JOURNAL OF UNDERGRADUATE RESEARCH

Control Methods for Dynamic Cellular Simulations of Fibrosis

Rohun Gargya

College of Medicine, University of Florida

Dr. Luis Sordo Vieira, Department of Medicine

Abstract

Agent-based modeling is widely used to simulate biological systems in fields such as mathematical oncology and systems immunology. However, due to their simulation-based nature, control methods are often scarce or computationally expensive, limiting their application in high-dimensional and dynamic environments. In this paper, we present a software-based approach that efficiently approximates these computationally expensive simulations, thereby enabling faster and more practical analysis for complex biological models. Our methodology integrates a suite of analysis techniques - including Random Forest Regressor, Sobol Analysis, SHAP (Shapley Additive Explanations), and Monte Carlo simulations - to systematically identify and quantify the key parameters that govern model behavior and biological outcomes. This hybrid framework leverages the strengths of both machine learning and mathematical analysis, reducing the data requirements compared to traditional methods while enhancing the interpretability of the simulation results. To evaluate our approach, we apply these techniques to an existing Agent-Based Model (ABM) coupled with Ordinary Differential Equations (ODEs). This integrated multi-scale framework simulates dynamic cellular interactions in fibrosis progression alongside detailed intracellular processes, providing a robust test case for our analysis strategy. Our preliminary results show that the combined methodology accurately isolates key parameter families driving significant biological behaviors, such as collagen production in fibrosis, while minimizing the number of simulation runs, reducing computational cost.

Keywords: Sensitivity Analaysis, Agent-Based Model (ABM), Fibrosis

Introduction

Agent-based models (ABMs) have emerged as powerful computational tools for simulating complex biological systems. By modeling individual entities and their interactions, ABMs offer a detailed and flexible representation of dynamic processes in fields such as mathematical oncology, systems immunology, and tissue modeling. These models are particularly valuable for capturing emergent behaviors arising from cellular and molecular interactions, providing insights into disease progression, treatment responses, and environmental influences.

However, the strengths of ABMs come with inherent challenges. Due to their simulation-based nature, ABMs are computationally expensive, especially when applied to high-dimensional systems or when numerous simulation runs are required to capture variability and uncertainty. This computational burden often limits their application in scenarios demanding real-time analysis or large-scale parameter exploration. Additionally, traditional control and optimization methods are rarely feasible due to the stochastic and non-linear dynamics inherent in ABMs.

To address these challenges, we present a software-based approach that significantly reduces the computational cost of ABM simulations while maintaining high fidelity to the original model dynamics. Our solution integrates advanced machine learning techniques and mathematical analysis to approximate the behavior of complex ABMs, enabling more practical and scalable investigations. Specifically, our hybrid framework leverages Random Forest Regressor for predictive modeling, Sobol Analysis for sensitivity quantification, SHAP (Shapley Additive Explanations) for interpretability, and Monte Carlo simulations for uncertainty analysis. By combining these techniques, we reduce the data requirements and computational effort typically associated with large-scale simulations.

To demonstrate the effectiveness of our approach, we apply our framework to an established agent-based model that simulates fibrosis progression, incorporating detailed cellular interactions and intracellular processes through a combination of ABMs and ordinary differential equations (ODEs). This integrated multi-scale model captures dynamic cellular interactions contributing to collagen production, a critical factor in fibrosis development. Through our methodology, we systematically identify key parameter families that drive significant biological behaviors while minimizing the number of simulation runs required to obtain reliable results.

