Journal of Rock Mechanics and Geotechnical Engineering 18 (2026) 4007–4022

Contents lists available at ScienceDirect

Journal of Rock Mechanics and Geotechnical Engineering

journal homepage: www.jrmge.cn

Full Length Article

Deflection of foundation pit retaining wall in heterogeneous soil strata: Intelligent prediction methods based on physics-informed machine learning

Cong Zhou [a,b], Lei He [a,b,*], Junchen He [a], Yi Zhang [a], Huaiguang Xiao [a], Chee Kiong Soh [a]

[a] School of Civil Engineering, Southeast University, Nanjing, 211189, China
[b] State Key Laboratory of Safety, Durability and Healthy Operation of Long Span Bridges, Southeast University, Nanjing, 211189, China

ARTICLE INFO

Article history:
Received 23 February 2025
Received in revised form
23 June 2025
Accepted 1 July 2025
Available online 26 October 2025

Keywords:
Heterogeneous soil strata
Braced excavation
Retaining wall deflection (RWD)
Soil-structure interaction
Physics-informed neural networks (PINNs)

ABSTRACT

Excavation-induced retaining wall deflection (RWD) significantly influences the safety of surrounding built environment. To predict the three-dimensional RWD in heterogeneous strata, a new partial differential equation (PDE) is derived in this study, and two prediction models are proposed, i.e. the physics-informed neural network (PINN) model and the data-driven PINN model. As a physical constraint, the new PDE is crucial to the loss functions of these models. Then, the validity of the models is verified and analysed using a subway deep-foundation pit. The results show that the training times of both models are controlled within 900 s, which is a significant reduction compared to that of the conventional numerical model. In addition, the prediction accuracy of the data-driven PINN model is higher than that of the numerical model, while that of the PINN model is slightly lower than that of the numerical simulation. However, in contrast to the data-driven PINN model, the PINN model can identify irregular soil interfaces in heterogeneous strata to learn the deflection continuity conditions at irregular interfaces and realize RWD prediction in non-uniform distributed strata. In practical applications in foundation pit engineering, the selection of the PINN and data-driven PINN models can be conducted according to the in situ distribution conditions of the strata to enable the early prediction of potential RWD, thereby providing a reliable basis for the further optimisation of retaining structures design.

1. Introduction

With rapid urbanisation, the development and utilisation of underground spaces have become crucial for sustainable urban development (Broere, 2016; Bobylev, 2016; Chen et al., 2022; Zhou et al., 2025). Braced foundation pits have emerged as a typical construction method for large-scale urban underground spaces, with the scale and complexity of such development increasing significantly (Zhang et al., 2018a; Feng, 2019). Notable examples include underground subway stations and basements for high-rise buildings. As a critical component of foundation pits, the deflection behaviour of the retaining wall is significantly influenced by excavation-induced soil disturbance (Kung et al., 2007; Zhao et al., 2021). Retaining wall deflection (RWD) not only directly impacts the safety of the foundation pit, but also poses a threat to the stability of the surrounding environment (Ibrahim et al., 2021; Dong et al., 2022), potentially triggering geological disasters and leading to substantial economic losses and social consequences. Therefore, understanding the deflection mechanism of retaining walls and effectively controlling RWD are crucial issues in foundation pit research.

Accurately predicting the deflection trends of retaining walls based on high-precision models during underground space planning or foundation pit design can have significant implications for engineering practice. Such predictions can facilitate the optimisation of site selection strategies for underground spaces and refinement of foundation pit design, and mitigate potential engineering risks (Zhang et al., 2019a; Hong et al., 2023). Moreover, it can significantly enhance the reliability and safety of underground

* Corresponding author. School of Civil Engineering, Southeast University, Nanjing, 211189, China.
E-mail address: helei_civil@seu.edu.cn (L. He).
Peer review under responsibility of Institute of Rock and Soil Mechanics, Chinese Academy of Sciences.

https://doi.org/10.1016/j.jrmge.2025.07.019

space development and offer a robust scientific foundation and valuable guidance for sustainable urban development.

In previous studies, driven by advances in numerical methods and the complexity of deflection behaviour, finite element (FE) modelling has been mainly used to estimate RWD. The interaction between the structure and the soil is a critical focal point for studying RWD (Lim et al., 2018; Liu et al., 2022; Li et al., 2024). Many studies have employed numerical methods to analyse the stress and deflection responses of retaining walls at different excavation stages, exploring deflection patterns and control strategies under structure-soil interaction (Zdravkovic et al., 2005; Finno et al., 2007; Iwata et al., 2007; Yeh et al., 2022). The impact of groundwater level variations on RWD has also attracted significant attention, with related studies examining the additional stress distribution caused by dewatering and its dynamic effects on RWD (Zheng et al., 2014; Zhang et al., 2018b; Zeng et al., 2021). Construction methods and sequencing are important influence factors, and have been optimised in relevant studies using numerical software to reveal the mechanisms via which different construction processes influence structural deformation (Lim and Ou, 2018; Maehara et al., 2020; Zheng et al., 2022). Furthermore, targeted analyses have been conducted on the complexities of geological conditions, and numerical models for excavation have evolved from conventional homogeneous soil models to heterogeneous soil models (Sert et al., 2016; Ding et al., 2021; Zhou et al., 2024), reflecting the significant impact of different geological conditions on RWD. Research findings indicate that RWD is often concentrated in weak soil layers, while it is relatively smaller in hard soil layers with higher cohesion (Li et al., 2024). Although numerical models can provide detailed RWD simulations, numerical models are generally characterized by cumbersome modelling processes, lengthy computation times, and limited applicability to other scenarios. Moreover, when comparing different site selection plans for underground spaces or alternative design schemes for foundation pits, the time cost associated with numerical methods is often significantly high.

To avoid large-scale numerical computations and enhance the general applicability, extensive field monitoring data have been analysed to develop empirical and semi-empirical solutions for RWD. For instance, Kung et al. (2007) proposed a simplified semi-empirical model for predicting the maximum RWD for a braced excavation in soft to medium clays. Wang et al. (2005) investigated the excavation variables affecting RWD at six deep-braced excavations in soft soils in Shanghai and derived the relationship between the maximum RWD and excavation depth from field measurements. Hsieh and Ou (2016) derived a simplified formula to predict the lateral RWD in deep excavation, considering the installation of cross walls. Goh et al. (2017) developed a simple RWD equation for estimating the maximum wall deflection considering three-dimensional (3D) effects through different ratios of excavation lengths and widths. Based on 30 case studies, Zhang et al. (2015) proposed a simple polynomial regression model for estimating the maximum RWD caused by deep excavations. Based on a series of 3D FE analysis results, Zhang et al. (2019b) developed a multivariate adaptive regression spline model to accurately learn the complicated implicit relationship between the maximum RWD and the influential factors. These empirical solutions are derived from the analysis and summarization of extensive measured data, providing a certain reference value. However, the deflection points of retaining walls that can be predicted using these empirical solutions are limited, typically focusing only on the maximum deflection value while failing to capture the overall deflection state of the structure.

Theoretical analytical approaches have been proposed to solve for the deflection values at more points on retaining walls. These methods simplify the structure as a pile embedded in the soil and derive a mechanical ordinary differential equation (ODE) to describe the RWD based on the force analysis of the pile (Zhu and Wei, 2010; Zhao et al., 2012; Abbasi et al., 2014). The ODE is then iteratively solved using the boundary conditions at the top and bottom of the pile and the deformation continuity conditions at the excavation surface, ultimately providing the horizontal deformation values. This analytical approach avoids lengthy numerical computations, improving the efficiency of RWD calculations. However, these methods are based on two-dimensional (2D) plane strain states and assume homogeneous soil conditions for the foundation pit. As a result, the computed results cannot reflect the 3D deflection behaviour of retaining walls in real-world engineering projects, and cannot quantify the impact of heterogeneous soil strata on RWD.

With the development of artificial intelligence (AI) technology, machine learning (ML) algorithms have been applied in underground engineering (Huang et al., 2021; Phoon and Zhang, 2022; Wu et al., 2023) to study RWD. For example, Chua and Goh (2005) adopted Bayesian neural networks to predict wall deflections; the developed model can account for various factors such as the wall stiffness, support stiffness, in situ stress state, non-homogeneous soil conditions, and variation of soil properties with depth. Cheng and Wu (2009) adopted a hybrid approach that fuses support vector machines and fast messy genetic algorithms to predict RWD in a deep excavation in the Taipei Basin; the results showed that the hybrid approach can successfully predict RWD. Zhao et al. (2021) proposed a convolution-neural-network-based prediction method for RWD capable of making dynamic predictions based on the extracted spatiotemporal features of wall deflection. To estimate wall deflection with high accuracy and efficiency, Tran et al. (2023) proposed several ML models, including feed-forward neural network back-propagation, long short-term memory (LSTM), bidirectional LSTM (Bi-LSTM), and support vector regression. The results show that the Bi-LSTM model has the highest robustness and can be used to support the safety monitoring and early warning of RWD. The application of AI technologies in predicting RWD is relatively limited, with existing studies primarily focusing on predicting the maximum deflection. This may be due to the strong spatial characteristics of RWD, in addition to the fact that monitoring points for RWD are mainly placed at key locations, resulting in a relatively small amount of field data available for research and analysis. As a result, it is difficult to fully capture the spatial characteristics of RWD with limited data. Currently, AI-based RWD predictions are closely aligned with data analysis and empirical methods.

From previous studies, it can be seen that while analytical methods can provide relatively rich deflection information for retaining walls, the assumptions are based on plane strain conditions and homogeneous soil strata. Empirical methods and AI models can only predict fewer deflection points, primarily excelling at predicting the maximum RWD. In order to quickly, accurately, and comprehensively predict RWD in heterogeneous strata, this study attempts to use physics-informed neural networks (PINNs) (Raissi et al., 2019; Zhu et al., 2019) for the predictive analysis of RWD. Compared to the conventional ML models with relatively limited generalization abilities, PINNs leverage a learning bias approach to make predictions across different datasets. By incorporating appropriate physical constraints into the neural network's loss function (Lan et al., 2024), PINNs can seamlessly integrate observational or experimental data with the physics and mechanics described by physical equations, reducing the need for prior knowledge of the underlying physics and mechanics (Borate et al., 2023; Wu et al., 2025). This enables comprehensive predictions of RWD with minimal data.

Furthermore, PINNs can solve forward-prediction problems without relying on prior knowledge, offering a novel approach for solving engineering problems without field data (Gong et al., 2024).

In this study, a new 3D partial differential equation (PDE) for forward RWD prediction is derived based on the existing 2D ODE, and two prediction models are proposed: the PINN and data-driven PINN models. The loss functions for both models are constructed based on the new PDE. The PINN model is not driven by actual engineering data, while the data-driven PINN model utilizes a small amount of engineering data for learning. Subsequently, the effectiveness of the PINN and data-driven PINN models is validated through a case study of a subway foundation pit. The applicability of the two models under different engineering geological conditions is also discussed. The remainder of this paper is organised as follows. Section 2 presents the derivation of the new PDE for 3D RWD and introduces the PINN and data-driven PINN models. Section 3 presents the validation of the PINN and data-driven PINN models using a subway foundation pit located in Nanjing, China. Section 4 discusses the use of the PINN model to predict RWD in non-uniformly distributed strata, and the main conclusions are summarised in Section 5.

2. Methodology

2.1. RWD under soil-structure interaction

Shallow foundation pits traverse relatively few geological strata and can generally be considered homogeneous (Fig. 1a). However, deep foundation pits, such as urban subway stations and large underground spaces, inevitably traverse multiple soil strata (Fig. 1b). The RWD of deep foundation pits under heterogeneous geological conditions becomes significantly more complex, requiring separate considerations in different soil strata.

