University of Nevada, Reno

Physical and Deep-Learning-Based Explorations of Microbe-Mediated Reactive Transport Processes in Porous Media Across Scales

A dissertation submitted in partial fulfillment of the requirements for the

degree of Doctor of Philosophy in Hydrology

By

Marc Berghouse

Advised by Dr. Rishi Parashar

December 2024

Acknowledgements

This thesis is the result of 4 and a half years of hard, passionate work. Not only hard work by myself, but hard work by each person involved in the steps along the way. First and foremost, I would like to thank my amazing advisor, Rishi Parashar. This thesis would simply not have been possible without his guidance, tireless work ethic, technical expertise, and moral support. From the first day I stepped into his office, he has done everything in his power to improve my research and general well-being. I would also like to thank my co-authors on various papers – Lazaro J. Perez, Filippo Miele, Verónica L. Morales, Timothy Schiebe, Andy Plymale, Ankur Bordoloi, George Bebis, and Alireza Tavakkoli, who have significantly improved the quality of this dissertation through revisions, suggestions, simulations, data analysis, and more. Similarly, I would like to thank my committee members – Veronica Morales, Timothy Schiebe, George Bebis, Eric Marchand, and Rishi Parashar (chair) for their years of support, insightful questions, and helpful criticisms.

My interest in reactive transport was mainly precipitated through my work as a research assistant for the SLAC SFA. Thus, I would also like to thank Zach Perzan, Kate Maher, and John Bargar for providing me with such an incredible opportunity and placing me on this amazing path. Specifically, Zach Perzan provided a huge amount of guidance in terms of lab work, field work, conceptual understandings of hydrology and reactive transport, and scientific data analysis.

I would next like to thank my parents, sister, and dog. My parents, Jack and Susan, and my sister, Brittany, have provided me with endless and incalculable support and love, and they are a joy to be around. In particular, this dissertation would have been nearly impossible without the financial support of my parents. My dog, Penny, is the shining light of my life that has allowed me to keep my sanity over the course of this PhD.

I would also like to thank my GPHS cohort and the GPHS program itself. My cohort provided me with a space to vent, discuss plans, and have fun. Without them the PhD experience would have been much less enjoyable. The GPHS program, and especially its director, Alexandra Lutz, have provided me with wonderful opportunities and have done much to help me achieve my research goals.

Finally, I would like to thank the grants that allowed me work as a paid PhD student. Specifically, I would like to acknowledge support from the U.S. Department of Energy (DOE) under award number DE- SC0019437, and career advancement support in the field of artificial intelligence based methods provided by NSF grant 2123481.

Table of Contents

Abstract

Reactive transport (RT) simulators are important tools often used by researchers to gain insights into subsurface processes. These multi-physics simulations attempt to represent many hydro-biogeochemical phenomena, but they often fall short in terms of computational speed and physical accuracy. This dissertation provides several tools and conceptual advancements that can be used to improve the speed and accuracy of RT simulations and further our understanding of their outputs. Specifically, this work investigates microbial motility in porous microfluidic devices, a comparison of particle tracking methods in porous media, and an investigation of biomass growth and chromium reduction in the hyporheic zone. Furthermore, this dissertation details the development and performance-testing of deep-learning-based tools for the extraction of motion statistics from videos of particles and the upscaling of RT simulations. Overall, new tools and insights are provided to help improve environmental management strategies, such as bioremediation of contaminated groundwater or improved understanding of nutrient cycles in water systems.

This dissertation advances our understanding of microbe-mediated reactive transport processes through a multi-scale approach that combines experimental observations, computational modeling, and innovative deep learning techniques. At the micro scale, experimental investigations reveal how different bacterial motility mechanisms respond to varying flow conditions, with peritrichous flagella enabling more resilient motility under higher flow rates compared to monotrichous flagella or pili. These findings provide crucial insights for developing more accurate models of microbial transport in subsurface environments.

To improve micro-scale investigations of bacterial transport, this dissertation gives a comparison of particle tracking (PT) methods and presents a novel deep-learning-based PT

method. The comparison between PT methods provides guidance for future researchers in terms of appropriate particle tracking linking algorithms to use for dispersive particles in porous media, conditions for desirable particle tracking experimental setups, and the limitations of particle tracking as it relates to analysis of bacterial transport. The novel deep learning method, DeepTrackStat (DTS), provides a framework for extracting motion statistics from particle tracking videos, addressing fundamental limitations in traditional tracking methods while significantly reducing computational demands. DTS shows especially strong performance for high-speed particles, giving it a clear spot for application within the pantheon of PT methods.

In addition to the work at the micro scale, this dissertation also provides improvements to microbe-mediated reactive transport modeling at the Darcy scale. The integration of novel physical approaches enables comprehensive investigation of coupled hydro-biogeochemical processes in the hyporheic zone, particularly focusing on the interactions between fluid flow, biomass development, and chromium reduction. Through extensive sensitivity analyses, this work reveals that while abiotic reduction dominates in high-electron-donor environments, biotic processes crucially influence the spatial distribution of reduction hotspots. Furthermore, the research demonstrates that speed-based biomass decay significantly impacts biomass growth only under specific conditions of high fluid velocity or weak biofilm cohesion, providing important constraints for environmental management strategies. Expanding on the Darcy-scale microbe-mediated reactive transport modeling, this dissertation presents STAMNet, a neural network for upscaling reactive transport simulations that enables efficient prediction of large-scale transport phenomena while preserving the essential dynamics observed at smaller scales. STAMNet has a simple MLP structure with a spatiotemporal attention mechanism (STAM) that uses cross-dimensional residual connections to improve both spatial and temporal feature extraction.

This work's multi-scale, multi-method approach provides a foundation for improving predictions of reactive transport in heterogeneous porous media while offering practical tools for environmental monitoring and remediation. The findings and methodologies presented here advance our ability to bridge scales in reactive transport modeling, from individual bacterial behavior to field-scale predictions, while the developed deep learning tools offer new possibilities for efficient analysis and upscaling of complex environmental processes. These contributions support more informed decision-making in environmental management and provide a framework for future investigations of coupled biological, chemical, and physical processes in porous media systems.

Chapter 1: Introduction

1.1 Background and Literature Review

1.1.1. Microbe-mediated Reactive Transport

Microbe-mediated reactive transport encompasses the interaction of biological, chemical, and physical processes within porous media, especially in subsurface environments. Microbial activities drive these processes primarily through their metabolic and motility behaviors, affecting both the transport of solutes [1-4] and the growth and dynamics of biofilms [5-8]. Some of the most studied aspects of microbe-mediated reactive transport include attachment/colonization [4, 9-12], chemotaxis [13-17], DLVO interactions [18, 19], biofilm formation [20] and breakup [21], microbial motility [22-31], nutrient cycling [32-34], biodegradation kinetics [35], weathering [36], and bioremediation [37-41]. To understand and predict the impacts of microbes on geochemical processes, an understanding of their transport is paramount. In other words, fundamental developments in transport modeling impact most of the models of biogeochemical interactions in the subsurface that are of interest to the research community. Furthermore, microbe-mediated reactive transport is a highly multidisciplinary field, meaning that developments in one area often lead to developments in different areas. For example, improvements in representations of biofilms at the pore scale can help inform how we simulate them in reactive transport models at the Darcy and field scales [42-44], or how we model transport of cancer drugs in human tissue [45].

As computers have become more powerful, allowing for more complex interactions to be simulated at larger scales, a large body of literature investigating the transport of microbes in porous media has recently developed [1-4, 9-13, 19, 21, 29-31, 42, 46-51]. Given the wide variety of literature, an almost equally wide variety of mathematical models have been developed to simulate various aspects of microbial transport. A brief introduction to the mathematics of

microbe-mediated reactive transport is given in sections 1.1.1.1 and 1.1.1.2 of this dissertation. Studies in this domain range from the micro-scale [52, 53] to the field scale [10, 54]. Micro and pore-scale experiments generally involve microfluidic devices, which are tiny (μm to mm range) transparent sections of porous media which contain small channels that allow fluid to be injected into the model [53]. Darcy and field scale experiments often use column experiments [10], acetate injection experiments [55], and reactive transport simulations [56, 57] to model transport.

Microbe-mediated reactive transport is a highly active area of research for two primary reasons – the complexity of the interactions means there are a wide variety of both fundamental and applied discoveries still to be made, and there is a clear human need for the development of better tools and more comprehensive theories that describe microbe-mediated reactions in the subsurface. This interaction is essential in environmental settings such as the hyporheic zone, where groundwater and surface water meet, creating a dynamic space for complex biogeochemical reactions [58]. For instance, microbes capable of reducing Cr(VI) to less toxic Cr(III) significantly affect the transport of chromium through groundwater systems [59, 60]. Understanding these processes is critical for fields such as bioremediation [61], water quality management [62], and nutrient cycling [63]. In addition to an increase in our modeling capabilities of bioremediation and nutrient cycling, two things that are of high importance for human governments around the world, many general theories and modeling methods of microbe-mediated reactive transport can be extended to the human body [64-68]. Some of the most important recent research in this domain includes studies that examine the dispersion of bacteria around cancerous tumors [69]. Specifically, it has been found that some bacteria, such as *Salmonella*, are attracted to chemicals emitted by necrotic regions of tumors and will try to burrow towards the middle of the tumor [70, 71]. Ultimately, the depth of our understanding of microbe-

mediated reactive transport processes, and our ability to simulate them, has a direct impact on our ability to effectively manage our environment and develop life-saving medical treatments.

1.1.1.1 **Reactive Transport Simulations**

General Overview

Reactive transport (RT) is the general term used to describe coupling of fluid flow, solute transport, and chemical reactions in porous media. Microbe-mediated reactive transport can be considered a subset of reactive transport wherein the primary point of investigation is to understand how microbes impact the environment. In natural environments, RT plays a pivotal role in determining the fate and distribution of nutrients, contaminants, and other chemical species. Models of RT, often referred to as reactive transport simulators, integrate hydrodynamic equations with chemical reaction networks, allowing researchers to simulate the evolution of species concentrations and mineral phases over time. RT simulations vary widely in terms of the hydro-biogeochemical and physical phenomena that are represented, but they all follow the same general structure and workflow.

In RT simulators, flow and transport are coupled. For a Darcy-scale (continuum) simulation, Richards' equation is generally used to represent flow, and the advection-dispersion equation (2) is used to represent transport [72]. Richards' equation can be thought of as making Darcy's law, which is a general representation of flow though porous media, a function of saturation. Darcy's law is given as:

(1) $q = -K\nabla h$

where q is the specific discharge, K is the hydraulic conductivity tensor (k/μ), and ∇h is the hydraulic head gradient. Making K a function of water saturation (θ) and splitting up the hydraulic

head into pressure head (h) and elevation head (z) terms then substituting this sum into Darcy's law gives

(2) q = $-K(\theta)\nabla(h+z) = -K(\theta)(\nabla h + \nabla z)$

The continuity equation, which represents the conservation of mass in the domain, is given as

(3) $\frac{\partial\theta}{\partial t} + \nabla \cdot q = 0$

Substituting q from (2) into (3) then gives

(4) $\frac{\partial\theta}{\partial t} + \nabla \cdot [K(\theta)(\nabla h + \nabla z)] = 0$

Expanding the divergence term and introducing the specific moisture capacity $C(h) = -\frac{d\theta}{dh}$ gives

(5) $C(h)\frac{\partial h}{\partial t} = \nabla \cdot (K(h)\nabla h) + \frac{\partial}{\partial z}[K(h)]$

This is commonly referred to as the mixed-form Richards equation, which is valid for both saturated and unsaturated conditions [72]. The Richards equation can also be represented in terms of soil water content (θ) by introducing the soil water diffusivity term $D(\theta) = -\frac{K(\theta)dh}{d\theta}$ which gives

(6) $\frac{\partial\theta}{\partial t} = \nabla \cdot (D(\theta)\nabla\theta) + \frac{\partial \mathrm{K}(\theta)}{\partial z}$

This equation allows the simulator to calculate the fluid velocity field throughout the domain by solving for $\mathrm{K}(\theta)$ and substituting the value into (2). Equation (6) is only valid for unsaturated conditions (since in saturated conditions $\frac{\partial\theta}{\partial t} = 0$), meaning equation (5) is the more general representation of Richards' equation. The ADE, on the other hand, governs the transport of solutes, and is often represented in its simplest 3D form as

(7) $\frac{\partial C}{\partial t} = -\nabla \cdot (vC) + \nabla \cdot (D\nabla C)$

Here, C is the solute concentration, t is time, v is the pore water velocity (derived from Darcy's law), and D is the hydrodynamic dispersion coefficient tensor. The 1D version of the ADE is given by

$$(8)\quad R\frac{\partial c}{\partial t} = D\frac{\partial^2 c}{\partial x^2} - v\frac{\partial c}{\partial x}$$

Here R represents retardation, which is a phenomena that describes the interactions between solute particles and the solid phase (AKA sorption). The retardation factor R is typically determined through laboratory batch or column experiments where both conservative and reactive tracers are used to measure the delay in breakthrough curves [73]. For linear sorption isotherms, R can be calculated as $R = 1 + \frac{k_d \rho_b}{n}$, where ρ_b is bulk density, n is porosity, and k_d is the soil partition coefficient which equals $f_{OC} k_{OC}$ where f_{OC} is the organic carbon fraction and k_{OC} is the carbon-water partition coefficient [74]. When dealing with nonlinear sorption behavior, such as Freundlich or Langmuir isotherms, the retardation factor becomes concentration-dependent and more complex to determine [75]. The presence of retardation effectively slows down the apparent velocity of the contaminant plume relative to the groundwater velocity by a factor of R. For example, if R = 2, the contaminant on average moves at half the speed of the groundwater, meaning the breakthrough curve is twice as long. Laboratory batch tests to determine k_d values often involve mixing known quantities of solution and solid material, allowing equilibration, and then measuring final concentrations to construct sorption isotherms [76].

The iterative solution of reactive transport equations involves a complex process of stepping through both time and space to solve equations (2), (5), and (7) (among others). As outlined in the pseudocode in Appendix A1, the process begins with initialization, setting up the grid dimensions, time steps, and initial conditions for concentration, water content, and velocity. The main simulation then proceeds through a series of time steps, each involving several key

computations across the entire spatial domain. For each time step, the algorithm first solves the Richards equation to update the water content at every grid point, followed by updating the hydraulic conductivity. Next, it computes the Darcy velocities based on the new water content. The core of the simulation involves solving the Advection-Dispersion Equation (ADE) for solute concentration, which accounts for advection, dispersion, and reactions in however many spatial dimensions have been prescribed to the domain (up to a max of 3). This step, as shown in A1, requires nested loops over all spatial dimensions, highlighting the computational intensity of the process. After updating concentrations, boundary conditions are applied, and results are output at specified intervals. In more complex reactive transport simulations, the concentrations of many species must be computed and the reaction functions have more terms, further increasing computational complexity. These reactions, in turn, can modify physical properties like porosity and permeability, so most reactive transport simulators will also update the permeability field with each temporal iteration.

The pseudocode in A1 demonstrates why speeding up reactive transport simulations is a high priority in the field. The nested loops for spatial dimensions (x, y, and z) within the main time-stepping loop result in a large number of calculations for each time step. For real-world applications with fine spatial and temporal resolutions, this can result in long computation times. This computational intensity is further increase when many species with reactions that depend on many other species. Consequently, researchers and developers in the field are continually seeking ways to optimize these simulations, such as through parallelization [77], adaptive mesh refinement [78], reduced order modeling [79], improved numerical schemes [80], adaptive time stepping [81], and deep learning [82]. These optimizations have allowed, and will hopefully

continue to allow, researchers to tackle more complex problems and provide timely insights for critical environmental and engineering decisions.

These simulators also incorporate chemical databases containing thermodynamic and kinetic data for various reactions. Geochemical calculations are performed to determine speciation, mineral saturation states, and redox conditions [83, 84]. Different coupling schemes, like sequential non-iterative approaches or global implicit methods, are employed to solve the transport and reaction calculations efficiently [85]. The ability to handle multiphase flow, temperature effects, and porosity-permeability feedback are also crucial features in many simulators [86]. These allow for modeling more complex scenarios involving multiple fluid phases, temperature-dependent reactions, and evolving medium properties. Some of the most popular reactive transport simulation frameworks include PFLOTRAN [87], STOMP [88], CrunchFlow [89], TOUGHREACT [90], PHREEQC [91], MIN3P [92], and OpenGeoSys [93]. Furthermore, especially advanced studies of phenomena not currently represented in these popular models are often developed through coupling with more advanced computational fluid dynamics models such as OpenFOAM [94], or the scale of the complexity of the simulation is reduced, allowing the simulation to be programmed from scratch to investigate a single/few RT phenomena.

Dispersion in Reactive Transport

One of the many complexities of reactive transport simulations involves the representation of dispersion. Dispersion generally describes the spreading of solutes due to the combined effects of molecular diffusion and non-uniform flow fields that cause variabilities in in the velocities of solute particles [95, 96]. Dispersion can be enhanced through phenomena such as particle motility [97], density driven flow [98], soil heterogeneity and anisotropy [99], pore-scale mixing [100], and any other phenomena that can increase the variability of solute pathlines, and it can be reduced

through high pore-network connectivity [101], attachment/adsorption [102], and low Peclet numbers [103]. The Peclet number, which represents the ratio of advective to diffusive transport, plays a crucial role in characterizing both microscopic and macroscopic dispersion behavior [103, 104]. A lower Peclet number means diffusion is the dominant form of transport, and a larger number means advection is the dominant form of transport. Thus, a greater Peclet number generally results in greater dispersion.

Dispersion (both micro and macro) can be quantified in terms of the direction of transport through the phrases “longitudinal” (in the direction of flow), “transverse” (perpendicular to direction of flow), and “vertical”. In micro-scale environments, the relative amounts of each dispersion coefficient are highly dependent on experimental conditions, such as porous geometry and solute particle properties. In macro-scale environments, dispersion is primarily dependent on flow and soil conditions and mainly occurs in the longitudinal direction, although these generalizations are subject to a number of other hydro-biogechemical conditions.

At the micro scale, dispersion is primarily governed by the variations in streamlines followed by particles, which are caused by the geometry of the porous medium or properties of the particles (such as motility and attachment). The flow field is generally solved via the Navier-Stokes equations, and particle transport can be solved in a variety of ways. In cases where the exact positions of the particles over time are known, such as in chapter 2 of this dissertation on bacterial transport in microfluidic devices, we can use the second centered moment of the positions of the bacteria to quantify dispersion. When the exact positions of the particles are not known, but the influent and effluent concentrations are known to allow construction of a breakthrough curve, then we can use the ADE to approximate dispersion by fitting the breakthrough curve to the analytical solution [105, 106]. However, the ADE is often an inadequate

representation of solute transport, as it only provides an analytical solution in Fickian regimes. For non-Fickian transport, models such as the fractional ADE [107] and the continuous time random walk [108] can achieve higher accuracy of breakthrough curve predictions and effective dispersion coefficients.

As briefly discussed above, one common approach to quantify dispersion at the micro scale is through the second centered moment of the average concentration [95, 109]. The second centered moment, denoted as $\sigma^2(t)$, is defined as: $\sigma^2(t) = \int [x - \mu(t)]^2 c(x,t) dx$, where $c(x,t)$ is the concentration of the solute at position x and time t, and $\mu(t)$ is the first moment (mean position) of the concentration distribution at time t. In chapter 2, we calculate the average spatial variance of bacteria over time evolving from a point-like injection, akin to the transport Green function defined as $D^e(t) = \frac{1}{w\phi} \int_0^w dy' D^e(t, y')$ [110, 111]. This function describes the calculation of the effective dispersion coefficient across the *y*-dimension of the domain. It can be solved discretely by calculating the asymptotic rate of spreading for each bin comprising the *y* dimension (assuming advection is in the *x* direction). The temporal evolution of the second centered moment provides valuable insights into the dispersive behavior of the solute plume. In the case of Fickian dispersion, where the dispersion coefficient is constant and the advection-dispersion equation (ADE) is valid, the second centered moment grows linearly with time [112, 113]. The slope of this linear relationship is directly proportional to the macroscopic dispersion coefficient. However, in heterogeneous media, the assumption of Fickian dispersion may not hold, particularly at early times or in the presence of strong heterogeneity [108, 114]. In such cases, anomalous (non-Fickian) transport may occur, characterized by non-linear growth of the second centered moment with time [107, 115].

Macroscopic dispersion, also commonly referred to as hydrodynamic dispersion, generally emerges as a result of fluid and solute velocity variations at larger scales, such as the Darcy and field scales [116, 117]. The hydrodynamic dispersion coefficients (in the longitudinal direction) are calculated as $D_x = D_m + va_L$, where D_m is molecular diffusion, v is the seepage velocity, and a_L is the longitudinal dispersivity (units of length). The mechanical dispersion term arises from the spatial averaging of pore-scale velocity fluctuations, while molecular diffusion is an intrinsic property of the solute-fluid system [116, 118]. Thus, in reactive transport simulations, the representation is determined by the scale. For micro-scale and pore-scale simulations, dispersion is often defined as the average spatial variance of particles over time. For Darcy-scale and field-scale simulations, dispersion is often defined with this simple function.

In addition to the calculation of dispersion coefficients, which allow for calculation of concentrations in time and space via the ADE or fADE, the temporal evolution of dispersion also provides a critical lens through which to view transport. Over time, dispersion in heterogeneous media often exhibits distinct pre-asymptotic and asymptotic regimes [96, 117]. In the pre-asymptotic regime, the dispersive behavior is strongly influenced by the local velocity fluctuations and the correlation structure of the heterogeneous medium [119]. In the asymptotic regime, the dispersive behavior is primarily governed by larger-scale heterogeneities such as geological units (and other spatially-large soil properties), flow inputs, and particle behavior (motility, attachment, etc.).

In general, both pre-asymptotic and asymptotic regimes can be defined as subdiffusive, diffusive, superdiffusive, or ballistic. These regimes are often defined through the mean-square displacement (MSD). Specifically, in the context of diffusion, the regime is considered subdiffusive if $\frac{d(MSD)}{dt} < 1$, diffusive if $\frac{d(MSD)}{dt} = 1$, superdiffusive if $\frac{d(MSD)}{dt} > 1$, and ballistic if $\frac{d(MSD)}{dt} \geq 2$.

Although the MSD can be formulated in a variety of different ways depending on the context of use, in the context of diffusion (no advection), the second centered moment of an ensemble of particles is exactly equal to the MSD. Thus, for diffusive regimes, the diffusion coefficient can be calculated by setting the slope of the MSD equal to $2dDt$ [120], where d is the number of dimensions, D is the diffusion coefficient, and t is time. This can be solved by taking the derivative on both sides, which gives $\frac{d(MSD)}{dt}\frac{1}{2d} = D$. Although it is important to understand both the pre-asymptotic and asymptotic behavior, which can reveal phenomena such as confinement, the overall transport (effective diffusion) is generally defined by the asymptotic behavior [119, 109].

During the early stages of the pre-asymptotic regime, the solute plume may exhibit ballistic or superdiffusive transport, characterized by a rapid growth of $\sigma^2(t)$ [115, 120]. This behavior arises from the coherent motion of solute particles along preferential flow paths, leading to enhanced spreading [121]. As the plume evolves and samples more of the heterogeneous velocity field, the pre-asymptotic regime may transition to a subdiffusive behavior, where the growth of $\sigma^2(t)$ slows down [108, 114]. This transition from high to low MSD slope is often emblematic of confined transport conditions. As particles spread out over time in low-porosity media, they are likely to encounter dead-end pores and weakly-connected paths, which causes a subsequent decrease in the MSD slope due to what is essentially a decrease in the degrees of freedom of particle motion [122-124].

1.1.1.2 Microbial Motility and Bacterial Transport

Types of Microbial Motility

Microbial motility is a field that describes the movements patterns of all microbes (bacteria, viruses, protists, algae, fungi, etc.), although the term is generally used to describe the motility of

bacteria. Researchers have identified five general modes of bacterial motility – swimming, twitching, gliding, sliding, and swarming [125].

Swimming motility represents the most well-characterized form of bacterial locomotion, primarily facilitated by rotating flagella that act as helical propellers [27, 126, 127], allowing bacteria to achieve speeds of up to 500 $\mu m \cdot s^{-1}$ in liquid environments [128]. Flagella come in many different shapes and sizes, but can generally be classified as monotrichous (one flagella), amphitrichous (flagella on both ends of the bacterium), lophotrichous (many flagella on one end), or peritrichous (many flagella all over the bacterium). When monotrichous flagella are located at one of the ends of the major axis of the bacteria, it may be referred to as a polar flagellum [129]. Swimming bacteria can respond to various environmental stimuli through chemotaxis, modulating their flagellar rotation to navigate towards attractants or away from repellents [130, 131]. The ability to swim provides bacteria with considerable advantages in surface colonization [12], biofilm formation [132], and establishing infections in host organisms [133].

Twitching motility constitutes a distinct form of bacterial translocation dependent on Type IV pili (T4P), protein filaments that undergo cycles of extension, substrate attachment, and retraction [134]. The molecular machinery responsible for twitching comprises a complex membrane-spanning apparatus that includes ATPase motors, which power pili extension and retraction through ATP hydrolysis [27]. This form of motility enables bacterial cells to move across solid surfaces at speeds of approximately 0.1-1 μm/s [135], significantly slower than swimming but highly effective for surface colonization. The coordinated expression and assembly of pili is regulated by multiple environmental signals and integrated into broader cellular processes including biofilm formation and virulence. Twitching motility plays a crucial role in host colonization by pathogenic bacteria, particularly in the early stages of infection when bacteria

must traverse host tissue surfaces [125, 136]. The biomechanical forces generated during pili retraction can exceed the nanonewton range [137], making twitching a remarkably powerful, although generally slow, form of bacterial movement.

Gliding motility represents a sophisticated form of bacterial locomotion that enables directed movement across solid surfaces without conventional motility appendages. This type of motility is particularly prevalent among phylogenetically diverse bacteria including members of the Bacteroidetes phylum, Myxococcales order, filamentous cyanobacteria, and mycoplasmas, with each group potentially employing distinct molecular mechanisms [138, 139]. Building upon this foundation, contemporary research has unveiled the intricate biophysical principles governing bacterial gliding at the single-cell level. Advanced microscopy and rheological studies have demonstrated that successful translocation depends on the mechanical properties of both the bacterial cell and its environment [140, 141]. The substrate's physical characteristics prove particularly crucial, as they determine whether movement is primarily driven by cellular deformation-induced matrix flow caused by slime production or by complex interfacial dynamics at the bacterial leading edge. This mechanistic understanding of gliding motility has broader implications for bacterial behaviors including predation, development, and biofilm formation, while also providing insights into how bacteria navigate diverse environmental conditions.

Sliding motility describes a passive form of bacterial surface translocation that occurs without the use of dedicated motility structures, where the primary driving force comes from the physical pushing of dividing cells against their neighbors [125]. This process manifests differently across bacterial species, which can be classified into distinct groups based on their sliding mechanisms. In almost all cases, however, the physical basis of sliding involves the reduction of surface tension through various secreted compounds including biosurfactants (e.g., surfactin,

serrawettin), exopolysaccharides, hydrophobic proteins, and glycopeptides, which collectively modify substrate surface properties [125, 142]. This form of passive motility, driven by the expansive force of cellular reproduction, proves particularly advantageous for bacterial colony expansion in environments where active motility mechanisms might be energetically unfavorable, and its regulation involves complex cellular processes that integrate environmental sensing with the coordinated production of surface-modifying compounds.

Swarming motility represents a sophisticated form of social bacterial movement characterized by the rapid and coordinated migration of dense populations across surfaces [27, 125, 143]. This process requires flagella and involves a complex differentiation process whereby vegetative cells transform into elongated, hyperflagellated swarmer cells capable of moving in multicellular groups. The initiation of swarming behavior typically involves sensing appropriate environmental conditions, including surface contact, nutrient availability, and cell density, followed by the coordinated expression of genes involved in swarmer cell differentiation and movement [144]. Critical to successful swarming is the production of surfactants that reduce surface tension and facilitate the expansion of the bacterial population across the substrate. Swarming motility has important implications for bacterial pathogenesis and antimicrobial resistance, as swarming cells often display elevated resistance to various antibiotics and host defense mechanisms [145].

Bacterial Transport

The transport of bacteria through fluid environments represents a complex interplay of physical, chemical, and biological factors operating across multiple spatial and temporal scales. At the microscale (1-10 μm), bacterial transport is governed by the mechanics of cell motility [27, 31, 126], Brownian motion [27, 146, 147], chemotaxis [13, 15, 16] and cell-surface interactions [18],

while larger scale phenomena (>100 μm) such as advection, dispersion, microbial growth/death, attachment rates, and bulk fluid dynamics dominate transport at the macro level [4, 46, 148, 149]. The relative contribution of these mechanisms varies significantly depending on environmental conditions, with the Reynolds number ($Re \approx 10^{-5} – 10^{-3}$ for swimming bacteria) and Péclet number serving as critical dimensionless parameters that characterize the balance between advective and diffusive transport processes. In natural systems, bacterial transport is further complicated by cellular responses to chemical gradients (chemotaxis), oxygen availability (aerotaxis), and various other environmental stimuli that can modify movement patterns and influence net displacement.

The presence of active motility mechanisms also substantially impacts bacterial transport dynamics and population distributions. Flagellar swimming, which generates speeds of 1-500 μm/s [128], enables enhanced diffusion coefficients of motile bacteria that typically range from 10^{-6} to 10^{-5} cm²/s, which is about 1000x greater than the typical diffusion coefficients for non-motile cells [150]. Surface-associated motility modes demonstrate distinct transport characteristics: twitching (0.1-1 μm/s) [135] facilitates surface colonization through pili-mediated movement, gliding (1-10 μm/s)[151] enables translocation along surfaces without appendages, and swarming (2-10 μm/s)[152] promotes rapid population-scale movement across surfaces. These various motility mechanisms exhibit distinct responses to fluid flow, with swimming cells demonstrating rheotaxis (orientation with respect to flow) and complex trajectories arising from the interplay between self-propulsion and ambient flow fields [27, 125]. For example, swimming bacteria often exhibit upstream migration in shear flows and accumulation near surfaces due to hydrodynamic interactions [153].

In porous media, bacterial transport is governed by multiple retention and mobilization mechanisms operating simultaneously. Physical straining occurs when cells become trapped in

pore throats smaller than a critical size (which primarily depends on the ratio of average pore throat size to bacteria size), while physiochemical filtration processes include van der Waals forces, electrostatic interactions, and hydrophobic effects that promote attachment to grain surfaces [4, 18, 19]. Fluid shear stress also impacts bacterial transport, resulting in phenomena such as shear trapping [147] and colonization of surfaces [12]. The efficiency of these shear and retention-based mechanisms depend strongly on cell size, shape, and surface properties (e.g., lipopolysaccharides, extracellular polymeric substances), as well as the pore space geometry and chemical conditions (pH, ionic strength) of the system.

Chemotaxis significantly modulates bacterial transport behavior through the generation of directed movement along chemical gradients [15-17]. This process operates through a well-characterized signal transduction pathway involving methylation-dependent adaptation, which allows bacteria to respond to relative changes in chemoeffector concentrations over several orders of magnitude [154]. The impact of chemotaxis on transport is often quantified using the chemotactic sensitivity coefficient χ_0, which can range from 10^{-5} to 10^{-4} cm²/s depending on the species and chemoattractant [155, 156]. In flowing systems, the interplay between chemotaxis and fluid flow creates complex transport phenomena, including the formation of bacterial plumes and bands that can significantly enhance dispersion rates. These effects are particularly pronounced in the presence of sharp chemical gradients, where the chemotactic drift speeds (typically 0.1-2.5 μm/s) can become comparable to or exceed the fluid velocity [157]. In addition, chemotaxis can dramatically influence bacterial transport through porous media by promoting attachment or detachment from surfaces based on local chemical conditions.

The impact of bacterial transport extends across numerous practical applications in both engineered and natural systems. In environmental engineering, bacterial transport models inform

bioremediation strategies and the design of water treatment systems. These applications often involve complex geometries and heterogeneous environments where traditional continuum models may break down, necessitating multiscale approaches that bridge molecular, cellular, and population-level dynamics. In addition to the research discussed above on transport in porous media and chemotaxis, recent studies have highlighted sophisticated behaviors such as surface sensing [158], collective motion [159, 160], and adaptation to mechanical and chemical constraints [161-163]. These insights are particularly relevant for understanding biofilm formation, infection processes, and microbial ecosystem dynamics in natural environments such as soil, groundwater, and marine systems, where bacterial transport plays a crucial role in nutrient cycling and ecosystem function.

Mathematics of Microbial Motility and Bacterial Trasnport

The mathematical description of bacterial motility spans multiple scales, from individual cell dynamics to Darcy-scale behavior. Although some mathematical descriptions of twitching exist, this dissertation section only provides a brief overview of some of the most relevant swimming-based models. At the microscale, the Langevin equation is often used to describe the movement of Lagrangian particles [164, 165].

(9) $\frac{dx(t)}{dt} = v[x(t)] + \xi(t)\sqrt{2D}$

where v is the velocity of the particle at position x during time t, $\xi(t)$ is a Gaussian white noise vector, and D is the molecular diffusion coefficient. This equation captures the inherent Brownian motion of bacteria, although it is a rough approximation of overall motility. To better represent motility, we can adjust the Langevin equation to represent run and tumble motion. Many Langevin-type representations of motility have been designed by researchers over the years to

better approximate bacterial motion. One example represents the Langevin equation as a potential differential [166]

(10) $\dot{x} = -\frac{dU}{dx} + \eta_x(t)$

where $\dot{x}$ is rate of change of position $\frac{dx}{dt}$, $\eta_x(t)$ is the gaussian noise function, and U is a phenomenological potential function described as

(11) $U(x) = U_0 - \rho\left[x - \frac{\gamma}{\delta}\cosh(\delta x)\right]$

where U_0 is an adjustable constant and the tuning parameters ρ, γ, and δ are used to define the run and tumble states of the bacteria according to the relationship $\rho^2\delta\sqrt{1+\gamma^2} = C$. C is a constant and $\rho(\delta,\gamma) = 1$ during tumbling and $\rho(\delta,\gamma) > 1$ during runs. More detailed run and tumble parameters are described through a control parameter $\beta(\delta,\gamma)$. The inclusion of β and steady speed v_s in the Langevin equation for speed leads to

(12) $\dot{v} = -\lambda_0(\beta)[v - v_s(\beta)] + \eta_v(t)$

where $\lambda_0(\beta)$ is the inverse of the characteristic run and tumble times, τ. The solution for equation (13) for the condition of $v(t_0) = v_0$ is the given as

(13) $v(\beta,t) = v_s(\beta) + [v_0 - v_s(\beta)]G_v^{(\beta)}(t,t_0) + \int_{t_0}^{t}\eta_v(s)G_v^{(\beta)}(t,s)ds$

where $G_v^{(\beta)}$ is the green function for run and tumble motion defined as

(14) $G_v^{(\beta)}(t,t') = e^{-\frac{|t-t'|}{\tau(\beta)}}$

In porous media, the Langevin approach is refined to account for a variety of dispersion-related phenomena [31]. A useful description of the 1D streamwise motion of non-motile bacteria known as the continuous time random walk approach (CTRW) can be formulated discretely as

(15) $x_{n+1} = x_n + \frac{\Delta s}{\chi}, \quad t_{n+1} = t_n + \frac{\Delta s}{v_n}$

where the transition length Δs represents the incremental distance that a bacterial cell travels along its actual path through the porous medium, and the advective tortuosity χ represents how much the trajectory of a particle deviates from a straight line and is defined as the total distance traveled by a particle divided by its displacement. To account for the run and tumble motility of bacteria, this approach is further amended to include trapping time τ.

(16) $x_{n+1} = x_n + \frac{\Delta s}{\chi}, \quad t_{n+1} = t_n + \frac{\Delta s}{v_n} + \tau \frac{\Delta s}{v_n}$

The principle of this model is that bacteria randomly attach to grain surfaces and will remain on the surface for a certain amount of time before moving (either due to motility or external forces). The modeling of trapping times is still an active area of research, but may be exponentially distributed as $\psi(t) = \frac{\exp\left(-\frac{t}{\tau_c}\right)}{\tau_c}$.

For chemotactic populations, the Keller-Segel model [167] offers a continuum description that has become fundamental in understanding bacterial pattern formation and biofilm development. The time-dependent density of cells $\rho(\boldsymbol{r}, t)$, assuming no changes due to cell growth and death, can be modeled as [168]

(17) $\partial_t \rho(\boldsymbol{r}, t) + \nabla \cdot \boldsymbol{j}(\boldsymbol{r}, t) = 0$

where the flux of cells is $\boldsymbol{j}(\boldsymbol{r}, t) = \boldsymbol{j}(\boldsymbol{r}, t)_{diffusion} + \boldsymbol{j}(\boldsymbol{r}, t)_{chemotaxis}$. $\boldsymbol{j}(\boldsymbol{r}, t)_{diffusion}$ is defined as $-D_\rho \nabla \rho$ where D_ρ is the macroscopic diffusion coefficient, and $\boldsymbol{j}(\boldsymbol{r}, t)_{chemotaxis}$ is defined as $\rho v_{chemotaxis}$. Here, $v_{chemotaxis}$ is the velocity of bacteria due to chemotaxis, also known as the chemotactic drift velocity, which is defined as $v_{chemotaxis} = \chi(c)\nabla c$ where c is concentration and $\chi(c)$ is the chemotactic sensitivity. Using these relationships to develop a reaction-diffusion equation results in

(18) $\partial_t \rho = D_\rho \nabla^2 \rho - \nabla[\rho \chi(c) \nabla c\,]$

(19) $\partial_t c = D_c \nabla^2 c + h\rho - kc$

where $h\rho$ and kc account for production and consumption of the chemotactic species. These two equations are thus referred to as the Keller-Segal model.

At the Darcy scale, bacterial transport is commonly described using modified forms of the Advection-Dispersion Equation (ADE) that incorporate biological processes and attachment/detachment kinetics. The 1D ADE for microbial transport in a saturated, homogeneous porous medium can be formulated in simple terms as [46]

(20) $$R\frac{\partial C}{\partial t}+\frac{\rho_b}{\varepsilon}\frac{\partial S}{\partial t}=D\frac{\partial^2 C}{\partial x^2}-v\frac{\partial C}{\partial x}$$

where S is the attached microbe concentration, D is the hydrodynamic dispersion coefficient, v is microbial velocity, R is the retardation factor, ρ_b is the bulk density, and ε is the bed porosity. Kinetic attachment and detachment rates are used to represent the $\frac{\rho_b}{\varepsilon}\frac{\partial S}{\partial t}$ term in the equation

(21) $$\frac{\rho_b}{\varepsilon}\frac{\partial S}{\partial t}=k_{att}C-\frac{\rho_b}{\varepsilon}k_{det}S$$

where $k_{att}=\frac{3(1-\varepsilon)v}{2d_c}\eta_0\alpha$. This formula comes from the domain of colloid filtration theory. Here, d_c is the average grain size, η_0 is known as the single-collector contact efficiency, and α is the attachment collision efficiency. Equation (20) is a relatively accurate representation of the subsurface transport of bacteria in fully saturated conditions. However, transport often takes place in unsaturated conditions and thus depends on air-water exchange. A more complex 1D formation of microbial transport in porous media for multiphase flow can be given as [148]

(22) $$\frac{\partial\theta_c C}{\partial t}+\rho_b\frac{\partial S}{\partial t}+\frac{\partial A_{aw}\Gamma}{\partial t}=\frac{\partial}{\partial z}\left(\theta_c D\frac{\partial C}{\partial z}\right)-\frac{\partial q_c C}{\partial z}+B_w$$

Where C is the microbe concentration in the aqueous phase, θ_c is the volumetric water content available for microbes, S is the microbe concentration in the solid phase, A_{aw} is the interfacial area per unit volume, Γ is microbe concentration in the gaseous phase, q_c is the volumetric water

flux density for colloids, and B_w is a source/sink term that represents microbial growth and death in the aqueous phase.

The mathematical frameworks presented here, ranging from microscale Langevin descriptions of individual bacterial motion to Darcy-scale transport equations, demonstrate the complex multiscale nature of bacterial transport in porous media. While these models have proven valuable in describing various aspects of bacterial movement and transport, significant challenges remain in bridging scales and incorporating the full complexity of biological responses, particularly in heterogeneous environments. Future work will likely focus on developing more sophisticated hybrid approaches that can better couple individual bacterial behavior with population-scale phenomena while maintaining computational tractability. Additionally, the incorporation of emerging understanding about bacterial surface interactions, biofilm formation, and collective behavior will continue to refine these mathematical descriptions.

1.1.1.3 Growth and Dynamics of Biofilms

Biofilms are communities of microorganisms that attach to surfaces and surround themselves with a self-produced sticky matrix substance. These microbial communities are ubiquitous in nature and can be found virtually anywhere moisture and nutrients are available, from tooth surfaces [169] to deep ocean hydrothermal vents [170]. The development of biofilms follows distinct stages: initial surface attachment, microcolony formation, maturation, and dispersal [171]. Surface attachment is mediated by specific adhesins and regulated by intracellular signaling molecules, particularly cyclic-di-GMP, which serves as a master regulator in the transition from planktonic to sessile states [172]. Upon attachment, bacteria initiate the production of extracellular polymeric substances (EPS), creating a complex matrix that comprises polysaccharides, proteins,

extracellular DNA, and lipids [173]. This matrix may account for up to 90% of the mass of the biomass, with the mass of the actual cells comprising the other 10% [173].

The biofilm matrix serves multiple critical functions beyond structural support. Cross-linked polymers within the matrix create a complex three-dimensional architecture featuring water channels that facilitate nutrient transport and waste removal [174-177]. Furthermore, the biofilm matrix, with its negatively charged eDNA, has been shown to bind to positively charged antibiotics, thereby reducing there efficiency in penetration and subsequent destruction of the biofilm [178]. The matrix composition demonstrates remarkable plasticity, with bacteria modifying EPS production in response to environmental conditions including hydrodynamic conditions [179], temperature, pH, and nutrient availability [180]. This adaptability enables biofilms to maintain structural integrity and function across diverse environments.

Spatiotemporal organization within biofilms exhibits sophisticated patterns driven by both chemical and physical factors. Chemical gradients of oxygen, nutrients, and signaling molecules create distinct microenvironments that influence bacterial metabolism and gene expression [181]. These gradients can lead to metabolic stratification, with aerobic processes generally dominating in upper regions (due to more O_2 and light) and anaerobic processes occurring in deeper layers. Population-level coordination is achieved through quorum sensing networks, which regulate gene expression based on cell density through small diffusible molecules [182]. Multiple quorum sensing circuits may operate simultaneously, enabling fine-tuned responses to environmental conditions [183].

In medical contexts, biofilms present significant challenges due to their enhanced resistance to antimicrobial treatments. This resistance arises through multiple mechanisms: physical protection by the EPS matrix [178], altered metabolic states within the biofilm [184], and

the emergence of persister cells that can survive high antibiotic concentrations [185, 186]. Biofilms are implicated in a large number of human diseases, including caries, gingivitis, periodontitis, lung infections, catheter-based infections, eye infections, and medical device-associated infections [187]. Novel therapeutic approaches targeting biofilm-specific mechanisms are being developed, including matrix-degrading enzymes [188, 189], quorum sensing inhibitors [190], and surface modifications that prevent initial attachment [191].

In porous media systems, biofilm growth creates complex feedback loops between biological activity and physical transport processes. As biofilms develop, they can significantly reduce pore spaces, which alters flow patterns and creates preferential flow channels [192, 193]. This modification of the pore structure influences both hydraulic conductivity and solute transport pathways. As a simple consequence of the Bernoulli principle, local flow velocities can significantly increase in channels adjacent to biofilm-occupied pores, while becoming negligible in fully blocked pores. These hydrodynamic changes affect nutrient delivery and waste removal, creating spatial heterogeneities in biofilm growth and metabolism.

The mechanical properties of biofilms in porous media depend strongly on flow conditions and nutrient availability. When biofilms are exposed to shear stresses of about 0.01 to 1 Pa, chemical signaling between cells results in rapid adaptations that often take the form of increased EPS production [194]. Shear forces at critical thresholds (typically 0.1-1 Pa) can cause periodic sloughing events, where portions of the biofilm detach and are transported downstream, potentially colonizing new regions [195, 196]. In terms of nutrient availability, organic carbon [197], phosphate, and calcium [198] concentrations have been determined to be primary limiting factors of biofilm growth.

In subsurface environments, biofilm-mediated processes significantly influence biogeochemical cycling and contaminant fate and transport. Within the hyporheic zone, biofilms create reactive hotspots where enhanced mixing between surface water and groundwater streams promotes rapid nutrient transformation [199-201]. Biofilm development also affects hyporheic exchange flows by modifying both the hydraulic conductivity and reactive surface area of the sediment. Studies have shown that mature biofilms can significantly reduce hydraulic conductivity, thereby altering residence time distributions and reaction kinetics [202-204].

The influence of biofilms on reactive transport processes extends beyond simple flow modification. The EPS matrix acts as a selective barrier that can concentrate certain solutes while excluding others, creating unique microenvironments for chemical reactions. Metal ions, for example, can accumulate within the matrix at concentrations much higher than the bulk fluid due to binding with negatively charged EPS components [205, 206]. This concentration effect, combined with the diverse metabolic capabilities of biofilm communities, enables efficient transformation of various contaminants including heavy metals, organic pollutants, and pesticides.

Biofilm growth in reactive transport models is typically represented through coupled equations describing biomass accumulation and fluid flow. A common approach is the multispecies biofilm model introduced by Wanner and Gujer [207] and detailed by Wang and Zhang [208]. A continuum description of biofilm can be given as:

(23) $\frac{\partial f_i}{\partial t} = \mu o_i f_i - \frac{1}{\rho_i}\frac{\partial g_i}{\partial z}$

where the i subscript corresponds to the specific biofilm species, f is the volume fraction, μo is the specific growth rate, ρ is the biofilm density, and g is the mass flux of the biofilm. This flux can be expressed as $g_i(t,z) = u(t,z)\rho_i f_i(t,z)$. Substituting this into equation (9) yields

(24) $\frac{\partial f_i}{\partial t} = \left(\mu o_i - \frac{\partial u}{\partial z}\right) f_i - u\frac{\partial f_i}{\partial z}$

Which allows us to establish a relationship between the volume fraction of biofilm and flow speed. Summing over all microbial species and integrating over the vertical dimension then gives us the fluid velocity at the film-water interface of the biofilm as a function of thickness (L).

(25) $u_L = \int_0^L \bar{\mu} o(t, z')dz' + \sigma(t)$

where $\sigma(t)$ is defined as the exchange velocity between the biofilm and the bulk liquid. More sophisticated models incorporate biofilm mechanical properties, detachment due to shear stress, motility of dispersed cells, and equations for substrate interactions [209].

1.1.1.4 Nutrient Cycling and Biomass Growth in the Hyporheic Zone

Introduction and Historical Context of the Hyporheic Zone

The hyporheic zone represents the dynamic interface beneath and alongside stream beds where shallow groundwater and surface water mix. This region was first conceptualized by Orghidan in 1959 as an interstitial habitat with distinct fauna [210], but its significance in ecosystem functioning wasn't fully appreciated until decades later when the hyporheic zone came to be known as a critical ecotone linking surface and groundwater systems [211]. Early studies also demonstrated the zone's crucial role in organic matter processing and nutrient retention [212] and revealed complex patterns of nitrification and denitrification [213].

As research progressed, Boulton and colleagues synthesized knowledge on hyporheic zone hydrology, emphasizing how flow paths and residence times determine biogeochemical transformations [214]. The integration of molecular techniques in the early 2000s highlighted complex spatial organizations of biofilms in streambed sediments and their fundamental role in carbon and nitrogen cycling [215]. Around this time the concept of hyporheic "hot moments" - periods of intense biogeochemical activity – was also introduced, emphasizing the temporal

dynamics of nutrient cycling and how hydrological events and seasonal changes alter biogeochemical processes [216].

The functional significance of the hyporheic zone has become increasingly apparent through studies demonstrating its role in ecosystem processes ranging from nutrient transformation to contaminant degradation. Recent work by Rittmann and McCarty has emphasized how understanding microbial growth and yield in these systems is crucial for developing accurate reactive transport models [217]. This understanding becomes particularly important when predicting ecosystem responses to perturbations and optimizing bioremediation strategies in complex environmental systems. Other recent studies have identified the hyporheic zone as an important source of methane and carbon dioxide [218, 219].

Physical and Hydrological Controls

The physical and hydrological characteristics of the hyporheic zone fundamentally control its biogeochemical functioning. Flow magnitude and direction significantly influence these processes. Research has revealed distinct patterns in gaining streams, where hyporheic flux points upward from groundwater, versus losing streams, where it points downward from river water [220]. These flow patterns create unique chemical environments, as river water typically contains higher concentrations of bioavailable carbon and dissolved oxygen, leading to greater biomass growth in losing streams [212, 214].

Flow dynamics operate across multiple scales, from pore-scale processes to reach-scale exchange patterns. Studies have shown that higher flow velocities enhance nutrient and oxygen transport, potentially supporting greater steady-state biomass concentrations, though excessive speeds can cause biomass scouring [21, 215, 221]. However, research from chapter 5 of this dissertation indicates weaker correlations between biomass and vertical (-.09) and longitudinal

(.02) velocities than previously thought, which highlights the difficulty in predicting general relationships between processes in the hyporheic zone.

Under high flow conditions, microbial communities exhibit remarkable adaptations. Specifically, bacteria have been shown to increase extracellular polymeric substances (EPS) production, forming stronger biofilms resistant to erosion [221, 222]. The EPS matrix serves multiple functions beyond structural stability, including nutrient and enzyme retention, creating favorable microenvironments for microbial growth. These adaptations are particularly important in systems experiencing variable flow regimes, where maintaining attachment and metabolic function under changing conditions is crucial [222].

Factors Impacting Biomass Growth in the Hyporheic Zone

Temperature propagation through the hyporheic zone, influenced by flow patterns, creates distinct thermal regimes that affect microbial community structure and function. Seasonal temperature variations have been shown to lead to shifts in microbial community composition [223]. Furthermore, temperature has been show to generally impact a variety of biogeochemical process rates, including organic matter decomposition and nutrient cycling [224]. The thermal characteristics of the hyporheic zone are particularly important in the context of climate change, as temperature shifts can fundamentally alter microbial community structure and function.

Temperature fundamentally controls microbial metabolism and growth through multiple simultaneous mechanisms. Ratkowsky and colleagues established that while mesophilic bacteria thrive between 20-45°C, distinct adaptations enable psychrophiles and thermophiles to function effectively outside this range [225]. Morita's research on psychrophilic bacteria revealed specialized adaptations allowing growth below 20°C [226], while Schoolfield et al. demonstrated

that growth rate-temperature relationships typically follow modified Arrhenius kinetics, with rates approximately doubling for every 10°C increase until reaching an optimal temperature [227].

At a molecular scale, deviations beyond optimal temperatures lead to protein denaturation, enzyme inactivation, and membrane damage [228]. Furthermore, temperature not only influences cellular components but also substrate diffusion rates and enzyme kinetics, creating complex feedback loops in microbial growth responses [229]. These temperature-dependent processes become particularly relevant in the hyporheic zone, where thermal gradients can create distinct zones of microbial activity.

The influence of pH on bacterial growth extends beyond simple growth rate effects. While most species prefer conditions between pH 6.5-7.5, Horikoshi's work on alkaliphiles demonstrated remarkable adaptations allowing growth in extreme pH conditions [230]. Furthermore, studies have shown how pH affects multiple cellular processes simultaneously, including membrane transport, protein stability, energy metabolism, and cell wall integrity [231]. Thus, changes in environmental pH require significant energy expenditure to maintain the normal function of these cellular processes (i.e., cellular homeostasis), reducing growth efficiency and yield [232].

The relationship between nutrient availability and microbial growth, initially described by Monod's classical equation [233], has been significantly refined through modern research. Following work challenged the concept of fixed growth constants, suggesting instead an "uncertainty principle" in bacterial growth kinetics that better reflects the complexity of natural systems [234]. Furthermore, investigations into nutrient-limited transport revealed how bacteria respond to nutrient limitations through various adaptive mechanisms, including changes in cell size, morphology, and metabolic pathways [235].

Oxygen availability dramatically influences bacterial growth patterns in the hyporheic zone. Aerobic bacteria typically require dissolved oxygen levels above 2 mg/L for optimal growth [236]. Without oxygen, anaerobic bacteria may still thrive, although they generally exhibit lower growth rates due to less efficient energy generation pathways [237]. Bacterial respiratory systems have also shown remarkable flexibility, with many bacteria being able to adapt to varying oxygen levels by modifying their respiratory chains or switching between aerobic and anaerobic metabolism [238]. Stolper and colleagues further demonstrated that some bacteria can maintain aerobic metabolism at extremely low oxygen concentrations - as low as nanomolar levels - through specialized high-affinity oxidases [239]. This adaptability becomes particularly important in the hyporheic zone, where oxygen gradients can shift rapidly with changes in flow conditions.

Environmental stressors significantly impact microbial growth and community structure. Studies have detailed how heavy metals can inhibit growth through multiple mechanisms, including direct enzyme inhibition, membrane damage, DNA/RNA damage, and oxidative stress [240]. However, many bacteria have also developed various resistance mechanisms to metal toxicity [241].

Competition between bacterial species adds another layer of complexity to community dynamics. Direct competition for nutrients, production of inhibitory compounds, and modification of shared environments have been shown to influence community structure [242]. Hibbing et al. further elaborated on bacterial competition strategies, demonstrating how bacteria may respond through enhanced substrate uptake systems, production of antimicrobial compounds, or metabolic specialization [243].

Temporal and Seasonal Dynamics

Temporal variations in hyporheic zone processes create distinct patterns of biogeochemical activity. During summer months, warm temperatures propagate throughout the hyporheic zone, increasing growth rates [221, 222]. Winter conditions lead to reduced growth due to cold temperature propagation, though research has shown that specialized cold-adapted communities can maintain significant activity [244].

Nonperennial streams present unique challenges for microbial communities. Dry periods often create isolated pools with varied redox conditions [245], resulting in the development of specialized microbial populations in these environments [246]. These conditions can favor anaerobic processes such as denitrification [247] and iron reduction processes [248]. Microbes in nonperennial stream have also shown special adaptations that allow them to go dormancy during dry events and reemerge during wetting events, thus stimulating rapid changes in microbial community composition and function [249].

1.1.2 Particle Tracking

1.1.2.1 Introduction and Fundamental Concepts

Particle tracking, the analysis of video data to reconstruct object or organism trajectories, has become an essential quantitative tool across scientific disciplines, from biology and physics to environmental science and engineering. The foundational framework for modern particle tracking emerged through Crocker and Grier's seminal work, which introduced a robust algorithm for linking particle positions across video frames by minimizing total squared displacement [250]. This approach provided solutions for key challenges in particle tracking, including noise handling, particle density limitations, and position linking across frames.

The technical challenges in particle tracking arise from several key factors that must be addressed simultaneously. First, the accurate detection of particles requires sophisticated image

processing to distinguish true particles from noise and background variations. Second, the linking of particle positions between frames becomes increasingly complex with higher particle densities and faster movement speeds. Third, measurement uncertainties and detection failures must be handled robustly to maintain tracking accuracy over long trajectories. These challenges have driven the development of increasingly sophisticated algorithms and methodologies [251, 252].

The evolution of experimental techniques, particularly in microscopy and video technology, has continuously expanded tracking capabilities. Super-resolution microscopy techniques, exemplified by Betzig's work, pushed the boundaries of spatial resolution in particle detection [253], while improvements in high-speed imaging enabled the capture of rapid dynamics previously impossible to resolve. These technological advances have been complemented by the development of sophisticated tracking algorithms, including multiple hypothesis tracking (MHT) and probabilistic data association filters (PDAF), which have significantly improved tracking reliability in complex environments [252, 254].

1.1.2.2 Development of Tracking Algorithms

The evolution of particle tracking algorithms reflects the increasing complexity of tracking applications and the need for robust performance under challenging conditions. The earliest successful approaches relied on nearest-neighbor linking methods, but these proved insufficient for dense particle fields or complex motions. A significant advance came with the development of global linking strategies that consider multiple frames simultaneously to optimize trajectory reconstruction [250]. These methods minimize the global displacement across all possible particle matches, significantly improving tracking accuracy in complex scenarios.

Multiple hypothesis tracking (MHT) represented a major algorithmic breakthrough, particularly for biological applications. Jaqaman and colleagues demonstrated that MHT could

effectively handle particle merging and splitting events, temporary particle disappearance, and dense particle fields [252]. The method works by maintaining multiple possible trajectory hypotheses and selecting the most probable set of tracks based on all available information, making it particularly robust for tracking in biological systems where object interactions are common and frequent. The probabilistic data association filter (PDAF) approach, adapted for biological applications by Smal et al., provided a framework for explicitly handling measurement uncertainties [254]. This method is particularly valuable when tracking particles in low signal-to-noise conditions or when dealing with closely spaced targets. PDAF works by computing association probabilities between measurements and tracks, allowing for multiple potential measurement-to-track assignments and effectively handling uncertainty in the measurement process.

Some of the biggest breakthroughs in imaging techniques for particle tracking revolve around 3D tracking [258-260]. 3D tracking is technically challenging due to the algorithmic complexity of matching trajectories, fundamental difficulties in accurate depth determination and particle localizations, and increased computation time. Furthermore, 3D tracking, especially in porous media, represents a challenging problem due to the refraction of light around clear, synthetic grains, which may obfuscate or distort bacteria [2, 259, 260].

Comprehensive evaluation efforts have been crucial in understanding the strengths and limitations of different tracking approaches. Chenouard et al.'s objective comparison of particle tracking methods provided the first standardized assessment of various algorithms across different experimental conditions [255]. This work established benchmark datasets and evaluation metrics that continue to guide algorithm development and selection for specific applications. Recent

algorithmic developments have focused on improving computational efficiency and robustness while maintaining high accuracy in challenging scenarios [256, 257].

1.1.2.3 Applications in Environmental Science

Particle tracking methods have become fundamental tools for studying transport phenomena in environmental systems, particularly in porous media where complex geometry and flow patterns significantly influence particle and microorganism movement. The application of these techniques has enabled direct observation of processes previously only understood through bulk measurements or theoretical predictions [251]. High-speed imaging combined with sophisticated tracking algorithms has revealed detailed dynamics of both abiotic particle transport and microbial motility in these complex environments.

In the study of solute transport, particle tracking methods have provided unprecedented insights into the spatial heterogeneity of porous media flow. By tracking individual particles or fluorescent tracers, researchers can directly observe preferential flow paths, stagnation zones, and mixing behavior at the pore scale [254, 255]. These observations have proven crucial for validating and refining theoretical models of dispersion and mixing in porous media, particularly in systems where traditional bulk measurements fail to capture important local phenomena.

The application of particle tracking to microbial systems has revolutionized our understanding of bacterial transport and behavior in environmental contexts. These methods enable researchers to analyze individual bacterial trajectories, revealing how motility patterns change in response to environmental gradients, flow conditions, and surface interactions [215]. Particularly valuable has been the ability to distinguish between active bacterial movement and passive transport, providing insights into how microorganisms navigate and colonize porous

environments. The integration of tracking techniques with microfluidic devices has allowed for controlled studies of bacterial behavior under various environmental conditions [223].

For biofilm studies, particle tracking has enabled detailed investigation of the early stages of bacterial attachment and colonization. By tracking individual cells as they transition from planktonic to sessile states, researchers can better understand the mechanisms controlling biofilm formation and development [225]. These observations have practical implications for various environmental applications, from bioremediation to the management of bacterial fouling in industrial systems.

1.1.2.4 Current Challenges and Future Directions

Several key challenges persist in particle tracking applications, particularly for environmental systems where complex geometries, multiple scales of motion, and variable imaging conditions create significant technical difficulties. One fundamental challenge involves maintaining tracking accuracy when particles or organisms move across different focal planes or temporarily disappear from view, especially in three-dimensional porous media systems [255]. The development of more robust algorithms for handling these scenarios remains an active area of research, with recent work focusing on probabilistic methods that can maintain track continuity despite temporary loss of information [252, 256].

Technical advances in imaging capabilities continue to drive new applications in environmental science. High-speed cameras with improved sensitivity and resolution enable tracking of faster processes and smaller objects, while advances in 3D imaging techniques [258] allow for better characterization of transport in complex porous media. Future areas of research also include automatic analysis platforms, more robust comparisons between different PT

algorithms, and deep learning models for both the segmentation of particles and linking of trajectories.

1.1.3 Deep Learning for Particle Tracking and Reactive Transport

1.1.3.1 Introduction and Fundamental Deep Learning Concepts

Deep learning has emerged as a transformative tool in scientific computing, particularly for analyzing complex physical systems and processing large-scale experimental data. The foundations for these advances were established through breakthroughs in neural network training methodologies, including efficient backpropagation algorithms and GPU acceleration [261]. The 2012 ImageNet competition marked a pivotal moment when convolutional neural networks (CNNs) demonstrated unprecedented performance in image recognition tasks [262], catalyzing widespread adoption across scientific domains.

The application of deep learning to scientific problems has been enabled by several key technological developments. First, the development of specialized architectures like region-based CNNs and the YOLO (You Only Look Once) framework provided efficient tools for real-time object detection [263]. Second, the introduction of self-supervised learning approaches, exemplified by Momentum Contrast (MoCo) [264] and SimCLR [265], enabled effective learning from unlabeled data - a crucial capability for scientific applications where labeled data is often scarce. These advances have proven particularly valuable in environmental science and hydrology, where complex, high-dimensional problems demand sophisticated analytical approaches.

Recent reviews have highlighted how deep learning is also revolutionizing microscopy and particle tracking applications [266]. The key advantage of deep learning approaches lies in their ability to handle noise, variability, and complex backgrounds while maintaining high processing speeds. These capabilities have proven especially valuable in biological imaging, where traditional

methods often struggle with low signal-to-noise ratios and variable imaging conditions. They have shown particular promise for the segmentation and localization of cells/particles, but have only shown mild benefits for the task of trajectory linking/extraction.

1.1.3.2 Deep Learning Methods in Particle Detection and Tracking

The integration of deep learning into particle tracking has revolutionized our ability to analyze complex dynamic systems at multiple scales. A significant breakthrough came with Newby et al.'s work using CNNs to detect and track particles in dense suspensions, demonstrating superior detection performance compared to traditional methods, particularly in challenging scenarios with high particle density and low signal-to-noise ratios [267]. This approach has since been expanded through various specialized architectures and methodologies designed specifically for biological and physical applications.

Recent advances in deep learning-based tracking have addressed several key technical challenges. Yao and colleagues developed sophisticated data association methods that significantly improve tracking accuracy in complex biological environments [268]. Their approach specifically tackles the challenging problem of linking particle detections across frames, a critical issue in biological tracking applications. Spilger et al. further advanced the field by introducing a deep particle tracker specifically designed for fluorescence microscopy, demonstrating robust performance across varying imaging conditions [269].

The application of deep learning to live-cell imaging has enabled unprecedented insights into cellular dynamics. Song et al. developed automated multidimensional tracking systems capable of following particle movement in living cells with high precision [270], while Ritter and colleagues enhanced particle detection and tracking capabilities in fluorescence microscopy

through specialized neural network architectures [271]. These advances have been particularly valuable for studying bacterial motility and cellular transport processes.

Cell tracking and lineage reconstruction represent another frontier where deep learning has made significant impacts. Moen et al. demonstrated accurate cell tracking and lineage construction in live-cell imaging experiments [272], while Lugagne and colleagues developed DeLTA, an automated system for cell segmentation, tracking, and lineage reconstruction [273]. These tools have proven invaluable for studying bacterial population dynamics and cell-cell interactions in complex environments.

1.1.3.3 Physics-Informed Neural Networks and Reactive Transport

The integration of physical principles with deep learning architectures has created powerful new tools for modeling transport phenomena. Physics-informed neural networks (PINNs), introduced by Raissi et al., provide a framework for incorporating physical laws directly into neural network architectures [274]. This approach ensures that predictions respect fundamental physical constraints while leveraging the flexibility and computational efficiency of deep learning models. Kang et al. extended this concept by coupling generative adversarial networks with physics-driven models for inverse groundwater modeling [275], demonstrating improved predictions of contaminant transport in heterogeneous aquifers.

Recent advances in reactive transport modeling have focused on bridging scales and reducing computational costs. You and Lee developed deep learning methods for upscaling reactive transport models from pore-scale to continuum-scale [276], while Wang and Battiato created a framework specifically for modeling reactive transport and mineral precipitation in fracture-matrix systems [277]. These approaches have significantly reduced computational requirements while maintaining accuracy. Leal et al. demonstrated the power of machine learning

for accelerating chemical equilibrium calculations in reactive transport modeling [278], while Prasianakis et al. developed neural network approaches for process coupling and parameter upscaling [279].

The application of deep learning to multiphysics problems has enabled new approaches to complex system modeling. Jagtap et al. developed a framework to capture mixing patterns in reactive-transport systems [280], while Lu et al. created data-informed emulators for multi-physics simulations [281]. These advances have been particularly valuable for environmental applications, where multiple physical and chemical processes interact across different scales. Recent work by Basha et al. has further advanced the integration of machine learning with physics-driven modeling for multiphase systems [282].

1.1.3.4 Advanced Architectures and Multiphysics Applications

Advanced neural network architectures have emerged as powerful tools for modeling complex physical systems. Graph neural networks, as demonstrated by Sanchez-Gonzalez et al., have proven particularly effective for modeling many-body systems and particle interactions [283]. The transformer architecture, introduced by Vaswani et al. [284], has been adapted to capture long-range dependencies in particle trajectories and chemical reaction networks, enabling more accurate predictions of system behavior over extended time periods.

Generative models have found novel applications in environmental and physical modeling. Laloy et al. demonstrated the use of spatial generative adversarial networks for creating geologically realistic permeability fields [285], while recent climate modeling work by Yu et al. has shown how deep learning can emulate high-resolution physics in hybrid multi-scale simulators [286]. These approaches have been particularly valuable for generating realistic subsurface property fields and modeling complex environmental systems.

The application of deep learning to computational fluid dynamics and multiphysics simulations has seen significant advances. Obiols-Sales et al. developed CFDNet, demonstrating substantial acceleration of fluid simulations through deep learning [287], while Koric and Abueidda applied deep learning sequence methods to multiphysics modeling of materials processes [288]. Gunawardena et al. extended these approaches to atmospheric transport models [289], showing how machine learning emulation can effectively capture complex spatial deposition patterns. Mo et al. demonstrated the use of deep convolutional encoder-decoder networks for uncertainty quantification in dynamic multiphase flow [290], while recent work has focused on developing hybrid approaches that combine the efficiency of deep learning with the physical accuracy of traditional numerical methods. These developments have enabled new applications in environmental monitoring, industrial process optimization, and scientific discovery.

1.1.3.5 Current Applications and Future Directions

Deep learning applications in particle tracking and reactive transport continue to expand, driven by both technological advances and practical needs. In biological imaging, recent developments have enabled automated analysis of increasingly complex cellular behaviors. The integration of deep learning with traditional microscopy has enhanced our ability to track multiple particles simultaneously while maintaining high temporal and spatial resolution [265, 270]. These advances have proven particularly valuable for studying bacterial motility patterns, cell-cell interactions, and biofilm formation dynamics in real-time.

The coupling of machine learning with physics-based modeling represents a rapidly evolving frontier. Recent work has demonstrated significant progress in developing hybrid approaches that combine the computational efficiency of neural networks with the physical accuracy of traditional numerical methods [281, 282]. These hybrid models are particularly

promising for environmental applications, where complex multiscale processes must be simulated efficiently. Emerging research focuses on developing interpretable deep learning models that can provide insights into the physical mechanisms underlying observed phenomena while maintaining computational tractability.

Several key challenges and opportunities in this domain include the development of self-supervised and few-shot learning approaches to address the scarcity of labeled data in scientific applications, integration of uncertainty quantification in deep learning predictions for improved reliability in scientific applications, creation of scalable architectures capable of handling increasingly large and complex datasets, and development of physics-informed approaches that can effectively bridge multiple spatial and temporal scales [276, 277]. The future of this field appears to be moving toward increasingly sophisticated hybrid approaches. These include the development of neural network architectures that can automatically map and incorporate physical laws [278, 279], the use of reinforcement learning for optimizing experimental designs, and the application of federated learning techniques for collaborative modeling of reactive transport processes. As computational capabilities continue to advance, these methods are expected to enable new insights into complex environmental and biological systems that were previously difficult to study.

1.2 Purpose and Scope

1.2.1 Original Intent

The original intent of this research was to explore and develop scalable models for microbe-mediated reactive transport processes, specifically focusing on bacterial motility, chemotaxis, and bacterial transport at the pore and field scales. Our hypothesis was (and still is) that an improved

understanding of these processes and the development of more accurate models would allow us to gain further insight into the complex interactions between hydrodynamics, microbial motility, and biofilm structure in porous media.

The study was initially designed to investigate how different microbial species contribute to reactive transport, particularly in subsurface environments where fluid flow and chemical gradients play a significant role in determining microbial activity and nutrient cycling. By integrating experimental observations with computational models, the goal was to provide mechanistic insights into microbial motility, biofilm development, and their collective impact on reactive transport. Specifically, our intent for future chapters was to upscale our experimental results at the micro scale to the pore and field scales, thus enhancing models of microbial motility across all scales. In addition, we sought to investigate the impact of micro-scale motility processes on field-scale bioremediation, which is a phenomena that is currently poorly understood.

1.2.2 Broadening of Scope

As the research progressed, it became clear that additional complexities related to microbial behavior and environmental conditions necessitated a broadening of the original scope. While the initial focus was on microbial motility across a range of scales, the research expanded to include more comprehensive analyses of particle tracking, biomass growth in the hyporheic zone, and deep learning tools for particle tracking and reactive transport simulation upscaling.

One of the key reasons for the broadening of scope was the fact that our micro-scale research turned out to be more challenging than what was originally predicted. These challenges primarily revolved around difficulties in image analysis that required the development of a novel background subtraction algorithm and experimentation with numerous particle tracking methods. This delay in the production of results shifted our timeline. Furthermore, we realized that our

experiments with comparing the performance of multiple particle tracking algorithms had produced interesting results that were worth trying to publish. Rather than continue with our original plans to move on to upscaling our micro-scale results to the pore scale, we decided to pursue the research that was ready to go.

After two and a half years of my PhD, both of these chapters were nearly complete, and the next object of our focus was to incorporate the results from our micro-scale investigations of microbial motility into field-scale reactive transport simulations. The plan for this chapter was to use published research on field scale bioremediation that uses some reactive transport simulator to model the in-situ data. By adapting the model previously used in the published research to include microbial motility parameters, we hoped to gain insight into how bacterial motion impacts the fit between the model and the experimental data. However, due to a variety of complications, we ultimately weren't able to achieve these goals. One of the primary barriers we faced is that there is only one study we could find that has both in-situ data and a reactive transport model in a well-supported framework for bioremediation at the field scale [291]. To further complicate matters, the model was originally created in STOMP. However, STOMP is a proprietary code that requires a license to run, meaning we needed to port the simulation to PFLOTRAN. We spent multiple months trying to rewrite the simulation in PFLOTRAN, but were ultimately unable to reproduce the time series from the published research given the complexities of the reactions that were modeled. Luckily, we were able to pivot our objectives as we were working on this chapter that never came to be, which resulted in a significant broadening of scope. We still wanted to explore microbe-mediate reactive transport processes at large scales, which resulted in our pivot to investigate the feedback mechanisms of biomass growth in the hyporheic zone and the velocity-based decay of biomass. Although this chapter no longer uses our micro-scale research to improve

Darcy and field-scale reactive transport simulations, we calibrated our velocity-based biomass decay based on others' micro-scale research, thus completing our objective to upscale microbial physics in some way.

In addition to our pivot in large-scale studies of microbial physics, we also actively decided to look for applications of deep learning within our research. Specifically, we identified significant gaps in the literature for applications of deep learning to particle tracking and reactive transport upscaling. Thus, two chapters of this dissertation are devoted to applications of deep learning within the broad domain of microbe-mediated reactive transport. Although other directions may have provided for more thorough investigations of microbial motility or reactive transport modeling, the expansion of this dissertation through comparative analysis and deep learning tools provides for a much more rigorous analysis of particle tracking, which is the primary tool used to understand microbial motility. Furthermore, this broadening of scope increases its relevancy to a wider audience and provides more points to design further studies from.

1.2.3 Organization of Disseration

This dissertation is organized into 7 separate chapters. Chapter 1 is the current chapter, which gives readers most of the background knowledge needed to understand the contributions of the dissertation (chapter 1). This introduction was created with the intention of being an in-depth summary of the basic theories explored throughout this dissertation. Instead of providing introductions that thoroughly link the concepts of each chapter, this introduction serves to establish current knowledge of these links, and the concluding chapter is used to link the different results of each chapter in the context of the studies cited in this introduction. If readers find themselves confused about chapter progression, they should refer back to the abstract and introduction. After the background and literature review, the introduction continues with sections

detailing the original plans for the thesis, the broadening of its scope over time, the organization of the dissertation (this section), which provides an overview of each chapter along with a conceptual diagram of the whole dissertation (Fig. 1), and the primary contributions it delivers to the research community.

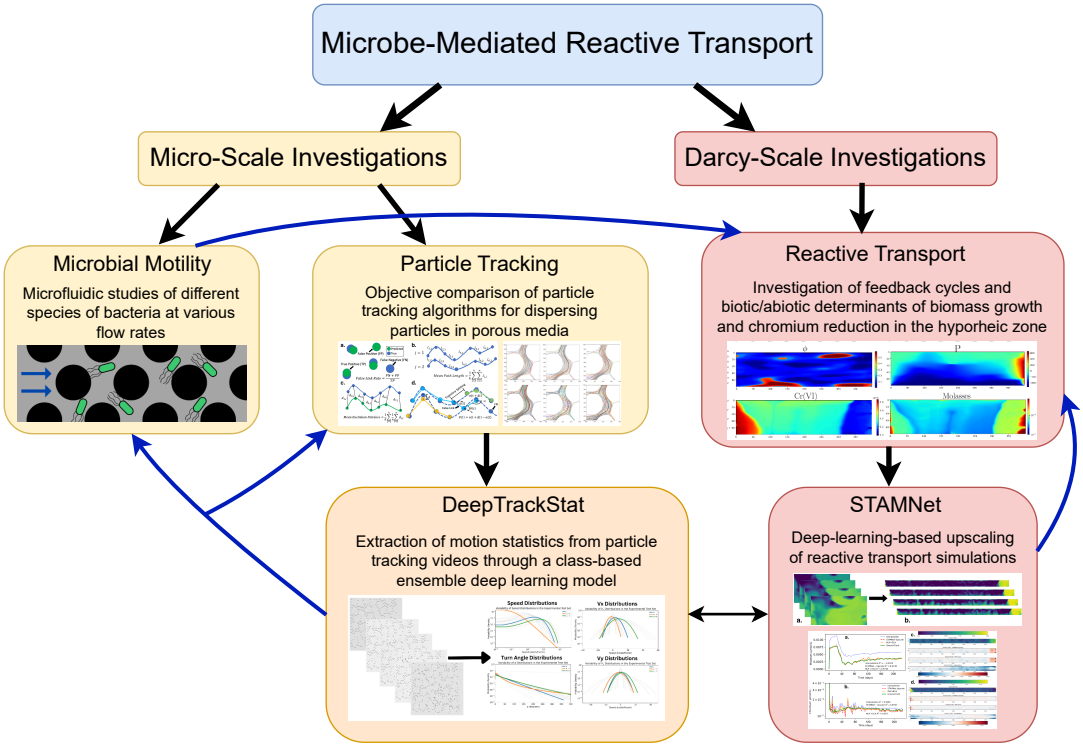


Figure 1. Conceptual diagram of the applications of the various chapters of this dissertation. Chapter 2 focuses on microbial motility, chapter 3 focuses on particle tracking (with applications to chapter 2), chapter 4 focuses on a novel deep learning model (named DeepTrackStat) for the extraction of motion statistics from videos of dispersing particles (with applications to chapters 2 and 3), chapter 5 focuses on the incorporation of novel physics in reactive transport simulations, and chapter 6 focuses on the deep-learning-based upscaling of reactive transport simulations. Chapter 6 also shares similarities with chapter 4 with regards to using deep learning to process spatiotemporal inputs, and chapter 2 contains information that, while not explicitly used in this dissertation for chapter 5, may be used to improve the reactive transport simulations featured in chapter 5. Chapter 4 is shaded a different color (orange) because although it was developed for micro-scale applications (yellow), it may also be applied to videos of particle transport at larger scales (red). These chapters act to improve various aspects of microbe-mediated reactive transport, culminating in a dissertation that contains a variety of new tools and insights that may be used to help improve environmental management strategies.