Methods/Body

Model Description

Our study leverages an existing fibrosis model comprising two components: a logic-based ODE model and an agent-based model. The logic-based ODE model is an intracellular signaling network includes 91 nodes and 142 reactions, constructed from experimental data. Node activities are governed by Hill ODEs with default parameters (weight = 1, Hill coefficient = 1.4, EC50 = 0.6). It processes input parameters (IL-6, IL-1β, TNFα, TGFβ) from the ABM to output collagen mRNA and other proteins (Rikard et al., 2019). The agent-based model simulates cellular interactions in the extracellular space, supplying input parameters to the ODE model. Outputs from the ODE are converted from normalized weights (0–1, representing 0–100% receptor activation) to ABM-compatible units (pg/mL) for iterative simulation over multiple time steps, equivalent to 6 weeks (Zeigler et al., 2016).

Modifications on Model

Originally implemented in MATLAB, we converted the model to Python to leverage robust libraries for analysis and control. This multi-scale framework accelerates simulations, compressing a 6-week process into hours, though interpreting results requires extensive runs to explore parameter combinations affecting collagen output.

Analysis Techniques

To tackle the model's complexity and high-dimensional parameter space, we employed a suite of sensitivity analysis methods:

1. **Random Forest Regressor:** A machine learning ensemble method that constructs multiple decision trees during training and outputs the average prediction of the individual trees. In our context, it performs sensitivity analysis by learning the relationship between the model's input parameters and its outputs (e.g., collagen production). It was selected for its prowess in modeling nonlinear relationships with limited data, we trained it on just 20 simulation runs to predict collagen output from input parameters.
2. **Sobol Sensitivity Analysis:** A variance-based method that decomposes the output variance into contributions from individual parameters and their interactions. This global

sensitivity analysis method quantifies individual and interactive parameter contributions to output variance. With 11 parameters (D = 11) and a base sample size of N = 2, we ran 48 simulations [N × (2D + 2)] to compute first-order and total-order indices.

3. **SHAP (Shapley Additive Explanations):** A powerful method used to interpret machine learning models by attributing the contribution of each feature to a model's predictions. It is based on Shapley values from cooperative game theory, which fairly distribute the contribution of each feature by considering all possible feature subsets. Applied to interpret the Random Forest model, SHAP provides local and global insights into parameter importance and interactions.
4. **Monte Carlo Simulations:** It serve as a robust statistical method to assess the impact of uncertainty in the model parameters. Used for uncertainty analysis, these simulations visualize parameter distributions and correlations with collagen output. By analyzing the ensemble of simulation results, we can compute summary statistics (e.g., mean, variance, confidence intervals) for the outputs. These statistics provide quantitative measures of how sensitive the model is to changes in input parameters.

These techniques were chosen for their complementary strengths: Random Forest for efficient prediction, Sobol for comprehensive sensitivity, SHAP for interpretability, and Monte Carlo for uncertainty assessment.

Results

Random Forest Results

Using only 20 simulation runs, the Random Forest Regressor accurately identified key parameters influencing collagen output (Antoniadis, Lambert-Lacroix, & Poggi, 2020). Feature importance analysis (Figure 1) highlighted mechanical and forskolin as the most dominant factors overall. Among controllable parameters (IL-6, IL-1β, TNFα, TGFβ), TGFβ and IL-6 stood out as the most influential.

