Data-driven symbolic regression reveals microbial growth laws

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

2025-11-12

Abstract

Connecting mathematical models with empirically measured microbial growth has remained challenging, as numerous competing models based on different theoretical approaches can fit the observations. Therefore, we develop a method to automatically propose growth models from microbial data alone. We validate it using an available dataset of *E. coli* grown on known resources, and study sixteen species across various concentrations of a complex medium. Even if we cannot track resources, we can indirectly observe them by exploiting the area under the growth curve (AUC). Indeed, Random Forest machine learning regressor predicts growth rates more accurately using the AUC than instantaneous population size. Then, we use the inherently interpretable approach of symbolic regression to infer explicit dynamical models directly from growth data. For a dataset on two metabolically non-substitutable resources, the algorithm indeed proposed two resources, linked by a multiplicative-effect Monod dynamics implying non-substitutability. Our data showed that, overall, models based on a Monod dynamics provide the best balance between data fit, model parsimony, and biological relevance. Thus, we propose an approach to read data-driven growth laws off from microbial batch cultures. Our study demonstrates that the AUC measured in routine experiments informs us on the effective resource dynamics governing microbial growth.

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. So far, 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 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 fully interpretable alternative to black-box machine learning is symbolic regression, a method providing explicit mathematical expressions directly from data [29], which can fruitfully serve as a hypothesis generation approach [30]. Symbolic regression has yielded impressive results in recovering and expanding known laws in physics [31] and, in the context of microbial ecology, in reverse-engineering Lotka-Volterra and logistic-based models [32, 33, 34, 35]. But empirically indistinguishable models [27, 28] imply that we cannot unambiguously infer any unique, true, generative model based on observations alone. Instead, a data-driven approach should ask which models suffice to reproduce the observed behaviour, so as to propose data-driven null-hypotheses compatible with observations.

The main strength of symbolic regression is the unbiased discovery of relationships between predictors and a dependent variable, done by directly searching the space of mathematical expressions [29]. At its basic usage, the algorithm requires no assumptions or any other inputs than the data itself. While computationally intensive, computing power advancements in the past decades have made this methodology viable and increasingly popular across a multitude of scientific disciplines. However, the true potential of the method is in the ability to specify various constraints for the mathematical expressions by defining allowed mathematical operations and partial functional expressions. The latter constraint is extremely important for utilisation of the known mathematical laws and general domain knowledge, as showcased in this study.

Here, we make minimal theoretical assumptions to better understand growth data obtained for sixteen bacterial species in R2A medium (figure 1). Besides key nonlinearities, assuming a constant-rate per capita resource consumption suffices to improve a black-box Random Forest regression. Next, we crack this

algorithmic black-box and 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 may offer a useful access to the latent resource space. Then, we use theoretical insights to inform the symbolic regression algorithm by specifying theory-based mathematical expression constraints and favour biologically interpretable models. We validate our approach using available data on *E. coli* with known chemical resources [14]. The algorithm helps us to reveal that bacterial growth widely follows a Monod dynamics in our experiments, with either one or two effective underlying resources.

