Discovering Biochemical Reaction Models by Evolving Libraries

Justin N. Kreikemeyer[1][0000−0002−4109−3608], Kevin Burrage[2,3][0000−0002−8111−1137], and Adelinde M. Uhrmacher[1][0000−0001−5256−4682]

[1] University of Rostock, Rostock, Germany
{justin.kreikemeyer,adelinde.uhrmacher}@uni-rostock.de
[2] Queensland University of Technology (QUT), Brisbane, Australia
[3] Department of Computer Science, University of Oxford, Oxford, UK (Visiting Professor) kevin.burrage@qut.edu.au

Abstract. In a time of data abundance, automatic methods increasingly support manual modeling. To this end, the Sparse Identification of Non-linear Dynamics (SINDy) provides a solid foundation for identifying non-linear dynamical systems in the form of differential equations. In biochemistry, reaction networks imply coupled differential equations. It has recently been demonstrated how this intrinsic coupling can be achieved within the SINDy framework, providing a straightforward interpretation of the learned equations as reaction systems with mass-action kinetics. However, this extension inherits from SINDy the requirement to enumerate all candidate reactions in a library, resulting in ill-posed optimization problems and long model descriptions, limiting its utility for identifying models with many species. Here, we elaborate on the recent advances in bringing SINDy to the biochemical domain by considering the subsampling of reaction libraries as part of an evolutionary optimization scheme. This enables the generation of parsimonious models, as well as the inclusion of model-level constraints, and allows the consideration of large numbers of candidate reactions. We evaluate the approach on two smaller case studies and the recovery of a large Wnt signaling model.

Keywords: sparse regression · genetic programming · automatic model generation · machine learning · reaction systems.

Manuscript

1 Introduction

In *regression*, models are used to infer the relationships among variables from a series of measurement points to, e.g., predict future measurements. This is done by choosing from various available models, from simple linear functions to deep neural networks, and fitting methods. However, the most accurate deep models quickly become opaque to the scientist and end-user, especially when learning complex and non-linear relationships. Thus, *symbolic regression* seeks to identify expressive yet human-interpretable expressions for the relationships inside a system. Instead of choosing a specific model structure, these approaches learn the

This version of the article has been accepted for publication, after peer review and is subject to Springer Nature's AM terms of use, but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections. The Version of Record is available online at: https://doi.org/10.1007/978-3-031-71671-3_10

structure alongside suitable parameters. For example, the Sparse Identification of Non-linear Dynamics (SINDy) [7] can often discover parsimonious differential equations that govern the observed evolution of dynamical systems. Remarkably, it achieves the learning of non-linear dynamics by solving a *linear* equation system, relying on a comprehensive set of time series for all variables of interest and a library of possible function evaluations on those variables (cf. Fig. 1, top).

For many engineering domains, ordinary differential equations (ODEs) are a suitable modeling formalism and, thus, a target for symbolic regression. In systems biology, reaction or rule-based modeling approaches prevail [11], also

