International Journal of Plasticity 201 (2026) 104687

Contents lists available at ScienceDirect

International Journal of Plasticity

journal homepage: www.elsevier.com/locate/ijplas

Differentiable constitutive modelling of granular media: From surrogate material model in explicit FEM-DEM to automatic elastoplastic calibration

Mengqi Wang [a], Tongming Qu [b], Shaoheng Guan [c,*], Y.T. Feng [d], Jidong Zhao [a,*]

[a] Department of Civil and Environmental Engineering, Hong Kong University of Science and Technology, Clearwater Bay, Kowloon, Hong Kong Special Administrative Region of China
[b] Institute of Water Engineering Sciences, Wuhan University, Wuhan, 430072, China
[c] School of Artificial Intelligence, Jingchu University of Technology, Jingmen, 448000, China
[d] Zienkiewicz Centre for Computational Engineering, Faculty of Science and Engineering, Swansea University, Swansea, Wales, SA1 8EP, UK

ARTICLE INFO

Keywords:
Automatic differentiation
Elastoplastic constitutive models
CSUH model
Stress-point integration
Parameter calibration
FEM-DEM
Granular materials

ABSTRACT

This study presents a novel differentiable framework for automated calibration of elastoplastic constitutive models using stress–strain data from various loading paths. The key innovation is a soft-gated stress-point integration scheme that smooths the elastic–plastic transition, enabling end-to-end automatic differentiation (AD) of phenomenological models. This framework leverages the Critical State Unified Hardening (CSUH) model as a principled 'surrogate' to replace computationally expensive discrete element method (DEM)-based representative volume element (RVE) in an explicit finite element method-discrete element method (exFEM-DEM) multiscale scheme. This CSUH-based surrogate is trained iteratively on extensive RVE data, with physics-based regularisation applied to the loss function to ensure all calibrated parameters remain physically interpretable. Results demonstrate that the trained surrogate accurately replicates granular responses from high-fidelity DEM simulations and generalises excellently across diverse boundary value problems (BVPs), achieving significant computational acceleration. A comparison with an alternative elastoplastic model (the enhanced-IME model) underscores the framework's versatility and the critical importance of model selection. This differentiable approach offers a robust pathway to substantially accelerate multiscale simulations and streamline the calibration of complex constitutive models.

1. Introduction

Granular media are ubiquitous in natural systems and engineering applications. Their engineering-scale response is typically analysed within a continuum framework, yet the macroscopic behaviour of granular media fundamentally originates from discrete particle-level interactions (Liu et al., 2020; Houdoux et al., 2021; Mei et al., 2023; W. Zhang et al., 2025). This discrete-continuum mismatch motivates multiscale strategies that bridge grain-scale mechanics within continuum solvers.

The FEM-DEM multiscale framework represents one such approach by coupling a discrete description of grain assemblies with a continuum-based FEM solver (Guo and Zhao, 2014, 2016; Guo et al., 2022). In such a multiscale method, the granular media

* Corresponding authors.
E-mail addresses: guanshaoheng@jcut.edu.cn (S. Guan), jzhao@ust.hk (J. Zhao).

https://doi.org/10.1016/j.ijplas.2026.104687
Received 7 December 2025; Received in revised form 25 March 2026
Available online 2 April 2026

are explicitly represented by DEM-based RVE models to compute the stress–strain response of each Gauss point, given the local deformation from the FEM mesh (Wang and Zhang, 2021; Hu et al., 2022). This setup enables the FEM-DEM approach to serve as an alternative to phenomenological constitutive models by directly bridging the transition from RVE-scale mechanical response to continuum-scale behaviour. However, the FEM-DEM multiscale approach is computationally intensive, as each Gauss point requires a concurrent DEM simulation, especially for irregularly shaped particles in the BVP analysis (Lai et al., 2022; Huang et al., 2024). Such computational cost increases rapidly with mesh refinement, severely limiting its applicability to large-scale simulations.

Recent advances in machine learning (ML) provide a promising alternative to alleviate this computational bottleneck by leveraging the data-fitting capabilities of neural networks (NNs) (Li et al., 2019; Fu et al., 2023). Various neural networks, such as single-step-based neural networks (Jordan et al., 2020; Wang et al., 2024b), recurrent neural networks (Qu et al., 2021; Ma et al., 2022; Tancogne-Dejean et al., 2021; Girard and Mohr, 2025), temporal convolutional neural networks (Wang et al., 2022; Q.Z. Guan et al., 2023), and graph neural networks (Mei et al., 2025), have been adopted to develop NN-based stress–strain surrogates or NN-based simulators. By replacing DEM solvers with such ML surrogates trained on RVE datasets, the resulting FEM-NN framework (Logarzo et al., 2021; Zhang et al., 2021) can achieve comparable performance to the corresponding FEM-DEM simulation while offering significant computational acceleration (Qu et al., 2023b).

A common feature of these pure data-fitting models is that they learn material behaviour directly from data without explicitly relying on established constitutive theories (Qu et al., 2023a). Therefore, their performance is highly sensitive to the training scenarios, and typically deteriorates markedly under unseen loading paths or boundary conditions (Qu et al., 2025). This limitation restricts their practical utility for solving diverse BVPs for granular materials, making their deployment to case-specific applications (S. Guan et al., 2023; Wang et al., 2024a; S. Guan et al., 2023), and raising concerns regarding interpretability and physical consistency.

In this work, rather than training a pure neural-network surrogate, an elastoplastic constitutive model, encapsulating elasticity, yield criteria, hardening laws, and plastic flow rules, is leveraged as a 'principled surrogate', serving as a simplified, fast-to-evaluate approximation of high-fidelity, computationally expensive granular systems. The term 'surrogate' is adopted here for two related reasons. First, the optimised constitutive model is deployed at Gauss points to replace repeated RVE calls in the FEM-DEM multiscale scheme, thereby reducing computational cost while retaining comparable macroscopic predictions, which aligns with the conventional objective of surrogate modelling. Additionally, the model parameters are identified from large collections of Gauss point stress–strain curves extracted from multiscale analyses in a training-like manner of a surrogate, so that the resulting constitutive update reproduces the RVE response with high accuracy in both interpolation and extrapolation regimes. Compared with black-box emulators, the proposed surrogate preserves mechanical consistency with the underlying granular assembly while remaining physically interpretable through its embedded elastoplastic constitutive skeleton.

Within the proposed framework, two primary challenges need to be addressed. Given numerous established constitutive models for granular materials, encompassing soils (Li et al., 2024; Yao et al., 2024; Zhao et al., 2024; Cheng and Yin, 2024) and rocks (Cao et al., 2020, 2022; Zhu et al., 2025), selecting a rational constitutive backbone is crucial. In parallel, conventional calibration strategies for constitutive parameters are generally designed for laboratory or macroscopic simulation data generated under triaxial or other regular loading scenarios (Mohr and Henn, 2007; Levasseur et al., 2008; Papon et al., 2012; Gras et al., 2017; Yin et al., 2018; Zhu and Chen, 2022; Gavris and Sun, 2025). By contrast, the RVE dataset generated by exFEM-DEM modelling at Gauss points exhibits complex, often non-smooth stress–strain histories, rendering traditional calibration strategies inadequate, necessitating the design of a robust parameter optimisation procedure for identifying the free parameters of the chosen principled surrogate with the collected RVE dataset.

In this work, to address these issues, a novel data-driven, automated calibration scheme for tunable parameters of constitutive models is proposed, in which different candidates are adopted as principled surrogates and systematically assessed for their ability to reproduce the mechanical response of granular assemblies. Furthermore, automatic differentiation (AD)-compatible modifications are also introduced into the constitutive skeleton to ensure a smooth and numerically stable optimisation process.

Automatic differentiation (AD) is not new as a computational technique and has been widely used in physics-informed neural networks (PINNs) and differentiable programming. In PINNs (Haghighat et al., 2021; Eghbalian et al., 2023; Eghtesad et al., 2024; Jadoon et al., 2025; Chen et al., 2025; N. Zhang et al., 2025), AD is primarily used to differentiate network-represented field variables with respect to space and time so that governing PDE residuals together with boundary and initial conditions can be constructed and minimised. The inferred quantities are typically a small set of coefficients or material parameters associated with relatively simple constitutive descriptions. In differentiable modelling, such as differentiable FE (Xue et al., 2023; Wang et al., 2025) or MPM (Zhao et al., 2026), AD is more commonly applied at the solver level by treating the simulator as a differentiable mapping and computing gradients of a misfit objective with respect to selected inputs, thereby enabling gradient-based inverse problems, usually for simplified material parameters or boundary and initial-condition variables. For critical-state-based, path-dependent soil elastoplastic constitutive models, however, systematic AD-based calibration remains relatively limited, and parameter identification is still more often performed using derivative-free global optimisers, such as genetic algorithms (Yin et al., 2018; Liu et al., 2024) and particle swarm optimisation (Tao et al., 2022; Granitzer et al., 2024), because non-smooth sensitivities arising from return mapping, yield-surface crossing, and internal-variable evolution can make gradient-based optimisation difficult to stabilise.

The present work targets this gap at the constitutive-update level rather than at the PDE-residual or whole-solver level. Specifically, the Gauss-point stress update of a physically structured critical-state constitutive model is recast into a stably differentiable computational graph with necessary AD-compatible modifications, so that its parameters can be calibrated end-to-end against RVE-generated stress–strain mappings under physical bounds and regularisation constraints. The resulting calibrated constitutive backbone is then reused across different macro-scale BVPs as a principled surrogate for the multiscale constitutive

mapping, which distinguishes the present framework from PINN-type equation-driven training and from solver-specific differentiable inverse analyses.

The remainder of the paper is organised as follows: Section 2 presents the basic framework for optimising the constitutive parameters of the principled surrogate, followed by a brief overview of the exFEM-DEM multiscale simulation method. Then it details the end-to-end parameter flow in the chosen surrogate and the implemented differentiability-oriented modifications to the surrogate to enable gradient descent-based calibration. Section 3 reports the sensitivity analysis of selected training hyperparameters and presents the calibration results for the tunable constitutive parameters of the surrogate, using datasets generated from exFEM–DEM footing simulations. In Section 4, the resulting surrogate is first validated at the RVE scale and then further applied to solve different BVPs to assess the accuracy and generalisation capability of the optimised surrogate. Section 5 reports the optimisation of two additional surrogate models (the MCC and the enhanced IME) and examines how the underlying constitutive skeleton of each surrogate affects the performance and behaviour of the proposed optimisation workflow. Conclusions are made in Section 6.

2. Methodology

2.1. A data-driven optimisation framework for determining elastoplastic constitutive parameters

This work aims to optimise parameters of an elastoplastic constitutive-based surrogate so that it faithfully reproduces the RVE stress–strain responses in exFEM–DEM analysis. Using the proposed data-driven optimisation scheme, the RVE stress–strain features extracted from explicit FEM-DEM modelling data are assimilated into the constitutive backbone, encompassing elasticity, yield criteria, plastic flow rules, and hardening. In contrast to developing a pure neural network surrogate to replace the DEM solver and accelerate multiscale modelling, the optimised constitutive surrogate maintains high computational efficiency and physical fidelity to the mechanical response of the RVE, and demonstrates strong generalisation beyond the training domain.

Fig. 1 outlines the optimisation workflow for the elastoplastic constitutive surrogate using data generated from exFEM-DEM modelling. In the proposed data-driven optimisation method, an exFEM-DEM strip footing modelling is first executed and, at each Gauss point (GP_i), the embedded RVE provides stress–strain sequences $[\boldsymbol{\epsilon}_i^{\mathrm{RVE}}, \boldsymbol{\sigma}_i^{\mathrm{RVE}}]$. The collected dataset $[\boldsymbol{\epsilon}_{\mathrm{RVE}}, \boldsymbol{\sigma}_{\mathrm{RVE}}]_{i=1}^{K}$ is then used to initialise and calibrate the free parameters $\boldsymbol{\varphi}^{(0)} = \left\{ \varphi_1^{(0)}, \varphi_2^{(0)}, \ldots, \varphi_m^{(0)} \right\}$ of the selected constitutive surrogate $\boldsymbol{\mathcal{M}}$, and K is the number of Gauss points in one exFEM-DEM modelling. Given $[\boldsymbol{\epsilon}_i^{\mathrm{RVE}}]_{i=1}^{K}$, the model prediction $\hat{\boldsymbol{\sigma}}_i = \boldsymbol{\mathcal{M}}(\boldsymbol{\epsilon}_i^{\mathrm{RVE}}, \boldsymbol{\varphi}^{(n)})$ is compared with $\boldsymbol{\sigma}_i^{\mathrm{RVE}}$ through a scalar loss $\boldsymbol{\mathcal{L}}$. The gradient of the trainable parameter vector $\boldsymbol{\varphi}^{(n)}$ is iteratively solved by the automatic differentiation (AD) and updated by the gradient descent algorithm until a convergence criterion is met, yielding the optimised model $\boldsymbol{\mathcal{OM}}$. The optimised surrogate $\boldsymbol{\mathcal{OM}}$ is finally validated on independent BVPs, including a biaxial compression, a retaining-wall and a cavity expansion cases, with the modelling result compared against corresponding exFEM-DEM simulations to assess its fidelity and generalisation. Additionally, the proposed framework is also validated under the 3D FE setting, where a 3D conventional triaxial compression (CTC) exFEM-DEM modelling is executed to generate the new training data and reoptimise the CSUH-based surrogate, which is then validated via a distinct 3D Plan-strain triaxial compression (PTC) BVP.

Such an optimisation method estimates all elastoplastic parameters jointly. This unified treatment avoids decoupled fitting of elastic and plastic parameters, accounting for parameter interactions consistently. Furthermore, the gradient descent-based optimisation process provides explicit sensitivity information of optimisable parameters to the mechanical behaviour of the RVE, thereby revealing how these elastic and plastic parameters interact to influence the mechanical response of the RVE. Additionally, the RVE at Gauss points experiences complex loading conditions, naturally forming a multi-pressure, multi-orientation, multi-path dataset that spans a wide range of loading paths and loading conditions. This ample training data facilitates the enhancement of the generalisation for an optimised model in the FE analysis.

2.2. The explicit FEM-DEM method

As outlined in Section 2.1, optimising constitutive model-based surrogate $\boldsymbol{\mathcal{M}}$ requires a reliable RVE-scale stress–strain dataset. In this work, such data are obtained through an FEM-DEM multiscale framework, which automatically produces diverse stress–strain loading paths associated with the specific BVP. These loading paths are kinematically affected by the surrounding deformation and thus are hard to obtain via pure element-based tests. Attributed to particle-scale description, these RVEs inherently capture mechanical behaviours of granular assemblies such as strain localisation and anisotropy evolution, thereby offering a more realistic representation of loading paths encountered in engineering simulations than those pre-designed loading paths.

To realise this data-generation strategy, an explicit FEM formulation is adopted within the multiscale framework. Different from the implicit FEM, explicit FEM advances the system solution without Newton–Raphson iterations, which involves the assembly of the tangent operator. This avoids the associated computational complexity, making the explicit multiscale approach better suited for generating training data for surrogate models.

The exFEM-DEM framework comprises two components: (i) the exFEM solver for computing nodal displacements in the FE mesh, and (ii) the DEM solver to evaluate the stress–strain response of each RVE embedded into the GP in the FE elements. In the exFEM solver, the displacement of FE nodes $\boldsymbol{u}$ over a domain Ω is governed by Newton's second law, which can be expressed in the weak

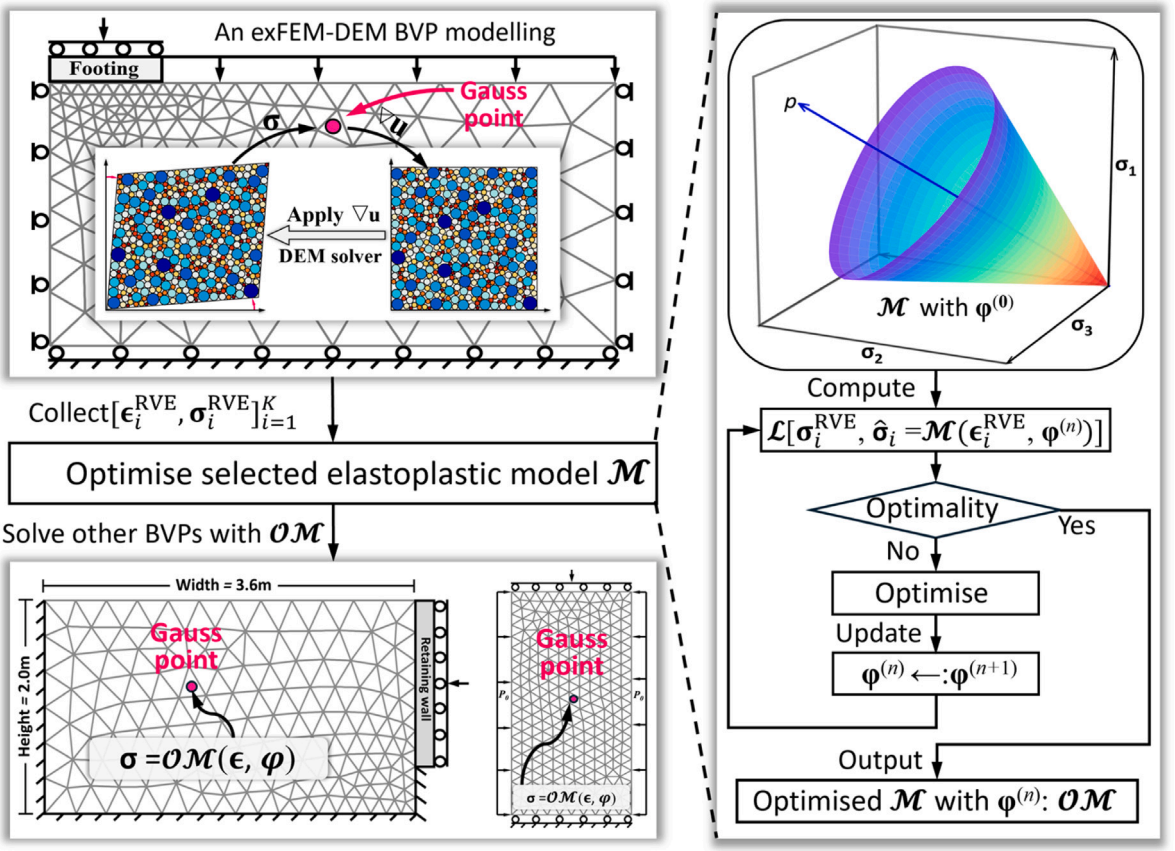


Fig. 1. The optimisation framework for tunable elastoplastic constitutive parameters in the principled surrogate.

form with the assistance of the Gauss theorem:

$$\int_{\Omega} \rho \ddot{\boldsymbol{u}}_t d\Omega = \int_{\Omega} \nabla \boldsymbol{u}_t : \boldsymbol{\sigma}_t dv - \int_{\Omega} \rho \boldsymbol{b} d\Omega - \int_{\Gamma} \boldsymbol{u}_t \cdot \boldsymbol{t} d\Gamma \tag{1}$$

where ρ is the bulk density; $\boldsymbol{b}$ and $\boldsymbol{t}$ represent body force applied to finite elements and traction imposed on the Neumann boundary Γ, respectively; and $\ddot{\boldsymbol{u}}$ denotes the nodal acceleration vector.

In the discretised FE system, element-level matrices, including mass matrix M^e, internal force vector F^e, boundary force T^e and body force vector B^e, are first computed for each element. These local matrices are then assembled into their corresponding global counterparts. Accordingly, the weak form in Eq. (1) can be expressed in the global matrix form to compute the $\ddot{\boldsymbol{u}}$:

$$\boldsymbol{M}\ddot{\boldsymbol{u}}|_t = \boldsymbol{F}|_t - \boldsymbol{B}|_t - \boldsymbol{T}|_t \tag{2}$$

where $\boldsymbol{M}$ is the global mass matrix; $\boldsymbol{F}|_t$, $\boldsymbol{B}|_t$ and $\boldsymbol{T}|_t$ correspond to right-hand side of Eq. (1) at the tth load step, respectively.

Given the obtained $\ddot{\boldsymbol{u}}$ in Eq. (2), the central difference method is leveraged to update the nodal displacement $\boldsymbol{u} \leftarrow \boldsymbol{u}|_{t+\Delta t}$ within a constant Δt as follows:

$$\begin{cases} \dot{\boldsymbol{u}}|_{t+0.5\Delta t} = \dot{\boldsymbol{u}}|_{t-0.5\Delta t} + \Delta t \ddot{\boldsymbol{u}}|_t \\ \boldsymbol{u}|_{t+\Delta t} = \boldsymbol{u}|_t + \Delta t \dot{\boldsymbol{u}}|_{t+0.5\Delta t} \end{cases} \tag{3}$$

where $\dot{\boldsymbol{u}}$ represents the velocity of the node, and the strain increment $\Delta\epsilon|_t$ can be obtained by:

$$\Delta\epsilon|_t = \nabla(\Delta \boldsymbol{u}|_t) = \nabla(\boldsymbol{u}|_{t+\Delta t} - \boldsymbol{u}|_t) \tag{4}$$

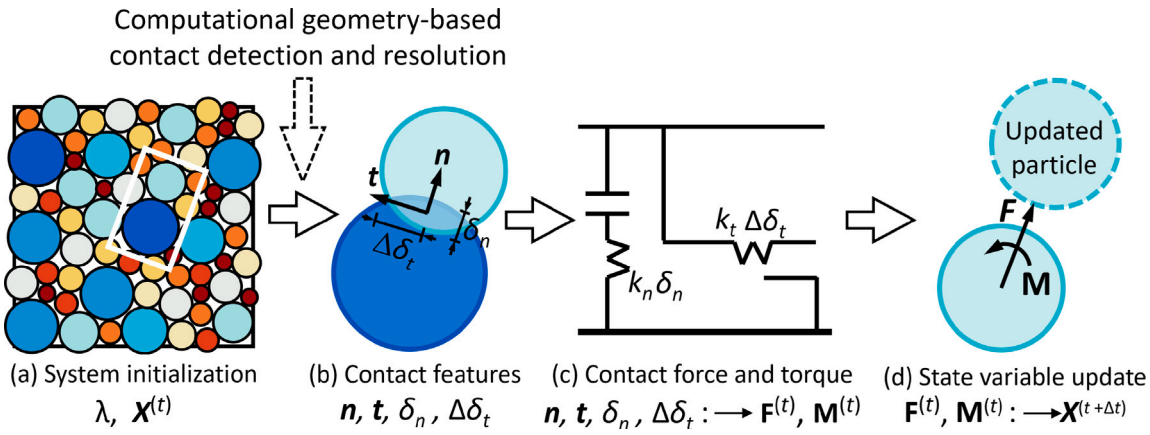


Fig. 2. The computation procedure of the DEM solver in the exFEM-DEM multiscale framework.

The detailed solving process of the exFEM solver is listed in Algorithm 1.

Algorithm 1 The workflow of the exFEM-DEM scheme

1: Discretise geometry domain into FE mesh
2: Initialise displacement field $\mathbf{u}|_{(0)}$
3: Get initial stress $\boldsymbol{\sigma}_{(0)}$ from the RVE
4: **for** t = 0, 1, 2, 3,..., N **do**
5: Apply the boundary condition of the t^{th} load step to the FE mesh
6: Compute the $\ddot{\boldsymbol{u}}|_t$ with Eq. (2)
7: Compute the $\Delta\boldsymbol{u}|_t$ and $\Delta\boldsymbol{\varepsilon}|_t$ with Eqs. (3) and (4)
8: Compute the $\Delta\boldsymbol{\sigma}|_t$ with the RVE and $\Delta\boldsymbol{\epsilon}|_t$
9: Update strain state: $\boldsymbol{\epsilon}|_{t+1} \leftarrow: \boldsymbol{\epsilon}|_t + \Delta\boldsymbol{\epsilon}|_t$
10: Update stress state: $\boldsymbol{\sigma}|_{t+1} \leftarrow: \boldsymbol{\sigma}|_t + \Delta\boldsymbol{\sigma}|_t$
11: Update nodal displacement: $\boldsymbol{u} \leftarrow: \boldsymbol{u}|_t + \Delta\boldsymbol{u}|_t$
12: $n = n + 1$
13: **end for**

As presented in Step 8 of Algorithm 1, the stress update at each Gauss point in the exFEM-DEM multiscale scheme is performed using a DEM-based RVE, which consists of different disks with a thickness of 0.1 mm under the 2D setting or different sphere particles under the 3D setting in this work. At the start of a macroscopic time step t, the RVE, characterised by material parameters λ such as the normal and tangential stiffnesses K_n and K_t, inherits grain-level particle configurations $\boldsymbol{X}^{(t)}$ (such as particle velocity and position) from the previous macroscopic step. The strain increment $\Delta\boldsymbol{\epsilon}|_t$, computed by the exFEM solver at the Gauss point, is applied as a boundary condition to the RVE.

Under the imposed deformation, the 2D-RVE system is advanced through multiple explicit micro-time steps. Within each micro-time increment $\Delta t'$, as shown in Fig. 2, interparticle contacts are first detected, and the corresponding geometric descriptors (such as contact normal $\boldsymbol{n}$, tangential direction $\boldsymbol{t}$, overlap δ_n) are computed. Then, contact forces $\boldsymbol{F}^{(t)}$ and torques $\boldsymbol{T}^{(t)}$ of each particle are evaluated based on the selected contact law (e.g., the Hertz–Mindlin model), which are then used to compute the stress tensor $\boldsymbol{\sigma}|_{t+\Delta t'} = \boldsymbol{\sigma}|_t + \Delta\boldsymbol{\sigma}|_{\Delta t'}$ of the RVE with the homogenisation assumptions. Finally, particle accelerations are determined by Newton's second law and are used to update particle configurations $\boldsymbol{X}^{(t)}$ to $\boldsymbol{X}^{(t+\Delta t')}$.

Such micro time-stepping continues until an equilibrium criterion is satisfied and the final RVE stress tensor $\boldsymbol{\sigma}|_{t+1}$ is passed back to the exFEM solver as the updated stress of Step 8 in Algorithm 1, thereby enabling the global equilibrium to be advanced to the next macroscopic time increment.