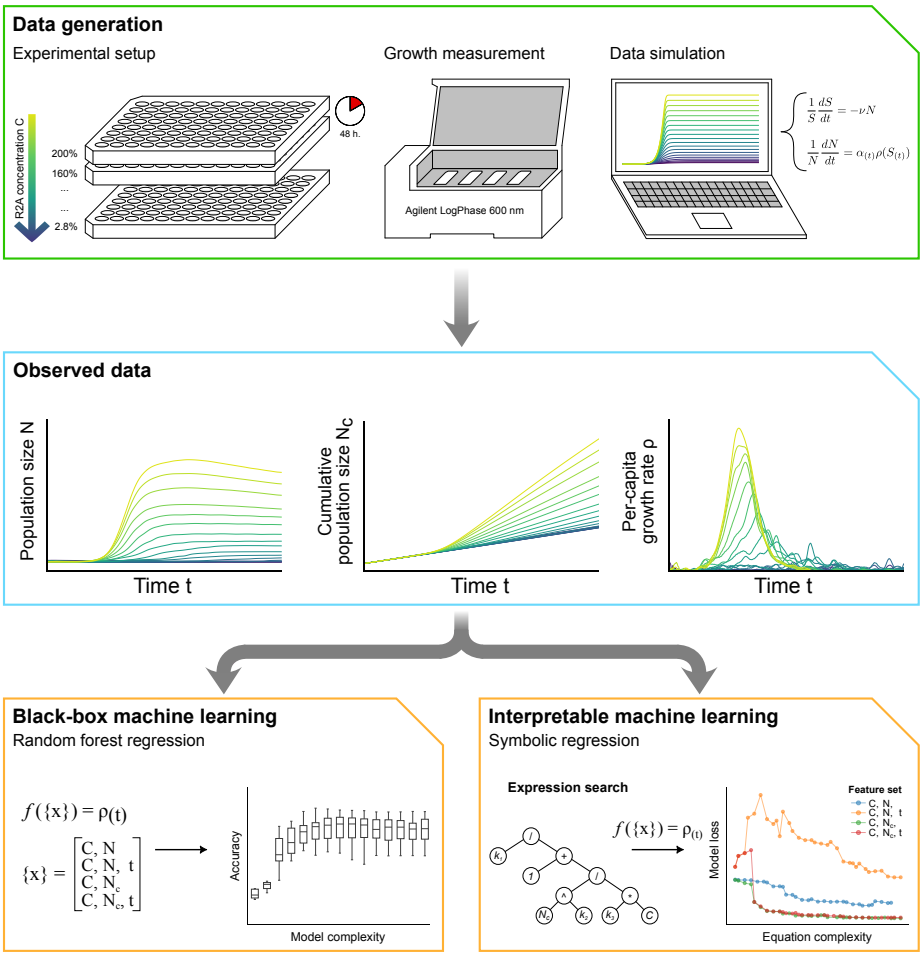

Figure 1: Workflow for the production and analysis of the datasets.

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 S1. All were able to grow in Reasoner's 2A medium and the following culture protocol yielded our data, summarised in Figure S1.

Solution M9 contained 11.28 g of M9 minimal medium salts (MP Biomedicals, lot `142662`) in 1.0 L of Milli-Q purified water. The culture medium (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 M9. We prepared a range of 20 different concentrations C of culture medium, separated by an incremental factor 0.8, from 200 % to 2.88 %, by successive dilutions into autoclaved solution M9. Bacterial pre-cultures were run by inoculating each species into 6.0 mL of culture medium (100 %), then incubating for 48 h in aerobic conditions at 30 °C under agitation at 70 rpm, and finally diluted into autoclaved solution M9 by a factor 10^{-4}.

Monocultures were run in four batch replicates by inoculating 110 µL of each culture medium concentration C with 20 µL of pre-culture 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.

2.2 Theoretical methods and models

Consider a bacterial population of size N growing in a substrate of concentration S. The law of mass action assumes that it uptakes resources at a constant-rate ν relative to its size:

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

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. The variables S, N and N_c are time-dependent. This constant-rate model for per capita resource consumption readily extends to several independent resources and naturally gives a paramount role to the cumulative population gain N_c. In laboratory terms, the cumulative population gain N_c corresponds to the area under the (growth) curve (AUC) sometimes reported in microbiology [15]. Conversely, the predictive value of N_c in explaining empirical growth data supports the constant-rate per capita 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]:

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

where ρ is the time-independent and potentially resource-dependent growth function. How physiological adjustment delays bacterial cells' growth dynamics is usually represented by similar mathematical forms [6, 7, 36, 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:

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

Monod showed that the maximum $\rho_{\max}$ of the empirical per capita growth rate ρ_{obs} was an increasing then saturating function of the resource concentration [10]. The demonstration concerns only the maximum per capita growth rate $\rho_{\max}(S)$ in a bacterial growth experiment [10, 37]. However, this type II functional response pattern [38] is usually modelled using a hyperbolic function [39, 10, 40] and integrated as a function of time into a dynamical model [24]:

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

where K is the substrate concentration inducing half-maximum response. Without further assumptions, this functional form cannot model multi-resource growth [10], which may be complicated by non-mutually-exclusive resources [41], colimitation [42, 14], and sequential preference [43]. 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 with multiplicative or additive effects [42, 13, 28, 14]:

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

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

The multiplicative-effect multi-resource model combines Eq. (1), Eq. (2) and Eq. (5). It assumes that growth requires all resources, which are thus biologically not substitutable. The additive-effect multi-resource model combines Eq. (1), Eq. (2) and Eq. (6). It assumes that any resource allows for some growth, which implies substitutable resources. 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, general support for either model over the other [27, 28], so that we have no prior hypothesis in our experimental setting involving several species and a rich medium.

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 [44]. Code is available at: `https://github.com/anthony-sun/SR-Monod`. Since the batch replicates were extremely similar, we averaged population sizes across the four replicates, for each time point. This has no impact on the study but facilitates computation. The supplementary information details the protocol used for simulated data and expands on the symbolic regression methods.

For Random Forest regression, the aim was to determine which observed variables would best predict growth rate and thus better reflected a latent resource dynamics. We split each dataset into a training set (75 %) and a testing set (25 %) comprising every fourth medium concentration 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. We used 100 estimators, with the coefficient of determination R^2 as a performance score, and different values for maximum tree depth.

For symbolic regression, the aim was to find explicit mathematical growth models which we could then interpret, in light of biological knowledge, as revealing latent resources. 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 using the default method (with crossover and mutation) and parameters in reference [44]. 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 used the following expression templates constraining the search for mathematical expressions:

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

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

where the algorithm looks for the undefined parameters $\hat{q}$ and $\hat{m}$, and for the undefined functions $\hat{\rho}$ and $\hat{h}$. Template (7) assumed Eq. (2) and Eq. (3) by imposing the search for adjustment parameters q and m. Template (8) inspired by Eq. (1) additionally assumes that the medium's relative concentration C has a multiplicative impact on the growth's resource-dependency.

3 Results

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

We hypothesised that it is reasonable to approximate resource consumption as a constant-per-capita-rate process. Following Eq. (1), this would give rise to

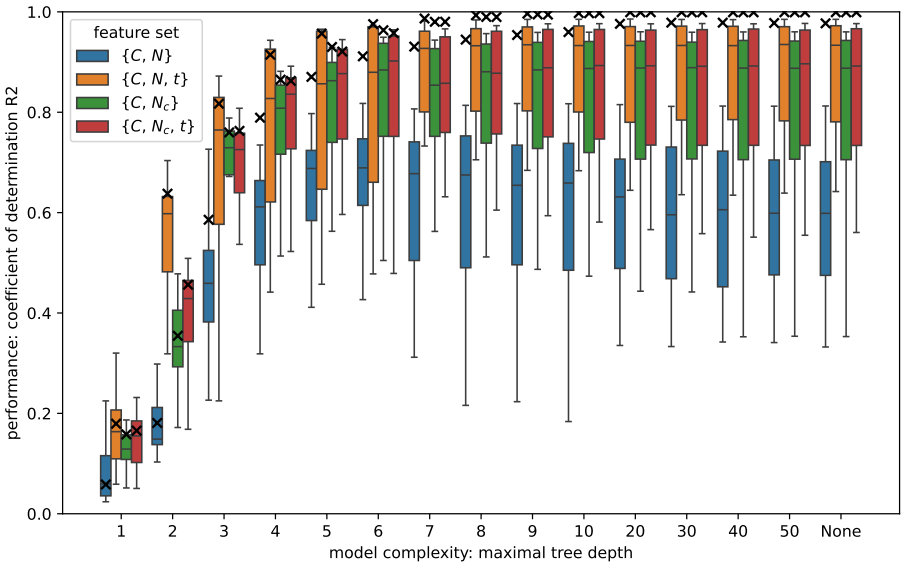

Figure 2: Performance of a black-box Random Forest regressor in predicting the per capita growth rate from different feature sets in experimental data, given by the coefficient of determination R^2. Crosses show the median train score across species. The cumulative population gain predictor N_c outperforms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

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, instantaneous population size N. To test that, we compared the performance of a black-box Random Forest regressor using either N or N_c. This cannot *prove* that resource consumption has a constant per capita rate but it can show that this is a reasonable approximation.

