

# Explicit physics-informed neural networks for nonlinear closure: The case of transport in tissues

Ehsan Taghizadeh<sup>a</sup>, Helen M. Byrne<sup>b</sup>, Brian D. Wood<sup>a,\*</sup>

<sup>a</sup> Chemical, Biological, and Environmental Engineering, Oregon State University, OR 97331, USA

<sup>b</sup> Mathematical Institute, University of Oxford, Oxford OX2 6GG, UK

## ARTICLE INFO

### Article history:

Available online 29 October 2021

### Dataset link:

<https://github.com/TaghizadehE/explicit-PINN>

### Keywords:

Explicit physics-informed neural networks  
Deep learning  
Tissue transport  
Nonlinear kinetics  
Upscaling  
Effectiveness factor

## ABSTRACT

In upscaling methods, closures for nonlinear problems present a well-known challenge. While a number of theoretical methods have been proposed for handling such closures, nonlinearities still remain a significant obstacle for many problems. In this work, we use a combination of formal upscaling and data-driven machine learning for explicitly closing a nonlinear transport and reaction process in multiscale tissues. The classical effectiveness factor model is used to formulate the macroscale reaction kinetics. We train a multilayer perceptron network using training data generated by direct numerical simulations over microscale examples. Once trained, the network is used in an algorithm for numerically solving the upscaled (coarse-grained) differential equation describing mass transport and reaction in two example tissues. The network is described as being *explicit* in the sense that the network is trained using macroscale concentrations and gradients of concentration as components of the feature space rather than incorporating them as part of a constraint in the optimization process.

Network training and solutions to the macroscale transport equations were computed for two different tissues. The two tissue types (brain and liver) exhibit markedly different geometrical complexity and spatial scale (cell size and sample size). The upscaled solutions for the average concentration are compared with numerical solutions derived from the microscale concentration fields by *a posteriori* averaging. There are three outcomes of this work of particular note.

1. Our overall approach results in an upscaled nonlinear PDE. The PDE is closed using a neural network, and our approach results in the definition of the classical effectiveness factor for effecting closure.
2. We identify particular *source terms* for the closure problem that are important for representing the structure of the closure. These source terms involve macroscale concentrations and their gradients. We adopt these source terms to use as explicit features in the learning algorithm. We find the trained networks that include the macroscale source terms generate models that are able to predict the correction factor with increased fidelity over those that do not.
3. We find that the trained network exhibits good generalizability, and it is able to predict the effectiveness factor with high fidelity for realistically-structured tissues despite the significantly different scale and geometrical complexity of the two example tissue types.

\* Corresponding author.

E-mail address: [brian.wood@oregonstate.edu](mailto:brian.wood@oregonstate.edu) (B.D. Wood).



This latter result emphasizes our purposeful connection between conventional averaging methods with the use of machine learning for closure; this contrasts with some machine learning methods for upscaling where the exact form of the macroscale equation remains unknown.

© 2021 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

Developing closures associated with the upscaling of nonlinear continuum mechanical problems is an enduring challenge. One solution to this problem, fostered by increased computational speed and capacity, has been the development of methods to directly resolve all relevant scales of the phenomena of interest. While this has been applied effectively to the problem of, for example, the momentum balances that describe turbulence, the approach is still too cost prohibitive (in terms of computational requirements) to be used routinely for many problems of interest. Thus, there is a continuing need for appropriately upscaled (or *coarse-grained* or *homogenized*) representations of nonlinear continuum mechanical problems.

Upscaling can be accomplished via a number of approaches ranging from formal averaging methods to various numerical schemes; reviews of these methods (with an emphasis on nonlinear problems) have been reported in [1–3]. Regardless of the approach, the process of eliminating the microscale variables in coarse-grained problems is known generally as the *closure problem*. For nonlinear problems, there are no general methods for exact closures. The use of machine learning (ML) methods represents a relatively new option for closing nonlinear problems. With sufficient training data, ML methods (such as any number of neural networks types, support vector methods, etc.) have the ability to *learn* how to represent such data.

In this paper, we focus specifically on coupling a *formal upscaling method* [4] for conducting the coarse-graining of a nonlinear transport and reaction problem, with the use of a deep learning approach for effecting the closure. For this problem, upscaling the transport and reaction problem in biological tissues is investigated by starting at the microscale level of representation of the quantities of interest. We adopt a simple two-phase description of the system with classical hyperbolic kinetics representing the reaction term. Because the reaction term is nonlinear, there is no formal exact scheme for closing the problem. We investigate the use of an artificial neural network (ANN) for learning the *effectiveness factor* representation of the reaction term with high fidelity; specifically, the architecture that we adopt is a classical deep multilayer perceptron (MLP) network. The trained neural network shows high generalizability even for problems with very different scales and geometry than the training data examples, implying that the learned mapping function can be used in a wide range of real applications.

### 1.1. Machine learning in upscaling

Machine learning (ML) is a discipline that has seen a renaissance in applications, mostly due to a combination of increasing processor speeds, computational and storage capacities, and the development of new algorithms. While there are still many questions regarding the interpretability of these methods [5], they also provide one of the few methods that have offered solution options for complex problems, in particular nonlinear ones.

There has been some success in both learning and predicting the complex spatio-temporal behavior of linear and nonlinear partial differential equations (PDEs) [6–10]. A subset of this work has focused on discovering coarse-grained representations of the PDEs under consideration using ML methods [11]. These approaches fall primarily into two major categories.

1. Methods for directly *learning* the appropriate coarse-grained PDE. The number of different approaches pursued here is vast. It includes the use of simple neural networks [12], recurrent neural networks [13], deep neural networks with dimension reduction via diffusion maps [14], recurrent neural network architectures with dimension reduction accomplished via an autoencoder scheme [15], multilayer perceptrons for generating reduced-order metamodels [2], and direct/constrained equation learning methods [16]. Because these models are designed to learn the appropriate macroscale behavior from an appropriate suite of ML approaches, the end result is frequently high-fidelity predictive ability from the generated network, but usually *without* the generation of an explicit macroscale equation for the process.
2. Methods in which the macroscale form of the PDE is specified (via any of a number of formal methods for upscaling), and then the problem is *closed* by learning from a large number of microscale examples. There are occasions where the development of an explicitly-defined macroscale balance equation is desirable. There are a number of classical upscaling methods that have been used to accurately average microscale balance equations to generate macroscale representations in a mathematically formal and robust way (reviews of such approaches can be found in [17,18]). These approaches often use physical information uncovered from the upscaling process (e.g., solution invariant, bounds for the variable scale, smoothness requirements, or the specification of source terms in the closure problem) to constrain the problem. This approach has recently had good success in predicting nonlinear closures for turbulence [19,20].



In this work, we adopt the second of these two approaches, with an application to the coarse-graining of chemical transport and reaction in biological tissues. In this application, coarse graining is conducted via formal volume averaging [4], and the process has many similarities with reduced order modeling (ROM). Although many popular ROM methods rely on expansion in an optimal basis function to generate a subspace with fewer degrees of freedom, in volume averaging one reduces the number of degrees of freedom by passing the underlying equations through a spatial (or temporal-spatial) filter. In this sense, the filtered model accomplishes the same goals as ROM; the high-order microscale model is projected onto a lower-order subspace to remove components of the solution that are considered less critical to representing the underlying physics, leading to a model with a reduced number of degrees of freedom.

Recently, the idea of enforcing constraints imposed by physical conservation laws has been incorporated into ANNs [21]; these networks are referred to as physics-informed neural networks (PINNs). The idea is that any number of known (or assumed) ancillary physical conditions can be implemented as constraints on the neural network by incorporating these additional conditions as Lagrange multipliers in the loss function. Minimization of the loss function then leads to solutions that also impose the applied constraints. This approach is particularly beneficial when the data sets are sparse; for feature combinations that are “far” from known target data points, the approach requires the neural network to at least honor the ancillary physical constraints. Because the dependent variables in these constraints are not necessarily evaluated directly, these conditions can be thought of as *implicit* constraints on the problem. These approaches, generally referred to as physics-informed neural networks, have been very successful in applications to a wide array of problems from fluid mechanics [9] to nonlocal (fractional) conservation laws [22].

In the work reported here, the nature of the workflow allows a neural network design to predict the macroscale nonlinear reaction rate as informed *explicitly* by the underlying physics. Because we consider a two-scale micro-macro system that must be trained using microscale fields, one can also compute the corresponding macroscale average, and gradient of the average as part of the training data set. Thus, these two macroscale terms can be incorporated explicitly as part of the feature space; because these quantities are computed directly from the microscale results, they automatically conserve mass and flux. This is in contrast to conventional PINNs, where the conservation law is usually implemented implicitly as a constraint. Explicit representation of the conservation laws as part of the feature space will not always be practical; rather, this approach is a useful one in problems such as upscaling where the number of features is relatively small. Essentially, the explicit approach represents a trade off requiring *a priori* computations to densely characterize the feature space, versus the more difficult (but much more flexible) problem of allowing for sparser data at the cost of having to compute the balance laws (often involving automatic differentiation algorithms because such laws are frequently PDEs) during the training of the neural network.

The explicit inclusion of dependent variables as a component of the feature space for ANNs has significant history. In an early example, the single-scale problem of learning the heat transfer coefficient was conducted using a simple 1-layer ANN applied to experimental data [23]. In multiscale systems, the explicit use of the upscaled values of the independent variable has only just begun to be recognized as a useful strategy for ML. The work has been done almost exclusively in the problem of turbulence closures, possibly because this represents one of the most critical and enduring nonlinear closure problems that arises in the upscaling of multiscale systems. As an example, the work reported in [24], various combinations of the known values of the spatially filtered velocity and gradient of the filtered velocity are implemented as components of the feature space for learning the closure for a large eddy simulation decomposition of the momentum balance; the inclusion of these terms is shown to improve the predictive fidelity of the resulting ANN model predicting the closure term. The explicit use of upscaled dependent variables has been also been proposed by [19], who parameterized the Reynolds-averaged Navier-Stokes equations using a neural network with a set of 10 features based on the macroscale strain and rotation tensor components. Similar studies continue to be developed using closures via ANN based on explicit inclusions of the macroscale velocity field (e.g., [25,26]).

Outside of turbulence theory, formal upscaling methods have generally not been combined with closure achieved via ANNs. In particular, formal upscaling techniques such as homogenization theory (see the reviews in [27,17]), thermodynamically constrained averaging theory (TCAT) [28], and the method of volume averaging [4] is often applied to problems with both multiple phases and a hierarchy of length scales. The usual approach in these methods for handling nonlinear terms has been to consider various linearizations of the problem to determine bounding behavior (e.g., [29]). To our knowledge, the use of ANNs has not been widely adopted for nonlinear closures associated with the upscaling in multiphase multiscale systems.

## 1.2. Objectives and outline

For continuum mechanical descriptions of biological tissues, microscale representations for describing both physical and biological processes have become significantly more widespread [30]. While sub-cell-level representations of the biophysical and biochemical processes in tissues have several advantages (e.g., allowing direct representation of cells with different phenotypes), such representations suffer from the same problems as those described for turbulence. For many practical applications, both the number of degrees of freedom and the presence of nonlinearities make the direct computation of the continuum mechanics at the microscale impractical [2]. Hence, upscaling methods have been developed and employed to allow resolution in spatial and temporal domains where it is needed, while providing accurate but more economical methods for domains where resolution is not a priority.



We develop a data-driven framework which combines upscaling, microscale numerical solutions, and deep neural networks in order to provide closure for the problem of transport and nonlinear reactions in tissues. Note that the proposed approach can be used for upscaling many nonlinear phenomena, as long as an unclosed mathematical form for the upscaled equation can be derived, and that the required training dataset is computationally feasible. Our approach is based on using the classical *effectiveness factor* [31–33] to define the reaction rate in terms of average concentrations. In this work we do not attempt to learn the form of the upscaled equation using machine learning [12–15,2,16]; rather, we use formal upscaling and an effectiveness factor representation for the nonlinear reaction rate term to determine the upscaled form. The formal closure of this upscaled form is then learned using machine learning methods.

In short, the effectiveness factor can be defined as follows. Consider a nonlinear kinetic reaction rate  $R = R(c)$  which depends upon the concentration,  $c$ . A spatial average of the reaction rate (where  $\langle \cdot \rangle$  represents the spatial averaging operation) is then specified by

$$\langle R(c) \rangle = \eta^*(\mathbf{x}) R(\langle c \rangle) \quad (1)$$

where  $\mathbf{x}$  is a vector of descriptive model parameters determined from the physics of the problem. Here,  $\eta^*(\mathbf{x})$  is a (potentially highly) nonlinear correction factor that accounts for the fact that the concentrations at the cell surface may be significantly different from the average concentration, especially when intracellular reactions are present. In the material following,  $\eta^*$  is a (unknown) function that is to be estimated;  $\eta$  represents the *observed values* computed from microscale simulations that form the training and validation data set, and  $\hat{\eta}$  indicates the *estimates* of  $\eta^*$  from the proposed ANN network. For future reference, particular estimates of the effectiveness factor from microscale numerical solutions are denoted by  $\eta^{(j)}$ .

