Proceedings of the 21st ICSMGE, Vienna, Austria, 14 – 19 June 2026. Pistrol, Adam & Schweiger (eds.)
Published by: ÖGG, Austrian Society for Geomechanics, Salzburg, Austria, ISBN 978-3-9503898-4-5
https://doi.org/10.53243/ICSMGE2026-596

Reduced Order Models for geotechnical predictions

Mao Ouyang, Alexandros Petalas, William Coombs, Charles Augarde
Department of Engineering, Durham University, United Kingdom, mao.ouyang@durham.ac.uk

Ahmed Alagha, Jonathan Knappett, Michael Brown
School of Science and Engineering, University of Dundee, United Kingdom

ABSTRACT: Finite Element (FE) modelling is routine in geotechnical engineering design and analysis; however, high-fidelity 3D models remain the preserve of highly complex or prestigious geotechnical construction projects, due to their computational cost. A sensitivity study using a complex model can be expensive and time-consuming even using modern hardware. Reduced Order Models (ROMs) offer a solution at potentially vastly lower cost. They are also numerical models of geotechnical problems, where the inputs are a series of defining parameters (geometry, soil properties etc.) and the output is a prediction of interest, e.g. the maximum deflection of a retaining structure. A ROM is developed by analysis of a number of "surrogates" or "snapshots" which are selected instances of results obtained from complex FE modelling. The surrogates are run only once but the ROM can be used to explore predictions in the parameter space in-between the surrogates. In this paper we explain the development of a ROM using a simple geotechnics problem as an example, where advanced neural operator approaches are advantageous. The framework is currently being developed to analyse the behaviour of large 3D braced excavations where the outputs are wall displacements and prop loads, with a particular emphasis on behaviour at the corners.

KEYWORDS: Reduced order models; surrogate; braced excavation.

1 INTRODUCTION

High-fidelity computational geotechnical models are complex and require considerable resources. Reducing the model dimensionality using Reduced Order Models (ROMs) offers a solution at potentially vastly lower cost but then the issue is accuracy. The challenges in designing a ROM are (a) to effectively reduce the model dimensionality, e.g., to achieve good performance of reconstruction to the Full Order Model (FOM) from a reduced (or "latent") space; (b) to accurately predict spatio-temporal behaviours, e.g., to obtain the results in the original space from the input in the latent space.

Machine learning has been widely used in modelling geotechnical problems (Sheil, et al., 2026). However, in most cases, learning/training is carried out in the original spaces, rendering intensive computational cost (Choi, et al., 2025). To develop more efficient ROMs for geotechnical predictions, the first task is to find low dimensional representations of high dimensional data. Pearson (1901) proposed the original form of Principal Component Analysis (PCA), which is a linear method to obtain low dimensional but representative subspaces of data and is familiar from image compression. The recent popularity of deep learning has introduced a new method for finding low dimensional subspaces, most notably using the autoencoder (Hinton & Salakhutdinov, 2006) which is a form of neural network (NN). The work presented here adopts the autoencoder for determining the latent spaces. The second task for ROM development is learning of the nonlinear behaviour in the latent space. Neural operators, e.g., Deep Neural Operator (*DeepONet*) (Lu, et al., 2021), Fourier Neural Operator (Li, et al., 2021), present a principal framework for learning functions in continuous domains, e.g., latent space (Azizzadenesheli, et al., 2024). Kontolati et al. (2024) proposed methods of learning neural operators in a latent space, facilitating the real-time prediction of complex nonlinear systems. Regazzoni et al. (2024) introduced the Latent Dynamic Network (LDN) to discover low dimensional subspaces while learning the system behaviour in the original high dimension space. Brivio et al. (2024) developed a low-cost pre-training procedure for ROMs based on both PCA and the autoencoder. In this study, we will implement the DeepONet due to the advantage that it substantially improves generalization based on a design of two subnetworks, the branch net for the input function, and the trunk net for the output spatio-temporal locations.

