# The Two Stress State Variable Link from Stress Path to Stress State Surface to Elastoplastic Modeling of Unsaturated Soil

Sandra Houston*

School of Sustainable Engineering and The Built Environment, Arizona State University, Tempe, AZ, USA

**Corresponding author's email:* sandra.houston@asu.edu

**Abstract:** Particularly after publication of Soil Mechanics for Unsaturated Soils by Fredlund and Rahardjo [1], which largely followed the seminal contributions of Matyas and Radhakrishna [2] and Fredlund and Morgenstern [3], a formal approach to unsaturated soil mechanics began to take root. Elastoplastic modeling of unsaturated soils evolved, beginning with the still-today most-adopted elastoplastic constitutive model of the Barcelona Basic Model (BBM) [4]. Zhang and Lytton [5,6,7] contributed an elastoplastic framework, the Modified State Surface Approach (MSSA), which established a clear link between the State Surface Approach of Fredlund and Rahardjo and the elastoplastic BBM of Alonso, Gens, and Josa. An MSSA view can be used to compare and contrast various approaches to modeling of unsaturated soil, and as an aid to judge the suitability and limitations of modeling approaches, as well as simplified stress path-based laboratory methods commonly used in routine practice today, e.g., ASTMD4546 [8]. A two stress variables approach is key for judging the appropriateness and limitations of simple to complex methods, whether the problem is one of unsaturated soils alone, or whether transition from unsaturated to saturated is a part of the problem.

## Introduction

In application, engineers have successfully faced the challenges of building on and in unsaturated soils throughout history. Nonetheless, application of theoretical principles to unsaturated soils has been slow to make its way into mainstream geotechnical engineering. Introduction of effective stress by Terzaghi [9] led to a deeper understanding of the mechanics of saturated soils compared to unsaturated soils, resulting in soil mechanics emerging largely as a discipline based on saturated soil mechanics theory. The role of total stress and pore water pressure in control of the behavior of saturated soils became understood primarily through use of effective stress principles. Unfortunately, in the early development stages of soil mechanics measurement of negative pore water pressure were challenging, resulting in unsaturated soils being treated under various "simplifying" assumptions of pore water pressure, such as a commonly-adopted "conservative" assumption that pore water pressures above the groundwater table are zero - a condition wherein total stresses and effective stresses are assumed equal. But pore water pressures are always negative, never zero, above the groundwater table and the assumption that it is always conservative to assume zero pore water pressures for unsaturated soils holds some risks.

Nonetheless, total stress approaches for unsaturated soil conditions have prevailed in engineering practice for decades, and total stress methods for unsaturated soil problems remain common today. This does not mean that geotechnical engineers failed to understand that total stress approaches held limitations in understanding the behavior of unsaturated soils, and for this reason stress-path appropriate testing in determination of unsaturated soil total stress properties are common. Afterall, if the stress path in the laboratory is the same as that in the field, one can assume (hope) that whatever changes occur in total stress and negative pore water pressure that the same changes will be the same.  An alternate "conservative" and common approach to dealing with unsaturated soils has been to assume that any unsaturated soil may become saturated at some time under field conditions, requiring testing of the soil in a saturated state for determination of effective stress properties for direct use in effective stress analyses, typically under assumed zero pore water pressure conditions. Of course, the reality of soil moisture state (suction values) is often intermediate between the in-situ (at time of sampling and testing) and fully wetted (saturated) conditions. Because unsaturated soil response is largely controlled by changes in soil moisture (due to the link between water content and negative pore water pressure for unsaturated soil conditions), it is not a trivial task to determine appropriate field conditions/paths for application of the stress path method to unsaturated soils. Thus, use of total stress approaches, including assumed zero pore water pressure approaches, have considerable limitations and varying consequences on design and performance of infrastructure [10].

In the 1950s and 1960s, geotechnical engineers began measuring and controlling negative pore water pressures (soil suction), primarily using axis translation methods [11,12].  It was the breakthrough in control of soil suction that led to early developments in unsaturated soil mechanics wherein it was recognized that unsaturated soil behavior was controlled by two independent stress variables of net stress ($\sigma$-$u_a$) and suction ($u_a$-$u_w$), or total stress and negative pore water pressure for atmospheric air pressure conditions.  Early approaches to modeling unsaturated soil volume change response were based on state surface approaches (e.g., Matyas and Radhakrishna [1]; Fredlund and Morgenstern[3]) under a simplifying assumption of nonlinear elasticity and incremental elasticity, required due to complexities of unsaturated soil response to changes in stress state variables.  Subsequent developments in critical state modeling of unsaturated soils (e.g., Alonso, et al. [4]) have also required simplifications and assumptions due to the complex elastoplastic and highly nonlinear response of unsaturated soils to changes in the stress state variables.

Herein, various simplified methods of unsaturated soil analyses are explored and compared. Through better understanding of how simplified approaches  fit into the "big picture" of unsaturated soil behavior, advantages and limitations, as well as similarities and differences, between methods, can be identified. A look at the big picture of unsaturated soil response requires use of a two independent stress variable approach which is facilitated using the integrated elastoplastic framework on the Modified State Surface approach (MSSA) of Zhang and Lytton [5, 6, 7].

**Two Independent Stress Variable Approaches to Unsaturated Soil Theory**

There were some challenges to development of an understanding of stress variables controlling unsaturated soil behavior compared to saturated soil behavior, largely due to challenges in measurement of negative pore water pressures. Hence, development of a more general continuum mechanics-based soil mechanics theory, applicable across the spectrum of unsaturated to saturated conditions, has been slow to advance. The study of unsaturated soil mechanics started in earnest along with adoption of methods for measurement and control of soil suction in the 1940s to 1950s. The axis-translation method, adopted from the soil science community, became the go-to method for control of matric soil suction in laboratory experiments [11,12,13]. The axis translation method entails increasing the pore air pressure, $u_a$, so that the pore water pressure, $u_w$, can be increased to avoid cavitation of water in the pore water pressure control system. The mean total stress, $\sigma_m$, is increased by the same amount as $u_a$ to maintain mean net stress ($\sigma_m$-$u_a$ = $p$) constant, and matric suction ($u_a$-$u_w$ = $s$) is controlled or measured. In the 1960s and 1970s, the role of suction and net stress in unsaturated soil behavior was deeply explored [14,15,16,17,2, 18, 3]. Due in part to the heavy reliance on axis translation in measurement of negative pore water pressure, the controlling stress variables that emerged in development of unsaturated soil mechanics theory were net stress ($\sigma$-$u_a$) and suction ($u_a$-$u_w$). Although not frequently discussed, it is worth noting that when $u_a$ is atmospheric that the controlling stress variables for unsaturated soils are the same as those for saturated soils – total stress and pore water pressure.

Although net stress and suction pressure were demonstrated, separately, to control the shear strength and volume change behavior of unsaturated soils, there were also early efforts to identify single-valued “effective stress” expressions (analogous to effective stress for saturated soils), incorporating both net total stress and soil suction [19, 20, 21]. A single-valued effective stress was demonstrated by experimental studies to be inadequate for describing unsaturated soil behavior [22,23,24]. Ultimately, failure of single-valued effective stress efforts for unsaturated soils led to general acceptance of a two independent stress state variable approach proposed by several unsaturated soil researchers such as Matyas and Radhakrishna [2], Fredlund and Morgenstern[3], Fredlund and Rahardjo [1], and Alonso et al. [4].

Initially, the failure of effective stress for unsaturated soils was demonstrated on the basis of soil volume change, most notably for soil collapse response to soil wetting under load. Recently Zhang and Houston [25] used an elastoplastic approach to evaluate the various categories of unsaturated soil effective stress equations that have been put forth in the literature: (1) volume change-based; (2) shear strength-based; (3) yield-based; or (4) degree of saturation-based.

