Data-driven symbolic regression reveals microbial growth laws

T. Anthony Sun[1], Dovydas Kičiatovas[1], Inga-Katariina Aapalampi[2], Teemu Kuosmanen[1], Teppo Hiltunen[2], and Ville Mustonen[1]

[1]Organismal and Evolutionary Biology Research Programme, Department of Computer Science, University of Helsinki, Helsinki, Finland.
[2]Department of Biology, University of Turku, Turku, Finland.

2025-04-29

Abstract

Microbial growth is intimately dependent on the resources that cells find in their environment. While resource consumption in simple systems is well understood in theory, connecting mathematical models with empirical observations in more complex environments remains difficult. To address this, we measure bacterial growth across various concentrations of a complex growth medium and analyse the data with the assumption that populations consume resources at a constant rate. This approach highlights cumulative population gain as a key observable, revealing otherwise hidden resource dynamics.

We show that using cumulative population gain allows machine learning algorithms to predict growth rates more accurately than instantaneous population size, supporting the constant-rate resource consumption model. However, black-box machine learning says very little explicitly about the actual growth laws. To overcome this, we apply symbolic regression in a biologically informed way to reveal that additive and multiplicative models based on a Monod dynamics provide the best balance between data fit, model parsimony, and biological relevance.

Our study demonstrates how combining theoretical knowledge with symbolic regression can advance the understanding of microbial growth, bridging the gap between mathematical modelling and empirical research in microbial ecology.

1 Introduction

Microbial growth is a key process in the ecology and evolution of microbes, and understanding it quantitatively is essential for public health in the face of rising antibiotic resistances in human pathogens [1, 2], and for various biotechnology applications [3, 4]. Yet, quantifying how microbial populations convert resources

into biomass remains challenging, with most experiments relying on chemically complex media and short culturing times. To our knowledge, there has been no simple way to use batch culture data to infer how microbial populations grow, without relying on strong prior modelling assumptions. We know for certain that a growing microbial population uptakes resources from the environment and converts them to new biomass, and previous research has established that growth in batch cultures involve two intrinsic nonlinearities. On the one hand, the typical lag phase observed in growth experiments [5] implies that minimal models must include a time-dependent adjustment function to account for physiological changes occurring in bacterial cells encountering a fresh culture medium [6, 7, 8, 9]. On the other hand, the maximum per capita growth rate has a saturating response to an increase in resource concentration, demonstrated by Monod's classical work on a single resource [10].

A wealth of theoretical work has proposed models for multi-resource growth [11, 6, 7, 12, 5, 13, 14, 15]. Yet, quantitative analyses of empirical bacterial growth have often circumvented any explicit account of resource dynamics by using traditional models based on logistic growth [8, 16], which might misrepresent microbiological processes at work [17, 18, 19, 20, 21, 22]. Our inability to fully use the biologically more appropriate resource-consumer models does not arise from a lack of theoretical tools [23], but from the fact that model complexity often far surpasses experimental evidence. This evidence typically covers observed population data but not resource dynamics, which creates a disconnect between observable phenotypes related to macroscopic growth (*e.g.* yield, effective growth rate, lag time) and underlying microscopic mechanisms (*e.g.* maximum growth rate, resource consumption) [24]. Indeed, several incompatible multi-resource models may fit experimental data equally well [25, 26, 27, 15, 28]. Thus, we lack decisive methods to assess models for microbial growth *in vivo* [15] that result from the consumption of potentially multiple resources.

Applying machine learning techniques to microbiology [16] is promising, but canonical examples, such as Random Forest regression, yield black-box models that are notoriously difficult to interpret. So far, an interpretable, yet automated, method to uncover dynamical growth models from experimental data has been missing. One promising and fully interpretable alternative to black-box machine learning is symbolic regression, a method providing explicit mathematical expressions directly from the data [29]. Symbolic regression has already yielded impressive results in recovering and expanding known laws in physics [30] and, in the context of microbial ecology, in reverse-engineering Lotka-Volterra and logistic-based models [31, 32, 33, 34].

Here, we make minimal theoretical assumptions to better understand experimental and simulated growth data. Besides key nonlinearities, we assume constant-rate resource consumption (figure 1). This suffices to capture experimental and simulated observations to a great extent with a black-box Random Forest regression algorithm. Next, we crack this algorithmic black-box open to reverse-engineer explicit dynamical models for the microscopic growth process. Using symbolic regression, we first confirm our black-box results that the cumulative population gain is a useful handle to access the latent resource dynamics. Then, we use theoretical insights to inform the symbolic regression algorithm by specifying theory-based mathematical expression constraints and select for biologically interpretable models. Trading off between parsimony and best fit, the algorithm helps us to reveal that bacterial growth widely follows a Monod

dynamics in our experiments, with either one or two underlying resources.

