Mechanistic Flow Matching for Data-Scarce Scientific Cohorts

Jiyeon Kim*
Ajou University
Department of Mathematics
Suwon, Republic of Korea
gkim0201@ajou.ac.kr

Daehong Min†
Ajou University
Department of Mathematics
Suwon, Republic of Korea
mdh3120@ajou.ac.kr

Minjung Gim†
National Institute for Mathematical Sciences
Daejeon, Republic of Korea
mjgim@nims.re.kr

Dongwook Shin
Ajou University
Department of Mathematics
Suwon, Republic of Korea
dws@ajou.ac.kr

Soon-Sun Kwon
Ajou University
Department of Mathematics
Suwon, Republic of Korea
qrio1010@ajou.ac.kr

Abstract

Generative modeling for scientific data is challenging in data-scarce settings ($N = 8$), where models must generalize beyond limited observations while respecting mechanistic constraints. In personalized oncology, longitudinal tumor growth data are often sparse, yet generated trajectories must remain consistent with biologically motivated dynamics for quantitative analysis. Conventional synthetic-data approaches can be useful, but under extreme data scarcity, statistical augmentation alone may not preserve consistency with the mechanistic dynamics governing the observations.

To address this challenge, we propose Physics-Guided Optimal Transport Conditional Flow Matching (Physics OT-CFM), a flow-based generative framework that rectifies the transport field using mechanistic growth dynamics. Our method operates on Gompertz growth parameters, which are decoded into tumor growth trajectories. It combines data-based non-dimensionalization with physics-guided modifications of training targets and sampling dynamics using constraints induced by the prescribed tumor growth ODE.

Experiments on glioma xenograft data with only eight untreated subjects show that Physics OT-CFM achieves near-zero constraint violations while retaining competitive distributional metrics against established synthetic-data baselines. The generated samples reduce simple replication of sparse training points and produce trajectories consistent with the prescribed mechanistic growth law. Together, these results suggest that physics-guided flow matching can serve as a data-centric augmentation component for AI Data Scientist pipelines in scientific domains, providing synthetic cohorts that downstream analyses can consume without sacrificing mechanistic consistency.

†These authors contributed equally.
*Corresponding author. Email: gkim0201@ajou.ac.kr.

Conference'17, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YYYY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn

CCS Concepts

• **Computing methodologies** → **Machine learning**; *Machine learning approaches*; • **Applied computing** → **Health informatics**.

Keywords

AI Data Scientist, Data-Centric AI, Generative Modeling, Physics-Guided Machine Learning, Conditional Flow Matching, Tumor Growth

ACM Reference Format:
Jiyeon Kim*, Daehong Min†, Minjung Gim†, Dongwook Shin, and Soon-Sun Kwon. 2026. Mechanistic Flow Matching for Data-Scarce Scientific Cohorts. In . ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction

Accurately modeling tumor growth dynamics is important for treatment response assessment, prognosis prediction, and clinical trial design [1, 2]. In clinical and preclinical studies, tumor size is typically measured at discrete time points and summarized as tumor growth curves, which guide therapeutic decisions and are widely used to quantify antitumor efficacy in xenograft and patient-derived xenograft (PDX) models [3–5]. However, in vivo tumor growth studies are often constrained by limited follow-up duration, tumor-burden considerations, and ethical regulations on animal use, making such experiments costly and restricting both sample size and measurement frequency [6, 7]. As a result, many xenograft datasets consist of short and sparsely observed tumor growth trajectories, often with only a small number of untreated control animals [7, 8]. Since untreated tumor growth provides the baseline dynamics against which treatment effects are interpreted, the scarcity of such trajectories makes it difficult to characterize inter-animal variability in tumor growth parameters with sufficient precision [9]. This represents an extreme regime of generative modeling where models must leverage mechanistic structure to avoid naive memorization and generalize reliably from very few observations [10, 11]. Such a challenge necessitates generative frameworks that can faithfully extrapolate beyond sparse, ethically restricted data while adhering to complex domain constraints.

In data-scarce tumor growth studies, synthetic cohorts are commonly generated using Bayesian models, kernel density estimation

(KDE), or interpolation-based methods such as SMOTE [12–16]. Although well established, these methods can exhibit structural limitations in the extreme small-sample regime considered here. Bayesian approaches incorporate prior knowledge and quantify uncertainty, but with only a few untreated trajectories, posterior predictive samples can be highly sensitive to informative prior specification. KDE smooths the empirical parameter distribution, but Euclidean kernel smoothing ignores mechanistic coupling among growth parameters and may place probability mass in biologically implausible off-manifold regions, especially when bandwidth selection is unstable under $N = 8$ [14, 15]. SMOTE preserves proximity to observed subjects through interpolation, but its samples remain confined to the empirical convex hull, limiting tail variability and generation beyond simple convex combinations of the training data [16]. These complementary limitations motivate a data-adaptive and physics-constrained generative framework that preserves agreement with the observed cohort while reducing prior sensitivity, off-manifold smoothing, and convex-hull confinement.

The challenge above also fits naturally into the broader picture of building AI Data Scientists for scientific domains. End-to-end scientific data analysis—spanning preprocessing, augmentation, modeling, and downstream interpretation—is increasingly delegated to automated pipelines, and in preclinical or biomedical settings the augmentation step is often the limiting factor when only a handful of observations are available. In such settings, synthetic samples that look statistically plausible but violate domain laws can degrade every downstream stage that consumes them. We therefore view physics-constrained generation in the data-scarce regime as a data-centric capability that an AI Data Scientist should provide: samples should match the observed distribution and at the same time remain compatible with the mechanistic dynamics that downstream analyses assume.

To address these limitations, we propose Physics OT-CFM, a generative framework that incorporates mechanistic growth dynamics into flow-based modeling through an explicit rectification of the transport field. Rather than generating raw tumor-volume time series directly, our method generates Gompertz growth parameters and decodes them into tumor growth trajectories. The underlying dynamics are modeled by a Gompertz-type tumor growth ODE, which captures key qualitative properties of untreated tumor growth, including positivity and growth deceleration [17, 18]. Inspired by physics-informed machine learning [11, 19], Physics OT-CFM integrates physics-derived gradients into both training and sampling, guiding generated parameters toward regions consistent with the prescribed growth law. This learned transport-based generation avoids relying solely on prior-driven sampling, unconstrained smoothing, or fixed convex interpolation, enabling the synthesis of tumor growth trajectories consistent with the prescribed mechanistic growth law under extreme data scarcity.

Contributions. Our main contributions are:

- We introduce a simple data-based non-dimensionalization for Gompertz parameters that stabilizes learning under severe scale imbalance while enforcing $V_0 < K$ by construction.
- We propose a physics-guided rectification of OT-CFM that modifies both training targets and sampling dynamics using gradients of a differentiable Gompertz constraint penalty derived from Gompertz growth properties.
- We compare Physics OT-CFM against four baselines spanning parametric and generative paradigms; our method achieves distributional fidelity competitive with established baselines in the extreme data-scarce regime ($N = 8$).
- Our framework requires no informative prior, attains deterministic constraint satisfaction through the rectified vector field rather than via post-hoc rejection or hard-coded manifold assumptions, and avoids prior misspecification by construction.

Taken together, these components provide a practical framework for learning physics-constrained generative models from extremely small scientific datasets, where synthetic samples are guided by a prescribed mechanistic ODE while maintaining statistical agreement with the observed cohort.

2 Preliminaries

Gompertz Growth Model. We model baseline tumor growth using the Gompertz growth law [17, 18], a parsimonious ODE-based model that captures key qualitative features of untreated tumor growth, including positivity, growth deceleration, and saturation. Let $V(t) > 0$ denote tumor volume at time t. We use the three-parameter Gompertz ODE:

$$\frac{dV(t)}{dt} = aV(t)\log\left(\frac{K}{V(t)}\right), \qquad V(0) = V_0, \tag{1}$$

where $a > 0$ is the growth rate, $K > 0$ is the carrying capacity, and V_0 is the initial tumor volume.

The qualitative properties of its closed-form solution (i.e., when $0 < V_0 < K$, the trajectory remains positive, increases monotonically, and approaches K as $t \to \infty$) directly motivate the physical constraints used in our generative framework. Detailed derivations, closed-form solutions, and the reduced Gompertz parameter (α, β) used for baseline comparisons are deferred to Appendix A.

Optimal Transport Conditional Flow Matching (OT-CFM). We consider a flow-based generative model that transforms samples from a simple base distribution p_0 into samples from a target data distribution p_{data} through the probability flow ODE $\frac{dz_t}{dt} = v_\theta(z_t, t)$ for $t \in [0, 1]$.

Conditional Flow Matching (CFM) trains the time-dependent neural vector field v_θ using linear interpolation paths $z_t = (1-t)z_0 + tz_1$ between a base sample $z_0 \sim p_0$ and a data sample $z_1 \sim p_{\text{data}}$. The standard CFM objective minimizes the distance to the target velocity $z_1 - z_0$:

$$\min_\theta \mathbb{E}\left[\|v_\theta(z_t, t) - (z_1 - z_0)\|_2^2\right]. \tag{2}$$

OT-CFM improves this construction by pairing z_0 and z_1 through minibatch optimal transport, which reduces path crossings and stabilizes training. However, standard OT-CFM does not enforce domain-specific physical constraints. We therefore introduce a physics-guided rectification of the transport field so that generated Gompertz parameters remain consistent with the prescribed growth law.

3 Method

In this section, we introduce Physics OT-CFM, a generative framework for biological growth parameters that explicitly incorporates Gompertz-model-based physical constraints. We begin by providing the design rationale for selecting OT-CFM over other generative paradigms in the context of extreme data-scarcity (Section 3.1). Then our approach combines a data-based non-dimensionalization (Section 3.2) of model parameters with physics-guided flow learning, enabling stable training and biologically consistent sample generation (Section 3.3).

3.1 Rationale for Flow-based Framework

To synthesize physically plausible parameters under extreme data-scarcity, we employ OT-CFM as a promising generative framework [20]. The rationale for our selection considering the limitations of other models is as follows: Diffusion models can enforce physical constraints through gradient guidance or manifold projection, their stochastic sampling and complex curved paths can accidentally lead to data memorization in low-data regimes. It prioritizes the replication of training points over manifold interpolation [21–23]. In contrast, OT-CFM learns a deterministic vector field $\frac{d\mathbf{z}_t}{dt} = v_\theta(\mathbf{z}_t, t)$ along straight optimal-transport paths, which (i) admits precise and stable rectification via physics gradients $\nabla_{\mathbf{z}_t}\Phi_{\text{phys}}$, and (ii) reduces stochastic variance to encourage genuine interpolation between sparse observations. We empirically compare against a physics-guided diffusion baseline in Section 4.

