The core of my research is data-driven mathematical modeling, which I typically use to study biological problems through interdisciplinary collaboration. My ongoing and future research projects focus on the development, analysis, and application of methods to enable robust parameter estimation, model selection, and uncertainty quantification directly from biological data [1, 2, 3, 4]. I combine concepts from machine learning, data science, and topology to aid in the model development process [5, 6, 7, 8]. This work stems from my doctoral and postdoctoral research on the derivation and analysis of partial differential equation (PDE) models from both biological data and stochastic agent-based models (ABMs). My research involves extensive interdisciplinary collaboration, and some of my ongoing efforts include work with:

- Chemical engineers to ascertain how individual interactions in a mathematical model scale to emergent collective phenomena,
- Topologists to automatically infer the characteristic morphologies of healthy and cancerous blood vessel networks, and
- Neuro-oncologists to determine optimal experimental design methods for equation learning algorithms.

A common factor in these biological problems is *collective behavior*, or the emergent population-wide patterns that arise from many interacting individuals. The biological questions that I investigate include how individual interactions scale to healthy or impaired wound healing dynamics [3, 7, 9, 10], how to distinguish different blood vessel network morphologies, and how to infer the individual interactions underlying biological flocking [5]. While this work has been designed with specific biological applications in mind, the ubiquity of collective behavior ensures that the resulting models and methods that I have developed are broadly applicable to many biological fields, ranging from microscales such as biomechanics and cellular biology to macroscales that include ecological foraging.

Below, I summarize some of my key research studies towards this problem as well as plans for future work. These areas of my research include (A) Mathematical Modeling of Cell Migration during Wound Healing, (B) Topological Data Analysis of Biological Phenomena, and (C) Discovering Dynamical System Models with Machine Learning .

(A) Mathematical Modeling of Cell Migration During Wound Healing

Background: Reaction-diffusion PDE models are frequently used to model the wound healing process to elucidate how cell populations collectively migrate into the wound space. A challenge for mathematicians is to develop models that capture cell interactions and can describe experimental data. *Some current wound healing dynamics that are poorly understood include how cells interact during with their neighbors and methods to quantify the biochemical regulation of collective migration.* My research aim is to inform our understanding of wound healing dynamics as a means to elucidate what goes wrong in nonhealing wounds, which cost the US Health Care system $15 billion annually.

I collaborated with experimentalists from the Department of Biochemistry at the University of Colorado, Boulder to study how cells interact during the wound healing process. Figure 1 depicts three snapshots of a scratch assay experiment conducted to mimic human

wound healing. Briefly, a sheet of human skin cells is grown to confluence. Half of this sheet is then scratched away, and the remaining cells migrate rightwards into the denuded area. Human skin cells maintain physical connections to their neighbors during these experiments, and how cells use these connections to interact with their neighbors is poorly understood.

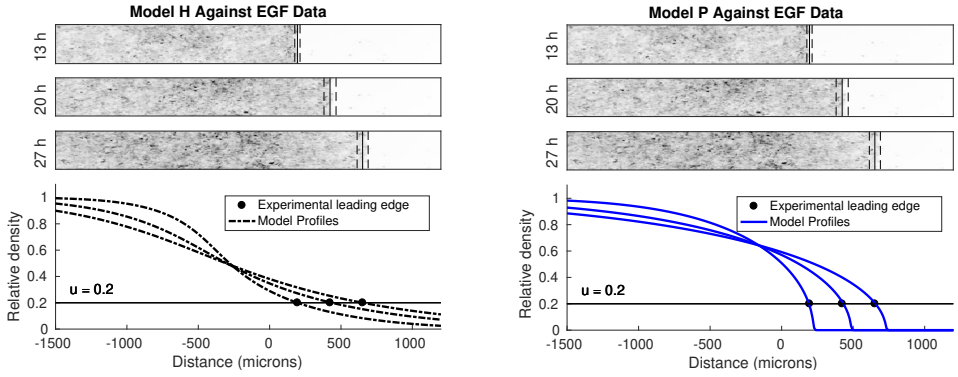

Figure 1: Experimental wound healing snapshots against best-fit profiles for Models H (left) and P (right) at $t = 13, 20$, and 27 hours.

