



Heriot-Watt University
Research Gateway

Machine learning-based multiscale constitutive modelling: Development and application to dual-porosity mass transfer

Citation for published version:

Ashworth, M, Elsheikh, AH & Doster, F 2022, 'Machine learning-based multiscale constitutive modelling: Development and application to dual-porosity mass transfer', *Advances in Water Resources*, vol. 163, 104166. <https://doi.org/10.1016/j.advwatres.2022.104166>

Digital Object Identifier (DOI):

[10.1016/j.advwatres.2022.104166](https://doi.org/10.1016/j.advwatres.2022.104166)

Link:

[Link to publication record in Heriot-Watt Research Portal](#)

Document Version:

Publisher's PDF, also known as Version of record

Published In:

Advances in Water Resources

Publisher Rights Statement:

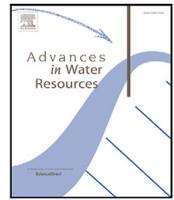
Crown Copyright © 2022.

General rights

Copyright for the publications made accessible via Heriot-Watt Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

Heriot-Watt University has made every reasonable effort to ensure that the content in Heriot-Watt Research Portal complies with UK legislation. If you believe that the public display of this file breaches copyright please contact open.access@hw.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Machine learning-based multiscale constitutive modelling: Development and application to dual-porosity mass transfer

Mark Ashworth^{*}, Ahmed H. Elsheikh, Florian Doster

Institute of GeoEnergy Engineering Heriot-Watt University, Edinburgh, United Kingdom

ARTICLE INFO

Keywords:

Multiscale constitutive modelling
Homogenisation
Dual-porosity
Machine learning for transient phenomena
Hybrid machine learning-physics-based modelling

ABSTRACT

In multiscale modelling, multiple models are used simultaneously to describe scale-dependent phenomena in a system of interest. Here we introduce a machine learning (ML)-based multiscale modelling framework for modelling hierarchical multiscale problems. In these problems, closure relations are required for the macroscopic problem in the form of constitutive relations. However, forming explicit closures for nonlinear and hysteretic processes remains challenging. Instead, we provide a framework for learning constitutive mappings given microscale data generated according to micro and macro transitions governed by two-scale homogenisation rules. The resulting data-driven model is then coupled to a macroscale simulator leading to a hybrid ML-physics-based modelling approach. Accordingly, we apply the multiscale framework within the context of transient phenomena in dual-porosity geomaterials. In these materials, the inter-porosity flow is a complex time-dependent function making its adoption within flow simulators challenging. We explore nonlinear feedforward autoregressive ML strategies for the constitutive modelling of this sequential problem. We demonstrate how to inject the resulting surrogate constitutive model into a simulator. We then compare the resulting hybrid approach to traditional dual-porosity and microscale models on a variety of tests. We show the hybrid approach to give high-quality results with respect to explicit microscale simulations without the computational burden of the latter. Lastly, the steps provided by the multiscale framework herein are sufficiently general to be applied to a variety of multiscale settings, using different data generation and learning techniques accordingly.

1. Introduction

When modelling we are often posed with the dilemma of choosing between models of varying levels of accuracy, complexity and efficiency. Such problems are exemplified in the subsurface. In these settings, strong heterogeneities exist within the rock fabric leading to multiscale phenomena. One solution is to model at the heterogeneity level (microscale), capturing the full range of scale-dependent phenomena with appropriate post-processing. However, at large scales, computational and information requirements preclude the use of such detailed simulation models. Instead, in situations where there is weak coupling between scales, a so-called separation of scales exists and we can use coarse representations. Accordingly, we have a hierarchy of models, where microscale information is incorporated implicitly within macroscale models in the form of constitutive relations. However, often these constitutive models are based on empiricism using simple considerations such as linearity (Weinan, 2011). Whilst practical, use of macroscopic models with simplified constitutive relations can lead

to undesirable inaccuracies. This deficiency motivates the use of multiscale modelling approaches. That is, to combine the detail of microscale models with the efficiency of macroscale models.

Here we are concerned with the use of multiscale methods with application to hierarchical multiscale problems. Generally, multiscale methods can be classified as analytical or computational (Weinan, 2011), although significant overlap may exist between the two. In situations where analytical methods are impractical or intractable, it is desirable instead to use computational approaches. One such computational approach, informed by developments in the latter part of the twentieth century, is computational homogenisation (Feyel, 1999; Kouznetsova et al., 2001; Miehe and Koch, 2002). In this approach, microscopic boundary value problems are nested within points in the macroscopic domain, with the solution of the former providing information for the latter. Notably, computational homogenisation has emerged as a powerful tool for modelling multiscale problems as diverse coupled multiphysics (Özdemir et al., 2008b; Su et al., 2011; Kaessmair and Steinmann, 2018) and transient diffusion problems (Larsson et al.,

^{*} Corresponding author.

E-mail addresses: ma174@hw.ac.uk (M. Ashworth), a.elsheikh@hw.ac.uk (A.H. Elsheikh), f.doster@hw.ac.uk (F. Doster).

2010; Ramos et al., 2017; Waseem et al., 2020) to name but a few. Nonetheless, whilst accurate, these approaches still remain expensive to use in practice (Wang and Sun, 2018). This expense motivates the use of so-called sequential methods. Here calculations are typically made offline to create a look-up table. The look-up table is then called online, using an appropriate interpolator where necessary. The downsides of the look-up table approach are we acquire little knowledge about the function, the quality of our dataset and how well we may generalise to samples not included in the table. Alternatively, we can learn a surrogate model offline, given our input-output data and other possible information. Learning such mappings is one of the fundamental goals of machine learning (ML). More generally, ML encompasses a range of algorithms to learn models from data, and as a discipline has seen widespread interest within the geoscience community to address a variety of modelling challenges (Razavi et al., 2012; Lary et al., 2016; Tahmasebi et al., 2020).

Machine learning approaches to generating surrogate constitutive models can be traced back several decades. For example, in the pioneering work of Ghaboussi et al. (1991) the authors used so-called neural networks to represent the constitutive behaviour of concrete under various loading conditions. From this and other similar early works (Goh, 1995; Ghaboussi et al., 1998), users have incorporated data-driven surrogate models within multiscale modelling strategies for a variety of applications (Unger and Könke, 2009; Hambli et al., 2011; Asproulis and Drikakis, 2013). More recently, work has been done for modelling time-dependent constitutive information using so-called recurrent neural networks (Wang and Sun, 2018; Ghavami and Simone, 2019; Chen, 2021). Despite the success of these recurrent algorithms for sequential modelling, they can be challenging to implement and to understand (Pascanu et al., 2013; Bai et al., 2018). Further, Bai et al. (2018) showed more recently autoregressive feedforward neural network architectures that give superior performance compared to recurrent structures for a wide variety of sequence modelling tasks. Given the variety of multiscale problems and machine learning methods, it would be useful to have a framework detailing the key questions and considerations for the generation and use of surrogate constitutive models within sequential multiscale approaches.

The aim and novelty of this work is to introduce and apply a machine learning-based multiscale framework for modelling complex constitutive behaviours such as those due to nonlinearity and hysteresis. Both of these phenomena are commonly observed in subsurface resource modelling. Our framework is broken down into four key components, with the starting point being a homogenisation procedure. Here we use a two-scale homogenisation approach such as those forming the basis for computational homogenisation methodologies. In this framework, the homogenisation procedure plays a fundamental role in generating consistent data to learn from. Further, the procedure also allows insight into the quantities involved in the constitutive mapping. With the scale transition rules from the homogenisation procedure in place, the remaining parts of the framework involve: Generating the data, learning the model and coupling the result to a physics-based simulator.

For our application we consider the diffusive mass transfer involved in dual-porosity (or dual-continuum more generally) flow modelling. In the dual-porosity (DP) setting, a strong contrast in conductivities exists between the different porosity levels. A classical example of such a material is naturally fractured rock. Here, the matrix has significantly lower permeability than the fractures. As a result, the matrix introduces a history-dependent memory effect associated with the mass exchange between the two porosity levels. Accordingly, in the presence of fracture dynamics the inter-porosity exchange is non-local in time, requiring a convolutional product for its description. Further, even for cases involving static fracture pressures, challenges still remain in describing inter-porosity exchange. For these latter cases, and for simple geometries, analytical solutions for the inter-porosity exchange term are possible. However, solutions for these static fracture pressure cases

are still explicit in time, and are formed from a complex infinite series. Accordingly, describing inter-porosity flow for both static and dynamic fracture behaviours remains difficult with traditional simulation methods. These challenges necessitate simpler descriptions for inter-porosity exchange, leading to the commonly used linear description involving the difference between averaged pressures (Warren and Root, 1963). However, this simplification is known to lead to measurable errors in flow behaviour, particularly at early-time (Zimmerman et al., 1993; Ashworth and Doster, 2020). Instead of making any such assumptions, we use a data-driven constitutive model learnt on time series data coming directly from the microscale. In line with the aims of this work to introduce and apply our framework, we consider only cases of static fracture pressure. Accordingly, we can give an overview of the framework on a problem without significant complexity, using easy-to-generate data and established learning algorithms. We subsequently compare various nonlinear feedforward autoregressive algorithms approaches to model the time-dependent problem. The successful model is then coupled to a dual-porosity simulator in the MATLAB Reservoir Simulation Toolbox (MRST) (Lie et al., 2012; Lie, 2019). Lastly, in the context of this work, it is worth mentioning the interesting work done by Andrianov and Nick (2021). For this, the authors use ML to learn the mapping between images of pixelated fracture geometries and DP constitutive parameters. In contrast, taking inspiration from computational homogenisation approaches, our goal here is to provide a framework to learn and apply a constitutive mapping itself.

We structure the remainder of this paper as follows: In Section 2, we introduce the framework, detailing the key elements and considerations therein. In Section 3, we start the application of the framework through the two-scale homogenisation approach to derive the macroscopic dual-porosity model. In Section 4, we introduce a dual-porosity model problem that will serve as the testbed for the remaining parts of the multiscale framework. Further, given this model problem we derive an often-used linear transfer model. In Section 5, we generate the dataset for learning. To do so, we use analytical solutions for the given model problem, whilst also considering the end use of the ML model in a numerical simulator. In Section 6, we learn our surrogate constitutive model, comparing a variety of learning algorithms for modelling the time-dependent problem. In Section 7, we couple the resulting ML model to a physics-based simulator, leading to a hybrid ML-physics approach. We then test the resulting hybrid approach against a dual-porosity model equipped with the traditionally used linear transfer model and microscale simulations. In Section 8, we finish with conclusions and recommendations for future work. Lastly, for reproducibility, the codes used in this work are available as open-source.¹

2. Machine learning-based constitutive modelling framework

Here we introduce the ML-based multiscale constitutive modelling framework and the key concepts therein. The framework itself can be broken into four key components: Homogenisation, data generation, surrogate constitutive model learning, and ML and physics model coupling (Fig. 1). However, we stress that these components are interrelated and should be understood in parallel as opposed to isolated procedures.

2.1. Homogenisation

Homogenisation is concerned with the question: How do we link between micro and macroscales, providing constitutive information from the former to the latter? Various approaches exist for linking between scales such as asymptotic homogenisation (Sánchez-Palencia, 1980; Auriault et al., 2010), the method of volume averaging (Quintard

¹ <https://github.com/mashworth11/ML-MM>

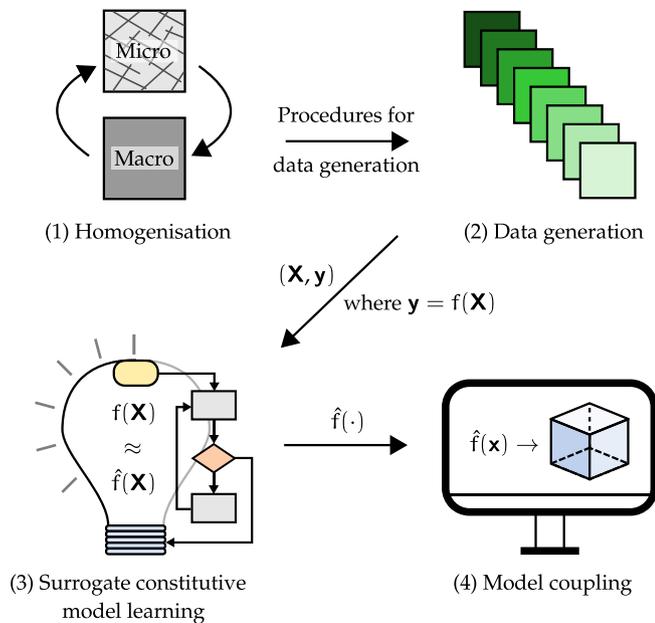


Fig. 1. The machine learning-based multiscale constitutive modelling framework.

and Whitaker, 1993; Whitaker, 1999), and computational homogenisation approaches (Feyel, 1999; Kouznetsova et al., 2001). Depending on underlying requirements (e.g. a-priori knowledge for scaling parameters in the case of asymptotic approaches) and resources, some upscaling strategies may be more suitable than others. Regardless, within the context of this framework, the homogenisation procedure provides the guide for generating consistent *offline* data for training and validating a machine learning model. We note, one could consider the homogenisation step as part of the data generation process. However, we choose to distinguish these two steps within the current framework to highlight their respective importance and considerations therein.

