Computers and Geotechnics 193 (2026) 107937

Contents lists available at ScienceDirect

Computers and Geotechnics

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

Research Paper

Extension of explicit Runge-Kutta substepping stress integration for viscoplastic model of saturated soils

Wei CHENG [a], Zhen-Yu YIN [a,b,*]

[a] Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, China
[b] State Key Laboratory of Climate Resilience for Coastal Cities, The Hong Kong Polytechnic University, Hong Kong, China

ARTICLE INFO

Keywords:
Fractional consistency viscoplasticity
Runge Kutta
Explicit stress Integration
Piezocone penetration
Gallery excavation

ABSTRACT

Stable integration schemes are critically important for rate-dependent constitutive models, serving as a cornerstone for ensuring accuracy, efficiency, and robustness in finite element implementations. This paper investigates the numerical performance of explicit stress integration schemes with adaptive substepping for integrating a newly proposed fractional consistency two-surface viscoplastic model for saturated clays. The incremental stress–strain-strain rate relation of the model can be linearized following the consistency condition of the rate-dependent loading surface and subsequently integrated using four distinct explicit Runge-Kutta substepping integration algorithms (i.e., RK12, RK23, RK34, RK45) with automatic error control and stress drift correction techniques. The overall numerical performance of the algorithms in terms of accuracy and efficiency is evaluated at both the material point level (i.e., isotropic, oedometric, and triaxial compression tests) and the boundary-value problem level (i.e., piezocone penetration and underground gallery excavation), which demonstrates that the RK23 and RK34 algorithms perform excellently in balancing accuracy and computational cost. The proposed algorithms provide a versatile and adaptive framework for integrating time-dependent constitutive equations, particularly those based on the consistency viscoplastic approaches commonly used in advanced rate-dependent modeling, allowing for a wide range of geotechnical engineering applications.

1. Introduction

In geotechnical numerical analysis, the mechanical response of soils is fundamentally characterized by the stress–strain relationships defined through constitutive models. To achieve accurate and reliable outcomes across diverse engineering scenarios, it is necessary to employ, based on the specific characteristics of the problem being analyzed, advanced soil constitutive models capable of capturing multifaceted soil behaviors, including nonlinearity, dilatancy, rate dependency, temperature sensitivity, degradation of inherent structure, anisotropy, unsaturated effects, and cyclic accumulation *etc*. Integrating these stress–strain relationships within a general-purpose nonlinear finite element framework presents a formidable challenge, as it requires simultaneous consideration of such complexities while ensuring precision, computational efficiency, and robustness in advanced constitutive modeling.

Over recent decades, significant advancements have proliferated in phenomenological soil constitutive modelling, spanning frameworks of elastoplasticity, viscoplasticity, hypoplasticity, and hyperplasticity, with seminal contributions including the modified Cam-clay model (Roscoe and Burland, 1968), bounding/subloading surface models (Dafalias, 1986; Hashiguchi, 1989), the unsaturated soil models (Alonso et al., 1990), and the thermomechanical models (Hueckel and Borsetto, 1990). In these widely adopted phenomenological constitutive models, the deformation behavior of soils is governed by the coexistence of elastic and plastic mechanisms. This introduces loading/unloading inequality constraints, which are mathematically formulated as the Karush-Kuhn-Tucker (KKT) complementarity conditions. To enable efficient numerical integration of the stress–strain relationships under such inequality constraints, numerous general strategies have been developed, broadly categorized as explicit and implicit integration schemes. Implicit integration methods typically employ an elastic–-plastic operator splitting technique with the elastic prediction by freezing the plastic flow, and correct the plastic parts by the backward Euler or midpoint/ trapezoidal rule method (Simo and Hughes, 1998) using returning mapping algorithms to ensure the model satisfies *KKT* conditions, including closest point projection method (Borja and Lee, 1990; Borja et al., 2001; Rouainia and Muir Wood, 2001; Foster et al., 2005; Hu and Liu, 2014; Zhang and Buscarnera, 2016; Zhao et al., 2018;

* Corresponding author at: Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, China.
E-mail addresses: w1cheng@polyu.edu.hk (W. CHENG), zhenyu.yin@polyu.edu.hk (Z.-Y. YIN).

https://doi.org/10.1016/j.compgeo.2026.107937
Received 17 November 2025; Received in revised form 2 January 2026; Accepted 18 January 2026
Available online 27 January 2026

Marinelli and Buscarnera, 2019; Petalas and Dafalias, 2019; Carow and Rackwitz, 2021; Zhou et al., 2022a; Su et al., 2023; Tian et al., 2024; Zhou et al., 2025), cutting plane method (Starman et al., 2014; Zhao et al., 2018; Yin et al., 2019; Li and Yin, 2020, 2021; Zhao et al., 2023), radial return method (Borja and Lee, 1990; Versino and Bennett, 2018), *etc*. Among these, the closest point projection method is one of the most widely used; it yields unconditionally stable stress updates by solving, via Newton–Raphson iterations, a set of nonlinear equations built from the yield (or plastic potential) function and its first-order derivatives with respect to stresses and internal variables, while second-order derivatives of the yield function enter only in the derivation of the consistent tangent operator for global Newton iterations. By contrast, the cutting plane algorithm obtains a simpler and more flexible formulation by evaluating derivatives at trial points. All these categories, however, rely on first-order derivatives of the residual functions and neglect higher-order terms, inherently limiting efficiency compared with the latest trust region-based returning-mapping algorithms (Lester and Scherzinger, 2017; Zhou et al., 2022b) and second-order cone programming-based approaches (Krabbenhoft and Lyamin, 2012). Additionally, the second-order derivatives of yield surfaces and plastic potential functions in complex constitutive models may be difficult or impractical to derive analytically, making the computation of a consistent tangent operator challenging.

Traditional explicit integration schemes, e.g., the forward Euler approach, the modified Euler approach, and Runge-Kutta methods, are intuitive and easy to implement. However, when faced with large strain increments, they often incur significant errors or even numerical failures, making it difficult to guarantee stability and convergence. To address this issue, Sloan (1987) and Sloan et al. (2001) proposed the use of an automatic substepping explicit integration method incorporating the relative error control through two lower-order and high-order single steps and a drift correction of stress point back to the yield surface at the end of each substep, thereby markedly improving the accuracy and convergence of the explicit algorithms in integrating MCC-type models. Pérez-Foguet et al. (2001) provided the consistent tangent matrix corresponding to explicit substepping schemes, thereby further improving the computational efficiency of constitutive models in finite element analyses. These approaches (Sloan et al., 2001) were subsequently extended to more complex constitutive models (Zhao et al., 2005; Pedroso et al., 2008; Farias et al., 2009; Andrianopoulos et al., 2010; Sołowski and Gallipoli, 2010a, 2010b; Hong et al., 2012; Hu and Liu, 2014; Zhou and Zhang, 2015; Zhang and Zhou, 2016; Lloret-Cabot et al., 2016; Sun et al., 2021; Carow and Rackwitz, 2021; Lloret-Cabot et al. 2021; Dong, 2023), becoming a key means to enhance the stability and reliability of explicit methods. Among those approaches, Pedroso et al. (2008) established a general root-finding framework based on the Kronecker-Picard formula to accurately locate all intersection points, while Zhou and Zhang (2015) employed a double cosine function method for their thermo-mechanical model for reliable stress integration with non-convex yield surfaces; Sołowski and Gallipoli (2010) and Hong et al. (2012) addressed cross-hardening issues between intersecting yield surfaces; Andrianopoulos et al. (2010) and Hu and Liu (2014) were the first to extend adaptive explicit integration algorithms to the integration of constitutive models of cyclic plasticity; Lloret-Cabot et al. (2016) and Lloret-Cabot and Sheng (2022) systematically explored the efficiency, error, and performance of different order explicit adaptive Runge-Kutta schemes in solving MCC model at the element level, and Dong (2023) further extended these discussions to the more complex MIT-E3 model at both the element and boundary-value problem levels. This approach remains viable within the rate-independent elastoplastic framework, even for moderately complex constitutive models. Certainly, there are other promising stress integration/updating algorithms (Scalet and Auricchio, 2018), including, but not limited to, mathematical programming, incremental energy minimization, and deep learning methods. It should be noted that the deep learning methods, gaining immense popularity, primarily utilize artificial neural networks to approximate the existing stress–strain equations or training data by minimizing the loss function towards zero, thereby achieving the update of stresses and internal variables. This approach aligns with solving the unconstrained optimization problems, but shows much less efficiency in direct solving forward problems. Its main advantages lie in providing an excellent surrogate model or computational framework for both forward and inverse problems (Zhang, 2026).

Adaptive explicit integration schemes have been demonstrated to be effective for rate-independent models and offer clear advantages as an efficient integration framework. However, their applications to rate-dependent models (Perzyna, 1963; Wang et al., 1997; Shi et al., 2019; Cheng and Yin, 2024, 2025; Cheng et al., 2024) require further discussion and refinement, and they have not been fully implemented successfully in finite element codes. This paper fully linearizes the novel simplified fractional consistency isotach two-surface viscoplastic model (FVP) into a system of nonlinear equations via its rate-dependent loading surface, subsequently solved through various adaptive explicit integration algorithms with dynamic adjustment of the substep size to ensure robust convergence and computational efficiency with minimal modifications. After this introduction, the following Section 2 introduces the general framework and existing limitations of consistency rate-dependent models. Section 3 presents the incremental stress–strain-strain rate relationship of the FVP model. Section 4 details the iterative procedures for various explicit adaptive Runge-Kutta substepping integration algorithms and implements their code within a commercial finite element software ABAQUS. All integration strategies are employed and tested by solving material point analyses under diverse rate-dependent loading paths in Section 5. Section 6 further simulates two initial boundary value problems (IBVPs) of piezocone penetration and gallery excavation to assess the performance of the algorithms, followed by conclusions in Section 7. It should be noted that bold-faced letters denote tensors and vectors, stress and strain in compression are positive, and stress quantities are to be defined as effective stress quantities.

2. Constitutive modeling of rate-sensitive geomaterials

For classical consistency elasto-viscoplastic constitutive models (Qiao et al., 2016; Shi et al., 2019; Cheng and Yin, 2024; Zambrano-Cruzatty et al., 2024), the progressive evolution of stresses and internal variables is conveniently described by an algebraic constraint, defined as:

$$E_\sigma = \left\{ (\boldsymbol{\sigma}', \boldsymbol{H}, \dot{\boldsymbol{\varepsilon}}^{vp}) : f_{LS}(\boldsymbol{\sigma}', \boldsymbol{H}, \dot{\boldsymbol{\varepsilon}}^{vp}) \le 0 \right\} \tag{1}$$

where $\boldsymbol{\sigma}'$ denotes the general effective stress tensor, $\boldsymbol{H}$ represents a set of internal variables, $\dot{\boldsymbol{\varepsilon}}^{vp}$ signifies the viscoplastic strain rate; $f_{LS}(\boldsymbol{\sigma}', \boldsymbol{H}, \dot{\boldsymbol{\varepsilon}}^{vp})$ defines the rate-dependent actual/loading yield surface; the admissible region E_σ encompasses the elastic domain, meaning that any admissible state $(\boldsymbol{\sigma}', \boldsymbol{H}, \dot{\boldsymbol{\varepsilon}}^{vp})$ must lie strictly within the interior of E_σ; the yield function $f_{LS}(\boldsymbol{\sigma}', \boldsymbol{H}, \dot{\boldsymbol{\varepsilon}}^{vp})$ explicitly depends on $(\boldsymbol{\sigma}', \boldsymbol{H}, \dot{\boldsymbol{\varepsilon}}^{vp})$ and evolves continuously along the boundary of E_σ, which is defined as $\partial E_\sigma = \left\{ (\boldsymbol{\sigma}', \boldsymbol{H}, \dot{\boldsymbol{\varepsilon}}^{vp}) : f_{LS}(\boldsymbol{\sigma}', \boldsymbol{H}, \dot{\boldsymbol{\varepsilon}}^{vp}) = 0 \right\}$.

Given the standard small strain assumption, the total strain increment ($d\boldsymbol{\varepsilon}$) is additively decomposed into elastic ($d\boldsymbol{\varepsilon}^e$) and viscoplastic ($d\boldsymbol{\varepsilon}^{vp}$) components. This allows the effective stress increment ($d\boldsymbol{\sigma}'$) to be described by an incremental Hooke's law, namely,

$$d\boldsymbol{\sigma}' = \boldsymbol{D}^e : d\boldsymbol{\varepsilon}^e = \boldsymbol{D}^e : (d\boldsymbol{\varepsilon} - d\boldsymbol{\varepsilon}^{vp}) \tag{2}$$

where $\boldsymbol{D}^e$ denotes the elastic stiffness tensor, and it is typically considered to be rate-independent.

Furthermore, within the framework of elasto-viscoplasticity, the increment of internal variables ($d\boldsymbol{H}$) and viscoplastic strain rate ($d\dot{\boldsymbol{\varepsilon}}^{vp}$) are defined to depend on viscoplastic strain ($\boldsymbol{\varepsilon}^{vp}$), specifically,

$$d\boldsymbol{H} = \frac{\partial \boldsymbol{H}}{\partial \boldsymbol{\varepsilon}^{vp}} d\boldsymbol{\varepsilon}^{vp} \text{ and } d\dot{\boldsymbol{\varepsilon}}^{vp} = \frac{d\boldsymbol{\varepsilon}^{vp}}{dt} - \dot{\boldsymbol{\varepsilon}}_n^{vp} \tag{3}$$

where dt is the given time increment, and $\dot{\boldsymbol{\varepsilon}}_n^{vp}$ is the viscoplastic strain rate at the previous step.

The viscoplastic strain increment ($d\boldsymbol{\varepsilon}^{vp}$) is determined by a specific form of viscoplastic potential function (g). If the form of g is the same as $f_{LS}(\boldsymbol{\sigma}', \boldsymbol{H}, \dot{\boldsymbol{\varepsilon}}^{vp})$, the viscoplastic flow rule is associated; otherwise, it is non-associated.

The fractional viscoplastic flow rule is a special type of non-orthogonal flow rule, based on the fractional order derivative of stress corresponding to $f_{LS}(\boldsymbol{\sigma}', \boldsymbol{H}, \dot{\boldsymbol{\varepsilon}}^{vp})$, which is the form adopted in our paper as follows:

$$d\boldsymbol{\varepsilon}^{vp} = \Lambda \cdot \boldsymbol{m} = \Lambda \cdot \left(\frac{\partial^{\mu} f_{LS}}{\partial p'^{\mu}} \frac{\partial p'}{\partial \boldsymbol{\sigma}'} + \frac{\partial^{\mu} f_{LS}}{\partial q^{\mu}} \frac{\partial q}{\partial \boldsymbol{\sigma}'} \right) \tag{4}$$

where $q = \sqrt{3/2 \cdot \boldsymbol{s} : \boldsymbol{s}}$ and $\boldsymbol{s} = \boldsymbol{\sigma}' - p'\mathbf{1}$ is the deviatoric stress invariant and tensor, respectively; μ is the fractional order, Λ is the viscoplastic multiplier, and $\boldsymbol{m}$ is actually the non-orthogonal gradient of $f_{LS}(\boldsymbol{\sigma}', \boldsymbol{H}, \dot{\boldsymbol{\varepsilon}}^{vp})$. When the fractional order $\mu = 1$, this non-orthogonal gradient reduces to the conventional first-order gradient, which coincides with the classical associated viscoplastic flow formulation. When $\mu \neq 1$, however, the flow direction is no longer strictly normal to the loading surface. In other words, the "fractional viscoplastic flow rule" provides, through the additional parameter μ, a tunable degree of freedom to capture the dilatancy behaviors that are commonly observed in overconsolidated clays. It should be noted that the fractional viscoplastic flow rule here is a simplified formulation, which does not account for the effect of Lode's angle and is primarily employed for the purpose of algorithm verification.

The consistency elasto-viscoplastic model is completed subject to the standard *KKT* complementarity inequality condition regarding the viscoplastic multiplier and the yield function:

$$\Lambda \geq 0,\ f_{LS}(\boldsymbol{\sigma}', \boldsymbol{H}, \dot{\boldsymbol{\varepsilon}}^{vp}) \leq 0,\ \text{and}\ \Lambda \cdot f_{LS}(\boldsymbol{\sigma}', \boldsymbol{H}, \dot{\boldsymbol{\varepsilon}}^{vp}) = 0 \tag{5}$$

The presence of the complementarity inequality condition makes it indispensable to enhance the operator splitting technique to freeze the viscoplastic strain components before performing the elastic prediction, particularly in the explicit substepping integration schemes for rate-dependent elasto-viscoplastic problems.

