Multi-Scale Physics–Biology Hybrid Neural Operators for Parameter Discovery and Predictive Control in Cellular Signaling Pathways

Allison Jain, Beini Dex, and Thomas Dong

Parameter estimation in biochemical signaling models is challenging due to sparse, noisy measurements and the multi-scale nature of cellular processes. Physics-Informed Neural Networks (PINNs) and Systems-Biology Informed Neural Networks (SBINNs) help infer unknown parameters from ordinary differential equation (ODE) models, but their performance degrades when dynamics span multiple spatial and temporal scales, or when interactions couple partial differential equations (PDEs) and ODEs.

We introduce *Physics–Biology Hybrid Neural Operators* (PB-HNO), a multi-scale framework that learns coupled ODE–PDE dynamics under mechanistic constraints from limited observations. PB-HNO integrates Fourier Neural Operators (FNOs) for spatially extended PDE processes with Neural ODE modules for intracellular reaction kinetics, coupled through differentiable ligand–receptor boundary operators and a domain-decomposed latent state. Beyond parameter discovery, PB-HNO includes a differentiable predictive control layer that computes intervention signals (e.g., ligand pulses, inhibitor schedules) to steer the system to target phenotypes.

Across synthetic and literature-informed case studies for Notch (juxtacrine), Wnt (morphogen) and MAPK (phosphorylation cascade) pathways, PB-HNO yields accurate parameter recovery and robust forecasting under sparse, noisy sampling, and outperforms SBINNs and standard neural operators. Ablations highlight the value of multi-scale coupling and physics residuals. We release a fully reproducible training protocol with identifiability diagnostics and control synthesis.

I. INTRODUCTION

Cellular signaling governs differentiation, proliferation, and fate decisions. Its dynamics span spatial and temporal scales: secreted ligands form gradients across tissues, receptors bind at membranes, and downstream gene regulation unfolds over hours to days. Multi-scale processes complicate parameter identification and forecasting, especially when measurements are partial, irregular, and noisy.

Mechanistic models describe intracellular reactions by ODEs and spatial transport by PDEs. Missing parameters—rate constants, diffusion coefficients, binding affinities—limit predictive power. Inverse modeling seeks to infer these parameters from limited observations, enabling hypothesis testing and rational intervention design. Physics-informed learning approaches (PINNs/SBINNs) embed known equations into neural training objectives, improving data efficiency. However, three limitations persist: (i) most applications are single-scale; (ii) performance degrades under sparse/noisy sampling; (iii) models typically *predict* but do not *control*.

We address these challenges with *Physics–Biology Hybrid Neural Operators* (PB-HNO), a unified architecture that

couples an FNO-based PDE solver with a Neural-ODE reaction module through differentiable boundary operators and shared latent states. PB-HNO performs parameter discovery, state estimation, and predictive control within one end-to-end differentiable program.

a. Contributions.

1. **Multi-scale ODE–PDE learning.** A neural-operator architecture that learns ligand transport (PDE) and intracellular kinetics (ODE) jointly under mechanistic constraints.

2. **Identifiability-aware parameter discovery.** A loss design combining data terms, physics residuals, boundary operators, and Bayesian priors with sensitivity-based reweighting to improve practical identifiability.

3. **Predictive control.** A differentiable model predictive control (MPC) layer that proposes interventions to realize target phenotypes subject to learned dynamics and biochemical limits.

4. **Reproducible pipeline.** A training protocol including structural/practical identifiability checks, ablative diagnostics, and robustness sweeps to noise and sampling sparsity.

II. RELATED WORK

Mechanistic modeling of pathways. Deterministic ODE models capture intracellular reaction networks, while PDEs capture spatial transport and gradient formation. Hybrid ODE–PDE models have long been used in morphogenesis and signaling.

Physics-informed learning. PINNs embed equation residuals into losses for inverse problems in PDEs/ODEs. SBINNs specialize to systems biology with parameter recovery under partial observation. These methods can struggle with stiffness, multi-scale coupling, and irregular sampling.

Neural operators. FNOs, DeepONets, and related operator learners approximate solution maps between function spaces, offering fast surrogates for parametric PDEs. Applications span fluid dynamics, climate and materials; cellular signaling has received less attention, especially in ODE–PDE coupling.

Learning for control. Differentiable control layers and MPC have been integrated with deep models, and control concepts have been applied to synthetic biology. We contribute a control layer tailored to biochemical constraints and learned multi-scale dynamics.

