Forks in the Road: Rethinking Modeling Decisions that Defined the Teaching and Practice of Geotechnical Engineering

R. Salgado

Purdue University, West Lafayette, IN, USA
rodrigo@ecn.purdue.edu

ABSTRACT: Geotechnical engineering is a discipline that evolved in response to the need to design structures on, in or of soil and rock. It spans a wide range of applications, including tunnels, foundations, dams, and retaining structures. It deals with a material known to be difficult to model: a particulate material whose mechanical response is affected by all three invariants of the stress tensor, by density and by fabric. This paper and the corresponding lecture focus on mechanics-based geotechnical engineering applications. The paper reviews some of the major decisions that were made by the engineers and researchers who developed geotechnical engineering to the point at which it was an identifiable separate discipline and the consequences that these decisions have had on the development of the discipline and on its teaching. The paper identifies some key modeling choices that were made that have had a disproportionate impact on the teaching and practice of geotechnical engineering. The focus of the paper is therefore on these decisions and choices, and what should be taught in their place today.

Keywords: geotechnical engineering, education, sand, clay, Mohr-Coulomb yield criterion, Tresca yield criterion, dilatancy

1 Empiricism, Science and Geotechnical Design

1.1 The Pre-Science Days

Construction in, on or with soil is nothing new: we have been building structures of the most varied types for millennia. One might infer from that that geotechnical engineering, which is the engineering of structures or systems of which soil is an integral part, would be a settled subject. In some ways it is, for geotechnical structures do get built; and these structures are typically designed with methods developed over the course of the last several decades. However, the fact that we can design and construct does not mean that we do these things as well as we could, and it does not mean that the models that we use in analysis and design are correct.

In any type of activity, improved processes and products result from trial and error, but only up to a point. This attempt to arrive at better ways of doing things without a full understanding of the factors at play and their interrelationships is known to us as empiricism, and progress can at times be painful. An interesting twist in how both individuals and populations learn and add to knowledge in an empirical manner resulted from the development of the World Wide Web, the internet and smart search engines. The combination of these three technologies, and the access by a large fraction of the Earth's population to them has given people much more access to knowledge and the possibility of experimenting with knowledge they find online, keeping what works, and discarding what does not. Whereas individuals in their daily lives and people working in the trades have benefited from the rapidly accumulating body of easily accessible specialized knowledge, it is possible to argue that the same is not true of a profession, which geotechnical engineering is. The reason for this is that, at least for the more challenging projects, the engineering profession today must rely on science, and science cannot be found or taught or developed so easily and so loosely. Not to rely on science would take us back a hundred years, when results in terms of economy and safety were far from satisfactory.

There is a common misconception that all engineering done before the advent of science was conservatively done. The inference seems to be common-sensical, because it would be natural to proceed cautiously when one does not know very well what one is doing, i.e., when one is proceeding by trial and error. However, that has not necessarily been so. While cases of serious engineering failures would not have appeared in geotechnical scientific journals – because they did not exist before the second half of the 20th century – we can still learn about how things could go wrong in the pre-science days of geotechnical engineering by referring, for example, to court decisions. An interesting case is that of *Stees v. Leonard*. Here is an excerpt of a pertinent part:

> The action was brought to recover damages for a failure of defendants to erect and complete a building on a lot of plaintiffs, on Minnesota street, between Third and Fourth streets, in the city of St. Paul, which, by an agreement under seal between them and plaintiffs, the defendants had agreed to build, erect, and complete, according to plans and specifications annexed to and made part of the agreement. The defendants *commenced the construction of the building, and had carried it to the height of three stories, when it fell to the ground. The next year, 1869, they began again and carried it to the same height as before, when it again fell to the ground,* whereupon defendants refused to perform the contract.
>
> *Stees v. Leonard,* 20 Minn. 494, 449 (1874). (emphasis added)

There are other cases like this recorded in court proceedings that show the inadequacy of a trial-and-error approach, which lacks a basis on the underlying science. The number of events is most certainly a multiple of those we can learn about from consulting such records. Starting a geotechnical engineering course with a case history like this, and following that with a discussion of the scientific method gives students an appreciation for what the subject is about, its importance, and why science matters.

1.2 The Development of the Science

The scientific method is the formulation of a hypothesis about some question or problem and then the idealization and execution of experiments to validate the hypothesis. If the hypothesis is properly validated, we have a model, which we can then use to guide further scientific inquiry or the development of engineering design methods. Until the early 20th century, all that anyone working with soil and rock could count on was empirical knowledge. It was not until scientists like Forcheimer (whom his student, Karl Terzaghi, later emulated in many respects) started seeking to frame some flow problems as boundary-value problems (Goodman, 1999) that the science of soil mechanics started coming into form. It was a natural step to go from flow problems to consolidation theory, and that development is credited as the birth of soil mechanics. Although Terzaghi's one-dimensional consolidation theory was imperfect – see, e.g., Goodman (1999) and Salgado (2008) for an account of why that is so and of the sad events involving Terzaghi and Forcheimer – its flaws were not fatal to its application to a range of practical problems, and it was by no means a misstep. It will not be discussed further in the present paper.

Once consolidation theory was in place, the same general approach – looking for the science to underpin design methods in the incipient engineering discipline that we now call geotechnical engineering – was followed for other problems. Bearing capacity theory, as an example, follows from work done during the industrial revolution on metal indentation (Prandtl, 1920, 1921; Reissner, 1924). This path was by no means easy, and, faced with hurdles, the pioneers took detours and made decisions that have had significant implications on how geotechnical engineering is practiced and how it is taught at universities even today.

1.3 What Does this Paper Cover?

This paper examines how these difficulties and resulting decisions, many related to how to model the mechanical response of soil, have shaped the development of the discipline and its teaching. Understanding of soil mechanics is vastly superior today. I will propose some ideas regarding key content that should be taught at the undergraduate and graduate level that is consistent with current understanding and that – contrary to opinions sometimes verbalized – is easily learned by students. Due to space limitations, the paper covers only three of the fundamental model choices that shaped soil mechanics and geotechnical engineering, but there are more.

The three topics addressed are the use of the Mohr-Coulomb and Tresca yield criteria to model soil shear strength, the use of an associated flow rule with these models, and the neglect of shear strain localization. These choices have guided the development of the discipline and have led to a significant body of work. Among topics not covered are the reliance of analyses on infinitesimal strains, the neglect of fabric effects on material response, and the use of total-stress undrained analyses in clays.

The paper is very much focused on the content that should be taught, rather than teaching approaches and pedagogy. However, I will also discuss, albeit somewhat superficially, possible approaches to better teaching the right content.

2 The Original Sin: Soil as a Mohr-Coulomb Material and Clay as a "Cohesive" Material

2.1 Background

To understand why, today, students learn that there are two types of soils – "cohesionless soils" and "cohesive soils" – we must travel back to the 1950s, when the science of soil mechanics was in development. After a relatively successful study of 1D consolidation using the coupling of deformation with flow, Terzaghi and co-workers set about dealing with problems involving shear strength, such as the calculation of the bearing capacity of foundations.

The state of the mechanics of foundations at the time was fundamentally this: little progress had been made over the practice of foundation engineering in the preceding century. We discussed earlier the case of *Stees v. Leonard*, in which a contractor tried, not once, but twice, to erect a building on soil that could not support it. In the lawsuit that followed, they misidentified the cause of the problem, which was a bearing capacity problem, as the existence of a "quick sand" at the site. But, even as the understanding that one must design against bearing capacity "failures" – i.e., bearing capacity ultimate limit states – started forming, the means to calculate this bearing capacity lagged behind.