3. Formulation of a fractional consistency two-surface viscoplastic model

Based on the modified isotach viscosity, a fractional consistency two-surface viscoplastic model (FVP) that strictly adheres to the *KKT* conditions has been proposed to characterize the mechanical behaviors of overconsolidated clay (Cheng et al., 2024). In the present study, this FVP model serves as the vehicle for testing the proposed integration algorithms. Its fundamental theoretical framework and systematic experimental validation have been detailed in the aforementioned independent study. The focus here is to outline its core formulations pertinent to numerical implementation, and subsequently, to conduct a performance analysis of the proposed integration schemes. Here, we have simplified the rate-dependent model by neglecting temperature effects. The schematic plots of the rate-dependent model are presented in Fig. 1, and the corresponding constitutive equations are encapsulated in Eqs. 6–20. In this model, the viscoplastic strain rate tensor ($\dot{\boldsymbol{\varepsilon}}^{vp}$) in general form is replaced by a generalized viscoplastic strain rate $\dot{\varepsilon}_m^{vp}$.

Within the FVP model formulation, a linear isotropic elastic deformation is postulated, with the elastic stiffness tensor $\boldsymbol{D}^e$ defined by

$$\boldsymbol{D}^e = (K - 2/3G)\mathbf{1} \otimes \mathbf{1} + 2G\boldsymbol{I} \tag{6}$$

where K and G are the elastic bulk modulus and the elastic shear modulus of soils, respectively, and $\boldsymbol{I}$ is the fourth-order identity tensor defined as $I_{ijkl} = \left(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}\right)/2$, where δ_{ij} is the Kronecker delta. With a pressure-dependent isotropic hypoelastic formulation, the expression of K and G is presented as follows:

$$K = \frac{(1+e_0)p'}{\kappa} \text{ and } G = \frac{3(1-2\nu)(1+e_0)p'}{2(1+\nu)\kappa} \tag{7}$$

with e_0 being the initial void ratio, $p' = \mathrm{tr}(\boldsymbol{\sigma}')/3$ being the mean effective stress, κ being the swelling slope in (υ, $\ln p'$) plane, and ν being the constant drained Poisson's ratio.

In the viscoplastic mechanism, a loading surface (*LS*), connected to a yield surface (*YS*) by the radial mapping rule ($r = p'/\overline{p'} = q/\overline{q} = p'_c/\overline{p'_c}$), is adopted to describe the progressive hardening behaviors at overconsolidated states:

$$f_{LS}(\boldsymbol{\sigma}', r, \overline{p'_c}) = \frac{q^2}{M_f^2} + p'(p' - r\overline{p'_c}) \tag{8}$$

with M_f being the stress ratio corresponding to the apex of the loading surface; the yield ratio r and the pre-consolidation pressure $\overline{p'_c}$ being the two internal variables regarding to the viscoplastic strain increment. When the soil turns into a normally consolidated state ($r = 1$), *LS* will coincide with *YS* using the same formulation of $f_{YS}(\boldsymbol{\sigma}', \overline{p'_c}) = f_{LS}(\boldsymbol{\sigma}', r, \overline{p'_c}) = q^2/M_f^2 + p'(p' - \overline{p'_c})$.

The modified exponential formed isotach viscosity is defined on the evolution of pre-consolidation pressure, where the current pre-consolidation pressure increases with the increase in the ratio of the current generalized viscoplastic strain rate $\dot{\varepsilon}_m^{vp}$ to the reference generalized viscoplastic strain rate $\dot{\varepsilon}_{mref}^{vp}$:

$$\overline{p'_c} = \overline{p'_{cref}} \left(\frac{\dot{\varepsilon}_m^{vp}}{\dot{\varepsilon}_{mref}^{vp}} \right)^{\alpha} \tag{9}$$

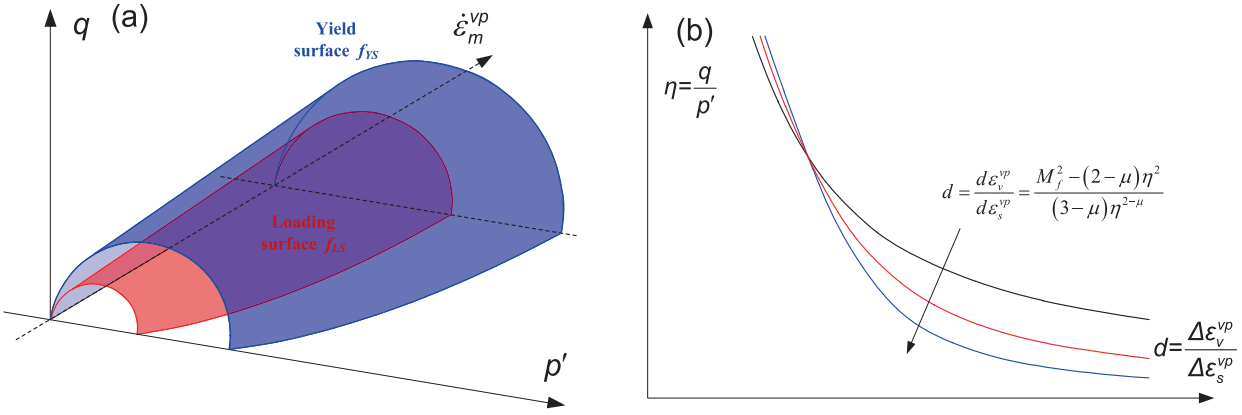

Fig. 1. Schematic plots of the fractional consistency two-surface viscoplastic model: (a) (p', q, $\dot{\varepsilon}_m^{vp}$) space; (b) dilatancy equation in (η, d) plane.

with

$$\dot{\varepsilon}_m^{vp} = \sqrt{\left(\dot{\varepsilon}_v^{vp}\right)^2 + \left(M_g \dot{\varepsilon}_s^{vp}\right)^2},\ \dot{\varepsilon}_v^{vp} = tr(\dot{\boldsymbol{\varepsilon}}^{vp}),\ \text{and}\ \dot{\varepsilon}_s^{vp} = \sqrt{2/3}\left\|\dot{\boldsymbol{\varepsilon}}^{vp} - 1/3\dot{\varepsilon}_v^{vp}\mathbf{1}\right\| \tag{10}$$

being the generalized, volumetric, deviatoric viscoplastic strain rate, $\alpha = \beta\{\kappa/(\lambda-\kappa) + \mathbf{1}/[1 - s_L(1-1/r)]\}$ being the creep index associated with the overconsolidation ratio, β being a material parameter controlling the creep rate, λ being the slope of the normal compression line in (υ, $\ln p'$) plane, s_L being the parameter controlling the hardening rate of *LS* approaching *YS*.

M_g is the critical state ratio that defined through the internal frictional angle φ. The values of M_g and M_f satisfy a certain relation of $M_f = \sqrt{2-\mu}\cdot M_g$ with μ being the fractional order, and the viscoplastic flow direction is defined through the dilatancy ratio (d) and the stress ratio (η) expressed in terms of fractional-order derivatives:

$$d = \frac{d\varepsilon_v^{vp}}{d\varepsilon_s^{vp}} = \frac{\partial^\mu f_{LS}/\partial p'^\mu}{\partial^\mu f_{LS}/\partial q^\mu} = \frac{M_f^2 - (2-\mu)\eta^2}{(3-\mu)\eta^{2-\mu}} \tag{11}$$

The increments of two internal variables $\boldsymbol{H} = \left\{\overline{p'_{cref}}, r\right\}$, namely the reference pre-consolidation pressure ($\overline{p}'_{cref}$) and the yield stress ratio (r) at the reference generalized viscoplastic strain rate ($\dot{\varepsilon}^{vp}_{mref}$) are presented as follows:

$$d\overline{p}'_{cref} = \frac{1+e_0}{\lambda-\kappa}\overline{p}'_{cref}\mathrm{d}\varepsilon_v^{vp} \tag{12}$$

$$dr = \frac{1+e_0}{\lambda-\kappa}s_L(1-r)\left(\mathrm{d}\varepsilon_v^{vp} + \mathrm{A_d}\mathrm{d}\varepsilon_s^{vp}\right) \tag{13}$$

with $\mathrm{A_d}$ being the parameter controlling the contribution of the viscoplastic deviatoric strain in the hardening process.

By applying the consistency condition for *LS* in the FVP model, the following expression can be derived at a given strain increment $d\boldsymbol{\varepsilon}$ and time increment dt:

$$df_{LS} = \frac{\partial f_{LS}}{\partial \boldsymbol{\sigma}} : \boldsymbol{D}^e[d\boldsymbol{\varepsilon} - d\boldsymbol{\varepsilon}^{vp}] + \frac{\partial f_{LS}}{\partial \boldsymbol{H}} : \frac{\partial \boldsymbol{H}}{\partial \boldsymbol{\varepsilon}^{vp}} : d\boldsymbol{\varepsilon}^{vp} + \frac{\partial f_{LS}}{\partial \dot{\boldsymbol{\varepsilon}}^{vp}} : \left[\frac{d\boldsymbol{\varepsilon}^{vp}}{dt} - \dot{\boldsymbol{\varepsilon}}_n^{vp}\right] = 0 \tag{14}$$

By incorporating the consistency equation (Eq. (14)) and the current viscoplastic strain increment (Eq. (4)) into Eq. (2), we can derive the complete general incremental form of the stress–strain-strain rate relationship and the change in the viscoplastic multiplier:

$$d\boldsymbol{\sigma}' = \boldsymbol{D}^{ep} : d\boldsymbol{\varepsilon} + \boldsymbol{D}^{er} : \dot{\boldsymbol{\varepsilon}}_n^{vp} \tag{15}$$

with

$$\boldsymbol{D}^{ep} = \boldsymbol{D}^e - \frac{\boldsymbol{D}^e : \boldsymbol{m} : \boldsymbol{b}^T : \boldsymbol{D}^e}{\boldsymbol{b}^T : \boldsymbol{D}^e : \boldsymbol{m} + h} \tag{16}$$

$$\boldsymbol{D}^{er} = \frac{\boldsymbol{D}^e : \boldsymbol{m} : \boldsymbol{d}^T}{\boldsymbol{b}^T : \boldsymbol{D}^e : \boldsymbol{m} + h} \tag{17}$$

$$\Lambda = \frac{\boldsymbol{b}^T : \boldsymbol{D}^e : d\boldsymbol{\varepsilon} - \boldsymbol{d}^T : \dot{\boldsymbol{\varepsilon}}_n^{vp}}{\boldsymbol{b}^T : \boldsymbol{D}^e : \boldsymbol{m} + h} \tag{18}$$

where:

$\boldsymbol{b} = \partial f_{LS}/\partial \boldsymbol{\sigma}, \boldsymbol{m} = \partial^\mu f_{LS}/\partial p'^\mu \cdot \partial p'/\partial \boldsymbol{\sigma}' + \partial^\mu f_{LS}/\partial q^\mu \cdot \partial q/\partial \boldsymbol{\sigma}', \boldsymbol{d} = \partial f_{LS}/\partial \dot{\boldsymbol{\varepsilon}}^{vp}$, $\boldsymbol{j} = \partial \boldsymbol{H}/\partial \boldsymbol{\varepsilon}^{vp}$, and.

h is the viscoplastic modulus giving the expression of

$$h = -\boldsymbol{d}^T : \boldsymbol{j} : \boldsymbol{m} - \frac{1}{dt}\boldsymbol{d}^T : \boldsymbol{m} \tag{19}$$

It should be noted that this model comprises 9 parameters (listed in Table. 2), the majority of which can be determined either by direct calibration or by fitting results from conventional drained/undrained triaxial tests and oedometric/isotropic compression tests. The detailed calibration procedure can be found in Cheng et al. (2024).

4. Explicit integration of constitutive equations

This section extends the explicit adaptive substepping integration algorithm, originally developed for rate-independent elastoplasticity, to the fractional consistency viscoplastic model (FVP). This extension retains the two core advantages inherent in the forward integration scheme with minimal modifications, namely substepping strategy with automatic error control and accurate stress drift correction to the rate-dependent loading surface, while, for the first time, systematically implementing and assessing different–order Runge–Kutta substepping schemes (RK12, RK23, RK34, RK45) for a fractional two–surface viscoplastic model at both the material–point and boundary–value–problem levels..

4.1. Main phase assessment

During the initialization phase of the stress updating, the prescribed total strain increment ($\Delta\boldsymbol{\varepsilon}$) and time increment ($\Delta t$) obtained from the finite element iteration are utilized for the elastic trial calculation on the current stress state ($\boldsymbol{\sigma}'_0$, $\boldsymbol{H}_0$, $\dot{\boldsymbol{\varepsilon}}_0^{vp}$):

$$\Delta\boldsymbol{\sigma}'^{tr} = \boldsymbol{D}^e : \Delta\boldsymbol{\varepsilon}\cos\theta = \frac{\partial f_{LS}/\partial \boldsymbol{\sigma}' : \Delta\boldsymbol{\sigma}'^{\mathrm{tr}}}{\left\|\partial f_{LS}/\partial \boldsymbol{\sigma}'\right\| \cdot \left\|\Delta\boldsymbol{\sigma}'^{\mathrm{tr}}\right\|} \tag{20}$$

where $\Delta\boldsymbol{\varepsilon}$ is assumed purely elastic, and θ denotes the angle between the loading surface gradient and the trial stress increment $\Delta\boldsymbol{\sigma}'^{tr}$.

The current phase is determined by comparing the relative magnitudes of $f_{LS}\left(\boldsymbol{\sigma}'_0 + \Delta\boldsymbol{\sigma}'^{\mathrm{tr}}, \boldsymbol{H}_0, \dot{\boldsymbol{\varepsilon}}_0^{vp}\right)$ and $\cos\theta$ against the tolerance limits FTOL and LTOL. Correspondingly, three scenarios are possible, namely.

(i) If $f_{LS}\left(\boldsymbol{\sigma}'_0 + \Delta\boldsymbol{\sigma}'^{\mathrm{tr}}, \boldsymbol{H}_0, \dot{\boldsymbol{\varepsilon}}_0^{vp}\right) <$ -FTOL, the stress state undergoes a purely elastic stage, the trial stress can be directly accepted as the updated stress. In this case, because the present FVP model is formulated within a rate-independent elastic and rate-dependent viscoplastic framework, no viscous strain is generated in scenario (i). For more general situations, one could adopt a viscoelastic-viscoplastic formulation, in which time-dependent deformation may also occur within the nominally elastic regime.

(ii) If $f_{LS}\left(\boldsymbol{\sigma}'_0 + \Delta\boldsymbol{\sigma}'^{\mathrm{tr}}, \boldsymbol{H}_0, \dot{\boldsymbol{\varepsilon}}_0^{vp}\right) >$ -FTOL and $\cos\theta <$ -LTOL, the stress state transitions from elasticity to elasto-viscoplasticity, and the iterative interpolation is then applied to ensure the condition $\left|f_{LS}\left(\boldsymbol{\sigma}'_0 + \alpha\,\Delta\boldsymbol{\sigma}'^{\mathrm{tr}}, \boldsymbol{H}_0, \dot{\boldsymbol{\varepsilon}}_n^{vp}\right)\right| <$ FTOL is satisfied. Here, the parameter α is the strain proportion factor for transitioning from an elastic state to the yield surface. It is determined by solving the nonlinear equation $f_{LS}\left(\boldsymbol{\sigma}'_0 + \alpha\,\Delta\boldsymbol{\sigma}'^{\mathrm{tr}}, \boldsymbol{H}_0, \dot{\boldsymbol{\varepsilon}}_n^{vp}\right) = 0$ via an iterative interpolation method, such as the Pegasus algorithm (Sloan et al., 2001).

(iii) If $f_{LS}\left(\boldsymbol{\sigma}'_0 + \Delta\boldsymbol{\sigma}'^{\mathrm{tr}}, \boldsymbol{H}_0, \dot{\boldsymbol{\varepsilon}}_0^{vp}\right) >$ -FTOL and $\cos\theta >$ -LTOL are satisfied, the stress state is purely elasto-viscoplastic. In this case, the stress updating requires the explicit substepping stress integration method detailed in Section 4.2, followed by the drift correction scheme described in Section 4.3 to ensure the stress point remains on the loading surface. The complete computational procedure is systematically outlined in Table 1. It is noteworthy that under certain loading paths (e.g., undrained creep), the imposed total strain increment ($\Delta\boldsymbol{\varepsilon}$) is zero. In this special case, the elastic trial stress increment vanishes ($\Delta\boldsymbol{\sigma}'^{tr} = 0$), and the stress state should be directly via the purely viscoplastic substepping procedure describe above.

Table 1
Procedures for explicit adaptive Runge-Kutta substepping with error control.