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 $\{C, N_c\}$). 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 1). 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, with a gap between train and test scores. Models relying instead on cumulative population gain N_c reduced this overfit. Both this overfit and its reduction support a resource-consumer framework against a logistic-based model (figure S4).

Our 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

Table 1: 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)

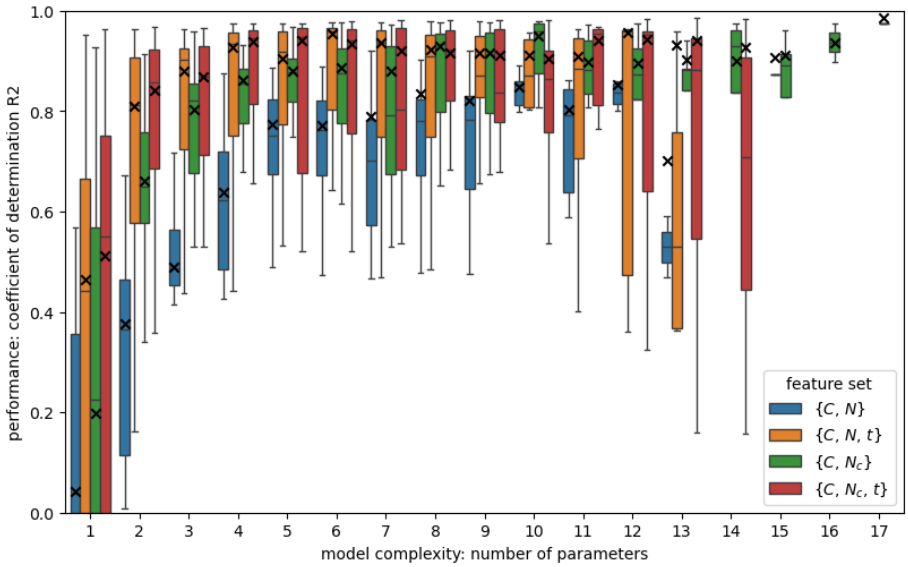

Figure 3: Performance of the unconstrained symbolic regressor in predicting the per capita growth rate from different feature sets in experimental data, given by the coefficient of determination R^2. Crosses show the median train score across species. The cumulative population gain predictor N_c outperforms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

per capita resource consumption in our experiments. The feature sets $\{C, N, t\}$, $\{C, N_c\}$ and $\{C, N_c, t\}$ yielded a similar improvement, with $\{C, N, t\}$ performing best. Thus, it seemed that N_c effectively offered a parsimonious, unidimensional approximation of the bidimensional space (N, t). But at this point, it remained unclear which feature combination best reflected the 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

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 [44]. 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 per capita consumption model.

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. Feature set $\{C, N_c\}$ yielded the best overall performance (figure 2, 3), especially with more complex models. This 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) implying 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 and provided mathematically explicit models. 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

So far, we had found solid support for the constant-rate assumption for per capita resource consumption, as formulated in Eq. (1). We now hypothesised that a symbolic regression algorithm could infer a biologically interpretable growth function, when guided by theory-based expectations. Thus, we ran three rounds of symbolic regression: one with minimal biological assumptions (unconstrained), and two informed by supplementary theoretical assumptions, using expression templates given first by Eq. (7) then additionally by Eq. (8). We validated our approach using available batch culture data where *E. coli* is grown on glucose as a carbon source and ammonium as a nitrogen source [14]. Our algorithms recovered a bidimensional Monod model, and the best-fitting, biologically interpretable model was compatible with the two inferred resources' metabolic non-substitutability (figure S5).

A data-driven approach cannot prove nor disprove any mathematical expression. Instead, the algorithms proposed a cohort of models compatible with the data, sufficient to approximate empirical observations. This included 1068 models learnt from experimental datasets in R2A. Then, to avoid focusing on well-fitting models with no biological interpretability, we manually annotated models compatible with theoretical expectations. Overall, we found few of the proposed models to be interpretable in light of the literature in microbiology (figure 4).

