Kinetic Model of Terminal B Cell Differentiation: Parameter Estimation and Bistability Analysis

Karolina Chłopicka [15716546] (karolina.chlopicka@student.uva.nl)

GitHub repository: `https://github.com/karolina-chl/bio_informatics.git`

Abstract

Terminal differentiation of B cells into antibody-secreting plasma cells is governed by a tightly regulated transcriptional network involving BLIMP1, BCL6, and IRF4. The main interest of this study revolved around a kinetic model of B cell differentiation, originally developed by Martínez et al. (2012). The model was numerically solved, and parameter fitting was performed. The aim was to reproduce the results and investigate the effects of parameter fitting under different constraints. Using gene expression data, parameter estimation was performed with and without enforcing bistability conditions. Enforcing bistability constraints improved biological plausibility by producing more switch-like behavior between the Germinal Center state and the Plasma Cell state. The study also discovered that some parameters are more sensitive than others to the width of the search bounds. While the model recapitulates key dynamics, the work highlighted a major shortcoming: the parameter estimation did not reproduce the same results as the original study.

Contents

1 Introduction

Understanding how B cells undergo terminal differentiation into plasma cells is essential for understanding the mechanisms of immunity. This process is regulated by a handful of transcriptional factors, including but not limited to BLIMP1, BCL6, and IRF4. Regulatory interactions determine whether B cells remain in the germinal center (GC) and when it transitions to a plasma cell (PC).

In 2012, Martínez et al. (2012) proposed a quantitative kinetic model capturing this transition using a system of nonlinear ordinary differential equations. The model incorporates known biological interactions and ensures the possibility of bistability—a mechanism critical for ensuring irreversible cell fate decisions.

The primary goal of this project was to reproduce the results of the Martínez model and carry out parameter estimation analyses. Additionally, the analyses was extended beyond the scope of Martínez's paper by investigating the impact of the width of the bounds on parameter estimation.

2 Theory

2.1 Biological background

When the body encounters pathyogens, it activates special immune cells called **B cells**. These B cells go through a rapid transformation process in special areas of the immune system called **germinal centers (GCs)**.

The **germinal center** has two main parts:

- The **dark zone**, where B cells mutate and improve their antibodies.
- The **light zone**, where B cells are tested—only the best ones that recognize the germ survive and move on.

To become either:

- **Memory B cells** (which remember the germ for future attacks), or
- **Plasma cells** (which produce large amounts of antibodies to fight the germ),

Key **proteins** that control this change:

- **BCL6**: Keeps B cells in the germinal center stage. It must be turned off for the B cell to move on. Highly expressed in GC B cells; represses genes to support proliferation and tolerance of SHM. Essential for GC formation.
- **IRF4**: Helps push B cells to become plasma cells. Critical for both GC responses and ASC differentiation. Acts dose-dependently (low = GC/CSR, high = ASC fate).
- **BLIMP1**: The final switch that makes a B cell become a plasma cell. Promotes terminal B cell differentiation. It is required for mature plasma cells but not initial ASC commitment. Works with IRF4/AP-1.

These proteins interact in a very controlled way—like a complex switch system—to ensure the right cells change at the right time. If this system breaks, it can lead to immune diseases or cancer (like B-cell lymphomas).

There are four cellular fates of B cells — division, death, CSR development and ASC development (Nutt et al. (2015)). How initially homogeneous cell types generate so many different cell types over a short period of time remains one of the compelling mysteries associated with antibody regulation. The proteins that play a key role in the model are BCL6, BLIMP1, and IRF4.

