Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

# Temporal convolutional network-based parameter inversion, deformation forecasting and dual-track early warning for a super- large excavation in water-rich sandy strata

**Shengyuan Qian**

qianshy@jsou.edu.cn

Jiangsu Open University

**Xiaodong Ni**

Hohai University

**Meng Kong**

Jiangsu Open University

**Article**

**Keywords:** Deep excavation, Temporal convolutional network, Back analysis, Latin hypercube sampling, Modified Mohr–Coulomb model, Deformation forecasting, Seepage, Anchorage foundation pit

**Posted Date:** November 20th, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-11594914/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

**Additional Declarations:** No competing interests reported.

# Temporal convolutional network-based parameter inversion, deformation forecasting and dual-track early warning for a super-large excavation in water-rich sandy strata

**Shengyuan Qian[1*], Xiaodong Ni[2] & Meng Kong[1]**

[1]College of Urban Construction, Jiangsu Open University, Nanjing 210036, China.
[2]College of Civil and Transportation Engineering, Hohai University, Nanjing 210024, China.
*email: qianshy@jsou.edu.cn

## Abstract

Deformation prediction for super-large excavations in water-rich sandy strata is limited by uncertain soil parameters, while data-driven forecasts alone cannot show whether a deformation trend is abnormal. This study couples the two approaches for the Φ130 m circular excavation of the Shiziyang Bridge west anchorage. Sensitivity analysis of a modified Mohr–Coulomb model showed that the effective friction angle and unloading–reloading reference modulus of the silty clay and fine-sand layers control the deformation. A temporal convolutional network (TCN), trained on 50 Latin hypercube finite-element samples, mapped nine measured responses to these six parameters with a cross-validated mean absolute percentage error (MAPE) of 5.0%, 24% lower than that of the best of five baseline regressors. Calibration reduced the MAPE of the simulated histories from 21.0% to 3.0% for wall displacement and from 29.2% to 5.8% for ground settlement. A second TCN forecast wall displacement 14 days ahead with a MAPE of 4.60%, against 13.09% for persistence. A dual-track rule issues a warning when the two forecasts differ by more than 10% and the forecast exceeds 70% of the 30 mm control value. At nine monitoring points, the rule flagged all three sections with seepage or accelerating deformation, without false alarms.

**Keywords** Deep excavation, Temporal convolutional network, Back analysis, Latin hypercube sampling, Modified Mohr–Coulomb model, Deformation forecasting, Seepage, Anchorage foundation pit

As deep excavations grow in plan size and depth, construction disturbance and uncertainty increase, and safety incidents remain frequent[1,2]. Water-rich sandy strata are particularly demanding. Their high permeability and strong hydraulic connectivity mean that dewatering raises the head difference across the retaining system, which can trigger piping, sand boiling and basal inrush[3,4]. Dissipation of pore pressure also produces large ground settlement and additional deflection of the retaining structure[5]. Minor defects in the waterproof curtain can

develop into through-wall leakage within weeks, and the early signs of such failures are often recorded but not acted on in time[6]. For super-large anchorage pits, whose excavation spans several months and whose deformation control values are strict, safe construction therefore depends on the observational method: deformation is predicted, measured, compared and acted on stage by stage[7].

Numerical simulation is the main tool for the quantitative assessment of deep excavations and their interaction with adjacent underground structures. Continuum and discrete models have been used to evaluate excavation-induced effects on existing tunnels[8], the efficiency of grouting reinforcement during shield tunnelling near operating lines[9], crack propagation around tunnels and dams[10], the creep of anchored walls[11] and the base stability of circular pits[12]. For the companion east anchorage of the Shiziyang Bridge, Zhang et al.[13] showed that the choice of constitutive model strongly affects the computed wall deflection and settlement. Whatever model is used, predictive accuracy depends on the input soil parameters[14,15]. Laboratory tests do not reproduce in situ stress paths, and borehole tests disturb the soil[16]. When laboratory-derived parameters are used directly, the discrepancy between simulated and measured deformation tends to accumulate with excavation depth[17].

Back analysis updates these parameters from field measurements. Optimisation-based approaches, using genetic algorithms[18] and single- or multi-objective formulations[14,17,19], search the parameter space by repeatedly running the forward model. Probabilistic approaches, such as Bayesian updating[20,21] and efficient probabilistic back analysis[15,16], also quantify parameter uncertainty. Both families require many forward analyses, which is costly for a three-dimensional coupled model. Machine-learning surrogates reduce this cost by learning the response–parameter mapping from a limited set of simulations; examples include back-propagation (BP) neural networks[22], optimised multi-output least-squares support vector regression[23] and multi-parameter intelligent inversion that accounts for path-dependent soil behaviour[24]. Three questions remain open. First, shallow surrogates treat the measurement points as unordered, independent features and thus discard the spatial shape of the wall-deflection profile and the settlement trough. Second, the training samples are usually generated by orthogonal or random designs whose coverage of the parameter space is rarely checked. Third, with such small sample sets, surrogate accuracy is seldom reported under cross-validation or against simpler baselines.

Data-driven forecasting of excavation deformation has advanced in parallel. Convolutional and spatiotemporal deep-learning models predict wall deflection[25,26], and hybrid models such as EMD-SSA-GRNN[27], ST-CNN-LSTM[28], CNN-LSTM-Transformer[29] and TCN-LSTM[30]

improve accuracy on monitoring series; physics-informed graph networks add mechanical constraints[31]. These developments are summarised in recent reviews[32–34]. The temporal convolutional network (TCN) uses dilated causal convolutions and residual connections, which provide parallel training, stable gradients and a receptive field that grows exponentially with depth[35,36]. Most of these models, however, are compared only with other neural networks. Monitoring series of wall displacement change slowly from day to day, so a naive forecast that repeats the last measurement is already accurate at a one-day horizon[37], and an apparent improvement among complex models may conceal the absence of real forecasting skill.

Early warning in practice relies on fixed thresholds of cumulative displacement or displacement rate, usually set as fractions of the design control value[38]. Such thresholds respond only after the deformation has become large, and a data-driven forecast used on its own cannot show whether a predicted trend is abnormal, because it carries no reference for the response expected under normal conditions. A calibrated numerical model provides exactly that reference. The disagreement between the two could therefore flag abnormal behaviour, yet it has rarely been quantified or used as a warning indicator.

This study addresses two questions: how the key soil parameters of a three-dimensional coupled model can be identified from a small number of costly simulations, and whether the disagreement between calibrated numerical and data-driven forecasts can serve as a physically interpretable indicator of abnormal deformation. For the Φ130 m circular excavation of the Shiziyang Bridge west anchorage, a TCN inverse model that convolves along the spatial order of the wall-deflection and settlement responses is trained on a verified Latin hypercube design and benchmarked against five regressors under repeated cross-validation. A TCN forecaster for daily wall displacement is then compared with naive, statistical and neural-network references at one-day and 14-day horizons. Finally, a dual-track warning rule that combines the relative deviation between the two forecasts with the absolute forecast displacement is verified against field observations at nine monitoring points.

# Methods

### Study site and ground conditions

The Shiziyang Bridge is a single-span steel truss suspension bridge with a main span of 2180 m and two circular gravity anchorages. The west anchorage foundation pit, 130 m in diameter, is the larger of the two and one of the largest anchorage pits ever constructed; it was selected as the research object. The pit is supported by a circular diaphragm wall with an outer diameter

of 130 m and a thickness of 1.5 m, combined with a ring-shaped reinforced concrete inner lining. The excavation proceeds in 12 stages of 3 m each to a maximum depth of 36.3 m and covers approximately 12,668 m² (Fig. 1).

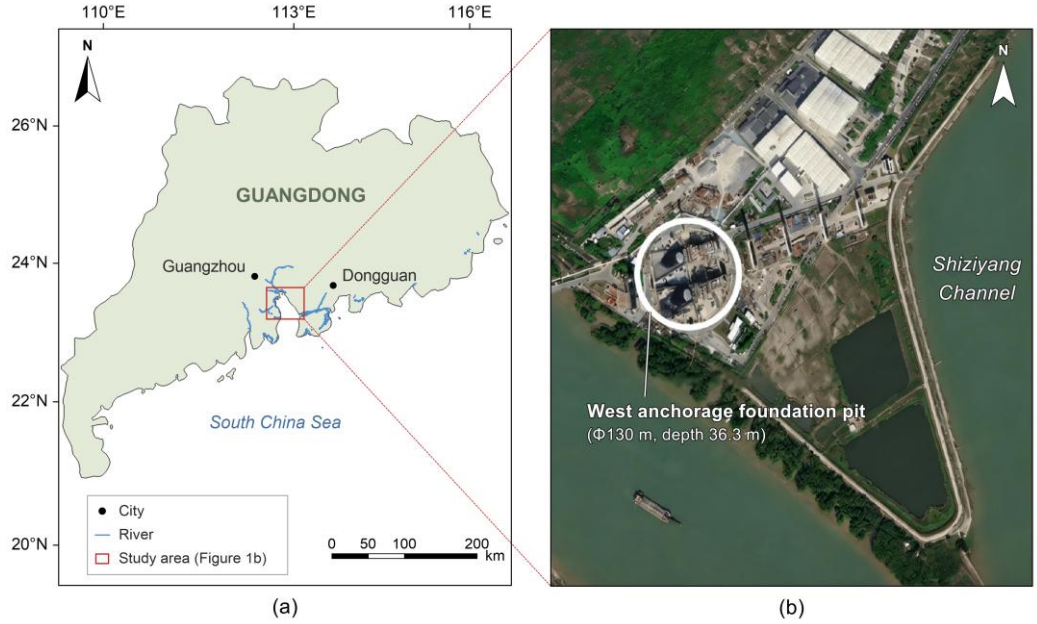



**Figure 1.** Location of the study area: (**a**) Guangdong Province and the Pearl River Estuary; (**b**) satellite image of the west anchorage site beside the Shiziyang Channel.

The subsurface consists of miscellaneous fill, an upper fine-sand layer, silty clay, a lower fine-sand layer, strongly weathered argillaceous sandstone and moderately weathered argillaceous sandstone, forming a binary structure with soft upper soils and stiff lower rock (Supplementary Figs. S1 and S2). The silty clay and the two fine-sand layers lie within the depth range of the maximum wall deflection, and the lower fine sand forms a confined aquifer above the bedrock. The site has a humid climate with abundant rainfall, and surface water, pore water and bedrock fissure water are hydraulically connected. Groundwater levels fluctuate seasonally with rainfall, evaporation and tides. In wet seasons, infiltration raises the phreatic level, increasing the lateral pressure and the seepage risk in the fine-sand layers, while recharge of the confined aquifer raises its head and reduces the safety margin against basal inrush. In dry seasons, tidal fluctuations cause transient head changes that alter the pressure difference across the waterproof curtain and may trigger piping.

