Conference Proceedings

Workshop

Numerical Methods in Geotechnics

AI in geotechnics, deformation based design, numerical optimization

17^{th} & 18^{th} of September 2026
Hamburg, Germany

Editor

J. Grabe

Hamburg University of Technology
Institute of Geotechnical Engineering and Construction Management

Publications by the Institute of
Geotechnical Engineering and Construction Management

64

Editor:

Univ.-Prof. Dr.-Ing. Jürgen Grabe
Hamburg University of Technology
Institute of Geotechnical Engineering and Construction Management
Schellerdamm 22 (Channel 7)
D – 21079 Hamburg
e-mail: *grabe@tuhh.de*

ISBN-13: 978-3-936310-66-5 (first edition)

DOI: 10.15480/882.18143

License: CC BY 4.0

Printing company

Druckzentrum Neumünster GmbH
Rungestraße 4
D – 24537 Neumünster

Published in the same series:

1. J. Grabe (eds.), 2000: Verbrennungsrückstände. Conference Proceedings, ISBN 3-936310-00-9
2. J. Grabe (eds.), 2001: Schaden- und Risikomanagement im Tiefbau. Conference Proceedings, ISBN 3-936310-01-7
3. J. Grabe, 2003. Bodenmechanik und Grundbau. ISBN 3-936310-03-3
4. J. Grabe (eds.), 2003: Euronormen in der Geotechnik – Was ändert sich? Conference Proceedings, ISBN 3-936310-04-1
5. J. Grabe (eds.), 2003: Bodenverdichtung, Experimente - Modellierung - Geräteentwicklung - Baustellenberichte - F+E-Bedarf. Conference Proceedings, ISBN 3-936310-05-X
6. M. Kelm, 2004: Nummerische Simulation der Verdichtung rolliger Böden mittels Vibrationswalzen. Dissertation, ISBN 3-936310-06-8
7. J. Grabe (eds.), 2004: Kaimauern - Messungen und Numerik. Conference Proceedings, ISBN 3-936310-07-6
8. J. Stein, 2005. Experimentelle und numerische Untersuchungen zum Düsenstrahlverfahren. Promotion, ISBN 3-936310-09-2
9. J. Grabe (eds.), 2005: Grenzschicht Wasser und Boden - Phänomene und Ansätze. Conference Proceedings, ISBN 3-936310-10-6
10. J. Grabe (eds.), 2005: FEM in der Geotechnik - Qualität, Prüfung, Fallbeispiele - Conference Proceedings, ISBN 3-936310-11-4
11. B. Mardfeldt, 2006: Zum Tragverhalten von Kaikonstruktionen im Gebrauchszustand. Dissertation, ISBN 3-936310-12-2
12. J. Grabe (eds.), 2006: Optimierung in der Geotechnik - Strategien und Fallbeispiele. Conference Proceedings, ISBN-13: 978-3-936310-13-9
13. T. Bierer, 2007: Bodenschwingungen aus Straßenverkehr auf unebener Fahrbahn im Zeitbereich - experimentelle und theoretische Untersuchungen. Dissertation, ISBN-13: 978-3-936310-14-6
14. J. Grabe (eds.), 2007: Bemessen mit Finite-Elemente-Methoden. Conference Proceedings, ISBN-13: 978-3-936310-15-3
15. K.-P. Mahutka, 2008: Zur Verdichtung von rolligen Böden infolge dynamischer Pfahleinbringung und durch Oberflächenrüttler. Dissertation, ISBN-13: 978-3-936310-16-0

16. J. Grabe (eds.), 2008: Seehäfen für Containerschiffe zukünftiger Generationen. Conference Proceedings, ISBN-13: 978-3-936310-17-7

17. F. König, 2008: Zur zeitlichen Traglastentwicklung von Pfählen und der nachträglichen Erweiterung bestehender Pfahlgründungen. Dissertation, ISBN-13: 978-3-936310-18-4

18. S. Henke, 2008: Herstellungseinflüsse aus Pfahlrammung im Kaimauerbau. Dissertation, ISBN-13: 978-3-936310-19-1

19. J. Grabe (eds.), 2009: Spundwände – Profile, Tragverhalten, Bemessung, Einbringung und Wiedergewinnung. Conference Proceedings, ISBN-13: 978-3-936310-20-7

20. J. Dührkop, 2009: Zum Einfluss von Aufweitungen und zyklischen Lasten auf das Verformungsverhalten lateral beanspruchter Pfähle in Sand. Dissertation, ISBN-13: 978-3-936310-21-4

21. O. Möller, 2009: Zum Langzeit-Kompressionsverhalten weicher organischer Sedimente. Dissertation, ISBN-13: 978-3-936310-22-1

22. J. Grabe (eds.), 2011: Ports of container ships of future generations. Conference Proceedings, ISBN-13: 978-3-936310-23-8

23. S. Kinzler, 2011: Zur Parameteridentifikation, Entwurfs- und Strukturoptimierung in der Geotechnik mittels numerischer Verfahren. Dissertation, ISBN-13: 978-3-936310-24-5

24. G. Qiu, 2012: Coupled Eulerian Lagrangian Simulations of Selected Soil-Structure Problems. Dissertation, ISBN-13: 978-3-936310-25-2

25. X. Ma, 2013: Nutzung der oberflächennahen Geothermie mittels Energiepfählen und Erdwärmesonden. Dissertation, ISBN-13: 978-3-936310-26-9

26. J. Grabe (eds.), 2013: Proceedings of the Conference on Maritime Energy COME 2013. Conference Proceedings, ISBN-13: 978-3-936310-28-3

27. J. Grabe (eds.), 2013: Bemessen mit numerischen Methoden. Conference Proceedings, ISBN-13: 978-3-936310-29-0

28. T. Pucker, 2013: Stoffmodell zur Modellierung von stetigen Materialübergängen im Rahmen der Optimierung geotechnischer Strukturen. Dissertation, ISBN-13: 978-3-936310-30-6

29. S. Henke, 2013: Untersuchungen zur Pfropfenbildung infolge der Installation offener Profile in granularen Böden. Habilitation, ISBN-13: 978-3-936310-31-3

30. J. Grabe (eds.), 2014: Ports for Container Ships of Future Generations. Conference Proceedings, ISBN-13: 978-3-936310-32-0

31. J. Grabe (eds.), 2014: Offshore Basishäfen, Conference Proceedings, ISBN-13: 978-3-936310-33-7

32. C. Rudolph, 2015. Untersuchungen zur Drift von Pfählen unter zyklischer, lateraler Last aus veränderlicher Richtung, Dissertation, ISBN-13: 978-3-936310-34-4

33. J. Grabe (eds.), 2015: Morphodynamics 2015, Conference Proceedings, ISBN-13: 978-3-936310-35-1

34. T. Hamann, 2015: Zur Modellierung wassergesättigter Böden unter dynamischer Belastung und großen Bodenverformungen am Beispiel der Pfahleinbringung, Dissertation, ISBN-13: 978-3-936310-36-8

35. B. Schümann, 2015: Beitrag zum dynamischen Dreiphasenmodell für Boden auf Basis der Finite-Elemente-Methode, Dissertation, ISBN-13: 978-3-936310-37-5

36. M. Milatz, 2015: Untersuchungen zum Einfluss der Kapillarität auf das hydraulisch-mechanische Verhalten von granularer Tragschichten für Verkehrswege, Dissertation, ISBN-13: 978-3-936310-38-2

37. H. Kaya, 2016: Bodenverschleppung und Spaltbildung infolge der Einbringung von Profilen in Dichtungsschichten aus Ton, Dissertation, ISBN-13: 978-3-936310-39-9

38. J. Grabe (eds.), 2017: Proceedings of the Conference on Maritime Energy COME 2017. Conference Proceedings, ISBN-13: 978-3-936310-40-5

39. B. Kocak, 2017: Zur numerischen Modellierung von hydraulisch-mechanisch gekoppelten Prozessen in gesättigten granularen Böden mittels Smoothed Particle Hydrodynamics, Dissertation, ISBN-13: 978-3-936310-41-2

40. K. Siegl, 2017: Zur Pfahldynamik von gerammten Großrohrpfählen und der daraus resultierenden Wellenausbreitung in Wasser und im Meeresboden, Dissertation, ISBN-13: 978-3-936310-42-9

41. J. Grabe (eds.), 2017: Numerical Methods in Geotechnics, Conference Proceedings, ISBN-13: 978-3-936310-43-6

42. J. Grabe (eds.), 2018: Digitale Infrastruktur und Geotechnik (DIG 2018), Conference Proceedings, ISBN-13: 978-3-936310-44-3

43. D. Osthoff, 2018: Zur Ursache von Schlosssprengungen und zu einbringbedingten Lageabweichungen von Spundwänden, Dissertation, ISBN-13: 978-3-936310-45-0

44. E. Heins, 2018: Numerical based identification of the pile-soil interaction in terms of the axial pile bearing capacity, Dissertation, ISBN-13: 978-3-936310-46-7

45. K.-F. Seitz, 2021: Zur Topologieoptimierung von geotechnischen Strukturen und zur Tragfähigkeitssteigerung des Baugrunds durch Scherfugenverfestigung, Dissertation, ISBN-13: 978-3-936310-47-4

46. D. Plenker, 2021: Physical and numerical investigations of the dynamic interaction of saturated granulates and fluid, Dissertation, ISBN-13: 978-3-936310-48-1

47. J. Grabe, J.-O. Backhaus & P. Vogel, 2021: Bauprojektmanagement, ISBN-13: 978-3-936310-49-8

48. M. Kanitz, 2021: Experimental and numerical investigations of particle-fluid systems in geotechnical engineering, Dissertation, ISBN-13: 978-3-936310-50-4

49. J.-O. Backhaus, 2021: A methodology for the numeric time-cost forecast and pareto optimization of large injection projects in tunneling, Dissertation, ISBN-13: 978-3-936310-51-1

50. S. N. Sinduri, 2021: Optimisation of deep compaction as liquefaction mitigation measure, Dissertation, ISBN-13: 978-3-936310-52-8

51. J. Bubel, 2022: Zum Versagen von Unterwasserböschungen im Seegang, Dissertation, ISBN-13: 978-3-936310-53-5

52. J. Grabe (eds.), 2022: Adaptive Structures at Shore, Conference Proceedings, ISBN-13: 978-3-936310-54-2

53. J. Grabe (eds.), 2022: Numerische Methoden in der Geotechnik, Conference Proceedings, ISBN-13: 978-3-936310-55-9

54. J. Grabe (eds.), 2022: Processes in natural and technical Particle-Fluid Systems (PintPFS), Conference Proceedings, ISBN-13: 978-3-936310-56-6

55. T. Törzs, 2022: Zum hydraulisch-mechanisch gekoppelten Verhalten teilgesättigter granularer Geomaterialien infolge hydraulischer und mechanischer Beanspruchungen, Dissertation, ISBN-13: 978-3-936310-57-3

56. M. Milatz, 2022: Investigation of capillary effects on the grain scale by means of in situ experiments, imaging and numerical simulations, Habilitation, ISBN-13: 978-3-936310-58-0

57. A. Chmelnizkij, 2023: Regularized MPM for porous media, Dissertation, ISBN-13: 978-3-936310-59-7

58. J. Grabe (ed.), 2023: Herausforderungen in der Geotechnik – 25 Jahre Institut für Geotechnik und Baubetrieb, TU Hamburg, Conference Proceedings, ISBN-13: 978-3-936310-60-3

59. J. Beuße, 2023: Zur Einbringung und zum Tragverhalten von kombinierten Spundwänden von Ufereinfassungen, Dissertation, ISBN-13: 978-3-936310-61-0

60. D. Zobel, 2023: Zur Implementierung höherwertiger Stoffmodelle am Beispiel der Hypoplastizität, Dissertation, ISBN-13: 978-3-936310-62-7

61. A. Hagemann, 2024: On cyclic and dynamic pile-bearing behaviour, Dissertation, ISBN-13: 978-3-936310-63-4

62. P. Kaminski, 2024: Experimental investigations on the mechanics of gassy sands – testing methodology, shear tests, and imaging, Dissertation, ISBN-13: 978-3-936310-64-1

63. J. Grabe (Ed.), 2025: smartGeotech 2025 - Künstliche Intelligenz, Automatisierung, Robotik, Conference Proceedings, ISBN-13: 978-3-936310-65-8

Table of Contents / Inhaltsverzeichnis

Preface/Vorwort 1

Geomechanics-informed artificial neural network for high-cyclic soil response 3
M. Tafili, R. Polo-Mendoza, J. Duque, D. Mašín and T. Wichtmann

Data and thermodynamics are all you need:
Constitutive model discovery under hard constraints 7
Filippo Masi

Insights into soil nonlinearity using surrogate models-based back-analyses
of geotechnical structures 9
David M. G. Taborda, António M. G. Pedro, Agustín Ruiz Lópe

AI-based prediction of soil behaviour using triaxial and oedometer data 23
Mohamed Abdennadher, Jürgen Grabe

Smarter Geotechnical Coding with AI 37
Patrick Staubach

An approach of BIM to FEM-coupling to support automation
in geotechnical engineering 49
Sascha Henke, Johannes Beck

Parametric finite element analyses incorporating stochastic
ground models for uncertainty propagation 59
Tobias Peterstorfer, Johannes Leo, Franz Tschuchnigg

Towards displacement-based design of quay walls 69
Luca Flessati

A comparative study of analytical and numerical based
design approaches for rigid inclusionss 81
Prerana Krishnaraj, Christopher Tinat

Installation method influence on the prediction of long-term monopile deformations –
A simplified numerical approach 93
Viet Hung Le, Frank Rackwitz

On the use of surrogate modelling in parameter selection 113
Ronald B.J. Brinkgreve, Majd Ahmad

Numerical study on the influence of mesh dependency on shear planes in Finite Element Method 129
Mazdak Atri, Tim Pucker

From Data to Design: Insights from the SAGE-SAND and EXCALIBUR Projects and a Class A Blind Prediction Test 141
Stijn François, George Anoyatis, Hadrien Rattez, Frédéric Collin, Kristof Maes, Luc Simonin, Michail Spyridis, Shijin Li, Pauline André, Thrasyvoulos-Konstantinos Elezoglou

Linking numerical modelling to construction work: challenges and opportunities 149
Raul Fuentes

Generative Design with Plaxis 157
Jia Lin

Saving time and money with mathematical optimisation - applied optimisation in geotechnical engineering 167
Jörg Meier

Surrogate Modelling for a Deep Excavation Problem with Increasing Feature and Target Complexity 179
Johannes Leo, Thomas Elmer, Franz M. Rohrhofer, Franz Tschuchnigg

A data-driven perspective on the vibrodriving process: Modelling, optimisation and control 193
Francisco Williams-Riquer, Mechthild Cramer

Physics enhanced and data driven modeling applied to structural and wave dynamics 205
Svenja Ehlers, Kerstin Kracht, Norbert Hoffmann

Application of the phase-field method for resource-efficient ground improvement through shear band strengthening 217
Elnaz Hadjiloo, Jürgen Grabe

Preface/Vorwort

In the past few decades, numerical methods have emerged as a key tool in the geotechnical engineering. Even though it is now state-ofthe-art to deepen our understanding of soil mechanics and soil-structure interaction, there remains a pressing need for further development. With that in mind, I wish to outline three areas that I believe needs further investigation and development.

1. **Artificial intelligence in geotechnics**
   AI is changing our world in numerous fields, and geotechnics is no exception. It provides an opportunity to improve or potentially replace the constitutive models, geotechnical designs, and procedures.

2. **Deformation based design**
   The traditional geotechnical design relies on limit state analysis and safety factors. However, most structures fail not because of reaching the limit state, but because of unacceptable deformations or settlements, for example. The FEM is well-suited for the prediction of such deformations under serviceability limits, so why is it not used routinely in geotechnical designs? Perhaps it is time to reconsider our design philosophy.

3. **Numerical optimization**
   Geotechnical design and construction are largely driven by experience and trial-and-error. This process can be optimized by the application of mathematics and control methods, which in turn leads to a reduction of costs, CO_2 emissions, and time duration of the project.

The aim of this workshop is to advance the role of numerical methods in geotechnical engineering by rethinking how we model, design, and optimize, moving from experience driven, limit state based practice toward approaches that are more predictive, deformation aware, and computationally efficient. By bringing together perspectives on artificial intelligence, deformation based design, and numerical optimization, the workshop seeks to stimulate research that closes the gap between advanced numerical modeling and routine practice, ultimately supporting more reliable, cost effective, and sustainable geotechnical solutions.

We hope this workshop sparks meaningful discussion and that all participants feel welcome and at ease.

Jürgen Grabe

organized by Stephanie Richter, Jeannine Eßmann, Milena Bauersachs and Lennart Sekulic

Hamburg, den 17. September 2026

Geomechanics-Informed Artificial Neural Network for High-Cyclic Soil Response

M. Tafili, R. Polo-Mendoza, J. Duque, D. Mašín and T. Wichtmann

Abstract: High-cyclic loading can generate relevant permanent deformation even when the strain accumulated per cycle is small. Laboratory databases are indispensable but too sparse to train transferable neural constitutive surrogates. We combine experiments, a calibrated high-cycle accumulation (HCA) model and a geomechanics-informed artificial neural network (GANN). Experimental small data are expanded into mechanics-based synthetic big data, while admissibility constraints and bounded outputs guide learning. A generalized GANN trained on 2.2 million simulations for 22 sands receives loading conditions, state variables and basic soil descriptors instead of soil-specific HCA parameters. Blind predictions for two excluded sands show generalization without material-specific recalibration.

1 Motivation and concept

Offshore wind foundations and other infrastructure may experience millions of load cycles. Direct discretization of each cycle and subsequent integration of an incremental constitutive model is computationally expensive. HCA models instead describe permanent-strain accumulation from cyclic amplitude, loading history, state and average stress (Niemunis et al., 2005). This enables long-term analysis, but material parameters normally require dedicated tests for every soil.

Machine learning provides fast mappings, but high-cyclic datasets are small and unconstrained extrapolation can violate basic trends. We assign complementary roles: experiments define admissible behavior, the HCA model (Niemunis et al., 2015) generates systematic variation, and the neural network learns under geomechanical restrictions. Figure 1 summarizes the path from experimental small data to mechanics-based synthetic big data, geomechanics-informed AI and unseen soils.

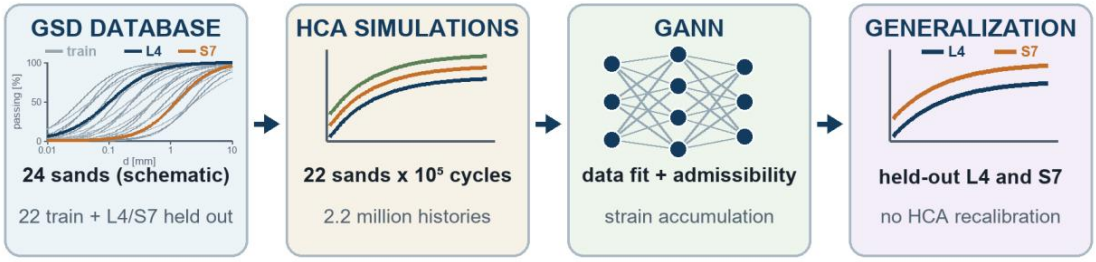


Figure 1: Schematic workflow from diverse sand grain size distributions (GSDs) through HCA-generated data and geomechanics-informed learning to blind predictions for L4 and S7 sands.

2 Geomechanics-informed learning framework

Soil-specific HCA calibrations for Karlsruhe Fine Sand (Wichtmann et al., 2014, Wichtmann et al., 2015) and Australian superfine silica sand (Canales-Brenlla et al., 2026) first are used to generate broad loading scenarios. The GANN (Tafili et al., 2026; Polo-Mendoza et al., 2026a; Polo-Mendoza et al., 2026b) maps state and loading variables to accumulated strain. Its loss penalizes decreasing accumulation, and output transformations bound the response. The accumulation direction retains the experimentally supported HCA flow rule linking the volumetric-to-deviatoric proportion mainly to average stress ratio (Wichtmann et al., 2014). Mechanics thus enters through data generation, input-output design and loss formulation.

The generalized model replaces seven soil-specific HCA constants by characteristic grain sizes, grading indices, minimum and maximum void ratios, and critical-state friction angle. The database contains 24 sands. L4 and S7 sands stay outside model development; for each of the other 22 sands, 100,000 admissible scenarios are simulated. The resulting 2.2 million simulations span material, state and loading variability, while the held-out sands are used to validate the material generalization.

3 Generalization and outlook

Across different sands, the generalized GANN reproduces the strain accumulation predicted by the HCA model very well. More than 99% of the evaluated cases reach R^2 values between 0.8 and 1.0. Even more importantly, the model also performs well for the two sands L4 and S7, which were not used during training. For these sands, the GANN captures the measured strain accumulation and its dependence on the loading conditions without using their soil-specific HCA parameters. The interpretability analyses also show physically reasonable trends: the coefficient of uniformity influences the behaviour at the beginning of cyclic loading, while the coefficient of curvature becomes more influential at larger numbers of cycles.

The framework therefore uses a limited amount of experimental data in a more efficient way and reduces the need for extensive soil-specific testing. However, its applicability is still limited to the range of soils and loading conditions represented in the database, and the current feed-forward network has no explicit memory of the loading history. Future work will therefore extend the database, investigate architectures that can better account for loading history, and implement the surrogate model in finite-element analyses. In this framework, experiments provide the physical basis, constitutive modelling expands the available data in a controlled way, and geomechanics-informed machine learning enables fast and more general predictions reducing the need for new experiments.

References

[1] Niemunis A., Wichtmann T. and Triantafyllidis T. (2005): A high-cycle accumulation model for sand. *Computers and Geotechnics*, 32:245-263.

[2] Wichtmann T., Niemunis A. and Triantafyllidis T. (2014): Flow rule in a high-cycle accumulation model backed by cyclic test data of 22 sands. *Acta Geotechnica*, 9:695-709.

[3] Wichtmann T., Niemunis A. and Triantafyllidis T. (2015): Improved simplified calibration procedure for a high-cycle accumulation model. *Soil Dynamics and Earthquake Engineering*, 70:118-132.
[4] Canales-Brenlla, L., Koch, F., Tafili, M., Prada-Sarmiento, L. F., & Wichtmann, T. (2026). Mechanical behaviour of Australian superfine silica sand: from monotonic to high-cyclic loading. *Acta Geotechnica*, 1-21.
[5] Tafili, M., Polo-Mendoza, R., Duque, J., Mašín, D., & Wichtmann, T. (2026). A Geomechanics-Informed Artificial Neural Network for High-Cyclic Soil Response – Part I: Core Model Architecture. *Soil Dynamics and Earthquake Engineering*, submitted.
[6] Polo-Mendoza, R., Tafili, M., Duque, J., Mašín, D., & Wichtmann, T. (2026a). A Geomechanics-Informed Artificial Neural Network for High-Cyclic Soil Response – Part II: Model Generalization. *Soil Dynamics and Earthquake Engineering*, submitted.
[6] Polo-Mendoza, R., Tafili, M., Duque, J., Mašín, D., & Wichtmann, T. (2026b). Towards physics-consistent machine learning models: A geomechanics-based artificial neural network for high-cyclic soil response. In *E3S Web of Conferences* (Vol. 730, p. 03007). EDP Sciences.

Author

Univ.-Prof. Dr.-Ing. Merita Tafili
Ruhr-Universität Bochum
Lehrstuhl für Bodenmechanik, Grundbau und Umweltgeotechnik
Universitätsstraße 150
44801 Bochum
Tel.: +49 (0) 234 32-17596
e-mail: merita.tafili@rub.de
Web: https://www.bgu.ruhr-uni-bochum.de/bgu/lehrstuhl/team/tafili.html.de

Dr. Rodrigo Ernesto Polo Mendoza
Charles University
Institute of Hydrogeology, Engineering Geology and Applied Geophysics
Albertov 2038/6
128 00 Praha, Czech Republic
Tel.: +420-2-2195 1552
e-mail: polomenr@natur.cuni.cz
Web: https://is.cuni.cz/webapps/whois2/osoba/1926791818556571/?lang=en

Prof. Dr. Jose Duque
Universidad de la Costa
Civil and Environmental Faculty
Calle 58 #55 – 66,
Barranquilla, Colombia
Tel.: +57 3052822632
e-mail: jduque@cuc.edu.co
Web: https://www.researchgate.net/profile/Jose-Duque-9

Prof. Dr. David Mašín
Charles University
Institute of Hydrogeology, Engineering Geology and Applied Geophysics

Albertov 2038/6
128 00 Praha, Czech Republic
Tel.: +420-2-2195 1552
e-mail: masin@natur.cuni.cz
Web: https://web.natur.cuni.cz/~masin/

Univ.-Prof. Dr.-Ing. habil. Torsten Wichtmann
Ruhr-Universität Bochum
Lehrstuhl für Bodenmechanik, Grundbau und Umweltgeotechnik
Universitätsstraße 150
44801 Bochum
Tel.: +49 (0) 234 32-26135
e-mail: torsten.wichtmann@rub.de
Web: https://www.bgu.ruhr-uni-bochum.de/bgu/lehrstuhl/team/wichtmann.html.de

Data and thermodynamics are all you need: Constitutive model discovery under hard constraints

Filippo Masi

Author

Filippo Masi
Univ. Grenoble Alpes, Inria,
CNRS, Grenoble INP, Institute of Engineering, LJK
38000 Grenoble, France
e-mail: filippo-masi@inria.fr

Insights into soil nonlinearity using surrogate models-based back-analysis of geotechnical structures

David M. G. Taborda, António M. G. Pedro, Agustín Ruiz López

Abstract: The use of surrogate models in the back-analysis of geotechnical structures is explored, using the deep excavation for Tottenham Court Road station (London, UK) as a case study. Artificial neural networks are trained to reproduce the wall displacements computed by finite element analyses, with a principal component analysis-based approach drastically reducing training times without loss of accuracy. A sensitivity study reveals that wall movements at each stage are controlled by the soil stiffness over a narrow, progressively increasing strain range. A genetic algorithm-based back-analysis of the nonlinear stiffness of London Clay shows that, while the tangent stiffness curve cannot be uniquely identified, the operational secant stiffness is consistently recovered, allowing the calibration of a conventional model.

1 Introduction

The design and construction of deep excavations in urban environments require accurate predictions of ground movements and structural forces, which are strongly influenced by the highly nonlinear stress-strain behaviour of the ground. Field monitoring data collected during construction provide a valuable opportunity to improve these predictions through back-analysis, whereby the parameters of the adopted numerical model are adjusted until the computed response matches the observed one. This process underpins the observational method (Peck, 1969), which has been successfully applied to deep excavations to achieve significant savings in construction time and cost (e.g. Yeow et al., 2014). However, the practical application of back-analysis to complex geotechnical structures is often hampered by the computational cost of the underlying finite element analyses, with optimisation algorithms typically requiring a very large number of evaluations of the numerical model. Surrogate models – computationally inexpensive approximations of the full numerical model – offer a means of overcoming this limitation, enabling extensive parametric studies, design optimisation, sensitivity analyses and (near) real-time back-analysis (e.g. Ferrero et al., 2023; Ruiz López et al., 2025).

A further difficulty concerns the interpretation of the back-analysed parameters. When advanced constitutive models capable of reproducing the variation of soil stiffness with strain level are employed, it is generally assumed that a single set of parameters can reproduce the observed behaviour across all stages of construction. In simpler constitutive models, this assumption may not be valid, and different sets of parameters may be required for different project

phases (Chen, 2018; Foo et al., 2024). This paper explores this problem by using surrogate models based on artificial neural networks (ANNs) in the back-analysis of the excavation for Tottenham Court Road station, London, inferring directly the stress-strain response of London Clay, i.e. without pre-imposing a specific constitutive model.

2 Case study

2.1 Details of the analysis

To illustrate the procedure of back-analysis using surrogate models, the case of the Tottenham Court Road (TCR) station excavation is considered herein. This station, located in the centre of London, United Kingdom, was extensively redeveloped as part of the Crossrail project (now the Elizabeth Line). The excavation is approximately $41\ m \times 31\ m$ in plan and $30\ m$ deep, and is supported by $40\ m$-deep, $1.0\ m$-thick diaphragm walls. The geotechnical profile is typical of central London: $8.5\ m$ of superficial deposits ($4.2\ m$ Made Ground, MG, and $4.3\ m$ of Terrace Gravels, TG), followed by $22.2\ m$ of London Clay (LC), $19.5\ m$ of Lambeth Group Clay (LGC) and $3.5\ m$ of Thanet Sand (TS). Chalk is found below $53.7\ m$ but is not included in the numerical models. The struts, as reported by Kwang and Chen (2025), consist of hollow circular steel sections, spaced at $4\ m$ in the out-of-plane direction and installed without prestress. Figure 1 illustrates the construction sequence and stratigraphy, highlighting the phases for which wall movement data were collected (Chen, 2018); Phase 1 (activation of the wall beam elements) and the prop installation phases (3, 5, 7 and 9) are omitted.

Two distinct piezometric levels are observed: the shallower coincides with the top of the TG stratum, while the deeper was, at the time of construction, close to the top of the TS stratum (Yeow et al., 2014). The permeability of LC and LGC – the only undrained materials in this stratigraphy – is simulated with the User-Defined Flow Model (UDFM) IC MAGE FM01 (Taborda et al., 2023), reproducing the variation proposed by Avgerinos et al. (2016) ($\log_{10} k_v = 0.0215\,(y - y_{LC}) - 9.427$, where y_{LC} is the elevation of the top of the LC layer), based on data reported in Hight et al. (2007). These hydraulic properties, together with the two piezometric levels described above, result in the underdrained pore water pressure profile shown in Figure 2(a).

The initial stress state is generated based on unit weight values, γ, of $18\ kN/m^3$ for the MG material and $20\ kN/m^3$ for the remaining materials, and the profiles of pore water pressure and at-rest coefficient of earth pressure, K_0, represented in Figure 2(a) (Gawecka et al., 2017). Assuming plane strain conditions and exploiting symmetry, only one half of the problem is modelled, with the left boundary coinciding with the axis of symmetry and the right boundary located $87.5\ m$ from the wall, sufficiently far not to influence the results. Horizontal displacements are restrained along the vertical boundaries and all displacements prevented at the base, located at a depth of $53.7\ m$. The domain is discretised into 5511 15-node triangular elements (Bentley Systems, 2025), refined near the wall, where larger stress gradients are expected, and progressively coarser towards the right and bottom boundaries. The diaphragm wall is simulated by elastic beam elements ($E = 30\ GPa, \nu = 0.2$), with its interaction with the soil being modelled using interface elements with a Mohr-Coulomb failure criterion ($R_{int} = \tan\varphi'_{int}/\tan\varphi'_{soil} = c_{int}/c_{soil} = 2/3$ and $\psi_{int} = 0°$). The stiffness of these elements is determined following the procedure outlined in Bentley Systems (2025).

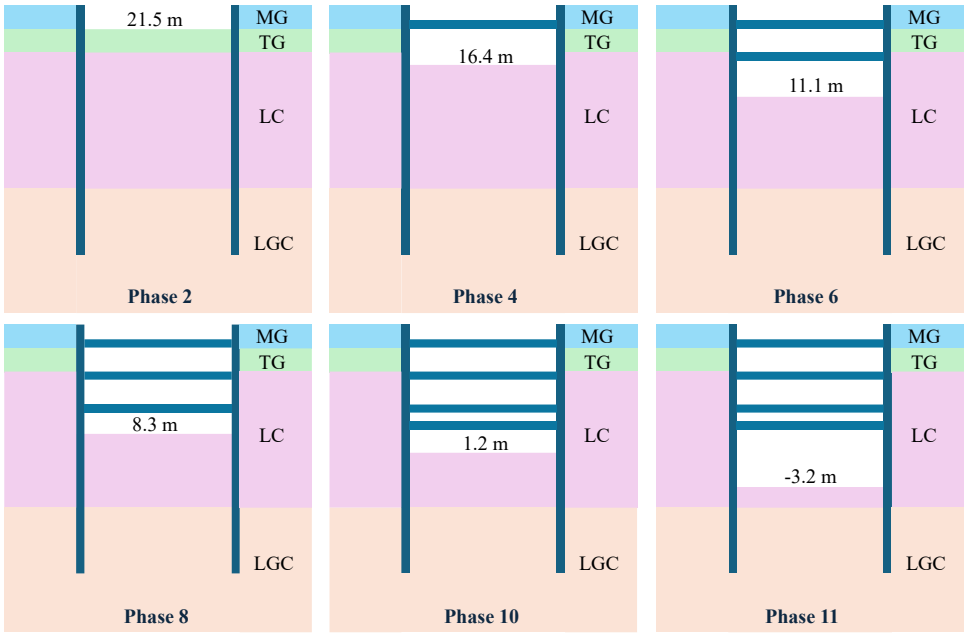
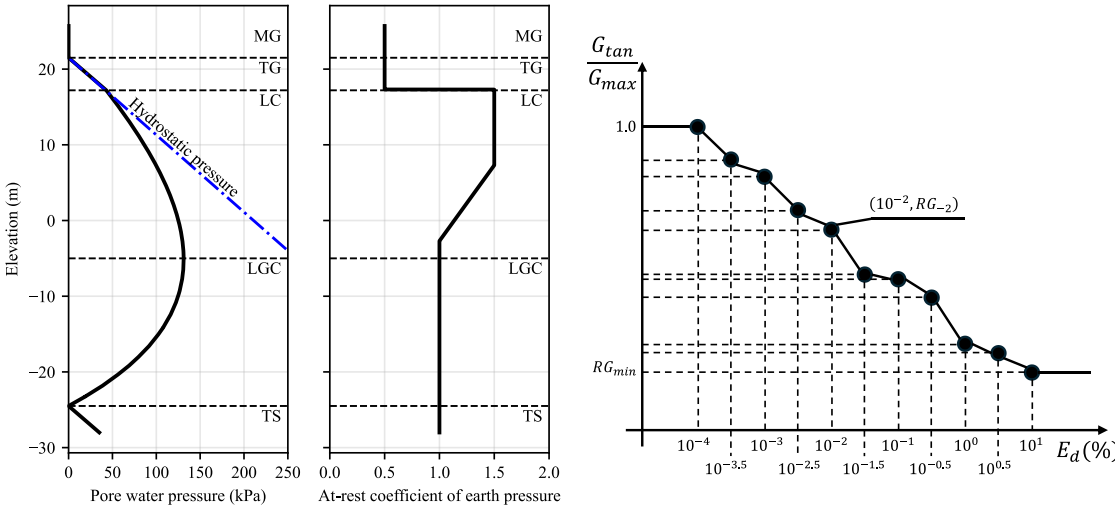


Figure 1: Construction sequence adopted in Tottenham Court Road station.



(a) (b)

Figure 2: (a) initial conditions of the deposit in terms of pore water pressure distribution and K_0 and (b) illustration of the elastic shear modulus simulated by the IC MAGE M34 model.

2.2 Soil modelling

The Made Ground (MG) is simulated as linear elastic-perfectly plastic, with its strength defined by the Mohr-Coulomb failure criterion and non-associated plasticity ($E = 10\,MPa, \nu = 0.2$, $\varphi' = 30°, \psi = 0°$). The behaviour of TG, LGC and TS is simulated using the User-Defined Soil Model (UDSM) IC MAGE M01 (Taborda et al., 2024), which combines the hypoelastic model proposed by Taborda et al. (2016) with the Mohr-Coulomb failure criterion. According

to this model, the tangent shear modulus, G_{tan}, varies with the mean effective stress, p', and the second invariant of the strain tensor, E_d:

$$G_{tan} = G_{ref}\left(\frac{p'}{p'_{ref}}\right)^{m_G}\left(RG_{min} + \frac{1 - RG_{min}}{1 + \left(\frac{E_d}{a}\right)^{b}}\right) \tag{1}$$

where G_{ref} is the maximum shear modulus for a reference mean effective stress, p'_{ref}, m_G is a parameter defining the nonlinearity of the increase in stiffness with mean effective stress, and a, b and RG_{min} are parameters characterising the reduction in stiffness with the strain level, E_d. The bulk modulus is defined by a similar formulation, with the volumetric strain being used instead of the deviatoric strain level, while parameters K_{ref}, m_K, r, s and RK_{min} replace G_{ref}, m_G, a, b and RG_{min}, respectively.

The behaviour of LC, which has the greatest influence on the response of the retaining wall, since approximately two-thirds of the excavation takes place in this material, is described by the UDSM IC MAGE M34 (Taborda, 2026). This model is analogous to M01 but, instead of a smooth function being used for defining the elastic shear modulus, this property is obtained using $G_{tan} = G_{ref}\left(p'/p'_{ref}\right)^{m_G} RG(E_d)$, where $RG(E_d)$ is defined through linear interpolation between a set of discrete points $(E_{d,i}, RG_i)$ equally distributed along a logarithmic axis – see Figure 2(b):

$$RG(E_d) = \begin{cases} 1, & E_d \le 10^{-4}\% \\ RG_{i-1} + (RG_i - RG_{i-1})\dfrac{\left(\log_{10} E_d - \log_{10} E_{d,i-1}\right)}{\left(\log_{10} E_{d,i} - \log_{10} E_{d,i-1}\right)}, & E_{d,i-1} \le E_d \le E_{d,i} \\ RG_{min}, & E_d \ge 10\% \end{cases} \tag{2}$$

3 Surrogate model

3.1 Generation of surrogate models and their evaluation

A set of six surrogate models was generated for the TCR station, one for each of the phases shown in Figure 1 (Phases 2, 4, 6, 8, 10 and 11). The first phase of this process involves defining the sampling of the parameter space. The number of samples required is difficult to determine, depending on the dimensionality of the space (higher dimensionality requires, in principle, more samples) and the nonlinearity of the function being approximated (steeper gradients require higher sample density). In the present case, Latin Hypercube Sampling (LHS) (McKay et al., 1979) is employed, in which each dimension is divided into N intervals (N being the desired number of samples, in the present case $N = 2000$) and one parameter combination is assigned to each, ensuring more uniform coverage than simple random sampling irrespective of sample count. Other alternative approaches include sequential sampling using MaxPro (see e.g. Ruiz López et al. (2025)) and adaptive sampling, such as CV-BASHES (Yang et al., 2026).

All parameters (stiffness of soil and struts, and $K_{0,top}$, see Figure 2(a)) were set to the values determined following a large-scale back-analysis reported in Taborda et al. (2026) for TCR, with the exception of those defining the stiffness of LC. The nonlinear behaviour of this material is the object of the present study and is therefore described using M34, with each of its 11 parameters (G_{ref} and the 10 values RG_i defining the normalised stiffness curve at strain values between $10^{-4}\%$ and 10%, as shown in Figure 2(b)) becoming one dimension of the parameter space. However, given the need to ensure that the stiffness is monotonically decreasing, the LHS was used to sample the reduction rate, generating the 2000 tangent shear modulus reduction curves for LC shown in Figure 3(a), with $G_{ref} \in [25\ MPa, 150\ MPa]$. These were automatically run and post-processed in PLAXIS 2024.3 using its Python API (Bentley Systems, 2025), extracting the wall displacements at 20 uniformly distributed points. After the completion of this process, all calculations are verified to ensure they completed successfully, with unsuccessful analyses being removed from the database. In the present study, a success rate of 100% was achieved and the resulting profiles are displayed in Figure 3(b), where it can be seen that the highest spreads of predictions are concentrated in elevations between $y = 0\ m$ and $y = 15\ m$, roughly coinciding with the location of the LC layer, as expected. Given that Phase 6 is the first where the excavation consists exclusively of LC, it is perhaps unsurprising that the spread of wall movement predictions almost doubles in this stage (from about $10\ mm$ to $20\ mm$), further increasing in subsequent stages, reaching over $30\ mm$ for Phase 11.

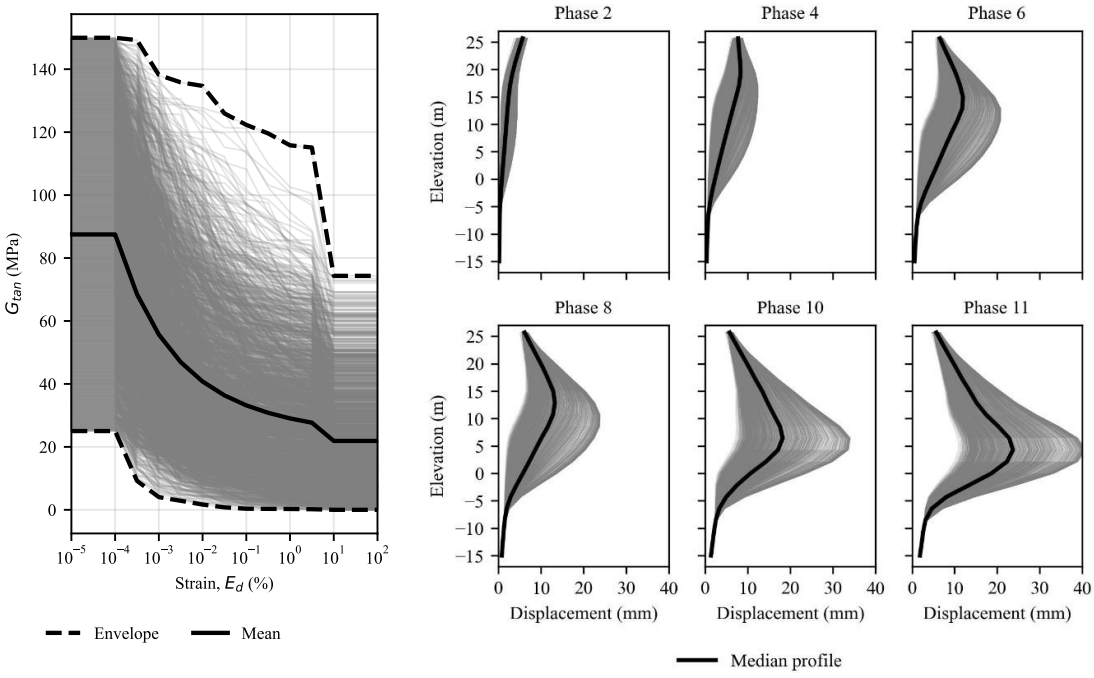


Figure 3: Generated database for training the surrogate models in terms of (a) shear stiffness curves for $p' = p'_{ref}$ and (b) wall displacement profiles.

All generated surrogate models are feedforward multi-layer perceptrons implemented in PyTorch v2.5.1 (Ansel et al., 2024), using the $FlexibleANN$ class. Two distinct approaches are adopted to investigate the potential gains in terms of training time arising from reducing the dimensionality of the output space for the specific case of a retaining wall. In the first approach, denoted $Full\ ANN$, the depth is added as an input and therefore each ANN predicts a single

output (displacement) based on 12 parameters (depth, G_{ref} and 10 values of RG_i). In the second approach, denoted $ANN\ PCA$, a principal component analysis (PCA) is first performed on the output database of the numerical model. The database is then rebuilt using only the first k components, to identify how many are required to explain a given percentage of the original variance. Given that wall displacement profiles are generally smooth and free of noise, a relatively small number of components is expected to be needed to achieve excellent accuracy. However, it is also likely that, since nonlinearity increases with excavation depth, the number of required components is unlikely to remain constant between stages. In the present case, 99% of the variance of the original database can be obtained using two components for phases 2 and 4, and three components for all subsequent phases. Conservatively, four components are used for the entire set of results. Therefore, in the $ANN\ PCA$ approach, the 11 inputs (G_{ref} and 10 values of RG_i) are used to predict the 4 outputs ($k = 4$ resulting from the PCA). The outputs are then inverted into wall displacements. This reduction yields smaller, faster-training networks: instead of a database of 2000 analyses × 20 elevations, only $2000\ analyses \times 4$ components are required.

For both sets of $ANNs$, the inputs and outputs are normalised using the $StandardScaler$, which imposes zero mean and unit variance. The loss function combines the mean squared error (MSE) with a bending-energy regularisation term weighted by a coefficient λ, the latter applied to the $Full\ ANN$ only. This term penalises the curvature of the wall displacement profile, thereby discouraging oscillatory behaviour that would be physically implausible for a displacement field in a continuum medium. The surrogates are trained using the Adam optimiser (Kingma and Ba, 2015), with L2 regularisation and early stopping to ensure better generalisation. Candidate combinations are drawn randomly, without replacement, from a hyperparameter grid spanning both deeper (more hidden layers) and wider (more neurons per layer) architectures, different activation functions, learning rates, dropout rates and batch sizes. Given the high computational effort required, the training is performed using an NVIDIA GPU with CUDA acceleration. Each combination is evaluated using five-fold cross-validation and, for the $Full\ ANN$, a stratified approach at the sample level is used to avoid displacement profiles of a given parameter combination appearing simultaneously in the training and validation folds. To ensure an unbiased assessment of performance, 20% of the samples were reserved from the outset of the process, being used exclusively for testing.

The trained $ANNs$ obtained in the two approaches have markedly different architectures. For the $Full\ ANN$, the selected hidden layer sizes are [4096,2048,1024] for phases 2, 4, 6, 8 and 10, and [1024,1024,1024,1024] for Phase 11, with training times ranging from $560\ s$ to $2500\ s$. For the $ANN\ PCA$ approach, smaller networks suffice, with hidden layers [512,256,128] for phases 2, 4, 10 and 11 and [128,64,32] for phases 6 and 8, with training times between $8\ s$ and $15\ s$. All R^2 values are above 0.999 for the testing database in both approaches.

Beyond these aggregate metrics, a second stage of evaluation consists in determining the R^2 value for each test sample and analysing its distribution. For each phase, the worst sample, as well as the samples corresponding to the 5th, 25th and 50th percentiles (i.e. the median error) of R^2 are selected. As proposed in Ruiz López et al. (2025), the underlying idea is that, if the errors associated with these samples are tolerable, it is possible to rapidly establish the proportion of samples for which the surrogate model performs well in terms of reproducing the original PLAXIS analyses. For brevity, only the results for Phase 11 (largest displacements) are shown (Figure 4), with the errors for the $ANN\ PCA$ being used to rank the samples. The errors are small, indicating that, even for samples to which the training process had no access, the

surrogate models perform well, both when using the larger networks ($Full\ ANN$) and smaller ones ($ANN\ PCA$). Given that no appreciable difference in performance is observed between the two sets of surrogate models, the latter will be used hereinafter.

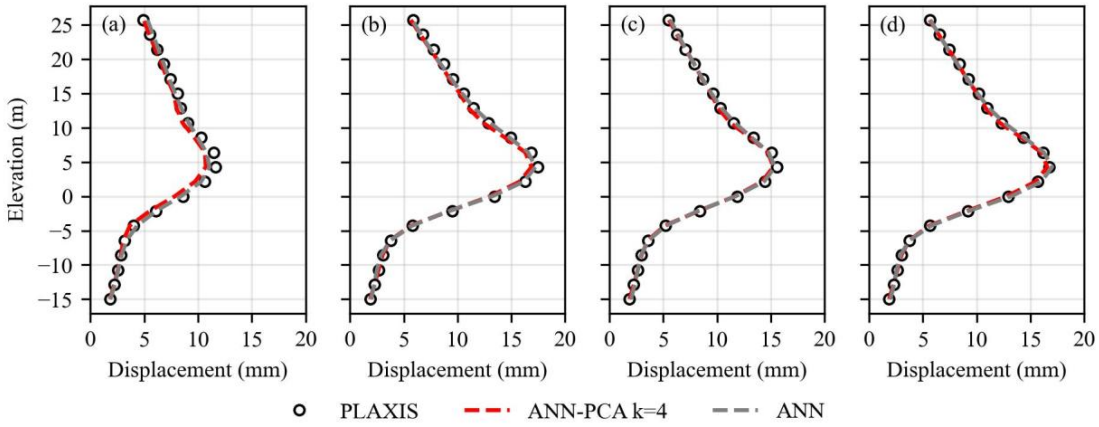


Figure 4: Comparison between the original results (PLAXIS) and the two trained neural networks ($Full\ ANN$ and $ANN\ PCA$ with $k = 4$) for Phase 11, corresponding to (a) the worst sample and the sample at the (b) 5th, (c) 25th and (d) 50th percentile of the R^2 distribution.

3.2 Sensitivity study

The negligible evaluation time of the trained surrogates ($<1s$), compared with that of the finite element analyses they approximate, enables extensive parametric studies of wall displacement, reliability-based analysis and design optimisation. Here, a sensitivity study using Sobol indices is carried out to illustrate the use of surrogate models in the interpretation of the behaviour of the simulated excavation. Based on a decomposition of the output variance into contributions attributable to each input, the method defines two indices: the first-order index S_i, quantifying the contribution of each parameter i to the total variance of the output when considered in isolation (Sobol', 2001), and the total-order index ST_i, which includes all additional interactions of that parameter with the remaining ones (Homma and Saltelli, 1996). A high ST_i therefore identifies an influential parameter, whether acting directly or through interactions, and the difference $ST_i - S_i$ measures the weight of those interactions. The Sobol indices are calculated in Python using the SALib library, via the Saltelli method (Saltelli et al., 2010).

Given that the values of parameters RG_i are normalised quantities, it is expectable that G_{ref} necessarily dominates, scaling the entire shear modulus reduction curve. For the depth-averaged Sobol indices, which give one value per phase, a fixed value of $G_{ref} = 100\ MPa$ is adopted, avoiding the distortion caused by the in-built importance of this parameter. The results, shown in Figure 5, reveal the role of soil nonlinearity in the wall response: as the excavation progresses, the strain range over which the stiffness governs wall displacements evolves from $10^{-2.5} - 10^{-2}$ % (Phase 2) to $10^{-1.5} - 10^{-1}$ % (phases 10 and 11). Moreover, it suggests, rather unambiguously, that the stiffness values at all other strain levels are unimportant to the obtained displacements, leading to an important conclusion: without multiple excavation stages, the back-analysis of stiffness parameters will be necessarily limited to a restricted strain range. This is the object of the following section.

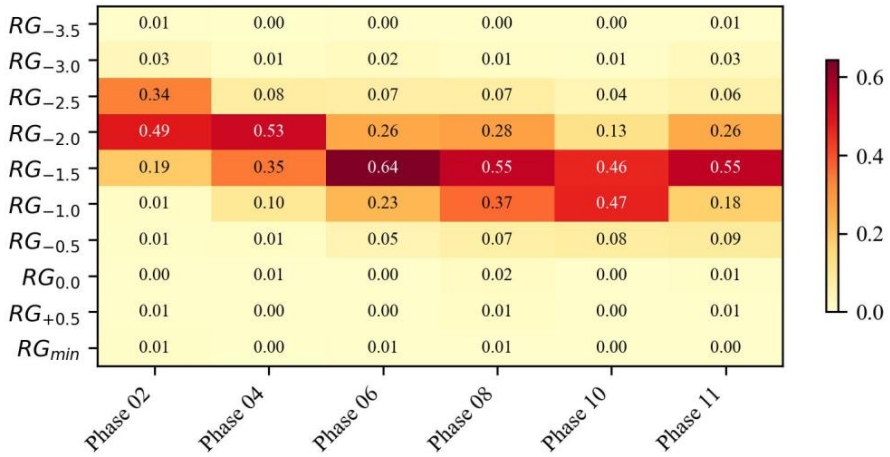


Figure 5: Depth-averaged values of ST_i for the various phases (assuming $G_{ref} = 100\ MPa$).

4 Back-analysis

4.1 Procedure

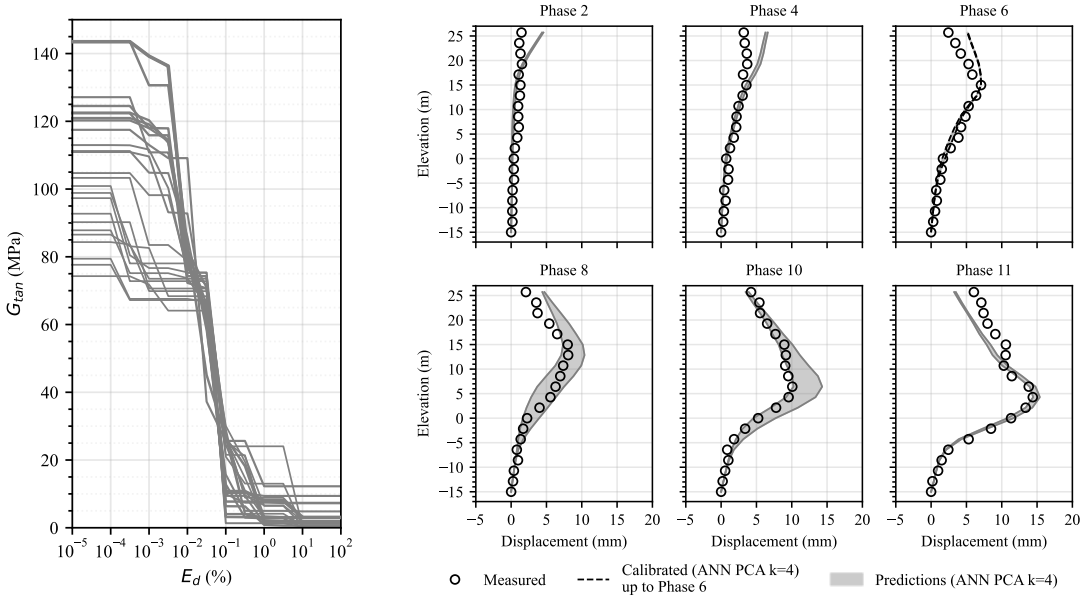
With the reduction in computation time achieved through the replacement of finite element models by surrogate models, it becomes possible to carry out the back-analysis of highly complex geotechnical structures in (near) real time, supporting, for example, the implementation of the observational method (Peck, 1969; Yeow et al., 2014; Chen et al., 2018; Ferrero et al., 2023). In the present case, the coupling of an optimisation algorithm with the developed surrogate models is intended to produce a back-analysis of the stress-strain response of the LC layer by trying to match the wall movements observed in the excavation. It is important to note that a certain degree of bias has already been introduced by setting the stiffness parameters for the struts and the remaining soils, as well as the at-rest coefficient of earth pressure to values that lead to a good reproduction of wall movements.

The back-analysis was performed by coupling the developed $ANN\ PCA$ surrogate models with a genetic algorithm implemented in the IC MAGE GenAlgo library (Ruiz López, 2024). This algorithm seeks the set of 11 parameters that minimises a weighted error metric that contains both the RMSE between the displacement profiles predicted by the surrogate model and the profiles measured in the field, and the absolute error measured at the depth of greatest wall movement. For simplicity, equal weights are given to these errors when assessing the fitness of each candidate solution. Moreover, due to the little impact that the behaviour of LC has on the response of the wall during the earlier excavation stages, only the field data collected for phases 6 to 11 are used. Lastly, two additional modifications were introduced that are specific to the problem being analysed: first, a penalty is introduced for any candidate solution that produces non-monotonic stiffness reduction curves; second, to limit the importance of G_{ref}, the genetic algorithm is defined such that it calibrates directly the value of the tangent shear modulus at the 11 strain levels, which are subsequently converted to $RG_i = G_{tan,i}/G_{ref}$ when running the surrogate models ($RG = 1$ for $E_d = 10^{-4}\%$). The search space is inferred from the sampling procedure and is represented by the dashed lines in Figure 3(a). To simulate the back-analysis process, a progressive approach is adopted: calibration is initiated considering only Phase 6 as the target and, subsequently, additional phases are incorporated incrementally until all four

phases are included. For each subset of active phases, the genetic algorithm is run 25 times with independent random initialisations, in order to evaluate the variability of the results.

4.2 Results

The detailed results of the adopted procedure are shown for Phase 6 in Figure 6, where it can be seen that there is a considerable variability in the obtained stiffness reduction curves (Figure 6(a)), particularly at low and large strain values, which agrees with the conclusions drawn from the Sobol indices study in Figure 5. Conversely, at intermediate strains ($10^{-2.5} - 10^{-1.5}$%), all the candidate solutions appear to converge. In terms of displacement profiles (Figure 6(b)), a good agreement between all the candidates and the measured displacements (Phase 6) can be observed, which is unsurprising given that the fitness of the various candidate solutions is evaluated based on these data. For subsequent phases, some spread is visible for phases 8 and 10, though the predicted range always encompasses the measured displacements. For Phase 11, a good prediction of the field data is obtained, and the range of predictions is narrower, perhaps reflecting the fact that the behaviour of LC has a diminishing influence for greater depths.

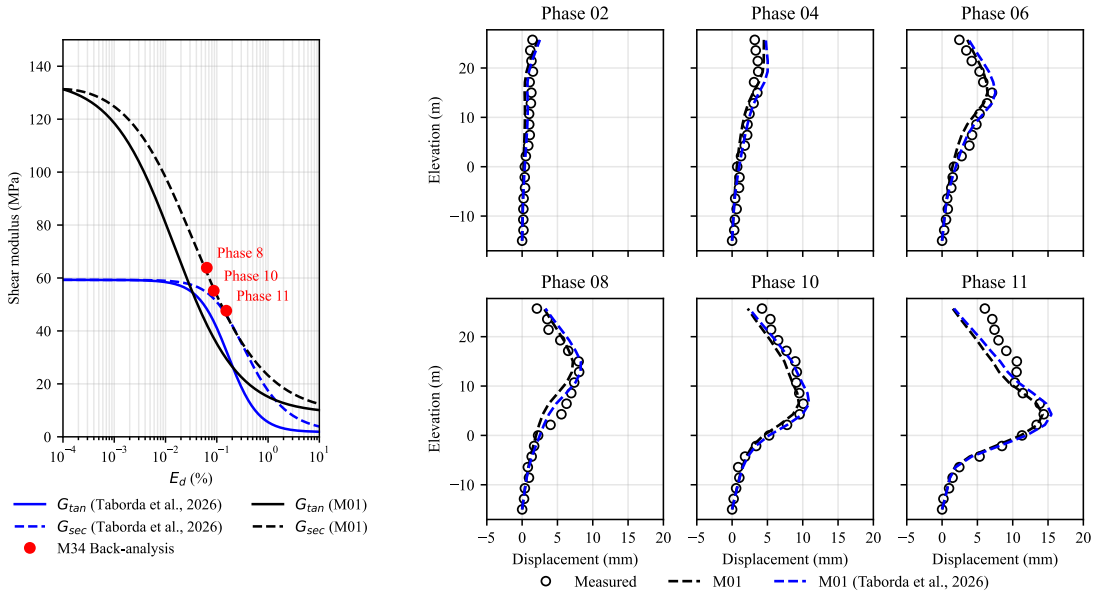


(a) (b)

Figure 6: Calibration using field data measured in Phase 6 in terms of (a) stiffness reduction curves and (b) corresponding wall profiles for all phases.

A further interpretation of the concept of 'operational stiffness' may suggest that, rather than variations of tangent shear moduli with strain, secant stiffness moduli should be considered. In effect, this quantity denotes the average stiffness from the start of the analysis and therefore encompasses, at least to a certain extent, the entire stress-strain history of the material up to the current state. Thus, for each phase, the obtained tangent shear moduli variations are integrated to calculate the corresponding stress-strain curve, which is then converted into secant shear moduli reduction curves. The 25 candidate solutions are considered and the point of least spread is determined, i.e. the strain value for which the difference $\max(G_{sec,i}) - \min(G_{sec,i})$ reaches a minimum. It is argued herein that this convergence of the candidate solutions identifies the

'operational stiffness' for the phase in question. The three points corresponding to phases 8, 10 and 11 are illustrated in Figure 7(a) together with two sets of curves obtained using the continuous function described by Equation (1): the first, in black, is calibrated as best fit to the three datapoints, while the second, in blue, is the back-analysed stiffness reduction curve reported by Taborda et al. (2026). It is interesting to notice that, despite being independent studies of the same case study, there is a good agreement between the Taborda et al. (2026) curve and the three datapoints. The two sets of parameters for M01 (calibrated, in black, and reported by Taborda et al. (2026), in blue) are subsequently used in finite element analyses and the calculated wall displacements are illustrated in Figure 7(b). Clearly, despite very different initial shear modulus values (135.0 MPa vs. 59.3 MPa), the good agreement in terms of 'operational stiffness' means that the predictions are very similar. This observation further reinforces the difficulties in identifying objectively material parameters based on limited observations (in this case, wall displacements) of complex geotechnical structures simulated using constitutive models with inherently flexible formulations.



(a) (b)

Figure 7: (a) stiffness reduction factors obtained through back-analysis for phases 8, 10 and 11 and curves obtained using calibrated M01 parameters, and (b) modelled wall response.

5 Conclusions

This paper demonstrates the use of surrogate models in the back-analysis of a deep excavation, using the Tottenham Court Road station (London, UK) as a case study, focussing on establishing directly the stress-strain response of London Clay from the measured wall movements. To achieve this, a database of 2000 finite element analyses was first generated through Latin Hypercube Sampling of the parameters controlling the nonlinear stiffness of this material and was employed to train artificial neural networks reproducing the computed wall displacements at six construction stages. Two approaches were investigated: a $Full\ ANN$, in which depth is included as an additional input, and an $ANN\ PCA$ approach, in which the dimensionality of the output space is first reduced through principal component analysis. Both approaches achieved

excellent accuracy, with R^2 values above 0.999 for the testing database, while the latter required considerably smaller networks and training times of just a few seconds.

The negligible computational cost of the trained surrogate models enabled a comprehensive sensitivity study based on Sobol indices, which showed that the wall displacements at each construction stage are controlled by the soil stiffness over a narrow band of strain levels, shifting towards larger strains as the excavation progresses, while the stiffness at all other strain levels has little influence on the computed response. It follows that a single excavation stage can constrain the stiffness over only one narrow strain window, meaning that continuous monitoring data are essential if the stiffness reduction curve is to be recovered over any meaningful range – a requirement that deserves careful attention when this approach underpins the observational method.

Lastly, the surrogate models were coupled with a genetic algorithm to back-analyse the stiffness reduction curve of London Clay from the measured wall movements. Considerable variability was observed in the identified tangent stiffness curves at low and large strains, with convergence occurring only at intermediate strain levels, in agreement with the findings of the sensitivity study. When the results were interpreted in terms of secant stiffness, however, a well-defined 'operational stiffness' emerged at each stage, in good agreement with a back-analysis of the same case study (Taborda et al., 2026). Indeed, two sets of parameters with markedly different initial shear moduli produced wall displacement predictions that differed only marginally. Where a flexible constitutive formulation is calibrated against a single class of observation, then, parameter identifiability rather than predictive accuracy becomes the binding limitation.

Acknowledgements

This research is part of project SIDeTools-Smart Infrastructure Design Tools (EP/X52446X/1), funded by the Engineering and Physical Sciences Research Council (EPSRC, UK).

References

[1] Ansel, J., Yang, E., He, H., Gimelshein, N., Jain, A., Voznesensky, M., Bao, B., Bell, P., Berard, D., Burovski, E., Chauhan, G., Chourdia, A., Constable, W., Desmaison, A., DeVito, Z., Ellison, E., Feng, W., Gong, J., Gschwind, M., Hirsh, B., Huang, S., Kalambarkar, K., Kirsch, L., Lazos, M., Lezcano, M., Liang, Y., Liang, J., Lu, Y., Luk, C. K., Maher, B., Pan, Y., Puhrsch, C., Reso, M., Saroufim, M., Siraichi, M. Y., Suk, H., Zhang, S., Suo, M., Tillet, P., Zhao, X., Wang, E., Zhou, K., Zou, R., Wang, X., Mathews, A., Wen, W., Chanan, G., Wu, P., Chintala, S. (2024) PyTorch 2: Faster Machine Learning Through Dynamic Python Bytecode Transformation and Graph Compilation. Proceedings of the 29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS '24), Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3620665.3640366

[2] Avgerinos, V., Potts, D. M., Standing, J. R. (2016) The use of kinematic hardening models for predicting tunnelling-induced ground movements in London Clay. *Géotechnique*. 66 (2), 106–120. https://doi.org/10.1680/jgeot.15.p.035

[3] Bentley Systems (2025) PLAXIS 2D 2024.3 – Reference Manual.

[4] Chen, Y. (2018) Application of new observational method on deep excavation retaining wall design in London Clay. PhD Thesis. Department of Engineering, University of Cambridge.

[5] Chen, Y., Pillai, A. K., Nicholson, D. P., Biscontin, G. (2018) Back analysis of Crossrail deep excavations using three-dimensional finite element modelling. Proceedings of 9th European Conference on Numerical Methods in Geotechnical Engineering, Porto, Portugal.

[6] Ferrero, J. A., Ruiz López, A., Taborda, D. M. G., Brasile, S. (2023) Applying the observational method to a deep braced excavation using an artificial neural network. 10th European Conference on Numerical Methods in Geotechnical Engineering, London, UK. https://doi.org/10.53243/NUMGE2023-303

[7] Foo, C. F., Le, T., Bailie, P., Standing, J. R. (2024) Back-analysis of an embedded retaining wall in stiff clay. *Proceedings of the Institution of Civil Engineers – Geotechnical Engineering*. 177 (6), 581–600. https://doi.org/10.1680/jgeen.22.00072

[8] Gawecka, K. A., Taborda, D. M. G., Potts, D. M., Cui, W. J., Zdravkovic, L., Kasri, M. S. H. (2017) Numerical modelling of thermo-active piles in London Clay. *Proceedings of the Institution of Civil Engineers – Geotechnical Engineering*. 170 (3), 201–219. http://doi.org/10.1680/jgeen.16.00096

[9] Hight, D. W., Gasparre, A., Nishimura, S., Minh, N. A., Jardine, R. J., Coop, M. R. (2007) Characteristics of the London Clay from the Terminal 5 site at Heathrow Airport. *Géotechnique*. 57 (1), 3–18. https://doi.org/10.1680/geot.2007.57.1.3

[10] Homma, T., Saltelli, A. (1996) Importance measures in global sensitivity analysis of nonlinear models. *Reliability Engineering & System Safety*. 52 (1), 1–17. https://doi.org/10.1016/0951-8320(96)00002-6

[11] Kingma, D. P., Ba, J. L. (2015) Adam: A method for stochastic optimization. Proceedings of the 3rd International Conference on Learning Representations (ICLR 2015), San Diego, CA, USA.

[12] Kwang, C. W., Chen, Y. (2025) A method of isolating different components of prop axial force through FFT&BWF signal filters. *Proceedings of the Institution of Civil Engineers – Geotechnical Engineering*. 178 (5), 574–587. https://doi.org/10.1680/jgeen.24.00241

[13] McKay, M. D., Beckman, R. J., Conover, W. J. (1979) Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. *Technometrics*. 21 (2), 239–245. https://doi.org/10.1080/00401706.1979.10489755.

[14] Peck, R. B. (1969) Advantages and Limitations of the Observational Method in Applied Soil Mechanics. *Géotechnique*. 19 (2), 171–187. http://doi.org/10.1680/geot.1969.19.2.171

[15] Ruiz López, A. (2024) IC MAGE GenAlgo – Genetic Algorithms for Geotechnics (Version 1.0). Zenodo. https://doi.org/10.5281/zenodo.13998194

[16] Ruiz López, A., Taborda, D. M. G., Tsiampousi, A., Go, J., Dingle, H. (2025) A data-driven macroelement model for suction buckets in sand. 5th International Symposium on Frontiers in Offshore Geotechnics, Nantes, France. https://doi.org/10.53243/ISFOG2025-347

[17] Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., Tarantola, S. (2010) Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Computer Physics Communications*. 181 (2), 259–270. https://doi.org/10.1016/j.cpc.2009.09.018

[18] Sobol', I. M. (2001) Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation*. 55 (1), 271–280. https://doi.org/10.1016/S0378-4754(00)00270-6.

[19] Taborda, D. M. G. (2026) IC MAGE Model 34 – a Mohr-Coulomb failure criterion with discrete variation of nonlinear isotropic elastic stiffness (Version 1.0). Zenodo. http://doi.org/10.5281/zenodo.18757193

[20] Taborda, D. M. G., Kontoe, S., Tsiampousi, A. (2023) IC MAGE Flow Model 01 – Non-linearly varying anisotropic permeability (Version 1.0). Zenodo. https://doi.org/10.5281/zenodo.7796383

[21] Taborda, D. M. G., Kontoe, S., Tsiampousi, A. (2024) IC MAGE Model 01 – Strain-hardening/softening Mohr-Coulomb failure criterion with isotropic small strain stiffness (Version 2.2). Zenodo. https://doi.org/10.5281/zenodo.13891265.

[22] Taborda, D. M. G., Pedro, A. M. G., Matos Fernandes, M., Almeida e Sousa, J. (2026) Metamodels in geotechnical engineering: an application to the back-analysis of geotechnical structures. Geotecnia (in Portuguese, accepted).

[23] Taborda, D. M. G., Potts, D. M., Zdravkovic, L. (2016) On the assessment of energy dissipated through hysteresis in finite element analysis. *Computers and Geotechnics*. 71, 180–194. https://doi.org/10.1016/j.compgeo.2015.09.001

[24] Yang, Y., Ruiz López, A., Tsiampousi, A., Taborda, D. M. G. (2026) A novel adaptive sampling approach with batch selection for the automatic generation of surrogate models in geotechnical engineering. *Data-Centric Engineering*. 7, e2. https://doi.org/10.1017/dce.2025.10036

[25] Yeow, H.-C., Nicholson, D., Man, C.-L., Ringer, A., Glass, P., Black, M. (2014) Application of observational method at Crossrail Tottenham Court Road station, UK.

Proceedings of the Institution of Civil Engineers – Geotechnical Engineering. 167 (2), 182–193. https://doi.org/10.1680/geng.13.00035

Authors

Dr David M G Taborda
Imperial College London
Department of Civil & Environmental Engineering
Skempton Building
London, United Kingdom
Tel.: +44 (0) 207594 6033
e-mail: d.taborda@imperial.ac.uk
Web: https://profiles.imperial.ac.uk/d.taborda

Prof António M G Pedro
Department of Civil Engineering, ISISE, ARISE
University of Coimbra
Coimbra, Portugal
Tel.: +351 239 797 224
e-mail: amgpedro@dec.uc.pt
Web: https://apps.uc.pt/mypage/faculty/amgpedro

Dr Agustín Ruiz López
Seequent, the Bentley Subsurface Company
Delft, The Netherlands
e-mail: agustin.ruizlopez@seequent.com

AI-based prediction of soil behaviour using triaxial and oedometer data

Mohamed Abdennadher, Jürgen Grabe

Abstract: The reliable numerical simulation of granular materials strongly depends on the appropriate selection of material and contact parameters. In particle-based numerical models, these parameters are often difficult to determine directly from laboratory measurements because the macroscopic response of a granular material results from complex interactions between particle geometry, contact behaviour, friction, and packing structure. This study presents a methodology for calibrating numerical material parameters using laboratory test data in combination with artificially generated particle geometries.

In previous work, an artificial intelligence-based framework was developed to generate realistic three-dimensional particle geometries. The approach combines a point-cloud autoencoder with a Denoising Diffusion Probabilistic Model (DDPM) to learn the morphological characteristics of real particles and generate new particle shapes. The generated geometries provide the basis for the numerical samples investigated in the present study.

The main objective of the present work is to use experimental oedometer and triaxial test data to systematically calibrate the parameters of the numerical material model using artificial intelligence. A neural-network-based surrogate model is trained using numerical simulations performed with different combinations of material parameters. The trained model learns the relationship between numerical material parameters and the resulting macroscopic soil response. The experimental laboratory data are subsequently used as target data in an inverse optimization procedure to identify the parameter combination that best reproduces the observed soil behaviour.

The proposed methodology provides a link between AI-based particle generation, numerical simulation, and experimental soil testing. It offers the potential to reduce the computational effort associated with conventional trial-and-error calibration and provides a systematic approach for identifying material parameters from laboratory data.