2.2 The Model

The aim of the model is to mimic a gene regulatory module controlling the germinal center pathway exit. The model is defined by the system of equations given in the Equation 1.

$$\begin{aligned}\frac{dp}{dt} &= \mu_p + \sigma_p \frac{k_b^2}{k_b^2 + b^2} + \sigma_p \frac{r^2}{k_r^2 + r^2} - \lambda_p p \\ \frac{db}{dt} &= \mu_b + \sigma_b \frac{k_p^2}{k_p^2 + p^2} \cdot \frac{k_b^2}{k_b^2 + b^2} \cdot \frac{k_r^2}{k_r^2 + r^2} - (\lambda_b + BCR)b \\ \frac{dr}{dt} &= \mu_r + \sigma_r \frac{r^2}{k_r^2 + r^2} + CD40 - \lambda_r r\end{aligned} \tag{1}$$

In the model in Equation 1, p, b, and r stand for the protein levels of BLIMP1, BCL6, and IRF4, respectively.

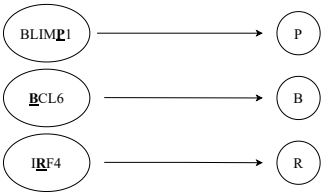

Figure 1: Notation used in the model for BLIMP1, BCL6, and IRF4 protein

The level of each of the proteins is further influenced by four parameters: λ, k, σ, and μ.

- λ_p, λ_b, λ_r indicate the degradation rates;
- μ_p, μ_b, μ_r are parameters indicating the basal production rate for each protein;
- k_p, k_b, k_r stand for the dissociation constant;
- σ_p, σ_b, σ_r indicate the maximum transcription rate for each protein.

Counting all the parameters above, in the model, there are 12 parameters total. Additionally, BCR and CD40 regulatory signals are included in the model. The phenomenological form of them is:

$$\begin{aligned}BCR &= brc_0 \frac{k_b^2}{k_b^2 + b^2} \\ CD40 &= cd_0 \frac{k_b^2}{k_b^2 + b^2}\end{aligned} \tag{2}$$

For the purpose of the model form Equation 1, it is assumed that the level of protein for each transcription factor is roughly proportional to the mRNA levels. It is also assumed that each transcriptional interaction can be modeled by the Hill function with a cooperative coefficient of 2. Finally, the last assumption about the model is that each transcription factor has the same binding affinity.

The model is experiencing the bistable regime under the following conditions:

$$\begin{aligned}\beta &= \frac{\mu_r + CD40 + \sigma_r}{\lambda_r k_r} > \sqrt{3} \\ \beta^3 - (\beta^2 - 3)^{\frac{2}{3}} + 9\beta &< \frac{27\sigma_r}{2\lambda_r k_r} \\ \beta^3 + (\beta^2 - 3)^{\frac{2}{3}} + 9\beta &> \frac{27\sigma_r}{2\lambda_r k_r}\end{aligned} \tag{3}$$

3 Methods

3.1 Numerically solving ODE system

The kinetic model described in the Equation 1 is given as a system of ordinary differential equations. To observe the behavior of the model over time, we need to numerically solve it.
Initial conditions for BLIMP1, BCL6, and IRF4, that were fed to the model are listed in the Table 1. The model was numerically integrated using the `odeint` function in Python. This function is de-

BLIMP1	BCL6	IRF4
0.1	5	0.1

Table 1: Initial conditions for BLIMP1, BCL6, IRF4 taken when numerically solving the model.

signed to solve the initial value problems for first-order ordinary differential equations and it's a part of `scipy.integrate`.

BLIMP1, BCL6, IRF4, as well as BCR and CD40 are dependent on time. All 12 other parameters used in the system are fixed; their value are constant over time. The specific values for each parameter were directly taken from the Martínez et al. (2012) paper and are presented in the Table 2

Parameter	**Value**	**Description**
μ_p	10^{-6}	Basal transcription rate
μ_b	2	Basal transcription rate
μ_r	0.1	Basal transcription rate
σ_p	9	Maximum induced transcription rate
σ_b	100	Maximum induced transcription rate
σ_r	2.6	Maximum induced transcription rate
k_p	1	Dissociation constant:
k_b	1	Dissociation constant
k_r	1	Dissociation constant
λ_p	1	Degradation rate
λ_b	1	Degradation rate
λ_r	1	Degradation rate

Table 2: Parameter values used in the model by Martínez et al. (2012).