It is challenging to develop models of wound healing that can describe experimental data due to the complexity of this process. I derived two multiscale PDE models that considered two plausible hypotheses on how cells interact with their neighbors during the wound healing [9]. The first model (Model H) assumes cells use physical connections to *hinder* the migration of others into the wound area, and the second model (Model P) assumes cells use physical connections to *promote* the migration of others into the wound area. Model P (solid blue lines) better matches the experimental cell density than Model H (black dashed lines) in Figure 1, which suggests that cells use physical connections to promote the migration of their neighbors. We validated this hypothesis experimentally and concluded that human skin cells use physical connections to promote the migration of their neighbors during wound healing. *This study was the first to derive a PDE model that matches experimental data while considering cell-cell interactions.*

Biologists understand that internal biochemical pathways influence cell speed during wound healing, but there is no accepted mathematical method to quantify the influence of biochemical activation on migration patterns. Towards this end, I developed and analyzed a biochemically-structured reaction-diffusion equation given by

$$\begin{aligned} \frac{\partial u}{\partial t} + \underbrace{\frac{\partial}{\partial m}\left[f(t)g(m)u\right]}_{activation} &= \underbrace{D(m)\frac{\partial^2 u}{\partial x^2}}_{diffusion} + \underbrace{\lambda(m)u\left(1 - w(t,x)\right)}_{population\ growth} \qquad (1) \\ w(t,x) &= \int_{m_0}^{m_1} u(t,x,m)dm. \end{aligned}$$

where $u(t, x, m)$ is a structured cell density with activation level m at the spatial dimension x over time t, $f(t)$ represents biochemical activation in response to a chemical, $g(m)$ provides nonlinear rates of biochemical activation, $D(m)$ denotes activation-dependent diffusion, $\lambda(m)$ denotes proliferation, and $w(t, x)$ represents cell density in time and space.

In some regimes, I proved that this equation admits unique self-similar traveling wave solutions characterizing how populations simultaneously migrate in space and activate along the biochemical dimension [10]. To analyze the full equation, I combined numerical results with analytical results to show that activating the entire population leads to the furthest-migrating simulations. *These results present novel methods to analyze and interpret structured PDE models which provides an avenue into richer dynamics capturing the interaction between biochemical pathways and population-wide migration rates during wound healing.*

Future work: Haptotaxis refers to the directed migration of cells towards adhesion sites on the ground. This process guides many steps of wound healing, but has not been mathematically explored due to difficulties in creating gradients of adhesion molecules. We are collaborating with the Haugh Lab (NCSU Chemical Engineering) to derive agent-based models (ABMs) of haptotaxis. We have developed algorithms to track cell locations over time and are now comparing ABM simulations to this data to understand the individual cell mechanisms that lead to healthy wound healing. *We have applied for a $1.2 million Joint NSF/NIH DMS/NIGMS grant (J. Nardini, Co-PI) to support this project from 2021-2025.*

(B) Topological Data Analysis of Biological Phenomena

Background: In many biological processes, complicated interactions between individuals lead to fascinating emergent behavior: the flocking of birds, branched blood vessel networks, or the spots dotting a leopard's skin. *Topological data analysis* (TDA) is an emerging field of research that is suited to extract and infer geometrical properties from experimental data on these phenomena. Agent-based models (ABMs) are suited model the interactions that lead to this patterns, but ABM analysis is challenging due to its discrete, stochastic, and computational nature. *I use TDA to enable common analyses for ABMs, including parameter estimation, model selection, and sensitivity analysis.* Below, I discuss two applications of this methodology for parameter estimation and sensitivity analysis for ABMs with application to biological flocking and blood vessel formation.

Parameter estimation for biological swarm ABMs

When comparing ABMs to experimental data, we may be interested in whether the emergent behavior from the ABM and data match each other. Topological data analysis (TDA) uses methods from topology to summarize such patterns underlying data. In particular, *crocker plots* are a topological signature that count the number of loops (H1) in a dataset over time. In Figure 3, I depict the crocker plot for a spatial ABM, and we see that $H1 = 2$ over many time points and proximity parameters, signifying the presence of two loops (one clockwise, one counter-clockwise) in the ABM simulation.

