Comparing Neural ODEs to Numerical Stochastic Differential Equations to Forecast Glioblastoma Tumour Growth

Jessica Fu,† Sara Parvaresh Rizi,‡ and Aarya Shah¶

†jessica.fu@mail.utoronto.ca | 1011048607
‡s.parvareshrizi@mail.utoronto.ca | 1010913451
¶aarya.shah@mail.utoronto.ca | 1010871157

Abstract

Glioblastoma, a neoplastic growth of the glial cell, is an aggressive, incurable, and the most common brain tumour. Modern oncology utilizes mathematical modelling, such as ODEs, PDEs, or Neural ODEs, to predict patient survival rates by understanding tumour growth trajectories. However, few studies compare numerical and neural models using stochastic differential equations (SDEs) to account for biological randomness. This research compared Neural ODEs to three stochastic numerical methods: Euler-Maruyama (EM), Milstein, and Fourth-order Stochastic Runge-Kutta (SRK). The models were evaluated on predictive accuracy, convergence, and computational efficiency. The Neural ODE utilized segmented MRIs from the LUMIERE Dataset and a hybrid Markov model to determine cancer stage progression. Statistical evaluation on a 10-patient testing set showed that the Milstein method achieved the highest representative predictive accuracy (MASE = 1.04 ± 0.29). While the Neural ODE demonstrated the fastest stabilization (TCT = 39.31 ± 0.58 weeks) and the lowest representative structural error ($\chi^2 = 4.64 \times 10^5$), it was susceptible to extreme failures in low-data regimes. Thus, this study concludes that the Milstein method is the most reliable for high-precision individual pathwise forecasting, while Neural ODEs offer a computationally efficient alternative for global trend analysis.

Keywords: Stochastic differential equations, ordinary differential equations (ODEs), glioblastoma tumour growth, neural ODEs, Runge-Kutta stochastics, Euler-Maruyama, Milstein, LUMIERE data set, mathematical modelling, Markov transition matrix,

Github Repository: https://github.com/snowxzf/MAT292_Project

1 Introduction

Glioblastoma is the deadliest and most common brain tumour, with 12,000 to 15,000 diagnoses a year in the States alone (MayoClinic, n.d.). Like other tumours, it uses chemotaxis, the directed movement of cells or organisms in response to a chemical gradient, which largely influences tumour growth speed and metastasis. It spreads rapidly and is incurable. Thus, the only possible course of action is to reduce symptoms and slow their growth.

Glioblastoma grows by forming microscopic branches. These branches spread throughout the brain, making it harder to extract the tumour during surgery. The average survival rate is 12-18 months, with only 5 - 10% of patients surviving for more than 5 years (Glioblastoma Foundation 2024).

Glioblastoma tumours are classified based on the width of tumour rings and tumour surface regularity, which assesses the extent to which a tumour resembles a sphere (Pérez-Beteta et al. 2019). Spherical tumours indicate slower progression compared to 'spikey' tumour surfaces (Lin et al. 2025). These factors alter the survival rate of patients, and the correlation can be predicted using partial differential equations (PDEs), ordinary differential equations (ODEs), or machine learning (ML) models.

Understanding glioblastoma growth trajectories through mathematical modelling and ordinary differential equations is useful in analyzing factors that impede or catalyze tumour growth, which can be used to model and hypothesize potential treatment methods. However, due to the large amount of randomness associated with each patient, stochastic methods were used instead.

The aim of this paper is to compare three different numerical, and one neural analysis methods using stochastic differential equations to find the optimal model for comparing predicted tumour growth trajectories, to be used to create more adept models for glioblastoma prevention.

2 Mathematical Model and Theoretical Foundation

2.1 The Core Model: Gompertz

ODEs and PDEs help determine the evolution rates of tumours and how factors like glucose levels, chemotaxis, chemo-radiation, and different treatments can perturb the proliferation of the cells. One prevalent model within computational biology is the Gompertz model, which describes tumour growth as experiencing an "initial exponential growth" before inflecting as it reaches a carrying capacity or maximum size (Z. Ma et al. 2020). The Gompertz model for tumour growth can be modelled by the following equation (Murphy, Jaafari, and Dobrovolny 2016):

$$\frac{\mathrm{d}V}{\mathrm{d}t} = a\,V\ln\left(\frac{b}{V}\right) \tag{1}$$

Where a is the intrinsic growth rate, V is the tumour volume, and b represents the carrying capacity where cell growth slows due to the depletion of oxygen and nutrients (Z. Ma et al. 2020). However, a stochastic differential equation (SDE) was used to represent this behaviour instead of a deterministic ODE, as it better accounts for varying uncertainty and randomness based on individual patient health. In clinical studies related to glioblastoma, the stochastic differential model notes an inverse relationship between tumour volume and "specific growth rate" (SGR) (Slavkova et al. 2023), which is incorporated into the Gompertz framework as follows (Murphy, Jaafari, and Dobrovolny 2016):

$$dV = aV\ln\left(\frac{b}{V}\right)dt + \frac{cV}{h+\sqrt{V}}dW(t) \tag{2}$$

Note that the standard SDE is defined as $dV_t = f(V_t,t)dt + g(V_t,t)dW_t$, where $f(V_t,t)$ is the drift term and $g(V_t,t)$ is the diffusion term (Lalley 2016). The diffusion term, in this case, $\frac{cV}{h+\sqrt{V}}$, accounts for "white noise" or measurement randomness inherent in MRI data, where $W(t)$ is a standard Wiener process representing the cumulative probability distribution of the noise (Brownian motion, n.d.). The drift function for this specific Gompertz model is: $aV\ln\left(\frac{b}{V}\right)$, showing the predictable growth of the tumour.

2.2 Transition to a Time-Dependent Gompertz Equation

While Equation (2) provides a strong biological basis, the "logarithmic" nature of the Gompertz drift can become numerically unstable when tumour volumes V are very small or approach zero in clinical data. As well, classical ODE tumour growth models (e.g., logistic or Gompertz) often treat the growth rate and carrying capacity as static parameters. In clinical reality, treatment interventions, microenvironment changes, and therapy windows cause both the effective growth rate and carrying capacity to vary in time (Giorno and Nobile 2025). To mitigate this while maintaining the sigmoidal growth characteristic, we modified the model using the following SDE:

$$dV = \left(a(t)V\ln(\frac{b(t)}{V}) - k(t)V\right)dt + \frac{cV}{h+\sqrt{V}}dW \tag{3}$$

This was mostly inspired by the work of Vaghi et al. 2020 and Gerlee 2012, who argued that effective growth and capacity are dynamic rather than constant. Building on these insights, we modified the model to include time-dependent parameters $a(t)$ and $b(t)$, and a piecewise chemotherapy decay term $k(t)$. Following the methodologies of Simeoni et al. 2004. and Schwartz 1961, $k(t)$ was implemented as a piecewise function to account for specific therapy windows. It represents the chemo-induced decay rate, defined piecewise to account for the treatment window:

$$k(t) = \begin{cases} k_0 & t \in [t_{\text{chemo, start}}, t_{\text{chemo, end}}] \\ 0 & \text{otherwise} \end{cases}$$

The transitions for $a(t)$ and $b(t)$ represent time-dependent growth rates and carrying capacities, respectively. c and h are stochastic parameters, and W is the Wiener process, such as in Equation (2). To capture gradual shifts in growth and capacity arising from chemotherapy onset or vascular adaptation, hyperbolic-tangent and logistic-sigmoid transitions were used:

$$a(t) = a_1 + a_2\tanh(a(t-t_c))$$

$$b(t) = b_1 + b_2\sigma(\beta(t-t_b)), \quad \sigma(x) = \frac{1}{1+e^{-x}}$$

While the hyperbolic tangent and sigmoid were chosen for smoothness, other monotone transition functions would yield qualitatively similar dynamics. For more details on the iterations used to achieve this formula, see Appendix 7.7. While the exact functional form is motivated by the desire for smooth transitions, the underlying concept of time-varying parameters remains consistent with current literature (Al-Huniti et al. 2020). For implementation within the numerical solvers, the discrete version of the formula is:

$$\begin{aligned} V_{n+1} = V_n + \left(a(t_n)V_n\ln\left(\frac{b(t_n)}{V_n}\right) - k(t_n)V_n\right)\Delta t \\ + \frac{cV_n}{h+\sqrt{V_n}}\sqrt{\Delta t}\,\xi_n, \quad \xi_n \sim N(0,1) \end{aligned} \tag{4}$$

Based on the modified model, the drift and diffusion functions for Equation (3) are:

- Drift term: $f(V,t) = a(t)V\ln\left(\frac{b(t)}{V}\right) - k(t)V$
- Diffusion term: $g(V,t) = \frac{cV}{h+\sqrt{V}}$

3 Methodology

3.1 Data Acquisition and Cohort Selection

To ensure clinical relevance, this study utilized the LUMIERE dataset, a comprehensive single-center collection of longitudinal Glioblastoma (GBM) MRI scans accompanied by expert ratings (Suter et al. 2022). Unlike many publicly available datasets that primarily feature pre-operative imaging, LUMIERE provides fully longitudinal follow-up data, which is important for refining treatment response assessment.