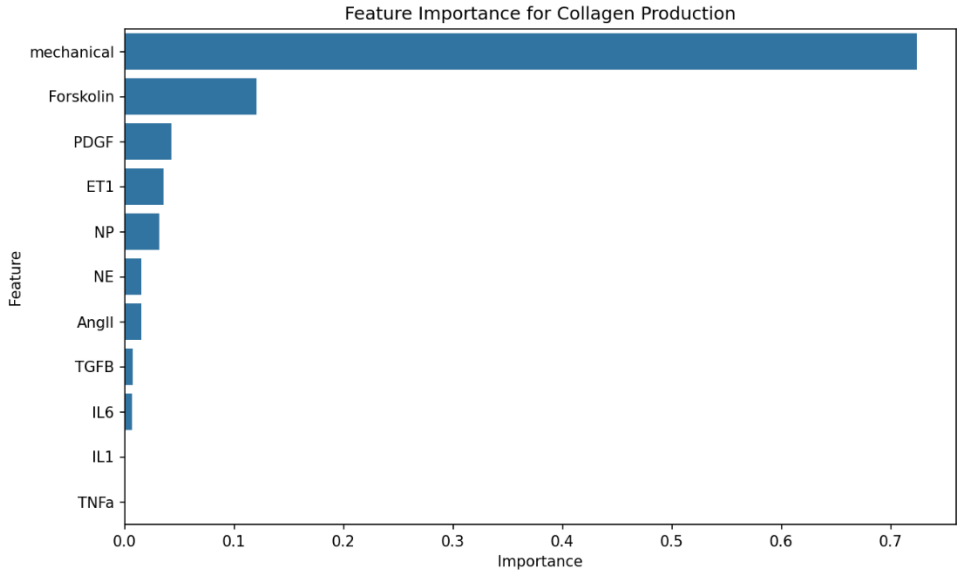

Figure 1. Feature importance plot from Random Forest Regressor, showing mechanical and forskolin as primary influencers of collagen output, with TGFβ and IL-6 leading among controllable parameters.

Sobol Analysis Results

Sobol Sensitivity Analysis, conducted over 48 simulation runs, reinforced these findings (Chen et al., 2024). First-order indices, which measure direct parameter effects, and total-order indices, which include interactions, both identified mechanical and forskolin as top contributors (Figure 2). Among the parameters of interest, TGFβ exhibited the highest sensitivity, underscoring its critical role in collagen production. High total-order indices relative to first-order indices suggest significant parameter interactions.

```
Sobol Sensitivity Analysis Results:
```

```
First Order Sensitivity:
AngII: 0.005159
TGFB: -0.011195
mechanical: 0.731732
IL6: -0.000056
IL1: -0.000437
TNFa: 0.000630
NE: 0.014879
PDGF: 0.000065
ET1: 0.007657
NP: 0.118277
Forskolin: 0.248073
```

```
Total Order Sensitivity:
AngII: 0.000157
TGFB: 0.000493
mechanical: 2.167955
IL6: 0.000000
IL1: 0.000000
TNFa: 0.000002
NE: 0.000839
PDGF: 0.000000
ET1: 0.000097
NP: 0.054656
Forskolin: 0.235522
```

Figure 2. Sobol Sensitivity Analysis results, listing first-order and total-order indices for model parameters, confirming the prominence of mechanical and forskolin, with TGFβ as the most impactful controllable parameter.

SHAP Analysis

- SHAP analysis elucidated parameter interactions, focusing on TGFβ and IL-6 (Duan & Ökten, 2023). The interaction plot (Figure 3) showed a weak positive effect, with SHAP values ranging from -0.0004 to 0.0004. Higher TGFβ values correlated with slightly increased collagen predictions when IL-6 was also elevated, though the interaction's magnitude remained modest.

