Transport in Porous Media
https://doi.org/10.1007/s11242-021-01628-9

Upscaling and Automation: Pushing the Boundaries of Multiscale Modeling through Symbolic Computing

An Introduction to **Symbolica**

Kyle Pietrzyk[1] · Svyatoslav Korneev[2] · Morad Behandish[2] · Ilenia Battiato[1]

Received: 2 February 2021 / Accepted: 21 May 2021

Abstract
Macroscopic differential equations that accurately account for microscopic phenomena can be systematically generated using rigorous upscaling methods. However, such methods are time-consuming, prone to error, and become quickly intractable for complex systems with tens or hundreds of equations. To ease these complications, we propose a method of *automatic* upscaling through symbolic computation. By streamlining the upscaling procedure and derivation of applicability conditions to just a few minutes, the potential for democratization and broad utilization of upscaling methods in real-world applications emerges. We demonstrate the ability of our software prototype, Symbolica, by reproducing homogenized advective-diffusive-reactive (ADR) systems from earlier studies and homogenizing a large ADR system deemed impractical for manual homogenization. Novel upscaling scenarios previously restricted by unnecessarily conservative assumptions are discovered and numerical validation of the models derived by Symbolica is provided.

Keywords Upscaling · Homogenization · Symbolic Computing · Reactive Systems

1 Introduction

In recent decades, a significant intellectual investment has been made by the scientific community towards developing *multiscale* abstractions, numerical schemes, and related theories

✉ Morad Behandish
moradbeh@parc.com

✉ Ilenia Battiato
ibattiat@stanford.edu

Kyle Pietrzyk
pietrzyk@stanford.edu

Svyatoslav Korneev
skorneev@parc.com

[1] Department of Energy Resources Engineering, Stanford University, 367 Panama St., Stanford, CA 94305, USA

[2] Palo Alto Research Center, 3333 Coyote Hill Rd., Palo Alto, CA 94304, USA

Published online: 28 August 2021

for physics-based modeling and simulation of porous media—see Battiato et al. (2019) for a broad, yet incomplete, review. Such efforts span scientific fields and engineering applications, contributing their own flavors and specific focuses from areas including energy storage (Li and Monroe 2020), environmental systems (Floudas et al. 2016), and medical sciences (Wood et al. 2002; Walpole et al. 2013). The paradigm shift initiated by multiscale modeling has produced promising results that demonstrate the handling of problems where "classical" single-scale approaches notoriously fail to capture experimentally observed features (e.g., reactions in mixing-controlled chemical transformations (Knutson et al. 2007; Li et al. 2006), anomalous transport (Tartakovsky and Neuman 1998), hysteretic behavior of capillary pressure and relative permeabilities in immiscible two-phase flow (Zagnoni et al. 2010)). Yet, the adoption and systematic utilization of multiscale models by practitioners for the purposes of prediction, design, control, and optimization remain limited. While a combination of often sector-specific factors contribute to this lack in deployment, the inability to reuse or generalize multiscale models without considerable re-derivation for different physical scenarios serves as a common obstacle. As a result, ample time and specialized expertise are required to appropriately formulate each model by navigating the "cost vs. accuracy" trade-offs specific to individual applications. Due to a scarcity in these resources, multiscale models are broadly avoided by practitioners.

Considering applications that require computational efficiency, the re-purposing of high-fidelity multiscale models that resolve multiple spatial and temporal scales would be deemed unaffordable and unnecessary. For example, in vehicle battery management systems, battery models primarily stem from equivalent circuit (EC) theory (Hu et al. 2012) and are governed by ordinary differential equations (ODEs) in time. While high-fidelity multiscale models involving partial differential equations (PDEs) in space and time could capture more accurate depictions of charge and mass transport, EC models are faster to solve and allow for real-time battery usage optimization during driving cycles (Plett 2004). However, with appropriate formulation, spatially-averaged multiscale models governed by ODEs could compete with EC models in functionality, accuracy, and system integration. Because EC models often require parameters obtained through empirical calibration, they do not offer guarantees in predictive error, nor seamless integration into additional systems. On the other hand, multiscale models have the potential to resolve these shortcomings by considering averaged physical behaviors occurring on smaller temporal and spatial scales. Though competitive multiscale models are already available (Smith et al. 2009; Arunachalam et al. 2015, 2017; Moura et al. 2017; Perez et al. 2017), their utilization remains in a primarily academic context, where some level of expertise is necessary to reap their benefits.

At the opposite end of the spectrum, some practical systems are combinatorially complex and require accurate solutions to immense systems of PDEs (e.g., PDEs with numerous coupling source terms). In these applications, computationally efficient models are eclipsed by those that are computationally *affordable*. For example, large chemical reaction networks (CRNs) consisting of hundreds of chemical species are ubiquitous in biogeochemical reactive systems and cellular/biological pathways (Ragsdale and Pierce 2008; Falkowski 2001). To make the simulations of such systems computationally affordable, spatial complexities are typically ignored and CRN topologies are used to generate large systems of coupled ODEs for modeling (Soliman and Heiner 2010; Mutlay and Restrepo 2015). However, this approach becomes inadequate when spatial gradients cannot be ignored or presupposed, e.g., when micro-scale mass transport limitations control the overall system reactivity. For this reason, an alternative modeling strategy often pursued in computational fluid dynamics (CFD) is to fully resolve both spatial and temporal scales, but limit the number of reacting species (Pantano 2004; Maas and Pope 1992). Although this strategy leads to a manageable system

of PDEs, the results may not accurately represent the complexity of the real-world reaction. Therefore, multiscale modeling is suggested as a means to account for both spatial gradients and a large number of reacting species; by averaging complex reactive transport at the micro-scale, accurate and affordable results may be obtained at the macro-scale. However, the inability to re-purpose developed models for general reactions and physical scenarios remains an obstacle.

The limited ability to re-purpose multiscale models in combination with the variety of model design criteria across engineering disciplines demonstrates a need for quick and informed multiscale model development engines. Such tools could reduce, or even eliminate, the required time and specialized expertise preventing practitioners from developing multiscale models according to the cost-accuracy trade-offs specific to individual applications. As a result, the potential to standardize multiscale modeling and disseminate its benefits would be enhanced. Due to its success in other aspects of engineering (e.g., algorithmic mesh generation (Ho-Le 1988) and adaptive grid refinement (Mansell et al. 2002)), we propose *automation* as an engine for generalizing and democratizing multiscale model development. Similar to how computational physics softwares provide communal access to numerical methods, we believe automating model development procedures could instigate the broad adoption of multiscale modeling by practitioners due to rapid and accessible model development strategies.

In recent decades, a number of approaches to multiscale model development known as *upscaling* techniques have gained popularity (e.g., the method of volume averaging (MVA) (Whitaker 1999), thermodynamically constrained averaging theory (TCAT) (Gray and Miller 2014), homogenization methods (Hornung 1997)). These approaches employ rigorous mathematical analyses of PDEs to derive *upscaled equations*, which consist of enhanced mathematical coefficients that account for multiscale behaviors. Developing multiscale models via upscaling hosts a number of advantages for obtaining accurate solutions in an efficient manner. Firstly, *applicability conditions* (or *scaling laws* for MVA (Wood 2009; Golfier et al. 2009), *permissibility conditions* in TCAT (Miller et al. 2018), etc.) that ensure the validity of the "separation of scales" assumption can be formulated with *a priori* error estimates for derived models (Battiato et al. 2009). Secondly, parameter-fitting at coarser scales is not required if sufficient information is known at finer scales (e.g., the unit-cell structure of a porous medium). Thirdly, the possibility to construct multiscale models that seamlessly connect adjacent scales exists (Iliev et al. 2020; Korneev and Battiato 2016), as implied in the left branch of Fig. 1. All of these aspects appeal to diminishing the "trial-and-error" process model developers undergo in practice. Furthermore, upon creating a synergy between upscaling and automation, additional time could be saved due to automatic model derivation. Ultimately, the implications of combining upscaling and automation are far-reaching for practitioners; while accelerated upscaling could provide computationally efficient models for *forward* problems, it could also support the resolving of complex *inverse* problems (e.g., real-time control, iterative design, shape/topology optimization, material discovery, etc.).

Despite the benefits upscaling and automation would bring to multiscale model development, practitioners also require effective strategies for model *deployment*. In the current context, model deployment refers to a model's ability to be implemented and resolved in analysis. Tension has historically existed between the development and deployment stages of modeling due to dissimilar objectives in the academic and application-based environments where models are often developed and deployed, respectively. A consequence stemming from the differing objectives is the inconsistent level of system complexity considered in each stage. This discrepancy has fostered differing foci and desired model features among the two stages of modeling. For example, in physics-based multiscale model development, a general

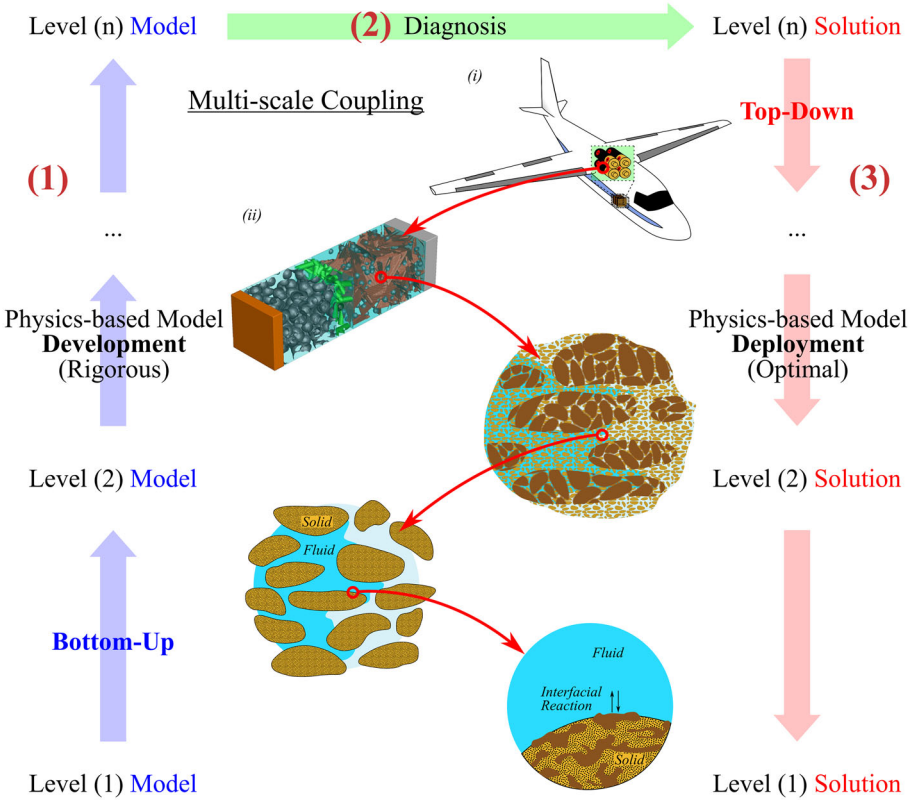

Fig. 1 A sketch of the workflow that connects (1) rigorous model development based on upscaling methods (e.g., volume averaging, thermodynamically constrained averaging, homogenization, etc.) to (3) optimal model deployment at different scales through (2) diagnosis criteria (e.g., applicability conditions), which guide algorithmic refinement strategies within the validity of the different upscaling approximations being used. Inspiration for the airplane image denoted by "*(i)*" was drawn from Locker (2018), the battery schematic denoted by "*(ii)*" is reprinted from Vasilyeva et al. (2018) with permission from Elsevier; and inspiration for the depictions of porous media was drawn from Battiato (2016)

objective is to precisely track the impact of small-scale physical processes across scales in less complex systems (Battiato et al. 2019). In these often academic investigations, *rigor* is desired and upscaling methods are employed analytically or with limited numerical support to develop accurate multiscale models. However, in real-world engineering applications, general objectives involving the efficient analysis of combinatorially and geometrically complex systems would render the development of such models extremely tedious, and their deployment intractable (e.g., PDEs of numerous non-linear terms coupled according to intricate CRNs (Chilakapati et al. 1998) and domains of highly heterogeneous microstructure shape and material properties). Rather than implement a rigorously developed model for complex system analysis, practitioners would prefer models developed for *optimality* in deployment with respect to sector-specific criteria and available resources (e.g., accuracy, tractability, and computational costs). As a result, practitioners are deterred from using multiscale models and upscaling methods, and pursue options better suited for optimal deployment.

The misalignment between idealized and realistic problem setups complicates the cooperation of model development and deployment strategies, especially for upscaling methods. However, automating model development procedures presents a unique opportunity to establish a fruitful connection between the two stages of modeling. While automating model development via upscaling resolves the issue of manually handling lengthy mathematical derivations, the tractability of the resulting upscaled models does not scale well with the system complexities considered upon deployment. Nevertheless, an advantage of automating model development procedures is the ability to "pipeline" the developed models directly into computational solvers using automated *syntax conversion* strategies. In conceptualizing further, an enticing deployment strategy for models after direct transfer to a solver could be envisioned as a "top-down" algorithmic refinement procedure, as portrayed in the right branch of Fig. 1. By exploiting the seamless connections between upscaled models from adjacent scales, finer-scale models could be used only within the spatio-temporal regions where coarser-scale models cease to remain valid (Battiato et al. 2011; Yousefzadeh and Battiato 2017). As a result, computational efficiency and affordability may be obtained while solution accuracy is preserved for complex systems. To diagnose the validity of coarser-scale models, a set of coarser-scale criteria could be defined and used to anticipate the violation of finer-scale applicability conditions from coarser-scale quantities (top branch of Fig. 1). Altogether, the harmony created between accelerated development and deployment strategies with automation would provide an invaluable analysis tool for engineering industries including defense, medicine, and energy, as their applications increasingly rely upon accurate predictions of complex multiscale behaviors in systems involving coupled physicochemical processes.

In light of the foreseen advantages, we aim to encourage the union between automation and multiscale model development and deployment. Easing the bottleneck associated with multiscale modeling for large, complex systems enhances the perceived relevance and penetration of multiscale modeling in the work of practitioners. While previous works have automated the numerical upscaling of material properties (Bahmani et al. 2019), permeability of porous media (Amaziane and Koebbe 2006), and river networks (Wu et al. 2011), we seek a generalized, symbolic approach to integrating automation and multiscale model development. Due to the series of "mechanical" steps found in symbolic upscaling procedures, which render manual upscaling implementations infeasible for complex systems, we pursue a means for automated symbolic upscaling to push the boundaries of multiscale model development.

In this work, we focus on the left branch of Fig. 1 (namely, model development by rigorous upscaling) and employ symbolic computation to automate the homogenization procedure for systems of coupled advective-diffusive-reactive (ADR) equations. For this purpose, we introduce Symbolica, a software for the systematic automation of symbolic upscaling. Based on classical homogenization procedures by the method of multiple-scale expansions (MSE) (e.g., Hornung 1997; Auriault and Adler 1995; Boso and Battiato 2013), Symbolica extends the homogenization procedure of simple ADR systems to ADR systems of arbitrary numbers of species and equations with linear and non-linear bulk and surface reactions. As a result, sets of homogenized equations, their applicability conditions, and corresponding closure problems are achieved in a timely manner with minimal human interaction. By automating upscaling procedures and allocating cumbersome symbolic manipulations to computational resources, large multiscale systems common in engineering applications become more practical to model and easier to optimize.

The manuscript is organized as follows. In Sect. 2, we present the problem formulation for the advective and diffusive transport of N solutes undergoing both homogeneous and heterogeneous reactions in porous media. Basic definitions of averaging operators and derivations

of fast and slow variables are also provided. In Sect. 3, we introduce our software prototype, Symbolica, and discuss its inputs, internal structure, and implemented strategy for symbolic homogenization by means of MSE. In Sect. 4, we provide numerical validation of the homogenized systems derived by Symbolica with a solution comparison between the homogenized models and pore-scale models for three example problems of increasing complexity. While two example problems stem from previously published works, allowing for direct comparison, the third example problem demonstrates Symbolica's ability to handle large systems deemed manually intractable. Additionally, we report Symbolica's uncovering of novel upscaling scenarios previously overlooked due to association with symbolic tedium. In particular, we find that sufficient applicability conditions posed in previous studies using manual upscaling were unnecessarily conservative, leaving additional upscaling pathways unexplored. We conclude with Sect. 5, where we propose generalizations of the program to include automated closure formulations, consideration of higher-order terms, and upscaling in time.

2 Problem Formulation

Let $\hat{\Omega}^\epsilon \subset \mathbb{R}^b$, where $b = \{1, 2, 3\}$, be a porous medium consisting of a multi-connected pore-space $\hat{\mathcal{B}}^\epsilon$, an impermeable solid matrix $\hat{\mathcal{G}}^\epsilon$, and a smooth interface between the two domains $\hat{\Gamma}^\epsilon$. Physical gradients within this medium are assumed to be adequately described on two separate length scales: a larger scale $\hat{\mathcal{L}}$ and a smaller scale $\hat{\ell}$. Then, a length scale ratio

$$\epsilon \equiv \frac{\hat{\ell}}{\hat{\mathcal{L}}} \tag{1}$$

is defined, where the assumption $\hat{\mathcal{L}} \gg \hat{\ell}$ implies $\epsilon \ll 1$. Unless otherwise stated, hats indicate variables with physical dimension, while variables without hats are assumed to be physically dimensionless.

2.1 Governing Equations and Boundary Conditions

We assume the pore-space to be fully saturated with an incompressible liquid whose velocity and pressure fields are governed by the Stokes equation, the incompressible continuity equation, and a no-slip boundary condition, such that

$$\hat{\mu}\hat{\nabla}^2\hat{\mathbf{u}}_\epsilon - \hat{\nabla}\hat{p}_\epsilon = \mathbf{0} \quad \text{in } \hat{\mathcal{B}}^\epsilon, \tag{2a}$$

$$\hat{\nabla}\cdot\hat{\mathbf{u}}_\epsilon = 0 \quad \text{in } \hat{\mathcal{B}}^\epsilon, \tag{2b}$$

$$\hat{\mathbf{u}}_\epsilon = \mathbf{0} \quad \text{on } \hat{\Gamma}^\epsilon, \tag{2c}$$

where $\hat{\mathbf{u}}_\epsilon = \hat{\mathbf{u}}_\epsilon\left(\hat{t}, \hat{\mathbf{x}}\right)$ is the fluid velocity at time $\hat{t} > 0$ and spatial coordinate $\hat{\mathbf{x}} \in \hat{\mathcal{B}}^\epsilon$, $\hat{\mu}$ is the dynamic viscosity, and $\hat{p}_\epsilon = \hat{p}_\epsilon\left(\hat{t}, \hat{\mathbf{x}}\right)$ is the pressure. Here, we do not assume a specific initial condition and only consider regions far from the boundaries of $\hat{\Omega}^\epsilon$, such that non-local effects caused by macro-scale boundary conditions on $\hat{\Omega}^\epsilon$ are negligible.