3.2 Data

Parameter fitting was performed based on micro-array gene expression datasets from normal, transformed, and experimentally manipulated human B cells. The data consisted of 45 data points, 15 for each of the proteins, i.e., 15 for BLIMP1, 15 for BCL6, and 15 for IRF4. Each of 15 data points consists of 5 data points for Centroblast, 5 for Centrocyte, and 5 for Plasma Cell. All the data points used for parameter fitting are in the Table 3. Overall, the data consists of Germinal Center related gene expression data - CB and CC data points and Plasma Cell related gene expression data - PC data points.
The only changes in data done, before performing parameter fitting was to rename the columns. The data was not subject to any further manipulation before performing parameter fitting.

Category	Sample	BLIMP1	BCL6	IRF4
Germinal Center	CB	2.72073	12.15459528	3.712181782
	CB	2.65367	11.90625967	5.414852183
	CB	2.70401	12.21045968	5.063578039
	CB	2.62238	12.17949077	4.685991127
	CB	2.61155	12.15459528	5.997029122
	CC	2.5806	11.93114681	5.831566224
	CC	2.55039	12.17949077	5.702515218
	CC	2.55361	12.07294855	7.124772756
	CC	2.62904	12.13623394	6.59974542
	CC	2.57203	12.00261001	6.374098322
Plasma Cell	PC	10.1235	7.035853737	9.103791089
	PC	8.44254	7.341818328	9.627906128
	PC	8.80529	6.773579006	10.24778136
	PC	10.1821	8.432358776	9.6111935
	PC	9.81398	7.258304827	9.964856276

Table 3: Microarray gene expression data for BLIMP1, BCL6, and IRF4.

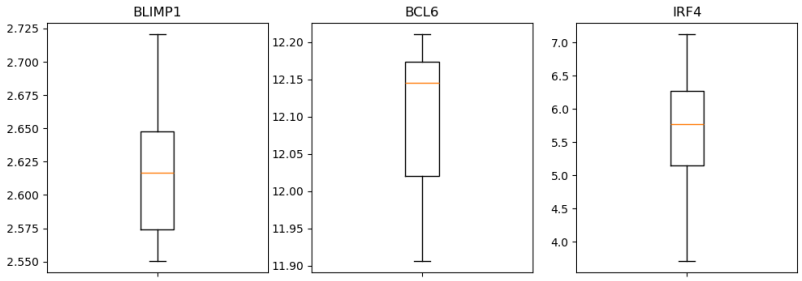

Figure 2: Boxplot of germinal Center related data for BLIMP1, BCL6, and IRF4.

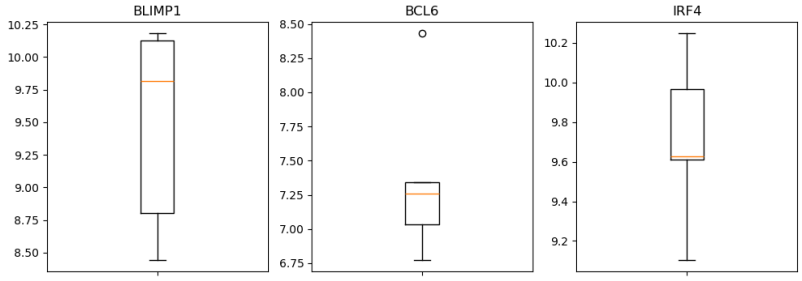

Figure 3: Boxplot of plasma Cell related data for BLIMP1, BCL6, and IRF4.

3.3 Parameter fitting

Simulated annealing is a method used to find a good solution to a difficult optimization problem. It's based on the idea of how metals are cooled slowly so that their atoms settle into a stable, low-energy state. In optimization, this concept is used to help find a solution that gives the best possible value (like the lowest cost or highest score) in problems where there might be many local optima—places that look good but aren't the absolute best.

