ML model described in Section 3. The capabilities of the proposed modeling approach are assessed using the following steps: 1) the local scale model was validated by comparing results obtained from physical experiments available in literature (Section 4.1), 2) the multilayer ML model was first constructed based on a dataset generated by exercising the validated local model. The ML model was validated by comparing its predictions with the local model predictions within a wide range of parameter values. 3) Once the two models were validated, a real case study is investigated by analyzing a large portion of the Nashville Metro Levee System under three selected scenarios.

**TABLE 1** Soil characteristics and parameters in IJkdijk experiments

Type	$d_{70}$	$d_{60}/d_{10}$	Porosity	Conductivity
[-]	[mm]	[-]	[-]	[mm/s]
Fine	0.18	1.6	0.35-0.45	0.08
Coarse	0.26	1.8	0.34-0.45	0.12-0.14

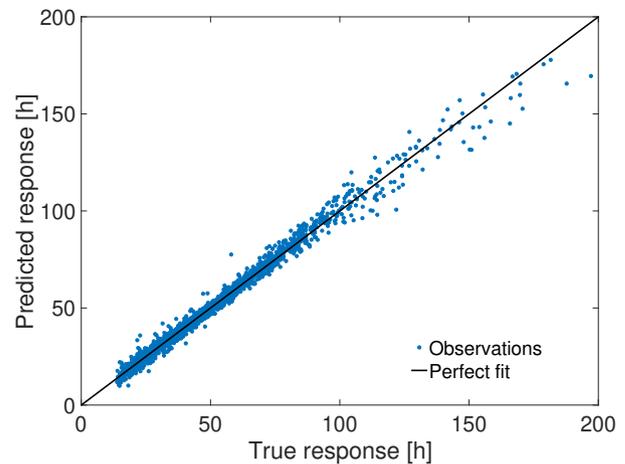
**FIGURE 8** Comparison between experimental and numerical results for the full-scale IJkdijk experiments.

#### 4.1 | Local scale model validation

The local scale model was first validated by comparing numerical results with the experimental results from tests carried out at IJkdijk facilities (Koelewijn et al., 2014; V. van Beek, Bezuijzen, & Sellmeijer, 2013; V. M. van Beek et al., 2011). The schematic setup for the tests is illustrated in Fig. 7 and the measured mechanical characteristics are reported in Table 1.

In the field tests, four different full scale levees were tested. In these tests, the applied hydraulic head on the wet side of the levee was increased at the rate of 100 mm/hr until piping occurred. The tests were simulated using the numerical approach described above to obtain the value for the critical head. The strategy adopted for the simulation of the different tests was the following: all the parameters that could be directly obtained from the results reported in (V. M. van Beek et al., 2011) were used in the simulations, while the other parameters were obtained by fitting the first test setup (i.e. the one that employed the soil with the lowest permeability). The calibrated values were kept constant in all the subsequent simulations. Consistently with the reported test setup, the initial hydraulic head distribution in the domain was imposed so that the hydraulic head value for the downstream side of the structure was equal to 0 (see fig. 7). This assumption

Output Class	full piping	1975 26.7%	84 1.1%	39 0.5%	94.1% 5.9%
	partial piping	52 0.7%	299 4.0%	47 0.6%	75.1% 24.9%
	no piping	51 0.7%	58 0.8%	4795 64.8%	97.8% 2.2%
		95.0% 5.0%	67.8% 32.2%	98.2% 1.8%	95.5% 4.5%
		full piping	partial piping	no piping	
		Target Class			

**FIGURE 9** Confusion matrix for the Quadratic Support Vector Machine model.**FIGURE 10** Real values vs predictions scatter plot for the "full piping" regression case.

is general and does not influence the results of the simulations. Figure 8 shows the comparison of the experimental and numerical results, exhibiting a good agreement and indicating that the model adequately captures piping progression at the local scale. All the parameters used in the validation procedure are reported in Table 2.

#### 4.2 | System scale model validation

Considering the calibrated local model as reference, we proceed with the construction of the system scale model, and

**TABLE 2** Model parameters used in the local model validation

Test no.	$L$ [mm]	$L_{exit}$ [mm]	$i_{global}$ [-]	$c_0$ [-]	$k$ [mm/s]	$\phi_0$ [-]	$\phi_{max}$ [-]	$\lambda$ [mm <sup>-1</sup> ]	$i_{crit}$ [-]
1	15000	10	-	0.009	0.08	0.408	0.8	0.08	0.34
2	15000	10	-	0.009	0.09	0.408	0.8	0.08	0.34
3	15000	20	-	0.009	0.12	0.395	0.8	0.08	0.34
4	15000	20	-	0.009	0.14	0.395	0.8	0.08	0.34