We consider the mass transport of N solutes subject to advection, diffusion, and both homogeneous and heterogeneous reactions. Without loss of generality, we assume that each species i, where $i \in \{1, 2, \cdots, N\}$, can undergo both bimolecular homogeneous reactions of the type $A + B \leftrightarrow C + D$ in the liquid phase and heterogeneous reactions of the type

$M_{(l)} \leftrightarrow M_{(s)}$ at the liquid-solid interface. Generalization to other types of reactions is straightforward. Within the pore-space, the transport of each reactive species is governed by a system of ADR equations of the form

$$\frac{\partial \hat{c}_\epsilon^{(i)}}{\partial \hat{t}} + \hat{\mathbf{u}}_\epsilon \cdot \hat{\nabla} \hat{c}_\epsilon^{(i)} - \hat{D}^{(i)} \hat{\nabla}^2 \hat{c}_\epsilon^{(i)} = \hat{R}_\epsilon^{(i)} \quad \text{in } \hat{\mathcal{B}}^\epsilon, \tag{3a}$$

$$\hat{R}_\epsilon^{(i)} = \sum_{j=1}^{N} (-1)^{p_L^{(i,j)}} \hat{\mathcal{K}}_L^{(i,j)} \hat{c}_\epsilon^{(j)} + \sum_{j=1}^{N} \sum_{k=j}^{N} (-1)^{p_{NL}^{(i,j,k)}} \hat{\mathcal{K}}_{NL}^{(i,j,k)} \hat{c}_\epsilon^{(j)} \hat{c}_\epsilon^{(k)}, \tag{3b}$$

subject to

$$-\mathbf{n} \cdot \hat{D}^{(i)} \hat{\nabla} \hat{c}_\epsilon^{(i)} = \hat{\mathcal{K}}_S^{(i)} \left(\hat{c}_\epsilon^{(i)^{n_i}} - \hat{C}^{(i)^{n_i}} \right) \quad \text{on } \hat{\Gamma}^\epsilon, \tag{3c}$$

where $\hat{c}_\epsilon^{(i)} = \hat{c}_\epsilon^{(i)}(\hat{t}, \hat{\mathbf{x}})$ is the concentration of species i at time $\hat{t} > 0$ and spatial coordinate $\hat{\mathbf{x}} \in \hat{\mathcal{B}}^\epsilon$, $\hat{D}^{(i)}$ is the diffusion coefficient for species i, $\hat{R}_\epsilon^{(i)}$ is the sum of all bulk reaction terms for species i, $p_L^{(i,j)}$ and $p_{NL}^{(i,j,k)}$ are either 0 or 1, $\hat{\mathcal{K}}_L^{(i,j)}$ is the reaction rate constant of the linear bulk reaction corresponding to species j in equation i, $\hat{\mathcal{K}}_{NL}^{(i,j,k)}$ is the reaction rate constant of the non-linear bulk reaction corresponding to species j and k in equation i, $\mathbf{n}$ is the normal vector to the liquid-solid interface pointed towards the solid, $\hat{\mathcal{K}}_S^{(i)}$ is the reaction rate constant at the liquid-solid interface corresponding to species i, n_i is a positive integer related to the order of reaction, and $\hat{C}^{(i)}$ is the threshold concentration of species i (Morse and Arvidson 2002). Again, we do not assume a specific initial condition and only consider regions far from the boundaries of $\hat{\Omega}^\epsilon$. While Eqs. (2a), (2b), and (2c) may be solved for the fluid velocity and pressure fields, we focus on automating the homogenization procedure of Eqs. (3a), (3b), and (3c) for a given $\hat{\mathbf{u}}_\epsilon$.

2.2 Unit-Cell Domain Formulation

We introduce a spatially-dependent variable $\hat{\boldsymbol{\xi}}(\hat{\mathbf{x}}) \equiv \hat{\mathbf{x}}$. Here, we scale $\hat{\boldsymbol{\xi}}(\hat{\mathbf{x}})$ with $\hat{\ell}$ and $\hat{\mathbf{x}}$ with $\hat{\mathcal{L}}$, such that

$$\boldsymbol{\xi}(\mathbf{x}) = \epsilon^{-1} \mathbf{x}, \tag{4}$$

where $\boldsymbol{\xi}(\mathbf{x})$ and $\mathbf{x}$ are referred to as the "fast" and "slow" variables, respectively (Hornung 1997). Then, we write any spatially-dependent function $f_\epsilon(\mathbf{x})$ as $f_\epsilon(\mathbf{x}) = f(\mathbf{x}, \boldsymbol{\xi}(\mathbf{x}))$ and use the chain rule when considering ∇, the total differential operator in space, to write

$$\nabla f_\epsilon \equiv \nabla_{\mathbf{x}} f + \frac{1}{\epsilon} \nabla_{\boldsymbol{\xi}} f. \tag{5}$$

As shown, ∇ is defined as a sum of two differential operators in space: $\nabla_{\mathbf{x}}$ and $\nabla_{\boldsymbol{\xi}}$, which scale with $1/\hat{\mathcal{L}}$ and $1/\hat{\ell}$, respectively.

Because we seek to homogenize Eqs. (3a), (3b), and (3c) using a "separation of scales" approach, we assume $\epsilon \ll 1$ and narrow our focus to systems accurately described as spatially periodic on the length scale $\hat{\ell}$. Under these restrictions, we treat $\boldsymbol{\xi}(\mathbf{x})$ as an independent variable $\boldsymbol{\xi}$ uncoupled from $\mathbf{x}$, and define $\boldsymbol{\xi}$ in the pore-space $\mathcal{B}$ of a spatially periodic "unit-cell" domain Y. A smooth interface Γ exists within the unit-cell domain between $\mathcal{B}$ and the solid impermeable matrix of the unit-cell, $\mathcal{G}$ (Boso and Battiato 2013). We note that $\mathcal{B}$ and $\mathcal{G}$ should be arranged within the unit-cell domain such that a collection of contiguously placed domains Y is representative of how $\hat{\mathcal{B}}^\epsilon$ and $\hat{\mathcal{G}}^\epsilon$ are arranged within $\hat{\Omega}^\epsilon$.

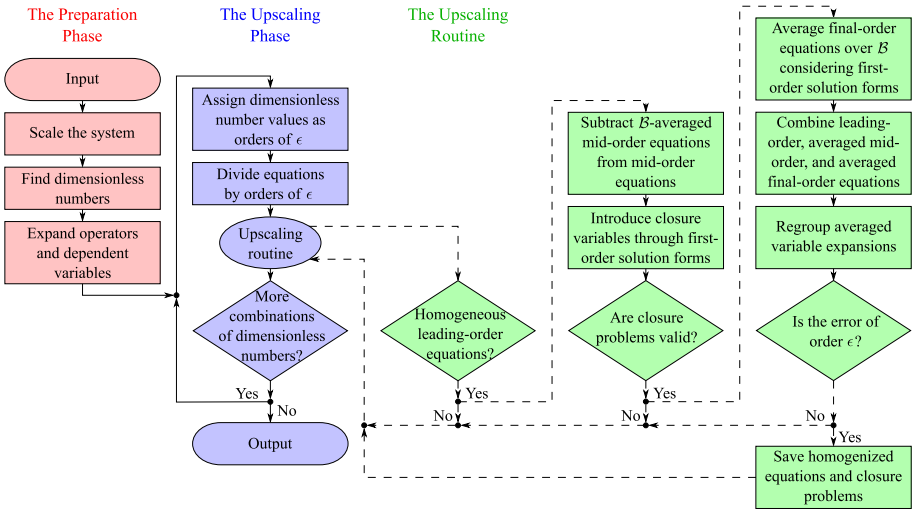

Fig. 2 A flowchart of the homogenization procedure completed by Symbolica. The program is described in two main parts: the preparation phase (red) and the upscaling phase (blue and green). In the preparation phase, the system is scaled, dimensionless numbers are found, and the differential operators and dependent variables are expanded using the "separation of scales" assumption and asymptotic expansions, respectively. In the upscaling phase, the homogenization procedure is tested for different combinations of dimensionless number values. As shown, the answer to all three questions in the upscaling routine must be "Yes" for Symbolica to save the valid homogenized model before exiting the upscaling routine

In light of the unit-cell formulation, we reconsider $\mathbf{x}$ as an element of a fictitious model domain $\Omega \equiv \Omega^\epsilon = \hat{\Omega}^\epsilon / \hat{\mathcal{L}}^b$, which is treated as an "$\hat{\ell}$-averaged" continuum under the previous assumption that $\epsilon \ll 1$ and with the intention of averaging the reactive transport over $\boldsymbol{\xi}$ in the unit-cell domain (Bachmat and Bear 1986). Therefore, we define averaging operators over the unit-cell Y, its pore-volume $\mathcal{B}$, and its liquid-solid interface Γ as

$$\langle \cdot \rangle \equiv \frac{1}{|Y|} \int_{\mathcal{B}} (\cdot)\ d\boldsymbol{\xi}, \quad \langle \cdot \rangle_{\mathcal{B}} \equiv \frac{1}{|\mathcal{B}|} \int_{\mathcal{B}} (\cdot)\ d\boldsymbol{\xi}, \quad \text{and} \quad \langle \cdot \rangle_{\Gamma} \equiv \frac{1}{|\Gamma|} \int_{\Gamma} (\cdot)\ d\boldsymbol{\xi}, \tag{6}$$

respectively. Depending on b, $|Y|$, $|\mathcal{B}|$, and $|\Gamma|$ are the volumes, areas, or segments of the unit-cell, the pore-space in the unit-cell, and the liquid-solid interface in the unit-cell, respectively. We also note that $\phi = |\mathcal{B}|/|Y|$ is the porosity of the unit-cell. By homogenizing ADR systems within this framework, we ultimately derive models that describe the system dynamics in an "$\hat{\ell}$-averaged" sense on a continuous domain at the larger scale $\hat{\mathcal{L}}$.

3 Symbolica and Automated Upscaling

Though upscaling procedures are generally comprised of routine steps, such procedures require careful mathematical manipulation and are often costly in time. In this section, we introduce Symbolica, a program for automating upscaling procedures. In general, Symbolica works in a symbolic domain and consists of functions tailored to automate routine steps found in symbolic upscaling procedures. This ultimately reduces the time and human interaction necessary to derive upscaled models. Although we use Symbolica to automate the homogenization procedure of Eqs. (3a), (3b), and (3c), other forms of reactive source terms

can be handled as well. Furthermore, generalization to other types of equations is possible and relatively straightforward to include. While Symbolica is currently written in Wolfram Mathematica (Version 12.1, 2020), it only requires a language capable of manipulating symbolic objects and handling basic symbolic computations. This allows Symbolica to be quite versatile in distribution. Finally, it is worth noting that Symbolica upscales in vector form, and therefore, can handle 1D, 2D, and 3D geometries.

As shown in the flowchart of Fig. 2, Symbolica determines the homogenized equations and closure problems in two phases: the preparation phase and the upscaling phase. In the preparation phase, Symbolica scales the system and extracts dimensionless numbers from the equations and boundary conditions. Then, considering the previous problem formulation, dependent variables are represented as power series in terms of the small length scale ratio ϵ, and differential operators are recast to consider both fast and slow variables.

After completing the preparation phase, the upscaling phase is initiated by assigning values to the previously defined dimensionless numbers. Then, a procedure similar to that carried out in the work of Boso and Battiato (2013) is pursued with modifications for finding applicability conditions and considering general ADR systems with arbitrary numbers of species and reactions. Alternative upscaling strategies, as those discussed in Auriault and Adler (1995), Allarie and Raphael (2007), Rubinstein and Mauri (1986), and Iliev et al. (2020), may also be implemented in this phase. Upon considering multiple combinations of dimensionless number values, Symbolica searches the dimensionless parameter space for upscaled equations and closure problems. As a result, homogenized models are derived in a timely manner and with minimal human interaction. In the following subsections, further details on the inputs, steps executed, and outputs of Symbolica are provided with respect to the considered problem.

3.1 Inputs

The input to Symbolica includes the symbolic definition of (1) a system of equations and boundary conditions, (2) variables, scales, and master variables, (3) dimensionless number forms, and (4) combinations of dimensionless number values for which to upscale. We discuss these inputs separately in the following.

Equations and Boundary Conditions Symbolica accepts a symbolically-formatted system of equations and boundary conditions (Step 1 in Fig. 3). Here, we assume the system to be of the form shown in equations (3a), (3b), and (3c).

Variables, Scales, and Master Variables For each variable used in the system, two additional associations are required: a scale and a master variable. As implied by the subtle shading in the left branch of Fig. 3, it is best to consider this triplet as a hierarchical structure. At the bottom of the hierarchy, *variables* are defined for symbolic representation of the system (Step 2 in Fig. 3). They are the main components manipulated by Symbolica in upscaling procedures. Just above variables in the hierarchy are *scales*, which describe the magnitude and physical dimension of an associated variable (Step 3 in Fig. 3). In general, a single scale may be associated to multiple variables, assuming the magnitudes and physical dimensions are appropriate. At the top of the hierarchy are *master variables*, which generalize scales and variables based on physical dimension, or physical interpretation in the case of dimensionless variables (Step 4 in Fig. 3). For example, as shown in Fig. 3, all concentration scales $\hat{\mathcal{C}}^{(i)}$ and $\hat{C}^{(i)}$, which scale variables $\hat{c}_\epsilon^{(i)}$ and $\hat{C}^{(i)}$ respectively, are associated with the concentration master variable $\hat{\mathbb{C}}$ due to similar physical interpretation. More formally, we define a general

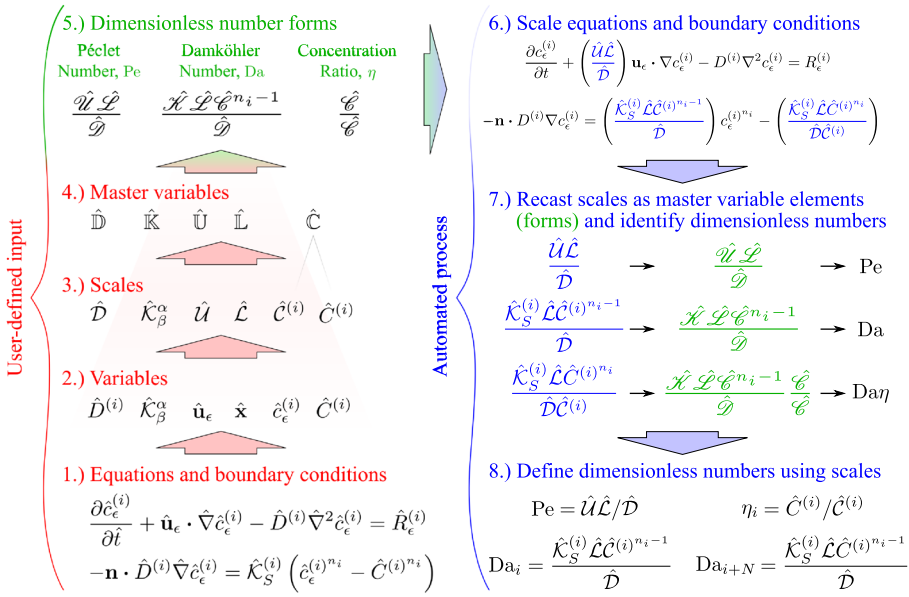

Fig. 3 A flowchart detailing the inputs to Symbolica and their use in the automated scaling and dimensionless number finding procedures. As an input (red), the user provides equations and boundary conditions (Step 1) using defined variables (Step 2), which are associated with scales (Step 3) and master variables (Step 4). The master variables, and specifically the corresponding dummy elements, are then used to define dimensionless number forms (Step 5, in green), which may be part of the input or retrieved from a previously created archive. Symbolica then preforms the following tasks automatically (blue): scales the system using the variable-associated scales (Step 6), recasts the dimensionless coefficients as groups of master variable dummy elements and identifies the contained dimensionless numbers using the dimensionless number forms (Step 7), and defines dimensionless numbers using the scales of the original dimensionless coefficients (Step 8). Here, the master variables of the diffusion scales, reaction scales, velocity scales, length scales, and concentration scales are $\hat{\mathbb{D}}$, $\hat{\mathbb{K}}$, $\hat{\mathbb{U}}$, $\hat{\mathbb{L}}$, and $\hat{\mathbb{C}}$, and their dummy elements are $\hat{\mathscr{D}}$, $\hat{\mathscr{K}}$, $\hat{\mathscr{U}}$, $\hat{\mathscr{L}}$, and $\hat{\mathscr{C}}$, respectively. Also, $i = \left\{m \in \mathbb{Z}^+ : m \leq N\right\}$, N is the number of species, $\alpha = \{(i,j),\ (i,j,k),\ (i)\}$, and $\beta = \{L,\ NL,\ S\}$

master variable $\hat{\mathbb{A}}$ as the set

$$\hat{\mathbb{A}} \equiv \left\{\hat{\mathscr{A}} : \hat{\mathscr{A}} \text{ are scales of similar physical dimension or interpretation}\right\}, \tag{7}$$

e.g., $\hat{\mathbb{C}} \equiv \{\hat{\mathscr{C}} : \hat{\mathscr{C}} = \hat{\mathcal{C}}^{(i)} \vee \hat{\mathscr{C}} = \hat{C}^{(i)}\}$ in our previous example related to concentration. In the next paragraph and in Sect. 3.2, we describe in detail the role of master variables in Symbolica's procedure to define dimensionless numbers.

Dimensionless Number Forms In addition to variables, scales, and master variables, general *forms* of dimensionless numbers should be specified (Step 5 in Fig. 3) for Symbolica to automatically identify and define the dimensionless numbers in a system. We characterize a *dimensionless number form* as a ratio between general physical phenomena. For example, the form of the Péclet number is a ratio between advective and diffusive transport time scales. It is worth emphasizing that dimensionless number forms abstract dimensionless numbers by not requiring a system, nor its scales, upon definition.

Formally, let us consider $N_{\hat{\mathbb{A}}}$ master variables $\hat{\mathbb{A}}_m$ defined by Eq. (7), where $N_{\hat{\mathbb{A}}} \geq m \in \mathbb{Z}^+$. We can then formulate all possible scale definitions of a general dimensionless number for a

given system in the set

$$\mathfrak{A} \equiv \left\{ \mathfrak{a}\left(\hat{\mathscr{A}}_1, \hat{\mathscr{A}}_2, ..., \hat{\mathscr{A}}_{N_{\hat{\mathbb{A}}}}\right) : \hat{\mathscr{A}}_m \in \hat{\mathbb{A}}_m, N_{\hat{\mathbb{A}}} \geq m \in \mathbb{Z}^+ \right\}, \tag{8}$$

where the differing structures of the dummy element $\mathfrak{a}(\hat{\mathscr{A}}_1, \hat{\mathscr{A}}_2, ..., \hat{\mathscr{A}}_{N_{\hat{\mathbb{A}}}})$, as oppose to the differing elements themselves, are considered the dimensionless number forms. As a relevant example, we can define the master variables of diffusion scales, velocity scales, and length scales as $\hat{\mathbb{D}}$, $\hat{\mathbb{U}}$, and $\hat{\mathbb{L}}$, respectively, using Eq. (7). We can then define all possible scale definitions of the Péclet number for a given system in the set $\mathfrak{Pe} \equiv \{\hat{\mathscr{U}}\hat{\mathscr{L}}/\hat{\mathscr{D}} : \hat{\mathscr{D}} \in \hat{\mathbb{D}}, \hat{\mathscr{U}} \in \hat{\mathbb{U}}, \hat{\mathscr{L}} \in \hat{\mathbb{L}}\}$ using Eq. (8). Here, the structure "$\hat{\mathscr{U}}\hat{\mathscr{L}}/\hat{\mathscr{D}}$", which is constructed from dummy elements $\hat{\mathscr{D}}$, $\hat{\mathscr{U}}$, and $\hat{\mathscr{L}}$ of the diffusion scale, velocity scale, and length scale master variables respectively, is the dimensionless number form that abstracts the Péclet number.

Because dimensionless number forms abstract dimensionless numbers, users can specify the structures of dimensionless numbers Symbolica should define without prior knowledge of the dimensionless numbers in a scaled system. For example, once Symbolica has scaled a system, as in Step 6 of Fig. 3, the emerging dimensionless coefficients are investigated. In Step 7 of Fig. 3, Symbolica recasts the dimensionless coefficients as groups of master variable dummy elements, which are then identified as dimensionless numbers using the dimensionless number forms. As a result, instead of defining a single dimensionless number from the third dimensionless coefficient, Symbolica uses the dimensionless number forms provided in Step 5 to identify the coefficient as a product of two more commonly used dimensionless numbers (i.e., Da and η). This allows Symbolica to produce results using dimensionless numbers that are familiar to the user for easy interpretation. Additionally, because master variables and dimensionless number forms can be defined in an abstract sense, it is possible to construct a general archive of dimensionless number forms to be used for arbitrary systems. The user then only needs to associate system scales and variables to master variables upon input.

Combinations of Dimensionless Number Values Once the dimensionless numbers of a system are identified, Symbolica probes different combinations of dimensionless number values for homogenized systems. While these combinations may be generated by Symbolica using a default range of values, user-defined combinations can focus Symbolica to homogenize a few desired cases.

