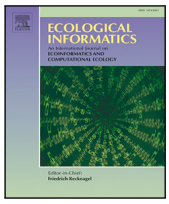
Ecological Informatics 99 (2026) 103988

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf



Symbolic regression for empirically realistic population dynamic time series

Cheyenne N. Jarman [a,*], Taal Levi [b], Mark Novak [a]

[a] Department of Integrative Biology, Oregon State University, Corvallis, OR, USA
[b] Department of Fisheries, Wildlife, and Conservation Sciences, Oregon State University, Corvallis, OR, USA

ARTICLE INFO

Dataset link: https://zenodo.org/records/19339107, https://github.com/cheyennejarman/SymbReg_KelpForest_Simulations.git

Keywords:
Population dynamics,
Scientific machine learning,
Time-series analysis,
Explanatory models,
Theoretical ecology,
Giant Kelp forests

ABSTRACT

Applications of machine learning in ecology are rapidly expanding. Symbolic regression is gaining particular attention for its success in reverse-engineering human-readable explanatory population models, including the logistic growth and Lotka–Volterra equations, from simulated and laboratory-based population time series. However, field-based populations often lack the characteristics of the idealized time series used in prior assessments. We evaluated the utility of symbolic regression for such time series by quantifying its success for synthetic data varying in sampling density, population cycle asymmetry, process noise, and the erroneous consideration of spurious variables. We further compared two data preprocessing options for estimating population growth rates (i.e., a discrete-time vs. continuous-time approach), and four evaluation workflows for selecting equations (two subjective, where gains in performance as a function of model complexity are assessed visually, and two objective, where quantitative metrics of performance determine model selection). Results indicate that a trade-off between sampling density and process noise primarily drives equation and variable recovery. Symbolic regression failed to recover the underlying equation at sampling densities below 25 points per cycle; however, at higher sampling densities, process noise did increase equation recovery rates. Importantly, although the true model was frequently recovered at sampling densities of 25 or more points per cycle, it was not consistently selected by the evaluation workflows. This discrepancy highlights a need for more robust post-algorithm selection criteria to identify the focal equation among competing candidates.

1. Introduction

Ecologists are motivated by the goal of describing patterns and process through observation, experimentation, and mathematical modeling. Mathematical models are particularly important for abstracting high-dimensional systems into equations that bring focus to key variables and mechanisms. For over a century, such models have been constructed based on *a priori* conceptualizations. For example, the variables, parameters, and formulations of the logistic growth (Verhulst, 1838) and Lotka–Volterra (Volterra, 1926; Lotka, 1925) models, as well as the myriad other population dynamics models descended from these, were specified by logic, intuition, and first principles (Hilborn and Mangel, 1997). Unfortunately, although this approach may work for well-studied or relatively simple systems, it is susceptible to biases stemming from preconceived notions, paradigms, and scientific inertia (Bonnaffé et al., 2021; Ellner et al., 2002; Kuhn, 1970; Novak et al., 2025). In recent years, several approaches have therefore begun to take a more data-driven approach to dynamical model building (Solomatine and Ostfeld, 2008). Tools like neural networks (Bonnaffé et al., 2021), empirical dynamical modeling (Sugihara and May, 1990; Munch et al., 2020), and symbolic regression (Bongard and Lipson, 2007) in particular are increasingly being applied to time series to infer the data-generating processes from the data, rather than specifying a particular model or set of models *a priori*.

As a form of scientific machine learning (Baker et al., 2019), symbolic regression stands out among these data-driven approaches for its goal of providing human-readable, mechanistically interpretable equations in the spirit of the classical equations of theoretical ecology (Gerwin, 1974; Langley, 1977). Given data on a response variable and a set of putative predictor variables, symbolic regression algorithms use a process akin to biological evolution to evolve generations of equations through mutation, the recombination of parental equation components, and the selection and transmission of equations based on their fit to the data. (See the *Supplementary Materials* for an overview.) The first ecological application of symbolic regression to population time series appears to have been Bongard and Lipson (2007) who found it to successfully recover the two-species Lotka–Volterra competition equations even when the simulated time series were subject to observation noise. They also found it able to produce similar equations and mechanistic interpretations to previous *a priori* models of lynx-hare interactions when applied to the classic Hudson Bay population time series. Gaucel

* Corresponding author.
E-mail address: jarmanc@oregonstate.edu (C.N. Jarman).

https://doi.org/10.1016/j.ecoinf.2026.103988
Received 30 March 2026; Received in revised form 11 August 2026; Accepted 11 August 2026
Available online 27 August 2026

et al. (2014) extended these findings to demonstrate symbolic regression's ability to recover the Lotka–Volterra predator–prey equations from simulated time series subject to observation noise under sampling strategies in which the time elapsed between observations was either regular or irregular. This was followed by Martin et al. (2018) who demonstrated symbolic regression's success in reverse-engineering the logistic growth equation, the Lotka–Volterra predator–prey equations, and a stage-structured population model from empirical time series of single and paired predator–prey laboratory populations. Finally, Chen et al. (2019) assessed the method's ability to recover and reverse-engineer the population growth and nonlinear functional response equations of predator–prey populations from simulated and laboratory time series, focusing specifically on the data's dynamical informativeness through a contrast of time series exhibiting a stable limit cycle versus a transient approach to a fixed-point equilibrium. Although symbolic regression was only successful at recovering the generative equation in the cycling case, they concluded that its failure in the transient case was a limitation of the data rather than of symbolic regression itself (Chen et al., 2019). Altogether, the current literature therefore provides a unanimous endorsement of symbolic regression for learning population dynamic models from time series data.

Despite prior demonstrations of its success, symbolic regression's ability to reverse-engineer the equations of more complex ecological time series with field-relevant sampling characteristics remains unknown. While symbolic regression offers a powerful framework for recovering mechanistic models, it has yet to be fully operationalized for generating insights from empirical systems. This lag in broad-scale application persists because we lack a clear understanding of how it handles the reality of field data compared to idealized simulations. In this regard it is particularly unclear whether the sampling densities of typical population time series are sufficient for symbolic regression to succeed. Prior assessments have assumed sampling intervals many times shorter than typical population cycle period lengths or generation times (e.g., Inchausti and Halley, 2001). Further knowledge gaps remain regarding time-series preprocessing practices (i.e., the use of time series interpolation methods) and how process noise rather than observation noise, the symmetry of population cycles, and the erroneous consideration of spurious predictor variables (previous benchmarks were restricted to *a priori* known drivers) affects the method's reliability. Finally, there has been little consideration of the alternative methods by which the suite of equations returned by symbolic regression can be assessed to identify and select the correct or "best" equation.

By generating empirically relevant time series from a mechanistic model of giant kelp populations, we evaluated symbolic regression's success in regards to each of these factors in six case studies, with each subjected to five different sampling densities. We included a contrast of four alternative evaluation workflows for selecting candidate equations from along the Pareto frontier, the set of equations representing the optimal trade-off between complexity and explanatory power. We also examined the diversity of returned equations along and behind the frontier to gain insight into the effectiveness of each workflow. We structure our manuscript by introducing our generative population model, further motivating and describing the case studies, and detailing our implementation of symbolic regression and means of quantifying its rates of success for each workflow. Overall, we identified significant limitations under certain field-relevant conditions, particularly in regards to post-algorithm selection criteria for identifying the true model among the equations returned by symbolic regression.