2.3. The end-to-end parameter flow in the chosen surrogate

2.3.1. Criteria for selecting principled surrogates

Selecting a constitutive model-based surrogate with sufficient expressiveness to reproduce the RVE stress–strain response is essential to implement the proposed scheme. In this work, the principled surrogate is chosen based on the following principles: (i) strong ability to describe different types of granular media under various states; (ii) fewer yet identifiable constitutive parameters to improve the stability of the calibration; (iii) strong physical interpretability; (iv) minimising square-root and division operations in

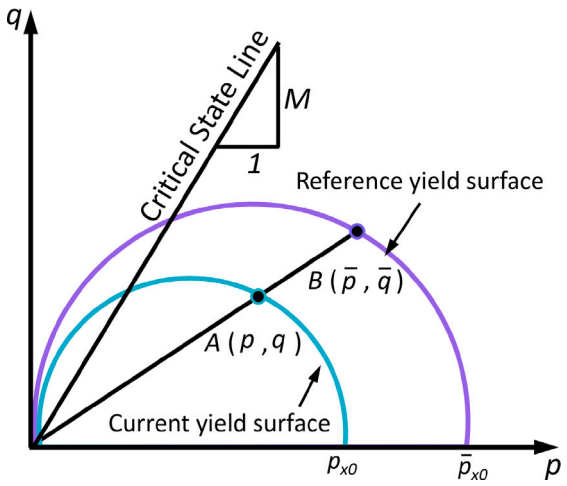


Fig. 3. The reference and current yield surface of the CSUH model.

constitutive formulations to enable a smoother stress-point integration, thereby remaining a numerically more stable optimisation process.

Guided by these criteria, an IME (Isotropic elasticity, von Mises yield surface, and Exponential hardening)-based surrogate is modified and first optimised within the proposed workflow. The optimisation procedure and comparative analyses for the enhanced IME surrogate are detailed in the Discussion section. These results demonstrate that, despite its computational simplicity, the enhanced IME surrogate is ultimately insufficiently expressive to reproduce the RVE stress–strain responses.

Alternatively, the Critical State Unified Hardening (CSUH) model (Yao et al., 2019) is adopted as the other principled surrogate in this work due to its ability to comprehensively describe the mechanical behaviour of diverse soil types and consolidation history, but with fewer material parameters. This unification is realised mainly by introducing a unified hardening function H, a modified normal compression line (NCL), and a density-dependent state variable ξ.

Briefly, the CSUH model comprises: (i) a reference surface $\bar{f}$ to represent the preconsolidation state; (ii) the current yield surface f to characterise the hardening-softening behaviour of soils, including over-consolidated clays and dense sands (see Fig. 3); (iii) the plastic potential function g with the non-association flow rule. The constitutive equations for the plastic part are formulated as follows:

$$\bar{f} = \ln\frac{\bar{p}}{\bar{p}_{x0}} + \ln\left(1 + \frac{\bar{q}^2}{M^2\bar{p}^2}\right) - \frac{\epsilon_v^p}{c_p} \tag{5}$$

$$f = \ln\frac{p + p_s}{p_{x0} + p_s} + \ln\left(1 + \frac{(1+\chi)q^2}{M^2p^2 - \chi q^2}\cdot\frac{p}{p + p_s}\right) - \frac{H}{c_p} \tag{6}$$

$$g = \ln\frac{p}{p_y} + \ln\left(1 + \frac{q^2}{M_c^2p^2}\right) \tag{7}$$

$$H = \int \frac{M_f^4 - \eta^4}{M_c^4 - \eta^4} d\epsilon_v^p \tag{8}$$

with stress invariants:

$$p = 1/3\mathrm{tr}\,(\boldsymbol{\sigma}), \quad q = \sqrt{3/2\mathbf{s}:\mathbf{s}}, \quad \mathbf{s} = \boldsymbol{\sigma} - p\mathbf{I}, \tag{9}$$

where, M is the critical state stress ratio; $c_p = (\lambda - \kappa)/(1 + e_0)$, with λ, κ and e_0 being the NCL slope, unloading slope and initial void ratio, respectively; p_s is the compressive hardening parameter; χ control the distortion of the current yield surface; and $\bar{p}_{x0}$, p_{x0} and p_y are the p-axis intersection of $\bar{f}$, f and g.

The potential failure ratio M_f, related to Hvorslev envelopes (Zhou and Yao, 2018), governs the hardening boundary and the hardening-softening transition behaviour of soils. The characteristic ratio M_c (a function of the state variable ξ) is specifically introduced to capture the state-dependent dilatancy behaviour of sands. Parameters M_f and M_c are respectively formulated as:

$$\begin{cases} M_f = 6\left(\sqrt{\frac{k}{R}(1+\frac{k}{R})} - \frac{k}{R}\right), \; k = \dfrac{M^2}{12(3-M)} \\ M_c = M\exp(-m\xi) \end{cases} \tag{10}$$

which extends the function of the unified hardening (UH) model (Yao et al., 2009) from clays to sands/granular soils. Where $\eta(= q/p)$ is the stress ratio; m is a dilatancy parameter; the over-consolidation parameter $R = p/\bar{p} = 1/OCR$ (over-consolidation-ratio) =

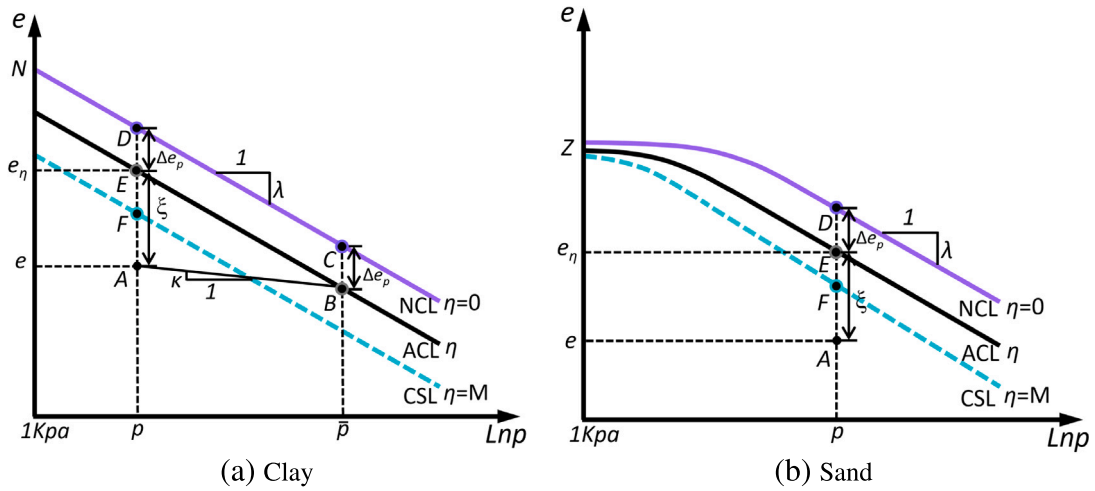


Fig. 4. The NCL, anisotropic compression line (ACL), critical state line (CSL), and state variable ξ.

$exp[-\xi/(\lambda-\kappa)]$, and $\xi = e_\eta - e$ as shown in Fig. 4(a). Furthermore, replacing the linear formulation in Eq. (11), a curved isotropic compression line (ICL), given in Eq. (12), is adopted as the NCL in the CSUH model, as shown in Fig. 4(b).

$$e = N - \lambda \ln p \tag{11}$$

$$e = Z - \lambda \ln\left(\frac{p + p_s}{1 + p_s}\right) \tag{12}$$

with parameter p_s being computed as:

$$p_s = \exp\left(\frac{N - Z}{\lambda}\right) - 1 \tag{13}$$

Correspondingly, e_η is formulated as:

$$e_\eta = Z - \lambda \ln\left(\frac{p + p_s}{1 + p_s}\right) - (\lambda - \kappa) \ln\left[\frac{\left(\frac{M^2+\eta^2}{M^2-\chi\eta^2}\right)p + p_s}{p + p_s}\right] \tag{14}$$

and according to the geometry relationship plotted in Fig. 4(a) e_0 at point A when $p = p_0$ can be computed as:

$$e_0 = e_{\eta,p=p_0} - (\lambda - \kappa)\ \ln\left(\frac{(p_0 * \mathrm{ocr}/1000 + p_s)}{(p_0/1000 + p_s)}\right) \tag{15}$$

where N and Z are the void-ratio at $p = 1$ kPa on the linear and curved NCLs, respectively.

Therefore, in the CSUH model, elastic and plastic parameters, including ν, λ, κ, N, Z, OCR, m, M, χ, as well as e_0 reflecting the OCR, are required to be calibrated. However, the RVE datasets reflect the mechanical response of the local material rather than a dense, well-distributed set of yield points in the $p-q$ space, making it difficult to calibrate a χ with physical meaning. Therefore, in this work, χ is directly fixed to zero.

2.3.2. *The automated parameter-optimisation workflow*

As outlined in Section 2.3.1, the parameter vector in the CSUH model is initialised as $\boldsymbol{\varphi}^{(0)}\ \left\{\kappa_0, \lambda_0, N_0, Z_0, M_0, m_0, \nu_0, e_0^{(0)}\right\}$ and subsequently optimised to reproduce the RVE stress–strain response. The stress-point integration process starts with an initial state: $\boldsymbol{\sigma}_0 = p\mathbf{I}, \boldsymbol{\epsilon}_0 = 0, H_0 = 0, \boldsymbol{\epsilon}_0^p = 0$, and $\eta_0 = 0$. The flow of these optimisable parameters within the whole integration at one Gauss point for a single $\Delta\boldsymbol{\epsilon}$ is presented in Fig. 5.

As shown in Fig. 5, constitutive constants in a surrogate $\boldsymbol{\varphi}^{(0)}$ are first converted to trainable $\boldsymbol{\varphi}$ by weights $\boldsymbol{W} = (w_\kappa, w_\lambda, w_N, w_z, w_m, w_\nu, w_{e0})$, and are used to computed two material constants $c_p(\lambda, \kappa, e_0)$ and $p_s(N, Z, \lambda)$. To accurately detect the onset of yielding and enhance the numerical stability of the optimisation process, the stress-point integration is iteratively executed in the sub-increment $\Delta\boldsymbol{\epsilon}_n$ for one strain increment $\Delta\boldsymbol{\epsilon}$:

$$T = \max\left(1,\ \left\lceil \frac{\|\Delta\boldsymbol{\epsilon}\|^2}{\epsilon_{\mathrm{ref}}} \right\rceil\right), \quad h = \frac{1}{T}, \quad \Delta\boldsymbol{\epsilon}_n = h\,\Delta\boldsymbol{\epsilon}, \quad n = 1, \ldots, T. \tag{16}$$

Coupled with current state variables $(\boldsymbol{\sigma}_n, \mathbb{D}_n, e_n, H_n, \xi_n, \boldsymbol{\epsilon}_n^p)$, the sub-strain $\Delta\boldsymbol{\epsilon}_n$ is then used to compute the trial stress $\boldsymbol{\sigma}_n^{tr}$, the current yield value f^{st} and trail yield value f^{tr}. These quantities determine the intersection factor r_{mid}, which locates the fraction of the sub-increment that remains within the yield surface. The computed r_{mid} are forwarded to the stress-point integration process to

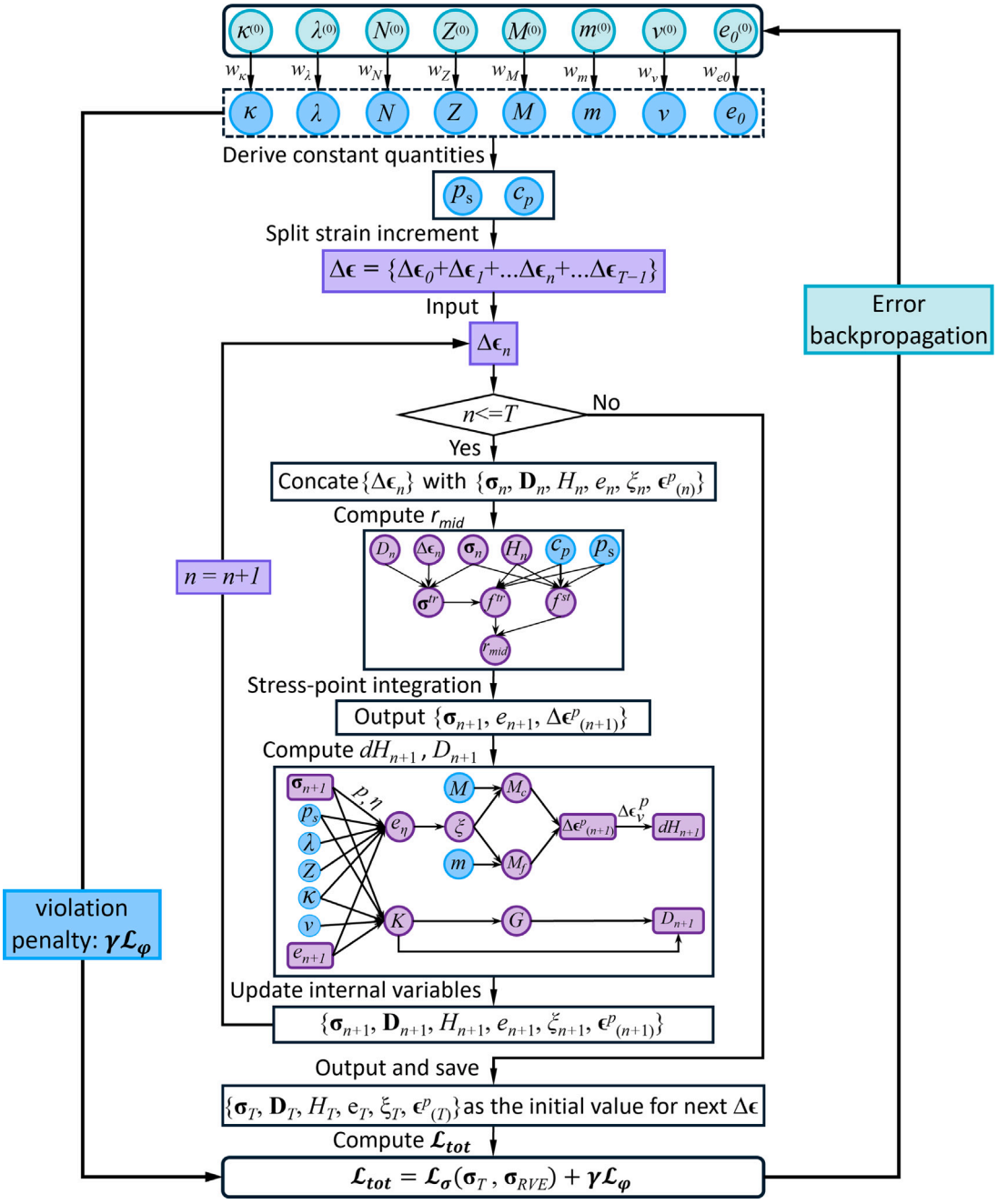


Fig. 5. The end-to-end parameter flow within the CSUH-based surrogate.

obtain $\boldsymbol{\sigma}_{n+1}$, e_{n+1}, and $\Delta\boldsymbol{\epsilon}^p_n$, which are further used to update the hardening parameter and material elastic matrix for the next sub-step. After iterating until $n > T$, the set $(\boldsymbol{\sigma}_T, \mathbb{D}^e_T, H_T, e_T, \xi_T, \boldsymbol{\epsilon}^p_T)$ is stored as the initial state for next $\Delta\boldsymbol{\epsilon}$ and used to compute the loss value.

To guarantee the optimised constitutive parameters within prescribed physical bounds, a penalty item $\mathcal{L}_{\varphi}$ is added to the total loss function, and only the out-of-range values are penalised. The normalised total loss function $\mathcal{L}_{tot}$ is expressed as:

$$\mathcal{L}_{tot} = \mathcal{L}_{\sigma} + \gamma\mathcal{L}_{\varphi} = \frac{\mathrm{MSE}(\boldsymbol{\sigma}_{\mathrm{T}}, \boldsymbol{\sigma}_{\mathrm{RVE}})}{\mathrm{mean}(\boldsymbol{\sigma}^2_{\mathrm{RVE}})} + \gamma \sum_{i \in params} \left[\left(\frac{\max(0, l_i - \varphi_i)}{(u_i - l_i)} \right) + \left(\frac{\max(0, \varphi_i - u_i)}{(u_i - l_i)} \right) \right] \tag{17}$$

Here, $\mathcal{L}_{\sigma}$ denotes the stress-misfit term, formulated as the mean-squared error (MSE) between the Gauss-point stress predicted by the surrogate CSUH model and the corresponding RVE stresses, given the current strain state and parameter set. The residuals are

Table 1
The limited domain for optimisable parameters in the CSUH-based surrogate.

Parameters	κ	λ	N	Z	m	ν	e_0	M
Lower bound	0.01	0.02	0.3	0.3	0.5	0.1	0.2	0.8
Upper bound	0.25	0.25	3.0	3.0	2.5	0.5	0.8	2.0

calculated component-wise at every retained load increment and Gauss point, and then averaged over the entire dataset. The loss is further normalised by $mean(\boldsymbol{\sigma}^2_{RVE})$ to obtain a dimensionless measure of stress discrepancy. Each optimisable parameter φ_i is bounded as $\varphi_i \in [l_i, u_i]$, with values of l_i and u_i referring to the work of Zhu and Chen (2022) and listed in Table 1. The range-normalisation term $(u_i - l_i)^2$ renders the penalty term dimensionless. The hyperparameter $\gamma > 0$, governing the strength of the bound enforcement and specified via an exponential rule that transitions from γ_{max} to γ_{min} by:

$$\gamma(\text{epoch}) = \gamma_{\max}\left(\frac{\gamma_{\min}}{\gamma_{\max}}\right)^{\text{epoch/total}} \tag{18}$$

2.4. Automatic differentiation-compatible modifications to the constitutive surrogate

To implement the flowchart in Section 2.3.2, a critical challenge is to render the stress-integration in the constitutive surrogate differentiable and numerically stable, thereby enabling an AD-based gradient-descent optimisation of the parameter vector $\boldsymbol{\varphi}$. In the CSUH model, however, the effect of the intermediate principal stress is incorporated via a stress-transformation procedure, which can amplify round-off errors and create singularities near degenerate stress states. In addition, the evaluation of r_{mid} in classical return-mapping calculations using stress-point integration is typically non-differentiable. Therefore, AD-compatible adjustments are introduced to the CSUH-based surrogate and its stress-point integration process, ensuring a stable data-driven calibration scheme.

2.4.1. Explicit parameterisation to intermediate principal stress ratio

In this section, the influence of the intermediate principal stress ratio $b = (\sigma_2 - \sigma_3)/(\sigma_1 - \sigma_3)$ (or equivalently the Lode angle θ) on the critical state stress is explicitly considered to ensure a stable and differentiable optimisation when optimising parameters with the RVE training data.

In the CSUH model, the influence of the intermediate principal stress coefficient b or Lode-angle θ on the critical state stress ratio is implicitly considered in a transformed stress (TS) space through mapping the curved-triangular SMP yield surface on the π-plane to a circular von Mises–type yield locus with a direction-preserving radial scaling $\alpha(\theta) = q_c/q$ (Yao and Wang, 2014; Yao et al., 2019):

$$\begin{cases} \tilde{\sigma}_{ij} = \sigma_{ij} & q = 0 \\ \tilde{\sigma}_{ij} = p\delta_{ij} + \alpha(\theta)\left(\sigma_{ij} - p\delta_{ij}\right) & q \neq 0 \end{cases} \tag{19}$$

where $q_c = 2I_1/(3\sqrt{(I_1I_2 - I_3)/(I_1I_2 - 9I_3)} - 1)$ with I_1, I_2 and I_3 are the invariants for stress tensor, and $\tilde{\sigma}_{ij}$ represents stress state in the TS space.

The TS mapping approach within the CSUH model offers a unified description of triaxial compression and extension on the π-plane, and the intermediate principal stress ratio or Lode-angle dependence is introduced implicitly via q_c, which is also referred to as $M_{\text{ts}}(\theta)$ in this study. However, a direct implementation of such an invariant-based TS formulation inside an AD-based data-driven calibration is often numerically fragile. The underlying reason is that the TS scaling factor $\alpha(\theta)$, evaluated from the invariant expression, depends on nested square-root operations and ratio terms, which can introduce ill-conditioned gradients and thereby induce noisy and unstable gradient updates in the proposed data-driven calibration framework.

To address this issue, an algebraically simpler and smoothly differentiable alternative $M_{\text{approx}}(\theta)$ is designed to explicitly describe the effects of the Lode angle θ on M in this work. A $M_{\text{approx}}(\theta)$ is proposed and used solely to improve AD robustness rather than to modify the underlying TS/SMP locus.

In the TS method, the transformed space can be mapped back to the original $p - q$ space with $p = \tilde{p}$ and $\tilde{q} = \alpha(\theta)q$ by:

$$\tilde{q} - M_{\text{com}}\tilde{p} = \alpha(\theta)q - M_{com.}p = q - M_{\text{approx}}(\theta)p = 0 \tag{20}$$

In this way, the relationship between M_{approx} and the Lode angle θ can be explicitly expressed:

$$M_{\text{approx}}(\theta) = M_{\text{com}}/(q_c/q) = M_{\text{com}}/\alpha(\theta) \tag{21}$$

where M_{com} is the stress ratio under triaxial compression, corresponding to the principal stress $\sigma_1 > \sigma_2 = \sigma_3$. To ensure the expression of $M_{\text{approx}}(\theta)$ smoother, the expression of stress ratio under compression M_{com} in the Mohr–Coulomb criterion is leveraged:

$$M_{\text{com}} = \frac{6\sin\phi}{3 - \sin\phi} \tag{22}$$

where ϕ is the frictional angle. Therefore, optimising the critical stress ratio is equivalent to calibrating the friction angle.

Once the $\alpha(\theta)$ is specified, the explicit $M_{\text{approx}}(\theta)$ formulation can generate the same yielding states as the circular locus in the SMP/TS space. Since the π-plane yield locus exhibits a $2\pi/3$ cyclic symmetry, the TS scaling factor $\alpha(\theta)$ should be representable as

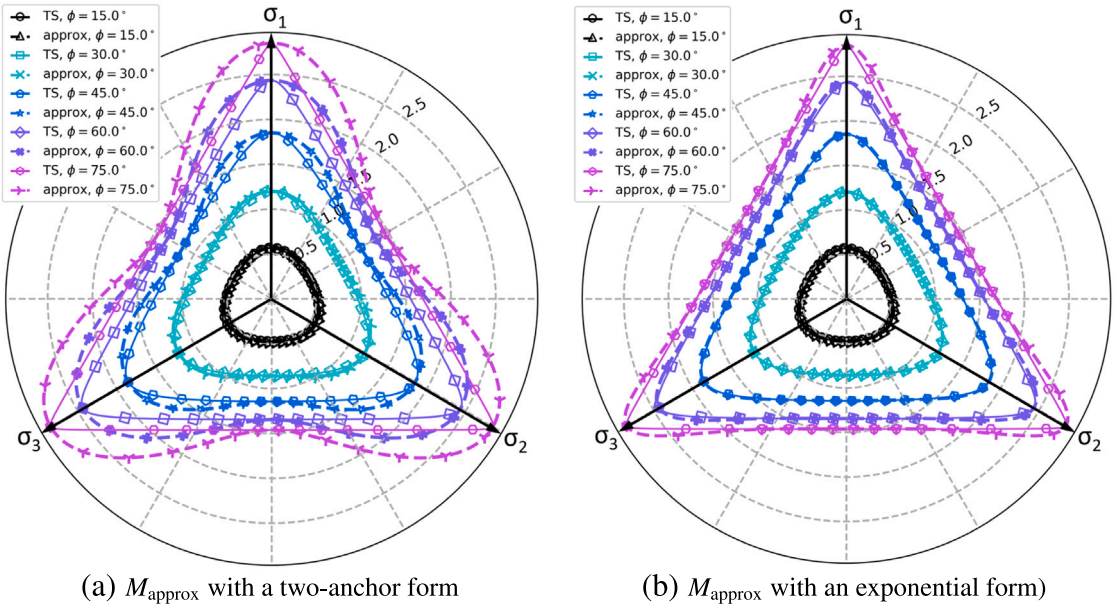


Fig. 6. Two different forms of the $M_{\text{approx}}(\theta)$ under different friction angle.

a periodic function of 3θ. To obtain an algebraically simple, continuously differentiable $M_{\text{approx}}(\theta)$, the following form is originally adopted to construct the formulation of $\alpha(\theta)$, referring to the work of (Pastor et al., 1990):

$$\alpha(\theta) = a + b\sin 3\theta \tag{23}$$

and the $M_{\text{approx}}(\theta)$ can be expressed as:

$$M_{\text{approx}}(\theta) = M_{\text{com}}/(a + b\sin 3\theta) \tag{24}$$

where a and b are determined from two anchors, specifically, the triaxial compression (TC) with $\theta = \pi/6$ and triaxial extension (TE) condition with $\theta = -\pi/6$:

$$\begin{cases} M_{\text{approx}}(\pi/6) = M_{\text{com}}/(a+b) = M_{\text{com}} \Rightarrow a + b = 1 & \text{TC}, \theta = \pi/6 \\ M_{\text{approx}}(-\pi/6) = M_{\text{com}}/(a-b) = M_e \Rightarrow a - b = r & \text{TE}, \theta = -\pi/6 \end{cases} \tag{25}$$

where r represents the compression-to-extension strength ratio $M_{\text{com}}/M_{\text{e}}$ and formulated as a friction-angle-dependent quantity $r = r(\phi)$:

$$r(\phi) = M_{\text{com}}/M_{\text{e}} \tag{26}$$

which is not treated as an independent trainable parameter in the AD framework, but as a derived shape-control variable. Therefore, the $M_{\text{approx}}(\theta)$ with two anchor points can be refreshed as:

$$M_{\text{approx}}(\theta) = \frac{2M_{\text{com}}}{(1+r) + (1-r)\sin 3\theta} \tag{27}$$

Specifically, once the friction-angle ϕ is determined, the M_{com} can be computed. Then, given the lode angle $\theta = -\pi/6$ and mean stress p, the corresponding value of q in the original stress space can be obtained from:

$$q_c(p, q, \theta) = M_{\text{com}} p \tag{28}$$

Since q_c is expressed in terms of stress invariants, its dependence on q is implicit and nonlinear, and thus the above equation cannot be inverted analytically to give an explicit expression of q. Therefore, given the prescribed $\theta = -\pi/6$, q is computed numerically by solving the nonlinear scalar equation:

$$F(q) = q_c(p, q, \theta) - M_{\text{com}} p = 0 \tag{29}$$

In this work, the root $q(\theta = -\pi/6)$ is located by a bracketing strategy and then solved robustly using the bisection method, which can be further used to compute the $M_{\text{e}} = r = M_{\text{ts}}(\theta = -\pi/6)$.

