Systems & Control Transactions

Research Article
Foundations of Computer Aided Process Design (FOCAPD 2024)
Breckenridge, Colorado, USA. July 14-18, 2024
Peer Reviewed Conference Proceeding

PSE PRESS

Improving Mechanistic Model Accuracy with Machine Learning Informed Physics

William Farlessyost[a*], Shweta Singh[a,b]

[a] Purdue University, Agricultural & Biological Eng., West Lafayette, Indiana, USA
[b] Purdue University, Ecological & Environmental Engineering, West Lafayette, Indiana, USA
* Corresponding Author: wfarless@purdue.edu.

ABSTRACT

Machine learning presents opportunities to improve the scale-specific accuracy of mechanistic models in a data-driven manner. Here we demonstrate the use of a machine learning technique called Sparse Identification of Nonlinear Dynamics (SINDy) to improve a simple mechanistic model of algal growth. Time-series measurements of the microalga Chlorella Vulgaris were generated under controlled photobioreactor conditions at the University of Technology Sydney. A simple mechanistic growth model based on intensity of light and temperature was integrated over time and compared to the time-series data. While the mechanistic model broadly captured the overall growth trend, discrepancies remained between the model and data due to the model's simplicity and non-ideal behavior of real-world measurement. SINDy was applied to model the residual error by identifying an error derivative correction term. Addition of this SINDy-informed error dynamics term shows improvement to model accuracy while maintaining interpretability of the underlying mechanistic framework. This work demonstrates the potential for machine learning techniques like SINDy to aid simple mechanistic models in scale-specific predictive accuracy.

Keywords: Dynamic Modelling, Machine Learning, System Identification, Surrogate Model, Batch Process

INTRODUCTION

Within process design and system analysis, mechanistic models have long stood as pillars for understanding and predicting the behavior of complex systems. Rooted in fundamental physical and chemical principles, these models offer a structured approach to dissecting system dynamics, providing insights that are crucial for industrial applications [1]. However, as we delve deeper into the nuances of these systems, particularly at varying scales, the limitations of mechanistic models become apparent. A key challenge lies in their generalization; while these models are adept at capturing overarching trends and basic interactions, they often fall short in accurately representing the intricate dynamics of specific systems under varied conditions.

This is especially true in biological and ecological modeling, where the complexity of interactions and the sensitivity to environmental variables pose significant hurdles. The modeling of microalgal growth, such as that of Chlorella vulgaris, exemplifies this challenge. Here, the transition from laboratory-scale observations to commercial-scale operations is fraught with complexities, as the models struggle to account for the myriad of factors influencing algal growth dynamics at different scales [3,4].

Yet the uptick of machine learning methods bolstering traditional system identification offer a beacon of hope in this landscape. Among the various machine learning techniques, Sparse Identification of Nonlinear Dynamics (SINDy) emerges as a particularly promising tool. SINDy is a novel algorithm that seeks to identify the governing equations of a dynamical system in a sparse, interpretable form [2-4]. This approach is particularly relevant in process modeling, where understanding the fundamental dynamics is as crucial as achieving predictive accuracy [5].

The objective of this study is to showcase the application of SINDy in refining a mechanistic model of algal growth. Focusing on the microalga Chlorella vulgaris, we utilize time-series data collected under controlled photobioreactor conditions at the University of Technology

Sydney (Unpublished Data).

The following paper delves into the methods, where we describe the initial mechanistic model based on light intensity and temperature, followed by the integration of the SINDy algorithm derived error term. The results section presents a comparative analysis between the original and the enhanced models, highlighting the improvements in accuracy and interpretability. In the discussion, we explore the implications of these findings for process design, particularly in the scaling of microalgal growth systems. Finally, the paper concludes with a summary of our key findings and their significance in the broader context of industrial applications.

BACKGROUND

Process Design

Historically, process design has been deeply rooted in the principles of chemical engineering, with a strong emphasis on understanding and manipulating material and energy balances [6-8]. Mechanistic models, characterized by their basis in fundamental scientific principles, have been extensively applied across various domains of process engineering [1, 9-13]. These models excel in their ability to predict system behavior under a range of conditions, providing a reliable foundation for process design and optimization. However, their application is not without limitations. One significant challenge is their generalization, which can lead to inaccuracies when applied to specific systems or at different scales [14]. Biological systems are inherently complex, and their dynamics can vary significantly with changing environmental conditions and scales, posing a substantial challenge to traditional mechanistic modeling approaches.