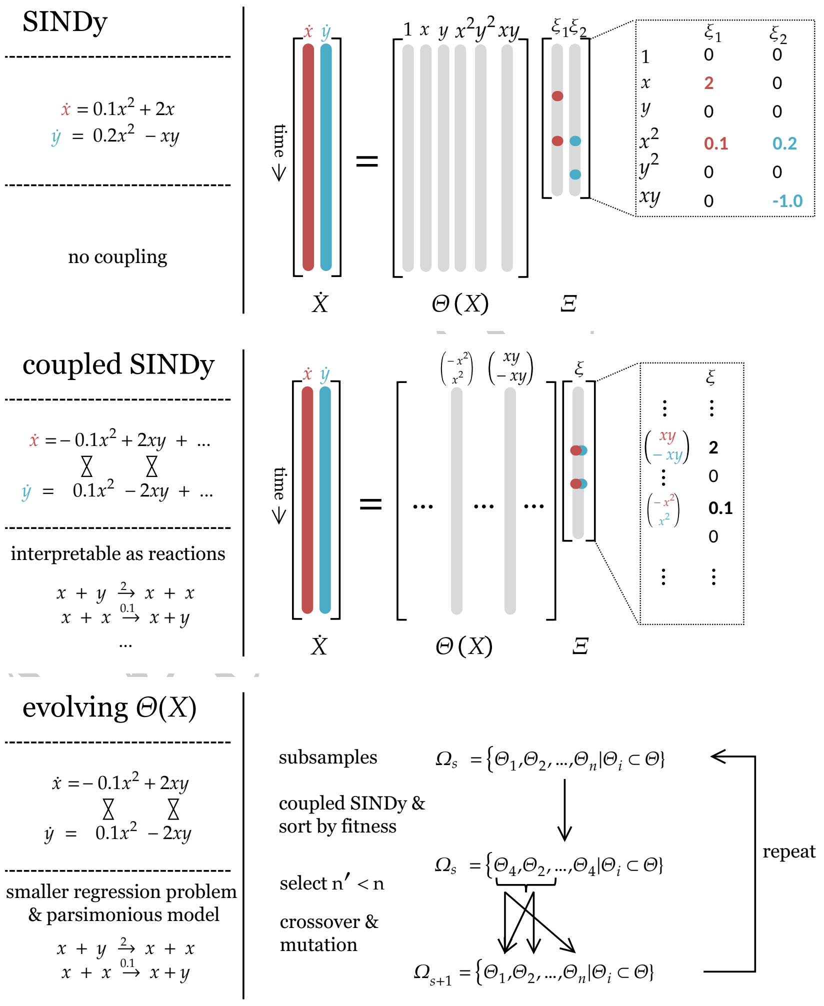

SINDy
$\dot{x} = 0.1x^2 + 2x$
$\dot{y} = 0.2x^2 - xy$
no coupling
$\dot{x}$ $\dot{y}$
time
1 x y x^2 y^2 xy
ξ_1 ξ_2
$\dot{X}$
$\Theta(X)$
Ξ
ξ_1 ξ_2
1 0 0
x 2 0
y 0 0
x^2 0.1 0.2
y^2 0 0
xy 0 -1.0
coupled SINDy
$\dot{x} = -0.1x^2 + 2xy + \dots$
$\dot{y} = 0.1x^2 - 2xy + \dots$
interpretable as reactions
$x + y \xrightarrow{2} x + x$
$x + x \xrightarrow{0.1} x + y$
...
$\begin{pmatrix} -x^2 \\ x^2 \end{pmatrix}$ $\begin{pmatrix} xy \\ -xy \end{pmatrix}$
ξ
$\begin{pmatrix} xy \\ -xy \end{pmatrix}$ 2
0
$\begin{pmatrix} -x^2 \\ x^2 \end{pmatrix}$ 0.1
0
evolving $\Theta(X)$
$\dot{x} = -0.1x^2 + 2xy$
$\dot{y} = 0.1x^2 - 2xy$
smaller regression problem & parsimonious model
$x + y \xrightarrow{2} x + x$
$x + x \xrightarrow{0.1} x + y$
subsamples
$\Omega_s = \{\Theta_1, \Theta_2, \dots, \Theta_n | \Theta_i \subset \Theta\}$
coupled SINDy & sort by fitness
$\Omega_s = \{\Theta_4, \Theta_2, \dots, \Theta_4 | \Theta_i \subset \Theta\}$
select n′ < n
crossover & mutation
$\Omega_{s+1} = \{\Theta_1, \Theta_2, \dots, \Theta_n | \Theta_i \subset \Theta\}$
repeat

Fig. 1: Schematic overview of SINDy ([7], top), coupled SINDy ([8], center) and evolving libraries (this paper, bottom). For simplicity, we neglect the 1/2 factor for x^2 and y^2 resulting from mass-action kinetics. Graphic inspired by [7].

reflected in systems biology standards [16]. Although reactions are frequently mapped to ODEs for simulation, the opposite mapping is not as straightforward [32]. Particularly, if ODEs are learned from data, they might easily miss an important feature of the modeled biochemical reactions that is intrinsically captured by reaction-based modeling approaches: reactions introduce a *coupling* between ODEs. For example, a single reaction $A + B \xrightarrow{k} C$, translates into three coupled ODEs $dA/dt = dB/dt = -dC/dt = -k \cdot A \cdot B$. Note how terms in the differential equation for one variable constrain the terms that may occur in the other variables' derivatives (cf. Fig. 1, center). Based on this observation, in [8], a coupled approach to SINDy is proposed. By using a library of possible reactions instead of possible functions, reaction-based models can be learned from data.

One problem with introducing such a coupling is the combinatorial nature of the possible couplings. Compared to standard SINDy, its coupled counterpart exhibits a much wider design matrix, which can lead to problems with linear solvers whose ability to find good solutions depends on a well-conditioned matrix (cf. Section 5.1). This motivates a subsampling approach, where only a well-chosen subset of the reaction candidates is used to fit a model. To this end, we here embed coupled SINDy inside a genetic algorithm (cf. Fig. 1, bottom) to find an optimal subset and thus model. This allows constraining the model size to a minimum, ensuring human interpretability, as well as handling large libraries that would otherwise result in extremely large regression problems. We evaluate this idea on three case studies (cf. Section 5): the discovery of disease spread, predator-prey dynamics, as well as the complex dynamics of a Wnt pathway [34] involving 19 variables, which results in a library of around 37 000 reactions. To the best of our knowledge, the latter is one of the largest search spaces tackled so far for the learning of reaction models. For example, the large case study in [17] searches only around 6000 possible reactions. Before introducing our method in greater detail in Section 4, we provide some background on biochemical reaction models and (sparse) symbolic regression (cf. Section 2), including related work (cf. Section 3). We conclude with a discussion of the results and an outlook in Sections 6 and 7.

