Estimating Growth Rate from Sparse and Noisy Data: A Bayesian Approach

David Andersson * **Anton Vernersson** * **Anne Richelle** **

* *Sartorius Corporate Research, Umea, Sweden (e-mails: david.andersson@Sartorius.com, anton.vernersson@Sartorius.com).*
** *Sartorius Corporate Research, Brussels, Belgium (e-mail: anne.richelle@sartorius.com)*

Abstract: Accurate estimation of growth rates from sparse and noisy concentration data is a significant challenge in bioprocess modeling. This paper presents a method that utilizes a modular non-linear interpolation framework combined with Bayesian parameter inference to address this issue. By incorporating prior knowledge of cell culture dynamics with differentiable basis functions, our approach generates credibility intervals for growth rates, enhancing the reliability of predictions. We validated the performance of our method using growth rate data generated *in silico* and demonstrated its application on two *in vitro* datasets, showcasing its robustness across various measurement conditions and practical applicability. Results indicate improvements in the reliability and credibility of predictions compared to traditional methods, making this framework a valuable resource for accurate growth rate estimations.

Keywords: Bayesian methods, Nested Sampling, Metabolic Rates, Bioprocess, Modeling and Identification

1. INTRODUCTION

Accurate estimation of growth rates is a critical aspect of bioprocess modeling, yet it remains a significant challenge due to the sparse and noisy nature of concentration data. In bioprocessing, data is collected through discrete measurements from analytical devices, which introduce noise. Cost constraints limit sampling frequency, leading to sparse time series with few repeated measurements. The stochastic nature of bioprocessing further complicates reliable growth rate estimation (Sonnleitner et al., 2000; Zahel et al., 2016; Bayer et al., 2020).

Traditional approaches, such as stepwise integral estimations, assume constant growth rates between sampling intervals (Wechselberger et al., 2013). While these approaches are straightforward, they oversimplify biological processes and may introduce inaccuracies, especially in non-linear fermentation systems where cell growth rates exhibit variability over time. Moving average filters have been employed to smooth such fluctuations, but they reduce the ability to capture dynamic transitions and can propagate errors (Paulsson et al., 2014).

To address these challenges, Bayer et al. (2020) proposed using cubic smoothing splines for continuous growth rate estimates. This method fits a spline across noisy data points, balancing smoothness and precision, and provides reliable results even with varying sampling frequencies and measurement noise. However, it requires careful parameter tuning to prevent overfitting or oversmoothing (Swain et al., 2016). Given the limitations of existing methods, more robust approaches are needed to capture dynamic changes in cell growth rates by accounting for non-linear behavior and addressing data sparsity and noise.

In this work, we propose a Bayesian inference framework to model cell growth using statistical models based on non-linear basis functions. By applying inference through Nested Sampling, we generate posterior distributions of model parameters and facilitate model comparison by calculating the Bayesian evidence, a reliable metric for model selection. To validate the effectiveness of our method, we first apply it to *in silico* data and subsequently demonstrate its performance on *in vitro* datasets.

2. METHODS

Our approach utilizes an interpolation model with differentiable basis functions tailored to discrete measurements of viable cell density (VCD). By fitting the interpolation directly to VCD data and differentiating the resulting function, we estimate cell growth rates. We explore various models by adjusting the number and type of basis functions, and employ Bayesian evidence to rank these models.

2.1 Prior-Informed Interpolation Model

To incorporate prior assumptions about the growth rate, we define interpolation models that translate our understanding of the growth rate behaviour into mathematical constraints.

We aim to interpolate a set of N_t observations: $\{(t_i, y_i)\}_{i=1}^{N_t}$, where t denotes time and y represents the observations. We define $\mathcal{I}(t)$ as a sum of basis functions $\{\phi_j(t)\}_{j=1}^{N_f}$ to approximate y:

$$\mathcal{I}(t_i) = \sum_{j=1}^{N_f} \alpha_j \phi_j(t_i) \approx y_i, \tag{1}$$

where α_j are undetermined coefficients $j = 1, 2, \ldots, N_f$ for N_f number of functions.

We evaluated several basis functions, including the Gaussian function $e^{-\rho(t-\kappa)^2}$, the hyperbolic tangent function $\tanh(\rho(t-\kappa))$, and the sigmoid function $\frac{2}{1+e^{-\rho(t-\kappa)}}$, where ρ is introduced to modulate the function slopes, while κ denotes the center, allowing the basis function to be shifted by the direct distance $t-\kappa$. We chose these functions based on the assumption that cell population growth typically converges to plateau-like levels, representing growth states. Although other functions could also represent this, we selected a subset to showcase our framework's capabilities.