The dissertation then moves to an investigation of the advection-dominated transport dynamics of pili and flagella-mediated motile bacteria in porous media (chapter 2). This chapter focuses on understanding the dynamics of bacterial transport in microfluidics experiments for three different species of motile bacteria under varying flow rates and porosities. In this chapter bacterial transport is characterized through turn angle and speed distributions, mean square displacements, and dispersion coefficients. Furthermore, OpenFOAM simulations are used to model the Eulerian flow field of our microfluidic geometries to allow for comparison between the bacterial speed distributions and the speed distributions of the flow fields. The results of this chapter show that higher flow rates suppress bacterial motility, leading to advection-dominated transport, where bacterial motion is primarily driven by the flow of fluid. Furthermore, it is shown that at high flow speeds bacteria with peritrichous flagella maintain their motility characteristics to a higher degree than bacteria with monotrichous flagella or pili do.

Next, the dissertation presents an evaluation of particle tracking (PT) codes used for dispersing particles in porous media (chapter 3). Specifically, this chapter primarily investigates tracking performance for particles in similar environments to our microfluidic experiments performed in chapter 2. Using a suite of classical and experimental comparison statistics, and visual analysis of trajectories, this chapter provides a rigorous comparison between four different PT algorithms: V-TrackMat, TrackMate-Kalman, TrackMate-LAP, and Trackpy. The results of this chapter show the importance of a wholistic and task-relevant approach to using statistical comparisons to measure differences in PT algorithm performance. Concomitantly, these results highlight the poor performance of all PT algorithms in situations of high speed, high density particle simulations, and further show the impact of this poor performance on statistics relevant for bacterial transport.

In response to the challenges identified in particle tracking, this dissertation next introduces DeepTrackStat, an end-to-end deep learning framework for the extraction of motion statistics from videos of particles (chapter 4). DeepTrackStat (DTS) is designed to automate and enhance the process of generating speed, velocity component, and turn angle statistics, particularly in complex and high-speed environments where traditional methods fall short. The deep learning framework efficiently processes large datasets, allowing for fast and accurate motion extraction from videos of dispersing particles in porous media. To make DTS as robust as possible, it was trained on variations in flow/motion type and speed, and particle density, shape, brightness. We show that DTS generally performs better and is also faster than classical particle tracking algorithms. Furthermore, DTS shows especially good performance for high-speed particles.

Next, this dissertation explores nutrient cycling and speed-based biomass decay in the hyporheic zone (chapter 5). Using 344 different simulations of biomass growth in the hyporheic zone, this chapter presents a comprehensive sensitivity analysis of the major feedback cycles present. Furthermore, this chapter presents a sensitivity analysis of speed-based biomass decay, finding that the calibrated parameters are not very sensitive, but that weakly cohesive biofilm connections, represented by a low value of our parameter β, will significantly impact biomass growth. In this chapter we further investigate general trends of our simulations in relationship to chromium reduction. Overall, we find that reduction is primarily dominated by abiotic processes, but that biomass growth can lead to reduction hotspots.

One of the limitations identified in chapter 5 is that many of the simulations are only for a small 1x2 meter scale, which reduces the generalizability of the results. To address this concern, and make general improvements to reactive transport modeling, chapter 6 of this dissertation

details the development of STAMNet, a spatiotemporal attention-based neural network for upscaling reactive transport simulations. STAMNet overcomes the traditional computational limitations of reactive transport simulations by leveraging deep learning to capture the spatiotemporal dependencies in reactive transport processes. The network is trained to recognize patterns at a 1x2 meter scale and use them to predict behavior at a 1x20 meter scale, offering a significant improvement in simulation efficiency compared to RT simulators. STAMNet was applied to scenarios involving biomass growth and the transport of contaminants like chromium (Cr(VI)), demonstrating its capability to handle real-world environmental problems. By incorporating a novel spatiotemporal attention mechanism, STAMNet was able to focus on the most relevant spatial and temporal features, providing more accurate predictions of large-scale transport phenomena than other simpler neural architectures could.

Chapter 7 then presents a summary of the information in, and synthesis of the relationships between, the five primary chapters (chapters 2-6), then finishes with the conclusion of this dissertation. This chapter discusses how the dissertation as a whole contributes to advancements in microbial transport understanding, particle tracking methodologies, and reactive transport modeling. Furthermore, this chapter highlights our contributions from the micro-scale to the field scale, discusses relevant contributions to bioremediation, and gives a road map for potential future research. This dissertation is then concluded with a summary, and a discussion of limitations and future work. Ultimately, this dissertation was created to provide researchers with new theories, models, and tools surrounding microbe-mediated reactive transport, with the hope that these advancements will allow for further development of knowledge in this domain.

1.2.4 Research Contributions

This dissertation makes several significant contributions across different areas of microbe-mediated reactive transport, particle tracking, and deep learning, advancing both theoretical understanding and practical applications. The key contributions are outlined below:

1. **Microbial Motility in Porous Media**
    a. ***Comprehensive Analysis of Multiple Bacterial Species***: Prior studies often focused on a single bacterial species or simplified flow conditions. This dissertation advances the field by directly comparing the motility of bacteria with different mechanisms (e.g., flagella, pili) under various flow rates and pore geometries. This comparative study enhances the understanding of microbial behavior in a variety of flow conditions and provides critical insights into how diverse bacterial communities may behave in subsurface conditions. Specifically, this chapter shows that bacteria with peritrichous flagella maintain their motility characteristics at higher flow speeds than bacteria without flagella and with monotrichous flagella.
    b. ***Linking Microscale Behavior to Macroscale Transport***: The identification of a transition from motility-driven to advection-dominated transport at higher flow rates bridges the gap between small-scale bacterial behavior and larger-scale transport phenomena. This is crucial for developing accurate upscaling techniques, informing when simplified, advection-based models are appropriate for describing bacterial transport in high-flow scenarios.
2. **Particle Tracking in Porous Media**
    a. ***Comprehensive Evaluation of Particle Tracking Methods***: This dissertation evaluates multiple particle tracking (PT) algorithms across a wide range of flow conditions for transport of particles in porous media. By rigorously comparing tracking methods under complex scenarios, this work offers valuable guidance for researchers in

selecting appropriate tracking methods, designing experimental setups, and making appropriate performance comparisons for the tracking of dispersive particles in porous media.

b. ***Development of DeepTrackStat***: The creation of DeepTrackStat, a class-based ensemble deep learning framework, represents a significant advancement in the general field of particle tracking. DeepTrackStat is specifically designed for the extraction of motion statistics, meaning it does not allow for trajectory reconstruction like traditional particle tracking methods do. However, for the task of motion statistic extraction, it overcomes many limitations of traditional PT methods. It dramatically enhances the speed and accuracy of extracting motion statistics from particle videos, especially for high-particle-speed scenarios, meaning it is a powerful tool that researchers can add to their particle tracking toolbox.

3. Nutrient Cycling and Biomass Growth in the Hyporheic Zone

a. ***Coupling Physical, Chemical, and Biological Processes***: This research provides a comprehensive analysis of feedback cycles in the hyporheic zone, explicitly considering the interactions between flow conditions, nutrient availability, and biomass growth. This holistic approach goes beyond previous studies, which often considered these factors in isolation, advancing the understanding of nutrient cycling and biomass dynamics in critical environmental interfaces.

b. ***Refining Paradigms in Chromium Reduction***: The finding that abiotic processes generally dominate chromium reduction in high-nutrient scenarios refines the existing paradigm of the competition between abiotic and biotic reduction. Furthermore, the finding that high biomass concentrations result in hot spots of chromium reduction adds additional nuance to our understanding of chromium remediation. These

insights underscore the importance of considering both biotic and abiotic pathways in contaminant models, especially for the design of more effective remediation strategies for redox-sensitive contaminants.

4. **Reactive Transport Simulations**

    a. ***Introducing a Velocity-Based Biomass Decay Model***: The development of a velocity-based biomass decay model used within PFLOTRAN is a significant contribution to the general field of reactive transport simulations (and specifically for the sub-field of microbe-mediated reactive transport). It captures the influence of fluid shear on biofilm stability and growth, which has often been oversimplified in prior research. This more realistic representation of biofilm dynamics is crucial for accurately predicting contaminant fate, transport and nutrient cycling in flowing subsurface systems.

    b. ***Creation of STAMNet for Scalable Simulations***: Another major contribution of this dissertation is the development of STAMNet, a spatiotemporal attention-based neural network for reactive transport simulations. STAMNet addresses one of the most persistent challenges in the field—scaling models to field-relevant sizes without losing important spatial and temporal details. By capturing complex dependencies in reactive transport processes, STAMNet opens new possibilities for efficient, large-scale simulations.

    c. ***Applying Deep Learning for Physical Upscaling***: The use of deep learning for physically meaningful upscaling is another notable contribution of this dissertation. STAMNet demonstrates that deep learning can maintain critical spatial and temporal relationships in reactive transport processes, something often lost in more simplified

upscaling methods. This advancement extends the growing field of deep learning applications in environmental science and geosciences.

These contributions represent significant advancements in microbial motility research, particle tracking methodologies, large-scale simulation models, and understanding the dynamic processes in the hyporheic zone. Ultimately, this dissertation provides a robust framework for future research and practical applications in the general fields of environmental management and bioremediation.

References

[1] Yarwood, R.R., Rockhold, M.L., Niemet, M.R., Selker, J.S. and Bottomley, P.J., 2006. Impact of microbial growth on water flow and solute transport in unsaturated porous media. Water resources research, 42(10).

[2] Sandrin, S.K., Jordan, F.L., Maier, R.M. and Brusseau, M.L., 2001. Biodegradation during contaminant transport in porous media: 4. Impact of microbial lag and bacterial cell growth. Journal of contaminant hydrology, 50(3-4), pp.225-242.

[3] Bellin, C.A. and Rao, P.S.C., 1993. Impact of bacterial biomass on contaminant sorption and transport in a subsurface soil. Applied and environmental microbiology, 59(6), pp.1813-1820.

[4] Ginn, T. R., et al. (2002). Processes in microbial transport in the natural subsurface. Advances in Water Resources, 25(8-12), 1017-1042.

[5] Bravo, P., Lung Ng, S., MacGillivray, K.A., Hammer, B.K. and Yunker, P.J., 2023. Vertical growth dynamics of biofilms. Proceedings of the National Academy of Sciences, 120(11), p.e2214211120.

[6] Mazza, M.G., 2016. The physics of biofilms—an introduction. Journal of Physics D: Applied Physics, 49(20), p.203001.

[7] Geesey, G.G., Stupy, M.W. and Bremer, P.J., 1992. The dynamics of biofilms. International biodeterioration & biodegradation, 30(2-3), pp.135-154.

[8] Hartmann, R., Singh, P.K., Pearce, P., Mok, R., Song, B., Díaz-Pascual, F., Dunkel, J. and Drescher, K., 2019. Emergence of three-dimensional order and structure in growing biofilms. Nature physics, 15(3), pp.251-256.

[9] Newby, D.T., Pepper, I.L. and Maier, R.M., 2009. Microbial transport. In Environmental microbiology (pp. 365-383). Academic Press.

[10] Mailloux, B.J., Fuller, M.E., Onstott, T.C., Hall, J., Dong, H., DeFlaun, M.F., Streger, S.H., Rothmel, R.K., Green, M., Swift, D.J. and Radke, J., 2003. The role of physical, chemical, and microbial heterogeneity on the field-scale transport and attachment of bacteria. Water Resources Research, 39(6).

[11] Mueller, R.F., 1996. Bacterial transport and colonization in low nutrient environments. Water Research, 30(11), pp.2681-2690.

[12] Secchi, E., Vitale, A., Miño, G.L., Kantsler, V., Eberl, L., Rusconi, R. and Stocker, R., 2020. The effect of flow on swimming bacteria controls the initial colonization of curved surfaces. Nature communications, 11(1), p.2851.

[13] de Anna, P., Pahlavan, A.A., Yawata, Y., Stocker, R. and Juanes, R., 2021. Chemotaxis under flow disorder shapes microbial dispersion in porous media. Nature Physics, 17(1), pp.68-73.

[14] Keegstra, J.M., Carrara, F. and Stocker, R., 2022. The ecological roles of bacterial chemotaxis. Nature Reviews Microbiology, 20(8), pp.491-504.

[15] Colin, R., Ni, B., Laganenka, L. and Sourjik, V., 2021. Multiple functions of flagellar motility and chemotaxis in bacterial physiology. FEMS microbiology reviews, 45(6), p.fuab038.

[16] Berg, H.C., 1975. Chemotaxis in bacteria. Annual review of biophysics and bioengineering, 4(00), pp.119-136.

[17] Segall, J.E., Block, S.M. and Berg, H.C., 1986. Temporal comparisons in bacterial chemotaxis. Proceedings of the National Academy of Sciences, 83(23), pp.8987-8991.

[18] Hermansson, M., 1999. The DLVO theory in microbial adhesion. Colloids and surfaces B: Biointerfaces, 14(1-4), pp.105-119.

[19] Bai, H., Cochet, N., Pauss, A. and Lamy, E., 2017. DLVO, hydrophobic, capillary and hydrodynamic forces acting on bacteria at solid-air-water interfaces: their relative impact on bacteria deposition mechanisms in unsaturated porous media. Colloids and Surfaces B: Biointerfaces, 150, pp.41-49.
[20] Cunningham, A.B., Characklis, W.G., Abedeen, F. and Crawford, D., 1991. Influence of biofilm accumulation on porous media hydrodynamics. Environmental science & technology, 25(7), pp.1305-1311.
[21] Wei, G. and Yang, J.Q., 2023. Impacts of hydrodynamic conditions and microscale surface roughness on the critical shear stress to develop and thickness of early-stage Pseudomonas putida biofilms. Biotechnology and bioengineering, 120(7), pp.1797-1808.
[22] Frymier, P.D., Ford, R.M., Berg, H.C. and Cummings, P.T., 1995. Three-dimensional tracking of motile bacteria near a solid planar surface. Proceedings of the National Academy of Sciences, 92(13), pp.6195-6199.
[23] Skerker, J.M. and Berg, H.C., 2001. Direct observation of extension and retraction of type IV pili. Proceedings of the National Academy of Sciences, 98(12), pp.6901-6904.
[24] Meister, M., Lowe, G. and Berg, H.C., 1987. The proton flux through the bacterial flagellar motor. Cell, 49(5), pp.643-650.
[25] Mitchell, J.G. and Kogure, K., 2006. Bacterial motility: links to the environment and a driving force for microbial physics. FEMS microbiology ecology, 55(1), pp.3-16.
[26] Budrene, E.O. and Berg, H.C., 1991. Complex patterns formed by motile cells of Escherichia coli. Nature, 349(6310), pp.630-633.
[27] Wadhwa, N. and Berg, H.C., 2022. Bacterial motility: machinery and mechanisms. Nature reviews microbiology, 20(3), pp.161-173.
[28] Raina, J.B., Fernandez, V., Lambert, B., Stocker, R. and Seymour, J.R., 2019. The role of microbial motility and chemotaxis in symbiosis. Nature Reviews Microbiology, 17(5), pp.284-294.
[29] Mitchell, J.G. and Kogure, K., 2006. Bacterial motility: links to the environment and a driving force for microbial physics. FEMS microbiology ecology, 55(1), pp.3-16.
[30] Picioreanu, C., Kreft, J.U., Klausen, M., Haagensen, J.A.J., Tolker-Nielsen, T. and Molin, S., 2007. Microbial motility involvement in biofilm structure formation–a 3D modelling study. Water science and technology, 55(8-9), pp.337-343.
[31] Dentz, M., Creppy, A., Douarche, C., Clément, E. and Auradou, H., 2022. Dispersion of motile bacteria in a porous medium. Journal of Fluid Mechanics, 946, p.A33.
[32] Mosley, O.E., Gios, E., Close, M., Weaver, L., Daughney, C. and Handley, K.M., 2022. Nitrogen cycling and microbial cooperation in the terrestrial subsurface. The ISME Journal, 16(11), pp.2561-2573.
[33] Chen, X., Ye, Q., Sanders, C.J., Du, J. and Zhang, J., 2020. Bacterial-derived nutrient and carbon source-sink behaviors in a sandy beach subterranean estuary. Marine Pollution Bulletin, 160, p.111570.
[34] Krause, S., Lewandowski, J., Grimm, N.B., Hannah, D.M., Pinay, G., McDonald, K., Martí, E., Argerich, A., Pfister, L., Klaus, J. and Battin, T., 2024. Ecohydrological interfaces as hotspots of ecosystem processes. Ecohydrological Interfaces, pp.1-28.
[35] Alexander, M. and Scow, K.M., 1989. Kinetics of biodegradation in soil. Reactions and movement of organic chemicals in soils, 22, pp.243-269.
[36] Banfield, J. F., & Nealson, K. H. (1997). Geomicrobiology: Interactions between Microbes and Minerals. Mineralogical Society of America.
[37] Yaashikaa, P.R., Kumar, P.S., Jeevanantham, S. and Saravanan, R., 2022. A review on bioremediation approach for heavy metal detoxification and accumulation in plants. Environmental Pollution, 301, p.119035.
[38] Bala, S., Garg, D., Thirumalesh, B.V., Sharma, M., Sridhar, K., Inbaraj, B.S. and Tripathi, M., 2022. Recent strategies for bioremediation of emerging pollutants: a review for a green and sustainable environment. Toxics, 10(8), p.484.
[39] Saeed, M.U., Hussain, N., Sumrin, A., Shahbaz, A., Noor, S., Bilal, M., Aleya, L. and Iqbal, H.M., 2022. Microbial bioremediation strategies with wastewater treatment potentialities–A review. Science of the total environment, 818, p.151754.
[40] Mehrotra, T., Dev, S., Banerjee, A., Chatterjee, A., Singh, R. and Aggarwal, S., 2021. Use of immobilized bacteria for environmental bioremediation: A review. Journal of Environmental Chemical Engineering, 9(5), p.105920.
[41] Pushkar, B., Sevak, P., Parab, S. and Nilkanth, N., 2021. Chromium pollution and its bioremediation mechanisms in bacteria: A review. Journal of Environmental Management, 287, p.112279.
[42] Perez, L.J., Parashar, R., Plymale, A. and Scheibe, T.D., 2022. Contributions of biofilm-induced flow heterogeneities to solute retention and anomalous transport features in porous media. Water Research, 209, p.117896.
[43] Orgogozo, L., Golfier, F., Buès, M. and Quintard, M., 2010. Upscaling of transport processes in porous media with biofilms in non-equilibrium conditions. Advances in Water Resources, 33(5), pp.585-600.
[44] Peszynska, M., Trykozko, A., Iltis, G., Schlueter, S. and Wildenschild, D., 2016. Biofilm growth in porous media: Experiments, computational modeling at the porescale, and upscaling. Advances in Water Resources, 95, pp.288-301.
[45] Forbes, N.S., 2010. Engineering the perfect (bacterial) cancer therapy. Nature Reviews Cancer, 10(11), pp.785-794.

[46] Tufenkji, N., 2007. Modeling microbial transport in porous media: Traditional approaches and recent developments. Advances in water resources, 30(6-7), pp.1455-1469.
[47] Ford, R.M. and Harvey, R.W., 2007. Role of chemotaxis in the transport of bacteria through saturated porous media. Advances in Water Resources, 30(6-7), pp.1608-1617.
[48] Corapcioglu, M.Y. and Haridas, A., 1984. Transport and fate of microorganisms in porous media: a theoretical investigation. Journal of Hydrology, 72(1-2), pp.149-169.
[49] Hornberger, G.M., Mills, A.L. and Herman, J.S., 1992. Bacterial transport in porous media: Evaluation of a model using laboratory observations. Water Resources Research, 28(3), pp.915-923.
[50] Bhattacharjee, T., Amchin, D.B., Ott, J.A., Kratz, F. and Datta, S.S., 2021. Chemotactic migration of bacteria in porous media. Biophysical Journal, 120(16), pp.3483-3497.
[51] Li, B.L., Loehle, C. and Malon, D., 1996. Microbial transport through heterogeneous porous media: random walk, fractal, and percolation approaches. Ecological Modelling, 85(2-3), pp.285-302.
[52] Ahmed, T. and Stocker, R., 2008. Experimental verification of the behavioral foundation of bacterial transport parameters using microfluidics. Biophysical journal, 95(9), pp.4481-4493.
[53] Hol, F.J. and Dekker, C., 2014. Zooming in to see the bigger picture: microfluidic and nanofabrication tools to study bacteria. Science, 346(6208), p.1251821.
[54] Scheibe, T.D., Dong, H. and Xie, Y., 2007. Correlation between bacterial attachment rate coefficients and hydraulic conductivity and its effect on field-scale bacterial transport. Advances in water resources, 30(6-7), pp.1571-1582.
[55] Wang, M., Ford, R.M. and Harvey, R.W., 2008. Coupled effect of chemotaxis and growth on microbial distributions in organic-amended aquifer sediments: Observations from laboratory and field studies. Environmental science & technology, 42(10), pp.3556-3562.
[56] Minto, J.M., Lunn, R.J. and El Mountassir, G., 2019. Development of a reactive transport model for field-scale simulation of microbially induced carbonate precipitation. Water Resources Research, 55(8), pp.7229-7245.
[57] Hron, P., Jost, D., Bastian, P., Gallert, C., Winter, J. and Ippisch, O., 2015. Application of reactive transport modeling to growth and transport of microorganisms in the capillary fringe. Vadose Zone Journal, 14(5), pp.vzj2014-07.
[58] Boano, F., Harvey, J.W., Marion, A., Packman, A.I., Revelli, R., Ridolfi, L. and Wörman, A., 2014. Hyporheic flow and transport processes: Mechanisms, models, and biogeochemical implications. Reviews of geophysics, 52(4), pp.603-679.
[59] Wang, Y.T. and Shen, H., 1995. Bacterial reduction of hexavalent chromium. Journal of industrial microbiology and biotechnology, 14(2), pp.159-163.
[60] Jiang, K., Zhang, J., Deng, Z., Barnie, S., Chang, J., Zou, Y., Guan, X., Liu, F. and Chen, H., 2021. Natural attenuation mechanism of hexavalent chromium in a wetland: Zoning characteristics of abiotic and biotic effects. Environmental Pollution, 287, p.117639.
[61] Peralta-Maraver, I., Reiss, J. and Robertson, A.L., 2018. Interplay of hydrology, community ecology and pollutant attenuation in the hyporheic zone. Science of the Total Environment, 610, pp.267-275.
[62] Lawrence, J.E., Skold, M.E., Hussain, F.A., Silverman, D.R., Resh, V.H., Sedlak, D.L., Luthy, R.G. and McCray, J.E., 2013. Hyporheic zone in urban streams: A review and opportunities for enhancing water quality and improving aquatic habitat by active management. Environmental Engineering Science, 30(8), pp.480-501.
[63] Bardini, L., Boano, F., Cardenas, M.B., Revelli, R. and Ridolfi, L., 2012. Nutrient cycling in bedform induced hyporheic zones. Geochimica et Cosmochimica Acta, 84, pp.47-61.
[64] Ahmad, I., Ilyas, H., Urooj, A., Aslam, M.S., Shoaib, M. and Raja, M.A.Z., 2019. Novel applications of intelligent computing paradigms for the analysis of nonlinear reactive transport model of the fluid in soft tissues and microvessels. Neural Computing and Applications, 31, pp.9041-9059.
[65] Beg, O.A., Debnath, S. and Roy, A.K., 2022. Reactive solute transport in blood flow through a permeable capillary. Archives of Mechanics.
[66] Möller, M.N., Cuevasanta, E., Orrico, F., Lopez, A.C., Thomson, L. and Denicola, A., 2019. Diffusion and transport of reactive species across cell membranes. Bioactive lipids in health and disease, pp.3-19.
[67] Zhang, L., Gardiner, B.S., Smith, D.W., Pivonka, P. and Grodzinsky, A., 2008. A fully coupled poroelastic reactive-transport model of cartilage. Molecular & Cellular Biomechanics, 5(2), p.133.
[68] Ishikawa, T., Sato, T., Mohit, G., Imai, Y. and Yamaguchi, T., 2011. Transport phenomena of microbial flora in the small intestine with peristalsis. *Journal of theoretical biology*, *279*(1), pp.63-73.
[69] Patyar, S., Joshi, R., Byrav, D.P., Prakash, A., Medhi, B. and Das, B.K., 2010. Bacteria in cancer therapy: a novel experimental strategy. Journal of biomedical science, 17, pp.1-9.
[70] Toley, B.J. and Forbes, N.S., 2012. Motility is critical for effective distribution and accumulation of bacteria in tumor tissue. Integrative Biology, 4(2), pp.165-176.
[71] Song, W.F., Zheng, D., Zeng, S.M., Zeng, X. and Zhang, X.Z., 2022. Targeting to tumor-harbored bacteria for precision tumor therapy. ACS nano, 16(10), pp.17402-17413.
[72] Charbeneau, R.J., 2006. Groundwater hydraulics and pollutant transport. Waveland Press.

[73] Domenico, P.A. and Schwartz, F.W., 1997. Physical and chemical hydrogeology. John wiley & sons.
[74] Fetter, C.W., Boving, T. and Kreamer, D., 2017. Contaminant hydrogeology. Waveland Press.
[75] Zheng, C. and Bennett, G.D., 2002. Applied contaminant transport modeling (Vol. 2, p. 353). New York: Wiley-Interscience.
[76] Freeze, R.A. and Cherry, J.A., 1979. Groundwater prentice-hall. Englewood Cliffs, NJ, 176, pp.161-177.
[77] Hammond, G.E. and Lichtner, P.C., 2010. Field-scale model for the natural attenuation of uranium at the Hanford 300 Area using high-performance computing. Water Resources Research, 46(9).
[78] Shuyu, S.U.N. and Wheeler, M.F., 2004. Mesh adaptation strategies for discontinuous Galerkin methods applied to reactive transport problems.
[79] Hoffmann, J., Kräutle, S. and Knabner, P., 2012. A general reduction scheme for reactive transport in porous media. Computational Geosciences, 16, pp.1081-1099.
[80] Yeh, G. T., & Tripathi, V. S. (1989). A critical evaluation of recent developments in hydrogeochemical transport models of reactive multichemical components. Water Resources Research, 25(1), 93-108.
[81] Carrayrou, J., Hoffmann, J., Knabner, P., Kräutle, S., De Dieuleveult, C., Erhel, J., Van der Lee, J., Lagneau, V., Mayer, K.U. and MacQuarrie, K.T., 2010. Comparison of numerical methods for simulating strongly nonlinear and heterogeneous reactive transport problems—the MoMaS benchmark case. Computational Geosciences, 14, pp.483-502.
[82] Mo, S., Zhu, Y., Zabaras, N., Shi, X. and Wu, J., 2019. Deep convolutional encoder-decoder networks for uncertainty quantification of dynamic multiphase flow in heterogeneous media. Water Resources Research, 55(1), pp.703-728.
[83] Molins, S. and Mayer, K.U., 2007. Coupling between geochemical reactions and multicomponent gas and solute transport in unsaturated media: A reactive transport modeling study. Water Resources Research, 43(5).
[84] Leal, A.M., Blunt, M.J. and LaForce, T.C., 2014. Efficient chemical equilibrium calculations for geochemical speciation and reactive transport modelling. Geochimica et Cosmochimica Acta, 131, pp.301-322.
[85] Saaltink, M.W., Carrera, J. and Ayora, C., 2001. On the behavior of approaches to simulate reactive transport. Journal of Contaminant Hydrology, 48(3-4), pp.213-235.
[86] Sin, I. and Corvisier, J., 2019. Multiphase multicomponent reactive transport and flow modeling. Reviews in Mineralogy and Geochemistry, 85(1), pp.143-195.
[87]. Lichtner, P. C., Hammond, G. E., Lu, C., Karra, S., Bisht, G., Andre, B., ... & Kumar, J. (2015). PFLOTRAN user manual: A massively parallel reactive flow and transport model for describing surface and subsurface processes. Technical Report (No. LA-UR-15-20403). Los Alamos National Laboratory.
[88] White, M. D., & Oostrom, M. (2006). STOMP subsurface transport over multiple phases version 4.0 user's guide. Pacific Northwest National Laboratory, PNNL-15782.
[89] Steefel, C.I. and Molins, S., 2009. CrunchFlow. Software for modeling multicomponent reactive flow and transport. User's manual, pp.12-91.
[90] Xu, T., Spycher, N., Sonnenthal, E., Zhang, G., Zheng, L. and Pruess, K., 2011. TOUGHREACT Version 2.0: A simulator for subsurface reactive transport under non-isothermal multiphase flow conditions. Computers & Geosciences, 37(6), pp.763-774.
[91] Parkhurst, D.L. and Appelo, C.A.J., 2013. Description of input and examples for PHREEQC version 3—a computer program for speciation, batch-reaction, one-dimensional transport, and inverse geochemical calculations. US geological survey techniques and methods, 6(A43), p.497.
[92] Su, D., Mayer, K.U. and MacQuarrie, K.T., 2021. MIN3P-HPC: A high-performance unstructured grid code for subsurface flow and reactive transport simulation. Mathematical Geosciences, 53(4), pp.517-550.
[93] Bilke, L., Lehmann, C., Naumov, D., Fischer, T., Wang, W., Buchwald, J., Rink, K., Grunwald, N., Zill, F., Kiszkurno, F.K. and Mollali, M., 2022. OpenGeoSys. Zenodo.
[94] Jasak, H., Jemcov, A. and Tukovic, Z., 2007, September. OpenFOAM: A C++ library for complex physics simulations. In International workshop on coupled methods in numerical dynamics (Vol. 1000, pp. 1-20).
[95] Kitanidis, P. K. (1988). Prediction by the method of moments of transport in a heterogeneous formation. Journal of Hydrology, 102(1-4), 453-473.
[96] Gelhar, L. W. (1993). Stochastic Subsurface Hydrology. Prentice-Hall.
[97] Creppy, A., Praud, O., Druart, X., Kohnke, P.L. and Plouraboué, F., 2015. Turbulence of swarming sperm. Physical Review E, 92(3), p.032722.
[98] Tenchine, S., & Gouze, P. (2005). Density contrast effects on tracer dispersion in variable aperture fractures. Advances in Water Resources, 28(3), 273-289.
[99] Gelhar, L. W., Gutjahr, A. L., & Naff, R. L. (1979). Stochastic analysis of macrodispersion in a stratified aquifer. Water Resources Research, 15(6), 1387-1397.
[100] Dentz, M., Le Borgne, T., Englert, A., & Bijeljic, B. (2015). Mixing, spreading and reaction in heterogeneous media: A brief review. Journal of Contaminant Hydrology, 120-121, 1-17.

[101] Bernabé, Y., Li, M., & Maineult, A. (2016). Permeability and pore connectivity: A new model based on network simulations. Journal of Geophysical Research: Solid Earth, 121(4), 3365-3392.
[102] Valocchi, A. J. (1985). Validity of the local equilibrium assumption for modeling sorbing solute transport through homogeneous soils. Water Resources Research, 21(6), 808-820.
[103] Bijeljic, B., Mostaghimi, P., & Blunt, M. J. (2011). Signature of non-Fickian solute transport in complex heterogeneous porous media. Physical Review Letters, 107(20), 204502.
[104] Javadpour, F. (2009). Nanopores and apparent permeability of gas flow in mudrocks (shales and siltstone). Journal of Canadian Petroleum Technology, 48(8), 16-21.
[105] Sahimi, M. (1995). Flow and Transport in Porous Media and Fractured Rock. VCH.
[106] Bijeljic, B., & Blunt, M. J. (2006). Pore-scale modeling and continuous time random walk analysis of dispersion in porous media. Water Resources Research, 42(1).
[107] Benson, D. A., Wheatcraft, S. W., & Meerschaert, M. M. (1998). The fractional-order governing equation of Lévy motion. Water Resources Research, 34(6), 1461-1469.
[108] Berkowitz, B., Cortis, A., Dentz, M., & Scher, H. (2006). Modeling non-Fickian transport in geological formations as a continuous time random walk. Reviews of Geophysics, 44(2).
[109] Dentz, M., Kinzelbach, H., Attinger, S., & Kinzelbach, W. (2000). Temporal behavior of a solute cloud in a heterogeneous porous medium: 1. Point-like injection. Water Resources Research, 36(12), 3591-3604.
[110] Perez, L. J., Hidalgo, J. J., & Dentz, M. (2019). Upscaling of mixing-limited bimolecular chemical reactions in Poiseuille flow. Water Resources Research, 55(1), 249-269.
[111] Puyguiraud, A., Gouze, P., & Dentz, M. (2020). Transport upscaling in highly heterogeneous aquifers and karst systems. Transport in Porous Media, 131(1), 1-26.
[112] Gelhar, L. W., & Axness, C. L. (1983). Three-dimensional stochastic analysis of macrodispersion in aquifers. Water Resources Research, 19(1), 161-180.
[113] Dagan, G. (1984). Solute transport in heterogeneous porous formations. Journal of Fluid Mechanics, 145, 151-177.
[114] Neuman, S. P., & Tartakovsky, D. M. (2009). Perspective on theories of non-Fickian transport in heterogeneous media. Advances in Water Resources, 32(5), 670-680.
[115] Dentz, M., Cortis, A., Scher, H., & Berkowitz, B. (2004). Time behavior of solute transport in heterogeneous media: transition from anomalous to normal transport. Advances in Water Resources, 27(2), 155-173.
[116] Bear, J. (2013). Dynamics of Fluids in Porous Media. Dover Publications.
[117] Dagan, G. (2012). Flow and Transport in Porous Formations. Springer Science & Business Media.
[118] Scheidegger, A. E. (1961). General theory of dispersion in porous media. Journal of Geophysical Research, 66(10), 3273-3278.
[119] Dagan, G. (1991). Dispersion of a passive solute in non-ergodic transport by steady velocity fields in heterogeneous formations. Journal of Fluid Mechanics, 233, 197-210.
[120] Le Borgne, T., Dentz, M., & Carrera, J. (2008). Lagrangian statistical model for transport in highly heterogeneous velocity fields. Physical Review Letters, 101(9), 090601.
[121] Kang, P. K., de Anna, P., Nunes, J. P., Bijeljic, B., Blunt, M. J., & Juanes, R. (2014). Pore-scale intermittent velocity structure underpinning anomalous transport through 3-D porous media. Geophysical Research Letters, 41(17), 6184-6190.
[122] Renner, J. (2017). Predicting transport in fractured and porous rock: Challenges and adventures. Procedia Earth and Planetary Science, 17, 57-60.
[123] Nicolaides, C., Jha, B., Cueto-Felgueroso, L., & Juanes, R. (2010). Impact of viscous fingering and permeability heterogeneity on fluid mixing in porous media. Water Resources Research, 46(9).
[124] Zhang, D., & Neuman, S. P. (1990). A quasi-linear theory of non-Fickian and Fickian subsurface dispersion: 2. Application to anisotropic media and the Borden site. Water Resources Research, 26(5), 903-913.
[125] Zegadło, K., Gieroń, M., Żarnowiec, P., Durlik-Popińska, K., Kręcisz, B., Kaca, W. and Czerwonka, G., 2023. Bacterial motility and its role in skin and wound infections. International Journal of Molecular Sciences, 24(2), p.1707.
[126] Nakamura, S. and Minamino, T., 2019. Flagella-driven motility of bacteria. Biomolecules, 9(7), p.279.
[127] Terashima, H., Kojima, S. and Homma, M., 2008. Flagellar motility in bacteria: structure and function of flagellar motor. International review of cell and molecular biology, 270, pp.39-85.
[128] Bente, K., Mohammadinejad, S., Charsooghi, M.A., Bachmann, F., Codutti, A., Lefèvre, C.T., Klumpp, S. and Faivre, D., 2020. High-speed motility originates from cooperatively pushing and pulling flagella bundles in bilophotrichous bacteria. Elife, 9, p.e47551.
[129] Leifson, E. and Hugh, R., 1953. Variation in shape and arrangement of bacterial flagella. Journal of Bacteriology, 65(3), pp.263-271.
[130] Spohn, G. and Scarlato, V., 2001. Motility, chemotaxis, and flagella. Helicobacter pylori: Physiology and Genetics, pp.239-248.

[131] Colin, R., Ni, B., Laganenka, L. and Sourjik, V., 2021. Multiple functions of flagellar motility and chemotaxis in bacterial physiology. FEMS microbiology reviews, 45(6), p.fuab038.
[132] Houry, A., Gohar, M., Deschamps, J., Tischenko, E., Aymerich, S., Gruss, A. and Briandet, R., 2012. Bacterial swimmers that infiltrate and take over the biofilm matrix. Proceedings of the National Academy of Sciences, 109(32), pp.13088-13093.
[133] Ushijima, B. and Häse, C.C., 2018. Influence of chemotaxis and swimming patterns on the virulence of the coral pathogen Vibrio coralliilyticus. Journal of bacteriology, 200(15), pp.10-1128.
[134] Mattick, J.S., 2002. Type IV pili and twitching motility. Annual Reviews in Microbiology, 56(1), pp.289-314.
[135] Jayathilake, P.G., Li, B., Zuliani, P., Curtis, T. and Chen, J., 2019. Modelling bacterial twitching in fluid flows: a CFD-DEM approach. Scientific reports, 9(1), p.14540.
[136] Zolfaghar, I., Evans, D.J. and Fleiszig, S.M., 2003. Twitching motility contributes to the role of pili in corneal infection caused by Pseudomonas aeruginosa. Infection and immunity, 71(9), pp.5389-5393.
[137] Biais, N., Ladoux, B., Higashi, D., So, M. and Sheetz, M., 2008. Cooperative retraction of bundled type IV pili enables nanonewton force generation. PLoS biology, 6(4), p.e87.
[138] Spormann, A.M., 1999. Gliding motility in bacteria: insights from studies of Myxococcus xanthus. Microbiology and Molecular Biology Reviews, 63(3), pp.621-641.
[139] McBride, M.J., 2001. Bacterial gliding motility: multiple mechanisms for cell movement over surfaces. Annual Reviews in Microbiology, 55(1), pp.49-75.
[140] Tchoufag, J., Ghosh, P., Pogue, C.B., Nan, B. and Mandadapu, K.K., 2019. Mechanisms for bacterial gliding motility on soft substrates. Proceedings of the National Academy of Sciences, 116(50), pp.25087-25096.
[141] Harshey, R.M., 2003. Bacterial motility on a surface: many ways to a common goal. Annual Reviews in Microbiology, 57(1), pp.249-273.
[142] Hölscher, T. and Kovács, Á.T., 2017. Sliding on the surface: bacterial spreading without an active motor. Environmental Microbiology, 19(7), pp.2537-2545.
[143] Kearns, D.B., 2010. A field guide to bacterial swarming motility. Nature reviews microbiology, 8(9), pp.634-644.
[144] Fraser, G.M. and Hughes, C., 1999. Swarming motility. Current opinion in microbiology, 2(6), pp.630-635.
[145] Lai, S., Tremblay, J. and Déziel, E., 2009. Swarming motility: a multicellular behaviour conferring antimicrobial resistance. Environmental microbiology, 11(1), pp.126-136.
[146] Duffy, K.J., Cummings, P.T. and Ford, R.M., 1995. Random walk calculations for bacterial migration in porous media. Biophysical journal, 68(3), pp.800-806.
[147] Rusconi, R., Guasto, J.S. and Stocker, R., 2014. Bacterial transport suppressed by fluid shear. Nature physics, 10(3), pp.212-217.
[148] Bradford, S.A., Morales, V.L., Zhang, W., Harvey, R.W., Packman, A.I., Mohanram, A. and Welty, C., 2013. Transport and fate of microbial pathogens in agricultural settings. Critical reviews in environmental science and technology, 43(8), pp.775-893.
[149] Bradford, S.A., Wang, Y., Kim, H., Torkzaban, S. and Šimůnek, J., 2014. Modeling microorganism transport and survival in the subsurface. Journal of environmental quality, 43(2), pp.421-440.
[150] Kim, Y.C., 1996. Diffusivity of bacteria. Korean Journal of Chemical Engineering, 13, pp.282-287.
[151] Sibley, L.D., Håkansson, S. and Carruthers, V.B., 1998. Gliding motility: an efficient mechanism for cell penetration. Current Biology, 8(1), pp.R12-R14.
[152] Harshey, R.M., 2003. Bacterial motility on a surface: many ways to a common goal. Annual Reviews in Microbiology, 57(1), pp.249-273.
[153] Kaya, T. and Koser, H., 2012. Direct upstream motility in Escherichia coli. Biophysical journal, 102(7), pp.1514-1523.
[154] Niwano, M. and Taylor, B.L., 1982. Novel sensory adaptation mechanism in bacterial chemotaxis to oxygen and phosphotransferase substrates. Proceedings of the National Academy of Sciences, 79(1), pp.11-15.
[155] Ford, R.M. and Lauffenburger, D.A., 1991. Measurement of bacterial random motility and chemotaxis coefficients: II. Application of single-cell-based mathematical model. Biotechnology and bioengineering, 37(7), pp.661-672.
[156] Lewus, P. and Ford, R.M., 2001. Quantification of random motility and chemotaxis bacterial transport coefficients using individual-cell and population-scale assays. Biotechnology and bioengineering, 75(3), pp.292-304.
[157] Pankratova, E.V., Kalyakulina, A.I., Krivonosov, M.I., Denisov, S.V., Taute, K.M. and Zaburdaev, V.Y., 2018. Chemotactic drift speed for bacterial motility pattern with two alternating turning events. PloS one, 13(1), p.e0190434.
[158] Tuson, H.H. and Weibel, D.B., 2013. Bacteria–surface interactions. Soft matter, 9(17), pp.4368-4380.
[159] Wioland, H., Lushi, E. and Goldstein, R.E., 2016. Directed collective motion of bacteria under channel confinement. New Journal of Physics, 18(7), p.075002.
[160] Wu, Y., 2015. Collective motion of bacteria in two dimensions. Quantitative Biology, 3, pp.199-205.
[161] Kotte, O., Zaugg, J.B. and Heinemann, M., 2010. Bacterial adaptation through distributed sensing of metabolic fluxes. Molecular systems biology, 6(1), p.355.

[162] Ramachandran, A., Stone, H.A. and Gitai, Z., 2024. Free-swimming bacteria transcriptionally respond to shear flow. Proceedings of the National Academy of Sciences, 121(42), p.e2406688121.
[163] Wheeler, J.D., Secchi, E., Rusconi, R. and Stocker, R., 2019. Not just going with the flow: the effects of fluid flow on bacteria and plankton. Annual review of cell and developmental biology, 35(1), pp.213-237.
[164] Risken, H. (1996) The Fokker-Planck Equation: Methods of Solution and Applications. 2nd Edition, Springer, New York, 1-7, 56-59.
[165] Puyguiraud, A., Perez, L.J., Hidalgo, J.J. and Dentz, M., 2020. Effective dispersion coefficients for the upscaling of pore-scale mixing and reaction. Advances in Water Resources, 146, p.103782.
[166] Fier, G., Hansmann, D. and Buceta, R.C., 2018. Langevin equations for the run-and-tumble of swimming bacteria. Soft Matter, 14(19), pp.3945-3954.
[167] Keller, E.F. and Segel, L.A., 1971. Model for chemotaxis. Journal of theoretical biology, 30(2), pp.225-234.
[168] Taktikos, J., 2013. Modeling the random walk and chemotaxis of bacteria: Aspects of biofilm formation.
[169] Mosaddad, S.A., Tahmasebi, E., Yazdanian, A., Rezvani, M.B., Seifalian, A., Yazdanian, M. and Tebyanian, H., 2019. Oral microbial biofilms: an update. European Journal of Clinical Microbiology & Infectious Diseases, 38, pp.2005-2019.
[170] O'Brien, C.E., Giovannelli, D., Govenar, B., Luther, G.W., Lutz, R.A., Shank, T.M. and Vetriani, C., 2015. Microbial biofilms associated with fluid chemistry and megafaunal colonization at post-eruptive deep-sea hydrothermal vents. Deep Sea Research Part II: Topical Studies in Oceanography, 121, pp.31-40.
[171] Tolker-Nielsen, T., 2015. Biofilm development. Microbial Biofilms, pp.51-66.
[172] O'Toole, G. A., & Kolter, R. (1998). Flagellar and twitching motility are necessary for Pseudomonas aeruginosa biofilm development. Molecular Microbiology, 30(2), 295-304.
[173] Flemming, H.C. and Wingender, J., 2010. The biofilm matrix. Nature reviews microbiology, 8(9), pp.623-633.
[174] Fish, K.E., Osborn, A.M. and Boxall, J., 2016. Characterising and understanding the impact of microbial biofilms and the extracellular polymeric substance (EPS) matrix in drinking water distribution systems. Environmental science: water research & technology, 2(4), pp.614-630.
[175] Ding, Z., Bourven, I., Guibaud, G., van Hullebusch, E.D., Panico, A., Pirozzi, F. and Esposito, G., 2015. Role of extracellular polymeric substances (EPS) production in bioaggregation: application to wastewater treatment. Applied microbiology and biotechnology, 99, pp.9883-9905.
[176] Karimi, A., Karig, D., Kumar, A. and Ardekani, A.M., 2015. Interplay of physical mechanisms and biofilm processes: review of microfluidic methods. Lab on a Chip, 15(1), pp.23-42.
[177] Melo, A., Quintelas, C., Ferreira, E.C. and Mesquita, D.P., 2022. The role of extracellular polymeric substances in micropollutant removal. Frontiers in Chemical Engineering, 4, p.778469.
[178] Liu, H.Y., Prentice, E.L. and Webber, M.A., 2024. Mechanisms of antimicrobial resistance in biofilms. npj Antimicrobials and Resistance, 2(1), p.27.
[179] Pan, M., Li, H., Han, X., Ma, W., Li, X., Guo, Q., Yang, B., Ding, C. and Ma, Y., 2022. Effects of hydrodynamic conditions on the composition, spatiotemporal distribution of different extracellular polymeric substances and the architecture of biofilms. Chemosphere, 307, p.135965.
[180] Tallawi, M., Opitz, M. and Lieleg, O., 2017. Modulation of the mechanical properties of bacterial biofilms in response to environmental challenges. Biomaterials science, 5(5), pp.887-900.
[181] Jo, J., Price-Whelan, A. and Dietrich, L.E., 2022. Gradients and consequences of heterogeneity in biofilms. Nature Reviews Microbiology, 20(10), pp.593-607.
[182] Solano, C., Echeverz, M. and Lasa, I., 2014. Biofilm dispersion and quorum sensing. Current opinion in microbiology, 18, pp.96-104.
[183] Miller, M.B. and Bassler, B.L., 2001. Quorum sensing in bacteria. Annual Reviews in Microbiology, 55(1), pp.165-199.
[184] Crabbé, A., Jensen, P.Ø., Bjarnsholt, T. and Coenye, T., 2019. Antimicrobial tolerance and metabolic adaptations in microbial biofilms. Trends in microbiology, 27(10), pp.850-863.
[185] Keren, I., Kaldalu, N., Spoering, A., Wang, Y. and Lewis, K., 2004. Persister cells and tolerance to antimicrobials. FEMS microbiology letters, 230(1), pp.13-18.
[186] Lewis, K., 2005. Persister cells and the riddle of biofilm survival. Biochemistry (Moscow), 70, pp.267-274.
[187] Wilson, M., 2001. Bacterial biofilms and human disease. Science progress, 84(3), pp.235-254.
[188] Torelli, R., Cacaci, M., Papi, M., Sterbini, F.P., Martini, C., Posteraro, B., Palmieri, V., De Spirito, M., Sanguinetti, M. and Bugli, F., 2017. Different effects of matrix degrading enzymes towards biofilms formed by E. faecalis and E. faecium clinical isolates. Colloids and Surfaces B: Biointerfaces, 158, pp.349-355.
[189] Kaplan, J.B., 2014. Biofilm matrix-degrading enzymes. Microbial Biofilms: Methods and Protocols, pp.203-213.
[190] Brackman, G. and Coenye, T., 2015. Quorum sensing inhibitors as anti-biofilm agents. Current pharmaceutical design, 21(1), pp.5-11.
[191] Bazaka, K., Jacob, M.V., Crawford, R.J. and Ivanova, E.P., 2012. Efficient surface modification of biomaterial to prevent biofilm formation and the attachment of microorganisms. Applied microbiology and biotechnology, 95, pp.299-311.

[192] Bottero, S., Storck, T., Heimovaara, T.J., van Loosdrecht, M.C., Enzien, M.V. and Picioreanu, C., 2013. Biofilm development and the dynamics of preferential flow paths in porous media. Biofouling, 29(9), pp.1069-1086.
[193] Kurz, D.L., Secchi, E., Carrillo, F.J., Bourg, I.C., Stocker, R. and Jimenez-Martinez, J., 2022. Competition between growth and shear stress drives intermittency in preferential flow paths in porous medium biofilms. Proceedings of the National Academy of Sciences, 119(30), p.e2122202119.
[194] Conrad, J.C. and Poling-Skutvik, R., 2018. Confined flow: consequences and implications for bacteria and biofilms. Annual review of chemical and biomolecular engineering, 9(1), pp.175-200.
[195] Fabbri, S. and Stoodley, P., 2016. Mechanical properties of biofilms. The perfect slime—microbial extracellular polymeric substances, pp.153-178.
[196] Wei, G. and Yang, J.Q., 2023. Impacts of hydrodynamic conditions and microscale surface roughness on the critical shear stress to develop and thickness of early-stage Pseudomonas putida biofilms. Biotechnology and bioengineering, 120(7), pp.1797-1808.
[197] Chandy, J.P. and Angles, M.L., 2001. Determination of nutrients limiting biofilm formation and the subsequent impact on disinfectant decay. Water Research, 35(11), pp.2677-2682.
[198] Rinaudi, L., Fujishige, N.A., Hirsch, A.M., Banchio, E., Zorreguieta, A. and Giordano, W., 2006. Effects of nutritional and environmental conditions on Sinorhizobium meliloti biofilm formation. Research in microbiology, 157(9), pp.867-875.
[199] Xian, Y., Jin, M., Zhan, H. and Liang, X., 2022. Permeable biofilms can support persistent hyporheic anoxic microzones. Geophysical Research Letters, 49(14), p.e2021GL096948.
[200] Krause, S., Lewandowski, J., Grimm, N.B., Hannah, D.M., Pinay, G., McDonald, K., Martí, E., Argerich, A., Pfister, L., Klaus, J. and Battin, T., 2024. Ecohydrological interfaces as hotspots of ecosystem processes. Ecohydrological Interfaces, pp.1-28.
[201] Li, Y., Liu, M. and Wu, X., 2022. Reclaimed water reuse for groundwater recharge: A review of hot spots and hot moments in the hyporheic zone. Water, 14(12), p.1936.
[202] Rosenzweig, R., Furman, A., Dosoretz, C. and Shavit, U., 2014. Modeling biofilm dynamics and hydraulic properties in variably saturated soils using a channel network model. Water Resources Research, 50(7), pp.5678-5697.
[203] Yan, Z., Liu, C., Liu, Y. and Bailey, V.L., 2017. Multiscale investigation on biofilm distribution and its impact on macroscopic biogeochemical reaction rates. Water Resources Research, 53(11), pp.8698-8714.
[204] Dennis, M.L. and Turner, J.P., 1998. Hydraulic conductivity of compacted soil treated with biofilm. Journal of Geotechnical and Geoenvironmental Engineering, 124(2), pp.120-127.
[205] Jasu, A. and Ray, R.R., 2021. Biofilm mediated strategies to mitigate heavy metal pollution: A critical review in metal bioremediation. Biocatalysis and Agricultural Biotechnology, 37, p.102183.
[206] Mishra, S., Huang, Y., Li, J., Wu, X., Zhou, Z., Lei, Q., Bhatt, P. and Chen, S., 2022. Biofilm-mediated bioremediation is a powerful tool for the removal of environmental pollutants. Chemosphere, 294, p.133609.
[207] Wanner, O. and Gujer, W., 1986. A multispecies biofilm model. Biotechnology and bioengineering, 28(3), pp.314-328.
[208] Wang, Q. and Zhang, T., 2010. Review of mathematical models for biofilms. Solid State Communications, 150(21-22), pp.1009-1022.
[209] D'Acunto, B., Frunzo, L., Klapper, I., Mattei, M.R. and Stoodley, P., 2019. Mathematical modeling of dispersal phenomenon in biofilms. Mathematical biosciences, 307, pp.70-87.
[210] Orghidan, T., 1959. Ein neuer Lebensraum des unterirdischen Wassers: der hyporheische Biotop. *Arch. Hydrobiol*, *55*(3), pp.392-414.
[211] Stanford, J. A., & Ward, J. V. (1993). An ecosystem perspective of alluvial rivers: Connectivity and the hyporheic corridor. Journal of the North American Benthological Society, 12(1), 48-60.
[212] Findlay, S. (1995). Importance of surface-subsurface exchange in stream ecosystems: The hyporheic zone. Limnology and Oceanography, 40(1), 159-164.
[213] Triska, F. J., Duff, J. H., & Avanzino, R. J. (1993). The role of water exchange between a stream channel and its hyporheic zone in nitrogen cycling at the terrestrial-aquatic interface. Hydrobiologia, 251(1), 167-184.
[214] Boulton, A. J., Findlay, S., Marmonier, P., Stanley, E. H., & Valett, H. M. (1998). The functional significance of the hyporheic zone in streams and rivers. Annual Review of Ecology and Systematics, 29(1), 59-81.
[215] Battin, T. J., Kaplan, L. A., Newbold, J. D., & Hansen, C. M. (2003). Contributions of microbial biofilms to ecosystem processes in stream mesocosms. Nature, 426(6965), 439-442.
[216] McClain, M. E., Boyer, E. W., Dent, C. L., Gergel, S. E., Grimm, N. B., Groffman, P. M., ... & Pinay, G. (2003). Biogeochemical hot spots and hot moments at the interface of terrestrial and aquatic ecosystems. Ecosystems, 6(4), 301-312.
[217] Rittmann, B. E., & McCarty, P. L. (2001). Environmental Biotechnology: Principles and Applications. McGraw-Hill.
[218] Schindler, J.E. and Krabbenhoft, D.P., 1998. The hyporheic zone as a source of dissolved organic carbon and carbon gases to a temperate forested stream. Biogeochemistry, 43, pp.157-174.

[219] Michaelis, T., Wunderlich, A., Coskun, Ö.K., Orsi, W., Baumann, T. and Einsiedl, F., 2022. High-resolution vertical biogeochemical profiles in the hyporheic zone reveal insights into microbial methane cycling. Biogeosciences, 19(18), pp.4551-4569.
[220] Cardenas, M. B., & Wilson, J. L. (2007). Dunes, turbulent eddies, and interfacial exchange with permeable sediments. Water Resources Research, 43(8), W08412.
[221] Flemming, H. C., & Wingender, J. (2010). The biofilm matrix. Nature Reviews Microbiology, 8(9), 623-633.
[222] Flemming, H. C., Wingender, J., Szewzyk, U., Steinberg, P., Rice, S. A., & Kjelleberg, S. (2016). Biofilms: An emergent form of bacterial life. Nature Reviews Microbiology, 14(9), 563-575.
[223] Febria, C. M., Beddoes, P., Fulthorpe, R. R., & Williams, D. D. (2012). Bacterial community dynamics in the hyporheic zone of an intermittent stream. The ISME Journal, 6(5), 1078-1088.
[224] Bardini, M., Boano, F., Cardenas, M. B., Revelli, R., & Ridolfi, L. (2012). Nutrient cycling in bedform induced hyporheic zones. Geochimica et Cosmochimica Acta, 84, 47-61.
[225] Ratkowsky, D. A., Olley, J., McMeekin, T. A., & Ball, A. (1982). Relationship between temperature and growth rate of bacterial cultures. Journal of Bacteriology, 149(1), 1-5.
[226] Morita, R. Y. (1975). Psychrophilic bacteria. Bacteriological Reviews, 39(2), 144-167.
[227] Schoolfield, R. M., Sharpe, P. J., & Magnuson, C. E. (1981). Non-linear regression of biological temperature-dependent rate models based on absolute reaction-rate theory. Journal of Theoretical Biology, 88(4), 719-731.
[228] Neidhardt, F. C., Ingraham, J. L., & Schaechter, M. (1990). Physiology of the bacterial cell: A molecular approach. Sinauer Associates.
[229] Beveridge, T. J., Lawrence, J. R., & Murray, R. G. E. (2010). Bacterial Structure and Function: A Guide to Morphology and Behavior. Cambridge University Press.
[230] Horikoshi, K. (1999). Alkaliphiles: Some applications of their products for biotechnology. Microbiology and Molecular Biology Reviews, 63(4), 735-750.
[231] Booth, I. R. (1985). Regulation of cytoplasmic pH in bacteria. Microbiological Reviews, 49(4), 359-378.
[232] Krulwich, T. A., Sachs, G., & Padan, E. (2011). Molecular aspects of bacterial pH sensing and homeostasis. Nature Reviews Microbiology, 9(5), 330-343.
[233] Monod, J. (1949). The growth of bacterial cultures. Annual Review of Microbiology, 3(1), 371-394.
[234] Ferenci, T. (1999). 'Growth of bacterial cultures' 50 years on: towards an uncertainty principle instead of constants in bacterial growth kinetics. Research in Microbiology, 150(7), 431-438.
[235] Button, D. K. (1985). Kinetics of nutrient-limited transport and microbial growth. Microbiological Reviews, 49(3), 270-297.
[236] Brock, T. D., Madigan, M. T., Martinko, J. M., & Parker, J. (2003). Biology of microorganisms. Prentice Hall.
[237] Thauer, R. K., Jungermann, K., & Decker, K. (1977). Energy conservation in chemotrophic anaerobic bacteria. Bacteriological Reviews, 41(1), 100-180.
[238] Richardson, D. J. (2000). Bacterial respiration: a flexible process for a changing environment. Microbiology, 146(3), 551-571.
[239] Stolper, D. A., Revsbech, N. P., & Canfield, D. E. (2010). Aerobic growth at nanomolar oxygen concentrations. Proceedings of the National Academy of Sciences, 107(44), 18755-18760.
[240] Nies, D. H. (1999). Microbial heavy-metal resistance. Applied Microbiology and Biotechnology, 51(6), 730-750.
[241] Roane, T. M., & Pepper, I. L. (2000). Microorganisms and metal pollutants. Environmental Microbiology, 403-423.
[242] Little, A. E., Robinson, C. J., Peterson, S. B., Raffa, K. F., & Handelsman, J. (2008). Rules of engagement: Interspecies interactions that regulate microbial communities. Annual Review of Microbiology, 62, 375-401.
[243] Hibbing, M. E., Fuqua, C., Parsek, M. R., & Peterson, S. B. (2010). Bacterial competition: Surviving and thriving in the microbial jungle. Nature Reviews Microbiology, 8(1), 15-25.
[244] Zeglin, L. H. (2015). Stream microbial diversity in response to environmental changes: Review and synthesis of existing research. Frontiers in Microbiology, 6, 454.
[245] Datry, T., Larned, S. T., & Tockner, K. (2014). Intermittent rivers: A challenge for freshwater ecology. BioScience, 64(3), 229-235.
[246] Zeglin, L. H., Dahm, C. N., Barrett, J. E., Gooseff, M. N., Fitpatrick, S. K., & Takacs-Vesbach, C. D. (2011). Bacterial community structure along moisture gradients in the parafluvial sediments of two ephemeral desert streams. Microbial Ecology, 61(3), 543-556.
[247] Gómez, R., Arce, M. I., Sánchez, J. J., & del Mar Sánchez-Montoya, M. (2012). The effects of drying on sediment nitrogen content in a Mediterranean intermittent stream: A microcosms study. Hydrobiologia, 679(1), 43-59.
[248] Doretto, A., Piano, E. and Larson, C.E., 2020. The River Continuum Concept: lessons from the past and perspectives for the future. *Canadian Journal of Fisheries and Aquatic Sciences*, *77*(11), pp.1853-1864.
[249] Merbt, S. N., Auguet, J. C., Casamayor, E. O., & Martí, E. (2016). Biofilm recovery in a wastewater treatment plant-influenced stream and spatial patterns of community composition. Frontiers in Microbiology, 7, 1772.

[250] Crocker, J. C., & Grier, D. G. (1996). Methods of digital video microscopy for colloidal studies. Journal of Colloid and Interface Science, 179(1), 298-310.
[251] Saxton, M. J., & Jacobson, K. (1997). Single-particle tracking: applications to membrane dynamics. Annual Review of Biophysics and Biomolecular Structure, 26(1), 373-399.
[252] Jaqaman, K., Loerke, D., Mettlen, M., Kuwata, H., Grinstein, S., Schmid, S. L., & Danuser, G. (2008). Robust single-particle tracking in live-cell time-lapse sequences. Nature Methods, 5(8), 695-702.
[253] Betzig, E., Patterson, G. H., Sougrat, R., Lindwasser, O. W., Olenych, S., Bonifacino, J. S., ... & Hess, H. F. (2006). Imaging intracellular fluorescent proteins at nanometer resolution. Science, 313(5793), 1642-1645.
[254] Smal, I., Draegestein, K., Galjart, N., Niessen, W., & Meijering, E. (2008). Particle filtering for multiple object tracking in dynamic fluorescence microscopy images: Application to microtubule growth analysis. IEEE Transactions on Medical Imaging, 27(6), 789-804.
[255] Chenouard, N., Smal, I., De Chaumont, F., Maška, M., Sbalzarini, I. F., Gong, Y., ... & Meijering, E. (2014). Objective comparison of particle tracking methods. Nature Methods, 11(3), 281-289.
[256] Tinevez, J. Y., Perry, N., Schindelin, J., Hoopes, G. M., Reynolds, G. D., Laplantine, E., ... & Eliceiri, K. W. (2017). TrackMate: An open and extensible platform for single-particle tracking. Methods, 115, 80-90.
[257] Ershov, D., Phan, M.S., Pylvänäinen, J.W., Rigaud, S.U., Le Blanc, L., Charles-Orszag, A., Conway, J.R., Laine, R.F., Roy, N.H., Bonazzi, D. and Duménil, G., 2021. Bringing TrackMate into the era of machine-learning and deep-learning. *BioRxiv*, pp.2021-09.
[258] Amat, F., Lemon, W., Mossing, D. P., McDole, K., Wan, Y., Branson, K., ... & Keller, P. J. (2014). Fast, accurate reconstruction of cell lineages from large-scale fluorescence microscopy data. Nature Methods, 11(9), 951-958.
[259] Bhattacharjee, T., Amchin, D.B., Ott, J.A., Kratz, F. and Datta, S.S., 2021. Chemotactic migration of bacteria in porous media. Biophysical Journal, 120(16), pp.3483-3497.
[260] Papadopoulos, C., 2023. A novel experimental platform for monitoring and imaging bacterial biofilm growth in porous media flows (Doctoral dissertation, Institut National Polytechnique de Toulouse-INPT).
[261] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444.
[262] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems, 25, 1097-1105.
[263] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. IEEE Conference on Computer Vision and Pattern Recognition, 779-788.
[264] He, K., Fan, H., Wu, Y., Xie, S., & Girshick, R. (2020). Momentum contrast for unsupervised visual representation learning. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 9729-9738.
[265] Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020). A simple framework for contrastive learning of visual representations. International Conference on Machine Learning, 1597-1607.
[266] Cheng, H.J., Hsu, C.H., Hung, C.L. & Lin, C.Y. (2022). A review for cell and particle tracking on microscopy images using algorithms and deep learning technologies. biomedical journal, 45(3), pp.465-471.
[267] Newby, J. M., Schaefer, A. M., Lee, P. T., Forest, M. G., & Lai, S. K. (2018). Deep neural networks automate detection for tracking of submicron-scale particles in 2D and 3D. Proceedings of the National Academy of Sciences, 115(36), 9026-9031.
[268] Yao, Y., Smal, I., Grigoriev, I., Akhmanova, A. & Meijering, E. (2020). Deep-learning method for data association in particle tracking. Bioinformatics, 36(19), pp.4935-4941.
[269] Spilger, R., Wollmann, T., Qiang, Y., Imle, A., Lee, J.Y., Müller, B., Fackler, O.T., Bartenschlager, R. & Rohr, K. (2018). Deep particle tracker: Automatic tracking of particles in fluorescence microscopy images using deep learning. In Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support (pp. 128-136).
[270] Song, D., Zhang, X., Li, B., Sun, Y., Mei, H., Cheng, X., Li, J., Cheng, X. & Fang, N. (2024). Deep Learning-Assisted Automated Multidimensional Single Particle Tracking in Living Cells. Nano Letters, 24(10), pp.3082-3088.
[271] Ritter, C., Spilger, R., Lee, J.Y., Bartenschlager, R. & Rohr, K. (2021). Deep learning for particle detection and tracking in fluorescence microscopy images. In 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI) (pp. 873-876).
[272] Moen, E., Borba, E., Miller, G., Schwartz, M., Bannon, D., Koe, N., Camplisson, I., Kyme, D., Pavelchek, C., Price, T. & Kudo, T. (2019). Accurate cell tracking and lineage construction in live-cell imaging experiments with deep learning. Biorxiv, p.803205.
[273] Lugagne, J.B., Lin, H. & Dunlop, M.J. (2020). DeLTA: Automated cell segmentation, tracking, and lineage reconstruction using deep learning. PLoS computational biology, 16(4), p.e1007673.
[274] Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. Journal of Computational Physics, 378, 686-707.

[275] Zhan, C., Dai, Z., Samper, J., Yin, S., Ershadnia, R., Zhang, X., Wang, Y., Yang, Z., Luan, X. and Soltanian, M.R., 2022. An integrated inversion framework for heterogeneous aquifer structure identification with single-sample generative adversarial network. *Journal of Hydrology*, *610*, p.127844.
[276] You, J. & Lee, K.J. (2024). Upscaling reactive transport models from pore-scale to continuum-scale using deep learning method. Geoenergy Science and Engineering, 238, p.212850.
[277] Wang, Z. & Battiato, I. (2024). A deep learning upscaling framework: Reactive transport and mineral precipitation in fracture-matrix systems. Advances in Water Resources, 183, p.104588.
[278] Leal, A.M., Kyas, S., Kulik, D.A. & Saar, M.O. (2020). Accelerating reactive transport modeling: on-demand machine learning algorithm for chemical equilibrium calculations. Transport in Porous Media, 133(2), pp.161-204.
[279] Prasianakis, N.I., Haller, R., Mahrous, M., Poonoosamy, J., Pfingsten, W. & Churakov, S.V. (2020). Neural network based process coupling and parameter upscaling in reactive transport simulations. Geochimica et Cosmochimica Acta, 291, pp.126-143.
[280] Jagtap, N.V., Mudunuru, M.K. & Nakshatrala, K.B. (2021). A deep learning modeling framework to capture mixing patterns in reactive-transport systems. arXiv preprint arXiv:2101.04227.
[281] Lu, H., Ermakova, D., Wainwright, H.M., Zheng, L. & Tartakovsky, D.M. (2021). Data-informed emulators for multi-physics simulations. Journal of Machine Learning for Modeling and Computing, 2(2).
[282] Basha, N., Arcucci, R., Angeli, P., Anastasiou, C., Abadie, T., Casas, C.Q., Chen, J., Cheng, S., Chagot, L., Galvanin, F. & Heaney, C.E. (2024). Machine learning and physics-driven modelling and simulation of multiphase systems. International Journal of Multiphase Flow, 179, p.104936.
[283] Sanchez-Gonzalez, A., Godwin, J., Pfaff, T., Ying, R., Leskovec, J., & Battaglia, P. W. (2020). Learning to simulate complex physics with graph networks. Nature Machine Intelligence, 2(11), 668-675.
[284] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30, 5998-6008.
[285] Laloy, E., Hérault, R., Jacques, D., & Linde, N. (2018). Training-image based geostatistical inversion using a spatial generative adversarial neural network. Water Resources Research, 54(1), 381-406.
[286] Yu, S., Hannah, W.M., Peng, L., Bhouri, M.A., Gupta, R., Lin, J., Lütjens, B., Will, J.C., Beucler, T., Harrop, B.E. & Hillman, B.R. (2023). ClimSim: An open large-scale dataset for training high-resolution physics emulators in hybrid multi-scale climate simulators. arXiv preprint arXiv:2306.08754.
[287] Obiols-Sales, O., Vishnu, A., Malaya, N. & Chandramowliswharan, A. (2020). CFDNet: A deep learning-based accelerator for fluid simulations. In Proceedings of the 34th ACM international conference on supercomputing (pp. 1-12).
[288] Koric, S. & Abueidda, D.W. (2021). Deep learning sequence methods in multiphysics modeling of steel solidification. Metals, 11(3), p.494.
[289] Gunawardena, N., Pallotta, G., Simpson, M. & Lucas, D.D. (2021). Machine learning emulation of spatial deposition from a multi-physics ensemble of weather and atmospheric transport models. Atmosphere, 12(8), p.953.
[290] Mo, S., Zhu, Y., Zabaras, N., Shi, X., & Wu, J. (2019). Deep convolutional encoder-decoder networks for uncertainty quantification of dynamic multiphase flow in heterogeneous media. Water Resources Research, 55(1), 703-728.
[291] Yabusaki, S.B., Fang, Y., Williams, K.H., Murray, C.J., Ward, A.L., Dayvault, R.D., Waichler, S.R., Newcomer, D.R., Spane, F.A. & Long, P.E. Variably saturated flow and multicomponent biogeochemical reactive transport modeling of a uranium bioremediation field experiment. Journal of contaminant hydrology, 126(3-4), 271-290 (2011).

Chapter 2: Advection-Dominated Transport Dynamics of Pili and Flagella-Mediated Motile Bacteria in Porous Media

2.1 Abstract

The transport of motile bacteria in porous media is highly relevant to many fields, ranging from ecology to human health. Still, critical gaps remain in our understanding of the impacts of motility, hydrodynamics, and pore structure on bacterial transport. Here, we present direct visualizations of three species of motile bacteria under variable flow rates and porosities. We find that at higher flow rates, motility is less critical to the transport of bacteria, as motion is controlled by hydrodynamic advection, making it difficult for bacteria to move across streamlines. We show that this lack of motion across streamlines results in increased velocity autocorrelation and bacterial spreading in the direction of flow. Furthermore, we find that transport of bacteria with different motility types are impacted by flow rates to different extents. At low flow rates, the transport of bacteria with pili-mediated twitching motility is strongly controlled by advection, whereas bacteria with flagella still display active motility. At higher flow rates, we show that bacteria with peritrichous flagella maintain their motility characteristics to a greater degree than bacteria with pili or monotrichous flagella. We also examine experimental net speeds of bacteria in relation to the simulated flow fields and find that the interactions between hydrodynamics, motility, and porous media geometry lead to oversampling of medium-velocity regions of a pore network by all three species. The study presents new perspectives on how different types of motile bacteria are transported and dispersed in porous media aided by strength of differentially advecting fluid.

2.2. Introduction

Motile bacteria often live in dynamic flow environments, and their migration involves complex self-propulsion strategies that are relevant to human health and ecology [1–3]. Navigating confined spaces of a pore network, motile bacteria employ diverse movement modalities (e.g., turn angles, run-and-tumble, or run-and-flick) that characterize their migration [4–9]. In porous media, the degree of confinement (i.e., porosity) and speed of the fluid flow strongly affect bacterial migration and modulate their interactions with the surrounding environment. This, in turn, has a broad range of effects on bacteria, such as altering their movement [10], behavior [11], resource acquisition [12], and signaling [13], thereby influencing their metabolic functions, spatial distribution, and diversity. Spatial variations in flow velocities and the related changes in shear add another level of complexity to the transport of bacteria. Transverse movement of bacteria from low-shear to high-shear regions located near surfaces has been shown to result in the accumulation of cells in low-velocity regions [14, 15]. This phenomenon, termed shear trapping, has been identified as one major mechanism that drives initial colonization of curved surfaces and microfluidic pore channels [16–18], leading to the formation of suspended biofilm structures [19, 20]. While these observations have led to improvements in our understanding of bacterial transport in idealized systems, there are significant gaps in our ability to quantify and predict transport behavior under complex conditions, such as in pore networks designed to produce the hydrodynamics of natural porous media. Understanding the motile behavior of bacteria in confined environments, in which they search for available physical space and move in response to fluid flow, has implications for a wide range of applications, such as bioremediation [21], biofilm formation [22, 23], and anticancer drug delivery [24].

In this chapter, we report direct bacterial transport visualizations, at single-cell resolution, of three different species of motile bacteria under variable flow conditions in a quasi-2D porous media with different levels of pore confinement. Recent research conducted in microfluidic porous chips has shown that while transport of non-motile bacteria is compact giving rise to a Gaussian distribution of traveled distances (i.e., follows streamlines with negligible retardation due to mass exchanges between fast and slow moving zones), the distribution of motile strains in the pore spaces show both active retention and enhanced dispersion due to exchanges between fast flow channels and low velocity regions closer to the grain surfaces [17, 25]. The presented work here examines the transport dynamics of three motile species with an overarching goal of highlighting key statistical differences in various transport metrics so that more informed modeling approaches can be developed for upscaled transport simulations. Having an improved understanding of dispersion rates and of the key factors that control dispersion, such as the velocity and turn angle distributions, would provide robust ingredients for development of random walk based approaches [25]. We focus on investigating *Acidovorax* strain JHL-9 [26], *Geobacter sulfurreducens* [27], and *Paenibacillus* strain 300A [28] due to their common attribute of metal-reducing capabilities. By studying these specific microorganisms, our research findings are especially relevant to bioremediation and biogeochemical cycling in terrestrial environments [29-31]. Furthermore, our selection of these species is based on their distinct modes of motility. Using their pili to attach to surfaces and pull themselves towards new locations [32], *Geobacter sulfurreducens* exhibit twitching motility [33]. *Paenibacillus* 300A exhibit swimming motility, presumably driven by peritrichous flagella [34]. *Acidovorax* JLH-9 [26] exhibit twitching motility, consistent with genomic analysis of the strain, though transmission electron microscopy (TEM) images of the strain suggest the presence of polar/monotrichous flagella, indicating the possibility

of swimming motility as well. Swimmers generally move much faster than twitchers [32, 35-37], providing a reasonable basis in this study to compare the two different motility types at different flow rates. The primary focus of this dissertation chapter is not to decipher the fundamental reasons for differences in the transport behaviors of the three selected species, but rather to evaluate transport characteristics of bioremediation relevant species with different motility types in order to help lay a framework for species-aware upscaling and macroscale transport simulations.

We find that regardless of the motility type, as flow rates increase, individual cells have trouble in moving across streamlines, resulting in weaker coupling between bacterial motility and their overall transport characteristics. We show that as flow speed increases, bacteria disperse faster in the direction of flow, due to a lower likelihood of motion across streamlines and an increase in longitudinal displacement driven by differentially advecting fluid in a porous environment. In other words, the distance between individual cells grows at a fast rate since cells are less likely to make transverse movements (i.e., displacements across streamlines), and are more likely to move longitudinally at a range of velocities produced by the parabolic nature of laminar flow profiles. Furthermore, we show that the motility of *Paenibacillus* is less impacted by flows in porous media than the motilities of *Geobacter* and *Acidovorax*, highlighting the strength of peritrichous flagella-driven motility. Additionally, we provide evidence that motile bacteria tend to oversample medium-velocity zones in porous media for the flow conditions tested in our experiments. This work thus provides an improved picture of the transport of motile bacteria in confined porous media under variable flow rates, especially in relation to the impact of flow on different motility types, with implications for several applications where an understanding of pore-scale transport and upscaling of bacterial transport is desired.

The work presented in the chapter is an enhanced version of the in-review article: "Berghouse, M. Perez, L.J., Plymale, A., Scheibe, T., & Parashar, R. Advection-Dominated Transport Dynamics of Pili and Flagella-Mediated Motile Bacteria in Porous Media. *Soft Matter*. 2024." Specifically, more information is provided in the discussion about the limitations of the research with respect to particle tracking, which is the primary motivation for chapters 3 and 4 of this dissertation, and more information is provided about the potential ways in which our work can contribute to mathematical descriptions of bacterial transport.