**TABLE 3** Definition of the model parameters and observation intervals

Parameter	Unit	Symbol	Minimum	Maximum
Length	[mm]	$L$	10000	20000
Exit Condition	[mm]	$L_{exit}$	0	500
Hydraulic Gradient	[-]	$i_{global}$	0.01	0.225
Initial Concentration	[-]	$c_0$	0.001	0.01
Initial Porosity	[-]	$\phi_0$	0.35	0.45
Soil Conductivity	[mm/s]	$k$	$1e^{-3}$	$2e^{-1}$
Maximum Porosity	[-]	$\phi_{max}$	0.7	0.85
Lambda	[mm <sup>-1</sup> ]	$\lambda$	0.01	0.1
Critical Gradient	[-]	$i_{crit}$	0.25	0.4

assessment of its accuracy characteristics compared with the physics-based local model. The parameter subspace for the ML model is defined based on continuous ranges for each geometric and soil parameter summarized in Table 3. All model parameters represent a set of continuous variables for which no prior knowledge is available. For this reason, Simple Random Sampling (SRS) is chosen as a suitable sampling technique (Tillé, 2006). Based on this approach, random sampling is performed inside the given intervals shown in Table 3 for each of the set of parameters that constitutes the input for the local model. Every value in the defined intervals has the same probability of being drawn, so that a uniform probability distribution is used. It is important to note that, in order for the classification model to correctly predict cases in which piping does not occur, the parameters space must contain values of the parameters for which the solution of the problem is trivial (as in the case of absence of upstream flow or exit condition). For this reason, low values of the global hydraulic gradient and exit condition width must be included in the training set to improve the accuracy of the classification model.

A total of 7400 simulations were used to train the multilayer ML model. For each training simulation, a set of 9 randomly drawn model parameters is used to construct a local model. A class is assigned to each simulation based on the numerical results, establishing whether erosion evolves completely ("full

piping" case), partially ("partial piping" case), or does not start at all ("no piping"). For the cases of full and partial piping, the output chosen as the response measures are the time to failure and the normalized pipe length (see Fig. 5). The former is evaluated as the time at which the simulated erosion process is fully developed ( $\phi_{upstream} = \phi_{max}$ ), while the latter is simply calculated as the distance of the furthest completely eroded point to the origin, divided by the total length of the domain (see Fig. 7). The outcome of the 7400 simulations shows that 4881 local models resulted in no piping cases, 2078 in full piping cases and 441 in partial piping cases. Differences in the numbers of the observed outcomes follow directly from the SRS of the model parameters, showing how the case of partial piping corresponds to a smaller portion of the parameters hyperspace when compared to the full or no piping cases. The observed imbalance in the dataset size for the different classes is due to the fact that the outcome of the local scale simulations is not known a-priori, so that the numerical results yield a higher number of no piping simulations with respect to the other two groups.

The first step in the proposed multiscale approach consists of the use of a classifier learner. For a given set of parameters, the trained model predicts the class of the current simulation (meaning that the model can select if the current set of

**TABLE 4** Model parameters used in the validation of the system scale model

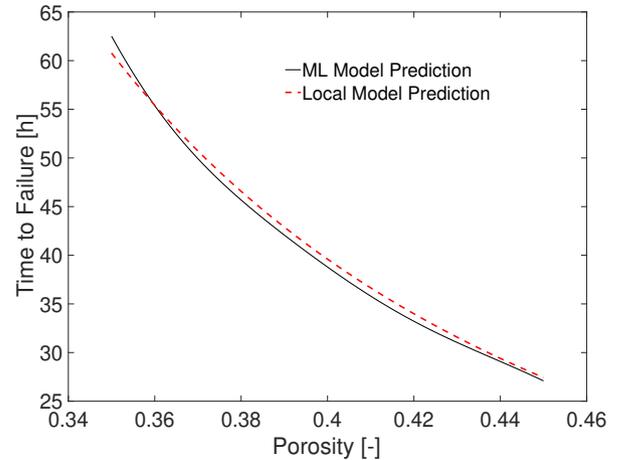
$i_{global}$	$c_0$	$\phi_0$	$\phi_{max}$	$\lambda$	$i_{crit}$
[-]	[-]	[-]	[-]	[mm <sup>-1</sup> ]	[-]
0.2043	0.009	0.35-0.45	0.81	0.08	0.34

parameters yields a full, partial or null piping process), allowing for the prediction of the current state of the structure. The Quadratic Support Vector Machine shows an accuracy of 95.5% when considering a 5-fold cross-validation for the obtained dataset, meaning it can confidently be used to classify the current state of the system based on a given set of parameters. Figure 9 shows the confusion matrix for the trained classifier model, which graphically describes the number of true positive samples for every class (on the matrix diagonal), together with false positive and negative samples (non-diagonal terms). Due to the difference in dataset sizes for the three different groups, the model exhibits higher confidence values for the "no piping" case and the lowest value for the "partial piping" one.

In the second step of the multiscale procedure, a set regression ML models were trained for the partial and full piping classes. The predictors used in the training phase are the parameters reported in Table 3, while the observed responses were the length of the observed pipe and the time to the complete evolution of the erosion process. It is worth noticing that for the case of full piping, the evaluation of the pipe length  $L_{pipe}$  is trivial (i.e.  $L_{pipe} = L$ ) and therefore it does not require direct evaluation (as shown in Fig. 5). A 5-fold cross-validation was used on these two dataset as well, and the GP regression model exhibits  $R^2$  values of 97% and 94% for the "full" and "partial" piping, respectively (see Fig. 10).

One interesting feature of the trained ML model is the capability of generating predictions of the influence of one or more parameters on the response of the system. In order to assess the accuracy of the proposed system scale model, obtained results are compared with the local scale model. Both tests involved the solution of 200 simulations in which all the parameters were kept constant except for the initial porosity, which varied linearly inside the interval. The summary of the model parameters involved is given in Table 4.