## Field monitoring and control value

The deformation control value of the pit, $U_c$ = 30 mm, was specified in the project design in accordance with the Specifications for Design of Foundation of Highway Bridges and Culverts

(JTG 3363-2019), which requires that the deformation of the retaining structure does not exceed the allowable deformation of adjacent buildings and underground facilities[13,39]. Thirty-six deep horizontal displacement points along the diaphragm wall and 36 ground-surface settlement points around the pit perimeter were monitored daily (Fig. 2). The quarter-circular sector covering ZQT01–ZQT09 and DBC01–DBC09 is the primary study area. Twelve dewatering wells were arranged uniformly on a circle 4 m inside the diaphragm wall, and four piezometric observation wells were installed 5 m outside the wall to monitor the groundwater response to dewatering. The record analysed here covers 92 days, from the start of excavation to the end of stage XI; stages I–XI ended on days 4, 10, 18, 29, 38, 47, 56, 69, 78, 85 and 92, respectively.

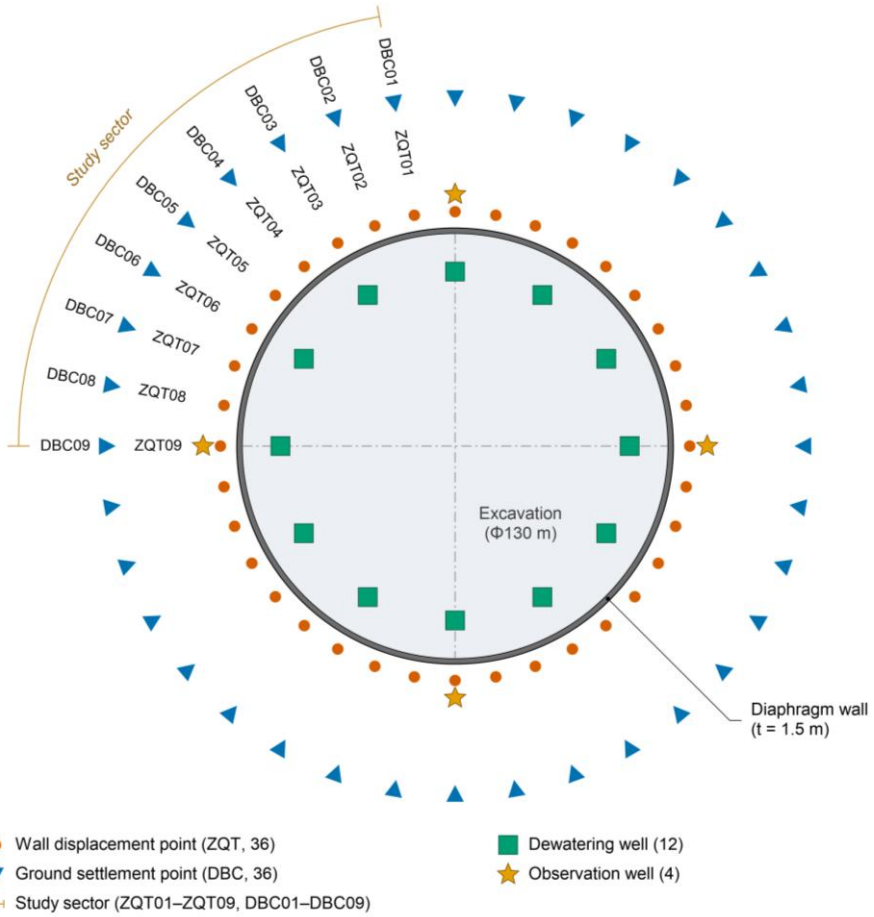



**Figure 2.** Layout of monitoring points and wells (schematic, not to scale). Monitoring points are spaced at 10°; the bracketed sector (ZQT01–ZQT09, DBC01–DBC09) is the study area.

## Overview of the TCN-driven framework

The workflow couples three-dimensional finite-element simulation with the TCN (Fig. 3). On the mechanistic track, field, geological and hydrogeological data are used to build the finite-element model, sensitivity analysis identifies the most influential soil parameters, and the TCN inverse model calibrates them from measured deformation. On the data-driven track, a second TCN forecasts deformation from the monitoring record. The two forecasts are then compared by the dual-track warning rule described below.

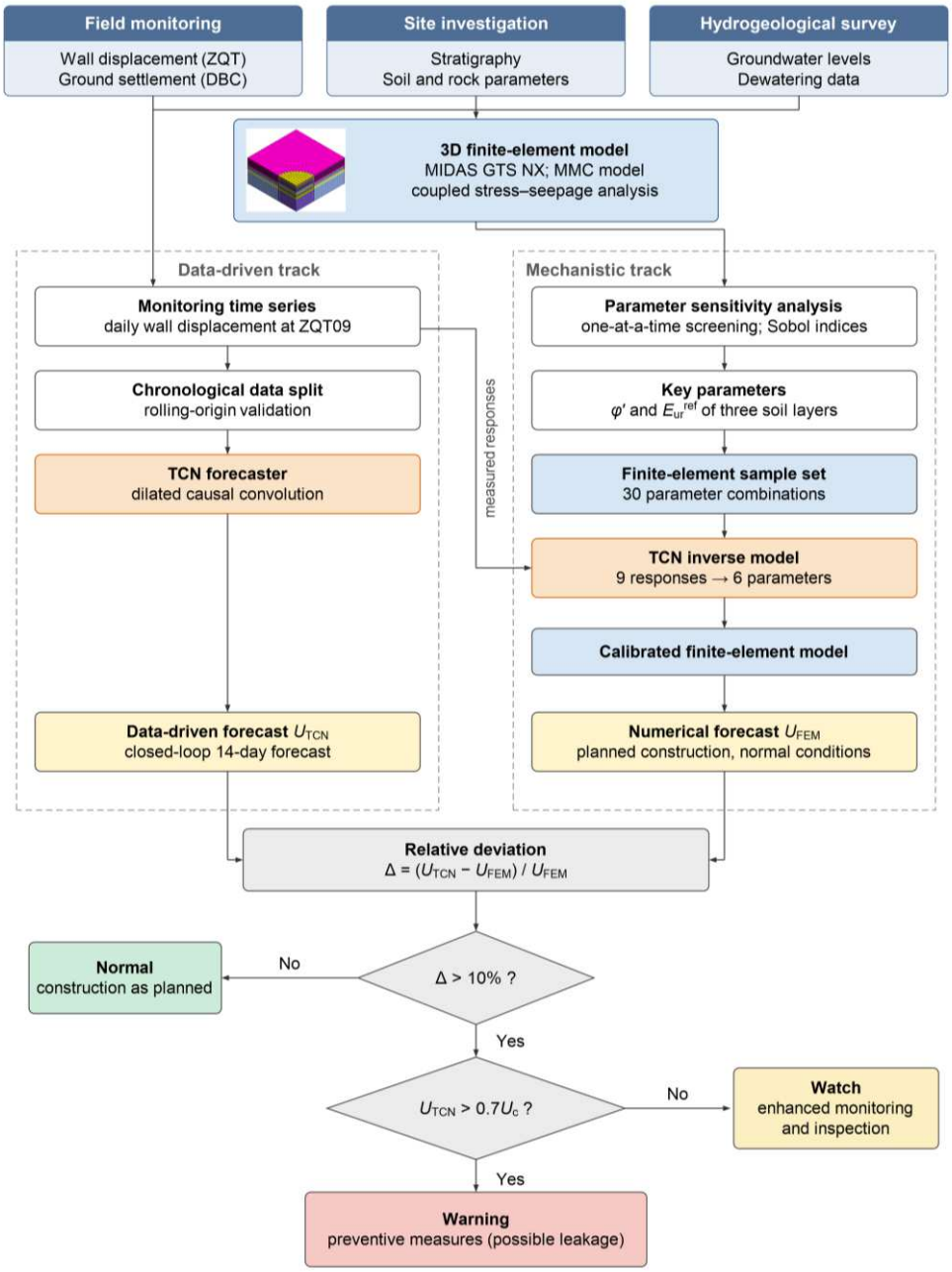



**Figure 3.** Workflow of the TCN-driven methodology: the mechanistic track calibrates the finite-element model by TCN inversion, the data-driven track forecasts deformation with a TCN, and the two forecasts are compared by the two-tier warning rule.

## Finite-element model and constitutive parameters

A three-dimensional model of the foundation pit was built in MIDAS GTS NX 2023. Owing to the circular symmetry, one quarter of the structure was modelled. The base and lateral boundaries were fixed, and the model extends 216 m horizontally (six times the excavation depth) and 108 m vertically (three times the excavation depth) to limit boundary effects. The soil, inner lining and diaphragm wall were represented by solid elements and the dewatering wells by truss elements. Interface elements reproduced the water-blocking effect of the muddy aquitard, with hydraulic heads of −1.4 m in the shallow layer and −5.79 m in the lower confined aquifer. The 12 excavation stages followed the sequence of dewatering, excavation and lining installation, and a coupled stress–seepage analysis was performed. The model (Supplementary Fig. S3) contains 24,786 elements and 28,143 nodes. Mesh convergence was checked with a coarse, the adopted and a refined mesh (12,450, 24,786 and 49,800 elements); refining the adopted mesh changed the maximum wall displacement and ground settlement at stage VIII by 0.9% and 2.4%, respectively.

The Mohr–Coulomb (MC) model is widely used because of its simplicity, but its elastic–perfectly plastic formulation uses a single modulus for loading and unloading, cannot capture nonlinear hardening and overestimates basal heave. The modified Mohr–Coulomb (MMC) model in MIDAS GTS NX introduces three independent stiffness parameters, namely the reference secant modulus $E_{50}^{\mathrm{ref}}$, the reference oedometer modulus $E_{\mathrm{oed}}^{\mathrm{ref}}$ and the unloading–reloading reference modulus $E_{\mathrm{ur}}^{\mathrm{ref}}$, and therefore describes the nonlinear response under loading and unloading and suits multi-stage excavation and dewatering. The MMC model was therefore adopted. The input parameters (Table 1) were determined from the site investigation and laboratory tests. The layer thickness $H$ and Poisson's ratio $v$ were taken from the geological data, the power exponent $m$ was set to 0.5, and $c'$ and $\varphi'$ were obtained from direct shear tests. The moduli were derived from the compression modulus $E_{\mathrm{s}}$ with the empirical correlations in Eqs. (1)–(4).

$$E_{\mathrm{oed}}^{\mathrm{ref}} \approx 0.8E_{\mathrm{s}} \quad (1)$$

$$E_{50}^{\mathrm{ref}} \approx E_{\mathrm{s}} \quad (2)$$

$$E_{\mathrm{ur}}^{\mathrm{ref}} \approx 6E_{50}^{\mathrm{ref}}\ (\text{sand}) \quad (3)$$

$$E_{\mathrm{ur}}^{\mathrm{ref}} \approx 8E_{50}^{\mathrm{ref}}\ (\text{clay}) \quad (4)$$

As in the hardening soil family[40], the unloading–reloading stiffness depends on the stress level:

$$E_{\mathrm{ur}} = E_{\mathrm{ur}}^{\mathrm{ref}}\left(\frac{c'\cos\varphi' + \sigma_3'\sin\varphi'}{c'\cos\varphi' + p^{\mathrm{ref}}\sin\varphi'}\right)^{m} \quad (5)$$

where $\sigma'_3$ is the minor effective principal stress and $p^{\mathrm{ref}}$ = 100 kPa.