This paper proposes a ROM where the inputs are a series of defining parameters (geometry, soil properties etc.) and the output is a prediction of interest, e.g. excess pore water pressure (PWP), or the maximum deflection of a retaining structure. The latent space is obtained through training using a number of "snapshots" selected from the high dimensional numerical results (the FOM). After obtaining the latent space of both input and output, we train a DeepONet to map the relationship between the input and output. Finally, the ROM is a combination of both the trained autoencoder and DeepONet. The whole process will be demonstrated on the classic 1D consolidation problem. Our key contributions can be summarised as follows:

- **Proposal of a ROM framework for geotechnical models:** We demonstrate the processes of training, testing and application of the developed state-of-the-art ROM where the prediction accuracy is comparable with the FOM.
- **Enhanced computational efficiency:** Through reducing the dimensionality, the resulting architecture is significantly smaller than the FOM. Both the training, testing, and application cost are greatly reduced compared with the FOM.

2 METHODOLOGY

2.1 *Problem statement*

The key difference between a DeepONet and more standard NN approaches is that unlike NNs, the DeepONet is seeking the *operator* that transforms the input (initial conditions for instance) to the output (excess PWPs for instance) rather than finding a link between two sets of data; a DeepONet learns the nonlinear relationships between the infinite dimensional functional spaces on bounded domains. We denote by $\boldsymbol{u}: \Omega$ a time-constant input field and $\boldsymbol{y}: \Omega \times [0, T]$ an output field, where Ω is the space domain and $T > 0$ is the final time. The goal is to find the nonlinear dependency of $\boldsymbol{y}$ on $\boldsymbol{u}$ through the following parametric mapping:

$$\mathcal{G}: \boldsymbol{u} \times \Theta \rightarrow \boldsymbol{y} \quad (1)$$

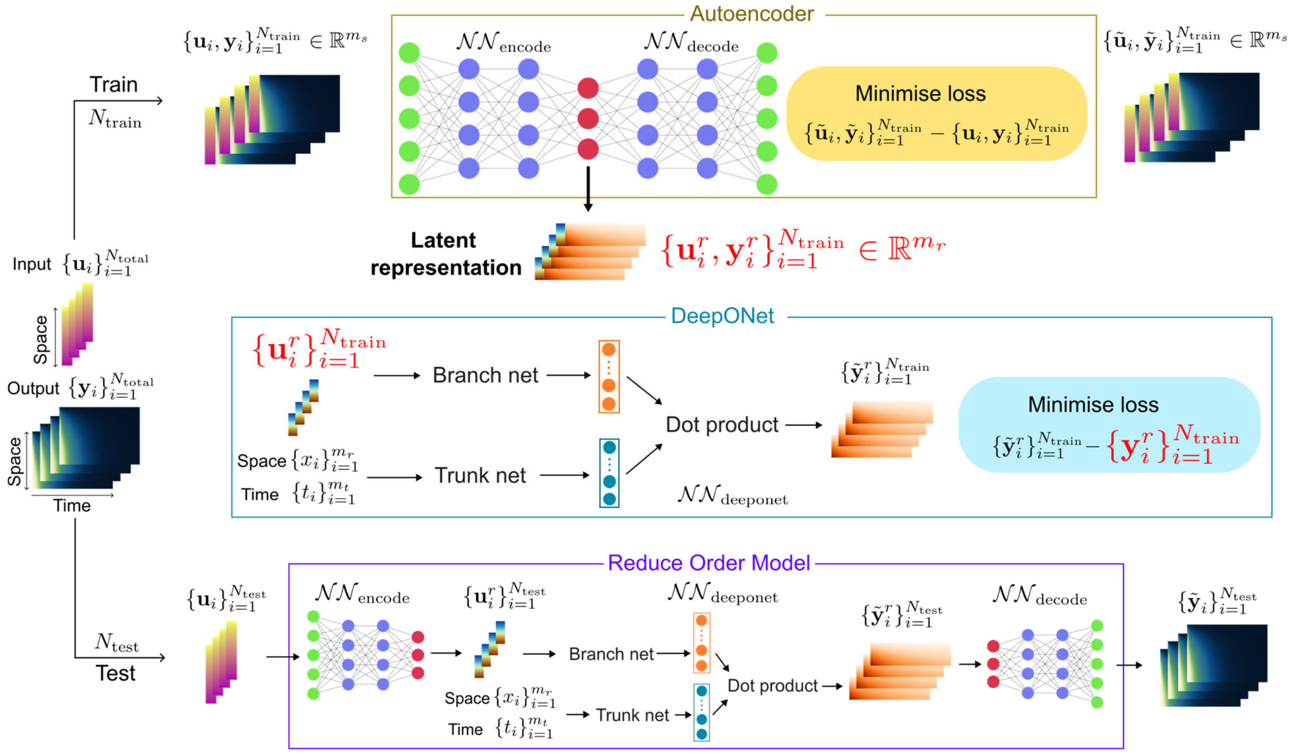

Figure 1. The Reduced Order Model (ROM) for learning neural operators in the latent spaces.

where Θ is a finite-dimensional parameter space.