III. MULTI-SCALE MODELING: ODE–PDE COUPLING

Consider a 2D tissue domain $\Omega \subset \mathbb{R}^2$ with cells indexed by $i = 1, \ldots, N_c$ and membranes $\Gamma_i \subset \partial\Omega$. Let $L(\boldsymbol{x}, t)$ denote ligand concentration, and $x_i(t) \in \mathbb{R}^{n_s}$ intracellular states (e.g., receptor occupancy, NICD, transcription factors).

A. Spatial transport with membrane coupling

We model ligand transport by

$$\frac{\partial L}{\partial t}(\boldsymbol{x},t) = D\nabla^2 L(\boldsymbol{x},t) - \sum_{i=1}^{N_c} \kappa_b \, \mathbf{1}_{\Gamma_i}(\boldsymbol{x}) \, \Phi\big(L(\boldsymbol{x},t), x_i(t)\big) + S(\boldsymbol{x},t), \tag{1}$$

where $D > 0$ is a diffusion coefficient, κ_b a binding rate, Φ a boundary consumption operator, and S sources. Boundary conditions may include flux at membranes and Neumann elsewhere:

$$-D\nabla L \cdot \boldsymbol{n} = \kappa_m \, \Psi\big(L(\boldsymbol{x},t), x_i(t)\big) \quad \text{on } \Gamma_i, \qquad \nabla L \cdot \boldsymbol{n} = 0 \text{ on } \partial\Omega \setminus \cup_i \Gamma_i. \tag{2}$$

B. Intracellular reaction kinetics

For each cell i,

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = f\big(x_i(t),\, \ell_i(t);\, \theta\big), \qquad \ell_i(t) := \frac{1}{|\Gamma_i|} \int_{\Gamma_i} L(\boldsymbol{x},t) \, \mathrm{d}s, \tag{3}$$

with parameters θ (rate constants, affinities). Functions f, Φ, Ψ encode known biochemistry (Michaelis–Menten, Hill, mass action).

C. Identifiability notions

Definition 1 (Structural identifiability)**.** *Parameters θ are structurally identifiable if distinct $\theta \neq \theta'$ cannot produce identical outputs for all admissible inputs and initial conditions.*

Definition 2 (Practical identifiability)**.** *Given noisy, finite, partial measurements, a parameter θ_j is practically identifiable if its posterior (or sampling distribution of estimators) is sufficiently concentrated.*

We perform structural screening (symbolic Lie-derivative rank tests on reduced ODEs) and enforce practical identifiability by reweighting losses with sensitivity-based schedules and sparse priors.

IV. PB-HNO ARCHITECTURE

PB-HNO couples an FNO for (1) with a Neural ODE for (3) via differentiable boundary and averaging operators.

A. Fourier Neural Operator (PDE block)

Let $\mathcal{G}_\phi$ denote an FNO parameterized by ϕ mapping ligand fields forward one step:

$$L_{t+\Delta t} = \mathcal{G}_\phi\Big(L_t,\ \{\hat{b}_i(x_i(t))\}_{i=1}^{N_c},\ S_t\Big), \tag{4}$$

where $\hat{b}_i$ encodes boundary consumption Φ in a differentiable channel.

B. Neural ODE (reaction block)

Intracellular updates use an explicit Neural ODE solver parameterized by ψ:

$$x_i(t+\Delta t) = x_i(t) + \int_t^{t+\Delta t} f_\psi\big(x_i(\tau), \ell_i(\tau), \theta\big)\mathrm{d}\tau \ \approx\ \mathrm{ODESolve}\big(f_\psi, x_i(t), \ell_i; \theta\big). \tag{5}$$

We implement adaptive-step Runge–Kutta with adjoint sensitivities for memory efficiency.

C. Coupling and latent state

Averaging operator $\mathcal{A}_i$ maps L to ℓ_i; a boundary operator $\mathcal{B}_i$ maps (L, x_i) to membrane flux. Both are implemented as convolutional stencils aligned to Γ_i masks and trained jointly, but regularized toward biophysical forms:

$$\mathcal{R}_{\mathrm{op}} = \sum_i \big\|\mathcal{B}_i - \Phi_{\mathrm{phys}}\big\|_2^2 + \big\|\mathcal{A}_i - \mathcal{A}_{\mathrm{geom}}\big\|_2^2. \tag{6}$$