Fig. 11 shows the comparison between the two models. The good agreement between the two curves indicates that the ML model can be used to obtain information on the system response. Both sets of simulations were run using an Intel i7 desktop computer. The total run-time for the physics-based local scale model was 58 minutes, while the run-time for the



**FIGURE 11** Influence of the initial porosity on the time to failure: numerical model vs ML model predictions

ML system scale model was 3.2 seconds, with a speedup of approximately three orders of magnitude.

### 4.3 | Sensitivity Analysis

The fast prediction capability of the ML model was also leveraged to perform a parameter sensitivity analysis. In contrast with traditional local sensitivity characterizations, which rely on varying one parameter, while all others are kept constant, we create response maps, where two parameters are simultaneously varied. All the possible combinations of two from the nine parameters involved in the study were investigated. The intervals chosen for the values of the parameters were the same reported as in Table 3. The remaining parameters not directly investigated at each step were taken equal to the mean value of the respective interval, for comparison purposes.

The subplots in Fig 12 were obtained by discretizing the plane of two variables into a regular grid of 200 by 200 points. The first step involved the use of the classifier model previously trained, in order to discern between cases of "no piping", "partial piping" or "full piping". This information allows for the creation of three different zones in the parameters space. Subsequently, for each point residing in the partial or full piping domains, a prediction was made based on the trained regression models. For the case of "partial piping", the ML model allows for the prediction of the extension of the pipe (reported as normalized to the length of the domain), while in the case of "full piping" it is possible to evaluate the time to failure (expressed in hours). The time to failure and length of partial piping are reported as contours in Figs 12.

In this way, several tables are obtained, which provide useful information on the response of the system based on the parameters values. The so-obtained tables provide a tool to assess the

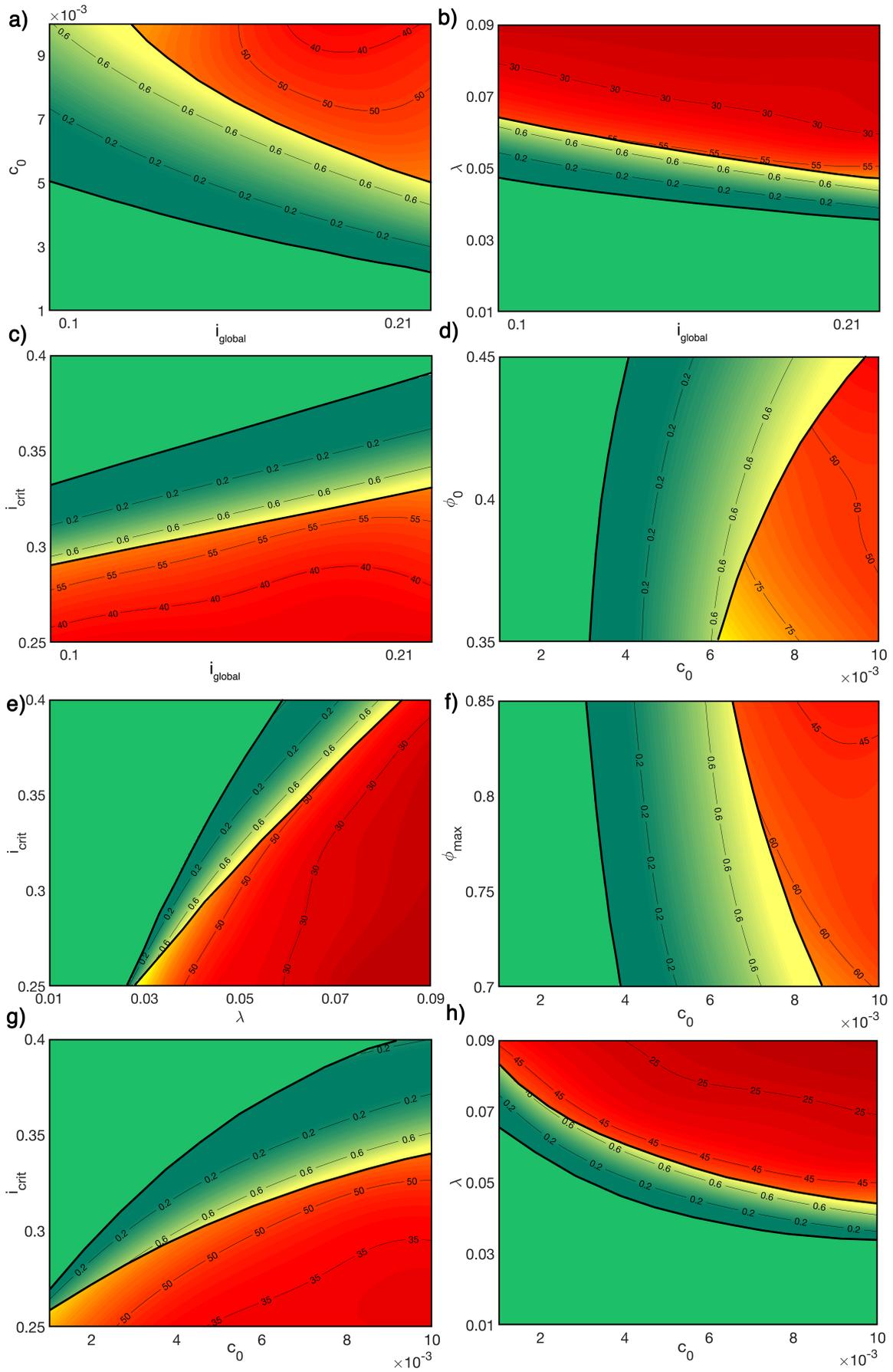


FIGURE 12 Sensitivity study performed by means of the proposed ML approach.