Table 2 summarises the best-fitting models that remained biologically interpretable in experiments. In many species, the algorithm initially proposed 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

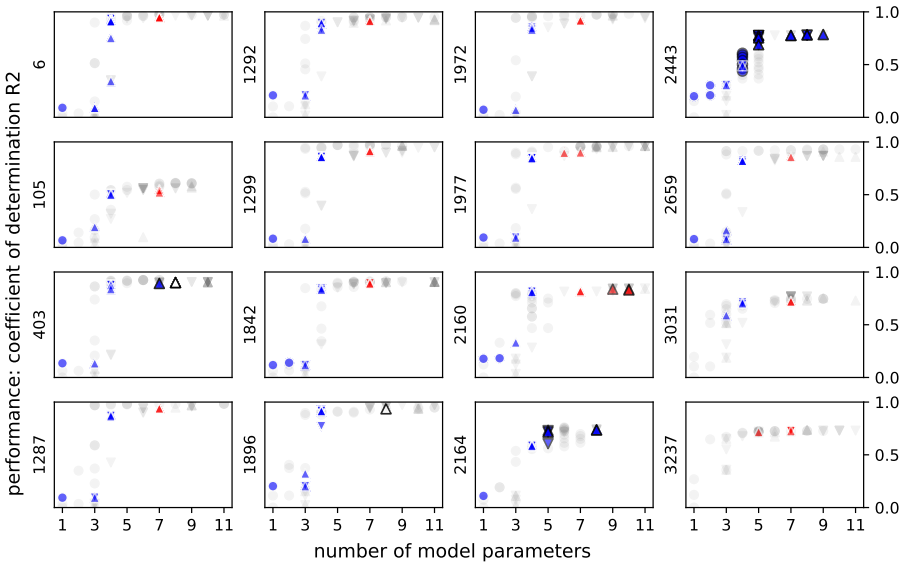

Figure 4: Fitting performance of models proposed by the symbolic regressor for R2A data. Blue: linear-based model; red: Monod-based model; white: both; gray: not biologically interpretable. White edge: unidimensional resource; black edge: multidimensional resource. Circle: no template; triangle pointing down: Template (7); triangle pointing up: Template (8).

for one of 16 experimental species in R2A medium, suggesting that experimental conditions did not cover the saturating part 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 multidimensional resource, with constant-per-capita-rate or Monod-like components.

Table 2: Formula, interpretation and inferred resource dimensionality of the best-fitting, biologically interpretable growth models for ρ in Eq. (2) and Eq. (4).

model for $\rho(S)$	interpretation	res. dim.	number of experimental species
$g_{\max} S$	linear	1	1 sp.: `2443`
$g_{\max} \frac{S}{K+S}$	Monod	1	11 sp.
$g_{\max} S_1 \times \frac{S_2}{K+S_2}$	linear $\times$ Monod	2	2 sp.: `403`, `1896`
$g_{\max} \frac{S_1}{K_1+S_1} \times \frac{S_2}{K_2+S_2}$	Monod $\times$ Monod	2	1 sp.: `2160`
more complex	linear and Monod	≥ 2	1 sp.: `2164`

4 Discussion

4.1 Biological implications

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). This implies that the study of microbial growth should systematically report the area under the curve (AUC) computed from growth curves [15]. Then, we used symbolic regression to learn explicit growth models from population size data, without relying on prior theory. We validated our approach using an available dataset in which resources were known [14]. While fits to the data were excellent, an unconstrained use of the method did not ensure biological relevance, thus emphasising the need of biology bespoke usage of the approach. Therefore, we combined this approach with the following theoretical expectations based on the literature, spelt out here: biological constants are positive; expressions combine through addition, multiplication, division and power; biological mechanisms involve only integer powers of the medium concentration C; the law of mass action of Eq. (1) applies, *i.e.,* N_c is an important state variable; the effective per capita growth rate follows Eq. (2) as the product of an adjustment function and a resource-dependent part.