The practice of foundation engineering was to try and build based on prior experience, an experience that was often not applicable to the conditions at hand. In this environment, in which scientific knowledge hardly existed, it is not surprising that Terzaghi believed that "[b]ecause of the unavoidable uncertainties involved in the fundamental assumptions of the theories and in the numerical values of the soil constants, simplicity is of much greater importance than accuracy." (Terzaghi & Peck, 1967 at 153). This thinking permeated much of Terzaghi's work at the time, and it is therefore no surprise that he also believed that "[i]n spite of the apparent simplicity of their general characteristics, the mechanical properties of real sands and clays are so complex that a rigorous mathematical analysis of their behavior is impossible." (Terzaghi, 1943 at 5).

We now know that there are three things that are incorrect in Terzaghi's two statements. First, simplicity and accuracy are not to be directly compared. Something can be both simple and inaccurate, and *vice-versa*. To state that something simple but inaccurate is superior to something not simple but accurate does not appear sensible. Second, the mechanical properties of sand and clay are not even apparently simple. Refer to Figure 1 and Figure 2 for stress-strain plots for sand and clay sheared under drained and undrained conditions in triaxial compression. The stress q in the figures is the Mises shear stress (a multiple of the octahedral shear stress). Without an understanding of the mechanics of these soils, it is impossible to make sense of transitions and reversals between contractive and dilative response, of the existence of a peak shear stress to normal effective stress ratio, of the existence of a critical state, or of the transition to a residual strength at large shear strains and sufficiently large normal effective stresses for clays. Lastly, the final part of Terzaghi's second statement is also (today) incorrect, because researchers are developing fairly rigorous relationships for modeling soil behavior. Monotonic mechanical response is not considered today a challenge to model (see e.g., Chakraborty et al., 2013b; Dafalias & Herrmann, 1986; Li & Dafalias, 2000; Loukidis & Salgado, 2009b; Manzari & Dafalias, 1997; Woo & Salgado, 2015). Figure 1 and Figure 2 show simulations done using an advanced constitutive model that clearly match the experimental response quite well.

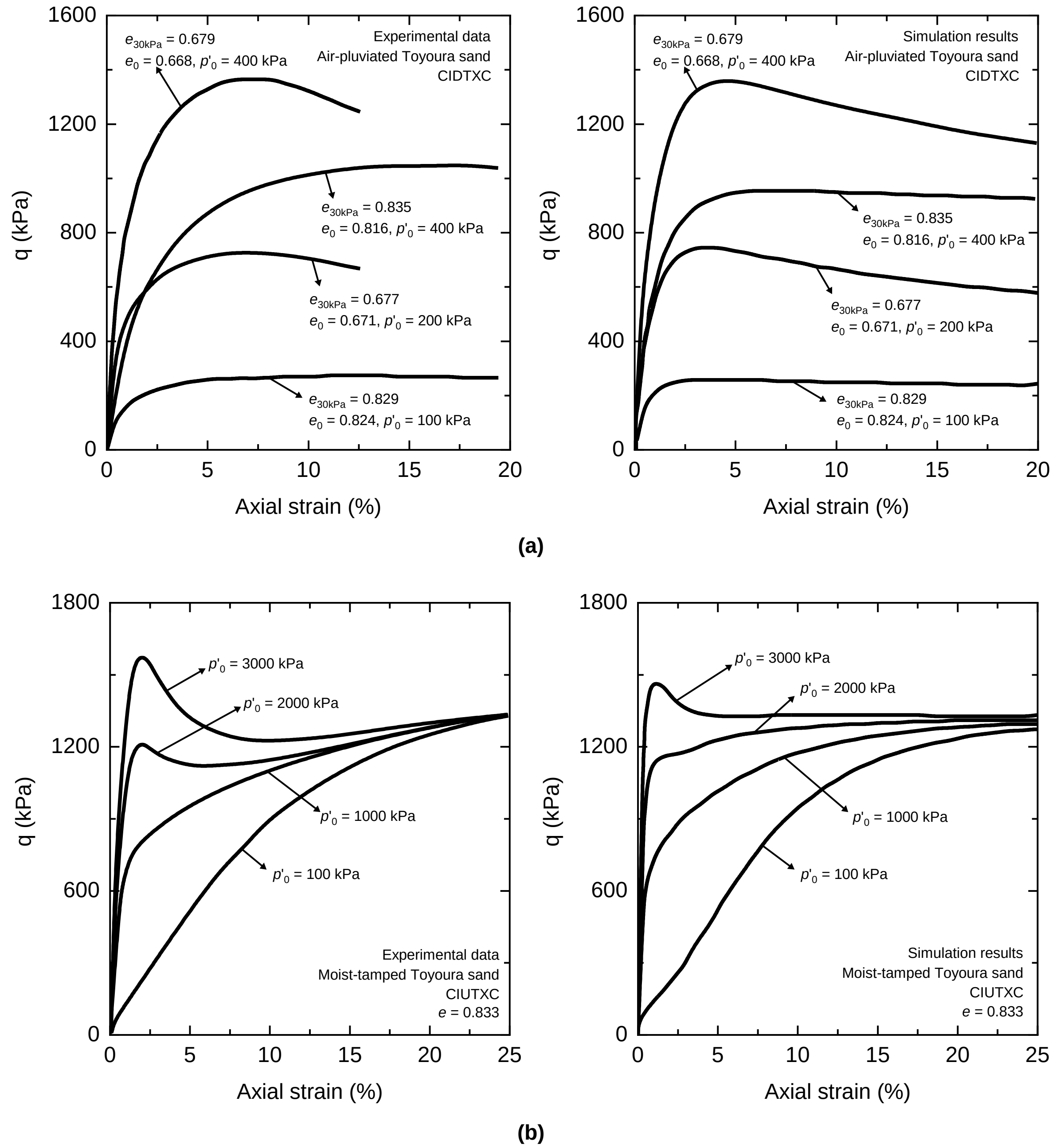
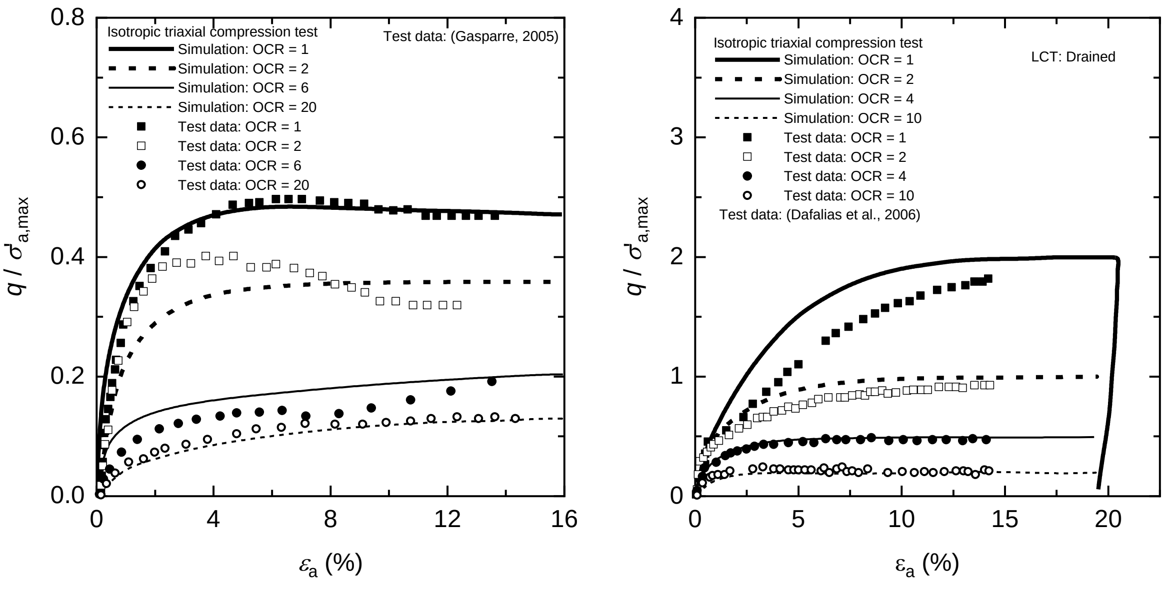


