

A multiscale multi-permeability poroplasticity model linked by recursive homogenizations and deep learning

Kun Wang · WaiChing Sun

Received: January 11, 2018/ Accepted: date

Abstract Many geological materials, such as shale, mudstone, carbonate rock, limestone and rock salt are multi-porosity porous media in which pores of different scales may co-exist in the host matrix. When fractures propagate in these multi-porosity materials, these pores may enlarge and coalesce and therefore change the magnitude and the principal directions of the effective permeability tensors. The pore-fluid inside the cracks and the pores of host matrix may interact and exchange fluid mass, but the difference in hydraulic properties of these pores often means that a single homogenized effective permeability tensor field is insufficient to characterize the evolving hydraulic properties of these materials at smaller time scale. Furthermore, the complexity of the hydro-mechanical coupling process and the induced mechanical and hydraulic anisotropy originated from the micro-fracture and plasticity at grain scale also makes it difficult to propose, implement and validate separated macroscopic constitutive laws for numerical simulations. This article presents a hybrid data-driven method designed to capture the multiscale hydro-mechanical coupling effect of porous media with pores of various different sizes. At each scale, data-driven models generated from supervised machine learning are hybridized with classical constitutive laws in a directed graph that represents the numerical models. By using sub-scale simulations to generate database to train material models, an offline homogenization procedure is used to replace the up-scaling procedure to generate cohesive laws for localized physical discontinuities at both grain and specimen scales. Through a proper homogenization procedure that preserves spatial length scales, the proposed method enables field-scale simulations to gather insights from meso-scale and grain-scale micro-structural attributes. This method is proven to be much more computationally efficient than the classical DEM-FEM or FEM² approach while at the same time more robust and flexible than the classical surrogate modeling approach. Due to the usage of bridging-scale technique, the proposed model may provide multiple opportunities to incorporate different types of simulations and experimental data across different length scales for machine learning. Numerical issues will also be discussed.

Keywords dual-porosity, data-driven modeling, directed graph, embedded discontinuity, recurrent neural network, multiscale method

1 Introduction

Many geological materials are porous media with a pore size distribution that spans several orders in magnitude. For instance, a crystalline rock may contain micro-pores filled with brine inclusion inside each crystal grain, while precipitation may exist in between grain boundaries. However, the initiation, propagation and coalescence of flaws, defects and cracks may also produce larger pores that become flow conduits. Natural geological process or human activities such as CO₂ storage or hydraulic fractures may also induce

Corresponding author: WaiChing Sun
Assistant Professor, Department of Civil Engineering and Engineering Mechanics, Columbia University, 614 SW Mudd, Mail Code: 4709, New York, NY 10027 Tel.: 212-854-3143, Fax: 212-854-6267, E-mail: wsun@columbia.edu

changes in micro-structural attributes and pore size distribution [Bai et al., 1993, Zimmerman et al., 1993, Kuhlman et al., 2013]. Since the hydro-mechanical coupling effects may vary across multiple temporal and spatial scales due to the large spectrum of the pore size distribution, the interaction of the pore fluid in micro- and macro-pores and that of the solid skeleton may lead to highly complex path-dependent behaviors. While it is possible to propose multiporosity and multipermeability models for deformable porous media, these models often require a large set of material parameters to calibrate the constitutive laws that characterize (1) the hydro-mechanical responses of the solid skeleton, (2) the permeability relations of the macro- and micro-pores, and (3) the fluid mass exchanges across the dual pore spaces [Lewis and Ghafouri, 1997, Callari and Federico, 2000, Choo et al., 2016].

This large amount of material parameters makes material parameter identification difficult and imposes high demand on experimental data that are not always available in practice [Liu et al., 2016]. Meanwhile, the apparent good fits between experiments and simulations may easily be attributed to the wrong reasons – excess curve-fitting and high dimensionality of the material parameter set [Wang et al., 2016]. This so-called curse of high dimensionality (cf. Friedman [1997]) is further complicated by the new set of information afforded by recent X-ray tomographic imaging and digital image correlation techniques. While these techniques are important for understanding new insight of porous media at the microscopic origins, creating a consistent phenomenological interpretation for the relations between macroscopic stress-strain curve and micro-scale grain-scale data with phenomenological model is also understandably more challenging. For instance, the experimental data can be used in the numerical modeling process in a number of different ways. The most trivial case is perhaps the parameter identification procedure, which can be regarded as an optimization or constrained optimization problem where an objective function is defined as a metric that measures the discrepancy between the benchmark (often the experimental data) and the simulation results. The optimized material parameters are determined via a misfit function of optimization problem subjected to a number of equality or inequality constraints (e.g. the valid range of Poisson ratio) [Schmidt et al., 2015, Liu et al., 2016]. The calibration process, therefore, produces the optimal set of material parameters that minimizes the errors measured by the objective function [Wang et al., 2016].

One possible extension of this approach is to use data-driven method to replace the constitutive law itself. The idea of using data-driven model obtained from supervised learning to replace constitutive laws for single-physics solid mechanics problem can at least be traced back to the 90s. For instance, Ghaboussi et al. [1991] discovered that artificial feed-forward neural network can be trained to replace constitutive laws. By utilizing the self-organization capabilities to adjust weight or strength of connections among neurons or processing units, machine can “learn” to reduce the error of the prediction made by the network of processing units via a procedure called back propagation. As a result, the neural network that completes the training against the training data set can be used to replace the constitutive laws for materials, including those exhibiting rate- and history-dependent behaviors [Graf et al., 2010, Furukawa and Yagawa, 1998]. Another different approach has recently been proposed in Kirchdoerfer and Ortiz [2016] where the supervised learning process typically required for the artificial neural network and the neural network itself is completely by-passed. Instead, the authors propose a new constrained optimization problem that minimizes the discrepancy between measured and predicted responses, while the knowledge that is of great certainty or of high degree of belief, such as the compatibility equation and balance principles are used as the constraints. Consequently, by finding the saddle point of the constrained optimization problem, local data sets that are closest to the satisfaction of compatibility and equilibrium can be located to generate incremental updates of elasticity problems. However, this method has not yet been expanded for history-dependent (e.g. plasticity, damage) behaviors and the proper way to incorporate it without internal variables is not clear at this point. In all the cases mentioned above, the data-driven technology is typically applied for one single purpose – replacing conventional constitutive laws with data-driven models either through supervised machine learning (cf. Ghaboussi et al. [1991], Lefik et al. [2009], ?) or recently variational principles (cf. Kirchdoerfer and Ortiz [2016]). In both cases, the hierarchy of the single-physics solid mechanics problems is simply a sequence illustrated in Figure 1.

In the single-physics solid mechanics problem, the relationship between strain and displacement, and between the balance of linear momentum and stress is considered “definition”, while the relationship between the stress and strain is clearly the only phenomenological component and therefore it is justified to be replaced by data-driven constitutive law [Ghaboussi et al., 1991, Lefik et al., 2009, Kirchdoerfer and Ortiz, 2016].



Fig. 1: Hierarchy of single-physics solid mechanics problem. Black arrow represents a definition or a “universal principle”; red arrow represents either a phenomenological relation or an operator that is defined not based on first principles.

This black-and-white classification between phenomenological law and universal principles is, nevertheless, not effective for multi-physical problems where physical quantities are linked by much complicated hierarchical relations, such as the example poromechanics problem shown in Figure 2.

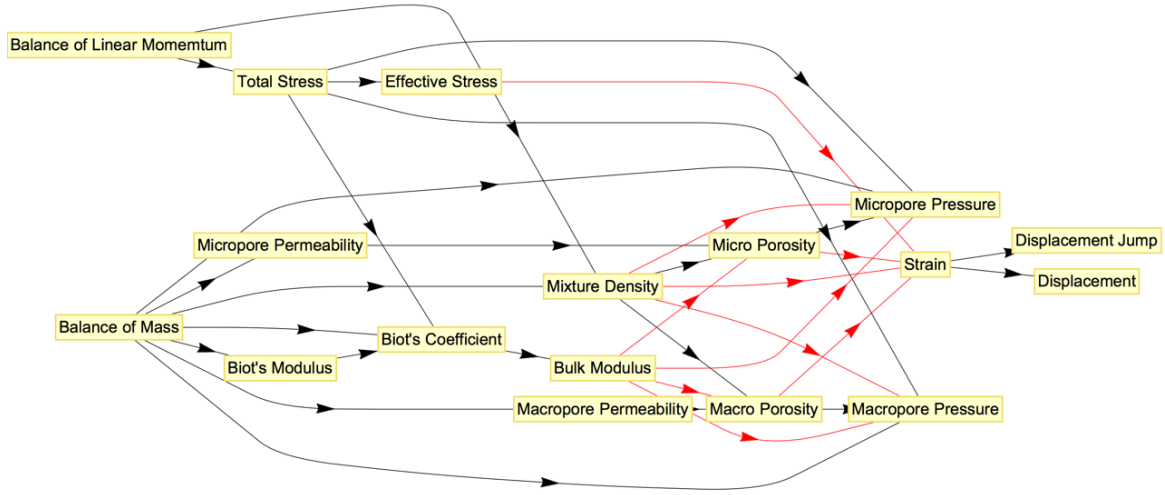


Fig. 2: Hierarchy of a multi-physics poromechanics problem for fluid-infiltrating dual-porosity media. Black arrow represents a definition or a “universal principle”; red arrow represents either a phenomenological relation or an operator that is defined not based on first principles.

This complexity leads to issues that are often neglected but could have important implication on the quality of simulations. One key issue is the consistency. In particular, computational models for multiscale poromechanics problem often rely on multiple constitutive laws to replicate the corresponding hydraulic and mechanical processes. For example, in a reservoir simulator, one may employ a cap-plasticity model such as those in Foster et al. [2005] and Sun et al. [2013a], to model the path-dependent behavior of porous rock, while use a retention model to predict the relations between degree of saturation and suction, and another hydraulic model to relate suction with the relative permeability [Choo et al., 2016]. However, as pointed out previously in a number of works, such as Zhu and Wong [1997] and Nuth and Laloui [2008], deformation of the solid skeleton may inevitably change the microstructural attributes, such as pore size distribution and tortuosity of the pore space and these changes may in return affect the pore pressure and hence the effective stress history. As a result, the assumption that the permeability, solid constitutive law and poroelasticity material parameters such as Biot’s modulus and Biot’s coefficient can be calibrated separately in a de-coupled manner may lack physical underpinning, especially when the microstructure of the solid skeleton is expected to undergo significant changes such as cracks, strain localization, and phase transition, that affect both the coupling mechanism and induce anisotropy in both the mechanical and hydraulic responses [Sun et al., 2011b, Guo and Zhao, 2013, Kuhn et al., 2015, Sun et al., 2016, Wang et al., 2016, Na and Sun, 2017].

One possible way to improve the consistency of the computational model is to directly incorporate the micro-mechanical data and direct numerical simulations on the microstructure to improve the consistency

of the hydro-mechanical coupling model. For instance, one may replace constitutive law with a calibrated micro-mechanical simulations for field scale problems as demonstrated in Miehe [2003], Nitka et al. [2011], Sun et al. [2013b], Guo et al. [2016] or they can be concurrently coupled with the macroscopic model via a handshake transition domain [Wellmann and Wriggers, 2012, Sun and Mota, 2014, Sun et al., 2016]. In both cases, the major technical challenge is the prohibitively high computational cost to run micro-mechanical simulations for field scale problems and the difficulty to store, post-process and analyze the large amount of data generated from numerical simulations.

The objective of this paper is to present a modeling framework in which a data-driven model is used as a mean to link information across multiple scales via offline training. This approach has two distinct advantages. First, unlike the classical hierarchical coupling methods such as Feyel and Chaboche [2000], Miehe and Bayreuther [2007], Geers et al. [2010] where incremental constitutive updates of each integration points in the macroscopic model are driven by imposing increment changes on the boundary conditions of the RVE simulations, the data-driven approach only requires the generation of database, which can take place offline and not during the numerical simulations. In other words, once the machine learning processes are completed, the users of the multiscale model may run multiscale simulations without running the RVE simulations again. As a result, the computational cost of the data-driven multiscale model can be so significantly reduced that it becomes possible to link simulations of more than two scales together for field applications. Second, unlike the other offline approach in which the accuracy, robustness and sophistication of the multiscale simulations are all limited by the quality of the macroscopic surrogate models such as the phenomenological models for interface element (e.g. Zhou et al. [2009]) or bulk element (e.g. Liu et al. [2016]), the data-driven neural network models can be adaptive and the level of complexity of the data-driven model can be easily changed by adding more nodes and layers into the neural networks. Thirdly, since both the mechanical and hydraulic properties are both updated directly from the same micro-mechanical simulations, the consistency, and compatibility of the data-driven mechanical and hydraulic material laws are guaranteed. Finally, the data-driven approach also allows data from experiments and micro-mechanical simulations both be directly incorporated into the machine learning procedure to generate and verify the data-driven models.

The organization of the paper is as follows. First, we will review the field theory and the mathematical formulation of the dual-porosity dual-permeability poromechanics problem in the infinitesimal regime. Then we will describe the various supervised machine learning techniques we used to replace phenomenological constitutive laws with data-driven models that are trained by a combination of experimental data and RVE simulations. Following this step, we will introduce further details on the training procedures and the selection of the right machine learning method. In particular, we will provide detailed account on the usage of both the classical artificial neural network and recently developed recurrent neural network architectures to generate data-driven models as surrogates for linking multiple scales. The potential of deep learning for computational poromechanics modeling will also be discussed. As for notations and symbols, bold-faced letters denote tensors; the symbol \cdot denotes a single contraction of adjacent indices of two tensors (e.g. $\mathbf{a} \cdot \mathbf{b} = a_i b_i$ or $\mathbf{c} \cdot \mathbf{d} = c_{ij} d_{jk}$); the symbol $:$ denotes a double contraction of adjacent indices of tensor of rank two or higher (e.g. $\mathbf{C} : \mathbf{e}^e = C_{ijkl} e_{kl}^e$); the symbol \otimes denotes a juxtaposition of two vectors (e.g. $\mathbf{a} \otimes \mathbf{b} = a_i b_j$) or two symmetric second order tensors (e.g. $(\boldsymbol{\alpha} \otimes \boldsymbol{\beta}) = \alpha_{ij} \beta_{kl}$). As for sign conventions, unless specified otherwise, we consider the direction of the tensile stress and dilative pressure as positive.

2 Problem Statements

To test the applicability of the graph-based machine learning model, we select the simulations of hydro-mechanical coupling effect of deformable multi-permeability porous media, a sufficiently complex multi-physical problem that has great demand and common in engineering applications, as the test bed. Multi-porosity/multi-permeability porous media are materials consisting of more than one co-existing pore systems. These pore systems can be isolated pores or interconnected, as shown in Figure 3 [Elsworth and Bai, 1992, Zimmerman et al., 1993, Ji et al., 2015, Kuhlman et al., 2015, Choo et al., 2016]. In a multi-permeability model, one does not seek to obtain a single effective permeability for the entire pore space. Instead, the entire pore space is partitioned into two or more sub-systems based on the distinctive pore space sizes (e.g. limestone, carbonate rock) or geometrical features (e.g. permeable host matrix with cracks and joints).

164 These co-existing pore systems may exchange fluid mass, depending on the pore connectivity. Multi-
 165 porosity/multi-permeability models are important for characterizing reservoirs, containment transport,
 166 re-activation of faults and geological disposal of CO₂ and nuclear waste.

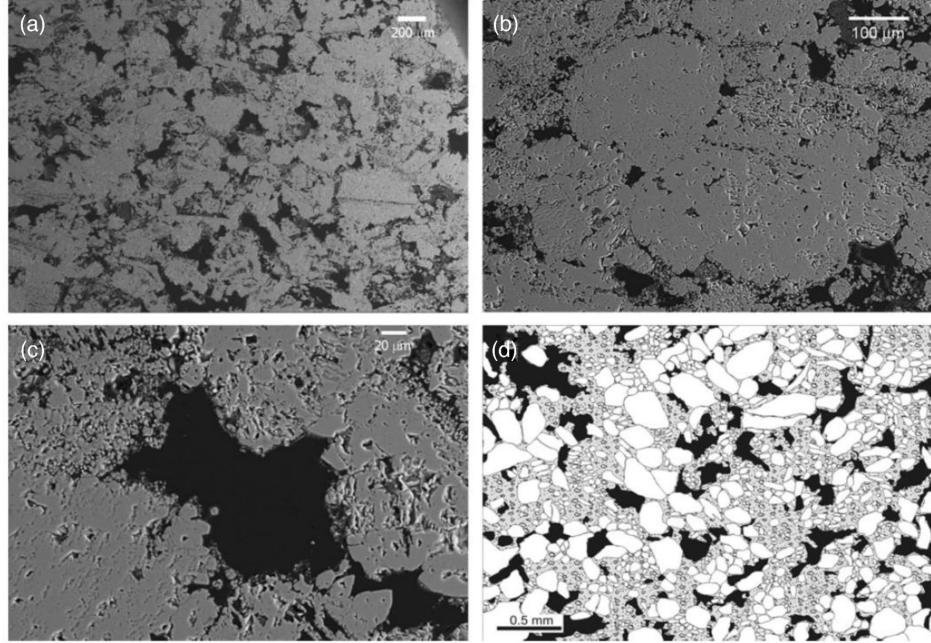


Fig. 3: Backscattered SEM images of an intact sample of Majella limestone where pores of multiple sizes are clearly observed. Figure adapted from [Ji et al. \[2015\]](#).

167 While multi-porosity and multi-permeability models are powerful approaches to incorporate micro-
 168 structural information for macroscopic modeling, the identification of material parameter and validation of
 169 the model is more difficult [[Köhne et al., 2006](#), [Auriault et al., 2010](#), [Lewandowska and Auriault, 2013](#)]. This
 170 issue becomes even more profound when the deformation of the solid skeleton becomes non-negligible
 171 due to plastic deformation, crack growths or other changes of microstructures that might lead to complex
 172 anisotropic evolutions of the effective permeability in multiple porous systems inside the porous media.
 173 While multiscale modeling has found to be able to provide insight on the evolving microstructural at-
 174 tributes for single-porosity system as shown in [Sun et al. \[2011b,a\]](#), [Liu et al. \[2015\]](#), [Wang and Sun \[2015,](#)
 175 [2016b,a\]](#), the hierarchical multi-scale computation is often too expensive to link simulations across more
 176 than one scales and hence not suitable for more complex problems in which information across multiple
 177 scales is important for the macroscopic outcome. In this work, the aforementioned issues are resolved by
 178 introducing deep learning as an offline scale-bridging tool for multi-porosity system. For completeness,
 179 we will provide a brief overview of the mathematical framework of the multiscale modeling framework,
 180 which is extended from the single-porosity counterpart documented in [Wang and Sun \[2016b\]](#).

181 2.1 Balance Principle

182 For completeness, we briefly outline the initial boundary value problem that simulates the deformation-
 183 diffusion coupling mechanism in deformable multi-porosity media. Here we consider a saturated porous
 184 media occupying a spatial domain $\mathcal{B} \subset \mathbb{R}^{n_{sd}}$ where $n_{sd} = 1, 2$, or 3 stands for the number of spatial di-
 185 mensions. The boundary of the body is denoted as $\partial\mathcal{B} \subset \mathbb{R}^{n_{sd}-1}$. The porous solid is treated as a double-
 186 porosity mixture and the two dominant pore scales are the macropores M (the pores in fissures, dila-

187 tant shear bands, cracks, ...) and the micropores m (the solid matrix pores). The elementary volumes of
 188 the total mixture, solid, macropores, micropores and void are therefore denoted as dV , dV_s , dV_M , dV_m ,
 189 $dV_v = dV - dV_s = dV_M + dV_m$, respectively. The pore fractions for macropores (ψ^M) and micropores (ψ^m)
 190 are defined as,

$$\psi^M = \psi = \frac{dV_M}{dV_v}, \quad \psi^m = 1 - \psi = \frac{dV_m}{dV_v}. \quad (1)$$

191 Meanwhile, the volume fractions for solid, macropores and micropores can be expressed as a function of
 192 porosity $\phi = \frac{dV_v}{dV}$ and pore fractions, i.e.,

$$\phi^s = \frac{dV_s}{dV} = 1 - \phi, \quad \phi^M = \frac{dV_M}{dV} = \phi\psi, \quad \phi^m = \frac{dV_m}{dV} = \phi(1 - \psi). \quad (2)$$

193 The partial densities of each constituent can be determined using the volume fractions and the intrinsic
 194 mass densities of solid ρ_s and fluid ρ_f :

$$\rho^s = (1 - \phi)\rho_s, \quad \rho^M = \phi\psi\rho_f, \quad \rho^m = \phi(1 - \psi)\rho_f. \quad (3)$$

195 Then the total mass density of the mixture is given by

$$\rho = \rho^s + \rho^M + \rho^m. \quad (4)$$

196 Assume that the solid skeleton deformation is infinitesimal, hence the difference between the reference
 197 and spatial configuration is neglected and $\det F = J \approx 1$. In this case, we may denote the material time
 198 derivative following the solid velocity $\mathbf{v} = \frac{\partial \mathbf{u}}{\partial t}$ as $(\dot{\bullet}) = \frac{\partial(\bullet)}{\partial t} + \nabla(\bullet) \cdot \mathbf{v}$. Assuming incompressible solid
 199 and fluid constituents, negligible inertial terms and no mass exchange between solid and fluid, following
 200 the formulation of [Choo and Borja \[2015\]](#), the strong form of the problem reads: find the displacement $\mathbf{u} : \mathcal{B} \rightarrow \mathbb{R}^{n_{sd}}$, the Cauchy macropore pressure $p_M : \mathcal{B} \rightarrow \mathbb{R}$ and the Cauchy micropore pressure $p_m : \mathcal{B} \rightarrow \mathbb{R}$
 201 such that the balance of linear momentum, the balance of fluid mass in macropores, the balance of fluid
 202 mass in micropores and the boundary conditions are satisfied:

$$\left\{ \begin{array}{ll} \nabla \cdot \boldsymbol{\sigma} + \rho \mathbf{g} = c_0(\tilde{\mathbf{v}}_m - \tilde{\mathbf{v}}_M) & \text{on } \mathcal{B}, \\ \rho_f \psi \nabla \cdot \mathbf{v} + \nabla \cdot \mathbf{q}_M = -c_0 & \text{on } \mathcal{B}, \\ \rho_f (1 - \psi) \nabla \cdot \mathbf{v} + \nabla \cdot \mathbf{q}_m = c_0 & \text{on } \mathcal{B}, \\ \mathbf{u} = \bar{\mathbf{u}} & \text{on } \partial \mathcal{B}_u \\ \boldsymbol{\sigma} \cdot \mathbf{n} = \bar{\mathbf{t}} & \text{on } \partial \mathcal{B}_t, \\ p_M = \bar{p}_M & \text{on } \partial \mathcal{B}_{p_M}, \\ \mathbf{q}_M \cdot \mathbf{n} = -\bar{q}_M & \text{on } \partial \mathcal{B}_{q_M}, \\ p_m = \bar{p}_m & \text{on } \partial \mathcal{B}_{p_m}, \\ \mathbf{q}_m \cdot \mathbf{n} = -\bar{q}_m & \text{on } \partial \mathcal{B}_{q_m}. \end{array} \right. \quad (5)$$

204 $\boldsymbol{\sigma}$ denotes the total Cauchy stress tensor and the effective Cauchy stress tensor is given by,

$$\boldsymbol{\sigma}' = \boldsymbol{\sigma} + \bar{p} \mathbf{1} = \boldsymbol{\sigma} + [\psi p_M + (1 - \psi) p_m] \mathbf{1}. \quad (6)$$

205 c_0 is the fluid mass transfer coefficient between the macropores and micropores. $\tilde{\mathbf{v}}_M = \mathbf{v}_M - \mathbf{v}$, $\tilde{\mathbf{v}}_m =$
 206 $\mathbf{v}_m - \mathbf{v}$ are the relative fluid velocities in macropores and micropores, respectively. A semi-empirical con-
 207 stitutive equation is adopted for c_0 :

$$c_0 = \frac{\bar{\alpha}}{\mu_f} (p_M - p_m), \quad (7)$$

208 where $\bar{\alpha}$ is a parameter characterizing the interface permeability between the macropores and micropores.
 209 μ_f is the dynamic viscosity of the fluid.

210 The relative fluid mass fluxes $\mathbf{q}_M, \mathbf{q}_m$ are related to Cauchy pore pressures via Darcy's law:

$$\begin{aligned}\mathbf{q}_M &= \rho_f \phi \psi \tilde{\mathbf{v}}_M = -\rho_f \frac{\mathbf{k}_M}{\mu_f} \cdot (\nabla p_M - \rho_f \mathbf{g}), \\ \mathbf{q}_m &= \rho_f \phi (1 - \psi) \tilde{\mathbf{v}}_m = -\rho_f \frac{\mathbf{k}_m}{\mu_f} \cdot (\nabla p_m - \rho_f \mathbf{g}).\end{aligned}\quad (8)$$

211 where \mathbf{k}_M and \mathbf{k}_m are intrinsic permeability tensors at macro-scale pore and micro-scale pore. \mathbf{g} is the
212 gravity acceleration vector.

213 The boundary $\partial\mathcal{B}$ having unit normal \mathbf{n} at $\mathbf{x} \in \partial\mathcal{B}$ admits the decomposition

$$\begin{cases} \partial\mathcal{B} = \overline{\partial\mathcal{B}_u} \cup \overline{\partial\mathcal{B}_t} = \overline{\partial\mathcal{B}_{p_M}} \cup \overline{\partial\mathcal{B}_{q_M}} = \overline{\partial\mathcal{B}_{p_m}} \cup \overline{\partial\mathcal{B}_{q_m}} \\ \emptyset = \partial\mathcal{B}_u \cap \partial\mathcal{B}_t = \partial\mathcal{B}_{p_M} \cap \partial\mathcal{B}_{q_M} = \partial\mathcal{B}_{p_m} \cap \partial\mathcal{B}_{q_m}, \end{cases}\quad (9)$$

214 where $\partial\mathcal{B}_u, \partial\mathcal{B}_{p_M}$ and $\partial\mathcal{B}_{p_m}$ are Dirichlet boundaries with solid displacement $\bar{\mathbf{u}}$, macropore pressure \bar{p}_M
215 and micropore pressure \bar{p}_m prescribed. $\partial\mathcal{B}_t, \partial\mathcal{B}_{q_M}$ and $\partial\mathcal{B}_{q_m}$ are Neumann boundaries with tractions $\bar{\mathbf{t}}$,
216 macropore flux \bar{q}_M and micropore flux \bar{q}_m prescribed.