The effectiveness factor is then learned by using a particular ANN architecture known as a feed-forward multi-layer perceptron (MLP) network to parameterize a least-squares fit of a training data set containing thousands of examples of features ( $\mathbf{x}$ ) and targets ( $\eta^{(j)}$ ). In short, we seek to approximate the function  $\eta^*$  by parameterizing the unknown weights for a high-dimensional *functional approximation*,  $\hat{\eta}$ , to the observed target values,  $\eta^{(j)}$ . The functional can be represented by the following schematic compositional form, indicating  $M$  sequential transformations of an initial linear combination,  $\ell^{(0)}$ , of the feature space  $\mathbf{x}$  (cf. [34])

$$\eta^*(\mathbf{x}) \approx \hat{\eta}(\mathbf{x}) = \ell^{(f)} \circ T^{(M)} \circ \dots \circ T^{(1)} \circ \ell^{(0)}(\mathbf{x}) \quad (2)$$

Each of the  $M$  transformations,  $T^{(i)}$ , is accomplished using any number of possible nonlinear activation (or basis) functions; conventionally these transformation layers are called *hidden layers*. For each such layer  $T^{(i)}$ , there are  $N^{(i)}$  neurons, which may vary from layer to layer depending upon the network structure. The output of each hidden layer is the input for the subsequent layer; usually the input to subsequent layers is constructed as simple linear combinations of the output from the previous layer, but other mappings are possible. The final output is again generated by a linear combination from the output,  $\ell^{(f)}$ , of the last hidden layer. Each layer will generally include one neuron of the  $N^{(i)}$  that represents a constant (bias) term. For ease of visualization, the compositional form of the network functional is usually displayed as a connected graph.

A graphical presentation of the workflow for this process is provided in Fig. 1. Our approach has many similarities with the closure scheme proposed by [35], but with an extension to neural networks for parameterizing the proposed closure and more intensive sampling of the microscale solutions.

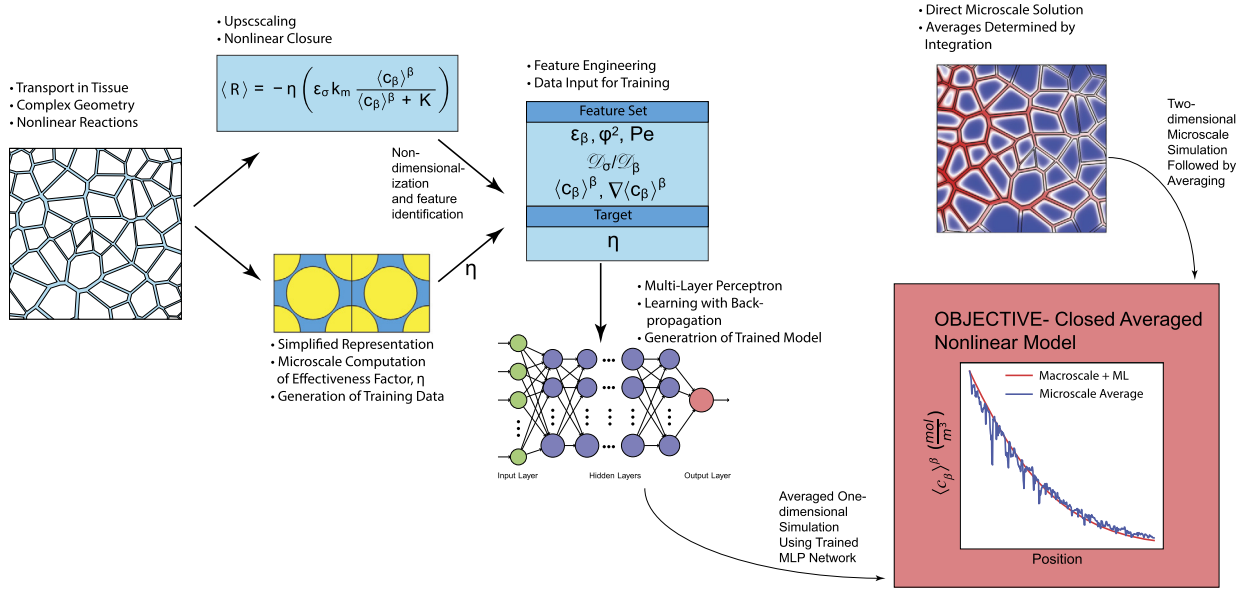
The remainder of the paper is organized as follows. In §1, we outline the sub-cell-scale continuum-mechanical description of mass and momentum transport in a multiphase (intracellular phase, extracellular phase, and cell membrane) tissue. In §2, we summarize previous work in which the microscale problem has been upscaled using the method of volume averaging, introduce the definition of the effectiveness factor, and explain how it can be computed from direct microscale simulation on a representative volume. In §3, we describe the upscaled tissue transport model; most of the results in this section reflect results summarized from previous work by our group. In §4, we describe the identification of the feature sets for learning, emphasizing how these features are selected using physically-driven assessments of the problem. In §5 the MLP network is described, and the process of developing the data sets and training the MLP are explored in detail. In §6, validation of the trained model for the effectiveness factor is conducted by applying the learned function to complex tissues that were not used as part of the training efforts. Finally, in §7, we offer a summary and conclusions.

## 2. Microscale description of tissue transport

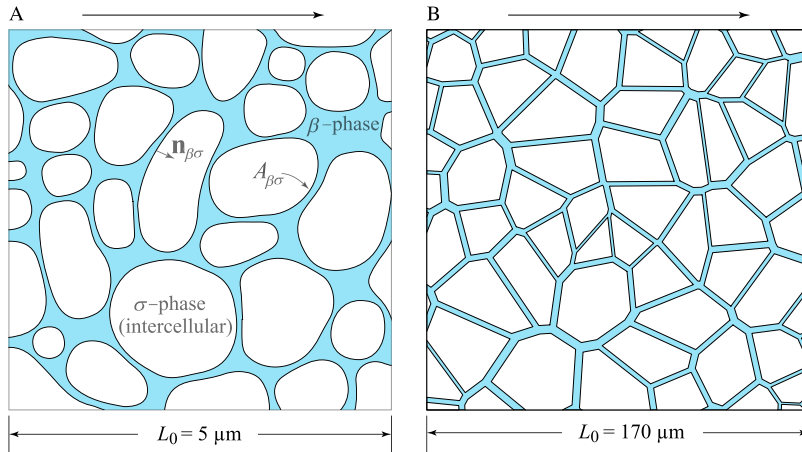
At the microscale level of resolution, we define the tissue system as being comprised of 1) an extracellular phase, 2) an intracellular phase, and 3) a cell membrane separating the two phases. Although the use of microscale models of tissues using continuum partial differential equations has been promoted for some time [36,37], it has only recently started to be routinely adopted as a method for understanding transport and reactions within complex tissue geometries [38–41].

For this research, we have adopted measured geometries that are representative of two substantively different tissue types (Fig. 2): brain cortex tissue [38], and liver lobule tissue [42]. For both tissues, Michaelis-Menten kinetics are adopted to describe the rate of chemical transformation, which is consistent with a number of previous studies [43,42]. Similar sub-cell-scale balance equations have been adopted by researchers studying oxygen transport in mesenchymal stem cells [44],





**Fig. 1.** Graphical representation for the upscaling and closure of transport and nonlinear reactions in biological tissues.



**Fig. 2.** Examples of the geometries of the domain  $\mathcal{V}$  (with superficial volume  $V$ ) showing extracellular ( $V_\beta$ ) and intracellular ( $V_\sigma$ ) phases, and the cell membrane ( $A_{\beta\sigma}$ ). (A) Image for brain tissue (neuropil), adapted from [38]. The extracellular volume fraction is about  $\epsilon_\beta = 0.22$ , which is consistent with estimates from [48]. (B) Image of engineered liver tissue spheroid, adapted from [42]. Arrows above images show direction of mean flow. These two types of tissues have different geometrical complexities.

diffusion of chemicals in brain tissue [38,43,45], drug and oxygen transport in liver lobules [42], transport of cytokines in tumors [46], and movement of solutes in interstitial cell spaces [47] among others.

To begin the process of upscaling, we first assume that a *representative volume* (RV) of tissue exists. This means that the tissue structure is organized spatially such that it is sensible to consider macroscopic (or *effective*) representations of the tissue processes. The details as to the conditions for which an RV exists are usually approached via statistics of the geometric structure. This issue is discussed in additional detail elsewhere [49,50]. Here, we assume an RV exists, and the volume can be divided into the three components, identified as follows (Fig. 2)

$$\mathcal{V}(\mathbf{x}) = \mathcal{V}_\beta(\mathbf{x}) \cup \mathcal{A}_{\beta\sigma}(\mathbf{x}) \cup \mathcal{V}_\sigma(\mathbf{x}) \quad (3)$$

where  $\mathcal{V}_\beta$  and  $\mathcal{V}_\sigma$  represent the extracellular and intracellular volumes, respectively, and  $\mathcal{A}_{\beta\sigma}$  represents the cell membrane separating the two phases. In the remainder of the paper, a subscript  $\beta$  indicates the extracellular phase (usually fluid phase), whereas a subscript  $\sigma$  represents the intracellular phase.

At the microscale, for the momentum balance in the extracellular phase we have adopted the commonly-used Darcy-Brinkman equations [40] as described by Whitaker [4, §4.2.6] as follows.



**Table 1**

Data amalgamated from literature sources for brain and liver tissues. These ranges incorporate both physiologic and experimental conditions [51,42,52–57].

Parameter	Brain	Liver (hepatocyte spheroids)
$\mathcal{D}_\beta \left( \frac{\text{m}^2}{\text{s}} \right)$	$6 \times 10^{-10} - 20 \times 10^{-10}$	$6 \times 10^{-10} - 20 \times 10^{-10}$
$\mathcal{D}_\sigma / \mathcal{D}_\beta$	0.1 – 1.0	0.1 – 1.0
$c_{\max} \left( \frac{\text{mol}}{\text{m}^3} \right)$	0 – 1.8	$1 \times 10^{-3} - 1.0$
$\varepsilon_\beta$	0.23 – 0.49	0.02 – 0.41
$k_m \left( \frac{\text{mol}}{\text{m}^2 \cdot \text{s}} \right)$	0.01 – 1667	$5 \times 10^{-6} - 0.45$
$K \left( \frac{\text{mol}}{\text{m}^3} \right)$	0.003 – 528	$5 \times 10^{-4} - 0.14$
$Pe$	0.01 – 8.0	0.01 – 117
$r_{\sigma,eff} \text{ (m)}$	$0.47 \times 10^{-6}$	$11.7 \times 10^{-6}$
$\varphi^2 \text{ (-)}$	0 – 100	0 – 77
$\kappa \text{ (m}^2\text{)}$	$2 \times 10^{-14} - 2 \times 10^{-8}$	$1 \times 10^{-10} - 7.5 \times 10^{-8}$

### Momentum: Extracellular phase

$$\mu_{eff} \mathbf{K}^{-1} \mathbf{v}_\beta = -\nabla p_\beta + \rho_\beta \mathbf{g} + \mu_\beta \nabla^2 \mathbf{v}_\beta \quad (4)$$

$$B.C. 1 \quad \mathbf{v}_\beta = \mathbf{0}, \text{ at cell surface} \quad (5)$$

Here,  $\mathbf{v}_\beta$  is the (intrinsic) fluid velocity of the extracellular phase,  $\mathbf{K}$  is the Darcy permeability for the extracellular phase,  $p_\beta$  is the pressure in the extracellular phase,  $\mu_\beta$  is the fluid viscosity,  $\mu_{eff} = \phi \mu_\beta$  is the effective viscosity for the Brinkman term,  $\mathbf{g}$  is the gravity vector, and  $\phi$  is the fluid volume fraction of the polymeric substances comprising the extracellular phase.

The microscale mass balance equations can be written as follows.

### Mass: Extracellular phase

$$\frac{\partial c_\beta}{\partial t} = -\mathbf{v}_\beta \cdot \nabla c_\beta + \nabla \cdot (\mathcal{D}_\beta \nabla c_\beta) \quad (6)$$

$$B.C. 1 \quad -\mathbf{n}_{\beta\sigma} \cdot \mathcal{D}_\beta \nabla c_\beta = -\mathbf{n}_{\beta\sigma} \cdot \mathcal{D}_\sigma \nabla c_\sigma, \text{ at cell surface}$$

$$B.C. 2 \quad c_\beta = c_\sigma, \text{ at cell surface} \quad (7)$$

$$I.C. 1 \quad c_\beta(\mathbf{x}, 0) = \mathcal{I}_\beta(\mathbf{x}) \quad (8)$$

### Mass: Intracellular phase

$$\frac{\partial c_\sigma}{\partial t} = \nabla \cdot (\mathcal{D}_\sigma \nabla c_\sigma) - k_m \frac{c_\sigma}{c_\sigma + K} \quad (9)$$

$$I.C. 2 \quad c_\sigma(\mathbf{x}, 0) = \mathcal{I}_\sigma(\mathbf{x}) \quad (10)$$

Note that external boundary conditions are needed to uniquely complete this description. In these expressions,  $c_\beta$  and  $c_\sigma$  are the concentration of the chemical species of interest in each phase, respectively,  $\mathcal{D}_\beta$  and  $\mathcal{D}_\sigma$  are diffusion coefficients of the chemical species in each phase,  $k_m$  is the maximum velocity kinetic coefficient, and  $K$  is the half-saturation constant. The model is fairly general, and the literature contains data for parameterizing for a number of different physiologically-relevant chemical species with these kinetics. The simplest interpretation of this model is one for which the membrane transport is dominated by diffusion rather than mediated by proteins, the diffusive flux is not limiting (compared to the reaction rate), and there is no thermodynamic partitioning between the intracellular and extracellular phases; thus there is no jump in concentration from one side to the other at the membrane interface. More complex models that represent nonlinear membrane transport as are also possible [37], but these would expand the number of independent parameters for the problem.

A sampling of the parameter ranges for the model described above, available from the literature, is given in Table 1 (N.B., the new symbol  $r_{\sigma,eff}$  describes the effective cell length, i.e. effective radius). From this perspective, we do not associate the concentrations  $c_\beta$  and  $c_\sigma$  with any particular chemical species; rather, we require only that the associated parameters adopted fall within the parameter ranges reported in the literature.

