Mesopotamian Journal of Civil Engineering
Vol.2026, **pp**. 139–173
DOI: https://doi.org/10.58496/MJCE/2026/012; ISSN: 3006-1148
https://mesopotamian.press/journals/index.php/MJCE

Research Article

Optimising Physics-Informed Neural Networks for Laterally Loaded Piles by Characteristic-Length Normalisation

Sabah Hassan Fartosy[1,*], Mohammed Khachi Hatem[2]

[1] *Department of Water Resources Engineering, Engineering College, Mustansiriyah University, Baghdad, Iraq,*

[2] *Department of Civil Engineering, Engineering College, Mustansiriyah University, Baghdad, Iraq,*

ARTICLE INFO

Article History

Received 06 Sep 2026
Revised 20 Oct 2026
Accepted 13 Nov 2026
Published 23 Nov 2026

Keywords

Physics-informed neural networks
Laterally loaded piles
Subgrade reaction
Characteristic length
Parameter identifiability
Soil-structure interaction

ABSTRACT

Physics-informed neural networks are increasingly applied to soil-structure interaction, but the effect of variable scaling on their accuracy has not been quantified. This study examines a laterally loaded pile in a Winkler medium, a fourth-order boundary-value problem, using a Hermite finite-element reference solver verified against an exact closed-form solution and published nondimensional coefficients. Four scaling formulations are compared over 1,222 forward trainings with ten random seeds per configuration. Scaling depth to the interval from minus one to plus one fails for flexible piles: the median relative error reaches 1.03 at relative slenderness 20.5, against 1.64×10^{-3} when depth is normalised by the characteristic length. The failure is geometric. The mechanically active zone must be at least about 0.3 input units wide and must lie close to the input origin at which the initial network features are centred. Characteristic-length normalisation satisfies both conditions by construction. Its median error stays below 3.8×10^{-3} for slenderness up to 30 and across 13 architecture and training variants, and its largest per-pile median over 24 randomly generated piles is 1.0×10^{-2}. It is not always the most accurate formulation, but it has the smallest worst case. A generalised characteristic length, defined by $k(T)T^4 = EI$, extends the scaling to power-law, layered and nonlinear p–y profiles. Inverse identification of a two-parameter subgrade profile is practically non-identifiable. A linearised sensitivity analysis predicts the pivot depth of the recovered parameters to within 0.06T in four configurations. Bending moments recovered from 4 to 12 noisy readings have a median error of 5–7 per cent when the governing equation is imposed. A conventional finite-element back-analysis gives similar accuracy at about one-thirtieth of the cost.

1. INTRODUCTION

The laterally loaded pile is a recurring design problem in foundation engineering. Horizontal load is transferred to the ground through deep foundations in wind turbines, bridge piers, quay walls and transmission towers. The model used in routine practice remains the beam on a Winkler medium, in which soil resistance is represented by distributed springs. Closed-form solutions for a uniform medium were given by Hetényi [1]. These were extended by Matlock and Reese [2] to a modulus increasing linearly with depth, and nondimensional coefficients that remain in use were published. Ultimate-capacity rules were provided by Broms [3, 4]. The springs were later replaced by a continuum. Boundary integrals were used by Poulos [5], nonhomogeneous media were treated by Banerjee and Davies [6], and closed-form expressions for flexible piles were derived by Randolph [7]. Elastic continuum solutions were also obtained by Sun [8], and a load-transfer formulation was developed by Guo and Lee [9]. Variational solutions for layered soils and rectangular sections were obtained by Basu and Salgado [10, 11] and by Basu et al. [12], and these were extended to properties varying with depth by Choi et al. [13]. The nonlinear response of rock-socketed shafts was analysed by Zhang et al. [14].

Machine learning entered geotechnical engineering through data-driven correlation. Nonlinear geotechnical behaviour was modelled with neural networks by Goh [15]. Failure modes around underground openings were classified by Lee and Sterling [16]. Foundation settlement was predicted by Sivakugan et al. [17], and that prediction was refined for shallow foundations on granular soils by Shahin et al. [18]. Networks were embedded inside a constitutive model by Ghaboussi and Sidarta [19] rather than being used for end-to-end regression. Liquefaction was classified from cone penetration data by Juang and Chen [20], and several algorithms were compared on the same interpretation task by Rauter and Tschuchnigg [21]. The undrained

*Corresponding author. Email: dr.sabah77@uomustansiriyah.edu.iq

lateral capacity of piles in clay was predicted by Das and Basudhar [22], and slope stability was treated by Kostić et al. [23]. A consistent limitation is reported in reviews of this literature. Early applications were surveyed by Shahin et al. [24] and the state of the art was documented later by the same authors [25]. Several hundred studies were synthesised by Yaghoubi et al. [26]. Three deficiencies recur. Large training sets are required by purely data-driven models. Generalisation outside the range of the training data is poor. No guarantee is provided that predictions satisfy equilibrium or compatibility. The third deficiency matters most in design, because admissible internal forces need not be produced by a model that fits measured deflections.

That limitation is addressed directly by physics-informed neural networks. Representation of a solution is guaranteed by the universal approximation property of feedforward networks [27]. Networks were first trained to satisfy differential equations by Lagaris et al. [28]. The approach was formalised by Raissi et al. [29], in which the equation residual is added to the loss, and the resulting field was surveyed by Karniadakis et al. [30]. The variants now in use were catalogued by Cuomo et al. [31] and by Faroughi et al. [32]. An energy functional is minimised instead of a residual in related formulations [33, 34], the domain is decomposed in others [35, 36], and solution operators rather than single solutions are learned in a further class [37]. The first demonstrations were provided in solid and fluid mechanics [38–43]. Reviews of the field are given by Cai et al. [44], Bai et al. [45] and Brunton and Kutz [46].

Training of these models is not automatic. It was shown by Wang et al. [47] that the residual and boundary terms of the loss evolve at different rates and that convergence can be stalled by that imbalance. Gradient-based weighting of the loss terms [48] and self-adaptive weights [49] were proposed as remedies, and collocation points were redistributed towards large residuals [50, 51]. Failure modes that are not repaired by more training were catalogued by Krishnapriyan et al. [52]. A separate line of work showed that networks fit smooth, low-frequency components first; this spectral bias was documented by Rahaman et al. [53] and by Xu et al. [54], and it was countered by rescaled input features [55, 56]. It was reported by Mandl et al. [57] that training for one-dimensional consolidation was accelerated by an affine transformation of the variables. The choice of variables is therefore not a cosmetic preprocessing step. That dependence is made explicit for a fourth-order problem in the present study. Applications in geomechanics followed quickly. One-dimensional consolidation was solved by Bekele [58]. An elasto-plastic surrogate was built by Eghbalian et al. [59]. Multi-dimensional consolidation was treated by Guo and Yin [60], and unsaturated expansive soils were modelled by Li et al. [61], for which the broader constitutive context is reviewed by Li et al. [62]. A general geoengineering solver was assembled by Chen et al. [63], settlement of an immersed tunnel was analysed by He et al. [64], and seismic site response was assessed by Liu et al. [65]. For piles specifically, axisymmetric and plane-strain pile–soil interaction was solved and inverted by Vahab et al. [66], who used a domain-decomposition network to handle abrupt changes in material properties between layers. A physics-informed framework for lateral soil–structure interaction was developed by Ouyang et al. [67], in which both surrogate construction and back analysis were demonstrated, and it was extended to large deflections with non-differentiable resistance by Ouyang et al. [68].

Two questions remain open. The first concerns scaling. Training is sensitive to the magnitude of inputs, outputs and loss terms. The pile equation is fourth order, so the leading term is rescaled by the fourth power of any change of depth variable. Flexural rigidity is of order 10^7 to 10^9 N·m² while deflection is of order 10^{-2} m, and the residual and boundary terms therefore differ by many orders of magnitude. Inputs are normalised to a fixed interval, usually $[-1, 1]$, in published studies, and success is reported. Whether that choice remains adequate as the pile becomes more flexible has not been tested, and no geometric criterion is available to predict when it fails. The second question concerns inversion. Back analysis of soil stiffness from measured deflection is presented as a capability of the method, but the limits of what the data can determine have not been established. Soil is mobilised by a pile only over its upper length. Stiffness below that depth cannot influence the measurements, so it cannot be recovered. The present study concerns the lateral response of a single pile and the identification of lateral subgrade stiffness. It does not address axial bearing capacity, for which data-driven correlations are the subject of a separate literature [22, 26].

Both questions are answered quantitatively in this study. A Hermite finite-element solver is verified against an exact closed-form solution and against published nondimensional coefficients, and it is used as ground truth. Four scaling formulations of the identical boundary-value problem are trained under an equal budget, with ten random seeds per configuration. The comparison is extended by a controlled sweep in which only the pile length is varied, by an ablation over network architecture, activation, collocation density and training budget, by adaptive weighting and adaptive sampling, and by 24 randomly generated piles that include steel and concrete sections, head moments and non-linear stiffness profiles. Inverse identification of a two-parameter subgrade profile is performed from sparse, noisy deflection readings in four configurations, under Gaussian, heavy-tailed and systematic measurement errors, and with data generated by an independent non-linear p–y model. The contributions are as follows. First, the conventional practice of scaling depth to a fixed interval is shown to fail for flexible piles, and the failure is explained by two geometric conditions on the width and location of the active zone in network input units. Second, normalisation by a generalised characteristic length is shown to satisfy both conditions by construction and to give the smallest worst-case error across all tests. Third, the subgrade profile is shown to be practically non-identifiable, and the pivot of the recovered parameters is predicted by a linearised sensitivity analysis. Fourth, the

recovery of bending moments from inclinometer-type data is compared with regularised differentiation and with a conventional finite-element back-analysis.

2. PROBLEM STATEMENT AND REFERENCE SOLUTION

2.1 Governing equation

A pile of flexural rigidity EI and embedded length L is loaded at its head by a lateral force H and a moment M_0. Soil resistance is modelled as a distributed reaction proportional to the local deflection. Equilibrium of a Euler–Bernoulli beam then gives Equation (1).

$$EI\frac{d^4y}{dz^4} + k(z)\,y = 0 \tag{1}$$

Here y is the lateral deflection, z is the depth measured from the ground surface, and k(z) is the modulus of subgrade reaction, which has dimensions of force per unit length per unit deflection. The profile is written as the power law of Equation (2), in which k_{ref} is the value at z = L.

$$k(z) = k_{ref}\left(\frac{z}{L}\right)^n \tag{2}$$

The two classical cases are covered by the exponent n. The uniform modulus appropriate to overconsolidated clay, treated analytically by Biot [69] and Hetényi [1], is recovered by setting n = 0. The linear increase appropriate to sand, as assumed by Matlock and Reese [2] and by Reese et al. [70], is recovered by setting n = 1. Layered or normally consolidated deposits are represented by intermediate values, for which layered solutions were given by Davisson and Gill [71]. Bending moment and shear follow as M = EI y″ and V = EI y‴.

The modulus of subgrade reaction is not a soil constant. It was shown by Terzaghi [72] to depend on the loaded width and on the stress level, and values were recommended by soil type rather than obtained by measurement. Inverse identification from a load test is attractive for that reason, and its limits must be established for the same reason. The same framework is adopted with nonlinear resistance curves in offshore design codes [73]. The conditions in Equation (3) are given by a free head loaded by H and M_0. Those in Equation (4) are given by a free toe.

$$EI\,y''(0) = -M_0\,,\quad EI\,y'''(0) = H \tag{3}$$

$$EI\,y''(L) = 0\,,\quad EI\,y'''(L) = 0 \tag{4}$$

2.2 Characteristic length

The relative stiffness of pile and soil is expressed by a characteristic length T. For an arbitrary stiffness profile, T is defined here as the depth at which the local subgrade stiffness and the pile rigidity are in balance, as given in Equation (5).

$$k(T)\,T^4 = EI \tag{5}$$

The function $k(z)z^4$ increases monotonically with depth for any non-decreasing profile, so Equation (5) has a single root. For a profile that decreases with depth the smallest root is taken. For the power law of Equation (2) the root is explicit and is given by Equation (6).

$$T_n = \left(\frac{EI\,L^n}{k_{ref}}\right)^{\frac{1}{4+n}} \tag{6}$$

Two classical lengths are recovered. For n = 1, k_{ref}/L equals the modulus gradient n_h, and Equation (6) reduces to $T = (EI/n_h)^{1/5}$ of Matlock and Reese [2]. For n = 0, Equation (6) reduces to the relative stiffness $(EI/k)^{1/4}$ of a uniform medium [1]. The original submission used $T = (EI/n_h)^{1/5}$ with $n_h = k_{ref}/L$ for every exponent. That length is exact only for n = 1. Its consequence for $n \neq 1$ is derived in Section 3.3. Where no subscript is given below, T denotes T_n; for the linear profiles of the forward study the two definitions coincide. The dimensionless group that controls the response is the relative slenderness of Equation (7).

$$\lambda = \frac{L}{T} \tag{7}$$

Piles with λ below about 2 behave as rigid bodies, and rotational restraint at the head governs their response [74]. Piles with λ above about 5 behave as flexible members, and the head response is not changed by further length. This distinction is standard in design [75, 76]. It is central to the present results because the geometry seen by a neural network is also controlled by λ. The fixed-point definition of Equation (5) also applies to a layered profile and to the secant stiffness of a nonlinear p–y curve. Both uses are tested in Section 4.7.

2.3 Reference solver

Equation (1) is discretised with two-node cubic Hermite Euler–Bernoulli beam elements. The Winkler term is integrated with six-point Gauss–Legendre quadrature within each element. The Hermite basis is the exact homogeneous solution of EI $y'''' = 0$, so rapid convergence is obtained. Loads are applied as consistent nodal terms. The banded system is solved directly. For the nonlinear p–y curves of Section 4.7, the internal force is assembled from the p–y law at the quadrature points and a full Newton–Raphson iteration with the consistent tangent is used, until the maximum nodal increment falls below 10^{-11} of the maximum nodal value. All reference results are computed with 2000 elements. A comparable element formulation for pile analysis was presented by Ouyang et al. [77]. The verification outcomes are listed in Table 1.

TABLE I. VERIFICATION OF THE HERMITE FINITE-ELEMENT REFERENCE SOLVER AGAINST AN EXACT CLOSED-FORM SOLUTION AND AGAINST PUBLISHED NONDIMENSIONAL COEFFICIENTS.

Benchmark quantity	Present solver	Reference value	Type of reference	Deviation
Head deflection, uniform medium (m)	4.9338851×10^{-3}	4.9338855×10^{-3}	Exact closed form [1]	7.9×10^{-6} %
Head rotation, uniform medium (rad)	$-1.8257417 \times 10^{-3}$	$-1.8257419 \times 10^{-3}$	Exact closed form [1]	6.2×10^{-6} %
Deflection profile, upper 0.6L (relative L_2 error)	—	—	Exact closed form [1]	8.8×10^{-8}
Deflection coefficient at surface, $A_y(0)$	2.4292	2.435	Published chart value [2]	0.24 %
Moment-case deflection coefficient at surface, $B_y(0)$	1.6194	1.623	Published chart value [2]	0.22 %
Peak moment coefficient, A_m	0.7718 at Z = 1.33	0.772 at Z ≈ 1.3	Published chart value [2]	0.03 %
Moment-case coefficient, $B_m(0)$	1.0000	1.000	Published chart value [2]	< 0.01 %

Two different kinds of reference are used. The first is exact. It is the closed-form solution of Hetényi [1] for a finite beam of length 25 m on a uniform medium with a free loaded end, with EI = 400 MN·m², k = 30 MN/m² and H = 200 kN; the exact values in Table 1 were evaluated at 30 significant digits. The computed head deflection deviates from the closed form by 7.9×10^{-6} per cent, and the deflection profile agrees to a relative error of 8.8×10^{-8} over the upper 60 per cent of the beam. The problem is defined in Fig. 1 and the comparison is shown in Fig. 2.

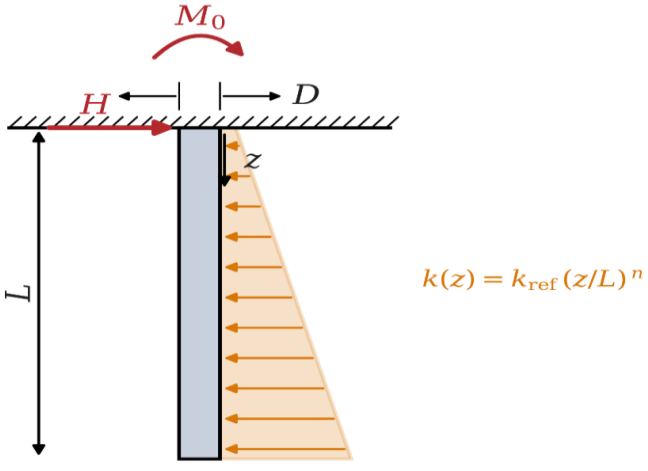

Fig. 1. Definition of the laterally loaded pile problem, showing the head load H, the head moment M_0, the embedded length L, the diameter D, the depth coordinate z and the subgrade reaction profile k(z).

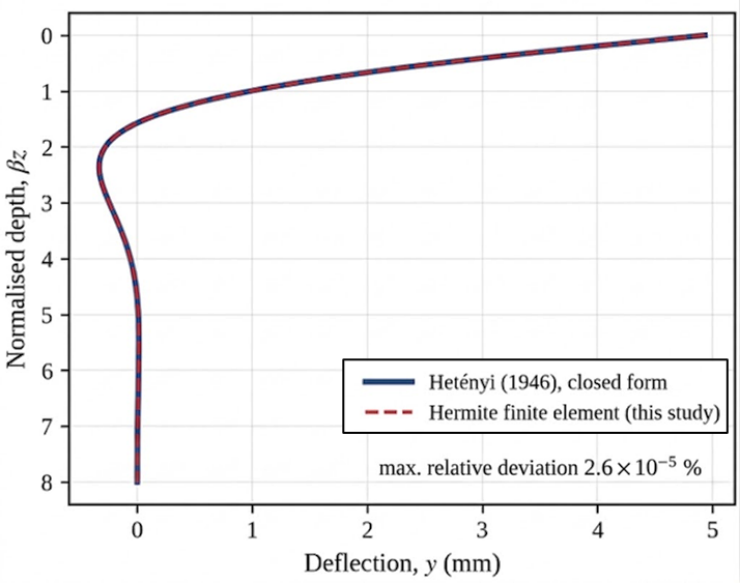

Fig. 2. Verification of the Hermite finite-element reference solver against the closed-form solution of Hetényi [1] for a uniform Winkler medium.