To represent changes in VCD, we used indefinite integrals of the basis functions for interpolation, enabling us to differentiate and achieve the desired form. The indefinite integral of the Gaussian function does not have a simple closed-form expression in terms of elementary functions. However, it can be expressed in terms of the error function, $\mathrm{erf}(t)$, as:

$$\int e^{-\rho(t-\kappa)^2}\,dt = \sqrt{\frac{\pi}{4\rho}}\mathrm{erf}\Big[\sqrt{\rho}(t-\kappa)\Big] + C, \tag{2}$$

where C is an arbitrary constant of integration. Note that the error function needs to be calculated using a separate integral:

$$\mathrm{erf}(z) = \frac{2}{\sqrt{\pi}}\int_0^z e^{-x^2}\,dx, \tag{3}$$

The indefinite integrals of the hyperbolic tangent and the sigmoid functions are respectively:

$$\int \tanh\big[\rho(t-\kappa)\big]\,dt = \frac{1}{\rho}\ln\Big[\cosh\big(\rho(t-\kappa)\big)\Big] + C, \tag{4}$$

and

$$\int \frac{1}{1+e^{-\rho(t-\kappa)}}\,dt = \frac{1}{\rho}\ln\big(1+e^{-\rho(t-\kappa)}\big) + C. \tag{5}$$

Finally, the interpolation function of VCD is defined as:

$$\tilde{\mathcal{X}}(t_i) = \sum_{j=1}^{N_f}\alpha_j\phi_j(t_i), \tag{6}$$

where ϕ corresponds to one of the following functions:

$$\phi_j(t) = \begin{cases} \sqrt{\frac{\pi}{4\rho_j}}\mathrm{erf}\Big[\sqrt{\rho_j}(t-\kappa_j)\Big] + C \text{ for Gaussian,} \\ \frac{1}{\rho_j}\ln\Big[\cosh\big(\rho_j(t-\kappa_j)\big)\Big] + C \text{ hyperbolic tangent,} \\ \frac{1}{\rho_j}\ln\big(1+e^{-\rho_j(t-\kappa_j)}\big) + C \text{ for sigmoid.} \end{cases} \tag{7}$$

For comparison purposes, we included the well-adapted logistic model from Jolicoeur and Pontier (1989) with an additional constant term:

$$\hat{\mathcal{X}}(t) = \frac{a}{e^{bt} - ce^{-dt}} + C, \tag{8}$$

with derivative

$$\frac{d\hat{\mathcal{X}}(t)}{dt} = -\frac{ae^{dt}(be^{bt(b+d)} + cd)}{(c - e^{t(b+d)})^2}. \tag{9}$$

2.2 Statistical framework

To approximate the parameter distributions of the VCD interpolation model, we developed a statistical framework that employs Bayesian inference using the Nested Sampling method.

Bayesian inference. Bayes theorem states that the posterior probability of event $A|B$ is equal to the product of the likelihood $P(B|A)$ and prior probability $P(A)$, divided by the marginal likelihood $P(B)$:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}. \tag{10}$$

We seek to generate the posterior probability for a model $P(\theta|D)$, given some parameters θ and some data D. Using equation (10), the posterior can be stated as:

$$P(\theta|D) = \frac{\mathcal{L}(D|\theta)\pi(\theta)}{\mathcal{Z}(D)}, \tag{11}$$

where $\mathcal{L}$ represents the likelihood, π the prior and $\mathcal{Z}$ the marginal likelihood or *evidence*. The evidence can be calculated by integrating the entire parameter space Ω_θ as:

$$\mathcal{Z} = \int_{\Omega_\theta}\mathcal{L}(D|\theta)\pi(\theta)\,d\theta. \tag{12}$$

The posterior predictive distribution for a new observation at an arbitrary timepoint $\tilde{y}(t)$ given the data is defined as:

$$P(\tilde{y}(t)|D) = \int\mathcal{L}(\tilde{y}(t)|\theta)P(\theta|D)\,d\theta, \tag{13}$$

where $\mathcal{L}(\tilde{y}(t)|\theta)$ is the likelihood of the new observation given posterior parameters, in this case generated by Nested Sampling (Sivia and Skilling, 2006).

Nested Sampling. Nested sampling is a computational method used to estimate Bayesian evidence and generate posterior parameter distributions (Skilling, 2004). The algorithm iteratively replaces the least likely sample points with new ones of higher likelihoods, while maintaining a likelihood constraint. This process allows for the exploration of high-dimensional parameter spaces and the precise estimation of Bayesian evidence.