**Table 1.** Soil parameters of the MMC model.

| Soil layer | $H$ (m) | $c'$ (kPa) | $\varphi'$ (°) | $E_{50}^{ref}$ (MPa) | $E_{oed}^{ref}$ (MPa) | $E_{ur}^{ref}$ (MPa) |
|---|---|---|---|---|---|---|
| Miscellaneous fill | 6.40 | 7.8 | 18.4 | 4.4 | 3.52 | 26.4 |
| Fine sand (upper) | 4.60 | 0 | 29.1 | 8.6 | 6.88 | 51.6 |
| Silty clay | 9.50 | 5.9 | 26.4 | 12.0 | 9.60 | 96.0 |
| Fine sand (lower) | 2.90 | 0 | 31.7 | 11.3 | 9.04 | 67.8 |

*H*, layer thickness; *c′*, effective cohesion; *φ′*, effective friction angle. The power exponent $m = 0.5$ for all layers; Poisson's ratio and unit weight are given in Supplementary Fig. S2.

Gravity loads were applied first, and the nodal displacements were reset to zero after equilibrium of the initial stress field. After the retaining structure and waterproof curtain were installed, a 20 kPa surcharge was applied to represent construction machinery. Dewatering and staged excavation were then simulated in sequence; each analysis step comprised three increments of 1 day, corresponding to 3 days of construction.

## Architecture of the TCN

The TCN is a convolutional network for sequence data[35]. As shown in Fig. 4, it consists of stacked residual modules, each containing two dilated convolutional layers with weight normalisation, ReLU activation and dropout, and a residual connection that prevents gradient degradation. For an input sequence $x$ and a filter $f$ of size $k$, the dilated convolution at position $s$ is

$$F(s) = \sum_{i=0}^{k-1} f(i) \cdot x_{s-d \cdot i} \tag{6}$$

where $d$ is the dilation factor. Each residual module outputs

$$o = \mathrm{ReLU}\big(x + \mathcal{F}(x)\big) \tag{7}$$

where $\mathcal{F}$ denotes the stacked convolution, normalisation, activation and dropout operations; a $1 \times 1$ convolution matches the channel dimensions of $x$ and $\mathcal{F}(x)$ when they differ. With the dilation doubling from one module to the next ($d_b = 2^{b-1}$, $b = 1, \ldots, B$) and two convolutions per module, the receptive field is

$$R = 1 + 2(k-1)\sum_{b=1}^{B} 2^{b-1} = 1 + 2(k-1)(2^B - 1) \tag{8}$$

so the context available to each output grows exponentially with depth, while the number of trainable weights grows only linearly. The two TCN models in this study share this backbone but differ in the axis along which the convolution operates, as described in the next two subsections.

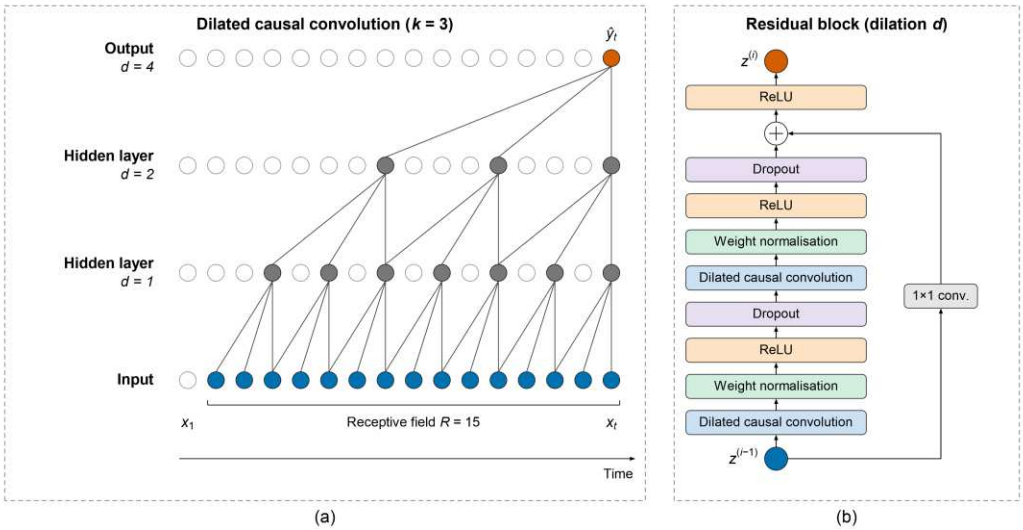



**Figure 4.** Basic architecture of the TCN: (**a**) dilated causal convolution with kernel size $k = 3$ and dilations $d = 1$, 2 and 4, showing the inputs on which the output at time $t$ depends (receptive field $R = 15$, Eq. (8)); (**b**) residual block, Eq. (7).

## TCN inverse model

The inverse problem is cast as a regression from the response vector $\mathbf{u} = [w(5), w(10), w(15), w(20), w(25), w(30), s(10), s(15), s(20)]$ to the parameter vector $\boldsymbol{\theta}$, which contains $\varphi'$ and $E_{\mathrm{ur}}^{\mathrm{ref}}$ of the upper fine sand, the silty clay and the lower fine sand. Here $w(z)$ is the wall deflection at depth $z$ (m) at ZQT09 and $s(x)$ is the ground settlement at distance $x$ (m) from the pit edge at DBC09, both at the end of stage VIII.

Strictly, this input is not a time series, and the TCN is used here as a dilated one-dimensional residual convolutional network. The choice rests on the structure of the data. The nine responses are not exchangeable features but ordered samples of two spatially continuous fields, the deflection profile $w(z)$ and the settlement trough $s(x)$. The information that distinguishes one parameter set from another lies largely in the shape of these fields, for example the depth and sharpness of the deflection bulge, which a fully connected network must learn from scratch. The vector $\mathbf{u}$ is therefore treated as a one-channel sequence of length nine and convolved along the spatial index, with non-causal ("same") padding because spatial order has no causal direction. With $k = 3$, $B = 2$ and dilations of 1 and 2, Eq. (8) gives $R = 13 \geq 9$, so each output feature aggregates the whole response vector. Weight sharing limits the network to 20,229 trainable parameters, which reduces the risk of over-fitting a training set of 40 samples. Whether this inductive bias is beneficial is an empirical question, which is answered in the Results by comparing the TCN with five simpler regressors under identical cross-validation.

The network uses 16 filters with a kernel size of 3, a dropout rate of 0.05, two residual modules, a global average pooling layer and a six-node fully connected output layer. Inputs and outputs were standardised with the statistics of the training fold only, and the network was trained with the Adam optimiser and a mean-squared-error loss (280 epochs with early stopping, initial learning rate 0.001). The training samples were generated by Latin hypercube sampling (LHS)[41], which divides the range of each parameter into as many equal-probability intervals as there are samples and draws one value from each interval, so that every marginal range is covered with few samples.

Because the uncalibrated model underestimated the measured deformation (see Results), indicating overly stiff parameters, the sampling ranges were shifted towards weaker values. The ranges of $\varphi'$ were [21.1°, 29.1°], [18.4°, 26.4°] and [23.7°, 31.7°] for the upper fine sand, the silty clay and the lower fine sand, respectively. The ranges of $E_{\mathrm{ur}}^{\mathrm{ref}}$ were 4–6 times $E_{50}^{\mathrm{ref}}$ for the sandy layers and 4–8 times $E_{50}^{\mathrm{ref}}$ for the clay, the upper bounds corresponding to Eqs. (3) and (4). An LHS design of 50 combinations was generated within these ranges (Supplementary Fig. S4). Each combination was analysed with the finite-element model, and the six wall displacements at ZQT09 and three settlements at DBC09 at the end of stage VIII formed the sample set (Supplementary Table S1). A single coupled analysis requires about 2 h, so the size of the design is a compromise between computational cost and coverage of the six-dimensional parameter space.

*Adequacy of the sample set.* Four properties of the design indicate that 50 samples cover the inversion space adequately. First, coverage: projected onto any single parameter, the samples occupy 48–50 of the 50 equal-width intervals and span 96.4–99.8% of the prescribed range, so every part of each range is represented (Supplementary Fig. S4). Second, independence: the largest pairwise correlation between parameters is $|r| = 0.27$ and the mean is 0.11, both below the critical value of 0.28 for 50 samples at the 5% significance level, so the effect of one parameter is not confounded with that of another. Third, density: 50 samples for six parameters is close to the 10 samples per input recommended for surrogate modelling of computer experiments[45], and the minimum normalised distance between any two samples, 0.30, shows that no two samples are near-duplicates. Fourth, the learning curve reported in the Results flattens beyond 30 training samples, and the inverted parameters lie inside the sampled ranges rather than at their bounds, so the design neither limits the accuracy nor constrains the solution.

## TCN forecasting model

The time-series model uses the daily record of lateral wall displacement at ZQT09, which covers 92 days from the start of excavation to the end of stage XI. Days 1–78 (stages I–IX) were used for model development, with days 1–62 for training and days 63–78 as a chronologically held-out test set; no shuffling was applied. Days 79–92 (stages X and XI) were not used in development. Two quantities are distinguished over days 79–92. The rolling one-day-ahead prediction is updated each day with the newly measured value and therefore carries a lead time of one day. The closed-loop 14-day forecast is issued once on day 78, each predicted value being fed back as the next input and no measurement after day 78 being used; only this quantity carries a 14-day lead time.

Hyperparameters were selected by rolling-origin validation within the training window[42], using four expanding-window folds validated on days 43–47, 48–52, 53–57 and 58–62; the configuration with the lowest mean validation error was retained, and the test set was evaluated once. Normalisation statistics were computed from the training window only.