Machine Learning

In recent years, the emergence of machine learning as a complementary tool in scientific research has opened new avenues in process design. Machine learning's data-driven nature allows it to uncover complex, non-linear relationships within high-dimensional data sets, offering a level of insight and prediction that is often beyond the reach of traditional models [15, 16]. This has been particularly advantageous in process modeling, where machine learning algorithms have been successfully integrated with mechanistic models to enhance their accuracy and adaptability [14].

Among the various machine learning techniques, SINDy has emerged as a promising tool for process design. Developed as an algorithm to identify governing equations of dynamical systems, SINDy distinguishes itself by its ability to distill complex data sets into sparse, interpretable models [2, 3]. This capability makes SINDy particularly suitable for process design applications where understanding the fundamental dynamics is as crucial as achieving predictive accuracy.

Microalgae

The modeling of microalgal growth, specifically the growth dynamics of Chlorella vulgaris, provides a compelling case study for the application of SINDy in process design [17, 18]. Microalgae, recognized for their potential in industrial applications, present significant challenges in scaling up from laboratory to commercial operations [18]. Traditional mechanistic models, while providing a basic understanding of algal growth, struggle to capture the full complexity of these biological systems at different scales [19, 20]

METHODOLOGY

Algae Mechanistic Model

We attempt to improve upon a simplified model that based on two key environmental factors: light intensity and temperature. This model combines the principles of Haldane kinetics and the Cardinal Temperature Model with Inflection (CTMI), offering a comprehensive understanding of algal growth dynamics. The model's first component, based on Haldane kinetics, addresses the effect of light intensity on algal growth. It recognizes that growth rate increases with light intensity up to an optimal level, after which it starts to decline due to photoinhibition. This relationship is mathematically represented by the equation:

$$\mu(I) = \mu opt * I\left(I + \left(\frac{\mu opt}{\alpha}\right) * \left(\left(\frac{I}{Iopt}\right) - 1\right)^2\right) \quad (1)$$

Here, $\mu(I)$ represents the growth rate as a function of light intensity I, μopt is the maximum growth rate, $Iopt$ is the optimal light intensity, and α is a parameter that moderates the effect of light intensity on the growth rate.

The second component of the model, the CTMI, quantifies the influence of temperature on algae growth. This model defines the growth rate within a range of temperatures, specifying minimum ($Tmin$), optimal ($Topt$), and maximum ($Tmax$) temperatures for growth. The CTMI model is encapsulated in the following equation:

$$\mu(T) = k * \frac{\left((T - Tmin) * (Topt - T)\right)}{\left((Topt - Tmin) * (Tmax - T)\right)} \quad (2)$$

In this formulation, $\mu(T)$ denotes the growth rate as a function of temperature T, with $Tmin$, $Topt$, and $Tmax$ being the minimum, optimal, and maximum temperatures for algal growth, respectively. k is a scaling constant.

The final model, encapsulating both the temperature and light components, is then the product of $\mu(T)$ and $\mu(I)$.

Data Source and Preparation

Our study utilizes data collected from the PBR 1250L photobioreactor system from Industrial Plankton as the primary cultivation equipment. This system provides a substantial working volume of 1250 liters, enabling a continuous harvesting capacity of up to 500 liters daily. Data collection was centered around minute measurements of pH level, irradiation, temperature, and relative density of the growth media within the photobioreactor. The pH control was managed by adjusting its setpoint and stabilized through automatic CO_2 injection. Irradiation was provided by LED light columns. Temperature control was maintained externally and the relative density of the algae was monitored through an optical density sensor.

The study involved a batch culture maintained over eleven days with alternating irradiation and pH setpoints, keeping the temperature constant and assuming a sufficient initial nutrient supply. The operational sequence was adjusted bi-daily, and the bioreactor was prepared with specific concentrations of culture media, followed by a regulated air bubbling process to maintain optimal growth conditions. This resulted in 22,301 data points for use in training and testing models.

