Learning Model Parameter Dynamics in a Combination Therapy for Bladder Cancer from Sparse Biological Data

Kayode Olumoyin
H. Lee Moffitt Cancer Center and Research Institute
Integrated Mathematical Oncology Department
Tampa, FL
kayode.olumoyin@moffitt.org

Lamees El Naqa
H. Lee Moffitt Cancer Center and Research Institute
IMO High School Internship Program
Tampa, FL
elnaqalamees@gmail.com

Katarzyna Rejniak
H. Lee Moffitt Cancer Center and Research Institute
Integrated Mathematical Oncology Department
Tampa, FL
kasia.rejniak@moffitt.org

Abstract

In a mathematical model of interacting biological organisms, where external interventions may alter behavior over time, traditional models that assume fixed parameters usually do not capture the evolving dynamics. In oncology, this is further exacerbated by the fact that experimental data are often sparse and sometimes are composed of a few time points of tumor volume. In this paper, we propose to learn time-varying interactions between cells, such as those of bladder cancer tumors and immune cells, and their response to a combination of anticancer treatments in a limited data scenario. We employ the physics-informed neural network (PINN) approach to predict possible subpopulation trajectories at time points where no observed data are available. We demonstrate that our approach is consistent with the biological explanation of subpopulation trajectories. Our method provides a framework for learning evolving interactions among biological organisms when external interventions are applied to their environment.

1 Introduction

Bladder cancer is the $9th$ leading cause of cancer death in men and women globally [4]. More than 50% of patients with intermediate and high-risk localized non-muscle invasive bladder cancer treated with bladder-sparing treatment fail these therapies. In recent years, there is a shift to immunotherapy and combination therapy in how bladder cancer is treated. One of the motivating factor is the success of immunotherapy and combination therapy in the treatment of solid tumors such as unresectable metastatic melanoma and cervical cancer [17, 12].

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: Learning from Time Series for Health.

Ordinary differential equation (ODE) models can be used to study the interactions between different types of cells in the bladder cancer microenvironment, where external interventions may alter interaction behavior over time. In [3, 10], it is suggested that these external interventions may cause the ODE model parameters to be time-varying. Another motivation for modeling some parameters as time-varying in a biological system is that, in some cases, we lack full mechanistic knowledge of the model. Here, time-varying parameters can act as surrogates for unobserved or unmodeled effects [5, 16]. Advancements in machine learning and data science have made it possible to learn complex dynamics directly from data. A successful approach in recent years is the physics-informed neural networks (PINNs), introduced in [14], in which the form of the differential equation is assumed to be known and the goal is to learn its parameters from data. Several recent advancements in modeling biological systems from data have been inspired by PINN [9, 13, 8, 11].

It is commonplace in differential equations model of oncology, that data are sparse. This sparsity occurs because: (1) data are collected at irregular time points; (2) the collected data may contain noise due to measurement errors; and (3) the interacting variables may only be partially observable. In this work, we propose a method that uses the known few discrete data points of the total tumor volume to predict admissible trajectories at unobserved time points for each of the interacting subpopulations. We demonstrate that these learned trajectories at the unobserved time points matches with the biological behavior of the subpopulations. We use spline-based interpolation to extend the observed total tumor volume. We are able to model continuous time-varying parameters from sparse data points.

2 ODE Model for a Combination Therapy

We describe an ODE model of a growing tumor exposed to a combination of two anticancer treatments: gemcitabine (GEM) chemotherapy and T cell immunotherapy, based on a study published in [2]. In this experimental study, the 1×10^5 MB49-OVA bladder cancer cells were instilled orthotopically into the mouse bladder. This was followed on day 10 by intravesical injection of $500\mu g$ of GEM to locally deplete the tumor microenvironment from the immunosuppressive cells, such as the myeloid-derived suppressive cells (MDSCs). Next, at day 14, the 5×10^6 OT-1 T cells were administered intravesically into the mouse bladder. The OT-1 T cells are engineered to recognize the MB49-OVA tumor. This immunotherapy treatment is a pre-clinical analogue of the adoptive cell therapy with tumor-infiltrating lymphocytes (ACT-TIL), that is a personalized immunotherapy using patients' own T cells expanded ex vivo and reinfused to the patient [15]. In the pre-clinical study, tumor overall volume was monitored using ultrasound imaging and recorded on days $6, 9, 13, 16, 20$ and 23.