217 Meanwhile, the initial conditions are imposed as

$$\mathbf{u}(\mathbf{x}, t = t_0) = \mathbf{u}_0(\mathbf{x}), \quad p_M(\mathbf{x}, t = t_0) = p_{M_0}(\mathbf{x}), \quad p_m(\mathbf{x}, t = t_0) = p_{m_0}(\mathbf{x}) \quad \text{for all } \mathbf{x} \in \mathcal{B} \text{ at } t = t_0 \quad (10)$$

218 2.2 Embedded strong discontinuities for displacement and pore-fluid flux jump

219 In this work, the supervised machine learning procedure is used to generate constitutive laws for the
220 strong discontinuities. As a result, we briefly review the kinematics and constitutive responses of dual-
221 permeability porous media with strong discontinuity.

222 Consider a given material point $\mathbf{X} \in \mathcal{B}$ and an associated local neighborhood $\mathcal{B}_X \subset \mathcal{B}$ crossed by a
223 strong discontinuity (fracture, shear band, fault, etc.). Denote the surface of the discontinuity as Γ and the
224 local domain \mathcal{B}_X is thus divided by Γ into subdomain pairs $\mathcal{B}_X = \mathcal{B}_X^+ \cup \mathcal{B}_X^-$. The motion of the particles
225 within \mathcal{B}_X is described by local displacement field $\mathbf{u}_\mu = \mathbf{x}_\mu - \mathbf{X} = \boldsymbol{\varphi}_\mu(\mathbf{X}) - \mathbf{X}$. Assume the following
226 relation to the large-scale (or conformal) displacement field \mathbf{u} ,

$$\mathbf{u}_\mu = \mathbf{u} + \llbracket \mathbf{u} \rrbracket (H_\Gamma - f_\Gamma), \quad (11)$$

227 where $\llbracket \mathbf{u} \rrbracket$ is the displacement jump across the interface Γ , H_Γ is the Heaviside step function across Γ and
228 f_Γ is a smooth ramp function in \mathcal{B}_X specified in [Borja, 2000]. It is also useful to define the continuous part
229 $\bar{\mathbf{u}}$ of motion \mathbf{u}_μ as

$$\bar{\mathbf{u}} = \mathbf{u} - \llbracket \mathbf{u} \rrbracket f_\Gamma. \quad (12)$$

230 The large-scale (or conformal) and continuous infinitesimal strains can be defined as,

$$\boldsymbol{\epsilon} = \nabla^s \mathbf{u} = \frac{1}{2} (\nabla \mathbf{u} + (\nabla \mathbf{u})^T), \quad \bar{\boldsymbol{\epsilon}} = \nabla^s \bar{\mathbf{u}}. \quad (13)$$

231 The local infinitesimal strain is given by, assuming relative uniformity of $\llbracket \mathbf{u} \rrbracket$ along Γ such that $\nabla^X \llbracket \mathbf{u} \rrbracket \rightarrow$
232 $\mathbf{0}$ in \mathcal{B}_X ,

$$\begin{aligned}\boldsymbol{\epsilon}_\mu &= \nabla^s \mathbf{u}_\mu \\ &= \bar{\boldsymbol{\epsilon}} + (\llbracket \mathbf{u} \rrbracket \otimes \mathbf{n})^s \delta_\Gamma \\ &= \boldsymbol{\epsilon} - (\llbracket \mathbf{u} \rrbracket \otimes \nabla f_\Gamma)^s + (\llbracket \mathbf{u} \rrbracket \otimes \mathbf{n})^s \delta_\Gamma,\end{aligned}\quad (14)$$

233 where the equation $\nabla H_\Gamma = \delta_\Gamma \mathbf{n}$ is employed. δ_Γ is the Dirac delta function across Γ and \mathbf{n} is the unit normal
234 of Γ pointing from \mathcal{B}_X^- to \mathcal{B}_X^+ .

235 The porous media in the vicinity of strong discontinuity Γ is composed of macro-scale voids gener-
236 ated by cracking, shear band formation and micro-scale voids inside the underlying solid matrix. In other
237 words, we assume that the strong discontinuities are only captured by the macropores and the micropores

remain continuous. The motion of the fluid flow in the local neighborhood \mathcal{B}_X in both scales are characterized by macropore fluid flow \mathbf{q}_{M_μ} and micropore fluid flow \mathbf{q}_{m_μ} [Callari and Armero, 2002]:

$$\begin{aligned}\mathbf{q}_{M_\mu} &= \mathbf{q}_M + \llbracket \mathbf{q}_M \rrbracket (H_\Gamma - f_\Gamma), \\ \mathbf{q}_{m_\mu} &= \mathbf{q}_m + \mathbf{0} (H_\Gamma - f_\Gamma).\end{aligned}\tag{15}$$

where \mathbf{q}_M and \mathbf{q}_m are regular flow in the macropores and micropores, respectively. We assume that the macro fluid flow field exhibit a jump $\llbracket \mathbf{q}_M \rrbracket$ across Γ , while the micro fluid flow does not. For convenience, the ramp function f_Γ used to embed the strong discontinuity of the displacement field is also used to reproduce the jump of the fluid flux. Assuming that the pore fluid is incompressible, the local balance equations of the macro-fluid content M_{M_μ} and local micro-fluid content M_{m_μ} (fluid mass increment per unit reference volume of porous solid) read,

$$\begin{aligned}\dot{M}_{M_\mu} &= -\nabla \cdot \mathbf{q}_{M_\mu} = -\nabla \cdot \mathbf{q}_M + \llbracket \mathbf{q}_M \rrbracket \cdot \nabla f_\Gamma - (\llbracket \mathbf{q}_M \rrbracket \cdot \mathbf{n}) \delta_\Gamma \\ \dot{M}_{m_\mu} &= -\nabla \cdot \mathbf{q}_{m_\mu} = -\nabla \cdot \mathbf{q}_m.\end{aligned}\tag{16}$$

with the assumption that $\nabla \cdot \llbracket \mathbf{q}_M \rrbracket \rightarrow 0$ in \mathcal{B}_X .

The solution of the local displacement jump field $\llbracket \mathbf{u} \rrbracket$ requires a local equilibrium equation relating the total stress field in $\mathcal{B}_X \setminus \Gamma$ and the total traction across Γ driving the mechanical inelastic effects inside the strong discontinuity. The weak form writes, for all variations $\delta \llbracket \mathbf{u} \rrbracket$, [Callari and Armero, 2002]

$$-\frac{1}{V_{\mathcal{B}_X}} \int_{\mathcal{B}_X} \delta \llbracket \mathbf{u} \rrbracket \cdot \boldsymbol{\sigma} \mathbf{n} dV + \frac{1}{L_\Gamma} \int_\Gamma \delta \llbracket \mathbf{u} \rrbracket \cdot \mathbf{T}_\Gamma d\Gamma = 0\tag{17}$$

where $V_{\mathcal{B}_X} = \text{measure}(\mathcal{B}_X)$ and $L_\Gamma = \text{measure}(\Gamma)$. Under the assumption that the macropore fluid flux is discontinuous, then the pressure is continuous while the pressure gradient is discontinuous across the strong discontinuity. The local equilibrium equation can be written in terms of effective stress and effective traction, i.e.,

$$-\frac{1}{V_{\mathcal{B}_X}} \int_{\mathcal{B}_X} \delta \llbracket \mathbf{u} \rrbracket \cdot \boldsymbol{\sigma}' \mathbf{n} dV + \frac{1}{L_\Gamma} \int_\Gamma \delta \llbracket \mathbf{u} \rrbracket \cdot \mathbf{T}'_\Gamma d\Gamma = 0\tag{18}$$

In the limit $\frac{V_{\mathcal{B}_X}}{L_\Gamma} \rightarrow 0$:

$$\mathbf{T}'_\Gamma(\llbracket \mathbf{u} \rrbracket) = (\boldsymbol{\sigma}'(\bar{\boldsymbol{\epsilon}}) \cdot \mathbf{n})|_\Gamma.\tag{19}$$

The constitutive relation for the effective traction across the discontinuity comes from the data-driven model trained with data from sub-scale simulations, as shown in Section 5. Meanwhile, the rate form of the constitutive relation for the stress in the continuum writes:

$$\boldsymbol{\sigma}'(\bar{\boldsymbol{\epsilon}}) = \mathbf{C}^e : \bar{\boldsymbol{\epsilon}} = \mathbf{C}^e : \boldsymbol{\epsilon} - \mathbf{C}^e : (\llbracket \mathbf{u} \rrbracket \otimes \nabla f_\Gamma)^s,\tag{20}$$

where \mathbf{C}^e is the rank-four elastic moduli tensor.

According to Borja [2000], static condensation can be performed on Eq. 19 and the balance of linear momentum in Eq. 5 for constant triangle elements used in this work. Thus, within each iteration step for the solution of the global equations Eq. 5, Eq. 19 can be solved locally to get $\llbracket \mathbf{u} \rrbracket$ at each Gauss point, assuming constant $\boldsymbol{\epsilon}$. Moreover, following the derivations in Callari and Armero [2002], which has the same assumption of fluid flux discontinuity as in this work, the balance of fluid mass in macropores and micropores in Eq. 5 remain unmodified.

Here, we define the displacement, macropore pressure and micropore pressure trial spaces as

$$\begin{aligned}V_u &= \{\mathbf{u} : \mathcal{B} \rightarrow \mathbb{R}^{n_{sd}} \mid \mathbf{u} \in H^1, \mathbf{u}|_{\partial B_u} = \bar{\mathbf{u}}\} \\ V_{p_M} &= \{p_M : \mathcal{B} \rightarrow \mathbb{R} \mid p_M \in H^1, p_M|_{\partial B_{p_M}} = \bar{p}_M\} \\ V_{p_m} &= \{p_m : \mathcal{B} \rightarrow \mathbb{R} \mid p_m \in H^1, p_m|_{\partial B_{p_m}} = \bar{p}_m\}\end{aligned}\tag{21}$$

266 and the corresponding admissible spaces of variations as

$$\begin{aligned} V_\eta &= \{\eta : \mathcal{B} \rightarrow \mathbb{R}^{n_{sd}} \mid \eta \in H^1, \eta|_{\partial B_u} = \mathbf{0}\} \\ V_{\zeta_M} &= \{\zeta_M : \mathcal{B} \rightarrow \mathbb{R} \mid \zeta_M \in H^1, \zeta_M|_{\partial B_{p_M}} = 0\} \\ V_{\zeta_m} &= \{\zeta_m : \mathcal{B} \rightarrow \mathbb{R} \mid \zeta_m \in H^1, \zeta_m|_{\partial B_{p_m}} = 0\} \end{aligned} \quad (22)$$

267 where H^1 denotes the Sobolev space of functions of degree one. Using the backward Euler implicit scheme
268 of step size $\Delta t = t_{n+1} - t_n$, the time-integrated variational form of the large-scale problem is constructed
269 as: find $\mathbf{u} \in V_u$, $p_M \in V_{p_M}$ and $p_m \in V_{p_m}$ such that for all $\eta \in V_\eta$, $\zeta_M \in V_{\zeta_M}$ and $\zeta_m \in V_{\zeta_m}$

$$\begin{aligned} G : V_u \times V_{p_M} \times V_{p_m} \times V_\eta &\rightarrow \mathbb{R} \\ G(\mathbf{u}, p_M, p_m, \eta) &= \underbrace{\int_{\mathcal{B}} \nabla^s \eta : \sigma' dV}_{g_1} - \underbrace{\int_{\mathcal{B}} \nabla \cdot \eta \bar{p} dV}_{g_2} \\ &\quad - \underbrace{\int_{\mathcal{B}} \eta \cdot \rho g dV}_{g_3} + \underbrace{\int_{\mathcal{B}} \eta \cdot c_0 (\tilde{v}_m - \tilde{v}_M) dV}_{g_4} - \underbrace{\int_{\partial B_t} \eta \cdot \bar{t} dA}_{g_{ext}} = 0 \end{aligned} \quad (23)$$

$$\begin{aligned} H_{p_M}^{\Delta t} : V_u \times V_{p_M} \times V_{p_m} \times V_{\zeta_M} &\rightarrow \mathbb{R} \\ H_{p_M}^{\Delta t}(\mathbf{u}, p_M, p_m, \zeta_M) &= \underbrace{\int_{\mathcal{B}} \zeta_M \rho_f \psi \nabla \cdot (\mathbf{u} - \mathbf{u}_n) dV}_{h_1^M} \\ &\quad + \underbrace{\Delta t \rho_f \int_{\mathcal{B}} \nabla \zeta_M \cdot \frac{\mathbf{k}_M}{\mu_f} \cdot (\nabla p_M - \rho_f \mathbf{g}) dV}_{h_2^M} \\ &\quad + \underbrace{\Delta t \int_{\mathcal{B}} \zeta_M c_0 dV}_{h_3^M} - \underbrace{\Delta t \int_{\partial B_{q_M}} \zeta_M \bar{q}_M dA}_{h_{ext}^M} = 0 \end{aligned} \quad (24)$$

$$\begin{aligned} H_{p_m}^{\Delta t} : V_u \times V_{p_M} \times V_{p_m} \times V_{\zeta_m} &\rightarrow \mathbb{R} \\ H_{p_m}^{\Delta t}(\mathbf{u}, p_M, p_m, \zeta_m) &= \underbrace{\int_{\mathcal{B}} \zeta_m \rho_f (1 - \psi) \nabla \cdot (\mathbf{u} - \mathbf{u}_n) dV}_{h_1^m} \\ &\quad + \underbrace{\Delta t \rho_f \int_{\mathcal{B}} \nabla \zeta_m \cdot \frac{\mathbf{k}_m}{\mu_f} \cdot (\nabla p_m - \rho_f \mathbf{g}) dV}_{h_2^m} \\ &\quad - \underbrace{\Delta t \int_{\mathcal{B}} \zeta_m c_0 dV}_{h_3^m} - \underbrace{\Delta t \int_{\partial B_{q_m}} \zeta_m \bar{q}_m dA}_{h_{ext}^m} = 0 \end{aligned} \quad (25)$$

270 To the best knowledge of the authors, this work is the first recorded implementation of assumed strain
271 hydro-mechanical model for dual-permeability dual-porosity system with embedded strong discontinu-
272 ities in both the displacement and flux fields. Nevertheless, due to the main focus of this work is on the
273 incorporation of the machine learning method for scale bridging, we decided to omit the details of the im-
274 plementation of the assumed strain dual-permeability model. Interested readers may refer to a number of
275 related works on assumed strain formulation such as [Callari and Armero \[2002\]](#), [Borja \[2000\]](#), [Mosler and](#)
276 [Meschke \[2003\]](#), [Callari et al. \[2010\]](#) for details. Furthermore, as we employ equal-order discretization for
277 the solid displacement and the pore pressure in the macro- and micro-pores, the lack of two-fold inf-sup
278 condition must be circumvented. For brevity, the discussion of the stabilization procedure is not included

but can be found in earlier works such as [White and Borja \[2008\]](#), [Sun et al. \[2013c\]](#), [Sun \[2015\]](#), [Choo and Borja \[2015\]](#), [Krischok and Linder \[2016\]](#).

3 Architecture of directed graph for data-driven poromechanics problems

In the classical hierarchical multiscale framework, such as FEM² (cf. [Frankenreiter et al. \[2011\]](#), [Fish \[2013\]](#)) and DEM-FEM ([Liu et al. \[2015\]](#), [Wang and Sun \[2016b,a\]](#)), the multiscale simulations are conducted by replacing constitutive law with representative elementary volume (REV) simulations that provide the incremental constitutive updates at each integration point. This method is typically much more cost efficient than the direct numerical simulation, as the micro-mechanical simulations are confined in the REV domain rather than conducted in the entire physical domain. However, the computational cost is typically much higher than the conventional constitutive law driven method [[Liu et al., 2015](#), [Guo et al., 2016](#)].

This computational cost is perhaps feasible for two-scale coupling simulations, but it may become a severe computational barrier if one attempts to link simulations across more than one scale. The remedy to this issue can be classified as two approaches – the usage of surrogate model [[Keshavarz and Ghosh, 2013](#), [Liu et al., 2016](#)] or the usage of reduced-order modeling [[Fish and Wu, 2011](#)]. In the former case where surrogate model is used, the smaller scale simulations will often be used to generate a database aimed to record the homogenized responses of the representative elementary volume. This database can also be experimental data or a combination of both “real” experimental data and the “virtual” simulation data. This database is then split into two *mutually exclusive* subsets – One used to calibrate and identify material parameters via inverse problems; another one used for validation and performance assessment of the numerical models [[Liu et al., 2016](#)]. Nevertheless, the primary drawback of the surrogate-based approach, in particular in the cases where phenomenological models are used as surrogate model, is that the accuracy and efficiency are highly dependent on the quality of the surrogate models that replace the direct numerical simulations (DNS). Furthermore, this approach often requires multiple surrogate models for multiphysics problem that might not be consistent with each other. This issue is particularly common for poromechanics problems (e.g. [Shahir et al. \[2012\]](#)) where the usage of kinematic hardening plasticity model coupling with isotropic permeability model often leads to the discrepancy that is hard to detect. Even worse, the introduction of multiple material parameters may make it easier to complete curve-fitting for a model that lacks prediction capacity otherwise.

Here we limit our focus on a hierarchical multiscale coupling approach in which recurrent neural network trained by a supervised deep learning is used as surrogate model to deliver constitutive responses, from solution database [[Tawhai et al., 2009](#), [Jain and Ghosh, 2009](#)]. Our major point of departure is the use of a graph-based concept previously presented in [Pawlowski et al. \[2012\]](#), [Sun \[2015\]](#) and [Salinger et al. \[2016\]](#) to design the information flow from smaller to larger scales and the use of recurrent neural network to automatically generate **one** surrogate model that provide the updates for both the effective stress and permeability tensors in the macroscopic and microscopic pore space.

This directed graph represents the hierarchy of information processed in a computational model that utilizes a combination of classical and data-driven models. In the directed graph representation, physical quantities are viewed as vertices, while the relations among physical quantities are considered as edges that link those building blocks together to form a computational model. In the case where the directed graph represents the hierarchy of information of the initial boundary value problem, the most upstream vertices (also called the root in graph theory literature) would be the governing equations (e.g. balance principles, phase field evolution equations, etc), while the downstream vertices (also called the leaves in graph theory literature) are the unknown variables (e.g. displacement, pore pressure, temperature). Close examination of the information flow may help us distinguish the edges into 1. definitions (e.g. relation between deformation gradient and displacement field, relation between Biot’s coefficient and bulk moduli) 2. universal principles (e.g. effective stress principle – relationship among total stress, effective stress and the pore pressure of the macro- and micro-pores, balance principles) and 3. material laws, phenomenological relations or empirical rules (e.g. Darcy’s law, water retention curve, stress-strain relation), as shown in Figure 4.

Unlike the model-free approach in which the entire computational model is replaced by neural network (cf. the model-free approach e.g. [Graf et al. \[2010\]](#)), our approach is to keep the edges identified

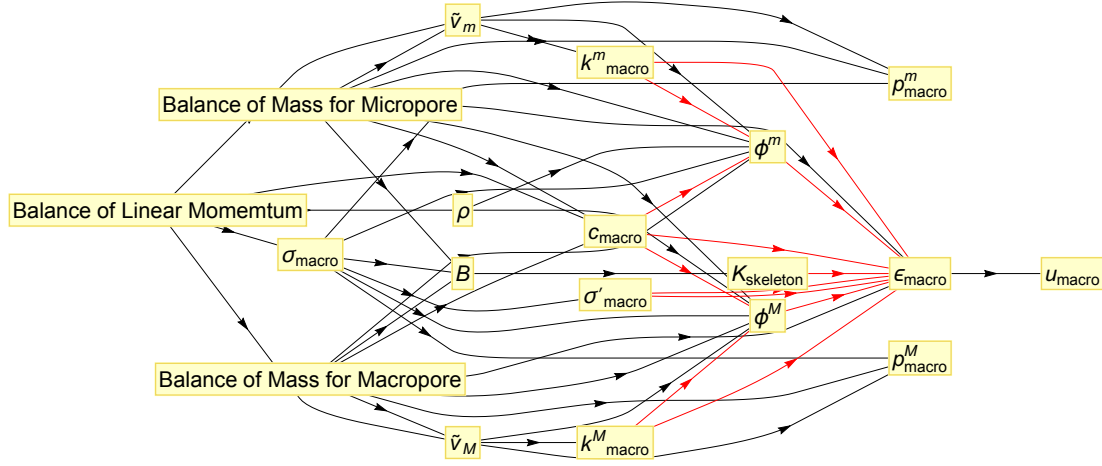


Fig. 4: Directed graph representing the information flow of the the multi-scale multi-physics dual-permeability poromechanics problem for fluid-infiltrating media. Red arrow represents either a phenomenological relation or an operator that is defined not by definition, universal law or first principle.

330 either as definitions or universal principles, but replace the edges that are commonly linked together by
 331 phenomenological models with data-driven model trained from deep learning. To do so, we first identify
 332 the subgraph (the graph formed by a subset of the vertices and edges of the graph that presents the com-
 333 puter model) in which the vertices are only connected by material laws. In this subgraph, we again identify
 334 the leaves and roots. For instance, in the elasticity problem example illustrated in 1, there will be only one
 335 leave (strain) and one root (stress).

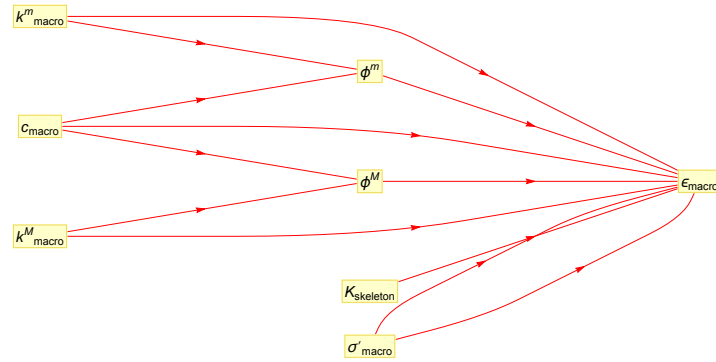


Fig. 5: Sub-graph of the multi-scale multi-physics poromechanics problem for fluid-infiltrating media. Red arrow represents either a phenomenological relation or an operator that is defined not based on first principles.

336 This information in return gives us an idea about what we should “learn” and what should be in
 337 the input and output in the supervised machine learning setting. In the dual-permeability poromechanics

problem, which we selected as test bed, we assume that the effective stress principle is valid for the dual-permeability system [Biot, 1941, Terzaghi, 1943, Callari and Federico, 2000, Coussy, 2004]. Furthermore, since we use a recurrent neural network for the supervised training, the time history of strain is not explicitly expressed as an additional vertex in the directed graph, as previously done in Lefik and Schrefler [2003] where feed-forward neural network is used. Rather, they are taken into account as internal state by default such that the path-dependent behavior of the dual-permeability porous material can be replicated. In the dual-permeability problem, we identify that macroscopic strain, ϵ_{macro} is the root and the effective stress σ'_{macro} , permeability of the macroscopic and microscope pores, k_{macro}^M , k_{macro}^m , the mass exchange rate c_{macro} , and the bulk modulus of solid skeleton K_{skeleton} , as shown in Figure 5. Once the input and output are determined, the rest of the task is to determine the appropriate model that gives us the output prediction when a specific input is given. Notice that it is also possible that the "material laws" of a multi-physical problem may also lead to multiple sub-graphs that share no vertex. In such sense, the procedure described above still applies, but the machine learning for each sub-graph will be independent to each other.

Another important observation is that it is not necessary to completely replace the subgraph with data-driven model. For instance, one may use the conventional material law to connect the strain and strain history with stress but use experimental data to generate a data-driven model that predicts the permeability from strain history. The optimization of the choice of the edges for the hybrid approach, especially when it is subjected to noisy data is an important topic but is out of the scope of this study.

3.1 Preparation of databases for offline hierarchical supervised machine learning

In the previous section, we discuss the anatomy of the mathematical model represented in a directed graph, and the method to identify the components to insert data-driven model properly in a multi-physical problem without altering the hierarchy and connectivity of the physical quantities. In this section, our goal is to focus on how to use sub-scale data to enhance the predictions via an offline hierarchical bridging scale method. In particular, we will review the difference of online and offline hierarchical multiscale approaches and procedure of generating pre-computation databases for fast or real-time multiscale simulations.

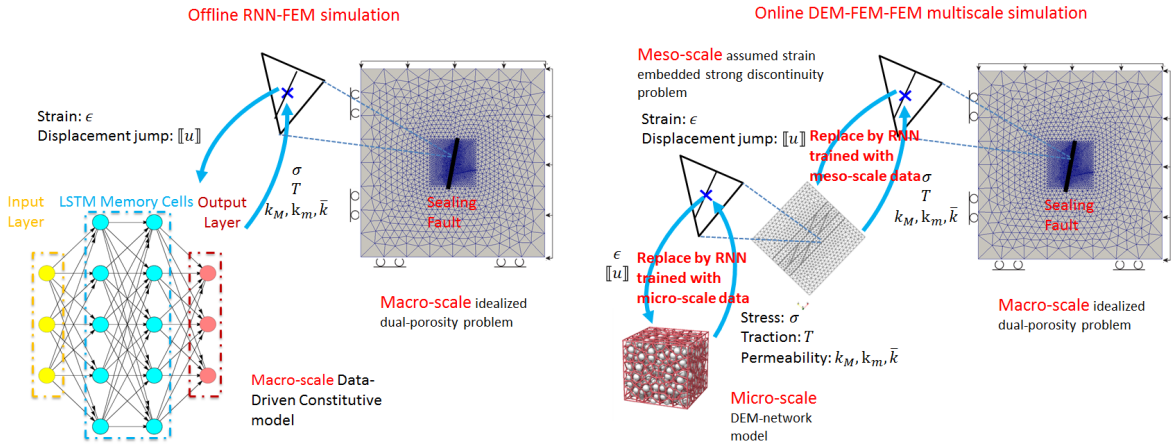


Fig. 6: Comparison between off-line pre-trained multiscale ANN-FEM simulations and online hierarchical multiscale simulations.

