Review

Review of fundamental principles in modelling unsaturated soil behaviour

Daichao Sheng*
Centre for Geotechnical and Materials Modelling, School of Engineering, The University of Newcastle, NSW 2308, Australia

A R T I C L E   I N F O
Article history:
Received 5 February 2011
Received in revised form 18 April 2011
Accepted 9 May 2011

Keywords:
Unsaturated soils
Constitutive modelling
Volume change
Shear strength
Yield stress
Water retention
Hydro-mechanical coupling

A B S T R A C T
An unsaturated soil is a state of the soil. All soils can be partially saturated with water. Therefore, constitutive models for soils should ideally represent the soil behaviour over entire ranges of possible pore pressure and stress values and allow arbitrary stress and hydraulic paths within these ranges. The last two decades or so have seen significant advances in modelling unsaturated soil behaviour. This paper presents a review of constitutive models for unsaturated soils. In particular, it focuses on the fundamental principles that govern the volume change, shear strength, yield stress, water retention and hydro-mechanical coupling. Alternative forms of these principles are critically examined in terms of their predictive capacity for experimental data, the consistency between these principles and the continuity between saturated and unsaturated states.

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction
Soils that are partially saturated with water are often referred to as unsaturated soils. Some soils exhibit distinctive volume, strength and hydraulic properties when become unsaturated. For these soils, a change in the degree of saturation can cause significant changes in volume, shear strength and hydraulic properties. Nevertheless, the distinctive volume, strength and hydraulic behaviour for unsaturated states should be treated as material nonlinearity and modelled consistently and coherently. In other words, a constitutive model for a soil should represent the soil

* Tel.: +61 2 49215746; fax: +61 2 49216991.
E-mail address: daichao.sheng@newcastle.edu.au

Contents lists available at ScienceDirect
Computers and Geotechnics
journal homepage: www.elsevier.com/locate/compgeo
behaviour over entire ranges of possible pore pressure and stress values and should allow arbitrary stress and hydraulic paths within these ranges. After all, any soil can be partially saturated with water and an unsaturated soil is only a state of the soil, not a new soil, as pointed out by Gens et al. [36].

Soil mechanics principles are more established for soils at saturated states. Generalisation of these principles to unsaturated soils requires careful consideration of these fundamental issues: (1) volume change behaviour associated with suction or saturation changes, (2) shear strength behaviour associated with suction or saturation changes, and (3) hydraulic behaviour associated with suction or saturation changes. Soils can experience significant volume changes upon changes of the degree of saturation or suction. Some soils expand upon wetting, some collapse and some do both depending on the stress level. The large volume changes associated with saturation change can lead to severe damages to foundations and structures. Shear strength of soils can also change dramatically as the degree of saturation changes, and a related engineering problem is slope failures caused by rainfall. Unsaturated soils also have distinctive hydraulic behaviour which has profound implications in designing cover and containment systems for various industrial and municipal wastes. These fundamental issues are indeed the main concerns of unsaturated soil mechanics and its engineering applications [26,45].

Constitutive modelling of unsaturated soils generally involves the generalisation of constitutive models for saturated states to unsaturated states, by incorporating the fundamental issues mentioned above. Research in this direction was pioneered by Alonso et al. [2] and it has since attracted extensive interest. A large number of constitutive models can now be found in the literature. There are several state-of-the-art reviews over the last 15 years or so, e.g. Gens [32], Wheeler and Karube [148], Kohgo [58], Wheeler [146], Gens et al. [37], Sheng and Fredlund [107], Sheng et al. [110], Gens [33], Cui and Sun [19] and Gens [34]. These papers may serve as good references for studying the topic. They usually provide (1) a thorough discussion of stress state or constitutive variables used to establish various models, (2) an in-depth analysis of specific constitutive models and their advantages and disadvantages, and (3) latest developments in the area of unsaturated soil modelling. The papers by Gens [33,34] also provide interesting discussions on the physical significance of different suction components and their roles in constitutive modelling.

Instead of a comprehensive review of existing constitutive models for unsaturated soils, this paper devotes its main attention to a number of specific issues: (1) volume change behaviour, (2) variation of yield stress and shear strength with suction, (3) water retention behaviour and hydro-mechanical coupling. These issues represent the most fundamental components of constitutive models for unsaturated soils. Alternative methods for tackling these fundamental issues are scrutinised, particularly against the principle that partial saturation is a state of soil and against their predictive capacity for experimental data. The issue of stress state variables, a seemingly unavoidable topic in unsaturated soil mechanics, is not specifically discussed in this paper. In other words, constitutive models are not judged based on the stress state variables they use, rather on their qualitative predictions of observed soil behaviour. In addition, a number of advanced topics are not discussed in this paper:

Thermodynamics of unsaturated soils. This topic often leads to interesting discussion of thermodynamically consistent stress-state variables and constitutive models. Interested readers may refer to, e.g. Houlsby [44], Hutter et al. [47], Gray and Schrefler [41], Sheng et al. [114], Li [63], Samat et al. [100], Coussy et al. [16] and Zhao et al. [157].

Micromechanical modelling of unsaturated soils. This approach can often lead to insights into soil behaviour and sometimes also validation of macroscopic (continuum) constitutive equations. Some good examples of this approach include the work by Gili and Alonso [39], Jiang et al. [49], Katti et al. [52] and Scholtès et al. [104]. A closely related approach is the micro-macro double-structure models by Gens and Alonso [35], Alonso et al. [4], Sánchez et al. [101] and Cardoso and Alonso [13].

Thermo-hydro-chemo-mechanical modelling of unsaturated soils. In this approach, additional environmental variables such as temperature and chemical concentrations are introduced to study soil behaviour. Some recent work in this area refers to Loret et al. [67], Guimarães et al. [43], Cleall et al. [15], Kimoto et al. [56], Gens et al. [36,37] and Gens [34].

Miscellaneous topics such as non-isothermal behaviour, anisotropy, rate-dependent behaviour, degradation and damage, liquefaction of unsaturated soils. These topics are usually related to specific soils or specific problems. Some representative work on these topics are Cui and Delage [17], Stropeit et al. [121], Romero and Jommi [97] and D’Onza et al. [24] for anisotropy; Modaressi and Modaressi [77] and Cui et al. [18] for non-isothermal behaviour; Cardoso and Alonso [13] for degradation modelling; Arson and Gatmiri [5] and Yang et al. [151] for damage modelling; Oldecop and Alonso [85] and Pereira and de Gennaro [91] for rate-dependent behaviour; Unno et al. [138] and Bian and Shahrour [7] for liquefaction.

The paper is organised as follows. It first presents a brief discussion of two basic concepts used in constitutive modelling, i.e. the net stress and the matric suction. The fundamental issues on volume change, yield stress, shear strength, water retention and hydro-mechanical coupling are then discussed. The paper finally outlines the challenges and possible solution strategies for implementing unsaturated soil models into the finite element method.

2. Net stress and suction

Net stress is commonly used in interpreting unsaturated soil behaviour and in constitutive modelling. It is defined as

$$\sigma_{ij} = \sigma_{ij} - \delta_{ij} u_a$$

where $\sigma_{ij}$ is the net stress tensor, $\sigma_{ij}$ the total stress tensor, $\delta_{ij}$ the Kronecker delta, and $u_a$ the pore air pressure. The net stress is often used to analyse laboratory data, particularly those based on the axis-translation technique where the air pressure is not zero. It is sometimes perceived to recover the effective stress when soils become saturated. Such a perception should however be avoided. Under natural ground conditions where the air pressure is atmospheric, the net stress is equivalent to the total stress. Indeed, we never use the net stress concept to describe the behaviour of dry sand. In other words, the atmospheric pore air pressure should be considered zero. Net stress is different from total stress only when the air pressure is not atmospheric.

The concept of net stress can be useful in interpreting experimental data based on axis-translation technique, if the technique is indeed valid for applying suction (see discussions on its validity in e.g. [80,6]). In this case, the air pressure is used as a reference value for stress measures and the net stress is the total stress in excess of the pore air pressure.

The soil suction in the literature of unsaturated soil modelling usually refers to the matric suction and is usually expressed as:

$$S = u_a - u_w$$

where $S$ is called the matric suction in soil physics terminology and also called the matrix suction [6], and $u_w$ the pore water pressure. The matric suction is used interchangeably with the matric potential in soil–water potential. The latter is a measurement of energy and consists of two parts: capillary and adsorptive potentials. When
pore water exists as capillary water at relatively high degrees of saturation, the capillary potential is dominant in the matric potential and the definition by (2) is then considered to be valid. However, when pore water exists as adsorbed water films at low degrees of saturation, the adsorptive potential ($\psi_{\text{ads}}$) becomes dominant in the matric potential. Consequently, questions have been raised regarding whether Eq. (2) is still valid for the matric potential [6]. In the case when water exists as adsorbed films to solid particles, the true water pressure is not well defined. It is not unique at one material point and is dependent on the proximity to the solid particle surface [70]. However, an apparent water pressure can be introduced: $u_w = u_\text{a} - \psi_{\text{ads}}$, i.e. the apparent water pressure represents the negative adsorptive potential measured in excess of air pressure. Such an apparent water pressure is then unique at one material point. With such a definition of $u_w$, the matric potential can be expressed by Eq. (2) over the full range of saturation. Nevertheless, the matric suction should be differentiated from capillary phenomena. Its two components may not easily be separable in a soil with double-porosity. As pointed out in Gens [34], it is more appropriate to think of matric suction as a variable that expresses quantitatively the degree of attachment of water to solid particles that results from the general solid/water/interface interaction.

