Efficient Parameter Estimation in Agent-Based Models of Collective Cell Invasion Via Gaussian Process Surrogates and Bayesian Optimization

Aneesh Krishna[1], Temitope O. Benson[1], Khansa Khanam Umer Sultan[1], Ashlee N. Ford Versypt[2]

[1]Institute of Artificial Intelligence and Data Science, University at Buffalo, The State University of New York

[2]Department of Chemical and Biological Engineering, University at Buffalo, The State University of New York

University at Buffalo
The State University of New York

NIH
NIGMS MIRA
R35GM133763

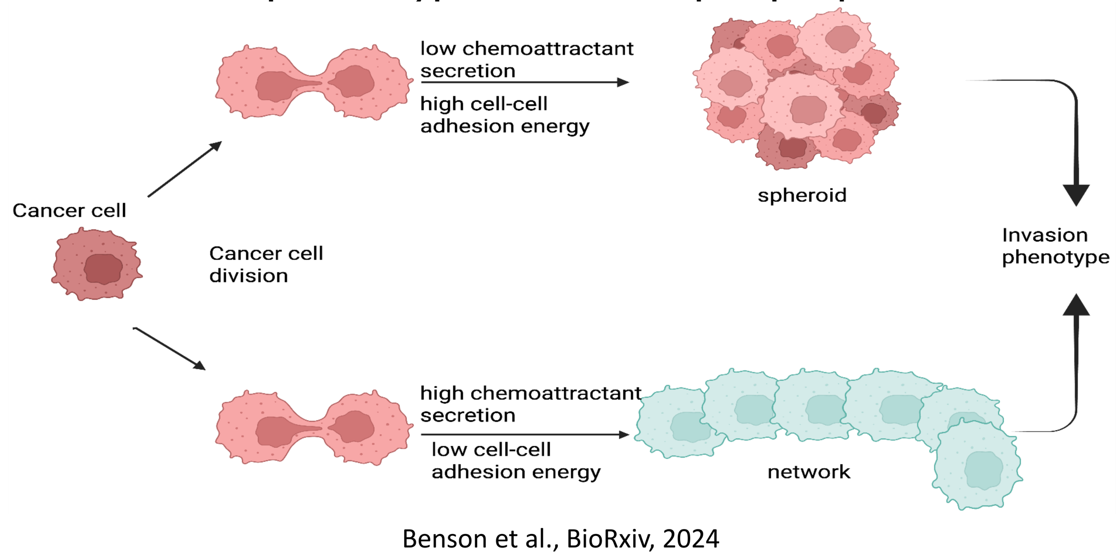
Objectives

- Calibrate agent-based model (ABM) parameters of collective cell invasion without brute-force simulation
- Use a Gaussian process (GP) surrogate steered by Bayesian optimization to find best-fit parameters in far fewer runs

Introduction

Cancer phenotypes

- The cancer cells grow into clusters forming different phenotypes with unique properties