In this standard setting, the ROM for learning the neural operators in the latent space (mapping the latent input to latent output), and then decode to original space is presented in Figure 1.

2.2 *Data generation*

Though the proposed ROM in Figure 1 can be used for a wide range of applications, here we present a detailed implementation for PWP dissipation due to one-dimensional consolidation problem (Terzaghi, 1925). The governing equation for the consolidation with initial conditions measured at m_s points (Figure 2) is

$$\frac{\partial y_i}{\partial t} = c_v \frac{\partial^2 y_i}{\partial x^2},\ i = 1, 2, \cdots, m_s \qquad (2)$$

where, y_i is the PWP of the ith point, and c_v is the coefficient of consolidation. Here, for simplicity we adopt a uniform value of $c_v = 2.24\, m^2/y$ with different starting profiles of excess PWPs. Boundary conditions are as shown in Figure 2 and $H = 30\ mm$ based on Mohamedelhassan & Shang (2002).

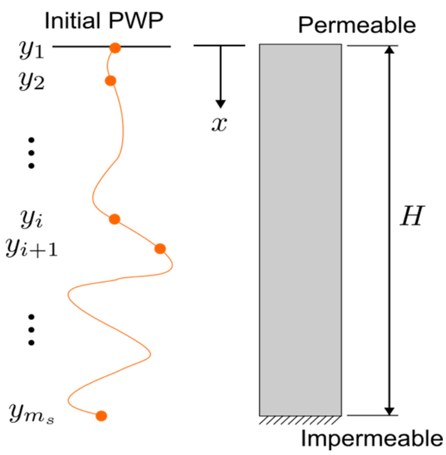

Figure 2. One-dimensional consolidation on arbitrary initial PWPs.

The ROM is built by learning the link between the initial conditions ($\boldsymbol{u} = \boldsymbol{y}(x, t = 0)$) to the evolution of the PWP ($\boldsymbol{y}(x, t \in (0, T])$) (Singh & Chakraborty, 2022).

The input ($\boldsymbol{u}$) is a vector of PWPs generated using a Gaussian Random Field (GRF) with the correlation length of 0.2 kPa (Lu, et al., 2021) at 2,048 equally distributed points on unit normalised vertical dimension ($m_s = 2{,}048$). The output ($\boldsymbol{y}(x, t \in (0, T])$) is obtained by running 300 second-order implicit Finite Difference approximations of Equation (2) for 1 hour with 256-time steps with respect to each input ($N_{total} = 300, m_t = 256$). We separate 80% of the data for training and 20% for testing.

2.3 *Model dimension reduction*

We train an unsupervised autoencoder model to find the low dimensional latent representation (m_r) for the concatenated "snapshots" ($\{\boldsymbol{u}_i, \boldsymbol{y}_i\}_{i=1}^{N_{train}}$) through reshaping the space and time dependent output to be the dimension of $m_t \times m_s$, the columns sharing the same dimension as $\boldsymbol{u}$ (m_s). A multi-layer perceptron autoencoder with three hidden layers each is used to train the encoder ($\mathcal{NN}_{encoder}$), and decoder ($\mathcal{NN}_{decoder}$), through minimising the error between the predicted and original "snapshot".

2.4 *Train neural operator on latent space*

Once we have obtained the latent representation of both input and output ($\{\boldsymbol{u}^r, \boldsymbol{y}^r\}$), the aim is to approximate the mappings from $\boldsymbol{u}^r$ to $\boldsymbol{y}^r$. An unstacked DeepONet with the branch net fed with the $\boldsymbol{u}^r$, and the trunk net fed with both the time and space coordinates are implemented. Here we consider a Convolutional Neural Network (CNN) for the branch net and a Feed-forward Neural Network (FNN) for the trunk net for the DeepONet architecture. The DeepONet ($\mathcal{NN}_{deeponet}$) is trained by minimising the loss of the predicted output ($\widetilde{\boldsymbol{y}^r}$) with the reduced output ($\boldsymbol{y}^r$).

