PSE PRESS

**FOPAM 2026**
*Proceedings of the 3rd Foundations of Process/Product Analytics and Machine Learning*
Decatur, USA, July 26-30 (2026)

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

Aneesh Krishna

*University at Buffalo, The State University of New York, Institute for Artificial Intelligence and Data Science*

## ABSTRACT

Cell migration and invasion are key processes underlying cancer metastasis, driven by cell–cell adhesion, chemotaxis, and matrix remodeling. Agent-based models (ABMs) are a powerful computational approach for simulating these complex, multicellular behaviors. In an ABM, each cell is represented as an autonomous agent that follows local rules governing its interactions with neighboring cells and the surrounding environment. This bottom-up framework captures the emergent, heterogeneous dynamics of cell populations that continuum models cannot easily reproduce. CompuCell3D is a widely used platform for ABMs in biological cell systems; however, a critical limitation is that it and most other ABM tools do not natively support systematic, black-box parameter estimation for their stochastic simulations. Identifying biologically meaningful parameter sets for ABMs is especially challenging because the stochastic nature of these simulations means that repeated runs with identical parameters produce different outputs, making it difficult to reliably assess how well any given parameter set matches experimental data. Existing calibration strategies, such as Monte Carlo sampling and grid search, address this challenge by simply running the model many times across a broad parameter space; an approach that is computationally expensive and scales poorly as model complexity grows.

What is needed is a more efficient approach that intelligently navigates the parameter space, reducing the number of costly simulations required to achieve accurate calibration. This work focuses on parameter estimation for agent-based modeling of collective cell invasion for two cancer cell phenotypes: a network (invasive) phenotype and a spheroid (non-invasive) phenotype emerging from growth of tumor spheroids simulated using CompuCell3D. We adopt a data-centric approach based on Gaussian process surrogate models and Bayesian optimization (BO). Unlike traditional sampling methods that treat parameter estimation as exhaustive search, BO constructs a probabilistic surrogate model of the simulation's behavior and uses it to sequentially propose new parameter sets that are expected to provide maximal information gain. This allows the framework to efficiently operate with limited simulation data, dramatically reducing the number of model evaluations needed. Leveraging simulation outputs alongside experimental data for invasion dynamics and circularity of cell clusters, the method iterates through sequential rounds of parameter selection, ABM simulation, and comparison with experimental observations.

The calibrated model successfully reproduces experimental invasion dynamics and circularity across both phenotypes. This framework provides a broadly applicable, efficient solution for parameter estimation in stochastic ABMs, addressing a gap in existing modeling tools for ABM calibration and reducing the need for extensive trial-and-error simulation of expensive models.