## 3. Upscaled tissue model

The macroscale mass balance equations are derived by averaging the microscale equations. This has been done previously via the method of volume averaging by forming appropriately weighted spatial averages over a representative volume of cells



and interfaces (e.g., Fig. 2); the results are reported in detail elsewhere [37,29,36]. The general result of upscaling in two-phase systems is the development of two macroscale models (one for each phase). However, such two-phase models are not necessarily the most frequently utilized or the most efficient model to adopt. In this section, we outline the upscaling process and associated constraints where a one-equation model can be derived to represent the tissue mass balance. In the remainder of this section, we draw on work developed in detail previously by our group [37,58,29,59,60]. Extensive details regarding the specifics of the upscaling and closure (for generally linearized processes) can be found in those references. We do point out that, while the upscaling approach is detailed fully in previous work, the specific method used to represent and close the reaction rate term presented in this study (i.e., a nonlinear closure involving the effectiveness factor) has not been previously detailed. In particular, this work was initiated to address some of the shortcomings noted for high reaction rates in those previous efforts (e.g., see the discussions of the reaction rate limitations described in [37,29]). We note that the classical separation of length scales,  $\ell \leq r_0 \leq L$  in the physical structure is assumed. Here,  $\ell$  is the characteristic length scale for the microscale structure (e.g.,  $\ell$  is on the order of a cell diameter),  $r_0$  is the characteristic size associated with an averaging volume (defined below), and  $L$  is the characteristic length associated with averaged properties in the tissue (this would be on the order of the scale of an organ).

### 3.1. One-equation model

While two-equation models can represent a wide range of tissue behaviors, they are significantly less efficient to use in applications than are single-equation models. One resolution to this problem is to develop a *one-equation macroscale* model that eliminates the coupling between the phases by positing an effective rate of reaction depending only on 1) the model parameters, and 2) the concentration field of the extracellular phase (Eq. (1)). Following results reported previously [37,29], the one-equation model takes the form

$$\frac{\partial \langle c_\beta \rangle^\beta}{\partial t} = -\langle \mathbf{v}_\beta \rangle^\beta \cdot \nabla \langle c_\beta \rangle^\beta + \nabla \cdot (\mathbf{D}^* \cdot \nabla \langle c_\beta \rangle^\beta) + \varepsilon_\beta^{-1} \langle R \rangle \quad (11)$$

Here, the volume average over the extracellular phase (the  $\beta$  phase occupying  $V_\beta$ ) is defined by the intrinsic average

$$\langle c_\beta \rangle^\beta|_{(\mathbf{x},t)} = \varepsilon_\beta^{-1} \int_{\mathbf{r} \in \mathcal{V}(\mathbf{x})} w(\mathbf{x} - \mathbf{r}) c_\beta(\mathbf{r}, t) I_\beta(\mathbf{r}) dV(\mathbf{r}) \quad (12)$$

where  $w$  represents a spatial weighting function. The characteristic size  $r_0$  defined above measures the width of this function. Note that a *superficial* volume average can also be defined; the two averages are related through the volume fraction by  $\langle c_\beta \rangle^\beta = \varepsilon_\beta^{-1} \langle c_\beta \rangle$ . Analogous definitions hold for the intracellular ( $\sigma$ ) phase. For the equations above,  $\langle \mathbf{v}_\beta \rangle^\beta$  is the intrinsic average fluid velocity in the extracellular phase,  $\mathbf{D}^*$  is the effective dispersion tensor for the extracellular phase,  $\varepsilon_\beta$  is the volume fraction of the extracellular phase,  $\varepsilon_\beta^{-1} \langle R \rangle$  is the averaged reaction rate,  $w$  is a compact spatial weighting function, and  $I_\beta$  is a phase indicator function (where  $I_\beta = 1$  for spatial points in  $V_\beta$ , and zero otherwise). By definition the two volume fractions are related by  $\varepsilon_\sigma = 1 - \varepsilon_\beta$ . Additional details regarding the upscaling process that leads to Eq. (11) can be found in the previous literature on volume averaging in biological systems [37,58,29,59,60].

There are practical reasons that one might adopt the one-equation (rather than two-equation) approach. For example, in real-world experiments, the extracellular concentration is often more readily measurable than the intracellular concentration (cf. reference [61, Chp. 6] and reference [32, Chp. 10]). From the computational standpoint, the use of a single-phase model decreases the number of independent balance equations that must be solved (and, hence, the number of degrees of freedom). Thus, alternatives to the two-equation model have substantial practical value. Macroscale equations using the effectiveness factor approach have been widely adopted for representing the transport and reaction process in tissues [62,44,63–68].

### 3.2. Computing the effectiveness factor and effective dispersion tensor

The superficial average reaction rate within the volume,  $\mathcal{V}(\mathbf{x})$ , is found by computing the integral

$$\langle R \rangle = -k_m \int_{\mathbf{r} \in \mathcal{V}(\mathbf{x})} w(\mathbf{x} - \mathbf{r}) I_\sigma(\mathbf{r}) \frac{c_\sigma(\mathbf{r})}{c_\sigma(\mathbf{r}) + K} dV(\mathbf{r}) \quad (13)$$

This provides a scheme for defining the effectiveness factor. We begin with the *proxy* average reaction rate expression,  $R_0$

$$R_0 = -\left( \varepsilon_\sigma k_m \frac{\langle c_\beta \rangle^\beta}{\langle c_\beta \rangle^\beta + K} \right) \quad (14)$$

Clearly,  $R_0$  is generally not the correct reaction rate (although it may be correct in certain limiting regimes). This definition is motivated by simply replacing the concentration terms (inside the intracellular phase,  $c_\sigma$ ) appearing in the non-linear



reaction term in Eq. (9) with the intrinsic average concentration,  $\langle c_\beta \rangle^\beta$ , in the one-equation model. It is apparent that this definition needs a correction term that can generate the actual reaction rate given by Eq. (13). Hence, the expression is adjusted by the effectiveness factor,  $\eta$ . This defines the effective rate of reaction by the relationship [33]

$$\langle R \rangle = \eta^* R_0 \quad (15)$$

$$= -\eta^* \left( \varepsilon_\sigma k_m \frac{\langle c_\beta \rangle^\beta}{\langle c_\beta \rangle^\beta + K} \right) \quad (16)$$

In this work, there are two effective parameters to be determined via closure,  $\mathbf{D}^*$  and  $\eta^*$ , using the solutions of Eqs. (5)–(10). The closure for the effective dispersion coefficient,  $\mathbf{D}^*$ , is a linear problem that has been extensively studied [4], and in most cases its prediction is decoupled from reaction effects. For isotropic media, the dispersion tensor is primarily a function of a two parameters, the Péclet number and the extracellular volume fraction,  $\epsilon_\beta$ . The computation for  $\mathbf{D}_A^*$  can be accomplished following a conventional analysis as outlined by Whitaker [4, Chp. 3]. For completeness, this computation is outlined in the Appendix.

The challenge at this juncture is to determine an appropriate method for predicting the effectiveness factor. For linear (or linearized) problems, closure can be achieved by a sequence of algebraic manipulations of the microscale and macroscale balance equations (cf. reference [29]). For nonlinear problems, there are no general methods; closures are usually developed through various linearizations of the problem to provide asymptotically valid solutions. If the goal of the closure is to obtain results that are valid under general conditions, one must resort to numerical approaches to compute  $\eta^*$ .

Unlike the effective dispersion tensor, the effectiveness factor is a complex nonlinear function of the average concentration fields, the kinetic rate parameters ( $k_m$  and  $K$ ) and the transport parameters  $\mathbf{D}_A^*$ , the average velocity  $\langle \mathbf{v}_\beta \rangle^\beta$ , and the geometry of the problem. Thus, the function defining  $\eta^*$  is of significantly higher dimension (i.e., it is described by more than one or two independent variables) than the one defining the effective dispersion tensor. Parsing out the particular independent variables on which  $\eta$  depends is discussed under the section on physics-driven feature selection; the details of feature engineering are described in §4.

### 3.3. Algorithm for computing the ensemble effectiveness factors for training

For a specified realization of the tissues involved in this work, the macroscale effectiveness factor can be numerically computed by solving the microscale balance equations (Eqs. (5)–(10)). To compute the example target data,  $\eta^{(j)}$ , one needs to first compute the microscale concentration fields over a representative region,  $\mathcal{V}(\mathbf{x})$ , for a specified set of parameters. One algorithm for computing  $\eta$  is as follows.

1. Within an averaging volume, compute the actual rate of reaction  $\langle R \rangle$ , from the microscale concentration field, as indicated by Eq. (13).
2. Compute the intrinsic average concentration,  $\langle c_\beta \rangle^\beta$ , in the extracellular phase within  $\mathcal{V}(\mathbf{x})$ .
3. Compute  $R_0$  using Eq. (14).
4. Compute *observations* of the effectiveness factor, from  $\eta^{(j)} = \langle R \rangle / R_0$ .

There is one additional facet to this computation that requires brief discussion. In general, the effectiveness factor is a function of the (generally transient) concentration field. However, there are often significant differences in the magnitude of the characteristic time scales for the microscale ( $t^*$ ) and macroscale ( $T^*$ ) processes (cf. references [4,37,29]) such that  $t^* \ll T^*$ . In other words, a small perturbation in any macroscale parameter (on time scale  $T^*$ ) is rapidly relaxed at the microscale (taking time  $t^*$ ). In such situations, the microscale problem can be treated as a quasi-steady one [29]. This indicates that the effectiveness factor can be computed from the steady-state versions of Eqs. (6)–(10). Regardless of this generality, in many applications the processes involved do occur at steady state. In the remainder of this work, we will focus on predicting the effectiveness factor for quasi-steady-state conditions.

## 4. Physics-driven feature selection and generation of example data ensemble

The ML problem for this application is a supervised learning process via MLP network. The first step in the process is to establish a vector of independent variables,  $\mathbf{x}$ , that determine the unknown function for the effectiveness factor, i.e.,  $\eta^* \approx \hat{\eta}(\mathbf{x})$  (N.B.: the distinction between the feature vector,  $\mathbf{x}$  and the coordinate vector,  $\mathbf{x}$ ). These independent variables are known as *features* in machine learning; we assume that there are  $k$  such features that are identified. The array of all  $j$  examples of  $k$  features is denoted  $\mathbf{X} = \mathbf{x}^{(j,k)}$ . As a reminder of the notation, we define the *observed* training (and validation) set of features and target values by  $\eta^{(j)} = \eta^{(j)}(\mathbf{X}; \theta)$ , where  $\theta$  is an array of parameters (weights and biases) used in the neural network (described in §5).

For this work, the features can be classified into two main categories: 1) features that are independently known parameters (or parameter groupings) from the nondimensionalized microscale balances and boundary conditions describing the system; we refer to this class of parameters as *parametric* physics-driven features, and 2) features that arise as *source terms* that drive the deviation quantities in the PDE describing their balance. Because these source terms are functions of



the macroscopic dependent variable, we refer to them as *macroscopic source term features*. These two types of features are strictly driven by the physics of the problem.

#### 4.1. Explicit parameters

Nondimensionalization of the microscale balance equations leads to a learning framework that is independent of the physical dimensions of the problem; this allows us to learn the model on a geometry from which the *scale* has been removed, increasing the generalizability of the learned model. The choice of brain and liver geometries was based in part on the large difference in their length scales. If a model trained on nondimensionalized data is able to describe the effectiveness factor for both systems, then it suggests that the rescaling has captured, to some extent, the universal behavior. The microscale balance equations were nondimensionalized as follows.

$$\frac{\partial C_\beta}{\partial \tau} = -Pe \frac{\mathbf{v}_\beta}{U} \cdot \nabla C_\beta + \nabla^2 C_\beta \quad (17)$$

$$I.C. 1 \quad C_\beta(\mathbf{Z}, 0) = I_\beta(\mathbf{Z}) \quad (18)$$

$$B.C. 1 \quad -\mathbf{n}_{\beta\sigma} \cdot \nabla C_\beta = -\mathbf{n}_{\beta\sigma} \cdot (D_r \nabla C_\sigma), \text{ at cell surface} \quad (19)$$

$$B.C. 2 \quad C_\beta = C_\sigma, \text{ at cell surface} \quad (20)$$

$$\frac{\partial C_\sigma}{\partial \tau} = D_r \nabla^2 C_\sigma - \varphi^2 \frac{C_\sigma}{C_\sigma + 1} \quad (21)$$

$$I.C. 2 \quad C_\sigma(\mathbf{Z}, 0) = I_\sigma(\mathbf{Z}) \quad (22)$$

where here

$$\begin{aligned} C_\beta &= \frac{c_\beta}{K}; & C_\sigma &= \frac{c_\sigma}{K}; & D_r &= \frac{\mathcal{D}_\sigma}{\mathcal{D}_\beta}; & Pe &= \frac{U r_{\sigma,eff}}{\mathcal{D}_\beta} \\ \varphi^2 &= \frac{k_m r_{\sigma,eff}^2}{K \mathcal{D}_\beta}; & \tau &= \frac{r_{\sigma,eff}^2}{\mathcal{D}_\beta t}; & \mathbf{Z} &= \frac{\mathbf{z}}{r_{\sigma,eff}}; \end{aligned}$$

Here we adopt the coordinate system  $\mathbf{z} = (x, y, z)$  [and its non-dimensional form  $\mathbf{Z} = (X, Y, Z) = (x/r_{\sigma,eff}, y/r_{\sigma,eff}, z/r_{\sigma,eff})$ ], where the  $z$ -axis is aligned with the mean direction of flow. Note that here the gradient operator is the dimensionless one  $\nabla \equiv \frac{\partial}{\partial \mathbf{Z}}$ . By the symmetry of the problem, we have that  $\nabla \langle C_\beta \rangle = (0, 0, \partial \langle C_\beta \rangle / \partial Z)$ . The symbol  $U$  represents the intrinsic volume average of the velocity over the entire domain, i.e.,  $U = (\langle \mathbf{v}_\beta \rangle^\beta \cdot \langle \mathbf{v}_\beta \rangle^\beta)^{\frac{1}{2}}$ . The effects of reaction rate, convection, and diffusion forces are manifest in the dimensionless numbers  $\varphi^2 = k_m r_{\sigma,eff}^2 / (K \mathcal{D}_\beta)$  (computed in the cell phase),  $Pe$  (computed in the extracellular phase), and  $D_r = \mathcal{D}_\sigma / \mathcal{D}_\beta$ , respectively; these quantities are then natural choices of *parametric variables* for the feature set.