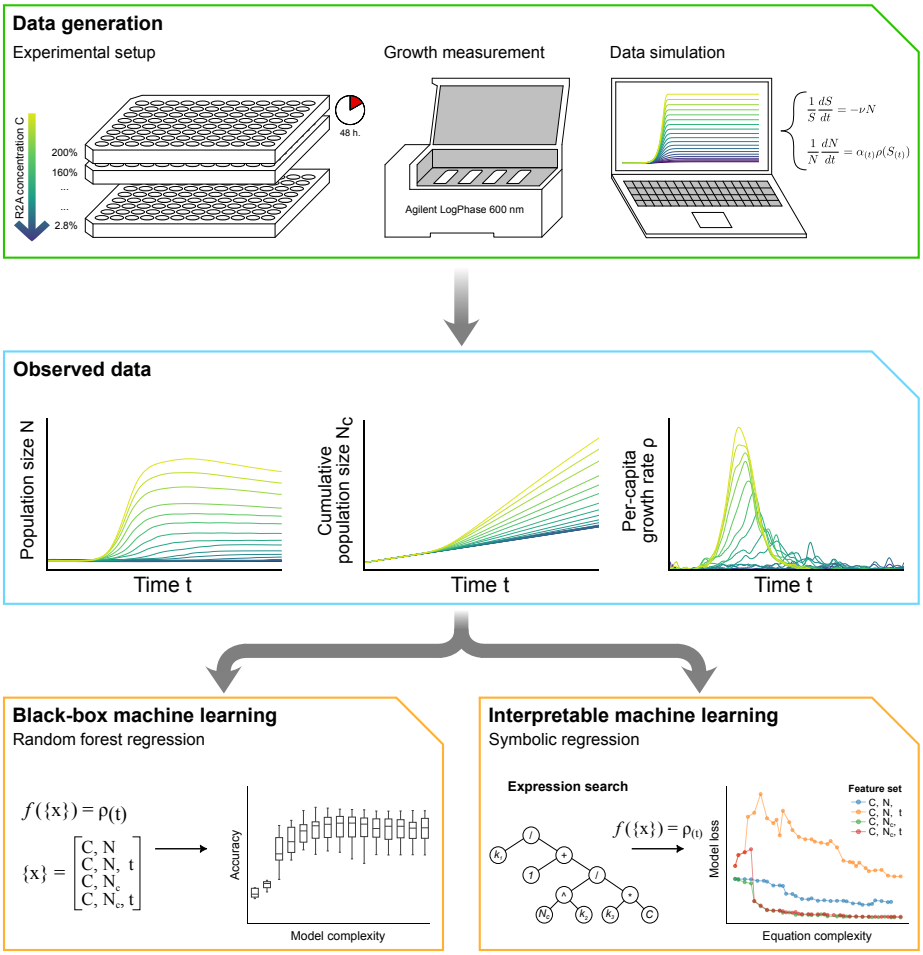

Figure 1: Workflow for the production and analysis of the datasets.

2 Materials and methods

2.1 Experimental methods for bacterial cultures

Our experiments involved 16 species from the HAMBI collection (Helsinki Institute of Life Science HiLIFE, Biodiversity Collections Research Infrastructure HUBCRI, Microbial Domain Biological Resource Centre HAMBI) summarised in table 1.

Solution A contained 11.28 g of M9 minimal medium salts (MP Biomedicals, lot `142662`) in 1.0 L of Milli-Q purified water. Culture medium B (100 %) was prepared by dissolving then autoclaving 3.0 g of Reasoner's 2A broth powder (Neogen, lot `UK208078/190`) into 1.0 L of solution A. We prepared a range of media B (concentrations C) by successive dilutions of medium B (200 %) into autoclaved solution A by an incremental factor of 0.8. Bacterial

Table 1: Bacterial species used in our experiments, and their HAMBI code.

HAMBI	identification	HAMBI	identification
1972	*Aeromonas caviae*	1292	*Morganella morganii*
105	*Agrobacterium tumefaciens*	3031	*Niabella yanshanensis sp. Nov.*
2160	*Bordetella avium*	2443	*Paracoccus denitrificans*
1287	*Citrobacter koseri*	1977	*Pseudomonas chlororaphis*
403	*Comamonas testosteroni*	6	*Pseudomonas putida*
2164	*Cupriavidus necator*	1896	*Sphingobacterium spiritivorum*
1299	*Kluyvera intermedia*	1842	*Sphingomonas yanoikuyae*
3237	*Microvirga lotononidis sp. Nov.*	2659	*Stenotrophomonas maltophilia*

pre-cultures D were run by inoculating each species into 6.0 mL of medium B (100 %), then incubating for 48 h in aerobic conditions at 30 °C under agitation at 70 rpm, and finally diluted into autoclaved solution A by a factor 10^{-4}.

Monocultures were run by inoculating 110 µL of each medium B (of appropriate concentration C) with 20 µL of pre-culture D onto a 96-well plate, then incubating for 48 h in aerobic conditions at 30 °C under agitation at 800 rpm. An Agilent BioTek LogPhase 600 device automatically recorded time and optical density at 600 nm every 10 min as a proxy to population size. We derived all other observations numerically. Figure S**??** summarises the data.

2.2 Theoretical methods and models

Consider a bacterial population of size N growing in a substrate of concentration S. It is commonly assumed that it uptakes resources at a constant-rate ν relative to its size:

$$\frac{1}{S}\frac{dS}{dt} = -\nu N \Rightarrow S = CS^0 \exp\left[-\nu \int_0^t N(u)du\right] := CS^0 \exp\left(-\nu N_c\right), \qquad (1)$$

where C is the relative initial concentration of the growth medium compared to a reference, and $S^0 = S_{(C=1,t=0)}$ is the resource density in the reference concentration. This constant-rate model for resource consumption readily extends to several independent resources and naturally gives a paramount role to the cumulative population gain N_c. Conversely, the predictive value of N_c in explaining empirical growth data supports the constant-rate resource consumption model.