Figure 1. Results of triaxial compression test results performed on sands: (a) drained (b) undrained (Woo & Salgado, 2015)

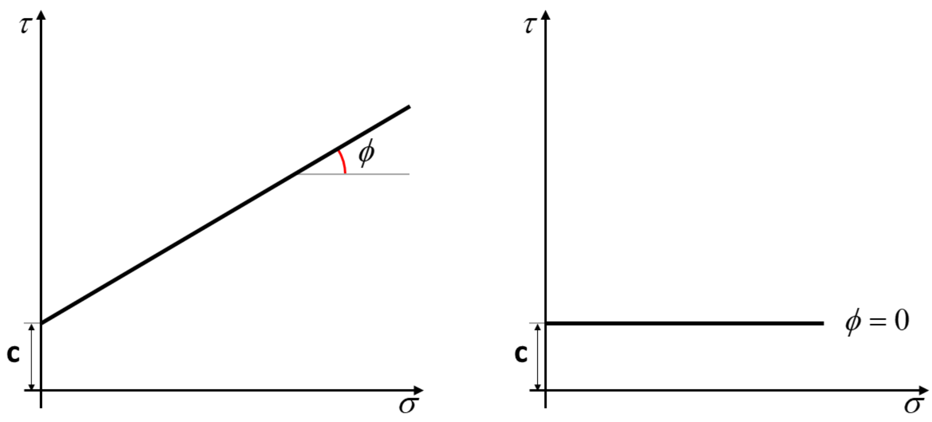
Faced with what he deemed an impossibility, it is not surprising that Terzaghi proposed the concepts of an "ideal sand" – a linear elastic, perfectly plastic Mohr-Coulomb type of material with non-zero friction angle and c = 0 – and an "ideal clay" – a linear elastic, perfectly plastic material following a Tresca yield criterion (Terzaghi, 1943). Terzaghi referred to this material as a "cohesive" material, a term that survives to this day. As to sand, engineers soon started assuming non-zero cohesion also for sand, deviating from the original "ideal sand" concept that Terzaghi had advanced.



(a) (b)

Figure 2. Results of triaxial compression tests performed on clays: (a) undrained (b) drained (Chakraborty et al. 2013b; Dafalias et al., 2006; Gasparre, 2005)

So a Mohr-Coulomb yield criterion (Figure 3a) would be used for sand, and a Tresca yield criterion (Figure 3b) would be used for clay. The only way to understand this postulation is to assume that Terzaghi observed increasing strengths for sand tested at increasing confining stresses under drained conditions, but constant strength for clay with increasing *total stresses* when samples were tested under undrained conditions. Based on this limited set of observations, Terzaghi postulated behaviors for soil that are not real. To this, Schofield (1988) later referred as "Terzaghi's error." This criticism is tempered by the recognition that the "ideal clay" model turned out to be an effective basis to build a body of analysis for problems involving saturated clay and that even erroneous models of soil behavior were better than the crude form of knowledge available in those days. Additionally, the greater harm concerning sands was the subsequent use of a Mohr-Coulomb material with nonzero cohesion for sand, rather than the original ideal sand concept. Consequently, some viable theories have evolved from these simple "ideal" soil models, but the failure to accurately describe the sources of shear strength in soils remained.



(a) (b)

Figure 3. Relationship between normal and shear stresses for (a) a Mohr-Coulomb material, idealized in the 1950s as an "ideal sand" if c= 0 and (b) a Tresca material, idealized in the 1950s as an "ideal clay"

2.2 The Error

We have argued that Terzaghi's "ideal sand" and "ideal clay" models led to an erroneous description of soil behaviour. This is true even if one is simply interested in calculating shear strengths and has no interest in realistically simulating any other aspects of behaviour. But why is it so? For the answer, we look to plasticity theory.

Perhaps nothing has been as damaging to the teaching of soil mechanics than the notion that soil can generally be considered to follow the Mohr-Coulomb yield criterion. A material that follows the Mohr-Coulomb yield criterion experiences plastic strains only when the stress state satisfies the relationship:

$$F = (\sigma_1 - \sigma_3) - (\sigma_1 + \sigma_3)\sin\phi - 2c\cos\phi = 0 \tag{1}$$

where σ_1 and σ_3 are the maximum and minimum principal stresses, respectively. The function F of stresses is referred to as the yield function, and $F = 0$ is referred to as the yield criterion. The parameters ϕ and c are the friction angle and the cohesion, respectively, of the material. Terzaghi's ideal sand has non-zero ϕ and $c = 0$, and the ideal clay has zero ϕ. In later work, engineers abandoned the original concept of zero c in sand and started using nonzero c and ϕ to describe sand. No explanation was provided for the source of what should amount to a frictional strength component and a stress-independent (frictionless or cohesive) strength component. What this step left both educators and practitioners with was a model that was not based on an understanding of soil behaviour, since ϕ and c were the starting point of the analysis: the model fundamental parameters.

Unfortunate implications of this paradigm were the misunderstanding that clean, uncemented sands could have non-zero c, and that clays had a constant c, a result directly implied by the "ideal clay" model. Initially, educators taught students that a set of tests had to be done at more or less the "appropriate" level of effective stresses, and straight-line fits to the corresponding data points would yield the correct values of ϕ and c. This presented a variety of questions, one of which regarded the applicable level of effective stress for a problem in which the soil experiences a wide range of stress levels, as in the bearing capacity problem. In some of these problems, stresses can be as high as several or even tens of megapascals. Clearly, performing shear strength tests at these elevated stress levels was not realistic.

Fortunately, even as Terzaghi made these influential choices, others (e.g., Taylor, 1948) were attempting to understand what the real sources of shear strength were. Taylor laid the foundation for what would later be known as critical-state soil mechanics. In this framework for the mechanics of soil, soil is a frictional material capable of volume change; a second source of shear strength results from this dilative response.

2.3 What Should Be Taught Instead

What emanated from the studies of Roscoe et al. (1958), Schofield (2006), Taylor (1948) and others was the understanding that soil is always a frictional material. In the absence of cementation, a fully saturated or completely dry soil derives its strength exclusively from friction if under sufficiently high confining stress and/or sufficiently low density (see, e.g., Salgado, 2008). If either density is sufficiently high or stress is sufficiently low, soil also derives its strength from dilatancy.

It follows that, whether teaching at the graduate or undergraduate level, we should teach our students that soil takes its strength from two sources: friction and dilatancy. It is essential to stress that unstructured soil (soil without cementation or any source of extraneous cohesion) is frictional, lacking cohesion. A good starting point for this discussion is plastic deformation in the absence of any tendency to change volume: the so-called critical state. Surprisingly, based on anecdotal evidence, this is a concept that undergraduate students are often not exposed to. The concept is however easy to teach. The easiest way to teach it is to show students that the critical state is simply a purely frictional state. At critical state, the soil derives its strength from the frictional strength between soil particles, there being no other source of shear strength. And frictional strength only exists in the presence of non-zero effective normal stress.