2.2. Data generation

Data generation is concerned with the question: How will we generate sample data for our learning algorithm? Microscale results can be expensive when numerically derived due to limited computational resources and/or high problem dimensionality. Managing this trade-off to ensure the dataset (sample set) is as informative as possible is referred to as experimental design (or design of experiments) (Razavi et al., 2012). Traditional approaches to experimental design involve the use of various sampling strategies (e.g. Monte Carlo, Latin hypercube, quasi-Monte Carlo methods) to build the dataset prior to learning. However, these so-called one-shot strategies can either contain too few samples for an accurate surrogate (undersampling) or more samples than we need (oversampling), thus wasting precious resources (Xiao et al., 2018). An alternative to one-shot approaches is to use feedback from the ML training process to find samples that provide the maximum information gain when labelled and used to update the model. These so-called active learning (or sequential sampling) approaches can address the under/oversampling trade-off described previously (Crombecq, 2011).

In addition to data-driven ML approaches, recent advances have seen the development of theory-driven ML approaches. Here the idea is to use physics such as partial differential equations, initial and boundary condition information, symmetry or invariance properties, and thermodynamical considerations (Ling et al., 2016; Wang and Sun, 2018; Raissi et al., 2019; Zhu et al., 2019; Masi et al., 2021), to constrain the learning process. Accordingly, we can supplement missing or difficult to obtain data. Further, theory-driven approaches also act

to regularise the learning process and ensure the physical consistency of our model. However, in this work we consider purely data-driven methods and save the incorporation of theory into the learning process for future studies.

2.3. Surrogate constitutive model learning

Surrogate constitutive model learning is concerned with the question: Given the properties of our data, the type of prediction problem and the end use for the resulting ML model, what is the most appropriate learning strategy to use? Training ML algorithms can be expensive from both data and computational points of view. If our sample set is poor, complex algorithms run the risk of failing to generalise when given new data (overfitting). Conversely, if our model is too simple we risk failing to extract meaningful relationships from our data (underfitting). Besides addressing underfitting and overfitting, we may also want algorithms that are intrinsically interpretable and quick to retrain in light of new data (e.g. linear regression, tree-based methods); give us an insight into aleatoric (data) and/or epistemic (model) uncertainties (e.g. Gaussian processes, Bayesian learning); and provide an embedded approach to feature (input) engineering in light of complex and/or high dimensional data (e.g. deep learning). Ultimately, within the context of multiscale modelling, the choice of learning method should be guided by considering the interdependence between data quality and availability, ML model accuracy and practicality, and the model's end use in physics-based simulation.

2.4. Model coupling

Model coupling is concerned with the question: How will we use our learnt constitutive model in our macroscale physics simulator? Ideally, the ML model can be implemented within a numerical solver without significant intrusion. Accordingly, important considerations at this stage may be the accessibility of data during simulation to pass to the ML model; the ML model inference time; how predictions will be fed into the numerical model; and in the case of nonlinear materials, how we extract state-dependent material properties such as permeability and stiffness from our surrogate constitutive model. These latter points will depend on whether the constitutive model is to be used in an explicit or implicit manner. We describe the former, as those predictions made using previously observed and converged quantities. Subsequently, in a nonlinear solver the resulting surrogate constitutive model will only appear in the residual. Alternatively, if the prediction also depends on quantities at the current iteration level, parameters from the surrogate associated with these quantities will be needed in the Jacobian. For example, in Hashash et al. (2004) the authors consider nonlinear deformation problems and derive a closed-form expression for calculating the tangent stiffness tensor from a neural network. Accordingly, the authors provide consistent Jacobian entries for use in a Newton–Raphson scheme. For this work, our surrogate will be explicitly coupled due to the nature of the constitutive relation being injected. Nonetheless, the example above highlights the type of considerations we need to make in order to embed ML models within traditional physics-based simulations.

3. Application of the framework: two-scale homogenisation for dual-porosity flow problems

With the details of the framework in-hand we now progress to its application. For our application we consider problems associated with dual-porosity flow, such as those observed in multiscale geomaterials. Accordingly, in this section we derive the dual-porosity flow model by way of a two-scale homogenisation approach developed for transient diffusion problems (Larsson et al., 2010; Ramos et al., 2017; Brassart and Stainier, 2019; Waseem et al., 2020).

3.1. Preliminaries

The homogenisation procedure relies on the definition of representative elementary volumes Ω (REV) taken from a macroscopic domain $\bar{\Omega}$. Accordingly, we are interested in quantities varying at two distinct scales: Those variations at points within an REV $\mathbf{x} \in \Omega$, and those variations at material points within the macroscopic body $\bar{\mathbf{x}} \in \bar{\Omega}$. Note, quantities written with $\bar{\cdot}$ denote macroscopic quantities.

The REV is defined according to a separation of scales in terms of geometrical information and the wavelengths of physical processes involved. Formally, the separation of scales is defined as

$$\ell \ll l \ll L, \quad (1)$$

where ℓ , l and L denote the characteristic lengths of the heterogeneities, REV and macroscopic body respectively.

Our reference material is naturally fractured rock. Accordingly, the material is composed of a low permeability matrix (denoted by subscript 1) and a high permeability fracture material (denoted by subscript 2). We assume homogeneity and isotropy of each material as well as isotropic geometrical distributions at the microscale.

Next, we introduce the following volume averaging relations

$$\langle z \rangle = \frac{1}{|\Omega|} \int_{\Omega} z dV, \quad \langle z \rangle_{\alpha} = \frac{1}{|\Omega_{\alpha}|} \int_{\Omega_{\alpha}} z dV, \quad (2)$$

where z is an arbitrary field quantity. Accordingly, $\langle \cdot \rangle$ denotes the volume average over the whole REV domain, whilst $\langle \cdot \rangle_{\alpha}$ denotes the volume average over the constituent volume $\Omega_{\alpha} \subset \Omega$. The two volume averages defined in Eq. (2) are related such that

$$\langle z \rangle = v_1 \langle z \rangle_1 + v_2 \langle z \rangle_2, \quad (3)$$

where $v_{\alpha} = |\Omega_{\alpha}|/|\Omega|$ is the volume fraction of phase α .

Lastly, we define the following relations between rates of change and their averages, such that

$$\langle \dot{z} \rangle = \left\langle \frac{\partial z}{\partial t} \right\rangle = \frac{\partial}{\partial t} \langle z \rangle = \langle \dot{z} \rangle, \quad (4)$$

3.2. Mass balance at the macro and microscales

The dual-porosity model is a multi-continuum model at the macroscale described by the continuum-wise mass balance as

$$\dot{\bar{m}}_1 + \nabla \cdot \bar{\mathbf{w}}_1 = \bar{r}_1, \quad (5)$$

$$\dot{\bar{m}}_2 + \nabla \cdot \bar{\mathbf{w}}_2 = \bar{r}_2, \quad (6)$$

where $\dot{\bar{m}}_{\alpha}$, $\bar{\mathbf{w}}_{\alpha}$ and \bar{r}_{α} denote the rate of macroscopic mass content change, mass flux vector, and inter-continuum rate of mass transfer for continuum α respectively. Further, mass conservation requires $\bar{r}_2 = -\bar{r}_1$.

In addition to appropriate boundary and initial conditions, we require constitutive relations for the macroscopic quantities \bar{m}_{α} , $\bar{\mathbf{w}}_{\alpha}$ and \bar{r}_{α} . The simplest approach to specify these closures is through phenomenological arguments. Alternatively, we can derive closure relations through homogenisation. In this regard, asymptotic homogenisation has been successful in deriving a closed dual-porosity model starting at the microscale (Royer and Auriault, 1994). These results thus serve as a useful reference to the approach introduced herein.

At the microscale we have the following mass balance

$$\dot{m} + \nabla \cdot \mathbf{w} = 0, \quad (7)$$

where m and \mathbf{w} denote the rate of microscopic mass content change and mass flux vector respectively. Additionally, at this scale we assume both materials are saturated by the same slightly compressible fluid of density ρ . Under isothermal flows we define the fluid compressibility relation

$$c_l = \frac{1}{\rho} \frac{\partial \rho}{\partial p}, \quad (8)$$

where p denotes the microscopic pressure. A similar compressibility relation is defined for the local porosity $\phi = |\Omega^p|/|\Omega|$, where Ω^p is the pore volume, such that

$$c_{\phi} = \frac{1}{\phi} \frac{\partial \phi}{\partial p}. \quad (9)$$

We assume small perturbations in fluid density and porosity allowing linearised evolutions of these quantities. Further, the small perturbation hypothesis permits the replacement of ρ and ϕ with ρ^0 and ϕ^0 respectively, where necessary. Notations with superscript 0 denote reference conditions. From $m = \rho\phi$ and $\mathbf{w} = \rho\mathbf{q}$, where \mathbf{q} is the microscopic volume flux, Eq. (7) becomes

$$\rho^0 \phi^0 c_p \dot{p} + \nabla \cdot (\rho^0 \mathbf{q}) = 0, \quad (10)$$

where we have made use of the small perturbation hypotheses and where $c = c_l + c_{\phi}$. Lastly, the microscopic volume flux is related to the microscopic pressure gradient through Darcy's law (neglecting gravitational effects)

$$\mathbf{q} = \frac{\mathbf{k}}{\mu} \cdot \nabla p, \quad (11)$$

where μ is the fluid viscosity. For the isotropic case $\mathbf{k} = kI$, in which I denotes the identity matrix. To link the macroscopic description in Eqs. (5) to (6) to the microscale description in Eq. (7) we make use of a two-scale homogenisation approach. This approach is split into two steps:

Downscaling: Involves translation of macroscopic loading parameters (macroscopic pressures and gradients) to consistent constraints and boundary conditions at the microscale.

Upscaling: Involves transformation of microscopic rates of change and fluxes to their macroscopic counterparts through averaging.

We address the downscaling step in the section to follow.

3.3. Downscaling

For a dual-porosity material, associated to a macroscopic point $\bar{\mathbf{x}}$ we have two macroscopic pressures \bar{p}_1 , \bar{p}_2 and gradients $\bar{\nabla} \bar{p}_1$, $\bar{\nabla} \bar{p}_2$. For each constituent we then define microscopic pressure fields according to Taylor series expansions of the first order about the macroscopic point

$$p(\bar{\mathbf{x}}, \mathbf{x}, t) = \bar{p}_1(\bar{\mathbf{x}}, t) + \bar{\nabla} \bar{p}_1(\bar{\mathbf{x}}, t) \cdot (\mathbf{x} - \bar{\mathbf{x}}) + p'(\bar{\mathbf{x}}, \mathbf{x}, t) \quad \text{in } \Omega_1, \quad (12)$$

$$p(\bar{\mathbf{x}}, \mathbf{x}, t) = \bar{p}_2(\bar{\mathbf{x}}, t) + \bar{\nabla} \bar{p}_2(\bar{\mathbf{x}}, t) \cdot (\mathbf{x} - \bar{\mathbf{x}}) + p'(\bar{\mathbf{x}}, \mathbf{x}, t) \quad \text{in } \Omega_2, \quad (13)$$

where p' denote higher order expansion terms corresponding to pressure fluctuations at the microscale. We note, the microscale is positioned relative to the macroscopic point such that $\langle \mathbf{x} - \bar{\mathbf{x}} \rangle = 0$, which also holds for constituent averages.

From Eqs. (12) to (13), we require the following compatibilities to hold

$$\langle p \rangle_1 = \bar{p}_1, \quad (14)$$

$$\langle p \rangle_2 = \bar{p}_2, \quad (15)$$

corresponding to the requirement $\langle p' \rangle_{\alpha} = 0$. From, Özdemir et al. (2008a), Eqs. (14) to (15) are related to requiring consistency between the microscopic and macroscopic mass contents such that

$$\bar{\rho} \bar{\phi}_{\alpha} \bar{c}_{\alpha} \bar{p}_{\alpha} = \langle \rho \phi c p_{\alpha} \rangle, \quad (16)$$

where $\bar{\phi}_{\alpha} = |\Omega_{\alpha}^p|/|\Omega| = v_{\alpha} \phi_{\alpha}$ is the effective continuum porosity. From Eq. (16), we propose the following consistency conditions for the mass storage

$$\bar{\rho} \bar{\phi}_{\alpha} \bar{c}_{\alpha} = \langle \rho \phi c \rangle, \quad (17)$$

which is inline with results obtained through asymptotic homogenisation (Auriault, 1983; Royer and Auriault, 1994).