Regardless the basis of definition, it was demonstrated that there is no Terzaghi-equivalent effective stress equation for unsaturated soils. Volume change, the first-explored basis long-recognized as problematic in finding any unsaturated soil effective stress, represents the most challenging aspect of unsaturated soil response and is the focus herein.

Based on the two independent stress variables, two seemingly different paths in the development of unsaturated soil mechanics theory have been taken for volume change: (1) State Surface [1, 2, 3] and (2) Critical State [4]. In addition, Stress-Path Approaches, being experimentally based and therefore of a two independent stress state variable nature, continue to hold a rightful position in unsaturated soils geotechnical applications. Stress path methods, where field-appropriate stress levels and stress paths are taken in performance of laboratory tests, are commonly used in practice for expansive and/or collapsible soil volume change estimation [8].

It is the author's opinion that, viewed within the appropriate context of complex elastoplastic response of soil behavior, variously-simplified approaches to unsaturated soil modeling can be shown to be appropriate and theoretically sound, and consistent with demonstrated behavior, provided the independent role of total stress and negative pore water pressure is properly recognized and taken into account. Across methods based on independent stress state variables of net stress and suction, there appear to be more similarities than differences to volume change estimation for some commonly encountered field stress paths. It is quite possible to address unsaturated soil engineering applications in a sound manner, with varying levels of simplifications. Understanding of the role of the two independent stress state variables is the key element to obtaining consistency across the varyingly simplified approaches, as in understanding the big picture of the complex unsaturated soil response to changes in stress variables. The Modified State Surface framework of Zhang and Lytton [5, 6] provides the integrated elastoplastic big picture required for judging reasonableness of differing methods (hierarchies) of unsaturated soil volume change analyses. The MSSA is recommended as an excellent thinking tool for exploration of appropriateness of a given constitutive model for the problem at-hand [26] and is used here to explore the link between stress path, stress state surface, and elastoplastic unsaturated soil approaches.

## Two Separate Stress State Variables as a Key Link Between Approaches to Unsaturated Soil Modeling

### *Background*

It is through consideration of the theory of unsaturated soil mechanics based on the two independent stress variables of net stress (total stress) and suction (negative pore water pressure) that comparisons of various methods for engineering of unsaturated soils can be properly made. Such comparisons require first a big picture view of unsaturated soil behavior. Three aspects of unsaturated soil behavior are critical to seeing the big picture: (1) recognition of the role of elastoplastic response in understanding the behavior of unsaturated soils, (2) separation of stress variables $\sigma$ and $u_w$ (i.e., adoption of the two independent stress variables), (3) recognition of the non-linear behavior of unsaturated soils in response to changes in total stress and the highly non-linear response of unsaturated soils to changes in negative pore water pressure (suction). It was Zhang and Lytton [5, 6, 7] who introduced the Modified State Surface Approach (MSSA) that represented the breakthrough thinking required to connect the dots between the various approaches. Although Zhang [27] and Riad and Zhang [28] have, respectively, demonstrated triaxial conditions and hysteresis within the MSSA framework, for simplification hysteresis will be neglected and only isotropic loading is considered herein.

***The State Surface Approach***

The concept of a state surface approach to unsaturated soil modeling is perhaps first attributable to the 1962 contribution of Coleman [18], who proposed incremental linear elastic equations for calculation of soil volume change ($dV$) as shown in Equation 1 [1]. The coefficients of compressibility for the soil skeleton ($C_{21}$, $C_{22}$, and $C_{23}$) were assumed dependent on the current values of the stress variables ($u_w$-$u_a$, $\sigma_m$ -$u_a$, and $\sigma_1 - \sigma_3$), as well as the stress history.

$$-dV/V = -C_{21}(du_w\text{-}du_a) + C_{22}\ (d\sigma_m\ \text{-}du_a) + C_{13}\ (\ d\sigma_1\text{-}d\sigma_3) \quad (1)$$

Although parallel treatment of the soil skeleton and water phase will not be directly addressed in this paper, it is of major significance that Coleman, using the same stress variables as those of Eq. 1, above, also presented equations for volume change associated with the water phase of the soil. The need for consideration of two state (constitutive) surfaces, one the for the soil skeleton and one for the water phase was demonstrated and promoted by Matyas and Radhakrishna [2], and subsequently by Fredlund and Morgenstern [3]. It was Matyas and Radhakrishna who presented the first experimentally-determined state surfaces for unsaturated soils, presenting their laboratory compression test results in 3-D plots of void ratio as a function of the two independent stress variables of $\sigma$-$u_a$ (net stress) and $u_a$-$u_w$ (suction). The state surface of Matyas and Radhakrishna was found to have a warped shape, due to the collapse response of the soil, and was demonstrated to be non-unique for some paths (Fig. 1). Lloret and Alonso [29] pointed out that a state surface approach is sufficiently general that it can be used as a unified approach for expansion or collapse volume change response.

Fredlund and Morgenstern [3] demonstrated through laboratory null tests that the controlling stress variables of net stress and suction were valid, and these variables were also shown to be consistent with multiphase continuum mechanics for a three-phase (solid, water, and air) soil. In their landmark contribution, Fredlund and Morgenstern presented constitutive equations for soil volume change, along with constitutive (state) surfaces. A state surface-based incremental elastic constitutive model for volume change of unsaturated soils (both void ratio and water content phases) was proposed, based on the demonstrated net stress and suction variables and the demonstrated essential uniqueness of the void ratio state surface for small, incremental changes in stress [1].

Void ratio state surfaces were found to be unique for monotonic loading paths only. Reasoning that an appropriately-applied incremental elastic model can be used to capture path-dependent nature of unsaturated soil volume change response, and in a manner consistent with classical soil mechanics, Fredlund and Rahardjo [1] presented a complete theory for volume change of unsaturated soils based on an incremental elastic formulation. Constitutive equations were provided for the void ratio (and water content) and plotted as a 3-D constitutive surface relating void ratio to changes in net stress and suction.

The instantaneous slopes of the state surface, as proposed by Fredlund and Rahardjo, are shown in Fig. 1. Fredlund and Rahardjo also presented their state surface volume change theory using the familiar void ratio versus log stress format and corresponding compressibility indices (Fig. 2), including different compression and rebound slopes.

Although the state surfaces were generally presented for monotonic loading conditions, Fredlund and Rahardjo, taking an incremental linear elastic approach, acknowledged the change in slope of the state surface with stress level (both net stress and suction) and required change in slope of the state surface upon unloading. Fredlund and Rahardjo presented both void ratio and water content constitutive surfaces – a key feature of their fundamental approach. It is of significant note that the first textbook on unsaturated soils, Soil Mechanics for Unsaturated Soils by Fredlund and Rahardjo [1], was presented in the context of classical soil mechanics, using primarily limit equilibrium and incremental linear elastic approaches to facilitate ease of adoption in established geotechnical engineering practice. Although the focus of this paper is on unsaturated soil volume change, the Fredlund and Rahardjo book contained a complete theory of unsaturated soil mechanics covering topics of shear strength, volume change, and unsaturated flow, along with applications and discussions on laboratory testing methods.