Step	Description
1	Enter initial state: effective stress $\boldsymbol{\sigma}'_n$, internal variables $\mathbf{H}_n$, viscoplastic strain rate $\dot{\boldsymbol{\varepsilon}}^{vp}_n$, void ratio e_n; Enter total strain increment $\Delta\boldsymbol{\varepsilon}$, real time increment Δt, local error tolerance *STOL*; Set the order of the Runge-Kutta method s; Initialize the accumulated pseudo-time $T = 0$ and the pseudo-step size $\Delta T = 1.0$
2	If $T < 1.0$, continue the integration (proceed to Steps 3–9)
3	For a given pseudo-step size ΔT, the strain increment and real time increment are subdivided into sub-increments: $\Delta\boldsymbol{\varepsilon}_j = \Delta T{\cdot}\Delta\boldsymbol{\varepsilon}$ and $\Delta t_j = \Delta T{\cdot}\Delta t$
4	According to the selected RK order s, compute the intermediate states ($\widetilde{\boldsymbol{\sigma}}'_i$, $\widetilde{\mathbf{H}}_i$, $\widetilde{\dot{\boldsymbol{\varepsilon}}}^{vp}_{n,i}$, $\widetilde{e}_i$) and the corresponding increments ($\Delta\boldsymbol{\sigma}'_i$, $\Delta\mathbf{H}_i$, $\Delta\dot{\boldsymbol{\varepsilon}}^{vp}_{n,i}$) for each stage $i = 1,\ldots, s$ using Eqs. together with the coefficients listed in Appendix A.
5	Assemble the lower-order solution ($\boldsymbol{\sigma}'_{n+1}$, $\mathbf{H}_{n+1}$, $\dot{\boldsymbol{\varepsilon}}^{vp}_{n,n+1}$, e_{n+1}) and the higher-order reference solution ($\widetilde{\boldsymbol{\sigma}}'_{n+1}$, $\widetilde{\mathbf{H}}_{n+1}$, $\widetilde{\dot{\boldsymbol{\varepsilon}}}^{vp}_{n,n+1}$, $\widetilde{e}_{n+1}$) via Eq. (22)
6	Estimate the relative local error R_n for the current pseudo-step according to the chosen RK order s. If $R_n \le STOL$: the substep is accepted. Update the stress and state variables, apply the drift correction described in Section 4.3, and advance the accumulated pseudo-time $T \leftarrow T + \Delta T$
7	If $R_n > STOL$: the substep is rejected. First compute the scaling factor: $\rho = \max\left\{0.9\left(\frac{STOL}{R_n}\right)^{1/s},\ 0.1\right\}$ Then update the pseudo-step size: $\Delta T \leftarrow \max\{\rho{\cdot}\Delta T,\ \Delta T_{\min}\}$ Return to step 2 to recompute the substep.
8	After a substep is accepted, predict the size of the next pseudo-step: $\rho = \min\left\{0.9\left(\frac{STOL}{R_n}\right)^{1/s},\ 1.1\right\}$ If the previous substep is rejected, limit the growth: $\rho \leftarrow \min\{\rho,\ 1.0\}$ Update the pseudo-step: $\Delta T \leftarrow \rho{\cdot}\Delta T$
9	To ensure numerical stability and to prevent the pseudo–time from exceeding 1.0, impose the following constraints: $\Delta T \leftarrow \max\{\Delta T,\ \Delta T_{\min}\}$, $\Delta T \leftarrow \min\{\Delta T,\ 1.0 - T\}$ Return to Step 2 and continue until $T = 1.0$.
10	When $T = 1.0$, the stress and state variables are all updated.

Table 2
Model parameters and state variables for both element tests and boundary value problems.

		Definition	Value
Model parameter	λ	Slope of the reference compression line	0.18
	κ	Initial slope of the swelling line in $v - \ln p'$ space	0.02
	ν	The vertical Poisson's ratio	0.3
	φ	Internal frictional angle	18.5
	μ	Fractional order	0.6
	s_L	Isotropic progressive hardening rate	8
	A_d	Relative ratio of viscoplastic shear to volumetric strain in deviatoric loading paths	1
	β	The ratio of secondary compression index to current tangent compression slope	0.06
	$\dot{\varepsilon}^{vp}_{mref}$	Reference generalized viscoplastic strain rate after primary consolidation	8×10^{-8} s^{-1}
State variables	e	Current void ratio	–
	$\overline{p'_{cref}}$	The reference pre-consolidation pressure	–
	r	The yield ratio	–
	$\dot{\varepsilon}^{vp}_m$	Current generalized viscoplastic strain rate	–

4.2. Explicit substepping integration schemes

By employing the smoothing technique, the differential stress–strain-strain rate relationship of the FVP model is generally formulated as Eq. (15). Thus, the relevant constitutive equations are linearized and solved by introducing a pseudo-time T for a given strain increment $\Delta\boldsymbol{\varepsilon}$ and a given real time increment Δt:

$$\begin{cases} \dfrac{d\boldsymbol{\sigma}'}{dT} = \boldsymbol{D}^{ep} : \Delta\boldsymbol{\varepsilon} + \boldsymbol{D}^{er} : \dfrac{\dot{\boldsymbol{\varepsilon}}^{vp}_n}{dT} & \dfrac{d\mathbf{H}}{dT} = \Lambda\boldsymbol{j} \\ \dfrac{d\dot{\boldsymbol{\varepsilon}}^{vp}}{dT} = \Lambda\boldsymbol{m} - \dot{\boldsymbol{\varepsilon}}^{vp}_n & \dfrac{de}{dT} = -(1+e)\Delta\varepsilon_v \end{cases} \tag{21}$$

where pseudo-time $T(0 < T \le 1)$ is defined as $T = (t - t_0)/\Delta t$ with t_0 and $t_0 + \Delta t$ being the start time and the end time of the imposed loading increment, and $\Delta\varepsilon_v$ is the volumetric strain increment. Here, the pseudo-time T is introduced purely as a numerical parameter to describe the progression of stress integration within a given loading increment, and it does not possess any independent physical meaning beyond the imposed real time increment Δt. For each increment, the stress update is performed by scaling both the strain increment $\Delta\boldsymbol{\varepsilon}$ and the real time increment Δt by T (i.e., using $T{\cdot}\Delta\boldsymbol{\varepsilon}$ and $T{\cdot}\Delta t$), so that the evolution of all internal variables remains consistent with the actual strain–time path while enabling convenient step-size control and error estimation in the numerical integration.

Following the general computational process of the explicit substepping integration, the given step size (strain increment and real time increment) is typically divided into several substeps by using four different order Runge-Kutta algorithms (RK12, RK23, RK34, and RK45) with truncation error control techniques. Under the known initial state ($\boldsymbol{\sigma}'_n$, $\mathbf{H}_n$, $\dot{\boldsymbol{\varepsilon}}^{vp}_n$, e_n) and given pseudo-time increment ΔT, the explicit substepping calculations are completed for s iterations and summed up:

$$\begin{cases} \boldsymbol{\sigma}'_{n+1} = \boldsymbol{\sigma}'_n + \sum_{i=1}^{s} b_i \Delta\boldsymbol{\sigma}'_i & \mathbf{H}_{n+1} = \mathbf{H}_n + \sum_{i=1}^{s} b_i \Delta\mathbf{H}_i \\ \dot{\boldsymbol{\varepsilon}}^{vp}_{n,n+1} = \dot{\boldsymbol{\varepsilon}}^{vp}_{n,n} + \sum_{i=1}^{s} b_i \Delta\dot{\boldsymbol{\varepsilon}}^{vp}_{n,i} & e_{n+1} = (e_n + 1)\exp(-\Delta T{\cdot}\Delta\varepsilon_v) - 1 \end{cases} \tag{22}$$

where s is determined by the specific order of the Runge-Kutta algorithm ($s = 2, 3, 5, 6$ corresponds to RK12, RK23, RK34, and RK45, respectively). b_i is a set of summation coefficients for calculating the current state ($\boldsymbol{\sigma}'_{n+1}$, $\mathbf{H}_{n+1}$, $\dot{\boldsymbol{\varepsilon}}^{vp}_{n,n+1}$, e_{n+1}), and $\widetilde{b}_i$ represents an alternative set of b_i to compute the current trial state ($\widetilde{\boldsymbol{\sigma}}'_{n+1}$, $\widetilde{\mathbf{H}}_{n+1}$, $\widetilde{\dot{\boldsymbol{\varepsilon}}}^{vp}_{n,n+1}$, $\widetilde{e}_{n+1}$). The stress and state variable increments ($\Delta\boldsymbol{\sigma}'_i$, $\Delta\mathbf{H}_i$, $\Delta\dot{\boldsymbol{\varepsilon}}^{vp}_{n,i}$) at i stage ($i = 1,\ldots, s$) are calculated as follows:

$$\begin{cases} \Delta\sigma'_i = \boldsymbol{D}^{ep}\left(\widetilde{\sigma}'_i, \widetilde{\boldsymbol{H}}_i, \widetilde{\dot{\boldsymbol{\varepsilon}}}^{vp}_{n,i}, \widetilde{e}_i\right) : \Delta T_n{\cdot}\Delta\boldsymbol{\varepsilon} + \boldsymbol{D}^{er}\left(\widetilde{\sigma}'_i, \widetilde{\boldsymbol{H}}_i, \widetilde{\dot{\boldsymbol{\varepsilon}}}^{vp}_{n,i}, \widetilde{e}_i\right) : \widetilde{\dot{\boldsymbol{\varepsilon}}}^{vp}_{n,i} \\ \Delta\mathbf{H}_i = \widetilde{\Lambda}_i\left(\widetilde{\sigma}'_i, \widetilde{\boldsymbol{H}}_i, \widetilde{\dot{\boldsymbol{\varepsilon}}}^{vp}_{n,i}, \widetilde{e}_i\right)\boldsymbol{j}\left(\widetilde{\sigma}'_i, \widetilde{\boldsymbol{H}}_i, \widetilde{\dot{\boldsymbol{\varepsilon}}}^{vp}_{n,i}, \widetilde{e}_i\right) \\ \Delta\dot{\boldsymbol{\varepsilon}}^{vp}_{n,i} = \widetilde{\Lambda}_i\left(\widetilde{\sigma}'_i, \widetilde{\boldsymbol{H}}_i, \widetilde{\dot{\boldsymbol{\varepsilon}}}^{vp}_{n,i}, \widetilde{e}_i\right) \Big/ (\Delta T_n{\cdot}\Delta t){\cdot}\boldsymbol{m}\left(\widetilde{\sigma}'_i, \widetilde{\boldsymbol{H}}_i, \widetilde{\dot{\boldsymbol{\varepsilon}}}^{vp}_{n,i}, \widetilde{e}_i\right) - \widetilde{\dot{\boldsymbol{\varepsilon}}}^{vp}_{n,i} \end{cases} \tag{23}$$

with

$$\begin{cases} \widetilde{\boldsymbol{\sigma}}'_i = \boldsymbol{\sigma}'_n + \sum_{k=1}^{i-1} a_{ik}\Delta\boldsymbol{\sigma}'_k & \widetilde{\mathbf{H}}_i = \mathbf{H}_n + \sum_{k=1}^{i-1} a_{ik}\Delta\mathbf{H}_k \\ \widetilde{\dot{\boldsymbol{\varepsilon}}}^{vp}_{n,i} = \dot{\boldsymbol{\varepsilon}}^{vp}_{n,n} + \sum_{k=1}^{i-1} a_{ik}\Delta\dot{\boldsymbol{\varepsilon}}^{vp}_{n,k} & \widetilde{e}_i = (e_n + 1)\exp\left(-\sum_{k=1}^{i-1} a_{ik}{\cdot}\Delta T{\cdot}\Delta\varepsilon_v\right) - 1 \end{cases} \tag{24}$$

where a_{ik} denotes a matrix of summation coefficients employed to determine the intermediate trial state ($\widetilde{\boldsymbol{\sigma}}'_i$, $\widetilde{\mathbf{H}}_i$, $\widetilde{\dot{\boldsymbol{\varepsilon}}}^{vp}_{n,i}$, $\widetilde{e}_i$). The three coefficient vectors/matrices (a_{ik}, b_i, $\widetilde{b}_i$) used in different Runge-Kutta order methods by Fehlberg (1969) are recommended and detailed in Appendix A.

Noting that the local truncation errors in the current state ($\boldsymbol{\sigma}'_{n+1}$, $\mathbf{H}_{n+1}$, $\dot{\boldsymbol{\varepsilon}}^{vp}_{n,n+1}$, e_{n+1}) and the current trial state ($\widetilde{\boldsymbol{\sigma}}'_{n+1}$, $\widetilde{\mathbf{H}}_{n+1}$, $\widetilde{\dot{\boldsymbol{\varepsilon}}}^{vp}_{n,n+1}$, $\widetilde{e}_{n+1}$) are $\boldsymbol{O}((\Delta T)^s)$ and $\boldsymbol{O}\left((\Delta T)^{s+1}\right)$, respectively, an estimate of the relative error R_n is defined to adjust the size of the substep:

$$R_n = \max\left\{ \frac{\|\widetilde{\boldsymbol{\sigma}}'_{n+1} - \boldsymbol{\sigma}'_{n+1}\|}{\|\widetilde{\boldsymbol{\sigma}}'_{n+1}\|}, \frac{\left\|\widetilde{\boldsymbol{H}}_{n+1} - \boldsymbol{H}_{n+1}\right\|}{\left\|\widetilde{\boldsymbol{H}}_{n+1}\right\|}, \frac{\|\widetilde{\dot{\boldsymbol{\varepsilon}}}^{vp}_{n,n+1} - \dot{\boldsymbol{\varepsilon}}^{vp}_{n,n+1}\|}{\|\widetilde{\dot{\boldsymbol{\varepsilon}}}^{vp}_{n,n+1}\|}, EPS \right\} \quad (25)$$

where *EPS* is a machine constant that denotes the minimum relative error calculable. The current substep is accepted if its relative error meets a small positive local tolerance STOL, and rejected if it does not. In both cases, the next pseudo-step size is updated according to

$$\Delta T^{i+1} = \rho \cdot \Delta T^{i} \quad (26)$$

with a required factor ρ given by

$$\rho = \min\{ \rho_{\max}, \quad \max\{ \rho_{\min}, \quad (\boldsymbol{STOL}/R_n)^{1/s} \} \} \quad (27)$$

with$\rho_{\min} = 0.1$, $\rho_{\max}$=1.1. To start the whole integration procedure, ΔT_n is initially set to 1. If a substep is rejected with $R_n >$ STOL, its size would be reduced with Eqs. (26) and (27) for a new integration until $\sum \Delta T_n = 1$. The complete algorithmic workflow, including error control and step–size adaptation, is summarized in Table 1.

4.3. Drift correction to the rate-dependent loading surface

After completing a successful pseudo-substep (i.e., accepted step 6 in Table.1), the updated stress state must strictly remain on the rate-dependent loading surface. However, due to the linearization inherent in the explicit Runge-Kutta integration of Eq. (23), the computed stress may drift away from the loading surface. This drift violates the *KKT* complementarity condition (Eq. (5) and can accumulate numerical errors if left uncorrected

$$\left| f_{LS}\left(\boldsymbol{\sigma}'_{n+1}, \boldsymbol{H}_{n+1}, \dot{\boldsymbol{\varepsilon}}^{vp}_{n,n+1}\right) \right| > FTOL \quad (28)$$

Consequently, various correction techniques, e.g., consistent correction (Potts and Gens, 1985), and normal correction scheme, are strongly recommended to project the stress point back onto the loading surface. In this work, a normal correction scheme (Sloan and Randolph, 1982) is employed after each accepted pseudo-substep to project the stress point back onto the current loading surface while keeping fixed the total strain increment $\Delta\boldsymbol{\varepsilon}$, the real time increment Δt, and the newly updated internal variables $\boldsymbol{H}_{n+1}$ and viscoplastic strain rate $\dot{\boldsymbol{\varepsilon}}^{vp}_{n,n+1}$. Because the preceding substepping integration already enforces the evolution laws for hardening and viscoplastic flow within a controlled error tolerance, the drift from the loading surface in each step is kept within a very small tolerance. This “stress–only” correction, with internal variables and viscoplastic strain rate kept fixed, simply projects this small drift back onto the loading surface in stress space. For such small drifts, its effect on the hardening law and flow rule is of the same order as the integration error and does not systematically distort the material response.