The algorithm starts with an initial solution and a high "temperature" value. It then makes small random changes to the solution to explore the space of possible answers. If a change leads to a better

solution, it accepts it. If the change makes things worse, it might still accept it, but only with a certain probability. This probability decreases as the temperature goes down. Over time, as the temperature gets lower, the algorithm becomes less likely to accept worse solutions and focuses more on refining what it has. This balance between exploration and refinement is what allows simulated annealing to sometimes find better solutions than more straightforward methods.

`dual.annealing` from `scipy.optimize` package, implemented in Python's, combines simulated annealing with a local search method. The idea is to get the best of both worlds: global exploration and local precision. The algorithm first uses a global simulated annealing strategy to broadly explore the solution space and avoid getting stuck in poor-quality solutions. Then it applies a local optimizer to fine-tune the solution it finds. This makes the algorithm more robust and efficient for finding near-optimal solutions in difficult optimization problems with many variables or complex landscapes. It's particularly useful when the search space is large and contains many local minima.

The goal of the optimization algorithm is to minimize the distance between the estimation and the data points. The objective that is being minimized is the error function. The error function evaluates the difference at two critical time points: $t = 25$ (representing the GC state) and $t = 175$ (representing the PC state). At those critical points, the error is calculated as a difference between the model estimation and the data. For the purpose of this calculation the below function is used:

$$\text{Error}_{\text{GC}} = \sqrt{(P_{25} - \hat{P}_{\text{GC}})^2} + \sqrt{(B_{25} - \hat{B}_{\text{GC}})^2} + \sqrt{(R_{25} - \hat{R}_{\text{GC}})^2} \tag{4}$$

$$\text{Error}_{\text{PC}} = \sqrt{(P_{175} - \hat{P}_{\text{PC}})^2} + \sqrt{(B_{175} - \hat{B}_{\text{PC}})^2} + \sqrt{(R_{175} - \hat{R}_{\text{PC}})^2} \tag{5}$$

$$\text{Total Error} = \text{Error}_{\text{GC}} + \text{Error}_{\text{PC}} \tag{6}$$

Parameter fitting was done both with and without bistability conditions. This was motivated by the discrepancy between model trajectories and the real-life data. When doing the parameter fitting with bistability constraints, Equation 4, Equation 5, and Equation 6 were applied, but parameter sets not following the bistability conditions were punished with a large error, resulting in only accepting bistable parameters.

3.4 Experimenting with the width of the bounds

This experiment was designed to evaluate how the parameter estimation changes depending on the width bounds. The goal of this experiment is to determine the robustness and sensitivity of parameter estimation and whether the estimation depends on the bounds chosen. This experiment demonstrates how the range of initial parameter guesses (bounds) influences the accuracy of parameter estimation in models with constraints such as bistability. The results can help identify the sensitivity of each parameter to initial conditions and guide the selection of informed prior bounds in future estimations.

Let $\boldsymbol{\theta} = (\mu_p, \mu_b, \mu_r, \sigma_p, \sigma_b, \sigma_r)$ denote the parameter values estimated in the paper Martínez et al. (2012). The bounds are gradually increased using the, the formula for the upper and lower bounds as given below:

$$\begin{aligned} \text{Lower bound} &= \max(0, \theta_i - \theta_i \cdot p) \\ \text{Upper bound} &= \theta_i + \theta_i \cdot p \end{aligned} \tag{7}$$

where p is the percentage by which the bound is extended. For the purpose of the experiment, we

took p equal to:

$$p \in \{0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5\}$$

For each p value following steps were followed.

1. Bounds are generated for each of the parameters $\mu_p, \mu_b, \mu_r, \sigma_p, \sigma_b, \sigma_r$ using Equation 7.
2. A parameter fitting function is called, and parameter estimation is carried out with the determined bounds and bistability conditions.
3. Steps 1 and 2 are repeated for each bound size and for each run optimal parameters and the error are recorded.