The dataset includes 2,487 images across 91 patients. For this research, the cohort was divided into a training set of 80 patients and a testing set of 10 patients (Sivakumar, Parthasarathy, and Padmapriya 2024). The testing patients were strategically selected based on their longitudinal depth, specifically choosing those with the highest number of available weeks to provide a lengthy temporal baseline for comparing predictive accuracy.

3.2 Data Processing and Segmentation

Accurate tumour burden quantification is required for the accurate forecasting of growth trajectories. Manual identification of abnormal tissue in complex cranium scans is tedious; therefore, this study employed state-of-the-art automated segmentation tools. Using the `hd-glio-auto` framework, based on the U-Net deep learning architecture, active tumour volumes were extracted from contrast-enhanced T1-weighted and T2-weighted MRI sequences (Walsh et al. 2022, Suter et al. 2022). Volumetric data were normalized to week counts relative to the pre-operative acquisition to blur precise follow-up intervals while preserving temporal precision.

3.3 Experimental Design and Model Training

The study compared the forecasting capabilities of Neural Ordinary Differential Equations (Neural ODEs) against classical numerical Stochastic Differential Equation (SDE) solvers. The data partitioning logic for each approach is summarized in Table 1.

Table 1. Logic for Data Partitioning and Model Evaluation

Model Type	Training Phase	Testing Phase
Neural ODE	Learns transition dynamics from 80-patient training set.	Evaluated on full temporal span of 10-patient test set.
Numerical	N/A (Calibrated per patient during testing phase).	Calibrated on initial 50% of points; forecasts remaining 50%.

Using the given `runhdglio` SH file and `run`.py file, the images were successfully segmented and stored. The tumour volume was extracted at each stage by running `extract_volumes` SH and `create_volumes` SH files. Running these files with the testing and training set produced `all_tumour_volumes_hdglio_test` and `all_tumour_volumes_hdglio_train`, which were used for all the following methods.

3.4 Neural Prediction of Tumour Growth

Neural ODE models view the underlying biological growth process as a system of ordinary differential equations where the vector field is parameterized by an artificial neural network (Blaom and Okon 2025). Following the *HybridStateODE* architecture, the model in this report utilized the training set to learn Markov state transition probabilities between various clinical states: Operative (OP), Stable Disease (SD), Progressive Disease (PD), and Partial/Complete Response (PR/CR) (Rausand, Garg 2025). Note that the definition of *HybridStateODE*, Markov transition matrix, and each state in the matrix is written in the glossary (Appendix 7.1). The network used the universal approximation theorem to interpolate continuous growth rates based on these learned discrete states.

The baseline training model provided with the LUMIERE study, implemented in statespace.py, incorporated tumour volume, patient demographic information (methylation status), and expert clinical annotations (Suter et al. 2022). This model was run with the training dataset, generating two Markov state transition matrices, one for methylated and one for unmethylated patients (see Glossary, 7.1). The resulting transition matrices quantified the probability of a patient transitioning to the next clinical state given their current clinical state.

To extend the Markov state transition matrix to a continuous tumour volume progression representation throughout the weeks rather than a state-based prediction, a hybrid modelling approach was implemented. This model combines state probability vectors with the initial tumour volumes to infer the patient's clinical state and corresponding volume progression at each time point.

In this hybrid neural model, the *SmallNet* and *HybridStateODE* components were used to train the neural model by assigning each of the six clinical states in the Markov transition matrix with a corresponding tumour growth rate (different rates for methylated and unmethylated patients). These state-specific growth rates, combined with the state transition probabilities, provide a weighted estimate of tumour growth that accounts for the patient's methylation status. The code then uses Runge-Kutta to perform numerical analysis with each step, determining how the growth rate changes with time and allowing for a more accurate growth trajectory, while having a sustainable runtime.

To quantify uncertainty in the tumour volume growth trajectories, the hybrid model integrates the Monte Carlo Prediction framework. It repeatedly generated predictions with injected stochastic noise, thus simulating the behaviour of a stochastic differential equation in numerical analysis. The model then created an average of these sums, allowing for the determination of the confidence intervals of the plots. All these procedures were performed exclusively on the training dataset, after which the fully trained models were evaluated on the testing dataset for comparison with numerical methods.

3.5 Numerical Implementation of Stochastic Solvers

The stochastic logistic growth model was numerically integrated using three distinct integration schemes: Euler-Maruyama (EM), the Milstein method, and a Strong Runge-Kutta (SRK) approximation. These solvers were implemented in Python to simulate tumour volume trajectories and evaluate predictive stability against clinical data.

3.5.1 Ensemble Simulation and Parameter Estimation

Similar to the neural method, to account for the inherent stochasticity of the diffusion term ($g(V,t)dW$), all numerical

solvers were implemented using a Monte Carlo ensemble approach. The adoption of a Monte Carlo ensemble approach is necessitated by the non-linear dynamics inherent in the stochastic logistic growth model (Wunsch). In systems where either the evolution equation or the measurement process is non-linear, standard linear estimation methods, such as the Kalman filter, fail to rigorously account for the sum of errors from initial conditions and unknown controls. For each patient in the test set, an ensemble of $M = 60$ independent trajectories was simulated. This ensemble size was chosen as a trade-off between statistical significance (to calculate the analytic mean and 95% confidence intervals) and computational runtime. Parameters for the time-varying growth transitions (a_1, a_2), the carrying capacity stages (b_1, b_2), and the stochastic volatility (c, h) were estimated for each patient. The estimation utilized a combination of Global Optimization (Differential Evolution) and local refinement (Least Squares) to minimize the residual sum of squares between the ensemble mean and observed MRI volumes. The inclusion of the piecewise decay parameter k_0 allowed the model to account for specific chemotherapy windows during the estimation process. The final forecast was derived from the arithmetic mean of these trajectories, while the 95th percentile was used to establish confidence intervals around the predicted tumour volume.

Following the out-of-sample evaluation framework established by Blaom and Okon (Blaom and Okon 2025), the simulation was split into two phases:

1. **Calibration Phase**: The first 50% of a patient's temporal data was used to estimate the model's dynamic parameters, including the growth transition coefficients (a_1, a_2), the carrying capacity stages (b_1, b_2), the chemotherapy decay rate (k_0), and the stochastic scaling factors (c, h). This was done using a combination of Differential Evolution for global search and Least Squares for local refinement.
2. **Prediction Phase**: The calibrated model then simulated the remaining 50% of the timeline, which was hidden from the optimizer to evaluate the method's predictive accuracy against clinical reality.

3.5.2 *Discretization and Step-Size Control*

For the baseline predictive fitting, a uniform simulation step size of $\Delta t = 0.1$ weeks was utilized. This choice was informed by exploratory experiments (see document in Appendix 7.7), which found that further refinement offered diminishing returns for predictive accuracy while significantly increasing the computational overhead of the Monte Carlo simulations. For formal convergence analysis, the solvers were evaluated across a refinement grid of $\Delta t \in \{0.8, 0.4, 0.2, 0.1, 0.05\}$.