SINDy-Derived Error Model

We employ the SINDy algorithm [21] to identify a differential equation to model the rate of change of error between the actual and predicted *Chlorella vulgaris* densities. Here, the error given by Equation 4 as

$$e = Density\_actual - Density\_model \quad (4)$$

varies over time. The state data, represented by X_{sindy} includes the error, capturing the essential aspects of the model's performance over time. Additionally, control data, denoted as U_{sindy}, incorporates key environmental factors, namely temperature T, light intensity I, and pH that significantly influence algal growth.

$$X_{sindy} = e \quad (5)$$

$$\dot{X}_{sindy} = \dot{e} \quad (6)$$

$$U_{sindy} = [T\ I\ pH] \quad (7)$$

We develop a control-only error model using the SINDy framework with the introduction of a dummy state variable. This dummy variable, a zero-valued array, is designed to match the dimensions of X_{sindy}, the derivative of the error. Its primary function is not to represent a 'real' state variable in the traditional sense; instead, it serves as a tool for facilitating the SINDy algorithm process of system dynamics identification, which typically expects state variables to influence the system dynamics. The SINDy algorithm is thus attempting to fit sparsity matrix Φ such that:

$$\dot{X}_{\{sindy\}} = U_{\text{sindy}}\Phi \quad (9)$$

To capture the complexity of biological growth processes in our SINDy model, we included a variety of custom functions in the feature library. Each function represents a specific biological growth dynamic, such as logistic growth, exponential growth, and other well-known biological models shown in Table 1.

Table 1: Functional basis for SINDy library.

Logistic growth	$f(x) = \frac{1}{\{1 + e^{\{-x\}}\}}$ (10)
Exponential growth	$f(x) = e^{\{x\}}$ (11)
Gompertz function	$f(x, c_1, c_2) = e^{\{c_1 e^{\{-c_2 x\}}\}}$ (12)
Allee effect	$f(x, A) = \frac{x}{\{A - x + \epsilon\}}$ (13)
Michaelis-Menten kinetics	$f(x, K_m) = \frac{x}{\{K_m + x + \epsilon\}}$ (14)
Holling's Type II function	$f(x, a) = \frac{ax}{\{1 + ax + \epsilon\}}$ (15)

Here, the logistic growth function represents a growth process that accelerates rapidly at first and then slows down as it approaches a maximum limit, commonly used to model population growth. The exponential growth function describes a process where the rate of growth is proportional to the current amount, leading to a rapid increase over time. The Gompertz function models growth that initially accelerates and then decelerates, commonly used in biology to describe the growth of tumors or certain populations [22]. The parameters c_1 and c_2 reflect the maximum value the function will reach and the rate at which that point will be reached respectively. The Allee effect function represents a scenario where population growth is positively correlated with the population size, highlighting the difficulties in growth at very low densities, with A representing this population threshold value [23]. The parameter ϵ is used here as well as the following functions to as a small value that prevents division by zero. The Michaelis-Menten kinetics function is typically used to describe the rate of enzymatic reactions, showing how reaction rate varies with substrate concentration [24]. The parameter K_m is the substrate concentration at which the growth rate is half of its maximum. Holling's Type II function, models the rate of predation in an ecosystem, increasing linearly at low prey density and saturating at higher densities [25]. The growth rate a here represents the efficiency of resource consumption by the growing population.

Mechanistic & Error Model Coupling

After the ODE model for the error derivative is recovered, this sum of functions representing the

derivative is added directly to the original mechanistic CTMI-light model, as shown in Figure 1. This updated, or "corrected," model can then be integrated overtime to return the specific density from the estimated growth rate.

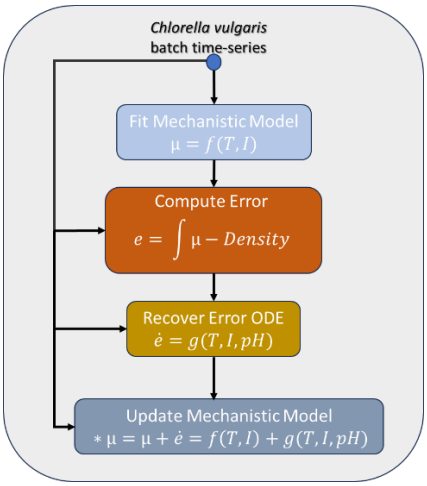

Figure 1: Methodological flow resulting in an updated mechanistic model for Chlorella vulgaris rate of change of density.

