

Network scaling and scale-driven loss balancing for intelligent characterization of poroelastic systems

Yang Xu¹, Fatemeh Pourahmadian^{1,2*}

¹ *Department of Civil, Environmental & Architectural Engineering, University of Colorado Boulder, USA*

² *Department of Applied Mathematics, University of Colorado Boulder, USA*

Abstract

A deep learning framework is developed for multiscale characterization of poroelastic media from full waveform data which is known as poroelastography. Special attention is paid to heterogeneous environments whose multiphase properties may drastically change across several scales. Described in space-frequency, the data takes the form of focal solid displacement and pore pressure fields in various neighborhoods furnished either by reconstruction from remote data or direct measurements depending on the application. The objective is to simultaneously recover the six hydromechanical properties germane to Biot equations and their spatial distribution in a robust and efficient manner. Two major challenges impede direct application of existing state-of-the-art techniques for this purpose: (i) the sought-for properties belong to vastly different and potentially uncertain scales, and (ii) the loss function is multi-objective and multi-scale (both in terms of its individual components and the total loss). To help bridge the gap, we propose the idea of *network scaling* where the neural property maps are constructed by unit shape functions composed into a scaling layer. In this model, the unknown network parameters (weights and biases) remain of $O(1)$ during training. This forms the basis for explicit scaling of the loss components and their derivatives with respect to the network parameters. Thereby, we propose the physics-based *dynamic scaling* approach for adaptive loss balancing. The idea is first presented in a generic form for multi-physics and multi-scale PDE systems, and then applied through a set of numerical experiments to poroelastography. The results are presented along with reconstructions by way of gradient normalization (GradNorm) and Softmax adaptive weights (SoftAdapt) for loss balancing. A comparative analysis of the methods and corresponding results is provided. The case of multi-scale reconstructions from noisy data is also numerically investigated.

Keywords: adaptive learning, multi-task learning, poroelastic characterization, full-waveform inversion, multiphysics system identification

1. Introduction

Multiphase processes are at the heart of many critical applications including renewable energy [1, 2], carbon capture [3, 4], and medicine [5, 6]. For instance, engineered injection of fluids in the subsurface, known as the stimulation process, has remarkably enhanced sustainable energy mining from geothermal reservoirs [7, 8] offering an important complement to other renewables with the potential of generating over 100 GW_e of cost-competitive capacity by 2050 [9, 10]. Realizing this potential relies on smart stimulation schemes that require *real-time feedback on the subsurface hydromechanical evolution* [11]. The latter is particularly crucial in enhanced geothermal systems owing to their continuous demand for stimulation and recharge for long-term production [9]. In magnetic resonance elastography [12], a medical imaging technique for in vivo mechanical characterization of soft tissues, it is well known that single-phase elastodynamic models fail to capture the true behavior of live tissues with high water content. This has reportedly led to unstable and unverifiable reconstructions of brain (and other organs) [13, 14, 15]. Recent developments highlight the utility

*Corresponding author: tel. 303-492-2027, email fatemeh.pourahmadian@colorado.edu

and importance of Biot’s poroelastodynamics theory [16, 17, 18] in capturing the fluid-solid interactions in live tissues and is shown to result in more accurate and stable reconstructions [19, 20, 21]. *In vivo hydrodynamical*
15 *characterization* of biological materials could significantly aid differential diagnosis and progress monitoring of neurological disorders such as hydrocephalus and brain tumors [22].

Early poroelastograms referred to time sequences of effective Poisson’s ratio calculated from the evolving ratio of radial to axial strains in poroelastic materials during quasi-static stress relaxation [23, 24]. This information could then be used to quantify the specimen’s permeability. However, this approach involved major
20 simplifying assumptions on the physics of fluid motion and solid deformation in the sample. Recent advances in magnetic resonance motion sensing has enabled a more comprehensive characterization of biphasic materials from volumetric time-harmonic displacement data (mainly in the low-frequency regime) where nonlinear minimization is employed to reconstruct piece by piece the distribution of (a subset of) Biot parameters along with the interstitial fluid pressure fields [13, 22, 19]. The remaining challenges in this vein include tardy reconstructions, and sensitivity of the results to the motion frequency, noise in data, and other (poroelastic) properties
25 of the specimen which are not directly identified but estimated a priori using other methods. To help bridge the gap, this study aims to carefully extend our recent work on ML-based elastography [25, 26] to poroelastodynamics in order to enable simultaneous reconstruction of all Biot parameters from full-field displacement and pore pressure data in a target region of the subsurface. Special attention is paid to challenges related to
30 multi-scale and multi-physics full waveform inversion as well as reconstruction from noisy data. It should be mentioned that this study is particularly related and may contribute to sequential remote sensing [27] where the subsurface is first (geometrically) imaged by way of inverse scattering solutions [28, 29] where subsurface regions of interest e.g., process zones in geo-energy systems or anomalous parts in soft tissues are identified for targeted hydromechanical characterization. Next, local (pore pressure and elastic displacement) fields
35 will be recovered in the identified zones of interest from remote measurements using recent developments in auto-focusing [30]. This may be accomplished by Marchenko-type integral equations where measurements on the surface are deployed to generate focal fields at arbitrary points in a heterogeneous subsurface of unknown properties [31]. Such focal fields yield components of the relevant Green’s function as in [30] (or its recent variants [32, 33]) where the source location coincides with the targeted focus point. On repeating the algorithm
40 for a number of focal points, one may generate sufficient instances of local wavefields in a region of interest for multi-phase characterization. It should be noted that in some applications such as convection MRI [34, 35], direct measurement of full poroelastic waveforms in the subsurface may also be an option. Given the above, the primary objective of this investigation is the piece-wise reconstruction of Biot properties from full poroelastic waveforms in a neighborhood of focal points.

The reconstruction entails multitask learning [36, 37] where the proposed multiscale neural network, modeling poroelastic properties of the subsurface, minimizes the cumulative residual of a system of partial differential equations (PDEs). More specifically, the loss function takes the form $\mathcal{L} = \frac{1}{N} \sum_{i=1}^N \|w_i \ell_i\|^2$ wherein ℓ_i denotes a single-task objective with the associated weight w_i , and the summations is over N PDEs. In multi-phase systems, ℓ_i s describe physical principles that involve spatial and temporal scales straddling multiple orders
50 of magnitude. Thus, careful tuning of the weights is paramount for a robust and successful reconstruction. In deep learning community, loss balancing for multitask learning goes beyond the limits of physics and has been the subject of mounting interest in a wide range of applications. In computer vision, multitask learning is leveraged to improve generalization capability of UberNet [38] and Mask R-CNN [39] through shared representations for instance segmentation and object detection. Traditional methods of loss balancing typically
55 use static and/or manually-tuned weights. For example, sparse autoencoders employ a small, constant weight, known as the sparsity parameter, for the regularization term which is usually found by trial and error that can be suboptimal and labor-intensive. Recent developments, however, have focused on optimizing the learning process by adaptively balancing the loss contributions of different tasks i.e., $w_i = w_i(t)$ where t indicates the training step. Kendall et al. [40] introduced a method that calculates the weights based on the homoscedastic uncertainty in each task. This approach has been shown to outperform static weight baselines and improve
60 overall task performance. Building on this idea, GradNorm [41] was recently developed as a gradient normalization algorithm that balances training by adjusting gradient magnitudes dynamically. GradNorm aims to equalize the training rates of different tasks, ensuring that no single task dominates the learning process. This method has been shown to improve accuracy and reduce overfitting across tasks for various network architectures and datasets. Another approach is the so-called SoftAdapt [42] which uses a family of Softmax-inspired
65

methods to adaptively update the weights based on the live statistics of each subloss. This technique makes use of the initial (or previous) iterations in each epoch to create a preconditioner matrix that normalizes the partial gradients of the loss function in the parameter space. SoftAdapt can be easily integrated into existing architectures, providing a flexible and efficient way to optimize multitask learning. GradNorm was recently adopted by [43] for multiphysics characterization of the subsurface using physics-informed neural networks and reportedly resulted in unstable reconstructions. This inspired a new training logic based on splitting the PDE system into parts dominated by distinct physics/scales. The latter will work in many situations but may pose challenges in highly-heterogeneous systems where the nature of coupling between various physical phenomena may rapidly change in a small neighborhood. In addition, proper decomposition of a PDE system may be particularly complicated in dynamic problems with highly oscillatory data where training PINNs proved to be relatively slow and challenging even in systems governed by single-scale PDEs [25]. In general, in multiscale and multiphysics systems, disparity in gradients exists and some features associated to small time and/or space scales i.e., high frequencies and wave numbers may be less observable and more impacted by noise compared to others. In light of this and given the errors that may be involved in network-estimated derivatives of data that appear in the loss components, especially in initial epochs, it seems that loss balancing based on normalized gradients may lead to amplifying noise or suppressing the signal that may explain the observed instabilities in recent works, and thus, may need to be revisited and adapted for such applications.

In our effort to extend ML-based elastography [25] to poroelastography, we encountered two major impediments: (1) the sought-for material properties, i.e., the outputs of neural network, belong to several different scales owing to their distinct physical nature; the scaling for some of the parameters may be known a priori but for others such as the permeability coefficient or porosity may be uncertain or vary on a set of scales, and (2) various equations in the governing PDE system, i.e., the loss components, describe the balance and coupling between physical phenomena at largely different scales so that the normalized equations remain multiscale, both individually and as a system, such that for instance the leading order in one equation may be $O(1)$, while in the other is $O(10^{-3})$. The first challenge remains true even after standard non-dimensionalization of the governing equations by introducing unit reference scales for relevant physical quantities. For instance, after normalization, the shear modulus in a specimen may be of $O(1)$ while the permeability coefficient is of $O(10^{-5})$. One solution is hardcoding via a change of variable so that all the PDE parameters are re-scaled to $O(1)$. This has been successfully implemented for instance in [43]. A caveat of this approach is that (a) it is not applicable in heterogeneous environments where some of the unknowns such as porosity or permeability could span several scales, and (b) the exact scaling of each unknown quantity should be known a priori.

This work presents a few ideas that could potentially address the above challenges. This paper is organized as the following. Section 2 introduces the proposed network scaling and scale-driven loss balancing in a generic framework for multi-scale PDE systems. Section 3 describes the problem statement for poroelastography and adapts the proposed approach for this application. In addition, the state-of-the-art GradNorm and SoftAdapt techniques for loss balancing are briefly introduced. This section also includes an analysis of the proposed dynamic scaling and its relation to GradNorm and SoftAdapt. Section 4 provides an account of numerical experiments where the reconstruction results from the three approaches are compared and the case of reconstruction from noisy data is investigated. This is then followed by a detailed discussion of the observations and conclusions.

2. Proposed approach

Network scaling embeds a generic change of variable in the architecture of neural networks that map the unknown physical properties of a system in space. In this approach a scaling layer is added right before the output while other layers are normalized i.e., scaled to $O(1)$. For instance, let us consider a multilayer perceptron (MLP) as a property map where the input consists of a grid of neighborhoods in space ξ_i , $i = 1, 2, \dots, N_\xi$, while the output is the associated material properties (or PDE parameters) in each neighborhood $\boldsymbol{\vartheta}_i = (\vartheta_1, \vartheta_2, \dots, \vartheta_{N_\boldsymbol{\vartheta}})_i$ where $N_\boldsymbol{\vartheta}$ is the number of PDE parameters. In this setting, every MLP layer except the last one is defined by the standard map $\Upsilon_\ell: \mathbf{x}^{\ell-1} \rightarrow \mathbf{x}^\ell$,

$$\mathbf{x}^\ell = \Upsilon_\ell(\mathbf{x}^{\ell-1}) := \sigma(\mathbf{W}^\ell \mathbf{x}^{\ell-1} + \mathbf{b}^\ell), \quad \ell = 1, 2, \dots, N_\ell - 1, \quad (1)$$

where \mathbf{W}^ℓ and \mathbf{b}^ℓ respectively designate the ℓ^{th} layer's weight and bias and $\sigma = \tanh, \text{ReLU}$ is the activation function. The network is constructed by coherent composition of Υ_ℓ for $\ell = 1, 2, \dots, N_\ell - 1$, wherein N_ℓ denotes the number of layers. In the last layer, the scaling is applied as the following

$$(\boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}_2, \dots, \boldsymbol{\vartheta}_{N_\xi}) = \text{diag}[(s_1, s_2, \dots, s_{N_\boldsymbol{\vartheta}})_1, \dots, (s_1, s_2, \dots, s_{N_\boldsymbol{\vartheta}})_{N_\xi}] \mathbf{x}^{N_\ell - 1}, \quad (2)$$

wherein $(s_i)_j$ denotes the scale of ϑ_i , $i = 1, \dots, N_\boldsymbol{\vartheta}$, in the j^{th} neighborhood, $j = 1, \dots, N_\xi$, such that $(\vartheta_i)_j = O((s_i)_j)$. In most physical systems, the majority of system's properties vary in space but remain of the same scale i.e., $(s_i)_j = s_i, \forall j, i \in I \subset \{1, \dots, N_\boldsymbol{\vartheta}\}$ except for a small subset of quantities in which case the scaling factors may be identified by a small set \mathbb{S}_i of relevant scales (indicating the likely range of scales that the associated property may assume), thereby the correct scaling factor in each neighborhood will be selected during the learning process such that $(s_i)_j \in \mathbb{S}_i, \forall j, i \in \{1, \dots, N_\boldsymbol{\vartheta}\} \setminus I$. For example, the scaling for permeability in rocks can assume values from the set $\{10^{-5}, 10^{-6}, 10^{-7}\}$ and the exact factors in various neighborhoods will be picked from this set during the optimization process. Note that in the proposed architecture, the network parameters i.e., weights and biases remain of $O(1)$ which aids their fast and robust identification, while the output is properly scaled by the last layer. Note that the scaling factors in this layer are either exactly determined by the nominal values or preliminary experiments, or belong to a small set of scales specified based on the expected range of variation for the associated physical quantities in the domain.

Given the system's response \mathbf{v} , over a set of neighborhoods, the objective is to learn the unknown parameters of the property map $(\{\mathbf{W}^\ell, \mathbf{b}^\ell\}, \ell = 1, 2, \dots, N_\ell - 1, \text{ and } (s_i)_j \in \mathbb{S}_i, \forall j, i \in \{1, \dots, N_\boldsymbol{\vartheta}\} \setminus I)$ by minimizing the residual of the system's governing equations $\mathcal{L} = \frac{1}{N} \sum_{i=1}^N \|w_i \ell_i\|^2$. It should be mentioned that here we consider data inversion in the frequency-domain i.e., $\mathbf{v} = \mathbf{v}(\mathbf{x}, \omega)$ and since multiscale physical quantities are frequency-dependent, we will focus on reconstructions at fixed frequencies. In presence of multi-frequency data, one may conduct the inversion process at each frequency separately and uncover the PDE parameters as a function of frequency (or their evolution in time by an inverse Fourier transform). We assume that each equation ℓ_i in the PDE system can be expressed as a summation of separable operators in terms of data \mathbf{v} and the unknown parameters $\boldsymbol{\vartheta}$ such that

$$\ell_i = \sum_{j=1}^{N_j} f_{ij}(\boldsymbol{\vartheta}, \omega) D_{ij}(\mathbf{v}), \quad D_{ij}(\mathbf{v}) = \sum_{k=1}^{N_k} \frac{\partial^{\alpha_1^k + \alpha_2^k + \alpha_3^k}}{\partial x_1^{\alpha_1^k} \partial x_2^{\alpha_2^k} \partial x_3^{\alpha_3^k}} h_{ij}^k[\mathbf{v}](\mathbf{x}), \quad \boldsymbol{\alpha}^k = (\alpha_1^k, \alpha_2^k, \alpha_3^k), \quad (3)$$

wherein $f_{ij} = f_{ij}(\boldsymbol{\vartheta}, \omega)$ and $h_{ij} = h_{ij}^k(\mathbf{v})$ for $i = 1, \dots, N$, $j = 1, \dots, N_j$ and $k = 1, \dots, N_k$, are linear or nonlinear functions of their arguments; D_{ij} denotes a differential operator, and the multi-index notation $\boldsymbol{\alpha}^k \in \mathbb{N}^3$ is used for partial differentiation with respect to $\mathbf{x} = (x_1, x_2, x_3)$ [44].

