Delft University of Technology

Document Version
Final published version

Licence
CC BY

Citation (APA)
Brinkgreve, R. B. J., & Ahmad, M. (2026). On the use of surrogate modelling in parameter selection. In J. Grabe (Ed.), *Conference Proceedings Workshop Numerical Methods in Geotechnics 2026: AI in geotechnics, deformaton based design, numerical optimization* (pp. 113-125). Hamburg University of Technology.

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
In case the licence states “Dutch Copyright Act (Article 25fa)”, this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership.
Unless copyright is transferred by contract or statute, it remains with the copyright holder.

Sharing and reuse
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.

On the use of surrogate modelling in parameter selection

Ronald B.J. Brinkgreve, Majd Ahmad

Abstract: Surrogate modelling is an application of Artificial Intelligence in which a Machine Learning model is trained to produce results that are originally obtained from a more complicated (numerical) model, especially when numerous such results are required. This article demonstrates two examples of using surrogate modelling in geotechnical applications involving (constitutive) model parameter selection as part of model fitting (parameter optimization) and probabilistic analysis. It gives some practical ideas on how to maintain the consistency of parameter sets in the case of parameter inter-dependencies, how to improve the quality of the surrogate in the case of large parameter variations, and how to improve model fitting to a target curve.

1 Introduction

The use of Artificial Intelligence (AI) and Machine Learning (ML) progresses very rapidly in all kinds of applications. Although Geotechnical Engineering is historically relatively slow in picking up new global trends, there are numerous applications one could think of in which AI and ML can improve the efficiency of the geotechnical workflow. To narrow down the scope of this contribution, focus is given here on Surrogate Modelling, as a particular application of ML. In Surrogate modelling, an ML model (Classification or Regression) is used to provide fast output results ('output features') based on a set of input parameters ('input features'), thereby replacing more complicated and time-consuming (calculation) models. A representative set of input features (e.g. geometric parameters, model parameters, model conditions, etc.), covering the domain of the intended application, is used a-priori in the more complex model to generate the corresponding model results, while input parameters and related output results serve as training data for the surrogate model, provided the model results are consistent and reliable. Once trained, the surrogate model can be applied instantaneously on a new ('unseen') set of input parameters, producing results as if they were obtained from the more complex calculation model, but much faster.

Several examples of surrogate modelling have been reported recently, where surrogate models replace more complex geotechnical calculation models (mostly finite element models), in cases where large numbers of model simulations are required. Mitelman et al. (2023) show how surrogate modelling can be used to support the observational method during the execution of a geotechnical construction project. Ruiz López et al. (2024) demonstrate the use of surrogate

modelling for an application involving a shaft excavation in London clay, highlighting their potential use in the geotechnical design process. Granitzer et al. (2025) describe a surrogate modelling framework for evaluating the stability of cantilever walls, considering uncertainty and sensitivity of the input parameters. They hint on the possibility of adopting a fully probabilistic approach in the design process.

In this article, surrogate modelling is used to speed up the process of parameter selection in two different types of applications, namely parameter optimization and probabilistic analysis.

Section 2 of this article describes an application in which parameter selection in model fitting is used to find the best set of constitutive model parameters to best match target stress-strain curves for a given series of laboratory tests. In this application, surrogate modelling is used to replace non-linear numerical stress point simulations of laboratory tests.

Section 3 describes an application in which parameter selection is used in probabilistic analysis to predict the probability of failure of a river dike under high-water design conditions. In this application, surrogate modelling is used to replace non-linear 2D finite element calculations.

Section 4 gives some conclusions and recommendations on the use of surrogate models in applications involving model parameter selection and optimization. The article ends with some references for further reading.

2 Parameter optimization

In finite element modelling of practical geotechnical applications, numerical single stress point simulations of soil lab tests can be used to calibrate constitutive model parameters against real lab testing data. Although single stress point simulations are rather quick, the number of runs needed in parameter optimization may still cost more time than desired within a broader workflow. Surrogate modelling can be a solution to reduce calculation time.

Starting point for the development of a framework for parameter selection and optimization using surrogate modelling was recent work on the UDCAM-S model (Ragni et al., 2023; Del Brocco et al., 2026). This model is based on the NGI-ADP constitutive model (Grimstad et al., 2012) in which the model parameters are adapted for cyclic loading effects using a cyclic accumulation procedure. The cyclic accumulation procedure results in target stress-strain curves for undrained triaxial compression (TXC), triaxial extension (TXE) and direct simple shear (DSS). The parameters of the NGI-ADP model are optimized in a numerical procedure in which the soil lab tests are numerically simulated, to provide the best match with the target stress-strain curves. For details and backgrounds on the UDCAM(-S) model and the accumulation procedure, reference is made to Jostad & Andresen (2009), Jostad et al. (2014) and Andersen (2015).