The step-by-step correction procedure is as follows:

Step 1: Evaluation surface violation

Let $\boldsymbol{\sigma}'_{n+1}$ be the stress obtained from the substepping algorithm. Compute

$$f_{LS,0} = f_{LS}\left(\boldsymbol{\sigma}'_{n+1}, \boldsymbol{H}_{n+1}, \dot{\boldsymbol{\varepsilon}}^{vp}_{n,n+1}\right) \quad (29)$$

If$\left| f_{LS,0} = f_{LS}\left(\boldsymbol{\sigma}'_{n+1}, \boldsymbol{H}_{n+1}, \dot{\boldsymbol{\varepsilon}}^{vp}_{n,n+1}\right) \right| <$ FTOL, the stress is already sufficiently close to the surface and no correction is needed. Otherwise, proceed.

Step 2: Compute the correction direction and magnitude.

A first-order Taylor expansion about the drifted stress gives

$$f_{LS}\left(\boldsymbol{\sigma}'_{\text{corrected}}, \boldsymbol{H}_{n+1}, \dot{\boldsymbol{\varepsilon}}^{vp}_{n,n+1}\right) \approx f_{LS,0} + \left.\frac{\partial f_{LS}}{\partial \boldsymbol{\sigma}'}\right|_{\left(\boldsymbol{\sigma}'_{n+1}, \boldsymbol{H}_{n+1}, \dot{\boldsymbol{\varepsilon}}^{vp}_{n,n+1}\right)} : \delta\boldsymbol{\sigma}' = 0 \quad (30)$$

The correction is applied in the direction normal to the loading surface at the drifted point, i.e.,

$$\delta\boldsymbol{\sigma}' = -\delta\Lambda \left.\frac{\partial f_{LS}}{\partial \boldsymbol{\sigma}'}\right|_{\left(\boldsymbol{\sigma}'_{n+1}, \boldsymbol{H}_{n+1}, \dot{\boldsymbol{\varepsilon}}^{vp}_{n,n+1}\right)} \quad (31)$$

where is a scalar multiplier. Substituting Eq. into Eq. yields the explicit stress correction:

$$\delta\boldsymbol{\sigma}' = -\frac{f_{LS,0}}{\left(\frac{\partial f_{LS}}{\partial \boldsymbol{\sigma}'} : \frac{\partial f_{LS}}{\partial \boldsymbol{\sigma}'}\right)} \left.\frac{\partial f_{LS}}{\partial \boldsymbol{\sigma}'}\right|_{\left(\boldsymbol{\sigma}'_{n+1}, \boldsymbol{H}_{n+1}, \dot{\boldsymbol{\varepsilon}}^{vp}_{n,n+1}\right)} \quad (32)$$

Step 3: Apply the correction and update the stress

The corrected stress state is then:

$$\boldsymbol{\sigma}'_{corrected} = \boldsymbol{\sigma}'_{n+1} + \delta\boldsymbol{\sigma}' \quad (33)$$

This corrected stress is guaranteed to satisfy $f_{LS}\left(\boldsymbol{\sigma}'_{\text{corrected}}, \boldsymbol{H}_{n+1}, \dot{\boldsymbol{\varepsilon}}^{vp}_{n,n+1}\right) = 0$ to first-order accuracy. It replaces $\boldsymbol{\sigma}'_{n+1}$ as the updated stress for the next computational step.

The combination of the explicit Runge-Kutta substepping (Section 4.2) with this efficient normal drift correction provides a complete, robust, and accurate stress-updating procedure for the fractional consistency viscoplastic model within an explicit finite element framework.

5. Numerical results: Element test Analyses

This section evaluates the numerical performance of the explicit substepping algorithms (e.g., RK12, RK23, RK34, and RK45) by conducting a series of rate-dependent element tests on saturated clays using the fractional consistency two-surface viscoplastic model. For each test, the results obtained from the explicit algorithms are validated by comparison with a highly accurate reference solution. The reference solution is generated using an adaptive high-order Runge-Kutta integration method (RK45), which ensures the required accuracy by enforcing stringent relative error tolerances ($\Delta t_{\min} = 10^{-12}$, LTOL = 1.0×10^{-12}, FTOL = 1.0×10^{-12}, and STOL = 1.0×10^{-12}), and is therefore regarded as exact. Meanwhile, the model parameters employed in the simulations, calibrated based on experimental data for natural Boom clay (Cheng et al, 2024), are listed in Table. 2.

5.1. Simulations of rate-dependent element tests

In this section, we first demonstrate the capabilities of the employed fractional consistency viscoplastic model in capturing time-dependency, nonlinearity, strain and strain-rate softening, and non-associativity of highly overconsolidated clays. It should be noted that all simulations are conducted under the initial conditions of an isotropic effective stress of $p' = 1.5$ MPa, an initial void ratio of $e_0 = 0.6$, and an initial overconsolidation ratio (OCR) of 4. Four typical element tests are used for simulation validation, including: (1) odeometric compression tests under stepwise strain rates or interrupted by drained creep phases; (2) isotropic compression tests under stepwise stress rates or interrupted by undrained stress relaxation phases; (3) triaxial drained triaxial compression tests under alternating axial strain loading rates; and (4) triaxial undrained shear tests at alternating axial strain loading rates. For the sake of clarity and conciseness, only the numerical results from RK23 ($\Delta t_{\min} = 10^{-6}$, LTOL = 1.0×10^{-9}, FTOL = 10^{-6}, and STOL = 10^{-6}) and the reference RK45 are respectively presented in Figs. 2-5.

The odeometric compression simulation results show that during the interrupted drained creep phases in Fig. 2(a), the compressive strain continues to develop under constant vertical effective stress, successfully reproducing the secondary compression observed in geomaterials. Similarly, when a constant volumetric strain is applied in Fig. 3(a), the mean effective stress exhibits a continuous decrease. During stepwise

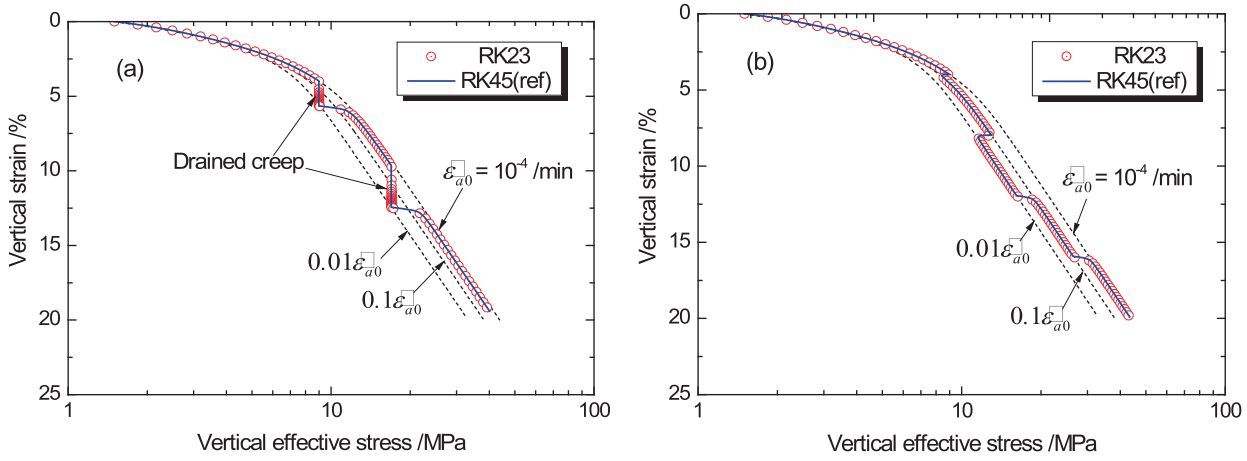

Fig. 2. Drained strain-rate controlled odeometric compression test results in (ε_v, lnσ_v') plane: (a) interrupted drained creep; (b) stepwise strain rates.

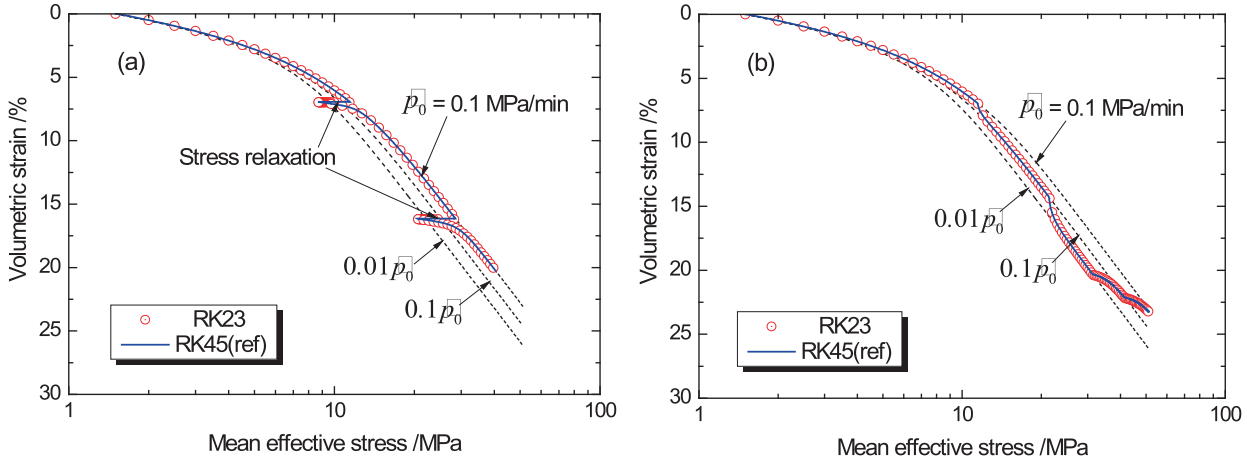

Fig. 3. Drained stress-rate controlled isotropic compression test results in (ε_v, lnp') plane: (a) interrupted undrained stress relaxation; (b) stepwise strain rates.

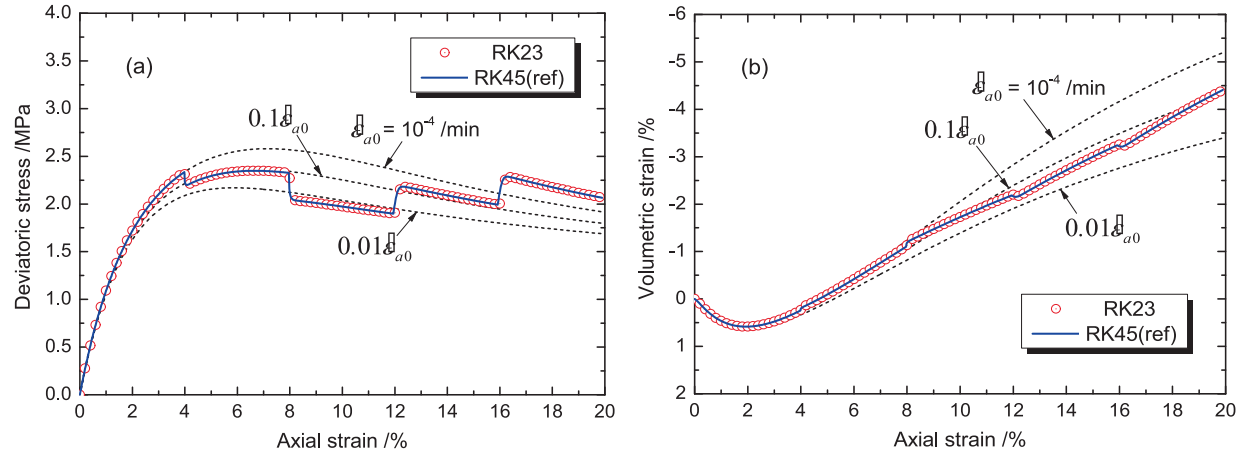

Fig. 4. Drained alternating strain-rate triaxial compression test: (a) Deviatoric stress-axial strain; (b) Volumetric strain-axial strain.

rate transitions in Figs. 2(b) and 3(b), both odeometric and isotropic compression tests exhibit isotach viscosity, which is generally consistent with experimental observations on natural Boom clays under isothermal conditions.

Under highly overconsolidated stress states, saturated clays typically exhibit pronounced nonlinearity and dilatancy behaviors. The numerical results in Figs. 4 and 5 effectively capture the stepwise features of deviatoric stress, volumetric strain, and excess pore pressure under varying axial strain rates. It should be noted that the drained triaxial test results in Fig. 4 do not fully satisfy the isotach phenomenon, which is closely related to the non-associativity of saturated clays. This aspect is well represented in the employed viscoplastic model through

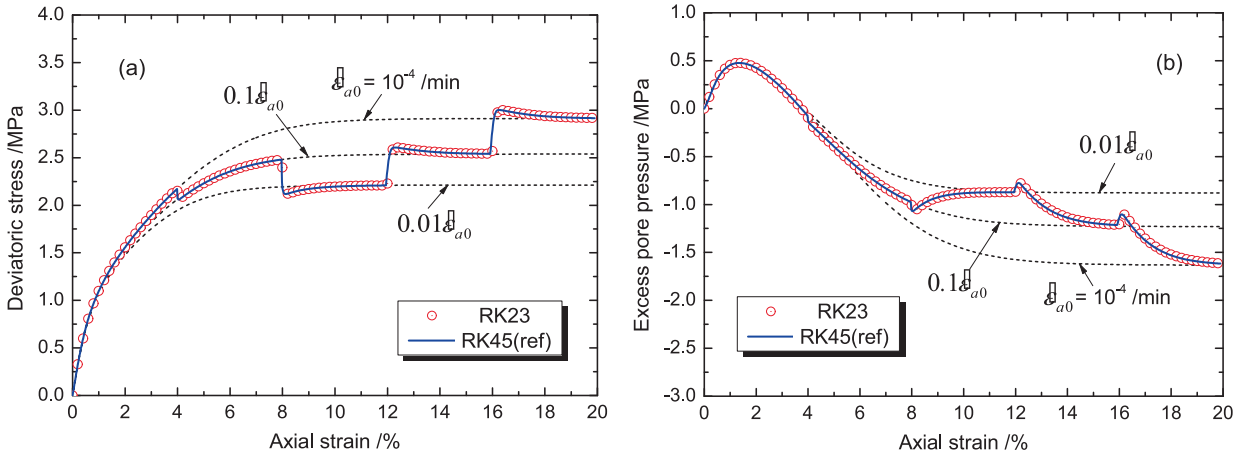

Fig. 5. Undrained alternating strain-rate triaxial compression test: (a) Deviatoric stress-axial strain; (b) Excess pore pressure-axial strain.

the non-orthogonal viscoplastic flow rule.

The comparison between the simulation results of the RK23 and reference RK45 methods shows that the two sets of results completely overlap, which also demonstrates the applicability and effectiveness of the explicit substepping integration algorithms when extended to rate-dependent viscoplastic model integration.

5.2. *Efficiency and accuracy analysis of the explicit substepping algorithms*

To quantitatively assess the applicability of explicit substepping integration algorithms for rate-dependent constitutive models, a detailed analysis is conducted to examine the influence of different Runge Kutta orders, error tolerance STOL, and imposed strain rates on accuracy and efficiency of the rate-dependent stress–strain responses.

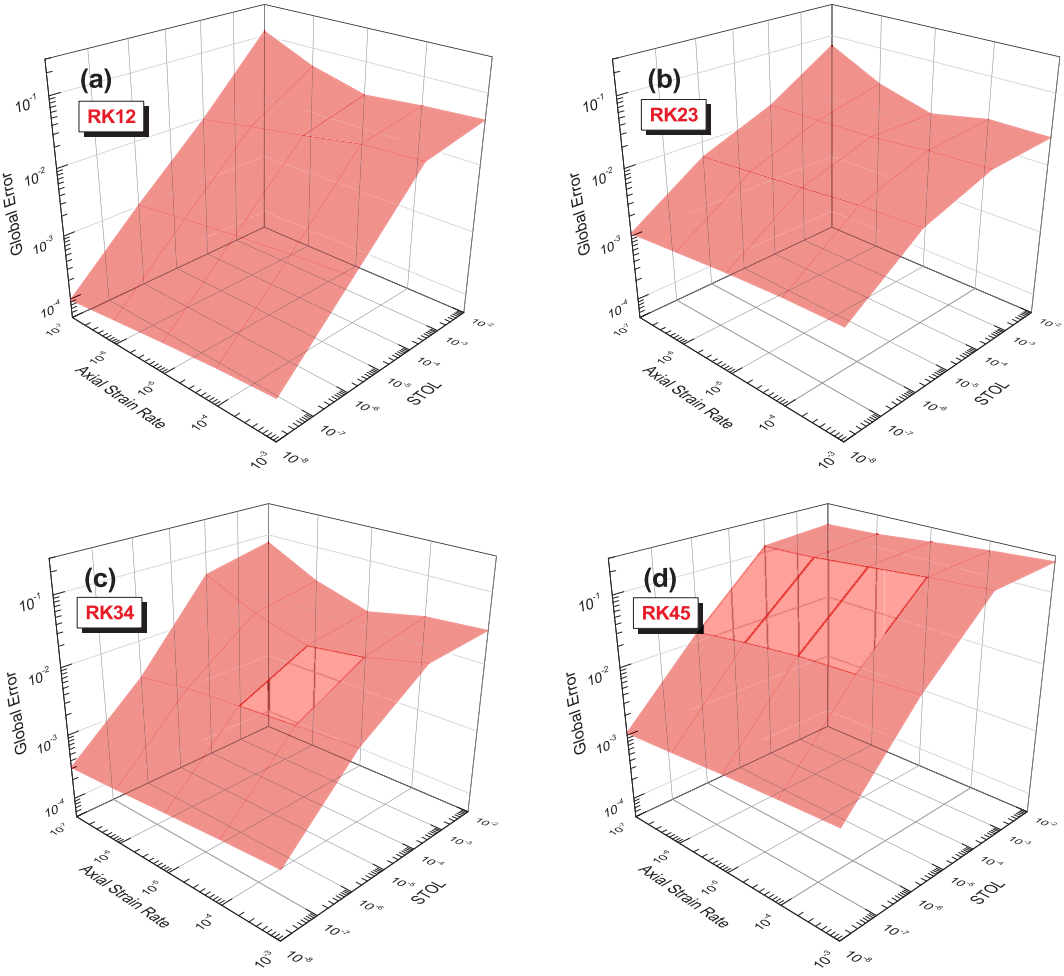

Fig. 6. Variation in the global error with axial strain rate and STOLs under drained isotropic loading paths: (a) RK12 with substepping; (b) RK23 with substepping; (c) RK34 with substepping; (d) RK45 with substepping.