4 Results and Discussion

4.1 Numerically solving ODE system

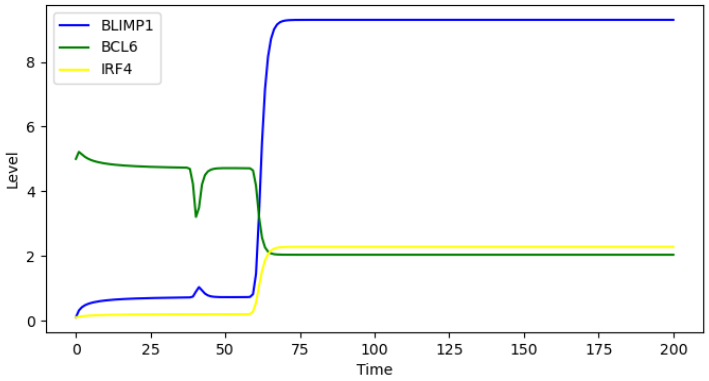

Figure 4: Visualization of the system of ODE equations for terminal B-cell differentiation numerically integrated.

Figure 4 presents the time-dependent dynamics of the system. Initially, BCL6 is highly expressed, while IRF4 and BLIMP1 levels are much lower. As time progresses, the level of BCL6 declines, while the levels of BLIMP1 and IRF4 increase. The increase in BLIMP1 is more rapid than that of IRF4.

Considering the overall behavior, there are two stable regimes in the model. The first stable regime corresponds to the Germinal Center state (GC state), and the second corresponds to the Plasma Cell state (PC state). The GC state is marked by high BCL6 expression and low BLIMP1/IRF4 levels. In contrast, the PC state features upregulation of BLIMP1 and IRF4, with strong repression of BCL6.

The behavior of BCL6, IRF4, and BLIMP1 also reflects what is expected from a biological point of view. The observed shift is consistent with known biological transitions and validates the model's ability to reproduce known features of B cell differentiation. BCL6 is highly expressed in Germinal Center B cells, but must be down-regulated to allow differentiation. IRF4 leads to the transcriptional repression of BCL6 and to the transactivation of BLIMP1. IRF4 binds to its own promoter, supporting a positive feedback mechanism that allows Plasma Cells to maintain high IRF4 expression. BLIMP1 drives the regulatory program associated with plasmacytic differentiation and immunoglobulin secretion.

Overall, while the model captures core system dynamics, it simplifies several biological processes. For example, BCR and CD40 pathways are represented by simple Hill-type functions. The model also assumes a constant degradation rate, set to 1 for all components: λ_p, λ_b, and λ_r. The same simplification is applied to all dissociation constants, with $k_p = k_b = k_r = 1$.

4.2 Parameter estimation with and without bistability conditions

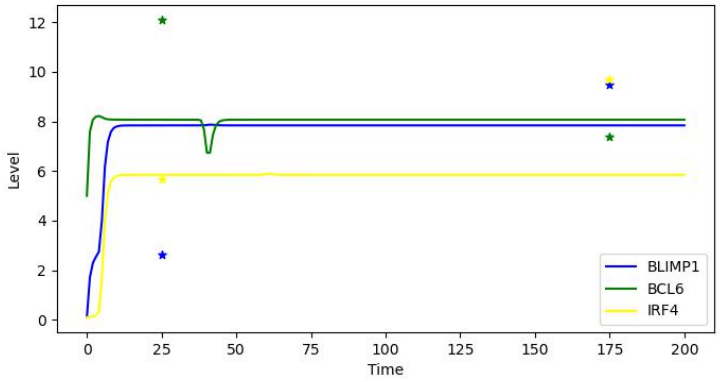

Figure 5: Parameter estimation without bistability conditions

Figure 5 shows the time-course output using the estimated parameters. The model successfully captured the overall transition dynamics from the GC to PC state. As expected, BCL6 expression was initially high and declined over time, while BLIMP1 and IRF4 levels increased. However, the transitions between states appeared more gradual and less switch-like.