In constitutive modelling, the matric suction is often treated as an additional variable in a stress space for establishing constitutive laws. This approach was pioneered in the Barcelona Basic Model (BBM) by Alonso et al. [2] and has then followed in most existing models, with a few exceptions where the suction is treated as an internal variable or hardening parameter (e.g. [9,68]). Again, these approaches are not differentiated in this paper. Since the matric suction can vary independently of stress, it is treated as an independent axis in the stress space in this paper. In addition, the matric suction is considered to coincide with the negative pore water pressure for fully saturated states and thus the suction axis extends from negative infinity to positive infinity in the stress space.

3. Volume change behaviour

The volume change behaviour is one of the most fundamental properties of soils. For unsaturated soils, the large volume changes associated with suction change can cause severe damages to foundations and structures. The volume change equation also underpins the yield stress–suction and shear strength–suction relationships [110]. It is indeed the only absolutely necessary component that is needed to extend a saturated soil model to unsaturated states. The model that defines the volume change caused by stress and suction changes should again be applicable to the entire range of possible pore water pressure or suction values. The discussions below are limited to isotropic stress states. The volume change associated with changes of deviatoric or shear stress has to be considered in a three-dimensional constitutive framework, which depends on the specific model used for saturated soils, and is outside the scope of this paper.

For saturated soils, a common starting point is the linear relationship between the specific volume ($v$) and the logarithmic effective mean stress ($\ln p'$) for normally consolidated soils:

$$v = N - \lambda \ln p' = N - \lambda \ln(p - u_w)$$  \hspace{1cm} (3)

where $p$ is the mean stress, $\lambda$ is the slope of the $v - \ln p'$ line, and $N$ is the intercept on the $v$ axis when $\ln p' = 0$. Eq. (3) is only valid for positive increments of the effective stress. For unloading and reloading, the volume change depends on the specific plasticity framework adopted in the constitutive model. For example, hypoplasticity and bounding surface plasticity adopts different volume change mechanisms than classical elastoplasticity. However, for normally consolidated soils subject to positive stress increments, Eq. (3) is usually used independently of the theoretical framework.

It should be noted that Eq. (3) represents a straight line in the $v - \ln p$ space only if the pore water pressure is zero. If the pore water pressure were kept at a negative value (suction), equation (3) would predict a smooth curve in the $v - \ln p$ space, as shown by Fig. 1. The air entry suction for the soil in Fig. 1 is assumed to be larger than 100 kPa, so that the soil remains saturated. Indeed, these compression lines look very much like those for overconsolidated soils. However, the curvature of the normal compression lines is purely due to the nature of the logarithmic function and the translation from the effective stress space ($v - \ln p'$) to the total stress space ($v - \ln p$), not due to overconsolidation.

Equation (3) can also be written in an incremental form as follows:

$$d v = -\lambda \frac{d p}{p - u_w} - \lambda \frac{d(-u_w)}{p - u_w}$$  \hspace{1cm} (4)

It is clear that a negative increment in pore water pressure has exactly the same effect on the volume change of a saturated soil as an equal positive increment in mean stress.

In the literature, Eq. (3) is extended to unsaturated states in one of the three approaches:

- **Approach A**: Separate stress and suction approach, or the net stress and suction approach.
- **Approach B**: Combined stress–suction approach, or the effective stress approach.
- **Approach C**: SFG approach, which is a middle ground between Approach A and B.

These approaches are discussed separately below in terms of their advantages and disadvantages.

3.1. Separate stress and suction approach

In Approach A, the volume change due to stress change is separated from that due to suction change. A typical example of the volume change equations in this approach is:

$$v = N - \lambda_{vp} \ln p - \lambda_{us} \ln \left(\frac{s + u_w}{u_{at}}\right)$$  \hspace{1cm} (5)

where $p$ is the net mean stress, $N$ is the specific volume when $\ln p = 0$ and $s = 0$, $\lambda_{vp}$ is the slope of an assumed $v - \ln p$ line or the compressibility due to stress change, $\lambda_{us}$ the slope of an assumed $v - \ln s$ line or the shrinkability due to suction change, and...
volume change caused by suction changes is independent of stress, which is at variance with experimental observation shown in Fig. 3. In addition, the atmospheric pressure ($u_{at}$) in (5) makes the suction change insignificant when the suction is less than the atmospheric pressure ($s < u_{at}$). The first point, i.e. the discontinuity at the transition between saturated and unsaturated states, was also one of the reasons that some researchers turned to the effective stress approach (e.g. [112]). A simple numerical example will illustrate this problem. Let a soil be compressed at the transition suction ($s_{t}$) from mean stress 1 kPa to 100 kPa. Let the air pressure remain atmospheric. In the saturated zone, the volume changes according to Eq. (3):

$$\Delta v_{vs} = -\lambda_{vp} \ln \left( \frac{100 + s_{a}}{1 + s_{a}} \right)$$

In the unsaturated zone, the volume changes according to Eq. (5):

$$\Delta v_{vs} = -\lambda_{vp} \ln 100$$

These two volume changes can be quite different, depending on the value of the transition suction. The transition suction is either the air-entry or the air-expulsion value, depending on the hydraulic path.

3.2. Combined stress–suction approach

In Approach B, the matric suction and the net mean stress are combined into one single variable, i.e. an effective stress, to define their effects on soil volume. A general form of the effective stress is

$$p' = p + f(s)$$

where $f$ is either a function of suction or a function of suction and degree of saturation. Such a definition of effective stress is very general and covers most existing definitions in the literature (but perhaps not the recent one by [144]). It is noted that function $f(s)$ usually involves material state variables such as the degree of saturation and the air entry value. As a consequence, the definition of the effective stress and hence the stress space can change with material states, a feature which is not shared by its counterpart for saturated soils.

With such an effective stress, Eq. (3) is usually assumed to be still valid for unsaturated states:

$$v = N - \lambda \ln p' = N - \lambda(s) \ln(p + f(s))$$

(7)

where $N$ is the specific volume when $\ln p' = 0$. If the effective stress is indeed effective in controlling soil volume, $\lambda$ should remain constant under constant $p'$. As such, parameters $N$ and $\lambda$ should be independent of suction. However, this is seldom the case in most effective stress models. In the literature, $\lambda$ is usually assumed to be a function of $s$, while $N$ is treated either as a constant or a variable. We first discuss the case where $N$ is a constant and then show that a varying $N$ with suction cannot be recommended.

Eq. (7) is widely used in the literature and in fact most effective stress models adopt it as the volume change equation (e.g. [9,12,50,57,59,68,82,103,112,114,123,125]). Eq. (7) generally recovers Eq. (3) when the soil becomes saturated. This is one of the greatest advantages of using the effective stress.

However, there are also some disadvantages with Eq. (7). The obvious issue is the difficulty in addressing the different compressibilities due to stress and suction changes, as shown in Fig. 2, since there is now only one compressibility in the volume change equation. The second issue is related to a constraint on the compressibility $\lambda$. Let a saturated slurry soil be dried from zero suction to a historically high value (above the suction-increase yield surface, see section below on yield stress).

$u_{at}$ is the atmospheric pressure and is added to avoid the singularity when $s = 0$. Again, Eq. (5) is only used for increasing mean stress or increasing suction. Indeed, $\lambda_{vs}$ is usually replaced by the elastic compression index $k_{vp}$ in most applications, unless the suction increases to a historically high value (above the suction-increase yield surface, see section below on yield stress).

Eq. (5) has been used in many models such as Alonso et al. [2], Wheeler and Sivakumar [150], Cui and Delage [17], Chiu & Ng [14], Georgiadis et al. [38] and Thu et al. [134]. The main advantage of Eq. (5) is that the compressibility due to stress and suction changes are dealt with separately. This does not only provide extra flexibility for modelling soil behaviour, but is also supported by experimental data. Toll [136] and Toll and Ong [137] showed that the two compressibilities $\lambda_{vp}$ and $\lambda_{vs}$ can be totally different (Fig. 2). It is usually true that the suction shrinks to a degree of saturation. On the other hand, the stress compressibility ($\lambda_{vp}$) can increase with increasing suction, particularly for compacted soils where highly compressible macropores (inter-aggregate pores) are present [96,30].

