

A NOTE ON APPLICABILITY OF ARTIFICIAL INTELLIGENCE TO CONSTITUTIVE MODELING OF GEOMATERIALS

Mehdi Pouragha,^{1,*} Mahdad Eghbalian,² & Richard Wan²

¹Department of Civil and Environmental Engineering, Carleton University, Ottawa, Canada, ON K1S 5B6

²Department of Civil Engineering, University of Calgary, Canada, AB T2N 1N4

*Address all correspondence to: Mehdi Pouragha, Department of Civil and Environmental Engineering, Carleton University, Ottawa, Canada, ON K1S 5B6, E-mail: mehdi.pouragha@carleton.ca

Original Manuscript Submitted: 9/20/2020; Final Draft Received: 11/17/2020

With the ever accelerating spread of artificial intelligence (AI) in virtually all disciplines of science and engineering, the geotechnical studies and practices have also adopted these approaches for exploring and modeling of complex problems whose thorough understanding often falls beyond the reach of analytical and even numerical methods. In the midst of the overwhelming appeal of AI during recent years, however, there remains some overlooked fundamental questions regarding the inherent ability of AI-based models to represent the constitutive behavior of materials in general, and geomaterials in particular. This brief communications explores, from a theoretical point of view, the question of if, and how, an AI-generated model can replace symbolic constitutive models for materials and what would be the future of theoretical constitutive modeling in the age of AI.

KEY WORDS: artificial intelligence, constitutive modeling, incrementally nonlinear models

1. INTRODUCTION

“[The] confusion of reduction as a tactic with reductionism as an ontological stance is like saying that a square wave is really the sum of a large number of sine waves because I can so represent it to an arbitrary degree of accuracy.”

Richard Levins and Richard Lewontin,
the dialectical biologist, 1985.

It was Roy Amara, the futurist scientist, who famously said, “We tend to overestimate the effect of a technology in the short run and underestimate the effect in the long run.” The statement is arguably accurate for artificial intelligence (AI) and its application to many problems, including among the materials sciences. History will indeed be a fairer judge, but when looking at the body of research and scientific perception of AI, we seem to be situated at the transition between what Amara meant by short run and long run. The overenthusiastic beliefs in AIs are being gradually adjusted by more realistic conceptions that outline the nonetheless impressive scope, as well as the limitations and pitfalls of such approaches.

In the realm of materials modeling, and geomechanics in particular, the AI methods, and specifically, artificial neural networks (ANNs), have been widely used for problems that span from simple multivariable calibration procedures to modeling of complex boundary value problems associated with ground excavation and slope stability (Moayedi et al., 2018). As expected, the developed ANNs for such case-specific loading programs remain, by nature, case specific and are applicable only within the scope of the dataset for which the models are trained.

Nonetheless, such inherent case specificity and dependency on the training scope did not prevent the percolation of ANN modeling approaches into the more general domain of problems such as constitutive modeling of materials. Most notably, in the early and mid 1990s, studies by Ghaboussi and coworkers (Ghaboussi et al., 1991), and others such as Ellis et al. (1995) attempted, with some success, to replace the constitutive models of concrete and geomaterials with ANN models. Ever since, neural networks have been adopted for representing various types of constitutive models, for instance, the elasticity and plasticity of foams (Liang and Chandrashekara, 2008; Settgaest et al., 2019), the elastoplasticity of solids (Zhang and Mohr, 2020), viscoplasticity (Furukawa and Yagawa, 1998; Xu et al., 2020b), and the multiscale response of fiber-reinforced composites (Liu et al., 2020). Huang et al. (2020) explored the performance of deep neural networks in representing a mechanical constitutive model, including the multiscale response of composites, by adopting boundary-value-problem FEM simulations as the training data. The comparison provided therein with other forms of function approximations clearly demonstrated the strength of neural networks rooted in their regularization and generalization capabilities. Moreover, in a noteworthy recent study, Xu et al. (2020a) outlined an ANN-based incremental framework adaptable to various classes of materials, including hyperelastic, elastoplastic, and multiscale composites. In other studies related to materials modeling, neural networks have been adopted as an offline upscaling tool to arrive at a representation of material behavior linked with its microstructural attributes. Notable works here include modeling of the complex path-dependent multiscale plasticity (Mozaffar et al., 2019) and poroplasticity (Wang and Sun, 2018).

From a broader historical perspective, such data-driven approaches to materials modeling have been dubbed as the “fourth paradigm” in materials characterization proceeding experimental, analytical, and numerical methods (Agrawal and Choudhary, 2016). Recent decades have witnessed the birth of numerous machine learning techniques, each exploring new horizons in the capabilities of AI, and many of which have been adopted by researchers in various fields of materials modeling. Studies such as Gao (2018), Bock et al. (2019), and Moayedi et al. (2018) provide a thorough comparative review of these techniques and their advantages and disadvantages specific to engineering problems and geomechanics in particular. The general verdict of such reviews is that, in the field of materials modeling, AI-based models are capable of capturing, to a good extent, the behavior of the target materials subjected to generic loading protocols. However, as in almost every field, no “one-size-fits-all” approach exists, and the modeling techniques are often tailored for the specific problem at hand (Bock et al., 2019; Gao, 2018). Also, not surprisingly, any attempt at generalization of AI-based models is strictly contingent on the availability of comprehensive good-quality training data (Bock et al., 2019; Peng et al., 2020; Xu et al., 2020a; Zhang and Mohr, 2020). While this might seem to be a practical shortcoming at first glance, it quickly turns into a fundamental issue recalling that the size of the ideal training dataset increases exponentially with the number of prominent model parameters (Alwosheel et al., 2018; Hestness et al., 2017).

Beyond such well-understood limitations, the applicability of AI-based constitutive models in capturing more complex loading conditions remains yet to be verified due to more fundamental

concerns (Mozaffar et al., 2019; Xu et al., 2020a). The difference between the two above-mentioned classes of problems (i.e., case-specific boundary value problems vs. constitutive modeling) is visualized for a pair of typical cases in Fig. 1. For a simple boundary value problem concerning the stability of a foundation, a generic ANN, as schematically depicted in Fig. 1(a), can be trained for inputs such as the vertical load and the size of the footing, while the output of the model is the factor of safety (FS) against failure [Fig. 1(b)]. Given a proper set of training observations, the trained model can then be used to predict the stability of the next footing whose properties lies within the range of data used for training. Expectedly, the model is only applicable for the loading types for which it is trained for, and any extension beyond the loading conditions included in the training dataset requires an update in training.