This result suggests that, although accurate in matching endpoint expression values, the model without bistability constraints lacks the sharp commitment behavior observed in biological differentiation. In real B-cell development, the transition from GC to PC state is not only directional but also irreversible, often associated with bistability and hysteresis—features that are not guaranteed under unconstrained parameter fitting.

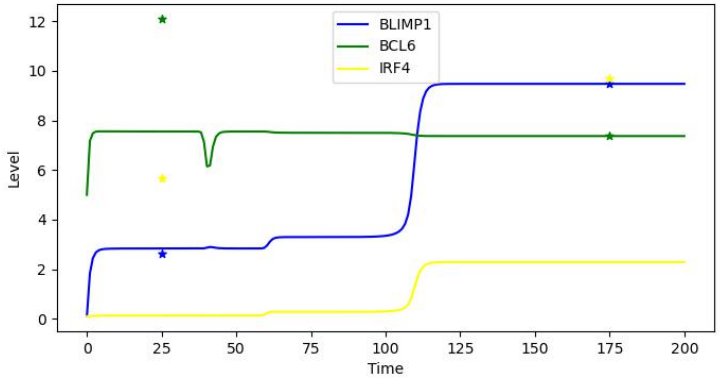

Figure 6: Parameter estimation with bistability conditions

To improve biological realism, a second parameter estimation was performed with added constraints enforcing bistability. Specifically, the model required satisfaction of analytical bistability conditions on IRF4 dynamics, based on the presence of a positive feedback loop and the influence of CD40 signaling. These constraints ensure that the system can exhibit two distinct stable states (GC and PC), separated by a threshold that enables a switch-like transition.

Figure 6 illustrates the resulting dynamics under bistability-constrained parameters. As in the unconstrained case, BCL6 levels were initially high and decreased over time, while IRF4 and BLIMP1 increased. However, the transition was more abrupt, consistent with the bistable switching mechanism. The model better reflects the biological irreversibility of the differentiation process, where once IRF4

reaches a critical level, BLIMP1 activation is reinforced and BCL6 repression is sustained.

The inclusion of bistability constraints improved the model's mechanistic fidelity without compromising fit quality. The results demonstrate that enforcing biologically motivated constraints can guide the model towards more interpretable and realistic parameter regimes, enhancing both predictive power and alignment with known regulatory mechanisms.

Parameter	**Martínez et al. (2012)**	**Estimated (No Bistability)**	**Estimated (Bistability)**
μ_p	1×10^{-6}	0.27	2.88
μ_b	2	9.44	7.37
μ_r	0.1	0.02	0.09
σ_p	9	7.50	7.54
σ_b	100	0.93	0.94
σ_r	2.6	8.51	2.62
Final error	—	15.64	18.17

Table 4: Comparison of estimated parameter values with and without bistability constraints, and reference values from Martínez et al. (2012)