The existing PLAXIS-implementation of UDCAM-S (Plaxis, 2018) as well as the updated procedure (Del Brocco et al., 2026) use Particle Swarm Optimization (PSO) to optimize the NGI-ADP model parameters. In this context, each ‘particle’ represents a set of model parameters. Starting from a certain number of particles (typically 25 – 100) a certain number of iterations (typically 25 – 100) are required to provide a converged solution. Each particle evaluation in each iteration requires three numerical lab test simulations. While in the existing procedure the

optimization is done only once as pre-processing, the updated procedure is performed for multiple soil layers in an iterative setting. This may eventually result in the order of 10^6 lab test simulations, all together.

2.1 Surrogate model

A surrogate model in the form of an artificial neural network with intermediate hidden layers (Multi-Layer Perceptron, MLP) was created to produce stress-strain curves for all three test types for a given set of model parameters. Input features of the MLP are the ten UDCAM-S model parameters, and output features are the corresponding (scaled) stress-strain curves represented by 3 x 101 value pairs (i.e. 101 curve points for each test with corresponding stress and strain value), giving 606 output features in total per parameter set. The data was conditioned by dividing stress values by the cyclic undrained shear strength in triaxial compression, τ^C. The input and output data are further scaled using the StandardScaler function in Python. Between the 10 input features and the 606 output features, the MLP comprises three hidden layers of 64, 256 and 512 neurons, respectively. The MLP model, used as `outputs = model(inputs)` and the associated data structure are designed to deal with multiple samples (`inputs`, i.e. a series of *10-parameters* sets) and corresponding results (`outputs`, i.e. a series of *606-values* sets) at once. Moreover, the Python code in which the whole procedure is implemented is designed to make use of the GPU (via the CUDA library) to gain optimum performance when used in the context of parameter selection and optimization.

A dataset composed of 47500 10-parameters sets, and corresponding results obtained from numerical single stress point lab test simulations (3x101 stress-strain pairs from TXC, TXE and DSS), was used to train the surrogate model. 80% of the data was used for training; 20% for testing. By using the surrogate on random samples from unseen parameter sets and comparing with numerical lab test simulations, it was verified that the surrogate curves almost perfectly match the curves obtained from the numerical simulations. This is essential to let the parameter optimization converge towards the right solution. Note that the surrogate verification only means that the surrogate is a good representation of the *numerical* lab test simulations with the selected constitutive model; it does not mean it is representative of *real* lab testing data.

Besides sufficient representative training data, covering the full domain of potential parameter values (considering their realistic combinations and correlations; see next section), proper scaling of input and output features is essential to create a high-quality surrogate that can be used in parameter optimization. This may also involve taking log-values instead of the actual parameter values when variations of more than one order of magnitude are encountered.

2.2 Particle Swarm Optimization

To generalize the parameter optimization process, the PSO algorithm is designed in terms of seven independent parameters $\{\xi_1, ..., \xi_7\}$ with values in the range [0, 1], instead of using the ten UDCAM-S model parameters directly. The independent parameters relate in the following way to the actual UDCAM-S model parameters, considering their search space in the parameter optimization process:

Table 1: Parameters in the UDCAM-S model and their respective ranges used in PSO

Parameter	Expression	Unit	Meaning
OCR	1.0	-	Over-consolidation ratio (fixed value)
ν	0.495	-	Undrained Poisson's ratio (fixed value)
τ^C	*select* (ξ_1, 5, 100)	kPa	Cyclic undrained shear strength in TXC
G_{ur} / τ^C	*select* (ξ_2, 50, 1000)	-	Unload/reload shear modulus ratio
τ^P / τ^C	*select* (ξ_3, 0.03, 2.3)	-	Undr. shear strength ratio in TXE/TXC
τ^{DSS} / τ^C	*select* (ξ_4, 0.3, 2.3)	-	Undr. shear strength ratio in DSS/TXC
γ_f^C	*select* (ξ_5, 400, 2000) / (G_{ur}/τ^C)	%	Equivalent failure shear strain in TXC
γ_f^E	*select* (ξ_6, 0.95, 5.0) · γ_f^C	%	Equivalent failure shear strain in TXE
γ_f^{DSS}	*select* (ξ_7, 0.95, 5.0) · γ_f^C	%	Equivalent failure shear strain in DSS
τ_0 / τ^C	0.5 (1–K_0) σ'_v / τ^C	-	Initial strength mobilization ratio

K_0 is obtained from the target curves. The *select* function returns an interpolated value between the second (lower bound) and third (upper bound) value in the list, based on the independent parameter ξ_i. The expressions in the second column with upper and lower bound values are the result of a preliminary (internal) research with the purpose of arriving at complete parameter sets that are considered realistic for the type of soil in this application.