1 Introduction

Numerical modelling has become an important tool for investigating the mechanical behaviour of granular materials. Particle-based approaches, such as the Discrete Element Method (DEM), provide

the possibility of representing the material at the particle scale and investigating the influence of particle geometry, packing structure, and contact interactions on the macroscopic response of the material.

However, one of the main challenges in numerical soil modelling is the selection of appropriate material and contact parameters. Parameters such as particle-particle friction, rolling resistance, contact stiffness, damping, and other contact properties have a direct influence on the simulated mechanical response. In many cases, these parameters cannot be obtained directly from conventional laboratory tests. Instead, they have to be estimated or calibrated by comparing numerical simulations with experimental observations.

This calibration process can be particularly challenging because several numerical parameters may influence the same macroscopic response. Consequently, manually changing individual parameters based on experience can require a large number of simulations and may not result in a unique or physically meaningful parameter set.

An additional challenge concerns the representation of particle geometry. Real granular materials contain particles with irregular and complex shapes. Simplifying these particles to spheres or other idealized geometries can influence the simulated packing and mechanical behaviour. Therefore, a realistic representation of particle morphology is important when the objective is to reproduce laboratory-scale soil behaviour numerically.

In previous work, an artificial intelligence-based framework was developed to generate realistic three-dimensional particle geometries. The framework uses real particle geometries as training data and combines a point-cloud autoencoder with a Denoising Diffusion Probabilistic Model (DDPM). The autoencoder provides a compact representation of the particle geometries, while the diffusion model learns the distribution of these representations and generates new particle geometries. The resulting particles represent newly synthesized geometries with morphological characteristics similar to the real particles.

The present study extends this approach by focusing on the calibration of the numerical material model. Instead of considering the artificial particles only as a method for generating geometries, they are incorporated into numerical laboratory tests. The simulated response is then compared with real oedometer and triaxial test data.

The main objective is to develop an AI-assisted workflow in which realistic artificial particles and experimental soil data are combined to automatically identify the material parameters of a numerical model.

The proposed workflow can be summarized as follows:

Real particles → AI-generated particles → Numerical simulations → AI surrogate model
→ Parameter optimization → Experimental validation.

2 AI-Based Generation of Artificial Particle Geometries

2.1 Generation Framework

The generation of artificial particles is based on a generative machine-learning framework developed in previous work. The original particle dataset consists of three-dimensional geometries of real particles. These geometries are converted into point clouds to provide a consistent representation for the neural networks. For each particle, 1000 points are sampled using Farthest Point Sampling. The point clouds are subsequently centred and normalized to obtain a common coordinate system. In addition to the spatial coordinates, geometric and topological information is incorporated into the representation. The first component of the framework is a point-cloud autoencoder. The encoder compresses the high-dimensional particle representation into a compact latent representation, while the decoder reconstructs the three-dimensional particle geometry. A Chamfer-distance-based loss function is used to evaluate the difference between the original and reconstructed point clouds. The second component is a Denoising Diffusion Probabilistic Model. The diffusion model is trained on the latent representations obtained from the autoencoder. During generation, the model starts from random noise and progressively removes the noise to obtain a new latent representation. This representation is subsequently passed through the decoder to generate a new three-dimensional particle. The previous study demonstrated that the resulting particles can represent new geometries that were not present in the original training dataset while maintaining morphological characteristics learned from the real particle dataset.

2.2 Application to Numerical Modelling

For the present study, the generated particle geometries are used to construct numerical samples for the simulation of laboratory tests. The use of artificially generated particles allows irregular particle geometries to be represented while maintaining control over the statistical characteristics of the particle population. The generated particles are assembled into numerical specimens that reproduce the dimensions and initial conditions of the corresponding laboratory samples. The artificial particle generation therefore forms the first AI component of the proposed methodology.

The second AI component, introduced in the following sections, is used to identify the numerical material parameters.

3 Numerical Model and Laboratory Data

3.1 Numerical Representation of the Soil

The numerical model used in this study is based on a particle-based representation of the soil. The particle geometries generated using the AI-based framework described in the previous section are used to construct numerical specimens. Although the particle geometry provides an important contribution to the macroscopic response, the mechanical behaviour of the numerical material is also controlled by a set of material and contact parameters. Depending on the selected numerical formulation, these parameters may include particle stiffness, friction coefficient, rolling resistance, damping, and other contact-related properties. A major challenge is that these parameters cannot generally be obtained directly from macroscopic laboratory measurements. Instead, they have to be identified indirectly by comparing numerical and experimental behaviour. This parameter identification problem is the main focus of the present study.

3.2 Oedometer Tests

Experimental oedometer data are used as one of the reference datasets for the calibration procedure. The oedometer test investigates the one-dimensional compression behaviour of the soil under constrained lateral deformation. The measured relationship between vertical stress and vertical strain provides information about the stiffness and compressibility of the material. The numerical model reproduces the dimensions, boundary conditions, and loading sequence of the laboratory test. The numerical response is recorded and compared with the corresponding experimental response.

3.3 Triaxial Tests

Triaxial test data provide additional information about the mechanical behaviour of the material under shear loading. In a triaxial test, the specimen is subjected to a confining pressure followed by axial loading. The resulting deviatoric stress, axial strain, and volumetric strain provide information about the stiffness, strength, and volumetric deformation behaviour of the soil. The numerical triaxial test is constructed using the same general stress conditions as the laboratory experiment. The use of

both oedometer and triaxial data is important because a parameter set calibrated against only one type of experiment may reproduce one aspect of the material behaviour while failing to reproduce another.

4 AI-Based Calibration of Numerical Material Parameters

4.1 Motivation for AI-Based Calibration

The calibration of numerical material parameters is commonly performed by manually modifying parameters and repeatedly running numerical simulations until an acceptable agreement with laboratory measurements is obtained. This procedure can become computationally expensive because a single numerical simulation may require a significant amount of computational time. Furthermore, several parameters can influence the same macroscopic response. The present study therefore proposes an AI-based approach for identifying the numerical material parameters. Instead of manually searching the parameter space, a neural network is trained to learn the relationship between the numerical material parameters and the resulting macroscopic response. The key idea is to use the numerical model itself as a data generator. A large number of numerical simulations are performed using different combinations of material parameters. The resulting simulated oedometer and triaxial responses form a training dataset for the machine-learning model.

The AI model learns the relationship

$$\boldsymbol{\theta} \rightarrow \mathbf{y}, \tag{1}$$

where $\boldsymbol{\theta}$ represents the numerical material parameters and $\mathbf{y}$ represents the resulting numerical response. Once trained, the AI model acts as a surrogate for the computationally expensive numerical model.

4.2 Generation of the Numerical Training Dataset

The first step of the proposed calibration procedure is the generation of a numerical training dataset. A physically reasonable range is defined for each material parameter that is expected to influence the numerical response. Parameter combinations are then generated within these ranges. For each parameter combination, numerical oedometer and triaxial simulations are performed. For example, the material parameter vector can be defined as

$$\boldsymbol{\theta} = [\mu, k_n, k_t, \mu_r, \ldots], \tag{2}$$

where μ represents the friction coefficient, k_n and k_t represent normal and tangential contact stiffness, and μ_r represents a rolling-resistance parameter. Each parameter combination results in a corresponding numerical response:

$$\mathbf{y} = [\sigma_o(\epsilon_o), q_t(\epsilon_t), \epsilon_v(\epsilon_t), \ldots], \tag{3}$$

where σ_o represents the oedometer stress response, q_t represents the triaxial deviatoric stress response, and ϵ_v represents the volumetric strain. Repeating this process for a large number of parameter combinations produces a database containing pairs of numerical parameters and simulated material responses.

4.3 Neural-Network Surrogate Model

A neural network is trained using the numerical simulation database. The input of the network consists of the numerical material parameters,

$$\mathbf{x} = [\theta_1,\theta_2, \ldots ,\theta_n], \tag{4}$$

while the output consists of characteristic points or features describing the simulated laboratory response,

$$\hat{\mathbf{y}} = f_{\mathrm{AI}}(\mathbf{x}). \tag{5}$$

The objective of the training process is to minimize the difference between the neural-network prediction and the corresponding numerical simulation. A mean squared error loss can, for example, be defined as

$$L = \frac{1}{N} \sum_{i=1}^{N} \left(y_i - \hat{y}_i\right)^2 . \tag{6}$$

After training, the neural network acts as a surrogate model of the numerical simulation. This means that the response of the numerical material can be predicted for a new parameter combination without performing a complete numerical simulation.

4.4 Inverse Identification of Material Parameters

The most important step of the proposed methodology is the use of the trained AI model for inverse parameter identification. The experimental laboratory measurements are used as the target response,

$$\mathbf{y}_{\mathrm{exp}}. \tag{7}$$

The objective is to find the parameter vector $\boldsymbol{\theta}$ for which the AI-predicted response is as close as possible to the experimental response. The parameter-identification problem can therefore be formulated as

$$\boldsymbol{\theta}^* = \underset{\boldsymbol{\theta}}{\operatorname{argmin}}\ J\left(f_{\mathrm{AI}}(\boldsymbol{\theta}), \mathbf{y}_{\mathrm{exp}}\right), \tag{8}$$

where f_{AI} is the trained neural network and J is an objective function describing the difference between the predicted and experimental responses. The objective function can combine the different laboratory tests:

$$J = w_o J_o + w_t J_t + w_v J_v, \tag{9}$$

where J_o represents the error in the oedometer response, J_t represents the error in the triaxial stress response, and J_v represents the error in the volumetric response. The weighting factors w_o, w_t, and w_v control the relative importance of the different experimental responses. The optimization algorithm searches the material parameter space to identify the parameter combination that minimizes the objective function. In this way, the AI is not only used to predict soil behaviour. It is used to identify the numerical material parameters required to reproduce the experimentally observed behaviour.

4.5 Iterative AI-Assisted Calibration

The complete calibration procedure can be implemented as an iterative loop. Initially, a parameter space is defined and numerical simulations are performed for different parameter combinations. These simulations are used to train the surrogate model. The trained AI model is then combined with an optimization algorithm to identify a parameter combination that is predicted to reproduce the experimental data. The resulting parameter combination is subsequently evaluated using the original numerical model. The numerical result is compared with the laboratory data. If a significant difference remains, the new numerical simulation can be added to the training dataset and the AI model can be retrained. The calibration procedure can therefore be summarized as:

1. Define physically reasonable ranges for the numerical material parameters.
2. Generate parameter combinations.
3. Perform numerical oedometer and triaxial simulations.
4. Store the material parameters and corresponding numerical responses.
5. Train the neural-network surrogate model.
6. Provide the experimental oedometer and triaxial data as target responses.
7. Use an optimization algorithm together with the trained AI model to identify promising material parameters.
8. Perform a full numerical simulation using the predicted parameters.
9. Compare the numerical results with the experimental data.
10. Add new numerical simulations to the training dataset if required.
11. Retrain the AI model and repeat the procedure until the desired agreement is achieved.

5 Training and Evaluation of the AI Model

5.1 Training Dataset

The quality of the AI-based calibration depends strongly on the numerical training dataset. The parameter ranges must be sufficiently broad to cover the physically reasonable region of the parameter space. At the same time, unrealistic parameter combinations should be avoided because they can reduce the efficiency of the training process. The numerical database is divided into training, validation, and testing subsets. The training data are used to optimize the neural-network weights. The validation dataset is used to monitor the generalization performance during training, while the testing dataset is kept separate and is used to evaluate the final surrogate model.

5.2 Prediction Accuracy

The objective of the surrogate model is not to permanently replace the numerical model. Instead, it provides a computationally efficient approximation that allows rapid exploration of the material

parameter space. The predicted and simulated oedometer and triaxial responses are therefore compared using appropriate error measures.

For example, the mean squared error can be expressed as

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2, \tag{10}$$

where y_i represents the numerical simulation result and $\hat{y}_i$ represents the prediction of the neural network. A sufficiently accurate surrogate model allows a large number of parameter combinations to be evaluated without performing a full numerical simulation for every combination.

6 Results

6.1 Overview of the Synthetic Demonstration

To demonstrate the feasibility of the proposed AI-assisted calibration methodology, a synthetic numerical dataset was considered covering a representative range of numerical material parameters. The purpose of this demonstration is to evaluate whether the proposed surrogate-model-based approach can reproduce the response of the numerical model and subsequently identify a parameter combination capable of reproducing a prescribed reference response.

The parameter space considered in the demonstration consisted of the inter-particle friction coefficient μ, the normal contact stiffness k_n, the tangential contact stiffness k_t, and the rolling-resistance coefficient μ_r. A total of 1,000 numerical parameter combinations were considered. The corresponding numerical responses consisted of oedometer compression curves, triaxial deviatoric stress-strain curves, and volumetric strain responses.

The reference response was generated using a predefined parameter combination that was subsequently treated as unknown during the inverse-identification stage. This provides a controlled proof-of-concept for the proposed methodology because the identified parameters can be compared with the parameters used to generate the reference response.

The results presented in this section should therefore be interpreted as a numerical proof-of-concept rather than as experimental validation.

6.2 AI Prediction of Numerical Responses

The numerical dataset was divided into training, validation, and testing subsets using a ratio of 70:15:15. The neural network was trained to approximate the relationship between the numerical material parameters and the resulting macroscopic responses. The trained surrogate model demonstrated a high level of agreement with the numerical simulations contained in the independent testing dataset. The coefficient of determination was approximately $R^2 = 0.98$ for the oedometer response and $R^2 = 0.97$ for the triaxial deviatoric stress response. The volumetric strain response showed a slightly larger prediction error, with an R^2 value of approximately 0.95. The mean absolute error remained relatively small for all three response quantities. The results indicate that the neural network was able to capture the nonlinear relationship between the material parameters and the macroscopic response of the numerical model.

6.3 Inverse Identification of Material Parameters

After training the surrogate model, the reference response was used as the target for inverse parameter identification. The optimization algorithm searched the parameter space using the AI model rather than repeatedly evaluating the full numerical model. The reference parameter combination consisted of a friction coefficient of 0.50, a normal contact stiffness of 1.00×10^6, a tangential contact stiffness of 5.00×10^5, and a rolling-resistance coefficient of 0.10. The optimization procedure identified a friction coefficient of approximately 0.48. The normal contact stiffness was identified as approximately 1.04×10^6, while the tangential contact stiffness was approximately 4.85×10^5. The identified rolling-resistance coefficient was approximately 0.11. The relative differences between the reference and identified parameters were approximately 4% for the friction coefficient, 4% for the normal stiffness, 3% for the tangential stiffness, and 10% for the rolling-resistance coefficient. Although the rolling-resistance parameter showed the largest relative difference, the resulting macroscopic response remained close to the reference response. This indicates that different combinations of microscopic parameters can potentially produce similar macroscopic behaviour.

6.4 Optimization Behaviour

The inverse optimization procedure was evaluated by monitoring the combined objective function during the search. The initial parameter combinations produced relatively large differences between the surrogate prediction and the reference response. As the optimization progressed, the objective function decreased continuously and reached a stable minimum after approximately 80 optimization iterations. The objective function decreased from an initial normalized value of approximately 1.00

to approximately 0.06. This corresponds to a reduction of approximately 94% in the overall response error. The rapid decrease during the first iterations indicates that the surrogate model was able to provide useful information about the direction of the parameter search. The subsequent slower decrease reflects the refinement of the parameter combination near the optimum.

6.5 Comparison with Oedometer Data

The identified parameter set was subsequently evaluated using the original numerical model. The resulting oedometer response was compared with the reference response. The normalized error of the oedometer response was approximately 3.2%. In comparison, the initial parameter set resulted in an error of approximately 12.5%. This represents a substantial improvement in the reproduction of the compression behaviour. The result indicates that the contact stiffness and friction-related parameters identified by the optimization procedure provide a more suitable representation of the reference response.

6.6 Comparison with Triaxial Data

The calibrated parameters were subsequently evaluated under triaxial loading conditions. Both the deviatoric stress-strain response and the volumetric strain response were considered. The calibrated model reproduced the general nonlinear stress-strain behaviour of the reference response. The initial stiffness and subsequent increase in deviatoric stress were captured satisfactorily. The normalized error of the triaxial stress response was approximately 4.1%, compared with 15.1% for the initial parameter set. The volumetric response also showed a substantial improvement after calibration. The normalized error was reduced from approximately 17.8% to 5.3%. These results demonstrate the importance of simultaneously considering stress and volumetric responses during parameter identification. A parameter set optimized only against the deviatoric stress response could potentially reproduce the strength behaviour while producing an unrealistic volumetric response.

6.7 Combined Evaluation of the Calibrated Model

The final calibrated parameter set was evaluated using both oedometer and triaxial loading conditions. The combined objective function was defined according to

$$J = w_o J_o + w_t J_t + w_v J_v, \tag{11}$$

with equal weighting assigned to the three response components in the present demonstration. The combined normalized error of the calibrated model was approximately 4.2%, compared with approximately 15.1% for the initial parameter set. The simultaneous improvement of the oedometer and triaxial responses indicates that the calibrated parameter set is not merely optimized for one individual loading condition. Instead, the parameter-identification procedure finds a compromise that provides a consistent representation of the different macroscopic responses. This is particularly important because the different numerical parameters have different effects on the simulated behaviour. For example, the friction and rolling-resistance parameters strongly influence shear behaviour, while the contact stiffness parameters have a more direct influence on the stiffness and deformation response.

6.8 Computational Efficiency

One of the principal motivations for using the surrogate model is to reduce the computational cost associated with inverse parameter identification. In a conventional calibration procedure, each new parameter combination requires a complete numerical simulation. In the present demonstration, the trained neural network was able to evaluate a large number of parameter combinations much more rapidly than the original numerical simulations. The optimization procedure required approximately 80 surrogate-model evaluations before reaching the final parameter combination. Only a small number of additional full numerical simulations were required for verification. This indicates that the surrogate model can substantially reduce the number of expensive numerical simulations required during the parameter-search process. The computational advantage becomes increasingly relevant as the number of material parameters and loading conditions increases.

6.9 Summary of Results

The synthetic demonstration provides several observations regarding the proposed AI-assisted calibration methodology. First, the neural-network surrogate was able to reproduce the numerical responses with relatively small prediction errors. The testing results yielded coefficients of determination between 0.95 and 0.98 for the considered response quantities. Second, the inverse optimization procedure was able to recover a parameter combination close to the predefined reference parameters. The relative differences between the reference and identified parameters were generally below 10%, although the rolling-resistance parameter showed a somewhat larger deviation. Third, the calibrated parameter set significantly reduced the difference between the numerical and reference responses. The error in the oedometer response was reduced from approximately 12.5% to 3.2%, while the triaxial stress-response error decreased from 15.1% to 4.1%. Finally, the simultaneous consideration of oedometer and triaxial responses resulted in a more robust parameter identification

than calibration against a single loading condition. Overall, the synthetic results indicate that the proposed workflow is capable of connecting numerical material parameters with macroscopic soil behaviour through an AI-based surrogate model. The results therefore provide a proof-of-concept for the proposed methodology and motivate its application to actual numerical simulations and experimental laboratory data.

Author

Mohamed Abdennadher
Technische Universität Hamburg
Institut für Geotechnik und Baubetrieb
Schellerdamm 22-24
21079 Hamburg
Tel.: +49(0)40/42878-4669
e-mail: mohamed.abdennadher@tuhh.de
Web: www.tu-harburg.de/gbt

Smarter Geotechnical Coding with AI

Patrick Staubach

Abstract: Large language models can support programming-related tasks in geotechnical simulations, but their usefulness depends strongly on the type of task. This contribution discusses four case studies from the development, inspection and optimisation of geotechnical simulation software. The first case concerns the AI-assisted development of a graphical preprocessor for cone penetration simulations in numgeo. The second case uses AI to inspect an Abaqus implementation of hypoplasticity with intergranular strain and to identify local inconsistencies. The third case deals with local optimisation of performance-critical routines in numgeo, including lookup structures, caching and contact-related search operations. The final case reports a partially unsuccessful attempt to implement the Hardening Soil model with AI assistance. The comparison shows that AI was most useful for bounded tasks with clear interfaces, existing reference files and modular code structure. The study is based on practical tests with ChatGPT 5.5 and GitHub Copilot.

1 Introduction

Artificial intelligence is increasingly used in geotechnics, but its role is not uniform. In some cases, AI becomes part of the simulation software itself, for example through language-based interfaces, documentation assistants, surrogate models, digital twins or automated design exploration. In other cases, AI is used outside the simulation software as a programming assistant, for example to generate scripts, inspect existing code, refactor routines, propose tests or support debugging. These two uses are related, but they have different implications for verification. AI embedded in engineering software mainly changes the interaction between the engineer and the simulation environment, whereas AI-assisted programming can directly affect the implementation of numerical models, input generators or solver routines.

The increasing role of large language model (LLM)-based interfaces in geotechnical engineering is reflected in both recent review literature and emerging commercial tools. Sheil et al. identify human-machine interaction via large language models as one of the relevant technological enablers for future geotechnical AI, while also emphasising the need for prompt engineering, validation, benchmarking and human oversight [3]. In commercial software, Rocscience's RSInsight represents a geotechnics-specific development, where an established provider of geotechnical analysis tools introduces an AI chatbot for documentation access and, prospectively, *Natural*

Engineering Languages for model generation, analysis execution and reporting [4]. StackAI provides a different example. It is a general AI automation platform, but offers dedicated tools for geotechnical engineering, including tools for site investigation data, borehole logs, report generation and compliance review [5]. This distinction is important. RSInsight illustrates how AI may become integrated into specialised geotechnical simulation software, whereas StackAI illustrates how general-purpose AI platforms are beginning to provide domain-specific tools for geotechnical practice. The current AI landscape in geotechnical engineering is heterogeneous. Table 1 summarises representative examples and distinguishes between language-guided simulation interfaces, document-centred workflow automation, analytical design assistants and general coding tools. The broader engineering software market shows a similar development. Large simulation and design vendors increasingly integrate AI into existing CAD, CAE and digital-twin environments. These tools are generally not geotechnics-specific.

Table 1: Examples of AI-supported software relevant to geotechnical engineering.

Example	Geotechnics-specific	Main AI mode	Main purpose
Rocscience RSInsight / RSInsight+ [4]	Yes, embedded in a geotechnical software ecosystem	Chatbot / Natural Engineering Language	Documentation access, technical dialogue, future model generation, analysis execution and reporting
Ansys AI portfolio [6]	No, general engineering simulation platform	AI-augmented simulation, engineering copilot and surrogate modelling	Simulation acceleration, design-space exploration, generative geometry, digital twins and product or physics support through Ansys-GPT
StackAI geotechnical engineering solution [5]	Partly. StackAI is a general enterprise AI platform with special geotechnical toolsets	Agentic document and workflow automation	Borehole-log extraction, site investigation, report generation, compliance review and project-specific document search

A second line of work concerns the use of LLMs for the implementation of numerical methods. Kim et al. investigated whether ChatGPT can implement finite-element models for geotechnical boundary-value problems and showed that high-level finite-element libraries as a basis can make AI-generated code more reliable, whereas lower-level implementations require substantially more prompt refinement and human intervention [2]. The use of AI for programming is also supported by a broader software-engineering literature. Studies on GitHub Copilot show that AI coding assistants are often used either in an acceleration mode, where the programmer already knows the intended solution, or in an exploration mode, where the assistant is used to search for possible implementations [1].

The focus of this paper is restricted to the use of AI for programming-related tasks in geotechnical simulations. It does not aim to review all possible applications of AI in geotechnics. Instead, the paper examines how large language models can support the development, modification, inspection and optimisation of code used for numerical geotechnical simulations. This includes the generation of pre- and postprocessing tools, the inspection of existing material routines, local refactoring and performance optimisation, and the attempted implementation of a complete constitutive model.

The paper is written as a case-study contribution. Four examples are considered. They represent different levels of AI involvement in simulation-related programming tasks. Table 2 summarises the four case studies used in this paper to assess AI-supported programming in geotechnical simulations. In the first case, AI was used to develop a Python-based graphical preprocessor for axisymmetric CPT simulations in numgeo (`www.numgeo.de`, [7, 8, 9, 10]). In the second case, AI was used to inspect an existing Fortran routine for hypoplasticity and to identify local inconsistencies in the treatment of void ratio limits and small stress states. In the third case, AI was used as a refactoring and optimisation assistant for local numgeo routines, with a focus on search operations, lookup structures and frequently executed contact-related code. In the fourth case, AI was used to implement the Hardening Soil model. This last example did not lead to a fully satisfactory reproduction of reference results. Together, the examples cover successful tool generation, bounded consistency checking, local code optimisation and a case where AI assistance became difficult to verify.

Table 2: Overview of the four case studies.

Case	AI task	Verification focus	Outcome
CPT preprocessor	Generate and refine Python GUI, mesh generator and numgeo input writer	Mesh conformity, numgeo syntax, output definitions and geometric edge cases	Successful prototype after several prompt refinements and follow-up testing
Hypoplastic routine	Inspect Fortran implementation and identify local inconsistencies	Void ratio limits, lower stress bounds and unchanged material response after small code changes	Useful consistency check and marginal performance improvement
numgeo optimisation	Refactor local routines and suggest lookup and caching improvements	Identical results, runtime, memory overhead and OpenMP safety	Suitable for bounded local improvements
Hardening Soil model	Implement complex constitutive model with AI support	Comparison with reference solution and component-level tests	Partly functional, but not fully validated

2 Case study 1: CPT preprocessor for numgeo

The first case study concerns a Python tool for generating axisymmetric CPT models in numgeo. The objective was to create a controlled preprocessor for a specific class of models. The tool generates a soil mesh, a CPT mesh and a master input file. It also provides a mesh preview and prepares print output for the evaluation of total resistance, shaft resistance and tip resistance. Figure 1 shows a part of the graphical user interface.

The geotechnical problem is challenging enough to be useful as a test case. A CPT simulation in a Lagrange setting requires a body-fitted representation of the cone shaft and tip, a contact surface between cone and soil, a penetration step and, in saturated analyses, hydro-mechanical

coupling. The implemented geometry contains a CPT shaft, a conical or rounded tip, an initial embedment and a narrow zipper extension below the tip.

In the first phase of Case study 1, the development was carried out with ChatGPT 5.5 (the modes ”pro” and ”thinking” were used). In a later follow-up phase, GitHub Copilot was additionally used to inspect the repository, reproduce a specific mesh artefact and support targeted testing and correction of the CPT geometry routines.

The development did not start from an empty code base. An existing script for a rigid-inclusions modelling task was used as a template and already provided a GUI structure, mesh-generation logic and numgeo input-writing routines. This was important for the success of the AI-assisted workflow, because it constrained the task and allowed the assistant to adapt an established software structure rather than inventing a complete preprocessor from scratch.

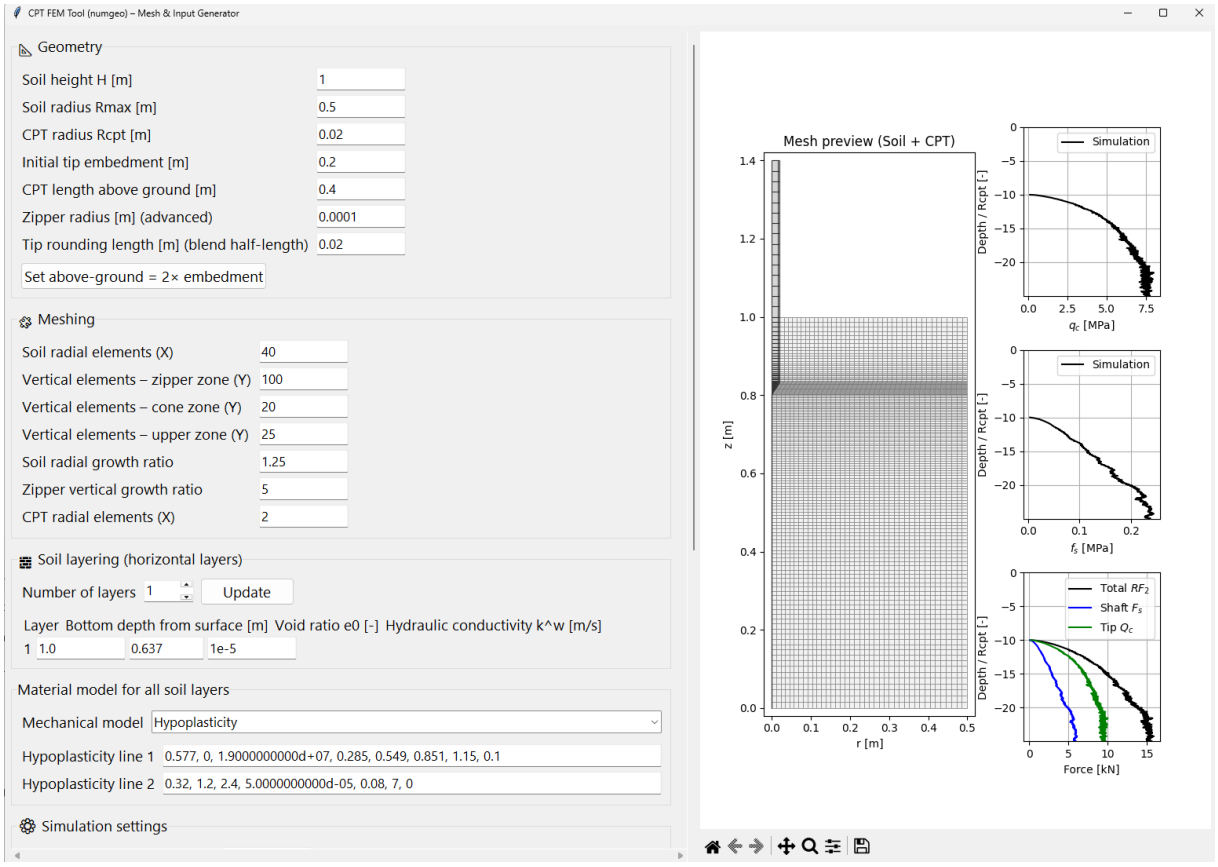


Figure 1: Graphical user interface for generating axisymmetric CPT simulations in numgeo

The final implementation was organised into three Python modules: `cpt_mesh_utils.py`, `cpt_input_builder.py` and `CPTFEMTool.py`. The first module contains the geometric description of the CPT, including `CPTGeometry`, `MeshControls`, `SoilLayer` and the mesh generation routines. A central point was that the radial CPT profile `r_profile(z)` was identical for both the CPT surface and the inner soil boundary. The second module, `cpt_input_builder.py`, writes the numgeo input files. It exports separate soil and CPT

part files, includes them as instances, creates node-based contact surfaces, assigns soil layers by bounding boxes and defines the geostatic and penetration steps. The generation of the numgeo input file was also guided by an uploaded template of an existing CPT simulation, which was used to show the AI assistant the general structure and syntax expected by numgeo. The online documentation was provided as an additional reference. The third module, `CPTFEMTool.py`, provides the graphical user interface and exposes only the relevant engineering parameters, including CPT radius, embedment, penetration depth, penetration velocity, friction coefficient, void ratio, hydraulic conductivity and drainage option.

This modularisation was important for the AI-assisted workflow. It allowed individual parts of the implementation to be modified, checked and regenerated without repeatedly changing the complete program. As a result, prompt iterations became faster and more controlled, because geometry and mesh generation, input-file writing and user-interface changes could be treated as separate subproblems.

The development did not work in one prompt. The first output was useful as a draft, but several CPT-specific details had to be refined and verified in later iterations. After the initial transfer from the existing template, the main refinement concerned the generalisation of the mesh generation for different CPT geometries, including varying rounding radii of the tip and the zipper extension. In parallel, the generated numgeo input structure had to be adapted to the requirements of coupled simulations, including separate soil and CPT parts, appropriate element types, initial stresses, density definitions, contact definitions and drainage-dependent pore-pressure boundary conditions. The implementation was then checked by generating and running CPT simulations for different geometrical settings and by inspecting whether all input-file segments were written consistently. Further prompt iterations focused on the graphical user interface, output definitions and the selection of different soil material models. Figure 2 summarises this iterative development process.

An additional follow-up step was carried out using GitHub Copilot. The objective was to check a specific geometric artefact observed in the generated CPT mesh: for small CPT or zipper radii, a visible edge could appear at the transition between the rounded CPT tip and the thin zipper extension, although the intended geometry was smooth. In contrast to the initial GPT-based workflow, Copilot was not only used to suggest code changes, but also to inspect the repository, identify the responsible part of the mesh-generation routine and reproduce the issue with a small test script. The relevant inconsistency was traced to a heuristic classification of mid-side nodes during the projection of the soil mesh onto the CPT boundary. The correction replaced this heuristic by a deterministic projection based on the analytical profile function ($r_{\text{profile}}(z)$), such that mid-side nodes are classified using the local profile radius rather than a global radius split. Copilot also generated a visual before-and-after comparison and supported the implementation of additional safeguards, including a smaller default value for the tip rounding length and validation rules to ensure that the rounding length does not exceed the CPT radius. Therefore, GitHub Copilot can support not only implementation, but also repository inspection, generation of tests and preparation of a pull request.

The main lesson from this case is simple. AI was useful because the task was bounded. Reference

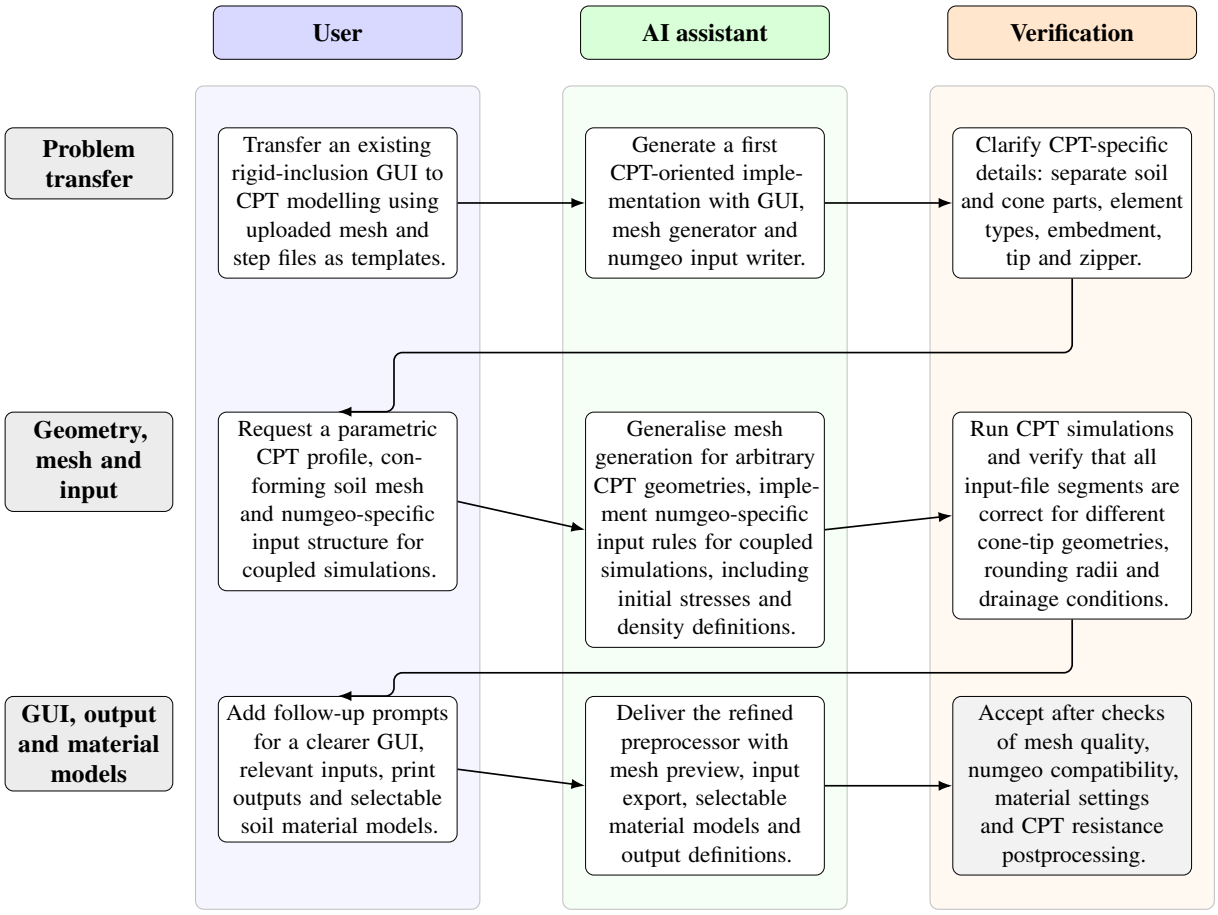


Figure 2: AI-assisted development workflow for the CPT preprocessor. The figure condenses the prompt refinement into three main iterations.

input files were available and the output could be checked by inspecting the mesh and the generated numgeo input.

3 Case study 2: hypoplastic Abaqus implementation

The second case study concerns an existing Fortran implementation of hypoplasticity with intergranular strain extension for Abaqus. The implementation is part of an open repository for hydro-mechanically coupled explicit simulations [11]. The observed problem was that simulations sometimes produced small time increments when the void ratio became very large and the stress state approached a lower limit.

In contrast to the CPT preprocessor, the objective was not to generate a new tool, but to inspect an already available constitutive implementation and to identify parts of the code where the numerical treatment was not fully consistent.

The prompt given to the AI assistant was deliberately diagnostic. The routine was provided together with the question whether obvious errors, contradictions or questionable implementation choices could be found. The most relevant findings concerned the treatment of the void ratio and the stress state at very small pressures. In different parts of the routine, the void ratio was limited or corrected in slightly different ways. Similarly, lower bounds on the mean stress were introduced in more than one location and not always with the same numerical value or interpretation. These local inconsistencies did not necessarily imply that the routine was wrong in all applications, but they made it difficult to understand which part of the implementation controlled the response in critical stress states. The AI assistant helped to identify implementation details that should be made explicit and consistent before further calibration or comparison with benchmark simulations. This case is an example of a useful but limited AI contribution. The AI did not validate the material model and did not replace the numerical algorithms. Its value was in locating a small set of plausible implementation mechanisms in a long Fortran routine. The final assessment still requires controlled runs and comparison of outputs before and after each change.

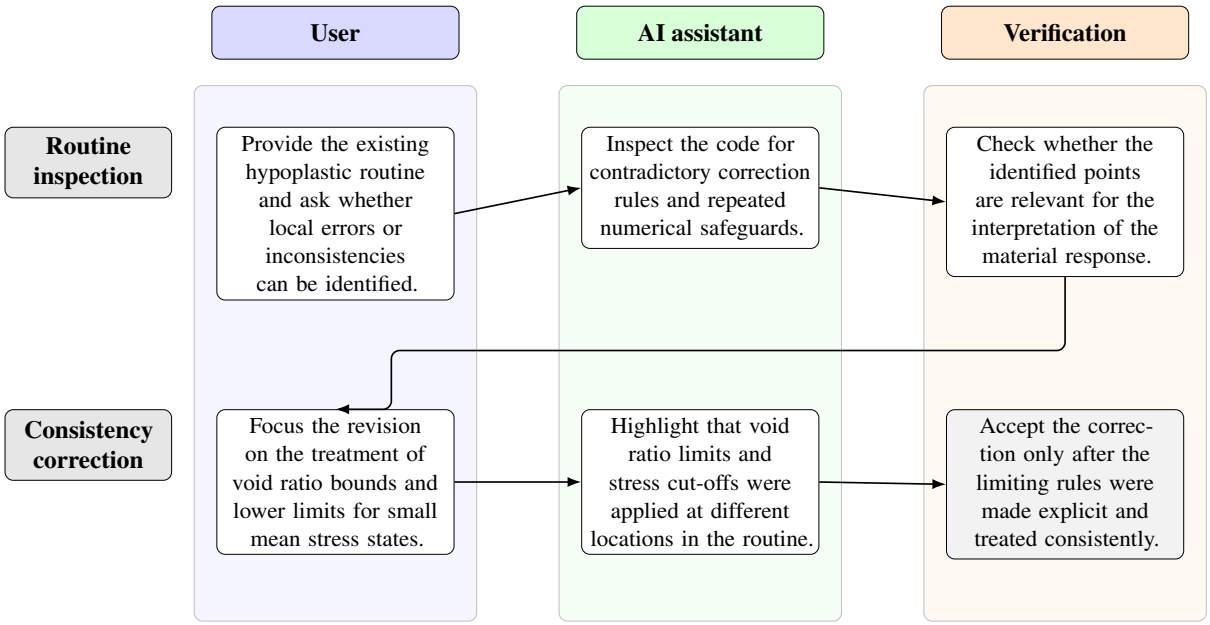


Figure 3: AI-assisted consistency check of the hypoplastic material routine

In a later follow-up, GitHub Copilot was also used to inspect the most recent commits of the hypoplastic Abaqus routine. This interaction differed from the initial consistency check because the focus was not on identifying conceptual inconsistencies, but on understanding and evaluating small local performance optimisations. Copilot summarised the recent changes and were assessed as mostly reasonable. A further small optimisation was then proposed and implemented with Copilot support. The relevant modification replaced selected `matmul` operations for small fixed-size vectors by explicitly unrolled operations and reduced repeated evaluations of expensive expressions where possible. This is only a marginal change in the overall routine and does not alter the constitutive model itself. Nevertheless, it can be useful because such operations are executed many times during a finite-element analysis. Beyond detecting inconsistencies in

the implementation, the assistant can also help to identify small performance improvements in frequently called material routines.

4 Case study 3: optimisation and parallelisation of local numgeo routines

The third case study concerns AI support for smaller optimisation tasks in numgeo. This is a different use case from model generation or constitutive debugging. The code already worked, but certain local routines involved repeated searches, temporary allocations and data movement. Such patterns are common in research codes: they are not necessarily wrong, but they can become expensive when called many times in contact search, assembly or postprocessing loops.

The AI was used as a refactoring assistant. The prompts were constrained: do not change the numerical formulation, preserve public interfaces where possible, keep a legacy path for comparison and make proposed OpenMP changes explicit. In this setting, AI can be useful because the optimisation target is local. The developer can compare the old and new routines with identical input data and can measure whether the runtime improves.

The most important suggested changes were to reduce repeated calls to search functions, to introduce precomputed lookup structures, to avoid repeated allocation and deallocation in inner loops and to check whether loops were safe for OpenMP parallelisation. Figure 4 summarises the process.

This case shows one of the stronger use cases for AI-assisted coding. The task is narrow, the code changes can be reviewed, and success can be measured. AI was useful for finding avoidable overhead and for proposing straightforward but programming time-consuming implementations that can be easily benchmarked.

5 Case study 4: inconclusive Hardening Soil implementation

The final case study is included because it did not fully succeed. The goal was to implement the Hardening Soil model with AI assistance and to reproduce reference calculations. Only publications on the equations of the Hardening Soil model were provided to the AI. The AI assistant therefore had to extract the relevant equations and algorithmic details from the documents and transfer them into the structure of an existing constitutive-model routine. As an additional template, an implementation of another constitutive model with an explicit forward Euler stress-integration scheme was provided.

The AI was useful for drafting code, translating equations, adding local checks and the stress integration scheme. After many iterations the implementation produced stress-strain curves that

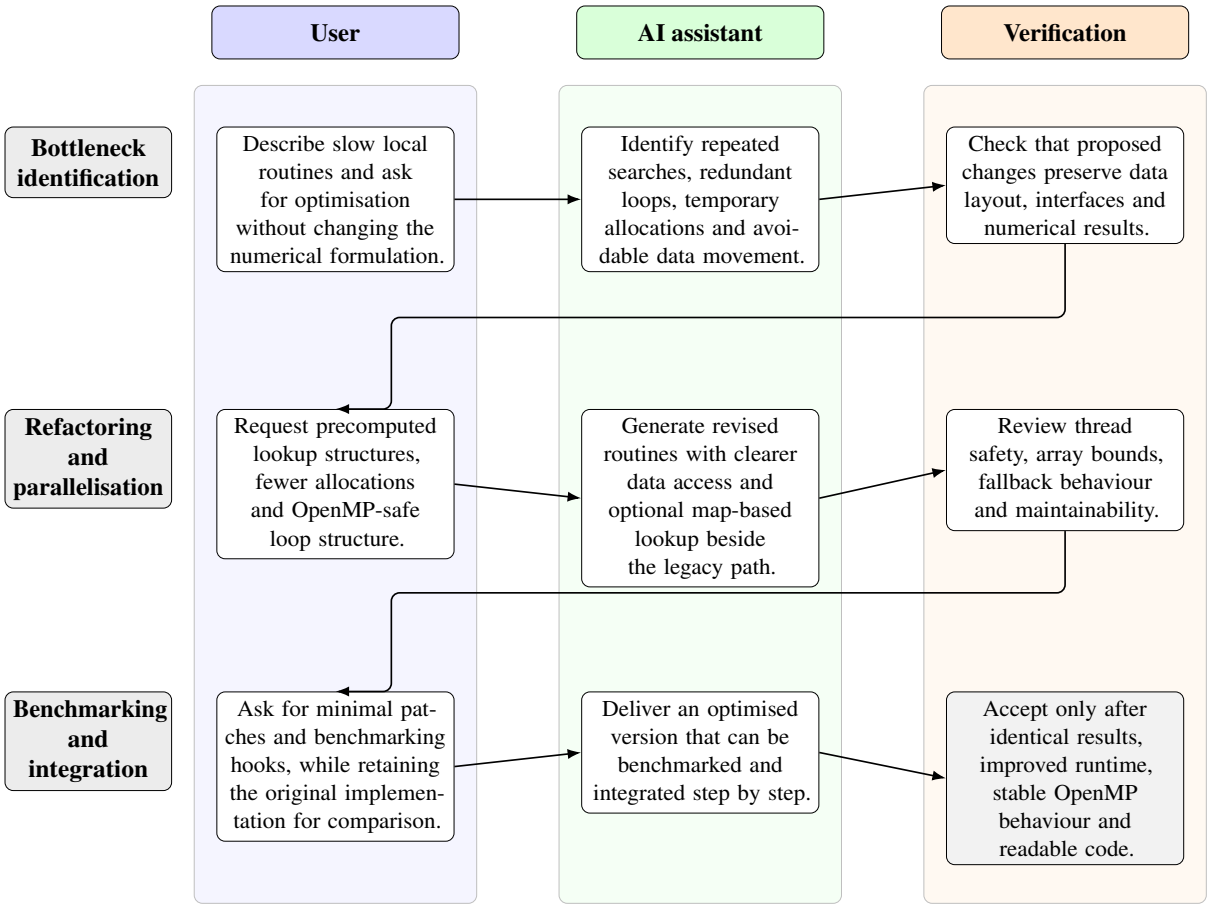


Figure 4: AI-assisted optimisation for local routines in numgeo

were often close to the reference. However, the remaining differences could not be explained reliably. A change in one part of the code improved one response but changed another. It became unclear whether the deviations were caused by an implementation error, a different integration algorithm, an undocumented convention in the reference software or an incomplete translation of the model equations.

The main problem was therefore not only accuracy. It was loss of interpretability during development. The implementation had grown through many AI-assisted modifications, and it was no longer clear which part of the code controlled which part of the response. The problems encountered in this case should not be interpreted as being specific to AI-generated code only. Similar difficulties also occur in conventional constitutive-model programming. The case therefore illustrates a general challenge of constitutive modelling that becomes more critical in AI-assisted programming that when the relation between individual code changes and the resulting material response cannot be isolated, verification becomes the limiting step.

6 Conclusions

This paper examined AI-assisted geotechnical coding through four case studies. The CPT preprocessor showed that AI can accelerate the development of a focused tool when reference files, modelling constraints and verification checks are provided. The hypoplastic material routine case showed that AI can help identify plausible code-level mechanisms in complex Fortran routines. The follow-up interactions with GitHub Copilot also indicate that coding assistants can support a higher degree of process automation than a purely chat-based workflow. In particular, Copilot was useful not only for suggesting code changes, but also for inspecting the repository, reproducing errors, generating small test scripts, checking the modified routines and preparing pull requests. This makes it a valuable complement to interactive LLM-based code generation, especially when debugging and testing can be carried out directly within the development environment. In addition, the automatic generation of documentation, commit messages and concise summaries of code changes can reduce the administrative burden of scientific software development and make the development history easier to trace. The numgeo optimisation case showed that local refactoring and parallelisation are suitable tasks if the original implementation is retained for benchmarking. The Hardening Soil case showed that AI assistance can also fail to deliver a fully validated result when the constitutive model is complex and the effect of individual implementation choices cannot be isolated.

The presented examples also indicate a broader shift in the role of the engineer. AI does not remove the need for engineering judgement. Rather, it can increase its relative importance. If routine tasks such as input-file generation, scripting, postprocessing and first drafts of documentation can be partly delegated to AI-assisted tools, the engineer can spend more time on the actual engineering decisions. This is a positive development, especially in geotechnics, where users of advanced numerical codes are usually trained as engineers rather than professional programmers. The responsibility for the numerical model remains with the engineer, but AI can reduce the amount of auxiliary programming work required to use such models effectively.

Literatur

[1] Barke S., James M. B. und Polikarpova N. (2023): Grounded Copilot: How programmers interact with code-generating models. *Proceedings of the ACM on Programming Languages*, 7(OOPSLA1):78–105. DOI: `https://doi.org/10.1145/3586030`.

[2] Kim T., Yun T. S. und Suh H. S. (2025): Can ChatGPT Implement Finite Element Models for Geotechnical Engineering Applications?. *International Journal for Numerical and Analytical Methods in Geomechanics*, 49(6):1747-1766. DOI: `https://doi.org/10.1002/nag.3956`.

[3] Sheil B., Anagnostopoulos C., Buckley R., Ciantia M. O., Febrianto E., Fu J., Gao Z., Geng X., Gong B., Hanley K., He P., Kolomvatsos K., Lopes B. C. F. L., Ninić J., Previtali M., Rezania M., Ruiz-Lopez A., Sun J., Suryasentana S., Taborda D., Utili S., Whyte

S. und Zhang P. (2026): Artificial intelligence transformations in geotechnics: Progress, challenges and future enablers. *Computers and Geotechnics*, 189:107604. DOI: https://doi.org/10.1016/j.compgeo.2025.107604.

[4] Hammah R. (2025): AI Digs Deep, Engineers Go Deeper: How AI is Radically Transforming Geotechnical Engineering. Rocscience article. Published 21 November 2025, updated 25 November 2025.

[5] StackAI (2026): The Agentic AI Platform for Geotechnical Engineering. Online resource. Accessed 7 July 2026.

[6] Ansys (2026): Ansys AI: Transforming Simulation at the Speed of AI. Online resource. Accessed 7 July 2026.

[7] Machaček J. (2020): Contributions to the Numerical Modelling of Saturated and Unsaturated Soils. Dissertation, Publications of the Institute of Soil Mechanics and Rock Mechanics, Karlsruhe Institute of Technology, Issue No. 187.

[8] Machaček J., Staubach P., Tafili M., Zachert H. und Wichtmann T. (2021): Investigation of three sophisticated constitutive soil models: From numerical formulations to element tests and the analysis of vibratory pile driving tests. *Computers and Geotechnics*, 138:104276. DOI: https://doi.org/10.1016/j.compgeo.2021.104276.

[9] Staubach P. (2022): Contributions to the numerical modelling of pile installation processes and high-cyclic loading of soils. PhD Thesis, Publications of the Chair of Soil Mechanics, Foundation Engineering and Environmental Geotechnics, Ruhr-University Bochum, Issue No. 73.

[10] Staubach P., Machaček J. und Wichtmann T. (2022): Mortar contact discretisation methods incorporating interface models based on Hypoplasticity and Sanisand: Application to vibratory pile driving. *Computers and Geotechnics*, 146:104677. DOI: https://doi.org/10.1016/j.compgeo.2022.104677.

[11] Staubach P. (2026): abaqus-explicit: Hydro-mechanically coupled explicit VUMAT for Abaqus using the hypoplastic model with intergranular strain extension. GitHub repository. https://github.com/patrickstaubach/abaqus-explicit. Accessed 2 July 2026.

Author

Univ.-Prof. Dr.-Ing. Patrick Staubach
Bauhaus-Universität Weimar
Professur Geotechnik
Coudraystrasse 11c
99423 Weimar, Germany
e-mail: patrick.staubach@uni-weimar.de
Web: https://www.uni-weimar.de/de/bau-und-umwelt/professuren/geotechnik/

An approach of BIM-to-FEM-coupling to support automation in geotechnical engineering

Sascha Henke, Johannes Beck

Abstract: In the course of ongoing digitalisation in the construction industry, the development of highly automated digital processes is becoming increasingly important to fully exploit the benefits of digitalisation. This paper presents an approach to BIM-to-FEM coupling that aims to contribute to advances in automation in geotechnical engineering.
Based on existing BIM information models, the challenges of transferring information models into numerical analysis models in geotechnics are first outlined. On the basis of these challenges, a process is developed that addresses them and thus enables a largely automated generation of two-dimensional finite element models from the data contained in the information models. In addition to the subsurface, the integration of structures (e.g. excavations) is also taken into account. The developed process is evaluated based on the resulting mesh quality.

1 Introduction

Construction projects are increasingly being delivered using the collaborative working methodology of Building Information Modelling (BIM). With the "Digital Planning and Construction implementation roadmap" and subsequent master plans for federal highways, waterways, and federal buildings, public clients are promoting and requiring the use of the BIM method. A key aspect of this is the cross-disciplinary use of information-rich, spatial digital models. Typically, a discipline-based modelling approach is followed, in which each specialist planner contributes one or more subject-specific or spatially separated models, which are then merged into coordination models or an overall model.

While BIM methodology is increasingly being adopted in geotechnics, making spatial information models of the subsurface and geotechnical structures available, these models are still rarely used as a basis for design and verification. As a result, numerical simulations are set up in isolation from the BIM-centric planning process. This implies (i) additional effort due to repeated modelling, (ii) no guaranteed consistency between simulation and design, since BIM does not function as a single source of truth (SSOT), (iii) limited traceability of model generation, and (iv) restrictions in process automation and in the reuse of simulation results.

The reasons for the lack of integration between the two technologies include missing or insufficient processes, a lack of methods for describing model generation, inadequate interfaces, and poor quality of information models.

In geotechnical models, particularly in the context of highly automated BIM-to-FEM integration, there are numerous discipline-specific challenges that make a direct transfer from the often significantly more fine-grained BIM model into a numerical (finite element) analysis model difficult or, in some cases, impossible. Especially in the case of complex subsurface and structural conditions, this regularly leads to a very high effort in numerical model creation, despite the availability of detailed BIM models, with extensive manual post-processing required in order to establish an adequate computational model. This paper addresses this issue. It presents the challenges involved in creating numerical analysis models based on BIM discipline models, and on this basis develops an automated workflow that enables automated BIM-to-FEM coupling for two-dimensional numerical models. An extension to three-dimensional models is possible but is not part of this contribution.

2 Integration of BIM in numerical simulations

The challenges arising from the automated generation of numerical finite element (FE) models from highly detailed ground models have already been addressed in several publications. For example, Brinkgreve and Brasile (2022), as well as Brinkgreve et al. (2024), describe a workflow in which a ground model extracted from Leapfrog is imported into the FE software PLAXIS. Their work primarily focuses on the poor mesh quality that results in areas containing thin soil layers.

The meshing challenges associated with deriving FE models from BIM models are also discussed by Aukenthaler et al. (2018) from the perspective of an FE software developer. The authors conclude that deficiencies in the quality of the FE mesh are primarily caused by inaccuracies in the underlying CAD model.

A study by Qiu et al. (2024) investigates a deep excavation including an existing underground metro station with tunnel. In this approach, the structural geometry is transferred from a Revit model to Abaqus. However, manual post-processing is still required, as certain structural components, such as the excavation support system, must be remodelled within the numerical software. The need for post-processing of geometries transferred from BIM software is also discussed by Sauerbrey et al. (2022) using the example of a deep excavation.

Further studies mainly address individual aspects of data exchange between BIM and numerical models, often in the context of tunnel engineering (e.g., Ninic et al., 2017; Ninic et al., 2020). In addition to the workflow from the BIM model to the computational model, these studies also consider the reverse process, namely how analysis results can subsequently be transferred back to and visualized within the BIM model.

It should be noted that the above review represents only a selection of existing studies in literature. Nevertheless, it illustrates that previous studies have generally focused on individual aspects of BIM-to-FEM integration. To date, however, no largely automated workflow has been developed that is capable of generating a broad range of geotechnical boundary value problems directly from existing discipline-specific BIM models. At the same time, the published work

clearly demonstrates that several key challenges must be addressed in order to enable such a process in a robust and reliable manner. These challenges include, but are not limited to, the following:

- Meshing problems resulting from the different levels of geometric detail required in BIM models compared with FE models.
- Challenges associated with complex geological conditions, such as thin soil layers, pinching-out strata, and other irregular stratigraphic features.
- Difficulties in integrating geotechnical structures and structural components into an automated workflow, with manual remodelling of these components still generally being required.

3 BIM-to-FEM coupling

3.1 General idea

Since BIM-based ground models are now being developed in many projects and typically contain a high level of data and information, while generally providing an accurate representation of subsurface geometry, it is a natural step to make use of these data by transferring them directly into downstream analysis tools, such as finite element (FE) software. The objective is to avoid redundant modelling of both the ground and structural components, thereby improving workflow efficiency and minimizing data transfer errors. However, as briefly discussed in Section 2, the differing levels of geometric granularity between BIM and FE models prevent such a direct transfer without additional processing. Consequently, extensive manual remodelling is currently required in most cases.

To optimize this process and to establish a BIM-to-FEM coupling workflow that requires as little manual intervention as possible, the challenges that arise, primarily related to geometry, must be addressed through geometry optimization techniques. The associated challenges and exemplary solution approaches are discussed in the following.

3.2 Challenges

The challenges associated with BIM-to-FEM coupling are predominantly geometry-related. An overview of various challenges arising from potential subsurface stratifications is presented in Figure 1**.**

In Figure 1, as well as in subsequent figures in which finite element meshes are evaluated with respect to mesh quality, the quality of the elements is assessed using the normalized shape ratio ρ_{NSR}, defined as the ratio of the inradius r to the circumradius R of a triangular element:

$$\rho_{NSR} = \alpha \cdot \frac{r}{R} \tag{1}$$

For triangular elements, the normalization factor is taken as $\alpha = 2$, whereas $\alpha = 3$ would be used for tetrahedral elements. Since the present study is restricted to two-dimensional models discretized with triangular elements, $\alpha = 2$ is adopted throughout. Accordingly, a value of $\rho_{NSR} = 1$ represents an ideal element, whereas $\rho_{NSR} = 0$ indicates maximum element distortion.

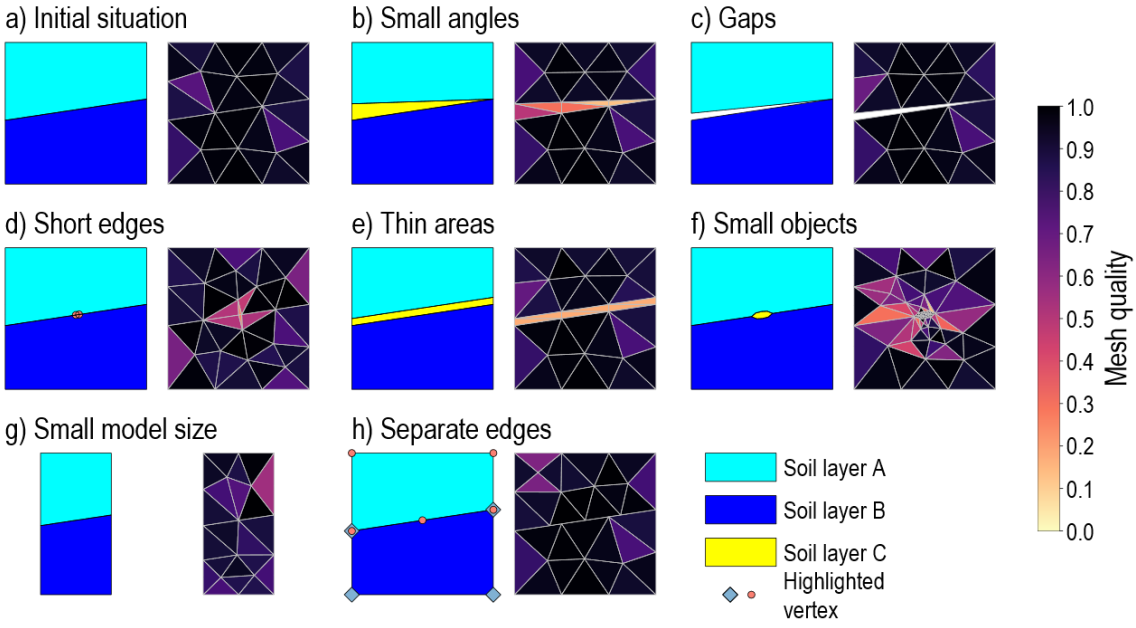


Figure 1: Different geometric challenges related to subsurface stratification in the context of BIM-to-FEM coupling

Figure 1 illustrates the principal geometric challenges that may arise when transferring CAD program to a finite element mesh generator. Although the underlying causes differ, they generally result either in poor mesh quality or, in some cases, prevent successful mesh generation altogether. Starting from a simple reference case consisting of two horizontally superimposed soil layers (Figure 1a), the following stratification-related geometric issues can be identified: (I) acute angles, for example at pinching-out layers (Figure 1b), (II) thin layers (Figure 1e), and (III) small inclusions such as lenses (Figure 1f). The remaining issues shown in Figure 1 are primarily caused by modelling or data transfer errors originating from the BIM model. These include, for example, gaps between adjacent layers (Figure 1c), short edges (Figure 1d), and disconnected edges (Figure 1h). Finally, the case illustrated in Figure 1g represents another common practical challenge. To minimize boundary effects, numerical models generally require a much larger spatial extent than is supported by the available site investigation data. Consequently, the required model extension is often not substantiated by actual ground information and therefore cannot be directly represented in the BIM ground model.

The mesh quality shown in Figure 1 clearly demonstrates that, with the exception of the insufficient model extent shown in Figure 1g, the aforementioned geometric deficiencies generally lead to a significant deterioration in mesh quality. Without appropriate geometry optimization, extensive manual pre-processing of the finite element model is therefore typically required to obtain a suitable numerical model based on the BIM data.

The challenges discussed so far relate exclusively to the subsurface geometry. Figure 2 additionally illustrates the influence of geotechnical structures that must be intersected with the ground model, using a tunnel excavation as an example. It can be seen that, depending on the position of the tunnel relative to soil layer boundaries, geometric issues similar to those shown in Figure 1 arise, including acute angles and thin layers. Whenever such geometric deficiencies are introduced by the intersection of the structure with the subsurface stratification, the number of finite elements increases considerably while the overall mesh quality is significantly reduced.

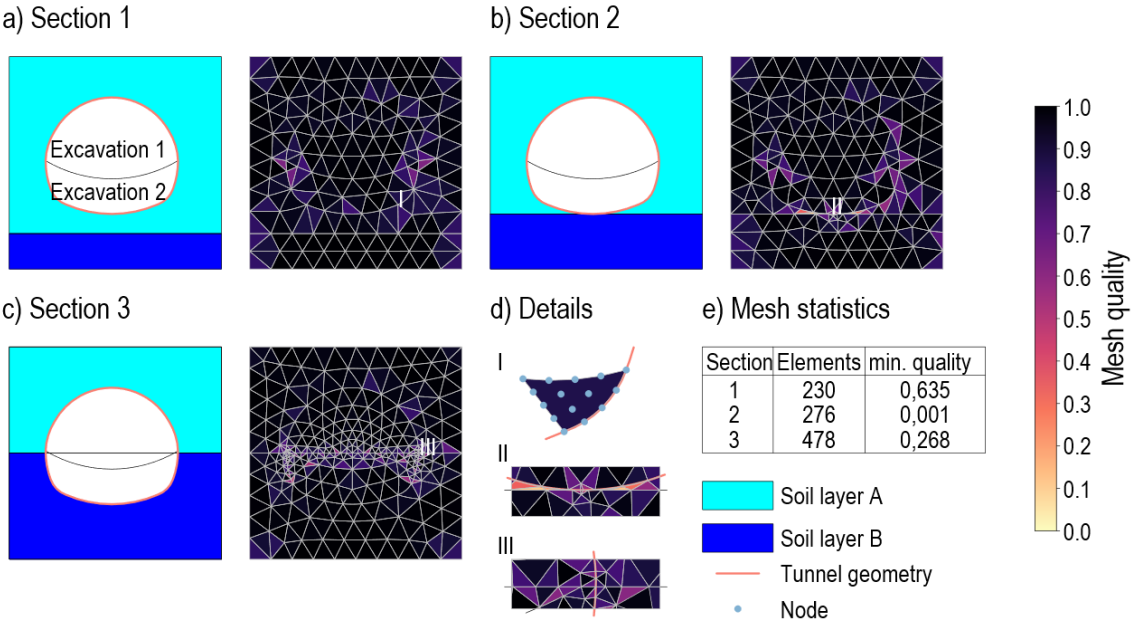


Figure 2: Influence of a geotechnical structure on the meshing of the subsurface in the numerical model

3.3 Geometry idealization supporting BIM-to-FEM coupling

In this section, selected algorithms are briefly introduced that address the aforementioned challenges in the context of geometry optimization and simplification. Due to space limitations, only a limited number of representative examples are presented.

3.3.1 Disconnected Edges, Small Gaps, and Overlaps

In Figure 3, the initial problem and reference geometry are illustrated by two curves comprising a total of five vertices, which may be interpreted as segments of polylines representing a subsurface layer. The endpoints of the two curves are not coincident, which would cause meshing problems due to the presence of non-coincident nodes in the resulting mesh.

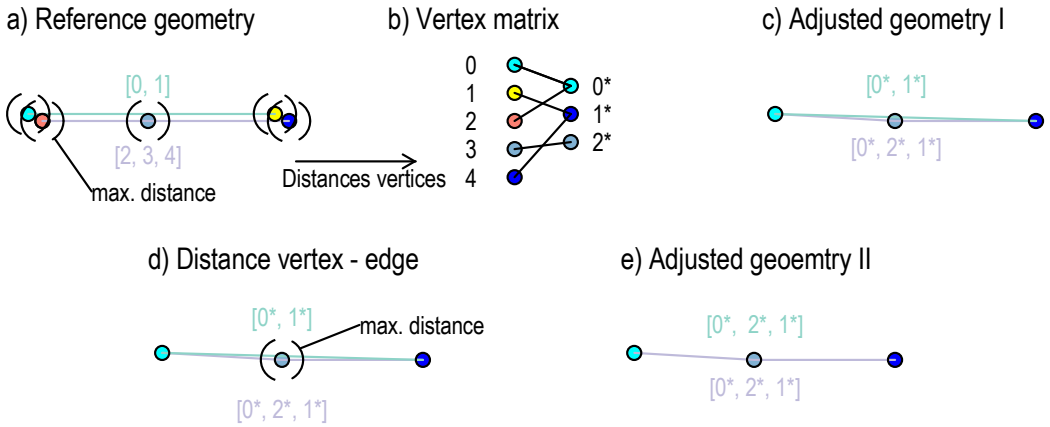


Figure 3: Adjustment of Similar Polyline Segments

First, a list of all vertices and a distance matrix containing the distances from each vertex to the vertices of other curves are constructed. In the subsequent step, vertices whose mutual distance falls below a predefined threshold are merged, thereby resolving the non-coincident endpoints in the example. Various strategies can be employed to determine which vertex is retained and which is moved or adjusted. This results in the configuration shown in Figure 3c. In subsequent steps, the intermediate vertex can be processed in an analogous manner (Figure 3d), ultimately yielding the cleaned geometry shown in Figure 3e.

3.3.2 Sharp angles

The problem of sharp corners is commonly encountered in subsurface layers. In these areas in particular, poorly conditioned finite element meshes are typically generated if no further geometric adjustment is performed. Figure 4 presents an algorithm illustrating how an appropriate simplification can be carried out for sharp angles.

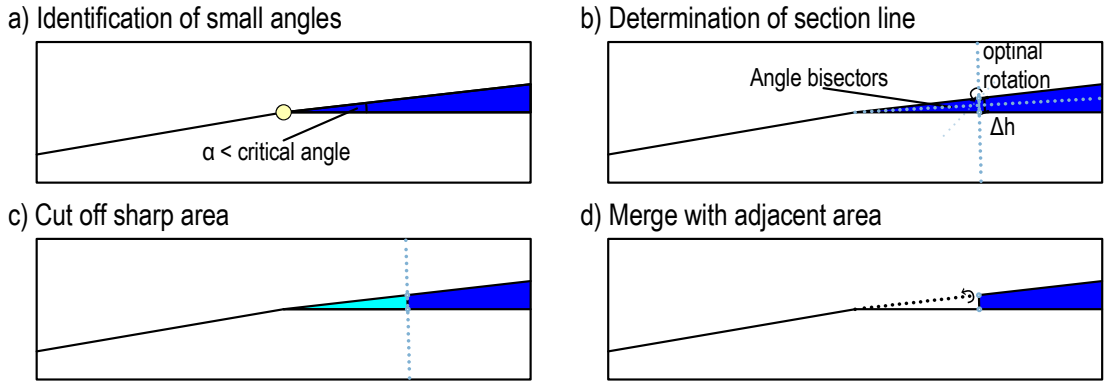


Figure 4: Algorithm for simplification of sharp edges

The procedure is as follows: (i) During an iteration over the polylines, the interior angles are identified and evaluated. As soon as an interior angle is detected that falls below a predefined threshold angle, the geometry simplification process is initiated at this location. (ii) At the identified interior angle, an angle bisector is first constructed between the adjacent segments. A line arranged orthogonally to the angle bisector is then shifted until the intersection points with the polyline are separated by a predefined distance Δh (Figure 4b). Optionally, a subsequent rotation can be applied, by which the shifted line (here representing the cutting edge) is inclined such that its orientation resembles the original erosion geometry. (iii) The acute-angled section adjacent to the cutting edge is then removed and connected to an adjacent subsurface layer (Figure 4c and d). In the present case, the connection could be established with either of the two polylines. Here, an approach was chosen in which the longer common edge is selected for merging the subsurface layers.

The result of this geometry simplification can be seen in Figure 5. It is evident that a significant improvement in mesh quality was achieved (cf. Figure 5b and 5d). Furthermore, this approach also allows for a significant reduction in the number of elements in the FE model, especially when applying a local mesh refinement considering skewness.

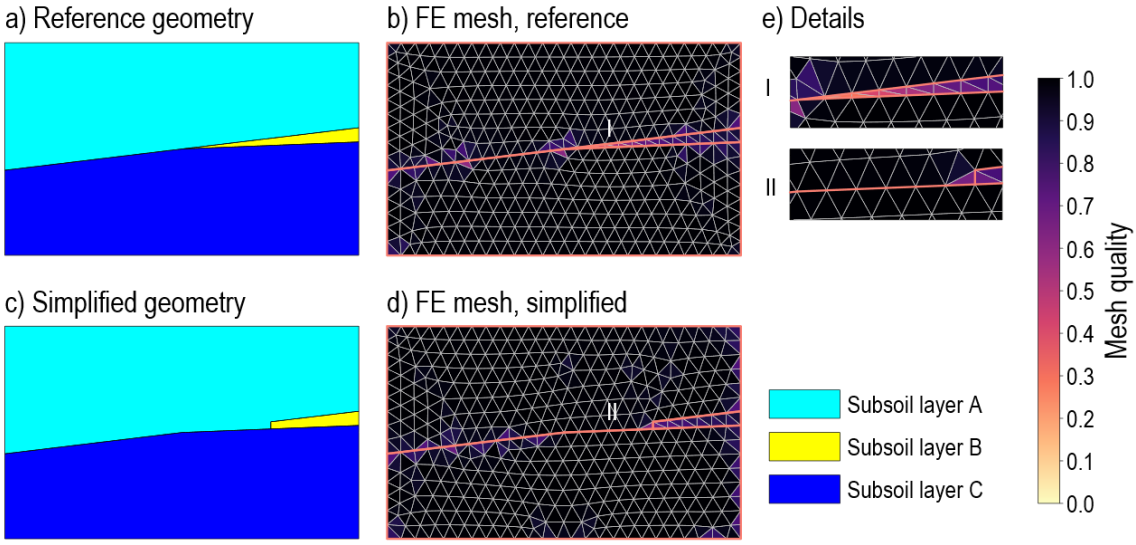


Figure 5: Development of mesh quality with and without geometric simplification of sharp edges

3.3.3 Small Angles at Intersections with Structural Elements

Furthermore, Figure 6 presents an algorithm for simplifying the geometry at intersections between subsurface layers and structural elements, such as anchors.