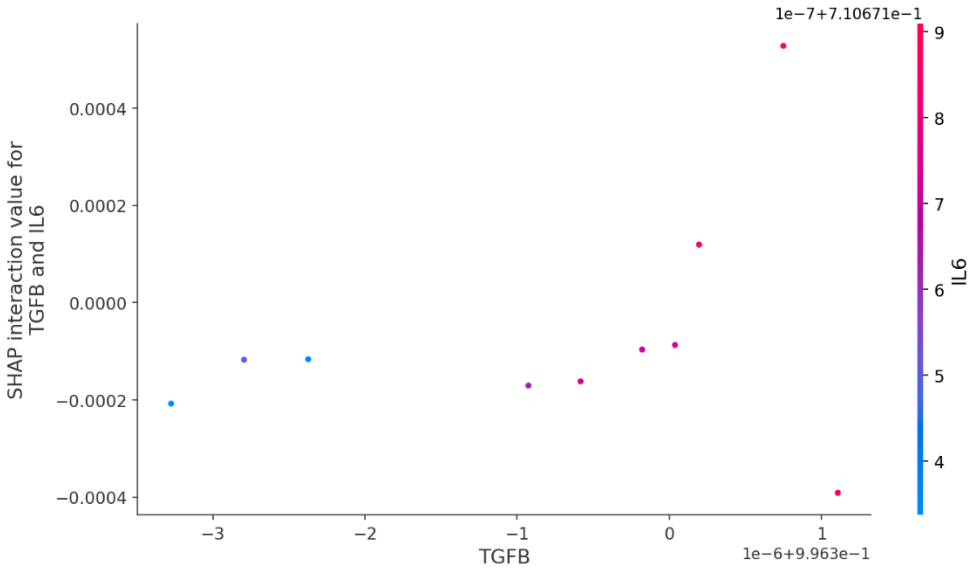

Figure 3. SHAP interaction plot for TGFβ and IL-6, illustrating a weak positive interaction affecting collagen output, with values clustering at higher TGFβ levels.

Monte Carlo Simulations

Monte Carlo simulations provided a pairplot (Figure 4) of parameter distributions and relationships with collagen output (Ökten & Liu, 2021). Kernel Density Estimation (KDE) plots along the diagonal revealed parameter spread, while scatter plots highlighted a strong correlation between IL-6 and collagen, aligning with prior analyses.