The advantages of this approach are:

1. The algorithm becomes model independent; it is easy to make the procedure suitable for other models with other parameters.
2. Correlations between model parameters can be easily accounted for.

It should be realized that advanced constitutive models, such as the NGI-ADP, UDCAM-S, or Hardening Soil model (Schanz et al., 1999), may incorporate some implicit limitations and dependencies, which means that parameter values cannot be selected fully independent from each other. Examples are the failure strains (γ_f^C, γ_f^E, γ_f^{DSS}) in Table 1. These are selected (with some variation, as expressed by the ranges) from the abstract parameters ξ_i in relation to the shear modulus ratio (G_{ur} / τ^C) as well as the failure strain in TXC (γ_f^C). Using the model parameters directly as independent parameters in PSO could lead to invalid parameter combinations, which could break down the optimization procedure.

The PSO algorithm is designed in such a way that all 'particles' (where each particle is a set of seven abstract parameters ξ_i relating to 10 model parameters) are calculated simultaneously. In this case, 50 particles are used, which means that `inputs` contains 50 parameter sets and `outputs` contains the 50 corresponding results (refer to Section 2.1 for a description of `inputs` and `outputs`). A total of 30 iterations is performed to find the optimum location of the particles within the search space. The particle giving the lowest 'cost' (i.e. smallest deviation from the target curves) at some point during the iteration process is considered the optimum set of parameters.

Since the initial stiffness (i.e. gradient in the stress-strain curve) is an important curve feature to match, the cost function is designed such that it not only considers the vertical distance between the target curve and the model result (as reflected by *MAE*, *MSE* or R^2 error measures), but it also considers differences in (initial) gradients.

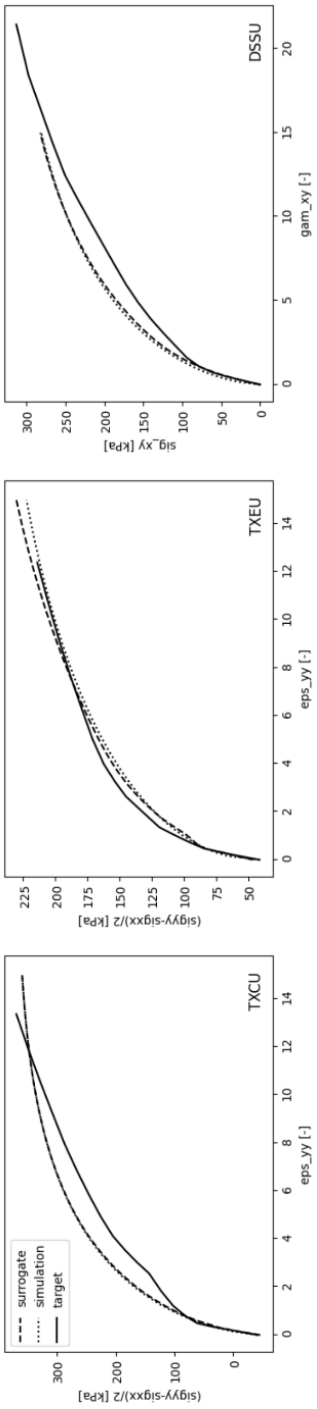


Figure 1. Comparison of target curves for TXC, TXE and DSS with stress-strain curves obtained from the surrogate model and numerical lab test simulations.

Figure 1 shows a comparison between a set of target curves for TXC, TXE and DSS, the results obtained from the surrogate model for the optimized parameter set, and the results obtained from numerical lab test simulations with the same parameter set.

The plots show that:

1. The results from the surrogate model match very well with those from the numerical lab test simulations
2. The model results fit quite well to the target curves (with special attention to the initial stiffness)

3 Probabilistic analysis of dike failure

Recent design guidelines for river dikes in The Netherlands (POVM, 2020) describe the procedure and requirements for the use of Finite Element calculations in the design process. Under new climatic conditions, many dikes need to be reinforced. At locations where space is limited and dikes cannot be increased in height in the conventional way with the required safety level, structures (e.g. piles, sheet piles, anchors) can be used as reinforcement elements to meet the safety requirements. Numerical analysis (e.g. finite element analysis) is the only way to take soil-structure interaction properly into account, which is essential in such cases.