3.2 Data-based Non-dimensionalization

The biological growth parameters of the Gompertz model, $\mathbf{x} = (a, K, V_0)^\top \in \mathcal{X} \subset \mathbb{R}^3$, naturally exhibit different physical scales. Directly learning a generative model in this raw parameter space often leads to optimization instability and violations of basic biological constraints. To address this issue, we draw inspiration from non-dimensionalization in classical mathematical biology [24]. In that context, system variables are typically scaled by characteristic constants (e.g., scaling volume by K or time by $1/a$) to derive a dimensionless form of the governing ODE, such as $d\hat{V}/d\hat{t} = \hat{V}\log(1/\hat{V})$ where $\hat{V} = V/K$. While such theoretical scaling simplifies the physics, it does not account for the statistical variance and scale imbalances inherent in multi-patient datasets.

Therefore, we introduce a data-based non-dimensionalization that maps $\mathbf{x}$ into a latent space $\mathcal{Z}$ through a domain-aware transformation $\mathcal{T} : \mathcal{X} \to \mathcal{Z}$. This transformation not only enforces physical constraints but also "centers" the parameter manifold based on the cohort's statistics, aligning with the broader paradigm of physics-informed modeling [11].

We compute reference scales from the training data using median statistics, denoted by a_{med} and K_{med}. The transformed latent variable $\mathbf{z} = (z_1, z_2, z_3)^\top \in \mathbb{R}^3$ is defined as

$$z_1 = \frac{a}{a_{\text{med}}}, \qquad z_2 = \log\left(\frac{K}{K_{\text{med}}}\right), \qquad z_3 = \log\left(\frac{V_0}{K}\right). \tag{3}$$

This transformation stabilizes learning by reducing scale imbalance and enforces the biological constraint $V_0 < K$ by construction. In particular, the ratio-based representation of z_3 allows the latent space to remain unconstrained while preserving valid initial conditions in the original physical parameter space.

3.3 Physics-Guided Flow Learning

While standard OT-CFM learns probability transport paths efficiently, it does not explicitly account for the underlying physical structure of biological growth models. As a result, generated samples may drift away from biologically valid regions of the parameter space. To address this limitation, we propose a physics-guided flow learning strategy that incorporates gradient information from physical constraints into the learned vector field.

Physics Penalty Function. We define a differentiable physics penalty function $\Phi_{\text{phys}}(\mathbf{z})$ that measures violations of Gompertz model based constraints after decoding latent variables via $\mathcal{T}^{-1}$. The penalty assigns high costs to states that contradict the Gompertz growth dynamics, such as negative growth rates or non-monotonic tumor growth trajectories. Specifically, Φ_{phys} is constructed as a sum of soft barrier terms enforcing three fundamental biological properties:

- valid initial conditions ($V_0 < K$),
- monotonic growth ($dV/dt \geq 0$),
- near-saturation at the terminal time ($V(T) \approx K$).

Defining the normalized volume ratio as $v(t) = V(t)/K$, the total physics penalty is given by

$$\begin{aligned}\Phi_{\text{phys}}(\mathbf{z}) = &\underbrace{\zeta\left(\frac{v(0)-1}{\tau}\right)}_{\text{Boundary}} + \underbrace{\mathbb{E}_{t\sim\text{Unif}[0,T]}\left[\zeta\left(\frac{-\dot{v}(t)}{\tau}\right)\right]}_{\text{Monotonicity}} \\ &+ \underbrace{\zeta\left(\frac{|v(T)-1|-\epsilon}{\tau}\right)}_{\text{Saturation}}\end{aligned} \tag{4}$$

where $\dot{v}(t)$ is the time derivative and $\zeta(x) = \log(1+\exp(x))$ is the Softplus function, serving as a smooth approximation of the ReLU function to ensure the differentiability of the vector field guidance. The parameter τ controls the sharpness of the barrier and ϵ is a small relaxation margin to account for biological variability.

Efficient evaluation. Since the Gompertz model admits a closed-form solution (Eq. (10)), we compute $v(t)$ and $\dot{v}(t)$ analytically after decoding $\mathbf{z}$ to (a, K, V_0). In practice, the expectation over t is approximated by Monte Carlo samples of $t \sim \text{Unif}[0, T]$, which adds negligible overhead compared to the neural vector field evaluation.

Training with Rectified Targets. Standard OT-CFM uses a straight-line target velocity $u^*(\mathbf{z}_t) = \mathbf{z}_1 - \mathbf{z}_0$. We modify this target to respect physical laws by adding a gradient term. We adjust the target velocity to guide the transport process toward the physically valid manifold as follows:

$$u_{\text{target}}(\mathbf{z}_t) = u^*(\mathbf{z}_t) - \lambda_{\text{train}}\nabla_{\mathbf{z}_t}\Phi_{\text{phys}}(\mathbf{z}_t). \tag{5}$$

Here, the term $-\nabla_{\mathbf{z}_t}\Phi_{\text{phys}}$ provides a geometry-aware correction that repels trajectories from high-penalty regions. Importantly, this modifies the *target* field rather than adding an additional loss,

allowing the neural vector field to internalize physics compliance as an intrinsic property of the learned transport map.

Refined Sampling. We apply a similar strategy during sampling to guarantee the biological plausibility of the generated parameters. As we solve the ODE from noise to data, we add a correction term $-\lambda_{\text{sample}}\nabla\Phi_{\text{phys}}$ to the predicted velocity. This test-time refinement ensures that the generated parameters stay on the biologically valid manifold. The detailed procedures for training and inference are summarized in Algorithm 1 and Algorithm 2, respectively.

The proposed physics-guided training and sampling procedures are designed to be computational efficient. The additional overhead arises from evaluating $\nabla_{\mathbf{z}_t}\Phi_{\text{phys}}$, whose complexity is negligible compared to the forward and backward passes of the neural vector field v_θ. This allows our framework to incorporate complex biological priors without significantly increasing the training time or sampling latency.

For completeness, we provide the full algorithmic descriptions of the physics-guided training and sampling procedures in Algorithm 1 and Algorithm 2 respectively.

Algorithm 1 Physics-Guided OT-CFM Training

Require: Dataset $\mathcal{D} = \{\mathbf{x}^{(i)}\}$ (physical space), transform $\mathcal{T}$, physics penalty Φ_{phys}, guidance weight λ

Initialize network parameters θ
for iteration $s = 1$ to S **do**
  Sample data batch $\mathbf{x}_1 \sim \mathcal{D}$
  Sample noise $\mathbf{z}_0 \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
  Transform to latent space: $\mathbf{z}_1 \leftarrow \mathcal{T}(\mathbf{x}_1)$
  Sample time $t \sim \text{Unif}(0, 1)$
  Compute interpolation: $\mathbf{z}_t = (1-t)\mathbf{z}_0 + t\mathbf{z}_1$
  Base target velocity: $u^* = \mathbf{z}_1 - \mathbf{z}_0$
  // **Physics-guided rectification**
  Compute physics gradient: $\mathbf{g} = -\nabla_{\mathbf{z}_t}\Phi_{\text{phys}}(\mathbf{z}_t)$
  Rectified target: $u_{\text{target}} \leftarrow u^* + \lambda\mathbf{g}$
  // **Update**
  Predict velocity: $v_{\text{pred}} = v_\theta(\mathbf{z}_t, t)$
  Loss $\mathcal{L} = \|v_{\text{pred}} - u_{\text{target}}\|_2^2$
  Update $\theta \leftarrow \theta - \eta\nabla_\theta\mathcal{L}$
end for
return Trained vector field v_θ

Algorithm 2 Physics-Guided Sampling

Require: Trained vector field v_θ, transform $\mathcal{T}$, physics penalty Φ_{phys}, number of steps N, sampling guidance weight λ_s

Sample $\mathbf{z}_0 \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
Set step size $\Delta t = 1/N$
for $k = 0$ to $N-1$ **do**
  $t_k = k \cdot \Delta t$
  Base velocity: $v_{\text{base}} = v_\theta(\mathbf{z}_k, t_k)$
  // **Physics refinement**
  Compute physics gradient: $\mathbf{g} = -\nabla_{\mathbf{z}_k}\Phi_{\text{phys}}(\mathbf{z}_k)$
  Effective velocity: $v_{\text{eff}} \leftarrow v_{\text{base}} + \lambda_s\mathbf{g}$
  Update state: $\mathbf{z}_{k+1} \leftarrow \mathbf{z}_k + \Delta t \cdot v_{\text{eff}}$
end for
Decode parameters: $\mathbf{x}_{\text{final}} = \mathcal{T}^{-1}(\mathbf{z}_N)$
return $\mathbf{x}_{\text{final}}$

4 Experiments

4.1 Experimental Setup

Dataset. We use the Tumor Growth Dataset from the Teaching Statistics in the Health Sciences (TSHS) repository [25]. The data originate from a preclinical glioma xenograft study ($N = 37$) with longitudinal tumor volume measurements. In this work, we focus on the untreated control group ($N = 8$), which received no therapeutic intervention. These trajectories reflect the intrinsic, natural tumor growth dynamics governed by purely biological factors, and provide a baseline against which treatment effects are usually interpreted. Focusing on this sparse subset allows us to study generative modeling under severely data-scarce conditions while avoiding confounding effects from therapy, and providing a rigorous testbed for our physics-guided framework.

Data Preprocessing. Our generative framework models the distribution of biological growth parameters rather than raw tumor volume time series. For each subject, we fitted the three-parameter Gompertz growth model to the observed tumor volume measurements using nonlinear least-squares optimization. This procedure yields a dataset of parameter vectors

$$\mathbf{x}^{(i)} = \left(a^{(i)}, K^{(i)}, V_0^{(i)}\right)^\top, \qquad i = 1, \ldots, 8.$$

As discussed in Section 3.2, these physical parameters exhibit strongly different scales, which can lead to unstable optimization in generative models. We therefore apply the proposed data-based non-dimensionalization (Eq. (3)) to map the parameters into a three-dimensional latent space. Unlike classical non-dimensionalization, which reduces degrees of freedom, our approach preserves the full parameter dimension while normalizing magnitudes based on statistics of the observed data. All subsequent experiments are conducted in this non-dimensionalized latent space.