In an online hierarchical coupling approach in which simulations of different scales are linked, we might define macroscopic problems and microscopic problems and consider different technique to link the two problems in a hierarchical manner. For instance, the mechanical and hydro-mechanical responses of

granular materials can be replicated by a DEM-FEM coupled model in which the macroscopic material laws are homogenized from discrete element simulations. By associating each integration point with an RVE, the coupling of the micro-problem (DEM) and macro-problem (FEM) is established by replacing macroscopic phenomenological relations with DEM simulations for each incremental time step. [Miehe and Bayreuther, 2007, Nitka et al., 2011, Guo and Zhao, 2014, Liu et al., 2015, Wang and Sun, 2016a,b]. This approach is nevertheless very expensive, as each constitutive update at each integration point would require an individual sub-scale DEM simulation performed on a different RVE at each incremental step. While parallel implementation is efficient for the hierarchical DEM-FEM coupling approach (as the DEM simulations can be embarrassingly parallel), the total number of required DEM simulations still grows proportionally with the mesh size used in the macroscopic problems. This cost becomes more profound when information across more than two scales become important, as shown in the dual-porosity, dual-permeability poromechanics problem illustrated in Figure 6.

On the other hand, the offline hierarchical coupling method does not directly use the DEM simulations **during** the macroscopic simulations. Rather, it involves an additional step in which the sub-scale simulations are used to generate a database. This database is then used to calibrate a surrogate model that is sufficiently efficient for macroscopic boundary value problems. The surrogate model can be simply a phenomenological model [Keshavarz and Ghosh, 2013, Liu et al., 2016], a reduced-order sub-scale model [Fish, 2013, Yvonnet and He, 2007, Zahr et al., 2017] or a data-driven model [Kirchdoerfer and Ortiz, 2016]. The offline hierarchical technique, if conducted properly, has at least two advantages. First, it costs less as the offline techniques as it does not need on-the-fly sub-scale simulations. Second, the offline treatment provides an opportunity for one to combine real and virtual data together to improve the accuracy of numerical simulations.

As shown in Figure 6, we will leverage these advantages to conduct a simulation that links the hydro-mechanical simulations of fractured porous media across three scales. In particular, our objective is to introduce a recursive training procedure where the database generated from small-scale simulations would be used to train a meso-scale RNN data-driven model, while the meso-scale RNN-FEM model will be used to generate another database to train the macroscopic data-driven model used for field scale simulations, as shown in Figure 7. Here we first assume that the principle of separation of scale is valid such that,

$$l_{\text{micro}} \leq l_{\text{meso}} \leq l_{\text{macro}}. \quad (26)$$

Previously, Wang and Sun [2016b] has established a finite strain DEM-FEM coupling model to simulate two-phase poromechanics problem. Here we extend this work and focus on the case where intense localization due to damage or fractures occurs across the micro-, meso- and macroscopic scales. As a result, we require two sets of numerical simulations to generate the appropriate database to first link micro-mechanical DEM simulations to the meso-scale RVE, then again link the meso-scale DEM-FEM simulation to macroscopic scale field problem.

The first micro-mechanical simulation database consists of simulation results obtained from a DEM-network model in which DEM assemblies are subjected to different loading paths. The constitutive laws (traction-separation law and anisotropic permeability of macroscopic pore space) obtained from homogenizing the DEM responses are used as the data set for training and validating the neural network models. Here we assume that the mesoscale model employs a finite element discretization with displacement, pore pressure and their corresponding jumps as unknown in each incremental time step. Applying the effective stress principle, we postulate that there exists an effective stress such that it solely depends on the deformation and deformation history of the solid skeleton [Terzaghi, 1943, Biot, 1941].

In the infinitesimal regime, the Hill-Mandel micro-heterogeneity condition requires the volume average of the virtual power in the RVE to equal the virtual power done by the volume averages of power-conjugate stress and deformation measures. In terms of stress σ' and infinitesimal strain ϵ :

$$\langle \sigma' \rangle : \langle \dot{\epsilon} \rangle = \langle \sigma' : \dot{\epsilon} \rangle \quad (27)$$

The traction $\langle T'_\Gamma \rangle$ averaged in the RVE representing the interface is given by,

$$\langle T'_\Gamma \rangle = \langle \sigma' \rangle \cdot \mathbf{n}, \quad (28)$$

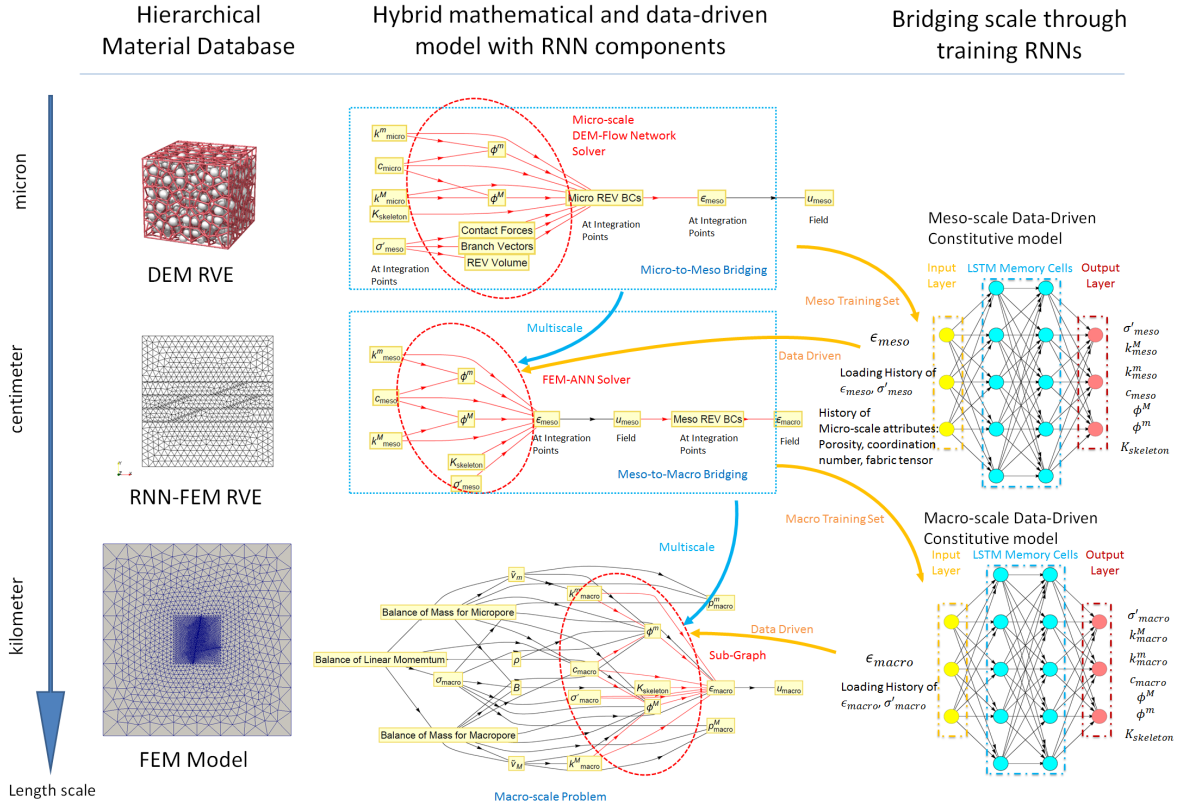


Fig. 7: Hierarchy of a multi-scale multi-physics poromechanics problem for fluid-infiltrating media. Black arrow represents a definition or a “universal principle”; red arrow represents either a phenomenological relation or an operator that is defined not based on first principles.

where \mathbf{n} is a unit vector normal to the interface and \mathbf{m} is a unit vector tangential to the interface. The average infinitesimal strain is defined in terms of $\llbracket \mathbf{u} \rrbracket$ and the initial height of the RVE h_0 :

$$\langle \epsilon \rangle = \text{sym} \left(\frac{1}{h_0} \llbracket \mathbf{u} \rrbracket \otimes \mathbf{n} \right). \quad (29)$$

Thus, the Hill-Mandel lemma in the interface in terms of $\langle T_{\Gamma}^t \rangle$ and $\llbracket \mathbf{u} \rrbracket$ is given by:

$$h_0 \langle \sigma' : \dot{\epsilon} \rangle = \langle T_{\Gamma}^t \rangle \cdot \llbracket \dot{\mathbf{u}} \rrbracket. \quad (30)$$

In the infinitesimal regime, the time derivative of displacement jump reads,

$$\llbracket \dot{\mathbf{u}} \rrbracket = \llbracket \dot{\mathbf{u}} \rrbracket_n \mathbf{n} + \llbracket \dot{\mathbf{u}} \rrbracket_m \mathbf{m}. \quad (31)$$

The Hill-Mandel lemma for the interface therefore can be rewritten as,

$$h_0 \langle \sigma' : \dot{\epsilon} \rangle = \langle T_n \rangle \llbracket \dot{\mathbf{u}} \rrbracket_n + \langle T_m \rangle \llbracket \dot{\mathbf{u}} \rrbracket_m. \quad (32)$$

According to Eq. (28), effective traction in the normal and tangential direction can be written as,

$$\begin{aligned} \langle T_n \rangle &= \langle \sigma' \rangle \cdot \mathbf{n} \cdot \mathbf{n} \\ \langle T_m \rangle &= \langle \sigma' \rangle \cdot \mathbf{n} \cdot \mathbf{m} \end{aligned} \quad (33)$$

where the overall effective stress is given by,

$$\langle \sigma' \rangle = \frac{1}{V_0} \sum_c^{N_c} f^c \otimes l^c \quad (34)$$

where V_0 is the initial volume of the RVE. f^c is the contact force at the grain contact $x + y_c \in \mathbb{R}^3$. l^c is the branch vector, the vector that connects the centroids of two grains forming the contact. N_c is the total number of particles in the RVE. Among the admissible boundary conditions fulfilling the Hill-Mandel micro-heterogeneity condition, we adopt the periodic boundary conditions [Miehe et al., 2010], where the motion of a particle on the boundary of the RVE is characterized by, assuming rigid particles,

$$x(X) = \langle F \rangle \cdot X_c + w_c + R \cdot (X - X_c) \quad (35)$$

where X_c is the initial position of particle center, w_c is the displacement fluctuation and R describes the particle rotation. For a pair of particles on opposite boundaries ∂V^+ and ∂V^- , the periodicity enforces the periodicity of fluctuations and rotations

$$w_c^- = w_c^+, \quad R^- = R^+, \quad (36)$$

and the anti-periodicity of support forces and couples

$$a_c^- = -a_c^+, \quad m_c^- = -m_c^+, \quad (37)$$

where a_c is the opposite of the resultant force on the boundary particle exerted by other particles, m_c is the opposite of the resultant couple about the center X_c on the boundary particle.

The theoretical basis and the calculation of homogenized permeability has been previously studied in Du and Ostoja-Starzewski [2006], Ostoja-Starzewski et al. [2007], Sun et al. [2011a]. Assuming that the DEM assembly is used to model the strong discontinuity which often becomes flow conduit or flow barrier, we may again use a Hill-Mandel lemma corresponding to the Darcy's law to determine the effective permeability of the assembly, i.e., [Du and Ostoja-Starzewski, 2006],

$$\langle \nabla p \cdot q \rangle = \langle \nabla p \rangle \cdot \langle q \rangle \quad (38)$$

Eq. 38 can be satisfied by the Dirichlet boundary condition in which the pore pressure difference across two opposite face is prescribed and Darcy's velocity is determined. As a result, the effective permeability tensor of a RVE can be determined via numerical flow experiment on the RVE. The fluid flux vector q within the RVE is computed when subjected to pressure gradient ∇p , and the permeability k_{RVE} is determined by Darcy's law

$$q = -\frac{1}{\mu} k_{RVE} \nabla p. \quad (39)$$

Among the solution strategies, the numerical solution of Stokes equations using finite element or Lattice-Boltzmann method yield accurate results but require large computational resources. This work resorts to the pore network flow model which simplifies the pores as nodes interconnected by edges allowing fluid flow [Bryant et al., 1993, Chareyre et al., 2012]. The first step of the homogenization procedure of permeability is the domain decomposition of the DEM sphere packing. This is achieved by well developed Delaunay triangulation and dual Voronoi graph algorithms [Edelsbrunner and Shah, 1996, Chareyre et al., 2012]. Using the particle centers as the triangulation nodes, the deformed micro-scale domain Ω_μ is decomposed into cells $\Omega_\mu = \bigcup_{i=1}^{N_t} \Omega_\mu^i$. Ω_μ^i is triangle in 2D analysis and tetrahedron in 3D analysis. The 2D concepts are adopted in the following descriptions. Each triangular cell Ω_μ^i encloses a pore space of volume V_v^i between three particles. The remaining solid space is the intersection of Ω_μ^i with the three particles and has the volume V_s^i . The dual domain decomposition of Ω_μ into Voronoi cells generates the pore network in the DEM assembly. Each node is regarded as the center of the pore space in a triangular cell Ω_μ^i and each edge serves as the flow pipe connecting two pore space centers.

The next step is to define the local conductance of each edge ("pipe") in the flow network. Consider an edge connecting the pore i and j of cells Ω_μ^i and Ω_μ^j , respectively. Suppose that the flux in this pipe is q^{ij}

when i at position \mathbf{x}^i has the pressure p^i and j at position \mathbf{x}^j has the pressure p^j , and that the length of the pipe is $L^{ij} = \|\mathbf{x}^i - \mathbf{x}^j\|_2$, the local conductance g^{ij} relates these quantities by,

$$q^{ij} \cdot (\mathbf{x}^i - \mathbf{x}^j) = q^{ij} = \frac{g^{ij}}{\mu} \frac{p^i - p^j}{L^{ij}}. \quad (40)$$

Extensive studies have been conducted on defining g^{ij} in pore network models [Piri and Blunt, 2005, Hilpert et al., 2003]. This work adopts the definition in Chareyre et al. [2012], in which a new method of determining hydraulic radius HR^{ij} of a cross-section of complex geometry is proposed. The local conductance admits the expression

$$g^{ij} = \alpha A^{ij} (HR^{ij})^2. \quad (41)$$

where α is a non-dimensional factor reflecting the pore throat shape ($\alpha = 0.5$ in this work), A^{ij} is the cross-sectional area of the throat. A domain Θ^{ij} around the throat between two pore spaces is defined based on the triangular cells and the dual Voronoi cells. Its volume is ϕ^{ij} and the total area of its boundaries is γ^{ij} . The hydraulic radius is given by

$$HR^{ij} = \frac{\phi^{ij}}{\gamma^{ij}}. \quad (42)$$

In the numerical example of this work, we adopt the micro-scale DEM assembly as the hydro-mechanical constitutive model for the sealing fault, which has an intrinsic permeability in the order of 10^{-14} m^2 . This low level of permeability is observed in porous media that often has a porosity of 0.1-0.2. However, in the DEM numerical assembly, where mono-disperse spheres are adopted to represent the particles, this porosity is not possible to achieve under confining stress in the order of Mega Pascals. The lowest porosity of the RVE is still higher than 0.3. This discrepancy comes from the poor representation of the actual shape of grains in real soils by the idealized spheres. Realistic and accurate porosity will be reached by using numerical particles with complex shapes, yet this requires more powerful DEM simulation tools and is out of the scope of this study. Hence the permeability obtained from the flow network simulation on the micro-scale DEM assembly is artificially scaled to the order of 10^{-14} m^2 . Our focus lies on the path-dependent changes in permeabilities in the normal and tangential directions of the strong discontinuity, following the displacement jumps applied to the interface.

Once the DEM assembly generates a sufficiently large database, the database can be used to train the data-driven model. Typically one would like to test a large variety of different loading paths such that different responses (torsion, shear, stretch, compression, loading & unloading) can be anticipated. The exact content of the database is often determined after a trial-and-error procedure. The size of database strongly depends on the exact configuration, type and the training process used for the neural network. Understandably, it could be counterproductive to generate a large database for a small neural network. On the other hand, it also does not make sense to have a very deep and complex neural network design while the data available for the supervised training and validation are limited. The detailed description on the design of the neural network will be presented in the next section.

One may think of the trained meso-scale data-driven model as a representation of the data or as a surrogate model. However, the advantage of the neural network as surrogate model is that one may easily adjust the neural network configuration, whereas changing from one surrogate model to another often require a substantial amount of work to identify material parameters.

After the completion of the training of the meso-scale data-driven model used to represent the strong discontinuity at the meso-scale, we then generate another set of RVE that uses the data-driven model as a replacement of the DEM model to capture the traction-separation law of the localized damage zones at the meso-scale. We then subject the meso-scale RVE to various loading paths and obtain the simulated responses from the hybrid neural network/finite element model as shown in Figure 8. If there is no comparable experimental data available, then the simulated responses of the dual-porosity material constitute the database for the data-driven model used in field-scale problems.

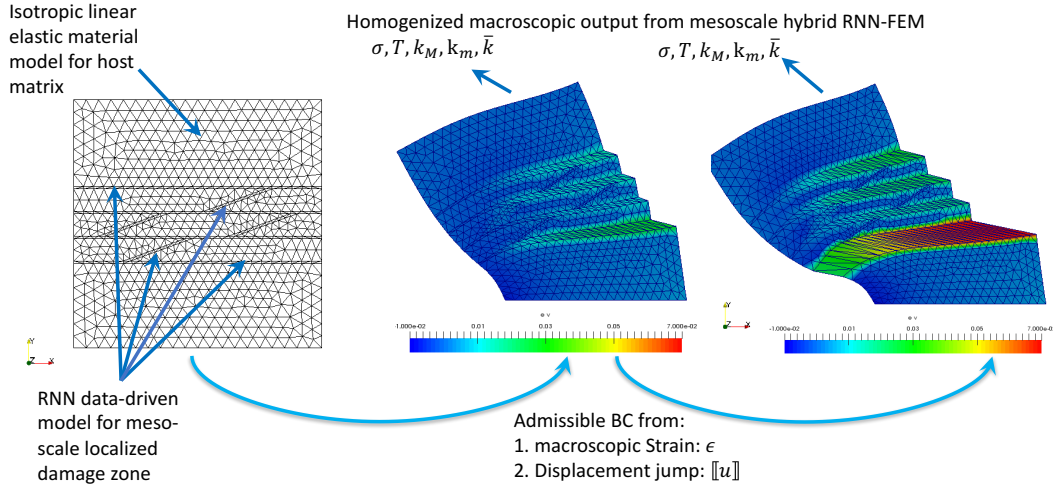


Fig. 8: The generation of database for macroscopic data-driven model using hybrid neural-network / finite element model. In the meso-scale simulations, we consider the localized damage zone as a dual-porosity material in which the data-driven model provides the traction-separation law and the flow prediction normal and orthogonal to the interface while the responses of the host matrix is captured by a simple elastic materials. The meso-scale RVEs are then subjected to various loadings, and the responses are recorded and used to train and validate the macroscopic data-driven model. In this figure, displacement field is scaled by a factor of 50.

4 Offline bridging scales via recurrent neural network

In a nutshell, machine learning refers to the ability of a computer to learn without being explicitly programmed. In the field of computational mechanics, machine learning has been widely used for finding the bases of the reduced dimensional space for reduced order modeling, and for replacing constitutive laws with trained artificial neural network. The latter tasks have found a various degree of success in previous work such as [Ghaboussi et al. \[1991\]](#) and [Lefik and Schrefler \[2003\]](#). One key aspect that is critical for the application of the solid mechanics applications is that the machine learning process must be able to generate path dependent responses such that the strain, strain rate and strain history may all affect the resultant stress responses. The ability of replicating history dependent behavior is equally important for capturing the hydraulic responses. For instance, the effective permeability of a porous rock may be influenced by the damage of the host matrix. Furthermore, water retention curve, the relation between degree of saturation and suction, is known to be dependent on the wetting and drying history of the pores. In the poromechanics literature [[Johari et al., 2006](#), [Lamorski et al., 2008](#)], the hydro-mechanical path-dependent behavior has been enforced by additional input in a feed-forward neural network or support vector machine.

In this work, we propose a number of new innovations for the neural network models built specifically for path-dependent materials with potential failures. First, we use a specific type of recurrent neural network called long-short-term-memory neural network to create the effect of internal states or memory to capture the path-dependent behavior (Section 4.2), while overcoming the gradient vanishing or exploding issues commonly encountered in training RNN with deep hidden layers (Section 4.3). Meanwhile, we also introduce a new technique in which micro-structural attributes, such as fabric tensors, higher-order deformation measures are used to improve the accuracy of the constitutive responses. To avoid overfitting and dealing with noisy data, the introduction of a dropout layer for the RNN network model is discussed (Sec-

tion 4.4). Finally, we also analyze the lack of objectivity in the previous neural network model and propose the usages of spectral form and Lie algebra to address this without extending the database (Section 4.5). Finally, the implementation of the algorithms that construct, train and validate the data-driven model will be discussed (Section 4.6). Throughout this section, we will provide a number of numerical experiments within each sub-section to demonstrate the rationale of these concepts and the execution of these ideas.

4.1 Overview of supervised machine learning with artificial neural network

For completeness, we provide a brief summary of the artificial neural network and the applications to computational mechanics problems. More comprehensive descriptions can be found in multiple sources across disciplines, ranging from weather forecasting (cf. Gardner and Dorling [1998], signal processing (cf. Cochocki and Unbehauen [1993]), renewable energy applications (cf. Kalogirou [2001]), to support decision for cancer treatment (cf. Lisboa and Taktak [2006]). Here our focus is on the application of computational mechanics problems. The classical modeling of poromechanics problems and more generally computational machine problems often involves the following steps:

1. First, one derive the field theory which is often derived from first principles (e.g. balance of linear and angular momentum, mass, energy and thermodynamic laws).
2. Then, proper constitutive laws are chosen to replicate the responses of a particular type of materials or microstructures [Zienkiewicz et al., 1999].
3. Experimental data are then used to calibrate the computational models. The identification of material parameters are sometime conducted manually, but general speaking the material parameter identification procedure can be recast as optimization problems [Liu et al., 2016].
4. Finally, the proposed model is verified against benchmarks and validated with experiment data that are **not used for calibrations** such that the forward prediction ability can be measured [Wang and Sun, 2016a,b].

In the previous work that applies neural network and supervised learning for computational mechanics problems, Steps 2 and 3 are often bypassed. Instead, data set obtained from experiments are first split into two mutually exclusively subsets, i.e. (1) the training set, which is used for machine learning and (2) the validation set, which is used for validation. Following this step, an artificial neural network is generated. The design of an artificial neural network and many of the evolution roles are inspired by the biological counterpart. Basically an artificial neural network consists of at least three types of layers, the input layer, in which the neuron carries the information of the input, the hidden layers, which are the components that carries out the neural computations and the output layer, which stores the output results, as shown in Figure 9.

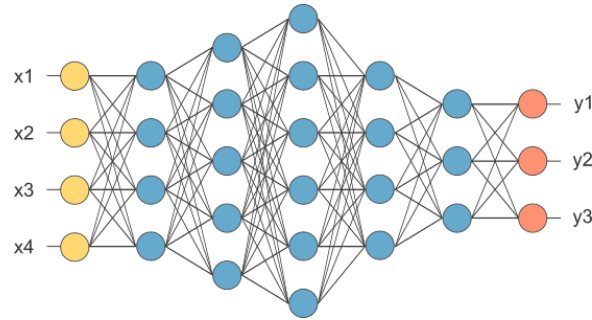


Fig. 9: An example of artificial neural network with 4 inputs, 5 hidden layers with different number of neurons and 3 outputs. Figure adopted from Open Neural Networks Library.

In the hidden layers, each artificial neuron or vertex is assigned an initial weight. A forward prediction can be made by propagating signals through the neuron network from the input layer. Depending on the

input signals coming from the connected neurons, each neuron may activate, de-activate or change. This change may in return affect the weight or signal of the connected neuron. This propagation of information can be hierarchical in which case the artificial neural network itself is a directed acyclic graph. In this case, the neural computation can be regarded as a propagation of information from the upstream to the downstream. At the very end of the neural computation, the output layer will contain the output data. In the case where supervised learning is used, the training data set will be used as the benchmark of the neural network.

The supervised learning is a procedure in which one attempts to adjust the weight of the neural network such that the discrepancy between the benchmark and prediction is minimized. It should be noticed that the discrepancy is measured by the cost (or objective) function. When the forward prediction for a given input signal is computed, the prediction stored in the output layer will be compared with the benchmark and the cost or error is updated. Following this step, the error values of each vertex in the output layer will be backward propagated such that each vertex in the hidden layer will receive an error value. This error value is then used to compute the gradient of the cost function with respect to the weight corresponding to each vertex. This gradient term is then used to update the weights such that the cost function is minimized.

For instance, if we replace the constitutive law with neural network for elastic materials, then the input will be a strain measure and the output can be stress measure. However, this input and output is not the only feasible choice. For instance, one may use both the strain and strain rate as input and output the stress for a rate-dependent material [Jung and Ghaboussi, 2006]. For material that exhibits history dependent behaviors, Lefik and Schrefler [2003] has demonstrated that an incremental form, which contains the strain components, the stress components and the incremental of strain expressed in a Cartesian coordinate system can be used as input to enforce "memory" such that the history-dependent behavior can be replicated. In this work, we will instead use a recurrent neural network to capture this memory effect exhibited in path-dependent porous media.