3.2 The Preparation Phase

Scaling the System. With an appropriate input, Symbolica begins the preparation phase by scaling the system with the provided scales. To do this, Symbolica multiplies each term in the system by the scales corresponding to the variables in each term. While multiple scale combinations may be considered for an arbitrary ADR system, we consider the following for demonstrative purposes:

$$\begin{aligned} \hat{c}_\epsilon^{(i)} &= \hat{\mathcal{C}}^{(i)} c_\epsilon^{(i)}, \quad \hat{t} = \frac{\hat{\mathcal{L}}^2}{\hat{\mathcal{D}}} t, \quad \hat{\mathbf{u}}_\epsilon = \hat{\mathcal{U}} \mathbf{u}_\epsilon, \\ \hat{\nabla} &= \frac{1}{\hat{\mathcal{L}}} \nabla, \quad \hat{D}^{(i)} = \hat{\mathcal{D}} D^{(i)}, \quad \hat{R}_\epsilon^{(i)} = \frac{\hat{\mathcal{D}}\hat{\mathcal{C}}^{(i)}}{\hat{\mathcal{L}}^2} R_\epsilon^{(i)}, \end{aligned} \tag{9}$$

where $\hat{\mathcal{C}}^{(i)}$ is the concentration scale for species i, $\hat{\mathcal{D}}$ is the diffusion coefficient scale for all species, $\hat{\mathcal{U}}$ is the fluid velocity scale, and the diffusion time scale $\hat{\mathcal{L}}^2/\hat{\mathcal{D}}$ is used to scale $\hat{t}$. With these scales, Eqs. (3a), (3b), and (3c) become

$$\frac{\partial c_\epsilon^{(i)}}{\partial t} + \left(\frac{\hat{\mathcal{U}}\hat{\mathcal{L}}}{\hat{\mathcal{D}}}\right)\mathbf{u}_\epsilon \cdot \nabla c_\epsilon^{(i)} - D^{(i)}\nabla^2 c_\epsilon^{(i)} = R_\epsilon^{(i)} \quad \text{in } \mathcal{B}^\epsilon, \tag{10a}$$

$$R_\epsilon^{(i)} = \sum_{j=1}^{N} (-1)^{p_L^{(i,j)}} \left(\frac{\hat{\mathcal{K}}_L^{(i,j)}\hat{\mathcal{L}}^2\hat{\mathcal{C}}^{(j)}}{\hat{\mathcal{D}}\hat{\mathcal{C}}^{(i)}}\right) c_\epsilon^{(j)} + \sum_{j=1}^{N}\sum_{k=j}^{N} (-1)^{p_{NL}^{(i,j,k)}} \left(\frac{\hat{\mathcal{K}}_{NL}^{(i,j,k)}\hat{\mathcal{L}}^2\hat{\mathcal{C}}^{(j)}\hat{\mathcal{C}}^{(k)}}{\hat{\mathcal{D}}\hat{\mathcal{C}}^{(i)}}\right) c_\epsilon^{(j)} c_\epsilon^{(k)}, \tag{10b}$$

$$-\mathbf{n} \cdot D^{(i)}\nabla c_\epsilon^{(i)} = \left(\frac{\hat{\mathcal{K}}_S^{(i)}\hat{\mathcal{L}}\hat{\mathcal{C}}^{(i)^{n_i-1}}}{\hat{\mathcal{D}}}\right) c_\epsilon^{(i)^{n_i}} - \left(\frac{\hat{\mathcal{K}}_S^{(i)}\hat{\mathcal{L}}\hat{\mathcal{C}}^{(i)^{n_i}}}{\hat{\mathcal{D}}\hat{\mathcal{C}}^{(i)}}\right) \quad \text{on } \Gamma^\epsilon. \tag{10c}$$

Due to the complexity of the dimensionless coefficients in Eqs. (10a), (10b), and (10c), many dimensionless numbers can exist in ADR systems of multiple species. Additionally, dimensionless numbers may repeat throughout a system due to reoccurring reaction rates and the use of similar concentration scales. To maintain unique dimensionless number definitions, previously defined dimensionless numbers must be consulted prior to defining additional dimensionless numbers. Because these difficulties make manual dimensionless number definition tedious, prone to error, and time-consuming for large ADR systems, it is advantageous to use the automated procedure implemented in Symbolica for finding and defining the dimensionless numbers of a system.

Finding Dimensionless Numbers After scaling the system, Symbolica is tasked with recasting the dimensionless coefficients, shown in the parentheses of Eqs. (10a), (10b), and (10c), in terms of dimensionless numbers. As demonstrated in Step 7 of Fig. 3, Symbolica first employs the user-defined associations between scales and master variables to recast the dimensionless coefficients as groups of master variable dummy elements. Then, Symbolica compares multiple products and quotients of dimensionless number forms with each group of master variable dummy elements to recognize the dimensionless numbers appearing within each group. Finally, Symbolica compares the scales of the identified dimensionless numbers in each group with those of previously defined dimensionless numbers to decide if new dimensionless numbers should be defined or if previously defined dimensionless numbers have appeared in the groups. As shown in Step 8, the result is a set of unique dimensionless numbers defined using system scales without prior knowledge of the dimensionless numbers appearing in the system. In the case a dimensionless coefficient cannot be recast in terms of a single dimensionless number, Symbolica attempts to use the least number of dimensionless numbers, up to a maximum threshold defined by the user, to recast the dimensionless coefficient. If the threshold is reached, Symbolica stops the procedure and notifies the user.

Expanding Operators and Dependent Variables With the scaled system, Symbolica adopts the previous problem formulation, where length scales $\hat{\mathcal{L}}$ and $\hat{\ell}$ are assumed to exist. The input system is then considered with the total differential operator in space ∇, as shown in Eq. (5), and with the unit-cell problem traversed by $\boldsymbol{\xi}$. Additionally, we note that Symbolica is capable of defining further variables in time according to relevant dimensionless numbers found in the system. For example, given a defined Péclet number Pe and Damköhler numbers Da_m, Symbolica introduces the time-dependent variables $\tau_{\mathrm{Pe}}(t) \equiv \mathrm{Pe}t$ and $\tau_{\mathrm{Da}_m}(t) \equiv \mathrm{Da}_m t$, respectively, where $m \in \mathbb{Z}^+$ such that $m \leq M$ and M is the number of Damköhler numbers defined. Considering general time-dependent functions $f_\epsilon(t)$ as $f_\epsilon(t) = f(t, \tau_{\mathrm{Pe}}(t), \boldsymbol{\tau}_{\mathrm{Da}}(t))$, where $\boldsymbol{\tau}_{\mathrm{Da}}(t)$ is a tuple with components $[\boldsymbol{\tau}_{\mathrm{Da}}(t)]_m = \tau_{\mathrm{Da}_m}(t)$, Symbolica assumes that the total differential operator in time takes the form

$$\frac{\partial f_\epsilon}{\partial t} \equiv \frac{\partial f}{\partial t} + \mathrm{Pe}\frac{\partial f}{\partial \tau_{\mathrm{Pe}}} + \sum_{m=1}^{M} \mathrm{Da}_m \frac{\partial f}{\partial \tau_{\mathrm{Da}_m}}. \tag{11}$$

While Symbolica introduces the new time variables and expands the total differential operator in time for completeness, we restrict our focus to the physics occurring on the diffusive time scale and assume no dependencies on additional time variables.

In restricting the analysis to a single time variable t, dependent variables $c_\epsilon^{(i)}(t, \mathbf{x})$ and $\mathbf{u}_\epsilon(t, \mathbf{x})$ are redefined as functions of $\boldsymbol{\xi}$ and expanded as power series in terms of ϵ, such that

$$c_\epsilon^{(i)}(t, \mathbf{x}) \equiv c^{(i)}(t, \mathbf{x}, \boldsymbol{\xi}) = \sum_{j=0}^{\infty} \epsilon^j c_j^{(i)}(t, \mathbf{x}, \boldsymbol{\xi}), \tag{12}$$

$$\mathbf{u}_\epsilon(t, \mathbf{x}) \equiv \mathbf{u}(t, \mathbf{x}, \boldsymbol{\xi}) = \sum_{j=0}^{\infty} \epsilon^j \mathbf{u}_j(t, \mathbf{x}, \boldsymbol{\xi}), \tag{13}$$

where $c_j^{(i)}(t, \mathbf{x}, \boldsymbol{\xi})$ and $\mathbf{u}_j(t, \mathbf{x}, \boldsymbol{\xi})$ are assumed to be periodic in $\boldsymbol{\xi}$. With the expanded, dimensionless system, the preparation phase concludes and Symbolica proceeds to the upscaling phase, as described in Fig. 2.

3.3 The Upscaling Phase

Assignment of Dimensionless Number Values Prior to collecting terms with similar orders of ϵ in each equation and boundary condition, the dimensionless numbers are assigned values. While multiple combinations of dimensionless number values must be considered to find applicability conditions for a homogenized system, it may also be of interest to analyze select combinations of dimensionless number values. In either case, Symbolica can appropriately accommodate the analysis.

In literature, a common strategy for finding applicability conditions is to upscale a system using dimensionless numbers recast as variable powers of the small parameter, ϵ (Battiato and Tartakovsky 2011; Boso and Battiato 2013). During the upscaling procedure, applicability conditions are established by assessing the ability to separate scales and formulate valid closure problems for a range of dimensionless number values. While this strategy saves time by requiring only one iteration of the upscaling procedure, assessing the formulation of valid closure problems for multiple cases of dimensionless number values can be complex and prone to arriving at unnecessarily restrictive conditions. Instead of implementing this strategy, we take advantage of the available computing power procured through automating the upscaling procedure and let Symbolica attempt to upscale the system for each combination of dimensionless number values within a defined range. This *direct sampling* approach gives rise to the loop seen in the upscaling phase of Fig. 2. While potentially unfeasible for manual implementation, this approach can be executed in a timely manner using Symbolica, as noted in Table 3, and can be easily parallelized.

Alternative to searching for applicability conditions, users may be interested in finding upscaled systems for only a few combinations of dimensionless number values. In this case, users can provide Symbolica with quantitative information about the system scales. This allows Symbolica to compute and assign values to the dimensionless numbers after finding them in the preparation phase.

Upscaling Routine After assigning a combination of values to the dimensionless numbers and collecting terms with similar orders of ϵ in each equation and boundary condition, a host of ordered equations and boundary conditions are created from the original system. The upscaling routine, as described in Fig. 2, is then pursued by following a similar homogenization procedure to that employed in the work of Boso and Battiato (2013). As previously

discussed, the method for finding applicability conditions is altered from the cited work to utilize the available computing power procured through automating the upscaling procedure. Additionally, a straightforward extension to the cited procedure for handling more reaction terms has been implemented in Symbolica. While details behind the algebraic manipulations of Symbolica may be found in the cited work, the three main tasks completed during the upscaling routine are detailed here.

In the first task, the homogeneity of the leading-order equations and boundary conditions is evaluated. Upon finding homogeneous equations and boundary conditions, the leading-order solutions are assumed to be independent of $\boldsymbol{\xi}$ and the next steps in the routine are pursued. If the leading-order equations and boundary conditions are non-homogeneous, the upscaling routine is exited.

In the second task, the validity of the closure problems derived from the equations and boundary conditions of the following order, which we refer to as the "mid-order", is assessed. As described in Fig. 2, this task involves averaging the mid-order equations over $\mathcal{B}$ and subtracting the result from the original mid-order equations to obtain a third set of equations. In an attempt to rid any $\mathbf{x}$-dependencies from the third set of equations and the mid-order boundary conditions, the following first-order solution form is assumed:

$$c_1^{(i)}(t, \mathbf{x}, \boldsymbol{\xi}) = \boldsymbol{\chi}^{(i)}(\boldsymbol{\xi}) \cdot \nabla_{\mathbf{x}} c_0^{(i)}(t, \mathbf{x}) + \overline{c}_1^{(i)}(t, \mathbf{x}), \tag{14}$$

where $\boldsymbol{\chi}^{(i)}(\boldsymbol{\xi})$ is widely known as the *closure variable* for species i and $\overline{c}_1^{(i)}(t, \mathbf{x}) = \langle c_1^{(i)} \rangle_{\mathcal{B}}$ is the average of the first-order concentration solution for species i over $\mathcal{B}$, which implies $\langle \boldsymbol{\chi}^{(i)} \rangle_{\mathcal{B}} = \mathbf{0}$ in light of the leading-order solution being independent of $\boldsymbol{\xi}$. Upon further simplification of the difference between the mid-order equations and their averages over $\mathcal{B}$, the equations and boundary conditions to solve for $\boldsymbol{\chi}^{(i)}(\boldsymbol{\xi})$, known as the *closure problems*, are found. If these problems allow for $\boldsymbol{\chi}^{(i)}(\boldsymbol{\xi})$ to be found as only a function of $\boldsymbol{\xi}$, which has been postulated, then the closure problems are deemed valid (i.e., consistent with the postulation) and Symbolica moves onto the final task. If the closure problems are not deemed valid, the upscaling routine is exited. Here, we acknowledge that assuming the solution form in equation (14) restricts the range of dimensionless number value combinations in which Symbolica can find homogenized models. We look to relax this restriction in the future by allowing Symbolica to adapt its assumed solution form based on the equation being analyzed.

Finally, in the third task, the error of the homogenized models is verified to be no greater than order ϵ. To accomplish this, the equations on the order following the mid-order, referred to as the "final-order" equations, are averaged over $\mathcal{B}$ and combined with the leading-order and averaged mid-order equations. Considering the sum of these three sets of equations, Symbolica attempts to construct homogenized equations by regrouping the averages of the variable expansions shown in Eq. (12) and (13). If leftover expansion terms in the homogenized equations are of order ϵ or smaller after regrouping, then the error is deemed at most first-order and Symbolica proceeds to save the homogenized equations before exiting the upscaling routine. Otherwise, the upscaling routine is exited upon failure to obtain the correct order of error.

Upon exiting the upscaling routine, Symbolica proceeds to analyze the next combination of dimensionless number values before ending the upscaling phase and providing the saved homogenized equations, closure problems, and defined dimensionless numbers to the user in symbolic form. In the following section we validate Symbolica's outputs against both published literature and numerical simulations for three ADR problems of increasing complexity.

Table 1 Definitions of the variables from the problem formulation considering a 2D channel geometry

Variable	Definition for a channel geometry
Scales	
$\hat{\ell}$	Channel width
$\hat{\mathcal{L}}$	Length scale along the channel
Full medium domain	
Ω^ϵ	$\left\{(x, y) : 0 < x < 1,\ -\frac{1}{2} < y < \frac{1}{2}\right\}$
$\partial\Omega_0^\epsilon$	$\left\{(x, y) : x = 0,\ -\frac{1}{2} < y < \frac{1}{2}\right\}$
$\partial\Omega_1^\epsilon$	$\left\{(x, y) : x = 1,\ -\frac{1}{2} < y < \frac{1}{2}\right\}$
$\mathcal{B}^\epsilon$	$\left\{(x, y) : 0 < x < 1,\ -\frac{1}{2} < y < \frac{1}{2}\right\}$
$\partial\mathcal{B}_0^\epsilon$	$\left\{(x, y) : x = 0,\ -\frac{1}{2} < y < \frac{1}{2}\right\}$
$\partial\mathcal{B}_1^\epsilon$	$\left\{(x, y) : x = 1,\ -\frac{1}{2} < y < \frac{1}{2}\right\}$
Γ^ϵ	$\left\{(x, y) : 0 < x < 1,\ y = \pm\frac{1}{2}\right\}$
Unit-cell domain	
Y	$\left\{y : -\frac{1}{2} < y < \frac{1}{2}\right\}$
$\mathcal{B}$	$\left\{y : -\frac{1}{2} < y < \frac{1}{2}\right\}$
Γ	$\left\{y : y = \pm\frac{1}{2}\right\}$
$\lvert Y\rvert$	1
$\lvert\mathcal{B}\rvert$	1
ϕ	1
"$\hat{\ell}-$ *averaged" continuum domain*	
Ω	$\{x : 0 < x < 1\}$
$\partial\Omega_0$	$\{x : x = 0\}$
$\partial\Omega_1$	$\{x : x = 1\}$
Spatial variables and operators	
$\mathbf{x}$	$x\mathbf{e}_x$
$\boldsymbol{\xi}$	$y\mathbf{e}_y$
$\nabla_{\mathbf{x}}$	$\frac{\partial}{\partial x}\mathbf{e}_x$
$\nabla_{\boldsymbol{\xi}}$	$\frac{\partial}{\partial y}\mathbf{e}_y$

Here, $\mathbf{e}_x$ and $\mathbf{e}_y$ are unit vectors in the x- and y-directions, respectively

4 Validation and Numerical Experiments

We now demonstrate the capabilities and benefits of using Symbolica to derive homogenized models in three example problems. In the first two examples, we directly compare the outputs of Symbolica with the classically homogenized reactive systems analyzed in Battiato and Tartakovsky (2011) and Boso and Battiato (2013), which involve non-linear heterogeneous and mixing-induced homogeneous reactions, respectively. In the first problem, we demonstrate the direct sampling approach implemented in Symbolica for finding applicability conditions and highlight the advantages over manual investigation. We extend this analysis into the

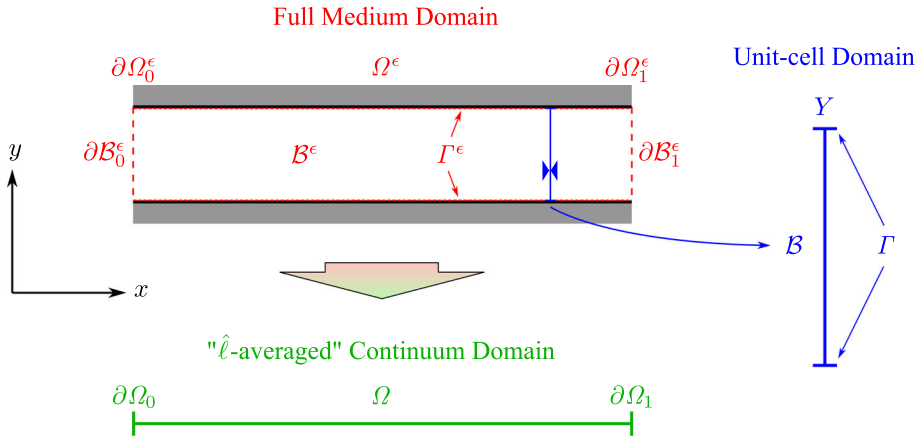

Fig. 4 A schematic of the three domains defined in Table 1 for the problem formulation considering a 2D channel geometry

second problem, where applicability conditions are again found for a more complex reactive system. In the third problem, we demonstrate Symbolica's handling of large, coupled systems of equations by homogenizing a ten-species system with non-linear homogeneous reactions in the liquid phase. For each example, we validate the derived models and procedures implemented in Symbolica numerically for transport in a channel flow. Specifically, we provide numerical evidence that each model achieves the predicted error by comparing the spatially-averaged, pore-scale solution (2D) with the homogenized solution (1D) derived by Symbolica. A description of the numerical domains is provided in the following section.

4.1 Domain and Mesh Setup

While Symbolica completes the upscaling procedure independently of geometry, numerical validation is conducted for the three example problems using a 2D channel geometry in a Cartesian plane, where the x-direction is parallel to the length of the channel, for simplicity. It is worth emphasizing that applying these models to such a geometry is identical to averaging across the cross-section of the channel. This eliminates the dependency in the vertical direction and results in a 1D upscaled model. We let $\hat{\ell}$ be the channel width, which scales the y-coordinate, and $\hat{\mathcal{L}}$ be a length scale of interest along the length of channel, which scales the x-coordinate. From these length scales, the definitions of the domains and operators described in the problem formulation (Sect. 2) then follow, as shown in Table 1. Additional variables related to domain boundaries have also been defined in Table 1. Sketches of the defined domains with the corresponding notation are provided in Fig. 4.