However, as shown in Fig. 6(a), such a two-anchor form of the $M_{\text{approx}}(\theta)$ remains limited. It can match M_{ts} exactly under TC and TE paths but still exhibit systematic discrepancies at intermediate Lode angles (e.g. near $\theta = 0$), which tend to be more visible when the calibrated friction angle is large.

Table 2
The detailed difference between two forms of the $M_{\text{approx}}(\theta)$.

$M_{\text{approx}}(\theta)$	Formula	The number of anchor points	Anchor positions	Parameters
Two-anchor form	$\frac{2M_{\text{com}}}{(1+r)+(1-r)\sin 3\theta}$	2	$\theta = 30°, -30°$	r
Exponential form	$M_{\text{com}}[\frac{2}{(1+r^{1/k})+(1-r^{1/k})\sin 3\theta}]^k$	3	$\theta = 30°, 0°, -30°$	$r,\ k$

Table 3
The absolute error of two M_{approx} formulations relative to M_{ts}.

$M_{\text{approx}}(\theta)$	$\theta = 15°$		$\theta = 0°$		$\theta = -15°$	
Two-anchor form	$\varphi = 15°$	Error = 0.014	$\varphi = 15°$	Error = 0.012	$\varphi = 15°$	Error = 0.011
	$\varphi = 45°$	Error = 0.164	$\varphi = 45°$	Error = 0.106	$\varphi = 45°$	Error = 0.095
	$\varphi = 75°$	Error = 0.486	$\varphi = 75°$	Error = 0.215	$\varphi = 75°$	Error = 0.189
Exponential form	$\varphi = 15°$	Error = 0.010	$\varphi = 15°$	Error = 0.000	$\varphi = 15°$	Error = 0.009
	$\varphi = 45°$	Error = 0.012	$\varphi = 45°$	Error = 0.000	$\varphi = 45°$	Error = 0.012
	$\varphi = 75°$	Error = 0.036	$\varphi = 75°$	Error = 0.000	$\varphi = 75°$	Error = 0.017

To improve global agreement while preserving smoothness and symmetry of $M_{\text{approx}}(\theta)$, a exponential transformation is introduced with:

$$M_{\text{approx}}(\theta) = \frac{M_{\text{com}}}{\alpha(\theta)^k}, \quad k > 0 \tag{30}$$

where k controls the curvature (sharpness/roundness) of deviatoric parts generated by the two-anchor form of $M_{\text{approx}}(\theta)$. However, a direct application of this transformation to Eq. (27) can alter the compression-to-extension strength ratio from r to r^k. To preserve the prescribed ratio $M_{\text{com}}/M_{\text{e}} = r$, an internal ratio r_{in} is introduced:

$$r_{\text{in}}^k = r \Rightarrow r_{\text{in}} = r^{1/k} \tag{31}$$

Then, Eq. (25) can be reformulated as:

$$\begin{cases} M_{\text{approx}}(\pi/6) = M_{\text{com}}/(a+b)^k = M_{\text{com}} \Rightarrow a+b=1 & \text{TC}, \theta = \pi/6 \\ M_{\text{approx}}(-\pi/6) = M_{\text{com}}/(a-b)^k = M_{\text{e}} \Rightarrow a-b = r_{\text{in}} = r^{1/k} & \text{TE}, \theta = -\pi/6 \end{cases} \tag{32}$$

and the modified two anchor form of the $M_{\text{approx}}(\theta)$ can be recomputed as:

$$M_{\text{approx}}(\theta) = M_{\text{c}}[\frac{2}{(1+r^{1/k})+(1-r^{1/k})\sin 3\theta}]^k \tag{33}$$

where k is also treated as a derived shape-control variable, and can be identified with another anchor point $M_{\text{ts}}(\theta = 0)$. The detailed difference between the two-anchor and the exponential form of $M_{\text{approx}}(\theta)$ is listed in Table 2

As shown in Fig. 6(b), introducing the curvature parameter k enables the $M_{\text{approx}}(\theta)$ to match the global shape of the TS/SMP locus more closely across the entire Lode-angle range, while retaining the required $2\pi/3$ cyclic symmetry to $\sin 3\theta$ and smooth differentiability under different friction angles.

Table 3 quantifies the absolute errors of the two M_{approx} formulations relative to M_{ts} under different Lode angles and friction angles. Since the M_{ts} values at $\theta = -30°$ and $30°$ are prescribed as the anchoring conditions in both two forms of M_{approx}, the errors at these two Lode angles are identically zero and therefore omitted. Accordingly, Table 3 reports only the errors at the Lode angles $\theta = -15°, 0°, 15°$ for friction angle $\varphi = 15°, 45°, 75°$ as shown in Fig. 6.

The results show that the exponential form of M_{approx} can significantly reduce the intermediate-Lode-angle errors, particularly under high friction angles where the two-anchor form exhibits large deviations. This demonstrates that the exponential mapping effectively compensates for the insufficient curvature/flexibility of the two-anchor interpolation between the prescribed compression–extension anchors.

2.4.2. A soft-gated stress-point integration scheme

A smooth, continuously differentiable stress-point integration process is crucial to the gradient-descent-based error backpropagation algorithm and to maintaining the numerical stability of the parameter calibration process.

Given a strain increment $\Delta\boldsymbol{\epsilon}$, the stress-point integration begins by computing the elastic trial stress by $\boldsymbol{\sigma}_{n+1}^{\text{tr}} = \boldsymbol{\sigma}_n + \mathbb{D}_n^{\text{e}} : \Delta\boldsymbol{\epsilon}$ with $\mathbb{D}_n^e$ being the elastic material matrix. The yield value is then evaluated at both the previous stress and the trial stress with $f^{\text{st}} = f(\boldsymbol{\sigma}_n, H_n)$ and $f^{\text{tr}} = f(\boldsymbol{\sigma}_{n+1}^{\text{tr}}, H_n)$, respectively.

Based on f^{st} and f^{tr}, the stress-point integration over a $\Delta\boldsymbol{\epsilon}$ is classified as (i) inside the yield surface, (ii) crossing the yield surface, and (iii) outside the yield surface, as illustrated in Fig. 7. Specifically, the increment is accepted elastically in Fig. 7(a). The plastic return mapping is required to correct the stress state when the stress-point integration crosses or over the current yield surface $f_n = 0$, as plotted in Fig. 7(b) and (c), respectively.

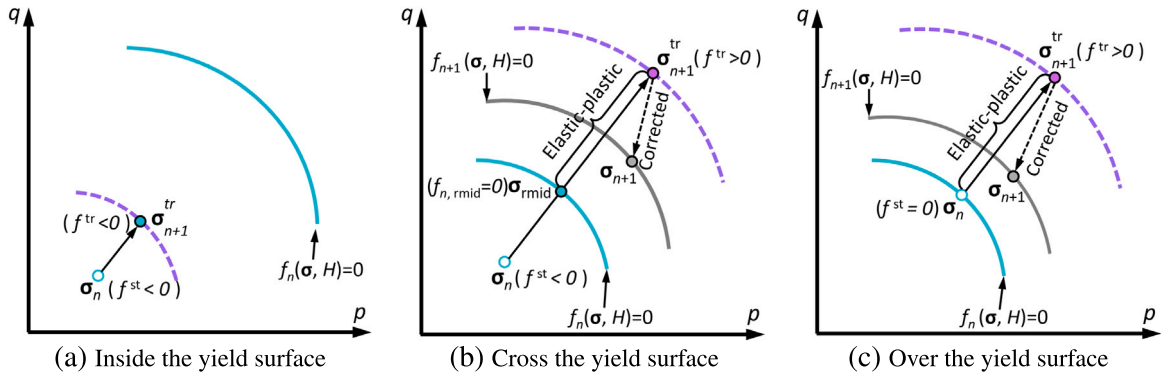


Fig. 7. Different situations within one stress-point integration.

As discussed in Section 2.3.2, the stress-point integration is iteratively executed in the sub-increment $\Delta\epsilon_n$ within a total strain increment $\Delta\epsilon$ with Eq. (16), and the stress-point integration can be unified into a single expression to describe three cases (inside, crossing, and outside the yield surface):

$$\sigma_{n+1} = \sigma_n + \mathbb{D}^{\mathrm{e}}_n : r_{\mathrm{mid}}\Delta\epsilon_n + \mathbb{D}^{\mathrm{e}}_{n,\mathrm{rmid}} : ((1 - r_{\mathrm{mid}})\Delta\epsilon_n - \Delta\epsilon^p_n) \tag{34}$$

where the plastic strain increment $\Delta\epsilon^p_t$ within $\Delta\epsilon_t$ follows the non-associated flow rule:

$$\Delta\epsilon^p = d\lambda\partial g_{n,\mathrm{rmid}}/\partial\sigma_{\mathrm{rmid}} \tag{35}$$

and the plastic factor $d\lambda$ is derived from the consistency condition:

$$df = \frac{\partial f_{n,\mathrm{rmid}}}{\partial\sigma_{\mathrm{rmid}}}d\sigma + \frac{\partial f_{n,\mathrm{rmid}}}{\partial H_n}dH = \frac{\partial f_{n,\mathrm{rmid}}}{\partial\sigma_{\mathrm{rmid}}}\mathbb{D}^{\mathrm{e}}_{n,\mathrm{rmid}}(d\epsilon_n - d\lambda\frac{\partial g_{n,\mathrm{rmid}}}{\partial\sigma_{\mathrm{rmid}}}) + \frac{\partial f_{n,\mathrm{rmid}}}{\partial H_n}\frac{\partial H_n}{\partial\epsilon^p_n}d\lambda\frac{\partial g_{n,\mathrm{rmid}}}{\partial\sigma_{\mathrm{rmid}}} = 0 \tag{36}$$

leading to $d\lambda$:

$$d\lambda = \left(\frac{\partial f_{n,\mathrm{rmid}}}{\partial\sigma_{\mathrm{rmid}}}\mathbb{D}^{\mathrm{e}}_{n,\mathrm{rmid}}d\epsilon_n\right) / \left(\frac{\partial f_{n,\mathrm{rmid}}}{\partial\sigma_{\mathrm{rmid}}}\mathbb{D}^{\mathrm{e}}_{n,\mathrm{rmid}}\frac{\partial g_{n,\mathrm{rmid}}}{\partial\sigma_{\mathrm{rmid}}} - \frac{\partial f_{n,\mathrm{rmid}}}{\partial H_n}\frac{\partial H_n}{\partial\epsilon^p_n}\frac{\partial g_{n,\mathrm{rmid}}}{\partial\sigma_{\mathrm{rmid}}}\right) \tag{37}$$

with $\sigma_{\mathrm{rmid}} = \sigma_n + \mathbb{D}^{\mathrm{e}}_n : r_{\mathrm{mid}}\Delta\epsilon_n$ and $f_{n,\mathrm{rmid}} = f(\sigma_{\mathrm{rmid}}, H_n)$

The elastic ratio $r_{\mathrm{mid}} \in [0, 1]$ in one $\Delta\epsilon_n$ in Eq. (34) serves as a switch between three situations within the stress-point integration and is controlled by values of both f^{st} and f^{tr}, and as:

$$r_{\mathrm{mid}} = \begin{cases} 1 & f^{tr} < 0 \\ (0, 1) & f^{st} < 0,\ f^{tr} > 0 \\ 1 & f^{st} > 0,\ f^{tr} > 0 \end{cases} \tag{38}$$

As the trial stress is computed under the elastic assumption, a linear interpolation along the elastic stress path is used to locate r_{mid} as an approximation to its exact solution within a tiny $\Delta\epsilon_t$ when $f^{\mathrm{st}} < 0$ and $f^{\mathrm{tr}}_{n+1} > 0$:

$$(1 - r_{\mathrm{mid}})f^{\mathrm{st}} + r_{\mathrm{mid}}f^{\mathrm{tr}} = 0 \Longrightarrow r_{\mathrm{mid}} = \frac{f^{\mathrm{st}}}{f^{\mathrm{st}} - f^{\mathrm{tr}}} \tag{39}$$

When $r_{\mathrm{mid}} = 0, or\ 1$, $\mathbb{D}^{\mathrm{e}}_{n,\mathrm{rmid}} = \mathbb{D}^{\mathrm{e}}_n$, $\sigma_{\mathrm{rmid}} = \sigma_n$, $f_{n,\mathrm{rmid}} = f^{st}$, and $g_{n,\mathrm{rmid}} = g_n$.

However, the above definition of r_{mid} renders the piecewise switch (inside/crossing/outside) non-differentiable at $r_{\mathrm{mid}} = 0$ and $r_{\mathrm{mid}} = 1$. Therefore, a cubic Hermite smooth operator $S(u)$ is applied to r_{mid} to ensure the interior of (0,1) remains unchanged while two endpoints are differentiable with an edge width δ_r:

$$S(u) = \begin{cases} 2u^2/\delta_r - u^3/\delta_r^2 & 0 < u \le \delta_r \\ u & \delta_r < u < 1 - \delta_r \\ 1 - [2(1-u)^2/\delta_r - (1-u)^3/\delta_r^2] & 1 - \delta_r \le u < 1 \end{cases} \tag{40}$$

and the rounded elastic ratio is therefore revised as:

$$r_{\mathrm{mid}} = \begin{cases} 1, & f^{\mathrm{tr}} < 0 \\ S(r_{\mathrm{mid}}), & f^{\mathrm{st}} < 0,\ f^{\mathrm{tr}} > 0 \\ 0, & f^{\mathrm{st}} > 0,\ f^{\mathrm{tr}} > 0 \end{cases} \tag{41}$$

Algorithm 2 The stress-point integration process with a local edge-rounding clipping

Given: strain increment $\Delta\boldsymbol{\epsilon}$ and void ratio e_n;
Compute elastic material matrix $\mathbb{D}_n^{\mathrm{e}}$
Split $\Delta\boldsymbol{\epsilon}$ into sub-increment $\Delta\boldsymbol{\epsilon}_n$
for $n = 0, 2, 3, ..., T-1$ **do**
  Compute $\boldsymbol{\sigma}_{n+1}^{\mathrm{tr}} \leftarrow: \boldsymbol{\sigma}_n + \mathbb{D}_n^{\mathrm{e}} : \Delta\boldsymbol{\epsilon}_n$
  Compute H_n with Eq. (8)
  Compute $f^{\mathrm{st}}(\boldsymbol{\sigma}_n^{\mathrm{tr}}, H_n)$, $f^{\mathrm{tr}}(\boldsymbol{\sigma}_{n+1}^{\mathrm{tr}}, H_n)$
  Compute r_{mid}
  if $r_{\mathrm{mid}} = 1$ **then**
    $\boldsymbol{\sigma}_{n+1} \leftarrow: \boldsymbol{\sigma}_{n+1}^{\mathrm{tr}}$
  else
    Update $\boldsymbol{\sigma}_n \leftarrow: \boldsymbol{\sigma}_{\mathrm{rmid}}$, $f_n \leftarrow: f_{n,\mathrm{rmid}}$, $e_n \leftarrow: e_{\mathrm{rmid}} = e_n - (1+e_n)r_{\mathrm{mid}} * \Delta\boldsymbol{\epsilon}_{n,v}$
    Compute $\mathbb{D}_{n,\mathrm{rmid}}^{\mathrm{e}}$
    Compute $d\lambda$ and $\Delta\boldsymbol{\epsilon}_n^p$,
    Correct $\boldsymbol{\sigma}_{n+1} \leftarrow: \boldsymbol{\sigma}_n + \mathbb{D}_n^{\mathrm{e}} : r_{\mathrm{mid}}\Delta\boldsymbol{\epsilon}_n + \mathbb{D}_{n,\mathrm{rmid}}^{\mathrm{e}} : ((1-r_{\mathrm{mid}})\Delta\boldsymbol{\epsilon}_n - \Delta\boldsymbol{\epsilon}_n^p)$
  end if
  Update $\boldsymbol{\sigma}_n \leftarrow: \boldsymbol{\sigma}_{n+1}$, $e_n \leftarrow: e_{n+1}$, $\boldsymbol{\epsilon}_n^p \leftarrow: \boldsymbol{\epsilon}_{n+1}^p$
  Update $\mathbb{D}_n^{\mathrm{e}} \leftarrow: \mathbb{D}_{n+1}^{\mathrm{e}}$
  Update $H_n \leftarrow: H_{n+1}$
  Update $f^{\mathrm{st}} \leftarrow: f(\boldsymbol{\sigma}_{n+1}, H_{n+1})$
  $n \leftarrow: n+1$
end for

The detailed implementation of the stress-point integration process in the CSUH model is listed in Algorithm 2.

2.4.3. Backpropagation-induced gradient flow during parameter updating

As exhibited in Fig. 5, the error-backpropagation scheme is leveraged to update the CSUH parameters after computing the normalised $\mathcal{L}_{tot}$. However, the stress-point integration in one $\Delta\boldsymbol{\epsilon}$ is iteratively executed within sub-strain increment $\Delta\boldsymbol{\epsilon}_n$. Furthermore, the supervision is available only at the terminal substep, i.e. a labelled stress $\boldsymbol{\sigma}_{\mathrm{RVE}}$ associated with $\Delta\boldsymbol{\epsilon}_n^T$. Therefore, the corresponding modification is required for the reverse AD process to ensure the gradient of the trainable weight can be transmitted around a total $\Delta\boldsymbol{\epsilon}$.

During the optimisation process, the normalised loss $\boldsymbol{\mathcal{L}}_{tot}$ is backforwarded to update the gradient of trainable weight $\boldsymbol{W}^{(t)}$ at the current moment t:

$$\frac{\partial \boldsymbol{\mathcal{L}}_{tot}}{\partial \boldsymbol{W}^{(t)}} = \frac{\partial \boldsymbol{\mathcal{L}}_{\boldsymbol{\sigma}}}{\partial \boldsymbol{\varphi}} : \frac{\partial \boldsymbol{\varphi}}{\partial \boldsymbol{W}^{(t)}} + \frac{\partial \boldsymbol{\mathcal{L}}_{\boldsymbol{\varphi}}}{\partial \boldsymbol{W}^{(t)}} = \boldsymbol{\varphi}^{(0)} \frac{\partial \boldsymbol{\mathcal{L}}}{\partial \boldsymbol{\sigma}_T} : \frac{\partial \boldsymbol{\sigma}_T}{\partial \boldsymbol{\varphi}} + \frac{\partial \boldsymbol{\mathcal{L}}_{\boldsymbol{\varphi}}}{\partial \boldsymbol{W}^{(t)}} \tag{42}$$

According to Eq. (34), $\boldsymbol{\sigma}_{n+1}$ is a function of $\boldsymbol{\sigma}_n$ and $\boldsymbol{\varphi}$:

$$\boldsymbol{\sigma}_{n+1} = \boldsymbol{\mathcal{F}}_n(\boldsymbol{\sigma}_n, \boldsymbol{\varphi}) \quad (n = 0, 1, \ldots, T-1) \tag{43}$$

As the computation of $\boldsymbol{\sigma}_{n+1}$ depends on $\boldsymbol{\sigma}_n$, thereby $\partial\boldsymbol{\sigma}_{n+1}/\partial\boldsymbol{\varphi}$ can be iteratively solved by:

$$\boldsymbol{J}_{n+1} = \frac{\partial \boldsymbol{\mathcal{F}}_n}{\partial \boldsymbol{\sigma}_n}\boldsymbol{J}_{\boldsymbol{n}} + \frac{\partial \boldsymbol{\mathcal{F}}_n}{\partial \boldsymbol{\varphi}} \quad with \quad \boldsymbol{J}_n \equiv \frac{\partial \boldsymbol{\sigma}_n}{\partial \boldsymbol{\varphi}}, G_n = \frac{\partial \boldsymbol{\mathcal{F}}_n}{\partial \boldsymbol{\sigma}_n} = \frac{\partial \boldsymbol{\sigma}_{n+1}}{\partial \boldsymbol{\sigma}_n}, C_n = \frac{\partial \boldsymbol{\mathcal{F}}_n}{\partial \boldsymbol{\varphi}}|_{\boldsymbol{\sigma}_n} = \frac{\boldsymbol{\sigma}_{n+1}}{\partial \boldsymbol{\varphi}}|_{\boldsymbol{\sigma}_n} \tag{44}$$

and accordingly, Eq. (42) can be rewritten as:

$$\frac{\partial \boldsymbol{\mathcal{L}}}{\partial \boldsymbol{W}^{(t)}} = \boldsymbol{\varphi}^{(0)} \frac{\partial \boldsymbol{\mathcal{L}}}{\partial \boldsymbol{\sigma}_T} : \sum_{n=0}^{T-1}\Big(\prod_{k=n+1}^{T-1} G_k\Big)C_n + \frac{\partial \boldsymbol{\mathcal{L}}_{\boldsymbol{\varphi}}}{\partial \boldsymbol{W}^{(t)}} \tag{45}$$

with:

$$C_n = \frac{\partial \mathbb{D}_n^{\mathrm{e}}}{\partial \boldsymbol{\varphi}} : r_{\mathrm{mid}}\Delta\boldsymbol{\epsilon}_n + \mathbb{D}_n^{\mathrm{e}} : \frac{\partial r_{\mathrm{mid}}}{\partial \boldsymbol{\varphi}}\Delta\boldsymbol{\epsilon}_n + \frac{\partial \mathbb{D}_{n,\mathrm{rmid}}^{\mathrm{e}}}{\partial \boldsymbol{\varphi}} : \boldsymbol{P}_n + \mathbb{D}_{n,\mathrm{rmid}}^{\mathrm{e}} : \Big(-\frac{\partial r_{\mathrm{mid}}}{\partial \boldsymbol{\varphi}}\Delta\boldsymbol{\epsilon}_n - \frac{\Delta\boldsymbol{\epsilon}_n^p}{\partial \boldsymbol{\varphi}}\Big) \tag{46}$$

$$G_n = \mathbf{I} + \frac{\partial \boldsymbol{\sigma}_{n+1}}{\partial r_{\mathrm{mid}}}\frac{\partial r_{\mathrm{mid}}}{\boldsymbol{\sigma}_n} + \frac{\partial \boldsymbol{\sigma}_{n+1}}{\partial \mathbb{D}_n^t}\frac{\partial \mathbb{D}_n^e}{\boldsymbol{\sigma}_n} + \frac{\partial \boldsymbol{\sigma}_{n+1}}{\partial \mathbb{D}_{n,\mathrm{rmid}}^{\mathrm{e}}}\frac{\partial \mathbb{D}_{n,\mathrm{rmid}}^{\mathrm{e}}}{\boldsymbol{\sigma}_n} + \frac{\partial \boldsymbol{\sigma}_{n+1}}{\partial \Delta\boldsymbol{\epsilon}_n^p}\frac{\partial \Delta\boldsymbol{\epsilon}_n^p}{\boldsymbol{\sigma}_n} \tag{47}$$

where $\boldsymbol{P}_n = (1-r_{\mathrm{mid}})\Delta\boldsymbol{\epsilon}_n - \Delta\boldsymbol{\epsilon}_n^p$. Then the gradient of $\boldsymbol{W}^{(t)}$ can be updated with the learning rate $\boldsymbol{\beta}$ until the final training epoch by:

$$\boldsymbol{W}^{(t+1)} = \boldsymbol{W}^{(t)} - \beta\frac{\partial \boldsymbol{\mathcal{L}}_{tot}}{\partial \boldsymbol{W}^{(t)}}_{tot} \tag{48}$$

where $\boldsymbol{\beta}$ is learning rate and related to the adopted optimiser.

Table 4
The parameters of each RVE in the exFEM-DEM simulation.

Parameter	Value
Particle number	1000
Particle size range (mm)	(3~7)
Young's modulus (MPa)	600
Density (kg/m^3)	2650
Particle frictional coefficient	0.5
Damping ratio	0.1

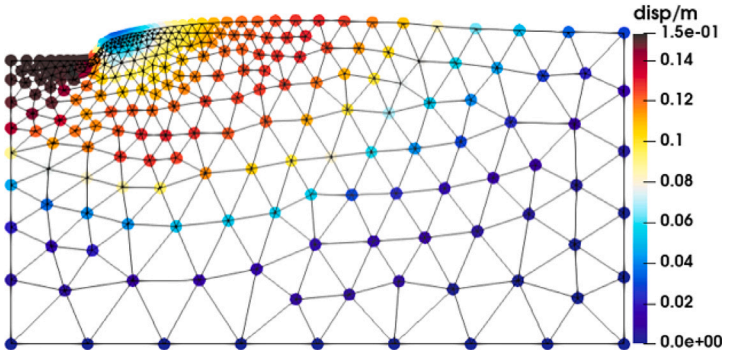


Fig. 8. The solved displacement field of the exFEM-DEM strip footing problem.

3. The optimisation for surrogate material mode

This section presents the detailed optimisation results for each elastic and plastic parameter mentioned in Section 2.3.1 using the collected RVE datasets generated by an exFEM-DEM modelling.

3.1. An explicit FEM-DEM strip footing modelling

In this section, a strip-footing problem is simulated within the exFEM-DEM framework under a 2D plane-stress FE setting with out-of-plane stress and strain components of zero, and the stress–strain curves of each RVE at the Gauss point are collected during the modelling, serving as the training set. The computational domain is a 4.0 m × 2.0 m rectangular domain with a 0.3 m rigid footing at its top-left corner. The domain is discretised using an unstructured triangular mesh; in total, 552 finite elements are generated with 552 Gauss-point RVEs embedded. The base boundary is fully fixed, whereas the two lateral boundaries are constrained in the horizontal direction and free in the vertical direction. A uniform surcharge P_0 = 100 kPa is applied to the rigid footing. A displacement-controlled vertical loading is imposed by prescribing a constant downward velocity of v = 0.1 m/s with a time increment of $\Delta t = 2 \times 10^{-4}$ s. The loading is advanced until a nominal loading strain of 7.5% is reached.