Under the influence of soil-structure interaction, the soil pressure exerted on the retaining wall lies between the active and passive soil pressures, representing a non-limit state pressure (Zhu and Wei, 2010). Therefore, the classical methods for calculating the Coulomb and Rankine soil pressures are not suitable for describing the stress state of the retaining wall. Previous studies employed the soil spring theory to quantify the soil pressure acting on the retaining wall under non-limit state conditions (Zhao et al., 2012; Abbasi et al., 2014). Considering the 2D excavation scenario shown in Fig. 2a as an example, the RWD towards the excavation is considered to be positive and that away from the excavation is considered to be negative. Then, the soil pressure at any point on the retaining wall can be expressed as

$$\sigma = e_0 \pm k_s\omega(z) \tag{1}$$

where e_0 is the static lateral soil pressure behind the wall; $\omega(z)$ is the RWD value at depth z; k_s is the stiffness coefficient of the soil spring, also known as the coefficient of lateral resistance (kPa/m); and $k_s\omega(z)$ is the incremental soil pressure behind the wall caused by the RWD.

The value of k_s is significantly influenced by factors such as the depth of the calculation point, proportional coefficient of lateral resistance m (kPa/m^2), and RWD of the calculation point; it can be calculated as follows (Zhao et al., 2012):

$$\left.\begin{aligned} k_s &= \left[\frac{2}{mz} + \frac{\omega(z)}{e_0 - e}\right]^{-1} \\ m &= \frac{1}{\omega(z=H)}\left(0.2\varphi^2 - \varphi + c\right) \end{aligned}\right\} \tag{2}$$

where H is the excavation depth of the foundation pit, as shown in Fig. 1a; φ is the soil internal friction angle; c is the soil cohesion; and e denotes the ultimate state soil pressure, which can be calculated using Rankine's soil pressure formula. When $\omega(z) > 0$, e corresponds to the active soil pressure e_a; and when $\omega(z) < 0$, e corresponds to the passive soil pressure e_p.

The deformation ODE for a pile under plane strain conditions can be expressed as (Zhu and Wei, 2010; Abbasi et al., 2014):

$$EI\frac{d^4\omega(z)}{dz^4} + P = 0 \tag{3}$$

where E is the Young's modulus, EI represents the bending stiffness of the pile, and P is the applied load.

By substituting Eq. (1) into Eq. (3), the deflection ODE for the retaining wall under plane strain conditions can be obtained using Eq. (4).

$$\left.\begin{aligned} EI\frac{d^4\omega(z)}{dz^4} + k_T\omega(z)b &= [e_0 - k_s\omega(z)]b \quad (0 < z < H) \\ EI\frac{d^4\omega(z)}{dz^4} + [e_{0b} + k_s\omega(z)]b &= \left[e_{0f} - k_s\omega(z)\right]b \quad (z \geq H) \end{aligned}\right\} \tag{4}$$

where k_T represents the average stiffness coefficient of the internal support, b is the length of deflection units, e_{0f} denotes the static lateral soil pressure outside the pit below the excavation surface, and e_{0b} refers to the static lateral soil pressure inside the pit below the excavation surface. Here, e_{0f} and e_{0b} are calculated as

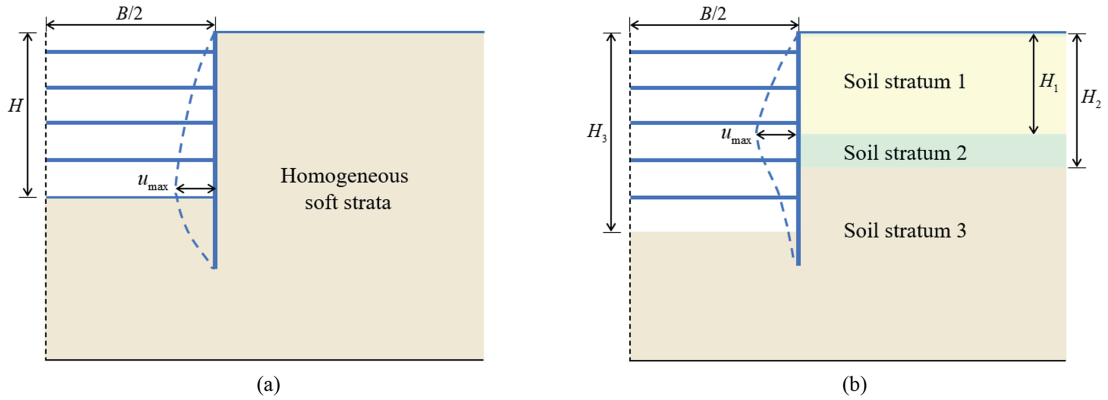

Fig. 1. RWD in different soil strata: (a) In homogeneous soil strata, and (b) In heterogeneous soil strata.

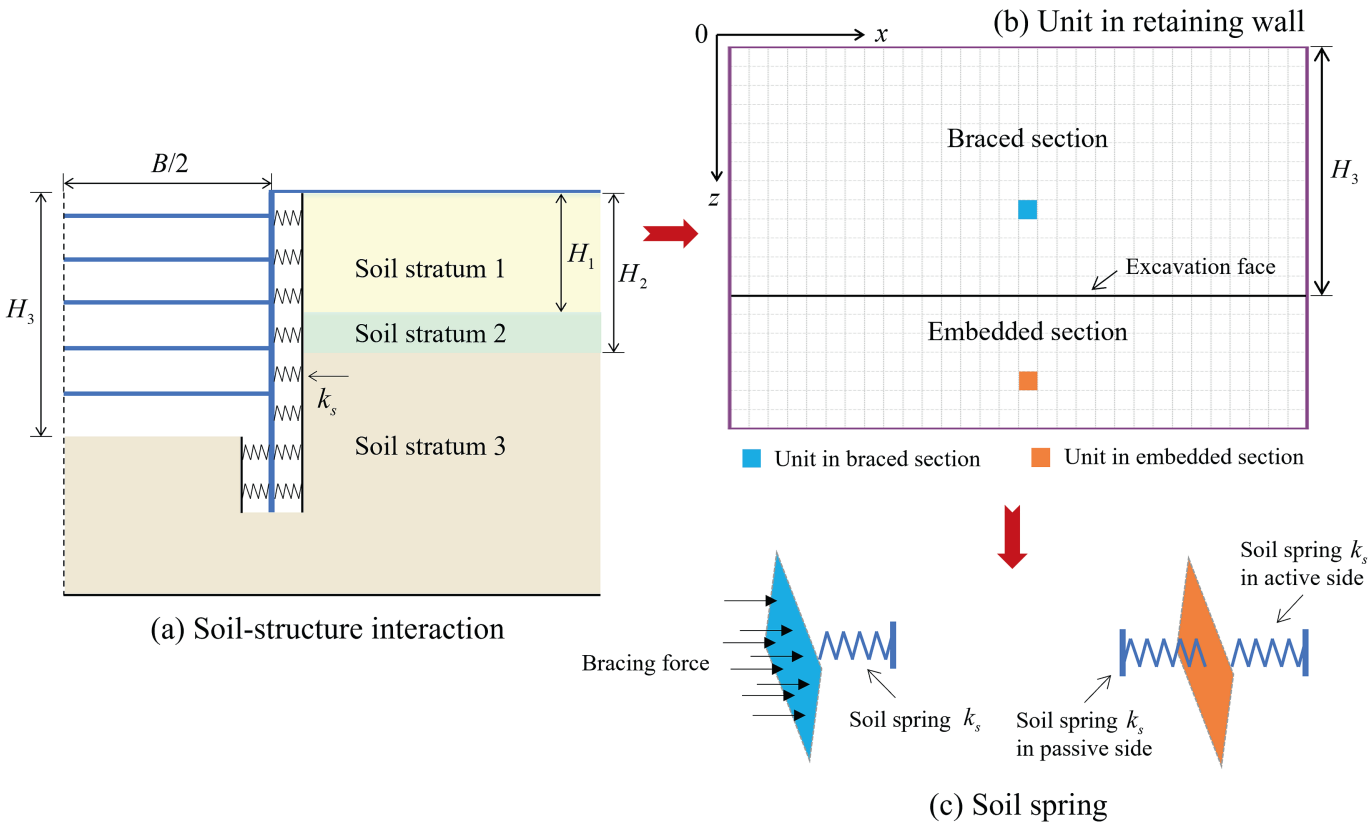

Fig. 2. Diagram of soil-structure interaction.

$$\left.\begin{aligned} e_{0f} &= e_0 = k_0\gamma z \\ e_{0b} &= k_0\gamma(z-H) \end{aligned}\right\} \tag{5}$$

where k_0 is the Rankine static soil pressure lateral coefficient, and γ is the soil weight.

Eq. (4) is a widely used deflection ODE for retaining walls. However, the assumptions of plane strain conditions and homogeneous soil strata clearly deviate significantly from the 3D RWD in practical applications. Moreover, Eq. (4) cannot capture the comprehensive deflection behaviour of retaining walls.

To solve the 3D deflection of retaining walls, it is necessary to conduct force analyses on structures. Considering an arbitrary unit on the retaining wall as an example (Fig. 2b), its deformation is influenced not only by the internal support force, but also by the external soil pressure and bending stiffness. Since the length and height of the retaining wall are significantly larger than the width, the retaining wall can be considered equivalent to a plate-like structure inserted into the soil. The interaction between the structure unit and the soil can be represented by soil springs (Fig. 2c). Subsequently, the 3D deflection PDE for retaining walls can be constructed by referring to Kirchhoff theory (Zhao et al., 2022) and existing 2D deflection ODE. The new deflection PDE is

$$\left.\begin{aligned} D\nabla^4\omega(x,z) + k_T\omega(x,z)A &= [e_0 - k_s\omega(x,z)]A \quad (0 \le z \le H) \\ D\nabla^4\omega(x,z) + [e_{0b} + k_s\omega(x,z)]A &= \left[e_{0f} - k_s\omega(x,z)\right]A \quad (z > H) \end{aligned}\right\} \tag{6}$$

where ∇ is the Laplace operator, D is the bending stiffness of the retaining wall (kN·m) calculated using Eq. (8), k_T is the average stiffness coefficient (kPa/m) of the internal supports to each computed unit (Zhao et al., 2012) calculated using Eq. (9), and A is the area of the deflection unit (m^2).

$$\nabla^4\omega = \frac{\partial^4\omega}{\partial x^4} + 2\frac{\partial^4\omega}{\partial x^2\partial z^2} + \frac{\partial^4\omega}{\partial z^4} \tag{7}$$

$$D = \frac{Et^3}{12(1-\upsilon^2)} \tag{8}$$

$$k_T = \frac{\alpha_T E A_T}{\lambda l_0 S_1 S_2} \tag{9}$$

where t is the thickness of the retaining wall (m); υ is the Poisson's ratio; λ is the support adjustment coefficient, which is taken as 0.5 when the soil properties, depth, and surrounding loads on both sides of the pit are similar; α_T is the support relaxation coefficient, typically ranging from 0.8 to 1.0 when the internal support does not consider pre-applied axial stress; A_T is the cross-sectional area of the support (m^2); l_0 is the length of the internal support (m); S_1 is the horizontal spacing of the support (m); and S_2 is the vertical spacing of the support (m).

If the soil at the excavation site comprises heterogeneous strata (Fig. 2a), further refinement of Eq. (6) is required based on the burial depth of different soil strata. The burial depth of the first soil stratum, the burial depth of the second soil stratum, and the excavation depth are denoted as H_1, H_2, and H_3, respectively. The deflection PDE for the retaining wall in heterogeneous strata can be expressed as

$$\left.\begin{aligned} D\nabla^4\omega_1(x,z)+k_T\omega_1(x,z)A&=[e_0-k_s\omega_1(x,z)]A \quad (0\le z\le H_1)\\ D\nabla^4\omega_2(x,z)+k_T\omega_2(x,z)A&=[e_0-k_s\omega_2(x,z)]A \quad (H_1<z\le H_2)\\ D\nabla^4\omega_3(x,z)+k_T\omega_3(x,z)A&=[e_0-k_s\omega_3(x,z)]A \quad (H_2<z\le H_3)\\ D\nabla^4\omega_4(x,z)+[e_{0b}+k_s\omega_4(x,z)]A&=\left[e_{0f}-k_s\omega_4(x,z)\right]A \quad (z>H_3) \end{aligned}\right\} \tag{10}$$

Eq. (10) can be updated according to the number of traversed soil strata during excavation.