Besides the use of partial factors, imposed by the national annex of the Dutch version of Eurocode 7, as used in standard design calculations, there is a tendency to move towards probabilistic analysis. The result of a probabilistic analysis is a probability of failure P_f (considering a certain failure mechanism or the overall probability of failure). However, probabilistic analysis, especially Monte-Carlo Simulation (MCS), may become extremely time-consuming if the required number of calculations are to be performed by finite element analysis. Surrogate modelling can be used to make such calculations feasible. In this case, model characteristics (geometry, soil parameters, load conditions, etc.) are parameterized as input features, and the surrogate is trained on certain results (output features) of a modest number of finite element calculations considering variations in (stochastic) input parameters. The trained surrogate model is then used to provide the large number of results that is required in the probabilistic analysis.

3.1 Case study: Lekdike near Bergambacht

In a research project at Delft University of Technology, a case study was elaborated with the purpose to implement a workflow and develop the necessary algorithms for finite element-based probabilistic analysis of dike failure. In a case study, a dike section along the river Lek near Bergambacht is assumed to be reinforced by means of a sheet pile wall. The wall starts about halfway the inside embankment slope at +2.6 m down to -12 m (with respect to the Dutch Reference Level, NAP), into the Pleistocene sand layer. The sand layer is overlain by Holland Peat and clay-type soil, similar as the dike itself. In the 2D finite element model, the sheet pile wall is modelled by plate elements with interfaces at both sides. In the present study, the original site geometry and subsurface profile are slightly simplified to facilitate the probabilistic analyses while retaining the key geotechnical features governing dike stability. Figure 2 shows a schematization of the Bergambacht case study with the most relevant dimensions.

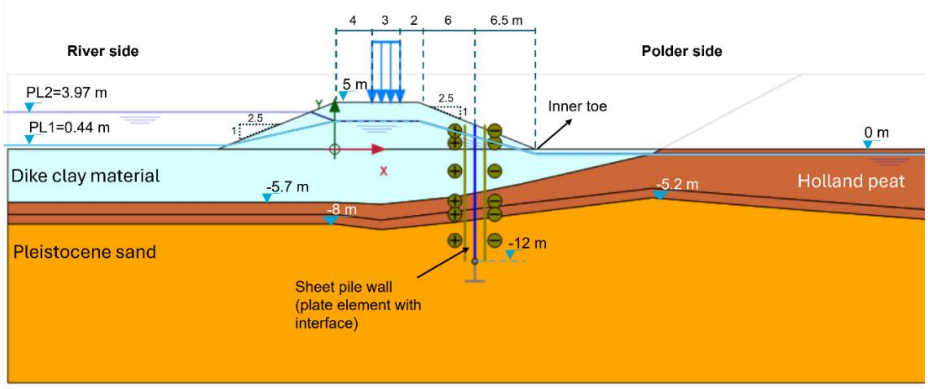


Figure 2. Schematization of the Bergambacht case study for probabilistic analysis

The sand layer is modelled by the Hardening Soil (HS) model. The peat and clay layers are initially modelled by the Soft Soil Creep (SSC) model to set up the stress state, install the sheet pile wall, apply high-water conditions (PL2 in Figure 2) and apply a predefined surface load. At the point where design requirements are to be checked, a switch is made to the NGI-ADP SHANSEP model. Some of the model parameters are used as stochastic variables in the probabilistic analysis, considering their mean values based on site investigation data and stochastic distributions (distribution functions, standard deviations) based on relevant literature.

A complete finite element calculation involves a total of 8 calculation phases, following the POVM recommendations (POVM 2020):

1. Initial stress generation using K_0-procedure on the original dike geometry and initial water level (PL1)
2. Stress redistribution, restoring equilibrium
3. 10-years creep period (*OCR* will evolve)
4. Dike reinforcement (heightening + sheet pile wall)
5. 100-years creep period (effects on structural forces for expected lifespan)
6. Increase to high-water level (PL2)
7. Clay and peat layers: Switch to NGI-ADP SHANSEP model
   - Apply distributed load
   - Apply partial factors on model parameters:
     - ➢ check if soil failure does not occur
     - ➢ check for maximum displacement
     - ➢ check if structural forces are within structural design values
8. Global safety factor (SF) analysis
     - ➢ check that SF is larger than required schematization factor

For further details of the finite element model, reference is made to Ahmad et al. (2025).