4.2 Deep learning with recurrent neural network

The recurrent neural network (RNN) is an umbrella term for artificial neural networks with connection topology possesses cycles [Lukoševičius and Jaeger, 2009]. In other words, the recurrent neural network considers data existing as sequences and the output of a layer in the previous step is added back as additional input and fed back into the same layer to produce the output (hence the name recurrent neural network).

The existence of the cycles leads to a profound difference, as it resembles how a history-dependent process evolve in time with cause-and-effect relationship (e.g. translating paragraphs of content between different languages, analysis of video surveillance). While classical artificial neural network can be regarded as a mathematical function, the recurrent neural is a dynamical system model of the biological neural networks that *possess memory* and is able to process arbitrary sequences of input and generate arbitrary sequences of output. These important features have made RNN versatile among many applications that require learning from temporal data such as speech recognition, machine translation, quick-type for smartphones and driver-less car technology. In structural engineering, RNN has been used to perform model-free structural analysis in which the structural behavior is predicted without any physical model [Graf et al., 2010]. Similar approach has been applied in geotechnical engineering in which an RNN is used to replicate stress-strain relation of sand subjected to monotonic triaxial compression loading [Zhu et al., 1998].

In all the application mentioned above, the RNN machine learning procedure is often used to produce data-driven model that completely replace the constitutive models based on human interpreted knowledge. In this work, we take a different approach in which the machine learning is not used to generate model-free prediction but to be used for generating links for bridging simulations across scales in an off-line fashion. Furthermore, we also retain the usage of the mechanics principles whenever possible in the multiphysical model conceptualized as a directed graph. Ultimately, the resultant model represents a hybridization of human- and machine-interpreted knowledge that can be used to generate predictions and as forecast engine. In the following subsections, we will describe the specific techniques we used to build

the recurrent neural network and how it can be trained using a combination of data generated from experiments and micro-mechanical simulations.

4.3 Overcoming gradient vanishing or exploding issues with long short-term memory architecture

Despite the exceptional prediction power of RNN as forecast engine, RNN is known to be relatively vulnerable to the vanishing and exploding gradient problems. While a vanishing gradient may lead to the change of weight of the nodes in the recurrent neural network that has no significant changes on the error measured by the objective function or cost function. By contrary, an exploding gradient may lead to the error very sensitive to any small change of weights in the nodes. Both issues can be resolved for mechanics data-driven model produced by very small architecture with limited number of hidden layers [Lefik and Schrefler \[2003\]](#), [Jung and Ghaboussi \[2006\]](#). However, since (1) the usage of RNN leads to a larger number of nodes in the input layers, (2) the multiscale coupling scheme uses additional microstructural attributes as input, and more importantly (3) we intend to study the usage of deep learning which requires a sizable number of hidden layers, the vanishing or exploding gradients could be a significant issue.

Here we take advantages of a technique commonly used in computational linguistics called Long Short Term Memory (LSTM). First introduced by [Hochreiter and Schmidhuber \[1997\]](#), the LSTM's major departure is the use of memory blocks, instead of the classical artificial neurons as nodes for RNN. Within a memory block, a new entity called "gate" is introduced to control the flow of information and the state of the block, as shown in Figure 10.

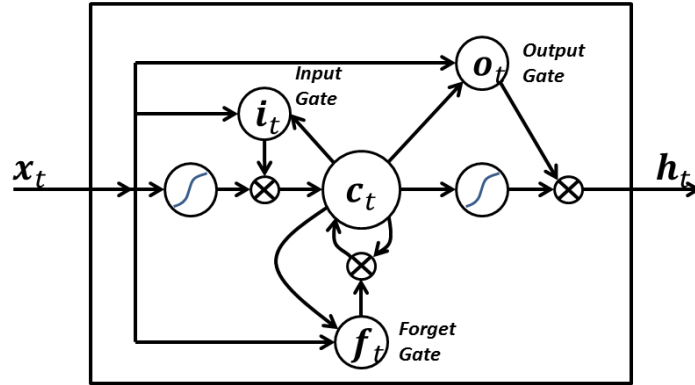


Fig. 10: A Long-Short-Term-Memory neuron with input, output, and forget gate to process sequence with memory effect.

A LSTM neuron possesses a state of the memory cells at time t C_t . Define x_t as the value of the input sequence at time t , and h_t as the value of the output sequence at time t . The signal through the forget gate is given by

$$f_t = \sigma(W_f \cdot x_t + U_f \cdot h_{t-1} + b_f), \quad (43)$$

where σ is the sigmoid function $\sigma(x) = \frac{1}{1+\exp(-x)}$, W_f and U_f are weight matrices, b_f is bias vector for the forget gate.

The new information to be stored in the cell state is given by the signal i_t through the input gate

$$i_t = \sigma(W_i \cdot x_t + U_i \cdot h_{t-1} + b_i), \quad (44)$$

where W_i and U_i are weight matrices, b_i is bias vector for the input gate.

The new candidate value cell state is given by a tanh layer

$$\tilde{C}_t = \tanh(W_C \cdot x_t + U_C \cdot h_{t-1} + b_C), \quad (45)$$

where \tanh is the hyperbolic tangent function $\tanh(x) = \frac{\exp(x) - \exp(-x)}{\exp(x) + \exp(-x)}$, W_C and U_C are weight matrices, b_C is bias vector.

The old cell state C_{t-1} is updated by the above forget and input information, i.e.,

$$C_t = f_t C_{t-1} + i_t \tilde{C}_t \quad (46)$$

Finally, for the output signal

$$h_t = o_t \tanh(C_t) \quad (47)$$

where o_t is the signal through the output gate

$$o_t = \sigma(W_o \cdot x_t + U_o \cdot h_{t-1} + b_o), \quad (48)$$

where W_o and U_o are weight matrices, b_o is bias vector for the output gate.

To showcase the advantage of using LSTM with micro-structure information as a part of input features, we examine the forward prediction capability of (1) the classical neural network "ANN" (2) LSTM neural network which is specially designed for memorizing sequences "LSTM" and (3) LSTM neural network that also memorizes micro-structure attributes "LSTM Microstructure Data". Numerical simple shear tests with loading-unloading under different confining pressure ($\sigma = 50$ MPa, 60 MPa, 70 MPa, 90 MPa and 100 MPa) are conducted on a discrete element assembly and serve as the training data to the three neural network models (Fig 11(a)). The "ANN" model design is similar to Ghaboussi et al. [1991], where the inputs are the confining pressure σ , the shear strains γ_{n-1} , γ_n , γ_{n+1} and the shear stresses τ_{n-2} , τ_{n-1} , τ_n . The subscripts indicate the time steps t_{n-2} , t_{n-1} , t_n , t_{n+1} . The output is the shear stress τ_{n+1} at the next time step t_{n+1} .

Different from the classical ANN model, LSTM neural network accepts sequences of history values of the physical parameters as inputs. Thus the input features now consist of the confining pressure σ , the sequence of history values of shear strains $[\gamma_{n-1}, \gamma_n, \gamma_{n+1}]$, and the sequence of history values of shear stresses $[\tau_{n-2}, \tau_{n-1}, \tau_n]$. The output features is again the shear stress τ_{n+1} . In addition to the strain and stress history, one key innovation we attempted in this paper is to incorporate the evolution of microstructural attributes as additional input for the neural network. The micro-structure data adopted in this example are the porosity ϕ of the DEM assembly, the coordination number CN and the fabric tensor A_F . $A_F = \frac{1}{N_c} \sum_c \mathbf{n}_c \otimes \mathbf{n}_c$, where \mathbf{n}_c is the normal of contact c . Here our goal is to check whether the incorporation of any of these additional data as input in the RNN network improves the prediction quality. If the answer is positive, it is likely that a human-derived phenomenological model could benefit from the inclusion of these physical quantities.

In a supervised machine learning setting, the LSTM neural network will be adjusted based on the portion of the dataset used for calibration. After the training or back propagation completes, the relationship between these averaged micro-scale attributes and the predicted stress state can be determined. For phenomenological models where history-dependent behavior is encoded in the evolution of internal variables, the influence of micro-structural attributes are often implicitly incorporated (except a few exception such as [Dafalias and Manzari, 2004, Fu and Dafalias, 2011]), this could be a difficult task.

All three neural network models have two hidden layers of 80 nodes, and dropout layers of rate 0.2 are placed after each LSTM layer. The function of the dropout layer will be discussed in details in Section 4.4. The sigmoid activation function is chosen for the output layer. We also set the same mean squared error as the training goal of all three models such that the errors are supposed to reduce to the same level (around $1e-5$).

Fig 11(b) compares their forward predictions of the loading-unloading behavior under confining pressure of $\sigma = 80$ MPa (not included in the training set). It is shown that the LSTM model performs better than the ANN model, in regard to the peak stress, softening and unloading-reloading cycles. Also, with micro-scale information, LSTM can yield closer prediction to the test data than LSTM that only process macroscopic strain and stress data.

Furthermore, determining how much and what types of micro-structure data to be incorporated into machine learning model is a challenging task. As an example, we compare three designs of LSTM network (1) LSTM with ϕ (2) LSTM with ϕ , CN and (3) LSTM with ϕ , CN , A_F (Fig 11(e)). For (1) and (2), the additional information on average number of interactions per particle does not improve the generalization capability of the LSTM model. The model with the fabric tensor, which describes the matrix of the porous

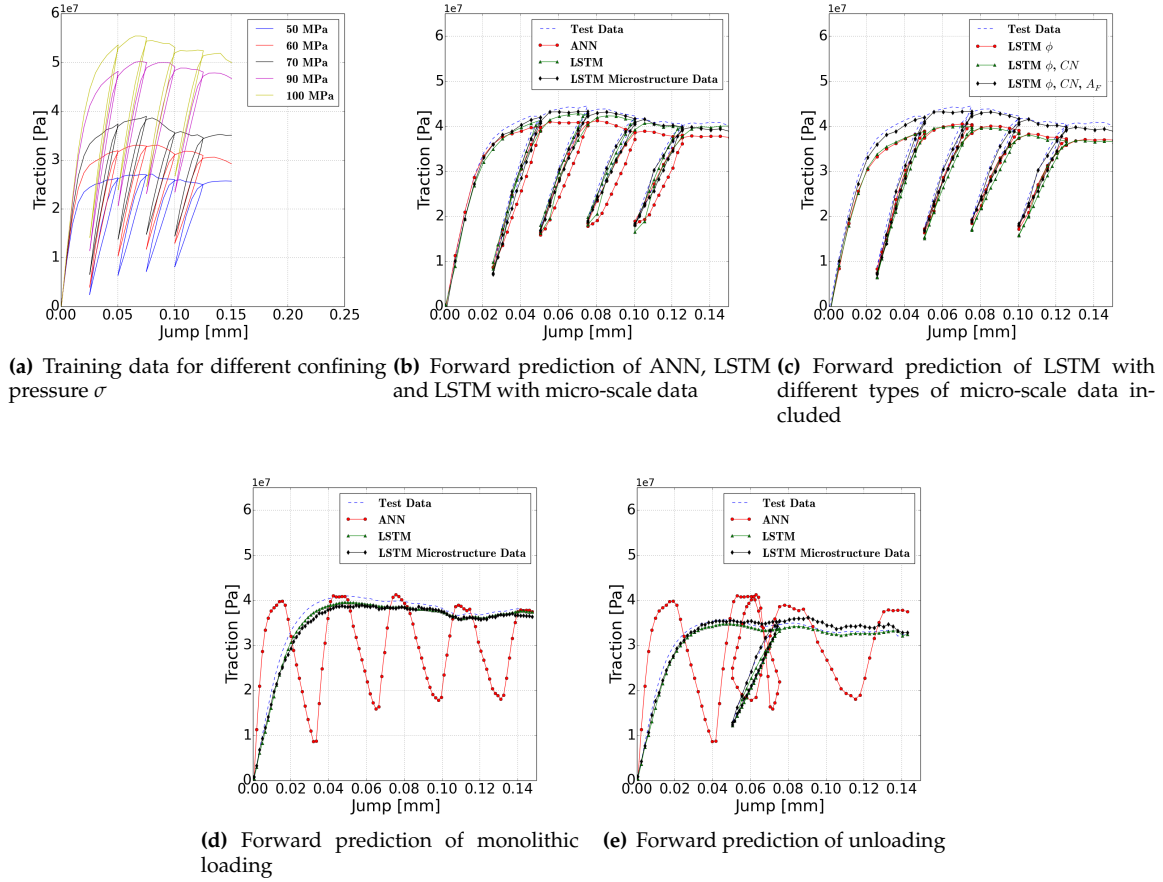


Fig. 11: Comparison of forward prediction capacity between different configurations of neural network models.

media, gives significantly more accurate results. This is due to the deformation mode of the DEM sample. The micro-structure are heavily distorted during the shearing. This is reflected in the change of principal values and rotation of principal directions of the fabric tensor. ϕ and CN , however, are hard to represent this induced anisotropy in the micro-structure because they are scalars. Thus in this example, the evolution history of the fabric tensor is crucial to the forward prediction capacity of the LSTM network.

Lastly, we study the performance of different LSTM architectures to determine the neural network parameters adopted in this work. Table 1 lists 5 neural network configurations that differ in the number of hidden layers, the number of neurons per hidden layer and the activation function for the output layer. The training data and testing data are the previous dataset from the numerical simple shear tests including the micro-scale attributes: porosity, coordination number and fabric tensor. The training data consists of 500 samples and the testing data contains 100 samples. The training phase consists of 5000 epochs and the batch size is 100. The loss function is the standard mean squared error. The value of loss on both training and testing data are recorded during the training epochs and are presented in Fig. 12. The performance curves show that all architectures can drive the training error down to the 10^{-5} level and the testing error down to the 10^{-4} level. Configurations 4 and 5 perform better in the training data and Configuration 5 is more accurate in predictions. The discrepancy is not significant. Thus for the neural network architecture used in this work, we choose the second configuration in Table 1, which gives good training and prediction

ANN Configuration	Number of hidden layers	Number of LSTM neurons per hidden layer	Activation function for output layer
1	2	50	Sigmoid
2	2	80	Sigmoid
3	2	100	Sigmoid
4	3	80	Sigmoid
5	2	80	ReLU

Table 1: Different ANN architectures for evaluation of training performance.

performances. The number of LSTM neurons in the network is small so that it will speed up the training and the calculation in the triple-scale online simulations.

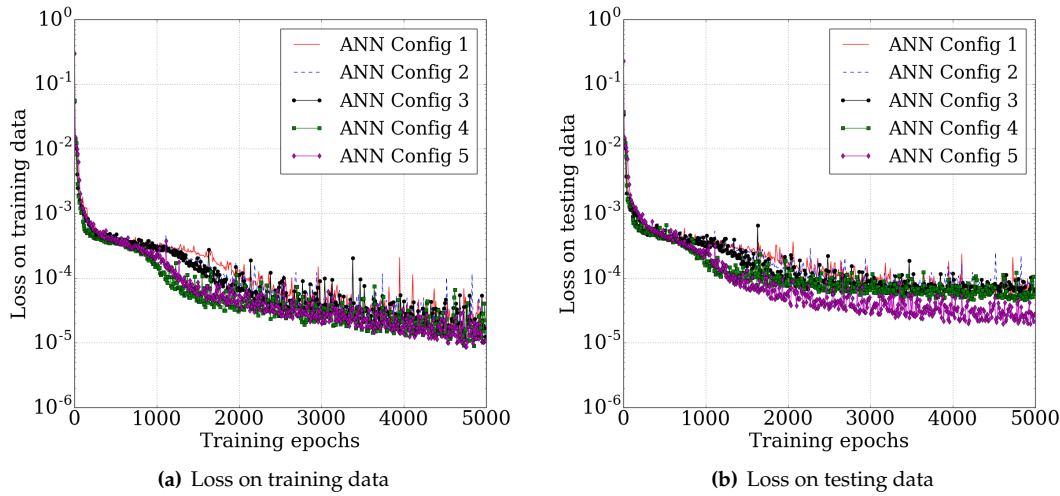


Fig. 12: Comparison of training performance of different ANN architectures in Table 1.

4.4 Performance of LSTM neural network on noisy data

In recent work such as [Kirchdoerfer and Ortiz \[2016\]](#), [Versino et al. \[2017\]](#), the authors have presented ways to either use variational principles or symbolic regression to build data-driven model for elastic and plastic material models with various amount of human knowledge. In particular, [Kirchdoerfer and Ortiz \[2016\]](#) has provided a systematic study on how reduction on data points in stress-strain curve affects the robustness and accuracy of data-driven model for hyperelastic problems and assess the effect of under-fitting. However, since most of the data-driven models are generated from optimization procedure, the data-driven models may also generate over-fitting predictions.

A neural network with LSTM layers and dropout layers is trained with the load-deformation curves obtained from experiments performed on Arkose sandstone at various deformation rates (Fig. 13(a) [Peng, 1973]). This experimental data set contains four displacement rates (0.08 mm/s, 0.16 mm/s, 0.32 mm/s, 0.64 mm/s). 0.08 mm/s, 0.32 mm/s, 0.64 mm/s as the training set and 0.16 mm/s as the test set. Random noise in force is added in a point-wise manner to the original data. The random noise follows a normal

distribution with zero mean and standard deviation in inverse proportion to the square root of the data size, following [Kirchdoerfer and Ortiz \[2017\]](#). It should be noted that since LSTM networks are designed for memorizing sequences, learning the rate-dependent behavior does not require any specific change and therefore more convenient than the feed-forward neural network counterpart. The input is composed of the next displacement $d_{t+\Delta t}$, the history of displacement $d_t, d_{t-\Delta t}, d_{t-2\Delta t}$ and the history of force $f_t, f_{t-\Delta t}, f_{t-2\Delta t}, f_{t-3\Delta t}$. The output is the force $f_{t+\Delta t}$ at the displacement $d_{t+\Delta t}$ based on these input sequences. Note that the components in the sequences are separated by the real time increment Δt , then the information of deformation rate is directly presented to the LSTM neural network by the input sequences. Thus using the deformation rate as an additional input to the neural networks is not necessary. Fig. 13(b) shows that the LSTM network has successfully acquired the rate-dependence behavior from the training data and returns accurate forward prediction on the testing data. Fig. 13(c) shows the effect of dropout layers on the forward predictions of the LSTM model, where different dropout rates of 0.2, 0.4 and 0.6 are compared. In this example, the models are capable of suppressing the over-fitting on noisy training data and provide accurate predictions. The effect of the dropout layer is minor. Interestingly, the recurrent neural network is able to deliver a very accurate forward prediction in a cross-validation study, even though the data available are quite limited, as shown in Figure 13. The in-depth study of dropout layers effect and the techniques to deal with noise for data-driven constitutive models are important, but is out of the scope of this paper. These studies will be conducted in a future dedicated work.

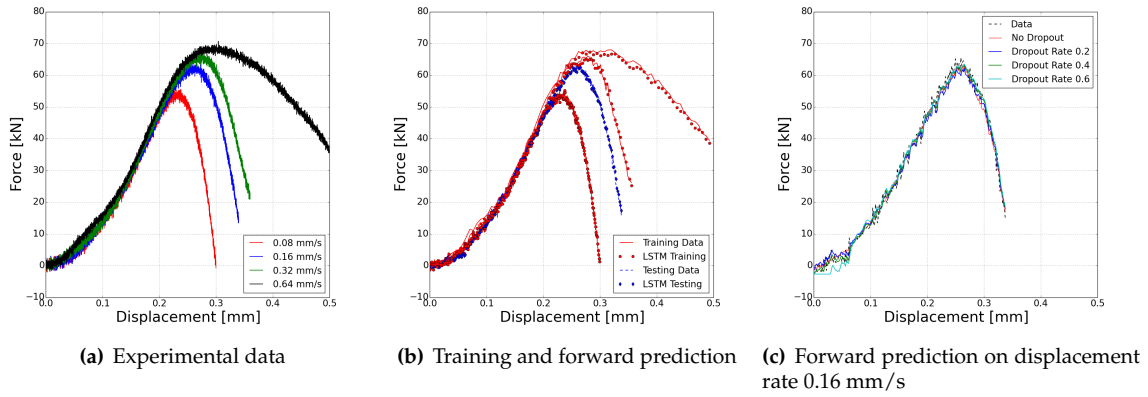


Fig. 13: Training of LSTM model with rate-dependent experimental data of Arkose sandstone with artificial noise (after [\[Peng, 1973\]](#)). The models with dropout layers of different dropout rates deliver similar forward predictions on the testing data in this example.

4.5 Enforcing objectivity and spectral decomposition

The data-driven constitutive law must preserve the principle of objectivity, i.e., the rotation and translation of the observer frame should not affect the material responses. This requirement is not automatically satisfied if we only train a neural network with the components of strain tensors expressed in terms of a particular coordinate system as input and components of stress tensors in the same coordinate system as output. In the 1990s, early works attempting to train artificial neural network to generate data-driven constitutive laws did not address this issue, as the goal of this material model is often used only to test whether the constitutive laws can generate stress-strain curve for selected experimental settings (e.g. uniaxial, biaxial and triaxial loadings) [\[Ghaboussi et al., 1991\]](#). This is reasonable as the frame indifference problem is not apparent (although it can be easily checked) in infinitesimal material point simulation.

This situation changes when [Lefik and Schrefler \[2003\]](#) attempts to put data-driven model into finite element code and discovers that superposing rigid body rotation with the deformation may lead to changes of the stress prediction. They then propose a feasible solution where one simply generates extra stress-strain curves from the same experimental data by rotating the material coordinate system multiple times such that the neural network can learn the objectivity requirement from the data. This remedy understandably enlarges the size of the training data set and potentially requires more neurons due to the additional rules needed to be learned from data. Furthermore, rather than enforcing the objectivity in a strict sense, the artificial network may only recognize the rules approximately, and hence objectivity can only be satisfied within a reasonable but not negligible tolerance.

In this paper, we propose an alternative new approach to enforce the objectivity in the infinitesimal regime, which could be advantageous over the previous attempts in the sense that (1) it does not require extension of the training sets through imposing rotation on the coordinate system and (2) the resultant data-driven model will enforce objectivity in a strict sense.

The key is to modify the way data are represented in the input and output layers such that the desired properties of a physical quantity (e.g. deformation gradient, rotation) in a specific Lie group (e.g. special orthogonal group, general linear group, symplectic group) can be retained while the incremental update is sought. In the proposed framework, we will first express all the tensor qualities in spectral form. For instance, the effective stress, the infinitesimal strain, and permeability tensors can be expressed in terms of their corresponding principal values and directions, i.e.,

$$\boldsymbol{\sigma}' = \sum_{A=1}^3 \sigma'_A \mathbf{n}_\sigma^{(A)} \otimes \mathbf{n}_\sigma^{(A)}, \quad \boldsymbol{\epsilon} = \sum_{A=1}^3 \epsilon_A \mathbf{n}_\epsilon^{(A)} \otimes \mathbf{n}_\epsilon^{(A)}, \quad \mathbf{k} = \sum_{A=1}^3 k_A \mathbf{n}_k^{(A)} \otimes \mathbf{n}_k^{(A)} \quad (49)$$

where σ'_A and $\mathbf{n}_\sigma^{(A)}$, ϵ_A and $\mathbf{n}_\epsilon^{(A)}$, and k_A and $\mathbf{n}_k^{(A)}$ are the eigenvalues and corresponding normalized eigenvectors of effective stress, infinitesimal strain, and permeability tensors respectively. Similarly, the forth-order tangential operator can be written in spectral form, i.e.,

$$\begin{aligned} \mathbf{c} = & \sum_{A=1}^3 \sum_{B=1}^3 \frac{\partial \sigma'_A}{\partial \epsilon_B} \mathbf{m}^{(A)} \otimes \mathbf{m}^{(B)} \\ & + \frac{1}{2} \sum_{A=1}^3 \sum_{B \neq A}^3 \frac{\sigma'_B - \sigma'_A}{\epsilon_B - \epsilon_A} (\mathbf{m}^{(AB)} \otimes \mathbf{m}^{(AB)} + \mathbf{m}^{(AB)} \otimes \mathbf{m}^{(BA)}) \end{aligned} \quad (50)$$

where $\mathbf{m}^{(A)} = \mathbf{n}^{(A)} \otimes \mathbf{n}^{(A)}$, $\mathbf{m}^{(AB)} = \mathbf{n}^{(A)} \otimes \mathbf{n}^{(B)}$. Note that the tangent is composed of two terms, one corresponding to the derivative of the principal stress with respect to principal strain at a fixed principal direction; another corresponding to the spin of the principal axes [\[Borja and Tamagnini, 1998\]](#). In the special case where strain and stress are coaxial and the permeabilities are isotropic, one may simply design a recurrent neural network with the principal strain time history as input and output the principal stresses. However, the co-axial assumption may be not appropriate for more complex situations. Here we consider the general case in which the strain, stress and permeability tensors are not necessarily coaxial. In this case, we must consider how to represent the rotation of the principal directions at each incremental time step.