In the strip-footing exFEM-DEM modelling, disk-based (2D) RVEs are adopted to provide a well-defined in-plane macroscopic stress–strain response, to match the plane-stress FE setting. The detailed micro-parameters adopted for the RVE assemblies in the exFEM–DEM simulations are summarised in Table 4. Specifically, each RVE comprises 1000 disks with diameters in the range 6–12 mm and a thickness of 0.1 m, modelled as linear-elastic solids with a Young's modulus of 600 MPa and a density of 2650 kg/m^3. Inter-particle contact behaviour is characterised by the linear elastic–frictional contact model in `Yade` with a friction coefficient of 0.5, and a viscous damping ratio of 0.1 is introduced to represent energy dissipation and improve numerical stability. Prior to their use in the explicit FEM–DEM coupling, each RVE packing is first isotropically consolidated under hydrostatic confinement to a target mean effective stress of $p_0 = \sigma_0$ = 100 kPa, so that a well-defined, stress-equilibrated initial state is obtained for subsequent loading.

The explicit FEM-DEM strip footing modelling is implemented on a workstation equipped with a 12th-generation Intel Core i7 CPU (16 cores) and an NVIDIA GeForce GTX 2060 Ti GPU. A shared-memory parallel mode using the `multiprocessing` module at the Python library is leveraged, enabling concurrent RVE evaluations across Gauss points with multiple CPU cores. With these hardware and parallel settings, the complete strip-footing analysis, comprising 7500 load steps, required a total wall-clock time of 12.37 h. The strain and stress states of each load step are collected as the training set used to optimise the CSUH-based surrogate. The solved displacement field by the exFEM-DEM scheme is plotted in Fig. 8.

To further demonstrate the training data, Fig. 9 presents recorded stress–strain curves from representative Gauss points during the strip footing simulation, providing insight into the space of the training sample. Such RVE datasets capture stress–strain responses of local Gauss points in which the stress components σ_{11}, σ_{12}, and σ_{22} evolve jointly, thereby sampling a richer region of the stress–strain space than typical triaxial loading cases. This joint coverage provides additional constraints for both elastic and plastic parameters, enhancing the generalisation of the optimised model under complex loading conditions. Additionally, calibration with RVE data

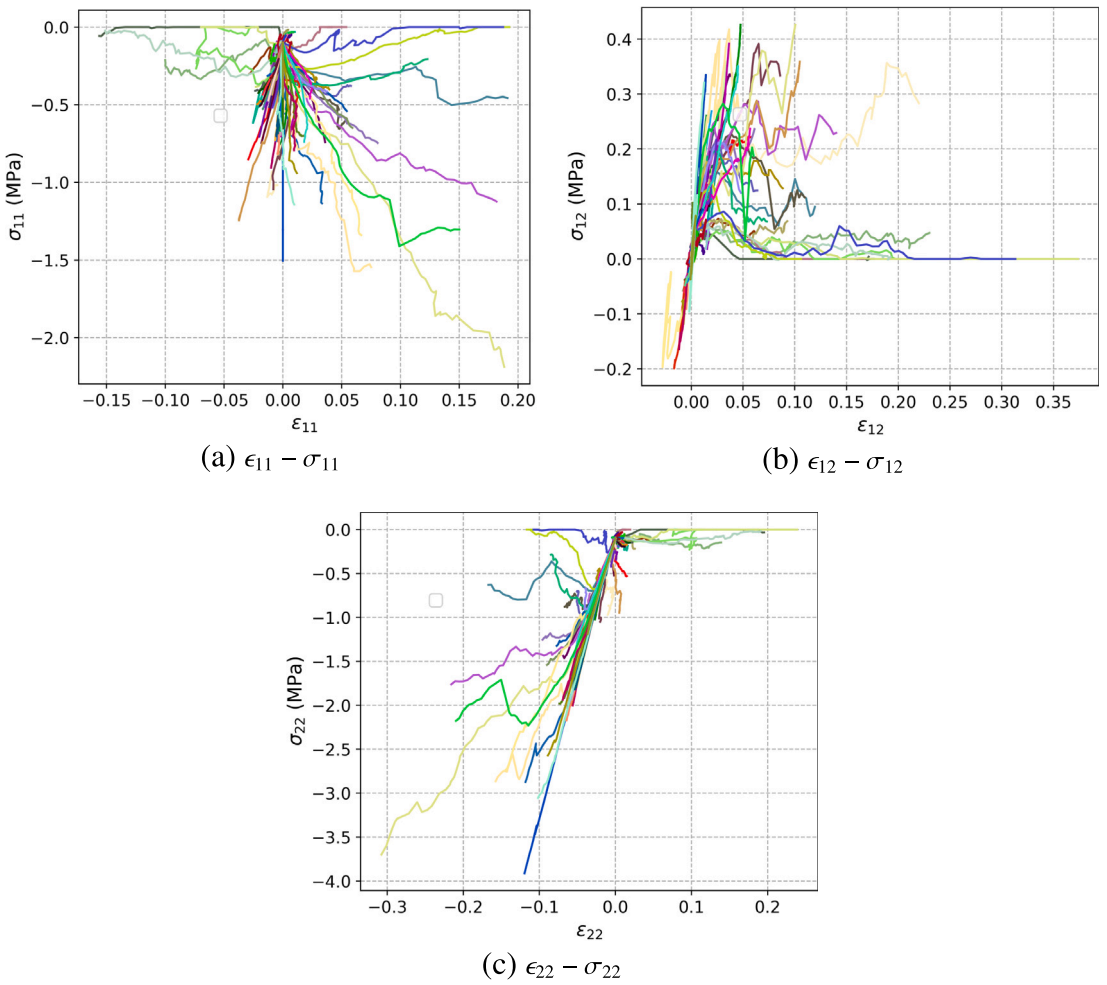


Fig. 9. Partial stress–strain space of training samples extracted from RVEs in the exFEM–DEM strip footing simulation.

requires minimising a sequence-level joint loss across all stress components over all time steps, rather than fitting a few feature points. This optimisation paradigm enforces tensorial consistency among the stress components, preserves path dependence and loading history, and avoids the information loss associated with feature selection, thereby yielding a more physically consistent and robust calibration.

3.2. The selection of the smoothing parameter in the soft-gated operator

The soft-gated stress-point integration in Section 2.4.2 employs a smooth operator $S(u)$ to replaces the non-differentiable, piecewise elastic–plastic switch at $r_{\text{mid}} = 0$ and $r_{\text{mid}} = 1$ with a smoothing parameter δ_r, serving as a hyper-parameter controlling the extent of smoothing. To select a suitable value of δ_r, an exFEM strip-footing simulation is performed using the original CSUH parameters with $\delta_r = 0$. During the simulation, the pairs $(f^{\text{st}}, f^{\text{tr}})$ are recorded and used as baseline samples to quantify how different δ_r affects the computed r_{mid}.

As the smoothing function $S(u)$ is activated only when the stress update involves an intersection with the yield surface within an increment, only the crossing cases where $f^{\text{st}} < 0$ and $f^{\text{tr}} > 0$ are focused as shown in Fig. 10(a) and (b), respectively. It is observed that more than 60% data fall within $f^{\text{st}} \in (-10^{-3}, 0)$ and $f^{\text{tr}} \in (0, 10^{-3})$. This indicates that the elastic trial stress typically crosses the yield surface only marginally, rendering r_{mid} numerically sensitive. This sensitivity can be amplified during backpropagation, necessitating the introduction of the smoothing operator $S(u)$.

Using the baseline dataset of f^{st}, f^{tr}, r_{mid} computed without smoothing ($\delta_r = 0$) are compared against the smoothed counterparts obtained with $\delta_r = 0.05$ and $\delta_r = 0.10$, and summarised in Fig. 11. The comparison result shows that increasing δ_r progressively modifies the near-endpoint region, where values of r_{mid} close to 0 are pushed upward, and values close to 1 are pulled downward. This implies that a systematic deviation from the original stress-point integration is introduced when δ_r is too large, which can

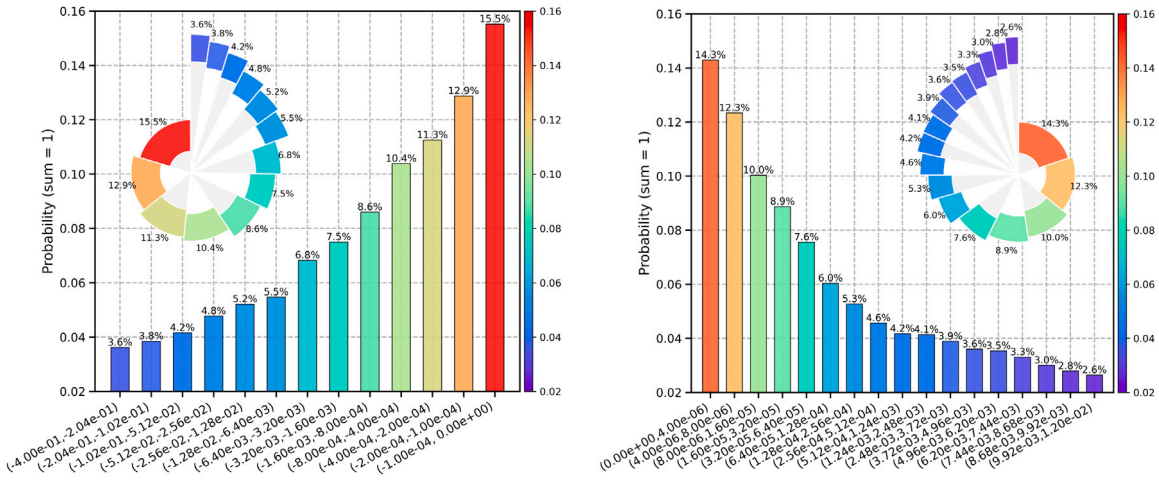


Fig. 10. The distribution of f^{st} and f^{tr} within crossing yield surface cases.

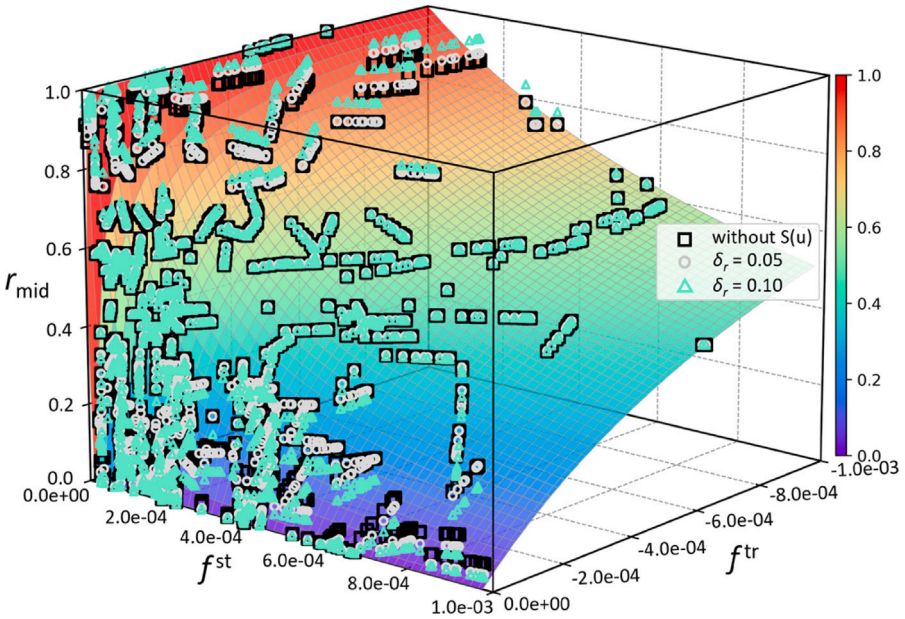


Fig. 11. The deviation of *rmid* with the increase of δ_r.

result in a noticeable bias in the stress update. Based on this sensitivity analysis, the upper bound for δ_r is restricted to 0.05, so that smoothing remains confined to a narrow neighbourhood of the endpoints and the induced integration bias is kept limited.

Then, the effect of δ_r on the smoothness of the elastic–plastic transition is assessed in Fig. 12, which compares the mapping $S(r_{\text{mid}}, \delta_r)$ for $\delta_r \in$ 0.01,0.02, 0.03, 0.04, 0.05 against the non-smoothed case. The comparison result shows that when δ_r is too small (e.g., δ_r = 0.01 or 0.02), the transition zone becomes extremely narrow, resulting in a limited practical smoothing effect. By contrast, δ_r values in the range of 0.03 to 0.05 produce a clearly smoother transition in the vicinity of r_{mid} = 0 and 1, consistent with the intended behaviour of the Hermite-based soft gate.

Taken together, the choice of δ_r reflects a balance between integration fidelity and differentiability. The sensitivity analysis in Fig. 11 indicates that δ_r should not exceed 0.05, as larger values introduce a systematic departure from the original stress-point integration and may bias the stress update. In contrast, the smoothness results in Fig. 12 show that excessively small δ_r yields insufficient smoothing, providing limited benefit for stabilising the AD process. Balancing these competing requirements, δ_r = 0.03 is finally adopted.

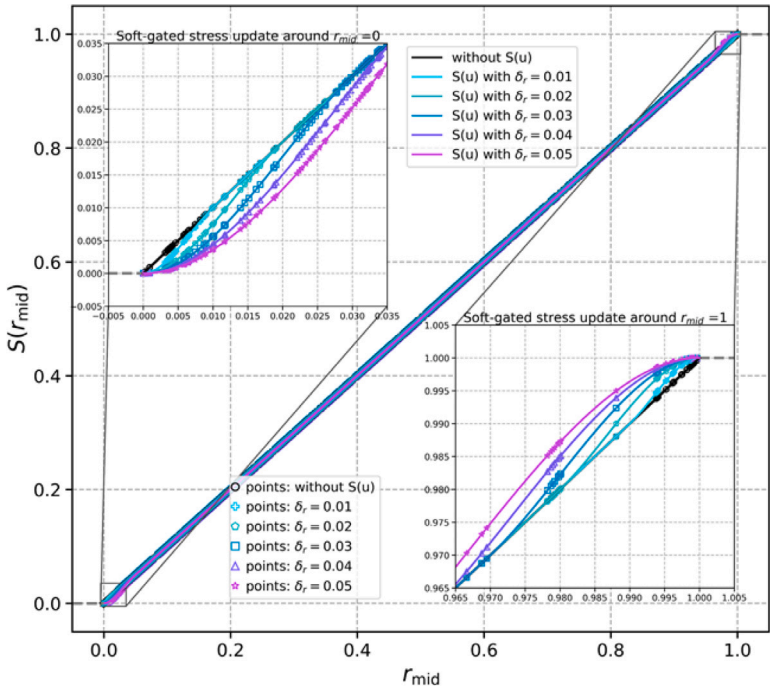


Fig. 12. The effect of δ_r on the smoothness of the elastic–plastic transition.

3.3. Optimisation results for elastoplastic parameters in CSUH-based surrogate

In Section 3.2, a numerically reasonable range for the smoothing width δ_r is identified using f^{st}, f^{tr} computed from the forward stress-point integration in the exFEM simulations. This range provides a principled guideline for selecting δ_r, thereby substantially reducing the need for extensive trial-and-error experiments against training-loss curves. In this section, to further refine the selection of δ_r and penalty coefficient γ, and to systematically assess the influence of the proposed AD-compatible modifications on optimisation smoothness, numerical stability, and convergence characteristics, three controlled comparative experiments are performed under distinct training schemes.

During the optimisation process, the early stopping technique is not adopted as the training loss curve exhibits frequent spike-like oscillations. An early-stopping criterion can therefore be triggered prematurely by transient loss fluctuations and terminate optimisation before the parameters reach a stable basin or before later-stage improvements are realised. Instead, each training scheme is run for up to 50,000 epochs with an initial learning rate of 0.01, and initial parameter values adopted from Yao et al. (2019), but with different treatment of the intermediate principal-stress coefficient and stress-point integration strategy. The surrogate in training scheme1 adopted the original CSUH transformed-stress (TS) formulation to account for the intermediate principal-stress coefficient and applies a stress-point integration strategy without the smoothing strategy; the stress-point integration is regularised using the soft-gated strategy in $scheme2$ while the original transformed-stress (TS) formulation is retained; both AD-compatible modifications in Section 2.4 are activated in training $scheme3$. For each training scheme, the optimisation is repeated with different random seeds, multiple values of δ_r and γ, thereby enabling robustness testing of the proposed AD-compatible modifications and providing a reference for selecting the final used δ_r and γ.

Fig. 13 presents training results for three training schemes. Each case is run for up to 50,000 epochs with an initial learning rate of 0.01. 80% RVE loading curves generated from the exFEM–DEM modelling are used to optimise the CSUH surrogate, with 10% of the stress–strain pairs reserved for validation to avoid overfitting during training. Before splitting, the data set is randomly shuffled. Training $scheme1$ is repeated five times in all data with different random seeds as shown in Fig. 13(a). The results show that training $scheme1$ consistently achieves a rapid reduction in early-stage loss in all random seeds, followed by a noticeably slower decay with intermittent needle-like oscillations.

In Fig. 13(b), $scheme2$ is assessed under five δ_r values within the identified range in Section 3.2, with the soft-gated stress-point integration enabled. Despite intermittent needle-like oscillations, the resulting loss is substantially lower, indicating improved optimisation stability due to the smoothed yield-state transition. An obvious δ_r-dependence of the training results is also identified. The final loss converges to comparable levels for δ_r = 0.01,0.02, and 0.03, whereas a larger smoothing value of δ_r = 0.05 yields a noticeably higher residual, with the training loss over the last 2000 epochs significantly increasing to approximately 0.020. This result indicates that, while the introduction of δ_r can improve the stability of the AD-based optimisation process, excessively large δ_r can adversely affect the performance of the final calibrated model. These observations provide further support for the choice of δ_r in Section 3.2.

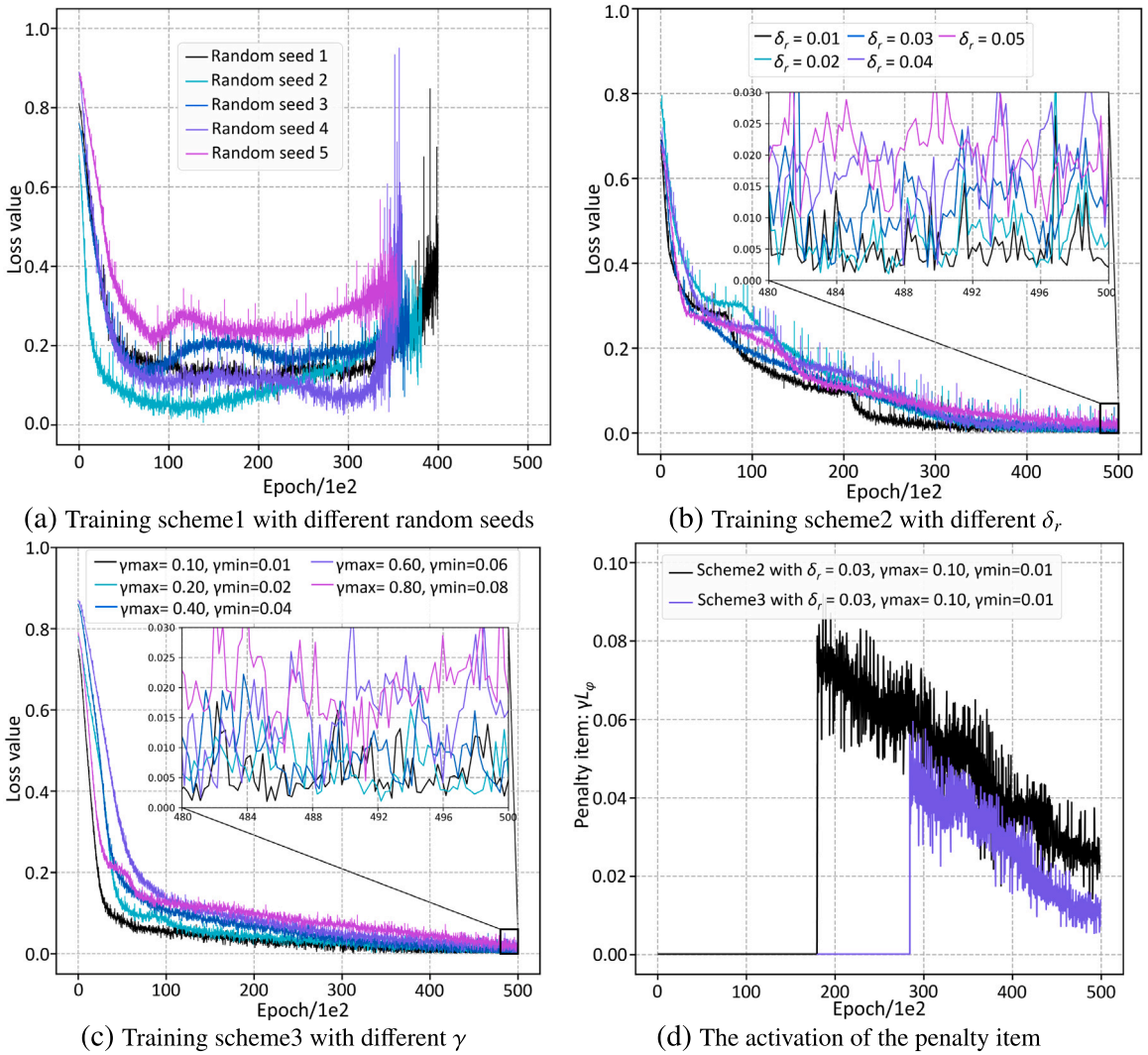


Fig. 13. The training procedure of different optimisation schemes for the CSUH-based surrogate.

With the surrogate further revised to include an explicit parameterisation of the intermediate principal-stress ratio, *scheme*3 exhibits faster convergence and the smallest loss variance across the tested penalty weights γ as shown in Fig. 13(c). The superior performance of *scheme*3 stems from the explicit treatment of the intermediate principal-stress coefficient, significantly reducing square-root and division operations embedded in the TS formulation, thereby providing more stable AD conditioning than *scheme*2.

The selection of γ used in *scheme*3 is guided by balancing the constraint enforcement against stress fitting by maintaining $\gamma\mathcal{L}_\varphi$ at a magnitude comparable to $\gamma\mathcal{L}_\sigma$ when violations occur. Using a penalty-free pre-training run as a reference scale, the order of γ used in *scheme*3 is set to 10^{-1}~10^{-2} to ensure an effective regularisation without degrading convergence stability. In both *scheme*2 and *scheme*3, the bound-penalty term $\gamma\mathcal{L}_\varphi$ is activated only when optimised parameters violate their prescribed ranges, otherwise it remains zero. As presented in Fig. 13(d), with the same δ_r and γ setting, $\gamma\mathcal{L}_\varphi$ is triggered earlier in training *scheme*2 than in *scheme*3. This earlier activation is consistent with the more prominent loss oscillations in *scheme*2, increasing the likelihood of overshooting bounded regions during parameter updates.

In addition to the loss evolution of the CSUH-based surrogate, Fig. 14 reports the epoch-wise evolution of several key elastoplastic parameters, including λ, κ, N, Z, M, and m for *scheme*2 and *scheme*3 under $\delta_r = 0.03$, $\gamma_{\min} = 0.01$ and $\gamma_{\max} = 0.1$. These results both confirm the effectiveness of the bound-penalty mechanism during optimisation and offer further insight into the physical interpretability and consistency of the training process.

Across all evolution results of constitutive parameters in Fig. 14, *scheme*2 and *scheme*3 exhibit parameter-update paths, yet their long-term trends are consistent, indicating the key CSUH parameters are identifiable under the RVE training constraints. Furthermore, several variables, including λ, Z, M, and m, in *scheme*2 repeatedly interact with the prescribed bounds, implying intermittent activation of the penalty mechanism when parameter updates leave the admissible domain, whereas the optimisation

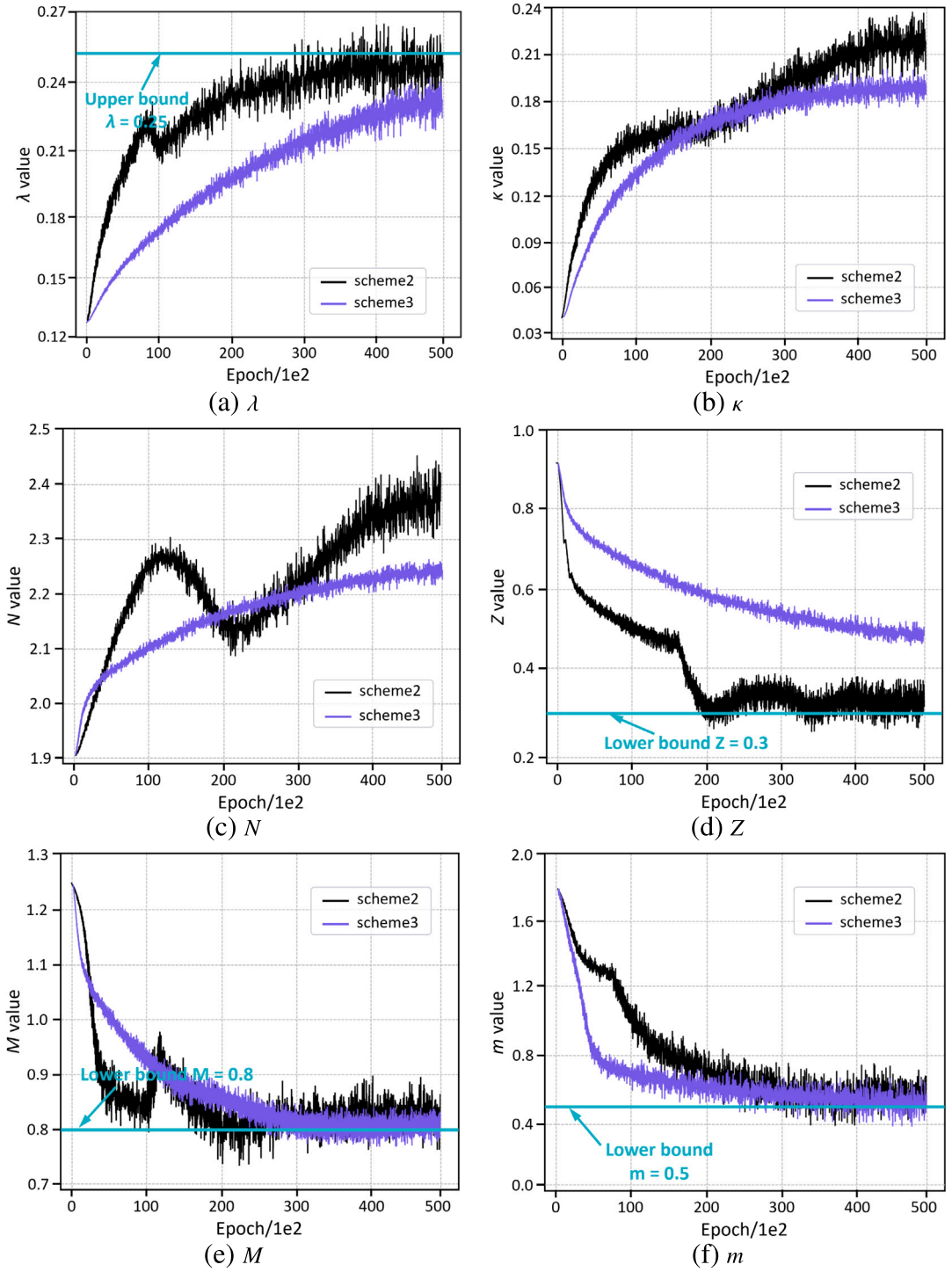


Fig. 14. The evolution process of optimisable constitutive parameters during the training process.