2. Methods

2.1. The generative model

We used the Bence-Nisbet delay-differential model for giant kelp (*Macrocystis pyrifera*) to generate population time series (Bence and Nisbet, 1989). This model qualitatively reproduces the cycles observed in empirical giant kelp populations by encapsulating an effect of adult kelp on the recruitment of juvenile kelp in competing for space. The model describes the population growth rate of adult kelp A at time t as

$$\frac{dA(t)}{dt} = se^{-\alpha\tau}[1 - a_A A(t-\tau)]_+ - mA(t), \quad (1)$$

where s represents the juvenile settlement rate, α and m represent the juvenile and adult per capita mortality rates, a_A represents the proportion of space occupied by an adult individual, and τ represents the temporal lag of juveniles maturing to adults (i.e., juveniles are represented implicitly by a time-delayed effect of adult population size). The operator $[x]_+ = x$ if $x \geq 0$ else 0 provides an abrupt stop in growth once no space is available. We modified this model to include the possibility of multiplicative process noise in juvenile settlement and adult mortality (i.e., $s \rightarrow s\exp[\epsilon_1]$ and $m \rightarrow m\exp[\epsilon_2]$ with $\epsilon_i \sim \mathcal{N}(0, \sigma_i)$). Previous studies have only considered unbiased observation noise (i.e., measurement error), the effect of which diminishes with increased sampling density and time-series replication, while neglecting process noise (i.e., the inherent stochasticity of population dynamic processes). Notably, although process noise may increase the challenge of reverse engineering equations just like observation noise, it may also make time series more informative of their underlying processes by pushing systems to explore a larger space of variable states and conditions (Boettiger, 2018).

We chose the Bence-Nisbet model for several additional reasons as well. First, because it can generate both symmetrical and asymmetrical population cycles in a biologically justified manner. Many real-world populations exhibit asymmetric cycles (e.g., a rapid increase followed by a slow decay Turchin, 2003), yet with the exception of Chen et al. (2019), all prior studies have considered only symmetrically cycling populations. This may be important because the combination of asymmetry and low sampling density could bias estimated rates of population change due to the higher chance of capturing the longer, slower phase, particularly under sampling regimes that are regular (i.e., periodic), as most field-based time series are. Second, because it is more complex than previously used generative models by virtue of a time-delayed variable and use of exponential and $[\ ]_+$ operators. Its inclusion of a time-delayed variable was of particular interest because (*i*) many population time series entail time-delayed processes arising from higher-dimensional multi-species and age- or stage-structured processes (Turchin, 2003; de Roos and Persson, 2003), (*ii*) several time-series methods (e.g., empirical dynamical modeling) rely on the representation of multi-dimensional systems using time-delay embedding (i.e., Takens' theorem Takens, 1981), and (*iii*) it provided a means to introduce non-trivial spurious variables in our assessments; all prior studies provided their symbolic regression algorithms only with variables known to influence dynamics, which is an unlikely scenario in field contexts.

2.2. Case studies

2.2.1. Time series generation

We used the model to generate time series for six case studies, evaluating both a discrete-time and a continuous-time approach to calculating the per capita growth rate from symmetrically and asymmetrically cycling deterministic time series (cases 1–4; Fig. 1 top four rows) and a continuous-time approach from asymmetrically cycling stochastic time series under low and high levels of process noise (cases 5–6; Fig. 1 bottom two rows). These cases were designed to isolate specific empirical data-relevant challenges (e.g., how often data points can be collected, a modeler's choice between discrete- and continuous-time growth rate estimations, whether a population exhibits symmetric or asymmetric cycles, and the inclusion of process noise). The juvenile maturation rate was set to $\tau = 2$ following Bence and Nisbet (1989) for all case studies. Most other parameter values were specific to

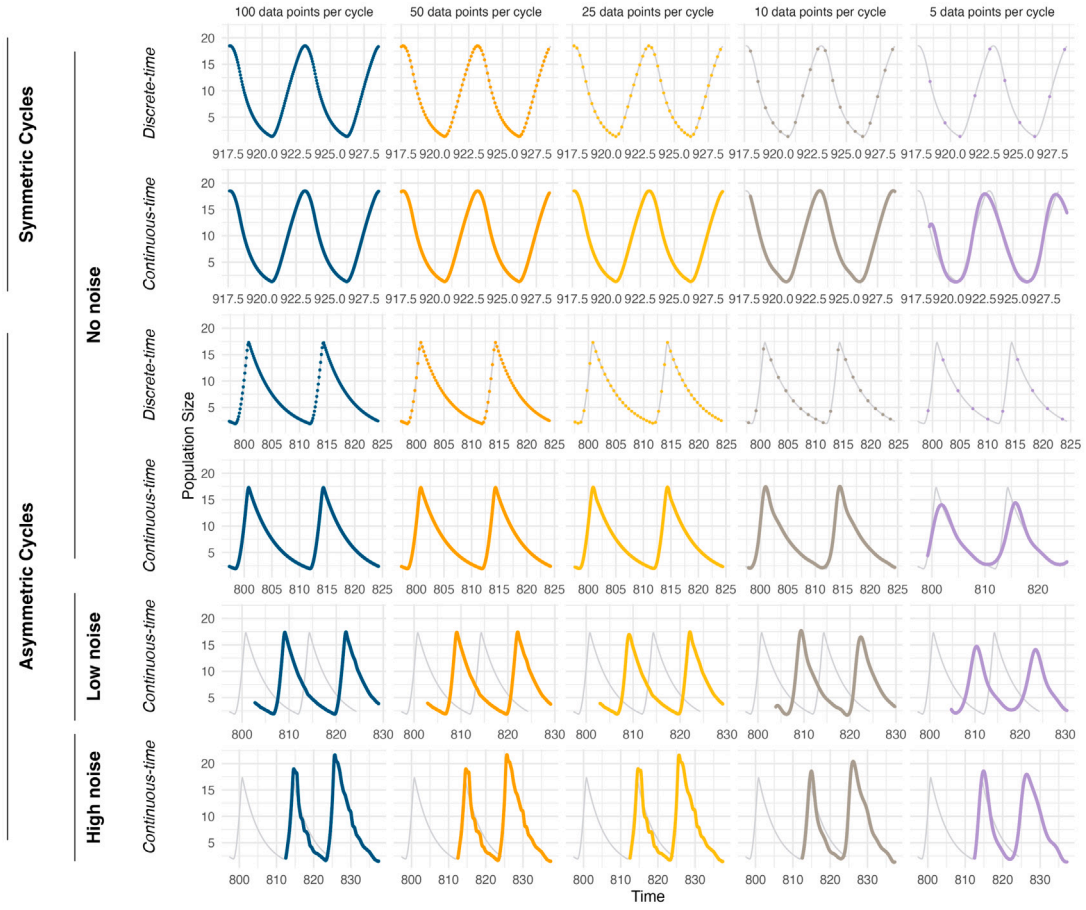


Fig. 1. Time series illustrating our six case studies and the consequences of varying sampling density within each. A total of 15 cycles were used in our analyses, but only two are plotted for visualization. The light gray lines show the dynamics of the instantaneous deterministic Bence-Nisbet model for reference. The overlaid colored points show the (downsampled and interpolated) data from which the per capita growth rates and (lagged) putative predictor variables were obtained and provided to symbolic regression. Per capita growth rates were calculated using a discrete- or continuous-time preprocessing approach (see main text).

each case study (Tables S.1-S.2). Each time series was simulated for 1000 time steps, from which the last 15 complete population cycles were retained. The cycle period lengths, as determined by Fourier analysis (see Supplemental Materials), were 5.5 and 13.5 time steps in the symmetric and asymmetric deterministic cases, and averaged 13.2 and 12.5 time steps for the symmetric stochastic cases under low and high process noise, respectively. The time series thereby excluded transient dynamics and were of a length representative of long-term empirical datasets. To generate the actual time series that we provided to symbolic regression, we downsampled each case study's time series at sampling densities of 100, 50, 25, 10, or 5 time points per cycle. Although not reported, prior studies used synthetic time series with sampling densities of hundreds of time points per cycle and laboratory time series with sampling densities no lower than approximately 12 time points per cycle.