The second reference is not exact. It is the set of nondimensional coefficients published by Matlock and Reese [2] for a modulus increasing linearly with depth, read from tables and charts of limited precision. The computed deflection coefficient at the ground surface is 2.4292 against the published 2.435. The computed deflection coefficient for the applied-moment case at the same location is 1.6194 against the published 1.623. The computed peak moment coefficient is 0.7718 at a depth coefficient of 1.33 against the published 0.772 at approximately 1.3. These comparisons are shown in Fig. 3 and Fig. 4. The residual differences are consistent with the graphical precision of the original charts. This comparison checks the solver for a linearly increasing modulus; it is not an independent exact benchmark.

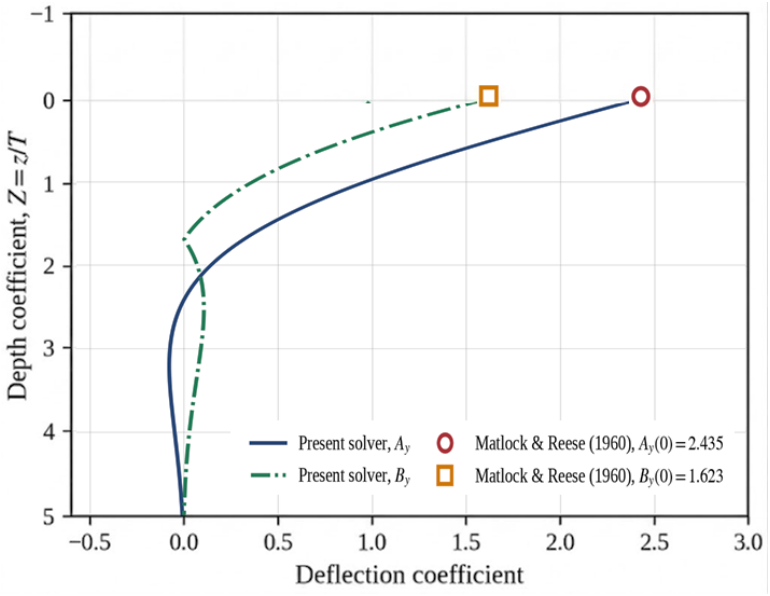

Fig. 3. Deflection coefficients A_y for the applied-load case and B_y for the applied-moment case from the present solver, compared with the ground-surface values published by Matlock and Reese [2].

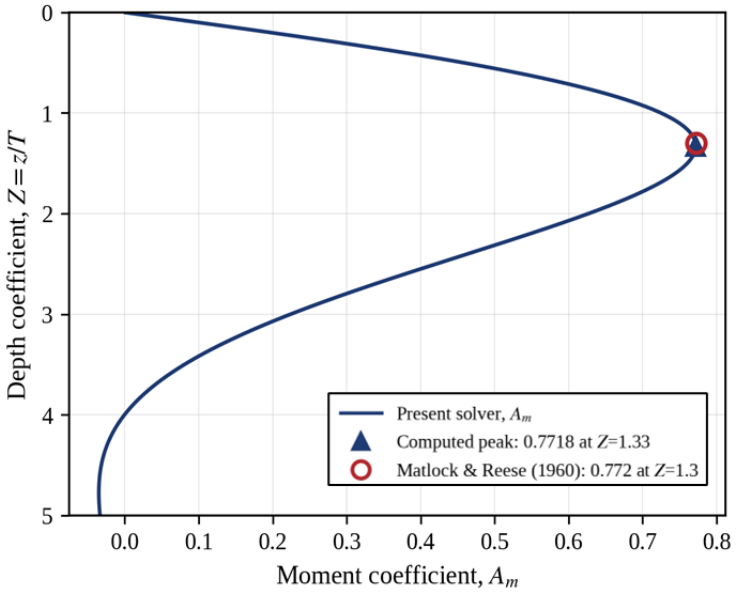

Fig. 4. Moment coefficient A_m from the present solver, compared with the peak value published by Matlock and Reese [2].

3. NEURAL FORMULATIONS

3.1 Network and derivatives

A fully connected network maps a scalar coordinate to a scalar response. Three hidden layers of 32 neurons with hyperbolic tangent activation form the baseline. Other depths, widths and activations are examined in Section 4.6. Weights are initialised from a normal distribution with the variance of Glorot and Bengio [78], and biases are initialised to zero. The fourth derivative of the network output is required by the residual of Equation (1). Four nested reverse-mode differentiations are expensive, and a deep graph is built by them. Derivatives are instead propagated analytically through each layer as a five-component jet (a, a′, a″, a‴, a⁗). The jet is transformed component-wise by a linear layer, and the bias enters only the zeroth component. For an activation f the propagation follows the Faà di Bruno formula, given to fourth order in Equation (8), in which f_k denotes the k-th derivative of f evaluated at a.

$$t' = f_1 a', \quad t'' = f_2 a'^2 + f_1 a'', \quad t''' = f_3 a'^3 + 3 f_2 a' a'' + f_1 a''', \quad t'''' = f_4 a'^4 + 6 f_3 a'^2 a'' + f_2 (3 a''^2 + 4 a' a''') + f_1 a'''' \tag{8}$$

For t = tanh(a) with g = 1 − t², the derivatives are $f_1 = g$, $f_2 = -2tg$, $f_3 = g(6t^2 - 2)$ and $f_4 = 8tg(2 - 3t^2)$. The scheme is exact. Its maximum relative difference from four nested automatic differentiations is 2.4×10^{-15} for tanh, 1.2×10^{-15} for the sine and 6.5×10^{-14} for the SiLU activation. The background to automatic differentiation is given by Baydin et al. [79]. All models are implemented in JAX 0.11.2 [80] with the Optax 0.2.8 optimisers [81], in double precision on a single processor core. The first submission used an equivalent PyTorch implementation [82]. The code was re-implemented to run the enlarged study reported here, and every neural-network result in this revision was regenerated with the new code.

3.2 Loss and optimiser

The loss is the sum of a mean-square residual over collocation points and a mean-square boundary term, as given in Equation (9), in which N_c is the number of collocation points and b_j are the four boundary residuals of Equation (3) and Equation (4).

$$\mathcal{L} = \left(\frac{1}{N_c}\right) \Sigma\, r(\xi_i)^2 + \left(\frac{1}{4}\right) \Sigma\, b_j{}^2 \tag{9}$$

Training is performed with 1500 iterations of Adam [83] at a learning rate of 3×10^{-3}, followed by up to 400 iterations of L-BFGS [84] with a memory of 100 and a zoom line search that enforces the strong Wolfe conditions [85]. L-BFGS is stopped earlier if the gradient norm falls below 10^{-14} or the loss stops changing. Collocation is performed at 121 uniformly spaced points. This budget is held fixed for every formulation and every case, so differences in accuracy are attributable to the formulation alone. Accuracy is measured by the relative L_2 error of Equation (10), evaluated at 401 equally spaced depths against the finite-element reference, separately for deflection and for bending moment.

$$e_y = \frac{\|\hat{y} - y\|_2}{\|y\|_2}, \qquad e_M = \frac{\|\hat{M} - M\|_2}{\|M\|_2} \tag{10}$$

3.3 The four scaling formulations

Four treatments of the identical boundary-value problem are compared. They differ only in the change of variables applied before the problem is seen by the network and in the units of the loss terms. They are summarised in Table 2 and illustrated in Fig. 5.

TABLE II. THE FOUR SCALING FORMULATIONS OF THE IDENTICAL BOUNDARY-VALUE PROBLEM.

Formulation	Coordinate	Response	Loss units	Input span
F1 raw	$\xi = z$	$u = y$	SI	[0, L]
F2 input-scaled	$\xi = 2z/L - 1$	$u = y$	SI	[−1, 1]
F3 fully scaled, length L	$\xi = 2z/L - 1$	$u = y/y_c$	dimensionless	[−1, 1]
F4 proposed, length T	$\zeta = z/T$	$u = y/y_c$	dimensionless	[0, λ]

Raw SI variables are used in formulation F1. Depth is in metres, deflection is in metres, and the loss is in newtons per metre and newton metres. Depth is scaled to the interval [−1, 1] in formulation F2 and everything else is left unchanged. This is the most common preprocessing choice in the applied literature. The response is additionally scaled by $y_c = H\,T^3/EI$ in formulation F3, the residual is divided by H/T, the shear conditions by H and the moment conditions by HT, but the geometric length L is retained as the coordinate scale. F3 and F4 therefore have identical loss definitions in identical physical units. They differ only in the coordinate. The mechanical characteristic length T is used as the coordinate scale in formulation F4, which is proposed here. The reduction of Equation (1) to dimensionless form is set out step by step below, because the parameter-free form is what makes the scaling transferable.

The dimensionless depth and deflection are defined in Equation (11). The chain rule then converts every derivative, as in Equation (12).

$$\zeta = \frac{z}{T}, \qquad u = \frac{y}{y_c}, \qquad y_c = \frac{H\,T^3}{EI} \tag{11}$$

$$\frac{d^m y}{dz^m} = \frac{y_c}{T^m}\frac{d^m u}{d\zeta^m}, \qquad m = 0,1,\dots,4 \tag{12}$$

Substitution of Equation (12) with m = 4 into Equation (1) and division by $EI\,y_c/T^4$ gives Equation (13). The factor $EI\,y_c/T^4$ equals H/T, which is the natural scale of the residual used in F3 and F4.

$$\frac{d^4 u}{d\zeta^4} + \tilde{k}(\zeta)\,u = 0, \qquad \tilde{k}(\zeta) = \frac{k(T\zeta)\,T^4}{EI} \tag{13}$$

For the power law of Equation (2), the dimensionless stiffness depends on which length is used, as Equation (14) shows.

$$\tilde{k}(\zeta) = \frac{k_{ref}\,T^{4+n}}{EI\,L^n}\,\zeta^n = \begin{cases} \zeta^n, & T = T_n \\ \lambda_1^{1-n}\,\zeta^n, & T = (EI/n_h)^{1/5} \end{cases} \tag{14}$$

With $T = T_n$ the stiffness term is ζ^n for every pile, and $\lambda = L/T_n$ and n are the only remaining parameters. With the Matlock–Reese length an extra factor λ_1^{1-n} appears, so the equation is parameter-free only for n = 1. This addresses the theoretical point raised on the original definition. The boundary conditions of Equation (3) and Equation (4) are divided by H and HT, which gives Equation (15).

$$u'''(0) = 1, \qquad u''(0) = -\frac{M_0}{HT}, \qquad u''(\lambda) = u'''(\lambda) = 0 \tag{15}$$

Every term in Equation (13) and Equation (15) is of order unity. The network input spans [0, λ] rather than a fixed interval. This distinction is the subject of Section 4.

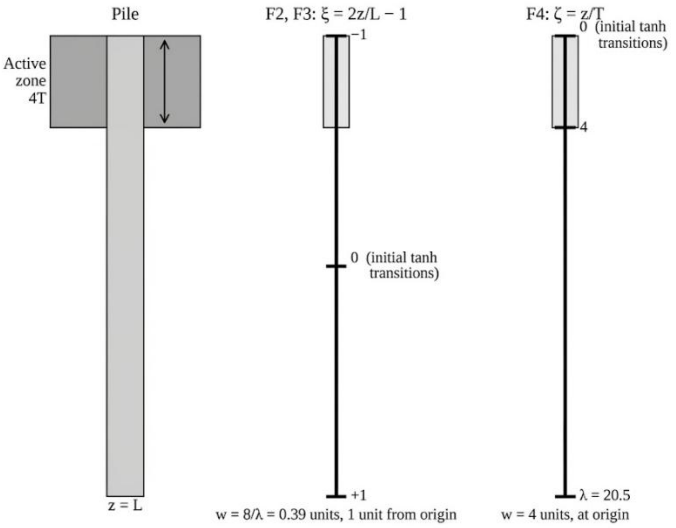

Fig. 5. Mapping of pile depth to network input for case C5 (λ = 20.5) under fixed-interval scaling (F2, F3) and characteristic-length normalisation (F4). The shaded band is the mechanically active zone, $0 \le z \le 4T$. With zero initial biases, all first-layer tanh transitions are centred at input 0. Under F2 and F3 the active zone is 0.39 input units wide and lies one unit from that origin; under F4 it is 4 units wide and starts at the origin.

3.4 Implementation and cost

One forward training takes a median of 1.77 s on a single processor core at the budget defined above, after a one-off compilation of about 15 s. The study reported here comprises 1,222 forward trainings and 360 inverse or data-only trainings. Section 7 compares this cost with that of the finite-element solver and of a conventional finite-element back-analysis. Speed on the forward problem is not the advantage of the neural formulation, and the comparison in Section 7 shows that it is not an advantage for this one-dimensional inverse problem either.

3.5 Statistical treatment and measurement-error models

Every forward configuration is trained with ten random seeds in Section 4.2 and with three to five seeds in the sensitivity studies. The seed changes the network initialisation only. Results are reported as the median and the interquartile range. Uncertainty in a median is expressed by a bootstrap 95 per cent confidence interval from 4000 resamples [86]. Differences between formulations are tested with the two-sided Mann–Whitney U test [87], which does not assume normality. With ten runs per group, the smallest attainable p value is 1.8×10^{-4}.

Synthetic readings for the inverse analyses are generated by Equation (16), in which y is the reference deflection, z_i (i = 1, …, N_s) are equally spaced depths between the head and the toe, y_{max} is the maximum absolute deflection and ε_i are independent random variables of zero mean and unit variance.

$$\hat{y}_i = y(z_i + \Delta z) + \eta\, y_{max}\, \varepsilon_i + \beta\, y_{max}\, \frac{L - z_i}{L} \qquad (16)$$

Four error models are obtained from Equation (16). The baseline model has Gaussian ε_i with $\Delta z = \beta = 0$ and a noise level η of 0, 2, 5 or 10 per cent. The heavy-tailed model draws ε_i from a Student t distribution with three degrees of freedom, rescaled to unit variance. The bias-drift model adds a linear ramp, β = 2 or 5 per cent, that vanishes at the toe. It represents the accumulated effect of a constant tilt offset when inclinometer readings are summed upwards from an assumed fixed toe [88–90]. The depth-offset model reads the true profile at $z_i + \Delta z$ with Δz = 0.10 m or 0.25 m and assigns the value to z_i. It represents a cable-depth or casing-settlement error. The last three models are combined with η = 2 per cent Gaussian noise. A seeded random generator is used for every realisation, so each dataset is reproducible and the same dataset is given to every estimator compared in Section 6.

4. EFFECT OF SCALING ON ACCURACY

4.1 Test cases

Five piles are analysed. Diameter ranges from 0.35 m to 1.20 m and embedded length from 10 m to 20 m. Young's modulus is 30 GPa throughout, representing concrete. The modulus gradient ranges from 5 MN/m³ to 25 MN/m³, covering loose to dense granular deposits. The profile is linear (n = 1), so T_n and the Matlock–Reese length coincide. The head load is 200 kN and no head moment is applied. The cases and their reference responses are listed in Table 3. Relative slenderness spans

2.77 to 20.50, so both rigid and flexible behaviour are included in the set. These five cases vary diameter, length and stiffness together. The effect of slenderness alone is isolated in Section 4.3, and broader conditions are examined in Section 4.7.

TABLE III. TEST CASES AND THEIR REFERENCE RESPONSES, COMPUTED WITH THE VERIFIED FINITE-ELEMENT SOLVER.

Case	D (m)	L (m)	EI (MN·m²)	n_h (MN/m³)	T (m)	λ = L/T	y_0 (mm)	M_{max} (kN·m)
C1	1.20	10	3053.6	5	3.607	2.77	9.03	483.6
C2	0.90	15	966.2	8	2.609	5.75	8.93	402.5
C3	0.60	12	190.9	10	1.804	6.65	14.94	278.4
C4	0.45	18	60.4	15	1.321	13.62	18.55	203.9
C5	0.35	20	22.1	25	0.976	20.50	20.42	150.6

The reference deflection profiles are plotted against normalised depth in Fig. 6. The five curves collapse onto a single shape. Deflection reverses sign at approximately ζ = 2.4 and is negligible below ζ = 5. The mechanically active zone is confirmed by this collapse to occupy a fixed extent in units of T rather than in units of L.

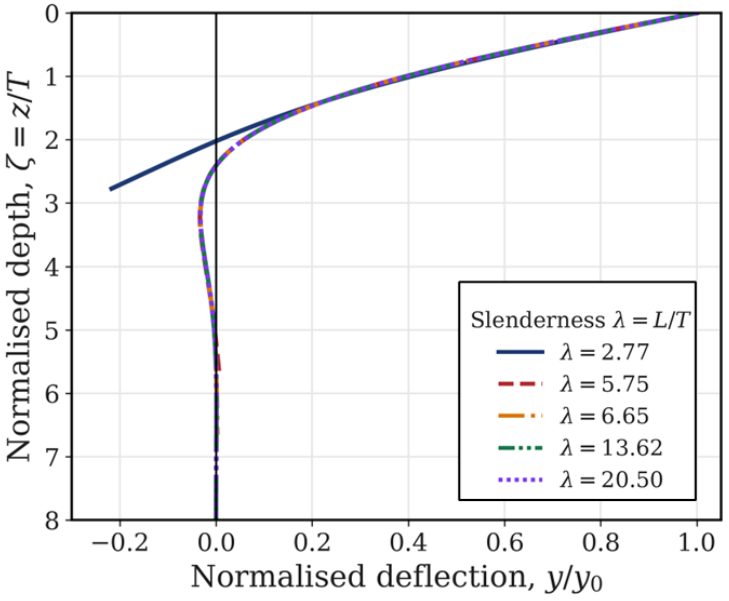

Fig. 6. Reference deflection profiles for the five cases plotted against normalised depth ζ = z/T, showing that the active zone occupies a fixed extent in units of the characteristic length. Deflection is normalised by its head value y_0.

