Predictioneer: Inferring a Gene Regulatory Network from Spatial Expression Dynamics

Solving the Inverse Problem in Biological Systems Modeling

Team ORDINAR
Jadavpur University, Kolkata
Sourish Senapati, Snehali Sadhukhan,
Bidhidipta Dutta, Asmita Bagchi

February 3, 2026

1 Introduction

The Predictioneer challenge involves the identification of a Gene Regulatory Network (GRN) governing the spatial-temporal dynamics of a four-gene system (S1, S2, S3, S4) in a 2D tissue. The system is exposed to two spatially varying morphogens, $M_1(x, y)$ and $M_2(x, y)$, defined as:

$$M_1(x, y) = e^{-2.5x}, \quad M_2(x, y) = \exp\left(-\frac{(x - 0.5)^2 + (y - 0.5)^2}{2(0.18)^2}\right) \tag{1}$$

The objective is twofold:

1. Reconstruct the interaction matrices (A and B) and the decay rate (α) using data from Experiment A (removal of M_2).
2. Predict the system behavior for Experiment B (removal of M_1).

2 Theoretical Model

The system is modeled as a set of coupled ordinary differential equations (ODEs), linearized around the baseline expression level:

$$\frac{dS_i}{dt} = -\alpha_i S_i + \sum_{j=1}^{4} A_{ij} S_j + \sum_{k=1}^{2} B_{ik} M_k \tag{2}$$

where A_{ij} represents the influence of gene j on gene i, and B_{ik} represents the coupling of morphogen k to gene i. The term $-\alpha_i S_i$ captures the auto-regulation/decay that restores the system to its baseline.

3 Methodology: GPU-Accelerated 6-Sigma Solver

To handle the high-dimensional search space and the large spatial dataset (thousands of grid points), we developed a custom **GPU-Accelerated Evolutionary Solver** using PyTorch.

3.1 Data Processing

We utilized the provided `GRN_experiment_M2_removal.csv`. Time-series derivatives $\frac{dS_i}{dt}$ were computed using smoothed finite differences at each spatial coordinate (x, y). The initial steady-state (at $t = 0$) provided a critical constraint:

$$\dot{S}_{ss} = 0 \implies \alpha S_{ss} = A S_{ss} + B M_{both} \tag{3}$$

3.2 Optimization Strategy

We employed a **Mutative Evolutionary Search** strategy (inspired by 6-Sigma refinement):

- **Population Size:** 100,000 candidate models per generation.
- **Loss Function:** A weighted combination of Prediction MSE (Experiment A dynamics) and Nullspace Loss (Steady-state constraint).
- **Hardware:** Accelerated via Tensor Matmul operations on an NVIDIA RTX 4050 GPU, allowing for evaluation of over 10 million models in under 5 minutes.
- **Convergence:** Achieved an RMSE of approximately 8.9×10^{-4} on the training set.

4 Results: Model Reconstruction

The inferred parameters were discretized following the competition guidelines (+1 for promotion, -1 for inhibition, 0 for no interaction).

4.1 Interaction Matrix (A)

The matrix A describes the gene-gene interactions (rows i, columns j representing influence of j on i):

	S1	S2	S3	S4
S1	-	+1	-1	+1
S2	-1	-	-1	+1
S3	+1	-1	-	-1
S4	-1	+1	-1	-

Table 1: Discretized Gene Interaction Matrix A

4.2 Morphogen Coupling Matrix (B)

The matrix B describes the influence of morphogens M_1 and M_2 on each gene:

	M_1	M_2
S1	+1	0
S2	+1	0
S3	-1	0
S4	+1	0

Table 2: Discretized Morphogen Coupling Matrix B

4.3 Decay Rate (α)

The inferred decay rate α was found to be approximately **0.1479**, indicating a relatively slow return to baseline, which allows for sustained spatial patterns during perturbation.

5 Experiment B Prediction

Using the optimized continuous parameters, we simulated the temporal evolution of the system for Experiment B (M_1 removal). The simulation was performed using an Euler-integration scheme across all spatial coordinates for the relevant time points. The results were exported as `predicted_experiment_b.csv`, matching the structure and point-ordering of the provided test files.

6 Conclusion

Our approach combined high-performance computing with biological constraints to yield a highly accurate reconstruction of the underlying Gene Regulatory Network. The extremely low RMSE indicates that the linear approximation of the system dynamics is highly robust for the given experimental range. The discovered network architecture reveals a complex feedback system where M_1 acts as a primary drive for widespread activation, while M_2 plays a more subtle role in fine-tuning the initial steady state.