3.2 Stochastic parameters

After an initial screening and sensitivity analysis of all parameters using the elementary effects method (Morris, 1991), a total of 16 parameters were found to be most relevant for the calculation results. These parameters were used in the probabilistic analysis. Their stochastic distributions are listed in Table 2.

As mentioned in the first example, some model parameters are inter-dependent and cannot be selected arbitrarily. Therefore, the parameters of the Hardening Soil model (Pleistocene sand layer) were all selected using the correlations by Brinkgreve et al. (2010) based on Relative Density (*RD*) as a stochastic parameter.

Table 2: Stochastic parameters used in probabilistic analysis, with their distributions

Parameter	Model	Soil	Distribution	Mean	*COV*
γ_{sat}	HS	Sand	Normal	20	0.04
φ‘	HS	Sand	Truncated normal	32.5	0.1
R_{inter}	HS	Sand	Truncated normal	0.67	0.065
RD	HS	Sand	Truncated normal	0.6	0.3
γ_{sat}	SSC	Clay	Normal	18.45	0.04
κ^*	SSC	Clay	Lognormal	0.005	0.3
φ‘	SSC	Clay	Truncated normal	35	0.1
POP	SSC	Clay	Lognormal	30	0.2
κ^*	SSC	Peat	Lognormal	0.025	0.3
s_u^{DSS} / s_u^A	NGI-ADP SHANSEP	Clay	Truncated normal	0.66	0.1
α	NGI-ADP SHANSEP	Clay	Normal	0.32	0.2
Power (m)	NGI-ADP SHANSEP	Clay	Normal	0.8	0.06
G / s_u^A	NGI-ADP SHANSEP	Peat	Lognormal	40.9	0.15
s_u^{DSS} / s_u^A	NGI-ADP SHANSEP	Peat	Truncated normal	0.89	0.1
α	NGI-ADP SHANSEP	Peat	Normal	0.36	0.2
Stiff.red.fac-	Elastic	Sheet pile	Uniform	0.9	-

For the Soft Soil Creep model (clay and peat layers), the parameters λ^* and μ^* were selected as:

$$\lambda^* = select\ (\xi_1, 5, 10) \cdot \kappa^* \qquad \text{Modified compression index} \qquad (1a)$$

$$\mu^* = select\ (\xi_2, 0.02, 0.08) \cdot \lambda^* \qquad \text{Modified creep index} \qquad (1b)$$

where the parameters ξ_i are randomly selected from a uniform distribution in the range [0, 1]. This additional variability is meant to account for the epistemic uncertainty in empirical correlations, while keeping the parameter sets in the surrogate model limited.

In the NGI-ADP SHANSEP model, s_u^A is based on the SHANSEP expression:

$$s_u^A = \sigma'_v\ \alpha\ OCR^m \qquad \text{Undrained shear strength in active loading} \qquad (2)$$

in which σ'_v and *OCR* in each stress point are based on the finite element results of the calculation phase preceding the switch to the NGI-ADP SHANSEP model. The power *m* is set to 0.8 and the passive strength ratio is set to $s_u^P/s_u^A = (s_u^{DSS}/s_u^A)^2$. The failure shear strains are selected as:

$$\gamma_f^C = 8 \text{ (clay) or } 10 \text{ (peat)} \qquad \text{Failure strain in TXC} \qquad (3a)$$

$$\gamma_f^{DSS} = select\ (\xi_3, 1.5, 2.0) \cdot \gamma_f^C \qquad \text{Failure strain in DSS} \qquad (3b)$$

$$\gamma_f^E = \gamma_f^C + 2\ (\gamma_f^{DSS} - \gamma_f^C) \qquad \text{Failure strain in TXE} \qquad (3c)$$

Geometric uncertainties in terms of variations in soil layer levels, water levels, etc., have not been considered as stochastic parameters in the probabilistic analysis.

A total of 25000 sets (i.e. 25000 'samples'), each containing the 16 parameters in Table 2, were generated using Sobol quasi-random sampling. With these parameter sets, complete finite element calculations were performed. The results (based on evaluations from calculation phase 7) comprising information about soil collapse, maximum displacement and maximum pile force, were stored for each parameter set. 21405 runs completed successfully without soil collapse or numerical failure, while soil collapse (macro instability) occurred in 3549 runs and numerical failure occurred in 46 runs. The latter (46) were excluded from the dataset, since these failures are purely numerical (non-physical).