**FIGURE 13** Spatial distribution of the water level control points for Cumberland River in the Nashville Metro area.

likelihood of piping, and the severity of the process evolution, in a compact yet systematic way. It is worth noticing that performing such sensitivity analysis was only possible because of the fast predictions guaranteed by the proposed multilayer ML approach, as the computational cost associated with the solution of the problem would be prohibitive for the physics-based approach.

The sensitivity analysis shows how the value of the different parameters affect the results of the simulations, detailing interesting features of the numerical model while describing the extent and severity of BEP in the infrastructure for different cases. For instance, Fig. 12 c shows the interaction between the critical gradient  $i_{crit}$  (i.e., a measure of the soil resistance to erosion) and the global gradient  $i_{global}$  (i.e., a measure of the hydraulic loads acting in the structure). The plot shows a linear relationship between the two parameters, with the expected result that higher values of the hydraulic loads or lower values of the critical gradient yield more critical states with respect to the progression of BEP in the system. On the other hand, other parameters show a clear non-linear trend, as in Fig. 12 h, where the interaction between the model parameter  $\lambda$  and the initial concentration  $c_0$  was investigated.

Information obtained from the sensitivity analysis can therefore be used to the twofold objective of: 1) evaluating the effect of different combinations of model parameters on the results of the simulation, 2) convey a graphical interpretation of the proposed model.

#### 4.4 | System Scale Predictions

The predictions obtained from the ML model are used in a multiscale framework in order to evaluate the system level response in an efficient fashion. By analyzing the entire levee system, it is possible to capture localized failures as a function of the geometrical and mechanical features of the system. The capabilities of the proposed multiscale model are demonstrated

by simulating the progression of BEP in a river levee system. The chosen testbed is the Nashville Metro Levee System (see Fig. 13 ). This infrastructure is of great importance for the Nashville Metro area, which has already experienced a severe flood event in 2010 (Moore, Neiman, Ralph, & Barthold, 2012). At each time step of the simulation the input parameters are evaluated and the output response of the system scale model is used to evaluate the state at every computational point in the system. The evaluation of the total length of eroded path at time step  $n_s$  at an arbitrary position in the system is performed as follows:

$$L_{pipe}(t = n_s dt) = \sum_{i=1}^{n_s} v_{pipe}(i) \cdot dt \quad (22)$$

where  $i$  is the counter for the time steps of amplitude  $dt$  and  $v_{pipe}$  represents the linearized speed of the process defined as:

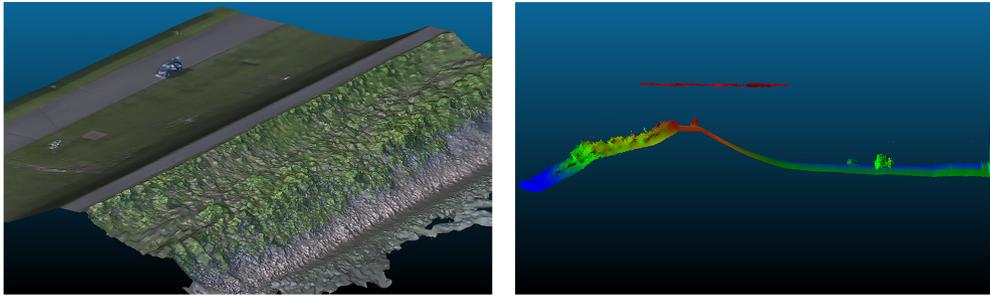
$$v_{pipe}(i) = \begin{cases} 0 & \text{if case = 'safe'} \\ \frac{L}{t^*} & \text{if case = 'partial' \& } L_{pipe} < L^* \\ 0 & \text{if case = 'partial' \& } L_{pipe} \geq L^* \\ \frac{L}{t^{**}} & \text{if case = 'full'} \end{cases} \quad (23)$$

where  $L^*$  and  $t^*$  are the length and time to failure obtained from the "partial" piping regression case,  $L$  is the total length of the erosion path and  $t^{**}$  is the time to failure obtained from the 'full piping' regression model.

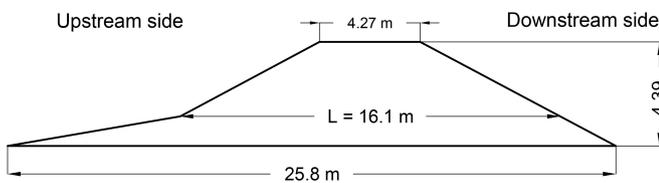
#### 4.4.1 | Construction of the computational domain

The reconstruction of the geometry of the system is a task associated with several levels of difficulty and uncertainties. The planar geometry of the levee system at the structural level (i.e. the levee centerline) can be reconstructed from publicly available satellite information with satisfactory precision. This information itself is not sufficient to reconstruct the 3-dimensional geometry of the structure, but only its arrangement in a horizontal plane. For this reason, data obtained from satellite images is enriched with the information obtained from a remote sensing campaign carried out by the authors on the Nashville Metro Levee System. A portion of the FPS in exam was scanned by means of both photogrammetric and LiDAR techniques, using Unmanned Aerial Vehicles (UAVs) to perform fast and accurate reconstructions of the infrastructure. Fig 14 shows the two 3-dimensional models obtained by means of the two techniques, while Fig. 15 shows the average reconstructed cross-section for the Nashville Metro Levee that was used in this study. Each one of the so-obtained sections represents a computational point in the domain of the FPS, which was discretized over the length of the levee centerline.