Inside the channel, we assume a fully-developed flow driven by a constant pressure gradient and solve equations (2a), (2b), and (2c) to gain

$$\hat{\mathbf{v}}\,(y) = \hat{\mathcal{U}} v\,(y)\,\mathbf{e}_x \quad \text{in } \mathcal{B}^\epsilon, \tag{15}$$

where $\hat{\mathcal{U}} = -\hat{\ell}^2(\partial\hat{p}/\partial\hat{x})/(8\hat{\mu})$ is the velocity scale, $\mathbf{e}_x$ is the unit vector in the x-direction, and

$$v\,(y) = 1 - 4y^2 \quad \text{in } \mathcal{B}^\epsilon. \tag{16}$$

Table 2 Simulation parameters and boundary conditions used to numerically solve the homogenized and pore-scale models for the first example problem

Homogenized model		Pore-scale model	
Simulation parameters			
Maximum $\Delta x = 1 \times 10^{-5}$		Maximum $\Delta x \Delta y = 5 \times 10^{-5}$	
Simulation boundary conditions			
$\left\langle c^{(1)} \right\rangle_{\mathcal{B}} = 0.7$	for $x \in \partial\Omega_0,\ t > 0$	$c_\epsilon^{(1)} = 0.7$	for $(x, y) \in \partial\mathcal{B}_0^\epsilon,\ t > 0$
$\partial \left\langle c^{(1)} \right\rangle_{\mathcal{B}} / \partial x = 0$	for $x \in \partial\Omega_1,\ t > 0$	$\partial c_\epsilon^{(1)} / \partial x = 0$	for $(x, y) \in \partial\mathcal{B}_1^\epsilon,\ t > 0$

All numerical calculations are completed using a packaged finite element method solver in the software Wolfram Mathematica (Version 12.1, 2020). The meshes of both 1D and 2D domains are composed of uniform, second-order elements, with quadrilateral elements used in the 2D domains. Details regarding the maximum element and time step sizes used in each example problem are reported in the corresponding sections that follow.

4.2 Example Problem 1: One Species, Non-linear Heterogeneous Reaction

4.2.1 Reaction Description and Preparation

In the first example problem, we use Symbolica to homogenize the system presented in the work of Battiato and Tartakovsky (2011). In this system, a single species in the liquid phase of a pore-space undergoes a non-linear, heterogeneous reaction at the liquid-solid interface of an arbitrary geometry. Using the dimensional equation and boundary condition from (Battiato and Tartakovsky 2011), we write the input system describing the transport as

$$\frac{\partial \hat{c}_\epsilon^{(1)}}{\partial \hat{t}} + \hat{\mathbf{u}}_\epsilon \cdot \hat{\nabla} \hat{c}_\epsilon^{(1)} - \hat{D}^{(1)} \hat{\nabla}^2 \hat{c}_\epsilon^{(1)} = 0 \quad \text{in } \hat{\mathcal{B}}^\epsilon, \tag{17a}$$

$$-\mathbf{n} \cdot \hat{D}^{(1)} \hat{\nabla} \hat{c}_\epsilon^{(1)} = \hat{\mathcal{K}}_S^{(1)} \left(\hat{c}_\epsilon^{(1)^{n_1}} - \hat{C}^{(1)^{n_1}} \right) \quad \text{on } \hat{\Gamma}^\epsilon, \tag{17b}$$

where we assume $\hat{\mathbf{u}}_\epsilon$ is a known function that allows for scales to be separated and $n_1 = 2$. With the scales defined in equation (9), and $\hat{C}^{(1)} = \hat{C}^{(1)}$, Symbolica finds the dimensionless system

$$\frac{\partial c_\epsilon^{(1)}}{\partial t} + \mathrm{Pe}_1 \mathbf{u}_\epsilon \cdot \nabla c_\epsilon^{(1)} - D^{(1)} \nabla^2 c_\epsilon^{(1)} = 0 \quad \text{in } \mathcal{B}^\epsilon, \tag{18a}$$

$$-\mathbf{n} \cdot D^{(1)} \nabla c_\epsilon^{(1)} = \mathrm{Da}_1 \left(c_\epsilon^{(1)^2} - 1 \right) \quad \text{on } \Gamma^\epsilon, \tag{18b}$$

prior to expanding the operators and dependent variables. Here, Symbolica has defined the Péclet and Damköhler numbers, Pe_1 and Da_1, as

$$\mathrm{Pe}_1 = \frac{\hat{\mathcal{U}} \hat{\mathcal{L}}}{\hat{\mathcal{D}}}, \quad \mathrm{Da}_1 = \frac{\hat{\mathcal{K}}_S^{(1)} \hat{\mathcal{L}} \hat{C}^{(1)}}{\hat{\mathcal{D}}}. \tag{19}$$

With the dimensionless equations, Symbolica proceeds to homogenize the system and find applicability conditions by probing the dimensionless parameter space as per user specifications. We emphasize that the conditions derived here correspond to the following hypothesis:

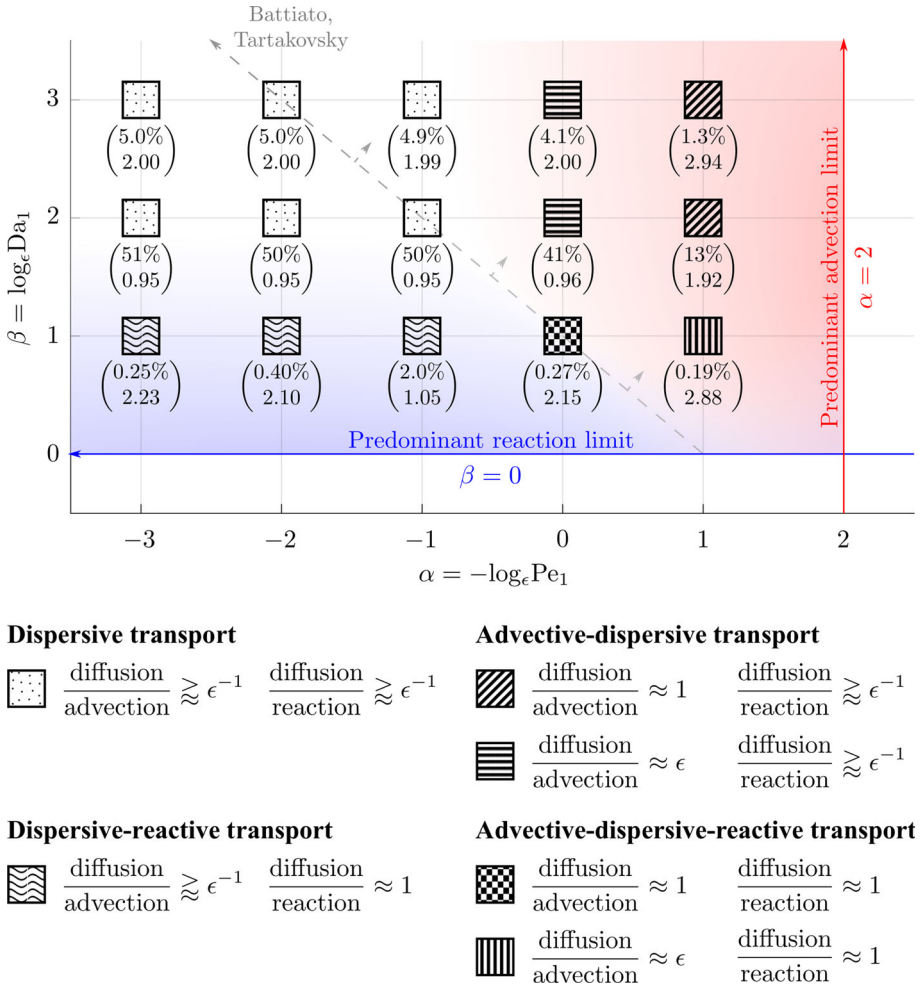

Fig. 5 The $\alpha - \beta$ phase diagram for the first example problem providing the applicability region of the homogenized models bounded by $\alpha = 2$ and $\beta = 0$, where α and β are defined in Eq. (20). The additional applicability condition and corresponding applicability region determined in the work of Battiato and Tartakovsky (2011) is indicated by the gray-dashed line. At each point probed in the (α, β)-space, the homogenized and pore-scale models are solved for multiple values of ϵ to calculate $\mathcal{E}_r^{(1)}$, defined in Eq. (23), and the order of convergence of $E_{\max}^{(1)}$, where $E_{\max}^{(1)}$ is defined in Eq. (21). The symbols indicate different combinations of magnitudes between the advective, diffusive, and reactive terms of the homogenized equations, as shown in the legend

(1) the first-order solutions have the form defined in Eq. (14), (2) the fluid velocity is only a function of the fast variables, and (3) only three orders of equations (the leading-order, mid-order, and final-order) are used to derive the homogenized model, as discussed in Sect. 3.3. During the direct sampling approach, Symbolica attempts to homogenize the system for all combinations of dimensionless numbers Pe_1 and Da_1 with assigned values ϵ^ω where $\omega \in \{-2, -1, 0, 1, 2, 3\}$ as set by the user. This range of values corresponds to sampling 36 discrete points in the $(\mathrm{Pe}_1, \mathrm{Da}_1)$ space.

4.2.2 Symbolic Upscaling Results and Numerical Validation

In its entirety, the automated upscaling procedure executed in roughly 13.43 seconds, where loading Symbolica and the inputs took 2.95 seconds, completing the preparation phase took 0.56 seconds, completing the upscaling phase took 8.83 seconds, and creating the output file took 1.09 seconds. In contrast to the work of Battiato and Tartakovsky (2011), where the applicability conditions

1. $\mathrm{Pe}_1 < \epsilon^{-2}$,
2. $\mathrm{Da}_1 < 1$,
3. $\mathrm{Da}_1/\mathrm{Pe}_1 < \epsilon$,

were found, our results show that condition 3 is unnecessary to achieve homogenized models with the correct error of order ϵ, as Symbolica successfully derives homogenized models for $\mathrm{Da}_1/\mathrm{Pe}_1 > \epsilon$. This suggests that the last condition, found through manual derivation, is overly restrictive.

For consistency with Battiato and Tartakovsky (2011), we define

$$\alpha = -\log_\epsilon \mathrm{Pe}_1 \quad \text{and} \quad \beta = \log_\epsilon \mathrm{Da}_1, \tag{20}$$

and provide numerical validation of the extended applicability region by comparing averaged solutions from the pore-scale model, described in Eqs. (18a) and (18b), and the homogenized models. We choose to solve for the steady-state concentration profile in a channel with constant input to mitigate computational complexity and focus on validating the extended applicability region. While the domains for the 2D pore-scale model and 1D homogenized models are shown in Fig. 4 and detailed in Table 1, the simulation parameters and boundary conditions applied to the models are found in Table 2. Additionally, the homogenized equations and closure problems corresponding to the 15 homogenized combinations of α and β can be found in Appendix A.

To compare the homogenized and pore-scale models, we define the maximum absolute error between the spatially averaged pore-scale solution $\langle c_\epsilon^{(i)} \rangle_{\mathcal{B}}$ and the homogenized solution $\langle c^{(i)} \rangle_{\mathcal{B}}$ as

$$E_{\max}^{(i)} = \max_{x \in \Omega} (E^{(i)}), \tag{21}$$

where

$$E^{(i)} = |\langle c_\epsilon^{(i)} \rangle_{\mathcal{B}} - \langle c^{(i)} \rangle_{\mathcal{B}}|, \tag{22}$$

is the absolute error. Here, the numerical solution of the pore-scale model $c_\epsilon^{(i)}$ is averaged over the channel width to obtain $\langle c_\epsilon^{(i)} \rangle_{\mathcal{B}}$ and we let $i = 1$ for the current example problem. In computing the error, we consider multiple values of $\epsilon \in [0.01, 0.1]$ at each sampling point in the (α, β)-space by changing the values of Pe_1 and Da_1 according to Eq. (20). This allows us to record the greatest relative maximum error $\mathcal{E}_r^{(1)}$, defined as

$$\mathcal{E}_r^{(1)} = 100\% \times \max_{\epsilon \in [0.01, 0.1]} \left(\frac{E_{\max}^{(1)}(\epsilon)}{\epsilon} \right), \tag{23}$$

at each sampling location.

To visualize and validate the applicability region found by Symbolica, where $\alpha < 2$ ($\mathrm{Pe}_1 < \epsilon^{-2}$) and $\beta > 0$ ($\mathrm{Da}_1 < 1$), we provide the $\alpha - \beta$ phase diagram in Fig. 5. As shown, the 15 sampling locations where Symbolica successfully homogenized the system are highlighted

Table 3 A direct comparison of the execution times and related parameters from each example problem

Example problem	Number of equations	Sampled combinations	Load Symbolica and inputs [sec]	Preparation phase [sec]	Upscaling phase [sec]	Write output [sec]	Total time [sec]
1	1	36	2.95	0.56	8.83	1.09	13.43
2	3	625	1.69	2.08	432.41	59.05	495.23
3	10	2	2.92	4.12	56.05	1.73	64.82

Here, we let $i \in \mathbb{M} = \{m \in \mathbb{Z}^+ : m \leq 10\}$ and $H(x)$ is the Heaviside function

Table 4 Exponents of ϵ that define the combinations of dimensionless number values in the three considered scenarios

	Scenario 1	Scenario 2	Scenario 3
α	-2	$1/2$	$1/2$
β	$1/4$	$1/4$	1
γ	$-7/4$	-1	$-7/4$
δ	$-7/4$	-1	$-7/4$
$\alpha+\beta$	$\mathbf{-7/4}$	$\mathbf{3/4}$	$3/2$
$\alpha+\gamma$	$\mathbf{-15/4}$	$-1/2$	$\mathbf{-5/4}$
$\alpha+\delta$	$\mathbf{-15/4}$	$-1/2$	$\mathbf{-5/4}$

Here, we let $\mathrm{Pe}_1 = \epsilon^{-\alpha}$, $\mathrm{Da}_1 = \epsilon^{\gamma}$, $\mathrm{Da}_2 = \epsilon^{\delta}$, and $\mathrm{Da}_3 = \epsilon^{\beta}$. The summed exponents that violate an additional applicability condition from the work of Boso and Battiato (2013) are indicated in boldface

with the corresponding results from the error analysis. The additional applicability condition found in Battiato and Tartakovsky (2011), $\mathrm{Da}_1/\mathrm{Pe}_1 < \epsilon$ ($\alpha+\beta > 1$), is visualized using the dashed line $\alpha+\beta = 1$ for comparative purposes. Each square symbol in the diagram indicates a different combination of magnitudes between the advection, diffusion, and reaction terms of the homogenized equations. Consequently, the symbols also represent different pairs of homogenized equations and closure problems found by Symbolica. For example, the dotted square symbol indicating diffusive transport represents Eqs. (39a)–(39d) in Appendix A. The two numbers below each symbol provide information regarding the error between the solutions of the homogenized model and the pore-scale model. The first number is $\mathcal{E}_r^{(1)}$ defined in Eq. (23). Because all percentages stay below 100%, the magnitudes of the absolute errors remain less than ϵ at all sampling locations for all values of ϵ considered, despite some (α, β) pairs violating the additional applicability condition, $\alpha+\beta > 1$ ($\mathrm{Da}_1/\mathrm{Pe}_1 < \epsilon$). The second number represents the order of convergence of $E_{\max}^{(1)}$ with respect to ϵ, which is calculated considering multiple values of ϵ for each sampling point. As shown, all convergence orders are approximately 1 or greater, regardless of the condition $\alpha+\beta > 1$ ($\mathrm{Da}_1/\mathrm{Pe}_1 < \epsilon$). Because consistency is found between the theoretically predicted and numerically calculated errors, the less restrictive applicability conditions found by Symbolica are deemed viable for rigorous upscaling.

The difference in reported applicability regions reveals a key advantage of finding applicability conditions through the direct sampling approach implemented in Symbolica. As previously stated, applicability conditions are often obtained by parameterizing the dimensionless numbers using variable powers of ϵ in manual investigation. Though time-efficient for manual implementation, this strategy is prone to finding unnecessarily restrictive conditions. Additionally, its time-efficiency can be outweighed by the time necessary to formulate complex applicability diagrams for large ADR systems. However, by automating the upscaling procedure and allocating the work to computational resources, Symbolica avoids these obstacles and brings time feasibility to the direct sampling approach. We believe this makes the automated direct sampling approach a better alternative for finding applicability conditions.

Table 5 Simulation parameters, boundary conditions, and initial conditions used to numerically solve the homogenized and pore-scale models for the second example problem

Homogenized model		Pore-scale model	
Simulation parameters			
Maximum $\Delta x = 1 \times 10^{-5}$		Maximum $\Delta x \Delta y = 5 \times 10^{-6}$	
Maximum $\Delta t = 1 \times 10^{-3}$		Maximum $\Delta t = 1 \times 10^{-3}$	
Simulation boundary conditions			
$\partial \left\langle c^{(i)} \right\rangle_{\mathcal{B}} / \partial x = 0$	for $x \in \partial\Omega_0 \cup \partial\Omega_1$, $t > 0$	$\partial c_\epsilon^{(i)} / \partial x = 0$	for $(x, y) \in \partial\mathcal{B}_0^\epsilon \cup \partial\mathcal{B}_1^\epsilon$, $t > 0$
Simulation initial conditions			
$\left\langle c^{(1)} \right\rangle_{\mathcal{B}} = H(0.5 - x)$	for $x \in \Omega$, $t = 0$	$c_\epsilon^{(1)} = H(0.5 - x)$	for $(x, y) \in \mathcal{B}^\epsilon$, $t = 0$
$\left\langle c^{(2)} \right\rangle_{\mathcal{B}} = H(x - 0.5)$	for $x \in \Omega$, $t = 0$	$c_\epsilon^{(2)} = H(x - 0.5)$	for $(x, y) \in \mathcal{B}^\epsilon$, $t = 0$
$\left\langle c^{(3)} \right\rangle_{\mathcal{B}} = 0$	for $x \in \Omega$, $t = 0$	$c_\epsilon^{(3)} = 0$	for $(x, y) \in \mathcal{B}^\epsilon$, $t = 0$

Here, we let $i \in \{1, 2, 3\}$ and $H(x)$ is the Heaviside function

4.3 Example Problem 2: Three Species, Non-linear Homogeneous Reaction

4.3.1 Reaction Description and Preparation

We now use Symbolica to homogenize the three-species ADR system investigated in the work of Boso and Battiato (2013), which includes a non-linear, reversible homogeneous biomolecular reaction and a linear heterogeneous reaction at the liquid-solid interface, i.e., $A + B \rightleftharpoons C \rightleftharpoons S$. Using the dimensional equations and boundary conditions from Boso and Battiato (2013), we write the input system describing the reactive transport as

$$\frac{\partial \hat{c}_\epsilon^{(1)}}{\partial \hat{t}} + \hat{\mathbf{u}}_\epsilon \cdot \hat{\nabla} \hat{c}_\epsilon^{(1)} - \hat{D}^{(1)} \hat{\nabla}^2 \hat{c}_\epsilon^{(1)} = -\hat{\mathcal{K}}_{NL}^{(1,2)} \hat{c}_\epsilon^{(1)} \hat{c}_\epsilon^{(2)} + \hat{\mathcal{K}}_L^{(3)} \hat{c}_\epsilon^{(3)} \quad \text{in } \hat{\mathcal{B}}^\epsilon, \tag{24a}$$

$$\frac{\partial \hat{c}_\epsilon^{(2)}}{\partial \hat{t}} + \hat{\mathbf{u}}_\epsilon \cdot \hat{\nabla} \hat{c}_\epsilon^{(2)} - \hat{D}^{(2)} \hat{\nabla}^2 \hat{c}_\epsilon^{(2)} = -\hat{\mathcal{K}}_{NL}^{(1,2)} \hat{c}_\epsilon^{(1)} \hat{c}_\epsilon^{(2)} + \hat{\mathcal{K}}_L^{(3)} \hat{c}_\epsilon^{(3)} \quad \text{in } \hat{\mathcal{B}}^\epsilon, \tag{24b}$$

$$\frac{\partial \hat{c}_\epsilon^{(3)}}{\partial \hat{t}} + \hat{\mathbf{u}}_\epsilon \cdot \hat{\nabla} \hat{c}_\epsilon^{(3)} - \hat{D}^{(3)} \hat{\nabla}^2 \hat{c}_\epsilon^{(3)} = \hat{\mathcal{K}}_{NL}^{(1,2)} \hat{c}_\epsilon^{(1)} \hat{c}_\epsilon^{(2)} - \hat{\mathcal{K}}_L^{(3)} \hat{c}_\epsilon^{(3)} \quad \text{in } \hat{\mathcal{B}}^\epsilon, \tag{24c}$$

$$-\mathbf{n} \cdot \hat{D}^{(1)} \hat{\nabla} \hat{c}_\epsilon^{(1)} = 0 \quad \text{on } \hat{\Gamma}^\epsilon, \tag{24d}$$

$$-\mathbf{n} \cdot \hat{D}^{(2)} \hat{\nabla} \hat{c}_\epsilon^{(2)} = 0 \quad \text{on } \hat{\Gamma}^\epsilon, \tag{24e}$$

$$-\mathbf{n} \cdot \hat{D}^{(3)} \hat{\nabla} \hat{c}_\epsilon^{(3)} = \hat{\mathcal{K}}_S^{(3)} \left(\hat{c}_\epsilon^{(3)} - \hat{C}^{(3)} \right) \quad \text{on } \hat{\Gamma}^\epsilon, \tag{24f}$$

where variables with superscripts "(1)", "(2)", and "(3)" correspond to species A, B, and C, respectively, and we assume $\hat{\mathbf{u}}_\epsilon$ is a known function that allows for scales to be separated. With respect to the input forms provided in Eqs. (3a) and (3b), we let $\hat{\mathcal{K}}_{NL}^{(1,1,2)} = \hat{\mathcal{K}}_{NL}^{(2,1,2)} = \hat{\mathcal{K}}_{NL}^{(3,1,2)} = \hat{\mathcal{K}}_{NL}^{(1,2)}$, $\hat{\mathcal{K}}_L^{(1,3)} = \hat{\mathcal{K}}_L^{(2,3)} = \hat{\mathcal{K}}_L^{(3,3)} = \hat{\mathcal{K}}_L^{(3)}$, $n_3 = 1$, and $p_L^{(i,j)}$ and $p_{NL}^{(i,j,k)}$ are chosen accordingly.