2.2.2. Response and predictor variables

Each downsampled time series was used to generate the response and predictor variables that we provided to symbolic regression. As the response variable, we used the per capita growth rate, estimated differently depending on the use of a discrete- or continuous-time approach (Fig. 2a). For the discrete-time approach we calculated the per capita growth rate using the log-ratio of successive abundances, $\ln\left(A_{t+\Delta t}/A_t\right)/\Delta t$, as this converges on the instantaneous per capita growth rate $\frac{1}{A}\frac{dA}{dt}$ as Δt decreases (see *Supplementary Materials*). For the continuous-time approach we calculated the first derivative of the cubic spline interpolated log-transformed abundances (using SciPy, Virtanen et al., 2020) since $\frac{d\ln(A)}{dt} = \frac{1}{A}\frac{dA}{dt}$. Our comparison of these two approaches was motivated by prior simulation- and laboratory-based studies that have used continuous-time preprocessing, relying on sufficiently high sampling densities to avoid issues associated with alternative interpolation techniques. Our use of log-transformed population size differs from prior studies as it avoided the interpolation of negative population sizes to which the use of natural-scale population sizes is susceptible when sampling density is low.

All prior studies of ecological applications of symbolic regression to population time series provided their algorithms with state variables that were known to pertain directly to the focal dynamic. Similarly, we conducted a preliminary study to ensure that the symbolic regression algorithm we used could recover the generative equation when only $A(t)$ and $A(t-2)$ were considered as putative predictors (see the *Supplementary Materials*). However, in empirical contexts the true causal variables are typically not known. Our analyses therefore also considered the extraneous predictor variables $A(t-1)$ and $A(t-3)$. Note that only $A(t)$ and $A(t-2)$ appear in the generating equation (Eq. (1)), making $A(t-1)$ and $A(t-3)$ autocorrelated but spurious determinants of the dynamics. For both the discrete-time and continuous-time approaches, the lagged variables were determined by back-transforming interpolated log-transformed population sizes.

2.3. Symbolic regression

Symbolic regression aims to find simple yet accurate equations to explain a given dataset. Using a process akin to biological evolution,

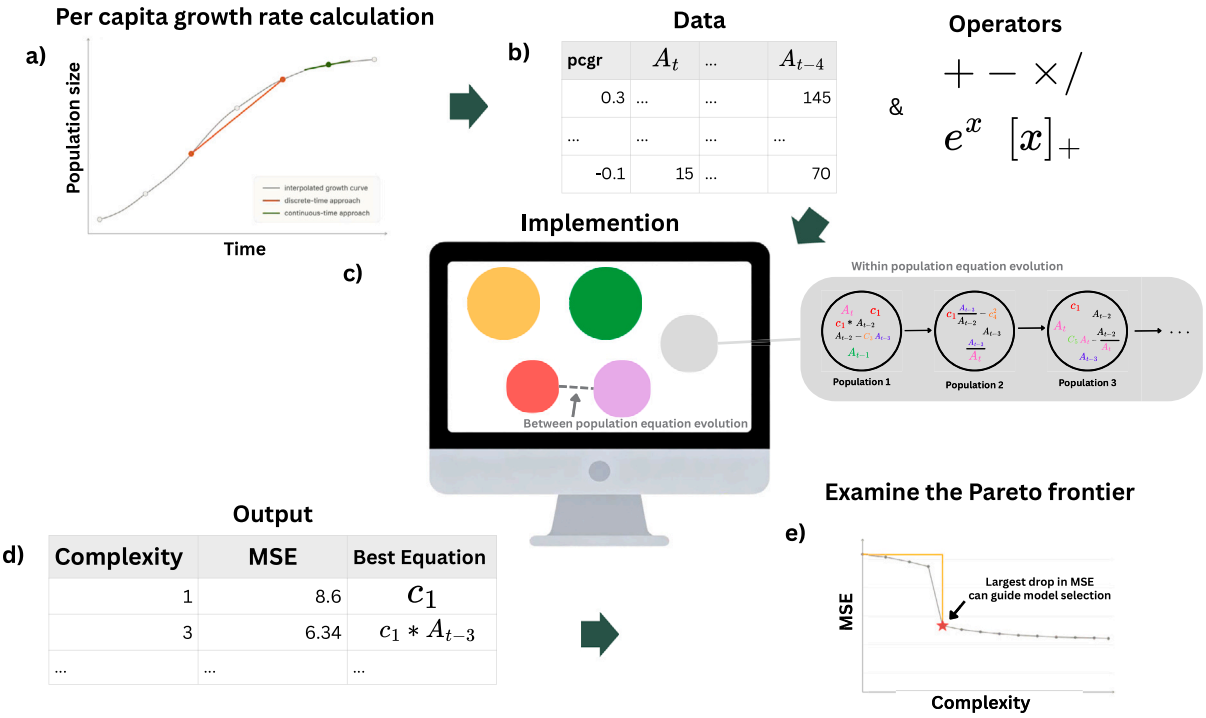


Fig. 2. (a) Time series data were analyzed using either a discrete-time approach (orange line, difference in population sizes) or a continuous-time approach (green line, tangent to interpolated population sizes) to calculate the response variable, the per capita growth rate (pcgr), that we provided to symbolic regression. (b) The additional inputs provided to symbolic regression were the suites of predictor variables ((lagged) population sizes) and mathematical operators (including the custom function $[x]_+$). (c) In PySR, equation evolution occurs in two stages, (*i*) "within population evolution" wherein multiple equations exist within a single generation and parental equations are selected based on their fitness to create new equations in subsequent generations, and (*ii*) "between population evolution" wherein equation components can migrate between concurrent populations. (d) The output at the end of a run entails a list of the best (lowest MSE) equation for each level of complexity, along with its MSE values and complexity level. (e) The Pareto frontier of equations may be examined to select the equation affecting the largest drop in MSE (represented by the orange lines) as done in the first of the four workflows that we evaluated.

generations of equations are evolved, where equations with a better fit to the data have a higher chance of producing offspring equations in the following generation. Symbolic regression begins with the creation of an initial generation of equations based on the input variables and operators (Fig. 2b). Parental equations of high fitness produce offspring equations for the next generation of equations with additions, deletions, recombinations, or mutations of their components. This equation evolution process occurs for a set amount of time or generations (Fig. 2c). The output is a set of top-performing equations that span a range of complexity values. For a more detailed overview, see the *Supplementary Materials*.

2.3.1. Implementation

Prior ecological studies used custom-built symbolic regression algorithms. We used PySR (v1.3.1; Cranmer, 2023), a Python library with a Julia backend that has seen frequent use in Physics. Unlike prior algorithms, PySR evolves equations using multiple semi-independent populations of equations between which migration can occur in order to explore a greater volume of equation space (see the *Supplementary Materials* for further details). With mean squared error (MSE, a.k.a. L2-loss) as the metric of equation fitness, we used 96 populations and ran each downsampled time series through PySR 100 independent times. Each of the $6 \times 5 \times 100 = 3000$ searches was run on a single compute node of a high performance computing cluster equipped with 64 CPU cores and a total of 514 GB of RAM for 22 h or until an equation reached an MSE of 2×10^{-9}, though in almost all cases the search time was exhausted first. (Preliminary experiments revealed that these stopping criteria approached the limit where performance gains were exhausted.) Running 100 independent searches allowed us to quantify rates of equation recovery in a probabilistic manner rather than the binary success/failure manner of prior studies.

2.3.2. Equation and variable recovery success

We compared four workflows — two subjective and two objective — to select a candidate equation from among the learned equations and assess symbolic regression's ability to recover the Bence-Nisbet model. Success rates were quantified in two ways for each workflow: as the proportion of 100 runs in which the selected equation (*i*) included only the two correct predictor variables of the Bence-Nisbet model (i.e., $A(t)$ and $A(t-2)$), and (*ii*) included only the two correct predictor variables and matched the Bence-Nisbet model in its functional form and parameter values (to three significant digits).