In this setting, for *loss balancing*, by assuming that f_{ij} is a Lipschitz function of network parameters (see Remark 1), we propose to scale each loss component ℓ_i by a dynamic weight w_i such that the average scale in the normalized equation $w_i \ell_i$ is $O(1)$. For this purpose, at every epoch, the scale of every loss term in (3) is determined as the following

$$e_{ij} = \text{round}(\log_{10}[f_{ij}(\boldsymbol{\vartheta}, \omega)]) + \text{round}(\log_{10}[|\langle D_{ij}(\mathbf{v}) \rangle_\xi|]), \quad i = 1, \dots, N, j = 1, \dots, N_j, \quad (4)$$

where $\text{round}()$ maps its argument to the nearest integer and $\langle |\cdot| \rangle_\xi$ denotes the mean of absolute value of its argument over the support ξ . Note that the scale of $f_{ij}(\boldsymbol{\vartheta}, \omega)$ is *in part* dictated by the last layer in the neural property map and can also be determined by the network's estimates for the PDE parameters $\boldsymbol{\vartheta}$. Here, ω is the frequency of probing waves and a known (input) parameter. Moreover, the scale of $D_{ij}(\mathbf{v})$ is directly computed from data using spectral methods for signal processing and differentiation [25, 26]. Given the above, the average scale and affiliated weight for each loss component is specified as follows,

$$w_i = 10^{-\beta_i}, \quad \beta_i = \frac{1}{N_j} \sum_{j=1}^{N_j} e_{ij}, \quad i = 1, \dots, N. \quad (5)$$

Here we assumed that 10 is an appropriate basis for separating the scales. One may replace 10 by 10^7 and

Algorithm 1 Pseudocode for the method of *dynamic scaling* for loss balancing.

Require: (1) network-predicted scaling for unknown physical parameters $\boldsymbol{\vartheta}$, (2) loss components ℓ_i recast according to (3), and (3) relevant field derivatives $D_{ij}(\mathbf{v})$ that appear in the loss

for $t = 0$ **to** epoch **do**

if f_{ij} is Lipchitz continuous $\forall i, j$ with respect to the network parameters **then**

 Compute $\text{round}(\log_{10}[f_{ij}(\boldsymbol{\vartheta}, \omega)])$ for $i = 1, \dots, N$ and $j = 1, \dots, N_j$ based on the network outputs

 Compute $\forall i, j$, $\text{round}(\log_{10}[\langle |D_{ij}(\mathbf{v})| \rangle_{\xi}])$ according to the loss

 Compute the average scale of each loss component

$$\beta_i = \frac{1}{N_j} \sum_{j=1}^{N_j} \left(\text{round}(\log_{10}[f_{ij}(\boldsymbol{\vartheta}, \omega)]) + \text{round}(\log_{10}[\langle |D_{ij}(\mathbf{v})| \rangle_{\xi}]) \right)$$

else

 Compute $\text{round}(\log_{10}[\langle |D_{ij}(\mathbf{v})| \rangle_{\xi}])$ for $i = 1, \dots, N$ and $j = 1, \dots, N_j$ according to the loss

 Compute $\forall i, j, n$, $\text{round} \left(\log_{10} \left[\frac{\partial f_{ij}}{\partial \vartheta_n}(\vartheta_n, \omega) \right] \right)$, $n = 1, \dots, N_{\boldsymbol{\vartheta}}$, based on the network outputs

 Compute $\forall n$, $\text{round}(\log_{10}[\vartheta_n])$, based on the network outputs

 Compute the average scale of each loss component's derivative with respect to a network parameter

$$\beta_i = \frac{1}{N_j} \sum_{j=1}^{N_j} \left(\text{round}(\log_{10}[\langle |D_{ij}(\mathbf{v})| \rangle_{\xi}]) + \frac{1}{N_{\boldsymbol{\vartheta}}} \sum_{n=1}^{N_{\boldsymbol{\vartheta}}} \left(\text{round} \left(\log_{10} \left[\frac{\partial f_{ij}}{\partial \vartheta_n}(\vartheta_n, \omega) \right] \right) + \text{round}(\log_{10}[\vartheta_n]) \right) \right)$$

end if

 Update weights $w_i = 10^{-\beta_i}$

 Update weighted loss $\mathcal{L} = \frac{1}{N} \sum_{i=1}^N \|w_i \ell_i\|^2$

 Optimize \mathcal{L}

end for

gauge the basis as needed in their application of interest. The pseudocode for the proposed approach is provided in Algorithm 1.

Remark 1. Let w be a network parameter, observe that if f_{ij} is Lipchitz continuous with respect to w such that the Lipchitz constant is of $\mathcal{O}(f_{ij})$, thanks to the proposed architecture for property maps and their scaling, then the proposed weights w_i in (5) automatically normalize the gradient of weighted loss components with respect to the network parameter w . This will be demonstrated in Section 3 within the context of poroelastography. In other applications where f_{ij} is not Lipchitz continuous, one may alternatively use the following weights for gradient normalization based on scaling

$$w'_i = 10^{-\beta'_i}, \quad \beta'_i = \frac{1}{N_j} \sum_{j=1}^{N_j} \left(\text{round}(\log_{10}[\langle |D_{ij}(\mathbf{v})| \rangle_{\xi}]) + \frac{1}{N_{\boldsymbol{\vartheta}}} \sum_{n=1}^{N_{\boldsymbol{\vartheta}}} \left(\text{round} \left(\log_{10} \left[\frac{\partial f_{ij}}{\partial \vartheta_n}(\vartheta_n, \omega) \right] \right) + \text{round}(\log_{10}[\vartheta_n]) \right) \right), \quad (6)$$

125 where the scale of ϑ_n is governed by the last layer in the property network. Note that given the architecture of the latter, $\frac{\partial \vartheta_n}{\partial w} = \mathcal{O}(\vartheta_n)$ for $n = 1, \dots, N_{\boldsymbol{\vartheta}}$.

3. Model problem

This section implements the proposed approach for poroelastic characterization of rocks using elastic wave-
form and acoustic pore pressure data. In what follows, we begin with the problem statement and highlight its
multiscale nature that introduced many challenges in our early attempts for data inversion using the available
130 techniques. This is followed by proper normalization and construction of the proposed neural architecture
with network scaling. For loss balancing, we provide a comparative analysis between the proposed dynamic
scaling and the current state-of-art methods, namely: the GradNorm and SoftAdapt, for multi-task learning.

3.1. Problem Statement

Consider a poroelastic domain \mathcal{B} characterized by the drained Lamé parameters λ and μ , Biot modulus M ,
total density ρ , fluid density ρ_f , apparent mass density ρ_a , permeability coefficient κ , porosity ϕ , and the Biot
effective stress coefficient α . With reference to Fig. 1 (a), it is assumed that the plane-strain approximation
holds and that the system's response to a fluid body force at $\mathbf{x}_o = (x_o, y_o)$ is reconstructed from far-field data (or
directly measured using for example convection MRI [34, 35]) in the neighborhood \mathcal{B}_o of the source location
in terms of the solid displacement $\mathbf{u} = \mathbf{u}(\mathbf{x}, \omega)$ and pore pressure $p = p(\mathbf{x}, \omega)$. On recalling that ω denotes the
frequency of wave motion, Biot equations [18, 45, 28] may be used to describe $[\mathbf{u}, p](\mathbf{x}, \omega)$ for $\mathbf{x} \in \mathcal{B}_o$ as

$$\begin{aligned}\Pi_1[\mathbf{u}, p; \delta, \omega] &:= \nabla \cdot (\mathbf{C} : \nabla \mathbf{u}) - a \nabla p + \omega^2 b \mathbf{u} - \mathbf{f}^u(\delta) = \mathbf{0}, \\ \Pi_2[\mathbf{u}, p; \delta, \omega] &:= \frac{c}{\omega^2} \nabla^2 p + M^{-1} p + a \nabla \cdot \mathbf{u} + \frac{c}{\omega^2} f^p(\delta) = 0,\end{aligned}\tag{7}$$

where $\mathbf{C} = \lambda \mathbf{I}_2 \otimes \mathbf{I}_2 + 2\mu \mathbf{I}_4$ is the fourth-order drained elasticity tensor with $\mathbf{I}_m (m = 2, 4)$ representing m th-order
symmetric identity tensor, and

$$\begin{aligned}\gamma &= \frac{\rho_a}{\phi^2} + \frac{\rho_f}{\phi} + \frac{i}{\omega \kappa}, \quad a = \alpha - \frac{\rho_f}{\gamma}, \quad b = \rho - \frac{\rho_f^2}{\gamma}, \quad c = \frac{1}{\gamma}, \\ \mathbf{f}^u &= \left[-\frac{\rho_f}{\gamma} \delta(\mathbf{x} - \mathbf{x}_o), 0 \right], \quad f^p = \nabla \cdot \delta(\mathbf{x} - \mathbf{x}_o), \quad \delta(\mathbf{x} - \mathbf{x}_o) = D e^{-\zeta \|\mathbf{x} - \mathbf{x}_o\|^2},\end{aligned}\tag{8}$$

135 wherein $\delta(\mathbf{x} - \mathbf{x}_o)$ is a time-harmonic fluid body source applied at \mathbf{x}_o characterized by its magnitude D and
spatial decay rate ζ . Given the response $[\mathbf{u}, p](\mathbf{x}, \omega)$ at a fixed frequency in the neighborhood \mathcal{B}_o , the objective
is to reconstruct the domain's hydromechanical properties i.e., $\{\mu_o, \lambda_o, M_o, \alpha_o, \phi_o, \kappa_o\}$ in the vicinity of source

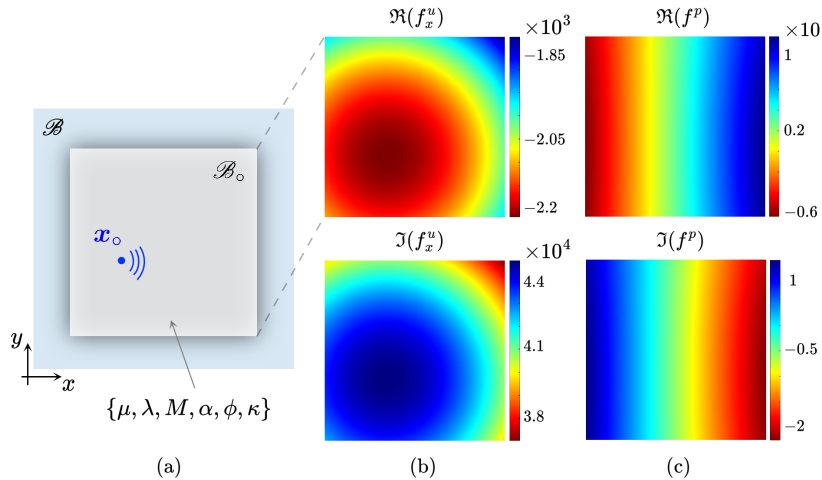


Figure 1: Synthetic experiments simulating wave motion in the poroelastic domain \mathcal{B} : (a) the model is harmonically excited at
frequency ω by a fluid body force $\delta(\mathbf{x} - \mathbf{x}_o)$ and the response is computed in a neighborhood of the source point \mathbf{x}_o i.e., in the
square \mathcal{B}_o of side $5\ell_r$, (b) the x component of effective body force f^u in the first of Biot equations in (7), and (c) the fluid source
term f^p in the generalized Darcy's law i.e., the second of (7). Here, \Re and \Im respectively indicate the real and imaginary parts
of a complex-valued quantity.

point \mathbf{x}_o . It should be mentioned that the source terms f^u and f^p in (7) play a key role in breaking the intrinsic symmetry in Biot equations and thus enable concurrent identification of all poroelastic parameters by eliminating the null space in the governing PDE system for every $\mathbf{x} \in \mathcal{B}_o$. Moreover, these terms locally enhance the observability of hydraulic transport in \mathcal{B}_o and thus improve the robustness when retrieving the permeability coefficient. The main challenge in practice pertains to characterization of tight or low-permeability formations. As such, here we assume that \mathcal{B} is comprised of Pecos sandstone according to [45, 46]. It should be mentioned that in our numerical experiments we assumed that the poroelastic properties are homogenized or constant in the vicinity of each source i.e., focal point. In what follows, the reconstructions are performed in two focal areas whose affiliated properties are listed in Table 1.

3.2. Dimensional platform

To facilitate data processing, we normalize the system by choosing $\rho_r = 10^3$ kg/m³, $\ell_r = 0.14$ m, and $\mu_r = 5.85$ GPA as the reference scales respectively for mass density, length, and stress. Note that here ρ_r is identified by the fluid density, ℓ_r is the drained shear wavelength, and μ_r is the drained shear modulus. In this setting, all quantities are non-dimensionalized as reported in Table 1. Given the multiphase nature of the domain, the normal system retains its multiscale character and further processing is required to unify the scale of all variables. This will be accomplished by the proposed network scaling in this study.

Table 1: Poroelastic properties of the domain \mathcal{B} in the vicinity of two focal points.

property	value in focal area $i = 1,2$	dimensionless values
first Lamé parameter (drained)	$\lambda'_1 = \lambda'_2 = 2.74$ GPA	$\lambda_1 = \lambda_2 = 0.47$
drained shear modulus	$\mu'_1 = \mu'_2 = 5.85$ GPA	$\mu_1 = \mu_2 = 1$
Biot modulus	$M'_1 = M'_2 = 9.71$ GPA	$M_1 = M_2 = 1.66$
total density	$\rho'_1 = \rho'_2 = 2270$ kg/m ³	$\rho_1 = \rho_2 = 2.27$
fluid density	$\rho'_{f_1} = \rho'_{f_2} = 1000$ kg/m ³	$\rho_{f_1} = \rho_{f_2} = 1$
apparent mass density	$\rho'_{a_1} = \rho'_{a_2} = 117$ kg/m ³	$\rho_{a_1} = \rho_{a_2} = 0.117$
permeability coefficient	$\kappa'_1 = 503, \kappa'_2 = 0.8$ mm ⁴ /N	$\kappa_1 = 1.5407 \times 10^{-5}, \kappa_2 = 2.45 \times 10^{-8}$
porosity	$\phi_1 = \phi_2 = 0.195$	$\phi_1 = \phi_2 = 0.195$
Biot effective stress coefficient	$\alpha_1 = \alpha_2 = 0.83$	$\alpha_1 = \alpha_2 = 0.83$
excitation frequency	$\omega' = 1.2$ MHz	$\omega = 391$
source amplitude	$D' = 8.36 \times 10^4$ m	$D = 5.97 \times 10^5$
source decay rate	$\varsigma' = 9.57 \times 10^4$ m ⁻²	$\varsigma = 187.52$

3.3. Data inversion

In every focal area, given the frequency ω and the fluid body source $\delta(\mathbf{x} - \mathbf{x}_o)$ – and thus, $f^u(\delta)$ and $f^p(\delta)$ as in Fig. 1 – the hydromechanical response is obtained in terms of the two-dimensional solid displacement $\mathbf{u} = (u_x, u_y)$ and pore pressure p as illustrated in Fig. 2. In this section, given the dataset $[\mathbf{u}_i^j, p_i^j; \delta_i, \omega]$ in two focal regions $i = 1,2$ with $j = 1, \dots, N_{p_i}$ data points in each neighborhood, the objective is to recover the affiliated hydromechanical properties $\{\mu_i, \lambda_i, M_i, \alpha_i, \phi_i, \kappa_i\}$ in each region. For this purpose, the real and imaginary parts of the Biot equations as well as every complex-valued quantity are separated such that (7) is transformed into a real-valued system of six PDEs. The real and imaginary parts of each quantity are denoted by $\Re(\cdot)$ and $\Im(\cdot)$ respectively. The six PDE parameters in each neighborhood are mapped by a scaled MLP that is identified by minimizing the weighted residuals of the Biot system. In what follows, the property map is constructed and trained using three adaptive methods for balancing the six components of the loss function, namely: the proposed dynamic scaling, GradNorm and SoftAdapt. Each method is briefly introduced followed by a comparative analysis of their logic and performance.