However, there are also a few shortcomings about Eq. (5). First, Eq. (5) does not recover Eq. (3) when the soil becomes saturated. It represents a linear $\nu - \ln p$ relationship for constant suctions, unless $\lambda_{vp}$ is assumed to be a function of stress (as in [38]. This is not consistent with the saturated soil model (Fig. 1). As a consequence, the volume change becomes undefined at the transition suction between saturated and unsaturated states. Second, the

![Fig. 2. Variation of compressibilities with degree of saturation (after [136]).](image)

![Fig. 3. Variation of shrinkability with stress (after [23]).](image)
stress (Fig. 9), this drying path is elastoplastic, not purely elastic. The volume of the soil then changes according to:

\[ v_h = N - \lambda(s) \ln(1 + f(s)) \]  

(8)

Now compress the soil under constant suction, i.e. stress path BC in Fig. 4. The compression line will be curved in the \( v - \ln p \) space, due to the \( f(s) \) term. If the suction at point B is above the air entry value, this compression line is expected to intersect with the initial compression line for saturated states. Let the intersection be point C (at net mean stress of \( p_b \)). The volume at C is then:

\[ v_C = v_h - \lambda(s) \ln \left( \frac{p_b + f(s)}{1 + f(s)} \right) \]  

(9)

Along path AC, the volume changes according to:

\[ v_C = N - \lambda(0) \ln(p_b + f(0)) = N - \lambda(0) \ln p_b \]  

(10)

Replacing (8) into (9) and then equating (9) with (10) lead to:

\[ \frac{\lambda(s)}{\lambda(0)} = \frac{\ln(p_b)}{\ln(p_b + f(s))} < 1 \]  

(11)

We usually anticipate the effective stress to increase with increasing suction, at least at low suction values. Therefore, we have: \( \lambda(s) < \lambda(0) \), meaning that the slope of the compression line decreases with increasing suction. Such a constraint on \( \lambda \) is however not supported by experimental data. In Alonso et al. [2], the slope of the compression lines decreases with increasing suction. In the data by Toll [136], Sivakumar and Wheeler [117], Toll and Ong [137] and Gallipoli et al. [30] for compacted soils, the slope of the compression lines increases with increasing suction. In addition, data on wetting-induced collapse (e.g. [126]) do not support an ever increasing collapse volume with increasing mean stress.

In some constitutive models that use the combined stress–suction approach, parameter \( N \) is assumed to vary with suction. If \( N \) decreases with increasing suction, the same constraint on \( \lambda \), i.e. Eq. (11) would apply. To avoid this constraint, \( N \) has to increase with increasing suction. However, the variation of \( N \) with suction has implications on the yield surface evolution. An increasing \( N \) with suction basically implies that drying a slurry soil under constant effective mean stress will cause the yield surface to expand, which is not consistent with the stress path. This inconsistency will be discussed in association with Fig. 9 below.

Gallipoli et al. [30] proposed the following volume change equation:

\[ v = (N - \lambda \ln p')(1 - a(1 - \exp(b \xi))) \]  

(12)

where \( N \) and \( \lambda \) are the two parameters of the normal compression line for saturated states, \( a \) and \( b \) fitting parameters, and \( \xi \) a positive variable representing the bonding effects of suction. The bonding variable \( \xi \) is a function of both \( s \) and \( S_r \). Eq. (12) is hence equivalent to Eq. (7) with \( N \) and \( \lambda \) being functions of \( s \) and \( S_r \). Gallipoli et al. [29,30] showed that Eq. (12) is able to predict the volume change at both normal compression and critical states for a variety of compacted soils. One challenge in using Eq. (12) is that the yield stress is likely to be functions of both suction and degree of saturation. Because the \( s-S_r \) relationship is usually not unique due to hydraulic hysteresis, the resulting loading-collapse surface may not be well defined. However, Gallipoli (personal communication) suggests that it is sufficient to define the bonding variable \( \xi \) in terms of \( S_r \) only (e.g. \( \xi = 1 - S_r \)), which would then resolve the non-uniqueness problem and lead to a unique loading-collapse surface in the \( S_r-p' \) space.

To avoid the constraint defined by Eq. (11), a possible augmentation to Eq. (7) is to assume that the compressibility \( \lambda \) is a function of degree of saturation, i.e. \( \lambda(S_r) \), while keeping \( N \) constant. This is similar to the approach by Gallipoli et al. [30] with \( \xi \) being a function of \( S_r \) only. Because \( S_r \) changes with soil volume even if the suction is kept constant, the slope of the compression line will then change and will most likely increase with increasing mean stress. Therefore, it is possible to have \( \lambda(S_r) \) decreasing with decreasing \( S_r \), but the slope of the compression line for constant suction increases with increasing stress (Fig. 5b). Al-Badran and Schanz [1] used an approach similar to Fig. 5b, but formulated their volume change equation in the net stress–suction space. Kikumoto et al. [55] and Zhang and Ikariya [152] assumed a saturation-dependent compressibility (\( \lambda(S_r) \)), but formulated their models in the effective stress–suction space. The same issue of non-uniqueness of the yield surface in the stress–suction space will arise, due to hydraulic hysteresis. To avoid this non-uniqueness problem, the yield surface and hence the shear strength can be defined in the \( S_r-p' \) space, thus eliminating suction as the additional variable of the stress space. Some work in this direction has recently been reported by Zhou et al. [159].

Clearly further research is required if the combined stress–suction approach is adopted for the volume change, particularly in terms of consistent explanation of the suction-caused and stress-caused volume changes for reconstituted soils. A worthwhile endeavour in this direction is perhaps to explore the possibilities...
of using \( s_a \) in Eq. (7) and the \( S_r - p' \) space to establish all constitutive equations.

3.3. SFG approach

Sheng et al. [108] proposed a third way to model the volume change for unsaturated soils under isotropic stress states. The new model, referred to as the SFG model, represents a middle ground between Approach A and Approach B and is expressed in an incremental form as follows:

\[
dv = -\lambda_{vp} \frac{dp}{p + f(s)} - \lambda_{vs} \frac{ds}{p + f(s)}
\]  

(13)

Eq. (13) is in the same form as Eq. (4). Similar to Approach A, Eq. (13) is defined in terms of net stress and suction and separates the compressibilities due to the two variables, i.e. \( \lambda_{vp} \) and \( \lambda_{vs} \). Similar to Approach B, it combines the suction with the net mean stress in the denominator, i.e. the term \( p + f(s) \), and recovers Eq. (4) for saturated states. The term \( p + f(s) \) represents the interaction between stress and suction and makes the normal compression lines for non-zero suction curved in the \( V - \ln p \) space. However, there is no constraint on parameter \( \lambda_{vp} \). As a first approximation, \( \lambda_{vp} \) can be assumed to be independent of suction, as indicated by the data of Jennings and Burland [48] for air-dry soils. More realistically it should depend on suction. For example, the data of Sivakumar and Wheeler [117] shows that \( \lambda_{vp} \) increases with increasing suction for compacted soils. Parameter \( \lambda_{vs} \) must equal \( \lambda_{vp} \) when the soil is fully saturated, because of Eq. (4). It generally decreases with increasing suction and approaches zero. Sheng et al. [108] suggested the following simple function for \( \lambda_{vs} \):

\[
\lambda_{vs} = \begin{cases} 
\lambda_{vp}, & s \leq s_{sa} \\
\lambda_{vp} \frac{s}{s_{sa}}, & s > s_{sa} 
\end{cases}
\]  

(14)

where \( s_{sa} \) is the transition suction and was also called the saturation suction in Sheng et al. [108]. It is the unique transition suction between saturated and unsaturated states in the SFG model.

We note that the number ’1’ in Eq. (14) is used to avoid the singularity when \( s_{sa} = 0 \) and is not truly needed if \( s_{sa} \) is not absolutely zero. A better expression would be:

\[
\lambda_{vs} = \begin{cases} 
\lambda_{vp}, & s \leq s_{sa} \\
\lambda_{vp} \frac{s}{s_{sa}}, & s > s_{sa} 
\end{cases}
\]  

(15)

The difference between Eqs. (14) and (15) is minimal, but Eq. (15) is preferred. Eq. (15) can be applied as long as the transition suction is not absolutely zero. Kurucuk et al. [61] used Eq. (13), but a different function for \( \lambda_{vs} \) than (15). Again, both \( \lambda_{vp} \) and \( \lambda_{vs} \) can vary with stress path and take different values on a loading and unloading path respectively.

The function \( f(s) \) in Eq. (13) can also take different forms. Sheng et al. [108] initially used the following function:

\[
f(s) = s
\]  

(16)

This is perhaps the simplest form possible for \( f(s) \) and yet guarantees the continuity between saturated and unsaturated states. Even with this simplest form, Zhou and Sheng [156] showed that the SFG model is able to predict a good set of experimental data on volume change and shear strength, both for soils reconstituted from slurry and for compacted soils. Due to the \( f(s) \) term in Eq. (13), the normal compression lines will be curved in the \( V - \ln p \) space (Fig. 6). Fig. 6a shows the compression curves for a reconstituted soil under various suction pressures. The compression curves for \( s = 400, 650, 1000 \) kPa are all normal compression lines that do not involve any unloading or reloading. The SFG predictions for these curves were obtained with one single \( \lambda_{vp} \) value. Fig. 6a also shows the reconstituted soil becomes stiffer as the suction increases. Fig. 6b shows the compression curves for a compacted soil. The difference in the estimated yield stresses by Eqs. (5) and (13) is clearly shown in the figure for \( s = 300 \) kPa. The yield stress is indicated by the meeting points of the unloading-reloading line and the normal compression lines. The parameter \( \lambda_{vp} \) was allowed to change with suction in Fig. 6b.