Table 5
The final adopted constitutive parameters in the CSUH-based surrogate.

Parameters	κ	λ	N	Z	m	ν	e_0	M
Original	0.04	0.135	1.9	0.9	1.8	0.2	0.721	1.25
Optimised	0.162	0.223	2.253	0.516	0.635	0.236	0.276	0.792

curves of most of these parameters smoothly converge to designated interior plateaus in $scheme3$. This observation is consistent with the loss and penalty evolutions in Fig. 13. Additionally, $scheme3$ shows smaller fluctuations and better monotonicity than $scheme2$. This indicates that the differentiability-oriented modification in Section 2.4 can effectively alleviate the non-smoothness gradient update during the error-backpropagation process, thereby improving optimisation stability and convergence efficiency.

More specifically, to reproduce the RVE response, the mild increases in λ and N together with the significant reduction in Z shown in Fig. 14(a), (c) and (d) lead to an increase of p_s via Eq. (13), indicating a larger curvature of the ICL, delaying the onset of prominent contractive behaviour (i.e., the later approach to the asymptotic NCL). This trend is consistent with RVE-scale micro-mechanisms in which early-stage contact-network rearrangement and evolution can absorb shear deformation with limited volumetric collapse, thereby postponing the emergence of apparent contractancy. Furthermore, the remarkable increase in κ (Fig. 14(b)) with a mild rise in λ steepens both the ACL and NCL while reducing their separation, yielding a smaller ξ at a given η. Together with this decrease in ξ, the strong reductions in M and m in Fig. 14(e) and (f) lead to an overall reduction in $M_c = M\exp(-m\xi)$, implying a shortened contractive regime and a more rapid transition from contractancy to dilatancy in the RVE response.

Finally, the parameter set at training epoch 48,690 of $Scheme3$, where the loss attains its minimum of 1.49e−3, is adopted and reported in Table 5. These constitutive parameters jointly govern the model's performance, and a comprehensive evaluation of the optimised model is presented using representative BVPs in the next section.

4. Numerical validation

In this section, the surrogate equipped with the calibrated constitutive parameters is validated against representative BVPs. Under the 2D setting, the assessment considers three benchmarks: biaxial compression, retaining-wall, and cavity expansion.

In addition, the proposed framework is extended to three dimensions by performing a conventional triaxial compression (CTC) exFEM–DEM simulation to generate the training dataset for recalibrating the CSUH-based surrogate. The CSUH parameters identified from the 2D strip-footing benchmark are not directly transferred to the 3D setting. The reason is that the training data produced by the 2D exFEM–DEM modelling with disk-based RVEs provides only an in-plane (2D) stress state, and no out-of-plane component is defined, which is fundamentally different from 3D cases. The re-optimised surrogate is subsequently evaluated on a distinct 3D BVP, namely plane-strain triaxial compression (PTC), to demonstrate its generalisation capability beyond the training loading paths.

For each BVP, predictions from the optimised and original constitutive models are benchmarked against the corresponding explicit FEM-DEM reference with the same mesh without introducing any implicit length scale or regularisation, evaluating the capability of the optimised model to reproduce the RVE mechanical response.

4.1. Biaxial compression

4.1.1. Validation results in the RVE scale

In this section, the biaxial compression problem is first leveraged to validate the optimised surrogate. Three biaxial compression modellings are conducted using the original CSUH model, the optimised CSUH model and the RVE model, providing the local stress–strain response, respectively. Each modelling is performed on a 0.05 m × 0.1 m rectangular specimen discretised with 484 unstructured triangular meshes shown in Fig. 15.

In the biaxial compression simulation, two lateral boundaries of the specimen are left traction-free, allowing for their deformation along both axial and lateral directions while a constant confining pressure of $P_0 = 100$ kPa is applied. Axial loading is imposed on the top surface in a displacement-controlled manner up to 10% axial strain, with a time increment of 2×10^{-4} and a loading rate of 0.01 m per step with 5000 increments.

The loading histories of six representative Gauss points marked in purple in Fig. 15 are plotted in Fig. 16. Compared with the original one, the optimised CSUH model reproduces the RVE stress–strain responses with markedly improved agreement. Minor discrepancies emerge in the later loading stage, where the predicted responses exhibit a mild drift from the RVE. Such deviations are consistent with post-peak behaviour, where strain localisation amplifies small perturbations in the evolution of internal variables, and the incremental stress-point updates accumulate the stepwise errors over long loading histories.

The stress–strain domain sampled under the biaxial compression modelling with the optimised-CSUH is compared with that represented by the strip-footing dataset employed for parameter optimisation in Fig. 17. The limited overlap between the two domains indicates that the two BVPs induce distinct deformation regimes and stress-path characteristics, supporting the extrapolative performance of the optimised CSUH model under loading conditions not represented in the calibration data.

Additionally, the orientation mismatch $\theta_\epsilon - \theta_\sigma$ of the optimised CSUH model is further evaluated, where θ_ϵ and θ_σ are defined by the principal directions of the strain and stress tensors, respectively. Fig. 18(a) shows that the original CSUH response yields a relatively smooth and weakly dispersed evolution along with equivalent strain ϵ_e. In comparison, the exFEM–DEM reference in Fig. 18(b) exhibits a larger and more scattered angular deviations. After data-driven calibration, the optimised CSUH model reproduces a misalignment-angle distribution that is closer to the DEM reference in both range and dispersion, as evidenced in Fig. 18(c), indicating improved agreement with the multiscale response.

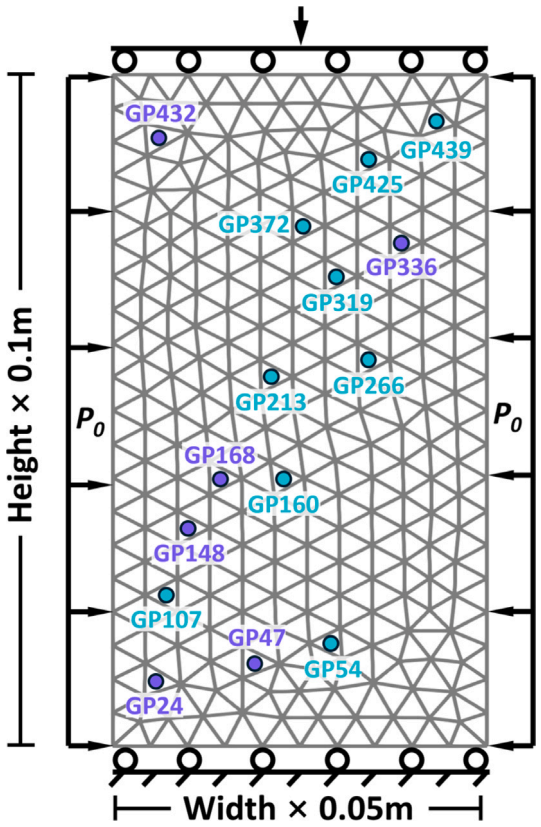


Fig. 15. The boundary conditions for the biaxial compression modelling.

4.1.2. The macroscopic modelling result of biaxial compression simulation

As discussed in the last section, the optimised CSUH model exhibits good consistency with both the physical characteristics and the stress–strain responses obtained from the RVE simulations. The following subsection further validates its performance through macroscopic modelling results of the biaxial compression BVP.

Fig. 19 demonstrates the shear strain localisation predicted by the exFEM-DEM strategy, the exFEM with the original CSUH, and the optimised CSUH model at different load steps. It is observed that the original CSUH model exhibits an overly narrow and strongly localised shear band with initial constitutive parameters. In contrast, the exFEM-optimised-CSUH approach, augmented with insights derived from RVE datasets, mitigates excessive peak localisation observed with the exFEM-Original-CSUH approach and predicts a more diffuse shear strain distribution on the same FE mesh, without introducing any explicit internal length scale, nonlocal formulation, or gradient regularisation. This response is more consistent with the shear band morphology observed in the reference exFEM-DEM simulation. The computational settings for each simulation, including the number of Gauss points, total loading steps, and wall-clock time, are listed in Table 6.

Such a transition is consistent with the optimised parameter set of the CSUH-based surrogate. A prominent increase in κ with a mild rise in λ steepens the ACL and NCL and reduces their separation, lowering sensitivity of $\xi = e_\eta - e = (\lambda - \kappa)\ln(1 + \frac{A(\eta)}{1+p_s})$ to the changes in the stress ratio η according to Eqs. (12) and (14). Consequently, ξ evolves more gradually along the loading path, limiting the minor perturbations of local states into abrupt changes in strength and flow characteristics and hence mitigating collapse into an overly thin localisation band. In parallel, the reduced M and m weaken the dependence of M_c on the state variable ξ (Eq. (10)), alleviating abrupt state-driven strength loss and suppressing the onset of a positive feedback loop in which local softening accelerates strain concentration. Finally, the mild increases in λ and N together with the significant reduction in Z increase p_s via Eq. (13), increasing the curvature of the ICL and allowing inelastic deformation to accumulate over a wider region, thereby delaying the onset of a distinct shear band.

Furthermore, the top force and volume strain obtained from three approaches are also presented in Fig. 20 to assess the performance of the optimised model in solving macroscopic BVPs. As shown in Fig. 20(a), the FEM-optimised-CSUH approach, enhanced with RVE-informed data features, displays a smoother post-peak decay, while its peak top-force is reduced and brought closer to the exFEM-DEM reference. This improvement is primarily attributable to the substantial decrease in M, which lowers the critical-state stress ratio and thereby reduces the peak resistance in the optimised CSUH. Meanwhile, the substantial increase in κ reduces the elastic bulk modulus K and consequently the shear modulus G, moderating pre-yield stress build-up and further limiting

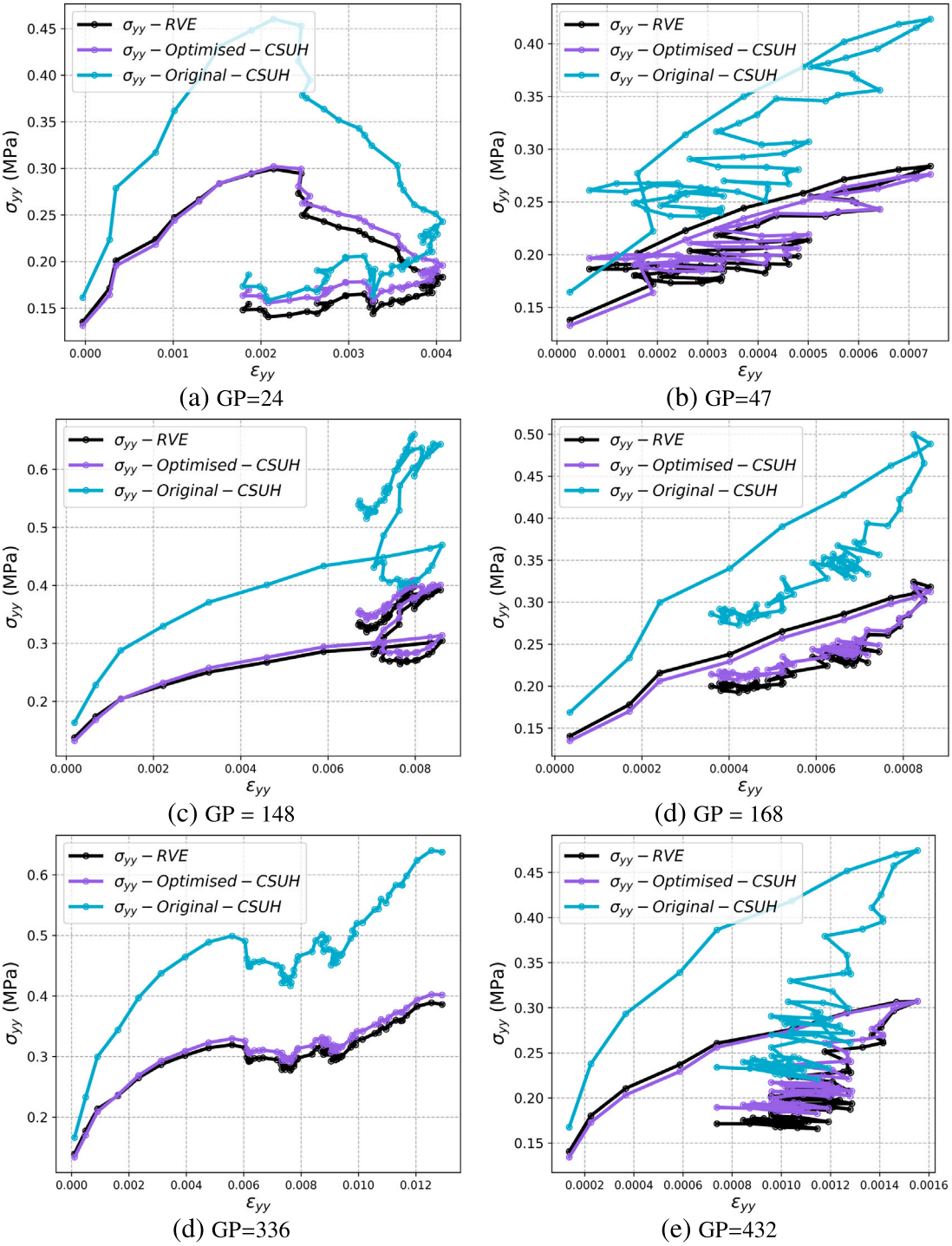


Fig. 16. The predicted stress of the CSUH-based data-driven model.

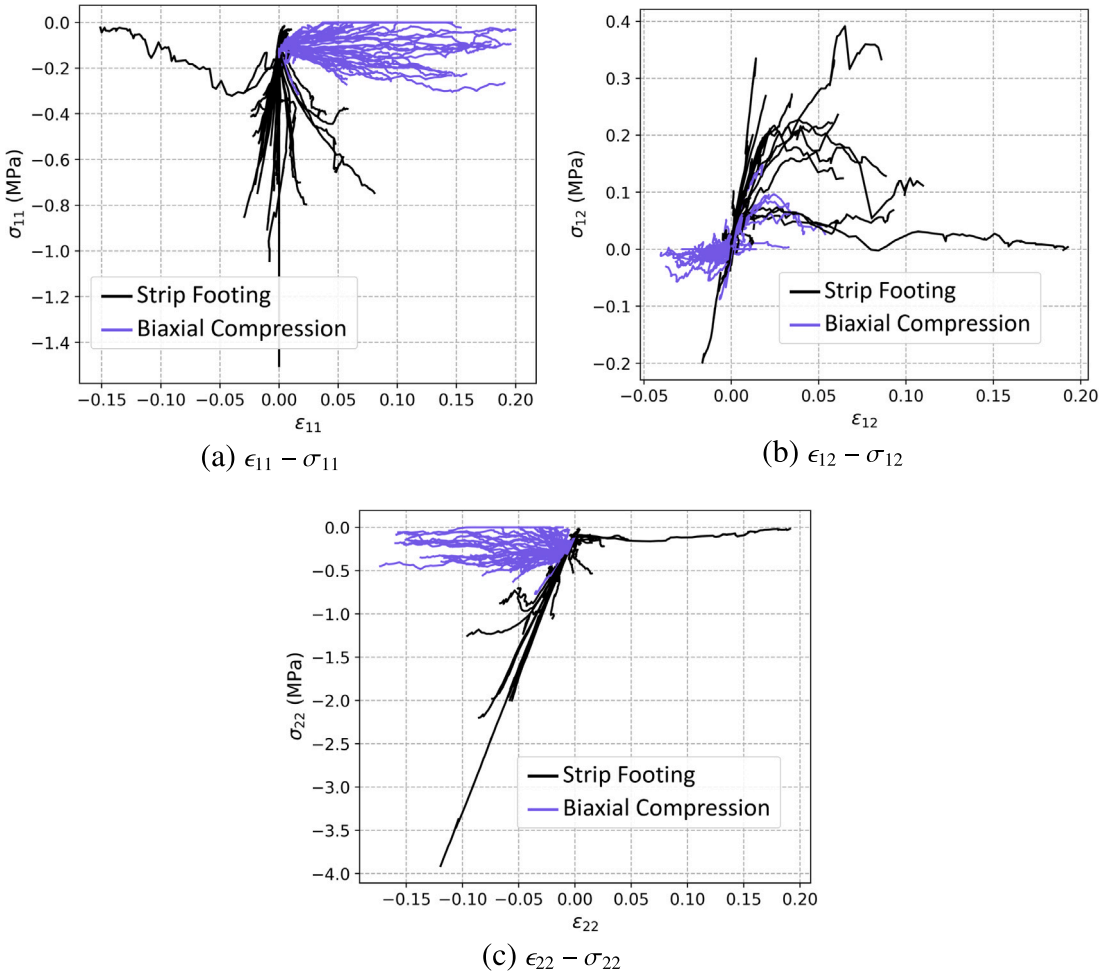


Fig. 17. The difference in loading space between the strip footing and the biaxial compression problems.

Table 6
The detailed experimental parameters of the three biaxial compression modelling.

Methods	Gauss points	Total steps	Running times/h	Overall MAE	Peak error/kN	Post-peak error/kN
exFEM-DEM	484	5000	10.974	\	\	\
exFEM-Original-CSUH	484	5000	0.146	24.31	33.02	22.76
exFEM-Optimised-CSUH	484	5000	0.142	6.92	3.57	8.23

the peak load. The overall mean average error (MAE), peak error, and post-peak error between the exFEM-DEM and exFEM-CSUH computation of the top force are listed in Table 6.

The volumetric response computed by the exFEM-optimised-CSUH approach also shows closer agreement with the exFEM-DEM benchmark in Fig. 20(b), characterised by an earlier contractancy-to-dilatancy transition and a smaller terminal volumetric strain. As the contractancy-dilatancy switch depends on the mobilisation of the stress ratio η relative to the M_c in Eq. (8), a smaller M_c triggers dilation at a smaller η. In addition, the decrease in $(\lambda - \kappa)$ attenuates the evolution of the state variable ξ. Given the state-dependent dilatancy in CSUH, the slower variation of ξ moderates post-peak dilation growth and results in a reduced terminal volumetric strain.

4.2. Retaining wall

The optimised CSUH model is also used to model the retaining wall problem to fully explore its capability to solve various BVPs. As shown in Fig. 21, the retaining wall problem is defined on a 3.6 m × 2.0 m domain discretised using 321 unstructured triangular

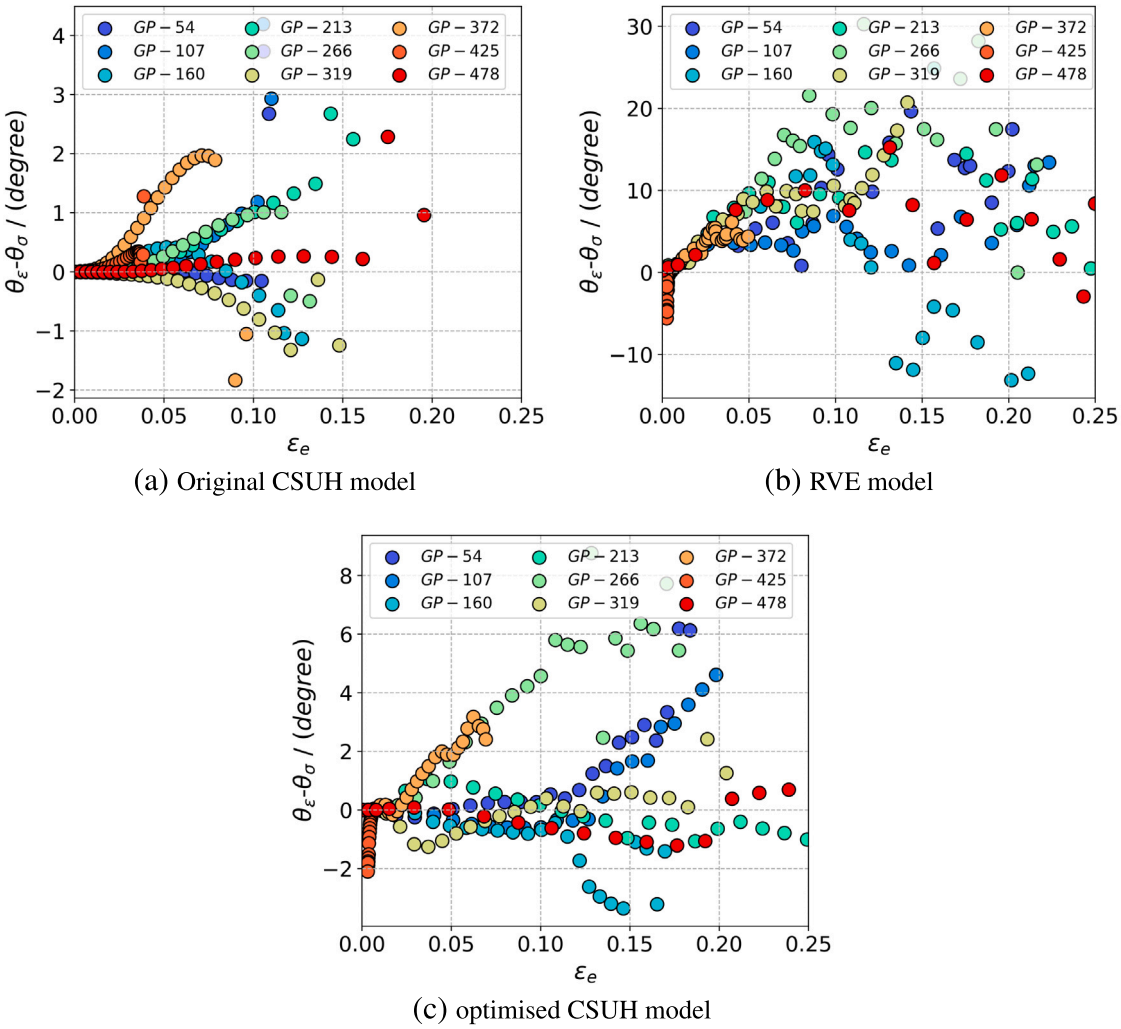


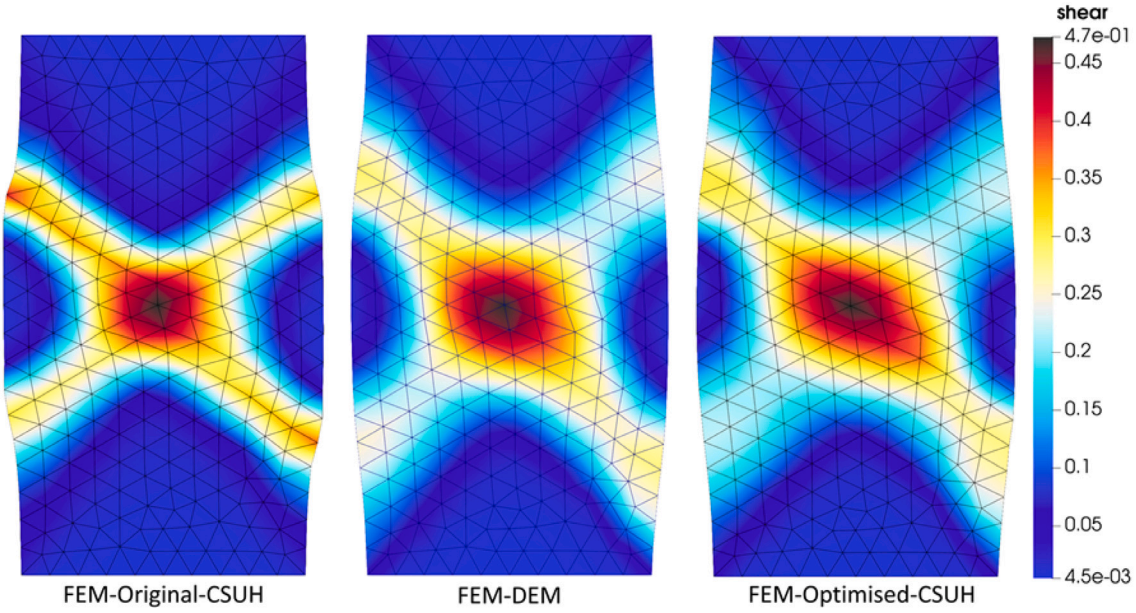
Fig. 18. Difference in angle between the principal direction of strain and stress tensor $\theta_\epsilon - \theta_\sigma$.

finite elements. The base of the geometry domain is fully fixed in both horizontal and vertical directions, and the top boundary is traction-free. To enforce plane strain conditions, the left boundary is restrained in the horizontal direction. A rigid retaining wall with a height of 1.5 m is positioned along the right boundary of the domain, and the left part of the right boundary is fully fixed. The interface between the wall and the soil is assumed to be rough, preventing the relative vertical displacement between the soil and the rigid wall with a uniformly distributed surcharge of $P_0 = 100$ kPa applied to the wall. To model active wall movement, a displacement-controlled loading is applied horizontally to the wall until a total displacement of 0.26 m is reached.