4.2 Accuracy across slenderness

TABLE IV. MEDIAN RELATIVE L_2 ERROR IN DEFLECTION OVER TEN RANDOM SEEDS, WITH THE INTERQUARTILE RANGE IN BRACKETS, AT A FIXED TRAINING BUDGET OF 1500 ADAM AND UP TO 400 L-BFGS ITERATIONS.

Case	λ = L/T	F1 raw	F2 input-scaled	F3 length L	F4 length T
C1	2.77	3.32×10^{-2} [1.49×10^{-2}, 4.91×10^{-2}]	2.78×10^{-3} [1.73×10^{-3}, 3.52×10^{-3}]	2.00×10^{-5} [1.34×10^{-5}, 2.96×10^{-5}]	2.11×10^{-5} [1.48×10^{-5}, 2.73×10^{-5}]
C2	5.75	3.28×10^{-3} [2.46×10^{-3}, 4.05×10^{-3}]	1.42×10^{-2} [1.06×10^{-2}, 2.32×10^{-2}]	1.42×10^{-3} [8.14×10^{-4}, 1.97×10^{-3}]	3.78×10^{-3} [3.49×10^{-3}, 4.27×10^{-3}]
C3	6.65	1.12×10^{-3} [6.11×10^{-4}, 1.39×10^{-3}]	1.73×10^{-2} [5.33×10^{-3}, 2.88×10^{-2}]	1.02×10^{-3} [6.68×10^{-4}, 1.24×10^{-3}]	1.37×10^{-3} [1.15×10^{-3}, 1.43×10^{-3}]

Case	λ = L/T	F1 raw	F2 input-scaled	F3 length L	F4 length T
C4	13.62	1.44×10^{-3} [1.20×10^{-3}, 2.29×10^{-3}]	1.02 [1.00, 1.03]	2.79×10^{-4} [2.04×10^{-4}, 3.97×10^{-4}]	7.15×10^{-4} [5.33×10^{-4}, 8.46×10^{-4}]
C5	20.50	4.38×10^{-3} [3.35×10^{-3}, 5.29×10^{-3}]	1.03 [1.03, 1.03]	0.30 [1.51×10^{-2}, 0.69]	1.64×10^{-3} [1.39×10^{-3}, 2.55×10^{-3}]
Worst median	—	3.32×10^{-2}	1.03	0.30	3.78×10^{-3}

Each formulation is trained ten times per case with different random seeds. The median relative L_2 error in deflection and its interquartile range are reported in Table 4. The same data are plotted in Fig. 7, and the corresponding moment error is plotted in Fig. 8. Pairwise comparisons with F4 are listed in Table 5.

TABLE V. RATIO OF THE MEDIAN DEFLECTION ERROR OF F1, F2 AND F3 TO THAT OF F4, WITH BOOTSTRAP 95 PER CENT CONFIDENCE INTERVAL, AND TWO-SIDED MANN–WHITNEY P VALUE (TEN SEEDS PER GROUP). A RATIO ABOVE 1 MEANS THAT F4 IS MORE ACCURATE.

Case	F1 / F4	F2 / F4	F3 / F4
C1	1572 (585–2772) $p = 1.8 \times 10^{-4}$	132 (68.6–626) $p = 1.8 \times 10^{-4}$	0.947 (0.538–1.76) $p = 0.79$
C2	0.867 (0.564–1.17) $p = 0.43$	3.76 (2.55–9.2) $p = 1.8 \times 10^{-4}$	0.376 (0.203–0.561) $p = 2.2 \times 10^{-3}$
C3	0.821 (0.431–1.22) $p = 0.31$	12.6 (1.95–23.8) $p = 3.3 \times 10^{-4}$	0.747 (0.354–1.06) $p = 0.04$
C4	2.01 (1.29–6.31) $p = 4.6 \times 10^{-3}$	1432 (1154–1949) $p = 1.8 \times 10^{-4}$	0.39 (0.246–0.721) $p = 2.2 \times 10^{-3}$
C5	2.67 (1.49–3.73) $p = 1.3 \times 10^{-3}$	628 (368–771) $p = 1.8 \times 10^{-4}$	183 (7.06–501) $p = 1.8 \times 10^{-4}$

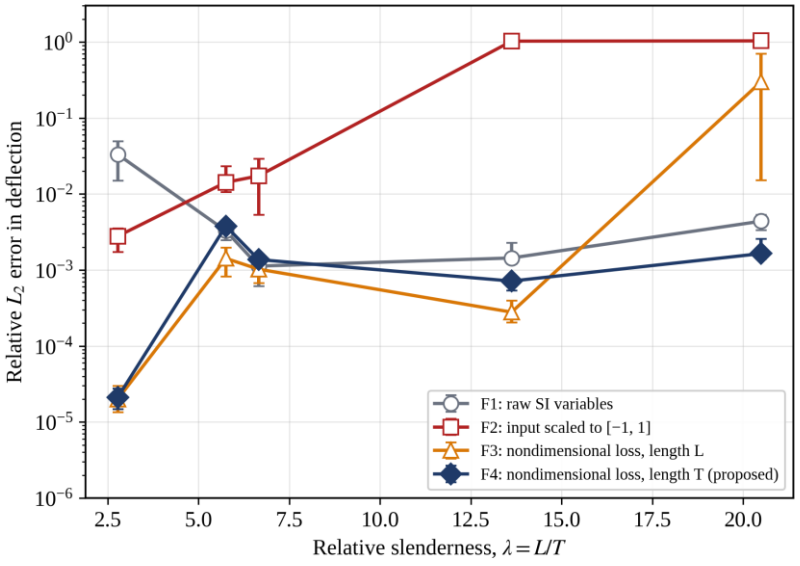

Fig. 7. Median relative L_2 error in deflection for the four scaling formulations across the range of relative slenderness, at a fixed training budget. Bars show the interquartile range over ten seeds.

Four findings follow. First, formulation F2 fails for the two most flexible piles. Its median error is 1.02 at $\lambda = 13.62$ and 1.03 at $\lambda = 20.50$, which means that the predicted deflection bears no useful relation to the reference, and all ten seeds fail in both cases. Its median error is 1430 and 628 times that of F4 at these two slenderness values, with $p = 1.8 \times 10^{-4}$, the smallest value the test can return. Second, formulation F3 fails at the highest slenderness despite a fully nondimensional loss. Its median error at $\lambda = 20.50$ is 0.300, five of its ten seeds exceed 0.1, and its bootstrap interval for the ratio to F4 is 7.1 to 500. The problem is therefore not repaired by loss normalisation alone. The cause lies in the coordinate, because F3 and F4 share the same loss. Third, the median error of formulation F4 stays below 3.8×10^{-3} in every case, and no single F4 run exceeds 4.4×10^{-3}. Fourth, F4 is not the most accurate formulation in every case. F3 is more accurate than F4 at $\lambda = 5.75$, 6.65 and 13.62, by factors of 2.7, 1.3 and 2.6. The advantage of F4 is a bounded worst case, not uniform superiority.

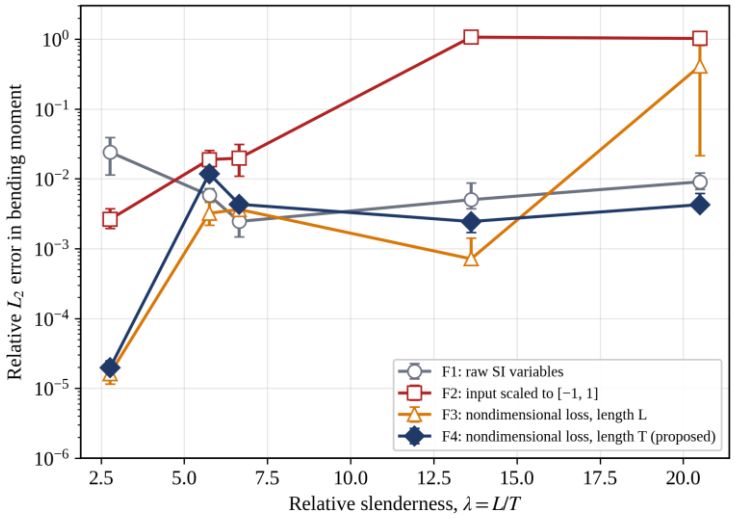

Fig. 8. Median relative L_2 error in bending moment for the four scaling formulations across the range of relative slenderness. Bars show the interquartile range over ten seeds.

The same ordering is obtained for bending moment in Fig. 8. Median moment errors of 1.02 and 0.411 are reached by F2 and F3 at the highest slenderness, whereas the median moment error of F4 remains between 2.0×10^{-5} and 1.2×10^{-2} throughout. Bending moment is the quantity of design interest, so the ordering matters more there than in deflection.

Formulation F1 is the least accurate at the lowest slenderness, where its median error is 3.3×10^{-2}. This is consistent with the mechanism proposed below, because depth is retained in metres by F1 and the active zone is therefore not compressed. Its weakness lies in the balance of the loss. The root-mean-square ratio of residual to boundary terms at initialisation for the most flexible case is listed in Table 6. The median ratio over ten initialisations is 61 for F1 and 285 for F2, against 2.35 for F3 and 9.2 for F4. Convergence is slowed by an unbalanced loss, which explains why F1 is weakest for the stiff pile, where the solution is dominated by the boundary terms.

TABLE VI. ROOT-MEAN-SQUARE RATIO OF RESIDUAL TO BOUNDARY TERMS AT INITIALISATION FOR CASE C5 (MEDIAN AND RANGE OVER TEN INITIALISATIONS), AND THE CORRESPONDING NETWORK INPUT SPAN.

Formulation	Input span	Response scale	Residual / boundary ratio
F1	20.00	1.00	61 (16–807)
F2	2.00	1.00	285 (43–1490)
F3	2.00	8.40×10^{-3}	2.35 (0.35–20.1)
F4	20.50	8.40×10^{-3}	9.2 (1.0–41.9)

4.3 Controlled slenderness sweep

The five cases of Table 3 change diameter, length and soil stiffness together. To isolate slenderness, a second set was built in which $EI = 60.4$ MN·m², $n_h = 15$ MN/m³ and $H = 200$ kN are held fixed and only L is varied, so that λ takes 13 values from 2 to 30. For λ above about 5 the reference solution in ζ is then identical for every pile, and only the length of the inactive tail changes. Each of F2, F3 and F4 was trained with five seeds per value. The results are plotted in Fig. 9.

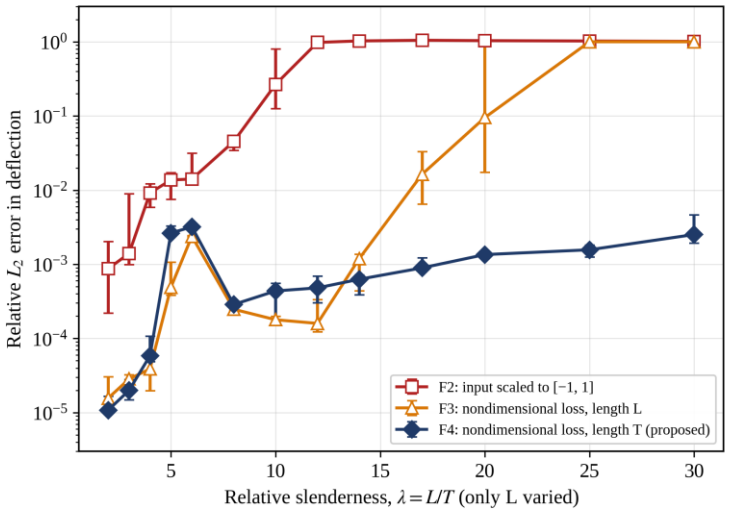

Fig. 9. Controlled slenderness sweep in which only the pile length is varied (EI = 60.4 MN·m², n_h = 15 MN/m³). Median relative L_2 error in deflection over five seeds, with interquartile range.

The sweep confirms that slenderness alone controls the failure. The median error of F2 rises from 8.7×10^{-4} at $\lambda = 2$ to 4.5×10^{-2} at $\lambda = 8$ and 0.27 at $\lambda = 10$, and it is at or above 0.98 for every $\lambda \geq 12$. The median error of F3 stays below 2.4×10^{-3} up to $\lambda = 14$, rises to 1.6×10^{-2} at $\lambda = 17$ and 9.5×10^{-2} at $\lambda = 20$, and reaches 1.0 at $\lambda = 25$ and 30. The median error of F4 stays between 1.1×10^{-5} and 3.2×10^{-3} over the whole range, and it is 2.5×10^{-3} at $\lambda = 30$. Between $\lambda = 8$ and 12, F3 is more accurate than F4, with median errors of 1.6 to 2.5×10^{-4} against 2.9 to 4.8×10^{-4}.

4.4 Mechanism

The failure of F2 and F3 has a geometric explanation. The coordinate maps are affine. An affine input map $x = cz + d$ is absorbed by the first network layer, as Equation (17) shows, so all four formulations can represent exactly the same functions.

$$W(c\,z + d) + b = (cW)z + (dW + b) \tag{17}$$

What differs is the starting point of training. With zero biases and weights of order unity, every first-layer unit $\tanh(W x)$ has its transition centred at $x = 0$ and a width of order one input unit. Networks fit such smooth, low-frequency components first [53, 54]. The initial features are therefore concentrated in a region about one input unit wide around the input origin. Lateral response is confined to the upper 4T, as Fig. 6 shows. The fraction of the embedded length that is mechanically active is given by Equation (18).

$$\phi = min(1, \frac{4T}{L}) = min(1, \frac{4}{\lambda}) \tag{18}$$

That fraction is a property of the pile and not of the formulation. Two quantities depend on the formulation. The first is the width of the active zone measured in input units. It is the product of ϕ and the input span S, as given in Equation (19), with $S = L$ for F1, $S = 2$ for F2 and F3, and $S = \lambda$ for F4.

$$w = \phi\, S \tag{19}$$

The second is the distance of the active zone from the input origin. Under F2 and F3 the pile head is mapped to $x = -1$, one input unit from the origin. Under F1 and F4 the head is mapped to the origin itself. The three cases of Equation (19) evaluate to $w = \min(L, 4T)$, $w = \min(2, 8/\lambda)$ and $w = \min(\lambda, 4)$ respectively. Values of w are plotted against relative slenderness in Fig. 10, and the five cases are marked.

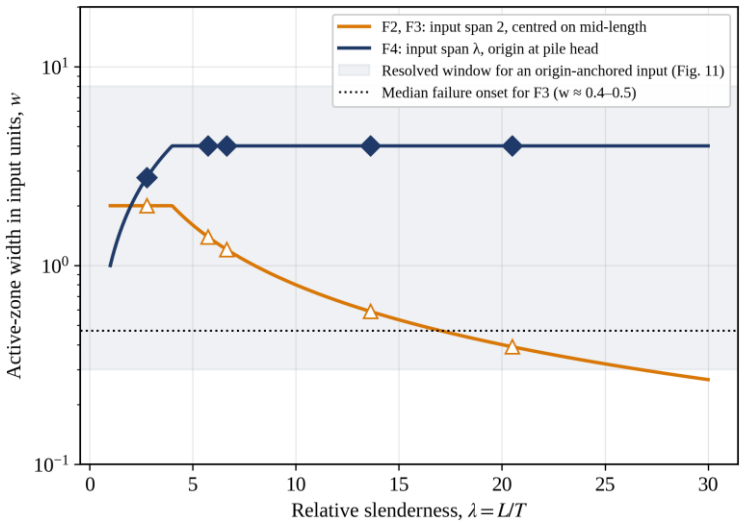

Fig. 10. Width of the mechanically active zone measured in network input units, plotted against relative slenderness. The shaded band is the width range over which an origin-anchored input trained successfully in Fig. 11. The dotted line marks the median failure onset of F3 in Table 7.

The two quantities were tested separately at the most flexible pile, C5. In the first test the loss of F4 was retained and the input was set to x = S z/L, so that the head stays at the origin and only the width w = S min(1, 4/λ) changes. S was varied from 0.5 to 80 with five seeds each. The same test was repeated for C4. The results are plotted in Fig. 11. The error is below 3.4×10^{-3} for every width between 0.29 and 7.8 units at C5 and between 0.29 and 11.8 units at C4. It rises to 0.19 at w = 0.20 and to 1.0 at w = 0.10. It also rises to 0.99 and 1.0 once w exceeds about 15. In that case the tail of the domain lies far from the origin, where the initial tanh features are saturated and nearly constant.

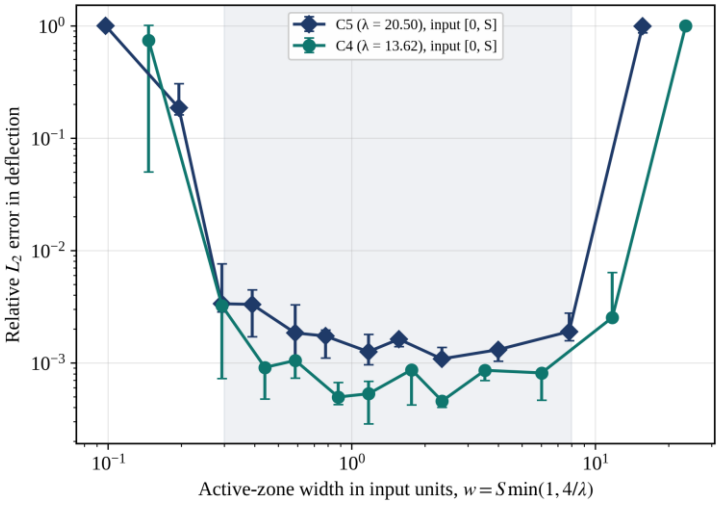

Fig. 11. Effect of the width of the active zone in input units, with the pile head held at the input origin. Median relative L_2 error in deflection over five seeds, with interquartile range, for cases C5 and C4.

In the second test the width was held at w = 4 and the head was moved away from the origin, using x = z/T + x_0. A centred input x ∈ [−S/2, S/2] was also tested, which places the head S/2 units from the origin. The results are plotted in Fig. 12. With w = 4, the median error is below 1.3×10^{-3} for head offsets up to 6 units, but individual seeds fail at offsets of 1 and 2 units, and the median error is 0.95 at 10.25 units. With the centred input, the median error is 7.0×10^{-3} at S = 2, which is F3 itself, and it is 0.94 at S = 20.5, where the head is 10.25 units from the origin.