Prior studies assessed success using several approaches, including by subjectively inspecting the Pareto frontier of the learned equations to search for the true or a recognizable equation. The Pareto frontier represents the equations that maximize fitness (e.g., minimize MSE) for their given level of complexity; it reflects the most efficient trade-off between equation fitness and complexity. PySR quantifies an equation's complexity as the total count of all parameters, variables, and operators it contains (which for the Bence-Nisbet model is 14). Typically, the simplest model that provides the greatest increase in fitness along the Pareto frontier is selected as the best performing model, although a user's goals may cause the selection of other models. For instance, a goal of high predictive performance may influence a user to select a more complex model than if the goal is mechanistic interpretation (Shmueli, 2010). For Workflow 1, we followed the common practice of visually inspecting plots of MSE versus complexity and selecting the simplest equation that affected the largest additive decrease in MSE relative to all other equations along the frontier (e.g., Fig. 2e). Workflow 2 repeated this visual inspection using ln(MSE) to select the simplest model affecting the largest multiplicative decrease in MSE relative to all other equations along the frontier. PySR itself provides an objective,

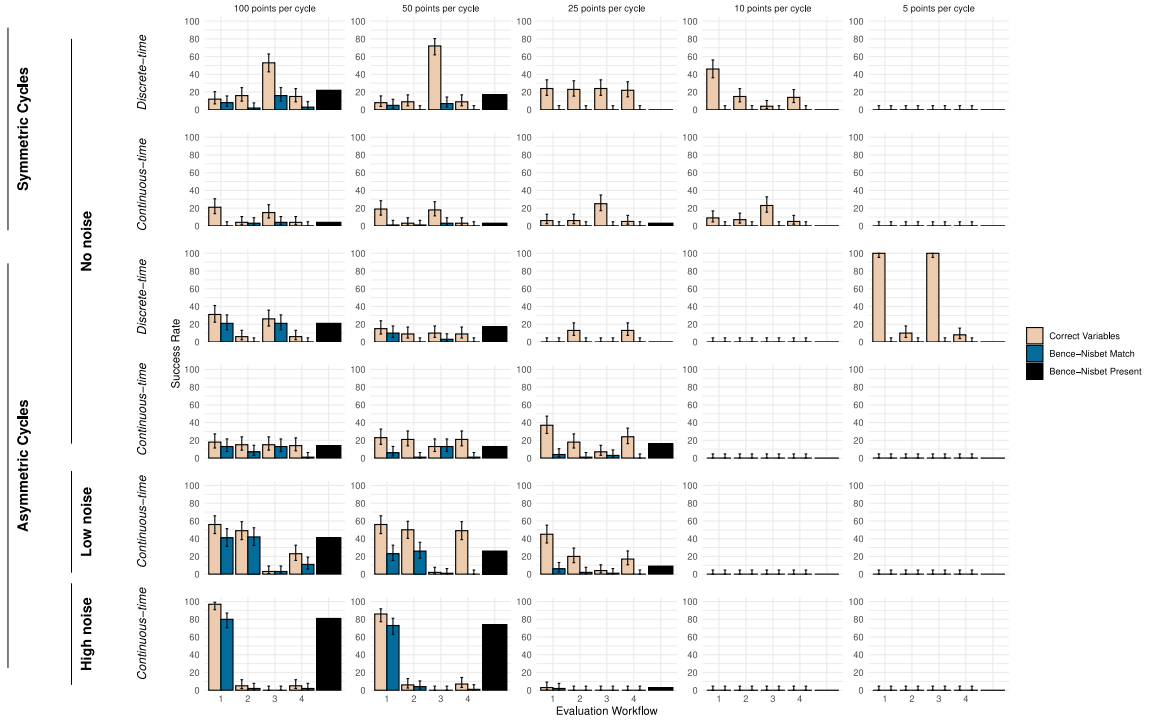


Fig. 3. The success rate of symbolic regression and the four model selection workflows in selecting equations that contained the correct variables or that contained the correct variables and matched the formulation of the Bence-Nisbet model with which we generated the data. Confidence intervals reflect binomial uncertainties given a sample size of 100 independent runs. The black bars indicate the proportion of the 100 independent runs in which the Bence-Nisbet model occurred, regardless of whether a workflow selected it or whether it occurred in the top 10 equations.

quantitative selection criterion equivalent to Workflow 2 called *score* that is calculated as the negative discrete log loss change per unit change in complexity (Cranmer, 2023), which is what we used for Workflow 3. In Workflow 4, we used the Bayesian Information Criteria (BIC) for model selection (Schwarz, 1978). BIC evaluates equation performance by penalizing the equation's fit (its MSE) by the count of its parameters and the sample size of the data (Höge et al., 2018; Hilborn and Mangel, 1997). Next to AIC, BIC remains the most commonly used information criterion but has the property of being asymptotically consistent as the sample size increases (i.e., will select the true model — rather than the most efficient model — when it is present). For all workflows, we only considered equations with a complexity of up to 20 as equations of greater complexity became challenging for human readability, algebraic manipulation, and interpretation.

2.3.3. Explanatory variable diversity and selection success

In addition to quantifying rates of equation and variable recovery success for each workflow as above, we also (*i*) looked for the Bence-Nisbet model and (*ii*) characterized the diversity of variable combinations present within the top (lowest MSE) equations of each complexity level returned by PySR, combining all 100 runs of each case-study sampling-density combination. We did the latter by counting the number of unique variable combinations that occurred in the up to top 10 equations, restricting ourselves to equations in the complexity range of 10–18. (While we considered up to 10 equations, fewer were recovered for certain complexities across the 100 searches. Because these instances were rare, we maintain the 'top 10' terminology throughout for brevity.) Without necessitating an equation selection workflow, this analysis was motivated by the context in which a user does not know the true driving variables or is not expecting classical population models to apply — hence would not look for or be able to identify these along the Pareto frontier or in any suite of returned equations — but could gain confidence and insight by virtue of symbolic regression returning a low diversity of equations that combined just a few dominant variables at or near the Pareto frontier.

Finally, in our most judicious assessment aligned with those of several prior studies, we determined whether the Bence-Nisbet model appeared in any of the returned equations of a given complexity in the range of 10–20, combining all 100 runs of each case study sampling-density combination. This permitted us to distinguish the ability of symbolic regression to evolve the correct generating equation from the ability of the four workflows to identify it along the Pareto frontier. Thus, when the Bence-Nisbet model did occur, we determined its maximum MSE-rank by counting the number of equations of the same complexity that exhibited a lower MSE than it.

3. Results

3.1. Equation and variable recovery success

We observed widely varying success rates (0%–75% of 100 runs) in recovering the Bence-Nisbet model using the four equation selection workflows (Fig. 3). Overall, success rates were most influenced by sampling density and the presence of process noise, with no substantive or consistent effects of cycle symmetry versus asymmetry or preprocessing method. Although none of the workflows proved reliable, the two objective workflows (3 and 4) generally performed more poorly than the two subjective workflows (1 and 2) when success was observed. The highest rates of success were observed when applying Workflow 1 to the continuous-time preprocessed asymmetric time series having high process noise subject to sampling densities of 50 or more time points per cycle. Modest rates of success (10%–20%) were also observed when applying Workflows 1 and 2 to otherwise equivalent time series having low process noise subject to sampling densities of 50 or more time points per cycle. However, rates of success were very low (<10%)

Fig. 4. The proportional frequency of variable combinations present among the up to top 10 (lowest MSE) equations of each complexity level in the range of 10–18 (combining all 100 runs of each case-study sampling-density combination). The Bence-Nisbet model with which the data were generated has a complexity of 14 and contains only $A(t)$ and $A(t-2)$, making the inclusion of $A(t-1)$ and $A(t-3)$ spurious. All possible variable combinations are depicted in the legend; however, the combination of $A(t)$ and $A(t-3)$ never occurred.

or zero for the vast majority of other case studies, workflows, and sampling densities. In fact, the Bence-Nisbet was never recovered by any workflow when sampling densities were less than 50 time points per cycle, except for a few instances in the presence of process noise (19 of 7200 instances; 18 cases × 100 runs × 4 workflows).