Model Evaluation

In the evaluation of the error ODE model, key metrics of RMSE, MAPE, and R^2 are calculated for each fold of the time series cross-validation to assess model accuracy. RMSE is determined by taking the square root of the average of squared differences between predicted and observed values. MAPE is calculated as the average of the absolute differences between predicted and observed values, divided by the observed values, and expressed as a percentage. R^2 is computed as the proportion of variance in the observed data that is predictable from the model inputs.

Visual comparisons of both the corrected and non-corrected models' predictions against the observed cumulative growth data are integrated and tested for each validation fold through a process of data preparation, plotting, visual analysis, and comparison. These plots allow for the assessment of how closely the predicted growth curves follow the observed data and the extent to which the corrected predictions improve alignment with the observed data compared to the non-corrected predictions.

RESULTS

Algae Mechanistic Model

We fit the combined CTMI-light model for density using the full 22,301 measurement points available for temperature, light intensity, and density. Since the mechanistic model is built for non-standardized relationships between components, we choose to pass in the raw data without standardization or normalization. We employed non-linear curve fitting techniques, utilizing the 'curve_fit' function from the Python 'scipy.optimize' library, to estimate the parameters of the model with $Topt = 27.329\,C$.

To check the performance of this model, we integrate the fitted model over time and compare the result to the measured specific density. As seen in Figure 2, the combined model predicts a linear increase in algae density overtime.

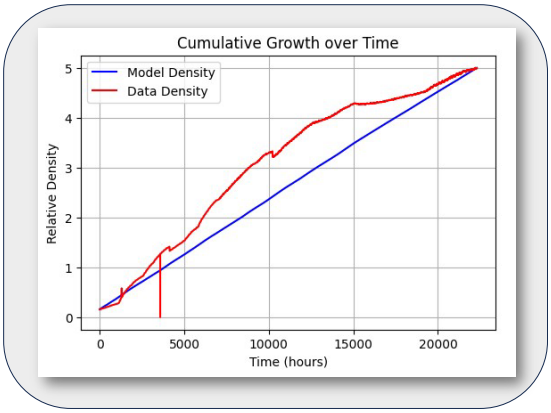

Figure 2: Combined CTMI-light model for growth rate integrated over time versus measured model specific density over all 22,301 data points. Combined model RMSE is 0.009.

Data Preparation

To make the SINDy optimization more robust against outliers in feature weights, the inputs, temperature, light intensity, and pH are each normalized by subtracting their time-series mean and dividing by standard deviation. This approach was chosen since the inputs to the batch were periodic, meaning their distribution was close to Gaussian.

The full timeseries data, with standardized control inputs, is then broken into twenty folds of sequential data. Cross-validation is employed such that in each iteration, the model is trained on an expanding window of past data and validated on a subsequent, non-overlapping window of future data. This strategy allows the model to learn from a progressively larger dataset, capturing more of the underlying temporal patterns and dynamics as it moves through the folds.

SINDy Error ODE Model Recovery

We leverage the sequentially thresholded least squares (STLSQ) algorithm as the optimization engine for Sparse Identification of Nonlinear Dynamics (SINDy). In this implementation, the STLSQ sparsity threshold is set at 0.0001, allowing the model to focus on the most impactful predictors and enhance its transparency. Additionally, the maximum iteration limit is set to 2000,

balancing computational efficiency with sufficient convergence time for the optimizer. These carefully chosen parameters ensure that the STLSQ algorithm effectively extracts the underlying dynamics of the system while maintaining interpretability and computational efficiency.

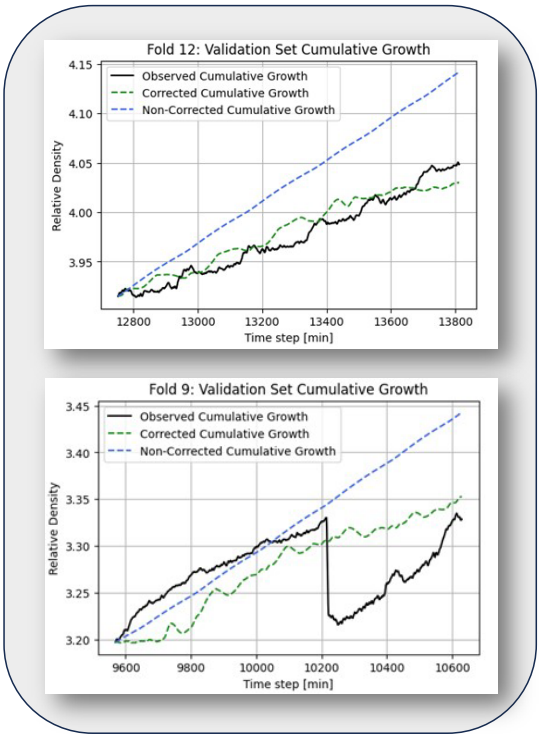

Figure 3: Updated growth rate model (integrated model versus actual specific density) showing better performance than original CTMI-light model.