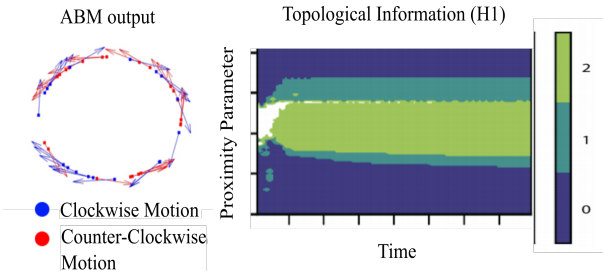

Figure 3: An ABM's patterning can be summarized with topological information.

Machine learning is suited to extract complex interactions from high-dimensional data. My research provided the first study to demonstrate how TDA-derived features can accurately train machine learning algorithms for predicting spatiotemporal dynamics. We simu-

lated an ABM for many parameter values and computed the corresponding crocker plots for each simulations. Using these crocker plots as the input feature vectors, we trained a linear support vector machine (SVM) that outputs a parameter prediction. Performing 5-fold cross validation led to accurate parameter classification for 90% of the held out ABM simulations. We also trained linear SVMs with order parameters (physics-based calculations summarizing the polarity, angular momentum, and distance between agents) to compare our topological approach against. Linear SVMs trained on order parameters only achieved at best 80% accuracy. We ultimately concluded that *TDA provides an efficient way to visually summarize the patterns underlying data and which can aid parameter estimation for an ABM.*

Topological Analysis for Spatial Blood Vessel Networks

Current TDA methods focus on analyzing point clouds of data, including collections of points in $\mathbb{R}^n$. Spatial networks demonstrate a range of interesting topological properties, but current methods to summarize these patterns from data are lacking. A striking example of such networks include blood vessels (Fig. 4), where we observe distinct network morphologies inside healthy and diseased human eyes. I have developed a novel TDA methodology to summarize the topological information present in network images and demonstrated that this methodology can accurately classify simulations from an ABM of the blood vessel development process. I plan to extend this framework in the future to serve as a computational tool to summarize network morphologies and aid medical specialists in diagnosing retinal diseases.

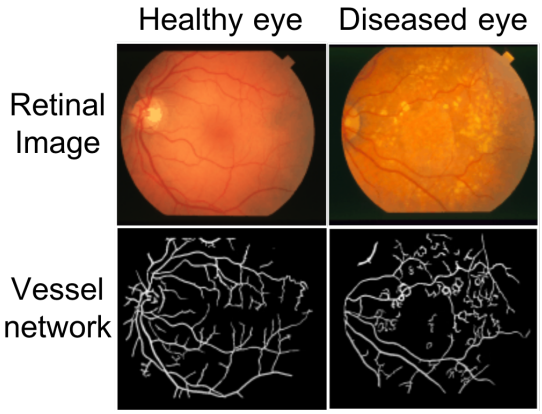

Figure 4: Healthy and diseased eyes lead to varied network structure and function.

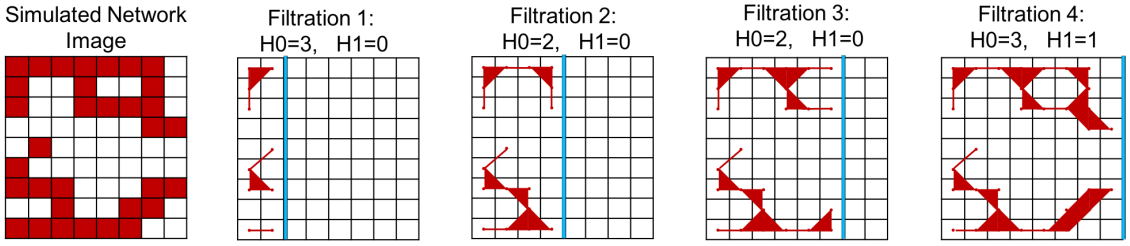

Figure 5: The plane sweeping approach in 2d: We begin with an image of a simulated network (red pixels denote the presence of a blood vessel) and sweep a line from left to right over the image to extract the network's topological information.