To marginalize the integral in equation (12), we used the dynamic Nested Sampling method proposed by Speagle (2020) with the implementation by Sergey Koposov et al. (2024). This requires defining the models' log-likelihood and transforming the prior from unit norm for each parameter. In our data, VCD measurements exhibit increasing variance with higher values, indicating a proportional relationship between the variance and observations. Hence, we used a log-normal distribution to account for the specific nature of the observations:

$$y(t) \sim \ln\mathcal{N}(\ln(y(t)), \sigma). \tag{14}$$

We define the log-likelihood as follows:

$$\mathcal{L}(y) = \begin{cases} -\frac{1}{2}\sum_{i=1}^{N_t}\Bigg[2\ln(\theta_{N_\sigma}) + \ln(2\pi) + \ln\big(\mathcal{X}(t_i,\theta)\big) \\ \qquad + \Bigg(\frac{\ln(\mathcal{X}(t_i,\theta)) - \ln(y_i)}{\theta_{N_\theta}}\Bigg)^2\Bigg] \text{if } \mathcal{X}(t_i,\theta) > 0, \\ -\infty \text{ otherwise.} \end{cases} \tag{15}$$

Model parameters are denoted by θ for $n = 1, 2, \ldots, N_\theta$, with N_θ being the total number of parameters. The measurement noise σ is approximated by θ_{N_θ}. Each parameter uses uniform bounded priors, $\theta \sim \mathrm{Unif}[\theta^l, \theta^u]$, which are

transformed from uniformly distributed variables $u_n \sim \text{Unif}[0,1]$ to the desired ranges $[\theta_n^l, \theta_n^u]$ as follows:

$$\theta_n = \theta_n^l + u_n(\theta_n^u - \theta_n^l). \tag{16}$$

If the measurement noise interval is known, prior bounds for σ are set as $\theta_{N_\sigma} \sim [\theta_{N_\sigma}^l, \theta_{N_\sigma}^u]$. If the noise is unknown, we have applied a fallback heuristic by solving the inverse of equation (8) using a quasi-Newton gradient optimization scheme (Fletcher, 1987). Since we are assuming log normally distributed noise on our measurements, we let the standard deviation of the log residuals between the observations and the Logistic model define the upper bound.

Error Propagation. Nested sampling accounts for error propagation by incorporating parameter uncertainties and measurement errors in the likelihood function. The resulting posterior distributions inherently reflect these uncertainties. When generating posterior predictive distributions, these errors are propagated through subsequent analyses, thereby influencing predictions and further inferences (Sivia and Skilling, 2006).

We used Nested Sampling to estimate the model's parameter distributions. By applying the log-likelihood (eq. (15)) and the prior transform (eq. (16)), we created instances of the models defined by equations (6) and (8). This approach allowed us to obtain N_s sampled parameter vectors for each model, effectively capturing the parameter distributions. For the model defined by equation (6), the vector of parameters $\tilde{\theta}$ is given by:

$$\tilde{\theta} = \left(\{\alpha_j\}_{j=1}^{N_p}, \{\rho_j\}_{j=1}^{N_p}, \{\kappa_j\}_{j=1}^{N_p}, \theta_{\mathcal{C}}, \sigma\right) \in \mathbb{R}^{N_s \times (3N_p+2)}. \tag{17}$$

For the model described by equation (8), the collected parameters $\hat{\theta}$ are given by:

$$\hat{\theta} = (\theta_a, \theta_b, \theta_c, \theta_d, \theta_{\mathcal{C}}, \sigma) \in \mathbb{R}^{N_s \times 6}, \tag{18}$$

where $\theta_{\mathcal{C}}$ represents the constant parameter C in both models. The posterior predictive trajectories of the VCD can then be produced by evaluating $\tilde{\mathcal{X}}(t_i, \tilde{\theta})$ or $\hat{\mathcal{X}}(t_i, \hat{\theta})$.

2.3 Standard rate approximation methods

For comparison, we have included two common rate approximation methods. The first method is a logarithmic constant rate approximation, defined as follows:

$$f(t,y) = \begin{cases} \ln(y_i) + \ln\left(\dfrac{y_i}{y_{i+1}}\right)\dfrac{(t-t_i)}{(t_{i+1}-t_i)} & \text{for } t \in [t_i, t_{i+1}], \\ 0 \qquad \text{otherwise.} \end{cases} \tag{19}$$

The second method is a cubic spline model (Dierckx, 1993; Bayer et al., 2020), a piecewise polynomial function composed of recursively defined basis splines of order 3 (see Appendix A).