2.3. Materials and Methods

2.3.1 Bacterial Transport in Microfluidic Devices

To investigate the impacts of porosity, flow rate, and motility on bacterial transport, we recorded high-resolution videos of three species of bacteria swimming in microfluidic devices [2,000 mm width × 20 mm height ($w \times h$)] at flow rates of 0 µL/h (no flow), 1 µL/h and 5 µL/h. The chosen flow rates allow for comparative analyses of bacterial transport for the control condition of no-flow, and when the magnitude of flow speeds and bacterial motility speeds are of a similar order. The micromodels were made from polydimethylsiloxane (PDMS) and contained staggered pillar arrays of different grain diameters and pore lengths, resulting in either low porosity ($\phi = 0.42$) or high porosity ($\phi = 0.60$) micromodels. The mean fluid speeds (v_m) determined from flow rate (Q), cross-sectional area (A), and porosity (ϕ) as $v_m = \frac{Q}{A\phi}$, in the low porosity geometries were 16.5 µ/s and 82.7 µm/s, and the mean fluid speeds in the high porosity geometries were 11.6 µm/s and 57.9 µm/s, for the low and high flow rates, respectively. These correspond to fluid speeds in the range of 1 m/day to 7.15 m/day which are medium to high speed values typically observed in bioremediation applications in alluvial aquifer settings [38, 39]. We found conducting

experiments at lower flow rates (< 1µL/h) in our micromodels challenging due to pump limitations in establishing extremely low uniform rates and difficulties encountered by emergence of small drift speeds even in absence of external flow gradients presumably due to imperfections in model fabrication, small pressure aberrations at the inlet/outlet ports, small axial tilts, or potential presence of extracellular polymeric substance gradient [40]. Using pore throat length (given in section 2.2) as the characteristic length, and using the values of dispersion coefficients presented later in section 3.1.2, the velocities generated in the experiments would result in Peclet numbers approximately in the range of 0.25 to 16 – a range that allow us to make broad observations though they may be not fully generalizable. Experiments at no-flow condition in an open environment (i.e., without granular obstacles) were conducted with a subgroup of species in a previous work [41], which provides insights into the species-aware departure from Fickian diffusion in unconfined environments. After recording the videos from several replicates of experiments, we used TrackMate [42] to track, extract, and reconstruct thousands of trajectories (ranging in length from low tens of microns to several hundred microns) of individual cells (**Fig. 1**).

To account for the small drift observed in no-flow experiments from one end of the micromodel to the other, we calculated the background flow (drift speed) by computing the rate of change in the location of the centroid as $v_d = \frac{1}{k}\sum_0^k \frac{d}{dt}(r_{cm})$, where $r_{cm}$is the center of mass (x and y positions) of all bacteria in a frame at no-flow, and k is the number of frames (in our case, $k = 30$). To include the greatest number of possible trajectories for this calculation, we reset the starting point of all trajectories to $t = 0$. We calculated mean drift speeds for *Acidovorax* (0.29 µm/s), *Geobacter* (1.41 µm/s) and *Paenibacillus* (0.59 µm/s) separately. These drift speeds are significantly small compared to the speeds produced in the low-flow and high-flow experiments

thus allowing us to treat the transport of bacteria in the two flow experiments as being solely driven by the interplay between motility and hydrodynamics.

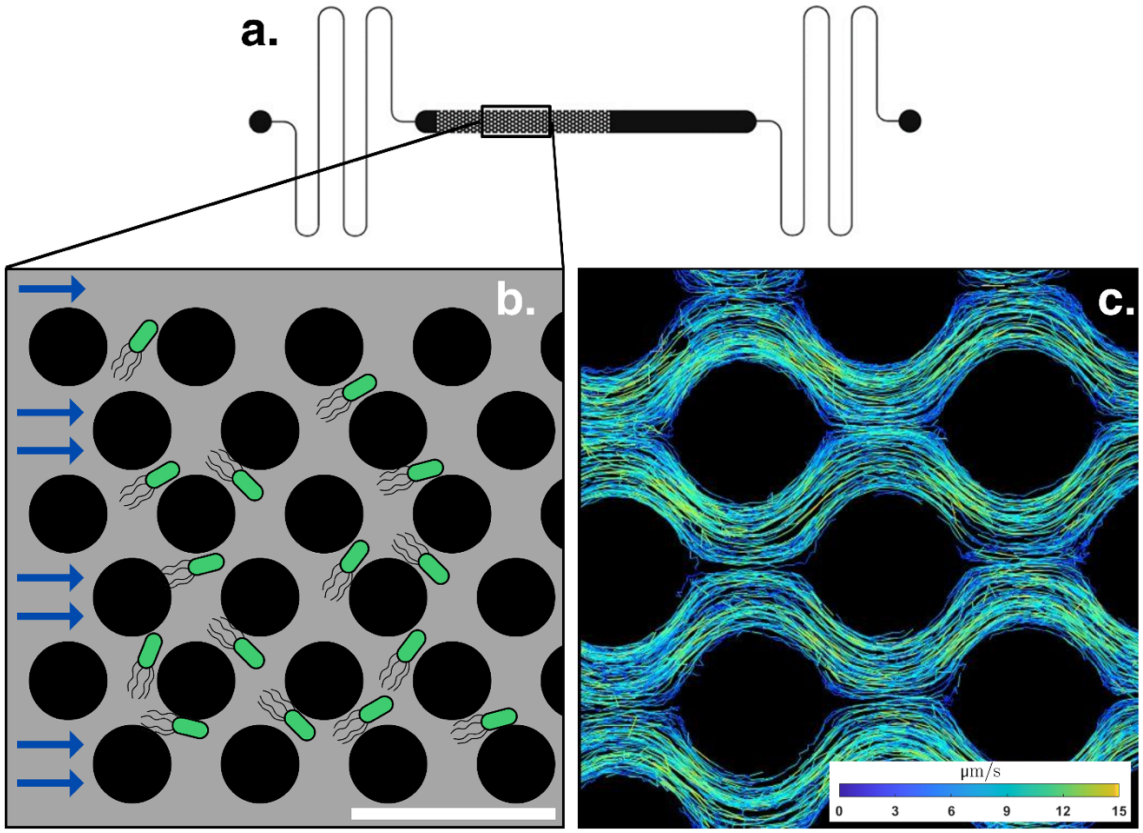


Figure 1. Experimental setup used to analyze bacterial transport in microfluidic devices. **(a)** A sketch of the full micromodel from [41], which used the same basic micromodel schematic as our experiments. The left black dot represents the inlet and the right dot represents the outlet. The black section represents an unobstructed part of the micromodel, and the gray section represents the part of the micromodel with cylinders. **(b)** Depiction of bacteria flowing (from left to right) through a section of the high porosity ($\phi = 0.60$) micromodel. The gray space represents the channels that fluid and bacteria travel through, and the black circles represent the cylinders (also referred to as "grains"). The scale bar represents 120 μm. Bacteria are not drawn to scale. **(c)** Bacterial trajectories for *Acidovorax* obtained in 5-minute interval over the course of the experiment in the high porosity micromodel at a flow rate of 1 μL/h. The colormap represents net speed of bacteria, with warm colors representing high speeds and cool colors representing low speeds.

Transport characteristics of bacteria were quantified using net speeds $|v_n| = \frac{\sqrt{(x_{t+1}-x_t)^2+(y_{t+1}-y_t)^2}}{\Delta t}$, turn angles $\alpha_t = tan^{-1}\left(\frac{y_{t+2}-y_{t+1}}{x_{t+2}-x_{t+1}}\right) - tan^{-1}\left(\frac{y_{t+1}-y_t}{x_{t+1}-x_t}\right)$, mean square displacement $MSD(t) = \frac{1}{N}\sum_{i=1}^{N}|r_i(t) - r_i(0)|^2$, velocity autocorrelations $C_v(\tau) = \langle|v_n|(t+\tau) \cdot v(t)\rangle$, effective dispersion coefficients $D^e(t) = \frac{1}{w\phi}\int_0^w D^e(t,y')dy'$, and bivariate angle-speed probability density contours. Here x_t and y_t are individual bacteria positions at time t, *N* is the total number of tracked cells, r_i is the displacement for bacterium i, and $|v_n|$ is the magnitude of the net velocity (i.e., speed) of the bacteria. The scripts used to calculate all statistics can be found in Supplementary Methods 2. Note that the net speeds are the speed of the bacteria determined through particle tracking. Since bacteria are displaced through the porous media both due to their own motility and the advection imparted by the background flow, the net speed obtained via particle tracking measures the combined effect of these two drivers.

2.3.2 Micromodel Construction

Micromodels for three porous geometries were constructed from PDMS using staggered arrays of grains to represent porous media (see Fig. 1). The three geometries used in this experiment were **(1)** arrays with a grain diameter (GD) of 80 µm and a pore throat length (PL) - minimum space between grains - of 20 µm ($\phi = 0.42$), **(2)** arrays with a GD of 40 µm and a PL of 20 µm ($\phi = 0.6$), and **(3)** arrays with a GD of 40 µm and a PL of 10 µm ($\phi = 0.42$). The micromodel dimensions were 2 mm in the transverse direction, 17 mm in the longitudinal direction (for the porous section), and 20 µm in the vertical direction. We chose a depth of 20 µm as we found that a larger depth causes bacteria to move in and out of the focal plane of our camera too often, and constraining the depth further would have caused excessive shear along the vertical plane.

2.3.3 Bacteria Culture

Bacterial strains *Acidovorax* JHL-9 and *Paenibacillus* 300A were grown in liquid culture aerobically at 30 °C on dextrose-free Trypticase Soy Broth (TSB). At late-log to stationary phase, cultures were diluted to an optical density at 600 nm (OD600) of ~ 0.1 – 0.15 and injected into the micromodel devices described above. *Geobacter sulfurreducens* was grown anaerobically (80:20 N_2:CO_2), in glass serum bottles or headspace vials, crimp-sealed with butyl-rubbers stoppers, on Freshwater Medium [43] with 50 mM sodium fumarate as electron acceptor in place of ferric citrate [44]. Stationary-phase *G. sulfurreducens* cells were injected, without dilution, into micromodel devices that had been de-oxygenated overnight in an H_2-free anoxic chamber (MBraun, O_2 < 10 ppm, 100% N_2). De-oxygenated micromodels were then removed from the anoxic chamber using an anaerobic jar and were kept in the anaerobic jar until immediately before use. *G. sulfurreducens* cells were removed from the serum bottle or headspace vial with a degassed (80:20 N_2:CO_2) 1-cc syringe and 22-gauge needle and immediately injected into the degassed micromodel. Though our video acquisition was generally restricted to about an hour after the injection of bacteria, it should be noted that *G. sulfurreducens* are also known to tolerate and grow with oxygen as a terminal electron acceptor for up to 24 hours [45].

2.3.4 Video Acquisition

All videos were collected with a confocal imaging technique on a Nikon Eclipse Ti2-U inverted microscope equipped (viewing vertically downwards and recording motion in the x-y plane) with a digital CMOS camera Hamamatsu Orca-Flash 4.0 controlled by NIS Elements imaging software. The sensor pixel size was 6.5 µm x 6.5 µm, and each recorded frame had a size of 2048 pixels x 2048 pixels. For videos at 10x magnification, the recorded domain size was 2048 x 6.5/10 = 1331.2 µm x 1331.2 µm, and for videos at 20x magnification the video domain was 665.6 µm x 665.6 µm.

Videos were recorded for 5 minutes at frame rates of about 10 frames per second (the interval time between frames varied slightly resulting in rates of 8-12 frames per second). The exact time interval between frames were recorded to allow for accurate computation of transport metrics.

2.3.5 Image Preprocessing

The raw videos were preprocessed with background subtraction using a lag method specifically developed in-house for these experiments. To capture trajectories of bacteria that may have not moved between two successive frames, the subtracted background must be more than a few frames back in time. Standard practice in background subtraction for such cases is to use the initial frame, or the mean frame, as a background for the rest of the video, but this was not possible in our case due to variability in image brightness throughout the duration of the video. To get around these problems, we used the 5^{th} previous frame to perform background subtraction. In other words, to subtract the background of frame 6, we calculated frame 6 minus frame 1. Thus, any bacteria that moved a little over the course of 5 frames could still be identified in particle tracking.

2.3.6 Particle Tracking

After background subtraction, the foreground was then loaded into ImageJ and particle tracking was performed with the plugin TrackMate. For feature detection the Laplacian of Gaussian (LoG) detector was used, and to link the features, we used the Linear Assignment Problem (LAP) tracker. A sample output of trajectories from TrackMate is given in the Supplementary Data.

2.3.7 Flow Field Simulations

Simulations of the flow field were conducted to understand which parts of the geometry bacteria are likely to oversample or undersample. The experimental geometry was initially digitized in Blender then refined in OpenFOAM to produce a regular grid consisting of 2400 x 2400 x 72 voxels

with a resolution of $\Delta x = \Delta y = 0.2773\ \mu\text{m},\ \Delta z = 0.2778\ \mu\text{m}$. The flow fields of the digitized geometries were obtained by solving the flow of incompressible Newtonian fluid governed by the Navier-Stokes equations using SimpleFoam. The steady-state solver belongs to the OpenFOAM package that uses semi-implicit methods for pressure linked equations algorithms. Constant flow rate at specific experimental values of 1 and 5 μL/h ($Q = 2.78 \times 10^{-13}\text{m}^3\text{s}^{-1}$ or $Q = 1.39 \times 10^{-12}\text{m}^3\text{s}^{-1}$) and constant pressure $P = 0\ \text{kg m}^{-1}\text{s}^{-2}$ were imposed at the inlet and outlet of the domain, respectively. No-slip conditions were assigned to the fluid-solid interface. We used a kinematic viscosity ν of 1.14x10^{-6} m^2/s for the fluid (TSB), given a calculated ratio of $\frac{\nu_{TSB}}{\nu_{water}}$ of 1.14 [46]. A sample case folder for the simulations, as well as the commands used to run it on a local machine, can be found in Supplementary Methods 1.

2.4 Results

2.4.1 Advection-Dominated Transport Dynamics

We use the term "advection-dominated transport" to highlight a regime wherein the variable shear forces within pore spaces, and dominance of flow speeds over motility speeds, restrict the ability of bacteria to move across streamlines, thus guiding their motion primarily along streamlines at differential velocities. Advection-dominated transport would occur in scenarios of high Peclet number [47] and persists in situations where weak coupling between motility and biofilm formation patterns are observed [48]. This type of transport, guided by shear induced cell rotation causing decreased transverse dispersion and increased lateral dispersion, has also been previously reported for bacteria in porous media flows [49]. Here, we provide additional relevant statistical information to characterize advection-dominated transport for three different species of motile bacteria relevant to bioremediation applications. In the following, we characterize

transport dynamics through the (MSD), turn angle distribution, $C_v(\tau)$, and $D^e(t)$. We use these statistics to develop a robust understanding of transport driven by differential advection, movement across streamlines, velocity decorrelation, and spreading [50, 51].

2.4.1.1 Turn Angle Analysis Reveals Impact of Flow Rate on Motility

To understand the motility of each species of bacteria, we primarily use their turn angle distributions. Note that turn angles as defined in this chapter are not the same as traditionally reported turn angles that reflect the body orientation during bacterial run and tumbles [52]. Because the bacterial speeds are not significantly greater than the background flow speeds, and because our experiments were performed at a relatively low frame rate, we use the term "turn angle" to capture the relative change in the trajectory of advected cells between successive video frames.

For no-flow conditions, *Paenibacillus* (peritrichous flagella-based motility) show a high probability of low or very high turn angles (Fig 2a). Low turn angles ($\alpha > -30° \; or \; \alpha < 30°$) would represent persistent forward motion (i.e., long run times), and high angle turns would represent reversals in direction (i.e., tumbling). Although not a necessary condition, a high probability of low turn angles implies a high probability of runs, and a high probability of medium to high turn angles implies a high probability of tumbling. *Geobacter* and *Acidovorax* (pili or monotrichous flagella-based motility) have a relatively low probability of both low and high turn angles because their motion is generally more random and is subject to slight changes in hydrodynamics.

To identify the differences in motility for each species more effectively, we also report the ratio of the turn angle PDF for each species to the turn angle PDF for *Geobacter* (Figs. 2d-2f). Essentially, *Geobacter* represents our twitching baseline, as their speed distributions (Supplementary Figure 1) and mean speed (about 2.3 μm/s after subtracting average drift) are

generally in agreement with previously reported twitching speeds of various bacteria [32, 35-37]. Thus, the PDFs of *Acidovorax* and *Paenibacillus* turn angles show departure of their motion behavior from a typical twitcher. When flow is absent (Fig. 2d), the PDF ratio for *Paenibacillus* shows exactly what we expect from a swimmer – high probability of low turn angles (persistent forward motion), low probability of medium turn angles (random motion), and high probability of high turn angles (direction reversal). Intriguingly, the genome of *Acidovorax* JHL-9 (see data availability) contains numerous genes related to twitching motility but not a complement of genes related to flagella-mediated motility. However, as previously discussed, wet mount TEM images of strain JHL-9 (Supplementary Figure 2) suggest the presence of polar flagella. Furthermore, the speed (Supplementary Figure 1) and turn angle ratio (Fig. 2d) distributions at no-flow indicate that *Acidovorax* behaves differently than *Geobacter* and closer to *Paenibacillus,* thus revealing distinct motility traits to the three species selected in this study.

Compared to the case of our no-flow experiments, the interpretation of our turn angle distributions in the presence of flow is slightly more complicated. We posit that in a viscous steady-state flow, non-motile bacteria would behave as inert particles transported by advection only, thus moving along streamlines of the pore-scale flow field, which would result in small turn angles between successive steps of the trajectories. In other words, in presence of a background flow, persistent forward motion means a high probability of low turn angle distributions. However, motile bacteria also move across streamlines, move in reverse direction, and explore the pore space under flow conditions, and as a result, large turn angles should be expected for highly active self-propelled bacteria [25, 53-55].

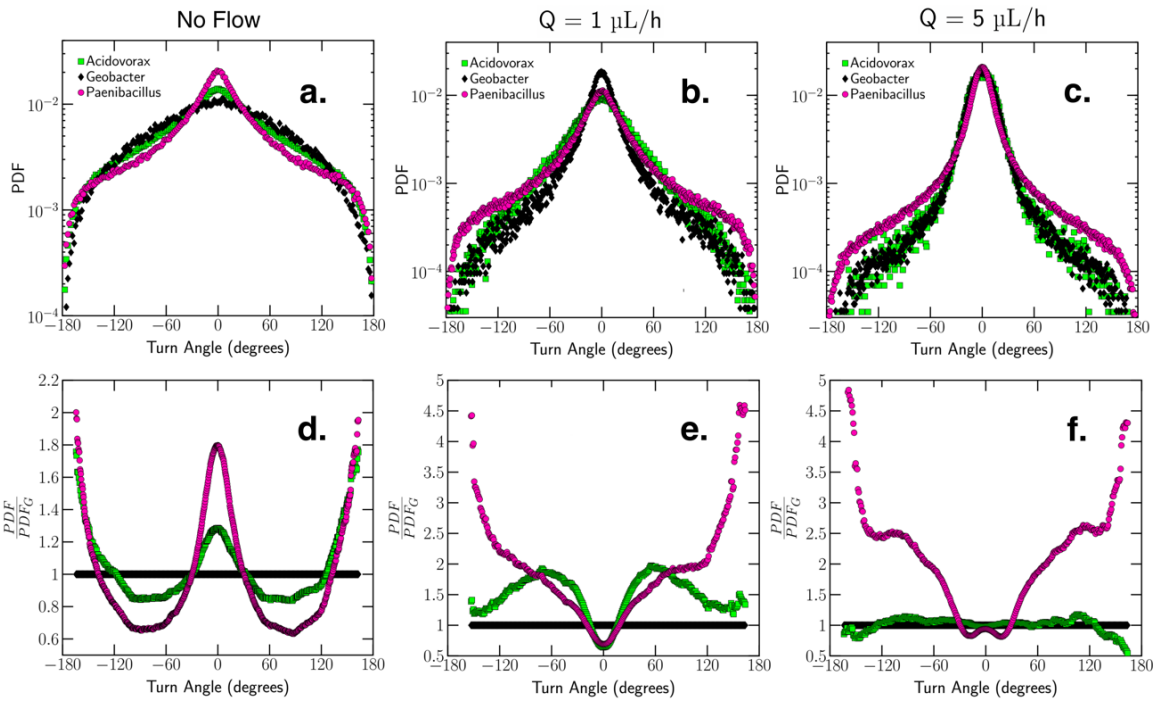


Figure 2. Turn angle PDFs for all three species in the low porosity geometry (grain diameter = 80 μm, pore length = 20 μm). **(a)** Turn angle PDF at flow rate of 0 μL/h. **(b)** Turn angle PDF at flow rate of 1 μL/h. (c) Turn angle PDF at a flow rate of 5 μL/h . (d) Turn angle PDF ratio at a flow rate of 0 μL/h. (e) Turn angle PDF ratio at a flow rate of 1 μL/h. (f) Turn angle PDF ratio at a flow rate of 5 μL/h. Turn angle PDF ratios are calculated as the PDF/PDFG, where PDFG represents the PDF of Geobacter. We choose to make the ratios relative to *Geobacter* as they move much less than the other bacteria. Convergence of the shape of the turn angle distribution and clustering of turn angles around 0° at 5 μL/h indicates strong advection-dominated transport.

We find that *Paenibacillus* have a higher probability of large turns ($\alpha < -90° \; or \; \alpha > 90°$) than the other two species at a flow rate of 1 μL/h, while at 5 μL/h this difference is even more noticeable (Figs. 2b and 2c). Examining the turn angle PDF ratios (Fig. 2e), we see that at 1μl/h *Paenibacillus* and *Acidovorax* have similar distributions for low to medium turn angles, but *Paenibacillus* has a much higher probability of large turn angles, indicating a greater potential for direction reversal than the other bacteria. At 5 μL/h (Fig. 2f) the similarities between *Paenibacillus* and *Acidovorax* completely disappear, and the turn angle PDFs for *Acidovorax* and *Geobacter* essentially converge. This implies that at high flow speeds, *Paenibacillus*, with its peritrichous

flagella, are either able to tumble more, or run faster, than *Acidovorax* or *Geobacter*. Furthermore, these results suggest that *Acidovorax,* with its monotrichous flagella, experience a greater impact on its motility due to flow speed than *Paenibacillus* do. While more experiments with a greater variety of monotrichous and peritrichous species are needed to confirm this trend, our initial results imply that peritrichous flagella enable motility at higher flow speeds than monotrichous flagella. Since the no-flow speed PDFs show that *Acidovorax* and *Paenibacillus* have similar max speeds, and monotrichous and amphitrichous bacteria have generally been shown to be capable of higher speeds than peritrichous bacteria [36, 58], it does not seem likely that the difference in motility between *Acidovorax* and *Paenibacillus* at high speed is due to run speeds. Thus, our results also imply that at high flow rates *Acidovorax* are unable to tumble, but *Paenibacillus* can. This is supported by research showing that increasing numbers of flagella increases the probability of tumbling [59]. However, we should also note that differences in the flagellar architecture are not the only possible explanations for differences in the turn angle distributions. Two other possible explanations for this include reorientation strategies, which may impact their preference to run or tumble [60], and size-related dynamics, which have been shown to influence hydrodynamic impacts on bacterial motility [61]. Regardless of the exact cause, our results do show that *Paenibacillus* can maintain swimming-like behavior at higher flow rates. *Acidovorax*, on the other hand, act like swimmers at low or no flow, and twitchers at high flow. In other words, advection-dominated transport, which causes trajectories of swimmers to appear similar as trajectories of twitchers, occurs at a lower flow rate for *Acidovorax* than for *Paenibacillus*.

2.4.1.2 Effect on Bacterial Spreading at Different Porosities and Flow Rates

Figure 3 confirms additional evidence of advection-dominated transport via the computed MSD. Here, we introduce the term “differential advection” to describe the MSD results, stemming from

the bacteria's mixed super-diffusive motions influenced by streamline shifts, trapping, and pore space exploration. This term aptly captures the relationship between velocity decorrelation events and bacterial advection and offers a nuanced understanding of the transport dynamics. As the flow rate increases, bacteria will, on average, have a larger range of displacements due to the magnitudes of velocities it can sample within the laminar profile of porous media flow [17]. Furthermore, smaller turn angles at higher flow rates (as described in the previous section) implies less streamline changing. This results in higher values of the MSD driven by increased differential advection as bacteria are transported by a range of velocities produced by converging and diverging streamlines within the pore network. These observations are further used for later comparisons in the context of Fig. 6, and generally align with enhanced dispersion reported due to transport of bacteria along faster flow paths than the local flow [17, 25].

Complementing the increased differential advection, as the flow rate increases, the MSDs of all species of bacteria in the low porosity geometry show signs of convergence (both in slope and magnitude as seen in Fig. 3e), likely driven by decorrelation of cell swimming as bacteria navigates pore structures [56]. Essentially, advection-dominated transport is thus revealed by the convergence (between different species of bacteria) of both turn angle distributions and MSD, and a shift toward greater differential advection. In contrast, with the high porosity geometry, we observe less evidence of MSD convergence, which indicates that the flow speeds are not high enough to suppress bacterial motility, leading to a reduction in differential advection as shown by lower MSD values (Figs. 3b and 3c). We also find that for a fixed porosity and flow rate, *Geobacter* and *Acidovorax* always advect more than *Paenibacillus*, further supporting the idea that peritrichous swimmers are differentially advected to a lesser degree than twitchers or polar swimmers.

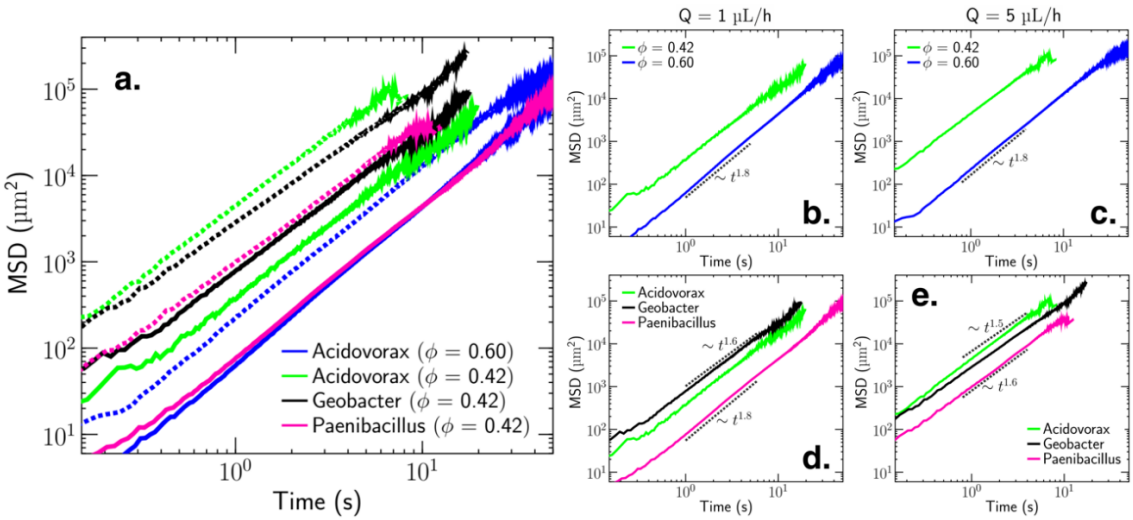


Figure 3. Mean square displacements (MSDs) at different porosities and flow rates for three different species of bacteria. **(a)** MSDs from relevant experiments (3 species at lower porosities and 1 at higher porosity, 2 flow rates). The 1 µL/h results are shown as solid lines and the 5 µL/h results are shown as dotted lines. **(b)** MSDs for *Acidovorax* for $\phi = 0.60$ and $\phi = 0.42$ at a flow rate of 1 µL/h (mean fluid speed of 11.6 µm/s and 16.5 µm/s respectively). **(c)** MSDs for *Acidovorax* for $\phi = 0.60$ and $\phi = 0.42$ at a flow rate of 5 µL/h (mean fluid speeds of 57.9 µm/s and 82.7 µm/s respectively). **(d)** MSDs for all species for $\phi = 0.42$ at a flow rate of 1 µL/h. **(e)** MSDs for all species for $\phi = 0.42$ at a flow rate of 5 µL/h. These figures show an increase in the impact of differential advection for motile bacteria as the flow rate increases. The rapid increase in MSD driven by differential advection, along with convergence of the MSDs in the low porosity geometry at 5 µL/h, provide evidence of advection-dominated transport. At both flow speeds, *Paenibacillus* show lower values of MSD than *Geobacter* or *Acidovorax*, indicating a stronger resistance to advection-dominated transport. All low porosity results in this figure are from the grain diameter = 80 µm, pore length = 20 µm geometry.

To further understand spreading in our experiments, we calculated the effective dispersion coefficient, D^e, based on the average spatial variance of the bacteria distribution evolving from a point-like injection, that is, the transport of Green function as defined in [62, 63]. At 1 µL/h, D^e is impacted by spreading in both the longitudinal and transverse directions for *Paenibacillus* and *Acidovorax*, but primarily in the longitudinal direction for *Geobacter* owing to its considerably lower twitching speed than the mean fluid speed. At 5 µL/h, D^e primarily represents longitudinal dispersion for all species. Our results show that when the flow rate increases from 1 µL/h to 5 µL/h, D^e increases the most for *Geobacter*, and the least for

Paenibacillus (Table 1). Because *Paenibacillus* are able to maintain some form of motility at 5 µL/h, and as a result are still able to change streamlines and explore the pore space, differential advection has less of an impact on their dispersion than it does for the dispersion of *Geobacter* and *Acidovorax*. In other words, the bacteria that follow streamlines or explore less space in the transverse direction to the flow, advect and spread more in the direction of flow. These results complement those presented in [17, 25, 49], which showed that hydrodynamic gradients in porous geometries reduce transverse dispersion. We further this research by showing that bacterial transport is advective-dominated for a wide variety of flow rates depending on the type of bacterial motility.

Table 1. Effective Bacterial Dispersion Coefficients D^e ($\mu\frac{m^2}{s}$) for all experiments conducted in the low porosity geometry ($\phi = 0.42$). As flow rate increases, *Geobacter* have the greatest increase in dispersion and *Paenibacillus* have the smallest increase in dispersion. As the motility speeds of the bacteria are less than the fluid speed at 5 µL/h, dispersion is almost entirely in the direction of flow for the 5 µL/h experiments.

D^e ($\mu\frac{m^2}{s}$)	*Paenibacillus*	*Geobacter*	*Acidovorax*
1 mL/h	107 ± 46	308 ± 97	204 ± 56
5 mL/h	217 ± 43	895 ± 148	466 ± 74

2.4.1.3 Velocity Autocorrelation to Examine Emergence of Advection-Dominated Transport

We use the velocity autocorrelation function (C_v) to further provide information on advection-dominated transport [64]. Generally, in porous media, bacteria show decorrelation in velocities over time due to a tendency for sampling different portions of the flow field and trapping events (i.e., pore confinement, occurrence of collisions and attachment to obstacles) [8, 63]. Previous research has shown that decorrelation of bacterial trajectories is more rapid at high flow rates

[49]. Our findings both confirm these trends and present new information on how motility type impacts decorrelation. We show that *Paenibacillus* and *Acidovorax* exhibit decorrelation faster than *Geobacter* at 1 µL/h (Fig. 4a), but that at 5 µL/h, all decorrelation times are essentially the same. This suggests that at low flow rates swimmers experience larger variations in velocity over time by sampling multiple streamlines and trapping events that decorrelate subsequent velocities. However, as flow rate increases, motility type no longer has significant impact on decorrelation events. These observations further support the presence of flagella-based swimming for *Acidovorax* at 1 µL/h. Convergence of C_v decorrelation times at high flow rate points to the emergence of advection-dominated transport. These trends generally agree with the observations of the MSD, D^e and turn angle distribution analysis, although C_v is slightly less sensitive to differences in motility than the other metrics.

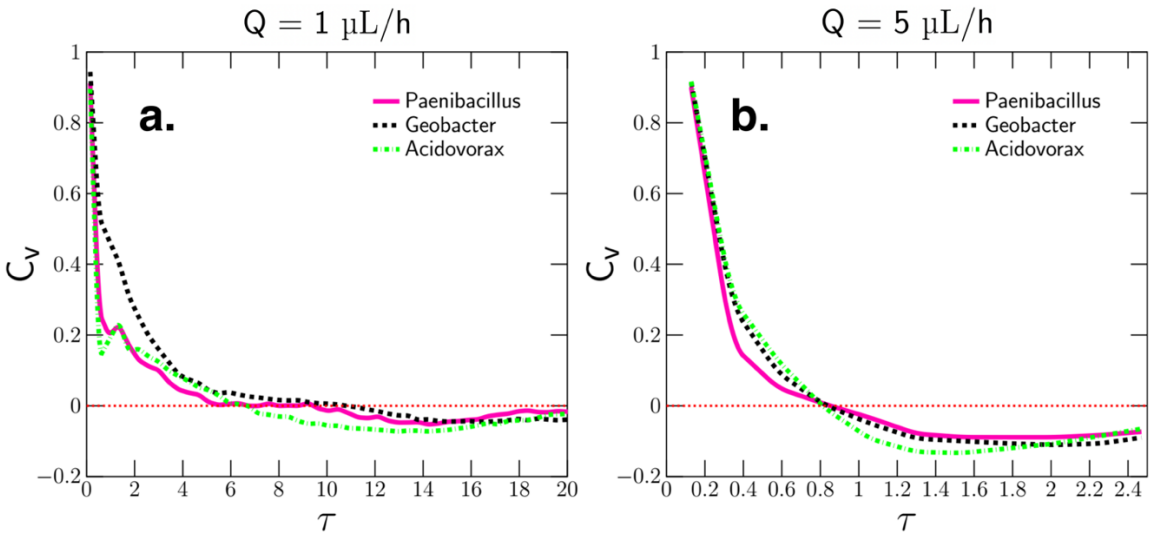


Figure 4. Velocity Autocorrelation functions (C_v) for the low porosity experiments at (a) 1 µL/h and (b) 5 µL/h. At 1 µL/h, the swimming species (*Paenibacillus* and *Acidovorax*) show faster decorrelation time (approximately 6.7 seconds) than the 10.8 second decorrelation time of the twitching species (*Geobacter*). At 5 µL/h the C_v for the plots of all species converges producing a decorrelation time of approximately 0.8 seconds, which is further evidence of advection-dominated transport.

2.4.2 Spatial Variations in Net Speeds

We digitized the experimental microfluidic geometries and simulated the steady-state viscous flow at high resolution using SimpleFoam [65, 66] to determine how bacteria may under or oversample different parts of a flow field (Fig. 5). By comparing the obtained simulated distribution of fluid speeds against the experimentally derived distribution of net bacterial speeds, we can develop an understanding of the zones within a pore network that bacteria may preferentially occupy. We recognize that a more accurate comparison would use flux weighting and particle tracking to compare the simulated fluid speed PDFs with the net bacterial speed PDFs. However, given the large number of trajectories (tens of thousands for each bacteria), and the periodic nature of our flow field, we posit that the trajectories of tracked bacteria adequately sample the domain space and thus provide basis for comparison to simulated speeds.

We observe that regardless of flow rate, motility type, or porosity, motile bacteria in porous media tend to undersample low-speed zones and oversample medium-speed zones (relative to the Eulerian fluid speed PDF) (Fig. 6). This provides additional insight challenging the notion of shear trapping which suggests bacteria in a shear flow will oversample low-speed zones [67, 68]. A plausible explanation for this observed difference lies in recognizing that studies reporting shear trapping were often conducted in simpler geometries (e.g., straight channels) [14] than in porous media geometries producing converging and diverging streamlines, although it has been reported that shear may lead to creation of hotspot of colonization instead of trapping close to walls of curved surfaces [18]. . An exact relationship between hydrodynamics and the observed speed sampling cannot be deduced in our study because of lack of high-resolution tracking

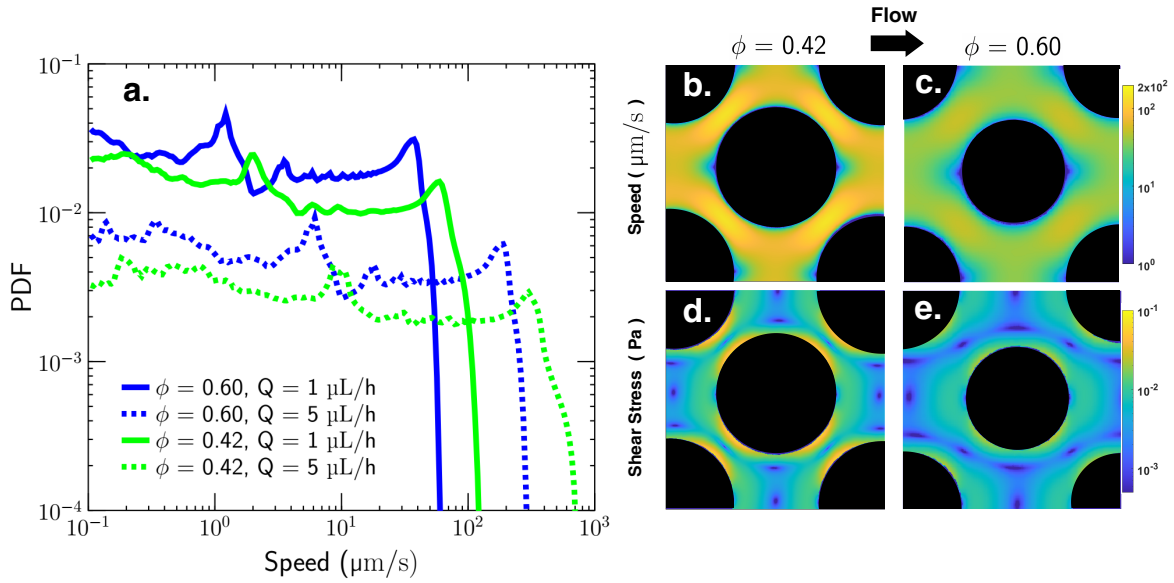


Figure 5. Results from the steady-state viscous flow simulations of our experimental microfluidic geometries. All low porosity results in this figure are from the grain diameter = 80 μm, pore length = 20 μm geometry. **(a)** Probability distribution functions (PDFs) of fluid speed from the simulated flow fields for each porosity and flow rate used in our experiments. **(b)** Zoomed-in velocity magnitude field for the low porosity simulation at 1 μL/h. **(c)** Zoomed-in velocity magnitude field for the high porosity simulation at 1 μL/h. **(d)** Zoomed-in shear stress magnitude field for the low porosity simulation at 1 μL/h. **(e)** Zoomed-in shear stress magnitude field for the high porosity simulation at 1 μL/h.

necessary to compute shear-induced lateral transport towards the walls due to Jeffrey orbits [69-71] and potential backward swimming in the leeward side of the grains, which has been reported has another form of shear trapping in more complex geometries [18].

In no-flow conditions (Supplementary Fig 1), after subtracting drift speeds, *Geobacter* has a mean speed of 2.3 μm/s, *Acidovorax* has a mean speed of 5.9 μm/s, and *Paenibacillus* has a mean speed of 7.2 μm/s. The 95th percentile speed for the three species are 4.5 μm/s, 17.9 μm/s, and 20.6 μm/s for *Geobacter*, *Acidovorax*, and *Paenibacillus* respectively. At the low flow rate of 1 mL/h (16.5 μm/s average flow speed at low porosity), the swimming speeds of *Acidovorax* and *Paenibacillus* can thus exceed the fluid flow speeds (although just barely in the case of

Acidovorax), but at 5 mL/h, none of the bacteria in our study can consistently exceed the fluid flow speed. Thus, advection-dominated transport in its simplest form is a result of flow speeds

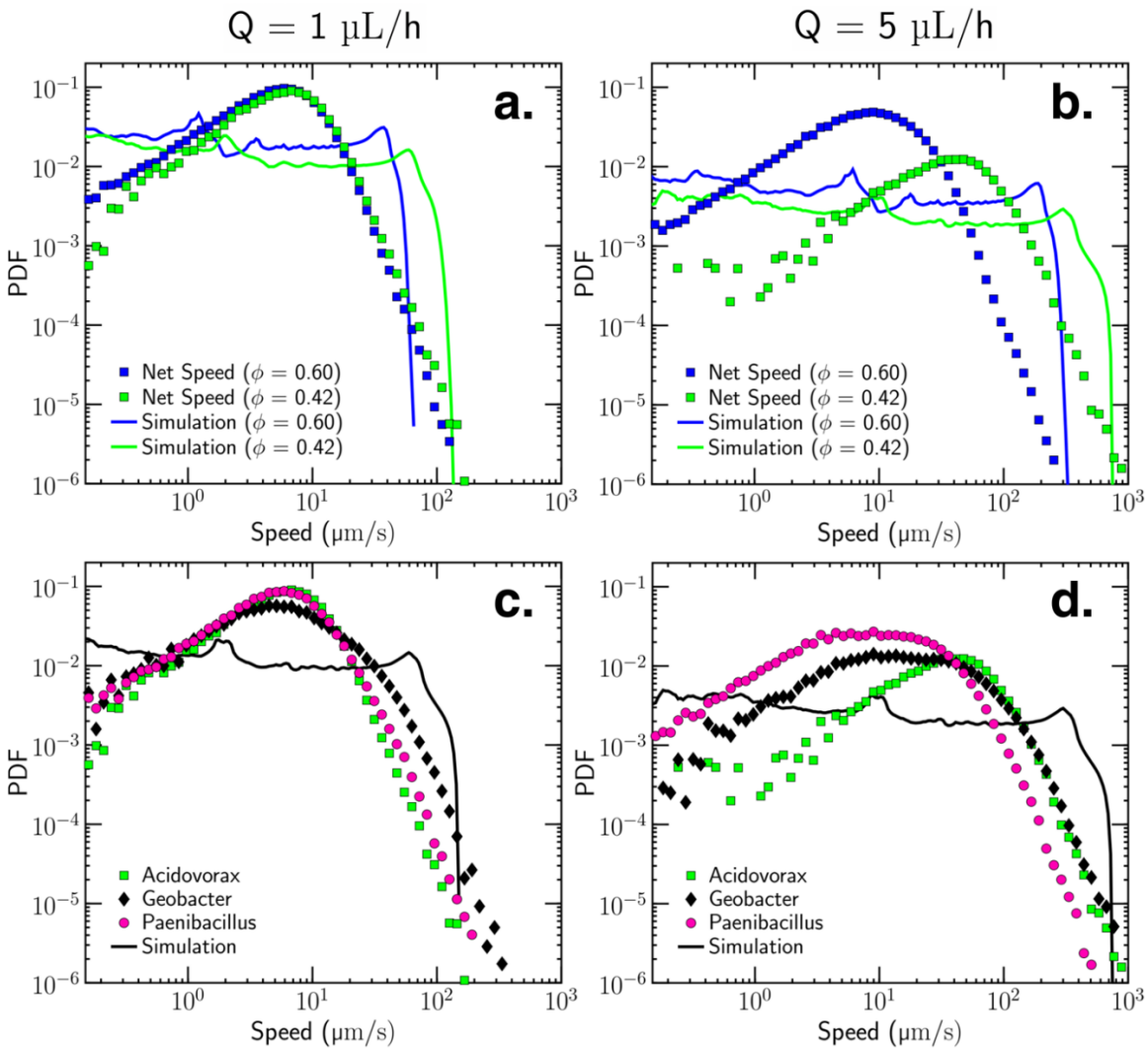


Figure 6. PDFs of speeds plotted to show the net speeds of the bacteria overlayed by the simulated speeds for the respective flow rate and porosity. The net speeds are represented by scatter points, whereas the simulated speeds are represented by solid lines. All low porosity results in this figure are from the grain diameter = 80 μm, pore length = 20 μm geometry. All simulated PDFs represent the distribution of Eulerian flow speeds for that geometry. **(a)** Net and simulated speed distributions for *Acidovorax* for $\phi = 0.60$ and $\phi = 0.42$ at a flow rate of 1 μL/h. **(b)** Net and simulated speed distributions for *Acidovorax* for $\phi = 0.60$ and $\phi = 0.42$ and flow rate of 5 μL/h. **(c)** Net and simulated speeds for all species for $\phi = 0.42$ at a flow rate of 1 μL/h. **(d)** Net and simulated speeds for all species for $\phi = 0.42$ at a flow rate of 5 μL/h. These figures show the tendency for motile bacteria to oversample medium-speed zones within a porous media.

exceeding motility speeds. However, shear adds another layer of complexity when considering the ability for bacteria to bundle/unbundle their flagella. Recent work has shown that at a shear magnitude of about 0.26 Pa, E. Coli lose control over this mechanism and can't effectively swim [72]. The 1 μL/h simulations in our study do not produce shear exceeding this value (Figs. 5d and 5e), but at 5 μL/h, the value of shear close to the grain is larger than this threshold value. Although an investigation of threshold shear magnitudes for bundling abilities in *Paenibacillus* and *Acidovorax* was beyond the scope of this chapter, the inability to control motility [73] through bundling/unbundling of flagella remains a likely explanation for the differences observed in the transport of *Paenibacillus* and *Acidovorax*. It should be noted that in our quasi-2D porous media, under uniform and laminar flow, there are no chemotactic or thermal gradients influencing the transport. Thus, the magnitude and distribution of shear within a porous media, which attains its maximum value at grain surfaces and is minimum along centerlines of a pore channel, is likely the primary physical mechanism that controls bacterial transport.

2.4.3 Combined Effect of Turn Angle and Net Speed on Spreading

We further analyze the combined influence of net speed and turn angle on the advective spreading of motile bacteria using a matrix of bivariate (speed-angle) joint probability density contours (Fig. 7). The probability density matrix allows us to observe general relationships between the differential advection driven spreading plotted in Fig. 3a, and the turn angles and net speeds of the bacteria. As net speed increases, bacteria have a narrower range of turn angles and, therefore, greater spreading in the longitudinal direction stems from strongly advective particle motion. In the top and middle rows of Fig. 7, larger turn angles and less spreading are seen from

left to right. In the bottom row, there is no significant change in large-angle turns or advective spreading. Thus, somewhere between the middle and bottom rows, or around a median speed of 50-100 mm/s, the impacts of advection-dominated transport increase to the extent that changes in fluid speed causes insignificant difference in advective spreading or turn angle for bacteria of the same motility type. This suppression in active dispersion in the case of strong fluid flow corroborates recent studies of transport of actively moving particles in porous media [74]. In addition to providing deeper insight into the transition to the advection-dominated regime, the joint probability density matrix also shows that bacteria are more likely to make large turns at low speeds than at high speeds. Conversely, small-angle turns are more likely to occur at high speeds than large-angle turns. When bacteria are moving with faster streamlines, their turn angles are smaller as they are more likely to go with the flow. When moving with slower streamlines, bacteria are more likely and more able to make large turns and cross transversely to other streamlines. This provides further evidence that pore space exploration and movement across streamlines require low fluid speeds and results in large turn angles.

2.5. Discussion

This study focuses on investigating the impact of flow rates and porosity on the transport of different species of motile bacteria in porous media. We show that *Geobacter*, with their purely twitching-based motility, understandably need surfaces to propel themselves forward and are unable to swim in suspended media, and as a result are not fast enough in a porous media domain to show any impacts of motility on their transport at low or high flow rates. *Paenibacillus*, with their peritrichous flagella, exhibit strong swimming motility. Although they exhibit weak motion

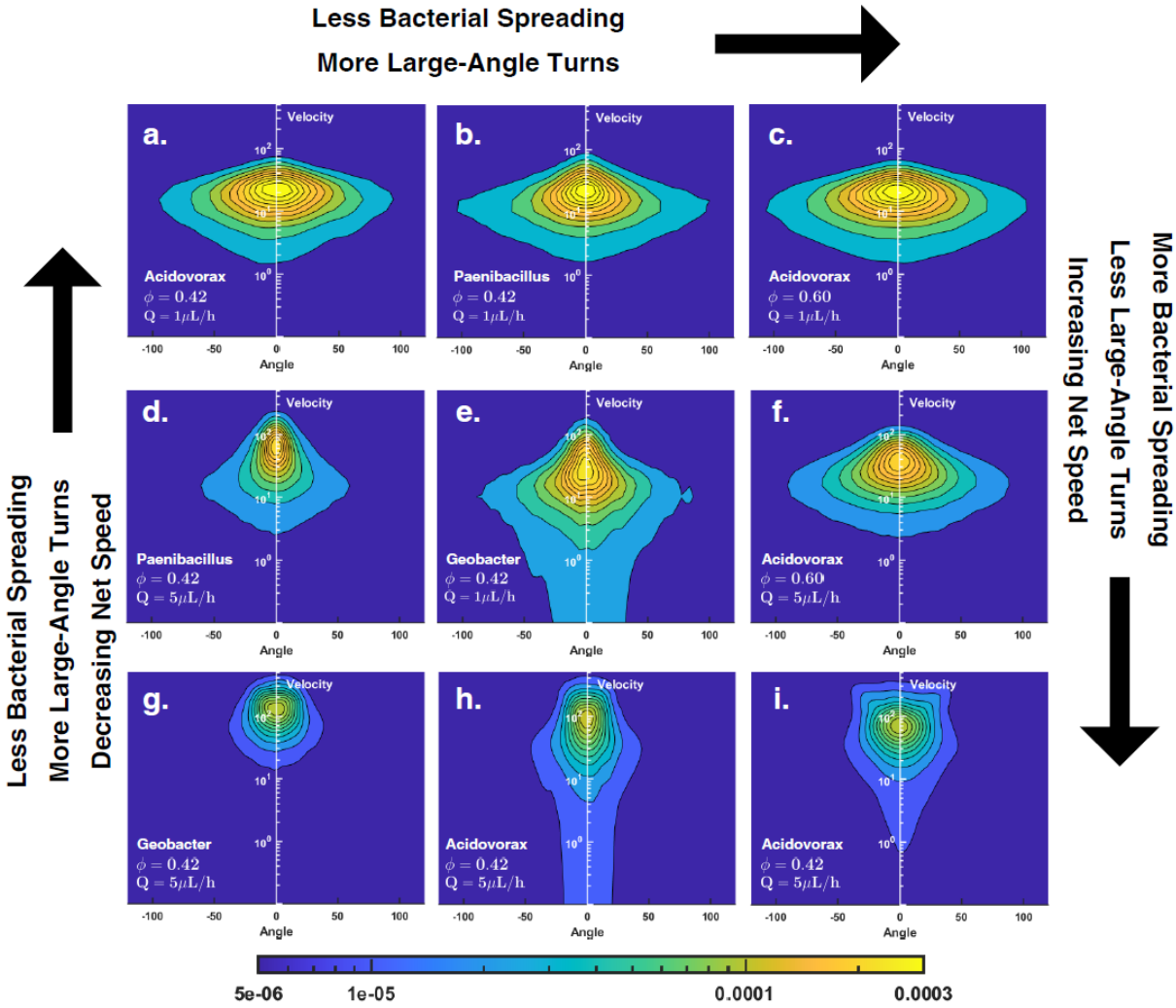


Figure 7. Velocity-angle joint probability density matrix. **(a)** *Acidovorax*, $\phi = 0.42$, 1 µL/h (grain diameter = 80 mm, pore length = 20 µm) **(b)** *Paenibacillus*, $\phi = 0.42$, 1 µL/h **(c)** *Acidovorax*, $\phi = 0.60$, 1 µL/h **(d)** *Paenibacillus*, $\phi = 0.42$, 5 µL/h **I** *Geobacter*, $\phi = 0.42$, 1 µL/h **(f)** *Acidovorax*, $\phi = 0.60$, 5 µL/h **(g)** *Geobacter*, $\phi = 0.42$, 5 µL/h **(h)** *Acidovorax*, $\phi = 0.42$, 5 µL/h (grain diameter = 80 mm, pore length = 20 mm) **(i)** *Acidovorax*, $\phi = 0.42$, 5 µL/h (grain diameter = 40 µm, pore length = 10 µm). Each figure in the density matrix shows probability density contours for net speed and turn angle for a particular set of conditions. As we move across the matrix from bottom to top, we see decreased net speeds, increased large-angle turns, and less spreading. As we move from left to right across the matrix, we see a slight increase in large-angle turns and decrease in spreading, but not as much as going from bottom to top. There is no significant change in net speed moving from left to right. This figure implies that fast net speeds are required for bacteria to be in the advection-dominated regime, which results in more small-angle turns. Furthermore, past a threshold speed of about 50-100 µm/s, motile bacteria are unlikely to have large-angle turns.

across streamlines and exploration of pore space at high flow rates, their turn angle distributions reflect a higher degree of activity. In the middle of the twitchers and swimmers are *Acidovorax*. Their transport metric are closer to swimmers at no-flow and at a flow rate of 1 µL/h, but at 5 µL/h, their transport metrics tend to appear closer to twitchers. Although a deep investigation of the motility type of *Acidovorax* is beyond the scope of this work, we show that differences in flagellar architecture offer a reasonable explanation for their behavior. Our results and previous imaging of *Acidovorax* suggest that they have a single polar flagellum as opposed to the peritrichous flagella of *Paenibacillus*. At high flow rates, it appears that peritrichous flagella are more able to facilitate tumbling behavior, and thus, movement across streamlines.

In the advection-dominated transport regime, lack of pore space exploration and streamline changing results in less transverse movement and thus leads to an overall increase in transported distance and spread in the direction of flow. We show that advection-dominated transport is revealed through convergence of turn angle distributions, MSDs, species independent velocity decorrelation times, and a clustering of turn angles around $0°$. This study also provides contrasting results to the notion of shear trapping wherein motile bacteria are expected to oversample low-velocity regions in a shear flow. In the case of our complex porous geometry, we observe bacteria oversampling medium-speed regions. When the geometry of pore channels allows for convergence and divergence of streamlines in 2D space, producing hydrodynamic patterns typically found in realistic porous media, wide-ranging values of shear forces emerge, leading to an interesting interplay between shear, motility, and the overall bacterial transport. Oversampling of medium-speed zones could be because of high levels of shear (closer to the walls) preventing bacteria from bundling their flagella [72]. In this unbundled state, the bacteria act as deformable objects resulting in a stokes lift force as they approach surfaces. In addition, size

exclusion and hydrodynamic chromatography have shown that the transport of microbes is dependent on size and shape [75, 76]. Size exclusion occurs because bacteria are too large to only occupy the slow speed zones around the grain [77]. Unless they are attached, bacteria will move more quickly around the grain than a solute will because part of their body is in higher speed zones. Finally, electrostatic repulsion, or the likely presence of energy barriers close to the grains, may prevent bacteria from getting too close to surfaces [18, 55, 78-81]. The wide variety of plausible explanations for absence of shear trapping in our study, or shear trapping being manifested at through bacterial reorientation in the leeward side of the grains, illustrate the complexities of analyzing bacterial transport in porous media. The oversampling of medium speeds can thus be a result of several different hydrodynamic or biophysical properties, and we identify shear as a likely physical phenomenon underpinning our observed transport patterns.

Our work complements previous studies that have shown advection to dominate the transport of bacteria at high flow rates [17, 25], effectively erasing the differences in motile behavior between different species of bacteria [14, 18, 67, 78, 82]. Our work also builds upon the body of evidence showing that there are significant transport differences between swimmers and twitchers [83, 84], and that bacteria with straighter paths (non-motile) spread more (in the direction of flow) than bacteria with exploratory paths (motile) at low flow speeds [17, 25]. We expect the results presented here to help future researchers in developing more robust experiments and models for not only bioremediation, but other applications where species-aware transport dynamics at small-scale can support and inform development of improved upscaled models. As evidenced by plots of various transport metrics which tend to move towards convergence at high flow rates, it can be argued at sufficiently high flow rates in porous media, different bacterial species will all exhibit uniform transport characteristics not too dissimilar than

those expected from passive tracers. The usefulness of the presented research is in recognizing that such high flow rates are rarely encountered in porous media applications, and the progression of bacterial species towards a uniform transport behavior depends on the flow rates, porosities, and the motility types.

Although it was beyond the scope of this work to rigorously investigate how our results may be applied to general micro-scale and pore-scale modeling of microbial motility, we offer a quick example here of how they may be simply applied to Darcy-scale bacterial transport. As discussed in section 1.1.1.2.3 of the dissertation, The 1D ADE for microbial transport in a saturated, homogeneous porous medium can be formulated in simple terms as [46] $R\frac{\partial C}{\partial t} + \frac{\rho_b}{\varepsilon}\frac{\partial S}{\partial t} = D\frac{\partial^2 C}{\partial x^2} - v\frac{\partial C}{\partial x}$, where S is the attached microbe concentration, D is the hydrodynamic dispersion coefficient, v is microbial velocity, R is the retardation factor, ρ_b is the bulk density, and ε is the bed porosity. Given this equation, a simple way to incorporate our results would to have D be a function of v_ϕ (pore water velocity) and motility type. Very roughly, our results imply that D doubles when v increases by a factor of five, and D is 2x greater for bacteria with twitching motility than for those with monotrichous flagella, and the D for bacteria with monotrichous flagella is 2x greater than for those with peritrichous flagella. However, given our lack of variety in species, our use of only two flow rates, and our incomplete understanding of the motility type of *Acidovorax*, more robust studies are needed to confirm the present trends before significant alterations to transport equations are warranted. Furthermore, a more rigorous mathematical understanding of the upscaling of these bacterial dispersion coefficients from micro-scale to Darcy-scale is needed to effectively use them in bioremediation and other field-scale efforts.

While we have tried to provide a robust analysis of bacterial transport in idealized porous media under different flow rates, we also recognize that our study contains many limitations. The

bacteria were difficult to image and required large exposure times, which resulted in low frame rates and significant light scattering around the grains, thus impacting the accuracy of particle tracking. The low frame rates were particularly challenging for situations with high-speed particles (such as Geobacter at 5 μL/h), which highlights the need for tracking methods that can more accurately extract motion statistics from high speed particles. Furthermore, the low frame rate prevented us from analyzing bacteria through traditional run and tumble statistics. We also recognize that a more expansive set of experiments would have included a wider variety of flow rates (especially lower flow rates so that motility driven diffusion is more dominant than flow driven advection), which would allow for more confidence in any trends observed. Also, although we have mainly attributed the differences in transport of our three species to their differences in motility, there are other phenomena, such as the impact of hydrodynamics on different cell lengths (i.e., size exclusion), and DLVO and steric interactions [7], which could offer supporting explanations. Finally, we recognize that a more complete study of the impacts of flow rate on the transport of different bacteria would examine the impact of shear on the ability for monotrichous and peritrichous flagella to bundle/unbundle. These limitations show that there is still significant work to be done to develop a mature theory of bacterial transport in porous media flows.

Data Availability

Besides the raw video and trajectory data, we have provided most of the other data and scripts required to replicate our findings in the supplementary materials. Raw video and trajectory data are available from the corresponding author upon reasonable request. The genome for *Acidovorax* JHL-9 can be found at: https://genome.jgi.doe.gov/portal/AcispJHL9_FD/AcispJHL9_FD.info.html.

Acknowledgments

This research is based upon work supported by the U. S. Department of Energy (DOE) under award number DE-SC0019437. The experiments were performed on a project award (https://doi.org/10.46936/ltds.proj.2020.51218/60006749) from the Environmental Molecular Sciences Laboratory (EMSL), a DOE Office of Science User Facility sponsored by the Biological and Environmental Research program, and were partially supported by the Laboratory Directed Research and Development (LDRD) Program at Pacific Northwest National Laboratory. A portion of these data were produced by the US Department of Energy Joint Genome Institute (https://ror.org/04xm1d337; operated under Contract No. DE-AC02-05CH11231) in collaboration with the user community. The authors would specifically like to acknowledge the micromodel fabrication help received from Dr. Hardeep Mehta at EMSL and microscopy assistance from EMSL's Tom Wietsma. PNNL is a multi-program national laboratory operated for the U.S. Department of Energy (DOE) by Battelle Memorial Institute under Contract No. DE-AC05-76RL0-1830.

References

[1] Persat, A. Nadell, C.D., Kim, M.K., Ingremeau, F., Siryaporn, A., Drescher, K., Wingreen, N.S., Bassler, B.L., Gitai, Z., & Stone, H.A. The mechanical world of bacteria, *Cell* **161**, 988-997 (2015).

[2] Yang, P. & Van Elsas, J. D. Mechanisms and ecological implications of the movement of bacteria in soil. *Applied Soil Ecology* **129**, 112–120 (2018).

[3] Jalili-Firoozinezhad, S., Gazzaniga, F.S., Calamari, E.L., Camacho, D.M., Fadel, C.W., Bein, A., Swenor, B., Nestor, B., Cronce, M.J., Tovaglieri, A. & Levy, O. A complex human gut microbiome cultured in an anaerobic intestine-on-a-chip. *Nat Biomed Eng* **3**, 520–531 (2019).

[4] Mitchell, J. G. & Kogure, K. Bacterial motility: links to the environment and a driving force for microbial physics: Bacterial motility. *FEMS Microbiology Ecology* **55**, 3–16 (2006).

[5] Jakuszeit, T. & Croze, O.A. Role of tumbling in bacterial scattering at convex obstacles. *Physical Review E*, ***109*(4)**, 044405 (2024).

[6] Ugolini, G.S., Wang, M., Secchi, E., Pioli, R., Ackermann, M. & Stocker, R. Microfluidic approaches in microbial ecology. *Lab on a Chip* (2024).

[7] Tokárová, V., Sudalaiyadum Perumal, A., Nayak, M., Shum, H., Kašpar, O., Rajendran, K., Mohammadi, M., Tremblay, C., Gaffney, E.A., Martel, S. & Nicolau Jr, D.V. Patterns of bacterial motility in microfluidics-confining environments. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2013925118 (2021).

[8] Perez, L. J., Bhattacharjee, T., Datta, S. S., Parashar, R. & Sund, N. L. Impact of confined geometries on hopping and trapping of motile bacteria in porous media. *Phys. Rev. E* **103**, 012611 (2021).

[9] Bhattacharjee, T. & Datta, S.S. Confinement and activity regulate bacterial motion in porous media. *Soft Matter*, *15*(48), pp.9920-9930 (2019).

[10] Scheidweiler, D., Miele, F., Peter, H., Battin, T. J. & De Anna, P. Trait-specific dispersal of bacteria in heterogeneous porous environments: from pore to porous medium scale. *J. R. Soc. Interface.* **17**, 20200046 (2020).
[11] Makarchuk, S., Braz, V. C., Araújo, N. A. M., Ciric, L. & Volpe, G. Enhanced propagation of motile bacteria on surfaces due to forward scattering. *Nat Commun* **10**, 4110 (2019).
[12] Perez, L. J., Parashar, R., Plymale, A. & Scheibe, T. D. Contributions of biofilm-induced flow heterogeneities to solute retention and anomalous transport features in porous media. *Water Research* **209**, 117896 (2022).
[13] Ariel, G., Rabani, A., Benisty, S., Partridge, J.D., Harshey, R.M. & Be'Er, A. Swarming bacteria migrate by Lévy Walk. *Nat Commun* **6**, 8396 (2015).
[14] Rusconi, R., Guasto, J. S. & Stocker, R. Bacterial transport suppressed by fluid shear. *Nature Phys* **10**, 212–217 (2014).
[15] Barry, M. T., Rusconi, R., Guasto, J. S. & Stocker, R. Shear-induced orientational dynamics and spatial heterogeneity in suspensions of motile phytoplankton. *J. R. Soc. Interface.* **12**, 20150791 (2015).
[16] Yawata, Y., Nguyen, J., Stocker, R. & Rusconi, R. Microfluidic Studies of Biofilm Formation in Dynamic Environments. *J Bacteriol* **198**, 2589–2595 (2016).
[17] Creppy, A., Clément, E., Douarche, C., D'Angelo, M. V. & Auradou, H. Effect of motility on the transport of bacteria populations through a porous medium. *Phys. Rev. Fluids* **4**, 013102 (2019).
[18] Secchi, E., Vitale, A., Miño, G.L., Kantsler, V., Eberl, L., Rusconi, R. & Stocker, R. The effect of flow on swimming bacteria controls the initial colonization of curved surfaces. *Nat Commun* **11**, 2851 (2020).
[19] Rusconi, R., Lecuyer, S., Guglielmini, L. & Stone, H. A. Laminar flow around corners triggers the formation of biofilm streamers. *J. R. Soc. Interface.* **7**, 1293–1299 (2010).
[20] Rusconi, R., Lecuyer, S., Autrusson, N., Guglielmini, L. & Stone, H. A. Secondary Flow as a Mechanism for the Formation of Biofilm Streamers. *Biophysical Journal* **100**, 1392–1399 (2011).
[21] Pieper, D. H. & Reineke, W. Engineering bacteria for bioremediation. *Current Opinion in Biotechnology* **11**, 262–270 (2000).
[22] Ginn, T.R., Wood, B.D., Nelson, K.E., Scheibe, T.D., Murphy, E.M. & Clement, T.P. Processes in microbial transport in the natural subsurface. *Advances in Water Resources* **25**, 1017–1042 (2002).
[23] Scheidweiler, D., Bordoloi, A.D., Jiao, W., Sentchilo, V., Bollani, M., Chhun, A., Engel, P. & de Anna, P. Spatial structure, chemotaxis and quorum sensing shape bacterial biomass accumulation in complex porous media. *Nature Communications*, ***15*****(1),** 191 (2024).
[24] Song, J., Zhang, Y., Zhang, C., Du, X., Guo, Z., Kuang, Y., Wang, Y., Wu, P., Zou, K., Zou, L. & Lv, J. A microfluidic device for studying chemotaxis mechanism of bacterial cancer targeting. *Sci Rep* **8**, 6394 (2018).
[25] Dentz, M., Creppy, A., Douarche, C., Clément, E. & Auradou, H. Dispersion of motile bacteria in a porous medium. *J. Fluid Mech.* **946**, A33 (2022).
[26] Lee, J.H., Fredrickson, J.K., Plymale, A.E., Dohnalkova, A.C., Resch, C.T., McKinley, J.P. & Shi, L. An autotrophic H_2-oxidizing, nitrate-respiring, Tc(VII)-reducing *Acidovorax* sp. isolated from a subsurface oxic-anoxic transition zone: H_2-oxidizing, Tc-reducing *Acidovorax* spp. *Environmental Microbiology Reports* **7**, 395–403 (2015).
[27] Caccavo Jr, F., Lonergan, D.J., Lovley, D.R., Davis, M., Stolz, J.F. & McInerney, M.J. Geobacter sulfurreducens sp. nov., a hydrogen- and acetate-oxidizing dissimilatory metal-reducing microorganism. *Appl Environ Microbiol* **60**, 3752–3759 (1994).
[28] Ahmed, B., Cao, B., McLean, J.S., Ica, T., Dohnalkova, A., Istanbullu, O., Paksoy, A., Fredrickson, J.K. & Beyenal, H. Fe(III) Reduction and U(VI) Immobilization by Paenibacillus sp. Strain 300A, Isolated from Hanford 300A Subsurface Sediments. *Appl Environ Microbiol* **78**, 8001–8009 (2012).
[29] Krawczyk-Bärsch, E., Gerber, U., Müller, K., Moll, H., Rossberg, A., Steudtner, R. & Merroun, M.L. Multidisciplinary characterization of U(VI) sequestration by Acidovorax facilis for bioremediation purposes. *Journal of Hazardous Materials* **347**, 233–241 (2018).
[30] Methe, B.A., Nelson, K.E., Eisen, J.A., Paulsen, I.T., Nelson, W., Heidelberg, J.F., Wu, D., Wu, M., Ward, N., Beanan, M.J. & Dodson, R.J. Genome of *Geobacter sulfurreducens:* Metal Reduction in Subsurface Environments. *Science* **302**, 1967–1969 (2003).
[31] Govarthanan, M., Mythili, R., Selvankumar, T., Kamala-Kannan, S., Rajasekar, A. & Chang, Y.C. Bioremediation of heavy metals using an endophytic bacterium Paenibacillus sp. RM isolated from the roots of Tridax procumbens. *3 Biotech* **6**, 242 (2016).
[32] Jin, F., Conrad, J.C., Gibiansky, M.L. & Wong, G.C. Bacteria use type-IV pili to slingshot on surfaces. *Proceedings of the National Academy of Sciences*, ***108*****(31),** 12617-12622 (2011).
[33] Speers, A.M., Schindler, B.D., Hwang, J., Genc, A. & Reguera, G. Genetic identification of a PilT motor in Geobacter sulfurreducens reveals a role for pilus retraction in extracellular electron transfer. *Frontiers in microbiology*, **7**, 1578 (2016).
[34] Grady, E.N., MacDonald, J., Liu, L., Richman, A. & Yuan, Z.C. Current knowledge and perspectives of Paenibacillus: a review. *Microbial cell factories*, **15**, 1-18 (2016).

[35] Jayathilake, P.G., Li, B., Zuliani, P., Curtis, T. & Chen, J. Modelling bacterial twitching in fluid flows: a CFD-DEM approach. *Scientific reports*, **9(1),** 14540 (2019).
[36] Nakamura, S. Spirochete flagella and motility. *Biomolecules*, **10(4),** 550 (2020).
[37] Bahar, O., De La Fuente, L. & Burdman, S. Assessing adhesion, biofilm formation and motility of Acidovorax citrulli using microfluidic flow chambers. *FEMS Microbiology letters*, **312(1),** 33-39 (2010).
[38] Wang, S., Jaffe, P.R., Li, G., Wang, S.W. & Rabitz, H.A. Simulating bioremediation of uranium-contaminated aquifers; uncertainty assessment of model parameters. *Journal of Contaminant Hydrology*, ***64*(3-4),** 283-307 (2003).
[39] Yabusaki, S.B., Fang, Y., Williams, K.H., Murray, C.J., Ward, A.L., Dayvault, R.D., Waichler, S.R., Newcomer, D.R., Spane, F.A. & Long, P.E. Variably saturated flow and multicomponent biogeochemical reactive transport modeling of a uranium bioremediation field experiment. *Journal of contaminant hydrology*, ***126*(3-4),** 271-290 (2011).
[40] Liu, Y., Du, M., Shu, S., Wei, J., Zhu, K. & Wang, G. Bacterial surface properties and transport behavior actively respond to an extracellular polymeric substance gradient in saturated porous media. *Science of The Total Environment*, 173889 (2024).
[41] Yang, X., Parashar, R., Sund, N.L., Plymale, A.E., Scheibe, T.D., Hu, D. & Kelly, R.T. On Modeling Ensemble Transport of Metal Reducing Motile Bacteria. *Sci Rep* **9**, 14638 (2019).
[42] Tinevez, J.Y., Perry, N., Schindelin, J., Hoopes, G.M., Reynolds, G.D., Laplantine, E., Bednarek, S.Y., Shorte, S.L. & Eliceiri, K.W. TrackMate: An open and extensible platform for single-particle tracking. *Methods* **115**, 80–90 (2017).
[43] Lovley, D.R., & Phillips, E.J.P. Novel mode of microbial energy metabolism: organic carbon oxidation coupled to dissimilatory reduction of iron or manganese. *Appl. Environ. Microbiol.* **54**, 1472-1480 (1988).
[44] Plymale, A.E., Bailey, V.L., Fredrickson, J.K., Heald, S.M., Buck, E.C., Shi, L., Wang, Z., Resch, C.T., Moore, D.A. & Bolton Jr, H. Biotic and Abiotic Reduction and Solubilization of Pu(IV)O2•xH2O(am) as Affected by Anthraquinone-2,6-disulfonate (AQDS) and Ethylenediaminetetraacetate (EDTA). *Environ. Sci. Technol*. **46**, 2132–2140 (2012).
[45] Lin, W. C., M. V. Coppi, & Dr. R. Lovely. Geobacter sulfurreducens Can Grow with Oxygen as a Terminal Electron Acceptor. *Appl. Environ. Microbiol.* **70 (4)**, 2525-2528 (2004).
[46] Thornton, M.M., Chung-Esaki, H.M., Irvin, C.B., Bortz, D.M., Solomon, M.J. & Younger, J.G. Multicellularity and antibiotic resistance in Klebsiella pneumoniae grown under bloodstream-mimicking fluid dynamic conditions. *The Journal of infectious diseases*, **206(4),** 588-595 (2012).
[47] Tuval, I., Cisneros, L., Dombrowski, C., Wolgemuth, C.W., Kessler, J.O. & Goldstein, R.E. Bacterial swimming and oxygen transport near contact lines. *Proceedings of the National Academy of Sciences*, **102(7),** 2277-2282 (2005).
[48] Rossy, T., Nadell, C.D. & Persat, A. Cellular advective-diffusion drives the emergence of bacterial surface colonization patterns and heterogeneity. *Nature communications*, **10(1),** 2471 (2019).
[49] Dehkharghani, A., Waisbord, N., Dunkel, J. & Guasto, J.S. Bacterial scattering in microfluidic crystal flows reveals giant active Taylor–Aris dispersion. *Proceedings of the National Academy of Sciences*, **116(23),** 11119-11124 (2019).
[50] Einstein, A. On the Movement of Small Particles Suspended in Stationary Liquids Required by the Molecular-Kinetic Theory of Heat. *Annalen Der Physik*, 549–560 (1905).
[51] Duffy, K. J. & Ford, R. M. Turn angle and run time distributions characterize swimming behavior for Pseudomonas putida. *J Bacteriol* **179**, 1428–1430 (1997).
[52] Berg, H.C. & Brown, D.A. Chemotaxis in Escherichia coli analysed by three-dimensional tracking. *Nature*, **239(5374),** 500-504 (1972).
[53] Becker, M. W., Metge, D. W., Collins, S. A., Shapiro, A. M. & Harvey, R. W. Bacterial Transport Experiments in Fractured Crystalline Bedrock. *Ground Water* **41**, 682–689 (2003).
[54] Liu, J., Ford, R. M. & Smith, J. A. Idling Time of Motile Bacteria Contributes to Retardation and Dispersion in Sand Porous Medium. *Environ. Sci. Technol.* **45**, 3945–3951 (2011).
[55] Conrad, J. C. & Poling-Skutvik, R. Confined Flow: Consequences and Implications for Bacteria and Biofilms. *Annu. Rev. Chem. Biomol. Eng.* **9**, 175–200 (2018).
[56] Dehkharghani, A., Waisbord, N. & Guasto, J.S. Self-transport of swimming bacteria is impaired by porous microstructure. *Commun Phys* **6**, 18 (2023).
[57] Ford, R.M. & Harvey, R.W. Role of chemotaxis in the transport of bacteria through saturated porous media, *Advances in Water Resources* **30**(6-7): 1608-1617 (2008).
[58] Bente, K., Mohammadinejad, S., Charsooghi, M.A., Bachmann, F., Codutti, A., Lefèvre, C.T., Klumpp, S. & Faivre, D. High-speed motility originates from cooperatively pushing and pulling flagella bundles in bilophotrichous bacteria. *Elife*, **9**, e47551 (2020).
[59] Najafi, J., Shaebani, M.R., John, T., Altegoer, F., Bange, G. & Wagner, C. Flagellar number governs bacterial spreading and transport efficiency. *Sci Adv* **4**: eaar6425 (2018).
[60] Mitchell, J.G. The energetics and scaling of search strategies in bacteria. *The American Naturalist,* **160(6),** 727-740 (2002).
[61] Kaya, T. & Koser, H. Direct upstream motility in Escherichia coli. *Biophysical journal*, **102(7),** 1514-1523 (2012).