We demonstrated that this novel filtration approach can accurately classify model simulations of the blood vessel development process. We generated over 1,000 simulated blood vessel networks with an ABM model using different mechanistic parameters for each network. A simple k-means clustering algorithm can robustly separate the large parameter space into distinct regions that lead to topologically similar networks.

Future Work: An open question from this research is the development of *stability criteria* to ensure these summaries of data are not corrupted by noise. I plan to perform

a stability analysis of the plane sweeping approach to ensure this methodology robustly identifies the topological features of a given network. Once we know this methodology is stable, I plan to use this methodology to train machine learning algorithms on a database of over 200 blood vessel networks (similar to those Figure 4) that have been labelled with professional disease diagnostics. This will lead to a useful computational tool that will allow medical specialists to easily diagnose patients.

(C) Discovering Dynamical System Models with Machine Learning

Background: Model selection is challenging because many models may be able to reproduce similar behavior. Equation learning is a recent field of research that uses concepts from machine learning to infer a mathematical model directly from noisy data. *Biological data presents many challenges for such methods with sparse and noisy data*, as small noise levels, as small noise levels (such as 5%) and infrequent sampling of data can corrupt derivative computations and, in turn, lead to inaccurate model recovery. I have developed an equation learning pipeline to enable accurate model selection from sparse and noisy data.

Suppose we have noisy observations, $y(t,x)$, of some smooth PDE, $u(t,x)$. I have developed an equation learning pipeline (Figure 6) to infer equations from noisy data using and artificial neural network for denoising, the PDE Functional Identification of Nonlinear Dynamics (PDE-Find) algorithm, and model selection.

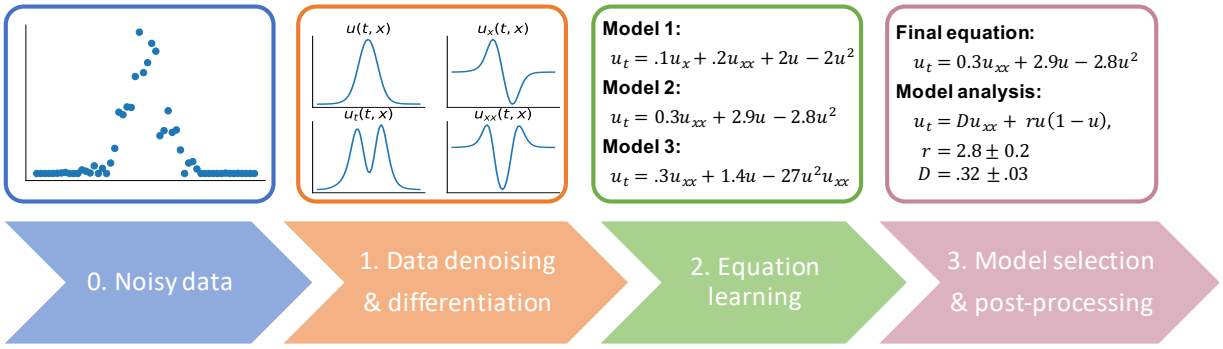

Figure 6: An equation learning pipeline consisting of data denoising, equation learning, and model selection allows for accurate model discovery with sparse and noisy biological data.

The first step of this pipeline denoises $y(t,x)$ to obtain derivative estimates u_t, u_x, u_{xx}, etc. The second step builds a large library of candidate terms for the learned model, such as

$$\Theta = \begin{bmatrix} u & u^2 & \dots & u^p & u_x & u_{xx} & \dots \end{bmatrix},$$

where each column of Θ is a vectorization of the written term. The learned model for y is then obtained by sparsely solving the linear system

$$u_t = \Theta\xi. \tag{2}$$

The nonzero elements of ξ provide parameter estimates for the relevant terms in the learned model. We obtain several candidate models from this process and perform model selection and uncertainty quantification in the final step.