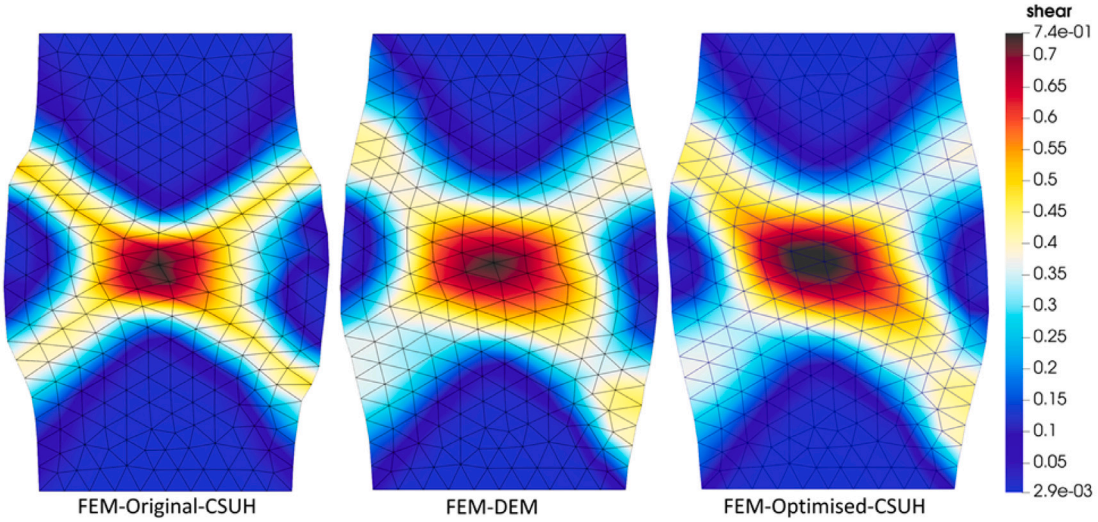
Three methods used in Section 4.1 are employed again to solve the displacement fields of the retaining wall problem, respectively, and the simulation results at different load steps are presented in Fig. 22. In both Fig. 22(a) and (b), the exFEM-Original-CSUH framework predicts a relatively smaller and more dispersed deformation zone, especially in the upper backfill region, where the displacement field is underestimated compared to the FEM-DEM benchmark. In contrast, the exFEM-optimised-CSUH model shows a significantly improved match with the exFEM-DEM modelling results, reproducing the magnitude pattern of the displacement field with greater accuracy. The computational settings for each simulation, including the number of Gauss points, total loading steps, and wall-clock time, are listed in Table 7.

This improvement is attributed to the optimised parameter set, which recalibrates the strength and stiffness scales of the CSUH surrogate using the RVE-derived stress–strain dataset. The reduced M, together with m, decreases the overall shear-strength level and weakens state sensitivity, thereby alleviating excessive post-yield shear localisation and enabling a more distributed deformation pattern along the top surface of the specimen. In addition, the substantial increase in κ lowers the overall stiffness of the system, which moderates pre-yield stress accumulation and limits the behaviour of deformation constraint. With the optimised parameters, the optimised continuum CSUH can produce surface displacements that are in closer agreement with the exFEM–DEM benchmark.

Fig. 23 compares the stress–strain ranges covered by the optimised-CSUH model in the retaining-wall analysis with those in the strip-footing dataset used for calibration. The overlap between the two is small, showing that the different boundary conditions



(a) Load step = 100



(b) Load step = 200

Fig. 19. The solved shear strain fields by three different schemes in the biaxial compression problem.

Table 7
The detailed experimental parameters of the three retaining wall modelling.

Methods	Gauss points	Total steps	Running times/h	Overall MAE	Peak error/kN	Post-peak error/kN
exFEM-DEM	321	6000	9.316	\	\	\
exFEM-Original-CSUH	321	6000	0.133	79.61	97.86	52.68
exFEM-Optimised-CSUH	321	6000	0.131	14.96	21.76	7.49

in these BVPs lead to different deformation patterns and stress-path trends. The retaining-wall case, therefore, evaluates the model mainly beyond the range of the calibration data, serving as a demanding test of its robustness under loading conditions not included in the parameter fitting.

Additionally, Fig. 24 reports the evolution of top force and volumetric strain to loading strain for the retaining-wall problem using three strategies. The trends observed in Fig. 24(a) and (b) closely resemble those previously reported for the biaxial compression

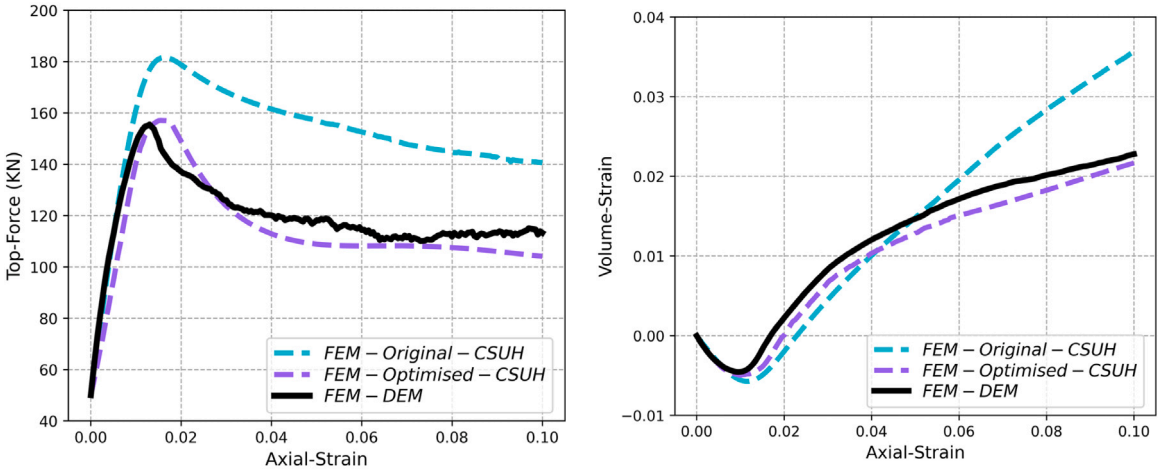


Fig. 20. The top force and volume strain of the sample in the biaxial compression problem.

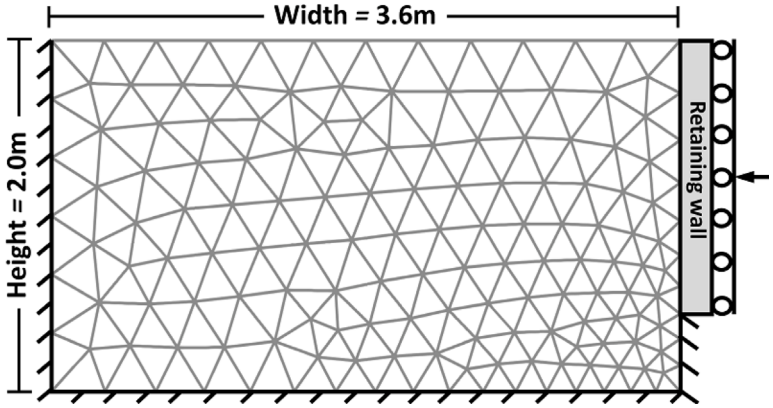


Fig. 21. The boundary condition of the retaining wall modelling.

problem in Fig. 20. The overall mean average error (MAE), peak error, and post-peak error between the exFEM-DEM and exFEM-CSUH computation of the top force are listed in Table 7. Specifically, the exFEM-optimised-CSUH approach, informed by RVE-scale data, consistently yields a smoother post-peak softening response in top force and a more gradual volumetric contraction. The underlying mechanisms are therefore analogous to those discussed for the biaxial compression problem and are not repeated here. These results further validate the capability of the optimised CSUH model in capturing the progressive deformation behaviour of granular materials under different BVPs.

4.3. Cylindrical cavity expansion

The optimised CSUH model is further assessed using the cavity expansion problem (Guo et al., 2016). As shown in Fig. 25, only a quarter annulus is modelled to reduce the computational cost. The specimen is idealised as an annulus with inner radius $r_c = 8$ mm and outer radius $r_o = 150$ mm. A uniform confining pressure $P_0 = 100$ kPa is imposed on the outer circular boundary ($r = r_0$) as a traction condition. The inner boundary ($r = r_c$) is subjected to a prescribed outward radial displacement, directed from the inner cavity towards the outer ring to enlarge the cavity, through a displacement-controlled monotonic loading $u_c(t)$, until the maximum radial displacement reaches 10 mm.

During the cavity expansion modelling, tangential traction and tangential displacement are not prescribed on either circular boundary, such that the response is primarily governed by radial loading. The normal displacement is constrained by $u_x = 0$ on $x = 0$ and by $u_y = 0$ on $y = 0$, with the corresponding tangential motion remaining free. To assess the robustness of the exFEM-optimised CSUH scheme under a more complex setting, an additional tangential constraint is imposed on the top corner of the upper straight boundary. This local restriction breaks the ideal radial symmetry and introduces an intentionally asymmetric boundary condition, thereby generating a more complicated shear-band pattern for comparison with the corresponding exFEM–DEM simulation results.

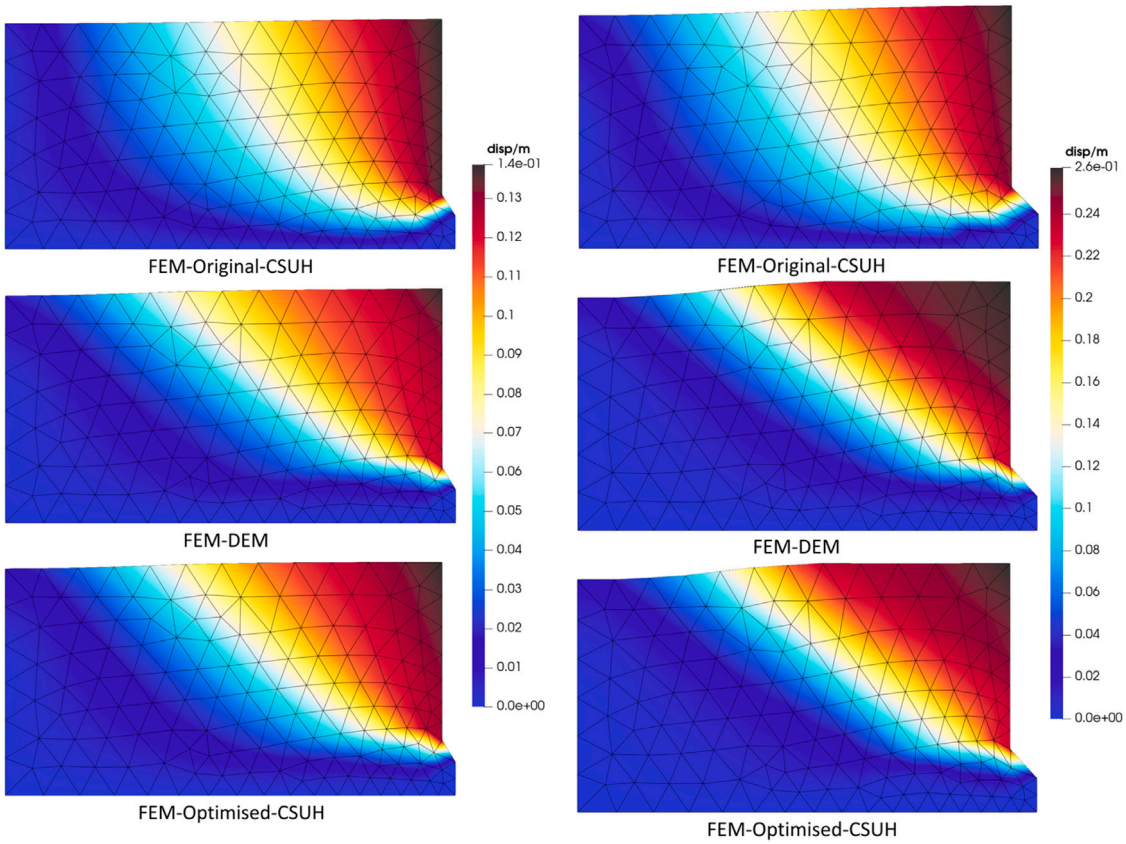


Fig. 22. The solved displacement fields by all three different schemes in the retaining wall problem.

Table 8
The detailed experimental parameters of the three cavity expansion modelling.

Methods	Gauss points	Total steps	Running times/h	Overall MAE	Peak error/kPa	Post-peak error/kPa
exFEM-DEM	900	7100	17.362	\	\	\
exFEM-Original-CSUH	900	7100	0.187	52.12	67.69	51.79
exFEM-Optimised-CSUH	900	7100	0.185	9.25	14.25	2.76

Three solution strategies introduced in Section 4.1 are again applied to the cavity-expansion problem, and the corresponding shear-field distributions at the selected loading stage are shown in Fig. 26. The results show that the exFEM-Original-CSUH strategy exhibits a more localised, patchy high-shear region near the cavity wall, with the shear band developing in a manner that differs from the exFEM-DEM benchmark. By contrast, the modelling results with the FEM-optimised-CSUH approach are in closer agreement with exFEM-DEM, capturing both the overall extent and the dominant orientation of the high-shear zone more consistently. The computational settings for each simulation, including the number of Gauss points, total loading steps, and wall-clock time, are listed in Table 8.

With the calibrated parameter set, the optimised CSUH model yields a Gauss-point stress–strain response that is more consistent with the RVE behaviour. The optimised-CSUH continuum simulation develops a similar local mesh distortion to the FEM-DEM results in the vicinity of the lower inner-cavity corner. This concentrated distortion promotes multiple localisation trigger points, from which several shear bands initiate and propagate upward with continued loading, reproducing the banding pattern observed in FEM–DEM. The improved agreement is consistent with the parameter-induced softening, whereby the increased κ reduces the elastic stiffness and the reduced M lowers the mobilised shear resistance, jointly mitigating excessive post-yield localisation. Consistent with the retaining-wall observations, these changes promote a more distributed deformation development, such that the dominant shear bands shift into the interior of the specimen rather than remaining constrained to the near-vertical boundary.

Similar to the biaxial compression and retaining wall cases, Fig. 27 compares the stress–strain paths sampled in the cavity-expansion simulations with those represented in the strip-footing dataset used for optimising the CSUH-based surrogate. These two sets of loading paths show that the Gauss points in the cavity expansion problem undergo stress paths which are not well covered

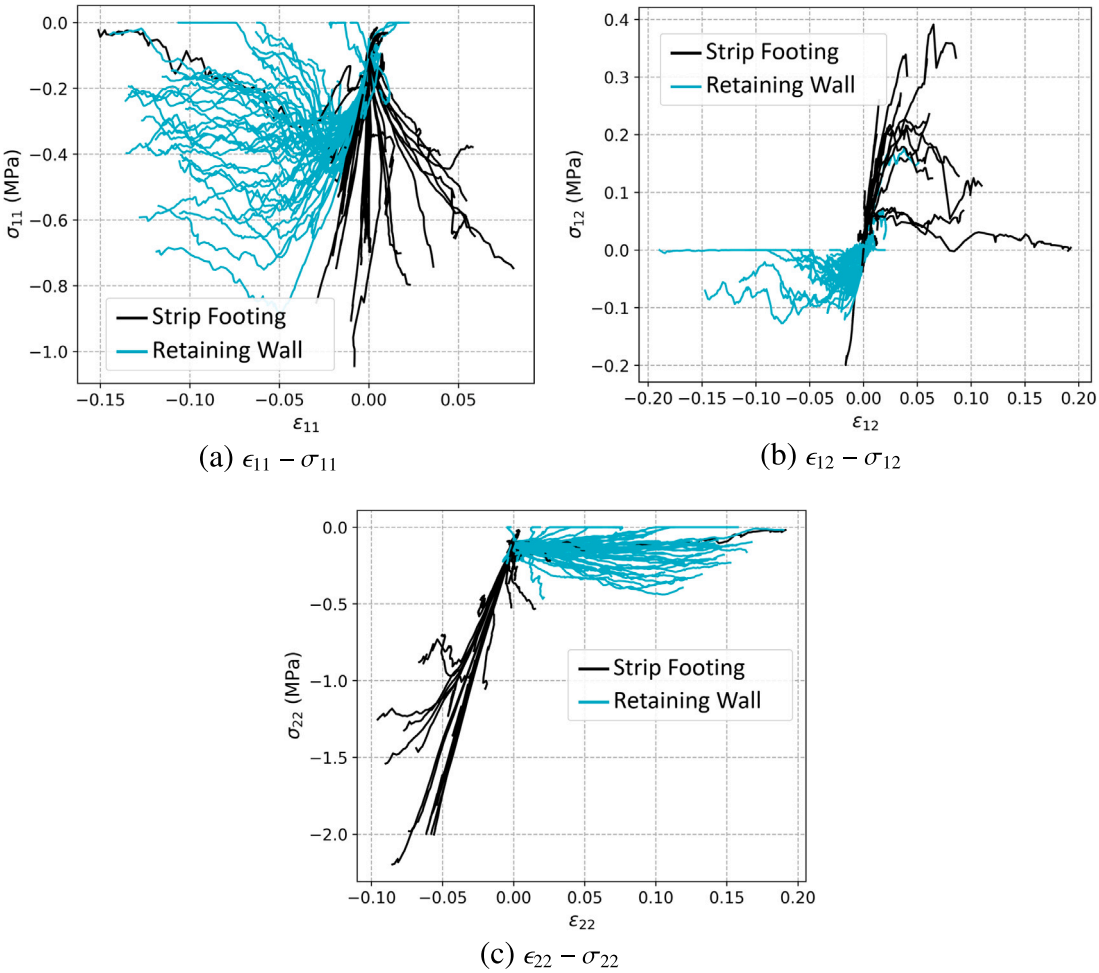


Fig. 23. The difference in loading space between the strip footing and the retaining wall expansion problems.

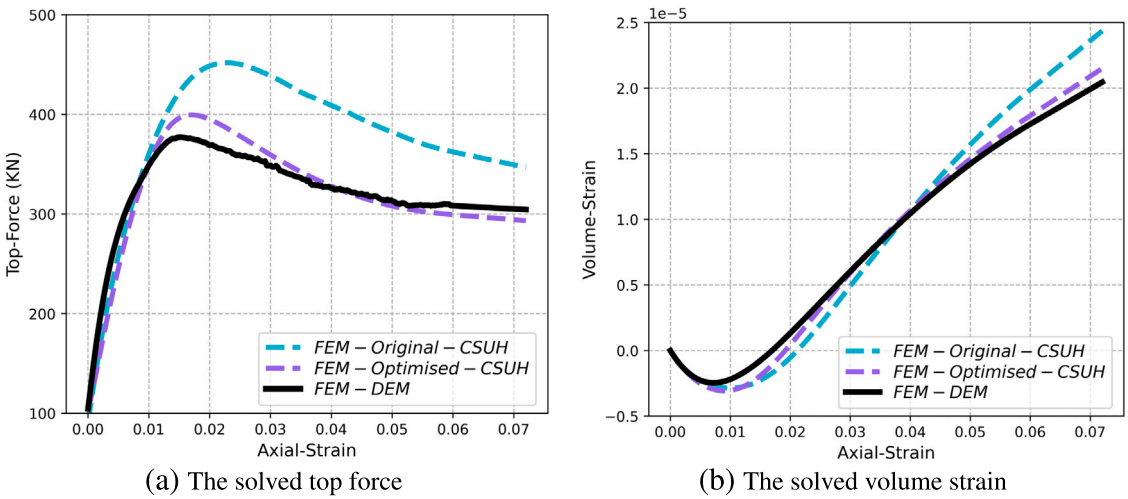


Fig. 24. The rigid wall's top force and volume strain in the retaining wall problem.

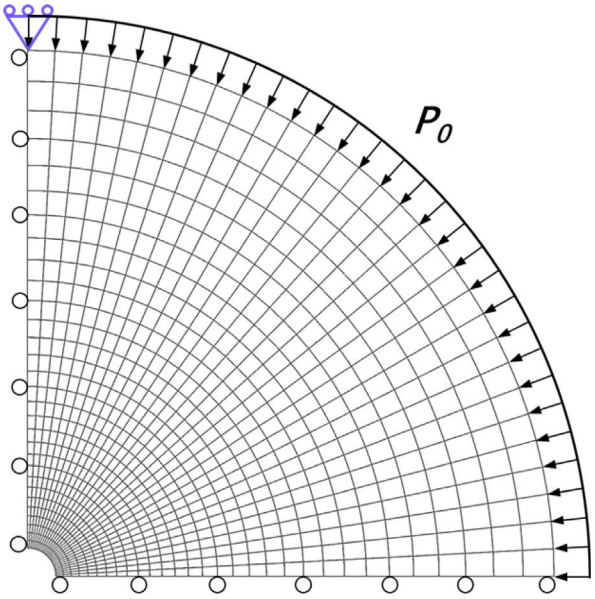


Fig. 25. The boundary conditions for the cavity expansion modelling.

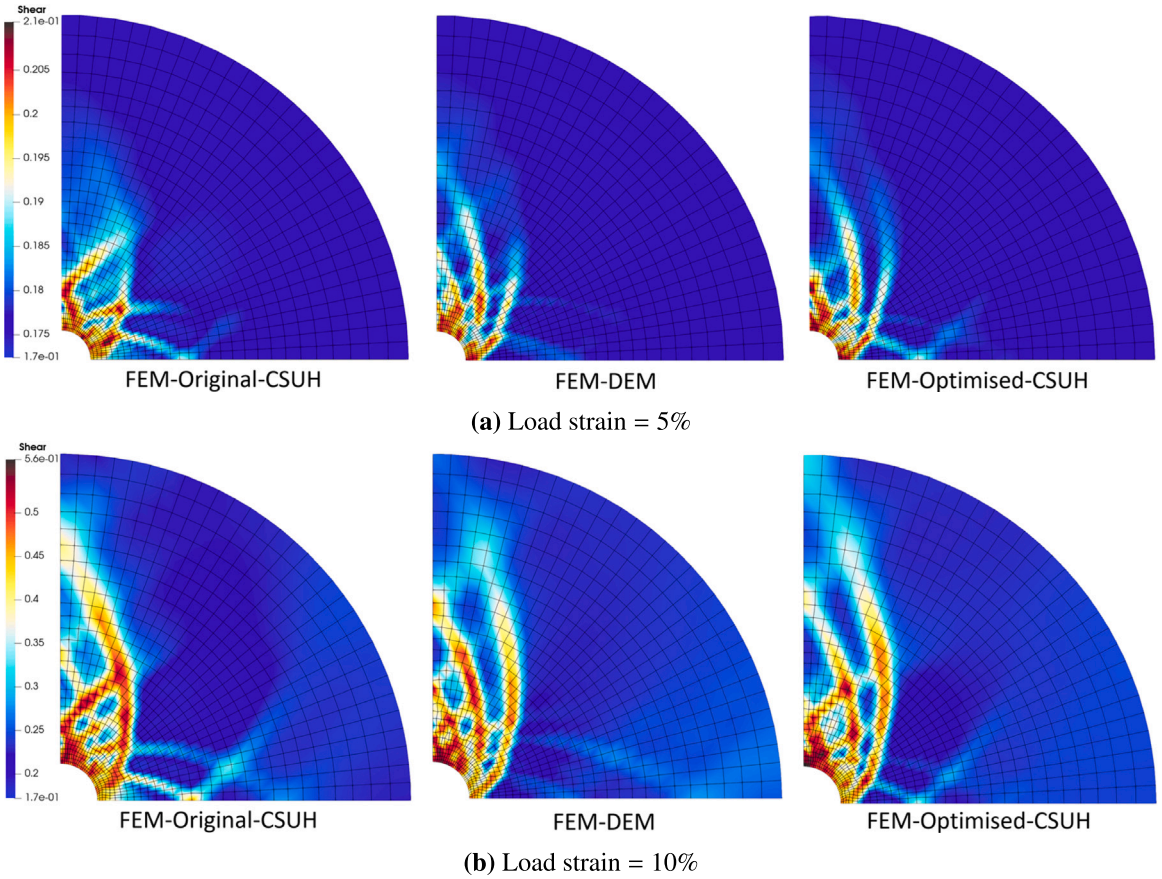


Fig. 26. The solved shear strain fields by three different schemes in the cavity expansion problem.