To account for the lag period observed in experiments, minimal models of bacterial growth assume a time-dependent physiological adjustment α in their population dynamics [11, 6, 15]:

$$\rho_{\text{obs}} := \frac{1}{N}\frac{dN}{dt} = \alpha(t) \times \rho(S), \qquad (2)$$

where ρ is the resource-dependent growth function. How physiological adjustment delays bacterial cells' growth dynamics is usually represented by similar mathematical forms [6, 7, 35, 9, 15]. From fits to real data, we know that just two phenomenological parameters suffice to capture the adjustment function [8, 9], namely the initial physiological state of the cell (population) q and the rate of change m in adjusting its physiology to the present environmental state:

$$\alpha(t) = \frac{q}{q + \mathrm{e}^{-mt}}. \qquad (3)$$

Monod showed that the maximum growth $\rho_{\max}$ as a function of increasing resource concentration was increasing then saturating [10]. The original demonstration, as well as automatic reverse-engineering of the model [36], concernonly the maximum per capita growth rate $\rho_{\max}(S)$ in a bacterial growth experiment. However, this type II functional response pattern [37] is usually modelled using a hyperbolic function [38, 10, 39] and integrated as a function of time into a dynamical model [24]:

$$\rho(S) = g_{\max}\, g(S) \quad \text{with} \quad g(S) := \frac{S}{K+S}, \tag{4}$$

where K is the substrate concentration inducing half-maximum response. Without further assumptions, this functional form cannot model multi-resource growth, which may be complicated by non-mutually-exclusive resources [40], colimitation [41, 14], and sequential preference [42]. As it is unclear how several resources $(S_0, \ldots, S_s)$ together impact bacterial growth, multi-resource growth models combine each resource's Monod contribution to the total growth in various ways, notably in a multiplicative or additive manner [41, 13, 28, 14]:

$$\rho_M(S_0, \ldots, S_s) = g_{\max} \prod_{x=0}^{s} g(S_x), \tag{5}$$

$$\rho_A(S_0, \ldots, S_s) = g_{\max} \sum_{x=0}^{s} g(S_x). \tag{6}$$

In what follows, the multiplicative multi-resource model M combines Eq. (1), Eq. (2) and Eq. (5) and assumes that growth requires all resources. The additive multi-resource model A combines Eq. (1), Eq. (2) and Eq. (6) and assumes that any resource allows for some growth. A common framework can unify several formulations [27], but at the cost of a combinatorial explosion of theoretical parameters, as the number of resources increases. Overall, experimental data provides no definitive support for either model over the other [27, 28].

2.3 Numerical methods

Each dataset consisted of a range of bacterial growth curves at different initial medium concentrations C for a given species. We used Python (3.10.14) for analysis, NumPy (1.26.4) and Pandas (2.2.2) for numerical protocols, Matplotlib (3.8.4) for visualisations, Scikit-learn (1.4.2) for Random Forest methods, PySR (1.3.1) for symbolic regression [43]. Code is available at: `https://github.com/anthony-sun/SR-Monod`.

For simulations, parameter ranges were chosen so as to match the orders of magnitude encountered in experimental datasets. Each dimension of the initial resource vector was set to 1; and the initial population size, to 0.1. We used the Runge-Kutta 4(5) method to integrate over the time span $[0, 50]$ with a time step of 1/6, matching the timing (h) of experimental samples. Cumulative population gain N_c was computed using the trapezoidal rule.

For Random Forest regression, we split each dataset into a training set (75 %) and a testing set (25 %) comprising every fourth growth curve along the range of medium concentrations. The predicted response variable was the per capita growth rate ρ_{obs}. Feature variables included: one population size

feature, namely N or N_c, the relative initial medium concentration C, the time t.

For symbolic regression, we split each dataset into a training set (50 %) and a testing set (50 %) comprising every other growth curve along the range of medium concentrations. For a given complexity bound on output expressions, the algorithm optimised data fits based on mean squared-error. More complex functional forms resulted from randomly combining feature variables and constants using: addition (+), multiplication (×), division (/), and constant power ($\hat{}^k, k > 0$). We imposed biology-grounded restrictions against negative constants and non-integer powers of the relative medium concentration C. To facilitate numerical computations, we normalised each dataset's cumulative population gain N_c by its maximum value, and we used the exponential negative scaled time $U = \mathrm{e}^{-t}$ and the exponential negative cumulative population gain $F = \mathrm{e}^{-N_c}$ as features. Finally, we considered the following expression templates:

$$\rho_{\text{obs}} = \alpha(t) \times \rho(C, N_c) = \frac{q}{q + \mathrm{e}^{-mt}} \times \rho(C, N_c), \tag{7}$$

$$\rho_{\text{obs}} = \alpha(t) \times \rho(C \times h(N_c)) = \frac{q}{q + \mathrm{e}^{-mt}} \times \rho(C \times h(N_c)). \tag{8}$$

Template (7) integrated Eq. (2) and Eq. (3) by imposing the search for adjustment parameters q and m. Template (8) inspired by Eq. (1) additionally constrained the mathematical form of the argument for resource-dependent growth ρ.

3 Results

3.1 Cumulative population gain allows for black-box prediction of growth

We hypothesised that we could approximate resource consumption as a constant-rate process. According to Eq. (1), this would give rise to cumulative population gain N_c as a key lens to study population dynamics. Therefore, if our hypothesis was true, then bacterial per capita growth rate ρ_{obs} should be better explained by the cumulative population gain N_c, than by the traditional approach through instantaneous population size N. To test that, we compared the performance of a black-box Random Forest regressor using either N or N_c.

Using N_c instead of N improved the performance of a Random Forest regressor in predicting per capita growth rate ρ_{obs} (figure 2, $\{C, N\}$ and $\{C, N_c\}$). The advantage of N_c alone over N alone did not result from some experimental growth curves being non-monotonic, because we obtained similar results with monotonic, simulated data (figure S**??**). Feature N_c effectively captured the impact of both N and t since feature sets $\{C, N, t\}$, $\{C, N_c\}$, and $\{C, N_c, t\}$ achieved a similar improvement. They also yielded simpler decision trees: while relying on N led to a median maximum tree depth of 35, relying on N_c instead reduced it to 25 (table 2). When capping the maximum depth of all decision trees from the start, all three yielded a systematic improvement of the Random Forest regression compared to feature set $\{C, N\}$ (figure 2).