The corner of the foundation pit is the connection position of two retaining walls in different directions, which leads to a higher rigidity and significantly small deflection (Iwata et al., 2007). Assuming that the length of the retaining wall is L, the boundary conditions for the deflection and deflection gradient at both ends of the retaining wall can be set as

$$\left.\begin{aligned} \omega(x=0,z)=0,\frac{\partial\omega(x=0,z)}{\partial x}=0\\ \omega(x=L,z)=0,\frac{\partial\omega(x=L,z)}{\partial x}=0 \end{aligned}\right\} \tag{11}$$

In addition, considering the bending stiffness of the retaining structure, its deflection variations are generally smooth and continuous. To prevent abrupt changes in deflection at the different soil interfaces during calculation, appropriate displacement continuity constraints should be applied. Thus, at the soil interfaces $z=H_1$ and H_2, and the excavation surface $z=H_3$, the deflection continuity conditions should be imposed:

$$\left.\begin{aligned} &[\omega_1(x,z)-\omega_2(x,z)]|_{z=H_1}=0\\ &\left[\frac{\partial\omega_1(x,z)}{\partial z}-\frac{\partial\omega_2(x,z)}{\partial z}\right]\Big|_{z=H_1}=0\\ &[\omega_2(x,z)-\omega_3(x,z)]|_{z=H_2}=0\\ &\left[\frac{\partial\omega_2(x,z)}{\partial z}-\frac{\partial\omega_3(x,z)}{\partial z}\right]\Big|_{z=H_2}=0 \end{aligned}\right\} \tag{12}$$

$$\left.\begin{aligned} &[\omega_3(x,z)-\omega_4(x,z)]|_{z=H_3}=0\\ &\left[\frac{\partial\omega_3(x,z)}{\partial z}-\frac{\partial\omega_4(x,z)}{\partial z}\right]\Big|_{z=H_3}=0 \end{aligned}\right\} \tag{13}$$

The above boundary conditions and deflection continuity conditions encompass both the Dirichlet and Neumann boundary conditions (Gong et al., 2024).

It should be noted that the applicability of Eq. (10) for 3D RWD is subject to certain practical constraints. The soil springs used to characterize soil-structure interaction in this PDE are linear elastic springs, primarily capturing the small deformation behaviour of the soil before it enters the plastic stage. For large deformations or highly nonlinear conditions, the applicability of this PDE may be limited. However, since this PDE is specifically designed for braced excavation in foundation pit engineering, it remains suitable. Compared to other excavation methods, braced excavation provides the best control over soil and retaining structure deformation, generally constrained to within 40 mm (Zhou et al., 2024). Therefore, it can be reasonably approximated that the soil deformation remains within the linear elastic range under braced excavation disturbances.

2.2. PINN and data-driven PINN models

Due to the complexity of Eq. (10), obtaining its general solution through known boundary and continuity conditions is extremely difficult. Additionally, m in Eq. (2) is determined using an empirical method (Zhao et al., 2012); it is assumed that when $\omega(x,z=H)$ <10 mm, $\omega(x,z=H)$ can be approximated as 10 mm. Therefore, even when the general solution to Eq. (10) is obtained, the precision is affected by the accuracy of the m value. Considering that the deflection PDE of the retaining wall has already been derived, PINNs can be used to predict and solve the RWD.

Conventional ML models such as the multi-layer perceptron (MLP) (Fig. 3) employ loss functions to quantify the inconsistency between the predicted and actual values (Kolay and Baser, 2014; Borate et al., 2023). The loss function is crucial for model training because it quantifies the network performance. It facilitates back-propagation by enabling the neural network to learn through adjusting biases and weights to minimize the loss, and it also serves as a critical evaluation metric.

PINNs constrain neural networks using physical and mechanical laws governed by PDEs. Instead of relying exclusively on extensive datasets for training, PINNs integrate these physical laws into the learning algorithm (Buzaev et al., 2024; Gong et al., 2024). This integration allows PINNs to infer from significantly smaller datasets compared with conventional ML models. Moreover, by embedding known physics into their structures, PINNs exhibit superior generalization capabilities beyond their training datasets (Meng et al., 2023), adhering to physical laws that enhance their robustness and reliability in prediction tasks (Cai et al., 2021; Sun et al., 2023).

This study proposes two RWD prediction models based on the newly established PDE: the PINN and data-driven PINN models

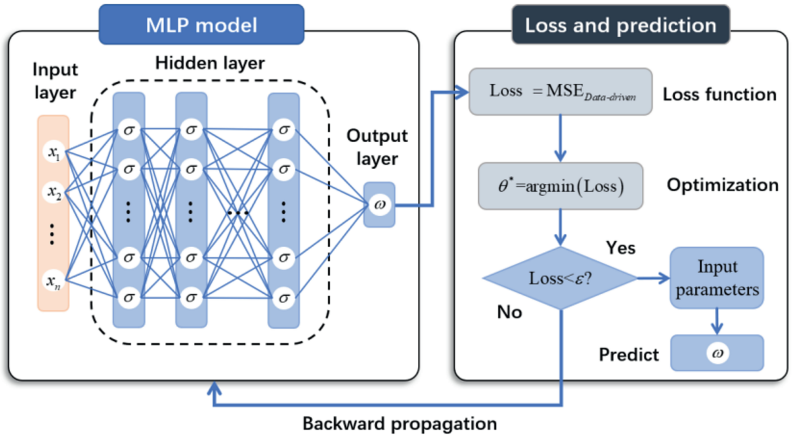

Fig. 3. MLP framework.

(Fig. 4). Both models are constructed based on MLP and share the same structure. They can automatically differentiate ("AD" in Fig. 4) variables during the training process based on PDE constraints (Gong et al., 2024). The PINN model does not rely on actual engineering data, but can be considered an intelligent PDE solver. The input of the PINN model comprises only the deflection coordinates x and z of the units on the retaining wall, which can be generated through random sampling. The other PDE parameters, such as the excavation dimensions, mechanical properties of the retaining wall and internal supports, and surrounding geological parameters, are known constants within the model. The PDE is incorporated into the PINN model as a physical constraint through the physical loss function, enabling optimisation of the network structure using only the coordinates x and z as parameters.

The data-driven PINN model requires learning from existing RWD case data. Based on the factors influencing RWD summarised in previous studies (Kung et al., 2007; Zhang et al., 2015) and the PDE parameters, the input for the training set of the data-driven PINN model can be determined. These input parameters include the coordinates x and z of the deflection units on the retaining wall, length L of the retaining wall, burial depths H_1 of the soil stratum interface, excavation depth H_2, soil internal friction angle φ, soil cohesion c, soil weight γ, bending stiffness D of the retaining wall, and stiffness k_T of the internal supports, amounting to 11 input parameters. When the number of soil strata traversed by the excavation increases, the number of soil stratum interfaces and input parameters increase correspondingly. The correlation analysis between the input parameters and RWD is shown in Fig. 5. It can be observed from Fig. 5 that there is a good correlation between the input parameters and RWD, which allows the model to effectively learn the relationship between the input parameters and RWD. The absolute values of the correlation coefficients between different input parameters are all less than 0.5, which meets the requirement of weak correlation.

Compared to the PINN model, the total loss function of the data-driven PINN model includes both the PDE-based physical loss and data loss. As shown in Fig. 4, their total loss function can be expressed as

$$\left.\begin{aligned}\text{Total loss}_{\text{PINN}} &= \lambda_{\text{PDE}}L_{\text{PDE}} + \lambda_{\text{B}}L_{\text{B}} + \lambda_{\text{C}}L_{\text{C}} \\ \text{Total loss}_{\text{Data-PINN}} &= \lambda_{\text{data}}L_{\text{data}} + \lambda_{\text{PDE}}L_{\text{PDE}} + \lambda_{\text{B}}L_{\text{B}} + \lambda_{\text{C}}L_{\text{C}}\end{aligned}\right\} \tag{14}$$

where L_{data} is the loss of the data; L_{PDE} is the loss of the PDE; L_{B} is the loss of the boundary condition; L_{C} is the loss of the deflection continuity condition; and λ_{data}, λ_{PDE}, λ_{B}, and λ_{C} are the weights for each loss function, respectively.

In addition, m is treated as a trainable neural network weight in both the PINN and data-driven PINN models, thereby optimizing the calculation accuracy for k_s and improving the prediction accuracy of the RWD.

The tanh function was used as the activation function for the hidden layer due to its balanced output range and smooth derivatives, which can help improve convergence during training (Gong et al., 2024). The proposed PINN and data-driven PINN models can be evaluated by comparing the predicted values to the field monitoring and numerical simulation results. Three metrics, i.e. the mean absolute error (MAE), root mean square error (RMSE), and coefficient of determination (R^2) are adopted:

$$MAE = \frac{\sum_{(x,z)} \left| y^{\text{PINNs}}_{(x,z)} - y^{\text{Field}}_{(x,z)} \right|}{N_{(x,z)}} \tag{15}$$

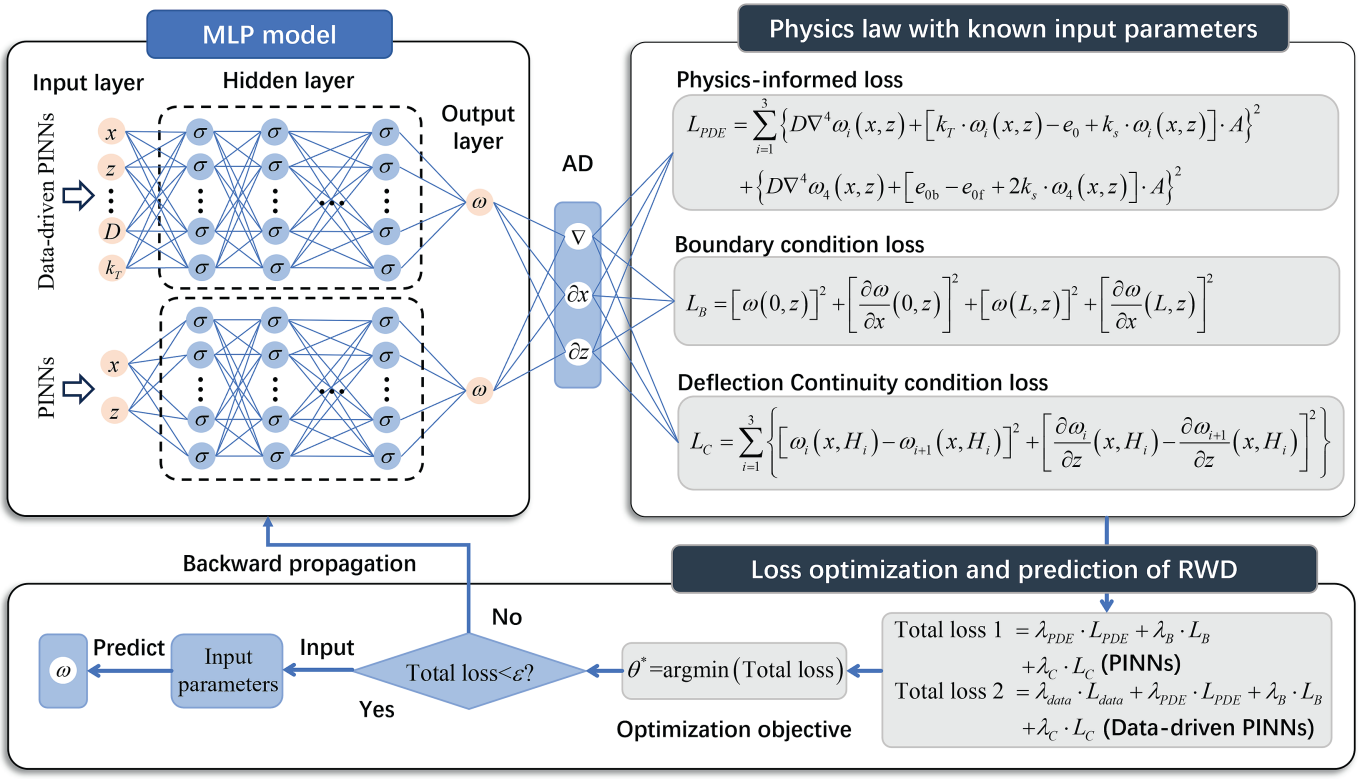

Fig. 4. PINN and data-driven PINN frameworks.