It is sometimes surprising to students who have somehow learned otherwise that even clays are purely frictional materials. An example that can be used to get this last point across is that of a clay deposit forming at the bottom of a lake (Salgado, 2008). It is easy for students to understand that the soil right at the surface of the bottom of the lake, composed of particles that have recently deposited out of water, lacks shear strength. The reason for that is that the clay there is almost a slurry: it is under zero effective stress and has very high void ratio. In the absence of a normal effective stress, that clay has zero shear strength because it is a frictional material. An example for sand that can be given, to which undergraduate students can easily relate, is that someone picking up some sand on the beach can easily manipulate the soil, for it lacks strength, and it lacks strength because it is under nearly zero normal effective stress.

The other component of shear strength is due to dilatancy, which can best be explained by referring to a figure such as Figure 4, which shows that spherical particles that are closely packed must separate in the direction normal to that of shearing. This separation must occur against an existing normal effective stress, which requires work to be done. Where does the work come from? From the applied shear stress. So the applied shear stress must overcome not only frictional strength to cause the material to deform plastically, but also this confining stress opposing the required soil dilation.

These two concepts are easy for students to understand. This basic understanding of the physical processes underlying shear strength development in soil can then be used throughout their course of study of geotechnical engineering applications (retaining structures, foundations, slopes and other structures), and should effectively inoculate them against the flawed concepts of "cohesive" or "cohesive-frictional" soils. From that point on, students will understand that soils are truly potentially dilative, frictional materials.

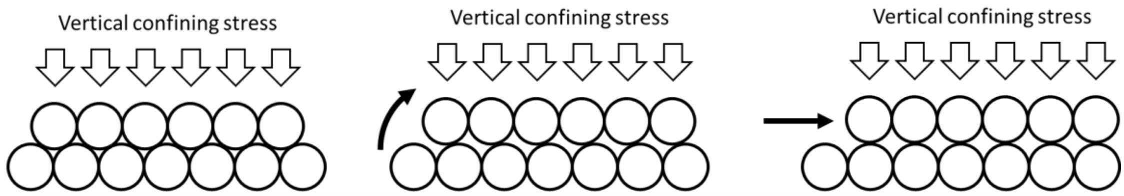


Figure 4. Particle climbing action for densely arranged particles (Salgado, 2008)

At the undergraduate level, one of the easiest ways to teach how dilatancy works is to use the Bolton (1986) framework for sands. This work has been extensively referred to and has been extended to apply to sands with fines (see, e.g., Carraro et al., 2009; Salgado et al., 2000) and sands at low confining stresses (Chakraborty & Salgado, 2010). Concisely, for a sand, the peak friction angle ϕ_p is written as the summation of a critical-state friction angle ϕ_c plus an angle due to dilatancy:

$$\phi_p = \phi_c + A_\psi I_R \qquad (2)$$

where A_Ψ is a parameter in Bolton's equation having value of 3 for triaxial conditions and 5 for plain-strain conditions, and I_R is the relative dilatancy index given by:

$$I_R = I_D(Q - \ln p') - R \qquad (3)$$

where I_D is relative density, p' is the mean effective stress and *Q and R* are fitting parameters.

At the Ph.D. level, one must go much beyond this. It is important then to cover constitutive modeling (mainly the most recent models, such as bounding-surface models) and particle-based methods.

3 Building on the Original Sin: Reliance on the Associated Flow Rule

3.1 Background

The teaching of geotechnical engineering tends to emphasize stresses, but strains are just as much a part of the solution to any boundary-value problems in geomechanics. The only exposure that students seem to get to strains is that stress-strain plots are typically shown or obtained in the laboratory and during the coverage of consolidation. A standard discussion surrounds the facts that loose sands contract or dilate less than dense sands and that dense sands may contract initially, but then end up being ultimately dilative. Strains are typically not linked back to stresses with any rigor, and that is sometimes true even at the graduate level. Yet, this link is crucial to the modeling of the mechanical response of soil.

The relationship is rather obvious to students in the context of elasticity. There is a general sense that application of a stress increment leads to a strain increment, and that its removal returns the body to its original configuration. When it comes to plasticity, matters turn more complex.

The rate of the plastic strain tensor in classical plasticity models is obtained from the plastic flow rule:

$$\dot{\varepsilon}_{ij}^{p} = \dot{\lambda} \frac{\partial G}{\partial \sigma_{ij}} \tag{4}$$

where *i* and *j* are indices taking values 1, 2 or 3; σ_{ij} are the six components of the (symmetric) stress tensor; $\dot{\lambda}$ is the plastic multiplier; and *G* is the plastic potential, a function of the stress tensor:

$$G = G(\sigma) \tag{5}$$

Given that there are six independent stress components, Equation (4) states that the plastic strain increments, or rates are determined by a six-dimensional surface defined by Equation (5). The meaning of the term $\partial G/\partial \sigma_{ij}$ is that of a gradient in that space. This can best be visualized if we represent the stress tensor using its three principal stresses, in which case we are able to represent these equations in 3-dimensional space (see Figure 5). The gradient can then be visualized as being normal to the 3-dimensional surface defined by Equation (5). This visualization of a 6-dimensional process in 3-dimensional space can only be taken so far, as discussed by Woo & Salgado (2014).

If the gradient is aligned with the σ_1 axis, for example, that means that only the ε_1 strain component will change, with $\dot{\varepsilon}_2 = \dot{\varepsilon}_3 = 0$. So $\partial G/\partial \sigma_{ij}$ determines the proportion or ratio between each pair of strain rate components.

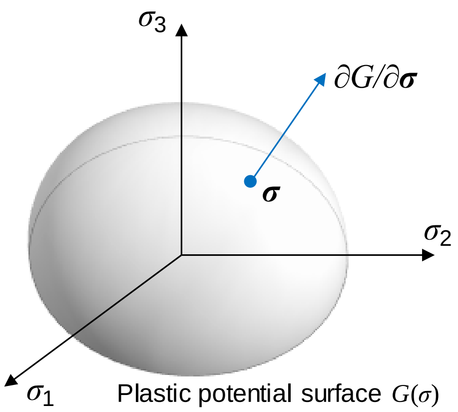


Figure 5. Plastic potential surface represented in principal stress space and its stress gradient, which enters the formualtions of the flow rule

In metal plasticity, which developed considerably during the industrial revolution, it was observed that there was no plastic volume change during plastic deformation. Although we don't show this here, this leads to the result that plastic strain rate is normal to the yield surface given by Equation (1) if plastic strains are plotted in the same space (with a separate scale) as stresses. This led to the adoption of what we now call an associated flow rule for the plastic strain rate, where *F* is used as the plastic potential:

$$\dot{\varepsilon}_{ij}^{p} = \dot{\lambda} \frac{\partial F}{\partial \sigma_{ij}} \tag{6}$$

If we are working with clays using total stresses in undrained loading simulations, we are in effect using Terzaghi's "ideal clay" model. There is then no volumetric strain, and Equation (6) is approximately applicable. In drained simulations or effective-stress simulations, an associated flow rule does not apply. This can be observed by performing experiments and observing the lack of normality between the plastic strain rate and the yield surface. However, it is important to understand what the fundamental error of use of an associated flow rule is in those cases.

3.2 The Error

A material undergoing plastic deformation (yielding), in contrast with only elastic deformation, dissipates energy. We can think of energy dissipation as the energy that has to be expended to change the material internally (i.e., to permanently deform it in some manner). The rate of plastic energy dissipation D_p per unit volume for infinitesimal-strain plasticity is given by:

$$D_p = \sigma_{ij}\dot{\varepsilon}_{ij}^{p} \tag{7}$$

where σ_{ij} is the stress, and $\dot{\varepsilon}_{ij}^{p}$ is the time rate of plastic strain.