The water level tables were obtained from the USGS National Water Information System, which is a database



**FIGURE 14** Two 3-dimensional reconstructions of a portion of the Nashville Metro Levee system executed by means of photogrammetry (left) and LiDAR-based SLAM algorithms (right).



**FIGURE 15** Cross-section of the Nashville Metro Levee system reconstructed by means of the UAV remote sensing campaign.

composed of approximately 1.5 million sites in the United States (Goodall, Horsburgh, Whiteaker, Maidment, & Zaslavsky, 2008). The data obtained from the database is the water level height at the six locations reported in Fig. 13 . The values of the water height recorded at the six discrete points reported in Fig. 13 were then interpolated in space in order to obtain values at the computational nodes. An observation period of 1 year was considered in the simulations (from January to December 2016). This time period was chosen so that the effect of seasonality can be properly taken into account.

Obtaining information on the mechanical properties of the soil is a challenging task for several reasons: 1) the levee system was built in the early 1920's so that original design plans are not readily available, 2) different stakeholders are responsible for the maintenance of the infrastructure, 3) several natural hazards required partial reconstructions of the structure leading to changes in materials and sectional arrangements over the length of the system. The information used in this study was obtained from the available technical reports on the Nashville Metro Levee System (Metropolitan Nashville-Davidson County, 2015; Nashville Area Metropolitan Planning Organization, 2015; Nashville Metro Water Services, 2013).

#### 4.4.2 | Safety factors

The Factor of Safety (FoS) against BEP failure is defined as follows:

$$FoS(\mathbf{x}, t) = 1 - \frac{L_{pipe}(\mathbf{x}, t)}{L(\mathbf{x})} \quad (24)$$

The FoS, defined as a function of time and position along the system domain, is equal to 0 when the pipe length is equal to the levee length, and 1 when the pipe length equals zero.

#### 4.4.3 | Considered scenarios and full scale results

The performed system scale simulations took into account the portion of the Nashville levee shown in Fig. 13 , for a total length of the structure of approximately 47 km. Three different fictional scenarios were considered as the initial conditions for the simulations. For all of them, the simulations were run for a full year. This choice was made to take into account seasonality on the evolution of BEP. The model input constitutes the nine parameters defined in Table 3 at every computational point in the levee. In the first scenario, spatially constant geometrical and mechanical parameters of the levee were considered along with the water data recordings. The values of the parameters are reported in Table 5 . This choice is equivalent to imposing a spatially constant initial damage on the downstream side of the FPS for the considered length. Even though this scenario is highly unlikely to happen in reality, it represents the first benchmark to test the numerical procedure. As expected, the evaluated FoS decreases in all the sections according to the measured water levels (points at a lower altitude have a higher value of the hydraulic load and therefore the pipe progression runs faster). Eventually, after enough time all the monitored sections of the levee undergo extensive piping, as shown in Fig. 16 .

In the second scenario, spatial variability of the input parameters was considered. The size of the structure is big enough that the mechanical and geometrical features of the FPS

**TABLE 5** Values of the parameters used in the system scale simulations

Parameter	Unit	Symbol	Mean Value (All Scenarios)	Distribution (Scenario 2)	Parameters (Scenario 2)	Variability (Scenario 3)
Length	[mm]	$L$	16100	Normal	$\sigma = 100$	-
Exit Condition	[mm]	$L_{exit}$	400	Weibull	$a = 5, b = 0.5$	500 (localized)
Hydraulic Gradient	[-]	$i_{global}$	Calculated	-	-	-
Initial Concentration	[-]	$c_0$	0.005	Normal	$\sigma = 1e^{-3}$	-
Initial Porosity	[-]	$\phi_0$	0.35	Normal	$\sigma = 0.01$	-
Soil Conductivity	[mm/s]	$k$	$1e^{-1}$	Normal	$\sigma = 1e^{-4}$	-
Maximum Porosity	[-]	$\phi_{max}$	0.8	Normal	$\sigma = 0.1$	-
Lambda	[mm <sup>-1</sup> ]	$\lambda$	0.3	Normal	$\sigma = 0.01$	-
Critical Gradient	[-]	$i_{crit}$	.275	Normal	$\sigma = 0.025$	-

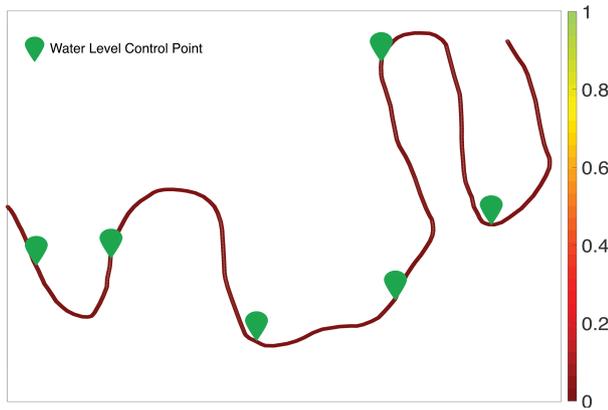
exhibit a significant variation throughout the system. Probability distributions were assumed for all the parameters listed in Table 3 and the value of the coefficients were either obtained through literature review (in the case of mechanical quantities) or numerical considerations (in the case of model parameters). The simulations show that the non-uniform input parameters distribution results in a variable FoS distribution along the structure. Fig. 17 displays the results of the numerical simulation at the end of the simulation, showing how several damage localization zones develop, indicating the presence of extensive BEP in the infrastructure.