In addition to constraints provide by Eqs. (14) to (15) we require averages of microscopic gradients to equal their macroscopic counterparts. Accordingly, taking the gradients of Eqs. (12) to (13) and averaging over each constituent leads to

$$\langle \nabla p \rangle_1 = \overline{\nabla p}_1 + \langle \nabla p' \rangle_1 \quad \text{in } \Omega_1, \quad (18)$$

$$\langle \nabla p \rangle_2 = \overline{\nabla p}_2 + \langle \nabla p' \rangle_2 \quad \text{in } \Omega_2, \quad (19)$$

where we use the following relation $\nabla(\mathbf{x} - \bar{\mathbf{x}}) = \mathbf{I}$. To fulfil the compatibility between gradients we require $\langle \nabla p' \rangle_\alpha = \mathbf{0}$ for both constituents. Application of Gauss' divergence theorem on averages of the fluctuation gradients leads to

$$\int_{\partial\Omega_\alpha} p' \mathbf{n} \, dS = \mathbf{0}, \quad (20)$$

where $\partial\Omega_\alpha$ is the boundary of Ω_α and \mathbf{n} is the outward unit normal vector. Condition Eq. (20) is satisfied by setting certain restrictions (boundary conditions) on the fluctuations at the boundaries. Examples of such boundary conditions include periodic boundary conditions and prescribed fluctuation boundary conditions. In this work we use the latter for demonstrative purposes, although the former is often used for flow-based upscaling procedures (Lie, 2019). Accordingly, prescribed fluctuation boundary conditions correspond to setting $p' = 0$ leading to

$$p(\bar{\mathbf{x}}, \mathbf{x}, t) = \bar{p}_\alpha(\bar{\mathbf{x}}, t) + \overline{\nabla p}_\alpha(\bar{\mathbf{x}}, t) \cdot (\mathbf{x} - \bar{\mathbf{x}}) \quad \text{on } \partial\Omega_\alpha, \quad (21)$$

from either Eq. (12) or Eq. (13). In the dual-porosity formulation we assume the matrix material to be completely permeated by the fracture material. Additionally, we assume continuity of the microscopic pressure field at the interface $\Gamma = \partial\Omega_1$ between the two materials. Accordingly, boundary conditions for both materials can be completely defined by considering the prescribed fluctuation boundary conditions for the fracture material

$$p(\bar{\mathbf{x}}, \mathbf{x}, t) = \bar{p}_2(\bar{\mathbf{x}}, t) + \overline{\nabla p}_2(\bar{\mathbf{x}}, t) \cdot (\mathbf{x} - \bar{\mathbf{x}}) \quad \text{on } \partial\Omega_2, \quad (22)$$

We note, Eq. (22) is somewhat similar to the boundary conditions specified by the asymptotic homogenisation approach on Γ , where $p = \bar{p}_2$ on Γ (Royer and Auriault, 1994). However, here the explicit size of the REV and macroscopic fracture pressure gradient is taken into account. Nonetheless, for small macroscopic gradients and vanishingly small REV sizes $p = \bar{p}_2$ on Γ is a reasonable approximation. With the required downscaling relations we can now proceed to upscaling.

3.4. Upscaling

Upscaling for the macroscopic quantities $(\dot{\bar{m}}_\alpha, \bar{\mathbf{w}}_\alpha)$ is achieved through a virtual power equivalence relation between the micro and macroscales. This relation is referred to as the generalised Hill–Mandel condition (Blanco et al., 2016), and requires the macroscopic virtual power to equal its volume averaged microscopic counterpart at a given material point. Here we introduce a virtual power relation for the dual-porosity material by summing the external powers coming from Eqs. (5) to (6) and using $\delta p_1 \bar{r}_1 + \delta p_2 \bar{r}_2 = 0$ to give

$$\delta \bar{p}_1 \dot{\bar{m}}_1 - \overline{\nabla \delta p}_1 \cdot \bar{\mathbf{w}}_1 + \delta \bar{p}_2 \dot{\bar{m}}_2 - \overline{\nabla \delta p}_2 \cdot \bar{\mathbf{w}}_2 = \langle \delta p \dot{m} - \nabla \delta p \cdot \mathbf{w} \rangle, \quad (23)$$

where we use Eq. (7) at the microscale for compactness. Use of the decomposition in Eq. (3) on Eq. (23) followed by substitution of the appropriate expansions in Eqs. (12) to (13) leads to

$$\begin{aligned} \sum_{\alpha=1,2} \delta \bar{p}_\alpha \dot{\bar{m}}_\alpha - \overline{\nabla \delta p}_\alpha \cdot \bar{\mathbf{w}}_\alpha &= v_1 \langle \delta \bar{p}_1 + \overline{\nabla \delta p}_1 \cdot (\mathbf{x} - \bar{\mathbf{x}}) + \delta p' \dot{m} \\ &\quad - [\nabla \delta \bar{p}_1 + \nabla \delta p'] \cdot \mathbf{w} \rangle_1 \\ &\quad + v_2 \langle \delta \bar{p}_2 + \overline{\nabla \delta p}_2 \cdot (\mathbf{x} - \bar{\mathbf{x}}) + \delta p' \dot{m} \\ &\quad - [\nabla \delta \bar{p}_2 + \nabla \delta p'] \cdot \mathbf{w} \rangle_2. \end{aligned} \quad (24)$$

Rearranging Eq. (24) to isolate terms involving fluctuations gives

$$\begin{aligned} \sum_{\alpha=1,2} \delta \bar{p}_\alpha \dot{\bar{m}}_\alpha - \overline{\nabla \delta p}_\alpha \cdot \bar{\mathbf{w}}_\alpha &= v_1 \langle \delta \bar{p}_1 \dot{m} - \overline{\nabla \delta p}_1 \cdot [\mathbf{w} + (\mathbf{x} - \bar{\mathbf{x}}) \dot{m}] \rangle_1 \\ &\quad + v_2 \langle \delta \bar{p}_2 \dot{m} - \overline{\nabla \delta p}_2 \cdot [\mathbf{w} + (\mathbf{x} - \bar{\mathbf{x}}) \dot{m}] \rangle_2 \\ &\quad + v_1 \langle \delta p' \dot{m} - \nabla \delta p' \cdot \mathbf{w} \rangle_1 \\ &\quad + v_2 \langle \delta p' \dot{m} - \nabla \delta p' \cdot \mathbf{w} \rangle_2. \end{aligned} \quad (25)$$

The terms in Eq. (25) involving the pressure fluctuations represents the weak form of the phase-wise microfluctuation balance of mass. These terms are subsequently addressed through the chain rule and the divergence theorem leading to

$$\begin{aligned} v_1 \langle \delta p' \dot{m} - \nabla \delta p' \cdot \mathbf{w} \rangle_1 + v_2 \langle \delta p' \dot{m} - \nabla \delta p' \cdot \mathbf{w} \rangle_2 \\ = \langle \delta p' (\dot{m} + \nabla \cdot \mathbf{w}) \rangle - \int_{\partial\Omega_1} p' \mathbf{n} \, dS - \int_{\partial\Omega_2} p' \mathbf{n} \, dS. \end{aligned} \quad (26)$$

Accordingly, the first term on the right-hand side of Eq. (26) is satisfied due to the microscopic conservation of mass Eq. (7). The remaining terms then vanish with the appropriate boundary conditions, in this case the prescribed fluctuation boundary conditions.

Having dealt with the fluctuation terms we seek to obtain a dual-porosity formulation from the remaining terms in Eq. (25). This goal is pursued in the following section.

3.5. Recovery of the dual-porosity model

To recover a dual-porosity model from the virtual power equivalence we assume quasi-steady state in the fracture material due to the high phase permeability. Accordingly,

$$\nabla \cdot (\mathbf{w}) = \rho^0 \nabla \cdot (\mathbf{q}) = 0 \quad \text{in } \Omega_2, \quad (27)$$

whilst transient diffusion effects remain in the matrix material at the microscale. However, due to the percolating fractures and the low intrinsic matrix permeability we assume $\langle \mathbf{w} \rangle_1 \approx 0$. With the following assumptions at the microscale we recover from Eq. (25)

$$\begin{aligned} \delta \bar{p}_1 \dot{\bar{m}}_1 - \overline{\nabla \delta p}_1 \cdot \bar{\mathbf{w}}_1 + \delta \bar{p}_2 \dot{\bar{m}}_2 - \overline{\nabla \delta p}_2 \cdot \bar{\mathbf{w}}_2 &= \delta \bar{p}_1 \langle \dot{m} \rangle_1 v_1 \\ &\quad - \overline{\nabla \delta p}_1 \cdot \langle (\mathbf{x} - \bar{\mathbf{x}}) \dot{m} \rangle_1 v_1 \\ &\quad - \overline{\nabla \delta p}_2 \cdot \langle \mathbf{w} \rangle_2 v_2. \end{aligned} \quad (28)$$

Comparing terms in Eq. (28) we identify the following relations

$$\dot{\bar{m}}_1 = v_1 \langle \dot{m} \rangle_1 = \rho^0 \overline{\phi}_1^0 c_1 \langle \dot{p} \rangle_1, \quad (29)$$

$$\bar{\mathbf{w}}_1 = v_1 \langle (\mathbf{x} - \bar{\mathbf{x}}) \dot{m} \rangle_1, \quad (30)$$

$$\bar{\mathbf{w}}_2 = v_2 \langle \mathbf{w} \rangle_2 = \rho^0 v_2 \langle \mathbf{q} \rangle_2, \quad (31)$$

where fluid density at reference conditions at the microscale corresponds to the equivalent quantity at the macroscale i.e. $\bar{\rho}^0 = \rho^0$. Further, for the small perturbation considerations used here $\bar{c}_\alpha = c_\alpha$.

Looking first at the averaged quantity in Eq. (30), we identify this as the first moment of the rate of pressure change in the matrix. Interestingly, this term is not present in the classical asymptotic homogenisation approach (Royer and Auriault, 1994). Physically, this term introduces a size-effect and represents the microscale mass inertia due to the low permeability of the matrix material. However, in general this inertial term is shown to be negligible with respect to the macroscopic flux from the fractures (Brassart and Stainier, 2019). Further, for isotropically distributed materials at the microscale this term is shown to vanish (Brassart and Stainier, 2019).

From Eq. (31) we can identify the macroscopic volumetric fracture flux as

$$\bar{\mathbf{q}}_2 = v_2 \langle \mathbf{q} \rangle_2 \approx \langle \mathbf{q} \rangle, \quad (32)$$

where the relation on the right-hand side of Eq. (32) arises due to the assumption $v_1(\mathbf{q})_1 \approx 0$ and from Eq. (3). Next, Using Eq. (11) in Eq. (32) we write

$$\bar{q}_2 = \frac{\bar{k}_2}{\mu} \cdot \bar{\nabla} \bar{p}_2, \quad (33)$$

where \bar{k}_2 denotes the (macroscopic) effective fracture permeability tensor. Accordingly, we can recover entries to this permeability tensor provided we can calculate the macroscopic fracture flux corresponding to the macroscopic pressure gradient. Such calculations are addressed in so-called flow-based upscaling procedures (Lie, 2019).

Finally, Eq. (29) shows the relation between the micro and macroscopic rate of mass content change. From Eq. (4) we see

$$\rho^0 \bar{\phi}_1^0 c_1 \langle \dot{p} \rangle_1 = \rho^0 \bar{\phi}_1^0 c_1 \langle \dot{p} \rangle_1. \quad (34)$$

Accordingly, with Eq. (29) (resp. Eq. (34)) and Eq. (31) in Eqs. (5) to (6), whilst neglecting the inertial effects from Eq. (30), we recover the dual-porosity model from the two-scale homogenisation procedure as

$$\rho^0 \bar{\phi}_1^0 c_1 \langle \dot{p} \rangle_1 = \bar{r}_1, \quad (35)$$

$$\rho^0 \bar{\phi}_2^0 c_2 \langle \dot{p} \rangle_2 + \rho^0 \bar{\nabla} \cdot (v_2 \langle \mathbf{q} \rangle_2) = -\bar{r}_1, \quad (36)$$

where we use the compatibility in Eq. (16) for the fracture rate of mass content change \bar{m}_2 .

In this work our focus is in the modelling of the inter-porosity transfer term \bar{r}_1 . From Eq. (35) we see $\bar{r}_1 = \rho^0 \bar{\phi}_1^0 c_1 \langle \dot{p} \rangle_1$. Subsequently, to compute \bar{r}_1 all we need is the solution to the microscale boundary value problem Eq. (10) in the matrix, given boundary conditions from Eq. (22) on Γ and some initial condition $p_1(\bar{\mathbf{x}}, \mathbf{x}, t = 0) = p_1^0$. However, often we not have direct access to the microscale within a macroscopic simulation. Further, even if the solution could be calculated directly a-priori, it is difficult to use in practice owing to its complex form. Specifically, for evolving fracture pressures, the matrix pressure solution is defined as a convolution product over the fracture pressure history (Royer and Auriault, 1994). Even for static fracture pressures following a step increase, the matrix pressure solution is still an infinite series defined in explicit time (shown in the following section). Due to the described complexities, it is common to adopt a simple linear relation to model \bar{r}_1 . Such simplification leads to measurable errors in flow behaviour. Accordingly, in the remainder of this work we aim to improve on linear transfer models through machine learnt constitutive models which incorporate temporal information using quantities directly available during simulation.

4. A dual-porosity model problem and recovery of a first-order inter-porosity mass transfer model

With the homogenisation step in-hand we now have a guideline for how to generate data to learn from. Accordingly, we proceed to the remaining steps of the multiscale framework. To do so, we first consider a model problem for which data can be generated using analytical solutions. Further, from these solutions we can obtain a first-order mass transfer model that is commonly used in dual-continuum flow modelling. This simplified transfer model will then serve as a reference for testing any data-driven model coupled to a numerical simulator. We note that whilst the problem considered here is not too complex, it serves as a useful demonstration for the remaining steps of the framework and to explore various established machine learning algorithms for modelling the time-dependent mass transfer problem. Future directions therefore concern numerically derived data involving fracture dynamics, more complicated microscale geometries, and more advanced methods for surrogate modelling (e.g. active learning, physics-informed models etc.). In this latter element, work combining the data-driven parameterisations from Andrianov and Nick (2021) with elements of the framework introduced herein could be an interesting path for further research.