The forecasting lag is one day: the value at day $t + 1$ is predicted from the series up to day $t$. This lag is not the length of the input window. The network is trained in sequence-to-sequence mode on the whole training series, and its causal convolutions give each prediction access to all past values within the receptive field. With $k = 3$, Eq. (8) gives $R = 29$, 61 and 125 days for $B = 3$, 4 and 5; with $B = 5$, $R = 125$ days, which covers the development record. To test whether this history is exploited, the TCN was also trained in a sliding-window mode with explicit look-back windows of $L = 1$, 3, 7 and 14 days. The remaining settings were 16 convolutional filters, a kernel size of 3, the Adam optimiser, 150 epochs, an initial learning rate of 0.01 with piecewise decay, a gradient threshold of 1 and L2 regularisation of 0.001.

## Dual-track warning rule

The two forecasts carry different information. The TCN forecast is driven by the monitoring record, so localised anomalies such as seepage, which leave a signature in the deformation series, are reflected in it. The numerical forecast, made with the calibrated parameters and the planned construction sequence, represents the expected response under normal conditions and cannot include unforeseen events. The TCN does not predict leakage as such; it extrapolates the trend present in the record. A forecast that departs from the numerical baseline therefore indicates that the observed trend is no longer explained by the calibrated mechanics. The relative deviation at a monitoring point is defined as

$$\Delta = \frac{U_{\mathrm{TCN}} - U_{\mathrm{FEM}}}{U_{\mathrm{FEM}}} \times 100\% \quad (9)$$

where $U_{\mathrm{TCN}}$ and $U_{\mathrm{FEM}}$ are the TCN and calibrated numerical forecasts of lateral wall displacement at the end of the forecast horizon. The numerical baseline is the denominator because it is the reference for normal behaviour. The rule has three states:

(1) *Normal*: $\Delta \leq T$; deformation is consistent with the calibrated model, including $\Delta < 0$, when the observed trend is slower than expected;

(2) *Watch*: $\Delta > T$ and $U_{\mathrm{TCN}} \leq \alpha U_{\mathrm{c}}$; the trend is anomalous but the forecast displacement is well below the control value, so monitoring frequency and site inspection are increased;

(3) *Warning*: $\Delta > T$ and $U_{\mathrm{TCN}} > \alpha U_{\mathrm{c}}$; the trend is anomalous and approaches the control value, so preventive measures are required.

The adopted values are $T$ = 10%, $\alpha$ = 0.7 and $U_{\mathrm{c}}$ = 30 mm; their basis and sensitivity are examined in the Results.

**Evaluation metrics and baseline models**

Accuracy was measured by the root-mean-square error (RMSE), the mean absolute error (MAE) and the mean absolute percentage error (MAPE):

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (10)$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (11)$$

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\% \quad (12)$$

where $y_i$ and $\hat{y}_i$ are the reference and predicted values. RMSE and MAE carry the unit of the predicted quantity (° for $\varphi'$, MPa for $E_{\mathrm{ur}}^{\mathrm{ref}}$ and mm for displacement).

*Inversion*. In addition to a single 40/10 split, accuracy was evaluated by five-fold cross-validation repeated 20 times with different random partitions, giving 100 train–test evaluations per model. Five baselines were trained on identical folds: multiple linear regression (MLR); a BP neural network with one hidden layer and Bayesian regularisation; support vector regression with a Gaussian kernel and hyperparameters tuned by Bayesian optimisation (BO-SVR), representing the optimiser-tuned support vector regression family; Gaussian process regression with an automatic relevance determination kernel (GPR); and random forest (RF) with 200 trees. Paired differences in MAPE between the TCN and each baseline were tested with the Wilcoxon signed-rank test.

*Time-series forecasting*. The TCN was compared with persistence (the last measured value), a random walk with drift, ARIMA with orders selected by the Akaike information criterion, BP and support vector regression with a three-day lagged input, and long short-term memory (LSTM) and gated recurrent unit (GRU) networks[43,44] of comparable size. All models used the windows defined above and the same rolling-origin selection; neural networks were trained with ten random seeds. Computations were performed in MATLAB R2023b on a desktop computer with an Intel Core i7-12700 CPU, 32 GB of RAM and an NVIDIA RTX 3060 GPU.

# Results

## Deviation of the uncalibrated model from field measurements

At the end of stage VIII, the wall displacement and ground settlement reached their maxima and the coupled seepage–stress effects and accumulated parameter errors were most pronounced, so this stage was selected as the control condition for the sensitivity analysis and the inversion. The ground settlement forms a trough with a maximum of 62.8 mm about 15 m from the pit edge (Supplementary Fig. S5a), and the wall displacement follows a bulging profile with a maximum of 22.3 mm near the excavation face (Supplementary Fig. S5b).

A stage-by-stage comparison at ZQT09 and DBC09 (Fig. 5) shows close agreement during stages I–III; thereafter the simulated values diverge from the measurements, and the divergence grows with depth. Two error measures are used. The point error $\varepsilon = |U_{\mathrm{sim}} - U_{\mathrm{meas}}|/U_{\mathrm{meas}} \times 100\%$ compares simulated and measured values at one instant, and the MAPE (Eq. (12)) summarises the agreement over the stage I–IX history. At the end of stage IX, the simulated wall displacement at ZQT09 was 18.94 mm against 21.87 mm measured ($\varepsilon$ = 13.4%), and the simulated settlement at DBC09 was 77.81 mm against 104.64 mm measured ($\varepsilon$ = 25.64%). Over the whole history, the MAPEs were 21.0% and 29.2% and the MAEs 2.25 mm and 18.0 mm, respectively. The consistent underestimation shows that the uncalibrated model cannot support safety assessment.

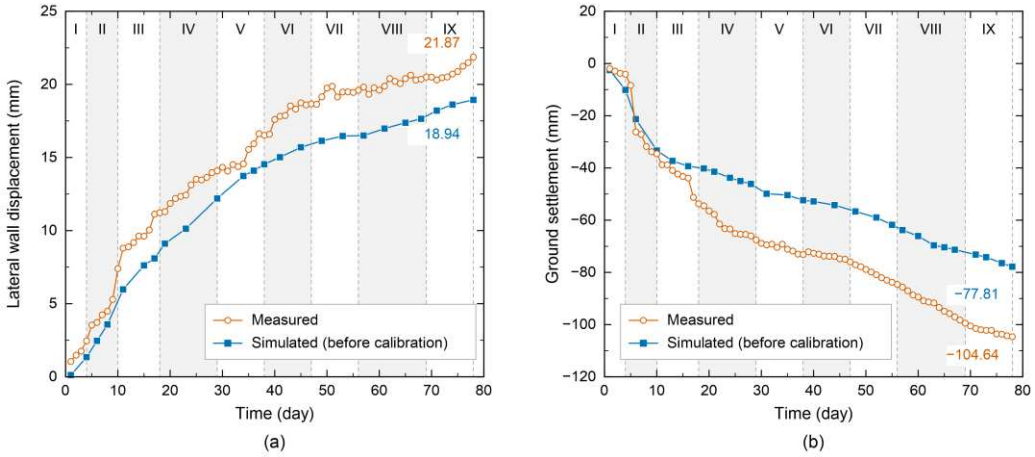



**Figure 5.** Simulated and measured deformation before calibration: (**a**) horizontal wall displacement at ZQT09; (**b**) ground surface settlement at DBC09. Roman numerals denote excavation stages.

## Sensitivity screening of soil parameters

A one-at-a-time sensitivity analysis was performed by varying each parameter of the silty clay and fine-sand layers between −30% and +30% while keeping the others unchanged, and recording the changes in maximum wall displacement and maximum ground settlement. Sensitivity was classified as extremely high (>30%), high (20–30%), moderate (10–20%) or low (<10%)[22,24].

For the silty clay (Supplementary Fig. S6), reducing $\varphi'$ by 30% increased the wall displacement by 18.39% and the settlement by 18.68%, and reducing $E_{\mathrm{ur}}^{\mathrm{ref}}$ by 30% increased them by 21.24% and 13.39%, respectively. For the fine-sand layers (Supplementary Fig. S7), the corresponding increases were 21.39% and 22.45% for $\varphi'$ and 24.85% and 20.85% for $E_{\mathrm{ur}}^{\mathrm{ref}}$. The cohesion of the silty clay and the other two moduli of all layers changed either measure by at most 12.03%. Thus $\varphi'$ and $E_{\mathrm{ur}}^{\mathrm{ref}}$ were the two most influential parameters of each layer for both deformation measures, with $E_{\mathrm{ur}}^{\mathrm{ref}}$ dominating the wall displacement and $\varphi'$ the settlement.

## Parameter interactions

Equation (5) shows that $\varphi'$ and $E_{\mathrm{ur}}^{\mathrm{ref}}$ do not act independently, so their effects need not be additive. The one-at-a-time analysis was therefore used only for screening, and the six parameters were inverted jointly. To quantify their interactions, a Gaussian-process surrogate was fitted to the 50 samples, and first-order ($S_1$) and total-order ($S_T$) Sobol indices were estimated with the Saltelli scheme[46,47]; the difference $S_T - S_1$ measures the share of variance due to interactions (Fig. 6). For the wall displacement, $E_{\mathrm{ur}}^{\mathrm{ref}}$ had the largest first-order indices in every layer (0.13–0.17), and for the settlement, $\varphi'$ did (0.11–0.14), in agreement with the screening. In every case $S_T$ was about twice $S_1$: interactions accounted for 0.08–0.17 of the

variance per parameter, and the first-order indices summed to only 0.69 for the wall displacement and 0.64 for the settlement.

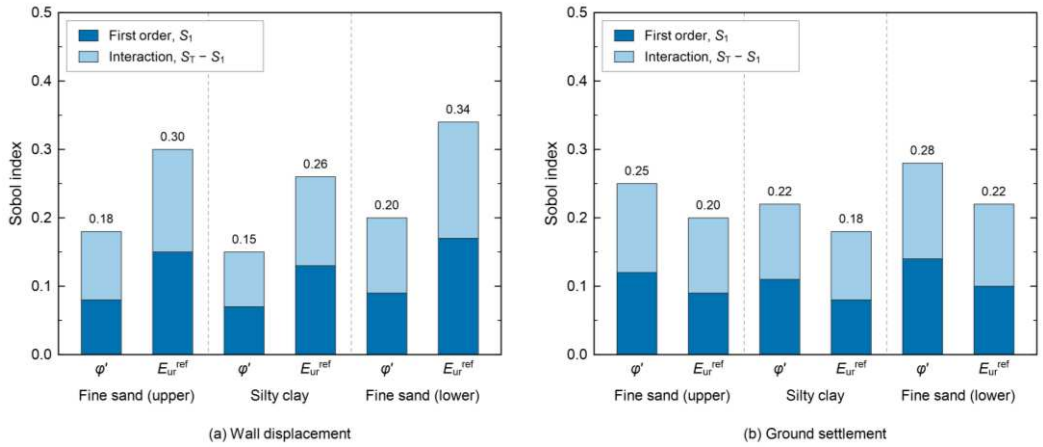



**Figure 6.** Variance-based (Sobol) sensitivity indices of the six inverted parameters for (**a**) wall displacement and (**b**) ground settlement. The dark part of each bar is the first-order index $S_1$, the light part the interaction share $S_T - S_1$, and the number the total-order index $S_T$.