Here, a relative error (ϑ) is defined as the ratio between each effective stress state computed by the current ($\boldsymbol{\sigma}'$) and the reference ($\boldsymbol{\sigma}'_{ref}$) solution, namely

$$\vartheta = \frac{\left\|\boldsymbol{\sigma}'_{ref} - \boldsymbol{\sigma}'\right\|}{\left\|\boldsymbol{\sigma}'_{ref}\right\|} \tag{34}$$

In a given stress path containing m stress sampling points, a global error function (ϑ_T) is given by

$$\vartheta_T = \sum_{i=1}^{m} |\vartheta_i| \tag{35}$$

First, the strain rate-controlled isotropic compression paths with an axial strain increment of 0.1 interpolated at 100 stress sampling points are used as examples to investigate the global cumulative error values generated by the same given strain increment under different STOLs and strain rates. The corresponding three-dimensional global error plots with four stress integration algorithms-RK12, RK23, RK34, and RK45-are presented and compared in Fig. 6. It is found that all four integration algorithms of different orders are applicable to the rate-dependent model and can maintain the global error ϑ_T at a relatively low level. As the value of STOL decreases, the global error ϑ_T correspondingly decreases rapidly; as the axial strain rate decreases, the global error ϑ_T would also slightly decrease, except for RK45.

To more thoroughly analyze the accuracy of the four different integration algorithms under rate-dependent stress paths, the global isoerror maps, an effective and reliable tool, are introduced to further evaluate the numerical results under different strain rate-controlled triaxial compression tests. Based on this method in Figs. 7 and 8, two typical STOL values, 10^{-4} and 10^{-6}, are chosen; the assigned axial strain rate spans a range from 10^{-8} min^{-1} to 10^{-3} min^{-1}; and the applied total axial strain varies from 0.005 to 0.1, with each incremental step set to be 0.005. In general, the global error ϑ_T increases with either an increase in the prescribed total axial strain or a decrease in the applied axial strain rate. When STOL is set to be 10^{-4}, the global error ϑ_T produced by the four integration algorithms is at a similar level, with the maximum global error ϑ_T ranging between 0.02 and 0.04 (the average relative error per stress point between 0.001 and 0.002). As STOL decreases exponentially to 10^{-6}, the maximum global errors ϑ_T among the four schemes would differ more noticeably but also decrease significantly to a range of 0.0002 to 0.002 (the average relative errors between 0.00001 and 0.0001). Notably, the errors associated with RK45 do not decrease despite its higher order, and it maintains a relatively small and tolerable error without further reduction. This is likely because the linearization of the plastic strain rate increment in Eq. (14) does not fully satisfy the first-order Taylor series expansion of the loading surface. The combined results from Figs. 7 and 8 highlight the importance of choosing an appropriate error tolerance and integration scheme to achieve reasonably accurate solutions.

In the assessment of computational costs associated with various integration algorithms, a common approach (Lloret-Cabot and Sheng, 2022; Dong, 2023) is to quantify the total number of the incremental constitutive relation evaluations required per computational step, explicitly incorporating both rejected substeps and subsequent stress drift correction iterations: the RK12 scheme necessitates two constitutive evaluations per substep; the RK23 requires three evaluations; the RK34 demands five evaluations; and the RK45 involves six constitutive evaluations per substep within this comprehensive assessment framework. Fig. 9 shows the relationship between the computational cost and the input axial strain increment size along two CRS odeometric loading paths, with the plots on the left representing the results under high strain rates and the plots on the right as the low strain rates. To illustrate the impact of the allowable error tolerance on computational cost in the substepping algorithms, different substepping tolerance values of STOL

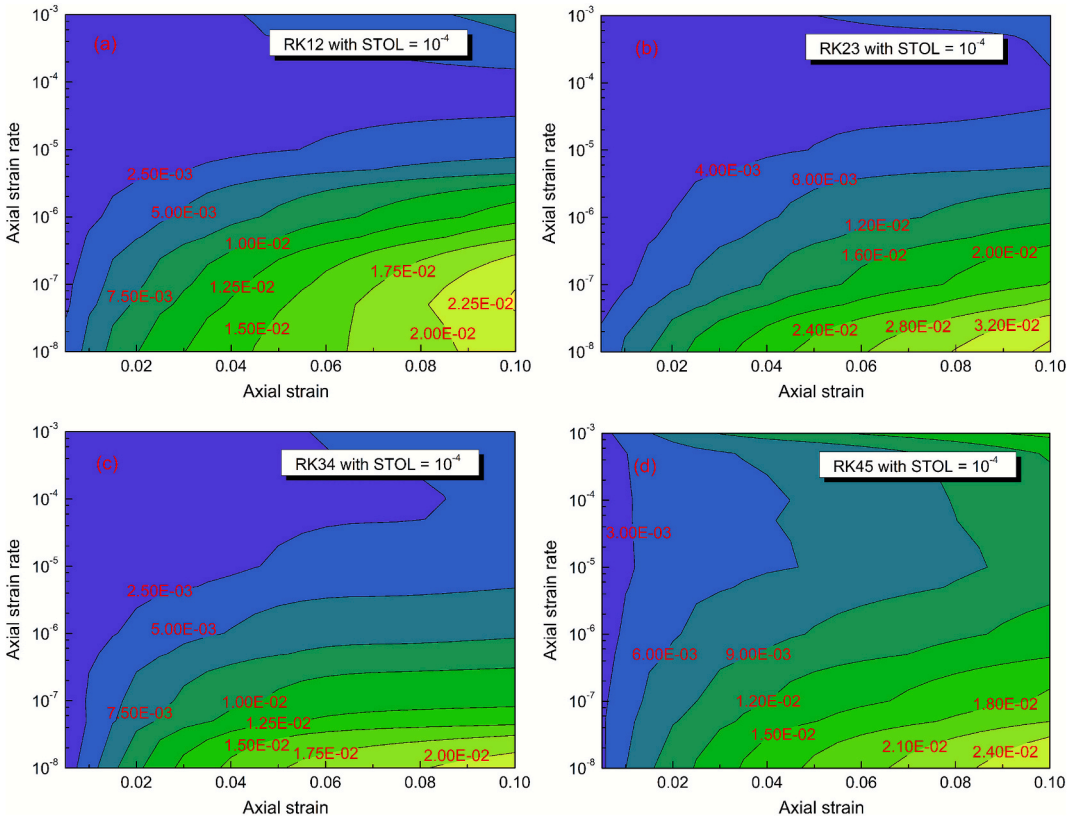

Fig. 7. Global cumulative *iso*-error maps with STOL = 10^{-4} for different strain rate-controlled drained triaxial loading paths: (a) RK12 with substepping; (b) RK23 with substepping; (c) RK34 with substepping; (d) RK45 with substepping.

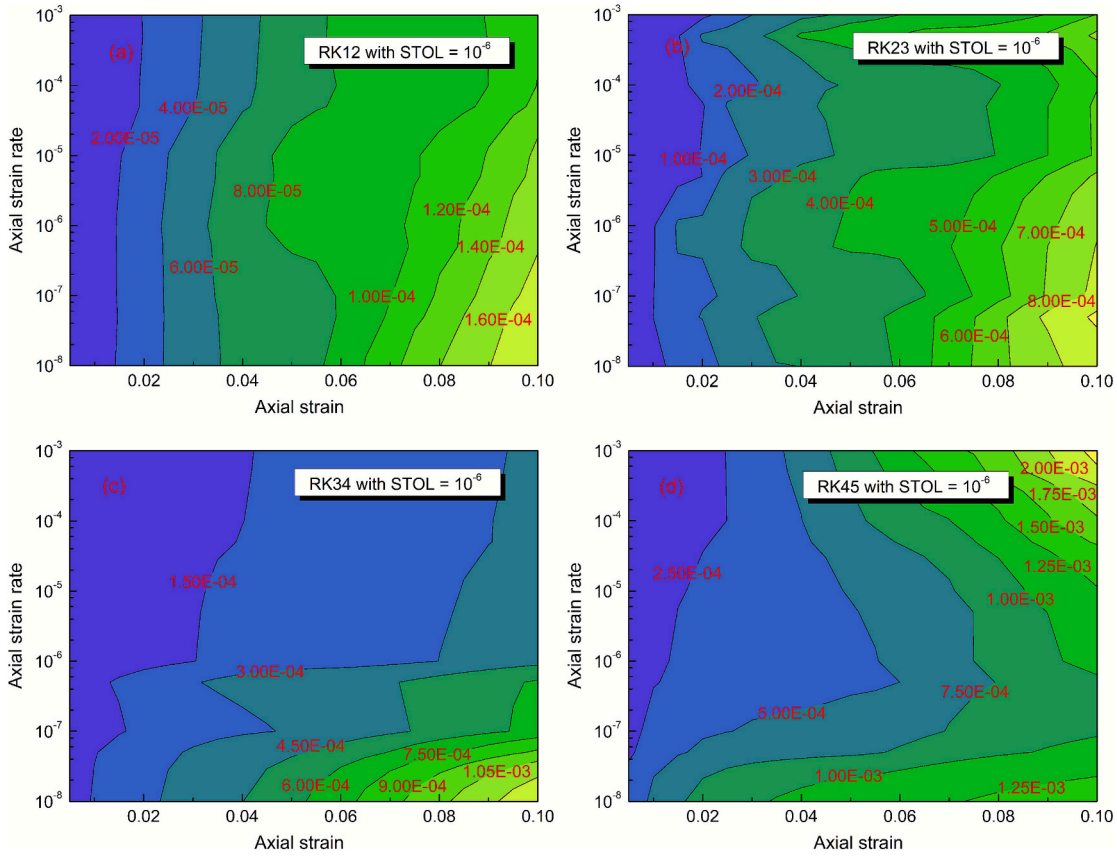

Fig. 8. Global cumulative *iso*-error maps with STOL = 10^{-6} for different strain rate-controlled drained triaxial loading paths: (a) RK12 with substepping; (b) RK23 with substepping; (c) RK34 with substepping; (d) RK45 with substepping.

$= 10^{-2}$, 10^{-4}, 10^{-6}, and 10^{-8} are selected.

Overall, the computational cost of explicit stress integration algorithms increases significantly as the value of the allowable error tolerance STOL decreases, which is especially evident in lower-order substepping algorithms (e.g., RK12, RK23), as they require more evaluations of the constitutive equations to meet the accuracy requirements. For the same input axial strain increment size, the number of constitutive relation evaluations decreases progressively as the order of the substepping integration method increases. For example, when $\Delta\varepsilon_a = 0.1$ and STOL = 10^{-6}, RK12 requires 3900 evaluations of the constitutive relations, RK23 requires 585, RK34 requires 300, and RK45 requires only 336; the difference becomes even more pronounced as the allowable error tolerance STOL becomes stricter. In other words, despite requiring one additional evaluation of the constitutive relation during each substep calculation, RK23 would reduce the overall computational cost by a factor of six. Similarly, RK34 would reduce the total computational cost by nearly half compared to RK23, even though it involves two more evaluations. RK45 and RK34 have similar computational costs and performance, indicating that simply increasing the order of the substepping scheme does not necessarily lead to improved computational efficiency and may instead increase the complexity of the integration code.

Regardless of the orders of the integration algorithms, slower loading rates would slightly increase the number of constitutive evaluations required, and this difference becomes particularly noticeable when the given axial strain increments are relatively small (e.g. $\Delta\varepsilon_a = 10^{-4}$, 10^{-5}). Therefore, in terms of integration efficiency and ease of code implementation, RK23 and RK34 are the substepping schemes that deserve more attention.

6. Numerical simulation: Boundary Value Problems

The principal objective of the simulation is to comprehensively evaluate the performance of the representative explicit RK23 and RK34 substepping integration schemes for solving time-dependent viscoplastic initial boundary value problems within the finite element framework. For this purpose, the finite element code ABAQUS/CAE is employed under quasi-static conditions to analyze the response of overconsolidated soils in two typical time-dependent scenarios: the penetration of a piezocone, as shown in Fig. 10, and the excavation of an underground gallery, as depicted in Fig. 14. In both cases, a typical set of model parameters and all state variables that evolve together with the stress and strain fields for overconsolidated soils are summarized and described in Table. 2. Meanwhile, it should be noted that the inertial effect is neglected for the sake of simplicity.

6.1. Piezocone penetration

The first numerical example is a piezocone (CPTU) penetration in an overconsolidated clay layer (OCR = 3), which is widely adopted to assess the numerical stability and efficiency of the integration algorithms for the rate-dependent model due to the significant principal stress rotation and the singularity at the edge of the cone. As plotted in Fig. 10, the ground surface load on the top of the soil layer is set to 5 kPa, the effective unit weight is 10 kN/m^3, the K_0 value is 0.8, the void ratio is 2, and the permeability is 1×10^{-7} m/s. The CPTU probe has a standard diameter of 35.7 mm (cross-sectional area 100 mm^2), a cone angle of 60°, and a penetration depth of 20 times the diameter, with two penetration speeds of 0.01 m/s ($T_0 = 714$ s) and 0.001 m/s ($T_0 = 7140$ s) to examine the rate effects. To preserve axisymmetry and improve computational efficiency, the finite element model is axisymmetric with