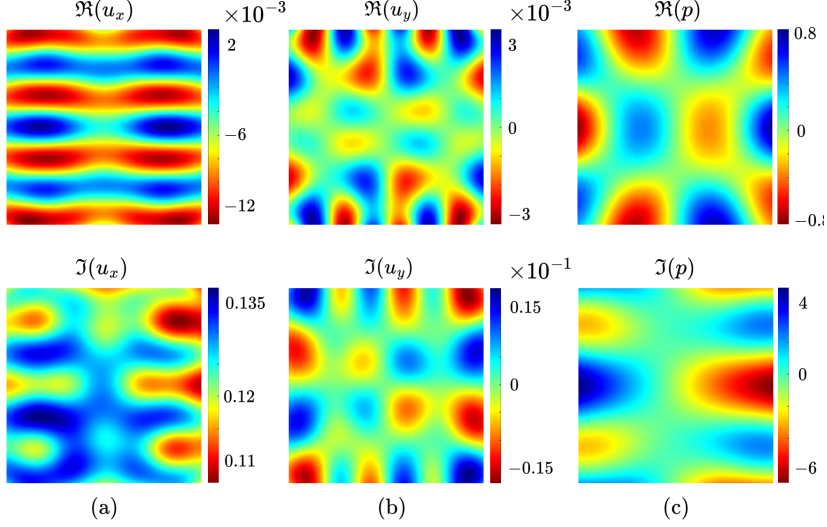


Figure 2: Simulated poroelastic response to the excitation shown in Fig. 1 in focal area ξ_i ; real (top row) and imaginary (bottom row) of (a) solid displacement u_x , (b) solid displacement u_y , and (c) interstitial pore pressure p .

Scaled neural networks as property maps

In this work, the unknown hydromechanical properties in the designated focal areas are mapped by a multilayer perceptron whose dense range allows for capturing arbitrary complex functions [47]. As depicted in Fig. 3, the input consists of a set of neighborhoods $\{\xi_i\}$, $i = 1, \dots, N_\xi$, here $N_\xi = 2$, while the output denotes the associated poroelastic properties. Every MLP layer (before the last one) is defined by (1), while the last scaling layer is defined by

$$(\boldsymbol{\vartheta}_1^*, \boldsymbol{\vartheta}_2^*) = (\{\mu_1^*, \lambda_1^*, M_1^*, \alpha_1^*, \phi_1^*, \kappa_1^*\}, \{\mu_2^*, \lambda_2^*, M_2^*, \alpha_2^*, \phi_2^*, \kappa_2^*\}) = \text{diag}[1, 1, 1, 1, 0.1, s_k^1, 1, 1, 1, 1, 0.1, s_k^2] \mathbf{x}^{N_\ell - 1}, \quad s_k^i \in \{10^{-5}, 10^{-6}, 10^{-7}, 10^{-8}\}, \quad i = 1, 2, \quad (9)$$

in this application based on the nominal (normalized) values for sandstone. The network is trained by minimizing the loss function \mathcal{L} as follows

$$\mathcal{L} = \sum_{i=1}^2 \sum_{j=1}^{N_{p_i}} \sum_{k=1}^6 \left\| w_k^i \ell_k(\mathbf{u}_i^j, p_i^j; \delta_i, \omega | \boldsymbol{\vartheta}_i^*, \{D_{kl}\}_{l=1, \dots, N_k}) \right\|^2, \quad (10)$$

where $\boldsymbol{\vartheta}_i^* = (\mu_i^*, \lambda_i^*, M_i^*, \alpha_i^*, \phi_i^*, \kappa_i^*)$ is the set of network-predicted properties in each focal region ξ_i ; $D_{kl} \in \{\partial/\partial x, \partial/\partial y, \partial^2/\partial x^2, \partial^2/\partial y^2, \partial^2/\partial x \partial y\}$ is the set of differential operators according to (3), and each loss component is

$$\{\ell_1, \ell_2, \ell_3, \ell_4, \ell_5, \ell_6\} = \{\Re(\Pi_1)_x, \Im(\Pi_1)_x, \Re(\Pi_1)_y, \Im(\Pi_1)_y, \Re(\Pi_2), \Im(\Pi_2)\}, \quad (11)$$

wherein Π_1 and Π_2 are defined in (7). Thereby, ℓ_1 can be written as

$$\begin{aligned} \ell_1 = & \mu \left(\frac{\partial^2 \Re(u_x)}{\partial x^2} + \frac{\partial^2 \Re(u_x)}{\partial y^2} + \frac{\partial^2 \Re(u_x)}{\partial x^2} + \frac{\partial^2 \Re(u_y)}{\partial x \partial y} \right) + \lambda \left(\frac{\partial^2 \Re(u_x)}{\partial x^2} + \frac{\partial^2 \Re(u_y)}{\partial x \partial y} \right) - \\ & - \Re(a) \frac{\partial \Re(p)}{\partial x} + \Im(a) \frac{\partial \Im(p)}{\partial x} + \omega^2 \Re(b) \Re(u_x) - \omega^2 \Im(b) \Im(u_x) - \Re(f_x^u), \end{aligned} \quad (12)$$

170 and $\ell_2 - \ell_6$ are provided in Appendix A for completeness.

Adaptive weights for loss balancing

In this section, we briefly describe implementation of the proposed Dynamic Scaling for adaptive loss balancing, and provide an overview of the SoftAdapt and GradNorm methods. This is followed by a comparative analysis of these approaches and reporting of the obtained results in Section 4.

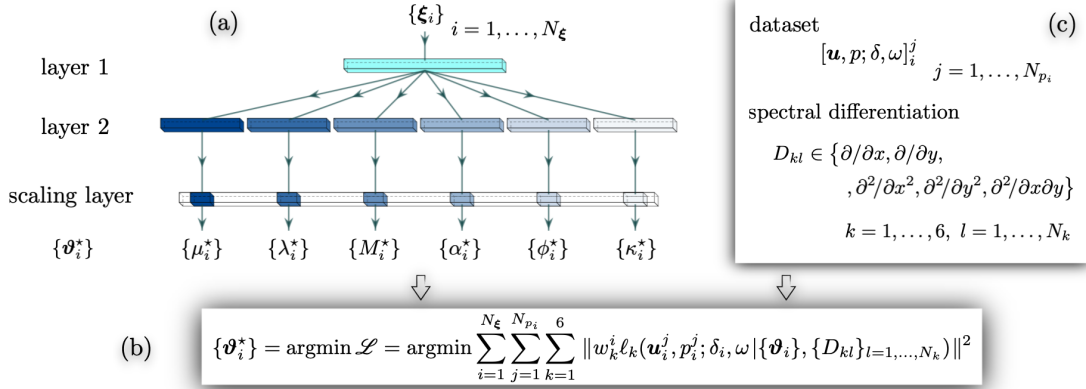


Figure 3: Application of the proposed approach for intelligent poroelastography: (a) mapping of the unknown hydromechanical properties in each focal neighborhood ξ_i , $i = 1, 2$, by a multiscale MLP such that the network parameters (i.e., weights and biases) remain of $O(1)$ while the outputs are properly scaled by the last layer, (b) identifying the network by minimizing the multi-objective loss function \mathcal{L} comprised of weighted PDE residuals associated with the Biot equations, and (c) the training dataset entailing the harmonic source function δ_i of frequency ω in each focal region and the associated displacement and pressure fields $[\mathbf{u}_i^j, \mathbf{p}_i^j]$ at $j = 1, \dots, N_{p_i}$ points in the vicinity of each source along with their spatial derivatives D_{kl} .

175 *Dynamic Scaling (DynScl)*

According to (3)-(5), this approach weights each task ℓ_k , $k = 1, \dots, 6$, such that the affiliated residue is normalized by the average scale in ℓ_k . For example, the weight w_1^i associated with ℓ_1 in (12) in the i^{th} neighborhood is specified as follows,

$$w_1^i = 10^{-\beta^i}, \quad \beta_1^i = \frac{1}{7} \sum_{l=1}^7 \operatorname{round}(\log_{10}[f_{1l}(\theta_i^*, \omega)]) + \operatorname{round}(\log_{10}[\langle |D_{1l}[u_x, u_y, p; \mathbf{f}^u, \mathbf{f}^p](\mathbf{x}_i^j)| \rangle_j]), \quad i = 1, 2. \quad (13)$$

Here, $\langle |\cdot| \rangle_j$ indicates the mean of absolute value of its argument over $j = 1, \dots, N_{p_i}$ points in every focal region ξ_i , and \mathbf{x}_i^j denotes the position vector affiliated with data point $(\mathbf{u}_i^j, \mathbf{p}_i^j)$ where $i = 1, 2$. Note that all derivatives $D_{1l}[u_x, u_y, p; \mathbf{f}^u, \mathbf{f}^p]$, $l = 1, \dots, 7$, are computed by way of spectral differentiation in each neighborhood from the set $\{(\mathbf{u}_i^j, \mathbf{p}_i^j, \delta_i, \omega)\}_{j=1, \dots, N_{p_i}}$. For completeness, the explicit form of operators f_{1l} and D_{1l} , $l = 1, \dots, 7$, according to (3) is provided in the following.

$$\{f_{11}, f_{12}, f_{13}, f_{14}, f_{15}, f_{16}, f_{17}\}(\theta_i^*, \omega) = \{\mu_i^*, \lambda_i^*, -\Re(a_i^*), \Im(a_i^*), \omega^2 \Re(b_i^*), -\omega^2 \Im(b_i^*), 1\}, \quad (14)$$

$$a_i^* = a(\alpha_i^*, \phi_i^*, \kappa_i^*), \quad b_i^* = b(\phi_i^*, \kappa_i^*), \quad i = 1, 2,$$

$$\{D_{11}, D_{12}, D_{13}, D_{14}, D_{15}, D_{16}, D_{17}\}[u_x, u_y, p; \mathbf{f}^u, \mathbf{f}^p] =$$

$$\left\{ \left(\frac{\partial^2 \Re(u_x)}{\partial x^2} + \frac{\partial^2 \Re(u_x)}{\partial y^2} + \frac{\partial^2 \Re(u_x)}{\partial x^2} + \frac{\partial^2 \Re(u_y)}{\partial x \partial y} \right), \left(\frac{\partial^2 \Re(u_x)}{\partial x^2} + \frac{\partial^2 \Re(u_y)}{\partial x \partial y} \right), \frac{\partial \Re(p)}{\partial x}, \frac{\partial \Im(p)}{\partial x}, \Re(u_x), \Im(u_x), -\Re(\mathbf{f}_x^u) \right\}. \quad (15)$$

Similar factorizations can be formulated for $\ell_2 - \ell_6$ in (A.1) to find $w_2^i - w_6^i$ in (10) for every focal region ξ_i , $i = 1, 2$.

Softmax adaptive weights (SoftAdapt)

In this approach, the weight of each loss component is calculated based on its live performance statistics. More specifically, the rate of change of each objective function relative to others is used to gauge its visibility to the minimizer such that the convergence of all loss components in the parameter space is approximately isotropic [42]. To this end, recall that the sought-for poroelastic properties θ_i^* in each focal area ξ_i , $i = 1, 2$, is a function of network parameters i.e., weights and biases $\{\mathbf{W}^m, \mathbf{b}^m\}_{m=1, 2, \dots, N_m}$ in N_m layers, which in turn are

a function of iteration steps t_n . Then, define the rate of change of the loss component $\ell_k = \ell_k(\boldsymbol{\vartheta}_i^*(t_n))$ by

$$s_k^i = \ell_k(\boldsymbol{\vartheta}_i^*(t_n)) - \ell_k(\boldsymbol{\vartheta}_i^*(t_{n-1})), \quad k = 1, \dots, 6, \quad i = 1, 2, \quad n = 1, \dots, N_{\text{epoch}}, \quad (16)$$

wherein N_{epoch} is the number of epochs. In this setting, the loss weights are defined by

$$w_k^i = \frac{e^{\eta(s_k^i - \max(\{s_k^i\}_{k=1, \dots, 6}))}}{\sum_{k=1}^6 e^{\eta(s_k^i - \max(\{s_k^i\}_{k=1, \dots, 6}))}}, \quad i = 1, 2, \quad (17)$$

where η is a tunable hyper-parameter with default value of $\eta = 0.1$. SoftAdapt assigns larger weights to loss components with slower convergence rate.

Gradient normalization (GradNorm)

GradNorm balances the *gradient of weighted loss components* with respect to network parameters to ensure that all objectives train at similar rates [41]. This method addresses the common issue of gradient imbalances by penalizing tasks with excessively large or small gradients. Let $\mathbf{w} \subset \{\mathbf{W}^m, \mathbf{b}^m\}_{m=1, 2, \dots, N_m}$ denote the parameters of shared MLP layer(s) in the property map, see Fig. 3. GradNorm finds the optimal weights $w_k^i(t_n)$, at every epoch t_n , by minimizing the L_1 difference between the actual and average norms of loss gradients with respect to \mathbf{w} as the following

$$\mathcal{L}_V^i = \sum_{k=1}^6 \left| \left\| \nabla_{\mathbf{w}} w_k^i(t_n) \ell_k(\boldsymbol{\vartheta}_i^*(t_n)) \right\|_2 - \frac{1}{6} \sum_{k=1}^6 (r_k(t_n))^{\tilde{\eta}} \left\| \nabla_{\mathbf{w}} w_k^i(t_n) \ell_k(\boldsymbol{\vartheta}_i^*(t_n)) \right\|_2 \right|, \quad r_k(t_n) = \frac{\ell_k(t_n)/\ell_k(t_1)}{\sum_{k=1}^6 \ell_k(t_n)/\ell_k(t_1)}, \quad (18)$$

where $\tilde{\eta}$ is a tunable hyperparameter that determines the strength of the restoring force that aligns the loss components to a common convergence rate. When loss components differ greatly in scale, causing significant variations in the loss distribution during training, a higher value of $\tilde{\eta}$ is needed to enforce a balance. In contrast, when the scale of loss component are more similar, a lower value of $\tilde{\eta}$ is suitable. Taking $\tilde{\eta} = 0$ will assign equal weights to all components. Note that minimizing \mathcal{L}_V^i is repeated at every epoch while optimizing \mathcal{L} . Since w_k^i directly influences the gradient magnitudes for each loss component, \mathcal{L}_V^i is differentiated with respect to w_k^i , and the resulting derivatives are used to update each weight w_k^i via standard gradient decent approaches.