To scale the system, we allow Symbolica to assume the scales in Eq. (9) and reduce the number of scales by letting $\hat{\mathcal{C}}^{(i)} = \hat{\mathcal{C}}^*$, where $i \in \{1, 2, 3\}$. In doing so, Symbolica finds the dimensionless system

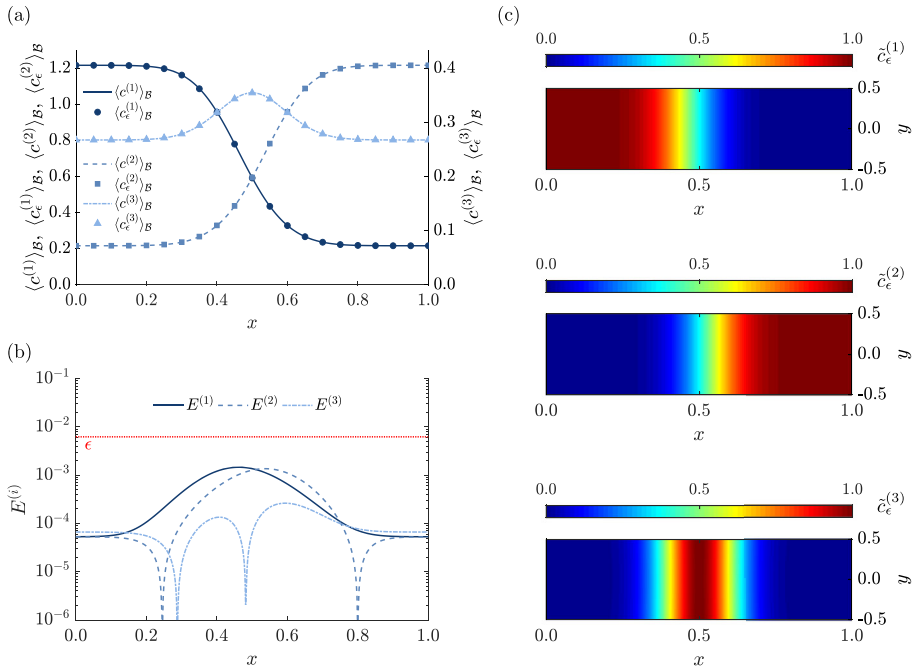

Fig. 6 The numerical results from scenario 1 of the second example problem at $t = \epsilon$, where $\epsilon = 0.00625$ and $(\alpha, \beta, \gamma, \delta) = (-2, 1/4, -7/4, -7/4)$. **(a)** The $\mathcal{B}$-averaged concentration profiles from the homogenized model and the spatially averaged pore-scale solutions from Eqs. (25a)–(25f). **(b)** The absolute error $E^{(i)}$ in the x-direction, defined in Eq. (22) where $i = \{1, 2, 3\}$. The red dotted line provides the upper limit on the error predicted by the homogenized model. **(c)** The normalized pore-scale concentration profiles $\tilde{c}_\epsilon^{(i)}$, defined in Eq. (27) for $i = \{1, 2, 3\}$. Here, $\min_{(x,y)\in\Omega^\epsilon}(c_\epsilon^{(i)})$ was found as 0.2168, 0.2168, and 0.2675, and $\max_{(x,y)\in\Omega^\epsilon}(c_\epsilon^{(i)})$ was found as 1.2168, 1.2168, and 0.3549, for $i = 1$, 2, and 3, respectively

$$\frac{\partial c_\epsilon^{(1)}}{\partial t} + \mathrm{Pe}_1 \mathbf{u}_\epsilon \cdot \nabla c_\epsilon^{(1)} - D^{(1)} \nabla^2 c_\epsilon^{(1)} = -\mathrm{Da}_1 c_\epsilon^{(1)} c_\epsilon^{(2)} + \mathrm{Da}_2 c_\epsilon^{(3)} \quad \text{in } \mathcal{B}^\epsilon, \tag{25a}$$

$$\frac{\partial c_\epsilon^{(2)}}{\partial t} + \mathrm{Pe}_1 \mathbf{u}_\epsilon \cdot \nabla c_\epsilon^{(2)} - D^{(2)} \nabla^2 c_\epsilon^{(2)} = -\mathrm{Da}_1 c_\epsilon^{(1)} c_\epsilon^{(2)} + \mathrm{Da}_2 c_\epsilon^{(3)} \quad \text{in } \mathcal{B}^\epsilon, \tag{25b}$$

$$\frac{\partial c_\epsilon^{(3)}}{\partial t} + \mathrm{Pe}_1 \mathbf{u}_\epsilon \cdot \nabla c_\epsilon^{(3)} - D^{(3)} \nabla^2 c_\epsilon^{(3)} = \mathrm{Da}_1 c_\epsilon^{(1)} c_\epsilon^{(2)} - \mathrm{Da}_2 c_\epsilon^{(3)} \quad \text{in } \mathcal{B}^\epsilon, \tag{25c}$$

$$-\mathbf{n} \cdot D^{(1)} \nabla c_\epsilon^{(1)} = 0 \quad \text{on } \Gamma^\epsilon, \tag{25d}$$

$$-\mathbf{n} \cdot D^{(2)} \nabla c_\epsilon^{(2)} = 0 \quad \text{on } \Gamma^\epsilon, \tag{25e}$$

$$-\mathbf{n} \cdot D^{(3)} \nabla c_\epsilon^{(3)} = \mathrm{Da}_3 \left(c_\epsilon^{(3)} - \eta_1 \right) \quad \text{on } \Gamma^\epsilon, \tag{25f}$$

prior to expanding the operators and dependent variables. Here, Symbolica has defined the Péclet number Pe_1, the Damköhler numbers Da_1, Da_2, and Da_3, and the concentration ratio η_1 as

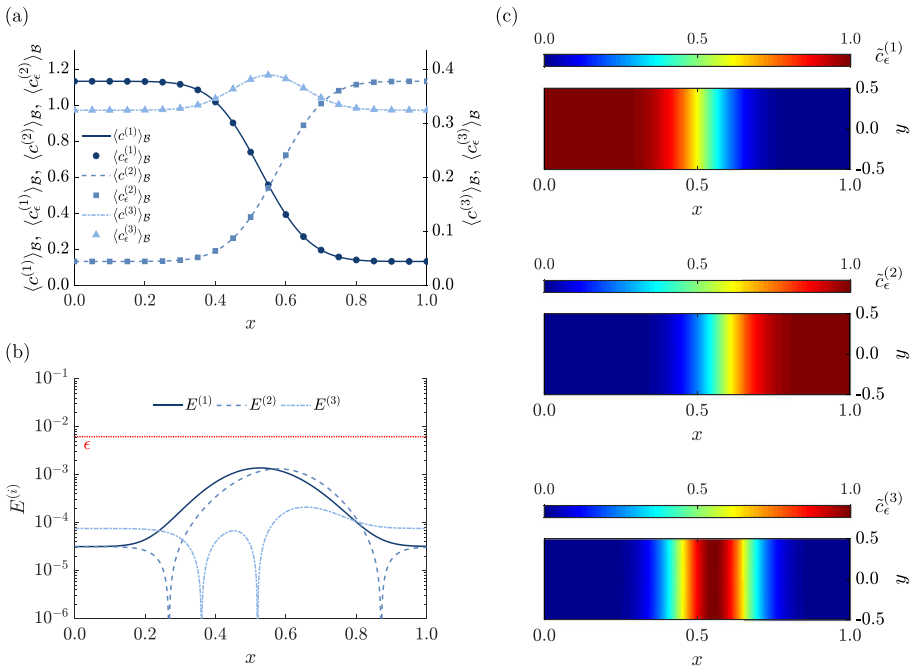

Fig. 7 The numerical results from scenario 2 of the second example problem at $t = \epsilon$, where $\epsilon = 0.00625$ and $(\alpha, \beta, \gamma, \delta) = (1/2, 1/4, -1, -1)$. **(a)** The $\mathcal{B}$-averaged concentration profiles from the homogenized model and the spatially averaged pore-scale solutions from Eqs. (25a)–(25f). **(b)** The absolute error $E^{(i)}$ in the x-direction, defined in Eq. (22) where $i = \{1, 2, 3\}$. The red dotted line provides the upper limit on the error predicted by the homogenized model. **(c)** The normalized pore-scale concentration profiles $\tilde{c}_\epsilon^{(i)}$, defined in Eq. (27) for $i = \{1, 2, 3\}$. Here, $\min_{(x,y)\in\Omega^\epsilon}(c_\epsilon^{(i)})$ was found as 0.1320, 0.1319, and 0.3237, and $\max_{(x,y)\in\Omega^\epsilon}(c_\epsilon^{(i)})$ was found as 1.1319, 1.1318, and 0.3888, for $i = 1$, 2, and 3, respectively

$$\mathrm{Pe}_1 = \frac{\hat{\mathcal{U}}\hat{\mathcal{L}}}{\hat{\mathcal{D}}}, \quad \mathrm{Da}_1 = \frac{\hat{\mathcal{K}}_{NL}^{(1,2)}\hat{\mathcal{L}}^2\hat{\mathcal{C}}^*}{\hat{\mathcal{D}}}, \quad \mathrm{Da}_2 = \frac{\hat{\mathcal{K}}_{L}^{(3)}\hat{\mathcal{L}}^2}{\hat{\mathcal{D}}}, \quad \mathrm{Da}_3 = \frac{\hat{\mathcal{K}}_{S}^{(3)}\hat{\mathcal{L}}}{\hat{\mathcal{D}}}, \quad \eta_1 = \frac{\hat{\mathcal{C}}^{(3)}}{\hat{\mathcal{C}}^*}. \tag{26}$$

Similar to the previous example problem, we now let Symbolica homogenize the system in Eqs. (25a)–(25f) and find applicability conditions through direct sampling of the 5D dimensionless parameter space. To conduct the direct sampling approach, Symbolica attempts to homogenize the system for all combinations of dimensionless numbers Pe_1, Da_1, Da_2, and Da_3 with assigned values ϵ^ω, where $\omega \in \{-2, -1, 0, 1, 2\}$, while assuming $\eta_1 = \epsilon^0$. This corresponds to sampling 625 points in the 5D parameter space.

4.3.2 Symbolic Upscaling Results and Numerical Validation

The execution time of the entire automated upscaling procedure was roughly 8 minutes and 15.23 seconds, where a detailed breakdown of the time is provided in Table 3. In general, the increased execution time with respect to the previous example is a result of the increase in system size and combinations of dimensionless number values considered for finding applicability conditions. It is worth noticing that most of the execution time is dedicated to the upscaling phase, which can be easily parallelized due to the independence of the upscaling

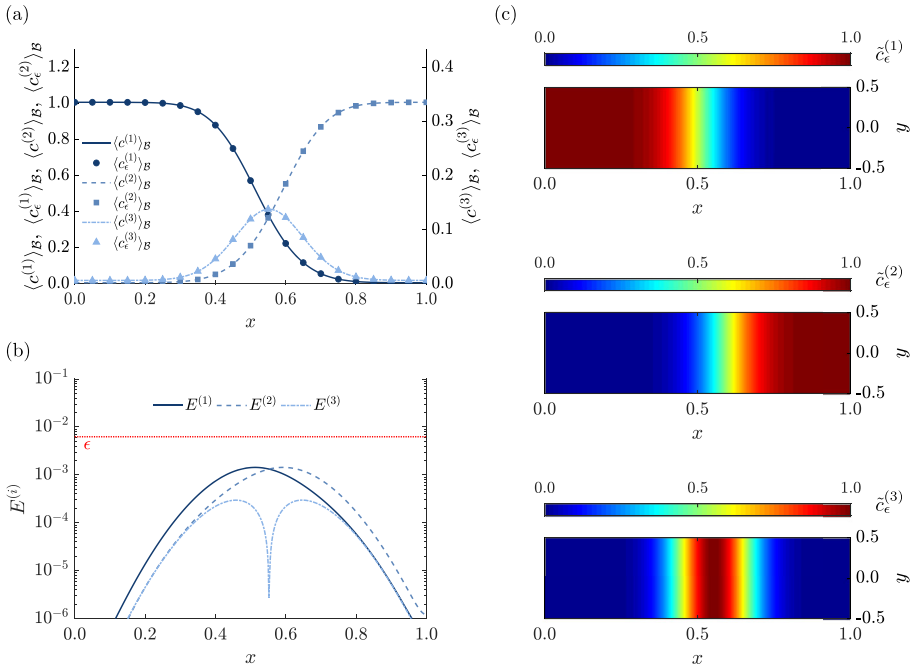

Fig. 8 The numerical results from scenario 3 of the second example problem at $t = \epsilon$, where $\epsilon = 0.00625$ and $(\alpha, \beta, \gamma, \delta) = (1/2, 1, -7/4, -7/4)$. **(a)** The $\mathcal{B}$-averaged concentration profiles from the homogenized model and the spatially averaged pore-scale solutions from Eqs. (25a)–(25f). **(b)** The absolute error $E^{(i)}$ in the x-direction, defined in Eq. (22) where $i = \{1, 2, 3\}$. The red dotted line provides the upper limit on the error predicted by the homogenized model. **(c)** The normalized pore-scale concentration profiles $\tilde{c}_\epsilon^{(i)}$, defined in Eq. (27) for $i \in \{1, 2, 3\}$. Here, $\min_{(x,y)\in\Omega^\epsilon}(c_\epsilon^{(i)})$ was found as 0.0062, 0.0061, and 0.0063, and $\max_{(x,y)\in\Omega^\epsilon}(c_\epsilon^{(i)})$ was found as 1.0061, 1.0061, and 0.1378, for $i = 1$, 2, and 3, respectively

procedure at each sampling point. As a result, parallelization can significantly expedite the search for applicability conditions.

Similar to the observation in the first example problem, less restrictive applicability conditions are found by Symbolica than those found in the work of Boso and Battiato (2013) for the same assumed first-order solution, the same assumption on the fluid velocity, and the same ordered equation sets considered. The sufficient conditions identified by Boso and Battiato (2013) are

1. $\mathrm{Pe}_1 < \epsilon^{-2}$,
2. $\mathrm{Da}_1 < \epsilon^{-2}$,
3. $\mathrm{Da}_2 < \epsilon^{-2}$,
4. $\mathrm{Da}_3 < 1$,
5. $\mathrm{Da}_3/\mathrm{Pe}_1 < \epsilon$,
6. $\mathrm{Da}_1/\mathrm{Pe}_1 < \epsilon^{-1}$,
7. $\mathrm{Da}_2/\mathrm{Pe}_1 < \epsilon^{-1}$.

However, Symbolica successfully homogenized systems while violating conditions 5, 6, and/or 7 by direct sampling, which suggests that such conditions are overly restrictive.

For consistency with Boso and Battiato (2013), we define α, β, γ, and δ such that $\mathrm{Pe}_1 = \epsilon^{-\alpha}$, $\mathrm{Da}_1 = \epsilon^{\gamma}$, $\mathrm{Da}_2 = \epsilon^{\delta}$, and $\mathrm{Da}_3 = \epsilon^{\beta}$. To provide numerical validation of the extended applicability region, we compare averaged solutions from the pore-scale model, described in Eqs. (25a) – (25f), and the homogenized models for the three combinations of $(\alpha, \beta, \gamma, \delta)$ reported in Table 4. We note that these combinations violate applicability conditions 5, 6, and/or 7, but satisfy conditions 1-4: scenario 1 violates conditions 5, 6, and 7; scenario 2 violates condition 5 ($\alpha + \beta > 1$); scenario 3 violates conditions 6 ($\alpha + \gamma > -1$) and 7 ($\alpha + \delta > -1$). Additionally, we note that scenarios 2 and 3 are taken directly from Boso and Battiato (2013). In solving the models, we use the same problem setup as that presented in Boso and Battiato (2013): species A and B are spatially separated in a channel at $t = 0$ with an initial absence of species C. While the general parameters, boundary conditions, and initial conditions applied to the models are found in Table 5, the homogenized equations and closure problems corresponding to each scenario are found in Appendix B.

To visualize the simulation results from scenario 1, we provide Fig. 6. A comparison between the homogenized solutions and spatially averaged pore-scale solutions is shown in Fig. 6a, where the homogenized model qualitatively captures the transport behavior at the continuum scale. Further quantitative evidence of this agreement is provided in Fig. 6b, where the absolute error $E^{(i)}$, as defined in Eq. (22), remains less than ϵ throughout the domain despite violating applicability conditions 5, 6, and 7. Finally, in Fig. 6c, the normalized pore-scale concentration profiles for the three concentrations, defined as

$$\tilde{c}_\epsilon^{(i)} = \frac{c_\epsilon^{(i)} - \min\limits_{(x,y)\in\Omega^\epsilon}(c_\epsilon^{(i)})}{\max\limits_{(x,y)\in\Omega^\epsilon}(c_\epsilon^{(i)}) - \min\limits_{(x,y)\in\Omega^\epsilon}(c_\epsilon^{(i)})}, \tag{27}$$

for $i \in \{1, 2, 3\}$, show that all species are well-mixed across the channel, which intuitively supports the expectation that the system is homogenizable.

Despite violating condition 5 in scenario 2, the homogenized model qualitatively captures the averaged pore-scale dynamics in Fig. 7a. Further support for this agreement is provided in Fig. 7b, where $E^{(i)}$ again remains less than ϵ throughout the domain. Lastly, the normalized concentration contours from the pore-scale model in Fig. 7c have shifted to the right, as compared to those in Fig. 6c, due to a greater advective effect. However, diffusive transport still controls the transverse mixing due to a small channel width. This causes the isoconcentration curves to remain as vertical lines for all species, as opposed to parabolic in shape.

Finally, Fig. 8a shows that the homogenized model of scenario 3 qualitatively captures the averaged transport behavior of the pore-scale solution, despite violating conditions 6 and 7. Compared to scenario 2, a weaker heterogeneous reaction has been applied; however, the gradients of $\langle c^{(3)} \rangle_{\mathcal{B}}$ around $x = 0.45$ and $x = 0.65$ are slightly steeper than those shown in Fig. 7a. This is due to the greater non-linear homogeneous reaction occurring in the liquid phase of scenario 3. Further quantitative evidence of model agreement is provided in Fig. 8b, where $E^{(i)}$ remains less than ϵ throughout the domain. Again, the normalized concentration contours from the pore-scale model in Fig. 8c display vertical isoconcentration lines for all species. Similar to those in Fig. 7c, the contours have been shifted to the right due to the greater advective effect.