In the third scenario, a constant distribution of the parameters was considered, with the difference that two portions of the levee (the Bordeaux area and a portion in-between the Inglewood and Omohundro areas) were considered as experiencing severe damage on the downstream side of the FPS. This choice represents the presence of extensive maintenance works on parts of other infrastructures that might impact the stability of the FPS, or the presence of a highly aggressive source of external damage (such as tree branches or fauna). This increased damage was modeled by imposing an exit condition length ( $L_{exit}$ ) of 500 mm in the two areas. This scenario represents a relatively common case of practical interest, as the infrastructure often experiences high levels of localized damage on the downstream side, as reported in previous studies (Foster et al., 2000b; Hagerty & Curini, 2004). The presence of such initial damage could be detected by human or automated inspections. Figure 18 shows the results of the third scenario. The localization of the downstream damage triggers higher erosion rates in the zone, resulting in localized damage and lower factor of safety in the triggered area. It is also possible to see different levels of erosion between the two localized areas. This is explained by the fact that the Inglewood area is situated in a depressed region and therefore experiences higher levels of

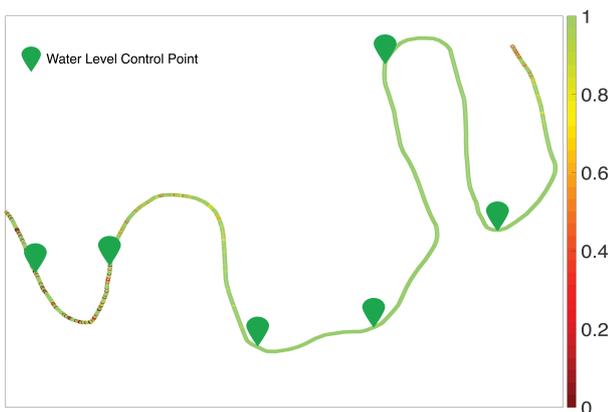
water height and hydraulic gradients, resulting in more severe piping.

## 5 | CONCLUSIONS

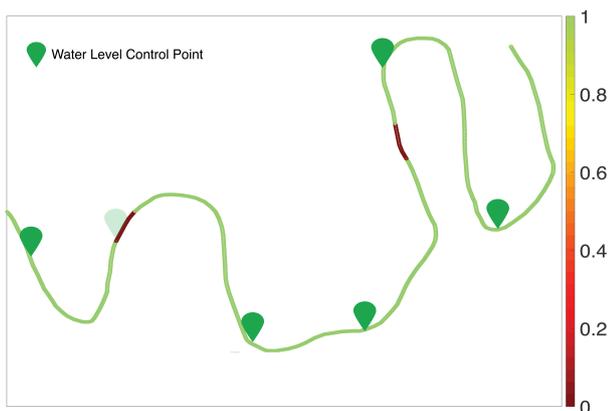
A novel multiscale procedure is proposed for the simulation of BEP progression in earthen embankments. A multiphase model is defined at the local scale, simulating the progression of erosion by means of a concentration-dependent constitutive law. Numerical solutions of such model are employed as database for the training of a ML model at the system scale, which predicts the possible states of failure in the domain and provides quantitative information regarding the damage in the system as a function of space and time. The proposed multiscale framework is computationally efficient enough to allow evaluation of health at the system scale. Moreover, full sensitivity studies are possible by means of the fast ML predictions, allowing for the evaluation of response maps that graphically represent the influence of different parameters on the results. As a case study, the proposed model was employed for the simulation of the portion of FPS that protects the Nashville area. Simulations were performed according to different scenarios that correspond to common practical situations, and the capability of the model to accept field data was demonstrated. The proposed multiscale framework could be embedded in resilience and risk analysis platforms, providing an accurate and efficient tool for the real-time prediction of damage progression in FPSs. Natural extension of our work for future research is the definition of a 3-dimensional numerical model for the solution of the local scale problem that could potentially lead to an increased level of description of the erosion process. This would allow for a greater accuracy in the description of the global scale behavior through the proposed multilayer ML model. This, however, would come at the cost of a significant



**FIGURE 16** Spatial distribution of the computed factor of safety for Scenario 1.



**FIGURE 17** Spatial distribution of the computed factor of safety for Scenario 2.



**FIGURE 18** Spatial distribution of the computed factor of safety for Scenario 3.

increase in the computational demand for the solution of the local scale problems and the creation of the results database.

## ACKNOWLEDGMENTS

This research is based upon work supported by Vanderbilt University under the "Intelligent Resilient Infrastructure Systems (IRIS)" Trans-Institutional Programs (TIPs) Initiative, which is greatly appreciated and acknowledged.