One shortcoming of Eq. (16) is that the soil compressibility approaches zero as suction increases to infinite. There is also a theoretical discontinuity between unsaturated states and completely dry state (\( S_r = 0 \)). To avoid these problems, an alternative form of \( f(s) \) could be used, for example:

\[
f(s) = S_r s
\]  

(17)

This equation will not only guarantee the continuity between saturated and unsaturated states, but also the continuity between unsaturated and the completely dry state (\( S_r = 0 \)). More interestingly, both Eqs. (16) and (17) will lead to the same shear strength–suction relationship. The degree of saturation in (17) can also be replaced by the effective degree of saturation (\( S' \)), as suggested by Pereira and Alonso [90] when discussing Bishop’s \( \chi \) parameter. However, the performance of Eq. (17) is yet to be
validated against experimental data, and the loading-collapse yield surface may become non-unique due to hydraulic hysteresis.

The SFG approach seems to be able to overcome some disadvantages of Approach A and Approach B. The main disadvantage of the SFG approach is that it exists only in an incremental form and its integration depends on stress path [154,109]. The stress path dependency requires special treatment in the stress integration of the constitutive model [111].

3.4 Stress-path dependent elastic behaviour

An interesting issue related to the volume change is that all the approaches discussed above are stress-path dependent. Zhang and Lytton [154] and Sheng et al. [109] showed that Approaches A and C lead to stress-path dependent elastic behaviour. Approach B can also result in stress-path dependent elastic behaviour, because a closed loop of net stress and suction changes do not necessarily lead to a closed loop of effective stress changes (Fig. 7). Such behaviour is caused by the material state dependency of the effective stress. For example, the change of the effective mean stress along path AB in Fig. 7 is usually different from that along path CD, because the degree of saturation (or the air entry value) has changed from A to C, even though the net mean stress is kept constant during both paths. Such an unclosed loop means that the model is stress-path dependent, even if the stress changes occur inside the elastic zone. A model that exhibits stress-path dependent elastic behaviour is at variance with classical elastoplasticity and thermodynamics, and hence should generally be avoided. However, it is still a challenge to develop a model that exhibits a stress-path independent elastic behaviour.

4. Yield stress versus suction

4.1 Relationship between yield stress and volume change equation

In the literature of unsaturated soil mechanics, the yield stress of an unsaturated soil is usually assumed to be a function of suction. The concept of yield stress in classical elastoplasticity theory refers to the stress level that causes plastic deformation. In other plasticity frameworks such as bounding surface plasticity [21,99,78], hypoplasticity [60,72] and generalised plasticity [69,86,101], plastic deformation occurs along all loading paths including reloading. In these cases, a loading function or a bounding surface is used to differentiate unloading from loading. In the discussion below, the yield surface concept is based on the classical elastoplasticity theory, but can be extended to the loading or bounding surface concepts.

Under isotropic stress states, the yield stress is also called the preconsolidation stress. For unsaturated soils, the yield net mean stress, denoted here by $p_c$, is conventionally determined from isotropic compression curves obtained at constant suctions. These compression curves are usually plotted in the space of void ratio versus logarithmic net mean stress. The initial portion of the curve is typically flatter than the ending portion of the curve, if the suction is larger than zero. Such a curve is then approximated by two straight lines, one representing the elastic unloading-reloading line and the other the elastoplastic normal compression line. The intersection point of the two straight lines is used to define the preconsolidation (yield) stress (Fig. 8a). The yield stress so determined is generally found to increase with increasing suction, irrespective of samples prepared from slurry states or from compacted soils, leading to the so-called loading-collapse yield surface that is widely used in constitutive models for unsaturated soils (Fig. 8b). The procedure outlined above for determining the yield stress is based on the assumption that the $e - \ln p$ relationship for normally consolidated soils at $s > 0$ is linear and may not be consistent with the definition of yield stress. To demonstrate this inconsistency, we should first realise that the isotropic compression curves shown in Fig. 8 are typical of unsaturated soils reconstituted from slurry (e.g.
as well as of compacted soils. Because it is easier to understand the preconsolidation stress for a slurry soil that for a compacted soil, we use a slurry soil as an example here. Let us assume that the slurry soil has not been consolidated (with a zero preconsolidation stress). The initial yield stress for the soil is then zero (Point A in Fig. 9a). We also note that the effective stress for saturated states is constant along the 135° line in the $p$–$s$ space.

Drying the slurry soil to suction $B$ under zero mean stress is similar to consolidating the soil to stress $E$ under zero suction (Fig. 9a). Indeed, if the air entry value of the soil is larger than suction at $B$, the length $AB$ is exactly the same as $AE$, due to the effective stress principle for saturated soils. However if the air entry value is lower than suction at $B$, the length $AB$ should generally be larger than $AE$, because a suction increment is generally less effective than

---

**Fig. 8.** Isotropic compression curves under constant suction ($s$) and derived yield stresses.

**Fig. 9.** Evolution of the elastic zone during drying and compression of a slurry soil (A → B → C: constant net mean stress).
an equal stress increment in terms of consolidating the unsaturated soil. Once the soil becomes unsaturated, the yield stress does not necessarily change with suction along the 135° line (denoted by the dashed line). However, the new yield surface always go through the current stress point, as shown in Fig. 9a. The elastic zone expands as the suction increases from A to B. The stress points (e.g. B and C) are on the current yield surfaces. Let now the soil be isotropically compressed under the constant suction at Point C (stress path CD in Fig. 9a). The isotropic compression path (CD) is again outside the initial elastic zone and the soil at point C is normally consolidated. Therefore, the isotropic compression path (CD) as well as the drying path from B to C is elastoplastic and does not involve a purely elastic portion as Fig. 8 indicates, suggesting that the method for determining the yield stress in Fig. 8 be conceptually inconsistent. Indeed, the suction-induced apparent consolidation effect should refer to the increase of the preconsolidation stress at zero suction ($p_c(0)$) moves from E to F as suction increases from B to C in Fig. 9a), not the preconsolidation stress at the current suction. We also note that as the slurry soil is dried under constant stress, the soil becomes stiffer. The slope of the compression line at point C is much smaller than that at point A in Fig. 9c.

The same analysis can be done in the effective stress–suction space (Fig. 9b). In the $p' – s$ space, the zero shear strength line or the apparent tensile strength surface is commonly assumed to be: $p' = 0$, i.e. the vertical line that goes through the origin of the space. This line must also be the initial yield surface for a soil that has not been consolidated, i.e. $p_c(0) = 0$. For a saturated soil that has a zero yield stress, i.e. a slurry, the size of the elastic zone remains zero as long as the effective mean stress remains zero, irrespective of the pore water pressure values. When the soil becomes unsaturated, the yield surface will continue along the zero shear strength line, if the effective mean stress remains zero. To keep the effective mean stress zero, a tensile net mean stress has to be applied to balance out the suction increase. As a consequence, the size of the elastic zone remains zero, which also reflects the effective stress principle. Furthermore, if the yield surface does not collapse to the zero shear strength line when $p_c(0) = 0$, there would be plastic deformation for loading along the yield surface, and the shear strength as well as the yield stress of the soil would become non-uniquely defined.

In the effective stress–suction space (Fig. 9b), the stress path for suction increase under constant net mean stress is initially inclined to horizontal by 45° for saturated states. Therefore, the stress path will cross the current yield surface and the drying path AB for a slurry soil is elastoplastic, not purely elastic. Once the soil becomes unsaturated, the stress path will drift away from the 45° line and the yield surface will also drift away from the vertical line (the dashed lines in Fig. 9b). Nevertheless, the current stress points A, B, C, and D stay on the current yield surfaces and the stress path ABCD causes elastoplastic volume change, in consistency with Fig. 9a.

Fig. 9b can also be used to illustrate that parameter $N$ in the volume change equation, i.e. Eq. (7), cannot be a function of suction. This is because increasing suction under constant effective mean stress, i.e. stress path AH in Fig. 9b, does not lead to the expansion of the elastic zone. To obtain stress path AH in Fig. 9b, a tensile net mean stress has to be applied as suction increases, so that the effective mean stress remains zero. Such a stress path corresponds to the so-called neutral loading and does not lead to yield surface expansion. The expansion of the elastic zone is a prerequisite for shifting the normal compression lines towards higher specific volume in $v – \ln p'$ space (Fig. 5a). Therefore, parameter $N$ in Eq. (7) cannot be a function of suction, if the zero shear strength line is assumed to be $p' = 0$. The corresponding volume change along path ABCD can still be explained consistently if the soil compression index ($\lambda$) is assumed to be a function of degree of saturation, as shown in Fig. 9d. Points A and H are then identical in the $v – \ln p'$ space.