Overall, the three numerical experiments provide evidence that the applicability of the homogenized models extends beyond conditions 5, 6, and 7 found in Boso and Battiato (2013). This further underscores the value of automating the direct sampling approach for finding applicability conditions.

Table 6 Values assigned to the dimensionless numbers in each of the two considered scenarios

	Scenario 1	Scenario 2
Pe_1	ϵ^2	ϵ^{-1}
Da_i	ϵ^{-1}	ϵ^0
Da_{i+10}	ϵ^1	ϵ^1
η_i	ϵ^0	ϵ^0

Here, we let $i \in \mathbb{M} = \{m \in \mathbb{Z}^+ : m \leq 10\}$

4.4 Example Problem 3: Ten Species, Non-linear Homogeneous Reactions

4.4.1 Reaction Description and Preparation

In the final example problem, we demonstrate Symbolica's capability to handle large, coupled systems of equations by homogenizing an arbitrary ten-species ADR system. We consider both linear and non-linear homogeneous reactions in the liquid phase and assume linear heterogeneous reactions at the liquid-solid interface for each species. In general, we write the model reaction as

$$A + B \rightleftharpoons C, \tag{28a}$$
$$C + D \rightleftharpoons S, \tag{28b}$$
$$S \rightarrow K + D, \tag{28c}$$
$$K \rightarrow T, \tag{28d}$$
$$T + A \rightarrow B, \tag{28e}$$
$$W + K \rightleftharpoons M, \tag{28f}$$
$$G + A \rightarrow C, \tag{28g}$$

where each letter indicates a unique species. The dimensional input system describing the reactive transport is written using Eqs. (3a) and (3c), i.e.,

$$\frac{\partial \hat{c}_\epsilon^{(i)}}{\partial \hat{t}} + \hat{\mathbf{u}}_\epsilon \cdot \hat{\nabla} \hat{c}_\epsilon^{(i)} - \hat{D}^{(i)} \hat{\nabla}^2 \hat{c}_\epsilon^{(i)} = \hat{R}_\epsilon^{(i)} \quad \text{in } \hat{\mathcal{B}}^\epsilon, \tag{29a}$$
$$-\mathbf{n} \cdot \hat{D}^{(i)} \hat{\nabla} \hat{c}_\epsilon^{(i)} = \hat{\mathcal{K}}_S^{(i)} \left(\hat{c}_\epsilon^{(i)^{n_i}} - \hat{C}^{(i)^{n_i}} \right) \quad \text{on } \hat{\Gamma}^\epsilon, \tag{29b}$$

where $i \in \mathbb{M} = \{m \in \mathbb{Z}^+ : m \leq 10\}$, $\hat{\mathbf{u}}_\epsilon$ is a known function that allows for scales to be separated, and $n_i = 1$. Here, we note that variables with superscripts "(1)", "(2)", ... "(10)" correspond to species A, B, C, D, G, K, M, S, T, and W, respectively. Additionally, we define $\hat{R}_\epsilon^{(i)}$ as

$$\hat{R}_\epsilon^{(1)} = -\hat{\mathcal{K}}_{NL}^{(1,2)} \hat{c}_\epsilon^{(1)} \hat{c}_\epsilon^{(2)} + \hat{\mathcal{K}}_L^{(3)} \hat{c}_\epsilon^{(3)} - \hat{\mathcal{K}}_{NL}^{(1,5)} \hat{c}_\epsilon^{(1)} \hat{c}_\epsilon^{(5)} - \hat{\mathcal{K}}_{NL}^{(1,9)} \hat{c}_\epsilon^{(1)} \hat{c}_\epsilon^{(9)}, \tag{30a}$$
$$\hat{R}_\epsilon^{(2)} = -\hat{\mathcal{K}}_{NL}^{(1,2)} \hat{c}_\epsilon^{(1)} \hat{c}_\epsilon^{(2)} + \hat{\mathcal{K}}_L^{(3)} \hat{c}_\epsilon^{(3)} + \hat{\mathcal{K}}_{NL}^{(1,9)} \hat{c}_\epsilon^{(1)} \hat{c}_\epsilon^{(9)}, \tag{30b}$$
$$\hat{R}_\epsilon^{(3)} = \hat{\mathcal{K}}_{NL}^{(1,2)} \hat{c}_\epsilon^{(1)} \hat{c}_\epsilon^{(2)} - \hat{\mathcal{K}}_L^{(3)} \hat{c}_\epsilon^{(3)} - \hat{\mathcal{K}}_{NL}^{(3,4)} \hat{c}_\epsilon^{(3)} \hat{c}_\epsilon^{(4)} + \hat{\mathcal{K}}_{NL}^{(1,5)} \hat{c}_\epsilon^{(1)} \hat{c}_\epsilon^{(5)} + \hat{\mathcal{K}}_L^{(8)_1} \hat{c}_\epsilon^{(8)}, \tag{30c}$$
$$\hat{R}_\epsilon^{(4)} = -\hat{\mathcal{K}}_{NL}^{(3,4)} \hat{c}_\epsilon^{(3)} \hat{c}_\epsilon^{(4)} + \hat{\mathcal{K}}_L^{(8)_1} \hat{c}_\epsilon^{(8)} + \hat{\mathcal{K}}_L^{(8)_2} \hat{c}_\epsilon^{(8)}, \tag{30d}$$
$$\hat{R}_\epsilon^{(5)} = -\hat{\mathcal{K}}_{NL}^{(1,5)} \hat{c}_\epsilon^{(1)} \hat{c}_\epsilon^{(5)}, \tag{30e}$$
$$\hat{R}_\epsilon^{(6)} = -\hat{\mathcal{K}}_L^{(6)} \hat{c}_\epsilon^{(6)} + \hat{\mathcal{K}}_L^{(7)} \hat{c}_\epsilon^{(7)} + \hat{\mathcal{K}}_L^{(8)_2} \hat{c}_\epsilon^{(8)} - \hat{\mathcal{K}}_{NL}^{(6,10)} \hat{c}_\epsilon^{(6)} \hat{c}_\epsilon^{(10)}, \tag{30f}$$
$$\hat{R}_\epsilon^{(7)} = -\hat{\mathcal{K}}_L^{(7)} \hat{c}_\epsilon^{(7)} + \hat{\mathcal{K}}_{NL}^{(6,10)} \hat{c}_\epsilon^{(6)} \hat{c}_\epsilon^{(10)}, \tag{30g}$$

$$\hat{R}_\epsilon^{(8)} = \hat{\mathcal{K}}_{NL}^{(3,4)} \hat{c}_\epsilon^{(3)} \hat{c}_\epsilon^{(4)} - \hat{\mathcal{K}}_L^{(8)_1} \hat{c}_\epsilon^{(8)} - \hat{\mathcal{K}}_L^{(8)_2} \hat{c}_\epsilon^{(8)}, \tag{30h}$$

$$\hat{R}_\epsilon^{(9)} = \hat{\mathcal{K}}_L^{(6)} \hat{c}_\epsilon^{(6)} - \hat{\mathcal{K}}_{NL}^{(1,9)} \hat{c}_\epsilon^{(1)} \hat{c}_\epsilon^{(9)}, \tag{30i}$$

$$\hat{R}_\epsilon^{(10)} = \hat{\mathcal{K}}_L^{(7)} \hat{c}_\epsilon^{(7)} - \hat{\mathcal{K}}_{NL}^{(6,10)} \hat{c}_\epsilon^{(6)} \hat{c}_\epsilon^{(10)}. \tag{30j}$$

With respect to the input form provided in Eq. (3b), we have let $\hat{\mathcal{K}}_L^{(3,8)} = \hat{\mathcal{K}}_L^{(8)_1}$, $\hat{\mathcal{K}}_L^{(6,8)} = \hat{\mathcal{K}}_L^{(8)_2}$, $\hat{\mathcal{K}}_L^{(4,8)} = \hat{\mathcal{K}}_L^{(8,8)} = \hat{\mathcal{K}}_L^{(8)_1} + \hat{\mathcal{K}}_L^{(8)_2} = \hat{\mathcal{K}}_L^{(8)}$, $\hat{\mathcal{K}}_{NL}^{(i,j,k)} = \hat{\mathcal{K}}_{NL}^{(j,k)}$, $\hat{\mathcal{K}}_L^{(i,q)} = \hat{\mathcal{K}}_L^{(q)}$, and $p_L^{(i,j)}$ and $p_{NL}^{(i,j,k)}$ have been chosen accordingly. Here, $i \in \mathbb{M}$, $j \in \mathbb{M}$, $k \in \{m \in \mathbb{Z}^+ : j \leq m \leq 10\}$ for a given j, and $q \in \mathbb{M} \setminus \{8\}$.

To scale the system, we allow Symbolica to assume the scales in Eq. (9) and reduce the number of scales by letting $\hat{\mathcal{C}}^{(i)} = \hat{\mathcal{C}}^{(*)}$ for $i \in \mathbb{M}$. In doing so, Symbolica finds the dimensionless system

$$\frac{\partial c_\epsilon^{(i)}}{\partial t} + \mathrm{Pe}_1 \mathbf{u}_\epsilon \cdot \nabla c_\epsilon^{(i)} - D^{(i)} \nabla^2 c_\epsilon^{(i)} = R_\epsilon^{(i)} \quad \text{in } \mathcal{B}^\epsilon, \tag{31a}$$

$$-\mathbf{n} \cdot D^{(i)} \nabla c_\epsilon^{(i)} = \mathrm{Da}_{i+10} \left(c_\epsilon^{(i)} - \eta_i \right) \quad \text{on } \Gamma^\epsilon, \tag{31b}$$

where $R_\epsilon^{(i)}$ is defined as

$$R_\epsilon^{(1)} = -\mathrm{Da}_1 c_\epsilon^{(1)} c_\epsilon^{(2)} + \mathrm{Da}_2 c_\epsilon^{(3)} - \mathrm{Da}_3 c_\epsilon^{(1)} c_\epsilon^{(5)} - \mathrm{Da}_4 c_\epsilon^{(1)} c_\epsilon^{(9)}, \tag{32a}$$

$$R_\epsilon^{(2)} = -\mathrm{Da}_1 c_\epsilon^{(1)} c_\epsilon^{(2)} + \mathrm{Da}_2 c_\epsilon^{(3)} + \mathrm{Da}_4 c_\epsilon^{(1)} c_\epsilon^{(9)}, \tag{32b}$$

$$R_\epsilon^{(3)} = \mathrm{Da}_1 c_\epsilon^{(1)} c_\epsilon^{(2)} - \mathrm{Da}_2 c_\epsilon^{(3)} - \mathrm{Da}_5 c_\epsilon^{(3)} c_\epsilon^{(4)} + \mathrm{Da}_3 c_\epsilon^{(1)} c_\epsilon^{(5)} + \mathrm{Da}_6 c_\epsilon^{(8)}, \tag{32c}$$

$$R_\epsilon^{(4)} = -\mathrm{Da}_5 c_\epsilon^{(3)} c_\epsilon^{(4)} + \mathrm{Da}_6 c_\epsilon^{(8)} + \mathrm{Da}_7 c_\epsilon^{(8)}, \tag{32d}$$

$$R_\epsilon^{(5)} = -\mathrm{Da}_3 c_\epsilon^{(1)} c_\epsilon^{(5)}, \tag{32e}$$

$$R_\epsilon^{(6)} = -\mathrm{Da}_8 c_\epsilon^{(6)} + \mathrm{Da}_9 c_\epsilon^{(7)} + \mathrm{Da}_7 c_\epsilon^{(8)} - \mathrm{Da}_{10} c_\epsilon^{(6)} c_\epsilon^{(10)}, \tag{32f}$$

$$R_\epsilon^{(7)} = -\mathrm{Da}_9 c_\epsilon^{(7)} + \mathrm{Da}_{10} c_\epsilon^{(6)} c_\epsilon^{(10)}, \tag{32g}$$

$$R_\epsilon^{(8)} = \mathrm{Da}_5 c_\epsilon^{(3)} c_\epsilon^{(4)} - \mathrm{Da}_6 c_\epsilon^{(8)} - \mathrm{Da}_7 c_\epsilon^{(8)}, \tag{32h}$$

$$R_\epsilon^{(9)} = \mathrm{Da}_8 c_\epsilon^{(6)} - \mathrm{Da}_4 c_\epsilon^{(1)} c_\epsilon^{(9)}, \tag{32i}$$

$$R_\epsilon^{(10)} = \mathrm{Da}_9 c_\epsilon^{(7)} - \mathrm{Da}_{10} c_\epsilon^{(6)} c_\epsilon^{(10)}, \tag{32j}$$

prior to expanding the operators and dependent variables. Here, Symbolica has defined one Péclet number, 20 Damköhler numbers, and 10 concentration ratios as

$$\begin{aligned}
&\mathrm{Pe}_1 = \frac{\hat{\mathcal{U}}\hat{\mathcal{L}}}{\hat{\mathcal{D}}}, \quad \mathrm{Da}_1 = \frac{\hat{\mathcal{K}}_{NL}^{(1,2)} \hat{\mathcal{L}}^2 \hat{\mathcal{C}}^{(*)}}{\hat{\mathcal{D}}}, \quad \mathrm{Da}_2 = \frac{\hat{\mathcal{K}}_L^{(3)} \hat{\mathcal{L}}^2}{\hat{\mathcal{D}}}, \quad \mathrm{Da}_3 = \frac{\hat{\mathcal{K}}_{NL}^{(1,5)} \hat{\mathcal{L}}^2 \hat{\mathcal{C}}^{(*)}}{\hat{\mathcal{D}}}, \\
&\mathrm{Da}_4 = \frac{\hat{\mathcal{K}}_{NL}^{(1,9)} \hat{\mathcal{L}}^2 \hat{\mathcal{C}}^{(*)}}{\hat{\mathcal{D}}}, \quad \mathrm{Da}_5 = \frac{\hat{\mathcal{K}}_{NL}^{(3,4)} \hat{\mathcal{L}}^2 \hat{\mathcal{C}}^{(*)}}{\hat{\mathcal{D}}}, \quad \mathrm{Da}_6 = \frac{\hat{\mathcal{K}}_L^{(8)_{(1)}} \hat{\mathcal{L}}^2}{\hat{\mathcal{D}}}, \quad \mathrm{Da}_7 = \frac{\hat{\mathcal{K}}_L^{(8)_{(2)}} \hat{\mathcal{L}}^2}{\hat{\mathcal{D}}}, \\
&\mathrm{Da}_8 = \frac{\hat{\mathcal{K}}_L^{(6)} \hat{\mathcal{L}}^2}{\hat{\mathcal{D}}}, \quad \mathrm{Da}_9 = \frac{\hat{\mathcal{K}}_L^{(7)} \hat{\mathcal{L}}^2}{\hat{\mathcal{D}}}, \quad \mathrm{Da}_{10} = \frac{\hat{\mathcal{K}}_{NL}^{(6,10)} \hat{\mathcal{L}}^2 \hat{\mathcal{C}}^{(*)}}{\hat{\mathcal{D}}}, \quad \mathrm{Da}_{i+10} = \frac{\hat{\mathcal{K}}_S^{(i)} \hat{\mathcal{L}}}{\hat{\mathcal{D}}}, \quad \eta_i = \frac{\hat{\mathcal{C}}^{(i)}}{\hat{\mathcal{C}}^{(*)}},
\end{aligned} \tag{33}$$

for $i \in \mathbb{M}$. Because of the high computational cost required to find the applicability conditions from the 31-dimensional space of values, we instead allow Symbolica to homogenize the system for the two scenarios of dimensionless number values listed in Table 6. We note that transport is controlled by reaction in scenario 1 and by advection in scenario 2.

Table 7 The simulation parameters, boundary conditions, and initial conditions used to numerically solve the homogenized and the pore-scale models for the third example problem

Homogenized model		Pore-scale model	
Simulation parameters			
Maximum $\Delta x = 1 \times 10^{-5}$		Maximum $\Delta x \Delta y = 1 \times 10^{-5}$	
Maximum $\Delta t = 1 \times 10^{-3}$		Maximum $\Delta t = 1 \times 10^{-3}$	
Simulation boundary conditions			
$\partial \left\langle c^{(i)} \right\rangle_{\mathcal{B}} / \partial x = 0$ for $x \in \partial\Omega_0 \cup \partial\Omega_1,\ t > 0$		$\partial c_\epsilon^{(i)} / \partial x = 0$ for $(x, y) \in \partial\mathcal{B}_0^\epsilon \cup \partial\mathcal{B}_1^\epsilon,\ t > 0$	
Simulation initial conditions			
$\left\langle c^{(1)} \right\rangle_{\mathcal{B}} = 0.1 + 0.8H\,(\tilde{x}_1 - x)$	for $x \in \Omega,\ t = 0$	$c_\epsilon^{(1)} = 0.1 + 0.8H\,(\tilde{x}_1 - x)$	for $(x, y) \in \mathcal{B}^\epsilon,\ t = 0$
$\left\langle c^{(2)} \right\rangle_{\mathcal{B}} = 0.1 + 0.8H\,(x - \tilde{x}_1)$	for $x \in \Omega,\ t = 0$	$c_\epsilon^{(2)} = 0.1 + 0.8H\,(x - \tilde{x}_1)$	for $(x, y) \in \mathcal{B}^\epsilon,\ t = 0$
$\left\langle c^{(3)} \right\rangle_{\mathcal{B}} = 0.2$	for $x \in \Omega,\ t = 0$	$c_\epsilon^{(3)} = 0.2$	for $(x, y) \in \mathcal{B}^\epsilon,\ t = 0$
$\left\langle c^{(4)} \right\rangle_{\mathcal{B}} = 0.4 + 0.4H\,(\tilde{x}_2 - x)$	for $x \in \Omega,\ t = 0$	$c_\epsilon^{(4)} = 0.4 + 0.4H\,(\tilde{x}_2 - x)$	for $(x, y) \in \mathcal{B}^\epsilon,\ t = 0$
$\left\langle c^{(5)} \right\rangle_{\mathcal{B}} = 0.4 + 0.5H\,(x - \tilde{x}_4)$	for $x \in \Omega,\ t = 0$	$c_\epsilon^{(5)} = 0.4 + 0.5H\,(x - \tilde{x}_4)$	for $(x, y) \in \mathcal{B}^\epsilon,\ t = 0$
$\left\langle c^{(6)} \right\rangle_{\mathcal{B}} = 0.5 + 0.2H\,(x - \tilde{x}_4)$	for $x \in \Omega,\ t = 0$	$c_\epsilon^{(6)} = 0.5 + 0.2H\,(x - \tilde{x}_4)$	for $(x, y) \in \mathcal{B}^\epsilon,\ t = 0$
$\left\langle c^{(7)} \right\rangle_{\mathcal{B}} = 0.1 + 0.8H\,(x - \tilde{x}_1)$	for $x \in \Omega,\ t = 0$	$c_\epsilon^{(7)} = 0.1 + 0.8H\,(x - \tilde{x}_1)$	for $(x, y) \in \mathcal{B}^\epsilon,\ t = 0$
$\left\langle c^{(8)} \right\rangle_{\mathcal{B}} = 0.2 + 0.5H\,(x - \tilde{x}_3)$	for $x \in \Omega,\ t = 0$	$c_\epsilon^{(8)} = 0.2 + 0.5H\,(x - \tilde{x}_3)$	for $(x, y) \in \mathcal{B}^\epsilon,\ t = 0$
$\left\langle c^{(9)} \right\rangle_{\mathcal{B}} = 0.3 + 0.3H\,(\tilde{x}_1 - x)$	for $x \in \Omega,\ t = 0$	$c_\epsilon^{(9)} = 0.3 + 0.3H\,(\tilde{x}_1 - x)$	for $(x, y) \in \mathcal{B}^\epsilon,\ t = 0$
$\left\langle c^{(10)} \right\rangle_{\mathcal{B}} = 0.2 + 0.5H\,(\tilde{x}_4 - x)$	for $x \in \Omega,\ t = 0$	$c_\epsilon^{(10)} = 0.2 + 0.5H\,(\tilde{x}_4 - x)$	for $(x, y) \in \mathcal{B}^\epsilon,\ t = 0$
$\tilde{x}_1 = \begin{cases} 0.3 \text{ for scenario 1} \\ 0.15 \text{ for scenario 2} \end{cases}, \quad \tilde{x}_2 = \tilde{x}_1 + 0.1, \quad \tilde{x}_3 = \tilde{x}_1 + 0.25, \quad \tilde{x}_4 = \tilde{x}_1 + 0.4$			