We use a combination of custom, Polynomial, and Fourier libraries in SINDy. The custom library included logistic growth, exponential growth, Gompertz, Allee effect, Michaelis-Menten kinetics, and Holling's Type II functions, with all parameters uniformly set to one. This simplified configuration allowed for a consistent and straightforward interpretation of the model's outcomes while retaining the flexibility to model complex biological phenomena. The Polynomial Library, configured to include terms up to the second degree, enables the model to identify both linear and quadratic relationships. Meanwhile, the Fourier Library plays a key role in capturing remaining periodicity.

An ensemble method is utilized to enhance the model's robustness and accuracy [26]. This approach involves fitting the model multiple times using subsets of the training data, allowing for a comprehensive capture of potential dynamics and mitigating the impact of outliers or anomalous data points. The ensemble method generates 1000 models, each trained on a random subset comprising half of the training data.

Model Performance

A SINDy model is trained for each of the 20 validation folds and the error metric scores for RMSE, MAPE, and R^2 are tabulated in **Table 2**. Scores across the early folds reflect completely unbounded and divergent behavior in the corrected model when integrated over time, due to the limited training data up to that point. Scores continue to improve (RMSE and MAPE decreasing and R^2 increasing) until Fold 6, where a measurement fluke in specific density temporarily biases the training data. These high RMSE and MAPE and highly negative R^2 are seen again in Folds 14 and 15, reflecting completely divergent behavior likely the result of similar poor data quality around that point in time.

Table 2: Error metric scores across full 20 validation folds.

Fold #	Uncorrected *RMSE* *MAPE* *R^2*	Corrected *RMSE* *MAPE* *R^2*
1	1.15e+8, 2.31e+10%, -7.52e+17	1.055e+12, 1.50e+14%, -6.29e+25
2	0.075, 5.353%, 0.585	1.80e+10, 1.25e+12%, -2.38e+22
3	0.0925, inf%, 0.160	1.64e+11, inf%, -2.64e+24
4	0.0300, 1.34%, 0.876	0.399, 18.7%, -20.9
5	0.117, 4.62%, 0.328	0.867, 32.5%, -36.1
6	0.0969, 3.77%, 0.0973	3.92, 125%, -1475
7	0.0884, 2.75%, 0.361	0.0817, 2.71%, 0.455
8	0.0364, 1.06%, 0.805	0.0909, 1.95%, -0.215
9	0.0827, 1.80%, -4.90	0.0479, 1.28%, -0.981
10	0.055, 1.40%, 0.639	0.0684, 1.47%, 0.443
11	0.0235, 0.541%, 0.913	0.0480, 1.16%, 0.635
12	0.0595, 1.35%, -1.17	0.0132, 0.278%, 0.893
13	0.0138, 0.242%, 0.948	0.0263, 0.549%, 0.811
14	0.120, 2.37%, -310	2.53e+5, 3.76e+6%, -1.38e+15
15	0.0949, 1.90%, -29.2	12.1, 177%, -492000
16	0.0720, 1.39%, -7.01	0.223, 3.97%, -76.1
17	0.0844, 1.65%, -10.8	0.216, 4.14%, -76.5
18	0.0144, 0.260%, 0.942	0.0642, 1.19%, -0.144
19	0.0192, 0.326%, 0.873	0.0484, 0.806%, 0.197
20	0.0586, 0.964%, -2.34	0.109, 1.81%, -10.6

The fold models not exhibiting this fully divergent behavior either outperform the original combined CTMI-light model both visually quantitatively (for example models for Folds 9 and 12, see Table 2) or capture some pattern or periodicity not found in the original mechanistic models, while simultaneously slowly diverging from the actual truth specific density (for example models for Folds 8 and 18, see **Table 2**). In Fold 9, we see that the corrected model outperforms the original mechanistic model even with experimental error due to the drop in

specific density.