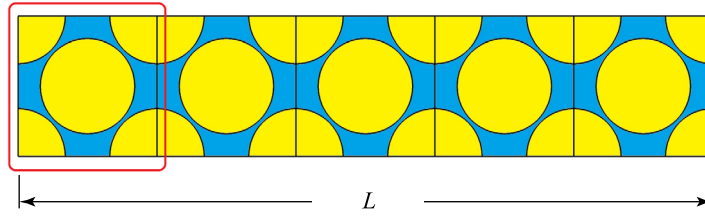
#### 4.2. Explicit macroscale source term features

The additional selection of the *average concentration* and *gradient of the average concentration* as features can be motivated in two ways. First, empirically we know that the effectiveness factor depends directly on the average concentration, as specified by Eq. (15). One might, on this basis, be motivated to try using the first few terms of a Taylor series expansion as features (equivalent to using  $\langle C_\beta \rangle$  and  $\partial \langle C_\beta \rangle / \partial Z$  as features). More directly, however, it can be shown that the average concentration and its gradient appear as *source terms* in the Green's function solution for the effective parameters in linearized versions of this problem. For example the quasi-steady closure for the linearized version of this problem (found by subtracting the average equation from the microscale equation) and the internal boundary condition can be expressed as (after a number of simplifications) [29]

$$\frac{\partial \tilde{c}_\beta}{\partial t} + \mathbf{v}_\beta \cdot \nabla \tilde{c}_\beta + \underbrace{\tilde{\mathbf{v}}_\beta \cdot \nabla \langle c_\beta \rangle^\beta}_{\text{source}} = \nabla \cdot (\mathcal{D}_\beta \nabla \tilde{c}_\beta) + \underbrace{\varepsilon_\beta^{-1} a_v \frac{k_m}{K} \langle c_\beta \rangle^\beta}_{\text{source}} + \varepsilon_\beta^{-1} a_v \frac{k_m}{K} \frac{1}{A_{\beta\sigma}} \int_{\mathbf{r} \in \mathcal{A}_{\beta\sigma}} \tilde{c}_\beta dA \quad (23)$$

$$B.C. 1 \quad -\mathcal{D}_\beta \nabla \tilde{c}_\beta \cdot \mathbf{n}_{\beta\sigma} - \frac{k_m}{K} \tilde{c}_\beta = \underbrace{\mathcal{D}_\beta \nabla \langle c_\beta \rangle^\beta \cdot \mathbf{n}_{\beta\sigma}}_{\text{source}} + \underbrace{\frac{k_m}{K} \langle c_\beta \rangle^\beta}_{\text{source}} \quad (24)$$





**Fig. 3.** The unit cell depicted inside the red box and an array of the unit cell. The flow and cell phases are distinguishable by color. The square's length is  $L_0 = 3 \times 10^{-6}(m)$  and the circle's radius is  $r_{\sigma,eff} = 1.02 \times 10^{-6}(m)$ . One of the advantages of designing a simplified geometry is that it uses less computational power to numerically solve the microscale balance equations. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

The terms denoted *source* in this expression are macroscale terms that arise in the microscale closure problem. When an integral solution in terms of Green's functions is constructed, these source terms appear in convolution integrals with the Green's functions (see Wood and Valdés-Parada [60] for additional details). Ultimately, the behavior of the effectiveness factor is determined by the solution to this closure problem. This establishes the rationale for including both  $\langle c_\beta \rangle^\beta$  and  $\nabla \langle c_\beta \rangle^\beta$  as features for the prediction of  $\eta^*$ .

For nonlinear problems, such an approach will lead to expansions in the averaged concentration that are infinite-order in the spatial derivatives (i.e., these expansions arise from a conventional power series). However, terms with spatial derivatives of order higher than one will lead to a local closed problem that has spatial derivatives of order higher than two. Any finite number of terms with spatial derivatives of order higher than two would lead to consistency problems in the resulting macroscale equation; such equations will not predict strictly positive values for the concentration [69,60]. In general a nonlocal approach (such as a convolutional or fractional derivative formulation) would be required to resolve the expansion. From a practical perspective, if a purely local equation is desired, then the truncation of the series must be done at first-order gradient terms if one is to have consistency in the macroscale concentration balance.

The example given above is strictly a description of how the averaged concentration and its gradient arise as source terms for the closure of the generally nonlinear problem. While no linearization is conducted in the closure described in this paper, the discussion does provide motivation for examining both  $\langle c_\beta \rangle^\beta$  and  $\nabla \langle c_\beta \rangle^\beta$  as features for the purposes of the closure described in §3.2. It also provides strong motivation to suggest that including higher-order terms in this case does not likely lead to improvements in the learning algorithm for predicting  $\eta^*$ , since such terms are ostensibly incompatible with the desired form (a second-order-in-space, local formulation) for the macroscale equation. While nonlocal formulations at the macroscale are possible [60,70], such a formulation is outside the scope of this paper.

#### 4.3. Representative geometry

Appropriate representation of the system geometry in the feature space is a somewhat complex problem. In many important processes in biological systems (e.g., flow in vascular networks), one would expect the details of the geometry to be an important component of the feature set describing the behavior of the system, and this information should be encoded into the feature set. However, for convection-diffusion in nearly isotropic (e.g., close to circular/spherical) geometries, it has been observed that the volume fraction alone provides a good representation of the geometrical information [71]; similarly, simple unit cells have been successfully used to compute the effectiveness factor [29] in systems with transport and reaction. While this is not to say that additional geometrical details could not be extracted by incorporating additional learning methods (e.g., using convolutional neural networks to extract geometrical information from images of representative volumes e.g., Li et al. [72], Wu et al. [73]), for the geometries expressed by these tissues, such additional efforts are an active area of investigation.

Because the geometry is reflected primarily by the porosity of this system, we adopted a *representative* geometry to model the system. The representative geometry provides sufficient similarity to the actual geometrical structure such that it captures the salient characteristics of the more realistic geometries being modeled (such as those presented in Fig. 2), but is simple enough such that the solutions to the unit cell problems are not overly costly to solve numerically. The advantage of using such simplified geometry is that not only numerically solving the microscale balance equations are computationally less expensive, but also the geometry itself does not have singular points such as cusps; the latter avoids possible numerical issues which are common problems in complex structures.

As with other elements of the learning process, the validity of any particular representative structure as a proxy for the actual geometry of the tissues involved can only be checked heuristically. The representative geometry adopted for this work is a simple set of unit cells illustrated in Fig. 3. To avoid the strong influence of the Dirichlet boundary conditions imposed at the inlet of the system, we used a sequence of five unit cells illustrated in Fig. 3 for each simulation, while the data from the first cell was discarded. Note that the values for  $\langle C_\beta \rangle^\beta$  were determined using Eq. (12), where the weighting function,  $w$ , was taken to be a uniform (top-hat) function with width equal to  $L/5$ . Estimates for the derivative  $\frac{\partial}{\partial z} \langle C_\beta \rangle^\beta$  were computed by a simple centered finite difference after computing the average concentration  $\langle C_\beta \rangle^\beta$  in each cell.



#### 4.4. Feature set: computability and consistency

Regardless of any other arguments supporting the set of features described above, it is necessary that the set of features be at least consistent with the available information and physics of the system being described. The microscale parameter set defining the problem physics is given by  $\varphi^2$ ,  $Pe$ ,  $D_r$ , and  $\varepsilon_\beta$ .

As described above, the two macroscale variables (the average concentration and gradient of the average concentration) have clear motivation for use as feature variables when examining the source terms in the differential balances that define the effectiveness factor. From a more practical perspective, one can easily compute estimates for the averaged concentration  $\langle C_\beta \rangle^\beta$  and average gradient  $\frac{\partial}{\partial Z} \langle C_\beta \rangle^\beta$  directly from the ensemble of microscale simulations used to generate the target data. In Fig. 3 we illustrate one example of a unit cell array used in the ensemble of microscale computations. The average concentration is estimated cell-wise by conducting the appropriate integration over each cell in the array. For determining estimates of the gradient, centered finite differences of the averages can then be employed on the spatially averaged concentrations defined for the array.

Based on these considerations, the following feature set was adopted for representation to feed in the MLP: 1)  $\varphi^2$ , 2)  $Pe$ , 3)  $D_r$ , 4)  $\varepsilon_\beta$ , 5)  $\langle C_\beta \rangle^\beta$ , and 6)  $\frac{\partial}{\partial Z} \langle C_\beta \rangle^\beta$ . We note that each of the chosen features has a high correlation with  $\eta$  as shown in Fig. 4 indicating that each contributes significantly to the reduction in variance for the model. Thus, we did not consider feature sets with fewer than the six features identified. Because the features are largely bounded by physically relevant parameters, no specific feature selection algorithms (e.g., LASSO regression [74] for dimension reduction) were required.

#### 4.5. Generation of the example data ensemble

To conduct the training, we developed an extensive data set by high-resolution modeling of the microscale balance equations. There are two common challenges in generating the training data such that the training process properly reflects the underlying functional relationship.

1. The distribution of the dataset should not be highly unbalanced; when the distribution of the dataset is biased or skewed, the results can be poor model performance because important subdomains of the feature space have relatively small representation in the data set.
2. The dataset should sample the feature space densely enough so that the resulting MLP predicts the value of  $\eta$  with high fidelity for any set of points in the feature space.

The first of these issues can be handled by careful generation of the feature space samples. For example, features are usually normalized to occur on an interval from zero to one [75]. For nonlinear problems, there is no obvious (e.g., linear) relationship between the distributions of the parameters and the resulting distribution of the target variables,  $\eta^{(j)}$ ; thus, the distribution of the feature space must be determined empirically. However, starting with close-to-uniform distributions for each feature can provide a reasonable starting point. For our case, we heuristically adjusted the initially uniform distributions of the independent parameters to generate a more uniform set of target variables. In order to get a close-to-uniform distribution for the target value  $\eta$  we found it was helpful to adopt *non-uniform distributions* for  $Pe$  and  $\varphi^2$ .

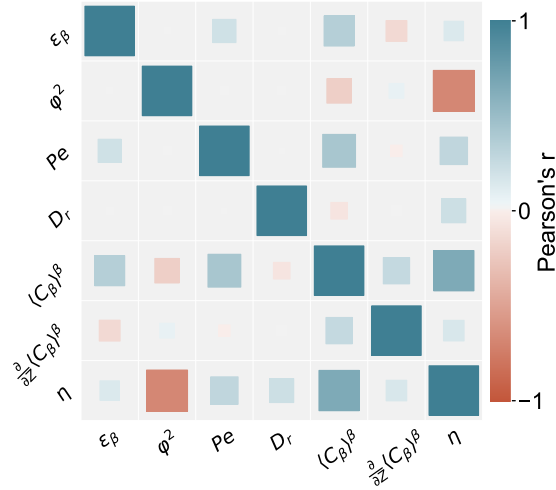
For the second issue, there is a trade off between the sampling density in the ensemble of examples, and the potential for overfitting the data. Generally, more densely sampling the feature space will allow the learned model to have good fidelity (low error) while avoiding overfitting [76].

Table 1 summarizes the range of physiological and experimental data found in the literature; these establish reasonable domains for each feature. Based on this table, we set the domain for training as follows:  $0.25 < \varepsilon_\beta < 0.85$ ,  $0.01 < \varphi^2 < 100$ ,  $0.01 < Pe < 100$ ,  $0.01 < D_r < 1$ , and  $0.1 < C_{max} < 10$  to ensure that they cover the reported literature data for brain and liver. We use the MATLAB's build-in *Latin hypercube* sampling function to generate random distributions, as close to uniform as practical, for  $\varepsilon_\beta$ ,  $D_r$ , and  $C_{max}$ .

To conduct the learning process, we need an appropriate ensemble of known values of  $\eta$  that are associated with a known vector of feature values. To generate such examples, we computed the microscale solutions to the balance equations over the representative geometry illustrated in Fig. 3. Realizations of feature sets were generated randomly from the distributions reported in the previous section. For each such sample of the feature space, we employed the finite element software COMSOL Multiphysics 5.5® to solve the steady-state balance equations implemented on the representative geometry (Fig. 3), and compute the corresponding target values,  $\eta^{(j)}$ .

For the flow problem, the external boundary conditions were periodic on the surfaces perpendicular to mean flow, and specified pressures at the inlet and outlet surfaces. For the mass transport and reaction problem, periodic conditions were used on the external surfaces perpendicular to the mean flow, the inlet boundary was set as a specified concentration, and the outlet boundary was specified by a zero-concentration-gradient condition. Internal boundaries were as specified in Eqs. (6)–(10). We performed a convergence analysis based on Richardson extrapolation on the simulations with the highest convection as well as lowest and highest reaction rates following Roache [77] in order to ensure that the numerical results are stable. We computed the grid convergence index (GCI), which provides a bound on the estimated error of the numerically converged solution, for the simulations. We imposed the condition that the GCI be on the order of  $1 \times 10^{-4}$  or





**Fig. 4.** The Pearson's correlation between each pair of features and target set. Note that the target space ( $\eta$ ) is fully covered. The positive and negative linear correlation are represented as blue and red squares; the bigger square, the strongest linear correlation. The  $\langle C_\beta \rangle^\beta \in$  macroscale features and  $\varphi^2 \in$  parametric features have the most influence on the  $\eta$ .

less, indicating a grid-independent solution. For each simulation (each with a unique set of feature data) we computed the effectiveness factor as described in §3.3.