Here, we let $i \in \mathbb{M} = \{m \in \mathbb{Z}^+ : m \leq 10\}$ and $H(x)$ is the Heaviside function

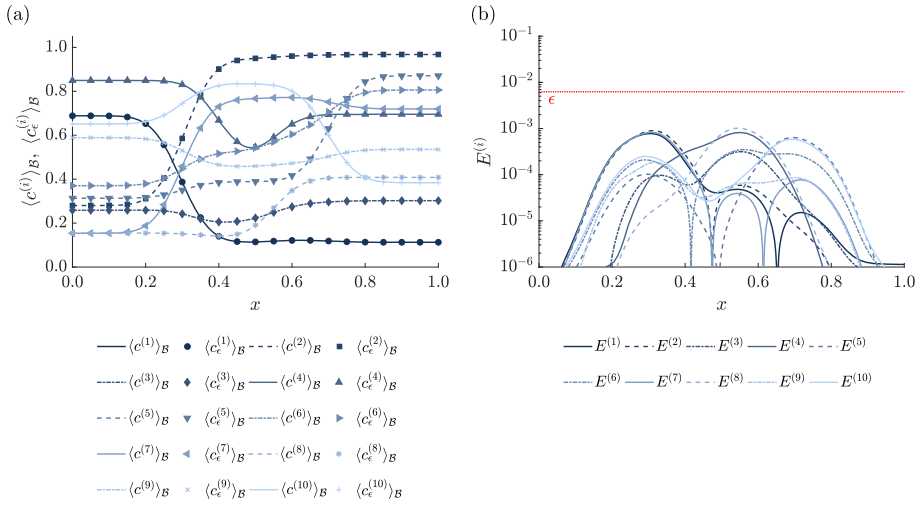

Fig. 9 The numerical results from scenario 1 of the third example problem at $t = 0.002$, where $\epsilon = 0.00625$. **(a)** The $\mathcal{B}$-averaged concentration profiles from the homogenized model and the spatially averaged pore-scale solutions from Eqs. (31a)–(32j). **(b)** The absolute error $E^{(i)}$ in the x-direction, defined in Eq. (22) where $i \in \mathbb{M} = \{m \in \mathbb{Z}^+ : m \leq 10\}$. The red dotted line provides the upper limit on the error predicted by the homogenized model

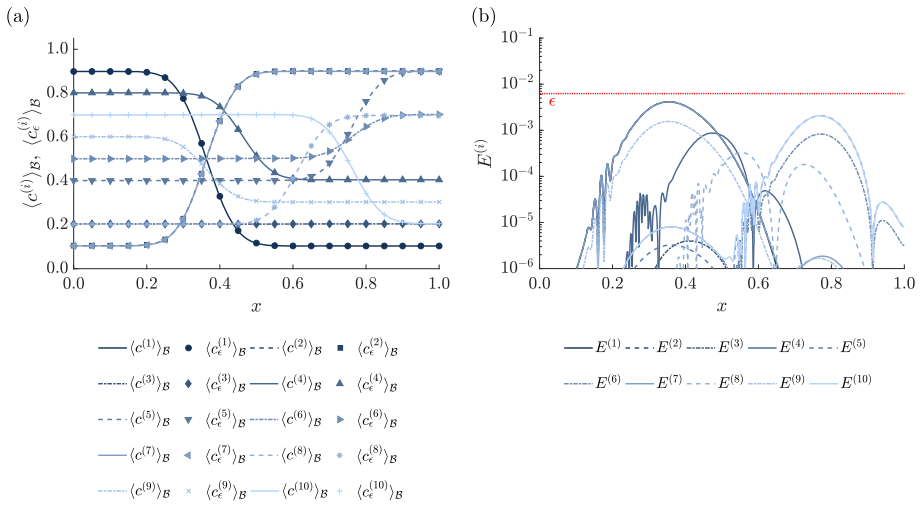

Fig. 10 The numerical results from scenario 2 of the third example problem at $t = 0.002$, where $\epsilon = 0.00625$. **(a)** The $\mathcal{B}$-averaged concentration profiles from the homogenized model and the spatially averaged pore-scale solution from Eqs. (31a)–(32j). **(b)** The absolute error $E^{(i)}$ in the x-direction, defined in Eq. (22) where $i \in \mathbb{M} = \{m \in \mathbb{Z}^+ : m \leq 10\}$. The red dotted line provides the upper limit on the error predicted by the homogenized model

4.4.2 Symbolic Upscaling Results and Numerical Verification

The execution time of the entire automated upscaling procedure was roughly 1 minute and 4.82 seconds, where a detailed breakdown of the time is provided in Table 3. We note that completing the upscaling phase for both scenarios took roughly 56.05 seconds, which averages to about 28.03 seconds per scenario. We attribute the increase in average time per iteration of the upscaling phase to the increased system size and reemphasize the potential of parallelization in cutting down computational time.

We now verify the homogenization procedure implemented in Symbolica for complex systems by comparing the averaged solutions from the pore-scale model, described in Eqs. (31a)–(32j), and the homogenized models for scenarios 1 and 2 in Table 6. We choose to simulate a transient problem where the initial concentration profiles contain discontinuities at various locations in the domain. The general simulation parameters, boundary conditions, and initial conditions applied to the models are found in Table 7. After solving both models, the numerical solution of the pore-scale model is averaged over the channel width to obtain $\langle c_\epsilon^{(i)} \rangle_{\mathcal{B}}$ and the absolute error is calculated using Eq. (22), where $i \in \mathbb{M}$.

To visualize the simulation results from scenario 1, we provide Fig. 9. A comparison between the homogenized solutions and spatially averaged pore-scale solutions is shown in Fig. 9a, where the homogenized model captures the complex gradients experienced by nearly all species throughout the domain. Quantitative validation of the homogenized model may be found in Fig. 9b, where all $E^{(i)}$ remain less than ϵ throughout the domain. Figure 10 shows the results from scenario 2 where advection dominates. As shown in Fig. 10a, the homogenized model qualitatively captures the advective behavior of the averaged pore-scale solution. Again, quantitative verification of the homogenized model may be found in Fig. 10b, where all $E^{(i)}$ remain less than ϵ throughout the domain as predicted by the upscaling procedure. These results further support the verification of Symbolica for homogenizing complex ADR systems.

In finding agreement between the homogenized and spatially averaged pore-scale solutions for the ten-species system, we show that Symbolica's homogenization procedure is robust. While manual homogenization of a ten-species system is time-intensive at best, utilization of Symbolica allocates the work to computational resources and allows for homogenized models to be produced in as little as 30 seconds. This shows that automation and upscaling can be successfully combined to study systems previously deemed unfeasible to model or upscale manually.

5 Conclusion

Multiscale model development is often time-intensive and requires specialized expertise to navigate cost-accuracy tradeoffs specific to individual applications. The complex systems found in practice create an obstacle for the development and deployment of such models, which leads practitioners to pursue more tractable alternatives optimized for deployment. Yet, multiscale models host many benefits desired in the work of practitioners. In particular, benefits including computational efficiency and affordability are provided through rigorous upscaling with applicability conditions, *a priori* error estimates, and the potential for lack of tuning parameters. To combat the challenges practitioners face associated with model development via upscaling, we developed and validated a software, Symbolica, able to perform automatic symbolic upscaling based on classical homogenization theory with minimal user

input. We then tested Symbolica's performance on three example problems of increasing complexity and demonstrated Symbolica's efficiency, accuracy, and robustness in handling reactive systems of numerous reacting species and different dynamical conditions (i.e., different combinations of dimensionless number values). Additionally, we showed that the direct sampling approach implemented by Symbolica allows for the derivation of applicability conditions that are less restrictive than those previously obtained through manual upscaling.

In validation of the automated upscaling procedure implemented by Symbolica for Eqs. (3a)–(3c), we consider future modifications for advancing the software:

1. Automating and generalizing the assumed closure forms;
2. Expanding the derivation to higher order terms;
3. Implementing time-averaging procedures.

In addition to the proposed technical modifications, which can increase the interdisciplinary impact of the software, an important step towards democratizing the utilization of upscaling is creating a version of Symbolica using an open-source symbolic framework. This will ensure equitable access to all, and as a result, expand the accessibility of valuable upscaling techniques to a wide, interdisciplinary community.

Acknowledgements This material is based upon work supported by the Defense Advanced Research Projects Agency (DARPA) under Agreement No. HR00112090061. The views, opinions and/or findings expressed are those of the authors and should not be interpreted as representing the official views or policies of the Department of Defense or U.S. Government. KP was also supported by the Stanford Graduate Fellowship in Science and Engineering.

Appendix A

The homogenized equations and closure problems solved for the first example problem at each indicated (α, β) coordinate in Fig. 5.

$(\alpha, \beta) : (1, 1)$

$$\frac{\partial \left\langle c^{(1)} \right\rangle_{\mathcal{B}}}{\partial t} + \frac{1}{\epsilon} \langle \mathbf{u} \rangle_{\mathcal{B}} \cdot \nabla_{\mathbf{x}} \left\langle c^{(1)} \right\rangle_{\mathcal{B}} - \nabla_{\mathbf{x}} \cdot \left[\mathbf{D}^{(1)} \cdot \nabla_{\mathbf{x}} \left\langle c^{(1)} \right\rangle_{\mathcal{B}} \right] + \frac{|\Gamma|}{|\mathcal{B}|} \left(\left\langle c^{(1)} \right\rangle_{\mathcal{B}}^2 - 1 \right) = 0 \quad \text{in } \Omega, \tag{34a}$$

$$\mathbf{D}^{(1)} = \left\langle D^{(1)} \left(\mathbf{I} + \nabla_{\boldsymbol{\xi}} \boldsymbol{\chi}^{(1)} \right) \right\rangle_{\mathcal{B}} - \left\langle \boldsymbol{\chi}^{(1)} \otimes \mathbf{u}_0 \right\rangle_{\mathcal{B}}, \tag{34b}$$

$$\mathbf{u}_0 \cdot \left(\mathbf{I} + \nabla_{\boldsymbol{\xi}} \boldsymbol{\chi}^{(1)} \right) - \langle \mathbf{u}_0 \rangle_{\mathcal{B}} - \nabla_{\boldsymbol{\xi}} \cdot \left[D^{(1)} \left(\mathbf{I} + \nabla_{\boldsymbol{\xi}} \boldsymbol{\chi}^{(1)} \right) \right] = \mathbf{0} \quad \text{in } \mathcal{B}, \tag{34c}$$

$$-\mathbf{n} \cdot \left[D^{(1)} \left(\mathbf{I} + \nabla_{\boldsymbol{\xi}} \boldsymbol{\chi}^{(1)} \right) \right] = \mathbf{0} \quad \text{on } \Gamma. \tag{34d}$$

$(\alpha, \beta) : (0, 1)$

$$\frac{\partial \left\langle c^{(1)} \right\rangle_{\mathcal{B}}}{\partial t} + \langle \mathbf{u} \rangle_{\mathcal{B}} \cdot \nabla_{\mathbf{x}} \left\langle c^{(1)} \right\rangle_{\mathcal{B}} - \nabla_{\mathbf{x}} \cdot \left[\mathbf{D}^{(1)} \cdot \nabla_{\mathbf{x}} \left\langle c^{(1)} \right\rangle_{\mathcal{B}} \right] + \frac{|\Gamma|}{|\mathcal{B}|} \left(\left\langle c^{(1)} \right\rangle_{\mathcal{B}}^2 - 1 \right) = 0 \quad \text{in } \Omega, \tag{35a}$$

$$\mathbf{D}^{(1)} = \left\langle D^{(1)} \left(\mathbf{I} + \nabla_{\boldsymbol{\xi}} \boldsymbol{\chi}^{(1)} \right) \right\rangle_{\mathcal{B}}, \tag{35b}$$

$$\nabla_{\boldsymbol{\xi}} \cdot \left[D^{(1)} \left(\mathbf{I} + \nabla_{\boldsymbol{\xi}} \boldsymbol{\chi}^{(1)} \right) \right] = \mathbf{0} \quad \text{in } \mathcal{B}, \tag{35c}$$

$$-\mathbf{n} \cdot \left[D^{(1)} \left(\mathbf{I} + \nabla_{\boldsymbol{\xi}} \boldsymbol{\chi}^{(1)} \right) \right] = \mathbf{0} \quad \text{on } \Gamma. \tag{35d}$$

$(\alpha, \beta):\ (-1,1)\,,\ (-2,1)\,,\ (-3,1)$

$$\frac{\partial \left\langle c^{(1)} \right\rangle_{\mathcal{B}}}{\partial t} - \nabla_{\mathbf{x}} \cdot \left[\mathbf{D}^{(1)} \cdot \nabla_{\mathbf{x}} \left\langle c^{(1)} \right\rangle_{\mathcal{B}} \right] + \frac{|\Gamma|}{|\mathcal{B}|} \left(\left\langle c^{(1)} \right\rangle_{\mathcal{B}}^2 - 1 \right) = 0 \quad \text{in } \Omega, \tag{36a}$$

$$\mathbf{D}^{(1)} = \left\langle D^{(1)} \left(\mathbf{I} + \nabla_{\boldsymbol{\xi}} \boldsymbol{\chi}^{(1)} \right) \right\rangle_{\mathcal{B}}, \tag{36b}$$

$$\nabla_{\boldsymbol{\xi}} \cdot \left[D^{(1)} \left(\mathbf{I} + \nabla_{\boldsymbol{\xi}} \boldsymbol{\chi}^{(1)} \right) \right] = \mathbf{0} \quad \text{in } \mathcal{B}, \tag{36c}$$

$$-\mathbf{n} \cdot \left[D^{(1)} \left(\mathbf{I} + \nabla_{\boldsymbol{\xi}} \boldsymbol{\chi}^{(1)} \right) \right] = \mathbf{0} \quad \text{on } \Gamma. \tag{36d}$$

$(\alpha, \beta):\ (1,2)\,,\ (1,3)$

$$\frac{\partial \left\langle c^{(1)} \right\rangle_{\mathcal{B}}}{\partial t} + \frac{1}{\epsilon} \langle \mathbf{u} \rangle_{\mathcal{B}} \cdot \nabla_{\mathbf{x}} \left\langle c^{(1)} \right\rangle_{\mathcal{B}} - \nabla_{\mathbf{x}} \cdot \left[\mathbf{D}^{(1)} \cdot \nabla_{\mathbf{x}} \left\langle c^{(1)} \right\rangle_{\mathcal{B}} \right] = 0 \quad \text{in } \Omega, \tag{37a}$$

$$\mathbf{D}^{(1)} = \left\langle D^{(1)} \left(\mathbf{I} + \nabla_{\boldsymbol{\xi}} \boldsymbol{\chi}^{(1)} \right) \right\rangle_{\mathcal{B}} - \left\langle \boldsymbol{\chi}^{(1)} \otimes \mathbf{u}_0 \right\rangle_{\mathcal{B}}, \tag{37b}$$

$$\mathbf{u}_0 \cdot \left(\mathbf{I} + \nabla_{\boldsymbol{\xi}} \boldsymbol{\chi}^{(1)} \right) - \langle \mathbf{u}_0 \rangle_{\mathcal{B}} - \nabla_{\boldsymbol{\xi}} \cdot \left[D^{(1)} \left(\mathbf{I} + \nabla_{\boldsymbol{\xi}} \boldsymbol{\chi}^{(1)} \right) \right] = \mathbf{0} \quad \text{in } \mathcal{B}, \tag{37c}$$

$$-\mathbf{n} \cdot \left[D^{(1)} \left(\mathbf{I} + \nabla_{\boldsymbol{\xi}} \boldsymbol{\chi}^{(1)} \right) \right] = \mathbf{0} \quad \text{on } \Gamma. \tag{37d}$$

$(\alpha, \beta):\ (0,2)\,,\ (0,3)$

$$\frac{\partial \left\langle c^{(1)} \right\rangle_{\mathcal{B}}}{\partial t} + \langle \mathbf{u} \rangle_{\mathcal{B}} \cdot \nabla_{\mathbf{x}} \left\langle c^{(1)} \right\rangle_{\mathcal{B}} - \nabla_{\mathbf{x}} \cdot \left[\mathbf{D}^{(1)} \cdot \nabla_{\mathbf{x}} \left\langle c^{(1)} \right\rangle_{\mathcal{B}} \right] = 0 \quad \text{in } \Omega, \tag{38a}$$

$$\mathbf{D}^{(1)} = \left\langle D^{(1)} \left(\mathbf{I} + \nabla_{\boldsymbol{\xi}} \boldsymbol{\chi}^{(1)} \right) \right\rangle_{\mathcal{B}}, \tag{38b}$$

$$\nabla_{\boldsymbol{\xi}} \cdot \left[D^{(1)} \left(\mathbf{I} + \nabla_{\boldsymbol{\xi}} \boldsymbol{\chi}^{(1)} \right) \right] = \mathbf{0} \quad \text{in } \mathcal{B}, \tag{38c}$$

$$-\mathbf{n} \cdot \left[D^{(1)} \left(\mathbf{I} + \nabla_{\boldsymbol{\xi}} \boldsymbol{\chi}^{(1)} \right) \right] = \mathbf{0} \quad \text{on } \Gamma. \tag{38d}$$

$(\alpha, \beta):\ (-1,2)\,,\ (-2,2)\,,\ (-3,2)\,,\ (-1,3)\,,\ (-2,3)\,,\ (-3,3)$

$$\frac{\partial \left\langle c^{(1)} \right\rangle_{\mathcal{B}}}{\partial t} - \nabla_{\mathbf{x}} \cdot \left[\mathbf{D}^{(1)} \cdot \nabla_{\mathbf{x}} \left\langle c^{(1)} \right\rangle_{\mathcal{B}} \right] = 0 \quad \text{in } \Omega, \tag{39a}$$

$$\mathbf{D}^{(1)} = \left\langle D^{(1)} \left(\mathbf{I} + \nabla_{\boldsymbol{\xi}} \boldsymbol{\chi}^{(1)} \right) \right\rangle_{\mathcal{B}}, \tag{39b}$$

$$\nabla_{\boldsymbol{\xi}} \cdot \left[D^{(1)} \left(\mathbf{I} + \nabla_{\boldsymbol{\xi}} \boldsymbol{\chi}^{(1)} \right) \right] = \mathbf{0} \quad \text{in } \mathcal{B}, \tag{39c}$$

$$-\mathbf{n} \cdot \left[D^{(1)} \left(\mathbf{I} + \nabla_{\boldsymbol{\xi}} \boldsymbol{\chi}^{(1)} \right) \right] = \mathbf{0} \quad \text{on } \Gamma. \tag{39d}$$

Appendix B

The homogenized equations and closure problems solved for the three scenarios of $(\alpha, \beta, \gamma, \delta)$ considered in the second example problem.