## Inversion accuracy and model calibration

*Repeated cross-validation and baselines.* Figure 7 summarises the 100 train–test evaluations for the TCN and the five baselines. The mean MAPE over the six parameters was 5.0% for the TCN (standard deviation 0.8%). The strongest baseline was RF, with a mean MAPE of 6.6%, followed by GPR (6.9%) and BO-SVR (7.3%); the TCN reduced the error by 24% relative to RF, and its paired differences from all baselines were significant (Wilcoxon signed-rank test, $p < 0.001$). The TCN was the most accurate model for every one of the six parameters. Errors were consistently larger for $E_{ur}^{ref}$ than for $\varphi'$ for all models.

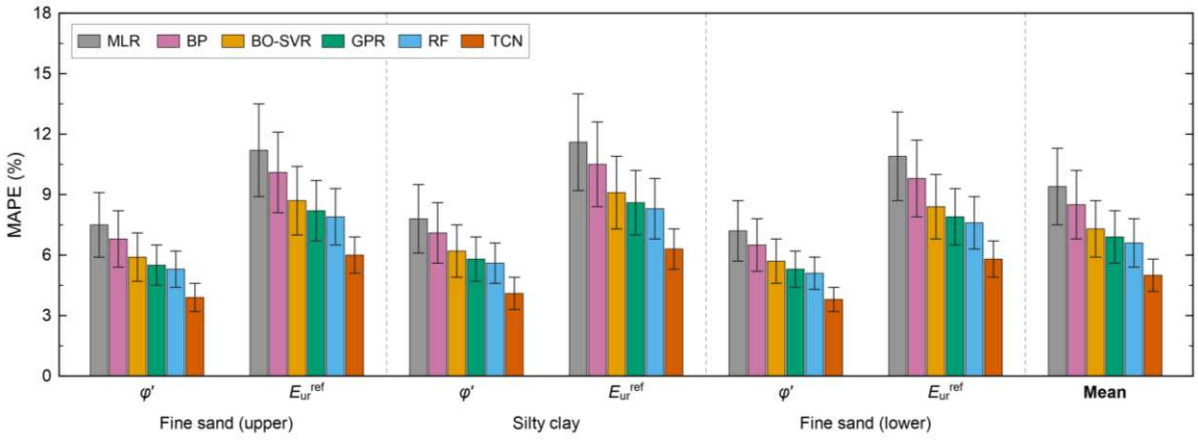



**Figure 7.** Inversion accuracy of the TCN and five baseline models under five-fold cross-validation repeated 20 times. Bars show the mean MAPE over the 100 test folds and error bars one standard deviation; "Mean" is the average over the six parameters.

*Sample adequacy and robustness.* Training on 20, 30 and 40 samples gave mean test MAPEs of 7.6%, 5.9% and 5.0%, respectively, indicating that the inversion accuracy improves with the number of training samples and that the marginal gain diminishes beyond 30 samples. Adding zero-mean Gaussian noise with standard deviations of 2% and 5% of each response to the test inputs increased the mean MAPE to 5.8% and 7.3%, respectively.

Feeding the field measurements into the trained model gave $\varphi' = 24.86°$ and $E_{\mathrm{ur}}^{\mathrm{ref}} = 45.32$ MPa for the upper fine sand, $\varphi' = 19.78°$ and $E_{\mathrm{ur}}^{\mathrm{ref}} = 74.59$ MPa for the silty clay, and $\varphi' = 25.45°$ and $E_{\mathrm{ur}}^{\mathrm{ref}} = 57.68$ MPa for the lower fine sand. All six values lie inside, not at the bounds of, the sampled ranges, so the identification is not constrained by the prior limits. Propagating 2% input noise through the ensemble of 100 cross-validation models gives standard deviations of 0.35° and 1.8 MPa for $\varphi'$ and $E_{\mathrm{ur}}^{\mathrm{ref}}$ of the upper fine sand, 0.42° and 3.5 MPa for the silty clay, and 0.58° and 2.9 MPa for the lower fine sand.

Substituting these parameters into the numerical model markedly reduced the deviation between simulated and measured values (Fig. 8). Over the stage I–IX history, the MAPE of wall displacement at ZQT09 decreased from 21.0% to 3.0% (MAE from 2.25 mm to 0.27 mm), and the point error at the end of stage IX decreased from 13.4% to 2.1%. For settlement at DBC09, the MAPE decreased from 29.2% to 5.8% (MAE from 18.0 mm to 1.94 mm), and the point error at the end of stage IX from 25.64% to 1.3%. This is an out-of-sample check, because the inversion used only the nine stage VIII responses, whereas the comparison covers the full histories, including stage IX.

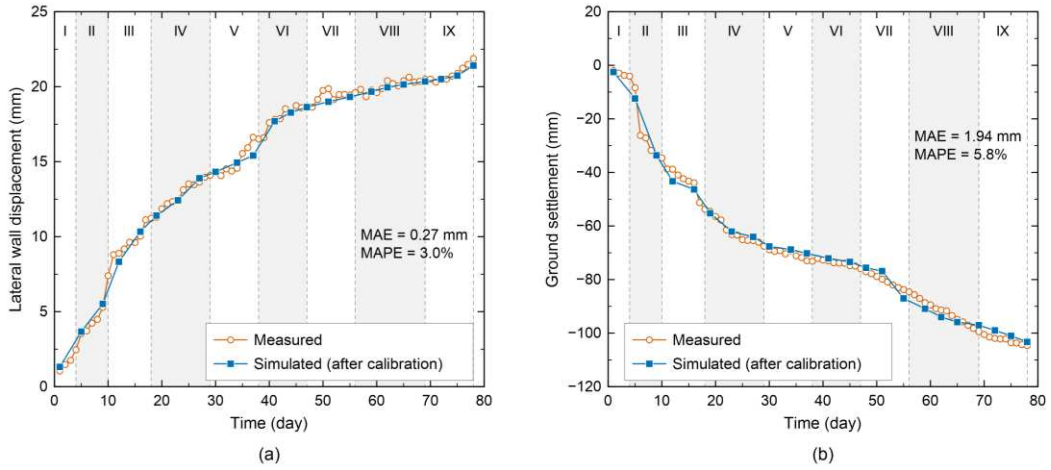



**Figure 8.** Simulated and measured deformation after parameter inversion: (**a**) horizontal wall displacement at ZQT09; (**b**) ground surface settlement at DBC09. MAE and MAPE are evaluated over stages I–IX.

### Time-series forecasting at ZQT09

Figure 9 shows the TCN predictions over the whole record. The MAE and MAPE were 0.48 mm and 4.52% over the training window and 0.46 mm and 2.26% over the test window. These values are not on their own evidence of skill. Over days 63–78 the measured displacement rose from 20.21 to 21.87 mm, a mean daily increment of 0.11 mm, so the series is close to a random walk and the one-day-ahead task is easy: persistence gives an MAE of 0.20 mm and a MAPE of 0.96% over the same window (Table 2), so the TCN has negative skill relative to persistence at this horizon. The comparison that separates the models is made at the 14-day horizon, where persistence gives an MAE of 3.41 mm and a MAPE of 13.09% and ARIMA gives 1.37 mm and 5.28%.

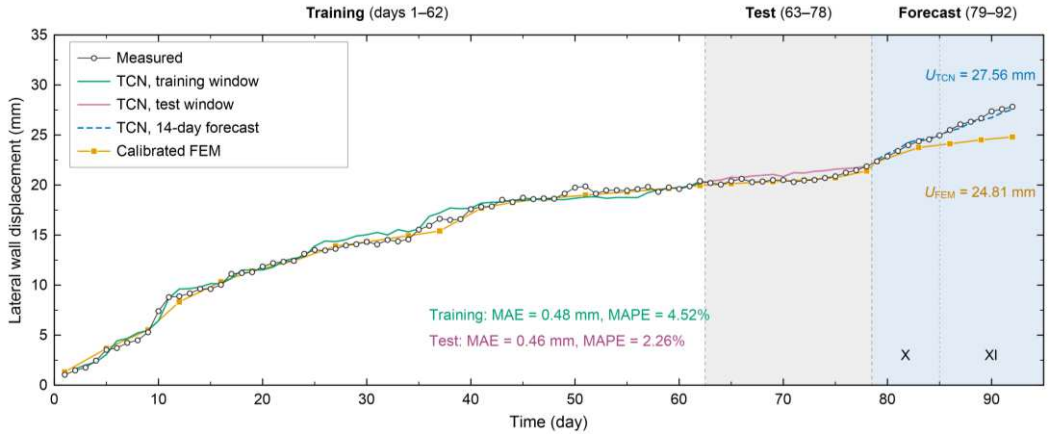



**Figure 9.** Prediction of horizontal wall displacement at ZQT09: TCN fit over the training window (days 1–62), TCN prediction over the chronologically held-out test window (days 63–78), TCN forecast for stages X and XI (days 79–92) and calibrated finite-element (FEM) response.

**Table 2.** Accuracy of the forecasting models at ZQT09 (mean over ten random seeds for neural networks).

| Model | Days 63–78, one-day-ahead | | | Days 79–92, one-day-ahead | | | Days 79–92, closed-loop 14-day | | |
|---|---|---|---|---|---|---|---|---|---|
| | MAE (mm) | RMSE (mm) | MAPE (%) | MAE (mm) | RMSE (mm) | MAPE (%) | MAE (mm) | RMSE (mm) | MAPE (%) |
| Persistence | 0.20 | 0.23 | 0.96 | 0.43 | 0.51 | 1.70 | 3.41 | 4.20 | 13.09 |
| Random walk with drift | 0.24 | 0.31 | 1.18 | 0.16 | 0.21 | 0.64 | 1.38 | 1.80 | 5.33 |
| ARIMA(2,1,0) | 0.22 | 0.28 | 1.08 | 0.25 | 0.33 | 1.00 | 1.37 | 1.80 | 5.28 |
| BP ($L$ = 3) | 0.55 | 0.68 | 2.80 | 0.35 | 0.45 | 1.35 | 2.80 | 3.50 | 10.50 |
| SVR ($L$ = 3) | 0.48 | 0.60 | 2.45 | 0.32 | 0.41 | 1.22 | 2.50 | 3.10 | 9.50 |
| LSTM | 0.42 | 0.53 | 2.15 | 0.28 | 0.36 | 1.08 | 2.20 | 2.75 | 8.30 |
| GRU | 0.40 | 0.51 | 2.05 | 0.27 | 0.35 | 1.05 | 2.10 | 2.65 | 8.00 |
| TCN, sequence mode | 0.46 | 0.58 | 2.26 | 0.23 | 0.29 | 0.88 | 1.20 | 1.50 | 4.60 |
| TCN, window $L$ = 1 | 0.50 | 0.62 | 2.40 | 0.24 | 0.31 | 0.92 | 1.35 | 1.70 | 5.10 |
| TCN, window $L$ = 3 | 0.47 | 0.59 | 2.30 | 0.23 | 0.30 | 0.90 | 1.28 | 1.60 | 4.90 |
| TCN, window $L$ = 7 | 0.45 | 0.57 | 2.22 | 0.22 | 0.29 | 0.87 | 1.22 | 1.53 | 4.70 |
| TCN, window $L$ = 14 | 0.44 | 0.56 | 2.18 | 0.21 | 0.28 | 0.85 | 1.18 | 1.48 | 4.55 |

Persistence, random walk with drift and ARIMA were fitted on days 1–78 only. $L$, explicit look-back window (days).