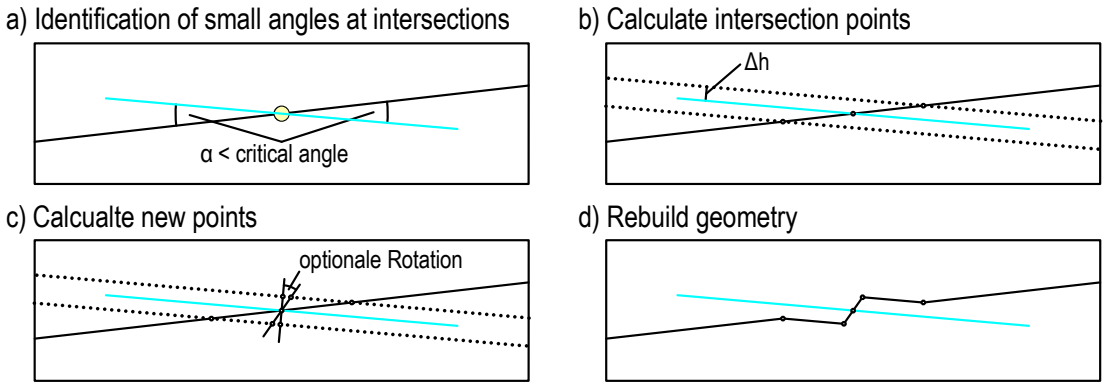


Figure 6: Algorithm for simplification of small angles due to intersection of structural elements and soil layers

The intersection of anchors and subsurface layers may result in small angles between the structural element and the subsurface layer, as shown in Figure 7. Particularly in the intersection region, highly distorted elements are generated due to the very acute angles occurring in this area.

Within the framework of the geometry optimization, the following procedure is applied: (i) In the first step, embedded 1D elements are identified that form angles with the boundary curves of the subsurface layers smaller than a predefined threshold angle (Figure 6a). At these locations, corresponding geometry simplifications are performed. (ii) In the second step (Figure 6b), two parallel lines are constructed on both sides of the structural element at a predefined minimum distance. (iii) These parallel lines are then intersected with the curves of the

subsurface layers (Figure 6b and c). A line passing through the central intersection point is subsequently added, which can optionally be rotated to account for the geological formation process of the subsurface layer (Figure 6c). (iv) Based on these intersection points, a new cleaned polyline is generated in which no acute angles below the predefined threshold angle remain (Figure 6d).

Figure 7 again illustrates the influence of this minor geometric adjustment on the mesh quality of the FE mesh in the corresponding area. Once again, it can be observed that no highly distorted elements remain after geometry optimization. In addition, the number of elements is also reduced by a value of around 50%.

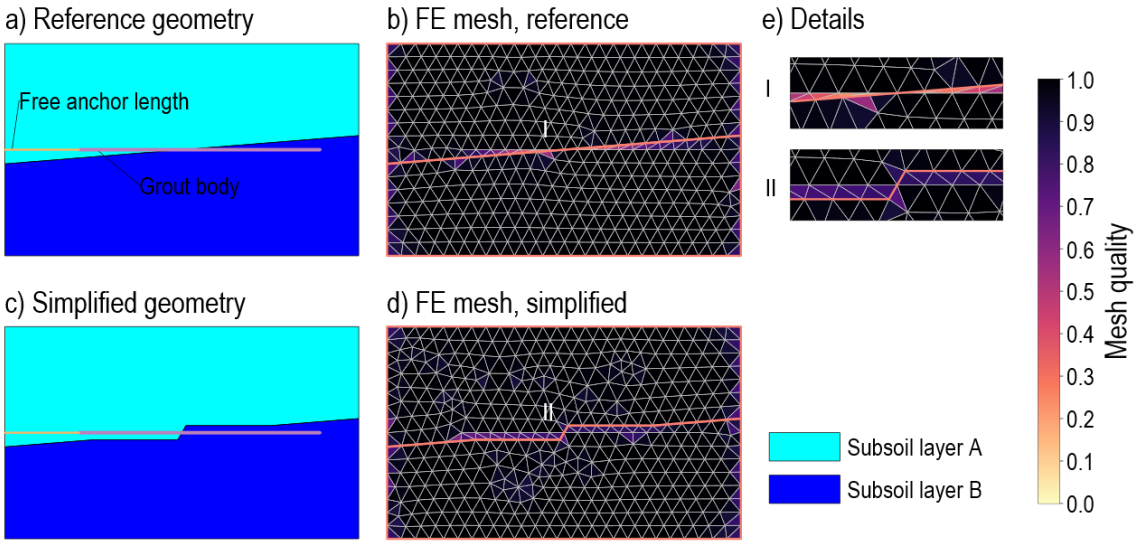


Figure 7: Example of an anchor-soil intersection without and with geometry optimization

Further examples and application cases are presented, for example, in Beck (2026).

4 Outlook

The present paper has addressed a workflow that enables the largely automated generation of numerical FE models based on data from existing BIM-based geotechnical models. A key challenge identified in this context is the differing levels of granularity between the various models, which regularly leads to problems during mesh generation in the creation of finite element models. As a consequence, a time-consuming manual remodelling process is currently often required, which is both inefficient and highly susceptible to errors.

Using various problem scenarios, the methods of geometry simplification and their influence on the quality of the resulting FE meshes are demonstrated. The results are promising and indicate that, through the application of such approaches, a reliable BIM-to-FEM coupling workflow may become feasible for practical applications in the future. Compared to other research activities, it should be emphasized that, in the context of a consistent use of BIM models, not only the subsurface conditions themselves but also structural elements embedded within the ground are considered.

At this stage, it should be noted that the development of the presented approach has so far been limited to two-dimensional finite element models. Further research will therefore focus on extending the methodology to three-dimensional models. In addition, the applicability and robustness of the developed methodology must be further validated through its consistent application to practice-oriented case studies. Initial examples of such applications are provided, for instance, by Beck (2026).

References

[1] Aukenthaler, M., Keizer, V. und Brinkgreve, R. B. (2018). PLAXIS Interoperability: From Customization to BIM compliance. In: *Workshop Digitale Infrastruktur & Geotechnik*. Hrsg. von J. Grabe. Veröffentlichungen des Institutes Geotechnik und Baubetrieb. Hamburg: Technische Universität Hamburg, Institut für Geotechnik und Baubetrieb.

[2] Beck J. (2026): Zur Berechnung geotechnischer Systeme mithilfe von multidimensionalen Informationsmodellen. Eingereichte Dissertation an der Helmut-Schmidt-Universität, Hamburg.

[3] Brinkgreve, R. B. und Brasile, S. (2022). Automatic Finite Element Modelling and Parameter Determination for Geotechnical Design. In: *Conference Proceedings Workshop Numerische Methoden in der Geotechnik*. Hrsg. von J. Grabe. Hamburg, S. 83–98

[4] Ninic, J., Koch, C. und Stascheit, J. (2017). An integrated platform for design and numerical analysis of shield tunnelling processes on different levels of detail. In: *Advances in Engineering Software* 112, S. 165–179. DOI: 10.1016/j.advengsoft.2017.05.012

[5] Ninic, J., Koch, C., Vonthron, A., Tizani, W. und Konig, M. (2020). Integrated parametric multi-level information and numerical modelling of mechanised tunnelling projects. In: *Advanced Engineering Informatics* 43, S. 101011. DOI: 10.1016/j.aei.2019.101011.

[6] Qiu, Y., Wang, J., Zhang, C., Hua, L. und Zhou, Z. (2024). Numerical Simulation Study on the Impact of Excavation on Existing Subway Stations Based on BIM-FEM Framework. In: *Buildings* 14.5, S. 1444. DOI: 10.3390/buildings14051444

[7] Sauerbrey, J., Schmitt, J. und Ramm, H. (2022). Erfahrungen zur 3D-, 4D- und 5D-Ebene im Rahmen der praktischen Anwendung von BIM bei der Planung einer tiefen Baugrube. In: *13. Kolloquium Bauen in Boden und Fels*. Hrsg. von C. Moormann und C. Vogt-Breyer. Tübingen: expert, S. 361–366.

Authors

Univ.-Prof. Dr.-Ing. habil. Sascha Henke
Helmut-Schmidt-Universität, Hamburg
Professur für Geotechnik
Holstenhofweg 85
22043 Hamburg
Tel.: +49 (0) 40 – 6541 / 3351
e-mail: sascha.henke@hsu-hh.de
Web: www.hsu-hh.de/geot

M.Eng. Johannes Beck
LEONHARD WEISS GmbH & Co. KG
Leonhard-Weiss-Straße 2-3
74589 Satteldorf
Tel.: +49 (0) 7951 33 / 3352
e-mail: joh.beck@leonhard-weiss.com

Parametric finite element analyses incorporating stochastic ground models for uncertainty propagation

Tobias Peterstorfer, Johannes Leo, Franz Tschuchnigg

Abstract: Stratigraphic uncertainty can substantially influence geotechnical finite element analyses but is commonly represented by only a single deterministic ground model. This paper presents a non-intrusive workflow for propagating geological uncertainty using multiple point statistics (MPS) and parametric finite element analyses. MPS is used to generate alternative stratigraphic realizations conditioned on a geological training image and available investigation data. A geometric post-processing procedure converts the categorical raster fields into consistent polygonal geometries and simplifies the resulting interfaces for efficient finite element mesh generation. The workflow is demonstrated for a complex slope-retaining system within a stratigraphic model comprising four geotechnical units. The ensemble results illustrate how stratigraphic uncertainty propagates to the geotechnical system response and provide information beyond a single deterministic prediction.

1 Introduction

Geotechnical finite element models include assumptions and simplifications regarding geometry, the representation of soil behaviour using constitutive models, and the applied boundary conditions. Ground-related uncertainties affect the FE-model in multiple ways, e.g., through unknown stratigraphic configuration and material properties assigned to the individual geotechnical units. The former is often termed *geological uncertainty* and concerns the geometries of the ground model, whereas *geotechnical uncertainty* describes the spatial variability of material properties within pseudo-homogeneous units. Both sources of uncertainty are pronounced in geotechnical engineering because the subsurface cannot be observed directly, geomaterials are naturally heterogeneous, and site investigations generally provide only sparse information.

Starting with the early work of Vanmarcke [15], numerous studies have investigated the influence of soil variability on the reliability of geotechnical structures. Fenton and Griffiths [4] pioneered the random finite element method by coupling random fields of geotechnical parameters with finite element analysis in a Monte Carlo framework. In comparison, the propagation of uncertainty in stratigraphic geometry has received less attention, although, as reported by Clayton [2], uncertainty in geological unit boundaries constitutes a comparable source of construction

risk. Therefore, a need for practical methods that integrate probabilistic ground models into geotechnical finite element analyses exists.

Current methods for probabilistic geological modelling may broadly be divided into interface-based and category-based approaches. Interface-based methods describe stratigraphic boundaries as random functions [8], whereas category-based methods simulate soil types directly on a discretized grid, for example using coupled Markov-chain models [3]. Both approaches can generate multiple subsurface configurations, but their application generally requires the calibration of statistical parameters (e.g., variogram parameters of random functions or soil type transition probabilities) from sparse and 'ugly' real-world investigation data [9]. Multiple point statistics (MPS) provide a promising alternative by reproducing complex spatial patterns and geological relationships from a training image.

Shi and Wang [10, 11] demonstrated the potential of using MPS-generated stratigraphic models in geotechnical analyses. The present study investigates the applicability to a complex slope-retaining system and addresses the geometric post-processing required to convert pixel-based MPS realizations into consistent polygonal geometries suitable for parametric finite element modelling. An ensemble of alternative stratigraphic realizations is then used to investigate the resulting variability in the predicted response of the geotechnical structure.

2 Methods

2.1 Multiple Point Statistics

Multiple point statistics (MPS) comprise a class of stochastic simulation methods used to reproduce complex spatial structures from a training image. Although rooted in geostatistics, MPS may therefore also be interpreted as a form of Bayesian supervised machine learning, in which the training image provides prior information on geological structures and site observations serve as conditioning data [10]. In contrast to conventional geostatistical approaches, which commonly describe spatial dependence using two-point measures such as variograms, MPS considers spatial patterns involving multiple locations simultaneously. The method has been widely applied in reservoir modelling and hydrogeology, while more recent studies have demonstrated its potential for geotechnical site characterization [10, 11].

In sequential simulation algorithms such as GENESIM [7], the simulation domain is populated one cell at a time. At each unsimulated location, the known values within a prescribed neighbourhood form a local data event consisting of conditioning observations and previously simulated cells. Similar spatial configurations are identified in the training image, and their frequency is used to estimate the conditional probability

$$\mathbb{P}\left(Z(\boldsymbol{x}) = z_k \mid \mathcal{D}_x\right) = \frac{\mathbb{P}\left(\mathcal{D}_y \cap Z(\boldsymbol{y}) = z_k\right)}{\mathbb{P}\left(\mathcal{D}_y\right)}, \tag{1}$$

where $\mathcal{D}_x$ denotes the data event around the target location $\boldsymbol{x}$, and $\mathcal{D}_y$ denotes a matching

configuration centred at location $\boldsymbol{y}$ in the training image. The category at the target cell is sampled from the resulting conditional distributions and subsequently becomes part of the conditioning information. Repeating the simulation with different random paths and draws produces alternative stratigraphic realizations that honour the available site observations while reproducing geological patterns from the training image. For a more comprehensive description of the theoretical background of MPS, the reader is referred to [7, 10].

2.2 Geometric Post-Processing for FE-Model Integration

The MPS realizations are generated as categorical raster grids (Figure 1a), whereas polygonal representations are usually preferable for integration into the finite element model. At fine simulation resolutions, however, direct polygonization (Figure 1b) produces boundaries with a large number of vertices, which can introduce excessive geometric constraints and adversely affect finite element mesh generation. To reduce this geometric complexity, the polygon boundaries are therefore simplified using the Visvalingam-Whyatt algorithm implemented in the Python package Shapely [5]. The degree of simplification can be controlled by a prescribed tolerance. As shown in Figure 1c, this procedure substantially reduces the number of boundary vertices while preserving a space-filling partition of the model domain without gaps or overlaps between adjacent units. The resulting polygons can subsequently be transferred to a finite element software and discretized efficiently.

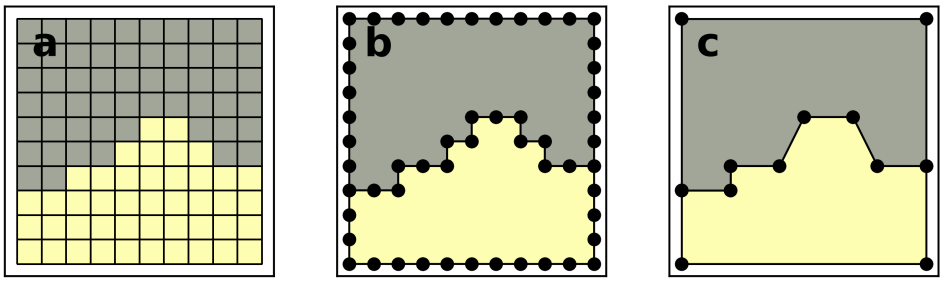


Figure 1: Post-processing workflow: (a) categorical raster field, (b) polygonized realization, and (c) simplified polygons

2.3 Parametric Finite Element Analysis

In a parametric finite element analysis, the numerical model is implemented as a callable function of a prescribed set of input parameters. Generally, these inputs may describe material properties, loads, boundary conditions, or geometric features and may be represented by scalar values, spatial fields, or more complex geometric data structures. In the present framework, the stratigraphic configuration is provided as a set of polygons delineating the individual geological units. For each input stratification, a Python-based routine automatically generates a calculation model in PLAXIS 2D [1], creates a mesh, performs the calculation, and extracts predefined output

quantities of interest. The parametric finite element analysis may therefore be expressed as a deterministic mapping, which maps a given set of input parameters and returns a unique response:

$$\mathcal{M}: \quad \mathbf{x} \mapsto \mathbf{y} = \mathcal{M}(\mathbf{x}), \tag{2}$$

where $\mathbf{x} \in \mathbb{R}^M$ denotes the model input and $\mathbf{y} \in \mathbb{R}^Q$ contains the corresponding response quantities, such as nodal displacements, structural forces, or factors of safety [12].

Such parameterized formulations enable the automated evaluation of multiple ground model configurations and are therefore suitable for sensitivity analysis, optimization, and uncertainty propagation. The approach is non-intrusive, as the finite-element solver itself does not require modification; uncertainty is propagated solely through repeated evaluations of the model with different input configurations.

3 Application

The presented workflow is applied to a slope-retaining system. To isolate the impact of stratigraphic uncertainty on the system response, all other model components, such as the design of the retaining system and material parameters, are treated deterministically. For demonstration purposes, 150 ground model realizations are generated and evaluated using a parametric FE-model.

3.1 Calculation Model

The case study considers the staged excavation of a 15 m high cut at the toe of a slope, as shown in Figure 2. The excavation is supported by a 25 cm thick shotcrete facing, eight rows of soil nails, and two levels of prestressed anchors. The soil nails (type R51-800) with lengths of 10, 12, and 15 m and an out-of-plane spacing of 1.5 m are represented by embedded beam elements [6] with an assumed grout-body diameter of 10 cm and layer-dependent skin resistance. The prestressed anchors have a free length of 34 m, a bonded length of 11 m, an out-of-plane spacing of 2 m, and a prestressing force of 800 kN. Their free sections are modelled using node-to-node anchors with $EA = 234$ MN, while the bonded sections are represented by embedded beam elements.

The excavation is simulated in six stages to depths of 3.0 m, 4.5 m, 7.5 m, 9.0 m, 12.0 m, and 15.0 m below the excavation crest. To approximate three-dimensional effects within the 2D plane-strain model, each excavation step is divided into an unsupported partial excavation stage using a stress-release approach ($M_{\mathrm{Stage}} = 0.7$), followed by a second stage in which the corresponding support elements are activated and the excavation is completed. The ground model comprises four geotechnical units, represented using the Hardening Soil model, with their constitutive parameters summarized in Table 1.

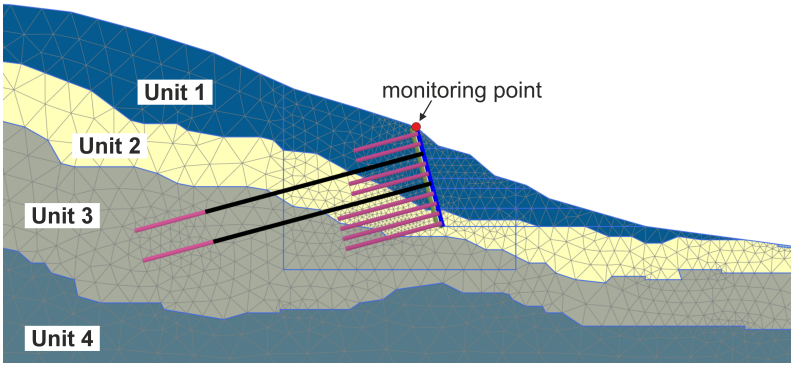


Figure 2: Calculation model

Table 1: Soil parameters ($p_{ref} = 100$ kPa)

Parameter	E_{oed}^{ref}	E_{50}^{ref}	E_{ur}^{ref}	m	ν	c'	φ'	ψ	γ
Unit	(MPa)	(MPa)	(MPa)	(-)	(-)	(kPa)	(°)	(°)	(kN/m^3)
Unit 1	30	30	90	0.5	0.2	2	36	4	20
Unit 2	18	23	85	0.7	0.2	5	24	0	20
Unit 3	40	40	120	0.5	0.2	0	35	0	21
Unit 4	60	60	180	0.5	0.2	0	38	5	21

3.2 Stochastic Stratigraphic Model

For the considered example, only two boreholes are assumed to be available from the site investigation; the stratigraphic configuration between and beyond these locations is therefore uncertain. Based on a geological interpretation, a training image is drawn using image-editing software, with each geological unit represented by a distinct coloured area. The image is subsequently rasterized onto a regular grid and converted into a categorical indicator field (Figure 3, left). MPS is then used to generate an ensemble of alternative ground models that reproduce the spatial patterns of the training image while honouring the available borehole and outcrop data (Figure 3, right).

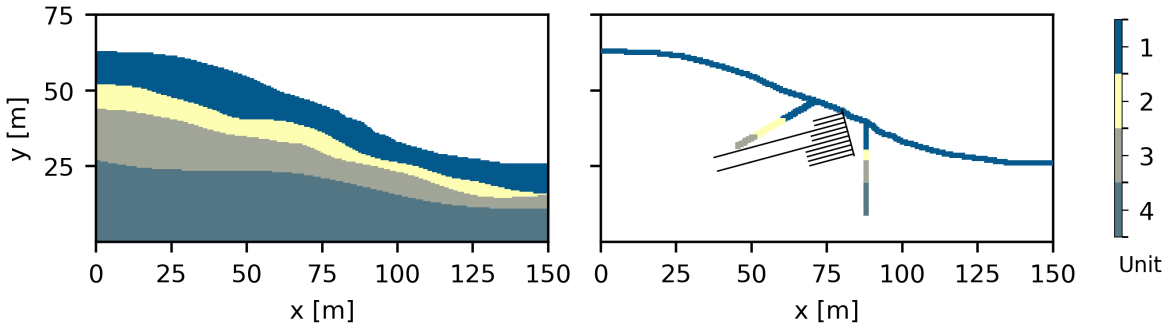


Figure 3: Training image (left) and hard constraints from boreholes and surface outcrop observations (right) used in the MPS simulations

The resulting ensemble is summarized by the modal soil type and the information entropy, shown in Figure 4. The modal soil type denotes the most frequently occurring category at each grid cell, while the information entropy $H(\boldsymbol{x}) = -\sum_k p_k(\boldsymbol{x}) \log_2 p_k(\boldsymbol{x})$ provides a measure of local categorical uncertainty. As shown in Figure 4, the modal soil type broadly reproduces the main stratigraphic features represented in the training image. The entropy distribution highlights zones of increased uncertainty along stratigraphic boundaries and shows a general increase in uncertainty with increasing distance from the conditioning data as expected.

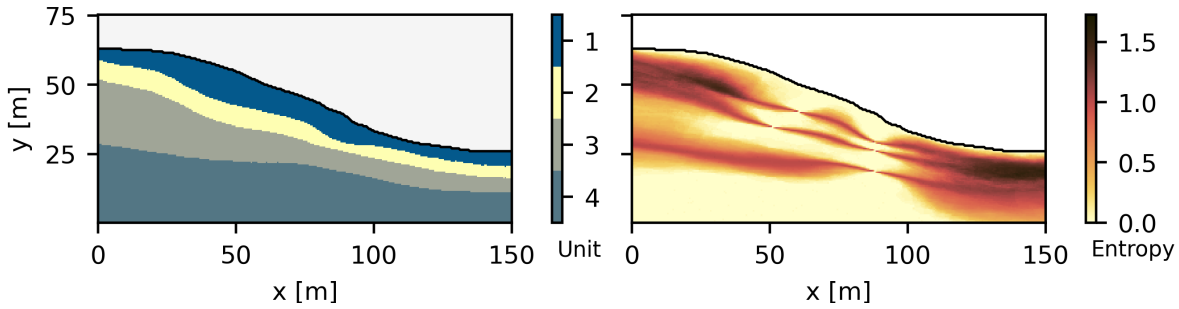


Figure 4: Modal soil type (left) and information entropy (right) obtained from the MPS ensemble

Figure 5 shows six exemplary post-processed realizations. Although most simulations produced geologically plausible configurations, approximately 10% exhibited artefacts that were inconsistent with the geological conceptual model. The realization shown in the bottom-right panel provides an example of such an unrealistic simulation outcome. These realizations were identified through visual screening and excluded from further evaluations. The issue is discussed in Section 4.

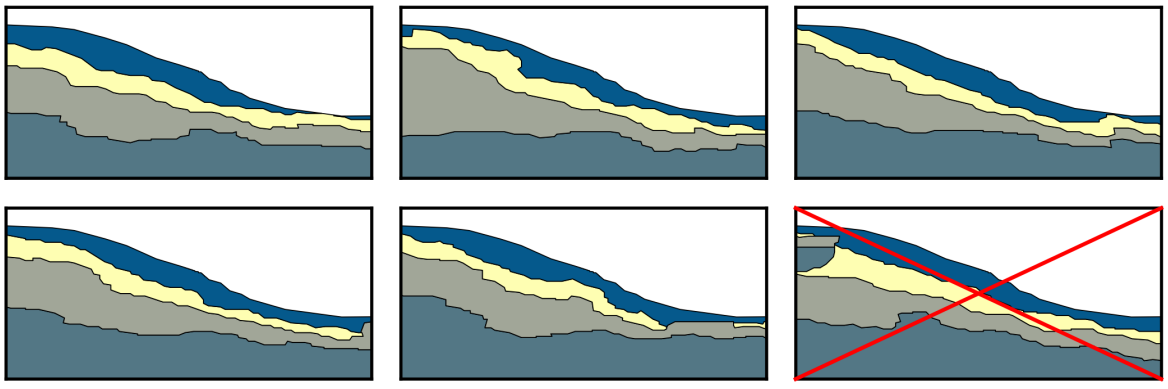


Figure 5: Six exemplary realizations of the ground model. The realization in the bottom-right panel was identified as implausible and excluded from further analysis

3.3 Response Variability

Selected response quantities obtained from the finite element analyses are presented in Figures 6 and 7. Figure 6 shows the development of the horizontal displacement at the top of the retaining structure during the staged excavation. The spread between the individual realizations increases as excavation progresses, indicating a growing influence of stratigraphic uncertainty on the predicted deformation response. The resulting response ranges may support a more informative comparison between numerical predictions and monitoring data. In contrast to a deterministic analysis, which provides only a single reference value, the ensemble indicates whether an observed deviation lies within the range associated with the adopted stratigraphic uncertainty. This information is particularly relevant to applications of the observational method and real-time back-analysis [14].

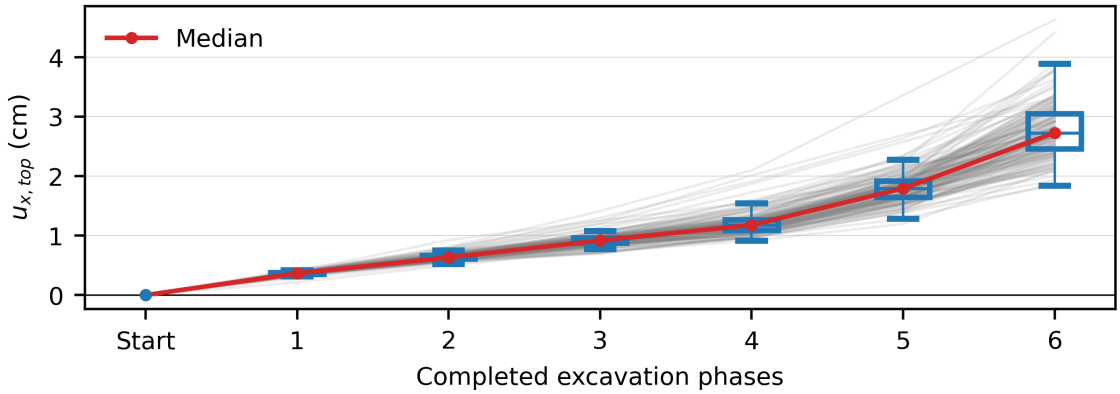


Figure 6: Evolution of the horizontal displacement at the top of the retaining wall during the excavation phases

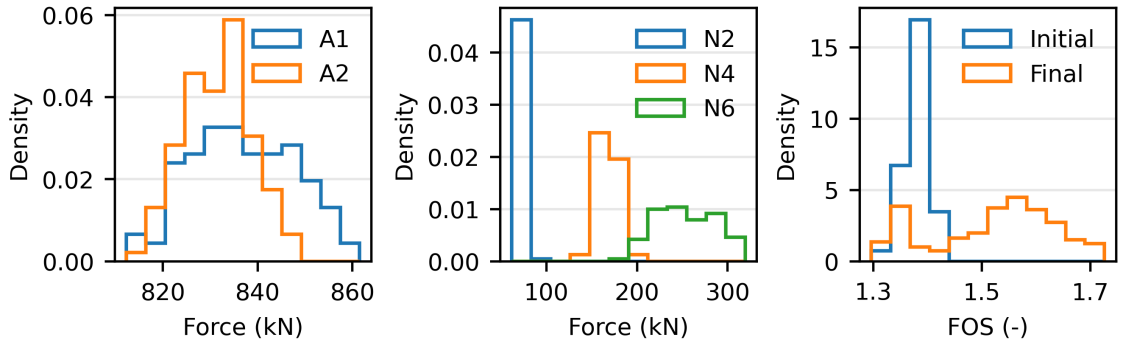


Figure 7: Exemplary response distributions. Left: anchor forces in the upper (A1) and lower (A2) anchor levels; centre: axial forces in the second (N2), fourth (N4), and sixth (N6) soil nails from the top; right: FOS before construction and after completion of the retaining system

Figure 7 presents the resulting distributions of anchor forces, soil nail forces, and factors of safety. The structural components exhibit different sensitivities to the stratigraphic realization. For example, the force in soil nail N2 varies only slightly across the ensemble, whereas soil nail N6 shows substantially greater variability. The factors of safety obtained from φ-c-reduction

analyses [13] are compared for the initial state and after the completed excavation. The distribution after excavation is wider and exhibits a pronounced bimodal shape, which may indicate the occurrence of two different failure mechanisms within the ensemble.

4 Discussion

The results demonstrate that MPS provides a feasible basis for representing stratigraphic uncertainty in geotechnical finite element analyses. Although the simulations are generated on a categorical raster grid, geometric post-processing enables their conversion into consistent polygonal models that can be transferred directly to a finite element software. A practical advantage of the approach is that prior geological knowledge is represented through a training image rather than through an explicitly parameterized stochastic model. Since conceptual geological models are commonly used in practice, this representation may be more accessible to practitioners than defining random interface functions using variograms or specifying transition probabilities.

Nevertheless, the present implementation exhibits several limitations. Some realizations contained geologically implausible features and required manual screening. While this was manageable for the relatively small ensemble considered here, it would be unsuitable for large-scale reliability analyses involving thousands of realizations. The issues occurred predominantly near the model boundaries, but their origin could not be conclusively identified. Possible causes include the limited spatial extent of the training image, incomplete conditioning neighbourhoods near the boundaries, and the specific GENESIM settings. Since Shi and Wang [11] did not report comparable artefacts, the issue may be related to the present implementation rather than to the general approach. Further work should test extended-domain simulations followed by cropping and include automated plausibility checks. In addition, the MPS grid resolution and geometric simplification settings affect stratigraphic boundary roughness and should be selected carefully. The sensitivity of the geotechnical response to interface roughness is problem-specific and should therefore be assessed for the application under consideration.

5 Conclusion

Multiple point statistics enable the generation of alternative stratigraphic models conditioned on site investigation data and a conceptual training image. Through geometric post-processing, the resulting categorical raster fields can be converted into consistent polygonal geometries and integrated into complex geotechnical finite element models without modification of the numerical solver. The presented case study demonstrates how uncertainty in the stratigraphic configuration can be quantitatively propagated to predicted deformations, structural forces, and factors of safety. The resulting ensemble provides additional information beyond a single deterministic prediction by defining a plausible range of numerical responses. The presented workflow therefore offers a practical basis for the non-intrusive consideration of stratigraphic uncertainty in geotechnical analysis. Further development is required to improve the geological realism of the simulations.

References

[1] Bentley Systems (2025): *PLAXIS 2D 2025.1 Reference Manual*.

[2] Clayton C. R. I. (2001): *Managing Geotechnical Risk: Improving Productivity in UK Building and Construction*. Thomas Telford Publishing, London, UK. ISBN: 0727729675

[3] Elfeki A. M. M. und Dekking F. M. (2005): Modelling subsurface heterogeneity by coupled Markov chains: Directional dependency, Walther's law and entropy. *Geotechnical and Geological Engineering*, 23(6):721–756. https://doi.org/10.1007/s10706-004-2899-z

[4] Fenton G. A. und Griffiths D. V. (2008): *Risk Assessment in Geotechnical Engineering*. Wiley. https://doi.org/10.1002/9780470284704

[5] Gillies S., van der Wel C., van den Bossche J., Taves M. W., Arnott J., Ward B. C. et al. (2025): Shapely, Version 2.1.2. https://doi.org/10.5281/zenodo.17193310

[6] Granitzer A.-N., Tschuchnigg F., Hosseini S. und Brasile S. (2024): Insight into numerical characteristics of embedded finite elements for pile-type structures employing an enhanced formulation. *International Journal for Numerical and Analytical Methods in Geomechanics*, 48(1):223–249. https://doi.org/10.1002/nag.3641

[7] Hansen T. M., Vu L. T. und Bach T. (2016): MPSLIB: A C++ class for sequential simulation of multiple-point statistical models. *SoftwareX*, 5:127–133. https://doi.org/10.1016/j.softx.2016.07.001

[8] Leo J., Peterstorfer T. und Tschuchnigg F. (2026): On the consideration of geological and geotechnical uncertainty in numerical slope stability analysis. *Engineering Geology*, 372(A):108952. https://doi.org/10.1016/j.enggeo.2026.108952

[9] Phoon K.-K., Ching J. und Shuku T. (2022): Challenges in data-driven site characterization. *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, 16(1):114–126. https://doi.org/10.1080/17499518.2021.1896005

[10] Shi C. und Wang Y. (2021a): Nonparametric and data-driven interpolation of subsurface soil stratigraphy from limited data using multiple point statistics. *Canadian Geotechnical Journal*, 58(2):261–280. https://doi.org/10.1139/cgj-2019-0843

[11] Shi C. und Wang Y. (2021b): Smart determination of borehole number and locations for stability analysis of multi-layered slopes using multiple point statistics and information entropy. *Canadian Geotechnical Journal*, 58(11):1669–1689. https://doi.org/10.1139/cgj-2020-0327

[12] Sudret B. (2012): Meta-models for structural reliability and uncertainty quantification. In: Phoon K., Beer M., Quek S. und Pang S. (Ed.), *Proceedings of the 5th Asian-Pacific Symposium on Structural Reliability and its Applications (APSSRA 2012)*, pp. 53–76.

[13] Tschuchnigg F. (2019): *A Contribution to Computational Geotechnics – Theory and Practice*. Graz University of Technology, Graz, Austria.

[14] Tschuchnigg F., Nicholson D., Canavate Grimal A., Chen Y., Le T., Ferrero Cafiero J. A., O'Brien T., Zania V. und Nguyen L. T. (2026): Introduction to real-time back analysis and recent developments related to the Observational Method. In: Pistrol J., Adam D. und Schweiger H. F. (Ed.), *Proceedings of the 21st International Conference on Soil Mechanics and Geotechnical Engineering (ICSMGE 2026)*. https://doi.org/10.53243/ICSMGE2026-975

[15] Vanmarcke E. H. (1977): Probabilistic modeling of soil profiles. *Journal of the Geotechnical Engineering Division*, 103(11). https://doi.org/10.1061/AJGEB6.0000517

Authors

Dipl.-Ing. Tobias Peterstorfer
Graz University of Technology
Institute of Soil Mechanics, Foundation Engineering and Computational Geotechnics
Rechbauerstraße 12
8010 Graz
Tel.: +43 316 873 - 4218
e-mail: peterstorfer@tugraz.at
Web: www.tugraz.at/en/institutes/ibg/home

Dipl.-Ing. Johannes Leo
Graz University of Technology
Institute of Soil Mechanics, Foundation Engineering and Computational Geotechnics
Rechbauerstraße 12
8010 Graz
Tel.: +43 316 873 - 6239
e-mail: leo@tugraz.at
Web: www.tugraz.at/en/institutes/ibg/home

Univ.-Prof. Dipl.-Ing. Dr.techn. Franz Tschuchnigg
Graz University of Technology
Institute of Soil Mechanics, Foundation Engineering and Computational Geotechnics
Rechbauerstraße 12
8010 Graz
Tel.: +43 316 873 - 6729
e-mail: franz.tschuchnigg@tugraz.at
Web: www.tugraz.at/en/institutes/ibg/home

Towards displacement-based design of quay walls

Luca Flessati

Abstract: Maritime quay walls are critical components of ports, enabling the safe berthing of vessels and efficient cargo handling. Their structural performance is governed by soil-structure interaction, where wall movements and soil response mutually influence each other. Conventional design methods simplify this interaction by assuming predefined active and passive earth pressures, resulting in analytically solvable but highly conservative designs. This paper demonstrates the application of displacement-based design methodologies that explicitly account for soil-structure interaction, with a particular focus on some of the approaches currently adopted in the Dutch practice and by the Port of Rotterdam Authority. By including realistic deformation behaviour into the design process, these methodologies offer the potential for more efficient, economical, and reliable quay wall designs.

1 Introduction

Maritime ports serve as essential nodes within global supply chains, facilitating the transport of nearly 90% of international trade (International Chamber of Shipping, 2024). Quay walls (QWs), the structures that allow the berthing of ships and facilitate the loading and unloading of cargo, are essential components within port infrastructures. Quay walls typically rely on slender structural elements to resist lateral soil pressure. When soil pressure acts on the wall, it induces wall movements. These movements lead to soil displacements, which in turn change the soil pressure. Therefore, the soil structure interaction (SSI) plays a fundamental role in the mechanical response of QWs.

Traditional QW design methods rely on a set of simplifying assumptions that effectively eliminate SSI from the analysis. Lateral earth pressures are prescribed a priori by assuming the soil to be at failure under either active (land side) or passive (water side) conditions. Consequently, the stress distribution acting on the wall cannot evolve in response to wall deformations, and the resulting solution is independent of soil and wall stiffness. Additional simplifications concern the anchors response (usually assumed to be rigid) and the kinematic conditions at the wall toe. Depending on the method, the wall base may be assumed to prevent horizontal displacement while allowing rotation (free base), or to restrain both displacement and rotation (fixed base). The method proposed by Blum (1931), the basis of the traditional Dutch design approach, imposes even stronger constraints by requiring displacement, rotation, and curvature to vanish at the wall tip.

These assumptions reduce the problem into a statically determinate one that can be solved analytically. Moreover, since lateral earth pressure is treated as a predefined action, rather than a result of a SSI problem, these traditional design approaches can easily be used in the design codes partial safety factor framework. Despite their simplicity, in the Dutch context these approaches are not now used for the final design of quay walls, but are sometimes used to have a first indication of the dimensions of the wall. In practice, these traditional approaches are usually very conservative, providing technical solutions characterized by very long and very stiff walls. Such solutions may substantially increase construction costs and potentially create practical difficulties during installation.

The goal of this paper is demonstrating the application of displacement-based design methodologies for quay walls, with particular emphasis on the design solutions currently adopted for the Port of Rotterdam Authority (PoR). For the sake of brevity, the focus will only be on the wall response and the response in the soil in the proximity. Other mechanisms, such as piping and global QW loss of stability are not hereafter considered.

2 Quay wall design solutions

At present, QWs in PoR are predominantly constructed using combi-walls, consisting of steel tubular piles combined with sheet piles. The tubular piles have both a retaining function and a bearing function for vertical loads. In contrast, the sheet piles serve only to transmit earth and water pressures to the primary elements. Combi-walls are used due to their large stiffness, strength, bearing capacity and for their economic convenience (de Gijt & Broeken, 2013). The most commonly adopted structural configurations are: (i) anchored walls (AW, Figure 1a) and (ii) walls with relieving platforms (RPW, Figure 1b).

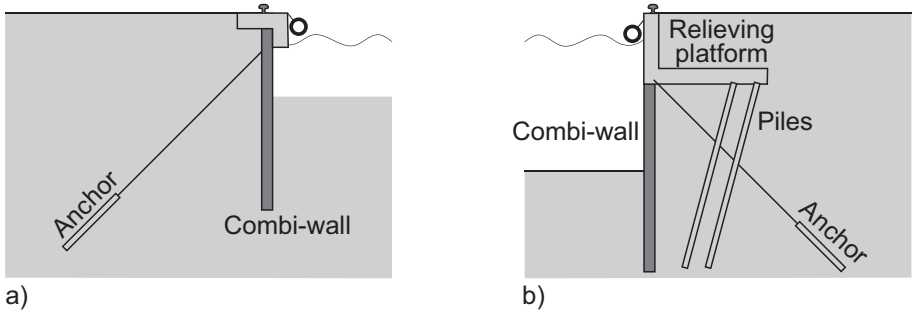


Figure 1: Type of walls currently adopted in the PoR: a) anchored walls and b) walls with relieving platform

Currently, AWs are typically designed with a single row of anchors, as the installation of underwater anchors remains technically challenging. However, new technologies aimed at the safe, reliable, and cost-effective installation of underwater anchors are currently under development (Roubos, 2026). To ensure adequate pull-out resistance, design guidelines (CUR 166) require the anchor bodies to be embedded in the deeper sand layers. Consequently, the

anchors are generally installed at relatively large inclinations. AWs are commonly adopted for quay walls with moderate retaining heights or where the operational loads from cranes, moored vessels, and stored cargo are relatively limited.

In contrast, RPWs are generally adopted for large retaining heights and the high operational loads associated with deep-sea ports. The relieving platform redistributes the vertical loads, transferring a significant portion of the backfill and crane loads directly to the wall and its supporting piles. As a result, the vertical stresses in the soil immediately behind the wall are substantially reduced, resulting to lower lateral earth pressures acting on the wall. This, in turn, limits the horizontal deflection of the wall and improves its overall structural performance. Despite these advantages, RPWs are associated with high construction costs and a substantial environmental impact due to their intensive use of concrete and steel. According to Roubos (2026), the embodied carbon footprint of an RPW is approximately 50 tonnes of CO_2 per metre of quay wall.

The optimisation of these structures therefore offers considerable potential for both economic and environmental benefits. However, achieving this remains challenging due to the complex interaction between the soil and the structure that cannot be accounted for with the traditional approaches. In the following, displacement based design approaches for quay walls with progressively increasing levels of detail and computational cost are presented.

3 Subgrade reaction based approaches

According to this approach, the retaining wall is schematized as a beam, while the soil reactions and anchor forces are concentrated at the beam nodes and represented by means of independent linear or non-linear springs (Figure 2a). The main challenge of this approach lies in defining the constitutive relationship of the non-linear springs modelling the lateral soil reaction. In the simplest applications, elastic-perfectly plastic laws are adopted, where the limiting resistances are typically derived from active and passive earth pressure theories, while the initial elastic stiffness depends on both the soil stiffness and the geometry of the wall.

The current Dutch design practice (CUR 166) adopts a more sophisticated constitutive model that aims to capture the non-linear soil behaviour more realistically. This constitutive law is piecewise linear as it is shown in Figure 2b. For nil displacements, the lateral earth pressure p_0 is calculated from the at rest lateral earth pressure coefficients. On the active side, the response is elastic perfectly plastic and the stiffness is defined k_{h1}, while the ultimate active pressure value p_a is calculated through the limit active pressure. On the passive side, the response is subdivided into 4 linear branches. The first one, from p_0 to $0.5p_p$, being p_p the ultimate passive pressure value, is characterized by stiffness k_{h1}. The second branch, from $0.5p_p$ to $0.8p_p$, is characterized by stiffness k_{h2}, while the third one, from $0.8p_p$ to p_p, is characterized by stiffness k_{h3}. Finally, the last branch for lateral pressure equal to p_p defines the failure condition (i.e. zero stiffness). The characteristic secant stiffness values k_{h1}, k_{h2}, k_{h3} (Figure 2b). These values were derived from the Dutch practice and put in relation with the type of soil and the cone tip resistance value q_c or the undrained strength S_u (Table 1). According to De Gijt & Broeken (2013), the values on

the table was only validated for retaining heights up to 8-10m. For larger structures the CUR 166 suggests to correct the spring stiffness values to account for wall dimensions.

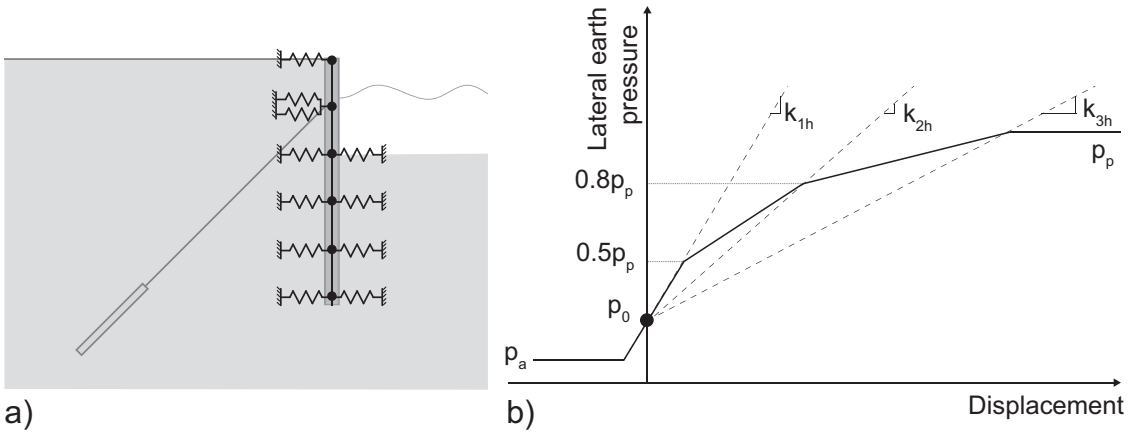


Figure 2: a) Schematization of wall as beam interacting with non-linear springs and b) Piecewise linear spring constitutive law

Table 1: Low and high secant characteristic subgrade reaction values

		k_{h1} (MN/m^3)	k_{h2} (MN/m^3)	k_{h3} (MN/m^3)
Sand	q_c (MPa)			
loose	5	12 - 27	6 - 13.5	3 - 6.75
medium	15	20 - 45	10 - 22.5	5 - 11.25
dense	25	40 - 90	20 - 45	10 - 22.5
Clay	S_u (kPa)			
soft	25	2 - 4.5	0.8 - 1.8	0.5 - 1.125
stiff	50	4 - 9	2 - 4.5	0.8 - 1.8
very stiff	200	6 - 13.5	4 - 9	2 - 4.5

From a theoretical point of view, this approach can also accommodate inclined walls, sloping ground surfaces and soil-wall interface friction, by properly calculating the active and passive earth pressure. In the practice, the influence of these factors is disregarded on the spring stiffness, and therefore the results have to be interpreted with particular attention.

This approach can account for very simple steady state seepage conditions (one-dimensional upward and downward seepage, typical when thin low permeability horizontal layers are present), but it cannot consider neither more complex seepage nor hydro-mechanical coupling. As a result, the method should not be used under partially drained conditions, where the rate of excess pore pressure dissipation is comparable to the rate of excavation.

Moreover, this approach cannot predict the vertical ground settlements associated with horizontal wall movements. Although some simplified procedures to estimate settlements from wall deflections exist, the results should be interpreted with particular caution. This limitation is

particularly relevant because, in many cases, the control of vertical ground movements is one of the governing design criterion.

Finally, the applicability of the approach is restricted to relatively simple two-dimensional geometries. It cannot adequately represent complex structural configurations, such as RPWs, irregular excavation geometries, complex loading conditions, strongly layered soil profiles, or all the situations where the stress distribution within the soil differs significantly from the assumptions underlying active and passive earth pressure theories.

Despite these limitations, the method remains extremely attractive in engineering practice. In fact, it provides results (in terms of wall deformations, bending moments, shear forces, and anchor loads) in a very short time while still accounting for SSI. For this reason, it is well suited to preliminary and conceptual design stages, where several alternative solutions must be sieved.

4 Finite element approaches

Finite element (FE) analyses are considerably more computationally demanding than the approaches discussed previously; however, they overcome the inherent limitations of the subgrade reaction method. In particular, FE models can realistically represent complex structural configurations, heterogeneous soil stratigraphies, hydraulic conditions (also hydro-mechanical coupling), and staged excavation processes. Also three-dimensional geometries and loads (e.g. moored vessels and cranes) can be reproduced, but, at present three dimensional models are not commonly adopted in the design.

In recent years, numerous studies have demonstrated the capability of FE analyses to accurately reproduce the mechanical response of QWs. Within the context of PoR, experience gained from several case studies (Lai et al., 2025; Hao et al., 2026; Roubos et al., 2026) has enabled the development of a site-specific framework for calibrating the parameters of the Hardening Soil constitutive model (Schanz et al., 1999). The calibration is based on literature q_c based correlations, providing a practical methodology for site characterization.

As an example, Figure 3 (adapted from Roubos et al., 2026) compares the measured displacements of the RPW following excavation (dotted line) with the preliminary FE predictions (solid line) developed within the CLARION project (CLARION, 2024). The dashed line represents an alternative prediction approach, which is discussed in the following section. The numerical results are in reasonably good agreement with the field measurements, particularly considering that the soil parameters were not calibrated against the in situ observations.

Despite the capability of FE analyses to reproduce the response of QWs, several aspects require careful consideration. In particular, the following topics are discussed: (i) the use of characteristic/design soil properties; (ii) the choice of an appropriate soil constitutive model and (iii) the use of FE analyses to develop and calibrate simplified wall calculation models.

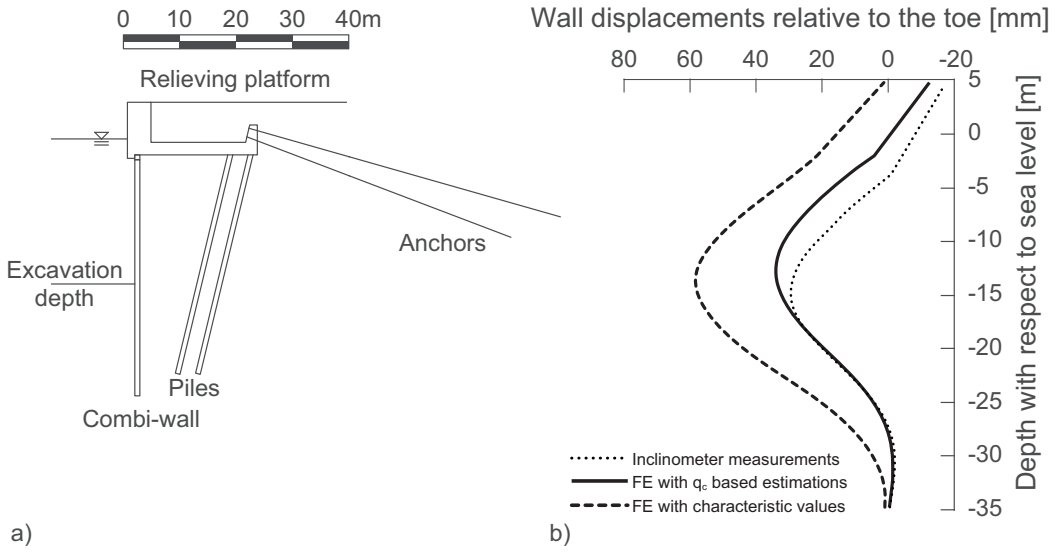


Figure 3: a) considered geometry and b) comparison of field measurements and numerical results, adapted from Roubos et al., (2026)

4.1 The use of characteristic/design values

According to the current partial factor design framework adopted in Eurocode 7 (EN 1997-1), the design should be performed by using characteristic or design values of the soil properties rather than best-estimate values. Consequently, the most representative estimates of the ground parameters are conservatively reduced to account for uncertainty and to satisfy the prescribed level of safety.

The implications of this approach are illustrated with reference to the case study of Roubos et al. (2026). Since the comparison is performed considering only permanent loads, the analysis is carried out using characteristic soil properties in accordance with the EC7 framework (serviceability limit state). The FE prediction obtained under these assumptions is shown by the dashed line in Figure 3b. Compared with both the field measurements and the prediction based on best-estimate parameters, the use of characteristic values leads to a significant underestimation of the system performance, resulting in predicted wall displacements that are, in this case, almost twice those observed in the field.

Moreover, it is worth noting that in non-linear SSI problems, the (arbitrary) reduction of soil strength parameters (it theory also stiffness should be reduced but this is not accounted for, Flessati & di Prisco, 2023) may lead to the development of non-realistic failure mechanisms. As a result, the design may be governed by mechanisms that are not representative of the actual system response. For this reason, in non-linear analyses (not only geotechnical but also structural), partial factor design approaches based on direct reduction of material properties (such as the M2 set of partial factors in Eurocode 7 terminology) should be applied with particular caution.

4.2 Soil constitutive law

The reliability of FE predictions is affected by uncertainties associated with both the constitutive model formulation and the calibration of its parameters. In the following, for the sake of brevity only the most diffused (and simple) classes of models will be presented and the limitation of these models will be discussed with the reference of a simplified case of an excavation of an AW (Figure 4). In particular it is considered a combi-wall constituted by tubular piles of diameter 1.42 m, thickness 0.02 m, length 33.5 m and spacing 3.29 m, connected to 3 x PU28 steel sheet-piles. The depth of excavation is equal to 18 m, the water table is at 4 m from the ground surface. The anchor is schematized as a rigid constraint at 2 m depth. The wall is assumed to be installed in a homogeneous medium-dense sand.

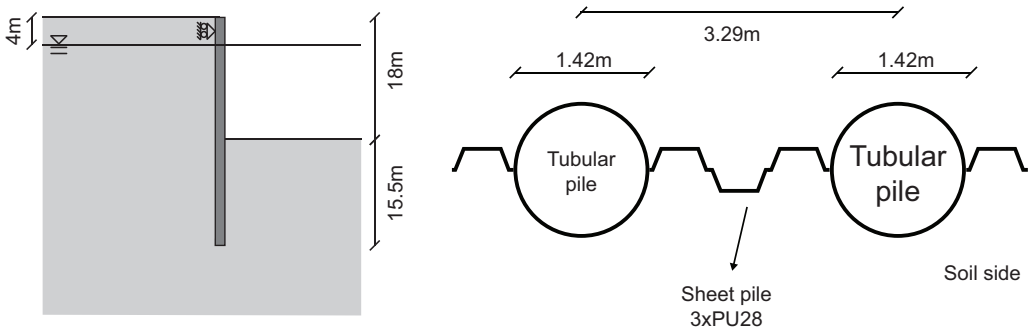


Figure 4: Geometry of the considered anchored QW

As a comparison, the soil behaviour is reproduced by means of the hardening soil model (HS) and an elastic-perfectly plastic constitutive law. In this second case, a Mohr-Coulomb failure surface with a non-associated flow rule is considered. The HS model parameters (Table 2) were chosen on the basis of the values adopted for a medium dense sand. The elastic perfectly plastic constitutive law properties were chosen by considering: (i) the same friction and dilatancy angle with respect to those of Table 2, (ii) the same Poisson ratio and (iii) an elastic stiffness that correspond to the HS elastic-plastic one (E_{50}) at the wall mid point.

Table 2: Hardening soil parameters

E_{50}^{ref} (MPa)	E_{oed}^{ref} (MPa)	E_{ur}^{ref} (MPa)	ν (-)	m (-)	ϕ' (°)	ψ (°)	R_f (-)
45	45	135	0.2	0.5	34	3	0.9

The results in terms of wall bending moment at the end of the excavation for the two cases (HS, solid line, whereas the elastic-perfectly plastic MC, dashed line) reported in Figure 5a are in good agreement. The results in terms of wall horizontal displacement are still in reasonable agreement, even if the MC results clearly cannot reproduce the response of the passive side at the wall toe Figure 5b. This is because with elastic perfectly plastic models exhibit the same stiffness under loading and reloading/unloading. However, the real problem concerning the MC results is related with the displacements at the ground surface in the land side (Figure 5c). In fact in this

case the numerical results clearly show an unrealistic heave. This unrealistic behaviour, that can also be observed for different geometries/soil mechanical properties (and also other excavation problems, such as tunnel excavations), highlight that for QWs the use of simple elastic-perfectly plastic constitutive laws should be avoided.

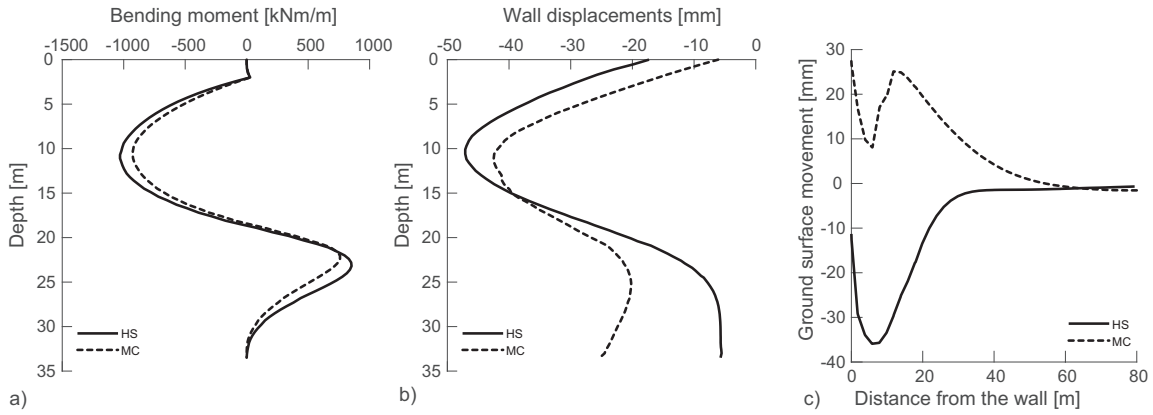


Figure 5: Influence of soil constitutive law on numerical results: a) wall bending moments, b) wall displacements and c) landside settlements

The results reported in this section highlight that, even when the same basic mechanical parameters (strength and stiffness) are adopted, different constitutive models may provide different predictions because they represent stiffness, stress dependency, unloading/reloading behaviour, and dilatancy in different ways. On the one hand, more sophisticated constitutive models can better reproduce some aspects of soil behaviour; on the other hand, they generally require the calibration of a larger number of parameters. In practical applications, however, the available in-situ and laboratory data may not always be sufficient to reliably determine all the required parameters. In such cases, the selection and calibration of the constitutive model necessarily rely, to a certain extent, on the engineering judgement and experience of the designer. Increasing the complexity of the constitutive model does not necessarily result in more reliable predictions if the available data do not support an adequate calibration.

4.3 Simplyfing the modelling approach through FE

As previously discussed, RPWs are highly complex structures that cannot be analysed using simplified methods. Consequently, their design relies computationally demanding FE analyses. This limits both the number of configurations (e.g. pile diameter, length, and spacing) that can be investigated and the application of probabilistic approaches for reliability assessment (Montali et al., 2026). The objective of this section is to demonstrate that, by adopting an appropriate yet simple modelling approach, the response of an RPW can be reproduced by an equivalent AW, allowing the analysis of the system by means of simplified subgrade reaction methods.

The reference case considered is the RPW presented in Section 4.1. The numerical results (Figure 6a, adapted from Montali et al., 2026) show that, at the end of the excavation process, the active

soil wedge on the landside (i) is only marginally influenced by the piles and (ii) detaches from the relieving platform. As a result, the vertical stresses are transferred entirely to the wall and the piles.

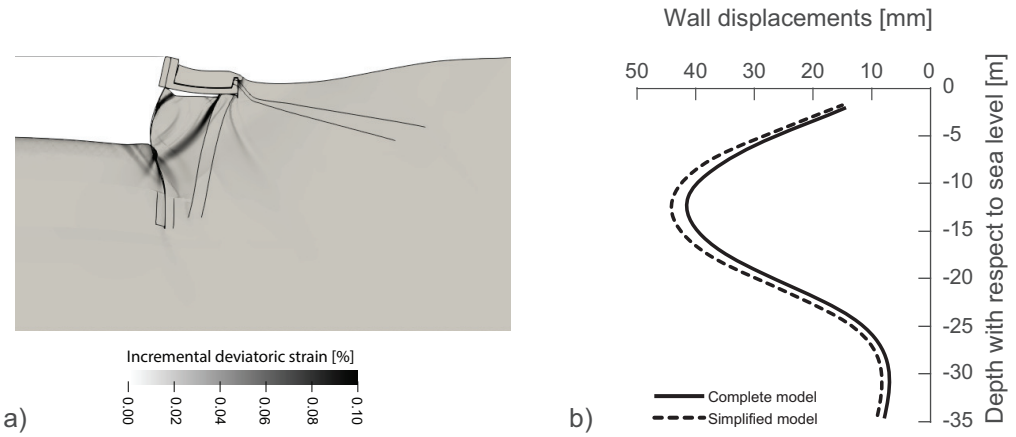


Figure 6: a) deviatoric plastic strain increments relative to the case of section 4.1 and b) comparison between the complete and the simplified model

Based on these observations, Montali et al. (2026) proposed that, when only the wall response is of interest, the RPW can be represented by a simplified structural model in which the relieving platform and the piles are omitted, while the anchors are modelled as springs. In this simplified model, (i) the combi-wall extends only to the foundation level of the relieving platform, (ii) an initial surcharge equivalent to the self-weight of the soil prior to construction is applied at this level, and (iii) this surcharge is removed before the first excavation stage. The application and subsequent removal of this surcharge leave the soil on the active side of the wall in an overconsolidated state. The comparison of wall horizontal displacements at the end of the excavation shown in Figure 6b demonstrates excellent agreement between the simplified (dashed line) and the complete FE models (solid lines), with a maximum discrepancy lower than 5%.

The range of applicability of this simplified approach, particularly with respect to soil properties, wall geometry, and stiffness of the relieving platform, is currently under investigation. In addition, ongoing research is focused on how the induced overconsolidation can be reliably incorporated into the constitutive formulations adopted in subgrade reaction models.

5 Conclusions

The design of modern quay walls increasingly requires methodologies that explicitly account for soil-structure interaction to achieve safe, economical, and sustainable solutions. Traditional analytical approaches remain valuable for preliminary design, but their simplifying assumptions often lead to (very) conservative designs.

Subgrade reaction methods provide an effective intermediate approach by incorporating soil-

structure interaction with limited computational effort, making them well suited for conceptual design and evaluating alternative solutions. Nevertheless, their applicability is restricted to relatively simple geometries and loading conditions.

Finite element analyses offer the most comprehensive framework for modelling quay walls. However, the direct application of characteristic/design soil parameters within non-linear finite element analyses may produce overly conservative predictions and unrealistic failure mechanisms, indicating the need for further developments in partial factor based design frameworks. The reliability of the results strictly depends on the adopted constitutive model, that have to be chosen with particular attention. Finally, recent developments in simplified finite element based modelling approaches show that even complex relieving platform quay walls can be represented by equivalent simplified models with excellent accuracy.

These developments provide a promising pathway towards displacement-based design methodologies that combine the accuracy of advanced numerical analyses with the efficiency required for engineering practice. Continued research on simplified modelling techniques, constitutive formulations, and probabilistic design approaches will further support the development of more economical, sustainable, and reliable quay wall infrastructure.

References

[1] Blum, H. (1931): Einspannungsverhältnisse bei Bohlwerken und deren vereinfachte Berechnung mit Hilfe von “ideelle” und “stellvertretender” Belastung, PhD thesis, Technische Hochschule Carolo Wilhelmina zu Braunschweig

[2] CLARION (2024). Climate resilient port infrastructures https://projectclarion.eu/

[3] CUR 166 Damwandconstructies, 6e herziene druk, CUR Bouw & Infra, Gouda 2012.

[4] de Gijt, J. G., & Broeken, M. L. (2013). Quay walls. CRC Press.

[5] EN 1997-1: 2004, Eurocode 7 - Geotechnical design, Part 1: General rules, European Committee for Standardization, Brussels.

[6] Flessati, L., & di Prisco, C. (2023). Deep tunnel faces in cohesive soils under undrained conditions: application of a new design approach. *European Journal of Environmental and Civil Engineering*, 27(8), 2630-2644.

[7] Hao, N., Duffy, K., Reale, C., and Gavin, K. (2026): Performance-based reliability assessment of quay walls. *Computers and Geotechnics*, 191, 107794.

[8] International Chamber of Shipping. (2024). Shipping and world trade: Global supply and demand for seafarers. http://www.ics-shipping. org/shipping-fact/

[9] Lai, F., Duffy, K., Gavin, K., Lu, D., & Roubos, A. (2025): Integration of Field Monitoring and Numerical Modelling to Evaluate the Construction Performance of a Deep-Sea Quay Wall. *Journal of Geotechnical and Geoenvironmental Engineering*, 151(12), 04025156

[10] Montali, F., Alver, O., Gavin, K., Jommi, C., Flessati, L. (2026): A simplified modelling approach for the analysis of port quay-walls with relieving platform International Conference on Advances and Innovations in Soft Soil Engineering 24 - 26 August, Delft

[11] Roubos, A. (2026): Impact Infra Innovation Soil - Structure - Materials, 2nd Ports and Waterways Conference, 4-6 February 2026, Delft

[12] Roubos, A., Flessati, L., Alver, O., Gavin, G. (2026): CLARION: Smart & sustainable climate resilient Quay Walls Proceedings of the 21st ICSMGE, Vienna, Austria, 14-19 June 2026.

[13] Schanz, T., Vermeer, P. A., and Bonnier, P. G. (1999): The hardening soil model: Formulation and verification. *Beyond 2000 in computational geotechnics*, Routledge, 281-296

6 Acknowledgements

The authors gratefully acknowledge the contributions of A. Roubos (Port of Rotterdam and TU Delft), O. Alver (TU Delft), F. Montali (TU Delft), and K. Gavin (TU Delft). This paper is supported by the Clarion project, which has received funding from the European Union's Horizon Europe research and innovation program under Grant Agreement 101147041.

Author

Dr. Luca Flessati
Technische Universiteit Delft
Faculteit Civiele Techniek en Geowetenschappen
Stevinweg 1
2628 CN Delft
e-mail: `l.flessati@tudelft.nl`

A comparative study of analytical and numerical based design approaches for rigid inclusions

Prerana Krishnaraj, Christopher Tinat

Abstract: Columns for ground improvement constructed from high-stiffness and high-strength materials are internationally known as rigid inclusions (RI). In the forthcoming generation of Eurocode 7 Part 3 (EC 7-3), rigid inclusions are classified as Ground Improvement Class BII.
Rigid inclusions may be installed beneath isolated foundations or arranged in a grid pattern to provide area-wide ground improvement, for example beneath slabs or slab-on-rafts. Unlike conventional deep foundation systems, load transfer is achieved either through a load transfer platform (LTP) placed above the inclusion heads or through direct contact between the supported structure and the inclusions without structural connection reinforcement.
The design of rigid inclusion systems requires a thorough understanding of the interaction mechanisms between the inclusions, the load transfer platform, and the supported structure. Accurate prediction of load sharing and deformation behavior is therefore essential for a reliable and economical design. Using the example of a reinforced foundation slab from a case study in Hamburg, this paper illustrates the governing load transfer mechanisms and demonstrates the importance of accurately modelling the rigid inclusion network and its interaction with the surrounding soil and overlying structure.

1 Introduction

1.1 Background and Motivation

Historically, structures constructed on soft soils were commonly founded on deep foundation elements, such as piles, often resulting in significant construction costs. In the early 1990s, Menard developed the innovative concept of Controlled Modulus Columns (CMC), the origin technique to produce rigid inclusions (RI). This technology consists of a network of rigid inclusions installed down to a load-bearing layer and structurally disconnected from the supported structure, typically by means of a load transfer platform (LTP). Both the inclusions and the surrounding soil contribute to the foundation system. Using rigid inclusions brings therefore similar effectiveness as piling while retaining the advantages of a shallow foundation system. When the soil, in its current state, cannot ensure the integrity of the structure, ground reinforcement by RI can provide an efficient foundation system by mitigating settlements, increasing the bearing capacity and/or ensuring the stability.

The foundation, the load transfer platform, the rigid inclusions, and the soil between the inclusions interact to form a combined inclusion-supported foundation system. Complex interaction mechanisms occur within this system and must be considered when assessing its deformation behaviour.
The motivation of this paper is to present current deformation-based design approaches and, in particular, to introduce numerical modelling techniques for the analysis and design of rigid inclusions. Furthermore, the paper aims to provide an outlook on areas where further research is required in order to improve the reliability and accuracy of settlement and load-transfer predictions.

1.2 Rigid Inclusions in the new Eurocode framework

In the future generation of Eurocode 7, which is currently in the final stages of publication, ground improvement is addressed in much greater detail in EC 7-3, "Geotechnical Structures", Chapter 12. In principle, the following four classes of ground improvement are distinguished:

Table 12.1 — Classification of ground improvement

Class	A – Diffused	B – Discrete
I	**AI – Diffused with no measurable unconfined compressive strength** The improved ground has an increased shear strength or stiffness higher than that of the original ground. The improved ground can be modelled as a ground with improved properties.	**BI – Discrete with non-rigid inclusions** Inclusions, installed in the ground, with higher shear capacity and stiffness compared to the surrounding ground. The unconfined compressive strength of the inclusion is not measurable.
II	**AII – Ground improvement zone with measurable unconfined compressive strength** The improved ground is modified from its original natural state, has a measurable unconfined compressive strength and is significantly stiffer than the surrounding ground. Usually, it comprises a composite of a binder and ground.	**BII – Discrete with rigid inclusions** Rigid inclusions, installed in the ground, with unconfined compressive strength and significantly higher stiffness than the surrounding ground. The inclusions can be an engineered material such as timber, concrete/grout or steel or a composite of a binder and ground.

Figure 1: Classification of ground improvement (prEN 1997-3)

While the methods belonging to Classes AI and AII aim to achieve an area-wide or volumetric change of soil parameters, such as increasing shear strength, increasing stiffness, or altering permeability, Classes BI and BII consist of inclusions, i.e. discrete elements whose properties differ from those of the surrounding soil.
The distinction between Class I and Class II ground improvement methods is based on the unconfined compressive strength (UCS). Ground improvement methods of Class I do not develop measurable compressive strength, whereas for Class II, material samples can be subjected to compressive strength testing.
Consequently, Class BI comprises granular columns (non-rigid inclusions). The behaviour of these columns is defined by conventional soil parameters, such as the Young- or oedometric modulus and shear parameters.
Class BII covers the rigid inclusions (RI) that are the focus of this paper. These columns are significantly stiffer than those of Class BI, and the column material possesses its own compressive strength. Their behaviour is not defined by soil parameters but rather by the load-settlement response during load-bearing, which for an individual inclusion is similar to that of a pile. Both shaft resistance and base resistance are mobilized.