In the literature, the yield stress variation with suction is a rather confusing part. Many early models adopt three yield surfaces in the net stress–suction space, namely the loading-collapse yield surface, the suction-increase yield surface and the apparent tensile strength surface. Fig. 10 shows some examples of these models. The loading-collapse yield surface is used to model the volume collapse when an unsaturated soil is first loaded under constant suction and then wetted under constant stress. The suction-increase yield surface is used to capture the plastic volume change when an unsaturated soil is dried to a historically high suction. The apparent tensile strength surface defines the zero shear strength or the apparent tensile strength due to suction increase. These yield surfaces are usually defined separately. For example, setting the preconsolidation stress ($p_{co}$) to zero in the loading-collapse yield function does not recover the apparent tensile strength function. The suction-increase yield surface is usually horizontal or gently sloped (Fig. 10) and is not related to the loading-collapse surface or the apparent tensile strength surface.

Sheng et al. [108] showed that the loading-collapse surface, the apparent tensile strength surface and the suction-increase yield surface are related to each other. In the SFG model, the yield stress–suction relationship, the apparent tensile strength–suction relationship and the shear strength–suction relationship are all derived from the volume change equation, i.e. Eq. (13). In this model, the yield stress for a slurry soil that has never been consolidated or dried varies with suction in a unique function. This function also defines the apparent tensile strength surface or the zero shear strength surface in the stress–suction space (the curve through point A in Fig. 11). The curve approaches the 45° line as the suction becomes zero or negative (positive pore water pressure). Drying this slurry soil under zero stress (stress path ABC) causes the expansion of the yield surface to point C in Fig. 10. Therefore, the suction-increase yielding is already included in the yield function and there is no need to define a separate function. If the unsaturated soil at point C is then compressed under constant suction (stress path CD in Fig. 11), the yield surface will evolve to the loading-collapse surface that passes through point D in Fig. 11. The yield surface in the stress space represents the contours of the hardening parameter, which is usually the plastic volumetric strain. The stress path CD will change the initial shape of the yield surface, because the plastic volumetric strain along CD depends on the suction level. The loading-collapse yield function recovers the apparent tensile strength function when the preconsolidation
stress at zero suction ($p_{00}$) is set to zero. All these yield surfaces are continuous and smooth in the stress–suction space. Models based on effective stresses also have these features of the SFG model, as shown in Fig. 9b. As pointed out by Wheeler and Karube [148] and shown by Sheng et al. [108] and Zhang and Lytton [156], the apparent tensile strength function, the suction-increase yield function and the loading-collapse yield function are all related to the volumetric model that defines the elastic and elastoplastic volume changes caused by stress and suction changes. If Approach A, i.e. Eq. (5), is adopted to describe the volume change, the loading-collapse yield surface will take the following form:

$$p_c = \begin{cases} p_{00} - s, & s \leq s_{sa} \\ \frac{p_{00} - s_{sa}}{\kappa} \ln \frac{s}{s_{sa}}, & s > s_{sa} \end{cases}$$

(18)

where $p_{00}$ is the yield stress at zero suction, $p_c$ a reference stress, and $\kappa$ the elastic compression index. The specific shape of this function depends on the variation of the compressibility with suction, i.e. the function $\lambda_{ap}(s)$. The suction-increase yield surface can also be derived from Eq. (5). If the shrinkability ($\lambda_{sa}$) is assumed to be independent of stress, the suction-increase yield surface is simply:

$$s = s_0$$

(19)

where $s_0$ is the yield suction. Eq. (19) represents a horizontal line in the stress–suction space. Because Eq. (5) is not defined at zero suction and zero mean stress, the apparent tensile strength surface cannot be derived from (18). A separate function is usually introduced:

$$\tilde{p}_0 = -ks$$

(20)

where $k$ was assumed to be constant in Alonso et al. [2], but vary with suction in Georgiadis et al. [38].

If Approach B, i.e. Eq. (7), is adopted to describe the volume change, the loading-collapse surface can then be written as:

$$p_c = \begin{cases} p_{00} - s, & s \leq s_{sa} \\ \frac{p_{00} - s_{sa}}{\kappa} \ln \frac{s}{s_{sa}}, & s > s_{sa} \end{cases}$$

(21)

The suction-increase surface is the same as Eq. (19). The apparent tensile strength surface is recovered from (21) by setting $p_{00} = 0$:

$$p_{00} = 0$$

(22)

which represents the vertical line going through the origin of the stress space.

If the SFG model, i.e. Eqs. (13) and (15), is adopted to describe the volume change, the yield stress takes the following form:

$$p_c = \begin{cases} p_{00} - s, & s \leq s_{sa} \\ p_{00} - s_{sa} \ln \frac{s}{s_{sa}}, & s > s_{sa} \end{cases}$$

(23)

The function defines the suction-increase surface and is independent of the specific function $f(s)$ used in (13). The apparent tensile strength function is also defined in (23) by setting $p_c = 0$. The loading-collapse surface takes a very similar form:

$$p_c = \begin{cases} p_{00} - s, & s \leq s_{sa} \\ \frac{\lambda_{sa}}{p_c} (p_{00} + f(s) - s_{sa} \ln \frac{s}{s_{sa}}) - f(s), & s > s_{sa} \end{cases}$$

(24)

where $p_{00}$ is the new yield stress at zero suction (Fig. 11). This function is however dependent on the specific forms of $f(s)$ used in (13).

The functions defined in (18)–(24) are all continuous over the entire ranges of possible suction or pore pressure values. However, functions (18) and (21) may not be smooth, dependent on functions $\lambda_{ap}(s)$ and $\lambda_{sa}(s)$, respectively. On the other hand, functions (23) and (24) are continuous and smooth. All these functions can be incorporated into existing constitutive models for saturated soils. For example, if the modified Cam clay model is used for saturated soil behaviour, the yield function can be generalised to unsaturated states along the suction axis:

$$f = \frac{q^2}{M^2} (\tilde{p} - \tilde{p}_0)(\tilde{p}_c - \tilde{p}) = 0$$

(25)

where $f$ is the yield function in the stress space, $q$ is the deviator stress, $M$ is the slope of the critical state line in $q$–$p$ space, and $\tilde{p}_0$ and $\tilde{p}_c$ are defined above. Again, Eq. (25) is valid for all pore pressure and suction values.

4.2. Reconstituted soils versus compacted soils

A soil can become partially saturated with water in different ways. Two types of unsaturated soil samples are often used in laboratory: (1) dry soil powders are statically or dynamically compacted at specified water contents, (2) samples reconstituted from slurry are dried to unsaturated states. The first type of samples (compacted soils) is far more common than the second type of samples (reconstituted soils), because it is more difficult to desaturate a slurry sample than a compacted sample. It has been noted that most constitutive models for unsaturated soils are based on experimental data for compacted soils [108]. Compacted soil samples can be prepared at an initial water content smaller than the optimum water content (compacted wet of optimum) or larger than the optimum water content (compacted dry of optimum). Reconstituted samples can be air-dried, heat-dried, freeze-dried or osmosis-dried. Different samples preparation methods usually result in different soil microstructures. For example, soils compacted dry of optimum usually have a double-porosity microstructure, meaning that the pore size distribution curve exhibits two or more peaks. In these soils, there are two types of pores: large inter-aggregates pores which are collapsible upon wetting and small intra-aggregates pore which are more stable. On the other hand, soils air-dried from slurry usually exhibit a uni-modal pore size distribution, at least at low stresses. Nevertheless, as recently pointed out by Tarantino [133], the boundary between compacted and reconstituted soils is not always clear and the microstructure of a soil can change with stress and hydraulic paths.

Unsaturated soils with a bi-modal pore size distribution (PSD) are usually collapsible. Wetting such a soil can collapse the inter-aggregates pores and results in a uni-modal pore size distribution when the soil becomes full saturated. In terms of constitutive modelling, a collapsible soil is characterised by a loading-collapse yield
The question is now about the reconstituted soils. A soil reconstituted from slurry is characterised by a uni-modal PSD, wetting it under constant stress usually does not cause volume collapse. However, can such a soil become collapsible if it is air-dried and compressed to high stresses? According to the SFG model [108], the yield surface for a slurry soil can evolve into a loading-collapse surface if the soil is dried and compressed to sufficiently high stresses (point C in Fig. 13), which means that a uni-modal reconstituted soil can evolve into a bi-modal collapsible soil (stress path B → A → C). Is there experimental evidence for such an evolution? In other words, can drying and compressing of a reconstituted soil lead to aggregation or cluster formation so that the soil becomes collapsible? Unfortunately there is very few data on reconstituted soils in the literature. Nevertheless, the classic reference by Jennings and Burland [48] seems to support such an evolution. Fig. 14 is a replot from Jennings and Burland [48] for an air-dry silt.
It is clear that the air-dry soil can become collapsible if it is compressed to sufficiently high stresses. Another set of data reported by Cunningham et al. [20] for a reconstituted silty clay also seem to indicate that compressing a soil at sufficiently high suction can eventually lead to a collapsible soil (Fig. 15). The limited data in the literature seem to support the evolution of LC surfaces as suggested in the SFG model (Fig. 13). However, experimental data on reconstituted soils are generally too few to be conclusive.