On the other hand, for the case of representing a constitutive model, the ANN is trained for a finite set of stress-strain responses associated with a finite number of loading paths, while the final model is expected to be applicable for all the possible loading paths which are not necessarily included in the training dataset, as illustrated in Fig. 1(c).

Thus, a fundamental difference arises between the modeling through ANN of case-specific problems, such as the one in Fig. 1(b), as compared to ANNs representation of a constitutive model [see Fig. 1(c)]; whereas the former class embraces its case specificity, the latter has an implicit claim to generality that is the basis of every conceptual “model.” In other words, a constitutive model differs from an ad-hoc fit precisely in that it claims to be applicable to loading conditions for which it is not necessarily calibrated/trained.

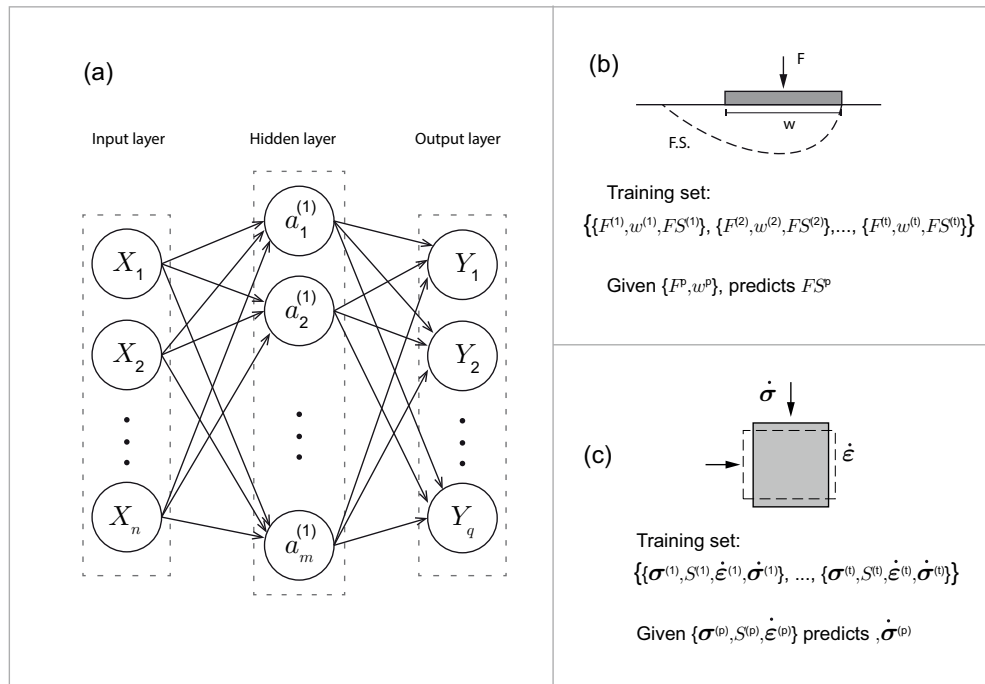


FIG. 1: (a) Structure of a generic three-layer neural network. (b) Boundary value problem of stability of a simple foundation with width of w bearing the vertical force of F . (c) Representing a constitutive model via an ANN. The model is expected to predict the response to loading paths that are not explicitly included in the training dataset.

Richard Levins, quoted at the beginning of this paper, was among the first to reflect upon different strategies for general model building in his seminal 1966 article where he outlined an almost zero-sum tradeoff among precision, realism, and generality of a model (Levins, 1966). A physical model, in Levins's opinion, is bound to make a compromise among these three basic elements with one or two elements often being sacrificed in favor of others. Levins's arguments, framed mainly for mathematical models in biology, was greeted by a mixture of approval and criticism by its successors with the core idea of the existence of such a trade-off being still argued to date (Lewis and Belanger, 2015; Orzack, 2012). Nevertheless, Levins's argument on the role and the trade-off thereof precision, realism, and generality becomes of particular relevance when evaluating the efficacy of AI models to fully represent broad concepts such as constitutive behavior.

In this paper we ponder, from a rather speculative point of view, the inherent capability of, and conditions required for, artificial intelligence and ANNs in particular to represent a mechanical constitutive model. The requirements regarding the claim of generality are scrutinized in the context of common and more elaborate incrementally nonlinear constitutive models for geomaterials. The idea of sufficient domain knowledge is introduced that ensures the generality of an AI-based model. The example of classical incrementally linear soil models is compared with incrementally nonlinear models with the possibility of the claim of generality only being apparent for the former. The paper concludes with armchair contemplation of the role of theoretical constitutive modeling in the age of artificial intelligence. Despite the contemplative nature of the study, the practical field of AI in geotechnical engineering can doubtlessly benefit from such a bird's eye overview of the theoretical basics in relation to the state-of-the-art research in AI. Through a more accurate charting of AI's capabilities in modeling of materials, such explorations can serve as a primary guideline for researchers and engineers to better realise the scope of practical possibilities in geotechnical modeling while also preventing crucial pitfalls associated with overextension of AI models without necessary precautions.

2. CLAIM OF GENERALITY

Explicit definitions for the concept of generality in the context of modeling, and the different types that it assumes, have been offered in Weisberg (2004) with some quantitative scales based on measure theory put forward in Lewis and Belanger (2015). Loosely put, generality of a model refers to its capability of being applicable to a range of problems for which the model is not directly trained or calibrated. For instance, a soil constitutive model that is calibrated using drained triaxial test data is implicitly claiming, and is indeed expected, to capture other common loading conditions such as undrained or simple shear. Notwithstanding the comparative notion adopted herein, the claim of generality, expressed or otherwise implied, is at the core of all constitutive models inasmuch as a constitutive model is different from an ad-hoc fit.