4.1. Model problem

For the model problem we consider a two-dimensional matrix domain surrounded by fractures (Fig. 2(a)). At a time t^0 the matrix and fractures are in hydrostatic equilibrium. Then, following a jump in fracture pressure at a time t^{0+} , the fracture pressure remains fixed leading to $\bar{\nabla} \bar{p}_2 = 0$ (Fig. 2(b)). Accordingly, from Eq. (21) and the pressure continuity across the interface we have

$$p(\bar{\mathbf{x}}, \mathbf{x}, t) = \bar{p}_2(\bar{\mathbf{x}}, t) \quad \text{on } \partial \Omega_1. \quad (37)$$

with initial conditions

$$p(\bar{\mathbf{x}}, \mathbf{x}, t = 0) = p_1^0(\bar{\mathbf{x}}, \mathbf{x}) \quad \text{in } \Omega_1. \quad (38)$$

Due to the considered (fracture pressure) boundary conditions and simple geometry, analytical solutions exist for calculating the solution. These solutions are introduced subsequently.

4.2. Analytical solution and first-order mass transfer model

The analytical solution to the 2D matrix diffusion problem described above is found by taking the product of the (dimensionless) solution to a 1D diffusion problem such as those arising in slab geometries (Crank, 1979; Holman, 1990; Lim and Aziz, 1995). Averaging the solution over the REV, and noting the equivalence with macroscopic quantities, then leads to

$$\bar{p}_1 = p_1^0 + (\bar{p}_2 - p_1^0) \left\{ 1 - \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} \left(\frac{8}{\pi^2} \right)^2 \frac{1}{(2n+1)^2(2m+1)^2} \exp \left(- \frac{\pi^2 k_1 t}{\phi_1^0 c_1 \mu \ell^2} [(2n+1)^2 + (2m+1)^2] \right) \right\}, \quad (39)$$

where we use $\bar{p}_1 = \langle p \rangle_1$. Next, taking the time derivative of Eq. (39) gives

$$\dot{\bar{p}}_1 = (\bar{p}_2 - p_1^0) \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} \left(\frac{8}{\pi^2} \right)^2 \frac{\pi^2 k_1}{\phi_1^0 c_1 \mu \ell^2} \left[\frac{1}{(2n+1)^2} + \frac{1}{(2m+1)^2} \right] \exp \left(- \frac{\pi^2 k_1 t}{\phi_1^0 c_1 \mu \ell^2} [(2n+1)^2 + (2m+1)^2] \right). \quad (40)$$

Substitution of Eq. (40) into Eq. (35) gives \bar{r}_1 . However, the resulting mass transfer expression is unsuitable for simulation purposes given the presence of explicit time and the infinite series. To alleviate these dependencies we can recover a first-order approximation to Eq. (40). Accordingly, taking the first term from Eq. (40) and eliminating t using the first term from Eq. (39) gives

$$\dot{\bar{p}}_1 = \frac{2\pi^2 k_1}{\phi_1^0 c_1 \mu \ell^2} (\bar{p}_2 - \bar{p}_1). \quad (41)$$

Substitution of Eq. (41) into Eq. (35) leads to the following first-order inter-porosity transfer model (Zimmerman et al., 1993; Lim and Aziz, 1995)

$$\bar{r}_1 = \rho^0 \frac{v_1 \eta k_1}{\mu} (\bar{p}_2 - \bar{p}_1), \quad (42)$$

where $\eta = (N_f \pi^2) / \ell^2$ denotes the so-called shape factor in which N_f is the number of characteristic fracture sets.

Several investigations have shown use of the linear constitutive model in Eq. (42) to lead to measurable errors in flow behaviour (Zimmerman et al., 1993; Ashworth and Doster, 2020). One common way to alleviate these inaccuracies is to use multi-rate transfer models (Haggerty and Gorelick, 1995; Geiger et al., 2013). However, in this work we treat the problem differently. Instead, we use machine learning to bridge the scales in the form of a surrogate constitutive model.

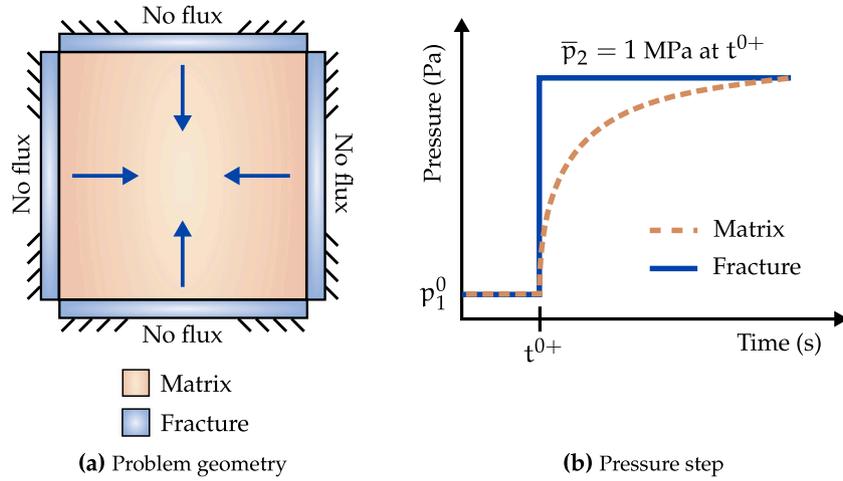


Fig. 2. Model problem setup: Geometry (a), and fracture pressure step (b).

5. Application of the framework: Data generation

Our goal for the supervised learning problem is to learn a mapping from quantities available during simulation to an output aligned with our constitutive property. To do so we consider a dataset \mathcal{D} of n samples, $\mathcal{D} = \{(\mathbf{x}_i, \mathbf{y}_i) \mid i = 1, \dots, n\}$, where $\mathbf{x}_i \in \mathbb{R}^{d_{in}}$ is a vector of inputs and $\mathbf{y}_i \in \mathbb{R}^{d_{out}}$ is a (vector) quantity derived from the microscale by way of our upscaling procedure. Learning a mapping between inputs and outputs then equates to learning a function of the form

$$\mathbf{y}_i = f(\mathbf{x}_i) \approx \hat{f}(\mathbf{x}_i). \quad (43)$$

As a result, we can incorporate local scale physics, such as those coming from the analytical solution, or numerical simulations, within macroscopic models without having to formulate explicit phenomenological expressions.

We frame our learning problem considering the physics of the process, inputs coming from simulation and an output that is aligned with the discrete structure of the numerical problem. For the latter, we specify our output following application of the backward Euler method commonly used for time discretisation in numerical modelling. Specifically

$$\frac{\partial \bar{p}_1}{\partial t} \approx \frac{\bar{p}_1(t+1) - \bar{p}_1(t)}{\Delta t}. \quad (44)$$

From Eq. (44), our goal is to predict the pressure at the next time step

$$y_i(t+1) = \bar{p}_1(t+1). \quad (45)$$

At simulation time we incorporate the resulting prediction into the finite-difference calculation in Eq. (44) at each time step. Accordingly, our learning problem is formulated on the basis of time series calculated with Eq. (39). Each individual time series corresponds to a different initial matrix pressure p_1^0 . We take 120 different p_1^0 , sampled with logarithmic spacing from the interval of $[1, 1 \times 10^6]$ Pa. Every series starts from 0 and finishes at a final time of $T = 100$ s, with a uniform stepping interval of 0.1 s. The remaining parameters in Eq. (39) are fixed as $k_1 = 0.5$ md, $\phi_1^0 = 0.2$, $c_1 = 1.4$ GPa $^{-1}$, $\mu = 1$ cp, $\ell = 1$ m, and following the step in \bar{p}_2 at t^{0+} , $\bar{p}_2 = 1$ MPa. A plot of each series s as time vs y^s is shown in Fig. 3(a). Further, Fig. 3(b) shows how this time series data is split into training and testing data using a 2/3 : 1/3 split respectively. Accordingly, every third sequence corresponds to a test series.

6. Application of the framework: Surrogate constitutive model learning

In this section we review and compare the learning algorithms used to generate our surrogate constitutive model. Specifically, we

introduce the various approaches and considerations for modelling the time-dependent problem. To follow, we train and test the different modelling approaches on our data. The results from this section will then be carried on to the next stage of the framework i.e. coupling the data-driven model to a numerical simulator.

6.1. Autoregressive approaches

To start, we introduce the modelling description used to account for time-dependency. That is, we model the sequential problem using an autoregressive description. In the following, we present autoregressive descriptions in the context of feedforward models. In feedforward approaches the information flow is unidirectional from the current inputs to outputs. This modelling approach is contrary to recurrent models, in which the information flow comes from both the current inputs and from calculations made on inputs at previous time steps.

For a (multivariable) feedforward autoregressive model, $\hat{f}(\cdot)$ uses previous series outputs to predict the next output in the series. Accordingly, for a two variable, univariate feedforward autoregressive model

$$\hat{y}_i(t+1) = \hat{f}(\mathbf{x}_i(t)) = \hat{f}([\hat{y}_i(t), \dots, \hat{y}_i(t-d_y+1), u_i(t), \dots, u_i(t-d_u+1)]), \quad (46)$$

where u_i denotes an external series, and d_y and d_u denote the number of sequence terms, or dependency length, for the internal and external series variables respectively. In this work we set $d_y = d_u = d$.

Our external series corresponds to the boundary conditions for the matrix provided by the homogenisation approach given the model problem described in Section 4. Specifically, previous sequence values for \bar{p}_2 corresponding to $\bar{p}_2 = p_1^0$ when $t < t^{0+}$, and $\bar{p}_2 = 1$ MPa (or any selected pressure step value) when $t \geq t^{0+}$. Finally, we note that d_y and d_u (or d) are defined a-priori. As a result, potentially important long-range information is lost if short dependency lengths are chosen. One option to alleviate the challenges posed by fixed dependency lengths would be to use recurrent neural networks such as the long short-term memory network.

For feedforward autoregressive training we treat the learning task as a *single-step ahead* prediction problem. In this strategy, the idea is to use lagged ground truths (targets) y_i in place of the predicted values \hat{y}_i in Eq. (46).

Autoregressive testing is treated according to how the ML model will be used when coupled to the (macroscopic) simulator. That is, as a *multi-step ahead* prediction problem. Accordingly, the trained single-step ahead models are used in a recursive manner, with predictions \hat{y} at previous time steps fed back into the inputs to predict the next

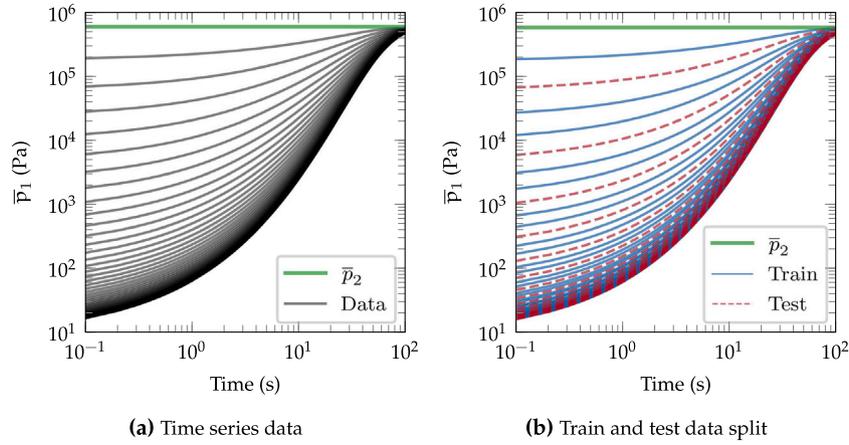


Fig. 3. Time series data for the learning problem: Raw data (a), and training and test data split (b).

time step, as per Eq. (46). The single-step ahead training and testing strategy can work provided the training error is sufficiently small such that accumulated errors do not explode when using the model as a multi-step ahead predictor. When this requirement on the accumulated errors cannot be satisfied (which is difficult to determine a-priori), iterative approaches are unstable. In such cases, multi-output strategies that predict the whole forecast horizon at once can be used. However, the stability of multi-output regressors often comes at the cost of predictive accuracy, particularly for long forecast horizons (Sangiorgio and Dercole, 2020).

In the section to follow we consider two ML algorithms for feed-forward autoregressive modelling. Specifically, polynomial regression (PR) and a fully connected neural network (FC-NN).

6.2. Polynomial regression

Here we introduce the polynomial regression algorithm. These algorithms are attractive because they are easy to understand and fast to train. However, a notable downside to these methods is the curse of dimensionality for large input vectors. Nonetheless, PR algorithms are a simple place to start for many regression tasks, and serve as a good baseline model. Accordingly, under PR $\hat{f}(\cdot)$ is formulated as

$$\hat{y}_i = \hat{f}(\mathbf{x}_i) = \mathbf{w}^\top \boldsymbol{\theta}(\mathbf{x}_i), \quad (47)$$

where $\boldsymbol{\theta} = [\theta_1, \dots, \theta_{n_p}]$ is the vector of polynomial basis functions acting on \mathbf{x}_i and $\mathbf{w} \in \mathbb{R}^{n_p}$ is a vector of weights (parameters) of the model. In this work we use a second order polynomial (feature) transformation, which comes as a standard tool in many machine/statistical learning packages.