In summary, the microstructure and particularly the pore size distribution of a soil is usually reflected in the yield surface and volume change equation of the constitutive model for the soil. A bi-modal pore size distribution can evolve to a uni-modal pore size distribution under proper stress and hydraulic paths, and vice versa.

5. Shear strength

The change of shear strength with suction or saturation is one of the main reasons behind rainfall-induced landslides. It is related to the volume change equation [110]. However, this relationship has been overlooked in most existing models for unsaturated soils. If the slope of the critical state line is assumed to be independent of suction, as supported by experimental data of Ng and Chiu [79], Thu et al. [135] and Nuth and Laloui [83], the shear strength–suction relationship can be derived uniquely from the volume change equation. If the slope of the critical state line depends on suction, as supported by data of Toll [136], Toll and Ong [137] and Merchán et al. [74], two equations are needed to define the shear strength–suction relationship, namely the volume change equation, and the $M(s)$ function, with $M$ being the slope of the critical state line in the deviator – mean stress space.

Bishop and Blight [8] first proposed an effective stress definition to interpret the shear strength of unsaturated soils:

$$
\tau = c' + \sigma_n \tan \phi' = c' + (\sigma_n + \chi s) \tan \phi' = \sigma_n \tan \phi' + c = \sigma_n \tan \phi' + \chi
$$

where $\tau$ is the shear strength, $c'$ is the effective cohesion for saturated states and is usually assumed to be zero, $\sigma_n$ and $\sigma_s$ are respectively the effective and net normal stress on the failure plane, $\phi'$ is the effective friction angle of the soil, $\chi$ is the well-known Bishop's effective stress parameter, and $c'$ is the apparent cohesion which includes the friction term due to suction, i.e. $c' = c' + \chi s \tan \phi'$.

Fredlund et al. [25] proposed the following relationship which conveniently separates the shear strength due to stress from that due to suction:

$$
\tau = \left[ c' + (\sigma_n - u_s) \tan \phi' \right] + \left[ (u_s - u_w) \tan \omega \right] = c' + (\sigma_n - u_s) \tan \phi' + \left[ (u_s - u_w) \tan \omega \right]
$$

(27)

where $\tau$ is the shear strength, $c'$ the effective cohesion for saturated states and is usually assumed to be zero, $\sigma_n$ the normal stress on the failure plane, $\phi'$ the effective friction angle of the soil, and $\omega$ the frictional angle due to suction. Obviously, if $\omega$ is set to $\phi'$ in Eq. (27), the Coulomb friction criterion in terms of effective stress for saturated soils is recovered.

A common conception is that the shear strength of an unsaturated soil can be sufficiently defined by a single effective stress [3,105,146]. However, this conception is only true when the slope of the critical state line or the friction angle of the soil does not change with suction. When the friction angle of the soil changes with suction, the shear strength can no longer be defined by a single effective stress. This is clear by comparing equation (26) with (27):

$$
\tau = \left[ c' + (\sigma_n + s) \tan \phi'(s) \right] + \left[ s \tan \omega \right] = c' + (\sigma_n + s) \tan \phi'(s) + \left[ s \tan \omega \right]
$$

(28)

where the Bishop effective stress parameter ($\chi$) is set to $\tan \omega / \tan \phi'$. It is clear that the variable ($s$) cannot be eliminated no matter how the effective stress is defined.

The shear strength of an unsaturated soil is fully defined if the apparent cohesion ($c'$) or the friction angle due to suction ($\omega$), and the friction angle due to stress $\phi'$ in Eq. (27) are known. A number of alternative equations have been used in the literature to define $c'$ or $\omega$ (e.g. [27,54,75,84,137,140]). Most of these equations are empirically based and are defined independently from the volume change equation, which, if incorporated into a constitutive model, may lead to inconsistencies with the yield surface as discussed in the Section above. Pires and Alonso [90] where $\chi$ is expressed in terms of the effective degree of saturation that depends on the micro- and macro-saturation.

Sheng et al. [115] recently compared various shear strength equations for unsaturated soils against a large number of data sets. The equations include those empirically based and those embedded in constitutive models. The equations studied by Sheng et al. [115] are listed in Table 1. Fig. 16 presents one example of such comparisons. The parameters used in the shear strength equations

<table>
<thead>
<tr>
<th>Equations</th>
<th>$\tan \phi'$</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Öberg and Sällfors [84]</td>
<td>$\gamma$</td>
<td>SWCC</td>
</tr>
<tr>
<td>2. Fredlund et al. [27]</td>
<td>$(c' + \kappa s)$</td>
<td>$\kappa = 1.3$, $s = 0.659$, SWCC</td>
</tr>
<tr>
<td>3. Vanapalli et al. [140]</td>
<td>$(\sigma_n + s) \tan \omega$</td>
<td>$s = 0.659$, SWCC</td>
</tr>
<tr>
<td>4. Toll and Ong [137]</td>
<td>$(\sigma_n + s) \tan \omega$</td>
<td>$k = 1.2$, SWCC</td>
</tr>
<tr>
<td>5. Alonso et al. [2]</td>
<td>$\chi$</td>
<td>$\kappa = 0.4$</td>
</tr>
<tr>
<td>6. Sun et al. [122]</td>
<td>$\mu$</td>
<td>$\mu = 110$ kPa</td>
</tr>
<tr>
<td>7. Khalili and Khakhaz [54]</td>
<td>$(\sigma_n + s) \tan \omega$</td>
<td>$s = 60$ kPa, $r = 0.55$</td>
</tr>
<tr>
<td>8. Sheng et al. [108]</td>
<td>$(\sigma_n + s) \tan \omega$</td>
<td>$s = 25$ kPa</td>
</tr>
</tbody>
</table>
are given in Table 1. The comparative study by Sheng et al. [115] reveals that:

1. If the friction angle of the soil is assumed to be independent of suction, all shear strength equations in the literature can be formulated in form of either Eq. (26) or Eq. (27). In this case, there is little difference in formulating shear strength in one single stress variable or in two independent stress variables. The real challenge is to find a single effective stress when the friction angle depends on suction.

2. The performance of the shear strength equations in predicting experimental data depends on the careful determination of material parameters and also on the specific data set. A shear strength equation may predict one data set better than other data sets.

3. The shear strength equations that incorporate the soil–water characteristic curve (SWCC) generally require more parameters. This group of equations seem to provide reasonable prediction of shear strength for unsaturated soils (e.g. [27,140,137]). However, some equations are sensitive to the residual suction [140,137], which can be difficult to determine accurately from the SWCC.

4. Simpler shear strength equations that are embedded in constitutive models appear to provide reasonable predictions of shear strength [54,108,122]. These equations contain typically one or two parameters. However, these equations usually do not predict any peak value of the shear strength attained at an intermediate suction level.

6. Water retention behaviour and hydro-mechanical coupling

The issue of interaction between the mechanical and hydraulic behaviour was perhaps first raised by Wheeler [145] and then by Dangla et al. [22]. The first complete model that deals with coupled hydro-mechanical behaviour of unsaturated soils was perhaps due to Vaunat et al. [143]. A number of coupled models soon followed (e.g. [114,149]). With respect to hydraulic behaviour of unsaturated soils, many models [28,139] take advantage of the fact that the influence of suction on degree of saturation is more significant than the influence of deformation. The dependency of degree of saturation on suction is described by a soil–water characteristic curve (also called soil–water retention curve, SWRC). Only until recently, the effects of deformation on SWCCs have been considered (e.g. [31,76,124,149,160]).

As pointed out by Wheeler et al. [149], the mechanical behaviour of an unsaturated soil depends on degree of saturation even if the suction, net stress and specific volume are kept the same for the soil. Separate treatment of mechanical and hydraulic components in modelling unsaturated soil behaviour has certain limitations in reproducing experimental observation. It would be difficult to consider the saturation dependency in a mechanical model that is independent of the hydraulic behaviour. Similarly, a hydraulic model that is independent of mechanical behaviour cannot easily take into account the effects of soil density on the SWCC. Experimental data generally demonstrate the following points:

1. A SWCC obtained under a higher net mean stress tends to shift towards the higher suction [31,62,73,81]. This means that the incremental relationship between degree of saturation ($S_i$) and suction ($s$) depends on net mean stress ($p$) or soil density.

2. When the suction is kept constant, isotropic loading or unloading can also change the degree of saturation of an unsaturated soil [149]. This implies that the degree of saturation is related to stress or soil density when the suction is kept constant.