Table 1: Statistical characteristics of the extracted Gompertz parameters from the U87 glioma control group ($N = 8$).

Parameter	Mean ± SD	Range [Min, Max]	Unit
Growth Rate (a)	0.09 ± 0.04	[0.03, 0.17]	day^{-1}
Carrying Capacity (K)	17,127 ± 7,508	[2,024, 24,057]	mm^3
Initial Volume (V_0)	23.4 ± 41.6	[2.4, 125.0]	mm^3

Table 1 highlights three challenges: severe scale imbalance ($K \gg a$), high heterogeneity in V_0 (CV > 1.7), and a wide K range under $N = 8$ that encourages memorization. These motivate our data-driven non-dimensionalization (Section 3.2) and physics-guided rectification to stabilize learning and promote physically consistent manifold interpolation.

Domain expertise and biological validation. Although the authors are affiliated with mathematics departments, the study was designed and validated based on the corresponding author's prior experience in mathematical oncology and tumor growth modeling.

For each generated parameter set, we verified $0 < V_0 < K$, monotonic growth on $[0, T]$, and near-saturation at T (via Φ_{phys}), and checked derived growth summaries for consistency with tumor-growth modeling practice.

4.2 Baselines

We compare against four baselines spanning parametric and generative paradigms. Compared with these alternatives, our framework offers three concrete advantages: (i) it requires no informative prior, learning the parameter manifold directly from the $N = 8$ samples; (ii) it attains deterministic constraint satisfaction through the rectified vector field rather than via post-hoc rejection or hard-coded manifold assumptions; and (iii) it is robust to prior misspecification by construction, since no external population-level parameters are imposed.

Following the reduced Gompertz parameterization [12], all main comparisons operate in (α, β), where $\alpha = a \log(K/V_0)$ and $\beta = a$. Because the Bayesian baselines do not natively output V_0, this reduced representation is the common ground for fair quantitative comparison. The two generative baselines (Physics Diffusion and Physics OT-CFM) are converted from their native (a, K, V_0) outputs to (α, β) for evaluation.

Bayesian approach with Gompertz model (Bayesian-Gompertz). This is our principal parametric competitor. In the first stage, per-subject parameters $\{(\alpha_i, \beta_i)\}_{i=1}^{8}$ are obtained via nonlinear least-squares fits to each animal's longitudinal volume measurements. In the second stage, we perform Bayesian inference on the population mean $\boldsymbol{\mu} = (\mu_\alpha, \mu_\beta)$ using log-normal priors centered at Vaghi's values [12], with scale parameters exaggerated relative to the published values to reflect the high uncertainty in our $N = 8$ cohort (full prior specification in Appendix B.1). We emphasize that this is a Bayesian approach, chosen to match the available data scale and to avoid overfitting nuisance variance components. Samples are drawn independently for α and β from the resulting log-normal posteriors, yielding a two-dimensional scatter in (α, β).

Bayesian approach with Reduced Gompertz model (Bayesian-Reduced Gompertz). We additionally consider a reduced variant that adopts Vaghi's deterministic relationship $\alpha = k\beta$ with $k = 7.87$ and zero residual variance. Here, only β is sampled stochastically and α is determined deterministically, yielding samples that lie on a one-dimensional manifold. This baseline represents the strongest prior-informed setting: it directly encodes Vaghi's reduced Gompertz structure [12] as a hard constraint.

Adopting an external prior requires verifying that Vaghi's published parameters are consistent with our cohort. Two checks support this. First, the empirical ratio α_i/β_i across the eight subjects has median 7.60 and mean 7.27, well-aligned with Vaghi's $k = 7.87$ (within 10% across three independent estimators: per-subject ratio, linear fit, and population-median ratio). Second, Vaghi's reduced model assumes $\alpha = k\beta + c$ with $c = 0$; a one-sample t-test on the per-subject estimates $c_i = \alpha_i - 7.87\beta_i$ yields $t(7) = -1.05$, $p = 0.33$ (95% CI for mean c: $[-0.16, +0.06]$), statistically indistinguishable from zero and supporting the $c = 0$ assumption. Importantly, k is taken from external literature rather than estimated on our $N = 8$. Therefore we can avoid circular use of the same data for both prior extraction and posterior comparison.

Physics Diffusion. As a generative-model counterpart, we implement a denoising diffusion probabilistic model [21] on the same scaled latent space defined in Section 3.2, with the identical physics penalty Φ_{phys} applied as gradient guidance [22, 23] during sampling. This baseline isolates the choice of generative paradigm (diffusion vs. flow matching) under same physics-aware design.

Non-parametric sanity-check baselines. We additionally include KDE (Gaussian kernel with Scott's-rule bandwidth) and SMOTE [16] ($k = 3$ nearest-neighbor convex interpolation) as non-parametric sanity checks. Both operate directly on the eight real samples in (α, β) with rejection sampling enforcing $\alpha, \beta > 0$. Definitions, parameter scatter, and trajectory visualizations for these two baselines are deferred to Appendix B; their full quantitative results appear alongside the four main methods in Appendix C.

4.3 Evaluation Metrics

We organize evaluation around two complementary scopes: *distribution fidelity* (comparing generated parameter distributions to the eight real subjects) and *training dynamics analysis* (tracking the evolution of generative models). Implementation details for all metrics are deferred to Appendix E.

Primary metrics: distribution fidelity on the Vaghi projection. The reduced Gompertz framework [12] expresses α and β via the deterministic relationship $\alpha = k\beta$ with $k = 7.87$, validated for our cohort in Section 4.2. To enable fair comparison across methods of different effective dimension—the Bayesian-Reduced Gompertz samples lie on this line by construction, while other methods produce 2D scatter—we orthogonally project every sample onto the Vaghi line and compare the resulting one-dimensional distributions. For a sample (α, β), the projected coordinate is:

$$\beta_{\text{proj}} = \frac{k\alpha + \beta}{k^2 + 1}, \qquad k = 7.87. \tag{6}$$

On this 1D projection, we evaluate fidelity using the **1-Wasserstein distance (W_1)** and **Tail-quantile errors (Δq_5, Δq_{95})**. The tail-quantile errors, defined as $\Delta q_p = |q_p^{\text{gen}} - q_p^{\text{real}}|$ for $p \in \{5, 95\}$, diagnose asymmetric failure modes such as hard-manifold tail amplification.

Complementary metrics: training dynamics in 3D latent. To track training dynamics in the full 3D scaled latent space (Section 4.4.3), we evaluate the generated distributions using **Sliced Wasserstein distance (SWD)**, unbiased **Maximum Mean Discrepancy (MMD^2)** with an RBF kernel, and the **Projected Kolmogorov–Smirnov (KS) statistic**. Formal definitions and hyperparameter settings for these standard metrics are provided in Appendix E.

4.4 Results

4.4.1 Distribution Fidelity. We compare three baselines and Physics OT-CFM on the 1D Vaghi projection β_{proj} against the eight real subjects projected the same way. Quantitative results are summarized in Table 2; the parameter scatter in (α, β) space, before projection, is visualized in Figure 1. Two additional non-parametric baselines (KDE, SMOTE) and complementary 1D metrics (KS, MMD^2) are reported in Appendix C.

Table 2: Distribution fidelity on the 1D Vaghi projection $\beta_{\text{proj}} = (k\alpha + \beta)/(k^2 + 1)$, $k = 7.87$. W_1 is the 1-Wasserstein distance to the projected eight real subjects; $\Delta q_5, \Delta q_{95}$ are tail-quantile errors. Physics OT-CFM achieves the lowest W_1 and Δq_{95} and ranks top-2 on every metric. All methods achieve 0% constraint violation by design; Physics OT-CFM additionally produces two numerical edge cases ($K \le V_0$). Non-parametric baselines (KDE, SMOTE) and complementary metrics are reported in Appendix C. Bold = best, underlined = second-best.

Method	$W_1 \downarrow$	$\Delta q_5 \downarrow$	$\Delta q_{95} \downarrow$
Bayesian-Gompertz	0.0135	0.0038	0.0192
Bayesian-Reduced Gompertz	0.0182	**0.0018**	0.0544
Physics Diffusion	0.0294	0.0182	0.0068
Physics OT-CFM (Ours)	**0.0082**	0.0027	**0.0019**

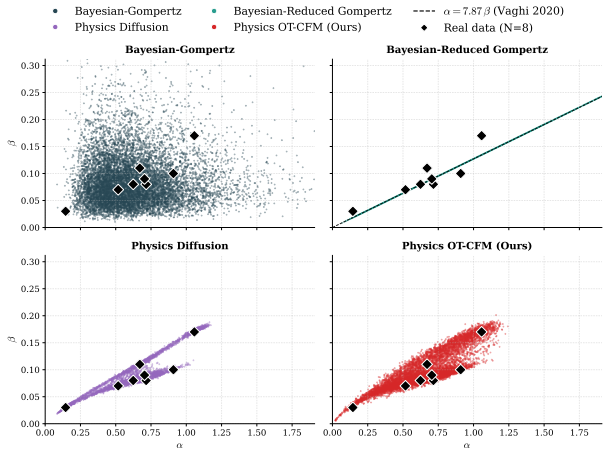


Figure 1: Generated parameter distributions in (α, β) space ($\alpha = a \log(K/V_0)$, $\beta = a$). Black diamonds: eight real subjects. Bayesian-Gompertz blows up the spread by inheriting Vaghi's ω from 66 breast-tumor subjects; Bayesian-Reduced Gompertz collapses to the line $\alpha = 7.87\beta$; Physics Diffusion exhibits bimodal collapse despite identical physics guidance to our method; Physics OT-CFM produces a smooth distribution that follows the real points and extends moderately beyond their convex hull. Two samples violating $K > V_0$ are omitted here and reported in Table 2.