Considering these results, there can be different definitions of 'failure' and 'non-failure'. Clearly, the 3549 runs in which soil collapse occurred in phase 7, can be classified as 'failure'. Moreover, runs without soil collapse in phase 7, but resulting in displacements larger than a given threshold value, can also be classified as 'failure'. Finally, runs giving a SF in phase 8 less than the required schematization factor of 1.1, can also be classified as 'failure'. Different definitions of failure were considered separately and subsequently processed into binary classes ('non-failure' / 'failure') for the training of surrogate models.

3.3 Surrogate model

Surrogate models were created based on a dataset composed of the 24954 usable results of the finite element calculations and their corresponding input parameter sets. The surrogate model is a neural network-based classification model with 16 input features and 2 output classes ('non-failure' (0) and 'failure' (1)). Input parameters were normalized to zero mean and unit covariance before training. The architecture and training settings of the neural network were determined through a hyperparameter tuning procedure. Different numbers of hidden layers and neurons per layer were tested, and the architecture yielding the lowest loss was selected. In addition to the network architecture, the training procedure included several measures to improve generalization and prevent overfitting. A dropout rate of 20% was applied during training, and a dynamic learning rate was used to adapt the learning rate during optimization. The number of training epochs was not fixed; instead, early stopping was applied based on the loss function evaluated on the validation/testing data, and training was terminated when further training no longer improved the loss. The final architecture consisted of two hidden layers with 64 and 32 neurons, respectively.

To enhance the training and predictive quality of the surrogate model in the probabilistic analysis, logistic regression is introduced. For this, a scalar value z (called 'logit') is defined as outcome for a given parameter set:

$$z = \log\left(\frac{p}{1-p}\right) \tag{4}$$

where p is the probability that failure occurs for the given parameter set. With this definition:

$z < 0$ means: 'non-failure' is more likely than 'failure' for this parameter set
$z > 0$ means: 'failure' is more likely than 'non-failure' for this parameter set
$z = 0$ means: 'non-failure' and 'failure' have equal probabilities for this parameter set

The sigmoid function can be applied to convert logits into failure probabilities p:

$$\sigma(z) = \left(\frac{1}{1 + e^{-z}}\right) \tag{5}$$

With logit z in the range ⟨-∞, ∞⟩ the sigmoid function σ(z) will give a result in the range ⟨0, 1⟩. A sigmoid close to 1.0 means a high probability of failure for the given parameter set, while a sigmoid close to 0.0 means a low probability of failure for that set. Based on a model threshold of 0.5, the output is classified into 'non-failure' ($\sigma(z) \leq 0.5$) or 'failure' ($\sigma(z) > 0.5$).

3.4 Probabilistic analysis

In probabilistic analysis, the probability of failure P_f for a given failure criterion can be estimated by means of crude Monte Carlo Simulations (MCS). Considering the use of surrogate modelling within MCS, each parameter set ('sample') gives a logit, which is turned into a sigmoid in the range ⟨0, 1⟩ and subsequently into the classification 'non-failure' or 'failure'. If n samples are evaluated resulting in m failures (and $n–m$ non-failures), the probability of failure for the given failure criterion is estimated as $P_f = m / n$.

To obtain good accuracy in MCS for low P_f-values, a (very) large number of samples need to be evaluated. For example, to be able to predict a P_f in the order of 10^{-4}, given a coefficient of variation (*COV*) of 5%, order 10^7 sample evaluations are needed. This makes MCS rather inefficient for low failure probabilities, albeit quite robust.

First Order Reliability Method (FORM) and Importance Sampling (IS) are two alternative techniques to estimate the probability of failure, giving similar accuracy with a lesser number of sample evaluations, especially for low P_f-values. In FORM, the limit state function (defining the boundary between 'failure' and 'non-failure') is linearized and the probability of failure is approximated in what is called the 'design point'. It is an optimization process that uses gradients to arrive at the design point. Importance Sampling is an improvement of MCS to reduce the variance and hence improve the accuracy around the limit state. After an initial estimate of the limit state with a modest number of samples, IS continues by sampling around the estimated limit state, using the 'likelihood ratio' of the reduced search space within the original distribution as weighting factor to evaluate the failure probability.