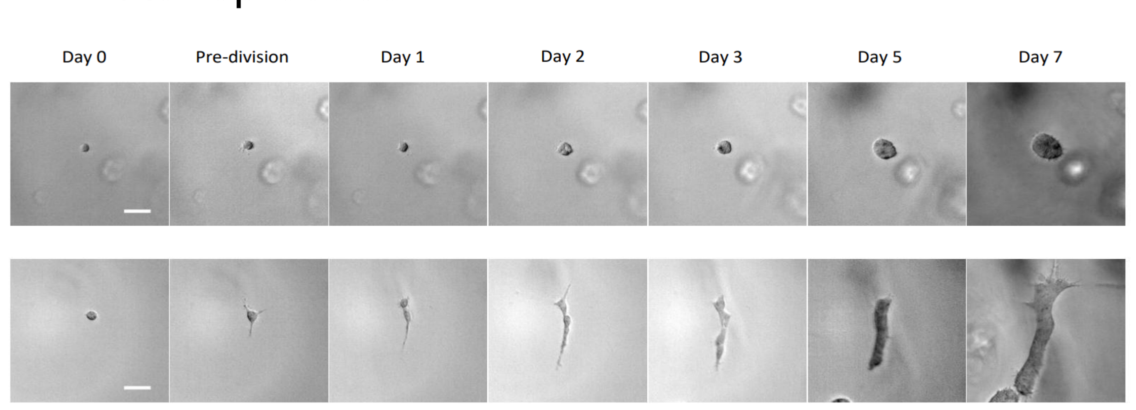


Benson et al., BioRxiv, 2024

- Network phenotype: branch-like shape, highly invasive in nature
- Spheroid phenotype: spherical in shape, less invasive in nature
- Mainly influenced by two features: cell adhesion strength (jcc) and chemical secretion rate (k)

Calibration target

- **What it is:** cancer cells invading, imaged over time
- **What it shows:** the two invasion phenotypes to calibrate to: spheroid (top) and network (bottom)
- **Its role:** this is the experimental target the ABM is fit to reproduce