In the case of traditional symbolic constitutive models, the claim of generality originates from the theoretical formalism upon which the model is based. By being established upon a physically sound basis such as thermodynamic requirements, potential functions, frame indifference, and tensorial isotropy, symbolic constitutive models claim that, once properly calibrated, they are applicable to loading program beyond the one for which they are initially calibrated.

Artificial intelligence, on the other hand, embraces quite wholeheartedly the ad-hocery associated with fitting in the absence of a formalism. The network weights in ANN are updated and eventually optimized without any regard to the possible formalism that the network might or might not represent. Therefore the question arises as to whence the claim of generality for

such models can originate. Indeed, it is well known that ANNs can be regarded as universal approximation functions which, in principle, are capable of estimating any multivariable, continuous function (Hartman et al., 1990; Hornik et al., 1989). Nevertheless, doubt still remains as to whether the accuracy required for the claim of generality can be achieved based on a finite set of observations used for training.

Implied here is the fact that the claim of generality refers only to the loading conditions and does not extend to the material's property. A model, symbolic or AI-based, can only be interpolated within the range of material properties for which it is calibrated or trained. Considering geomaterials, for instance, if a model is calibrated for a given range of void ratios and stresses, it cannot be expected *a priori* to be applicable beyond that range. The generality in this case refers to the fact that the model should be predictive of all combinations of stress and strains applied at the boundary.

2.1 Sufficient Domain Knowledge

The question of generality, as formulated herein, lies at the heart of the skepticism that proponents of theoretical methods sometimes exhibit toward artificial intelligence. However, we argue that the claim of generality can be justified for artificial intelligence based models of materials through what we may call "sufficient domain knowledge," or SDK for short. Simply put, the idea of sufficient domain knowledge envisages a set of experiments which collectively describes the behavior of material completely. Hence, it follows that any model (whether or not it is based upon a formalism) that can be uniquely calibrated (or is successfully trained) to capture SDK, will ineluctably be capable of predicting all the other loading conditions possible within the domain of the physics being studied. Worth noting is that the interpretation of generality adopted here is similar to what is dubbed as *p-generality* by Weisberg (2004). Of course, it is crucial to notice that the domain of the physics being concerned dictates the size of SDK. For instance, the SDK required for the mechanical behavior of materials is considerably smaller than that required for, for instance, thermomechanical behaviors.

For the mechanical behavior of geomaterials, a plausible SDK includes the stress-strain and volumetric responses of the soil to tests such as drained and undrained triaxial and simple shear experiments. Given the scope of the physics intended to be captured, the SDK may include static and dynamic experiments, as well as loadings that capture more intricate properties such as the noncoaxial loadings represented by hollow cylinder experiments. It is important to notice that we do not intend here to define the exact SDK but rather to examine the possibility of its existence. What is included in SDK can then be determined through theoretical consideration of the physics domain or simply through trial and error.

Based on the above argument, the generality of an AI-based constitutive model is ensured if a finite-size SDK is shown to exist, meaning that the concerned constitutive behavior, in its entirety, can be encapsulated into a finite set of observations.

2.2 Constitutive Modeling, Reducibility, and Sufficient Domain Knowledge

We argue in this section that the existence of a symbolic constitutive model, by and in itself, implies the existence of a finite-size sufficient domain knowledge.

Envisaging soil sample as a system of interacting particles, from a micromechanical perspective, the behavior of the assembly is governed by the position and velocity of numerous particles and different physics that govern their equally numerous interactions. Thus, for all practical

purposes, the amount of information required to interpret the collective behavior in terms of particle-scale properties can be assumed to be infinite. In much the same manner as discrete elements method (DEM) simulations, the overall behavior at the sample (macroscopic) level originates from equations of interactions and Newton's laws of motion solved algorithmically for all the particles. While often taken for granted, it is not given that the collective behavior arising from such repetitive algorithms should be explicable via a so-called "covering law," a.k.a a constitutive model for the case of multiagent materials (Sawyer, 2013). Indeed, the chaos theory of dynamic systems argues that deterministic constitutive models can only be formulated for variables that do not exhibit sensitive dependence on initial conditions (Prokopenko and Einav, 2015; Reisch, 1991) and as a result, can be reduced to statistical descriptions. For the case of geomaterials, for instance, as much as constitutive models can be set up for averaged properties such as stress-strain relationships, they cannot be extended to predict characteristics such as location and the shape of strain localization patterns (e.g., shear bands), which depend sensitively on local parameters.

Therefore the existence of a constitutive model implies the reducibility of the material's behavior with respect to the considered physics domain; the fact that the collective response of millions of particles is assumed to be encapsulated in a formulation involving a mere handful of state and material variables is inevitably based upon such a reductionism.

Coming back to the concept of sufficient domain knowledge, if the minimum size of SDK is infinite, it entails that no finite set of parameters should be able to describe the intended physical behavior. Combined with the reducibility argument provided above, it can then be deduced, via proof by contradiction, that if a constitutive model exists the behavior is then reducible, which also means that a finite-size SDK should exist. Hence it follows that

if a constitutive model is believed to exist, a proper AI-based model, ad hoc as it is, should be able to represent the constitutive model in its entirety.

There would indeed be processes (chaotic events, for instance) for which the AI-based methods cannot produce a reliable model. However, this is not indicative of an inherent shortcoming of AI but is rather due to the fact that a constitutive model simply does not exist.

3. EXAMPLE: INCREMENTALLY LINEAR AND NONLINEAR CONSTITUTIVE MODELS FOR SOILS

The reduction of symbolic constitutive models to a finite set of inputs, i.e., sufficient domain knowledge or SDK, is discussed here through two general classes of models commonly used for geomaterials—incrementally linear and incrementally nonlinear models.

Due to their inherent dissipative nature and path-dependent response, soil constitutive models are often formulated in an incremental form whereby the rate of stresses is related to that of strains. Considering small strain ranges for simplicity, the constitutive relation takes the following general form:

$$\dot{\sigma}_{ij} = C_{ijkl} \dot{\epsilon}_{kl}, \quad (1)$$

where the overdot denotes the time derivative (rate), σ and ϵ are the second-order stress and strain tensors, and C is the fourth-order constitutive stiffness tensor. Einstein notation is adopted where repetition of indices indicates summation.