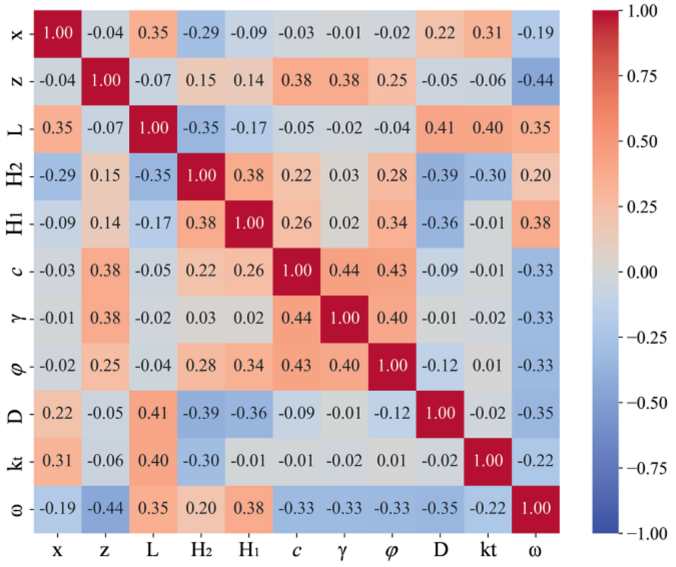

Fig. 5. Correlation among input parameters.

$$RMSE = \sqrt{\frac{\sum_{(x,z)} \left(y^{\text{PINNs}}_{(x,z)} - y^{\text{Field}}_{(x,z)}\right)^2}{N_{(x,z)}}} \tag{16}$$

$$R^2 = 1 - \frac{\sum_{(x,z)} \left(y^{\text{PINNs}}_{(x,z)} - y^{\text{Field}}_{(x,z)}\right)^2}{\sum_{(x,z)} \left(y^{\text{PINNs}}_{(x,z)} - \overline{y}^{\text{Field}}\right)^2} \tag{17}$$

where $N_{(x,z)}$ is the number of coordinate points, $y^{\text{PINNs}}_{(x,z)}$ is the predicted value of coordinate point (x, z), $y^{\text{Field}}_{(x,z)}$ is the field value of coordinate point (x, z), and $\overline{y}^{\text{Field}}$ is the mean value of the field data.

3. PINNs for forward problems: Prediction of RWD

3.1. Case setting

In this section, the effectiveness of the proposed PINN and data-driven PINN models is analysed based on the deflection of the long-side retaining wall in an actual foundation pit. The case studied is the No. 1 foundation pit project of the Qingliangshan (QLS) underground station of Nanjing Metro Line 7, China. This station was constructed using the cut-and-cover method, with an excavation length of approximately 50 m, width of approximately 23.5 m, and total excavation depth of approximately 45 m. The foundation pit is divided into ten excavation layers, and the depth of each excavation is approximately 5 m. The bottom end of the retaining wall is embedded 5 m in the soil.

As shown in Fig. 6, the area enclosed by the dashed lines represents the geological strata where the long-side retaining wall is located. The upper soil strata at the excavation site mainly comprise soft plastic silty clay with poor geological properties, and the lower strata comprise plastic to hard-plastic medium mudstone and sandstone. The key geological parameters are listed in Table 1.

The first five excavation strata are supported by ϕ609 mm steel pipes, while the remaining strata are supported by concrete bracings. The steel pipe has an elasticity modulus of 210 MPa, and

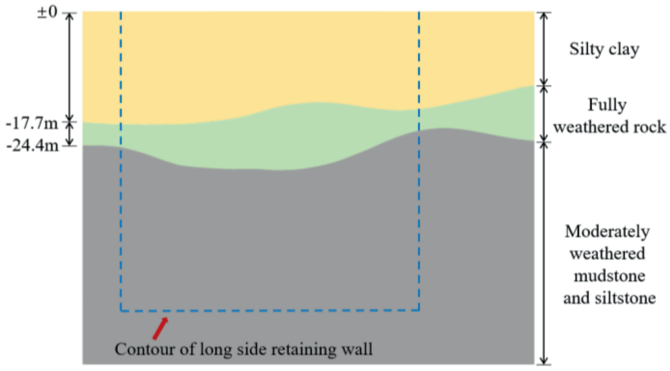

Fig. 6. Diagram showing the stratigraphic distribution.

Table 1
Soil parameters of the QLS foundation pit.

Soil layer	Depth (m)	γ (kN/m^3)	E (MPa)	φ (°)	c (kPa)
Silty clay	0–17	19.6	14	29.1	28.3
Fully weathered rock	17–27	21.4	70	30.6	70.2
Mudstone and siltstone	>27	24.6	900	38.3	500

the concrete bracing has a cross-sectional size of 700 mm × 800 mm with an elasticity modulus of 30 MPa. The lateral spacing between the bracings in each layer is 4 m. The vertical retaining structure of the foundation pit is a diaphragm wall with a thickness of 0.9 m and an elasticity modulus of 30 MPa.

This foundation pit is located in an old urban district surrounded by dense buildings. The construction team conducted frequent monitoring of the RWD, surface settlement, and deep-soil horizontal displacement. The deflection monitoring positions for the long-side walls are shown in Fig. 7. The clinometers CX1 and CX7 are located at the corners, CX4 is located at the centre, CX2 and CX6 are 9 m away from the corners, and CX3 and CX5 are 7 m away from the centre. The vertical distance between monitoring points in each clinometer tube is 1 m.

Fig. 8a shows the monitoring values of the RWD when the excavation reaches the bottom (45 m). The inverse distance-weighting interpolation method (Zhou et al., 2024) was used to generate the deflection cloud map of the entire retaining wall. As shown in Fig. 8b, each interpolated displacement coordinate point is spaced at 1 m. This interpolation helps in visualizing the deflection distribution across the entire structure and serves as a valuable reference for evaluating the performance of the proposed prediction models.

3.2. Datasets for model training and testing

3.2.1. PINN model

Based on the soil-structure interaction analysis discussed in Section 2.1, the retaining wall is discretized into deflection units (Fig. 9a). When using soil springs to characterize the soil-structure interaction, the stiffness coefficients of the springs vary across different soil strata, and thus, the soil springs must not cover multiple soil strata (Abbasi et al., 2014). Moreover, to facilitate the construction of displacement-continuity conditions at the soil interfaces, the discretized grids should also avoid covering the soil layer interface. After comprehensive consideration, the discretization grid was set to 1.0 m × 1.0 m in this study, with each unit's interaction with the soil being represented by a soil spring. Each unit's deflection is represented by the deflection value at the centre of grid.

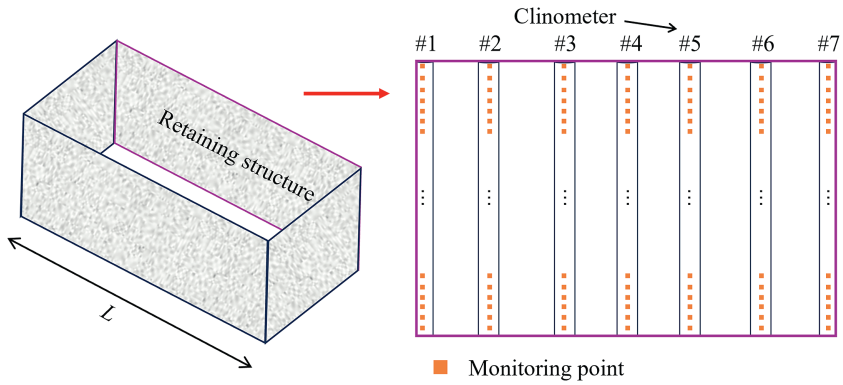

Fig. 7. Distribution of RWD monitoring points.

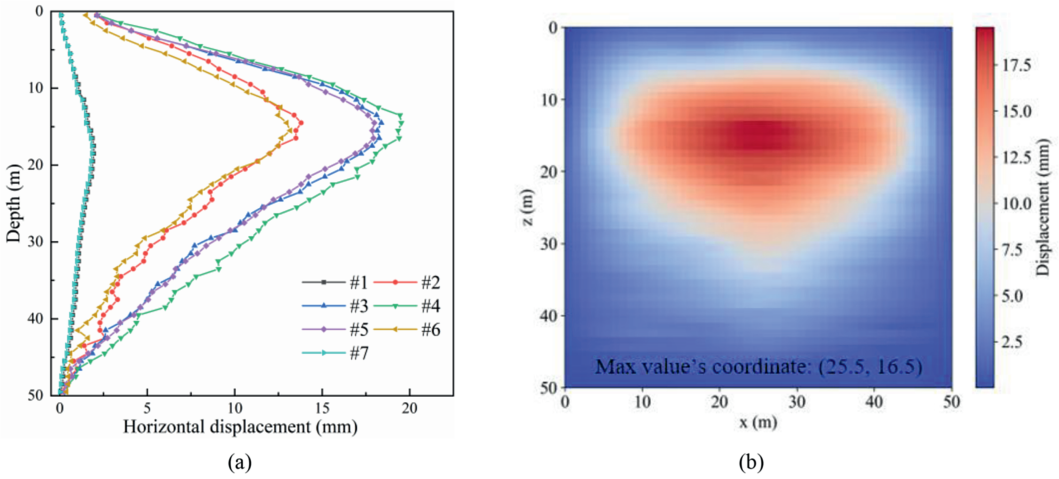

Fig. 8. Monitoring values of RWD: (a) Field monitoring value, and (b) Deflection cloud map of the retaining wall.

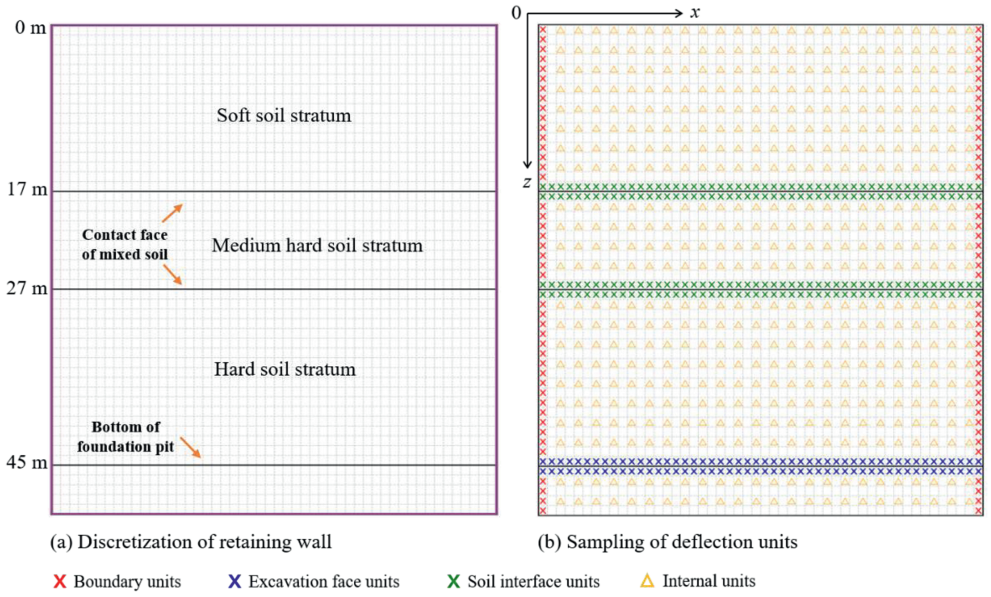

Fig. 9. Diagram of deflection units on the retaining wall.