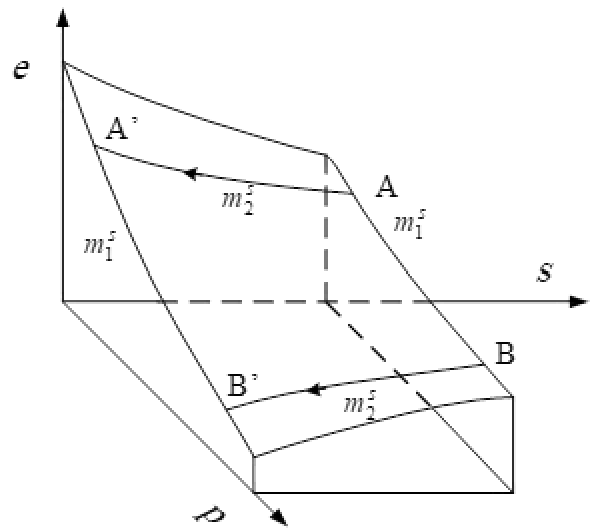



Figure 1. Warped state surfaces for void ratio of Matyas and Radhakrishna [2], showing instantaneous state surface slopes of Fredlund and Rahardjo [1] (from Houston and Zhang [26])

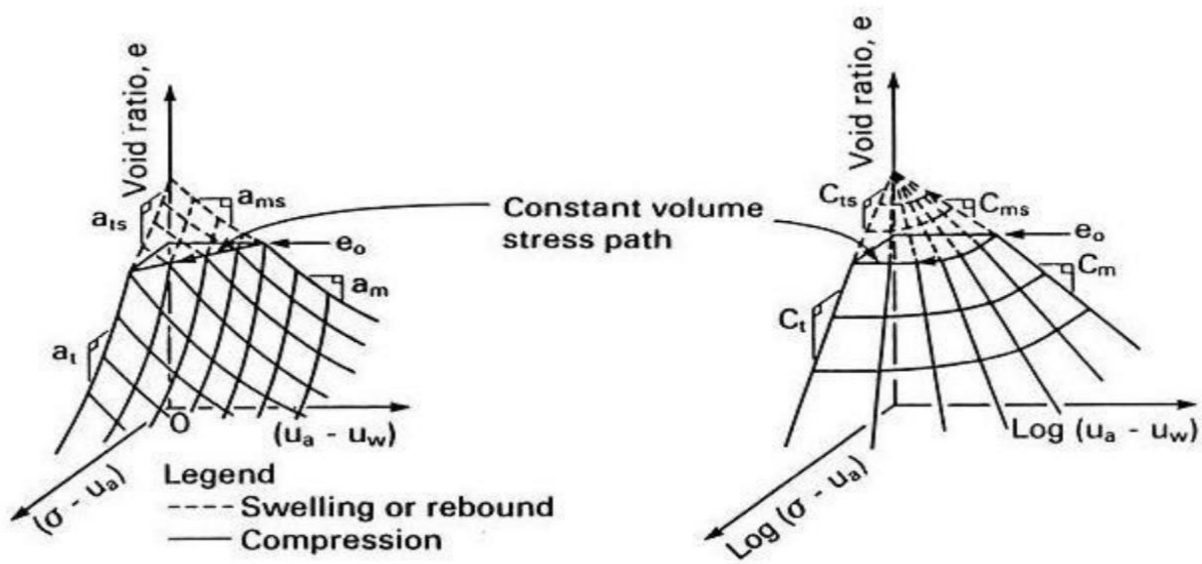



Figure 2. Void Ratio Constitutive Surface (from Fredlund and Rahardjo [1])

***The Barcelona Basic Model (BBM)***

Unsaturated soils exhibit highly nonlinear and elastoplastic response to loading and unloading associated with changes in net stress and/or suction. Thus, unsaturated soil behavior is stress path and stress history dependent – behavior that is only indirectly (via incremental elasticity) captured by the State Surface Approach. A breakthrough in elastoplastic constitutive modeling of unsaturated soil came with the Barcelona Basic Model (BBM) developed by Alonso, et al. [4]. The BBM, intended to be a simplified elastoplastic model capturing some of the key known aspects of response, has been exceptionally well-received by unsaturated soil elastoplastic modelers, and quite notably was designed to capture the collapse response (compression of the soil due to reduction of suction under applied net stress).

The BBM is a complete critical state model for unsaturated soils that is built on the framework of the Modified Cam Clay model for saturated soils. Fig. 3 shows the BBM for isotropic loading conditions, with Fig. 3(a) being the 2-D *e-p* plane projection of suction controlled isotropic loading of four "identical" soil specimens tested at various suction values. The yield points are interpreted to be the knee of the e-log p curve. For the *s=0* curve, absent sample disturbance, the yield point is the preconsolidation pressure. In the BBM, the yield points are used to describe the shape of the LC yield curve, as shown in the *s-p* projection of the yield curve of Fig. 3b. Fig. 3b is based on the underlying assumption that the test specimens are identical and share the same yield curve.

The yield curve is referred to as the Loading Collapse (LC) curve because a collapse response (reduced void ratio upon wetting under load) is obtained for an initial condition on the yield curve followed by, for example, a loading condition corresponding to decreasing suction under constant *p*. In the BBM as suction increases the yield stress increases, as shown in Fig. 3(b). In the elastic region, a reduction in soil suction (wetting) results in soil expansion, but reduction in soil suction (wetting) from a point on the yield surface results in an outward expansion of the yield curve and a collapse response. Evolution of the LC yield curve occurs in response to increase in net mean stress and in response to reduction in soil suction. Due to elastoplastic response, volume change of the soil can be path dependent. Under the original simplifying assumptions of the BBM, as intended by the developers, the model has been shown to capture soil response well in the collapse range. In the original BBM a suction increase (SI) yield surface was also included, however most researchers have now dropped the SI curve.

Fig. 4 shows the slopes of the suction -controlled isotropic compression curves of Sivakumar and Wheeler [31]. Consistent with terminology of the Modified Cam Clay model, the unload-reload slope is designated as $\kappa$ (saturated) or $\kappa(s)$ (unsaturated) and the virgin loading curve is designated by $\lambda$ or $\lambda(s)$. Unfortunately, in development of the BBM the concept of the void ratio state surface was essentially abandoned, as discussed by Zhang and Lytton [5]. The void ratio state surface was only indirectly considered in the BBM via use of the isotropic compression lines obtained through laboratory testing at different controlled values of suction.

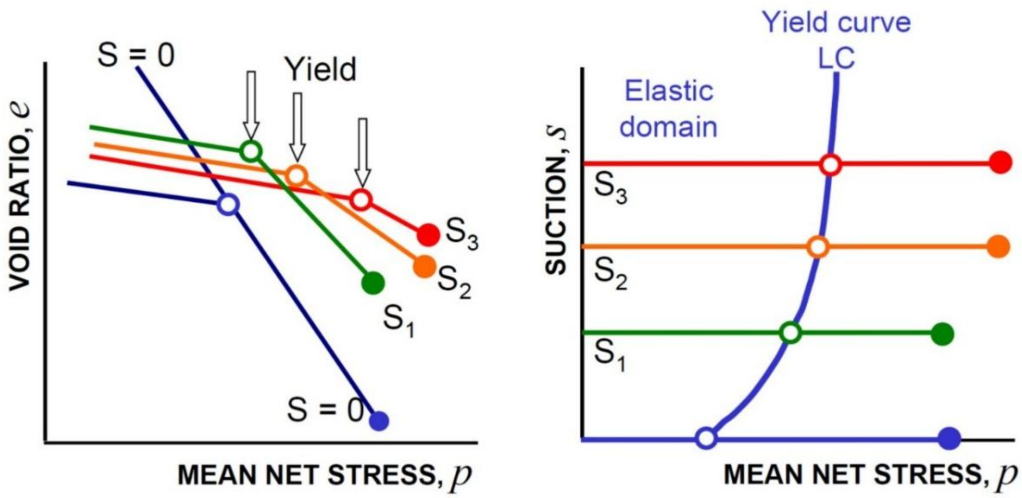



Figure 3.(a) Isotropic suction-controlled compression, (b) Traditional assumption that the specimens are identical and share the same yield curve(from Gens[30]).