Comparative analysis

The proposed approach based on dynamic scaling of the loss components is closely related to GradNorm. To observe this, recall that the weights and biases remain of $O(1)$ at every epoch in the property map and thus

$$\frac{\partial(\boldsymbol{\vartheta}_i^*)_m}{\partial w_n} = O((s_m)_i), \quad m = 1, \dots, 6, \quad n = 1, \dots, N_{\mathbf{w}}, \quad i = 1, 2, \quad (\boldsymbol{\vartheta}_i^*)_m \in \boldsymbol{\vartheta}_i^*, \quad w_n \in \mathbf{w}, \quad (19)$$

wherein $(s_m)_i$ is the relevant scale in the MLP's last layer. Whereby, it is straightforward from (13)-(15) to prove that

$$\begin{aligned} \frac{\partial w_1^i(t_n) \ell_1(\boldsymbol{\vartheta}_i^*(t_n))}{\partial w_n} &= \sum_{l=1}^7 w_1^i(t_n) D_{1l}[u_x, u_y, p; f^u, f^p] \frac{\partial f_{1l}(\boldsymbol{\vartheta}_i^*, \omega)}{\partial w_n} = \\ &= \sum_{l=1}^7 w_1^i(t_n) D_{1l}[u_x, u_y, p; f^u, f^p] \nabla_{\boldsymbol{\vartheta}_i^*} f_{1l}(\boldsymbol{\vartheta}_i^*, \omega) \frac{\partial \boldsymbol{\vartheta}_i^*}{\partial w_n} = O(1), \quad n = 1, \dots, N_{\mathbf{w}}, \quad i = 1, 2. \end{aligned} \quad (20)$$

This remains the case for all other weighted components of the loss $w_k^i(t_n) \ell_k(\boldsymbol{\vartheta}_i^*(t_n))$, for $k = 2, \dots, 6$, as one may prove that $f_{kl}(\boldsymbol{\vartheta}_i^*, \omega)$, corresponding to the factorization in (3) for the Biot system, are Lipchitz continuous with respect to $w_n \in \mathbf{w}$ such that the Lipchitz constant is of $O(f_{kl})$. In other words,

$$\frac{\partial f_{kl}}{\partial w_n} = O(f_{kl}), \quad k = 2, \dots, 6, \quad n = 1, \dots, N_{\mathbf{w}}, \quad i = 1, 2. \quad (21)$$

Therefore, the Dynamic Scaling approach automatically achieves the GradNorm’s objective in (18) without requiring a separate minimization procedure per epoch. This is accomplished by *approximating*, and thereby normalizing, the scale of relevant derivatives instead of their explicit calculation which may involve significant computational cost and complications during the optimization process. This solution is afforded by the proposed architecture of the property network and its explicit scaling via the last layer. Note that in many physical systems given the frequency of wave motion and network-predicted physical properties, the scale of spatial derivatives D_{kl} in the PDE system can be estimated without explicit differentiation. This will be particularly helpful for verification and validation of the computed derivatives, especially in presence of noise in data, and in certain scenarios can help expedite the loss balancing process.

Softmax adaptive weights carry the advantage of being computationally efficient and flexible so that they can be applied to a wide range of network architectures and multi-objective loss functions. There are two caveats, however, when it comes to using SoftAdapt for balancing multi-physics PDE systems: (i) the heuristic nature of the Softmax function that may not cater, in terms of its range, for systems that span several scales in time-space, and (ii) this approach uniformly assigns larger weights to loss components with lower rate of change. This is typically done without differentiating whether, for instance, the computed rate of change is large due to scaling or due to convergence. In physical systems, one objective may assume much larger values compared to others and shows a significant decay rate without actually converging, while other loss components could be fast converging but due to their smaller scale show much smaller decay rates. This could lead to training instability, especially in presence of noise.

4. Implementation results

4.1. Reconstruction from noiseless data

In this section, we report the reconstruction results related to two focal regions introduced in Section 3. Here, the proposed network scaling is coupled with three different loss balancing techniques, described in Section 3, namely: the proposed dynamic scaling (DS), GradNorm (GN), and SoftAdapt (SA). We compare the performance of these methods based on accuracy and robustness of the associated results. In all cases, the property maps are constructed by MLPs comprised of two parts; the first part is a single fully-connected tanh-activated layer with thirty two neurons as a common trunk, while the second part involves six individual towers, each of which composed of sixteen neurons whose output is separately scaled by the last layer to predict a designated material parameter. This architecture is consistent with the schematic shown in Fig. 3. All models are trained by the Adam optimizer, and the learning rate is tuned during the training process. We also compare the results to those obtained without network scaling and/or loss balancing. The poroelastodynamic simulations, germane to the configuration shown in Fig. 1 (a), are conducted using the FreeFem++ [48] platform to generate the training data. In this vein, we built on an existing code that was recently developed as part of [28] to model poroelastic wave motion in the subsurface. In what follows, the reported reconstruction error is normalized according to

$$\Xi((\vartheta_i^\star)_n) = \frac{|(\vartheta_i^\star)_n - (\vartheta_i)_n|}{|(\vartheta_i)_n|}, \quad i = 1, 2, \quad n = 1, \dots, N_{\vartheta}, \quad \star = \{\text{EW, DS, GN, SA}\}, \quad (22)$$

where $(\vartheta_i^\star)_n$ is the neural network prediction for a quantity whose true value is $(\vartheta_i)_n$. Note that $(\vartheta_i)_n$ indicates the n^{th} component of the property vector ϑ_i in the i^{th} focal region when $i = 1, 2$ in this study. It should be mentioned that in this section for clarity the superscript \star which indicates network prediction in Section 3 is replaced by an abbreviation of the method used for training.

The reconstruction results using the proposed dynamic scaling approach is shown in Figs. 4 and 5. Note that here all data from both low-permeability and high-permeability neighborhoods $\{\xi_i\}$, $i = 1, 2$, is used for training to simultaneously reconstruct the six unknown poroelastic parameters in each neighborhood leading to twelve network outputs $(\vartheta_1^{\text{DS}}, \vartheta_2^{\text{DS}}) = (\{\mu_1^{\text{DS}}, \lambda_1^{\text{DS}}, M_1^{\text{DS}}, \alpha_1^{\text{DS}}, \phi_1^{\text{DS}}, \kappa_1^{\text{DS}}\}, \{\mu_2^{\text{DS}}, \lambda_2^{\text{DS}}, M_2^{\text{DS}}, \alpha_2^{\text{DS}}, \phi_2^{\text{DS}}, \kappa_2^{\text{DS}}\})$. Our comparative analysis with GradNorm and SoftAdapt predictions is however conducted in each neighborhood separately. This is motivated by the logical comparison of these approaches, provided at the end of Section 3, realizing that SoftAdapt associates the variation of loss components only to their convergence behavior and may not differentiate the impact of distinct physical scales on variation magnitudes of different objectives. Moreover,

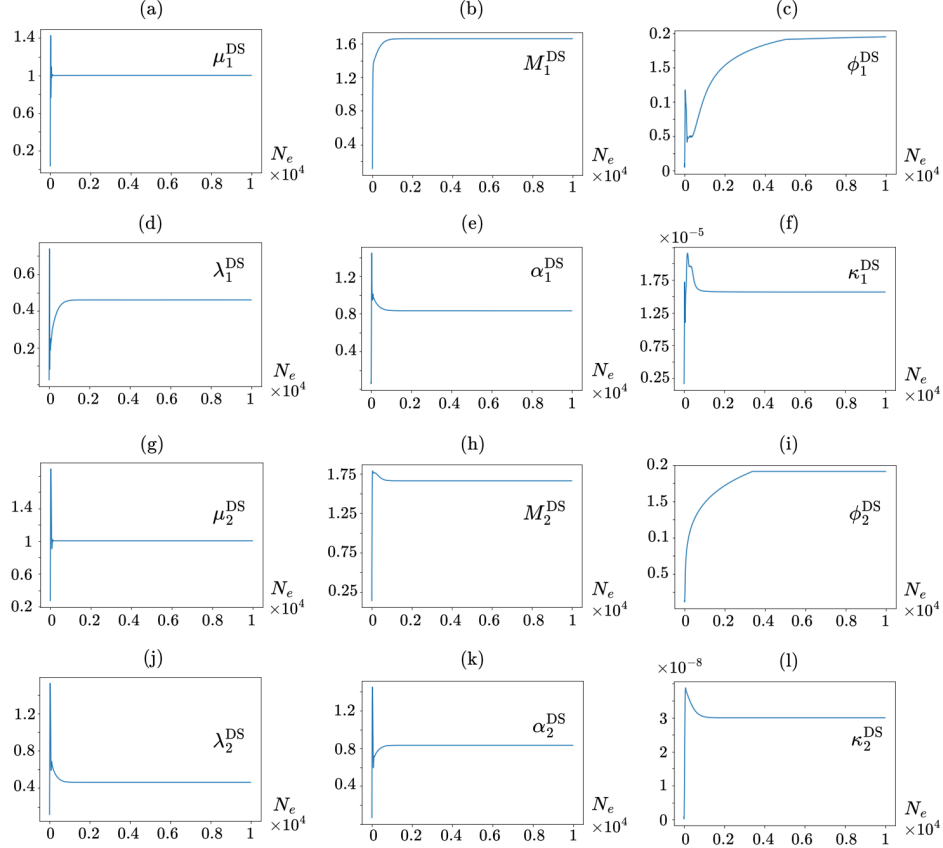


Figure 4: Network-predicted poroelastic properties *vs.* number of epochs N_e when the reconstruction is simultaneously conducted in the focal regions $\{\xi_i\}$, $i = 1, 2$. The network is a scaled MLP and the loss is balanced using the proposed Dynamic Scaling (DS) approach: (a, g) drained shear modulus μ_i^{DS} , (b, h) Biot modulus M_i^{DS} , (c, i) porosity ϕ_i^{DS} , (d, j) drained first Lamé parameter λ_i^{DS} , (e, k) Biot effective stress coefficient α_i^{DS} , (f, l) permeability coefficient κ_i^{DS} .

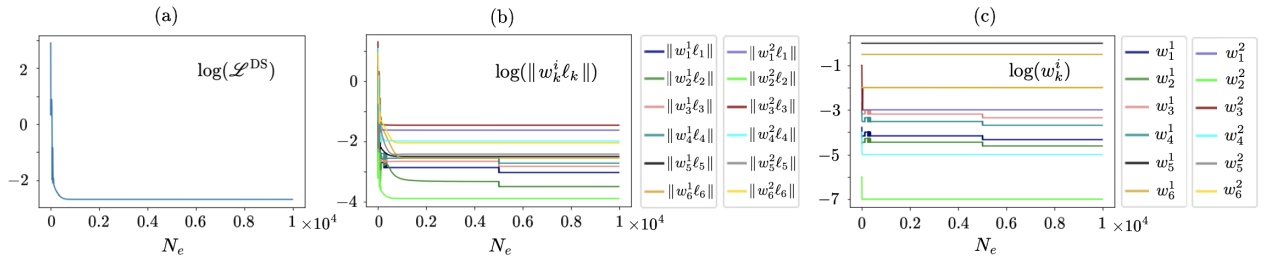


Figure 5: Convergence plots *i.e.*, loss *vs.* epoch corresponding to the results of Fig. 4 where the reconstruction is simultaneously conducted in the focal regions $\{\xi_i\}$, $i = 1, 2$. The network is a scaled MLP and the loss is balanced using the proposed Dynamic Scaling (DS) approach: (a) total loss $\log(\mathcal{L}^{\text{DS}})$ trajectory against the number of epochs, (b) weighted loss components $\log(\|w_k^i \ell_k\|)$ with ℓ_k , $k = 1, \dots, 6$, denoting the k^{th} loss component and w_k^i the associated weight in the i^{th} focal region, (c) loss weights $\log(w_k^i)$.

for computational efficiency, the GradNorm typically normalizes the gradients of weighted loss components with respect to the parameters in a shared layer in the neural network (as opposed to all network parameters) which may create some stiffness in loss balancing depending on the network architecture and input/output. Given this, it may be more insightful to compare the reconstruction results in each neighborhood separately. This may also shine some light on the impact of permeability on the reconstruction results using different methods.

Table 2 provides the reconstruction results and associated normal errors, with respect to the ground truth, in the high-permeability neighborhood ξ_1 . Here, the property map is a scaled MLP as shown in Fig. 3. Four loss balancing schemes are implemented, namely: DynScl (DS), GradNorm (GN), SoftAdapt (SA), and equal weights (EW) or no balancing. The reconstruction results are then accordingly denoted by $\boldsymbol{\vartheta}_1^{\text{DS}}$, $\boldsymbol{\vartheta}_1^{\text{GN}}$, $\boldsymbol{\vartheta}_1^{\text{SA}}$, and $\boldsymbol{\vartheta}_1^{\text{EW}}$ where

$$\boldsymbol{\vartheta}_1^\star = \{\mu_1^\star, \lambda_1^\star, M_1^\star, \alpha_1^\star, \phi_1^\star, \kappa_1^\star\}, \quad \star = \{\text{EW}, \text{DS}, \text{GN}, \text{SA}\}.$$

In all cases, the training data and network architecture remains the same. DS, GN, and SA methods outperform the case of equal weights, highlighting the importance of loss balancing in identification of multiphysics and multiscale systems. Note that the maximum normal error associated with EW, DS, GN, and SA reconstructions respectively read 100%, 2.2%, 18.36%, and 9.8%. The largest error does not correspond to the same parameter in different reconstructions. More specifically, the maximum error using equal weights and dynamic scaling is related to the drained first Lamé parameter λ_1^{EW} and λ_1^{DS} , while in the case of GradNorm and SoftAdapt, the largest error pertains to porosity ϕ_1^{GN} and Biot modulus M_1^{SA} respectively. The convergence plots along with network predictions as a function of optimizer step N_e are shown in Figs 6, 7, 8 for reconstructions by way of DynScl, GradNorm, and SoftAdapt respectively.

Table 3 reports the reconstruction results when the exercise in Table 2 is repeated without network scaling. Here, the property map is an MLP without the scaling layer i.e., the scaling layer is replaced by an affine map with unknown parameters that is standard in MLP architectures. In this case, the reconstruction fails regardless of the choice of loss balancing technique. This exercise highlights the importance of proper scaling of the model in tandem with loss balancing for a successful data inversion. It is note worthy that the maximum normalized error is first associated with the permeability coefficient κ_1^\star and next porosity ϕ_i^\star both related to the fluid dynamics in the interstitial pore space that is distinct (in terms of scale) from the physics of wave motion in the solid rock.

Table 4 presents the reconstruction results and affiliated normalized errors in the low-permeability neighborhood ξ_2 . Similar to Table 2, the property map is a scaled MLP as in Fig. 3 and four loss balancing techniques are deployed for training. Again, DS, GN, and SA methods outperform the case of equal weights. The largest error using equal weights, dynamic scaling, GradNorm and SoftAdapt are respectively related to porosity ϕ_2^{EW} , permeability coefficient κ_2^{DS} , porosity ϕ_2^{GN} and Biot modulus M_2^{SA} and read 71%, 20%, 35%, and 40%. In the SoftAdapt reconstruction, there are multiple terms with large associated normal error. The convergence plots along with network predictions as a function of the number of epochs N_e are shown in Figs 9, 10, 11 for reconstructions by way of DynScl, GradNorm, and SoftAdapt respectively.