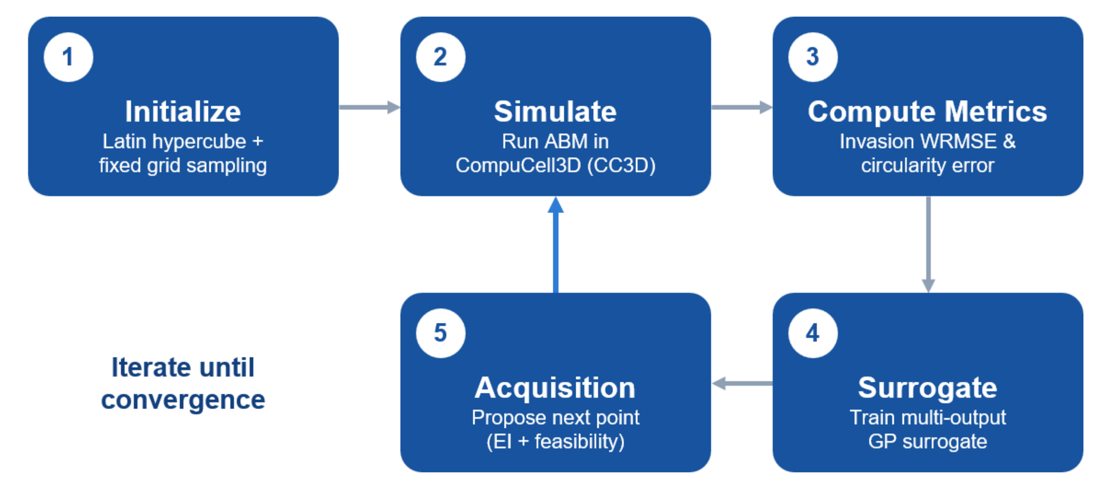
Chen et al. *iScience* 2020

Challenges with parameter estimation

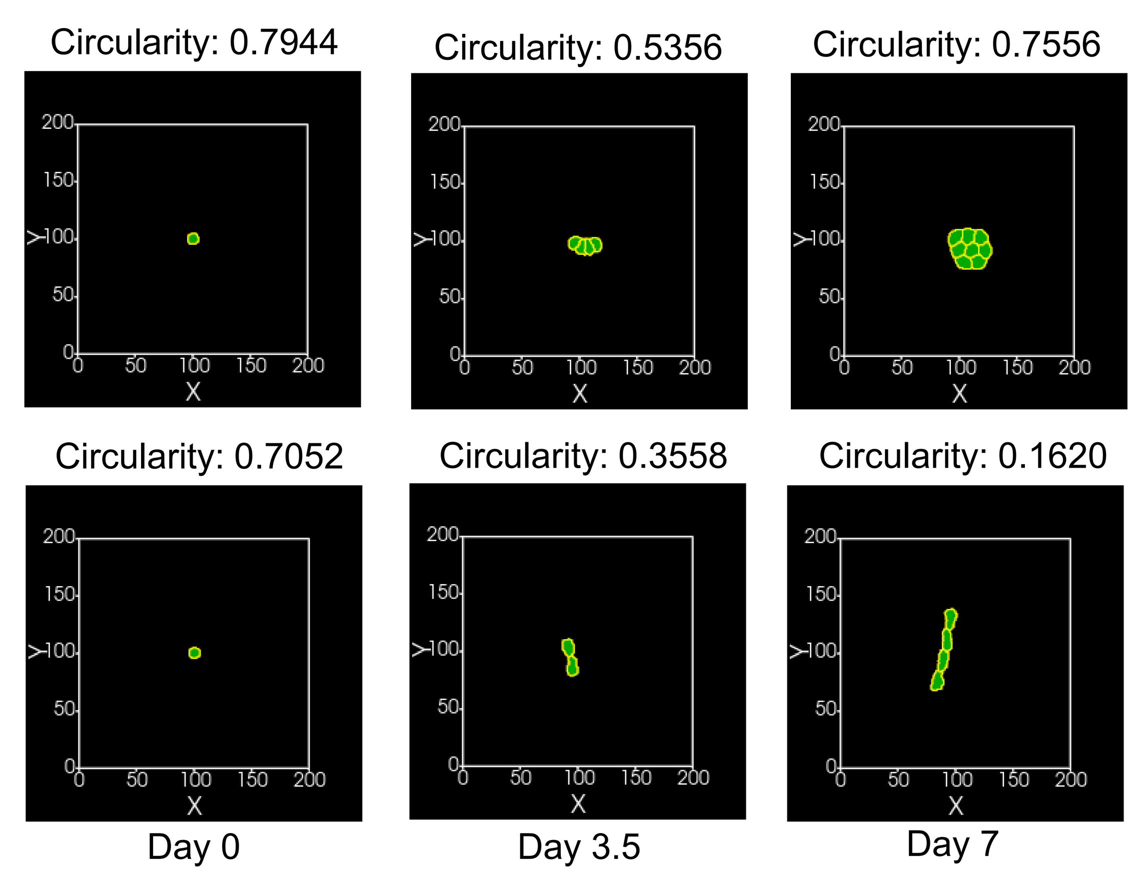
- Intrinsic variability
- Noisy objective landscape
- Expensive evaluations
- Complex search space

Metrics and Calibration

Calibration workflow



CompuCell3D for ABM simulation

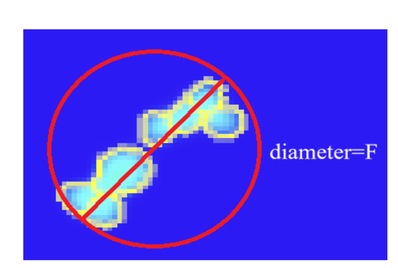


Metrics

- **Max invasion distance** (at time t):

$$d_{max}(t) = \max_i \| x_i(t) - x_1(0) \|$$

  - $x_i(t)$: coordinates of cell i at time t
  - $x_1(0)$: initial position of the reference cell at $t = 0$
  - $\|\cdot\|$: Euclidean distance
- **Circularity**: $C(t) = \frac{4A_{occ}(t)}{\pi F^2(t)}$