Inaccurate data denoising in step 1 of the pipeline leads to inaccurate differentiation and incorrect learned equations. Many current methods use polynomial splines for denoising, which can be corrupted in the presence of noise and few time samples (such as 5% noise with 5 time samples). We address the challenge of robust denoising and learning of sparse noisy data in [6, 8] by using a simple hidden-layer artificial neural network (ANN) to denoise data and perform model selection over several learned equations. I demonstrated in collaboration with neuro-oncologists that this pipeline accurately recovers the equations underlying noisy data from cancer-related models with as few as three data samples [8]. *The ability to learn equations in the presence of high noise levels and sparse time sampling will further our understanding of many biological phenomena where data collection is challenging,* e.g., *brain cancer imaging.*

Future work: Mean-field differential equation models are often used as surrogates for discrete and stochastic ABM simulations. Such models are advantageous for ABM analysis because they can be analyzed with analytical techniques (*e.g.*, bifurcation & traveling wave analysis), but they often lead to inaccurate ABM dynamics. I am collaborating with Ruth Baker and Matthew Simpson to determine how our equation learning pipeline can learn novel differential equation models that approximate ABMs better than mean-field models and capture how spatial patterns between agents affect the population's overall dynamical rates. As an example, I depict the output of an ABM for the growth of a population of cells during a proliferation assay against its continuum limit model and learned model in Figure 7. The mean-field model does a poor job describing the ABM output while the learned model nicely matches the ABM. *The development of learned differential equations models will enable accurate inference of ABM dynamics by using analytical techniques to understand the learned equation behavior.*

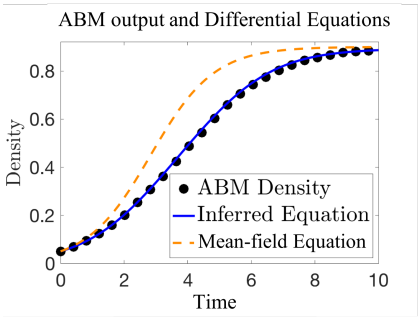

Figure 7: Comparing ABM output to differential equation models.

References

[1] H. T. Banks, A. Choi, T. Huffman, J. Nardini, L. Poag, and W. C. Thompson. Quantifying CFSE label decay in flow cytometry data. *Applied Mathematics Letters*, 26(5):571–577, May 2013.

[2] T. Huffman, K. Link, and et al. Nardini, J. A Mathematical Model of RNA3 Rrecruitment in the Replication Cycle of Brome Mosaic Virus. *Int. J. of Pure and Appl. Math.*, 92(1):27–49, 2014.

[3] John T. Nardini and D. M. Bortz. The influence of numerical error on parameter estimation and uncertainty quantification for advective PDE models. *Inverse Problems*, 35(6):065003, June 2019.

[4] Rebecca Everett and et al. Flores, Kevin B. A tutorial review of mathematical techniques for quantifying tumor heterogeneity. *Mathematical Biosciences and Engineering*, 17(4):3660, 2020.

[5] Dhananjay Bhaskar, Angelika Manhart, Jesse Milzman, John T. Nardini, Kathleen M. Storey, Chad M. Topaz, and Lori Ziegelmeier. Analyzing collective motion with machine learning and topology. *Chaos*, 29(12):123125, December 2019.

[6] John H. Lagergren, John T. Nardini, G. Michael Lavigne, Erica M. Rutter, and Kevin B. Flores. Learning partial differential equations for biolxogical transport models from noisy spatio-temporal data. *Proc. R. Soc. A*, 476(2234), February 2020.

[7] John H. Lagergren, John T. Nardini, Ruth E. Baker, Matthew J. Simpson, and Kevin B. Flores. Biologically-informed neural networks guide mechanistic modeling from sparse experimental data. *arXiv:2005.13073 [math, q-bio]*, May 2020. arXiv: 2005.13073.

[8] John T. Nardini and et al. Lagergren, John H. Learning Equations from Biological Data with Limited Time Samples. *Bull Math Biol*, 82(9):119, September 2020.

[9] John T. Nardini, Douglas A. Chapnick, Xuedong Liu, and David M. Bortz. Modeling keratinocyte wound healing: cell-cell adhesions promote sustained migration. *Journal of Theoretical Biology*, 400:103–117, July 2016.

[10] J. Nardini and D. Bortz. Investigation of a Structured Fisher's Equation with Applications in Biochemistry. *SIAM J. Appl. Math.*, 78(3):1712–1736, January 2018.