We observed higher — albeit still low and widely varying (0%–100% of 100 runs) — rates of success in using the workflows to select equations which contained only the correct variables of the Bence-Nisbet model along the Pareto frontier (Fig. 3). A rate of 100% correct variable recovery occurred anomalously for Workflows 1 and 3 in the case of discrete-time preprocessed deterministic asymmetric cycles subject to a sampling density of only 5 time points per cycle. (None of these equations matched the Bence-Nisbet model's structure, see below.) Similarly, an anomalously high rate of correct variable recovery occurred when applying Workflow 3 in the case of discrete-time preprocessed deterministic symmetric cycles at a sampling density of 50 time points per cycle. Beyond these exceptions, rates of correct variable recovery generally paralleled the patterns observed in the equation recovery success, tending to be low overall (<20%) and higher in the presence of process noise when applying Workflow 1. Unlike for equation recovery, correct variable recovery success rates remained above zero at sampling densities less than 50 time points per cycle in several deterministic cases.

3.2. Explanatory variable diversity and selection success

The diversity of variable combinations among the top 10 (lowest MSE) equations at each complexity level was low for a given case study (Fig. 4). No set of the top 10 equations contained more than 7 of the 11 possible pairwise-or-higher combinations of variables that could occur at complexities above 9, with agreement among all top equations for a single combination of variables or just two combinations of variables being by far the most frequent (48.7% and 27.7% of 267, respectively). Out of the 2581 equations examined in this analysis, $A(t)$ and $A(t-2)$ occurred most frequently (31.8%), followed by $A(t)$, $A(t-1)$ and $A(t-2)$ (23.2%). For most case studies, the frequency of equations that contained only the two correct variables, $A(t)$ and $A(t-2)$, increased substantially at all complexity levels as sampling density increased, with the majority of the top equations per complexity containing only the two correct variables in several high sampling density case studies. (The notable exceptions were the case study considering deterministic symmetric cycles preprocessed with the continuous-time approach, for which the frequency of the correct variable equation remained low for all sampling densities, and the previously mentioned anomalous case of discrete-time preprocessed deterministic asymmetric cycles at a sampling density of 5 time points per cycle.) For all case studies, sampling densities less than 50 points per cycle tended to result in equations containing spurious variables, with the frequency of equations containing the combinations of $A(t)$, $A(t-1)$ and either $A(t-2)$ or $A(t-3)$ increasing particularly in the presence of process noise. Equations containing $A(t)$, $A(t-1)$, and $A(t-2)$ specifically also occurred more frequently for the two symmetric cycle case studies at high sampling densities. Equations that included only the two spurious variables, $A(t-1)$ and $A(t-3)$, never occurred, except in four cases of asymmetric cycles (deterministic, discrete- and continuous-time preprocessed at a sampling density of 10 points per cycle, deterministic, discrete-time preprocessed at a sampling density of 5 points per cycle, and low process noise, continuous-time preprocessed at a sampling density of 10 points per cycle; 178 of 36,000 equations (4 cases × 9 complexity levels × 100 runs × 10 top equations).

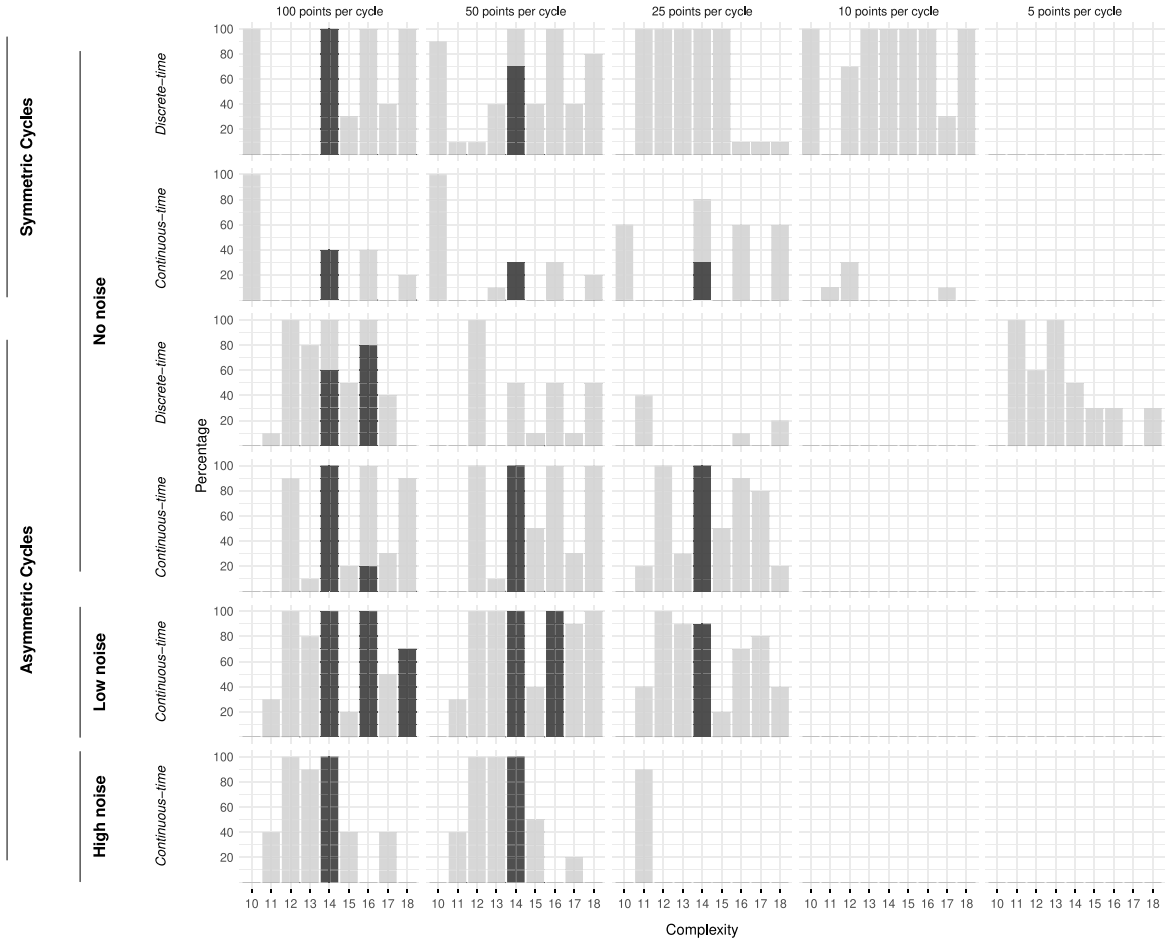


Fig. 5. The proportion of the top 10 (lowest MSE) equations that correctly matched the Bence-Nisbet model in each complexity level in the range 10–18 (combining all 100 runs of each case-study sampling-density combination). Lightgray backgrounds indicate instances in which at least one equation occurred that contained only the two variables of the Bence-Nisbet model, $A(t)$ and $A(t-2)$, as shown in Fig. 4. The complexity level of the Bence-Nisbet model is 14. Matches at complexities greater than 14 occurred when the equations could be algebraically simplified to the Bence-Nisbet model. Note that matches at odd-valued complexity levels are not possible given the even-valued complexity of the Bence-Nisbet model and the necessity of including an operator for every additional parameter or variable.

Despite not being consistently selected from along the Pareto frontier of any given run using the four workflows, the Bence-Nisbet model was evolved by symbolic regression for nearly all case studies (black bars in Fig. 3), and occurred in the top 10 (lowest MSE) equations of all case studies, given a sampling density of 25 or more points per cycle (Fig. 5). In the case studies in which the Bence-Nisbet model occurred within the top 10 as complexity 14, it had the lowest MSE, ranking first among all returned equations, 60% of the time. However, in four instances it held a lower rank (higher MSE), and in two cases it failed to appear entirely, not placing along the Pareto frontier (Fig. 6). In most of the cases where the Bence-Nisbet model occurred, equations of higher complexity which were equivalent to the Bence-Nisbet model (i.e., equations that had extraneous constants that could be removed by algebraic simplification) also occurred (Fig. 5). These equations often also ranked first at their complexity level, but were more likely to drop in rank and out of the top 10 as their complexity increased (Fig. 6).