#### 4.6. Feature correlation structure

To assess the quality of our selected feature set, we computed correlations among the features for the entire set of simulations comprising the training data set. In Fig. 4 the Pearson's correlation between each pair of features is illustrated. Given a pair of features,  $X^k = x^{(j,k)}$  and  $Y^\ell = x^{(j,\ell)}$ , the bivariate correlation, also known as the Pearson product-moment correlation coefficient, is defined by

$$\rho(X^k, Y^\ell) = \frac{\text{cov}(X^k, Y^\ell)}{\sigma_{X^k} \sigma_{Y^\ell}} \quad (25)$$

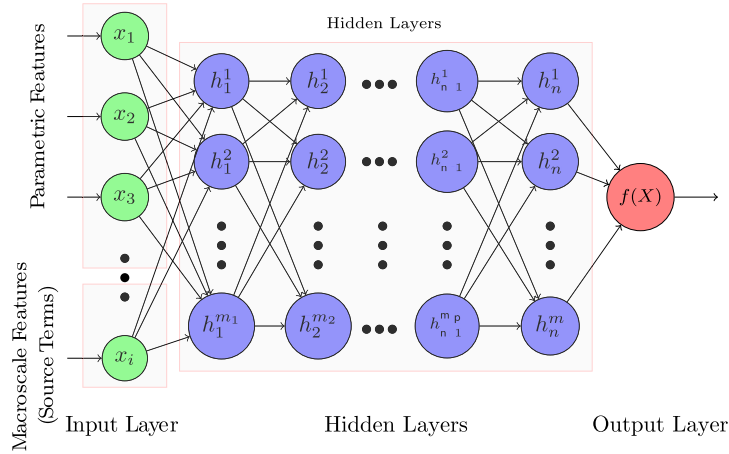
where cov is the covariance and  $\sigma^k$  is the standard deviation for feature  $k$ . The positive and negative linear correlation are represented as blue and red squares; the bigger square, the strongest linear correlation. From the last row (or last column), it is evident that

1. Each of the selected features substantially influence the target value; the corresponding learned weights for each feature set will not be close to zero throughout the learning process.
2. The order in which the features affect  $\eta$  is  $\varepsilon_\beta < \frac{\partial}{\partial z} \langle C_\beta \rangle^\beta < D_r < P_e < \langle C_\beta \rangle^\beta < \varphi^2$  for the generated dataset. These relations reveal the importance of the dimensionless reaction,  $\phi^2 \in$  the set of parametric features.
3. Training with the macroscale source terms as part of the feature space dramatically lowers the residuals for the trained network (see Figs. 7A and C). It is clear that including the source terms (or, in this case, the first two terms in the infinite sequence of source terms) has a significant positive influence on the ability to predict  $\eta^*$  with fidelity.

### 5. Deep feedforward MLP

Our MLP is instantiated by a plain stack of layers, known as a sequential model, where each layer has exactly one input tensor and one output tensor. Our designed network consisted of an input layer with a uniform kernel initializer, 4 dense fully connected hidden layers, and an output layer. The number of neurons in each layer were organized with 4 hidden layers using a conventional powers of two [78] for numbers of neurons; this led to layer widths of 512, 256, 64, and 16 neurons, respectively. A graphical representation of a MLP network using both the parametric features and macroscale source term features is given in Fig. 5. We opted for layers with decreasing numbers of neurons to (in principle) improve performance of the back propagation step. For each hidden layer a rectified linear unit (ReLU) function was utilized as activation function to nonlinearly transform the input functions. The input data were normalized to ensure that they are contained in the same order of magnitude to help prevent overfitting [74]. An adaptive moments ("Adam") gradient-based optimization scheme was adopted, with a learning rate (multiplier,  $\epsilon$ , in the gradient-based optimizer) equal to  $\epsilon = 10^{-3}$  and a time-based decay rate of learning rate/iteration number. For  $N$  training examples, the mean absolute percentage error (MAPE), defined as





**Fig. 5.** A schematic of a multilayer perceptron designed for regression task with  $n$  hidden layers each containing a different number of nodes (neurons) denoted by  $m$ . The MLP used in this work contained 4 hidden layers, each with a nonlinear ReLu (rectified linear unit) transformation. The number of neurons in each of the hidden layers was equal to 512, 256, 64, and 16, respectively.

$$MAPE = \frac{1}{N} \sum_{j=1}^{j=N} \left| \frac{\eta^{(j)} - \hat{\eta}(\mathbf{x}^{(j)}, \theta)}{\eta^{(j)}} \right| \times 100 \quad (26)$$

(where  $\eta^{(j)}$  is the true value from the data vector, and  $\hat{\eta}(\mathbf{x}^{(j)}; \theta)$  is the value predicted for input  $\mathbf{x}^{(j)}$  from the network), was used as the loss function. Note that the model seeks to minimize the defined loss function during the training process. Because our data comes from an ensemble that densely samples the feature space, the data are smooth rather than noisy. We found that regularization of the loss function did not improve the performance of the network, which should be expected for such data.

### 5.1. Network training and structure optimization

For training the network, of the total data set, 76% was used as training data, 4% as validation, and 20% as test data; the later was used to determine the generalization error [74] based on the mean squared error (MSE) defined by

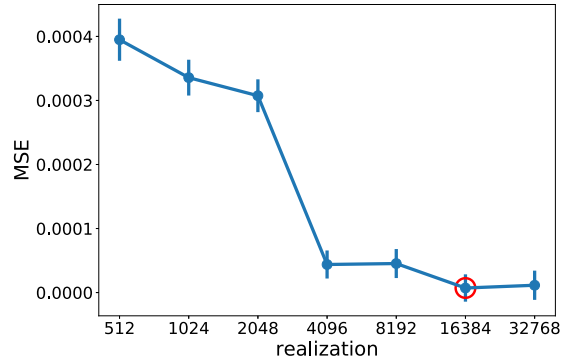
$$MSE = \frac{1}{N} \sum_{j=1}^{j=N} [\eta^{(j)} - \hat{\eta}(\mathbf{x}^{(j)}; \theta)]^2 \quad (27)$$

Note that neither the validation nor the test dataset contribute to the learning. Since the dataset is randomly split into the training, validation, and test components with each run, it yielded very slightly different results; as a result, for the remainder of this study, we present the best results obtained from 10 runs. The tuning of the network's hyperparameters (number of layers, type of activation function, dimensionality of the output space for each layer, etc.) was achieved during the validation step. The maximum number of neurons and number of hidden layers was determined heuristically. We adjusted these values manually over a number of validation runs for our data. A total of 4 layers with the maximum width of 512 neurons led to network performance that resulted in a small residual error but was still parsimonious from the perspective of total number of layers and neurons. In other words, the structure of our data and the small number of features and layers allowed for our network to be optimized using simple steps in which we changed the number of layers and neurons per layer manually. The average running time for the resulting network computing over 2000 epochs was about 1200 seconds on a Geforce GTX 1080 Ti GPU. We used Python 3.7 and Tensorflow 2.3.0 for developing the architecture.

To determine the appropriate size for our ensemble of macroscale simulations, we systematically increased the number of realizations during the Monte-Carlo simulations to find a balance between the computation cost and model accuracy. Fig. 6 shows the MSE against increasing numbers of realizations. Based on this information, we set the number of realizations generating the ensemble of examples to be  $N = 2^{14} = 16384$ . The number of realizations needed to fully cover the feature space is generally determined by some factors such as 1) the number of features in the space, and 2) the spanning width (domain) of each feature. As illustrated in Fig. 6,  $N = 16384$  represents a good compromise between computation costs and minimizing the MSE.

Figs. 7A and C illustrate the loss function versus the number of epochs for the cases where (A) both parametric and macroscale source features and (C) only the parametric features are included. While the network does not suffer from overfitting in either case, the loss of the validation is comparable to those of the training for the case where both the parametric and macroscale source term features are part of the training data. When the macroscale source term features





**Fig. 6.** The MSE vs. the number of realizations. We chose  $N = 16384$  as the number of realizations used in training. Error bars represent standard deviation from the result of 10 independent runs of the learning algorithm.

are excluded from training, case (C), we observe that 1) the loss of the training and validation are noticeably higher than those of case (A) and 2) the validation loss is higher than the training one.

Figs. 7B and D depict the predicted  $\hat{\eta}$  against ground truth  $\eta^{(j)}$  for the two types of tissues studied. When the macroscale source term features are not part of the training set, the network's prediction shows large deviations compared with the case where both the parametric and macroscale source term features are included. The primary significance of these plots is to show that both feature types play important roles in minimizing the loss function and, accordingly, improving the prediction's accuracy. These results are in accordance with the feature space correlation structure reported in Fig. 4, where the  $\eta^{(j)}$  variable is significantly improved by the addition of the macroscale source term features.

The use of macroscale source terms is a somewhat novel aspect of this work. The significance of the source terms in unveiling the underlying physics of the complex transport phenomena has been specifically addressed [79]. We examined the influence of including the macroscale features  $\langle C_\beta \rangle^\beta$  and  $\partial \langle C_\beta \rangle^\beta / \partial Z$  on the resulting error in the trained network; the inclusion of the physically-motivated source terms dramatically improves the performance of the trained network. Without the macroscale features, the MAPE and MSE error are 12% and  $6.0 \times 10^{-3}$ , respectively, whereas when these features are included, the MAPE and MSE are 0.55% and  $7.1 \times 10^{-6}$ , respectively. This implies that both classes of features (parametric and macroscale) are useful for obtaining a model with low training error.

## 6. Testing the learned model

To test the accuracy of the generated MLP closure model for the effectiveness factor (trained on the simplified geometry illustrated in Fig. 3, we computed numerical “experiments” on two tissues (brain and liver) that illustrate realistic structure based on measured tissue geometries. These geometries are based on images collected from tissues in the lab, and the geometrical complexity of these tissues is significantly different than the simple unit cells illustrated in Fig. 3. The goal of the validation was to compare 1) averaged concentrations computed by spatially filtering the direct microscale simulations on the complex geometries, and 2) averaged concentrations computed from the upscaled balance on the simple unit cells combined with the prediction of the effectiveness factor from the trained MLP network. The validation process involved the following sequence of activities.

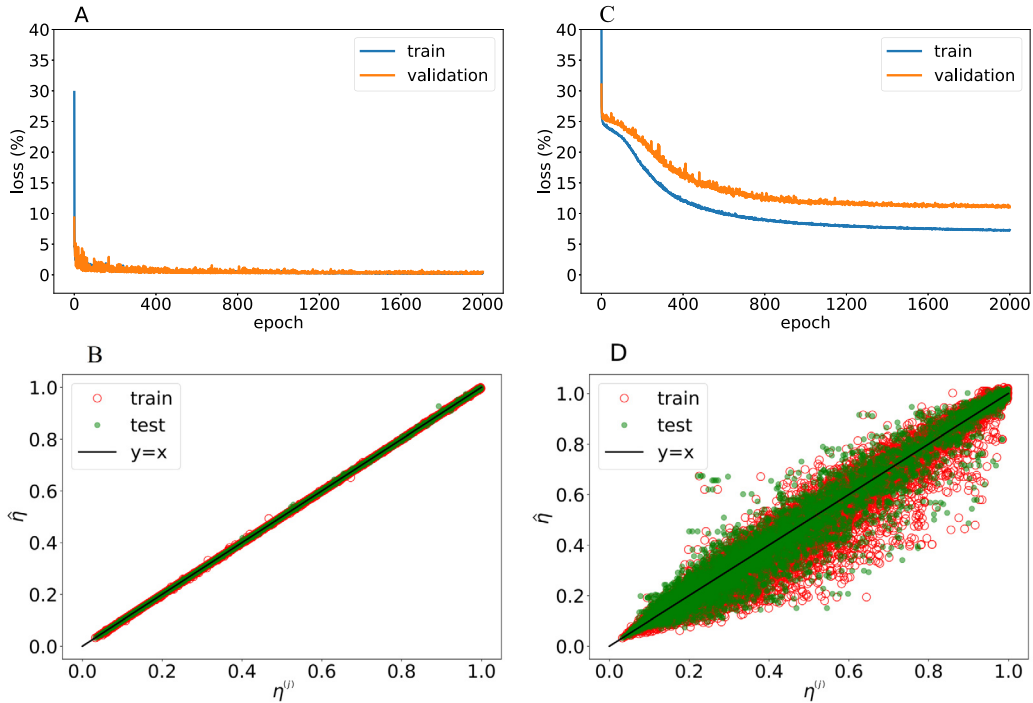
1. Generation of realistic microscale test problems with appropriate parameters and tissue geometry in 2 dimensions (i.e., Table 1 and Fig. 2).
2. Solving the test problems for specified values of the Péclet number, and spatially averaging them along the direction of the mean velocity to develop a 1-dimensional projection of the averaged concentration.
3. Developing and solving the equivalent 1-dimensional representation of the upscaled equation given by Eq. (11). This equation contains a reaction rate term of the form given by Eq. (16). This requires determining estimates of  $\langle C_\beta \rangle^\beta$  and  $\partial \langle C_\beta \rangle^\beta / \partial Z$  during the computation, and appropriately matching the Péclet number on the complex geometries and the training domains.
4. Comparison of the average concentration transects computed directly from the averaged microscale equations, and from the upscaled equation using the trained MLP model for the effectiveness factor.