Taking Equation (1) and Equation (6) into Equation (7), we obtain the following for the rate of plastic dissipation:

$$D_p = \dot{\lambda}[2c\cos\phi] \tag{8}$$

What Equation (8) tells us is that the rate of plastic energy dissipation is entirely due to the existence of a cohesion c. If c = 0, then no energy is dissipated during plastic flow. If we think of a sand in realistic terms, it has no cohesion. So Equation (8) is telling us that sand does not dissipate energy, which we know to be incorrect. This result is also baffling for the typical graduate student. How can a cohesive-frictional material, for that is what a Mohr-Coulomb material is supposed to be, dissipate no energy upon plastic deformation when c = 0? Is friction not intricately linked to energy dissipation?

The inescapable conclusion is that the use of the Mohr-Coulomb yield criterion with an associated flow rule to model real soils in effective-stress analysis is simply wrong. A sand loaded under drained conditions, which corresponds to the vast majority of applications involving sands, and is correspondingly taught quite often, cannot be modeled with a Mohr-Coulomb model even as an approximation, unless a flow rule that is not associated is used. Unfortunately, drained analysis with a Mohr-Coulomb material and an associated flow rule is what a large body of work in geotechnical engineering is based on. This is the content that many, if not most, geotechnical engineering students get in the classroom.

3.3 What Must Be Taught Instead

If one must use the Mohr-Coulomb model, it is important not to teach any of the theories in which an associated flow rule was assumed and, where needed, stress that the flow rule for a Mohr-Coulomb material cannot be associated if realism is to be achieved. This difference is far from just conceptual, with important numerical consequences.

Consider, for example, the bearing capacity problem in sand. The unit bearing capacity q_{bL} in sand can be seen as the summation of two terms:

$$q_{bL} = q_0 N_q + \frac{1}{2}\gamma B N_\gamma \tag{9}$$

where q_0 = overburden stress, γ = unit weight, and N_q and N_γ are bearing capacity factors. We ignore any depth correction factor that might be incorporated into Equation (9) for the purposes of the discussion that follows. The classical equations for the two bearing capacity factors are:

$$N_\gamma = 1.5\left(N_q - 1\right)\tan\phi \tag{10}$$

and

$$N_q = \frac{1+\sin\phi}{1-\sin\phi}e^{\pi\tan\phi} \tag{11}$$

Equation (10) is due to Brinch Hansen (1970), who proposed it based on results from the method of characteristics. The method of characteristics assumes an associated flow rule, as does most of the work published using limit analysis. We now know that these two equations cannot be correct, for sand does not follow an associated flow rule. How wrong are the results? We can answer this by referring to the equations proposed by Loukidis & Salgado (2009a) for a sand with a non-associated flow rule:

$$N_q = \frac{1+\sin\phi}{1-\sin\phi} e^{J(\phi,\psi)\pi\tan\phi} \tag{12}$$

and

$$N_\gamma = \left(N_q - 1\right)\tan(1.34\phi) \tag{13}$$

where J is a function given by

$$J(\phi,\psi) = 1 - \tan\phi\left[\tan(0.8(\phi-\psi))\right]^{2.5} \tag{14}$$

and ψ is the dilatancy angle.

The dilatancy angle is defined as:

$$\sin\psi = -\frac{\dot{\varepsilon}_v}{\dot{\gamma}_{max}} \tag{15}$$

where $\dot{\varepsilon}_v$ is the time rate of volumetric strain and $\dot{\gamma}_{max}$ is the rate of the maximum shear strain.

The dilatancy angle is a measure of how much volumetric strain results from shearing of the material. A flow rule associated with the Mohr-Coulomb yield function leads to $\psi = \phi$. It is more realistic for sands to assume $\psi < \phi$. This would correspond to a non-associated flow rule. Figure 6 illustrates the impact that the choice of an associated instead of a non-associated flow rule has on engineering computations related to the bearing capacity problem. The figure shows value of N_γ resulting from realistic pairings of ψ and ϕ and from $\psi = \phi$. Values for $\psi = \phi$ significantly exceed value for $\psi < \phi$.

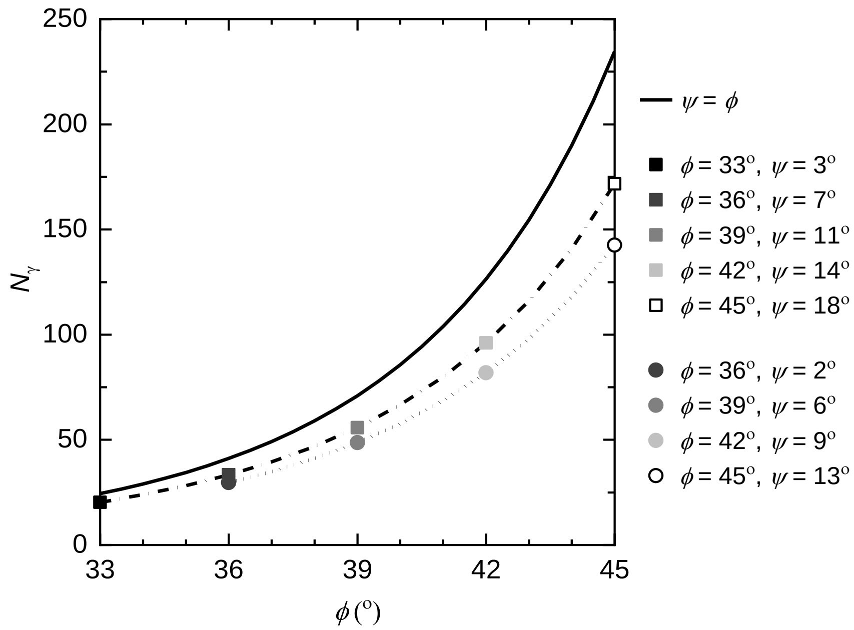


Figure 6. Comparison of values of bearing capacity factor N_γ calculated based on the assumption of associated flow ($\psi = \phi$) with values calculated based on non-associated flow ($\psi < \phi$)

How much difference does the choice of flow rule make in the calculation of the bearing capacity of a footing? Let us consider the bearing capacity factors and the limit bearing capacity q_{bL} of strip footings calculated using the two sets of equations. As an example, we take a friction angle ϕ = 45°; dilatancy angle ψ = 45° and 18°; and unit weight of sand = 19 kN/m^3. Table 1 presents the computed bearing capacity factors – N_γ and N_q – and the bearing capacity q_{bL} of two strip footings with width B = 1 m and 2 m, with an embedment of 0 m and 1 m, with the depth factor on the overburden term of the bearing capacity equation neglected.

Table 1. Effect of flow rule non-associativity on bearing capacity of strip footings: results of calculations using Equations (12) and (13)

Flow rule	ϕ (°)	ψ (°)	N_q	N_γ	q_{bL} (kN/m^2)			
					embedment = 0 m		embedment = 1 m	
					B = 1 m	B = 2 m	B = 1 m	B = 2 m
Associated	45	45	135	235	2230	4459	4792	7022
Non-associated		18	99	172	1631	3262	3511	5142

The resulting bearing capacity for footing on the surface of a deposit of the material following the associated flow rule is 37% greater than that calculated for a material following the non-associated flow rule. This very significant overestimation of the bearing capacity of a strip footing resulting from use of the associated flow rule is an error that is unconservative. Given the nature of shallow foundation design, with serviceability controlling in the vast majority of design cases, this error does not have the detrimental impact that it otherwise would.