2.4 Dynamic cell growth model

We generated *in silico* data using a cell growth model to test our rate estimation method (Richelle et al., 2022). This model comprises ordinary differential equations (ODEs) that describe the dynamics of cell populations across three phases: live cells, dead cells, and lysed cells, as defined below.

$$\frac{dX_v}{dt} = (\mu_{\text{eff}} - \mu_d - \frac{F}{V})X_v, \tag{20}$$

$$\frac{dX_d}{dt} = \mu_d X_v - (k_l + \frac{F}{V})X_d, \tag{21}$$

$$\frac{dX_l}{dt} = k_l X_d - \frac{F}{V}X_l, \tag{22}$$

where X_v is the VCD, X_d and X_l are the dead and lysed cells concentration. F is the feed flow rate and V is the bioreactor volume. μ_{eff} is the effective growth rate, μ_d is the effective death rate, and k_l is the lysis rate constant. The model also includes a "biomaterial" variable, $\emptyset_b$, representing a bulk set of metabolic byproducts secreted by the cells:

$$\frac{d\emptyset_b}{dt} = k_{\emptyset_b} X_v - \frac{F}{V}\emptyset_b, \tag{23}$$

where $k_{\emptyset_b}$ is the biomaterial secretion rate constant. To ensure identifiability, we set this parameter to 1. The cell growth rate, μ_{eff}, is represented as:

$$\mu_{\text{eff}} = \mu_{\max} \frac{1}{(\frac{\emptyset_b}{K_{I,\emptyset_b}})^3 + 1}, \tag{24}$$

where $\mu_{\max}$ is the maximum growth rate and $K_{I,\emptyset_b}$ is the inhibition constant for the biomaterial. The effective death rate, μ_d, is modeled as: $\mu_d = k_d + k_t X_l$, where k_d is the base death rate and k_t is the toxicity rate constant. The parameter values used to simulate the growth model were taken from Richelle et al. (2022).

3. RESULTS

In the next sections, we present the results of fitting VCD data using the models from Section 2. We label the models in equations (7) as *Gaussian*, *Tanh*, and *Sigmoid*, the model in equation (19) as *Constant*, and the one in Appendix A as *Spline*.

When fitting a model, we used Nested Sampling to estimate the Bayesian evidence Z, which quantifies the model's fit relative to its complexity. For each model type, we tested several configurations with varying numbers of basis functions, selecting the one with the superior $\log Z$ score. This approach ranks models by penalizing excessive complexity, ensuring a balance between fit and simplicity. Table 1 presents the final number of parameters used and the corresponding $\log Z$ scores for each model type.

3.1 In Silico Data Application

We simulated the VCD evolution using the cell growth model (Section 2.4) and sampled observation points to generate data for analysis. The simulation was integrated using an explicit fifth-order Runge-Kutta method with fourth-order error correction. Absolute and relative tolerances were both set to 1e-6. To account for measurement noise, we added log-normally distributed noise to the sampled data points, defining each noisy observation as follows:

$$\tilde{y}_i = y_i e^{\epsilon_i}, \tag{25}$$

with $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$ and σ representing the noise level (set to 0.1), for $i = 1, 2, \ldots, N_\eta$, with N_η denoting the number of sampled points. Finally, we encapsulated these noisy

Table 1. Model Performance Comparison

Dataset	Model	Parameters	log Z	SSE	Time [s]
In Silico	*Logistic*	6	-25.93	6.68	282
	Gaussian	$4{\cdot}3+2$	-23.23	9.06	22
	Tanh	$8{\cdot}3+2$	-31.04	379.17	88
	Sigmoid	$5{\cdot}3+2$	-22.79	3.53	7
	Spline	-	-	1116.85	-
	Constant	-	-	34.24	-
Low	*Logistic*	6	-18.37	-	592
	Gaussian	$3{\cdot}3+2$	-17.32	-	20
	Tanh	$5{\cdot}3+2$	-32.69	-	23
	Sigmoid	$4{\cdot}3+2$	-14.92	-	18
High	*Logistic*	6	-21.56	-	331
	Gaussian	$3{\cdot}3+2$	-14.83	-	31
	Tanh	$7{\cdot}3+2$	-24.85	-	98
	Sigmoid	$3{\cdot}3+2$	-2.04	-	21