Recall that a rigid body rotation can be viewed as a mapping where $\mathbf{x} = \varphi(\mathbf{X}, t) = \mathbf{R}(\mathbf{X}, t) \cdot \mathbf{X}$ where the rotation tensor \mathbf{R} belongs to the special orthogonal group $SO(3)$, i.e.,

$$SO(3) = \{\mathbf{A} \in GL(3) | \det \mathbf{A} = 1, \mathbf{A}\mathbf{A}^T = \mathbf{I}\} \quad (51)$$

where $GL(3)$ is the general linear group of 3 dimensions. As explained in previous work such as [Lefik and Schrefler \[2003\]](#), [Lefik et al. \[2009\]](#), directly using the components of the stress and strain with respect to a particular Cartesian coordinate system will make the predicted responses of the data-driven model lacking objectivity. The key underlying reason is that the neural network computation contains addition (and minus) operations, but the special orthogonal group $SO(3)$ does not belong to the additive abelian groups. In other words, adding the components of two rotation tensors together may lead to a tensor that does not belong to the $SO(3)$ group (e.g. the tensor resulted from adding two identity tensors together where $\mathbf{I} \in SO(3)$). As explained in previous work such as [Mota et al. \[2013\]](#), one way to resolve this issue

is to perform the machine learning using the infinitesimal version of the corresponding Lie group, i.e., the Lie algebra, i.e.,

$$so(3) = \{B \in M(3) | B = -B^T\} \quad (52)$$

where $M(3)$ denotes the space of the 3-by-3 tensors. Note that there are more than one way to parametrize the rotation tensor and hence relates a mapping of Lie group with the Lie Algebra counterpart. For instance, the finite rotation tensors and the infinitesimal spin tensor can be related via exponential and logarithm mapping, i.e., [Krysl and Endres, 2005, Tuzel et al., 2008, Mota et al., 2013]:

$$R = \exp[\tilde{\Psi}] = \sum_{k=0}^{\infty} \frac{\tilde{\Psi}^k}{k!} ; \Psi = \log R = \sum_{k=0}^{\infty} \frac{(-1)^{k-1}}{k} (R - I)^k \quad (53)$$

where $\tilde{\Psi}$ is a skew-symmetric infinitesimal spin tensor that can be defined by three off-diagonal components $\tilde{\Psi}_{23}, \tilde{\Psi}_{13}, \tilde{\Psi}_{12}$ whereas the diagonal components are zero. In the incremental form, the rotation matrix at time t_n is updated by

$$R_n = R_{n-1} \exp[\Delta \tilde{\Psi}_n] \quad (54)$$

Therefore, the input for a data-driven constitutive model trained by the recurrent neural network can be the history of the principal strains $\epsilon_1, \epsilon_2, \epsilon_3$ and the history of the incremental rotation parameters of strain directions $\Delta \tilde{\Psi}_{23}^\epsilon, \Delta \tilde{\Psi}_{13}^\epsilon, \Delta \tilde{\Psi}_{12}^\epsilon$. The outputs are the principal stresses $\sigma_1, \sigma_2, \sigma_3$ and the incremental rotation parameters of stress directions $\Delta \tilde{\Psi}_{23}^\sigma, \Delta \tilde{\Psi}_{13}^\sigma, \Delta \tilde{\Psi}_{12}^\sigma$. It should be noted that for anisotropic materials such as single crystals, the supervised machine learning still requires data generated from multiple stress path in different orientations, regardless of whether the spectral form is used or not.

Here, we present a simple numerical test to check whether the data-driven model provides objective predictions. The idea is to simply run the same tests multiple time, each with a different reference frame. If all these numerical simulations provide the same tensorial output (e.g. stress, permeability) and yield the same amount of strain energy. The test to check the objectivity consists of testing whether the strain energy density is invariant upon rotation of observer frame and if the stress measured in rotated frame follows the tensor rotation transformation rule. The procedure is presented in Algorithm 1.

Algorithm 1 Objectivity test

- 1: Generate random Strain ϵ .
 - 2: Compute Strain Energy $W(\epsilon)$, Cauchy stress $\sigma(\epsilon)$.
 - 3: Generate random Rotation $R \in SO(3)$.
 - 4: Compute $\epsilon^* = R \cdot \epsilon \cdot R^T$, Strain Energy $W(\epsilon^*)$, Cauchy stress $\sigma^*(\epsilon^*)$.
 - 5: **if** $W(\epsilon^*) = W(\epsilon)$ and $\sigma^* = R \cdot \sigma \cdot R^T$ **then**
 - 6: The material model is considered objective.
 - 7: **Exit.**
-

We test the objectivity of (1) recurrent neural network only fed with stress and strain components ("Direct RNN model"), (2) recurrent neural network with extra data of material responses observed in rotated coordinate system ("Objectivity enhanced RNN model") and (3) the proposed model using spectral decomposition ("Spectral RNN model"). As described in the procedure of the test, a random strain ϵ is imposed on the three RNN material models. 500 groups of Euler angles $\{\phi, \theta, \psi\}$ ("x-convention", $\phi \in [0, 2\pi], \theta \in [0, \pi], \psi \in [0, 2\pi]$) are randomly generated for the rotation of observer frame via corresponding rotation matrices R . 20 of them are randomly picked and presented to the "Objectivity enhanced RNN model" to provide the knowledge of frame indifference, i.e., $\epsilon^* = R \cdot \epsilon \cdot R^T$ serves as additional input samples and $R \cdot \sigma \cdot R^T$ as outputs. The relative error on the stress and strain energy read,

$$e_\sigma = \frac{\|\sigma^*(\epsilon^*) - R \cdot \sigma \cdot R^T\|}{\|R \cdot \sigma \cdot R^T\|} ; e_W = \frac{|W(\epsilon^*) - W(\epsilon)|}{|W(\epsilon)|} \quad (55)$$

for all 500 rotation cases in Fig. 14.

809 The "Direct ANN model" is shown not to be objective. The maximum errors in stress tensor and
 810 strain energy both exceed 100%, which are unacceptable for material constitutive models. With material
 811 responses in several rotated observer frames included in training data, the maximum errors are largely
 812 reduced to about 10%. However, the training data are enlarged twenty times in this numerical test and the
 813 rigorous objectivity is yet not achieved. The proposed spectral RNN model is demonstrated to be objective,
 814 since the inputs and outputs are independent on the choice of coordinate systems.

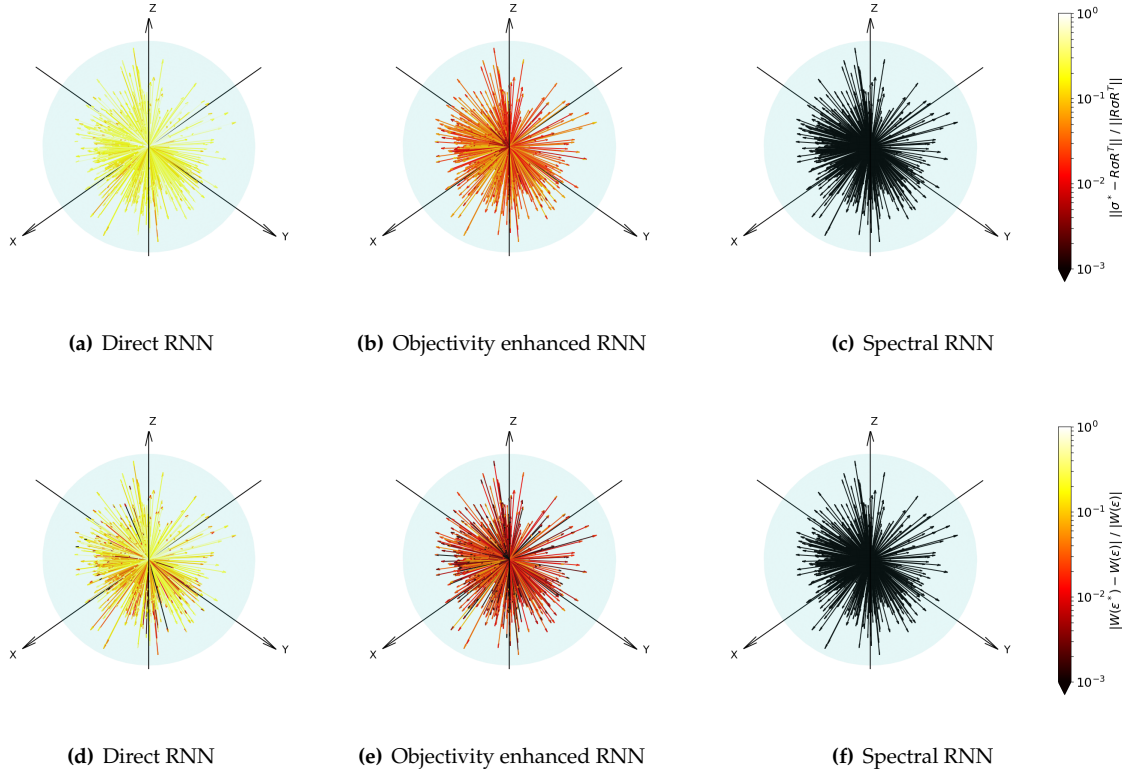


Fig. 14: Material frame indifference test for "Direct RNN model", "Objectivity enhanced RNN model" and "Spectral RNN model". A random strain ϵ is imposed on the three RNN material models. 500 randomly generated cases of rotation of observer frame are illustrated by the rotation vectors of the Euler's rotation theorem. The Euler angles $\{\phi, \theta, \psi\}$ are converted to Euler parameters $\{e_0, [e_1, e_2, e_3]\}$ and then to the rotation vector. In each case, the vector direction \hat{n} represents the axis of rotation, while the vector length l represents the angle of rotation $l \in [0, 2\pi]$ (counterclockwise). The sphere of radius 2π is drawn for guidance. The color of the vectors illustrate the relative errors for the norm of stress tensor (a-c) and the strain energy (d-f).

815 4.6 Highlights of Implementation

816 As for the implementation, we have leveraged Keras (cf. [Chollet et al. \[2015\]](#)), a high-level Python deep
 817 learning library, to build the LSTM neural networks and complete the training procedure. This model-
 818 level library allows for easy and fast prototyping of machine learning models. The low-level operations
 819 (such as tensor calculus) for machine learning are handled by Tensorflow, an open-source symbolic tensor
 820 manipulation library developed by Google, Inc [\[Abadi et al., 2016\]](#), serving as the "backend engine" of

Keras. One upshot of Tensorflow is that it has a flexible architecture based on data flow graphs that enable easy GPU accelerated training of various types of neural networks.

The building and training of the LSTM data-driven model in this paper contains four steps. Firstly, the data acquired from lower-scale numerical simulations are preprocessed and converted to specific data structure compatible with the LSTM training and validation algorithms. The data of numerical simulations are stored in comma-separated values (CSV) file and are imported by an open-source Python data analysis library Pandas [McKinney et al.]. The data are split to input features and outputs. These data are of different scales: 10^6 for traction, 10^{-2} for jump, 10^{-1} for porosity and 10^{-14} for permeabilities. Thus, each sequence of input and output is re-scaled to be within $[0, 1]$ using the MinMaxScaler class in sklearn.preprocessing toolkit [Pedregosa et al., 2011]. The input data structure that can be processed by the LSTM model must be an array of dimension 3, where the entries for the first dimension are the samples, the second dimension is the time history steps and the last dimension are the input features.

Secondly the multi-layered neural network is constructed with a few and easy-to-modify lines of codes. Keras offers a simple way to establish neural networks that incorporate input, LSTM, dropout, output layers. Adding or deleting a layer, modifying the number of nodes, changing the activation functions are very convenient thanks to the high-level library of Keras.

Then we are ready to launch the training epochs. We feed the LSTM model with the preprocessed input and output data. The back propagation algorithm will modify the weights of the neural network iteratively and the loss will be reduced to a small number (about $2e-6$ in this work). The learning rate can be reduced when the convergence becomes slow. Finally, the performance of the fully-trained LSTM material model is assessed on a set of test data that has not been provided to the model in the training phase.

5 Numerical Experiments

In this section, we present a triple-scale simulation which links the grain-scale simulations, the meso-scale assumed strain simulations and the macroscopic fault simulator together, as shown in Figure 15.

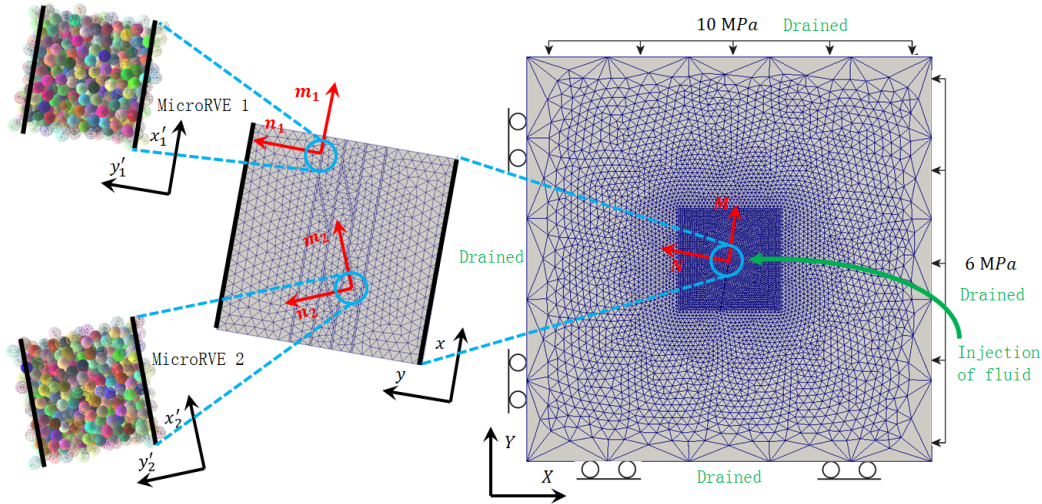


Fig. 15: Triple-scale data-driven fault reactivation simulations. $\{X, Y\}$ constitutes the coordinate system of the macro-scale problem. $\{M, N\}$ constitutes the local coordinate system of the strong discontinuity in macro-scale. For meso-scale problem, the coordinate system is $\{x, y\}$ (co-axial to $\{M, N\}$), and the internal structure distinguishes two coordinate systems of interface: $\{m_1, n_1\}$ and $\{m_2, n_2\}$. The corresponding coordinate systems for micro-scale RVEs are $\{x'_1, y'_1\}$ and $\{x'_2, y'_2\}$.

Scale & Model	Location	Parameter	Value
Grain-scale DEM	micro-discontinuities	Particle Young's modulus E	0.5 GPa
Grain-scale DEM	micro-discontinuities	Particle Poisson's ratio ν	0.3
Grain-scale DEM	micro-discontinuities	Particle Friction Angle	$\frac{\pi}{6}$
Grain-scale DEM	micro-discontinuities	Particle density	2600 kg/m ³
Grain-scale DEM	micro-discontinuities	Particle mean diameter	5 mm
Meso-scale FEM	meso-scale host matrix	Young's modulus	0.2 GPa
Meso-scale FEM	meso-scale host matrix	Poisson's ratio ν	0.2
Meso-scale FEM	meso-scale host matrix	Intrinsic permeability κ	$2e^{-14}$ m ²
Meso-scale FEM	meso-scale host matrix	Dynamic viscosity μ	$1e^{-3}$ Pa · s
Marco-scale FEM	marco-scale host matrix	Young's modulus	0.2 GPa
Marco-scale FEM	marco-scale host matrix	Poisson's ratio ν	0.2
Marco-scale FEM	marco-scale host matrix	Porosity of macropore ϕ_M	0.1
Marco-scale FEM	marco-scale host matrix	Porosity of micropore ϕ_m	0.2
Marco-scale FEM	marco-scale host matrix	Intrinsic permeability of macropore k_M	$1e^{-12}$ m ²
Marco-scale FEM	marco-scale host matrix	Intrinsic permeability of micropore k_m	$5e^{-17}$ m ²
Marco-scale FEM	marco-scale host matrix	Parameter of mass transfer $\bar{\alpha}$	$\rho_f * k_m$
Marco-scale FEM	marco-scale host matrix	Dynamic viscosity μ	$1e^{-3}$ Pa · s

Table 2: Material parameters for the grain-, meso- and macro-scale poromechanics problem with embedded strong discontinuities across three length scales.

Instead of directly replacing phenomenological laws with sub-scale simulations to generate constitutive responses as done in Wang and Sun [2016b], we introduce a data-driven offline coupling method in which numerical results from sub-scale simulations first constitute material databases. These databases are then used to train the recurrent neural network models. Once the training and validation is completed, the neural network is then used to replace the phenomenological traction-separation law and the interface conductivity models of the dual-porosity systems. This process is applied recursively across length scales twice such that the responses of the fault is predicted by a data-driven model trained and validated by data set generated with another set of data-driven models at smaller scale.

As emphasized in Belytschko et al. [2013], macroscopic responses of a material system are often dominated by the evolution of microstructural attributes, especially after the material bifurcation occurs. Yet, the traction-separation law and the conductivity law are often highly simplistic due to the difficulty to propose a proper model that captures the phenomenology. By incorporating the micro-structural information via deep learning, more realistic and complex constitutive laws can be generated automatically such that more accurate simulations of the localized responses can lead to more reliable macroscopic predictions.

In the following simulations, we assume that embedded strong discontinuities are pre-existed and do not propagate. In total, we construct two material databases. One contains the material responses of the DEM-network simulations that replicate the grain-scale interface between two bulk materials. In the second material database, the data are obtained via running RNN-FEM simulations where the neural network trained by the DEM-network database are re-used to model strong discontinuities, while the bulk material is idealized as isotropic elastic material. This recursive training strategy allows one to use machine learning as a mean to incorporate sub-scales information in an offline material. As a result, the triple-scale simulation only requires grain-scale material parameters for the DEM and flow network simulations as well as the material parameters used to model the bulk responses. All the path-dependent behaviors are therefore originated from the meso-scale interfaces and the macroscopic fault. The material parameters used in the numerical example are summarized in Table 2.

The initial and boundary conditions of the macroscopic 2D fault reactivation problem is shown in Figure 15. Note that this initial boundary value problem is a highly simplified model used for demonstration and testing purpose. A more dedicated case study intended to capture the actual complex operations of fluid injection in an actual field will be conducted in the future but is out of the scope of this study. The size

of the macroscopic domain is $1\text{km} \times 1\text{km}$ and it is assumed that the field is under plane strain condition. To simulate an anisotropic stress condition, the traction applied on the two opposite faces of the square domain is 10 MPa and 6 MPa accordingly. These values are held constant during the simulations. Meanwhile, we prescribe the Darcy's velocity of the macropores at the injection well about 250 meters from the fault line. To test the capacity of the data-driven model and to generate path-dependent responses at the field-scale level, this Darcy's velocity is not held constant but allowed to change over time, with the initial Darcy's velocity equals to 50m/s.

5.1 Training and validation of material laws for meso-scale interface

To train the recurrent neural network such that it can replicate the meso-scale hydro-mechanical responses of the embedded strong discontinuities, we first conduct 21 grain-scale simulations. The time history of the traction, displacement jump, and permeability in the normal and tangential direction, as well as the major and minor principal values of the fabric tensor are recorded. 16 of the simulation results are used as the training data set and rest 5 of them are used as the validation data set.

In each RVE simulation, the displacement boundary conditions are prescribed as shown in Figure 16. The DEM RVE is of the size of $10\text{cm} \times 10\text{cm} \times 5\text{cm}$, while the averaged grain diameter is 0.5cm. The micro-scale traction-separation law and the relation between the micro-structure and the permeability tensor on the interface are homogenized from a micro-scale RVE of discrete element particles. A set of displacement jump paths $\{u_n, u_s\}$ are applied to the micro-scale RVE, and the tractions $\{t_n, t_s\}$ are homogenized at each incremental deformation step. Furthermore, at each incremental step, we also construct a flow network inferred from the deformed configuration of the DEM assembly and use an inverse problem to compute the effective permeability in the tangential and normal directions, as shown in Figure 16.

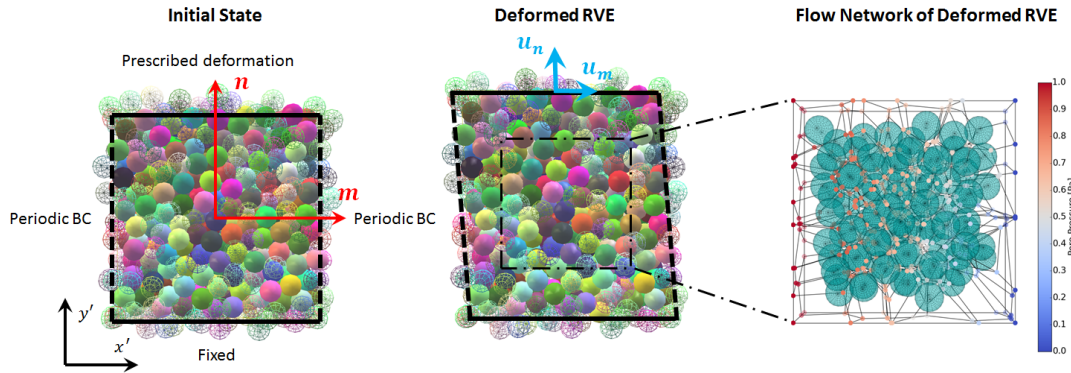


Fig. 16: Micro-scale RVE. The initial configuration of the granular assembly (LEFT), the deformed configuration of the granular assembly (MIDDLE), and the flow network generated from the deformed configuration used to predict the anisotropic effective permeability (RIGHT).

Before the displacement-driven grain-scale simulation begins, the DEM assembly must be in the stress state consistent to the macroscopic boundary condition. This is achieved by subjecting the DEM assembly with the right amount of shear and normal tractions along the boundaries.

The initial state of the micro-scale RVE is determined by the initial state of the macro-scale problem. The macro-scale fault with the inclination angle of 80° is under a confining pressure of 6 MPa in the X direction and 10 MPa in the Y direction.

$$\sigma_{\text{macro}}^{\text{Init}} = \begin{bmatrix} -6 & 0. \\ 0. & -10 \end{bmatrix}_{XY} \text{ MPa}, \quad (56)$$

where the subscript XY refers to the frame depicted in Figure 15. To introduce the proper initial stress state to the DEM assemblies, we first express this stress tensor in the local frame of the meso-scale RVE as

depicted in Figure 15 via coordinate transformation, i.e., $\sigma^{xy} = \mathbf{R}^T \cdot \sigma^{XY} \cdot \mathbf{R}$ such that,

$$\sigma_{\text{meso}}^{\text{Init}} = \begin{bmatrix} -9.88 & -0.68 \\ -0.68 & -6.12 \end{bmatrix}_{xy} \text{ MPa}, \quad (57)$$

where the subscript xy refers to the rotated frame for the meso-scale RVE. Note that since the DEM assemblies are aligned with the strong discontinuities in the meso-structures, the stress state is re-expressed in the two local coordinate systems such that the correct traction can be applied to the DEM assemblies to generate the correct initial stress state. Recall that the two local coordinate systems are 20 degrees apart from each other (one 10 degrees clockwise to the meso-scale frame, one 10 degrees counterclockwise to the meso-scale frame). As a result, we have,

$$\sigma_{\text{microRVE1}}^{\text{Init}} = \begin{bmatrix} -9.88 & -0.68 \\ -0.68 & -6.12 \end{bmatrix}_{x'_1 y'_1} \text{ MPa} ; \sigma_{\text{microRVE2}}^{\text{Init}} = \begin{bmatrix} -9.88 & 0.68 \\ 0.68 & -6.12 \end{bmatrix}_{x'_2 y'_2} \text{ MPa}. \quad (58)$$

Once the initial stress state of the DEM assemblies are set, we then run multiple simulations and collect the results to form the database for supervised machine learning. The choice of loading cases to be included in the training data set is often based on empirical knowledge. In this work, we adopt such design of proportional loading paths: in each loading case the ratio between the normal displacement u_n (along the unit vector \mathbf{n} in Fig. 16) and the tangential displacement u_s (along the unit vector \mathbf{m} in Fig. 16) remains a constant. In total the training data set contains 16 ratios: $\frac{u_n}{u_s} = \frac{i\pi}{8}, i = 0, 1, 2, \dots, 15$. Similar proportional loading paths have also been used to train constitutive laws for bulk materials in [Lefik and Schrefler, 2003]. In order for the data-driven model to learn the path-dependent behavior of the interface, we, for each loading ratio in our cases, prescribe the displacement such that, the norm of the displacement $u = \sqrt{u_n^2 + u_s^2}$ are prescribed with the following loading-unloading sequences: u first increases to 0.2 of the maximum displacement magnitude 0.01 m, then decreases to 0.1 of 0.01 m, and rise again to 0.4, then to 0.2, 0.6, 0.3, 0.8, 0.4, 1.0, 0.5. Note that, this design of training data set is suitable (but not necessarily optimized) for data-driven model used in finite element simulations in which the deformation paths of strong discontinuities are not known a priori. In many simulation cases, the major deformation paths could be anticipated. For instance, in shear band simulations, the shear effects predominate over the opening or closing of the interface. Thus the training data set to be constructed for these simulations should incorporate more shear-dominate loading paths. Nevertheless, the optimization of training data set is a challenging task and will be studied in future work. 5 additional loading paths for testing are also constructed. Some of them are monotonic loading, some have different loading-unloading sequences, and in some cases the ratio $\frac{u_n}{u_s}$ is not constant. Figure 17 shows a portion of the loading paths designed for machine learning. For brevity, we did not include all the available simulations in the database in the paper. Instead, we only show the results of 3 training sets and 3 testing sets, which are denoted as TR1, TR2 and TR3 and TE1, TE2 and TE3 respectively. Nevertheless, the training and validation algorithm as well as the database itself will be made available in an open source repository. The discrepancy between the data from micro-scale DEM simulation X_{DEM} and the results predicted by LSTM neural network X_{LSTM} is quantified by the scaled mean squared error given by

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N [\text{MinMaxScaler}(X_{\text{DEM}_i}) - \text{MinMaxScaler}(X_{\text{LSTM}_i})]^2, \quad (59)$$

where N is the number of data points. X_{DEM} and X_{LSTM} are re-scaled to be within the range $[0, 1]$ using the MinMaxScaler as described in the data preprocessing for machine learning in Section 4.6.

The physical parameters for the input of the LSTM neural network are the sequence of history values at time $[t_{n-1}, t_n, t_{n+1}]$ of the normal and tangential components of displacement jump, the sequence of history values at time $[t_{n-2}, t_{n-1}, t_n]$ of the normal and tangential components of traction, and the sequence of history values at time $[t_{n-2}, t_{n-1}, t_n]$ of the maximum and minimum principal values of the fabric tensor of the DEM RVE. The outputs of the LSTM neural network are the normal and tangential components of traction at time t_{n+1} , the maximum and minimum principal values of the fabric tensor at time t_{n+1} , and the permeabilities normal and tangential to the strong discontinuity at time t_{n+1} .