3.5.3 *Stochastic Integration Schemes*

The selection of numerical solvers was designed to evaluate a progression of complexity and accuracy, moving from a simple first-order baseline to higher-order schemes capable of handling the non-smoothness of Wiener processes. The **Euler-Maruyama (EM)** method was selected as the essential first-order baseline. While it is widely popular due to its ease of implementation, it possesses a weak convergence order of 1.0 and a low strong convergence order of 0.5, which can lead to significant path-wise errors as step sizes increase. To address these limitations, the **Milstein method** and **Stochastic Runge-Kutta (SRK)** variants were implemented as the two most common "improved versions" of the EM baseline. These methods improve accuracy by accounting for higher-order terms in the SDE expansion that the EM method ignores (Zilliz, n.d.). The specific mathematical implementation and justification for each solver are summarized as follows:

1. **Euler-Maruyama (EM)**: Serves as the fundamental first-order baseline for numerically solving Stochastic Differential Equations (SDEs) (Higham and -Maruyama). While computationally simple, it is limited by a strong convergence order of 0.5, meaning errors decrease slowly as time steps are refined (Higham 2001). It updates the volume V by combining a deterministic drift term and a stochastic diffusion term scaled by a Wiener process increment:

$$V_{n+1} = V_n + \left[a(t_n) V_n \ln\left(\frac{b(t_n)}{V_n}\right) - k(t_n) V_n \right] \Delta t + \left[\frac{cV_n}{h + \sqrt{V_n}} \right] \Delta W_n \tag{5}$$

2. **Milstein Method**: Preferred over the EM scheme because it offers significantly improved path-wise accuracy (strong order 1.0 convergence) by incorporating a second-order Itô-Taylor expansion term (Giles). This is a correction term, involving the derivative of the diffusion coefficient, $g'(V, t)$, which effectively reduces errors caused by the stochastic term in systems with non-negligible noise dependencies. This method is crucial for handling multiplicative noise, where the random fluctuations scale with the tumour volume. For our model, the derivative of the noise term with respect to V is $g'(V) = \frac{c(h + \frac{1}{2}\sqrt{V})}{(h+\sqrt{V})^2}$, therefore:

$$V_{n+1} = V_n + f(V_n, t_n)\Delta t + g(V_n, t_n)\Delta W_n + \frac{1}{2}\left[\frac{cV_n}{h + \sqrt{V_n}}\right]\left[\frac{c(h + \frac{1}{2}\sqrt{V_n})}{(h + \sqrt{V_n})^2}\right]\left((\Delta W_n)^2 - \Delta t\right) \tag{6}$$

   By accounting for the interaction between Brownian motion and the system's state through the gradient of the diffusion function, Milstein achieves higher precision more efficiently than the EM baseline.
3. **Strong Runge-Kutta (SRK)**: The SRK (Order 1.0) is a derivative-free method. (Solin 2014). It is included as a generalization of the Milstein method that uses intermediate stages to approximate higher-order terms. It approximates the higher-order terms by evaluating the diffusion at an auxiliary "predictor" point, $\Upsilon_n = V_n + f(V_n, t_n)\Delta t + g(V_n, t_n)\sqrt{\Delta t}$. In our case, SRK reduces to:

$$V_{n+1} = V_n + f(V_n, t_n)\Delta t + g(V_n, t_n)\Delta W_n + \frac{1}{2\sqrt{\Delta t}}\left[g(\Upsilon_n, t_n) - g(V_n, t_n)\right]\left((\Delta W_n)^2 - \Delta t\right) \tag{7}$$

   where $\Upsilon_n = V_n + f(V_n, t_n)\Delta t + g(V_n, t_n)\sqrt{\Delta t}$. This method is particularly useful when precise trajectory simulation is required, but explicit analytical derivatives are computationally expensive or unavailable. SRK methods generalize Runge-Kutta stages to account for the low regularity (Hölder continuity of 1/2) of Wiener processes. This formulation is particularly advantageous as it provides the same higher-order accuracy as Milstein but is derivative-free, eliminating the need for an analytical derivative of the diffusion term while remaining computationally efficient. As seen in the Github Repository, a four-step SRK was used, as it balances computational accuracy and efficiency (Runge-kutta methods and butcher tableau, n.d.).

By limiting the scope to these three methods, the study focuses on the most robust and standard tools implemented in modern

libraries like `torchsde` or `Diffrax`, allowing for a direct comparison of the trade-offs between computational cost, stability, and trajectory accuracy.

Detailed implementation scripts for the fitting loops, global optimization bounds, and Monte Carlo prediction logic are available in the project's Github Repository.

3.6 Analysis of Performance

The numerical stochastic methods and the Neural ODE framework were evaluated based on predictive accuracy, computational efficiency, and convergence stability. The following metrics were selected to provide a multi-faceted assessment of model performance, addressing accuracy, bias, and structural fit.

3.6.1 Accuracy and Predictive Metrics

To quantify the "goodness of fit" between the simulated trajectories and the clinical MRI observations, four statistical metrics were utilized:

1. **Kling-Gupta Efficiency (KGE)**: KGE is a comprehensive indicator designed to address the limitations of traditional correlation-based metrics by decomposing the model-data fit into three independent components: correlation (r), variability (α), and bias (β) (Schoups and Vrugt 2010). This formulation enables a more informative comparison in comparing simulations to real observations by specifically quantifying these factors. It is expressed as:

$$KGE = 1 - \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2} \tag{8}$$

where r is the correlation coefficient, α is the ratio of standard deviations, and β is the ratio of means. A higher KGE score indicates a superior fit.

2. **Nash-Sutcliffe Efficiency (NSE)**: NSE is a normalized statistic that determines the relative magnitude of the residual variance compared to the measured data variance (Nash and Sutville 1970). While useful in calibrating to a "naive" model baseline, it is sensitive to small changes in input parameters. Thus, it was used in conjunction with KGE to provide further insights into fit uncertainty. It is defined as:

$$NSE = 1 - \frac{\sum(y_{obs} - y_{sim})^2}{\sum(y_{obs} - \bar{y}_{obs})^2} \tag{9}$$

where $y_{obs,i}$ is the observed tumour volume from MRI at time i, $y_{sim,i}$ is the simulated tumour volume (ensemble mean) predicted by the model at time i, $\bar{y}_{obs}$ is the mean of all observed tumour volumes for that specific patient, and n is the total number of observations in the evaluation period. For this metric, values closer to 1 indicate better predictive ability.

3. **Mean Absolute Scaled Error (MASE)**: the MASE metric is used as the primary measure of forecast accuracy due to its scale-invariance and robustness against outliers (Hyndman and Koehler 2006). This metric uses the in-sample mean absolute error and a naive forecast method to quantify the accuracy of extrapolation and predictions. It is expressed as:

$$MASE = \frac{\frac{1}{n}\sum_{t=1}^{n}|e_t|}{\frac{1}{n-1}\sum_{t=2}^{n}|y_t - y_{t-1}|} \tag{10}$$

where e_t is the forecast error at time t, n is the number of observations in the prediction (testing) phase, and $|y_t - y_{t-1}|$ is the "naive" forecast error, representing the absolute difference between consecutive historical observations. MASE allows for comparing accuracy across different patient data scales, which is critical for tumour modelling where growth scales differ between individuals. A lower MASE score indicates higher accuracy.

4. **Chi-Squared (χ^2)**: The χ^2 statistic is employed to evaluate the goodness-of-fit between observed results and expected simulated results to determine if a significant relationship exists between the studied variables (Ruiz-Arrebola et al. 2021). It assesses whether deviations are statistically significant or the result of stochastic noise. It is defined as:

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} \tag{11}$$

where O_i is the observed value and E_i is the expected value. A lower χ^2 value indicates that the model more closely reflects the observed data.

3.6.2 Convergence and Stability Analysis

Numerical stability and computational efficiency were evaluated through the following metrics:

1. **Weak Convergence**: Weak approximations are used to assess the average behavior of the statistical convergence (Breheny 2024). A numerical method is said to have weak convergence of order β if, for a sufficiently smooth function g, there exists a constant C such that:

$$|E[g(V_T)] - E[g(\hat{V}_T)]| \leq C(\Delta t)^{\beta} \tag{12}$$

where V_T is the true solution and $\hat{V}_T$ is the numerical approximation at time T, Δt is the discretization time step, and $E[\cdot]$ is the expectation operator, calculated in this study as the arithmetic mean across the $M = 60$ independent Monte Carlo trajectories. This metric assesses how the ensemble mean behaviour stabilizes as Δt decreases.

2. **Strong Convergence**: Strong approximation focuses on the accuracy of individual sample paths (Breheny 2024). A method has strong convergence of order γ if there exists a constant C such that:

$$E[|V_T - \hat{V}_T|] \leq C(\Delta t)^{\gamma} \tag{13}$$

This method evaluates how quickly the numerical approximation approaches the "true solution" on a path-by-path basis using mean-square convergence.