Both MCS and Importance Sampling with predictions from a trained surrogate model were utilized to calculate the probability of failure for the Bergambacht case. Separate surrogate models were trained for two failure criteria as obtained from the results of phase 7 of the finite element calculations, namely:

1. Soil collapse occurring
2. Maximum displacement exceeding 0.20 m

The accuracy and convergence of the surrogate modelling was checked by varying the number of data points taken from the entire dataset to train a surrogate model, and by subsequently calculating the failure probability using MCS and IS with the respective surrogate. This was done for both failure criteria by evaluating and plotting *Recall* and *Precision* as classification

metrics (Figure 3a), as well as the failure probability P_f (Figure 3b); both as a function of the number of data points used for training, where:

$$Recall = \frac{TP}{TP + FN} \qquad Precision = \frac{TP}{TP + FP} \tag{6a,b}$$

and:

TP (True Positive) meaning: 'failure' correctly predicted
FP (False Positive) meaning: 'non-failure' predicted as 'failure'
FN (False Negative) meaning: 'failure' predicted as 'non-failure'

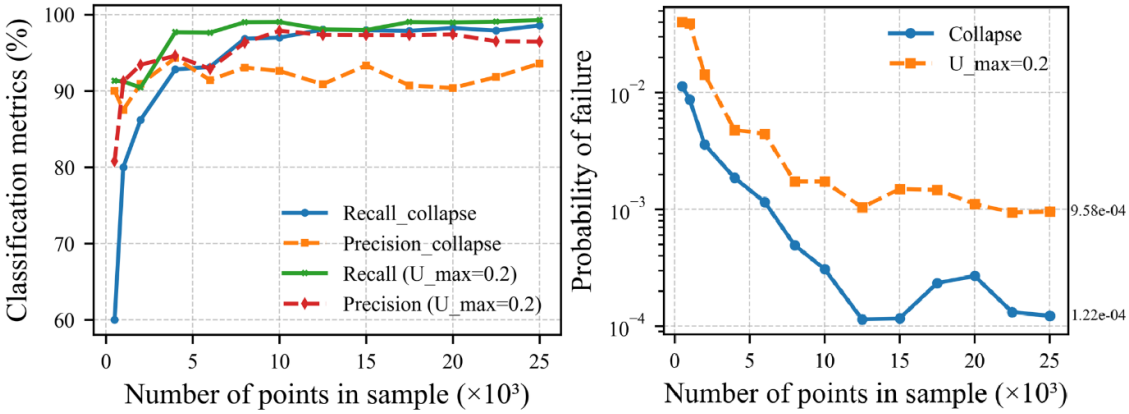


Figure 3. Accuracy and convergence of the solution
a) Classification metrics b) Probability of failure (log scale)

Figure 3 shows that a reasonable accuracy and convergence is obtained when using the entire dataset in the training process (24954 usable data points), although there is room for further improvement. Starting from an over-estimated P_f for a low number of training data points, the reduction of P_f seems to stabilize beyond 12.500 data points used for training.

A further comparison is made between the results from Importance Sampling considering $5 \cdot 10^5$ samples, and those from MCS considering $1 \cdot 10^7$ samples, when the surrogate is trained with a varying number of data points (Figure 4). Figure 4 shows that IS can obtain similar results as MCS with a significantly reduced number of samples (factor 20 less) used in the probabilistic analysis, and hence a significantly reduced calculation time.

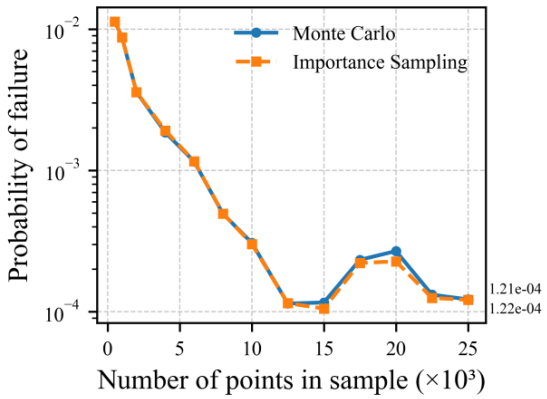


Figure 4. Comparison of P_f (soil collapse) from MCS (10^7 samples) and IS ($5 \cdot 10^5$ samples) using surrogates trained by a varying number of data points

4 Conclusions and recommendations

Surrogate modelling can be an efficient alternative for calculation processes involving large numbers of complex calculations. This article demonstrates the use of surrogate modelling in parameter selection for model fitting and probabilistic analysis. Based on the findings acquired in the research related to these topics, the following conclusions and recommendations are formulated:

- Rather than using actual (constitutive) model parameters, it is useful to define abstract parameters relating to the actual model parameters, to facilitate the selection of consistent parameter sets.
- Using sufficient data points and proper scaling of input and output features are essential in obtaining accurate surrogate results.
- Depending on the application, it is efficient to integrate the surrogate in other calculation processes in such a way that it can handle multiple parameter sets ('particles', 'samples', 'data points') at the same time.
- Gradient-based error measures in addition to *MAE*, *MSE* and R^2 can be helpful in model fitting-based parameter optimization.
- When using binary classification surrogates (involving output classes such as 'non-failure'/ 'failure'), it is useful to introduce logistic regression (using logit and sigmoid function) to improve the performance and predictive quality of the surrogate.
- Importance Sampling combined with surrogate modelling makes it feasible to perform finite element-based probabilistic analysis for geotechnical applications.