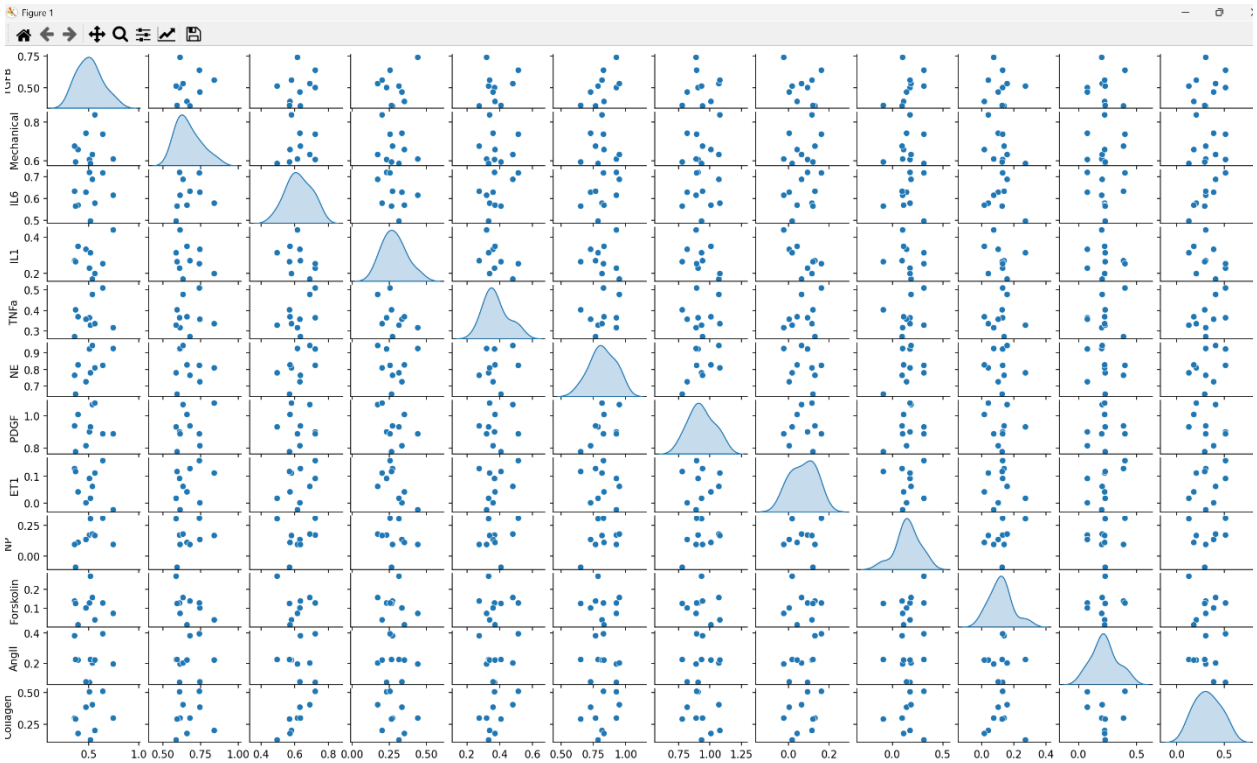

Figure 4. Monte Carlo pairplot showing parameter distributions (diagonal KDE plots) and correlations with collagen output, with a notable IL-6-collagen relationship.

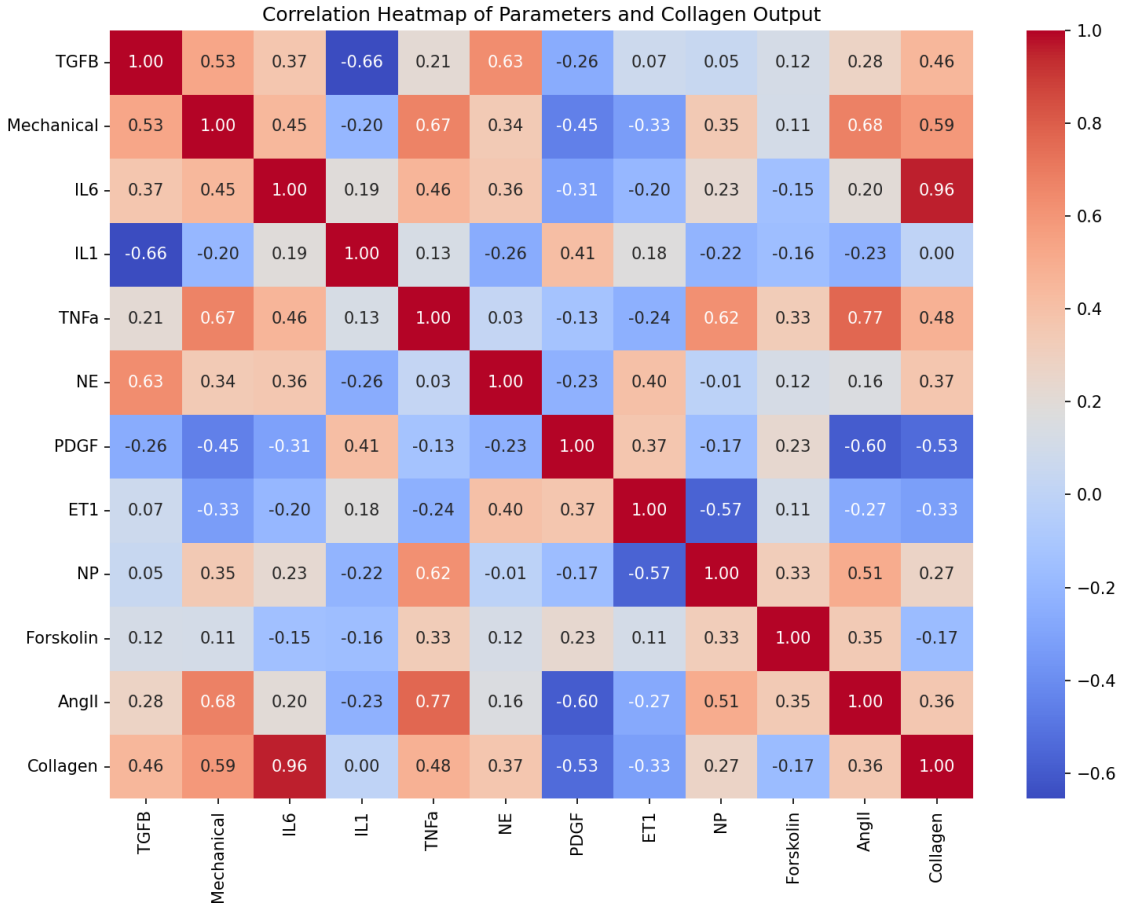

Figure 5. This shows a clearer correlation than the Monte Carlo pairplot above. It shows parameter corrleation and correlations with collagen output, with a notable IL-6-collagen relationship.