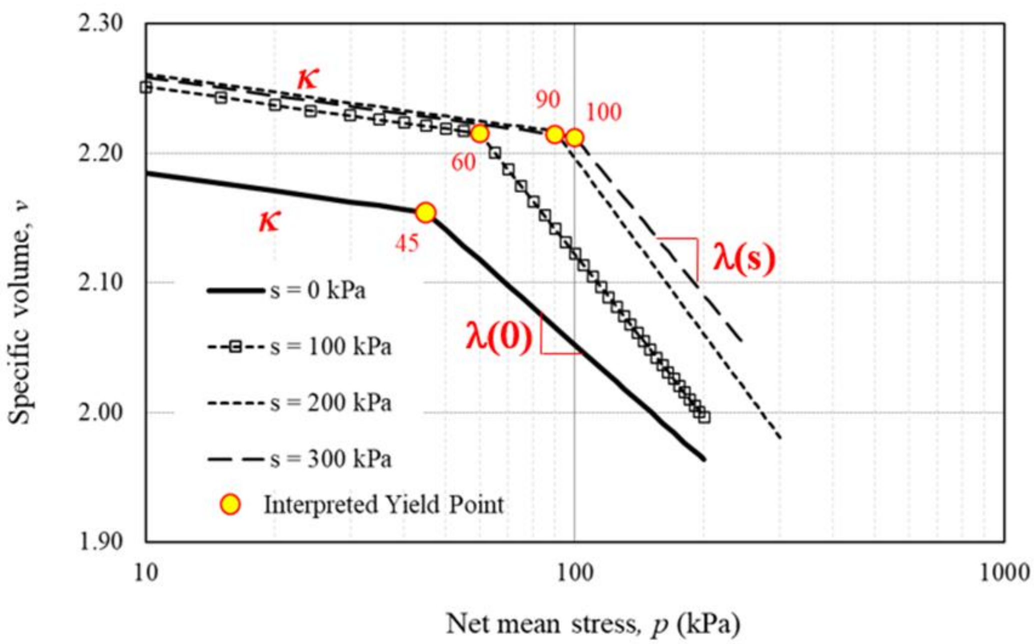



Figure 4. Isotropic compression curves of Wheeler and Sivakumar [31] showing interpreted yield points and slopes of the $v$ vs *log p* curves for suctions of 0, 100, 200, and 300 kPa under the traditional BBM assumption of identical specimens sharing the same LC (from Zhang and Houston [25]).

**The Modified State Surface Approach as the Link Between State Surface and Elastoplastic Models**

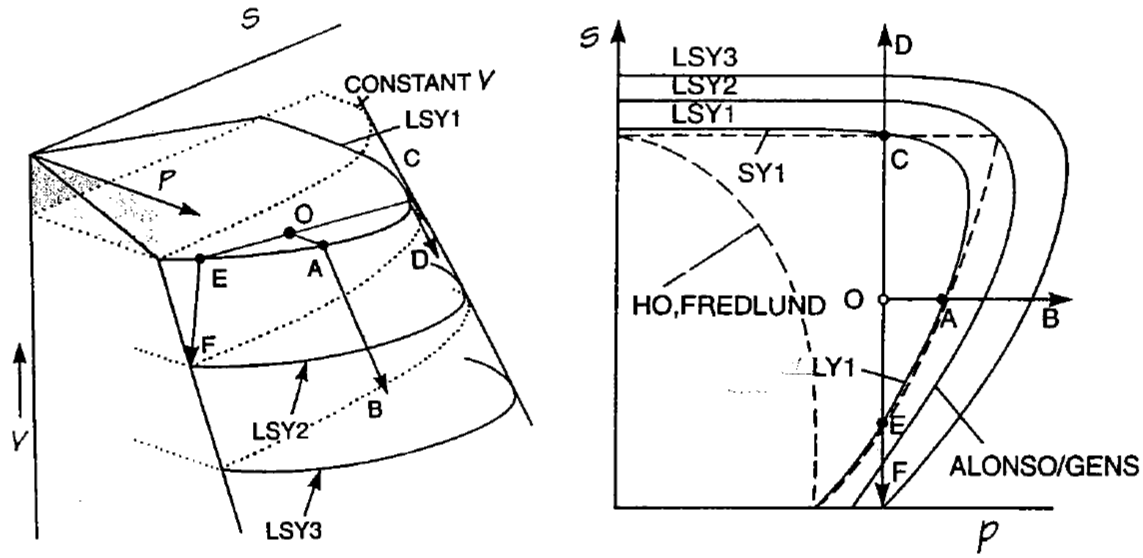
***Background***

As pointed out by Zhang and Lytton [5], after introduction of the BBM (Alonso, et al [4]) essentially all unsaturated soil elastoplastic modelers dropped state surface (constitutive surface) terminology in favor of such phrases as plastic hardening surface, state boundary surface, or virgin normal compression surface. After Alonso, et al. [4], unsaturated soil modelers began to clearly distinguish between state surface approaches and elastoplastic approaches – resulting in a likely unintended impression that the differences between state surface methods and elastoplastic methods could not be rectified, i.e., that their approaches were very, if not totally, different. Regardless the "camp" within which an unsaturated soil researcher operated ("state surface" or "critical state") it appears as if little effort was made to explore the two volume change modeling approaches within a unified framework. Yet state surface methods and elastoplastic methods modeled the same observed soil volume change responses.

Despite limited attention, it was recognized early-on that there was a connection between the isotropic suction-controlled normal compression lines and the LC curve of the BBM [32,33]. As observed by Zhang and Lytton [5], Wheeler and Karube [32] comment that there is "an inextricable link between the form of the normal compression lines in the *v:p* plane and the shape of the loading–collapse (LC) yield curve as it expands in the *s:p* plane." Delage and Graham [33] presented a 3-D plot, Fig. 5a, of the variation of specific volume in terms of net stress and suction, clearly showing a link between the yield points pulled from the suction-controlled isotropic compression lines and the shape of the LC. As noted by Zhang and Lytton [6], Delage and Graham [33] also correctly discussed that constant volume curves (e.g., the Ho and Fredlund curves of Fig. 5) are different than yield curves (e.g., the Alonso/Gens curves of Fig. 5). As recognized by Delage and Graham, the Ho, et. al. [34] curve of Fig. 5b does appear to come from a locus of points of constant volume, consistent with Fig. 2, rather than being a locus of yield points. It has been subsequently established that constant volume curves are not the same as yield curves for unsaturated soil, as cautioned by Delage and Graham [6,24].

Although discussed to a limited extent in the mid-1990s literature, as presented above, the link between the virgin loading isotropic compression lines and the observed shape of the yield curves appears not to have been further explored as elastoplastic modeling of unsaturated soils progressed. State surface methods for modeling of volume change of unsaturated soils also continued to evolve but separate from elastoplastic modeling efforts. It was the insightful contribution of Zhang and Lytton [5, 6] that explicitly recognized that the virgin loading state surface was exactly the trace of the yield curve. In development of their Modified State Surface Approach (MSSA), Zhang and Lytton [5,6] provided a clear link between the state surface approach and elastoplastic models for unsaturated soils, plainly demonstrating the connection between shape and evolution of the yield curve and the virgin loading state surface. Another important point of Zhang and Lytton is that the shape of yield curves from isotropic suction controlled compression tests must account for the fact that specimens at different suction are not "identical" because of differing stress histories [5,6].

As discussed in the section below, the MSSA further provided an elastoplastic framework for deeper understanding of complex unsaturated soil response to changes in the two controlling stress variables, net stress and suction.