2.5 *Error metric*

To assess the performance of ROM, we consider the Mean Square Error (MSE) on N_{test} test realisations:

$$\text{MSE} = \frac{1}{N_{test}} \sum_{j=1}^{N_{test}} (\boldsymbol{y}_j - \boldsymbol{y}_j^{rom})^2 \quad (3)$$

where $\boldsymbol{y}$ is the reference responses, and $\boldsymbol{y}^{rom}$ is the output from the ROM in the full dimension.

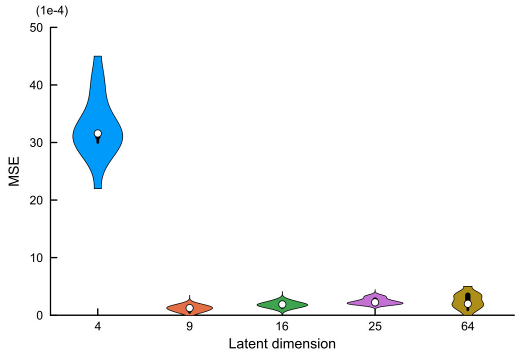

Figure 3. The reconstruction error of autoencoder with different latent dimensionalities.

3 RESULTS

3.1 *Reconstruction error of autoencoder*

To examine the performance of the autoencoder in dimensionality reduction, we show the MSE calculated by Equation (3) between the reference data and the model reconstructions in Figure 3. The violin plot shows the distributions of MSE for 5 independent trials with different random seed numbers. We observe that a latent dimension equal/greater than 9 can well represent the concatenated data.

3.2 *Results of the ROM*

For the test procedure, as indicated in Figure 1, the input from the test dataset is encoded to the latent space via the encoder ($\mathcal{NN}_{encoder}$), and then fed to the DeepONet ($\mathcal{NN}_{deeponet}$) to obtain the reduced output, which is decoded to the full dimension by the decoder ($\mathcal{NN}_{decoder}$).

Figure 4 shows a comparison among the reference response and model predictions of a representative sample from the test data. Figure 4(a) shows a reference response of PWP profile (in kPa), which is regarded as "ground truth" here. The ROM predictions based on the trained model (shown as Figure 1) are shown in Figure 4(b) with 16 latent dimensions.

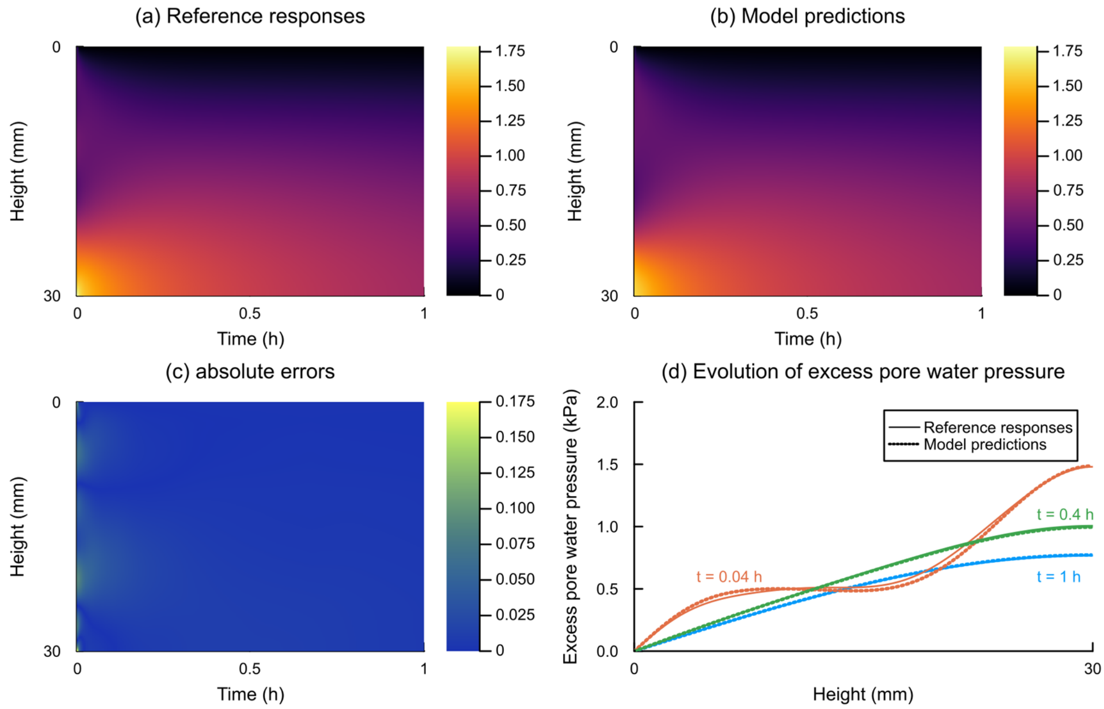

Figure 4. A representative example of the reference responses and ROM predictions on 16 latent dimensions.