[62] Perez, L.J., Hidalgo, JJ., & Dentz, M. Upscaling of mixing-limited bimolecular chemical reactions in Poiseuille flow. *Water Resources Research* **55**, no. 1: 249-269 (2019).
[63] Puyguiraud, A., Perez, L.J., Hidalgo, JJ., & Dentz, M. Effective dispersion coefficients for the upscaling of pore-scale mixing and reaction. *Advances in Water Resources* **146**: 103782 (2020).
[64] Weber, S.C., Thompson, M.A., Moerner, W.E., Spakowitz, A.J., & Theriot, J.A. Analytical tools to distinguish the effects of localization error, confinement, and medium elasticity on the velocity autocorrelation function. *Biophysical journal*, **102**(11), 2443-2450 (2012).
[65] Weller, H. G., Tabor, G., Jasak, H. & Fureby, C. A tensorial approach to computational continuum mechanics using object-oriented techniques. *Comput. Phys.* **12**, 620 (1998).
[66] OpenFOAM Foundation. OpenFOAM: Open-source CFD software. Retrieved from https://openfoam.org/ (2023).
[67] Vennamneni, L., Nambiar, S., & Subramanian, G. "Shear-induced migration of microswimmers in pressure-driven channel flow." *Journal of Fluid Mechanics* 890: A15 (2020).
[68] Lee, M., Lohrmann, C., Szuttor, K., Auradou, H. & Holm, C. The influence of motility on bacterial accumulation in a microporous channel. *Soft Matter*, **17(4),** 893-902 (2021).
[69] Jeffery, G.B. The motion of ellipsoidal particles immersed in a viscous fluid. *Proceedings of the Royal Society of London.* Series A, Containing papers of a mathematical and physical character, **102(715),** 161-179 (1922).
[70] Pedley, T.J. & Kessler, J.O. Hydrodynamic phenomena in suspensions of swimming microorganisms. *Annual Review of Fluid Mechanics*, **24(1),** 313-358 (1992).
[71] Choudhary, A., Paul, S., Rühle, F. & Stark, H. How inertial lift affects the dynamics of a microswimmer in Poiseuille flow. *Communications Physics*, **5(1),** 14 (2022).
[72] Yang, J., Kikuchi, K. & Ishikawa, T. High shear flow prevents bundling of bacterial flagella and induces lateral migration away from a wall. Communications Physics, **6(1),** 354 (2023).
[73] Licata, N.A., Mohari, B., Fuqua, C. & Setayeshgar, S. Diffusion of bacterial cells in porous media. *Biophysical journal*, **110(1),** 247-257 (2016).
[74] Alonso-Matilla, R., Chakrabarti, B. a& nd Saintillan, D. Transport and dispersion of active particles in periodic porous media. *Physical Review Fluids*, **4(4),** 043101 (2019).
[75] Weiss, T.H., Mills, A.L., Hornberger, G.M. & Herman, J.S. Effect of bacterial cell shape on transport of bacteria in porous media. *Environmental science & technology*, **29(7),** 1737-1740 (1995).
[76] Fontes, D.E., Mills, A.L., Hornberger, G.M. & Herman, J. Physical and chemical factors influencing transport of microorganisms through porous media. *Applied and Environmental Microbiology,* **57(9),** 2473-2481 (1991).
[77] F. Grant Ferris; Natalie Szponar; & Brock A. Edwards. Groundwater Microbiology. *The Groundwater Project* (2021).
[78] Molaei, M. & Sheng, J. Succeed escape: Flow shear promotes tumbling of Escherichia colinear a solid surface. *Sci Rep* **6**, 35290 (2016).
[79] Johnson, W. P., Blue, K. A., Logan, B. E. & Arnold, R. G. Modeling Bacterial Detachment During Transport Through Porous Media as a Residence-Time-Dependent Process. *Water Resour. Res.* **31**, 2649–2658 (1995).
[80] Dong, H., Scheibe, T. D., Johnson, W. P., Monkman, C. M. & Fuller, M. E. Change of Collision Efficiency with Distance in Bacterial Transport Experiments. *Ground Water* **44**, 415–429 (2006).
[81] Hermansson, M. The DLVO theory in microbial adhesion. *Colloids and Surfaces B: Biointerfaces* **14**, 105–119 (1999).
[82] de Anna, P., Pahlavan, A.A., Yawata, Y., Stocker, R. & Juanes, R.. Chemotaxis under flow disorder shapes microbial dispersion in porous media. *Nat. Phys.* **17**, 68–73 (2021).
[83] Son, K., Brumley, D.R., & Stocker, R. Live from under the lens: exploring microbial motility with dynamic imaging and microfluidics. *Nature Reviews Microbiology*. **13**, 761–775 (2015).
[84] Lu, N., Bevard, T., Massoudieh, A., Zhang, C., Dohnalkova, A.C., Zilles, J.L. & Nguyen, T.H. Flagella-Mediated Differences in Deposition Dynamics for Azotobacter *vinelandii* in porous media. *Environ. Sci. Technol*. **47**, 5162–5170 (2013).

Chapter 3: Evaluation of particle tracking codes for dispersing particles in porous media

3.1 Abstract

Particle tracking (PT) is a popular technique in microscopy, microfluidics and colloidal transport studies, where image analysis is used to reconstruct trajectories from bright spots in a video. The performance of many PT algorithms has been rigorously tested for directed and Brownian motion in open media. However, PT is frequently used to track particles in porous media where complex geometries and viscous flows generate particles with high velocity variability over time. Here, we present an evaluation of four PT algorithms for a simulated dispersion of particles in porous media across a range of particle speeds and densities. Of special note, we introduce a new velocity-based PT linking algorithm (V-TrackMat) that achieves high accuracy relative to the other PT algorithms. Our findings underscore that traditional statistics, which revolve around detection and linking proficiency, fall short in providing a holistic comparison of PT codes because they tend to underpenalize aggressive linking techniques. We further elucidate that all codes analyzed show a decrease in performance due to high speeds, particle densities, and trajectory noise. However, linking algorithms designed to harness velocity data show superior performance, especially in the case of high-speed advective motion. Lastly, we emphasize how PT error can influence transport analysis.

3.2 Introduction

Particle Tracking (PT) employs detection and linking algorithms to reconstruct the trajectories of objects within time-lapse image data. PT has vast applicability, spanning across any video data with moving entities, and it is one of the principal methods used to decipher microscale

transport processes. This includes phenomena such as particle diffusion [1–3], nano and micro particle transport in saturated [4, 5] or multiphases flow [6], bacterial dispersion [7–17] , chemotaxis [18], biofilm formation [19, 20], viral transport [12], transport in porous media [8–11, 21], colloid filtration [22], and computation of DLVO interactions through accurate trajectory analysis and measurements of hindered diffusion [23, 24]. For a granular understanding of these processes, the precision and speed of PT are paramount.

The PT landscape boasts a plethora of open source and proprietary codes [25]. This multitude underscores the pressing need for robust comparative analyses between codes. Notably, the seminal comparative study in this domain centered on particles exhibiting Brownian and directed motion within open media [26]. However, a void persists in the exploration of PT methods tailored for particles navigating porous media flows. In these flows, spatial confinement (i.e. obstacles or grains) and dispersion result in complex flow paths, leading to pronounced variability in velocity fields over small temporal and spatial scales [27]. We should note here that particle tracking, in the context used throughout this chapter, refers to time lapse image acquisition and subsequent particle identity assignment between consecutive frames [28].

The common strategy to reconstruct single particle trajectories by time lapse image acquisition first requires the particle detection in a single image at each frame and, then, the particle identity assignment (also termed as pairing or linking assignment) between particles detected in two consequent frames. Thus, key challenges in PT revolve around detection, localization, and linking errors [26, 29]. Detection errors often stem from overlapping particles, particles out of focus, particles indistinguishable from the background, or varied particle sizes [30, 31]. Linking errors can similarly be attributed to a confluence of factors: high particle speeds and densities, algorithmic inaccuracies, and preexisting detection errors [32]. Localization error

predominantly emerges from the detection algorithm and the signal-to-noise ratio of the imagery [26]. Linking errors generally represent the most significant source of error, although large errors can occur for images where detection is especially difficult. Localization errors, which are often sub-pixel, minimally influence PT performance for a large group of particles.

Here, we compare the performance of four PT linking algorithms for a simulated dispersing of particles in porous media. To understand the impacts across a range of particle speed distributions, we simulated tracer particle dispersion in two different porous geometries (further discussed in Methods section). A critical facet of particle tracking is the particle spacing displacement ratio [33], $PSDR = \frac{ipd}{U\Delta t}$, where ipd is the average inter-particle distance detected within each frame (averaged over all frames), U is the average particle speed, and Δt is the time interval between two consequent frames that corresponds to the inverse of the frame rate. Hence, PSDR is a measure of the mean particle spacing relative to the average jump length of particles between frames. This statistic can be considered a general constraint on the strength of PT algorithms, as it is directly related to the number of probable links each particle can make with other particles. For $PSDR << 1$, PT has been shown to be extremely challenging [34], and high particle densities have been show to increase the sensitivity of PT parameters [35]. In fact, one of the primary motivations of this chapter is to develop a rigorous understanding of particle tracking in these challenging scenarios, given that we observed jumps of up to 120 pixels per frame in our videos of Geobacter. As particles get very close together or have large displacements between frames, PT algorithms are not able to confidently determine accurate links to respective trajectories between frames. Thus, our analysis covers multiple mean speeds, particle densities, and speed distribution shapes to gauge PT codes across varied PSDRs (Table 1). Our approach not only evaluates PT codes for porous media, but also refines the standard PT comparison

benchmark. We show that "classical statistics," which exclusively focus on particle localization, detection, and linking, might not penalize aggressive linking adequately. Specifically, classical statistics aren't effective for understanding the error associated with PT algorithms that "force" links between trajectories under improbable circumstances. To provide a more accurate comparison between PT methods, we use a suite of experimental statistics that offer significant depth in understanding of PT results compared to classical statistics. Moreover, we shed light on the potential for PT error to skew transport analyses of tracer particles in porous media.

Concomitantly, we unveil a novel PT code (V-TrackMat), tailored for microfluidics experiments, crafted by the collaborative efforts of some of the coauthors. The other three algorithms we tested for this chapter (Trackpy, TrackMate-LAP, and TrackMate- Kalman) are described in further detail in the methods section. Our findings highlight V-TrackMat's ability to strike a balance between accuracy and judicious tracking at the expense of speed. Through this exploration, we endeavor to amplify the discourse in PT – emphasizing the shortcomings of traditional metrics, unveiling the intricacies of various PT algorithms, illuminating the impacts of PT error on transport analysis, and introducing a robust PT method.

The work presented in the chapter is an enhanced version of the published article: "Berghouse, M., Miele, F., Perez, L.J., Bordoloi, A., Morales, V.L., & Parashar, P. Evaluation of particle tracking codes for dispersing particles in porous media. *Sci Rep* **14**, 24094 (2024). https://doi.org/10.1038/s41598-024-75581-0". This chapter builds on the work presented in the published paper by establishing the motivation through reflections on chapter 2, discussing the generation of pathlines in greater detail, discussing how the limitations of particle tracking may have impacted the results of chapter 2, and presenting a rudimentary comparison of different PT methods for experimental data.

Table 1. PSDR values for all simulations in this chapter.

Bimodal	S_p **= 0.9**	S_p **= 2.6**	S_p **= 9.9**	S_p **= 19.6**
$\rho_p \approx 1.25e^{-4}$	42.8	15.5	4.6	2.5
$\rho_p \approx 2.50e^{-4}$	24.5	10.4	3.0	1.7
$\rho_p \approx 5.00e^{-4}$	14.9	6.6	2.3	1.1
Unimodal	S_p **= 0.8**	S_p **= 1.8**	S_p **= 6.7**	S_p **= 11.3**
$\rho_p \approx 1.25e^{-4}$	34.3	17.4	5.3	3.1
$\rho_p \approx 2.50e^{-4}$	22.8	12.6	4.0	2.4
$\rho_p \approx 5.00e^{-4}$	15.8	8.9	2.6	1.6

3.3 Methods

3.3.1 Simulations

PT codes have been extensively compared for Brownian motion and constant-velocity motion in open media [26]. However, both 3D natural [36] and 2D engineered porous materials [37] are characterized by complex pore structures that result in broadly distributed velocity fields which are known to challenge tracking algorithms, so we chose to simulate dispersing particles in porous media for our comparison. We used simulations to create our ground truth imagery and trajectories, because although the gold standard for PT comparison is experimental data, there are no manually-labeled videos of dispersing particles in porous media that can be used as the ground truth.

For each of the 2D geometries (described in the paragraph below), we used OpenFOAM [38] to solve the flow fields, then calculated the pathlines via the Matlab function "interpstreamspeed" (Fig. 1a). The “interpstreamspeed” function in MATLAB refines streamline vertices by interpolating them based on the speed of a vector field, accommodating both two-dimensional (2D) and three-dimensional (3D) datasets. Given spatial coordinates (X, Y, Z) on a

regular grid, velocity components (U, V, W), and an initial set of streamline vertices V_i, the function calculates or accepts a speed field S, defined as S = $\sqrt{U^2 + V^2}$ to guide vertex density.

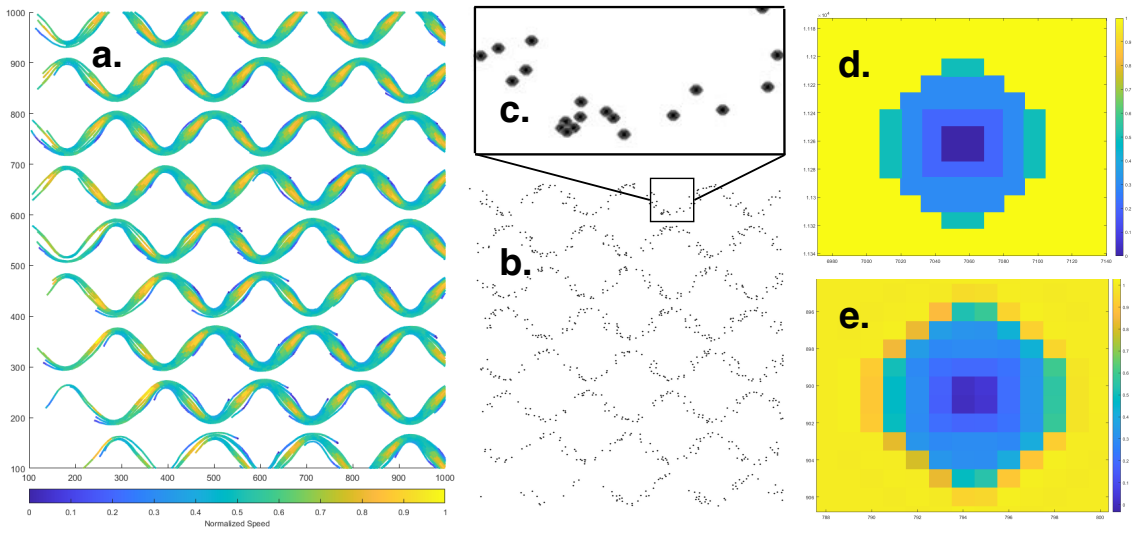


Figure 1. This figure shows the general workflow for the construction of simulated trajectories. (a) Pathlines for the bimodal simulation at *PSDR* = 1.1 colored by normalized speed. (b) Simulated imagery analyzed by various particle tracking methods. (c) Zoomed in section of simulated particles showing overlapping particles. Most issues with detection occur due to this overlapping, which results in false negatives. (d) Particle at full resolution (10000x10000) during simulation creation. (e) Particle at final resolution (2000x2000) after interpolation.

The speed at each vertex is interpolated using MATLAB's "interp3" (for 3D) or "interp2" (for 2D) functions, yielding speed values S_i at each streamline position, which are scaled by a user-defined factor SF to control vertex spacing. The function then resamples vertex positions by computing cumulative distances $d_i = \sum_{j=1}^{i} |V_j - V_{j-1}|$ along the streamline and placing new vertices P_k at fractional positions based on interpolated speeds, such that $d_k = S_i \cdot SF$ governs the spatial resolution. The resulting output is a refined streamline that adapts to local velocity variations, producing denser sampling in regions of slower flow and sparser sampling in regions of faster flow. The obtained flow fields exhibited large variations in speed distributions, resulting in simulations that we termed as "bimodal" and "unimodal". The pathlines were seeded with

relatively equally spaced particles throughout the whole domain (with minor fluctuations due to grain positions), and their motion along the pathlines over time resulted in the final ground truth imagery and trajectories (Fig. 1b). A small amount of random movement (Gaussian distribution with μ = 0 and σ = 0.5 pixels) was added on top of the ground truth advective trajectories. This amount of random motion can be considered a reasonable representation of diffusion in advection-dominated conditions ($Pe \approx 500$). Given a characteristic length of about 61.5 pixels (roughly equal to the average pore throat length), and a mean particle speed of between 2.6 and 10.2 px/frame, we can calculate $D = \frac{[2.6,10.2]*6.15}{500}$, which gives us a D in the range of $[0.32, 1.25]$ px^2/frame. Because we wanted to primarily focus on the linking abilities of different PT algorithms, we removed the background (cylinders of the geometry). However, due to overlapping particles in our simulations (Fig. 1c), we weren't able to completely remove the influence of detection accuracy. The particles with a diameter of 8 pixels (for a 2000x2000 pixel domain) were defined with a 2x2 black dot in the center with slightly increasing brightness for the surrounding pixels (Figs. 1d and 1e).

Our bimodal simulations were derived from experimental digital microscopy imagery of a quasi-2D porous media microfluidic device. The geometry of the microfluidic device, and thus our bimodal simulations, consists of a staggered array of equally sized and spaced grains (Figs. 1a and 2c), also referred to as a microfluidic lattice [8]. The unimodal simulation was derived from a geometry with similar average porosity, but with random placement and sizing of grains and pore throats (Fig. 2d). We term the first group of simulations as "bimodal" (Fig. 2a), because there are two clear peaks in the speed distributions for $\overline{S_{sim}} \geq 2.6$. We term the other group of simulations as "unimodal" (Fig. 2b), because there is only one clear peak at all mean speeds.

In addition to examining the impact of particle speed distribution on PT performance, we also vary the number of particles in our simulations. We use the initial number of particles at the beginning of each simulation to calculate the particle density (particles/pixels^2). Particle densities decrease over time as individual particle trajectories disappear or exit the simulation bounds.

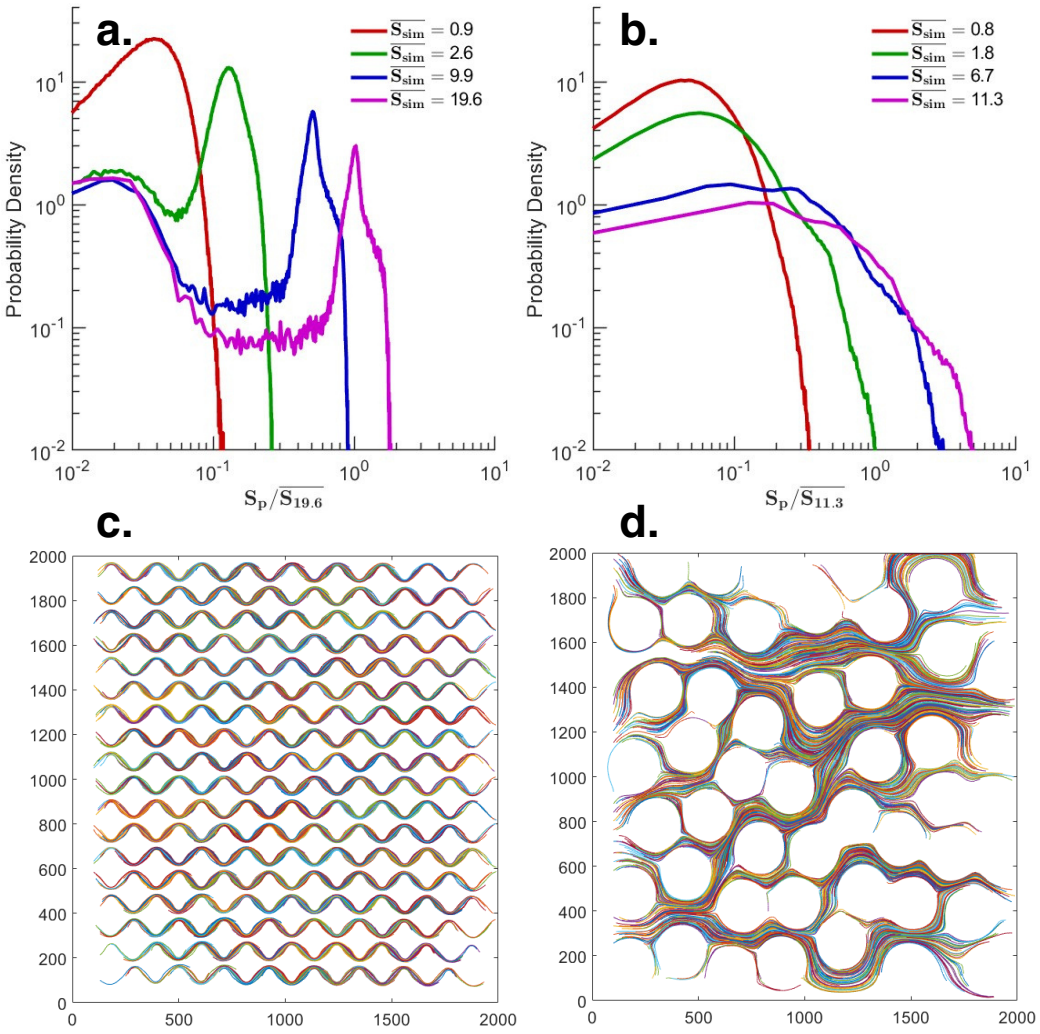


Figure 2. This figure shows the particle speed distributions for the bimodal (a) and unimodal (b) simulations, and the simulated pathlines for a bimodal (c) and unimodal (d) simulation for *PSDR* = 2.3 and *PSDR* = 2.6 respectively. The speed distributions in (a) are normalized by the mean speed of the lowest-PSDR bimodal simulation (19.6 pixels/frame). The speed distributions in (b) are normalized by the mean speed of the lowest-PSDR unimodal simulation (11.3 pixels/frame). The simulated pathlines and variety of speed distributions illustrate the large range of conditions our PT codes were tested in.

The primary goal of this chapter is to provide a rigorous comparison of different linking algorithms for simulations of dispersing particles in varying geometry and at different particle density and speed. Although this does not cover the full range of variability expected in videos of dispersing particles, we chose to focus on PSDR and geometry because they highlight differences in the linking capabilities of each PT code. However, performance of PT codes can also be impacted because of limitations of methods and devices to capture and process videos of dispersing particles. These limitations could potentially inject random fluctuations and intermittency (blinking) in particle positions. Thus, we also provide an analysis of simulations where we increase the magnitude of the random displacement on top of our purely dispersing particles, and randomly set 2% of particles to be invisible for each frame. Specifically, we enhance the random motion by using $\sigma = 2$ pixels (instead of 0.5) in the method to represent normally distributed weak diffusion as described above. The displacements were restricted to be less than 6 pixels (three standard deviation). This random motion accounts for the combined effects of diffusion, camera jitter, and oscillations in particle brightness. The intermittency generally accounts for particles moving in and out of the focal plane of the camera, which can be caused by a variety of phenomena such as diffusion, particle-particle interactions, particle-wall interactions, and camera exposure time. The magnitude of the intermittency (2%) was calculated through an analysis of microfluidic experiments of colloids (detailed methods provided in Supplementary Methods). Both of these additions can be thought of as increasing the noise of the trajectories, so in this chapter we refer to these experiments as the "noisy simulations". All other simulations in the chapter, as described above, contain a minimal amount of noise to ensure tracking capabilities. We thus refer to the main simulations as the "minimal-noise" simulations when comparing their results with those of the noisy simulations.

3.3.2 Particle Tracking

In order to focus on the depth of our PT comparisons, we chose to analyze four PT algorithms. For this study, we compared the outputs of TrackMate [39,40], Trackpy [41], and a new PT method developed by our co-authors named "V-TrackMat". We chose TrackMate (TM) and Trackpy (TP) due to their high popularity in bio-image analysis, and because they use different linking algorithms. Each algorithm tested in this dissertation chapter uses a nearest-neighbors-based method to link particles. To try and expand the variety of investigated algorithms, we also tested two deep learning (DL) methods [42, 43], but found that they either couldn't be run on our hardware (more than 8 GB VRAM or too slow for CPU-based models) or did not perform as well as TM, TP, and V-TrackMat. Likely, the high resolution of our images (2000x2000 pixels), and small size of our particles (8 pixels) precludes the effective use of convolutional networks and other common DL-based architectures. However, we did not test any standard models for object detection such as Yolov8 or FairMOT [44], so it is possible there are available architectures that outperform the traditional methods. A more rigorous investigation of all available DL models is needed to determine state of the art performance, and thus develop an accurate comparison between traditional and DL-based methods.

TrackMate is one of the most popular methods for particle tracking in the field of biological imaging. TM runs through ImageJ [45], which makes it challenging to script PT analysis. However, TM has shown high levels of accuracy [26], and its use within ImageJ means it is well suited for quick analytical workflows where visual inspection of results is necessary. TM allows the user to pick from a variety of linking methods, but we only chose to analyze the Kalman and Linear Assignment Problem (LAP) methods. The Kalman method [46] uses the autocorrelative tendencies of trajectories to predict the velocities of particles, and therefore their positions in subsequent

frames. The LAP method creates a cost matrix that finds the best match for each particle between two frames [47]. The cost matrix can be assigned additional variable-specific penalties that can improve linking, although this feature was not explored in our study. Both methods allow for gap-filling of particles that were missed in one frame and appeared up to a threshold number of frames after they were lost. TM also offers advanced filtering options that allows for easy removal of trajectories appearing to be erroneous. For example, TM allows for filtering based on the number of spots, track length, mean, min, and max speed, directional change rate, and linearity of forward progression.

Trackpy (TP) is another popular PT method written in Python, which makes it generally more scriptable than TM. However, this also means that some programming abilities are needed to effectively use this PT method. Furthermore, many of the trajectory filtering and visualization features that TM has are not part of the TP API, and would need to be manually coded from scratch. One area that TP excels in is its analysis functions. The API has functions to calculate pair correlations, MSDs, particle drift, van Hove correlations, and velocity correlations (amongst other functions). The linking algorithm for TP is based off the Crocker-Grier algorithm [48], one of the fundamental algorithms that many PT codes use in some variation. TP also has a special linking function that incorporates some velocity prediction element ("NearestVelocityPredict"), which was used for all our experiments. This velocity-based linking algorithm differs from the Kalman filter in that the Kalman filter considers the history of a trajectory (accounts for velocity variation in time), whereas the TP linking algorithm considers the velocity of the nearest particle (accounts for velocity variation in space).

V-TrackMat is a new Matlab-based PT method developed by some of our co-authors that has been successfully applied to tracking particles in three-dimensional and bioclogged

environment [49, 50]. The version of the code used in this chapter can be found at https://github.com/mberghouse/V-TrackMat. The development of this code was motivated by the need for a customized MATLAB-based code for 2D and 3D PT to both allow a secondary linking phase between anachronistic trajectories and further overcome the current limitations of TM and TP for crowded suspension and long-time image acquisitions. Indeed, accuracy of TP has been reported to suffer for crowded suspension [51] while TM resulted in several crash episodes during the linking step for benchmark experiments of particle tracking in microfluidics-disordered media for a total number of frames above 2000 at $PSDR \cong 1.4$. Although V-TrackMat has been developed for PT of 1 µm diameter latex particles in microfluidics application under laminar flow, it can be used for a variety of other PT applications. It first uses a nearest-neighborhood criteria by calling the ipdm routine between coordinates of centers detected in two consecutive frames, named as parents for particles detected at frame n and daughter for frame n + 1. The pairing has been optimized by assuming that the frame rate is high enough so that the mean particle's jump is lower than the mean inter-particle distance, mitigating the effect of intermittent behavior of a single particle's velocity under flow in confined media. Thus, the 2-frame velocity is computed for each pairing and the ipdm function is computed between daughters and the projection of the future position for the parents displaced by the quantity v dt along the tangent direction. After the first loop over the full set of frames is computed, the set of reconstructed trajectories is then processed by gluing anachronistic trajectories in the 2D+2D space. This is reasonable for high Pe and low Re where particles passing through the same position with the same velocity are, in fact, following the same streamline. To glue trajectories, a pairing was first assigned by minimizing distances between parent ending points and daughter starting points. The glue is accepted only for pairing whose distance is compatible with a jump allowed by both parents ending velocity and

daughter starting velocity. For multiple pairings, the criteria of the minimum in of $\left|v_{p_f} - v_{d_i}\right|$ is then applied. This second loop of gluing anachronistic particles can potentially be iterated multiple times until no new pairings are assigned. The algorithm can be applied to track particles from time lapse images acquired over different fields of view. A common challenge for any PT code is the increasing memory cast with both the increasing time and new incoming particles as a new ID must be assigned while keeping track of the already existing ones. This means that the size of the matrix composed by the number of trajectories by the number of frames increases linearly with time and flow rate. To save computing time and avoid memory dredge, V-TrackMat code considers particles as lost if, for 5 consequent frames, no pairing has been assigned. In this case, V-TrackMat saves the trajectories on the fly into binary files and finally removes them from the matrix.

Although our goal in this chapter was to test the ability of the linking algorithms for each code, it worth noting that detection is an important part of the PT workflow. V-TrackMat's detector was specifically designed to find bacteria in particular sets of experimental imagery, and we found that it did not perform well for the simulated imagery used in this chapter. Both TP and TM use robust detection methods, and no significant differences were observed from each of the detection methods upon initial inspection. For V-TrackMat, we detected spots with TM, then exported these spots to Matlab to perform linking with V-TrackMat.

3.3.3 Comparative Metrics

To understand the true performance of each PT method, we used multiple types of comparative metrics. We used "classical" statistical methods, which are similar to those described in [26]. In addition, we used experimental statistics and visual analysis of trajectories. The classical statistics used in this study are the false link rate (Fig. 3a), mean path length (Fig. 3b), and the Euclidean

distance (Fig. 3c). These statistics are aimed at diagnosing basic problems with particle detection, linking frequency, and linking accuracy (the percent of correct links between two frames). Since we simulate the movement of tracer particles in a microfluidic device under constant flow, we also chose to use experimentally relevant statistics as PT comparison metrics.

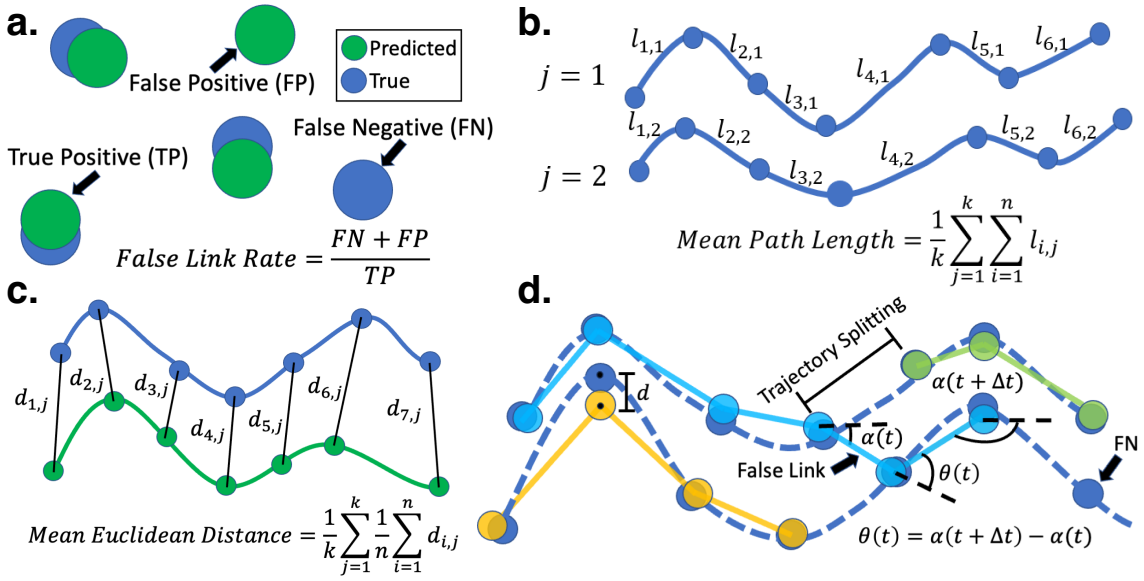


Figure 3. This figure shows visual representations of all classical statistics, and one experimental statistic, used to compare particle tracking performance in this chapter. In all panels, predicted points and trajectories are green, and ground truth points and trajectories are blue. (a) Snapshot of particles in a single frame to illustrate the false link rate (FLR). A false positive implies an erroneous point that was linked to another point in a previous frame. A false negative implies a missing link for a true point. Thus, FLR tests particle detection and missed links on a frame-by-frame basis. (b) Mean path length (MPL), shown here for two ground truth trajectories of length 7. Subscript i represents each successive line segment in time, and subscript j represents each trajectory. MPL tests missed links over time, also known as trajectory splitting. (c) Mean Euclidean distance (ED), shown here for one ground truth trajectory and one predicted trajectory (each of length 7). ED tests localization, detection, and linking accuracy (but in our case of simple particles and no background, primarily linking accuracy). To compute this statistic, we search for ground truth trajectories that are on average less than two pixels away from the predicted trajectory. From the set of matching trajectories, such as those pictured in the figure, we can calculate the ED. (d) Diagram of two ground truth trajectories (shown in dark blue) with potential PT-based trajectories (shown in light blue, yellow and green, with each color signifying a different trajectory predicted by a PT method) on top to illustrate the classical errors (and concept of the turn angle $\vartheta\,(t)$) discussed in this chapter.

To compare each PT code for the specific scenario of microbial transport, we used velocity autocorrelations $C_v(\tau) = \langle |v_n|(t+\tau) \cdot v(t) \rangle$, speed-angle joint probability density contours, which, are calculated from the particle speed $v_p = \frac{\sqrt{(x_{t+1}-x_t)^2+(y_{t+1}-y_t)^2}}{\Delta t}$ and turn angle $\alpha_t = tan^{-1}\left(\frac{y_{t+2}-y_{t+1}}{x_{t+2}-x_{t+1}}\right) - tan^{-1}\left(\frac{y_{t+1}-y_t}{x_{t+1}-x_t}\right)$ data, and mean square displacements $MSD(t) = \frac{1}{N}\sum_{i=1}^{N}|r_i(t) - r_i(0)|^2$ (Fig. 3d) as our experimental metrics. The MSD and C_v for each PT output were calculated via the MSDAnalyzer [52], a companion postprocessing program for TrackMate. All other statistics were calculated via scripts written in-house.

3.4 Results

3.4.1 Trajectory Patterns Illuminate PT Method Variations

To understand PT performance at a visual level, we plotted a small window of trajectories for the bimodal and unimodal simulations for PSDR $\leq$ 1.7 (Fig. 4). Some selected trajectories for PSDR $\geq$ 3.0 are also given in Supplementary Fig 3. For the bimodal simulation at PSDR = 3, trajectories from all PT codes reasonably mirror the simulations. However, this congruence quickly diminishes at PSDR = 1.7 (Fig. 4b), with marked deviations underscoring the nuances of each linking algorithm. It's noteworthy that for PSDR $\leq$ 1.7 (Figs. 4a and 4b), trajectories close to the cylinders (i.e., slow trajectories) are detected more accurately compared to the faster trajectories in the pore throat. In addition to increased speed, particles in the pore throat exhibit spatial convergence, which results in a large local decrease in PSDR. This observation carries over to the unimodal simulations (Fig. 4c), which exhibit consistent patterns across PT codes. A pervasive trend emerges: PT codes tend to underestimate the likelihood of particles moving at high

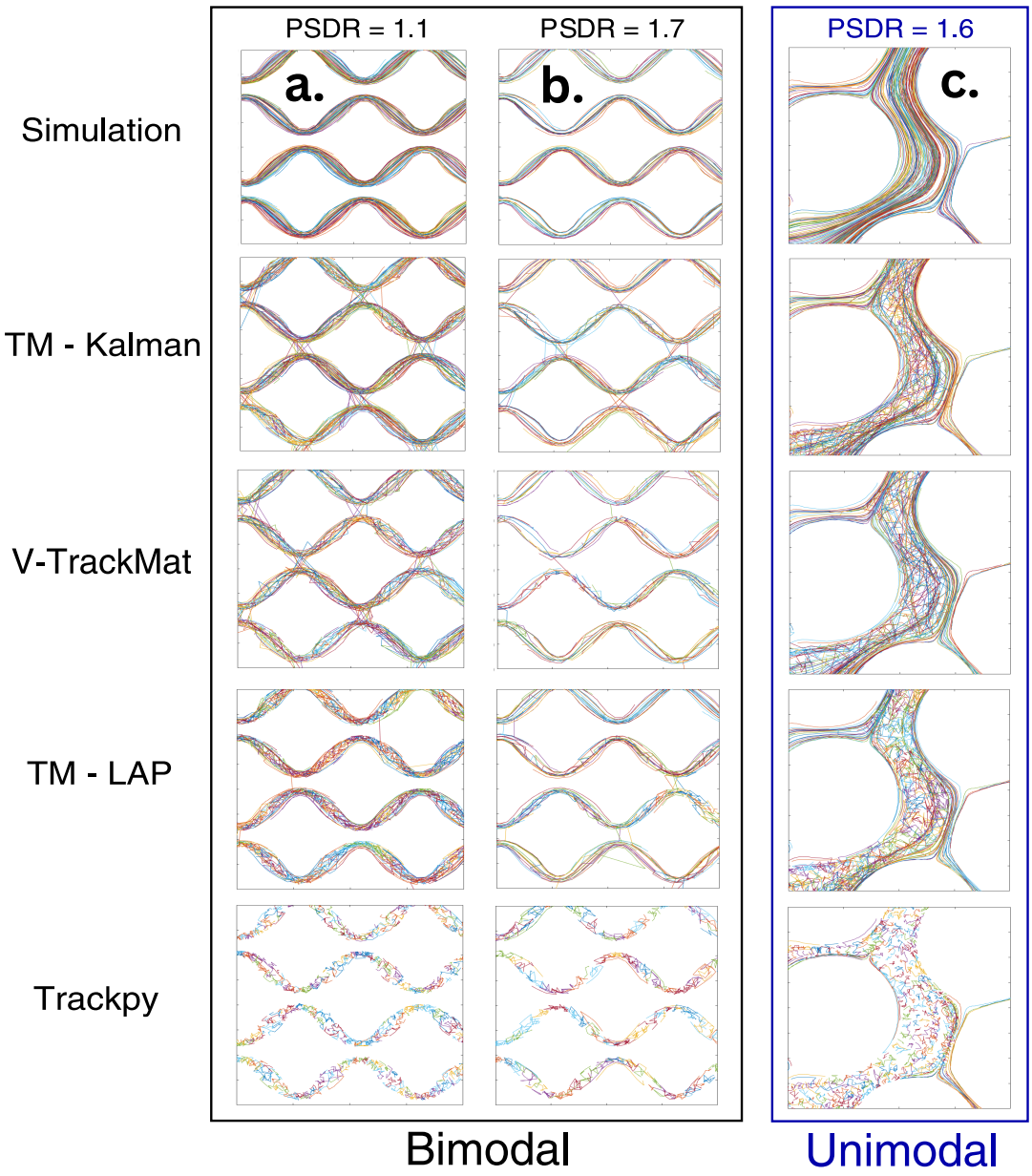


Figure 4. Sample trajectories for all PT codes for the low-PSDR bimodal and unimodal simulations. Each plot shows a 400x400 pixel section of the whole domain. Within a plot, each line corresponds to a unique trajectory (with random colors used to show the contrast between individual trajectories). (a) Bimodal simulations for *PSDR* = 1.1. (b) Bimodal simulations for *PSDR* = 1.7. (c) Unimodal simulations for *PSDR* = 1.6. All algorithms suffer from trajectory splitting and erroneous linking for these low-PSDR simulations. TM-Kalman, TM-LAP and V-TrackMat clearly outperform TP in all scenarios. Although TM-Kalman and V-TrackMat have more clearly false links that stretch across the pore space (jump from one group of streamlines to another), TM-LAP has a much larger amount of zig-zagging trajectories caused by erroneous links between close particles. The low-PSDR unimodal simulations generally show the same trends as the bimodal simulations; however, the differences between TM-Kalman, TM-LAP, and V-TrackMat are less significant.

velocities for a variety of simulated speed distributions, especially when spatial convergence further reduces PSDR.

Probing deeper into individual PT code performances for the bimodal simulations, especially at lower simulated PSDRs, TM–Kalman stands out with superior accuracy, although it's not exempt from erroneous links at elevated speeds. In particular, TM-Kalman shows a significant amount of erroneous long links across pore spaces (streamlines don't cross the pore space in our bimodal simulations, so any link across the pore space is a false link). TP shows the greatest amount of false links and split trajectories (further explained in next section) at PSDR $\leq$ 1.7 (Figs. 4a and 4b). TM-Lap similarly exhibits pronounced difficulties in linking high-speed particles, though not as significant as TP. At PSDR = 1.7, V-TrackMat trajectories generally resemble those of TM-LAP and TM-Kalman in terms of accuracy, although there are a smaller number of V-TrackMat trajectories. At PSDR = 1.1 (Fig. 4a), TM-Kalman and V-TrackMat appear to outperform TM-LAP. Although TM-Kalman and V-TrackMat may have more erroneous links across the pore space, TM-LAP has a much greater number of zig-zagging trajectories (links going back and forth between two or more different particles), and less true trajectories that last a significant distance. Thus, although V-TrackMat's linking algorithm is less aggressive than either TM algorithm, V-TrackMat captures a substantial portion of accurate trajectories.

The unimodal simulations at PSDR = 4 (Supplementary Figure 3) further show that all algorithms besides TP have robust performance regardless of geometry. At PSDR = 1.6 (Fig. 4c), all algorithms show problems with false links and split trajectories. Similar to the bimodal results, V-TrackMat and TM-Kalman seem to have a larger amount of accurate trajectories than TM-LAP. Thus, general algorithm performance is largely independent of the geometry in which the particles are tracked. However, it should be noted here that the range of possible particle speeds in our

simulations, which is largely impacted by geometry and flow conditions, only spans 3-4 magnitudes (Figs. 2a and 2b). High fidelity simulations of Lagrangian particles in porous media show speed distributions that range up to 8 orders of magnitude [53], so we can't be confident that our findings (relative rankings of PT performance) would remain accurate for transport in any geometry or flow condition. Furthermore, to focus on linking, we didn't include any background. However, in real experiments that image bacteria in microfluidic devices, the geometry has a significant impact on tracking performance due to the presence of light scattering around grains [21].

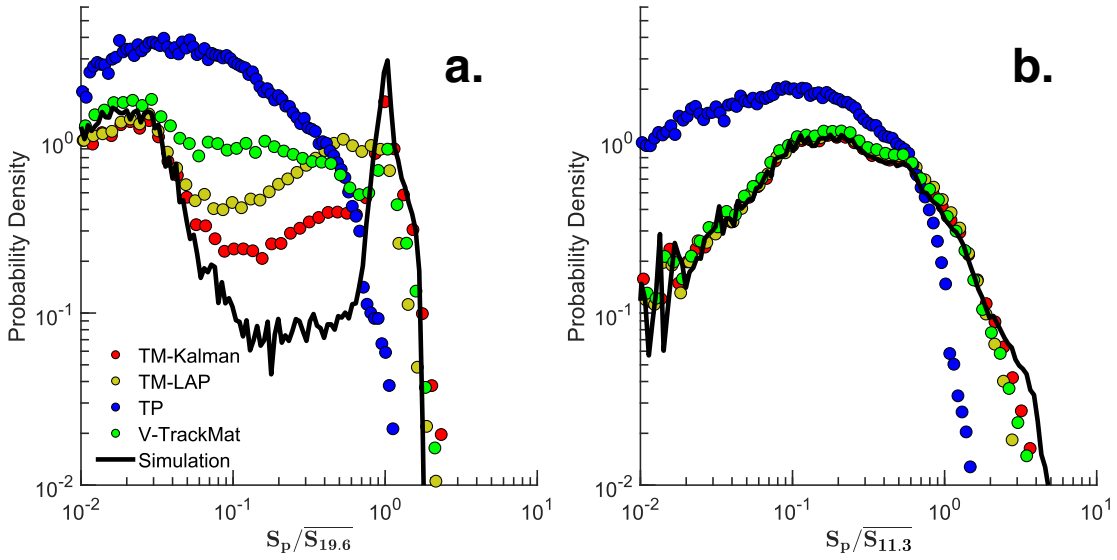


Figure 5. This figure shows the a comparison between the speed distributions for each PT code for the lowest-PSDR bimodal (a) and lowest-PSDR unimodal (b) simulations. The speed distributions in (a) are normalized by the mean speed of the lowest-PSDR bimodal simulation (19.6 pixels/frame). The speed distributions in (b) are normalized by the mean speed of the lowest-PSDR unimodal simulation (11.3 pixels/frame). This figure shows the ability of each PT code to handle significantly different distributions of particle speeds.

To further understand differences in our PT codes, we plot both the simulated (ground truth) and PT-generated normalized speed distributions for our lowest PSDR bimodal (Fig. 5a) and unimodal (Fig. 5b) simulations. To quantify these differences, we calculate the 1-Wasserstein

distance (W_1) between each ground-truth and tracked speed PDF (Table 2). Visual inspection of the PDFs, as well as the trends in (W_1), indicate that TM-Kalman is able to reproduce the simulated speed distributions the best, followed by V-TrackMat, then TM-LAP, then TP. Interestingly, each PT code besides TP overpredicts the fastest speeds for the bimodal simulation, but underpredicts the fastest speeds for the unimodal simulations. During tracking, an effort was made to use the highest possible linking distance that did not result in a significant number of mislinks. Because the range of speeds for the unimodal simulations is greater than that of the bimodal simulations, we were unable to capture the fastest speeds in the unimodal simulation without causing significant false links.

Table 2. W_1 between ground truth speed distributions and the speed distributions from each PT method for the lowest PSDR bimodal and unimodal simulations.

	Bimodal	Unimodal
TM-Kalman	0.1146	0.1363
TM-LAP	0.3400	0.2489
TP	0.7536	0.6349
V-TM	0.3434	0.2158

3.4.2 Relationship Between Classical Statistics and PSDR

To develop a more large-scale understanding of the performance of each PT code, we use a variety of classical and experimental statistics (Fig. 3). Each of these classical statistics target different potential sources of linking error. Because the imagery had a high signal to noise ratio, there were not many errors in the detection stage of PT for each simulation (only occurring due to overlapping particles). Therefore, the false link rate (FLR) primarily shows the potential for a particle to be unlinked, meaning there were no probable candidates for linking in nearby frames (Fig. 3a). The mean path length (MPL) shows the propensity for trajectories to be fractured due to lack of linking (Fig. 3b), and the Euclidean distance (ED) indicates the likelihood for links to move

back and forth between particles, sampling a large number of particles for a single trajectory (Fig. 3c). A realistic diagram of each of these potential errors is shown in Figure 3d. Plotting these statistics over a range of PSDRs reveals that TM–Kalman and TM-LAP consistently eclipse the performance of other PT methods (Fig. 6). In particular, the bimodal simulations reveal several task-relevant patterns. The mean path lengths (Fig. 6a) illuminate the tendency for V-TrackMat and TP to generally have shorter trajectories compared to either TM method. This shortening in TP's trajectories is significantly accentuated, especially at low PSDR levels. We attribute this phenomenon to 'trajectory splitting', where a particle is tracked for only a fragment of its presence in the field of view. Intricacies of TP's linking algorithm, which narrows the search space when inundated with potential particles for the ensuing frame, underpin this observation. While effective for slower-moving particles, especially in terms of memory requirements and algorithm speed, this linking strategy is less adept at tracking high-velocity particles in a directed flow. For V-TrackMat, the trajectory splitting seems to be a result of its more stringent linking algorithm. Although all PT codes try to match all trajectories during linking, V-TrackMat seems to have more extreme criteria that prevent incorrect links, as shown from the sample trajectories (Fig. 4). Thus, many trajectories are lost by V-TrackMat due to the algorithm's necessity for high-probability links. Furthermore, the FLRs (Fig. 6b) point towards V-TrackMat's propensity to either miss or inaccurately record a particle in a frame. However, because this error is likely a result of careful linking, the classical statistics may exaggerate the experimental errors for tracking algorithms such as V-TrackMat's. The EDs (Fig. 6c) further highlight that TP and V-TrackMat often record the most substantial discrepancies between the actual and tracked positions. This observation, particularly for V-TrackMat, implies that a rigorous linking algorithm doesn't invariably lead to precise

trajectory reconstructions. Although untested, it is theoretically plausible that during velocity-based linking or gluing, particles are incorrectly linked because they have similar velocities.

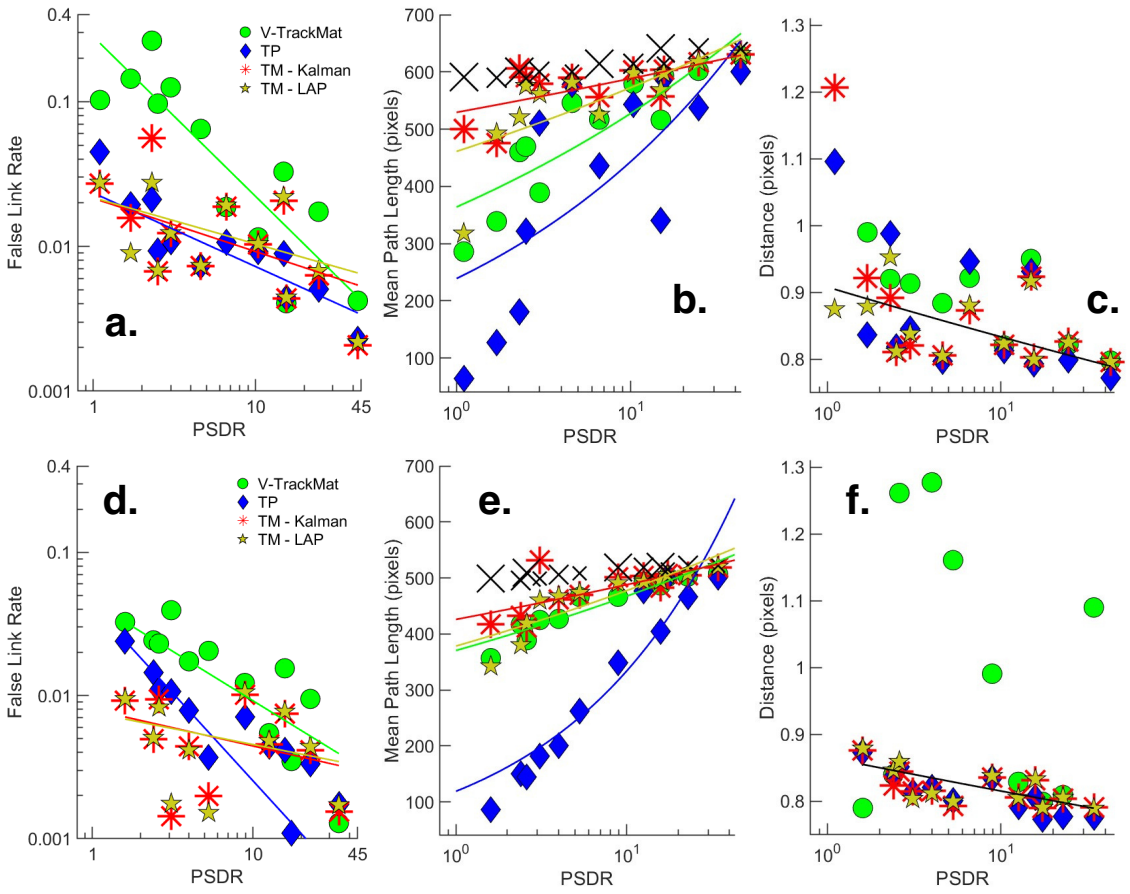


Figure 6. This figure shows the results of the classical comparative statistics for both the bimodal and unimodal simulations. For all plots, the size of the scatter points represent the particle density of the simulation (larger points means greater particle density). a-c correspond to the bimodal simulations, and d-f correspond to the unimodal simulations. (a) Mean false link rate (error due to detection and temporally local missed links). (b) Mean path length of all PT-obtained and simulated trajectories. The ground truth is shown as a black X. This statistic describes how often full trajectories are split (linking error over time). (c) Mean Euclidean distance between true and predicted trajectories (error due to localization and linking error). (d-f) Repeat of a-c but for the unimodal simulations. These figures generally indicate that V-TrackMat and TP have the worst "classical" performance. Furthermore, classical statistics tend to follow a power law trend as a function of PSDR. Power law fit equations and goodness of fit are given in Table 3.

For unimodal simulations, classical statistics (Figs. 6d-6f) generally perform better than their bimodal counterparts. The bimodal simulations have higher mean speeds than the unimodal solutions (Table 1). Furthermore, the bimodal simulations (Fig. 1a) have a larger number of particles at high speeds, which causes more difficulty in particle tracking. In addition, the unimodal simulations show a greater range of speeds and are generally more reminiscent of speed distributions of particles in porous media [54]. Thus, the unimodal simulations likely offer a more comprehensive representation of generic PT code efficacy in porous media. While the general trends mirror those in the bimodal findings, V-TrackMat performs comparatively better in the FLR metric (Fig. 6d) and worse in the ED metric (Fig. 6f), and TP performs better in the ED metric (Fig. 6f). TP's aforementioned challenges with fast-moving particles mean its performance slightly improves in unimodal settings, which aren't dominated by high speed trajectories. Still, TP's mean path lengths (Fig. 6e) depict a sharp decline as PSDR decreases, implying the persistent issue of trajectory splitting in both bimodal and unimodal settings.

Table 3. Power law fit equations and goodness of fit for ED, MPL and FLR.

Bimodal	***a***	***b***	R^2	***RMSE***
ED	0.922	-0.039	0.941	0.0210
MPL (TP)	238.9	0.268	0.510	141.00
MPL (TM-Kalman)	529.3	0.046	0.434	36.890
MPL (TM-LAP)	510.0	0.056	0.797	39.500
MPL (V-TrackMat)	363.9	0.161	0.767	55.250
FLR (TP)	0.023	-0.508	0.884	0.0041
FLR (TM-Kalman)	0.019	-0.342	0.813	0.0067
FLR (TM-LAP)	0.022	-0.319	0.261	0.0080
FLR (V-TrackMat)	0.212	-0.783	0.875	0.0288
Unimodal	***a***	***b***	R^2	***RMSE***
ED	0.849	-0.019	0.923	0.0323
MPL (TP)	118.9	0.448	0.885	55.620
MPL (TM-Kalman)	425.9	0.059	0.503	29.770
MPL (TM-LAP)	432.0	0.052	0.940	14.160
MPL (V-TrackMat)	370.4	0.101	0.881	17.960
FLR (TP)	0.043	-1.226	0.901	0.0022
FLR (TM-Kalman)	0.008	-0.255	0.142	0.0032
FLR (TM-LAP)	0.007	-0.223	0.110	0.0033
FLR (V-TrackMat)	0.045	-0.690	0.931	0.0032

The results of the classical statistics imply that TM-Kalman and TM-LAP outperform V-TrackMat in all cases, but from the sample trajectories (Fig. 4), we have shown this to not be true. Also, the trajectories show TP performs much worse than the other algorithms at low PSDR, but this is not reflected by the FLR and ED metrics. We posit that the primary reason for the disconnect between the classical statistics and the sample trajectories is that the FLR and ED metrics underpenalize aggressive linking. The FLR will always be lower when more links are forced, since the probability of false positive detection is very low. The ED metric will always be higher when more links occur between different trajectories, but if the trajectories are nearest neighbors, then the error will be relatively small. Thus, long trajectories and links across the pore space (such as those of V-TrackMat and TM-Kalman) will result in more ED error than zig-zagging trajectories between close neighbors (such as TM-LAP) will. Ultimately, the FLR and ED underpredict PT error for nearest-neighbor based algorithms with little constraint for linking. As a result, these statistics fail to grasp the nuanced differences between PT codes.

Beyond comparing PT codes' performances, we also demonstrate that all classical statistics have a power law relationship with PSDR, although some relationships are more significant than others (Table 3). As PSDR is reduced, all PT codes generally exhibit increased ED and FDR, and decreased MPL. V-TrackMat and TP show a steeper relationship between FDR and PSDR than either TM algorithm, which generally indicates that the TM algorithms are more robust with respect to FLR performance over a range of PSDRs (Figs. 6a and 6d). V-TrackMat and TP also show steeper relationships between MPL and PSDR, further demonstrating the resilience of the TM algorithms when considering classical linking failures. V-TrackMat and TP also generally show more significant (lower RMSE) power-law relationships than the TM algorithms, indicating that classical PT error for V-TrackMat and TP is more predictable. Furthermore, classical statistics from

unimodal simulations (Figs. 6d-6f) present slightly different power law relationships compared to those from bimodal simulations (Figs. 6a-6c). Thus, the choice of PT algorithm, and variations in ground truth particle speed distributions, can influence the specifics of these power law relationships.

3.4.3 Experimental Statistics Highlight Task-Specific PT Performance

The classical statistics from bimodal simulations (Fig. 6) echo many patterns observed in the sample trajectories (Fig. 4). However, there are notable deviations. The sample trajectories, for instance, present V-TrackMat as clearly superior to TP and comparable or superior to TM-LAP. To discern which mode of analysis — comparative statistics or visual trajectory inspection — offers a more accurate picture of PT performance, we used a variety of experimental statistics. In the bimodal simulations, the normalized speed-angle joint probability density difference heatmaps rank TM–Kalman as the top performer, with V-TrackMat and TM–LAP occupying intermediate positions and TP trailing (Fig. 7). All codes demonstrate strong tracking performance at PSDR $\geq$ 2.5, but V-TrackMat and TP's limitations become evident at PSDR $\leq$ 2.3. TM-LAP and TM-Kalman significantly outperform V-TrackMat for PSDR $\geq$ 1.7. However, at PSDR = 1.1, V-TrackMat performs better than TM-LAP, as shown by the large amount of overprediction for the probability of low speed and high turn angle particles (Fig. 7). This disparity is likely rooted in the LAP algorithm's propensity for aggressive linking that doesn't take particle velocities into account, in contrast to V-TrackMat's more conservative velocity-based approach. Consequently, at PSDR = 1.1, while LAP is prone to errant predictions for high speed particles and forces links with large turn angles, V-TrackMat is more likely to keep particles unlinked, and only significantly overpredicts low turn angles. In other words, V-TrackMat often refrains from making connections altogether, and when V-TrackMat does have false links, its reliance on expected particle velocities,

akin to TM-Kalman, ensures that the errors are relatively benign (with respect to velocity and angle distributions) compared to TM-LAP.

In the context of unimodal simulations (Supplementary Figure 4), both the V-TrackMat and TM algorithms predict speed and angle statistics with near perfection. V-TrackMat and TM-Kalman perform slightly better than TM-LAP, which can be seen from the slightly greater underprediction of high speed and low turn angle particles for TM-LAP at PSDR $\leq$ 2.6. TP shows relatively poor performance for all PSDR $\leq$ 3.1. These observations further reinforce the general trends seen in the sample trajectories (Fig. 4). They confirm the case presented by the classical statistics that TM-Kalman has superior performance, but they significantly contrast the relative classical results of V-TrackMat and TM-LAP. Specifically, the speed-angle distributions (both bimodal and unimodal) show that TM-LAP may be favorable for PSDR $\geq$ 1.7, but that V-TrackMat is superior for PSDR $\leq$ 1.6.

Velocity autocorrelation function (C_v) and mean squared displacement (MSD) analysis (Fig. 8) further corroborates the trends evident in the speed-angle heatmaps. It should be noted here that we only present the first 20 frames of the lowest and highest-PSDR simulations in the main text of this chapter, although the full C_v and MSDs for all simulations can be observed in Supplementary Figures 5 and 6. Because our simulations don't use reinjection to keep the number of particles in the field of view relatively constant, the C_v and MSDs for our simulated particles are unrealistic past 20-30 frames. Since the focus of our analysis is on the relatively accurate simulation of dispersing particles in porous media, we chose to focus on the subset of our results that are the most realistic.

At $PSDR = 34.3 - 42.8$, all PT methods align closely with the simulated autocorrelations and MSD ratios. There is some slight deviation for the MSD ratio at late times for the bimodal

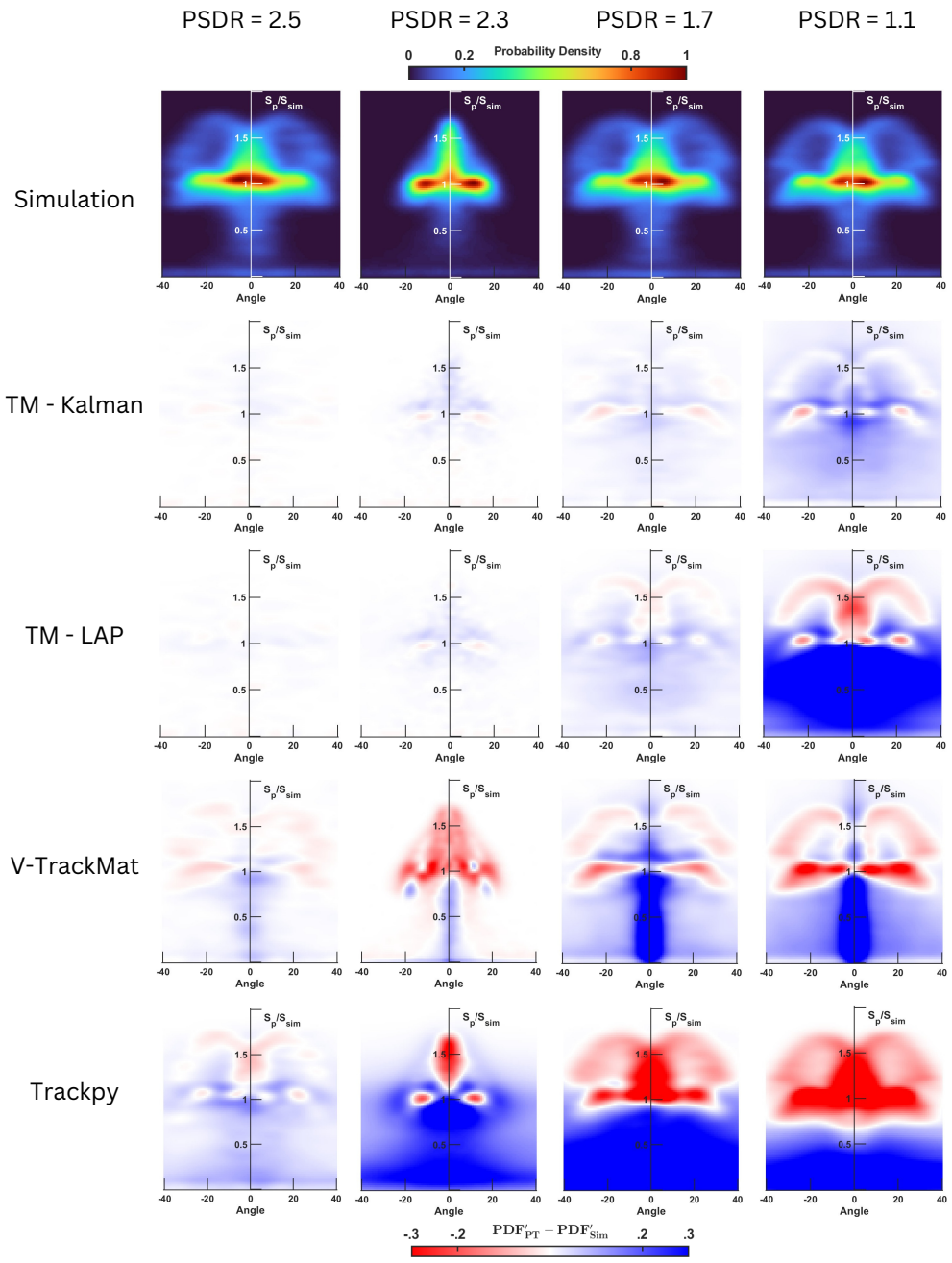


Figure 7. Speed-angle joint probability density difference heatmaps for the bimodal simulation. Speeds determined from particle tracking (S_p) are normalized by the mean speed of the respective simulation (S_{sim}). Red corresponds to an underprediction of probability density, blue corresponds to an overprediction of probability density, and white corresponds to an accurate probability density prediction within the speed-angle feature space. These results generally show the same trends as the sample trajectories (Fig. 1). At *PSDR* = 2.5, all algorithms show strong performance as indicated by the lack of strong color. All PT methods besides Trackpy and V-TrackMat show good replication of the simulation for $PSDR \geq 1.7$. At *PSDR* = 1.1, TM-Kalman still performs best and TP performs worst, but V-TrackMat surprisingly outperforms TM-LAP. Thus, at very low PSDR, velocity-based algorithms result in significant improvements to PT performance.

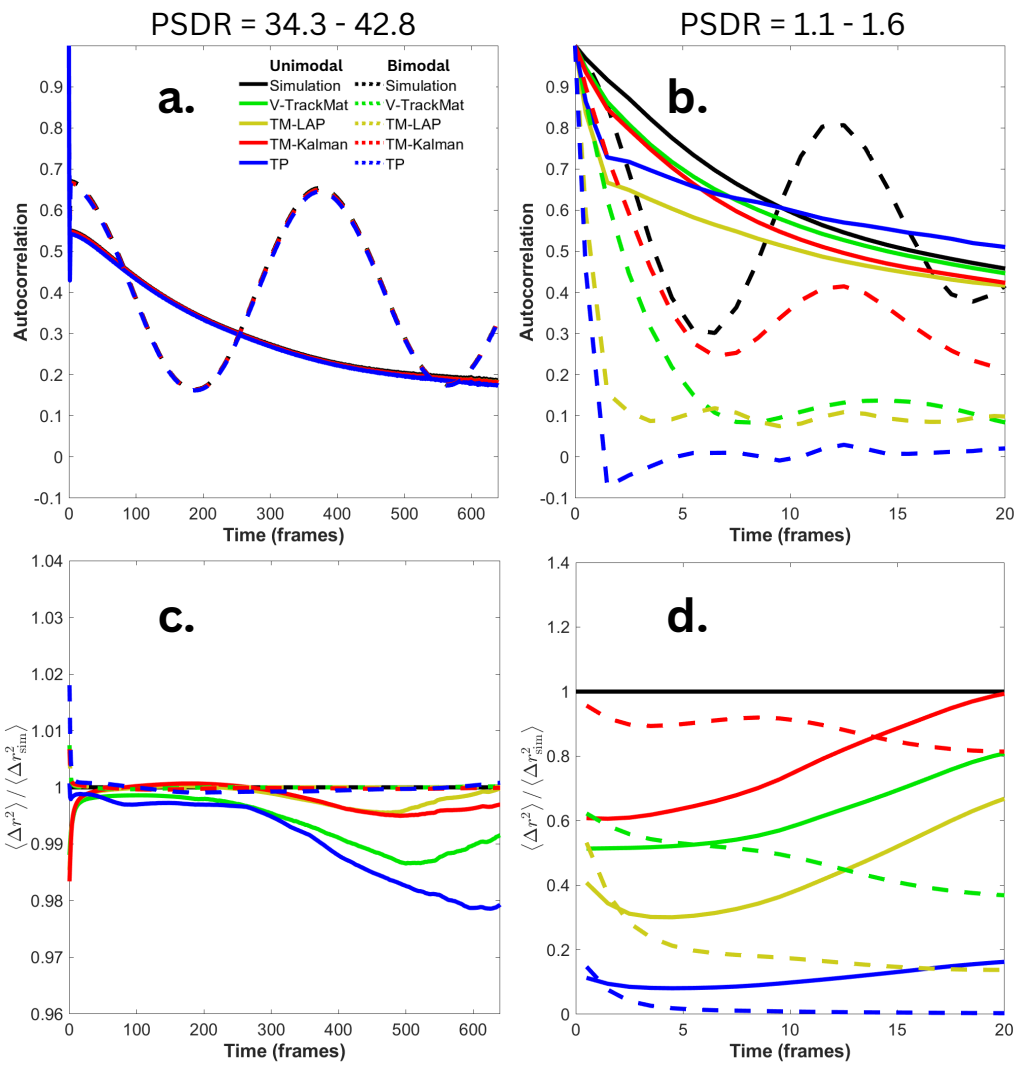


Figure 8. MSD ratios and VACFs for unimodal and bimodal simulations for high (a and c) and low (b and d) PSDRs. The MSD ratio is caluculated as the MSD obtained from particle tracking divided by the simulated MSD. An MSD ratio of 1 implies perfect accuracy. The bimodal MSD ratios are shown by dashed lines, and the unimodal MSD ratios are shown by solid lines. The simulation, or ground truth, is black, and the results from each PT method are different colors. For the unimodal simulations, *PSDR* = 34.3 (a and c) or *PSDR* = 1.6 (b and d). For the bimodal simulations, *PSDR* = 42.8 (a and c) or *PSDR* = 1.1 (b and d). These figures generally confirm trends present in the other experimental results. Furthermore, the MSDs and VACFs generally show the same trends, implying that a good prediction of MSD allows for a good prediction of C_v. However, unlike the other experimental statistics, the C_v is not a reliable proxy for general PT performance.

simulation for TP and V-TrackMat (Fig. 8c), but generally, all results are highly accurate. However, at $PSDR = 1.1 - 1.6$, all PT methods show large deviations in autocorrelation and MSD ratio. The

autocorrelation for the low PSDR bimodal simulation (Fig. 8b) shows decent performance for TM-Kalman, but poor performance for all other PT methods. The repetitive motion of the C_v is indicative of the wave-like periodic movement of the particles dispersing through the lattice-like geometry of the bimodal simulations. TM-Kalman is slightly able to capture this feature of the autocorrelation, but the other PT codes are not. The most likely explanation for this lies in the false links and splitting of fast trajectories. As previously discussed, as particles travel through the pore throat, they get closer together and speed up, which causes a decrease in the local PSDR. Thus, the C_v reveals that TM-Kalman is more likely to capture these fast/dense particles in the pore throats than the other PT codes are. The unimodal results for the C_v at low PSDR (Fig. 8b) surprisingly show that V-TrackMat outperforms TM-Kalman, and TP outperforms TM-LAP. However, the full C_v (Supplementary Figure 4) indicates the TM-LAP outperforms TP at $t \geq 30$. Likely, the C_v for TP is relatively accurate at early times because TP can only track very slow particles, so there are no significant false links that would cause velocity decorrelation between successive timesteps. TM-LAP, on the other hand, can track much faster particles, but may also erroneously link these fast particles, meaning a greater amount of velocity decorrelation. Thus, although it is important to know how accurate the C_v is for general analysis of particle transport, the C_v accuracy can't be used as a general proxy for total particle tracking accuracy.

The MSD ratios for low PSDR (Fig. 8d) show significant deviations from the simulated MSD for each PT code. Both the bimodal and unimodal results show TM-Kalman is able to most closely follow the true MSD (i.e., have an MSD ratio of 1), then V-TrackMat, then TM-LAP, and finally TP shows a complete disconnection from the true MSD. Interestingly, the unimodal simulations show an improvement in the MSD ratio over time, which indicates that for each PT code, the history of previous particle positions and links can improve the accuracy of tracking. For the bimodal

simulations, we see a decrease in the accuracy of the MSD ratio over time (Fig. 8d). However, the full time-series for the lowest PSDR bimodal simulation (Supplementary Figure 5) shows a significant improvement in the MSD accuracy over time for both V-TrackMat and TM-Kalman. Thus, velocity-based algorithms show a clear advantage in late time prediction of MSDs for low-PSDR scenarios, regardless of geometry.

Generally, our experimental statistics reveal that while rudimentary comparative statistics can offer broad insights into PT code competencies across various tracking scenarios, they might fall short in pinpointing optimal codes for specific particle motions with particular analytical objectives. In our bacterial dispersion simulation within porous media, these statistics fail to elucidate speed, angle, autocorrelation or displacement distribution accuracies — all crucial for comprehending bacterial transport. Furthermore, these comparative statistics tend to underpenalize aggressive linking. Thus, basic comparative statistics might not capture the full spectrum of PT code capabilities. A more complete analysis, which can be done through a variety of statistical and visual methods, is indispensable for discerning the optimal PT code tailored to specific conditions.

3.4.4 PT Performance for Simulations with Noisy Trajectories

While the primary analysis in this chapter revolves around simulations where trajectories only vary in speed and particle density, we have also provided an analysis of PT performance for simulations that contain more noise depicting experimental errors in video capture and processing. Specifically, we analyzed PT performance for simulations in which the particles had enhanced random displacement (Gaussian distribution with $\mu = 0$ and $\sigma = 2$ pixels) added to the purely advective tracks, and 2% of the particles were dropped in any given frame to account for particle intermittency. The random displacement is a simple representation of a variety of experimental

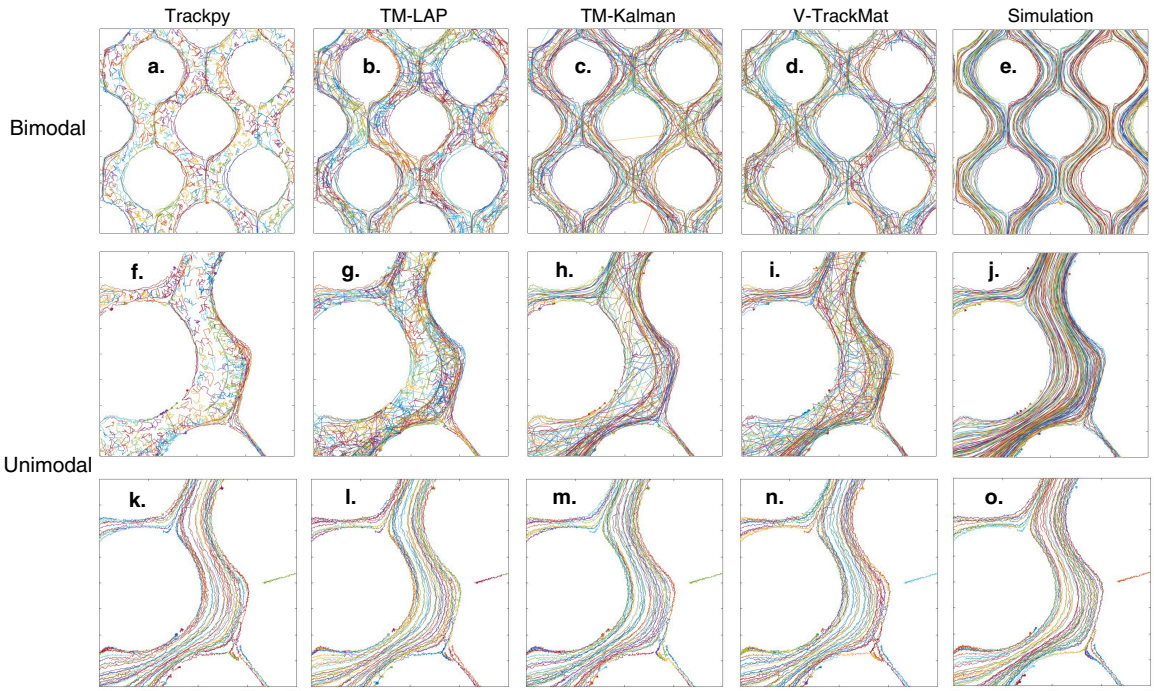


Figure 9. Sample trajectories for all PT codes for the low-PSDR bimodal and unimodal simulations with random motion and 2% particle intermittency. Each plot shows a 400x400 pixel section of the whole domain. Within a plot, each line corresponds to a unique trajectory (with random colors used to show the contrast between individual trajectories). (a-e) Bimodal simulations for *PSDR* = 1.5. (f-j) Unimodal simulations for *PSDR* = 1.5. (k-o) Unimodal simulations for *PSDR* = 8.1. Compared with Fig. 4, this figure (specifically the top left of f-j) shows a slight decrease in tracking performance for similar PSDR due to the addition of trajectory noise.

phenomena/positioning errors such as diffusion, camera jitter, and/or oscillations in particle brightness. The intermittency represents particles moving in and out of the focal plane, which can also be impacted by diffusion, particle-particle interactions, particle-wall interactions, and camera exposure time. Both of these changes can be generalized as increasing the noise of the trajectories in the simulations. The sample trajectories of the lowest PSDR simulations with the intermittent and random-motion particles (Fig. 9) show the same general trends as those of the simulations with minimal noise (Fig. 4), but for each PT code the errors are slightly higher in the case of the noisy trajectories. The speed-angle distributions for the noisy unimodal simulations (Fig. 10) show

that TP clearly has the worst performance for PSDR ≤ 2.3 . For the highest PSDR noisy unimodal simulation, the performance of all PT codes are comparable. At the lowest PSDR, TM-Kalman once again shows the best performance. Similar to the unimodal results with minimal trajectory noise (Supplementary Figure 4), V-TrackMat performs better than TM-LAP in all cases besides the highest-PSDR simulation.

Ultimately, these results indicate that trajectory noise such as large random motions and particle intermittency make the tracking process more error-prone, although the rankings of the linking algorithms are not impacted by these potential experimental issues. However, a more robust analysis of potential experimental errors would deal with a number of other factors such as signal to noise ratio and particle shape/size. This would also require a rigorous comparison of detection methods, which was beyond the scope of our work, but we recommend that future researchers compare PT codes in the context of more diverse simulations.

3.4.5 Consequences of Particle Tracking Errors on Transport Analysis

Building on our comparative exploration of PT codes, the findings from the bivariate speed-angle heatmaps, MSDs, and C_v (Figs. 7 and 8, and Supplementary Figures 4, 5, and 6) shed light on the dispersion dynamics of tracer particles within porous media. Specifically, they underscore how inaccuracies introduced by PT errors can skew transport analysis. A predominant manifestation of PT error arises from false links (Fig. 4d), leading to a systematic underestimation of high-speed particles (Figs. 4, 5 and 7). This, in turn, results in a conservative estimation of particle speeds (Figs. 5 and 7) and MSDs (Fig. 8). TP, which shows the most significant error due to trajectory splitting, underestimates the particle speeds and MSDs to an extreme degree for low PSDR.