We propose the following systems of ODEs to describe the interactions between cancer cells (C), OT-1 T cells (T), MDSCs (M), and GEM (G). These are presented in Eqs (1), with the summary of the model parameters presented in Table (1) in Section A.1.

$$
\begin{aligned}
\frac{dC(t)}{dt} &= p_C C(t) - k_{TC} C(t) T(t) - k_{GC} C(t) G(t) \\
\frac{dT(t)}{dt} &= U_T(t) + n_T T(t) - s_{CT} T(t) C(t) - s_{MT} T(t) M(t) - k_{GT} T(t) G(t) \\
\frac{dM(t)}{dt} &= r_M C(t) - k_{GM} M(t) G(t) - d_M M(t) \\
\frac{dG(t)}{dt} &= U_G(t) - d_G G(t)
\end{aligned}
\tag{1}
$$

Both chemo- and immune-therapy can affect the interactions between cancer cells, OT-1 T cells, and MDSCs. To investigate how the trajectories of these altered dynamics can be inferred from the experimental data, we will focus on the MDSCs suppression of T cells s_{MT}, assuming that this model parameter vary over time due to the external intervention of chemo- and immune-therapy.

3 Modeling Subpopulations and time-varying Parameters at Unobserved time points using PINN

3.1 Problem Statement

Consider a system of ODEs in the following form

$$\frac{du(t)}{dt} = f(t, u(t); \lambda_1, \lambda_2, \ldots, \lambda_n)\,, \tag{2}$$

where $u : [t_0, t_F] \to \mathbb{R}^m$ consists of m state variables and n parameters. At most $p \leq n$ of the parameters are time-varying and denoted by $\Lambda(t) = (\lambda_1(t), \lambda_2(t), \ldots, \lambda_p(t))$. The remaining $(n-p)$ parameters are constant. The function f is assumed to be nonlinear. Given experimental data $\{(t_i, u(t_i))\}_{i=1}^{M^{\text{data}}}$, which are assumed to approximately satisfy Eq. (2), we seek to model u and the time-varying parameters Λ using neural network surrogates, denoted by u_{NN} and Λ_{NN}, respectively.

3.2 PINN Framework and Hyperparameter Tuning

A physics-informed neural network (PINN) was constructed using a feedforward neural network (FNN) as described in Section A.2. First, we augment the experimental data points $\{(t_i, u(t_i))\}_{i=1}^{M^{\text{data}}}$ with an interpolation function $\hat{u}$ that interpolate $u(t_i)$ at times t_i, generating M^{interp} additional points within the same time domain. The PINN takes as input the interpolated time domain. Subpopulation trajectories are approximated through the surrogate u_{NN}, while the interaction dynamics are inferred through Λ_{NN}. The `SiLU` activation function was used in the hidden layers, and the `Softplus` activation function was applied to the output layers to enforce non-negativity in the solutions. The interpolated experimental data were normalized prior to training and optimization was performed using the `Adam` optimizer with 20,000 training epochs to ensure convergence. For the other PINN hyperparameters, u_{NN} was implemented with three hidden layers of 100 neurons each, while Λ_{NN} was implemented with two hidden layers of 200 neurons each. All implementations were carried out in the `PyTorch` library.

3.3 PINN Training and Loss Function

The total loss, L_{total} (Eq. (3)), is defined as a weighted sum of four components: the ODE residual loss L_r, the data loss L_d, the initial condition loss L_{IC}, and the biological constraint loss L_{bc}. The corresponding weights w_r, w_d, w_{IC}, w_{bc} were adaptively adjusted following the multi-task learning of the objective function during training to balance the contributions of the different loss terms [7]. The complete formulation of $\mathrm{L}_{\text{total}}$ is given in Eq. (4), where $\|\cdot\|$ denotes the L^2 norm. In Eq. (4), the derivative $\frac{d}{dt_j}u_{NN}$ is computed using automatic differentiation as described in [1] and evaluated at the time points t_j where $j \in \{1, 2, \ldots, M^{\text{data}} + M^{\text{interp}}\}$. The data loss L_d penalizes the difference between u_{NN} and the experimental data $u(t_i)$ at times t_i. The residual loss L_r enforces consistency between the neural network outputs and the governing ODE system, i.e.,

$$\tfrac{d}{dt_j}u_{NN}(t_j) = f\big(t_j, u_{NN}(t_j); \Lambda_{NN}(t_j), \lambda_{NN\,p+1}, \ldots, \lambda_{NN\,n}\big).$$