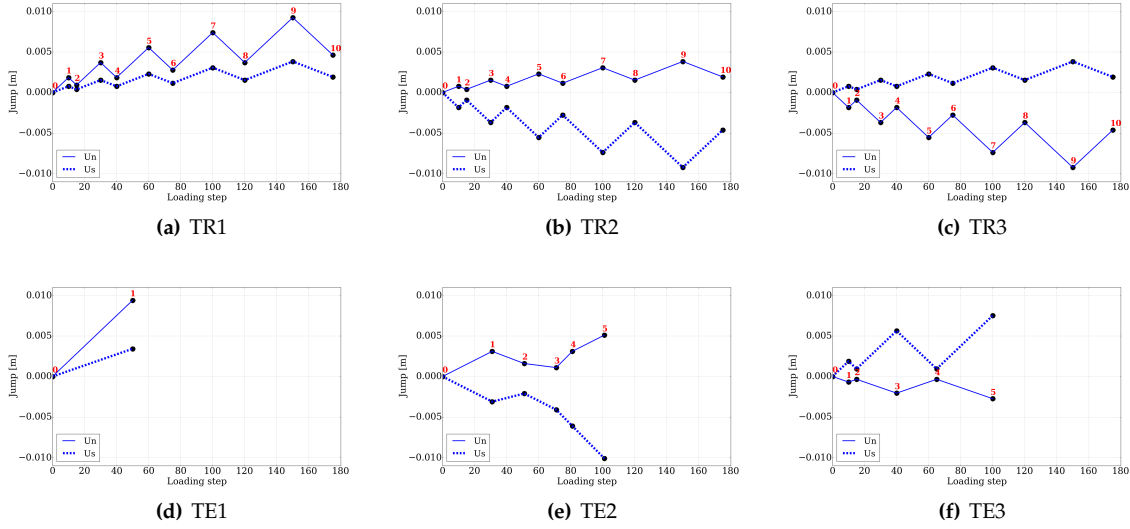


Fig. 17: Loading path of three selected training cases TR1, TR2, TR3 and three selected testing cases TE1, TE2, TE3 on the micro-scale RVE 1. u_n and u_s are the normal and tangential displacement jumps. The coordinate system is $\{m_1, n_1\}$ (or $\{x'_1, y'_1\}$) depicted in Fig. 15. It can be seen that TR1 and TR2 represent tensile-shear loading cases (as u_n is positive) and TR3 represents a compressive-shear loading case (u_n negative). The numbers mark the sequence of loading-unloading cycles.

Since in many engineering applications, the flow injection rate is transient and changes with time, the data-driven traction-separation laws must be able to capture the resultant combined isotropic and kinematic hardening mechanisms. Figures 18 and 19 show the comparisons between the DEM simulations and the simulated mechanical responses generated from the recurrent neural network in the normal and tangential directions. Except for the testing case TE2 in which there are notable discrepancy when the thin DEM layers are reloaded, the meso-scale data-driven traction-separation law is able to replicate both the cyclic and monotonic loading responses with negligible errors. Remarkably, this is achieved without using any internal variables to capture the history-dependent effect. Furthermore, we also show that the predicted responses are able to simulate both the damage-plastic flow and the elastic unloading in the cyclic responses. This coupled damage-plastic response is attributed to the evolution of the fabric tensors.

Figures 20 and 21 show the maximum and minimum eigenvalue of the fabric tensors following the prescribed displacements obtained from DEM and from the RNN predictions. The RNN generated responses are able to deliver very accurate predictions of the fabric tensor evolution. This good match is important for predicting induced anisotropy and may explain why the traction predictions in Figures 18 and 19 match well with the database.

The predictions of normal and tangential permeabilities following the prescribed displacements are shown in Figures 22 and 23 respectively. Again, with the help of characteristic microstructure information, the match is satisfying. To sum up, the trained data-driven model is capable of representing the micro-scale DEM-flow network model, in terms of traction, permeability, and invariants of fabric tensor. It is ready to be used as constitutive law for the strong discontinuity of the meso-scale RVE in the FEM-LSTM coupled simulations.

5.2 Training and validation of material laws for dual-porosity fault

The path-dependent constitutive model governing the displacement jump induced traction and permeability changes in the macroscopic sealing fault is provided by the macro-scale data-driven LSTM model. The data used to train and test this model are generated from multiscale simulations of the meso-scale

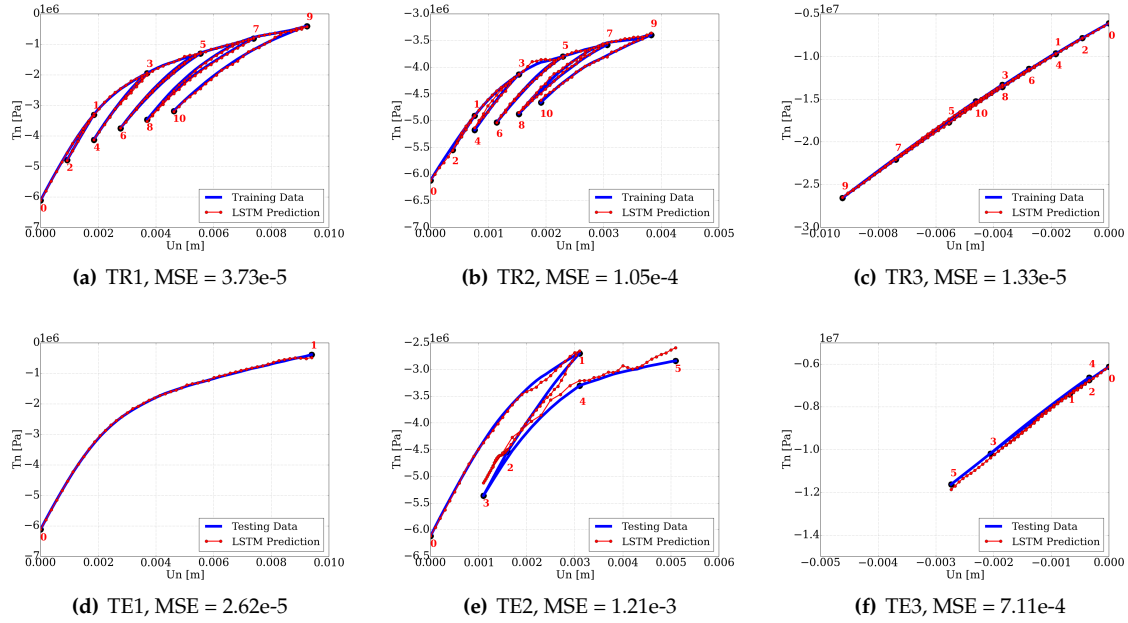


Fig. 18: Comparison of the micro-scale DEM simulation data and the trained meso-scale data-driven model. Normal traction against normal displacement jump for the selected training and testing cases. The numbers mark the sequence of loading-unloading cycles. MSE refers to the scaled mean squared error defined in Eq. 59.

RVE, where the interface behavior comes from the micro-scale RVE. The multiscale model is FEM-LSTM coupled, using the data-driven model trained in the previous section.

In each RVE simulation, the displacement boundary conditions are prescribed as shown in Figure 24. The meso-scale RVE is 2D, and has the size of 1m X 1m. A set of displacement jump paths $\{u_n, u_s\}$ are applied to the meso-scale RVE, and the tractions $\{t_n, t_s\}$ are homogenized at each incremental deformation step. Furthermore, at each incremental step, we also conduct an inverse problem to compute the effective permeability in the tangential and normal directions, as shown in Figure 24.

It is important to investigate how well the FEM-LSTM coupled scheme represents the FEM-DEM multiscale scheme, where the interface constitutive law comes from the DEM assembly from the previous section. An example of comparison is presented, where the meso-scale RVE is subjected to a displacement loading path with unloading. Fig. 25 compares the tractions in the normal and tangential directions following the prescribed displacements. The results are close to each other. Thus FEM-LSTM model could approximately represent the FEM-DEM model in the generation of a database. Another alternative is using hybrid database. In other words, a portion of the data is from FEM-LSTM model simulations, while the other portion is from FEM-DEM model. In the extreme case, all data are from FEM-DEM model, then the numerical example in this work will be two-scale data-driven simulation, instead of triple-scale.

We then run multiple meso-scale simulations and collect the results to form the database for supervised machine learning. The design of the training and testing data set is similar to the design in the previous section. In the meso-scale RVE, there is no definition of fabric tensor, thus the input data only consists of the displacement jumps and tractions in normal and tangential directions. Figure 26 shows a portion of the loading paths: TR1, TR2 and TR3 in the training sets and TE1, TE2 and TE3 in the testing sets. The physical parameters for the input of the LSTM neural network are the sequence of history values at time $[t_{n-1}, t_n, t_{n+1}]$ of the normal and tangential components of displacement jump, and the sequence of history values at time $[t_{n-2}, t_{n-1}, t_n]$ of the normal and tangential components of traction. The outputs of the LSTM neural network are the normal and tangential components of traction at time t_{n+1} , and the permeabilities normal and tangential to the strong discontinuity at time t_{n+1} .

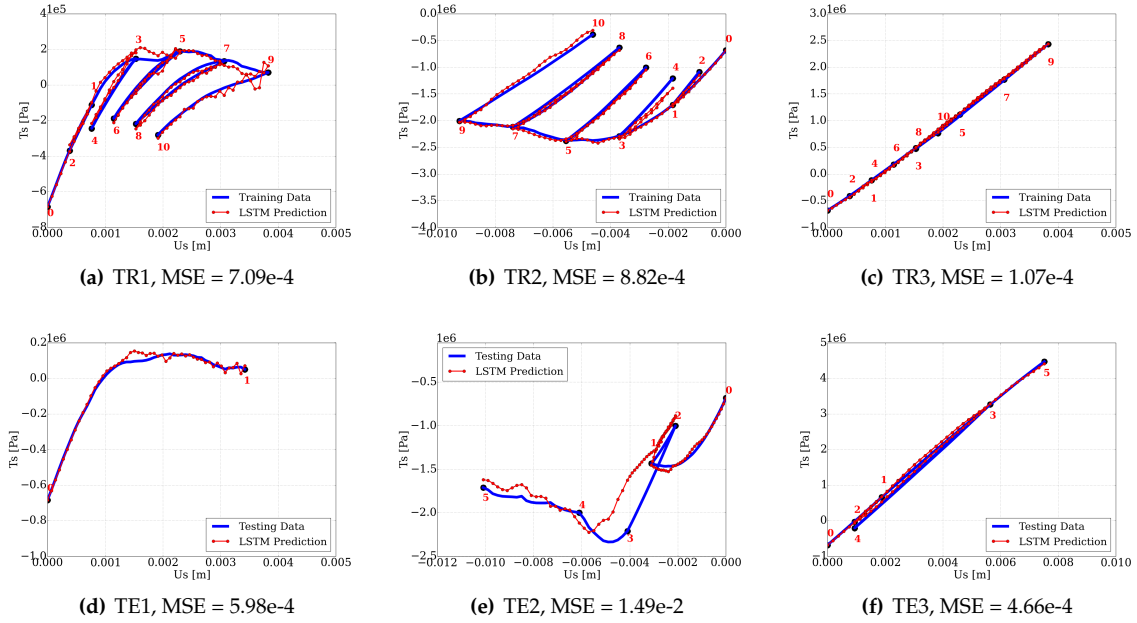


Fig. 19: Comparison of the micro-scale DEM simulation data and the trained meso-scale data-driven model. Tangential traction against tangential displacement jump for the selected training and testing cases. The numbers mark the sequence of loading-unloading cycles. MSE refers to the scaled mean squared error defined in Eq. 59.

Figures 27 and 28 shows the comparison between the FEM-LSTM simulations and the simulated mechanical responses generated from the recurrent neural network in the normal and tangential directions. The predicted responses are able to simulate both the damage-plastic flow and the elastic unloading in the cyclic responses. The predictions of normal and tangential permeabilities following the prescribed displacements are shown in Figures 29 and 30 respectively. The trained data-driven model is ready to be used as constitutive law for the strong discontinuity of the macro-scale problem.

5.3 Simulation of macro-scale fault reactivation problem

Water is injected to the macro-scale field through the source S located to the right of the sealing fault. The distance between S and the fault is about 250 m (Fig. 32(a)). The prescribed Darcy velocity at the source is shown in Fig. 31(a). The injection profile is composed of three injection-pause cycles, where water supply is provided for 40 hours under constant rate of 50 m/s, followed by a pause for 10 hours before the next cycle of injection. We simulate the hydro-mechanical dual-porosity problem with the traction-separation law and macropore permeability tensors along the sealing fault given by the meso-scale data-driven model. The pore pressure in both scales at the source S is presented in Fig. 31(b). The fluid is injected to the macropore space. Upon injection or pause, the macropore injection pressure jumps up or plunge immediately, while the micropore pressure at the injection point has the opposite behavior. This is caused by the low mass transfer permeability between the macropores and micropores. Then in the transient regime, when fluid gradually diffuses into the micropores by mass transfer, micropore pressure slowly approaches the macropore pressure. The two pressure will eventually be identical when the diffusion between pores reaches equilibrium. To show the influence of the inter-pore transfer, we present an additional case where the transfer parameter $\bar{\alpha}$ is ten times higher (Fig. 31(c)). The discrepancy between the pressures is significantly reduced. The following results are from the low inter-pore transfer case.

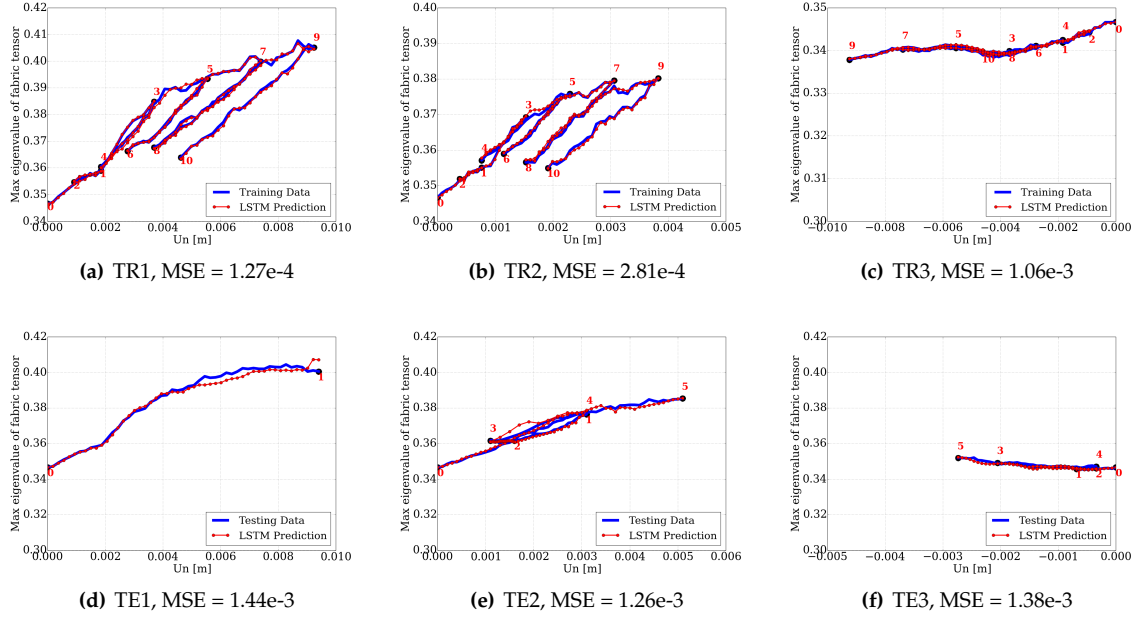


Fig. 20: Comparison of the micro-scale DEM simulation data and the trained meso-scale data-driven model. Maximum eigenvalue of fabric tensor against normal displacement jump for the selected training and testing cases. The numbers mark the sequence of loading-unloading cycles. MSE refers to the scaled mean squared error defined in Eq. 59.

To illustrate the hydraulic response, the macropore pressure field at time 40 h, 100 h and 140 h are presented in Fig. 32. The pressure plume is in the form of circle and expands with increasing amount of water injected through the source ($t = 40$ h). The pore pressure drops when the injection pauses, but the plume is still expanding ($t = 100$ h) driven by the excess pore pressure that has not been entirely diffused. When the injection is resumed, the pore pressure also rises again. The form of the pressure plume is disturbed when it reaches the sealing fault, which has a two-order-lower macropore permeability ($t = 140$ h). As for the micropore pressure field, it has a similar but delayed evolution behavior, due to the time required for the fluid transfer between macropores and micropores. The local pressure responses are illustrated in Fig. 33 for three locations in the fault indicated in Fig. 32(a). The distances to the source point $d_{SB} > d_{SA} > d_{SC}$. It is seen that the closer a point is to the source, the faster the pressure increases upon injection. There is also significant pressure gradient across the less-permeable fault. The difference between macropore and micropore pressure is due to the different permeability in macropores and micropores for the fluid to diffuse in the macro-scale field, and also the low transfer permeability between pores.

Due to the fully coupled nature of the problem, the mechanical responses of the porous solid, especially the displacement jump and traction at the strong discontinuity, strongly depend on how pore fluid diffuses inside the pore space. The evolution of macro-scale mean effective stress field during the fluid injection cycles is shown in Fig. 34.

It is clear that this field has the same pattern as the pore pressure field. The increase in the mean effective stress is due to the increase in pore pressure, in agreement to the effective stress principle. The evolution of macro-scale differential stress field (Fig. 35) is a combined effect of the far field differential stress, fluid injection and presence of sealing fault. The decrease in normal compression traction makes the fault surface unable to sustain the shear traction, and the fault starts to mobilize.

The local displacement and traction responses are illustrated in Fig. 36 and Fig. 37 respectively for three locations in the fault indicated in Fig. 32(a). The traction states that the material at A, B, C experienced during the injection-pause cycles are depicted in Fig. 38.

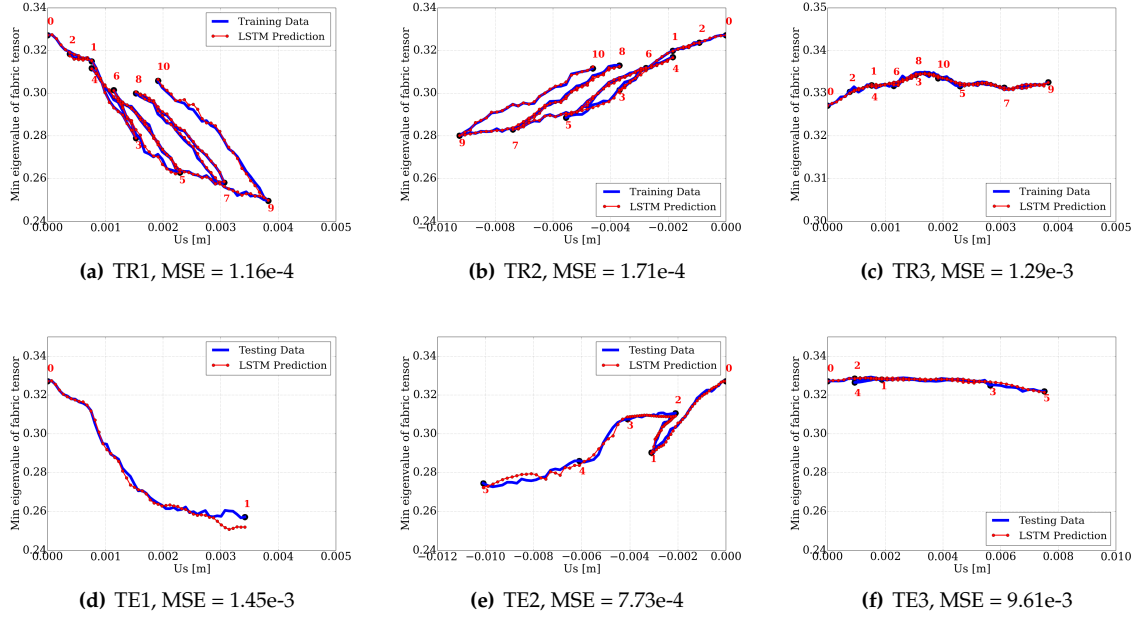


Fig. 21: Comparison of the micro-scale DEM simulation data and the trained meso-scale data-driven model. The minimum eigenvalue of fabric tensor against tangential displacement jump for the selected training and testing cases. The numbers mark the sequence of loading-unloading cycles. MSE refers to the scaled mean squared error defined in Eq. 59.

These results clearly demonstrate the capacity of our data-driven model in capturing the complex and path-dependent interface behaviors. This is a significant improvement over the phenomenological traction-separation laws where idealized tensile and shear (linear or exponential) behavior is often adopted [Park and Paulino, 2011]. The data-driven model can preserve important mechanical properties of the interface from sub-scale structures while reducing the computational costs compared to full micro-scale models such as DEM.

6 Conclusion

We introduce a recursive multiscale framework that captures the hydro-mechanical responses of multi-permeability porous media with embedded strong discontinuities across different length scales. Using the directed graph that represents the hierarchy of the numerical models as the starting point, we identify the knowledge gap and the weakest link of a multiscale multiphysics multi-permeability model and replace this portion of the computational model with a data-driven counterpart. By creating, training and validating recurrent neural network that has the capacity to memorize and interpret history-dependent events, we established a new recursive data-driven approach where information from multiple sub-scales can be used sequentially to generate macroscopic prediction in a cost-efficient manner. The triple-scale coupling simulations are validated at each sub-scale level where data set for training and validation are mutually exclusive to each other. Finally, we also address a number of technical issues, such as lack of objectivity, vanishing and exploding gradients and the over-fitting issues to ensure the robustness and accuracy of the numerical simulations. This hybrid data-driven modeling approach may play a critical role for analyzing problems where human-interpretable knowledge is sufficient to bring closure for forward predictions and for linking simulations across more than two scales in a cost-efficient manner.

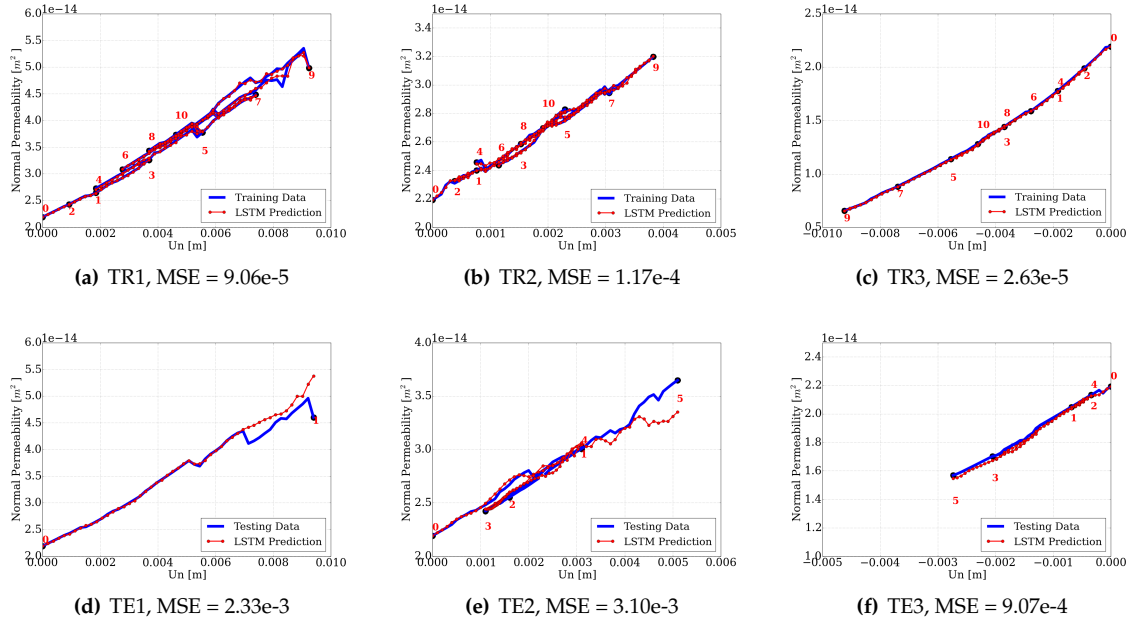


Fig. 22: Comparison of the micro-scale DEM simulation data and the trained meso-scale data-driven model. Normal permeability against normal displacement jump for the selected training and testing cases. The numbers mark the sequence of loading-unloading cycles. MSE refers to the scaled mean squared error defined in Eq. 59.

7 Acknowledgments

This research on the data-driven model is partially supported by the Earth Materials and Processes program from the US Army Research Office under grant contract W911NF-15-1-0581 and the Dynamic Materials and Interactions Program from the Air Force Office of Scientific Research under grant contract FA9550-17-1-0169, while on analysis on the multi-permeability porous media is supported by the nuclear energy university program from department of energy under grant contract DE-NE0008534. These supports are gratefully acknowledged.

The views and conclusions contained in this document are those of the authors, and should not be interpreted as representing the official policies, either expressed or implied, of the sponsors, including the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for government purposes notwithstanding any copyright notation herein.

References

- Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, et al. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:1603.04467*, 2016.
- Jean-Louis Auriault, Claude Boutin, and Christian Geindreau. *Homogenization of coupled phenomena in heterogeneous media*, volume 149. John Wiley & Sons, 2010.
- Mao Bai, Derek Elsworth, and Jean-Claude Roegiers. Multiporosity/multipermeability approach to the simulation of naturally fractured reservoirs. *Water Resources Research*, 29(6):1621–1634, 1993.
- Ted Belytschko, Wing Kam Liu, Brian Moran, and Khalil Elkhodary. *Nonlinear finite elements for continua and structures*. John Wiley & Sons, 2013.