The sampling points for the training set of the PINN model are shown in Fig. 9b. The sampling points comprise boundary points at $x = 0.5$ and 49.5 m; deflection continuous points at $z = 16.5$, 17.5, 26.5, 27.5, 44.5 and 45.5 m; and internal points (x, z) within the

domain. At the boundaries, the coordinates of the centre points of the units are extracted for the model to learn the boundary conditions (see red "x" in Fig. 9b). At the interfaces of soil strata and the excavation face, the coordinates of the centre points of the units from the two adjacent rows above and below these surfaces are extracted for the model to learn the deflection continuity conditions (green "x" and blue "x" in Fig. 9b). In the other areas, uniform sampling is performed to allow the model to learn the overall deflection characteristics of the structure under the PDE constraints (yellow "Δ" in Fig. 9b). The test set for the PINN model comprises the coordinates of the centre points of all units, as shown in Fig. 9b, amounting to 2500 input samples.

3.2.2. Data-driven PINN model

In the data-driven PINN model, the parameters of the training dataset must match the test set (QLS excavation case). When merging adjacent soil stratum into a unified stratum with similar mechanical properties, the number of strata the excavation traverses after merging should not exceed three. When there are fewer than three original strata, further refinement can be applied to transform them into three strata. However, if the mechanical property differences between adjacent soil stratum are substantial and cannot be simplified into three primary strata through merging, the excavation is not suitable to form a training set.

The training data for the data-driven PINN model used in this study are derived from actual foundation pit cases or refined numerical results supported by actual projects in previous studies. Five foundation pit cases, with a total of 462 datasets, are collected (Wang et al., 2005; Orazalin et al., 2015; Goh et al., 2017; Zhang et al., 2019b; Zhao et al., 2021). The statistical information of these 462 datasets is summarised in Table 2, including the minimum value (Min), maximum value (Max), mean, and standard deviation (St.D).

To reduce the impact of numerical differences between the different physical quantities on model training, all input and output variables in both the training and test datasets were standardized based on their ranges of variation, with all parameters normalized to the [0, 1] interval. The test set input parameters x and z are the coordinates of the centre points of all units in Fig. 9b. The other input parameters, such as γ, c, and φ, represent the soil parameters behind the wall corresponding to each grid unit. The parameters L, H_1, H_2, H_3, D, and k_T are fixed parameters for the QLS pit with the same value assigned to each unit.

3.3. PINN hyper-parameter optimisation

The primary objective of training the PINN and data-driven PINN models was to minimize the loss function, with the final loss value serving as a quantitative measure of convergence improvement. However, different hyperparameter settings can significantly affect the predictive accuracy. Therefore, before using the models for prediction, different learning rates, network layers, and numbers of neurons were explored to assess their impact on the convergence of the proposed models.

Table 2
Summary of RWD datasets.

Parameter	Min	Max	Value of mean	St.D
L (m)	45	110	74.3	27.6
H_1 (m)	10	30	21.5	8.4
H_2 (m)	20	45	31.5	10.9
H_3 (m)	15	50	34.7	14.9
γ (kN/m^3)	18.3	23.2	19.9	1.6
c (kPa)	15	55.4	30.2	15.8
φ (°)	20	30.2	28.5	5.4
D (MPa)	1.5	3	2.2	0.9
k_T (kPa)	0.8	1.5	1.1	0.2
ω (mm)	15	65	36.3	21.6

To study the impact of different learning rates on the two models, the MLP network structure was set as (128, 64, 32, 16, 8). The weights for the data loss, PDE loss, boundary loss, and deflection continuity loss were initially set to 1.0, 0.4, 10 and 10, respectively. Fig. 10 shows the effect of learning rates of 1×10^{-2}, 1×10^{-3}, 1×10^{-4} and 1×10^{-5} on the convergence ability of the two models. In the PINN model, convergence improved with increasing number of iterations. A learning rate of 1×10^{-2} enhanced the training stability, but the loss value was relatively large, which was detrimental to model convergence. However, when the learning rates were set to 1×10^{-3}, 1×10^{-4} and 1×10^{-5}, the model's convergence ability was improved significantly. At a learning rate of 1×10^{-3}, the total loss value was minimized, indicating optimal convergence, and the model reached its optimal state within 15,000 training epochs. In the data-driven PINN model, different learning rates all converged at approximately 16,000 epochs. Similar to the PINN model, the total loss value was minimized at a learning rate of 1×10^{-3}. Therefore, the learning rate of 1×10^{-3} was selected for both the PINN and data-driven PINN models.

The impact of hidden layers and neurons per hidden layer on the performance of the proposed models was examined. Given the large variety of input parameters and a single output parameter in this study, as well as the need for the models to learn the boundary and continuity characteristics of the existing RWD data, an MLP structure was employed in which the number of neurons decreased progressively with each layer (Borate et al., 2023). Specifically, the number of neurons in each subsequent hidden layer was set to 0.5 times that of the previous layer. In investigating the impact of the different hidden layers and neurons per hidden layer on the model convergence, the learning rates for both models were uniformly set to 1×10^{-3}, and the training epochs were set to 16,000. The weights for the data loss, PDE loss, boundary loss, and deflection continuity loss were also set to 1.0, 0.4, 10 and 10, respectively.

Both the PINN and data-driven PINN models featuring three, four, five, six, and seven hidden layers, each with eight neurons of the last layer, were developed to investigate the impact of hidden layers. Fig. 11a shows that the training time increased with increasing number of hidden layers, but the models' convergence did not improve with more hidden layers. With the same number of hidden layers, the training time of the data-driven PINN model was approximately 200 s longer than that of the PINN model. The total loss value was minimized for the PINN and data-driven PINN models when the number of hidden layers was five and six, respectively. Therefore, the recommended hidden layers for the PINN and data-driven PINN models are five and six, respectively.

To assess the impact of neuron numbers, the PINN and data-driven PINN models were constructed with two, four, six, eight, and ten neurons in the last hidden layer, while maintaining five hidden layers and six hidden layers, respectively. As shown in Fig. 11b, the convergence of both the models improved gradually with the increasing number of neurons in the last hidden layer. The total loss value of both models decreased significantly when the number of neurons in the last hidden layer increased from two to six. However, when the number of neurons in the last hidden layer was set to eight and ten, the change in the total loss value was less noticeable, with differences within the range of 0.3×10^{-3}. Considering the efficiency of model training, it is reasonable to set the number of neurons in the last hidden layer to eight for both models.

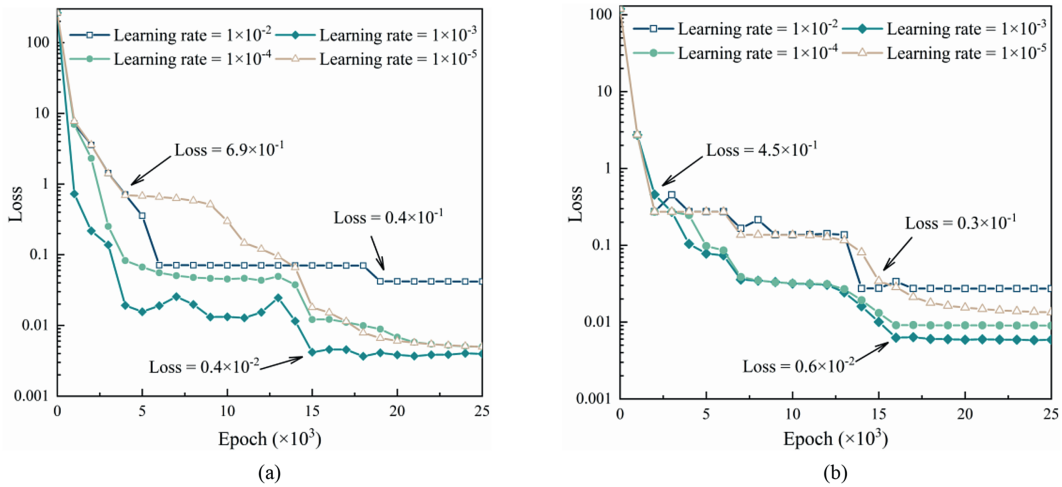

Fig. 10. Impact of different learning rates: (a) PINN model, and (b) Data-driven PINN model.

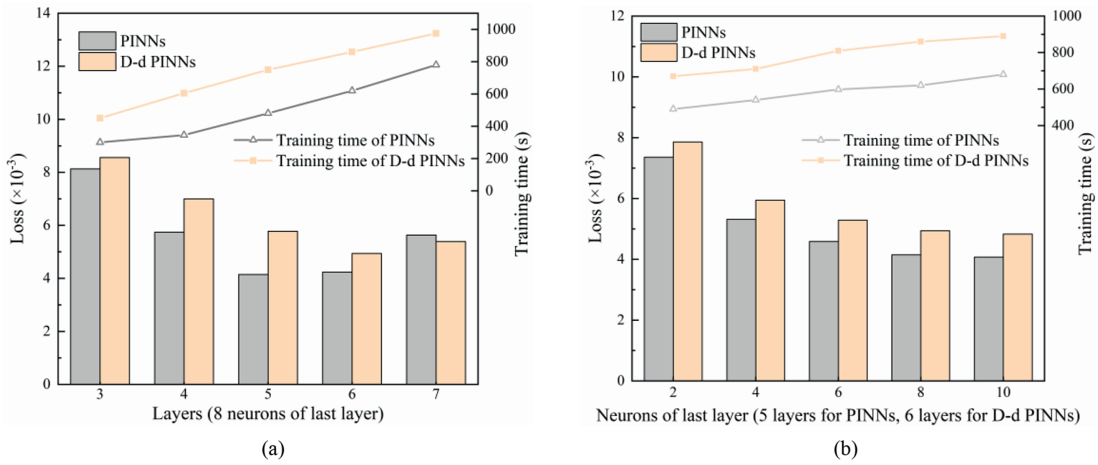

Fig. 11. Impact of changing hidden layers and neurons: (a) Different hidden layers, and (b) Different neurons of last layer.

3.4. Performance of the PINN and data-driven PINN models

Based on the hyperparameters analysis discussed in Section 3.3, the PINN model was configured with five hidden layers, with neurons set as (128, 64, 32, 16, 8); and the data-driven PINN model was configured with six hidden layers, with neurons set as (256, 128, 64, 32, 16, 8). Both models adopted a learning rate of 1×10^{-3}, with weights of the data loss, PDE loss, boundary loss, and deflection continuity loss set at 1.0, 0.4, 10 and 10, respectively. Subsequently, two models were used to predict the deflections of the QLS long retaining wall. Both models were trained for 16,000 epochs, with the entire training and prediction process implemented using the PyTorch framework. Additionally, a numerical model of the QLS foundation pit was developed using the PLAXIS 3D software (Fig. 12), providing a basis for comparing the advantages of the PINN model, data-driven PINN model, and the numerical model.