These steps are described in the material following.

### 6.1. Generating the microscale representations from observed tissue structures

One of the goals of this work is to illustrate that the learned model for the effectiveness factor is robust in the sense that it provides a direct macroscale prediction of the average concentration that is consistent with the concentration fields





**Fig. 7.** (A) and (B): The MAPE loss function and the machine-learned predicted  $\eta$  vs. ground truth  $\eta$ , determined from the microscale simulation, for the case where both macroscale and parametric features are included in the training process. (C) and (D): The MAPE loss function and the machine-learned predicted effectiveness factor ( $\hat{\eta}$ ) vs. ground truth effectiveness factor ( $\eta^{(j)}$ ), determined from the microscale simulation, for the case where only parametric features are included in the training process, i.e. the macroscale features are excluded.

averaged from microscale simulations. Good correspondence between these two measures indicates that the closure has been effectively carried out, and the micro-macro connection is well represented by the closure. Toward that end, we computed the microscale solution for transport and reaction problems on two tissues whose structure was adapted from images reported in the literature.

The detailed 2-dimensional simulations for validation were based on the brain [38] and liver [42] geometries illustrated in Fig. 2; the arrows denote the direction of mean flow (left to right) for these two domains. In order to assure that a large range of concentrations and concentration gradients were generated in our validation simulations, we extended the domain for both geometries using a sequence of reflections and concatenations. First, each domain was reflected about the right-hand axis perpendicular to the mean flow direction to generate a 1 by 2 domain, periodic on the boundaries perpendicular to mean flow. A second such reflection created a 1 by 4 domain, periodic at the inlet (left) and outlet (right). Additionally, to create reasonably unconstrained boundaries in the direction perpendicular to flow, the entire set of 1 by 4 cells was reflected about the axis parallel to flow. The final results were the creation of 4 by 2 arrays of the original geometries as shown in Fig. 8.

Solving the transport problem on these 2D domains was done using the finite element code COMSOL; the approach and error metrics were as described in §4. The set of parameters used for the simulations of brain and liver is provided in Table 2; these data are consistent with values determined from the literature listed in Table 1.

In order to compare the results obtained by the machine learned effectiveness factor with 2-D microscale simulation results (Eqs. (6)–(10)), we calculated the averaged concentration at the cross-sectional direction in the flow phase for 2-D microscale simulations. The values for  $\langle C_\beta \rangle^\beta$  were determined using the average defined by Eq. (12), where the weighting function,  $w$ , was taken to be a uniform (top-hat) function with width equal to  $2r_{\sigma,eff}$  (approximately one cell diameter), and height equal to the domain height.

We present one computation example for each different geometries (brain and liver) using physiologically reasonable parameter values (Table 2). The green lines in Figs. 8B and 8E illustrate the averaged concentration profile generated directly from the microscale solving Eqs. (5)–(10), and averaging these fields a posteriori.

## 6.2. The 1-dimensional averaged equation

We examine the macroscale effectiveness factor under the conditions of steady-state transport and reaction. As previously discussed in §3.3, we assume that there is a separation of time scales between the microscale and macroscale transport processes; that is, we assume  $t^* \ll T^*$ , where  $t^*$  represents the characteristic time scale for microscale processes, and  $T^*$  represents the characteristic time scale for macroscale processes. This is a common approximation for such systems which



**Table 2**  
Parameters used for direct microscale simulations for brain and liver.

Parameter	Brain	Liver
$r_{\sigma, \text{eff}}$ ( $\mu\text{m}$ )	0.47	11.67
$L_0$ ( $\mu\text{m}$ )	5	170
$L$ ( $\mu\text{m}$ )	20	680
$D_\beta$ ( $\frac{\text{m}^2}{\text{s}}$ )	$1 \times 10^{-10}$	$1 \times 10^{-10}$
$D_\sigma / D_\beta$	0.1	0.1
$c_{\text{inlet}}$ ( $\frac{\text{mol}}{\text{m}^3}$ )	2	2
$\varepsilon_\beta$	0.26	0.187
$k_m$ ( $\frac{\text{mol}}{\text{m}^3 \cdot \text{s} \cdot \text{T}}$ )	226.34	1.5
$K$ ( $\frac{\text{mol}}{\text{m}^3}$ )	1	1
$Pe$	8.16	52.09
$\varphi^2$	0.5	2.04
$\kappa$	$1 \times 10^{-10}$	$1 \times 10^{-10}$

have a separation of length scales, and there is substantial analysis to support this approximation (e.g., see Whitaker [4, Chp. 3.3]). If this separation of time scales were not valid, one would require a temporally non-local formulation [60].

Under conditions of a separation of time scales, the learned effectiveness factor should be applicable to both transient and steady transport at the *macroscopic scale* (cf. §3.3 above). However, to avoid the complications of possible different macroscale initial conditions (see Taghizadeh et al. [79] for discussion of the influence of initial conditions), we have opted for a strictly steady-state analysis of the macroscale behavior. Under steady-state conditions, with the average velocity aligned with the  $z$ -axis (direction of mean flow), our closed version of the macroscale transport equation (11) takes the form

$$0 = -U \frac{\partial \langle c_\beta \rangle^\beta}{\partial z} + D_{zz}^* \frac{\partial^2 \langle c_\beta \rangle^\beta}{\partial z^2} - \eta^* \frac{\varepsilon_\sigma}{\varepsilon_\beta} k_m \frac{\langle c_\beta \rangle^\beta}{\langle c_\beta \rangle^\beta + K} \quad (28)$$

This represents the upscaled (coarse-grained) mass balance problem for the tissue. The advantage to using the upscaled balance is that it requires significantly fewer degrees of freedom to resolve compared with the direct microscale solutions in 2 dimensions.

A custom-made steady-state finite difference code was generated to solve the macroscale transport problem directly, using our trained MLP for prediction of the appropriate value of the effectiveness factor,  $\eta$ . For these simulations, the external boundaries were set as follows.

1. Inlet. Dirichlet boundary condition specifying the macroscale concentration.
2. Outlet. Neumann boundary condition imposing a zero-gradient in the concentration.

The solution algorithm used second-order centered differences for spatial derivatives and conventional Picard iteration for solving the system. The stopping criterion for convergence was a global pointwise tolerance of  $\epsilon = 5 \times 10^{-10}$  between successive iterations. The value of  $\eta$  was determined by calling the trained network prediction once every five Picard iterations for the purpose of saving computation time. The value of  $\eta$  was updated using the current value of the features  $\varepsilon_\beta$ ,  $Pe$ ,  $\varphi^2$ ,  $D_\sigma / D_\beta$ ,  $\langle C_\beta \rangle^\beta$ , and  $\frac{\partial}{\partial z} \langle C_\beta \rangle^\beta$ ; note that only the two features based on concentration must be updated during computation. The grid convergence analysis (GCI) was performed in accordance with conventional practice [77]; the resulting GCI was of the order of  $1 \times 10^{-7}$ , indicating that the spatial discretization was sufficiently fine as to produce very little error in the numerical solution.

In addition to the effectiveness factor, the upscaled model also requires the effective dispersion coefficient,  $D_{zz}^*$ . We computed  $D_{zz}^*$  using the conventional methods of volume averaging [4,59]. For completeness, the prediction of  $D_{zz}^*$  is described in detail in the Appendix.

In order to compare the microscale and macroscale computations, it was necessary to match the values of the Péclet number ( $Pe$ ) and the Thiele modulus ( $\varphi^2$ ) between the microscale and macroscale simulations. Each of these unitless numbers require a characteristic length estimate for the effective cell radius  $r_{\sigma, \text{eff}}$ . While  $r_{\sigma, \text{eff}}$  is easy to estimate for the simple learning domain illustrated in Fig. 3, it is challenging to determine it for the complex geometries illustrated in Fig. 2. This is a non-trivial issue, since the accurate values of both  $Pe$  and  $\varphi^2$  are necessary as input (along with the other four features comprising the feature vector) to the trained network so that the effective reaction rate can be accurately predicted. Recall,  $Pe$  and  $\varphi^2$  are defined by

$$Pe = \frac{U r_{\sigma, \text{eff}}}{D_\beta} \quad (29)$$



$$\varphi^2 = \frac{k_m r_{\sigma,eff}^2}{K \mathcal{D}_\beta} \quad (30)$$

Note that each of these parameters requires an estimate of an *effective* value for the cell radius; we denote this quantity by  $r_{\sigma,eff}$ . While each of the other parameters comprising  $Pe$  and  $\varphi^2$  are set independently of the geometry,  $r_{\sigma,eff}$  must come from the geometry itself. The challenge at this juncture is to develop a reasonable method for predicting an effective radius,  $r_{\sigma,eff}$ , for the complex geometries illustrated in Fig. 3.

Following previous work [80], we used a Voronoi decomposition method to determine this value. Voronoi tessellation provides one method to assign an effective radius to complex geometries that are not inherently circular; the approach is well documented elsewhere (e.g., [81, §17.4]). As a summary, given a finite set of points in a plane, for each point the corresponding convex polygons Voronoi cell is formed by all the locations closer to that point than to any of the other points. To compute the effective radius of the tessellated domain, maximum-sized circles were inscribed in each Voronoi cell. The effective radius,  $r_{\sigma,eff}$ , was taken as the *area weighted* radius of these inscribed circles. The resulting values for brain and liver were computed to be  $r_{\sigma,eff} = 0.47$  and  $11.67 \mu\text{m}$ , respectively. With these values established, it was possible to compute  $Pe$  and  $\varphi^2$  uniquely for each of the two domains. This allowed matching both the Péclet number and Thiele modulus, which are necessary parameters for computing  $D_{zz}^*$  and  $\eta^*$ .

### 6.3. Comparison of the results

Fig. 8 shows a comparison of 1) the averaged concentration calculated from the averaged microscale concentration profile, and 2) the concentration obtained by solving the coarse-grained mass balance, Eq. (28), coupled with the trained network for predicting  $\eta^*$ . Figs. 8C and 8F show the corresponding effectiveness factor computed by the trained network. There is generally a good correspondence between the two averaged concentrations. Note that the geometries for the brain and liver systems have dramatically different length scales ( $L_0 = 5$  and  $L_0 = 170 \mu\text{m}$ , respectively). Interestingly, the learned model is able to capture the 1-D upscaled concentration regardless of the difference in their length scales. This is, in part, because we have used the dimensionless feature set to train the model. Even though the model was trained on a resembled geometry in which all the necessary data for training were collected within that geometry, it can predict the upscaled concentrations.

We defined the following error metric for the averaged spatial concentration. First, the average concentration computed from the 2-dimensional microscale solutions were interpolated linearly to the same grid used for the finite-difference computation of Eq. (28). This provided 1) the average concentration predicted from the 2-dimensional microscale simulations, and 2) the average concentration predicted from the macroscale balance given in Eq. (28) on a uniform, 1-dimensional grid with  $N$  total grid points. The  $L^1$  error metric for each point was then computed from

$$\epsilon(Z_i) = \frac{|\langle c_\beta \rangle_{\text{macro+ML}}^\beta - \langle c_\beta \rangle_{\text{micro}}^\beta|}{\langle c_\beta \rangle_{\text{macro+ML}}^\beta} \quad (31)$$

where  $\langle c_\beta \rangle_{\text{macro+ML}}^\beta$  denotes the average concentration predicted from the macroscale balance that uses the trained MLP network for the prediction of the effectiveness factor and  $\langle c_\beta \rangle_{\text{micro}}^\beta$  denotes the average concentration predicted from the 2-dimensional microscale simulations. The average fractional error was then computed from

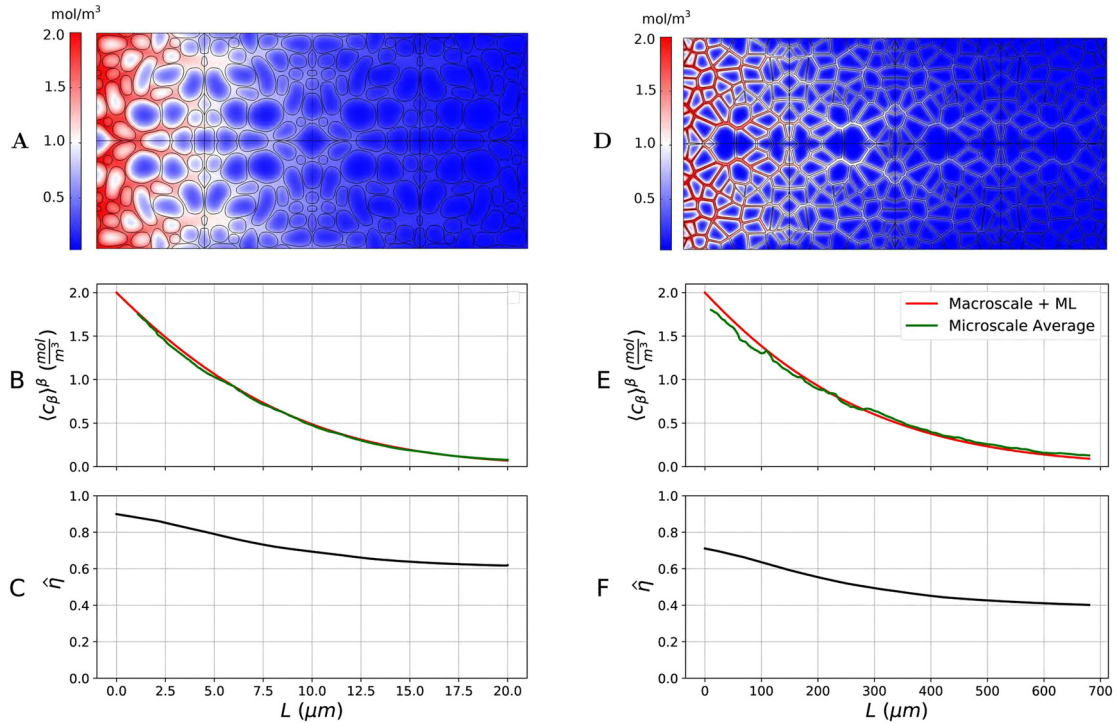
$$\epsilon_{avg} = \frac{1}{N} \sum_{i=1}^N \epsilon(Z_i) \quad (32)$$

This result computes the average error which are 2.8% for the brain and 10% for the liver geometry.