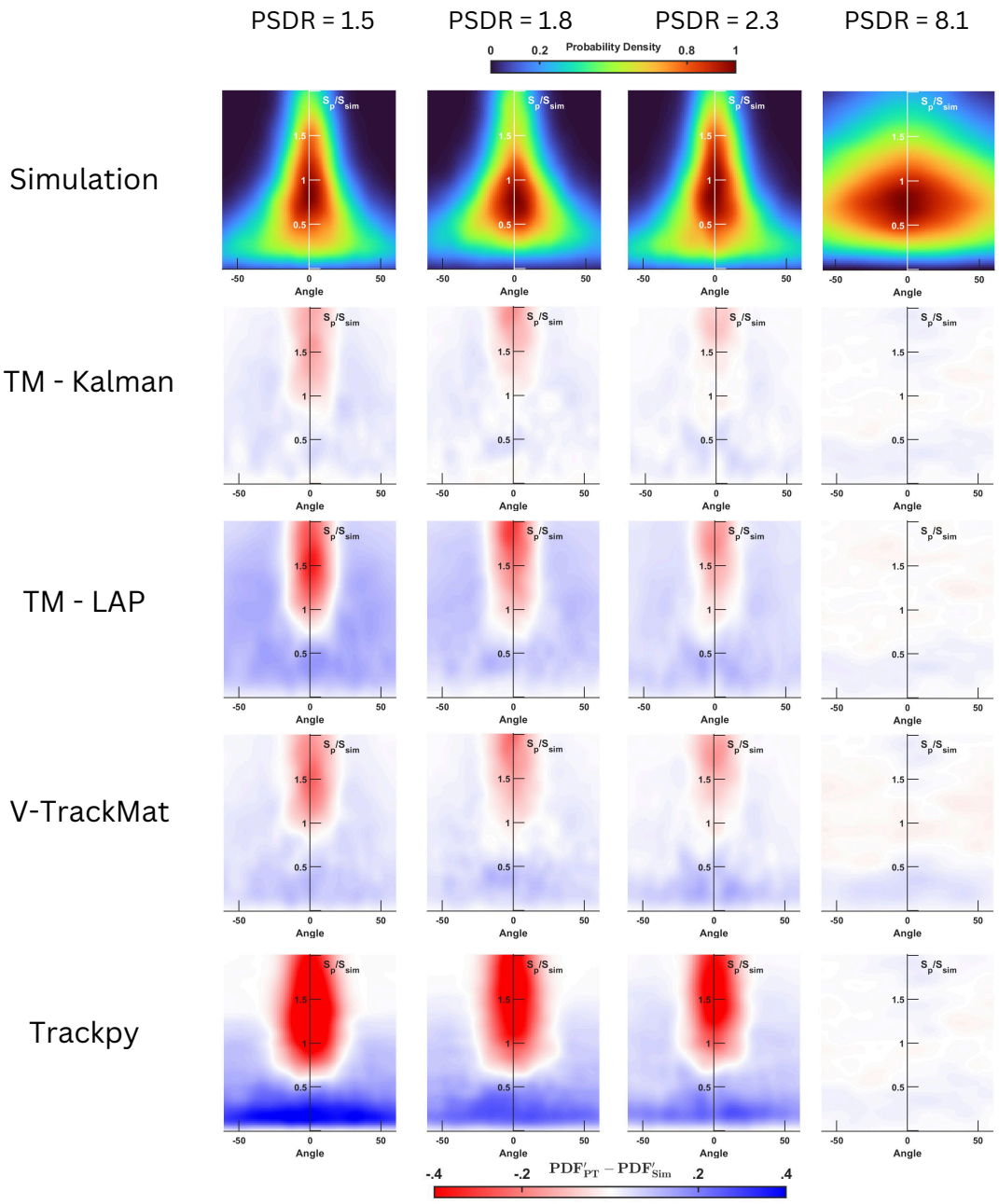


Figure 10. Speed-angle joint probability density difference heatmaps for the unimodal simulations with random motion and 2% particle intermittency. Speeds determined from particle tracking (S_p) are normalized by the mean speed of the respective simulation (S_{sim}). Red corresponds to an underprediction of probability density, blue corresponds to an overprediction of probability density, and white corresponds to an accurate probability density prediction within the speed-angle feature space. Although these noisy simulations are slightly harder to track, the general trends in PT performance remain the same.

Other consequences of PT error, which can primarily be observed in the TP results, are inflated turn angles (Fig. 7) and diminished or enhanced C_v (Fig. 8a and 8b), attributable mainly to trajectory splitting and erroneous linking. Furthermore, we find that in the case of the bimodal geometry, where there is a periodic nature to the velocity of particles over time, only TM-Kalman is able to slightly capture the periodicity of this autocorrelation.

Our analysis also emphasizes the paramount importance of experimental conditions (related to particle speed and density) in achieving reliable PT outcomes. In the case of minimal-noise simulations, a PSDR exceeding 3 ensures nearly flawless tracking, regardless of PT algorithm. Conversely, a PSDR near or below 1 presents challenges for all PT codes. In scenarios characterized by low PSDR coupled with directed particle movements, algorithms that harness velocity-based linking emerge as the more prudent choice. In addition, the results for noisy simulations are worse than for simulations with minimal noise (Fig. 4 and Fig. 9), which highlights the need for tight experimental controls to improve the visual quality of the particles. Although some amount of noise is unavoidable, these results show the importance of trying to ensure that all particles remain in the focal plane of the acquisition device.

3.4.6 PT Algorithm Speed Comparison

In addition to performance analysis, we also report how long each PT code takes to link trajectories (Fig. 11). Generally, TM – LAP is the fastest linking algorithm, then TP, then TM – Kalman, and V-TrackMat is the slowest. Thus, we observe a significant trade-off between performance and computation time – the best PT methods at low PSDR also take the longest. However, we must also note that each PT code is developed in a different programming language (Python, Matalb and Javascript), so we are unable to fairly assess the speed of the underlying algorithms.

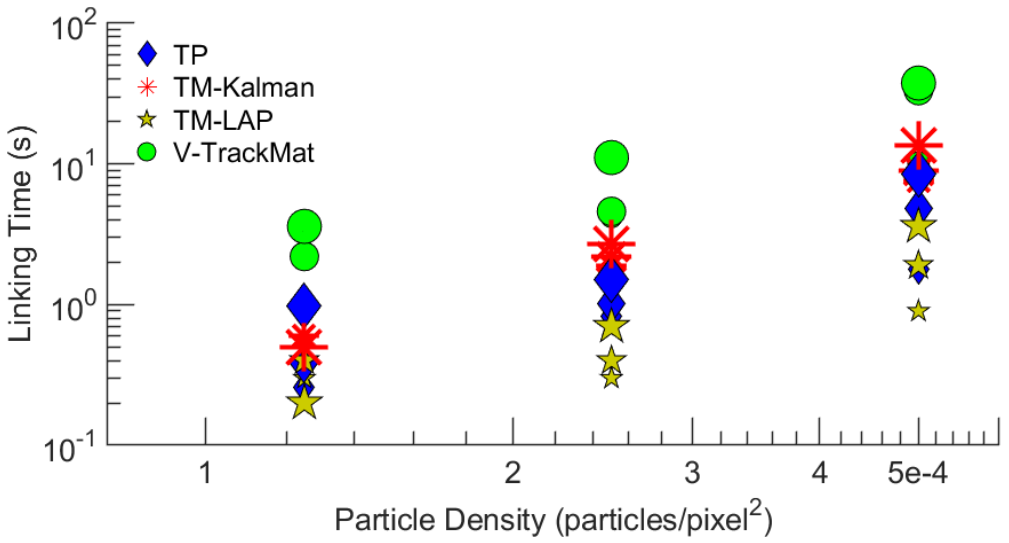


Figure 11. Amount of time each PT code takes during the linking stage for selected unimodal simulations. Simulation speed is represented by scatter point size (large = 9.9 px/frame, medium = 2.6 px/frame, small = 0.9 px/frame). V-TrackMat consistently has the longest linking times, and TM-LAP has the shortest linking times. All algorithms show a power law relationship between linking time and particle density. High speed simulations generally take longer to link than low speed ones, and this difference increases at higher particle density.

3.4.7 Connection to Chapter 2

The results presented in this chapter have significant connections to chapter 2. The flow field used to generate our bimodal simulations in this chapter are the same as the $\phi = 0.42$ flow fields simulated in chapter 2 (chapter 2, Figs. 6c and 6d). Thus, although the simulated particles are different than the bacteria in the experimental videos, the results of this chapter in reference to the bimodal simulations should generally apply to our experiments in chapter 2. Given that some of the videos for *Geobacter* and *Acidovorax* used for analysis in chapter 2 displayed jump lengths of up to 120 pixels per frame, it is reasonable to believe that some of our analysis suffers from the problems identified in this chapter.

To further probe this theory, we present zoomed-in trajectories (Figs. 12a-12d) and a comparison of speed distributions (Fig. 12e) for the different PT methods identified in this chapter

for one of our experimental videos of *Geobacter* at 5ul/h that was used in chapter 2. The sample trajectories show that all algorithms do surprisingly well, although TP results in significantly more erroneous links than the other algorithms do. Although it was not possible to calculate PSDR for the experimental trajectories given the lack of ground truth data, comparison with our simulated data would indicate the PSDR of this simulation was between 2 (Fig. 4b) and 3 (Supplementary Figure 3). The speed pdfs (Fig. 12e) further indicate that each algorithm was able to, for the most part, accurately track the bacteria. Although this comparison is not exactly accurate, because the simulation PDF comes from the Eulerian flow fields and the distributions for each PT method come from Lagrangian particle tracking. However, we posit that the flow field is adequately sampled as shown by the density of the experimental trajectories (Figs. 12a-12d), meaning that flux corrections are not necessary to make adequate comparisons between the two distributions. Furthermore, even if flux-weighting were to shift the location of the simulation PDF, it is unlikely that this shift would result in different rankings of algorithm accuracy. The results of the speed PDFs indicate that TM-Kalman is especially good at tracking high-speed particles, which backs up the claim other parts of this chapter that velocity-based algorithms offer improvements for particles with directed motion. Furthermore, these results show how all PT algorithms tend to underpredict high-speed particles in this experimental setting with fluid speeds of $v_m = 82.7$ µm/s, which further highlights the need for new developments in particle tracking that can improve the extraction of motion statistics from high-speed particles.

3.5 Discussion

Our comprehensive analysis of PT methods underscores TM–Kalman as the leading PT algorithm in terms of accuracy and robustness. While V-TrackMat emerges as a strong contender

at low PSDR, TM–LAP stands out at high PSDR. TP, although impressive at high PSDR, falters notably with disconnection challenges at PSDR ≤ 1.7. Despite the TP authors suggesting that

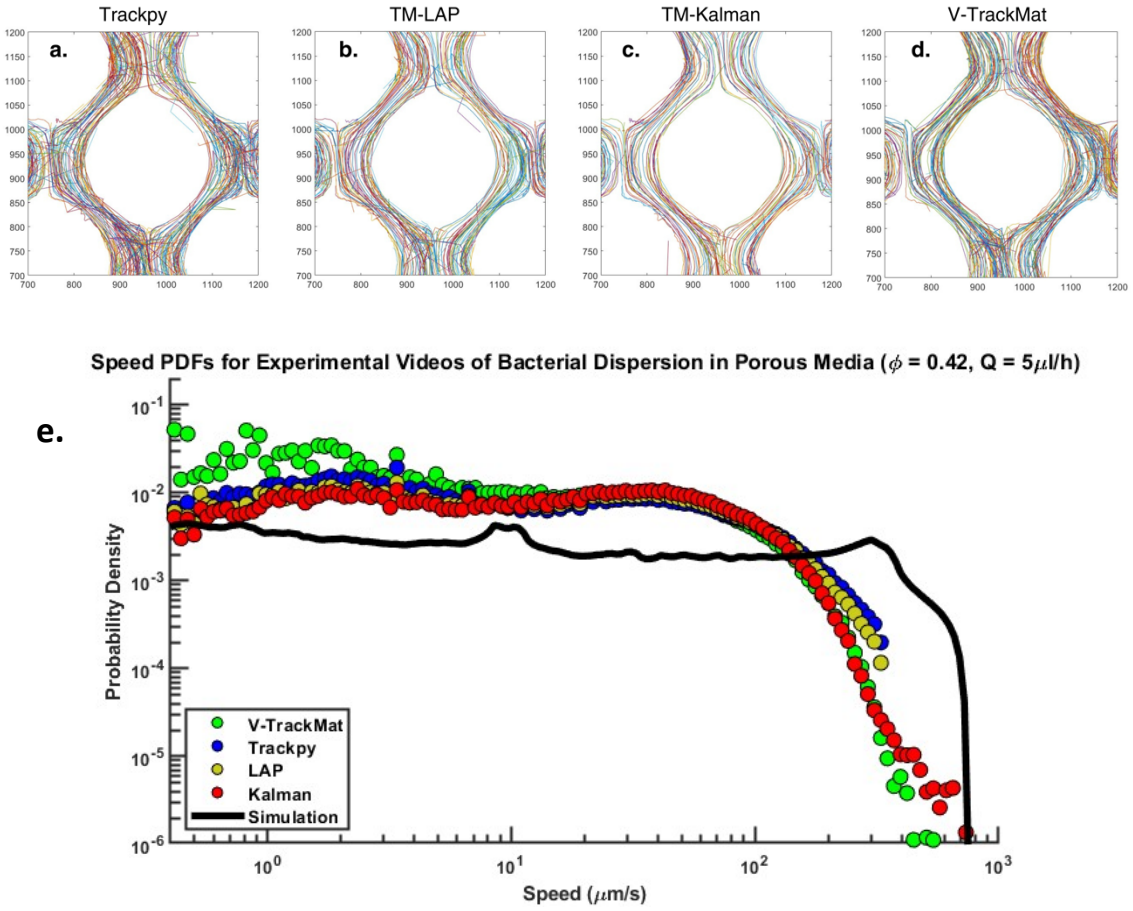


Figure 12. Comparison of PT methods for one of the experimental videos used for the analysis of Geobacter in a $\phi = 0.42$ porosity geometry with a flow rate of $Q = 5\frac{\mu L}{h}$ in chapter 2. Zoomed-in trajectories for Trackpy (TP) (a), TM-LAP (b), TM-Kalman (c), and V-TrackMat (d). The trajectories generally show strong performance for all PT algorithms except TP. (e) Comparison of speed PDFs for each PT algorithm in relation to the PDF obtained from the simulated flow field from chapter 2, Fig. 6d. TM-Kalman shows slight improvements over other methods, but all representations of the speed PDF are generally accurate.

adjusting the "SubnetOversizeException" variable could rectify the constrained search space at low PSDR, our attempts in this direction were unsuccessful. However, it should be noted that we used TP version 0.5.0, and that a newer version may have more potential to rectify this error.

Barring this error, TP would, in all probability, align more closely with the performance of other methods at low PSDR.

It's evident across the board that PT methods grapple with high FLRs, high ED, and trajectory splitting/fragmentation, especially at PSDR $\leq$ 1.7. However, we show that poor performance in classical statistics doesn't necessarily imply poor performance in experimental statistics. Specifically, classical statistics underpenalize aggressive linking algorithms, and overpenalize careful linking agorithms. In addition, we show that TM–Kalman, and V-TrackMat, which both use particle velocities to enhance predictions, exhibit marked improvements at PSDR = 1.1 relative to TM-LAP. Although TP also uses particle velocity information to make linking predictions, the SubnetOversizeException issue was much more significant than any potential gain due to velocity-based predictions. Thus, the leveraging of velocity data, especially for particles dispersing in porous media which exhibit a wide range of potential speeds, emerges as a critical factor in bolstering PT predictions.

Beyond algorithmic evaluations, this work highlights some of the common errors in transport analysis that emerge as a result of PT error. We find that all PT codes underestimate particle speeds and overestimate turn angles, and that poor tracking causes a significant loss in the accuracy of reproducing cyclical autocorrelations. Furthermore, we advocate for the recording of video data at a minimum threshold of PSDR $\geq$ 3, a benchmark that promotes reliable tracking irrespective of the algorithm employed. Finally, we show that experimental noise can reduce the quality of the predicted trajectories from each PT code, but it doesn't significantly impact the relative rankings of each PT code.

While we do provide robust statistical comparisons between some of the best-known open source PT codes, we recognize that our study is somewhat limited in scope. A more robust

analysis would use a wider variety of PT codes (including DL-based ones), more task-relevant statistics to test performance across a variety of domains, more variety of particle motions and noise, more variety in imagery type, and would enlist the creators of each code to submit the trajectories to be scored. Such an investigation was far beyond the scope of this chapter and would require a large collaborative effort spanning multiple disciplines. Looking forward, we envision our research catalyzing advancements in PT theory in three pivotal aspects. First, we urge future studies to transcend the boundaries of classical statistics in PT code comparisons, emphasizing the integration of experimental outcomes pertinent to specific applications. Second, our findings echo the effectiveness of velocity-centric PT methods for dispersing particles in porous media, extending their proven efficacy from constant velocity scenarios [26] to contexts characterized by large velocity fluctuations in time and space. Third, we highlight the importance of maintaining a reasonably high PSDR to achieve precise particle transport analysis.

Data Availability

Most of the data used in this study can be found at https://doi.org/10.5281/zenodo.10891931. Please send any data requests to the corresponding author (Rishi Parashar).

References

1. Saxton, M. Single-particle tracking: the distribution of diffusion coefficients. Biophys. journal 72(4), 1744–1753 (1997).
2. Hong, Q., Sheetz, M. & Elson, E. Single particle tracking. analysis of diffusion and flow in two-dimensional systems. Biophys. journal 60(4), 910–921 (1991).
3. Daumas, F. et al. Confined diffusion without fences of a g-protein-coupled receptor as revealed by single particle tracking. Biophys. journal 84(1), 356–366 (2003).
4. Forier, K. et al. Transport of nanoparticles in cystic fibrosis sputum and bacterial biofilms by single-particle tracking microscopy. Nanomedicine 8(6), 935–949 (2013).
5. Morales, V. L., Dentz, M., Willmann, M. & Holzner, M. Stochastic dynamics of intermittent pore-scale particle motion in three-dimensional porous media: Experiments and theory. Geophys. Res. Lett. 44, 9361–9371 (2017).
6. Bultreys, T., De Boever, W. & Cnudde, V. Imaging and image-based fluid transport modeling at the pore scale in geological materials: A practical introduction to the current state-of-the-art. Earth-Science Rev. 155, 93–128 (2016).

7. Scheidweiler, D., Miele, F., Peter, H., Battin, T. J. & de Anna, P. Trait-specific dispersal of bacteria in heterogeneous porous environments: from pore to porous medium scale. J. The Royal Soc. Interface 17, 20200046, DOI: 10.1098/rsif.2020.0046.
8. Dehkharghani, A., Waisbord, N., Dunkel, J. & Guasto, J. Bacterial scattering in microfluidic crystal flows reveals giant active taylor–aris dispersion. Proc. Natl. Acad. Sci. 116(23), 11119–11124 (2019).
9. Bhattacharjee, T. & Datta, S. Bacterial hopping and trapping in porous media. Nat. communications 10(1), 2075 (2019).
10. Dentz, M., Creppy, A., Douarche, C., Clément, E. & Auradou, H. Dispersion of motile bacteria in a porous medium. J. Fluid Mech. 946, A33 (2022).
11. Creppy, A., Clément, E., Douarche, C., D'angelo, M. & Auradou, H. Effect of motility on the transport of bacteria populations through a porous medium. Phys. Rev. Fluids 4(1), 013102 (2019).
12. Vallotton, P. et al. Diatrack particle tracking software: Review of applications and performance evaluation. Traffic 18(12), 840–852 (2017).
13. Germain, D., Leocmach, M. & Gibaud, T. Differential dynamic microscopy to characterize brownian motion and bacteria motility. Am. J. Phys. 84(3), 202–210 (2016).
14. Ewers, H. & Schelhaas, M. Effect of motility on the transport of bacteria populations through a porous medium. Methods enzymology 506, 63–80 (2012).
15. Dehkharghani, A., Waisbord, N. & Guasto, J. S. Self-transport of swimming bacteria is impaired by porous microstructure. Commun. Phys. 6, 18, DOI: 10.1038/s42005-023-01136-w (2023).
16. Rusconi, R., Guasto, J. S. & Stocker, R. Bacterial transport suppressed by fluid shear. Nat. Phys. 10, 212–217, DOI: 10.1038/nphys2883 (2014).
17. Bhattacharjee, T. & Datta, S. S. Bacterial hopping and trapping in porous media. Nat. Commun. 10, 2075, DOI: 10.1038/s41467-019-10115-1 (2019).
18. Jeon, H. et al. Quantitative analysis of single bacterial chemotaxis using a linear concentration gradient microchannel. Biomed. microdevices 11, 1135–1143 (2009).
19. Birjiniuk, A. et al. Single particle tracking reveals spatial and dynamic organization of the escherichia coli biofilm matrix. New journal physics 16(8), 085014 (2014).
20. Secchi, E. et al. The effect of flow on swimming bacteria controls the initial colonization of curved surfaces. Nat. Commun. 11, 2851, DOI: 10.1038/s41467-020-16620-y (2020).
21. Wu, H. & Schwartz, D. Nanoparticle tracking to probe transport in porous media. Accounts Chem. Res. 53(10), 2130–2139 (2020).
22. Linkhorst, J., Beckmann, T., Go, D., Kuehne, A. J. & Wessling, M. Microfluidic colloid filtration. Sci. reports 6(1), 22376 (2016).
23. Bevan, M. & Prieve, D. Hindered diffusion of colloidal particles very near to a wall: Revisited. The J. Chem. Phys. 113(3), 1228–1236 (2000).
24. Sholl, D., Fenwick, M., Atman, E. & Prieve, D. Brownian dynamics simulation of the motion of a rigid sphere in a viscous fluid very near a wall. The J. Chem. Phys. 113(20), 9268–9278 (2000).
25. Cheng, H., Hsu, C., Hung, C. & Lin, C. A review for cell and particle tracking on microscopy images using algorithms and deep learning technologies. biomedical journal 45(3), 465–471 (2022).
26. Chenouard, N. et al. Objective comparison of particle tracking methods. Nat. methods 11(3), 281–289 (2014).
27. Residori, M., Praetorius, S., de Anna, P. & Voigt, A. Influence of finite-size particles on fluid velocity and transport through porous media. Phys. Rev. Fluids 8, 7 (2023).
28. Meijering, E., Dzyubachyk, O. & Smal, I. Methods for cell and particle tracking. Methods enzymology 504, 183–200 (2012).
29. Clarke, D. & Martin-Fernandez, M. A brief history of single-particle tracking of the epidermal growth factor receptor. Methods protocols 2(1), 12 (2019).
30. Wiggins, C., Santos, R. & Ruggles, A. A feature point identification method for positron emission particle tracking with multiple tracers. Nucl. Instruments Methods Phys. Res. Sect. A: Accel. Spectrometers, Detect. Assoc. Equip. 843, 22–28 (2017).
31. Travers, T., Colin, V., Loumaigne, M., Barillé, R. & Gindre, D. Single-particle tracking with scanning non-linear microscopy. Nanomaterials 10(8), 1519 (2020).
32. Maska, M. & Matula, P. Particle tracking accuracy measurement based on comparison of linear oriented forests. In Proc. IEEE Int. Conf. on Comput. Vis. Work. 11–17 (2017).
33. Malik, N., Dracos, T. & Papantoniou, D. Particle tracking velocimetry in three-dimensional flows: Part ii: Particle tracking. Exp. fluids 15, 279–294 (1993).
34. Baek, S. & Lee, S. A new two-frame particle tracking algorithm using match probability. Exp. Fluids 22, 23–32 (1996).
35. Shuang, B., Chen, J., Kisley, L. & Landes, C. Troika of single particle tracking programing: Snr enhancement, particle identification, and mapping. Phys. Chem. Chem. Phys. 16(2), 624–634 (2014).

36. Bijeljic, B., Mostaghimi, P. & Blunt, M. J. Signature of Non-Fickian Solute Transport in Complex Heterogeneous Porous Media. Phys. Rev. Lett. 107, 204502, DOI: 10.1103/PhysRevLett.107.204502 (2011).
37. De Anna, P., Quaife, B., Biros, G. & Juanes, R. Prediction of the low-velocity distribution from the pore structure in simple porous media. Phys. Rev. Fluids 2, 124103, DOI: 10.1103/PhysRevFluids.2.124103 (2017).
38. Weller, H., Tabor, G., Jasak, H. & Fureby, C. A tensorial approach to computational continuum mechanics using object-oriented techniques. Comput. Phys. 12, 620 (1998).
39. Tinevez, J. et al. Trackmate: An open and extensible platform for single-particle tracking. Methods 115, 80–90 (2017).
40. Ershov, D. et al. Trackmate 7: integrating state-of-the-art segmentation algorithms into tracking pipelines. Nat. Methods 19(7), 829–832 (2022).
41. Allan, D., Caswell, T., Keim, N., van der Wel, C. & Verweij, R. soft-matter/trackpy: Trackpy v0. 5.0., DOI: https://doi.org/10.5281/zenodo.4682814 (2021).
42. Midtvedt, B. et al. Quantitative digital microscopy with deep learning. Appl. Phys. Rev. 8(1) (2021).
43. Fatemi, B., Halcrow, J. & Jaqaman, K. Geometric deep learning of particle motion by magik. Nat. Mach. Intell. 5(5), 483–484 (2023).
44. Zhang, Y., Wang, C., Wang, X., Zeng, W. & Liu, W. Fairmot: On the fairness of detection and re-identification in multiple object tracking. Int. J. Comput. Vis. 129, 3069–3087 (2021).
45. Abràmoff, M., Magalhães, P. & Ram, S. Image processing with imagej. Biophotonics international 11(7), 36–42 (2004).
46. Kalman, R. A new approach to linear filtering and prediction problems. Basic Eng. 82, 35–45 (1960).
47. Jaqaman, K. et al. Robust single-particle tracking in live-cell time-lapse sequences. Nat. methods 5(8), 695–702 (2008).
48. Crocker, J. & Grier, D. Methods of digital video microscopy for colloidal studies. J. Colloid Interface Sci. 179(1), 298–310 (1996).
49. Carrel, M. et al. Pore-scale hydrodynamics in a progressively bioclogged three-dimensional porous medium: 3-d particle tracking experiments and stochastic transport modeling. Water resources research 54(3), 2183–2198 (2018).
50. Carrel, M. et al. Biofilms in 3d porous media: Delineating the influence of the pore network geometry, flow and mass transfer on biofilm development. Water research 134, 280–291 (2018).
51. A. Alam, E. Crowd of Microswimmers. Phd thesis, Condensed Matter [cond-mat]: Université Grenoble Alpes, Université Grenoble Alpes (2022). Available at https://theses.hal.science/tel-03894880/document.
52. Tarantino, N. et al. Tnf and il-1 exhibit distinct ubiquitin requirements for inducing nemo–ikk supramolecular structures. The J. cell biology 204(2), 231 (2014).
53. Puyguiraud, A., Gouze, P. & Dentz, M. Upscaling of anomalous pore-scale dispersion. Transp. Porous Media 128, 837–855 (2019).
54. Nguyen, V. & Papavassiliou, D. Velocity magnitude distribution for flow in porous media. Ind. & Eng. Chem. Res. 60(38), 13979–13990 (2021).

Chapter 4: DeepTrackStat: an End-to-End Deep Learning Framework for Extraction of Motion Statistics from Videos of Particles

4.1 Abstract

As discussed in the previous chapter, particle tracking (PT) is a mature area of research that traditionally has used Gaussian filtering and nearest neighbors-based algorithms to detect and link features in a sequence of images. PT shares many similarities with the general task of object tracking, although it is specifically designed for tracking objects that are usually small and have a distinct shape shared amongst all particles in images that have little to no background. Although object tracking is also a mature area of research, transferring the computer vision techniques from general object tracking to PT presents significant challenges due to the sparsity and high resolution of PT videos. To remedy these issues, along with problems of classical PT methods presented in chapter 2, we present DeepTrackStat (DTS), a model that is able to bypass the tracking process entirely and generate accurate statistics on speed, velocity components (Vx and Vy), and turn angle for a wide range of PT scenarios including trajectories derived from Brownian motion, Poiseuille flow, and porous media flow. In addition to its ability to handle a variety of flow types, the model is robust to large variations in particle size, shape, brightness, speed, density, and signal to noise ratio. One of the primary advantages of DTS is that it can reduce the time required to obtain the target statistics from videos of moving particles by 6x (when compared with classical methods). Furthermore, we show that DTS is able to predict the motion statistics of particles with higher accuracy than a number of classical PT methods and simple deep learning (DL) based methods. In addition, we show that DTS' performance is comparable to Trackmate (TM), which was shown to be the top-performing PT algorithm from chapter 2, for a wide variety of simulated trajectories and experimental datasets of motile bacteria dispersing in porous media

under a range of flow conditions. We further highlight that DTS significantly outperforms TM for prediction of motion statistics for high-speed particles. We hope this work can be used to help advance DL-based methods for particle tracking, and that our model provides significant time savings and allows for improved analysis of particle trajectories.

4.2 Introduction

At a basic level, Particle Tracking (PT) is a set of algorithms used to detect bright spots and determine their trajectories across multiple frames of a video. PT can be considered a subset of the object tracking task, and it can be applied to any video data with moving objects, but it is especially relevant for tracking small, spherical particles. This includes phenomena such as bacterial dispersion and transport in porous media [1-8], cellular diffusion [9,10], biofilm formation [11], chemotaxis [12-14], viral transport [15], and colloid filtration [16]. These applications of PT are highly dependent on accurate measurements. For example, when PT algorithms tend to miss fast-moving particles, this can cause an underestimate of the mean square displacement [17].

Most PT frameworks consist of detection, linking, and filtering stages [18]. The detection stage generally uses a Gaussian filter to filter and normalize an image, thus revealing local maxima that correspond to the centroids of spherical objects of a certain diameter. Many strategies have been developed to improve upon this general detection method [19, 20], and the most state of the art detection algorithms use CNNs to improve particle recognition [21-23]. The linking stage consists of connecting the detected bright spots across multiple frames in time to form the most probable trajectory for each particle, and is generally where most PT algorithms differ from each other. For linking of fast and dense particles, TrackMate (TM) [24] has been shown to be one of

the best performing PT algorithms [18]. However, TM is also slow, and it may take significant domain knowledge and experimentation to achieve accurate tracking results. Unlike detection, very few DL-based methods have been developed for the linking stage. The only algorithm that has been presented as an end-to-end DL-based framework for linking particle trajectories is MAGIK [25], which uses a graph neural network to capture the spatiotemporal relationships present in PT coordinate data. Although this algorithm boasts strong performance for the tested scenarios, it is untested in scenarios of high particle speed and density, and requires a large amount of VRAM (a coordinates shape of 100 frames by 1000 particles requires at least more than 24GB). Furthermore, MAGIK still requires another algorithm to perform the detection stage, and then the coordinate data must be converted to node-edge format for model use (which is not a trivial step), meaning it does not improve the ease, speed, or accuracy of the overall particle tracking process. MAGIK can also be thought of as one of the few models that bridge the gap between the fields of particle tracking and object tracking. Object tracking is a mature field within the domain of computer vision that uses CNNs and vision transformers (VTs) to primarily track people and cars [26, 27]. Although object tracking methods are robust in their task-relevant performance, few have been trained to track the kinds of objects generally found in particle tracking experiments.

To bridge the gap between the particle tracking and object tracking domains and improve upon the extraction of motion statistics from PT data, we propose DeepTrackStat (DTS), a novel end-to-end DL-based framework. Specifically, our model offers the ability to predict speed, velocity (Vx and Vy) and turn angle distributions from a raw image sequence input. DTS is designed to be as general as possible, meaning it can accurately predict statistics from a variety of particle shapes, sizes, brightness, density, speed, and SNR, and a variety of trajectory motion types

such as dispersive, straight, and Brownian. DTS is a two-stage system that consists of a speed classifier (SC) and statistics-specific models (SSMs). An input image sequence is first classified according to mean speed, then based on this classification, different ensembles of models are used to generate the final predictions for each set of statistics. We show that our proposed class-based ensembling method largely outperforms a simple ensembling method and multiple classical PT algorithms. Furthermore, our method significantly outperforms three popular classical PT algorithms and slightly outperforms TM (a SOTA classical algorithm) over the whole test set, and it significantly outperforms TM when only analyzing videos with high-speed particles. Finally, we find that DTS can offer significant time savings for the extraction of motion statistics compared to classical PT algorithms as it's measured to be around 6x faster than TM.

The work presented in the chapter is an enhanced version of the in-review article: "Berghouse, M. & Parashar, R. DeepTrackStat: an End-to-End Deep Learning Framework for Extraction of Motion Statistics from Videos of Particles. *Engineering Applications of Artificial Intelligence*. 2024." This chapter adds to the work currently in review by exploring connections with chapters 2 and 3.

4.3 Data and methods

4.3.1 Simulated data

DTS was developed with the goal of extracting motion statistics from a wide range of videos of particles. One of the challenges of this task is that there are no common benchmarks that currently exist for the specific task of measuring particle motion extraction capabilities. Thus, we developed a novel dataset containing a wide variety of particle tracking cases to properly test DTS. To create a highly general model, we generated over 2000 simulations of moving particles that were used

for training. All simulations were 40 frames long, 2000 by 2000 pixels, and 1 channel (grayscale). The simulations differed in image and motion properties to train the model on a wide variety of spatiotemporal conditions. Samples of the types of imagery and the distributions of motion statistics generated from our simulations can be seen in Figure 1. The image properties we varied were particle shape, size, density, seeding location, brightness, and signal to noise ratio (SNR). To change the SNR of the simulations, we used varying combinations of Gaussian, speckled, and salt and pepper noise. The motion properties we varied were particle speeds and pathlines. The pathlines were either generated randomly (to represent Brownian motion) or from flow fields representative of flow in porous media, straight advective flow, or Poiseuille flow. The porous media pathlines were largely generated from flow fields of heterogeneous geometries (created both in OpenFOAM [28] and via the Lattice Boltzman Method). A small percentage of the simulated data represents that of the experimental microfluidic geometry used in chapter 2 (chapter 2, Fig. 1c). This type of homogeneous geometry has been used for particle tracking studies in the fields of bacterial motility and deterministic lateral displacement [2, 29]. Likewise, Brownian motion, and Poiseuille and heterogeneous porous media flows represent a large range of the motion observed in the body of cell microscopy data. Thus, our image and particle motion varieties aim to capture the most commonly encountered types of video data for both microfluidics and general microscopy experiments. The simulated training and testing set don't significantly differ. Different parameters (particle speed, density, SNR, shape and seeding location) were used to generate the testing simulations than the training simulations, but the range of distributions of variables from the testing simulations generally falls within the range of distributions of variables from the training simulations (Fig. 1). In an effort to display the

robustness of DTS, the simulated test set aims to replicate most of the variation in the simulated training set.

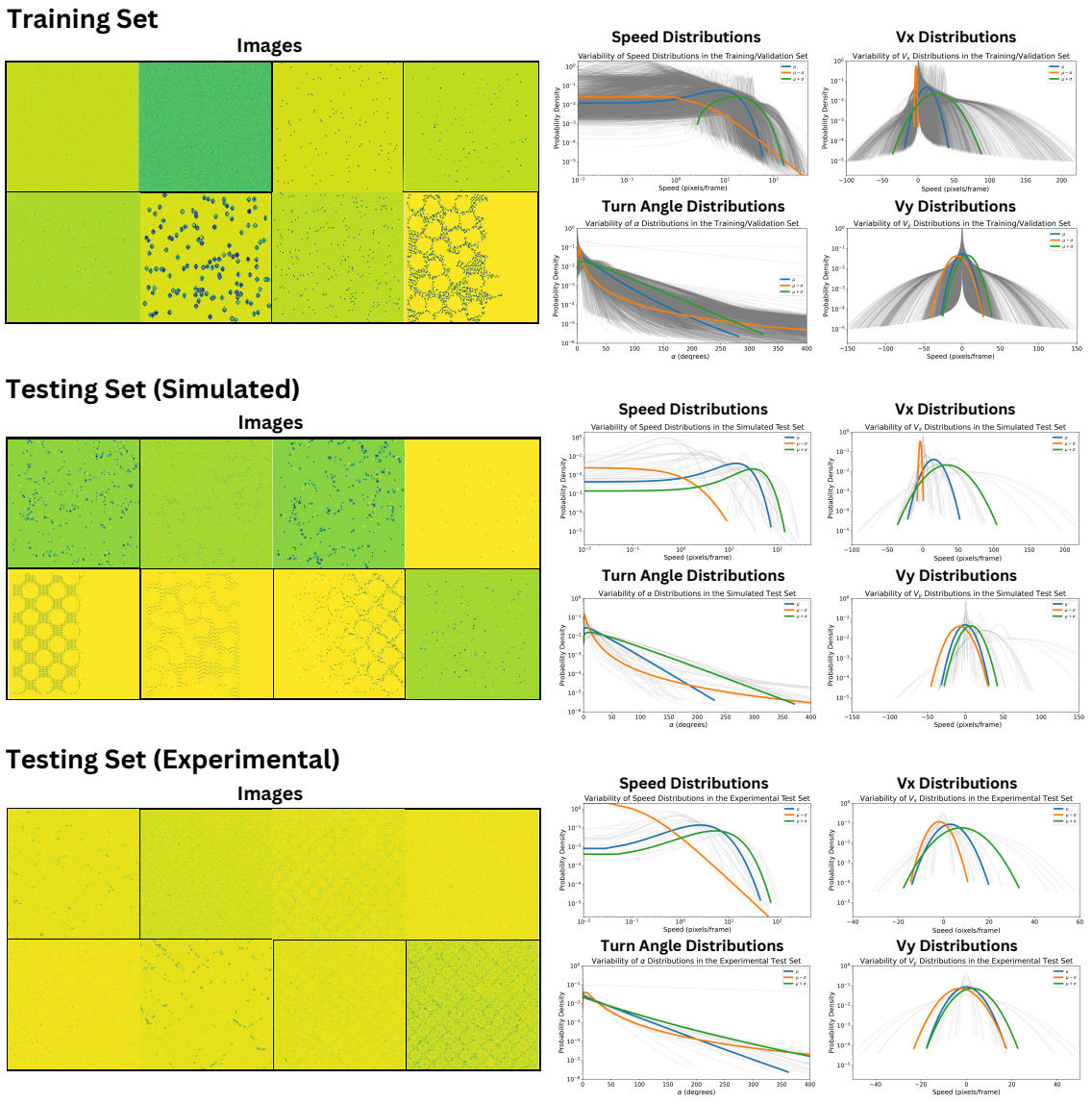


Figure 1. Images and motion statistic distributions for the training, simulated test, and experimental test data. The images are grayscale (1 channel) and are presented here as false-color images to highlight differences in brightness. The wide variety of images (in terms of particle density, shape, size, and image noise) illustrate the range of inputs that DTS is able to extract accurate predictions from. The blue distribution is the mean of the respective set (training, simulated testing, experimental testing), the orange is the mean minus one standard deviation, and the green is the mean plus one standard deviation. The grey distributions show the full range of variability for the respective set. The distributions for the simulated and experimental test datasets are mostly captured in the training dataset.

The aim of DTS is to predict speed, velocity component, and turn angle distributions directly from videos of moving particles. We chose to focus on the prediction of these motion statistics because they are important baseline measurements to understand the advective-diffusive transport of particles such as colloids and bacteria. However, we believe that frameworks like DTS can be extended to other statistics, so our work also serves as a proof of concept for researchers who may be looking for more task-relevant statistics such as dispersion coefficients or mean square displacements. For speed, we predict the magnitude of the ensemble velocity of the particles in pixels per frame as $S = \frac{\sqrt{(x_{t+1}-x_t)^2+(y_{t+1}-y_t)^2}}{\Delta t}$, where t represents time (in frames) and Δt = 1. For velocity, we predict the ensemble x and y velocity components (Vx and Vy) in pixels per frame. For turn angle, we predict the relative change in direction of the ensemble of particles between two successive frames as $\alpha_t = tan^{-1}\left(\frac{y_{t+2}-y_{t+1}}{x_{t+2}-x_{t+1}}\right) - tan^{-1}\left(\frac{y_{t+1}-y_t}{x_{t+1}-x_t}\right)$. Low average turn angle corresponds to particles that primarily move straight, and a high average turn angle corresponds to particles that have a high probability of changing directions between frames. One of the primary motivations behind this work lies in the results presented in chapter 3, which indicate a general inability for current state-of-the-art (SOTA) particle tracking algorithms to accurately high-speed particles. Thus, many of the simulations that were generated for training and testing have max speeds greater than 200 px/frame (Fig. 1), which is about twice the speed (in px/frame) of the fastest experimental data from chapter 2.

4.3.2 Experimental data

In addition to our simulated imagery/trajectories, we also test the performance of DTS on experimental videos. These videos are from microfluidics experiments of motile bacteria in porous and open media. Specifically, we use videos of Acidovorax [30], Geobacter [31], Paenibacillus [32],

and Shewanella [33] moving through structures with varying levels of porosity (ϕ = 0, ϕ = 0.42, and ϕ = 0.6) and at varying flow rates (0, 1, and 5 $\mu l/h$). Nine out of twenty-three of these experimental videos were also used in chapter 2 to analyze microbial motility. The other 14 videos come from various related experiments, meaning that they also have a high degree of similarity (in terms of image and particle motion properties) with the experimental videos from chapter 2. Because this is an experimental dataset, there is no ground truth. Thus, we use the results from TM as a relative ground truth to gauge the performance of DTS. The experimental videos range from 200 to 3000 frames, and are 2048 by 2048 pixels.

4.3.3 Model development

We present a novel end-to-end framework that consists of two stages: the speed classifier (SC) and the statistics- specific models (SSMs). Previous studies have presented similar class-based ensemble methods for various tasks [34, 35, 36]. However, our model is novel in its combination of architectures used, the use of a speed classifier to improve predictions of motion statistics, and the task of extracting information from videos of moving particles. The framework splits an input video into 40-frame chunks and averages the motion statistics predictions across all chunks, meaning the model can process any grayscale video input with at least 40 frames. The SC uses an ensemble of convolutional neural networks (CNNs) and vision transformers (ViTs) to classify the input into one of five classes based on speed. All models used in the ensemble can be seen in the publicly-available testing script, but the models that carry the most weight in the SC ensemble are VoloD1-384 [37], Pyramid Vision Transformer V2-b1 [38], RegnetX-032 [39], and VoloD3-448 [37], which were chosen for their high single-model performance. CNNs are well known to be able to capture spatial features within the images of a video [40], but may have trouble learning the temporal relationships in the data [41]. Thus, we also use ViTs, which are especially suited to learn

features in sequences of images [42], and have shown high performance on video classification tasks [43]. The speed classifier tries to predict ranges of mean particle speeds. Specifically, class 1 corresponds to a mean speed of 0-2 pixels/frame, class 2 is 2-5 pixels/frame, class 3 is 5-10 pixels/frame, class 4 is 10-18 pixels/frame, and class 5 corresponds to a mean particle speed of greater than 18 pixels/frame.

The second stage of the DTS framework, the SSMs, consists of a variety of ensemble models for each statistic and each class. The specific models used in each ensemble were determined by their single-model performance. DTS outputs a sorted 500-length vector of probable values for particle speeds, turn angles, and velocity components (Vx and Vy). Through this framework, the speed classifier has a large impact on the final results, with each specific SSM only slighlty shifting the value of the outputs. The ensemble weights for the SC and SSMs were determined through calibration of 50% of the simulated test data. In addition to calibration, we used simple boolean logic to improve DTS' performance on Brownian trajectories and trajectories that are relatively straight, but this feature has to be manually specified by the user. If the user knows a particular video contains primarily Brownian trajectories and sets this flag, then DTS will ensure that the output for Vx has a mean value of 0. Likewise, if the user observes that their video contains a large majority of particles that move straight, DTS will use a different ensemble for the turn angle distribution. Although DTS still has comparable performance to TrackMate without the use of these special flags, their use improves results for the specific cases of Brownian and straight particles. For all results discussed in this chapter, both flags were used to improve performance for these trajectory types.

The speed SSM primarily consists of a 4xVoloD1-224 patch model, VoloD3-448, VoloD1-384, and VoloD2-384. The patch model takes in a downsampled video input (448x448 pixels) and

splits it into four 224x224 patches. Each of these patches is then fed into a VoloD1-224 model with an output size of [B,500]. The outputs from each VoloD1-224 model are then concatenated and fed into a fully connected layer to get the final desired output shape of 500. For all other models, the final classification layer is simply replaced to get an output shape of 500. After each model generates its outputs, the class-based ensemble weights are used to generate the final model outputs. The exact models and ensemble weights used in each stage are given in Supplementary Figure 7. All base models were constructed with the PyTorch Image Models (TIMM) repository [44], meaning the final classification layer was changed via the "num_classes" flag. All models used the default pre-trained weights from TIMM (model-specific, but mostly ImageNet [45]). The turn angle, Vx, and Vy SSMs are constructed from similar ensembles, the details of which can be viewed in the code or Supplementary Figure 7.

4.3.4 Training and testing process

We used 1923 simulations for training and 481 simulations for validation, which was used to reduce overfitting during training via early stopping. The simulated test set contains 43 simulations and the experimental test set contains 23 pairs of images and trajectories. For TIMM models, the set dropout rate applies increasingly larger amounts of dropout in the transition layers of the model, with the final transition layer having the set amount of dropout. All patch-based models were trained with dropout of 0.3, and all other models were trained with dropout of 0.4. All models were trained with the AdamW optimizer at a learning rate (lr) of between 1.2e-5 and 2e-4. The speed classifier was trained with an lr of 2e-4, the speed SSM was trained with an lr of 1.2e-5, the Vy SSM was trained with an lr of 1e-4, the Vx SSM was trained with an lr of 1.2e-4, and the α SSM was trained with an lr of 6e-5. Hyperparameter tuning (dropout, learning rate, and number of classes) was done in a two step process. We tuned the hyperparameters automatically via

Optuna [46] for a few models for each statistic, then used the best range of learning rates to manually test a few sets of hyperparameters for each of the other models. All speed classifier

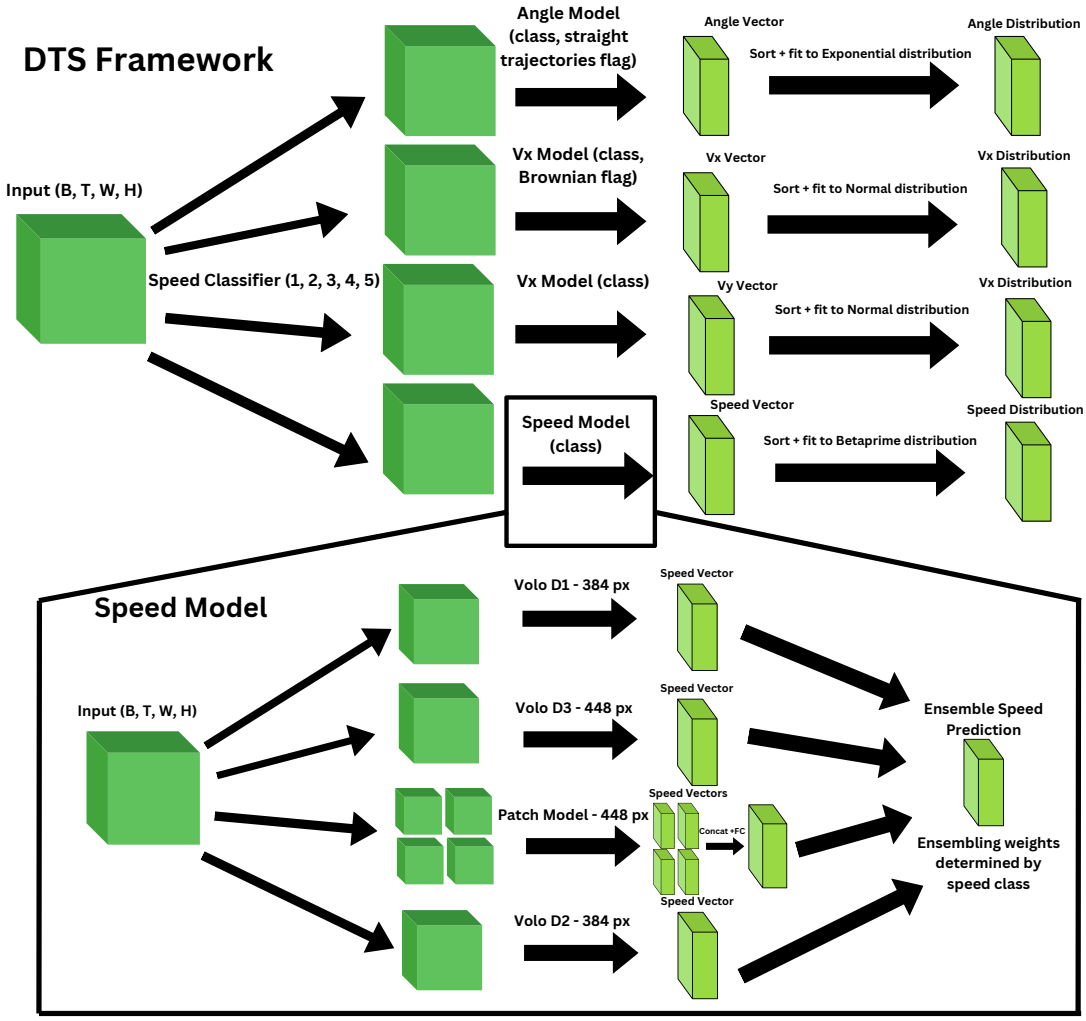


Figure 2: Overall framework for the proposed model (DTS). The model accepts a grayscale video as input (T must be 40, W and H must be equal) and first send it through the speed classifier (SC), which is an ensemble model used to classify the video of particles into 5 speed categories. The input is then sequentially sent to statistics-specific models (SSMs), which are each individually trained for their specific prediction task. Based on the output of the speed classifier, each SSM uses a different conditional ensemble model to generate the predictions. The outputs of DTS are the raw 500-length vectors of values for each statistic, and the respective distribution for each statistic.

models were trained for 400 epochs, speed prediction models were trained for 90 epochs, α models were trained for 300 epochs, Vy models were trained for 95 epochs, and Vx models were

trained for 150 epochs. All “224” (i.e. models that take in an input of 224x224) models were trained with a batch size of 32, all “384” models were trained with a batch size of 16, and all 448 models were trained with a batch size of 8. All training and testing for DTS, and all PT experiments, were performed on a CUDA-capable computer with an Nvidia 4090 GPU, Intel i9-14900KF CPU, and 96GB of RAM.

In this chapter we compare the performance of DTS to four other algorithms (TrackMate, Trackpy [47], TracTrac [48], and LapTrack [49]). For TrackMate, we used the Kalman filter linking algorithm for trajectories with directed motion and the LAP linking algorithm for trajectories with Brownian motion. Aggregate motion statistics for each classical PT algorithm (and the ground truth) were computed by ensemble averaging methods over all trajectories and frames. Essentially, each tracker outputs a csv of the trajectories that are sorted and looped through to calculate ensemble statistics. Statistics for DTS are calculated as the ensemble of all outputs from a single video.

Each model was calibrated to achieve the best results on the testing set. For TM, Trackpy, TracTrac and LapTrack, calibration was performed through a cycle of visual and statistical analysis to inform the adjustment of tracking parameters. For DTS, calibration entailed adjusting the class-based ensemble weights to achieve the best possible results on 50% of the testing data. The point of the calibration step is to simulate the scenario of using DTS to extract motion statistics from multiple videos. Given the rigorous testing in scientific literature that classical PT algorithms have gone through, it is reasonable to first use a classical PT framework (such as TM) to produce motion statistics in order to verify the accuracy of DTS for a particular dataset. If there are any significant discrepancies between DTS and TM, the ensemble weights of DTS can then be adjusted to match TM (or any other SOTA tracking algorithm), ensuring accurate statistics for the particular data

being analyzed. In addition to our calibrated results, we also provide uncalibrated results for DTS. In the uncalibrated version of DTS, the ensemble weights were determined through optimizing predictions of the validation set.

4.4 Results and discussion

4.4.1 Decreased run time

One of the primary advantages of DTS over classical PT algorithms (such as TM) is the reduced computation time for generation of motion statistics. Particle trajectory analysis often requires many imaging trials at high resolution, meaning the time required to extract motion statistics is an important concern. For one of our experimental videos with dense particles that contains 2480 frames, TM takes 2.5 minutes for loading the images into ImageJ [50], 5 minutes to perform the detection step, 4 minutes to perform the linking step, 0.5 minutes to filter and export the trajectories, and 0.5 minutes to calculate the statistics, which means the TM framework in total takes 12.5 minutes to extract statistics from the video data. This is assuming that OOM errors aren't encountered (a 2480 frame video of 2048x2048 resolution with >1000 particles in each frame may cause TM to crash) and that the tracking parameters used on the first try are optimal, which is unlikely for anyone besides an expert in the field. Even for someone experienced with PT codes, a 2480 frame video with over 1000 particles per frame will likely require 30 minutes to get good results. In stark contrast, DTS only takes 2 minutes to make its predictions for the same video, and requires significantly less domain knowledge to get accurate predictions of motion statistics. Additionally, DTS always takes two minutes for a 2480 frame video of 2048x2048 resolution, whereas the time required to generate statistics via classical PT methods significantly depends on

the number of trajectories. For videos with very few particles, DTS may not save much time, but for videos with a large number of particles (>1000), DTS will save a significant amount of time.

4.4.2 Ablation experiments

In order to show the benefit of our proposed model structure, we performed ablation experiments for each statistic. We report the mean average error (MAE) plus or minus the standard deviation of the mean value of each statistic across all simulated test data for the four best sets of single models, a simple ensemble of the best models, a single model that outputs all four variables at once, and our proposed class-based model (Table 1). The set of single models represent the top 4 single models for each variable. For example, MS-1 gives the results for a VoloD3-448 model used to calculate speed, a VoloD1-384 model used to calculate Vx, a VoloD4-448 model used to calculate Vy, and a VoloD3-448 model used to calculate α. In the 4-Var model, there are four separate VoloD1-224 models that generate the feature maps for each variable, then these four feature maps are concatenated and passed through a linear layer to generate the final output of shape [B, 500, 4]. In this case, we see performance is dramatically worse than that of the set of single models, the simple ensemble and DTS, which indicates the need to develop an ensemble of single models.

For all statistics besides Vx, we find that the class-based model largely outperforms the best sets of single models. Furthermore, DTS significantly ($p < 0.05$) outperforms a simple ensemble of the best single models for the speed and turn angle prediction tasks. In the case of Vx, although the MAE is greater for DTS than for model 1, other metrics (RMSE and W1) indicate that DTS has a better overall fit to the ground truth data. Thus, we illustrate that a class-based ensembling method can lead to significant performance increases for the task of predicting motion

Table 1. Ablation experiments on the simulated test set (n=43 samples). Scores are reported as the MAE for all samples in the test set plus or minus one standard deviation of the MAE between all samples of the test set. The metrics are non-negative with a large positive skew that often results in a standard deviation greater than the mean. For all proceeding tables, results are reported as the mean error with a 10th-90th range to clear any potential confusion. Here we give the results for the top 4 sets of single models, a simple ensemble of MS-1, MS-2, and MS-3, and a 4-var model that uses a single model (VoloD1-384) to predict all statistics at once, and DTS (the proposed framework). The best performing single models for speed are VoloD3-448 (MS-1 & MS-3) and the 4xVoloD1-224 patch model (MS-2 & MS-4). For Vx the best performing models are VoloD1-384 (MS-1), RegnetX-016 (MS-2), VoloD3-448 (MS-3), and RegnetX-032 (MS-4). For Vy the best performing models are VoloD4-448 (MS-1), VoloD3-448 (MS-2), VoloD1-384 (MS-3) and VoloD1-224 (MS-4). For turn angle the best performing models are VoloD3-448 (MS-1 & MS-3) and VoloD1-384 (MS-2 & MS-4). The class-based model significantly ($p < 0.05$) outperforms the simple ensemble for the speed predictions and slightly outperforms the ensemble in all other metrics.

Stat	MS-1	MS-2	MS-3	MS-4	Ensemble	4-Var Model	DTS
S	3.7 ± 5.0	4.2 ± 6.1	4.3 ± 6.0	9.7 ± 15	5.2 ± 7.8	10 ± 12	$\mathbf{2.8 \pm 3.3}$
Vx	5.7 ± 4.6	5.8 ± 4.8	6.0 ± 5.5	6.8 ± 6.4	5.7 ± 4.8	9.9 ± 12	5.1 ± 9.0
Vy	0.8 ± 0.7	0.9 ± 0.6	0.9 ± 0.9	1.0 ± 0.9	0.8 ± 0.8	6.5 ± 7.2	0.5 ± 0.5
α	3.4 ± 2.9	3.5 ± 2.7	3.7 ± 3.9	3.9 ± 4.1	3.0 ± 2.7	5.4 ± 5.3	2.5 ± 2.1

statistics from videos of particles. Furthermore, the class-based method contains a large number of parameters that can be manually fine- tuned (such as the Brownian and straight motion flags, and easily modifiable weights for the SC and SSMs), which allows for more precise calibration depending on the range of videos that need to be analyzed.

The full DTS framework has around 2 billion (1,994,005,238) parameters. The single models used in this study have between 8 and 200 million parameters. Using all the best single models, the full prediction framework would have around 330 million parameters. The ensemble framework uses 3 models for the prediction of each statistic, which gives it around 1 billion parameters. The 4-var model is just a single VoloD1-384 model for all statistics, so this would only have 26 million parameters. Although the 4-Var model performs significantly worse than all others,

this framework could be advantageous for situations where rapid predictions and low computational cost are required or preferred over accuracy.

4.4.3 Simulated test set

To ensure that DTS can handle a wide variety of simulated data, we included test simulations that had large variations in image and trajectory properties (Fig. 1). Our results indicate that the performance of DTS can match that of TM across this wide variety of simulations (Table 2). Specifically, DTS significantly outperforms TM in 2 out of 3 of the speed and angle metrics. Furthermore, DTS vastly outperforms other well-known PT methods such as Trackpy (Table 3), TracTrac (Table 4), and Laptrack (Table 5). In addition, we tried to compare the performance of DTS with a SOTA optical flow method [51], but determined that the model would require fine-tuning to perform the desired task (Supplementary Figure 8).

Complementary to our calibrated results for DTS, we also present results for an uncalibrated framework (Table 6). While the uncalibrated framework doesn't perform as well as the calibrated framework (Table 5), it does indicate that DTS can be used out-of-the-box to predict motion statistics with much greater accuracy than the calibrated predictions of Trackpy (Table 2), TracTrac (Table 3), and Laptrack (Table 4), and slightly better accuracy than the calibrated predictions of TrackMate (Table 5). Additionally, these results show that the class-based framework has a clear advantage over simple ensembling for speed and angle predictions (Table 1).

DTS shows especially strong performance for simulations with high particle speeds (Table 7). We define high-speed simulations as having a mean ensemble speed of greater than 25 pixels/frame. For our 12 test simulations that meet this criteria, DTS dramatically outperforms TM, showing improvement in every metric for the speed, Vx, and turn angle statistics. Once again, the

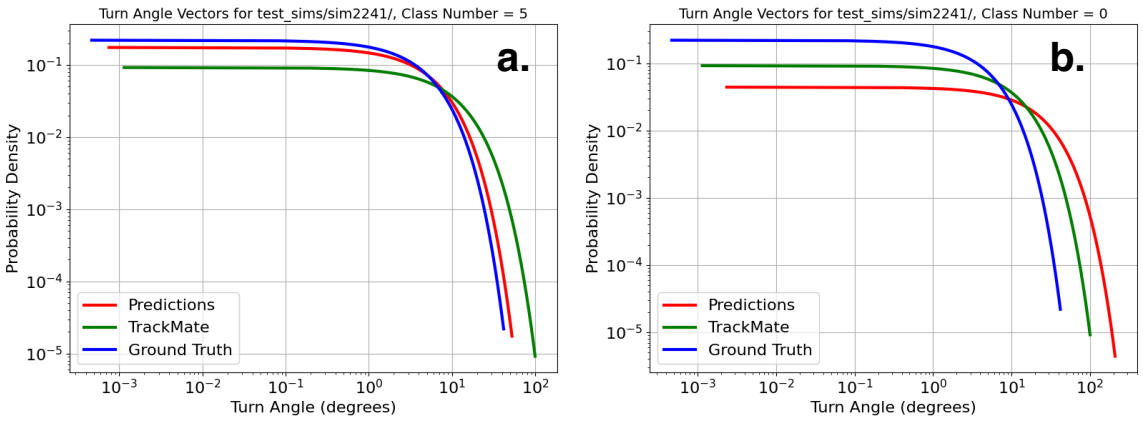


Figure 3: Impact on turn angle predictions of using the "straight trajectories flag" for a simulation of straight-moving particles. (a) Predictions with the flag on. (b) Predictions with the flag off. DTS is unable to accurately predict the turn angle distribution without manual help, illustrating the benefit of the flag, and the need to visually inspect the inputs before using DTS.

Table 2. Results from the simulated test set (n=43 samples) for each statistic comparing DTS and TM. The mean of each statistic is given along with the 90th-10th percentile error range. For both frameworks, we give the MAE, RMSE, and 1-Wasserstein distance (W1) for each statistic (relative to the ground truth). Statistically significantly better performances are bolded.

	DeepTrackStat			**TrackMate**		
Stat	MAE	RMSE	W1	MAE	RMSE	W1
Speed	**2.80** [**.116**, **7.18**]	**5.37** [**.839**, **11.6**]	.026 [.0005, .082]	7.41 [.030, 28.0]	11.1 [.610, 31.8]	.016 [.0015, .053]
Vx	5.05 [.047, 15.7]	8.48 [.828, 20.6]	.076 [.0013, .113]	7.95 [.004, .342]	12.0 [.712, 34.2]	.087 [.0010, .154]
Vy	.547 [.012, 1.27]	4.63 [.609, 12.9]	.026 [.0005, .108]	**.291** [**.005**, **.846**]	4.95 [.231, 11.9]	.021 [.0004, .090]
α	**2.51** [**.529**, **5.23**]	**5.41** [**1.57**, **9.63**]	.002 [.00002, .002]	5.49 [.987, 9.59]	10.5 [3.90, 20.8]	.002 [.00017, .004]

speed and angle predictions stand out, with both the MAE and RMSE showing a statistically significant improvement from TM. For traditional particle tracking methods (ie not based in deep learning methods) such as TM, the quality of the extracted trajectories is mainly determined by particle spacing displacement ratio (PSDR), which is the ratio of the average spacing between any two particles and the average speed of a particle. Since TM, and most other classical PT methods, are all roughly based on some kind of nearest neighbors approach, the lower the PSDR, the harder it is for them to accurately track the particles. At high particle speeds, the PSDR is low, so TM is unable to extract accurate trajectories. DTS, since it is not based on any kind of nearest neighbors algorithm and does not actually perform tracking, doesn't suffer from this issue. For speed, The MAE for DTS is about 5x less than that of TM, showing that DTS has a clear application for improving the accuracy of speed predictions for videos of high-speed particles.

In addition to our numerical performance comparison of DTS and TM, we also present a graphical performance comparison of the distributions of the simulated and experimental test sets (Fig. 4) for each statistic. The distributions obtained from the simulated test sets (Figs. 4a, 4b, 4c, and 4d) show that, on average, the distribution shapes obtained from DTS closely resemble the ground truth trajectories. Furthermore, for Vx, although the fit between distributions (as measured by W1) is essentially equal for TM and DTS (Table 2), the predictions by DTS are much more accurate for the high-speed simulations (Table 7). In addition, when looking at the distribution for all data, the predictions from DTS show much better alignment with the ground truth than TM does for the Vx statistic. This importantly shows that DTS has an equal chance of correctly predicting the true Vx distribution for individual videos, but a higher chance of predicting the true Vx distribution for a group of videos.

Table 3. Results from the simulated test set (n=43 samples) for each statistic for Trackpy. The mean of each statistic is given along with the 90th-10th percentile error range. The performance of Trackpy is considerably worse than that of DTS or TM.

Stat	MAE	RMSE	W1
S	8.44 [.306, 26.0]	10.8 [.508, 30.4]	.038 [.0008, .142]
Vx	8.87 [.392, 27.2]	11.6 [.662, 32.9]	.032 [.0015, .049]
Vy	3.42 [.174, 13.5]	4.87 [.364, 16.9]	.037 [.0002, .172]
α	8.42 [1.06, 20.4]	13.2 [1.77, 28.8]	.003 [.0002, .005]

Table 4. Results from the simulated test set (n=43 samples) for each statistic for TracTrac. The mean of each statistic is given along with the 90th-10th percentile error range. TracTrac is the lowest-performing PT method that was tested in this study.

Stat	MAE	RMSE	W1
S	12.0 [.029, 32.2]	14.2 [.068, 32.9]	.032 [.0017, .046]
Vx	12.8 [.045, 39.9]	14.8 [.156, 45.2]	.026 [.0022, .031]
Vy	4.06 [.023, 13.0]	5.32 [.179, 16.4]	.032 [.0015, .075]
α	14.5 [.548, 42.3]	14.5 [.548, 42.3]	.002 [.00007, .006]

Table 5. Results from the simulated test set (n=43 samples) for each statistic for Laptrack. The mean of each statistic is given along with the 90th-10th percentile error range. The performance of Laptrack is considerably worse than that of DTS or TM.

Stat	MAE	RMSE	W1
S	8.14 [.533, 22.3]	11.1 [1.33, 25.7]	.035 [.0006, .085]
Vx	8.69 [.584, 22.7]	12.5 [3.20, 29.6]	.032 [.0015, .063]
Vy	3.63 [.500, 10.4]	6.29 [1.19, 13.3]	.041 [.0006, .171]
α	10.4 [.964, 20.1]	15.5 [2.23, 29.0]	.003 [.0002, .005]

Table 6. Uncalibrated results from the simulated test set (n=43 samples) for each statistic for DTS. For these results, the ensemble weights and chosen models for DTS were determined via performance on the validation set, meaning these represent the general performance capabilities of DTS for completely unseen data. While the results aren't as strong as the calibrated ones, the errors in speed and angle prediction are still less than any other method tested in this chapter.

Stat	MAE	RMSE	W1
S	3.23 [.25, 7.7]	5.32 [.30, 12.1]	.021 [.00046, .039]
Vx	5.20 [.03, 20.2]	8.62 [.91, 22.4]	.026 [.00104, .077]
Vy	.627 [.03, 1.8]	4.25 [.39, 10.4]	.025 [.00153, .081]
α	2.58 [.53, 5.2]	5.35 [1.3, 9.81]	.002 [.00002, .002]

4.4.4 Experimental test set

The results from the experimental test set (Table 8) further indicate DTS' ability to accurately extract motion statistics from videos of moving particles. DTS shows strong alignment with TM for all statistics. The values are close enough that, given DTS' strong performance on simulated test set, it's unclear which distributions are more accurate. For example, the high speeds predicted by TM (Fig. 4e), such as up to 1000 pixels/frame, are highly unlikely for a video of bacteria in porous media flows at 2048x2048 resolution. This would mean that a particle could move across the camera's field of view in only two frames, which is not possible for the max flow speed (up to 800$\mu m/s$), frame rate (10 FPS), image magnification (0.325 px/μm) of our specific experiments.

A quick calculation shows the max speed a particle should be able to achieve in pixels/frame is about 246, meaning that DTS' estimate of a max speed of about 350 pixels/frame is in all likelihood more accurate than TM's estimate of 1000 pixels/frame. For the turn angle statistic derived from

Table 7. Average results from a high-speed subset (n=12 samples) of the simulated test data. Statistically significantly better performances are bolded. In the case of high-speed PT data, DTS performs better than TM in every metric.

	DeepTrackStat			**TrackMate**		
Stat	MAE	RMSE	W1	MAE	RMSE	W1
Speed	**4.33** [**1.25**, **9.35**]	**7.93** [**2.21**, **15.5**]	.002 [.0003, .003]	22.8 [2.36, 46.4]	27.5 [9.31, 51.9]	.004 [.002, .006]
Vx	12.9 [.718, 35.6]	18.6 [7.31, 43.4]	.050 [.0060, .027]	24.7 [2.40, 51.9]	28.9 [8.58, 56.4]	.146 [.0540, .161]
Vy	.882 [.249, 1.38]	8.70 [3.06, 16.3]	.004 [.0003, .004]	.775 [.190, 1.41]	9.51 [3.46, 17.2]	.003 [.0004, .009]
α	**2.85** [**.400**, **5.36**]	**5.03** [**1.24**, **7.88**]	.002 [.00005, .002]	5.60 [2.71, 7.68]	13.5 [5.35, 20.2]	.003 [.0006, .002]

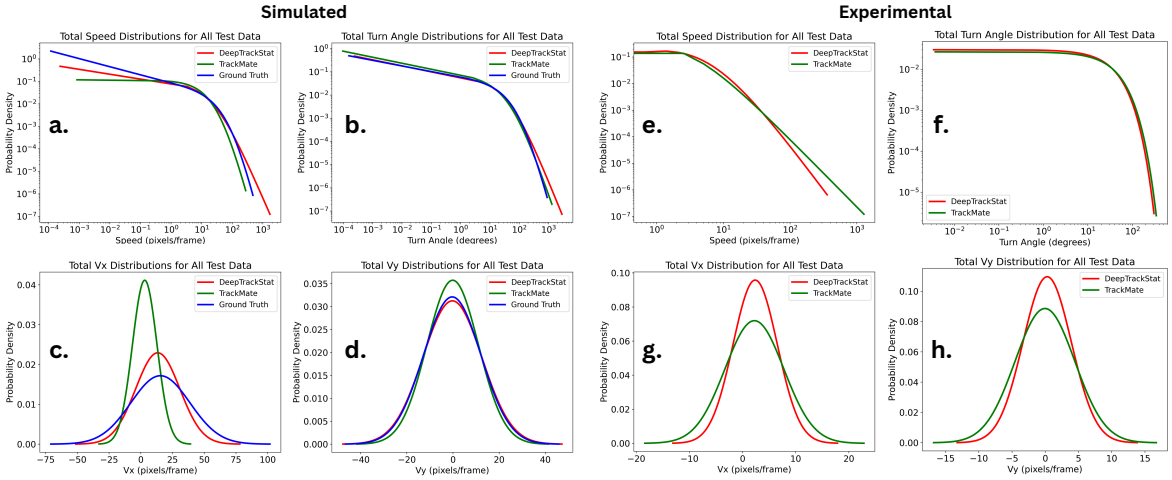


Figure 4: Distribution comparisons for simulated (a-d) and experimental (e-h) test sets for speed (a & e), turn angle (b & f), Vx (c & g), and Vy (d & h). For the simulated test set, we compare DTS and TM to the ground truth. For the experimental test set, we only compare DTS with TM, since there is no ground truth. Each distribution is obtained from a concatenated list of all values (from each individual simulation) for the respective statistic.

Table 8. Average results from the experimental test set (n=23 samples) for each statistic. Since there is no ground truth data for the experimental data, errors for DTS are calculated relative to TM.

Stat	MAE	RMSE	W1
S	1.98 [.312, 4.19]	5.36 [2.17, 11.8]	.026 [.006, .053]
Vx	1.26 [.072, 2.11]	2.29 [.757, 4.38]	.026 [.0056, .058]
Vy	.298 [.008, .793]	1.89 [.407, 2.95]	.033 [.002, .073]
α	7.05 [.724, 14.1]	10.8 [4.12, 17.1]	.001 [.00006, .002]

the average experimental tracks, DTS and TM have nearly identical exponential distributions (Fig. 4f). The Vx $(\mathrm{Fig.\ 4g})$ and $V\mathrm{y}$ distributions (Fig. 4h) show DTS accurately matches the mean of the TM output, but has a more narrow distribution.

To further explore the use of DTS on experimental data, we compare the turn angle distributions extracted via DTS with the turn angle distributions calculated from the PT-derived trajectories for videos of *Geobacter* and *Paenibacillus* in a $\phi = 0.42$ geometry for Q = 1 μL/h and Q = 5 μL/h (Fig. 5). These videos represent samples of the videos used in chapter 2 to investigate microbial transport at varying flow rates for bacterial of different motility types. Recalling the information from chapter 2 of this dissertation, *Geobacter* have twitching motility, whereas *Paenibacillus* have peritrichous flagella. For these tests, we recalibrated DTS by optimizing the class-based ensemble for performance on a separate set of videos for *Geobacter* and *Paenibacillus* at Q = 1 μL/h.

For the videos of *Geobacter* transport at Q = 5 μL/h and *Paenibacillus* transport at Q = 1 μL/h, we see a degree of agreement between DTS and TM. For *Geobacter* at Q = 1 μL/h, DTS predicts a smaller amount of turning (i.e., lower mean turn angle) than TM does, although the predictions are within the same ballpark relative to the other bacteria/flow rate combinations. For

Paenibacillus at Q = 5 μL/h, DTS predicts a larger amount of turning than TM does. This combination of over and underprediction by DTS results in slightly different conclusions than if one were to use TM. As discussed in chapter 2, a turn angle more closely centered around 0 (lower absolute mean) implies straight trajectories, which are more likely to occur with the twitching species due to their motility being dominated by advection. We observe these trends looking at

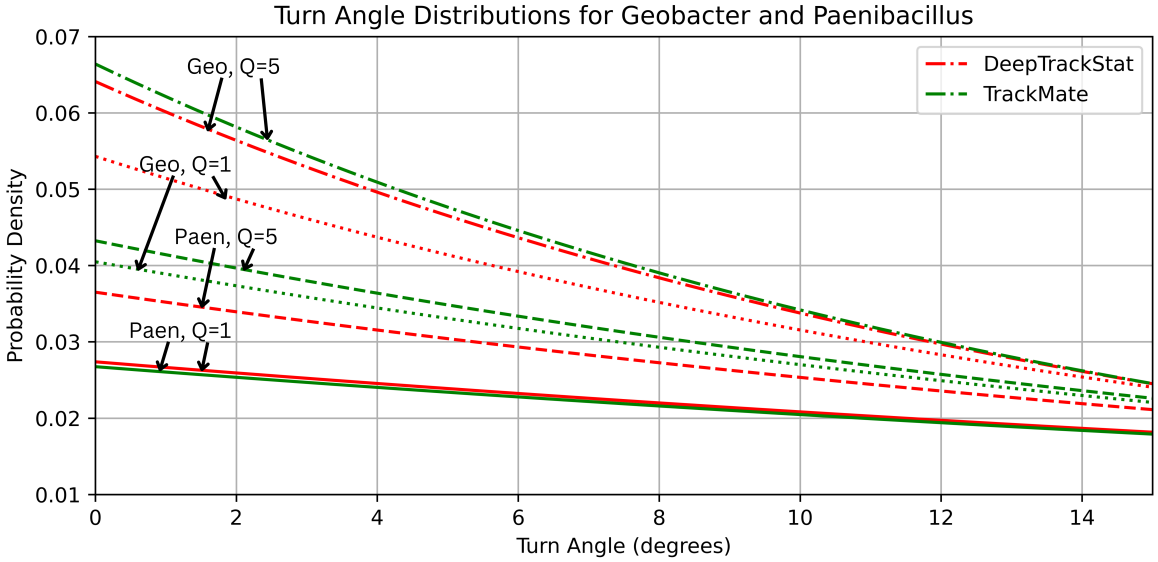


Figure 5. Turn angle distributions for Geobacter and Paenibacillus at Q = 5 μL/h and Q = 5 μL/h for DTS and TM. These distributions represent a subset of the results presented in chapter 2 to investigate bacterial transport in porous media. The results from DTS generally lead to the same conclusions about transport as the results from TM do.

the results for either DTS or TM, although DTS implies an even greater difference in the turn angles between species than TM does. Furthermore, the results from DTS indicate that there is less difference between the turn angles of the same species at different flow rates. Ultimately, because neither DTS or TM represent the ground truth, we can't comment on which method is more accurate for the experimental data. However, given the similarities of the conclusions, and the fact that there is no theoretical reason *Paenibacillus* should make less turns at Q = 5 μL/h than

Geobacter at Q = 1 μL/h, we posit that DTS, especially when calibrated, is accurate enough to replace particle tracking for the determination of turn angle distributions from experimental videos of bacteria in microfluidics.