In both neighborhoods, dynamic scaling seems to stabilize the training process and improves the network predictions compared to GradNorm and SoftAdapt. This is in particular evident from Fig. 5 (b) where all weighted loss components according to DynScl appear to converge at an approximately similar rates. This is consistent with our analysis at the end of Section 3. SoftAdapt appears to be generally unstable. This may be observed from the loss trajectory in Figs. 8 and 11 and the fact that the network prediction for porosity ϕ_1^{SA} and ϕ_2^{SA} does not converge in both cases when the training process is apparently stalled. GradNorm demonstrates greater stability in terms of the loss behavior. This balancing scheme, however, runs the risk of diminishing the loss sensitivity to a few model parameters. This, for instance, has led to the lack of convergence of porosity ϕ_1^{GN} in Fig. 10 (c), and Biot modulus M_2^{GN} in Fig. 7 (e).

4.2. Reconstruction from noisy data

This section aims to investigate the impact of noise in data on the poroelastic properties recovered by way of the proposed approach for model scaling and loss balancing. In this vein, the field data in Section 4.1 in

Table 2: Reconstructed poroelastic properties in the high-permeability neighborhood ξ_1 . Network scaling is applied here.

$\boldsymbol{\vartheta}_1$	μ_1	λ_1	M_1	α_1	ϕ_1	κ_1
ground truth (dimensionless value)	1	0.47	1.66	0.83	0.195	1.5407×10^{-5}
Equal Weights ($\boldsymbol{\vartheta}_1^{\text{EW}}$)	1.0003	0.0005	1.4937	1.1618	0.0492	1.9866×10^{-5}
DynScl ($\boldsymbol{\vartheta}_1^{\text{DS}}$)	1.0015	0.4599	1.6603	0.8307	0.1971	1.5667×10^{-5}
GradNorm ($\boldsymbol{\vartheta}_1^{\text{GN}}$)	0.99995	0.4721	1.72156	0.8276	0.2308	1.5369×10^{-5}
SoftAdapt ($\boldsymbol{\vartheta}_1^{\text{SA}}$)	0.9999	0.4665	1.4974	0.832	0.1861	1.5447×10^{-5}
$\Xi(\boldsymbol{\vartheta}_1^{\text{EW}})$	0.03%	100%	10%	40%	75%	29%
$\Xi(\boldsymbol{\vartheta}_1^{\text{DS}})$	0.15%	2.2%	0.018%	0.084%	1.4%	1.7%
$\Xi(\boldsymbol{\vartheta}_1^{\text{GN}})$	0.005%	0.45%	3.7%	0.29%	18.36%	0.25%
$\Xi(\boldsymbol{\vartheta}_1^{\text{SA}})$	0.01%	0.74%	9.8%	0.24%	4.6%	0.26%

Table 3: Reconstructed poroelastic properties in the high-permeability neighborhood ξ_1 without network scaling.

$\boldsymbol{\vartheta}_1$	μ_1	λ_1	M_1	α_1	ϕ_1	κ_1
ground truth	1	0.47	1.66	0.83	0.195	1.5407×10^{-5}
Equal Weights ($\boldsymbol{\vartheta}_1^{\text{EW}}$)	0.991	0.0006123	1.5014	1.1456	0.0214	2.0242
DynScl ($\boldsymbol{\vartheta}_1^{\text{DS}}$)	0.662	1.813	1.491	1.617	0.121	0.987
GradNorm ($\boldsymbol{\vartheta}_1^{\text{GN}}$)	0.9901	0.249	1.507	0.974	0.00056	1.938
SoftAdapt ($\boldsymbol{\vartheta}_1^{\text{SA}}$)	0.984	0.6187	1.3479	0.856	0.00003614	1.436
$\Xi(\boldsymbol{\vartheta}_1^{\text{EW}})$	0.9%	99.86%	9.55%	38.02%	89.025641%	13138083%
$\Xi(\boldsymbol{\vartheta}_1^{\text{DS}})$	33.8%	285.745%	10.181%	94.819%	37.949%	6406079%
$\Xi(\boldsymbol{\vartheta}_1^{\text{GN}})$	0.990%	47%	9.217%	17.349%	99.713%	12578597%
$\Xi(\boldsymbol{\vartheta}_1^{\text{SA}})$	1.6%	31.64%	18.801%	3.133%	99.981%	9320338%

Table 4: Reconstructed poroelastic properties in the low-permeability neighborhood ξ_2 . Network scaling is applied here.

$\boldsymbol{\vartheta}_2$	μ_2	λ_2	M_2	α_2	ϕ_2	κ_2
ground truth	1	0.47	1.66	0.83	0.195	2.45×10^{-8}
Equal Weights ($\boldsymbol{\vartheta}_2^{\text{EW}}$)	1.001	0.4456	1.6551	0.8395	0.0571	2.631×10^{-8}
DynScl ($\boldsymbol{\vartheta}_2^{\text{DS}}$)	1.0036	0.4652	1.6596	0.8306	0.191	2.934×10^{-8}
GradNorm ($\boldsymbol{\vartheta}_2^{\text{GN}}$)	1.002	0.4713	1.569	0.833	0.262	2.485×10^{-8}
SoftAdapt ($\boldsymbol{\vartheta}_2^{\text{SA}}$)	0.9998	0.4702	2.3189	0.8297	0.2164	2.0235×10^{-8}
$\Xi(\boldsymbol{\vartheta}_2^{\text{EW}})$	0.001%	5.2%	0.3%	1.1%	71%	7.4%
$\Xi(\boldsymbol{\vartheta}_2^{\text{DS}})$	0.36%	1%	0.024%	0.072%	2.1%	20%
$\Xi(\boldsymbol{\vartheta}_2^{\text{GN}})$	0.02%	0.27%	5.5%	0.036%	35%	1.4%
$\Xi(\boldsymbol{\vartheta}_2^{\text{SA}})$	0.02%	0.043%	40%	0.036%	11%	17%

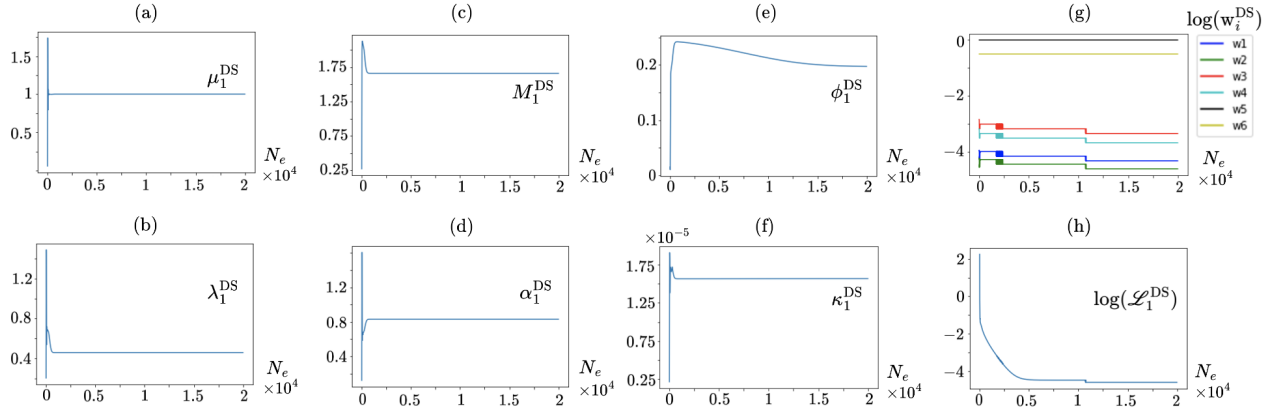


Figure 6: Network-predicted poroelastic properties *vs.* the number of epochs N_e when the reconstruction is conducted in the high-permeability neighborhood ξ_1 . The network is a scaled MLP and the loss is balanced using the proposed Dynamic Scaling (DS) approach: (a) drained shear modulus μ_1^{DS} , (b) drained first Lamé parameter λ_1^{DS} , (c) Biot modulus M_1^{DS} , (d) Biot effective stress coefficient α_1^{DS} , (e) porosity ϕ_1^{DS} , (f) permeability coefficient κ_1^{DS} , (g) DS weights w_i^{DS} , $i = 1, \dots, 6$, versus the minimizer step N_e , (h) weighted total loss $\log(\mathcal{L}_1^{\text{DS}})$ trajectory against the number of epochs.

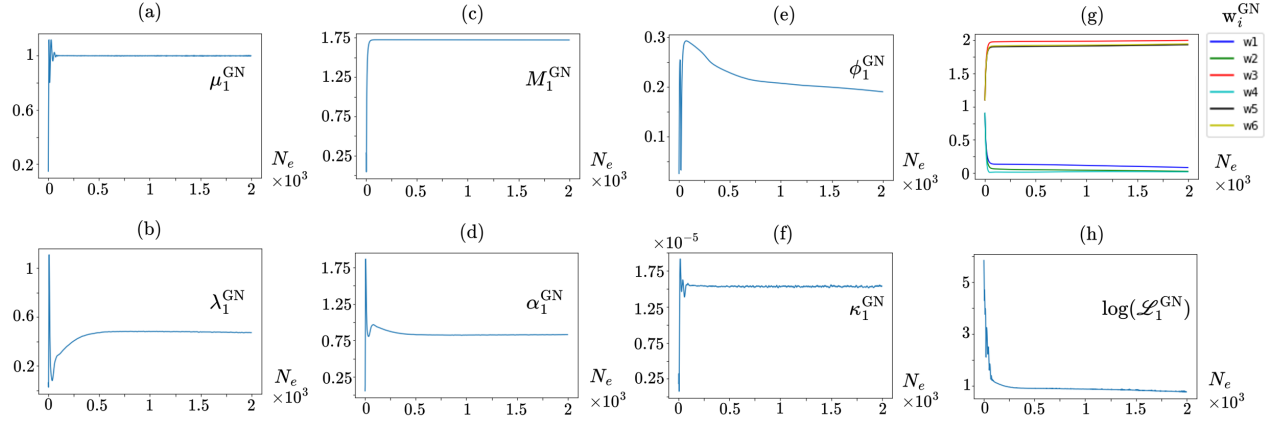


Figure 7: Network-predicted poroelastic properties *vs.* the number of epochs N_e in the high-permeability neighborhood ξ_1 . The network is a scaled MLP and the loss is balanced using the GradNorm (GN) approach: (a) drained shear modulus μ_1^{GN} , (b) drained first Lamé parameter λ_1^{GN} , (c) Biot modulus M_1^{GN} , (d) Biot effective stress coefficient α_1^{GN} , (e) porosity ϕ_1^{GN} , (f) permeability coefficient κ_1^{GN} , (g) GN weights w_i^{GN} , $i = 1, \dots, 6$, versus the minimizer step N_e , (h) weighted total loss $\log(\mathcal{L}_1^{\text{GN}})$ *vs.* epoch.

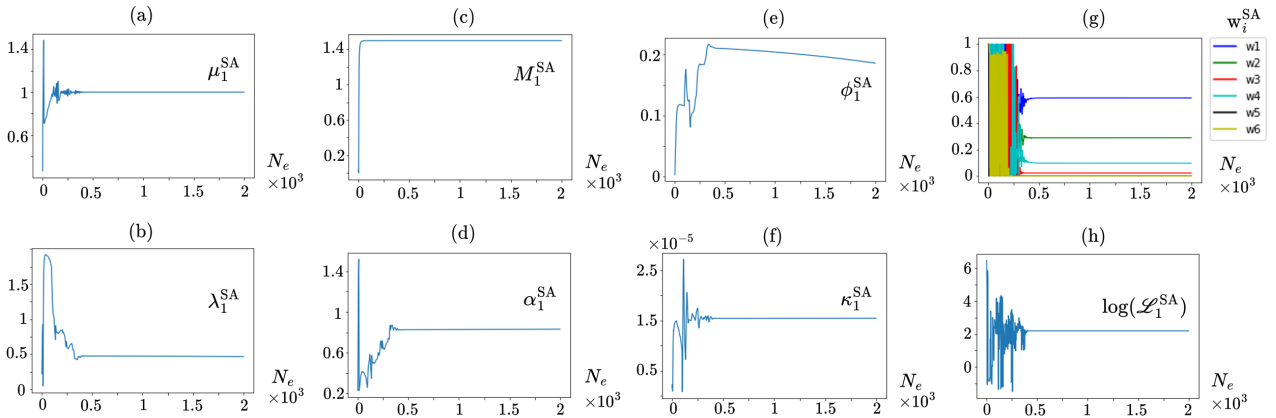


Figure 8: Network-predicted poroelastic properties *vs.* the number of epochs N_e in the high-permeability neighborhood ξ_1 . The network is a scaled MLP and the loss is balanced using the SoftAdapt (SA) approach: (a) drained shear modulus μ_1^{SA} , (b) drained first Lamé parameter λ_1^{SA} , (c) Biot modulus M_1^{SA} , (d) Biot effective stress coefficient α_1^{SA} , (e) porosity ϕ_1^{SA} , (f) permeability coefficient κ_1^{SA} , (g) SA weights w_i^{SA} , $i = 1, \dots, 6$, versus the minimizer step N_e , (h) weighted total loss $\log(\mathcal{L}_1^{\text{SA}})$ *vs.* epoch.

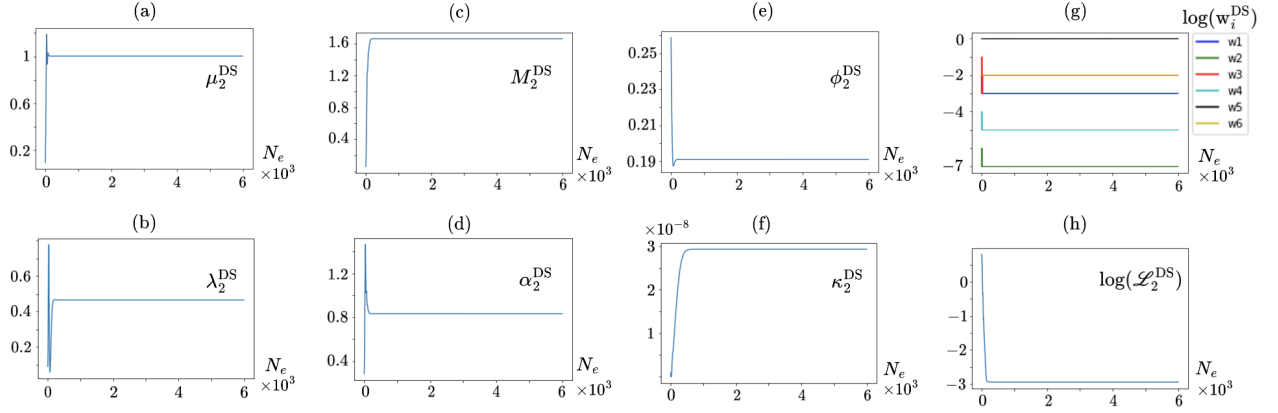


Figure 9: Network-predicted poroelastic properties *vs.* the number of epochs N_e when the reconstruction is conducted in the low-permeability neighborhood ξ_2 . The network is a scaled MLP and the loss is balanced using the proposed Dynamic Scaling (DS) approach: (a) drained shear modulus μ_2^{DS} , (b) drained first Lamé parameter λ_2^{DS} , (c) Biot modulus M_2^{DS} , (d) Biot effective stress coefficient α_2^{DS} , (e) porosity ϕ_2^{DS} , (f) permeability coefficient κ_2^{DS} , (g) DS weights w_i^{DS} , $i = 1, \dots, 6$, versus the minimizer step N_e , (h) weighted total loss $\log(\mathcal{L}_2^{\text{DS}})$ trajectory against the number of epochs.