3. **Trajectory Convergence Time (TCT)**: The TCT highlights the time t^* required for a process to reach stable, predictive behaviour within a predefined tolerance ϵ (Wasserman). It is defined as the earliest time point such that the relative change in the ensemble mean remains below ϵ for a sustained window ω:

$$TCT = \min\{t \in [0,T] \mid \forall \tau \in [t, t+\omega], \frac{|\bar{V}_\tau - \bar{V}_{\tau-\Delta t}|}{\bar{V}_{\tau-\Delta t}} < \epsilon\} \tag{14}$$

where $\bar{V}_\tau$ is the ensemble mean of the predicted tumour volume at time τ, calculated across all $M = 60$ Monte Carlo trajectories. In this study, $\epsilon = 10^{-4}$, representing a 0.01% maximum relative change, and $\omega = 5$ consecutive steps, to ensure the convergence is not a temporary fluctuation.

4. **Trajectory Runtime Analysis**: This serves as the primary indicator for model efficiency by measuring the computational "wall-clock" time required for a trajectory to complete or converge (H. Ma et al. 2025):

$$\text{Runtime} = t_{\text{end}} - t_{\text{start}}$$

where t_{end} is the end time and t_{start} is the start time of the script. This metric allows for a direct comparison between the computational overhead of training Neural ODEs versus the iterative steps of stochastic solvers.

4 Results

The performance of the four modelling approaches, being Euler-Maruyama (EM), Milstein, Strong Runge-Kutta (SRK), and Neural ODE, was evaluated across ten test patients. The results are categorized into predictive accuracy, convergence stability, and computational efficiency. A sample patient, 67, from the testing set was selected to portray the performance of the four methods, shown in Figure 1.

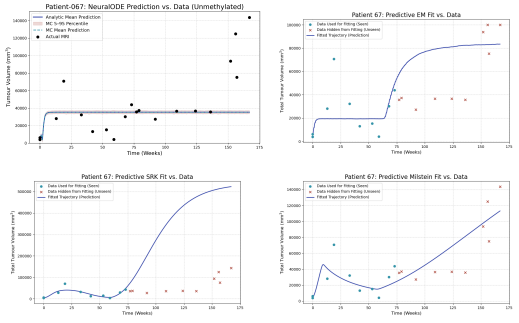

Figure 1. Outputs of each method for selected patient 67.

As seen by Figure 1, the numerical methods had access to the first half of the patient's data (in blue), and were allowed to approximate the second half (in red). The **Neural ODE** converged rapidly to a stable asymptotic regime governed by the learned Markov state dynamics. In this example, SRK seems to overestimate the tumour growth; however, this was not the case for all patients (see Appendix 7.5 for another SRK sample). As well, the performance of the neural ODE on the batch of all testing patients is shown in Appendix 7.6.

4.1 Predictive Accuracy

The stochastic numerical methods were fitted to the initial 50% of patient data, with the remaining 50% used to test extrapolation accuracy.

The statistical evaluation of the models, summarized in Table 2, indicates that the **Milstein** method achieved the highest raw predictive accuracy. It produced the lowest Mean Absolute Scaled Error (MASE = 1.04 ± 0.29) and the best Nash-Sutcliffe Efficiency (NSE = -0.00 ± 0.32), suggesting that its second-order approximation of the diffusion term is more effective at capturing the stochastic growth paths of glioblastoma than first-order or neural alternatives.

However, the aggregate means for several metrics—most notably the χ^2 statistic—were heavily skewed by astronomical errors in a small subset of outlier patients where simulated trajectories diverged significantly from MRI observations. This is particularly evident in the **Neural ODE** results, where the global mean χ^2 reached 2.69×10^{20}. To determine the representative performance for the majority of the patients, a secondary analysis was performed using Interquartile Range (IQR) clipping to remove these extreme statistical outliers, as shown in Table 3 (Sharma, n.d.). See Appendix 7.4 for more details on the clipped data.

Table 2. Predictive Performance Summary (Mean ± Std)

Method	MASE	Chi-Sq (χ^2)	NSE	KGE
EM	1.05 ± 0.28	$(1.25 \pm 3.43) \times 10^{10}$	-0.01 ± 0.33	0.06 ± 0.50
Milstein	1.04 ± 0.29	$(2.03 \pm 6.43) \times 10^{15}$	-0.00 ± 0.32	0.05 ± 0.50
SRK	1.26 ± 0.50	$(7.79 \pm 24.5) \times 10^{10}$	-0.40 ± 1.06	-0.04 ± 0.52
NeuralODE	1.44 ± 1.23	$(2.69 \pm 8.51) \times 10^{20}$	-0.50 ± 1.42	0.05 ± 0.75

Table 3. Clipped Predictive Performance Summary (Outliers Removed via IQR)

Method	MASE	Chi-Sq (χ^2)	NSE	KGE
EM	1.06 ± 0.28	$(2.83 \pm 3.83) \times 10^{6}$	-0.02 ± 0.33	0.05 ± 0.50
Milstein	1.04 ± 0.29	$(2.74 \pm 3.78) \times 10^{6}$	-0.00 ± 0.32	0.05 ± 0.50
SRK	1.26 ± 0.50	$(4.83 \pm 4.76) \times 10^{6}$	-0.16 ± 0.39	-0.04 ± 0.52
NeuralODE	1.21 ± 0.67	$(4.64 \pm 4.26) \times 10^{5}$	-0.19 ± 0.82	0.08 ± 0.68

Under these clipped conditions, the **Neural ODE** demonstrated a drastic reduction in structural error, yielding the lowest representative mean χ^2 ($(4.64 \pm 4.26) \times 10^5$). This suggests that while Neural ODEs may be prone to extreme failures in low-data regimes, they provide a highly efficient fit for typical growth patterns. In contrast, the stochastic numerical solvers provided superior path-wise tracking, though the first-order **EM** baseline yielded significantly higher variance in its χ^2 results compared to higher-order methods.

The plots for the performance of each one based on the four prediction metrics are shown in Figure 2. The red dots indicate the aggregate mean, while the plots themselves are representative of the mean of the clipped data.

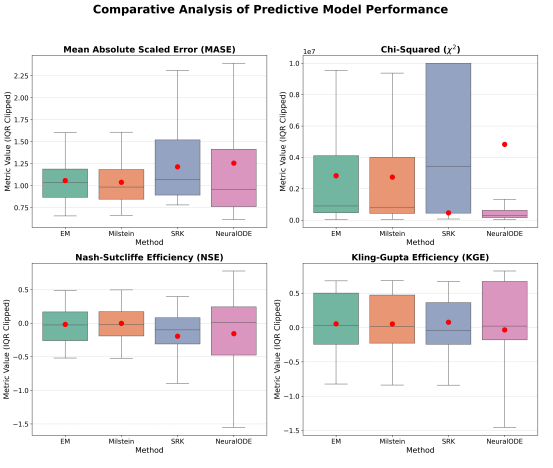

Figure 2. Predictive performance across MASE, χ^2, NSE, and KGE.

Among the four methods:

1. **Milstein Method**: Achieved the highest representative predictive accuracy with a clipped average MASE of 1.04 ± 0.29 and a mean NSE of -0.00 ± 0.32. It demonstrated a superior ability to capture sigmoidal growth trajectories while maintaining stable confidence intervals during Monte Carlo simulations.
2. **SRK Method**: While producing a higher MASE (1.26 ± 0.50) compared to Milstein, the SRK method remained competitive in terms of structural fit. Its higher-order derivative-free structure provided a consistent representative χ^2 value on the order of 10^6, demonstrating relative robustness across the majority of the testing cohort.
3. **Neural ODE**: The neural framework showed significant improvement when extreme outliers were removed, yielding the lowest representative mean χ^2 ($(4.64 \pm 4.26) \times 10^5$) and a competitive KGE of 0.08 ± 0.68. While it successfully learned general growth trends (e.g., Patient-067 MASE: 1.015), it demonstrated higher variance in path-wise accuracy compared to the Milstein solver.
4. **EM Baseline**: The first-order EM method consistently underperformed relative to higher-order schemes, yielding a mean MASE of 1.06 ± 0.28. It exhibited the highest sensitivity to stochastic noise, resulting in the largest variance in predictive trajectories and a representative χ^2 of $(2.83 \pm 3.83) \times 10^6$.