4.5 Conclusions

We show that our proposed model, DeepTrackStat, achieves SOTA performance at comparatively rapid speeds for the general task of predicting speed, Vx, Vy, and turn angle distributions for a wide variety of particle tracking situations. Specifically, our model is capable of predicting these motion statistics for a large range of particle, image, and trajectory types (dispersive, Brownian, Poiseuille) about 6 times faster than via classical particle tracking algorithms. Through ablation experiments we show that our novel class-based ensembling method outperforms a simple ensembling method. We then show that DTS outperforms all classical PT algorithms used in this study for the prediction of motion statistics for our simulated test set, and we confirm the applicability of our models to real-world data by showing that the average outputs of DTS are comparable to the average outputs produced through TM. Furthermore, we show that the insights gained from analysis of turn angle distributions extracted by DTS are similar to those gained from the analysis of turn angle distributions extracted by TM. In addition, we highlight DTS' strong performance for the specific task of predicting statistics from videos of particles moving at high speeds. In this case, the performance of DTS greatly exceeds that of TM (the top-performing classical algorithm). Thus, we present a novel method for extraction of motion statistics and apply it to videos of particles. Although our models are specifically trained for the task of extracting statistics from videos of moving particles, our class-based ensembling framework can theoretically

be extended to extract motion statistics of any set of objects where the speed of the object is significantly correlated with the other motion statistics to be predicted.

Although we have shown the robust performance of DTS across a wide variety of image and motion types, there are many limitations present in this chapter that primarily revolve around scope. First and foremost, we recognize that a more rigorous study would include more simulated and experimental test sets. In addition, for most of the results DTS was calibrated on 50% of the simulated test data, so for particles with motion statistics that greatly fall outside of the training or calibration range, it is unlikely that DTS will perform well. This can be seen in the case of the straight trajectories (Fig. 3) - although DTS was trained on trajectories of similar types, straight-trajectory simulations made up a small percentage the entire training set. Thus, without the addition of a manually set flag to indicate that the particles are moving straight, DTS is not able to make accurate turn angle predictions. Another primary limitation of our work is that we have only compared DTS to classical PT algorithms. A more robust study might use fine tuning of SOTA object tracking algorithms to more effectively combine the domains of particle tracking and object tracking and determine more optimal network architectures. Finally, our work is limited in that it can only be used to predict four motion statistics. Rigorous transport studies often need more statistical evidence to make insightful claims, so our work could be improved by increasing the number of statistics DTS can accurately predict.

We hope that DTS is of practical use to researchers interested in applications of particle tracking. We have supplied the model weights at https://zenodo.org/records/11245477, and all data and scripts needed for training and testing can be found at https://github.com/mberghouse/DeepTrackStat. Furthermore, our training data represents one of the most comprehensive sets ground-truth particle tracking data publicly available on the

internet, and we believe researchers will find it useful for the development of even more robust applications related to PT. Thus, we hope that our work generally sparks interest within the research community about applications of computer vision for particle tracking and motion statistics predictions. Neither of these are solved problems yet, and improving these tasks can greatly improve research capabilities in the wide variety of fields that make use of them.

Acknowledgements

This research is based upon work supported by the U.S. Department of Energy (DOE) under award number DE- SC0019437. The second author (R.P) acknowledges career advancement support in the field of artificial intelligence based methods provided by NSF grant 2123481.

Data Availability

Model weights can be found at https://zenodo.org/records/11245477, and all data and scripts needed for training and testing can be found at https://github.com/mberghouse/DeepTrackStat. For any other inquiries related to data or code please contact the corresponding author.

References

[1] Scheidweiler, D., Miele, F., Peter, H., Battin, T. J. & de Anna, P. Trait-specific dispersal of bacteria in heterogeneous porous environments: from pore to porous medium scale. J. The Royal Soc. Interface 17, 20200046, DOI: 10.1098/rsif.2020.0046 (2020).

[2] Dehkharghani, A., Waisbord, N., Dunkel, J. & Guasto, J. Bacterial scattering in microfluidic crystal flows reveals giant active taylor–aris dispersion. Proc. Natl. Acad. Sci. 116(23), 11119–11124 (2019).

[3] Bhattacharjee, T. & Datta, S. Bacterial hopping and trapping in porous media. Nat. communications 10(1), 2075 (2019).

[4] Dentz, M., Creppy, A., Douarche, C., Clément, E. & Auradou, H. Dispersion of motile bacteria in a porous medium. J. Fluid Mech. 946, A33 (2022).

[5] Creppy, A., Clément, E., Douarche, C., D'angelo, M. & Auradou, H. Effect of motility on the transport of bacteria populations through a porous medium. Phys. Rev. Fluids 4(1), 013102 (2019).

[6] Dehkharghani, A., Waisbord, N. & Guasto, J. S. Self-transport of swimming bacteria is impaired by porous microstructure. Commun. Phys. 6, 18, DOI: 10.1038/s42005-023-01136-w (2023).

[7] Rusconi, R., Guasto, J. S. & Stocker, R. Bacterial transport suppressed by fluid shear. Nat. Phys. 10, 212–217, DOI: 11 10.1038/nphys2883 (2014).

[8] Bhattacharjee, T. & Datta, S. S. Bacterial hopping and trapping in porous media. Nat. Commun. 10, 2075, DOI: 413 10.1038/s41467-019-10115-1 (2019).
[9] Saxton, M. Single-particle tracking: the distribution of diffusion coefficients. Biophys. journal 72(4), 1744–1753 (1997).
[10] Hong, Q., Sheetz, M. & Elson, E. Single particle tracking. analysis of diffusion and flow in two-dimensional systems. Biophys. journal 60(4), 910–921 (1991).
[11] Forier, K., Messiaen, A.S., Raemdonck, K., Deschout, H., Rejman, J., De Baets, F., Nelis, H., De Smedt, S.C., Demeester, J., Coenye, T. & Braeckmans, K. Transport of nanoparticles in cystic fibrosis sputum and bacterial biofilms by single-particle tracking microscopy. Nanomedicine 8(6), 935–949 (2013).
[12] Jeon, H., Lee, Y., Jin, S., Koo, S., Lee, C.S. & Yoo, J.Y. Quantitative analysis of single bacterial chemotaxis using a linear concentration gradient microchannel. Biomed. microdevices 11, 1135–1143 (2009).
[13] Partridge, J.D., Nhu, N.T., Dufour, Y.S. & Harshey, R.M. Escherichia coli remodels the chemotaxis pathway for swarming. MBio, 10(2), 10-1128 (2019).
[14] Kim, D., Liu, A., Diller, E. & Sitti, M. Chemotactic steering of bacteria propelled microbeads. Biomedical microdevices, 14, 1009-1017 (2012).
[15] Vallotton, P., Van Oijen, A.M., Whitchurch, C.B., Gelfand, V., Yeo, L., Tsiavaliaris, G., Heinrich, S., Dultz, E., Weis, K. & Grünwald, D., 2017. Diatrack particle tracking software: Review of applications and performance evaluation. Traffic 18(12), 840–852 (2017).
[16] Linkhorst, J., Beckmann, T., Go, D., Kuehne, A. J. & Wessling, M. Microfluidic colloid filtration. Sci. reports 6(1), 22376422 (2016).
[17] Berghouse, M., Miele, F., Perez, L., Bordoloi, A., Morales, V., Parashar, R. Evaluation of Particle Tracking Codes for Dispersing Particles in Porous Media. Scientific Reports, (2024).
[18] Chenouard, N., Smal, I., De Chaumont, F., Maška, M., Sbalzarini, I.F., Gong, Y., Cardinale, J., Carthel, C., Coraluppi, S., Winter, M. & Cohen, A.R. Objective comparison of particle tracking methods. Nature methods 11(3), 281–289 (2014).
[19] Midtvedt, B., Pineda, J., Skärberg, F., Olsén, E., Bachimanchi, H., Wesén, E., Esbjörner, E.K., Selander, E., Höök, F., Midtvedt, D. & Volpe, G. Single-shot self-supervised particle tracking. arXiv preprint arXiv:2202.13546 (2022).
[20] Midtvedt, B., Pineda, J., Skärberg, F., Olsén, E., Bachimanchi, H., Wesén, E., Esbjörner, E.K., Selander, E., Höök, F., Midtvedt, D. & Volpe, G. Single-shot self-supervised object detection in microscopy. Nature communications, 13(1), 7492 (2022).
[21] Ershov, D., Phan, M.S., Pylvänäinen, J.W., Rigaud, S.U., Le Blanc, L., Charles-Orszag, A., Conway, J.R., Laine, R.F., Roy, N.H., Bonazzi, D. & Duménil, G. Bringing TrackMate into the era of machine-learning and deep-learning. BioRxiv, (2021).
[22] Kapoor, V. & Carabaña, C. Cell Tracking in 3D using deep learning segmentations. In Python in Science Conference, 154-161 (2021).
[23] Schmidt, U., Weigert, M., Broaddus, C. & Myers, G. Cell detection with star-convex polygons. Springer International Publishing. In Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part II 11, 265-273, (2018).
[24] Tinevez, J.Y., Perry, N., Schindelin, J., Hoopes, G.M., Reynolds, G.D., Laplantine, E., Bednarek, S.Y., Shorte, S.L. and Eliceiri, K.W., 2017. TrackMate: An open and extensible platform for single-particle tracking. Methods, 115, pp.80-90.
[25] Fatemi, B., Halcrow, J. & Jaqaman, K. Geometric deep learning of particle motion by MAGIK. Nature Machine Intelligence, 5(5), 483-484 (2023).
[26] Zou, Z., Chen, K., Shi, Z., Guo, Y. and Ye, J., 2023. Object detection in 20 years: A survey. Proceedings of the IEEE, 111(3), pp.257-276.
[27] Amosa, T.I., Sebastian, P., Izhar, L.I., Ibrahim, O., Ayinla, L.S., Bahashwan, A.A., Bala, A. and Samaila, Y.A., 2023. Multi-camera multi-object tracking: a review of current trends and future advances. Neurocomputing, 552, p.126558.
[28] Weller, H., Tabor, G., Jasak, H. & Fureby, C. A tensorial approach to computational continuum mechanics using449 object-oriented techniques. Comput. Phys. 12, 620 (1998).
[29] Zeming, K.K., Salafi, T., Chen, C.H. & Zhang, Y. Asymmetrical deterministic lateral displacement gaps for dual functions of enhanced separation and throughput of red blood cells. Scientific reports, 6(1), 22934 (2016).
[30] Lee, J.H., Fredrickson, J.K., Plymale, A.E., Dohnalkova, A.C., Resch, C.T., McKinley, J.P. & Shi, L. An autotrophic H 2-oxidizing, nitrate-respiring, Tc(VII)-reducing Acidovorax sp. isolated from a subsurface oxic-anoxic transition zone: H 2 -oxidizing, Tc-reducing Acidovorax spp. Environmental Microbiology Reports 7, 395–403 (2015).
[31] Caccavo Jr, F., Lonergan, D.J., Lovley, D.R., Davis, M., Stolz, J.F. & McInerney, M.J. Geobacter sulfurreducens sp. nov., a hydrogen- and acetate-oxidizing dissimilatory metal-reducing microorganism. Appl Environ Microbiol 60, 3752–3759 (1994).
[32] Ahmed, B., Cao, B., McLean, J.S., Ica, T., Dohnalkova, A., Istanbullu, O., Paksoy, A., Fredrickson, J.K. & Beyenal, H. Fe(III) Reduction and U(VI) Immobilization by Paenibacillus sp. Strain 300A, Isolated from Hanford 300A Subsurface Sediments. Appl Environ Microbiol 78, 8001–8009 (2012).
[33] Hau, H.H. & Gralnick, J.A. Ecology and biotechnology of the genus Shewanella. Annu. Rev. Microbiol., 61, 237-258 (2007).

[34] Ohi, A.Q., Mridha, M.F., Hamid, M.A., Monowar, M.M. and Kateb, F.A. Fabricnet: A fiber recognition architecture using ensemble convnets. IEEE Access, 9, 13224-13236 (2021).
[35] Al-Khateeb, T., Masud, M.M., Khan, L., Aggarwal, C., Han, J. and Thuraisingham, B. Stream classification with recurring and novel class detection using class-based ensemble. In 2012 IEEE 12th international conference on data mining, 31-40, IEEE (2012).
[36] Ali, M., Zhu, P., Jiang, R., Huolin, M., Ehsan, M., Hussain, W., Zhang, H., Ashraf, U. and Ullaah, J. Reservoir characterization through comprehensive modeling of elastic logs prediction in heterogeneous rocks using unsupervised clustering and class-based ensemble machine learning. Applied Soft Computing, 148, 110843 (2023).
[37] Yuan, L., Hou, Q., Jiang, Z., Feng, J. & Yan, S. Volo: Vision outlooker for visual recognition. IEEE transactions on pattern analysis and machine intelligence, 45(5), 6575-6586 (2022).
[38] Wang, W., Xie, E., Li, X., Fan, D.P., Song, K., Liang, D., Lu, T., Luo, P. & Shao, L. Pvt v2: Improved baselines with pyramid vision transformer. Computational Visual Media, 8(3), 415-424 (2022).
[39] Radosavovic, I., Kosaraju, R.P., Girshick, R., He, K., & Dollár, P.: Designing Network Design Spaces. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 10425-10433, Seattle, WA, USA (2020).
[40] Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R. & Fei-Fei, L. Large-scale video classification with convolutional neural networks. In Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, 1725-1732 (2014).
[41] Chen, J. & Ho, C.M. MM-ViT: Multi-modal video transformer for compressed video action recognition. In Proceedings of the IEEE/CVF winter conference on applications of computer vision, 1910-1921 (2022).
[42] Arnab, A., Dehghani, M., Heigold, G., Sun, C., Lučić, M. & Schmid, C. Vivit: A video vision transformer. In Proceedings of the IEEE/CVF international conference on computer vision, 6836-6846 (2021).
[43] Li, Y., Wu, C.Y., Fan, H., Mangalam, K., Xiong, B., Malik, J. & Feichtenhofer, C. Mvitv2: Improved multiscale vision transformers for classification and detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 4804-4814 (2022).
[44] Ross Wightman. PyTorch Image Models. Github, Github repository (2019). doi: 10.5281/zenodo.4414861.
[45] Deng, J., Dong, W., Socher, R., Li, L.J., Li, K. and Fei-Fei, L. Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition, 248-255 (2009).
[46] Akiba, T., Sano, S., Yanase, T., Ohta, T. and Koyama, M. Optuna: A next-generation hyperparameter optimization framework. In Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining, 2623-2631 (2019).
[47] Allan, D. B. Trackpy. “soft-matter/trackpy: v0.6.4”. Zenodo (2024). doi: 10.5281/zenodo.12708864.
[48] Heyman, J. TracTrac: A fast multi-object tracking algorithm for motion estimation. Computers & Geosciences, 128, 11-18 (2019).
[49] Fukai, Y.T. and Kawaguchi, K. LapTrack: linear assignment particle tracking with tunable metrics. Bioinformatics, 39(1), 799 (2023).
[50] Abràmoff, M.D., Magalhães, P.J. and Ram, S.J., 2004. Image processing with ImageJ. Biophotonics international, 11(7), 36-42, (2004).
[51] Harley, A.W., Fang, Z., Fragkiadaki, K. Particle Video Revisited: Tracking Through Occlusions Using Point Trajectories. In ECCV (2022).

Chapter 5: Investigation of Feedback Cycles and the Impacts of Speed-Based Biomass Decay on Biomass Growth and Chromium Reduction in the Hyporheic Zone

5.1 Abstract

Within the hyporheic zone, a complex interplay of abiotic processes dictates the growth conditions of biomass. Given the hyporheic zone's potential role to bioremediate contaminants through biotic and abiotic reduction, decoding these growth determinants has broader ecological significance. In this study, we present a Monte-Carlo-style exploration into how varied initial conditions influence biomass growth and reduction of heavy metals using chromium as an example. Our modeling approach simulates a two-dimensional, meter scale cross-section of the hyporheic zone, integrating heterogeneous permeabilities and accurate hyporheic flux, and modeling chromium reduction through Monod kinetics.

To effectively capture bioclogging dynamics and soil respiration, we've enhanced the reactive transport model, PFLOTRAN. Our expanded model accounts for biomass decay influenced by fluid speed, and the dependency of biomass growth on temperature. We examine the speed-based biomass decay function by providing a sensitivity analysis of the impact of different parameter values on biomass growth and chromium reduction. Additionally, we conduct a large number of simulations to offer holistic insights into microbial growth dynamics through mean trend analysis, mean spatial distribution analysis, sensitivity analysis, PCA and clustering, and correlation heatmaps. This analysis reveals several insights into the feedback cycles and trends of biomass growth in the hyporheic zone under varying hydro-biogeochemical settings and shows that while abiotic reduction is generally more dominant than biotic reduction, high biomass concentrations give rise to reduction hotspots. While abiotic reduction largely determines the temporal distribution of average chromium concentrations in the domain, biotic reduction

controls the average spatial distribution of chromium in the domain. In summary, this chapter contains a theoretical model for biomass growth in saturated porous environments, a working version of our biomass growth model in PFLOTRAN, and analysis of the feedback cycles and variable relationships produced through 2D hyporheic zone simulations.

5.2 Introduction

Historically, studies in the hyporheic zone (HZ) have revolved around understanding flow dynamics [1, 2, 3], nutrient flux [4, 5, 6], and biofilm growth [7, 8, 9, 10]. Most of the current research on biomass growth in the HZ relates to the ecological significance of microbial communities [11, 12, 13, 14], but recent research has also illuminated the impact of biofilms on physical hydrologic properties such as permeability and local flow speed [9, 10, 15, 16]. Specifically, the emergence of biofilms, which causes a decrease in permeability known as bioclogging, has been recognized as a significant factor affecting water flow and solute transport [7, 17, 18]. Furthermore, the properties of biofilms such as strength, density and stickiness (adhesion and cohesion), and variations in the geometry of the porous media such as grain shape, impact the relative changes in permeability [17]. The degree of bioclogging is also dependent on flow speed. As speed increases, the increased fluid shear causes a breakup of the biofilm which represents a local decrease in biomass concentration.

Biomass growth in the HZ can be understood as a complex feedback loop dependent on a myriad of individual phenomena. Flow, as an initial causal variable in most cases, acts as the foundation for this feedback loop. It instigates a flux of temperatures and nutrient concentrations within the hyporheic zone. This flux, in turn, triggers changes in biomass concentrations. Generally, higher nutrient levels and increased temperatures tend to promote growth [19, 20, 21, 22].

However, the relationship between temperature and biomass is somewhat nuanced. For instance, an uptick in temperature leads to reduced water viscosity, which subsequently results in increased flow speeds. Increased fluid speed may bring in nutrient-rich or nutrient-poor water, resulting in changes to the biomass growth within the domain. Furthermore, high shear rates caused by increased fluid speed can result in the breaking apart and subsequent transport of biomass. However, the coupling of these multiple phenomena in Darcy-scale remediation simulations has not yet been investigated, meaning the exact impact of temperature and other abiotic factors on biomass growth in the HZ is not fully understood.

Biomass growth can also influence a soil's permeability [23, 24]. As biofilms become denser, they lead to bioclogging, reducing soil permeability. This impedes the nutrient dispersion across the HZ, subsequently slowing down biomass growth. This dynamic between biomass and permeability creates a negative feedback loop within the overarching feedback mechanisms. More biomass leads to reduced permeability, which in turn diminishes biomass growth. Many studies have investigated how bioclogging alters permeability [51-53], but few studies have simulated the impacts of permeability on biomass growth [54]. Furthermore, bioclogging simulations are often done in the context of homogeneous permeability simulations, meaning the impacts of soil heterogeneity on bioclogging and biomass growth are not well understood.

Reactive transport (RT) simulators [25, 26, 27] have emerged as powerful tools that allow for investigation and prediction of phenomena within the hyporheic zone [28, 29, 30, 31]. Field and lab studies are often resource-intensive and require a large amount of time. Although not able to capture the same novelty and richness of information available from in-situ measurements, RT simulations offer a highly analytical viewpoint of known physical and chemical phenomena that can be accurately described through systems of equations. RT simulators are constantly being

updated to include newly understood interactions. However, given their vast complexities, it is inevitable that there will always be new relevant models that have yet to be applied to RT simulators, such as the speed-based decay of biofilms.

In this chapter, we aim to provide two major contributions. First, we seek to further our understanding of the hyporheic zone by investigating the complex relationships governing biomass growth for a variety of input conditions. Furthermore, we highlight the nuanced relationships between abiotic reduction and bacterial proliferation and highlight how these relationships impact heavy meatal reduction (using chromium as an example) in the hyporheic zone. Second, we seek to improve the representation of microbial physics in reactive transport simulations by adding temperature-based biomass growth and speed-based biomass decay to PFLOTRAN, a popular reactive transport simulator. To illustrate the impacts of our augmentations, we highlight the sensitivity of temperature-based biomass growth and the sensitivity of speed-based biomass decay in terms of the overall biomass growth and the reduction of chromium.

5.3 Methods

5.3.1 Description of RT Simulations

5.3.1.1 Description of Chrotran

Our simulations are built upon PFLOTRAN [26], a sophisticated multi-physics reactive transport simulator developed collaboratively by multiple national laboratories. Specifically, we have adapted the Chrotran [32] version of PFLOTRAN to create a high-complexity simulation of biomass growth in the hyporheic zone at the Darcy scale. Chrotran models the dynamics of five key species: a heavy metal contaminant (in our case, Cr(VI)), an electron donor, biomass, a non-toxic conservative bio-inhibitor, and a biocide. It incorporates both direct abiotic reduction through

donor-metal interaction and biotic reduction driven by donor-induced biomass growth. Chrotran uses Monod kinetics to define biomass growth as a function of electron donor concentration, with additional factors accounting for biomass crowding, inhibition, and decay. We chose to make the input concentrations of bio-inhibitor and biocide low to simplify the analysis of our simulations. Chrotran also includes crucial processes such as donor sorption (through a mobile-immobile mass transfer system), bio-fouling, and biomass death. Furthermore, Chrotran allows for bioclogging modeling by dynamically updating porosity and hydraulic conductivity based on biomass concentration. The software can handle heterogeneous flow fields and arbitrarily many chemical species and amendment injection points, featuring full coupling between flow and reactive transport. For a comprehensive description of Chrotran's capabilities and mathematical formulation, please refer to Hansen et. al [32].

5.3.1.2 Speed-Based Biomass Decay

We made certain modifications to the published version of Chrotran by adjusting its parameters. Specifically, we calibrated biomass growth rates to align with the limited data available on biomass growth in the hyporheic zone [33, 34, 35]. Additionally, we adjusted steady-state concentrations to be broadly representative of those found in wetland environments [36].

The original Chrotran model (Equation 1) defines biomass decay using a simple linear function that only depends on a natural decay factor (λ_{B_2}) and the difference between the biomass concentration (B) and a minimum biomass threshold (B_0).

(1) $\lambda_b = \lambda_{B_2}(B - B_0)$

However, recent research suggests that biomass decay can also be influenced by shear stress [17]. Specifically, it has bee shown that the thickness of biofilms has an inverse logarithmic relationship with shear stress. In our study, we consider biofilm and biomass to be interchangeable. At high

flow speeds, significant shear stresses can form potentially dislodging and transporting biofilms and bacterial deposits. While ideally, we would incorporate shear stress values directly into Chrotran's biomass decay function, PFLOTRAN lacks the capability to calculate these stresses. As an alternative, we developed a function that calculates cell-specific biomass decay based on cell-specific Darcy velocity magnitude and biomass concentration (Equation 2). The decay rate λ_b is given in units of $\frac{mol}{m^3 \cdot hr}$. Since B is given in $\frac{mol}{m^3}$ and v is given in m/hr, β must have units of $\frac{m^3}{mol}$ and α must have units of m/h. The units of the natural decay factor, λ_{B_2}, depend on $\beta(B - B_0)$. For example, if $\beta(B - B_0) = 1$, then λ_{B_2} must have units of $\frac{mol}{m^4}$. This contrasts with the original biomass decay equation for Chrotran, where λ_{B_2} has units of t^{-1}. Thus, in our interpretation of biomass decay, the natural decay parameter can be considered a function of any environmental factors not represented by α and β.

(2) $\lambda_b = \lambda_{B_2}(v - \alpha)^{\beta(B-B_0)}$

For the main parameters of the equations ($\beta, \alpha, \lambda_{B_2}$), we provide a sensitivity analysis with $B = 1$ mol/m^3, $\lambda_{B_2} = 1.8 \times 10^{-5}$ m^3/mol, and $B_0 = 1 \times 10^{-10}$ mol/m^3 for v in the range of $[1 \times 10^{-8},\ 100]$ m/hr (Fig. 1). The sensitivity analysis of λ_{B_2} shows that the natural decay parameter simply increases the amount of decay for all valid velocity magnitudes (greater than α). The sensitivity analysis for β shows that low values of β correspond to a relatively low slope (i.e. small increase in decay as speed increases) but relatively high values of decay at low speeds (Fig. 4b). We interpret the β parameter to represent the mechanics of the biofilm decay based on factors such as adhesive and cohesive forces. In general, bacterial adhesion to surfaces (such as soil grains) is weaker than cell-to-cell cohesion in mature biofilms. This is because initial attachment to surfaces (adhesion) relies primarily on Van der Waals forces, electrostatic interactions, and specific molecular interactions between bacterial surface proteins and the

substrate. On the other hand, cell-to-cell cohesion in biofilms involves production of extracellular polymeric substances (EPS) that form a strong matrix, multiple types of chemical bonds between cells, physical entanglement of cells and matrix components, and the development of specialized structures for cell-cell attachment. The stronger cohesive forces in biofilms explain why mature biofilms often detach as whole chunks (i.e., sloughing) rather than individual cells when exposed to shear forces. The bacteria essentially create their own reinforced "community structure" that's more robust than their initial surface attachment.

The case of low β describes an immature biofilm in which the individual bacteria are unable to rapidly colonize surfaces and have weak adhesive forces that cause detachment at low flow speeds. As speed increases, the decay also increases due to this weak adhesion which results in sloughing. The case of high β describes a mature biofilm that has stronger adhesive forces at low flow speed (since most of the shear force impacts the outer walls of the biofilm), but weak adhesive and cohesive forces at high speed. In this case, the mature biofilm is thicker, and high shear stresses may lead to sloughing events, thus causing a larger increase in biomass decay (relative to the case of low β). While these phenomena may accurately describe situations for specific biofilms, we recognize that a more robust and flexible version of this parameter would also allow for changes in the slope (i.e., different values of β) for different values of velocity, since there are a variety of valid decay-speed slopes depending on the physiochemical characteristics of the biofilm and the environmental conditions. However, when we attempted to make the exponent term a function of velocity as well, the simulations in PFLOTRAN became divergent.

The sensitivity analysis for α shows it has the least impact on biomass decay of the three fitting parameters. Physically, α can be thought of as representing the threshold fluid speeds required to initiate shearing at the periphery of the biofilm. For our implementation in PFLOTRAN,

we use a conditional statement that returns the value of $(v - \alpha)$ if $v > \alpha$, else we return 10^{-10} in place of $(v - \alpha)$. The underlying premise of our biomass decay equation is that initial scouring establishes a steady-state biomass concentration at a given flow velocity. Subsequent increases in flow velocity lead to enhanced scouring, attributed to both the elevated flow speed and the formation of preferential flow channels within the biofilm matrix. We calibrated the fitting parameters α and β through a comparative sensitivity analysis with published research on biofilm thickness as a function of shear stress [17]. It is well-established that the distribution of micro and pore-scale

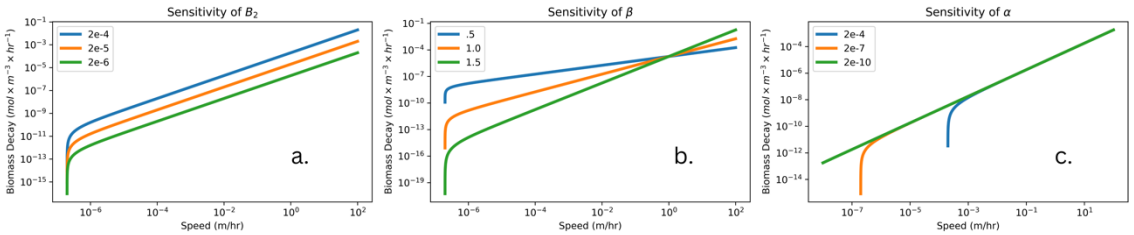


Figure 1: Sensitivity analysis of parameters in the augmented biomass decay equation for a biomass concentration of 1 mol/m^3. (a) Sensitivity of the λ_{B_2} parameter. Higher values of λ_{B_2} result in a larger amount of initial decay for speed>1e-4, but don't impact the long-term slope of the relationship between speed and decay. (b) Sensitivity of the β parameter. Higher values of β result in a smaller amount of initial decay but a greater slope. (c) Sensitivity of the α parameter. Higher values of α result in a greater increase in initial decay, but the speed required to produce the initial decay is greater.

velocities can diverge significantly from Darcy-scale velocities [37]. Given that biofilms primarily develop within pore spaces, they are subject to pore-scale velocity variations, resulting in higher shear stresses on the biofilms compared to Darcy-scale calculations. To address this scale discrepancy, we employed OpenFOAM [38, 39] to simulate two homogeneous porous geometries, each measuring 665 μm by 665 μm, with a porosity of 25% but differing in pore length and grain diameter (Fig. 2). Both simulations revealed an approximate order of magnitude difference between mean and maximum shear stresses. Consequently, we applied a pore-scale correction

factor of 10, as described in the caption of Figure 2, to adjust the shear stresses derived from PFLOTRAN's Darcy-scale simulations.

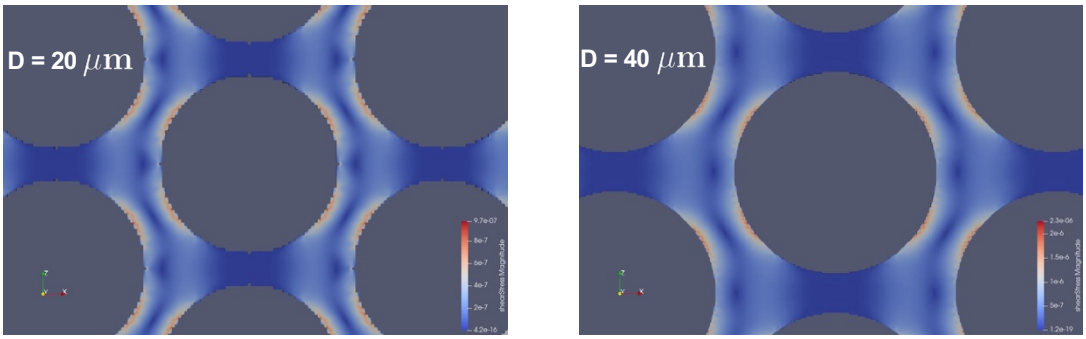


Figure 2: Shear stress fields derived from OpenFoam simulations of a 25% porosity geometry with a grain diameter of 20 μm and 40 μm. The D = 20 μm simulation has a mean shear stress of $1.27e^{-7}$ and a max shear stress of $7.89e^{-7}$. The D = 40 μm simulation has a mean shear stress of $2.71e^{-7}$ and a max shear stress of $1.86e^{-6}$. The ratio of max to mean shear stress increases as grain diameter increases (6.21 to 6.86), indicating that a pore scale correction factor of 10 is reasonable for our simulations, where average grain diameters may theoretically range from less than 1 μm (clay) to 100 mm (gravel).

In our shear stress calculations from PFLOTRAN, we opted to use the maximum shear stress rather than the average. This decision is justified by the sparse spatial distribution of significant shear stresses, which primarily occur at interfaces between units of differing permeability where large velocity gradients exist (Fig. 3). While our Gaussian permeability fields contain few sharp discontinuities, real soil systems typically exhibit high heterogeneity at the pore scale, suggesting that our simulated shear stress fields are likely more uniform than the stress fields in real soil systems. Therefore, we posit that the maximum shear stress serves as a reasonable proxy for the true mean shear stress in more heterogeneous systems.

To validate our approach, we created three baseline simulations and calculated their shear stresses using the aforementioned methodology. We then correlated these values with those presented in Figure 5 of [17], computing the percentage difference in biomass concentrations for each increment in shear stress (Fig. 4). These percentage differences were compared to the corresponding changes in biomass thickness reported in the literature, allowing for a final

calibration. This comprehensive calibration process yielded optimal values of $\beta = 0.8$ and $\alpha = 2 \times 10^{-7}$. This rigorous approach to parameterization, incorporating both theoretical considerations and empirical data, enhances the robustness and applicability of our biomass decay model across various hydrodynamic conditions in the hyporheic zone.

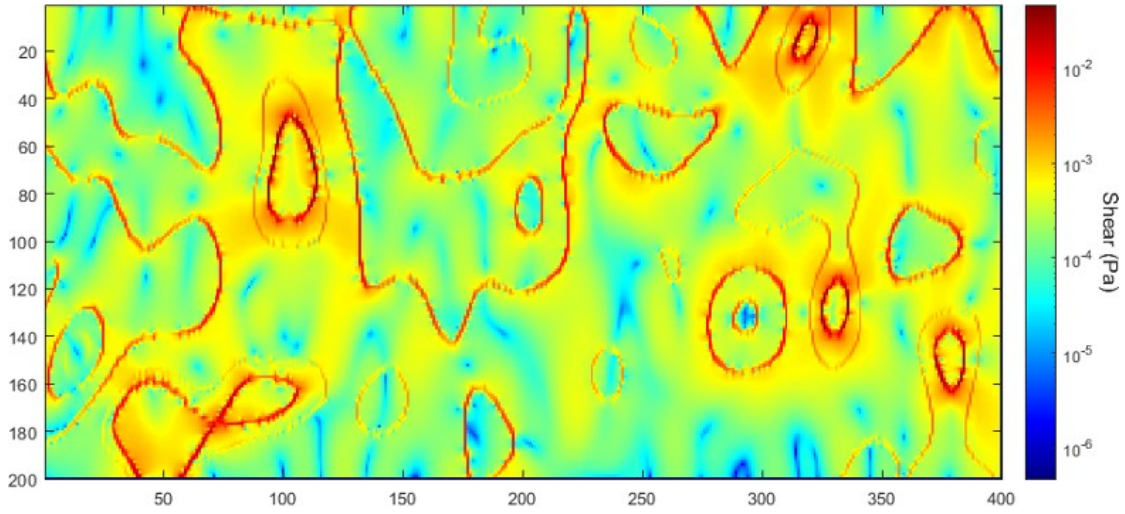


Figure 3. Average shear stress field for the high-speed calibration simulation. High shear values ($5e^{-3}$ to $5.4e^{-2}$), indicated by red color, occur due to sharp transitions in permeability and high velocity magnitude. The top right section of the domain contains preferential channels with high local velocities. When these high velocities encounter low-permeability transitions, their directions change and velocities significantly reduce, resulting in large velocity gradients (and thereby shear magnitudes). Areas of the spatial domain with lower velocity or less transitions between high and low permeability, which comprise the majority of the total area, generally have shear values of $1e^{-5}$ to $6e^{-4}$.

In addition to our augmentation of the biomass decay equation, we also altered the standard Chrotran biomass growth function to be a function of temperature. Numerous studies have shown that microbial growth generally increases with increasing temperature [40-43], and that this dependence may be modeled by the Ratkowski function [44]. The standard implementation of the Ratkowski function was difficult to implement in PFLOTRAN, so we use a sixth degree polynomial that accurately represents the Ratkowski function from 1 to 30 ℃. Specifically, we parameterized the Ratkowski function as

(3) $\lambda_{B_T} = P_1T^6 + P_2T^5 + P_3T^4 + P_4T^3 + P_5T^2 + P_6T + P_7$

where P1 = -2.9×10^{-9}, P2 = 2.3×10^{-7}, P3 = -6.5×10^{-6}, P4 = 6×10^{-5}, P5 = 1.4×10^{-4}, P6 = 2.4×10^{-2}, and P7 = 0.196. Accurate in-situ temperature time series were given for each of the in-situ hyporheic flux time series' that we chose to simulate (gaining, high gaining, losing, high losing, and Hanford), which allowed use to make use of this temperature-based growth equation in the context of realistic temperature fluctuations. Plugging equations (2) and (3) into the original Chrotran biomass growth equation gives

(4) $\mu_b = \lambda_B\lambda_{B_T}B\frac{D}{K_D+D}\left(\frac{K_B}{K_B+B}\right)^{\alpha_c} - \lambda_b$

where μ_b is the biomass growth rate, λ_B is a growth rate parameter, D is the electron donor concentration, K_D is the half-saturation constant for the electron donor, K_B is a Monod constant, and α_c is the crowding parameter.

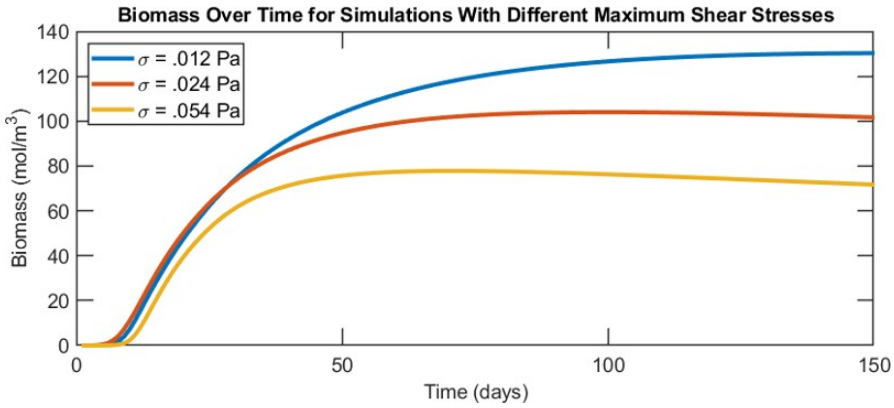


Figure 4. Biomass time series for a low-speed (blue), medium-speed (orange), and high speed (yellow) simulation. The legend gives the maximum shear for each simulation. With the pore-scale to darcy-scale correction of 10, we get max shear values of .054, .024, and .012. From the micro-scale experiments of shear-based biofilm breakup [17], these shear values correspond to biofilm thicknesses (after a 14 hour growth period) of 18, 25 and 31, or a 19% decrease in thickness going from σ = .012 Pa to σ = .024 Pa and and 28% decrease in thickness going from σ = .024 Pa to σ = .054 Pa. Our steady-state biomass concentrations show similar percentage differences (22% and 29%), indicating a relatively accurate calibration of our speed-based biomass decay.

5.3.1.3 Permeability and Flow

Studies have shown that biomass growth in the hyporheic zone is strongly linked to the hyporheic flux [9, 7, 45, 46]. Depending on the concentration of nutrients and the flow speed of the groundwater and surface water flows, a positive or negative flux can have different impacts on growth. To gain deeper insight into how exactly these differences affect biomass growth, we simulated the hyporheic zone under a variety of realistic flow conditions (Fig. 5a).

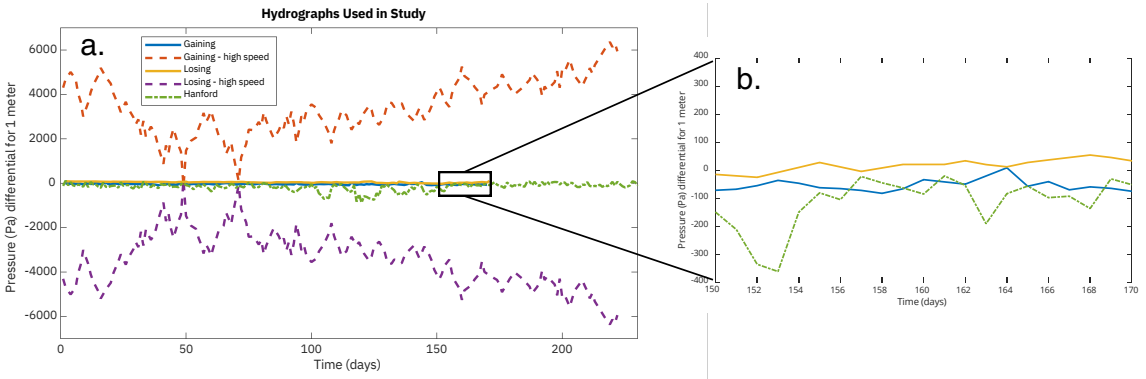


Figure 5: Time-series of pressure boundary condition groups used in the simulations. For each group, slight variations to the time-series were introduced to develop a much wider variety of potential flow conditions for the simulations.

The gaining and losing flow conditions at high speed represent the largest hyporheic fluxes we were able to find in the literature [35], and the low-speed gaining and losing conditions represent much smaller fluxes. In addition, we use hydrographs that come from hyporheic flux data measured at the Hanford site [34]. Thus, we examine feedback cycles and statistical relationships in the hyporheic zone from a general perspective that can be considered the average of a variety of previously published data. Furthermore, although it is unclear how much each base time series hydrograph influences our overall analysis, the results we present are indirectly relevant to chromium bioremediation simulations of the Hanford site.

In addition to the general direction and magnitude of flow, permeability has a significant influence on the general transport of nutrients and biomass within the hyporheic zone. At low

permeabilities, biomass and nutrients are less able to disperse throughout the hyporheic zone, so average biomass concentrations over a large area will be less. However, hotspots of high biomass concentration may still form in low permeability zones, which may also result in bioclogging. To understand how exactly biomass growth is impacted by a variety of permeabilities, we created a variety of heterogeneous permeability fields with different covariance ratios (Fig. 6). Although the mean permeability was similar for most simulations (around $2e^{-10}$ m^2), the effective permeability (K_{eff}) would change based on the covariance ratio. Furthermore, we also ran simulations at extremely low permeability to understand biomass growth in a significantly different environment.

5.3.1.4 Boundary Conditions

We model a segment of the hyporheic zone as a vertical slice measuring 1 meter in height and 2 meters in length. This slice represents the interface between groundwater and surface water, with the top and bottom boundaries corresponding to surface flow and groundwater respectively. The left and right boundaries extend the hyporheic zone longitudinally along the direction of river or

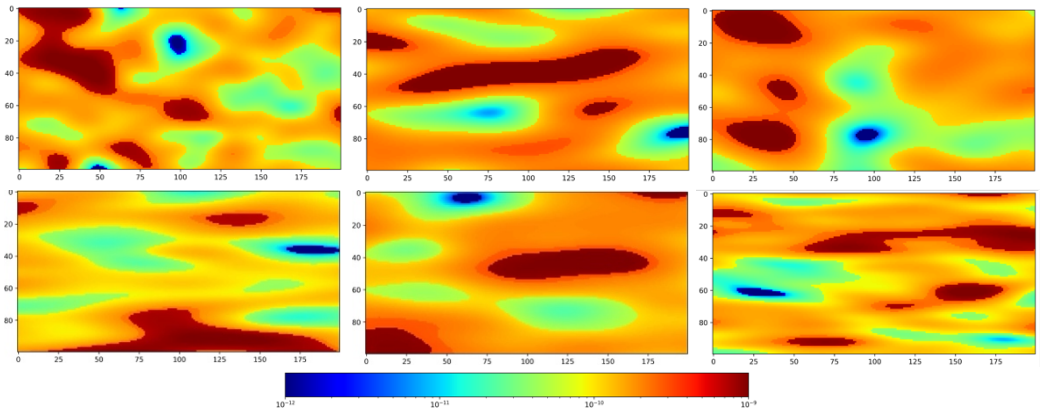


Figure 6. Sample heterogeneous permeability distributions (at t=0) used in simulations featured in this study. This figure is also featured in some of our unpublished work on upscaling the simulations discussed in this study.

groundwater flow. Flow conditions at the top and bottom are controlled by Dirichlet boundaries, with pressure differentials set according to the simulation type: gaining, losing, or Hanford. The lateral boundaries allow for bidirectional flow, regulated by a Dirichlet pressure boundary of zero on one side and a constant value on the other side. Transport boundary conditions mirror those used for flow. The transport boundary conditions for each chemical species were varied over our multitude of simulations, but the groundwater (transport from bottom into the domain) generally contained higher concentrations of nutrients and chromium than the surface water.

A variety of scales were used for the simulations in our study. For our analysis of feedback cycles and general relationships observed in our simulations, we created simulations that were 1x2 meters, 1x4 meters, and 1x20 meters that all had a dx and dy of 0.01 meters. For our sensitivity analysis of the impacts of our velocity-based biomass decay on biomass growth and chromium reduction, we created simulations that were 1x2 meters and had a dx and dy of 0.005 meters. Scale is not a main point of the analysis presented in this chapter, and the simulations of multiple scales were primarily generated for the purpose of training the deep-learning-based upscaling model presented in chapter 6 of this dissertation. However, our inclusion of simulations at multiple scales in this chapter does act to add to the overall variability between simulations, resulting in a more general and scale-agnostic analysis of biomass growth in the hyporheic zone.

5.3.1.5 Simulation Variables

The variables (also referred to in this work as features) of the simulations, as well as their range of possible values, are given in Table 1. All of the features that depend on the time-evolution of the simulation, such as biomass, chemistry, and flow speed, are referred to as the “physio-chemical

Table 1. Description of physio-chemical features and their ranges of possible values used in the simulations. B: biomass, ED: electron donor (molasses), Cr(VI): chromium, V_y: horizontal velocity, V_x: vertical velocity, P: pressure, T: temperature, ϕ: porosity, k: permeability.

Var	B (mol/m^3)	ED (mol/L)	Cr(VI) (mol/L)	V_y (m/hr)	V_x (m/hr)	P (Pa)	T (°C)	ϕ	K (m^2)
Min	1e-10	1e-20	1e-20	-632	-486	-1214	4.8	1e-4	1e-15
Max	768	5.5e-3	7.6e-3	671	651	7099	25	0.6	1e-9
Mean	58	8.1e-6	1.4e-5	-5.8e-2	-1.4e-2	786	11.5	0.13	2e-10

features." The variables that are prescribed at the beginning of the simulation and don't change value over time are referred to as "input variables".

5.3.2 Sensitivity Analysis

One of the primary goals of this study is to gain deeper insight on the abiotic controls of biomass growth in the hyporheic zone. To this extent, we employ a variety of sensitivity and correlation analyses to understand how different features may impact growth. We use classical sensitivity analysis, changing one feature and keeping all others equal, to determine the individual impacts that each feature may have on biomass growth. Furthermore, we use a Monte-Carlo-type sensitivity analysis to understand feature relationships at a more global level. Specifically, we ran 344 simulations of the hyporheic zone, each under slightly different input conditions, then used PCA and cluster analysis to understand physio-chemical feature relationships and groupings. In addition, we used a correlation heatmap to identify correlations between all the simulation features (both physio-chemical features and input variables).

5.4 Results

5.4.1 Feedbacks and Mechanisms of Biomass Growth

The primary known physio-chemical features that impact biomass growth at scale, as well as their

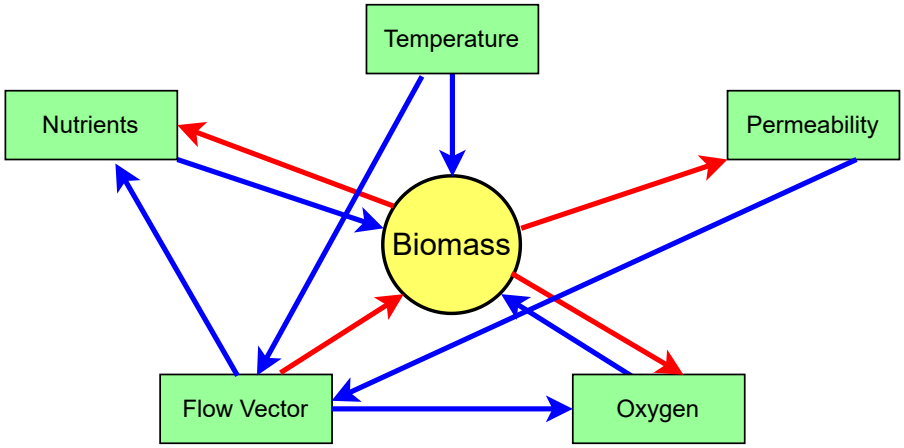


Figure 7. Known feedback cycles from physical features present in simulations of biomass growth in the hyporheic zone. A blue arrow signifies a positive causative effect (increase leads to increase), and a red arrow indicates a negative causative effect (increase leads to decrease).

general relationships with each other, are shown in Figure 7. While modeling of oxygen concentrations was beyond the scope of this study, we included it as a primary variable in the biomass feedback loop because of the vast literature detailing the impacts of oxygen depletion on bacteria [45, 47, 48]. These interactions were also discussed in the introduction, but essentially, changes in flow cause changes in nutrient concentrations and temperature which then cause changes in biomass growth. As biomass grows, it consumes nutrients and oxygen, and reduces the local permeability of its substrate. This decrease in permeability in turn causes a decrease in local fluid speed. The only relationship shown in this feedback loop that hasn't been widely reported on is that an increase in the flow vector may cause a decrease in biomass. Specifically, an increase in flow speed, which thereby leads to an increase in pore-scale shear forces, can result in the breakup and dispersal, and/or sloughing, of biofilms. This feedback loop represents some of the primary intuitive relationships for biomass growth, but the relative importance of each relationship depends on a number of other factors specific to the hydro-biogeochemical settings of the domain in which the biomass growth occurs. In our own simulations, which are further

discussed in greater detail throughout the rest of the chapter, we find many of the intuitive relationships presented in Figure 7 do not show high correlation values, further showing the specificity of biomass feedback loops in relation to the characteristics of the domain. The results presented in this chapter thus intend to refine the assumptions of the standard biomass feedback loop.

5.4.2 Spatial Trends

The average spatial distributions of the physio-chemical features for one of our losing simulations (Fig. 8) can be used to elucidate some of the relationships in our biomass growth feedback loop. Biomass is greatest in the high-porosity zones around the edges of the domain, and concentrations are smallest in the low-porosity zone in the middle of the domain. Where fluid speeds are largest, we see the greatest amount of lateral extension of biomass towards the middle of the domain. In addition, a small amount of velocity-based biomass decay can be observed in the lower right corner of the domain for the biomass distribution. The spatial distribution of molasses looks similar to that of biomass except the gradients are smoother. The temperature seems to be primarily determined by the horizontal flow (V_y). The spatial distribution of Cr(VI) seems to be primarily determined by molasses and V_y. Because pressure is lowest at the bottom boundary and highest at the right boundary, this results in a top-right to bottom-left pressure gradient. However, since the entire left boundary of the domain has higher pressure than the bottom boundary, there is also a top-left to bottom-right pressure gradient. Thus, fluid and nutrient flux enters in the top, right and left sides of the domain and primarily exits the domain at the middle of the bottom boundary. Thus, the high Cr(VI) concentrations on the left are a result of inward flux from the left side of the domain, and the low Cr(VI) concentrations are a result of both inward flux from the

right side of the domain and high biomass/molasses concentrations. The impacts of biomass growth on Cr(VI) reduction can be seen in the slightly elevated concentrations of Cr(VI) near the right boundary where biomass concentrations are high. Although molasses concentrations are also high here, most of it is immediately consumed by biomass and won't be available for abiotic reduction. Thus, where biomass concentrations are lower, even though molasses is also lower, we see greater reduction in chromium concentrations. In other words, the scale of the variability of biomass concentrations (x=400 to x=350 is a 200 times decrease in concentration) is greater than that of molasses concentrations (x=400 to x=350 is a 7 times decrease). Thus, at least for this simulation, abiotic reduction is more efficient that biotic reduction (i.e., 1 mol of molasses spent on abiotic reduction results in more total reduction than 1 mol of molasses spent on biomass growth).

In addition to the temporally averaged spatial distributions of key physio-chemical features for a single simulation, we present the average normalized spatial distributions of the physio-chemical features over all of our simulations (Fig. 9). This figure represents the general spatial trends that persist after averaging over each 1x2 meter simulation. The averaged spatial distributions of biomass, molasses (labeled as ED in Figure 9) and Cr(VI) all have a significant degree of similarity due to the boundary conditions of the simulations. Although the concentrations of these features were randomized in our simulations, the predominant direction of flow over the entirety of our simulations is from top (surface water) to bottom (groundwater). However, on average the surface water contains decreased levels of ED relative to the groundwater (not realistic, but a facet of the random variations of simulations), meaning biomass growth is generally greater in the gaining simulations. Thus, as groundwater flows into the domain from the bottom, and because nutrient and biomass concentrations had low initial values and the

transport conditions are greater than those at the top boundary, each of these variables shows high values near the bottom of the domain. The high concentrations of biomass on the right side of the domain are a result of the fact that most simulations featured lateral flow moving from the right to left side of the domain. Porosity, although impactful at the level of a single simulation, seems to have no impact on biomass or ED concentrations when averaged over all simulations. However, porosity does have a large impact on immobile molasses concentrations, with higher amounts of ED_{im} appearing where porosity is high. Similar to the average spatial distributions for a single simulation, the average spatial distributions over all simulations show the impacts of fluid speed on biomass decay as a slightly lower concentration of biomass in the lower right corner.

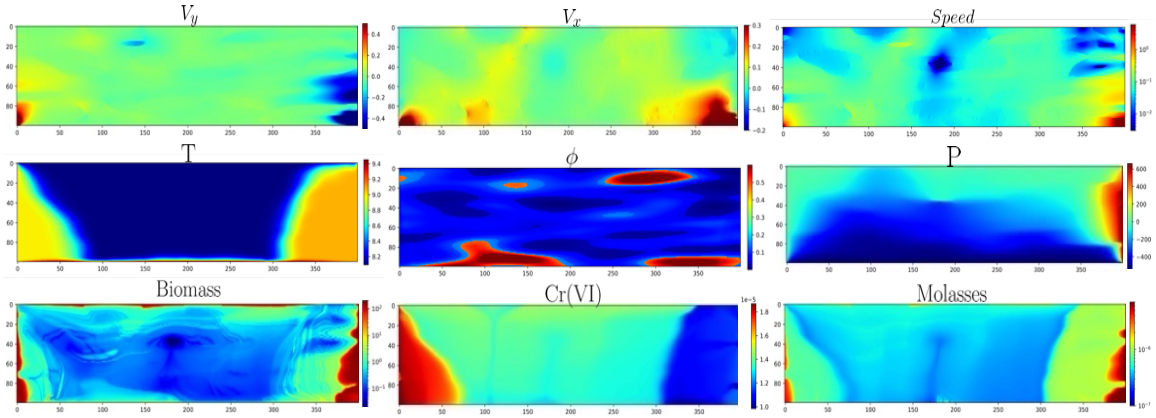


Figure 8. General (averaged over all simulations and time steps) trends for key physio- chemical features. Each variable is normalized, with 0 (dark blue) corresponding to low values and 1 (dark red) corresponding to high values.

These results highlight the complex interplay between physical, chemical, and biological processes in the HZ, demonstrating the importance of considering spatial heterogeneity and temporal dynamics in modeling efforts. The observed dominance of abiotic reduction over biotic reduction, even in areas of high biomass concentration, provides a nuanced view to the relationship between biotic and abiotic reduction and has significant implications for bioremediation strategies.

5.4.3 Temporal Trends

The temporal trends (Fig. 10) of the physiochemical features reveal a variety of insights related to the feedback mechanisms involved in these simulations. The red dotted lines in each time series plot show the time value of the inflection points for biomass growth (Fig. 10e). Average biomass concentrations increase very slowly for the first 18 days, increase rapidly for the next 36 days, then increase at a slightly lower growth rate for the rest of the simulation (Figs. 10d and 10e). The first inflection point (at t = 18 days) corresponds to the time when biomass growth starts to

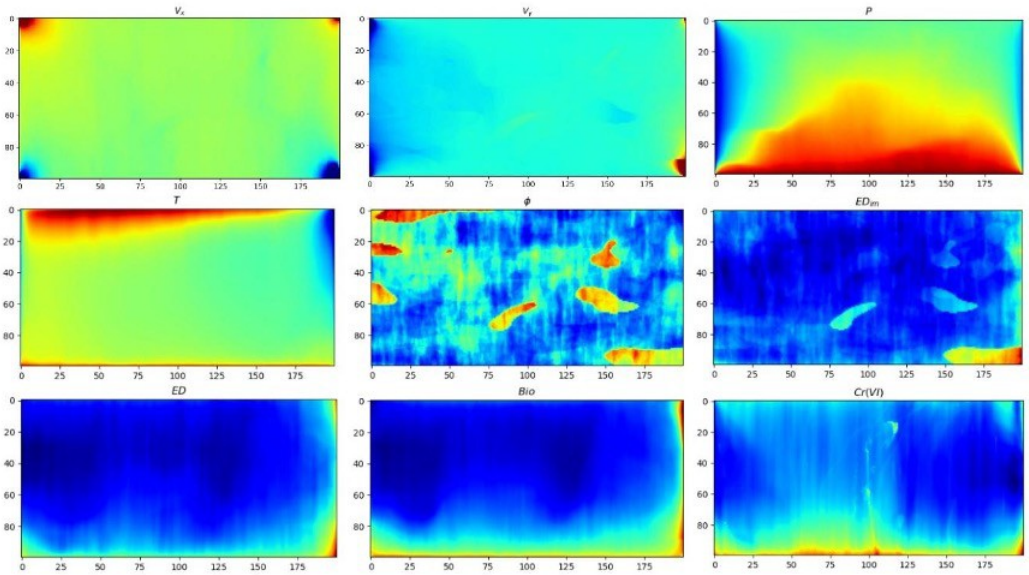


Figure 9. General (averaged over all simulations and time steps) trends for key physio- chemical features. Each variable is normalized, with 0 (dark blue) corresponding to low values and 1 (dark red) corresponding to high values.

dramatically increase. This also represents the point that ED and ED-immobile start to significantly decrease, Cr(VI) starts to increase, and porosity starts to decrease. We find that the relationship between biomass concentrations (in mol/m^3) and porosity can be described via the equation

(5) $\phi = .141 - 10^{-4} C_{bio}$

where C_{bio} is the biomass concentration. This equation may be useful for predicting general trends in porosity as a function of biomass concentrations, although it is important to note that this

function only represents our average simulation, and the exact value of this function will depend on a variety of factors such as soil type, microbial community structure, flow rates, and nutrient concentrations. At the highest amounts of growth (around day 46- 52), biomass growth has a clear correspondence with pressure and V_x (Figs. 10b and 10a). The increase in pressure corresponds to a large decrease in the magnitude of V_x and a small increase in the magnitude of V_y. These changes in flow then cause an increase in molasses, chromium and the rate of biomass biomass growth. Around t = 52, the pressure starts to drop again and the magnitude of V_x increases, resulting in decreases in biomass growth, Cr(VI) and ED. At t = 60, the molasses concentrations and biomass growth rate start to increase again, but the chromium concentration continues to decrease. The vertical flow (V_x) thus seems to have a large effect due to the transport input conditions at the top boundary. Specifically, greater negative vertical velocities, representing flow from the top of the domain to the bottom, seem to result in a increase in molasses concentrations, which thus causes an increase in biomass growth. Although biomass growth and molasses concentrations are enhanced due to increased nutrient flux from the top boundary, the amount of Cr(VI) is also increased from the greater top-boundary flow, resulting in a small increase in Cr(VI) concentrations from day 42 to day 52. At t = 70, we approach the steady-state trends of the simulations. From this point onward, the magnitude of V_y slowly decreases, V_x and pressure slowly increase, biomass growth and Cr(VI) slowly decrease, and molasses slowly increases. At this point, biomass concentrations are relatively high, meaning the increase in V_x and pressure cause a corresponding decrease in growth due to velocity-based biomass decay. While at t = 50 the spike in V_x causes an increase in the biomass growth rate, by t = 70 it starts to cause a decrease in the growth rate. The greater negative values of V_x also should result in increased concentrations of Cr(VI) considering the transport boundary conditions. However, because ED concentrations

steadily increase due to the decrease in biomass growth, the increased amount of abiotic reduction is enough to cause a decrease in Cr(VI). This indicates some dominance of abiotic reduction over biotic reduction, although it is not possible to completely parse each impact from these general time series. For a more robust analysis of the relative strengths of biotic and abiotic reduction, we provide a sensitivity analysis of the molasses consumption stoichiometry coefficient $D1$, discussed in the following section.

These findings reveal the intricate temporal dynamics of biomass growth, nutrient cycling, and chromium reduction in the HZ. The identification of key inflection points and the quantification of relationships between variables, such as biomass concentration and porosity, provide valuable insights for predicting system behavior and optimizing remediation efforts.

5.4.4 Drivers of Biomass Growth and Chromium Reduction in the Hyporheic Zone

5.4.4.1 Sensitivity Analysis

To gain deeper insights into the biotic and abiotic determinants of biomass growth in the hyporheic zone, we used sensitivity and correlation analysis. The sensitivity analysis (Fig. 11) shows the biomass time series and spatial distributions of the last time step for 6 equally spaced values of a simulation input variable (keeping all other input variables constant). We performed sensitivity analysis for five key variables (temperature, carbon reuse efficiency/molasses consumption stoichiometry coefficient, homogeneous permeability value, vertical velocity (V_x), and the biomass crowding parameter). The temperature sensitivity analysis (Fig. 11a) shows that at higher temperatures the biomass growth rate increases, leading to differences in biomass concentrations that remain constant after about day 60. The spatial distributions show more variation in biomass concentrations for higher-temperature simulations. The carbon reuse efficiency ($D1$), is a Chrotran parameter that defines the stoichiometric relationship between the

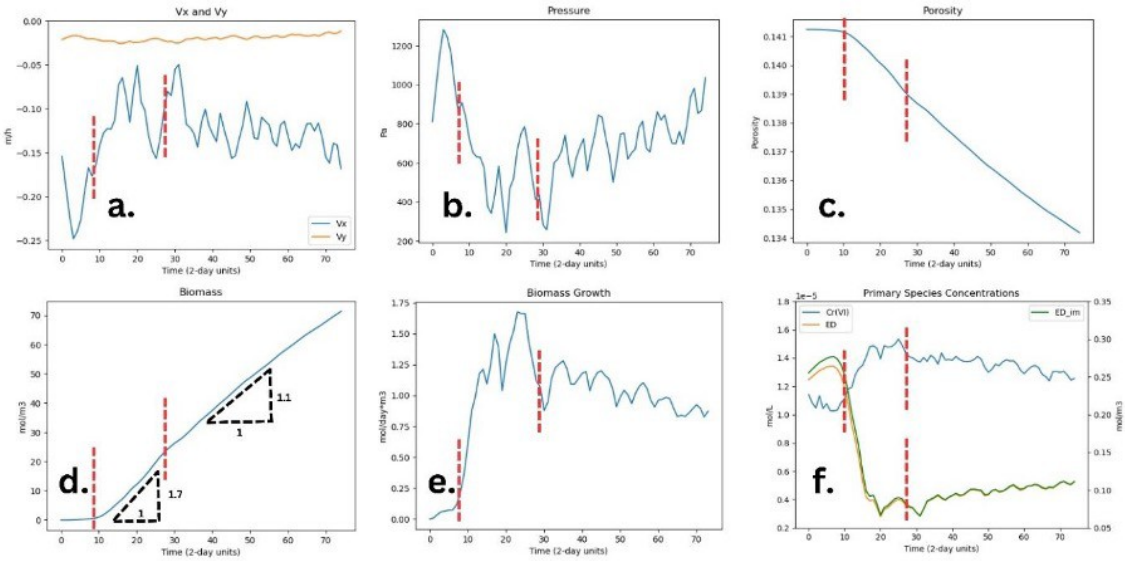
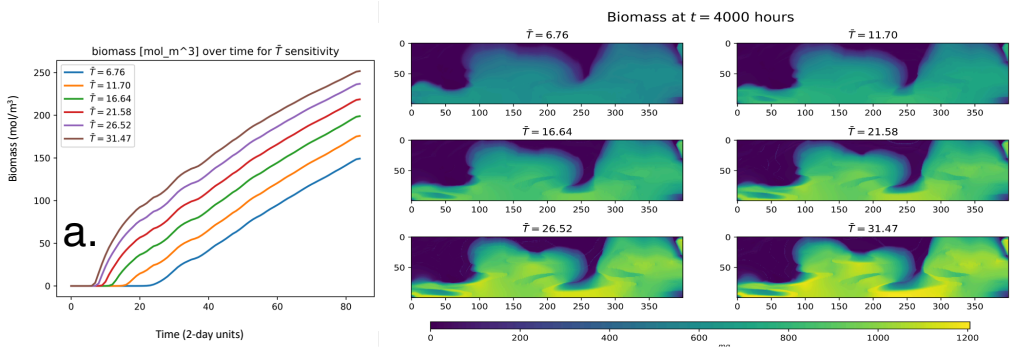


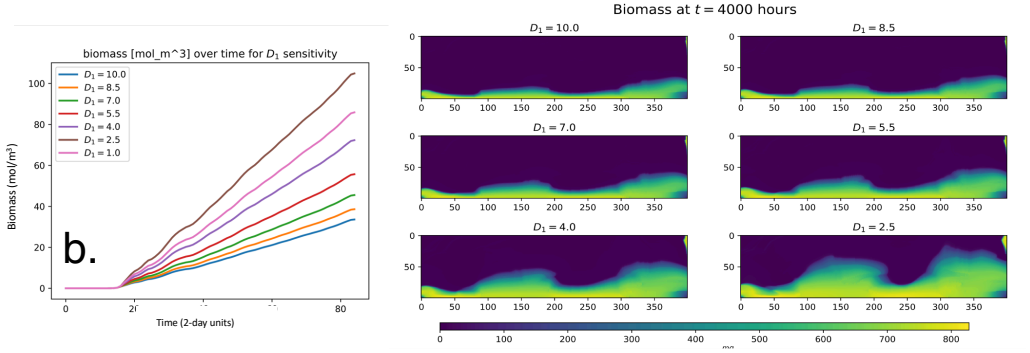
Figure 10. General (averaged over all simulations and spatial dimensions) temporal trends for key physio-chemical features. The red dotted lines show the approximate time of the biomass growth inflection points. (a) V_x (blue) and V_y (orange). (b) Pressure. (c) Porosity. (d) Biomass. (e) Biomass growth. (f) Cr(VI) (blue), ED (orange), and ED-immobile (green). Here, biomass growth is shown to primarily be dependent on ED concentration. The Cr(VI) timeseries shows an increase in concentration as molasses decreases, indicating the dominance of abiotic reduction over biotic reduction.

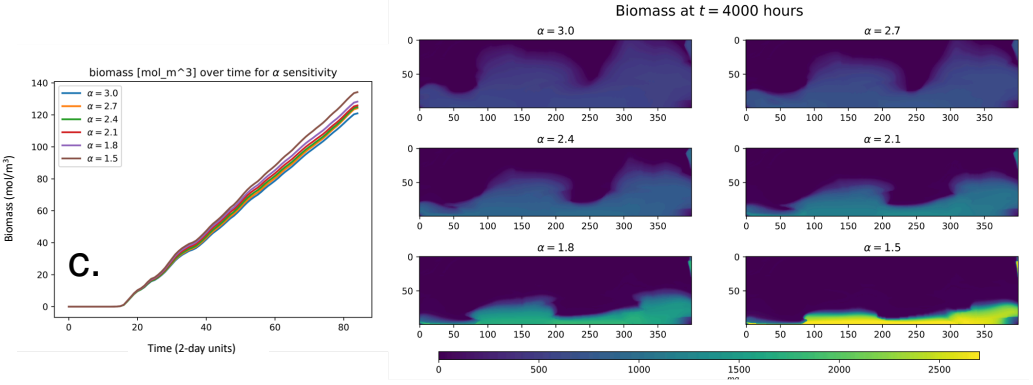
ED and biomass. So for $D1 = 1$, one mol of ED creates one mol of biomass. The sensitivity analysis for $D1$ (Fig. 11b) shows that lower values of $D1$ result in an increase in biomass growth that increases over time. The spatial distributions show that for lower values of $D1$, the biomass spreads further throughout the domain. The permeability sensitivity analysis (Fig. 11c) indicates that mean permeability of the domain has a huge impact on biomass growth. For $k = 1 \times 10^{-12}$ there is a minute amount of biomass growth that can be observed in the bottom corners of the spatial distribution, but for lower permeability values, we don't observe any biomass growth. A permeability of 1×10^{-12} m^2 is representative of silty sand or permeable basalt, implying that aquifers primarily composed of these materials (or lower-permeability materials) are not likely to house large concentrations of biomass. The V_x sensitivity analysis (Fig. 11d) indicates that greater vertical velocity contributes to a more sigmoidal (as opposed to linear or exponential) growth

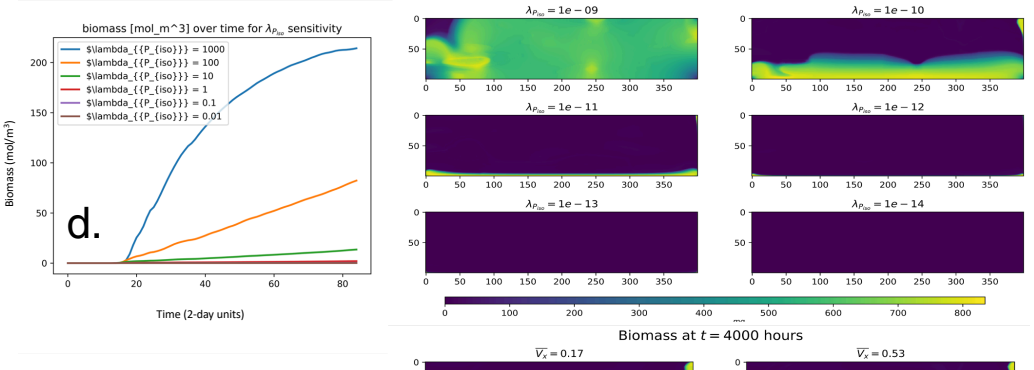
curve, results in generally greater biomass growth, and causes the growth curve to have small undulations. This wave-like behavior is a result of spikes in flow speed causing significant shearing of biomass, thus briefly decreasing the rate of biomass growth. α is another Chrotran parameter that describes biomass crowding. For higher values of alpha, we see slightly lower biomass concentrations over time, although the main difference is in the spatial distributions (Fig. 11e). When crowding is high (low α), we get much higher concentrations of biomass that are constrained to the first few centimeters of the domain. When crowding is low, we get a lower maximum biomass concentration, but the biomass is spread throughout the entire domain. Overall, the sensitivity analysis shows that many input features result in large changes to biomass concentrations, illustrating the complexity of determining the most important general impacts on biomass. The analysis also identifies mean permeability as the most critical factor in biomass growth (compared to the other variables in the sensitivity analysis).

In addition to the sensitivity analysis of biomass, we also provide a sensitivity analysis of chromium reduction (Fig. 12). Specifically, we provide the average spatial distributions (Fig. 12a-12d) and time series (Fig. 12e) for simulations with: no biomass (Fig. 12d), biomass with low molasses consumption stoichiometry (Fig. 12a), medium consumption stoichiometry (Fig. 12b), and high consumption stoichiometry (Fig. 12c). The molasses consumption stoichiometry, given here as $D1$, gives the number of mols of molasses require to produce 1 mol of biomass. Thus, for $D1 = 0.1$, 1 mol of molasses produces 10 mols of biomass, and for $D1 = 10$, 1 mol of molasses produces 0.1 mols of biomass. Although it is unlikely for 1 mol of molasses to produce 10 mols of biomass (since the molar mass of a microbial cell is on the order of 10^9, about 7 magnitudes larger than the molar mass of molasses), the $D1$ parameter can also be thought of as carbon reuse efficiency. When microbes die, they release bioavailable carbon that may be used by other


biomass [mol_m^3] over time for $\bar{T}$ sensitivity
$\bar{T} = 6.76$
$\bar{T} = 11.70$
$\bar{T} = 16.64$
$\bar{T} = 21.58$
$\bar{T} = 26.52$
$\bar{T} = 31.47$
Biomass (mol/m³)
Time (2-day units)
a.
Biomass at $t = 4000$ hours
$\bar{T} = 6.76$
$\bar{T} = 11.70$
$\bar{T} = 16.64$
$\bar{T} = 21.58$
$\bar{T} = 26.52$
$\bar{T} = 31.47$


biomass [mol_m^3] over time for D_1 sensitivity
$D_1 = 10.0$
$D_1 = 8.5$
$D_1 = 7.0$
$D_1 = 5.5$
$D_1 = 4.0$
$D_1 = 2.5$
$D_1 = 1.0$
Biomass (mol/m³)
Time (2-day units)
b.
Biomass at $t = 4000$ hours
$D_1 = 10.0$
$D_1 = 8.5$
$D_1 = 7.0$
$D_1 = 5.5$
$D_1 = 4.0$
$D_1 = 2.5$


biomass [mol_m^3] over time for α sensitivity
$\alpha = 3.0$
$\alpha = 2.7$
$\alpha = 2.4$
$\alpha = 2.1$
$\alpha = 1.8$
$\alpha = 1.5$
Biomass (mol/m³)
Time (2-day units)
c.
Biomass at $t = 4000$ hours
$\alpha = 3.0$
$\alpha = 2.7$
$\alpha = 2.4$
$\alpha = 2.1$
$\alpha = 1.8$
$\alpha = 1.5$


biomass [mol_m^3] over time for $\lambda_{P_{iso}}$ sensitivity
$\lambda_{P_{iso}} = 1e-09$
$\lambda_{P_{iso}} = 1e-10$
$\lambda_{P_{iso}} = 1e-11$
$\lambda_{P_{iso}} = 1e-12$
$\lambda_{P_{iso}} = 1e-13$
$\lambda_{P_{iso}} = 1e-14$
Biomass (mol/m³)
Time (2-day units)
d.
Biomass at $t = 4000$ hours

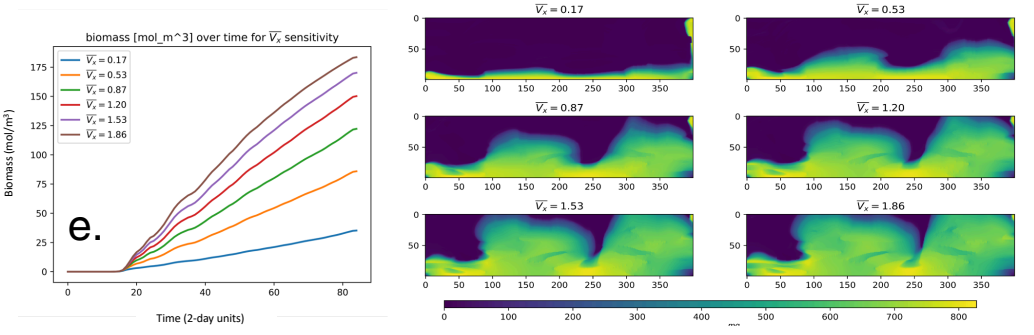

biomass [mol_m^3] over time for $\overline{V_x}$ sensitivity
$\overline{V_x} = 0.17$
$\overline{V_x} = 0.53$
$\overline{V_x} = 0.87$
$\overline{V_x} = 1.20$
$\overline{V_x} = 1.53$
$\overline{V_x} = 1.86$
Biomass (mol/m³)
Time (2-day units)
e.
Biomass at $t = 4000$ hours
$\overline{V_x} = 0.17$
$\overline{V_x} = 0.53$
$\overline{V_x} = 0.87$
$\overline{V_x} = 1.20$
$\overline{V_x} = 1.53$
$\overline{V_x} = 1.86$

Figure 11. Sensitivity analysis for selected features with large impacts on biomass growth. For each feature, we show the time series, as well as the spatial distributions of biomass for the final time step. The colorbars show biomass concentration in mol/m^3. (a) Temperature sensitivity. (b) Carbon reuse efficiency ($D1$) sensitivity. (c) Biomass permeability sensitivity. (d) Mean V_x (vertical flow speed) sensitivity. (e) Biomass crowding coefficient (α) sensitivity. Biomass concentrations for the temporal sensitivity plots are given in mol/m^3, and concentrations for the spatial sensitivity plots are given in mg/kg.

microbes to grow [50]. In the situation with no biomass growth (Fig. 12), chromium reduction is entirely dominated by abiotic reduction. Over time, the amount of incoming chromium at the domain boundaries exceeds the amount of incoming molasses, meaning we see fast reduction of chromium for the first 50 days, but once the molasses is consumed and starts to run out, the chromium concentration increases (Fig. 12e). The situation with $D1 = 0.1$ unsurprisingly shows a large amount of chromium reduction (Fig. 12a), since the small amount of molasses consumption results in significant biotic and abiotic reduction. For $D1 = 1$, the molasses consumption is greater, resulting in a smaller amount of biotic and abiotic reduction compared to the $D1 = 0.1$. Compared to the no biomass growth situation, there is less chromium reduction throughout most of the spatial domain, but there are biomass-induced hotspots of reduction near the bottom of the domain (Fig. 12b). The time series for the $D1 = 1$ case shows that chromium reduction starts off slower than the case with no biomass growth, but towards the end of the simulation, the $D1 = 1$ case has a negative slope, whereas the no biomass growth case has a positive slope. This can be explained by the fact that abiotic reduction is largely responsible for reduction at early times, but becomes less of an important factor over time as the electron donor concentration decreases. At later simulation times (t > 60 days), the biotic reduction starts to make a significant impact on chromium concentrations, resulting in similar end-simulation concentrations for the no biomass growth case and the $D1 = 1$ case. For the $D1 = 10$ case, all the molasses is quickly consumed by chromium and biomass, but not enough biomass grows to cause significant biotic reduction, resulting a small amount of initial reduction, then no reduction,

with a flat curve reflecting the chromium concentration for the constant transport boundary condition of these simulations.