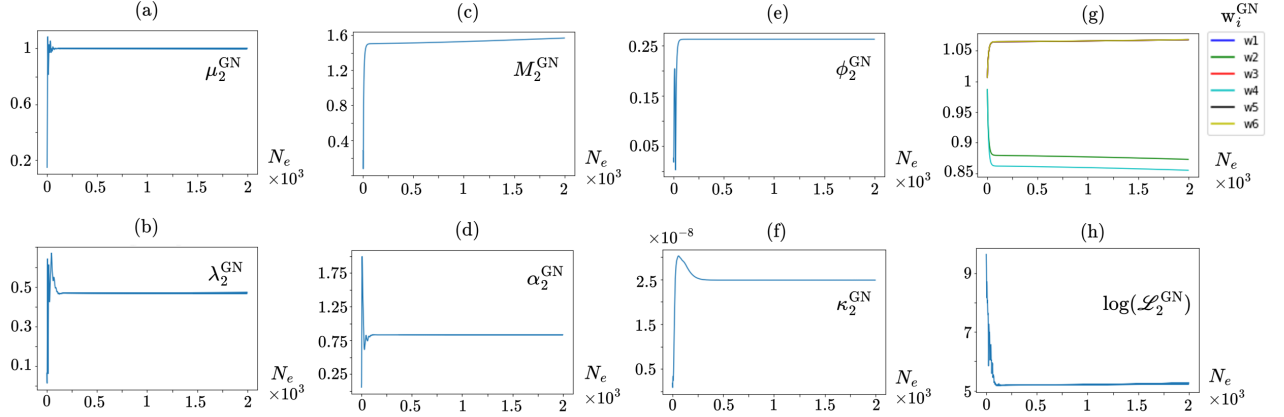


Figure 10: Network-predicted poroelastic properties *vs.* the number of epochs N_e in the low-permeability neighborhood ξ_2 . The network is a scaled MLP and the loss is balanced using the GradNorm (GN) approach: (a) drained shear modulus μ_2^{GN} , (b) drained first Lamé parameter λ_2^{GN} , (c) Biot modulus M_2^{GN} , (d) Biot effective stress coefficient α_2^{GN} , (e) porosity ϕ_2^{GN} , (f) permeability coefficient κ_2^{GN} , (g) GN weights w_i^{GN} , $i = 1, \dots, 6$, versus the minimizer step N_e , (h) weighted total loss $\log(\mathcal{L}_2^{\text{GN}})$ *vs.* epoch.

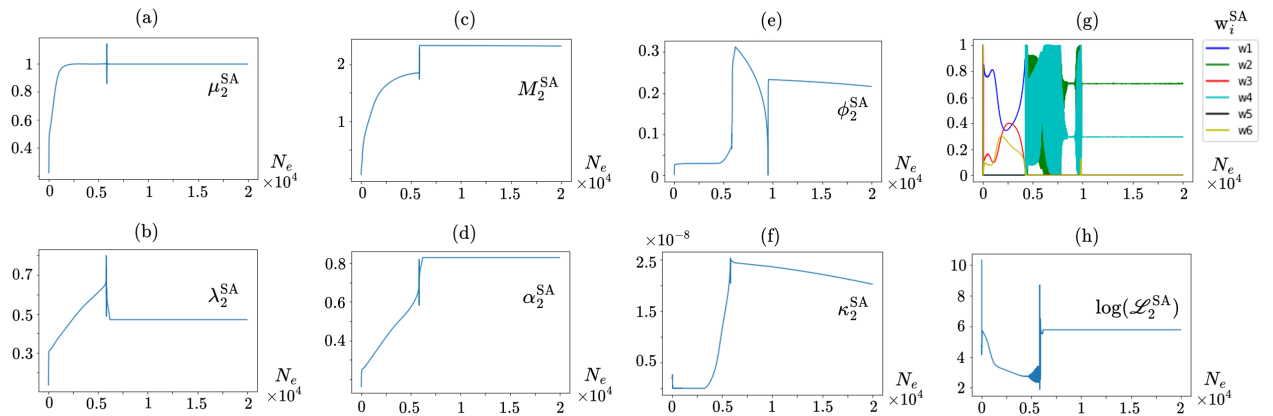


Figure 11: Network-predicted poroelastic properties *vs.* the number of epochs N_e in the low-permeability neighborhood ξ_2 . The network is a scaled MLP and the loss is balanced using the SoftAdapt (SA) approach: (a) drained shear modulus μ_2^{SA} , (b) drained first Lamé parameter λ_2^{SA} , (c) Biot modulus M_2^{SA} , (d) Biot effective stress coefficient α_2^{SA} , (e) porosity ϕ_2^{SA} , (f) permeability coefficient κ_2^{SA} , (g) SA weights w_i^{SA} , $i = 1, \dots, 6$, versus the minimizer step N_e , (h) weighted total loss $\log(\mathcal{L}_2^{\text{SA}})$ *vs.* epoch.

focal neighborhood ξ_1 is perturbed by 5% noise as the following

$$\begin{aligned}\tilde{\mathbf{Z}} &= \mathbf{Z} + \mathfrak{n} \max(|\mathbf{Z}|) \left(I_{N_x \times N_y} + i I_{N_x \times N_y} \right), \quad \mathfrak{n} = 0.05, \\ \mathbf{Z}(r, s) &= z(x_r, y_s), \quad z = u_x, u_y, p, \quad r = 1, \dots, N_x, \quad s = 1, \dots, N_y,\end{aligned}\tag{23}$$

where N_x and N_y are respectively the number of samples in x and y directions in the high-permeability neighborhood ξ_1 . Note that introducing the noise as in (23) as a percentage of the signal range is consistent with how measurement noise is quantified in practice. In multiscale systems, however, these perturbations are not equally reflected on the real and imaginary parts of the field variables. This is shown in Fig. 12 where the normal misfit between the noiseless and noisy fields are demonstrated for the decomposed solid displacement u_x and pressure p according to

$$\begin{aligned}\Theta(\tilde{\mathbf{Z}}) &= \frac{|\tilde{\mathbf{Z}} - \mathbf{Z}|}{\max(|\mathbf{Z}|)}, \quad \mathbf{Z}(r, s) = z(x_r, y_s), \quad z = \Re(u_x), \Im(u_x), \Re(u_y), \Im(u_y), \Re(p), \Im(p), \\ &r = 1, \dots, N_x, \quad s = 1, \dots, N_y.\end{aligned}\tag{24}$$

Observe that the misfit in the major components of u_x and p , namely: $\Im(u_x)$ and $\Im(p)$, reflects the added 5% noise. However, the normal difference for the minor components $\Re(u_x)$ and $\Re(p)$ could surpass 50%. This indicates that in multiscale systems, 5% noise in the signal (or data) could imply more than 50% of misfit on certain components of the wave field. It is evident from (12) and (A.1) that all components of the field variables and their spatial derivatives actively participate in the loss function (10) and depending on the frequency of wave motion the contribution of highly distorted components by noise could be significant. Thus, even moderate levels of noise could present a critical challenge in multiphysics system identification. This issue is not as problematic in one-scale physical systems. For instance, it is illustrated in our recent study [25] using both synthetic and experimental data that in elastography of composites i.e., when the PDE system is the (one-scale) Navier equations, 5% noise in data can be easily addressed using the existing tools of signal

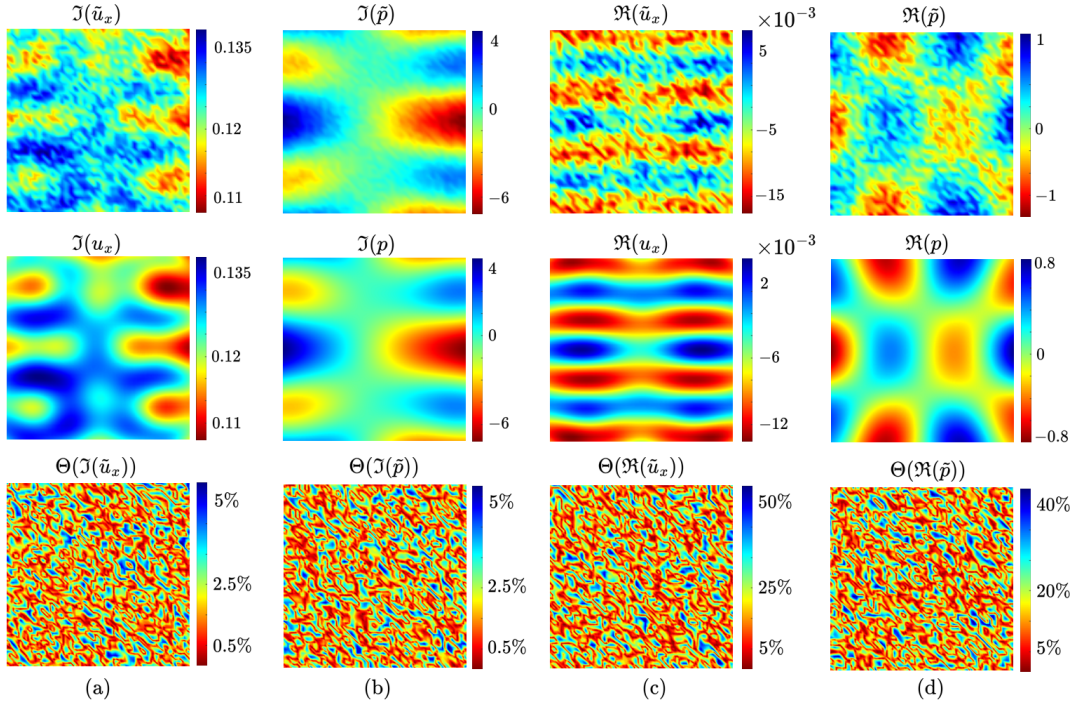


Figure 12: Normal misfit between the noiseless and noisy field components when the overall signal-to-noise ratio is 20: (a) imaginary part of solid displacement in x direction, (b) imaginary part of pore pressure, (c) real part of solid displacement in x direction, and (d) real part of pore pressure. In each column, top row is the noisy wave field, middle row is its noiseless counterpart, and the bottom row is the normalized different according to (24).

processing. To address this issue there are several options. For instance, one may switch to the weak (or other energy) form of the governing equations and take advantage of inherent averaging and thus smoothing nature of integral equations, e.g., see [44, 49, 50, 51] for this type of formulation that so far has been implemented in the case of one-scale PDE systems. Here the choice of test functions and boundary contributions will be the main questions. Another approach that is more consistent with the strong formulation of loss function, which is adopted in this study, is to take advantage of signal averaging couple with spectral denoising and FFT-based differentiation of the focal fields. The latter two methods are already deployed in the processing of noiseless signals in Section 4.1 and a detailed description of them may be found in [25, 26]. For averaging, we assume each test could be repeated N_T times so that the noisy focal fields \tilde{z}_t , $t = 0, \dots, N_T$ can be averaged and used for training in the following form

$$\langle \tilde{z} \rangle_{N_T} = \frac{1}{N_T} \sum_{t=1}^{N_T} \tilde{z}_t, \quad \tilde{z}_t = \Re(\tilde{u}_{x_t}), \Im(\tilde{u}_{x_t}), \Re(\tilde{u}_{y_t}), \Im(\tilde{u}_{y_t}), \Re(\tilde{p}_t), \Im(\tilde{p}_t), \quad (25)$$

265 In what follows, we set $N_T = 250, 1500, 2500$. In laboratory experiments, signal averaging is typically set anywhere between 100 to 1000 averages in every experiment [25, 26]. We also selected a significantly larger number of realizations at $N_T = 2500$ to examine whether by using a sufficiently large ensemble, one could push the reconstruction error arbitrary close to zero regardless of practical considerations.

Table 5 presents the reconstruction results from noisy data and associated normal errors, with respect to the ground truth, in the high-permeability neighborhood ξ_1 . Here, the property map remains the scaled MLP shown in Fig. 3. The proposed dynamic scaling for loss balancing is implemented. The reconstruction results are then accordingly denoted by

$$\boldsymbol{\vartheta}_{<1>N_T}^{\text{DS}} = \{\mu_1^{\text{DS}}, \lambda_1^{\text{DS}}, M_1^{\text{DS}}, \alpha_1^{\text{DS}}, \phi_1^{\text{DS}}, \kappa_1^{\text{DS}}\}_{<1>N_T}, \quad N_T \in \{250, 1500, 2500\}.$$

In all cases, the network architecture remains the same. μ_1^{DS} remains very close to the ground truth, with minimal deviation across all N_T values. The normal deviation of μ_1^{DS} from the ground truth decreases from 0.99% at 250 averages to 0.0795% at 2500 averages, indicating an improved accuracy with an increase in averaging. λ_1^{DS} shows a slight underestimation compared to its true value. The relative error of λ_1^{DS} demonstrates a significant reduction from 6.38% at 250 averages to 1.14% at 2500 averages, suggesting a substantial enhancement in precision with larger ensembles. M_1^{DS} is close to the ground truth, with minor fluctuations. 270 The relative error of M_1^{DS} diminishes from 3.27% at 250 averages to 0.0881% at 2500 averages, reflecting a similar trend in the reconstructions. α_1^{DS} shows minor deviations from the ground truth. The relative error of α_1^{DS} reduces significantly from 3.97% at 250 averages to 0.435% at 2500 averages. κ_1^{DS} values are fairly close to the ground truth at 1.5407×10^{-5} with small variations. The relative error decreases from 33.7% at 250 averages to 2.55% at 2500 averages, showing a marked improvement in accuracy with increased sample size. 280 ϕ_1^{DS} exhibits considerable variations from the ground truth, especially at lower N_T . The relative error drops from 65.2% at 250 averages to 34.9% at 2500 averages. However, the true value was never recovered regardless of the ensemble size. Overall, the analysis of Table 5 indicates that the accuracy of network estimates improves with the increase in averaging coupled with signal denoising and FFT-based differentiation. However, these measures do not fundamentally address the challenges of reconstructions from noisy data which could be the subject of future studies. 285 The convergence plots along with network predictions as a function of optimizer step N_e are shown in Figs 13, 14, 15 for the reconstructions reported in Table 5. Similar to reconstructions from noiseless data, training by way of dynamic scaling shows stable trajectories for both loss and network predictions.