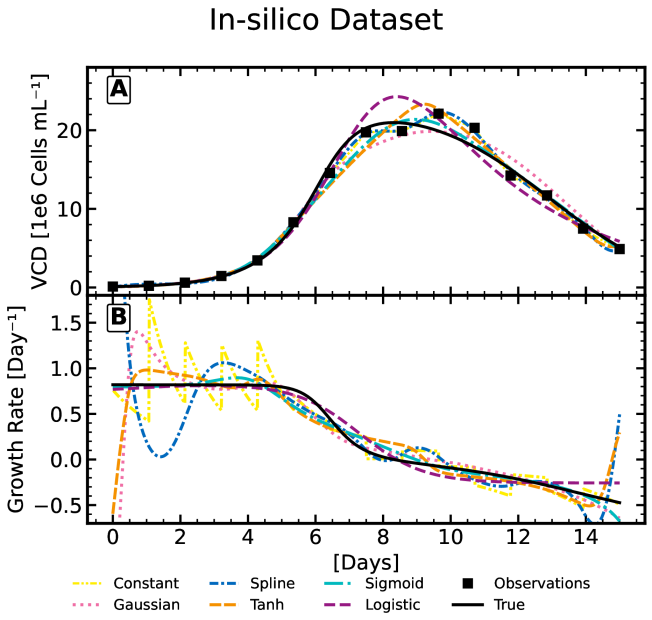

Fig. 1. VCD estimates (**A**), growth rate estimates (**B**) and *in silico* observations sampled from cell growth model.

data points in a discrete observation set $\{(t_i, \tilde{y}_i)\}_{i=1}^{N_\eta=15}$ as seen in Figure 1A.

VCD is recovered from the observed growth rate by integrating equation (20). To address the non-identifiability between the effective growth rate and the cell death rate, we introduced the parameter μ to represent the total growth rates recovered in our posterior predictions. For clarity, equation (20) is thus rewritten as $\frac{dX_v}{dt} = \mu X_v$, where $\mu = \mu_{\text{eff}} - \mu_d$.

Figure 1 presents the estimated VCD and growth rates for all models. Table 1 reports the calculation time and sum of squared errors (SSE) for the growth rates. Figure 1**B** shows that the *Tanh*, *Sigmoid*, *Gaussian*, and *Logistic* models provide similar growth rate estimates, calculated using the mean of their posterior predictive distributions. The *Sigmoid* model effectively captures growth rate dynamics with minimal oscillation. In contrast, the *Tanh* and *Gaussian* models show oscillations in the initial days, conflicting with our assumptions about cell growth. The *Logistic* model initially overestimates and subsequently

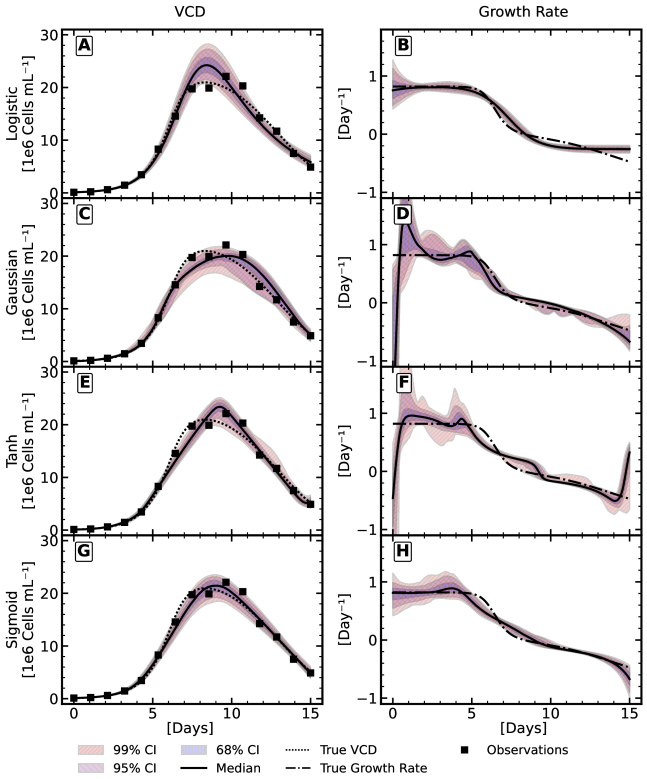

Fig. 2. Posterior predictive distributions of VCD and growth rates derived from *in silico* data generated with cell growth model.

underestimates the growth rate in the latter half of the period. The *Spline* method shows significant oscillatory behavior, while the *Constant* rate model consistently misestimates growth due to the sparsity of measurements.

In Figure 2, the posterior predictive distributions for growth rate and VCD are shown for *Logistic*, *Gaussian*, *Tanh* and *Sigmoid* models. The *Sigmoid* model shows the best overall fit to the true growth rate (Figure 2**H**). All interpolation models place the true rate trajectory in close proximity to credibility bounds except for the *Gaussian* model which overestimates the growth rate between days 1-2. The *Logistic* model, while approximating the VCD trajectory well, misestimates the growth rate indicating it is not adapted to the cell model (Figure 2**B**).