One of the early models that fully couple the hydraulic and mechanical components of unsaturated soil behaviour is that by Wheeler et al. [149]. Models that appeared before or soon after Wheeler et al. [149] tend to accentuate the influences of the hydraulic component on the mechanical component, not vice versa (e.g. [83,114,143]). The interaction between the mechanical and hydraulic components in the model by Wheeler et al. [149] was realised through the use of the average soil skeleton stress (effective stress), the modified suction and the coupling between the loading-collapse and suction-increase and suction-decrease surfaces. The average soil skeleton stress ($\sigma_s$) is an amalgam of stress, suction and degree of saturation. The modified suction ($s^*$) is a combination of suction and porosity. Therefore, the influence of hydraulic behaviour on the stress–strain relationship is considered via the definition of the average stress. The influence of porosity on the saturation–suction relationship is considered via the definition of the modified suction. The model by Wheeler et al. [149] is one of the few models that are qualitatively tenable in terms of coupling mechanical behaviour with hydraulic behaviour for unsaturated
soils. However, the use of the modified suction and the soil skeleton stress, which is one of the advantages that makes the model rigorously consistent in thermodynamics, can become a disadvantage as well, particularly in terms of quantitative prediction and the application of the model. For example, one of the difficulties in using this model is to quantify the synchronised movement between the loading-collapse (LC) surface and the suction-increase (SI) and suction-decrease (SD) surfaces. This synchronicity cannot easily be calibrated by laboratory experiments [93] or defined theoretically.

In more recent models, the influences of mechanical properties on the hydraulic behaviour are usually modelled via the dependency of the SWCC on soil volume [31,132], soil density [71,124], or volumetric strain [82]. Gallipoli et al. [31] suggested including a function of specific volume (v) in the SWCC equation of van Genuchten [139]. Tarantino [132] showed there is a unique relationship between the water ratio (product of $S_\text{r}$ and $e$) and the matrix suction and used this relationship to modify van Genuchten’s equation. The modified van Genuchten’s equation takes a similar form as Gallipoli et al. [31]. It is also common to express the SWCC equation in incremental forms. For example, Sun et al. [124] proposed a hydraulic model in the following form:

$$dS_r = \lambda_d e - \lambda_x ds/s$$  \hspace{1cm} (29)

where $\lambda_d$ is the slope of main drying or wetting curve, and $\lambda_x$ is the slope of degree of saturation versus void ratio curve under constant suction. Nuth and Laloui [82] used a similar equation as (29). In his model both air entry value ($S_{ae}$) and the slope of main drying curve ($\lambda_{dp}$) vary with void ratio. Nuth and Laloui [82] provided an alternative approach of modelling SWCC for a deforming soil. They assume there is an intrinsic SWCC for a specific suction and used this relationship to modify van Genuchten’s equation. The modified van Genuchten’s equation takes a similar form as Gallipoli et al. [31]. This synchronicity cannot easily be calibrated by laboratory experiments [93] or defined theoretically.

The models by Sun et al. [124], Nuth and Laloui [82], Mašín [71] and many others (e.g. [53,83,108,114,128,160]) essentially all adopt a water retention equation in the following form:

$$dS_r = (\ldots)ds + (\ldots)de$$  \hspace{1cm} (30)

This equation is not wrong, but the embedded $S_r$–$s$ relationship is for constant volume ($de = 0$). Therefore, it does not recover the conventional SWCC equations, which are obtained under constant stress. The volume change along a conventional SWCC can be significant, particularly at low suctions. For expansive soils, it is also common to study the water retention behaviour under constant volume [65,96,98]. On the other hand, if one specific $S_r$–$s$ relation is valid for constant stress, the $S_r$–$s$ relation for constant volume would inevitably involve soil compressibility and hence stress (suction) level and stress (suction) history. Therefore, it is unlikely that the same $S_r$–$s$ relation can be used to calibrate water retention data both for constant stress and constant volume. Furthermore, neglecting the volume change along SWCC can lead to inconsistent prediction of the degree of saturation during undrained compression, an issue raised by Zhang and Lytton [154].

Sheng and Zhou [116] proposed a new method for coupling hydraulic with mechanical behaviour. This new method is based on the fact that SWCCs are obtained under constant stress. In Sheng and Zhou [116], the volume change behaviour and the water retention behaviour under isotropic stress states are represented by the following incremental equations, respectively:

$$dS_r = Adp + Bds$$  \hspace{1cm} (31)

$$dS_r = Eds + \frac{S}{n}(1 - S)^2Adp$$

$$= \left( E - B\frac{S}{n}(1 - S)^2 \right) ds - \frac{S}{e}(1 - S)^2 de$$  \hspace{1cm} (32)

where parameters $A$ and $B$ are related to the specific volume change equation as discussed in Section 3, parameter $E$ is the gradient of the conventional SWCC, $e$ is the void ratio, $n$ is the porosity, and $\zeta$ is a fitting parameter. If the SFG model is used to describe the volume change, parameters $A$ and $B$ would then take the following form respectively:

$$A = \frac{\lambda_{dp}}{p + f(s)} \quad B = \frac{\lambda_x}{p + f(s)}$$  \hspace{1cm} (33)

The $S_r$–$s$ relationship in Eq. (32) is defined for constant stress ($dp = 0$) and hence parameter $E$ refers to the gradient of the conventional SWCC:

$$E = -\frac{dS_{SWCC}(s)}{ds}$$  \hspace{1cm} (34)

where $S_{SWCC}$ represents the conventional SWCC equation. The void ratio in Eq. (32) refers to the value at the current stress and suction.

It is clear from Eq. (32) that the $S_r$–$s$ relationship for constant volume ($de = 0$) is more complex than the conventional SWCC equation ($dS_r = Eds$). Eq. (32) was proposed based on experimental observation as well as the intrinsic phase relationship for undrained condition:

$$dS_r = \frac{S}{n}de, \quad dw = 0$$  \hspace{1cm} (35)

![Fig. 17. Qualitative analysis of isotropic compression under undrained condition.](image-url)
where \( w \) is the gravimetric water content. Eq. (35) actually imposes a constraint on suction change under undrained compression. This constraint is obtained by substituting Eq. (35) into Eq. (32):

\[
(S_r - S_f (1 - S_f)^e) \, d\psi = (nE - B) \, ds, \quad dw = 0
\]

(36)

Zhang and Lytton [154] recently noted that some models fail short in predicting undrained behaviour of unsaturated soils. The specific issue raised is illustrated in Fig. 17. Assume the initial state of a soil is inside the elastic zone, but on the main wetting curve, i.e. point A in Fig. 17a. Compressing the soil under undrained condition will lead to some suction decrease [127,131], say to point B. Assume B is still inside the elastic zone. Unloading from B to A will recover the initial volume of the soil and hence the initial degree of saturation should be recovered as well. However, if the volume change along the SWCC is neglected, the change of \( S_r \) along path ACB would follow the main wetting curve and is hence ‘elastoplastic’, whereas the change of \( S_r \) along path BDA would follow the scanning curve and is hence ‘elastic’, leading to inconsistent change of \( S_r \) (Fig. 17b). It would then seem unlikely that a model where the irreversible volume change is not synchronised with the irreversible saturation change could lead to a consistent prediction of saturation change over the closed path ABA [155]. This inconsistency is actually due to the assumption that the main wetting curve is defined for constant volume.

If Eq. (32) is used, the inconsistence in Fig. 17b will be avoided. Because of Eq. (35), the model will predict no change of \( S_r \) as long as \( \psi_0 = 0 \), irrespective of evolution of the suction-decrease surface. Eq. (35) is satisfied as long as the suction changes according to (36). It is easy to understand that the loading path ACB is not on the initial main wetting curve and the unloading path BDA is not on the scanning curve, because the mean stress is not constant along those stress paths. The main wetting curve at point A also shifts to that at point B, as the mean stress changes (Fig. 17c). The synchronised evolution of the LC, SI and SD surfaces (as in [149]) is not necessary for the consistent prediction in Fig. 17c. Indeed, the suction path can be ‘elastoplastic’ even though the stress path is elastic.

The following equation can be derived from Eq. (32):

\[
\frac{\partial S_r}{\partial e} = -\frac{S_r (1 - S_f)^e}{e} \, ds = 0
\]

(37)

The void ratio \( (e) \) in Eq. (37) refers to the initial void ratio at the current stress and its variation is purely due to stress change. It can also be interpreted as the initial void ratio at the start of the SWCC tests. Eq. (37) shows that the SWCC for a soil shifts with its initial void ratio. This is similar to the approach by Gallipoli et al. [31] where the van Genuchten equation was modified to include the initial void ratio. Indeed, their SWCC equation can be re-written as:

\[
\frac{\partial S_r}{\partial e} = -mn \, \psi \frac{S_r (1 - S_f)^{e/m}}{e}
\]

(38)

where \( m \) and \( n \) are two fitting parameters in the original van Genuchten equation, and \( \psi \) is another parameter introduced by Gallipoli et al. [31]. If the product \( mn \psi \) was set to 1, Eq. (38) would be equivalent to (37). Sheng and Zhou [116] showed that the intrinsic phase relationship requires:

\[
1 - \frac{S_r}{e} \geq \frac{\partial S_r}{\partial e} \geq -\frac{S_r}{e}
\]

(39)

The above constraint is satisfied if \( mn \psi = 1 \) in Eq. (38). It is also interesting to note that a value of 1.1 for \( mn \psi \) was used in all the numerical examples in Gallipoli et al. [31].