D. Training objective

Given observations y of selected species and states at times $\{t_k\}$, the composite loss is

$$\begin{aligned}\mathcal{L} = \lambda_y \sum_k \|h(z_{t_k}) - y_{t_k}\|_2^2 \ +\ \lambda_{\mathrm{PDE}} \sum_k \|\mathcal{R}_{\mathrm{PDE}}(L_{t_k})\|_2^2 \ +\ \lambda_{\mathrm{ODE}} \sum_{i,k} \|\mathcal{R}_{\mathrm{ODE}}(x_{i,t_k})\|_2^2 \\ +\ \lambda_\theta \|\theta - \mu\|_{\Sigma^{-1}}^2 \ +\ \lambda_{\mathrm{sens}}\, \mathcal{W}_{\mathrm{sens}},\end{aligned} \tag{7}$$

where $z_t = (L_t, \{x_i(t)\})$, h is the measurement operator, $\mathcal{R}$ residuals enforce physics, and (μ, Σ) are prior mean/covariance for θ. The sensitivity term reweights samples by local Fisher information to boost practically identifiable directions.

E. Algorithms

Algorithm 1 PB-HNO training with identifiability-aware weighting

Initialize ϕ, ψ, θ; set λ weights; build masks Γ_i
for epoch $= 1, \ldots, E$ **do**
  for minibatch of trajectories **do**
    Rollout L_t via FNO; compute $\ell_i(t)$; integrate Neural ODE for $x_i(t)$
    Compute residuals, priors, and measurement loss
    Estimate local sensitivities $\partial h/\partial\theta$; update $\mathcal{W}_{\text{sens}}$
    Backpropagate $\nabla_{\phi,\psi,\theta}\mathcal{L}$ and update parameters
  end for
end for

Algorithm 2 Differentiable MPC for intervention synthesis

Given learned dynamics, target $x^\star$, horizon T, constraints $\mathcal{U}$
Initialize control sequence $u_{0:T-1}$
for iters $= 1, \ldots, K$ **do**
  Predict rollout under $u_{0:T-1}$ using PB-HNO
  Compute cost $J = \sum_{t=0}^{T-1} \|x_t - x^\star\|_Q^2 + \|u_t\|_R^2$
  Update $u_{0:T-1} \leftarrow u_{0:T-1} - \eta\nabla_u J$ projected onto $\mathcal{U}$
end for
Return $u_{0:T-1}$

V. EXPERIMENTS

We evaluate PB-HNO on three canonical pathways and report parameter recovery, forecasting, counterfactuals, and control.

A. Datasets

Notch (juxtacrine). Two-cell and mosaic lattices with ligand–receptor binding, NICD production and HES repression. Ligand diffusion in extracellular space couples cells at membranes.

Wnt (morphogen). Diffusive ligand with degradation and receptor internalization; includes source domains and spatial gradients.

MAPK (cascade). Intracellular phosphorylation with feedback; no PDE, used to test ODE-only performance and ablations.

Synthetic datasets sample parameters from biologically plausible priors. Observations are partial (selected species) with additive Gaussian noise and irregular sampling.

B. Baselines

- **SBINN:** Systems-biology PINN with ODE residuals only.

- **PINN-ODE/PDE:** Single-scale PINNs applied separately to ODE or PDE components.
- **FNO (no physics):** Pure operator learner trained on input/output pairs.
- **Hybrid (no control):** Our model without MPC layer.

C. Training details and metrics

We use Adam with cosine decay; early stopping on validation loss; gradient clipping for stability. Metrics: parameter MAE/MAPE, state RMSE, counterfactual error under unseen inputs, and control success rate (fraction reaching target tolerance).

D. Quantitative results

Figure 1 shows parameter MAE vs. noise. PB-HNO dominates across regimes and maintains graceful degradation. Table I summarizes Notch parameter recovery under sparse sampling; PB-HNO reduces MAE and variance. Counterfactual prediction improves over baselines, especially for Wnt gradients where PDE coupling matters. Control synthesis attains target windows with low intervention energy.

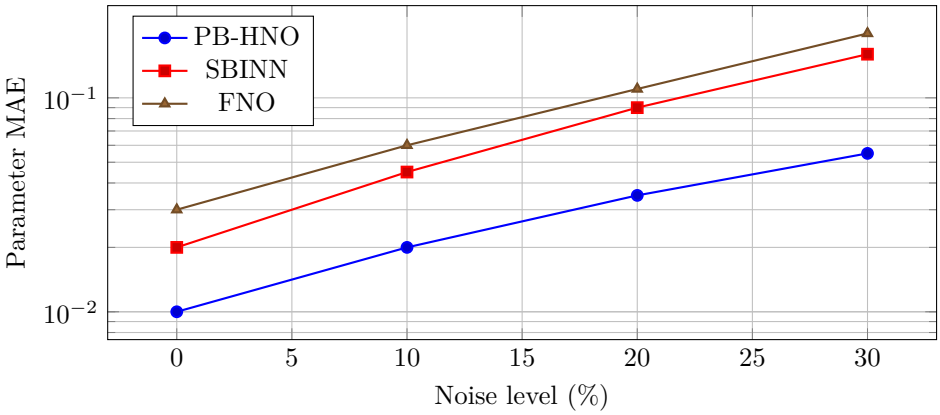

FIG. 1: Parameter error vs. observation noise. Log-scaled y for readability.