Quantitative comparison. Physics OT-CFM achieves the lowest W_1 and the lowest Δq_{95}, and the second-lowest Δq_5, ranking in the top-2 on every primary metric. The two Bayesian variants illustrate complementary failure modes of informative-prior approaches: the Bayesian-Gompertz uses Vaghi's variability components increased to accommodate $N = 8$ uncertainty. It produces an unrealistically wide spread reflected in elevated W_1 and Δq_{95}. The Bayesian-Reduced Gompertz achieves the lowest Δq_5 but the highest Δq_{95}, a 30× asymmetry between the two tails arising because its fixed deterministic relationship $\alpha = 7.87\beta$ amplifies the upper tail of the β posterior. Physics Diffusion exhibits the highest W_1, consistent with the bimodal collapse seen in Figure 1; the temporal appearance of this failure is examined in Section 4.4.2.

Manifold structure. Figure 1 visualizes the (α, β) parameter scatter for the four methods compared in the main table. The Bayesian-Reduced Gompertz collapses to the one-dimensional line $\alpha = 7.87\beta$, producing a characteristic line that cannot accommodate the off-line spread of the real samples. The Bayesian-Gompertz relaxes this to a two-dimensional cloud but expands the spread substantially. It leaves the real samples concentrated near its center rather than spanning its support—a direct visualization of the variance inflation noted above. Physics Diffusion exhibits a striking bimodal collapse, with samples concentrated along two thin bands that pass through a subset of the real points while leaving the intermediate region empty; this structural pathology underlies its elevated W_1 in Table 2. Physics OT-CFM produces a smooth distribution that follows the eight real points and extends moderately beyond their convex hull along the principal manifold direction. This indicates learned manifold structure rather than memorized interpolation.

4.4.2 $V(t)/V_0$ Trajectory Comparison. While Section 4.4.1 compares parameter distributions, the practical utility of synthetic samples depends on the growth dynamics they generate. We therefore compare normalized volume trajectories $V(t)/V_0$ across the four (α, β)-equivalent baselines and Physics OT-CFM. The normalization is essential: the Bayesian and flow-based baselines parameterize growth in (α, β) and do not natively encode V_0, so $V(t)$ in absolute units would require an external V_0 assumption. Using $V(t)/V_0$ cancels this dependence (the closed-form solution gives $V(t)/V_0 = \exp((\alpha/\beta)(1 - e^{-\beta t}))$) and isolates the intrinsic growth dynamics encoded by each method.

Trajectory diversity. Figure 2 shows 50 randomly sampled trajectories per method against the eight real curves. All methods reproduce the characteristic monotone-saturating shape of Gompertz growth, but differ markedly in the breadth and density of their generated bundles. Physics OT-CFM produces a continuous family of trajectories that envelop the eight real curves throughout $[0, 30]$ days, with sample density decreasing smoothly from the center to the edge—consistent with the smooth manifold observed in Figure 1.

Failure modes. The baselines exhibit distinct, identifiable failure modes that complement the parameter-space observations. The Bayesian-Gompertz generates a substantial fraction of nearly flat trajectories with $V(30)/V_0 \lesssim 30$, well below the slowest real subject; these arise from samples with very small β that the Vaghi prior assigns nontrivial mass to, and they are the trajectory-space manifestation of the over-spread (α, β) cloud. The Bayesian-Reduced Gompertz inherits the same trajectory range as Physics OT-CFM but with a coarser distribution, since its 1D manifold cannot populate intermediate (α, β) combinations off the line. Physics Diffusion shows the most pronounced pathology: trajectories form two disjoint bundles—one near the upper shell and one near the lower shell of the real data—with the intermediate range nearly empty. This bimodal pattern is the temporal expression of the parameter-space mode collapse seen in Figure 1, providing direct visual confirmation of the elevated W_1 and Δq_5 values reported in Table 2.

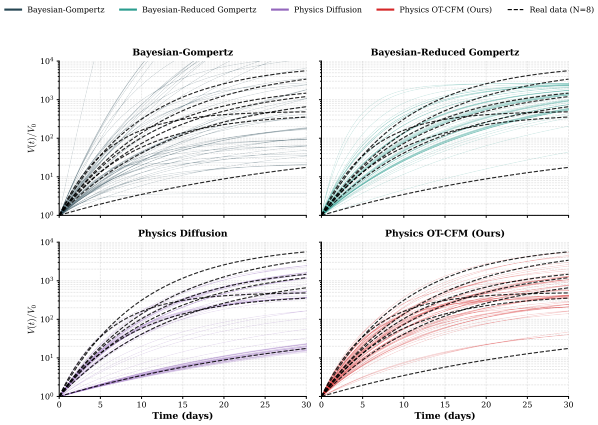


Figure 2: Normalized growth trajectories $V(t)/V_0$ over 30 days. Each panel shows 50 randomly sampled trajectories per method against the eight real curves (black dashed). Normalization removes the dependence on V_0, which is not encoded by the (α, β). Bayesian-Reduced Gompertz and Physics OT-CFM both envelop the real curves, while Bayesian-Gompertz generates substantial data at near-flat trajectories below the slowest real subject, reflecting its over-spread β posterior. Physics Diffusion forms two disjoint bundles—the temporal manifestation of the parameter-space bimodal collapse in Figure 1.

4.4.3 Training Dynamics. We track the evolution of physical compliance and distributional fidelity across 20,000 training steps for the two physics-guided generative models, Physics Diffusion and Physics OT-CFM. Complementary metrics (parameter violation, CVaR_{90}, SWD, MMD^2, projected KS) operate in the full 3D scaled latent space, exploiting the native 3D outputs of both models. Results are shown in Figure 3.

Physical compliance. Figure 3a tracks the parameter violation rate and the conditional tail expectation $\text{CVaR}_{90}(\phi)$ of the physics penalty. Physics Diffusion exhibits a pronounced early-phase instability: the violation rate spikes to nearly 3% around step 2,500 before decaying to zero by step 5,000, and CVaR_{90} correspondingly peaks above 19. Physics OT-CFM, in contrast, remains near zero throughout training, with a single brief excursion around step 9,000 that resolves within the next few thousand steps.

Distributional Convergence. Figures 3b and 3c present complementary distributional metrics. Physics Diffusion attains lower SWD (~ 1.0) than Physics OT-CFM (~ 1.13 at convergence), which is consistent with its tendency to capture projected first and second moments. However, MMD^2 and the projected KS statistic tell a different story: Physics OT-CFM maintains MMD^2 near zero throughout training, while Physics Diffusion remains around 0.10–0.20. Negative values are the standard small-sample artifact of the unbiased estimator and indicate near-identical distributions. The KS statistic is similarly lower for OT-CFM (median $D \approx 0.20$ versus ≈ 0.40), with correspondingly higher Monte-Carlo p-values.

5 Limitations and Ethical Considerations

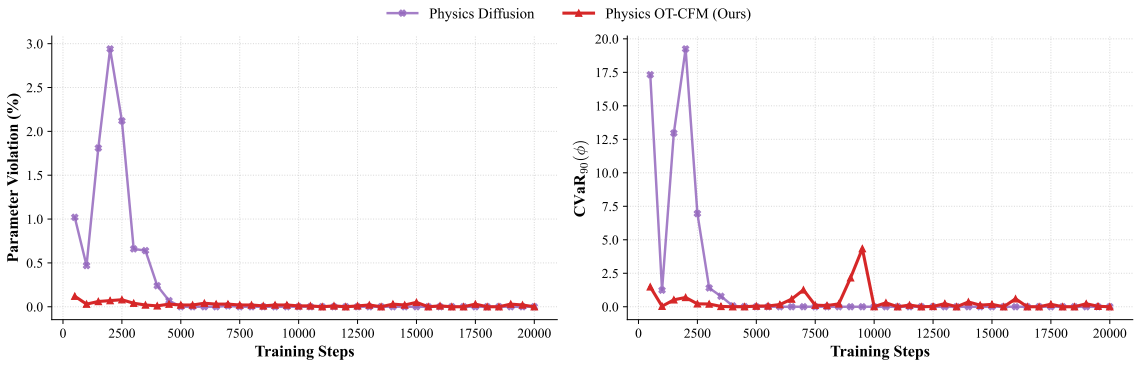
Limitations. While Physics OT-CFM demonstrates robust performance in data-scarce regimes, it has several limitations. First, with only $N = 8$ samples, distributional fidelity metrics have inherently limited statistical power; we therefore frame our quantitative findings as observations supporting the qualitative claim of physically-consistent manifold learning, rather than statistically conclusive comparisons. Larger preclinical cohorts will be needed to establish the generality of the observed ranking. Second, our framework relies on the closed-form solution of the Gompertz growth model; extending it to systems without a closed-form solution would require numerical ODE integration during training, which we have not evaluated in this work. Third, the current study is limited to a single cancer type (glioma) and a single experimental protocol (untreated controls from the TSHS repository); further validation on diverse cancer types and experimental conditions is necessary. Lastly, while our physics-guided rectification ensures consistency with the chosen ODE, it may bias generation toward dynamics expressible by that ODE, potentially overlooking complex biological phenomena that the Gompertz model does not account for, such as angiogenesis or immune system interactions.

Ethical Considerations. Our data come from the publicly available TSHS repository, where the original preclinical study followed institutional animal welfare guidelines and the 3Rs principles; no human identifiers are involved. The generated trajectories are intended for hypothesis generation and preclinical data augmentation only, and should not inform clinical decisions without rigorous validation and regulatory approval.