In the future, the verification of ultimate limit state (ULS) will be based on a system resistance, see Eq. 1. This approach considers both the inclusions and the soil located between the columns, each providing a design resistance that is combined to obtain the overall system resistance. Consequently, if the stability of the structure can be ensured solely by the surrounding soil, it is not necessary to assess the load-bearing capacity of the columns themselves. In such cases, the design may be governed exclusively by serviceability criteria, particularly deformation-related requirements.

$$R_{sys,d} = \frac{\sum_i^n R_{ri,i}}{\gamma_{Rd,ri}\gamma_{R,ri}} + \frac{R_g}{\gamma_{Rg}} \quad (1)$$

$R_{ri,i}$ is the representative value of the vertical resistance of the i-th rigid inclusion;
R_g is the representative value of the vertical resistance of the ground after installation of inclusions;
$\gamma_{Rd,ri}$ is a model factor;
$\gamma_{R,ri}$ is a partial resistance factor for the rigid inclusion system and
γ_{Rg} is a partial factor for the ground after treatment

1.3 Rigid Inclusions vs Pile Foundations

Structural loads are traditionally transferred to the ground through pile groups, for which group interaction effects must be considered during design. In conventional pile group foundations, the applied loads are assumed to be carried exclusively by the piles, which are connected by a pile cap. Although the pile cap is often in direct contact with the surrounding soil and may contribute to load transfer in practice, this contribution is generally neglected in design. Consequently, the design of pile groups is primarily governed by Ultimate Limit State (ULS) requirements of the piles.
In contrast, the piled raft concept considers the piles and the raft as a single integrated foundation system. Both components actively participate in load transfer, with the piles serving to supplement the bearing capacity of the raft and, more importantly, to reduce total and differential settlements. As a result, a significant proportion of the structural load can be carried directly by the raft, unlike conventional pile groups where the piles are assumed to support the entire load. Accordingly, the design of piled raft foundations is predominantly governed by Serviceability Limit State (SLS) considerations, with settlement performance constituting the primary design criterion.

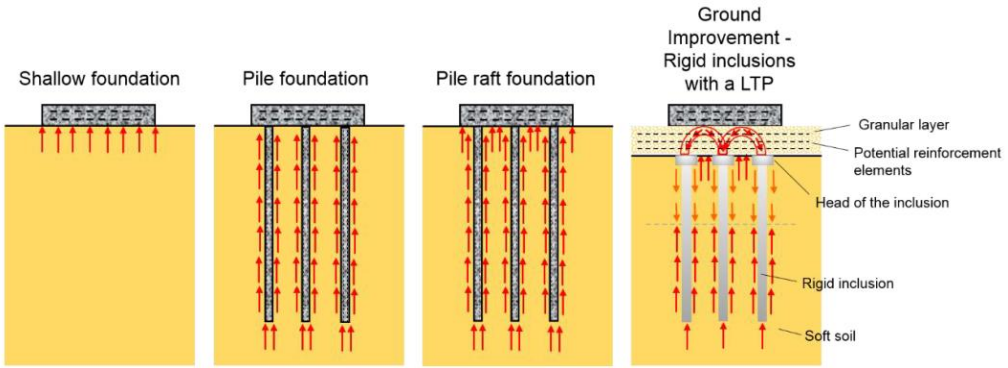


Figure 2: Type of load transfer in the different usual foundation concepts according to Varaksin et al. (2016)

In the new Eurocode, rigid inclusions (RI) are distinguished from pile verifications in Section 12.1(3) by means of two defining criteria:

- the presence of a load transfer platform (LTP) between the inclusions and the supported structural element or structure, or
- the absence of structural connection reinforcement in the case of direct support without a load transfer platform.

2 Deformation based Design Approaches for Rigid Inclusions

2.1 Analytical Design

2.1.1 Load Transfer Models

For a grid of rigid inclusions, load distribution can be analysed using the load transfer method based on a unit cell approach. In this method, the unit cell is divided into two interacting calculation domains: the rigid inclusion and the surrounding soil. The load shares carried by the inclusion and the soil are determined iteratively until both domains exhibit the same settlement response, thereby satisfying the condition of uniform settlement beneath a relatively rigidly loaded structure. This procedure corresponds to the MV2 calculation method described in IREX (2012).

The interaction between the rigid inclusion and the surrounding soil is governed by the mobilized shaft shear stresses along the inclusion and the mobilized base resistance at its toe. These mechanisms develop as a function of the relative displacement between the inclusion and the surrounding soil at a given depth. The deformation behaviour of the inclusion is evaluated using load-transfer relationships (load-transfer curves), which describe the progressive mobilization of shaft and base resistance with displacement.

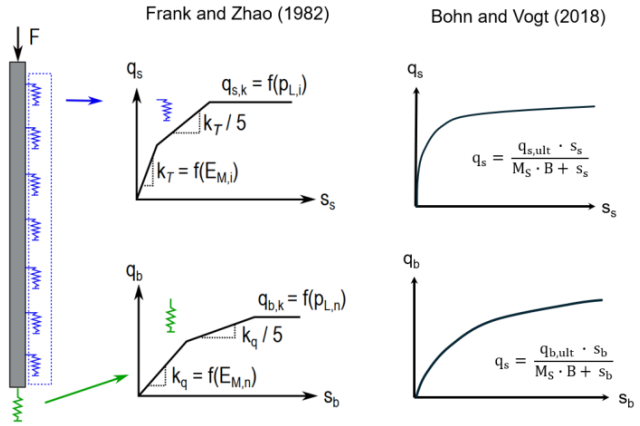


Figure 3: Mobilisation laws for RI. Left: Frank and Zhao (1982), Right: Hyperbolic approach by Bohn and Vogt (2018)

Various load transfer curve approaches have proven suitable for the design of rigid inclusions (RI). An overview and comparison of these methods are provided by Burtin and Racinais (2024). In Germany, it is common practice to use the empirically derived cube-root or hyperbolic functions proposed by Bohn and Vogt (2018), or the semi-empirical load transfer curves

developed by Frank and Zhao (1982). The corresponding modelling approaches are illustrated in Figure 3.

2.1.2 LTP Verification

The stress concentration at the head of the rigid inclusion, as determined using the load transfer method, must be verified against the load transfer platform (LTP) capacity through an appropriate consistency check. In principle, this verification is analogous to the plausibility checks performed in the Priebe method for Class BI granular columns.
The limiting stress of a Class BII rigid inclusion corresponds to the maximum stress that can be sustained at the inclusion head. It therefore defines the upper bound for which linear-elastic deformation analyses, and consequently the load-sharing calculations described above, remain applicable.
Methods for evaluating the ultimate limit state (ULS) resistance at the interface between the rigid inclusion and the load transfer platform are generally based on classical bearing failure theory. According to Tinat et al. (2021), the maximum stress at the inclusion head is governed by a conventional bearing capacity failure mechanism developing within the load transfer platform. From a mechanical perspective, this mechanism can be interpreted as an inverted bearing capacity failure, in which the rigid inclusion acts as the loading element and the surrounding load transfer platform provides the resisting medium.

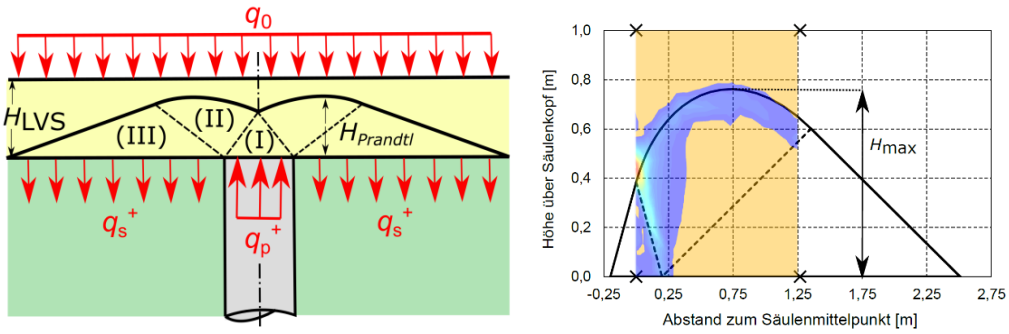


Figure 4: Consistency Check of the stress at the RI head. Left: Analytical approach using Prandtl's failure mechanism, Right: FE Analysis of an unit cell according to Tinat, Kirstein and Grabe (2021)

Analogous to conventional bearing capacity calculations, it is assumed that failure develops along a distinct shear surface. In cross-section, the theoretical failure surface is represented by a logarithmic spiral.
If the maximum allowable stress is exceeded by the stress obtained from the load-sharing calculations, for example when applying the above-mentioned load transfer curves, the column spacing or column diameter must be adjusted such that this condition is no longer violated and the compatibility requirement is satisfied. In other words, the mobilized stress at the column head must remain lower than the limiting stress defined by the LTP.
If this condition is not fulfilled, the assumed load-sharing model is simply no longer valid, since once the limiting stress has been reached, any additional load can only be carried by the surrounding soil. Alternatively, plastic deformation analysis at the interface between the column head and the load transfer platform may be taken into account. This can be achieved analytically by reducing the stiffness of the load transfer platform until the limiting stress is no longer exceeded, or by performing finite element analyses using appropriate constitutive soil models.

2.2 Numerical Design

2.2.1 Finite Element Modelling

The 3D numerical modelling was carried out using the finite element software PLAXIS 3D. The soil layers were modelled using both Mohr-Coulomb (MC) and Hardening Soil (HS) models. The rigid inclusions were modelled as volume elements. Interface elements were provided along the inclusion circumference and the inclusion tip. Wished-in-place approach was opted for the numerical modelling. This approach does not consider the disturbance and the soil property changes that occur during the inclusion installation. The rigid inclusions were modelled as non-porous linear elastic material.

3 Case Study: Hamburg Project

3.1 Project description and soil model

The project under consideration is a logistics platform located in Hamburg. The site is characterized by the geotechnical conditions that are typical for the Hamburg region, including highly compressible organic soft soils, which rendered a shallow foundation solution infeasible. RI were therefore installed using the Controlled Modulus Columns (CMC) method. The CMC technique is a full-displacement process that minimizes soil excavation and improves installation efficiency.
Due to the typically grid-shaped arrangement of the CMC columns, the ground was investigated in great detail during construction through the continuous recording of drilling parameters, providing valuable information in addition to the available site investigations. Variations in geological conditions could be identified during installation, allowing the column lengths to be adjusted accordingly. This approach ensures a uniform load-settlement behaviour even under conditions of significant stratigraphic variability or rapidly changing soil properties.

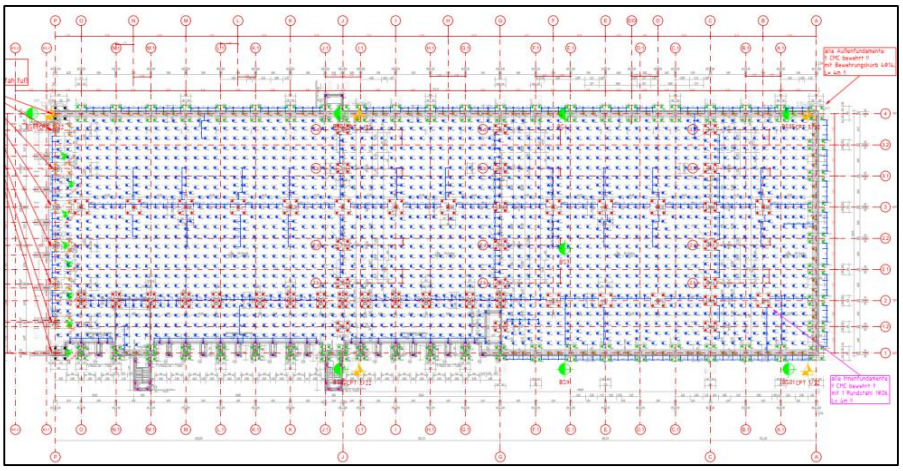


Figure 5: Layout of the RI installed beneath the logistics platform in Hamburg (grid: 2.25 m, D = 360 mm)

In the present project, CMC columns with a diameter of at least 360 mm were installed in a grid spacing of 2.25 m beneath the slab. Column lengths ranging from 5 m to 10 m due to

heterogeneous layering enabled the column toes to be embedded in the underlying competent sand and gravel layers. In total, more than 14,000 linear metres of columns were installed to support the logistics platform.

Figure 6 shows the schematic representation of the soil stratigraphy considered in the present study.

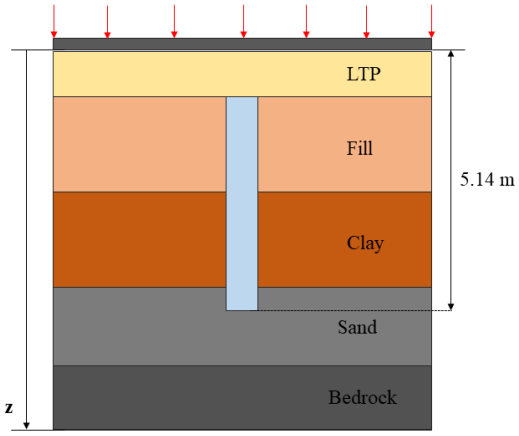


Figure 6: Schematic representation of soil stratigraphy

Table 1 summarises the material properties to be used in both analytical and numerical modelling.

Table 1: Properties of the soil used in analytical and numerical modelling

Layer	**z_{top} (m)**	**z_{bottom} (m)**	**E (kPa)**	**Poisson's ratio**	**Unit weight (kN/m^3)**	**q_s (kPa)**	**ktanδ**
LTP	0.0	0.9	59,514	0.3	20		1
Fill	0.9	2.64	18,501	0.3	18	160	0.45
Clay	2.64	4.14	1,858	0.3	12	30	0.20
Sand	4.14	7.14	29,700	0.3	19	220	1.00

3.2 Model selection

For numerical modelling, the soil layers were modelled using both Mohr Coulomb (MC) and Hardening Soil (HS) models. The parameters required for the MC model were obtained directly from Table 1, and the parameters required for the HS model were interpolated from the parameters given in Table 1.

The rigid inclusions considered were cylindrical concrete columns of diameter of 0.36 m and a centre to centre spacing of 2.25 m. They had a Young's modulus value of 12.1 GPa and 28th day compressive strength of 35 MPa. They extended throughout the layer of the compressible soil with their top coinciding with the top of the fill layer (Figure 6).

Two different numerical modelling approaches were adopted in the study. The first approach was inspired by unit cell method for settlement calculation in a rigid inclusion reinforced system. This approach employed a modified unit-cell method, where a rigid inclusion groups of 1 x 1 (1 inclusion), 3 x 3 (9 inclusions) and 5 x 5 (25 inclusions) were analysed. The entire surface of the model was loaded. Similar to the unit cell approach, the loading area extended up to a

distance equal to 0.5 times the inclusion spacing beyond the centre of the outer most inclusions. The second approach modelled the footing resting on the reinforced soil with the lateral soil boundaries not coinciding with the footing boundary. This gives a better representation of the behaviour of an actual footing-soil system. Figure 7 shows the profile considered for both the approaches for a single rigid inclusion system. A static surcharge pressure of 55 kPa was applied at the top of the LTP through rigid concrete footing.

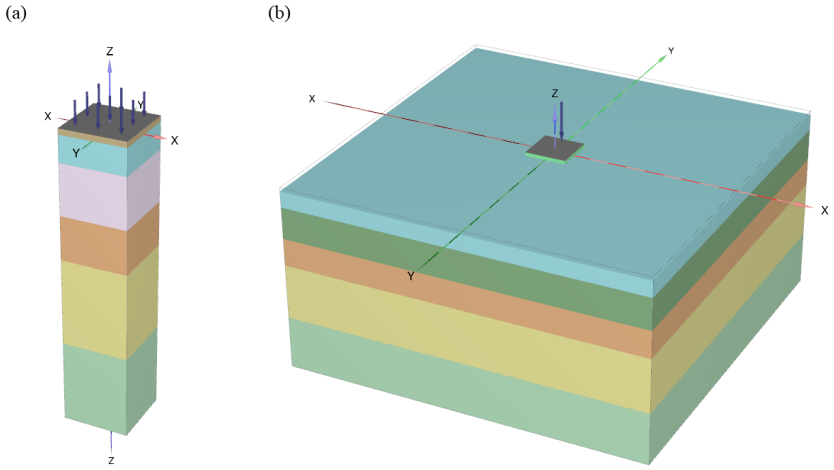


Figure 7: Soil profile used for (a) modified unit cell system and (b) footing-soil system

Table 2 summarizes the maximum surface settlements obtained using these two numerical modelling approaches with both MC and HS constitutive models. For the modified unit cell approach, the surface settlement was found to be least influenced by the number of inclusions, with MC model yielding a settlement of about 23 mm and the HS model yielding a settlement of about 41 mm. The analytical solution predicted the surface settlement of 24 mm. As compared to the analytical solution, the HS model significantly overestimated the surface settlement, while the MC model showed close agreement with the analytical results. The analytical solution assumes homogeneous soil with constant Young's modulus with soil behaving as a linear elastic material before yielding. These assumptions are similar to those for the elastic range in the MC model. This may explain the close agreement between the analytical and MC model based numerical modelling results.

In contrast, for the footing approach, the predicted settlements were significantly influenced by the number of inclusions considered, with the settlements increasing with the increase in the number of inclusions for both the material models. As the footing area increases, the stress influence zone extends deeper into the soil and hence larger volume of the compressible soil experiences significant stresses. This leads to higher surface settlement (Das and Sivakugan, 2017).

Table 2: Surface settlements obtained from different approaches in numerical modelling

Inclusion grid	**Maximum surface settlement (mm)**			
	Unit cell		**Footing**	
	MC model	**HS model**	**MC model**	**HS model**
1 x 1	23.08	41.57	6.16	16.69
3 x 3	23.11	41.41	20.07	37.27
5 x 5	22.93	48.66	24.76	53.63

Based on the results obtained from both the approaches considering the MC and HS model, the modified unit cell approach with soil modelled using MC constitutive model is selected for further analysis.

3.3 Comparison of settlement and load acting on the inclusion

The variation of the settlement in the inclusion and the surrounding soil obtained from the numerical and analytical approach is plotted in Figure 8. The model successfully predicted the soil settlement variation with depth, consistent with that of the analytical solution. However, the model overestimated the inclusion settlement, with the inclusion settlement being about 6 mm from the analytical model and 14 mm from the numerical model. The numerical model overestimated the inclusion settlement by 2.33 times. This might be attributed to the fact that the columns are modelled in the numerical analyses as wished-in-place elements, without accounting for the changes in soil properties induced by the full-displacement installation process. In contrast, the analytical approach is based on empirical load-transfer relationships derived from field measurements on full-displacement columns, which inherently capture the soil improvement effects caused by the installation process.

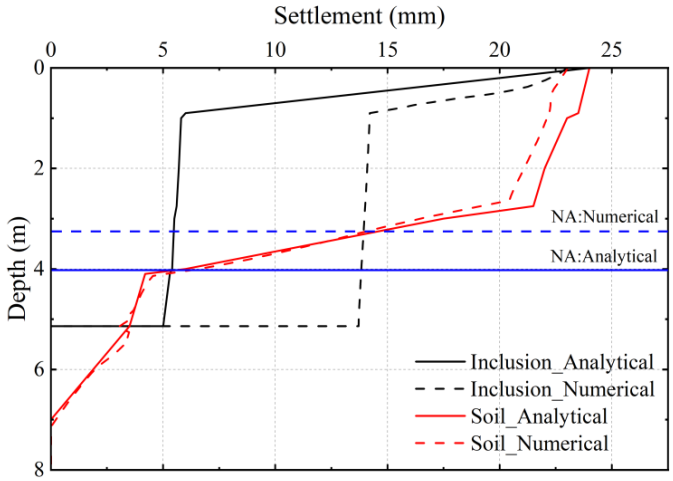


Figure 8: Comparison of the inclusion and the soil settlement obtained from the analytical and the numerical approach (NA: Neutral axis)

The variation of the vertical load acting on the inclusion predicted by the numerical and the analytical model is shown in Figure 9. It shows that the numerical model underpredicts the maximum load acting in the inclusion by 22.22%. Tinat et al. (2021) already demonstrated that, in highly compressible soils, mobilized stresses on the column head can reach the maximum stress within the LTP. Consequently, plastic deformations may develop, resulting in substantial deformations of the overall system. This phenomenon is associated with the punching of the column into the LTP, whereas no such effect is observed in stiffer soil conditions. To address this, a new series of analytical and numerical modelling was carried considering the clay with Young's modulus of 12.5 MPa. The analysis was carried out considering the LTP of two different thicknesses, 0.6 m and 0.9 m.

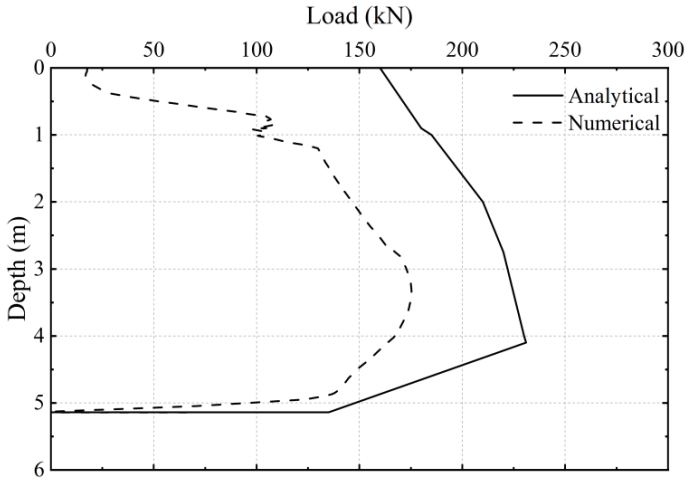


Figure 9: The analytical and numerical modelling results for the vertical load acting on the inclusion

3.4 Analysis of the system with clay having Young's modulus of 12.5 MPa

The plots for inclusion and soil settlement for LTP of thickness 0.6 m and 0.9 m are shown in Figure 10. It is evident from the plots that the numerical model predicted the maximum surface settlements similar to that of the analytical model. However, overprediction of inclusion settlement was observed in these cases as well, with numerically predicted inclusion settlement being 1.75 and 1.5 times the analytical prediction for the LTP thicknesses of 0.6 m and 0.9 m, respectively.

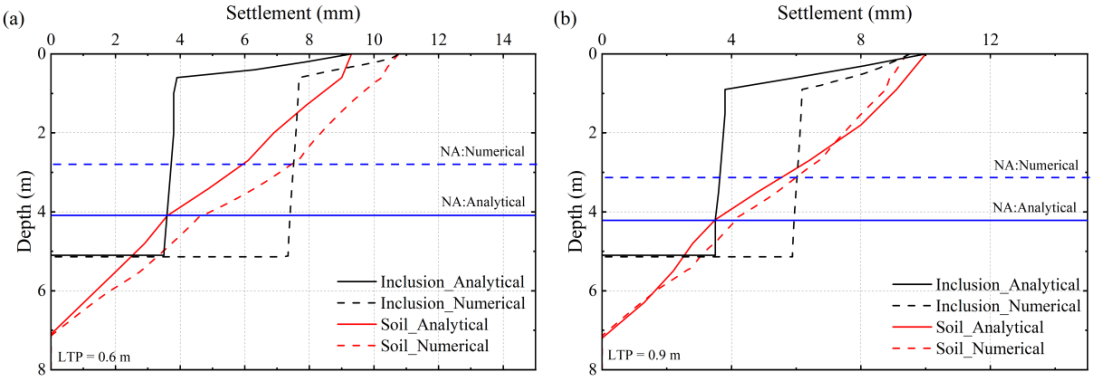


Figure 10: The inclusion and the soil settlement variation for the LTP thickness of (a) 0.6 m and (b) 0.9 m

The vertical load acting on the inclusion is plotted in Figure 11. Unlike the case of low strength clay, the plots obtained for these two cases are in good agreement with each other. However, the depth at which the maximum load was observed differed due to the difference in the neutral axis location for both numerical and analytical methods (Figure 10).

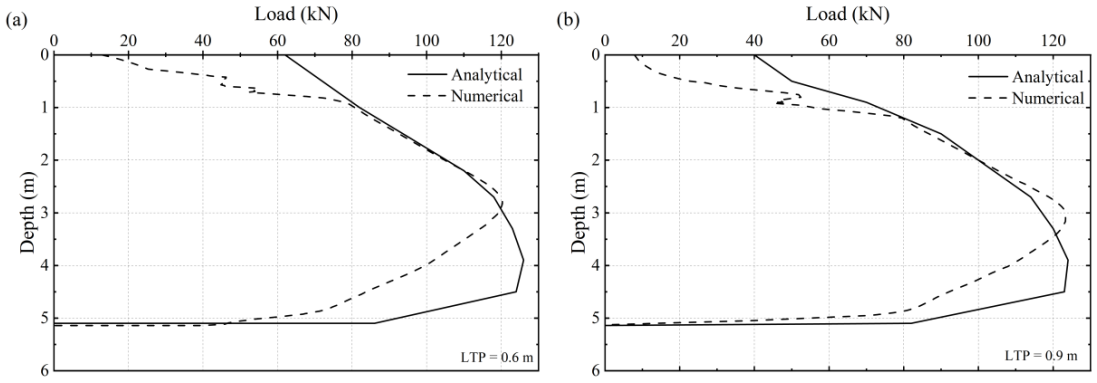


Figure 11: Comparison of the load acting on the inclusion for the LTP of thickness of (a) 0.6 m and (b) 0.9 m

4 Summary and Outlook

Rigid inclusions (RI) are a ground improvement technique that, in most practical applications, is designed primarily with respect to settlement and deformation criteria, particularly when the ULS verification of the structure can already be ensured by the natural ground. Consequently, accurate prediction of the deformation behaviour is of paramount importance for the design and assessment of the composite ground-improvement system.

Based on the comparison between the numerical simulations and the analytical calculations, the following conclusions can be drawn:

- Both approaches, analytical and numerical, show good agreement with respect to the settlement behaviour of the soil between the columns.
- However, regarding the settlement response of the columns, the analytical methods predict a significantly stiffer load-settlement behaviour than the numerical analyses. It is assumed that this discrepancy results from the fact that the columns are modelled in the numerical analyses as wished-in-place elements, without accounting for the changes in soil properties induced by the full-displacement installation process. In contrast, the analytical calculations are based on empirical load-transfer relationships derived from field measurements on full-displacement columns, thereby implicitly incorporating installation-induced soil improvement effects. A similar observation was reported by Racinais et al. (2017).
- The use of advanced constitutive soil models is particularly recommended when very soft soils are present and the load concentration at the column head reaches the maximum allowable stress within the LTP, according to Tinat et al. (2021). Under such conditions, plastic deformations may develop and should be appropriately captured in the analysis.
- When the stresses acting at the column head remain significantly below the limiting stress level of the LTP, the use of advanced constitutive models is not necessarily required. Simpler modelling approaches, such as the Mohr-Coulomb model, are generally sufficient for practical engineering applications.

The study highlights a significant need for further research regarding the incorporation of installation effects into numerical analyses of rigid inclusions. This is particularly relevant for full-displacement installation techniques, where substantial modifications of the surrounding soil properties occur during construction. These effects include soil densification beneath the

column toe and increased lateral confinement within the surrounding ground. Such installation-induced changes can lead to a significantly stiffer load-settlement response of the columns and should therefore be considered in future numerical modelling approaches.

References

[1] Bohn, C., Vogt, N. (2018): Lasttransfermethode zur Berechnung von Gründungen und Baugrundverbesserungen mit starren Säulen. *Bautechnik*, 96(9): 597-606

[2] Das, B. M., & Sivakugan, N. (2017). Fundamentals of geotechnical engineering. Cengage Learning.

[3] Eurocode 7 - Geotechnical design - Part 3: Geotechnical structures, Final Draft, October 2024.

[4] Frank, R., Zhao, S.-R. (1982): Estimation par les paramètres pressiometriques de l'enforcement sous charge axiale de pieux forés dans des sols fins. *Bulletin de liaison des LPC*, 119, 17-24.

[5] Irex (2012): Recommendations for the design, construction and control of rigid inclusion ground improvements, ASIRI National Project.

[6] Racinais, J., Maucotel, F., Varaksin, S., & Babak, H. (2017). Beneficial Use of Pressuremeter Tests for Accurate Modelling by Finite Elements of a Rigid Inclusion Ground Improvement Solution. In Proc. 19th Int. Conf. Soil Mechanics and Geotechnical Engineering.

[7] Tinat, C., Kirstein, J., Grabe J. (2021): Zur Tragfähigkeit von granularen Lastverteilungsschichten über Stabilisierungssäulen. *Geotechnik,* 44(2): 92-101

[8] Varaksin, S., Hamidi, B., Huybrechts, N., Denies, N. (2016): Ground improvement vs. pile foundations. ISSMGE–ETC, 3, 1-48.

Authors

Prerana Krishnaraj, PhD
Technische Universität Hamburg
Institut für Geotechnik und Baubetrieb
Schellerdamm 22 (Channel 7)
21079 Hamburg
Tel.: +49 (0) 40 - 42878 / 3890
e-mail: prerana.krishnaraj@tuhh.de
Web: www.tu-harburg.de/gbt

Christopher Tinat, MSc
Member of Working Group 2.8 DGGT
MENARD GmbH
Hittfelder Kirchweg 2
21220 Seevetal
e-mail: ctinat@menard.gmbh
Web: www.menard.gmbh

Installation method influence on the prediction of long-term monopile deformations – A simplified numerical approach

Viet Hung Le, Frank Rackwitz

Abstract: The installation method has a significant impact on the deformation behaviour of the monopile. However, these effects are often not considered in simulations due to the complexities of pile penetration processes. This study therefore develops a simplified numerical model to consider the effects of monopile installation using an impact and vibratory hammer. This model mimics the changes in stress conditions and void ratio in different zones around the pile. The long-term deformation of the pile under lateral cyclic loading was calculated using an explicit calculation method and a high cyclic accumulation model. The developed model was successfully validated using the results of small-scale model tests. For demonstration purpose, the performance of a full-scale monopile under vibratory driving was predicted.

1 Introduction

For the installation of large offshore monopiles, impact driving remains the most widely used method. However, in recent years, vibratory driving has emerged as an attractive alternative because it reduces underwater noise emissions, lowers the risk of pile run, and significantly decreases the stresses induced in the pile during installation. While impact driving moves the pile through a sequence of high-energy impulsive loads, vibratory driving relies on continuous oscillatory loading with comparatively small amplitudes. These fundamentally different penetration mechanisms produce different stress states and soil densities in the vicinity of the pile, resulting in differences in the deformation behaviour and the bearing capacity of the installed pile.

Despite these differences, the influence of installation effects on the subsequent deformation behaviour of monopiles is still not fully understood. Owing to the extensive experience and large field data available for impact-driven piles, their behaviour can generally be predicted with reasonable accuracy. In contrast, the performance of vibratory-driven piles remains more difficult to assess, particularly with regard to the long-term response under lateral cyclic loading. Uncertainties associated with vibratory driving have led to the continued use of impact driving for the installation to the final depth in several offshore projects (De Neef & Middendorp, 2013).

Numerical methods are well established in geotechnical engineering for analysing complex boundary value problems. Nevertheless, accurately simulating pile installation and its influence

on the long-term cyclic response remains challenging because of the highly complex penetration process and the accumulation of soil deformations under cyclic loading. This paper presents a simplified numerical approach for incorporating the effects of the impact and vibratory driving into the analysis of laterally loaded monopiles. Combined with the High-Cycle-Accumulation (HCA) model, the proposed approach enables the influence of installation effects on the long-term behaviour of monopiles to be investigated efficiently.

2 Calculation concept

2.1 Soil conditions due to impact and vibratory driving

During pile installation, the soil conditions surrounding the pile change differently depending on the installation method employed. The cyclic and dynamic movement of the pile induces highly complex mechanical processes in the soil adjacent to the pile shaft and the pile tip. The resulting pile/soil interaction may involve fluidisation, liquefaction, large local deformations, and cavitation. To improve the understanding of these mechanisms, a series of experimental and numerical investigations with sand have been carried out by Le et al. (2019), Remspecher et al. (2019), Labenski et al. (2019), Daryaei et al. (2020), and Remspecher (2025).

In small-scale model tests conducted with sand, changes in the void ratio surrounding the pile during impact and vibratory driving were investigated using the digital image correlation (DIC) technique (Remspecher, 2025). Figure 1 compares the void ratio distributions following installation by the two driving methods.

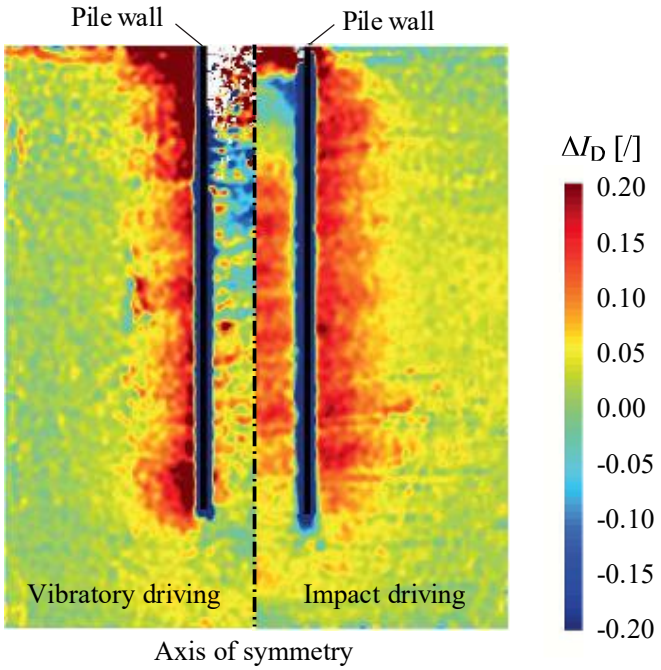


Figure 1: Experimental investigation of vibratory and impact pile driving: Change in relative density in small-scale model tests, results of the digital image correlation (DIC) analysis according to Remspecher (2025)

In both cases in Figure 1, a distinct loosening zone developed immediately adjacent to the pile shaft, where the soil density decreased significantly. Outside this zone, the surrounding soil

became denser as a result of the installation process. For vibratory driving, the increase in soil density was more pronounced but confined to a relatively smaller region around the pile. In contrast, impact driving produced a less intense but more spatially extensive compaction zone.

These observations were also confirmed by numerical investigations. Figure 2 presents the back-analysis of the small-scale model tests using the Multi Material Arbitrary Lagrangian Eulerian (MMALE) formulation to account for large deformations in combination with a hypoplastic constitutive model for the sand (Daryaei et al., 2020). The simulations reproduced the formation of a thin loosening zone immediately adjacent to the pile shaft, surrounded by a compaction zone with increased soil density. The average density within the compaction zone was higher by vibratory driving than by impact driving. Furthermore, the radial extent of the compaction zone remained nearly constant with depth for vibratory driving, whereas for impact driving it was greatest near the ground surface and decreased progressively with increasing depth.

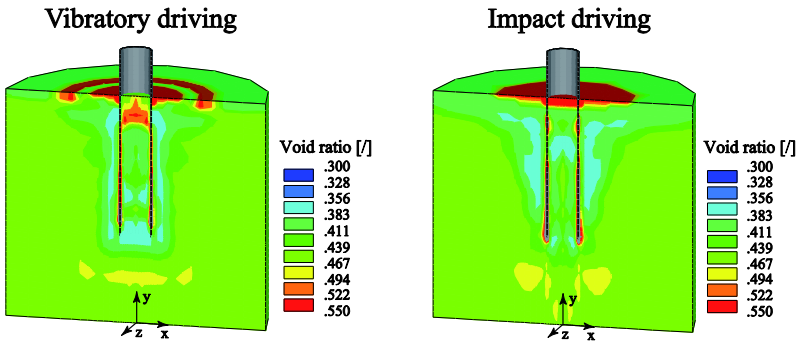


Figure 2: Distribution of the void ratio by vibratory and impact driven pile. Results of the numerical simulation (Daryaei et al., 2020; Le et al., 2021)

Figure 3 presents the distribution of horizontal stress after the pile had reached its final penetration depth. Both installation methods substantially altered the initial stress state and soil density during pile installation. Compared with vibratory driving, impact driving generated considerably higher horizontal stresses beneath the pile tip. This behaviour can be primarily attributed to the well-known plugging effect that develops during impact pile driving, whereby the resistance beneath the pile tip increases (White et al., 2005; Henke & Grabe, 2008).

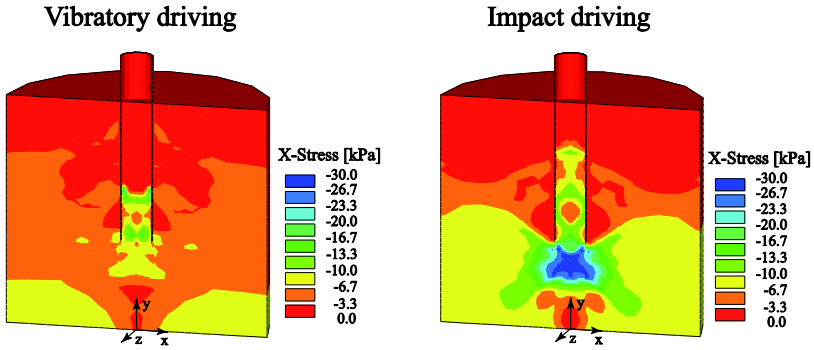


Figure 3: Distribution of the horizontal stress component by vibratory and impact driven pile. Results of the numerical simulation (Daryaei et al., 2020; Le et al., 2021)

2.2 Simplified numerical approach

In most numerical studies, the initial soil conditions are assumed to be homogeneous, with a uniform void ratio and a geostatic stress distribution that increases linearly with depth. The pile

is typically introduced using the wish-in-place method, whereby the pile is included in the numerical model without explicitly simulating the installation process. Consequently, the effects of pile installation on the surrounding soil are neglected. Only a limited number of studies have explicitly simulated pile installation and incorporated the resulting changes in the soil state into the subsequent loading analysis (Staubach et al., 2020; Le et al., 2021).

Owing to the highly complex pile/soil interaction and the large deformations involved during installation, explicit simulation of the penetration process requires advanced numerical formulations and is generally performed as a separate analysis prior to the subsequent loading simulation. To overcome these limitations, a simplified modelling approach is proposed. Based on the experimental and numerical observations presented above, the following initial soil conditions are introduced to represent the installation effects associated with vibratory and impact driving:

- The disturbed zone resulting from the installation process (Figure 4) will be characterised by a loosening zone directly on the pile and a compaction zone around it.
- The thicknesses of these zones remain constant with depth (t_l for the loosening zone and t_c for the compaction zone). Impact driving results in larger t_l and t_c than those of the vibratory driven pile.
- The change in the density in the compaction zone by vibratory driving is more significant than in the impact driving.
- The vertical stress corresponds to the geostatic stress generated by the overburden pressure. Owing to pile installation, the horizontal/vertical stress ratio K^* is assumed to be higher than the at-rest earth pressure coefficient K_0, reflecting the increase in lateral stresses induced by the installation process.

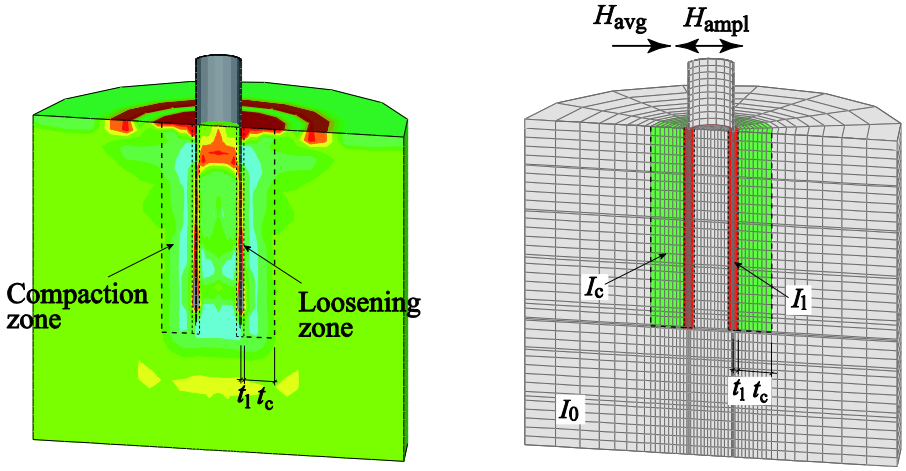


Figure 4: Simplified approach for considering the effects of installation with a compaction and loosening zone around the pile, different relative densities of the zones I_c, I_l, I_0

3 Material models for long-term cyclic behaviour

For simulating cyclic boundary conditions involving a large number of loading cycles, conventional calculations with repeated loading and unloading forces reach their limit. Due to computing efforts and cumulative numerical errors of the integration algorithm, this method is only applicable to a small number of cycles (N < 50). For a large number of cycles (N > 1,000), the

accumulation strain induced by cyclic loading must be calculated explicitly using an accumulation model.

In the explicit calculation concept, conventional constitutive models are employed to simulate the first two loading cycles. In the present study, the hypoplastic model with intergranular strain (Niemunis & Herle, 1997) is adopted for the first loading cycle, while an iso-elastic model is applied during the second cycle. For the subsequent explicit calculation phase, the High-Cycle Accumulation (HCA) model for sand according to Niemunis et al. (2005) is employed to capture the long-term accumulation of deformations under cyclic loading. These constitutive models have been successfully validated and applied to a wide range of boundary value problems involving granular soils (Zachert et al., 2014; Zachert, 2015; Le et al., 2021).

4 Validation of the numerical model

In the experimental investigation illustrated in Figure 1, the piles were subjected to two packages of lateral cyclic loading, each comprising 1,000 load cycles, after installation in sand. In these model tests, the steel model pile had an outer diameter of 200 mm, a wall thickness of 4 mm, and an embedded length of 870 mm. Details of the sand preparation procedure and the experimental test programme are provided by Remspecher (2025) and Le et al. (2021). The cyclic loading programme was identical for both impact-driven and vibratory-driven piles. The results of the cyclic loading were used to validate the numerical model. In these model tests, medium to coarse Berlin sand was used. Table 1 toTable 3 present the parameters of the material models used in the numerical simulations. These parameters are calibrated based on the laboratory tests according to Le (2015).

Table 1: Parameters of Berlin sand for HCA-Model

C_{ampl}	C_{N1}	C_{N2}	C_{N3}	C_p	C_Y	C_e
2.05	0.002	0.022	$8.5 \cdot 10^{-5}$	0.55	1.96	0.325

Table 2: Parameter of Berlin sand for Hypoplasticity with intergranular strain

φ_c [°]	h_s [MPa]	n	e_{d0}	e_{c0}	e_{i0}
31.5	2300	0.30	0.391	0.688	0.791

α	β	m_{R}	m_{T}	R	χ	β_{r}
0.13	1.0	4.4	2.2	10^{-4}	6.0	0.2

Table 3: Parameter of Berlin sand for isotropic elasticity

$C_{1\mathrm{c}}$	C_3	C_4	C_5
725	1.70	0.50	0.40

The numerical simulations were conducted using the FEM program ANSYS 17.0. The material models were implemented in the USERMAT subroutine. The standard SOLID185 3D continuum element with eight nodes was used. Table 5 shows the results of the FEM calculations compared to the model tests. The diagram contains three simulations with different initial conditions for the soil material. In the first simulation, the initial condition was homogeneous. The

pile was generated using the wish-in-place technique. There is no difference between impact and vibratory driving. The accumulated deformation of the pile in the simulation (Figure 5: black curves) differs significantly from the model test results.

In the second simulation, the initial void ratio and stress conditions were applied from the FEM simulations of the penetration process. The impact and vibratory pile driving simulations were performed using a separate model with the software package LS-Dyna (Daryaei et al., 2020). Deformation of the pile by impact and by vibratory driving shows very good agreement with the model tests (Figure 5, blue curves).

The results of the simplified model are presented in Figure 5 as red curves. The modifications of the models for impact and vibratory driving are presented in Table 4. These are based on observations of the model tests and numerical simulations of the penetration processes. It can be seen that the numerical results are in good agreement with the experimental results.

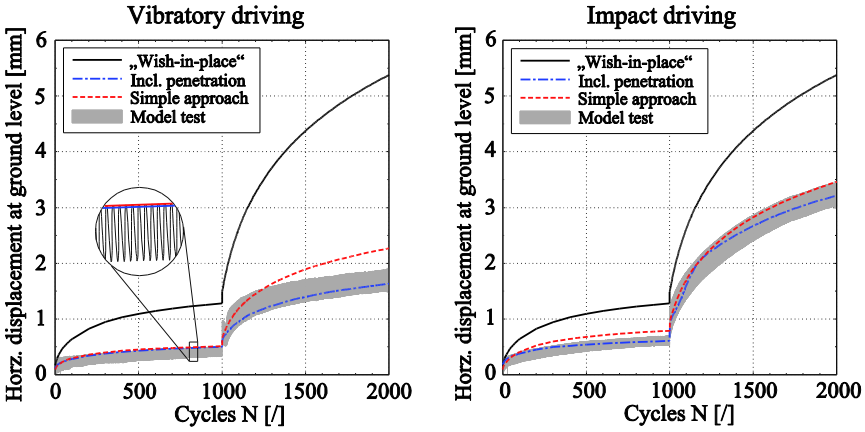


Figure 5: Comparison of the numerical and experimental results

Table 4: Modifications of the simplified models for vibratory and impact driving

	I_0 (%)	I_c (%)	I_l (%)	K_0 (/)	K^* (/)	t_c/t	t_l/t
Vibratory driving	75	100	65	0.463	0.875	40	1.0
Impact driving	75	85	65	0.463	0.875	50	1.5

5 Results for a real scale monopile

The performance of the simplified approach was evaluated using a full-scale monopile from an offshore wind project in the North Sea. The proposed approach was then applied to predict the accumulation of permanent pile displacement induced by lateral cyclic loading. The monopile has an outer diameter of 6.5 m, a wall thickness of 68 mm, and was installed with the vibratory method and has a final embedded length of 29 m. In the simulations, cyclic loading was applied at a point 1.0 m above the ground surface, consisting of a mean bending moment of $M_0 = 12$ MNm with an amplitude of $M_{ampl} = 24$ MNm, and a mean horizontal load of $H_0 = 300$ kN with an amplitude of $H_{ampl} = 600$ kN. The material parameters adopted for the Cuxhaven sand are summarised in Table 5, 6 and 7 according to Zachert (2015). The numerical models for vibratory driving were configured in accordance with the corresponding model tests, as summarised in Table 4.

Table 5: Parameters of Cuxhaven sand for HCA-Model according to Zachert (2015)

C_{ampl}	C_{N1}	C_{N2}	C_{N3}	C_p	C_Y	C_e
1.07	$4.38 \cdot 10^{-4}$	0.103	$9.12 \cdot 10^{-6}$	0.01	2.04	0.37

Table 6: Parameter of Cuxhaven sand for Hypoplasticity with intergranular strain according to Zachert (2015)

φ_c [°]	h_s [MPa]	n	e_{d0}	e_{c0}	e_{i0}
32.6	4030	0.289	0.612	0.948	1.090

α	β	m_R	m_T	R	χ	β_r
0.125	0.40	6.8	3.4	10^{-4}	6.0	0.10

Table 7: Parameter of Cuxhaven sand for isotropic elasticity

C_{1c}	C_3	C_4	C_5
130	4.30	0.60	0.20

Figure 6 presents the initial stress distribution and void ratio adopted in the simplified model. These modified initial conditions are representative of a vibratory-driven pile (Figure 4). The predicted accumulation of permanent pile displacement obtained using the different numerical models is shown in Figure 7. For demonstration purposes, the simulations were performed for up to 10^6 loading cycles.

The results show that the conventional wish-in-place approach predicts generally larger permanent pile displacements than both the simplified approach and the simulation in which the pile installation process was explicitly modelled and the resulting stress state and void ratio were adopted as the initial conditions. The simplified approach shows very good agreement with the predictions obtained from the installation model. This indicates that the proposed method successfully reproduces the effects of pile installation on the initial soil condition while requiring significantly less computational effort.

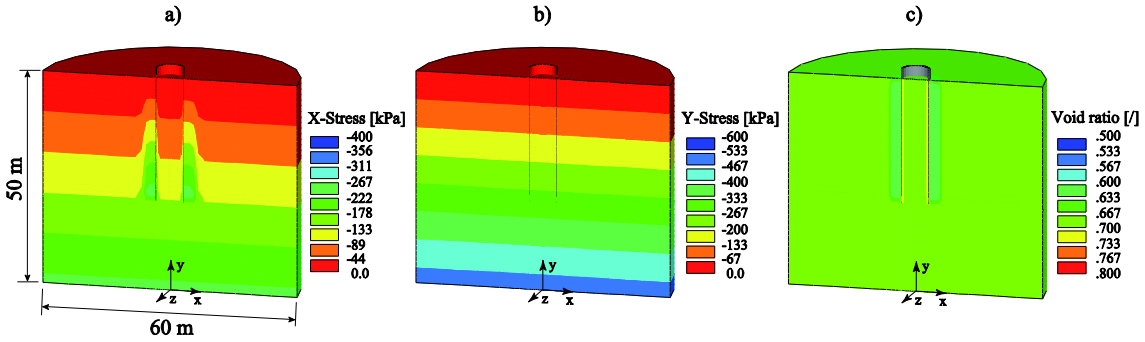


Figure 6: Initial condition for the vibratory driven pile in full-scale – horizontal stress (a), vertical stress (b) and void ratio (c)

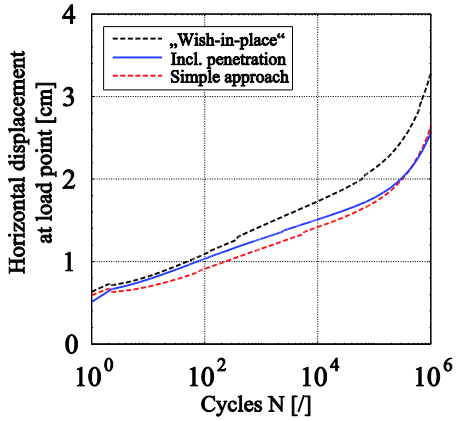


Figure 7: Numerical prediction for vibratory driven full-scale monopile in sand

6 Conclusions and outlook

Based on small-scale model tests and finite element simulations of the installation process, a simplified numerical approach was developed in which the initial soil conditions are modified within different zones surrounding the pile. The proposed approach was successfully validated against the results of small-scale model tests on impact- and vibratory-driven piles subjected to lateral cyclic loading. The post-installation soil conditions for the two installation methods differ in both the relative density and the spatial extent of the affected zones. Following vibratory driving, the compaction zone surrounding the pile exhibits a higher relative density, whereas both the loosening and compaction zones extend over a larger region after impact driving.

The applicability of the proposed approach was further evaluated for a full-scale offshore monopile. The results show that the simplified approach can effectively reproduce the effects of vibratory pile installation on the long-term response under lateral cyclic loading, while significantly reducing the computational effort compared with an explicit simulation of the installation process. Nevertheless, further validation against large-scale or full-scale experimental data is required before the approach can be adopted with confidence for practical engineering applications.

References

[1] Daryaei R., Bakroon M., Aubram D., Rackwitz F. (2020): Numerical evaluation of the soil behavior during pipe-pile installation using impact and vibratory driving in sand. *Soil Dynamics and Earthquake Engineering* 2020;134

[2] De Neef, L., Middendorp, P. (2013): Installation of Monopiles by Vibrohammers for the Riffgat Project. In: Proceedings of the Pfahlsymposium 2013, Institute for Soil Mechanics and Foundation Engineering, Technische Universität Braunschweig

[3] Henke S., Grabe J. (2008): Numerical investigation of soil plugging inside open-ended piles with respect to the installation method. *Acta Geotechnica*, 2008; 3:215-23

[4] Labenski J., Remspecher F., Le V. H., Moormann C., Rackwitz F. (2019): Lateral-bearing behaviour of vibratory driven monopiles in different model test set-ups. *International*

Journal of Physical Modelling in Geotechnics, 2019;20(6):355-364. https://doi.org/10.1680/jphmg.18.00090

[5] Le V.H. (2015): Zum Verhalten von Sand unter zyklischer Beanspruchung mit Polarisationswechsel im Einfachscherversuch. PhD Thesis. Publication of Soil Mechanics and Geotechnical Division, Technische Universität Berlin, Shaker Aachen

[6] Le V. H., Remspecher F., Rackwitz F. (2019): Numerical investigation of installation effects on the cyclic behaviour of monopile foundation under horizontal loading. In: Randolph M., Doan D., Tang A., Bui M., Dinh V. (eds) Proceedings of the 1st Vietnam Symposium on Advances in Offshore Engineering. VSOE 2018. Lecture Notes in Civil Engineering, vol 18. Springer, Singapore, pp. 389-394. https://doi.org/10.1007/978-981-13-2306-5_54

[7] Le V. H., Remspecher F., Rackwitz F. (2021): Development of numerical models for the long-term behaviour of monopile foundations under cyclic loading considering the installation effects. *Soil Dynamics and Earthquake Engineering* 150 (2021) 106927. https://doi.org/10.1016/j.soildyn.2021.106927

[8] Niemunis, A., Herle, I. (1997): Hypoplastic model for cohesionless soils with elastic strain range. *Mechanics of Cohesive-Frictional Materials*, Vol. 2, pp. 279-299

[9] Niemunis A., Wichtmann T., Trianstafyllidis T. (2005): A high-cycle accumulation model for sand. *Computers and Geotechnics*, 2005;32(4):245-63

[10] Remspecher F., Le V. H., Rackwitz F., Herwig V., Matlock B. (2019): Vibratory driven installation of monopiles - an experimental investigation of the soil-pile interaction. In: Randolph M., Doan D., Tang A., Bui M., Dinh V. (eds) Proceedings of the 1st Vietnam Symposium on Advances in Offshore Engineering. VSOE 2018. Lecture Notes in Civil Engineering, vol 18. Springer, Singapore, pp. 171-176. https://doi.org/10.1007/978-981-13-2306-5_22

[11] Remspecher F. (2025): On the effects of pile installation in granular soil. PhD Thesis. Technische Universität Berlin. https://doi.org/10.14279/depositonce-23356

[12] Staubach P., Machacek J., Moscoso M. C., Wichtmann T. (2020): Impact of the installation on the long-term cyclic behaviour of piles in sand: a numerical study. *Soil Dynamics and Earthquake Engineering*, Vol. 138, November 2020, 106223 https://doi.org/10.1016/j.soildyn.2020.106223

[13] White D., Schneider J.A., Lehane B.M. (2005): The influence of effective area ratio on shaft friction of displacement piles in sand. In: First international Symposium on Frontiers in offshore construction.

[14] Zachert H., Wichtmann T., Kudella P., Triantafyllidis T., Hartwig U. (2014): Validation of a high cycle accumulation model via FE-simulations of a full-scale test on a gravity base foundation for offshore wind turbines. In: International wind engineering conference - IWEC

[15] Zachert H. (2015): Zur Gebrauchstauglichkeit von Gründungen für Offshore Windenergieanlagen. PhD Thesis, Pubplication of the Institute for Soil and Rock Mechanics, Karlsruher Institute for Technology (KIT). Triantafyllidis T. (edit.), No. 180

Author

Dr.-Ing. Viet Hung Le
Technische Universität Berlin
Chair of Soil Mechanics and Geotechnical Engineering
Gustav-Meyer-Allee 25,
13355 Berlin
Tel.: +49 (0) 30 – 314 72064
e-mail: v.le@tu-berlin.de
Web: www.tu.berlin/grundbau

Univ.-Prof. Dr.-Ing. Frank Rackwitz
Technische Universität Berlin
Chair of Soil Mechanics and Geotechnical Engineering
Gustav-Meyer-Allee 25,
13355 Berlin
Tel.: +49 (0) 30 – 314 72341
e-mail: frank.rackwitz@tu-berlin.de
Web: www.tu.berlin/grundbau

Data-driven back-analysis and forward prediction of quay wall behaviour

Kacper Cerek[1], **Jürgen Grabe**[1]
[1]Institute of Geotechnical Engineering and Construction Management
Hamburg University of Technology, Hamburg, Germany

The long-term structural health of port infrastructure faces complex degradation from tidal loading, structural corrosion, and seabed scour. Standard monitoring techniques are often discontinuous or restricted by adverse environmental conditions. This paper proposes a data-driven framework leveraging a hybrid neural network architecture to perform real-time back-analysis and long-term state prediction of quay walls. Combining a Long Short-Term Memory (LSTM) network for capturing spatial deflection profiles and a Feedforward Neural Network (FNN) for cyclic tidal variations, the surrogate model substitutes computationally expensive high-cycle accumulation (HCA) finite element simulations. Training data is generated via a stochastic Latin Hypercube Sampling framework applied to a validated numerical model of a representative container terminal quay wall. The proposed methodology offers an efficient, non-invasive digital twin approach capable of decoding hidden structural state variables from easily accessible wall displacement scanning data.

1 Introduction

The long-term structural stability of quay walls in tidal ports is crucial for the operational efficiency of maritime infrastructure. These structures operate under severe and simultaneous degradation mechanisms, including tide-dependent deformation accumulation, corrosive cross-sectional loss, and seabed scouring. While cyclic loading induces cumulative strains in the surrounding subsoil, scouring and corrosion directly weaken the static system.

A primary engineering challenge is that critical state variables, such as the instantaneous scour depth and localized sub-surface corrosion rates, are exceptionally difficult to measure directly below the waterline. Classical structural health monitoring (SHM) systems are often cost-prohibitive, discontinuous, or highly error-prone due to environmental factors like siltation and water turbidity. Conversely, high-fidelity numerical approaches, such as the Finite Element Method (FEM) integrated with advanced constitutive relations like the High-Cycle Accumulation (HCA) model, require prohibitive computational time, which prevents real-time diagnostic evaluation.

To bridge this gap, this paper introduces a data-driven surrogate modeling framework. By designing a hybrid neural network that merges spatial deep learning with static parametric inputs, a high-performance digital twin is established. This network interprets readily available structural deflection profiles to back-calculate latent parameters (scour depth and structural degradation) in real time, ultimately empowering predictive, pro-active maintenance strategies.

2 State of the Art and Theoretical Background

2.1 Traditional Quay Wall Monitoring Techniques

Current observational methods rely on distinct physical principles, each with inherent technical limitations:

- **Deformation Measurements:** Horizontal displacements are conventionally recorded using embedded inclinometers or fiber-optic systems. Modern terrestrial scanning can map top-of-wall profiles, but sub-surface monitoring remains constrained by water turbidity and sediment coverage.
- **Scour Monitoring:** Acoustic methods (e.g., multibeam echosounders) struggle to differentiate fluid mud layers from the true load-bearing seabed. Numerical methods to determine scour depth exist but require complex sensor networks like pore-water pressure transducers. Mechanical tracking devices provide maximum scour depth but lack spatial resolution.
- **Corrosion Assessment:** Manual ultrasonic testing conducted by divers remains dominant but is highly time-consuming, hazardous, and weather-dependent. Automated inspection via Remote Operated Vehicles (ROVs) equipped with ultrasonic sensors is an emerging field for mapping wall-thickness losses.

2.2 Surrogate Modeling in Geotechnical Engineering

Geotechnical engineering decisions often require complex structural simulations. Advanced constitutive models (e.g., hypoplasticity, HCA) capture realistic soil behaviors but scale poorly when applied to multi-scenario planning or real-time evaluation. Surrogate models act as mathematical approximations that mimic high-fidelity numerical outputs at a fraction of the computational expense. Artificial Neural Networks (ANNs) have demonstrated exceptional capabilities in exploring dense design spaces and performing real-time inverse analyses. While conventional approaches fail to seamlessly integrate combined non-linearities (soil plasticity, high-cycle degradation, and geometric changes), deep-learning-based surrogates offer a robust alternative for continuous, real-time SHM.

3 Numerical Modeling and Data Generation

3.1 Case Study: Container Terminal

A representative container terminal quay wall situated within a large tidal port environment serves as the reference case study. The structure consists of a combined sheet pile wall utilizing PSp 1000 king piles and PZi 610 intermediate sheet piles. The wall was installed landside within a slurry-stabilized trench via a single-phase construction method. The reinforced concrete superstructure is deep-founded on three rows of vertical and inclined displacement bored piles. Horizontal tied-back support is provided via high-strength steel round anchors connected to an opposing landward structure (Figure 1).

The subsoil profile reflects a typical estuarine glacial geology, characterized by multiple layers of sand filling, underlain by a low-permeability, stiff clay layer, which acts as a hydraulic boundary.

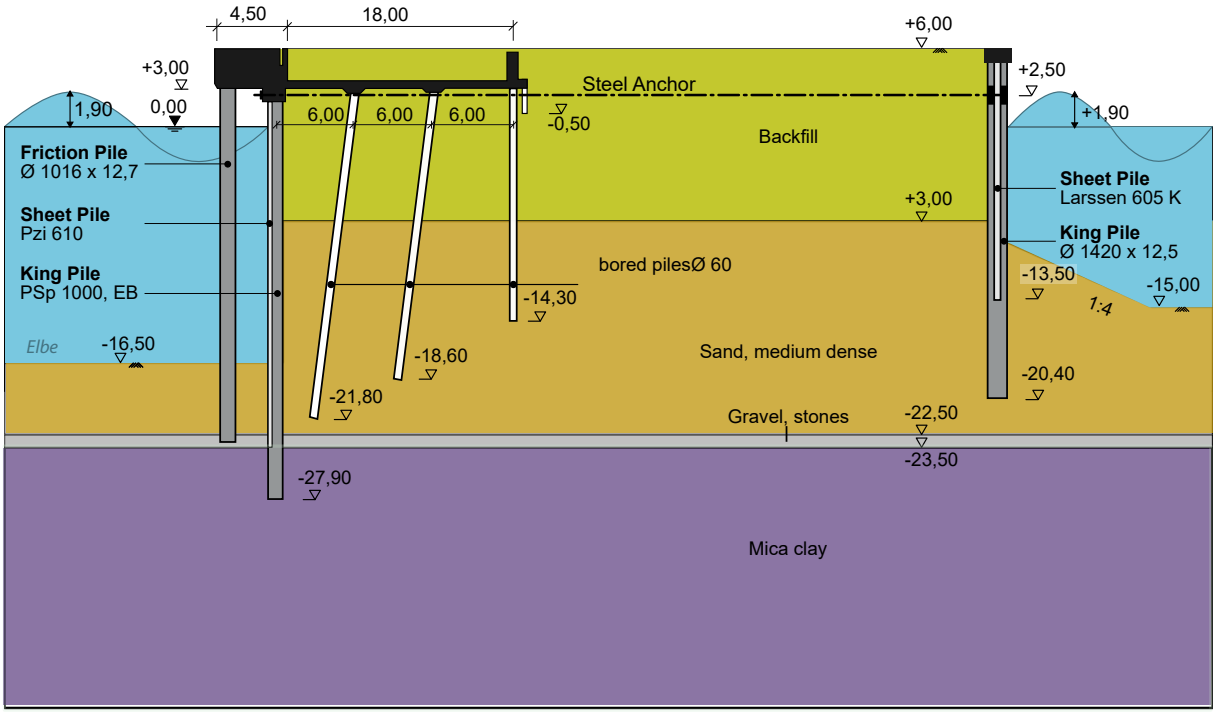


Figure 1: Quay wall cross section (Tetla et al. 2026).

3.2 Advanced Constitutive Modeling and Optimization

To handle high-cycle tidal actions, the High-Cycle Accumulation (HCA) model is utilized for the sand layers. This model uses an explicit calculation strategy to simulate the effects of thousands of tidal cycles efficiently without tracking individual load oscillations. To calibrate the model

parameters in the absence of comprehensive laboratory tests, an inverse identification procedure was implemented by coupling PLAXIS 2D with a Particle Swarm Optimization (PSO) algorithm (Grabe et al. 2007). The PSO algorithm iteratively minimizes the residual error between historical inclinometer measurements and the numerical predictions, yielding a highly validated baseline model (Tetla et al. 2026).

3.3 Finite Element Mesh and Boundary Conditions

The two-dimensional finite element model is developed in PLAXIS 2D. Structural elements are discretized using specialized components: the combined sheet pile wall is represented as a plate element, whereas the bored piles and friction piles are modeled as Embedded Beam Elements. The soil is modeled using 6-node triangular elements under plane strain conditions, with a heavily refined mesh within the structure-soil interaction zone.

3.4 Stochastic Framework for Dataset Generation

The simulation sequence is divided into a monotonic construction phase (Phases 0–8) and a cyclic accumulation phase (Phases 9–21), as outlined in Table 1.

Table 1: Finite Element Model Calculation Phases

Phase	Description
0	Initial equilibrium under at-rest earth pressure
1	Installation of the combined sheet pile wall
2	Installation of bored piles and horizontal anchor
3–5	Sequential excavation
6	Friction pile installation and superstructure execution
7	Landward backfilling
8	Nil-phase (relaxation)
9–10	Initial tidal cycles
11	Accumulation reference cycle
12–21	Accumulation intervals (10-year steps) with coupled corrosion and scour

To populate the surrogate training matrix, a stochastic Latin Hypercube Sampling (LHS) routine was executed across 100 simulations. The annual scour rates and waterside corrosion rates were varied within literature-defined bounds over a 100-year lifecycle split into 10-year increments. At each increment, the sheet pile's bending stiffness (EI) and axial stiffness (EA) were reduced, while the harbour bed was systematically lowered according to the sampled scour rate to represent the developing scour depth (t_s). Figure 2 shows results obtained for selected samples with various scour and corrosion rates for a period of 100 years.

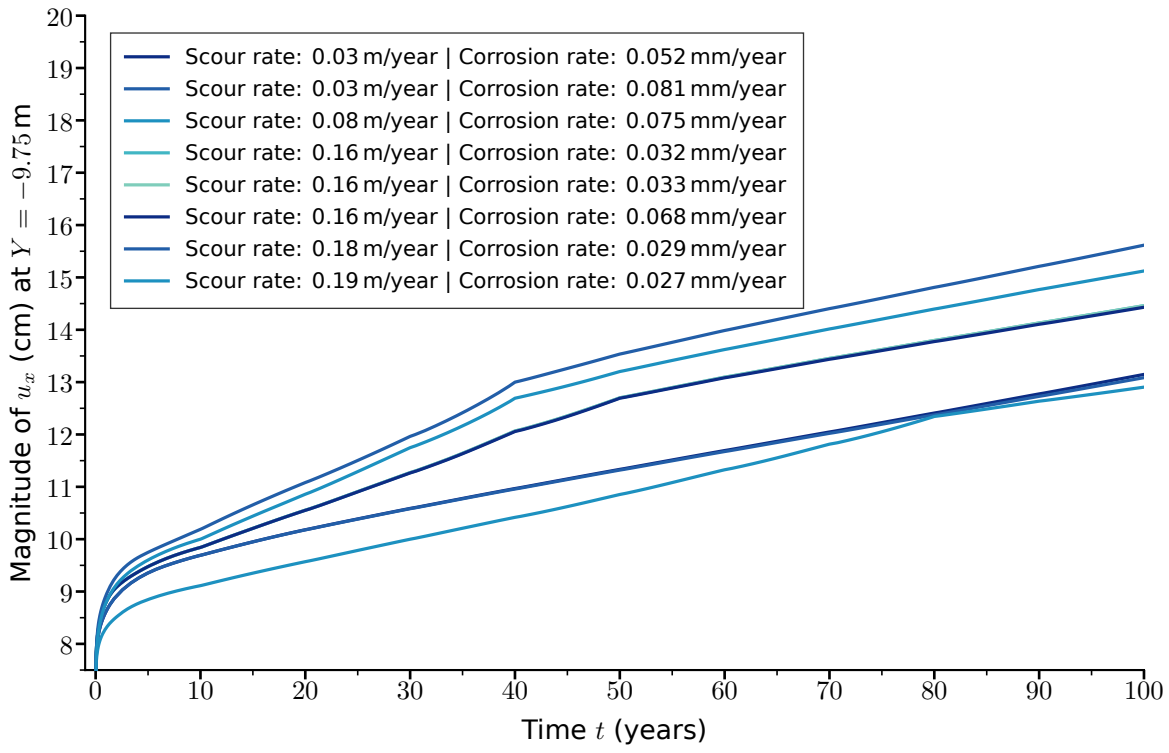


Figure 2: Development of the horizontal displacement magnitude $|u_x|$ at elevation $Y = -9.75$ m over 100 years for selected combinations of annual scour and corrosion rates.

4 Hybrid Surrogate Model Architecture

The core philosophy of the proposed surrogate model relies on the geotechnical principle that the horizontal deflection profile (u_x) of a sheet pile wall behaves as an integrated systemic response (Cerek et al. 2026a; Cerek et al. 2026b). The structural bending line (its curvature, inflection points, and displacement magnitudes) uniquely encodes latent state parameters, including the structural stiffness (EI), the depth of seabed embedment (t_s), and the cumulative loading history.

4.1 Network Layout: Dual-Branch Architecture

Because the input vector contains both a spatial sequence (the wall profile) and static scalar markers (the number of tidal cycles), a multi-input dual-branch network architecture was engineered:

- **Recurrent Branch (Spatial Deep Learning):** The wall deflection curve $u_x(y)$, discretized uniformly from the top of the wall down to the baseline mudline, is processed as a spatial

sequence using a Long Short-Term Memory (LSTM) network. LSTMs successfully avoid vanishing gradient problems, allowing the model to extract complex curvature features indicative of scour-induced loss of embedment or corrosion-induced bending.

- **Static Branch (Temporal Parameterization):** A Feedforward Neural Network (FNN) branch takes the number of tidal cycles as a scalar input to map the time-dependent, non-linear deformation accumulation.

The features extracted from both branches are concatenated and passed through post-fusion dense layers to output concurrent real-time predictions for the targets: scour depth (t_s) and remaining bending stiffness (EI_{corr}).

4.2 Data Preprocessing and Training Protocol

The generated dataset was divided into 60% for training, 20% for validation, and 20% for testing. Static inputs and target variables were normalized to the range $[0, 1]$. To accelerate gradient descent stability and ensure optimal convergence, the sequential displacement vectors were standardized using a standard Z-score transformation.

The comprehensive hyperparameters and activation configurations are detailed in Table 2.

Table 2: Hybrid Surrogate Model Architecture

Branch	Layer Type	Units / Neurons	Activation
Recurrent	LSTM 1	448	Sigmoid
	LSTM 2	256	Softplus
Static	FNN 1	480	Sinus
Fusion	Concatenation	256 + 480 = 736	–
Post-Fusion	FNN 2	192	ELU
	FNN 3	320	GELU
Output	FNN 4	2 (t_s, EI)	Linear

The model training was driven by the Adam optimizer with a fine-tuned learning rate of 3.7×10^{-5} across a maximum of 50,000 epochs. To mitigate overfitting, an early stopping mechanism monitored the validation Mean Squared Error (MSE), terminating execution if no improvement greater than 10^{-4} occurred over 200 consecutive epochs.

5 Results

To evaluate the predictive accuracy of the trained hybrid surrogate model, the inverse network was tested against the test dataset. The model's capacity to concurrently back-calculate the remaining structural bending stiffness EI_{corr} and the localized scour depth t_s solely from horizontal

displacement profile sequences u_x was quantitatively assessed using parity plots and coefficients of determination R^2.

Figure 3 illustrates the comparison between the actual bending stiffness values generated by the finite element model and the corresponding values predicted by the surrogate model. Across the range of cross-sectional degradation, the predicted outputs align closely with the line of perfect equality. The model achieves a coefficient of determination of $R^2_{95\%} = 0.9742$, demonstrating high precision in capturing structural stiffness reductions caused by progressive waterside corrosion.

The performance of the surrogate model in predicting scour depths (t_s) up to 6 m is presented in Figure 4. The model yields a coefficient of determination of $R^2_{95\%} = 0.9960$. Minor variances are observed primarily at larger scour depths (≥ 4 m), where complex non-linear redistribution of earth pressures and high displacement gradients increase sensitivity. Nevertheless, the tight clustering around the ideal prediction line confirms that spatial deflection curvature serves as an indicator for identifying loss of structural embedment below the mudline.

Overall, the high predictive accuracy confirms that the dual-branch hybrid network LSTM-FNN successfully disentangles the coupled effects of geometric scour degradation and material loss, rendering it highly effective for real-time digital twin structural health monitoring.