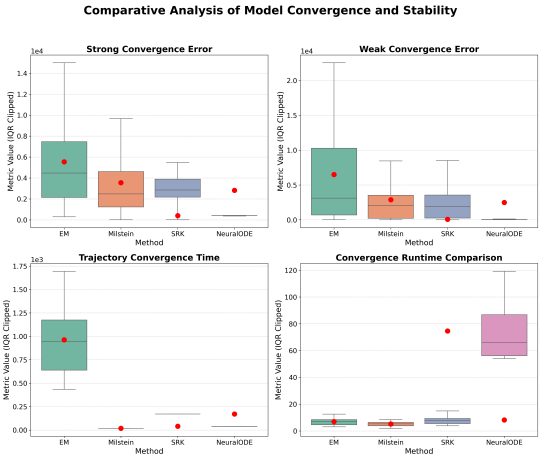

Figure 3. Comparison of Strong/Weak Error and TCT. Red dots indicate the clipped mean, while median values are indicated by the central box lines.

4.2 Convergence and Stability Results

Numerical stability was assessed using the Trajectory Convergence Time (TCT) and discretization error scaling. The global performance of each method across the full test set is detailed in Table 7, in Appendix 7.4, while representative behaviour—following the removal of extreme statistical outliers via IQR clipping—is presented in Table 4. The visual performance of each method on stochastic convergence is further illustrated in Figure 3.

Table 4. Clipped Convergence Performance Summary (Outliers Removed via IQR)

Method	Strong Error	Weak Error	TCT (wks)	Runtime (s)
EM	$5.54 \times 10^3 \pm 4.79 \times 10^3$	$6.49 \times 10^3 \pm 7.61 \times 10^3$	962.50 ± 395.52	7.01 ± 2.91
Milstein	$3.55 \times 10^3 \pm 3.55 \times 10^3$	$2.88 \times 10^3 \pm 3.23 \times 10^3$	$17.70 \pm$ N/A	5.16 ± 1.98
NeuralODE	404.07 ± 15.71	66.11 ± 27.05	39.31 ± 0.58	74.64 ± 23.70
SRK	$2.82 \times 10^3 \pm 1.73 \times 10^3$	$2.49 \times 10^3 \pm 2.74 \times 10^3$	$170.90 \pm$ N/A	8.15 ± 3.48

In Table 4, certain Trajectory Convergence Time (TCT) entries—specifically for the **Milstein** and **SRK** methods—report a standard deviation of "N/A". This statistical artifact occurs when the numerical solvers exhibit near-identical convergence behaviour across the valid patient cohort, resulting in a single unique TCT value that precludes the calculation of variance. This lack of variance indicates that the numerical TCT is tied to the deterministic nature of the carrying capacity once parameters are calibrated, rather than reflecting a flexible adaptation to the data, therefore, values with this error (no matter how low) will not be considered to be well-performing.

Specifically, the **Milstein** method consistently converged at 17.70 weeks for all stable trajectories, demonstrating a high degree of structural rigidity in its long-term growth limit. In contrast, the **Neural ODE** reported a standard deviation of 0.58 weeks, indicating that while it stabilizes significantly faster than the baseline integrators, its convergence is more sensitive to the specific state-transition dynamics learned from individual patient histories. The lack of variance in numerical TCT values shows the deterministic nature of the carrying capacity K in the stochastic logistic growth model once parameters are calibrated.

1. **Stability (TCT)**: The **Neural ODE** demonstrated the fastest representative stabilization, with a clipped mean TCT of 39.31 ± 0.58 weeks. In contrast, the **EM** method often failed to reach the stability threshold within the observed clinical window, resulting in a significantly higher mean TCT of 962.50 ± 395.52 weeks.
2. **Strong and Weak Convergence**: The **Neural ODE** framework exhibited the highest degree of numerical consistency, with clipped strong errors (404.07 ± 15.71) and weak errors (66.11 ± 27.05) an order of magnitude lower than those of the stochastic integrators. Within the numerical solvers, the **Milstein** and **SRK** methods demonstrated improved stability over the **EM** baseline, aligning with theoretical expectations for higher-order schemes.

4.3 Computational Efficiency

The computational "wall-clock" time was recorded for each method. Although the **Neural ODE** had a fast inference time, the initial training overhead and MLP backpropagation made it computationally more expensive than the direct numerical solvers for individual patient simulations. Notably, the **Neural ODE** framework proved computationally the most efficient overall for the test cohort, with an average FitTime of 156.78 ± 41.24 seconds, compared to over 900 seconds for the **EM** baseline. This suggests a trade-off where **Neural ODEs** are optimized for rapid inference and typical cases, while explicit higher-order stochastic solvers are superior for managing high volatility and noisy time-series data characteristic of longitudinal MRI-based tumour tracking.

4.4 Overall Comparison of Models

The comparative evaluation reveals distinct performance trade-offs among the four modelling approaches. The **Euler-Maruyama (EM)** method served as a baseline, exhibiting the highest sensitivity to stochastic noise and the poorest stability, with a mean Trajectory Convergence Time (TCT) of 962.50 weeks. While its simplicity is advantageous, its high variance in predictive trajectories (MASE 1.06 ± 0.28) limits its clinical reliability compared to higher-order schemes.

The **Milstein method** consistently achieved the highest representative predictive accuracy, yielding the lowest clipped MASE (1.04 ± 0.29) and the best Nash-Sutcliffe Efficiency (NSE = -0.00 ± 0.32). Its second-order diffusion correction provided superior pathwise tracking of stochastic growth, making it the most reliable predictor for individual patient trajectories. Similarly, the **SRK method** demonstrated comparable robustness; while its MASE was slightly higher (1.26 ± 0.50), it maintained a consistent structural fit on the order of 10^6 for χ^2, suggesting that derivative-free higher-order schemes are effectively resilient against the volatility of MRI-based volume data.

The **Neural ODE** framework exhibited behaviour distinct from the numerical solvers. Although it suffered from catastrophic divergence in certain sparse datasets (e.g., Patient 43, outlier), it achieved the lowest representative structural error once outliers were removed, with a clipped χ^2 of $(4.64 \pm 4.26) \times 10^5$—an order of magnitude better than the numerical methods. Furthermore, it demonstrated the fastest stabilization (TCT = 39.31 ± 0.58 weeks), indicating that while Neural ODEs may struggle with short-term stochastic fluctuations, they are highly effective at capturing global, smooth long-term growth trends.

Computationally, a clear trade-off exists: the **Neural ODE** was significantly more efficient for large-scale inference (156.78 ± 41.24 s) compared to the exhaustive fitting required for the **Milstein** method (777.0 ± 522.0 s). These distinctions are further synthesized in the Pugh Chart (Appendix 7.3), which serves as a qualitative scoring matrix to compare various concepts against a baseline across multiple criteria (Burge). Note that in this case, the baseline is the **Euler-Maruyma** method, as it is the simplest and most basic stochastic differential equation (SDE) solver, providing a fundamental benchmark for both complexity and accuracy (Higham and -Maruyama). By using this tool, the study confirms that while **Milstein** is optimal for high-precision individual forecasting, **Neural ODEs** offer a more stable and computationally efficient alternative for population-level trend analysis.

5 Discussion and Conclusion

The comparative analysis of four modelling frameworks for glioblastoma growth reveals distinct advantages depending on the clinical metric prioritized. The **Milstein method** achieved the highest overall representative accuracy with a clipped MASE of 1.04 ± 0.29. While the **SRK method** yielded a slightly higher MASE (1.26 ± 0.50), its derivative-free structure maintained high-order robustness, effectively minimizing extreme structural divergence as reflected in its consistent 10^6 range χ^2 values. Notably, the **Neural ODE** framework exhibited the fastest stabilization with a clipped Trajectory Convergence Time (TCT) of 39.31 ± 0.58 weeks, significantly outperforming the **Euler-Maruyama** baseline, which often failed to stabilize within the clinical window (TCT: 962.50 ± 395.52). These findings suggest a fundamental trade-off: higher-order SDE solvers are superior for short-term pathwise accuracy, while **Neural ODEs** provide more stable, smooth long-term growth trajectories.