4.3. Discussion

290 Section 4 showcases the capability of scaled neural networks in simultaneous reconstruction of poroelastic properties in distinct neighborhoods whose permeability coefficients differ by multiple orders of magnitude. This is thanks to the last layer in the proposed architecture which allows the network predictions to straddle a range of potential scales in heterogenous environments and does not require an exact a priori knowledge of the scaling of unknown physical properties. The latter would be required if the scaling were to be hardcoded. 295 Note that in this model, the neural map is factorized into two operators; the first operator, which involves the

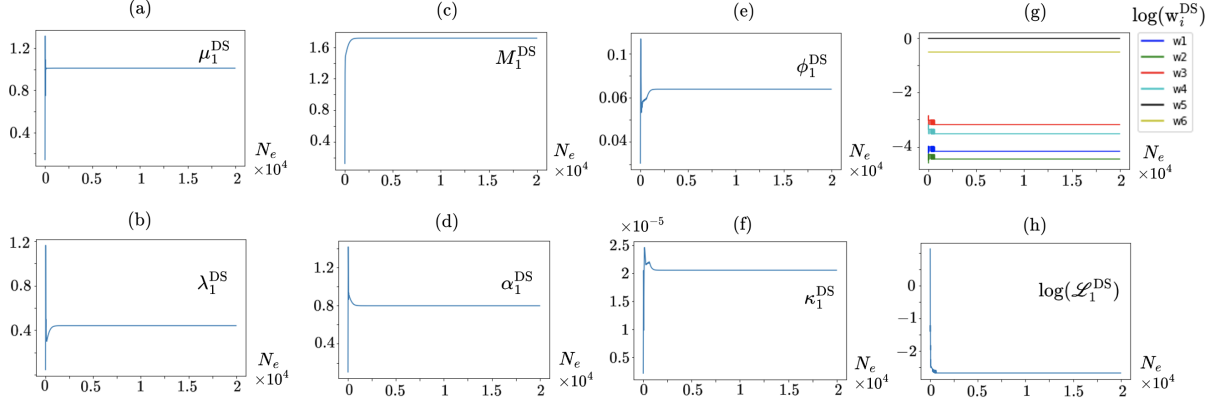


Figure 13: Network-predicted poroelastic properties *vs.* the number of epochs N_e in the high-permeability neighborhood ξ_1 . The network is a scaled MLP and trained by noisy waveforms that are averaged $N_T = 250$ times. The loss is balanced using the dynamic scaling (DS) approach: (a) drained shear modulus μ_1^{DS} , (b) drained first Lamé parameter λ_1^{DS} , (c) Biot modulus M_1^{DS} , (d) Biot effective stress coefficient α_1^{DS} , (e) porosity ϕ_1^{DS} , (f) permeability coefficient κ_1^{DS} , (g) DS weights w_i^{DS} , $i = 1, \dots, 6$, versus the minimizer step N_e , (h) weighted total loss $\log(\mathcal{L}_1^{\text{DS}})$ *vs.* epoch.

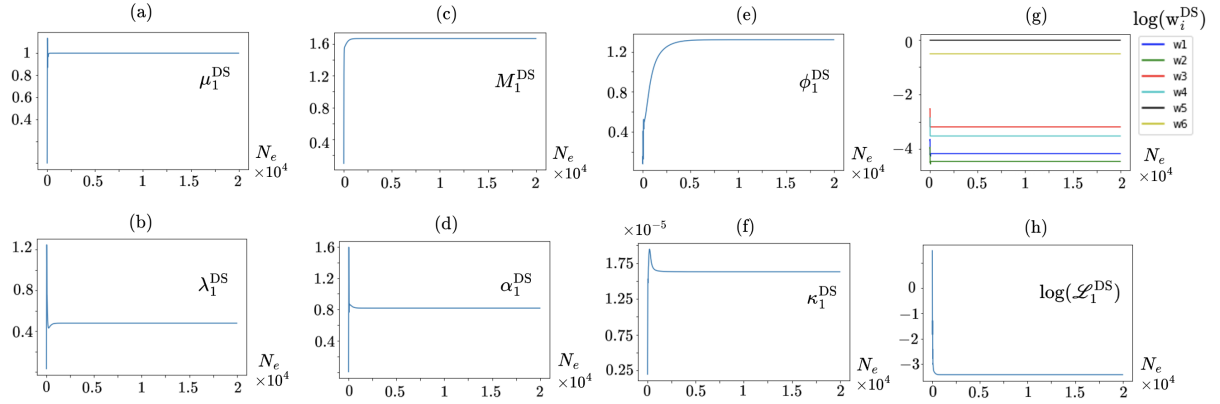


Figure 14: Network-predicted poroelastic properties *vs.* the number of epochs N_e in the high-permeability neighborhood ξ_1 . The network is a scaled MLP and trained by noisy waveforms that are averaged $N_T = 1500$ times. The loss is balanced using the dynamic scaling (DS) approach: (a) drained shear modulus μ_1^{DS} , (b) drained first Lamé parameter λ_1^{DS} , (c) Biot modulus M_1^{DS} , (d) Biot effective stress coefficient α_1^{DS} , (e) porosity ϕ_1^{DS} , (f) permeability coefficient κ_1^{DS} , (g) DS weights w_i^{DS} , $i = 1, \dots, 6$, versus the minimizer step N_e , (h) weighted total loss $\log(\mathcal{L}_1^{\text{DS}})$ *vs.* epoch.

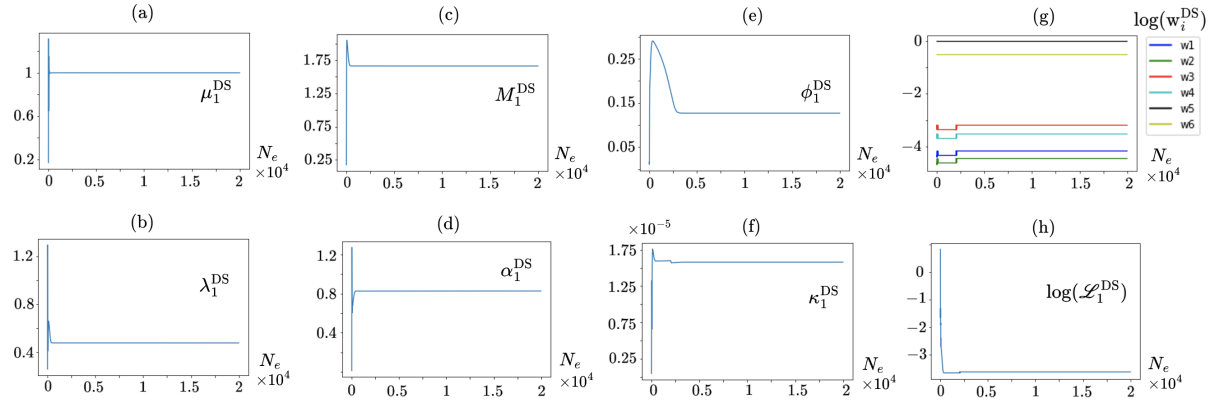


Figure 15: Network-predicted poroelastic properties *vs.* the number of epochs N_e in the high-permeability neighborhood ξ_1 . The network is a scaled MLP and trained by noisy waveforms that are averaged $N_T = 2500$ times. The loss is balanced using the dynamic scaling (DS) approach: (a) drained shear modulus μ_1^{DS} , (b) drained first Lamé parameter λ_1^{DS} , (c) Biot modulus M_1^{DS} , (d) Biot effective stress coefficient α_1^{DS} , (e) porosity ϕ_1^{DS} , (f) permeability coefficient κ_1^{DS} , (g) DS weights w_i^{DS} , $i = 1, \dots, 6$, versus the minimizer step N_e , (h) weighted total loss $\log(\mathcal{L}_1^{\text{DS}})$ *vs.* epoch.

Table 5: Reconstruction of poroelastic properties in the high-permeability neighborhood ξ_1 from noisy data. The waveforms are averaged $N_T \in \{250, 1500, 2500\}$ times. Network scaling is applied here.

$\boldsymbol{\vartheta}_1$	μ_1	λ_1	M_1	α_1	ϕ_1	κ_1
ground truth	1	0.47	1.66	0.83	0.195	1.5407×10^{-5}
DynScl ($\boldsymbol{\vartheta}_{<1>250}^{\text{DS}}$)	1.00991	0.44	1.714	0.7971	0.06786	2.06×10^{-5}
DynScl ($\boldsymbol{\vartheta}_{<1>1500}^{\text{DS}}$)	1.00003	0.479	1.6696	0.8175	0.13199	1.6295×10^{-5}
DynScl ($\boldsymbol{\vartheta}_{<1>2500}^{\text{DS}}$)	0.9992	0.475	1.6614	0.8264	0.12678	1.5799×10^{-5}
$\Xi(\boldsymbol{\vartheta}_{<1>250}^{\text{DS}})$	0.99%	6.38%	3.27%	3.97%	65.2%	33.7%
$\Xi(\boldsymbol{\vartheta}_{<1>1500}^{\text{DS}})$	0.00312%	1.98%	0.577%	1.51%	32.3%	5.76%
$\Xi(\boldsymbol{\vartheta}_{<1>2500}^{\text{DS}})$	0.0795%	1.14%	0.0881%	0.435%	34.9%	2.55%

unknown weights and biases, acts as a unit (or normal) shape function describing the spatial distribution of each property, while the second operator magnifies the first operator’s output according to a pre-determined set of likely scales for each physical property. It should be mentioned that in this formulation since the scaling of output is isolated and transferred as a whole to the last layer, the unknown network parameters (weights and biases) can be clamped at $O(1)$ during the optimization process. This accelerates the training and paves the way toward a *physics-based loss balancing approach* through establishing explicit estimates for the scale of each loss component and its gradients with respect to the model parameters. In the case of Biot equations, used for poroelastography, we took advantage of this architecture to show that each weighted loss component is a Lipschitz function of the network weights and the proposed dynamic scaling (DynScl) of the loss – based on the average scale of each loss component – ensures that the Lipschitz constant remains of $O(1)$ for all network parameters during training. This manifests itself in three advantages that we observed in the reconstructions from noiseless data, especially when the results were compared with that of Softmax adaptive weights (SoftAdapt) and gradient normalization (GradNorm). The first advantage is *stability and robustness*; compared to both SoftAdapt and GradNorm, the training by DynScl weights is more stable in that (a) all weighted loss components and the total loss converge with a unified rate, (b) the proper weight for each objective is quickly identified and remains stable throughout the optimization process, (c) the total loss converges to zero, and (d) all the network-predicted properties show convergence to certain limits. (a) and (b) remain the case in GradNorm, even though the estimated loss weights by the DynScl and GradNorm differ by orders of magnitude. Keep in mind that GradNorm optimizes the weights in each epoch in order to unify the rate of change of weighted loss components with respect to a shared layer in the neural network (not all the network parameters due to computational considerations). In Section 4, all GradNorm-estimated weights are of $O(1)$, in both focal regions, which could explain the remarkably large magnitude of the total loss and its potential insensitivity to the physics at smaller scales. In this case, while the total loss uniformly decreases, we observe that a few network-estimated parameters related to the physics of fluid flow in the pore space do not converge i.e., (c) and (d) do not seem to hold in GradNorm-based balancing. Another point worth mentioning is the impact of noise in data on the loss balancing procedure. Note that DynScl only requires an estimated scale for the field derivatives. While direct FFT-based differentiation of focal fields provides the best scaling estimate, as implemented in this study, this is not generally required and an estimate on the scale of derivatives can be obtained by an elementary FFT analysis of the measurements coupled with network-predicted scales for the physical properties at every epoch. GradNorm relies on the estimated derivatives furnished by automatic differentiation to optimize the loss weights at every epoch. The latter when combined with noisy or incomplete data may lead to instability as reported by [43]. In this study, since the data is dense and the waveforms were carefully processed prior to training for any of the loss balancing techniques, this problem was not observed. SoftAdapt is generally unstable in the reconstructions of Section 4. We believe that this is due to the limited range and sensitivity of the Softmax function for multiphysics objectives. Varying between zero and one, SoftAdapt weights are assigned such that objectives with small decay rates are more visible to the optimizer. In multiphysics systems, the decay rates are largely influenced by the scale of each loss component such that a fast converging task could assume a much smaller decay rate compared to a slow converging objective of larger magnitude. This could confuse or mislead the optimizer to focus on the objectives that are already converging

335 and reduce the loss sensitivity to certain outputs. This may explain the behavior observed in the SoftAdapt
 convergence plots wherein while the loss converges, some of the network-predicted properties are diverging.
 The second advantage of the proposed approach is in its *flexibility and efficiency*. Network scaling is a quite
 simple idea and can be applied to any architecture. Moreover, DynScl weights can be explicitly computed for
 340 both Lipschitz and non-Lipschitz objectives (see Remark 1) and does not require a parallel optimization process
 per epoch. In this respect, the computational cost of DynScl and SoftAdapt is similar. The third advantage
 of loss balancing by way of dynamic scaling is *clarity*. The DynScl weights are directly driven by the physics
 of each objective such that each loss component represents the balance of physical quantities of $O(1)$. In this
 framework, the total loss converges to zero ensuing that all objectives are met. In reconstructions of Section 4,
 the SoftAdapt and GradNorm weights are all of $O(1)$ and do not seem to be cognizant of the physics of loss
 345 components. This resulted in quite large-scale total losses that while significantly decrease by multiple orders
 during training, they do not converge to zero (or a near-zero value), causing an ambiguity in interpretation
 of the results as to whether the global convergence is achieved and all objectives are sufficiently met. In the
 reconstructions from noisy data, we showed that DynScl weights remain stable. We also implemented signal
 averaging coupled with spectral denoising and FFT-based differentiation to improve the reconstructions. The
 350 results suggest that while the proposed measures significantly enhance the accuracy of network predictions,
 they do not fully address the challenges involved in multiphysics system identification from noisy data. These
 challenges seem to be more related to errors in the calculation of derivatives and the question of observability
 of small-scale physics in presence of noise, rather than loss balancing which could be the subject of a future
 work.

355 5. Conclusion

An intelligent framework is established for identification of multiphysics systems from (reconstructed or
 measured) local wave fields. As an application, we focused on comprehensive characterization of poroelastic
 materials from solid displacement and pore pressure data. To this end, the neural maps, predicting the six
 unknown parameters of Biot equations, are constructed by a normal map describing the spatial distribution of
 360 each property that is composed into a scaling layer. This architecture allows its various outputs to assume a
 set of likely scales and does not require an exact knowledge of the scaling of unknown physical properties. In
 this model the unknown network parameters in the normal map remain of $O(1)$ during training. This forms
 the basis for physics-based balancing of the loss function through the proposed dynamic scaling. The latter
 is formulated separately for Lipschitz and non-Lipschitz objectives (with respect to the network’s weights and
 365 biases) through establishing explicit estimates for the scale of each loss component and its gradients with
 respect to the model parameters. The proposed approach that integrates the idea of network scaling with
 scale-driven loss balancing is then put to test by a set of numerical experiments. Special attention is paid to
 tight formations (low-permeability regions) and simultaneous reconstruction of all PDE parameters in highly
 heterogeneous environments where some unknown properties could straddle multiple scales in various neigh-
 370 borhoods. We demonstrated that the proposed method is successful in addressing some of these challenges.
 A comparative analysis is conducted with two state-of-the-art techniques for loss balancing, namely: Grad-
 Norm and SoftAdapt. We demonstrated both logically and computationally that scale-based modeling and
 loss balancing offer three advantages in terms of stability, efficiency, and accuracy that may be particularly
 relevant in multiphysics data processing. The proposed method is formulated in a generic platform so that its
 375 application to other physical systems is straightforward.

Acknowledgments

This study was funded by the National Science Foundation (Grant No. 1944812) and the University of
 Colorado Boulder through Fatemeh Pourahmadian’s startup. This work deployed resources from the University
 of Colorado Boulder Research Computing Group, which is supported by the National Science Foundation
 380 (awards ACI-1532235 and ACI-1532236), the University of Colorado Boulder, and Colorado State University.
 Special thanks are due to Kevish Napal for facilitating the use of FreeFem++ code developed as part of [28]
 for poroelastodynamic simulations.