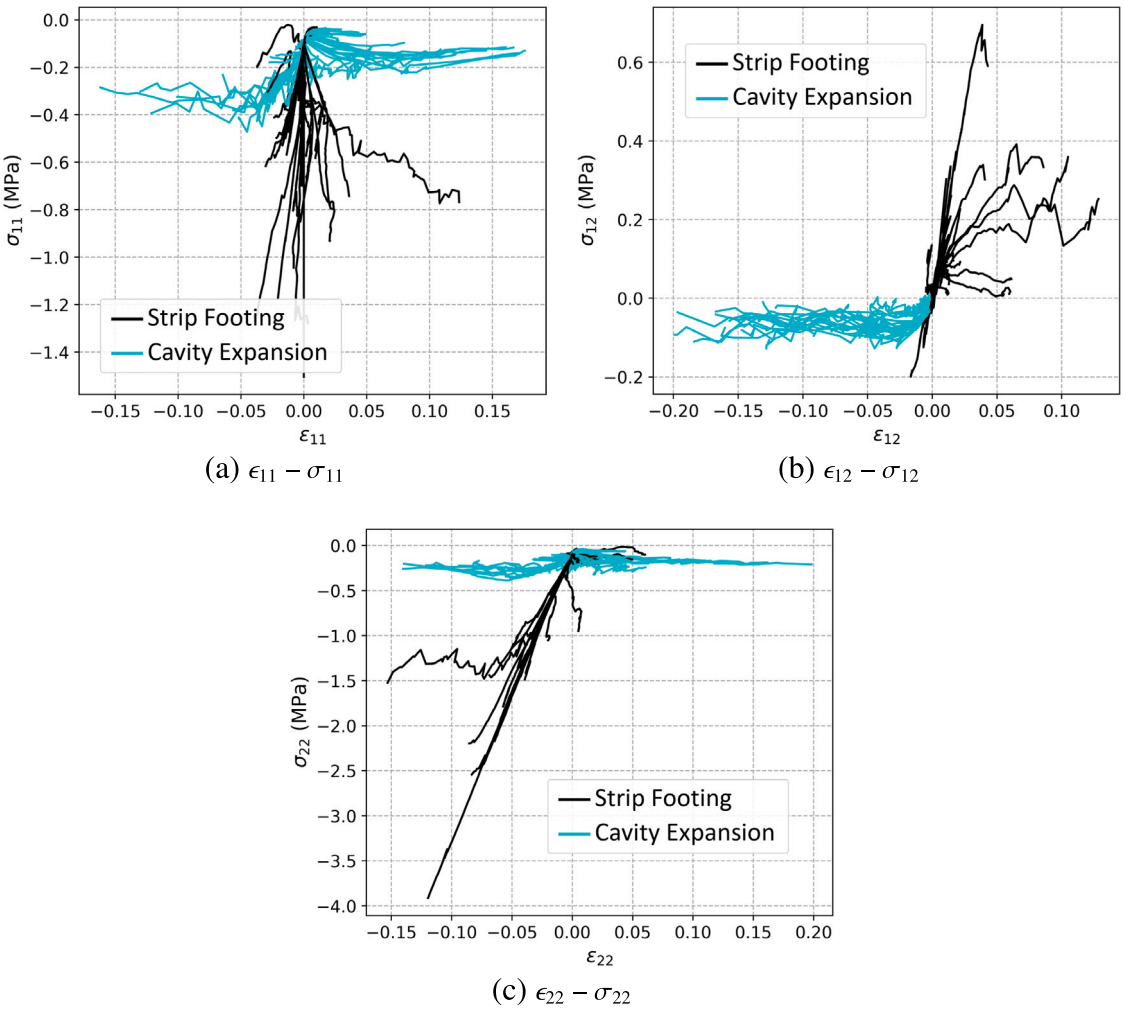


Fig. 27. The difference in loading space between the strip footing and the cavity expansion problems.

by the calibration data. The cavity-expansion test therefore evaluates the optimised-CSUH response largely beyond the calibrated stress–strain range under a more complex loading condition.

Fig. 28 presents the evolution of the cavity pressure with loading strain for the cavity-expansion problem obtained using the three strategies. The overall mean average error (MAE), peak error, and post-peak error between the exFEM-DEM and exFEM-CSUH computation of the cavity pressure are listed in Table 8. Similar to the biaxial compression and retaining-wall cases, the result obtained by FEM-optimised-CSUH follows the FEM-DEM response more closely, exhibiting a reduced peak resistance, whereas the original CSUH model markedly overestimates the peak level and maintains an overly high residual resistance. The improved agreement is consistent with the RVE data-informed optimisation. These observations support the robustness of the optimised CSUH surrogate under a distinct BVP and a different stress-path history.

4.4. Triaxial compression

In the 2D study, the calibrated surrogate is validated against a set of representative BVPs, including the biaxial compression, retaining wall and cavity expansion. In this section, the proposed optimisation framework is further examined in 3D problems. Notably, the CSUH parameters identified from the 2D strip-footing benchmark are not transferred to the 3D setting, since the 2D exFEM–DEM training data, generated with disk-based RVEs, provides only a plane-stress state with no defined out-of-plane component, which is incompatible with the fully three-dimensional stress response required in 3D analyses.

Specifically, a conventional triaxial compression (CTC) exFEM–DEM simulation is conducted to generate the 3D training dataset, based on which the CSUH-based surrogate is recalibrated. The resulting re-optimised surrogate is subsequently assessed on a distinct 3D BVP, i.e. a plane-strain triaxial compression (PTC) test, to evaluate its generalisation capability beyond the training loading paths.

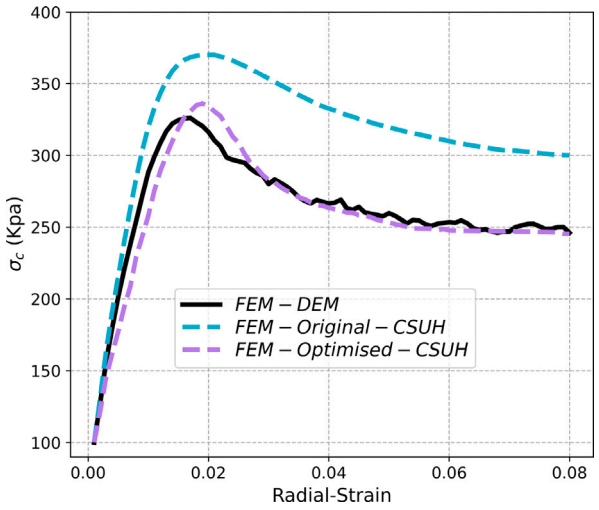
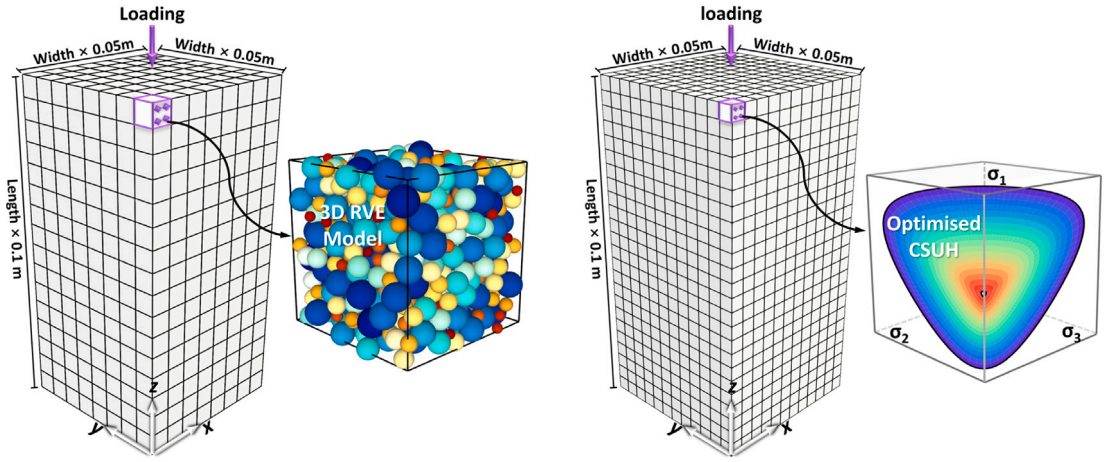


Fig. 28. The cavity pressure computed in the cavity expansion modelling.



(a) The conventional triaxial compression (CTC) loading (b) The plain-strain triaxial compression (PTC) loading

Fig. 29. The geometry domain for the CTC and PTC modelling with different mesh densities.

Table 9
The parameters of each 3D RVE in the exFEM-DEM simulation.

Parameter	Value
Particle number	900
Particle size range (mm)	(5~9)
Young's modulus (MPa)	600
Density (kg/m^3)	2650
Particle frictional coefficient	0.5
Damping ratio	0.1

As illustrated in Fig. 29, both the CTC and PTC simulations consider cubic specimens of height 0.1 m and cross-section 0.05 m × 0.05 m. The training dataset is generated from an exFEM-DEM CTC simulation discretised with a coarse mesh of 8 × 8 × 16 elements (Fig. 29a). A 3D RVE comprising 900 particles is embedded into each Gauss integration point and is initially isotropically consolidated to 100 kPa to provide the local constitutive response. The detailed micro-parameters adopted for the 3D RVE model used in the exFEM–DEM CTC simulations are summarised in Table 9.

In the CTC modelling, the bottom face is fully constrained, the top face is prescribed as rough by suppressing lateral displacements, and a constant confining pressure P_0 = 100 kPa is applied to all lateral faces. The axial compression is imposed via a uniform vertical displacement until the loading strain reaches 0.07. The loading curves from a randomly selected 70% subset

Table 10
The final adopted constitutive parameters in the CSUH-based surrogate in the 3D problem.

Parameters	κ	λ	N	Z	m	ν	e_0	M
Original	0.04	0.135	1.9	0.9	1.8	0.2	0.721	1.25
optimised	0.119	0.167	2.653	0.628	0.596	0.249	0.366	0.895

Table 11
The detailed experimental parameters of the three PTC modelling.

Methods	Gauss points	Total steps	Running times/h	Overall MAE	Peak error/kPa	Post-peak error/kPa
exFEM-DEM	27 648	7100	49.316	\	\	\
exFEM-Original-CSUH	27 648	7100	0.486	9.16	10.03	9.19
exFEM-Optimised-CSUH	27 648	7100	0.485	2.39	3.55	2.26

of RVEs are then assembled as the training set to re-optimise the CSUH model. The initial and re-optimised CSUH parameters are listed in Table 10 with the training hyperparameters of $\delta_r = 0.03$, $\gamma_{max} = 0.16$ and $\gamma_{min} = 0.016$. The re-optimised CSUH surrogate is subsequently embedded into the Gauss points and evaluated in the PTC case using the same specimen geometry but a refined mesh of $12 \times 12 \times 24$ elements (Fig. 29b). Plane strain is enforced by constraining one lateral direction, and the confining pressure is applied to the remaining pair of lateral faces, with all other boundary and loading prescriptions identical to those adopted in the CTC case.

The exFEM-DEM simulation results for the 3D CTC case are shown in Fig. 30(a). The computation is accelerated using multi-processor parallel execution and completed in 41.32 h on a 16-core workstation with 128 GB RAM. The result demonstrates that there is no distinct shear band; instead, shear deformation develops as a diffuse localisation zone. This behaviour is consistent with the mesh dependence feature of strain localisation, where the shear bands tend to be ambiguous when a relatively coarse mesh is used. With the re-optimised CSUH surrogate, the PTC problem is simulated and compared with both the results obtained by the exFEM-DEM and original-CSUH methods. The comparison results are exhibited in Fig. 30(b). The computational settings for each PTC simulation, including the number of Gauss points, total loading steps, and wall-clock time, are listed in Table 11.

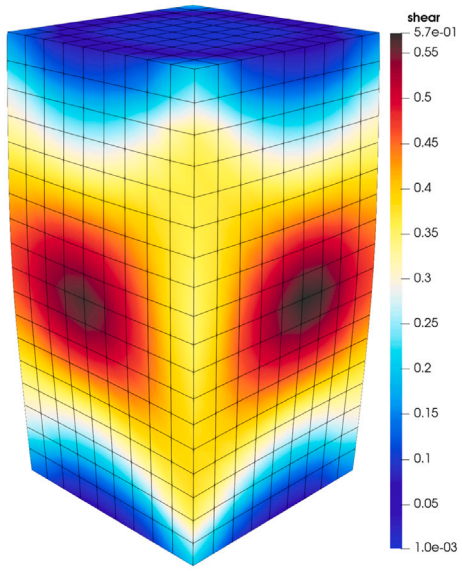
The PTC modelling results in Fig. 30(b) exhibit a comparable shear-localisation pattern to that observed in the biaxial compression simulations across the three modelling schemes. The exFEM analysis with the original CSUH parameters produces a thin, high-intensity shear band, reflecting an over-localised constitutive response. In contrast, the CSUH surrogate, re-optimised from the 3D RVE dataset, alleviates this excessive peak localisation and predicts a more diffuse shear-strain field. The resulting localisation pattern is in closer agreement with the reference exFEM–DEM response, evidencing improved constitutive fidelity under three-dimensional plane-strain triaxial loading.

Similar to the 2D problems, Fig. 31 contrasts the stress–strain loading curves sampled by Gauss points in the 3D CTC and PTC simulations. In particular, the PTC loading explores much larger positive variations in ϵ_{11} and broader ranges in the shear component ϵ_{13}, accompanied by distinct evolutions of σ_{22} and σ_{33} compared with the CTC loading. The limited overlap indicates that the PTC case tests the optimised CSUH model under 3D stress paths that are only partially represented by the CTC-based calibration data.

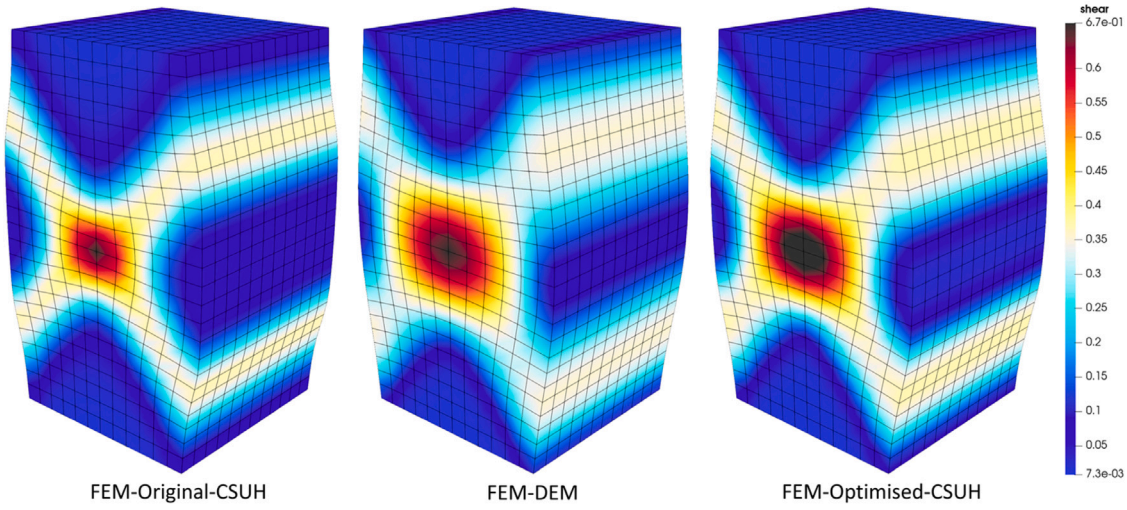
Additionally, Fig. 32 presents the top-force and volumetric-strain evolutions with axial strain for the PTC problem obtained using the three strategies. The overall response trends in Fig. 32(a) and (b) are consistent with those reported for the previous BVPs, showing comparable pre-peak hardening followed by post-peak decay in top force and a contractive-to-dilatant volumetric transition. The overall mean average error (MAE), peak error, and post-peak error between the exFEM-DEM and exFEM-CSUH computation of the top force are listed in Table 11. The parameter optimisation in the 3D setting presents a comparable overall trend to those identified in the 2D studies. The associated improvement of the optimised-CSUH predictions can therefore be interpreted as those discussed for the 2D problems and are therefore not repeated here.

5. Discussion

The ultimate purpose of this work is to develop a constitutive model-based surrogate to reproduce the mechanical behaviour of granular assemblies. However, there is no clear consensus on which elastoplastic template is both expressive and lightweight for the constitutive backbone. Therefore, two conventional elastoplastic backbones are selected for baseline evaluation. The first is an IME model (isotropic elasticity, a von Mises yield surface, and exponential hardening), which is modified and tested as an initial surrogate. The second baseline is the Modified Cam-Clay (MCC) model, assessed to represent the RVE stress–strain response within the proposed multiscale framework. Following the calibration workflow in Section 2.3 and using the RVE datasets described in Section 3.1, the constitutive parameters of both baselines are optimised to quantify their fidelity in reproducing the RVE response. As shown later, both calibrated IME and MCC models remain insufficiently expressive to match the RVE stress–strain behaviour with adequate accuracy, providing a controlled point of reference, and thereby motivating the adoption of CSUH as the surrogate backbone for an RVE-replacement model.



(a) The solved shear strain field by the exFEM-DEM for the CTC case



(b) The solved shear strain field by all three different schemes in the PTC problem.

Fig. 30. The simulation results of CTC and PTC cases.

5.1. The enhanced IME model

The original IME model adopts a standard associated flow rule with isotropic linear elasticity, and governs the yield surface of plasticity relying solely on the von Mises stress. It can be formulated as:

$$\begin{cases} \sigma_{ij} = p\delta_{ij} + s_{ij} = K\epsilon_{kk}\delta_{ij} + 2Ge_{ij} \\ \quad = \left[K\delta_{ij}\delta_{kl} + 2G(\delta_{ik}\delta_{jl} - \frac{1}{3}\delta_{ij}\delta_{kl})\right]\epsilon_{kl} \\ \quad = D^e_{ijkl}\epsilon_{kl} \\ f = \sigma_{von} - H - \sigma_0 \\ H = A(\epsilon_0 + \|\epsilon^p_{ij}\|_F)^n \\ \dfrac{\partial f}{\partial \sigma_{ij}} = \dfrac{\partial g}{\partial \sigma_{ij}} \end{cases} \tag{49}$$

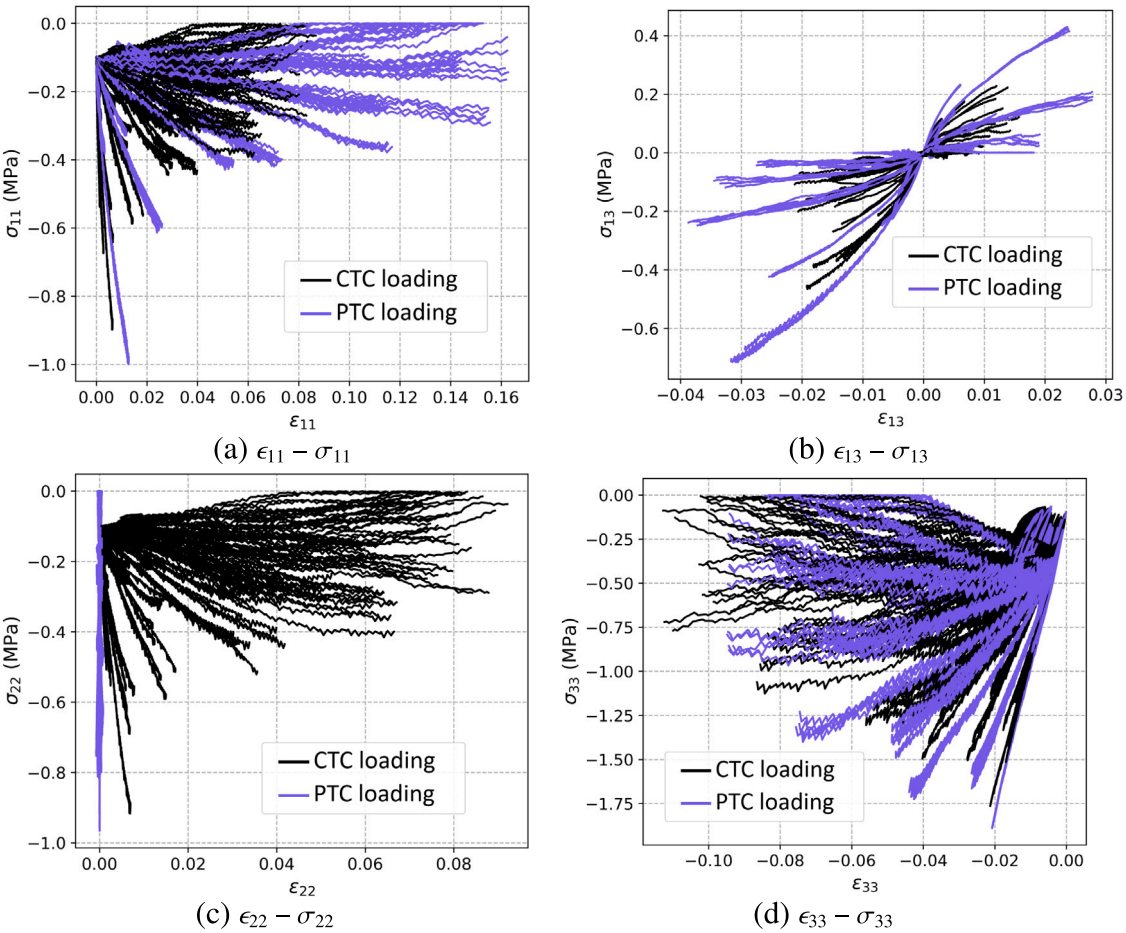


Fig. 31. Difference in loading space between CTC loading and PTC loading.

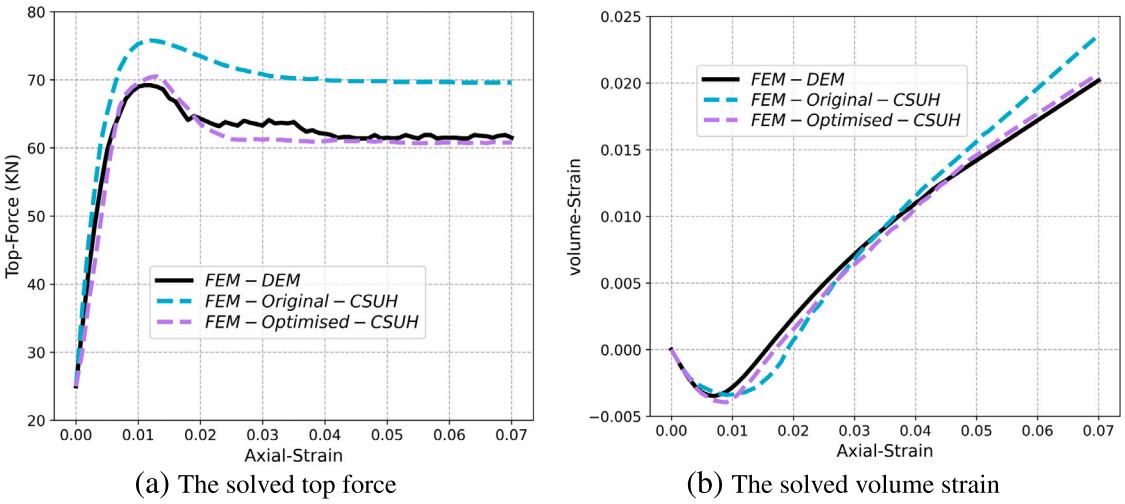


Fig. 32. The top force and volume strain in the PTC problem.

Table 12
Optimisable parameters and their corresponding initial values in the enhanced IME model.

Parameters	E	ν	A	B	ϵ_0	σ_0
Initial value	2e7	0.2	3e5	0.2	0.02	1e4
Parameters	C_d	C_p	C	D	ϵ_{0p}	n_E
Initial value	0.1	−0.1	3e5	0.2	0.02	0.1

where $\sigma_{von} = \sqrt{\frac{3}{2}s_{ij}s_{ij}}$ is the von Mises stress, $\|\epsilon^p_{ij}\|_F$ is the Frobenius norm of plastic strain tensor, which can be expressed as $\|\epsilon^p_{ij}\|_F = \sqrt{\sum_{i,j}|\epsilon_{ij}|^2}$. The plastic shear strain, which is evaluated as $\epsilon^p_s = \sqrt{\frac{2}{3}e^p_{ij}e^p_{ij}}$ where $e^p_{ij} = \epsilon^p_{ij} - \delta_{ij}\epsilon^p_{kk}/3$, is always used for metals. The von-Mises stress σ_{von} can be represented as:

$$\begin{cases} \sigma_{von} = \sqrt{3J_2} \\ J_2 = \frac{1}{2}s_{ij}s_{ij} \\ s_{ij} = \sigma_{ij} - p\delta_{ij} \end{cases} \tag{50}$$

where J_2 is the second invariant of the derivative stress.

In Eq. (37), $\partial f/\partial\sigma_{ij}$ and $\partial H/\epsilon^p_{ij}$ are required for the computation of plastic factor. The differentiation of σ_{von} to stress can be expressed as:

$$\frac{\sigma_{von}}{\sigma_{ij}} = \frac{3(\sigma_{ij} - p\delta_{ij})}{2\sigma_{von}} \tag{51}$$

And the differentiation of the Frobenius norm of the plastic tensor can be shown as:

$$\frac{\partial\|\epsilon^p_{ij}\|_F}{\partial\epsilon^p_{ij}} = \frac{\epsilon^p_{ij}}{\|\epsilon^p_{ij}\|_F} \tag{52}$$

where $\partial f/\partial\sigma_{von} = 1$ and $\partial f/\partial H = -1$ and $dH/d\|\epsilon^p_{ij}\|_F = nA(\epsilon_0 + \|\epsilon^p_{ij}\|_F)$. According to the chain rule, after substituting all of the derivatives into Eq. (37), the plastic factor $d\lambda$ is obtained. Due to these basic functions, the original IME model is incapable of reproducing nonlinear elasticity or yield criteria associated with mean stress, etc.

Several modifications are introduced to enhance the IME model for reproducing the RVE response, including a nonlinear formulation, a mean stress-related yield surface, and the non-association flow rule. Specifically, the elastic response remains isotropic with parameters E and ν, but E is extended to a nonlinear function of the plastic volumetric strain:

$$\begin{cases} E = E_0(1 + \epsilon^p_v)^{n_E} \\ f(p, \sigma_v) = C_p p + \sigma_{von} - H - \sigma_0 \\ H(\epsilon^p_s, \epsilon^p_v) = A(\epsilon_0 + \epsilon^p_s)^B + C(\epsilon_{0p} + \epsilon^p_v)^D \\ \frac{\partial g}{\partial p} = C_d + \frac{\partial f}{\partial p} \end{cases} \tag{53}$$

where E_0 is the initial Young's modulus, and n_E governs the evolution of E. C_p introduces mean-stress dependence in the yield surface, with $C_p = 0$ recovering the original IME yield condition. The constants C, ϵ_{0p}, and D quantify the contribution of plastic volumetric deformation to the hardening law. C_d controls the deviation from associated flow, enabling a non-associated flow rule with volumetric plasticity effects.

With these modifications, the enhanced IME model incorporates nonlinear elasticity, mean-stress-dependent yielding, and a non-associated flow rule. The optimisable parameters and their corresponding initial values are summarised in Table 12

5.2. The Modified Cam-Clay (MCC) model

The Modified Cam–Clay (MCC) model is also considered a conventional critical-state elastoplastic template, serving as a baseline surrogate for the RVE response. In the MCC model, the association-flow rule and an isotropic hardening mechanism with a constant critical-state stress ratio M are adopted in the $p-q$ plane. This provides a numerically robust baseline but offers limited flexibility in representing more general hardening features typically embedded in unified-hardening-type formulations. Briefly, the yield,

Table 13
The initial adopted constitutive parameters in the MCC-based surrogate.