At the 14-day closed-loop horizon, all models except persistence achieved a MAPE below 11%. LSTM (8.30%) and GRU (8.00%) were less accurate than the statistical references, ARIMA (5.28%) and the random walk with drift (5.33%), because their closed-loop errors accumulated over the horizon. The TCN variants gave the lowest errors (4.55–5.10%); in sequence mode, its MAPE of 4.60% was 13% below that of ARIMA. At the one-day horizon, all neural-network models underperformed persistence over days 63–78, because the record is close to a random walk and offers little temporal structure to learn. In the sliding-window mode, the test MAPE was 2.40%, 2.30%, 2.22% and 2.18% for $L$ = 1, 3, 7 and 14 days, respectively, a marginal improvement with diminishing returns beyond seven days.

At the end of stage XI, the TCN prediction at ZQT09 reached 27.56 mm without convergence, whereas the calibrated numerical forecast reached 24.81 mm with a flattening trend, giving $\Delta$ = 11.08% (Eq. (9)). The measurements departed progressively from the numerical forecast: the relative gap was 0.18% on day 79, 2.65% on day 83, 5.72% on day 86, 8.81% on day 89 and 12.17% on day 92, so the divergence developed throughout the stage rather than appearing at a single instant. A water leakage incident occurred at ZQT09 during stage XI, and the resulting deformation increase could not be reproduced by the numerical model, which does not include such events. The rolling one-day-ahead prediction tracked the measurements over days 79–92 with an MAE of 0.23 mm and a MAPE of 0.88%.

### Multi-point verification of the warning rule

The warning rule was applied to ZQT01–ZQT08 (Supplementary Fig. S8). Table 3 lists Δ, the ratio $U_{TCN}/U_c$, the resulting state and the field observations for all nine points, and Fig. 10 shows them in the decision plane.

**Table 3.** Dual-track evaluation at the end of the forecast horizon (stage XI) and field observations.

| Point | $U_{TCN}$ (mm) | $U_{FEM}$ (mm) | Δ (%) | $U_{TCN}/U_c$ | State | Field observation |
|---|---|---|---|---|---|---|
| ZQT01 | 9.03 | 8.25 | 9.45 | 0.30 | Normal | No anomaly |
| ZQT02 | 10.52 | 9.88 | 6.48 | 0.35 | Normal | No anomaly |
| ZQT03 | 15.75 | 10.25 | 53.66 | 0.53 | Watch | No accelerating trend or seepage |
| ZQT04 | 23.56 | 17.86 | 31.91 | 0.79 | Warning | Accelerating deformation; slight seepage outside the curtain |
| ZQT05 | 17.59 | 15.73 | 11.82 | 0.59 | Watch | No anomaly |
| ZQT06 | 24.63 | 22.58 | 9.08 | 0.82 | Normal | No anomaly |
| ZQT07 | 19.22 | 21.02 | −8.56 | 0.64 | Normal | No anomaly |
| ZQT08 | 29.74 | 25.45 | 16.86 | 0.99 | Warning | Accelerating deformation; slight seepage outside the curtain |
| ZQT09 | 27.56 | 24.81 | 11.08 | 0.92 | Warning | Water leakage during stage XI |

$U_c$ = 30 mm; states defined in Methods with $T$ = 10% and $\alpha$ = 0.7.

Four points (ZQT01, ZQT02, ZQT06 and ZQT07) have $|\Delta| \leq 9.45\%$ and are classified as normal; ZQT07 has a negative deviation, its record developing more slowly than the calibrated model predicts. ZQT03 (Δ = 53.66%, 0.53$U_c$) and ZQT05 (Δ = 11.82%, 0.59$U_c$) exceed the relative threshold but remain at low absolute levels and are classified as watch. ZQT04 (31.91%, 0.79$U_c$), ZQT08 (16.86%, 0.99$U_c$) and ZQT09 (11.08%, 0.92$U_c$) exceed both tiers and are classified as warning.

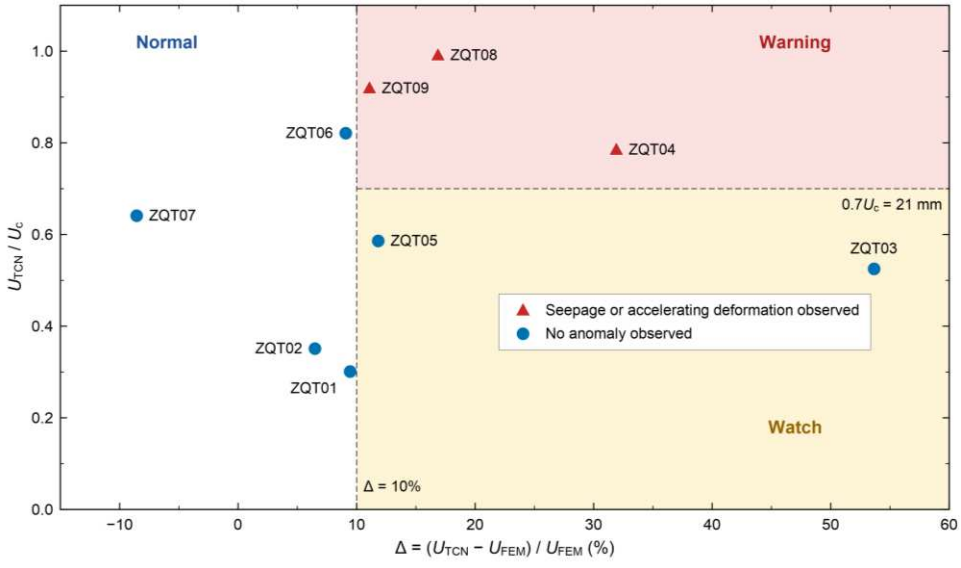



**Figure 10.** Decision plane of the dual-track warning rule for ZQT01–ZQT09: relative deviation Δ (Eq. (9)) against the TCN forecast normalised by the control value $U_c$ = 30 mm. Dashed lines indicate $T$ = 10% and $\alpha U_c$ = 21 mm.

*Basis of the 10% threshold.* Nine points cannot support a statistical confidence interval, so the threshold is treated as a heuristic supported by three lines of evidence. First, under normal conditions the deviation between the two forecasts arises from the errors of both models. Using the MAPE of the calibrated model for wall displacement (3.0%) and that of the TCN (2.26% on the test window and 4.52% on the training window), and converting MAPE to standard deviation by the factor $(\pi/2)^{1/2} \approx 1.25$ that holds for normally distributed errors, the combined standard deviation is 4.7–6.8% and the one-sided 95% bound is 7.7–11.2%; a deviation above about 10% is therefore unlikely to arise from model error alone. Second, the four normal-condition points all lie below 10% (maximum 9.45%). Third, Supplementary Table S2 shows how the classification changes with $T$.

With $\alpha = 0.7$, every threshold between 9.08% (ZQT06) and 11.08% (ZQT09) reproduces the field observations, and 10% lies near the middle of this range. With $T = 10\%$, every absolute tier between $0.59U_c$ (ZQT05) and $0.79U_c$ (ZQT04) gives the same result; $0.7U_c$ is consistent with the practice of setting warning levels at a fixed fraction of the control value[38]. Without the absolute tier, two false alarms occur. These windows are narrow (about ±1 percentage point for $T$) and rest on nine points, three of them positive, so $T = 10\%$ and $\alpha = 0.7$ should be regarded as calibrated for this site.

During construction, the measured responses at ZQT04 and ZQT08 confirmed the warnings. The displacement rates, cumulative settlements and settlement rates at these points were markedly higher than at the other sections, and both displacement–time and settlement–time curves accelerated without converging. Field inspection and water-level data revealed larger groundwater fluctuations and slight seepage on the outer side of the waterproof curtain, first recorded on day 86 at ZQT04 and day 88 at ZQT08. At the two watch points, ZQT03 and ZQT05, no accelerating trend or seepage was observed, so increased monitoring was sufficient. Over the nine points, the rule therefore gave three correct warnings, two watch states without consequence, four correct normal states, no false alarms and no missed cases. Excluding ZQT09, which was used to develop the method, the independent set (ZQT01–ZQT08) contains two positive and six negative cases.

## Discussion

### Deformation mechanism and parameter identifiability

The dominance of $\varphi'$ and $E_{\mathrm{ur}}^{\mathrm{ref}}$ follows from the stress paths of excavation. Excavation is predominantly an unloading process: the soil inside the pit and behind the wall follows

unloading or reloading paths, on which the MMC model uses $E_{\mathrm{ur}}^{\mathrm{ref}}$, whereas $E_{50}^{\mathrm{ref}}$ and $E_{\mathrm{oed}}^{\mathrm{ref}}$ govern primary deviatoric and oedometric loading, which occurs only locally. The friction angle acts through two mechanisms: it sets the mobilised shear strength and hence the active earth pressure on the wall, and it enters the stress-dependent stiffness of Eq. (5). The cohesion is small in the silty clay (5.9 kPa) and zero in the sands, so the frictional term controls both. The silty clay and fine-sand layers together span the depth of the maximum wall bulge and the zone of dewatering-induced consolidation, which explains why their parameters dominate both deformation measures.