4. Discussion

We evaluated the utility of symbolic regression for recovering the generative population model of synthetic time series that varied in sampling density, process noise, cycle asymmetry, and preprocessing approach, and assessed four workflows for selecting candidate equations from among those returned. In contrast to prior studies based on densely sampled time series, our results indicate that both equation recovery and equation selection are substantially less reliable under characteristics typical of field-based time series. In particular, recovery success depended strongly on sampling density and, to a lesser extent, on the presence of process noise, while population cycle asymmetry and preprocessing approach had comparatively little influence. Even when the true equation was evolved, none of the four workflows proved consistently reliable at identifying it.

Nevertheless, our findings also suggest reasons for cautious optimism. Indeed, a central insight of this study is the need to distinguish among three related but distinct components of symbolic regression analyses: the algorithm's ability to evolve the correct equation, the position of that equation relative to and along the Pareto frontier, and the performance of workflows in selecting an equation. Although the generative model we used was more complex than those of prior studies, symbolic regression frequently did evolve the correct equation at least once across repeated runs when sampling density was sufficiently high. In these cases, the correct equation often occurred among the top-performing equations of its complexity level and typically ranked first or near-first in terms of MSE across all runs. Moreover, the diversity of equations near the Pareto frontier was generally low and dominated by models containing the correct variables. Together, these results underscore the value of long-term ecological time series of high density (Hughes et al., 2017) and indicate that symbolic regression can

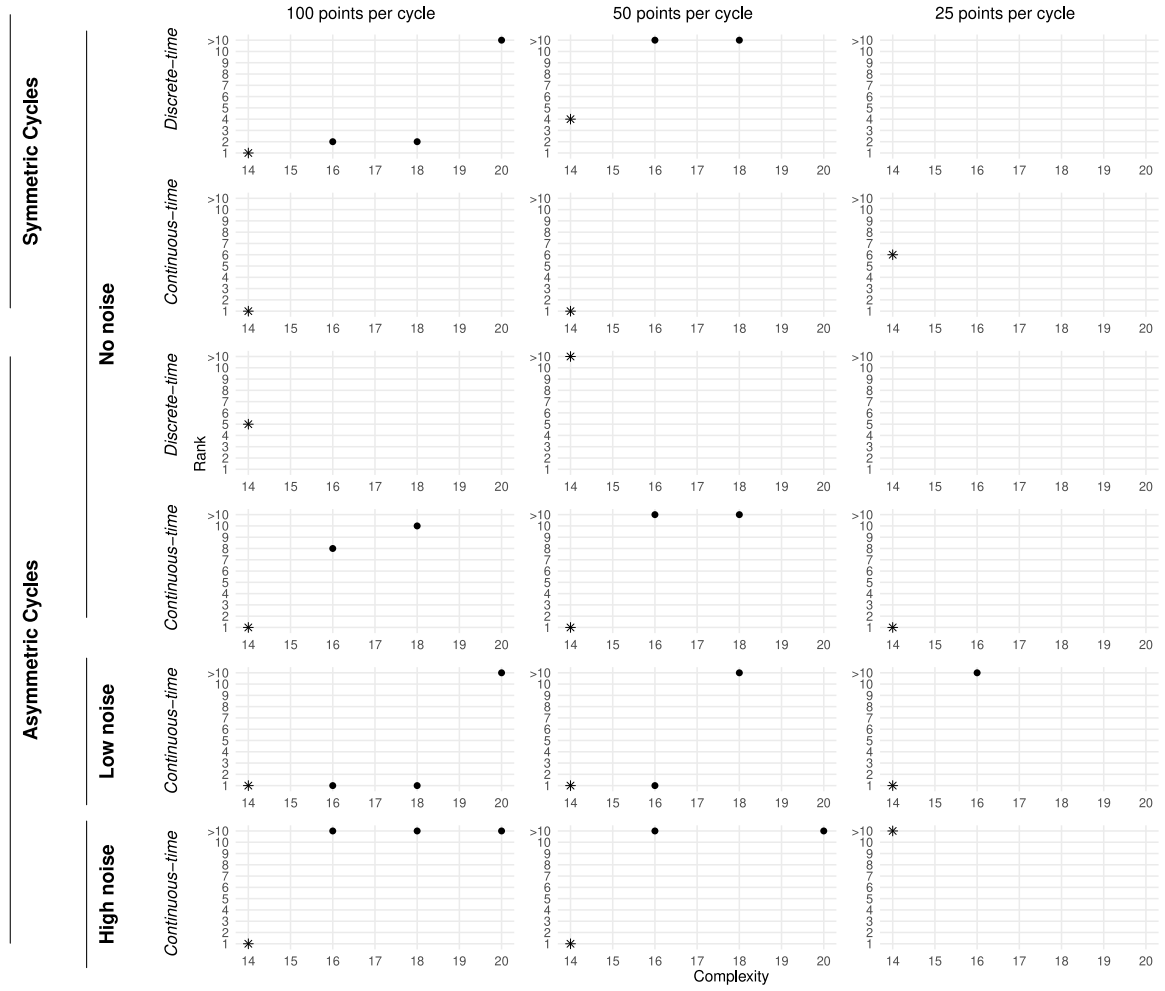


Fig. 6. The maximum MSE-rank of the Bence-Nisbet model when it was recovered, reflecting the number of equations that exhibited a lower MSE than it in the set of up to 100 equations returned by the 100 runs of each case-study sampling-density combination. The complexity level of the Bence-Nisbet model is 14 (indicated by asterisks), hence lower levels are not possible. Ranks at complexities greater than 14 (indicated by points) occurred when the equations could be algebraically simplified to the Bence-Nisbet model. The Bence-Nisbet model was never recovered for time series of 10 and 5 points per cycle or for the deterministic discrete-time processed case study at 25 points per cycle (see Fig. 5).

recover meaningful mechanistic structure from them, but also reveal that identifying this structure requires improved criteria for robust equation selection.

4.1. Equation generation versus equation selection

The low rate of successful recovery on a per-run basis raises the question of whether failures arose from limitations of our symbolic regression implementation or from limitations of the data and evaluation workflows. Our results suggest that the latter two, rather than computational constraints, were the primary factors. All searches were run for substantially more runs, longer durations, and across more populations of equations than in prior ecological studies, making it unlikely that recovery failures reflect insufficient exploration of equation space.

Instead, symbolic regression frequently evolved incorrect equations that fit the data as well as or better than the true model. This outcome is consistent with extensive prior work demonstrating that structurally distinct models can generate indistinguishable dynamics when data do not sufficiently constrain the mechanism (e.g., Perretti et al., 2013; Novak and Stouffer, 2021; Song and Levine, 2025). In such cases, equation selection criteria that rely on fit and various measures of complexity alone cannot be depended upon to identify the true equation, even when it is present. In addition, in our study, the correct equation not only occurred behind the Pareto frontier but more commonly was present on the frontier, yet was not selected by an equation selection workflow, particularly for the workflows that were metric-based and hence entirely objective. Together, these findings highlight the need for criteria that better reflect structural identifiability rather than goodness-of-fit and traditional measures of complexity alone (e.g., Novak and Stouffer, 2021). Approaches that explicitly consider equations both on and behind the Pareto frontier, or that incorporate additional dynamical diagnostics like the ability to reproduce key statistical measures of the time series (e.g., Song and Levine, 2025), may also help to improve inference.