5.4.4.2 Statistical Analysis

To further investigate the impacts of our simulation variables on biomass concentrations in a more general sense, we used PCA and cluster analysis (Fig. 13a) to identify groupings and large-scale relationships across all of our simulations. The first principal component, which explains 25% of the variance of the dataset, seems to be primarily determined by biomass, nutrients, chromium, molasses consumption stoichiometry coefficient (previously represented by $D1$, but in Figure 13 is represented by λ_D), and λ_C, which controls the bio-reduction reaction rate. Thus, the first principal component can be summarized as representing the primary factors impacting chromium remediation and biomass growth. The second principal component, which accounts for 17% of the total variation in the data, is primarily determined by flow vectors, pressure, permeability covariance, λ_C, α (the biomass crowding coefficient), porosity, and temperature. Generally, these variables represent the various boundary conditions of the simulations.

We also see interesting sets of groups that form as a result of cluster analysis in the space of the first two principal components. The right cluster can be described as the biomass and nutrient cluster, the top is the flow and porosity cluster, and the left is the Cr(VI) cluster. These clusters also represent most of the strongest correlations present in our simulations. Loosely, the features within a cluster are likely to be positively correlated with each other, the right and left clusters are likely to be negatively correlated with each other, and the top cluster has both positive and negative correlations with the features of the left and right clusters.

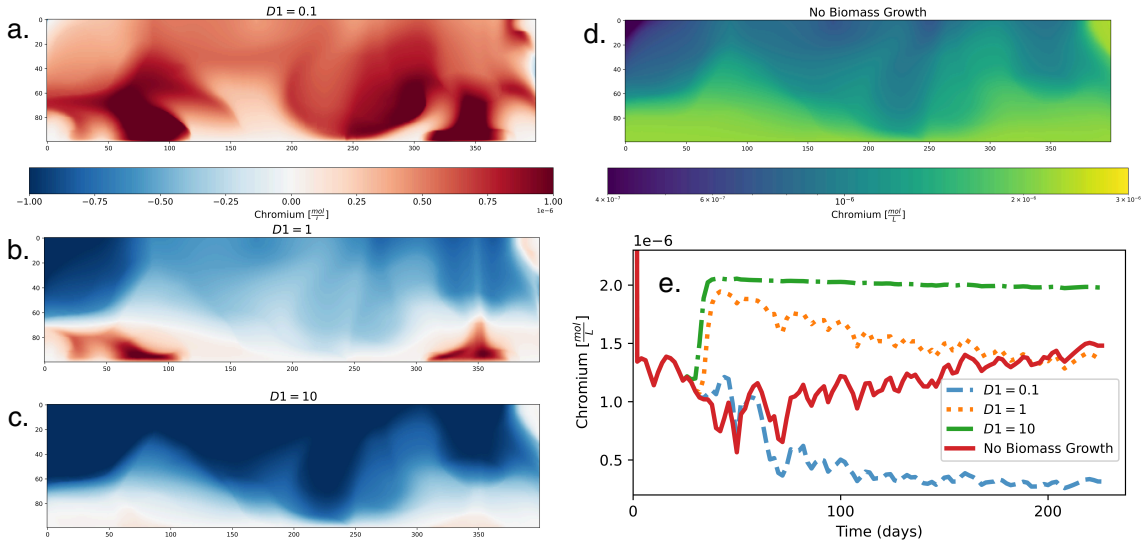


Figure 12. Sensitivity analysis for molasses consumption stoichiometry $D1$. (a) Spatial chromium concentration difference distribution ($C_{no-growth} - C_{D1}$) for $D1 = 0.1$, which means 1 mol of molasses produces 10 mols of biomass. The concentration for (a) is mostly positive because chromium remediation is greater in the case of a low stoichiometric coefficient compared to the situation with no biomass growth. (b) Spatial difference distribution for $D1 = 1$. (c) Spatial difference for $D1 = 10$. (d) Average spatial concentrations of chromium in the case of no biomass growth. (e) The colorbars show chromium concentration (or difference in chromium concentration for the plots on the left) in mol/L.

These correlations are further explored in the correlation matrix heatmap (Fig. 13b). The correlations indicated in the PCA (Fig. 13a) differ from those shown in the correlation matrix heatmap (Fig. 13b) for three key reasons. First, while PCA captures the directions of maximum variance in the data, these principal components are not direct measures of correlation between variables - rather, they represent linear combinations of variables that explain the most variance in the dataset. Second, our PCA analysis was restricted to physio-chemical features, while the correlation matrix encompasses a broader set of parameters, including a variety of biomass growth parameters from Chrotran (α, λ_B, λ_C, and λ_D), permeability covariance ratio (k_{CV}), and time. Finally, for the PCA we use the time-and-space-averaged features of each simulation as input (since PCA requires a 2D input), but for the correlation matrix, since we also include time, we use the space-averaged features for each time point for each simulation as input.

For biomass, the strongest correlations (as determined by magnitude in the correlation matrix) with other variables determined over all 344 simulations are σ_{bio} (0.58), time (0.50), σ_P (0.35), λ_D (-0.30), pressure (0.30), ED (0.28), ED_{im} (0.28), porosity (-0.24), λ_C (0.22), and finally chromium (-0.18). Surprisingly, biomass is more strongly correlated with the standard deviation of pressure than the pressure itself (except for chromium). The reason for this is likely because the standard deviations of features are more strongly correlated with the simulation boundary conditions than the features are. These correlations also indicate that biomass concentrations are highly dependent on time and nutrient concentrations, which are both normal assumptions made about biomass growth. Furthermore, the mathematical dependence of biomass growth on time and an electron donor is explicitly defined through Chrotran, so it is of no surprise that these variables are strongly positively correlated with biomass in our simulations. Perhaps more surprising, the augmentations we have introduced do not seem to result in significant correlations. Biomass is positively correlated with temperature with a value of 0.04, meaning that the impact of incorporating the Ratkowski function into the biomass growth equation is negligible. However, this may be a result of the temperature boundary conditions of the simulations since the sensitivity analysis of biomass growth shows a clear positive correlation between temperature and biomass concentration (Fig. 11a). The pressure, and the standard deviation of pressure (which loosely represent the pressure gradient), are the primary driving force of the flux of fluid and solutes in the domain, meaning it should have significant correlation with biomass with or without the speed-based biomass decay augmentation. V_x shows a small negative correlation with biomass, indicating that an increase in V_x leads to a decrease in biomass concentrations. Thus, there is some indication that the speed-based biomass decay influences the general

spatiotemporal distribution of biomass in the hyporheic zone, although by a relatively small amount.

The correlation between biomass and porosity is also interesting. One intuitive way to think about this relationship is that they should have a positive correlation between each other because a larger porosity means more pore space in which biomass can grow. However, biomass growth causes a decrease in local porosity meaning the correlation may be negative. The results of our general temporal trends, as well as the correlation heatmap, show a negative relationship, implying that for the simulations in this chapter, the main relationship is defined by the growth of biomass causing a decrease in porosity. However, this largely has to do with the density of the biomass, which we defined as 10g/L. A greater density would mean less volume change per mol of bacteria produced.

The correlation heatmap (Fig. 12a) also reveals important relationships for chromium. A large majority of correlations between chromium and other values have magnitudes below 0.1. The only significant correlations ($\geq$0.1) that chromium has are with biomass (-0.18), electron donor (-0.15), porosity (0.10), and λ_D (0.16). As we saw from the sensitivity analysis of the impact of λ_D on chromium concentrations (Fig. 12e), a higher value of λ_D leads to higher amounts of chromium concentration as most of the molasses is taken up to grow biomass but a large number of mols are required to produce one mol of biomass. The correlation with porosity is also related to its correlation with biomass. As biomass concentrations increase, they reduce the local porosity where growth is high. Porosity reduction only happens at a measurable level once biomass concentrations exceed about 1 mol/m^3, meaning that porosity reduction, at least in the case of our simulations, is almost always associated with a decrease in chromium concentrations.

The PCA (Fig. 13a) and correlation heatmap (Fig. 13b) add significant information to the feedback cycles of known relationships identified in section 5.4.1. To update the feedback cycles with the most relevant relationships identified in our simulations, we use the correlation matrix to identify any correlations with magnitude greater than 0.15. To portray the fact that this new feedback cycle (Fig. 14) purely represents our simulations, we have removed oxygen from the cycle and added Cr(VI). Also, to indicate the difference between speed and flow direction, we have added pressure to the cycle, and changed ”flow vector” to V_x. Finally, we have changed permeability to porosity since we didn't include permeability in the PCA and correlation analysis due to issues of multicollinearity between permeability and porosity. This new feedback cycle shows more negative relationships than positive ones. Pressure is responsible for most of the large positive correlations between these key variables. It has large positive correlations with biomass and nutrients, which in turn influence many of the other key simulation variables. Based on the number of large correlations present, biomass and nutrient concentrations have the greatest number of variables with significant relationships (4 each), indicating the general complexity interactions that impact biomass growth and nutrient concentrations in the hyporheic zone.

Compared to the feedback figure designed from knowledge of intuitive relationships (Fig. 7), the new feedback figure presents many differences. In fact, the only relationships that remain unchanged are the positive correlations between nutrients and biomass, and pressure (or flow vector) and nutrients, and the negative correlation between biomass and porosity (or permeability). The most surprising deletion of relationships is between temperature and biomass. Biomass growth in our simulations was explicitly made a function of temperature through the Ratkowski function, but this impact was not large enough to results in a high amount of correlation between biomass and temperature. The most surprising correlation sign flip is between porosity

and flow speed. Initially, we assumed porosity would increase as flow speed/pressure increases, due to the speed-based decay of biomass (increased flow speeds result in high rates of biomass decay which thus leads to increases in porosity). However, our results imply that this speed-based biomass decay is not one of the biggest factors impacting porosity. Rather, porosity is primarily controlled by the influx of nutrients that result in increased biomass growth and thus a reduction in porosity, meaning the correlation between V_x and porosity is generally negative.

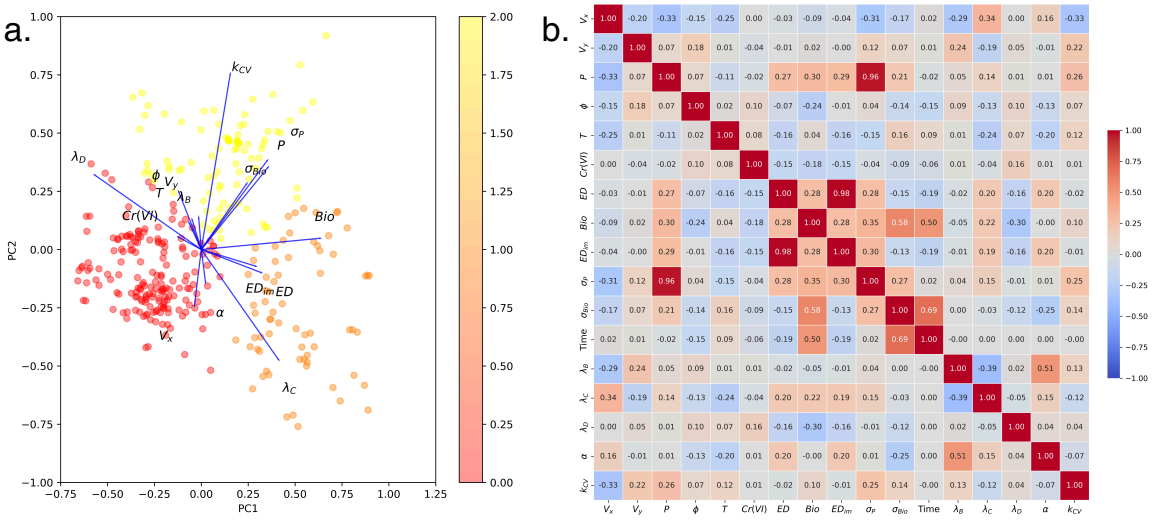


Figure 13. PCA for physio-chemical features that were varied for each simulation. The x-axis shows the first principal component, the y-axis shows the second principal component, and the color bar represents the KMeans clustering output in the 2D PCA space. ED, although the largest correlations for biomass are with σ_{Bio} and σ_{ED}. Cr(VI) shows complete anticorrelation with biomass, which is somewhat surprising given our other results that show the dominance of abiotic reduction over biotic reduction (Fig. 3). σ_ϕ and ϕ aren't strongly correlated with anything, although the show slight positive correlation with Cr(VI) and negative correlation with biomass, indicating that higher porosity soils may have less reductive capacity.

The most surprising new connection is between nutrients and temperature. Given that temperature has no correlation with biomass, it is somewhat surprising that a negative correlation exists between temperature and nutrients. Most likely, this correlation is the result of increased nutrient consumption by biomass at higher temperatures, although it is unclear why the correlation is so much stronger for nutrient concentrations than biomass concentrations. Overall,

the complex relationships between flow conditions, nutrient availability, and biomass distribution provide essential insights for designing effective bioremediation strategies and improving predictive models of HZ processes.

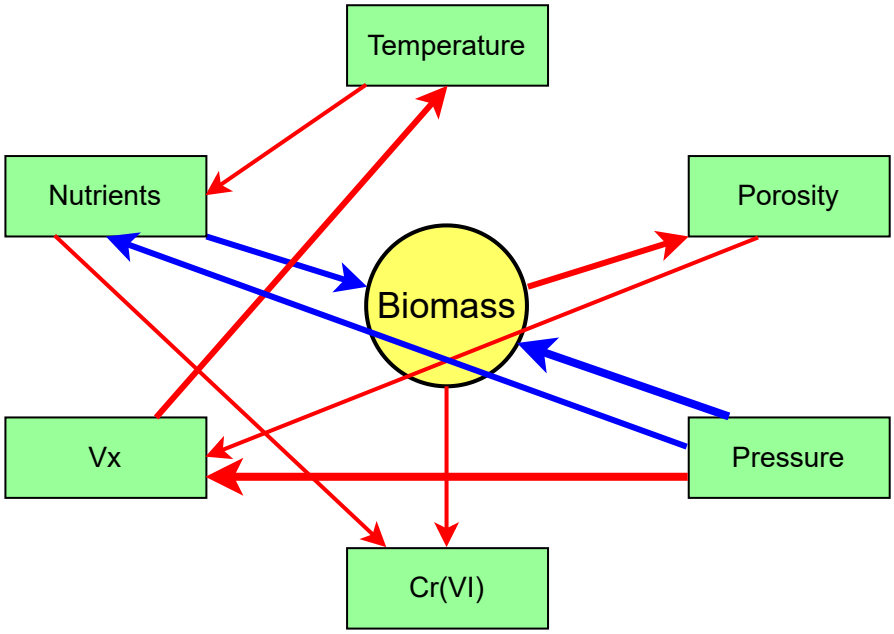


Figure 14. Primary feedback cycles determined from correlation heatmap of simulations of biomass growth in the hyporheic zone. A blue arrow signifies a positive correlation (increase leads to increase), and a red arrow indicates a negative correlation (increase leads to decrease).

5.4.5 Velocity-Based Biomass Decay Sensitivity Analysis

In addition to using sensitivity analysis to understand the relationships between the variables of our simulations, we also use sensitivity analysis to illustrate the impact of changing the parameters of our velocity-based biomass decay equation. Specifically, we show the spatial distributions at t=5400 hours for biomass (Figs. 15a-15f) and Cr(VI) (Figs. 15g-15l) for the situations of the calibrated decay parameters, no decay, high β, low β, high α, and low α. The spatial distributions for biomass generally show that biomass concentrations, as well as the spatial extent of the biomass plume, are highly dependent on the velocity-based biomass decay parameters. The

simulation with no decay has the highest concentrations of biomass, followed by the β = 1.2 simulation, then the $\alpha = 2 \times 10^{-4}$ simulation. The $\alpha = 2 \times 10^{-10}$ simulation shows essentially no difference from the calibrated simulation, and the β = 0.4 simulation shows much lower concentrations than the calibrated simulation. For the case of no decay, the spatial distributions aren't very different from those of the calibrated decay, but the low beta case has a dramatically different spatial distribution. The Cr(VI) distributions follow the same patterns as the biomass, except the no decay, $\alpha = 2 \times 10^{-4}$ and β = 1.2 simulations show less chromium (or more Cr(VI) reduction) than the other simulations.

We also show the spatial distributions of porosity (Fig. 15m) and V_x (Fig. 15n) to gain deeper insight into the spatial distributions of biomass and Cr(VI). The spatial distribution of biomass for the calibrated decay parameters (Fig. 15a) are largely determined by the low-porosity area at x=250, y=50 and the high porosity zone at x=175, y=80. Decay is highest where fluid speed is highest, which is in the corners of the simulation and the low-porosity zones in the middle and left side of the domain (Fig. 15n). Unsurprisingly, these are the areas where we see the greatest amount of difference from the simulation with calibrated parameters. Specifically, the no decay (Fig. 15b) and β = 1.2 (Fig. 15e) simulations for biomass show much greater biomass concentrations in the corners and middle high-porosity area of the domain compared to that of the calibrated decay simulation.

Surprisingly, although our general results (Fig. 10) show increased biomass results in a relative decrease in Cr(VI) reduction, we observe the opposite trend here. Looking at the time series outputs for the sensitivity analysis can help clear this confusion (Fig. 16). For biomass (Fig. 16a) and chromium (Fig. 16b), the time series plots show the relative differences highlighted by the spatial distribution plots. The time series plots for molasses (Fig. 16c), on the other hand,

shows very little difference due to changes in the velocity-based biomass decay parameters. Thus, we see that for each case of different parameter values, the same amount of molasses is consumed by biomass (besides the case of β = 0.4). Thus, greater biomass concentrations result in greater chromium reduction, since each simulation has a relatively equal amount of abiotic reduction. In other words, the increases/decreases in bioreduction due to parameter changes in the velocity-based biomass decay equation don't result in large decreases/increases in abiotic reduction. Overall, this sensitivity analysis of the novel speed-based biomass decay model demonstrates its generally small impact (besides the case of β = 0.4) on biomass distribution and chromium reduction. For β = 0.4, which represents a weakly adhesive biofilm in which the bacteria are poor initial surface colonizers, we see biomass concentrations may be dramatically reduced. The results highlight the importance of accurately representing biofilm dynamics in response to hydrodynamic conditions, which can substantially influence the overall effectiveness of bioremediation processes in the HZ.

5.5 Conclusions

This study provides valuable insights into the complex dynamics of biomass growth and chromium reduction in the hyporheic zone. Through a series of sophisticated simulations and comprehensive analyses, we have demonstrated the intricate interplay between various physio-chemical features, including flow characteristics, nutrient concentrations, temperature, and permeability. Our novel velocity-based biomass decay model, incorporated into PFLOTRAN, offers a more nuanced representation of biofilm dynamics in response to hydrodynamic conditions. This advancement allows for more accurate predictions of biomass distribution and its impact on contaminant reduction processes. Key findings from this study include:

- The identification of abiotic reduction as the dominant process in chromium remediation (for situations with $D1 \geq 1$).
- The identification of biomass growth as the primary process controlling the spatial distributions of remediation hotspots, especially for situations with $D1 \leq 1$.
- The quantification of relationships between biomass concentration and porosity, providing a useful predictive tool for future modeling efforts.

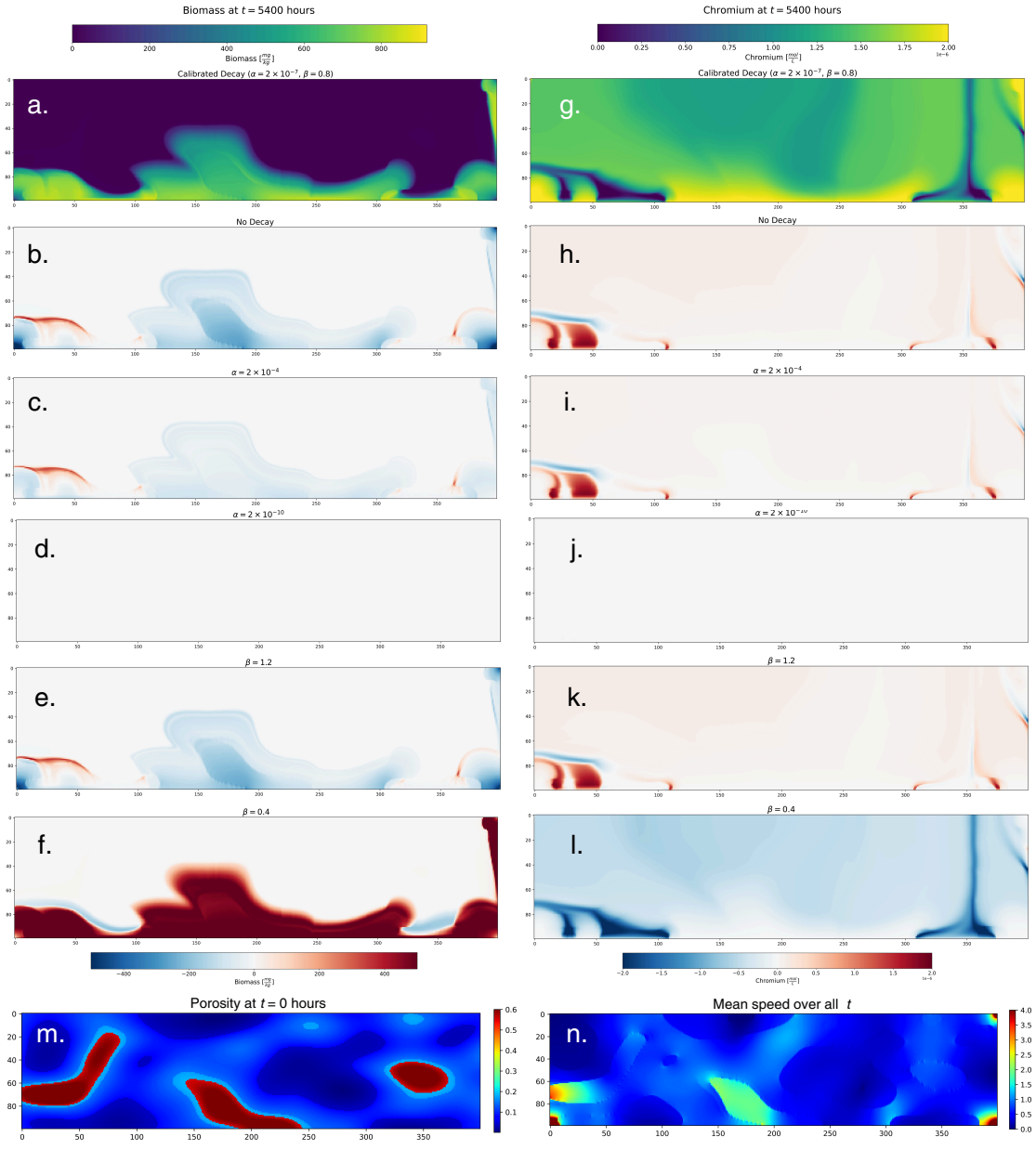


Figure 15. Spatial distributions of biomass (a-f) and chromium (g-l) at t = 5400 hours for the calibrated decay values (a & g), no decay (b & h), $\alpha = 2 \times 10^{-4}$ (c & i), $\alpha = 2 \times 10^{-10}$ (d & j), $\beta = 1.2$ (e & k), and $\beta = 0.4$ (f & l). All figures besides a and g show the difference between the calibrated decay and the respective statistic. Spatial distributions of porosity at $t = 0$ (m) and V_x averaged over all t (n) are used to further understand the differences in simulation output due to changes in the velocity-based biomass decay parameters. The top colormaps represents the concentrations shown in the calibrated decay plots (a & g), and the bottom colormaps represent the feature (biomass or chromium) concentration of the calibrated decay minus the feature concentration for the particular change in parameter value.

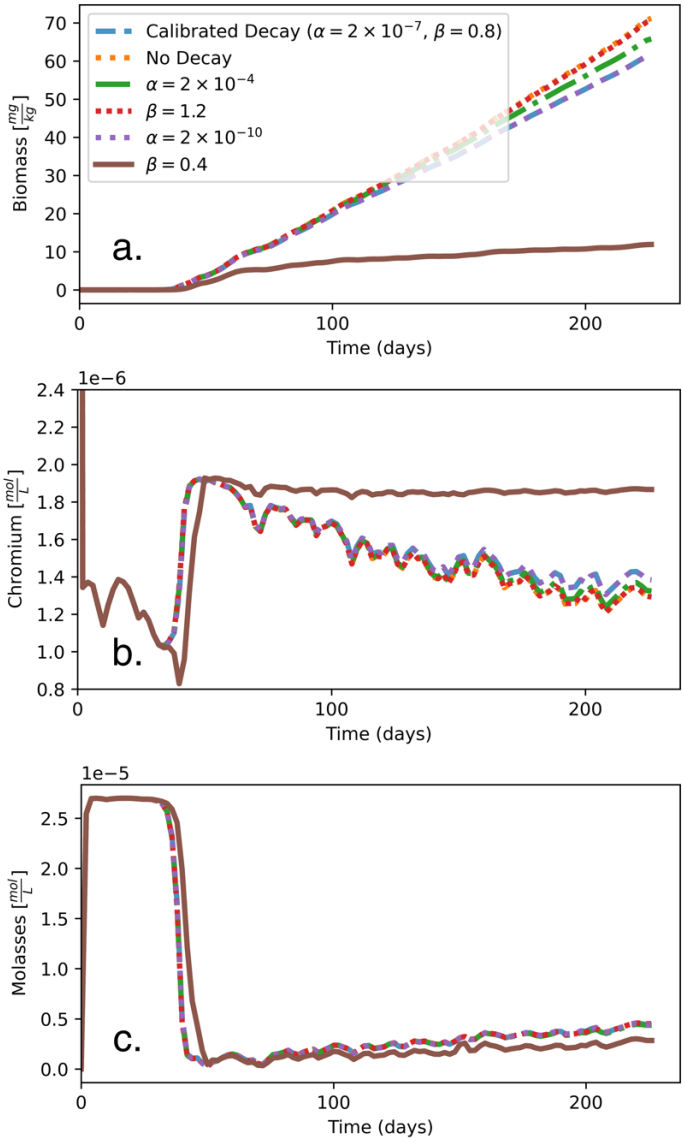


Figure 16. Time series plots (averaged over both spatial dimensions) for biomass (a), chromium (b), and molasses (c) for simulations with the calibrated decay parameters ($\alpha = 2 \times 10^{-7}$, $\beta = 0.8$), no decay, $\alpha = 2 \times 10^{-7}$, $\alpha = 2 \times 10^{-10}$, $\beta = 1.2$, and $\beta = 0.4$. Differences due to parameter changes are relatively small except for $\beta = 0.4$, and are largest for biomass, then chromium, then molasses.

- The revelation of complex feedback mechanisms between flow conditions, nutrient availability, and biomass growth, highlighting the importance of considering spatial heterogeneity in hyporheic zone modeling.
- The demonstration of the impact of speed-based biomass decay on overall system behavior and contaminant reduction efficiency, showing that the impacts are largely negligible except in cases of extremely high fluid speeds or weakly adhesive bacteria ($\beta \leq 0.4$) that make poor initial surface colonizers.

These results have far-reaching implications for bioremediation strategies in contaminated aquifers. By providing a more accurate representation of biomass dynamics and contaminant reduction processes, this study enables better prediction and optimization of remediation efforts. However, we also recognize that this study contains many limitations. For example, we acknowledge that a more robust equation for speed-based biomass decay should be derived from first principles (such as continuity and mass balance) and incorporate a variety of upscaled parameters, but this was beyond the scope of our work, especially considering the complexities of integrating such an equation into PFLOTRAN. Thus, we opted to solve this problem through a method of adapting a framework already present in PFLOTRAN and calibrating fitting parameters with pore-scale research of biofilm decay as a function of shear stress. This chapter also reflects one of the primary goals to improve bioremediation. Through improvements in our understanding of abiotic and biotic chromium reduction, and the sensitivity analysis of the impact of speed-based biomass decay on chromium reduction, we provide valuable insights and tools that can help improve contaminant remediation efforts.

Future research should focus on validating these findings with field studies and exploring the implications for bioremediation strategies in diverse hyporheic zone environments. Additionally, the integration of this advanced biomass decay model with other reactive transport processes could further improve our understanding and prediction of complex subsurface biogeochemical dynamics.

While this chapter is not directly influenced by any of the previous three chapters, it seeks to directly improve microbe-mediated reactive transport methods, which is the fundamental and overall goal of this dissertation. Furthermore, this chapter uses micro-scale physics to examine the transport of biomass at the Darcy scale, thus adding breadth to this dissertation in terms of its potential applications to microbe-mediated reactive transport.

References

[1] T. Singh, L. Wu, J. D. Gomez-Velez, J. Lewandowski, D. M. Hannah, S. Krause, Dynamic hyporheic zones: Exploring the role of peak flow events on bedform-induced hyporheic exchange, Water Re- sources Research 55 (1) (2019) 218–235.

[2] M. H. Kaufman, M. B. Cardenas, J. Buttles, A. J. Kessler, P. L. Cook, Hyporheic hot moments: Dissolved oxygen dynamics in the hyporheic zone in response to surface flow perturbations, Water Resources Research 53 (8) (2017) 6642–6662.

[3] F. Boano, J. W. Harvey, A. Marion, A. I. Packman, R. Revelli, L. Ridolfi, A. W¨orman, Hyporheic flow and transport processes: Mechanisms, models, and biogeochemical implications, Reviews of geophysics 52 (4) (2014) 603–679.

[4] L. Bardini, F. Boano, M. Cardenas, R. Revelli, L. Ridolfi, Nutrient cycling in bedform induced hyporheic zones, Geochimica et Cosmochimica Acta 84 (2012) 47–61.

[5] A. J. Boulton, S. Findlay, P. Marmonier, E. H. Stanley, H. M. Valett, The functional significance of the hyporheic zone in streams and rivers, Annual review of ecology and systematics 29 (1) (1998) 59–81.

[6] J. V. Kunz, M. D. Annable, J. Cho, W. von Tu¨mpling, K. Hatfield, S. Rao, D. Borchardt, M. Rode, Quantifying nutrient fluxes with a new hyporheic passive flux meter (hpfm), Biogeosciences 14 (3) (2017) 631–649.

[7] S. Roy Chowdhury, J. P. Zarnetske, M. S. Phanikumar, M. A. Briggs, F. D. Day-Lewis, K. Singha, Formation criteria for hyporheic anoxic microzones: Assessing interactions of hydraulics, nutrients, and biofilms, Water Resources Research 56 (3) (2020) no–no.

[8] S. Cook, O. Price, A. King, C. Finnegan, R. van Egmond, H. Sch¨afer, J. M. Pearson, S. Abolfathi, G. D. Bending, Bedform characteristics and biofilm community development interact to modify hyporheic exchange, Science of the Total Environment 749 (2020) 141397.

[9] A. Caruso, F. Boano, L. Ridolfi, D. L. Chopp, A. Packman, Biofilm-induced bioclogging produces sharp interfaces in hyporheic flow, redox conditions, and microbial community structure, Geophysical Research Letters 44 (10) (2017) 4917–4925.

[10] X. Ping, M. Jin, Y. Xian, Effect of bioclogging on the nitrate source and sink function of a hyporheic zone, Journal of Hydrology 590 (2020) 125425.

[11] L. Wang, Z. Wang, Y. Li, W. Cai, Y. Zou, C. Hui, Deciphering solute transport, microbiota assembly patterns and metabolic functions in the hyporheic zone of an effluent-dominated river, Water Research 251 (2024) 121190.

[12] Y. Wang, Y. Wang, J. Shang, L. Wang, Y. Li, Z. Wang, Y. Zou, W. Cai, L. Wang, Redox gradients drive microbial community assembly patterns and molecular ecological networks in the hyporheic zone of effluent-dominated rivers, Water Research 248 (2024) 120900.

[13] D. Liang, J. Song, J. Xia, J. Chang, F. Kong, H. Sun, D. Cheng, Y. Zhang, et al., Effects of heavy metals and hyporheic exchange on microbial community structure and functions in hyporheic zone, Journal of Environmental Management 303 (2022) 114201.
[14] L. Zhang, C. Zhang, K. Lian, C. Liu, Effects of chronic exposure of antibiotics on microbial community structure and functions in hyporheic zone sediments, Journal of Hazardous Materials 416 (2021) 126141.
[15] Y. Xian, M. Jin, H. Zhan, X. Liang, Permeable biofilms can support persistent hyporheic anoxic microzones, Geophysical Research Letters 49 (14) (2022) e2021GL096948.
[16] F. Kazemifar, G. Blois, M. Aybar, P. Perez Calleja, R. Nerenberg, S. Sinha, R. J. Hardy, J. Best, G. H. Sambrook Smith, K. T. Christensen, The effect of biofilms on turbulent flow over permeable beds, Water Resources Research 57 (2) (2021) e2019WR026032.
[17] G. Wei, J. Q. Yang, Impacts of hydrodynamic conditions and microscale surface roughness on the critical shear stress to develop and thickness of early-stage pseudomonas putida biofilms, Biotech- nology and bioengineering 120 (7) (2023) 1797–1808.
[18] E. Tsagkari, S. Connelly, Z. Liu, A. McBride, W. T. Sloan, The role of shear dynamics in biofilm formation, npj Biofilms and Microbiomes 8 (1) (2022) 33.
[19] N. S. Panikov, Microbial growth kinetics (1995).
[20] T. D. Brock, Microbial growth rates in nature, Bacteriological reviews 35 (1) (1971) 39–58.
[21] P. Gikas, S. Sengor, T. Ginn, J. Moberly, B. Peyton, The effects of heavy metals and temperature on microbial growth and lag, Global Nest J 11 (3) (2009) 325–32.
[22] M. Peleg, A new look at models of the combined effect of temperature, ph, water activity, or other factors on microbial growth rate, Food Engineering Reviews 14 (1) (2022) 31–44.
[23] Y. Zhang, Q. Tang, P. Shi, T. Katsumi, Influence of bio-clogging on permeability characteristics of soil, Geotextiles and Geomembranes 49 (3) (2021) 707–721.
[24] L. Allison, Effect of microorganisms on permeability of soil under prolonged submergence, Soil Science 63 (6) (1947) 439–450.
[25] C. I. Steefel, S. Molins, Crunchflow, Software for modeling multicomponent reactive flow and trans- port. User's manual (2009) 12–91.
[26] P. C. Lichtner, G. E. Hammond, C. Lu, S. Karra, G. Bisht, B. Andre, R. Mills, J. Kumar, Pflotran user manual: A massively parallel reactive flow and transport model for describing surface and subsurface processes, Tech. rep., Los Alamos National Laboratory (LANL), Los Alamos, NM (United States . . . (2015).
[27] M. D. White, M. Oostrom, Stomp subsurface transport over multiple phases version 3.0 user's guide, Tech. rep., Pacific Northwest National Lab., Richland, WA (US) (2003).
[28] Z. Hou, H. Ren, C. J. Murray, X. Song, Y. Fang, E. V. Arntzen, X. Chen, J. C. Stegen, M. Huang, J. D. Gomez-Velez, et al., A novel construct for scaling groundwater–river interactions based on machine-guided hydromorphic classification, Environmental Research Letters 16 (10) (2021) 104016.
[29] D. Dwivedi, I. C. Steefel, B. Arora, G. Bisht, Impact of intra-meander hyporheic flow on nitrogen cycling, Procedia Earth and Planetary Science 17 (2017) 404–407.
[30] S. B. Yabusaki, Y. Fang, K. H. Williams, C. J. Murray, A. L. Ward, R. D. Dayvault, S. R. Waichler, D. R. Newcomer, F. A. Spane, P. E. Long, Variably saturated flow and multicomponent biogeochemi- cal reactive transport modeling of a uranium bioremediation field experiment, Journal of contaminant hydrology 126 (3-4) (2011) 271–290.
[31] Y. Fang, X. Chen, J. Gomez Velez, X. Zhang, Z. Duan, G. E. Hammond, A. E. Goldman, V. A. Garayburu-Caruso, E. B. Graham, A multirate mass transfer model to represent the interaction of multicomponent biogeochemical processes between surface water and hyporheic zones (swat-mrmt-r 1.0), Geoscientific Model Development 13 (8) (2020) 3553–3569.
[32] S. K. Hansen, S. Pandey, S. Karra, V. V. Vesselinov, Chrotran 1.0: A mathematical and compu- tational model for in situ heavy metal remediation in heterogeneous aquifers, Geoscientific Model Development 10 (12) (2017) 4525–4538.
[33] K. Chen, X. Chen, J. C. Stegen, J. A. Villa, G. Bohrer, X. Song, K.-Y. Chang, M. Kaufman, X. Liang,
[34] Z. Guo, et al., Vertical hydrologic exchange flows control methane emissions from riverbed sediments, Environmental Science & Technology 57 (9) (2023) 4014–4026.
[35] H. Le Lay, Z. Thomas, F. Rouault, P. Pichelin, F. Moatar, Characterization of diffuse groundwater inflows into stream water (part ii: quantifying groundwater inflows by coupling fo-dts and vertical flow velocities), Water 11 (12) (2019) 2430.
[36] J. Ren, B. Chen, W. Zhang, L. Men, J. Yang, Y. Li, Quantification of the temporal–spatial distribu- tions characteristics of streambed hyporheic exchange fluxes with the seasonal variation using heat as a tracer, Environmental Earth Sciences 79 (2020) 1–19.
[37] Z. Wang, M. Zhao, Z. Yan, Y. Yang, K. J. Niklas, H. Huang, T. D. Mipam, X. He, H. Hu, S. J. Wright, Global patterns and predictors of soil microbial biomass carbon, nitrogen, and phosphorus in terrestrial ecosystems, Catena 211 (2022) 106037.

[38] A. Golparvar, Y. Zhou, K. Wu, J. Ma, Z. Yu, A comprehensive review of pore scale modeling methodologies for multiphase flow in porous media, Advances in Geo-Energy Research 2 (4) (2018) 418–440.
[39] H. Jasak, A. Jemcov, Z. Tukovic, et al., Openfoam: A c++ library for complex physics simulations, in: International workshop on coupled methods in numerical dynamics, Vol. 1000, Dubrovnik, Croatia), 2007, pp. 1–20.
[40] H. G. Weller, G. Tabor, H. Jasak, C. Fureby, A tensorial approach to computational continuum mechanics using object-oriented techniques, Computers in physics 12 (6) (1998) 620–631.
[41] D. S. Esser, J. H. Leveau, K. M. Meyer, Modeling microbial growth and dynamics, Applied microbi- ology and biotechnology 99 (2015) 8831–8846.
[42] C. Cruz-Paredes, D. T´ajmel, J. Rousk, Can moisture affect temperature dependences of microbial growth and respiration?, Soil Biology and Biochemistry 156 (2021) 108223.
[43] A. T. Nottingham, E. Bååth, S. Reischke, N. Salinas, P. Meir, Adaptation of soil microbial growth to temperature: using a tropical elevation gradient to predict future changes, Global change biology 25 (3) (2019) 827–838.
[44] K.-Y. LI, J. A. TORRES, Microbial growth estimation in liquid media exposed to temperature fluctuations, Journal of food science 58 (3) (1993) 644–648.
[45] D. A. Ratkowsky, J. Olley, T. McMeekin, A. Ball, Relationship between temperature and growth rate of bacterial cultures, Journal of bacteriology 149 (1) (1982) 1–5.
[46] S. P. Hendricks, Microbial ecology of the hyporheic zone: a perspective integrating hydrology and biology, Journal of the North American Benthological Society 12 (1) (1993) 70–78.
[47] M. E. Newcomer, S. S. Hubbard, J. H. Fleckenstein, U. Maier, C. Schmidt, M. Thullner, C. Ulrich, N. Flipo, Y. Rubin, Influence of hydrological perturbations and riverbed sediment characteristics on hyporheic zone respiration of co2 and n2, Journal of Geophysical Research: Biogeosciences 123 (3) (2018) 902–922.
[48] M. Azizi-Rad, G. Guggenberger, Y. Ma, C. A. Sierra, Sensitivity of soil respiration rate with respect to temperature, moisture and oxygen under freezing and thawing, Soil Biology and Biochemistry 165 (2022) 108488.
[49] H. Monterroso, M. A. Widdowson, W. S. Lotts, K. B. Strom, E. T. Hester, Effects of boundary hydraulics, dissolved oxygen, and dissolved organic carbon on growth and death dynamics of aerobic microbes in riverbed dune-induced hyporheic zones, Science of the Total Environment 906 (2024) 167401.
[50] Stuart, R.K., Mayali, X., Lee, J.Z., Craig Everroad, R., Hwang, M., Bebout, B.M., Weber, P.K., Pett-Ridge, J. and Thelen, M.P., 2016. Cyanobacterial reuse of extracellular organic carbon in microbial mats. The ISME Journal, 10(5), pp.1240-1251.
[51] Zhang, Y., Tang, Q., Shi, P. and Katsumi, T., 2021. Influence of bio-clogging on permeability characteristics of soil. Geotextiles and Geomembranes, 49(3), pp.707-721.
[52] Tang, Q., Gu, F., Zhang, Y., Zhang, Y. and Mo, J., 2018. Impact of biological clogging on the barrier performance of landfill liners. Journal of Environmental Management, 222, pp.44-53.
[53] Newcomer, M.E., Hubbard, S.S., Fleckenstein, J.H., Maier, U., Schmidt, C., Thullner, M., Ulrich, C., Flipo, N. and Rubin, Y., 2016. Simulating bioclogging effects on dynamic riverbed permeability and infiltration. Water resources research, 52(4), pp.2883-2900.
[54] Scholl, M.A., 2000. Effects of heterogeneity in aquifer permeability and biomass on biodegradation rate calculations—Results from numerical simulations. Groundwater, 38(5), pp.702-712.

Chapter 6: STAMNet - A Spatiotemporal Attention Module and Network for Upscaling Reactive Transport Simulations of the Hyporheic Zone

6.1 Abstract

Reactive transport (RT) simulations are important tools for understanding and predicting phenomena in the subsurface. However, RT is computationally intensive and complex simulations can be numerically unstable. Here, we present STAMNet, a low-parameter attention-based suite of neural nets that can upscale and upsample reactive transport simulations, applied to example problem of biomass growth in the hyporheic zone. We show that a simple MLP offers 30x speedup over standard multiphysics RT simulations and can accurately (90% R^2) predict the output of multiple variables of a 1x20 meter RT simulation by using the output from a 1x2 meter simulation as input. We add efficient channel attention to our optimized MLP which significantly improves the mean average error but doesn't impact the R^2. We further develop a novel spatiotemporal attention module (STAM), which results in significant improvements both in mean square error and R^2 (92.5%). Finally, we present a network architecture that utilizes STAM to accurately (99.9% R^2) upsample simulations in two dimensions. Specifically, our model allows for the 2x upsampling of simulations in the x and y dimensions to convert a coarse-grained input into a fine-grained output. These models have potential use for Monte-Carlo-style RT simulations and the work presented serves as a proof-of-concept for accurate prediction of large sets of spatiotemporal outputs.

6.2 Introduction

In the vast realm of environmental science, the hyporheic zone (HZ) stands out as a complex interface that has captured the attention of researchers for decades [1-3]. This subsurface

region, generally defined as the interface between river water and groundwater, hosts a myriad of complex interactions [4], with biofilms serving as a central character influencing broader hydrological and geochemical cycles [5].

Multiphysics simulators that use analytical and numerical methods to solve systems of equations that describe hydro-biogeochemical interactions in the subsurface environments, also known as reactive transport (RT) simulators, such as PFLOTRAN [6], STOMP [7], and CrunchFlow [8] are generally considered the gold standard for simulations of phenomena in the HZ. However, large scale RT simulations, and Monte-Carlo-type investigations of RT simulations, have high computational complexity and cost which are sensitive to convergence criteria [9-11] causing numerical instability and challenges in supporting hydro-biogeochemical research efforts.

Several recent studies in the field of computer science have shown that accurate multi-physics simulation emulation is possible with deep learning [12-16]. Thus, to alleviate the common shortcomings of RT simulators, some studies have attempted to apply these emulation frameworks to RT data [17-22]. Laloy and Jacques presented some of the earliest studies that looked into RT emulation with deep learning. They found deep neural networks (DNNs) outperform polynomial chaos expansion networks for the prediction of a target RT variable given some input variables of the RT timeseries. Although emulation is still a popular topic, much of the current research in this domain also seeks to upscale micro and pore-scale models to the macro/continuum scale. Wang and Battiato (2024) provide a comprehensive framework to upscale RT in fracture- matrix systems. Their framework uses a combination of traditional RT algorithms with a recurrent neural network (RNN) to capture the impact of small-scale features, which they show results in improved accuracy compared to a pure macroscale model. You et al. (2024) used convolutional neural networks (CNNs) to upscale pore-scale simulations to continuum-scale

simulations. They found that the effective surface area and effective diffusion coefficient could be predicted with high accuracy, but permeability is difficult to predict. These frameworks represent significant advances in the field of RT modeling, although they suffer from a lack of easy integration with current popular methods, and are constrained in their scope. General models that can be easily implemented would increase access of reactive transport simulation tools to a larger community of researchers.

In this chapter, we provide a deep-learning-based method for the upscaling and upsampling of RT simulations. We chose to model biomass growth in the hyporheic zone due to the complexity of the simulations and its importance for many biogeochemical functions they serve. Furthermore, simulations of biomass growth result in outputs at large scales that may be very different than outputs at small scales, which necessitates a more careful upscaling than a simple interpolation or polynomial fit. In addition, we consider our simulations of biomass growth in the hyporheic zone to be a proxy for general reactive transport modeling, since the biomass growth simulations take advantage of most of the modeling capabilities in PFLOTRAN.

One of the primary motivations for this chapter comes from the high computational time requirements experienced while generating experiments for chapter 5. Given that these simulations would crash about 15% of the time, and were more likely to crash for larger scale simulations, our generation of 344 simulations took over one month of constant 12-core computer use. Thus, the primary goal of this chapter to provide a tool capable of speeding up the process of generating many simulations at large scales.

The upscaling task we try to improve in this chapter is specifically the prediction of the spatiotemporal output for a 1x20 meter simulation given the output of a 1x2 meter simulation as input. To achieve this task, we use a subset of the simulations featured in chapter 5 for training

and testing. We test the performance of an optimized MLP, the MLP + efficient channel attention, and the MLP + our spatiotemporal attention module, STAM, which we find to generally outperform the other models. This model allows for a 30x speedup in the generation of large-scale simulations with an R^2 of the predicted mean time series of 0.925. We also devise an optimized linear architecture that incorporates STAM for the task of upsampling, which takes a 1x2 meter simulation with a resolution of 100 voxels/m as input and outputs a 1x2 meter simulation with a resolution of 400 voxels/m.

The work presented in this chapter is an enhanced version of the in-review article: "Berghouse, M. & Parashar, R. STAMNet—- A Spatiotemporal Attention Module and Network for Upscaling Reactive Transport Simulations of the Hyporheic Zone. *Computers and Geosciences*. 2025." This chapter adds to the work currently in review by exploring connections with chapters 4 and 5 of this dissertation and expanding the discussion on general contributions to microbe-mediated reactive transport.

6.3 Methods

This chapter uses multiphysics simulations to explore biomass growth in the HZ, and deep learning models to upscale and upsample these simulations. The simulations used in this chapter represent a subset of the simulations featured in chapter 5. While chapter 5 uses 344 simulations, many are not separated into low and high scale pairs, and some of the pairs show very little difference between outputs at low and high scales. We originally trained and tested models on all possible simulation pairs, but we found that performance of STAMNet was not significantly better than an optimized MLP without attention since a majority of the simulations were easy to predict. Thus, we arbitrarily dropped some simulations with small differences between low and high scale,

resulting in 138 simulations (out of the 344 total generated for chapter 5) used to train and test STAMNet-Upscale. In this section, we describe the boundary conditions and parameters used for our simulations, and the model architectures and training/testing procedures used for our upscaling and upsampling frameworks.

6.3.1 Simulations of the Hyporheic Zone

6.3.1.1 General Description of Simulations

Our simulations are based in PFLOTRAN, a multi-physics reactive transport simulator developed by multiple national laboratories [6]. PFLOTRAN represents a state-of-the-art computational framework for simulating coupled subsurface flow and reactive transport processes across multiple spatial and temporal scales. This massively parallel reactive transport code integrates sophisticated numerical methods to resolve multi-phase and hydro-biogeochemical interactions. The code's architecture enables the simulation of various subsurface processes, including density-dependent flow, variable saturation conditions, and non-isothermal phenomena, alongside comprehensive biogeochemical reactions such as aqueous complexation, mineral precipitation/dissolution kinetics, surface complexation, ion exchange, and microbially mediated transformations. As discussed in the introduction, we seek to use this reactive transport simulator to model biomass growth in the hyporheic zone. To this end, we have specifically adapted the Chrotran [22] version of PFLOTRAN to represent growth of biomass in the hyporheic zone at the Darcy scale. Chrotran defines biomass growth as a function of electron donor (ED) concentration through simple Monod kinetics. It uses biotic and abiotic reactions to model Cr(VI) reduction, defines a mobile-immobile mass transfer system for biomass and ED, and allows for bioclogging

modeling capabilities via the dependence of porosity and permeability on biomass concentration. For a full description of the biomass growth model, please refer to the original Chrotran paper.

The simulations described in this paper were created for the purpose of modeling complex interactions in the hyporheic zone. We simulate different flow conditions, permeability conditions, and concentration inputs to train our model on a general representation of simulations of biomass growth in the hyporheic zone. The simulations all contained high levels of nutrients that allowed for relatively linear growth throughout the time frame of the simulations (up to 228 days). For cases with low concentrations of nutrients, biomass growth leveled off more towards the end of the simulations. We chose to not investigate nutrient-limited scenarios because we observed less differences between small scale and large-scale simulations in cases of nutrient limitation, meaning a model that allows mapping between the two would be less useful. Thus, we focused on relatively high-nutrient simulations which show significant differences between small scale and large-scale outputs.

6.3.1.2 Boundary Conditions

For the baseline, we simulate a 1 meter (in vertical, or direction of hyporheic flow) by 2 meter longitudinal (in direction of river/groundwater flow) slice of a synthetic hyporheic zone represented by 100 by 200 voxels (dx = dy = 0.01 m). The top and bottom boundaries (1 m difference) respectively represent the surface and bottom-HZ pressures (which controls the amount and direction of vertical flow), and the left and right boundaries (2m difference) represent the pressure gradient in the longitudinal direction, thus controlling the vector of groundwater flow (also referred to here as the horizontal flow). The horizontal pressure gradient is constant over the duration of any given simulation, and the vertical pressure gradient for any given simulation is derived from three different sets of in-situ hyporheic flux data [23-25] (Chapter 5, Figure 5). As

discussed in further sections of the methods, all simulation variables, including the horizontal and vertical pressure gradients, take on different values for different simulations. From these base time series, we introduce random variations (large variations for the "high speed" time series and small variations for the other time series in Figure 5 of chapter 5) to increase the variability in potential flow conditions for our models to be trained on. The base set of pressure gradients for the horizontal flow was determined a range of realistic groundwater flow rates. Both horizontal flow (V_y) and vertical flow (V_x), and transport, are regulated by Dirichlet boundary conditions.

The primary motivation of this study is to develop upscaling and upsampling methods for RT simulations using deep learning (Fig. 1). Thus, we generated pairs of simulations for training that are identical in every way except scale (for upscaling) or resolution (for upsampling). As discussed above, the baseline simulations represent a 1 meter by 2 meter slice of the HZ. For the upscaling task, all models use the baseline 1x2 meter simulation as input to predict a 1x20 meter simulation. For the upsampling task, all models use the baseline 1x2 meter simulation with dx = dy = 0.01 m as input to generate a 1x2 meter simulation with dx = dy = 0.005 m. All upscaling simulations ran for 228 days (114 timesteps) and all upsampling simulations ran for 86 days (43 time steps). It should be noted that for the gaining and losing simulations the in-situ hyporheic flux data (Chapter 5, Figure 5) only extended to 170 days. We therefore applied constant flow boundary conditions to the last 58 days. The upscaling simulations were also different from the upsampling simulations in that they are based on heterogeneous permeability distributions whereas the upsampling simulations contain homogeneous permeability distributions. Sample permeability fields for the upscaling simulations are given in Figure 6 of chapter 5.

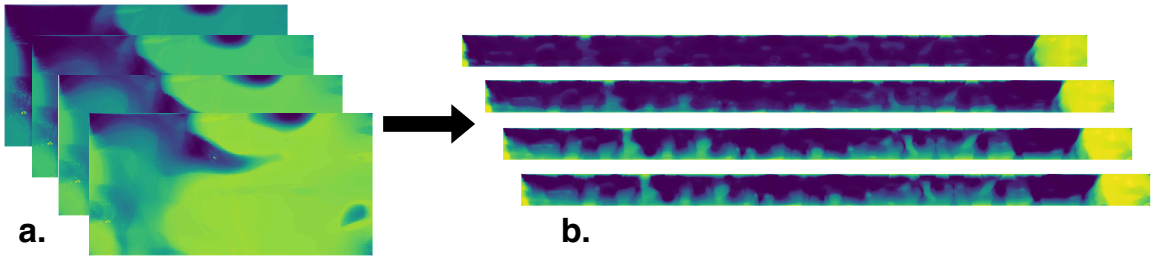


Figure 1. Sample ground truth snapshots of the 1x simulation (a) and 10x simulation (b) outputs. From top to bottom, the snapshots represent normalized biomass concentrations at t = 40, 80, 120, and 160 days. The primary motivation of this work is to provide a model that allows accurate mapping from 1x to 10x.

Table 1 Description of variables and their ranges of possible voxel-specific values used in the simulations. From left to right, these variables signify biomass, electron donor (molasses), and chromium concentrations, vertical velocity, horizontal velocity, pressure, temperature, porosity, permeability, biomass crowding parameter, and biomass growth parameter.

Variable	**B** $\left(\frac{mol}{m^3}\right)$	**ED** $\left(\frac{mol}{L}\right)$	**Cr(VI)** $\left(\frac{mol}{L}\right)$	V_y (m/hr)	V_x (m/hr)	**P** (Pa)	**T** (°C)	ϕ	**k** (m^2)	α	λ_b
Min	1e-10	1e-20	1e-20	-632	-486	-1214	4.8	1e-4	1e-15	0.5	1e-5
Max	765	5.5e-3	7.6e-3	671	651	7099	24.9	0.6	1.1e-9	3.0	1e-4
Mean	58	8.1e-6	1.4e-5	-5.8e-2	-1.4e-2	786	11.5	0.13	2e-10	2.8	1e-5

6.3.1.3 Simulation Variables

To train and test our model on a large variety of simulations, we added random variations to each of the variables of the simulations. The primary simulation variables, as well as their voxel-specific min, max, and mean values across all simulations, are given in Table 1. The average spatial distributions (in time and across all simulations) of the output features of the 1x2 meter upscaling simulations are shown in Figure 2. Molasses, biomass, and Cr(VI) all have similar distributions due to their coupling via chemical equilibria. In addition to the features listed in Table 1 and Figure 2, less consequential features that varied between simulations included S_c , S_d, and λ_c , which can all be classified as biomass growth parameters.

6.3.2 Deep-Learning-Based Upscaling

6.3.2.1 Model Architectures

Our initial model selection process was to look for published architectures that have been shown to be effective for spatiotemporal data [26-28]. However, our input and output tensors have shape [b, t, h, w], where b is batch size, t is the temporal dimension, and h and w are spatial dimensions. Thus, given the irregular shape of our inputs ([b,114,100,200]), and the large shape of our outputs ([b,114,100,2000]) for the upscaling task, we found that these published spatiotemporal models, which are often used for classification or object detection/tracking in video data, either were too large, or would not work well with our input shape. Thus, we moved to a smaller and simpler

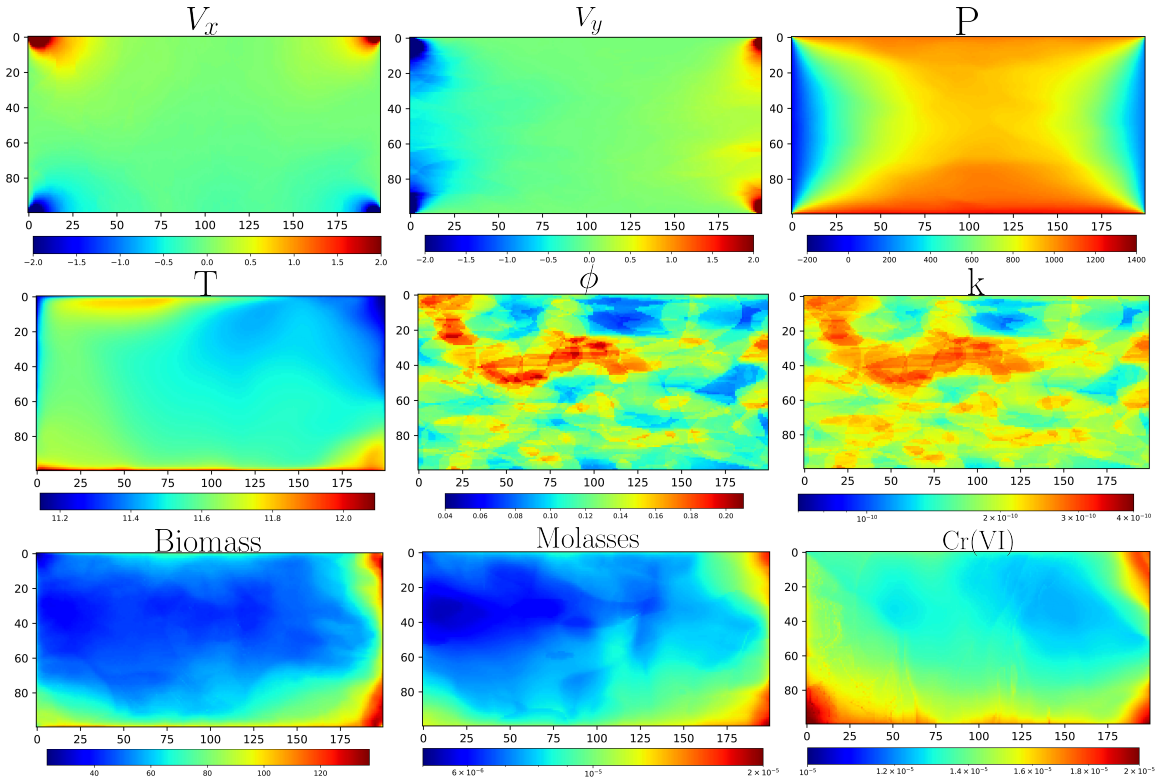


Figure 2. Mean spatial distributions for all simulations for (going from left to right and top to bottom) V_x (Darcy velocity in the vertical direction, measure in meters/hr), V_y (Darcy velocity in the direction of flow parallel to the river), pressure (Pa), temperature (◦C), porosity, permeability ($m2$), biomass (mol/$m3$), molasses (the electron donor, measured in mol/L), and Cr(VI) (mol/L).

MLP-based structure of our own design (Fig. 3a). This architecture takes in a 4D input (including batch size) and passes it through a series of linear layers with nonlinear activation functions to progressively increase the size of the final dimension to the desired number. The best number and sizes of linear layers, and the best activation function, were determined via automated hyperparameter tuning with Optuna [29]. After optimizing the structure of the MLP, we used ablation experiments with different variations of the first layer to determine the best method for initial upscaling (Table 2).

For both the upscaling and upsampling models, we also investigate the impact of attention on model performance. Specifically, we integrate efficient channel attention (ECA) [30], and a novel attention method (STAM), into the optimized MLP structure after the first layer (Figs. 3b & 3c). Our efficient channel attention method uses 1D convolution in the temporal dimension, allowing the model to focus on more relevant temporal features. STAM uses convolutions in multiple dimensions (fully described in section 6.3.2.2) to improve focus on task-relevant spatial and temporal features. The resulting architecture with the inclusion of STAM is called STAMNet. For the rest of the paper, we refer to the upscaling version of STAMNet as STAMNet-Upscale, and the upsampling version of STAMNet as STAMNet-Upsample. STAMNet-Upsample has a different architecture than STAMNet-Upscale because the task of upsampling requires a doubling in size for both of the spatial dimensions (Fig. 3c). At a basic level, STAMNet-Upscale increases the last spatial dimension by 10x, whereas STAMNet-Upsample increases both spatial dimensions by 2x.

6.3.2.2 Spatiotemporal Attention Module (STAM)

The STAM architecture (Fig 4) applies attention across multiple dimensions of the input tensor. It consists of four main branches (M1 - M4) that process the input in different permutations, allowing the network to capture dependencies across various dimensions.

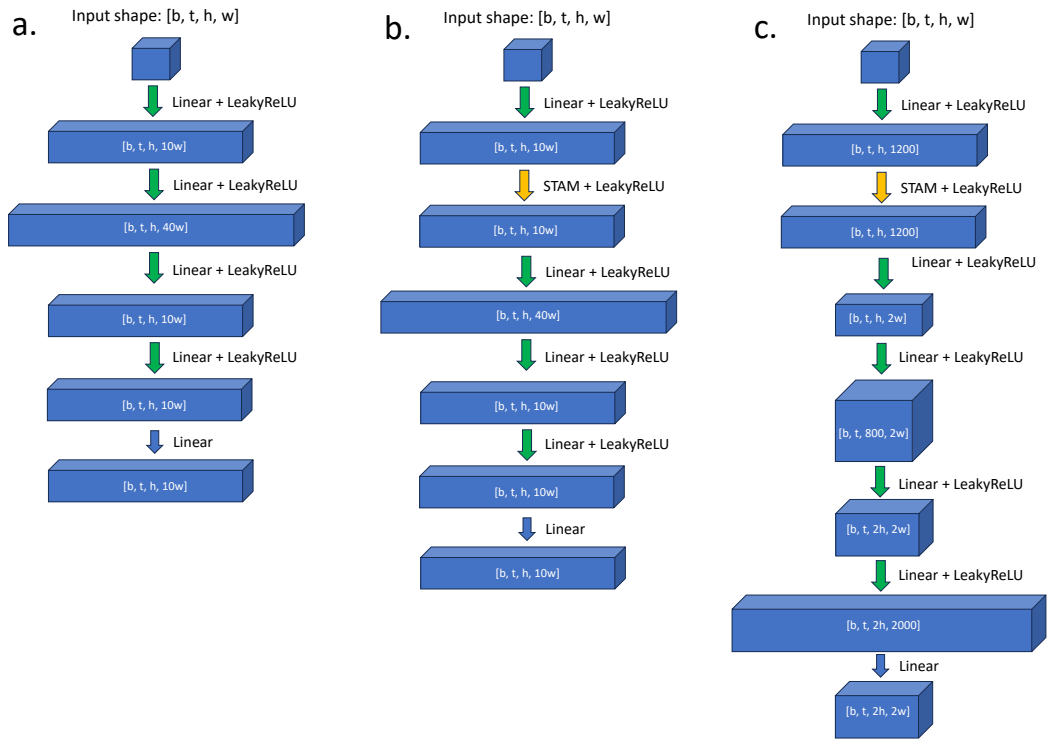


Figure 3. Model architectures for the optimal MLP (a), STAMnet-Upscale (b), and STAMNet-Upsample (c). Each block represents an intermediate output stage, and the arrows represent the layers of the model. The optimal MLP (a) and STAMNet-Upscale (b) take inputs of shape [b, t, h, w] and return outputs of shape [b, t, h, 10w]. STAMNet-Upsample takes inputs of shape [b, t, h, w] and returns outputs of shape [b, t, 2h, 2w]. STAM is a modular attention method that returns an output with the same shape as the input. The architecture of STAM is given in Figure 5. For the STAMNet-Upsample (c), permutations are used after the 2nd and 4th linear layers to have the appropriate dimensions. This model takes an input of shape [b, t, h, w] and gives an output of shape [b, t, 2h, 2w].

The four attention branches can be summarized as follows - M1 processes the input along the temporal dimension through convolutional layers and reduces the size of the width dimension through a linear layer, which results in an attention map of shape [b, t, h, 1]. M2 processes the input along the width dimension through convolutional layers and reduces the size of the time dimension (or the horizontal length being upscaled), which results in an attention map of shape [b, 1, h, w]. M3 processes the input along the height dimension and results in an attention map of shape [b, t, h, 1], and M4 processes the input along the temporal dimension and results in an

attention map of shape [b, t, 1, w]. M3 has a similar structure to M1 (and the same output shape) except it processes the height dimension through convolution instead of the weight dimension. Each branch follows a similar pattern: Conv2D (5x5) → LeakyReLU → Conv2D (1x1) → LeakyReLU → Linear → Sigmoid. The output of each branch is multiplied with the input and the resulting product is added back to the input, creating two levels of residual connections. The outputs from all branches are then averaged and passed through a sigmoid activation function and multiplied and then added to the input to get the final attention map, thus creating additional residual connections.

STAM incorporates several architectural features that enhance its ability to map spatiotemporal relationships. By processing the input tensor along different dimensions, it captures complex spatial-temporal dependencies that simpler. attention mechanisms or non-attentive models might overlook. The combination of 5x5 and 1x1 convolutions enables STAM to integrate both local and global context within each dimension [31-33]. Through residual connections and a final aggregation step, the model adaptively refines features, highlighting important patterns while attenuating less relevant information [34, 35]. The incorporation of LeakyReLU activations and dropout (in M2) introduces non-linearity and regularization, potentially enhancing the model's generalization capabilities [36]. Furthermore, by processing the input through different permutations, STAM generates complementary attention maps, effectively capturing diverse data patterns [37]. Lastly, the addition of the input to the attention-weighted features preserves original information while facilitating the learning of residual representations, thus providing a comprehensive approach to spatiotemporal feature extraction and refinement.

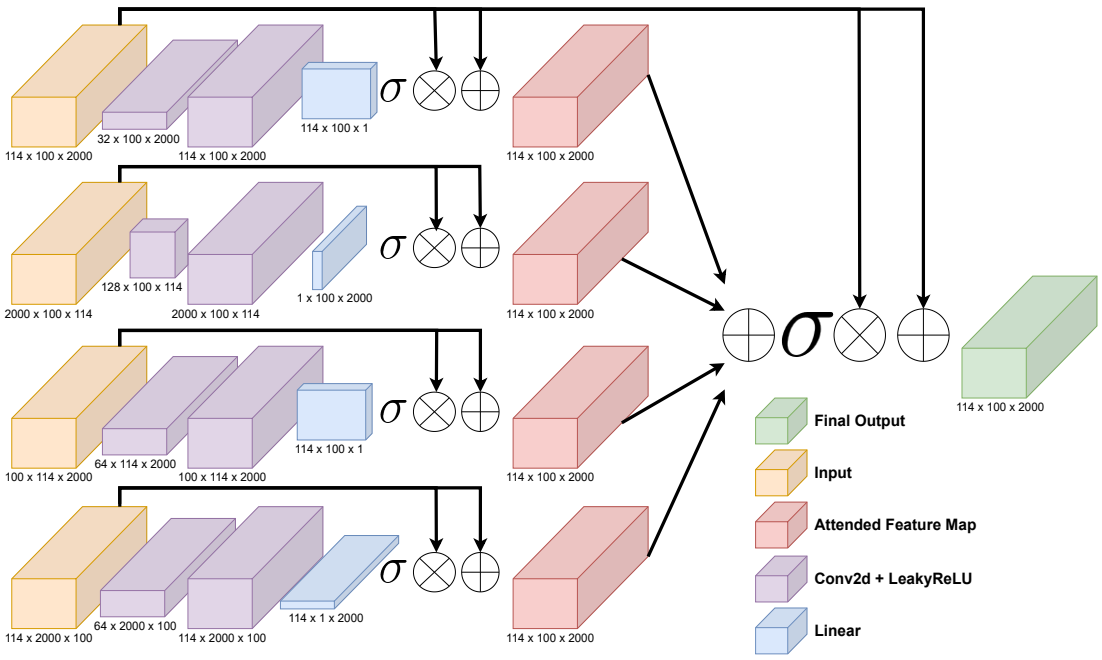


Figure 4. Architecture of the Spatiotemporal Attention Module (STAM). The model consists of four separate attention arms. Each attention arm has two 2D convolutional layers and one linear layer followed by sigmoid activation, then multiplication and addition with the original input. The attention arms differ in their shapes, which results in feature maps that are able to capture complex cross-dimensional relationships. The attended feature maps from each attention arm are then averaged, passed through a sigmoid layer, multiplied by the input and finally added to the input to get the final output of the modular attention method.

6.3.2.3 Training, Validation, and Testing Process

Although the RT simulations contain multiple output features (Fig. 2), we chose to focus on upscaling biomass, Cr(VI), and molasses. All other variables either have little variation between 1x and 10x scale simulations (such as temperature and pressure) or can be easily upscaled through physics-based methods (such as flow [38-40] and permeability [41]). With about 48 GB of VRAM, models could be developed to upscale all three variable at once. However, we were restricted to 24 GB of VRAM, and at this amount of VRAM we weren't able to effectively train multi-feature models. Thus, we trained a suite of models that separately upscale our three target variables.

For the upscaling task, 48 pairs of simulations were used for training, 8 pairs of simulations were used for validation, and 13 pairs of simulations were used for testing. Validation scores were

used to optimize hyperparameters and determine layer placement within STAMNet. Once the best model architectures were determined through hyperparameter optimization and ablation studies, the validation data was also used for training, resulting in 56 pairs of training simulations and 13 pairs of testing simulations for the final calculation of scores. We used the AdamW optimizer, dropout of 0.2 after the first linear layer (or after the attention layer for models with attention), and a learning rate (lr) between $1.6e^{-4}$ and $4.6e^{-4}$ with a cosine annealing warm restarts scheduler. For biomass upscaling we used a lr of $1.6e^{-4}$ and trained for 25 epochs, for biomass upsampling we used a lr of $4.6e^{-4}$ and trained for 900 epochs, and for ED and CR(VI) upscaling we used a lr of $3.0e^{-4}$ and trained for 110 epochs. The number of epochs used for each feature was determined based on when the validation set stopped showing improvement. For all biomass upscaling experiments, we trained and tested each model type 14 times and report the averages of each performance metric (MSE, MAE, and R^2). Each metric is calculated between all elements of the output tensor (x) and the ground truth (y). For example, the MAE is the sum of errors between each element of x and y divided by the number of elements in x and y. We also plot the mean time series and spatial distributions for each model to provide a visual understanding of the prediction errors. Specifically we use time series to investigate the average temporal distributions of the predictions of biomass, molasses (ED), and chromium for a simple interpolation model, an MLP+ECA model, and our STAMNet model. The spatial distributions are presented in two ways. The blue-green-yellow spatial distributions show the absolute concentrations of the feature in question, while the blue-white-red spatial distributions show the difference between the ground truth and the prediction for that particular model. For this visual analysis, we use a simple ensemble of the best-scoring variations of each model.

For the upsampling task, 40 pairs were used for training, 8 for validation, and 12 for testing. After optimal model structures were determined, the 8 simulation pairs used for validation were included in the training set, resulting in a final 48 simulation pairs for training and 12 simulation pairs for testing. All results for the upsampling task are a comparison of the average of 8 separately trained and tested models. For both upscaling and upsampling, the loss function used for training was $MSE + 0.6 \times MAE$, which was used over a standard MSE loss function as we found that only using MSE tends to result in a higher degree of overfitting. Furthermore, we found Huber loss to not weight the MAE strongly enough, which resulted in decreased MAE and R^2 scores.

6.4 Results

6.4.1 Ablation Experiment

To determine the best simple method of upscaling, we experimented with three model variations (Table 2). The structure of the linear model is shown in Figure 3a. This structure was determined through optimization of validation scores via Optuna. In the linear model, the first layer is a linear layer that increases the size of the final dimension of the input by 10x. To try and reduce the number of parameters, or have roughly the same number of parameters with a deeper first layer, we tried to replace the first linear layer with a 10x interpolation layer and a 20x interpolation layer. The reasoning behind this is that we had observed simple interpolation often allows for reasonably accurate upscaling compared to other simple models, so we thought it might be a parameter-efficient way to upscale the final dimension. The 10x interpolation layer takes the input of shape [b, t, h, w] and outputs a tensor of shape [b, t, h, 10w], while the 20x interpolation layer takes the same shape of input and outputs a tensor of shape [b, t, h, 20w]. Thus, the interpolation 10x model has the same structure as the linear model besides the 1st layer, which is instead a 10x repeat

interleave layer. Similarly, the 20x interpolation model has an initial layer that interpolates the final dimension of the input to 20x size. Because the linear layer of the optimal MLP upscales the final dimension to 10x, the 20x model has a slightly different structure of second layer since it takes an input of [b, t, h, 20w]. The linear model performed best in the MAE and R^2 metrics. Thus, although interpolation allows for model parameter savings, it is not much, and the reduced accuracy is not worth these savings in most cases, so we developed STAMNet on top of this optimal linear architecture.

Table 2. Ablation experiments for biomass upscaling to determine the best method of increasing dimension size. MAE is given in $\frac{mol}{m^3}$ and MSE is given in $\left(\frac{mol}{m^3}\right)^2$.

	Linear	**Interpolation (10x)**	**Interpolation (20x)**
MSE	2508	2471	2471
MAE	21.62	22.25	22.02
R^2	0.897	0.880	0.875

6.4.2 STAMNet-Upscale Performance

6.4.2.1 Biomass Upscaling

The results of our upscaling models for the biomass prediction task are given in Table 3. The interpolation model here is different than the interpolation models used in the ablation experiments. In the ablation experiments, the interpolation was used as an initial layer of a model with multiple linear layers and an activation function after the interpolation. For the interpolation model in table 3, there are no linear layers after the interpolation. In other words, it is just a simple interpolation of the final dimension, which is the most simple and rapid way to generate reasonably accurate results for the task of upscaling as defined in this paper. The simple MLP is a one-layer MLP that increases the size of the final dimension by 10x. The optimal MLP is the fully

optimized MLP structure given in Figure 3a. The structure of STAMNet, our proposed best-performing model, can be seen in Figure 3c. The MLP+ECA model has the same structure as STAMNet, but with the ECA attention module instead of the STAM attention module.

For all models tested, STAMNet-upscale shows the strongest performance by a statistically significant margin for the both MSE and R^2 metrics. Both models with attention modules outperform the optimal MLP, further indicating that attention is a useful tool for developing robust upscaling model architectures. STAMNet-Upscale performs better than the MLP+ECA model, indicating that cross-dimensional feature refinement offers performance benefits over single dimensional (temporal) feature refinement. All trained models perform better than simple interpolation, showing the general benefit to the approach of using deep learning for upscaling of reactive transport simulations.

To further investigate the performance of our different models for the task of upscaling biomass, we plot the spatial error (Fig. 5) and the mean time series error (Fig. 6). The spatial errors show that the simple interpolation (Fig. 5d), the MLP+ECA (Fig. 5c), and STAMNet-Upscale (Fig. 5b) all fail to capture fine variations in the ground truth spatial distribution. Instead, they achieve a low MSE/MAE by averaging out the variabilities in space. This is to be expected, however. Without a method that specifically constrains the spatial distributions of biomass concentrations, the model lacks the ability to predict exactly what the upscaled version will look like, so the model just makes an average guess. In other words, the neural nets may learn to approximate the vertical variability in biomass well, since this doesn't change much between small and large-scale simulations, but have no ability to predict the horizontal variability in biomass as this may change significantly based on scale and more strongly depends on the differences between the small and large- scale permeability fields. To compensate for this lack of knowledge, the neural nets make

predictions that represent averages across many horizontal voxels. We experimented with loss functions to try to add this constraint to the spatial distribution of the outputs, but found it had too negative of an impact on the outputs of time series distributions and did not improve the accuracy of the spatial distributions (either in exact value or "look") enough to warrant further investigation. Thus, although there are some differences in the spatial error between different models, no model we tested provides an adequate representation of physically realistic spatial variations, and more robust techniques are needed to achieve high-fidelity spatial predictions.