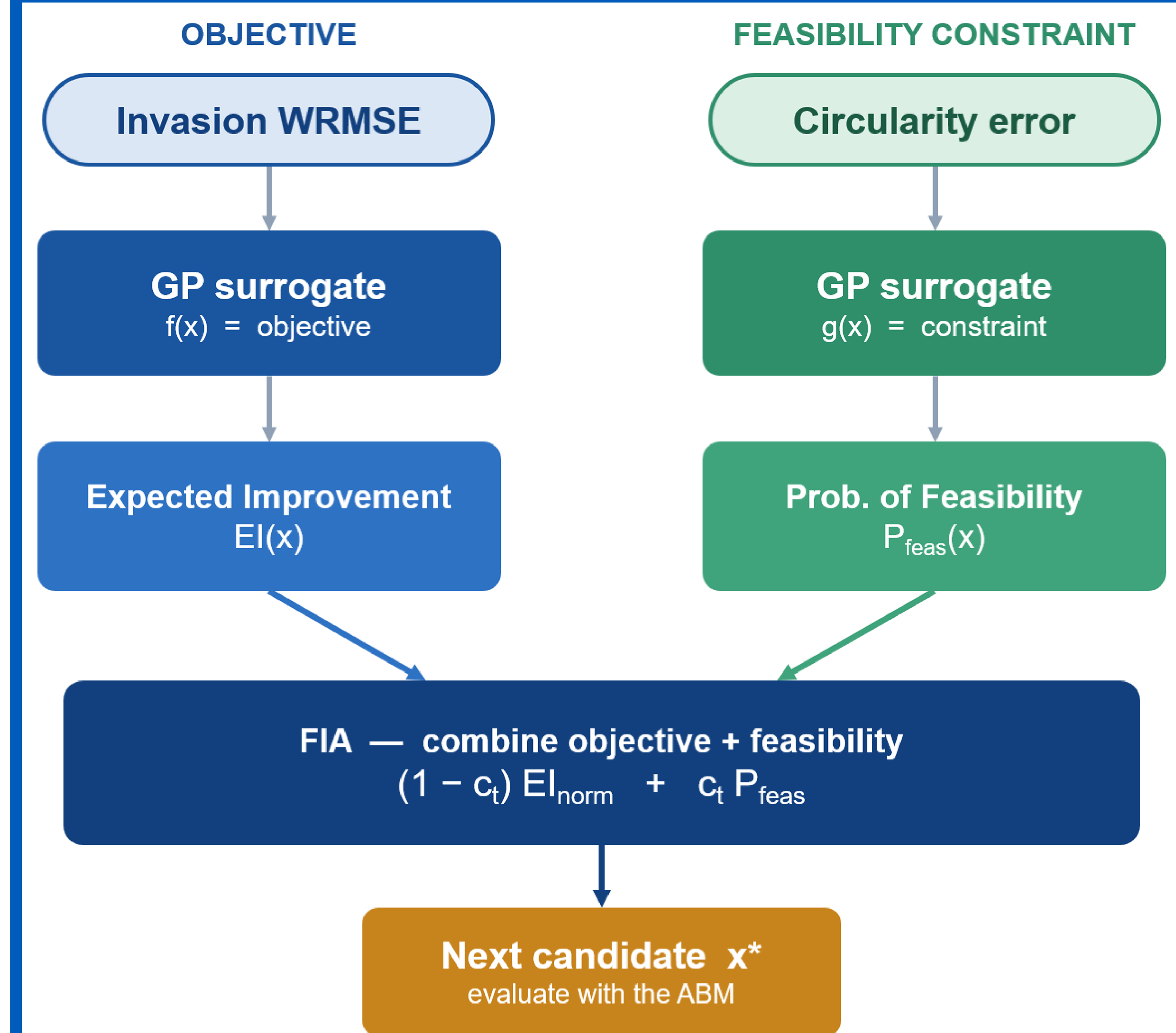


  - $A_{occ}(t)$: area occupied by the cells (green region) at time t
  - $F(t)$: maximum span of the cluster at time t
- **Invasion WRMSE**:

$$I_{WRMSE} = \sqrt{\frac{1}{T}\sum_t \frac{\left(d_{sim}(t) - d_{exp}(t)\right)^2}{\sigma^2(t)}}$$