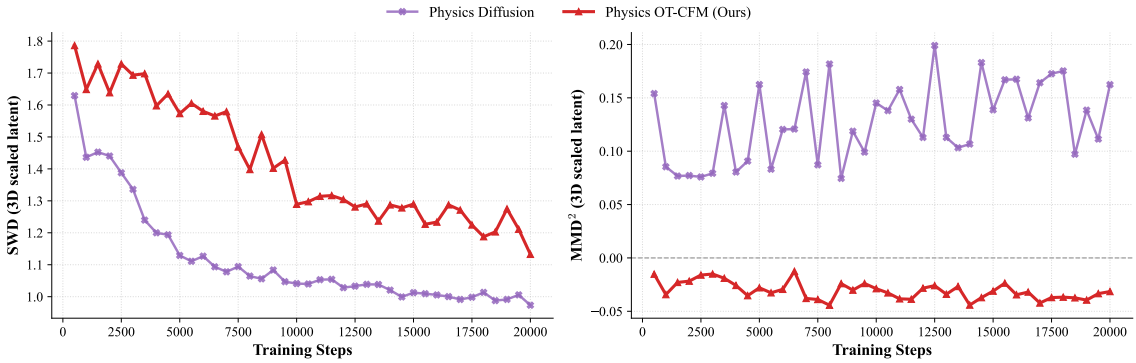
6 Conclusion and Future Work

We introduced **Physics-Guided Optimal Transport Conditional Flow Matching (Physics OT-CFM)** for scientific generation in extreme data-scarce regimes, rectifying the OT-CFM transport field with differentiable gradients induced by Gompertz growth constraints to enforce mechanistic validity during both training and sampling. On glioma xenograft controls with only $N = 8$ subjects, Physics OT-CFM achieves the lowest W_1 and Δq_{95} in the Vaghi-projected 1D distribution and ranks in the top-2 on every primary metric, while maintaining near-zero physics violation throughout training. Bayesian-Gompertz with Vaghi 2020 informative priors inflates the generated spread, showing that informative priors do not suffice in this regime. The Bayesian-Reduced Gompertz variant, which encodes Vaghi's deterministic relationship $\alpha = 7.87\beta$ as a hard constraint, achieves the lowest Δq_5 but the highest Δq_{95}, demonstrating that hard manifold assumptions amplify posterior tails in directions the data do not support. Comparison against a matched physics-guided diffusion baseline shows that the generative paradigm also matters: the diffusion baseline collapses to a bimodal pattern visible in both parameter space and trajectory bundles, illustrating that goodness-of-fit metrics alone can mis-rank generative models in scarce regimes when projected moments mask underlying mode collapse.

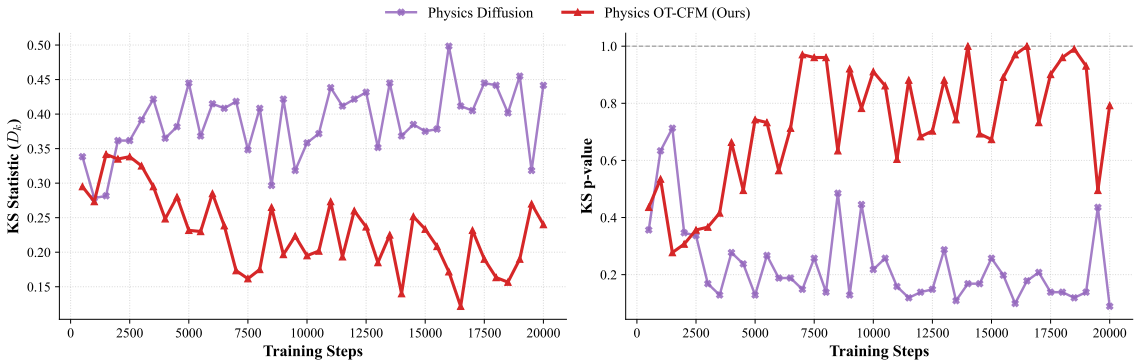
Future work will extend the framework to mechanistic models without closed-form solutions (SDE/PDE), intervention-conditional treatment response generation, and adaptive guidance schedules across varying measurement sparsity.



(a) Physical Violation Rate (%) over training steps.



(b) Global distributional metrics: SWD and MMD^2 in scaled latent space.



(c) Local statistical metrics: KS distance and p-value in scaled latent space.

Figure 3: Training dynamics of the two physics-guided generative models, Physics Diffusion and Physics OT-CFM, evaluated over 20,000 optimization steps within the full 3D scaled latent space. Regarding physical compliance, Physics OT-CFM remains consistently near zero on both the parameter violation rate and the conditional tail expectation (CVaR$_{90}$) throughout the entire training process, whereas Physics Diffusion exhibits a pronounced early-phase instability that only resolves around step 5,000 (a). On the complementary distributional metrics, Diffusion attains a marginally lower SWD, but Physics OT-CFM successfully maintains an unbiased MMD^2 near zero and achieves a substantially lower KS statistic with robustly higher p-values (b, c). This distinct metric dissociation directly reflects the diffusion model's severe bimodal mode collapse and its ultimate failure to capture the full underlying data manifold, as previously documented in Figures 1 and 2.

7 GenAI Disclosure

Generative AI tool (ChatGPT, Claude) was used in this paper for two purposes. First, it was used to check grammar and make English sentences more natural. Second, it was used to get some help for debugging and improving the implementation code. However, all main ideas, model design, experiments, and analysis were done and carefully checked by the authors, and the authors take full responsibility for any remaining errors.

Acknowledgments

Minjung Gim was supported by National Institute for Mathematical Sciences (NIMS) grant funded by the Korea government (MSIT) (No. B26810000). Dongwook Shin was supported by the National Research Foundation of Korea (NRF) grants funded by the Korea governments (Ministry of Science and ICT and Ministry of Education) (2020R1F1A1A01076151, RS-2023-00211854, and RS-2023-00285390). Soon-Sun Kwon was supported by the Ministry of Education and the NRF (RS-2025-00564343) and by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (RS-2025-02310331).

References

[1] Nidal Al-Huniti, Yan Feng, Jingyu Yu, Zheng Lu, Mario Nagase, Diansong Zhou, and Jennifer Sheng. Tumor growth dynamic modeling in oncology drug development and regulatory approval: Past, present, and future opportunities. *CPT: Pharmacometrics & Systems Pharmacology*, 9(8):419–427, 2020.

[2] René Bruno, Dean Bottino, Dinesh P. de Alwis, Antonio T. Fojo, Jérémie Guedj, Chao Liu, Kristin R. Swanson, Jenny Zheng, Yanan Zheng, and Jin Y. Jin. Progress and opportunities to advance clinical cancer therapeutics using tumor dynamic models. *Clinical Cancer Research*, 26(8):1787–1795, 2020.

[3] Elizabeth A. Eisenhauer, Patrick Therasse, Jan Bogaerts, Lawrence H. Schwartz, Daniel Sargent, Robert Ford, Janet Dancey, Stewart Arbuck, Simon Gwyther, Margaret Mooney, Larry Rubinstein, Lalitha Shankar, Lori Dodd, Richard Kaplan, Denis Lacombe, and Jaap Verweij. New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). *European Journal of Cancer*, 45(2):228–247, 2009.

[4] Fang-Shu Ou, Jun Tang, Ming-Wen An, and Sumithra J. Mandrekar. Modeling tumor measurement data to predict overall survival (OS) in cancer clinical trials. *Contemporary Clinical Trials Communications*, 23:100827, 2021.

[5] Michele Simeoni, Paolo Magni, Claudia Cammia, Giuseppe De Nicolao, Valeria Croci, Enrico Pesenti, et al. Predictive pharmacokinetic–pharmacodynamic modeling of tumor growth kinetics in xenograft models after administration of anticancer agents. *Cancer Research*, 64(3):1094–1101, 2004.

[6] Anne Rodallec, Cristina Vaghi, Joseph Ciccolini, Raphaelle Fanciullino, and Sebastien Benzekry. Tumor growth monitoring in breast cancer xenografts: A good technique for a strong ethic. *PLOS ONE*, 17(9):e0274886, 2022.

[7] Gregory Hather, Ray Liu, Syamala Bandi, Jerome Mettetal, Michael Manfredi, Wen-Chen Shyu, Jane Donelan, and Arijit Chakravarty. Growth rate analysis and efficient experimental design for tumor xenograft studies. *Cancer Informatics*, 13(Suppl 4):65–72, 2014.

[8] Jianrong Wu and Peter J. Houghton. Interval approach to assessing antitumor activity for tumor xenograft studies. *Pharmaceutical Statistics*, 9(1):46–54, 2010.

[9] Zinnia P. Parra-Guillen, Annette Berndt, and Ignacio F. Troconiz. Systematic modeling and design evaluation of unperturbed tumor dynamics in xenografts. *Journal of Pharmacology and Experimental Therapeutics*, 366(1):96–104, 2018.

[10] Amy Z. X. Zhu et al. Quantitative translational modeling to facilitate preclinical to clinical efficacy & toxicity translation in oncology. *Future Science OA*, 4(5):FSO306, 2018.

[11] George Em Karniadakis, Ioannis G. Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440, 2021.

[12] Cristina Vaghi, Anne Rodallec, Raphaëlle Fanciullino, Joseph Ciccolini, Jonathan P. Mochel, Michalis Mastri, Clair Poignard, John M. L. Ebos, and Sébastien Benzekry. Population modeling of tumor growth curves and the reduced Gompertz model improve prediction of the age of experimental tumors. *PLOS Computational Biology*, 16(2):e1007178, 2020.

[13] Lei Zhao, Mark A. Morgan, Laura A. Parsels, Jonathan Maybaum, Theodore S. Lawrence, and Daniel Normolle. Bayesian hierarchical changepoint methods in modeling the tumor growth profiles in xenograft experiments. *Clinical Cancer Research*, 17(5):1057–1064, 2011.

[14] Bernard W. Silverman. *Density Estimation for Statistics and Data Analysis*. Monographs on Statistics and Applied Probability. Chapman and Hall, London, 1986.

[15] David W. Scott. *Multivariate Density Estimation: Theory, Practice, and Visualization*. John Wiley & Sons, New York, 1992.

[16] Nitesh V. Chawla, Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer. SMOTE: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16:321–357, 2002.

[17] Sébastien Benzekry, Clare Lamont, Afshin Beheshti, Amanda Tracz, John M. Ebos, Lynn Hlatky, and Philip Hahnfeldt. Classical mathematical models for description and prediction of experimental tumor growth. *PLoS Computational Biology*, 10(8):e1003800, 2014.

[18] Narmin Ghaffari Laleh, Chiara Maria Lavinia Loeffler, Julia Grajek, Kateřina Staňková, Alexander T. Pearson, Hannah Sophie Muti, Christian Trautwein, Heiko Enderling, Jan Poleszczuk, et al. Classical mathematical models for prediction of response to chemotherapy and immunotherapy. *PLOS Computational Biology*, 18(2):e1009822, 2022.

[19] Maziar Raissi, Paris Perdikaris, and George E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019.

[20] Alexander Tong, Kilian Fatras, Nikolay Malkin, Guillaume Huguet, Yanlei Zhang, Jarrid Rector-Brooks, Guy Wolf, and Yoshua Bengio. Improving and generalizing flow-based generative models with minibatch optimal transport. *Transactions on Machine Learning Research*, pages 1–34, 2024.

[21] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 33, pages 6840–6851, 2020.

[22] Hyungjin Chung, Jeongsol Kim, Michael T. McCann, Marc L. Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. In *International Conference on Learning Representations (ICLR)*, 2023.