The two deformation measures respond to different mechanisms, which is why they rank the parameters differently. The wall deflection is governed by the stiffness of the retained soil under lateral unloading and is therefore most sensitive to $E_{\mathrm{ur}}^{\mathrm{ref}}$ (Supplementary Figs. S6 and S7; Fig. 6a). The ground settlement integrates the shear deformation of the soil wedge behind the wall and the volumetric compression caused by dewatering; both depend on the mobilised strength, so $\varphi'$ dominates (Fig. 6b). This difference is useful for the inversion. Because the wall-deflection profile and the settlement trough weight the two parameters differently, combining them in one response vector constrains a combination of $\varphi'$ and $E_{\mathrm{ur}}^{\mathrm{ref}}$ that neither measure alone could separate, which reduces the non-uniqueness of the inverse problem.

The Sobol analysis shows that this separation is incomplete. Interactions account for roughly half of the total-order indices (Fig. 6), because a lower $\varphi'$ both increases the earth pressure and reduces the stiffness through Eq. (5), so the effect of $E_{\mathrm{ur}}^{\mathrm{ref}}$ depends on the value of $\varphi'$ in the same layer. Two consequences follow. First, a one-at-a-time analysis, which by construction cannot detect such interactions, is suitable for screening but not for calibration, and the parameters must be inverted jointly, as done here. Second, the larger inversion errors for $E_{\mathrm{ur}}^{\mathrm{ref}}$ (Fig. 7) reflect the shared influence: part of the information about $E_{\mathrm{ur}}^{\mathrm{ref}}$ in the responses is attributed to $\varphi'$, and $E_{\mathrm{ur}}^{\mathrm{ref}}$ also spans a wider relative range (a factor of 1.5–2.0 between its bounds, against 1.3–1.4 for $\varphi'$).

The inverted parameters are effective values conditional on the MMC model, the fixed non-inverted parameters, the boundary and hydraulic conditions and the sampled ranges. The inverse TCN learns the mapping of the numerical model, not of the ground, so model-form error is absorbed into these values. Three observations show that the calibrated set nonetheless represents the site response: the values lie inside the sampled ranges; the calibrated model reproduces the full stage I–IX histories, which were not used in the inversion; and its forecasts agree with the TCN within 10% at the four normal points. The choice of the MMC model rather

than the hardening soil model with small-strain stiffness (HSS)[13,48] is consistent with this evidence. Small-strain stiffness raises the stiffness at small strains and would in general reduce the computed displacements, so its omission cannot explain the 25.64% underestimation of settlement before calibration; the bias is instead consistent with overly stiff correlation-based values of $E_{\mathrm{ur}}^{\mathrm{ref}}$ and $\varphi'$, which the inversion corrected.

**Sources of the TCN advantage in inversion and forecasting**

The 24% lower inversion error of the TCN relative to RF, the strongest baseline, is consistent with the inductive bias of the spatial convolution. RF, GPR and BO-SVR treat the nine responses as independent inputs and must infer the relations between neighbouring measurement points from the data. The TCN shares one set of weights along the deflection profile and the settlement trough, so features such as the curvature of the profile and the depth of its maximum are extracted with few parameters. This matters most in the small-sample regime of coupled three-dimensional analyses: with 40 training samples, a model that encodes the spatial order uses the samples to learn the parameter mapping rather than the geometry of the response. It is, however, moderate, and with larger sample sets the difference to GPR and RF may narrow.

For forecasting, the decisive factor is the horizon rather than the architecture. Over the test window, the mean daily increment of wall displacement (0.11 mm) is small relative to the day-to-day fluctuations of the record, so the one-day-ahead task is dominated by persistence, and the apparent accuracy of all learning models (MAPE 2.05–2.80%) reflects no real skill. Recent studies report improvements of about 20% in RMSE and MAE between hybrid and simpler networks[29], but such comparisons are made among neural networks, and a naive reference is rarely included. At the 14-day horizon, the persistence error (13.09%) exceeds the warning threshold itself, so a forecast is needed. Here the recurrent networks accumulated errors in closed-loop mode and fell behind ARIMA, whereas the TCN, whose causal convolutions process the whole sequence in parallel and whose receptive field covers the development record, gave the lowest error. The data-driven track therefore contributes multi-day trend extrapolation at the warning horizon, which is exactly the information the warning rule requires.

**Physical basis of the dual-track warning**

Existing studies typically use data-driven forecasts on their own, or use measurements only to calibrate a numerical model that is then trusted. This study instead treats the calibrated

numerical model as a hypothesis about normal behaviour and the TCN forecast as an observation-driven test of that hypothesis. The warning signal is the disagreement between the two, and it has a physical interpretation. The numerical model assumes an intact waterproof curtain and the design hydraulic heads. When the curtain leaks, groundwater flows into the pit, the head outside the wall falls locally, and the effective stress in the fine sand and silty clay rises. The resulting consolidation, together with the loss of fines from the sand near the defect, increases the settlement behind the wall and the lateral load on it, while seepage forces inside the pit reduce the passive resistance. None of these processes is represented in the calibrated model, but all of them accelerate the measured deformation, which the TCN extrapolates.

The field records match this mechanism. At ZQT09, the gap between the measurements and the numerical forecast grew steadily from 0.18% on day 79 to 12.17% on day 92, which is consistent with a progressively developing leak rather than a sudden event, as in the excavation leakage case analysed by Tan and Lu[6], where minor leaks preceded the outburst by two weeks. At ZQT04 and ZQT08, seepage outside the curtain and larger groundwater fluctuations were recorded on days 86 and 88, within the forecast horizon, and the measured–simulated deviations at these sections were larger than elsewhere. This suggests that local geological variability made the single calibrated parameter set unrepresentative of those sections and that the deviation captured the early stage of the seepage process. Conversely, the negative deviation at ZQT07 shows that the rule does not flag points that deform less than expected.

The two tiers of the rule reflect two different sources of error. A relative measure is sensitive at small displacements, where a small absolute difference produces a large $\Delta$, as at ZQT03 (53.66% at $0.53U_c$); the absolute tier prevents such cases from producing alarms while still raising the monitoring frequency. The relative threshold, in turn, is tied to the accuracy of the two tracks: the one-sided 95% bound of their combined error (7.7–11.2%) brackets the adopted 10%. The numerical values of $T$ and $\alpha$ are site-specific, but this derivation is general. At another site, $T$ should be set from the combined error of the calibrated model and the forecaster, and $\alpha$ from the warning levels of the applicable code, before the rule is verified against field observations.

**Computational cost, applicability and limitations**

The computational effort divides into an offline and an online part. Offline, the 50 finite-element analyses required about 2 h each (about 100 h in total) and can be run in parallel; training the inverse TCN took 25 s, and inference for a new set of measurements takes 2 ms, so re-inversion at a new stage is limited by the forward analyses rather than by the network.

Online, retraining the forecasting TCN on the updated daily record took 18 s, and a 14-day forecast took 0.1 s. Because monitoring is daily and each excavation stage lasts 6–13 days, the daily cycle of data upload, TCN update and comparison with the stored numerical baseline is completed within minutes. The framework therefore fits the stage-by-stage feedback loop of the observational method[7]: the calibrated model provides the predicted response, the forecaster provides the observed trend, and the warning rule converts their disagreement into a construction decision.

The verification covers nine monitoring points with three positive cases, which confirms that the rule is consistent with the field record but cannot establish detection or false-alarm rates with statistical confidence. Extending the evaluation to the remaining 27 monitoring points and to earlier forecast origins would quantify these rates. The warning lead time is bounded by the closed-loop forecast, whose accuracy decreases with the horizon, and the inverse model inherits the assumptions of the numerical model. Sequential updating of the parameters at each stage, surrogate-assisted adaptive sampling and physics-informed networks that embed the seepage–deformation coupling are promising ways to relax these constraints and to transfer the framework to other water-rich sites.

## Conclusions

This study developed a TCN-driven framework that couples parameter inversion, time-series forecasting and dual-track early warning, and applied it to the Φ130 m Shiziyang Bridge west anchorage in water-rich sandy strata. The effective friction angle and the unloading–reloading reference modulus of the silty clay and fine-sand layers control the deformation and interact strongly, so they must be inverted jointly. Trained on 50 Latin hypercube finite-element samples, the TCN inverse model identified them with a cross-validated MAPE of 5.0%, 24% lower than the best baseline, and calibration reduced the MAPE of the simulated histories from 21.0% to 3.0% for wall displacement and from 29.2% to 5.8% for settlement. The forecasting skill of the TCN lies at the warning horizon: no model outperformed persistence one day ahead, whereas the closed-loop 14-day MAPE was 4.60%, against 13.09% for persistence and 5.28% for ARIMA. The disagreement between the two forecasts is a physically interpretable indicator of seepage-induced deformation; with a relative threshold of 10% and an absolute tier of 70% of the control value, the rule flagged all three sections with seepage or accelerating deformation without false alarms. The threshold values are site-specific and should be recalibrated elsewhere by the same procedure.

## Data availability

The field monitoring records and finite-element model files belong to the project owner and are not publicly available owing to confidentiality restrictions. They are available from the corresponding author on reasonable request with the permission of the project owner. The Latin hypercube sample set used to train the inverse model is provided in Supplementary Table S1.

## References


1. Guo, C., Yang, X., Guo, C. & Li, P. Analysis of deformation of deep and large foundation pit support structure and impact on neighbouring buildings in complex environments. *Buildings* **15**, 1435. https://doi.org/10.3390/buildings15091435 (2025).
2. Lee, M. L. Design and construction of soil nail reinforced soldier pile wall as a temporary shoring system for deep excavation—a case study. *Int. J. Geotech. Eng.* **16**, 850–864. https://doi.org/10.1080/19386362.2020.1822571 (2022).
3. Wang, Z., Guo, X. & Wang, C. Field monitoring analysis of construction process of deep foundation pit at subway station. *Geotech. Geol. Eng.* **37**, 549–559. https://doi.org/10.1007/s10706-018-0626-4 (2019).
4. Xiang, X., Xu, P., Qiao, S. & Dong, H. Deformation characteristics of excavation under hydrogeological bias and optimization of support structures. *Int. J. Geotech. Eng.* **20**, 156–170. https://doi.org/10.1080/19386362.2025.2608360 (2026).
5. Liu, B., Zhang, D., Wang, Y., Wang, N. & Xu, W. Design optimization and observed performance of a super-large foundation pit excavation subjected to unsymmetrical loading in water-rich floodplain: A case study. *Soils Found.* **63**, 101329. https://doi.org/10.1016/j.sandf.2023.101329 (2023).
6. Tan, Y. & Lu, Y. Forensic diagnosis of a leaking accident during excavation. *J. Perform. Constr. Facil.* **31**, 04017061. https://doi.org/10.1061/(ASCE)CF.1943-5509.0001058 (2017).
7. Peck, R. B. Advantages and limitations of the observational method in applied soil mechanics. *Géotechnique* **19**, 171–187. https://doi.org/10.1680/geot.1969.19.2.171 (1969).
8. Tabaroei, A., Sarfarazi, V., Moaveni, M., Vakili, A. H. & Pham, T. A. Deep excavation-induced stability evaluation of a triple tunnel using discrete and continuum numerical modeling. *Int. J. Geomech.* **25**, 04024299. https://doi.org/10.1061/IJGNAI.GMENG-9963 (2025).
9. Zhang, Z. et al. Shield construction effect on the existing subway tunnel structures and efficiency of grouting reinforcement measures: A 3D model and a case study. *Mater. Sci.* **32**, 114–122. https://doi.org/10.5755/j02.ms.41087 (2026).
10. Fu, J. et al. Numerical modeling of crack propagation patterns and the influence of joint configuration on tunnel and dam stability. *Comput. Part. Mech.* **12**, 5859–5881. https://doi.org/10.1007/s40571-025-01103-y (2025).
11. Moazami, A., Soroush, A., Fattahi, S. M. & Sasanian, S. Numerical study of creep in soil anchor walls—a case study. *Int. J. Geotech. Eng.* **18**, 164–176. https://doi.org/10.1080/19386362.2024.2351608 (2024).