The authors wish to express their gratitude to the editor and the six anonymous reviewers for their valid and constructive suggestions, which greatly improved the quality of the manuscript. The authors would like to emphasize that the scenarios included in the Nashville Metro Levee System study are fictional and do not necessarily indicate areas of high flood risk.

## References

- Adeli, H. (2001). Neural networks in civil engineering: 1989–2000. *Computer-Aided Civil and Infrastructure Engineering*, 16(2), 126–142.
- Aizerman, M., Braverman, E., & Rozonoer, L. (1964). The probability problem of pattern recognition learning and the method of potential functions. *Automation and Remote Control*, 25, 1175–1190.
- Bendahmane, F., Marot, D., & Alexis, A. (2008). Experimental parametric study of suffusion and backward erosion. *Journal of Geotechnical and Geoenvironmental Engineering*, 134(1), 57–67.
- Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.
- Bligh, W. (1910). Dams, barrages and weirs on porous foundations. *Engineering News*, 64(26), 708–710.
- Bogdanor, M. J., Mahadevan, S., & Oskay, C. (2013). Uncertainty quantification in damage modeling of heterogeneous materials. *International Journal for Multiscale Computational Engineering*, 11(3).
- Bogdanor, M. J., Oskay, C., & Clay, S. B. (2015). Multiscale modeling of failure in composites under model parameter uncertainty. *Computational mechanics*, 56(3), 389–404.
- Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on computational learning theory* (pp. 144–152).
- De Wit, J., Sellmeijer, J., & Penning, A. (1981). Laboratory testing on piping. In *Proceedings of the 10th international conference on soil mechanics and foundation engineering, stockholm, sweden* (pp. 15–19).
- El Shamy, U., & Aydin, F. (2008). Multiscale modeling of flood-induced piping in river levees. *Journal of Geotechnical and Geoenvironmental Engineering*, 134(9), 1385–1398.

- Fascetti, A., & Oskay, C. (2019). Dual random lattice modeling of backward erosion piping. *Computers and Geotechnics, 105*, 265–276. doi:
- Fell, R., & Wan, C. (2002). Investigation of internal erosion and piping of soils in embankment dams by the slot erosion test and the hole erosion test-interpretative report. *Technical Report R-412*.
- Fleshman, M. S., & Rice, J. D. (2014). Laboratory modeling of the mechanisms of piping erosion initiation. *Journal of Geotechnical and Geoenvironmental Engineering, 140*(6), 04014017.
- Foster, M., Fell, R., & Spannagle, M. (2000a). A method for assessing the relative likelihood of failure of embankment dams by piping. *Canadian Geotechnical Journal, 37*(5), 1025–1061.
- Foster, M., Fell, R., & Spannagle, M. (2000b). The statistics of embankment dam failures and accidents. *Canadian Geotechnical Journal, 37*(5), 1000–1024.
- Fujisawa, K., Murakami, A., & Nishimura, S.-i. (2010). Numerical analysis of the erosion and the transport of fine particles within soils leading to the piping phenomenon. *Soils and foundations, 50*(4), 471–482.
- Goodall, J. L., Horsburgh, J. S., Whiteaker, T. L., Maidment, D. R., & Zaslavsky, I. (2008). A first approach to web services for the national water information system. *Environmental Modelling & Software, 23*(4), 404–411.
- Hagerty, D., & Curini, A. (2004). Impoundment failure seepage analyses. *Environmental & Engineering Geoscience, 10*(1), 57–68.
- Ke, L., & Takahashi, A. (2014). Triaxial erosion test for evaluation of mechanical consequences of internal erosion. *Geotechnical Testing Journal, 37*(2), 347–364.
- Koelewijn, A., De Vries, G., Van Lottum, H., Förster, U., Van Beek, V., & Bezuijen, A. (2014). Full-scale testing of piping prevention measures: Three tests at the ijdijk. *G. & White, Physical modeling in geotechnics*. London: Taylor & Francis Group.
- Lane, E. W. (1934). Security from under-seepage masonry dams on earth foundations. In *Proceedings of the american society of civil engineers* (Vol. 60, pp. 929–966).
- Lin, Y.-z., Nie, Z.-h., & Ma, H.-w. (2017). Structural damage detection with automatic feature-extraction through deep learning. *Computer-Aided Civil and Infrastructure Engineering, 32*(12), 1025–1046.
- Lominé, F., Scholtes, L., Sibille, L., & Poullain, P. (2013). Modeling of fluid–solid interaction in granular media with coupled lattice boltzmann/discrete element methods: application to piping erosion. *International Journal for Numerical and Analytical Methods in Geomechanics, 37*(6), 577–596.
- Luo, Y. (2013). A continuum fluid-particle coupled piping model based on solute transport. *International Journal of Civil Engineering, 11*(1B), 38–44.
- Metropolitan Nashville-Davidson County. (2015, January). *Metropolitan Nashville-Davidson County Mulyi-Hazard Mitigation Plan*.
- Moore, B. J., Neiman, P. J., Ralph, F. M., & Barthold, F. E. (2012). Physical processes associated with heavy flooding rainfall in nashville, tennessee, and vicinity during 1–2 may 2010: The role of an atmospheric river and mesoscale convective systems. *Monthly Weather Review, 140*(2), 358–378.
- Nabian, M. A., & Meidani, H. (2018). Deep learning for accelerated seismic reliability analysis of transportation networks. *Computer-Aided Civil and Infrastructure Engineering, 33*(6), 443–458.
- Nashville Area Metropolitan Planning Organization. (2015). *Building Resilience: A Climate Adaptation Plan*.
- Nashville Metro Water Services. (2013, January). *Unified Flood Preparedness Plan*.
- Rafiei, M. H., & Adeli, H. (2018). A novel unsupervised deep learning model for global and local health condition assessment of structures. *Engineering Structures, 156*, 598–607.
- Rasmussen, C. E. (2004). Gaussian processes in machine learning. In *Advanced lectures on machine learning* (pp. 63–71). Springer.
- Reich, Y. (1997). Machine learning techniques for civil engineering problems. *Computer-Aided Civil and Infrastructure Engineering, 12*(4), 295–310.
- Richards, K., & Reddy, K. R. (2012). Experimental investigation of initiation of backward erosion piping in soils. *Géotechnique, 62*(10), 933.
- Richards, K. S., & Reddy, K. R. (2007). Critical appraisal of piping phenomena in earth dams. *Bulletin of Engineering Geology and the Environment, 66*(4), 381–402.
- Robbins, B. (2016). Numerical modeling of backward erosion piping. *Applied Numerical Modeling in Geomechanics, 551–558*.
- Rotunno, A. F., Callari, C., & Froiio, F. (2017). Computational modeling of backward erosion piping. In *Models, simulation, and experimental issues in structural mechanics* (pp. 225–234). Springer.
- Sakthivadivel, R., & Irmay, S. (1966). *A review of filtration theories*. University of California, Hydraulic Engineering Laboratory, College of Engineering.
- Schmertmann, J. H. (2000). The no-filter factor of safety against piping through sands. In *Judgment and innovation: The heritage and future of the geotechnical engineering profession* (pp. 65–132). American Society of Civil Engineers.
- Sellmeijer, H., de la Cruz, J. L., van Beek, V. M., & Knoeff,