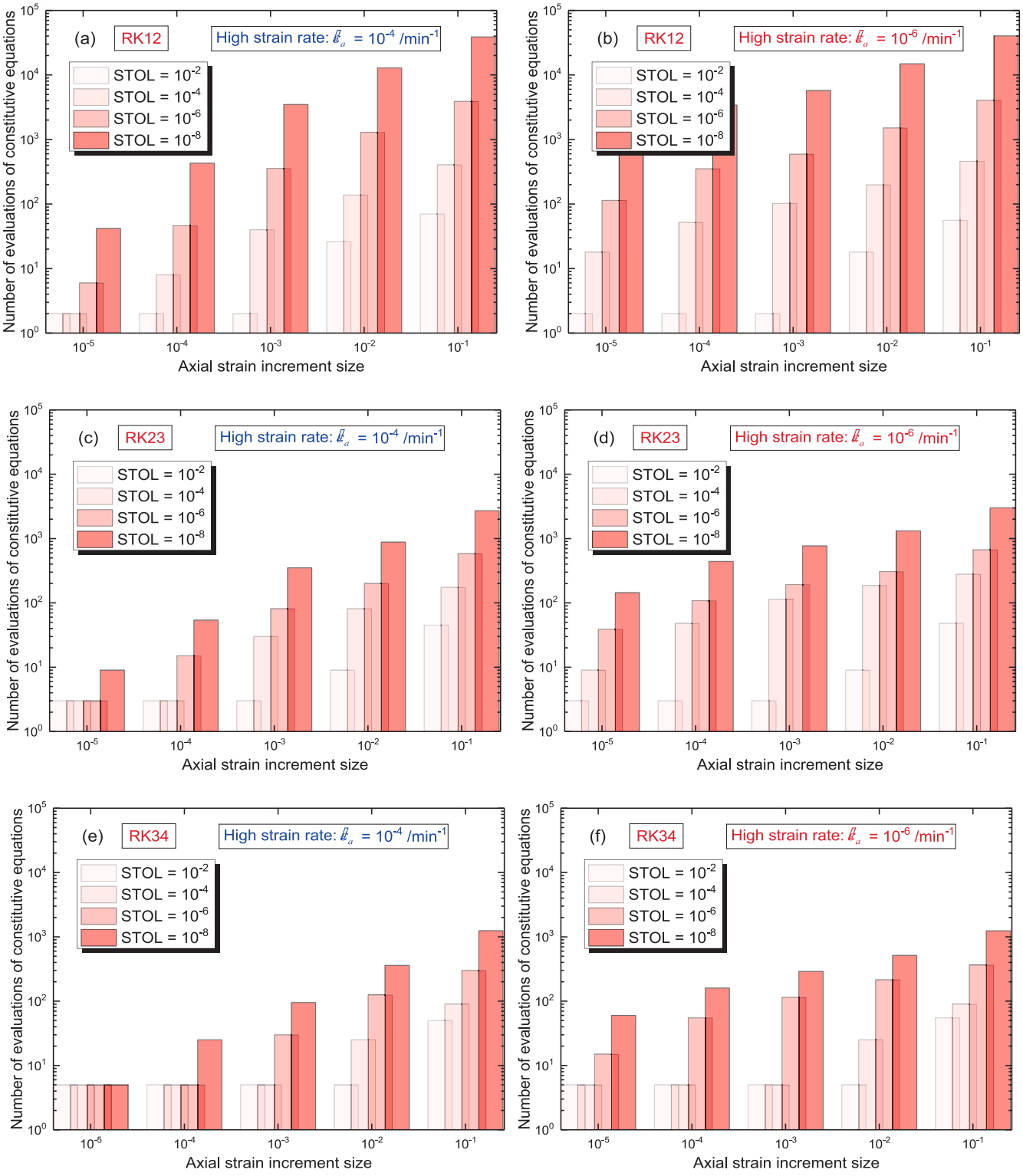

Fig. 9. Computational cost for two strain rates against imposed axial strain increment sizes: (a), (c), (e), (g) RK12, RK23, RK34, RK45 at a high strain rate of 10^{-4} min^{-1}; (b), (d), (f), (h) RK12, RK23, RK34, RK45 at a low strain rate of 10^{-6} min^{-1}.

the domain dimensions 1.5 m by 1.2 m, and the mesh sizes are refined linearly from 5 mm near the symmetry axis to 20 mm at the outer boundary. Meanwhile, to enhance the numerical stability of the contact analysis, an extremely small initial gap (e.g. 0.05 mm) is introduced between the cone and the soil to facilitate penetration. The soil domain comprises 2,690 CAX4P elements, the cone uses 85 CAX4 elements, and the normal contact between the cone and soil is modeled with a hard-contact formulation, allowing normal stress transmission only when the interfacing bodies are in compression.

Fig. 11 presents the development and distributions of excess pore pressure during different stages of piezocone penetration ($0.25T_0$, $0.5T_0$, $0.75T_0$, and T_0). Fig. 11(a-d) and 11(e-h) correspond to high and low

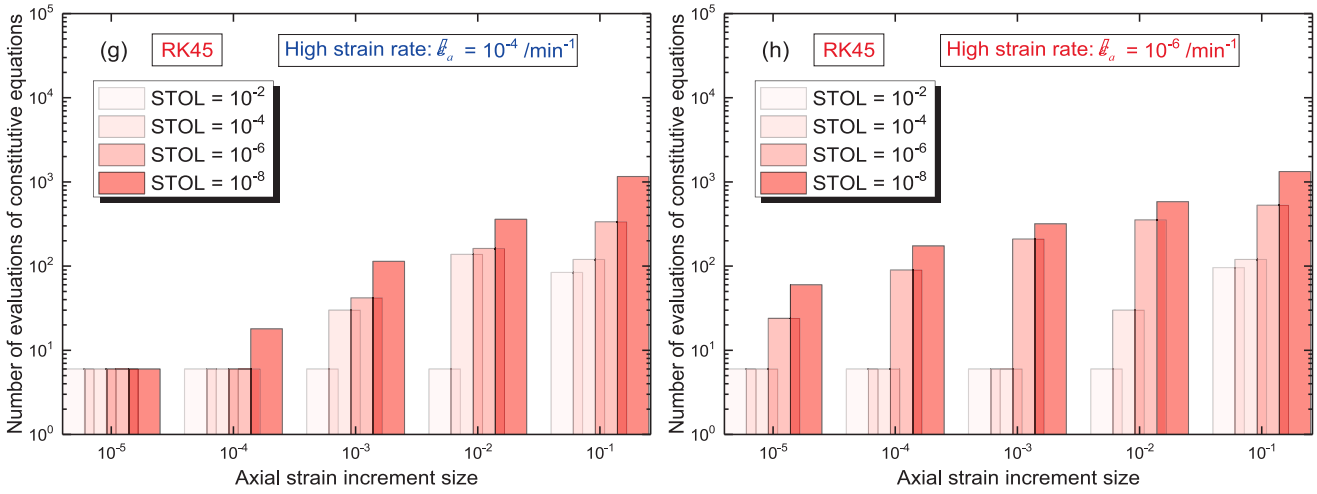

Fig. 9. (*continued*).

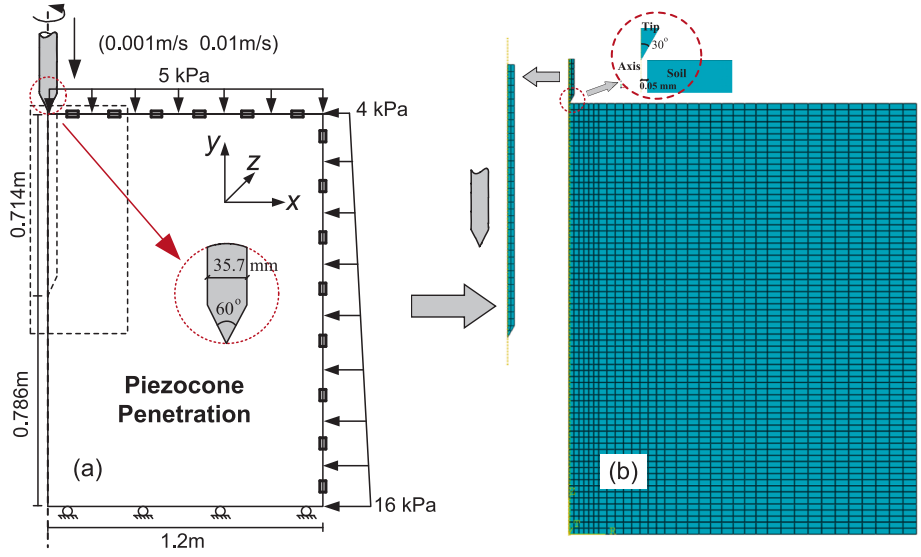

Fig. 10. Schematic representation of piezocone penetration: (a) piezocone penetration model; (b) Finite element discretization.

penetration rates, respectively. The simulations indicate that loading rate has a pronounced effect on the results: at higher rates, negative pore pressure tends to form ahead of the cone tip and positive excess pore pressure tends to develop along the sidewalls, with a wider range of pore pressure changes and stronger local concentration near the cone tip. In contrast, slower penetration allows excess pore pressure to develop more gradually and distribute over a broader region, reducing localized concentrations and producing more uniform diffusion of excess pore pressure.

Fig. 12 illustrates the time evolution of the radial effective stress at three observation points, A_1, A_2, and A_3, during the piezocone penetration, where these points are initially characterized by an approximately linear increase of p' and q with depth under a constant OCR of 3. Overall, for each observation point, the radial effective stress increases with penetration depth, then decreases and tends to stabilize. At the high penetration rate in Fig. 12(a), the magnitude of stress changes across stages is slightly greater, as reflected by wider spacing between the curves, indicating more pronounced nonuniformity in stress distribution. Additionally, the change in radial effective stress exhibits a clear lag with depth: deeper observation points experience later and larger changes. Additionally, under both high and low penetration rates, the simulation results obtained from RK23 and RK34 are in excellent agreement.

The computed results from Fig. 13 show that, under different penetration rates, the number of constitutive equation evaluations for the same observation point increases rapidly at first and then gradually levels off as time progresses. Eventually, the total counts stabilize within the ranges of 4,500 to 6,500 for the RK23 algorithm and 2,000 to 3,200 for the RK34 algorithm. There is also a noticeable lag effect with depth: deeper observation points require more equation evaluations. Furthermore, slower penetration rates tend to increase the number of constitutive equation calls required at each observation point, thereby extending the overall computation time.

In order to better highlight the sensitivity of the algorithms to penetration rates and convergence tolerances (STOL), Table. 3 compares the computational costs of RK23 and RK34 for solving piezocone pore pressure tests, including the total number of constitutive equation evaluations for all integration points and CPU time. Overall, tightening the convergence tolerance leads to substantial increases in both the number of constitutive-equation calls and CPU time. Lower penetration rates tend to produce more constitutive-equation evaluations and longer CPU time. Moreover, under looser tolerances (10^{-4} and 10^{-6}), RK34 requires more total constitutive equation evaluations than RK23. This is largely because under a given small step, large amounts of integration

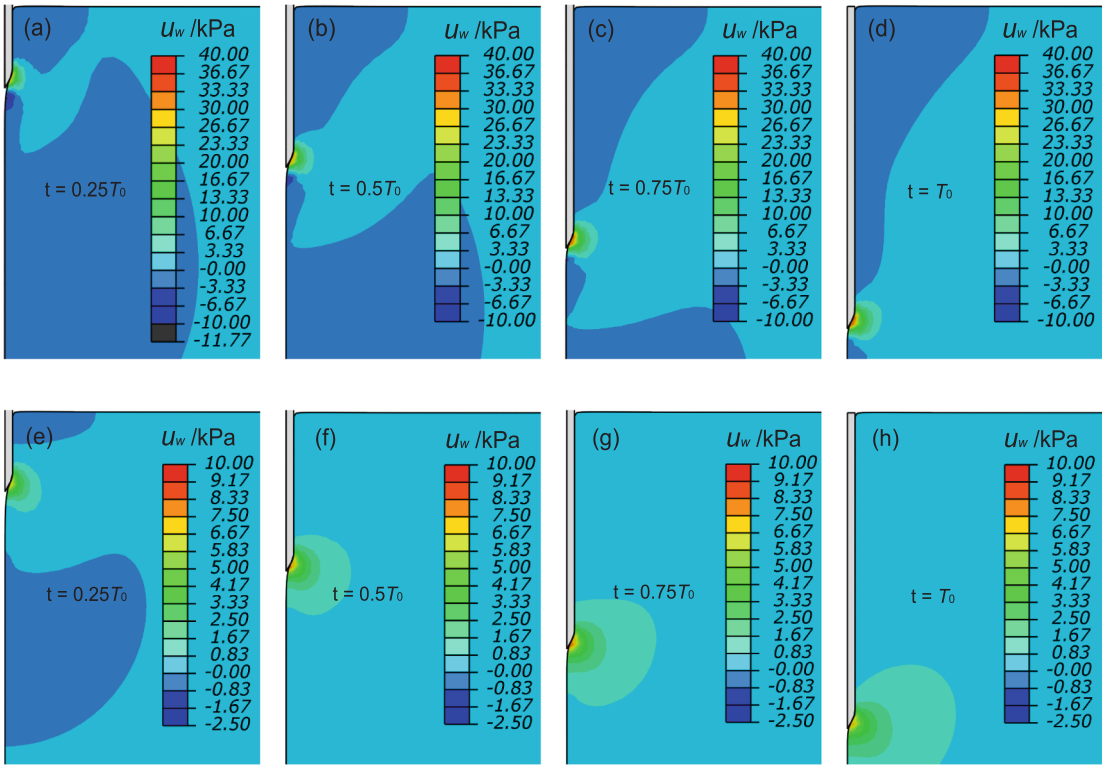

Fig. 11. Effect of penetration rates on the development of pore pressure (u_w) during piezocone penetration: (a), (b), (c) and (d) high penetration rate (0.01 m/s); (e), (f), (g) and (h) low penetration rate (0.001 m/s).

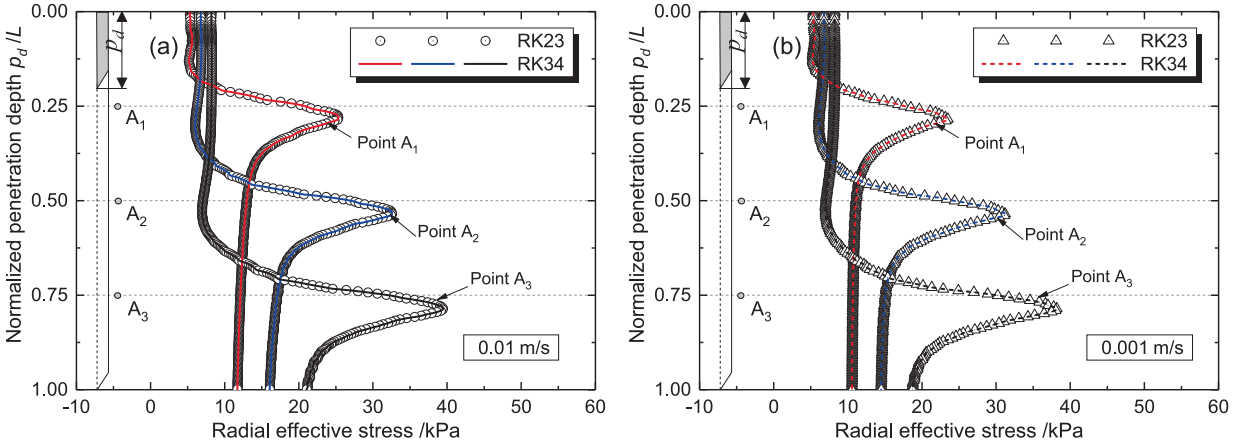

Fig. 12. Evolution of radial effective stresses for observation points A_1, A_2, A_3 during piezocone penetration: (a) high penetration rate (0.01 m/s); (b) low penetration rate (0.001 m/s).

points far from the CPTU region must be iterated five times (RK23 uses three iterations), reducing RK34's efficiency. Conversely, when the tolerance is very small (10^{-8}), even a small given step demands more constitutive-equation evaluations, further elevating the computational cost (as illustrated in Fig. 9).

6.2. Underground gallery excavation

In the second case, a quarter-domain, two-dimensional axisymmetric plane strain finite element model, as illustrated in Fig. 14, with domain dimensions of 100 m in width and 100 m in height and exhibiting symmetry about both the x- and y-axes, is developed to eliminate boundary effects induced by underground gallery excavation. For the sake of simplicity, the surrounding soil is fully saturated with in-situ vertical effective stress of 2.3 MPa, pore pressure of 2.2 MPa, K_0 value ($\sigma'_{xx}/\sigma'_{yy}$) of 0.8, initial void ratio of 0.61, initial isotropic pre-consolidation pressure of 6 MPa, and saturated permeability of 3×10^{-12} m/s. The soil layer is discretized into 2720 four-node quadrilateral plane strain elements (CPE4P), each containing four integration points, with mesh refinement applied near the excavation region to improve computational accuracy. Hydro-mechanical boundary conditions enforce fixed horizontal displacements along boundaries BE and CD, fixed vertical displacements along boundaries AC and DE, and prescribe hydraulic impermeability at all boundaries. The excavation gallery has a diameter of 2.94 m with a radial inward contraction of 0.09 m. Given the extremely low permeability and rate-dependent behaviors of deeply buried overconsolidated clay, the excavation process can be suitably approximated as an undrained contraction of a cylindrical cavity (Cheng

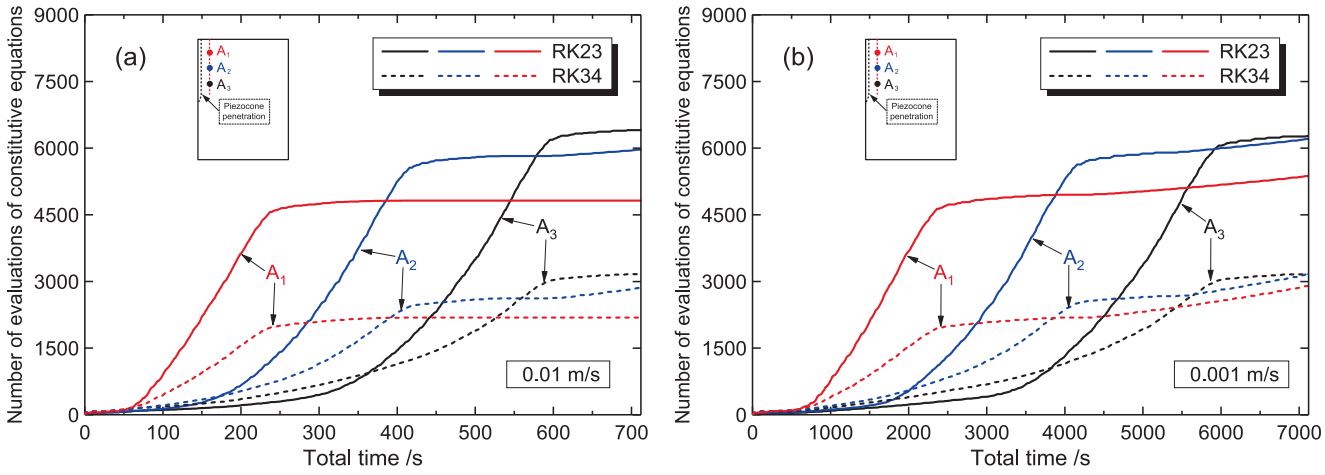

Fig. 13. Computational cost for observation points A_1, A_2, A_3 during piezocone penetration: (a) high penetration rate (0.01 m/s); (b) low penetration rate (0.001 m/s).