(a) (b)

Figure 5. 3-D plots of volume change response of unsaturated soils in response to changes in the two independent stress variables of *p* and *s* (from Delage and Graham[33])

***The Modified State Surface Approach***

In contrast to traditional incremental approaches to constitutive model development, Zhang and Lytton [5,6,7] presented an integrated elastoplastic framework, the Modified State Surface Approach (MSSA), from which to study various two independent stress variable (total stress and negative pore water pressure) unsaturated soil constitutive models. The MSSA embraces the notion that unsaturated soil behavior is elastoplastic, that behavior is controlled by two independent stress state variables (net stress and suction) and that behavior is highly nonlinear, and thus requires that laboratory results be compared from multiple specimens with differing stress histories. An integrated approach facilitates full understanding of model features, whereas when using incremental approaches relationships between different components of unsaturated soil models are not always obvious. Furthermore, the link between elastoplastic/critical state models and state surface approaches, relatively disguised by incremental approaches, are illuminated through the MSSA lens.

The integrated approach taken in the MSSA provides a big-picture framework from which investigation of the relationship between various model components can be undertaken, and from which different modeling approaches can be compared. The basic principles of the MSSA are simple to understand: (1)Unsaturated soil void ratio response is controlled by two independent stress variables of net stress and suction (total stress and negative pore water pressure for atmospheric air pressure conditions); (2) The elastoplastic void ratio response (surface) is unique; (3) The elastic surface is assumed fixed in configuration but moves downward (to lower void ratio) as plastic deformations occur; (4) The intersection of the elastic and elastoplastic surfaces is the yield curve.

Thus, the infinite set of all possible isotropic loading compression curves become elastic and elastoplastic surfaces when void ratio (specific volume) is plotted as a function of the two separate variables of total stress and pore water pressure, as shown in Fig. 6. Importantly, the elastoplastic surface represents virgin loading of the soil, as it is comprised of the full series of possible compression curves with slope $\lambda(s)$. The elastic surface represents unloading and reloading, corresponding to the infinite set of curves with slope $\kappa_s$. Zhang and Lytton demonstrate that the evolution of the yield curve is always described by the shape of the virgin loading surface because the intersection between the elastic and elastoplastic surface is the yield curve. For example, in Fig. 6, loading path E to V is elastoplastic, resulting in the yield curve moving from BEH to WVU as established by the shape of the virgin loading surface (BEHIFC). Virgin loading E to V changes the shape of the void ratio state surface because the elastic surface moves downward to a new, lower elastic state surface A'G'UVW.

The MSSA takes full advantage of the uniqueness of the virgin loading response, recognizing that the virgin loading void ratio state surface is the same as the elastoplastic loading surface and thus describes that evolution of the yield curve for unsaturated soils. It is this connection between the virgin loading state surface and the evolution of the yield curve that is the necessary link between the two independent stress variable state surface approach and two independent stress variable elastoplastic constitutive models. Essentially, a void ratio state surface, such as depicted in Figs. 1 and 2, provides a (typically laboratory-determined) combined elastic and elastoplastic unsaturated soil response for monotonic loading of a soil having a specific stress-history.

Using the principles of the MSSA, Zhang and Houston [4] presented a 3-D big-picture schematic of the elastoplastic response of soil for transition from unsaturated to saturated under isotropic loading conditions, as shown in the arithmetic plot of Fig. 7. The air-entry value shown in Fig. 7, is the negative pore water pressure at which air enters the largest soil pores and the soil becomes unsaturated. Although often assumed to be constant, the air-entry value is affected by stress path and stress history. In Fig. 7, the mean total stress, $\sigma_m$, and pore water pressure, $u_w$, are kept separate (that is, not combined into a single-valued effective stress) for both saturated and unsaturated soil conditions. Although the void ratio state surface is commonly plotted using an arithmetic scale for suction and net stress, in general it is easier to detect the transition from elastic to elastoplastic soil response using the familiar semi-log plot format. Using a semi-log plot for the stress variables of net stress and suction, the transition point between elastic and elastoplastic response (i.e., yield) is typically associated with a significant "knee" in the compression curve – consistent with the BBM interpretation of yield.

The simple principles of the MSSA show that the virgin loading state surface is the same as the elastoplastic surface and that the intersection of the elastic and elastoplastic surfaces provides the evolution of the yield curve. Thus, using the Fig. 7 MSSA schematic a comparison of yield curves for expansion response region to collapse response region of the virgin loading state surfaces can be made. The MSSA-generated framework of Fig. 7 demonstrates that yield curves are always inclined at $45^{o}$ for suction values less than the air-entry, but for collapse the yield curves are curved towards decreasing net stress and yield curves in the expansion region are curved towards increasing net stress. The LC curves of the BBM are appropriately shaped for collapse response, as intended by Alonso, et al. [4].

As noted above, Lloret and Alonso [29] observed that the state surface approach allowed for modeling of unsaturated soil volume change whether the response to wetting under load is expansive or collapsible. When using a state surface approach, the shape of the state surface is tracked via an incremental adjustment to moduli. Separation of elastic and elastoplastic response is not direct in a traditional state surface approach but can be handled by an incremental adjustment to moduli to account for differences in moduli between elastic (reloading-unloading) and elastoplastic (virgin) loading conditions. Importantly, the unique virgin loading portion of a state surface is identical to the elastoplastic (virgin loading) surface used in elastoplastic constitutive models. Using the MSSA, Zhang and Lytton observed that by allowing more flexibility in the shape of the yield curve that relatively simple modifications to existing unsaturated soil elastoplastic models, such as the BBM, could be made that allowed for modeling of both expansion and collapse elastoplastic response. Houston and Zhang [26] expanded on and demonstrated the concept of elastoplastic modeling of expansive and collapsible soils under a unified elastoplastic framework based on MSSA principles.

Comparing the traditional state surface approach to elastoplastic/critical state models through the lens of the MSSA, it can be easily seen that these two approaches are modeling the same soil behavior. It is the author's opinion that either method can work when understood within the big picture framework of elastoplastic soil response (Fig. 6 and Fig. 7).

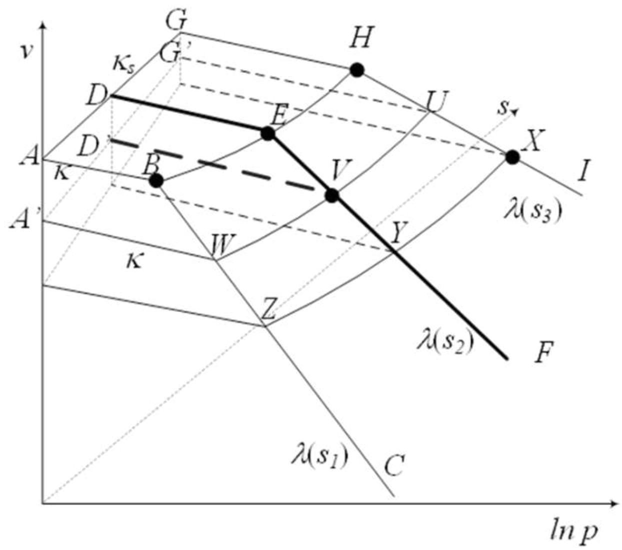



Figure 6. MSSA view of specific volume elastic and elastoplastic state surfaces based on conventional BBM interpretation of suction-controlled isotropic compression tests [9].