12. Zhu, X. et al. Deformation characteristics and base stability of a circular deep foundation pit with high-pressure jet grouting reinforcement. *Appl. Sci.* **15**, 6825. https://doi.org/10.3390/app15126825 (2025).

13. Zhang, T. K., Shi, Q. L., Xu, L. W., Fu, B. Y. & Xue, X. Y. Research on deformation analysis of anchor foundation pit excavation for Shiziyang Bridge. *Highway* (6), 183–190 (2025) (in Chinese).

14. Huang, Z. H., Zhang, L. L., Cheng, S. Y., Zhang, J. & Xia, X. H. Back-analysis and parameter identification for deep excavation based on Pareto multiobjective optimization. *J. Aerosp. Eng.* **28**, A4014007. https://doi.org/10.1061/(ASCE)AS.1943-5525.0000464 (2015).

15. Qi, X. H. & Zhou, W. H. An efficient probabilistic back-analysis method for braced excavations using wall deflection data at multiple points. *Comput. Geotech.* **85**, 186–198. https://doi.org/10.1016/j.compgeo.2016.12.032 (2017).

16. Jiang, S. H., Huang, J., Qi, X. H. & Zhou, C. B. Efficient probabilistic back analysis of spatially varying soil parameters for slope reliability assessment. *Eng. Geol.* **271**, 105597. https://doi.org/10.1016/j.enggeo.2020.105597 (2020).

17. Jin, Y. F., Yin, Z. Y., Zhou, W. H. & Liu, X. Intelligent model selection with updating parameters during staged excavation using optimization method. *Acta Geotech.* **15**, 2473–2491. https://doi.org/10.1007/s11440-020-00936-6 (2020).

18. Levasseur, S., Malécot, Y., Boulon, M. & Flavigny, E. Soil parameter identification using a genetic algorithm. *Int. J. Numer. Anal. Methods Geomech.* **32**, 189–213. https://doi.org/10.1002/nag.614 (2008).

19. Jin, Y. F., Yin, Z. Y., Zhou, W. H. & Huang, H. W. Multi-objective optimization-based updating of predictions during excavation. *Eng. Appl. Artif. Intell.* **78**, 102–123. https://doi.org/10.1016/j.engappai.2018.11.002 (2019).

20. Juang, C. H., Luo, Z., Atamturktur, S. & Huang, H. Bayesian updating of soil parameters for braced excavations using field observations. *J. Geotech. Geoenviron. Eng.* **139**, 395–406. https://doi.org/10.1061/(ASCE)GT.1943-5606.0000782 (2013).

21. Tao, Y., Sun, H. & Cai, Y. Predictions of deep excavation responses considering model uncertainty: Integrating BiLSTM neural networks with Bayesian updating. *Int. J. Geomech.* **22**, 04021250. https://doi.org/10.1061/(ASCE)GM.1943-5622.0002245 (2022).

22. Ma, H. H., Yuan, S., Zhang, Z. Z., Tian, Y. H. & Dong, S. S. Application of soil parameter inversion method based on BP neural network in foundation pit deformation prediction. *Appl. Geophys.* **20**, 299–309. https://doi.org/10.1007/s11770-023-1029-8 (2023).

23. Chen, C. et al. Soil parameter inversion in dredger fill strata using GWO-MLSSVR for deep foundation pit engineering. *Buildings* **15**, 1864. https://doi.org/10.3390/buildings15111864 (2025).

24. Liu, W. et al. Multi-parameter intelligent inverse analysis of a deep excavation considering path-dependent behavior of soils. *Comput. Geotech.* **174**, 106597. https://doi.org/10.1016/j.compgeo.2024.106597 (2024).

25. Zhao, H. J., Liu, W., Shi, P. X., Du, J. T. & Chen, X. M. Spatiotemporal deep learning approach on estimation of diaphragm wall deformation induced by excavation. *Acta Geotech.* **16**, 3631–3645. https://doi.org/10.1007/s11440-021-01264-z (2021).

26. Wu, C. et al. Prediction of wall deflection induced by braced excavation in spatially variable soils via convolutional neural network. *Gondwana Res.* **123**, 184–197. https://doi.org/10.1016/j.gr.2022.06.011 (2023).

27. Li, T., Shu, J. & Chang, D. Time-series prediction of settlement deformation in shallow buried tunnels based on EMD-SSA-GRNN model. *Sci. Rep.* **14**, 324. https://doi.org/10.1038/s41598-024-51165-w (2024).

28. Li, T., Shu, J. & Wang, Y. Deformation prediction of underground engineering support structures via the ST-CNN-LSTM model. *J. Civ. Struct. Health Monit.* **15**, 1901–1919. https://doi.org/10.1007/s13349-025-00917-x (2025).

29. Liao, Y. et al. Deformation prediction of deep foundation pit support piles based on a CNN-LSTM-Transformer model with spatiotemporal feature fusion. *Buildings* **16**, 2123. https://doi.org/10.3390/buildings16112123 (2026).

30. Ye, J. F. Research on deformation monitoring and prediction of deep foundation pits based on TCN-LSTM model. *Guangzhou Archit.* **54**, 72–77 (2026) (in Chinese).

31. Wei, J., Pan, Y. & Chen, J. J. Physics-informed edge-enhanced temporal graph convolutional network for multi-risk evolution prediction in deep excavation. *Adv. Eng. Inform.* **71**, 104391. https://doi.org/10.1016/j.aei.2026.104391 (2026).

32. Zhang, W. G. & Phoon, K. K. Editorial for advances and applications of deep learning and soft computing in geotechnical underground engineering. *J. Rock Mech. Geotech. Eng.* **14**, 671–673. https://doi.org/10.1016/j.jrmge.2022.01.001 (2022).

33. Hong, C., Luo, G. & Chen, W. Safety analysis of a deep foundation ditch using deep learning methods. *Gondwana Res.* **123**, 16–26. https://doi.org/10.1016/j.gr.2022.05.015 (2023).

34. Wang, X. et al. Research progress of machine learning in deep foundation pit deformation prediction. *Buildings* **15**, 852. https://doi.org/10.3390/buildings15060852 (2025).

35. Bai, S., Kolter, J. Z. & Koltun, V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. Preprint at https://arxiv.org/abs/1803.01271 (2018).

36. Li, S., Zhang, S., Zhang, C., Liu, L. & Zhang, X. Temporal convolutional network for gas concentration prediction based on weighted loss and channel coupling attention. *IEEE Sens. J.* **25**, 9802–9816. https://doi.org/10.1109/JSEN.2025.3529930 (2025).

37. Hyndman, R. J. & Koehler, A. B. Another look at measures of forecast accuracy. *Int. J. Forecast.* **22**, 679–688. https://doi.org/10.1016/j.ijforecast.2006.03.001 (2006).

38. Ministry of Housing and Urban-Rural Development of the People's Republic of China. *Technical Standard for Monitoring of Building Excavation Engineering (GB 50497-2019)* (China Planning Press, 2019) (in Chinese).

39. Ministry of Transport of the People's Republic of China. *Specifications for Design of Foundation of Highway Bridges and Culverts (JTG 3363-2019)* (China Communications Press, 2019) (in Chinese).

40. Schanz, T., Vermeer, P. A. & Bonnier, P. G. The hardening soil model: Formulation and verification. In *Beyond 2000 in Computational Geotechnics* (ed. Brinkgreve, R. B. J.) 281–296 (Balkema, 1999).

41. McKay, M. D., Beckman, R. J. & Conover, W. J. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* **21**, 239–245. https://doi.org/10.1080/00401706.1979.10489755 (1979).

42. Bergmeir, C. & Benítez, J. M. On the use of cross-validation for time series predictor evaluation. *Inf. Sci.* **191**, 192–213. https://doi.org/10.1016/j.ins.2011.12.028 (2012).
43. Hochreiter, S. & Schmidhuber, J. Long short-term memory. *Neural Comput.* **9**, 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735 (1997).
44. Cho, K. et al. Learning phrase representations using RNN encoder–decoder for statistical machine translation. In *Proc. 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* 1724–1734 (Association for Computational Linguistics, 2014).
45. Loeppky, J. L., Sacks, J. & Welch, W. J. Choosing the sample size of a computer experiment: A practical guide. *Technometrics* **51**, 366–376. https://doi.org/10.1198/TECH.2009.08040 (2009).
46. Sobol, I. M. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Math. Comput. Simul.* **55**, 271–280. https://doi.org/10.1016/S0378-4754(00)00270-6 (2001).
47. Saltelli, A. et al. Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Comput. Phys. Commun.* **181**, 259–270. https://doi.org/10.1016/j.cpc.2009.09.018 (2010).
48. Benz, T. *Small-Strain Stiffness of Soils and Its Numerical Consequences*. PhD thesis, Univ. Stuttgart (2007).

## Acknowledgements

The authors gratefully acknowledge the College of Civil and Transportation Engineering, Hohai University, and the Key Laboratory of Ministry of Education for Geomechanics and Embankment Engineering, Hohai University, for providing the necessary facilities and technical support for this research.

## Author contributions

X.N. conceived the study. X.N. and S.Q. developed the methodology. S.Q. and M.K. developed the software. M.K. performed the validation. S.Q. carried out the formal analysis and data curation and wrote the original draft. X.N. reviewed and edited the manuscript and acquired the funding. M.K. supervised and administered the project. All authors reviewed the manuscript.

## Funding

This research was funded by the Open Fund of Zhejiang Provincial Engineering Research Center for Urban Infrastructure Intelligence (grant number IUI2022-ZD-02).

## Competing interests

The authors declare no competing interests.

## Additional information

Supplementary Information accompanies this paper. Correspondence and requests for materials should be addressed to S.Q.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementary.docx