Whilst the inputs are polynomial (following a transformation) in Eq. (47), the weights remain linear. Finding \mathbf{w} is simply then a linear regression problem and is the goal of training. In supervised learning problems, training corresponds to minimising a loss function $\mathcal{L}(\cdot)$ that represents the error in approximating $f(\cdot)$ by $\hat{f}(\cdot)$ on our data. For the regression problem we use a squared loss such that

$$\mathcal{L}(\mathbf{w}) = \frac{1}{n} \|\mathbf{y} - \hat{\mathbf{f}}(\boldsymbol{\theta}; \mathbf{w})\|_2^2, \quad (48)$$

where $\boldsymbol{\theta} = [\boldsymbol{\theta}(\mathbf{x}_1), \boldsymbol{\theta}(\mathbf{x}_2), \dots, \boldsymbol{\theta}(\mathbf{x}_n)]^\top$ and $\mathbf{y} = [y_1, y_2, \dots, y_n]^\top$ for the univariate target case. Finding the optimal set of weights that minimise Eq. (48) is defined formally as

$$\arg \min_{\mathbf{w}} \mathcal{L}(\mathbf{w}) = \arg \min_{\mathbf{w}} \frac{1}{n} \|\mathbf{y} - \hat{\mathbf{f}}(\boldsymbol{\theta}; \mathbf{w})\|_2^2. \quad (49)$$

The squared loss in linear regression is a convex function which admits a global minimum in \mathbf{w} at the point $\nabla \mathcal{L} = 0$. Accordingly

$$\boldsymbol{\theta}^\top \boldsymbol{\theta} \mathbf{w} = \boldsymbol{\theta}^\top \mathbf{y}, \quad (50)$$

leading to

$$\mathbf{w} = \boldsymbol{\theta}^\dagger \mathbf{y}, \quad (51)$$

where $\boldsymbol{\theta}^\dagger = (\boldsymbol{\theta}^\top \boldsymbol{\theta})^{-1} \boldsymbol{\theta}^\top$.

PR is well known to suffer from high degrees of multicollinearity (Dalal and Zickar, 2012). As a result, $\boldsymbol{\theta}^\top \boldsymbol{\theta}$ is ill-conditioned and the solution in Eq. (51) is non-unique. For prediction problems such ill-conditioning is generally not a concern (Dormann et al., 2013), since new data comes from the same population as the training data. However, for problems of extrapolation such as multi-step ahead prediction, collinearity between inputs can change due to accumulating errors. Such changes can affect predictions on models arising from ill-conditioned matrices and should be addressed (Dormann et al., 2013). We therefore address multicollinearity between inputs using regularisation, which aims to restrict the space of viable models that fit our data (e.g. in the face of ill-conditioning). There are various forms of regularisation, however in this work we add the following penalisation $\gamma \|\mathbf{w}\|_2^2$ term to Eq. (48) leading to so-called ridge regression. The parameter γ in the penalty then controls the degree of weight penalisation.

6.3. Fully connected neural networks

Here we introduce the fully connected neural network model. Aside from the FC-NN considered in this work, neural networks more generally have become a popular choice for surrogate modelling due to their great flexibility, scalability and significant advancement within research communities. Further, contrary to the polynomial regression where the feature transforms are defined a-priori, in neural networks these transformations are implicit within the algorithm and learnt from the data.

To proceed, within an FC-NN information is fed forward from an input layer through hidden layers ($l = 1, \dots, L-1$) finishing at an output layer ($l = L$). Accordingly, under an FC-NN $\hat{f}(\cdot)$ is formulated as

$$\hat{y}_i = \hat{f}(\mathbf{x}_i) = s^{[L]}(s^{[L-1]}(\dots(s^{[1]}(\mathbf{x}_i)\dots))). \quad (52)$$

For a given layer l , $s^{[l]}$ is given as an affine transformation followed by an element-wise nonlinear function such that

$$s^{[l]}(\mathbf{x}) = h^{[l]}(\mathbf{W}^{[l]\top} \mathbf{x} + \mathbf{b}^{[l]}), \quad (53)$$

where $\mathbf{W}^{[l]} \in \mathbb{R}^{n_{l-1} \times n_l}$ and $\mathbf{b}^{[l]} \in \mathbb{R}^{n_l}$ are the weight matrix and bias parameter vector associated with layer l respectively. Evidently, for the univariate prediction considered herein, the weight matrix for the last layer will be a weight vector $\mathbf{W}^{[L]} = \mathbf{w}^{[L]}$. Notation $h(\cdot)$ is the nonlinear function referred to as an activation function. In this work, we take $h^{[L]}$ as a linear activation and all others as so-called exponential linear

units. In training the FC-NN we minimise the customary squared loss of the form

$$\mathcal{L}(\lambda) = \frac{1}{n} \|\mathbf{y} - \hat{f}(\mathbf{X}; \lambda)\|_2^2, \quad (54)$$

where $\lambda = [\mathbf{W}^{[1]}, \mathbf{b}^{[1]}, \dots, \mathbf{W}^{[L]}, \mathbf{b}^{[L]}]$ and $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]^\top$. However, unlike Eq. (48), with neural networks the loss function is non-convex. As a result, weights and biases are updated iteratively using a chosen optimiser together with the backpropagation algorithm (see Goodfellow et al., 2016 for further details). In this work, we found stochastic gradient descent with Nesterov momentum to work the best (Nesterov, 1983). We experimented with various model depths and widths, and found the best results came using a simple single hidden layer FC-NN with 12 hidden units. Additionally, with this architecture there was little overfitting for the single-step problem, and hence no need for regularisation. Our FC-NN was implemented using the Keras API (Chollet, 2015).

Lastly, we construct the datasets used by these models as $\mathcal{D} = (\mathbf{X}, \mathbf{y})$ such that $\mathbf{X} = [\mathbf{x}_1^1, \mathbf{x}_2^1, \dots, \mathbf{x}_{1000}^S]^\top$ and $\mathbf{y} = [y_1^1, y_2^1, \dots, y_{1000}^S]^\top$. Superscripts on the input matrix and target vector entries denote the time series label, with $S (= 120)$ being the total number of series. Subscripts then denote the time step within a given series.

6.4. Training and testing

In the following we present some additional considerations for model training and interpretation. Then, we present the results and discussions for the model training (single-step ahead) and testing (multi-step ahead) procedures.

Considerations

For our models we consider a dependency length of $d = 3$ for the inputs. Although not studied in detail here, d is a hyperparameter that can be tuned according to the trade-off between gain in accuracy and complexity.

The weight penalty term γ for the regularised polynomial regression (RPR) was found to give the best results for $\gamma = 1 \times 10^{-7}$. We present results comparing the impact of different values for γ for the multi-step ahead problem on both training and testing sets in the section to follow.

When training our FC-NN model we used a batch size of 128, run over 500 epochs. Further, we standardise our data to have a mean of zero and a standard deviation of one. Standardisation during testing is then done using the mean and standard deviation calculated on the training set.

To evaluate between the different models we consider one minus the normalised root-mean-square error given as

$$1 - NRMSE = 1 - \frac{RMSE}{\sigma} = \frac{\sqrt{(n-1)^{-1} \sum_{i=1}^n (y_i - \bar{y})^2}}{\sqrt{(n-1)^{-1} \sum_{i=1}^n (y_i - \bar{y})^2}}, \quad (55)$$

where σ and \bar{y} represent the sample standard deviation and mean of the (train or test) target set respectively. From Eq. (55) the closer the value to one, the more accurate the algorithm. Lastly, to highlight the effects of multicollinearity, we present results for the polynomial regression model both with and without regularisation.

Results

Fig. 4 shows the train and test results using the unregularised polynomial regression algorithm. From Fig. 4(a), we can see there is a good match when training as a single-step ahead prediction problem. The quality of this match is corroborated by the high evaluation metric (and low RMSE respectively) shown in Table 1. However, during testing, the prediction quality deteriorates, particularly for sequences associated with high initial pressures (Fig. 4(b)). Further, for the series corresponding to the highest initial pressure, we observe unbounded error growth with the prediction rapidly going negative. Accordingly,

this latter unbounded prediction leads to the missing entries for the error metrics for the test problem shown in Table 1. We hypothesised errors in the multi-step ahead case with this model to be caused by the multicollinearity existing between the transformed features, as described in Section 6.2. Accordingly, to address this problem we introduced a regularised PR by way of a penalisation term on the loss.

Results for the regularised PR are shown in Fig. 5 and Table 1. The training results shown in Fig. 5(a) are virtually identical to those shown in Fig. 4(a) for the unregularised PR case (Table 1). However, we can observe measurable differences on the multi-step ahead test set between the two polynomial regression models. From Fig. 5(b), we see the regularised PR predictions are qualitatively closer to the targets, contrary to Fig. 4(b). Further, we do not observe the unbounded predictions shown in the latter figure. Looking at the error metrics for the regularised PR model in Table 1, we see the RMSE is significantly higher for the testing set. Such differences are expected given the accumulation of errors, with the higher test error likely due to the small discrepancies observed at late times in Fig. 5(b). Nonetheless, the results using the simple (regularised) polynomial regression model are both qualitatively and quantitatively reasonable, serving as a good benchmark for the more complex methods to follow.

Fig. 6 shows the penalisation parameter tuning for the regularised PR. Accordingly, we observe as γ increases from $\gamma = 1 \times 10^{-9}$ to $\gamma = 1 \times 10^{-7}$ the RMSE on both the training and testing sets (treated as both multi-step ahead problems for this demonstration) decreases. However, from $\gamma = 1 \times 10^{-7}$ to $\gamma = 1 \times 10^{-6}$, we can see a sharp increase in RMSE. Indeed, past $\gamma = 1 \times 10^{-6}$ our RMSE becomes unbounded as the accumulating errors become significant. The results from Fig. 6 support our use of $\gamma = 1 \times 10^{-7}$ for the weight penalisation.

Lastly, Fig. 7 and Table 1 show the train and test results using the FC-NN. From Fig. 7(a), we see a good match between the training predictions and targets. However, comparison of the quantitative training results in Table 1 between the FC-NN and the RPR suggests underfitting with the former. From Fig. 7(b), we can observe good qualitative matches between the predictions and targets over early-time. However, from middle- to late-times the predictions diverge from the targets, suggesting the effects of accumulating errors on the multi-step predictions over these time periods. The resulting degradation in performance is reflected in the error metrics in Table 1 for the FC-NN on the testing case. We anticipate this diverging prediction behaviour to be a consequence of the underfitting observed with the FC-NN. Possible solutions to the underfitting problem include increasing the complexity of the model and/or using more data. For the former, we found increasing the width and depth of the network to add little benefit. Alternatively, whilst increasing the availability of data is trivial for this problem, it may not be practical for expensive-to-generate data regimes. Instead, we hypothesise the underfitting in this case to be down to the optimisation process used when learning the FC-NN. Specifically, even for convex loss landscapes, solutions achieved by gradient-based optimisers are unlikely to represent the 'optimum' solution due to the sensitivity of their performance on hyperparameters such as the learning rate. Nonetheless, when learning on more complex loss landscapes, these solutions are often 'good enough'. For example, in computer vision and natural language processing, deep neural networks remain the state of the art. However, for current purposes, in light of accumulating errors in the multi-step ahead prediction problem, we require a high degree of accuracy from our learnt surrogate constitutive model. Accordingly, in contrast to the FC-NN, the weights in the RPR are typically obtained using highly accurate matrix factorisation methods that lead to globally optimal solutions to the convex optimisation problem. It is worth noting, however, despite its performance compared to the PR model, the FC-NN does not suffer from the same multicollinearity problems as the latter. Such a feature may be useful for future investigations involving neural networks in such scientific problems.

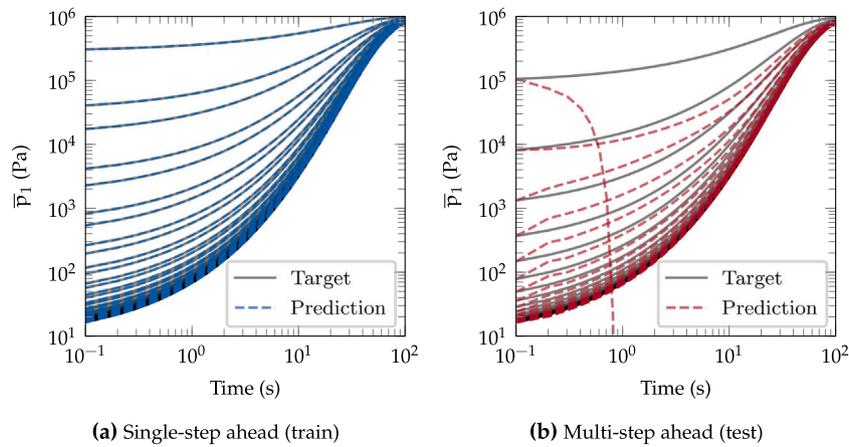


Fig. 4. Results for the unregularised polynomial regression: Single-step ahead training (a), and multi-step ahead testing (b).