Fig. 13a, 14a and 15a show the results from the numerical simulations, PINN model, and the data-driven PINN model, respectively. It can be observed that the deflection maps of the retaining wall derived from the three models exhibit a similar core shape (red region), which is approximately elliptical. Additionally, the coordinates of the maximum deflection point on the retaining

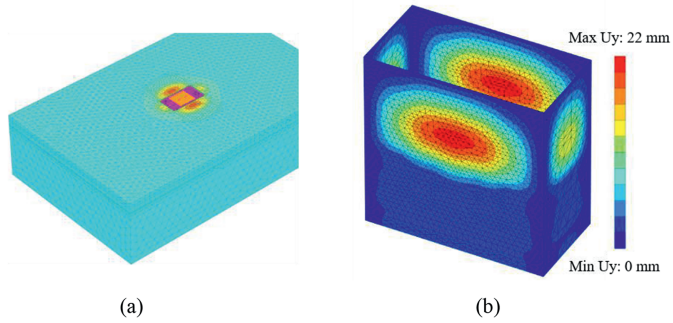

Fig. 12. Numerical simulation results based on PLAXIS 3D: (a) Numerical model of QLS pit, and (b) Cloud map of RWD.

wall are relatively consistent across the models. The main differences among the deflection maps lie in the hard soil stratum ($z > 27$m), where the numerical model shows relatively smaller deflection. In contrast, the PINN model tends to overestimate the deflection, while the data-driven PINN model provides predictions that are closer to the actual values. Figs. 13b, 14b and 15b illustrate the relative error between each model and the actual deflection

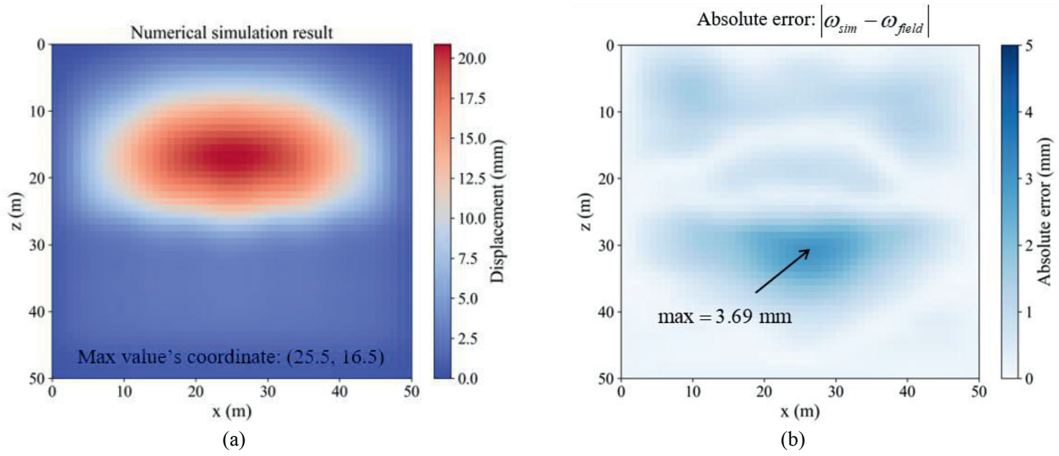

Fig. 13. Numerical results and absolute error: (a) Simulation results of PLAXIS 3D, and (b) Absolute error.

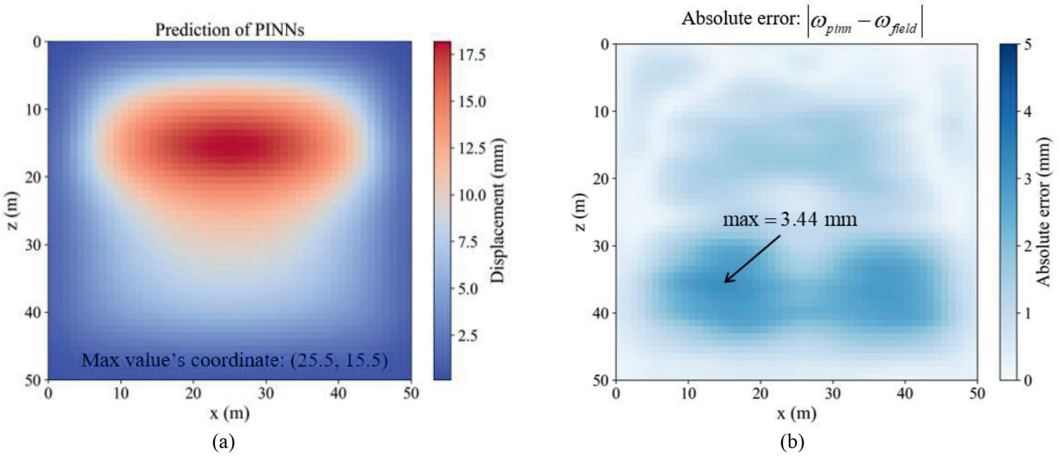

Fig. 14. Prediction of the PINN model and absolute error: (a) Prediction of the PINN model, and (b) Absolute error.

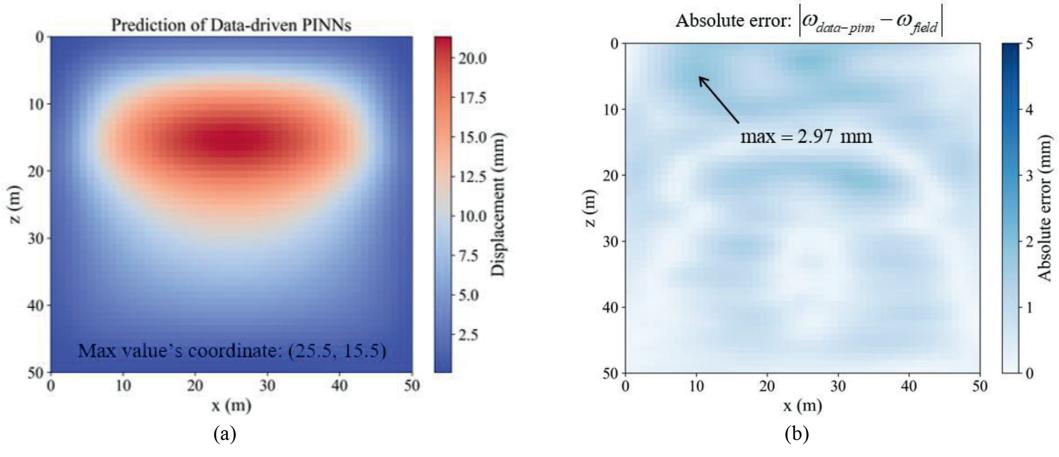

Fig. 15. Prediction of the data-driven PINN model and absolute error: (a) Prediction of the data-driven PINN model, and (b) Absolute error.

surface of the retaining wall. The maximum error between the numerical calculation results and actual values is 3.69 mm. For the PINN model, the maximum error is 3.44 mm, with a larger deep-blue area in the absolute error map compared to the numerical model and the data-driven PINN model. The data-driven PINN model achieves the smallest maximum error of 2.97 mm, with the smallest deep-blue area in its error map.

The evaluation of the computational and predictive accuracy of each model is summarised in Table 3. It can be observed that when comparing the deflection surface of the retaining wall or the

Table 3
Error statistics of each model.

Metrics	PLAXIS 3D	PINN	Data-driven PINN
MAE	0.863	1.081	0.827
RMSE	1.09	1.262	0.968
R^2	0.852	0.813	0.864
MAE (Actual value)	0.902	1.116	0.875
RMSE (Actual value)	1.071	1.357	0.994
R^2 (Actual value)	0.839	0.801	0.847

monitored values at actual inclinometer points, the data-driven PINN model exhibits the highest predictive accuracy, followed by the numerical and PINN models. Moreover, in the data-driven PINN model, the correlation between the predicted and actual values (R^2) is the highest, followed by the numerical and PINN models. Compared to the numerical model and the PINN model, the data-driven PINN model improves the MAE and RMSE by 9.3 % and 15.8 %, and 27.1 % and 27.2 %, respectively. In routine research on excavation deformations, the numerical model is as a conventional research tool that provides valuable reference data due to its reliable computational accuracy. However, compared to the PINN and data-driven PINN models, the numerical model involves a more complex modelling process and longer computation time. For instance, as shown in Fig. 12, the numerical model requires approximately 2 h of backend computation and extensive modelling. In contrast, the data-driven PINN model completes the predictive analysis in 15 min (approximately 900 s) without requiring any complex modelling process, demonstrating excellent computational efficiency and accuracy for engineering applications, which is suitable for practical applications. Using a computer equipped with a high-performance GPU, the model training and prediction speed can be improved further. The computer used in this study is equipped with an RTX 3060 Ti, which allows enabling CUDA in the Torch framework to run both the PINN and data-driven PINN models on the GPU. For the QLS case, with GPU support, the runtime of the PINN model can be reduced from 620.3 s to 416.9 s, and the runtime of the data-driven PINN model can be reduced from 896.7 s to 568.2 s.

Furthermore, under the condition of fixing the weights of the data loss, boundary loss, and deflection continuity condition loss, the impact of changing the PDE loss weights (0.1–0.9 for PINN and 0–0.9 for data-driven PINN) on the models' predictive accuracy is further analysed. As shown in Fig. 16, the prediction error of both models decreases initially and then increases with increasing PDE loss weight. The PINN model achieves the highest prediction accuracy when the PDE loss weight is in the range of 0.2–0.3, and the data-driven PINN model achieves optimal accuracy at a PDE loss weight of 0.5–0.6. Therefore, a PDE loss weight of 0.3 can be recommended for the PINN model, and 0.5 can be recommended for the data-driven PINN model.

4. Discussion: PINN model applied in non-uniform strata

4.1. RWD in non-uniform strata

Similar to the conventional numerical methods, the aforementioned RWD prediction models simplify the strata into uniformly distributed soil strata when addressing the influence of soft-hard soil interfaces. This simplification is suitable for geological bodies with relatively stable stratigraphic undulations within the excavation range. However, excavation projects inevitably encounter geological bodies with significant stratigraphic fluctuations. Simplifying such conditions into idealized strata may lead to a substantial discrepancy between the computed RWD and actual monitored values. For instance, in the case of the QLS excavation, the strata exhibited notable undulations (Fig. 17a). Therefore, this section discusses the application of the PINN model to predict RWD in complex stratigraphic distributions.

Compared to RWD in uniformly distributed strata, deflection in non-uniformly distributed strata requires defining new deflection regions and deflection continuity conditions. In non-uniformly distributed strata, the soil interfaces are irregular, making it difficult to characterize them with a clear depth parameter. Therefore, to align with the retaining wall's grids, we employed a pixel-based segmentation method for soil meshing (Peng et al., 2023). Specifically, the grid division size was determined based on the number of pixels. Each grid was then uniformly coloured according to the RGB value of the most dominant pixel category within it, ensuring that each divided grid corresponded to a single pixel block. In this study, a 1.0-m grid corresponded to ten pixels, a 0.5-m grid to five pixels, and a 0.2-m grid to two pixels (Fig. 17b and c). After discretizing the retaining wall, it is necessary to construct indices for all grid units using the coordinates of the unit centre points to determine the deflection region (corresponding soil stratum) of each unit and assign different PDE loss functions. Additionally, based on the unit indices, it is essential to sample

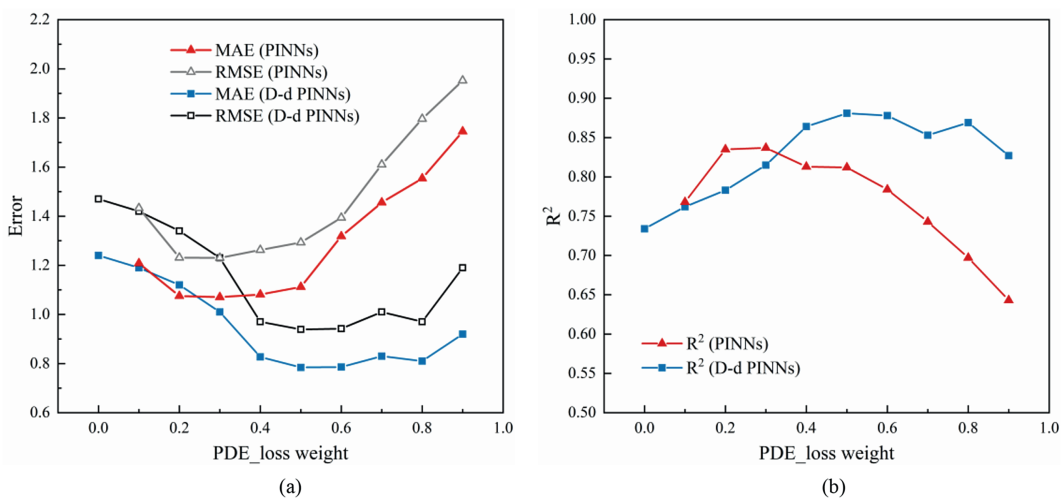

Fig. 16. Impact of PDE loss weight on (a) MAE and RMSE, and (b) R^2.