3.2 In Vitro Data Application

We utilized two *in vitro* experimental datasets from fed-batch CHO cell cultures, both generated using Sartorius' proprietary CHO cell line under the same operating conditions. The High Frequency Experiment dataset includes 31 VCD measurements over 12 days, using Sartorius proprietary media. The Low Frequency Experiment dataset consists of 12 VCD measurements over 15 days, with a slightly altered media formulation. We chose these datasets to test our approach's ability to capture growth evolution under different sampling frequencies and media conditions, despite using the same cell line and consistent operating conditions.

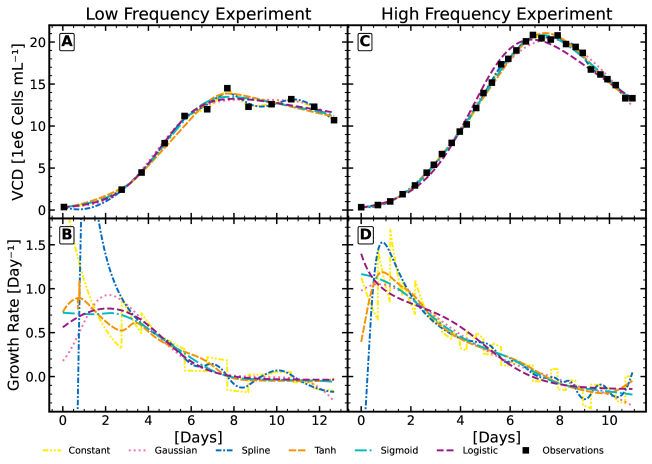

Fig. 3. VCD estimates (**A** and **C**) and growth rate estimates (**B** and **D**), derived from the Low and High Frequency Experiments, respectively.

In the High Frequency Experiment (Figure 3**C**), VCD estimates show that most models generate similar responses, except for the *Logistic* model, which is noticeably shifted. Regarding growth rates (Figure 3**D**), the models generally agree, except for the *Spline* model, which exhibits oscillation, and the *Constant* model, which displays periodic transitions. Figure 4 demonstrates that while all models provide satisfactory VCD estimates, the *Sigmoid* model most closely aligns with our assumptions for growth rates, particularly in terms of convergence when data frequency is high.

The Low Frequency Experiment challenges the models with sparse data as seen in Figure 3**A**. Consequently the *Spline* model's oscillations are more pronounced when inspecting the growth rates in Figure 3**B**. In Figure 5, the *Gaussian* model displays large initial swings, while the *Logistic* and *Tanh* models maintain confidence bounds that anticipate a small error, possibly being overly confident. The *Sigmoid* model provides a similar growth rate approximation to the *Logistic* model but with a more conservative confidence interval during the initial period where there is very little information.

Overall, the *Sigmoid* model proves robust for both experiments, supported by its strong evidence value and low SSE. In the High Frequency Experiment, we observe consistent model behavior, while the Low Frequency Experiment highlights the challenges of data sparsity and the need for balanced model confidence.

4. CONCLUSIONS

The proposed method provides a robust way to estimate credibility-bounded growth rates from sparse and noisy data. The *Sigmoid* model aligned well with our assumptions and showed high Bayesian evidence, offering more flexibility due to its ability to accommodate varying growth dynamics compared to the *Logistic* model. The estimates generated by the *Tanh* model tended to diverge significantly in areas of sparse data, highlighting its limitations in such conditions. The *Gaussian* model represented a middle ground and may prove useful in scenarios where moderate flexibility is required.

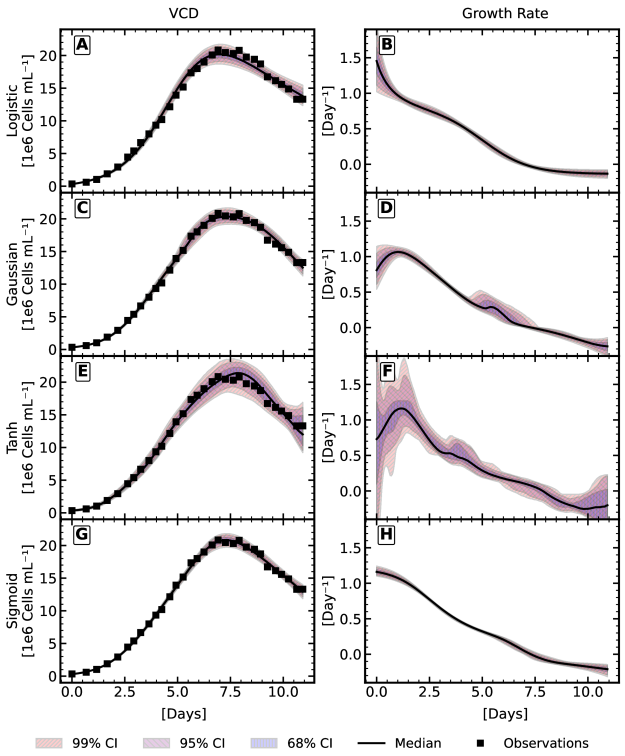

Fig. 4. Posterior predictive distributions of VCD and growth rates derived from the High Frequency Experiment data.