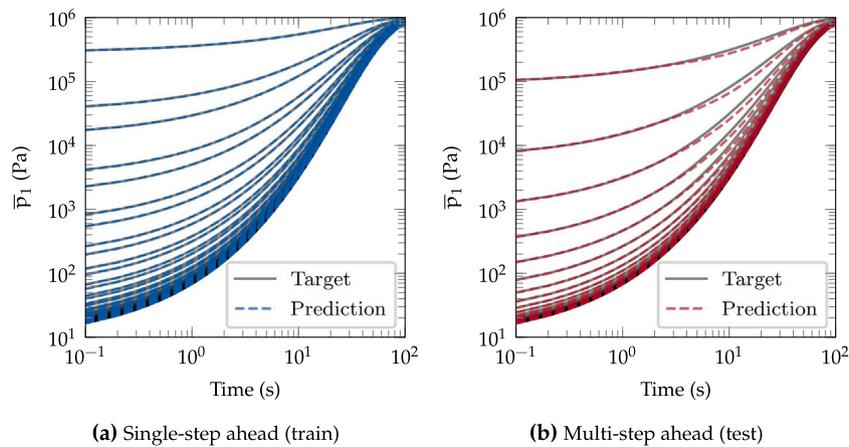


Fig. 5. Results for the regularised polynomial regression: Single-step ahead training (a), and multi-step ahead testing (b).

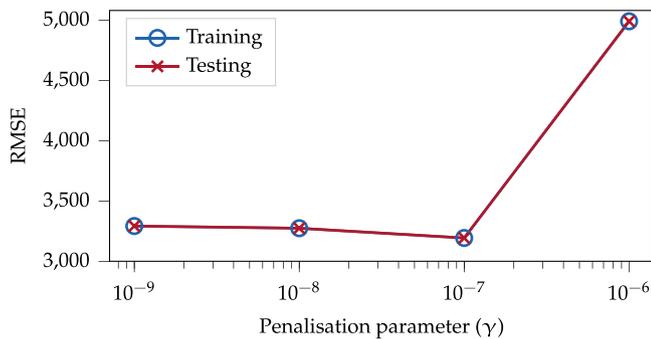


Fig. 6. Multi-step ahead training and testing RMSE for different penalisation parameter values in the regularised polynomial regression.

In summary, we have shown training and testing results for the time-dependent problem using several data-driven modelling strategies. We have seen that complexity does not always result in superior performance, with the simple regularised polynomial regression model outperforming the neural network based approach on both training and testing problems. Indeed, as part of this work, we also experimented

Table 1

Train and test evaluations for the different learning algorithms.

Algorithm	$(RMSE)_{train}$	$(1 - NRMSE)_{train}$	$(RMSE)_{test}$	$(1 - NRMSE)_{test}$
UPR	3.43	0.999	–	–
RPR	3.57	0.999	3200	0.983
FC-NN	266	0.998	13700	0.926

UPR, RPR, and FC-NN denote the unregularised and regularised polynomial regressors, and fully connected neural network respectively. The missing entry for UPR on the test is due to unbounded predictions.

with fitting even more complex neural networks, so-called encoder–decoder long short-term memory networks.² However, we found these recurrent neural networks more involved to fit in practice, with the results falling short of the performance of the polynomial regression models. For more complex problems, the additional effort required to fit such models may be warranted. However, when learning from expensive-to-generate data, we envisage a key challenge for machine learning-based multiscale modelling to be in the sample-efficiency of learning. Given the results shown in this section, we use the regularised PR model for the next stage of the framework application.

² See <https://github.com/mashworth11/ML-MM> for further details.

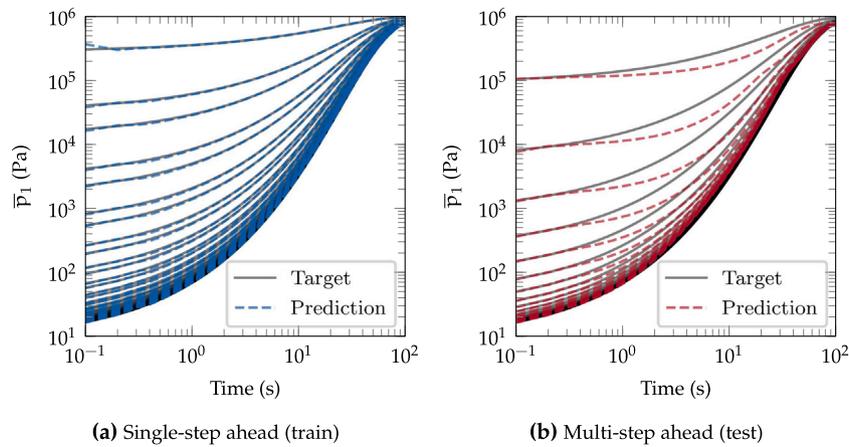


Fig. 7. Results for the fully connected neural network: Single-step ahead training (a), and multi-step ahead testing (b).

7. Application of the framework: Model coupling

Here we couple our trained surrogate constitutive model to a macroscopic physics-based model. We test the subsequent hybrid ML-physics model on a variety of cases against a traditional physics-based model that uses the linear transfer, and a microscale model.

7.1. Hybrid ML-physics model

Our physics-based numerical model is the discretised counterpart to Eqs. (35) to (36). Time discretisation is achieved using the backward Euler method, as per Eq. (44). Discretisation for the pressure gradient is done using the finite-volume method by way of the two-point flux approximation. The latter is implemented as standard within MRST. Given the propensity for nonlinear phenomena in subsurface flow (e.g. due to compressibility) it is common to use nonlinear solvers. Accordingly, we give Eqs. (35) to (36) in discrete block matrix form following application of Newton's method as

$$\begin{bmatrix} F_1 & G_1 \\ G_2 & F_2 \end{bmatrix}^{(k)} \begin{bmatrix} \delta \hat{p}_1 \\ \delta \hat{p}_2 \end{bmatrix}^{t+1,(k+1)} = - \begin{bmatrix} R_1 \\ R_2 \end{bmatrix}^{t+1,(k)}, \quad (56)$$

where δ and k now denote the change in solution and current iteration levels respectively. The flow matrix in the Jacobian is $F_\alpha = Q_\alpha + \Delta t T_\alpha$, where Q_α is the compressibility matrix and T_α is the transmissibility matrix (see Lie 2019 for details). For the matrix material in the dual-porosity setting, the transmissibility matrix has zero entries. Matrix G_α is the coupling matrix arising from an inter-porosity flow expression such as the linear approximation in Eq. (42) (see Ashworth and Doster, 2020 for details). Note, terms arising from the linear transfer approximation will also appear in the flow matrices. The solution vectors are $\hat{p}_\alpha = [\hat{p}_\alpha^1, \dots, \hat{p}_\alpha^{n_{cell}}]^\top$, where \hat{p}_α denotes the cell-wise pressure solution for continuum α . Notation n_{cell} is then the number of grid cells (control volumes). Next, $R_\alpha^{t+1,(k)} = [R_\alpha^1, \dots, R_\alpha^{n_{cell}}]^\top$ are the residual vectors for continuum α per grid cell. Intuitively, $R_\alpha^{t+1,(k)}$ provides the difference between the rate of change of fluid mass within a cell volume to that generated by sources and/or fluid movement through its boundaries at iteration level k . Finally, we inject our machine learning model into Eq. (56) through an explicit approach. As a result, following a Newton iteration the coupling matrices G_α in Eq. (56) are zero entries, whilst the machine learning-based \bar{r}_1 appears in the residuals.

7.2. Tests

To test the hybrid ML-physics model we consider several realisations of the 2D diffusion problem described in Section 4. For these realisations, we test our framework on initial conditions and time lengths not used for training our model. Accordingly, we compare the hybrid

approach against a microscale model and a physics-based model that uses the first-order mass transfer constitutive relation. For brevity in the remainder of this section, we refer to the latter method as the 'traditional' approach to dual-continuum modelling. In addition to the 2D problem, we also consider an application of the framework to a geological model. For this case, we only compare the hybrid and traditional dual-porosity approaches.

For the 2D problem test cases we discretise the DP problem using a single control volume. For the microscale model we resolve the matrix using 40×40 grid cells. Within the DP model we initialise the fracture pressure as $\bar{p}_2^0 = 1$ MPa. Further, to mimic the fracture boundary within the DP model, we set fracture permeability to 10 d. As a result, the fracture pressure equilibrates almost instantaneously with the external boundary in response to changes induced by inter-continuum flow. The remaining properties are as described in Section 5. Specifically, $\bar{\phi}_2^0 = 1 \times 10^{-3}$, $c_2 = c_1 = 1.4$ GPa $^{-1}$, $v_1 = 0.999$, $\phi_1^0 = 0.2$, $\mu = 1$ cp, $N_f = 1$, $\ell = 1$ m, $\rho^0 = 1000$ kgm $^{-3}$ and $k_1 = k' = 1$ md. With respect to $\bar{\phi}_2^0$ and c_2 , these are chosen somewhat arbitrarily since we do not consider fracture dynamics for the problems herein. Lastly, we make use of a discrete equivalent to Eq. (2) to compare the (averaged) microscopic matrix pressure field with its dual-continuum counterpart.

For the geological model, we use a grid taken from the (Open Porous Media Initiative, 2015). Accordingly, each control volume now corresponds to a 3D domain involving three orthogonal fracture sets leading to a 3D diffusion problem within the matrix. As a result, the analytical solution used to generate data is a higher dimensional version of Eq. (39). Further details of this solution, and the parameterisations used for this test, are described as part of the test case descriptions introduced below.

Case 1: $p_1^0 \in [1, 1 \times 10^6]$ Pa

For the first set of test cases we consider two initial pressures not used when constructing D , but coming from $[1, 1 \times 10^6]$ Pa. Specifically, we sample the initial pressures such that $p_1^{0,I} < 10$ Pa and 0.1 MPa $< p_1^{0,II}$, where superscripts I and II correspond to the two tests respectively. The boundary pressure is fixed as $\bar{p}_2 = 1$ MPa. The parameters for this 2D problem are as described above.

Case 2: $T \in \{1, 10\}$ s

For the second set of tests we consider shorter sequences for training and use the resulting models for testing on the full sequence. Specifically, we consider datasets up to $T = 1$ s and $T = 10$ s for training, instead of a 'full' dataset corresponding to $T = 100$ s. The motivation behind this test case is to see how the models perform when trained under limited data and used beyond the training data range. Accordingly, extrapolation outside of the training conditions is

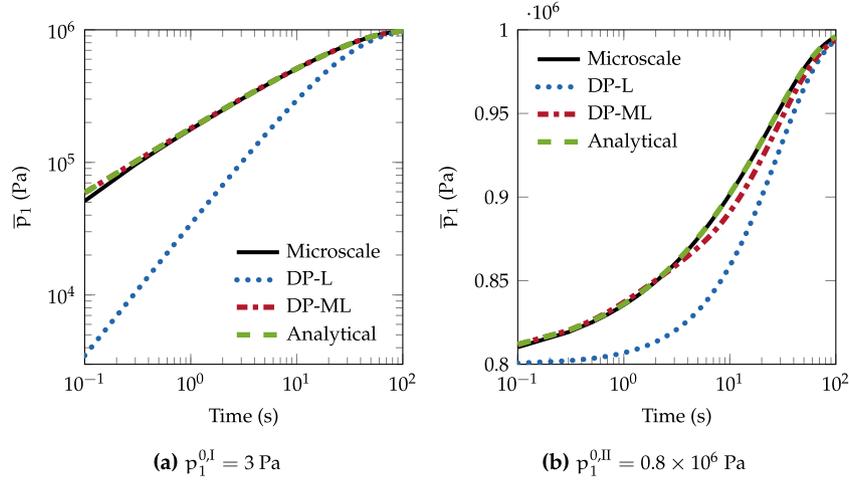


Fig. 8. Average matrix pressure evolutions for test case 1: $p_1^{0,I} = 3$ Pa (a) and $p_1^{0,II} = 0.8 \times 10^6$ Pa (b). Notations ‘DP-L’ and ‘DP-ML’ denote the dual-porosity model equipped with the linear and machine learning constitutive transfer models respectively.

now caused by the length of the testing sequence length relative to the training sequence length, as well as due to the multi-step ahead approach. The parameters for this 2D problem are as described above, with $p_1^0 = 1$ Pa.

Case 3: Application to a geological model

For the 3D problem, Eq. (39) is extended as (Lim and Aziz, 1995; Zhou et al., 2017),

$$\bar{p}_1 = p_1^0 + (\bar{p}_2 - p_1^0) \left\{ 1 - \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} \sum_{o=0}^{\infty} \left(\frac{8}{\pi^2} \right)^3 \frac{1}{(2n+1)^2(2m+1)^2(2o+1)^2} \exp\left(-\frac{\pi^2 k_1 t}{\phi_1^0 c_1 \mu \ell^2} [(2n+1)^2 + (2m+1)^2 + (2o+1)^2] \right) \right\} \quad (57)$$

where we parameterise Eq. (57) using $\bar{p}_2 = 1$ MPa, $c_1 = 3$ GPa $^{-1}$, $\phi_1^0 = 0.2$, $\mu = 5$ cp, $\ell = 1$ m and $k_1 = 1 \times 10^{-4}$ md. As before, we take 120 different p_1^0 sampled from the interval $[1, 1 \times 10^6]$ Pa. Further, we generate series starting from $t = 0$ and finishing at $T = 100$ h, with a uniform stepping interval of $\Delta t = 1$ h. Finally, we train and test on the resulting data using an autoregressive regularised polynomial regression model, which is subsequently injected into our physics-based simulator.