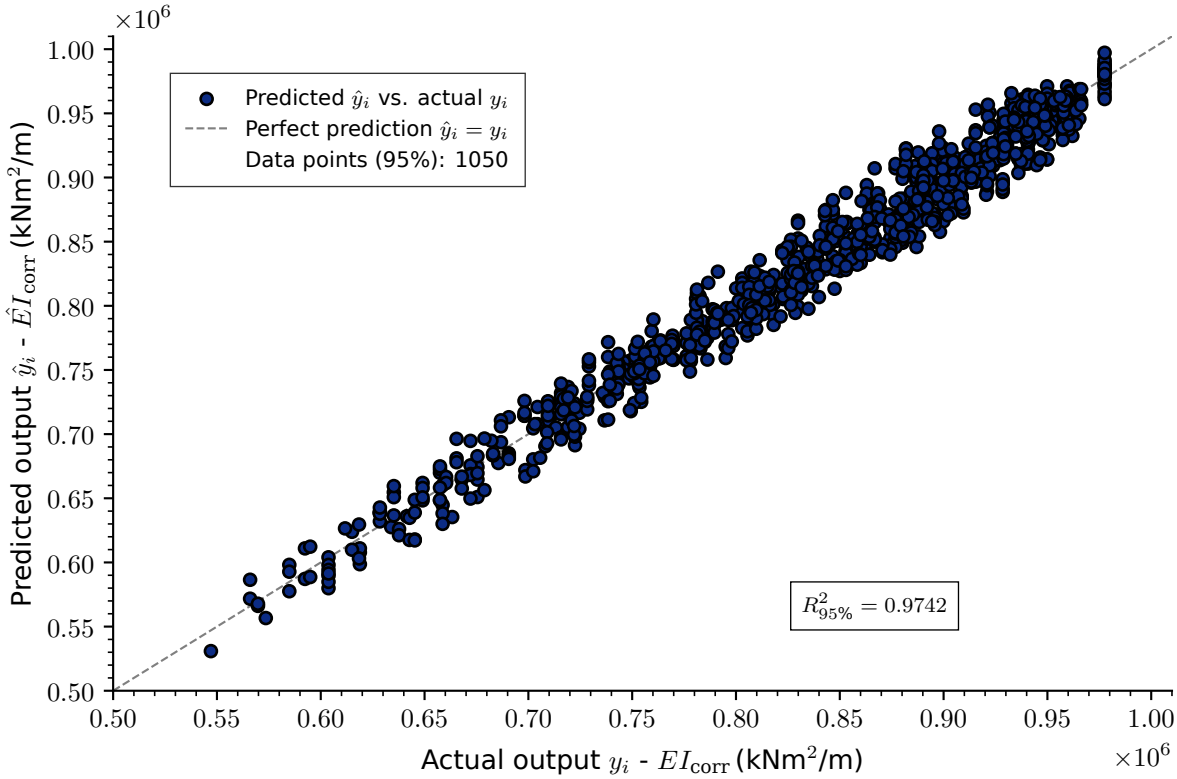


Figure 3: Comparison of predicted vs. actual bending stiffness after corrosion at the 95% quantile, including the coefficient of determination R^2.

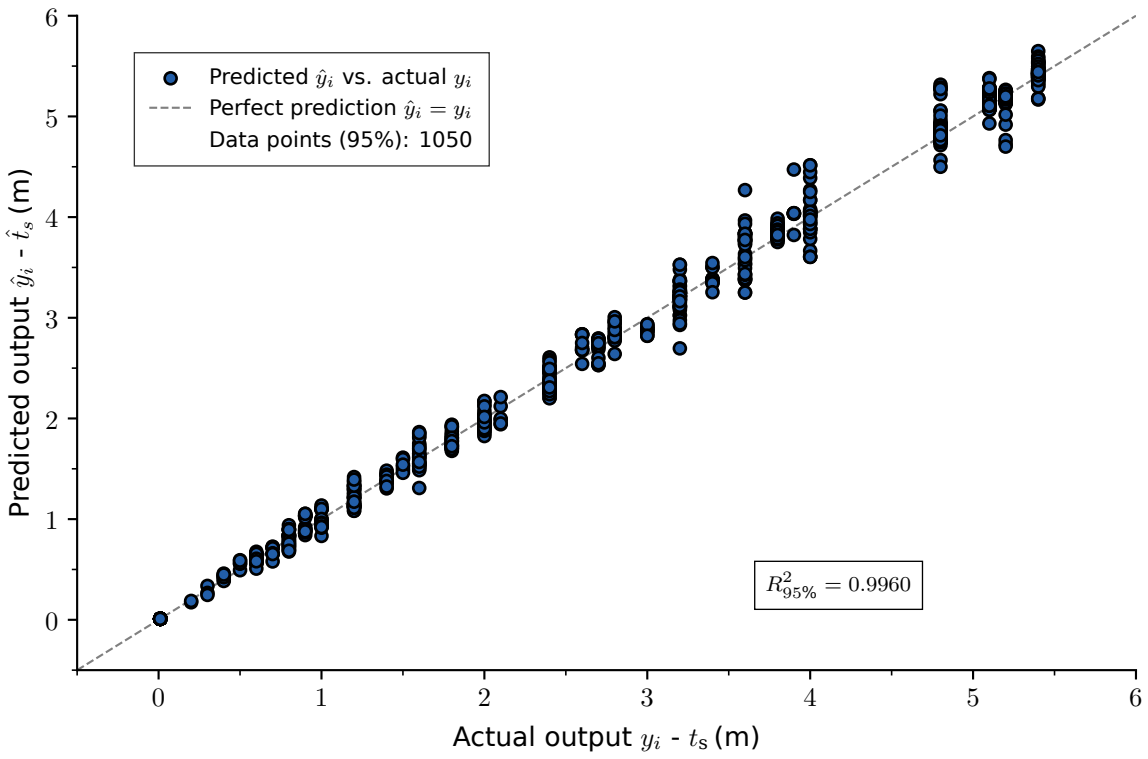


Figure 4: Comparison of predicted vs. actual scour depth at the 95% quantile, including the coefficient of determination R^2.

6 Conclusions

This paper presented a data-driven framework using a hybrid LSTM-FNN neural network for back-analyzing and forecasting the structural condition of cyclically loaded quay walls, leveraging a synthetic numerically generated dataset. The following core conclusions are drawn:

- Hybrid surrogate modeling acts as a powerful, computationally efficient tool capable of replacing intensive, non-linear finite element routines for long-term health monitoring.
- By interpreting the spatial deflection profile as an integrated system response, the inverse neural network successfully decouples and decodes underground and underwater structural degradation features (scour and corrosion) that are otherwise unmeasurable in real-time.
- The combination of spatial recurrent layers (LSTM) and static parametric layers (FNN) provides a flexible framework that can easily adapt to other complex soil-structure interaction scenarios.

Literatur

[1] Cerek, Kacper, Elnaz Hadjiloo und Jürgen Grabe (2026a). “Hybrid Surrogate Modeling of a Quay Wall: An Automatically Tuned Framework Integrating Long Short-Term Memory and Feedforward Neural Networks”. In: *Geo-Congress 2026*. ASCE, S. 332–341. DOI: 10.1061/9780784486726.032. eprint: https://ascelibrary.org/doi/pdf/10.1061/9780784486726.032.

[2] Cerek, Kacper, Dagmara Klos, Elnaz Hadjiloo und Jürgen Grabe (März 2026b). “Hybrid surrogate models of quay walls”. In: *Proceedings of the Institution of Civil Engineers - Civil Engineering*. DOI: 10.1680/jcien.25.00451.

[3] Grabe, J., S. Henke und B. Schümann (2007). “Numerische Simulation von Rammarbeiten im Erdwiderstandsbereich von Baugrubenwänden”. In: *Bautechnik* 84.8, S. 519–524.

[4] Tetla, Daniel, Duy Anh Dao, Kacper Cerek und Jürgen Grabe (2026). “Numerische Untersuchung hochzyklisch belasteter Kaimauern”. In: *geotechnik* n/a.n/a. DOI: https://doi.org/10.1002/gete.70043.

Author

Kacper Cerek, M. Sc.
Technische Universität Hamburg
Institut für Geotechnik und Baubetrieb
Schellerdamm 22-24
21079 Hamburg
Tel.: +49(0)40/42878-3912
e-mail: kacper.cerek@tuhh.de
Web: www.tu-harburg.de/gbt

Univ.-Prof. Dr.-Ing. Jürgen Grabe
Technische Universität Hamburg
Institut für Geotechnik und Baubetrieb
Schellerdamm 22-24
21079 Hamburg
Tel.: +49(0)40/42878-3762
Fax: +49(0)40/42878-3721
e-mail: grabe@tuhh.de
Web: www.tu-harburg.de/gbt

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.

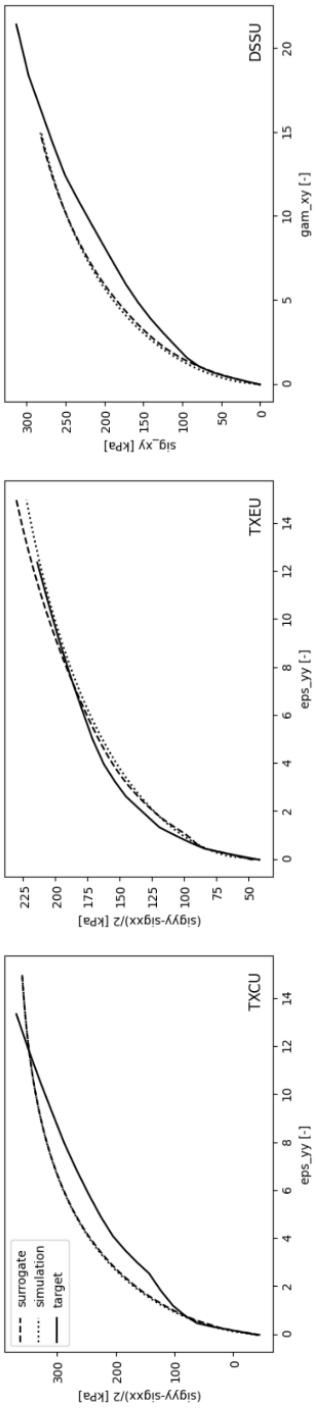


Figure 1. Comparison of target curves for TXC, TXE and DSS with stress-strain curves obtained from the surrogate model and numerical lab test simulations.

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.

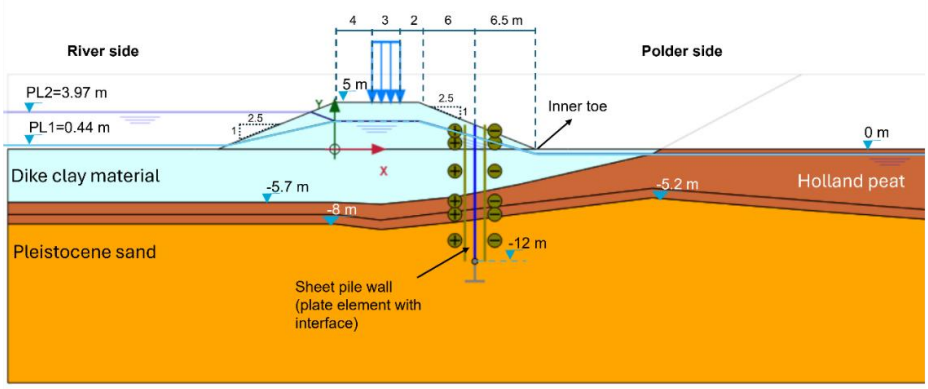


Figure 2. Schematization of the Bergambacht case study for probabilistic analysis

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'

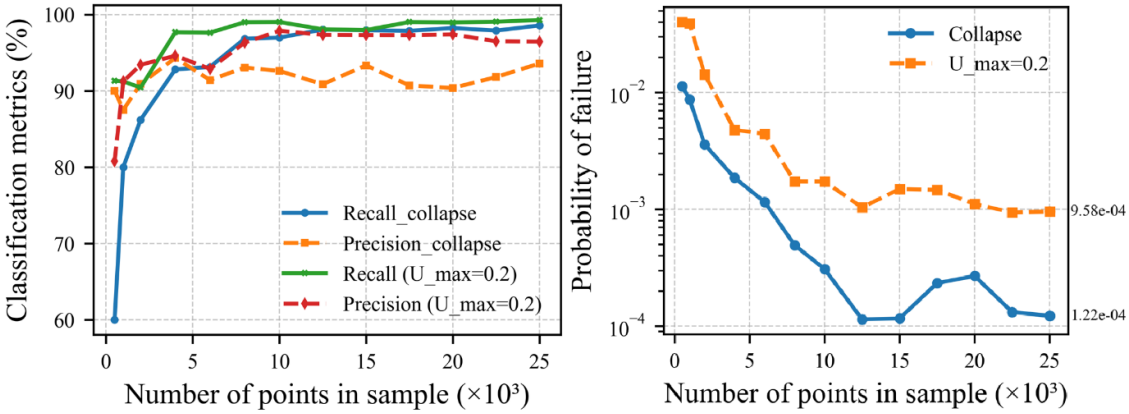


Figure 3. Accuracy and convergence of the solution
a) Classification metrics b) Probability of failure (log scale)

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.

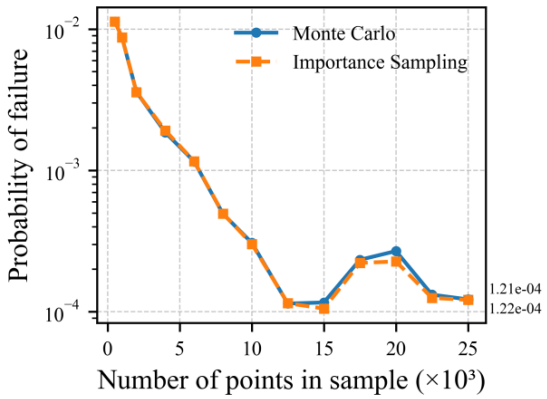


Figure 4. Comparison of P_f (soil collapse) from MCS (10^7 samples) and IS ($5 \cdot 10^5$ samples) using surrogates trained by a varying number of data points

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

Appendix A: Screening and input influence:

The total number of parameters in the finite model for the three soil layers including the three used soil models exceeds 60. Some of these parameters are specified as deterministic, as their variance has a negligible effect on the model outcome (e.g., ν, m, ψ', and c'). Other parameters are dependent on the inputs specified in the previous section. This reduces the number of independent input variables to 23, which is still relatively high for probabilistic design. To make the analysis computationally feasible, a screening study is performed to identify the parameters with the greatest influence on the output of interest (displacements and structural forces).

The Elementary Effects (EE) method, commonly known as the Morris method, is a global sensitivity analysis technique used for screening input factors in complex models to identify the most influential parameters with minimal computational cost. In this method, the elementary effect of an input factor x_i on the model output Y is defined as:

$$EE_i = \frac{f(x_1, \dots, x_i + \Delta, \dots, x_k) - f(x_1, \dots, x_i, \dots, x_k)}{\Delta} \qquad \text{Equation 1}$$

where k is the number of input factors and Δ is a predetermined step size in the normalized input space [0-1] (Morris, 1991). A set of randomly sampled trajectories through the input space is generated, and for each factor, the mean (μ) and standard deviation (σ) of the elementary effects are computed

$$\mu_k^* = \frac{1}{r}\sum_{i=1}^{r} |EE_k^r|, \qquad \sigma_k = \sqrt{\frac{1}{r}\sum_{i=1}^{r}(EE_k^r - \mu_k)^2} \qquad \text{Equation 2}$$

where r is the number of trajectories. The mean μ_i measures the overall influence of factor i, while the standard deviation σ_i reflects non-linear effects or interactions with other factors. Factors with high μ_i and/or σ_i are deemed important, whereas those with low values can be fixed, thereby reducing model complexity (Figure A1). This method is particularly useful in geotechnical modeling, where computationally expensive simulations preclude full variance-based sensitivity analyses (Campolongo et al., 2007; Saltelli et al., 2008).

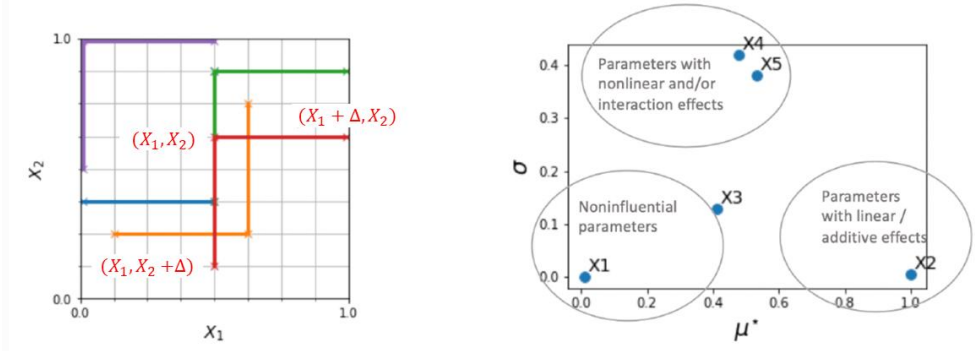


Figure A1 Screening using elementary effect method (Morris method)

Based on the calculated μ_k^* and σ_k shown in Figure A2 we found that the SHANSEP model parameter $\boldsymbol{\alpha}$ and the NGI-ADP model parameter Su_{DSS}/S_{uA} for the peat and clay soils have the highest influence on the outcome as expected since they represent the shear strength of the soil in this model. On the other hand, 8 parameters were found to have marginal influence on the change in displacement and structural forces where μ_k^* and σ were below 5% and therefore can be excluded from the probabilistic design. The excluded parameters are(POP_{peat}, ϕ'_{peat}, $\gamma_{fc\,\text{Dikematerial}}$, $G/Su_{\text{Dikematerial}}$, m_{Peat}, $\gamma_{sat\,Peat}$, $\gamma_{fc_{Peat}}$, OCR_{Sand}).

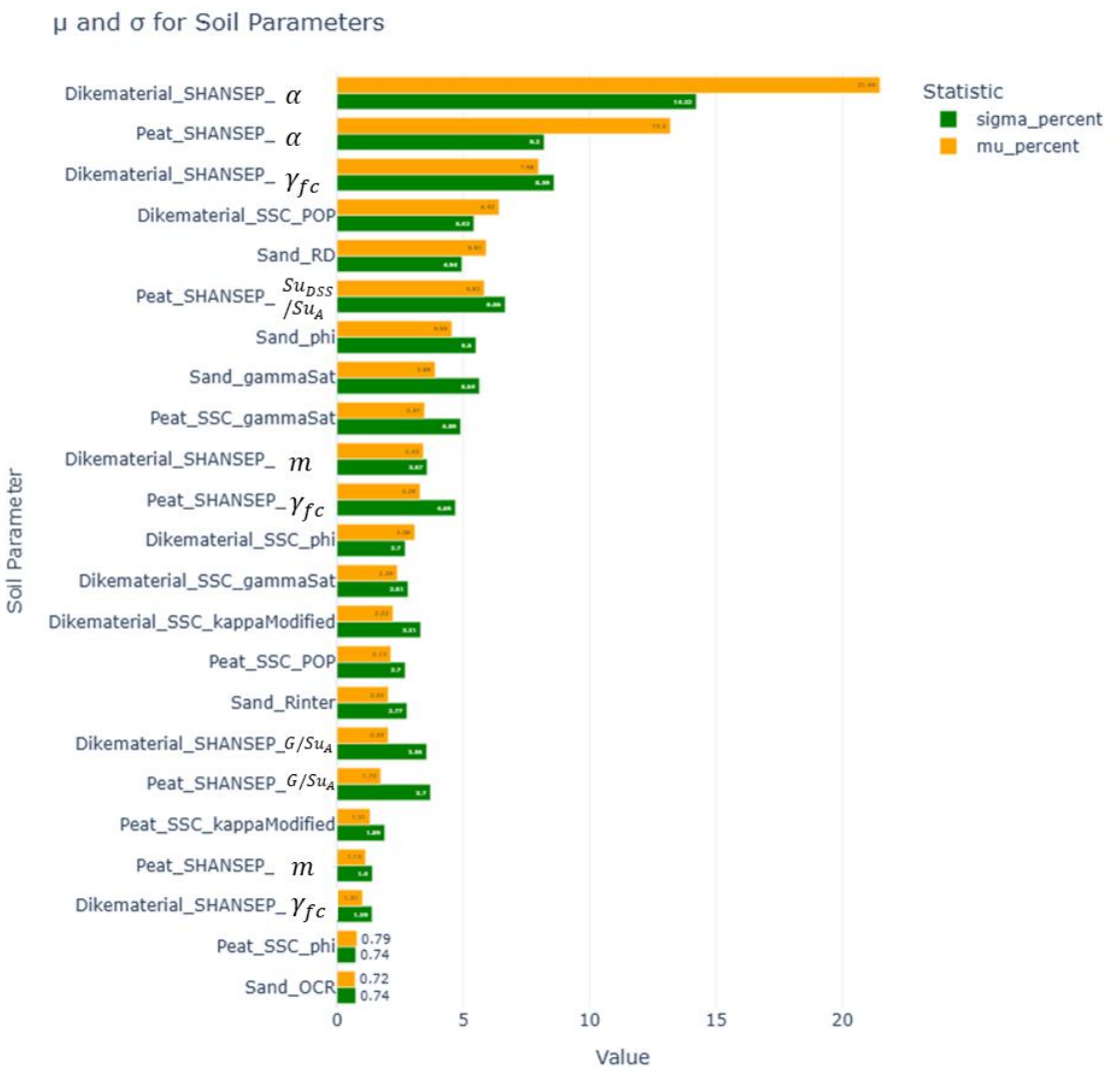


Figure A2 Screening results (μ_k^* and σ_k) for each of the parameters measured for the effect on the deformations in the inner crest

Numerical study on the influence of mesh dependency on shear planes in Finite Element Method

Mazdak Atri, Tim Pucker

Abstract: This paper investigates mesh dependency in geotechnical finite element simulations using Mohr–Coulomb plasticity. Two calibrated model sets are studied: a confined axisymmetric glass bead model with hardening and softening, and a confined plane strain sand model with linear hardening. In both cases, the material parameters are kept unchanged while the mesh size is varied. The results show that the elastic or prelocalization response is only weakly affected by the mesh size, whereas the plastic response and localization indicators are mesh dependent. The predicted equivalent plastic strain patterns become increasingly localized with mesh refinement. The results highlight the limitations of classical local continuum models and motivate regularized approaches with an internal length scale.

1 Introduction

Shear plane formation is an important aspect of the mechanical response of granular materials and geomaterials. Under loading, deformation may localize into narrow zones that influence the observed failure mechanism, mobilized strength, and global stress strain response. Experimental studies have shown that shear band initiation, thickness, and inclination depend on material state, density, stress conditions, anisotropy, and testing method (Tatsuoka et al., 1990; Han and Drescher, 1993; Alshibli and Sture, 1999; Alshibli et al., 2003).

In the finite element solutions of the classical Cauchy continuum, strain localization may lead to mesh dependent results. Once deformation starts to concentrate into narrow zones, the governing boundary value problem may lose its elliptic character, and the predicted width, intensity, and orientation of the localized deformation pattern may become affected by the discretization. This can influence not only the local deformation field, but also the global stress strain response. The relation between strain localization, loss of ellipticity, numerical instability, and mesh dependency has been discussed in previous studies (Ehlers et al., 2001; Stefanou and Gerolymatou, 2019).

The aim of this paper is not to introduce mesh dependency as a new phenomenon, but to demonstrate how it appears in selected geotechnical finite element simulations based on a Cauchy continuum. The Mohr–Coulomb model is selected deliberately because the focus of this study is on the influence of mesh size on the global response and localization pattern, rather

than on developing the most realistic constitutive description. For this reason, more sophisticated constitutive frameworks, such as critical state elastoplasticity or hypoplasticity, are outside the scope of this study. Two cases are investigated: a confined axisymmetric glass bead model with cohesion hardening and softening, and a confined plane strain sand model with linear cohesion hardening. The central contribution of this study is to demonstrate that successful calibration of the constitutive response under homogeneous loading does not guarantee mesh objective predictions at specimen scale once strain localization develops. The results are used to highlight the limitations of Cauchy continua in strain localization problems and to motivate regularized models with an internal length scale, such as micropolar or Cosserat continua (Liu et al., 2022; Mühlhaus and Vardoulakis, 1987; De Borst, 1991; Alshibli et al., 2006).

2 Confined axisymmetric glass bead model

The first model set represents the confined axisymmetric compression test on very dense glass beads reported by Alshibli and Cil (2017). The model follows the experimental geometry and boundary conditions, with a radius of 35.55 mm and a height of 71.10 mm. The global response is evaluated using the principal stress ratio σ_1/σ_3 and axial strain ε_a.

The specimen is discretized using eight node biquadratic axisymmetric quadrilateral elements with reduced integration (CAX8R). A Mohr–Coulomb model with cohesion hardening and softening is used. The critical state friction angle is assigned as the friction angle, while the peak and post peak response is represented through the cohesion hardening softening law. The material parameters and hardening softening input are given in Tables 1 and 2, respectively.

Table 1: Parameters of the glass bead model based on the work of Alshibli and Cil (2017)

Parameter	Symbol	Value
Young's modulus	E	61.1674 MPa
Poisson's ratio	ν	0.25
Critical state friction angle	φ	$29.6°$
Assigned dilation angle	ψ	$0.2°$
Contact friction coefficient	μ	0.5
Confining pressure	σ_3	100 kPa
Target axial strain	ε_a	20%

The analysis consists of initial confinement followed by axial compression. During the mesh study, only the discretization is varied; the geometry, constitutive parameters, boundary conditions, contact definition, and loading procedure remain unchanged. The four mesh configurations are summarized in Table 3.

A one element axisymmetric material point simulation using the same constitutive parameters is additionally performed as a homogeneous calibration check. It is not considered a mesh refinement case because strain localization cannot develop in the single element model.

Table 2: Cohesion hardening softening input used for the confined axisymmetric glass bead model

Equivalent plastic strain, $\bar{\varepsilon}^{pl}$ [-]	Cohesion, c [kPa]	Equivalent plastic strain, $\bar{\varepsilon}^{pl}$ [-]	Cohesion, c [kPa]
0.000000	0.5000	0.829428	12.7777
0.041248	11.5787	1.035120	10.3798
0.067943	23.4751	1.283320	8.1448
0.093215	27.4561	1.431150	7.0128
0.117312	30.0227	1.602070	5.8663
0.140853	31.6291	1.805800	4.7430
0.164439	32.1878	2.058040	3.6692
0.188191	32.1878	2.386460	2.6768
0.236393	31.5011	2.844670	1.7951
0.286001	30.1304	3.552400	1.0559
0.393534	25.3579	4.832200	0.5000
0.516459	20.4370	6.802860	0.5000
0.658912	16.6917	8.759970	0.5000

Table 3: Mesh configurations used for the confined axisymmetric glass bead model

Mesh	Elements ($N_r \times N_z$)	Number of elements	Approx. element size $h_r \times h_z$ [mm]
Very coarse	3×6	18	11.85×11.85
Coarse	4×8	32	8.89×8.89
Medium	6×12	72	5.93×5.93
Fine	12×24	288	2.96×2.96

The global response of the confined axisymmetric glass bead model is evaluated in terms of σ_1/σ_3 as a function of ε_a. Figure 1 compares the experimental reference response with the numerical results obtained using four different mesh sizes. The figure also includes the response of the one element material point simulation.

The deviation from the experimental response at the final loading state is quantified by the absolute error at 20 % axial strain

$$e_{20} = \left| \left(\frac{\sigma_1}{\sigma_3} \right)_{\text{exp}} - \left(\frac{\sigma_1}{\sigma_3} \right)_{\text{sim}} \right|_{\varepsilon_a = 20\%} . \tag{1}$$

The material point result is not included in this error plot because it is used only as a homogeneous constitutive calibration check.

The material point response follows the experimental reference curve closely over the full strain range. This indicates that the selected constitutive parameters and cohesion hardening softening law are consistent with the intended homogeneous triaxial response. However, the multi element simulations deviate from the material point and experimental responses after the peak. The numerical curves for different mesh sizes start to separate, and the post peak response does not show a clear convergence trend with mesh refinement. The error measure shown in Figure 2

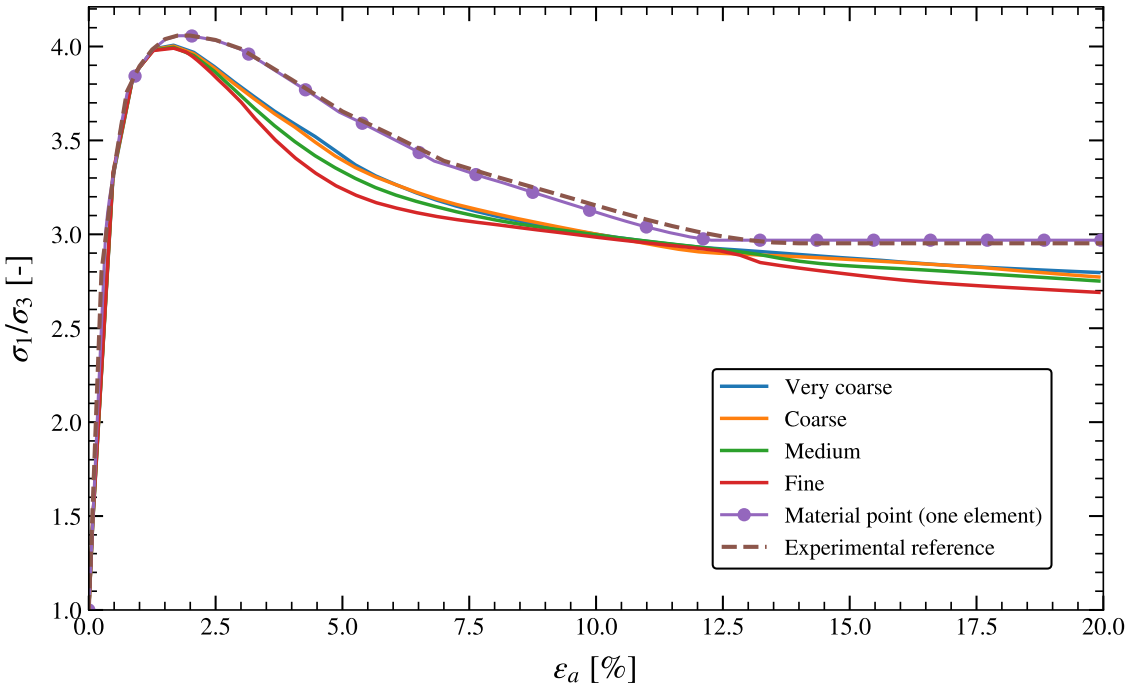


Figure 1: Principal stress ratio σ_1/σ_3 versus axial strain ε_a for the confined axisymmetric glass bead model.

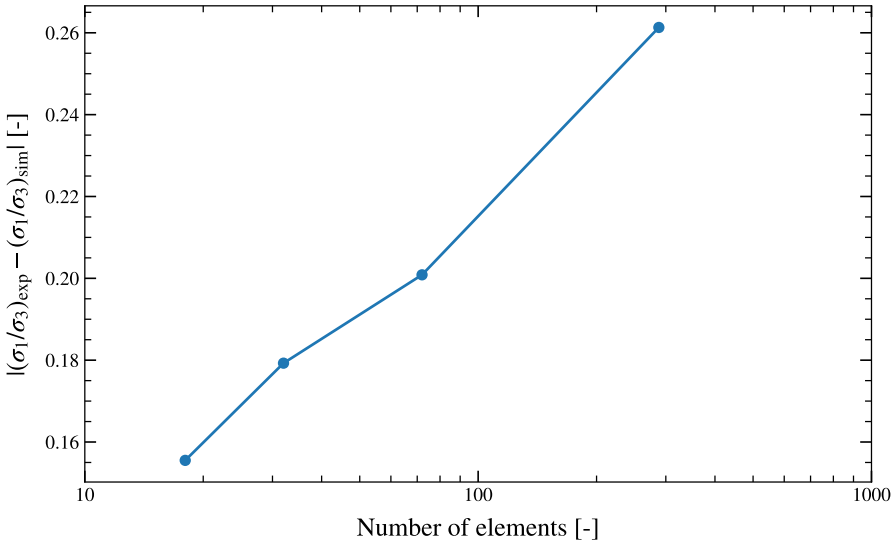


Figure 2: Absolute error in σ_1/σ_3 at $\varepsilon_a = 20\%$ versus number of elements for the axisymmetric glass bead simulations.

further illustrates that the deviation from the experimental stress ratio at $\varepsilon_a = 20\%$ changes with the number of elements. Therefore, the global response becomes dependent on mesh size after localization starts, even though the differences between the numerical curves remain moderate. In this axisymmetric model, the curves remain comparatively smooth for all mesh sizes, indicating that no pronounced numerical oscillations are observed in the global response when using the quadratic elements with reduced integration.

The influence of mesh size is more clearly visible in the spatial distribution of equivalent plastic strain. In the Abaqus Mohr–Coulomb formulation, the equivalent plastic strain $\bar{\varepsilon}^{pl}$ is the internal variable governing the isotropic evolution of cohesion, i.e. $c = c(\bar{\varepsilon}^{pl})$ (Dassault Systèmes, 2024a). The same equivalent plastic strain is reported by Abaqus as the output variable PEEQ (Dassault Systèmes, 2024c). For Mohr–Coulomb plasticity, Abaqus defines this equivalent plastic strain through the accumulated plastic work as

$$\bar{\varepsilon}^{pl} = \int \frac{1}{c}\, \boldsymbol{\sigma} : \mathrm{d}\boldsymbol{\varepsilon}^{pl}, \tag{2}$$

where c is the current cohesion yield stress, $\boldsymbol{\sigma}$ is the stress tensor, and $\mathrm{d}\boldsymbol{\varepsilon}^{pl}$ is the plastic strain increment. In the present study, PEEQ is used as a scalar indicator of accumulated plastic deformation and strain localization. Figure 3 compares the PEEQ contours for the four mesh sizes. In the very coarse and coarse meshes, plastic deformation is distributed over a relatively broad central region of the specimen. With mesh refinement, the plastic strain becomes more localized. In the fine mesh, a distinct X shaped localization pattern develops, and the highest values of PEEQ are concentrated in much narrower zones.

These results show that the calibrated hardening softening law reproduces the experimental stress ratio response under homogeneous material point loading. In the specimen scale simulations, however, the boundary conditions and contact constraints allow a nonuniform deformation field to develop. After the peak, shear strain and plastic strain localize, leading to stress redistribution within the specimen. Consequently, both the post peak global response and the PEEQ pattern become dependent on the mesh size.

3 Confined plane strain sand model

The second model set represents a confined plane strain compression test on dense silica sand. The reference response is obtained from a DEM simulation with $D_{50} = 0.5\,\mathrm{mm}$ and $C_u = 2$, using spherical particles. The global response is evaluated using the principal stress ratio σ_1/σ_3 and axial strain ε_a.

The FEM specimen is $30 \times 30\,\mathrm{mm}$, laterally confined, and compressed between rigid plates. It is discretized using four node bilinear plane strain quadrilateral elements with reduced integration (CPE4R).

The material response is described using a Mohr–Coulomb model with linear cohesion hardening, consistent with the continuously increasing DEM stress ratio response. In Abaqus, cohesion

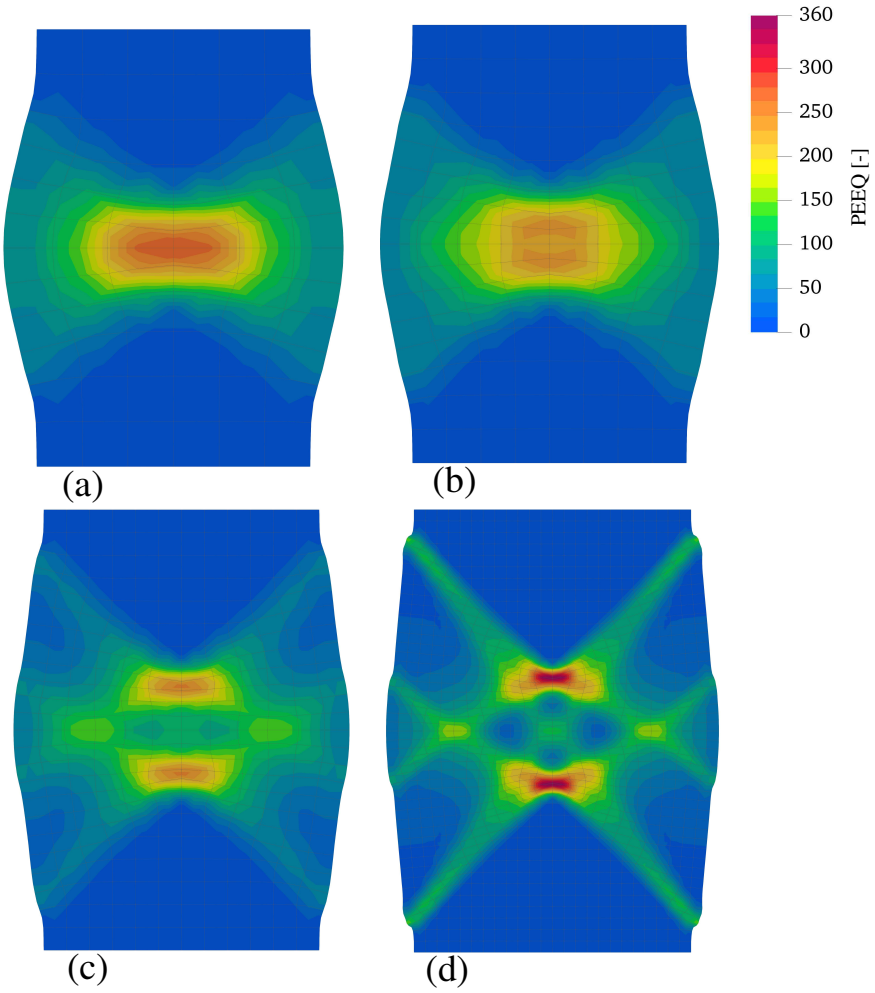


Figure 3: Equivalent plastic strain (PEEQ) contours for the confined axisymmetric glass bead model at $\varepsilon_a = 20\%$: (a) very coarse, (b) coarse, (c) medium, and (d) fine mesh. The same PEEQ contour scale is used for all four meshes.

evolves as a function of the equivalent plastic strain $\bar{\varepsilon}^{pl}$. The fine mesh is fitted to the DEM reference response, after which only the mesh size is varied; the material parameters, hardening law, boundary conditions, contact definition, and loading procedure remain unchanged.

The material and loading parameters, cohesion hardening law, and mesh configurations are summarized in Tables 4–6.

Figure 4 compares the DEM reference response with the Abaqus simulations for the four mesh sizes in terms of σ_1/σ_3 and ε_a. The fine mesh, which is used as the fitted reference mesh, reproduces the overall trend of the DEM curve reasonably well.

Figure 5 shows the cumulative particle rotation in the DEM midsection at $\varepsilon_a = 10\%$. Larger rotations concentrate in broad intersecting inclined zones, indicating localized particle rearrangement and shear deformation. Although cumulative particle rotation is not equivalent to continuum

Table 4: Material and loading parameters used for the confined plane strain sand model

Parameter	Symbol	Value
Young's modulus	E	140 MPa
Poisson's ratio	ν	0.45
Friction angle	φ	35°
Dilation angle	ψ	0°
Contact friction coefficient	μ	0.15
Lateral confinement	σ_3	110 kPa
Target axial strain	ε_a	10%

Table 5: Cohesion hardening law used for the confined plane strain sand model

Equivalent plastic strain, $\bar{\varepsilon}^{pl}$ [-]	Cohesion, c [kPa]
0.0	0.5
4.0	600

plastic strain, it provides a particle scale indicator of shear zone development for qualitative comparison with the FEM localization patterns.

The FEM localization patterns are evaluated using PEEQ, defined in Eq. (2). The DEM rotation field and FEM PEEQ contours are compared qualitatively in terms of the location, orientation, and width of the localized zones.

Figure 6 shows the PEEQ contours at $\varepsilon_a = 10\%$ for the four mesh sizes. The very coarse mesh produces a broad and relatively smooth plastic zone near the centre of the specimen. With mesh refinement, the plastic strain distribution becomes more structured and localized. In the medium and fine meshes, inclined bands become more visible, and the high PEEQ zones are concentrated in narrower regions. This shows that the predicted localization pattern depends on the mesh size, even though the same material parameters are used in all simulations.

Both DEM and FEM exhibit intersecting inclined localization zones with broadly similar orientations. The DEM zones are comparatively broad and diffuse, whereas the FEM high PEEQ zones narrow with mesh refinement and are substantially thinner in the fine mesh. Since cumulative particle rotation and PEEQ represent different physical measures, band inclination and width are compared only qualitatively.

Table 6: Mesh configurations used for the confined plane strain sand model

Mesh	Elements ($N_x \times N_y$)	Number of elements	Approx. element size $h_x \times h_y$ [mm]
Very coarse	6×6	36	5.00×5.00
Coarse	12×12	144	2.50×2.50
Medium	24×24	576	1.25×1.25
Fine	36×36	1296	0.83×0.83

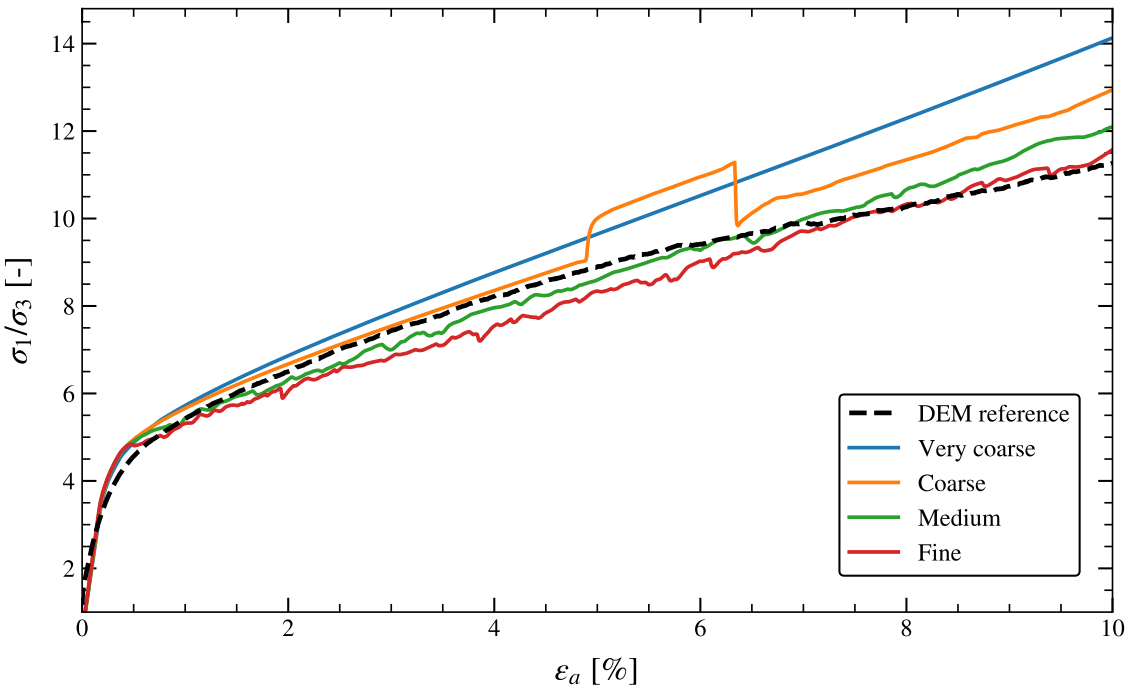


Figure 4: Principal stress ratio σ_1/σ_3 versus axial strain ε_a for the confined plane strain sand model. The fine mesh is fitted to the DEM reference response, while the coarser meshes are calculated using the same material parameters.

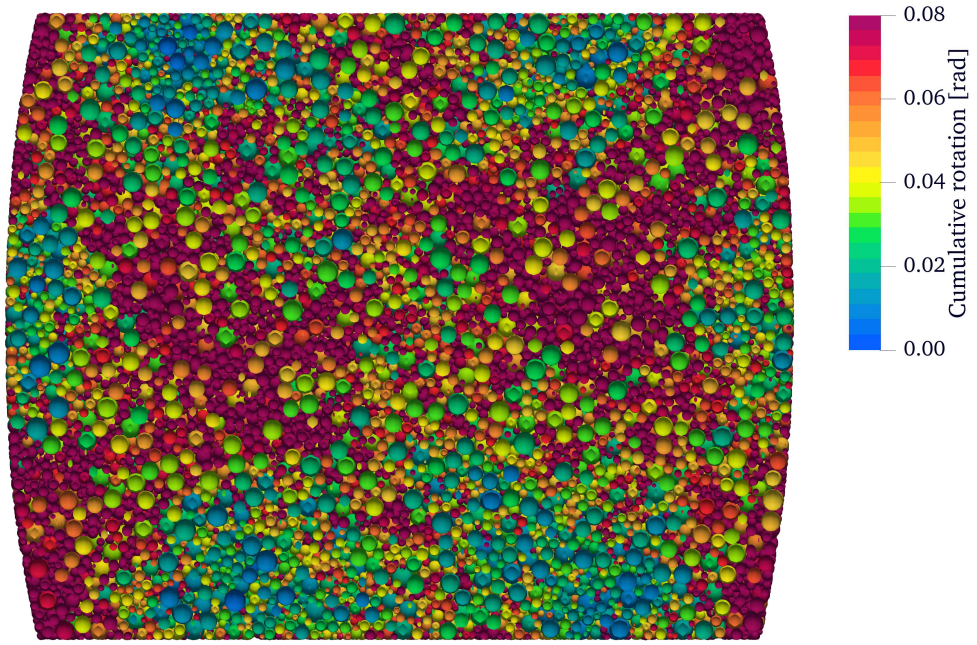


Figure 5: Cumulative particle rotation in the midsection of the DEM reference simulation at $\varepsilon_a = 10\%$.

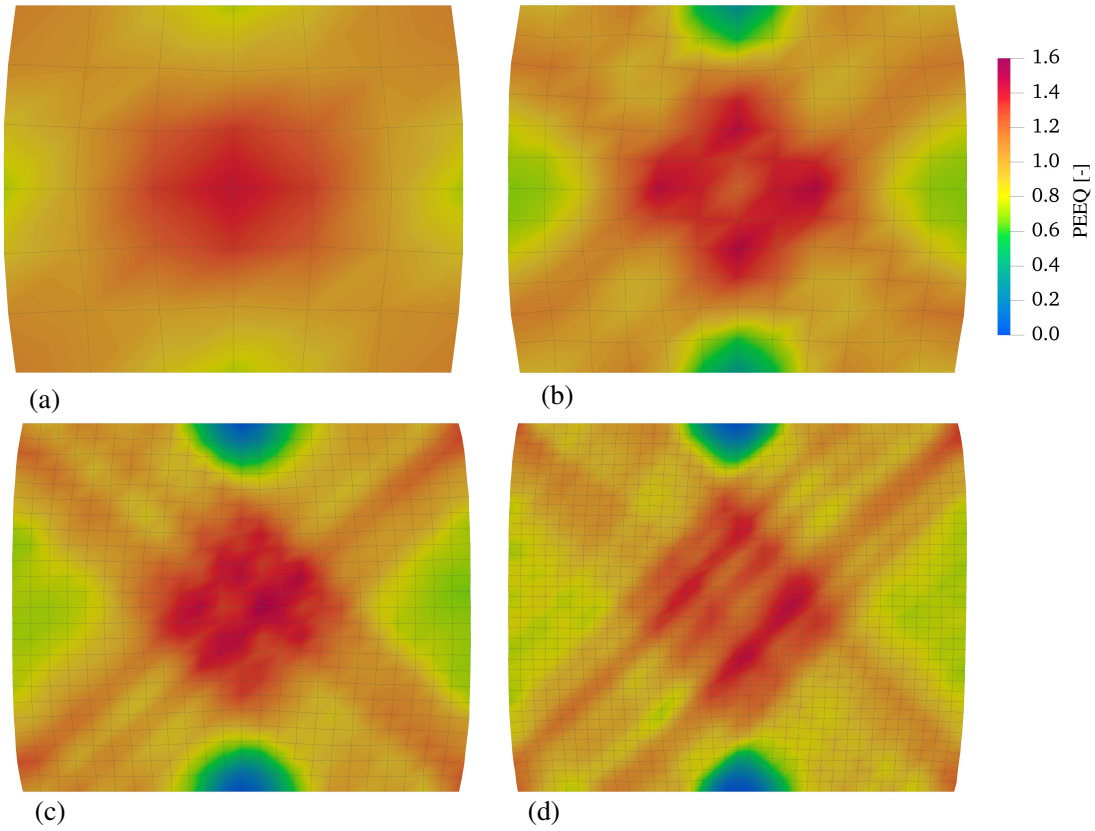


Figure 6: Equivalent plastic strain (PEEQ) contours for the confined plane strain sand model at $\varepsilon_a = 10\%$: (a) very coarse, (b) coarse, (c) medium, and (d) fine mesh. The same PEEQ contour scale is used for all four meshes.

All Abaqus curves almost coincide in the initial elastic part of the response. The influence of mesh size becomes visible after the response enters the plastic range. The coarser meshes predict higher stress ratio levels, and the deviation increases with axial strain. In addition, the plastic part of the response becomes less smooth as the mesh is refined. This numerical noise is most visible in the finer meshes and indicates the increasing sensitivity of the localized plastic regime. Overall, the same material parameters do not produce a mesh objective global response after plastic deformation develops.

4 Discussion

The two model sets illustrate the distinction between homogeneous constitutive calibration and specimen scale boundary value behavior. Before localization, the responses for different meshes remain comparatively close. Once localization develops, however, the global response becomes sensitive to the discretization.

Since the classical Cauchy continuum does not introduce an internal length scale, the width and intensity of localized deformation are controlled partly by the discretization. This explains the mesh dependence of both the post peak stress ratio response and the PEEQ patterns.

The specific position and inclination of localized bands should not be attributed to mesh size alone. They may also be influenced by boundary conditions and contact constraints, element formulation, axisymmetric or plane strain assumptions, symmetry restrictions, and numerical perturbations that select between competing localization modes. The present study therefore focuses on the loss of mesh objectivity rather than on predicting a unique band position or inclination.

The axisymmetric CAX8R simulations remain comparatively smooth, whereas the plane strain CPE4R simulations exhibit stronger numerical noise in the plastic regime. The DEM rotation and FEM PEEQ fields provide qualitatively similar indications of concentrated deformation but represent different physical quantities.

Overall, the observed mesh dependency motivates regularized formulations with an internal length scale, such as nonlocal, gradient enhanced, or micropolar/Cosserat continua (Liu et al., 2022).

5 Conclusions

This paper investigated mesh dependency in two calibrated geotechnical finite element model sets using Mohr–Coulomb plasticity, with the material parameters kept unchanged while the mesh size was varied.

For the axisymmetric glass bead model, the material point simulation reproduced the experimental response closely, whereas the post peak specimen response and PEEQ localization pattern changed with mesh refinement.

In the plane strain sand model, the elastic response was nearly mesh independent, while differences in the global response and PEEQ localization developed after plastic deformation.

The DEM and FEM results both indicated inclined zones of concentrated deformation, although cumulative particle rotation and PEEQ represent different physical quantities.

The results demonstrate that successful homogeneous calibration does not guarantee mesh objective predictions once strain localization develops in a local Cauchy continuum. Objective simulation of localized deformation therefore requires regularization with an internal length scale, for example through nonlocal, gradient enhanced, or micropolar/Cosserat formulations.

References

[1] Alshibli K. A., Alsaleh M. I. and Voyiadjis G. Z. (2006): Modelling strain localization in granular materials using micropolar theory: Numerical implementation and verification. *International Journal for Numerical and Analytical Methods in Geomechanics*, 30:1525–1544.

[2] Alshibli K. A., Batiste S. N. and Sture S. (2003): Strain localization in sand: Plane strain versus triaxial compression. *Journal of Geotechnical and Geoenvironmental Engineering*, 129(6):483–494.

[3] Alshibli K. A. and Cil M. B. (2017): Influence of particle morphology on the friction and dilatancy of sand. *Journal of Geotechnical and Geoenvironmental Engineering*, 143(9).

[4] Alshibli K. A. and Sture S. (1999): Sand shear band thickness measurements by digital imaging techniques. *Journal of Computing in Civil Engineering*, 13(2):103–109.

[5] Dassault Systèmes (2024a): Abaqus 2024 Documentation: Mohr–Coulomb plasticity. SIMULIA User Assistance.

[6] Dassault Systèmes (2024b): Abaqus 2024 Documentation: *MOHR COULOMB HARDENING. SIMULIA User Assistance.

[7] Dassault Systèmes (2024c): Abaqus 2024 Documentation: Output variable identifiers. SIMULIA User Assistance.

[8] De Borst R. (1991): Simulation of strain localization: A reappraisal of the Cosserat continuum. *Engineering Computations*, 8:317–332.

[9] Ehlers W., Ellsiepen P. and Ammann M. (2001): Time- and space-adaptive methods applied to localization phenomena in empty and saturated micropolar and standard porous materials. *International Journal for Numerical Methods in Engineering*, 52:503–526.

[10] Han C. and Drescher A. (1993): Shear bands in biaxial tests on dry coarse sand. *Soils and Foundations*, 33(1):118–132.

[11] Liu J., Wu L., Yin K., Song C., Bian X. and Li S. (2022): Methods for solving finite element mesh-dependency problems in geotechnical engineering–a review. *Sustainability*, 14:2982.

[12] Mühlhaus H.-B. and Vardoulakis I. (1987): The thickness of shear bands in granular materials. *Géotechnique*, 37(3):271–283.

[13] Stefanou I. and Gerolymatou E. (2019): Strain localization in geomaterials and regularization: Rate-dependency, higher order continuum theories and multi-physics. *ALERT Doctoral School*, 780:47–85.

[14] Tatsuoka F., Nakamura S., Huang C. and Tani K. (1990): Strength anisotropy and shear band direction in plane strain tests of sand. *Soils and Foundations*, 30(1):35–54.

Author

Mazdak Atri, M.Sc.
HafenCity Universität Hamburg
Geotechnik
Henning-Voscherau-Platz 1
20457 Hamburg
Tel.: +49 40 300880-4383
email: mazdak.atri@hcu-hamburg.de
Web: https://www.hcu-hamburg.de

Univ.-Prof. Dr.-Ing. Tim Pucker
HafenCity Universität Hamburg
Geotechnik
Henning-Voscherau-Platz 1
20457 Hamburg
Tel.: +49 40 300880-5494
email: tim.pucker@hcu-hamburg.de
Web: https://www.hcu-hamburg.de

From Data to Design: Insights from the SAGE-SAND and EXCALIBUR Projects and a Class A Blind Prediction Test

Stijn François, George Anoyatis, Hadrien Rattez, Frédéric Collin, Kristof Maes, Luc Simonin, Michail Spyridis, Pauline André, Thrasyvoulos-Konstantinos Elezoglou

Abstract: The growth of offshore wind energy has increased demands on the design, installation, operation and decommissioning of monopile foundations. This paper synthesises recent developments within the Belgian Energy Transition Fund projects SAGE-SAND (Soil Ageing Around Offshore Wind Turbine Foundations in Sandy Soils) and EXCALIBUR (EXperimental and Computational Analysis of monopile extraction), which combine full-scale field testing, laboratory characterisation and multi-scale numerical modelling to investigate soil ageing, installation effects and extractability of monopile foundations. Particular attention is given to a Class A blind prediction competition organised within SAGE-SAND, in which international teams predicted the lateral response of full-scale monopiles from pre-test geotechnical data alone. While several advanced numerical models reproduced key aspects of the measured response with good accuracy, substantial variability was observed among submissions, underlining the importance of parameter selection, constitutive modelling and uncertainty treatment. The combined outcomes of both projects and the blind prediction exercise illustrate how field data, micromechanical understanding and physics-based models can be integrated within a model-informed geotechnical design framework, encompassing installation, operation, ageing and decommissioning.

1 Introduction

Offshore wind energy is a central pillar of the European energy transition, and continuing increases in turbine capacity, rotor diameter and hub height are driving a corresponding growth in foundation size and loading demands. Monopiles remain the dominant foundation type due to their structural simplicity and well-established installation procedures (Byrne and Houlsby, 2003), but accurate prediction of soil–structure interaction is increasingly critical as diameters grow and sites become geotechnically more challenging (Arany et al., 2017).

Pile installation substantially disturbs the surrounding soil, altering stress state, density, fabric and pore pressure (White and Bolton, 2004). Impact driving remains the standard technique, while vibratory driving is gaining interest for its lower noise emissions and faster installation, with a well-documented influence on subsequent pile performance (Heins et al., 2020; Bienen

et al., 2021). Following installation, soil properties continue to evolve through hydro-mechanical and physicochemical processes collectively known as soil ageing, long recognised for axial pile capacity (Chow et al., 1998; Gavin and Igoe, 2021) but still poorly understood for the lateral response of large-diameter monopiles (Mitchell, 2008; Castilla-Barbosa et al., 2024).

Major research programmes such as PISA (Byrne et al., 2020; Taborda et al., 2020) and GDP (Tsetas et al., 2023) have substantially advanced monopile design and vibratory installation technology, respectively, but installation effects and soil ageing on lateral behaviour remain open questions. The Belgian Energy Transition Fund therefore supports two complementary projects: SAGE-SAND, combining full-scale field testing, laboratory characterisation and multi-scale numerical modelling at a dedicated site in the Port of Zeebrugge (Maes et al., 2025; Andries, 2025; Desrues et al., 2019); and EXCALIBUR, which currently extends this framework towards monopile extraction and decommissioning. Within SAGE-SAND, an international Class A blind prediction competition was also organised with ISSMGE Technical Committee 212 on Deep Foundations, building on the long tradition of blind prediction exercises as an objective benchmark of predictive capability.

This paper synthesises the main outcomes of SAGE-SAND (Section 2) project, outlines the EXCALIBUR project (Section 3), presents the blind prediction competition and its results (Section 4), and discusses the combined implications for model-informed offshore geotechnical design (Section 5), before concluding in Section 6.

2 The SAGE-SAND project

SAGE-SAND was initiated to understand how soil properties around offshore monopiles evolve after installation. While ageing effects on axial pile capacity are widely documented, their implications for the lateral response of large-diameter monopiles remain largely unquantified, despite being increasingly governed by serviceability and stiffness requirements. The project investigates the coupled effects of installation method, soil ageing and cyclic loading through an integrated field, laboratory and numerical programme, with particular attention to differences between impact- and vibratory-driven monopiles, as illustrated in figure 1.

A dedicated full-scale test site was established in the Port of Zeebrugge, representative of southern North Sea sandy conditions. Following a comprehensive CPT, sampling and laboratory characterisation campaign (Andries, 2025), four instrumented monopiles were installed using different techniques and monitored during installation, ageing and a subsequent series of lateral load tests at different times after installation (Maes et al., 2025). This combination of installation records, geotechnical data and full-scale load tests under controlled conditions remains rare in offshore geotechnics and provides an important benchmark dataset. Laboratory testing on recovered samples, combined with advanced microstructural investigation, targeted the mechanisms behind ageing: stress redistribution, pore pressure dissipation, particle rearrangement, contact evolution and physicochemical bonding. These observations were complemented by a multi-scale numerical programme, from continuum finite-element analyses of installation

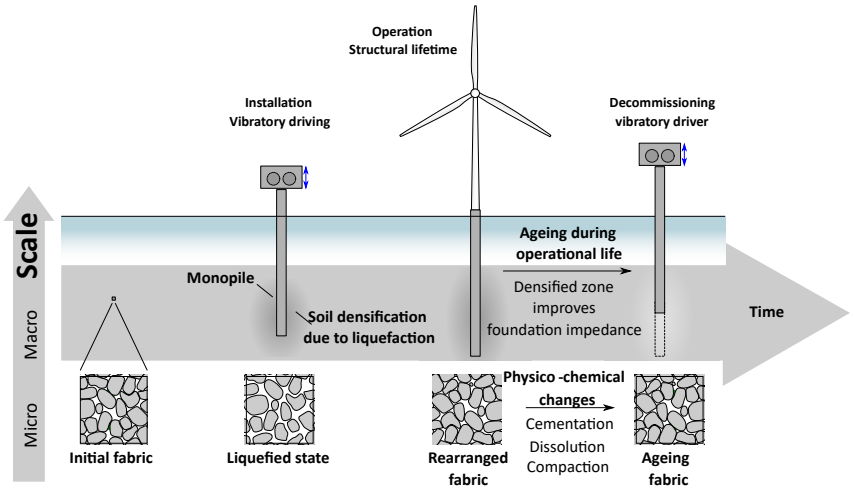


Figure 1: Influence of installation and soil aging effects on the response of offshore wind turbines over its lifespan.

and loading to FEM–DEM approaches capturing grain-scale rearrangement and stress transfer during cyclic loading and ageing. The Zeebrugge dataset and this modelling chain provide the experimental and mechanistic foundation for the blind prediction exercise discussed in Section 4.

3 EXCALIBUR: Towards Predictive Decommissioning

As the first generation of offshore wind farms approaches the end of its design life, sustainable decommissioning is becoming an important engineering and environmental challenge. Current practice commonly relies on partial pile removal, leaving significant steel volumes in the seabed and complicating future redevelopment (Topham et al., 2019; Al Mowafy et al., 2024). EXCALIBUR (EXperimental and Computational Analysis of monopile extraction: Leveraging Innovation and Better Understanding for Renewable offshore wind energy) builds directly on the SAGE-SAND framework to address this gap, hypothesising that extraction behaviour is strongly governed by installation history and subsequent soil ageing, including the geochemical cementation processes identified within SAGE-SAND, which can increase interface stiffness and shaft resistance over time and thus extraction forces many years later.

The experimental programme follows the same multi-scale logic as SAGE-SAND, from laboratory-scale mechanisms (vibration-induced liquefaction, cyclic degradation, interface behaviour) through intermediate-scale soil-pit testing to full-scale extraction trials at the Zeebrugge piles, using conventional vibratory extraction, combined vibration/water-jetting, and new industry-developed techniques. Particular attention is paid to characterising cemented crusts observed on piles recovered from marine environments, attributed to coupled corrosion, dissolution–precipitation and carbonate cementation, using advanced imaging and geochemical modelling of pore-water evolution around steel monopiles.

Numerically, engineering-scale FE and Material Point Method (MPM) simulations target the large deformations and complex soil flow of extraction, while discrete-element and hybrid FEM–DEM approaches resolve grain rearrangement and cemented-contact failure under repeated vibratory loading. As such, SAGE-SAND and EXCALIBUR thus span the full monopile lifecycle (installation, operation, ageing and decommissioning) within one coherent research and dataset framework.

4 Class A Blind Prediction Competition

Blind prediction exercises, in which participants analyse a problem before experimental results are released, eliminate retrospective calibration and provide an objective benchmark of modelling approaches, constitutive assumptions and parameter selection procedures (Jardine, 2001). Within SAGE-SAND, such a competition was organised with ISSMGE TC212 to evaluate current practice in predicting the lateral response of large-diameter monopiles in sand, and to identify the principal sources of uncertainty in advanced numerical analyses.

The exercise was based on two instrumented Zeebrugge monopiles, referred to as P3 (impact-driven) and P4 (vibratory-driven), each with an outer diameter of about 2 m and an embedded length of about 17 m, subsequently subjected to large-scale lateral load tests (Figure 2). Participants received the full geotechnical database (CPTs, laboratory tests, pile geometry, loading configuration, installation method) but no load-test measurements, so that the exercise evaluated the complete workflow from site interpretation to numerical implementation, not just the constitutive model. Predictions were scored against five indicators: the initial lateral stiffness K_0; the loads $H_{2\%}$ and $H_{7.5\%}$ at pile-head displacements of 2% and 7.5% of the diameter; the depth and magnitude of the maximum axial pile strain at $y/D = 7.5\%$; and report quality.

(a)  (b) 

Figure 2: (a) instrumented piles and (b) lateral load test.

Ten independent submissions were received from researchers across nine countries, spanning academia, research institutes and practice. Approaches ranged from fully 3D finite-element models (shell, beam or brick pile elements; constitutive models from Mohr–Coulomb to advanced hypoplastic and stiffness-degrading formulations) to simplified beam-on-p–y Winkler models. Soil parameters were derived variously from CPT correlations, triaxial data, or combinations

thereof, and only about half of the submissions explicitly accounted for installation effects, the rest assuming a wished-in-place pile.

The measured backbone response gave $K_0 = 71.0$ and 63.0 MN/m, the horizontal force at a displacement of 2% and 7.5% of the diameter, i.e. $H_{2\%} = 1.63$ and 1.26 MN, and $H_{7.5\%} = 3.71$ and 3.05 MN for piles P3 and P4, respectively, with maximum axial strains around 1560 $\mu\varepsilon$ (7.5 m depth, P3) and 1200 $\mu\varepsilon$ (7.6 m, P4). Comparison against the submitted predictions (Figure 3) reveal a very large scatter, with maximum-to-minimum ratios of about 100–110 on K_0, 30–32 on $H_{2\%}$ and 13–14 on $H_{7.5\%}$ across all submissions. Restricting the comparison to the seven teams using advanced 3D models narrowed this considerably (factors of about 11–19 on K_0 and 4–6 on $H_{2\%}/H_{7.5\%}$), confirming that deformation-based modelling with rigorous parameter identification and explicit treatment of installation effects markedly improves predictive fidelity, even though notable uncertainty remains. Almost all reports were nonetheless clear and well organised, regardless of numerical accuracy.

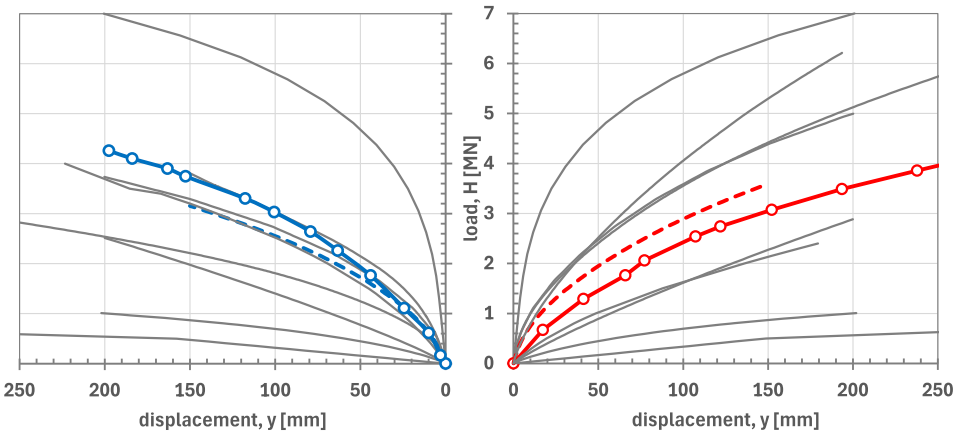


Figure 3: Comparison of submitted predictions (grey lines) against the measured backbone curves (dotted line) for pile P3 (impact driven, left) and pile P4 (vibratory driven, right). The average of the predictions is indicated with a dashed line.

It is important to note that the dominant source of this scatter was not the numerical solutiIon method itself but the interpretation of the shared geotechnical dataset: friction angle and stiffness profiles derived from the same CPTs already varied substantially between teams, and this variability propagated, largely unchanged, through the choice of constitutive model and treatment of installation effects into the final predicted response.

5 From Data to Design

The full-scale WP1-type tests within SAGE-SAND show measurable, but moderate, differences between impact- and vibratory-driven piles, smaller than several blind predictions anticipated, while the micromechanical and macroscale modelling chain indicates that both mechanical and geochemical ageing act, at least partially, to compensate the lower initial capacity typically

associated with vibratory installation. This is precisely the mechanism EXCALIBUR aims to quantify further, since it governs whether vibratory installation and extraction can be considered a genuinely lower-impact, circular-economy alternative over the full life of a wind farm.

The blind prediction competition shows that this promise is currently limited less by a lack of adequate numerical tools than by the difficulty of consistently deriving soil parameters from shared site data: teams using advanced 3D models with a rigorous, physically grounded parameter derivation and an explicit account of installation effects achieved markedly closer agreement with measurements than those relying on looser assumptions, even though the same raw CPT and laboratory data were available to everyone. This mirrors earlier prediction exercises for laterally loaded piles, where scatter was likewise dominated by differing assumptions rather than by the modelling framework itself, and it points to systematic parameter identification and transparent uncertainty reporting, rather than more sophisticated constitutive models alone, as the priority for closing the gap between predictive capability and practice.

6 Conclusions

SAGE-SAND and EXCALIBUR combine full-scale field testing, advanced laboratory characterisation and multi-scale numerical modelling to quantify how installation method and long-term soil ageing jointly govern monopile performance, and to assess vibratory driving as a lower-impact alternative for both installation and decommissioning. The Class A blind prediction competition, associated to the SAGE-SAND project, demonstrates that deformation-based numerical modelling, combined with rigorous, physically informed parameter identification and explicit treatment of installation effects, can reproduce the essential features of measured pile response with good fidelity. At the same time, the very large scatter among the ten submissions confirms that reliable, consistently applicable prediction of laterally loaded offshore monopiles is not yet achieved, and that the largest source of uncertainty lies in the derivation of soil parameters rather than in the numerical methods themselves. Continued investment in integrated, data-informed research such as SAGE-SAND and EXCALIBUR, and in benchmarking exercises such as this blind prediction competition, is needed to close this gap and move towards more robust, model-informed offshore geotechnical design.

References

[1] Al Mowafy A., Lotfian S. and Brennan F. (2024): Decommissioning Offshore Wind Fixed Steel Pile Foundations: A Critical Review. *Energies*, 17(21).

[2] Andries J. (2025): Factual Report on the Geotechnical Laboratory Testing on Soil Samples from the SAGE-SAND Project Site in Zeebrugge. Technical Report GEO-24/189, Version 2.0, NeXperta – Eigen Vermogen Xperta, Geotechniek, Gent, Belgium.

[3] Arany L., Bhattacharya S., Macdonald J. and Hogan S.J. (2017): Design of monopiles for

offshore wind turbines in 10 steps. *Soil Dynamics and Earthquake Engineering*, 92:126–152.

[4] Bienen B., Fan S. and Randolph M.F. (2021): Effect of the installation process on monopile lateral response. *Proceedings of the Institution of Civil Engineers – Geotechnical Engineering*, 174(5):530–548.

[5] Byrne B.W. and Houlsby G.T. (2003): Foundations for offshore wind turbines. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 361:2909–2930.

[6] Byrne B.W., McAdam R.A., Beuckelaers W.J.A.P., Burd H.J., Gavin K., Houlsby G.T., Igoe D.J.P., Jardine R. and Martin C.M. (2020): Cyclic Laterally Loaded Medium Scale Field Pile Testing for the PISA Project. Technical report.

[7] Castilla-Barbosa M., Ocampo-Terreros M. and Rincón-Arango O. (2024): A Review of Sand Aging: Mechanisms and Impacts. *Geotechnical and Geological Engineering*.

[8] Chow F.C., Jardine R.J., Brucy F. and Nauroy J.F. (1998): Effects of time on capacity of pipe pile in dense marine sand. *Journal of Geotechnical and Geoenvironmental Engineering*, 124(3):254–264.

[9] Desrues J., Argilaga A., Caillerie D., Combe G., Nguyen T.K., Richefeu V. and Dal Pont S. (2019): From discrete to continuum modelling of boundary value problems in geomechanics: An integrated FEM–DEM approach. *International Journal for Numerical and Analytical Methods in Geomechanics*, 43(5):919–955.

[10] Gavin K. and Igoe D. (2021): A field investigation into the mechanisms of pile ageing in sand. *Géotechnique*, 71(2):120–131.