The initial condition loss L_{IC} ensures that the learned trajectories satisfy prescribed initial conditions, thereby improving stability during training. Finally, the biological constraint loss L_{bc} incorporates domain-specific constraints on the subpopulations at M^{bc} specified time points.

$$\mathrm{L}_{total} = w_r\mathrm{L}_r + w_d\mathrm{L}_d + w_{IC}\mathrm{L}_{IC} + w_{bc}\mathrm{L}_{bc} \tag{3}$$

$$\mathrm{L}_r = \tfrac{1}{M^{\text{data}}+M^{\text{interp}}} \sum_{j=1}^{M^{\text{data}}+M^{\text{interp}}} \left\| \tfrac{d}{dt_j}u_{NN}(t_j) - f\big(t_j, u_{NN}(t_j); \Lambda_{NN}(t_j), \lambda_{NN\,p+1}, \ldots, \lambda_{NN\,n}\big) \right\|,$$

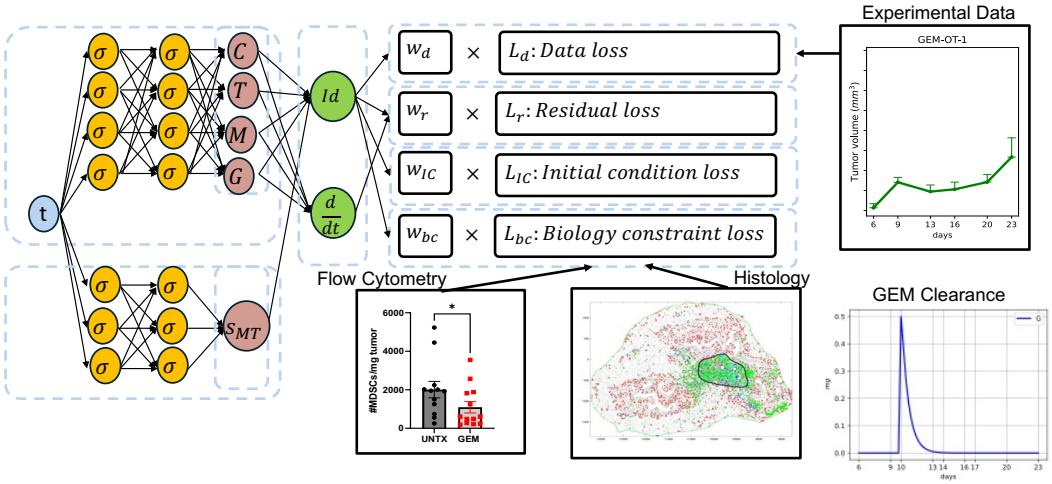
$$\mathrm{L}_d = \tfrac{1}{M^{\text{data}}} \sum_{i=1}^{M^{\text{data}}} \|u_{NN}(t_i) - \hat{u}(t_i)\|, \quad \mathrm{L}_{IC} = \|u_{NN}(t_0) - \hat{u}(t_0)\|, \quad \mathrm{L}_{bc} = \tfrac{1}{M^{\text{bc}}} \sum_{k=1}^{M^{\text{bc}}} \|u_{NN}(t_k) - \hat{u}(t_k)\|. \tag{4}$$

4 Result

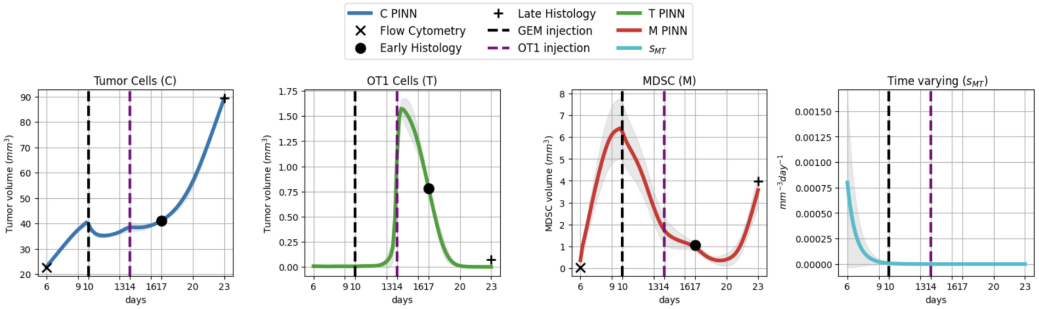
In our PINN implementation, we assumed the total tumor volume to be the sum of the subpopulations C, T, and M, which are unknown at the sparse experimental time points. We increased the number of data points within the points where the experimental data points were collected using spline-based interpolation.