We noted inconsistencies during repeated sampling, especially in areas with large credibility bounds, mainly due to balancing effective regularization from measurement noise. To address this, refining prior distributions with domain-specific knowledge, such as typical growth patterns or environmental conditions, can improve accuracy. Additionally, using models with more interpretable parameters can enhance the reliability of predictions.

While our framework is based on established statistical techniques, its novelty lies in the integration of prior biological knowledge and the use of Nested Sampling for efficient parameter space exploration. This sets it apart from traditional methods that rely on constant growth rate assumptions, providing a more dynamic approach to modeling. Our method can be applied to a wide range of reaction networks but may struggle with extreme non-linear behaviors. The computational complexity of Bayesian inference and Nested Sampling could affect scalability and practical implementation in real-time bioprocess monitoring. Future work could focus on optimizing computational efficiency, additional basis functions, exploring hybrid models, include additional bioprocess covariates, such as volume change, to enhance its applicability.

In conclusion, our Bayesian inference framework offers a promising alternative for estimating growth rates from sparse, noisy data. With potential for further refinement, it can improve accuracy and reliability in bioprocess modeling.

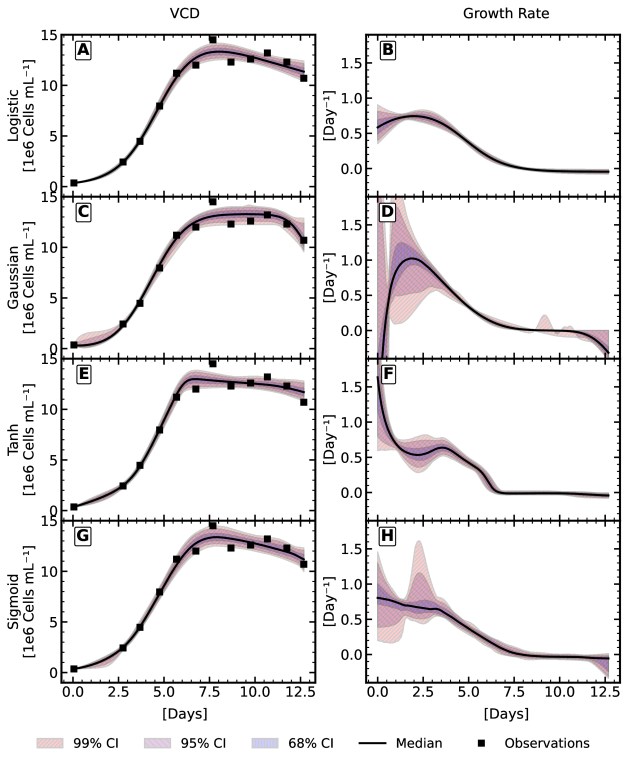

Fig. 5. Posterior predictive distributions of VCD and growth rates derived from the Low Frequency Experiment data.

ACKNOWLEDGEMENTS

We would like to acknowledge Oliver Cloarec for his valuable input on Bayesian statistics in this work.

REFERENCES

Bayer, B., Sissolak, B., Duerkop, M., von Stosch, M., and Striedner, G. (2020). The shortcomings of accurate rate estimations in cultivation processes and a solution for precise and robust process modeling. *Bioprocess and Biosystems Engineering*, 43(2), 169–178. doi: 10.1007/s00449-019-02214-6.

Dierckx, P. (1993). Univariat Splin S. In *Curve and Surface Fitting with Splines*, 3–22. Oxford University PressOxford. doi:10.1093/oso/9780198534419.003.0001.

Fletcher, R. (1987). *Practical methods of optimization*. Wiley, Chichester ; New York, 2nd ed edition.

Jolicoeur, P. and Pontier, J. (1989). Population growth and decline: a four-parameter generalization of the logistic curve. *Journal of Theoretical Biology*, 141(4), 563–571. doi:10.1016/S0022-5193(89)80237-1.