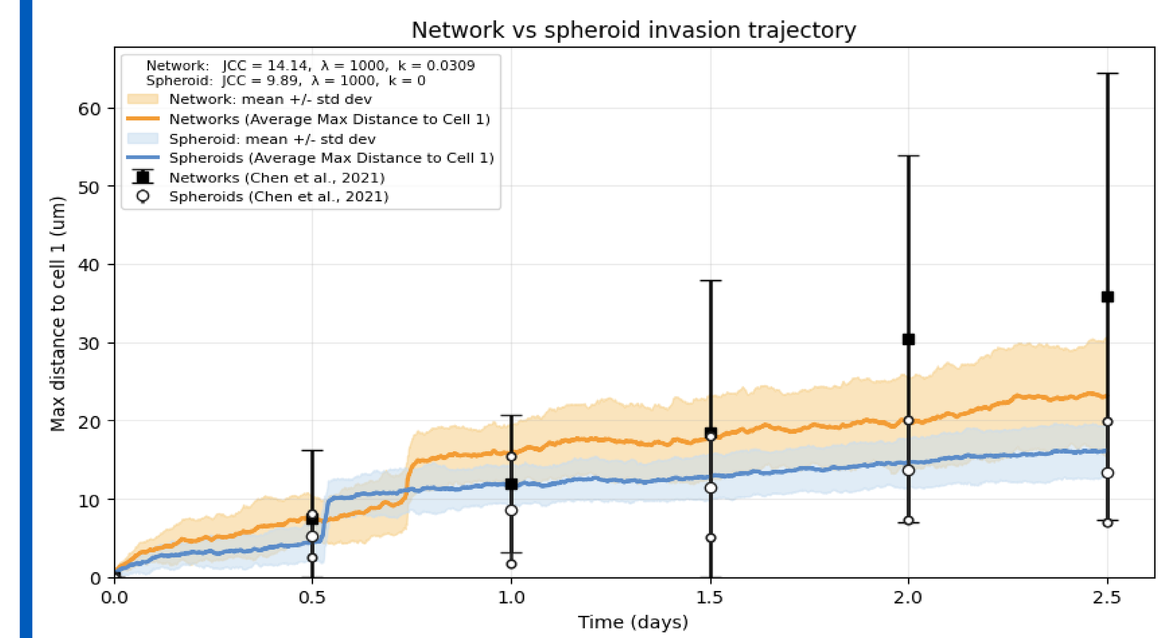
  - $d_{sim}(t)$: invasion from the simulation at time t
  - $d_{exp}(t)$: invasion from experiment at time t
  - $\sigma^2(t)$: experimental variance at time t
  - T: number of time points
- **Circularity error**: $Err_{circ} = |C_{sim} - C_{exp}|$
  - C_{sim}: circularity from the simulation
  - C_{exp}: circularity from experiment
  - $|\cdot|$: Absolute error