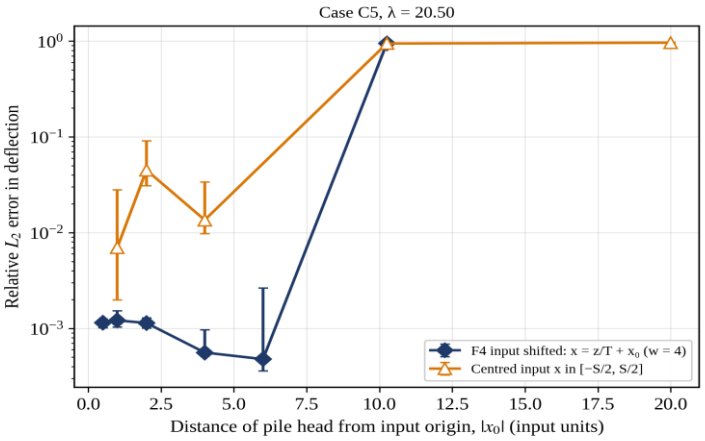

Fig. 12. Effect of the distance between the pile head and the input origin for case C5. Diamonds: F4 input shifted by x_0 at a fixed width w = 4. Triangles: centred input x ∈ [−S/2, S/2]; S = 2 reproduces F3. Median relative L_2 error over five seeds, with interquartile range.

These tests refine the criterion of the original submission, which attributed failure to w alone and placed the threshold at w ≈ 1. Two conditions must hold. The active zone must be at least about 0.3 input units wide, and it must lie within a few input units of the origin. The uncentred input [0, 2], which has the same width as F3, trains successfully at C5 with a median error of 3.3×10^{-3}, whereas F3 fails in half of its seeds. The failure of F3 at λ = 20.5 therefore combines a narrow zone, w = 0.39, with an off-origin location. F4 satisfies both conditions by construction, because w = 4 for every λ ≥ 4 and the head is always at the origin. Under F1, w = min(L, 4T) never falls below 3.90 across the five cases and the head is at the origin, which is why the boundary-layer failure is not observed for F1 and why its weakness is confined to loss imbalance at low slenderness.

4.5 Training loss is not a certificate of accuracy

The consequence for the most flexible pile is shown in Fig. 13 and Fig. 14. The head deflection is missed by formulation F2, and the peak of the bending moment profile is lost. The reference is tracked by formulation F4 in both quantities. The corresponding loss histories are shown in Fig. 15. At C5, F2 reduced its loss to a median of 1.2×10^{-5} of its initial value while its median deflection error was 1.03. F3 reduced its loss only to 4.0×10^{-3} of its initial value with a median error of 0.30. F4 reduced its loss to 2.0×10^{-6} of its initial value with a median error of 1.6×10^{-3}. A reduction of the loss by five orders of magnitude therefore coexisted with a completely wrong solution.

The relation between loss and error is examined over all 100 F3 and F4 runs of Table 4 in Fig. 16. These two formulations share the same dimensionless loss, so their losses can be compared directly. Within this pair the final loss and the error are strongly rank-correlated, with a Spearman coefficient of 0.98. The loss therefore ranks runs of one formulation. It does not bound the error. A final loss of 1.3×10^{-3} corresponded to an error of 9.5×10^{-2} for F3, and nearly equal losses of 1.2×10^{-4} and 1.4×10^{-4} corresponded to errors of 1.3×10^{-2} for F3 and 3.2×10^{-3} for F4. Losses in different units, such as the SI loss of F2, cannot be compared with them at all. A converged loss cannot be used to certify a physics-informed solution. The trained solution must be checked against an independent numerical solution, a closed-form limit or measurements.

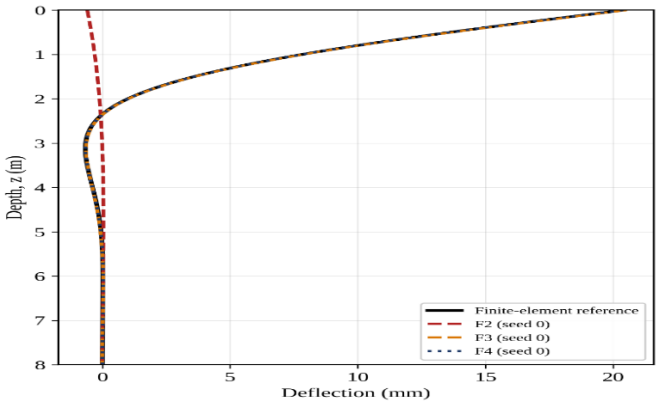

Fig. 13. Deflection profile for case C5 at relative slenderness 20.50, comparing the finite-element reference with formulations F2, F3 and F4 (seed 0). Only the upper 8 m is shown.

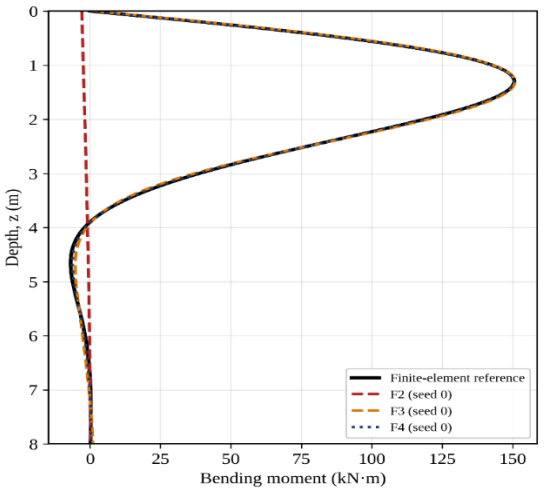

Fig. 14. Bending moment profile for case C5 at relative slenderness 20.50, comparing the finite-element reference with formulations F2, F3 and F4 (seed 0). Only the upper 8 m is shown.

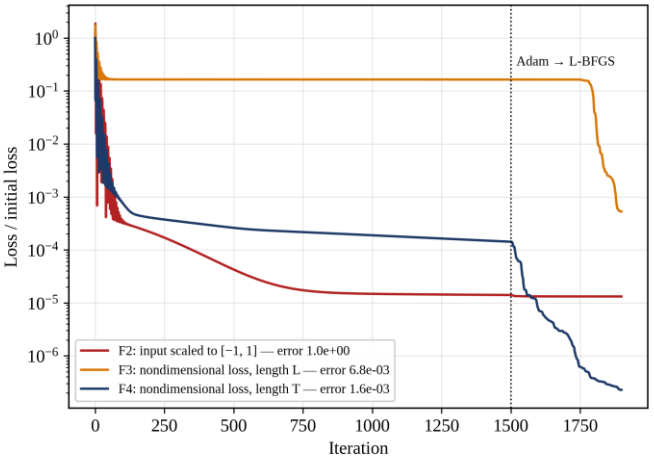

Fig. 15. Training loss histories for case C5 (seed 0) under formulations F2, F3 and F4, each normalised by its own initial value. The legend gives the final relative L_2 error in deflection.

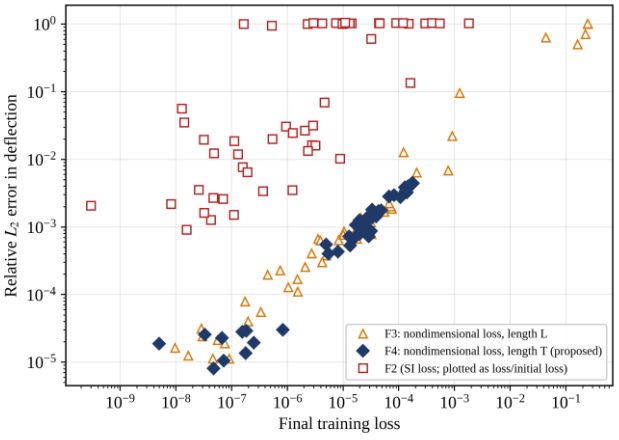

Fig. 16. Final training loss against relative L_2 error in deflection for all forward runs of the five test cases. F3 and F4 share the same dimensionless loss; F2 is plotted as the ratio of final to initial loss because its SI loss is not comparable.

4.6 Dependence on architecture, activation, collocation and training strategy

The comparison above uses one architecture and one training schedule. Thirteen variants were therefore run at case C5 with five seeds each, changing one setting at a time from the baseline. The variants cover depth (2, 4 and 5 hidden layers), width (16 and 64 neurons), activation (sine [91] and SiLU [92] instead of tanh), collocation density (61, 241 and 481 points), a four-fold training budget (6000 Adam and 1600 L-BFGS iterations) and a lower learning rate. F3 and F4 were compared because they differ only in the coordinate. The results are plotted in Fig. 17.

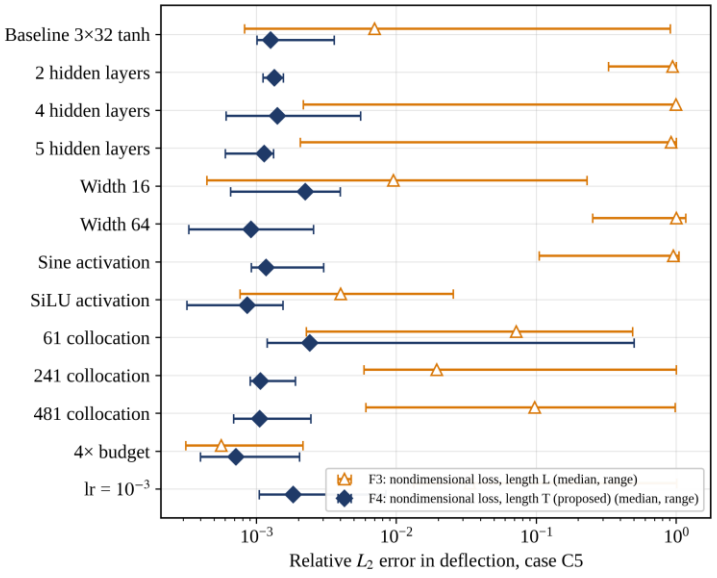

Fig. 17. Relative L2 error in deflection at case C5 (λ = 20.5) for thirteen architecture and training variants, each changed one at a time from the baseline. Markers show the median and bars the range over five seeds.

The median error of F4 lies between 7.2×10^{-4} and 2.4×10^{-3} in all thirteen variants. One seed of the 61-point variant failed, with an error of 0.50. The median error of F3 lies between 5.6×10^{-4} and 1.01. F3 fails, with median error above 0.9, for 2, 4 and 5 hidden layers, for 64 neurons and for the sine activation. It succeeds with the SiLU activation (4.0×10^{-3}), with 16 neurons (9.5×10^{-3}) and with the four-fold budget (5.6×10^{-4}). The failure of fixed-interval scaling is therefore not specific to tanh or to the chosen depth, and neither a deeper, wider nor denser network removes it. A four-fold longer training does remove it at λ = 20.5. The finding is therefore that F4 reaches a given accuracy at lower and more predictable cost, not that F3 cannot reach it eventually.

The failure threshold of F3 was then measured for four architectures by repeating the controlled sweep of Section 4.3 from λ = 6 to 30 with three seeds each. The slenderness at which the median error first exceeds 10^{-2} and the slenderness at which it first exceeds 0.9 are listed in Table 7, together with the corresponding w.

TABLE VII. FAILURE ONSET OF FORMULATION F3 IN THE CONTROLLED SLENDERNESS SWEEP FOR FOUR ARCHITECTURES (MEDIAN OF THREE SEEDS).

Architecture	λ at median error > 10^{-2}	w at onset	λ at median error > 0.9	w at failure
3 × 32, tanh (baseline)	17	0.47	25	0.32
5 × 32, tanh	20	0.40	25	0.32
3 × 64, tanh	17	0.47	20	0.40
3 × 32, sine	17	0.47	20	0.40

The onset lies at w = 0.40 to 0.47 for all four architectures, and complete failure occurs at w = 0.32 to 0.40. The threshold moves by one step of the sweep between architectures. It is therefore not an artefact of one network, but it is a property of this budget and should be read as w ≈ 0.4 to 0.5 for a centred input, not as a universal constant.

Adaptive training strategies were also tested at C4 and C5 with five seeds. Learning-rate annealing of the boundary weight [48] updates the weight every ten iterations from the ratio of the largest residual gradient to the mean boundary gradient. Residual-based adaptive distribution (RAR-D) [51] resamples the 121 collocation points every 250 Adam iterations from 2001 candidates, with probability proportional to the residual magnitude plus its mean. The results are listed in Table 8.

TABLE VIII. MEDIAN RELATIVE L_2 ERROR IN DEFLECTION WITH ADAPTIVE TRAINING STRATEGIES (FIVE SEEDS), COMPARED WITH THE BASELINE OF TABLE 4 (TEN SEEDS).

Case	Formulation	Baseline	Adaptive weighting [48]	Adaptive sampling [51]
C4 ($\lambda = 13.62$)	F2	1.02	0.36	1.00
	F3	2.79×10^{-4}	3.27×10^{-3}	3.15×10^{-3}
	F4	7.15×10^{-4}	1.30	7.96×10^{-4}
C5 ($\lambda = 20.50$)	F2	1.03	3.48	1.03
	F3	0.300	5.09×10^{-2}	0.977
	F4	1.64×10^{-3}	0.281	7.74×10^{-4}

Neither strategy repairs F2. Adaptive weighting reduced the median error of F3 at C5 from 0.30 to 5.1×10^{-2}, but it degraded F4, to 1.30 at C4 and 0.28 at C5, because the gradient-ratio rule drove the boundary weight to large values once the dimensionless loss was already balanced. Adaptive sampling halved the median error of F4 at C5 but did not rescue F3 there. The two strategies therefore act on different causes. Weighting acts on loss imbalance, sampling acts on the placement of collocation points, and the coordinate acts on the representation seen by the network. Coordinate normalisation is complementary to them, not a substitute, and a weighting scheme tuned for an unbalanced loss can harm a balanced one.

4.7 Generality of the normalisation

Five further tests examine conditions outside the five linear cases. Their results are summarised in Table 9:

The first test concerns an error in the soil stiffness used to compute T. For cases C2, C4 and C5, F4 was trained with T computed from a stiffness 0.1 to 10 times the true value, with five seeds each. The results are plotted in Fig. 18. Because of the exponent of one fifth, a stiffness error by a factor of 10 changes T by a factor of 1.58. The median error stays between 3.6×10^{-4} and 5.8×10^{-3} for every underestimate and for overestimates up to a factor of 2. At C5, overestimates by factors of 4 and 10 raise the median error to 2.7×10^{-3} and 8.4×10^{-3}, and single seeds then fail with errors of 0.25 and 0.65. An overestimate shortens T and lengthens the input span, and the domain tail then moves far from the origin, which is the second failure condition of Section 4.4. When the stiffness is uncertain, the lower estimate is therefore the safer choice.

The second test concerns profiles with $n \neq 1$. For a pile with $EI = 60.4$ MN·m² and $L = 18$ m, k_{ref} was chosen so that $\lambda = L/T_n = 15$ for $n = 0$, 0.5, 1.5 and 2.0. F4 was trained with T_n from Equation (6) and with the Matlock–Reese length T_1 used in the original submission, and F3 was trained for comparison. For $n = 0$ and 0.5 all three give median errors between 2.4×10^{-4} and 9.7×10^{-3}. For $n = 1.5$, F3 fails with 0.92 while F4 gives 1.0×10^{-2} with T_n and 1.5×10^{-2} with T_1. For $n = 2.0$ all three fail: F4 gives 0.23 with T_n and 0.31 with T_1, and F3 gives 1.0. The generalised length restores the parameter-free form and is modestly more accurate, but the gain over T_1 is small because T_1 differs from T_n by less than a factor of 1.8 in these cases. The failure at $n = 2$ is a limitation of the method at this budget. The dimensionless stiffness then grows as ζ^2 and reaches 225 at the toe, and the residual below the active zone dominates the loss.

The third test concerns layered soil. Two two-layer profiles with an interface at 3 m were built from a hyperbolic-tangent transition 0.1 m thick: a soft layer of 2 MN/m² over a stiff layer of 60 MN/m², and the reverse with 60 MN/m² over 5 MN/m². T was computed from Equation (5) as 2.34 m and 1.00 m. F4 gave median errors of 6.6×10^{-2} and 2.2×10^{-2}, against 0.24 and 3.5×10^{-2} for F3 and 0.46 and 0.44 for F2. F4 is the most accurate formulation, but its errors of 2 to 7 per cent are an order of magnitude above those for continuous profiles. The stiffness jump produces a jump in the fourth derivative of the deflection, which a single smooth network resolves poorly. Domain decomposition with interface conditions, as used by Vahab et al. [66] and in the methods of [35, 36], is the established remedy and is outside the scope of this study.

The fourth test concerns nonlinear soil. Case C4 ($D = 0.45$ m, $L = 18$ m) was embedded in sand described by the API p–y curves [70, 73], with $\phi' = 35°$, $\gamma' = 10$ kN/m³ and an initial modulus gradient of 24 MN/m³. The residual of Equation (1) was replaced by $EI\, y'''' + A\, p_u \tanh(k\, z\, y / A\, p_u) = 0$, and T was computed from the initial modulus as 1.20 m ($\lambda = 15.0$). At $H = 50$ kN the head deflection is 1.15 times the linear value, and at $H = 150$ kN it is 2.3 times the linear value. F4 gave median errors of 1.9×10^{-3} and 9.9×10^{-3}, F3 gave 4.9×10^{-3} and 4.4×10^{-2}, and F2 failed with 1.04 and 1.01. The moment profile

at the higher load is shown in Fig. 19. Normalisation by T computed from the initial stiffness therefore remains effective when the secant stiffness is well below the initial value.