From a practical standpoint, the choice of modelling approach should align with the specific clinical objective. For applications requiring precise short-term predictions, such as monitoring rapid progression during active treatment, the **Milstein method** is recommended due to its superior handling of multiplicative noise. For long-term prognosis and population-level trend analysis, **Neural ODEs** are preferable as they offer rapid inference (156.78 ± 41.24 s) and greater stability once trained. The use of a **Monte Carlo ensemble approach** ($M = 60$) was critical for both frameworks to account for the non-linear dynamics of tumour growth and the inherent measurement noise in MRI data, providing a robust approximation of the state probability density.

Despite these successes, several areas for improvement remain. The current **Neural ODE** architecture, which integrates Markov transition probabilities to determine growth rates, tends to plateau over long temporal horizons. The Neural ODE tends to plateau over long temporal horizons because its growth rates are constrained by the discrete states (SD, PD, etc.) defined in the Markov transition matrix. This suggests that a more continuous state-transition architecture could further improve long-term forecasting. This behaviour, identified as a structural bias during outlier analysis of patients like Patient 43 (see Appendix 7.4), suggests that future models should integrate temporal dynamics more explicitly to avoid horizontal growth trajectories that may not reflect real-world metastasis.

Additionally, the incorporation of patient-specific demographic features could significantly enhance personalization. While the LUMIERE dataset provides age, sex, and IDH mutation status, our current implementation primarily utilized methylation status for state-specific growth rates. For the neural framework, including these covariates in the input layer would allow the network to learn more nuanced trajectories. For numerical models, where direct training on covariates is not feasible, a potential solution is to stratify parameters by patient characteristics. For example, parameters could be tailored to create steeper growth slopes for older patient subgroups to reflect clinically observed differences in progression rates.

In conclusion, a hybrid modelling approach—combining the short-term pathwise accuracy of higher-order stochastic solvers with the global trend stability and computational efficiency of Neural ODEs—offers a promising path forward. By leveraging structured decision-making tools like the Pugh Chart, clinicians can select the optimal mathematical framework to support earlier interventions and more reliable treatment planning, ultimately improving outcomes for glioblastoma patients.

6 References

Blaom, Anthony D., and Samuel Okon. 2025. New tools for comparing classical and neural ODE models for tumor growth. *arXiv preprint arXiv:2502.07964* (February).

Breheny, Patrick. 2024. *Stochastic convergence,* -09-16.

Burge, Stuart. *The systems engineering tool box.* https://www.burgehugheswalsh.co.uk/uploaded/1/documents/pugh-matrix-v1.1.pdf.

Eisenhauer, E. A., P. Therasse, J. Bogaerts, L. H. Schwartz, D. Sargent, R. Ford, J. Dancey, et al. 2009. *New response evaluation criteria in solid tumours: revised recist guideline (version 1.1),* 2, -01. https://doi.org/10.1016/j.ejca.2008.10.026.

Garg, Jatin. 2025. How state space models improve time series prediction and dynamic systems, -09-19. https://www.gocodeo.com/post/how-state-space-models-improve-time-series-prediction-and-dynamic-systems#:~:text=Emerging%20Trends:%20SSMs%20Meet%20Deep%20Learning%20In,SSMs%20and%20expressive%20power%20from%20neural%20networks..

Gerlee, P. 2012. The model muddle: in search of tumour growth laws. *Quantitative Biology* (Fri, 14 Sep). https://doi.org/https://doi.org/10.48550/arXiv.1209.3170.

Giles, Mike. *Numerical methods ii - milstein.*

Giorno, Virginia, and Amelia G. Nobile. 2025. Richards and gompertz stochastic growth models with time-varying carrying capacity. *Computational and Applied Mathematics* 44, no. 8 (January). https://doi.org/https://doi.org/10.1016/j.cnsns.2021.106100. https://link.springer.com/article/10.1007/s40314-025-03351-3.

Glioblastoma Foundation. 2024. *What is the prognosis of glioblastoma?* https://glioblastomafoundation.org/news/what-is-the-prognosis-of-glioblastoma.

Higham, Des, and Euler -Maruyama. *Numerical simulation of stochastic differential equations: lecture 2, part 1.* Euler Maruyama Method.

Higham, Desmond J. 2001. *An algorithmic introduction to numerical simulation of stochastic differential equations *,* -08-01.

Al-Huniti, Nidal, Yan Feng, Jingyu (Jerry) Yu, Zheng Lu, Mario Nagase, Diansong Zhou, and Jennifer Sheng. 2020. Tumor growth dynamic modeling in oncology drug development and regulatory approval: past, present, and future opportunities. *CPT: Pharmacometrics & Systems Pharmacology* 9, no. 8 (August): 419–427. https://doi.org/10.1002/psp4.12542. https://cir.nii.ac.jp/crid/1874242817949989888.

Hyndman, Rob J., and Anne B. Koehler. 2006. Another look at measures of forecast accuracy. *International Journal of Forecasting* 22, no. 4 (October): 679–688. https://doi.org/https://doi.org/10.1016/j.ijforecast.2006.03.001.

Runge-kutta methods and butcher tableau. n.d. https://www.johndcook.com/blog/2020/02/13/runge-kutta-methods/.

Lalley, Steven P. 2016. *Stochastic differential equations,* -12-02.

Lin, Peng, Jin-Shu Pang, Ya-Dan Lin, Qiong Qin, Jia-Yi Lv, Gui-Qian Zhou, Tian-Ming Tan, Wei-Jia Mo, and Gang Chen. 2025. Tumour surface regularity predicts survival and benefit from gross total resection in IDH-wildtype glioblastoma patients. *Insights into Imaging* 16 (-2-17): 42. https://doi.org/10.1186/s13244-025-01900-2. https://pmc.ncbi.nlm.nih.gov/articles/PMC11833034/.

Ma, Hao, Zhiqiang Pu, Shijie Wang, Boyin Liu, Huimu Wang, Yanyan Liang, and Jianqiang Yi. 2025. Stochastic trajectory prediction under unstructured constraints (18 Mar).

Ma, Ziwei, Ben Niu, Tuan Anh Phan, Anne Line Stensjøen, Chibawanye Ene, Timothy Woodiwiss, Tonghui Wang, Philip K. Maini, Eric C. Holland, and Jianjun Paul Tian. 2020. Stochastic growth pattern of untreated human glioblastomas predicts the survival time for patients. *Scientific reports* 10, no. 1 (April): 6642. https://doi.org/10.1038/s41598-020-63394-w. https://link.springer.com/article/10.1038/s41598-020-63394-w.

MayoClinic. n.d. Glioblastoma: Symptoms and causes. *MayoClinic,* no. Oct 2.

Murphy, Hope, Hana Jaafari, and Hana M. Dobrovolny. 2016. Differences in predictions of ODE models of tumor growth: a cautionary example. *BMC Cancer* 16, no. 1 (February): 163. https://doi.org/10.1186/s12885-016-2164-x. https://link.springer.com/article/10.1186/s12885-016-2164-x.

Nash, J. E., and J. V. Sutville. 1970. River flow forecasting through conceptual models part i — a discussion of principles. *Journal of Hydrology* 10, no. 3 (April). https://doi.org/10.1016/0022-1694(70)90255-6.

Let's talk about mgmt. n.d. https://braintumor.org/news/lets-talk-about-mgmt/#:~:text=Methylated%20MGMT%20promoter:%20Gene%20is,t%20have%20as%20much%20benefit.%E2%80%9D.

Pérez-Beteta, Julián, David Molina-García, Alicia Martínez-González, Araceli Henares-Molina, Mariano Amo-Salas, Belén Luque, Elena Arregui, et al. 2019. Morphological MRI-based features provide pretreatment survival prediction in glioblastoma. *European radiology* 29, no. 4 (April): 1968–1977. https://doi.org/10.1007/s00330-018-5758-7. https://link.springer.com/article/10.1007/s00330-018-5758-7.

Rausand, Marvin. *Introduction markov process transition rates kolmogorov equations chapter 8 markov processes.*

Ruiz-Arrebola, Samuel, Damián Guirado, Mercedes Villalobos, and Antonio M. Lallena. 2021. Evaluation of classical mathematical models of tumor growth using an on-lattice agent-based monte carlo model. *Applied sciences* 11, no. 11 (June): 5241. https://doi.org/10.3390/app11115241. https://www.proquest.com/docview/2635405771.

Schoups, Gerrit, and Jasper A. Vrugt. 2010. A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-gaussian errors. *Water resources research* 46, no. 10 (October): n/a. https://doi.org/10.1029/2009WR008933. https://api.istex.fr/ark:/67375/WNG-HDNNVW1L-B/fulltext.pdf.