Given the path dependence of geomaterial response, the constitutive stiffness tensor C is often a function of current stress and state of material, i.e., $C = C(\sigma, S)$, with S being the

set of state variables defining the current state of materials. The constitutive model is deemed incrementally linear when the constitutive tensor C does not depend on loading rate, in this case $\dot{\epsilon}$, which means that the constitutive behavior can be understood and explored independently from the applied loading.

The typical stress response of an incrementally linear model to a so-called strain probing program is shown in Figs. 2(a)–2(c), where increments of strains with the same magnitude but different ratios of principal values are applied [Figs. 2(a) and 2(b)], with the incremental stress response as given in Fig. 2(c). While the applied strain probes fall along a circle, the stress response envelope traces a rotated ellipse whose characteristics depend on the particularities of C . Figure 2(c) also shows how mapping of the strain probe onto the corresponding stress response is represented by the fourth-order tensor C and the tensorial product in Eq. (1).

For the case of incrementally linear models in D dimensions, the constitutive behavior can be fully captured if the response of the material is known for at least D linearly independent strain increments whose combination can span the space of possible strain increments. Given D linearly independent strain rates $\{\dot{\epsilon}^{(1)}, \dots, \dot{\epsilon}^{(D)}\}$ and their respective stress responses $\{\dot{\sigma}^{(1)}, \dots, \dot{\sigma}^{(D)}\}$, the response to any arbitrary strain rate can now be found by treating the selected D strain rates as basis:

$$\dot{\sigma}_{ij} = C_{ijkl} \dot{\epsilon}_{kl} = C_{ijkl} \sum_{q=1}^D a^{(q)} \dot{\epsilon}_{kl}^{(q)} = \sum_{q=1}^D a^{(q)} \dot{\sigma}_{ij}^{(q)}, \quad (2)$$

where $\dot{\epsilon}^{(q)}$ are the D linearly independent strain rates and $\dot{\sigma}^{(q)}$ are the associated stress responses for which $\dot{\sigma}_{ij}^{(q)} = C_{ijkl} \dot{\epsilon}_{kl}^{(q)}$. Here $a^{(q)}$'s are coefficients describing the decomposition of strain

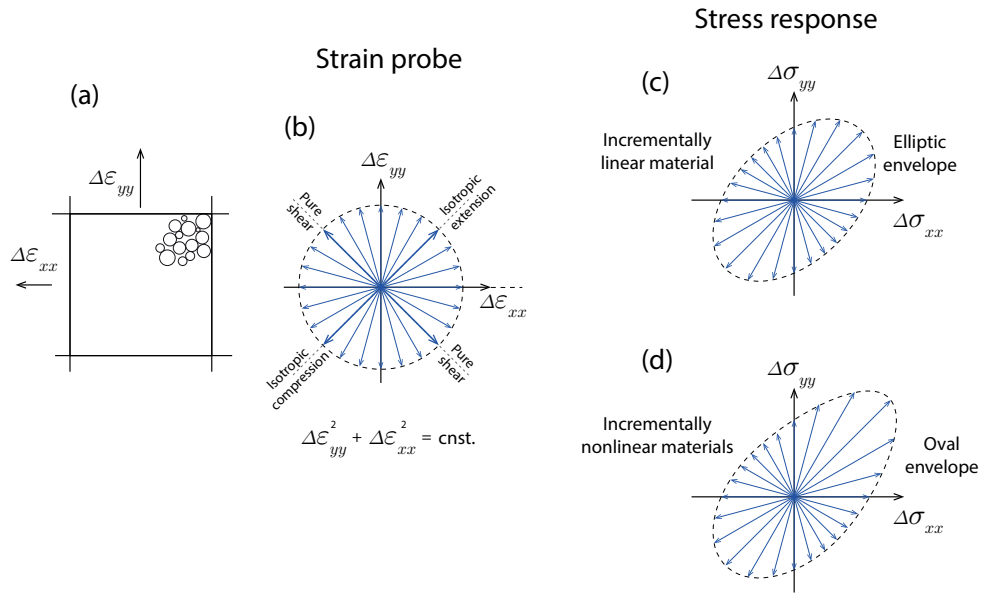


FIG. 2: (a) Vertical and horizontal strain increments applied to an element of soil. (b) Strain probing program; vertical and horizontal strain increments with different ratios are applied while the magnitude of strain is kept constant. (c, d) Stress response to strain probing of incrementally linear and nonlinear materials, respectively.

tensors into their basis. Simple though it may seem, Eq. (2) demonstrates how, for the case of incrementally linear models, the constitutive relation at any particular state of material can be encapsulated into D separate observations. Hence the set of observations $\{\dot{\epsilon}^{(q)}, \dot{\sigma}^{(q)}\}$ can be said to form a sufficient domain knowledge, SDK, for the given stress and material state. The proof procedure can be readily extended to common elastoplastic models where two separate tensorial zones, associated with loading and unloading paths, are recognized.

It is important to notice that the relation in Eq. (2) could not have been obtained were it not for the fact that the constitutive stiffness C remains constant for different strain increments. Moreover, the provided proof only shows the possibility of a finite SDK that encapsulates the constitutive behavior in a limited set of observations. It is crucial to notice, in practice, that the SDK does not represent a sufficient training set for all of an arbitrary ANN model, which possibly involves significantly more degrees of freedom than can be optimized by such a limited number of observations.

Returning to the types of geomechanical constitutive models, it is known that for granular material such as sand, the constitutive behavior does depend on loading direction, which leads to a class of advanced models called incrementally nonlinear, as studied by Darve and coworkers (Darve, 1990; Nicot and Darve, 2007). For such models the constitutive stiffness depends on the applied strain rate, or rather its direction, i.e.,

$$C = C\left(\sigma, S, \frac{\dot{\epsilon}}{\|\dot{\epsilon}\|}\right), \quad (3)$$

where $\|\dot{\epsilon}\| = \sqrt{\dot{\epsilon}_{ij}\dot{\epsilon}_{ij}}$ is the magnitude of strain increment and the term $\dot{\epsilon}/\|\dot{\epsilon}\|$ designates the direction of applied strain rate. Crucially, the form of constitutive tensor in Eq. (3), while nonlinear does not violate the first-order homogeneity requirement of Eq. (1) with respect to strain rates. The typical response of an incrementally nonlinear model to strain probing is given in Fig. 2(d), where, unlike the incrementally linear case [Fig. 2(c)], the response stress envelope no longer falls along an ellipse[†].