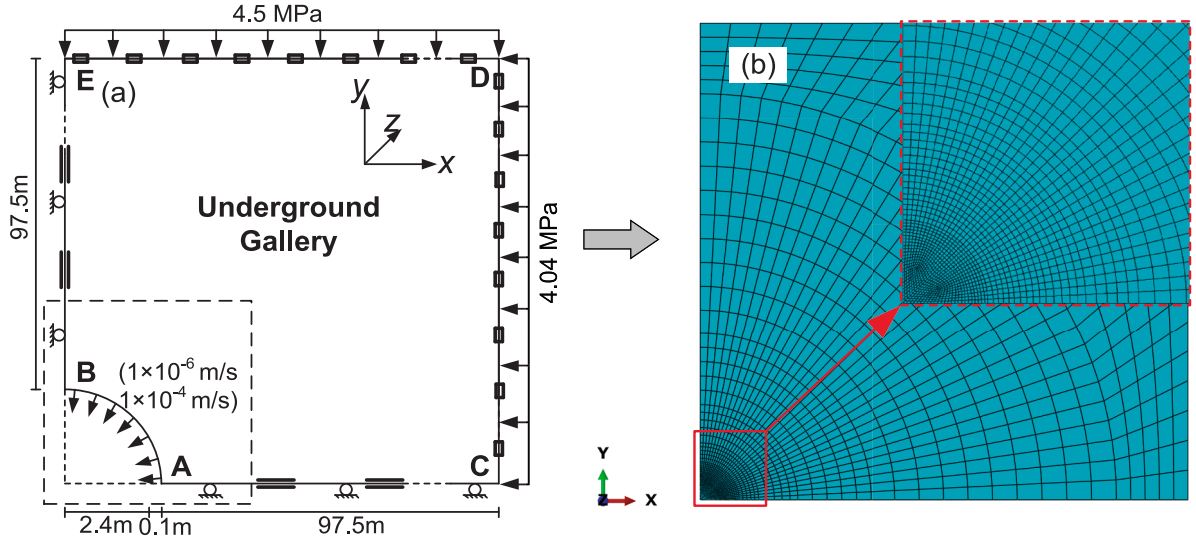

Fig. 14. Schematic representation of underground gallery excavation: (a) underground gallery model; (b) Finite element discretization.

Table. 3
. Computational costs of different substepping methods in solving piezocone penetration.

Numerical information		RK23		RK34	
penetration rate	STOL	Number of evaluations	CPU time /s	Number of evaluations	CPU time /s
High (0.01 m/s)	10^{-4}	2,648,157	1304.8	4,190,135	1873.3
	10^{-6}	4,309,653	1864.5	4,713,315	2110.8
	10^{-8}	12,322,728	1994.2	7,017,955	2269.8
Low (0.001 m/s)	10^{-4}	2,888,046	1716.9	4,602,995	2256.0
	10^{-6}	4,594,206	2051.4	5,125,165	2369.4
	10^{-8}	13,074,807	2151.8	7,504,895	2615.8

et al., 2022). Two different contraction rates, namely 1×10^{-4} m/s (high rate) maintained for 900 s, and 1×10^{-6} m/s (low rate) maintained for 90,000 s, are applied to equivalently simulate varying excavation rates.

Fig. 15 illustrates the distribution of deviatoric stress, pore pressure, and pre-consolidation pressure at the end of underground gallery excavation under two different loading rates. The comparison clearly shows that the loading rate markedly influences the simulation results. Overall, under high excavation rates in Fig. 15(a)-(c), the distributions of pore pressure, deviatoric stress, and pre-consolidation pressure all exhibit more pronounced localized concentrations, with steeper spatial gradients and greater non-uniformity. Specifically, high pore pressure is concentrated near the horizontal profile and the boundaries of the gallery for both rates, but this effect is notably amplified when the loading rate is higher. For deviatoric stress, the maximum values consistently appear at the gallery boundaries, though the peak value under high rate is approximately 20% greater than that under low rate. Additionally, a high excavation rate promotes the accumulation of pre-consolidation pressure in local regions, resulting in a distinctly uneven distribution and thereby intensifying the tendency of arching effects to develop.

Fig. 16 illustrates the distribution of three principal effective stresses along the horizontal profile (AC) after the excavation of the underground gallery. As the radial distance increases, both tangential principal effective stress σ'_y and the intermediate principal effective stress σ'_z gradually decrease to their initial values, whereas radial principal effective stress σ'_x first decreases and then increases back to the initial value. A higher excavation rate leads to greater maximum values of σ'_y and σ'_z, but has relatively little influence on σ'_x. Fig. 17 presents the effective stress path of point A at the gallery wall in the horizontal profile during the underground gallery excavation process. Starting

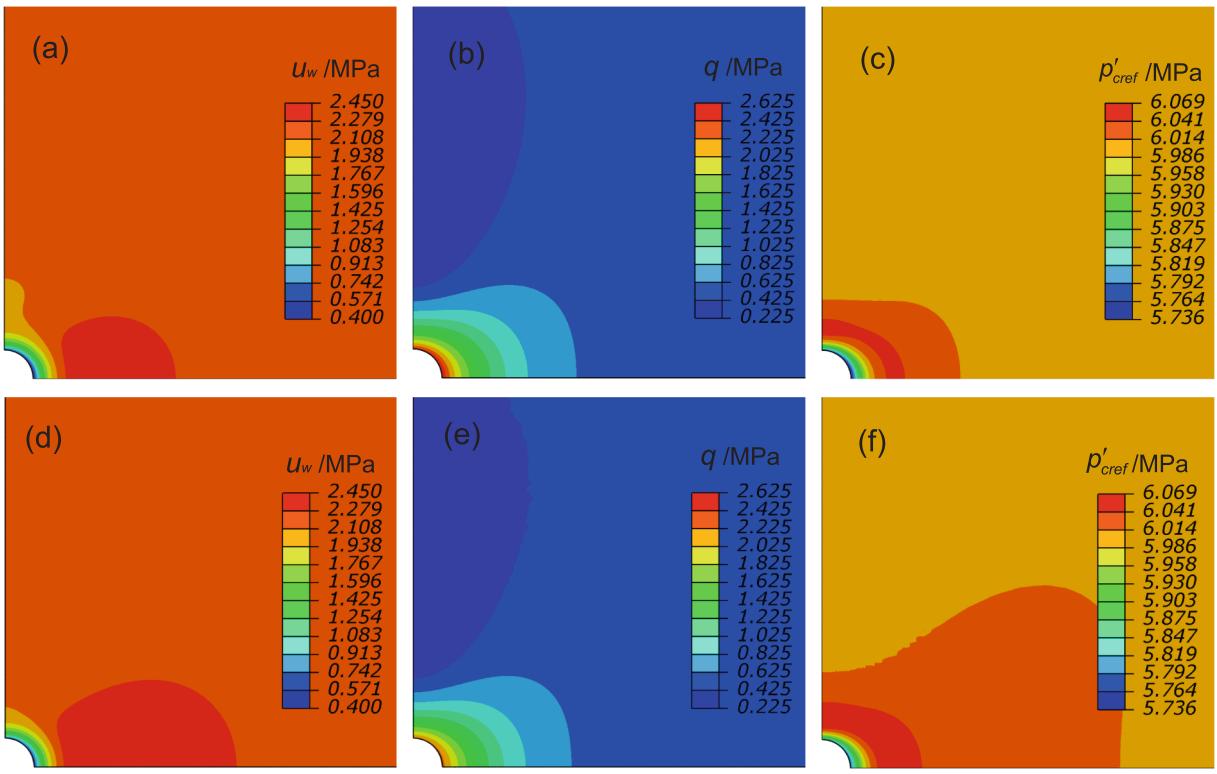

Fig. 15. Effect of excavation rates on the distribution of pore pressure (u_w), deviatoric stress (q) and pre-consolidation pressure ($\overline{p'_{cref}}$) after underground gallery excavation: (a), (c) and (e) high excavation rate (1×10^{-4} m/s); (b), (d) and (f) low excavation rate (1×10^{-6} m/s).

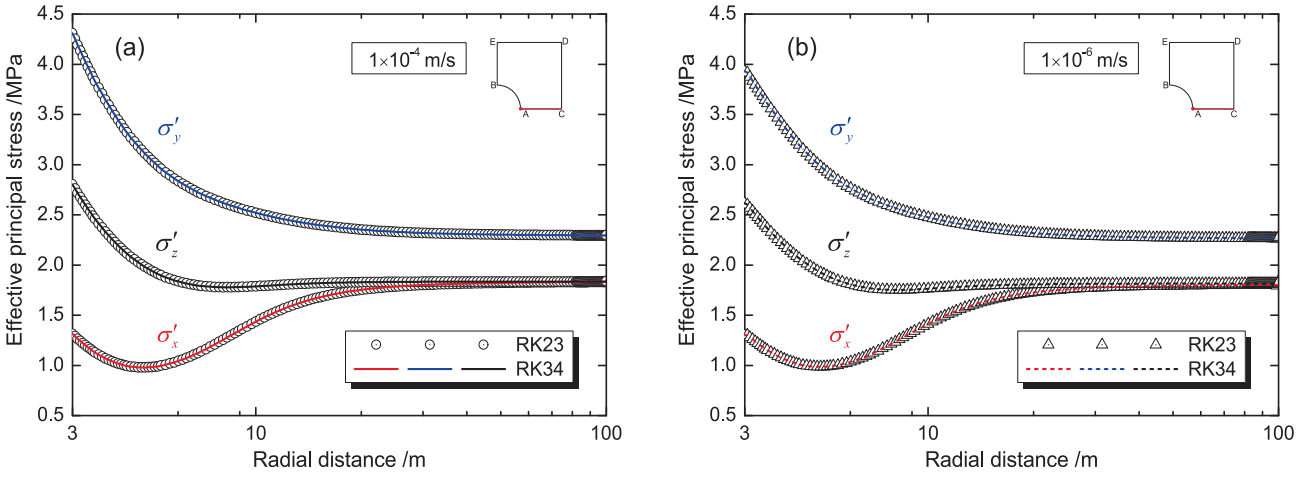

Fig. 16. Distributions of three principal effective stresses along horizontal profile after underground gallery excavation: (a) high excavation rate (1×10^{-4} m/s); (b) low excavation rate (1×10^{-6} m/s).

from the same initial stress state, the stress state traverses the critical state line, moving from the shear contraction zone into dilatancy. A higher excavation rate results in greater values of mean effective stress (p'), deviatoric stress (q), pre-consolidation pressure ($\overline{p'_{cref}}$), yield ratio (r), and the generalized viscoplastic strain rate ($\dot{\varepsilon}_m^{vp}$). In addition, the simulated results for RK23 and RK34 in Figs. 16 and 17 are fully consistent under both high and low excavation rates.

Fig. 18 shows the number of constitutive equation evaluations per step required at point A on the gallery wall during excavation at different rates, which exhibits a trend similar to that in Fig. 9. It can be observed that RK23 requires more model evaluations than RK34, especially during the early excavation stage, where this difference is more pronounced. Under the same excavation rate, the CPU time required by RK34 is approximately 2/3 of that needed by RK23. Furthermore, a higher excavation rate leads to fewer constitutive model evaluations and less CPU time consumption.

7. Conclusion

This paper presents an efficient and robust explicit adaptive numerical integration framework for rate-dependent constitutive models, systematically evaluating the performance of explicit stress integration schemes of varying orders applied to a fractional consistency two-surface viscoplastic model (FVP). Based on an explicit substepping strategy, we firstly reformulate the general stress–strain-strain rate incremental form of the FVP model, and subsequently explore the

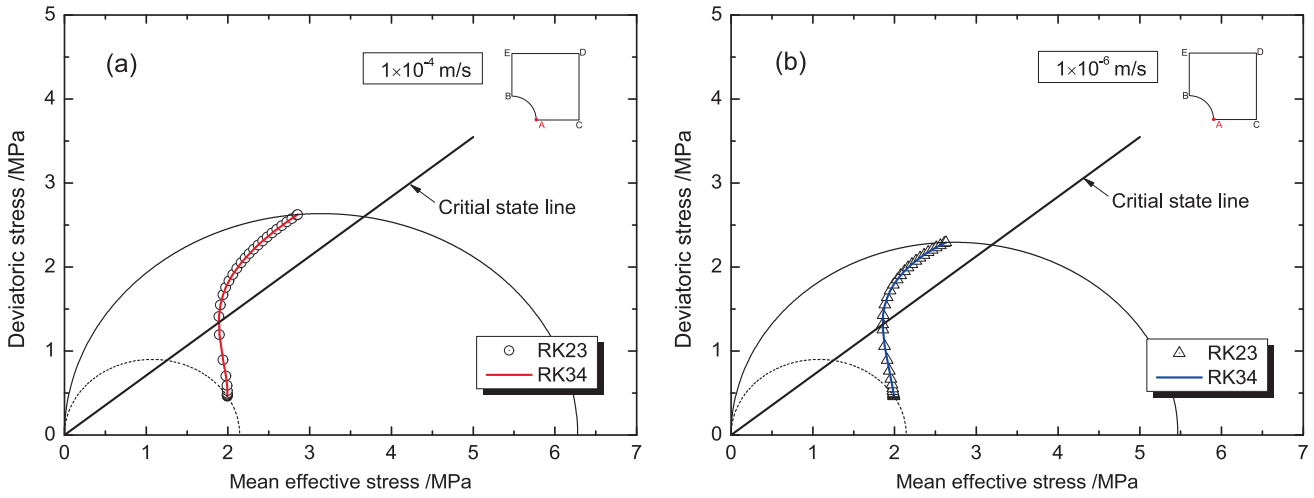

Fig. 17. Effective stress path of point A at the gallery wall on horizontal profile during underground gallery excavation: (a) high excavation rate (1×10^{-4} m/s); (b) low excavation rate (1×10^{-6} m/s).

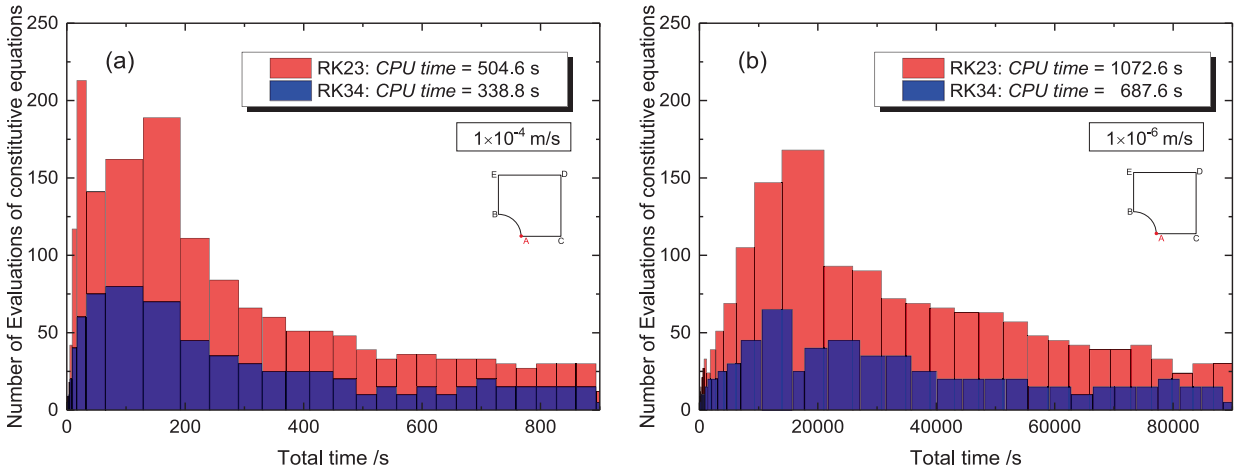

Fig. 18. Computational cost at point A for different excavation rates during underground gallery excavation: (a) high excavation rate (1×10^{-4} m/s); (b) low excavation rate (1×10^{-6} m/s).

accuracy and efficiency of different substepping methods with regard to the typical element tests and initial boundary-value problems. General conclusions can be drawn as follows:

(1) In contrast to Perzyna's typed viscoplastic models based on overstress concepts, the rate-dependent models formulated using a general consistency approach strictly adhere to the Karush-Kuhn-Tucker (*KKT*) complementarity condition, thereby providing a more direct evaluation of main phases through the explicit stress integration algorithm.