[11] Heins E., Randolph M.F., Bienen B. and Grabe J. (2020): Effect of installation method on static and dynamic load test response for piles in sand. *International Journal of Physical Modelling in Geotechnics*, 20(1):1–23.

[12] Jardine R.J. (2001): A competition to assess the reliability of pile prediction methods. Technical report.

[13] Maes K., François S., Simonin L., Bayart C., Bertholet A., Rattez H. and Anoyatis G. (2025): Experimental investigation of the installation and lateral capacity of instrumented monopiles at the SAGE-SAND test site in Zeebrugge. In: Proceedings of ISFOG 2025, 5th International Symposium on Frontiers in Offshore Geotechnics, Nantes, France.

[14] Mitchell J.K. (2008): Aging of Sand – a Continuing Enigma? 6th International Conference on Case Histories in Geotechnical Engineering, pp. 1–21.

[15] Taborda D.M.G., Zdravković L., Potts D.M., Burd H.J., Byrne B.W., Gavin K.G., Houlsby G.T., Jardine R.J., Liu T., Martin C.M. and McAdam R.A. (2020): Finite-element modelling of laterally loaded piles in a dense marine sand at Dunkirk. *Géotechnique*, 70(11):1014–1029.

[16] Topham E., Gonzalez E., McMillan D. and João E. (2019): Challenges of decommissioning offshore wind farms: Overview of the European experience. *Journal of Physics: Conference Series*, 1222(1).

[17] Tsetas A., Tsouvalas A., Gómez S.S., Pisanò F., Kementzetzidis E., Molenkamp T., Elkadi A.S.K. and Metrikine A.V. (2023): Gentle Driving of Piles (GDP) at a sandy site combining axial and torsional vibrations: Part I – installation tests. *Ocean Engineering*, 270.

[18] White D.J. and Bolton M.D. (2004): Displacement and strain paths during plane-strain model pile installation in sand. *Géotechnique*, 54(6):375–397.

Authors

Stijn François, Kristof Maes, Michail Spyridis
KU Leuven, Department of Civil Engineering, Structural Mechanics Section
Kasteelpark Arenberg 40
3001 Leuven
Belgium

George Anoyatis, Thrasyvoulos-Konstantinos Elezoglou, Pauline André*
KU Leuven, Department of Civil Engineering, Hydraulics and Geotechnics Section, Bruges Campus
Spoorwegstraat 12
8200 Bruges
Belgium

Hadrien Rattez, Luc Simonin, Pauline André*
UCLouvain, Vinci Building
Place du Levant 1
1348 Louvain-la-Neuve
Belgium

Frédéric Collin
Université de Liège, ArGEnCo Department
Allée de la Découverte 9 (Bât. B52/3)
4000 Liège
Belgium

*Also affiliated with KU Leuven (Hydraulics and Geotechnics Section, Bruges Campus) and UCLouvain.

Linking numerical modelling to construction work: challenges and opportunities

Raul Fuentes

Abstract: Advanced numerical models reproduce only the construction process represented in them. This paper examines this limitation through deep excavations and mechanised shield tunnels. In a London excavation, a temporary ramp produced 2 mm reverse wall movement, around 30% of the late-stage displacement, while delayed movement after prop removal approached 9 mm. In TULIP, a 0.2 bar grout-pressure difference coincided with a 41% difference in peak settlement. During a 100-day TBM stoppage, a Soft Soil model predicted 0.5 mm face displacement against 9 mm measured until construction-induced disturbance was included. Constitutive sophistication cannot compensate for incorrect geometry, sequence, timing or operational boundary conditions. As-built records, machine data and monitoring provide the main opportunities.

1 Introduction

Numerical modelling in geotechnical engineering has advanced rapidly. Three-dimensional geometry, non-linear constitutive laws, coupled hydro-mechanical calculations and increasingly automated workflows are now available for routine or specialist design. This development can create the impression that predictive reliability is primarily a question of selecting a sufficiently advanced constitutive model and calibrating its parameters. The field evidence considered here shows that this view is incomplete. A calculation also reproduces the geometry, sequence, timing, operational controls and temporary states that the modeller chooses to represent. Processes omitted from that description cannot be recovered by constitutive sophistication.

The principle is consistent with the observational method. Peck (1969) treated design predictions as conditional expectations that must be checked during construction, while Simpson (1992) emphasised that displacements depend on stiffness, stress history and the path followed during excavation and support. Monitoring is most effective when the model state corresponds to the actual construction state; otherwise, differences between calculation and measurement may be incorrectly absorbed into soil parameters.

This paper uses two types of underground construction to examine the issue quantitatively. Deep excavations are dominated by discrete construction activities such as excavation lifts, temporary ramps, prop installation and removal, and periods of inactivity. Mechanised tunnels are governed by moving and continuously changing boundary conditions, including face and grout pressures, advance rate, overcut, shield gap and stoppages. The examples are not intended as a comprehensive review. They illustrate a single proposition: realistic predictions require sufficient fidelity to the construction process, and construction information should be treated as a primary modelling input rather than as contextual project information. This may appear almost self-evident; however, the examples below show that movements generated by temporary or operational construction activities can be of the same order as the total deformation that the numerical analysis is intended to predict.

2 Construction-process fidelity

Construction-process fidelity is defined here as the degree to which a model represents the conditions that generate the actual stress path. Five components are particularly important: (i) geometry, including temporary geometry; (ii) sequence of excavation and support activation; (iii) timing and duration of stages; (iv) operational parameters that act as boundary conditions; and (v) interfaces and transient material states. A further distinction is necessary between the planned construction process, the process represented in the model, the process actually executed, and the part of that process recorded with sufficient resolution.

The limiting effect is analogous to other forms of model uncertainty. A refined constitutive law cannot correct an excavation level applied at the wrong time, an omitted temporary support, a nominal face pressure used in place of measured pressure, or a continuous tunnel advance assumed through a prolonged stoppage. The TULIP benchmark illustrates the broader difficulty. Nineteen numerical submissions used abundant site data, but the modelling of the TBM process varied substantially; even after common soil parameters were proposed, large differences remained and were attributed in part to assumptions about the tunnelling process (Bourgeois et al., 2023). The question is therefore not whether every construction detail must be represented, but which details control the response being assessed.

This also creates an identifiability problem when numerical models are calibrated against monitoring data. Different combinations of model parameters can reproduce the same measured response, particularly when calibration is based on a single response quantity; inverse analyses have consequently emphasised the value of multiple independent measurements for constraining the inferred ground behaviour (Ledesma et al., 1996; Finno and Calvello, 2005; Osouli et al., 2010). The problem is broader when construction assumptions are also uncertain. An omitted or incorrectly represented construction process may be compensated through changes in stiffness, strength or other model parameters until a measured displacement is reproduced. Agreement with one observable therefore does not demonstrate that the model has reproduced the correct mechanical state: another combination may match the same settlement while predicting different structural loads, pore pressures or subsequent behaviour. Calibration should therefore update construction assumptions as well as ground parameters and, where possible, be tested against several independent response quantities.

3 Deep excavations: sequence, temporary works and time

The St Botolph Street excavation in London provides a useful case because a sophisticated three-dimensional finite-element analysis was compared with comprehensive monitoring (Fuentes et al., 2018). The approximately 90 m by 65 m excavation was retained by secant and contiguous piled walls and supported by different prop systems. London Clay was represented with the relatively complex BRICK constitutive model using most-probable and characteristic parameter sets. The analysis included the previous basements and site history, wall installation and fifteen stages up to the completion of the main excavation. Importantly, the modelled construction process was reconstructed from available as-built records; it therefore approximated the executed process to the resolution permitted by those records. Figure 1 shows that the site did not progress as a uniform horizontal excavation: access, local temporary works, cores, raking props and raft construction created a strongly three-dimensional and evolving system.

3.1 Construction operations with measurable effects

The temporary access ramp is a particularly clear example. Its construction after the first prop stage (Stage 10; approximately 21 May 2008 in Figure 1) caused the nearby wall to move backwards by approximately 2 mm; subsequent excavation reversed the direction of movement. The significance is not simply that the wall moved out and then back again. For a path-dependent constitutive relationship such as BRICK (Simpson, 1992), the loading and subsequent unloading associated with the temporary ramp modify the stress-strain path and therefore the state from which later stages evolve. The response after removal of the ramp is not necessarily identical to that of a model in which the ramp never existed. Moreover, any construction loading acting while the ramp is present, for example from plant or traffic, acts on this modified state. Temporary geometry can therefore affect both the intermediate displacement and the subsequent response, even when the final geometry is unchanged.



Figure 1: Actual construction sequence of the St Botolph Street excavation. Reproduced from Fuentes et al. (2018).

For stages explicitly represented in the model, the most-probable BRICK parameters reproduced wall movements well, generally within approximately 2 mm of the measurements. This agreement is important: it shows that the constitutive framework was capable of reproducing the response generated by the modelled construction stages when these were tied to the available as-built record. Later discrepancies therefore identify processes that the analysis did not contain, rather than simply demonstrating inadequate constitutive complexity.

3.2 Time-dependent movements and three-dimensional geometry

After excavation, ground settlements behind the wall continued by approximately 2 mm and stabilised after about 60 days. Following prop removal, the accumulated wall movement approached 9 mm after roughly 160 days; before stabilisation, rates were generally within +/-0.5 mm/day. An undrained calculation using an advanced soil model reproduced the response before these delayed movements, but could not represent pore-pressure dissipation, construction inactivity and other time-dependent mechanisms excluded from the analysis. Trigger values derived from a numerical stage therefore require a construction chronology, not only a displacement value.

Geometry creates a related limitation. Three-dimensional stiffening towards excavation corners is well established in numerical analyses (St John et al., 2005). Measurements at St Botolph Street were likewise markedly lower close to the corners than near the centre of the

wall. The empirical method of Fuentes and Devriendt (2010), applied to this case, gave a corner-to-maximum displacement ratio of 34.2%. Applying one plane-strain prediction along the whole wall would consequently overestimate movements close to corners and could lead to inappropriate monitoring limits or mitigation. The relative magnitude of the construction effects is also significant. At Stage 15, the most-probable analysis predicted 7.09 mm wall displacement and 6.47 mm was measured. The approximately 2 mm ramp-induced movement is therefore of the order of 30% of the late-stage wall response, while the approximately 9 mm delayed movement following prop removal is of the same order as, or larger than, the displacement predicted during excavation. Although these values refer to different monitoring locations and should not be interpreted as point-by-point comparisons, they show that construction-related movements are not secondary perturbations.

4 Mechanised tunnels: operational data as boundary conditions

Mechanised tunnelling presents the same problem in a continuous form. The ground is affected by face pressure, shield interaction, overcut and taper, tail-void grouting, ring installation and the rate at which these actions move along the alignment. Process-oriented three-dimensional simulations can represent many of these mechanisms (Kasper and Meschke, 2004), but they require extensive data and contain modelling choices that are difficult to validate. The Advanced Displacement-Confinement Method (ADCM) makes the construction-to-response chain explicit: TBM operation controls circumferential confinement, which controls ground relaxation, lining loading, volume loss and surface settlement (Almog-Goldreich and Fuentes, 2026).

4.1 A quantitative grout-pressure comparison

The TULIP (Tunnelling and Limitation Impact on Piles) project in Paris provides an unusually direct comparison through two monitoring sections, SMRN and SMRS, along the same EPBM case study (Bourgeois et al., 2023; Almog-Goldreich and Fuentes, 2026). The two sections had the same tunnel depth (21.0 m), excavated diameter (9.87 m), 35 mm shield gap and recorded face pressure of 1.8 bar, with comparable face ground conditions. This comparison also illustrates the distinction between nominal, modelled and recorded construction conditions. If both sections were represented using a single nominal grout pressure, they would be essentially indistinguishable in terms of the principal TBM inputs considered here. The recorded operational data, however, show different grouting conditions: the grout pressure was 2.4 bar at SMRN and 2.2 bar at SMRS. Peak settlement was 8.6 mm at SMRN and 12.1 mm at SMRS, a difference of 3.5 mm or approximately 41%, although the grout-pressure difference was only 0.2 bar. Using the recorded values, ADCM calculated 7.9 and 12.4 mm respectively. Within that back-analysis, the lower recorded grout pressure is the simplest explanation for the different settlements of two otherwise largely comparable sections.

The comparison is not a universal settlement-per-pressure relationship; it shows that operational records can be first-order geotechnical inputs. Pressure measured at a grout pipe or pump is not necessarily the pressure acting uniformly on the ground. Tail-void grout flows, sets, contracts and evolves in stiffness and permeability, and hydro-mechanical analyses show that its representation affects deformation, pore pressure and lining response (Mohammadzamani et al., 2023). Grouting is therefore both an operational boundary condition and a time-dependent material process.

4.2 Stoppages and construction-induced disturbance

A 100-day unscheduled stoppage of an EPB machine in Ningbo soft clay gives an even stronger test (Yao et al., 2019). The 338 t machine settled 9 mm at the face while its tail rose by 6 mm. A three-dimensional Soft Soil analysis reproduced the rotation direction but predicted only 0.5 mm face settlement. Adding creep increased the prediction but remained far from the measurement. Good agreement was obtained only after introducing a construction-induced disturbance zone with strength and stiffness reduced to approximately 50% of the intact values; maximum calculated surface settlement was then 7.3 mm compared with about 10 mm measured.

The lesson is sharper than a generic statement that stoppages are undesirable. An advanced time-dependent constitutive model did not compensate for omission of the disturbed state created by tunnelling and the unusual 100-day loading condition. This case also illustrates the difference between the modelled and executed construction processes: a model assuming continuous advance would omit the 100-day stoppage entirely, even though the stoppage and the associated disturbed ground state controlled the observed response.

The magnitudes in Table 1 should not be interpreted as statistical ranges or typical values. They are project-specific responses to particular construction conditions. The underlying mechanisms, however, are not unusual. For deep excavations, published case histories compiled by Fuentes et al. (2018) report maximum time-dependent wall-movement rates between approximately 0.05 and 1.07 mm/day for different soils, excavation depths and construction methods, encompassing the rates observed at St Botolph Street. In mechanised tunnelling, independent studies similarly identify face pressure, tail-void grouting pressure and grout volume as first-order controls on surface settlement (Chen et al., 2024). The 100-day stoppage considered here is deliberately an extreme case rather than a representative operating condition. The examples therefore demonstrate the possible engineering significance of construction-process effects, rather than defining universal numerical sensitivities.

Table 1: Quantitative evidence of construction-process effects

Construction-process component	Observed or calculated evidence	Magnitude / relative scale	Implication
Temporary geometry (access ramp)	Deep excavation wall moved away from the excavation	Approx. 2 mm; ~30% of the 6-7 mm late-stage wall response	Temporary geometry changes the stress path and later state.
Sequence and timing (prop removal and idle time)	Post-removal wall movement continued before stabilising	Approx. 9 mm over 160 d; comparable to or larger than the 6-7 mm late-stage response	Stage timing is required for realistic response and trigger values.
Geometry (excavation corner)	Corner movement compared with maximum along wall	34.2% of maximum movement	A single plane-strain result is not representative.
Operational parameter and transient state (grout pressure)	TULIP SMRN compared with SMRS	0.2 bar difference; settlement +41%	Operational pressure is a model boundary condition.
Timing and transient state (100-day TBM stoppage)	Measured face settlement compared with Soft Soil prediction	9 mm measured vs 0.5 mm modelled (18x)	Disturbance and stoppage state must be represented.

5 Challenges and opportunities

5.1 Challenges in obtaining model-ready construction information

The first challenge is completeness. Site diaries may identify a date but not the spatial progression of an excavation; TBM systems may record pump pressure but not the circumferential pressure transferred to the ground. The second is temporal resolution. Daily monitoring may miss a response to an operation occurring over minutes, while machine data sampled at high frequency must be reduced without removing relevant events. The third is translation. Prop force, grout pressure, advance speed and steering corrections are measurements of construction systems, not ready-made finite-element boundary conditions. Their mechanical meaning must be established.

The identifiability issue described above is particularly important during model updating. Several independent measurements should therefore be used to distinguish deficiencies in the assumed construction process from deficiencies in ground parameters. Responsibilities must also be clear: the construction team needs to know which information is critical, and the modeller needs timely access to departures of the executed construction process from the process assumed in the analysis.

5.2 Opportunities for construction-aware modelling

The immediate opportunity is therefore not necessarily a more complex model. It is a structured construction record linked to numerical stages: actual excavation surfaces, support activation and removal, stage dates, idle periods, measured loads and relevant operational statistics. A simple model-ready schema could associate each construction event with: a stage or event identifier; start and end times; spatial extent and geometry; excavation, support and interface state; measured loads or operational variables; corresponding monitoring observations; and any recorded deviation from the planned or modelled condition. The planned, modelled, executed and recorded construction processes should therefore be maintained as distinct descriptions of the project. The recorded process provides the evidence available for reconstructing the executed process, but its completeness and spatial and temporal resolution are themselves sources of uncertainty. Differences between planned, modelled and executed conditions should be identified explicitly rather than hidden through subsequent calibration of soil parameters. Sensitivity studies should include construction variables, particularly where measurements show that small operational changes can produce large response changes.

Digital workflows make this increasingly feasible. Automated design-through-analysis methods can transfer tunnel geometry and process information into numerical models (Ninić et al., 2021). Reduced-order models and simulation-based assistants can evaluate face pressure, grout pressure and advance speed rapidly enough to support operational decisions (Zendaki et al., 2024). Monitoring and operational data then become a means of updating the modelled construction state towards the best available representation of the executed state, rather than only providing retrospective validation of the design model. The same limitation applies to data-driven approaches: if temporary stages, stoppages or operational variables are absent from the explanatory data, a learning algorithm cannot distinguish their effects from ground variability or noise and may absorb them into spurious correlations. The appropriate level of sophistication is therefore the simplest model that captures the mechanisms controlling the

decision; a construction-aware semi-analytical or two-dimensional model may be more reliable than a construction-blind three-dimensional calculation.

6 Conclusions

The examples support five conclusions. First, construction information is part of the geotechnical model because it defines the stress path and boundary conditions. Second, advanced constitutive models can reproduce only the processes included in the analysis; they cannot correct omitted temporary geometry, time-dependent stages, operational pressures or construction-induced disturbance. Third, the effects are quantitatively significant. In the excavation example, the approximately 2 mm movement associated with a temporary ramp was about 30% of the late-stage 6-7 mm wall response, while delayed movement following prop removal approached 9 mm. In tunnelling, a 0.2 bar grout-pressure difference coincided with a 41% settlement difference, and omission of the disturbed construction state during a long stoppage led to an 18-fold underprediction of TBM face displacement. Fourth, monitoring and operational records should be aligned with the executed construction process and used to update both ground parameters and construction assumptions. Finally, maintaining planned, modelled, executed and recorded process descriptions as distinct but connected information layers provides a practical route to improve predictions without assuming that increasing numerical complexity alone will improve realism.

References

[1] Almog-Goldreich E. and Fuentes R. (2026): The advanced displacement-confinement method for design of mechanized shield tunnels in soft ground. *Tunnelling and Underground Space Technology*, 174:107668.

[2] Bourgeois E., Berthoz N., Mohamad W., Szymkiewicz F., Le Kouby A., Branque D., Michalski A., Kreziak C. and Soyez L. (2023): Results of a benchmark exercise of prediction of tunnel-pile interaction: the TULIP project. Proceedings of the 10th European Conference on Numerical Methods in Geotechnical Engineering, DOI:10.53243/NUMGE2023-82.

[3] Chen T., Xu G. and Cong Y. (2024): Surface settlement of shallow shield tunneling: face pressure, tail-void grouting pressure, and grout volume. *International Journal of Geomechanics*, 24(6):04024110.

[4] Finno R.J. and Calvello M. (2005): Supported excavations: observational method and inverse modeling. *Journal of Geotechnical and Geoenvironmental Engineering*, 131(7):826-836.

[5] Fuentes R. and Devriendt M. (2010): Ground movements around corners of excavations: empirical calculation method. *Journal of Geotechnical and Geoenvironmental Engineering*, 136(10):1414-1424.

[6] Fuentes R., Pillai A. and Ferreira P. (2018): Lessons learnt from a deep excavation for future application of the observational method. *Journal of Rock Mechanics and Geotechnical Engineering*, 10:468-485.

[7] Kasper T. and Meschke G. (2004): A 3D finite element simulation model for TBM tunnelling in soft ground. *International Journal for Numerical and Analytical Methods in Geomechanics*, 28(14):1441-1460.

[8] Ledesma A., Gens A. and Alonso E.E. (1996): Estimation of parameters in geotechnical backanalysis—I. Maximum likelihood approach. *Computers and Geotechnics*, 18(1):1-27.

[9] Mohammadzamani D., Alimardani Lavasan A. and Wichtmann T. (2023): Tail void grouting material: a parametric study on the role of hydro-mechanical characteristics in mechanized tunneling. *Tunnelling and Underground Space Technology*, 135:105053.

[10] Ninić J., Alsahly A., Vonthron A., Bui H.-G., Koch C., König M. and Meschke G. (2021): From digital models to numerical analysis for mechanised tunnelling: a fully automated design-through-analysis workflow. *Tunnelling and Underground Space Technology*, 107:103622.

[11] Osouli A., Hashash Y.M.A. and Song H. (2010): Interplay between field measurements and soil behavior for capturing supported excavation response. *Journal of Geotechnical and Geoenvironmental Engineering*, 136(1):69-84.

[12] Peck R.B. (1969): Advantages and limitations of the observational method in applied soil mechanics. *Géotechnique*, 19(2):171-187.

[13] Simpson B. (1992): Retaining structures: displacement and design. 32nd Rankine Lecture. *Géotechnique*, 42(4):541-576.

[14] St John H.D., Zdravkovic L. and Potts D.M. (2005): Modelling of a 3D excavation in finite element analysis. *Géotechnique*, 55(7):497-513.

[15] Yao Y., Lu N., Yang Y., Cao Y. and Yu H.-S. (2019): Study of long-termed displacements of a tunnel boring machine during its stoppage. *Tunnelling and Underground Space Technology*, 84:432-439.

[16] Zendaki Y., Cao B.T., Alsahly A., Freitag S. and Meschke G. (2024): A simulation-based software to support the real-time operational parameters selection of tunnel boring machines. *Underground Space*, 14:176-196.

Author

Univ.-Prof. Dr.-Ing. Raul Fuentes
RWTH Aachen University
Chair of Geotechnical Engineering and Institute of Geomechanics and Underground Technology (GUT)
Im Süsterfeld 9
52072 Aachen, Deutschland
Tel.: +49 241 80 25247
e-mail: raul.fuentes@gut.rwth-aachen.de
Web: www.gut.rwth-aachen.de

Generative Design with Plaxis

Jia Lin

Abstract: This paper presents GD Plaxis, a Python-based generative design tool that automates the creation, evaluation, and optimization of PLAXIS 2D models for excavation pits. The tool integrates modules for model generation, numerical analysis, cost and CO_2 assessment, and optimization within a unified workflow. This enables systematic exploration of the design space and identification of efficient and sustainable solutions. Optimization studies demonstrate the applicability of the framework, showing improvements in design efficiency as well as reductions in cost and environmental impact while maintaining required performance criteria. The results highlight the potential of generative design approaches to enhance geotechnical engineering workflows and decision-making.

1 Introduction

Urban excavation projects are becoming increasingly complex due to space constraints, adjacent infrastructure, and stringent safety requirements. The design of excavation support systems, such as diaphragm walls or sheet pile structures, typically relies on advanced numerical methods, with finite element software like PLAXIS widely used in engineering practice. These tools allow for detailed simulation of soil–structure interaction and construction sequences; however, their application in daily engineering workflows remains largely manual and iterative.

Recently, however, optimization techniques have begun to attract attention in geotechnical engineering. For instance, Cerek & Grabe (2023) employed a multi-objective evolutionary algorithm coupled with the PLAXIS API to optimize dike geometry, while Cerek et al. (2024) explored the use of optimization algorithms as decision-making support tools in geotechnical practice. These studies demonstrate the feasibility of integrating numerical simulations with automated search strategies.

Nevertheless, the application of Generative Design—which goes beyond single-objective optimization by systematically generating and evaluating a vast number of design variants under predefined constraints—remains limited in geotechnical engineering, especially when combined with established numerical tools such as PLAXIS.

To bridge this gap, a Python-based tool named GD Plaxis (Generative Design PLAXIS) has been developed. The tool leverages the PLAXIS 2D remote scripting API to automate model generation, analysis, and post-processing, enabling a fully integrated workflow for the design and optimization of excavation pits. GD Plaxis combines multiple functionalities, including

parametric model generation, automated simulation, performance evaluation, and optimization with respect to engineering, economic, and environmental criteria.
The key contribution of this work is the development of a modular and extensible framework that allows for systematic exploration of alternative designs for excavation support systems. In contrast to conventional trial-and-error approaches, the proposed tool enables automated optimization and provides quantitative insights into the influence of design parameters through integrated sensitivity analyses. Additionally, the inclusion of cost and CO_2 assessment modules supports more informed and sustainable decision-making.

2 GD Plaxis Framework Overview

The workflow starts from a predefined PLAXIS model containing the geological profile, material properties, and loading conditions. Design parameters such as retaining structures, anchors, and geometric dimensions can then be modified parametrically. Using the PLAXIS Remote Scripting API, GD Plaxis automatically updates the model, executes the calculations, extracts the results, and evaluates the performance of the design.
In addition to conventional geotechnical criteria such as deformations, internal forces, and safety factors, the framework incorporates cost and CO_2 calculations. These quantities can be used directly as optimization objectives, enabling the identification of economical and environmentally efficient solutions.

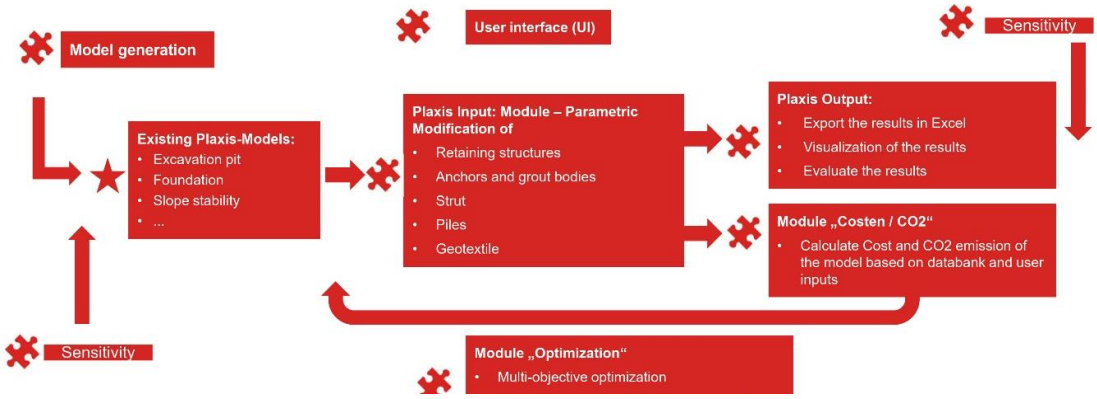


Figure 1: GD Plaxis workflow

3 GD Plaxis Modules

The system consists of seven modules: Connection, Generate, Model, Cost/CO_2, Optimization, Sensitivity, and Report. Together, these modules provide a continuous workflow from model creation and modification to optimization and result documentation. The functionality of each module is described in the following.

3.1 Connection Module

The connection module establishes communication between GD Plaxis and PLAXIS 2D via the Remote Scripting API. The user provides the server port and password to connect to an active PLAXIS session.

3.2 Generate Module

The generate module enables the automated creation of PLAXIS 2D models for excavation pits based on a set of parametric inputs. The process builds upon an existing PLAXIS model, in which soil stratigraphy, material properties, and initial loading conditions are predefined. This ensures consistency while allowing flexible modification of structural and geometrical components.
The user defines the excavation setup through a graphical interface. Key inputs include: Excavation geometry, Slope, Anchors, Retaining structures, Anchor position distribution and Mesh. Based on these inputs, the module automatically generates fully defined PLAXIS models, including geometry, structural elements, and calculation phases.
A key feature of the generate module is its capability to operate in two modes, either single model generation for rapid setup, or iterative generation and evaluation of multiple models.
In the iterative mode, simplified rule-based optimization is applied. Starting from an initial design, the model is evaluated with respect to stability and deformation criteria.
This approach enables the rapid identification of feasible and near-optimal design configurations with minimal user intervention, forming the basis for further detailed optimization within the framework.

3.3 Model Module

The model module provides a direct interface for modifying existing PLAXIS models in a parametric and efficient manner. Once a PLAXIS model is open and a connection is established, GD Plaxis automatically reads the model and identifies all relevant structural elements, including plates, node-to-node anchors, fixed-end anchors, embedded beams, and their associated material properties.
These elements are displayed in a structured overview, allowing the user to adjust key parameters such as anchor angle, position, length, and spacing. Modifications are performed by simply entering new numerical values in the user interface. When the “Update Model” function is executed, the PLAXIS model is automatically updated according to the specified inputs.
This functionality significantly simplifies model modification compared with conventional workflows, where geometry coordinates must be updated manually within the PLAXIS user interface.

3.4 Cost and CO_2 Module

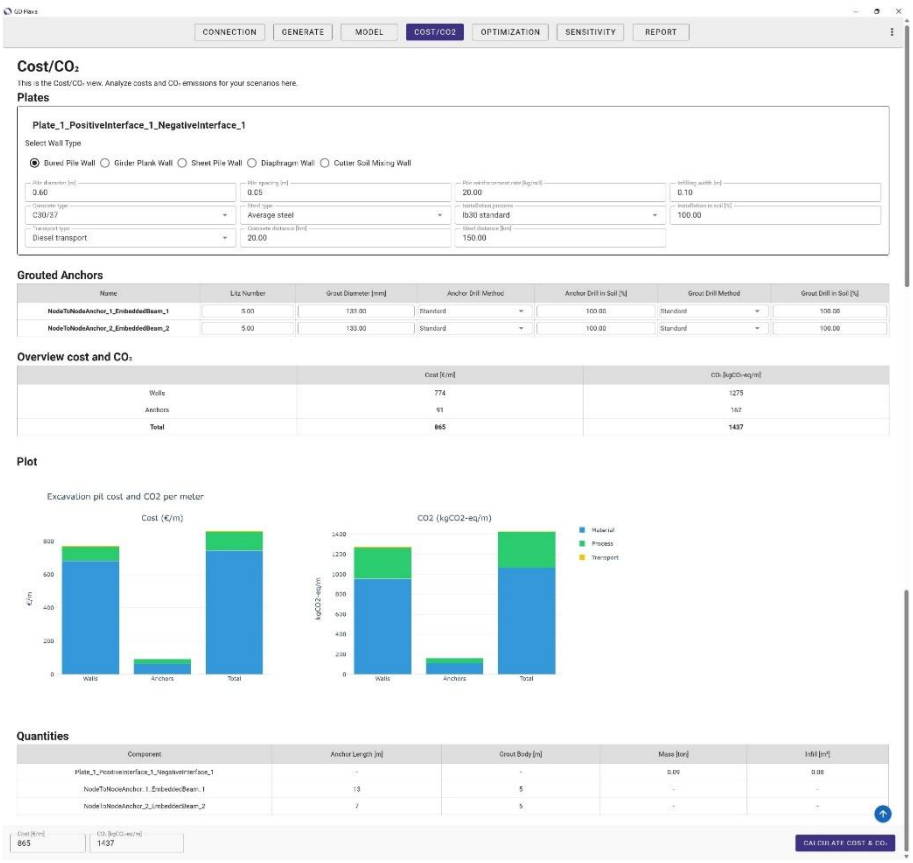


Figure 2: Cost and CO2 module

The cost and CO_2 module enable an approximate quantification of economic and environmental impacts directly from the PLAXIS model. As shown in Figure 2, the user selects structural elements such as plates, anchors, or embedded beams and specifies relevant parameters depending on the selected type. These inputs include geometric properties or weight, material characteristics, installation processes, and transport type and distance, etc.

Based on the provided inputs, GD Plaxis calculates unit values per meter of excavation structure using an internal cost and CO_2 database. The resulting values are subdivided into three categories:

- Material production (A1–A3)
- Transport (A4)
- Construction and installation processes (A5)

in accordance with the life cycle stages defined in DIN EN 17472. The results are aggregated to provide an overall estimate of cost and CO_2 emissions for the modelled system.

It should be emphasized that these values represent simplified estimates derived from generalized input parameters and database assumptions. Consequently, they are not intended for detailed cost estimation or formal life cycle assessment in real projects. Instead, the module provides a consistent and efficient basis for comparing design alternatives.

By integrating cost and environmental metrics into the modelling workflow, this module enables their direct use as objective functions in the subsequent optimization process and supports more informed and sustainability-oriented design decisions.

3.5 Optimization Module

The optimization module represents the core component of the GD Plaxis framework by linking parametric model definition, numerical analysis, and performance evaluation within an automated optimization process. It enables the systematic improvement of excavation designs with respect to technical, economic, and environmental criteria.



Figure 3: Optimization module

Two optimization approaches are implemented:

- Brute-force method: This method evaluates all possible combinations of the selected design parameters using user-defined parameter ranges and step sizes. Because the number of combinations increases rapidly with the number of variables, this approach is primarily intended for relatively small design spaces.

- Surrogate-based Bayesian optimization: For problems with a larger number of continuous design variables, GD Plaxis employs Bayesian optimization implemented using the Ax platform and its BoTorch modular optimization framework. Rather than evaluating the complete design space, a statistical surrogate representation based on the previously calculated designs is used to guide the selection of subsequent candidate solutions. This substantially reduces the number of computationally expensive PLAXIS analyses required to identify promising regions of the design space.

For the surrogate-based optimization, each selected design variable is defined as a continuous bounded parameter using the minimum and maximum values specified by the user. The initial exploration of the design space is performed using a Sobol low-discrepancy sequence. The number of initial Sobol trials is defined as 20% of the prescribed maximum number of optimization iterations, with a minimum of five initial trials. After completion of these initial evaluations, the optimization strategy switches automatically to the BoTorch modular Bayesian optimization generator. In the present implementation, no specific surrogate architecture is prescribed manually; instead, surrogate construction and model-based candidate generation are handled internally by the Ax/BoTorch framework. This approach was selected because each function evaluation requires a nonlinear finite-element analysis in PLAXIS, and Bayesian optimization is therefore well suited to reducing the number of costly numerical simulations compared with exhaustive exploration of the design space.

Within the optimization module (Figure 3), the user defines optimization objectives, constraints, and design variables. Available performance criteria include cost and CO_2 emissions, safety factors, structural response quantities such as bending moments and axial forces, and deformations. The design variables may include anchor inclination, length, vertical position and spacing, embedded-beam properties, or plate dimensions. For each selected variable, admissible lower and upper bounds are specified.

Once the optimization is initiated, GD Plaxis generates a candidate parameter set, automatically updates the corresponding PLAXIS model, executes the numerical calculation, extracts the required results, and evaluates the resulting design. The calculated quantities are subsequently returned to the optimization algorithm and used to determine the next candidate design. This creates a sequential feedback loop in which information obtained from previous PLAXIS analyses is used to focus subsequent evaluations on increasingly promising regions of the parameter space.

For multi-objective problems, the objectives are treated simultaneously rather than being combined into a single weighted objective function. The resulting designs are evaluated according to Pareto optimality. A solution is Pareto-optimal when improvement of one objective would require deterioration of at least one other objective. The final output therefore consists of a set of non-dominated solutions representing the trade-offs between competing objectives, for example construction cost and structural forces. This Pareto set allows the engineer to select an appropriate design according to project-specific priorities rather than relying on a single predefined weighting between objectives.

This combination of initial space-filling sampling, surrogate-assisted Bayesian optimization, automated finite-element evaluation, and Pareto-based multi-objective assessment enables efficient exploration of complex excavation design spaces while limiting the number of computationally expensive PLAXIS simulations.

3.6 Sensitivity Module

The sensitivity module is a supplementary component developed to address specific user requirements for parameter sensitivity studies in PLAXIS models.
The user selects the parameters to be analyzed (e.g., strength or stiffness parameters) and defines corresponding minimum and maximum values. Based on these inputs, GD Plaxis automatically generates two modified models: one using all minimum parameter values and one using all maximum values.
Both models are calculated sequentially, and the resulting outputs are compared with respect to relevant performance indicators such as deformations, internal forces, or safety factors. This enables a rapid assessment of the potential variation range of results due to parameter uncertainty.

3.7 Report Module

The report module facilitates efficient extraction, visualization, and documentation of numerical results from PLAXIS Output. Once connected to an active PLAXIS 2D Output session, GD Plaxis reads the model results automatically when the user initiates the “Get Results” function.
The module allows the user to evaluate structural responses of selected elements, particularly plates. Results can be visualized for chosen construction phases, for example in the form of bending moment distributions along the wall. In addition, the module identifies and lists maximum values of key response quantities (e.g., bending moments, shear forces, axial forces, and displacements) together with the corresponding phase in which they occur.

4 Application Examples

To demonstrate the capabilities of GD Plaxis, two excavation projects with different retaining systems and geological conditions were optimized. In both cases, the surrogate-based multi-objective optimization algorithm was applied. The optimization objectives were the minimization of construction cost and CO_2 emissions, while compliance with structural constraints was maintained.
The design variables considered in both studies included:

- Anchor inclination angle
- Anchor free length
- Grout body length
- Anchor position
- Wall embedment length

The following constraints were imposed:

- Maximum wall bending moment
- Maximum wall shear force
- Maximum wall displacement
- Maximum anchor normal force

For each design iteration, GD Plaxis automatically updated the PLAXIS model, executed the numerical analysis, extracted the results, and evaluated compliance with all optimization constraints. Candidate solutions were generated iteratively until a Pareto-optimal set of feasible designs was obtained.

4.1 Example 1 Sheet pile wall with one anchor level

The first example consists of a 5.9 m deep excavation supported by a sheet pile wall with one level of grouted anchors. The subsurface profile comprises clay, gravel, sand, and weathered bedrock. Table 1 summarizes the optimization results.

Table 1: Example1 optimization results

Parameters	Reference design	Optimized design	Reduction
Cost [€/m]	2115	1926	8.9%
CO_2 [$kgCO_2$-eq/m]	943	869	7.8%

The optimization achieved a reduction of approximately 9 % in construction cost and 8 % in CO_2 emissions while satisfying all prescribed performance criteria.

4.2 Example 2: Bored pile wall with two anchor levels

The second example consists of an 8.2 m deep excavation supported by a bored pile wall with two levels of grouted anchors. The geological profile includes weakly weathered rock, highly weathered rock, and loose soil deposits.

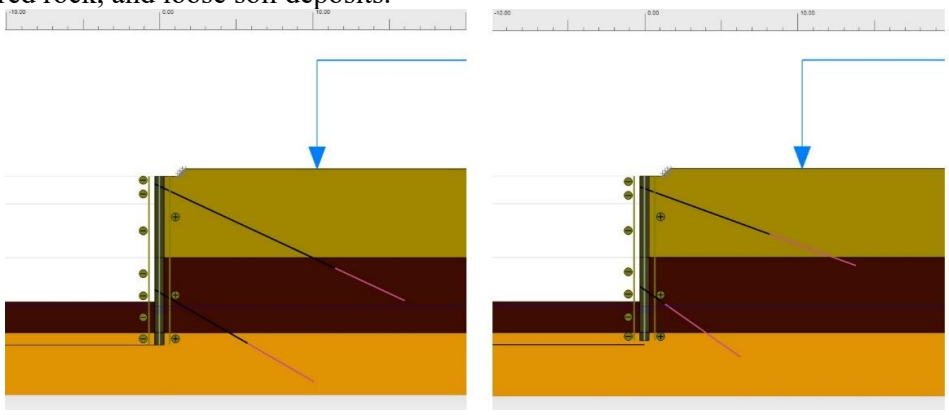


Figure 3: Original and optimized PLAXIS model

Table 2: Example2 optimization results

Parameters	Reference design	Optimized design	Reduction
Cost [€/m]	865	805	6.9%
CO_2 [$kgCO_2$-eq/m]	1437	1329	7.5%

The optimized solution reduced construction costs by approximately 7 % and CO_2 emissions by approximately 8 % while fulfilling all structural constraints.

4.3 Comparison of Results

Both examples show consistent reductions in cost and environmental impact while maintaining compliance with all structural requirements. The results indicate that optimization potential exists for different retaining systems and geological conditions. Furthermore, the studies demonstrate the ability of GD Plaxis to automatically evaluate a large number of design alternatives and identify efficient solutions that would be difficult to obtain through manual trial-and-error procedures.

5 Conclusions

This paper presented GD Plaxis, a Python-based generative design framework developed for the automated generation, evaluation, and optimization of excavation support systems in PLAXIS 2D. By integrating model generation, parametric model modification, numerical analysis, cost and CO_2 assessment, optimization, sensitivity analysis, and reporting within a unified workflow, the tool significantly reduces manual modelling effort and facilitates systematic exploration of design alternatives.

Two optimization examples involving sheet pile and bored pile wall systems demonstrated the practical applicability of the framework. In both cases, GD Plaxis successfully identified solutions with lower construction costs and reduced CO_2 emissions while satisfying all prescribed structural constraints. The results show that optimization can reveal design improvements that may not be readily identified through conventional manual trial-and-error procedures.

The developed framework combines established finite element modelling techniques with modern generative design and optimization methods, enabling more efficient and sustainability-oriented decision-making in geotechnical engineering. Future developments will focus on extending the framework to additional geotechnical applications, enhancing the optimization algorithms, and improving the accuracy of the integrated cost and environmental assessment models.

Acknowledgements

The author would like to express sincere gratitude to Jelle Van Den Berg for the programming and software development work that made the GD Plaxis framework possible.
Special thanks are extended to Annerose Bierer, Philipp Dörflinger, and Jacob Wegerer for their support in managing and guiding the project.
The author also gratefully acknowledges the valuable professional advice and technical discussions provided by colleagues Abdulaziz Abdelkadr, Haydar Al-Majadi, Gebreselassie Berhane, Sandor Fesues, Gisa Kleine Vennekate, Claudia Klotz, Andrej Mey, Attila Sarosi, Jörg Schreiber, Barbara Spuler, Marijn De Volder, Thomas Wieser, Nikoletta Zsiros, and Tommy Zottl. Their expertise and feedback contributed significantly to the development and validation of the tool.
The support and collaboration of all contributors are greatly appreciated.

References

[1] Bentley Systems, 2024. PLAXIS 2D Manuals (2024.2): Reference, Material Models, Scientific and Tutorial Manuals. [Online]
[2] Jürgens, H. and Henke, S. (2024). Numerical Optimisation of Excavation Pit Design Using Finite Element Analyses. *Geotech Geol Eng* 42:1659–1673
[3] Martins J. R. R. A. and Ning A. (2021): Engineering Design Optimization. Cambridge University Press, Cambridge
[4] DIN EN 17472:2022-08. Sustainability of construction works – Sustainability assessment of civil engineering works – Calculation methods. German Version EN 17472:2022. Beuth Verlag, Berlin.
[5] Cerek K. and Grabe J. (2023): Numerical simulation and optimization of dike geometry using multi-objective evolutionary algorithm NSGA-II. In: Proceedings of 10th European Conference on Numerical Methods in Geotechnical Engineering 2023 in London/UK
[6] Cerek K., Hadjiloo E., Grabe J. (2024): Application of optimization algorithms in geotechnical engineering as decision-making support tool., in: Proceedings of the XVIII European Conference on Soil Mechanics and Geotechnical Engineering (ECSMGE), Lisbon, Portugal
[7] Hvarfner C., Hellsten E. O., and Nardi L. (2024) Vanilla Bayesian Optimization Performs Great in High Dimensions, Proceedings of the 41st International Conference on Machine Learning (ICML)

Author

Dr. nat. techn. Jia Lin
Strabag AG
Zentrale Technick, Technische Büro Wien
Donau-City-Strasse 9
1220 Vienna
Tel.: +43 (0) 1 - 22422 / 1828
e-mail: jia.lin@strabag.com

Saving Time and Money with Mathematical Optimisation - Applied Optimisation in Geotechnical Engineering

Jörg Meier

Abstract:

When designing geotechnical structures, engineers have a certain amount of leeway. In the case of an excavation pit, for example, depth and contour of the volume to be excavated are fixed, whereas the design of the shoring is left to the engineer. A skillful choice of the design parameters often has a positive impact on cost, construction time and environmental footprint. Optimisation is therefore always part of the planning task. However, due to the numerous dependencies and non-linearities involved in these problems, manual optimisation is usually difficult and time-consuming. This article showcases how automated mathematical optimisation can be applied to real-world geotechnical problems.

1 Motivation

Designing more sophisticated geotechnical structures can quickly become complex. Verifying stability and serviceability of those structures is usually achieved using numerical models (e.g. see fig. 1), making this kind of models ubiquitous in geotechnical practice. Depending on the requirements of the structure or the task at hand, simple lamella methods, 2D strut and tie models, up to coupled time-dependent 3D finite element (FE) simulations are employed. To 'optimise a geotechnical structure' must therefore be aligned to this set of geotechnical design tools.

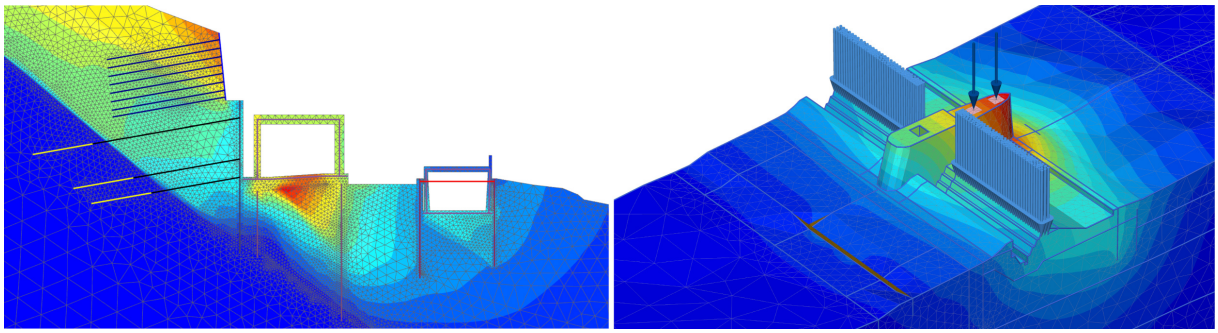


Figure 1: Numerical models of geotechnical structures; left: excavation pit of a cut and cover tunnel (2D FE model), right: part of the Geesthacht barrage near Hamburg (3D slice FE model)
source: Gruner AG, by author

The main challenge in manually optimising geotechnical structures is typically the large number of parameters, their complex interdependencies, and the wall time required for evaluation. Common characteristics of geotechnical numerical models are outlined in section 3. As individual parameters often cannot be optimised independently, manual optimisation quickly becomes impractical, with the engineer losing oversight or requiring excessive time to reach a solution. Consequently, manual optimisation is often feasible only to a limited extent. Mathematical optimisation, which is largely automated, provides an effective solution.

For geotechnical structures, the objective is usually to minimize financial cost (see section 2). As discussed in section 3, the technical characteristics of the relevant models are often not well suited for optimisation and vary significantly depending on the specific project. Section 4 provides brief notes on the optimisation process itself, while section 5 presents two illustrative examples.

2 Optimisation Objective

In the authors' experience, the primary objective of optimisation in geotechnical practice is usually the minimisation of total financial cost. This is mainly because financial optimisation aligns with the client's interests and, in practice, all aspects of a geotechnical structure can be expressed in terms of a (direct or indirect) monetary figure. This applies, for example, to the tonne of steel, the CO_2 equivalent and the construction time - all to be minimized. Therefore, throughout this article, financial cost functions are used as objective functions.
For a given set of design parameter values, a numerical model is built and evaluated (forward calculation). The results are used to estimate total financial cost, which serve as the optimisation criterion.

In this context, an interesting and beneficial effect emerges for many optimisation tasks in geotechnical engineering: if the number of components can be reduced, this typically leads to a simultaneous reduction in costs, construction time, *and* ecological footprint. This effect eliminates the need for multi-criteria optimisation (e.g. Pareto optimisation) in geotechnical engineering.

In addition to standard costs, such as those for materials and labour, the objective function may incorporate various penalty terms to avoid undesired solutions (e.g. overlapping anchors or excessive deformation of the excavation wall). For consistency within the objective function, these penalty terms are also represented as monetary values.

3 Geotechnical Models from an Optimisation Perspective

Although a vast variety of numerical models are used in geotechnical practice, certain common characteristics can be identified which are of interest from an optimisation perspective and are frequently encountered. These characteristics include:

Reproducibility: The same input parameters produce the same result when the calculation is performed again. This characteristic significantly facilitates verification.

Black box, closed source: Although the theoretical principles underlying most software systems used in geotechnical engineering are well known, the specific implementation - including the handling of special cases - is often not disclosed. The numerical models based on these systems must therefore be regarded as 'black boxes'.

Runtime: Apart from calculations in spreadsheet applications, which are usually completed in fractions of a second, most numerical models have runtimes in the order of a few minutes; larger 3D FE models can take up to several hours.

Interdependencies between parameters: Design parameters often influence one another and may have non-linear effects and dependencies.

Invalid parameter combinations: Not all combinations of design parameters yield valid calculation results (e.g. if load-bearing capacity of a component is insufficient). Many numerical methods are unable to complete calculation for such invalid parameter combinations and terminate with an error message.

Numerical stability / Numerical noise: Beside physically justified jumps in the system response (e.g. jump in the external load-bearing capacity for anchors in different soil layers), physically unjustifiable jumps in the system response also occur. A typical pattern of behaviour is the seemingly chaotic alternation between two or more different system responses for almost identical parameter values. Reasons for this behavior are manifold and may lie in the underlying numerical concept (e.g. mesh dependency), in the termination criteria used in iterative algorithms, or in locally incompatible transitions between special cases.

Number of design parameters: For geotechnical models, the number of parameters whose values are to be optimised often ranges from 5 to 20 (but may be higher for complex models). This observation already takes into account the fact that certain parameters are defined by the project boundary conditions and do not need to be varied.

To illustrate the characteristics 'numerical noise' and 'invalid parameter combinations' mentioned above, figure 2 shows, for the excavation shoring described in section 5.1 in the variant with 4 anchor layers, how the objective function value $f(x)$ (without penalty terms) depends on the position of 'AL 2'. The values of all other model parameters were fixed. Both, DC-Pit (beam analysis model) and Plaxis 2D (FE model) were used as forward calculations. Various effects are clearly visible in figure 2.

Magnitudes of $f(x)$: Magnitudes of the objective function values $f(x)$ differ significantly between DC-Pit and Plaxis 2D. This is primarily due to different calculation assumptions, such as the application of a minimum earth pressure in DC-Pit.

Objective function shapes, location of the optima: In some sections of the graphs, DC-Pit and Plaxis show the expected continuous and thus 'good-natured' dependence of $f(x)$ on the parameter varied (marked with ①). But, due to the different computational approaches used, different system behaviours emerge, leading to variations in the shape of $f(x)$ - even

though they describe the same physical system. This goes to the extend that the (calculated) minimum differs: for DC-Pit, the minimum corresponds to a position of AL2 of ∼12.4 m; for Plaxis, ∼11.3 m (each marked with ⓪).

Alternating between solutions: In some sections, both forward calculations show a back-and-forth oscillation ('alternation', marked with ②) between two or more solutions, which in this example appears to be numerically, but not physically, justified.

Numerical noise: Similar to the alternation between solutions, a degree of 'noise' in $f(x)$ sometimes occurs. In some cases, this noise is clearly visible, e.g. in the area marked with ③. However, due to the finite precision used in the numerical models, slight noise is present everywhere, even in the 'good-natured' regions, e.g. marked with ①. Other common causes of numerical noise include mesh-dependency and convergence criteria.

Jumps: Often attributable to code paths that are not fully compatible with one another (e.g. for special cases), jumps of varying magnitudes may occur, as indicated by ④.

Divergent objective function values: Often in the vicinity of failed forward calculations due to numerical problems, (non-physical) local extrema can be observed, as marked with ⑤.

Failing forward calculations: Both programmes used show failed forward calculations, some of which occur in clusters (marked by vertical lines). But the location of the clusters differs between DC-Pit and Plaxis. Given the range of variation used for the parameter varied, these have no physical basis but are likely attributable to purely numerical problems.

Apart from the first point (reproducibility), the remaining points listed hinder optimisation. Given these constraints, manual or (automated) mathematical optimisation can only be carried out through 'directed trial and error': certain combinations of parameters are tested, the resulting system responses are analysed, and new combinations of parameters are selected for a further optimisation based on this knowledge. This iterative process is repeated until either no further (meaningful) improvement can be achieved, or until the solution is otherwise deemed 'sufficient'.

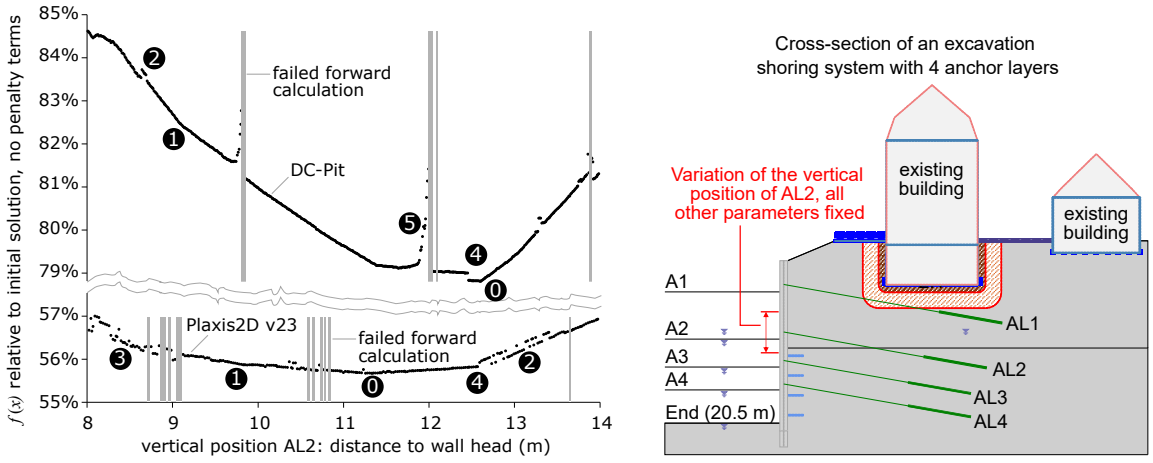


Figure 2: Dependence of the objective function value on the position of 'AL 2' (section 5.1)
(fixed parameters are not optimal; therefore, ⓪ does not correspond to the optimum from section 5.1)
source: Gruner AG, by author

4 Optimisation Algorithms for Geotechnical Problems

The field of applied mathematics offers a class of algorithms for the 'black box optimisation' problem outlined above, of which a subset is capable of handling the characteristics described in section 3. A description of these algorithms would go far beyond the scope of this article; therefore, readers are referred to corresponding literature (e.g. Meier, 2008; Gens et al., 1996).

It should be noted that this approach offers no guarantee that the exact 'global optimum' can be found (nor that such an optimum even exists). From a geotechnical practice perspective, however, this is generally neither required nor necessarily desirable. Rather, in geotechnical optimisation tasks, it is also important that the chosen design exhibits a certain degree of robustness with regard to the design parameters. If, for example, the anchor point or angle of a strand anchor needs to be adjusted slightly during construction, the system must be able to tolerate this without issue. This would not be the case, for example, if a strand anchor were positioned exactly between two closely spaced pipes.

To avoid hindering the project workflow, the result of an optimisation must be available within a reasonable timeframe. In many projects, this acceptable timeframe ranges from a few days to a few weeks - provided that the manual workload required are significantly smaller. To meet this timeframe, assuming a wall-time of $\sim$3 to 5 minutes per forward calculation and without parallelisation, an optimisation algorithm is required to converge within 100 to 1000 forward calculations.

5 Examples

5.1 Back-anchored Excavation Shoring

The Basel University Hospital (USB) is continuously developing its facilities in the centre of Basel, strengthening healthcare provision, clinical teaching and research. As part of this development, the existing clinical building K2, which dates from the 1970s, is to be replaced by a new building. Figure 3 provides an overview of the project site.

For the construction of the new building, an excavation pit up to $\sim$26.5 m in depth is required directly adjacent to the existing hospital buildings. Simplified, the ground below a thin layer of artificial fill consists of Lower gravel (NTS, down to approx. 13 m below ground level) and, beneath that, Meletta beds. For the design of the excavation pit, the groundwater level is to be assumed $\sim$11 m below ground level.

Over large sections, the associated excavation shoring consists of a back-anchored bored pile wall. As part of a specialist consultancy commissioned by the client, various optimisation measures were proposed for the excavation shoring of this project. These measures included conventional and technical optimisations as well as the use of mathematical optimisation methods.



Figure 3: Overview of the project perimeter
source: Google, GeoBasis-DE/BKG (left); Gruner AG, by author (right)

5.1.1 Conventional or Technical Optimisation Measures

Although conventional or technical optimisation measures are not the focus of this article, these should be briefly mentioned here:

Anchor bearing capacity: First pull-out tests carried out by the project author on vertical anchors yielded surprisingly low external load-bearing capacities R_{ak}. By adjusting construction parameters for repeated tests, the R_{ak} values could be more than doubled leading to more than 1800 kN in the NTS, and 1680 kN in the Meletta beds (8 m long grouting sections; borehole diameter of 178 mm).

Relief drillings: To reduce the water pressure acting on the bored pile wall, relief drillings have been utilized in the low-permeability Meletta beds. Details on this concept can be found in Moormann (2006) and Steiger & Pitteloud (2001). Furthermore, vertical relief drillings were arranged within the excavation pit to prevent hydraulic heave. The embedment length of the retaining wall is thus decoupled from the hydraulic heave.

Shortened primary piles: Since the wall length is no longer dependent on the hydraulic heave and given the low permeability of the Meletta beds, the length of the (unreinforced) primary piles could be reduced to just below the bottom of the excavation pit, since the (reinforced) secondary piles alone are structurally sufficient.

5.1.2 Mathematical Optimisation Measures

Figure 4 shows a cross-section through the southern part of the 20.5 m deep excavation pit and the existing building of the 'Markgräflerhof' in the immediate vicinity, which is of particular heritage value. An overlapping, back anchored bored pile wall is planned to act as the supporting structure. In this case, strand anchors are to be employed. The original project set-up was designed by the project author with DC-Pit (structural design) and Plaxis 2D (estimation of deformation field).

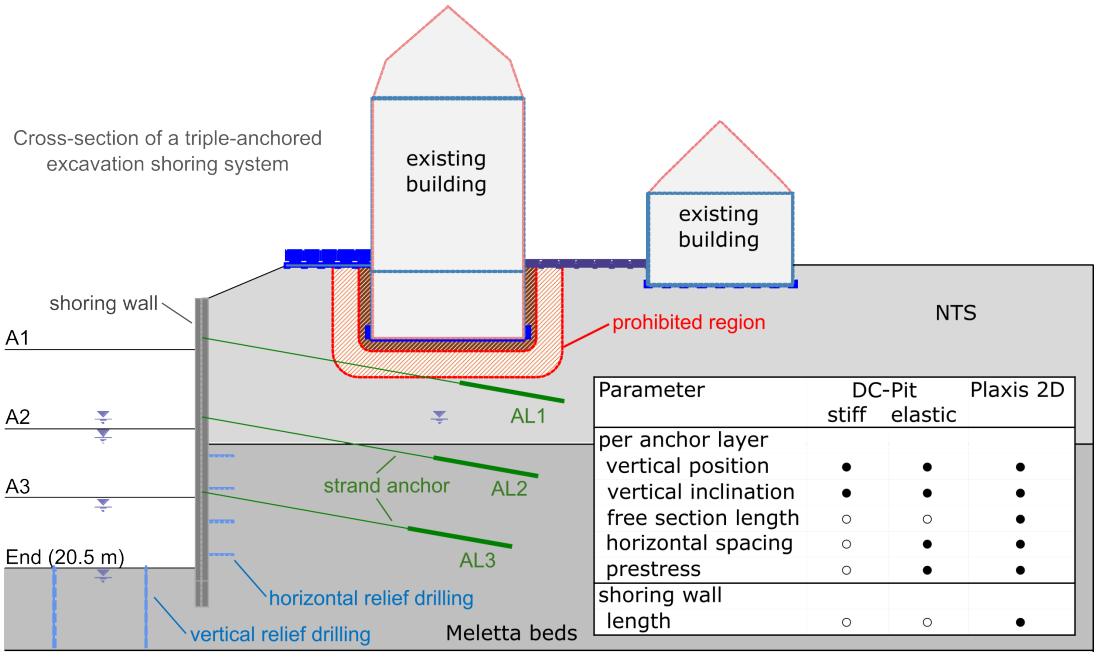


Figure 4: Cross-section of the back-anchored excavation shoring, including associated parameters (○ Parameters with no influence or calculation results; ● variable parameters)
source: Gruner AG, by author

The number of anchor layers, their respective parameters (vertical position, vertical inclination, horizontal spacing, etc.) and certain parameters of the bored pile wall (length of the secondary piles and reinforcement content) are - within certain limits - subject to optimisation; experience shows that these have a significant influence on costs.

The optimisation approach follows Meier (2019) and extends it by incorporating a Plaxis 2D FE model alongside DC-Pit. However, this article does not assess whether FE models can be reliably used for design verification according to the standards. Exemplary, one issue is the treatment of minimum earth pressure, which is required in conventional design but is not enforced in FE models. As a result, FE analyses can produce smaller internal forces than traditional methods under otherwise identical conditions. Therefore, the FE-based optimisation results are seen as numerical experiment and were not used for final design here.

For optimisation with DC-Pit, the project author's model was adopted directly. Using the 'elastic' option to model anchors, horizontal spacing and prestress become optimisation variables (see table in figure 4).

In the FE model (Plaxis 2D, plane-strain plastic analysis, steady-state flow), all material parameters and loads were used with characteristic (unfactored) values to ensure a realistic stress-strain field and initial state. To ensure an adequate margin to failure, a $\phi - c$ reduction was performed in the final state, with a minimum required factor of safety (FOS) of 1.4. In the FE model, free section length and wall length are not a calculation result and must be specified as input parameters (see table in fig. 4), leading to 5 parameters per anchor layer plus the wall length.

In accordance with the considerations of section 2, a cost function including various penalty terms (e.g. overlapping anchors, deformation of the excavation wall) was used as the objective function, which automatically estimates the cost per metre. 'Invariant' cost components, which are the same for all parameter combinations or vary only minimally, are not taken into account.

The objective function values are given as a percentage relative to the initial design prior to optimisation. Therefore, 100 % corresponds to the initial design with 8 anchor layers using much lower R_{ak} values, prior to the repeated anchor pull-out tests.

Several optimisation algorithms were used for optimisation for both, DC-Pit and Plaxis 2D, including PSO (Kennedy & Eberhardt, 1995) and ADEQFS (Yin et al., 2012). Since, for all optimisation algorithms used, the number of parameters to be optimised must remain constant during an optimisation run, several runs were carried out with different numbers of anchor layers.

Due to the optimisation algorithms used, parameters are employed as floating point values. After a plausibility check, these values must be rounded for practical use, leading to slightly suboptimal solutions. This aligns with the aim outlined at in section 4 of identifying robust parameter sets that remain effective despite minor adjustments.

Figure 5 shows the configurations for the excavation shoring found using PSO for 5, 4 and 3 anchor layers, both for DC-Pit and Plaxis 2D, derived on a standard PC (Intel i9, 11th gen, 3.50 GHz). To counteract convergence at local optima, several runs were carried out in each case and the result from the best run was selected. The configurations were found after approximately 600 to 1500 forward calculations, with the optimisation being terminated if no significant improvement in $f(x)$ could be achieved within the last 10 cycles. With forward calculations under one minute, DC-Pit optimisations typically finished within a day. In contrast, Plaxis 2D simulations required nearly 10 minutes each cumulating to more than a week in wall-time.

While for the optimisation runs penalty terms have been used, the costs in figure 5 (shown as percentages relative to the project author's initial model) do not include penalty costs. From figure 5, the following statements can be derived:

- Except for the case using DC-Pit with 3 anchor layers - which is structurally and geotechnically unsuitable - both approaches yield initially plausible results.
- The lack of a solution with DC-Pit with 3 anchor layers is due to the application of a minimum earth pressure (in the FE model, the earth pressure is a calculation result and does not reach the 'minimum earth pressure' to be applied in DC-Pit).
- As expected, anchors are oriented as horizontally as possible to maximise load transfer.
- The FE solution with 5 anchor layers shows a large spacing in the third layer, indicating it is unnecessary from an optimisation standpoint. Since the number of variables cannot change during a run, the layer remains but with minimal contribution. This confirms the need to perform multiple optimisations with varying numbers of anchor layers.
- The optimal parameter values obtained clearly differ for DC-Pit and Plaxis (e.g. the vertical locations of the anchor layers). This result corresponds to the behaviour observed in figure 2 attributed to the varying numerical approaches used.

- DC-Pit solutions are significantly more expensive than those from Plaxis 2D. This is mainly because the FE model uses characteristic values (but satisfies the FOS) and omits certain loads (e.g. minimum earth pressure), resulting in lower internal forces and thus reduced costs.
- Within the different number of anchor layer configurations for Plaxis 2D one can observe that the costs stay close to 53-54%. The same applies to DC-Pit with costs fluctuating between 75-77%. This is because fewer anchor layers lead to smaller horizontal spacing, keeping the total number of anchors nearly constant but redistributed. Additionally, excavation stage costs were not included in the objective function, so fewer stages do not reduce total costs.

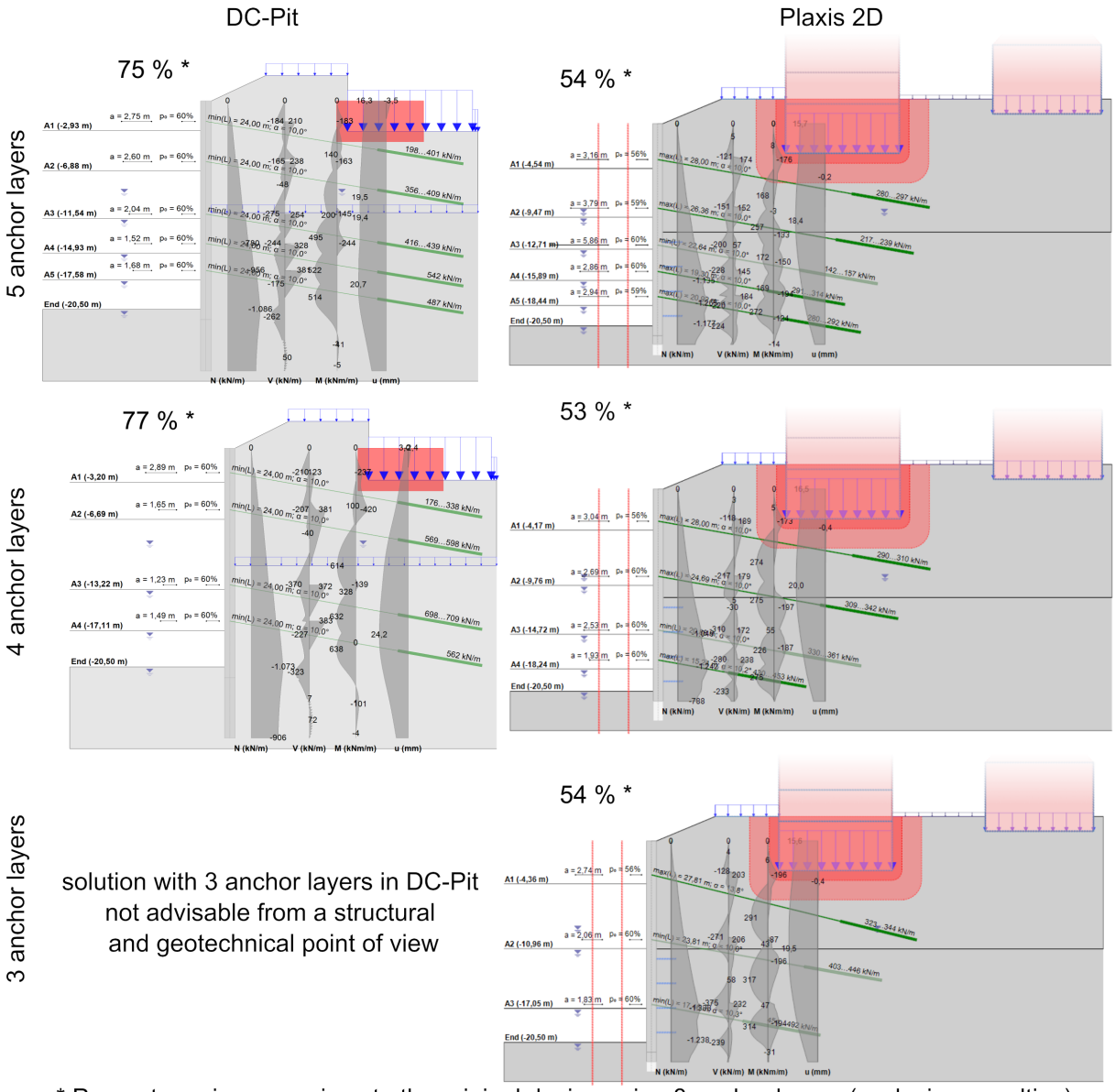


Figure 5: Optimal configurations for 5, 4 and 3 anchor rows with DC-Pit and Plaxis 2D
source: Gruner AG, by author

Compared to traditional manual optimisation, the automated optimisation approach adopted here demonstrated significant savings, thereby delivering clear advantages in terms of costs, deadlines and resources. For comparable excavation pit optimisations, cost savings of approximately 10 to 25% are frequently achieved.

5.2 Inner Lining of the Second Tube of the Lötschberg Base Tunnel

To realize the integrated synchronized timetables in long-distance traffic and to increase the capacity in freight traffic between Brig and Bern, the two-lane extension of the Lötschberg Base Tunnel (LBT) is necessary. This demands the installation of an inner lining over 14 km in the second LBT tube, which is currently secured with shotcrete. Considering the available excavation profile, a skillful selection of the sequence of formwork carriage geometries (FCGs) can lead to a reduction in the required concrete volume.

As a 'proof of concept', the use of mathematical optimisation algorithms for the automated determination of a cost-optimised sequence of formwork carriage geometries was investigated between km 31.500 and km 32.000. The basis for this is the reduced point cloud from a laser scan (figure 6) and geometric specifications for the formwork carriage geometry. Furthermore, it is specified that the adjustment of the formwork carriage is to be carried out exclusively by the arrangement of a horizontal insert at the apex (figure 7). Other variations (e.g. the inclination of the lateral formwork elements) were not desired by the client. A comprehensive description of this cost optimisation can be found in Meier et al. (2021).

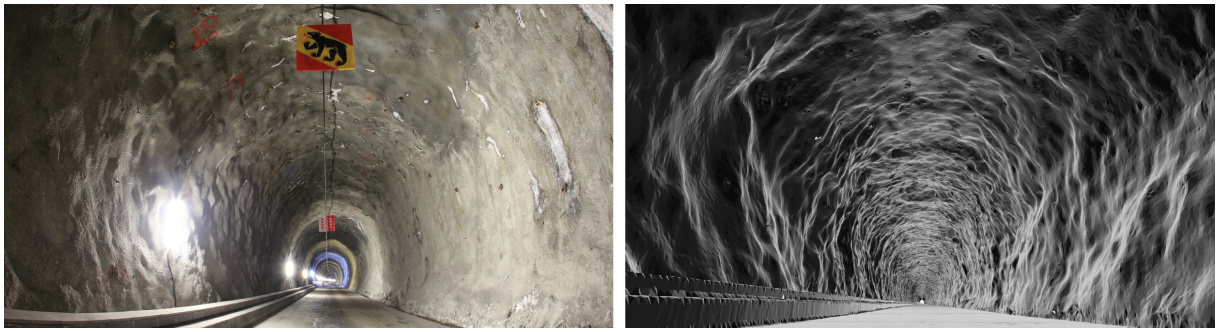


Figure 6: Photo taken inside the tunnel (left) and visualisation based on the 3D scan (right)
source: BLS Netz AG (left), Gruner AG (right)

In accordance with the terms of reference, it was assumed that the structural analysis of the tunnel lining was valid throughout the range of inset widths permitted within the scope of the optimisation. Accordingly, the forward calculation used determines exclusively geometric parameters, as shown in figure 7 (e.g. volume of the overlying concrete, volume of rock projections to be tapered and excavation surfaces requiring new support). Based on these geometric parameters and realistic unit prices, the costs can be calculated using a cost function $f(x)$.