It is obvious right away that the calculation in Eq. (2) no longer holds since the constitutive stiffness C varies for each strain rate, and thus the question arises as to whether a sufficient domain knowledge can, in principle, be realized for such cases. One can still argue, rather loosely we admit, that if a symbolic constitutive model is found to capture the directional dependency of C with a finite number of variables, then there would exist a finite number of observations based on which the parameters of the symbolic model can be uniquely identified. The set of such observations may be a primary candidate for sufficient domain knowledge to provide an AI-based model with the claim of generality. A more thorough theoretical investigation is nonetheless called for here to convincingly prove the existence of such SDK for incrementally nonlinear models. One can imagine a proof alluding to series expansion of $C(\sigma, S, \dot{\epsilon}/\|\dot{\epsilon}\|)$ in terms of $\dot{\epsilon}/\|\dot{\epsilon}\|$ and attempting to achieve relations similar to those in Eq. (2), a detailed examination of which is, however, beyond the scope of this article.

It is also worth mentioning that a symbolic constitutive model such as the one in Eq. (1), still remains more general compared to AI-based models in that it provides the possibility of rearranging stress and strain increments into new control/response conjugate sets, similar to the

[†]Note that the foregoing discussion is only a crude and incomplete presentation of incrementally nonlinear models that nevertheless serves to convey the main points. Similar behaviors to those presented in Fig. 2(d) can be produced by common incrementally bilinear elastoplastic models that are often considered to be incrementally linear within their so-called tensorial zone.

boundary conditions applied for strain-controlled triaxial tests where the incremental control parameters are vertical strains and horizontal stresses. A trained ANN, on the other hand, does not provide the possibility of such crossovers among input and output layers. Nevertheless, this shortcoming is not crucial, mainly because ANN models are almost never used as standalone entities and are instead oftentimes implemented into finite-element or other numerical solvers which formulate the mechanical response as strain-driven processes and any imposed stress requirement is achieved through proper iteration loops.

4. INCORPORATING FORMALISM INTO AI-BASED MODELS

As mentioned earlier, the size of a sufficient knowledge domain does not reflect the amount of observation required for training an AI-based model, which is exactly due to its lack of a formalism. In the absence of *a priori* structure, the number of model parameters to be optimized in an AI-based model (e.g., neurons' weight in the case of ANNs) is significantly larger than the number of calibration parameters in an equivalent symbolic model. As such, more observations are naturally required for the AI-based model to achieve, or at least approach, unique trained values.

The natural question then arises as to whether the AI-based model can be endowed with some of the characteristics that originate from having a formalism, or, in other words, whether the AI-model can be predisposed with *a priori* physical knowledge of the intended behavior, such as frame indifference, isotropy, and first-order homogeneity with respect to strain increments. Taking the case of ANNs, three different methods can be envisaged in order to encode such information into the AI-based model.

4.1 Artificial Expansion of Training Data

As a common approach already introduced in Lefik and Schrefler (2003), this method expands the original training data in order to restrain, through brute force, the ANN to conform with the intended criteria of the formalism. For instance, given an initial set of training data, the isotropy of the trained ANN can be imposed by conjoining the initial set with its rotated counterpart:

$$\begin{aligned} \text{Initial training set} & \xrightarrow{\text{Isotropy}} \text{Expanded training set} \\ \{\boldsymbol{\sigma}^{(q)}, S^{(q)}, \dot{\boldsymbol{\sigma}}^{(q)}, \dot{\boldsymbol{\epsilon}}^{(q)}\} & \rightarrow \\ & \left\{ \{\boldsymbol{\sigma}^{(q)}, S^{(q)}, \dot{\boldsymbol{\sigma}}^{(q)}, \dot{\boldsymbol{\epsilon}}^{(q)}\}, \{Q^T \boldsymbol{\sigma}^{(q)} Q, S^{(q)}, Q^T \dot{\boldsymbol{\sigma}}^{(q)} Q, Q^T \dot{\boldsymbol{\epsilon}}^{(q)} Q\} \right\}, \end{aligned} \quad (4)$$

where Q is a rotation tensor. The number of different rotations to be considered in order to ensure the accuracy of a model is studied in Ling et al. (2016).

The same can be applied to ensure first-order homogeneity with respect to strain/stress increments, i.e., the training set can be extended to include linearly scaled input data:

$$\begin{aligned} \text{Initial training set} & \rightarrow \text{Expanded training set} \\ \{\boldsymbol{\sigma}^{(q)}, S^{(q)}, \dot{\boldsymbol{\sigma}}^{(q)}, \dot{\boldsymbol{\epsilon}}^{(q)}\} & \rightarrow \left\{ \{\boldsymbol{\sigma}^{(q)}, S^{(q)}, \dot{\boldsymbol{\sigma}}^{(q)}, \dot{\boldsymbol{\epsilon}}^{(q)}\}, \{\boldsymbol{\sigma}^{(q)}, S^{(q)}, \alpha \dot{\boldsymbol{\sigma}}^{(q)}, \alpha \dot{\boldsymbol{\epsilon}}^{(q)}\} \right\}, \end{aligned} \quad (5)$$

with $\alpha > 0$ being a scaling factor. One of the drawbacks of the data augmentation method is the ensuing “black box” problem, leading to emergence of redundant weights in the neural network (Worrall et al., 2017).