Further improvement on the use of surrogate models in processes involving parameter selection is subject of continuing research.

References

[1] Ahmad M., Brinkgreve R.B.J. and Jonkman S.N. (2025): Effects of displacement piles for dike reinforcement on adjacent buildings. Geotech. Geol. Eng. (2025) 43:446 (https://doi.org/10.1007/s10706-025-03438-y)

[2] Andersen K.H. (2015): Cyclic soil parameters for offshore foundation design. Frontiers in Offshore Geotechnics III, 5–82. Meyer (ed.). London: Taylor & Francis Group.

[3] Brinkgreve R.B.J., Engin E. and Engin H.K. (2010): Validation of empirical formulas to derive model parameters for sands. Numerical Methods in Geotechnical Engineering 2010, T. Benz & S. Nordal (eds.) Leiden: CRC Press, 137-142.

[4] Del Brocco I., Brinkgreve R.B.J. and Bui T.A. (2026): An updated explicit approach to account for undrained cyclic loading effects in offshore monopile foundations. Int. Conf. on Advances and Innovations in Soft Soil Eng.

[5] Granitzer A.-N., Aas Nøst H., Monsås E., Erharter G. and Leo J. (2025): Surrogate modeling framework for cantilever sheet pile wall stability assessment: Case study campus Ullevål. Research Square. (https://doi.org/10.21203/rs.3.rs-8163979/v1)

[6] Grimstad G., Andresen L. and Jostad H.P. (2012): NGI-ADP: Anisotropic shear strength model for clay. Int. J. Num. Anal. Meth. Geomech., 36: 483-497. (https://doi.org/10.1002/nag.1016)
[7] Jostad H.P. and Andresen L. (2009): A FE procedure for calculation of displacements and Capacity of foundations subjected to cyclic loading. Proc. COMGEO I, Juan-les-Pins, Cote d'Azur, France.
[8] Jostad H.P., Grimstad G., Andersen K.H., Saue M., Shin Y. and You D. (2014): A FE procedure for foundation design of offshore structures–applied to study a potential OWT monopile foundation in the Korean Western Sea. Geotech. Eng. J. of the SEAGS & AGS-SEA, 45, 63-72. (https://doi.org/10.14456/seagj.2014.7)
[9] Mitelman A., Yang B., Urlainis A. and Elmo D. (2023): Coupling geotechnical numerical analysis with Machine Learning for Observational Method projects. Geosciences 13, 196. (https://doi.org/10.3390/geosciences13070196)
[10] Morris, M. D. (1991): Factorial sampling plans for preliminary computational experiments. Technometrics, 33(2), 161-174.
[11] Plaxis (2018): PLAXIS 2D Material Models Manual.
[12] POV Macrostabiliteit (2020): POVM Eindige-elementenmethode. (www.povmacrostabiliteit.nl)
[13] Ragni R., Sivasithamparam N., Jostad H.P., Zhang Y., Okuno T., Sugiyama H. and Fukutake K. (2025): Efficient modelling of cyclic degradation in monopile design. 9th Int. Conf. on Offshore Site Investigation Geotechnics, London, 1050-1056 (2023). https://doi.org/10.3723/NYJT4115)
[14] Ruiz López A., Taborda D.M.G., Tsiampousi A., Pedro A.M.G., and Hardy S. (2024): Applying machine learning to the development of surrogate models for shafts in clay. Proc. XVIII Eur. Conf. on Soil Mech. and Geotechnical Eng. Taylor & Francis. 795-798. (https://doi.org/10.1201/9781003431749-133)
[15] Schanz T., Vermeer P.A. and Bonnier P.G. (1999): The hardening-soil model: Formulation and verification. In: Beyond 2000 in Computational Geotechnics. Rotterdam: Balkema, 281-290.

Author

Dr.ir. Ronald B.J. Brinkgreve
Delft University of Technology
Faculty of Civil Engineering and Geosciences
Department of Geoscience & Engineering, Geo-Engineering Section
Stevinweg 1
2628 CN Delft
The Netherlands
Tel.: +31 (0)15 - 2783327
E-mail: r.b.j.brinkgreve@tudelft.nl
Web: https://www.tudelft.nl/staff/r.b.j.brinkgreve