The amount of computational time per iteration for both the microscale and macroscale approaches depends on the particular set of parameters that are adopted for each simulation. However, for the set of parameters used for the brain and liver simulations illustrated in Fig. 8, the ratio of the microscale to macroscale simulation time (roughly the throughput speedup ratio) is on the order of  $t_{\text{micro}}/t_{\text{macro}} \approx 2 - 5$ . Note, these estimates are among the smallest speedup ratios that we should expect because the aspect ratio of the vertical to horizontal sizes is small in our microscale simulations. We would expect these speedup ratios to increase for increasing domain sizes in the vertical direction since the number of unknowns in the microscale simulations would increase substantially, but those of the effective 1-dimensional simulation would remain constant.

One might note that our error metric for training the MLP (MAPE on the order of 0.5%) and the error for the final upscaled solution ( $\epsilon_{avg}$  on the order of 3-10%) are of different orders. There are a number of potential reasons for this difference, but the two primary sources are (1) information loss from upscaling, and (2) overfitting in the MLP. We note that upscaling is, by definition, a lossy process that cannot generally recover exact statistics for complex geometries. This is a problem that is inherent to the nature of upscaling [50]; higher-order statistical correlations are often neglected to make the problem tractable (i.e., to make it possible to close the problem). Although several efforts have been made to reduce overfitting (dense sampling, reducing the number of layers), this is also still a potential source of error in the upscaled numerical simulation. While not further explored here, one might potentially reduce overfitting errors by generating even





**Fig. 8.** (A) and (D): The 2D concentration profiles resulting from solving Eqs. (5)–(10) on a 4 by 2 array of the brain and liver geometries, respectively. (B) and (E): The comparison of the vertically averaged concentrations calculated from the microscale numerical simulations (green line) and concentrations obtained by solving Eq. (28) coupled with the trained network for predicting the effectiveness factor (red line). (C) and (F): The learned effectiveness factors,  $\hat{\eta}$ , as a function of distance along the direction of mean flow.

more densely sampled data (so that erroneous interpolations are discouraged), or by attempting further MLP simplifications so that the resulting network has fewer degrees of freedom. Finally, it is possible that including additional geometric could yield improved performance if this information is highly correlated with the effectiveness factor. While it is true that simple geometries frequently capture the effects of transport-reaction systems with nearly isotropic geometry (cf. Ochoa-Tapia et al. [82]), this is not necessarily the case when there are low values for the porosity of the extracellular phase, high reaction rates, and/or anisotropy [83]. For such cases, it may be useful to include geometric information either by computing additional features from the geometry (two and three-point correlations, etc.), or by direct encoding of the geometry into the ANN. An example of this latter approach has been presented by Wu et al. [73] for the case of diffusion in porous materials via a convolutional neural network (CNN) architecture. Such an approach has the potential to improve the accuracy of the nonlinear closure scheme outlined above.

## 7. Conclusions

In this work, we were able to provide a novel framework to *closure* for upscaling nonlinear problems. In particular, for the problem of upscaling nonlinear, hyperbolic (Michaelis-Menten-type) kinetics in tissues, we were able to present a workflow that defined the upscaling process and the subsequent development and training of a MLP network for closing the upscaled problem. The MLP network was trained on a model with particularly simple geometry; however, this did not appreciably limit the accuracy of the trained network for making predictions in more complex, but nearly isotropic, geometries. This result is consistent with previous observations in upscaled media. The result is significant in that it suggests for nearly isotropic tissue structures, the exact geometry of the tissue (e.g., spherical versus polygonal) is not as important as more basic metrics such as volume fraction. The practicality of this observation is that the trained networks exhibit high generalizability. This strongly suggests (as validated by our results), that a single trained network can be effectively applied for an entire class of problems defined by the form of the nondimensionalized equations, and the rough geometrical attributes. For investigations where an explicit form for the upscaled balance equation is desired, the approach reported here provides one possible route for providing a closed and explicit macroscale balance equation. More importantly, we showed that the macroscale features derived from the source terms appearing in the microscale closure problem are essential for the network in order to predict the correction factor with high fidelity. Although the computational cost of the proposed framework for closing the nonlinear problems is still significant (one has to generate a large number of examples to covering the associated parameter space), once the network is trained, it is able to predict the correction factor for a variety of realistic structures for a wide range of realistic parameters.



Finally, we note that there are several areas for additional study. For problems where the cells are not nearly isotropic, it is likely that the geometric structure of the tissue will be important. Including such information would require adding additional capabilities (and information) to the network. For example, the problem of accounting for geometry more generally could be handled by adding a convolution neural network component [73] to extract the geometric information for processing in parallel with the parametric and source term features.

### CRediT authorship contribution statement

**Ehsan Taghizadeh:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Helen M. Byrne:** Conceptualization, Methodology, Writing – review & editing. **Brian D. Wood:** Conceptualization, Formal analysis, Investigation, Methodology, Supervision, Visualization, Writing – original draft, Writing – review & editing.

### Declaration of competing interest

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

### Data and code availability

All data, finite different, and deep learning codes used in this paper are available at <https://github.com/TaghizadehE/explicit-PINN>.

### Acknowledgement

This work was supported by the School of Chemical, Biological and Environmental Engineering at Oregon State University through the Graduate Teaching Assistantship given to ET, and in part by the US Department of Energy Basic Energy Sciences Program (award DE-SC0021626). BDW acknowledges Steve Whitaker for collaborative work on this topic started several decades ago. We are grateful to Dr. Joseph Leedale (Univ. Liverpool) for graciously sharing his segmented geometries for the hepatocyte spheroids. We thank Dr. Malgorzata Peszynska for her comments on this work.

### Appendix A. Theoretical calculations of the effective dispersion

In 1-D, the macroscale dispersion tensor is represented by two independent parts

$$D_{zz}^* = D_{eff} - \langle \tilde{v}_z b_z \rangle^\beta \quad (A.1)$$

The effective dispersion tensor was determined using conventional volume averaging methods [4, Chp. 3]. The value of  $b_z$  is determined from the following boundary value problem solved over the representative domain (unit cell) described previously by [59]

$$\tilde{v}_z + \mathbf{v}_\beta \cdot \nabla b_z = \mathcal{D}_\beta \nabla^2 b_z \quad (A.2)$$

$$-\mathbf{n}_{\beta\sigma} \cdot \mathcal{D}_\beta \nabla b_z = n_z \mathcal{D}_\beta \quad (A.3)$$

$$b_z(\mathbf{r} + \ell_i) = b_z(\mathbf{r}) \quad (A.4)$$

$$\langle b_z \rangle = 0 \quad (A.5)$$

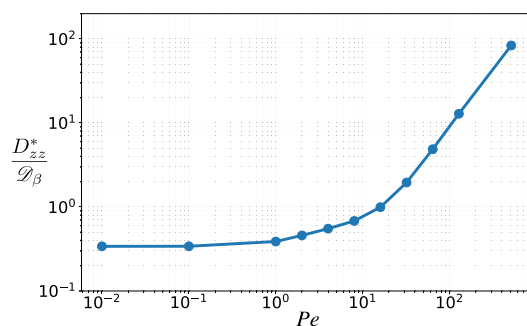
This problem was solved numerically using the same code and approach as described for the solution to the microscale balance equations. The result of this computation is the  $b_z$  field, from which the effective dispersion coefficient  $D_{zz}^*$  is computed by taking the average of the solution as shown in Eq. (A.1).

To compute the effective diffusion coefficient  $D_{eff}$  we adopted the solution described by [82].

$$\frac{D_{eff}}{\mathcal{D}_\beta} = \frac{2\mathcal{D}_\sigma/\mathcal{D}_\beta - \varepsilon_\beta(\mathcal{D}_\sigma/\mathcal{D}_\beta - 1)}{2 + \varepsilon_\beta(\mathcal{D}_\sigma/\mathcal{D}_\beta - 1)} \quad (A.6)$$

where  $\mathcal{D}_\sigma$  and  $\mathcal{D}_\beta$  are extracellular and intracellular diffusion coefficient, respectively. The effective diffusion coefficient can also be computed slightly more accurately by solving a *closure problem* as described in [4, Chp. 1]. Because the diffusion coefficient is not strongly affected by geometry for nearly isotropic materials, the analytical approximation was determined to be sufficiently accurate. The normalized (by  $\mathcal{D}_\beta$ ) effective diffusion coefficient is depicted in Fig. A.9. Note that this analysis assumes that the dispersion tensor is not significantly affected by the reaction process. For cases where the reaction affects the dispersion tensor, this solution would need to be reformulated to couple the reaction and dispersion processes.





**Fig. A.9.** Theoretical calculations of the effective dispersion vs. Pé for the representative unit cell. Note that the normalized effective diffusion coefficient ( $D_{zz}^* / \mathcal{D}_\beta$ ) is depicted in Fig. A.9. This analysis assumes that the dispersion tensor is not significantly affected by the reaction process.

## References

- [1] K. Matouš, M.G. Geers, V.G. Kouznetsova, A. Gillman, A review of predictive nonlinear theories for multiscale modeling of heterogeneous materials, *J. Comput. Phys.* 330 (2017) 192–220.
- [2] L. Burzawa, L. Li, X. Wang, A. Buganza-Tepole, D.M. Umlis, Acceleration of PDE-based biological simulation through the development of neural network metamodels, *Curr. Pathobiol. Rep.* (2020) 1–11.
- [3] G.C. Peng, M. Alber, A.B. Tepole, W.R. Cannon, S. De, S. Dura-Bernal, K. Garikipati, G. Karniadakis, W.W. Lytton, P. Perdikaris, et al., Multiscale modeling meets machine learning: what can we learn?, *Arch. Comput. Methods Eng.* (2020) 1–21.
- [4] S. Whitaker, *Theory and Applications of Transport in Porous Media: The Method of Volume Averaging*, Kluwer Academic Publishers, The Netherlands, 1999.
- [5] C. Molnar, *Interpretable Machine Learning*, <https://christophm.github.io/interpretable-ml-book/>, 2019.
- [6] R. Ranade, C. Hill, J. Pathak, Discretizationnet: a machine-learning based solver for Navier-Stokes equations using finite volume discretization, *arXiv:2005.08357*, 2020.
- [7] J. Zhuang, D. Kochkov, Y. Bar-Sinai, M.P. Brenner, S. Hoyer, Learned discretizations for passive scalar advection in a 2-D turbulent flow, *arXiv:2004.05477*, 2020.
- [8] S.H. Rudy, S.L. Brunton, J.L. Proctor, J.N. Kutz, Data-driven discovery of partial differential equations, *Sci. Adv.* 3 (2017) e1602614.
- [9] M. Raissi, G.E. Karniadakis, Hidden physics models: machine learning of nonlinear partial differential equations, *J. Comput. Phys.* 357 (2018) 125–141.
- [10] M. Raissi, P. Perdikaris, G.E. Karniadakis, Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, *J. Comput. Phys.* 378 (2019) 686–707.
- [11] Y. Bar-Sinai, S. Hoyer, J. Hickey, M.P. Brenner, Learning data-driven discretizations for partial differential equations, *Proc. Natl. Acad. Sci. USA* 116 (2019) 15344–15349.
- [12] S. Pawar, S.E. Ahmed, O. San, Interface learning in fluid dynamics: statistical inference of closures within micro-macro-coupling models, *Phys. Fluids* 32 (2020) 091704.
- [13] H. Arbab, J.E. Bunder, G. Samaey, A.J. Roberts, I.G. Kevrekidis, Linking machine learning with multiscale numerics: data-driven discovery of homogenized equations, *arXiv:2008.11276*, 2020.
- [14] S. Lee, M. Kooshkbaghi, K. Spiliotis, C.I. Siettos, I.G. Kevrekidis, Coarse-scale PDEs from fine-scale observations via machine learning, *Chaos* 30 (2020) 013141.
- [15] P.R. Vlachas, G. Arampatzis, C. Uhler, P. Koumoutsakos, Learning the effective dynamics of complex multiscale systems, *arXiv:2006.13431*, 2020.
- [16] J. Bakarij, D.M. Tartakovsky, Data-driven discovery of coarse-grained equations, *J. Comput. Phys.* (2021) 110219.
- [17] Y. Davit, C.G. Bell, H.M. Byrne, L.A. Chapman, L.S. Kimpton, G.E. Lang, K.H. Leonard, J.M. Oliver, N.C. Pearson, R.J. Shipley, et al., Homogenization via formal multiscale asymptotics and volume averaging: how do the two techniques compare?, *Adv. Water Resour.* 62 (2013) 178–206.
- [18] I. Battiato, D. O'Malley, C.T. Miller, P.S. Takhar, F.J. Valdés-Parada, B.D. Wood, et al., Theory and applications of macroscale models in porous media, *Transp. Porous Media* 130 (2019) 5–76.
- [19] J. Ling, A. Kurzawski, J. Templeton, Reynolds averaged turbulence modelling using deep neural networks with embedded invariance, *J. Fluid Mech.* 807 (2016) 155–166.
- [20] K. Duraisamy, G. Iaccarino, H. Xiao, Turbulence modeling in the age of data, *Annu. Rev. Fluid Mech.* 51 (2019) 357–377.
- [21] M. Raissi, P. Perdikaris, G.E. Karniadakis, Physics informed deep learning (part I): data-driven solutions of nonlinear partial differential equations, *arXiv:1711.10561*, 2017.
- [22] H. Antil, R. Khatri, R. Löhner, D. Verma, Fractional deep neural network via constrained optimization, *Mach. Learn.* 2 (2020) 015003.
- [23] K. Jambunathan, S. Hartle, S. Ashforth-Frost, V. Fontana, Evaluating convective heat transfer coefficients using neural networks, *Int. J. Heat Mass Transf.* 39 (1996) 2329–2332.
- [24] A. Beck, D. Flad, C.-D. Munz, Deep neural networks for data-driven LES closure models, *J. Comput. Phys.* 398 (2019) 108910.
- [25] R. Maulik, O. San, A. Rasheed, P. Vedula, Subgrid modelling for two-dimensional turbulence using neural networks, *J. Fluid Mech.* 858 (2019) 122–144.
- [26] A. Pal, Deep learning emulation of subgrid-scale processes in turbulent shear flows, *Geophys. Res. Lett.* 47 (2020) e2020GL087005.
- [27] B. Hassani, E. Hinton, A review of homogenization and topology optimization I—homogenization theory for media with periodic structure, *Comput. Struct.* 69 (1998) 707–717.
- [28] W.G. Gray, C.T. Miller, *Introduction to the Thermodynamically Constrained Averaging Theory for Porous Medium Systems*, Springer, 2014.
- [29] B.D. Wood, K. Radakovich, F. Golfier, Effective reaction at a fluid–solid interface: applications to biotransformation in porous media, *Adv. Water Resour.* 30 (2007) 1630–1647.
- [30] Y. Wang, E. Brodin, K. Nishii, H.B. Frieboes, S. Mumenthaler, J.L. Sparks, P. Macklin, Impact of tumor-parenchyma biomechanics on liver metastatic progression: A multi-model approach, *bioRxiv*, <https://doi.org/10.1101/2020.05.04.074989>, 2020.
- [31] E.W. Thiele, Relation between catalytic activity and size of particle, *Ind. Eng. Chem.* 31 (1939) 916–920.
- [32] G.A. Truskey, F. Yuan, D.F. Katz, *Transport Phenomena in Biological Systems*, Pearson/Prentice Hall, Upper Saddle River, NJ, USA, 2004.
- [33] M. Shuler, R. Aris, H. Tsuchiya, Diffusive and electrostatic effects with insolubilized enzymes, *J. Theor. Biol.* 35 (1972) 67–76.
- [34] E. Kharazmi, Z. Zhang, G.E. Karniadakis, Variational physics-informed neural networks for solving partial differential equations, *arXiv preprint*, *arXiv:1912.00873*, 2019.