When relying on population size N to predict the per capita growth rate ρ_{obs}, the Random Forest regressor showed systematic overfitting, yielding a gap

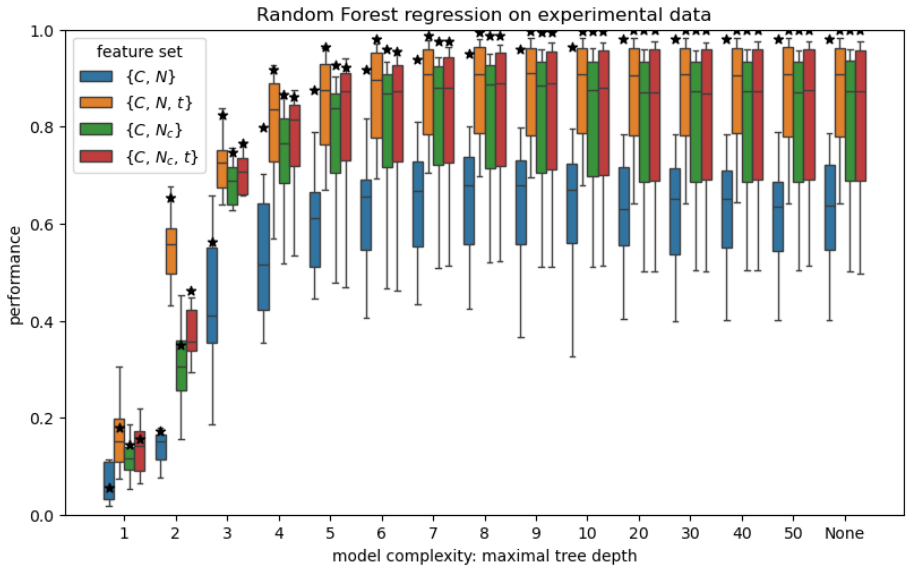

Figure 2: Performance of a black-box Random Forest regressor in predicting the per capita growth rate from different feature sets in experimental data, given by test score × coefficient of determination R^2. Black stars show the median train score across species. The cumulative population gain predictor N_c outperforms the population size predictor N.

Table 2: Unrestricted complexity of a Random Forest regressor using 100 trees to predict the per capita growth rate: median across 16 experimental datasets, each comprising 5780 points.

feature set	train score	test score	median maximum tree depth across species (SD)
$\{C, N\}$	98.1 %	63.7 %	35 (5)
$\{C, N, t\}$	99.9 %	90.8 %	27 (3)
$\{C, N_c\}$	99.9 %	87.2 %	25 (3)
$\{C, N_c, t\}$	99.9 %	87.4 %	26 (3)

between train and test scores. Models relying instead on cumulative population gain N_c reduced this overfit. We found that growth data simulated with resource-consumer models intrinsically produced both this overfit and its reduction by N_c (figure S**??**). On the one hand, they occurred even in the absence of the time-dependent adjustment α. On the other hand, they did not occur with simulations from a non-resource-consumer model. Their occurrence in experimental data supports a resource-consumer framework against a logistic-based model of bacterial growth (figure S**??**).

Experimental and simulated data confirmed that, compared to population size N, cumulative population gain N_c improved the performance of a Random Forest black-box, including by reducing the overfitting. This supported the constant-rate approximation for resource consumption in our experimental setting. As the three feature sets $\{C, N, t\}$, $\{C, N_c\}$ and $\{C, N_c, t\}$ yielded a similar improvement, N_c seemed to be an effective unidimensional approximation of the

bidimensional space (N, t). But at this point, it remained unclear which feature combination best reflected the actual mechanisms underlying biological growth. Connecting these observable features to the latent space of resources that bacteria convert into new biomass still required cracking the growth black-box and obtaining interpretable models.

3.2 Unconstrained symbolic regression predicts growth explicitly

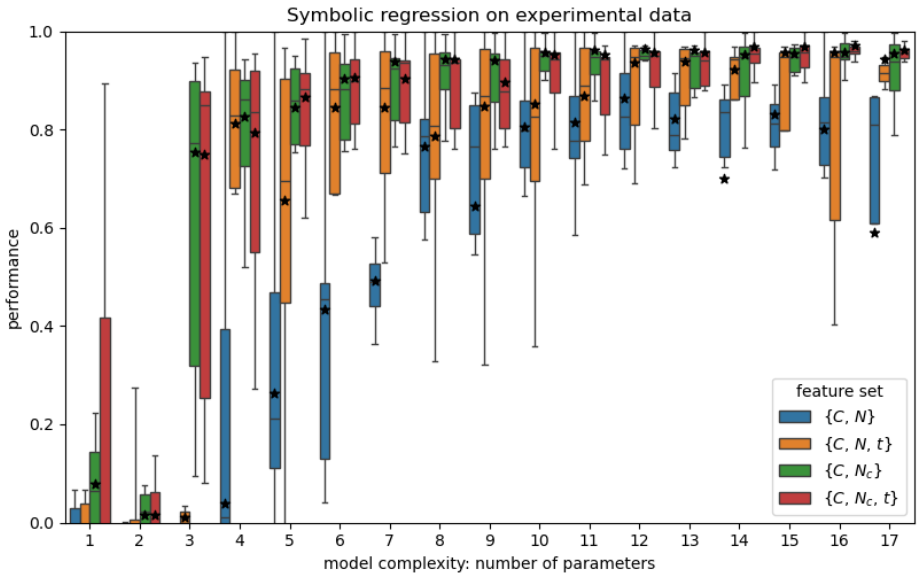

Figure 3: Performance of the unconstrained symbolic regressor in predicting the per capita growth rate from different feature sets in experimental data, given by test score $\times$ coefficient of determination R^2. Black stars show the median train score across species. The cumulative population gain predictor N_c outperforms the population size predictor N as well as its combination with time t.