Bayesian Optimization

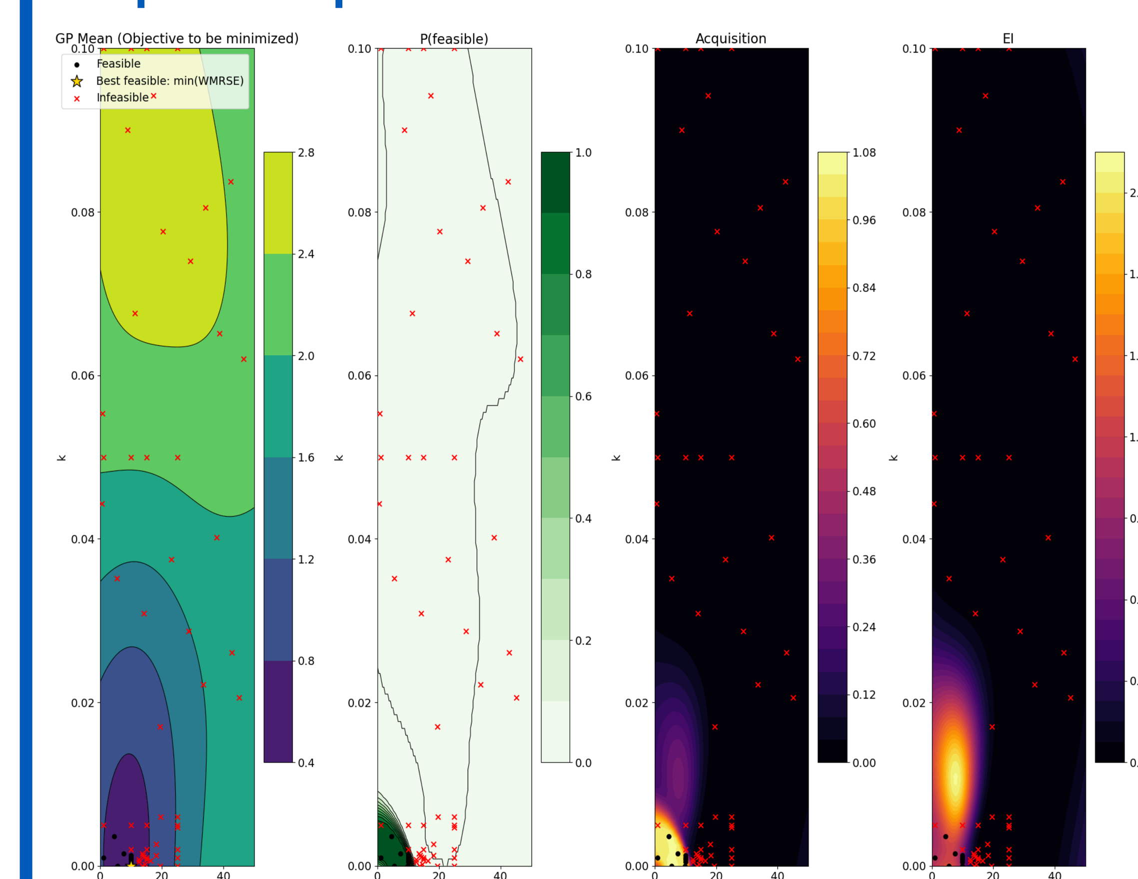


- **GP surrogate**
  - predicts **mean** and **uncertainty**
  - $f(x) \sim \mathcal{GP}\left(m(x), k(x, x')\right)$
  - $f(x)$: invasion WRMSE being modeled
  - $m(x)$: prior mean function
  - $k(x, x')$: RBF kernel
- **Expected improvement** (EI)
  - Evaluated for optimal candidate solution
  - $EI(x) = \left(f^* - \mu(x)\right)\Phi(Z) + \sigma(x)\phi(Z)$
  - $Z = \frac{f^* - \mu(x)}{\sigma(x)}$
  - $\mu(x), \sigma(x)$: GP posterior mean and std at x
  - f^*: best objective observed so far
  - Φ, ϕ: standard-normal CDF and PDF
  - Z: standardized improvement
- **Feasibility constraint**
  - Probability that circularity difference is within threshold (0.15 for spheroid and 0.1 for network)
  - $\mathrm{P_{feas}}(x) = \Phi\left(\frac{g_{thr} - \mu_g(x)}{\sigma_g(x)}\right) - \Phi\left(\frac{-\mu_g(x)}{\sigma_g(x)}\right)$
  - g_{thr}: feasibility threshold
  - $\mu_g(x), \sigma_g(x)$: GP posterior mean and std of the constraint at x
- **Feasibility-interpolated acquisition** (FIA)
  - $\mathrm{FIA}(x) = (1 - c_t)\,\mathrm{EI_{norm}}(x) + c_t\,P_{feas}(x)$
  - $c_t = \left(\frac{L}{N}\right)^t$
  - $\mathrm{EI_{norm}}(x)$: EI scaled to [0,1]
  - c_t: ratio of infeasible measurements (L) to total measurements (N)
  - t: user-set risk exponent, smaller t de-emphasizes feasibility (more risk); larger t emphasizes it (less risk)