A particle swarm optimiser (PSO; Eberhardt & Kennedy, 1995) with 10 particles was used as the optimisation algorithm. To counteract the PSO's tendency to converge to local optima,

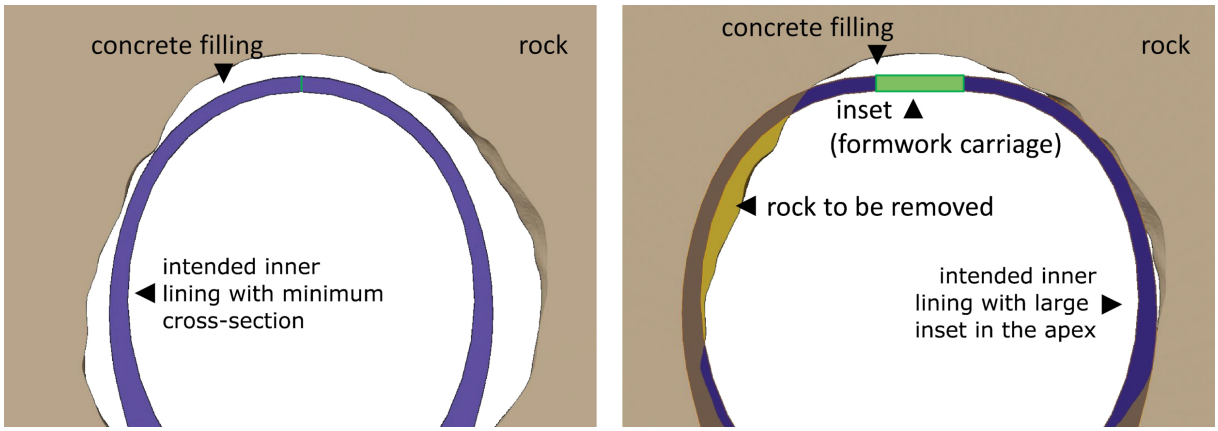


Figure 7: Cross-section with two extreme values for the width of the insert (source: by author)

several optimisation runs were carried out using different combinations of initial parameters, and the most favourable resulting solution was selected. Furthermore, an ADEQFS optimisation algorithm (Yin et al., 2012) with analogous approaches was used.

Neither PSO nor ADEQFS support adjusting the number of parameters during an optimisation run. However, the number of parameters depends on how often the formwork carriage is reconfigured: for each configuration, the width of the inset and the length of the configuration must be defined as parameters. Several optimisation runs were therefore carried out here with different numbers of reconfigurations.

Compared to traditional manual optimisation, the automated optimisation approach used here yielded a saving of approximately 1000 m^3 of concrete per 500 m of tunnel, leading to significant advantages in terms of costs, schedules and resources.

6 Summary

The objective of optimisation is generally to minimise costs - conveniently this usually coincides with structural, construction-time and environmental optimisation. However, manual optimisation is often only possible to a limited extent. The integration of automated mathematical optimisation algorithms offers an interesting alternative here, which is outlined in this paper and illustrated for two specific problems.

The list of various practical, structural and mathematical optimisation approaches in section 5.1 underscores the following recommendation: Geotechnical projects should be optimised in their entirety. A powerful tool for this is mathematical optimisation. However, this tool can only realise its full potential in conjunction with a deep understanding of the underlying numerical models, experience and geotechnical expertise.

References

[1] Eberhardt R. C., Kennedy J (1995): A new optimizer using particle swarm theory. *Proc. 6. International Symp. on Micro Machine and Human Science*, pp. 39-43, Nagoya, Japan.

[2] Gens A., Ledesma A., Alonso E. E. (1996): Estimation of Parameters in Geotechnical Backanalysis - II. Application to a Tunnel Excavation Problem. *Computers and Geotechnics*, Elsevier Science, Vol. 18, No. 1, 29-46

[3] Kennedy J., Eberhardt R. C. (1995): Particle Swarm Optimization. *Proc. of IEEE International Conf. on Neural Networks*, Vol. IV, pp. 1942-1948, Perth, Australia.

[4] Meier J. (2008): Parameterbestimmung mittels inverser Verfahren für geotechnische Problemstellungen. *PhD thesis*, Bauhaus-Universität Weimar.

[5] Meier J. (2019): Optimierung von Baugrubenabschlüssen. *26. Darmstädter Geot.-Koll.*.

[6] Meier J., Pitteloud L., Irngartinger S. (2021): Lötschberg Base Tunnel, Second Tube: Automated Cost Optimization of the Inner Lining. *Tunnel* 5/2021, Bauverlag, S. 22-35.

[7] Moormann C. (2006): Trag- und Verformungsverhalten tiefer Baugruben in bindigen Böden unter besonderer Berücksichtigung der Baugrund-Tragwerk- und der Baugrund-Grundwasser-Interaktion. *Mitteilungen des Instituts und der Versuchsanstalt für Geotechnik der Technischen Universität Darmstadt*, Heft 59, Darmstadt, PhD thesis.

[8] Steiger H., Pitteloud L. (2001): Neue Entwicklungen bei der Optimierung der Bemessung von tiefen Baugruben in Frankfurt am Main. *Mitteilungen der TU Darmstadt*.

[9] Yin J., Wang Y., Hu J. (2012): Free Search with Adaptive Differential Evolution Exploitation and Quantum-Inspired Exploration. *J. Netw. Comput. Appl.*, Vol. 35, Issue 3, pp. 1035-1051.

Author

Dr.-Ing. Jörg Meier
Gruner AG
St. Jakobs-Strasse 199
4020 Basel, Schweiz
Tel.: +41 61 317 62 53
e-mail: joerg.meier@gruner.ch
Web: gruner.ch

Surrogate Modelling for a Deep Excavation Problem with Increasing Feature and Target Complexity

Johannes Leo, Thomas Elmer, Franz M. Rohrhofer, Franz Tschuchnigg

Abstract: Surrogate models, or metamodels, are mathematical representations of computational models that establish a relationship between input features and quantities of interest. The underlying computational model commonly represents a relatively complex, deterministic, and computationally expensive system that must be solved repeatedly for optimization or uncertainty quantification problems. Consequently, its direct application in uncertainty quantification or optimization tasks is often impractical, as these analyses typically require thousands to millions of model evaluations. Surrogate models are therefore developed for a specific feature and target space to replace the computationally expensive component of the analysis and enable the efficient generation of new predictions. The feature and target spaces can exhibit varying levels of complexity, ranging from homogeneous to spatially varying material parameters for the input features and from one-dimensional to multi-dimensional quantities of interest for the outputs. This work addresses these variations in complexity in the context of surrogate model development for a deep excavation problem and investigates different surrogate modeling approaches with respect to their performance and applicability.

1 Introduction

Numerical geotechnical analyses are characterized by a high level of complexity and uncertainty (Leo et al., 2025) arising from (i) the nonlinear behavior of soils, (ii) complex loading and boundary conditions, (iii) soil-structure interaction, and (iv) complicated geometries and intersections, to name only the most important factors. Consequently, the computational time required for such simulations can range from several minutes to several hours or even days, despite significant advances in computational efficiency over recent years. This renders these analyses relatively unattractive for applications such as back-analysis, optimization-based design, and uncertainty quantification, or otherwise restricts the parameter space to problems of low dimensionality. One approach to overcome these limitations is to replace the computationally expensive numerical geotechnical analysis with a mathematical representation that maps input features to quantities of interest (QoI). This mathematical representation is commonly referred to as a surrogate model or metamodel (Forrester et al. 2008). Once established, the surrogate model can be evaluated thousands of times within seconds while maintaining an approximation of the original numerical

model response. This substantial gain in computational efficiency enables its integration into uncertainty quantification and design optimization frameworks (Granitzer et al. 2026).

In general, the development of a surrogate model involves several steps (Forrester et al. 2008), as illustrated in Figure 1. The starting point is the definition of the input feature space, i.e., the identification of the parameters of the numerical geotechnical model that are considered variable / random. This feature space is subsequently sampled to generate a training dataset and must therefore provide sufficient coverage of the entire parameter space. For each sample, the QoIs, defined in advance, are obtained by repeatedly evaluating the numerical geotechnical model using the corresponding set of input features. The generated samples and their associated QoI responses are then combined into a dataset, which forms the basis for the data-driven surrogate model training. During the training phase, the selected surrogate modeling approach is employed to learn the relationships between the input features and the QoIs. In this phase, the parameters of the surrogate model are optimized by minimizing the discrepancy between the surrogate model predictions and the responses obtained from the numerical geotechnical model. Once the training process satisfies predefined convergence and quality criteria, the trained surrogate model is tested on data points that were not used during training in order to assess its predictive performance. If the surrogate model demonstrates satisfactory performance during testing, one can proceed to the application phase. In this phase, the input feature space is sampled according to the requirements of the analysis, and the trained surrogate model is used to efficiently predict the corresponding QoIs.

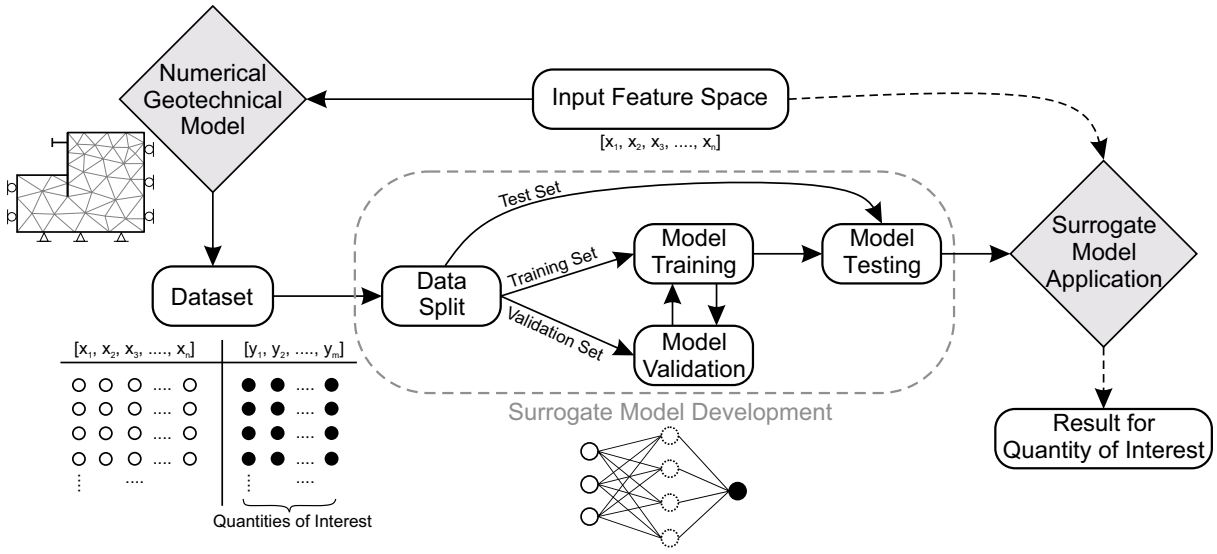


Figure 1: General workflow for surrogate model development with respect to a dataset from numerical geotechnical analysis.

When selecting a surrogate modeling approach, a wide range of methods with varying levels of formulation complexity is available. These methods range from computationally efficient and comparatively simple Response Surface Methods (RSM) (Wong, 1985) to statistical interpolation techniques such as Gaussian Processes (GP) (Marrel et al., 2024) and Polynomial Chaos Expansion (PCE) (Sudret, 2015), as well as deep learning approaches such as Artificial Neural

Networks (ANNs) (Cerek et al., 2026a). A variety of these methods and related approaches have been applied in geotechnical engineering, demonstrating the potential of surrogate models for different applications.

For example, Wong (1985) employed a RSM to investigate a slope stability problem, treating the effective friction angle (φ') and effective cohesion (c') as homogeneous random variables. An application of PCE is presented by Granitzer et al. (2026), whom assessed a cantilever sheet pile wall considering variable geometrical and constitutive model parameters and their influence on the factor of safety. Similarly, Yang et al. (2026) proposed a framework combining PCE with Principal Component Analysis to construct a surrogate model for high-dimensional outputs, demonstrated through the prediction of the deformation behaviour of a horizontally loaded pile. ANNs have likewise been increasingly applied in geotechnical engineering, as discussed by Cerek et al. (2026a). For instance, Soranzo et al. (2023) used Convolutional Neural Networks (CNN) to predict the factor of safety of slopes with arbitrary geometries and soil layering. Cerek et al. (2026b), on the other hand, employed both a Feedforward Neural Network and a Physics-Informed Neural Network to reproduce the deformation behaviour of a cofferdam while considering several construction parameters within the feature space. A further example is provided by Wang et al. (2021), who developed a CNN-based surrogate modeling framework incorporating spatially variable soil parameters for reliability-based slope assessment.

Although the scope of this literature review is intentionally limited, it nevertheless demonstrates that a variety of surrogate modeling approaches have been successfully applied to a broad spectrum of geotechnical problems involving different levels of feature and target complexity. However, a systematic assessment of how the choice of surrogate modeling approach is affected by the complexity of the input feature space and the target representation is still lacking. Therefore, this contribution compares the performance of different surrogate modeling methods using a deep excavation case study with varying feature and target complexities in order to identify reliable and computationally efficient approaches. Section 2 introduces the surrogate models considered in this study, while Section 3 evaluates and compares their performance for scenarios involving both scalar and vector-valued targets as well as different input feature complexities, from homogeneous to spatially variable soil parameters.

2 Surrogate models

In addition to the well-established surrogate modeling approaches GP (Marrel et al., 2024) and PCE (Sudret, 2015), ANN-based methods are also considered in this study. Since standard formulations of GP and PCE are employed, these methods are not described further in the following. Interested readers are referred to Marrel et al. (2024) and Sudret (2015) for detailed information. Furthermore, both approaches are only applied to problems with scalar target quantities. In contrast, a variety of architectures can be employed for ANNs. Therefore, the ANN architectures investigated in this work are illustrated in Figure 2 and described in greater detail below. ANNs are applied to problems with both scalar and vector-valued target quantities.

The most straightforward architecture considered is the Discrete Neural Network, which predicts the target quantity at a fixed number of locations (e.g., along a sheet pile wall). The model learns a mapping from the environmental parameters (variable features of the numerical model) to the solution vector using two hidden layers with 1024 neurons each. This architecture serves as a baseline for the more advanced neural network formulations and is illustrated in Figure 2a. Figure 2b shows a modified version of the discrete architecture that treats the target quantity as a continuous function by providing the spatial coordinate as an additional network input. This architecture is referred to as Continuous and learns a point-wise mapping $y = f(z,\text{env}),\ y \in \mathbb{R}$ thereby enabling predictions at arbitrary locations along the target function. The Continuous Two-Step architecture separates the learning process into two stages. First, the environmental parameters are encoded into a latent representation, effectively reducing the dimensionality of the input space. Subsequently, the spatial coordinate (e.g., depth along a sheet pile wall) is combined with the latent variables and used as input to a second network stage. The motivation behind this architecture is to simplify the learning task by separating environmental and spatial information. A schematic representation is shown in Figure 2c. The set of investigated architectures is completed by the Deep Operator Network by de Jong et al. (2025) (DeepONet), shown in Figure 2d. In this approach, the environmental and spatial information are processed by two independent subnetworks. The environmental parameters are treated by the *branch network*, whereas the spatial coordinate is processed by the *trunk network*. The outputs of both subnetworks are subsequently combined through a dot product to obtain the final prediction. Besides the flexibility in the design of the branch and trunk networks, a key advantage of DeepONet is that it learns operators rather than direct mappings between predefined inputs and outputs, making it particularly suitable for the prediction of continuous functions.

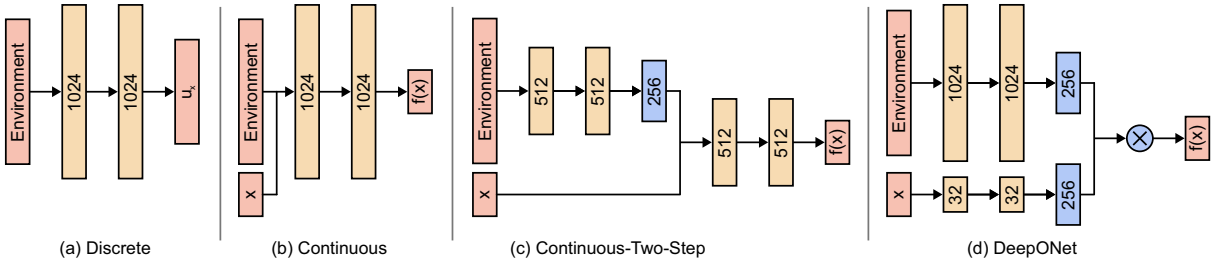


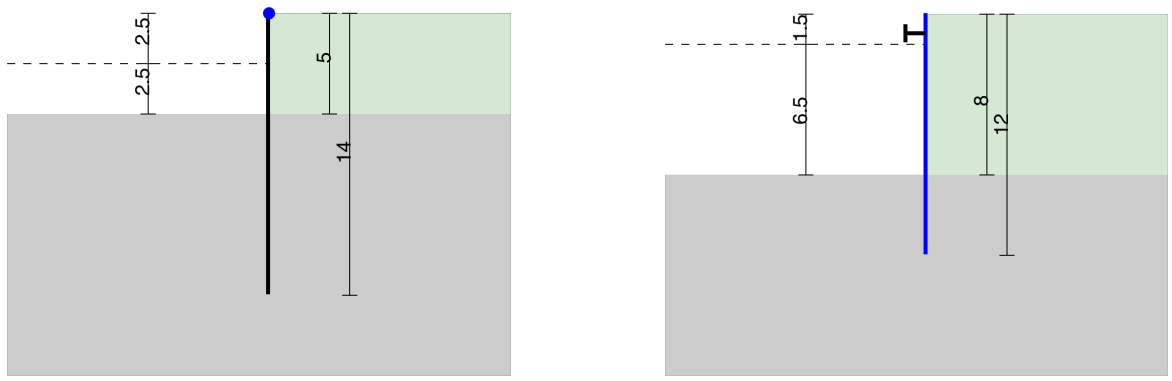
Figure 2: ANN surrogate model architectures considered for scalar and vector target predictions.

3 Experimental Setup

For the experimental setup, two deep excavation boundary value problems, shown in Figure 3, are considered. In both scenarios, the subsurface consists of two soil layers, of whom the soil parameters are either considered as homogeneous random variables or spatially correlated random variables. For the analyses assuming homogeneous soil conditions, the soil behavior is described using the Hardening Soil (HS) model (Schanz et al., 1999) with the parameters listed in Table 1 (Leo et al., 2026a). Parameters treated as random variables can also be identified from Table 1 by

their associated coefficients of variation (COV). As no computationally efficient implementation of the HS model capable of representing spatially variable soil parameters was available to the authors, a user-defined linear-elastic perfectly-plastic soil model with a Mohr-Coloumb failure criterion, adopted from Leo et al. (2026b), is employed. This material model allows the assignment of material parameters c', $\varphi'$and young‘s modulus (E) at the stresspoint level, thereby enabling the consideration of spatial variable soil material parameters. A detailed description of the adopted formulation is provided by Leo et al. (2026b). Corresponding material and random field parameters are found in Table 1. All random fields are generated using an exponential correlation model with a horizontal correlation length of 50 m and a vertical correlation length of 2 m. Lognormal distributions are assumed for c‘ and φ‘, to circumvent infeasible negative values, whereas a normal distribution is adopted for E. To avoid unrealistic stiffness values, the random fields of E are limited to a minimum value of 100 kPa and are defined using a depth dependent mean value according to $E = E_{\text{ref}} \left(\frac{\sigma'_1}{p_{\text{ref}}} \right)^m$. Cross-correlation between material parameters is not considered in the present study. For each scenario, a dataset comprising 1,000 finite element simulations is generated.

The first scenario depicts a 5 m deep excavation supported by a 14 m long sheet pile wall. The QoI is defined by the horizontal deflection of the sheet pile wall head, representing a scalar target problem. The numerical simulation sequence consists of the initialization of in-situ stresses using the K0-procedure, followed by the wished-in-place installation of the sheet pile wall and two excavation stages until the final excavation depth is reached. The QoI is extracted after completion of the final calculation stage. The remaining model dimensions are shown in Figure 3a.



(a) Scenario 1 - scalar target problem

(b) Scenario 2 - vector target problem

Figure 3: Layout of the excavation scenarios investigated, highlighting the different target complexities.

In the second scenario, the complexity of the boundary value problem is increased by considering an excavation depth of 8 m. Consequently, additional support measures are required, as illustrated in Figure 3b. The excavation is supported by a 12 m long sheet pile wall and a strut located 1.5 m below the initial ground surface. The simulation sequence comprises the initialization

of stresses using the K0-procedure, the wished-in-place installation of the sheet pile wall, and the installation of the strut combined with excavation to the final depth. Material properties of the structural elements are adopted from Jürgens and Henke (2024). In this scenario, the QoI is defined as the continuous deformation profile of the sheet pile wall, accordingly used in the studies dealing with vector target predictions.

Additionally, it should be noted that the assumption of an axis of symmetry does not strictly hold for analyses involving spatially variable soil parameters. Nevertheless, this assumption is retained in the present study for the sake of simplicity.

Table 1: Hardening soil (HS) and Mohr-Coloumb (MC) material parameters and random variables with their associated variability.

Parameter	Symbol	Layer 1	COV_1	Layer 2	COV_2	Unit
(Un-)saturated unit weight	$\gamma_{\text{unsat}}, \gamma_{\text{sat}}$	15	–	18	–	kN/m^3
Effective friction angle	φ'	25	0.20	30	0.20	°
Effective cohesion	c'	10	0.10	0	–	kPa
Dilatancy angle	ψ	0	–	0	–	°
Secant reference stiffness	$E_{50,\text{ref}}$	3.5	0.35	20	0.35	MPa
Oedometric reference stiffness	$E_{\text{oed,ref}}$	3.0	0.35	20	0.35	MPa
Un-/reloading reference stiffness	$E_{\text{ur,ref}}$	7.0	0.35	60	0.35	MPa
Young‘s modulus	E_{ref}	4.65	0.35	60.00	0.35	MPa
Poisson’s ratio (un-/reloading)	ν_{ur}	0.2	–	0.2	–	–
Stress dependency index	m	0.9	–	0.5	–	–
Reference pressure	p_{ref}	100	–	100	–	kPa
Strength reduction factor	R_{inter}	0.64	–	0.63	–	–
Overconsolidation ratio	OCR	1.0	–	1.0	–	–

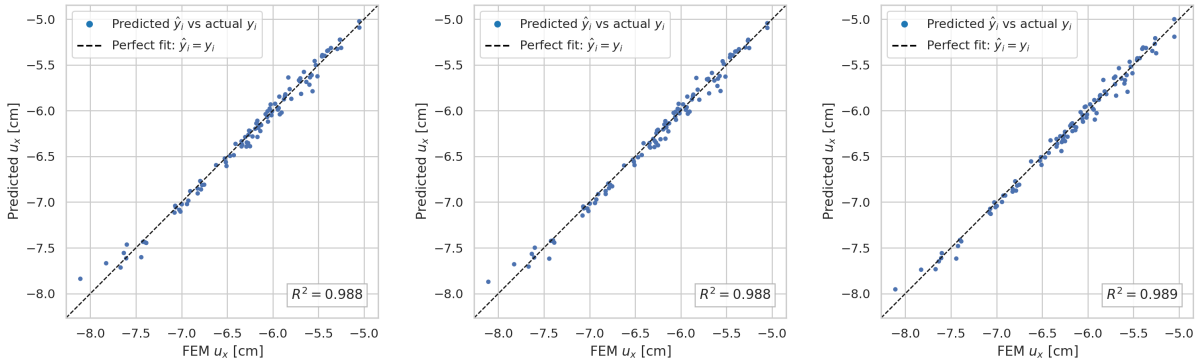
3.1 Scalar target prediction

As stated earlier, Scenario 1 is adopted for the investigation of surrogate models with scalar target quantities. Since the deformation profile of the sheet pile wall is continuous and reaches its maximum at the wall head, the horizontal displacement at this location is selected as the QoI. The results for the homogeneous soil case are presented and discussed in Section 3.1.1, whereas the corresponding analyses considering spatially variable (non-homogeneous) soil parameters are presented in Section 3.1.2.

3.1.1 Homogeneous soil parameters

The input features considered in the homogeneous studies are $\{\varphi', c', E_{50,\text{ref}}, E_{\text{oed,ref}}, E_{\text{ur,ref}}\}$, sampled using the Latin Hypercube Sampling method, and are used to compare the predictive performance of the GP, PCE, and Discrete Neural Network surrogate models. Figure 4 shows

the predicted horizontal displacements of the sheet pile wall head ($\hat{y}_i$) plotted against the corresponding FEM reference values (y_i) of the test set. The dashed line represents perfect agreement between prediction and reference ($\hat{y}_i = y_i$). As can be seen, all three approaches accurately reproduce the response of the numerical geotechnical model, with the predictions closely clustered around the identity line over the entire range of observed displacements ($u_x \approx -8.0$ to -5.0 cm). This visual impression is confirmed by the coefficients of determination, which amount to $R^2 = 0.988$ for both the GP and the PCE, and $R^2 = 0.989$ for the Discrete Neural Network. Furthermore, no substantial or practically relevant differences in the training times of the three surrogate models were observed. Consequently, for scalar target quantities in combination with moderate feature dimensionality, all three surrogate modeling approaches exhibit comparable predictive accuracy and computational efficiency.



(a) GP (b) PCE (c) DNN

Figure 4: Predicted versus reference horizontal displacement of the sheet pile wall head (u_x) for the test set of Scenario 1 with homogeneous soil parameters using the GP, PCE, and Discrete Neural Network (DNN) surrogate models.

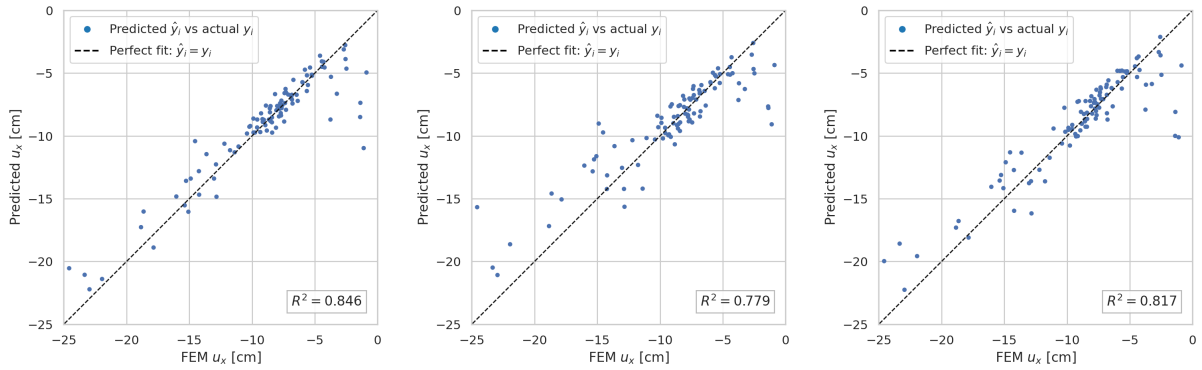
3.1.2 Spatially variable soil parameters

Having demonstrated that the investigated surrogate modeling approaches are capable of accurately reproducing the response under homogeneous soil conditions, the complexity of the input feature space is subsequently increased by considering spatially variable soil parameters. Since the discretized random fields constitute a high-dimensional input space, the environmental parameters can no longer be processed efficiently using standard GP and PCE formulations. Therefore, CNN-based surrogate architectures are employed, in which the random field realizations are first encoded by convolutional layers before being passed to the respective neural network architecture.

The predictive performance of the CNN-Discrete, CNN-Continuous, and CNN-DeepONet architectures is shown in Figure 5. Compared to the homogeneous case, all three approaches exhibit a considerably larger scatter around the identity line. This observation is reflected by the corresponding coefficients of determination, yielding $R^2 = 0.846$ for the CNN-Discrete,

$R^2 = 0.779$ for the CNN-Continuous, and $R^2 = 0.817$ for the CNN-DeepONet. Accordingly, the substantial increase in feature space dimensionality results in reduced predictive accuracy while simultaneously increasing the computational cost of training.

A more detailed assessment based on the prediction error distributions reveals that all three models exhibit an interquartile range of approximately ± 0.8 cm, whereas the whiskers extend to approximately ± 3 cm and individual outliers reach absolute errors of up to 10 cm. Here, the whisker length is defined as 1.5 times the interquartile range. Consistent with the coefficients of determination, the CNN-Discrete architecture achieves the highest overall predictive accuracy. Nevertheless, all three models provide reasonable predictions in the vicinity of the mean response, corresponding to horizontal displacements between approximately 5–10 cm. Outside this range, however, the predictive performance deteriorates noticeably. This behavior is attributed to the Monte Carlo sampling strategy employed for dataset generation for spatially variable soil parameters, as the tails of the response distribution are underrepresented in the training data. These findings highlight the need for enriched training datasets or tailored sampling strategies in future work.



(a) CNN-Discrete (b) CNN-Continuous (c) CNN-DeepONet

Figure 5: Predicted versus reference horizontal displacement of the sheet pile wall head (u_x) for the test set of Scenario 1 with spatially variable soil parameters using the CNN-Discrete, CNN-Continuous, and CNN-DeepONet architectures.

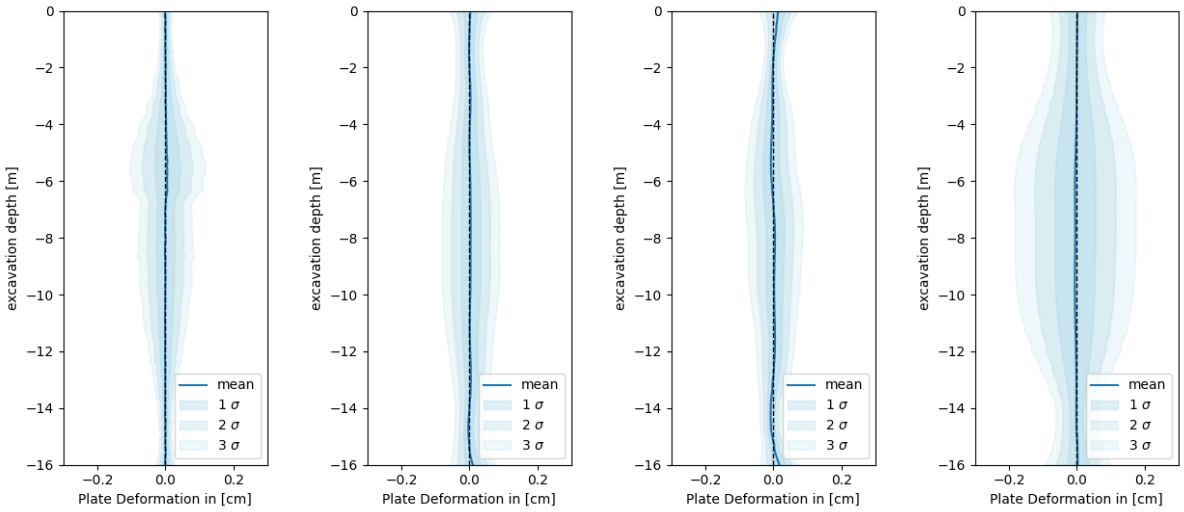
3.2 Vector target prediction

Moving from scalar to vector target prediction, Scenario 2 (Figure 3b) is considered, where the QoI is defined as the continuous deformation profile of the sheet pile wall. In contrast to the scalar target problem, the surrogate models must now reproduce the complete deformation profile, including the characteristic bulging below the strut level and the response along the embedment length. Accordingly, only ANN-based surrogate architectures are considered, namely the Discrete, the Continuous, the Continuous-Two-Step, and the DeepONet architectures.

3.2.1 Homogeneous soil parameters

Consequently, the same input feature space as introduced in Section 3.1.1 is adopted, with the spatial coordinate provided as an additional input for architectures predicting continuous target functions.

Figure 6 compares the depth-resolved prediction error distributions of the four investigated architectures by showing the mean error together with the one-, two-, and three-standard-deviation bands. For all models, the mean prediction error remains close to zero over the full wall length, indicating no systematic depth-dependent bias. The Discrete architecture exhibits the narrowest error bands, with the largest spread occurring near the maximum wall deflection and decreasing towards the wall head and toe. In contrast, the Continuous architecture shows a wider and more uniform error distribution, suggesting that the point-wise prediction strategy increases the learning complexity. This spread is considerably reduced by the Continuous-Two-Step architecture, indicating that separating environmental and spatial information simplifies the learning task. The DeepONet achieves a comparable error level but exhibits a slight oscillation of the mean error, indicating minor local deviations in the predicted deformation profile. Overall, prediction errors remain below approximately 0.2 cm for all architectures and are therefore negligible compared to the predicted deformation magnitudes of 5–10 cm, demonstrating that each surrogate model accurately reproduces the continuous wall deformation.



(a) Discrete (b) Continuous (c) Continuous-Two-Step (d) Deep-O-Net

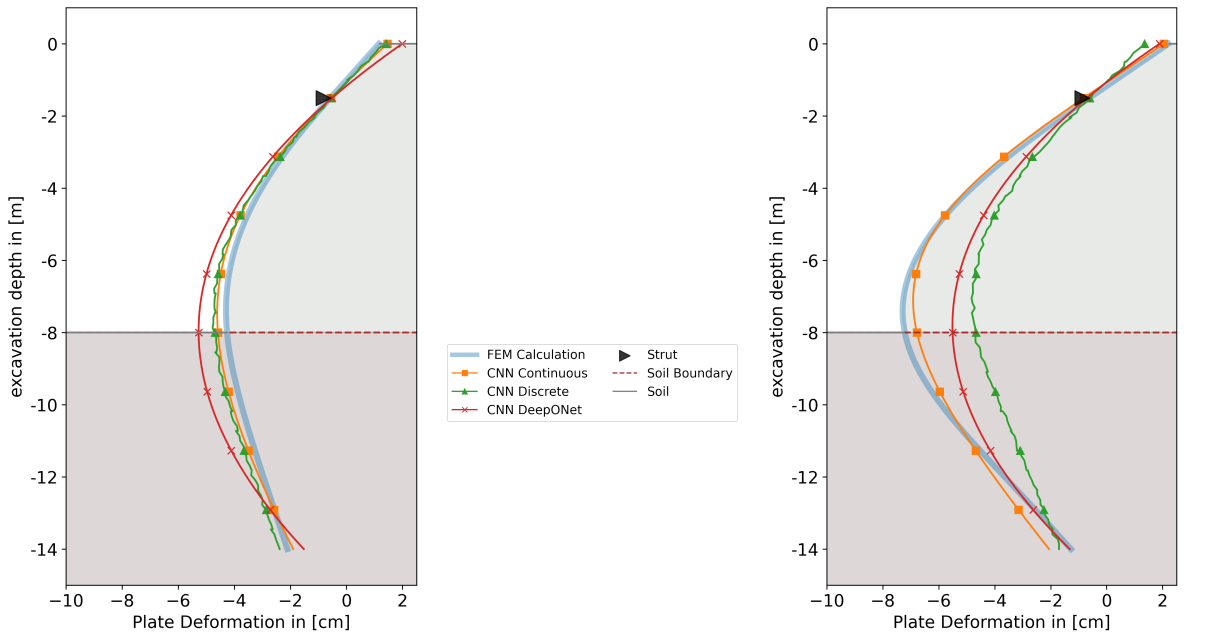
Figure 6: Depth-resolved error distributions of different surrogate modeling approaches for the test set of Scenario 2 with homogeneous soil parameters.

3.2.2 Spatially variable soil parameters

Finally, the most complex configuration is considered, combining vector target prediction with spatially variable soil parameters. As in the scalar case with non-homogeneous soil conditions, the CNN-Discrete, CNN-Continuous, and CNN-DeepONet architectures are investigated. Figure 7 compares the predicted and reference deformation profiles for two representative test samples.

Figure 7a shows a realization close to the mean wall deformation. All three architectures capture the overall deformation shape, with the CNN-Discrete and CNN-Continuous closely matching the reference solution, whereas the CNN-DeepONet slightly overestimates the deformation. In contrast, extreme realizations associated with larger wall displacements remain more challenging, as illustrated in Figure 7b. Here, the CNN-DeepONet underestimates the maximum deflection by approximately 2 cm, while the CNN-Discrete also underestimates the deformation over the entire wall and exhibits a non-smooth deformation profile due to its discrete output formulation.

These observations do not follow a consistent pattern across the test set, and no architecture consistently outperforms the others for extreme realizations. Overall, the CNN-based architectures reliably reproduce the mean deformation behavior of the sheet pile wall even for the most complex feature-target combination considered. However, accurately predicting extreme responses remains challenging and will likely require enriched training datasets or sampling strategies that better represent the distribution tails.



(a) Sample with high prediction accuracy

(b) Sample with reduced prediction accuracy

Figure 7: Predicted versus reference results of sheet pile wall deformations for two representative test samples with spatially variable soil parameters.

4 Conclusion

This contribution investigated the performance of different surrogate modeling approaches for a deep excavation problem by systematically increasing the complexity of both the input feature space, from homogeneous to spatially variable soil parameters, and the target representation, from a scalar wall-head displacement to the continuous deformation profile of the sheet pile wall. For homogeneous soil conditions, all investigated approaches achieved excellent predictive accuracy. For scalar targets, GP, PCE, and the Discrete NN performed nearly identically ($R^2 \geq 0.988$), indicating that computationally efficient surrogate models are sufficient for low-dimensional problems. Likewise, all ANN architectures accurately reproduced the continuous deformation profile. Introducing spatially variable soil parameters, however, reduced the predictive accuracy for both target types, with prediction errors increasing by more than one order of magnitude and extreme realizations remaining challenging for all CNN-based architectures.

This performance loss is primarily attributed to the sampling strategy used for database generation. While Latin Hypercube Sampling provides an efficient coverage of the homogeneous feature space, Monte Carlo sampling of random fields under represents extreme realizations. Consequently, the training dataset is concentrated around the mean response, leading to larger prediction errors in the distribution tails, which are particularly relevant for reliability-based applications.

Overall, the results demonstrate that surrogate model selection is primarily governed by the complexity of the feature space and target representation. While simple surrogate models are sufficient for homogeneous low-dimensional problems, high-dimensional spatially variable inputs require CNN-based architectures whose performance is currently limited more by the training data than by the network architecture itself. Future work should therefore focus on improved sampling strategies for random field problems and on extending the framework towards uncertainty quantification and reliability analysis.

References

[1] Cerek, K., Dao, D.A., Hadjiloo, E., and Grabe, J. (2026a): Development of neural network-based surrogate models in geotechnics: A state-of-the-art review. *Machine Learning and Data Science in Geotechnics*, 2(1). 10.1108/MLAG-04-2025-0011

[2] Cerek, K., Wojciechowski, M., Hadjiloo, E., and Grabe, J. (2026b): Surrogate model of cofferdams based on physics-informed neural networks. In *Proceedings of the 21st ICSMGE*, June 2026, Vienna, Austria. 10.53243/ICSMGE2026-670

[3] de Jong, T. O., Shukla, K., and Lazar, M. (2025): Deep Operator Neural Network Model Predictive Control. arXiv preprint arXiv:2505.18008. `https://arxiv.org/abs/2505.18008`

[4] Forrester, A. I. J., Sóbester, A., and Keane, A. J. (2008): *Engineering Design via Surrogate Modelling: A Practical Guide*. Wiley, New York.

[5] Granitzer, A., Nøst, H. A., Monsås, E., Erharter, G., and Leo, J. (2026): Surrogate Modeling Framework for Cantilever Sheet Pile Wall Stability Assessment: Case Study Campus Ullevål. *Geotechnical and Geological Engineering*, 44(4). 10.1007/s10706-026-03740-3

[6] Jürgens, H., and Henke, S. (2024): Numerical Optimisation of Excavation Pit Design Using Finite Element Analyses. *Geotechnical and Geological Engineering*, 42(3). 10.1007/s10706-023-02639-7

[7] Leo, J., Peterstorfer, T., and Tschuchnigg, F. (2025): Prognosis of Discrete Layer Boundaries for Synthetic Geological Models and Their Influence on Geotechnical Structures. In *Proceedings of the 9th ISGSR*, August 2025, Oslo, Norway.

[8] Leo, J., Erharter, G., and Tschuchnigg, F. (2026a): Geological Uncertainty and Its Consideration in Geotechnical Analysis by Means of the Finite Element Method. In *Proceedings of the 21st ICSMGE*, June 2026, Vienna, Austria. 10.53243/ICSMGE2026-1337

[9] Leo, J., Peterstorfer, T., and Tschuchnigg, F. (2026b): On the Consideration of Geological and Geotechnical Uncertainty in Numerical Slope Stability Analysis. *Engineering Geology*, 372. 10.1016/j.enggeo.2026.108952

[10] Marrel, A., and Iooss, B. (2024): Probabilistic surrogate modeling by Gaussian process: A review on recent insights in estimation and validation. *Reliability Engineering and System Safety*, 247. 10.1016/j.ress.2024.110094

[11] Schanz, T., Vermeer, P. A., and Bonnier, P. G. (1999): The Hardening Soil Model: Formulation and Verification. In: R. B. J. Brinkgreve (Ed.), *Beyond 2000 in Computational Geotechnics: 10 Years of PLAXIS*. Balkema, Rotterdam.

[12] Soranzo, E., Guardiani, C., Chen, Y., Wang, Y., and Wu, W. (2023): Convolutional neural networks prediction of the factor of safety of random layered slopes by the strength reduction method. *Acta Geotechnica*, 18(6). 10.1007/s11440-022-01783-3

[13] Sudret, B. (2015): Polynomial Chaos Expansions and Stochastic Finite Element Methods. Chapter 6 in *Risk and Reliability in Geotechnical Engineering*. Taylor & Francis.

[14] Wang, Z. Z., Xiao, C., Goh, S. H., and Deng, M. M. (2021): Metamodel-Based Reliability Analysis in Spatially Variable Soils Using Convolutional Neural Networks. *Journal of Geotechnical and Geoenvironmental Engineering*, 147(3). 10.1061/(ASCE)GT.1943-5606.0002486

[15] Wong, F. S. (1985): Slope Reliability and Response Surface Method. *Journal of Geotechnical Engineering*, 111(1). 10.1061/(ASCE)0733-9410(1985)111:1(32)

[16] Yang, N., Le, T., Zdravkovic, L., Potts, D., and Altmeyer, R. (2026): Sequential Calibration of Soil Parameters Using a Two-Step Surrogate for High-Dimensional Geotechnical Outputs. *Computers and Geotechnics*, 190. 10.1016/j.compgeo.2025.107688

Authors

Dipl.-Ing. Johannes Leo
Graz University of Technology
Institute of Soil Mechanics, Foundation Engineering and Computational Geotechnics
Rechbauerstrasse 12
8010 Graz, Austria
e-mail: leo@tugraz.at

Thomas Elmer BSc
Graz University of Technology
Institute of Soil Mechanics, Foundation Engineering and Computational Geotechnics
Rechbauerstrasse 12
8010 Graz, Austria

Dr.-techn. Franz M. Rohrhofer
Know-Center GmbH
Research Center for Data-Driven Business & Big Data Analytics
Sandgasse 36/4
8010 Graz, Austria
e-mail: frohrhofer@know-center.at

Univ.-Prof. Dr.-techn. Franz Tschuchnigg
Graz University of Technology
Institute of Soil Mechanics, Foundation Engineering and Computational Geotechnics
Rechbauerstrasse 12
8010 Graz, Austria
e-mail: franz.tschuchnigg@tugraz.at

A data-driven perspective on the vibrodriving process: Modelling, optimisation and control

Francisco Williams-Riquer[1], **Mechthild Cramer**[1]

[1]Institute of Geotechnical Engineering and Construction Management
Hamburg University of Technology, Hamburg, Germany

Data-driven methods offer potential to improve the vibratory pile-driving process through modelling, drivability assessment, and automation. However, their development requires high-quality datasets that are often difficult to obtain, as full-scale field measurements and centrifuge experiments demand considerable financial and logistical effort. To address this limitation, a model variable-moment vibrodriver was designed and constructed for operation in controlled laboratory and field conditions. This work provides an overview of the design and construction of the model vibrodriver as well as recent results leveraging data generated with this platform. Two different examples of data-driven approaches are presented. First, the modelling and identification of ground vibrations in the surrounding area during installation are investigated using field data from experiments in which model tubular piles were installed. This identification is performed using Dynamic Mode Decomposition with Time Delays (DMD-TD), an equation-free modal analysis method. Second, the identification and classification of vibrodriver regimes (contact and non-contact vibrodriving) is done directly from acceleration measurements. This classification is done using a supervised learning workflow based on a convolutional neural network. Overall, the presented approaches demonstrate that data-driven strategies, supported by purpose-built experimental hardware, can substantially improve the understanding, monitoring, and control of vibratory pile driving, paving the way toward more autonomous, efficient, and reliable pile installation in geotechnical practice.

1 Introduction

Vibrodriving has established itself as an effective alternative to impact driving for pile installation, offering faster penetration, reduced noise emissions, and a lower risk of structural damage to the pile. Despite these benefits, the technique has not achieved full acceptance in geotechnical practice, primarily because the nonlinear and nonsmooth dynamics of the coupled vibrator–pile–soil system remain insufficiently understood (Holeyman et al. 2017; Massarsch, Fellenius, et al. 2017). This knowledge gap manifests in two persistent challenges. First, the bearing capacity of vibrodriven piles is generally lower than that of impact-driven piles and depends on driving parameters, which has led several industrial standards to recommend completing the final driving phase with impact hammers (Rausche 2002). Second, the ground vibrations radiated during installation pose a concern in urban environments, where regulatory thresholds protect adjacent structures and sensitive infrastructure. Although machine parameters and soil

conditions are known to govern the process, the magnitude of their influence remains partly contradictory in the literature (Achmus et al. 2020).

Given these complexities, data-driven modelling has emerged as a promising path forward. Unlike physics-based models, which require explicit assumptions about constitutive behaviour, data-driven approaches can extract dominant patterns and predictive relationships directly from measurements, thereby bypassing modelling assumptions that are difficult to justify for vibrodriving. Their effectiveness, however, depends on the availability of high-quality datasets that span a representative range of operating conditions. Improving the understanding of vibrodriving has motivated extensive experimental work, ranging from early large-scale campaigns (Cudmani, Huber, et al. 2000; Dierssen 1994; Viking 2002) to more recent investigations of the influence of frequency and eccentric moment on penetration kinematics (Martinelli et al. 2025; Massarsch, Wersäll, et al. 2021). Nonetheless, field tests using an industrial vibrodriver, while indispensable, are costly, logistically demanding, and constrained to conditions tolerated by real construction projects, which limits the variability of the resulting datasets. Contractors are understandably reluctant to explore limiting or extreme operating scenarios, further narrowing the range of data available for analysis. Small-scale laboratory experiments (Rodger et al. 1980; Vogelsang et al. 2017) offer greater control but cannot realistically reproduce wave propagation in the surrounding soil, limiting geophone data.

To bridge this gap, this work proposes generating datasets using a model vibrodriver as an intermediate state to conceptually test data-driven strategies. A model variable-moment electrical vibrodriver was designed to combine laboratory-level reproducibility with field-level realism. The device was deployed during a field campaign in medium-to-dense sand and in laboratory conditions with fully saturated sand while simultaneously recording acceleration, displacement, force-torque, and (for the field case) geophone signals. The resulting datasets form the basis for the two data-driven case studies presented in this work, which illustrate the potential of such strategies for the analysis and monitoring of vibrodriving.

The first case study discusses the equation-free modelling and identification of ground vibrations transmitted to the surroundings employing Dynamic Mode Decomposition (DMD). DMD has gained traction as a technique for extracting dominant spatio-temporal patterns from measurements and producing physically interpretable modes (Kutz et al. 2016). Initially developed for fluid dynamics, DMD has since been applied across a range of engineering domains, including soil and structural dynamics (Abdennadher et al. 2025; Shioi, Otake, et al. 2023; Shioi and Otake 2026; Williams-Riquer, Chmelnizkij, Alkateeb, et al. 2024; Williams-Riquer, Abdennadher, et al. 2025; Williams-Riquer, Chmelnizkij, Abdennadher, et al. 2026). However, it encounters two limitations for direct application to vibrodriving: the standard DMD assumes stationary dynamics, and it is highly sensitive to measurement noise, leading to artificial damping in the reconstructed modes (Dawson et al. 2016). Both limitations can be mitigated by augmenting the data with time-delay coordinates, which enrich the observable space and improve the separation between coherent dynamics and noise (Arbabi et al. 2017; Dylewsky et al. 2022), and by deploying the algorithm in a recursive moving time-window manner. The resulting DMD with time-delay embeddings (DMD-TD), formulated recursively and adaptively, enables continuous tracking of the evolving wave field from noisy geophone measurements.

The second case study addresses the identification of the driving regime. A central concept for describing vibrodriving behaviour is the distinction between contact and non-contact regimes

(Cudmani 2001): in contact vibrodriving, the pile tip maintains permanent contact with the soil throughout the vibration cycle, whereas in non-contact vibrodriving, the tip detaches during the upward phase and re-engages during downward motion (Cudmani and Manthey 2019; Massarsch, Wersäll, et al. 2022; Vogelsang et al. 2017). Distinguishing between these regimes is advantageous for assessing installation efficiency; however, existing identification approaches rely on force or pressure measurements (Cudmani 2001), which are not routinely available during installation. Accelerometers, by contrast, are robust, compact, and straightforward to integrate into vibrodriving equipment. This work presents a supervised convolutional neural network (CNN) workflow (Fawaz et al. 2019; LeCun et al. 1998; Wang et al. 2017) to classify contact and impact regimes from accelerometer data alone.

The remainder of this paper is organised as follows. Section 2 describes the model vibrodriver, its instrumentation, and the laboratory and field campaigns. Section 3 presents the recursive DMD-TD approach for modelling and identifying ground vibrations. Section 4 introduces the CNN-based classification of driving regimes. Section 5 summarises the main findings and outlines directions for future work.

2 Design and construction of the model vibrodriver

The vibratory action transmitted to the pile is commonly decomposed into a dynamic component generated by the oscillating mass and a static one arising from the oscillating free weight acting on the pile. In the model vibrodriver presented here, no static component acts on the pile, since all relevant masses, including the vibrator housing and the pile itself, are assumed to oscillate jointly along a guide rail.

The dynamic excitation force is given by

$$F_{\mathrm{dyn}} = m_e \, l \, \omega^2 \cos\!\left(\frac{\alpha}{2}\right) \tag{1}$$

where m_e is the eccentric mass, l is the eccentricity radius, ω is the angular rotation velocity, and α is the phase difference between the upper and lower pairs of eccentric masses. By adjusting α, the magnitude of the applied dynamic force can be systematically varied without modifying the rotation speed, enabling independent control of frequency and excitation amplitude.

The designed model consists of four identical eccentric masses, each directly driven by an individual servomotor. The gearless arrangement eliminates the need for mechanical synchronisation through gears or belts and enables electronic regulation of the phase difference. The main mechanical parameters are summarised in Table 1, and a CAD rendering is shown in Figure 1.

The device is instrumented with several sensors. Each servomotor is fitted with an absolute encoder to measure angular position and velocity. The vibrodriver housing carries a triaxial accelerometer, a linear encoder attached to the guide, and a six-axis force–torque sensor between the vibrator and the pile. All sensor data are recorded synchronously at a 4 kHz sampling rate.

For field applications, the vibrodriver is mounted on a portable steel frame to facilitate trans-

Table 1: Mechanical parameters of the model vibrodriver.

Symbol	Description	Value
m_e	Eccentric masses	0.772 kg
m_p	Pile mass	5.0 kg
m_h	Housing mass	19.23 kg
l	Equivalent radius	17 mm
d	Pile diameter	30 mm

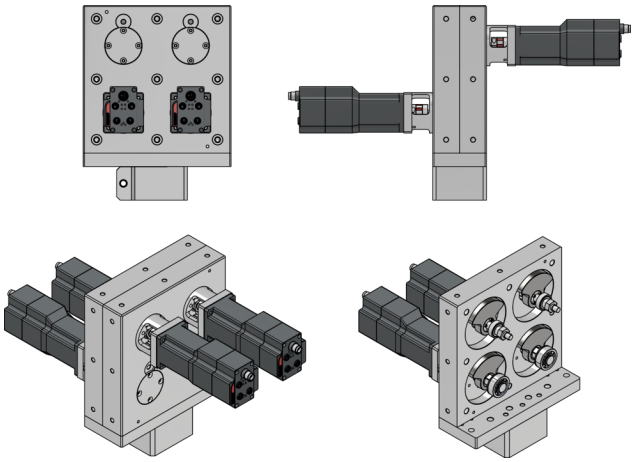


Figure 1: CAD model of the constructed vibrodriver.

port. A rope-and-pulley mechanism is integrated into the frame to enable pile extraction after each test. For field operation, the electronic components are powered by a battery pack and four triaxial geophones are arranged in a straight line from the vibrodriver to measure ground vibration at different radial distances.

3 Case Study I: Modelling of ground vibrations using DMD-TD

This section presents the first data-driven case study, in which Dynamic Mode Decomposition with Time-delay Embeddings (DMD-TD) is applied to model and identify the ground vibrations transmitted to the surroundings during vibrodriven pile installation. The objectives are: to obtain a low-dimensional, physically interpretable representation of the wave field captured by the geophones, and to adapt this representation continuously as the dynamics of the vibrator–pile–soil system evolve during installation.

DMD is a data-driven technique that identifies the dominant linear dynamics of a system from time-resolved snapshots, producing a low-dimensional modal decomposition in which each mode corresponds to an independent linear oscillator with a frequency, a growth or decay rate, and a spatial distribution across measurement channels.

Two practical limitations affect the direct application of standard DMD to vibrodriving. First,

the method assumes stationarity, which is violated as the pile penetrates evolving soil layers. Second, the eigenvalue estimates are biased by measurement noise, introducing artificial damping in the reconstructed modes. Both limitations can be addressed by augmenting the snapshot matrices with time-delayed copies, forming so-called Hankel matrices. As the number of delay coordinates increases, a spectral gap develops between signal-related and noise-related singular values, providing a natural criterion for rank truncation and denoising (Williams-Riquer, Chmelnizkij, Abdennadher, et al. 2026).

To track time-varying dynamics during installation, the DMD-TD procedure is deployed in a recursive sliding-window framework with automatic selection of the truncation rank and the number of delay coordinates. For each window, the dominant frequencies are first identified via FFT-based peak detection, and the truncation rank is set to twice the number of detected peaks to account for complex-conjugate mode pairs. The number of delays is then iteratively increased until a clear spectral gap between the last singular value corresponding to the selected rank and the next singular value is achieved. The window is subsequently shifted forward, and the procedure is repeated.

3.1 Results

A field measurement campaign was conducted using the model vibrodriver, during which three open-ended tubular model piles (30 mm in diameter) were installed. The driving frequency (up to 30 Hz) and eccentric moment (0–100%) were manually varied to achieve a penetration depth of approximately 80 cm. The site consisted of fine sand with up to 25% coarser sand and 8–11% fines, with $D_{50} = 0.144\,\text{mm}$, $C_u = 2.38$, $e_{\max} = 1.162$, $e_{\min} = 0.346$, and a partially saturated in-situ relative density of $D_r = 0.66$–0.74. During each installation, four radially placed triaxial geophones were used to capture particle velocity, as shown in Figure 2.

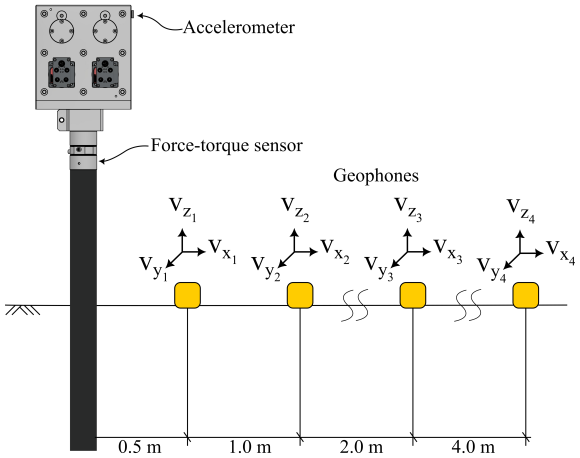


Figure 2: Positioning of geophones, an accelerometer, and a force–torque sensor during pile installation.

The adaptive DMD-TD algorithm was applied to the collected data. Figure 3 illustrates the resulting modal decomposition for the closest and farthest geophones. Each mode represents

an independent oscillator with a characteristic frequency and a spatial distribution across geophones at different radial distances. The mode shapes thus encode the spatial distribution of vibration energy, allowing wave-related features such as attenuation and phase relationships, which are characteristic of surface and body waves during pile installation, to be extracted directly from the modal decomposition.

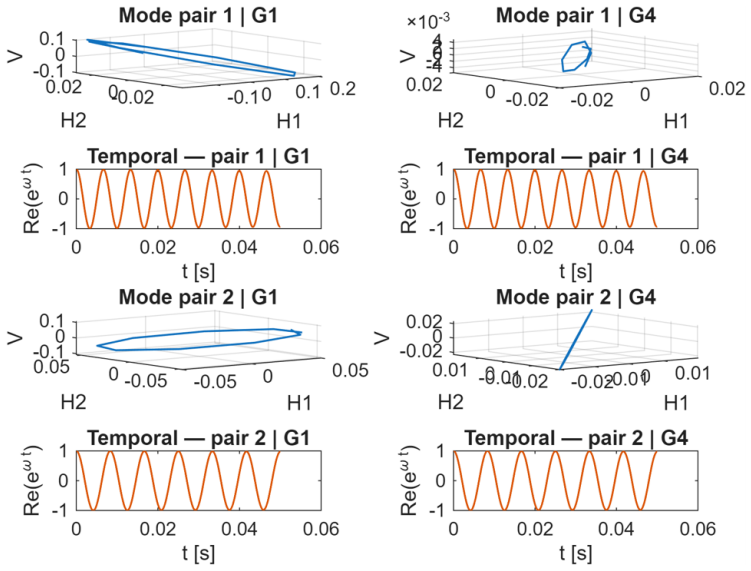


Figure 3: Shows the first four modes for G1 and G4.

4 Case Study II: Classification of vibrodriving regimes

This second case study presents a CNN for distinguishing contact from non-contact vibrodriving using only accelerometer data, thereby eliminating the need for force or pressure sensors, which are rarely available on-site. CNNs are well-suited to this task because they learn temporal patterns directly from one-dimensional signals (Fawaz et al. 2019). The network for this study consists of three convolutional blocks, followed by a classification head with dropout and a softmax output layer. The dataset was split 70/15/15 into training, validation, and test sets, with normalisation based only on the training data.

4.1 Results

Laboratory tests were conducted using the model vibrodriver in saturated coarse sand ($D_{50} = 0.656$ mm). Six tests were performed in total: three at 20 Hz and three at 30 Hz, with signals sampled at 4 kHz. A piezoelectric pressure sensor installed beneath the pile toe served as an independent reference for labelling the driving regime.

Figure 4 presents the classification procedure for a representative 30 Hz test based on the force measured at the pile head and the tip pressure measured at the pile toe. The upper diagram shows

the force–displacement curve of the entire installation process. Three sections of the force-displacement response are highlighted (Subplots 1.1–1.3). Section 1 corresponds to contact driving, Section 3 to non-contact driving, and Section 2 captures the transition between the two modes. At the start of the test (1.1), the displacement difference per cycle is small, with $\Delta d \approx 0.2\,\text{mm}$. As penetration progresses, Δd grows (1.2) and reaches $\Delta d \approx 1.4\,\text{mm}$ in Section 3 (1.3). Sections 1 and 3 can therefore be unambiguously assigned to the contact and non-contact driving modes, respectively.

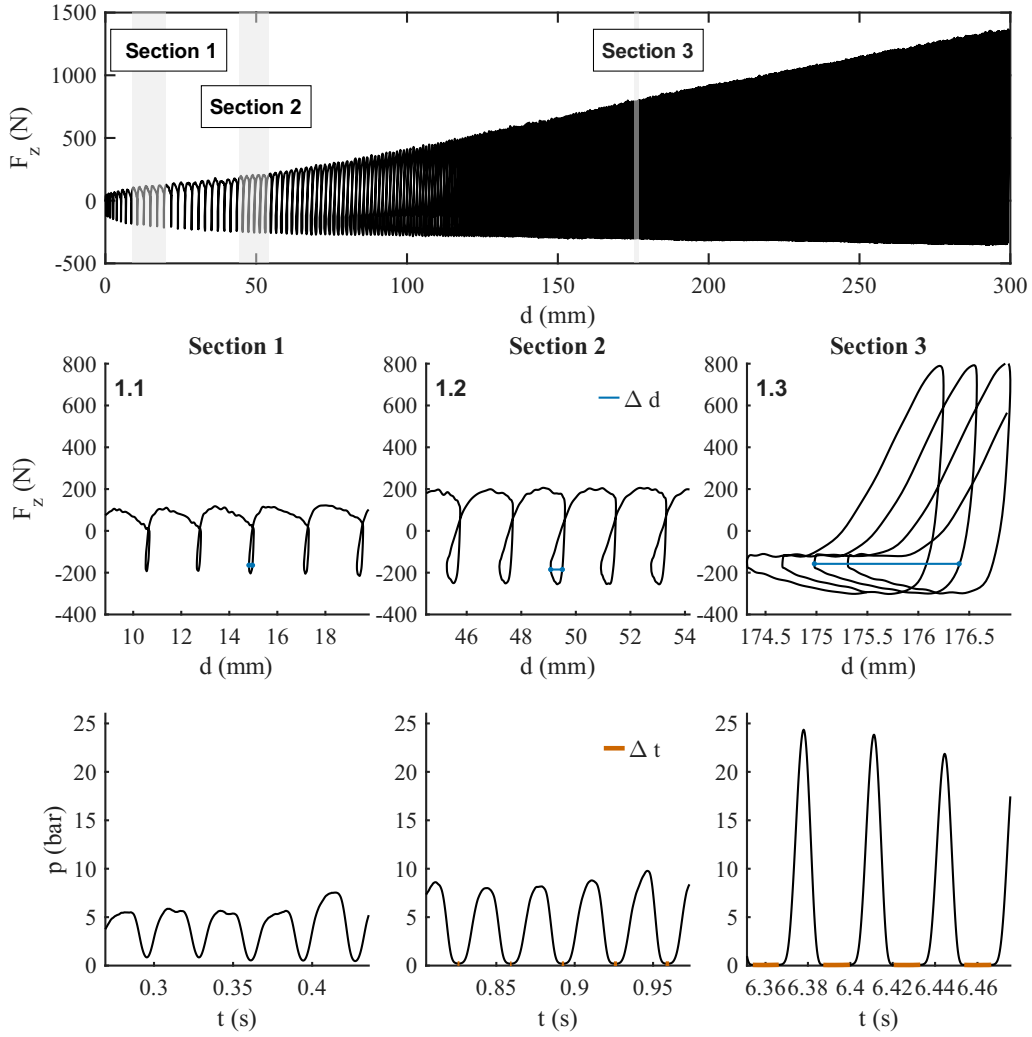


Figure 4: Driving mode classification.

The corresponding tip-pressure signals are shown in Subplots 2.1–2.3. During non-contact driving (2.3), the tip pressure exhibits a plateau at 0 bar, indicating a temporary loss of contact between the pile toe and the soil. During contact driving (2.1), the tip pressure remains positive throughout each vibration cycle, confirming sustained tip-soil contact.

For pressure-signal classification, the raw pressure signal is smoothed using a moving average with a window size of 20 samples. The time derivative of the smoothed signal is then computed around each local pressure minimum to obtain the rate of pressure change for every cycle. If the derivative remains within a $\pm$ 15 bar/s threshold for a time greater $\Delta t = 0.002$ s, a loss

of contact between the pile tip and the soil is assumed. The corresponding cycle is therefore classified as a non-contact cycle and vice versa.

All classified cycles from the six vibration tests (10,827 cycles in total) are subsequently segmented into individual snippets, each starting and ending at the upper turning point of the pile motion. The corresponding acceleration measurements are segmented in the same way. These classified acceleration snippets are then used as input for the convolutional neural network (CNN).

Figure 5 shows the resulting confusion matrices for the CNN classifier. Training and validation accuracies reach 100% for both classes. On the test set, only a single non-contact cycle is misclassified as contact, yielding recalls of 100% (contact) and 99.9% (non-contact), and precisions of 99.8% (contact) and 100% (non-contact). The CNN, therefore, reliably separates the two driving regimes from accelerometer data alone, despite the gradual nature of the transition, and captures patterns that simple threshold-based methods applied to a single channel would likely miss.

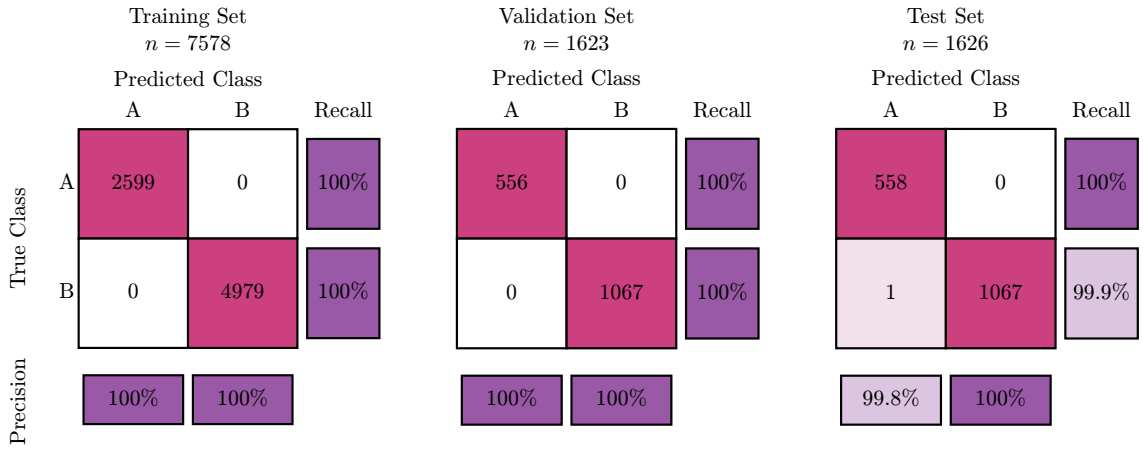


Figure 5: Confusion matrices with class-wise recall (row-normalised) and precision (column-normalised) for the training, validation, and test datasets of the CNN classifier.

5 Conclusions

This work has presented a model variable-moment vibrodriver developed for generating datasets from both field and laboratory experiments, suitable for the development and testing of data-driven applications in vibratory pile driving.

The transferability of these results to full-scale industrial vibrators remains an open question, and the present work should be understood as a proof of concept. Nevertheless, the data generated by the model vibrodriver already reproduces key qualitative features of those collected with industrial hardware: the contact and non-contact regimes can be clearly identified from the recorded signals, and the propagation of vibrations into the surrounding soil can be characterised in both the near and far field of the geophone array. Further studies are required to confirm the scalability of the proposed methods to field equipment and to a broader range of

soil conditions and driving parameters.

The experimental platform and the associated datasets pave the way towards partial and, eventually, fully automated systems for the optimisation of the vibrodriving process. Ongoing work focuses on extending the experimental campaign, on coupling the two data-driven methods to enable simultaneous identification of the radiated wave field and the prevailing driving regime, and on integrating these tools into a real-time monitoring and control framework for industrial pile installation.

References

[1] Abdennadher, M. A., F. Williams-Riquer, D. A. Dao, and J. Grabe (2025). "Data-driven modeling of structural dynamics: Improving measurement noise robustness with augmented Dynamic Mode Decomposition". In: *Proceedings of the 10th ECCOMAS Thematic Conference on Computational Methods in Structural Dynamics and Earthquake Engineering (COMPDYN 2025)*.

[2] Achmus, M., A. S. Schmoor, V. Herwig, and B. Matlock (2020). "Lateral bearing behaviour of vibro- and impact-driven large-diameter piles in dense sand". In: *Geotechnik* 43.3, pp. 147–159.

[3] Arbabi, H. and I. Mezić (2017). "Ergodic theory, Dynamic Mode Decomposition, and computation of spectral properties of the Koopman operator". In: *SIAM Journal on Applied Dynamical Systems* 16.4, pp. 2096–2126.

[4] Cudmani, R. (2001). "Statische, alternierende und dynamische Penetration in nichtbindigen Böden". PhD thesis. Universität Karlsruhe.

[5] Cudmani, R., G. Huber, and G. Gudehus (2000). "A mechanical model for the investigation of the vibro-drivability of piles in cohesionless soils". In: *Proceedings: Vibratory pile driving and deep soil compaction – TRANSVIB2000*.

[6] Cudmani, R. and S. Manthey (2019). "A novel vibro-penetration test (VPT) for the investigation of cohesionless soils in the field". In: *Soil Dynamics and Earthquake Engineering* 126, p. 105760.

[7] Dawson, S. T. M., M. S. Hemati, M. O. Williams, and C. W. Rowley (2016). "Characterizing and correcting for the effect of sensor noise in the Dynamic Mode Decomposition". In: *Experiments in Fluids* 57.3, p. 42.

[8] Dierssen, G. (1994). "Ein bodenmechanisches Modell zur Beschreibung des Vibrationsrammens in körnigen Böden". PhD thesis. Universität Karlsruhe.

[9] Dylewsky, D., E. Kaiser, S. L. Brunton, and J. N. Kutz (2022). "Principal component trajectories for modeling spectrally continuous dynamics as forced linear systems". In: *Physical Review E* 105.1, p. 015312.

[10] Fawaz, H. I., G. Forestier, J. Weber, L. Idoumghar, and P.-A. Muller (2019). "Deep learning for time series classification: A review". In: *Data Mining and Knowledge Discovery* 33, pp. 917–963.

[11] Holeyman, A. and V. Whenham (2017). "Critical review of the Hypervib1 model to assess pile vibro-drivability". In: *Geotechnical and Geological Engineering* 35.5, pp. 1933–1951.

[12] Kutz, J., S. Brunton, B. Brunton, and J. Proctor (2016). *Dynamic Mode Decomposition: Data-driven modeling of complex systems*. Society of Industrial and Applied Mathematics.

[13] LeCun, Y., L. Bottou, Y. Bengio, and P. Haffner (1998). "Gradient-based learning applied to document recognition". In: *Proceedings of the IEEE* 86.11, pp. 2278–2324.

[14] Martinelli, M., A. Tsetas, A. B. Fărăgău, A. Metrikine, and A. Tsouvalas (2025). "Soil resistance during vibratory pile installation: Experimental findings from lab-scale tests". In: *Soil Dynamics and Earthquake Engineering* 199, p. 109692.

[15] Massarsch, K. R., B. H. Fellenius, and A. Bodare (2017). "Fundamentals of the vibratory driving of piles and sheet piles". In: *Geotechnik* 40.2, pp. 126–141.