Our hypothesis was that growth curves in a range of medium concentrations should allow for the inference of dynamical growth models. We demonstrated it with a symbolic regression algorithm. Symbolic regression consists in optimising a functional form predicting a response variable under constraints on formula complexity, thus providing explicit mathematical models [43]. We used this tool to reproduce the results of the black-box Random Forest approach, and confirmed their consistency in supporting an N_c-focused constant-rate consumption model.

Our symbolic regression engine provided mathematically explicit models. We found that models relying on cumulative population gain N_c systematically outperformed models relying on population size N (figure 3, $\{C, N\}$ and $\{C, N_c\}$), which reproduced our black-box results in laboratory experiments and in simulations. Feature set $\{C, N_c\}$ yielded the best predictive performance overall (figure 2, 3), and we noted that the N_c-over-N improvement did not depend on the time feature t (figure 3, $\{C, N, t\}$, $\{C, N_c\}$ and $\{C, N_c, t\}$).

These results suggested that, instead of using N_c to approximate the bidimensional space (N, t), algorithms conversely used N and t together to computationally approximate N_c. Indeed, this aligned with the theoretical formulation of Eq. (1), which implied a map from resource S onto the unidimensional space N_c. Thus, cumulative population gain N_c appeared to best reflect the actual biological mechanisms at work, and we subsequently focused on feature set $\{C, N_c\}$.

Our naïve symbolic regression algorithm confirmed the potential for N_c to inform us on microbial growth resources. It was able to mimic the behaviour of our black-box Random Forest regressor in experimental and simulated data. Yet, it could not yield substantial biological insight beyond the importance of N_c, because the models that it inferred did not generally reflect theoretical expectations from the literature.

3.3 Constrained symbolic regression infers biological growth laws

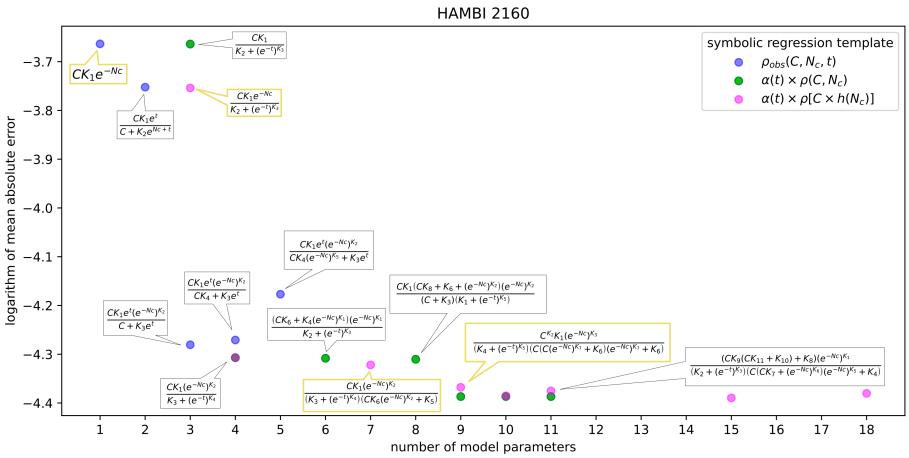

Figure 4: Symbolic regression Pareto front optimising model complexity and fitting performance against real data in *Bordetella avium* (HAMBI 2160), obtained with three templates: blue indicates no constraints; green indicates time-adjustment as in Eq. (2) and Eq. (3); pink additionally indicates concentration- and resource-dependency suggested by Eq. (1). From left to right, the four models highlighted with a yellow frame correspond to: 1) a constant-rate resource consumption without time-dependent adjustment as in Eq. (1); 2) a time-adjusted constant-rate resource consumption adding in Eq. (2) and Eq. (3); 3) a time-adjusted, Monod-like resource consumption learning Eq. (4); 4) a time-adjusted, multiplicative, bidimensional Monod-like resource consumption integrating Eq. (5). This provides evidence for the automatic inference of Eq. (4) and Eq. (5), with a substantial drop in fitting error.

So far, we had found solid support for the constant-rate assumption for resource consumption, as formulated in Eq. (1). We now hypothesised that the symbolic regression algorithm could infer a biologically interpretable growth

function, when guided by theory-based expectations. We constrained the equation search space by imposing functional forms that the symbolic regression algorithm should be looking for. Thus, we ran two rounds of symbolic regression informed by theoretical assumptions, using expression templates given first by Eq. (7) then by Eq. (8).

For each species, the algorithm proposed models of increasing complexity, which included 168 models learnt from experimental datasets. As expected, models with more parameters generally improved fits to data (figure 4). Then, to avoid focusing on models with no biological interpretability, we manually selected models compatible with literature-based theoretical expectations, such as Eq. (1). This task could not be automated as it implied critically assessing parsimonious, well-fitted models against biological relevance in current research (supplementary table). For instance, all formulas shown in figure 4 trade off fit to data and model simplicity, but only a few (highlighted in yellow) are interpretable in light of the literature in microbiology.

Table 3: Formula, interpretation and inferred resource dimensionality of the best-fitting, biologically interpretable growth models for ρ in Eq. (2) and Eq. (4). The symbol L represents linear model converting resources to growth.

model for $\rho(S)$	interpretation	resource dim.	HAMBI bacterial species
$g_{\max} S$	L	1	2 species: `2164` and `2443`
$g_{\max} \frac{S}{K+S}$	Monod	1	10 species
$g_{\max} S_1 \times \frac{S_2}{K+S_2}$	L × Monod	2	2 species: `403` and `1896`
$g_{\max} \frac{S_1}{K_1+S_1} \times \frac{S_2}{K_2+S_2}$	Monod × Monod	2	1 species: `2160`
$g_{\max} \left(\frac{S_1}{K_1+S_1} + \frac{S_2}{K_2+S_2} \right)$	Monod + Monod	2	1 species: `1287`

Table 3 summarises the best-fitting models that remained biologically interpretable in experiments. In all species, the algorithm started with a linear model, where the resource-dependent growth ρ was a linear function of a latent environmental resource S. This remained the best interpretable model for two of 16 experimental species, suggesting that experimental conditions did not cover the saturating side of Monod's formulation. In all other species indeed, we retrieved a dynamical formulation of the Monod model corresponding to Eq. (1) and Eq. (2). This implied an effectively unidimensional resource space where the per capita growth rate was a saturating function of a single environmental resource. In four cases, the algorithm further found that the observed growth was better explained by an effectively bidimensional resource, with constant-rate or Monod-like components. The two species with bidimensional Monod components yielded exactly Eq. (5) and Eq. (6).