The absolute error between the reference responses and model predictions (Figure 4(c)) suggests that the ROM can well describe the nonlinear development and dissipation of PWP during the consolidation process. The errors are concentrated at the initial stage of the consolidation. Figure 4(d) shows a detailed comparison between the excess PWP at $t = 0.04,\ 0.4$ and $1\ h$ for reference responses and model predictions. Again, the results of the ROM agree well with the reference.

In Figure 5, the results of ROMs with different latent dimensionalities are presented as violin plots, using 5 independent trials. The MSE is obtained via Equation (3). We observe that the latent dimensions of 25 and 64 perform worse than those of 9 and 16, suggesting that for the proposed ROMs, larger subspaces may not better describe the nonlinear behaviour. For a small latent space, e.g., 4, the performance is poor due to the large reconstruction error of the autoencoder (Figure 3), although the nonlinear behaviours on the latent space can be learned by the DeepONet in the latent space, the performance of ROMs is generally worse than the latent dimensions of 9 and 16. In this study, the latent dimension of 16 achieves the best performance with the mean MSE equal to 0.0015.

3.3 *Computational cost*

Here we compare the computational cost for training the DeepONet in the original space and latent space, as shown in Table 1. The differences between the Full DeepONet and Latent DeepONet are shown in Figure 6. For the DeepONet in the original space, the input in the full dimension ($\boldsymbol{u}$) is applied, whereas for the Latent DeepONet, the reduced input ($\boldsymbol{u}^r$) is used. The spatial information for the DeepONet in the original space and DeepONet in the latent space is also different. The trainable parameters for the DeepONet in the latent space are significantly reduced compared with the DeepONet in the original space, as well as the computational time. All the models were identically trained on an NVIDIA A100 GPU. For the online process of geotechnical predictions, the ROM is around 1,000 times faster than the FOM.

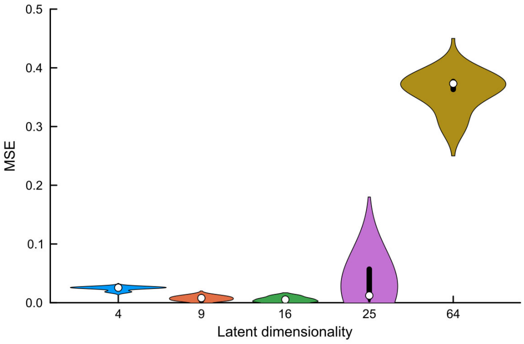

Figure 5. Results of ROMs with different latent dimensionalities.

It is possible to choose different network architectures for the autoencoder and the DeepONet, but our investigations suggest that the trunk net should be different from the encoder to avoid repeat convolution of the input data.

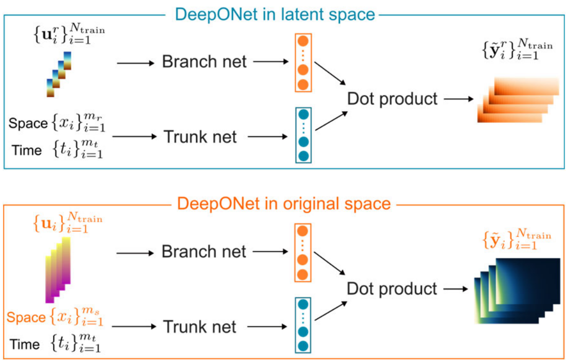

Figure 6. The differences between DeepONet in the original space and the DeepONet in the latent space.