This simple example, for one of the classical problems of soil mechanics, illustrates the level of error resulting from use of theories based on a Mohr-Coulomb material following an associated flow rule. Ideally, these would not be taught but for providing historical perspective. The teaching of methods of analysis and design that rely on realistic soil models would be the best approach, and it is possible in many instances. Failing that, whenever the Mohr-Coulomb yield criterion is used, it must be used with a non-associated flow rule.

Lastly, use of a non-associated flow rule does not heal the defects of a model relying on the Mohr-Coulomb yield criterion. The model is still exceedingly simple – having constant ϕ and ψ – and will not be realistic for calculations requiring a higher degree of realism. In such cases, use of a more sophisticated constitutive model is required.

4 Shear Strain Localization and its Implications

4.1 Background

In undergraduate laboratory classes, students typically see or perform triaxial tests on dense sand specimens; they observe the resulting "failure plane" that eventually develops through the specimen. In most classrooms, that observation leads to nothing more, but it should. That is the best time to make a number of crucial points that are today essential for a well-rounded geotechnical engineer to understand.

The first important point regarding that "failure plane" is that it is not a plane at all. The second is that "failure" is too vague a term, and it confuses students to use it. It is better to speak of what has happened as the shearing of the sand or, if one is especially attached to the word, as a shear "failure" of the sand specimen. Back to the first point, today it is possible to show students videos taken of the shearing of sand. In these videos, we can clearly see that a band of particles, with thickness of the order of 5 to as many as 20 particle diameters, is what constitutes that "plane." The "plane" is what we know today as a shear band.

Shear bands in soil have been studied as early as the 1970s (Vardoulakis et al., 1978). It is however very important to teach students this for the following reason: a plane is an abstraction from which no pattern of soil behaviour can be inferred, but a band, containing a number of soil particles, has a behaviour that results from the interactions of the particles in it. This interaction of particles in the band directly produces the constitutive behaviour of the soil. Once students understand this, it is much easier for them to understand how shaft resistance develops along a pile or why the pressure on a retaining wall is what it is.

The localization of shearing in a band results from the mechanical behaviour of soil: from the softening, i.e., loss of shear strength that occurs with the progression of shearing. With continuing shearing, the soil will tend to weaken at the location where this process first starts, shear strain then localizes there, sparing regions surrounding the band of further deformation. It is vital to understand this process because any simulations that we attempt of boundary-value problems involving such materials depends on correctly capturing the width of the shear bands. Mechanicians speak of the "length scale" of the material as determinative or intrinsically linked to the material behavior.

Shear bands are also seen in a Mohr-Coulomb soil following a non-associated flow rule. This is closely linked to the fact that, in these materials, plastic energy does dissipate – due to friction – once plastic shearing starts. It is then natural for shearing to continue where it started instead of diffusing to surrounding regions because that would require greater energy expenditure because of the plastic energy dissipation requirement.

Shear band thickness depends on essentially two things: (a) soil particle size and (b) whether it forms entirely within the soil or at an interface with a structural element. If the interface is rough, the shear band thickness will be of the order of the thickness that forms entirely within soil; however, if the interface is smooth, there is no shear band that forms along the interface: there is only clean sliding of the interface with respect to the soil (Tehrani et al., 2016; Tovar-Valencia et al., 2018). Images of strain localization can be collected through an exposed (transparent) window that allows visualization of soil during loading or, for small specimens, through X-Ray CT (e.g., Desrues et al., 2018). In approximate terms, shear bands in sand are of the order of 5 times the mean particle size for rough interfaces (Tehrani et al., 2016; Tovar-Valencia et al., 2018) to the order of 10 times the mean particle size for shear bands entirely contained in soil (Alshibli & Sture, 1999).

The simplest examples of localization and its impact on the solution of a boundary-value problem can be seen in the context of axially loaded piles, for which localization is known *a priori* to occur along the pile shaft (Han et al., 2017, 2018; Loukidis & Salgado, 2008; Salgado et al., 2017). Figure 7 shows the results of finite element analyses of an axially loaded pile in sand modelled using an advanced constitutive model in terms of the ratio K of the lateral effective stress on the pile shaft to the initial (free-field) vertical effective stress during shearing (Loukidis & Salgado, 2009b). It is seen in the figure that the shaft resistance calculated for a pile depends on the width of the finite elements used immediately next to the pile. As the finite element simulation progresses, shear strain localizes next to the pile in that "column" of elements. Consequently, the shear stress along the pile shaft at any given level of pile settlement depends on the response of that band of soil and how it responds to shearing. Pre-knowledge of what the shear band thickness is in a soil allows the correct calculation of the shaft resistance of the pile. The alternative is more difficult: use of a constitutive model and computational method that inherently have the correct length scale so that the correct final shear band pattern and thickness will result.

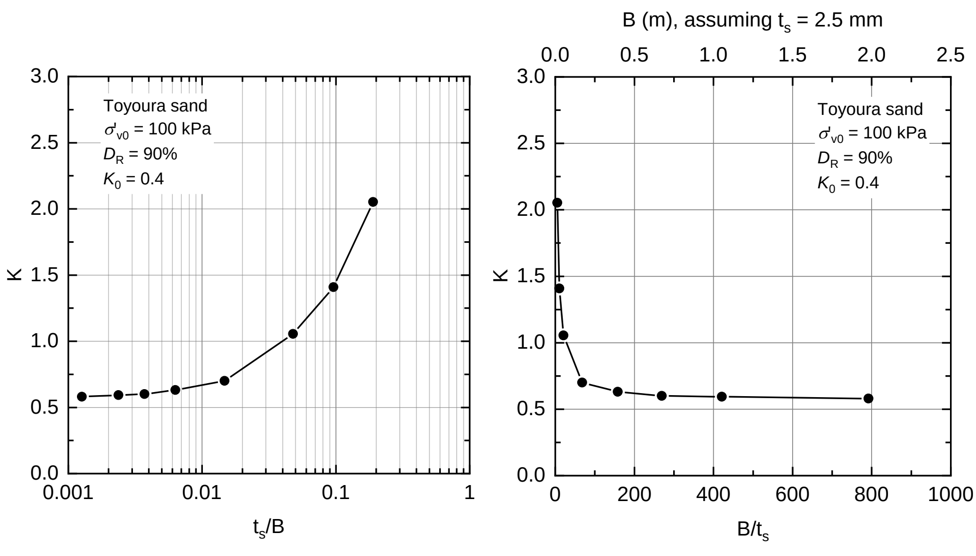


Figure 7. **Effect of ratio of shear band thickness t_s to pile diameter B on shaft resistance (Salgado et al., 2017): (a) K vs. t_s/B and (b) K vs. B/t_s**

4.2 The Shortcoming of Not Considering Shear Strain Localization

Students are often inundated with coverage of "elastic soil" or elasto-plastic soil following the Mohr-Coulomb or Tresca yield criteria. These are often observed in naïve use of commercial finite element software. An interesting illustration of how analyses using either an elastic soil model or an elasto-plastic soil model without realistic strength representation, strain softening and strain localization fall short comes again from foundation engineering.

Traditional models of pile group interaction relied on soil as an elastic material that transferred stresses between piles in a pile group (Poulos, 1968; Randolph & Wroth, 1979). This led to pile interaction and group efficiency coefficients that are unrealistic because the models did not account for strain localization, which significantly reduces interaction between neighbouring piles (Han et al., 2019). Figure 8 shows the significant difference in pile interaction within a group and group efficiency resulting from finite element analyses assuming a linear elastic soil, an elasto-plastic soil with a Mohr-Coulomb yield criterion, and a realistic sand model with an appropriately fine finite element mesh. These results show clearly that shear strain localization cannot be ignored if we desire accurate, realistic solutions to geotechnical boundary-value problems.