4 Discussion

First, we showed that cumulative population gain N_c provides a powerful and simple way to indirectly observe the latent resource space experienced by growing microbial populations, as captured by Eq. (1). Then, we used symbolic regression to learn explicit growth models from population size data, without relying on prior theory. While fits to the data were excellent, an unconstrained use of the method did not ensure biological relevance. Therefore, we combined

this approach with theoretical expectations based on the literature by templating expressions to constrain equation search. We found that the dynamical resource-consumer model based on Monod's formulation in Eq. (4) best trades off among fit to data, symbolic simplicity, and biological interpretability, which extends previous non-dynamical findings [36]. Furthermore, when two underlying resources were inferred, the evidence supported both a multiplicative and an additive combination of Monod components, as in Eq. (5) and Eq. (6) respectively. Our results validate the general approximation, formulated by Eq. (1) and Eq. (2), that bacterial growth follows a dynamical Monod model with latent resources consumed at a constant rate. Overall, they highlight the potential of symbolic regression as a data-driven discovery tool while emphasising the need for integrating existing theory and biological knowledge as informative priors.

Despite its success in physics [30], symbolic regression has mainly recovered logistic-based models in microbiology [31, 32, 33, 34], thus overlooking two limitations. First, such models fail [17, 18, 19, 20, 21, 22] to account for key resource dynamics mediating the interactions [44, 24] that explain how bacterial communities assemble [45, 46, 47], grow, organise and change [48], through competition and resource utilisation traits [49, 50]. Secondly, assuming a functional form *a priori* leaves the plethora [11, 6, 7, 12, 5, 13, 14, 15] of growth functions proposed in the literature unexploited. Rather than developing more and more complex theory, it is essential to better connect empirical data on microbial growth with available models offering easy computation and best understanding. We demonstrated how to combine pre-existing knowledge with the potential for symbolic regression to automatically infer an effective growth function from data, with little to no prior on its functional form.

Our method provides an access to the effective resource dimensionality experienced by a growing microbial population, *i.e.* the minimal number of virtual, independent resources theoretically required to reproduce the observed population dynamics. But since inferred resource dimensions are not reducible to one another [51], they do not, in principle, reflect physical substrates, such as carbon sources. On the one hand, departing from such a chemical view allows us to translate population dynamics into effective ecological niches, which are generally disconnected from chemical resources [52]. Thus, our approach connects microbiology experiments with a rich literature on multidimensional niches [53, 54, 55] and their overlap, which determine species coexistence in communities [56, 57, 58, 59]. On the other hand, it is unclear how modelled resource dimensions relate to chemically defined substrates in the laboratory.

Our study demonstrates that symbolic regression offers a promising, empirically viable, and scalable way to analyse routine experiments in microbiology to infer growth laws and effective resource dimensionalities. Future research could modulate the concentration of chemical resources individually, in a culture medium, to allow for a high-definition view of the growth process in a complex environment.

References

[1] Antimicrobial Resistance Collaborators. 2022 Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *The Lancet* **399**, 629–655. (10.1016/s0140-6736(21)02724-0)

[2] Wong C. 2024 Antibiotic resistance is a growing threat: is climate change making it worse?. *Nature*. (10.1038/d41586-023-04077-0)

[3] Lässig M, Mustonen V, Nourmohammad A. 2023 Steering and controlling evolution — from bioengineering to fighting pathogens. *Nature Reviews Genetics* **24**, 851–867. (10.1038/s41576-023-00623-8)

[4] Yeoman K, Fahnert B, Lea-Smith D, Clarke T. 2020 Microbial Biotechnology and Agriculture. In Yeoman K, Fahnert B, Lea-Smith D, Clarke T, editors, *Microbial Biotechnology* , . Oxford: Oxford University Press. (10.1093/hesc/9780198822813.003.0008)

[5] Allen RJ, Waclaw B. 2018 Bacterial growth: a statistical physicist's guide. *Reports on Progress in Physics* **82**, 016601. (10.1088/1361-6633/aae546)

[6] Baranyi J, Roberts TA. 1994 A dynamic approach to predicting bacterial growth in food. *International Journal of Food Microbiology* **23**, 277–294. (10.1016/0168-1605(94)90157-0)

[7] Huang L. 2011 A new mechanistic growth model for simultaneous determination of lag phase duration and exponential growth rate and a new Bělehrádek-type model for evaluating the effect of temperature on growth rate. *Food Microbiology* **28**, 770–776. (10.1016/j.fm.2010.05.019)

[8] Ram Y, Dellus-Gur E, Bibi M, Karkare K, Obolski U, Feldman MW, Cooper TF, Berman J, Hadany L. 2019 Predicting microbial growth in a mixed culture from growth curve data. *Proceedings of the National Academy of Sciences* **116**, 14698–14707. (10.1073/pnas.1902217116)

[9] Borse F, Kičiatovas D, Kuosmanen T, Vidal M, Cabrera-Vives G, Cairns J, Warringer J, Mustonen V. 2024 Quantifying massively parallel microbial growth with spatially mediated interactions. *PLOS Computational Biology* **20**, e1011585. (10.1371/journal.pcbi.1011585)