4.2 Imposing Physics-Based Symmetries

Some properties pertaining to inherent symmetries of the constitutive models can be encoded into ANNs by properly modifying the input and output training data. In particular, in order to ensure isotropy and frame indifference, simply an invariant form of the training stress/strain data can be used, which axiomatically secures isotropy (Lefik and Schrefler, 2003; Ling et al., 2016). The increased accuracy and efficiency of the neural network in this method, however, comes at the expense of higher computational costs (Worrall et al., 2017). A similar though more involved method can be adopted for anisotropic materials whose behavior can be described in terms of more joint invariants of a stress, strain, and structural anisotropy matrix through a proper representation theorem. Incorporating such symmetries in ANNs via invariant representation is discussed in detail in Cohen and Welling (2016) and Worrall et al. (2017) among others, mainly in the context of image, video, and audio recognition. However, the main ideas can be readily translated to materials modeling applications. Along a similar line, Xu et al. (2020a) showed how symmetries pertaining to the positive definiteness of a constitutive model can be accounted for in a neural network.

4.3 *A Priori* Structured ANNs

The last and more interesting method to encode formalism into ANN is to predispose the ANN with a structure that resembles the symbolic formalism. Here we quickly approach the edge of barely charted territories in theoretical studies of neural networks. In principle, such a background provides insight regarding the behavior of systems which can be coherently encoded into ANNs through a so-called “knowledge distilling” process (Hinton et al., 2015; Tartakovsky et al., 2018). Alternatively, the method of graph partition neural networks (Liao et al., 2018) can be used to mimic a flow of intermediary variables similar to the represented symbolic constitutive model. A primitive scheme of such partitioned networks can look like Fig. 3, where the ANN can be imagined as a collection of connected sub-networks that are structured according to the flow of information in a generic elastoplastic constitutive model, with the training procedure being carried out for the global network. The flow of information among the partitions can now be predetermined and excluded from training procedure. An interesting instance can be found in the recent work of Masi et al. (2020), where the ANN has been predisposed by an internal structure so as to ensure thermodynamic requirements. A more detailed assessment of such methods and their applications involves technicalities that fall well beyond our current purview.

5. THE ROLE OF CONSTITUTIVE MODELING IN THE AGE OF AI[‡]

If we accept the potency of AI in auspiciously representing the constitutive behavior of materials, then why do we need our precious elegantly forged symbolic theoretical models? Time and again the question has been brought up in discussions with colleagues in the field of constitutive modeling, and justifiably so since the same question is being encountered in many other fields as AIs rapidly take over functions that have traditionally been considered to be the sacred realm of conceptual model building.

Here we delve back into Levins’s concept of trade-off among precision, realism, and generality. In this context, the current state of AI-based models has eyes only for the precision; from an

[‡]This section is motivated by a conversation of the first author with Professor James Jenkins (Cornell University) during a recent Lorentz Center meeting in Leiden, The Netherlands.

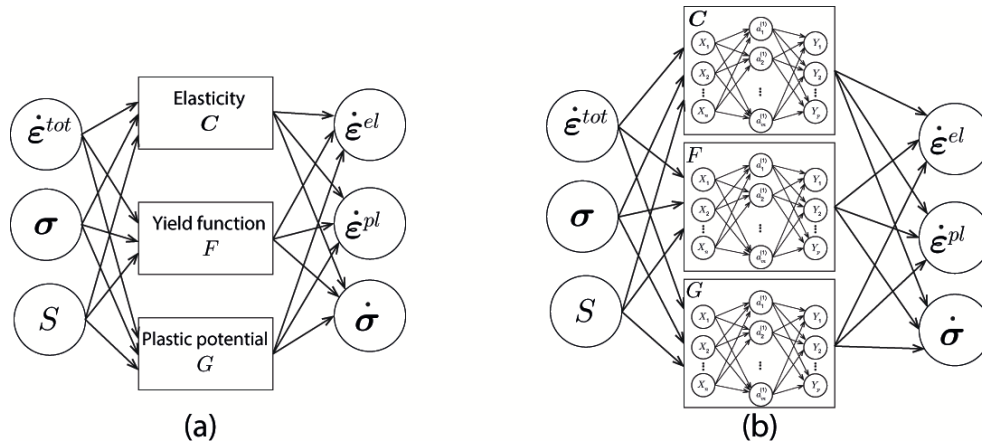


FIG. 3: (a) Structure of a typical elastoplastic constitutive model. (b) Structured ANN where each component of the constitutive model is represented by a subnetwork.

AI point of view, as long as the predictions are correct, we can dispose with the other two components for all practical purposes. The concern is perhaps more aptly framed as the dichotomy of comprehension and competence given in Dennett (2009). Put in this context, the AI modeling is depicted as “competence without comprehension,” which mirrors, to a good extent, the sacrifice of realism and generality in favor of precision, as Levins puts it.

It is no secret that human intelligence has lost the battle on precision front to artificial intelligence. Even on the constitutive modeling front, the studies show that ANN models not only outperform the symbolic models in precision, but they also exhibit more robustness as well as an easier implementation into finite element and other numerical solvers (Lefik and Schrefler, 2003; Shin and Pande, 2000). Other black-box models—based time series analysis are also shown to be capable of capturing the details of geomaterial behaviors that fall beyond the reach of conventional symbolic constitutive models (Small et al., 2013).

From such a perspective, it transpires that the future of theoretical constitutive modeling studies lies in producing comprehension about the nature of mechanical behaviors rather than quantitative results. At the risk of being proven wrong over time, we believe that those constitutive models aiming at explaining and understanding the underlying behaviors, rather than producing more accurate results with feeble formalism would be those that retain their relevance over time. Instances of such comprehension-oriented models are the micromechanical constitutive models that are often predicated upon microvariables that cannot be directly measured and calibrated, and as such, their quantitative predictive efficacy has always been their Achilles heel when compared against more traditional models (Pouragha and Wan, 2018). However, the remarkable explanatory power of such micromechanical approaches provides a great deal of what can be deemed as deeper comprehension of the overall behavior of materials.