- [35] I.G. Kevrekidis, C.W. Gear, J.M. Hyman, P.G. Kevrekidis, O. Runborg, C. Theodoropoulos, et al., Equation-free, coarse-grained multiscale computation: enabling macroscopic simulators to perform system-level analysis, *Commun. Math. Sci.* 1 (2003) 715–762.
- [36] J.A. Ochoa, S. Whitaker, P. Stroeve, Determination of cell membrane permeability in concentrated cell ensembles, *Biophys. J.* 52 (1987) 763–774.
- [37] B.D. Wood, S. Whitaker, Diffusion and reaction in biofilms, *Chem. Eng. Sci.* 53 (1998) 397–425.
- [38] K.C. Chen, C. Nicholson, Changes in brain cell shape create residual extracellular space volume and explain tortuosity behavior during osmotic challenge, *Proc. Natl. Acad. Sci. USA* 97 (2000) 8306–8311.
- [39] K.E. Holter, B. Kehlet, A. Devor, T.J. Sejnowski, A.M. Dale, S.W. Omholt, O.P. Ottersen, E.A. Nagelhus, K.-A. Mardal, K.H. Pettersen, Interstitial solute transport in 3D reconstructed neuropil occurs by diffusion rather than bulk flow, *Proc. Natl. Acad. Sci. USA* 114 (2017) 9894–9899.
- [40] M. Mosharaf-Dehkordi, A fully coupled porous media and channels flow approach for simulation of blood and bile flow through the liver lobules, *Comput. Methods Biomech. Biomed. Eng.* 22 (2019) 901–915.
- [41] T. Ricken, D. Werner, H. Holzhütter, M. König, U. Dahmen, O. Dirsch, Modeling function–perfusion behavior in liver lobules including tissue, blood, glucose, lactate and glycogen by use of a coupled two-scale PDE–ODE approach, *Biomech. Model. Mechanobiol.* 14 (2015) 515–536.
- [42] J.A. Leedale, J.A. Kyffin, A.L. Harding, H.E. Colley, C. Murdoch, P. Sharma, D.P. Williams, S.D. Webb, R.N. Bearon, Multiscale modelling of drug transport and metabolism in liver spheroids, *Interface Focus* 10 (2020) 20190041.
- [43] E. Vendel, V. Rottschäfer, E.C. de Lange, The need for mathematical modelling of spatial drug distribution within the brain, *Fluids Barriers CNS* 16 (2019) 12.
- [44] F. Zhao, P. Pathi, W. Grayson, Q. Xing, B.R. Locke, T. Ma, Effects of oxygen transport on 3-D human mesenchymal stem cell metabolic activity in perfusion and static cultures: experiments and mathematical model, *Biotechnol. Prog.* 21 (2005) 1269–1280.
- [45] K.E. Holter, B. Kehlet, A. Devor, T.J. Sejnowski, A.M. Dale, S.W. Omholt, O.P. Ottersen, E.A. Nagelhus, K.-A. Mardal, K.H. Pettersen, Interstitial solute transport in 3D reconstructed neuropil occurs by diffusion rather than bulk flow, *Proc. Natl. Acad. Sci. USA* 114 (2017) 9894–9899.
- [46] Y. Kim, A. Friedman, Interaction of tumor with its micro-environment: A mathematical model, *Bull. Math. Biol.* 72 (2010) 1029–1068.
- [47] W.J. Polacheck, J.L. Charest, R.D. Kamm, Interstitial flow influences direction of tumor cell migration through competing mechanisms, *Proc. Natl. Acad. Sci. USA* 108 (2011) 11115–11120.
- [48] J.P. Kinney, J. Spacek, T.M. Bartol, C.L. Bajaj, K.M. Harris, T.J. Sejnowski, Extracellular sheets and tunnels modulate glutamate diffusion in hippocampal neuropil, *J. Comp. Neurol.* 521 (2013) 448–464.
- [49] Y. Bachmat, J. Bear, On the concept and size of a representative elementary volume (REV), in: *Advances in Transport Phenomena in Porous Media*, Springer, 1987, pp. 3–20.
- [50] B.D. Wood, E. Taghizadeh, A primer on information processing in upscaling, *Adv. Water Resour.* (2020) 103760.
- [51] K.A. Lettmann, M. Hardtke-Wolenski, The importance of liver microcirculation in promoting autoimmune hepatitis via maintaining an inflammatory cytokine milieu—a mathematical model study, *J. Theor. Biol.* 348 (2014) 33–46.
- [52] E. Syková, C. Nicholson, Diffusion in brain extracellular space, *Physiol. Rev.* 88 (2008) 1277–1340.
- [53] M.J. Powers, K. Domansky, M.R. Kaazempur-Mofrad, A. Kalezi, A. Capitano, A. Upadhyaya, P. Kurzawski, K.E. Wack, D.B. Stolz, R. Kamm, et al., A microfabricated array bioreactor for perfused 3D liver culture, *Biotechnol. Bioeng.* 78 (2002) 257–269.
- [54] M.K. Sharp, R.O. Carare, B.A. Martin, Dispersion in porous media in oscillatory flow between flat plates: applications to intrathecal, periarterial and paraarterial solute transport in the central nervous system, *Fluids Barriers CNS* 16 (2019) 13.
- [55] C. Debbaut, J. Vierendeels, J.H. Siggers, R. Repetto, D. Monbaliu, P. Segers, A 3D porous media liver lobule model: the importance of vascular septa and anisotropic permeability for homogeneous perfusion, *Comput. Methods Biomech. Biomed. Eng.* 17 (2014) 1295–1310.
- [56] L. Ray, J.J. Iliff, J.J. Heys, Analysis of convective and diffusive transport in the brain interstitium, *Fluids Barriers CNS* 16 (2019) 6.
- [57] R. Maass-Moreno, C.F. Rothe, Distribution of pressure gradients along hepatic vasculature, *Am. J. Physiol., Heart Circ. Physiol.* 272 (1997) H2826–H2832.
- [58] B.D. Wood, M. Quintard, S. Whitaker, Calculation of effective diffusivities for biofilms and tissues, *Biotechnol. Bioeng.* 77 (2002) 495–516.
- [59] B.D. Wood, F. Golfier, M. Quintard, Dispersive transport in porous media with biofilms: local mass equilibrium in simple unit cells, *Int. J. Environ. Waste Manag.* 7 (2011) 24–48.
- [60] B.D. Wood, F.J. Valdés-Parada, Volume averaging: local and nonlocal closures using a Green’s function approach, *Adv. Water Resour.* 51 (2013) 139–167.
- [61] J. Bailey, F. Ollis David, *Biochemical Engineering Fundamentals*, McGraw-Hill Book Company, 1986.
- [62] M.M. Nava, M.T. Raimondi, R. Pietrabissa, A multiphysics 3D model of tissue growth under interstitial perfusion in a tissue-engineering bioreactor, *Biomech. Model. Mechanobiol.* 12 (2013) 1169–1179.
- [63] B. Dey, G.R. Sekhar, S.K. Mukhopadhyay, In vivo mimicking model for solid tumor towards hydromechanics of tissue deformation and creation of necrosis, *J. Biol. Phys.* 44 (2018) 361–400.
- [64] H. Xie, Y. Jiao, Q. Fan, M. Hai, J. Yang, Z. Hu, Y. Yang, J. Shuai, G. Chen, R. Liu, et al., Modeling three-dimensional invasive solid tumor growth in heterogeneous microenvironment under chemotherapy, *PLoS ONE* 13 (2018).
- [65] S. Khakpour, A. Di Renzo, E. Curcio, F.P. Di Maio, L. Giorno, L. De Bartolo, Oxygen transport in hollow fibre membrane bioreactors for hepatic 3D cell culture: a parametric study, *J. Membr. Sci.* 544 (2017) 312–322.
- [66] S. Dasika, S.T. Kinsey, B. Locke, Reaction–diffusion constraints in living tissue: effectiveness factors in skeletal muscle design, *Biotechnol. Bioeng.* 108 (2011) 104–115.
- [67] R.J. Shipley, S.L. Waters, Fluid and mass transport modelling to drive the design of cell-packed hollow fibre bioreactors for tissue engineering applications, *Math. Med. Biol.* 29 (2012) 329–359.
- [68] L.A. Chapman, J.P. Whiteley, H.M. Byrne, S.L. Waters, R.J. Shipley, Mathematical modelling of cell layer growth in a hollow fibre bioreactor, *J. Theor. Biol.* 418 (2017) 36–56.
- [69] R. Pawula, Generalizations and extensions of the Fokker-Planck-Kolmogorov equations, *IEEE Trans. Inf. Theory* 13 (1967) 33–41.
- [70] A. Eringen, *Nonlocal Continuum Field Theories*, Springer-Verlag, New York, 2002.
- [71] M. Quintard, Diffusion in isotropic and anisotropic porous systems: three-dimensional calculations, *Transp. Porous Media* 11 (1993) 187–199.
- [72] A. Li, R. Chen, A.B. Farimani, Y.J. Zhang, Reaction diffusion system prediction based on convolutional neural network, *Sci. Rep.* 10 (2020) 1–9.
- [73] H. Wu, W.-Z. Fang, Q. Kang, W.-Q. Tao, R. Qiao, Predicting effective diffusivity of porous media from images by deep learning, *Sci. Rep.* 9 (2019) 1–12.
- [74] I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning*, MIT Press, 2016.
- [75] J. Friedman, T. Hastie, R. Tibshirani, *The Elements of Statistical Learning*, second edition, Springer-Verlag, New York, 2009, volume 12th corrected printing 2017.
- [76] F. Chollet, et al., *Deep Learning with Python*, vol. 361, Manning, New York, 2018.
- [77] P.J. Roache, Perspective: a method for uniform reporting of grid refinement studies, *J. Fluids Eng.* 116 (1994) 405–413.
- [78] A. Zhang, Z.C. Lipton, M. Li, A.J. Smola, Dive into deep learning, <https://d2l.ai>, 2021.
- [79] E. Taghizadeh, F.J. Valdés-Parada, B.D. Wood, Preasymptotic Taylor dispersion: evolution from the initial condition, *J. Fluid Mech.* 889 (2020) A5.
- [80] S. Ostvar, B.D. Wood, A non-scale-invariant form for coarse-grained diffusion-reaction equations, *J. Chem. Phys.* 145 (2016) 114105.
- [81] S.S. Skiena, *The Algorithm Design Manual*, vol. 1, Springer-Verlag London Limited, 2008.
- [82] J.A. Ochoa-Tapia, P. Stroeve, S. Whitaker, Diffusive transport in two-phase media: spatially periodic models and Maxwell’s theory for isotropic and anisotropic systems, *Chem. Eng. Sci.* 49 (1994) 709–726.
- [83] J.-H. Kim, J.A. Ochoa, S. Whitaker, Diffusion in anisotropic porous media, *Transp. Porous Media* 2 (1987) 327–356.