4.3 Experimenting with the width of the bounds

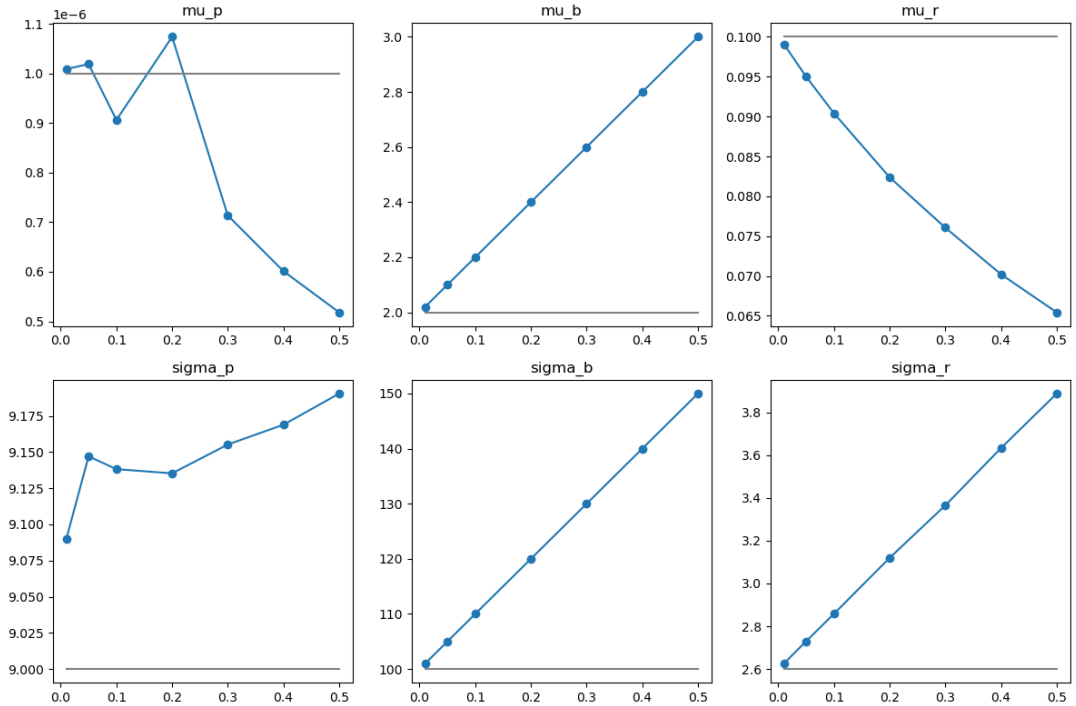

Figure 7: Experimenting with the width of the bounds

The results show different behavior among parameters. Notably, μ_b, σ_b, and σ_r, μ_r exhibit high sensitivity to the expansion of bounds. Their estimated values change approximately linearly with the bound width. This indicates that these parameters are not tightly constrained by the error function and can drift significantly while still satisfying the bistability constraints.

In contrast, the variation of σ_p parameter is much smaller. This implies that this parameter is more tightly constrained by the data and system's bistable behavior.

Interestingly, μ_p remains relatively stable for small increases in bounds but drops sharply as the bounds are enlarged by more than 20%.

Overall, this experiment highlights that no parameter remains constant when the bounds are increased. We observe that certain parameters like μ_p σ_p are less susceptible to change, while others, like μ_b, μ_r σ_b, σ_r, change almost linearly with the width of the bounds. These findings underscore the importance of careful constraint selection and support the use of biologically informed priors in systems with limited or noisy data.

5 Conclusion

In this project, the kinetic model of B cell terminal differentiation developed by Martínez et al. (2012) was successfully reimplemented, and its key dynamical features were reproduced. By numerically solving the model using the parameter values estimated by Martínez et al. (2012), it was confirmed that the basic biological behavior of BLIMP1, BCL6, and IRF4 can be replicated by the model.

The comparison between unconstrained and bistability-constrained parameter estimation demonstrated that it is not possible to simultaneously fit the data and preserve biologically realistic model behavior. During parameter estimation, there was a consistent trade-off between minimizing the error and accepting a larger error in order to satisfy the bistability constraints.

Despite the successful reimplementation, a major challenge was encountered regarding model reproducibility. It was not possible to recover the same parameter values as those reported by Martínez et al. (2012). In conclusion, ensuring consistent and transparent modeling pipelines—including access to data and code—is essential for building reliable and reusable models in systems immunology, and more broadly, in the field of computational science.

References

Martínez, M. R., Corradin, A., Klein, U., Álvarez, M. J., Toffolo, G. M., di Camillo, B., Califano, A., and Stolovitzky, G. A. (2012). Quantitative modeling of the terminal differentiation of b cells and mechanisms of lymphomagenesis. *Proceedings of the National Academy of Sciences of the United States of America*, 109(7):2672–2677.

Nutt, S. L., Hodgkin, P. D., Tarlinton, D. M., and Corcoran, L. M. (2015). The generation of antibody-secreting plasma cells. *Nature Reviews Immunology*, 15:160–171.