Schwartz, M. 1961. A biomathematical approach to clinical tumor growth. *European Journal of Cancer* 44, no. 1 (November): 142–150. https://doi.org/10.1016/j.ejca.2007.10.008.

Sharma, Aman. n.d. Outlier removal using iqr method. https://www.kaggle.com/code/aman2626786/outlier-removal-using-iqr-method.

Simeoni, Monica, Paolo Magni, Cristiano Cammia, Giuseppe De Nicalao, Valter Croci, Enrico Pesenti, and Massimiliano Germani. 2004. Predictive pharmacokinetic-pharmacodynamic modeling of tumor growth kinetics in xenograft models after administration of anticancer agents. *Cancer research* (February): 1094–1101. https://doi.org/10.1158/0008-5472.can-03-2524.

Sivakumar, Muthuramalingam, Sudhaman Parthasarathy, and Thiyagarajan Padmapriya. 2024. Trade-off between training and testing ratio in machine learning for medical image processing. *PeerJ. Computer science* 10 (September): e2245. https://doi.org/10.7717/peerj-cs.2245. https://www.ncbi.nlm.nih.gov/pubmed/39314694.

Slavkova, Kalina P., Sahil H. Patel, Zachary Cacini, Anum S. Kazerouni, Andrea L. Gardner, Thomas E. Yankeelov, and David A. Hormuth. 2023. Mathematical modelling of the dynamics of image-informed tumor habitats in a murine model of glioma. *Scientific reports* 13, no. 1 (February): 2916–14. https://doi.org/10.1038/s41598-023-30010-6. https://link.springer.com/article/10.1038/s41598-023-30010-6.

Solin, Arno. 2014. *Lecture 5: stochastic runge-kutta methods,* -11-25.

Suter, Yannick, Urspeter Knecht, Waldo Valenzuela, Michelle Notter, Ekkehard Hewer, Philippe Schucht, Roland Wiest, and Mauricio Reyes. 2022. The LUMIERE dataset: longitudinal glioblastoma MRI with expert RANO evaluation. *Scientific Data* 9, no. 1 (December): 768. https://doi.org/10.1038/s41597-022-01881-7.

Brownian motion. n.d. Wiener process description, *UChicago,* no. Oct 2, https://galton.uchicago.edu/~lalley/Courses/313/BrownianMotionCurrent.pdf.

Vaghi, Cristina, Anne Rodallec, Raphaëlle Fanciullino, Joseph Ciccolini, Jonathan P. Mochel, Michalis Mastri, Clair Poignard, John M. L. Ebos, and Sébastien Benzekry. 2020. Population modeling of tumor growth curves and the reduced gompertz model improve prediction of the age of experimental tumors. *Computational Biology* (February). https://doi.org/https://doi.org/10.1371/journal.pcbi.1007178.

Walsh, Jason, Alice Othmani, Mayank Jain, and Soumyabrata Dev. 2022. Using U-Net network for efficient brain tumor segmentation in MRI images. *Healthcare Analytics* 2 (November): 100098. https://doi.org/10.1016/j.health.2022.100098.

Wasserman, Larry. *Stochastic convergence.* Department of Statistics and Data Science Machine Learning Department Carnegie Mellon University. https://www.stat.cmu.edu/~larry/=stat700/Lecture7.pdf.

Wunsch, Carl. *Monte carlo/ensemble methods.* https://ocw.mit.edu/courses/12-864-inference-from-data-and-models-spring-2005/916736773602d3e5d1d8be3b79c509dc_monte_carlo_5_1.pdf.

Zilliz. n.d. What numerical solvers (like euler–maruyama) are used in continuous-time diffusion models? https://milvus.io/ai-quick-reference/what-numerical-solvers-like-eulermaruyama-are-used-in-continuoustime-diffusion-models.

7 Appendix

The following appendices provide supporting diagnostics and are not required to interpret the main conclusions.

7.1 Glossary

Methylated/Unmethylated: Glioblastoma tumours can either be methylated or unmethylated, where a methylated tumour has a methyl group on the DNA that controls which genes in the DNA get turned on (expressed) or off (silenced) in the promoter region of the DNA (Let's talk about mgmt, n.d.). Tumours are more likely to survive chemotherapy drugs if they have an unmethylated promoter region, prompting different growth trajectories and different treatment procedures.

Markov Transition Matrix: a square matrix used to describe the probabilities of moving from one clinical state to another. In this framework, each element P_{ij} represents the probability that a patient in state i (e.g., Stable Disease) will transition to state j (e.g., Progressive Disease) in the next time step, assuming the future state depends only on the current observation (Rausand).

HybridStateODE: The *HybridStateODE* combines the discrete probabilistic jumps from the Markov chain (see above) with continuous dynamics (Garg 2025). While the Markov matrix handles the categorical shifts in disease status, the Neural ODE governs the smooth, volumetric growth between those transitions. This "hybrid" approach allows the model to capture both the sudden clinical changes (like treatment response) and the underlying biological proliferation. This *HybridStateODE* is a function within the code in the Github Repository. Note that the definition of each state in the matrix is written below. The network used the universal approximation theorem to interpolate continuous growth rates based on these learned discrete states.

The next four definitions below are based on Eisenhauer et al. 2009:

- **OP (Operative)**: The immediate post-surgery state, used as the baseline measurement before assessing treatment response.
- **CR (Complete Response)**: The tumour has disappeared or is nearly undetectable, representing the most favourable outcome.
- **PR (Partial Response)**: The tumour has shrunk significantly but has not completely disappeared, indicating a good response to treatment.
- **SD (Stable Disease)**: The tumour shows little to no change in size (holding steady).
- **PD (Progressive Disease)**: The tumour is growing or spreading, indicating worsening disease and failure of current treatment.

7.2 Markov State Transition Matrices

Table 5. State Transition Probability Matrix for Unmethylated (left) and Methylated (right) Patients

	OP	CR	PR	SD	PD	Death
OP	0.435	0.011	0.000	0.250	0.239	0.065
CR	0.000	0.250	0.000	0.000	0.750	0.000
PR	0.000	0.000	0.167	0.000	0.500	0.333
SD	0.074	0.037	0.000	0.148	0.667	0.074
PD	0.133	0.013	0.067	0.000	0.387	0.400
Death	—	—	—	—	—	—

	OP	CR	PR	SD	PD	Death
OP	0.452	0.027	0.014	0.233	0.192	0.082
CR	0.000	0.688	0.000	0.000	0.250	0.063
PR	0.000	0.143	0.429	0.000	0.286	0.143
SD	0.000	0.000	0.000	0.472	0.444	0.083
PD	0.116	0.023	0.035	0.023	0.581	0.221
Death	—	—	—	—	—	—

Each table above shows how transitions between states are modelled depending on whether the patient is methylated or unmethylated. The leftmost column lists the current state the patient is in, while the top row lists the possible next states. The values inside the table represent the probability of moving from the current state (row) to a given next state (column).

7.3 Pugh Chart

To synthesize the performance of each modelling approach, a Pugh Decision Matrix was constructed (Table 6). The **Euler–Maruyama (EM)** baseline is used as the datum (0). The evaluation confirms that the **Milstein method** is the most balanced and accurate approach for individual patient forecasting, while the **Neural ODE** offers a highly efficient and stable alternative for capturing population-level growth trends.

7.4 Clipped Patients

7.4.1 Identification and Analysis of Statistical Outliers

A detailed investigation into the divergence between raw and clipped performance metrics revealed that three specific patients—43, 54, and 77—were primarily responsible for the extreme variance in the global cohort statistics. These individuals were flagged as outliers through Interquartile Range (IQR) analysis and subsequently clipped to better reflect representative model behaviour.

1. **Patient 43 (Neural ODE Failure)**: This patient represented the most significant divergence in the study, yielding a Neural ODE χ^2 value of 2.69×10^{21} and a MASE of 4.59. This indicates a complete failure of the neural framework to learn the specific progression dynamics for this individual, likely due to a trajectory that was poorly represented in the training cohort.