[23] Jacob K. Christopher, Stephen Baek, and Ferdinando Fioretto. Constrained synthesis with projected diffusion models. arXiv preprint arXiv:2402.03559, 2024.

[24] James D. Murray. *Mathematical Biology I: An Introduction*. Springer, 2002.

[25] Constantine Daskalakis. Tumor Growth Data Set. Consortium for the Advancement of Undergraduate Statistics Education (CAUSE), TSHS Resources, 2019. Data originally derived from human glioma xenograft experiments.

[26] B. Ribba, N. H. Holford, P. Magni, I. Trocóniz, I. Gueorguieva, P. Girard, C. Sarr, M. Elishmereni, C. Kloft, and L. E. Friberg. A review of mixed-effects models of tumor growth and effects of anticancer drug treatment used in population analysis. *CPT: Pharmacometrics & Systems Pharmacology*, 3:e113, 2014.

[27] Manfred Kühleitner, Norbert Brunner, Werner-Georg Nowak, Katharina Renner-Martin, and Klaus Scheicher. Best fitting tumor growth models of the von bertalanffy-pütter type. *BMC Cancer*, 19:683, 2019.

[28] Anna Kane Laird. Dynamics of tumour growth: Comparison of growth rates and extrapolation of growth curve to one cell. *British Journal of Cancer*, 19(2):278–291, 1965.

[29] L. Norton, R. Simon, H. D. Brereton, and A. E. Bogden. Predicting the course of Gompertzian growth. *Nature*, 264(5586):542–545, 1976.

[30] Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David Duvenaud. Neural ordinary differential equations. In *Advances in Neural Information Processing Systems*, volume 31, 2018.

[31] Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. In *Proceedings of the International Conference on Learning Representations*, 2023.

[32] Oriol Abril-Pla, Virgile Andreani, Colin Carroll, Larry Dong, Christopher J Fonnesbeck, Maxim Kochurov, Ravin Kumar, Junpeng Lao, Christian C Luhmann, Osvaldo A Martin, et al. PyMC: A modern, and comprehensive probabilistic programming framework in Python. *PeerJ Computer Science*, 9:e1516, 2023.

[33] Aki Vehtari, Andrew Gelman, Daniel Simpson, Bob Carpenter, and Paul-Christian Bürkner. Rank-normalization, folding, and localization: An improved $\widehat{R}$ for assessing convergence of MCMC. *Bayesian Analysis*, 16(2):667–718, 2021.

[34] Juan Antonio Cuesta-Albertos, Ricardo Fraiman, and Thomas Ransford. Random projections and goodness-of-fit tests in infinite-dimensional spaces. *Bulletin of the Brazilian Mathematical Society, New Series*, 37(4):1–25, 2006.

A Preliminaries

A.1 Gompertz Growth Model

To model baseline tumor growth trajectories, we consider four classical ODE-based growth laws that are commonly used in the tumor growth literature: exponential, logistic, Gompertz, and von Bertalanffy-type models. Our study operates in an extreme data-scarce regime with sparse longitudinal measurements, so we prioritize a parsimonious phenomenological model that captures the key qualitative behaviors of baseline tumor growth, including positivity, growth deceleration, and saturation, while remaining practical under limited data.

In this context, exponential growth can be useful as a short-time approximation but does not represent growth slowdown or saturation at later times. Logistic growth introduces a carrying capacity and deceleration, yet its specific saturation pattern may be less flexible for certain experimental tumor growth datasets. More general families, including von Bertalanffy-type dynamics, can incorporate additional mechanisms at the cost of increased model flexibility, which may lead to practical parameter identifiability issues when follow-up is sparse [17, 26, 27]. Comparative investigations of experimental tumor growth have frequently reported strong descriptive performance for Gompertz dynamics relative to simpler alternatives [12, 17]. Gompertzian growth also has a long history as a standard phenomenological model in tumor growth studies [28, 29].

Let $V(t) > 0$ denote the tumor volume at time t. The Gompertz growth ODE is given by

$$\frac{dV(t)}{dt} = a\,V(t)\,\log\left(\frac{K}{V(t)}\right), \qquad V(0) = V_0, \tag{7}$$

where $a > 0$ represents the growth rate and $K > 0$ denotes the carrying capacity. Here, K represents the long-term plateau level implied by the Gompertz model. Since our longitudinal measurements are sparse, K should be interpreted as an effective asymptotic size inferred from limited observations rather than a directly observed biological maximum.

A useful interpretation follows from the per-capita growth rate:

$$\frac{1}{V(t)}\frac{dV(t)}{dt} = a\,\log\left(\frac{K}{V(t)}\right). \tag{8}$$

Unlike exponential growth, which assumes a constant per-capita growth rate, Gompertz growth implies that the per-capita rate decreases as $V(t)$ approaches K. This provides a compact phenomenological representation of growth deceleration as resource constraints become more pronounced.

Moreover, Gompertz dynamics become linear after a log transform. Define $y(t) = \log V(t)$. Since $dV(t)/dt = V(t)\,dy(t)/dt$, Eq. (7) implies

$$\frac{dy(t)}{dt} = a\big(\log K - y(t)\big). \tag{9}$$

Thus, $\log V(t)$ relaxes exponentially toward $\log K$, making explicit the characteristic behavior of early exponential-like growth followed by progressive slowdown and saturation. In this representation, the parameter a sets the timescale of relaxation in log-space, thereby controlling how rapidly deceleration and saturation emerge.

Equation (7) admits a closed-form solution:

$$V(t) = K\exp\left(\log\left(\frac{V_0}{K}\right)\exp(-at)\right). \tag{10}$$

If $0 < V_0 < K$, the solution remains positive, is monotone increasing, and satisfies $\lim_{t\to\infty} V(t) = K$. Defining the normalized volume $v(t) = V(t)/K$ yields $v(0) = V_0/K \in (0, 1)$ and $v(t)$ increases toward 1 as t increases. These basic qualitative properties are used later to define physically meaningful constraints in our generative modeling framework (valid initial condition $V_0 < K$, monotonic growth, and near-saturation at terminal time).

A.2 Optimal Transport Conditional Flow Matching (OT-CFM)

We consider a flow-based generative model for continuous random variables $z \in \mathbb{R}^d$ with target distribution p_{data}. The goal is to transform samples from a simple base distribution p_0 (e.g., a standard Gaussian) into samples following p_{data} by solving an ordinary differential equation of the form

$$\frac{dz_t}{dt} = v_\theta(z_t, t), \qquad t \in [0, 1], \qquad z_0 \sim p_0, \tag{11}$$

where v_θ is a time-dependent neural vector field. New samples are generated by numerically integrating (11) from $t = 0$ to $t = 1$ [30]. Conditional Flow Matching (CFM) trains the vector field v_θ by matching it to a target velocity defined along interpolation paths between paired samples (z_0, z_1), where $z_0 \sim p_0$ and $z_1 \sim p_{\text{data}}$ [31]. Specifically, one samples $t \sim \text{Unif}[0, 1]$ and defines the linear interpolation

$$z_t = (1 - t)z_0 + tz_1,$$

whose associated velocity is $u^\star(z_t, t; z_0, z_1) = z_1 - z_0$. The CFM objective is given by

$$\min_\theta\ \mathbb{E}\Big[\big\|v_\theta(z_t, t) - (z_1 - z_0)\big\|_2^2\Big], \tag{12}$$

where the expectation is taken over the sampling of (z_0, z_1) and t.

Optimal Transport Conditional Flow Matching (OT-CFM) improves upon CFM by selecting the pairing of (z_0, z_1) using an optimal transport coupling computed over minibatches [20]. By minimizing the expected matching cost, OT-CFM reduces path crossings and leads to more stable training and higher-quality samples, especially in low-sample regimes.

While OT-CFM provides a flexible framework for learning complex distributions, it does not by itself enforce any domain-specific structural constraints. In the context of tumor growth modeling, this may lead to generated samples that violate basic biological properties. In the next section, we introduce Physics OT-CFM by rectifying the transport field with physics-derived gradients so that generated samples remain in biologically meaningful regions of the parameter space.

B Extended Baseline Details

This appendix provides full reproducibility information for the baselines compared in Section 4.2, including the two non-parametric sanity-check baselines (KDE, SMOTE) deferred from the main text. Quantitative results and visualizations covering all six methods are reported in Appendix C.

B.1 Bayesian Baselines: Prior Specification

Both Bayesian baselines (Bayesian-Gompertz and Bayesian-Reduced Gompertz) use log-normal priors on the population mean parameters $\boldsymbol{\mu} = (\mu_\alpha, \mu_\beta)$, with location parameters adopted from Vaghi 2020 [12].

$$\log \alpha \sim \mathcal{N}\left(\log 0.58,\ 0.4^2\right), \tag{13}$$

$$\log \beta \sim \mathcal{N}\left(\log 0.072,\ 0.5^2\right). \tag{14}$$

The location parameters are taken directly from Vaghi 2020 without modification. The scale parameters (0.4 and 0.5) are larger than the published inter-individual variability values, which were estimated from $N = 66$ breast-tumor subjects; we amplify them to reflect the additional uncertainty arising from our $N = 8$ cohort, where a tightly informative prior would dominate the likelihood and produce overconfident posteriors. Prior specifications are summarized in Table B.1.

Table B.1: Prior specifications for Bayesian baselines.

Parameter	Location $\mu_{\log}$	Scale $\sigma_{\log}$
α (population)	$\log(0.58) \approx -0.545$	0.4
β (population)	$\log(0.072) \approx -2.631$	0.5
Bayesian-Reduced Gompertz additionally fixes:		
k (deterministic ratio $\alpha = k\beta$)	7.87	
c (offset)	0	

The Bayesian-Reduced Gompertz variant additionally adopts Vaghi's deterministic relationship $\alpha = k\beta + c$ with $k = 7.87$ and zero residual variance, $c = 0$, yielding samples on the one-dimensional manifold $\alpha = 7.87\beta$. The choice $c = 0$ is justified by the sanity check reported in Section 4.2: a one-sample t-test on per-subject estimates of c yields $t(7) = -1.05$, $p = 0.33$, statistically indistinguishable from zero.