In the field of geomaterials, the advent of discrete element methods and their counterpart multiscale theories have, over the past few decades, greatly contributed to better understanding of physical processes that govern the stress-strain response of soils by interconnecting the many macroscopic facets of continuum mechanics to each other through their mutual microscopic underpinnings. Such comprehension-oriented models produce knowledge that can be employed, at the lowest level, in feature engineering of ANNs, while at higher levels they can provide the *a priori* information necessary for predesigning the internal structure of ANNs. Interestingly,

the reverse is also true, whereby ANNs can be employed to better bridge between micro- and macroscale properties. Recent studies indeed demonstrate the potency of machine learning techniques in pinpointing relevant macrolevel physics (Rudy et al., 2017), as well as recognizing prominent microscopic variables (such as local time and length scales) to be considered during the homogenization and localization procedures (Mozaffar et al., 2019; Peng et al., 2020; Vasilyeva et al., 2020; Wang and Sun, 2018).

In the end, it seems that the future of constitutive modeling involves a continuous dialogue between the comprehension obtained from theoretical constitutive modeling and the quantitative competence of AIs, for, as Levins put it, “understanding is not achieved by generality alone but by a relation between the general and the particular” (Levins, 1966).

6. CONCLUSION

With the ever-increasing appeal of artificial intelligence methods in the field of geotechnical modeling, the current study undertakes the timely task of exploring the inherent potentials of AI-based models, such as ANNs, to represent a constitutive model in its entirety. Notwithstanding their ad hoc nature, we investigated the possibility of AI-based models to have a claim of generality similar to that of symbolic models. By introducing the concept of sufficient domain knowledge, we demonstrated that generality of an AI-based representation of a constitutive model can indeed be achieved for incrementally linear models, with possible extensions sketched for incrementally nonlinear models.

Finally, the quantitative competence of AI-based models is contrasted with the comprehension-oriented outcome of theoretical constitutive models, whereby better knowledge and more accurate models are expected to form in a dialogue between the two counterparts.

ACKNOWLEDGMENTS

The ideas laid out in this article benefited from numerous stimulating discussions with Professor Felix Darve (Laboratoire 3SR, Grenoble) and Professor Francois Nicot (INRAE, Grenoble) over past few years. The first author would also like to acknowledge many in-depth conversations with Dr. Zaman Forootan of Deloitte, Canada, on the nature of AI models and neural networks. Research funding provided by the Natural Sciences and Engineering Research Council of Canada (RGPIN-2020-06480 and DGEER-2020-00411 held by M. Pouragha, and RGPIN-2016-04086 held by R. Wan) is gratefully acknowledged.

REFERENCES

- Agrawal, A. and Choudhary, A., Perspective: Materials Informatics and Big Data: Realization of the “Fourth Paradigm” of Science in Materials Science, *Appl. Mater.*, vol. **4**, no. 5, p. 053208, 2016.
- Alwosheel, A., van Cranenburgh, S., and Chorus, C.G., Is Your Dataset Big Enough? Sample Size Requirements when Using Artificial Neural Networks for Discrete Choice Analysis, *J. Choice Model.*, vol. **28**, pp. 167–182, 2018.
- Bock, F.E., Aydin, R.C., Cyron, C.J., Huber, N., Kalidindi, S.R., and Klusemann, B., A Review of the Application of Machine Learning and Data Mining Approaches in Continuum Materials Mechanics, *Front. Mater.*, vol. **6**, p. 110, 2019.
- Cohen, T. and Welling, M., Group Equivariant Convolutional Networks, *Int. Conf. Machine Learning*, pp. 2990–2999, 2016.

- Darve, F., The Expression of Rheological Laws in Incremental Form and the Main Classes of Constitutive Equations, *Geomaterials: Constitutive Equations and Modeling*, Boca Raton, FL: CRC Press, pp. 139–164, 1990.
- Dennett, D., Darwin's Strange Inversion of Reasoning, *Proc. Nat. Acad. Sci. USA*, vol. **106**, no. 1, pp. 10061–10065, 2009.
- Ellis, G., Yao, C., Zhao, R., and Penumadu, D., Stress-Strain Modeling of Sands Using Artificial Neural Networks, *J. Geotech. Eng.*, vol. **121**, no. 5, pp. 429–435, 1995.
- Furukawa, T. and Yagawa, G., Implicit Constitutive Modeling for Viscoplasticity Using Neural Networks, *Int. J. Numer. Methods Eng.*, vol. **43**, no. 2, pp. 195–219, 1998.
- Gao, W., A Comprehensive Review on Identification of the Geomaterial Constitutive Model Using the Computational Intelligence Method, *Adv. Eng. Informatics*, vol. **38**, pp. 420–440, 2018.
- Ghaboussi, J., Garrett, J., Jr., and Wu, X., Knowledge-Based Modeling of Material Behavior with Neural Networks, *J. Eng. Mech.*, vol. **117**, no. 1, pp. 132–153, 1991.
- Hartman, E.J., Keeler, J.D., and Kowalski, J.M., Layered Neural Networks with Gaussian Hidden Units as Universal Approximations, *Neural Comput.*, vol. **2**, no. 2, pp. 210–215, 1990.
- Hestness, J., Narang, S., Ardalani, N., Diamos, G., Jun, H., Kianinejad, H., Patwary, M., Ali, M., Yang, Y., and Zhou, Y., Deep Learning Scaling Is Predictable, Empirically, 2017. arXiv:1712.00409
- Hinton, G., Vinyals, O., and Dean, J., Distilling the Knowledge in a Neural Network, 2015. arXiv:1503.02531
- Hornik, K., Stinchcombe, M., and White, H., Multilayer Feedforward Networks Are Universal Approximators, *Neural Networks*, vol. **2**, no. 5, pp. 359–366, 1989.
- Huang, D.Z., Xu, K., Farhat, C., and Darve, E., Learning Constitutive Relations from Indirect Observations Using Deep Neural Networks, *J. Comput. Phys.*, p. 109491, 2020.
- Lefik, M. and Schrefler, B., Artificial Neural Network as an Incremental Non-Linear Constitutive Model for a Finite Element Code, *Comput. Methods Appl. Mech. Eng.*, vol. **192**, nos. 28-30, pp. 3265–3283, 2003.
- Levins, R., The Strategy of Model Building in Population Biology, *Am. Sci.*, vol. **54**, no. 4, pp. 421–431, 1966.
- Lewis, C.T. and Belanger, C., The Generality of Scientific Models: A Measure Theoretic Approach, *Synthese*, vol. **192**, no. 1, pp. 269–285, 2015.
- Liang, G. and Chandrashekhara, K., Neural Network Based Constitutive Model for Elastomeric Foams, *Eng. Struct.*, vol. **30**, no. 7, pp. 2002–2011, 2008.
- Liao, R., Brockschmidt, M., Tarlow, D., Gaunt, A.L., Urtasun, R., and Zemel, R., Graph Partition Neural Networks for Semi-Supervised Classification, 2018. arXiv:1803.06272
- Ling, J., Jones, R., and Templeton, J., Machine Learning Strategies for Systems with Invariance Properties, *J. Comput. Phys.*, vol. **318**, pp. 22–35, 2016.
- Liu, X., Tao, F., and Yu, W., A Neural Network Enhanced System for Learning Nonlinear Constitutive Law and Failure Initiation Criterion of Composites Using Indirectly Measurable Data, *Composite Struct.*, vol. **252**, p. 112658, 2020.
- Masi, F., Stefanou, I., Vannucci, P., and Maffi-Berthier, V., Thermodynamics-Based Artificial Neural Networks for Constitutive Modeling, 2020. arXiv:2005.12183
- Moayedi, H., Mosallanezhad, M., Rashid, A.S.A., Jusoh, W.A.W., and Muazu, M.A., A Systematic Review and Meta-Analysis of Artificial Neural Network Application in Geotechnical Engineering: Theory and Applications, *Neural Comput. Appl.*, pp. 1–24, 2018.
- Mozaffar, M., Bostanabad, R., Chen, W., Ehmann, K., Cao, J., and Bessa, M., Deep Learning Predicts Path-Dependent Plasticity, *Proc. Nat. Acad. Sci.*, vol. **116**, no. 52, pp. 26414–26420, 2019.