Parameters	κ	λ	N	ν	e_0	M
Original	0.04	0.135	1.9	0.2	0.721	1.25

Table 14
The final adopted constitutive parameter in the enhanced IME-based surrogate.

Parameters	E	ν	A	B	ϵ_0	σ_0
Original	2e7	0.2	3e5	0.2	0.02	1e4
Optimised	4.81e8	0.216	34.5	−2.23	2.79e−2	5.17e−6
Parameters	C_d	C_p	C	D	ϵ_{0p}	n_E
Original	0.1	−0.1	3e5	0.2	0.02	0.1
Optimised	0.155	−0.512	8.35e−5	0.424	8.83e−4	−3.54

hardening, and plastic potential function f, H, and g can be formulated as follows:

$$\begin{cases} f(p,q,\varepsilon_v^p) = \ln\left(\dfrac{p}{p_0}\right) + \ln\left(1+\dfrac{q^2}{M^2p^2}\right) - \dfrac{H}{c_p} = 0 \\ H = \displaystyle\int d\epsilon_v^p \\ g(p,q,\varepsilon_v^p) = f(p,q,\varepsilon_v^p) \end{cases} \tag{54}$$

where $c_p = (\lambda-\kappa)/(1+e_0)$.

Compared to the CSUH model, the standard MCC expressions in Eq. (54) provide a markedly more restrictive set of constitutive mechanisms. It introduces neither a distinct reference surface $\bar{f}$ nor an explicit internal descriptor R to track overconsolidation state, and retains a linear compression relation $e = N - \lambda \ln p$ in the $e - \ln p$. Therefore, standard MCC is primarily validated for normally consolidated clays and is generally less suitable for reproducing the sand-like materials, particularly under overconsolidated states.

Furthermore, the MCC hardening variable is governed exclusively by the plastic volumetric strain ϵ_v^p, leading to a strictly monotonic hardening response and lacking an intrinsic constitutive softening mechanism like the CSUH model, where a stress-ratio-controlled hardening law with a Hvorslev-type failure ratio M_f defines a hardening boundary through $(M_f^4 - \eta^4)/(M_c^4 - \eta^4)$ and enables hardening–softening transition through the integral evolution of H.

Additionally, MCC lacks an explicit state-variable framework to govern the contractancy-dilatancy transition like the CSUH with ξ and M_C. Under the standard associated-flow MCC formulation, the volumetric plastic strain increment is dictated by the fixed potential geometry, which tends to enforce a predominantly dilative response in shear, thereby limiting its capability to reproduce shear-induced contraction in overconsolidated soil materials. The optimisable parameters and their corresponding initial values are summarised in Table 13.

Due to the aforementioned limitations, employing the MCC model as the constitutive backbone to emulate the RVE stress–strain response is expected to introduce systematic deficiencies. These implications are not merely conceptual and will be examined explicitly in the subsequent numerical/experimental assessments.

5.3. Calibration and validation of the enhanced IME and MCC-based surrogate

The enhanced IME and MCC models are calibrated independently using varying fractions of the RVE dataset in Section 3.1 to assess the data efficiency of the proposed framework. The stress-integration scheme described in Section 2.4.2 is incorporated into a gradient-descent optimisation procedure, and the corresponding training histories of two models are exhibited in Fig. 33, respectively. The training results show that both models exhibit a rapid loss reduction at the early stage, followed by a gradual saturation towards a quasi-steady regime. Increasing the training fraction from 60% to 70% leads to a clear reduction in the converged training loss, whereas further expanding the training set up to 90% yields only marginal additional improvement. Finally, the enhanced IME model attains a minimum loss of 0.0146 at step 59,727 while the MCC model reaches a minimum loss of 0.0078 at step 58,976 out of 59,800 epochs with 70% training data.

Compared with the training result of $scheme3$ exhibited in Fig. 13, the optimisation of the enhanced IME and MCC model exhibits a smoother, less oscillatory loss curve, yet converges to a higher minimal loss. This trend is attributable to their simpler constitutive skeletons and superior differentiability, which promote stable gradient updates and regular convergence but limit the attainable fidelity to the RVE response. The final adopted parameter set for the enhanced IME model is listed in Tables 14 and 15.

To explicitly assess the capability of the optimised enhanced IME and MCC to replicate the stress–strain response of the RVE, two additional biaxial compression simulations are performed for each model using the original and optimised parameter sets. Fig. 34 reports the stresses predicted by the FE computations against the corresponding RVE benchmarks for the stress component σ_{yy} at Gauss-points marked in purple colour in Fig. 15. The optimised models exhibit a clear improvement over their original counterparts, indicating an enhanced ability to reproduce the RVE stress–strain response, although systematic biases remain. In particular, Fig. 34(a) shows that the optimised enhanced IME remains overly stiff at small strains, leading to overestimation in the initial stress

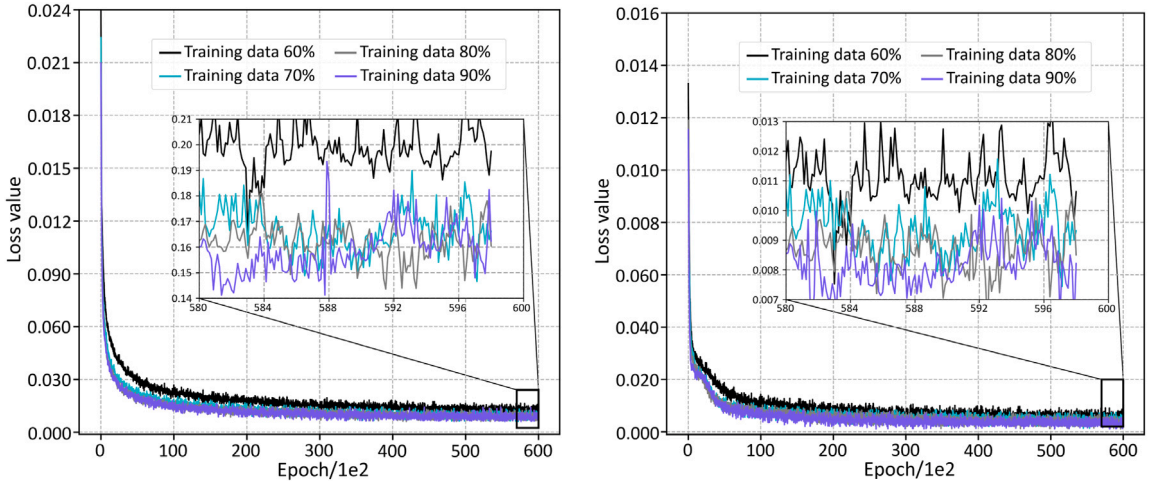


Fig. 33. The obtained training results of the enhanced IME and MCC models using varying fractions of the training dataset.

Table 15
The final adopted constitutive parameters in the MCC-based surrogate.

Parameters	κ	λ	N	ν	e_0	M
Original	0.04	0.135	1.9	0.2	0.721	1.25
Optimised	0.126	0.206	3.153	0.226	0.297	0.906

Table 16
The detailed experimental parameters of the three biaxial compression modelling.

Methods	Gauss points	Total steps	Running times/h	Prediction MAE
exFEM-DEM	484	5000	10.974	\
exFEM-Optimised-ENIME	484	5000	0.122	4.257
exFEM-Optimised-MCC	484	5000	0.129	7.913

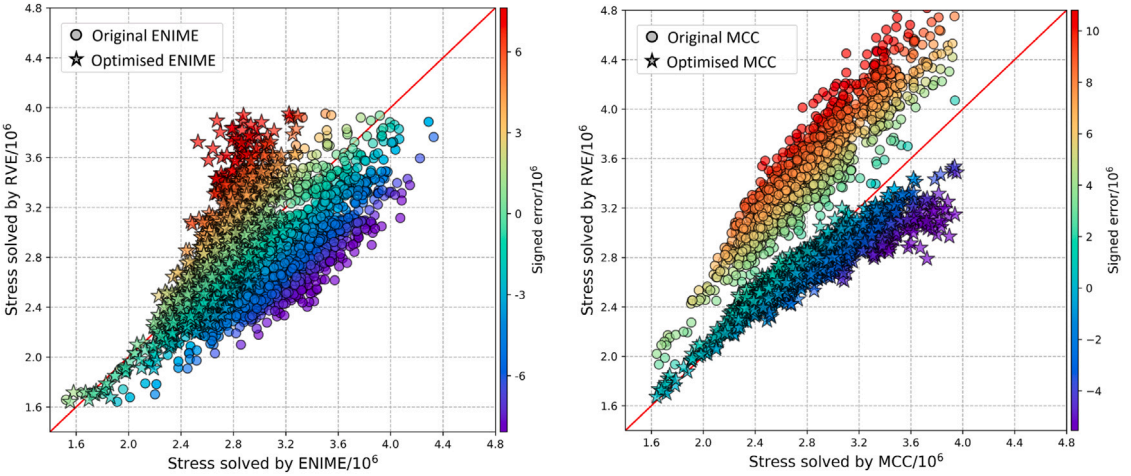
states, while its peak stress level is under-predicted. The optimised MCC captures the pre-peak evolution more faithfully but tends to overestimate stresses near the peak in Fig. 34(b). Table 16 reports the modelling time for the biaxial compression case using the optimised MCC and enhanced IME, benchmarks it against the exFEM–DEM runtime, and summarises the corresponding prediction MAE for both surrogates.

These residual discrepancies are primarily attributed to model-form limitations rather than insufficient optimisation. For the enhanced IME, the underlying Mises-type (J_2) plasticity skeleton lacks state-dependent dilatancy and pressure-dependent strength mobilisation mechanism. Therefore, it cannot reproduce key granular features observed in the RVE responses, including contractancy-dilatancy evolution and the hardening-to-softening transition. The optimised enhanced-IME model, therefore, compensates by adjusting the available parameters to fit the dominant portion of the dataset, which typically yields an overly initial stiffness while underestimating peak strength compared to the RVE model under the biaxial stress path. For MCC, although pressure dependence is embedded, its hardening is solely governed by ϵ_v^p and remains monotonic; the absence of a stress-ratio-controlled degradation law limits its ability to regulate post-peak tendencies, leading to peak overshoot despite good agreement in the pre-peak regime.

Accordingly, the optimisation of constitutive parameters in a principled surrogate should be conceived not merely as numerical convergence but as an integrated framework that incorporates constitutive knowledge, enhances structural expressiveness, and improves numerical robustness, thereby paving a pathway for developing an RVE-replacement surrogate based on the CSUH foundation.

6. Conclusion

This study presents a differentiable elastoplastic modelling framework that automates the calibration of a principled surrogate using a gradient descent-based data-driven optimisation strategy with DEM-based RVE data produced from multiscale FEM-DEM modelling. By implementing differentiability-oriented adjustments, notably a smoothed stress-point integration scheme, we enable stable, end-to-end automatic differentiation throughout the optimisation process. The resulting surrogate reproduces the RVE



(a) The prediction results of the enhanced-IME model (b) The prediction results of the standard MCC model

Fig. 34. The comparison of optimised enhanced IME and MCC predictions against RVE benchmarks for all Gauss points in biaxial compression.

responses with substantial computational acceleration while avoiding the poor extrapolation behaviour of pure neural-network surrogates.

We evaluated three constitutive models as candidate surrogates within the proposed framework. Although the enhanced IME model incorporates several enhancements, it remains insufficient for reproducing the complex response of the granular RVE, primarily because it lacks both state-dependent dilatancy and a pressure-dependent strength-mobilisation mechanism. The MCC model introduces pressure dependence, but its constitutive structure does not include a stress-ratio-controlled degradation law, which in turn limits its ability to capture the post-peak evolution of the granular RVE adequately. In contrast, the Critical State Unified Hardening (CSUH) model demonstrated the necessary representational capacity. However, its original formulation contained algebraic complexity, including nested division, square-root operations, and non-differentiable stress-point integration, that led to gradient instability. Our proposed adjustments successfully mitigated these issues, yielding a stable and convergent calibration. The optimised CSUH surrogate was subsequently validated on various boundary value problems, where it accurately reproduced the results of high-fidelity exFEM-DEM simulations with a substantial computational speedup.

This work proves to be a promising optimisation methodology to construct computationally efficient and physically faithful surrogates for multiscale modelling. The framework successfully balances numerical stability, generalisation capability, and computational efficiency, providing a reliable and automated pathway for calibrating complex constitutive models. These advances enhance the interpretability of data-driven soil mechanics and offer a practical tool for accelerating engineering simulations.

CRediT authorship contribution statement

Mengqi Wang: Writing – review & editing, Writing – original draft, Software, Methodology, Investigation. **Tongming Qu:** Writing – review & editing, Supervision, Funding acquisition. **Shaoheng Guan:** Software, Methodology, Investigation. **Y.T. Feng:** Validation, Supervision. **Jidong Zhao:** Writing – review & editing, Supervision, Funding acquisition.

Declaration of competing interest

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

Acknowledgements

The study was financially supported by the National Natural Science Foundation of China (NSFC)(#52439001, #212100001); Research Grants Council of Hong Kong (under GRF #16208224, GRF #16217225, CRF #C7085-24G, RIF #R6008-24, TRS #T22-607/24-N, and TRS #T22-606/23-R); Doctoral Startup Foundation of Jingchu University of Technology (#YYZ202513).

Data availability

Data will be made available on request.

References

Cao, Y.J., Shen, W.Q., Shao, J.-F., Wang, W., 2020. A novel fft-based phase field model for damage and cracking behavior of heterogeneous materials. Int. J. Plast. 133, 102786.

Cao, Y., Wang, W., Shen, W., Cui, X., Shao, J., 2022. A new hybrid phase-field model for modeling mixed-mode cracking process in anisotropic plastic rock-like materials. Int. J. Plast. 157, 103395.

Chen, X.-X., Zhang, P., Yin, Z.-Y., 2025. A comprehensive investigation of physics-informed learning in forward and inverse analysis of elastic and elastoplastic footing. Comput. Geotech. 181, 107110.

Cheng, W., Yin, Z.-Y., 2024. Fractional order viscoplastic modeling of anisotropically overconsolidated clays with modified isotach viscosity. Int. J. Plast. 172, 103858.

Eghbalian, M., Pouragha, M., Wan, R., 2023. A physics-informed deep neural network for surrogate modeling in classical elasto-plasticity. Comput. Geotech. 159, 105472.

Eghtesad, A., Tan, J., Fuhg, J.N., Bouklas, N., 2024. NN-EVP: A physics informed neural network-based elasto-viscoplastic framework for predictions of grain size-aware flow response. Int. J. Plast. 181, 104072.

Fu, J., Wang, M., Chen, B., Wang, J., Xiao, D., Luo, M., Evans, B., 2023. A data-driven framework for permeability prediction of natural porous rocks via microstructural characterization and pore-scale simulation. Eng. Comput. 39 (6), 3895–3926.

Gavris, G.B., Sun, W., 2025. Discovering neural elastoplasticity from kinematic observations. Proc. Natl. Acad. Sci. 122 (38), e2508732122.

Girard, A., Mohr, D., 2025. Recurrent neural network model predicting elasto-plasticity and matrix fracture in fiber-reinforced composites. Int. J. Solids Struct. 113703.

Granitzer, A.-N., Leo, J., Tschuchnigg, F., 2024. Particle Swarm optimization of interface constitutive model parameters for embedded beam formulations. Int. J. Geomech. 24 (11), 04024255.

Gras, J.-P., Sivasithamparam, N., Karstunen, M., Dijkstra, J., 2017. Strategy for consistent model parameter calibration for soft soils using multi-objective optimisation. Comput. Geotech. 90, 164–175.

Guan, S., Qu, T., Feng, Y., Ma, G., Zhou, W., 2023. A machine learning-based multi-scale computational framework for granular materials. Acta Geotech. 18 (4), 1699–1720.

Guan, Q.Z., Yang, Z.X., Guo, N., Hu, Z., 2023. Finite element geotechnical analysis incorporating deep learning-based soil model. Comput. Geotech. 154, 105120.

Guan, S., Zhang, X., Ranftl, S., Qu, T., 2023. A neural network-based material cell for elastoplasticity and its performance in FE analyses of boundary value problems. Int. J. Plast. 171, 103811.

Guo, N., Chen, L.F., Yang, Z.X., 2022. Multiscale modelling and analysis of footing resting on an anisotropic sand. Géotechnique 72 (4), 364–376.

Guo, N., Zhao, J., 2014. A coupled FEM/DEM approach for hierarchical multiscale modelling of granular media. Internat. J. Numer. Methods Engrg. 99 (11), 789–818.

Guo, N., Zhao, J., 2016. 3D multiscale modeling of strain localization in granular media. Comput. Geotech. 80, 360–372.

Guo, N., Zhao, J., Sun, W., 2016. Multiscale analysis of shear failure of thick-walled hollow cylinder in dry sand. Géotech. Lett. 6 (1), 77–82.

Haghighat, E., Raissi, M., Moure, A., Gomez, H., Juanes, R., 2021. A physics-informed deep learning framework for inversion and surrogate modeling in solid mechanics. Comput. Methods Appl. Mech. Engrg. 379, 113741.

Houdoux, D., Amon, A., Marsan, D., Weiss, J., Crassous, J., 2021. Micro-slips in an experimental granular shear band replicate the spatiotemporal characteristics of natural earthquakes. Commun. Earth Environ. 2 (1), 90.

Hu, Z., Yang, Z.X., Guo, N., Zhang, Y.D., 2022. Multiscale modeling of seepage-induced suffusion and slope failure using a coupled FEM–DEM approach. Comput. Methods Appl. Mech. Engrg. 398, 115177.

Huang, S., Wang, P., Lai, Z., Yin, Z.-Y., Huang, L., Xu, C., 2024. Machine-learning-enabled discrete element method: The extension to three dimensions and computational issues. Comput. Methods Appl. Mech. Engrg. 432, 117445.

Jadoon, A.A., Meyer, K.A., Fuhg, J.N., 2025. Automated model discovery of finite strain elastoplasticity from uniaxial experiments. Comput. Methods Appl. Mech. Engrg. 435, 117653.

Jordan, B., Gorji, M.B., Mohr, D., 2020. Neural network model describing the temperature-and rate-dependent stress-strain response of polypropylene. Int. J. Plast. 135, 102811.

Lai, Z., Chen, Q., Huang, L., 2022. Machine-learning-enabled discrete element method: Contact detection and resolution of irregular-shaped particles. Int. J. Numer. Anal. Methods Geomech. 46 (1), 113–140.

Levasseur, S., Malécot, Y., Boulon, M., Flavigny, E., 2008. Soil parameter identification using a genetic algorithm. Int. J. Numer. Anal. Methods Geomech. 32 (2), 189–213.

Li, W.L., Guo, N., Yang, Z.X., 2024. STZ-Clay: A shear-transformation-zone theory based constitutive model for clay. Int. J. Plast. 176, 103958.

Li, X., Roth, C.C., Mohr, D., 2019. Machine-learning based temperature-and rate-dependent plasticity model: application to analysis of fracture experiments on DP steel. Int. J. Plast. 118, 320–344.

Liu, J., Wautier, A., Bonelli, S., Nicot, F., Darve, F., 2020. Macroscopic softening in granular materials from a mesoscale perspective. Int. J. Solids Struct. 193, 222–238.

Liu, M., Zhuang, P., Lai, F., 2024. A Bayesian optimization-genetic algorithm-based approach for automatic parameter calibration of soil models: Application to clay and sand model. Comput. Geotech. 176, 106717.

Logarzo, H.J., Capuano, G., Rimoli, J.J., 2021. Smart constitutive laws: Inelastic homogenization through machine learning. Comput. Methods Appl. Mech. Engrg. 373, 113482.

Ma, G., Guan, S., Wang, Q., Feng, Y., Zhou, W., 2022. A predictive deep learning framework for path-dependent mechanical behavior of granular materials. Acta Geotech. 17 (8), 3463–3478.

Mei, J., Ma, G., Cao, W., Wu, T., Zhou, W., 2025. Graph neural network unveils the spatiotemporal evolution of structural defects in sheared granular materials. Int. J. Plast. 184, 104218.

Mei, J., Ma, G., Tang, L., Gao, K., Cao, W., Zhou, W., 2023. Spatial clustering of microscopic dynamics governs the slip avalanche of sheared granular materials. Int. J. Plast. 163, 103570.

Mohr, D., Henn, S., 2007. Calibration of stress-triaxiality dependent crack formation criteria: a new hybrid experimental–numerical method. Exp. Mech. 47 (6), 805–820.

Papon, A., Riou, Y., Dano, C., Hicher, P.-Y., 2012. Single-and multi-objective genetic algorithm optimization for identifying soil parameters. Int. J. Numer. Anal. Methods Geomech. 36 (5), 597–618.

Pastor, M., Zienkiewicz, O.C., Chan, A.H.C., 1990. Generalized plasticity and the modelling of soil behaviour. Int. J. Numer. Anal. Methods Geomech. 14 (March 1988), 151–190.

Qu, T., Di, S., Feng, Y., Wang, M., Zhao, T., 2021. Towards data-driven constitutive modelling for granular materials via micromechanics-informed deep learning. Int. J. Plast. 144, 103046.

Qu, T., Guan, S., Feng, Y., Ma, G., Zhou, W., Zhao, J., 2023a. Deep active learning for constitutive modelling of granular materials: From representative volume elements to implicit finite element modelling. Int. J. Plast. 164, 103576.

Qu, T., Zhao, J., Feng, Y.T., 2025. Artificial intelligence for computational granular media. Comput. Geotech. 185, 107310.
Qu, T., Zhao, J., Guan, S., Feng, Y.T., 2023b. Data-driven multiscale modelling of granular materials via knowledge transfer and sharing. Int. J. Plast. 171, 103786.
Tancogne-Dejean, T., Gorji, M.B., Zhu, J., Mohr, D., 2021. Recurrent neural network modeling of the large deformation of lithium-ion battery cells. Int. J. Plast. 146, 103072.
Tao, Y., He, W., Sun, H., Cai, Y., Chen, J., 2022. Multi-objective optimization-based prediction of excavation-induced tunnel displacement. Undergr. Space 7 (5), 735–747.
Wang, M., Feng, Y., Guan, S., Qu, T., 2024a. Multi-layer perceptron-based data-driven multiscale modelling of granular materials with a novel frobenius norm-based internal variable. J. Rock Mech. Geotech. Eng..
Wang, M., Kumar, K., Feng, Y., Qu, T., Wang, M., 2024b. Machine learning aided modeling of granular materials: A review. Arch. Comput. Methods Eng. 1–38.
Wang, M., Qu, T., Guan, S., Zhao, T., Liu, B., Feng, Y., 2022. Data-driven strain–stress modelling of granular materials via temporal convolution neural network. Comput. Geotech. 152, 105049.
Wang, X., Yin, Z.-Y., Wu, W., Zhu, H.-H., 2025. Differentiable finite element method with Galerkin discretization for fast and accurate inverse analysis of multidimensional heterogeneous engineering structures. Comput. Methods Appl. Mech. Engrg. 437, 117755.
Wang, M., Zhang, D.Z., 2021. Deformation accommodating periodic computational domain for a uniform velocity gradient. Comput. Methods Appl. Mech. Engrg. 374, 113607.
Xue, T., Liao, S., Gan, Z., Park, C., Xie, X., Liu, W.K., Cao, J., 2023. JAX-FEM: A differentiable GPU-accelerated 3D finite element solver for automatic inverse design and mechanistic data science. Comput. Phys. Comm. 291, 108802.
Yao, Y.-P., Hou, W., Zhou, A.-N., 2009. UH model: three-dimensional unified hardening model for overconsolidated clays. Geotechnique 59 (5), 451–469.
Yao, Y.-P., Liu, L., Luo, T., Tian, Y., Zhang, J.-M., 2019. Unified hardening (UH) model for clays and sands. Comput. Geotech. 110, 326–343.
Yao, Y., Tian, Y., Cui, W., Luo, T., Li, S., 2024. Unified hardening (UH) model for unsaturated expansive clays. Acta Geotech. 19 (6), 3655–3669.
Yao, Y.-P., Wang, N.-D., 2014. Transformed stress method for generalizing soil constitutive models. J. Eng. Mech. 140 (3), 614–629.
Yin, Z.-Y., Jin, Y.-F., Shen, J.S., Hicher, P.-Y., 2018. Optimization techniques for identifying soil parameters in geotechnical engineering: Comparative study and enhancement. Int. J. Numer. Anal. Methods Geomech. 42 (1), 70–94.
Zhang, W., Ma, G., Mei, J., Wang, R., Zhang, D., Cao, W., Zhou, W., 2025. Dynamic heterogeneity of irregular granular materials captured by in situ X-ray imaging. Int. J. Plast. 104532.
Zhang, N., Xu, K., Yin, Z.Y., Li, K.-Q., Jin, Y.-F., 2025. Finite element-integrated neural network framework for elastic and elastoplastic solids. Comput. Methods Appl. Mech. Engrg. 433, 117474.
Zhang, P., Yang, Y., Yin, Z.-Y., 2021. BiLSTM-based soil–structure interface modeling. Int. J. Geomech. 21 (7), 04021096.
Zhao, Y., Li, X., Jiang, C., Choo, J., 2026. GeoWarp: An automatically differentiable and GPU-accelerated implicit MPM framework for geomechanics based on NVIDIA Warp. Adv. Eng. Softw. 212, 104072.
Zhao, R., Zhou, A., Yao, Y.-P., 2024. The UTUH model: a time-dependent unified hardening constitutive model for unsaturated soils. Acta Geotech. 19 (4), 1649–1666.
Zhou, A., Yao, Y., 2018. Revising the unified hardening model by using a smoothed Hvorslev envelope. J. Rock Mech. Geotech. Eng. 10 (4), 778–790. http://dx.doi.org/10.1016/j.jrmge.2017.10.008.
Zhu, B., Chen, Z., 2022. Calibrating and validating a soil constitutive model through conventional triaxial tests: an in-depth study on CSUH model. Acta Geotech. 17 (8), 3407–3420.
Zhu, Q.-Z., Yu, Q.-J., Shao, J.-f., Zhao, X.-G., 2025. Micromechanical analysis of stress-induced anisotropic frictional damage, spontaneous localization and post-peak snapback problem in quasi-brittle geomaterials. Int. J. Plast. 104549.