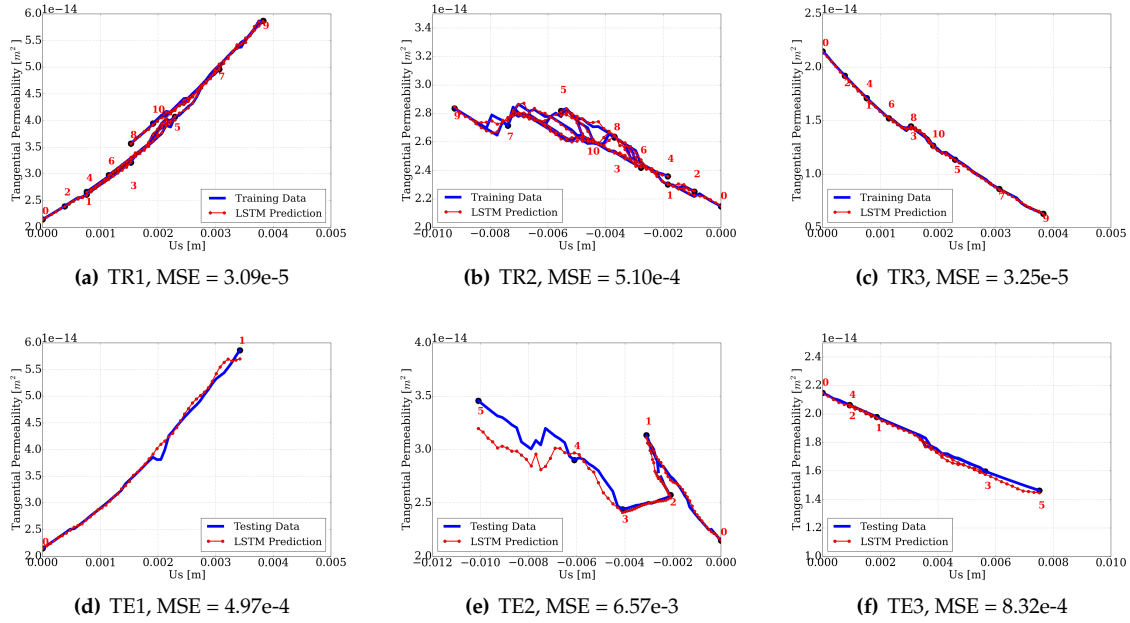


Fig. 23: Comparison of the micro-scale DEM simulation data and the trained meso-scale data-driven model. Tangential permeability against tangential displacement jump for the selected training and testing cases. The numbers mark the sequence of loading-unloading cycles. MSE refers to the scaled mean squared error defined in Eq. 59.

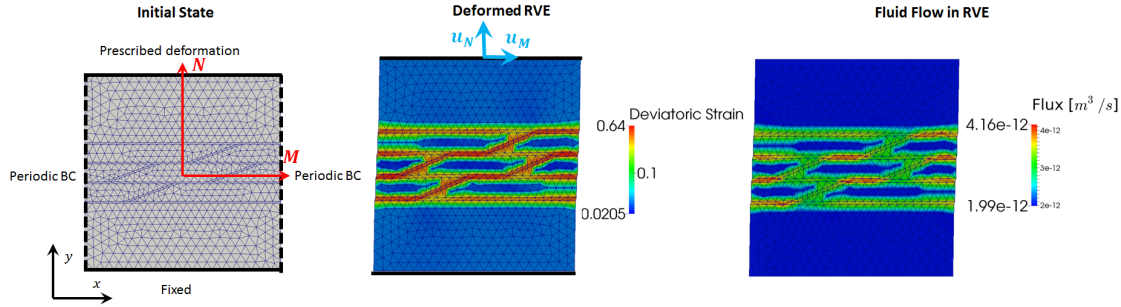


Fig. 24: Meso-scale RVE. The initial configuration of the meso-scale RVE and its pre-embedded interfaces (LEFT), the deformed configuration and the deviatoric strain field of the meso-scale RVE (MIDDLE), and the fluid flux calculation used to predict the anisotropic effective permeability (RIGHT).

- 1085 Maurice A Biot. General theory of three-dimensional consolidation. *Journal of applied physics*, 12(2):155–164,
1086 1941.
- 1087 Ronaldo I Borja. A finite element model for strain localization analysis of strongly discontinuous fields
1088 based on standard galerkin approximation. *Computer Methods in Applied Mechanics and Engineering*, 190
1089 (11):1529–1549, 2000.
- 1090 Ronaldo I Borja and Claudio Tamagnini. Cam-clay plasticity part iii: Extension of the infinitesimal model
1091 to include finite strains. *Computer Methods in Applied Mechanics and Engineering*, 155(1-2):73–95, 1998.
- 1092 Steven L Bryant, Peter R King, and David W Mellor. Network model evaluation of permeability and spatial
1093 correlation in a real random sphere packing. *Transport in Porous Media*, 11(1):53–70, 1993.
- 1094 C Callari and Francisco Armero. Finite element methods for the analysis of strong discontinuities in cou-
1095 pled poro-plastic media. *Computer Methods in Applied Mechanics and Engineering*, 191(39):4371–4400, 2002.

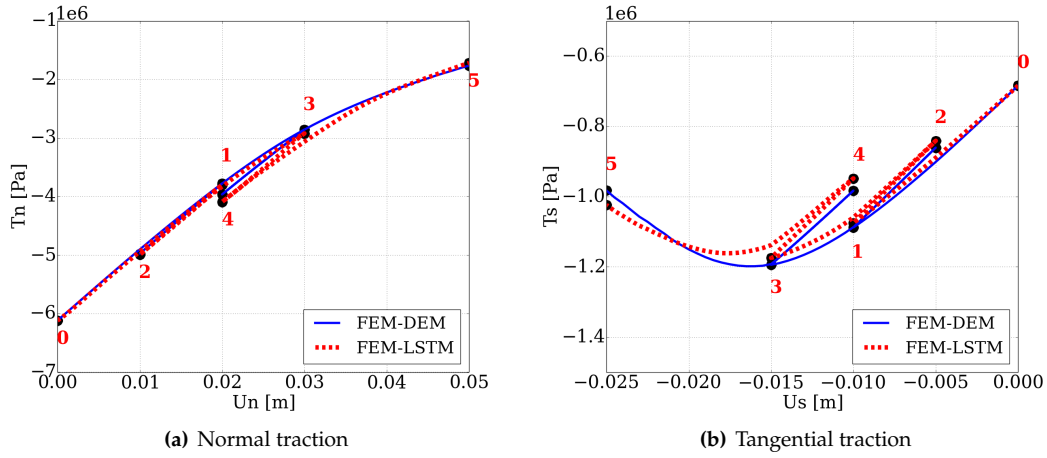


Fig. 25: Comparison of FEM-LSTM coupled model and FEM-DEM coupled model for meso-scale RVE. The numbers mark the sequence of loading-unloading cycles.

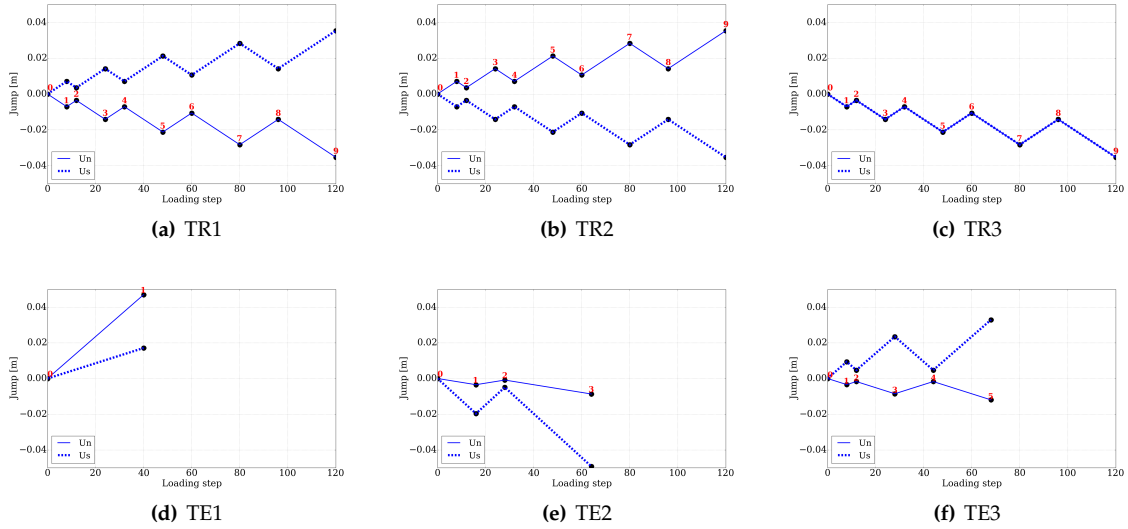


Fig. 26: Loading path of three selected training cases TR1, TR2, TR3 and three selected testing cases TE1, TE2, TE3 on the meso-scale RVE. u_n and u_s are the normal and tangential displacement jumps. The coordinate system is $\{M, N\}$ (or $\{x, y\}$) depicted in Fig. 15. It can be seen that TR2 represents a tensile-shear loading case (as u_n is positive), TR1 and TR3 represent compressive-shear loading cases (u_n negative). The numbers mark the sequence of loading-unloading cycles.

- 1096 C Callari and F Federico. Fem validation of a double porosity elastic model for consolidation of structurally
 1097 complex clayey soils. *International journal for numerical and analytical methods in geomechanics*, 24(4):367–
 1098 402, 2000.
- 1099 C Callari, F Armero, and A Abati. Strong discontinuities in partially saturated poroplastic solids. *Computer
 1100 Methods in Applied Mechanics and Engineering*, 199(23):1513–1535, 2010.
- 1101 Bruno Chareyre, Andrea Cortis, Emanuele Catalano, and Eric Barthélemy. Pore-scale modeling of viscous
 1102 flow and induced forces in dense sphere packings. *Transport in porous media*, 94(2):595–615, 2012.

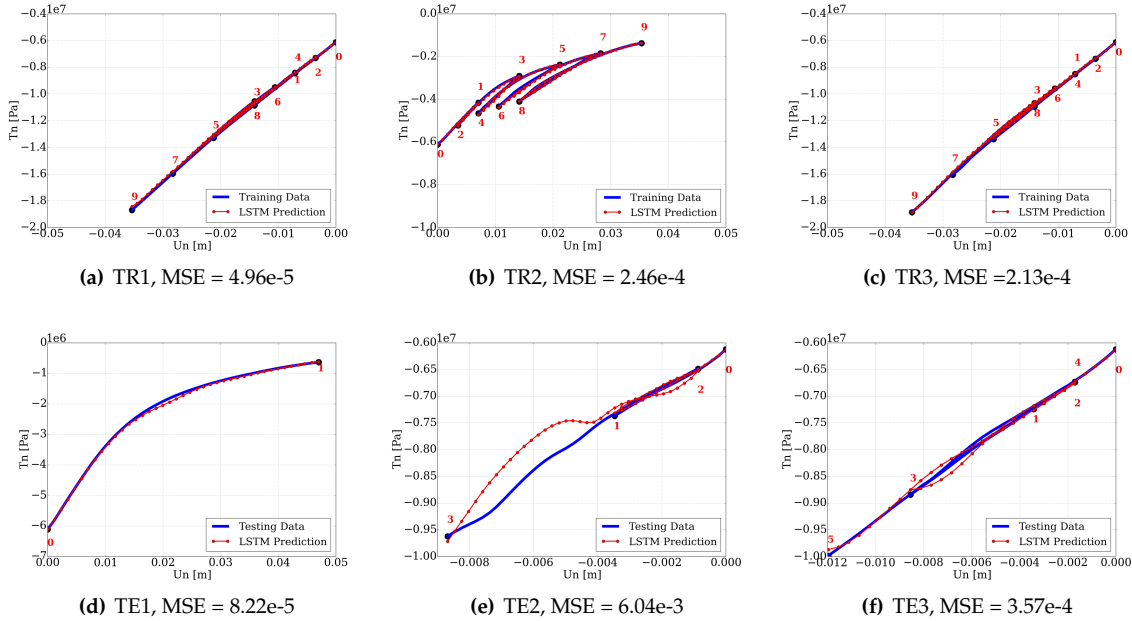


Fig. 27: Comparison of the meso-scale FEM-LSTM simulation data and the trained macro-scale data-driven model. Normal traction against normal displacement jump for the selected training and testing cases. The numbers mark the sequence of loading-unloading cycles. MSE refers to the scaled mean squared error defined in Eq. 59.

- 1103 François Chollet et al. Keras. <https://github.com/fchollet/keras>, 2015.
- 1104 Jinhyun Choo and Ronaldo I Borja. Stabilized mixed finite elements for deformable porous media with
1105 double porosity. *Computer Methods in Applied Mechanics and Engineering*, 293:131–154, 2015.
- 1106 Jinhyun Choo, Joshua A White, and Ronaldo I Borja. Hydromechanical modeling of unsaturated flow in
1107 double porosity media. *International Journal of Geomechanics*, page D4016002, 2016.
- 1108 A Cochocki and Rolf Unbehauen. *Neural networks for optimization and signal processing*. John Wiley &
1109 Sons, Inc., 1993.
- 1110 Olivier Coussy. *Poromechanics*. John Wiley & Sons, 2004.
- 1111 Yannis F Dafalias and Majid T Manzari. Simple plasticity sand model accounting for fabric change effects.
1112 *Journal of Engineering mechanics*, 130(6):622–634, 2004.
- 1113 X Du and M Ostoj-Starzewski. On the size of representative volume element for darcy law in random
1114 media. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*,
1115 volume 462, pages 2949–2963. The Royal Society, 2006.
- 1116 Herbert Edelsbrunner and Nimish R Shah. Incremental topological flipping works for regular triangula-
1117 tions. *Algorithmica*, 15(3):223–241, 1996.
- 1118 Derek Elsworth and Mao Bai. Flow-deformation response of dual-porosity media. *Journal of Geotechnical*
1119 *Engineering*, 118(1):107–124, 1992.
- 1120 Frédéric Feyel and Jean-Louis Chaboche. Fe 2 multiscale approach for modelling the elastoviscoplastic
1121 behaviour of long fibre sic/ti composite materials. *Computer methods in applied mechanics and engineering*,
1122 183(3):309–330, 2000.
- 1123 Jacob Fish. *Practical multiscale modeling*. John Wiley & Sons, 2013.
- 1124 Jacob Fish and Wei Wu. A nonintrusive stochastic multiscale solver. *International Journal for Numerical*
1125 *Methods in Engineering*, 88(9):862–879, 2011.
- 1126 CD Foster, RA Regueiro, AF Fossum, and RI Borja. Implicit numerical integration of a three-invariant,
1127 isotropic/kinematic hardening cap plasticity model for geomaterials. *Computer Methods in Applied Me-*
1128 *chanics and Engineering*, 194(50):5109–5138, 2005.

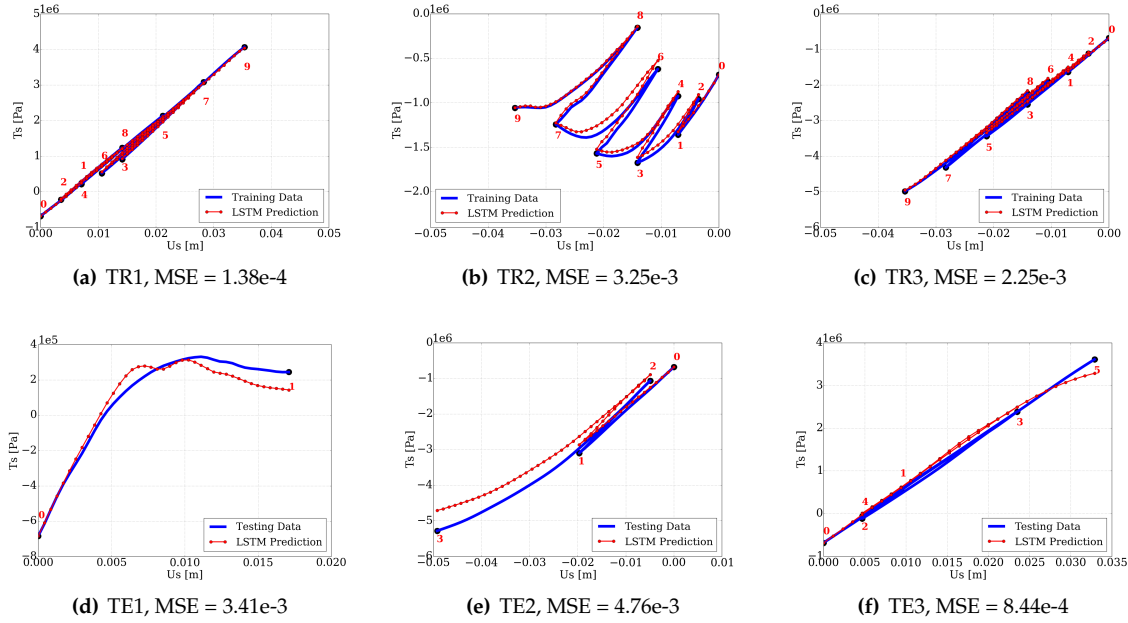


Fig. 28: Comparison of the meso-scale FEM-LSTM simulation data and the trained macro-scale data-driven model. Tangential traction against tangential displacement jump for the selected training and testing cases. The numbers mark the sequence of loading-unloading cycles. MSE refers to the scaled mean squared error defined in Eq. 59.

- Ilona Frankenreiter, Daniele Rosato, and Christian Miehe. Hybrid micro-macro-modeling of evolving anisotropies and length scales in finite plasticity of polycrystals. *PAMM*, 11(1):515–518, 2011.
- Jerome H Friedman. On bias, variance, 0i—loss, and the curse-of-dimensionality. *Data mining and knowledge discovery*, 1(1):55–77, 1997.
- Pengcheng Fu and Yannis F Dafalias. Fabric evolution within shear bands of granular materials and its relation to critical state theory. *International Journal for numerical and analytical methods in geomechanics*, 35(18):1918–1948, 2011.
- Tomonari Furukawa and Genki Yagawa. Implicit constitutive modelling for viscoplasticity using neural networks. *International Journal for Numerical Methods in Engineering*, 43(2):195–219, 1998.
- Matt W Gardner and SR Dorling. Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmospheric environment*, 32(14):2627–2636, 1998.
- Marc GD Geers, Varvara G Kouznetsova, and WAM Brekelmans. Multi-scale computational homogenization: Trends and challenges. *Journal of computational and applied mathematics*, 234(7):2175–2182, 2010.
- J Ghaboussi, JH Garrett Jr, and Xiping Wu. Knowledge-based modeling of material behavior with neural networks. *Journal of engineering mechanics*, 117(1):132–153, 1991.
- Wolfgang Graf, Steffen Freitag, Michael Kaliske, and J-U Sickert. Recurrent neural networks for uncertain time-dependent structural behavior. *Computer-Aided Civil and Infrastructure Engineering*, 25(5):322–323, 2010.
- Ning Guo and Jidong Zhao. The signature of shear-induced anisotropy in granular media. *Computers and Geotechnics*, 47:1–15, 2013.
- Ning Guo and Jidong Zhao. A coupled fem/dem approach for hierarchical multiscale modelling of granular media. *International journal for numerical methods in engineering*, 99(11):789–818, 2014.
- Ning Guo, Jidong Zhao, and WaiChing Sun. Multiscale analysis of shear failure of thick-walled hollow cylinder in dry sand. *Géotechnique Letters*, 6(1), 2016.
- Markus Hilpert, Roland Glantz, and Cass T Miller. Calibration of a pore-network model by a pore-morphological analysis. *Transport in Porous Media*, 51(3):267–285, 2003.

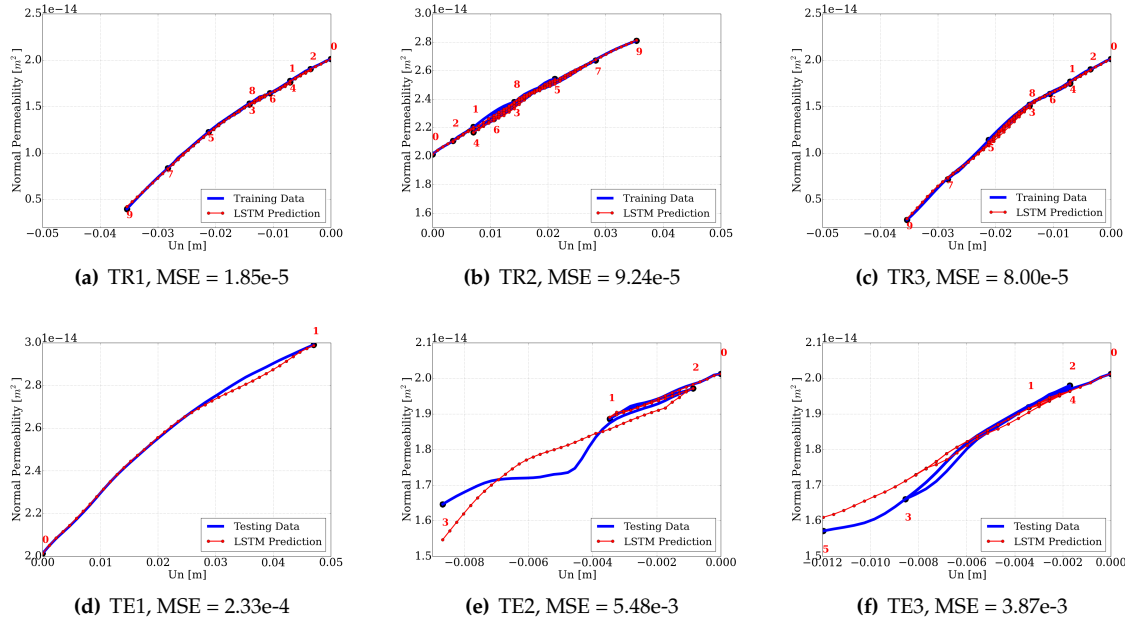


Fig. 29: Comparison of the meso-scale FEM-LSTM simulation data and the trained macro-scale data-driven model. Normal permeability against normal displacement jump for the selected training and testing cases. The numbers mark the sequence of loading-unloading cycles. MSE refers to the scaled mean squared error defined in Eq. 59.

- Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.
- Jayesh R Jain and Somnath Ghosh. Damage evolution in composites with a homogenization-based continuum damage mechanics model. *International Journal of Damage Mechanics*, 18(6):533–568, 2009.
- Yuntao Ji, Stephen A Hall, Patrick Baud, and Teng-fong Wong. Characterization of pore structure and strain localization in majella limestone by x-ray computed tomography and digital image correlation. *Geophysical Journal International*, 200(2):701–719, 2015.
- A Johari, G Habibagahi, and A Ghahramani. Prediction of soil–water characteristic curve using genetic programming. *Journal of Geotechnical and Geoenvironmental Engineering*, 132(5):661–665, 2006.
- Sungmoon Jung and Jamshid Ghaboussi. Neural network constitutive model for rate-dependent materials. *Computers & Structures*, 84(15):955–963, 2006.
- Soteris A Kalogirou. Artificial neural networks in renewable energy systems applications: a review. *Renewable and sustainable energy reviews*, 5(4):373–401, 2001.
- Shahriyar Keshavarz and Somnath Ghosh. Multi-scale crystal plasticity finite element model approach to modeling nickel-based superalloys. *Acta Materialia*, 61(17):6549–6561, 2013.
- Trenton Kirchdoerfer and Michael Ortiz. Data-driven computational mechanics. *Computer Methods in Applied Mechanics and Engineering*, 304:81–101, 2016.
- Trenton Kirchdoerfer and Michael Ortiz. Data driven computing with noisy material data sets. *arXiv preprint arXiv:1702.01574*, 2017.
- J Maximilian Köhne, Binayak P Mohanty, and Jirka Šimůnek. Inverse dual-permeability modeling of preferential water flow in a soil column and implications for field-scale solute transport. *Vadose Zone Journal*, 5(1):59–76, 2006.
- Andreas Krischok and Christian Linder. On the enhancement of low-order mixed finite element methods for the large deformation analysis of diffusion in solids. *International Journal for Numerical Methods in Engineering*, 106(4):278–297, 2016.

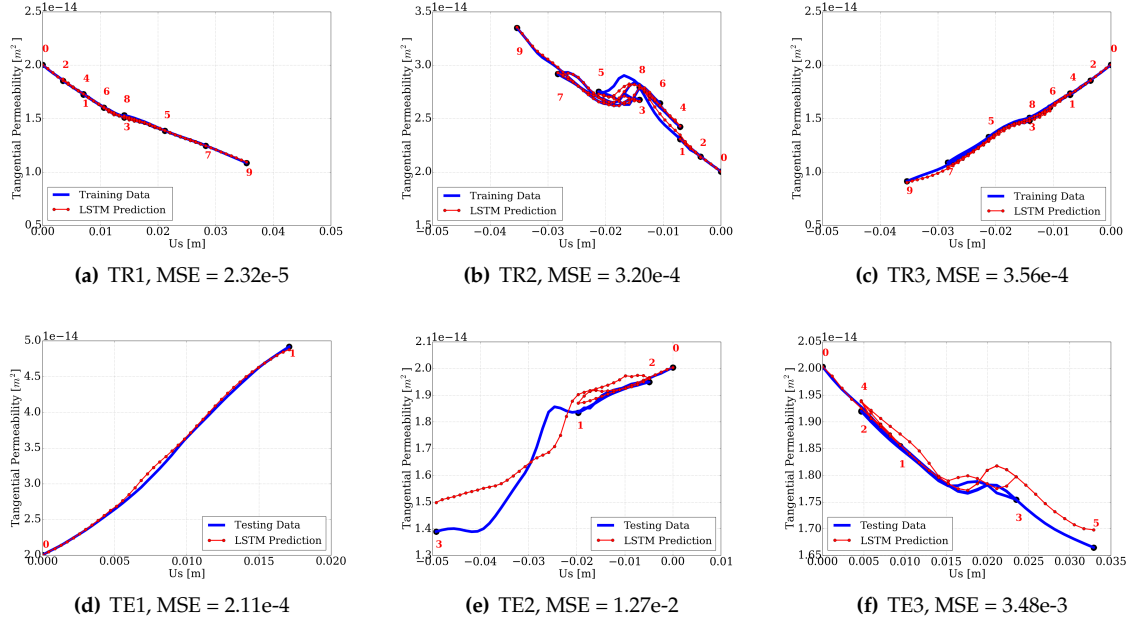


Fig. 30: Comparison of the meso-scale FEM-LSTM simulation data and the trained macro-scale data-driven model. Tangential permeability against tangential displacement jump for the selected training and testing cases. The numbers mark the sequence of loading-unloading cycles. MSE refers to the scaled mean squared error defined in Eq. 59.