Posterior inference. We perform posterior sampling in PyMC [32] using the No-U-Turn Sampler (NUTS) with 4 chains of 2,000 post-warmup draws each (warmup: 1,000), yielding 8,000 total posterior samples. Convergence is assessed via $\hat{R}$ statistics following Vehtari et al. [33]; all parameters satisfy $\hat{R} < 1.01$.

B.2 Non-Parametric Sanity-Check Baselines

We include two non-parametric baselines as sanity checks against which parametric and learning-based methods can be assessed.

KDE. We fit a Gaussian kernel density estimator on the eight real samples in (α, β) with bandwidth selected by Scott's rule [15], and apply rejection sampling to enforce $\alpha, \beta > 0$ (equivalent to $K > V_0$ under $a > 0$). With $N = 8$, the bandwidth oversmooths the empirical support and extends the tails beyond the observed range.

SMOTE. SMOTE [16] generates synthetic samples via convex interpolation between each real point and its k nearest neighbors ($k = 3$). By construction, all samples lie within the convex hull of the eight real points, providing strict in-sample interpolation but no extrapolation. The same constraint filtering ($\alpha, \beta > 0$) is applied for consistency with KDE.

B.3 Implementation Details

Physics OT-CFM and Physics Diffusion. The time-dependent vector field v_θ is parameterized by a multilayer perceptron (MLP) with three hidden layers (64 units each) and SiLU activations. We use the Adam optimizer with learning rate 3×10^{-4} and batch size 64. To assess training stability under data scarcity, we evaluate 40 checkpoints at 500-step intervals from 500 to 20,000 steps. For physics-guided rectification, we set the barrier sharpness $\tau = 3 \times 10^{-3}$ and guidance weights $\lambda_{\text{train}} = \lambda_{\text{sample}} = 0.2$. Sampling guidance is applied selectively in the later stage of transport ($t \geq 0.5$). The probability flow ODE is integrated via the Euler method with 150 steps. Physics Diffusion uses the identical MLP architecture and physics penalty Φ_{phys}, with gradient guidance applied during denoising. All experiments were implemented in PyTorch and executed on an Apple M3 Max system using the Metal Performance Shaders (MPS) backend.

Table B.2: Hyperparameters for Physics OT-CFM. Physics Diffusion uses the same MLP architecture, optimizer, and physics penalty configuration.

Hyperparameter	**Value**
MLP architecture	3 layers × 64 units
Activation function	SiLU
Optimizer	Adam
Learning rate	3×10^{-4}
Batch size	64
Training step range	500 – 20,000
Euler steps (inference)	150
Guidance weights (λ)	0.2
Barrier sharpness (τ)	3×10^{-3}
Guidance start time	$t = 0.5$
Hardware backend	Apple M3 Max (MPS)

C Extended 1D Projection Analysis

This appendix extends the primary 1D Vaghi-projection analysis (Section 4.4.1, Table 2) along three directions: **(i)** we add the two non-parametric sanity-check baselines (KDE and SMOTE) defined in Appendix B.2 to yield a complete six-method comparison; **(ii)** we report additional distributional metrics (KS statistic with Monte Carlo p-value, unbiased MMD^2) that test for statistical distinguishability; **(iii)** we visualize all six methods in both the (α, β) parameter space and in $V(t)/V_0$ trajectory space.

C.1 Extended Quantitative Comparison

Table C.1 extends Table 2 with the two non-parametric baselines and three additional metrics on the same 1D Vaghi projection β_{proj}:

- KS statistic (D): the supremum of the absolute difference between the two empirical CDFs;
- KS p-value: the right-tail probability of D under the null that both samples come from the same distribution;
- MMD^2 (unbiased): the unbiased U-statistic estimator with an RBF kernel and median-heuristic bandwidth.

Table C.1: Extended 1D distribution fidelity on the Vaghi projection β_{proj} across all six methods. KS p-values > 0.05 indicate samples are not statistically distinguishable; MMD^2 values close to zero indicate near-identical distributions, while a positive MMD^2 indicates a detectable difference (negative values are small-sample artifacts of the unbiased estimator and also indicate near-identical distributions). Bold = best, underlined = second-best.

Method	$W_1 \downarrow$	$\Delta q_5 \downarrow$	$\Delta q_{95} \downarrow$	KS$\downarrow$	$p \uparrow$	MMD^2
KDE	0.0095	0.0111	0.0211	0.223	0.745	−0.042
SMOTE	**0.0063**	0.0036	0.0024	**0.148**	**0.984**	−0.060
Bayesian-Gompertz	0.0135	0.0038	0.0192	0.303	0.381	−0.000
Bayesian-Reduced Gompertz	0.0182	**0.0018**	0.0544	0.265	0.544	−0.001
Physics Diffusion	0.0294	0.0182	0.0068	0.433	0.071	+0.149
Physics OT-CFM (Ours)	0.0082	0.0027	**0.0019**	0.171	0.944	−0.052

Ranking preservation. The qualitative ranking from Table 2 is preserved across all six metrics. SMOTE attains the strongest values on W_1 and KS-derived statistics by virtue of strict convex-hull confinement; Physics OT-CFM ranks second on those metrics and first on Δq_{95}. It achieves top-2 status on every primary and secondary metric. KDE places third on W_1 but exhibits inflated Δq_5 and Δq_{95}. This reflects symmetric over-smoothing of the empirical support under Scott's-rule bandwidth at $N = 8$.

Statistical distinguishability via MMD^2. Among the six methods, only Physics Diffusion yields a positive MMD^2 (+0.149): under the unbiased estimator with median-heuristic bandwidth, this indicates samples that are statistically distinguishable from the eight real subjects. All other methods yield MMD^2 values indistinguishable from zero, including all baselines and Physics OT-CFM. This is consistent with the bimodal collapse seen in Physics Diffusion's parameter scatter (Figures 1 and C.1): even when projected to one dimension, Physics Diffusion's distribution remains structurally separated from the real cohort.

C.2 Visualization of All Six Methods

Figure C.1 visualizes the parameter scatter in (α, β) for all six methods, with the eight real subjects plotted as black diamonds. Figure C.2 shows the corresponding $V(t)/V_0$ trajectory bundles, with the eight real curves shown as black dashed lines. These figures extend Figures 1 and 2 of the main text to include the two non-parametric baselines.

KDE produces a smoothed cloud extending beyond the empirical support, while SMOTE samples are strictly bounded within the convex hull of the eight real points. These structural patterns manifest in the time domain (Figure C.2): KDE generates a wide trajectory bundle extending beyond the real range, while SMOTE produces a narrower bundle confined to interpolation between the real curves.

Structural restriction. The non-parametric baselines achieve their apparent metric advantage through structural mechanisms unique to non-learning approaches. SMOTE's strict convex hull restriction guarantees closeness to the eight training points under any convex distance, which mechanically produces W_1 values. The same property prevents generation of samples outside the empirical support. It

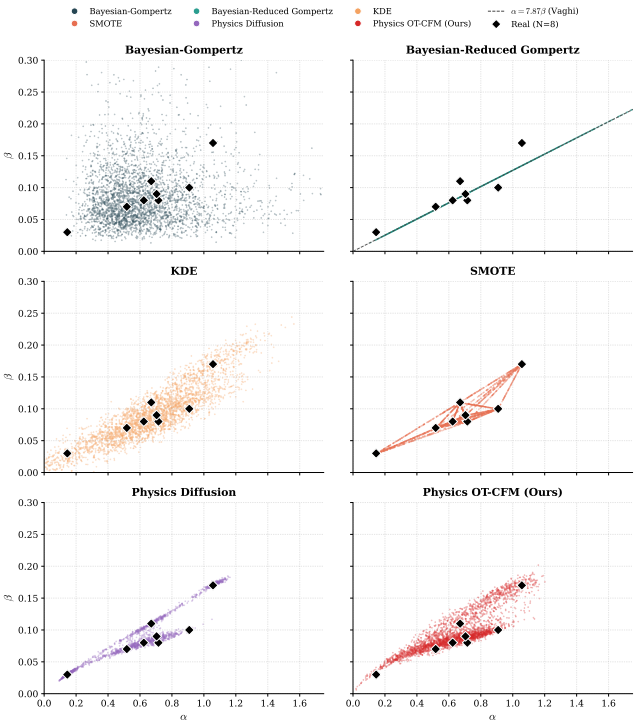


Figure C.1: Parameter scatter in (α, β) for all six methods. Eight real subjects shown as black diamonds; the dashed reference line on the Bayesian-Reduced Gompertz panel indicates the Vaghi relationship $\alpha = 7.87\beta$.

limits practical utility for cohort augmentation or hypothesis driven simulation. KDE's bandwidth controlled smoothing permits mild interpolation, but lacks any mechanism to respect the underlying biological manifold and instead increases both tails symmetrically. In contrast, Physics OT-CFM learns a continuous parameter manifold that follows the empirical support and extends modestly beyond it along directions consistent with the Gompertz dynamics (Figure 1, main text). It provides a generative model in the meaningful sense. It is capable of producing novel yet physically plausible samples beyond the training set.

D Mathematical Derivations

In this section, we provide the detailed mathematical derivations for the physics-guided rectification used in Physics OT-CFM. These derivations ensure the differentiability and stability of the vector field guidance throughout the training and sampling processes.