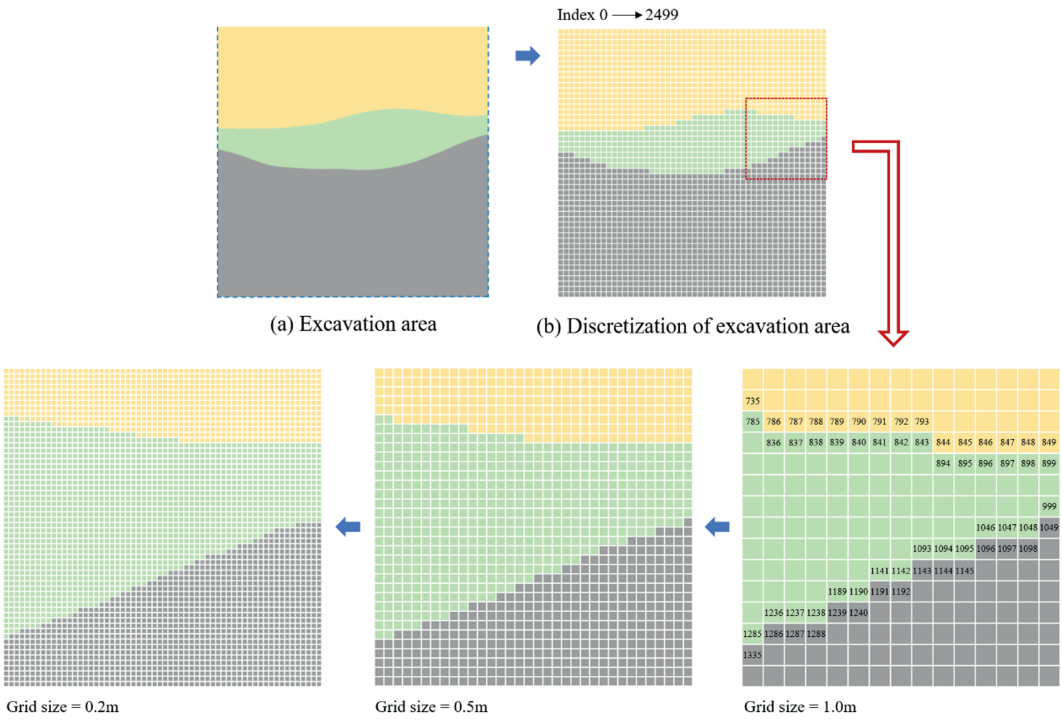

Fig. 17. Strata condition of the QLS foundation pit.

units at the irregular interfaces (Fig. 17c) to construct the loss function for deflection continuity conditions.

In the data-driven PINN model, the deflection continuity loss function at the soil interfaces and the PDE loss within different soil strata are constructed using input parameters H_1 and H_2. For a given excavation condition, H_1 and H_2 are fixed depth parameters, and the data-driven PINN model cannot learn the RWD at irregular soil interfaces during the training process. Therefore, the data-driven PINN model is not suitable for RWD prediction in non-uniformly distributed strata. In contrast, the input parameters of the PINN model are derived from random sampling within the retaining wall coordinate domain. These sampled units can cover the grid units at irregular soil interfaces, thereby enabling the construction of accurate loss functions for deflection continuity conditions at irregular soil interfaces.

4.2. Comparative analysis of prediction results

Fig. 18 shows the PINN model prediction results in non-uniformly distributed strata when the size of grid unit is 1.0 m × 1.0 m. The corresponding MLP structure is (128, 64, 34, 16, 8), with the learning rate set at 1×10^{-3}, and the PDE weight set at 0.4. As shown in Fig. 18a, the shape of the core deflection area (red region) of the retaining wall changes slightly. Due to the "concave" distribution of the third hard-soil stratum, the core deflection area on both sides of the retaining wall exhibits an inward trend, transforming from an "elliptical" shape (as shown in Figs. 13–15) to a "heart" shape, which is consistent with the actual deflection surface (Fig. 8b). Based on the absolute error analysis in Fig. 18b, it can be observed that the maximum absolute error decreased from 3.44 mm in Figs. 14b to 3.06 mm, with a reduction in the area of the deep blue region. Compared to the PINN model prediction shown in Fig. 14, the MAE and RMSE decreased from 1.081 and 1.262 to 0.942 and 1.017, respectively, and R^2 increased from 0.813 to 0.825. The MAE and RMSE of the actual values also decreased from 1.216 and 1.357 to 1.004 and 1.141, respectively, and R^2 increased from 0.801 to 0.819.

The training time for the PINN model in non-uniformly distributed strata is 714 s. Notwithstanding this increase in training time compared to the 620 s required for the uniform distribution strata (Fig. 14), the prediction accuracy of the PINN model was further improved. Therefore, for RWD prediction with approximately uniform strata, the data-driven PINN model can be prioritized. However, for RWD in non-uniformly distributed strata, only the PINN model can achieve fast and accurate prediction.

The impact of different grid sizes (pixel blocks) on the prediction accuracy and efficiency of the PINN model (Fig. 17c) was also studied. When the grid size was reduced to 0.5 m × 0.5 m, the MAE, RMSE, and R^2 were 0.927, 1.008, and 0.832, respectively, with a total training and prediction time of 22 min and 57 s. When the grid size was further reduced to 0.2 m × 0.2 m, the MAE, RMSE, and R^2 improved slightly to 0.916, 0.997, and 0.841, respectively, but the total training and prediction time increased significantly to 39 min and 13 s. It can be observed that as the grid resolution increases, the PINN model requires more training time; however, the improvement in prediction accuracy is relatively limited. Furthermore, when establishing the displacement continuity conditions at soil interfaces, the selection of corresponding soil elements (pixel blocks) is performed manually. As the grid refinement increases, more soil elements need to be extracted, making the construction of the displacement continuity conditions more complex. Therefore, for excavation projects without special requirements, a soil grid size of 1.0 m × 1.0 m is recommended for predicting the RWD in heterogeneous soil strata.

5. Conclusions

To predict the 3D RWD of braced excavation within heterogeneous strata, this study proposes two RWD prediction models: the

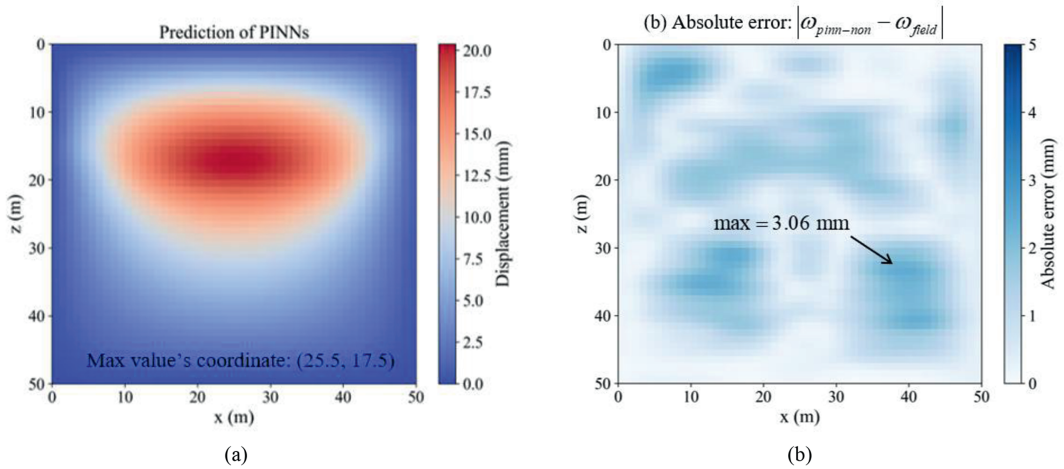

Fig. 18. Prediction of the PINN model (a) and absolute error (b) in non-uniformly distributed strata.

PINN and data-driven PINN models. The loss functions of both models are constructed based on a newly derived PDE for 3D RWD. Subsequently, the effectiveness of the proposed models was analysed using an actual excavation case from a subway project in Nanjing, China. The prediction results from both models were compared with the actual monitored data and numerical simulation results, exploring the predictive advantages and potential engineering applications of these models. The following conclusions are drawn:

(1) To characterize the interaction between the retaining wall and soil, this study derives a new 3D deflection PDE for retaining walls based on the soil-spring theory and the existing 2D deflection ODE. This new PDE is updated according to the different strata encountered during excavation, providing a foundation of physical information for the construction of both the PINN and data-driven PINN models.
(2) The loss function of the PINN model comprises the PDE loss, boundary condition loss, and deflection continuity condition loss. The loss function of the data-driven PINN model includes an additional data loss. For RWD prediction, when the weights of other loss functions are set at 1.0, the recommended range for the PDE loss weight in the PINN model is 0.2–0.3, and the recommended range for the PDE-loss weight in the data-driven PINN model is 0.5–0.6.
(3) The training times of both the PINN and data-driven PINN models show significant improvement compared to that of the numerical model. The prediction accuracy of the data-driven PINN model (MAE = 0.827, RMSE = 0.968, R^2 = 0.864) surpasses that of the numerical simulation results (MAE = 0.863, RMSE = 1.090, R^2 = 0.852), while the numerical simulation results are more accurate than those of the PINN model (MAE = 1.081, RMSE = 1.262, R^2 = 0.813). However, due to the influence of the form of the input parameters, the data-driven PINN model can only be applied to predict RWD under uniform strata distribution conditions.
(4) The randomly sampled coordinate points within the retaining wall coordinate domain serve as the input parameters for the PINN model. Therefore, in non-uniformly distributed strata, the PINN model can identify different RWD regions and irregular soil interfaces based on the grid unit indices. This enables the PINN model to learn the deflection continuity conditions at irregular interfaces by relying on the deflection continuity constraints, thereby improving the prediction accuracy of RWD in non-uniformly distributed strata.

The two models proposed in this study demonstrate their advantages considering different stratum conditions. They achieve accurate RWD predictions in heterogeneous strata and significantly reduce the computational time, thus addressing an important research gap in this field. However, the case in this study involves a regular rectangular excavation where the boundary conditions of the retaining structure are relatively simple. For excavations with irregular shapes, the complex forms of the retaining structures impose higher demands on the boundary condition definitions and deflection continuity conditions. For such cases, the applicability of both the PINN and data-driven PINN models needs further verification and adjustment based on actual monitored conditions. Optimizing these models for more complex scenarios is a key focus of future research.

CRediT authorship contribution statement

Cong Zhou: Writing – original draft, Software, Methodology. **Lei He:** Supervision, Resources, Project administration, Funding acquisition, Conceptualization. **Junchen He:** Visualization, Investigation. **Yi Zhang:** Data curation. **Huaiguang Xiao:** Writing – review & editing. **Chee Kiong Soh:** Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the Fundamental Research Funds for the Central Universities (Grant No. 2242023K5006) and the Jiangsu Civil Defense Office Program (Grant No. 7605009117).