Appendix A. Loss components

The expanded loss components ℓ_2 to ℓ_6 in (11) are listed in the following.

$$\begin{aligned}
\ell_2 &= \mu \left(\frac{\partial^2 \Im(u_x)}{\partial x^2} + \frac{\partial^2 \Im(u_x)}{\partial y^2} \right) + (\lambda + \mu) \left(\frac{\partial^2 \Im(u_x)}{\partial x^2} + \frac{\partial^2 \Im(u_y)}{\partial x \partial y} \right) - \\
&\quad - \Re(a) \frac{\partial \Im(p)}{\partial x} - \Im(a) \frac{\partial \Re(p)}{\partial x} + \omega^2 \Re(b) \Im(u_x) + \omega^2 \Im(b) \Re(u_x) - \Im(f_x^\mu), \\
\ell_3 &= \mu \left(\frac{\partial^2 \Re(u_y)}{\partial x^2} + \frac{\partial^2 \Re(u_y)}{\partial y^2} \right) + (\lambda + \mu) \left(\frac{\partial^2 \Re(u_y)}{\partial y^2} + \frac{\partial^2 \Re(u_x)}{\partial x \partial y} \right) - \\
&\quad - \Re(a) \frac{\partial \Re(p)}{\partial y} + \Im(a) \frac{\partial \Im(p)}{\partial y} + \omega^2 \Re(b) \Re(u_y) - \omega^2 \Im(b) \Im(u_y) - \Re(f_y^\mu), \\
\ell_4 &= \mu \left(\frac{\partial^2 \Im(u_y)}{\partial x^2} + \frac{\partial^2 \Im(u_y)}{\partial y^2} \right) + (\lambda + \mu) \left(\frac{\partial^2 \Im(u_y)}{\partial y^2} + \frac{\partial^2 \Im(u_x)}{\partial x \partial y} \right) - \\
&\quad - \Re(a) \frac{\partial \Im(p)}{\partial y} - \Im(a) \frac{\partial \Re(p)}{\partial y} + \omega^2 \Re(b) \Im(u_y) + \omega^2 \Im(b) \Re(u_y) - \Im(f_y^\mu), \\
\ell_5 &= \Re(a) \left(\frac{\partial \Re(u_x)}{\partial x} + \frac{\partial \Re(u_y)}{\partial y} \right) - \Im(a) \left(\frac{\partial \Im(u_x)}{\partial x} + \frac{\partial \Im(u_y)}{\partial y} \right) + \\
&\quad + \frac{\Re(c)}{\omega^2} \left(\frac{\partial \Re(p)}{\partial x^2} + \frac{\partial \Re(p)}{\partial y^2} \right) - \frac{\Im(c)}{\omega^2} \left(\frac{\partial \Im(p)}{\partial x^2} + \frac{\partial \Im(p)}{\partial y^2} \right) + \frac{\Re(p)}{M} + \Re(c) \frac{\Re(f^P)}{\omega^2} - \Im(c) \frac{\Im(f^P)}{\omega^2}, \\
\ell_6 &= \Re(a) \left(\frac{\partial \Im(u_x)}{\partial x} + \frac{\partial \Im(u_y)}{\partial y} \right) + \Im(a) \left(\frac{\partial \Re(u_x)}{\partial x} + \frac{\partial \Re(u_y)}{\partial y} \right) + \\
&\quad + \frac{\Re(c)}{\omega^2} \left(\frac{\partial \Im(p)}{\partial x^2} + \frac{\partial \Im(p)}{\partial y^2} \right) + \frac{\Im(c)}{\omega^2} \left(\frac{\partial \Re(p)}{\partial x^2} + \frac{\partial \Re(p)}{\partial y^2} \right) + \frac{\Im(p)}{M} + \Re(c) \frac{\Im(f^P)}{\omega^2} + \Im(c) \frac{\Re(f^P)}{\omega^2},
\end{aligned} \tag{A.1}$$

References

- 385 [1] S. K. Tangirala, F. Parisio, I. Vaezi, V. Vilarrasa, Effectiveness of injection protocols for hydraulic stimulation in enhanced geothermal systems, *Geothermics* 120 (2024) 103018.
- [2] A. Boyet, S. De Simone, S. Ge, V. Vilarrasa, Poroelastic stress relaxation, slip stress transfer and friction weakening controlled post-injection seismicity at the basel enhanced geothermal system, *Communications Earth & Environment* 4 (1) (2023) 104.
- 390 [3] A. I. Yerima, H. Tsegab, M. Hermana, L. H. Piccoli, A. A. Al-Masgari, Feasibility of co2 plume geothermal systems in miocene carbonate reservoirs of central luconia province, offshore sarawak: A systematic review of technologies, challenges, mitigation mechanisms and economic prospects, *Journal of Cleaner Production* (2024) 143032.
- [4] E. Anthony, N. Vedanti, Seismic low-frequency shadows and their application to detect co2 anomalies on time-lapse seismic data: A case study from the sleipner field, north sea, *Geophysics* 87 (2) (2022) B81–B91.
- 395 [5] D. R. Sowinski, M. D. J. McGarry, E. E. W. Van Houten, S. Gordon-Wylie, J. B. Weaver, K. D. Paulsen, Poroelasticity as a model of soft tissue structure: Hydraulic permeability reconstruction for magnetic resonance elastography in silico, *Frontiers in physics* 8 (2021) 617582.
- 400 [6] L. Tan, M. D. J. McGarry, E. E. W. Van Houten, M. Ji, L. Solamen, J. B. Weaver, K. D. Paulsen, Gradient-based optimization for poroelastic and viscoelastic mr elastography, *IEEE transactions on medical imaging* 36 (1) (2016) 236–250.

- [7] M. A. Sares, P. Morgan, V. J. Grauch, Geothermal resources of colorado and the potential for electrical power generation, in: AAPG Annual Convention and Exhibition, Denver, Colorado, 2009.
- 405 [8] L. Shevenell, M. Coolbaugh, R. Zehner, G. Johnson, W. Ehni, D. Noel, Geothermal exploration at emerson pass, pyramid paiute tribal lands, nevada, GRC Transactions 38 (2014) 29–38.
- [9] J. Renner, The future of geothermal energy, Tech. rep., Idaho National Laboratory (INL) (2006).
- [10] N. R. Council, Geological and geotechnical engineering in the new millennium: opportunities for research and technological innovation, National Academies Press, 2006.
- 410 [11] S. De Simone, C. Darcel, H. A. Kasani, D. Mas Ivars, P. Davy, The impact of size-dependent and stress-dependent fracture properties on the biot and skempton coefficients of fractured rocks, Rock Mechanics and Rock Engineering (2024) 1–22.
- [12] R. Muthupillai, D. J. Lomas, P. J. Rossman, J. F. Greenleaf, A. Manduca, R. L. Ehman, Magnetic resonance elastography by direct visualization of propagating acoustic strain waves, science 269 (5232) 415 (1995) 1854–1857.
- [13] P. R. Perriñez, F. E. Kennedy, E. E. W. Van Houten, J. B. Weaver, K. D. Paulsen, Magnetic resonance poroelastography: an algorithm for estimating the mechanical properties of fluid-saturated soft tissues, IEEE transactions on medical imaging 29 (3) (2010) 746–755.
- 420 [14] M. D. J. McGarry, C. L. Johnson, B. P. Sutton, J. G. Georgiadis, E. E. W. Van Houten, A. J. Pattison, J. B. Weaver, K. D. Paulsen, Suitability of poroelastic and viscoelastic mechanical models for high and low frequency mr elastography, Medical physics 42 (2) (2015) 947–957.
- [15] L. Tan, M. D. J. McGarry, E. E. W. Van Houten, M. Ji, L. Solamen, W. Zeng, J. B. Weaver, K. D. Paulsen, A numerical framework for interstitial fluid pressure imaging in poroelastic mre, Plos one 12 (6) (2017) e0178521.
- 425 [16] M. A. Biot, Theory of propagation of elastic waves in a fluid-saturated porous solid. i. low frequency range, The Journal of the acoustical Society of America 28 (2) (1956) 168–178.
- [17] M. A. Biot, Theory of propagation of elastic waves in a fluid-saturated porous solid. ii. higher frequency range, The Journal of the acoustical Society of America 28 (2) (1956) 179–191.
- 430 [18] M. A. Biot, Mechanics of deformation and acoustic propagation in porous media, Journal of applied physics 33 (4) (1962) 1482–1498.
- [19] T. Meyer, J. Castelein, J. Schattenfroh, A. S. Morr, R. V. da Silva, H. Tzschätzsch, R. Reiter, J. Guo, I. Sack, Magnetic resonance elastography in a nutshell: Tomographic imaging of soft tissue viscoelasticity for detecting and staging disease with a focus on inflammation, Progress in Nuclear Magnetic Resonance Spectroscopy (2024).
- 435 [20] F. Galarce, K. Tabelow, J. Polzehl, C. P. Papanikas, V. Vavourakis, L. Lilaj, I. Sack, A. Caiazzo, Displacement and pressure reconstruction from magnetic resonance elastography images: Application to an in silico brain model, SIAM Journal on Imaging Sciences 16 (2) (2023) 996–1027.
- 440 [21] M. Hosseini-Farid, M. Ramzanpour, J. McLean, M. Ziejewski, G. Karami, A poro-hyper-viscoelastic rate-dependent constitutive modeling for the analysis of brain tissues, Journal of the Mechanical Behavior of Biomedical Materials 102 (2020) 103475.
- [22] A. J. Pattison, M. McGarry, J. B. Weaver, K. D. Paulsen, Spatially-resolved hydraulic conductivity estimation via poroelastic magnetic resonance elastography, IEEE transactions on medical imaging 33 (6) (2014) 1373–1380.
- 445 [23] E. E. Konofagou, T. P. Harrigan, J. Ophir, T. A. Krouskop, Poroelastography: imaging the poroelastic properties of tissues, Ultrasound in medicine & biology 27 (10) (2001) 1387–1397.

- [24] G. P. Berry, J. C. Bamber, C. G. Armstrong, N. R. Miller, P. E. Barbone, Towards an acoustic model-based poroelastic imaging method: I. theoretical foundation, *Ultrasound in medicine & biology* 32 (4) (2006) 547–567.
- [25] X. Yang, F. Pourahmadian, J. Song, C. Wang, Deep learning for full-field ultrasonic characterization, *Mechanical Systems and Signal Processing* 201 (2023) 110668.
- [26] F. Pourahmadian, B. B. Guzina, On the elastic anatomy of heterogeneous fractures in rock, *International Journal of Rock Mechanics and Mining Sciences* 106 (2018) 259–268.
- [27] F. Pourahmadian, B. B. Guzina, H. Haddar, A synoptic approach to the seismic sensing of heterogeneous fractures: From geometric reconstruction to interfacial characterization, *Computer Methods in Applied Mechanics and Engineering* 324 (2017) 395–412.
- [28] F. Pourahmadian, K. Napal, Poroelastic near-field inverse scattering, *Journal of Computational Physics* 455 (2022) 111005.
- [29] X. Liu, J. Song, F. Pourahmadian, H. Haddar, Time-versus frequency-domain inverse elastic scattering: theory and experiment, *SIAM Journal on Applied Mathematics* 83 (3) (2023) 1296–1314.
- [30] K. Wapenaar, M. Dukalski, C. Reinicke, R. Snieder, Propagator and transfer matrices, marchenko focusing functions and their mutual relations, *Geophysical Journal International* 235 (2) (2023) 1403–1419.
- [31] K. Wapenaar, R. Snieder, S. de Ridder, E. Slob, Green’s function representations for marchenko imaging without up/down decomposition, *Geophysical Journal International* 227 (1) (2021) 184–203.
- [32] N. Boullé, C. J. Earls, A. Townsend, Data-driven discovery of green’s functions with human-understandable deep learning, *Scientific reports* 12 (1) (2022) 4824.
- [33] C. R. Gin, D. E. Shea, S. L. Brunton, J. N. Kutz, Deepgreen: deep learning of green’s functions for nonlinear boundary value problems, *Scientific reports* 11 (1) (2021) 21614.
- [34] H. Salavati, C. Debbaut, P. Pullens, W. Ceelen, Interstitial fluid pressure as an emerging biomarker in solid tumors, *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer* 1877 (5) (2022) 188792.
- [35] S. Walker-Samuel, T. A. Roberts, R. Ramasawmy, J. S. Burrell, S. P. Johnson, B. M. Siow, S. Richardson, M. R. Gonçalves, D. Pendse, S. P. Robinson, R. B. Pedley, M. F. Lythgoe, Investigating low-velocity fluid flow in tumors with convection-mri, *Cancer research* 78 (7) (2018) 1859–1872.
- [36] R. Caruana, Multitask learning, in: *Learning to learn*, Springer, 1998, pp. 95–133.
- [37] B. Bakker, T. Heskes, Task clustering and gating for bayesian multitask learning, *Journal of Machine Learning Research* 4 (May) (2003) 83–99.
- [38] I. Kokkinos, Ubernet: training a universal convolutional neural network for low-, mid-, and high-level vision using diverse datasets and limited memory, *arXiv preprint arXiv:1609.02132* (2016).
- [39] K. He, G. Gkioxari, P. Dollár, R. Girshick, Mask r-cnn, in: *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 2980–2988.
- [40] A. Kendall, Y. Gal, R. Cipolla, Multi-task learning using uncertainty to weigh losses for scene geometry and semantics, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 7482–7491.
- [41] Z. Chen, V. Badrinarayanan, C.-Y. Lee, A. Rabinovich, Gradnorm: gradient normalization for adaptive loss balancing in deep multitask networks, in: *Proceedings of the 35th International Conference on Machine Learning*, PMLR, 2018, pp. 794–803.
- [42] A. A. Heydari, C. A. Thompson, A. Mehmood, Softadapt: Techniques for adaptive loss weighting of neural networks with multi-part loss functions, *arXiv preprint arXiv:1912.12355* (2019).

- [43] D. Amini, E. Haghighat, R. Juanes, Inverse modeling of nonisothermal multiphase poromechanics using physics-informed neural networks, *Journal of Computational Physics* 490 (2023) 112323.
- 490 [44] A. C. Schmid, A. Doostan, F. Pourahmadian, Ensemble wsimdy for data-driven discovery of governing equations from laser-based full-field measurements, arXiv preprint arXiv:2409.20510 (2024).
- [45] B. Ding, A. H.-D. Cheng, Z. Chen, Fundamental solutions of poroelastodynamics in frequency domain based on wave decomposition, *Journal of Applied Mechanics* 80 (6) (2013) 061021.
- 495 [46] C. H. Yew, P. N. Jogi, Study of wave motions in fluid-saturated porous rocks, *The Journal of the Acoustical Society of America* 60 (1) (1976) 2–8.
- [47] K. Hornik, Approximation capabilities of multilayer feedforward networks, *Neural networks* 4 (2) (1991) 251–257.
- [48] F. Hecht, New development in freefem++, *Journal of Numerical Mathematics* 20 (3-4) (2012) 251–265. URL <https://freefem.org/>
- 500 [49] W. Aquino, M. Bonnet, Analysis of the error in constitutive equation approach for time-harmonic elasticity imaging, *SIAM Journal on Applied Mathematics* 79 (3) (2019) 822–849.
- [50] M. Bonnet, P. Salasiya, B. B. Guzina, Modified error-in-constitutive-relation (mecr) framework for the characterization of linear viscoelastic solids, *Journal of the Mechanics and Physics of Solids* (2024) 105746.
- 505 [51] C. Bellis, M. Trabelsi, F. Fremy, Reconstructing material properties by deconvolution of full-field measurement images: The conductivity case, *Inverse Problems* 33 (3) (2017) 035012.