Discussion/Conclusion

This study introduces a hybrid framework that melds machine learning and mathematical analysis to streamline the simulation and interpretation of computationally intensive agent-based models (ABMs) in biological systems. Applied to an ABM-ODE model of fibrosis, our approach efficiently identified key parameters driving collagen production—a central feature of fibrosis progression—while minimizing computational overhead.

The Random Forest Regressor, trained on just 20 simulations, captured nonlinear parameter relationships, pinpointing mechanical and forskolin as the most influential factors overall, and TGFβ and IL-6 as the dominant controllable parameters. Sobol Sensitivity Analysis, with 48 runs, offered a global view, confirming these findings and revealing significant parameter interactions via total-order indices. SHAP analysis enriched interpretability,

uncovering a subtle positive interaction between TGFβ and IL-6, while Monte Carlo simulations underscored IL-6's correlation with collagen levels, enhancing uncertainty assessment.

This integrated methodology drastically reduces the simulation burden of traditional ABMs, providing actionable insights into complex biological dynamics with fewer resources. By isolating critical parameters like TGFβ and IL-6, our framework supports targeted exploration of therapeutic strategies in fibrosis and beyond.

The approach's success in this context highlights its potential for broader application across mathematical oncology, systems immunology, and other ABM-driven fields. Future enhancements could incorporate real-time data integration, optimize treatment scenarios, or scale to higher-dimensional models. Ultimately, this work advances the scalability and interpretability of computational modeling, forging a path toward practical biomedical solutions.

Acknowledgements

Thank you so much Dr. Luis Sordo Vieira for this assistance and help in conducting research over the past couple semetsers. I have learned so much from you and your guidance has been invaluable. I truly appreciate the time and effort you've invested in helping me grow academically and professionally. Thank you again for your support!

References

Rikard, S. M., Athey, T. L., Nelson, A. R., Saucerman, J. J., & et al. (2019). Multiscale coupling of an agent-based model of tissue fibrosis and a logic-based model of intracellular signaling. *Frontiers in Physiology, 10*, 1481. https://doi.org/10.3389/fphys.2019.01481

Zeigler, A. C., Richardson, W. J., Holmes, J. W., & Saucerman, J. J. (2016). A computational model of cardiac fibroblast signaling predicts context-dependent drivers of myofibroblast differentiation. *Journal of Molecular and Cellular Cardiology, 94*, 72-81. https://doi.org/10.1016/j.yjmcc.2016.03.008

Chen, L., Xu, Z., Huang, D., & Chen, Z. (2024). An improved Sobol sensitivity analysis method. *Journal of Physics: Conference Series*, *2747*(1), 012025. https://iopscience.iop.org/article/10.1088/1742-6596/2747/1/012025/pdf

Antoniadis, A., Lambert-Lacroix, S., & Poggi, J. M. (2020). Random forests for global sensitivity analysis: A selective review. *Reliability Engineering & System Safety*, *196*, 107312. https://doi.org/10.1016/j.ress.2020.107312

Ökten, A. G., & Liu, Y. (2021). Randomized quasi-Monte Carlo methods in global sensitivity analysis. *Reliability Engineering & System Safety, 210*, 107520. https://doi.org/10.1016/j.ress.2021.107520

Duan, H., & Ökten, G. (2023). Derivative-based Shapley value for global sensitivity analysis and machine learning explainability. *arXiv Preprint*, arXiv:2303.15183. https://arxiv.org/pdf/2303.15183