As a final illustration of the importance of capturing shear strain localization correctly, consider again the bearing capacity problem discussed earlier. Assume that a student or engineer decides to use a modern method of analysis or a commercial computational package to perform calculations for the same problem we discussed earlier. Table 2 shows results for calculations using SNAC (Abbo & Sloan, 2000), OptumG2 (Krabbenhoft et al., 2015) and the material point method (MPM) (Bisht & Salgado, 2018; Woo & Salgado, 2018). The values shown in the table are in reasonable agreement because consistent size for the mesh elements were chosen in these calculations. The SNAC and OptumG2 analyses were done using 15-node triangles with 12-point Gauss quadrature. The MPM analyses were done using Q4 elements with an initial number of material points per element equal to 4 and a B-bar scheme. The MPM analysis with the smallest element size e = 0.025m has approximately the same Gauss point density as the SNAC analysis, and the match between the two is evident. However, use of a coarser mesh, whether in SNAC, OPTUM or MPM would produce higher values of bearing capacity. For example, in the table, MPM with the smallest element size e = 0.1m yields a bearing capacity of 3055 kPa instead of 2241 kPa. This results from the fact that strain localization can only take place to the degree that the mass is discretized. A coarse mesh will lead to thick shear bands and a stiffer response.

Table 2. Strip footing bearing capacity computed using different numerical schemes and element sizes e

Flow rule	ϕ (°)	ψ (°)	q_{bL} (kN/m^2) (embedment = 0 m, B = 1 m)				
			SNAC	OptumG2	MPM		
					e = 0.1 m	e = 0.05 m	e = 0.025 m
Associated	45	45	2230	2307	3055	2301	2241
Non-associated	45	18	1631	1646	1924	1650	1611

4.3 What Should Be Taught Instead

Students should be acquainted with realistic stress-strain relationships under various loading paths, both drained and undrained, and should be provided with the opportunity to understand the role density, initial effective stress, dilatancy, and fabric evolution have in shaping these relationships. When exposed to problems in which shear strain localization occurs, and therefore the stress-strain history before localization is determinative of soil response, it is important to explain this and provide students with solutions and design methods based on analyses that do take localization into consideration.

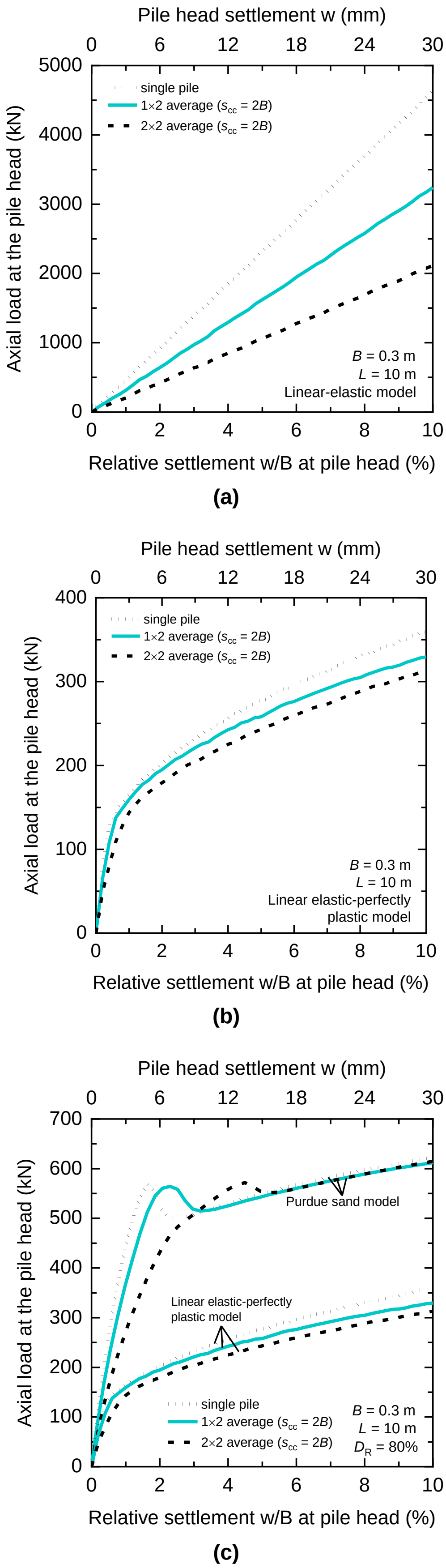


Figure 8. Load-settlement curves obtained from analyses using: (a) a linear-elastic model; (b) a linearly elastic-perfectly plastic model with the Mohr-Coulomb yield criterion; and (c) the Purdue sand model and the linearly elastic-perfectly plastic model with the Mohr-Coulomb yield criterion (Han et al., 2019)

Taking piles again as an example, teaching an analysis that ignores the shear strain localization along the pile shaft will be ineffective in that the value of pile shaft resistance cannot be calculated with any accuracy using such an analysis. Thus, one could teach using directly the results of analysis for piles in sand (e.g., Han et al., 2017; Loukidis & Salgado, 2008) or clay (e.g., Basu et al., 2014; Chakraborty et al. 2013a) that do account for localization and realistic soil response. For undergraduates, the teaching might consist of presenting the equations, explaining why they were formulated with those particular forms, and then having the students apply the equations directly to design problems. At the graduate level, one could go beyond that, and ask the student to read the papers, reproduce results and apply them to more challenging design problems.

As a final illustration of how strain localization can be included in our teaching, we turn again to the pile group example. It is advantageous to introduce students to these problems using the classical papers assuming linear elastic soil (Poulos, 1968; Randolph & Wroth, 1979), which facilitate understanding of the concepts of group pile interaction and group efficiency, but then share with them new results (Han et al., 2019; Salgado et al., 2017) that show that the interaction between the piles is considerably reduced when shear strains localize along the shafts of the piles.

5 Conclusions

The pioneers of soil mechanics faced some difficult choices. Faced with hard challenges and limited knowledge, they made some decisions on how to model soil and analyze the boundary-value problems of soil mechanics that have had a significant impact on how the discipline and its teaching evolved.

The three choices that were made that are highlighted in the paper are the use of Terzaghi's "ideal sand" and "ideal clay" models, the use of an associated flow rule with these models, and the neglect of shear strain localization in the solution of boundary-value problems. These choices led to some confusion regarding how soil responds to load, left engineers at a loss as to how to estimate shear strength parameters, and produced solutions to core problems in soil mechanics – such as the bearing capacity problems, the axial loading of a pile or the response of pile groups – that are either incorrect or unrealistic.

The discipline has overcome these initial modeling choices, and there are now better models and better theories for modeling both soil – the material – and the various engineering problems of interest. These better approaches need to be included in textbooks and shared with the community. With the right way of presenting these newer theories, it is possible to teach them to undergraduate, as well as graduate students.

Acknowledgments

I appreciate the comments and suggestions received from Drs. Marina Pantazidou, Monica Prezzi, Kenichi Soga and Fei Han. I have illustrated the main themes explored in connection with the teaching of geotechnical engineering using the results of research that I have done with colleagues – including students and former students. I also acknowledge the assistance of Rameez Raja, Eshan Ganju and Vibhav Bisht with certain details of the paper.

References

Abbo, A. J., Sloan, S. W. (2000). SNAC, User manual, Version 2.0. Department of Civil, Surveying and Environmental Engineering, University of Newcastle, Callaghan, Australia.