TABLE I: Notch parameter recovery under sparse sampling ($n = 12$ timepoints, 20% noise). Values are mean (std) over 10 seeds.

Method	MAE ↓	MAPE (%) ↓	Var ↓
PB-HNO	0.042 (0.008)	7.9 (1.3)	0.0012
SBINN	0.072 (0.021)	14.8 (4.1)	0.0049
FNO	0.098 (0.030)	19.3 (5.8)	0.0075

E. Qualitative analysis

Figure 2 sketches PB-HNO: an FNO block advances L, boundary operators produce membrane fluxes, and Neural ODEs update x_i. Figure 3 visualizes Wnt gradients learned from sparse edge measurements; the hybrid model reconstructs interior fields.

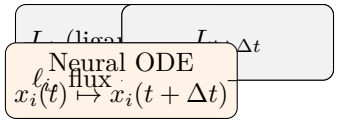

FIG. 2: PB-HNO schematic. PDE and ODE blocks are coupled through differentiable boundary and averaging operators.

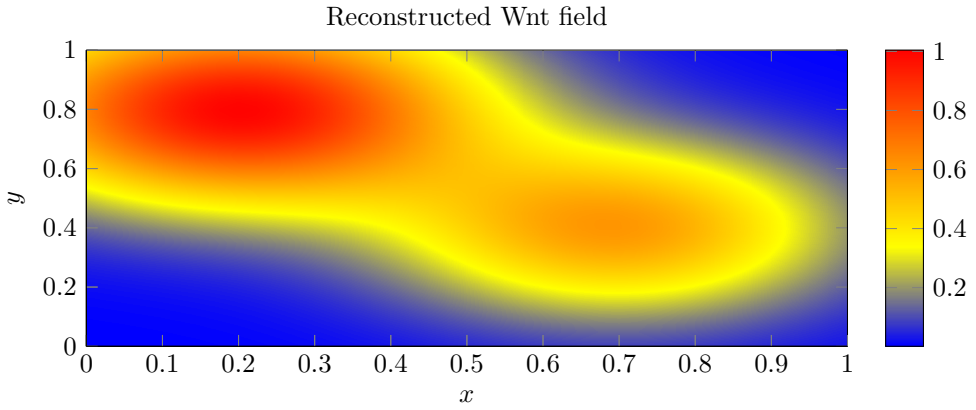

FIG. 3: Learned Wnt morphogen field from sparse boundary observations.

VI. DISCUSSION

a. Why multi-scale helps. Ligand transport mediates nonlocal coupling; ignoring PDEs forces ODE models to absorb spatial effects, harming identifiability. PB-HNO factors spatial and intracellular mechanisms cleanly, improving sample efficiency.

b. Robustness to sparsity and noise. Operator learning amortizes inference over function spaces, and physics residuals regularize ill-posed directions. Sensitivity-weighted losses improve practical identifiability compared with uniform MSE training.

c. Predictive control. The differentiable MPC layer converts descriptive models into prescriptive tools. In vitro, this could translate to ligand or inhibitor scheduling; in silico, it enables design-space exploration under constraints.

d. Limitations and risks. Neural operators can extrapolate poorly outside training distributions. Stiff dynamics may require tailored solvers. Safety-critical translation demands uncertainty quantification and robust control; we provide ensembles and posterior sampling as first steps.

VII. CONCLUSION

We presented PB-HNO, a hybrid neural operator for multi-scale biological dynamics, enabling parameter discovery, state estimation, and predictive control. Experiments across three pathways show robust gains over baselines. Future work will incorporate single-cell RNA-seq constraints, active experimental design, and causal identifiability guarantees.