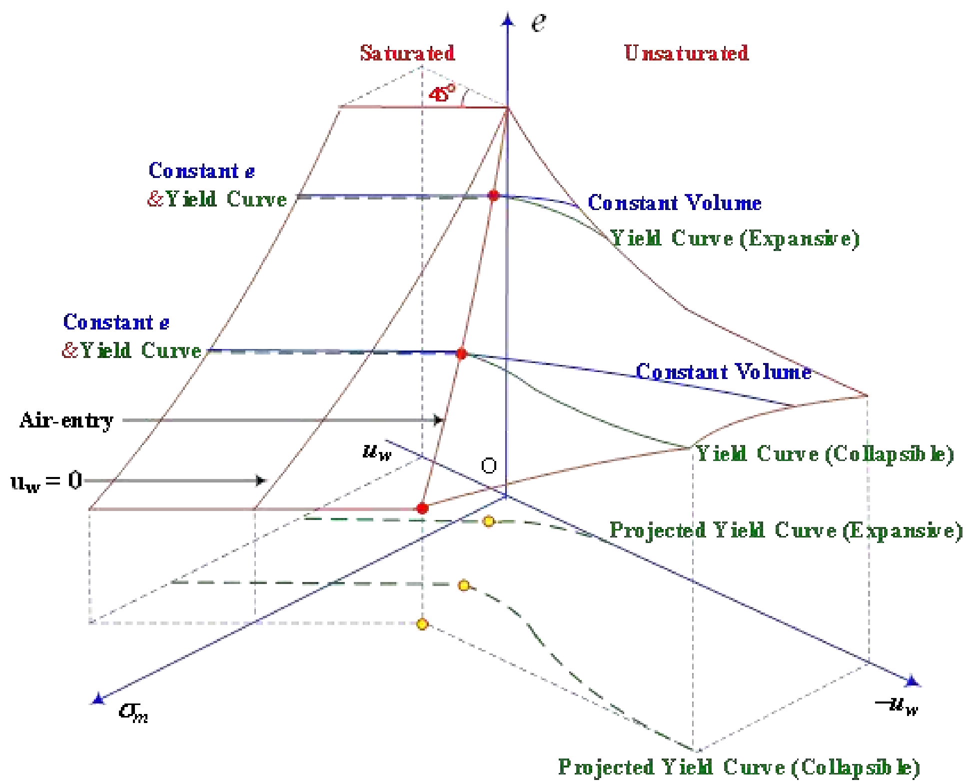


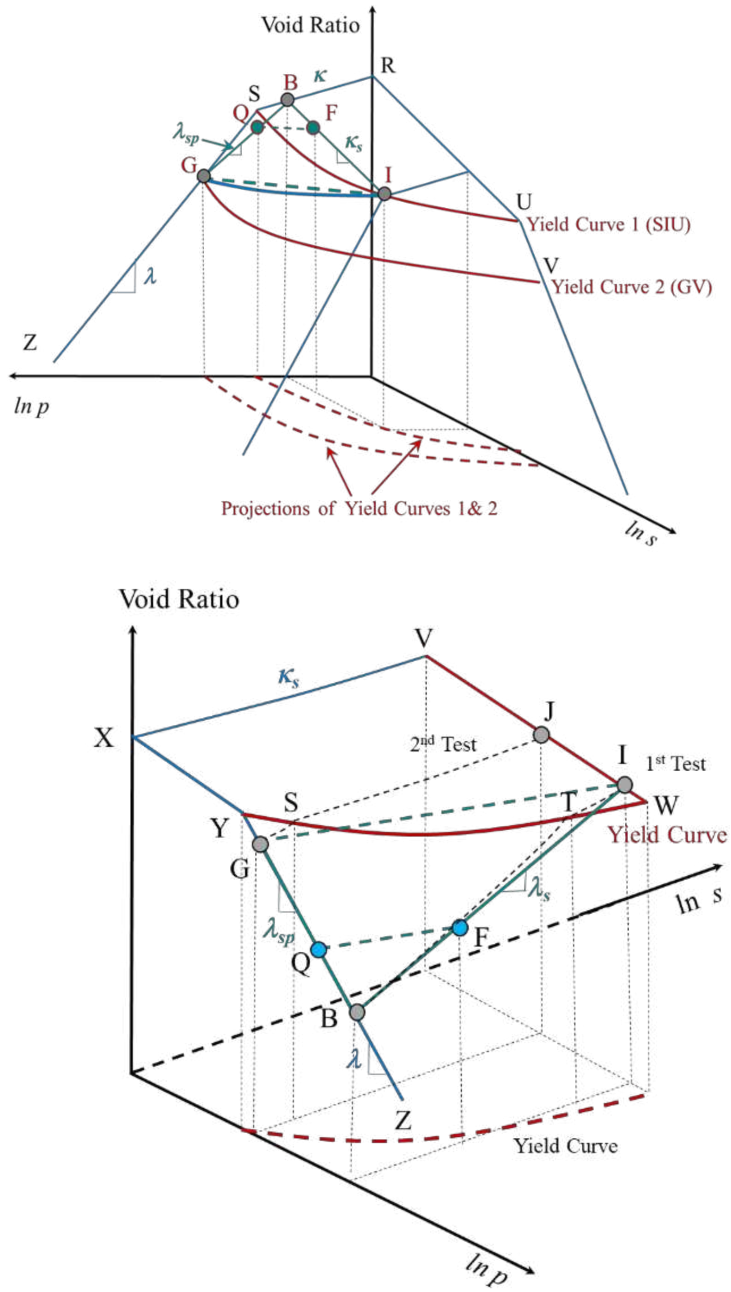

Figure 7. Modified State Surface view of soil elastoplastic behavior from unsaturated to saturated (from Houston and Zhang [9])

**Stress Path Approaches for 1-D $K_o$ Loading Conditions**

Stress path approaches, being laboratory testing-based, must follow a two independent stress variable approach to unsaturated soil modeling. This is because unsaturated soil testing must be performed by separate control of net stress and suction (total stress and negative pore water pressure). Not all laboratory test methods attempt to follow field-appropriate stress paths. Fortunately, with regard to volume change of unsaturated soils, there are 1-D, $K_o$ stress-path appropriate standard test methods available within the ASTM D-4546 test methods, whether the response of the soil is expansion or collapse. Often the field-appropriate path is one of constant total stress. One challenge in the use of non-suction-controlled stress path testing is how to interpret test results for suction values intermediate between in-situ and full wetting ($s=0$). Where suction values do not correspond to full wetting, it is important to understand how the interpretation of the test fits within the big picture of unsaturated soil response. Again, the MMSA provides an appropriate lens for assessment of appropriateness of the laboratory test method and any interpretation of test results.

The anticipated field stress paths for the ASTMD4546 test are shown in Fig. 8 for soils exhibiting expansion (Fig. 8a) and collapse (Fig. 8b) under a simplifying assumption (for plotting purposes only) of $K_o$ equal to 1. The field stress path begins at point I and follows path IFB. In the laboratory, compacted test specimens are first prepared according to field specifications and then loaded to point I corresponding to initial field state. For natural soils or existing fill soils, the best undisturbed specimen possible is collected from the field (with in-situ moisture state preserved) and then the specimen is loaded to field stress levels, corresponding to point I. The oedometer test proceeds by holding the net vertical stress constant and monitoring specimen volume change while submerging the specimen to achieve an $s=0$ condition, yielding an estimate of soil volumetric strain that would be expected if full wetting were to occur in the field.

To estimate field strains for values of suction intermediate between in-situ and full wetting, Houston and Zhang [26] demonstrate that the Surrogate Path Method (SPM) of Houston and Houston and Singhal [35,36] is consistent with the MSSA big-picture response of the soil for the field stress path IFB. The SPM uses a stress path method and it is appropriate for soils exhibiting expansion, collapse, or both, and uses laboratory results from conventional oedometer devices having no suction control or measurement capability. The SPM remains as true as possible to the MSSA representation of unsaturated soil elastoplastic response while using routine oedometer test results. The laboratory ASTMD4546 testing approach [8] is an integral part of the SPM because the full-wetting response of the specimen is used to anchor the volume change estimate to the $s=0$ elastoplastic soil response (RSG, Fig. 8a and XYZ, Fig. 8b). Thus, the SPM avoids challenges often encountered where slopes of the suction-change path are represented as straight line on semi-log plots.