- Nicot, F. and Darve, F., Basic Features of Plastic Strains: From Micro-Mechanics to Incrementally Nonlinear Models, *Int. J. Plasticity*, vol. **23**, no. 9, pp. 1555–1588, 2007.
- Orzack, S.H., The Philosophy of Modeling or Does the Philosophy of Biology Have Any Use?, *Philos. Transact. Roy. Soc. B: Biol. Sci.*, vol. **367**, no. 1586, pp. 170–180, 2012.
- Peng, G.C., Alber, M., Tepole, A.B., Cannon, W.R., De, S., Dura-Bernal, S., Garikipati, K., Karniadakis, G., Lytton, W.W., and Perdikaris, P., Multiscale Modeling Meets Machine Learning: What Can We Learn?, *Archives Comput. Methods Eng.*, pp. 1–21, 2020.
- Pouragha, M. and Wan, R., μ -GM: A Purely Micromechanical Constitutive Model for Granular Materials, *Mech. Mater.*, vol. **126**, pp. 57–74, 2018.
- Prokopenko, M. and Einav, I., Information Thermodynamics of Near-Equilibrium Computation, *Phys. Rev. E*, vol. **91**, no. 6, p. 062143, 2015.
- Reisch, G.A., Chaos, History, and Narrative, *History Theory*, vol. **30**, no. 1, pp. 1–20, 1991.
- Rudy, S.H., Brunton, S.L., Proctor, J.L., and Kutz, J.N., Data-Driven Discovery of Partial Differential Equations, *Sci. Adv.*, vol. **3**, no. 4, p. e1602614, 2017.
- Sawyer, R.K., Interpreting and Understanding Simulations: The Philosophy of Social Simulation, *Simulating Social Complexity*, Berlin: Springer, pp. 273–289, 2013.
- Settgast, C., Abendroth, M., and Kuna, M., Constitutive Modeling of Plastic Deformation Behavior of Open-Cell Foam Structures Using Neural Networks, *Mech. Mater.*, vol. **131**, pp. 1–10, 2019.
- Shin, H. and Pande, G., On Self-Learning Finite Element Codes Based on Monitored Response of Structures, *Comput. Geotech.*, vol. **27**, no. 3, pp. 161–178, 2000.
- Small, M., Walker, D.M., Tordesillas, A., and Tse, C.K., Characterizing Chaotic Dynamics from Simulations of Large Strain Behavior of a Granular Material under Biaxial Compression, *Chaos Interdisc. J. Nonlinear Sci.*, vol. **23**, no. 1, p. 013113, 2013.
- Tartakovsky, A.M., Marrero, C.O., Perdikaris, P., Tartakovsky, G.D., and Barajas-Solano, D., Learning Parameters and Constitutive Relationships with Physics Informed Deep Neural Networks, 2018. arXiv: 1808.03398
- Vasilyeva, M., Leung, W.T., Chung, E.T., Efendiev, Y., and Wheeler, M., Learning Macroscopic Parameters in Nonlinear Multiscale Simulations Using Nonlocal Multicontinua Upscaling Techniques, *J. Comput. Phys.*, p. 109323, 2020.
- Wang, K. and Sun, W., A Multiscale Multi-Permeability Poroplasticity Model Linked by Recursive Homogenizations and Deep Learning, *Comput. Methods Appl. Mech. Eng.*, vol. **334**, pp. 337–380, 2018.
- Weisberg, M., Qualitative Theory and Chemical Explanation, *Philos. Sci.*, vol. **71**, no. 5, pp. 1071–1081, 2004.
- Worrall, D.E., Garbin, S.J., Turmukhambetov, D., and Brostow, G.J., Harmonic Networks: Deep Translation and Rotation Equivariance, *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 5028–5037, 2017.
- Xu, K., Huang, D.Z., and Darve, E., Learning Constitutive Relations Using Symmetric Positive Definite Neural Networks, 2020a. arXiv: 2004.00265
- Xu, K., Tartakovsky, A.M., Burghardt, J., and Darve, E., Inverse Modeling of Viscoelasticity Materials Using Physics Constrained Learning, 2020b. arXiv: 2005.04384
- Zhang, A. and Mohr, D., Using Neural Networks to Represent Von Mises Plasticity with Isotropic Hardening, *Int. J. Plasticity*, p. 102732, 2020.