[10] Monod J. 1949 The Growth of bacterial cultures. *Annual Review of Microbiology* **3**, 371–394. (10.1146/annurev.mi.03.100149.002103)

[11] Zwietering MH, Jongenburger I, Rombouts FM, van 't Riet K. 1990 Modeling of the Bacterial Growth Curve. *Applied and Environmental Microbiology* **56**, 1875–1881. (10.1128/aem.56.6.1875-1881.1990)

[12] De Jong H, Casagranda S, Giordano N, Cinquemani E, Ropers D, Geiselmann J, Gouzé JL. 2017 Mathematical modelling of microbes: metabolism, gene expression and growth. *Journal of The Royal Society Interface* **14**, 20170502. (10.1098/rsif.2017.0502)

[13] Letten AD. 2022 rescomp: An R package for defining, simulating and visualizing ODE models of consumer-resource interactions. *BioRxiv*. (10.1101/2022.01.11.475574)

[14] Held NA, Krishna A, Crippa D, Battaje RR, Devaux AJ, Dragan A, Manhart M. 2024 Nutrient colimitation is a quantitative, dynamic property of microbial populations. *Proceedings of the National Academy of Sciences* **121**. (10.1073/pnas.2400304121)

[15] Ghenu AH, Marrec L, Bank C. 2024 Challenges and pitfalls of inferring microbial growth rates from lab cultures. *Frontiers in Ecology and Evolution* **11**. (10.3389/fevo.2023.1313500)

[16] Asnicar F, Thomas AM, Passerini A, Waldron L, Segata N. 2023 Machine learning for microbiologists. *Nature Reviews Microbiology* **22**, 191–205. (10.1038/s41579-023-00984-1)

[17] Turchin P. 2001 Does population ecology have general laws?. *Oikos* **94**, 17–26. (10.1034/j.1600-0706.2001.11310.x)

[18] Balsa-Canto E, Alonso-del-Real J, Querol A. 2019 Mixed growth curve data do not suffice to fully characterize the dynamics of mixed cultures. *Proceedings of the National Academy of Sciences* **117**, 811–813. (10.1073/pnas.1916774117)

[19] Ram Y, Obolski U, Feldman MW, Berman J, Hadany L. 2019 Reply to Balsa-Canto et al.: Growth models are applicable to growth data, not to stationary-phase data. *Proceedings of the National Academy of Sciences* **117**, 814–815. (10.1073/pnas.1917758117)

[20] Picot A, Shibasaki S, Meacock OJ, Mitri S. 2023 Microbial interactions in theory and practice: when are measurements compatible with models?. *Current Opinion in Microbiology* **75**, 102354. (10.1016/j.mib.2023.102354)

[21] Hatton IA, Mazzarisi O, Altieri A, Smerlak M. 2024 Diversity begets stability: Sublinear growth and competitive coexistence across ecosystems. *Science* **383**. (10.1126/science.adg8488)

[22] Ishizawa H, Tashiro Y, Inoue D, Ike M, Futamata H. 2024 Learning beyond-pairwise interactions enables the bottom–up prediction of microbial community structure. *Proceedings of the National Academy of Sciences* **121**. (10.1073/pnas.2312396121)

[23] Cui W, Marsland R, Mehta P. 2024 Les Houches Lectures on Community Ecology: From Niche Theory to Statistical Mechanics. *arXiv*. (10.48550/arXiv.2403.05497)

[24] Miguel Trabajo T, Guex I, Dubey M, Sarton-Lohéac E, Todorov H, Richard X, Mazza C, van der Meer JR. 2024 Inferring bacterial interspecific interactions from microcolony growth expansion. *microLife* **5**. (10.1093/femsml/uqae020)

[25] Breiman L. 2001 Statistical Modeling: The Two Cultures. *Statistical Science* **16**, 199–215. (10.1214/ss/1009213725)

[26] Peleg M, Corradini MG. 2011 Microbial Growth Curves: What the Models Tell Us and What They Cannot. *Critical Reviews in Food Science and Nutrition* **51**, 917–945. (10.1080/10408398.2011.570463)

[27] Tang J, Riley WJ. 2021 Finding Liebig's law of the minimum. *Ecological Applications* **31**. (10.1002/eap.2458)

[28] Held NA, Manhart M. 2024 Are microbes colimited by multiple resources?. *Current Opinion in Microbiology* **80**, 102509. (10.1016/j.mib.2024.102509)

[29] Kronberger G, Burlacu B, Kommenda M, Winkler SM, Affenzeller M. 2025 *Symbolic Regression*. Boca Raton: Chapman and Hall, CRC Press. (10.1201/9781315166407)

[30] Course K, Nair PB. 2023 State estimation of a physical system with unknown governing equations. *Nature* **622**, 261–267. (10.1038/s41586-023-06574-8)

[31] Gaucel S, Keijzer M, Lutton E, Tonda A. 2014 Learning Dynamical Systems Using Standard Symbolic Regression. In *Genetic Programming* , p. 25–36. Springer Berlin Heidelberg. (10.1007/978-3-662-44303-3_3)

[32] Martin BT, Munch SB, Hein AM. 2018 Reverse-engineering ecological theory from data. *Proceedings of the Royal Society B: Biological Sciences* **285**, 20180422. (10.1098/rspb.2018.0422)

[33] Regalado D. 2021 Approximate Analytic Solution to the Three Species Lotka-Volterra Differential Equation Model. *Recoletos Multidisciplinary Research Journal* **9**, 123–128. (10.32871/rmrj.2109.02.09)