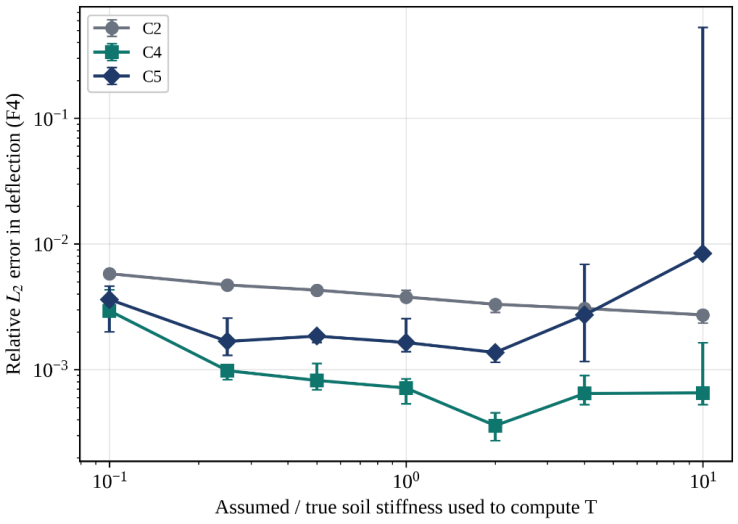

Fig. 18. Sensitivity of formulation F4 to the soil stiffness assumed when computing T. Median relative L_2 error in deflection over five seeds (ten at a ratio of 1), with interquartile range.

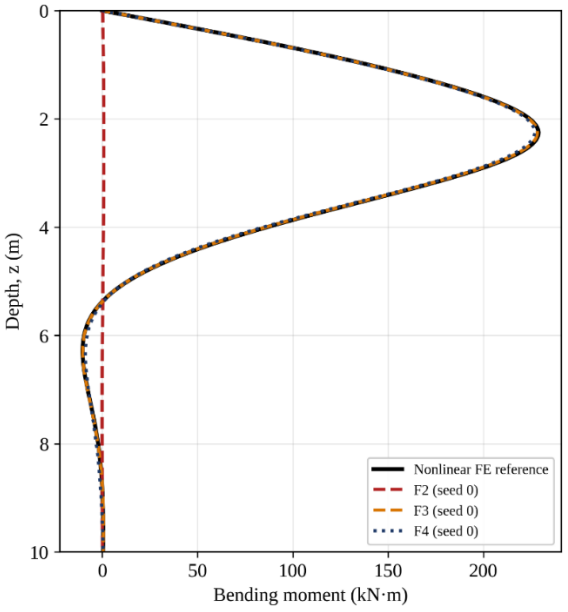

Fig. 19. Bending moment for case C4 in sand with nonlinear API p–y curves at H = 150 kN (head deflection 2.3 times the linear value), comparing the nonlinear finite-element reference with formulations F2, F3 and F4 (seed 0).

The fifth test uses 24 randomly generated piles. Diameter was drawn from 0.32 to 1.86 m, the section was either solid concrete (E = 30 GPa) or a steel pipe (E = 200 GPa, wall thickness 6.35 mm + D/100), length was drawn from 8 to 40 m, the exponent from {0, 0.5, 1, 1.5}, $\lambda = L/T_n$ log-uniformly from 2 to 30 (realised range 2.4 to 18.7), H from 50 to 1000 kN and a head moment $M_0 = mHT_n$ with m from 0 to 2. Each pile was trained with F2, F3 and F4 and two seeds. The per-pile median errors are plotted in Fig. 20. Over the 24 piles, the median of F4 is 2.6×10^{-3}, its 90th percentile is 6.2×10^{-3} and its maximum is 1.0×10^{-2}. The median of F3 is 1.7×10^{-3}, lower than that of F4, but its 90th percentile is 1.2×10^{-2} and its maximum is 0.52. F2 exceeds 10^{-2} for half of the piles and reaches 5.7. F4 was never the worst formulation and had the smallest maximum; F3 was more accurate for the typical pile.

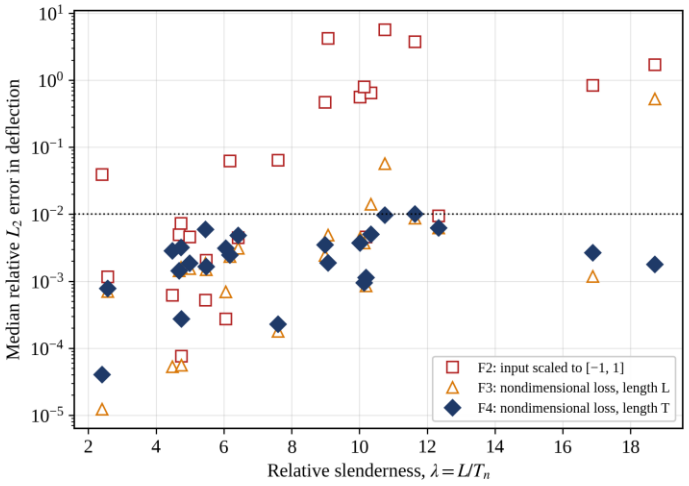

Fig. 20. Median relative L_2 error in deflection for 24 randomly generated piles (concrete and steel sections, n from 0 to 1.5, head moments), plotted against relative slenderness $\lambda = L/T_n$. The dotted line marks 10^{-2}.

TABLE IX. SUMMARY OF THE GENERALITY TESTS: MEDIAN RELATIVE L_2 ERROR IN DEFLECTION.

Test	F2	F3	F4
Stiffness for T wrong by 0.1× to 2× (C2, C4, C5)	—	—	3.6×10^{-4} to 5.8×10^{-3}
Stiffness for T overestimated 4× to 10× (C5)	—	—	2.7×10^{-3} to 8.4×10^{-3}
n = 0 / 0.5, λ = 15	—	2.4×10^{-4} / 9.7×10^{-3}	2.7×10^{-4} / 7.0×10^{-3}
n = 1.5 / 2.0, λ = 15	—	0.92 / 1.0	1.0×10^{-2} / 0.23
Two layers, soft over stiff	0.46	0.24	6.6×10^{-2}
Two layers, stiff over soft	0.44	3.5×10^{-2}	2.2×10^{-2}
API sand p–y, H = 50 kN	1.04	4.9×10^{-3}	1.9×10^{-3}
API sand p–y, H = 150 kN	1.01	4.4×10^{-2}	9.9×10^{-3}
24 random piles, median / maximum of per-pile medians	2.4×10^{-2} / 5.7	1.7×10^{-3} / 0.52	2.6×10^{-3} / 1.0×10^{-2}

4.8 Design procedure

A procedure that requires no tuning is obtained from these results. First, T is computed from Equation (5) using the best available estimate of soil stiffness; for a power law Equation (6) gives it directly, and for nonlinear soil the initial modulus may be used. Where the stiffness is uncertain, the lower estimate should be preferred. Second, the network input is set to $\zeta = z/T$, so that the input spans $[0, \lambda]$ with the pile head at the origin. Third, the deflection is scaled by $y_c = H\,T^3/EI$ and each loss term is divided by its natural magnitude, giving Equation (13) and Equation (15). Fourth, the trained solution is verified against an independent numerical solution rather than against the loss value. The estimate of T need not be accurate. A stiffness estimate wrong by a factor of two shifts T by 13 per cent, and underestimates by up to a factor of ten were tolerated in Section 4.7. A prior stiffness deliberately set to half the true value is used in the inverse analyses of Section 5, and the resulting normalisation is still effective.

5. IDENTIFICATION OF THE SUBGRADE REACTION PROFILE

5.1 Inverse problem

The modulus k(z) is assumed known in the forward problem. In practice it is the least certain input. Deflection is measured at discrete depths by inclinometers installed in test piles and stiffness is inferred from those readings; the interpretation of

such tests was addressed by Reese and Cox [93]. Physics-informed networks are well suited to this task because the forward and inverse problems are served by the same loss. Consolidation and seepage parameters were estimated in this way by Ito et al. [94], and material and load parameters were identified through a differentiable solver by Cheng et al. [95]. Inverse analysis of granular flows was performed on the same principle by Choi and Kumar [96]. Maximum-likelihood back-analysis in geotechnics was formalised by Ledesma et al. [97]. What the measurements can and cannot determine was not established for the lateral pile problem in these studies. It is established here.

Configuration A uses the geometry of case C3 as the truth, with k_{ref} = 120 MN/m² and n = 0.70, so that T_n = 1.598 m and λ = 7.51. Deflection is sampled at N_s equally spaced depths from the head to the toe and perturbed by the baseline Gaussian model of Equation (16). Both ln k_{ref} and n are treated as trainable scalars alongside the network weights, and the data misfit, normalised by y_c, is added with unit weight to the loss of Equation (9). The prior is deliberately incorrect: half the true k_{ref} and a linear profile, n = 1. The normalising length T = 2.07 m is computed from that prior, so the answer is not supplied to the scheme through its scaling. Sensor counts of 3, 4, 6, 8 and 12 are combined with noise levels of 0, 2, 5 and 10 per cent, and each combination is repeated with ten seeds, giving 200 runs. The seed changes both the initialisation and the noise realisation.

Three further configurations test whether the results depend on this one setting. Configuration B uses the pile of case C4 (L = 18 m) with n = 0.3 and T_n = 1.40 m (λ = 12.9). Configuration C uses the pile of case C2 (L = 15 m) with n = 1.2 and T_n = 2.60 m (λ = 5.8). Configuration D repeats A with a head moment $M_0 = HT_n$ acting against the lateral load, which represents partial head restraint. Each is run with eight readings at 0 and 2 per cent noise and ten seeds.

5.2 Practical non-identifiability

At zero noise with twelve sensors the median recovered k_{ref} is 40.8 per cent above the truth and the median recovered exponent is 0.835 against 0.700. The data are fitted almost exactly by these values. The reason is shown in Fig. 21. The 200 recovered parameter pairs of configuration A cluster along a single locus rather than scattering about the truth. That locus is the one along which stiffness at one particular depth is preserved. If the same stiffness at a depth z* is given by two profiles, their parameters are related by Equation (20).

$$ln\left(\frac{\hat{k}_{ref}}{k_{ref}}\right) = -(\hat{n} - n)\, ln\left(\frac{z^*}{L}\right) \tag{20}$$

Regression of the 200 results on Equation (20) gives z* = 0.80 m with a coefficient of determination of 0.984. Expressed in mechanical terms, $z^* = 0.50T_n$. The stiffness recovered at that depth has a median ratio of 0.974 to the truth. The two parameters are individually biased by tens of per cent, while their combination at z* is recovered to within 3 per cent.

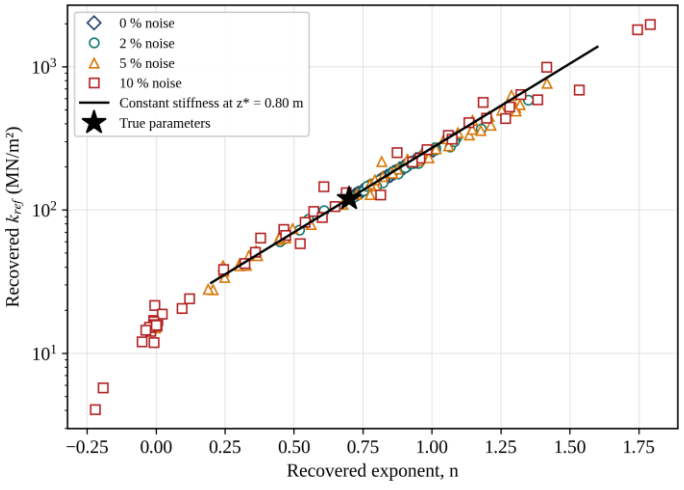

Fig. 21. Recovered stiffness and exponent from 200 inverse runs of configuration A, showing that the results cluster along the locus of constant stiffness at the pivot depth z* rather than scattering about the true parameters.

The original submission called this behaviour structural non-identifiability. That term is too strong [98, 99]. For exact, noise-free data the two parameters are identifiable in principle, because the sensitivity matrix of Section 5.3 is not singular. It is, however, ill-conditioned, and a gradient-based optimiser stops at a point on a flat valley of nearly equal misfit. The correct term is practical non-identifiability, and the parameter set is sloppy in the sense of Gutenkunst et al. [100].

5.3 Linearised identifiability analysis

A linearised analysis explains the locus and predicts the pivot without any neural network. Let $\theta = (\ln k_{ref}, n)$. The sensitivity matrix J and the Fisher information matrix F are given by Equation (21), in which σ is the noise standard deviation. The derivatives are computed by central differences of the finite-element solution.

$$J_{ij} = \frac{1}{\sigma}\frac{\partial y(z_i)}{\partial \theta_j}, \qquad F = J^{\mathrm{T}}J \tag{21}$$

Let $v = (v_1, v_2)$ be the eigenvector of F with the smallest eigenvalue. A parameter change along v changes the data least. It changes ln k(z) by $v_1 + v_2 \ln(z/L)$, which vanishes at the depth given by Equation (22). That depth is the predicted pivot.

$$z^* = L \exp\left(-\frac{v_1}{v_2}\right) \tag{22}$$

The square root of the ratio of the two eigenvalues measures how much better the stiff combination is determined than the sloppy one. The prediction is compared with the network inversions in Table 10. The linearised pivot and the pivot regressed from the network inversions agree to within $0.06T_n$ in all four configurations. The pivot recovered by a conventional finite-element least-squares back-analysis (Section 6) on the 40 datasets of configuration A at 2 per cent noise is 0.98 m, or $0.61T_n$, which lies between the other two estimates. The locus is therefore a property of the measurement, not of the neural formulation.

TABLE X. PIVOT DEPTH OF THE RECOVERED STIFFNESS PARAMETERS: REGRESSION OF NETWORK INVERSIONS COMPARED WITH THE LINEARISED PREDICTION OF EQUATION (22) (EIGHT READINGS).

Configuration	n	λ	Peak reaction depth	Network pivot (R^2)	Linearised pivot	Eigenvalue ratio, √(F-ratio)
A	0.7	7.51	0.68T	0.50T (0.984)	0.56T	74
B	0.3	12.9	0.35T	0.37T (0.999)	0.40T	101
C	1.2	5.77	0.98T	0.74T (0.997)	0.72T	68
D, head moment	0.7	7.51	1.07T	0.74T (0.992)	0.71T	34

The linearised pivot was then mapped over λ from 3 to 20 and n from 0 to 1.5 with 25 readings, as plotted in Fig. 22. For $\lambda \geq 5$ the pivot is almost independent of slenderness and depends on the exponent: $z^*/T_n \approx 0.30$ for n = 0, 0.45 for n = 0.35, 0.58 for n = 0.7, 0.67 for n = 1.0 and 0.81 for n = 1.5. The value 0.67T reported in the original submission is therefore the value for a linear profile, not a general constant. The eigenvalue ratio grows from about 20 at λ = 3 to between 90 and 177 at λ = 20, so identification becomes more sloppy as the pile becomes more flexible.

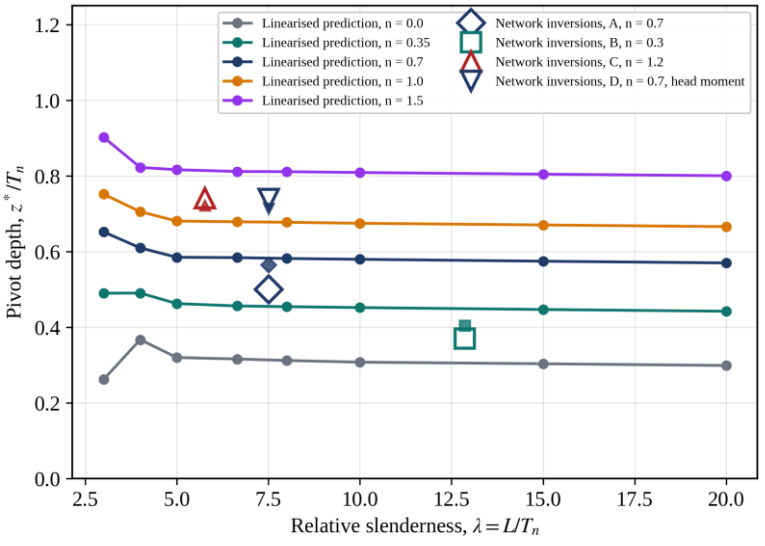

Fig. 22. Pivot depth z^*/T_n predicted by the linearised analysis as a function of slenderness and profile exponent (lines, 25 readings), compared with the pivot regressed from network inversions (open markers) and the linearised prediction for the same eight-reading configurations (filled markers).

Two further analyses test whether the limitation is specific to the power law. First, a different two-parameter family, $k = k_0 + n_h z$, was analysed in the same way. It has a pivot at 0.48T and an eigenvalue ratio of 26, so it too is practically non-identifiable. Second, a non-parametric analysis was made. The stiffness was allowed to vary independently in 40 depth bins, and the singular values of the resulting sensitivity matrix were computed for configuration A. They are plotted in Fig. 23. Normalised by the largest, the next three are 0.12, 0.035 and 0.014 with eight readings, and 0.13, 0.042 and 0.016 with 25 readings. At a noise level of 2 per cent, at most three independent combinations of the stiffness profile rise above the noise, whatever the parameterisation. Tripling the number of readings does not change this, because all readings see the same active zone. The dominant combination is a weighted average of stiffness centred at $0.74T_n$. This is a general limitation of lateral pile back-analysis from deflection data, not a feature of the power law or of the network.

5.4 Physical meaning of the pivot

The mobilised soil reaction for configuration A is plotted in Fig. 24. The reaction peaks at 1.09 m, which is $0.68T_n$. The depth containing 90 per cent of the mobilised reaction is 6.11 m, which is $3.83T_n$. The pivot lies at 0.80 m, in the upper part of the zone of high reaction and slightly above the peak. The same pattern holds in the other configurations, where the pivot lies at 0.69 to 1.06 times the depth of peak reaction (Table 10). Stiffness is constrained by the measurements where reaction is mobilised and is left free elsewhere. Individual recovered profiles are shown in Fig. 25 to converge near the pivot and to diverge above and below it.

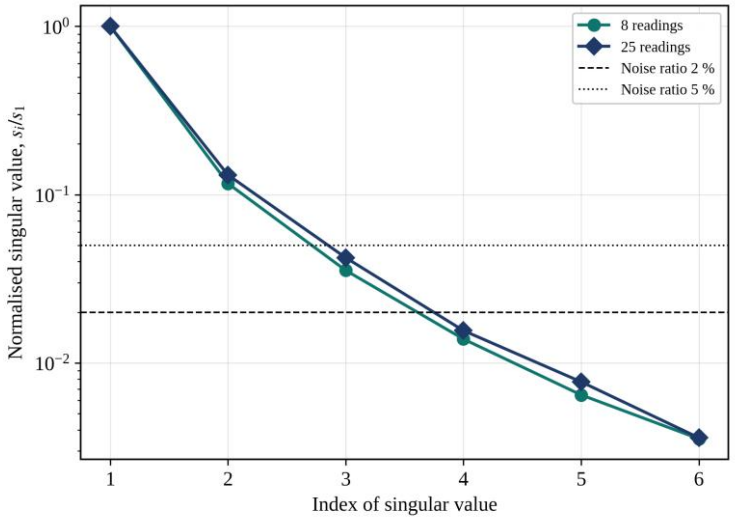

Fig. 23. Normalised singular values of the sensitivity of the readings to the stiffness in 40 independent depth bins (configuration A). Horizontal lines mark noise ratios of 2 and 5 per cent.