We denote the experimental data as $u_{\text{GEM-OT1}}$, which are tumor volume measurements from ultrasound images recorded on days $6, 9, 13, 16, 20$, and 23, see Figure (1)(**a**). We obtained the proportions of C and M at $t_0 = 6$ from flow cytometry data (Figure (1)(**a**)). Using two histology time points data collected on days 17 and 23, we obtained digital histology of bladder tissue from a mouse treated with GEM and OT-1 T cells. In the digital histology presented in Figure (1)(**a**), tumors (C) are outlined in black, MDSCs (M) are represented as green dots, and T cells (T) are blue dots. See Section A.3 for how we set the initial conditions and the biology constraints in the PINN implementation of the combination therapy in Eq. (1).

We demonstrate our approach by applying the PINN framework and implementation (sections (3.2) and (3.3)) to an ODEs model of a growing tumor exposed to a combination of two anticancer treatments: gemcitabine (GEM) chemotherapy and T cell immunotherapy described in section (2). In Figure (1)(**a**), we present the schematic of the implementation of the PINN algorithm to the system of ODEs given in Eq. (1). First, PINN learns the dynamics of the subpopulations C, T, M, and G, which are potential solutions to Eq. (1). Next, in an unsupervised way, PINN learns the time-varying parameter s_{MT} to infer interaction dynamics resulting from the external interventions of GEM (chemotherapy) and OT-1 T cells (immunotherapy).



(a) PINN framework to learn subpopulations and a time-varying parameter.



(b) Subpopulations C, T, M and time-varying s_{MT} learned by the PINN.

Figure 1: PINN implementation to the combination therapy in Eq. (1).

As shown in Figures (1)(**b**), PINN learns a possible dynamics for the subpopulations C, T, and M, aligning with biological constraints at the flow cytometry point (represented as black cross) and histology time points (represented as black circle dot at the early histology time point and black plus at the late histology time point). At the GEM injection (black vertical dash line) we see the effect of local depletion of the MDSCs and some killing of the tumor cells. The OT-1 injection (purple vertical dash line) indicate some killing of the tumor cells. The learned clearance of the drug GEM (G) is shown in Figure (1)(**a**). In Figures (1)(**b**), PINN learns possible interaction dynamics for the model parameter s_{MT} in an unsupervised fashion. We ran the PINN algorithm 10 times to generate uncertainty bound for each of the PINN solutions. The sum of the Subpopulations C, T, M matches the experimental tumor volume data at the time points where measurements were collected (Section A.4).

5 Discussion

The results we present in this work captures the changing dynamics in a limited-data scenario involving three different types of interacting cells and drug Gemcitabine. Additionally, we demonstrate how to use the PINN framework to incorporate biological knowledge (histology and flow cytometry data) as a form of regularization. This enables us to reduce the space of admissible solutions. The approach is applicable to many other differential equation models in biology, particularly in systems where interacting organisms alter their behavior in response to external interventions.

Acknowledgments and Disclosure of Funding

This work was supported by the Department of Defense grant W81XWH-22-1-0340 and the US National Institutes of Health, National Cancer Institute grant R01-CA259387. This work was supported in part by the Shared Resources at the H. Lee Moffitt Cancer Center & Research Institute an NCI designated Comprehensive Cancer Center under the grant P30-CA076292 from the National Institutes of Health. LEN was supported by the NIH grant U54 CA274507-02A1 through the Moffitt High-School Internship Program in Integrated Mathematical Oncology (HIP-IMO).

References

[1] A. G. Baydin, B. A. Pearlmutter, A. A. Radul, and J. M. Siskind. Automatic differentiation in machine learning: a survey. *Journal of Machine Learning Research*, 18(153):1–43, 2018.

[2] S. Bazargan, B. Bunch, A. M. E. Ojwang, J. Blauvelt, A. Landin, J. Ali, D. Abrahams, C. Cox, A. M. Hall, M. S. Beatty, M. Poch, K. Rejniak, and S. Pilon-Thomas. Targeting myeloid-derived suppressor cells with gemcitabine to enhance efficacy of adoptive cell therapy in bladder cancer. *Front Immunol*, 2023.

[3] M. Benaïm, C. Lobry, T. Sari, and E. Strickler. Population growth on a time varying network. *arXiv preprint arXiv:2411.07821v1*, 2024.

[4] J. Ferlay, M. Ervik, F. Lam, M. Laversanne, M. Colombet, L. Mery, M. Piñeros, A. Znaor, I. Soerjomataram, and F. Bray. Global cancer observatory: Cancer today. `https://gco.iarc.who.int/today`, 2024. accessed: [August 18, 2025].