Model Trends

The SINDy error-based models are provided as Supplementary Material. The natural functions are given only by name and variable they operate on (see Table 1). For example, $13.512\ gompertz(Temperature, Light)$ refers to the Gomperz function applied to both temperature and light variables (separately) but with a coefficient of 13.512 for each.

Early fold equation coefficients (Folds 1-7) show noticeable fluctuation in the impact of temperature. Initially, the effect is strongly negative (e.g., -104689.0 in Fold 1) but shows variability in subsequent folds, indicating changing sensitivity to temperature (see Supplementary Material). Light maintains a predominantly negative coefficient, with large magnitudes (e.g., -43231100.0 in Fold 1). pH coefficients show significant changes, starting positive and large (8179580.0 in Fold 1) but becoming negative by Fold 4, indicating a shifting role of pH in these systems. Squared terms like Temperature^2 and Light^2 indicate evolving non-linear effects, with coefficients varying and reflecting changing dynamics. Interaction terms start to appear and vary in significance, suggesting the evolving interplay between factors. Logistic and exponential growth models are applied with varying coefficients, indicating differing growth dynamics under changing conditions.

Folds (Folds 8-14) see the coefficient of Temperature stabilizing somewhat but continuing to show variation, possibly reflecting a nonlinear role not captured by the supplied function basis for SINDy. Light consistently shows a negative impact but with varying magnitudes, indicating a sustained but dynamically changing influence. The role of pH continues to evolve with coefficient changes. Cyclical terms become more pronounced, with sin and cos functions indicating more significant seasonal or cyclic effects.

Late Folds (Folds 15-20) show temperature coefficients become more moderate, suggesting a refined understanding or a stabilizing effect. The influence of Light remains significant with large negative coefficients, indicating a consistently crucial role. pH effects continue to show variability but within a narrower range, suggesting a more consistent but still variable role. The complexity in terms of diverse interactions and non-linear effects shows signs of convergence, with key relationships and factors becoming more defined. There is a tendency towards more consistent use of growth models, indicating a stabilization in how growth or change processes are being represented.

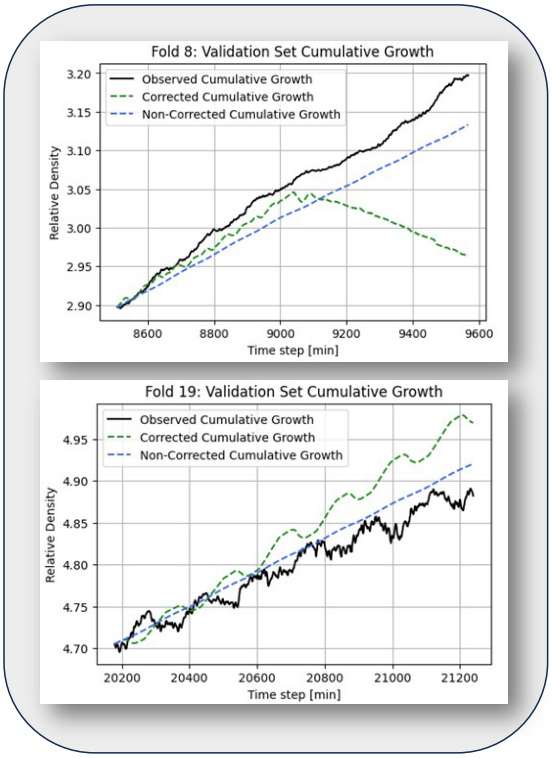

Figure 4: Updated growth rate model (integrated model versus specific density) showing divergent behavior.

CONCLUSION

This study demonstrates the potential of sparse machine learning techniques like SINDy to enhance the accuracy of mechanistic models while retaining interpretability. However, it also highlights the limitations and challenges that can arise in applying data-driven methods to improve mechanistic insights, especially given constrained data availability.

By focusing solely on a single experimental dataset for the microalga Chlorella vulgaris cultivated under tightly controlled photobioreactor conditions, the model's flexibility was restricted. While providing high temporal resolution measurements, spanning just one batch limited the diversity of growth dynamics captured. Periodic disturbances in the data, stemming from operational issues, further hampered efforts to train robust models.