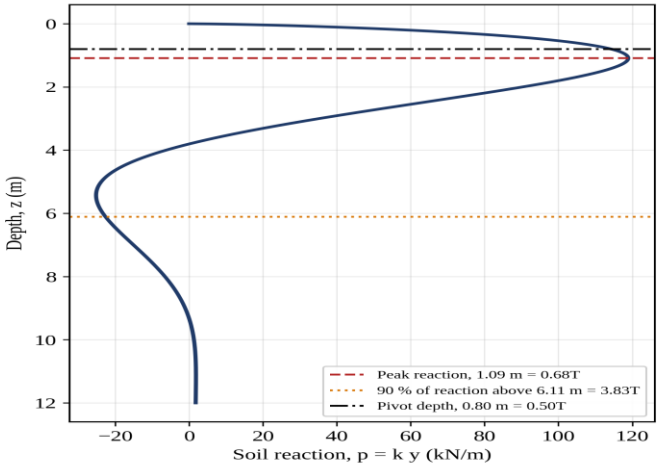

Fig. 24. Mobilised soil reaction for configuration A, showing the depth of peak reaction, the depth containing 90 per cent of the total mobilised reaction, and the pivot depth.

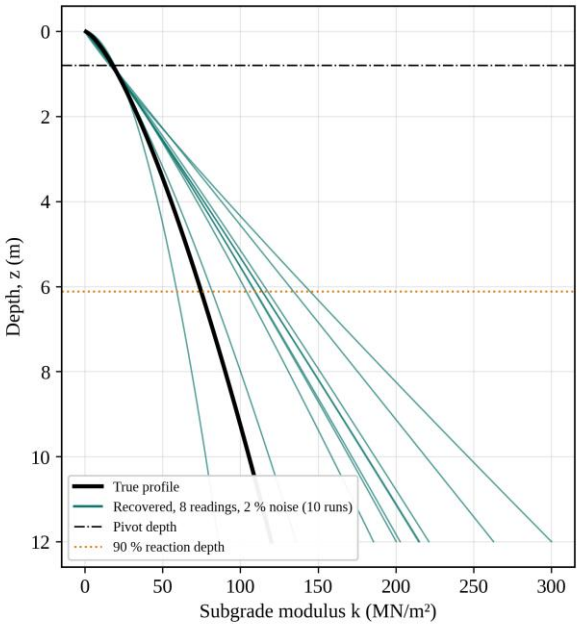

Fig. 25. Recovered subgrade reaction profiles from eight readings with 2 per cent noise (ten runs) compared with the true profile, showing convergence near the pivot depth and divergence away from it.

The practical consequence is a change in what is reported. The stiffness at the pivot depth, which is determined, should be stated by a back analysis, rather than the pair of profile parameters, which are not. Profiles that differ by tens of per cent in k_{ref} and by 0.1 in n are indistinguishable from the data. The information content of the test is overstated if either is presented as the identified profile. The same degeneracy is met by any method that fits the same readings, whether a network, a least-squares solver or a Bayesian sampler. A single point on the ridge is returned by a gradient-based optimiser, and no warning is given that the ridge exists. It becomes visible only when many independent runs are made or when the sensitivity matrix is examined.

5.5 Data requirements

TABLE XI. MEDIAN RELATIVE ERROR (%) IN SUBGRADE STIFFNESS AT THE DEPTH OF PEAK SOIL REACTION, WITH INTERQUARTILE RANGE, FROM TEN RUNS PER COMBINATION (CONFIGURATION A).

Readings N_s	0 % noise	2 % noise	5 % noise	10 % noise
3	3.0 [2.5–3.6]	2.7 [2.1–4.9]	8.1 [2.2–19.1]	17.5 [8.8–24.7]
4	4.2 [3.3–4.7]	4.2 [3.6–5.8]	8.5 [2.8–11.7]	25.1 [11.4–32.4]
6	2.4 [1.8–2.5]	3.4 [3.2–4.0]	10.7 [6.5–12.8]	16.4 [6.3–22.6]
8	1.9 [1.1–2.5]	1.8 [0.9–4.1]	6.5 [2.9–8.2]	28.3 [21.9–36.4]
12	1.8 [1.2–2.2]	4.0 [2.5–5.1]	3.9 [3.3–4.7]	14.6 [7.4–24.0]

The median error in stiffness at the depth of peak reaction is reported in Table 11 and plotted in Fig. 26. The error over the whole mobilised depth is plotted in Fig. 27. With noise not exceeding 2 per cent, the stiffness at the peak-reaction depth is recovered with a median error of 1.8 to 4.2 per cent for every sensor count, including three readings. At 5 per cent noise the median error is 3.9 to 10.7 per cent, and at 10 per cent noise it is 14.6 to 28.3 per cent, with no clear trend with sensor count. The noise level therefore matters more than the number of readings, which is consistent with Fig. 23. These values revise the thresholds of the original submission, in which three readings appeared inadequate at any nonzero noise; that result was obtained with three seeds and is not reproduced with ten.

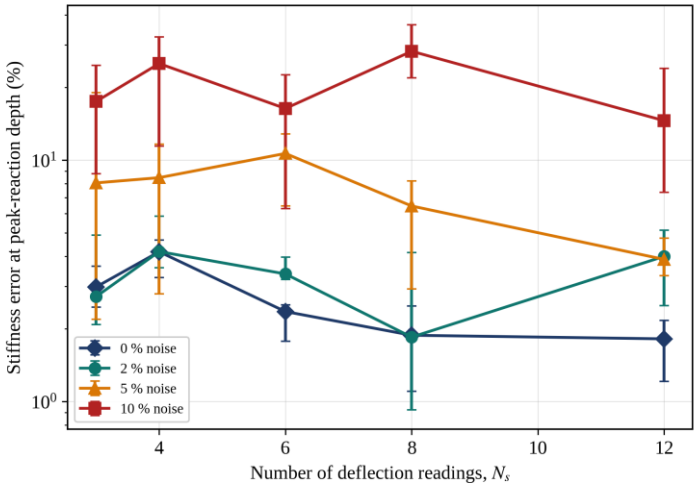

Fig. 26. Median relative error in subgrade stiffness at the depth of peak soil reaction as a function of sensor count and noise level (configuration A, ten runs, interquartile range shown).

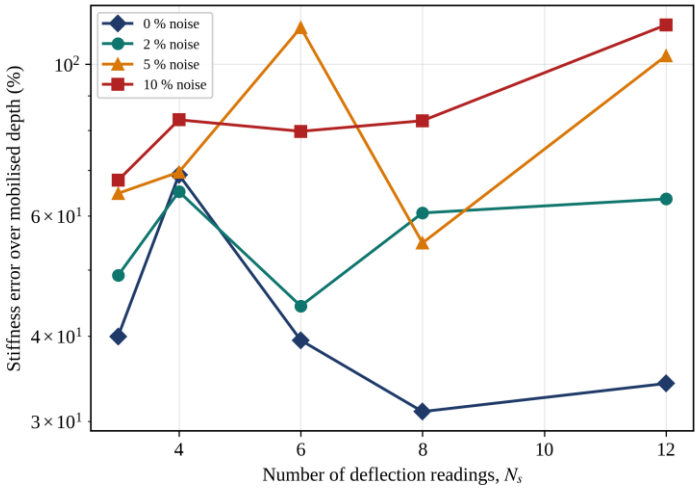

Fig. 27. Median relative L_2 error in the subgrade profile over the depth containing 90 per cent of the mobilised reaction, as a function of sensor count and noise level (configuration A).

The profile over the mobilised depth is not recovered, even without noise. Its median error is 21 to 44 per cent at zero noise and 32 to 75 per cent at 2 to 10 per cent noise. Over the full embedded length the error is larger still. The original submission stated that stiffness within the mobilised depth is recoverable. The present results show that only a weighted average concentrated near the pivot is recoverable, and an identified profile should not be reported as a function of depth.

5.6 Systematic measurement errors and model discrepancy

The Gaussian model does not represent the systematic errors of real inclinometers [88–90]. Configuration A with eight readings was therefore repeated with the heavy-tailed, bias-drift and depth-offset models of Section 3.5, with ten realisations each. A test of model discrepancy was also made. Readings were generated from the nonlinear API sand model of Section 4.7 for case C4 at H = 150 kN and inverted with the linear power-law network, so that the data come from a different model from the one assumed in the inversion. The results are listed in Table 12, together with those of the finite-element back-analysis and the smoothing spline of Section 6 on the same data.

TABLE XII. MEDIAN ERRORS (%) UNDER SYSTEMATIC MEASUREMENT ERRORS AND MODEL DISCREPANCY, EIGHT READINGS, TEN REALISATIONS. K: STIFFNESS AT THE PEAK-REACTION DEPTH; M: BENDING MOMENT.

Error model	k, network	k, FE-LS	M, network	M, FE-LS	M, spline
Gaussian, 2 %	1.8	1.3	5.9	6.5	31.4
Student t (3 d.o.f.), 2 %	3.1	2.1	4.0	5.2	29.1
Bias drift β = 2 % + Gaussian 2 %	1.8	4.3	3.2	6.2	31.9
Bias drift β = 5 % + Gaussian 2 %	9.8	11.3	6.4	8.1	30.2
Depth offset 0.10 m + Gaussian 2 %	11.5	9.5	7.3	8.7	35.4
Depth offset 0.25 m + Gaussian 2 %	29.6	27.7	10.3	12.8	34.8
Nonlinear p–y data, linear model, 2 %	—	—	6.9	9.5	19.8

Heavy-tailed noise and a 2 per cent bias drift changed the stiffness error at the peak-reaction depth by less than 3 percentage points. A 5 per cent bias drift raised it to 9.8 per cent, and a depth offset of 0.10 m and 0.25 m raised it to 11.5 and 29.6 per cent. Depth registration is therefore the most damaging of the errors tested, because the deflection gradient near the head is large and a small shift in depth produces a large apparent change in curvature. The moment error of the physics-constrained estimators stayed between 3.2 and 12.8 per cent in every scenario, whereas the smoothing spline stayed between 29 and 35 per cent. With data from the nonlinear model the network recovered the moment with a median error of 6.9 per cent and the deflection with 2.7 per cent. The identified linear stiffness at the peak-reaction depth, 2.33 m, was 0.62 times the secant modulus of the nonlinear truth at that depth. A linear back-analysis of a nonlinear test therefore yields an effective, load-dependent stiffness that should not be interpreted as a soil property.

6. RECOVERY OF BENDING MOMENTS FROM SPARSE MEASUREMENTS

Bending moments, not deflections, are required for design. Moments are obtained in field practice by differentiating measured deflection twice. Noise is amplified by that operation [101]. Six estimators are compared in this section, all given identical data: 4 to 12 readings of configuration A contaminated by 2 per cent Gaussian noise, with ten realisations. The data-only network fits the readings with the network of Section 3 and a data misfit alone, and the fit is then differentiated. It carries no regularisation. The polynomial fit uses least squares with the degree, from 2 to $\min(6, N_s - 2)$, chosen by leave-one-out cross-validation. The cubic smoothing spline [102] has its smoothing parameter chosen by generalised cross-validation [103]; it requires at least five readings. The Tikhonov estimator [104] fits a deflection field on 241 grid points with a penalty on its third derivative, so that curvature itself is not penalised, and the weight is chosen by generalised cross-validation. The finite-element back-analysis (FE-LS) fits $\ln k_{ref}$ and n by Levenberg–Marquardt least squares [105] with a 400-element solver, starting from the same prior as the network, and evaluates the moment from the fitted model. The physics-constrained network is that of Section 5. The results are reported in Table 13 and shown in Fig. 28, Fig. 29 and Fig. 30. The data-only network reproduces the readings but not the moment. Its median moment error is 58 per cent with four readings and 156 to 513 per cent with six or more, because a closer fit to noisy readings produces larger spurious curvature. Regularisation removes most of that instability but not the bias. The smoothing spline gives 26 to 46 per cent, the Tikhonov estimator 39 to 80 per cent and the polynomial 56 to 80 per cent. Their errors do not fall as readings are added, because the smoothing that suppresses noise also flattens the moment peak, and six to twelve readings do not resolve a peak about 2T wide. The two physics-constrained estimators give 2.7 to 9.9 per cent (FE-LS) and 5.4 to 6.6 per cent (network). The improvement over the best regularised estimator is a factor of 4.8 to 12.

TABLE XIII. MEDIAN RELATIVE L_2 ERROR (%) IN BENDING MOMENT RECOVERED FROM NOISY DEFLECTION READINGS AT 2 % NOISE, WITH INTERQUARTILE RANGE (TEN REALISATIONS; N/A: THE SPLINE REQUIRES AT LEAST FIVE READINGS).

Estimator	$N_s = 4$	$N_s = 6$	$N_s = 8$	$N_s = 12$
Data-only network	57.5 [43.4–70.0]	513.1 [264.9–872.4]	155.6 [62.5–331.0]	231.9 [61.5–1457.6]
Polynomial (LOOCV)	79.9 [79.8–79.9]	56.4 [55.0–63.4]	71.8 [67.2–74.2]	64.5 [52.7–70.0]
Smoothing spline (GCV)	n/a	25.8 [24.1–26.8]	31.4 [24.3–41.3]	45.9 [32.4–84.5]
Tikhonov, 3rd derivative (GCV)	79.7 [79.6–79.8]	40.9 [37.0–44.8]	38.9 [32.8–55.1]	47.4 [36.9–83.2]

Estimator	$N_s = 4$	$N_s = 6$	$N_s = 8$	$N_s = 12$
FE least-squares back-analysis	9.9 [4.3–14.3]	2.7 [1.7–8.2]	6.5 [4.2–8.1]	4.1 [1.9–8.0]
Physics-constrained network	6.6 [5.5–7.2]	5.4 [4.1–6.7]	5.9 [4.8–6.9]	5.9 [4.5–6.4]

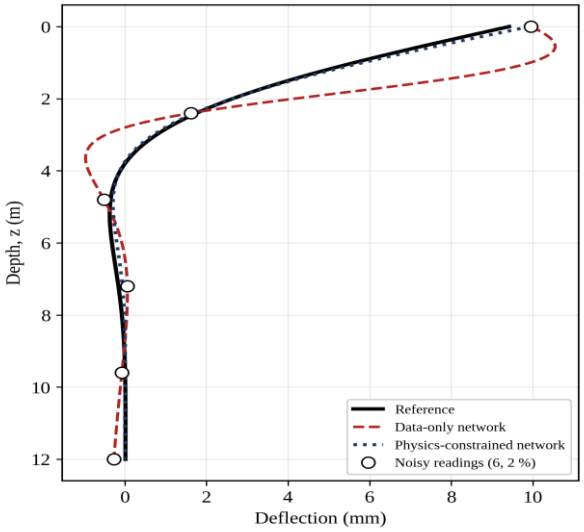

Fig. 28. Deflection fits from the data-only network and the physics-constrained network using six noisy readings, compared with the finite-element reference.

The comparison changes the conclusion of the original submission in one respect. The gain comes from imposing the governing equation, not from the neural network. The conventional finite-element back-analysis imposes the same equation and is as accurate. Regularised differentiation, which uses smoothness rather than mechanics, is not sufficient for the moment peak of a flexible pile. Bending moments should therefore be obtained from inclinometer readings through a mechanical model, by a network or by a conventional solver, and not by differentiating a smoothed curve.

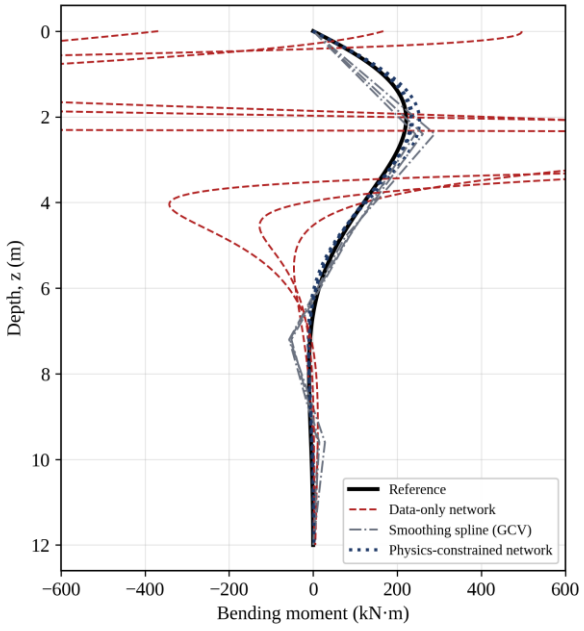

Fig. 29. Bending moment profiles recovered from six noisy deflection readings by the data-only network, the smoothing spline and the physics-constrained network, compared with the finite-element reference. Three realisations are shown for each estimator.

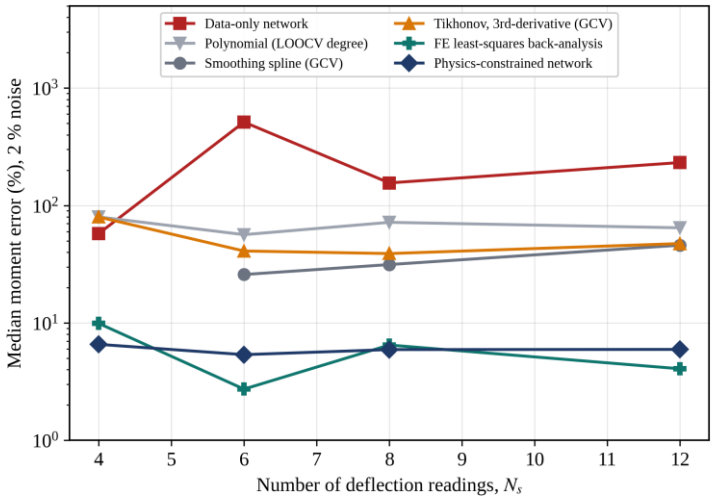

Fig. 30. Median relative error in recovered bending moment as a function of the number of deflection readings for six estimators, at 2 per cent noise.