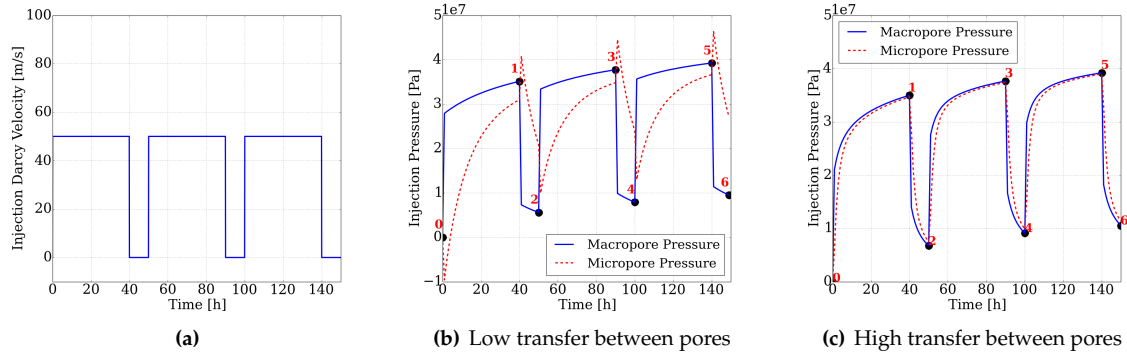


Fig. 31: Water supply in the macro-scale fault reactivation problem. (a) Time history of the prescribed injection velocity in macropores at the source point. (b) Computed responses of injection pressure in macropore and micropore at the source point (transfer parameter $\bar{\alpha} = \rho_f * k_m$). (c) Computed pressures in a comparison simulation where the transfer parameter $\bar{\alpha} = 10 * \rho_f * k_m$. The numbers mark the sequence of injection-pause cycles.

- 1180 Petr Krysl and Lance Endres. Explicit newmark/verlet algorithm for time integration of the rotational
 1181 dynamics of rigid bodies. *International journal for numerical methods in engineering*, 62(15):2154–2177, 2005.
 1182 Kristopher L Kuhlman, Jason E Heath, W Payton Gardner, and David G Robinson. Multiporosity flow of
 1183 gases in fractured shale formations. *Journal of Coal Geology*, 109(110):101–146, 2013.
 1184 Kristopher L Kuhlman, Bwalya Malama, and Jason E Heath. Multiporosity flow in fractured low-
 1185 permeability rocks. *Water Resources Research*, 51(2):848–860, 2015.

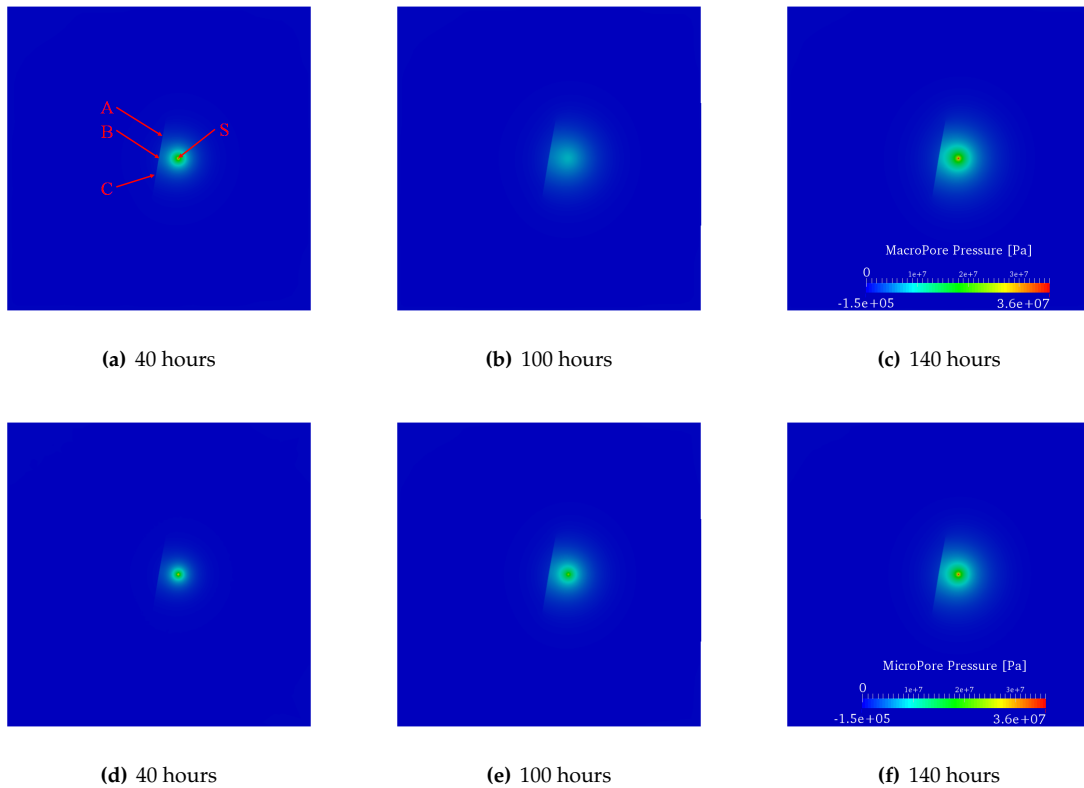


Fig. 32: Evolution of macropore pressure (a-c) and micropore pressure (d-f) field. S denotes the fluid source. A, B, C are three locations on the sealing fault. 40h is the end of the first injection, 100h is the end of the second pause, and 140h is the end of the third injection.

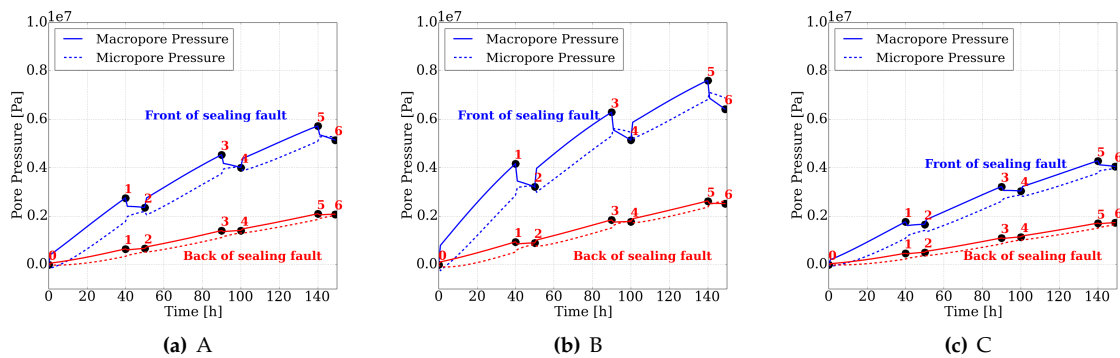


Fig. 33: Time history of local macropore and micropore pressure at location A, B, C (Fig. 32(a)) of the sealing fault. "Front" refers to the side of fault that is facing the source point. "Back" is another side that is away from the source. The numbers mark the sequence of injection-pause cycles.

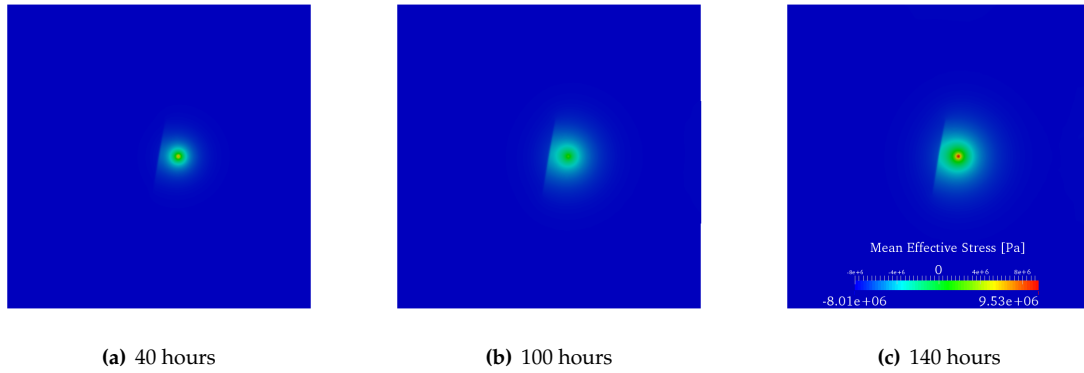


Fig. 34: Evolution of the mean effective stress field in the macro-scale simulation. 40h is the end of the first injection, 100h is the end of the second pause, and 140h is the end of the third injection.

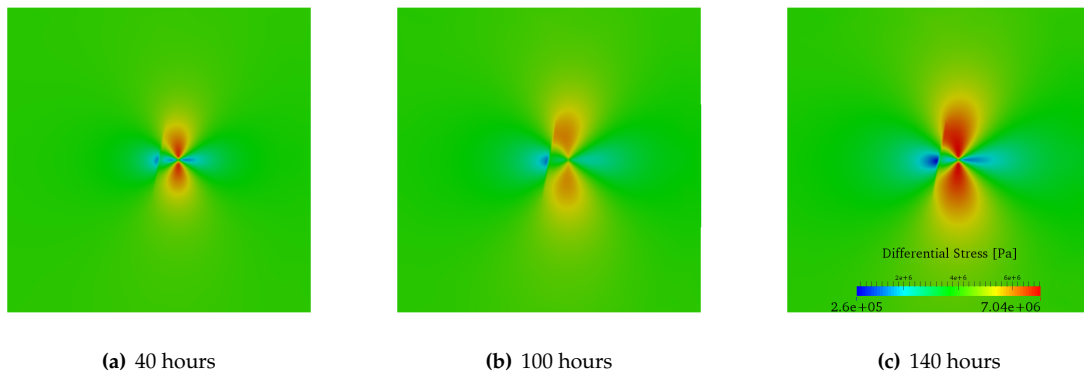


Fig. 35: Evolution of the differential stress field in the macro-scale simulation. 40h is the end of the first injection, 100h is the end of the second pause, and 140h is the end of the third injection.

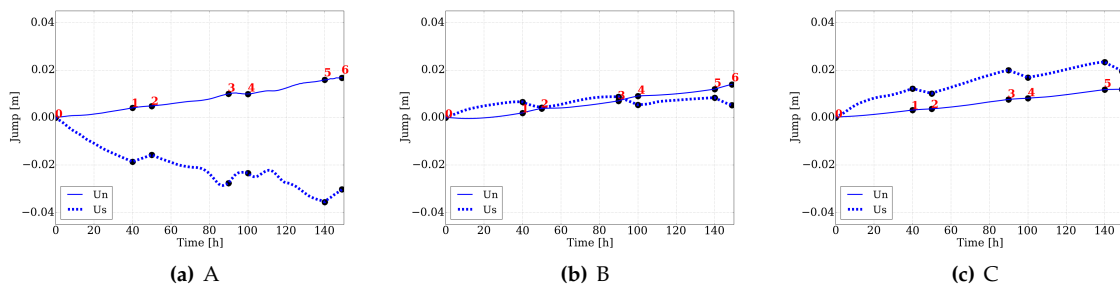


Fig. 36: Time history of normal and tangential displacement jumps at location A, B, C (Fig. 32(a)) of the sealing fault. The coordinate system is $\{M, N\}$ (or $\{x, y\}$) depicted in Fig. 15. The numbers mark the sequence of injection-pause cycles.

1188 Krzysztof Lamorski, Yakov Pachepsky, C Sławinski, and RT Walczak. Using support vector machines to
 1189 develop pedotransfer functions for water retention of soils in poland. *Soil Science Society of America*
 1190 *Journal*, 72(5):1243–1247, 2008.

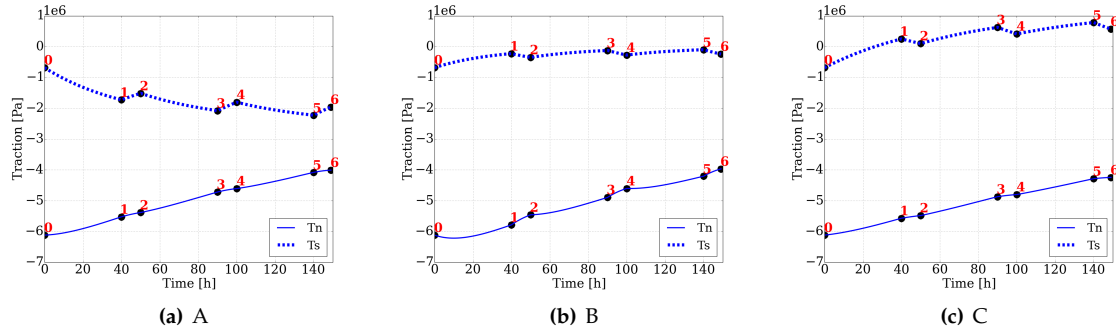


Fig. 37: Time history of normal and tangential displacement tractions at location A, B, C (Fig. 32(a)) of the sealing fault. The coordinate system is $\{M, N\}$ (or $\{x, y\}$) depicted in Fig. 15. The numbers mark the sequence of injection-pause cycles.

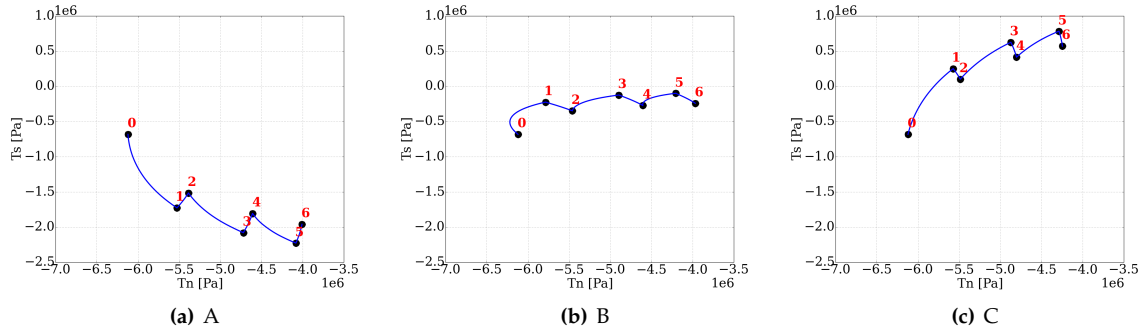


Fig. 38: Traction Path at location A, B, C (Fig. 32(a)) of the sealing fault. The coordinate system is $\{M, N\}$ (or $\{x, y\}$) depicted in Fig. 15. The numbers mark the sequence of injection-pause cycles.

- 1191 M Lefik and BA Schrefler. Artificial neural network as an incremental non-linear constitutive model for a
 1192 finite element code. *Computer methods in applied mechanics and engineering*, 192(28):3265–3283, 2003.
- 1193 M Lefik, DP Boso, and BA Schrefler. Artificial neural networks in numerical modelling of composites.
 1194 *Computer Methods in Applied Mechanics and Engineering*, 198(21):1785–1804, 2009.
- 1195 Jolanta Lewandowska and J-L Auriault. Extension of biot theory to the problem of saturated microporous
 1196 elastic media with isolated cracks or/and vugs. *International Journal for Numerical and Analytical Methods
 1197 in Geomechanics*, 37(16):2611–2628, 2013.
- 1198 Roland W Lewis and Hamid R Ghafouri. A novel finite element double porosity model for multiphase flow
 1199 through deformable fractured porous media. *International journal for numerical and analytical methods in
 1200 geomechanics*, 21(11):789–816, 1997.
- 1201 Paulo J Lisboa and Azzam FG Taktak. The use of artificial neural networks in decision support in cancer:
 1202 a systematic review. *Neural networks*, 19(4):408–415, 2006.
- 1203 Yang Liu, WaiChing Sun, Zifeng Yuan, and Jacob Fish. A nonlocal multiscale discrete-continuum model
 1204 for predicting mechanical behavior of granular materials. *International Journal for Numerical Methods in
 1205 Engineering*, 2015.
- 1206 Yang Liu, WaiChing Sun, and Jacob Fish. Determining material parameters for critical state plasticity
 1207 models based on multilevel extended digital database. *Journal of Applied Mechanics*, 83(1):011003, 2016.
- 1208 Mantas Lukoševičius and Herbert Jaeger. Reservoir computing approaches to recurrent neural network
 1209 training. *Computer Science Review*, 3(3):127–149, 2009.
- 1210 Wes McKinney et al. Data structures for statistical computing in python.

- C Miehe, J Dettmar, and D Zäh. Homogenization and two-scale simulations of granular materials for different microstructural constraints. *International Journal for Numerical Methods in Engineering*, 83(8-9):1206–1236, 2010.
- Christian Miehe. Computational micro-to-macro transitions for discretized micro-structures of heterogeneous materials at finite strains based on the minimization of averaged incremental energy. *Computer Methods in Applied Mechanics and Engineering*, 192(5):559–591, 2003.
- Christian Miehe and CG Bayreuther. On multiscale fe analyses of heterogeneous structures: from homogenization to multigrid solvers. *International Journal for Numerical Methods in Engineering*, 71(10):1135–1180, 2007.
- J Mosler and G Meschke. 3d modelling of strong discontinuities in elastoplastic solids: fixed and rotating localization formulations. *International Journal for Numerical Methods in Engineering*, 57(11):1553–1576, 2003.
- Alejandro Mota, WaiChing Sun, Jakob T. Ostien, James W Foulk, and Kevlin N Long. Lie-group interpolation and variational recovery for internal variables. *Computational Mechanics*, pages 1–19, 2013.
- SeonHong Na and WaiChing Sun. Computational thermo-hydro-mechanics for multiphase freezing and thawing porous media in the finite deformation range. *Computer Methods in Applied Mechanics and Engineering*, 318:667–700, 2017.
- Michał Nitka, Gaël Combe, Cristian Dascalu, and Jacques Desrues. Two-scale modeling of granular materials: a dem-fem approach. *Granular Matter*, 13(3):277–281, 2011.
- Mathieu Nuth and Lyesse Laloui. Advances in modelling hysteretic water retention curve in deformable soils. *Computers and Geotechnics*, 35(6):835–844, 2008.
- Martin Ostoj-Starzewski, X Du, ZF Khisaeva, and W Li. Comparisons of the size of the representative volume element in elastic, plastic, thermoelastic, and permeable random microstructures. *International Journal for Multiscale Computational Engineering*, 5(2), 2007.
- Kyoungsoo Park and Glaucio H Paulino. Cohesive zone models: a critical review of traction-separation relationships across fracture surfaces. *Applied Mechanics Reviews*, 64(6):060802, 2011.
- Roger P Pawlowski, Eric T Phipps, and Andrew G Salinger. Automating embedded analysis capabilities and managing software complexity in multiphysics simulation, part i: Template-based generic programming. *Scientific Programming*, 20(2):197–219, 2012.
- Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12(Oct):2825–2830, 2011.
- SS Peng. Time-dependent aspects of rock behavior as measured by a servocontrolled hydraulic testing machine. In *International Journal of Rock Mechanics and Mining Sciences & Geomechanics Abstracts*, volume 10, pages 235IN21237–236IN22246. Elsevier, 1973.
- Mohammad Piri and Martin J Blunt. Three-dimensional mixed-wet random pore-scale network modeling of two-and three-phase flow in porous media. i. model description. *Physical Review E*, 71(2):026301, 2005.
- Andrew Salinger, Roscoe Bartlett, Andrew Bradley, Qiushi Chen, Irina Demeshko, Xujiao Gao, Glen Hanson, Alejandro Mota, Richard Muller, Erik Nielsen, et al. Albany: Using component-based design to develop a flexible, generic multiphysics analysis code. *International Journal for Multiscale Computational Engineering*, 2016.
- U Schmidt, J Mergheim, and P Steinmann. Identification of elastoplastic microscopic material parameters within a homogenization scheme. *International Journal for Numerical Methods in Engineering*, 2015.
- Hadi Shahir, Ali Pak, Mahdi Taiebat, and Boris Jeremić. Evaluation of variation of permeability in liquefiable soil under earthquake loading. *Computers and Geotechnics*, 40:74–88, 2012.
- WaiChing Sun. A stabilized finite element formulation for monolithic thermo-hydro-mechanical simulations at finite strain. *International Journal for Numerical Methods in Engineering*, 103(11):798–839, 2015.
- WaiChing Sun and Alejandro Mota. A multiscale overlapped coupling formulation for large-deformation strain localization. *Computational Mechanics*, pages 1–18, 2014.
- WaiChing Sun, José E. Andrade, and John W. Rudnicki. Multiscale method for characterization of porous microstructures and their impact on macroscopic effective permeability. *International Journal for Numerical Methods in Engineering*, 88(12):1260–1279, 2011a.
- WaiChing Sun, José E Andrade, John W Rudnicki, and Peter Eichhubl. Connecting microstructural attributes and permeability from 3d tomographic images of in situ shear-enhanced compaction bands

- using multiscale computations. *Geophysical Research Letters*, 38(10), 2011b.
- WaiChing Sun, Qiushi Chen, and Jakob T Ostien. Modeling the hydro-mechanical responses of strip and circular punch loadings on water-saturated collapsible geomaterials. *Acta Geotechnica*, 2013a.
- WaiChing Sun, Matthew R. Kuhn, and John W. Rudnicki. A multiscale dem-lbm analysis on permeability evolutions inside a dilatant shear band. *Acta Geotechnica*, pages 1–16, 2013b.
- WaiChing Sun, Jakob T. Ostien, and Andrew G. Salinger. A stabilized assumed deformation gradient finite element formulation for strongly coupled poromechanical simulations at finite strain. *International Journal for Numerical and Analytical Methods in Geomechanics*, 37(16):2755–2788, 2013c.
- WaiChing Sun, Zhijun Cai, and Jinhyun Choo. Mixed arlequin method for multiscale poromechanics problems. *International Journal for Numerical Methods in Engineering*, 2016.
- Merryn Tawhai, Jeff Bischoff, Daniel Einstein, Ahmet Erdemir, Trent Guess, and Jeff Reinbolt. Multiscale modeling in computational biomechanics. *IEEE Engineering in medicine and biology magazine*, 28(3), 2009.
- Karl Terzaghi. *Theory of consolidation*. Wiley Online Library, 1943.
- Oncel Tuzel, Fatih Porikli, and Peter Meer. Learning on lie groups for invariant detection and tracking. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE, 2008.
- Daniele Versino, Alberto Tonda, and Curt A Bronkhorst. Data driven modeling of plastic deformation. *Computer Methods in Applied Mechanics and Engineering*, 318:981–1004, 2017.
- Kun Wang and WaiChing Sun. Anisotropy of a tensorial bishop coefficient for wetted granular materials. *Journal of Engineering Mechanics*, page B4015004, 2015.
- Kun Wang and Waiching Sun. A semi-implicit micropolar discrete-to-continuum method for granular materials. In M Papadarakakis, V Papadopoulos, G Stefanou, and V Plevris, editors, *Proceedings of European Congress on Computational Methods in Applied Science and Engineering*, number June, pages 5–10, Crete Island, 2016a.
- Kun Wang and WaiChing Sun. A semi-implicit discrete-continuum coupling method for porous media based on the effective stress principle at finite strain. *Computer Methods in Applied Mechanics and Engineering*, 304:546–583, 2016b.
- Kun Wang, WaiChing Sun, Simon Salager, SeonHong Na, and Ghonwa Khaddour. Identifying material parameters for a micro-polar plasticity model via x-ray micro-computed tomographic (ct) images: Lessons learned from the curve-fitting exercises. *International Journal for Multiscale Computational Engineering*, 14(4), 2016.
- Christian Wellmann and Peter Wriggers. A two-scale model of granular materials. *Computer Methods in Applied Mechanics and Engineering*, 205:46–58, 2012.
- Joshua A White and Ronaldo I Borja. Stabilized low-order finite elements for coupled solid-deformation/fluid-diffusion and their application to fault zone transients. *Computer Methods in Applied Mechanics and Engineering*, 197(49):4353–4366, 2008.
- Julien Yvonnet and Q-C He. The reduced model multiscale method (r3m) for the non-linear homogenization of hyperelastic media at finite strains. *Journal of Computational Physics*, 223(1):341–368, 2007.
- Matthew J Zahr, Philip Avery, and Charbel Farhat. A multilevel projection-based model order reduction framework for nonlinear dynamic multiscale problems in structural and solid mechanics. *International Journal for Numerical Methods in Engineering*, 2017.
- XW Zhou, NR Moody, RE Jones, JA Zimmerman, and ED Reedy. Molecular-dynamics-based cohesive zone law for brittle interfacial fracture under mixed loading conditions: Effects of elastic constant mismatch. *Acta Materialia*, 57(16):4671–4686, 2009.
- Jian-Hua Zhu, Musharraf M Zaman, and Scott A Anderson. Modeling of soil behavior with a recurrent neural network. *Canadian Geotechnical Journal*, 35(5):858–872, 1998.
- Wenlu Zhu and Teng-fong Wong. The transition from brittle faulting to cataclastic flow: Permeability evolution. *Journal of Geophysical Research: Solid Earth*, 102(B2):3027–3041, 1997.
- Olgierd C Zienkiewicz, AHC Chan, M Pastor, BA Schrefler, and T Shiomi. *Computational geomechanics*. Wiley Chichester, 1999.
- Robert W Zimmerman, Gang Chen, Teklu Hadgu, and Gudmundur S Bodvarsson. A numerical dual-porosity model with semianalytical treatment of fracture/matrix flow. *Water resources research*, 29(7): 2127–2137, 1993.