Across 20 validation folds, adding the data-driven error dynamics equation resulted in improved error metrics. However, early folds suffered due to insufficient training data leading to a diverging trajectory with exploding error. Later folds demonstrated closer alignment with observed growth trajectory or responded to oscillation over time neglected by the linear mechanistic model. Inaccuracy in the phase of the oscillatory trajectory predicted may be due to sensitivity to initial conditions and warrants further exploration. The variation in types of biological growth functions applied over folds potentially indicated the model homing in on appropriate

representations for different mechanisms or growth phases as training data and timespan was increased.

These outcomes highlight the power of refining first-principles models with nuanced insights from real system observations. However, the instability across folds underscores the need for more comprehensive datasets before operational deployment. Future efforts should focus on generating measurements encapsulating a wider array of cultivation conditions and durations. Increased scale and variance in datasets may enable machine learning to extract more robust patterns. Beyond SINDy, exploring alternative algorithms could further improve the mechanistic correction.

DIGITAL SUPPLEMENTARY MATERIAL

The full set of error ODEs generated can be found here. Link: [Error_ODEs].

ACKNOWLEDGEMENTS

The author William Farlessyost acknowledges the National Science Foundation for support under the Graduate Research Fellowship Program (GRFP) undergrant number DGE-1842166. Thanks to Sebastian Oberst (UTS) for his help and support on this project.

REFERENCES

1. L. Mears, S. M. Stocks, M. O. Albaek, G. Sin, and K. V. Gernaey, "Mechanistic Fermentation Models for Process Design, Monitoring, and Control," *Trends Biotechnol*, vol. 35, no. 10, pp. 914–924, Oct. 2017, doi: 10.1016/J.TIBTECH.2017.07.002.
2. M. Quade, M. Abel, J. Nathan Kutz, and S. L. Brunton, "Sparse identification of nonlinear dynamics for rapid model recovery," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 28, no. 6, p. 63116, 2018.
3. S. L. Brunton, J. L. Proctor, and J. N. Kutz, "Discovering governing equations from data by sparse identification of nonlinear dynamical systems," *Proceedings of the national academy of sciences*, vol. 113, no. 15, pp. 3932–3937, 2016.
4. E. Kaiser, J. N. Kutz, and S. L. Brunton, "Sparse identification of nonlinear dynamics for model predictive control in the low-data limit," *Proceedings of the Royal Society A*, vol. 474, no. 2219, p. 20180335, 2018.
5. F. Harirchi, "On sparse identification of complex dynamical systems: A study on discovering influential reactions in chemical reaction networks," *Fuel*, vol. 279, p. 118204, 2020.
6. G. Towler and R. Sinnott, *Chemical Engineering Design: Principles, Practice and Economics of Plant and Process Design*. Elsevier, 2021. doi: 10.1016/B978-0-12-821179-3.01001-3.
7. M. Sharifzadeh, "Integration of process design and control: A review," *Chemical Engineering Research and Design*, vol. 91, no. 12, pp. 2515–2549, Dec. 2013, doi: 10.1016/J.CHERD.2013.05.007.
8. B. W. Bequette, "Prentice Hall PTR : Process Control: Modeling, Design, and Simulation," 2002.
9. K. V. Gernaey, A. E. Lantz, P. Tufvesson, J. M. Woodley, and G. Sin, "Application of mechanistic models to fermentation and biocatalysis for next-generation processes," *Trends Biotechnol*, vol. 28, no. 7, pp. 346–354, Jul. 2010, doi: 10.1016/J.TIBTECH.2010.03.006.
10. S. Esplugas, S. Contreras, and D. F. Ollis, "Engineering Aspects of the Integration of Chemical and Biological Oxidation: Simple Mechanistic Models for the Oxidation Treatment," *Journal of Environmental Engineering*, vol. 130, no. 9, pp. 967–974, Sep. 2004, doi: 10.1061/(ASCE)0733-9372(2004)130:9(967).
11. R. T. Kapoor, M. Danish, R. S. Singh, M. Rafatullah, and A. K. Abdul, "Exploiting microbial biomass in treating azo dyes contaminated wastewater: Mechanism of degradation and factors affecting microbial efficiency," *Journal of Water Process Engineering*, vol. 43, p. 102255, Oct. 2021, doi: 10.1016/J.JWPE.2021.102255.
12. A. K. Datta, "Toward computer-aided food engineering: Mechanistic frameworks for evolution of product, quality and safety during processing," *J Food Eng*, vol. 176, pp. 9–27, May 2016, doi: 10.1016/J.JFOODENG.2015.10.010.
13. Q. Lu and F. Jiao, "Electrochemical CO2 reduction: Electrocatalyst, reaction mechanism, and process engineering," *Nano Energy*, vol. 29, pp. 439–456, Nov. 2016, doi: 10.1016/J.NANOEN.2016.04.009.
14. M. Mozaffar *et al.*, "Mechanistic artificial intelligence (mechanistic-AI) for modeling, design, and control of advanced manufacturing processes: Current state and perspectives," *J Mater Process Technol*, vol. 302, p. 117485, Apr. 2022, doi: 10.1016/J.JMATPROTEC.2021.11748
15. J. Huang, Y. Chen, S. Jiang, and C. Yuan, "Machine learning-enabled intelligent modeling and optimization of chemical processes," *Chem Eng Sci*, vol. 203, pp. 290–310, 2019.
16. H.-J. Song, D.-Y. Jeong, and J.-H. Ryu, "Prediction of chemical reaction yield using nonparametric regression and machine learning methods," *Ind Eng Chem Res*, vol. 54, no. 44, pp. 10868–10876, 2015.
17. C. Safi, B. Zebib, O. Merah, P. Y. Pontalier, and C. Vaca-Garcia, "Morphology, composition, production, processing and applications of Chlorella vulgaris: A review," *Renewable and*