7. COMPUTATIONAL COST

Wall-clock times on one processor core are listed in Table 14. The finite-element forward solution with 2000 elements takes 3.0 ms. A forward network training takes 1.77 s, about 590 times longer. The finite-element back-analysis takes a median of 0.05 s, including all forward solutions of the Levenberg–Marquardt iterations. A network inversion takes 1.69 s, about 35 times longer. The original submission stated that the inverse problem cannot be solved by a conventional solver. That statement was incorrect and has been removed. For this one-dimensional linear problem a conventional back-analysis is faster and equally accurate.

TABLE XIV. WALL-CLOCK COST ON A SINGLE PROCESSOR CORE (MEDIAN).

Task	Finite element	Physics-informed network
Forward solution, one pile	0.0030 s (2000 elements)	1.77 s (1900 iterations)
Inverse solution, two parameters	0.05 s (LM, 400 elements)	1.69 s
One-off compilation	—	≈ 15 s
Moment error, 8 readings, 2 % noise	6.5 %	5.9 %

The value of the network formulation lies elsewhere. The same code solves the forward and inverse problems, a change of profile or p–y law changes only the residual, and derivatives with respect to any parameter are available without a separate sensitivity code. These advantages matter in higher dimensions and for parameter fields rather than scalars, where conventional back-analysis requires adjoint or surrogate models. They were not tested here. The present study concerns the conditions under which the network formulation is accurate, not whether it is the fastest tool for a one-dimensional pile.

8. COMPARISON WITH PREVIOUS WORK

The reference solver agrees with the closed-form solution of Hetényi [1] to 7.9×10^{-6} per cent and with the coefficients of Matlock and Reese [2] to between 0.03 and 0.24 per cent. The latter deviations are smaller than the reading precision of the original design charts. It was established by Ouyang et al. [67] that the lateral pile problem can be solved and back analysis performed by physics-informed networks. The present results are consistent with that conclusion and extend it in two directions. Relative slenderness was not varied systematically by Ouyang et al. [67], and accuracy is shown here to depend strongly on it once a fixed, centred input interval is used. Back analysis of soil parameters was reported by Ouyang et al. [68] and, for axisymmetric and layered pile–soil systems, by Vahab et al. [66], but identifiability limits were not quantified. A two-parameter profile is shown here to be practically non-identifiable, and the combination that is identifiable is located and predicted.

The pivot depth can be compared with established design knowledge. The peak bending moment was placed by Matlock and Reese [2] at a depth coefficient of about 1.3, which the present solver reproduces as 1.33. The pivot depth is shallower, between 0.3T and 0.8T depending on the exponent. This is consistent with the distinction between the depth at which the pile is most stressed and the depth at which the soil is most mobilised. The depth containing 90 per cent of the mobilised reaction, $3.83T_n$ in configuration A, is consistent with the common design assumption that lateral resistance is derived from the upper part of a flexible pile [3, 76].

The scaling result connects to work outside geotechnics. Training failure was attributed by Wang et al. [47, 48] to imbalance between loss terms, and adaptive weighting was proposed. Table 8 and the failure of F3, whose loss is balanced, show that loss balancing alone is insufficient here, because part of the difficulty is set by the coordinate. The coordinate effect is a form of the spectral bias documented by Rahaman et al. [53] and Xu et al. [54]: a localised feature far from the input origin or narrower than the initial feature scale is learned slowly. Fourier-feature embeddings counter the same bias by rescaling the input [55, 56]; characteristic-length normalisation achieves the rescaling from mechanics, without a tuned frequency. It was found by Mandl et al. [57] that training for consolidation was accelerated by an affine change of variables. The same phenomenon is reported here in a fourth-order setting, in which the sensitivity to the coordinate scale is stronger because the fourth power of that scale is carried by the leading term.

On data requirements, it is reported by Shahin et al. [25] and by Yaghoubi et al. [26] that training sets of several hundred records are typically needed by purely data-driven geotechnical networks. A moment error of about 6 per cent is reached here from four to twelve readings of a single pile. The comparison is not exact, because regression across many sites is concerned in those reviews while one boundary-value problem with a known governing equation is concerned here. The direction of the difference is consistent with the data-efficiency argument of Raissi et al. [29], and Section 6 shows that the same efficiency is available from a conventional mechanical model.

9. LIMITATIONS

Eight limitations qualify these results:

First, all data are synthetic. Deflections are generated by the verified finite-element solver, and in Section 5.6 by a nonlinear p–y model, not by load tests. The scaling findings concern the numerical solution of a stated equation, and they carry over to field data only insofar as that equation describes the pile. The identifiability findings concern the information content of deflection readings, and Section 5.3 shows them to be a property of the sensitivity of the readings rather than of the network, so they apply to any back-analysis of the same measurements. What synthetic data cannot establish is the size of the errors that real tests would produce. Real inclinometer data contain bias shifts, depth-registration errors, casing deformation and installation effects [88–90]. Three of these were modelled in Section 5.6, where a 0.25 m depth offset raised the stiffness error to about 30 per cent. The errors of real tests are not known in advance, and the thresholds of Table 11 and Table 12 should be treated as lower bounds on what a field test would achieve. Validation against an instrumented full-scale test is the necessary next step.

Second, the soil model of the main study is a linear Winkler medium. The nonlinear p–y test of Section 4.7 used static API sand curves only [70, 73]; soft-clay curves [106] were not tested. Gapping, cyclic degradation and the accumulation of rotation under long-term cyclic loading [107, 108] were not modelled. Under cyclic loading the secant stiffness, and therefore T, changes with the number of cycles, and a normalisation fixed at the start of an analysis would need to be updated.

Third, the Winkler idealisation ignores the coupling between springs and the three-dimensional geometry of the pile. Continuum and three-dimensional finite-element analyses [5, 12, 109] and strain-wedge models [110] show that the subgrade modulus depends on diameter and on pile shape, and large-diameter monopiles require additional base and distributed-moment reactions [111]. The characteristic length would then be defined from an equivalent stiffness, and the geometric criterion of Section 4.4 would apply to that length. This was not tested.

Fourth, the normalisation was not effective for every profile. Sharp stiffness contrasts between layers gave errors of 2 to 7 per cent, and a profile growing as the square of depth ($n = 2$) gave errors above 20 per cent at $\lambda = 15$. Domain decomposition [35, 36, 66] or a truncated computational domain would be needed in those cases.

Fifth, the failure thresholds of Section 4.4 were measured at a fixed training budget and for four architectures. A four-fold longer training removed the failure of F3 at $\lambda = 20.5$. The thresholds $w \approx 0.3$ to 0.5 and the location condition are therefore statements about cost at a given budget, not about what a network can eventually represent.

Sixth, adaptive weighting and adaptive sampling were each tested in one standard form. Other forms, such as self-adaptive weights [49], adaptive activations [112] or Fourier-feature embeddings [55], may interact differently with the coordinate choice.

Seventh, only two profile parameters were identified in the parametric inversions. The non-parametric analysis of Section 5.3 shows that at most about three combinations are resolvable at 2 per cent noise, so richer layered profiles would be less identifiable still, unless deflection is combined with other measurements such as strain or head rotation.

Eighth, the cost comparison of Section 7 is for a one-dimensional problem on one processor core. It does not predict the relative cost for two- or three-dimensional problems, where the network formulation may compare differently.

10. CONCLUSIONS

The manner in which variable scaling controls the accuracy of physics-informed neural networks for a laterally loaded pile on a Winkler medium was examined, and the limits of subgrade stiffness identification from sparse deflection readings were established. The conclusions below are demonstrated for this model, for the ranges tested (λ from 2 to 30, n from 0 to 1.5, linear and static nonlinear p–y soil) and for synthetic data. Statements (1) to (4) concern the numerical method; statements (5) to (8) concern what deflection readings can determine.

1. Scaling depth to the centred interval [−1, 1] failed for flexible piles at the baseline budget. With an SI loss (F2) the median deflection error exceeded 0.98 for $\lambda \geq 12$ in the controlled sweep and was 1.03 at $\lambda = 20.5$. With a nondimensional loss (F3) it rose from 1.6×10^{-2} at $\lambda = 17$ to 1.0 at $\lambda = 25$. Loss normalisation alone did not repair the failure.

2. The failure is geometric. Training succeeded when the active zone, about 4T deep, was at least about 0.3 to 0.5 input units wide and lay within a few units of the input origin, where the initial network features are centred. Fixed-interval scaling violates both conditions as λ grows; normalisation by the characteristic length satisfies both by construction.

3. Characteristic-length normalisation gave the smallest worst-case error in every test set: a worst median of 3.8×10^{-3} for the five test cases, below 3.2×10^{-3} up to $\lambda = 30$, between 7.2×10^{-4} and 2.4×10^{-3} across 13 architecture variants, and a maximum per-pile median of 1.0×10^{-2} for 24 random piles. It was not the most accurate formulation for every pile. The generalised length defined by $k(T)T^4 = EI$ makes the equation parameter-free for any power law and applies to layered and nonlinear profiles, but it did not prevent failure for sharp layer contrasts or for $n = 2$.

4. A converged loss does not certify a correct solution. At $\lambda = 20.5$ the loss of F2 fell by five orders of magnitude while its error was 1.03. Within one dimensionless formulation the loss ranked runs by accuracy, but it did not bound the error.

5. The two-parameter subgrade profile is practically non-identifiable. Across 200 inversions the recovered parameters lie on the locus of constant stiffness at a pivot depth ($R^2 = 0.984$), with individual parameter biases of tens of per cent and a stiffness at the pivot within 3 per cent of the truth.

6. The pivot depth is predicted by a linearised sensitivity analysis to within 0.06T in four configurations. It depends mainly on the profile exponent, from about 0.3T for $n = 0$ to 0.8T for $n = 1.5$, and lies near and slightly above the depth of peak soil reaction. At 2 per cent noise at most about three independent stiffness combinations are resolvable from deflection readings, whatever the parameterisation. Identified stiffness should be reported at the pivot depth, not as a profile.

7. At noise levels up to 2 per cent, the stiffness at the peak-reaction depth was recovered with a median error below 5 per cent from three to twelve readings. The noise level mattered more than the number of readings. Depth-registration errors of 0.25 m raised the error to about 30 per cent.

8. Bending moments were recovered with a median error of 5 to 7 per cent when the governing equation was imposed, against 26 to 80 per cent for regularised differentiation and 58 to 513 per cent for an unregularised network. A conventional finite-element back-analysis gave the same accuracy at about one-thirtieth of the cost of the network inversion.

The practical recommendations are direct. When a physics-informed network is used for a laterally loaded pile, depth should be normalised by the characteristic length with the pile head at the input origin, the trained solution should be verified against an independent solution rather than against the loss, and identified soil stiffness should be reported at the pivot depth only. These recommendations are demonstrated for the Winkler pile model; their transfer to other geotechnical problems requires the same kind of verification.

Statements & Declarations

Acknowledgement:
The authors would like to acknowledge the great support provided by Mustansiriyah- University, Baghdad- Iraq, which allowed them to achieve this research study.

Funding
The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

CRediT authorship contribution statement
All authors achieved all tasks, including writing, review & editing, writing the original draft, visualization, validation, supervision, methodology, investigation, formal analysis, conceptualization.

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

Ethical Approval
This study did not involve human or animal subjects, and ethical approval was not required.

Data and code availability
The finite-element solver, the physics-informed network code and the generated datasets that support the findings of this study are available from the corresponding author on reasonable request.

References
[1] M. Hetényi, Beams on Elastic Foundation: Theory with Applications in the Fields of Civil and Mechanical Engineering. Ann Arbor, MI, USA: Univ. of Michigan Press, 1946.

[2] H. Matlock and L. C. Reese, "Generalized solutions for laterally loaded piles," Journal of the Soil Mechanics and Foundations Division, ASCE, vol. 86, no. 5, pp. 63--91, 1960.

[3] B. B. Broms, "Lateral resistance of piles in cohesionless soils," Journal of the Soil Mechanics and Foundations Division, ASCE, vol. 90, no. 3, pp. 123--156, 1964.

[4] B. B. Broms, "Lateral resistance of piles in cohesive soils," Journal of the Soil Mechanics and Foundations Division, ASCE, vol. 90, no. 2, pp. 27--63, 1964.

[5] H. G. Poulos, "Behaviour of laterally loaded piles: I -- Single piles," Journal of the Soil Mechanics and Foundations Division, ASCE, vol. 97, no. 5, pp. 711--731, 1971.

[6] P. K. Banerjee and T. G. Davies, "The behaviour of axially and laterally loaded single piles embedded in nonhomogeneous soils," Géotechnique, vol. 28, no. 3, pp. 309--326, 1978.

[7] M. F. Randolph, "The response of flexible piles to lateral loading," Géotechnique, vol. 31, no. 2, pp. 247--259, 1981.

[8] K. Sun, "Laterally loaded piles in elastic media," Journal of Geotechnical Engineering, vol. 120, no. 8, pp. 1324--1344, 1994.

[9] W. D. Guo and F. H. Lee, "Load transfer approach for laterally loaded piles," International Journal for Numerical and Analytical Methods in Geomechanics, vol. 25, no. 11, pp. 1101--1129, 2001.

[10] D. Basu and R. Salgado, "Elastic analysis of laterally loaded pile in multi-layered soil," Geomechanics and Geoengineering, vol. 2, no. 3, pp. 183--196, 2007.

[11] D. Basu and R. Salgado, "Analysis of laterally loaded piles with rectangular cross sections embedded in layered soil," International Journal for Numerical and Analytical Methods in Geomechanics, vol. 32, no. 7, pp. 721--744, 2008.

[12] D. Basu, R. Salgado, and M. Prezzi, "A continuum-based model for analysis of laterally loaded piles in layered soils," Géotechnique, vol. 59, no. 2, pp. 127--140, 2009.

[13] Y. S. Choi, D. Basu, R. Salgado, and M. Prezzi, "Response of laterally loaded rectangular and circular piles in soils with properties varying with depth," Journal of Geotechnical and Geoenvironmental Engineering, vol. 140, Art. no. 04013049, 2014.

[14] L. Zhang, H. Ernst, and H. H. Einstein, "Nonlinear analysis of laterally loaded rock-socketed shafts," Journal of Geotechnical and Geoenvironmental Engineering, vol. 126, no. 11, pp. 955--968, 2000.

[15] A. T. C. Goh, "Nonlinear modelling in geotechnical engineering using neural networks," Transactions of the Institution of Engineers, Australia: Civil Engineering, vol. 36, no. 4, pp. 293--297, 1994.

[16] C. Lee and R. Sterling, "Identifying probable failure modes for underground openings using a neural network," International Journal of Rock Mechanics and Mining Sciences and Geomechanics Abstracts, vol. 29, no. 1, pp. 49--67, 1992.

[17] N. Sivakugan, J. D. Eckersley, and H. Li, "Settlement predictions using neural networks," Austral. Civil Eng. Trans., vol. CE40, pp. 49--52, 1998.

[18] M. A. Shahin, H. R. Maier, and M. B. Jaksa, "Predicting settlements of shallow foundations using neural networks," Journal of Geotechnical and Geoenvironmental Engineering, vol. 128, no. 9, pp. 785--793, 2002.

[19] J. Ghaboussi and D. E. Sidarta, "New nested adaptive neural networks (NANN) for constitutive modeling," Computers and Geotechnics, vol. 22, no. 1, pp. 29--52, 1998.

[20] C. H. Juang and C. J. Chen, "CPT-based liquefaction evaluation using artificial neural networks," Computer-Aided Civil and Infrastructure Engineering, vol. 14, no. 3, pp. 221--229, 1999.

[21] S. Rauter and F. Tschuchnigg, "CPT data interpretation employing different machine learning techniques," Geosciences, vol. 11, no. 7, Art. no. 265, 2021.

[22] S. K. Das and P. K. Basudhar, "Undrained lateral load capacity of piles in clay using artificial neural network," Computers and Geotechnics, vol. 33, no. 8, pp. 454--459, 2006.

[23] S. Kostić, N. Vasović, K. Todorović, and A. Samčović, "Application of artificial neural networks for slope stability analysis in geotechnical practice," in Proc. 13th Symp. Neural Netw. Appl. (NEUREL), 2016, pp. 89--94.

[24] M. A. Shahin, M. B. Jaksa, and H. R. Maier, "Artificial neural network applications in geotechnical engineering," Australian Geomechanics, vol. 36, no. 1, pp. 49--62, 2001.

[25] M. A. Shahin, M. B. Jaksa, and H. R. Maier, "State of the art of artificial neural networks in geotechnical engineering," Electron. J. Geotech. Eng., vol. 8, pp. 1--26, 2008.

[26] E. Yaghoubi, E. Yaghoubi, A. Khamees, and A. H. Vakili, "A systematic review and meta-analysis of artificial neural network, machine learning, deep learning, and ensemble learning approaches in the field of geotechnical engineering," Neural Computing and Applications, vol. 36, pp. 12655--12699, 2024.

[27] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," Neural Networks, vol. 2, no. 5, pp. 359--366, 1989.

[28] I. E. Lagaris, A. Likas, and D. I. Fotiadis, "Artificial neural networks for solving ordinary and partial differential equations," IEEE Transactions on Neural Networks, vol. 9, no. 5, pp. 987--1000, 1998.

[29] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," Journal of Computational Physics, vol. 378, pp. 686--707, 2019.

[30] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, "Physics-informed machine learning," Nature Reviews Physics, vol. 3, no. 6, pp. 422--440, 2021.

[31] S. Cuomo, V. S. Di Cola, F. Giampaolo, G. Rozza, M. Raissi, and F. Piccialli, "Scientific machine learning through physics-informed neural networks: Where we are and what's next," Journal of Scientific Computing, vol. 92, Art. no. 88, 2022.

[32] S. A. Faroughi, N. M. Pawar, C. Fernandes, M. Raissi, S. Das, N. K. Kalantari, and S. Kourosh Mahjour, "Physics-guided, physics-informed, and physics-encoded neural networks and operators in scientific computing: Fluid and solid mechanics," Journal of Computing and Information Science in Engineering, vol. 24, Art. no. 040802, 2024.