Paulsson, D., Gustavsson, R., and Mandenius, C.F. (2014). A Soft Sensor for Bioprocess Control Based on Sequential Filtering of Metabolic Heat Signals. *Sensors*, 14(10), 17864–17882. doi:10.3390/s141017864.

Richelle, A., Corbett, B., Agarwal, P., Vernersson, A., Trygg, J., and McCready, C. (2022). Model-based intensification of CHO cell cultures: One-step strategy from fed-batch to perfusion. *Frontiers in Bioengineering and Biotechnology*, 10. doi:10.3389/fbioe.2022.948905.

Sergey Koposov, Josh Speagle, Kyle Barbary, Gregory Ashton, Ed Bennett, Johannes Buchner, Carl Scheffler, Ben Cook, Colm Talbot, James Guillochon, Patricio Cubillos, Andrés Asensio Ramos, Matthieu Dartiailh, Ilya, Erik Tollerud, Dustin Lang, Ben Johnson, jtmendel, Edward Higson, Thomas Vandal, Tansu Daylan, Ruth Angus, patelR, Phillip Cargile, Patrick Sheehan, Matt Pitkin, Matthew Kirk, Joel Leja, joezuntz, and Danny Goldstein (2024). joshspeagle/dynesty: v2.1.4. doi: 10.5281/ZENODO.3348367.

Sivia, D. and Skilling, J. (2006). *Data Analysis: A Bayesian Tutorial*. Oxford University Press. Google-Books-ID: lYMSDAAAQBAJ.

Skilling, J. (2004). Nested Sampling. 735, 395–405. doi:10.1063/1.1835238. ADS Bibcode: 2004AIPC..735..395S.

Sonnleitner, B., Scheper, T., Babel, W., Blanch, H.W., Cooney, C.L., Enfors, S.O., Eriksson, K.E.L., Fiechter, A., Klibanov, A.M., Mattiasson, B., Primrose, S.B., Rehm, H.J., Rogers, P.L., Sahm, H., Schügerl, K., Tsao, G.T., Venkat, K., Villadsen, J., Von Stockar, U., and Wandrey, C. (eds.) (2000). *Bioanalysis and Biosensors for Bioprocess Monitoring*, volume 66 of *Advances in Biochemical Engineering/Biotechnology*. Springer Berlin Heidelberg, Berlin, Heidelberg. doi:10.1007/3-540-48773-5.

Speagle, J.S. (2020). DYNESTY: a dynamic nested sampling package for estimating Bayesian posteriors and evidences. *Monthly Notices of the Royal Astronomical Society*, 493, 3132–3158. doi:10.1093/mnras/staa278. ADS Bibcode: 2020MNRAS.493.3132S.

Swain, P., Stevenson, K., Leary, A., Montano-Gutierrez, L.F., Clark, I., Vogel, J., and Pilizota, T. (2016). Inferring time derivatives, including cell growth rates, using Gaussian processes. doi:10.7488/DS/1405.

Wechselberger, P., Sagmeister, P., and Herwig, C. (2013). Real-time estimation of biomass and specific growth rate in physiologically variable recombinant fed-batch processes. *Bioprocess and Biosystems Engineering*, 36(9), 1205–1218. doi:10.1007/s00449-012-0848-4.

Zahel, T., Sagmeister, P., Suchocki, S., and Herwig, C. (2016). Accurate Information from Fermentation Processes – Optimal Rate Calculation by Dynamic Window Adaptation. *Chemie Ingenieur Technik*, 88(6), 798–808. doi:10.1002/cite.201500085.

Appendix A. B-SPLINE RECURSION

$S(t)$ is represented by a linear combination of basis splines $B_{j,k}(t)$: $S(t) = \sum_j \alpha_j B_{j,k}(t)$, where α_j are the coefficients. The basis splines $B_{j,k}(t)$ are defined recursively over a sequence of knots r_j. For $k = 0$:

$$B_{j,0}(t) = \begin{cases} 1 & \text{if } r_j \leq t < r_{j+1} \\ 0 & \text{otherwise} \end{cases}, \text{ and for } k > 0\text{:}$$

$$B_{j,k}(t) = \frac{t - r_j}{r_{j+k} - r_j} B_{j,k-1}(t) + \frac{r_{j+k+1} - t}{r_{j+k+1} - r_{j+1}} B_{j+1,k-1}(t).$$

To fit the coefficients α_j, the objective function includes a second-order derivative regularization term controlled by a smoothing parameter λ. The objective function to minimize is: $\sum_i (y_i - S(t_i))^2 + \lambda \int (S''(t))^2 \, dt$ where y_i are the observed data points, t_i are the corresponding positions, and $S''(t)$ is the second derivative of the spline function $S(t)$.