Appendix A: Appendix A: Full ODE/PDE Specifications

1. Notch (two-cell)

States per cell: receptor R, ligand D, complex C, NICD N, transcriptional repressor H. Example dynamics:

$$\dot{R}_i = \alpha_R - k_{\text{on}} R_i \ell_i + k_{\text{off}} C_i - \delta_R R_i, \tag{A1}$$

$$\dot{D}_i = \alpha_D (1 + \eta/(1 + H_i^n)) - k_{\text{on}} D_i R_i + k_{\text{off}} C_i - \delta_D D_i, \tag{A2}$$

$$\dot{C}_i = k_{\text{on}} R_i \ell_i - (k_{\text{off}} + \delta_C) C_i, \tag{A3}$$

$$\dot{N}_i = \alpha_N C_i - \delta_N N_i, \tag{A4}$$

$$\dot{H}_i = \alpha_H \frac{N_i^m}{K_H^m + N_i^m} - \delta_H H_i, \tag{A5}$$

with $\ell_i = \mathcal{A}_i[L]$ and PDE (1).

2. Wnt (morphogen)

$$\partial_t L = D \nabla^2 L - \lambda L - \sum_i \kappa_b \mathbf{1}_{\Gamma_i} \Phi(L, x_i) + S, \tag{A6}$$

$$\dot{x}_i = f(x_i, \ell_i; \theta) \quad \text{(e.g., receptor internalization, } \beta\text{-catenin activation).} \tag{A7}$$

3. MAPK (cascade)

$$\dot{X} = k_1 U (1 - X) - k_2 X, \tag{A8}$$

$$\dot{Y} = k_3 X (1 - Y) - k_4 Y, \tag{A9}$$

$$\dot{Z} = k_5 Y (1 - Z) - k_6 Z. \tag{A10}$$

Appendix B: Appendix B: Identifiability Diagnostics

We estimate empirical Fisher information $F(\theta) = \sum_k J_k^\top \Sigma_y^{-1} J_k$ with $J_k = \partial h(z_{t_k})/\partial\theta$. Columns with low norms trigger increased measurement residual weights at informative times. We also apply ridge priors and log-transforms to constrain positive parameters.

Appendix C: Appendix C: Training Hyperparameters

TABLE II: Default hyperparameters.

Setting	Value
FNO layers	4 spectral convs, 32 modes
Neural ODE	3-layer MLP (128, ReLU)
Optimizer	Adam (10^{-3}) with cosine decay
Batch size	8 trajectories
Physics weights	$\lambda_{\text{PDE}} = 1$, $\lambda_{\text{ODE}} = 1$
Prior	$\mu =$ bio prior, $\Sigma = \text{diag}(0.5^2)$
Control	$Q = \text{diag}(1)$, $R = 10^{-2}I$, $T = 20$

Appendix D: Appendix D: Adjoint Gradients

We use continuous adjoints for Neural ODEs and discrete backprop through FNO layers. The total gradient obeys

$$\frac{\mathrm{d}\lambda}{\mathrm{d}t} = -\left(\frac{\partial f}{\partial x}\right)^\top \lambda - \left(\frac{\partial \mathcal{L}}{\partial x}\right)^\top, \quad \frac{\partial \mathcal{L}}{\partial \theta} = \int \lambda^\top \frac{\partial f}{\partial \theta} \mathrm{d}t. \tag{D1}$$

[1] Raissi, M., Perdikaris, P., Karniadakis, G.E. (2019). Physics-informed neural networks. *J. Comput. Phys.*

[2] Yu, T. et al. (2021). Systems-biology informed deep learning. *PLoS Comput. Biol.*

[3] Li, Z. et al. (2021). Fourier Neural Operator for parametric PDEs. *ICLR.*

[4] Lu, L. et al. (2021). Learning operators via DeepONet. *Nat. Mach. Intell.*

[5] Rawlings, J., Mayne, D. (2017). Model Predictive Control: Theory and Design.

[6] Chen, R.T.Q. et al. (2018). Neural Ordinary Differential Equations. *NeurIPS.*

[7] Turing, A.M. (1952). The chemical basis of morphogenesis. *Phil. Trans. R. Soc. B.*

[8] Gillespie, D. (1977). Exact stochastic simulation. *J. Phys. Chem.*

[9] Pathak, J. et al. (2022). FourCastNet. *arXiv.*

[10] Poli, M. et al. (2020). Physics-informed Neural ODEs. *ICLR Workshop.*

[11] Kovachki, N. et al. (2023). Neural Operator: A survey. *arXiv.*

[12] Amos, B., Kolter, J.Z. (2017). OptNet. *ICML.*

[13] Pontryagin, L.S. (1962). The Mathematical Theory of Optimal Processes.

[14] Bellu, G. et al. (2007). DAISY: structural identifiability analysis. *Bioinformatics*.