For our geological DP model, the boundaries are specified as pressure boundaries set at 1 MPa. Next, the (macroscopic) fracture permeability is set at 10 d. Accordingly, the high fracture permeability ensures our numerical test problem is consistent with our learning problem in terms of pressure conditions for the fracture. We initialise the matrix as $p_1^0 = 5$ Pa. To complete the DP model we assign $\bar{\phi}_2^0 = 1 \times 10^{-3}$, $c_2 = 3$ GPa $^{-1}$ and $\rho^0 = 850$ kgm $^{-3}$. We compare results over the geological model between the hybrid ML-physics model and the physics-based model using the linear transfer. For the latter, the transfer function is parameterised using $N_f = 3$ and $k' = 1 \times 10^{-4}$ md. Finally, despite only training for a series corresponding to 100 h we run each simulation for a 1000 h, thus extrapolating outside of the training conditions.

7.3. Results

Here we present the matrix pressure evolution results for the test cases described above.

Case 1: $p_1^0 \in [1, 1 \times 10^6]$ Pa

Results for the first test case are shown in Fig. 8. In both cases we see pressure increasing with time as the matrix equilibrates with

the boundary pressure. We can see for both tests there is a measurable discrepancy between the DP model using a linear transfer and the equivalent microscale description. This discrepancy is particularly pronounced for test I (Fig. 8(a)). For this test case, the linear transfer is a particularly poor early-time model when the difference between continuum pressures is several orders of magnitude. In contrast, the DP approach with the ML transfer produces high quality matches for both test cases without the computational expense taken to run the microscale models. In Fig. 5(b), we observed small errors in the middle- to late-time regularised PR predictions for test series corresponding to high initial pressures. However, Fig. 8(b) shows despite these discrepancies, the DP model using the data-driven transfer still outperforms a linear transfer-based DP model with respect to the microscale solution.

Lastly, for both tests we also provide the analytical solutions (Fig. 8). The benefit of this addition is clearer in Fig. 8(a). From Fig. 8(a) we observe a small discrepancy between the microscale model and the analytical solution for the first few time steps. This discrepancy arises due to the time step discretisation error in the microscale model. Remarkably, we do not see this discrepancy between the ML-based approach and the analytical solution, since the learnt model comes from data unaffected by these discretisation errors. This subtle result corroborates the aim of multiscale approaches. Specifically, to combine the accuracy of microscopic representations with the practicality of macroscopic models.

Case 2: $T \in \{1, 10\}$ s

Fig. 9 shows the results for the different training and testing sequence length test cases. Fig. 9(a) shows the test case when using up to $T = 1$ s of data (as denoted by the grey vertical line). In this figure, we observe a good qualitative match relative to the microscale model using the hybrid approach over the first 1 s. Further, even for time steps within a few seconds after the 1 s training limit, the match between the hybrid and microscale models is good. However, from middle- to late-times, Fig. 9(a) shows we see the hybrid model diverging from the microscale model solution. Nonetheless, at late-times the model converges back to the microscale solution. In Fig. 9(b), the sequence limit is now at $T = 10$ s. As one would expect, more data used for our surrogate model results in improved hybrid model performance across the sequence length. Further, in Fig. 9(b) there is a noticeably smaller divergence between the hybrid and microscale model results following the 10 s limit, compared to that shown in Fig. 9(a) following the 1 s limit. This result is consistent with the general trend between surrogate model accuracy and data volume, leading to a lower accumulation of errors. Nevertheless, even with a surrogate constitutive model trained

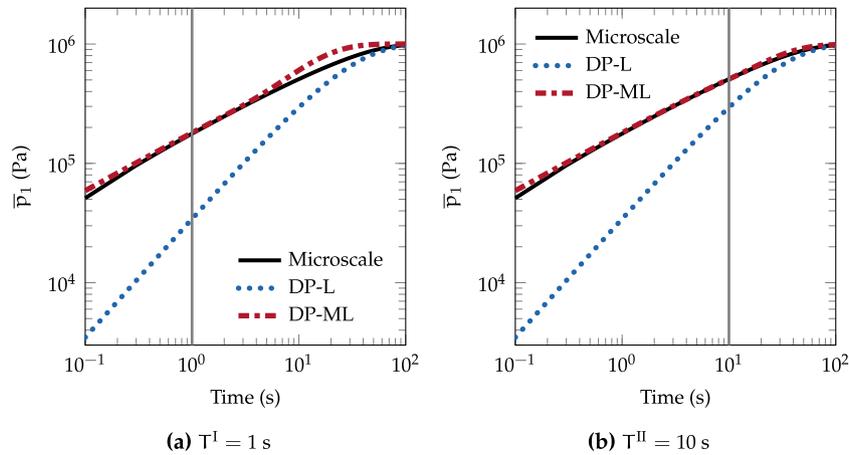


Fig. 9. Average matrix pressure evolutions for test case 2: $T^I = 1$ s (a) and $T^{II} = 10$ s (b). Notations ‘DP-L’ and ‘DP-ML’ denote the dual-porosity model equipped with the linear and machine learning constitutive transfer models respectively.

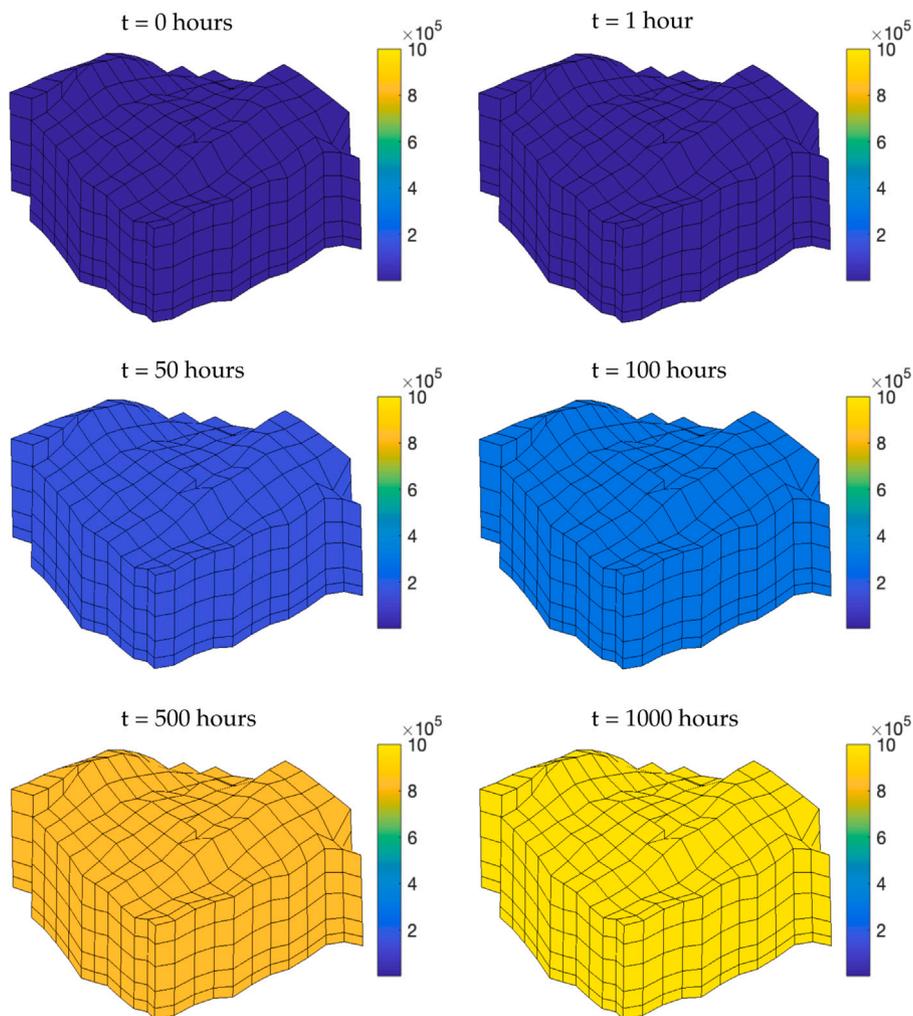


Fig. 10. Matrix pressure evolutions on a geological model using the linear transfer model.

using $T = 1$ s of sequence data, the hybrid model still performs well over the whole simulation time.

From the first two test cases, we have seen a traditional dual-porosity model approach gives the worst results at early-times. At these times, there is a significant different between matrix and boundary pressures. However, at later times, the linear transfer relation is a reasonable approximation of inter-continuum mass exchange. Accordingly,

it is wasteful to use a surrogate constitutive model in regions where a simple constitutive model performs well, given the data requirements to train the former. An interesting development, therefore, would be a hybrid model that can use different constitutive models in different regions (e.g. nonlinear vs linear) of a problem space (March et al., 2016).

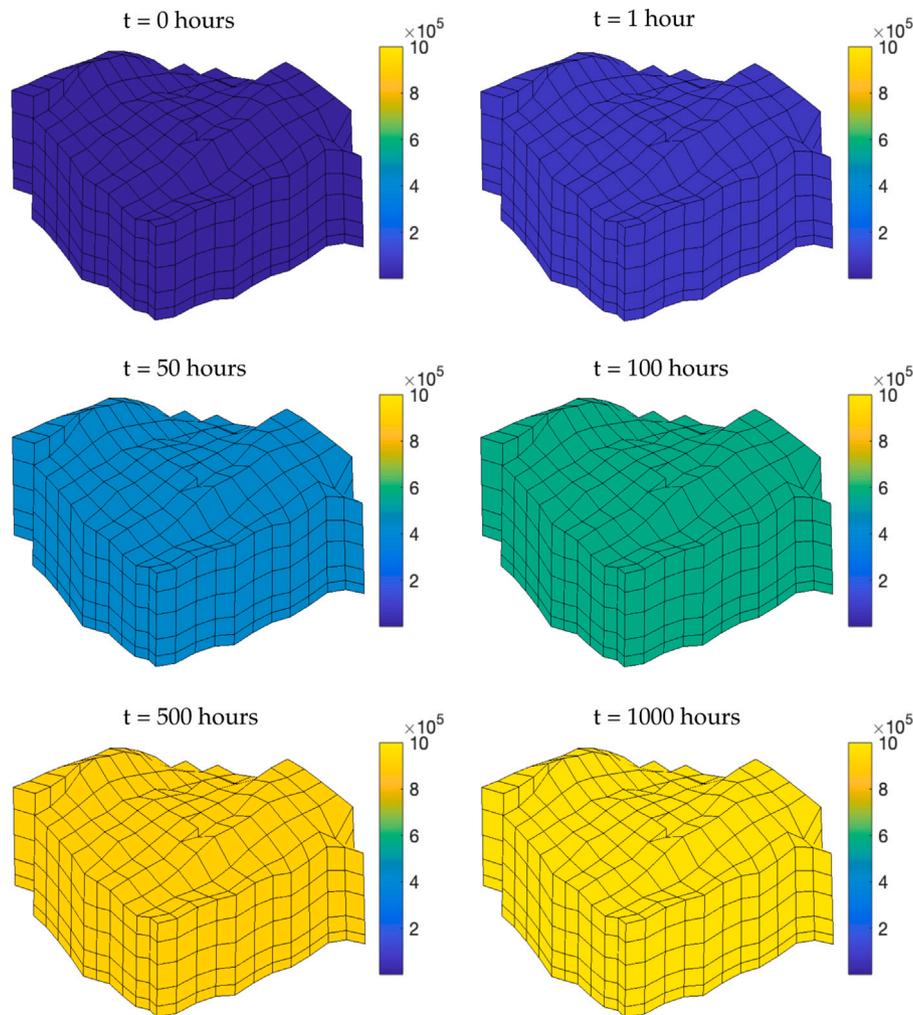


Fig. 11. Matrix pressure evolutions on a geological model using the machine learning-based transfer model.

Case 3: Application to a geological model

Figs. 10 to 11 show the results for the geological model test case using the traditional and ML-physics model respectively. In both figures we can see matrix pressure rising over time in response to diffusion driven equilibration with the fractures. However, we can see significant differences between the two modelling approaches following the initial conditions up until equilibrium conditions. Comparing Figs. 10 to 11 at $t = 1$ h, we can see the matrix pressures are noticeably lower when using the linear transfer compared to the hybrid ML-physics approach. This observation is consistent with the results in the previous tests, which showed the matrix pressures to be significantly underestimated when using a linear approximation. Accordingly, matrix fluxes are underpredicted when using this linear transfer approach. Such errors could be significant in geological settings such as groundwater remediation (Haggerty and Gorelick, 1995). Similar differences between the two approaches to those just described are observed at $t = 50, 100$ h, with the ML-physics approach predicting higher pressures than the traditional approach.

When extrapolating outside of the training conditions, the hybrid approach shows physically reasonable results (consistent with the previous time steps) at $t = 500$ h. Accordingly, at this time step, matrix pressures coming from the hybrid approach are higher compared to the traditional approach. However, by $t = 1000$ h, the two models have converged to the equilibrated pressure state.