[34] Hsin J, Agarwal S, Thorpe A, Sentis L, Fridovich-Keil D. 2023 Symbolic Regression on Sparse and Noisy Data with Gaussian Processes. *arXiv*. (10.48550/arXiv.2309.11076)

[35] Lo Grasso A, Fort A, Mahdizadeh FF, Magnani A, Mocenni C. 2023 Generalized logistic model of bacterial growth. *Mathematical and Computer Modelling of Dynamical Systems* **29**, 169–185. (10.1080/13873954.2023.2236681)

[36] Angaroni F, Peruzzi A, Alvarenga EZ, Pinheiro F. 2024 Kinbiont: From time series to ecological and evolutionary responses in microbial systems. *BioRxiv*. (10.1101/2024.09.09.611847)

[37] Holling CS. 1959 The Components of Predation as Revealed by a Study of Small-Mammal Predation of the European Pine Sawfly. *The Canadian Entomologist* **91**, 293–320. (10.4039/ent91293-5)

[38] Michaelis L, Menten ML. 1913 Die Kinetik der Invertinwirkung. *Biochemische Zeitschrift* **49**, 333–369.

[39] Holling CS. 1959 Some Characteristics of Simple Types of Predation and Parasitism. *The Canadian Entomologist* **91**, 385–398. (10.4039/ent91385-7)

[40] Perrin E, Ghini V, Giovannini M, Di Patti F, Cardazzo B, Carraro L, Fagorzi C, Turano P, Fani R, Fondi M. 2020 Diauxie and co-utilization of carbon sources can coexist during bacterial growth in nutritionally complex environments. *Nature Communications* **11**. (10.1038/s41467-020-16872-8)

[41] Saito MA, Goepfert TJ, Ritt JT. 2008 Some thoughts on the concept of colimitation: Three definitions and the importance of bioavailability. *Limnology and Oceanography* **53**, 276–290. (10.4319/lo.2008.53.1.0276)

[42] Kremling A, Geiselmann J, Ropers D, de Jong H. 2018 An ensemble of mathematical models showing diauxic growth behaviour. *BMC Systems Biology* **12**. (10.1186/s12918-018-0604-8)

[43] Cranmer M. 2023 Interpretable Machine Learning for Science with PySR and SymbolicRegression.jl. *arXiv*. (10.48550/arXiv.2305.01582)

[44] Meroz N, Livny T, Friedman J. 2024 Quantifying microbial interactions: concepts, caveats, and applications. *Current Opinion in Microbiology* **80**, 102511. (10.1016/j.mib.2024.102511)

[45] Coyte KZ, Rao C, Rakoff-Nahoum S, Foster KR. 2021 Ecological rules for the assembly of microbiome communities. *PLOS Biology* **19**, e3001116. (10.1371/journal.pbio.3001116)

[46] Lee H, Bloxham B, Gore J. 2023 Resource competition can explain simplicity in microbial community assembly. *Proceedings of the National Academy of Sciences* **120**. (10.1073/pnas.2212113120)

[47] Ho PY, Nguyen TH, Sanchez JM, DeFelice BC, Huang KC. 2024 Resource competition predicts assembly of gut bacterial communities in vitro. *Nature Microbiology* **9**, 1036–1048. (10.1038/s41564-024-01625-w)

[48] Pacheco AR, Osborne ML, Segrè D. 2021 Non-additive microbial community responses to environmental complexity. *Nature Communications* **12**. (10.1038/s41467-021-22426-3)

[49] Litchman E, Edwards KF, Klausmeier CA. 2015 Microbial resource utilization traits and trade-offs: implications for community structure, functioning, and biogeochemical impacts at present and in the future. *Frontiers in Microbiology* **06**. (10.3389/fmicb.2015.00254)

[50] Ho PY, Good BH, Huang KC. 2022 Competition for fluctuating resources reproduces statistics of species abundance over time across wide-ranging microbiotas. *eLife* **11**. (10.7554/elife.75168)

[51] Morozov A, Petrovskii S. 2013 Feeding on Multiple Sources: Towards a Universal Parameterization of the Functional Response of a Generalist Predator Allowing for Switching. *PLoS ONE* **8**, e74586. (10.1371/journal.pone.0074586)

[52] Meszéna G, Gyllenberg M, Pásztor L, Metz JAJ. 2006 Competitive exclusion and limiting similarity: A unified theory. *Theoretical Population Biology* **69**, 68–87. (10.1016/j.tpb.2005.07.001)

[53] Hutchinson GE. 1961 The Paradox of the Plankton. *The American Naturalist* **95**, 137–145. (10.1086/282171)

[54] MacArthur R, Levins R. 1967 The Limiting Similarity, Convergence, and Divergence of Coexisting Species. *The American Naturalist* **101**, 377–385. (10.1086/282505)

[55] Posfai A, Taillefumier T, Wingreen NS. 2017 Metabolic Trade-Offs Promote Diversity in a Model Ecosystem. *Physical Review Letters* **118**. (10.1103/physrevlett.118.028103)

[56] Chesson P. 2018 Updates on mechanisms of maintenance of species diversity. *Journal of Ecology* **106**, 1773–1794. (10.1111/1365-2745.13035)

[57] Badali M, Zilman A. 2020 Effects of niche overlap on coexistence, fixation and invasion in a population of two interacting species. *Royal Society Open Science* **7**, 192181. (10.1098/rsos.192181)

[58] Pastore AI, Barabás G, Bimler MD, Mayfield MM, Miller TE. 2021 The evolution of niche overlap and competitive differences. *Nature Ecology & Evolution* **5**, 330–337. (10.1038/s41559-020-01383-y)

[59] Rozas Garcia E, Crumpton MJ, Galla T. 2023 Niche overlap and Hopfield-like interactions in generalized random Lotka-Volterra systems. *Physical Review E* **108**. (10.1103/physreve.108.034120)