[16] Massarsch, K. R., C. Wersäll, and B. H. Fellenius (2021). "Dynamic ground response during vibratory sheet pile driving". In: *Journal of Geotechnical and Geoenvironmental Engineering* 147.7, p. 04021043.

[17] — (2022). "Vibratory driving of piles and sheet piles – State of practice". In: *Proceedings of the Institution of Civil Engineers – Geotechnical Engineering* 175.1, pp. 31–48.

[18] Rausche, F. (2002). "Modeling of vibrator driving". In: *Proceedings of the International Conference on Vibratory Pile Driving and Deep Soil Compaction*. Balkema, pp. 21–32.

[19] Rodger, A. A. and G. S. Littlejohn (1980). "A study of vibratory driving in granular soils". In: *Géotechnique* 30.3, pp. 269–293.

[20] Shioi, A. and Y. Otake (2026). "A fundamental study on introducing time-delay embedding to enhance the applicability of Dynamic Mode Decomposition for seismic response analysis". In: *Soil Dynamics and Earthquake Engineering* 200, p. 109806.

[21] Shioi, A., Y. Otake, I. Yoshida, S. Muramatsu, and S. Ohno (2023). "Data-driven approximation of geotechnical dynamics to an equivalent single-degree-of-freedom vibration system based on Dynamic Mode Decomposition". In: *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards* 17.1, pp. 77–97.

[22] Viking, K. (2002). "Vibro-driveability – A field study of vibratory driven sheet piles in non-cohesive soils". PhD thesis. Royal Institute of Technology (KTH).

[23] Vogelsang, J., G. Huber, and T. Triantafyllidis (2017). "Experimental investigation of vibratory pile driving in saturated sand". In: *Holistic simulation of geotechnical installation processes: Theoretical results and applications*. Springer, pp. 101–123.

[24] Wang, Z., W. Yan, and T. Oates (2017). "Time series classification from scratch with deep neural networks: A strong baseline". In: *2017 International Joint Conference on Neural Networks*, pp. 1578–1585.

[25] Williams-Riquer, F., M. A. Abdennadher, A. Chmelnizkij, J. Grabe, and K. Smarsly (2025). "Modeling and predicting soil vibrations during vibrodriving installation using noisy data and Dynamic Mode Decomposition". In: *Experimental vibration analysis for civil engineering structures*. Vol. 164. Lecture Notes in Civil Engineering, pp. 710–718.

[26] Williams-Riquer, F., A. Chmelnizkij, M. A. Abdennadher, D. A. Dao, A. Datar, J. Grabe, K. Smarsly, A. Shioi, and O. Otake (2026). "Dynamic Mode Decomposition with time-delay embedding for noise-robust modelling of soil vibrations during vibro-driven pile installation". Manuscript submitted for publication.

[27] Williams-Riquer, F., A. Chmelnizkij, D. Alkateeb, and J. Grabe (2024). "Prediction of induced soil vibration during pile vibrodriving using Dynamic Mode Decomposition (DMD)". In: *Journal of Physics: Conference Series* 2909.1, p. 012002.

Author

Francisco Williams-Riquer
Institute of Geotechnical Engineering and Construction Management
Hamburg University of Technology (TUHH)
Schellerdamm 22
21079 Hamburg, Germany
E-mail: francisco-williams@tuhh.de

Mechthild Cramer
Institute of Geotechnical Engineering and Construction Management
Hamburg University of Technology (TUHH)
Schellerdamm 22
21079 Hamburg, Germany
E-mail: mechthild.cramer@tuhh.de

Physics enhanced and data driven modeling applied to structural and wave dynamics

Svenja Ehlers, Kerstin Kracht, Norbert Hoffmann

Abstract: Physics-informed machine learning combines measurement data with governing physical equations to solve forward and, in particular, inverse problems in engineering. By embedding physical knowledge directly into the learning process, physics-informed neural networks (PINNs) enable physically consistent solutions even for sparse or incomplete observations. This paper introduces the basic concept of PINNs and demonstrates their application to two representative inverse problems: phase-resolved ocean wave reconstruction from sparse wave buoy measurements and distributed Young's modulus identification of a vibrating canvas painting from full-field vibration data. Although originating from different engineering disciplines, both examples illustrate the same general methodology of combining partial differential equations with measurement data, which is readily transferable to a wide range of applications and disciplines.

1 Introduction

Many geophysical and engineering systems are complex dynamical processes whose accurate identification, state reconstruction and prediction are essential for safe operation, control, and structural integrity. Examples range from nonlinear ocean waves to structural vibrations in mechanical components and cultural heritage objects such as canvas paintings. Despite their different physical nature, these systems share a common challenge: their dynamics evolve in space and time, while available measurements are often sparse, indirect, and affected by noise. This results in inverse problems, where either the full system state or unknown system parameters must be inferred from incomplete measurements.

Traditionally, such problems are addressed using physics-based numerical models derived from established governing equations, often formulated as partial differential equations (PDEs). In ocean wave dynamics, for example, nonlinear potential flow models can provide a highly accurate phase-resolved modeling approach, but require physically consistent initial conditions reconstructed from sparse measurements. Similarly, in structural dynamics, identifying distributed material properties or internal vibration states from limited sensor data often requires repeated forward simulations and iterative parameter optimization. While these approaches are physically correct, they often result in computationally expensive optimization problems.

From a mathematical perspective, many inverse problems can be interpreted as approximation tasks: the goal is to identify an unknown mapping from sparse observations to hidden states or parameters. This viewpoint naturally motivates the use of machine learning (ML), and in particular deep neural networks, as universal function approximators capable of representing highly nonlinear dependencies (Hornik et al., 1989). However, purely data-driven ML approaches typically require large amounts of high-quality labeled data, which are rarely available in scientific and engineering applications. In many cases, full-field ground-truth data cannot be measured directly and can only be approximated through expensive simulations or controlled experiments. Moreover, unconstrained black-box models may produce solutions that violate known physical principles.

Scientific machine learning (SciML) has emerged as a promising framework to address these limitations by combining data-driven machine learning with prior physical knowledge. Within this broader paradigm, one common approach is hybrid modeling, where machine learning models and physics-based solvers are treated as separate interacting components, for example by learning unresolved closure terms, correcting model errors, or accelerating computational subroutines (Willard et al., 2022). Another important branch focuses on system identification and equation discovery, where machine learning is used to infer unknown parameters, constitutive relations, or even governing equations directly from observational data.

In contrast, physics-informed machine learning (PIML) embeds physical knowledge directly into the learning process itself, as shown in Fig. 1. Here, governing equations, boundary conditions, and conservation principles are integrated into the optimization problem as observational, inductive or learning biases that constrain the admissible solution space. This tighter integration enables the learning of physically plausible solutions even in regimes with sparse or incomplete data (Karniadakis et al., 2021). Among the most prominent realizations of the learning bias concept are physics-informed neural networks (PINNs), which encode physical consistency through residual-based loss formulations (Raissi et al., 2019). In PINNs, the unknown solution of a PDE is directly parameterized by a neural network, transforming the problem of solving differential equations into an optimization problem over network parameters. In addition to fitting available measurements, the training objective includes residual terms of the governing equations and associated initial and boundary conditions, which are evaluated through automatic differentiation. This allows the network to interpolate between sparse observations while remaining consistent with the underlying physics. As a result, PINNs are particularly well suited for inverse problems in dynamical systems, where observational data alone are insufficient to uniquely determine the solution.

This work explores physics-enhanced and data-driven modeling for two representative classes of dynamical systems: phase-resolved ocean wave reconstruction from sparse measurements and vibration-based state and parameter identification in structural systems, exemplified by the dynamic behavior of canvas paintings. Although these applications differ in scale and governing physics, they illustrate a common methodological principle: leveraging physical structure within machine learning to solve challenging inverse problems more efficiently and robustly than conventional approaches alone.

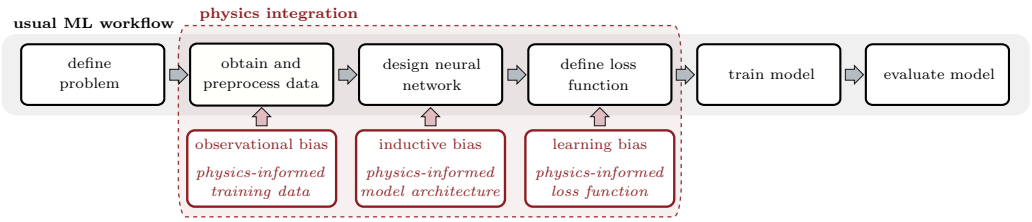


Figure 1: Physics-informed ML may integrate prior physical knowledge through three bias types in the conventional ML workflow: observational bias, where physics is implicitly introduced via physics-consistent training data (e.g from numerical simulations); inductive bias, where physical principles are embedded explicitly into the neural network architecture; and learning bias, where governing equations, boundary conditions, or conservation laws are incorporated directly into the optimization objective through physics-informed loss functions.

2 Example from wave dynamics

The phase-resolved modeling and prediction of ocean waves requires fully-resolved initial conditions. In practice, however, only sparse or indirect measurements of the sea surface are available from typical wave sensors. For example, wave buoys record time series of the surface elevation only at a few discrete spatial locations. Consequently, the reconstruction of complete ocean wave fields η from sparse elevation measurements η_{m} constitutes a classical inverse problem in fluid dynamics, commonly referred to as wave data assimilation, which may be formulated as the mapping

$$\eta_{\mathrm{m}} \in \mathbb{R}^{n_{\mathrm{buoy}} \times n_t} \rightarrow \eta \in \mathbb{R}^{n_x \times n_t}. \tag{1}$$

Here n_x and n_t denote the spatial and temporal grid points, respectively, and $n_{\mathrm{buoy}} \ll n_x$ denotes the number of available wave buoys.

Assuming incompressible, inviscid, and irrotational flow, ocean wave evolution can be described by potential flow theory. In this framework, the fluid velocity is represented by a scalar velocity potential $\Phi(x,t,z)$, which satisfies the Laplace equation inside the fluid domain

$$\nabla^2 \Phi = \frac{\partial^2 \Phi}{\partial x^2} + \frac{\partial^2 \Phi}{\partial z^2} = 0, \tag{2}$$

while nonlinear kinematic and dynamic free-surface boundary conditions govern the free surface elevation $\eta(x,t)$ and a bottom boundary condition imposes an impermeable sea bed at depth d

$$\frac{\partial \eta}{\partial t} + \frac{\partial \eta}{\partial x}\frac{\partial \Phi}{\partial x} - \frac{\partial \Phi}{\partial z} = 0 \quad \text{on } z = \eta(x,t) \tag{3}$$

$$\frac{\partial \Phi}{\partial t} + g\eta + \frac{1}{2}\left[\left(\frac{\partial \Phi}{\partial x}\right)^2 + \left(\frac{\partial \Phi}{\partial z}\right)^2\right] = 0 \quad \text{on } z = \eta(x,t) \tag{4}$$

$$\frac{\partial \Phi}{\partial z} = 0 \quad \text{on } z = -d. \tag{5}$$

Although these equations are well established for forward problems, solving the inverse problem with conventional variational or sequential data assimilation techniques requires repeated forward simulations and iterative optimization, which is computationally expensive (Köllisch et al., 2018).

To overcome this limitation, a physics-informed neural network (PINN) framework is proposed in Ehlers et al. (2025), shown in Fig. 2. Instead of solving the governing equations numerically on a predefined grid, the unknown solution fields are represented by two coupled neural networks. The first network approximates the free-surface elevation $\tilde{\eta}(x,t)$, while the second represents the velocity potential $\tilde{\Phi}(x,z,t)$. This dual-network setup is essential because both quantities are physically coupled through the nonlinear free-surface boundary conditions (Eq. (3- 4)). The experimental setup used for validation is illustrated in Fig. 3. A wave flume is equipped with synchronized cameras that provide highly resolved measurements of the wave surface, serving as the reference solution η_{camera} after PINN training. To emulate realistic offshore measurements, only three buoys are placed to extract measurements η_{m} as input for PINN training. The buoy measurements are incorporated into the $\tilde{\eta}$-network as hard constraints, ensuring exact agreement at measurement locations. To reconstruct the wave field between these measurements, the PINN is trained by minimizing the residuals of the governing equations as loss terms ($\mathcal{L}_{\text{Lap}}, \mathcal{L}_{\text{BC,kin}}, \mathcal{L}_{\text{BC,dyn}}, \mathcal{L}_{\text{BC,bot}}$), where the derivatives are evaluated efficiently by automatic differentiation. The physics therefore acts as a regularization mechanism, constraining the solution space to find the physically most consistent solution field.

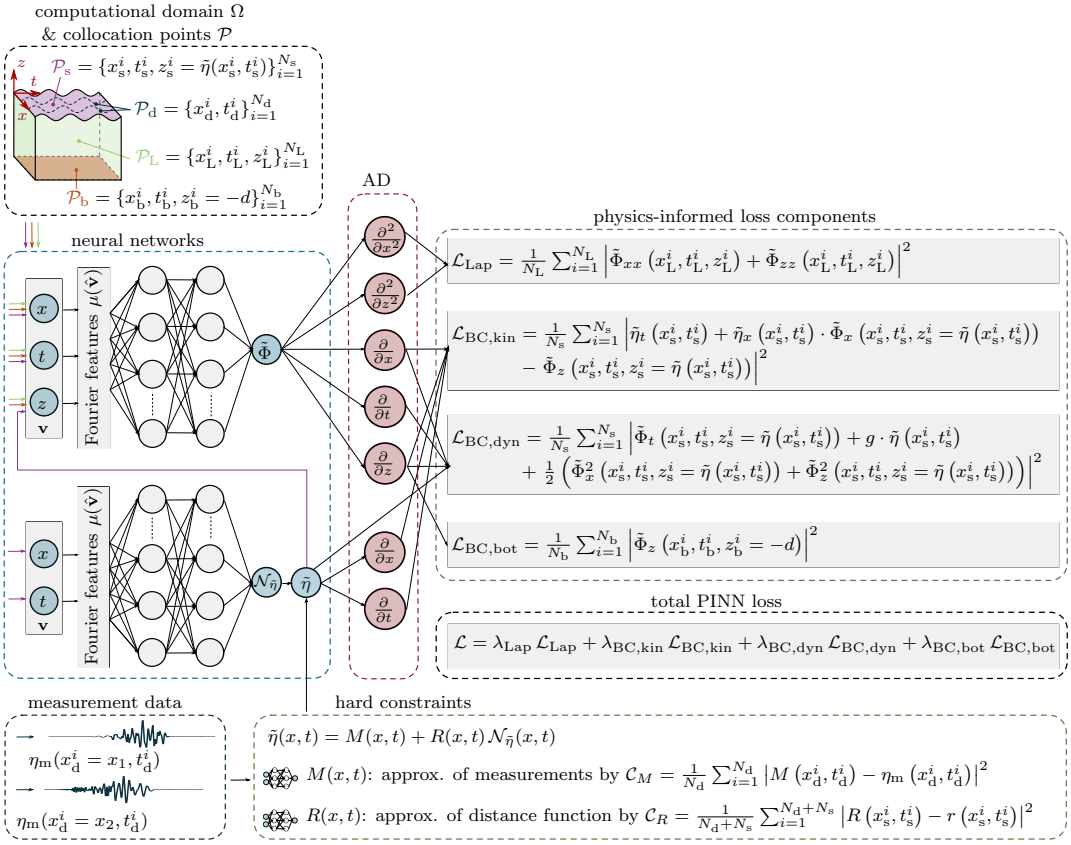


Figure 2: PINN framework for phase-resolved ocean wave data assimilation with two coupled neural networks to approximate the free-surface elevation $\tilde{\eta}(x,t)$ and the velocity potential $\tilde{\Phi}(x,z,t)$, respectively. Sparse wave buoy measurements η_{m} are combined with the governing equations of potential flow theory (Eqs. (2-5)) to reconstruct the complete wave field. Automatic differentiation is employed to evaluate these residuals, which are incorporated as physics-informed loss terms. During training, the network parameters are optimized such that the reconstructed wave field simultaneously satisfies the sparse measurements and the governing physical equations.

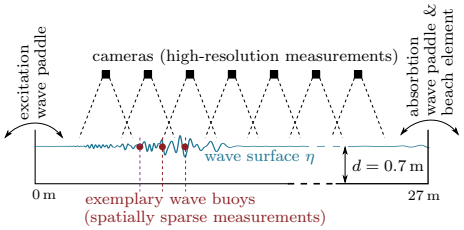


Figure 3: Experimental wave tank setup for wave data acquisition, which is equipped with synchronized cameras providing highly-resolved reference measurements of the free-surface elevation. Moreover, three buoys extract sparse wave surface measurements η_m that serve as input for the PINN. The complete camera-based wave field η_camera is used as the reference solution for evaluating the PINN reconstruction accuracy of $\tilde{\eta}$ after training.

The reconstruction results are shown in Fig. 4. Despite receiving only three sparse buoy time series η_m as input (column 2), the PINN accurately reconstructs the complete spatio-temporal wave field $\tilde{\eta}$ (column 3), closely matching the camera-based reference measurements η_camera (column 1), which is especially visualized in the cross-sections in the last column. Moreover, the PINN simultaneously infers the complete corresponding velocity potential $\tilde{\Phi}$ throughout the fluid domain (row 2), although no potential measurements are available from most experiments. This latent quantity enables the computation of physically relevant variables such as fluid velocities and dynamic pressures, which are important i.e. for wave-structure interaction analyses.

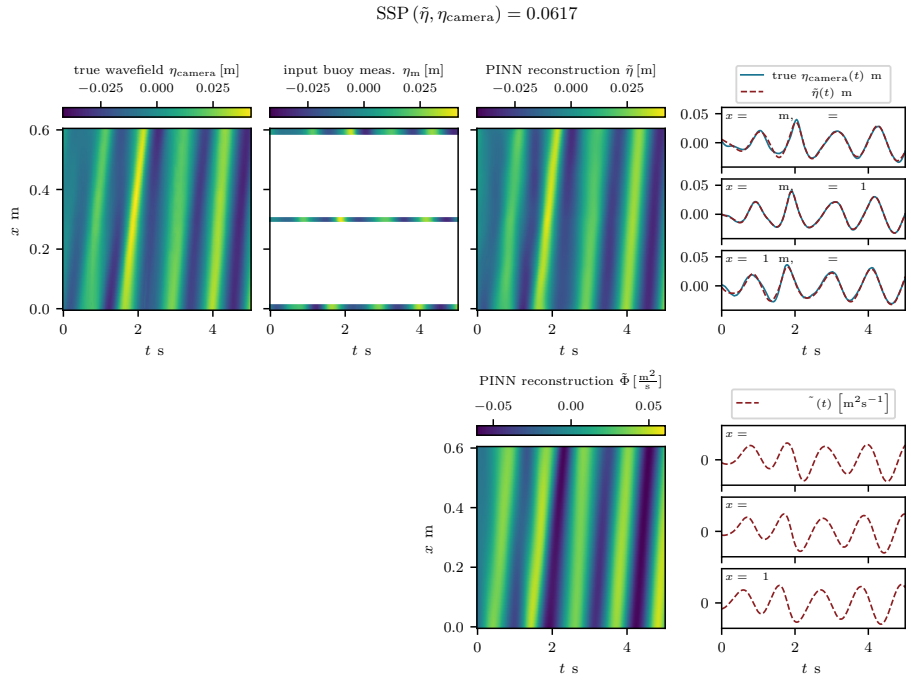


Figure 4: Reconstruction results of the PINN framework for wave data assimilation. The first column shows the camera-based reference wave field η_camera, while the second column depicts the three sparse buoy measurements η_m provided as the only input to the PINN. Despite these limited buoy data, the reconstructed free-surface elevation $\tilde{\eta}$ in the third column closely matches the reference solution η_camera. In addition, the PINN also infers the latent velocity potential $\tilde{\Phi}$, which is not available from measurements. Some temporal cross sections in the last column further demonstrate the close agreement between the reconstructed and reference wave elevations at different locations.

This example demonstrates one of the key strengths of physics-informed learning: the simultaneous reconstruction of observable ($\tilde{\eta}$) and unobservable ($\tilde{\Phi}$) quantities from sparse measurements by using the governing equations. Importantly, this concept is not restricted to fluid dynamics. The same methodology can be transferred to structural dynamics, where unknown material properties rather than unknown flow fields become the target of the inverse problem.

3 Example from structural dynamics

Besides fluid dynamics, many problems in structural dynamics can also be formulated as inverse problems. A representative example is the vibration analysis of canvas paintings in cultural heritage preservation. Unlike ocean wave applications, where only sparse measurements are typically available, the dynamic response w_m of a canvas painting can nowadays be measured with high spatial resolution, i.e. using optical measurement techniques. A suitable measurement method and setup was developed in Kracht (2011). However, the underlying material properties remain largely unknown and are commonly approximated by literature values and/or assumed to be homogeneous across the entire canvas painting. In reality, however, canvas paintings are highly heterogeneous structures due to pre-tensioning and as they may consist of multiple materials, including canvas, wooden stretcher bars, primer, various paint layers, and local damages as well as restoration treatments. Consequently, mechanical properties such as the Young's modulus may vary spatially across the artwork and cannot generally be represented by a single constant value. The inverse problem therefore shifts from reconstructing the system state to identifying distributed material parameters and can be formulated as

$$w_\mathrm{m} \in \mathbb{R}^{n_x \times n_y \times n_t} \rightarrow E \in \mathbb{R}^{n_x \times n_y}, \tag{6}$$

where w_m denotes the dense measured transverse vibration response and E the unknown spatial distribution of the Young's modulus. Here, n_x, n_y, and n_t represent the number of spatial and temporal sampling points, respectively.

In this case, a canvas mounted on a stretcher and primed multiple times was examined. The size of the freely vibrating surface of the test painting is $40\,\mathrm{cm} \times 30\,\mathrm{cm}$ (width $\times$ height). Before the vibration test, the object was subjected to a shock sequence on a hydropulse system to release the prestress. The measurement grid for recording spatially high-resolution vibration responses of the test painting is shown in Fig. 5. During the tests, the object is excited at one reference point at the stretcher. The response of the canvas is sequentially measured at 100×73 locations with a triangulation laser.
In the present example, the dynamic behavior of the painting is approximated as a thin Kirchhoff plate according to Kracht (2011)

$$\frac{Eh^3}{12\,(1-\nu^2)}\left(\frac{\partial^4 w}{\partial x^4} + 2\frac{\partial^4 w}{\partial x^2 \partial y^2} + \frac{\partial^4 w}{\partial y^4}\right) + \rho h \frac{\partial^2 w}{\partial t^2} = 0, \tag{7}$$

where w denotes the transverse displacement, E is the Young's modulus, $h = 2.5\mathrm{mm}$ is the average thickness, $\nu = 0.3$ is the average contraction coefficient, and $\rho = 450\frac{\mathrm{kg}}{\mathrm{mm}^2}$ is the density.

Classical identification approaches typically assume a homogeneous Young's modulus E and determine a single effective parameter through repeated forward simulations until the simulated vibration response matches the measurements. Such approaches are unable to capture local variations caused by the heterogeneous composition of the artwork.

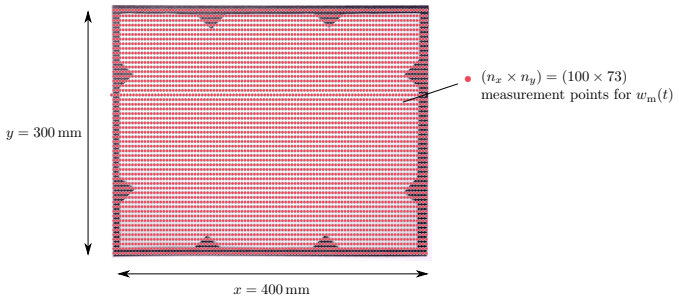


Figure 5: Painting used for vibration measurements. Although the canvas is almost uniformly painted and contains no intentionally introduced inhomogenities, local pre-stresses, wrinkles, and tension variations are apparent and produce spatially varying effective stiffness. The painting is sequentially excited at each point of the 100×73 measurement grid, where the transverse vibration response $w_{\mathrm{m}}(x,y,t)$ is subsequently used as input for the PINN.

To address this limitation, the physics-informed neural network shown in Fig. 6 is employed. Instead of learning only the displacement field $\tilde{w}(x,y,t)$, a second neural network is introduced to approximate the unknown spatial distribution of the Young's modulus $\tilde{E}(x,y)$. To obtain a physically meaningful optimization process, the Young's modulus is initially assumed to be homogeneous and initialized with the conventionally identified value of $E = 225\,\mathrm{N/mm^2}$. During the first 7000 training epochs, only the displacement network is optimized, allowing $\tilde{w}$ to accurately reproduce the measured vibration response while keeping the material properties fixed. After this warm-up phase, the Young's modulus network is activated and both neural networks are optimized simultaneously. During training, the densely measured vibration response w_{m} is incorporated as a data constraint through the loss term $\mathcal{L}_{\mathrm{data}}$, ensuring agreement between the predicted and measured displacement field. Simultaneously, the predicted Young's modulus field is embedded into the Kirchhoff plate equation, whose residual is evaluated using automatic differentiation and minimized via the physics-informed loss term $\mathcal{L}_{\mathrm{Kirchhoff}}$. Consequently, the vibration field and the spatially varying material properties are reconstructed simultaneously within a single optimization problem.

The PINN result for the identified $\tilde{E}$-distribution is presented in Fig. 7 and reveals distinct spatial variations across the painting. At the same time, the reconstructed displacement field $\tilde{w}$ accurately reproduces the measured vibration responses w_{m}, as demonstrated by the displacement profiles along representative cross-sections at $y = 75\,\mathrm{mm}$ and $y = 225\,\mathrm{mm}$, shown in Fig. 8 and Fig. 9, respectively. The identified $\tilde{E}$-distribution is further supported by the measured vibration amplitudes along the selected cross-sections. As shown in Fig. 7, the average $\tilde{E}$ is higher along the cross-section at $y = 75\,\mathrm{mm}$ than at $y = 225\,\mathrm{mm}$. Owing to the increased bending stiffness associated with a higher E, smaller vibration amplitudes are expected under identical excitation conditions. This behavior is confirmed by the displacement profiles, where the maximum amplitudes along the cross-section at $y = 75\,\mathrm{mm}$ (Fig. 8) are lower on average than those observed at $y = 225\,\mathrm{mm}$ (Fig. 9), which further proves the identified material distribution.

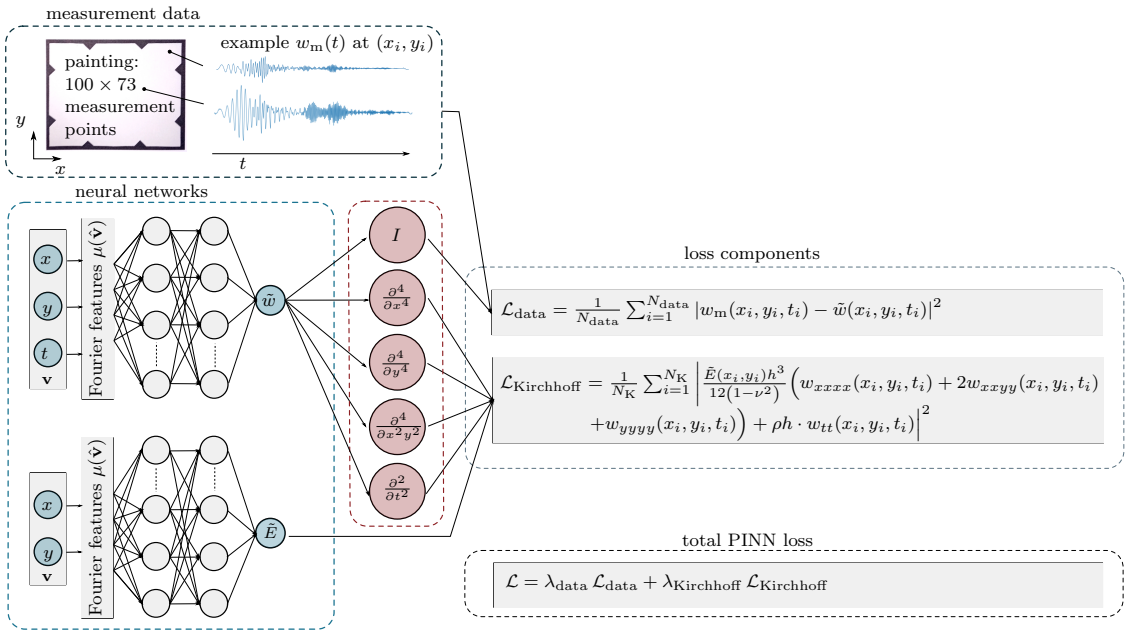


Figure 6: PINN framework for distributed material parameter identification. Two coupled neural networks approximate the vibration response $\tilde{w}(x,y,t)$ and the spatial Young's modulus distribution $\tilde{E}(x,y)$. Dense vibration measurements w_{m} are incorporated as data constraints, while the residual of the Kirchhoff plate equation is included as a physics-informed loss term. During training, both neural networks are optimized simultaneously such that the reconstructed vibration field agrees with the measurements while satisfying the governing equation.

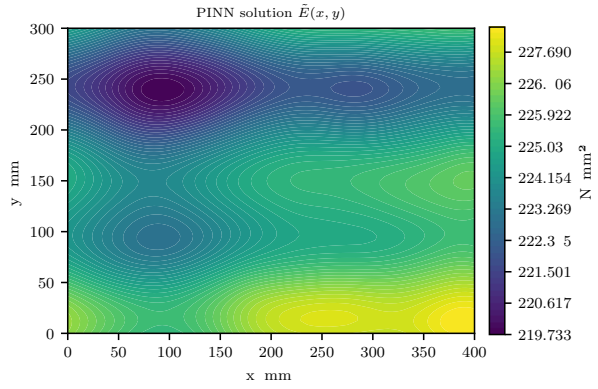


Figure 7: PINN reconstruction of the spatial Young's modulus distribution $\tilde{E}(x,y)$ obtained from the measured vibration responses. Starting from the homogeneous initial estimate of $E = 225\,\mathrm{N/mm^2}$, the PINN identifies spatial variations across the painting.

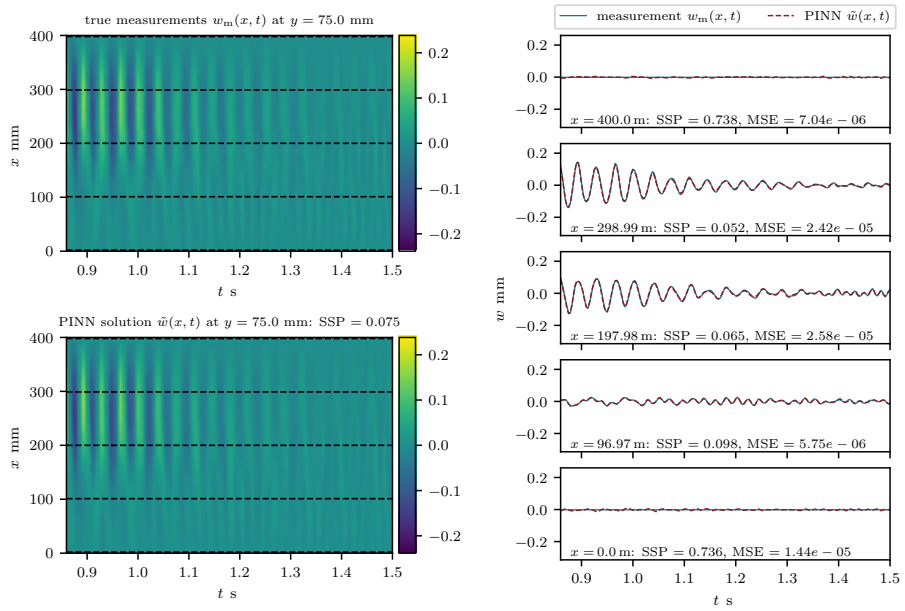


Figure 8: Comparison between the measured vibration response w_m (top left) and the reconstructed PINN solution $\tilde{w}$ (bottom left) along the cross-section at $y = 75\,\mathrm{mm}$. The right column shows representative temporal displacement histories at selected spatial locations (x), demonstrating the close agreement between measurements and the reconstructed displacement field. Owing to the higher average Young's modulus identified along this cross-section (Fig. 7), the vibration amplitudes are comparatively small.

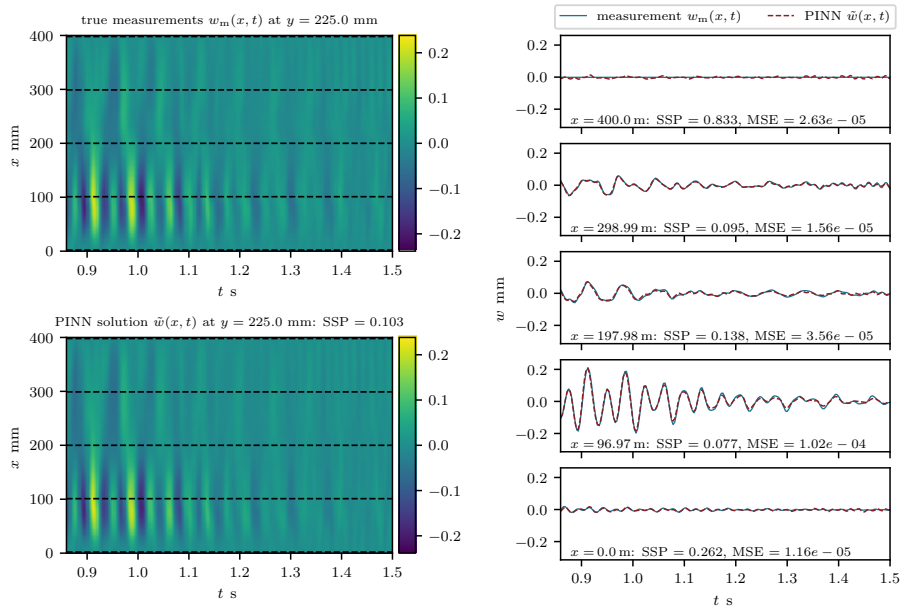


Figure 9: Comparison between the measured vibration response w_m (top left) and the reconstructed PINN solution $\tilde{w}$ (bottom left) along the cross-section at $y = 225\,\mathrm{mm}$. The right column shows representative temporal displacement histories at selected spatial locations (x), demonstrating the close agreement between measurements and the reconstructed displacement field. Compared to the cross-section at $y = 75\,\mathrm{mm}$, the lower average Young's modulus at $y = 225\,\mathrm{mm}$ (Fig. 7) results in reduced bending stiffness and larger maximum vibration amplitudes.

Together, these results demonstrate that physics-informed learning is not restricted to state reconstruction but can be readily extended to simultaneous distributed parameter identification within a single optimization framework. Such capabilities are attractive for structural health monitoring, when destructive material testing is impossible. Taken together, the wave and structural dynamics examples therefore illustrate a general physics-informed learning paradigm that is transferable to a wide range of inverse problems governed by partial differential equations.

4 Conclusion and outlook

This work presented physics-informed neural networks as a promising general framework for solving inverse problems in dynamical systems. Rather than replacing established physical models, PINNs combine governing equations with measurement data to infer unknown system states or parameters in a physically consistent manner. Two representative applications were discussed to illustrate this concept. In ocean wave dynamics, sparse wave buoy measurements were assimilated to reconstruct the complete free-surface elevation together with the latent velocity potential. In structural dynamics, the same methodology was transferred to identify the spatially varying Young's modulus of a painting directly from measured painting vibration responses. Although these applications originate from different engineering disciplines, they demonstrate the common capability of PINNs to exploit physical knowledge for solving challenging inverse problems under potentially limited measurement data.

Ultimately, the presented examples from wave and structural dynamics should not be regarded as isolated applications, but rather as representative case studies demonstrating a transferable methodology. Whenever the governing physics is (at least partially) known and measurements are (at least sparsely ord incompletely) available, physics-informed machine learning provides a systematic framework for combining measurement data with established physical models. This paradigm thus is also particularly promising for geotechnical engineering, where heterogeneous subsurface materials and limited accessibility often prevent direct characterization of relevant properties. PINNs may enable the identification of distributed material parameters, permeability fields, or evolving subsurface states. As sensing technologies and computational resources continue to advance, physics-informed machine learning is expected to become an increasingly important component of digital twins and data-driven decision support also for geotechnical systems.

References

S. Ehlers, N. Hoffmann, T. Tang, A. H. Callaghan, R. Cao, E. M. Padilla, Y. Fang, and M. Stender, "Physics-informed neural networks for phase-resolved data assimilation and prediction of nonlinear ocean waves," *Physical Review Fluids*, vol. 10, no. 9, p. 094901, 2025. doi: 10.1103/ytyy-pvys.

K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal

approximators," *Neural Networks*, vol. 2, no. 5, pp. 359–366, 1989. doi: 10.1016/0893-6080(89)90020-8.

G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, "Physics-informed machine learning," *Nature Reviews Physics*, vol. 3, no. 6, pp. 422–440, 2021. doi: 10.1038/s42254-021-00314-5.

N. Köllisch, J. Behrendt, M. Klein, and N. Hoffmann, "Nonlinear real time prediction of ocean surface waves," *Ocean Engineering*, vol. 157, pp. 387–400, 2018. doi: 10.1016/j.oceaneng.2018.03.048.

K. Kracht, "Die Untersuchung des Schwingungsverhaltens von Ölgemälden in Abhängigkeit der Alterung," *Dissertation*, Shaker, Düren, 2011.

M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," *Journal of Computational Physics*, vol. 378, pp. 686–707, 2019. doi: 10.1016/j.jcp.2018.10.045.

J. Willard, X. Jia, S. Xu, M. Steinbach, and V. Kumar, "Integrating scientific knowledge with machine learning for engineering and environmental systems," *ACM Computing Surveys*, vol. 55, no. 4, article 66, pp. 1–37, 2022. doi: 10.1145/3514228.

Authors

M. Sc. Svenja Ehlers
Institut für Strukturdynamik (M-14)
Technische Universität Hamburg
Schlossmühlendamm 30
21073 Hamburg
e-mail: svenja.ehlers@tuhh.de

Dr.-Ing. (habil.) Kerstin Kracht
Institut für Strukturdynamik (M-14)
Technische Universität Hamburg
Schlossmühlendamm 30
21073 Hamburg
e-mail: kerstin.kracht@tuhh.de

Prof. Dr. rer. nat. Norbert Hoffmann
Institut für Strukturdynamik (M-14)
Technische Universität Hamburg
Schlossmühlendamm 30
21073 Hamburg
e-mail: norbert.hoffmann@tuhh.de

Application of the phase-field method for resource-efficient ground improvement through shear band strengthening

Elnaz Hadjiloo, Jürgen Grabe

Abstract: Shallow foundations fail through localised shearing. Strengthening only the potential shear band instead of the whole failure body is a patented alternative to conventional ground improvement, and in model tests it raised the failure load by 76 %. Its numerical realisation has so far used strengthening criteria, which are state measures and need a holding rule for the band to grow and strengthen at the same time, whereas the present work makes growth and irreversibility part of the model. A phase-field model is coupled to a hypoplastic constitutive law and driven by the acoustic tensor, so that onset, position and orientation of the band follow from the loss of ellipticity. Mesh objectivity is shown at element and structure level, the driving force is compared with two alternatives, and the strengthening effect is reproduced at a strip footing.

1 Introduction

Geotechnical structures rarely fail everywhere at once. Under increasing load the deformation concentrates into thin shear zones, and the bearing capacity of a shallow foundation is governed by the mechanism that these zones form. Where that capacity has to be increased, established practice places structural elements such as piles, anchors or soil nails across the expected shear zone, or improves the ground over a large volume by jet grouting or injection. Both routes treat a volume that is large compared with the zone that actually governs failure.

Shear band strengthening follows the opposite logic. Only the beginning of the most probable shear band is strengthened, typically by a local cement injection, so that this band can no longer form and the mechanism is forced onto a kinematically less favourable path of higher capacity. The idea was developed and patented at Hamburg University of Technology (Grabe and Pucker, 2012). Seitz (2021) demonstrated it on model tests. Cement placed in the straight initial section of the shear band beneath a strip footing raised the failure load from a mean of 383 N to 674 N, an increase of 76 % against a scatter of 19 % in the reference tests.

Numerical work on the method has so far used a hypoplastic model with a continuous material transition from sand to cemented soil (Pucker, 2013), combined with strengthening criteria that select the elements to be treated (Hadjiloo and Grabe, 2023a, 2023b). The criteria are accumulated shear dissipation, degree of non-linearity, shear strain and mobilised friction angle, and increases in vertical force were obtained for treated areas of the foundation cross-section. Each of them is a state measure filtered by a threshold and describes the system at the moment it is evaluated. The band is grown by repeating the calculation on the strengthened system, and because a zone can appear uncritical once the injection is in place, that growth has to be secured by a holding rule which keeps strengthened points strengthened. Growth is therefore imposed from outside rather than following from the formulation. A second point is numerical. Localisation in a rate-independent continuum is mathematically ill-posed once ellipticity is lost, so the computed band width is set by the element size unless a regularisation is introduced (de Borst et al., 1993; Cui et al., 2023).

This contribution therefore couples a phase-field model of the shear band to the hypoplastic constitutive law and drives it with the acoustic tensor. The phase field supplies the regularisation, so the band width follows a material length scale rather than the mesh. The acoustic tensor supplies the criterion, so onset, position and orientation of the band are a prediction of the constitutive model rather than a calibrated threshold. The paper introduces the formulation and its implementation, shows that it is mesh-objective at element and at structure level, compares the acoustic driving force with two alternatives, and demonstrates at a strip footing that the strengthening effect is captured.

2 State of the art

2.1 Localisation and regularisation

The classical description of shear banding in frictional materials goes back to Rudnicki and Rice (1975) and Rice (1976). A band with normal n can form when the acoustic tensor $A(n) = n \cdot C \cdot n$, built from the tangent stiffness C, becomes singular. From that instant the boundary value problem loses ellipticity, the solution is no longer unique, and a standard discretisation localises the deformation into a single row of elements, so the dissipated energy scales with the element size. Regularisation restores a length scale. Non-local and gradient formulations, Cosserat continua and viscous regularisation are all used for this purpose in geomechanics (de Borst et al., 1993; Cui et al., 2023), and for hypoplastic models the polar extension is the established remedy (Tejchman and Bauer, 1996). The phase-field method belongs to the same family, but it introduces the length scale together with an explicit field variable for the damaged zone, which makes the band itself an output of the calculation.

2.2 Phase-field modelling of fracture

The phase-field approach originates in the variational formulation of brittle fracture by Francfort and Marigo (1998) and its regularised, numerically tractable form by Bourdin et al. (2000). A scalar field $\varphi \in [0,1]$ describes the transition from intact ($\varphi = 0$) to fully damaged ($\varphi = 1$) material over a regularisation length ℓ, and the surface energy of a discrete crack is replaced by a volume integral over the diffuse zone. Miehe et al. (2010a, 2010b) established the thermodynamically consistent multi-field formulation, the history field enforcing irreversibility and the operator-split solution scheme.

For geomaterials the method has been used mainly for brittle rock and for hydraulic fracturing, and has been extended towards frictional and plastic behaviour (Choo and Sun, 2018; Fei and Choo, 2020). Its use for granular soils described by an incrementally non-linear constitutive law, and specifically for the design of a ground improvement measure, has to the authors' knowledge not been reported. One implementation aspect makes such an application practicable. Navidtehrani et al. (2021) showed that the phase-field balance equation is formally identical to steady-state heat conduction, so that it can be solved in a commercial code through a user material subroutine alone (Section 3.4).

2.3 Shear band strengthening

The principle is shown in Figure 1a. For a strip footing, the shear band that would form under the lowest external load is identified and its straight initial section near the surface is strengthened, so that the critical mechanism is suppressed and a deeper mechanism with a higher failure load has to develop. Because only the initial section is treated, the intervention is

shallow, geometrically simple and small in volume, and this is what makes the method resource-efficient.

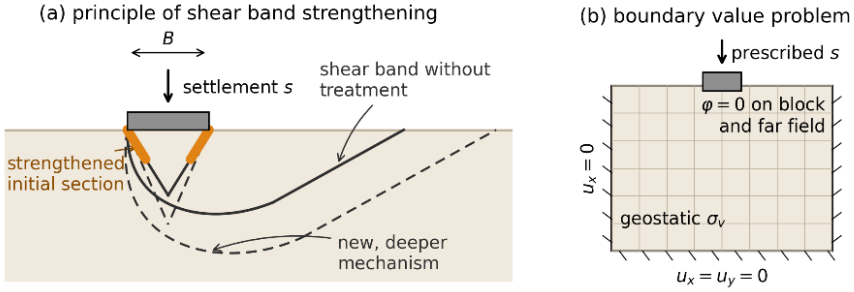


Figure 1: (a) Principle of shear band strengthening, after Hadjiloo and Grabe (2023a). Only the straight initial section of the wedge is treated, so a deeper mechanism has to develop. (b) Boundary value problem of Section 5.

3 Method

The phase-field method is not an alternative to the finite element method but an additional field that is solved with it (Ambati et al., 2015). In a standard finite element calculation the shear band is not an object of the model. It forms where the mesh allows it to form, it is one element row wide, it has to be read off the strain field afterwards, and the energy dissipated in it shrinks with the element size (de Borst et al., 1993). The phase field adds one unknown at every node, a damage variable φ between zero and one, and its balance equation carries a length ℓ, so the damaged zone has a width that belongs to the material rather than to the discretisation. Onset, position and orientation of the band are then results of the calculation, and the band can be passed directly to the strengthening map. What has to be supplied in return is a driving force that tells the field where damage is admissible, which for soil is the open question of Section 3.3.

3.1 Constitutive model and strengthening map

The soil is described by a hypoplastic constitutive law (von Wolffersdorff, 1996) in the extension of Pucker (2013), which interpolates linearly between an unstrengthened and a strengthened parameter set through a single scalar $\kappa \in [0,1]$. Here $\kappa = 0$ is the natural soil, $\kappa = 1$ the cemented material, and intermediate values are soil-cement mixtures. The strengthening geometry therefore enters the calculation as a field of κ, not as a separate material or a change of mesh. The map is generated from the computed band of the preceding run, as in the earlier work with strengthening criteria (Hadjiloo and Grabe, 2023a). This closes the design loop the method needs, because the calculation itself proposes where to inject.

3.2 Phase field and degradation

The step from a classical finite element calculation to a phase-field calculation is one additional term in the energy. In the regularised formulation of brittle fracture the energy of a body Ω with the strain energy density ψ, the degradation function $g(\varphi)$, the crack surface energy G_c and the regularisation length ℓ reads (Francfort and Marigo, 1998; Bourdin et al., 2000)

$$E(u, \varphi) = \int g(\varphi)\, \psi \, d\Omega + G_c \int [\varphi^2/(2\ell) + (\ell/2)|\nabla\varphi|^2] \, d\Omega \qquad (1)$$

The first integral is the mechanical part that the classical calculation already contains, reduced by g(φ) where damage has grown. The second integral is the fracture part, and it is the part that is added. It replaces the surface energy of a discrete crack by a volume integral over the diffuse zone and is what brings the length ℓ into the model (Miehe et al., 2010b; Ambati et al., 2015). In the discretisation it costs one degree of freedom per node and one further balance equation, and nothing else in the finite element problem changes. For the material used here the first integral cannot be evaluated in this form, because a hypoplastic law is a rate relation and has no strain energy density. The driving term is therefore replaced by the history field H of Section 3.3 while the fracture part is kept unchanged. With the AT2 choice the stationarity condition in φ gives the field equation solved in every increment,

$$G_c \, (\varphi/\ell - \ell \, \Delta\varphi) = 2 \, (1 - \varphi) \, H \tag{2}$$

where H is a history field that accumulates the driving force and is enforced to be non-decreasing, $H = \max(H_{old}, H_{new})$, so that damage does not heal (Miehe et al., 2010a). Irreversibility is thereby part of the formulation and does not have to be imposed from outside. The stiffness is degraded by $g(\varphi) = (1 - \varphi)^2 + k$ with a small residual k. A volumetric-deviatoric split after Amor et al. (2009) is applied so that only the deviatoric part of the response is degraded, because for soil the capacity to carry mean compressive stress must not be removed by shear damage. The spectral split of Miehe et al. (2010b) was tested and discarded, because under the isotropic confinement of a soil element all principal strains are compressive and the driving force collapses. In the biaxial test of Section 4 a modified isotropic degradation with a residual compression factor of 0.30 was used in the shear step, while Section 5 uses the Amor split throughout.

3.3 Driving force: the acoustic tensor

In the classical phase-field model of brittle fracture the driving force comes from the free energy of the material itself. The history field accumulates the tensile part of the elastic strain energy density, so that the crack grows where stored energy can be released (Miehe et al., 2010a). Neither ingredient is available here. A hypoplastic law is written as a rate relation and has no free energy from which a tensile part could be split off, and soil does not fail by opening in tension but by shearing under compression. A substitute is therefore required. The strengthening criteria of the earlier work are the obvious candidate, but as state measures they describe the current state rather than its evolution, they carry no information about the direction in which a discontinuity could form, and they have no relation to the loss of ellipticity that makes the computation mesh-dependent in the first place. The acoustic tensor addresses exactly that event.

A shear band is kinematically a jump of the deformation rate across a plane, a slip in some direction on a plane with a normal n. Requiring the tractions to stay continuous across that plane, together with the rate form of the constitutive law, leads to the classical localisation condition of Rudnicki and Rice (1975) and Rice (1976). A band with normal n becomes mechanically possible when the acoustic tensor formed from the tangent stiffness C loses its regularity,

$$\det A(n) = 0 \, , \quad A(n) = n \cdot C \cdot n \tag{3}$$

The tangent stiffness used here is the hypoplastic one, taken without the viscous regularization. The band normal is not known in advance, so in plane strain all orientations are scanned on a two degree grid, which is the standard evaluation of the localisation condition (Rice, 1976), and

the smallest determinant found is normalised by its elastic reference value. This leaves one scalar per integration point, which reads like a traffic light. Near one the material is far from localisation, approaching zero the bifurcation is close, and at or below zero a band is mechanically possible. The orientation at which the minimum occurs falls out of the same scan at no extra cost and is stored as a state variable, so the inclination of the band is a result of the calculation and not an input.

The source of the history field is assembled from this indicator by two factors with clearly separated roles. The first is a smooth switch on the indicator, evaluated on the already degraded material rather than on the intact one. It is closed while the material is far from localisation and opens as the indicator approaches zero, so that damage is admitted only where the constitutive model itself allows a discontinuity. The second factor vanishes in undamaged material and grows where the phase field is already present. It keeps the source from igniting spontaneously and lets the damage advance as a connected front from the existing band tip into material that the criterion has released. Both factors are non-negative and the source is a rate, so the history field can only increase and the dissipation inequality is satisfied by construction. The position and the width of the switch are numerical smoothing constants, declared as fixed and not fitted, and the only remaining calibration parameter is a single scale factor multiplying the crack surface energy.

3.4 Implementation and resolution requirement

The model is implemented in Abaqus/Standard. Following Navidtehrani et al. (2021), equation (2) is recast in the form of steady-state heat conduction and solved with the built-in coupled temperature-displacement element, CPE4T in plane strain. The nodal temperature is the phase field, the conductivity carries $G_c\,\ell$, and the volumetric heat source with its derivatives carries the right-hand side. No user element is required and one subroutine returns both mechanics and phase field, which is what makes it practicable to couple an existing hypoplastic implementation (Pucker, 2013) to a phase-field model.

This convenience has a price that has to be declared. The temperature degree of freedom is occupied by φ, so it is no longer available for a thermal field, and any other quantity that a legacy implementation transports through the same channel must be moved. In the code used here the material transition parameter κ was originally carried by the temperature. It is therefore separated from the degree of freedom and stored as a solution-dependent state variable, initialised element by element from the strengthening map. Temperature is φ and only φ, and κ is a field of its own. Without this separation the strengthening map and the damage field would overwrite one another.

The same length ℓ also sets the resolution requirement. The regularised band has to be covered by several elements, and the recommendations in the literature vary with the quantity of interest and reach up to about eight elements per length scale where the peak load matters (Ambati et al., 2015; Kristensen et al., 2021). That requirement is not adopted here but checked against the calculation itself in Section 4.

3.5 Comparison of driving forces

Two further driving forces were implemented and compared under otherwise identical conditions, varying only the driver flag. Driver A is the accumulated plastic shear dissipation, which stands for the strengthening criteria of the earlier work (Hadjiloo and Grabe, 2023a,

2023b), and driver B is the second-order work $d^2W = \dot{\sigma} : D$ along the realised loading path. Driver B evaluates stability only along the direction the solver has already taken, and for non-associated materials that path can remain stable while unstable directions exist next to it, so B stayed silent in all calculations. Table 1 summarises the comparison, and the evidence for its last row is given in Section 5.

Table 1: Comparison of the three driving forces at identical model, mesh and loading.

	A, shear dissipation	**B, second-order work**	**C, acoustic tensor**
Quantity	accumulated plastic shear work	sign of d^2W on the realised path	min over band normals of $\det(n \cdot C \cdot n)$
Question	where has shearing occurred?	is the realised path stable?	is a discontinuity possible, and in which direction?
Directional information	None	realised direction only	complete, θ_{crit} is the band angle
Under self-weight	source follows the stress level, hence a depth bias	no effect	barotropy stabilises, more confinement moves det A further from zero
Observed at the footing	forward loading, right place but about 30 times too weak. Reloading, floods to $\varphi = 0.41$	forward loading, no response. Reloading, frozen at the inherited seed	forward and reloading, right place, effective, grows and stays narrow

4 Element level: mesh objectivity

The requirement of Section 3.4 is checked here by asking whether the computed band keeps a width of its own when the mesh is refined. A plane-strain biaxial test (0.5 × 1.0 m, confining pressure 0.1 MPa, displacement-controlled shearing) serves as an element test bench (Figure 2), and three meshes with $h = \ell/2$, $\ell/4$ and $\ell/8$ are compared at $\ell = 0.05$ m. The soil is dense, $e_0 = 0.53$, so that the response shows dilatancy, a peak and softening.

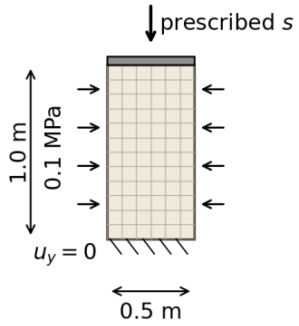


Figure 2: Biaxial element test. The specimen is compressed by a rigid platen while the lateral pressure is held constant.

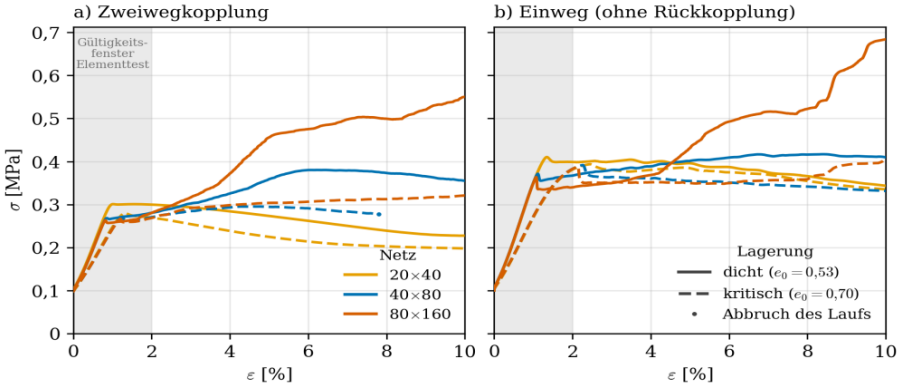


Figure 3: Biaxial test, nominal stress over nominal strain for the three meshes. (a) Two-way coupled, (b) one-way coupled control calculation without mechanical feedback. Solid lines are the dense configuration used here, dashed lines the critical one. The shaded band marks the validity window of the element test.

The load response is objective up to the peak (Figure 3). Beyond it the three meshes separate, because they are then in three different regimes, namely softening, peak and renewed hardening, so a stress read at a fixed strain is not a meaningful measure there. The cause lies in the specimen and not in the model. The band runs from corner to corner and ends in the loading platens, so sliding forces dilatancy against the cell pressure. A one-way coupled control calculation, in which φ is computed but does not act back on the mechanics, shows the effect most strongly, which rules out the coupling loop as its source. The load response is therefore read only up to band formation, below about 2 % strain.

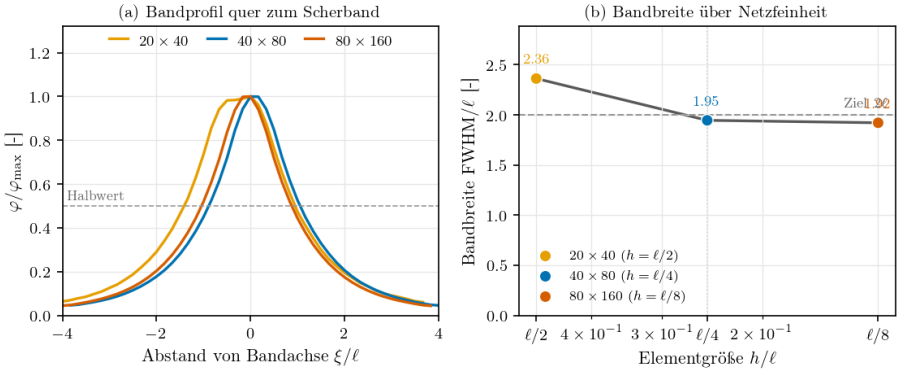


Figure 4: Band width in the dense configuration. (a) Phase-field profile across the band, normalised and centred at the peak. (b) Full width at half maximum over element size, target value 2ℓ.

The band width is the direct measure (Figure 4). The full width at half maximum of the phase-field profile across the band converges from above towards the theoretical value 2ℓ and stays within a few per cent of it from h ≤ ℓ/4, which is the range reported for phase-field models of brittle fracture (Ambati et al., 2015; Kristensen et al., 2021). The width is therefore tied to ℓ and not to h, which is the direct evidence that the regularisation works as intended, and it holds in the dense, dilatant case where it is actually needed. Figure 5 shows the band itself on the three meshes.

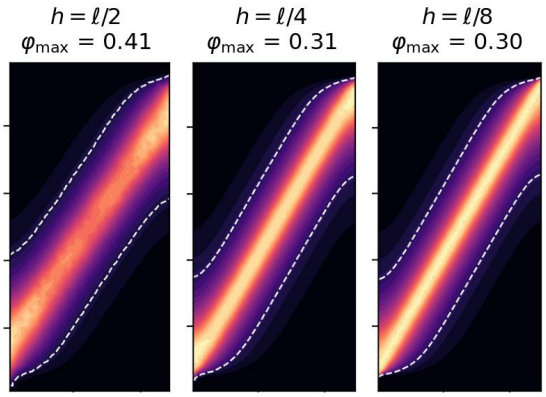


Figure 5: Phase field in the biaxial test at about 5 % strain for the three meshes, dense configuration with e0 = 0.53. The dashed line is the half-maximum contour used for the band width.

5 Boundary value problem: strip footing

The second question is whether this carries over to a structure, and which driving force is able to produce a band there. The footing is modelled in plane strain as a rigid, displacement-controlled block of width B = 0.5 m with an embedment of 0.10 m on a dense sand layer (e_0 = 0.53), geostatic initial stress, K_0 = 0.5, and φ = 0 on the block and the far-field boundaries (Figure 1b).

Mesh objectivity at the structure is reached in stages. The global bearing force at small settlement is already objective at h = ℓ, the area of the process zone needs h ≤ ℓ/2, and the peak load is the most demanding quantity. On a larger domain with four meshes the coarsest is unusable while the two finest differ by less than half a per cent. One element per length scale is therefore enough for the global stiffness but not for the failure load, which is consistent with the element-level requirement. The computed shear zone is shown in Figure 6.

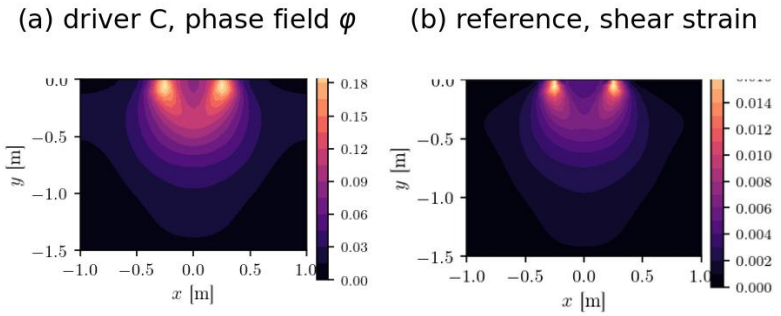


Figure 6: Shear zone at the strip footing. The phase field of driver C reproduces the two lobes at the foundation edges that the reference calculation without phase field shows in the accumulated shear strain, BVP according to Figure 1b

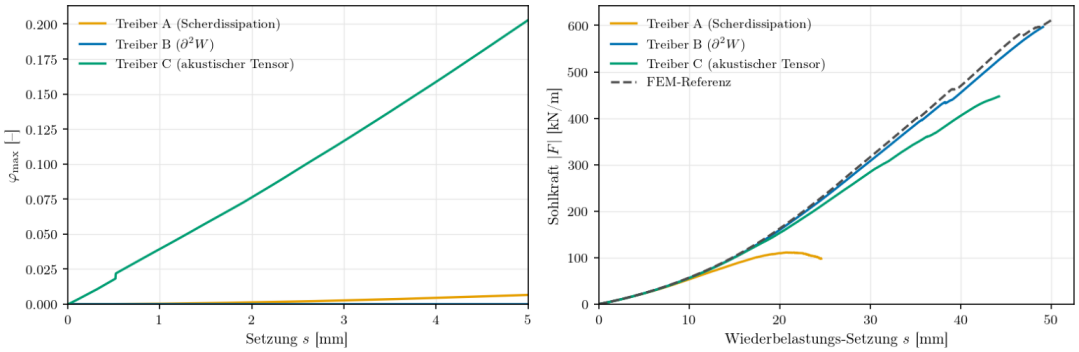


Figure 7: Comparison of the three driving forces at identical model, mesh and loading. (a) Maximum phase field over settlement in forward loading. (b) Load-settlement curves on reloading of a strengthened system, against the reference without phase field.

Figure 7a shows the forward calculation. Driver C grows to $\varphi_{max} = 0.203$ at 5 mm settlement and reduces the bearing force by 22 % against a reference in which the phase field is switched off. Driver A reaches a maximum about thirty times too small to change the mechanics, and driver B stays at exactly zero. Measured against the accumulated deviatoric shear strain of the reference, the neutral yardstick for where localisation actually occurs, driver C reproduces the two lobes at the foundation edges, while A is correctly placed but negligible. Driver C also inherits the band orientation from the acoustic tensor, 63° against the horizontal next to the Coulomb value of 61°, so the inclination lies inside the classical corridor (Vermeer, 1990) without any angle having been prescribed.

The decisive test is the reloading of an already strengthened system, the configuration for which shear band strengthening is designed (Figure 7b). On an identical strengthening map and with an identical seed, driver A collapses. The bearing capacity drops by about two thirds and the phase field floods the domain over several hundred nodes, which is diffusion instead of a band. At equal maturity driver A has damaged about three times the area of driver C, so the difference is not one of timing. Driver B lies on the reference curve, frozen at the inherited seed. Driver C loses about 13 % of the reference capacity and keeps the damaged area bounded, and it is the only formulation that both grows and stays narrow.

Figure 8 shows what the method is meant to produce. The unstrengthened footing reaches a peak of 35.1 kN/m at 28 mm settlement and softens afterwards. Strengthening the initial section of the band computed in that run and repeating the calculation raises the force at the same settlement by a factor of 2.3, for 0.0625 m² of treated ground per metre of footing. Four further repetitions reach a factor of 4.9 for 0.525 m² in total, while the efficiency per treated area falls to about a third of that of the first step. The first step, at the root of the band, is therefore by far the most effective. None of the strengthened runs reaches a peak within the computed path, so the comparison is made at the failure settlement of the unstrengthened run and is not directly comparable with results stated at failure (Hadjiloo and Grabe, 2023a; Seitz, 2021).

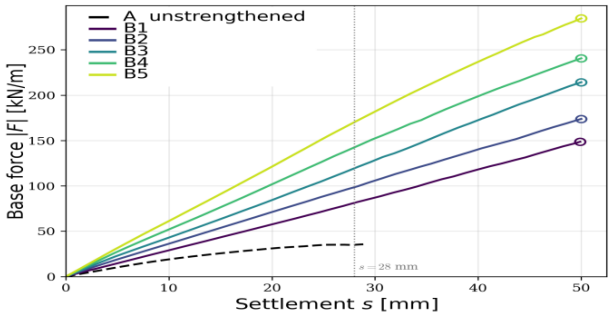


Figure 8: Reload chain at 0.10 m embedment. A is the unstrengthened reference, B1 to B5 successive strengthening steps, each generated from the band of the preceding run. The dotted line marks the failure settlement of A, the common evaluation anchor.

6 Conclusions and outlook

A phase-field model coupled to a hypoplastic constitutive law has been presented for the analysis of shear band strengthening, and three conclusions can be drawn. First, the formulation is mesh-objective. The band width is controlled by the regularisation length and converges to 2ℓ from $h \leq \ell/4$, in the dense and dilatant case where a regularisation is actually needed, and the failure load of the footing becomes mesh-independent from $h \leq \ell/3$. The mesh dependence that limits the interpretation of localisation in standard calculations is removed, and the resolution requirement is a result, not an assumption.

Second, the acoustic tensor is the appropriate driving force for this problem. It is the only one of the three formulations examined that supplies directional information, stays located correctly and narrow under reloading, and addresses the same mathematical event, the loss of ellipticity, that the regularisation is designed to cure. Onset, position and orientation of the band are consequently a prediction of the constitutive model, not the result of a threshold fitted to the expected answer.

Third, the coupling itself is the technical result. A hypoplastic constitutive law and a phase-field model have been joined inside a single user material subroutine of a commercial code, without a user element and without a separate mesh for the treated zone, which enters only as a field of κ. In this form the model reproduces the effect the method is built on. Strengthening the initial section of the potential band raises the load carried at a given settlement well above the unstrengthened reference while the treated volume stays small, and the first intervention at the root of the band is the most effective. A small, shallow cement injection thereby replaces a large improved volume or additional structural elements, with the corresponding savings in material and embodied CO_2.

The next step is the quantitative validation against the model tests of Seitz (2021), in which both the unstrengthened reference and several strengthening configurations were measured.

Acknowledgement

The authors thank the Deutsche Forschungsgemeinschaft for funding the project “Traglaststeigerung geotechnischer Konstruktionen durch Scherfugenverfestigung” (GR 1024/23-2).

References

[1] Ambati M., Gerasimov T. and De Lorenzis L. (2015): A review on phase-field models of brittle fracture and a new fast hybrid formulation. *Computational Mechanics*, 55(2):383–405

[2] Amor H., Marigo J.-J. and Maurini C. (2009): Regularized formulation of the variational brittle fracture with unilateral contact: numerical experiments. *Journal of the Mechanics and Physics of Solids*, 57(8):1209–1229

[3] Bourdin B., Francfort G.A. and Marigo J.-J. (2000): Numerical experiments in revisited brittle fracture. *Journal of the Mechanics and Physics of Solids*, 48(4):797–826

[4] Choo J. and Sun W. (2018): Coupled phase-field and plasticity modeling of geological materials: from brittle fracture to ductile flow. *Computer Methods in Applied Mechanics and Engineering*, 330:1–32

[5] Cui W., Potts D.M., Zdravković L., Gawecka K.A. and Tsiampousi A. (2023): Nonlocal strain regularisation for critical state models with volumetric hardening. *Computers and Geotechnics*, 157:105350

[6] de Borst R., Sluys L.J., Mühlhaus H.-B. and Pamin J. (1993): Fundamental issues in finite element analyses of localization of deformation. *Engineering Computations*, 10(2):99–121

[7] Fei F. and Choo J. (2020): A phase-field method for modeling cracks with frictional contact. *International Journal for Numerical Methods in Engineering*, 121(4):740–762

[8] Francfort G.A. and Marigo J.-J. (1998): Revisiting brittle fracture as an energy minimization problem. *Journal of the Mechanics and Physics of Solids*, 46(8):1319–1342

[9] Grabe J. and Pucker T. (2012): Bodenertüchtigungsverfahren sowie Anordnung dafür. Patent EP 2 497 863 A2, Technische Universität Hamburg-Harburg

[10] Hadjiloo E. and Grabe J. (2023a): Investigation of shear band strengthening by using different strengthening criteria. Proceedings of the 8th International Symposium on Deformation Characteristics of Geomaterials, Porto

[11] Hadjiloo E. and Grabe J. (2023b): Numerical shear band strengthening with contact modelling for the increase of load capacity of geotechnical structures. Geotechnics for Sustainable Infrastructure Development, Geotec Hanoi 2023

[12] Kristensen P.K., Niordson C.F. and Martínez-Pañeda E. (2021): An assessment of phase field fracture: crack initiation and growth. *Philosophical Transactions of the Royal Society A*, 379(2203):20210021

[13] Miehe C., Hofacker M. and Welschinger F. (2010a): A phase field model for rate-independent crack propagation: robust algorithmic implementation based on operator splits. *Computer Methods in Applied Mechanics and Engineering*, 199(45–48):2765–2778

[14] Miehe C., Welschinger F. and Hofacker M. (2010b): Thermodynamically consistent phase-field models of fracture: variational principles and multi-field FE implementations. *International Journal for Numerical Methods in Engineering*, 83(10):1273–1311

[15] Navidtehrani Y., Betegón C. and Martínez-Pañeda E. (2021): A unified Abaqus implementation of the phase field fracture method using only a user material subroutine. *Materials*, 14(8):1913

[16] Pucker T. (2013): Stoffmodell zur Modellierung von stetigen Materialübergängen im Rahmen der Optimierung geotechnischer Strukturen. Dissertation, Technische Universität Hamburg, Veröffentlichungen des Instituts für Geotechnik und Baubetrieb, Heft 28

[17] Rice J.R. (1976): The localization of plastic deformation. In: Koiter W.T. (ed.): Theoretical and Applied Mechanics, North-Holland, Amsterdam, 207–220
[18] Rudnicki J.W. and Rice J.R. (1975): Conditions for the localization of deformation in pressure-sensitive dilatant materials. *Journal of the Mechanics and Physics of Solids*, 23(6):371–394
[19] Seitz K.-F. (2021): Zur Topologieoptimierung von geotechnischen Strukturen und zur Tragfähigkeitssteigerung des Baugrunds durch Scherfugenverfestigung. Dissertation, Technische Universität Hamburg, Veröffentlichungen des Instituts für Geotechnik und Baubetrieb, Heft 45
[20] Tejchman J. and Bauer E. (1996): Numerical simulation of shear band formation with a polar hypoplastic constitutive model. *Computers and Geotechnics*, 19(3):221–244
[21] Vermeer P.A. (1990): The orientation of shear bands in biaxial tests. *Géotechnique*, 40(2):223–236
[22] von Wolffersdorff P.-A. (1996): A hypoplastic relation for granular materials with a predefined limit state surface. *Mechanics of Cohesive-Frictional Materials*, 1(3):251–271

Author

Elnaz Hadjiloo, M.Sc.
Technische Universität Hamburg (TUHH)
Institut für Geotechnik und Baubetrieb (B-5)
Schellerdamm 22 (Channel 7)
21079 Hamburg, Germany
Tel.: +49 (0) 40 30601 3811
e-mail: elnaz.hadjiloo@tuhh.de
Web: https://www.tuhh.de/gbt/institut/team/elnaz-hadjiloo