[33] E. Samaniego, C. Anitescu, S. Goswami, V. M. Nguyen-Thanh, H. Guo, K. Hamdia, X. Zhuang, and T. Rabczuk, "An energy approach to the solution of partial differential equations in computational mechanics via machine learning: Concepts, implementation and applications," Computer Methods in Applied Mechanics and Engineering, vol. 362, Art. no. 112790, 2020.

[34] J. Bai, Z. Lin, Y. Wang, J. Wen, Y. Liu, T. Rabczuk, Y. Gu, and X.-Q. Feng, "Energy-based physics-informed neural network for frictionless contact problems under large deformation," Computer Methods in Applied Mechanics and Engineering, vol. 437, Art. no. 117787, 2025.

[35] A. D. Jagtap, E. Kharazmi, and G. E. Karniadakis, "Conservative physics-informed neural networks on discrete domains for conservation laws: Applications to forward and inverse problems," Computer Methods in Applied Mechanics and Engineering, vol. 365, Art. no. 113028, 2020.

[36] B. Moseley, A. Markham, and T. Nissen-Meyer, "Finite basis physics-informed neural networks (FBPINNs): A scalable domain decomposition approach for solving differential equations," Advances in Computational Mathematics, vol. 49, no. 4, Art. no. 62, 2023.

[37] L. Lu, P. Jin, G. Pang, Z. Zhang, and G. E. Karniadakis, "Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators," Nature Machine Intelligence, vol. 3, no. 3, pp. 218--229, 2021.

[38] E. Haghighat, M. Raissi, A. Moure, H. Gomez, and R. Juanes, "A physics-informed deep learning framework for inversion and surrogate modeling in solid mechanics," Computer Methods in Applied Mechanics and Engineering, vol. 379, Art. no. 113741, 2021.

[39] C. Rao, H. Sun, and Y. Liu, "Physics-informed deep learning for computational elastodynamics without labeled data," Journal of Engineering Mechanics, vol. 147, no. 8, Art. no. 04021043, 2021.

[40] L. Sun, H. Gao, S. Pan, and J.-X. Wang, "Surrogate modeling for fluid flows based on physics-constrained deep learning without simulation data," Computer Methods in Applied Mechanics and Engineering, vol. 361, Art. no. 112732, 2020.

[41] X. Jin, S. Cai, H. Li, and G. E. Karniadakis, "NSFnets (Navier--Stokes flow nets): Physics-informed neural networks for the incompressible Navier--Stokes equations," Journal of Computational Physics, vol. 426, Art. no. 109951, 2021.

[42] G. Kissas, Y. Yang, E. Hwuang, W. R. Witschey, J. A. Detre, and P. Perdikaris, "Machine learning in cardiovascular flows modeling: Predicting arterial blood pressure from non-invasive 4D flow MRI data using physics-informed neural networks," Computer Methods in Applied Mechanics and Engineering, vol. 358, Art. no. 112623, 2020.

[43] M. Raissi, A. Yazdani, and G. E. Karniadakis, "Hidden fluid mechanics: Learning velocity and pressure fields from flow visualizations," Science, vol. 367, no. 6481, pp. 1026--1030, 2020.

[44] S. Cai, Z. Mao, Z. Wang, M. Yin, and G. E. Karniadakis, "Physics-informed neural networks (PINNs) for fluid mechanics: A review," Acta Mechanica Sinica, vol. 37, no. 12, pp. 1727--1738, 2021.

[45] J. Bai, H. Jeong, C. P. Batuwatta-Gamage, S. Xiao, Q. Wang, C. M. Rathnayaka, L. Alzubaidi, G.-R. Liu, and Y. Gu, "An introduction to programming physics-informed neural network-based computational solid mechanics," International Journal of Computational Methods, vol. 20, no. 10, Art. no. 2350013, 2023.

[46] S. L. Brunton and J. N. Kutz, "Promising directions of machine learning for partial differential equations," Nature Computational Science, vol. 4, no. 7, pp. 483--494, 2024.

[47] S. Wang, X. Yu, and P. Perdikaris, "When and why PINNs fail to train: A neural tangent kernel perspective," Journal of Computational Physics, vol. 449, Art. no. 110768, 2022.

[48] S. Wang, Y. Teng, and P. Perdikaris, "Understanding and mitigating gradient flow pathologies in physics-informed neural networks," SIAM Journal on Scientific Computing, vol. 43, no. 5, pp. A3055–A3081, 2021.

[49] L. D. McClenny and U. M. Braga-Neto, "Self-adaptive physics-informed neural networks," Journal of Computational Physics, vol. 474, Art. no. 111722, 2023.

[50] L. Lu, X. Meng, Z. Mao, and G. E. Karniadakis, "DeepXDE: A deep learning library for solving differential equations," SIAM Review, vol. 63, no. 1, pp. 208–228, 2021.

[51] C. Wu, M. Zhu, Q. Tan, Y. Kartha, and L. Lu, "A comprehensive study of non-adaptive and residual-based adaptive sampling for physics-informed neural networks," Computer Methods in Applied Mechanics and Engineering, vol. 403, Art. no. 115671, 2023.

[52] A. S. Krishnapriyan, A. Gholami, S. Zhe, R. M. Kirby, and M. W. Mahoney, "Characterizing possible failure modes in physics-informed neural networks," in Adv. Neural Inf. Process. Syst., 2021, vol. 34, pp. 26548–26560.

[53] N. Rahaman, A. Baratin, D. Arpit, F. Draxler, M. Lin, F. Hamprecht, Y. Bengio, and A. Courville, "On the spectral bias of neural networks," in Proc. 36th Int. Conf. Mach. Learn. (ICML), PMLR, vol. 97, 2019, pp. 5301–5310.

[54] Z.-Q. J. Xu, Y. Zhang, T. Luo, Y. Xiao, and Z. Ma, "Frequency principle: Fourier analysis sheds light on deep neural networks," Communications in Computational Physics, vol. 28, no. 5, pp. 1746–1767, 2020.

[55] M. Tancik, P. P. Srinivasan, B. Mildenhall, S. Fridovich-Keil, N. Raghavan, U. Singhal, R. Ramamoorthi, J. T. Barron, and R. Ng, "Fourier features let networks learn high frequency functions in low dimensional domains," in Adv. Neural Inf. Process. Syst., 2020, vol. 33, pp. 7537–7547.

[56] S. Wang, H. Wang, and P. Perdikaris, "On the eigenvector bias of Fourier feature networks: From regression to solving multi-scale PDEs with physics-informed neural networks," Computer Methods in Applied Mechanics and Engineering, vol. 384, Art. no. 113938, 2021.

[57] L. Mandl, A. Mielke, S. M. Seyedpour, and T. Ricken, "Affine transformations accelerate the training of physics-informed neural networks of a one-dimensional consolidation problem," Scientific Reports, vol. 13, no. 1, Art. no. 15566, 2023.

[58] Y. W. Bekele, "Physics-informed deep learning for one-dimensional consolidation," Journal of Rock Mechanics and Geotechnical Engineering, vol. 13, no. 2, pp. 420--430, 2021.

[59] M. Eghbalian, M. Pouragha, and R. Wan, "A physics-informed deep neural network for surrogate modeling in classical elasto-plasticity," Computers and Geotechnics, vol. 159, Art. no. 105472, 2023.

[60] H. Guo and Z.-Y. Yin, "A novel physics-informed deep learning strategy with local time-updating discrete scheme for multi-dimensional forward and inverse consolidation problems," Computer Methods in Applied Mechanics and Engineering, vol. 421, Art. no. 116819, 2024.

[61] K.-Q. Li, Z.-Y. Yin, N. Zhang, and J. Li, "A PINN-based modelling approach for hydromechanical behaviour of unsaturated expansive soils," Computers and Geotechnics, vol. 169, Art. no. 106174, 2024.

[62] K.-Q. Li, Z.-Y. Yin, J.-L. Qi, and Y. Liu, "State-of-the-art constitutive modelling of frozen soils," Archives of Computational Methods in Engineering, vol. 31, pp. 3801--3842, 2024.

[63] X. Chen, P. Zhang, and Z.-Y. Yin, "Physics-informed neural network solver for numerical analysis in geoengineering," Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards, vol. 18, no. 1, pp. 33--51, 2024.

[64] S.-Y. He, W.-H. Zhou, and C. Tang, "Physics-informed neural networks for settlement analysis of the immersed tunnel of the Hong Kong--Zhuhai--Macau bridge," International Journal of Geomechanics, vol. 24, no. 1, 2024.

[65] C. Liu, J. E. Macedo, and A. Rodríguez, "Leveraging physics-informed neural networks in geotechnical earthquake engineering: An assessment on seismic site response analyses," Computers and Geotechnics, vol. 182, Art. no. 107137, 2025.

[66] M. Vahab, B. Shahbodagh, E. Haghighat, and N. Khalili, "Application of physics-informed neural networks for forward and inverse analysis of pile–soil interaction," International Journal of Solids and Structures, vol. 277–278, Art. no. 112319, 2023.

[67] W. Ouyang, G. Li, L. Chen, and S.-W. Liu, "Machine learning-based soil--structure interaction analysis of laterally loaded piles through physics-informed neural networks," Acta Geotechnica, vol. 19, no. 7, pp. 4765--4790, 2024.

[68] W. Ouyang, G. Li, L. Chen, and S.-W. Liu, "Physics-informed neural networks for large deflection analysis of slender piles incorporating non-differentiable soil--structure interaction," International Journal for Numerical and Analytical Methods in Geomechanics, vol. 48, no. 6, pp. 1802--1823, 2024. doi: 10.1002/nag.3679.

[69] M. A. Biot, "Bending of an infinite beam on an elastic foundation," Journal of Applied Mechanics, vol. 4, no. 1, pp. A1--A7, 1937.

[70] L. C. Reese, W. R. Cox, and F. D. Koop, "Analysis of laterally loaded piles in sand," in Proc. 6th Offshore Technol. Conf., Houston, TX, USA, 1974, Paper OTC 2080, pp. 473--483.

[71] M. T. Davisson and H. L. Gill, "Laterally loaded piles in a layered soil system," Journal of the Soil Mechanics and Foundations Division, ASCE, vol. 89, no. 3, pp. 63--94, 1963.

[72] K. Terzaghi, "Evaluation of coefficients of subgrade reaction," Géotechnique, vol. 5, no. 4, pp. 297--326, 1955.

[73] American Petroleum Institute, "Recommended practice for planning, designing and constructing fixed offshore platforms -- working stress design," API, Washington, DC, USA, Rep. RP 2A-WSD, 2011.

[74] W. D. Guo, "Laterally loaded rigid piles with rotational constraints," Computers and Geotechnics, vol. 54, pp. 72--83, 2013.

[75] H. G. Poulos and E. H. Davis, Pile Foundation Analysis and Design. New York, NY, USA: Wiley, 1980.

[76] L. C. Reese and W. F. Van Impe, Single Piles and Pile Groups Under Lateral Loading, 2nd ed. London, U.K.: CRC Press, 2011.

[77] W. Ouyang, S.-W. Liu, J. Wan, and Y. Yang, "Euler--Bernoulli pile element for nonlinear buckling analysis of single piles in slope," International Journal of Geomechanics, vol. 21, no. 9, Art. no. 04021170, 2021.

[78] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in Proc. 13th Int. Conf. Artif. Intell. Statist. (AISTATS), 2010, vol. 9, pp. 249--256.

[79] A. G. Baydin, B. A. Pearlmutter, A. A. Radul, and J. M. Siskind, "Automatic differentiation in machine learning: A survey," Journal of Machine Learning Research, vol. 18, pp. 1--43, 2018.

[80] J. Bradbury, R. Frostig, P. Hawkins, M. J. Johnson, C. Leary, D. Maclaurin, G. Necula, A. Paszke, J. VanderPlas, S. Wanderman-Milne, and Q. Zhang, "JAX: Composable transformations of Python+NumPy programs," 2018. [Online]. Available: https://github.com/jax-ml/jax

[81] DeepMind et al., "The DeepMind JAX ecosystem," 2020. [Online]. Available: https://github.com/google-deepmind

[82] A. Paszke et al., "PyTorch: An imperative style, high-performance deep learning library," in Adv. Neural Inf. Process. Syst., 2019, vol. 32, pp. 8024--8035.

[83] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, arXiv:1412.6980.

[84] D. C. Liu and J. Nocedal, "On the limited memory BFGS method for large scale optimization," Mathematical Programming, vol. 45, no. 1--3, pp. 503--528, 1989.

[85] J. Nocedal and S. J. Wright, Numerical Optimization, 2nd ed. New York, NY, USA: Springer, 2006.

[86] B. Efron and R. J. Tibshirani, An Introduction to the Bootstrap. New York, NY, USA: Chapman & Hall, 1993.

[87] H. B. Mann and D. R. Whitney, "On a test of whether one of two random variables is stochastically larger than the other," Annals of Mathematical Statistics, vol. 18, no. 1, pp. 50–60, 1947.

[88] J. Dunnicliff, Geotechnical Instrumentation for Monitoring Field Performance. New York, NY, USA: Wiley, 1988.

[89] P. E. Mikkelsen, "Advances in inclinometer data analysis," in Proc. 6th Int. Symp. Field Measurements in Geomechanics (FMGM), Oslo, Norway, 2003.

[90] T. D. Stark and H. Choi, "Slope inclinometers for landslides," Landslides, vol. 5, no. 3, pp. 339–350, 2008.

[91] V. Sitzmann, J. N. P. Martel, A. W. Bergman, D. B. Lindell, and G. Wetzstein, "Implicit neural representations with periodic activation functions," in Adv. Neural Inf. Process. Syst., 2020, vol. 33, pp. 7462–7473.

[92] S. Elfwing, E. Uchibe, and K. Doya, "Sigmoid-weighted linear units for neural network function approximation in reinforcement learning," Neural Networks, vol. 107, pp. 3–11, 2018.

[93] L. C. Reese and W. R. Cox, "Soil behavior from analysis of tests of uninstrumented piles under lateral loading," in Performance of Deep Foundations, ASTM STP 444. West Conshohocken, PA, USA: ASTM, 1969, pp. 160--176.

[94] S. Ito, R. Fukunaga, and K. Sako, "Inverse analysis for estimating geotechnical parameters using physics-informed neural networks," Soils and Foundations, vol. 64, no. 6, Art. no. 101533, 2024.

[95] J. Cheng, K. Kumar, and M. J. DeJong, "Structural identification of material properties and load parameters using a differentiable material point method," Computer-Aided Civil and Infrastructure Engineering, vol. 40, no. 24, pp. 3856--3874, 2025.

[96] Y. J. Choi and K. Kumar, "Inverse analysis of granular flows using differentiable graph neural network simulator," Computers and Geotechnics, vol. 171, Art. no. 106374, 2024.

[97] A. Ledesma, A. Gens, and E. E. Alonso, "Estimation of parameters in geotechnical backanalysis — I. Maximum likelihood approach," Computers and Geotechnics, vol. 18, no. 1, pp. 1–27, 1996.

[98] R. Bellman and K. J. Åström, "On structural identifiability," Mathematical Biosciences, vol. 7, no. 3–4, pp. 329–339, 1970.

[99] A. Raue, C. Kreutz, T. Maiwald, J. Bachmann, M. Schilling, U. Klingmüller, and J. Timmer, "Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood," Bioinformatics, vol. 25, no. 15, pp. 1923–1929, 2009.

[100] R. N. Gutenkunst, J. J. Waterfall, F. P. Casey, K. S. Brown, C. R. Myers, and J. P. Sethna, "Universally sloppy parameter sensitivities in systems biology models," PLoS Computational Biology, vol. 3, no. 10, Art. no. e189, 2007.

[101] R. Chartrand, "Numerical differentiation of noisy, nonsmooth data," ISRN Applied Mathematics, vol. 2011, Art. no. 164564, 2011.

[102] C. H. Reinsch, "Smoothing by spline functions," Numerische Mathematik, vol. 10, no. 3, pp. 177–183, 1967.

[103] P. Craven and G. Wahba, "Smoothing noisy data with spline functions," Numerische Mathematik, vol. 31, no. 4, pp. 377–403, 1979.

[104] A. N. Tikhonov and V. Y. Arsenin, Solutions of Ill-Posed Problems. Washington, DC, USA: Winston & Sons, 1977.

[105] D. W. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," Journal of the Society for Industrial and Applied Mathematics, vol. 11, no. 2, pp. 431–441, 1963.

[106] H. Matlock, "Correlations for design of laterally loaded piles in soft clay," in Proc. 2nd Offshore Technol. Conf., Houston, TX, USA, 1970, Paper OTC 1204, pp. 577--594.

[107] J. H. Long and G. Vanneste, "Effects of cyclic lateral loads on piles in sand," Journal of Geotechnical Engineering, vol. 120, no. 1, pp. 225–244, 1994.

[108] C. LeBlanc, G. T. Houlsby, and B. W. Byrne, "Response of stiff piles in sand to long-term cyclic lateral loading," Géotechnique, vol. 60, no. 2, pp. 79–90, 2010.

[109] D. A. Brown and C.-F. Shie, "Three dimensional finite element model of laterally loaded piles," Computers and Geotechnics, vol. 10, no. 1, pp. 59–79, 1990.

[110] M. Ashour and G. Norris, "Modeling lateral soil–pile response based on soil–pile interaction," Journal of Geotechnical and Geoenvironmental Engineering, vol. 126, no. 5, pp. 420–428, 2000.

[111] B. W. Byrne, G. T. Houlsby, H. J. Burd, K. G. Gavin, D. J. P. Igoe, R. J. Jardine, C. M. Martin, R. A. McAdam, D. M. Potts, D. M. G. Taborda, and L. Zdravković, "PISA design model for monopiles for offshore wind turbines: Application to a stiff glacial clay till," Géotechnique, vol. 70, no. 11, pp. 1030–1047, 2020.

[112] A. D. Jagtap, K. Kawaguchi, and G. E. Karniadakis, "Adaptive activation functions accelerate convergence in deep and physics-informed neural networks," Journal of Computational Physics, vol. 404, Art. no. 109136, 2020.