[5] A. Glyn-Davies. *Statistical Inference and Learning for Stochastic and Partial Differential Equations*. PhD thesis, Apollo - University of Cambridge Repository, 2025.

[6] K. Hornik. Approximation capabilities of multilayer feedforward networks. *Neural Networks*, 4:251–257, 1991.

[7] A. Kendall, Y. Gal, and R. Cipolla. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. *in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7482–7491, 2018.

[8] E. Kharazmi, M. Cai, X. Zheng, Z. Zhang, G. Lin, and G. E. Karniadakis. Identifiability and predictability of integer- and fractional-order epidemiological models using physics-informed neural networks. *Nature Computational Science*, pages 744–753, 2021.

[9] J. H. Lagergren, J. T. Nardini, R. E. Baker, M. J. Simpson, and K. B. Flores. Biologically-informed neural networks guide mechanistic modeling from sparse experimental data. *PLoS Comput Biol*, 16, 2020.

[10] N. F. Müller, G. Dudas, and T. Stadler. Inferring time-dependent migration and coalescence patterns from genetic sequence and predictor data in structured populations. *Virus Evolution*, 5, 2019.

[11] K. D. Olumoyin, A. Q. M. Khaliq, and K. M. Furati. Data-driven deep-learning algorithm for asymptomatic covid-19 model with varying mitigation measures and transmission rate. *Epidemiologia*, 2:471–489, 2021.

[12] S. Pilon-Thomas, L. Kuhn, S. Ellwanger, W. Janssen, E. Royster, S. Marzban, R. Kudchadkar, J. Zager, G. Gibney, V. K. Sondak, J. Weber, J. J. Mulé, and A. A. Sarnaik. Efficacy of adoptive cell transfer of tumor-infiltrating lymphocytes after lymphopenia induction for metastatic melanoma. *Journal of immunotherapy*, 35(8):615–620, 2012.

[13] L. Podina, A. Ghodsi, and M. Kohandel. Learning chemotherapy drug action via universal physics-informed neural networks. *arXiv preprint arXiv:2404.08019v1*, 2024.

[14] M. Raissi, P. Perdikaris, and G. Karniadakis. Physics informed deep learning: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal Of Computational Physics*, pages 686–707, 2019.

[15] S. A. Rosenberg and N. P. Restifo. Adoptive cell transfer as personalized immunotherapy for human cancer. *Science*, pages 62–68, 2015.

[16] Y. Sawada. An efficient estimation of time-varying parameters of dynamic models by combining offline batch optimization and online data assimilation. *Journal of Advances in Modeling Earth Systems*, 14:e2021MS002882, 2022.

[17] S. Stevanović, L. M. Draper, M. M. Langhan, T. E. Campbell, M. L. Kwong, J. R. Wunderlich, M. E. Dudley, J. C. Yang, R. M. Sherry, U. S. Kammula, N. P. Restifo, S. A. Rosenberg, and C. S. Hinrichs. Complete regression of metastatic cervical cancer after treatment with human papillomavirus-targeted tumor-infiltrating t cells. *Journal of clinical oncology*, 33(14):1543–1550, 2015.

A Technical Appendices and Supplementary Material

A.1 Table of parameters for the ODE in Eq. (1)

Table 1: Model parameters in Eq. (1)

Parameter	Description [units]	Parameter	Description [units]
k_{TC}	T cell killing of cancer [$mm^{-3}\,day^{-1}$]	n_T	T cell net growth [day^{-1}]
k_{GC}	GEM killing of cancer [$mg^{-3}\,day^{-1}$]	U_G	GEM injection [$mg\,day^{-1}$]
k_{GM}	GEM killing of MDSCs [$mg^{-3}\,day^{-1}$]	U_T	T cell injection [$mm^3\,day^{-1}$]
k_{GT}	GEM killing of T cells [$mg^{-3}\,day^{-1}$]	r_M	MDSCs recruitment [day^{-1}]
d_G	GEM clearance [day^{-1}]	s_{MT}	MDSCs suppression of T cells [$mm^{-3}\,day^{-1}$]
d_M	MDSCs death [day^{-1}]	s_{CT}	Cancer suppression of T cells [$mm^{-3}\,day^{-1}$]
p_C	Cancer net growth [day^{-1}]		

A.2 Feedforward Neural Network (FNN)