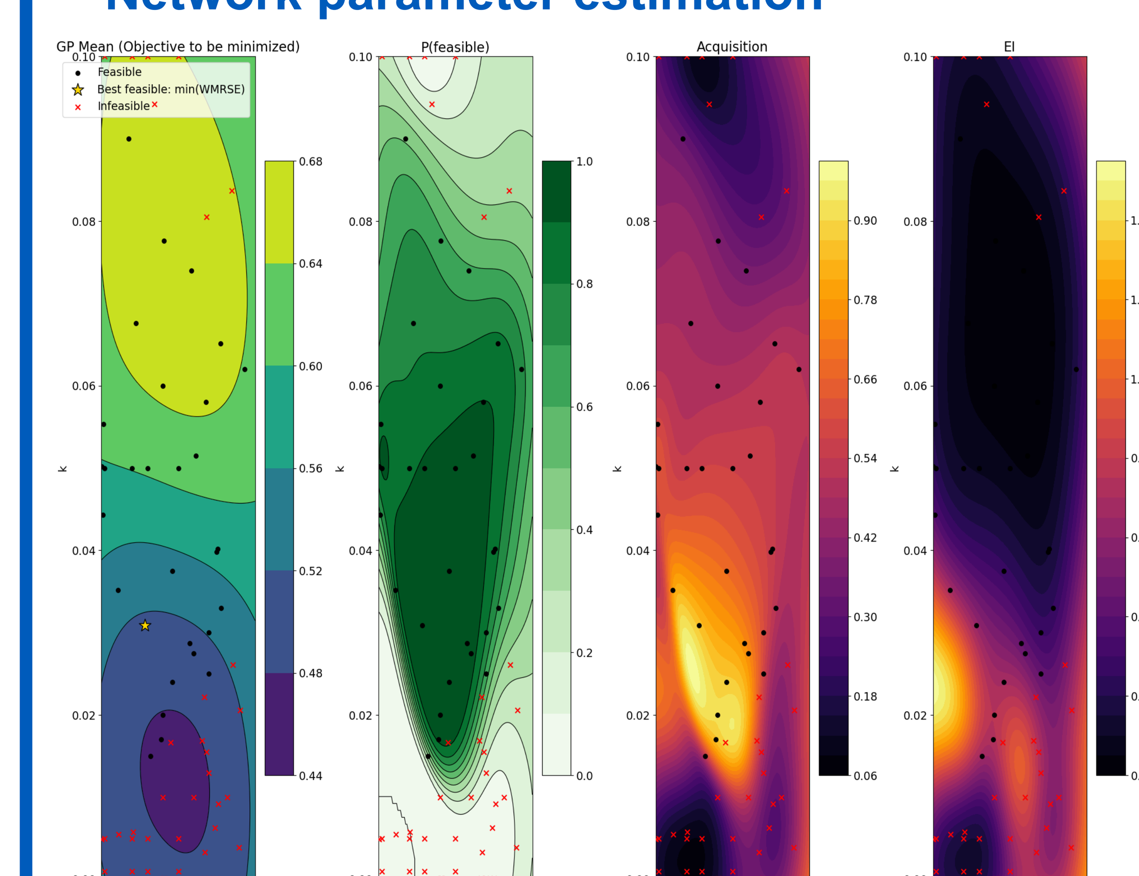
Results



Spheroid parameter estimation



Network parameter estimation



- Network cells invade farther and faster than spheroids, the gap widening over time
- Both fit at low k, but the spheroid sits at k = 0 while the network needs k ≈ 0.03
- The spheroid's feasible region is a tight low-jcc, low-k pocket; the network's parameter space is much broader