References

Abbasi, O., Ghanbari, A., Hosseini, S.A.A., 2014. An analytical method for calculating the natural frequency of reinforced retaining walls with soil-structure interaction effect. Geosynth. Int. 21 (1), 53–61.
Bobylev, N., 2016. Underground space as an urban indicator: measuring use of subsurface. Tunn. Undergr. Space Technol. 55, 40–51.
Borate, P., Rivière, J., Marone, C., Mali, A., Kifer, D., Shokouhi, P., 2023. Using a

physics-informed neural network and fault zone acoustic monitoring to predict lab earthquakes. Nat. Commun. 14 (1), 3693.
Broere, W., 2016. Urban underground space: solving the problems of today's cities. Tunn. Undergr. Space Technol. 55, 245–248.
Buzaev, F., Gao, J.X., Chuprov, I., Kazakov, E., 2024. Hybrid acceleration techniques for the physics-informed neural networks: a comparative analysis. Mach. Learn. 113 (6), 3675–3692.
Cai, S., Mao, Z., Wang, Z., Yin, M., Karniadakis, G.E., 2021. Physics-informed neural networks (PINNs) for fluid mechanics: a review. Acta Mech. Sinica-Prc. 37 (12), 1727–1738.
Chen, Y.L., Chen, Z.L., Guo, D.J., et al., 2022. Underground space use of urban built-up areas in the central city of Nanjing: insight based on a dynamic population distribution. Undergr. Space 7 (5), 748–766.
Cheng, M.Y., Wu, Y.W., 2009. Prediction of diaphragm wall deflection in deep excavations using Evolutionary Support Vector Machine Inference Model (ESIM). In: Proceedings of the 26th International Symposium on Automation and Robotics in Construction (ISARC 2009), pp. 176–182.
Chua, C.G., Goh, A.T.C., 2005. Estimating wall deflections in deep excavations using Bayesian neural networks. Tunn. Undergr. Space Technol. 20 (4), 400–409.
Ding, S.R., Yang, H.Q., Xu, J.B., 2021. Probabilistic analysis of a braced excavation considering soil spatial variability. In: Proc. of 8th International Conference on Civil Engineering, pp. 151–159.
Dong, Y.P., Burd, H.J., Houlsby, G.T., 2022. Finite-element investigation of excavation-induced settlements of buildings and buried pipelines. J. Geotech. Geoenviron. Eng. 148, 04022072.
Feng, C.L., 2019. Study on the Safety and Control of Deep Foundation Pit Engineering in Subway Station Based on Complex Stratigraphic Conditions. Beijing Jiaotong University, Beijing. Ph. D. Dissertation (in Chinese).
Finno, R.J., Blackburn, J.T., Roboski, J.F., 2007. Three-dimensional effects for supported excavations in clay. J. Geotech. Geoenviron. Eng. 133, 30–36.
Goh, A.T.C., Zhang, F., Zhang, W.G., Zhang, Y.M., Liu, H.L., 2017. A simple estimation model for 3D braced excavation wall deflection. Comput. Geotech. 83, 106–113.
Gong, W.B., Zuo, L.L., Li, L., Wang, H., 2024. Prediction of stratified ground consolidation via a physics-informed neural network utilizing short-term excess pore water pressure monitoring data. Comput-Aided. Civ. Inf. https://doi.org/10.1111/mice.13326.
Hong, L., Wang, X.Y., Zhang, W.G., 2023. Reliability-based robust geotechnical design of braced excavations considering multiple failure modes. Undergr. Space 9, 43–52.
Hsieh, P.G., Ou, C.Y., 2016. Simplified approach to estimate the maximum wall deflection for deep excavations with cross walls in clay under the undrained condition. Acta Geotech. 11, 177–189.
Huang, M.Q., Ninic, J., Zhang, Q.B., 2021. BIM, machine learning and computer vision techniques in underground construction: current status and future perspectives. Tunn. Undergr. Space Technol. 108, 103677.
Ibrahim, N., Shahaboddin, Y., Ali, A.G., Mehdi, J., 2021. Damage estimation in masonry buildings based on excavation-induced ground movements. Geotech. Geol. Eng. 39, 4365-4285.
Iwata, N., Nakai, T., Zhang, F., Inoue, T., Takei, H., 2007. Influences of 3D effects, wall deflection process and wall deflection mode in retaining wall problems. Soils Found. 47 (4), 685–699.
Kolay, E., Baser, T., 2014. Estimating of the dry unit weight of compacted soils using general linear model and multi-layer perceptron neural networks. Appl. Soft Comput. 18, 223–231.
Kung, G.T.C., Juang, C.H., Hsiao, E.C.L., Hashash, Y.M.A., 2007. Simplified model for wall deflection and ground-surface settlement caused by braced excavation in clays. J. Geotech. Geoenviron. Eng. 133, 731–747.
Lan, P., Su, J.J., Zhang, S., 2024. Surrogate modeling for unsaturated infiltration via the physics and equality-constrained artificial neural networks. J. Rock Mech. Geotech. 16 (6), 2282–2295.
Li, M.G., Xiao, Q.Z., Liu, N.W., Chen, J.J., 2024. Predicting wall deflections for deep excavations with servo struts in soft clay. J. Geotech. Geoenviron. Eng. 150 (1), 04023124.
Li, P., Liu, S.W., Ji, J., Ding, X.M., Bao, M.D., 2024. Stochastic analysis of excavation-induced wall deflection and box culvert settlement considering spatial variability of soil stiffness. J. Rock Mech. Geotech. 15 (12), 3256–3270.
Lim, A., Ou, C.Y., 2018. Performance and three-dimensional analyses of a wide excavation in soft soil with strut-free retaining system. Int. J. GeoMech. 18 (9), 05018007.
Lim, A., Ou, C.Y., Hsieh, P.G., 2018. Investigation of the integrated retaining system to limit deformations induced by deep excavation. Acta Geotech. 13 (4), 973–995.
Liu, Y., Zhao, Y.Z., Zhang, D., Liu, Z.Y., 2022. The long-term mechanical performance of geogrid-reinforced soil retaining walls under cyclic footing loading. Case Stud. Constr. Mater. 17, e01642.
Maehara, K., Hamanaka, A., Sasaoka, T., Shimada, H., Sakuma, S., 2020. Study on control of wall deflection in earth stepped-twin retaining wall using anchor method by means of numerical simulation. Adv. Civ. Eng., 2710954
Meng, Z., Qian, Q.C., Xu, M.Q., et al., 2023. PINN-FORM: a new physics-informed neural network for reliability analysis with partial differential equation. Comput. Methods Appl. Mech. Eng. 414, 116172.
Orazalin, Z., Whittle, A., Olsen, M., 2015. Three-dimensional analyses of excavation support system for the Stata Center basement on the MIT campus. J. Geotech. Geoenviron. Eng. 141 (7). https://doi.org/10.1061/(ASCE)GT.1943-5606.0001326.
Peng, L.Y., He, L., Zhang, Y., et al., 2023. Planning urban underground space from urban emergency evacuation: a digital layout planning method. Tunn. Undergr. Space Technol. 140, 105271.
Phoon, K.K., Zhang, W.G., 2022. Future of machine learning in geotechnics. Georisk 17 (1), 7–22.
Raissi, M., Perdikaris, P., Karniadakis, G.E., 2019. Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. J. Comput. Phys. 378, 686–707.
Sert, S., Luo, Z., Xiao, J.H., Gong, W.P., Juang, C.H., 2016. Probabilistic analysis of responses of cantilever wall-supported excavations in sands considering vertical spatial variability. Comput. Geotech. 75, 182–191.
Sun, Y.B., Sengupta, U., Juniper, M., 2023. Physics-informed deep learning for simultaneous surrogate modeling and PDE-constrained optimization of an airfoil geometry. Comput. Methods Appl. Mech. Eng. 411, 116042.
Tran, D., Nguyen, H., Wang, Y.R., et al., 2023. Analysis of artificial intelligence approaches to predict the wall deflection induced by deep excavation. Open Geosci. 15 (1), 20220503.
Wang, Z.W., Ng, C.W.W., Liu, G.B., 2005. Characteristics of wall deflections and ground surface settlements in Shanghai. Can. Geotech. J. 42, 1243–1254.
Wu, C.Z., Hong, L., Wang, L., et al., 2023. Prediction of wall deflection induced by braced excavation in spatially variable soils via convolutional neural network. Gondwana Res. 123, 184–197.
Wu, X.X., Zhang, Y., Mao, S., 2025. Learning the physics-consistent material behavior from measurable data via PDE-constrained optimization. Comput. Methods Appl. Mech. Eng. 437, 117748.
Yeh, T.Y., Ou, C.Y., Lim, A., 2022. A case study of strut-free excavation retaining system. Acta Geotech. 17 (12), 5557–5571.
Zdravkovic, L., Potts, D.M., St John, H.D., 2005. Modelling of a 3D excavation in finite element analysis. Geotechnique 55, 497–513.
Zeng, C.F., Powrie, W., Xue, X.L., Li, M.K., Mei, G.X., 2021. Effectiveness of a buttress wall in reducing retaining wall movement during dewatering before bulk excavation. Acta Geotech. 16 (10), 3253–3267.
Zhang, W.G., Goh, A.T.C., Xuan, F., 2015. A simple prediction model for wall deflection caused by braced excavation in clays. Comput. Geotech. 63, 67–72.
Zhang, W.G., Hou, Z.J., Goh, A.T.C., Zhang, R.H., 2018a. Estimation of strut forces for braced excavation in granular soils from numerical analysis and case histories. Comput. Geotech. 106, 286–295.
Zhang, R.H., Zhang, W.G., Goh, A.T.C., Hou, Z.J., Wang, W., 2018b. A simple model for ground surface settlement induced by braced excavation subjected to a significant groundwater drawdown. Geomech. Eng. 16 (6), 635–642.
Zhang, R.H., Goh, A.T.C., Zhang, W.G., 2019a. System reliability assessment on deep braced excavation adjacent to an existing upper slope in mountainous terrain: a case study. SN Appl. Sci. 1 (8), 867.
Zhang, W.G., Zhang, R.H., Wang, W., Zhang, F., Goh, A.T.C., 2019b. A Multivariate Adaptive Regression Splines model for determining horizontal wall deflection envelope for braced excavations in clays. Tunn. Undergr. Space Technol. 84, 461–471.
Zhao, D., Yuan, J.Y., Yuan, Y., 2012. Application of *p-y* curves in excavation engineering. Chin. J. Geotech. Eng. 34, 222–224 (in Chinese).
Zhao, H.J., Liu, W., Shi, P.X., et al., 2021. Spatiotemporal deep learning approach on estimation of diaphragm wall deformation induced by excavation. Acta Geotech. 16, 3631–3645.
Zhao, X., Zhu, W.D., Li, H.Y., Li, M., Li, X.Y., 2022. Review, classification, and extension of classical soil-structure interaction models based on different superstructures and soils. Thin-Walled Struct. 173, 108936.
Zheng, G., Zeng, C.F., Diao, Y., Xue, X.L., 2014. Test and numerical research on wall deflections induced by pre-excavation dewatering. Comput. Geotech. 62, 244–256.
Zheng, G., He, X.P., Zhou, H.Z., Diao, Y., Li, Z.W., Liu, X.M., 2022. Performance of inclined-vertical framed retaining wall for excavation in clay. Tunn. Undergr. Space Technol. 130, 104767.
Zhou, C., Su, Y.Q., He, L., et al., 2024. An intelligent resilience evaluation model for the development of urban underground space with safety concern of surrounding existing built environment. Tunn. Undergr. Space Technol. 149, 105783.
Zhou, C., He, L., He, J.C., Zhang, Y., Soh, C.K., 2025. Suitability evaluation of underground space development based on Bayesian network model: a case study considering population evacuation, excavation safety, and spatial connectivity. Expert Syst. Appl. 291, 128399.
Zhu, Y., Wei, S.H., 2010. Research on interaction between deep excavation supporting pile and soil. Rock Soil Mech. 31 (9), 2840–2844 (in Chinese).
Zhu, Y.H., Zabaras, N., Koutsourelakis, P.S., Perdikaris, P., 2019. Physics-constrained deep learning for high-dimensional surrogate modeling and uncertainty quantification without labeled data. J. Comput. Phys. 394, 56–81.

Lei He obtained his BEng degree in Naval Architecture and Marine Engineering from Tianjin University in 2004 and his PhD in Civil Engineering from Nanyang Technological University, Singapore, in 2011. In 2013, he was a visiting scholar at École Polytechnique Fédérale de Lausanne (EPFL), Switzerland. He is currently a Principal Young Professor at Southeast University. His research focuses on innovative planning, design, and construction methods for urban underground space, with a commitment to advancing the informatization and intelligent integration of underground infrastructure systems.