(a)

(b)

Figure 8. (a) Stress path for oedometer tests on soil exhibiting expansion response and associated SPM interpretation, (b) Stress path for oedometer tests on soil exhibiting collapse response and associated SPM interpretation (from Houston and Zhang [26])

Because suction is not controlled in the SPM laboratory test, suction can only be estimated or measured for initial and final test conditions. In the SPM, a surrogate path in the *s=0* plane is obtained through a mapping process of the actual stress path to estimate wetting induced volume change for less than full saturation (to suction values between initial and *s=0*). Although not required for using the suction interpolation approach of the SPM, the mapping onto the *s=0* plane provides a comfortable format for geotechnical practitioners who have become accustomed to using methods of unsaturated soil volume change analyses that are presented in the *s=0* plane (e.g., most heave computation methods).

Because field pore water pressures are always negative above the groundwater table, field suction rarely reaches *s=0* in the absence of perched water or groundwater table rise conditions. Therefore, a volume change analysis consistent with unsaturated soil mechanics theory must include estimation of field initial and final suction design values. The suction-based mapping from the *s-p* space to the *s=0* plane is a unique feature of the SPM which anchors the result of the volume change computation to fall between the full-wetting (*s=0*) strain (ASTMD-4546 single specimen test) and zero strain for the case of no wetting. In addition to the D-4546 test result, the mapping process requires an estimate of the net total stress corresponding to zero volume change upon wetting (e.g., the zero volume change swell pressure).

Fig. 8a shows the surrogate path BG with slope $\lambda_{sp}$ for wetting under constant load from point I. Point B is obtained from ASTM D4546, Method B (wetting under field overburden load). Point G (swell pressure), is most often obtained by load-back of the Method B specimen, but could also be obtained using the multiple specimen method, ASTM D4546 Method A. Mapping from path IFB (the actual path) to GB (the surrogate path) is accomplished through a simple ratio of initial and final soil suction, with the intent that, for partial wetting, the void ratio at point Q on the surrogate path matches the void ratio at point F on the actual path. The SPM computation (in the *e-log p, s=0* plane) for expansion response to wetting (Fig. 8a) proceeds as follows.

$$(\sigma_p)_{exp} = \sigma_o + R_w\,(\sigma_{cv} - \sigma_o) \qquad (2)$$

where $R_w$ = suction ratio = $s_f/s_i$; $\sigma_i$ =initial matric suction; $\sigma_f$ =final matric suction; $R_w = 1$ for no wetting; $R_w = 0$ for full wetting; $\sigma_p$ = surrogate final stress corresponding to final suction; $\sigma_o$ = initial total stress in the field corresponding to initial suction.

$$\varepsilon_{pw} = (C_H)_{SP} \log\,(\sigma_{cv}/(\sigma_p)_{exp} \text{ = partial wetting swell strain for expansion} \qquad (3)$$

where $(C_H)_{SP}$ =slope of the surrogate path on the log *p* (net total stress) and *s=0* plane.

The SPM mapping approach has been shown through some limited amount of suction-controlled oedometer testing to result in reasonably good agreement between swell strain (void ratio) at points F (on the actual path) and void ratio at point Q (on the surrogate path) [36, 37]. An advantage of the SPM is that whether the mapping is exact the result, being anchored to the full-wetting specimen response, will always be reasonable, falling between the full wetting strain and zero, and avoiding gross over-estimates or underestimates that are known to occur with some other oedometer-based volume change estimation procedures.

The MSSA view of the SPM for collapse response is shown in Fig. 8b. For this case, the field specimen is assumed to exist at point I, on the elastic plane WVXY. Test method ASTMD4546 follows the actual field path ITFB, and the surrogate path in *s-p* space is IFB. Due to transition from elastic to elastoplastic response at point T, the surrogate path deviates slightly from the actual path. However, the surrogate path still represents a very good approximation of the actual path. A second "identical" specimen is tested at a reduced stress level, following path JSG. Assuming, for convenience in this discussion, that there is essentially no volume change along path JSG, *p* at point J represents the transition between expansion and collapse for the field soil. The surrogate path in the *s=0* plane is GB, and the partial wetting collapse strain at point F (on the actual stress path) is estimated from point Q on the surrogate path using the initial and final soil suction ratio-based mapping similar to that discussed above for expansive soils, except

$$(\sigma_p)_{col} = \sigma_G + (1\text{-}R_w)\ (\sigma_o - \sigma_G) \qquad (4)$$

$$\varepsilon_{pw} = (C_H)_{SP} \log\ (\sigma_{cv}/(\sigma_p)_{col} = \text{partial wetting strain for collapse} \qquad (5)$$

Partial wetting strains are anchored between zero at point I for no wetting and B for full wetting.

**Hierarchy of Modeling Approach**

There is an unspoken but typically assumed hierarchy of unsaturated soil modeling, with elastoplastic/critical state models often being assumed to be at the top. However, it is a rare occurrence, and generally limited to only high risk, high dollar, and/or research, that elastoplastic unsaturated soil models are employed in real-world unsaturated soil applications. Further, it is typically unknown how well elastoplastic models can be extended outside of the range of laboratory data (and associated unsaturated soil responses) upon which such models were developed and validated. Hence, the use of a "higher level" constitutive model does not always assure improved predictions, and there are some common applications where substantial simplification is appropriate.

On the hierarchical scale of volume change modeling, most unsaturated soil researchers would likely categorize a direct stress path approach, such as the 1-D response-to-wetting under constant load (e.g., ASTM D4546), as a lower level approach, falling below elastoplastic approaches and state surface approaches. However, in the context of the big picture, it has been demonstrated that even simple non-suction controlled oedometer tests can lead to sound unsaturated soil volume change analyses where field conditions can be reasonably considered to result in 1-D, $K_0$ volume change.

**Summary of The Link between Unsaturated Soil Volume Change Modeling Approaches**

Three unsaturated soil volume change estimation methods have been discussed herein. It is fair to say that most unsaturated soil researchers would consider elastoplastic/critical state models as a hierarchical Level 1 approach, with state surface approaches being categorized perhaps as a Level 2. Undoubtably, under such hierarchical ranking, unsaturated soil researchers would consider methods based on non-suction controlled oedometer tests as Level 3. However, using the integrated elastoplastic framework of the MSSA, all three approaches have been shown to be consistent with known complex elastoplastic/path-dependent unsaturated soil behavior provided consistency with field stress conditions and stress paths are considered. Rather than emphasizing differences in approach, it is actually more useful to advancement of unsaturated soils into practice for the unsaturated soils community to recognize commonalities of methods – asking questions, for example, such as "Does the method properly consider the role of the two independent stress variables (net stress and suction) and is the elastoplastic response of the soil taken into account in a field condition-appropriate manner?" It is the author's view that differences (real and perceived) in laboratory stress-path, stress state surface, and elastoplastic models can be readily studied using the Modified State Surface approach.

The key link between the various hierarchical levels of volume change estimation is found in the use of two independent stress variables of net stress and suction. The clear link between the state surface approach and an elastoplastic/critical state approach is that the virgin loading state surface is identical to the elastoplastic surface. This link highlights an important aspect of unsaturated soil response for elastoplastic modeling of unsaturated soils whereby laboratory-demonstrated virgin loading state surfaces are used as a tool for describing the evolution of the yield curve.

## References


[1] D.G. Fredlund and H. Rahardjo. S*oil mechanics for unsaturated soils*. New York: John Wiley & Sons, 1993. http://dx.doi.org/10.1002/9780470172759 .

[2] E. L. Matyas and H.S.D. Radhakrishna. Volume change characteristics of partially saturated soils. *Geotechnique*, 18(4): 432-448 , 1968. http://dx.doi.org/10.1680/geot.1968.18.4.432.