We found that the dynamical resource-consumer model based on Monod's formulation in Eq. (4), which we never assumed, best trades off among fit to data, symbolic simplicity, and biological interpretability, which extends previous non-dynamical findings [37]. 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 per capita rate. Overall, they highlight the potential of symbolic regression as a data-driven hypothesis generation approach while emphasising the need to integrate existing theory and biological knowledge as informative priors.

Besides, our method provides an access to the effective resource dimensionality experienced by a growing microbial population, shown in Table 2. The effective resource dimensionality represents the minimal number of effective, independent resources theoretically required to reproduce the observed population dynamics. Among multidimensional resource models, multiplicative effects indicate substitutable colimiting resources [28], whereas additive effects suggest that the resources are not substitutable.

Modelled effective resources are defined purely from their impact on the growth dynamics. Departing from a chemical view allows us to translate population dynamics into effective ecological niches, which can be generally disconnected from chemical resources [45]. Thus, our approach connects microbiology experiments with a rich literature on multidimensional niches [46, 47, 48] and their overlap, which determine species coexistence in communities [49, 50, 51, 52]. Future research could modulate the concentration of single chemical resources in a culture medium, to allow for a high-definition view of the growth process in a complex environment, and discuss the interpretation of symbolic regression outputs in light of chemical evidence.

4.2 Methodological impact of symbolic regression in microbial ecology

Despite its success in physics [31], symbolic regression has mainly been constrained to search logistic-based models in microbiology [32, 33, 34, 35], thus overlooking two limitations. First, such models fail [17, 18, 19, 20, 21, 22] to account for key resource dynamics mediating the interactions [53, 24] that explain how bacterial communities assemble [54, 55, 56], grow, organise and change [57], through competition and resource utilisation traits [58, 59]. 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 constraints on its functional form. It is important to note that symbolic regression cannot decisively prove that any model is the true, generative model, nor disprove that alternative models could be generative. This method simply proposes approximations sufficient to reproduce data, but no interpretation. Notably, our proposed models involve exclusively observable variables. Only scientific interpretation in light of pre-existing knowledge expressed by Eq. (1) allows for a translation into a latent resource space S.

Since symbolic regression does not provide automatic scientific interpretation, it is useful in generating, but not testing, hypotheses [30]. Indeed, symbolic regression outcomes cannot reject alternative hypotheses, which may still receive other empirical support, *e.g.,* through knowledge about specific metabolites involved. Again, future research should further investigate the link between effective resources inferred by symbolic regression and specific chemicals.

Our study demonstrates that symbolic regression offers a promising, empirically viable, and scalable way to analyse routine experiments in microbiology to infer null-models for growth laws and effective resource dimensionalities. It is reasonable to assume a constant-rate per capita resource consumption and, based on this assumption, we expect the dynamical Monod model to capture the growth behaviour of most species in batch monocultures. Microbiologists can now assess this assumption using experimental data reporting the area under growth curves, and knowledge of resource consumption dynamics could inform our predictions and control of microbial community interactions.

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 Bělehdrádek-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] Radwan YA, Kronberger G, Winkler S. 2025 A Comparison of Recent Algorithms for Symbolic Regression to Genetic Programming. In Quesada-Arencibia A, Affenzeller M, Moreno-Díaz R, editors, *Computer Aided Systems Theory – EUROCAST 2024* , number 15172 in Lecture Notes in Computer Science pp. 157–171. Cham: Springer Nature Switzerland. (10.1007/978-3-031-82949-9$_1$5)

[31] 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)

[32] 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)

[33] 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)

[34] 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)

[35] 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)

[36] 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)

[37] Angaroni F, Peruzzi A, Alvarenga EZ, Pinheiro F. 2025 Translating microbial kinetics into quantitative responses and testable hypotheses using Kinbiont. *Nature Communications* **16**. (10.1038/s41467-025-61592-6)

[38] 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)

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

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

[41] 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)

[42] 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)

[43] 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)

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

[45] 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)

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

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

[48] 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)

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

[50] 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)

[51] 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)

[52] 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)

[53] 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)

[54] 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)

[55] 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)

[56] 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)

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

[58] 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)

[59] 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)