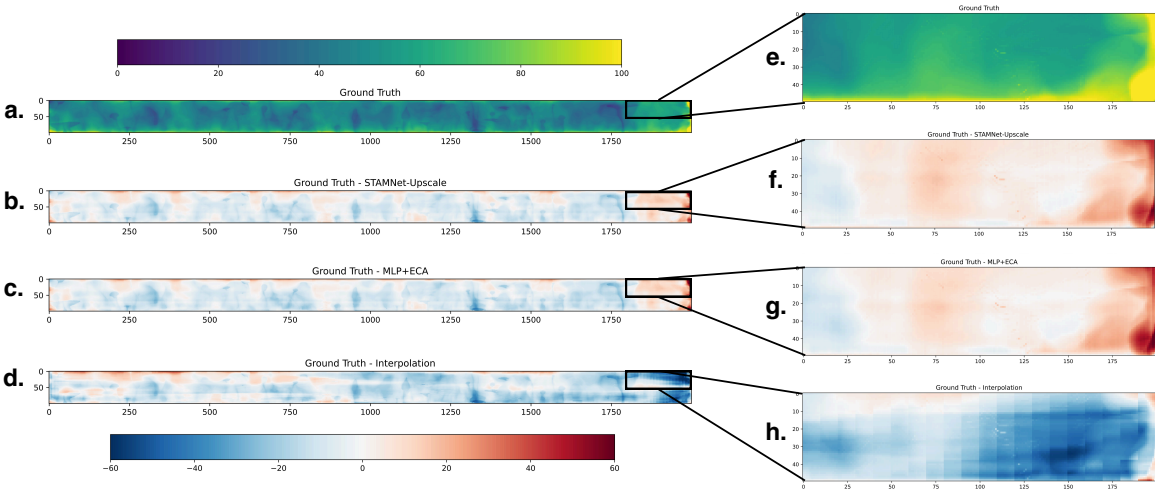


Figure 5: Spatial error distribution of biomass averaged over all test simulations and time steps. (a) Ground truth spatial distribution. (b) Ground truth minus output from STAMNet-Upscale. (c) Ground truth minus the MLP+ECA model. (d) Ground truth minus simple interpolation. (e-h) Zoomed in versions of a-d. These figures show that STAMNet has difficulty predicting fine spatial variations but is more accurate than a simple interpolation.

In terms of comparison between the methods, STAMNet-Upscale and the MLP+ECA clearly outperform the interpolation, which can be better seen from the zoomed-in sections of the spatial error distributions (Figs. 5e-5h). There is a very slight difference between error for STAMNet-Upscale and the MLP+ECA, but it is essentially negligible with regards to the overall accuracy of the predictions of spatial distributions. One big difference between the interpolation and the attention-based neural nets is that the interpolation model greatly overcalculates biomass

concentrations, especially near the right boundary of the domain. The right boundary of the domain is often a source of nutrients and thus a location of dense biomass growth. In the 1x2 meter simulation, these nutrients are able to reach into and cause biomass growth in about half of the domain, meaning a simple interpolation to the 1x20 meter simulation leads to high biomass concentrations that extend too far into the domain. Although we scale up the horizontal flow speed (V_y) for the 1x20 meter simulation, we still see that biomass tends to cover a smaller horizontal portion of the domain at larger scales. This is due to a variety of hydro-biogeochemical phenomena such as mixing, uptake and dispersion of nutrients, and the amount of space the biomass growth can cover given a certain amount of time and nutrients. In other words, because growth is largest at the domain boundaries due to the influx of nutrients, there is a limit to how far this boundary-enhanced growth can extend towards the middle of the domain. The neural nets, on the other hand, undercalculates biomass at the boundaries (right, top, and left) due to their tendency to average local variations in concentration.

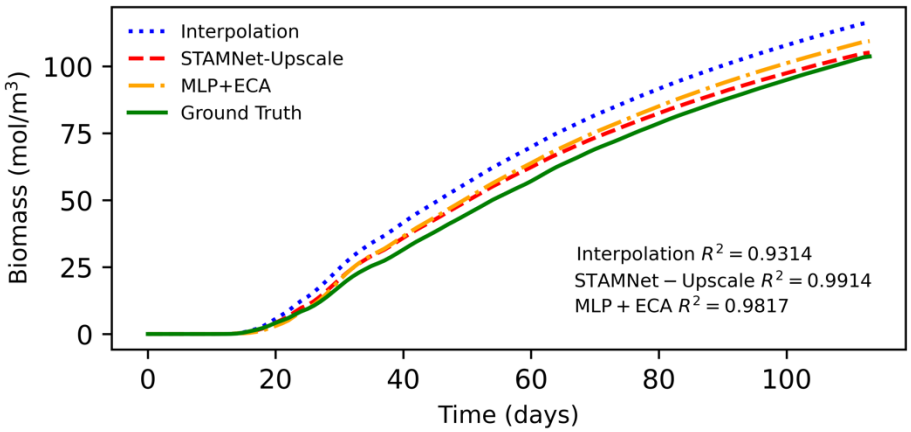


Figure 6. Biomass time series averaged over all upscaling test simulations and time steps. The blue dotted line corresponds to the time series for a simple interpolation of the input, the red dashed line corresponds to the output for STAMNetUpscale, and the green dash-dotted line corresponds to the ground truth. This figure shows STAMNet outperforms simple interpolation and achieves a high level of accuracy in terms of time series prediction.

In addition to our analysis of the spatial errors of the simple interpolation method and attention-based models, we also investigate their performance in terms of the average time series prediction (Fig. 6). Unlike the spatial distributions, all models perform quite well at the task of capturing the average upscaled time series. Both attention-based models clearly outperform a simple interpolation, and STAMNet-Upscale slightly outperforms the MLP+ECA model.

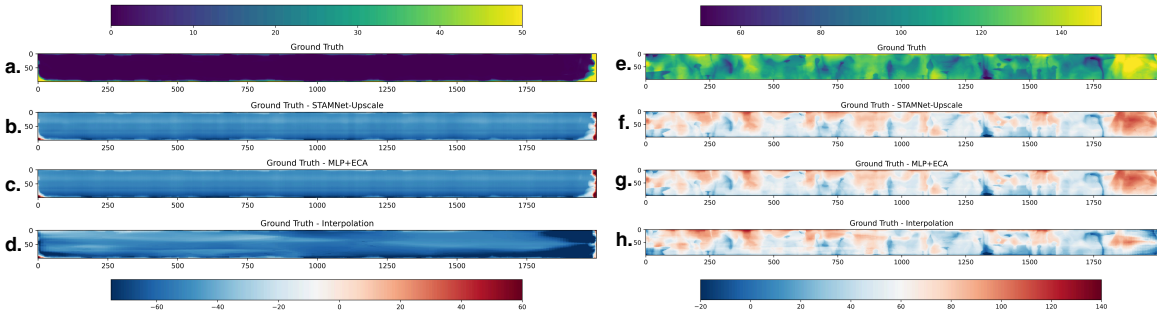


Figure 7. Spatial error distribution of biomass averaged over low and high-concentration test simulations.(a) Ground truth spatial distribution for low-concentration simulations.(b) Ground truth minus output from STAMNet-Upscale for low-concentration simulations.(c) Ground truth minus the MLP+ECA model for the low-concentration simulations. (d) Ground truth minus simple interpolation for low-concentration simulations. (e-h) High-concentration versions of a-d. These figures show greater difference between the spatial errors of the interpolation and our trained networks for low-concentration simulations than high-concentration simulations.

To further refine our general investigation of the upscaling potential of STAMNet for RT simulations, we split this analysis up to investigate performance on high-concentration and low-concentration simulations. Of the 13 test simulations, 5 simulations can be categorized as high-concentration (mean biomass greater than 50 mol/m^3), and 5 simulations can be categorized as low-concentration (mean biomass less than 15 mol/m^3). The spatial errors for both the low and high concentrations (Fig. 7) generally show the same trends as the spatial errors for the full set of results (Fig. 5). For the low concentrations, STAMNet-Upscale (Fig. 7b) and the MLP+ECA model (Fig. 7c) are completely indistinguishable, and both clearly outperform the interpolation (Fig. 7d). For the high concnetrations, there is similarly very little difference between the spatial errors for

STAMNet (Fig. 7f) and the MLP+ECA model (Fig. 7g). Both attention-based neural nets outperform simple interpolation (Fig. 7h), although similar to the full set of results for biomass upscaling, these differences are negligible compared to the overall error of the spatial distributions.

Table 3. Results for upscaling experiments. Each model was trained and tested 14 times, and values here correspond to the average scores over all 14 model iterations. MAE is given in $\frac{mol}{m^3}$ and MSE is given in $\left(\frac{mol}{m^3}\right)^2$. Values in bold indicate statistically significantly better performance than all other models.

	Interpolation	Simple MLP	Optimal MLP	MLP + ECA	STAMNet-Upscale
MSE	4089	2691	2508	2525	**2480**
MAE	26.42	28.08	21.62	**21.35**	21.60
R^2	0.727	0.903	0.897	0.892	**0.925**

Table 4. Results for upscaling experiments with molasses (ED) and Chromium. MAE is given in $\frac{mol}{m^3}$ and MSE is given in $\left(\frac{mol}{m^3}\right)^2$. Values in bold indicate statistically significantly better performance than all other models.

Cr(VI)	Interpolation	MLP + ECA	STAMNet-Upscale
MSE	5.32×10^{-4}	1.02×10^{-4}	8.92×10^{-5}
MAE	3.31×10^{-3}	2.45×10^{-3}	2.23×10^{-3}
R^2	0.9093	0.9508	0.9227
ED	Interpolation	MLP + ECA	STAMNet-Upscale
MSE	1.20×10^{-4}	2.77×10^{-5}	$\mathbf{2.62 \times 10^{-5}}$
MAE	2.72×10^{-3}	1.81×10^{-3}	1.80×10^{-3}
R^2	0.2937	0.888	**0.913**

The time series plots for the low and high-concentration upscaling (Fig. 8) reveal slightly more interesting deviations from the analysis of all test simulations. The low-concentration time series (Fig. 8a) shows dramatically better performance for the attention-based neural nets when compared to the simple interpolation. The high-concentration time series (Fig. 8b), on the other hand, shows relatively small differences between each model. Thus, our results indicate that a simple interpolation is a generally accurate way to upscale high-concentration RT simulations of

biomass growth, but not for low-concentration simulations as it results in a large amount of overprediction of biomass concentrations.

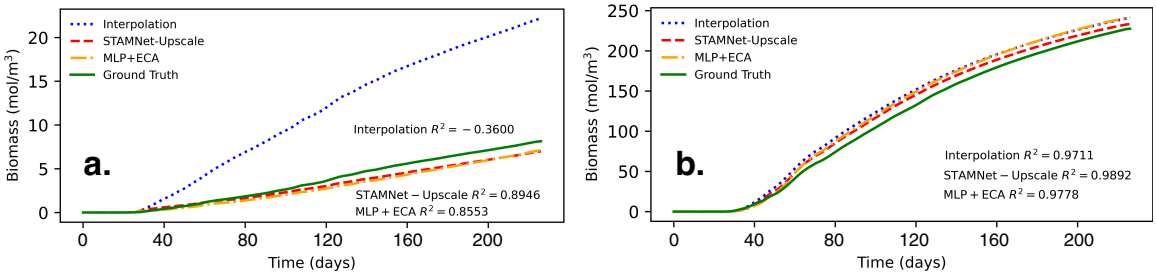


Figure 8. Biomass time series averaged over low (a) and high-concentration (b) upscaling test simulations. This figure shows STAMNet-Upscale outperforms simple interpolation and our MLP+ECA architecture, especially for lowconcentration simulations.

Across all biomass upscaling experiments, our results generally show that compared to models without attention, STAM can selectively focus on important features in the input, potentially leading to better performance on tasks that require understanding complex spatial-temporal relationships. Compared to ECA, which focuses on attention in the temporal dimension, STAM provides a more comprehensive attention mechanism that considers both spatial and temporal dimensions. STAM generally performs better than ECA, which indicates the cross-dimensional attention, which improves feature mapping in both the spatial and temporal dimensions, is advantageous for the task of upscaling.

6.4.2.2 Molasses and Cr(VI) Upscaling

We further show the strong performance of STAMNet by upscaling RT simulations for the molasses (electron donor - ED) and Cr(VI) features. We find that STAMNet significantly outperforms the simple interpolation and MLP+ECA model for ED prediction in the MSE and R^2 metrics (Table 4). Furthermore, the MLP+ECA model also significantly outperforms the simple interpolation in all metrics, which once again shows the benefit of neural architectures, and especially those with

attention, for the task of upscaling. Specifically, these results show that the benefits of our neural architectures for upscaling are not restricted to biomass, and can be extended to other features. The results for Cr(VI) similarly show high upscaling performance for both STAMNet-Upscale and the MLP+ECA model. Also, in this case, the MLP+ECA model outperforms STAMNet-Upscale in the R^2 metric, but performs worse than STAMNet-Upscale for the MSE and MAE metrics (although none of these differences are significant), indicating that the differences in performance between STAM and ECA may depend on the particular task. Thus, in addition to providing trained models for biomass, ED, and Cr(VI) upscaling, we provide multiple frameworks with which future researchers can train their own upscaling models for specific features they may be interested in. Although we find STAMNet to outperform our optimized MLP+ECA model, we encourage researchers to run their own experiments with their data to determine which model works best for their task.

Similar to our results for the biomass upscaling, we also present spatial and time series errors for chromium and molasses (Fig. 9). Like all the spatial errors for biomass, we find that STAMNet-Upscale and the MLP+ECA model outperform the simple interpolation but are not able to capture the fine-grained details of the spatial distributions for molasses (Fig. 9c) and chromium (Fig. 9d). For the mean time series comparison, we see STAMNet-Upscale and the MLP+ECA model perform equally well at molasses upscaling (Fig. 9a). Surprisingly, although the MLP+ECA model gives a higher R^2 for Cr(VI) when calculated as the mean over the set of R^2 values for each time series, when the R^2 is calculated from the mean time series, we see STAMNet-Upscale has a slightly higher R^2 (Fig. 9b). The simple interpolation performs well for Cr(VI), but not for molasses, further showing its inconsistent performance compared to that of the attention-based neural nets.

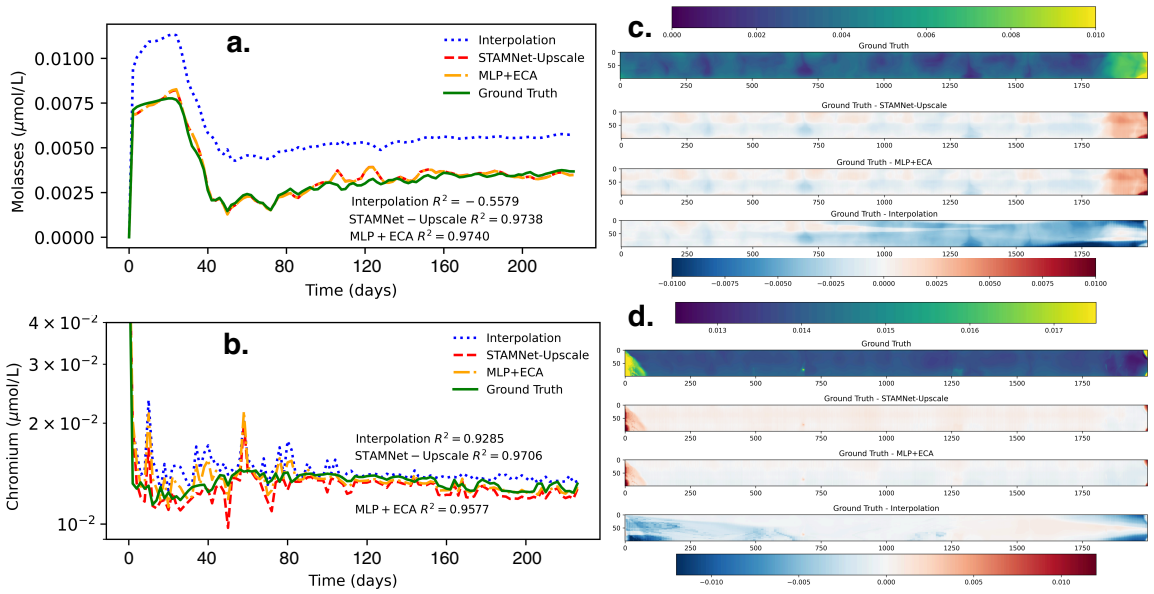


Figure 9. STAMNet and interpolation performance for (a) molasses time series, (b) chromium time series, (c) molasses spatial error distributions, and (d) chromium spatial error distributions. STAMNet outperforms interpolation in the molasses time series and spatial distribution and the chromium spatial distribution, but it is nearly indistinguishable from the interpolation for the chromium time series.

6.4.3 STAMNet-Upsample Performance

In addition to our investigation of upscaling in the sense of increasing the lateral domain of the simulation output, we also use another variant of the STAMNet architecture (Fig. 3c) to increase the resolution of the simulation output, which we refer to as the task of upsampling. We find that STAMNet-Upsample shows significantly better performance than all other models for the upsampling task in MSE and MAE, although a simple interpolation gives the best performance for R^2 (Table 5). Considering we trained and tested on a smaller number of simulations for the task of upsampling compared to the task of upsampling, and that we only used homogeneous permeability fields for the upsampling simulations, it is likely that the upsampling task was not difficult enough to result in large performance differences between models. Furthermore, some degree of overfitting on the MSE and MAE during training may have resulted in a model that focuses more on the spatial aspects of the upsampling task than the average temporal ones.

Looking at the spatial distributions for STAMNet-Upsample and the interpolation (Fig. 10), we can see that STAMNet-Upsample does indeed have more accurate spatial approximations (especially at the domain boundaries). Thus, while it could be argued the simple interpolation may be more appropriate for tasks that don't care about the accuracy of the spatial distribution, for tasks where spatial accuracy is important, STAMNet-Upsample is clearly advantageous to simple interpolation. We also find that the addition of STAM to the optimized MLP improves performance, but the addition of ECA to the optimized MLP generally decreases performance. These results contrast those of the upscaling task, where both ECA and STAM were found to improve MLP performance. As seen by the high R^2 of all models for the average time series, the task of upsampling does not result in large differences in the average temporal trends. Thus, ECA, which uses 1D convolution to improve temporal feature extraction, results in a tiny improvement in R^2 (relative to the optimal MLP), but causes a decrease in MAE and MSE due to the extra focus on temporal features. This trend is not observed in the upscaling results (Table 3), however, as ECA shows significant improvements to MAE for biomass upscaling. Given the significant differences in task and input tensors, these differences are likely a result of imperfect training hyperparameters. Thus, although ECA generally doesn't perform as well as STAMNet-Upsample for the task of upsampling, it is possible these results would be different if hyperparameters were individually tuned for each model, which strengthens our suggestion for future researchers to experiment with both the STAMNet and MLP+ECA architectures.

Table 5 Results for upsampling experiments. MAE is given in $\frac{mol}{m^3}$ and MSE is given in $\left(\frac{mol}{m^3}\right)^2$.

	Interpolation	Simple MLP	Optimal MLP	MLP+ECA	STAMNet-Upsample
MSE	73.50	54.45	28.58	29.76	**26.22**
MAE	1.443	1.557	1.263	1.268	**1.238**
R^2	0.9997	0.9987	0.9984	0.9987	0.9986

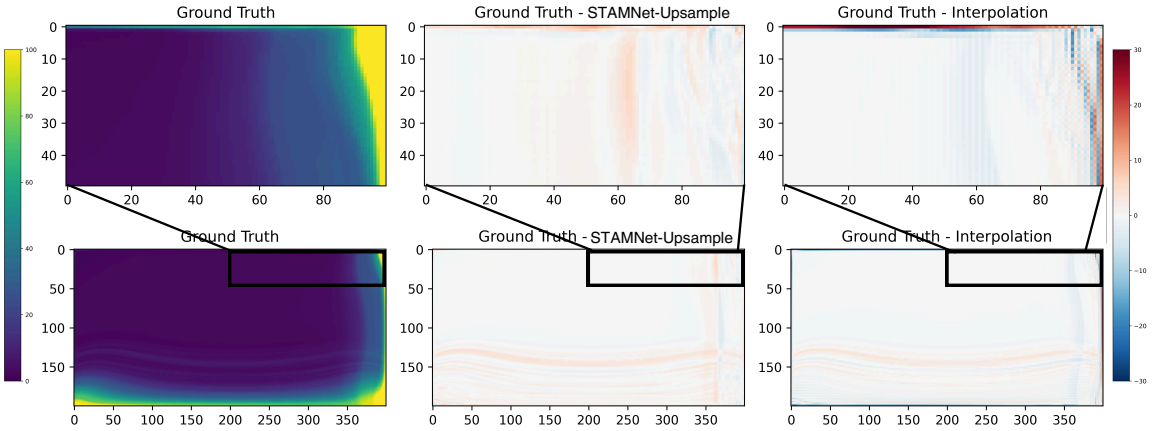


Figure 10. Spatial error distribution of biomass averaged over all test simulations and time steps for the upsampling task. (a) Ground truth spatial distribution. (b) Ground truth minus output from STAMNet-Upscale. (c) Ground truth minus simple interpolation. (d-f) Zoomed in versions of a-c. These figures show the superior spatial performance of STAMNet over interpolation for the task of upsampling, especially near the boundaries of the domain.

6.5 Conclusions

This study presents STAMNet, a novel deep learning architecture for upscaling and upsampling reactive transport simulations in the hyporheic zone. Our results demonstrate that STAMNet outperforms traditional interpolation methods and simpler neural network architectures across multiple tasks and metrics. STAMNet-Upscale significantly im- proved upon simple interpolation and optimized MLP models for biomass upscaling, achieving the highest R^2 (0.925) and lowest MSE among tested models. Furthermore, the spatiotemporal attention module (STAM) consistently out- performed efficient channel attention (ECA), indicating the benefits of cross-dimensional feature refinement for reactive transport modeling. Notably, STAMNet showed robust performance across different concentration regimes, with particularly strong improvements over interpolation for low-concentration simulations. The architecture performed well with other reactive transport variables, showing significant improvements for molasses (electron donor) upscaling and competitive performance for chromium upscaling. Furthermore,

STAMNet-Upsample demonstrated superior performance in increasing simulation resolution, particularly in capturing spatial details more accurately than simple interpolation. By enabling rapid upscaling and upsampling of simulations, this approach has the potential to accelerate research in hyporheic zone processes and enhance our understanding of complex subsurface biogeochemical dynamics. As the field continues to evolve, the integration of advanced deep learning architectures like STAMNet with domain-specific knowledge promises to unlock new possibilities in environmental modeling and decision-making.

While STAMNet shows promise for accelerating and enhancing reactive transport simulations, some limitations should be noted. The current implementation struggles to capture fine-grained spatial variations in upscaled simulations, instead producing averaged distributions. Additionally, the model's performance may vary depending on the specific reactive transport variable being predicted, as seen in the differences between biomass, molasses, and chromium results. It's also important to acknowledge that the study focused on a specific hyporheic zone scenario, and further testing is needed to confirm generalizability to other subsurface environments and reactive transport systems. Future research directions should address these limitations and expand upon the current work. Investigators should explore methods to incorporate physical constraints or multi-scale approaches to improve the spatial fidelity of upscaled predictions. Extending the model to handle a wider range of reactive transport variables and scenarios, including more complex biogeochemical reactions and heterogeneous subsurface environments, would further enhance its applicability. The integration of STAMNet with physics-based models to create hybrid approaches that leverage both data-driven and mechanistic insights is also an exciting avenue for development. Finally, investigating the potential of STAMNet for

other spatiotemporal prediction tasks beyond reactive transport, such as climate modeling or ecosystem dynamics, could open up new applications for this innovative architecture.

References

[1] Cardenas, M.B., 2015. Hyporheic zone hydrologic science: A historical account of its emergence and a prospectus. Water Resources Research 51, 3601–3616. doi:https://doi.org/10.1002/2015WR017028.

[2] Krause, S., Hannah, D.M., Fleckenstein, J.H., Heppell, C.M., Kaeser, D., Pickup, R., Pinay, G., Robertson, A.L., Wood, P.J., 2011. Inter-disciplinary perspectives on processes in the hyporheic zone. Ecohydrology 4, 481–499. doi:https://doi.org/10.1002/eco.176.

[3] Rode, M., Hartwig, M., Wagenschein, D., Kebede, T., Borchardt, D., 2015. The Importance of Hyporheic Zone Processes on Ecological Functioning and Solute Transport of Streams and Rivers. Springer Netherlands, Dordrecht. pp. 57–82. doi:10.1007/978-94-017-9846-4_4.

[4] Boulton, A.J., Datry, T., Kasahara, T., Mutz, M., Stanford, J.A., 2010. Ecology and management of the hyporheic zone: stream–groundwater interactions of running waters and their floodplains. Journal of the North American Benthological Society 29, 26–40. doi:10.1899/08-017.1.

[5] Roy Chowdhury, S., Zarnetske, J.P., Phanikumar, M.S., Briggs, M.A., Day-Lewis, F.D., Singha, K., 2020. Formation criteria for hyporheic anoxic microzones: Assessing interactions of hydraulics, nutrients, and biofilms. Water Resources Research 56, no–no.

[6] Hammond, G., Lichtner, P., Lu, C., Mills, R.T., 2012. Pflotran: Reactive flow & transport code for use on laptops to leadership-class supercomputers. Groundwater reactive transport models 5, 141–159.

[7] White, M., Oostrom, M., 1997. STOMP subsurface transport over multiple phases: Users guide. Technical Report. Pacific Northwest National Lab.(PNNL), Richland, WA (United States).

[8] Steefel, C.I., Molins, S., 2009. Crunchflow. Software for modeling multicomponent reactive flow and transport. User's manual , 12–91.

[9] Steefel, C.I., McQuarrie, K., 1996. Approaches to modeling of reactive transport in porous media. Reviews in mineralogy 34, 83–130.

[10] Su, D., Mayer, K.U., MacQuarrie, K.T., 2017. Parallelization of min3p-thcm: A high performance computational framework for subsurface flow and reactive transport simulation. Environmental Modelling & Software 95, 271–289.

[11] Yu, S., Hannah, W., Peng, L., Lin, J., Bhouri, M.A., Gupta, R., Lütjens, B., Will, J.C., Behrens, G., Busecke, J., et al., 2024. Climsim: A large multi-scale dataset for hybrid physics-ml climate emulation. Advances in Neural Information Processing Systems 36.

[12] Hennigh, O., Narasimhan, S., Nabian, M.A., Subramaniam, A., Tangsali, K., Fang, Z., Rietmann, M., Byeon, W., Choudhry, S., 2021. Nvidia simnet™: An ai-accelerated multi-physics simulation framework, in: International conference on computational science, Springer. pp. 447–461.

[13] Sun, R., Jeong, H., Zhao, J., Gou, Y., Sauret, E., Li, Z., Gu, Y., 2024. A physics-informed neural network framework for multi-physics coupling microfluidic problems. Computers & Fluids , 106421.

[14] Sun, Y., Lu, J., Liu, Q., Shuai, W., Sun, A., Zheng, N., Han, Y., Xiao, G., Xuan, J., Ni, M., et al., 2022. Multi-objective optimizations of solid oxide co-electrolysis with intermittent renewable power supply via multi-physics simulation and deep learning strategy. Energy conversion and management 258, 115560.

[15] Botache, D., Decke, J., Ripken, W., Dornipati, A., Götz-Hahn, F., Ayeb, M., Sick, B., 2024. Enhancing multi-objective optimisation through machine learning-supported multiphysics simulation, in: Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Springer. pp. 297–312.

[16] Laloy, E., Jacques, D., 2019. Emulation of cpu-demanding reactive transport models: a comparison of gaussian processes, polynomial chaos expansion, and deep neural networks. Computational Geosciences 23, 1193–1215.

[17] Demirer, E., Coene, E., Iraola, A., Nardi, A., Abarca, E., Idiart, A., de Paola, G., Rodríguez-Morillas, N., 2023. Improving the performance of reactive transport simulations using artificial neural networks. Transport in Porous Media 149, 271–297.
[18] Leal, A.M., Kyas, S., Kulik, D.A., Saar, M.O., 2020. Accelerating reactive transport modeling: on-demand machine learning algorithm for chemical equilibrium calculations. Transport in Porous Media 133, 161–204.
[19] Prasianakis, N.I., Haller, R., Mahrous, M., Poonoosamy, J., Pfingsten, W., Churakov, S.V., 2020. Neural network based process coupling and parameter upscaling in reactive transport simulations. Geochimica et Cosmochimica Acta 291, 126–143.
[20] Wang, Z., Battiato, I., 2024. A deep learning upscaling framework: Reactive transport and mineral precipitation in fracture-matrix systems. Advances in Water Resources 183, 104588.
[21] You, J., Lee, K.J., 2024. Upscaling reactive transport models from pore-scale to continuum-scale using deep learning method. Geoenergy Science and Engineering 238, 212850.
[22] Hansen, S.K., Pandey, S., Karra, S., Vesselinov, V.V., 2017. Chrotran 1.0: A mathematical and computational model for in situ heavy metal remediation in heterogeneous aquifers. Geoscientific Model Development 10, 4525–4538.
[23] K. Chen, X. Chen, J. C. Stegen, J. A. Villa, G. Bohrer, X. Song, K.-Y. Chang, M. Kaufman, X. Liang, Z. Guo, et al., Vertical hydrologic exchange flows control methane emissions from riverbed sediments, Environmental Science & Technology 57 (9) (2023) 4014–4026.
[24] J. Ren, B. Chen, W. Zhang, L. Men, J. Yang, Y. Li, Quantification of the temporal–spatial distributions characteristics of streambed hyporheic exchange fluxes with the seasonal variation using heat as a tracer, Environmental Earth Sciences 79 (2020) 1–19.
[25] H. Le Lay, Z. Thomas, F. Rouault, P. Pichelin, F. Moatar, Characterization of diffuse groundwater inflows into stream water (part ii: quantifying groundwater inflows by coupling fo-dts and vertical flow velocities), Water 11 (12) (2019) 2430.
[26] Yang, Z., Ding, M., Xu, B., Yang, H., Tang, J., 2022. Stam: A spatiotemporal aggregation method for graph neural network-based recommendation, in: Proceedings of the ACM Web Conference 2022, pp. 3217–3228.
[27] Gangopadhyay, T., Tan, S.Y., Jiang, Z., Meng, R., Sarkar, S., 2021. Spatiotemporal attention for multivariate time series prediction and interpretation, in: ICASSP 2021-2021 IEEE international conference on acoustics, speech and signal processing (ICASSP), IEEE. pp. 3560–3564.
[28] Chang, Z., Zhang, X., Wang, S., Ma, S., Gao, W., 2022. Stam: A spatiotemporal attention based memory for video prediction. IEEE Transactions on Multimedia 25, 2354–2367.
[29] Akiba, T., Sano, S., Yanase, T., Ohta, T., Koyama, M., 2019. Optuna: A next-generation hyperparameter optimization framework, in: Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining, pp. 2623–2631.
[30] Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., Hu, Q., 2020. Eca-net: Efficient channel attention for deep convolutional neural networks, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 11534–11542.
[31] Chansong, D., Supratid, S., 2021. Impacts of kernel size on different resized images in object recognition based on convolutional neural network, in: 2021 9th international electrical engineering congress (iEECON), IEEE. pp. 448–451.
[32] Georgescu, M.I., Ionescu, R.T., Miron, A.I., Savencu, O., Ristea, N.C., Verga, N., Khan, F.S., 2023. Multimodal multi-head convolutional attention with various kernel sizes for medical image super-resolution, in: Proceedings of the IEEE/CVF winter conference on applications of computer vision, pp. 2195–2205.
[33] Tan, M., Le, Q.V., 2019. Mixconv: Mixed depthwise convolutional kernels. arXiv preprint arXiv:1907.09595 .
[34] Liu, T., Chen, M., Zhou, M., Du, S.S., Zhou, E., Zhao, T., 2019. Towards understanding the importance of shortcut connections in residual networks. Advances in neural information processing systems 32.
[35] Drozdzal, M., Vorontsov, E., Chartrand, G., Kadoury, S., Pal, C., 2016. The importance of skip connections in biomedical image segmentation, in: International workshop on deep learning in medical image analysis, international workshop on large-scale annotation of biomedical data and expert label synthesis, Springer. pp. 179–187.

[36] Wei, C., Kakade, S., Ma, T., 2020. The implicit and explicit regularization effects of dropout, in: International conference on machine learning, PMLR. pp. 10181–10192.
[37] Woo, S., Park, J., Lee, J.Y., Kweon, I.S., 2018. Cbam: Convolutional block attention module, in: Proceedings of the European conference on computer vision (ECCV), pp. 3–19.
[38] Lasseux, D., Valdés-Parada, F.J., Wood, B.D., 2021. Recent developments in upscaling and characterization of flow and transport in porous media.
[39] Szymkiewicz, A., 2013. Upscaling from darcy scale to field scale. Modelling Water Flow in Unsaturated Porous Media: Accounting for Nonlinear Permeability and Material Heterogeneity , 139–175.
[40] Zhang, X., Ma, F., Yin, S., Wallace, C.D., Soltanian, M.R., Dai, Z., Ritzi, R.W., Ma, Z., Zhan, C., Lü, X., 2021. Application of upscaling methods for fluid flow and mass transport in multi-scale heterogeneous media: A critical review. Applied energy 303, 117603.
[41] Horgue, P., Guibert, R., Gross, H., Creux, P., Debenest, G., 2015. Efficiency of a two-step upscaling method for permeability evaluation at darcy and pore scales. Computational Geosciences 19, 1159–1169.

Chapter 7: Conclusion and Synthesis of Findings

7.1 Introduction

This dissertation presents a comprehensive exploration of microbe-mediate reactive transport with a focus on microbial motility, particle tracking, and biomass growth in the hyporheic zone. Through a combination of experimental observations, physics-based computational modeling, and deep learning approaches, this work advances our understanding of complex biogeochemical interactions across multiple scales. Furthermore, this work addresses critical gaps in our ability to model and predict reactive transport processes in heterogeneous subsurface environments. Ultimately, each chapter of this dissertation may be used to develop more robust bioremediation models and experiments.

7.2 Synthesis of Key Findings and Advancements

7.2.1 Advancing Microbial Transport Understanding

The investigation of advection-dominated transport dynamics of pili and flagella-mediated motile bacteria in porous media (Chapter 2) provides fundamental insights into how different bacterial species navigate complex flow environments. This work builds upon previous studies of bacterial motility in porous media [1-11] by explicitly comparing the behavior of different motility mechanisms under varying flow conditions. The research reveals a critical transition to advection-dominated transport as flow rates increase, with bacteria exhibiting peritrichous flagella (e.g., Paenibacillus) demonstrating superior maintenance of their motility characteristics at higher flow rates compared to bacteria with pili or monotrichous flagella. Furthermore, this chapter challenges previous assumptions about bacterial behavior in pore networks by discovering that

motile bacteria tend to oversample medium-velocity regions, contrary to earlier notions of preferential sampling in low-velocity zones [3].

These findings carry significant implications for understanding microbial transport in natural and engineered porous media systems. The observation of differential transport behavior among bacterial species with varying motility mechanisms suggests that models of microbial transport in the subsurface must account for species-specific motility characteristics. This consideration becomes particularly crucial in bioremediation applications, where the transport and distribution of specific bacterial strains may significantly impact treatment efficacy [4]. Additionally, the identification of advection-dominated transport regimes provides valuable insights into conditions where simplified transport models may suffice. In high-flow scenarios, where bacterial motility plays a diminished role, the dispersion coefficient will be almost entirely a function of the flow speed and porous geometry. In low-flow regimes, the dispersion coefficient will have less dependence on speed, similar dependence on porous geometry, and much greater dependence on motility. Thus, the work from this chapter indicates that at low speeds the dispersion coefficient should also be calculated as a function of motility type.

7.2.2 Enhancing Particle Tracking Methodologies

The evaluation of particle tracking codes for dispersing particles in porous media (Chapter 3) and the development of DeepTrackStat (Chapter 4) represent significant advancements in our ability to analyze and interpret experimental data on particle and microbial transport. Through rigorous analysis, the work of chapter 3 demonstrates that tracking algorithm performance for dispersing particles is largely dependent on the particle spacing displacement ratio (PSDR), with all methods exhibiting decreased accuracy at low PSDR values. The research also revealed that traditional metrics [12] for evaluating tracking performance may significantly underestimate errors in certain

scenarios, particularly when employing aggressive linking algorithms. Lastly, this chapter illustrates some of the errors that may have impacted the results from chapter 2, such as the tendency to miss high-speed particles and therefore underpredict the max speeds and mean square displacement (MSD). As such, this finding has important implications for the design and interpretation of microfluidics experiments.

The limitations identified in existing algorithms motivated the development of DeepTrackStat, a novel deep learning framework for extracting motion statistics from particle tracking videos. This innovative approach offers substantial advantages over traditional particle tracking methods [13-17], demonstrating improved accuracy in extracting speed, velocity, and turn angle distributions, particularly in challenging scenarios involving high-speed and high-density particle movements. Moreover, DeepTrackStat achieves significantly reduced computation time compared to classical particle tracking methods, enabling the analysis of larger datasets while maintaining high precision. Together, Chapters 3 and 4 provide essential tools and knowledge for improving the extraction of useful information from microfluidic particle tracking experiments, a crucial task for validating and refining computational models of microbial transport and reactive processes in porous media [2, 7, 9-11].

7.2.3 Improvements in Reactive Transport Modeling

While each chapter of this dissertation contributes to the advancement of reactive transport modeling, the contributions manifest in different ways across scales and applications. Chapter 2's improved understanding of bacterial motility for species with different motility types indicates potential new avenues for the refinement of bacterial transport models. The advancements in particle tracking accuracy and methodology presented in Chapters 3 and 4 enable more precise

parameter estimation for bacterial transport models, ultimately enhancing the accuracy of reactive transport simulations through improved foundational measurements.

Chapters 5 and 6 represent the most direct applications to reactive transport modeling. Chapter 5 introduces an experimentally-calibrated physics-based augmentation to PFLOTRAN that incorporates speed-based decay and temperature-based growth of biomass. Through comprehensive sensitivity analysis, this work demonstrated that speed-based biomass decay significantly impacts biofilm development only under specific conditions: when fluid speeds exceed 10 meters/day or in cases of "weakly cohesive" biofilms with exponential decay parameters below 0.4. Chapter 6 presents STAMNet, a neural framework for upscaling reactive transport simulations, which addresses aspects of the long-standing problem of computational intensity in reactive transport simulations. This innovative approach demonstrates the potential for deep learning techniques to bridge the gap between detailed process understanding and practical computational requirements.

7.2.4 Bridging Scales in Reactive Transport Modeling

A fundamental contribution of this dissertation lies in its comprehensive approach to bridging scales in reactive transport processes, from individual bacterial trajectories to field-scale reactive transport. The insights gained from microbial motility studies provide a foundation for understanding how individual bacterial behaviors translate to larger-scale transport patterns. While these microscale findings were not directly incorporated into the larger-scale models presented in later chapters, they inform our conceptual understanding of how microscale processes may influence macroscale outcomes.

The advancements in particle tracking methodologies serve as a crucial bridge between experimental observations and computational models. By improving our ability to extract accurate

motion statistics from experimental data, these tools enable better validation and refinement of both pore-scale and continuum-scale transport models. The broad applicability of these tools, particularly DeepTrackStat, extends beyond their original design for microfluidics studies. Through training on diverse simulations, these tools demonstrate utility for analyzing particles under various motion regimes across multiple scales.

The investigation of hyporheic zone processes integrates understanding from multiple scales by considering how pore-scale flow dynamics and biomass growth processes manifest in Darcy-scale reactive transport behavior. The velocity-based biomass decay model represents a particularly significant achievement in incorporating microscale biofilm dynamics [18] into continuum-scale reactive transport simulations. STAMNet further advances this multi-scale integration by directly addressing the challenge of upscaling reactive transport simulations, demonstrating how machine learning techniques can effectively capture sub-grid scale processes in large-scale predictions without imposing overwhelming computational demands.

7.2.5 Implications for Environmental Management and Bioremediation

The findings presented in this dissertation have substantial implications for environmental management and bioremediation strategies. The observation of differential transport characteristics among bacterial species with varying motility mechanisms suggests that bioremediation strategies must carefully consider the specific motility traits of target microorganisms when selecting bacterial strains for bioaugmentation or designing flow conditions to optimize the distribution of beneficial microorganisms in contaminated aquifers.

The development of DeepTrackStat provides environmental managers and researchers with a powerful tool for analyzing visual data of moving particles for a variety of scales and flow conditions, enabling more accurate characterization of contaminant transport behavior and more

robust modeling of bioremediation processes. Furthermore, STAMNet's capability for rapid large-scale predictions enables comprehensive Monte Carlo and sensitivity analyses that can support more informed decision-making in scenarios such as assessing contaminant spread in groundwater systems under different scenarios, evaluating long-term remediation effectiveness, and predicting climate change impacts on subsurface biogeochemical cycles.

The insights gained from the hyporheic zone study reveal that speed-based biomass decay significantly impacts remediation outcomes only under specific conditions of high fluid speed or weak biomass development. This understanding can inform the optimization of nutrient delivery and bioremediation strategies based on the identified feedback mechanisms between flow, nutrient transport, contaminant degradation, and biomass growth. Additionally, the novel insights into the relationship between biotic and abiotic reduction demonstrate that while abiotic reduction generally dominates in high-electron-donor environments, biotic reduction plays a crucial role in determining the spatial distribution of remediation hotspots.

7.3 Limitations and Technical Challenges

The experimental and computational approaches developed in this dissertation, while advancing our understanding of reactive transport processes, face several important limitations. The bacterial motility studies, though revealing important relationships between motility mechanisms and transport behavior, were necessarily limited to a subset of bacterial species and environmental conditions. A more robust study would investigate transport for a wider variety of flow rates, and the simplified porous media geometries used in experiments, while providing valuable insights, cannot fully replicate the complexity of natural sediments.

Technical limitations in particle tracking methodologies persist despite the advances made through DeepTrackStat. The fundamental trade-off between tracking accuracy and computational efficiency remains a challenge, particularly when dealing with high particle densities or rapid motion. While DTS successfully addresses many limitations of traditional tracking approaches, its reliance on training data means that performance may degrade when encountering particle behaviors significantly different from those represented in the training set. In addition, DTS is only designed to extract speed, velocity component and turn angle statistics, meaning researchers who want to derive more complex statistics from the raw trajectories will still need to resort classical particle tracking methods.

The PFLOTRAN augmentation faces limitations related to the simplifying assumptions inherent in continuum-scale modeling. The speed-based decay model, while incorporating important microscale processes, necessarily homogenizes complex spatial heterogeneities in biofilm structure and bacterial behavior. The identified threshold values for significant speed-based decay effects may vary in natural systems with more complex geometry and chemical conditions. Furthermore, the model's treatment of biofilm strength as a uniform parameter may not adequately capture the spatial and temporal variability in biofilm properties observed in natural systems. Finally, the equation used to model speed-based biomass decay used fitting parameters for calibration, but a more robust method would derive this equation from fundamental physical principles such as the continuity equation and/or conservation of mass.

STAMNet's current implementation, while demonstrating significant potential for upscaling reactive transport simulations, faces constraints in its generalizability. The framework's fixed input and output dimensions limit its direct application to different spatial scales or problem geometries. The requirement for consistent tensor shapes during training creates challenges for

developing truly scale-agnostic models. Additionally, computational resources remain a limiting factor across multiple aspects of this work, particularly in the generation of training data through high-resolution reactive transport simulations and the validation of model predictions against experimental data.

7.4 Future Research Directions

Several promising avenues for future research emerge from the findings and limitations identified in this dissertation. Future studies of bacterial motility should expand to investigate a broader range of bacterial species under varying environmental conditions. A particularly promising approach would involve the development of experimental systems using genetically engineered bacterial strains that differ only in motility mechanisms. Such controlled comparisons would provide crucial insights for developing more accurate transport models while eliminating confounding variables present in cross-species comparisons.

The advancement of particle tracking capabilities requires development of more sophisticated deep learning architectures that can maintain accuracy while reducing computational demands. Future iterations of DeepTrackStat could incorporate adaptive training approaches to handle a broader range of particle behaviors and experimental conditions. Extension to three-dimensional tracking applications represents a particularly important direction, as many natural systems exhibit complex three-dimensional flow patterns and bacterial behaviors that cannot be fully captured in planar analysis.

The reactive transport modeling framework could be enhanced through incorporation of additional biogeochemical processes and improved representation of spatial heterogeneity. Future work should focus on developing more sophisticated approaches for representing biofilm

mechanical properties and their spatial variation within continuum-scale models. Integration of advanced imaging techniques with reactive transport modeling could enable better characterization of biofilm spatial structure and its influence on local flow fields and reaction rates. STAMNet's capabilities could be significantly expanded through architectural modifications enabling true scale and feature agnosticism. Development of training strategies that can handle variable tensor shapes, potentially through implementation of adaptive neural network architectures or novel batching approaches, would greatly enhance the framework's applicability. Creation of a feature-agnostic version would require extensive training on diverse reactive transport scenarios, potentially utilizing transfer learning approaches to efficiently capture common patterns across different chemical systems.

Integration of these various research directions with emerging technologies represents a particularly promising avenue for advancement. Real-time sensing networks could provide validation data for model predictions while enabling adaptive optimization of bioremediation strategies. Climate change impacts on subsurface biogeochemical processes could be better understood through application of these modeling frameworks to scenarios incorporating projected environmental changes.

7.5 Concluding Remarks

This dissertation advances our understanding of reactive transport processes across multiple scales through the integration of experimental observations, computational modeling, and innovative deep learning techniques. The multifaceted approach developed here spans from microscale observations of bacterial motility to field-scale predictions of biogeochemical processes, providing a comprehensive framework for improving predictions of microbe-mediated reactive transport in heterogeneous porous media.

The development of novel tools such as DeepTrackStat and STAMNet, coupled with mechanistic insights into microbial transport and hyporheic zone processes, provides a robust framework for improving predictions of reactive transport in heterogeneous porous media. These advancements have significant implications for environmental management and bioremediation, offering the potential for more informed decision-making and optimized remediation strategies. The challenges and opportunities identified point toward an increasingly integrated approach to understanding and modeling reactive transport processes. Future advances will likely emerge from the confluence of high-resolution experimental techniques, sophisticated data analysis methods, and innovative modeling frameworks. Success in these endeavors will require continued collaboration across disciplines, from molecular biology to computer science, as well as sustained investment in both experimental and computational infrastructure.

As we continue to face global challenges related to water quality and ecosystem health, the approaches developed in this dissertation offer promising avenues for advancing our understanding and stewardship of critical environmental interfaces. By bridging scales and integrating diverse scientific approaches, this work lays a foundation for more comprehensive and predictive models of coupled hydrological, geochemical, and biological processes in complex environmental systems.

References

[1] Datta, S. S., Steinberg, A. P., & Ismagilov, R. F. (2016). Polymers in the gut compress the colonic mucus hydrogel. Proceedings of the National Academy of Sciences, 113(26), 7041-7046.

[2] Bhattacharjee, T., & Datta, S. S. (2019). Bacterial hopping and trapping in porous media. Nature Communications, 10(1), 2075.

[3] Rusconi, R., Guasto, J. S., & Stocker, R. (2014). Bacterial transport suppressed by fluid shear. Nature Physics, 10(3), 212-217.

[4] Ginn, T. R., Wood, B. D., Nelson, K. E., Scheibe, T. D., Murphy, E. M., & Clement, T. P. (2002). Processes in microbial transport in the natural subsurface. Advances in Water Resources, 25(8-12), 1017-1042.

[5] Tufenkji, N. (2007). Modeling microbial transport in porous media: Traditional approaches and recent developments. Advances in Water Resources, 30(6-7), 1455-1469.

[6] Scheidweiler, D., Miele, F., Peter, H., Battin, T. J. & De Anna, P. Trait-specific dispersal of bacteria in heterogeneous porous environments: from pore to porous medium scale. *J. R. Soc. Interface.* **17**, 20200046 (2020).

[7] Perez, L. J., Bhattacharjee, T., Datta, S. S., Parashar, R. & Sund, N. L. Impact of confined geometries on hopping and trapping of motile bacteria in porous media. *Phys. Rev. E* **103**, 012611 (2021).

[8] Tokárová, V., Sudalaiyadum Perumal, A., Nayak, M., Shum, H., Kašpar, O., Rajendran, K., Mohammadi, M., Tremblay, C., Gaffney, E.A., Martel, S. & Nicolau Jr, D.V. Patterns of bacterial motility in microfluidics-confining environments. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2013925118 (2021).

[9] Creppy, A., Clément, E., Douarche, C., D'Angelo, M. V. & Auradou, H. Effect of motility on the transport of bacteria populations through a porous medium. *Phys. Rev. Fluids* **4**, 013102 (2019).

[10] Secchi, E., Vitale, A., Miño, G.L., Kantsler, V., Eberl, L., Rusconi, R. & Stocker, R. The effect of flow on swimming bacteria controls the initial colonization of curved surfaces. *Nat Commun* **11**, 2851 (2020).

[11] Dentz, M., Creppy, A., Douarche, C., Clément, E. & Auradou, H. Dispersion of motile bacteria in a porous medium. *J. Fluid Mech.* **946**, A33 (2022).

[12] Chenouard, N. et al. Objective comparison of particle tracking methods. Nat. methods 11(3), 281–289 (2014).

[13] Tinevez, J. et al. Trackmate: An open and extensible platform for single-particle tracking. Methods 115, 80–90 (2017).

[14] Ershov, D. et al. Trackmate 7: integrating state-of-the-art segmentation algorithms into tracking pipelines. Nat. Methods 19(7), 829–832 (2022).

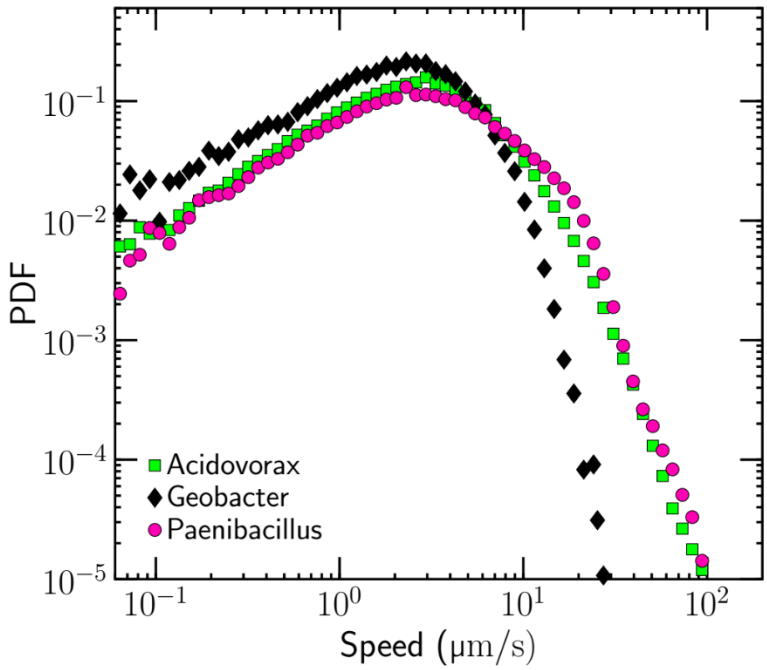
[15] Allan, D., Caswell, T., Keim, N., van der Wel, C. & Verweij, R. soft-matter/trackpy: Trackpy v0. 5.0., DOI: https://doi.org/10.5281/zenodo.4682814 (2021).

[16] Heyman, J. TracTrac: A fast multi-object tracking algorithm for motion estimation. Computers & Geosciences, 128, 11-18 (2019).

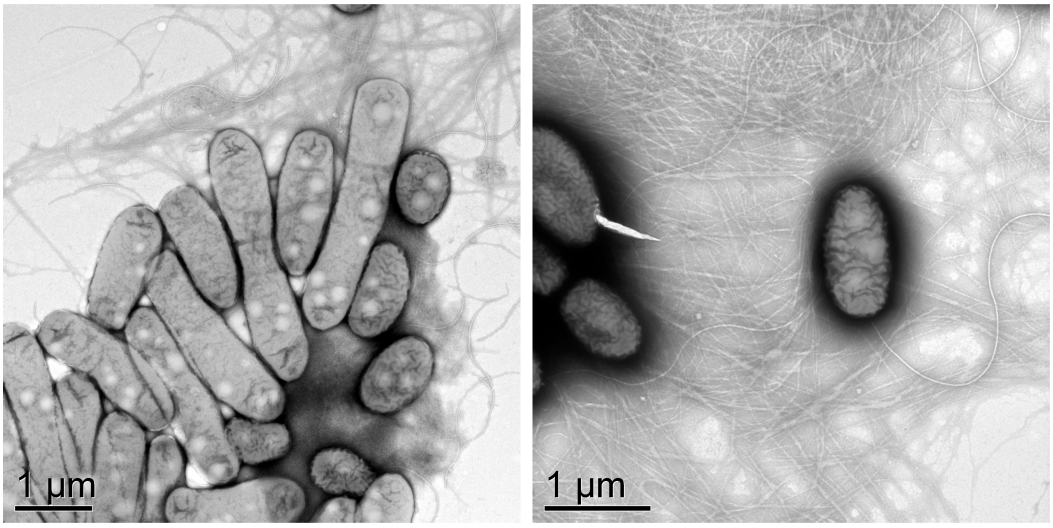
[17] Fukai, Y.T. and Kawaguchi, K. LapTrack: linear assignment particle tracking with tunable metrics. Bioinformatics, 39(1), 799 (2023).

[18] G. Wei, J. Q. Yang, Impacts of hydrodynamic conditions and microscale surface roughness on the critical shear stress to develop and thickness of early-stage pseudomonas putida biofilms, Biotech- nology and bioengineering 120 (7) (2023) 1797–1808.

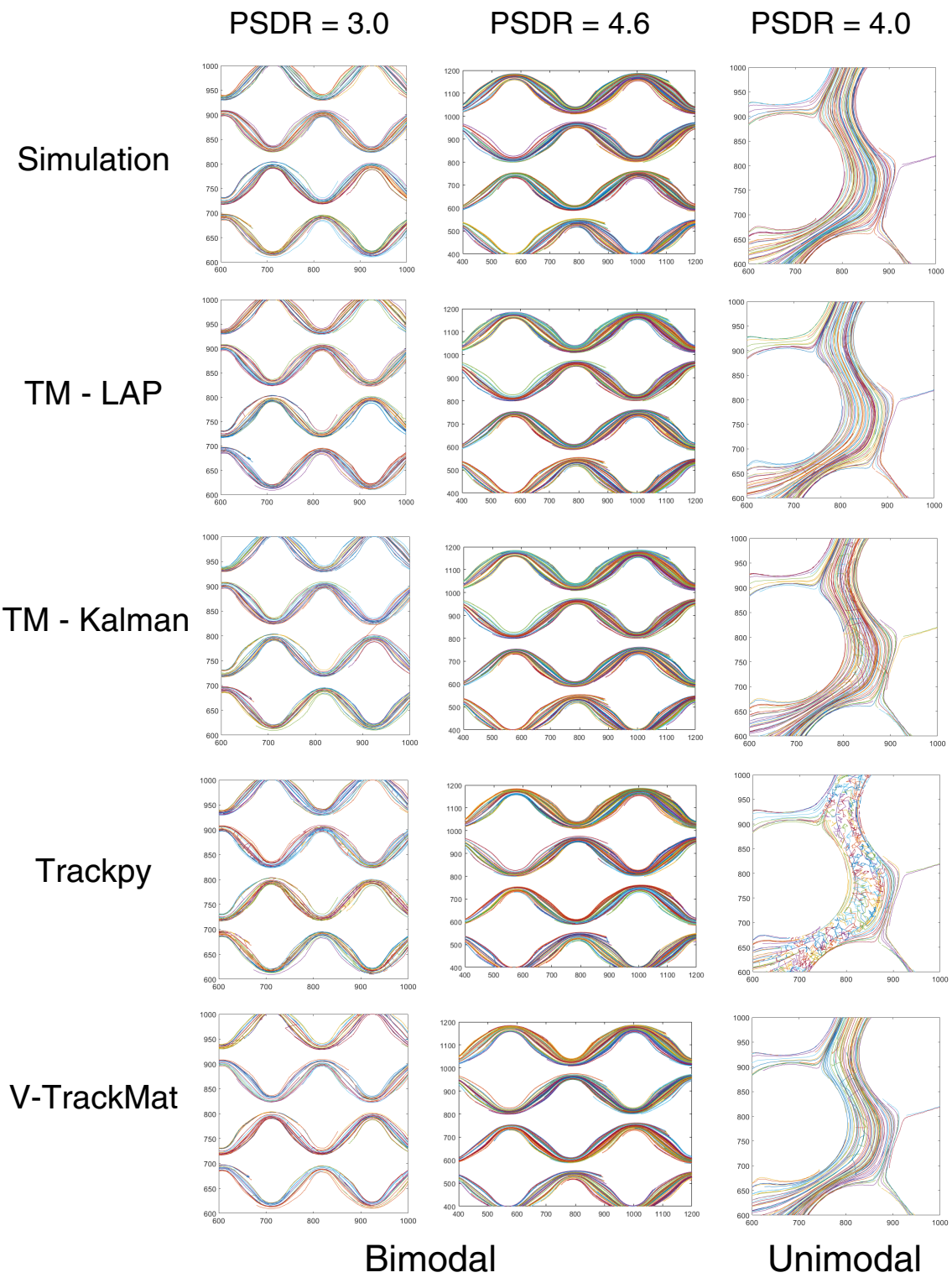
Supplementary Figures



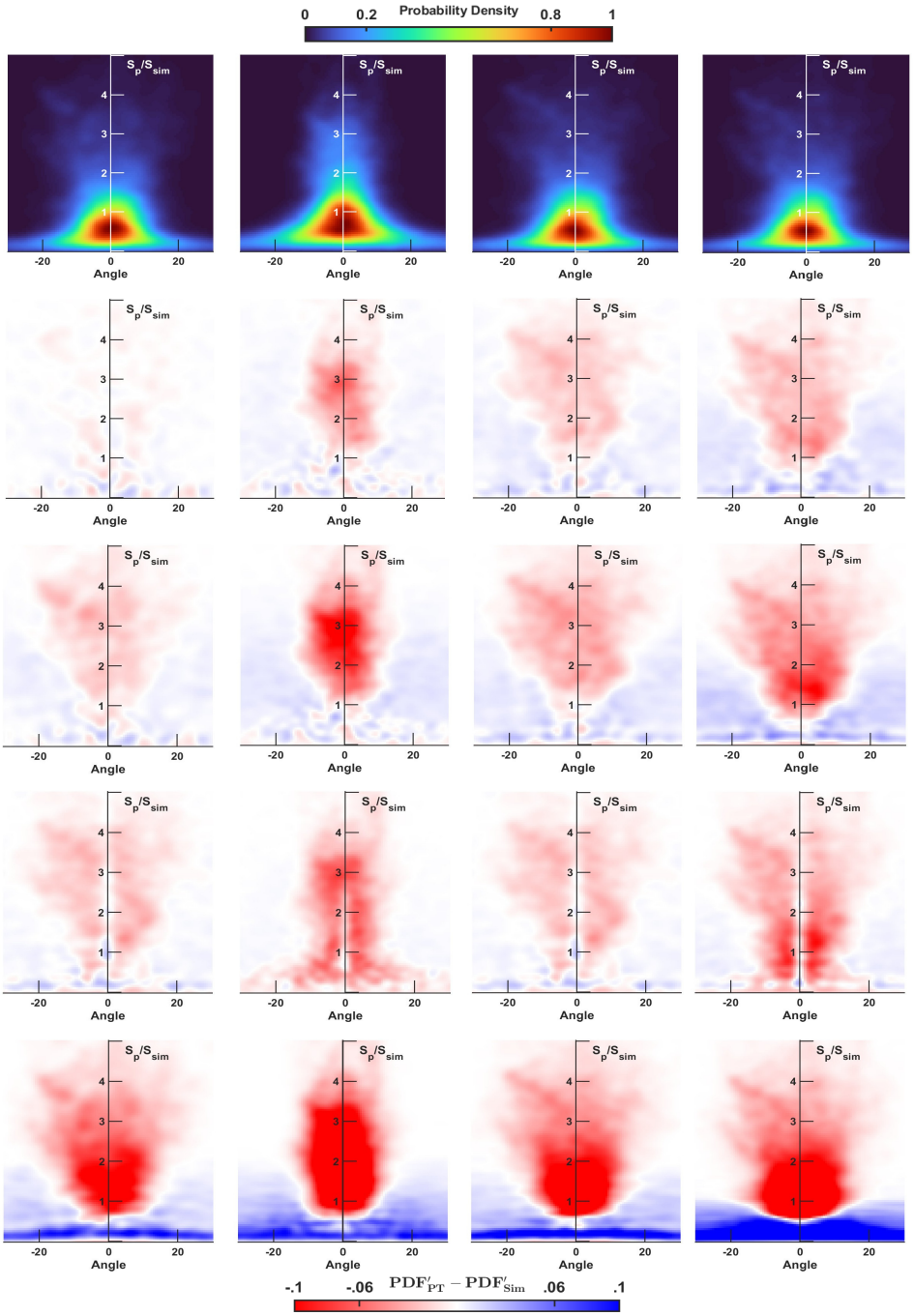
Supplementary Figure 1. Speed PDF for no-flow experiments in the high porosity geometry (grain diameter = 40 mm, pore length = 20 mm).



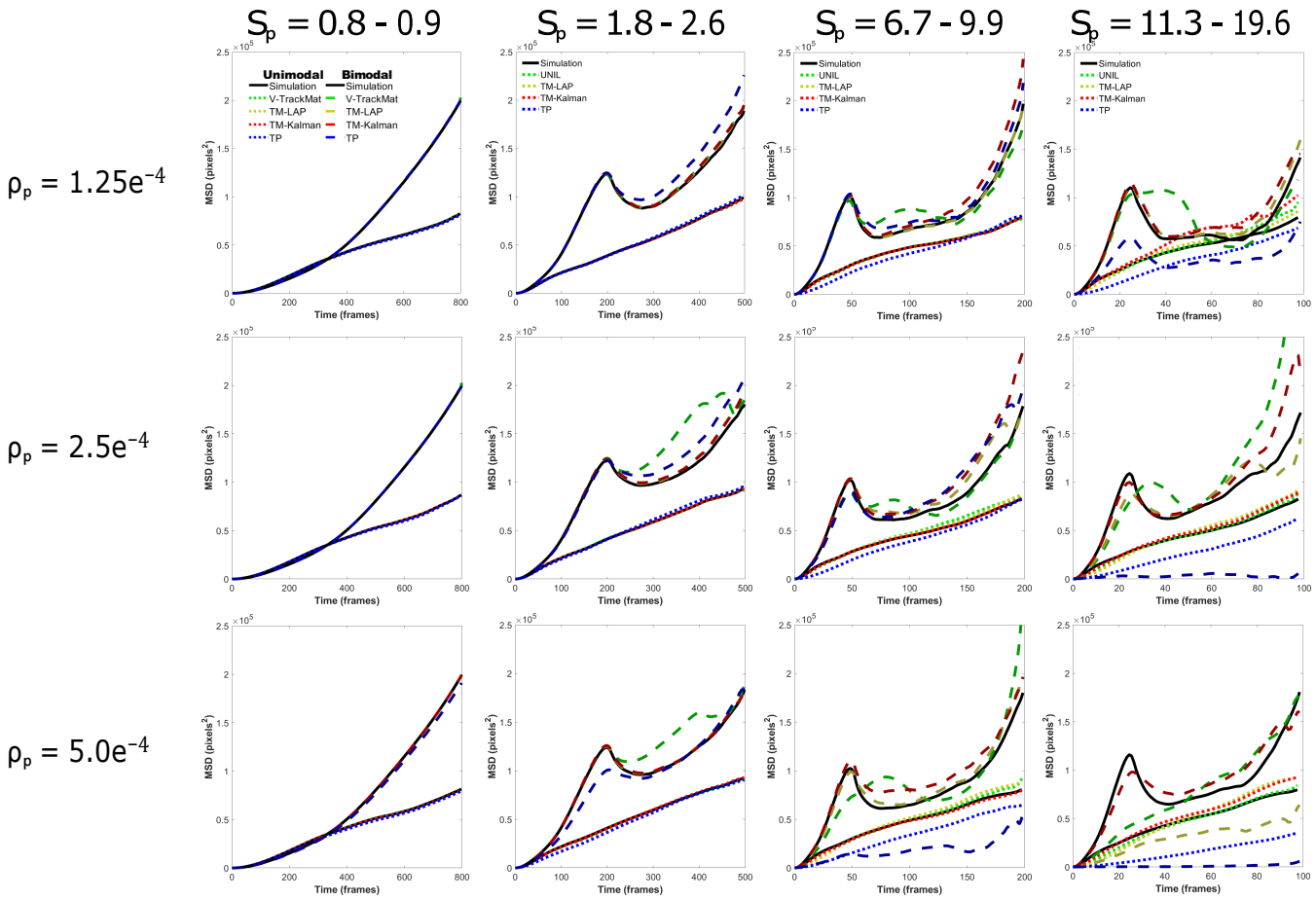
Supplementary Figure 2. Whole-mount transmission electron microscopy (TEM) images of *Acidovorax* JHL-9 (unpublished images from [22], courtesy of Alice Dohnalkova). The whole-mount images were prepared by adding JHL-9 liquid culture to a copper electron microscopy grid and examining by TEM at 200 kV using a JEOL 2010 high-resolution TEM.



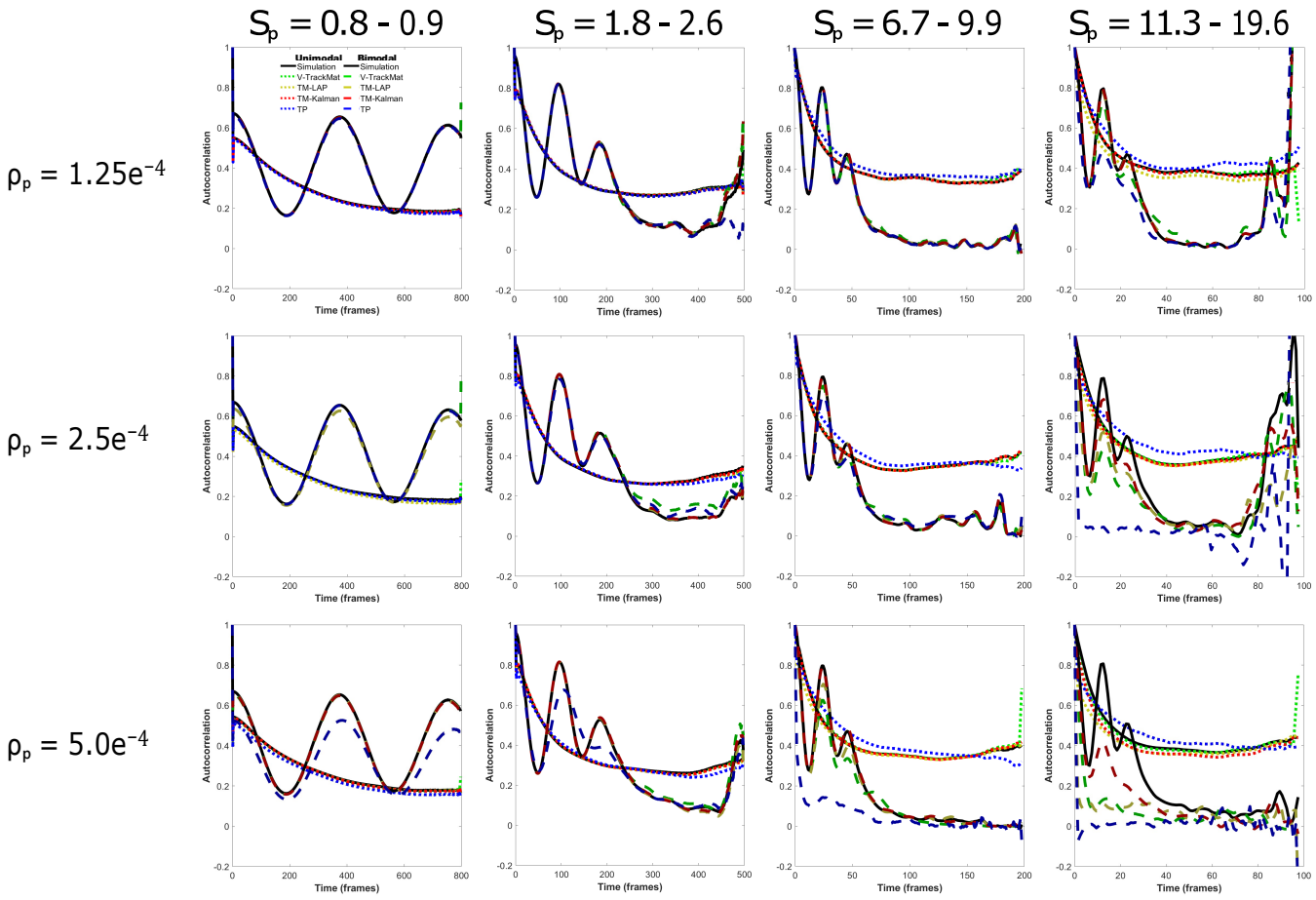
Supplementary Figure 3. Sample trajectories for all PT codes for selected high-PSDR bimodal and unimodal simulations. Each line corresponds to a unique trajectory (with random colors used to show the contrast between individual trajectories). For these high-PSDR simulations, all PT codes besides TP show near-perfect replication of the ground truth trajectories.



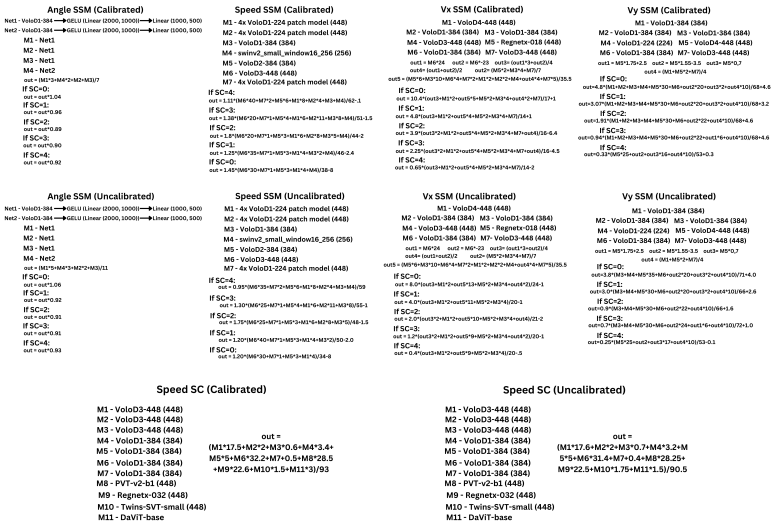
Supplementary Figure 4. Speed-angle joint probability density difference heatmaps for the unimodal simulation. Speeds determined from particle tracking (S_p) are normalized by the mean speed of the respective simulation (S_{sim}). Red corresponds to an underprediction of probability density, blue corresponds to an overprediction of probability density, and white corresponds to an accurate probability density prediction within the speed-angle feature space. These results show strong performance for TM-Kalman and V-TrackMat, slightly weaker performance for TM-LAP, and bad performance for TP.



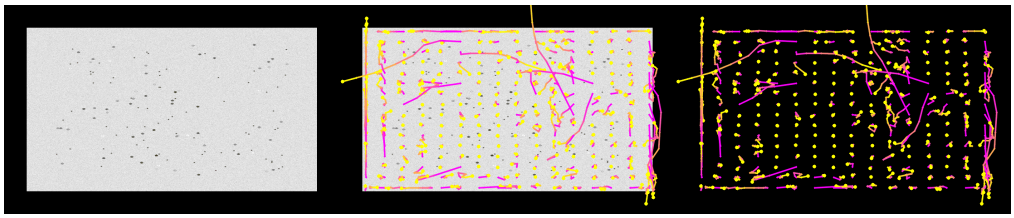
Supplementary Figure 5. All MSDs. Simulated particle speeds increase going from left to right, and particle densities increase going from top to bottom. The bottom right corner represents the lowest-PSDR simulation. The speed values for each column are given as a range because the unimodal simulations always have lower mean speeds than the bimodal simulations. The bimodal MSDs are shown by dashed lines, and the unimodal MSDs are shown by dotted lines. These figures confirm trends present in the other results - TP performs the worst, and other algorithms start to fail at $S_p \geq 11.3$.



Supplementary Figure 6. All C_v. Simulation speed increases from left to right, and particle density increases from top to bottom. The bottom right corner represents the lowest-PSDR simulation. Autocorrelations from the heterogenous simulations are dotted and slightly brighter colored than the bimodal simulations. The bimodal simulations show a periodicity in autocorrelation that is generated due to the periodic converging and diverging flowpaths of the bimodal geometry. The unimodal simulations show a clear trend of decorrelation over time. All PT codes show better performance for the unimodal simulations, although this is potentially due to differences in the mean speeds of the simulations. In terms of PT performance, these results follow the same general trends as the MSDs.



Supplementary Figure 7. Ensemble weightings for the calibrated and uncalibrated SSMs and SC. Each SSM is comprised of several models and the specific ensemble used to make a prediction is determined by the output of the SC.



Supplementary Figure 8. Results from PIPs (https://github.com/aharley/pips). Some amount of tracking seems to occur, but it is not useful for the task of motion statistic extraction. Further research should seek to fine tune this method (and other similar optical flow methods) to see if the models offer any underlying advantages for transfer learning.

Appendix

A1 – Pseudocode for simple reactive transport simulator

```
## Initialize variables and parameters
## Initialize grid dimensions: nx, ny, nz
## Initialize time steps: nt, dt
## Initialize spatial steps: dx, dy, dz
## Initialize arrays: C[nx][ny][nz], θ[nx][ny][nz], v[nx][ny][nz], K[nx][ny][nz], R[nx][ny][nz]
## Where C is the concentration of the species, θ is the water content, v is the pore water velocity, K is the ## hydraulic conductivity, and R is a reaction source/sink term that represents the chemical reactions with  ## other species
## Set initial/boundary conditions for C, theta, v

## Main time-stepping loop
for t = 1 to nt:
  ## Solve Richards equation for water content
  for x = 1 to nx:
    for y = 1 to ny:
      for z = 1 to nz:
        θ[x][y][z] = solve_richards_equation(θ[x][y][z], K(θ))
        update_hydraulic_conductivity(θ[x][y][z])

  ## Compute velocities using Darcy's law
  for x = 1 to nx:
    for y = 1 to ny:
      for z = 1 to nz:
        v[x][y][z] = compute_darcy_velocity(θ[x][y][z], ...)

  ## Solve Advection-Dispersion-Reaction Equation for concentration
  for x = 1 to nx:
    for y = 1 to ny:
      for z = 1 to nz:
        C[x][y][z] = solve_ADE(C[x][y][z], v[x][y][z], θ[x][y][z], R[x][y][z], D, …)

  ## Apply boundary conditions
  apply_boundary_conditions(C, θ, v)

  ## Output results at specified intervals
  if t % output_interval == 0:
```

output_results(C, θ, v, t)

End of simulation
finalize_output()

Helper functions (to be implemented separately)
function solve_richards_equation(θ[x][y][z], K(θ), $\frac{\partial \mathrm{K}(\theta)}{\partial \mathrm{z}}$, D):
  ## Solve Richards' equation based on water content, hydraulic conductivity, and dispersion

function update_hydraulic_conductivity(θ):
  ## Update hydraulic conductivity based on water content

function compute_darcy_velocity(ϕ, K, $\frac{dh}{dl}$):
  ## Compute velocity using Darcy's law

function solve_ADE(C, v, θ, R, D):
  ## Compute ADE as a function of concentration, pore water velocity, water content, chemical reactions, and dispersion

function apply_boundary_conditions(C, θ, v, ϕ, K):
  ## Apply appropriate boundary conditions

function output_results(C, θ, v, t, ϕ, K):
  ## Output or save results at the current time step