Sustainable Energy Reviews, vol. 35, pp. 265–278, Jul. 2014, doi: 10.1016/J.RSER.2014.04.007.
18. M. T. Ahmad, M. Shariff, F. Md. Yusoff, Y. M. Goh, and S. Banerjee, "Applications of microalga Chlorella vulgaris in aquaculture," *Rev Aquac*, vol. 12, no. 1, pp. 328–346, Feb. 2020, doi: 10.1111/RAQ.12320.
19. M. F. Blair, B. Kokabian, and V. G. Gude, "Light and growth medium effect on Chlorella vulgaris biomass production," *J Environ Chem Eng*, vol. 2, no. 1, pp. 665–674, Mar. 2014, doi: 10.1016/J.JECE.2013.11.005.
20. K. Rezaei, A. Javanshir, R. Barghbani, K. Rezaei, and A. Javanshir, "Investigating the Effects of Several Parameters on the Growth of Chlorella vulgaris Using Taguchi's Experimental Approach," *Article in International Journal of Biotechnology for Wellness Industries*, vol. 1, pp. 128–133, 2012, doi: 10.6000/1927-3037/2012.01.02.04.
21. S. L. Brunton, J. L. Proctor, and J. N. Kutz, "Sparse identification of nonlinear dynamics with control (SINDYc)," *IFAC-PapersOnLine*, vol. 49, no. 18, pp. 710–715, 2016.
22. P. Waliszewski and J. Konarski, "A Mystery of the Gompertz Function," *Fractals in Biology and Medicine*, pp. 277–286, Jan. 2005, doi: 10.1007/3-7643-7412-8_27.
23. J. C. Gascoigne and R. N. Lipcius, "Allee effects driven by predation," *Journal of Applied Ecology*, vol. 41, no. 5, pp. 801–810, Oct. 2004, doi: 10.1111/J.0021-8901.2004.00944.X.
24. B. Srinivasan, "A guide to the Michaelis–Menten equation: steady state and beyond," *FEBS J*, vol. 289, no. 20, pp. 6086–6098, Oct. 2022, doi: 10.1111/FEBS.16124.
25. H. N. Agiza, E. M. ELabbasy, H. EL-Metwally, and A. A. Elsadany, "Chaotic dynamics of a discrete prey–predator model with Holling type II," *Nonlinear Anal Real World Appl*, vol. 10, no. 1, pp. 116–129, Feb. 2009, doi: 10.1016/J.NONRWA.2007.08.029.
26. U. Fasel, J. N. Kutz, B. W. Brunton, and S. L. Brunton, "Ensemble-SINDy: Robust sparse model discovery in the low-data, high-noise limit, with active learning and control," *Proceedings of the Royal Society A*, vol. 478, no. 2260, 2022, doi: 10.1098/RSPA.2021.0904.