[3] D.G. Fredlund and N.R. Morgenstern. Stress state variables for unsaturated soils. *Journal of Geotechnical and Geoenvironmental Engineering*, 103(5), 1977. http://dx.doi.org/10.1061/ajgeb6.0000423.

[4] E.E. Alonso, E.E., A. Gens, and A. Josa. A constitutive model for partially saturated soils. *Geotechnique*, 40(3): 405-430, 1990. http://dx.doi.org/10.1680/geot.1990.40.3.405.

[5] X. Zhang and R.L. Lytton. Modified state-surface approach to the study of unsaturated soil behavior. Part I: basic concept. *Canadian Geotechnical Journal*, 46(5): 536-552, 2009a. http://dx.doi.org/10.1139/T08-136.

[6] X. Zhang and R.L. Lytton. Modified state-surface approach to the study of unsaturated soil behavior. Part II: general formulation. *Canadian Geotechnical Journal*, 46(5): 553-570, 2009b. http://dx.doi.org/10.1139/T08-137.

[7] X. Zhang, and R.L. Lytton. Modified state-surface approach to the study of unsaturated soil behavior. Part III: modeling of coupled hydromechanical effect. *Canadian Geotechnical Journal*, 49(1): 98-120, 2012. http://dx.doi.org/10.1139/t11-089.

[8] ASTM D4546. S*tandard test methods for one-dimensional swell or collapse of soils.* ASTM International, West Conshohocken, PA, 2014. https://doi.org/10.1520/D4546-14E01

[9] K. Terzaghi. *Theoretical soil mechanics*, Wiley, New York, 1943.

[10] S.L. Houston. It is time to use unsaturated soil mechanics in routine geotechnical engineering practice. *Journal of Geotechnical and Geoenvironmental Engineering*, 145(5), 2519001, 2019. http://dx.doi.org/10.1061/(ASCE)GT.1943-5606.0002044.

[11] L. Richards and M. Fireman. Pressure plate apparatus for measuring moisture sorption and transmission by soils, *Soil Science*, V. 56, 395-404, 1943.

[12] J.W. Hilf. An Investigation of Pore-water Pressure in Compacted Cohesive Soils, Technical Memorandum No. 654, U.S. Department of the Interior, Bureau of Reclamation, Design and Construction Division, Denver, CO, 1956.

[13] D. Croney, D., J. Coleman, and W. Black. Movement and distribution of water in soil, in relation to highway design and performance, Highway Research Board Special Report, 1958.

[14] A.W. Bishop and I.B. Donald. The experimental study of a partly saturated soil in a triaxial apparatus. Proc. 5th Int'l Conf on Soil Mech and Fdn. Engr., Paris, V. 1, 13-21, 1961.

[15] L. Barden, A.O. Madedor, and G. R. Sides. Volume change characteristics of unsaturated clay, ASCE*, J. Soil Mech and Fdn Div*., V. 95, SM1: 33-52, 1969.

[16] G.D. Aitchison. *Moisture equilibria and moisture changes in soils beneath covered areas*, A Symposium in Print, G. D. Atchison, ed., Butterworths, Australia, 1965.

[17] D.G. Fredlund and N.R. Morgenstern. "Stress state variables for unsaturated soils." *Journal of Geotechnical and Geoenvironmental Engineering*, 103, 1977. https://doi.org/10.1061/ajgeb6.0000423

[18] J.D. Coleman. Stress/strain relations for partly saturated soils, *Geotechnique*, 12 (4): 348-350, 1962.

[19] A.W. Bishop. The principle of effective stress, *Teknisk Ukeblad* 106: 859–863, 1959.

[20] D. Atchison. Relationship of moisture and effective stress functions in unsaturated soils, in *Pore Pressure and suction in soils*, British National Soc of Int'l Sec. of SMF Engr. Conf., London, Butterworths: 47-52, 1961.

[21] J.E. Jennings. A revised effective stress law for use in the prediction of behavior of unsaturated soils, in *Pore Pressure and Suction in Soils*, British National Soc of Int'l Sec. of SMF Engr. Conf., London, Butterworths: 26 – 30, 1961.

[22] A.W. Bishop and G.E. Blight. "Some aspects of effective stress in saturated and unsaturated soils." *Geotechnique*, 13: 177–197, 1963.

[23] G.E. Blight. A study of effective stress for volume change. *Moisture equilibria and moisture changes in soils beneath covered areas*, A Symposium in Print, G. D. Atchison, ed., Butterworths, Australia: 259-269, 1965.

[24] J.E. Jennings and J.B. Burland. Limitations to the use of effective stresses in partly saturated soils. *Geotechnique*, 12(2):125-144, 1962. http://dx.doi.org/10.1680/geot.1962.12.2.125.

[25] X. Zhang and S. Houston. Indefinability of effective stress for unsaturated soils, ASCE *Journal of Geotechnical and Geoevironmental Engineering*, Volume 150, Issue 8, August, 2024. https://ascelibrary.org/doi/10.1061/JGGEFK.GTENG-12435

[26] S.L. Houston and X. Zhang. Review of Expansive and Collapsible Soil Volume Change Models Within a Unified Elastoplastic Framework." *Soils and Rocks*, 44 (3), 2021.

[27] X. Zhang, X. Analytical solution of the Barcelona basic model. In *Experimental and Applied Modeling of Unsaturated Soils*, Shanghai, China: 96–103, 2010.

[28] B. Riad and X. Zhang. Modified state surface approach to study unsaturated soil hysteresis behavior. *Transportation Research Record: Journal of the Transportation Research Board*, 2674(10): 484-498, 2020. http://dx.doi.org/10.1177/0361198120937014.

[29] A. Lloret and E.E. Alonso. State surfaces for partially saturated soils. In *Proceedings of the 11th International Conference on Soil Mechanics and Foundation Engineering,* San Francisco: CRC Press, V. 2: 557-562, 1985.

[30] A. Gens, A. Part 1- Unsaturated Soils, short course, New Zealand Geotechnical Society, https://fl-nzgs-media.s3.amazonaws.com/uploads/2022/06/2010-05-Part-1-Antonio-Gens-unsaturated-soils-3.pdf, 2010.

[31] S.J. Wheeler and V. Sivakumar. "An elastoplastic critical state framework for unsaturated soils.'" *Geotechnique,* 45: 35–53, 1995.

[32] S.J. Wheeler and D. Karube, Constitutive modelling. *Proc. of the First International Conference on Unsaturated Soils (UNSAT)*, Paris, September 1995. Vol. 3, 1996.

[33] P. Delage and J. Graham. Mechanical behaviour of unsaturated soils: understanding the behaviour of unsaturated soils requires reliable conceptual models, *Proceedings of the 1st International Conference on Unsaturated Soils (UNSAT),* Vol. 3: 1223-1256. Paris: Balkema, 1996.

[34] D.Y.F. Ho, D.G. Fredlund, and H. Rahardjo. Volume change indices during loading and unloading of an unsaturated soil. *Canadian Geot. J.*, V. 29: 195-217, 1992.

[35] S.L. Houston and W.N. Houston. Suction-oedometer method for computation of heave and remaining heave, *Proceedings of the 2nd Pan-American Conference on Unsaturated Soils*. Dallas, GSP 300, ASCE: 93-116, 2018. http://dx.doi.org/10.1061/9780784481677.005.

[36] S. Singhal, Expansive soil behavior: property measurement techniques and heave prediction methods, PhD dissertation, Arizona State University, Tempe, AZ (USA), USA, 2010.

[37] A.H. Olaiz. Evaluation of testing methods for suction-volume change of natural clay soils. Doctoral dissertation. Arizona State University, Tempe, AZ (USA), 2017.