In summary, the results presented for the described tests show the potential for the ML-physics approach for multiscale modelling. With

this framework we can capture the accuracy of microscopic representations with the practicality of macroscopic models. We envision a number of interesting extensions to be investigated using the basis of the framework introduced herein.

8. Conclusions and future work

This paper introduces and applies a machine learning-based multiscale constitutive modelling framework. Accordingly, we detailed the key components of the framework as: Homogenisation, data generation, surrogate constitutive model learning, and ML-physics model coupling. In doing, we described various considerations that the user should make at each step. To test the framework we considered a multiscale pressure diffusion problem. Specifically, that of describing inter-continuum mass transfer in dual-porosity materials. In practice, it is common to model this process using an overly simple linear constitutive relation. Instead of making any such assumption, we use our framework to build a surrogate constitutive model based on microscale data. In applying the framework we introduced the various machine learning methods used to create this surrogate constitutive model. We account for time-dependence using autoregressive approaches, considering polynomial and neural network-based regressors. We found the former to provide higher accuracy when used as a multi-step ahead predictor, provided regularisation is used to deal with multicollinearity effects. Accordingly, we integrated the resulting data-driven surrogate into a physics-based dual-porosity model, creating a hybrid ML-physics approach. We showed the resulting hybrid approach to give high quality

results for a number of test cases compared to a linear transfer-based dual-porosity model, without the computational expense of explicit microscale models.

We hypothesise a number of exciting possibilities for further work when using data coming from more complex microscale scenarios (e.g. dynamic boundary conditions) than the ones considered here. In such cases, data is likely to be obtained numerically. However, numerical data can be expensive, leading to sparse datasets. Accordingly, in scientific computing domains, we see sample-efficiency as a key challenge for the uptake of machine learning-based multiscale modelling. In addressing this challenge, future developments could involve exploring different sampling and learning strategies such as active learning, theory-driven ML and probabilistic approaches to improve learning efficiency, generalisability and to understand uncertainty in multiscale settings.

CRedit authorship contribution statement

Mark Ashworth: Conceptualisation, Methodology, Investigation, Writing, Review. **Ahmed H. Elsheikh:** Supervision, Review. **Florian Doster:** Supervision, Review.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors are grateful for the funding provided to them by the Natural Environmental Research Council to carry out this work, and to Anne-Sophie Ruget for advice on figures.

References

- Andrianov, N., Nick, H.M., 2021. Machine learning of dual porosity model closures from discrete fracture simulations. *Adv. Water Resour.* 147, 103810.
- Ashworth, M., Doster, F., 2020. Anisotropic dual-continuum representations for multiscale poroelastic materials: Development and numerical modelling. *Int. J. Numer. Anal. Methods Geomech.*
- Asproulis, N., Drikakis, D., 2013. An artificial neural network-based multiscale method for hybrid atomistic-continuum simulations. *Microfluid. Nanofluid.* 15 (4), 559–574.
- Auriault, J., 1983. Effective macroscopic description for heat conduction in periodic composites. *Int. J. Heat Mass Transfer* 26 (6), 861–869.
- Auriault, J.-L., Boutin, C., Geindreau, C., 2010. Homogenization of Coupled Phenomena in Heterogeneous Media, Vol. 149. John Wiley & Sons.
- Bai, S., Kolter, J.Z., Koltun, V., 2018. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. arXiv preprint arXiv:1803.01271.
- Blanco, P.J., Sánchez, P.J., de Souza Neto, E.A., Feijóo, R.A., 2016. Variational foundations and generalized unified theory of RVE-based multiscale models. *Arch. Comput. Methods Eng.* 23 (2), 191–253.
- Brassart, L., Stainier, L., 2019. Effective transient behaviour of heterogeneous media in diffusion problems with a large contrast in the phase diffusivities. *J. Mech. Phys. Solids* 124, 366–391.
- Chen, G., 2021. Recurrent neural networks (RNNs) learn the constitutive law of viscoelasticity. *Comput. Mech.* 67 (3), 1009–1019.
- Chollet, F., 2015. Keras. URL: <https://github.com/fchollet/keras>.
- Crank, J., 1979. *The Mathematics of Diffusion*. Oxford University Press.
- Crombecq, K., 2011. Surrogate Modeling of Computer Experiments with Sequential Experimental Design (Ph.D. thesis). Ghent University.
- Dalal, D.K., Zickar, M.J., 2012. Some common myths about centering predictor variables in moderated multiple regression and polynomial regression. *Organ. Res. Methods* 15 (3), 339–362.
- Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J.R.G., Gruber, B., Lafourcade, B., Leitao, P.J., et al., 2013. Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography* 36 (1), 27–46.
- Feyel, F., 1999. Multiscale Fe2 elastoviscoplastic analysis of composite structures. *Comput. Mater. Sci.* 16 (1–4), 344–354.
- Geiger, S., Dentz, M., Neuweiler, I., et al., 2013. A novel multi-rate dual-porosity model for improved simulation of fractured and multiporosity reservoirs. *SPE J.* 18 (04), 670–684.
- Ghaboussi, J., Garrett, Jr., J., Wu, X., 1991. Knowledge-based modeling of material behavior with neural networks. *J. Eng. Mech.* 117 (1), 132–153.
- Ghaboussi, J., Pecknold, D.A., Zhang, M., Haj-Ali, R.M., 1998. Autoprogressive training of neural network constitutive models. *Internat. J. Numer. Methods Engrg.* 42 (1), 105–126.
- Ghavamian, F., Simone, A., 2019. Accelerating multiscale finite element simulations of history-dependent materials using a recurrent neural network. *Comput. Methods Appl. Mech. Engrg.* 357, 112594.
- Goh, A.T., 1995. Back-propagation neural networks for modeling complex systems. *Artif. Intell. Eng.* 9 (3), 143–151.
- Goodfellow, I., Bengio, Y., Courville, A., 2016. *Deep Learning*. MIT Press.
- Haggerty, R., Gorelick, S.M., 1995. Multiple-rate mass transfer for modeling diffusion and surface reactions in media with pore-scale heterogeneity. *Water Resour. Res.* 31 (10), 2383–2400.
- Hambli, R., Katerchi, H., Benhamou, C.-L., 2011. Multiscale methodology for bone remodelling simulation using coupled finite element and neural network computation. *Biomech. Model. Mechanobiol.* 10 (1), 133–145.
- Hashash, Y., Jung, S., Ghaboussi, J., 2004. Numerical implementation of a neural network based material model in finite element analysis. *Internat. J. Numer. Methods Engrg.* 59 (7), 989–1005.
- Holman, J.P., 1990. *Heat Transfer*. McGraw-Hill.
- Kaessmair, S., Steinmann, P., 2018. Computational first-order homogenization in chemo-mechanics. *Arch. Appl. Mech.* 88 (1–2), 271–286.
- Kouznetsova, V., Brekelmans, W., Baaijens, F., 2001. An approach to micro-macro modeling of heterogeneous materials. *Comput. Mech.* 27 (1), 37–48.
- Larsson, F., Runesson, K., Su, F., 2010. Variationally consistent computational homogenization of transient heat flow. *Internat. J. Numer. Methods Engrg.* 81 (13), 1659–1686.
- Lary, D.J., Alavi, A.H., Gandomi, A.H., Walker, A.L., 2016. Machine learning in geosciences and remote sensing. *Geosci. Front.* 7 (1), 3–10.
- Lie, K.-A., 2019. *An Introduction to Reservoir Simulation using MATLAB/GNU Octave: User Guide for the MATLAB Reservoir Simulation Toolbox (MRST)*. Cambridge University Press.
- Lie, K.-A., Krogstad, S., Ligaarden, I.S., Natvig, J.R., Nilsen, H.M., Skaflestad, B., 2012. Open-source matlab implementation of consistent discretisations on complex grids. *Comput. Geosci.* 16 (2), 297–322.
- Lim, K., Aziz, K., 1995. Matrix-fracture transfer shape factors for dual-porosity simulators. *J. Pet. Sci. Eng.* 13 (3–4), 169–178.
- Ling, J., Jones, R., Templeton, J., 2016. Machine learning strategies for systems with invariance properties. *J. Comput. Phys.* 318, 22–35.
- March, R., Doster, F., Geiger, S., 2016. Accurate early-time and late-time modeling of countercurrent spontaneous imbibition. *Water Resour. Res.* 52 (8), 6263–6276.
- Masi, F., Stefanou, I., Vannucci, P., Maffi-Berthier, V., 2021. Thermodynamics-based artificial neural networks for constitutive modeling. *J. Mech. Phys. Solids* 147, 104277.
- Miehe, C., Koch, A., 2002. Computational micro-to-macro transitions of discretized microstructures undergoing small strains. *Arch. Appl. Mech.* 72 (4–5), 300–317.
- Nesterov, Y., 1983. A method for unconstrained convex minimization problem with the rate of convergence $\alpha(1/k^2)$. *Dokl. USSR* 269, 543–547.
- Open Porous Media Initiative, 2015. *Open datasets*. URL <https://github.com/OPM/opm-data>.
- Özdemir, I., Brekelmans, W., Geers, M., 2008a. Computational homogenization for heat conduction in heterogeneous solids. *Internat. J. Numer. Methods Engrg.* 73 (2), 185–204.
- Özdemir, I., Brekelmans, W., Geers, M.G., 2008b. Fe2 computational homogenization for the thermo-mechanical analysis of heterogeneous solids. *Comput. Methods Appl. Mech. Engrg.* 198 (3–4), 602–613.
- Pascanu, R., Mikolov, T., Bengio, Y., 2013. On the difficulty of training recurrent neural networks. In: *International Conference on Machine Learning*. pp. 1310–1318.
- Quintard, M., Whitaker, S., 1993. One-and two-equation models for transient diffusion processes in two-phase systems. 23, Elsevier, pp. 369–464.
- Raissi, M., Perdikaris, P., Karniadakis, G.E., 2019. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.* 378, 686–707.
- Ramos, G.R., Santos, T.dos., Rossi, R., 2017. An extension of the hill-mandel principle for transient heat conduction in heterogeneous media with heat generation incorporating finite RVE thermal inertia effects. *Internat. J. Numer. Methods Engrg.* 111 (6), 553–580.
- Razavi, S., Tolson, B.A., Burn, D.H., 2012. Review of surrogate modeling in water resources. *Water Resour. Res.* 48 (7).
- Royer, P., Auriault, J.-L., 1994. Transient quasi-static gas flow through a rigid porous medium with double porosity. *Transp. Porous Media* 17 (1), 33–57.
- Sánchez-Palencia, E., 1980. Non-homogeneous media and vibration theory. In: *Lecture Notes in Physics*, vol. 127.
- Sangiorgio, M., Dercole, F., 2020. Robustness of lstm neural networks for multi-step forecasting of chaotic time series. *Chaos Solitons Fractals* 139, 110045.

- Su, F., Larsson, F., Runesson, K., 2011. Computational homogenization of coupled consolidation problems in micro-heterogeneous porous media. *Internat. J. Numer. Methods Engrg.* 88 (11), 1198–1218.
- Tahmasebi, P., Kamrava, S., Bai, T., Sahimi, M., 2020. Machine learning in geo-and environmental sciences: From small to large scale. *Adv. Water Resour.* 142, 103619.
- Unger, J.F., Könke, C., 2009. Neural networks as material models within a multiscale approach. *Comput. Struct.* 87 (19–20), 1177–1186.
- Wang, K., Sun, W., 2018. A multiscale multi-permeability poroplasticity model linked by recursive homogenizations and deep learning. *Comput. Methods Appl. Mech. Engrg.* 334, 337–380.
- Warren, J.E., Root, P.J., 1963. The behavior of naturally fractured reservoirs. *Soc. Pet. Eng. J.* 3 (03), 245–255.
- Waseem, A., Heuze, T., Stainier, L., Geers, M., Kouznetsova, V., 2020. Model reduction in computational homogenization for transient heat conduction. *Comput. Mech.* 65 (1), 249–266.
- Weinan, E., 2011. *Principles of Multiscale Modeling*. Cambridge University Press.
- Whitaker, S., 1999. *The Method of Volume Averaging*, Vol. 13. Springer Science & Business Media.
- Xiao, N.-C., Zuo, M.J., Zhou, C., 2018. A new adaptive sequential sampling method to construct surrogate models for efficient reliability analysis. *Reliab. Eng. Syst. Saf.* 169, 330–338.
- Zhou, Q., Oldenburg, C.M., Spangler, L.H., Birkholzer, J.T., 2017. Approximate solutions for diffusive fracture-matrix transfer: Application to storage of dissolved CO₂ in fractured rocks. *Water Resour. Res.* 53 (2), 1746–1762.
- Zhu, Y., Zabarab, N., Koutsourelakis, P.-S., Perdikaris, P., 2019. Physics-constrained deep learning for high-dimensional surrogate modeling and uncertainty quantification without labeled data. *J. Comput. Phys.* 394, 56–81.
- Zimmerman, R.W., Chen, G., Hadgu, T., Bodvarsson, G.S., 1993. A numerical dual-porosity model with semianalytical treatment of fracture/matrix flow. *Water Resour. Res.* 29 (7), 2127–2137.