$(\alpha, \beta, \gamma, \delta) : (-2, 1/4, -7/4, -7/4)$

$$\frac{\partial \left\langle c^{(1)} \right\rangle_{\mathcal{B}}}{\partial t} - \nabla_{\mathbf{x}} \cdot \left(\mathbf{D}^{(1)} \cdot \nabla_{\mathbf{x}} \left\langle c^{(1)} \right\rangle_{\mathcal{B}} \right) = -\epsilon^{-\frac{7}{4}} \left\langle c^{(1)} \right\rangle_{\mathcal{B}} \left\langle c^{(2)} \right\rangle_{\mathcal{B}} + \epsilon^{-\frac{7}{4}} \left\langle c^{(3)} \right\rangle_{\mathcal{B}} \quad \text{in } \Omega, \tag{40a}$$

$$\frac{\partial \left\langle c^{(2)} \right\rangle_{\mathcal{B}}}{\partial t} - \nabla_{\mathbf{x}} \cdot \left(\mathbf{D}^{(2)} \cdot \nabla_{\mathbf{x}} \left\langle c^{(2)} \right\rangle_{\mathcal{B}} \right) = -\epsilon^{-\frac{7}{4}} \left\langle c^{(1)} \right\rangle_{\mathcal{B}} \left\langle c^{(2)} \right\rangle_{\mathcal{B}} + \epsilon^{-\frac{7}{4}} \left\langle c^{(3)} \right\rangle_{\mathcal{B}} \quad \text{in } \Omega, \tag{40b}$$

$$\frac{\partial \left\langle c^{(3)} \right\rangle_{\mathcal{B}}}{\partial t} - \nabla_{\mathbf{x}} \cdot \left(\mathbf{D}^{(3)} \cdot \nabla_{\mathbf{x}} \left\langle c^{(3)} \right\rangle_{\mathcal{B}} \right) = \epsilon^{-\frac{7}{4}} \left\langle c^{(1)} \right\rangle_{\mathcal{B}} \left\langle c^{(2)} \right\rangle_{\mathcal{B}} - \epsilon^{-\frac{7}{4}} \left\langle c^{(3)} \right\rangle_{\mathcal{B}} - \epsilon^{-\frac{3}{4}} \frac{|\Gamma|}{|\mathcal{B}|} \left(\left\langle c^{(3)} \right\rangle_{\mathcal{B}} - \eta_1 \right) \quad \text{in } \Omega, \tag{40c}$$

$$\mathbf{D}^{(i)} = \left\langle D^{(i)} \left(\mathbf{I} + \nabla_{\boldsymbol{\xi}} \boldsymbol{\chi}^{(i)} \right) \right\rangle_{\mathcal{B}}, \tag{40d}$$

$$\nabla_{\boldsymbol{\xi}} \cdot \left[D^{(i)} \left(\mathbf{I} + \nabla_{\boldsymbol{\xi}} \boldsymbol{\chi}^{(i)} \right) \right] = \mathbf{0} \quad \text{in } \mathcal{B}, \tag{40e}$$

$$-\mathbf{n} \cdot \left[D^{(i)} \left(\mathbf{I} + \nabla_{\boldsymbol{\xi}} \boldsymbol{\chi}^{(i)} \right) \right] = \mathbf{0} \quad \text{on } \Gamma. \tag{40f}$$

$(\alpha, \beta, \gamma, \delta) : (1/2, 1/4, -1, -1)$

$$\frac{\partial \left\langle c^{(1)} \right\rangle_{\mathcal{B}}}{\partial t} + \epsilon^{-\frac{1}{2}} \nabla_{\mathbf{x}} \cdot \left(\langle \mathbf{u} \rangle_{\mathcal{B}} \left\langle c^{(1)} \right\rangle_{\mathcal{B}} \right) - \nabla_{\mathbf{x}} \cdot \left(\mathbf{D}^{(1)} \cdot \nabla_{\mathbf{x}} \left\langle c^{(1)} \right\rangle_{\mathcal{B}} \right) = -\epsilon^{-1} \left\langle c^{(1)} \right\rangle_{\mathcal{B}} \left\langle c^{(2)} \right\rangle_{\mathcal{B}} + \epsilon^{-1} \left\langle c^{(3)} \right\rangle_{\mathcal{B}} \quad \text{in } \Omega, \tag{41a}$$

$$\frac{\partial \left\langle c^{(2)} \right\rangle_{\mathcal{B}}}{\partial t} + \epsilon^{-\frac{1}{2}} \nabla_{\mathbf{x}} \cdot \left(\langle \mathbf{u} \rangle_{\mathcal{B}} \left\langle c^{(2)} \right\rangle_{\mathcal{B}} \right) - \nabla_{\mathbf{x}} \cdot \left(\mathbf{D}^{(2)} \cdot \nabla_{\mathbf{x}} \left\langle c^{(2)} \right\rangle_{\mathcal{B}} \right) = -\epsilon^{-1} \left\langle c^{(1)} \right\rangle_{\mathcal{B}} \left\langle c^{(2)} \right\rangle_{\mathcal{B}} + \epsilon^{-1} \left\langle c^{(3)} \right\rangle_{\mathcal{B}} \quad \text{in } \Omega, \tag{41b}$$

$$\frac{\partial \left\langle c^{(3)} \right\rangle_{\mathcal{B}}}{\partial t} + \epsilon^{-\frac{1}{2}} \nabla_{\mathbf{x}} \cdot \left(\langle \mathbf{u} \rangle_{\mathcal{B}} \left\langle c^{(3)} \right\rangle_{\mathcal{B}} \right) - \nabla_{\mathbf{x}} \cdot \left(\mathbf{D}^{(3)} \cdot \nabla_{\mathbf{x}} \left\langle c^{(3)} \right\rangle_{\mathcal{B}} \right) = \epsilon^{-1} \left\langle c^{(1)} \right\rangle_{\mathcal{B}} \left\langle c^{(2)} \right\rangle_{\mathcal{B}} - \epsilon^{-1} \left\langle c^{(3)} \right\rangle_{\mathcal{B}} - \epsilon^{-\frac{3}{4}} \frac{|\Gamma|}{|\mathcal{B}|} \left(\left\langle c^{(3)} \right\rangle_{\mathcal{B}} - \eta_1 \right) \quad \text{in } \Omega, \tag{41c}$$

$$\mathbf{D}^{(i)} = \left\langle D^{(i)} \left(\mathbf{I} + \nabla_{\boldsymbol{\xi}} \boldsymbol{\chi}^{(i)} \right) \right\rangle_{\mathcal{B}} - \epsilon^{\frac{1}{2}} \left\langle \boldsymbol{\chi}^{(i)} \otimes \mathbf{u}_0 \right\rangle_{\mathcal{B}}, \tag{41d}$$

$$\epsilon^{\frac{1}{2}} \left[\mathbf{u}_0 \cdot \left(\mathbf{I} + \nabla_{\boldsymbol{\xi}} \boldsymbol{\chi}^{(i)} \right) - \langle \mathbf{u}_0 \rangle_{\mathcal{B}} \right] - \nabla_{\boldsymbol{\xi}} \cdot \left[D^{(i)} \left(\mathbf{I} + \nabla_{\boldsymbol{\xi}} \boldsymbol{\chi}^{(i)} \right) \right] = \mathbf{0} \quad \text{in } \mathcal{B}, \tag{41e}$$

$$-\mathbf{n} \cdot \left[D^{(i)} \left(\mathbf{I} + \nabla_{\boldsymbol{\xi}} \boldsymbol{\chi}^{(i)} \right) \right] = \mathbf{0} \quad \text{on } \Gamma. \tag{41f}$$

$(\alpha, \beta, \gamma, \delta) : (1/2, 1, -7/4, -7/4)$

$$\frac{\partial \left\langle c^{(1)} \right\rangle_{\mathcal{B}}}{\partial t} + \epsilon^{-\frac{1}{2}} \nabla_{\mathbf{x}} \cdot \left(\langle \mathbf{u} \rangle_{\mathcal{B}} \left\langle c^{(1)} \right\rangle_{\mathcal{B}} \right) - \nabla_{\mathbf{x}} \cdot \left(\mathbf{D}^{(1)} \cdot \nabla_{\mathbf{x}} \left\langle c^{(1)} \right\rangle_{\mathcal{B}} \right) = -\epsilon^{-\frac{7}{4}} \left\langle c^{(1)} \right\rangle_{\mathcal{B}} \left\langle c^{(2)} \right\rangle_{\mathcal{B}} + \epsilon^{-\frac{7}{4}} \left\langle c^{(3)} \right\rangle_{\mathcal{B}} \quad \text{in } \Omega, \tag{42a}$$

$$\frac{\partial \left\langle c^{(2)} \right\rangle_{\mathcal{B}}}{\partial t} + \epsilon^{-\frac{1}{2}} \nabla_{\mathbf{x}} \cdot \left(\langle \mathbf{u} \rangle_{\mathcal{B}} \left\langle c^{(2)} \right\rangle_{\mathcal{B}} \right) - \nabla_{\mathbf{x}} \cdot \left(\mathbf{D}^{(2)} \cdot \nabla_{\mathbf{x}} \left\langle c^{(2)} \right\rangle_{\mathcal{B}} \right)$$
$$= -\epsilon^{-\frac{7}{4}} \left\langle c^{(1)} \right\rangle_{\mathcal{B}} \left\langle c^{(2)} \right\rangle_{\mathcal{B}} + \epsilon^{-\frac{7}{4}} \left\langle c^{(3)} \right\rangle_{\mathcal{B}} \quad \text{in } \Omega, \tag{42b}$$

$$\frac{\partial \left\langle c^{(3)} \right\rangle_{\mathcal{B}}}{\partial t} + \epsilon^{-\frac{1}{2}} \nabla_{\mathbf{x}} \cdot \left(\langle \mathbf{u} \rangle_{\mathcal{B}} \left\langle c^{(3)} \right\rangle_{\mathcal{B}} \right) - \nabla_{\mathbf{x}} \cdot \left(\mathbf{D}^{(3)} \cdot \nabla_{\mathbf{x}} \left\langle c^{(3)} \right\rangle_{\mathcal{B}} \right)$$
$$= \epsilon^{-\frac{7}{4}} \left\langle c^{(1)} \right\rangle_{\mathcal{B}} \left\langle c^{(2)} \right\rangle_{\mathcal{B}} - \epsilon^{-\frac{7}{4}} \left\langle c^{(3)} \right\rangle_{\mathcal{B}}$$
$$- \frac{|\Gamma|}{|\mathcal{B}|} \left(\left\langle c^{(3)} \right\rangle_{\mathcal{B}} - \eta_1 \right) \quad \text{in } \Omega, \tag{42c}$$

$$\mathbf{D}^{(i)} = \left\langle D^{(i)} \left(\mathbf{I} + \nabla_{\boldsymbol{\xi}} \boldsymbol{\chi}^{(i)} \right) \right\rangle_{\mathcal{B}} - \epsilon^{\frac{1}{2}} \left\langle \boldsymbol{\chi}^{(i)} \otimes \mathbf{u}_0 \right\rangle_{\mathcal{B}}, \tag{42d}$$

$$\epsilon^{\frac{1}{2}} \left[\mathbf{u}_0 \cdot \left(\mathbf{I} + \nabla_{\boldsymbol{\xi}} \boldsymbol{\chi}^{(i)} \right) - \langle \mathbf{u}_0 \rangle_{\mathcal{B}} \right] - \nabla_{\boldsymbol{\xi}} \cdot \left[D^{(i)} \left(\mathbf{I} + \nabla_{\boldsymbol{\xi}} \boldsymbol{\chi}^{(i)} \right) \right] = \mathbf{0} \quad \text{in } \mathcal{B}, \tag{42e}$$

$$-\mathbf{n} \cdot \left[D^{(i)} \left(\mathbf{I} + \nabla_{\boldsymbol{\xi}} \boldsymbol{\chi}^{(i)} \right) \right] = \mathbf{0} \quad \text{on } \Gamma. \tag{42f}$$

References

Allarie, G., Raphael, A.L.: Homogenization of a convection-diffusion model with reaction in a porous medium. C. R. Math. **344**, 523–528 (2007)

Amaziane, B., Koebbe, J.: JHomogenizer: a computational tool for upscaling permeability for flow in heterogeneous porous media. Comput. Geosci. **10**, 343–359 (2006)

Arunachalam, H., Korneev, S., Battiato, I., Onori, S.: Multiscale modeling approach to determine effective lithium-ion transport properties. In: Proceedings of the 2017 IEEE American Control Conference, pp. 92–97 (2017)

Arunachalam, H., Onori, S., Battiato, I.: On veracity of macroscopic lithium-ion battery models. J. Electrochem. Soc. **162**, A1940–A1951 (2015)

Auriault, J.L., Adler, P.M.: Taylor dispersion in porous media: Analysis by multiple scale expansions. Adv. Water Resour. **18**, 217–226 (1995)

Bachmat, Y., Bear, J.: Macroscopic modeling of transport phenomena in porous media. 1: The continuum approach. Transport Porous Med. **1**, 213–240 (1986)

Bahmani, B., Yang, M., Nagarajan, A., Clarke, P.L., Soghrati, S., Abedi, R.: Automated homogenization-based fracture analysis: Effects of SVE size and boundary conditions. Comput. Methods Appl. Mech. Engrg. **345**, 701–727 (2019)

Battiato, I.: Multiscale models of flow and transport. In: Cushman, J.H., Tartakovsky, D.M. (eds.) The Handbook of Groundwater Engineering, Chapter 12, pp. 359–381. CRC Press, Boca Raton (2016)

Battiato, I., Tartakovsky, D.M.: Applicability regimes for macroscopic models of reactive transport in porous media. J. Contam. Hydrol. **120–121**, 18–26 (2011)

Battiato, I., Tartakovsky, D.M., Tartakovsky, A.M., Scheibe, T.: On breakdown of macroscopic models of mixing-controlled heterogeneous reactions in porous media. Adv. Water Resour. **32**, 1664–1673 (2009)

Battiato, I., Tartakovsky, D.M., Tartakovsky, A.M., Scheibe, T.D.: Hybrid models of reactive transport in porous and fractured media. Adv. Water Resour. **34**, 1140–1150 (2011)

Battiato, I., Ferrero, P.T., O'Malley, D., Miller, C.T., Takhar, P.S., Valdés-Parada, F.J., Wood, B.D.: Theory and applications of macroscale models in porous media. Transp. Porous Med. **130**, 5–76 (2019)

Boso, F., Battiato, I.: Homogenizability conditions for multicomponent reactive transport. Adv. Wat. Pol. **62**, 254–265 (2013)

Chilakapati, A., Ginn, T., Szecsody, J.: An analysis of complex reaction networks in groundwater modeling. Water Resour. Res. **34**, 1767–1780 (1998)

Falkowski, P.G.: Biogeochemical cycles. In: Levin, S.A. (ed.) Encyclopedia of Biodiversity, vol. 1, pp. 437–453. Academic Press, Cambridge (2001)

Floudas, C.A., Niziolek, A.M., Onel, O., Matthews, L.R.: Multi-scale systems engineering for energy and the environment: challenges and opportunities. AIChE J. **62**, 602–623 (2016)

Golfier, F., Wood, B.D., Orgogozo, L., Quintard, M., Buès, M.: Biofilms in porous media: development of macroscopic transport equations via volume averaging with closure for local mass equilibrium conditions. Adv. Water Resour. **32**, 463–485 (2009)
Gray, W.G., Miller, C.T.: Introduction to the Thermodynamically Constrained Averaging Theory for Porous Medium Systems. Springer, Switzerland (2014)
Ho-Le, K.: Finite element mesh generation methods: a review and classification. Comput. Aided Des. **20**, 27–38 (1988)
Hornung, U.: Homogenization and Porous Media. Springer, New York (1997)
Hu, X., Li, S., Peng, H.: A comparative study of equivalent circuit models for Li-ion batteries. J. Power Sources **198**, 359–367 (2012)
Iliev, O., Mikelić, A., Prill, T., Sherly, A.: Homogenization approach to the upscaling of a reactive flow through particulate filters with wall integrated catalyst. Adv. Water Resour. **146**, 103,779 (2020)
Knutson, C., Valocchi, A., Werth, C.: Comparison of continuum and pore-scale models of nutrient biodegradation under transverse mixing conditions. Adv. Water Resour. **30**, 1421–1431 (2007)
Korneev, S., Battiato, I.: Sequential homogenization of reactive transport in polydisperse porous media. Multiscale Model. Simul. **14**, 1301–1318 (2016)
Li, G., Monroe, C.W.: Multiscale lithium-battery modeling from materials to cells. Annu. Rev. Chem. Biomol. Eng. **11**, 277–310 (2020)
Li, L., Peters, C.A., Celia, M.A.: Upscaling geochemical reaction rates using pore-scale network modeling. Adv. Water Resour. **29**, 1351–1370 (2006)
Locker, M.: A hybrid-electric plane will get a boost from a french engine giant (2018). https://www.fastcompany.com/90246210/zunum-aero-and-safran-helicopters-team-up-for-hybrid-electric-planes
Maas, U., Pope, S.B.: Simplifying chemical kinetics: Intrinsic low-dimensional manifolds in composition space. Combust. Flame **88**, 239–264 (1992)
Mansell, R.S., Ma, L., Ahuja, L.R., Bloom, S.A.: Adaptive grid refinement in numerical models for water flow and chemical transport in soil: a review. Vadose Zone J. **1**, 222–238 (2002)
Miller, C.T., Valdés-Parada, F.J., Ostvar, S., Wood, B.D.: A priori parameter estimation for the thermodynamically constrained averaging theory species transport in a saturated porous medium. Transp. Porous Med. **122**, 611–632 (2018)
Morse, J.W., Arvidson, R.S.: The dissolution kinetics of major sedimentary carbonate minerals. Earth-Sci. Rev. **58**, 51–84 (2002)
Moura, S.J., Argomedo, F.B., Klein, R., Mirtabatabaei, A., Krstic, M.: Battery state estimation for a single particle model with electrolyte dynamics. IEEE T. Contr. Syst. T. **25**, 453–468 (2017)
Mutlay, İbrahim, Restrepo, A.: Complex reaction networks in high temperature hydrocarbon chemistry. Phys. Chem. Chem. Phys. **17**, 7972–7985 (2015)
Pantano, C.: Direct simulation of non-premixed flame extinction in a methane-air jet with reduced chemistry. J. Fluid Mech. **514**, 231–270 (2004)
Perez, H.E., Dey, S., Hu, X., Moura, S.J.: Optimal charging of Li-ion batteries via a single particle model with electrolyte and thermal dynamics. J. Electrochem. Soc. **164**, A1679–A1687 (2017)
Plett, G.L.: Extended Kalman filtering for battery management systems of LiPB-based HEV battery packs part 3. State and parameter estimation. J. Power Sources **134**, 277–292 (2004)
Ragsdale, S.W., Pierce, E.: Acetogenesis and the Wood-Ljungdahl pathway of CO_2 fixation. Biochim. Biophys. Acta **1784**, 1873–1898 (2008)
Rubinstein, J., Mauri, R.: Dispersion and convection in periodic porous media. SIAM J. Appl. Math. **46**, 1018–1023 (1986)
Smith, M., García, R.E., Horn, Q.C.: The effect of microstructure on the galvanostatic discharge of graphite anode electrodes in $LiCoO_2$-based rocking-chair rechargeable batteries. J. Electrochem. Soc **156**, A896–A904 (2009)
Soliman, S., Heiner, M.: A unique transformation from ordinary differential equations to reaction networks. PLoS ONE **5**, e14,284 (2010)
Tartakovsky, D.M., Neuman, S.P.: Transient flow in bounded randomly heterogeneous domains. Water Resour. Res. **34**, 1–12 (1998)
Vasilyeva, M., Mistry, A., Mukherjee, P.P.: Multiscale model reduction for pore-scale simulation of Li-ion batteries using GMsFEM. J. Comput. Appl. Math. **344**, 73–88 (2018)
Walpole, J., Papin, J.A., Peirce, S.M.: Multiscale computational models of complex biological systems. Annu. Rev. Biomed. Eng. **15**, 137–154 (2013)
Whitaker, S.: The Method of Volume Averaging. Kluwer Academic Publishers, Dordrecht (1999)
Wolfram Research, Inc., Mathematica, Version 12.1, Champaign, IL (2021).
Wood, B.D.: The role of scaling laws in upscaling. Adv. Water Resour. **32**, 723–736 (2009)

Wood, B.D., Quintard, M., Whitaker, S.: Calculation of effective diffusivities for biofilms and tissues. Biotechnol. Bioeng. **77**, 495–516 (2002)
Wu, H., Kimball, J.S., Mantua, N., Stanford, J.: Automated upscaling of river networks for macroscale hydrological modeling. Water Resour. Res. **47**, W03,517–1–W03,517–18 (2011)
Yousefzadeh, M., Battiato, I.: Physics-based hybrid method for multiscale transport in porous media. J. Comput. Phys. **344**, 320–338 (2017)
Zagnoni, M., Anderson, J., Cooper, J.M.: Hysteresis in multiphase microfluidics at a T-junction. Langmuir **26**, 9416–9422 (2010)