4 CONCLUSIONS

We propose a ROM framework for geotechnical problems, where firstly, the data generated by a FOM are collected to find the latent representations via an autoencoder. Then, a neural operator is trained to map the input to output on the latent space. Finally, the predicted output from the neural operator is recovered to full dimension through the decoder. Our demonstration using the 1D consolidation problem has shown that acceptable results can be predicted for a problem spatially divided into 2,048 points using 9 and 16 latent dimensions with a high accuracy. Both the trainable parameters and computational time are significantly reduced in the ROM. The proposed approach is widely applicable to other geotechnical constructions which need high-fidelity modelling, indeed there is evidence of activity elsewhere in geotechnics making use of DeepONets (Xu et al. 2024).

5 ACKNOWLEDGEMENTS

This work was inspired by funding made available from UKRI for Durham and Dundee Universities [grant numbers EP/X024849/1 \& EP/X024768/1]. This work has used Durham University's NCC cluster. NCC has been purchased through Durham University's strategic investment funds and is installed and maintained by the Department of Computer Science.

Table 1. Comparison of computational costs for training the DeepONet in the original and latent spaces.

Model	m_r	parameters	Wall time (s)
DeepONet in original space	-	*1,253,004*	*895,039*
DeepONet in latent space	*4*	*19,552*	*3,550*
DeepONet in latent space	*9*	*25,952*	*4,769*
DeepONet in latent space	*16*	*34,912*	*5,291*
DeepONet in latent space	*25*	*46,432*	*5,840*
DeepONet in latent space	*64*	*61,372*	*6,993*

6 REFERENCES

Azizzadenesheli, K. et al., 2024. Neural operators for accelerating scientific simulations and design. *Nature Review Physics,* Volume 6, pp. 320-328.

Brivio, S., Fresca, S. & Manzoni, A., 2024. PTPI-DL-ROMs: Pre-trained physics-informed deep learning-based reduced order models for nonlinear parametrized PDEs. *Computer Methods in Applied Mechanics and Engineering,* Volume 432, p. 117404.

Choi, Y., Liu, C. & Macedo, J., 2025. Enhanced DeepONet for 1-D consolidation operator learning: an architectural investigation. arXiv preprint, 2507.10368.

Hinton, G. & Salakhutdinov, R., 2006. Reducing the dimensionality of data with neural networks. *Science,* Volume 313, pp. 504-507.

Kontolati, K., Goswami, S., Karniadakis, G. E. & Shields, M. D., 2024. Learning nonlinear operators in latent spaces for real-time predictions of complex dynamics in physical systems. *Nature Communications,* Volume 15, p. 5101.

Lee, P., Xie, K. & Cheung, Y., 1992. A study on one-dimensional consolidation of layered systems. *International Journal for Numerical and Analytical Methods in Geomechanics,* Volume 16, pp. 815-831.

Li, Z. et al., 2021. *Fourier Neural Operator for Parametric Partial Differential Equations.* s.l.:arXiv.

Lu, L. et al., 2021. Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. *Nature Machine Intelligence,* Volume 3, pp. 218-229.

Mohamedelhassan, E. & Shang, J., 2002. Vacuum and surcharge combined one-dimensional consolidation of clay soils. *Canadian Geotechnical Journal,* Volume 39, pp. 1126-1138.

Pearson, K., 1901. On lines and planes of closest fit to systems of points in spaces. *The London, Edinburgh, and Dublin philosophical magazine and journal of science,* 2(11), pp. 559-572.

Regazzoni, F. et al., 2024. Learning the intrinsic dynamics of spatio-temporal processes through Latent Dynamic Networks. *Nature Communications,* Volume 15, p. 1834.

Schiffman, R. L. & Stein, J. R., 1970. One-dimensional consolidation of layered systems. *Journal of Soil Mechanics and Foundation Divisions, ASCE,* Volume 96, pp. 1499-1504.

Sheil, B. et al., 2026. Artificial intelligence transformations in geotechnics: progress, challenges and future enablers. *Computers and Geotechnics,* Volume 189, p. 107604.

Singh, A. & Chakraborty, M., 2022. Numerical Analysis of 1-D Consolidation of Two-layered Soils with Different Initial Excess Pore Water Pressure. *Geotechnical and Geological Engineering,* Volume 40, pp. 1023-1040.

Terzaghi, K., 1925. *Erdbaumechanik.* Vienna: Deuticke.

Xu, C. et al. 2024. A multi-fidelity deep operator network (DeepONet) for fusing simulation and monitoring data: Application to real-time settlement prediction during tunnel construction. Eng Appln of Artificial Intelligence, V133, Part B, July 2024, 108156