Table 6. Pugh Decision Matrix for Concept Selection

Criterion	Baseline: EM	Milstein	SRK	Neural ODE
MASE (Accuracy)	0	+	-	-
χ^2 (Typical Fit)	0	+	-	+
χ^2 (Robustness)	0	-	+	-
NSE (Fit Quality)	0	+	-	-
KGE (Bias/Variance)	0	0	-	+
Fit Time (Efficiency)	0	+	+	+
Strong Conv. Error	0	+	+	+
Weak Conv. Error	0	+	+	+
Stability (TCT)	0	+	+	+
Total Score	**0**	**+5**	**+1**	**+3**

2. **Patient 54 (Accuracy Outlier)**: Flagged for high predictive error in the neural set (MASE = 2.45), Patient 54 exhibited structural inconsistencies that challenged both neural and numerical models. While not as extreme as Patient 43, the error was sufficient to significantly inflate the representative MASE mean.
3. **Patient 77 (Numerical Integrator Divergence)**: This patient was the primary cause for the astronomical raw χ^2 means in the numerical solvers, producing a value of 2.03×10^{16} in the Milstein method and 7.79×10^{11} in the SRK scheme. This divergence illustrates the inherent risks of multiplicative noise in stochastic solvers when calibration parameters (ρ, σ) fail to bound long-term growth for specific clinical patterns.

By removing these extreme cases, the clipped performance metrics (Table 3) provide a more accurate representation of how these models would perform for a typical glioblastoma patient within the LUMIERE cohort.

7.4.2 Raw Convergence Data

Below is the raw data for the convergence of the four methods (Table 3), before the outliers in Appendix 7.4 were clipped.

Table 7. Raw Convergence Performance Summary (Mean ± Std)

Method	Strong Error	Weak Error	TCT (wks)	Runtime (s)
EM	$5.54 \times 10^3 \pm 4.79 \times 10^3$	$6.49 \times 10^3 \pm 7.61 \times 10^3$	962.5 ± 395.5	7.01 ± 2.91
Milstein	$4.11 \times 10^3 \pm 4.78 \times 10^3$	$4.54 \times 10^3 \pm 6.96 \times 10^3$	$17.7\pm$ N/A	5.16 ± 1.98
NeuralODE	404.07 ± 15.71	66.11 ± 27.05	40.13 ± 2.47	74.64 ± 23.70
SRK	$2.82 \times 10^3 \pm 1.73 \times 10^3$	$2.55 \times 10^3 \pm 2.88 \times 10^3$	$170.9\pm$ N/A	8.15 ± 3.48

7.5 More Sample Patient Outputs

Below is another plot of SRK for patient 31 (Figure 4), where the predicted volume is more akin to the patient's actual growth, compared to the one discussed in the text, where there was blowup.

7.6 Batches

Below, in Figure 5, are the batch plots for the ten testing patients (including the clipped outliers) for the three numerical methods:

7.7 Iterations to Main Formula

This section documents the empirical evolution of our model design choices. Note that these modifications are phenomenological and motivated by preserving qualitative biological realism rather than deriving from first-principles cellular dynamics.

MODIFICATION 1: Noticed that just the original equation, Gompertz, wasn't holding up – tumour would shrink then grow back Added a degrowth term, then regrowth, then fitted it to sigmoidal through trial and error:

$$dV = (1 - s(t))a_1 V log(b_1/V) + s(t)a_2 V log(b_2/V) - k(t) * V)dt + (cV/(h + \sqrt{V}))dW \tag{15}$$

Sigmoidal $y = s(x) = \frac{1}{1+e^{-x}}$

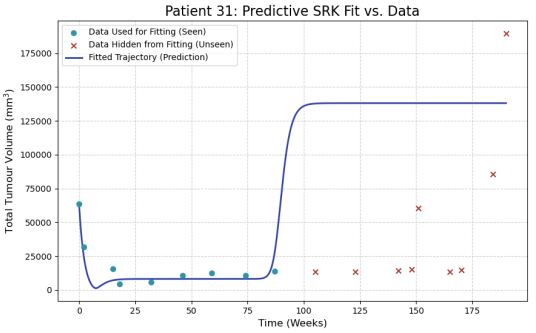

Figure 4. SRK output for Patient 31.

MODIFICATION 2: high runtime, but more accurate than before.

Modified to:

$$dV = \left(\sum_{i=1}^{n(t)} a_i V \log(b_i/V) - k(t)V \right) dt + \left(\frac{cV}{h + \sqrt{V}} \right) dW \tag{16}$$

where the number of terms $n(t)$ changes dynamically: Whenever the drift changes sign ($dV/dt = 0$), a new term is added with the opposite growth coefficient. Here, a_{new} is the negative of a_{last}, so the model self-adapts: growth → shrinkage → growth, etc.

MODIFICATION 3: "Classical ODE tumour growth models (e.g., logistic or Gompertz) treat the growth rate and carrying capacity as static parameters (Vaghi et al. 2020, Gerlee 2012). In reality, treatment, microenvironment changes and therapy windows cause both the effective growth-rate and carrying-capacity to vary in time. Accordingly, we model

$$dV = (a(t)Vln(b(t)/V) - k(t)V)dt + cV/(h + \sqrt{V})dW \tag{17}$$

for V tumour volume, $k(t)$ chemo-induced decay rate, defined piecewise in Equation 18, c and h stochastic parameters, W the Wiener process, and $a(t)$, $b(t)$ defined as: $a(t) = a_1 + a_2 tanh(a(t - t_c))$, $b(t) = b_1 + b_2((t - t_b))$, $\sigma = \frac{1}{1+e^{-x}}$ where the hyperbolic-tangent and logistic-sigmoid transitions capture gradual shifts in growth-rate and capacity arising from e.g. chemotherapy onset or vascular adaptation.

$$k(t) = \begin{cases} k_0 & t \in [t_{\text{chemo, start}}, t_{\text{chemo, end}}] \\ 0 & \text{otherwise} \end{cases} \tag{18}$$

While the exact functional form is our choice (motivated by smooth transitions), the underlying concept of time-varying parameters is consistent with the literature.

- Use of a time-dependent carrying capacity: Giorno and Nobile 2025
- Gompertz model no longer has a constant: Al-Huniti et al. 2020

The discrete version in the code: For more information on the iterations, go to this document.

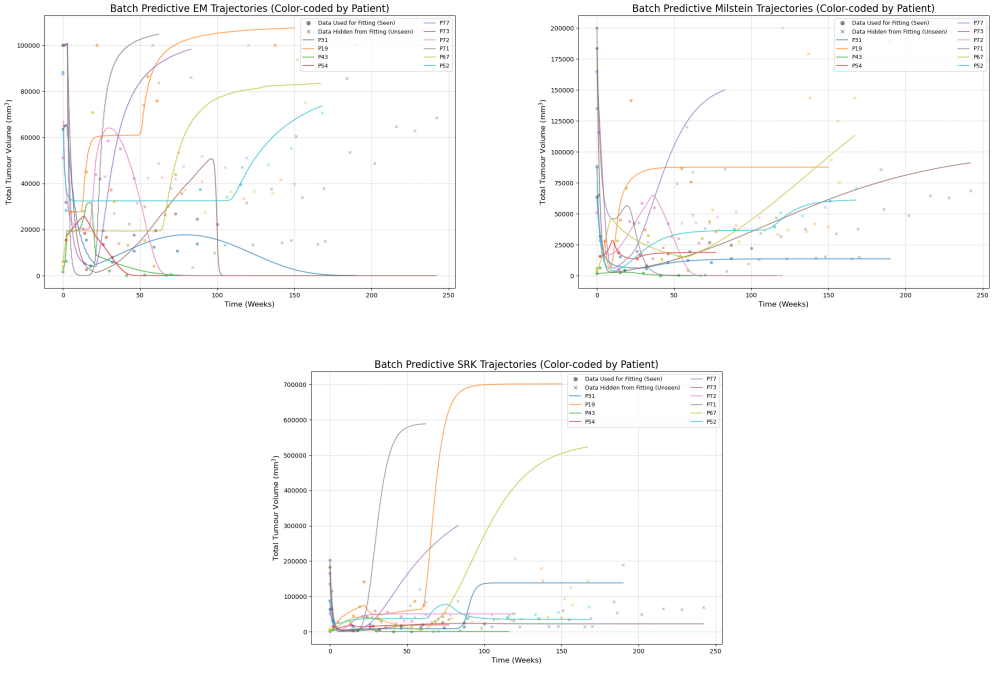

Figure 5. Batch output for Testing set for EM, Milstein, SRK respectively.