An FNN is an artificial neural network where the forward pass is in one direction, and it is usually fully connected. It can be represented as a composition of functions with L layers, where the input layer is an affine transformation of a vector t followed by an activation function. The output of the input layer becomes input to the next layer, continuing until the Lth layer as shown in Eq. (5).

$$NN(t;\theta) = \sigma_L(W_L\sigma_{L-1}(W_{L-1}\ldots\sigma_2(W_2\sigma_1(W_1t+b_1)+b_2)\ldots+b_{L-1})+b_L), \tag{5}$$

where $\theta := (W_1,\ldots,W_L,b_1,\ldots,b_L)$ denotes the set of neural network weight matrices W_k and bias vectors b_k for $(k=1,\ldots,L)$, and $\sigma_k(\cdot)$ denotes the activation function. It is widely known that a feedforward neural network with at least one hidden layer containing a finite number of neurons can approximate any continuous function on compact subsets of $\mathbb{R}^m$ [6].

A.3 Setting the Initial Conditions and Biology Constraints for PINN

We assume that $u_{\text{GEM-OT1}}$ is a sum of the subpopulations C, T, and M, so we implement the data loss L_d in Eq. (4) as $\frac{1}{M^{\text{data}}}\sum_{i=1}^{M^{\text{data}}}((C(t_i)+T(t_i)+M(t_i))-u_{\text{GEM-OT1}}(t_i)\,)^2$. In order to initialize C, T, M, at $t_0=6$, we define L_{IC} in Eq. (4) as follows:

$$\mathrm{L}_{IC} = (C(t_0)-q_0\times u_{\text{GEM-OT1}}(t_0)\,)^2 + (T(t_0)-q_1\times u_{\text{GEM-OT1}}(t_0)\,)^2 + (M(t_0)-q_2\times u_{\text{GEM-OT1}}(t_0)\,)^2\,,$$

where $q_0=0.99887$, $q_1=0$, and $q_2=1-0.99887$.

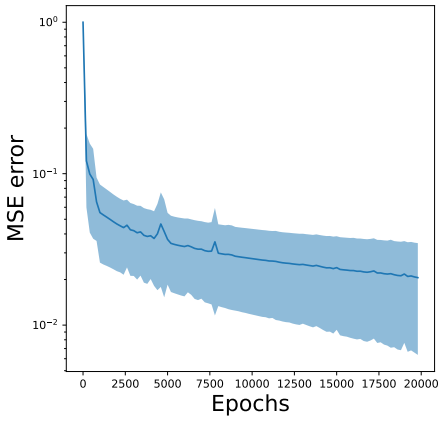
Similarly, we obtain the biology constraints loss L_{bc} in Eq. (4) as follows:

$$\mathrm{L}_{bc} = \sum_{k\in\{1,2\}}(C(t_k)-q_{0_k}\times u_{\text{GEM-OT1}}(t_k))^2+(T(t_k)-q_{1_k}\times u_{\text{GEM-OT1}}(t_k))^2+(M(t_k)-q_{2_k}\times u_{\text{GEM-OT1}}(t_k))^2$$

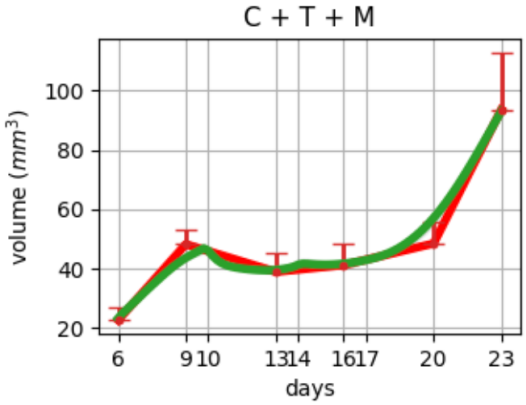
Where $q_{0_1}=0.95755$, $q_{1_1}=0.01818$, $q_{2_1}=0.02427$ are the proportions of C, T, and M respectively at day 17 (histology time point). Similarly for C, T, and M at day 23 (histology time point), we have that $q_{0_2}=0.95665$, $q_{1_2}=0.00078$, and $q_{2_2}=0.04256$ respectively.

A.4 Loss Convergence, Comparison of learned Subpopulations with Experimental data

The loss convergence and the comparison of the sum of the subpopulations $C+T+M$ vs the experimental tumor volume data at the time points where measurements were collected.



(a) Convergence plot for the PINN implementation.



(b) Total tumor volume (red) vs learned $C+T+M$ (green).