(2) As a representation of consistency viscoplastic models, the adopted fractional consistency two-surface viscoplastic model (FVP) is explicitly linearized into an incremental form of the stress–strain-strain rate relation following the consistency conditions of the loading surface in the elastic/elasto-viscoplastic steps.

(3) Four different order explicit stress integration algorithms (i.e., RK12, RK23, RK34, and RK45) are effectively integrated for the rate-dependent model integration. With adaptive step size adjustments and local error control, the accuracy and efficiency of the algorithms are ensured. From the numerical analysis (computational costs and error accumulation) of element tests (isotropic or triaxial compressions interrupted by creep/stress relaxation), RK23 and RK34 are the two recommended robust and efficient algorithms for balancing efficiency and accuracy.

(4) Two typical time-dependent scenarios, namely piezocone penetration and underground gallery excavation, effectively demonstrate the stability and flexibility of the explicit adaptive substepping algorithm in solving complex boundary-value problems using advanced consistency viscoplastic models. This approach also ensures an accurate representation of rate-dependent behaviors, making it well-suited for practical geotechnical applications requiring multi-stage analyses.

Data Availability Statement.

All data that support the findings of this study are available from the corresponding author upon reasonable request.

CRediT authorship contribution statement

Wei CHENG: Writing – original draft, Validation, Software, Methodology, Conceptualization. **Zhen-Yu YIN:** Writing – review & editing, Supervision, Conceptualization.

Declaration of competing interest

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

Acknowledgments

This research is financially supported by the Research Grants Council (RGC) of the Hong Kong Special Administrative Region Government (HKSARG) of China (Grant N_PolyuU534/20; Grant No. 15210322; Grant No. 15226322), and by the State Key Laboratory of Climate Resilience for Coastal Cities at the Hong Kong Polytechnic University.

Appendix A

For Runge-Kutta methods of different orders (RK12, RK23, RK34, and RK45), the common coefficients are given separately in the A-1, A-2, A-3, and A-4 expressions:

$$a=\begin{bmatrix}0 & 0\\ 1 & 0\end{bmatrix}, b=\begin{bmatrix}1\\ 0\end{bmatrix}, \widetilde{b}=\begin{bmatrix}1/2\\ 1/2\end{bmatrix} \tag{A-1}$$

$$a=\begin{bmatrix}0 & 0\\ 1 & 0\\ 1/4 & 1/4\end{bmatrix}, b=\begin{bmatrix}1/2\\ 1/2\\ 0\end{bmatrix}, \widetilde{b}=\begin{bmatrix}1/6\\ 1/6\\ 2/3\end{bmatrix} \tag{A-2}$$

$$a=\begin{bmatrix}0 & 0 & 0 & 0\\ 1/4 & 0 & 0 & 0\\ 4/81 & 32/81 & 0 & 0\\ 57/98 & -432/343 & 1053/686 & 0\\ 1/6 & 0 & 27/52 & 49/156\end{bmatrix}, b=\begin{bmatrix}1/6\\ 0\\ 27/52\\ 49/156\\ 0\end{bmatrix}, \widetilde{b}=\begin{bmatrix}43/288\\ 0\\ 243/416\\ 343/1872\\ 1/12\end{bmatrix} \tag{A-3}$$

$$a=\begin{bmatrix}0 & 0 & 0 & 0 & 0\\ 1/2 & 0 & 0 & 0 & 0\\ 1/4 & 1/4 & 0 & 0 & 0\\ 0 & -1 & 2 & 0 & 0\\ 7/27 & 10/27 & 0 & 1/27 & 0\\ 28/625 & -125/625 & 546/625 & 54/625 & -378/625\end{bmatrix}, b=\begin{bmatrix}1/6\\ 0\\ 4/6\\ 1/6\\ 0\\ 0\end{bmatrix}, \widetilde{b}=\begin{bmatrix}14/336\\ 0\\ 0\\ 35/336\\ 162/336\\ 125/336\end{bmatrix} \tag{A-4}$$

Data availability

Data will be made available on request.

References

Alonso, E.E., Gens, A., Josa, A., 1990. A constitutive model for partially saturated soils. Géotechnique 40 (3), 405–430.
Andrianopoulos, K.I., Papadimitriou, A.G., Bouckovalas, G.D., 2010. Explicit integration of bounding surface model for the analysis of earthquake soil liquefaction. Int. J. Numer. Anal. Meth. Geomech. 34 (15), 1586–1614.
Borja, R.I., Lee, S.R., 1990. Cam-clay plasticity, part 1: implicit integration of elasto-plastic constitutive relations. Comput. Methods Appl. Mech. Eng. 78 (1), 49–72.
Borja, R.I., Lin, C.H., Montáns, F.J., 2001. Cam-Clay plasticity, Part IV: Implicit integration of anisotropic bounding surface model with nonlinear hyperelasticity and ellipsoidal loading function. Comput. Methods Appl. Mech. Eng. 190 (26–27), 3293–3323.
Carow, C., Rackwitz, F., 2021. Comparison of implicit and explicit numerical integration schemes for a bounding surface soil model without elastic range. Comput. Geotech. 140, 104206.
Cheng, W., Yin, Z.Y., 2024. Fractional order viscoplastic modeling of anisotropically overconsolidated clays with modified isotach viscosity. Int. J. Plast 172, 103858.
Cheng, W., Chen, R.P., Pereira, J.M., Cui, Y.J., 2022. Undrained cylindrical cavity expansion/contraction in stiff clays using a two-surface plasticity model. Int. J. Numer. Anal. Meth. Geomech. 46 (3), 570–593.
Cheng, W., Yin, Z.Y., 2025. A non-isothermal fractional consistency two-surface viscoplasticity model for gas hydrate-bearing sediments. Int. J. Plast. 104547.
Cheng, W., Yin, Z.Y., Cui, Y.J., Peng, M.Z., 2024. Nonisothermal fractional order two-surface viscoplastic model for stiff clays. Can. Geotech. J.
Dafalias, Y.F., 1986. Bounding surface plasticity. I: Mathematical foundation and hypoplasticity. J. Eng. Mech. 112 (9), 966–987.
Dong, Y., 2023. Performance of explicit substepping integration scheme for complex constitutive models in finite element analysis. Comput. Geotech. 162, 105629.
Farias, M.M., Pedroso, D.M., Nakai, T., 2009. Automatic substepping integration of the subloading tij model with stress path dependent hardening. Comput. Geotech. 36 (4), 537–548.
Fehlberg, E., 1969. Low-order classical Runge-Kutta formulas with stepsize control and their application to some heat transfer problems (Vol. 315). National aeronautics and space administration.
Foster, C.D., Regueiro, R.A., Fossum, A.F., Borja, R.I., 2005. Implicit numerical integration of a three-invariant, isotropic/kinematic hardening cap plasticity model for geomaterials. Comput. Methods Appl. Mech. Eng. 194 (50–52), 5109–5138.
Hashiguchi, K., 1989. Subloading surface model in unconventional plasticity. Int. J. Solids Struct. 25 (8), 917–945.
Hong, P.Y., Pereira, J.M., Cui, Y.J., Tang, A.M., 2012. Explicit integration of a thermo-mechanical model for clays. Comput. Geotech. 46, 13–25.
Hu, C., Liu, H., 2014. Implicit and explicit integration schemes in the anisotropic bounding surface plasticity model for cyclic behaviours of saturated clay. Comput. Geotech. 55, 27–41.
Hueckel, T., Borsetto, M., 1990. Thermoplasticity of saturated soils and shales: constitutive equations. J. Geotech. Eng. 116 (12), 1765–1777.
Lester, B.T., Scherzinger, W.M., 2017. Trust-region based return mapping algorithm for implicit integration of elastic-plastic constitutive models. Int. J. Numer. Meth. Eng. 112 (3), 257–282.
Li, J., Yin, Z.Y., 2020. A modified cutting-plane time integration scheme with adaptive substepping for elasto-viscoplastic models. Int. J. Numer. Meth. Eng. 121 (17), 3955–3978.
Li, J., Yin, Z.Y., 2021. Time integration algorithms for elasto-viscoplastic models with multiple hardening laws for geomaterials: enhancement and comparative study. Arch. Comput. Meth. Eng. 28 (5), 3869–3886.
Lloret-Cabot, M., Sheng, D., 2022. Assessing the accuracy and efficiency of different order implicit and explicit integration schemes. Comput. Geotech. 141, 104531.
Lloret-Cabot, M., Sloan, S.W., Sheng, D., Abbo, A.J., 2016. Error behaviour in explicit integration algorithms with automatic substepping. Int. J. Numer. Meth. Eng. 108 (9), 1030–1053.
Lloret-Cabot, M., Wheeler, S.J., Gens, A., Sloan, S.W., 2021. Numerical integration of an elasto-plastic critical state model for soils under unsaturated conditions. Comput. Geotech. 137, 104299.
Marinelli, F., Buscarnera, G., 2019. A Generalized Backward Euler algorithm for the numerical integration of a viscous breakage model. Int. J. Numer. Anal. Meth. Geomech. 43 (1), 3–29.
Pedroso, D.M., Sheng, D., Sloan, S.W., 2008. Stress update algorithm for elastoplastic models with nonconvex yield surfaces. Int. J. Numer. Meth. Eng. 76 (13), 2029–2062.
Pérez-Foguet, A., Rodrıguez-Ferran, A., Huerta, A., 2001. Consistent tangent matrices for substepping schemes. Comput. Methods Appl. Mech. Eng. 190 (35–36), 4627–4647.
Perzyna, P., 1963. The constitutive equations for rate sensitive plastic materials. Q. Appl. Math. 20 (4), 321–332.
Petalas, A.L., Dafalias, Y.F., 2019. Implicit integration of incrementally non-linear, zero-elastic range, bounding surface plasticity. Comput. Geotech. 112, 386–402.
Potts, D.M., Gens, A., 1985. A critical assessment of methods of correcting for drift from the yield surface in elasto-plastic finite element analysis. Int. J. Numer. Anal. Meth. Geomech. 9 (2), 149–159.
Qiao, Y., Ferrari, A., Laloui, L., Ding, W., 2016. Nonstationary flow surface theory for modeling the viscoplastic behaviors of soils. Comput. Geotech. 76, 105–119.

Roscoe, K.H., Burland, J.B., 1968. On the generalized stress-strain behaviour of wet clay. In: Heymann, G., Leckie, F.A. (Eds.), Engineering Plasticity. Cambridge University Press, Cambridge, UK.
Rouainia, M., Muir Wood, D., 2001. Implicit numerical integration for a kinematic hardening soil plasticity model. Int. J. Numer. Anal. Meth. Geomech. 25 (13), 1305–1325.
Scalet, G., Auricchio, F., 2018. Computational methods for elastoplasticity: an overview of conventional and less-conventional approaches. Arch. Comput. Meth. Eng. 25, 545–589.
Shi, Z., Hambleton, J.P., Buscarnera, G., 2019. Bounding surface elasto-viscoplasticity: a general constitutive framework for rate-dependent geomaterials. J. Eng. Mech. 145 (3), 04019002.
Simo, J.C. and Hughes, T.J., 1998. Computational inelasticity. New York, NY: Springer New York.
Sloan, S., 1987. Substepping schemes for the numerical integration of elastoplastic stress–strain relations. Int. J. Numer. Meth. Eng. 24 (5), 893–911.
Sloan, S.W., Abbo, A.J., Sheng, D., 2001. Refined explicit integration of elastoplastic models with automatic error control. Eng. Comput. 18 (1/2), 121–194.
Sloan, S.W., Randolph, M.F., 1982. Numerical prediction of collapse loads using finite element methods. Int. J. Numer. Anal. Meth. Geomech. 6 (1), 47–76.
Sołowski, W.T., Gallipoli, D., 2010. Explicit stress integration with error control for the Barcelona Basic Model: Part I: Algorithms formulations. Comput. Geotech. 37 (1–2), 59–67.
Sołowski, W.T., Gallipoli, D., 2010. Explicit stress integration with error control for the Barcelona Basic Model.: Part II: Algorithms efficiency and accuracy. Comput. Geotech. 37 (1–2), 68–81.
Starman, B., Halilovič, M., Vrh, M., Štok, B., 2014. Consistent tangent operator for cutting-plane algorithm of elasto-plasticity. Comput. Methods Appl. Mech. Eng. 272, 214–232.
Su, C., Lu, D., Zhou, X., Wang, G., Zhuang, X., Du, X., 2023. An implicit stress update algorithm for the plastic nonlocal damage model of concrete. Comput. Methods Appl. Mech. Eng. 414, 116189.
Sun, Z., Chu, J., Xiao, Y., 2021. Formulation and implementation of an elastoplastic constitutive model for sand-fines mixtures. Int. J. Numer. Anal. Meth. Geomech. 45 (18), 2682–2708.
Tian, Y., Li, L., Yao, Y., Lu, D., Du, X., 2024. Approaches to improve the convergence and efficiency of the implicit stress integration algorithm for the unified hardening model. Comput. Geotech. 166, 105992.
Versino, D., Bennett, K.C., 2018. Generalized radial-return mapping algorithm for anisotropic von Mises plasticity framed in material eigenspace. Int. J. Numer. Meth. Eng. 116 (3), 202–222.
Wang, W.M., Sluys, L.J., de Borst, R., 1997. Viscoplasticity for instabilities due to strain softening and strain-rate softening. Int. J. Numer. Meth. Eng. 40 (20), 3839–3864.
Yin, Z.Y., Li, J., Jin, Y.F., Liu, F.Y., 2019. Estimation of robustness of time integration algorithms for elasto-viscoplastic modeling of soils. Int. J. Geomech. 19 (2), 04018197.
Zambrano-Cruzatty, L.E., Yerro, A., Macedo, J., 2024. Constitutive modelling of non-cohesive soils under high-strain rates: a consistency approach. Géotechnique 74 (8), 743–760.
Zhang, P., 2026. Neural network based numerical integration for elastoplastic constitutive relations. Comput. Geotech. 190, 107721.
Zhang, Y., Buscarnera, G., 2016. Implicit integration under mixed controls of a breakage model for unsaturated crushable soils. Int. J. Numer. Anal. Meth. Geomech. 40 (6), 887–918.
Zhang, Y., Zhou, A., 2016. Explicit integration of a porosity-dependent hydro-mechanical model for unsaturated soils. Int. J. Numer. Anal. Meth. Geomech. 40 (17), 2353–2382.
Zhao, C.F., Yin, Z.Y., Hicher, P.Y., 2018. Integrating a micromechanical model for multiscale analyses. Int. J. Numer. Meth. Eng. 114 (2), 105–127.
Zhao, J., Sheng, D., Rouainia, M., Sloan, S.W., 2005. Explicit stress integration of complex soil models. Int. J. Numer. Anal. Meth. Geomech. 29 (12), 1209–1229.
Zhao, L.Y., Liu, L.H., Lai, Y.M., 2023. An enhanced cutting plane algorithm of elastoplastic constitutive models for geomaterials. Comput. Geotech. 155, 105253.
Zhou, A., Zhang, Y., 2015. Explicit integration scheme for a non-isothermal elastoplastic model with convex and nonconvex subloading surfaces. Comput. Mech. 55 (5), 943–961.
Zhou, X., Lu, D., Su, C., Gao, Z., Du, X., 2022a. An unconstrained stress updating algorithm with the line search method for elastoplastic soil models. Comput. Geotech. 143, 104592.
Zhou, X., Lu, D., Zhang, Y., Du, X., Rabczuk, T., 2022b. An open-source unconstrained stress updating algorithm for the modified Cam-clay model. Comput. Methods Appl. Mech. Eng. 390, 114356.
Zhou, X., Lu, D., Zhao, J., Zhang, Y., Gao, Z., Rabczuk, T., Du, X., 2025. Material characteristic length insensitive nonlocal modelling: a computationally efficient scaled nonlocal integral method. Comput. Geotech. 188, 107587.