Eq. (37) can be integrated either analytically for certain \( \zeta \) values or numerically in more general cases. Because Eq. (37) is in an incremental form, integration of the equation requires one specific SWCC equation that corresponds to a reference initial void ratio. In other words, the conventional SWCC equation is only used for the reference initial void ratio and the new SWCC for a new initial void ratio is obtained by integration of (37).

The model by Sheng and Zhou [116] is validated against a variety of data sets. In Fig. 18, the results from suction-controlled oedometer compression tests by Jotisankasa [51] are used to validate Eq. (37). Since the suction remains constant for each curve in Fig. 18, no SWCC equation is required here and the only parameter required is \( \zeta \). Eq. (37) can directly be used to predict the variation of degree of saturation due the variation of stress-induced volume change. In this case, the void ratio \( (e) \) in Eq. (37) is the current void ratio and it changes due to stress variation in the oedometer tests. Fig. 18 clearly shows that Eq. (37) predicts very well the relationship between the degree of saturation and the void ratio when the suction is kept constant. Indeed, the experimental data are almost exactly on the predicted \( S_r–e \) curves. In Fig. 19, the data by Vanapalli et al. [141] on a compacted till were used. The SWCC for \( \psi_0 = 0.517 \) was fitted by the van Genuchten equation, while the other SWCCs are predicted by Eq. (37) with \( \zeta = 0.03 \). Fig. 19 shows that both the slope and the air entry value of the SWCCs change with the initial void ratio. In the two cases studied, the model by Sheng and Zhou [116] seems to be able to capture the effect of initial void ratio on the soil water retention behaviour. One
advantage of this approach is that Eq. (37) can be used with any existing SWCC equations, including those for uni-modal pore size distribution (e.g. [28,139]) and those for bi-modal pore size distribution (e.g. [11,40]), as well as those advanced models that adopt hysteretic scanning loops within the main drying-wetting loop [64,87–89].

7. Finite element implementation

One of the ultimate goals of constitutive modelling is to implement the model in a numerical method to solve boundary value problems. A constitutive model can generally be formulated in the following incremental form [66,110,114]:

\[
\frac{d\sigma}{dt} = \left( \begin{array}{c} D^p \\ W^p \\ T \\ H \end{array} \right) \left( \begin{array}{c} d\varepsilon \\ ds \end{array} \right) 
\]

where \( \sigma \) is the stress vector, \( \varepsilon \) is the vector of soil skeleton strain, \( \theta \) is the volumetric water content, \( D^p \), \( W^p \), \( T \) and \( H \) are constitutive matrices [110].

In the displacement finite element method, the nodal displacements and pore pressures are first solved from the equilibrium and continuity equations, based on the current stress states and the current volumetric water content. The strain and suction increments are then derived from the displacements and pore pressures. For given strain and suction increments, the current stress vector, the volumetric water content and the internal variables must be updated according to Eq. (40). This update is generally carried out by numerical integration.

One of the main challenges in integrating Eq. (40) arises from the non-convexity of the yield surface around the transition between saturated and unsaturated states [106,113,147]. The non-convexity seems to persist irrespective of the stress variables and is demonstrated in Fig. 20 [111].

Both implicit and explicit schemes have been used to integrate unsaturated soils models. In implicit schemes, all gradients and functions are estimated at advanced unknown stress states and the solution is achieved by iteration. This group of stress integration schemes include Vannut al. [142], Borja [10], Hoyos and Arduino [46], Zhang and Zhou [153], and Tamagnini and De Gennaro [130]. They usually do not deal with the non-convex problem by avoiding the transition between saturated and unsaturated states. Borja [10] was perhaps the only one in this group who noticed the problem. He suggested keeping the step size sufficiently small to avoid overshooting the plastic zone by the elastic trial stress (Fig. 20). Otherwise, there is currently no general method used to tackle the non-convexity problem in implicit schemes.

On the other hand, explicit schemes estimate the gradients and functions at the current known stress states and proceed in an incremental fashion. These methods are theoretically more appropriate for non-convex models [87,88]. González and Gens [42] compared both an implicit and an explicit scheme for integrating the BBM and found that the latter is more robust and efficient. However, explicit schemes usually require to determine the intersection between the current yield surface and the elastic trial stress path and some substepping methods to control the integration error [102,110,111,112,113,119,120].

A key issue in integrating the incremental stress–strain relationships using an explicit scheme is thus to find the intersection between the elastic trial stress path and the current yield surface. The most complex situation occurs when the yield surface is crossed more than once, such as shown in Fig. 20. However, it is not possible to know a priori how many times the yield surface is crossed. Therefore, for non-convex yield surfaces, the key task is to find the very first intersection for any possible stress and hydraulic path.

Finding the intersection between the elastic trial stress increment and the current yield surface can be cast as a problem of finding the multiple roots of a nonlinear equation:

\[
f(\alpha) = f(\sigma, s, z) = 0 
\]

where \( 0 \leq \alpha \leq 1 \). \( f(\sigma, s, z) \) is the yield function, \( z \) is a set of internal variables, and subscript \( \alpha \) indicates the quantity is estimated at strain increment \( \alpha \Delta s \) and suction increment \( \alpha \Delta z \). Pedroso et al. [87,88] proposed a novel method to bracket the roots (\( \alpha \)). The method is illustrated in Fig. 21. For a given increment (\( \alpha = 1 \)), the number of roots of \( f(\alpha) \) is first computed. If there is more than one root, the increment is divided into two equal sub-increments. The number of roots of each sub-increment is then computed. If the first sub-increment contains more than one root, it is further divided into two sub-increments. This process is repeated until the first sub-increment contains at most one root (Fig. 21). Once the roots are bracketed, the solution of the first root can be found by using numerical methods such as the Pegasus method [118].

Sheng et al. [111] applied the root bracketing method by Pedroso et al. [87,88] to integrate the SFG model and found that the method can indeed provide an accurate solution of the intersection

---

Fig. 20. Non-convexity of yield surfaces in unsaturated soil models in the suction–stress space.

Fig. 21. Bracketing the roots for nonlinear function according to Pedroso et al. [87,88].
problem. However, this method is found to be computationally expensive. The formula used to find the number of roots of a non-linear function requires both the first and second orders of gradients of the function and requires numerical integration. Furthermore, the root-finding procedure must be applied for all suction increments near the non-convexity. Further research in this direction is clearly required.

8. Conclusions

Some conclusions can be drawn from this review of constitutive modelling of unsaturated soils:

1. Partial saturation is only a state of soil. Constitutive models for soils should represent the soil behaviour over entire ranges of possible stress and pore pressure values. This requires a consistent and coherent merge of fundamental soil mechanics principles for both saturated and unsaturated states.

2. The volume change behaviour is one of the most fundamental properties of soils. For soils in unsaturated states, the volume change equation also underpins the yield stress–suction and shear strength–suction relationships. It also affects the soil water retention behaviour. Indeed, one of the most essential differences amongst various existing models is the volume change equation. Other differences are often the consequences of the volume change equation.

3. Three groups of volumetric models are compared and it is shown that each has advantages and disadvantages. There is certainly room for improvement of all these models. One observation here is that it seems difficult to describe the volume change of unsaturated soils in terms of a single stress variable.

4. The methods used to define the loading-collapse, suction-increase and apparent tensile strength surfaces should be consistent with the volume change equation. In addition, these surfaces and the shear strength function are all related to each other and hence should be defined consistently with each other.

5. It seems possible to model reconstituted soils and compacted soils within the same theoretical framework. The pore size distribution evolves with mechanical and hydraulic loading and is reflected by the loading collapse yield surface and the volume change equation.

6. If the friction angle of the soil is assumed to be independent of suction, all shear strength equations in the literature can be formulated either in terms of one single stress variable or in terms of two independent stress variables. The real challenge is to find a single effective stress when the friction angle depends on suction.

7. The performance of the shear strength equations in predicting experimental data depends on the careful determination of material parameters and also on the specific data set. A shear strength equation may predict one data set better than others.

8. When coupling the hydraulic component with the mechanical component in a constitutive model, it is recommended to take into account the volume change along soil–water characteristic curves. Neglecting this volume change can lead to inconsistent predictions of volume and saturation changes.

9. Unsaturated soil models are characterised by non-convex yield surfaces at the transition between saturated and unsaturated states. This non-convexity, if handled rigorously, can significantly complicate the implementation of these models into finite element codes. In this case, an explicit stress integration scheme incorporating an efficient root finding algorithm is preferred.

Acknowledgements

The author would like thank the editor of Computers and Geotechnics for the invitation of this review paper. This paper is based on the General Report presented at the 5th International Conference on Unsaturated Soils held at Barcelona during 6–8 September 2010 [105]. A number of people provided valuable comments and contributions to the General Report: Y.J. Cui, D.G. Fredlund, D. Gallipoli, A. Gens, S.L. Houston, J. Kodikara, J.M. Pereira, W.T. Solowsky, D.A. Sun, C. Yang, X. Zhang and A.N. Zhou. The author is grateful to all of them. The General Report has since been substantially revised to produce this paper, partly inspired by questions and informal discussions during and after the Barcelona conference. Finally the financial support from the Australian Research Council is greatly appreciated.

References


M原文缺失