Alshibli, K. A., Sture, S. (1999). Sand Shear Band Thickness Measurements by Digital Imaging Techniques. Journal of Computing in Civil Engineering, 13(2), 103–109.

Basu, P., Prezzi, M., Salgado, R., Chakraborty, T. (2014). Shaft Resistance and Setup Factors for Piles Jacked in Clay. Journal of Geotechnical and Geoenvironmental Engineering, 140(3), 04013026.

Bisht, V., Salgado, R. (2018). Local transmitting boundaries for the generalized interpolation material point method (GIMP). International Journal for Numerical Methods in Engineering, 114(11), 1228–1244.

Bolton, M. D. (1986). The strength and dilatancy of sands. Géotechnique, 36(1), 65–78.

Brinch Hansen, J. (1970). A revised and extended formula for bearing capacity. Danish Geotechn Inst Bull, 28, 5–11.

Carraro, J. A. H., Prezzi, M., Salgado, R. (2009). Shear strength and stiffness of sands containing plastic or nonplastic fines. Journal of Geotechnical and Geoenvironmental Engineering, 135(9), 1167–1178.

Chakraborty, T., Salgado, R. (2010). Dilatancy and shear strength of sand at low confining pressures. Journal of Geotechnical and Geoenvironmental Engineering, 136(3), 527–532.

Chakraborty, T., Salgado, R., Basu, P., Prezzi, M. (2013a). Shaft Resistance of Drilled Shafts in Clay. Journal of Geotechnical and Geoenvironmental Engineering, 139(4), 548–563.

Chakraborty, T., Salgado, R., Loukidis, D. (2013b). A two-surface plasticity model for clay. Computers and Geotechnics, 49, 170–190.

Dafalias, Y. F., Herrmann, L. R. (1986). Bounding Surface Plasticity. II: Application to Isotropic Cohesive Soils. Journal of Engineering Mechanics, 112(12), 1263–1291.

Dafalias, Y. F., Manzari, M. T., Papadimitriou, A. G. (2006). SANICLAY: Simple anisotropic clay plasticity model. International Journal for Numerical and Analytical Methods in Geomechanics, 30(12), 1231–1257.

Desrues, J., Andò, E., Mevoli, F. A., Debove, L., Viggiani, G. (2018). How does strain localise in standard triaxial tests on sand: Revisiting the mechanism 20 years on. Mechanics Research Communications, Elsevier Ltd, 92, 142–146.

Gasparre, A. (2005). Advanced laboratory characterization of London Clay. Imperial College London, Department of Civil and Environmental Engineering, London, UK.

Goodman, R. (1999). Karl Terzaghi: the engineer as artist. ASCE Press.

Han, F., Salgado, R., Prezzi, M. (2018). Numerical and experimental study of axially loaded non-displacement piles in sand. International Conference on Deep Foundations and Ground Improvement, NJ: Deep Foundations Institute, Hawthorne, New Jearsey, 221–229.

Han, F., Salgado, R., Prezzi, M., Lim, J. (2017). Shaft and base resistance of non-displacement piles in sand. Computers and Geotechnics, Elsevier Ltd, 83(83), 184–197.

Han, F., Salgado, R., Prezzi, M., Lim, J. (2019). Axial Resistance of Nondisplacement Pile Groups in Sand. Journal of Geotechnical and Geoenvironmental Engineering, 145(7), 04019027.

Krabbenhoft, K., Lyamin, A., Krabbenhoft, J. (2015). Optum computational engineering (OptumG2). <www.optumce.com> (Jun. 6, 2020).

Li, X. S., Dafalias, Y. F. (2000). Dilatancy for cohesionless soils. Géotechnique, 50(4), 449–460.

Loukidis, D., Salgado, R. (2008). Analysis of the shaft resistance of non-displacement piles in sand. Géotechnique, 58(4), 283–296.

Loukidis, D., Salgado, R. (2009a). Bearing capacity of strip and circular footings in sand using finite elements. Computers and Geotechnics, 36(5), 871–879.

Loukidis, D., Salgado, R. (2009b). Modeling sand response using two-surface plasticity. Computers and Geotechnics, 36(4), 166–186.

Manzari, M. T., Dafalias, Y. F. (1997). A critical state two-surface plasticity model for sands. Géotechnique, 47(2), 255–272.

Poulos, H. G. (1968). Analysis of the Settlement of Pile Groups. Géotechnique, 18(4), 449–471.

Prandtl, L. (1920). Über die Härte plastischer Körper. Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse, 74–85.

Prandtl, L. (1921). Hauptaufsätze: Über die Eindringungsfestigkeit (Härte) plastischer Baustoffe und die Festigkeit von Schneiden. ZAMM - Journal of Applied Mathematics and Mechanics / Zeitschrift für Angewandte Mathematik und Mechanik, 1(1), 15–20.

Randolph, M. F., Wroth, C. P. (1979). An analysis of the vertical deformation of pile groups. Géotechnique, 29(4), 423–439.

Reissner, H. (1924). Zum Erddruckproblem. Proceedings of the 1st International Congress for Applied Mechanics, C. B. Biezeno and J. M. Burgers, eds., Delft, The Netherlands, 295–311.

Roscoe, K. H., Schofield, A. N., Wroth, C. P. (1958). On the yielding of soils. Géotechnique, 8(1), 22–53.

Salgado, R. (2008). The engineering of foundations. McGraw-Hill, New York.

Salgado, R., Bandini, P., Karim, A. (2000). Shear Strength and Stiffness of Silty Sand. Journal of Geotechnical and Geoenvironmental Engineering, American Society of Civil Engineers, 126(5), 451–462.

Salgado, R., Han, F., Prezzi, M. (2017). Axial resistance of non-displacement piles and pile groups in sand. Rivista Italiana di geotecnica, 51(4), 35-46.

Schofield, A. N. (1988). Mohr Coulomb error correction. Ground Engineering, (August), 30–32.

Schofield, A. N. (2006). Interlocking, and peak and design strengths. Géotechnique, 56(5), 357–358.

Taylor, D. W. (1948). Fundamentals of soil mechanics. John Wiley & Sons, New York, USA.

Tehrani, F. S., Han, F., Salgado, R., Prezzi, M., Tovar, R. D., Castro, A. G. (2016). Effect of surface roughness on the shaft resistance of non-displacement piles embedded in sand. Géotechnique, 66(5), 386–400.

Terzaghi, K. (1943). Theoretical soil mechanics. John Wiley & Sons, New York.

Terzaghi, K., Peck, R. B. (1967). Soil mechanics in engineering practice. Wiley, New York.

Tovar-Valencia, R. D., Galvis-Castro, A., Salgado, R., Prezzi, M. (2018). Effect of surface roughness on the shaft resistance of displacement model piles in sand. Journal of Geotechnical and Geoenvironmental Engineering, 144(3), 04017120.

Vardoulakis, I., Goldscheider, M., Gudehus, G. (1978). Formation of shear bands in sand bodies as a bifurcation problem. International Journal for Numerical and Analytical Methods in Geomechanics, 2(2), 99–128.

Woo, S. I., Salgado, R. (2014). Determination of an image point on a surface based on a π plane-based algorithm. Computational Mechanics, 53(5), 1033–1046.

Woo, S. I., Salgado, R. (2015). Bounding Surface Modeling of Sand with Consideration of Fabric and its Evolution During Monotonic Shearing. International Journal of Solids and Structures, Elsevier Ltd, 63, 277–288.

Woo, S. I., Salgado, R. (2018). Simulation of penetration of a foundation element in Tresca soil using the generalized interpolation material point method (GIMP). Computers and Geotechnics, 94, 106–117.