4.2. Data informativeness

4.2.1. Effects of sampling density

Sampling density exerted the strongest influence on symbolic regression's performance. At sampling densities below 25–50 points per cycle, the correct equation was rarely evolved or reliably selected. Because all our time series were sampled with densities determined on a per-cycle basis and the lowest assessed sampling frequency exceeded the Nyquist-Shannon frequency of the simulated dynamics (Shannon, 1949), differences in recovery across case studies are unlikely to reflect classical aliasing effects (see the *Supplementary Materials*). Instead,

reduced sampling density appears to limit recovery by diminishing data informativeness, affecting the accuracy of interpolated abundances and the estimation of growth rates (particularly under a continuous-time approach), and consequently the ability of the data to discriminate among alternative equation variables and structures.

Our analyses therefore emphasize that sampling density must be considered relative to a system's dynamical frequencies or characteristic timescales when applying equation-discovery methods. For strongly seasonal systems, sampling once or even twice a year, as is often the case for field-based population time series, will not be enough. Because achieving high sampling density is often challenging in field-based contexts, alternative strategies — such as irregular sampling (Gaucel et al., 2014), compressive sampling (Candes and Wakin, 2008), adaptive sampling (Henrys et al., 2024), or the use of state-space rather than rate-versus-state equation formulations that avoid the error-sensitive estimation of growth rates (Munch et al., 2020) — may be useful in mitigating these constraints and warrant further investigation.

4.2.2. Effects of process noise

A more favorable result for field-based contexts was that process noise consistently increased recovery success. Indeed, as increasingly recognized in other contexts (Boettiger, 2018; Coulson et al., 2004; Dennis et al., 2006), stochastic perturbations appear to enhance data informativeness by expanding coverage of the system's attractor (see the *Supplementary Materials*) and revealing responses to variation that are absent with deterministic cycles. These effects likely reduce the set of dynamically equivalent models and help distinguish among competing equations. That said, contrary to our expectations, the highest noise level that we assessed often performed as well as or better than the low-noise case. Identifying the threshold at which process noise transitions from informative to detrimental thus remains an important direction for future work.

4.3. Cycle asymmetry and data preprocessing

Population cycle asymmetry had little effect on symbolic regression's performance, with asymmetric cycles performing comparably to or slightly better than symmetric ones. This result contrasts with our expectation that asymmetric cycles would be more vulnerable to undersampling, but suggests that asymmetry alone does not strongly interact with sampling density to constrain recovery when sampling density is sufficiently high (i.e., >25 points) on a per-cycle basis.

Similarly, we observed no consistent advantage of either discrete-time or continuous-time preprocessing for estimating per capita growth rates. Although both approaches can introduce distinct sources of error, these effects were generally secondary to those of sampling density and process noise. Stronger differences may emerge under alternative dynamics or more extreme asymmetries.

4.4. Variable identification

Introducing spurious predictor variables revealed that symbolic regression can correctly identify the true driving variables when sampling density is sufficiently high (i.e., $\gtrsim$25 points per cycle), even when the true equation structure is not evolved or identified. At lower sampling densities, however, autocorrelation among lagged predictors likely favored the inclusion of spurious but dynamically redundant variables. This pattern suggests that variable selection failures at low sampling densities again arise from data informativeness rather than from an inherent limitation of symbolic regression. The implementation of methods like convergent cross-mapping to remove spurious variables during preprocessing may be useful for improving inference (Munch et al., 2020; Sugihara and May, 1990). That said, the observation that the combination of $A(t)$, $A(t-1)$, and $A(t-2)$ specifically was the second most common combination of variables is particularly interesting given that it reflects the variables by which the true dynamics of the our presumed model may be faithfully represented by time-delay embedding using Takens' theorem (Takens, 1981). Future work should examine the degree to which selection biases arise when spurious state variables differ in their degree of autocorrelation with the system's true driving variables, as well as the extent to which time-delay embedding may interfere with the identification of a system's true causal structure. Additionally, investigating the frequency of specific equation structures, rather than just variable occurrences, would be valuable. Such an effort will necessitate the automation of algebraic and structural analyses given the large number of equations involved.

4.5. Conclusions

Our results indicate that symbolic regression can recover mechanistically meaningful population models, but only under conditions of sufficiently informative data and with careful attention to equation selection. Given sufficient sampling, the primary limitation identified here lies not in equation generation but in identifying the correct model from among dynamically equivalent alternatives. Improving post hoc evaluation criteria — particularly those that move beyond exclusive reliance on the Pareto frontier — thus represents a key opportunity for methodological advancement. Such improvements are needed for robust generic guidelines to be developed regarding the specific time series characteristics that are necessary to ensure correct ecological inferences using symbolic regression.

CRediT authorship contribution statement

Cheyenne N. Jarman: Writing – original draft, Visualization, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Taal Levi:** Writing – review & editing, Conceptualization. **Mark Novak:** Writing – review & editing, Funding acquisition, Conceptualization.

Funding

CNJ was supported by an NSF Graduate Research Fellowship (Grant No. 2234662) and the Gaulke Center for Marine Innovation and Technology.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank Ricardo Noe Gerardo Reyes Grimaldo, Rebecca Hutchinson, Bryan K. Lynn, Dave Lytle, Bruce Menge, Steve Munch, and the Model Enable Machine Learning group for their insight throughout the project development. We also thank members of the Novak Lab for their feedback on early drafts of the manuscript.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.ecoinf.2026.103988.

Data availability

All code and data are available at https://zenodo.org/records/19339107 and https://github.com/cheyennejarman/SymbReg_KelpForest_Simulations.git.

References

Baker, N., Alexander, F., Bremer, T., Hagberg, A., Kevrekidis, Y., Najm, H., Parashar, M., Patra, A., Sethian, J., Wild, S., Willcox, K., Lee, S., 2019. Workshop Report on Basic Research Needs for Scientific Machine Learning: Core Technologies for Artificial Intelligence. Tech. Rep. 1478744.

Bence, J.R., Nisbet, R.M., 1989. Space-limited recruitment in open systems: The importance of time delays. Ecology 70 (5), 1434–1441. http://dx.doi.org/10.2307/1938202, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.2307/1938202 https://onlinelibrary.wiley.com/doi/abs/10.2307/1938202.

Boettiger, C., 2018. From noise to knowledge: how randomness generates novel phenomena and reveals information. Ecol. Lett. 21 (8), 1255–1267. http://dx.doi.org/10.1111/ele.13085, https://onlinelibrary.wiley.com/doi/10.1111/ele.13085.

Bongard, J., Lipson, H., 2007. Automated reverse engineering of nonlinear dynamical systems. Proc. Natl. Acad. Sci. 104 (24), 9943–9948. http://dx.doi.org/10.1073/pnas.0609476104, http://www.pnas.org/cgi/doi/10.1073/pnas.0609476104.

Bonnaffé, W., Sheldon, B.C., Coulson, T., 2021. Neural ordinary differential equations for ecological and evolutionary time-series analysis. Methods Ecol. Evol. 12 (7), 1301–1315. http://dx.doi.org/10.1111/2041-210X.13606, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/2041-210X.13606 https://onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.13606.

Candes, E.J., Wakin, M.B., 2008. An introduction to compressive sampling. IEEE Signal Process. Mag. 25 (2), 21–30. http://dx.doi.org/10.1109/MSP.2007.914731.

Chen, Y., Angulo, M.T., Liu, Y.-Y., 2019. Revealing complex ecological dynamics via symbolic regression. BioEssays 41 (12), 1900069. http://dx.doi.org/10.1002/bies.201900069, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/bies.201900069 https://onlinelibrary.wiley.com/doi/abs/10.1002/bies.201900069.

Coulson, T., Rohani, P., Pascual, M., 2004. Skeletons, noise and population growth: the end of an old debate? Trends Ecol. Evolut. 19 (7), 359–364. http://dx.doi.org/10.1016/j.tree.2004.05.008, https://linkinghub.elsevier.com/retrieve/pii/S0169534704001466.

Cranmer, M., 2023. Interpretable machine learning for science with PySR and SymbolicRegression. jl. http://dx.doi.org/10.48550/arXiv.2305.01582, http://arxiv.org/abs/2305.01582.

de Roos, A.M., Persson, L., 2003. Competition in size-structured populations: mechanisms inducing cohort formation and population cycles. Theor. Popul. Biol. 63 (1), 1–16.