D.1 Analytical Gradients of the Gompertz Growth Model

While the physics penalty Φ_{phys} is computed in the latent space $\mathcal{Z}$, its components depend on the physical volume dynamics $v(t)$. To avoid the numerical instability and computational overhead associated with finite difference methods, we utilize the analytical solution of the Gompertz ODE. Given the latent vector $\mathbf{z}$, the decoded physical parameters $\mathbf{x} = \{a, K, V_0\}$ are obtained via the

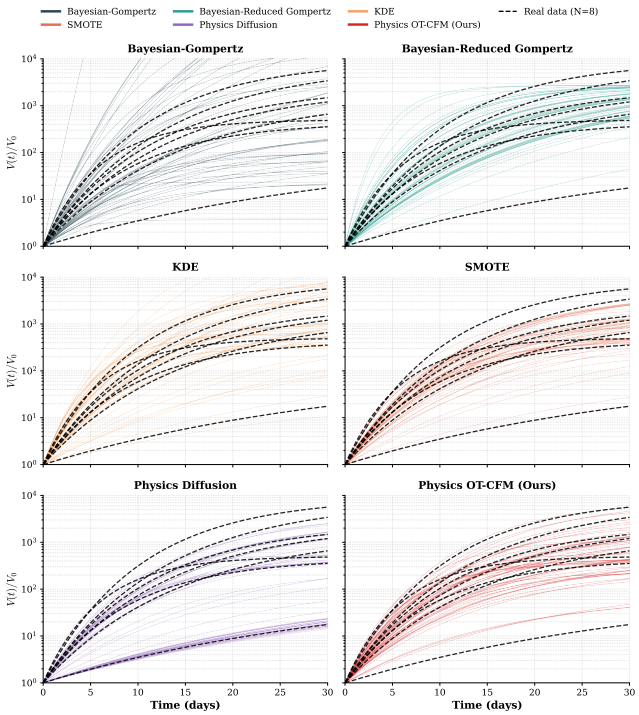


Figure C.2: Normalized growth trajectories $V(t)/V_0$ over 30 days for all six methods. Real curves shown as black dashed.

inverse transform $\mathcal{T}^{-1}$ defined in Eq. (3). The normalized volume $v(t) = V(t)/K$ is expressed as:

$$v(t) = \exp\left(\log\left(\frac{V_0}{K}\right)\exp(-at)\right) \tag{D.1}$$

To enforce the monotonicity constraint $\dot{v}(t) \geq 0$ as described in Section 3.3, we derive the exact time derivative:

$$\dot{v}(t) = \frac{\partial v(t)}{\partial t} = v(t) \cdot \left[a \cdot \log\left(\frac{K}{V_0}\right)\exp(-at)\right] \tag{D.2}$$

Since $v(t) \in (0, 1]$ and $V_0 < K$ is enforced by our latent mapping (Eq. (3)), the term in the bracket remains positive for any $a > 0$. This analytical formulation allows the neural network to efficiently internalize the direction of physically valid growth without relying on numerical approximations.

D.2 Differentiable Barrier via Softplus Function

As noted in Section 3.3, we utilize the Softplus function $\zeta(x) = \log(1 + \exp(x))$ to construct a differentiable barrier for the constraints. For any biological constraint $C(\mathbf{x}) \leq 0$, the penalty $\mathcal{P}(\mathbf{x})$ is formulated as:

$$\mathcal{P}(\mathbf{x}) = \tau \cdot \log\left(1 + \exp\left(\frac{C(\mathbf{x})}{\tau}\right)\right) \tag{D.3}$$

The sharpness parameter τ (set to 3×10^{-3} as detailed in Appendix B.3) ensures the following asymptotic behavior:

$$\lim_{\tau \to 0^+} \mathcal{P}(\mathbf{x}) = \max(0, C(\mathbf{x})) \tag{D.4}$$

This derivation justifies the behavior of our physics-guided vector field: it receives near-zero guidance in physically valid regions ($C(\mathbf{x}) < 0$) while experiencing a strong, differentiable push back toward the manifold in violated regions ($C(\mathbf{x}) > 0$). The choice of a small τ provides a smooth yet tight approximation of the biological bounds. This ensures that the physics-guided gradients do not vanish and remain differentiable, consistent with the smooth activation properties of the SiLU used in our model.

D.3 Chain Rule for Physics-Guided Rectification

The rectified target velocity u_{target} in Algorithm 1 requires the gradient $\nabla_{\mathbf{z}_t}\Phi_{\text{phys}}(\mathbf{z}_t)$. This is computed via the chain rule through the decoding transform $\mathcal{T}^{-1}$:

$$\nabla_{\mathbf{z}}\Phi_{\text{phys}}(\mathbf{z}) = \left(\frac{\partial \mathbf{x}}{\partial \mathbf{z}}\right)^{\top} \nabla_{\mathbf{x}}\Phi_{\text{phys}}(\mathbf{x}) \tag{D.5}$$

where $\frac{\partial \mathbf{x}}{\partial \mathbf{z}}$ is the Jacobian of the inverse transformation derived from the data-based non-dimensionalization in Eq. (3). By internalizing this gradient directly into the target velocity u_{target}, the neural vector field v_θ learns to approximate a transport map that is inherently physics-aware, effectively bridging the gap between sparse observations and biological laws. The Jacobian $\frac{\partial \mathbf{x}}{\partial \mathbf{z}}$ consists of simple diagonal and basic partial derivative elements, ensuring that the gradient computation remains numerically stable and computationally efficient throughout training.

E Evaluation Metrics: Implementation Details

This appendix provides estimator definitions and Monte Carlo calibration details for the metrics used in Section 4.3 and Appendix C. We organize the metrics by scope: primary metrics on the 1D Vaghi projection (Section E.1) and complementary metrics in the 3D scaled latent space (Section E.2).

E.1 Primary Metrics on the 1D Vaghi Projection

For each method, we project samples onto the Vaghi line $\alpha = k\beta$ ($k = 7.87$) using

$$\beta_{\text{proj}} = \frac{k\alpha + \beta}{k^2 + 1}, \tag{E.1}$$

and compare the resulting 1D distributions against the projection of the eight real subjects.

1-Wasserstein distance (W_1). For one-dimensional empirical distributions P and Q with sorted samples $\{a_{(i)}\}_{i=1}^{n}$ and $\{b_{(i)}\}_{i=1}^{m}$, we compute W_1 via the closed-form quantile expression

$$W_1(P, Q) = \int_0^1 \left|F_P^{-1}(u) - F_Q^{-1}(u)\right| du, \tag{E.2}$$

implemented using `scipy.stats.wasserstein_distance`, which handles unequal sample sizes via the empirical CDF integration formulation. No subsampling is applied; all available generated samples (typically ~10,000 per method) are compared against the eight projected real subjects.

Tail-quantile errors ($\Delta q_5, \Delta q_{95}$). We compute empirical quantiles via linear interpolation between order statistics(`numpy.quantile`

with default `linear` interpolation):

$$\Delta q_p = \left|\widehat{q}_p^{\,\mathrm{gen}} - \widehat{q}_p^{\,\mathrm{real}}\right|, \qquad p \in \{5, 95\}. \tag{E.3}$$

With $N_{\mathrm{real}} = 8$, the empirical 5th and 95th percentiles correspond to the lowest and highest real subjects, respectively, by definition of linear interpolation at the extremes. This is the most direct test of whether a generative model recovers the observed support boundary.

Kolmogorov–Smirnov statistic and Monte Carlo p-value. For Appendix C, we additionally report the two-sample KS statistic

$$D = \sup_{u\in\mathbb{R}} \left|F_{\mathrm{gen}}(u) - F_{\mathrm{real}}(u)\right|, \tag{E.4}$$

computed with `scipy.stats.ks_2samp` (asymptotic distribution by default). The associated p-value is the probability of observing a KS statistic at least as extreme as D under the null hypothesis that both samples are drawn from the same underlying distribution.

Maximum Mean Discrepancy (MMD^2, 1D). We use the unbiased U-statistic estimator with an RBF kernel,

$$\begin{aligned}\widehat{\mathrm{MMD}}^2(X, Y) = &\frac{1}{n(n-1)} \sum_{i\neq i'} k(x_i, x_{i'}) \\ &+ \frac{1}{m(m-1)} \sum_{j\neq j'} k(y_j, y_{j'}) \\ &- \frac{2}{nm} \sum_{i,j} k(x_i, y_j),\end{aligned} \tag{E.5}$$

where $k(x, y) = \exp\left(-\gamma\|x - y\|^2\right)$ is the RBF kernel. The bandwidth is selected by the median heuristic on pairwise distances within the combined sample, $\gamma = 1/(2\sigma^2_{\mathrm{med}})$. With $m = 8$ real subjects, the estimator's variance is dominated by the $1/m(m-1)$ term; small-sample artifacts can yield slightly negative values, which are interpreted as "no detectable difference" rather than "samples are closer than identical".

E.2 Complementary Metrics in the 3D Scaled Latent Space

The complementary metrics in Section 4.4.3 operate in the full 3D scaled latent space $\mathcal{Z}$ defined in Section 3.2. They are used only for training-dynamics analysis between Physics Diffusion and Physics OT-CFM, both of which produce native 3D output; they are not applied to the (α, β)-only baselines.

Sliced Wasserstein distance (SWD). For $K = 100$ random unit directions $\{v_k\}_{k=1}^K \subset \mathbb{S}^{d-1}$, we project samples and compute the squared 1D 2-Wasserstein distance between the projected empirical distributions. For equal-sized projected samples $\{a_i\}_{i=1}^n$ and $\{b_i\}_{i=1}^n$, the closed-form expression is

$$W_2^2(P_{v_k}, Q_{v_k}) = \frac{1}{n} \sum_{i=1}^{n} \left(a_{(i)} - b_{(i)}\right)^2, \tag{E.6}$$

and the SWD is the square root of the average over the K projections:

$$\mathrm{SWD}(P, Q) = \left(\frac{1}{K} \sum_{k=1}^{K} W_2^2(P_{v_k}, Q_{v_k})\right)^{1/2}. \tag{E.7}$$

Random directions are drawn uniformly on the sphere by sampling $g_k \sim \mathcal{N}(0, I_d)$ and setting $v_k = g_k/\|g_k\|$.

Maximum Mean Discrepancy (MMD^2, 3D). The same unbiased U-statistic estimator as in Eq. (E.5) is used, applied to the full 3D scaled latent samples with the same RBF kernel and median-heuristic bandwidth.

Projected Kolmogorov–Smirnov statistic. For each random direction v_k, we compute the two-sample KS statistic on projected samples following the random-projection goodness-of-fit framework [34]:

$$D_k = \sup_{u\in\mathbb{R}} \left|F_{X,k}(u) - F_{Y,k}(u)\right|, \qquad D^{(K)} = \max_{1\le k\le K} D_k. \tag{E.8}$$

We additionally compute a Monte Carlo p-value by repeating $B = 1000$ times: drawing independent synthetic sample pairs $(X^{(b)}, Y^{(b)})$ from the model, computing $D^{(K,b)}$, and estimating

$$p_{\mathrm{MC}} = \frac{1 + \sum_{b=1}^{B} \mathbb{I}\left(D^{(K,b)} \ge D^{(K)}_{\mathrm{obs}}\right)}{B + 1}. \tag{E.9}$$

At $N_{\mathrm{real}} = 8$, all methods pass the KS test at $\alpha = 0.05$ owing to limited statistical power; we therefore treat KS as a relative ranking diagnostic across training, not a binary acceptance criterion.