- H. (2011). Fine-tuning of the backward erosion piping model through small-scale, medium-scale and ijdijk experiments. *European Journal of Environmental and Civil Engineering*, 15(8), 1139–1154.
- Sellmeijer, J. B. (1988). On the mechanism of piping under impervious structures. *Technical University of Delft*.
- Sharif, Y. A., Elkholy, M., Hanif Chaudhry, M., & Imran, J. (2015). Experimental study on the piping erosion process in earthen embankments. *Journal of Hydraulic Engineering*, 141(7), 04015012.
- Stavropoulou, M., Papanastasiou, P., & Vardoulakis, I. (1998). Coupled wellbore erosion and stability analysis. *International journal for numerical and analytical methods in geomechanics*, 22(9), 749–769.
- Terzaghi, K., Peck, R. B., & Mesri, G. (1996). *Soil mechanics in engineering practice*. John Wiley & Sons.
- Tillé, Y. (2006). *Sampling algorithms*. Springer.
- US Army Corps of Engineers. (2000, April). *Design and Construction of Levees*.
- USACE. (1956). *Investigation of underseepage and its control, lower mississippi rivers*. Vicksburg, MS.
- Van Beek, V. (2015). *Backward erosion piping: Initiation and progression* (Unpublished doctoral dissertation). TU Delft.
- van Beek, V., Bezuijen, A., & Sellmeijer, H. (2013). Backward erosion piping. *Erosion in geomechanics applied to dams and levees*, 193–269.
- van Beek, V. M., Knoeff, H., & Sellmeijer, H. (2011). Observations on the process of backward erosion piping in small-, medium- and full-scale experiments. *European Journal of Environmental and Civil Engineering*, 15(8), 1115–1137.
- Vandenboer, K., van Beek, V., & Bezuijen, A. (2014). 3d finite element method (fem) simulation of groundwater flow during backward erosion piping. *Frontiers of Structural and Civil Engineering*, 8(2), 160–166. doi:
- Vapnik, V. (1998). The support vector method of function estimation. In *Nonlinear modeling* (pp. 55–85). Springer.
- Vardoulakis, I., Papanastasiou, P., & Stavropoulou, M. (2001). Sand erosion in axial flow conditions. *Transport in Porous Media*, 45(2), 267–280.
- Vardoulakis, I., Stavropoulou, M., & Papanastasiou, P. (1996). Hydro-mechanical aspects of the sand production problem. *Transport in porous media*, 22(2), 225–244.
- Wang, D.-y., Fu, X.-d., Jie, Y.-x., Dong, W.-j., & Hu, D. (2014). Simulation of pipe progression in a levee foundation with coupled seepage and pipe flow domains. *Soils and Foundations*, 54(5), 974–984.
- Wang, Y., & Ni, X. (2013). Hydro-mechanical analysis of piping erosion based on similarity criterion at micro-level by pfc3d. *European Journal of Environmental and Civil Engineering*, 17(sup1), s187–s204.
- Zhang, X., & Oskay, C. (2018). Material and morphology parameter sensitivity analysis in particulate composite materials. *Computational Mechanics*, 62(3), 543–561.
- Zhang, X., Wong, H., Leo, C. J., Bui, T. A., Wang, J., Sun, W., & Huang, Z. (2013). A thermodynamics-based model on the internal erosion of earth structures. *Geotechnical and Geological Engineering*, 31(2), 479–492.