Dennis, B., Ponciano, J.M., Lele, S.R., Taper, M.L., Staples, D.F., 2006. Estimating density dependence, process noise, and observation error. Ecol. Monograph. 76 (3), 323–341. http://dx.doi.org/10.1890/0012-9615(2006)76[323:EDDPNA]2.0.CO;2, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1890/0012%2D9615%282006%2976%5B323%3AEDDPNA%5D2.0.CO%3B2 https://onlinelibrary.wiley.com/doi/abs/10.1890/0012-9615%282006%2976%5B323%3AEDDPNA%5D2.0.CO%3B2.

Ellner, S.P., Seifu, Y., Smith, R.H., 2002. Fitting population dynamic models to time-series data by gradient matching. Ecology 83 (8), 2256–2270. http://dx.doi.org/10.1890/0012-9658(2002)083[2256:FPDMTT]2.0.CO;2, http://doi.wiley.com/10.1890/0012-9658(2002)083[2256:FPDMTT]2.0.CO;2.

Gaucel, S., Keijzer, M., Lutton, E., Tonda, A., 2014. In: Nicolau, M., Krawiec, K., Heywood, M.I., Castelli, M., García-Sánchez, P., Merelo, J.J., Santos, V.M. Rivas, Sim, K. (Eds.), Learning Dynamical Systems using Standard Symbolic Regression. In: Genetic Programming, Vol. 8599, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 25–36. http://dx.doi.org/10.1007/978-3-662-44303-3_3, series Title: Lecture Notes in Computer Science.

Gerwin, D., 1974. Information processing, data inferences, and scientific generalization. Behav. Sci. 19 (5), 314–325. http://dx.doi.org/10.1002/bs.3830190504, https://onlinelibrary.wiley.com/doi/10.1002/bs.3830190504.

Henrys, P.A., Mondain-Monval, T.O., Jarvis, S.G., 2024. Adaptive sampling in ecology: Key challenges and future opportunities. Methods Ecol. Evol. 15 (9), 1483–1496. http://dx.doi.org/10.1111/2041-210X.14393, arXiv:https://besjournals.onlinelibrary.wiley.com/doi/pdf/10.1111/2041-210X.14393 https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.14393.

Hilborn, R., Mangel, M., 1997. The Ecological Detective: Confronting Models with Data. Princeton University Press, Princeton, USA.

Höge, M., Wöhling, T., Nowak, W., 2018. A primer for model selection: The decisive role of model complexity. Water Resour. Res. 54 (3), 1688–1715.

Hughes, B.B., Beas-Luna, R., Barner, A.K., Brewitt, K., Brumbaugh, D.R., Cerny-Chipman, E.B., Close, S.L., Coblentz, K.E., de Nesnera, K.L., Drobnitch, S.T., Figurski, J.D., Focht, B., Friedman, M., Freiwald, J., Heady, K.K., Heady, W.N., Hettinger, A., Johnson, A., Karr, K.A., Mahoney, B., Moritsch, M.M., Osterback, A.-M.K., Reimer, J., Robinson, J., Rohrer, T., Rose, J.M., Sabal, M., Segui, L.M., Shen, C., Sullivan, J., Zuercher, R., Raimondi, P.T., Menge, B.A., Grorud-Colvert, K., Novak, M., Carr, M.H., 2017. Long-term studies contribute disproportionately to ecology and policy. BioScience 67 (3), 271–281.

Inchausti, P., Halley, J., 2001. Investigating long-term ecological variability using the global population dynamics database. Science 293 (5530), 655–657.

Kuhn, T.S., 1970. The structure of scientific revolutions. In: International Encyclopedia of Unified Science. In: Foundations of the Unity of Science, Vol. 2, University of Chicago Press, Chicago, no. 2.

Langley, P., 1977. BACON : A production system that discovers empirical laws. IJCAI https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=66b3069d5f92fb9372c3c3eaefd4cd4b26c5add2.

Lotka, A.J., 1925. Elements of Physical Biology. Williams & Wilkins.

Martin, B.T., Munch, S.B., Hein, A.M., 2018. Reverse-engineering ecological theory from data. Proc. R. Soc. B 285 (1878), 20180422. http://dx.doi.org/10.1098/rspb.2018.0422, https://royalsocietypublishing.org/doi/10.1098/rspb.2018.0422.

Munch, S.B., Brias, A., Sugihara, G., Rogers, T.L., 2020. Frequently asked questions about nonlinear dynamics and empirical dynamic modelling. ICES J. Mar. Sci. 77 (4), 1463–1479. http://dx.doi.org/10.1093/icesjms/fsz209, https://academic.oup.com/icesjms/article/77/4/1463/5643857.

Novak, M., Coblentz, K.E., DeLong, J.P., 2025. In defense of type i functional responses: The frequency and population dynamic effects of feeding on multiple prey at a time. Amer. Nat. 206 (4), 347–361. http://dx.doi.org/10.1086/737023.

Novak, M., Stouffer, D.B., 2021. Geometric complexity and the information-theoretic comparison of functional-response models. Front. Ecol. Evol. 9, 776. http://dx.doi.org/10.3389/fevo.2021.740362, https://www.frontiersin.org/article/10.3389/fevo.2021.740362.

Perretti, C.T., Munch, S.B., Sugihara, G., 2013. Model-free forecasting outperforms the correct mechanistic model for simulated and experimental data. Proc. Natl. Acad. Sci. 110 (13), 5253–5257. http://dx.doi.org/10.1073/pnas.1216076110, arXiv:https://www.pnas.org/content/110/13/5253.full.pdf https://www.pnas.org/content/110/13/5253.

Schwarz, G., 1978. Estimating the dimension of a model. Ann. Stat. 6 (2), 461–464. http://dx.doi.org/10.1214/aos/1176344136.

Shannon, C.E., 1949. Communication in the presence of noise. Proc. I. R. E..

Shmueli, G., 2010. To explain or to Predict? Statist. Sci. 25 (3), http://dx.doi.org/10.1214/10-STS330, https://projecteuclid.org/journals/statistical-science/volume-25/issue-3/To-Explain-or-to-Predict/10.1214/10-STS330.full.

Solomatine, D.P., Ostfeld, A., 2008. Data-driven modelling: some past experiences and new approaches. J. Hydroinformatics 10 (1), 3–22. http://dx.doi.org/10.2166/hydro.2008.015, https://iwaponline.com/jh/article/10/1/3/2913/Datadriven-modelling-some-past-experiences-and-new.

Song, C., Levine, J.M., 2025. Rigorous validation of ecological models against empirical time series. Nat. Ecol. Evol.

Sugihara, G., May, R.M., 1990. Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. Nature 344 (6268), 734–741. http://dx.doi.org/10.1038/344734a0, http://www.nature.com/articles/344734a0.

Takens, F., 1981. Detecting strange attractors in turbulence. vol. 898, Springer, Berlin, Heidelberg.

Turchin, P., 2003. Complex Population Dynamics: A Theoretical/Empirical Synthesis. Princeton University Press, Princeton, USA.

Verhulst, P., 1838. Notice on the law that a population follows in its growth. Corresp. Math. Phys. 10, 113–121.

Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., S.J. van der Walt, M. Brett, Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., Carey, C.J., Polat, İ., Feng, Y., Moore, E.W., J. VanderPlas, D. Laxalde, Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A.H., Pedregosa, F., van Mulbregt, P., SciPy 1.0 Contributors, 2020. SciPy 1.0: Fundamental algorithms for scientific computing in Python. Nature Methods 17, 261–272. http://dx.doi.org/10.1038/s41592-019-0686-2.

Volterra, V., 1926. Fluctuations in the abundance of a species considered mathematically. Nature 118 (2972), 558–560. http://dx.doi.org/10.1038/118558a0, https://www.nature.com/articles/118558a0.