Manuscript

2 Background

This section reviews the formalism of reaction systems, which is widely used to develop models in biochemistry. Typically, these models are hand-crafted, making extensive use of domain knowledge. The reaction systems' parameters are often calibrated based on measurement data. The second part discusses how to additionally identify a model structure from (time-series) data, which is the main topic of this paper and further discussed in Section 3.

2.1 Biochemical Reaction Models

In systems biology, reaction networks are important in describing system dynamics. Their simple syntax allows a natural specification of simulation models in terms of reactants, products, and reaction rate constants:

$$R_j\colon\ l_{0,j}S_0 + \cdots + l_{i,j}S_i \xrightarrow{c_j} r_{0,j}S_0 + \cdots + r_{i,j}S_i \tag{1}$$

where S is a vector of species, $l_{\cdot,j}$ and $r_{\cdot,j}$ are vectors of coefficients determining the stoichiometry and change $\nu_j = r_{\cdot,j} - l_{\cdot,j}$ of reaction j, and c_j determines the rate at which the reactants are consumed to form products. It is common to restrict $\sum_i l_{i,j} = 2$ for all reactions j, i.e., only consider *binary reactions*, as a collision of more than two reactants is unlikely. Starting from an initial vector of amounts for each species, their quantities evolve over time t, which we denote as $S(t)$. Many systems obey the *law of mass action*, such that the effective rate $a_j(S(t))$ (also called propensity) of a reaction depends on the reactant's amounts at the current time. For example, for deterministic systems $a_j(S(t)) = A(t)B(t)c_j$ in the case of binary reactions with reactants A and B ($A \neq B$) and $a_j = 0.5A^2(t)c_j$ ($A = B$).

Assuming a *homogeneous mixture* of species, the semantics of this representation are defined by a system of ODEs called the chemical master equation (CME) [35]. However, as this system is intractable for many relevant cases, simulation is often applied in practice. The CME can, for example, be understood as a continuous-time Markov chain (CTMC) over the evolution of the amounts of species $S(t)$, accounting for the discreteness and stochasticity of small amounts. Using the stochastic simulation algorithm, sample trajectories of the CTMC can be obtained, approaching the exact distribution in the limit of samples. Here, we focus on approximating the CME under the assumption of a deterministic system so that reactions can be directly translated to (coupled) ODEs over time. An example reaction system and its translation to a system of differential equations are depicted in Fig. 1 (bottom left).

Many models use rule-based formalisms, which offer even more expressiveness by introducing the concepts of attributes and compartments [15,4,11]. In cases where attributes take on only a finite set of values and compartments are not dynamic, it is possible to derive the underlying "flattened" reaction network. We use this approach in Section 5 when considering a Wnt signaling model.

2.2 Symbolic Regression for Reaction Systems

The field of symbolic regression develops methods to derive symbolic expressions fitting a given set of measurement points. When the fitting is successful, these are powerful alternatives to black-box regression models, such as neural networks. Symbolic regression methods have also been applied to reaction systems. Early approaches relying on genetic programming [18,26], start with a random population of reaction systems and converge to increasingly fit solutions by applying evolutionary operators.

Whereas genetic programming can, in principle, work with any symbolic expression including imperative programs, the sparse identification of non-linear dynamics (SINDy) [7] is tailored toward differential equations describing the behavior of a system over time, as are common in computational science. This

Manuscript

is made possible by constructing a linear equation system over derivative measurements and possible functions involved in them and solving a least-squares problem. The latter is defined over (1) N measured variables and M discrete measurements of their time derivatives $Y' \in \mathbb{R}^{M \times N}$ (which may be determined numerically from Y) and (2) a matrix of candidate derivatives (design matrix) $\Theta(Y) \in \mathbb{R}^{M \times K}$, determined by the application of a vector of candidate functions $\Theta \in \mathbb{R}^K$ with $\theta \in \Theta \colon \mathbb{R}^N \to \mathbb{R}$ to each measurement in Y:

$$\min_{\Xi} ||Y' - \Theta(Y)\Xi||_2 \tag{2}$$

The weights or parameters $\Xi \in \mathbb{R}^{K \times N}$ then provide a factor with which each function from the *library* Θ contributes to each of the N ODEs. A schematic depiction is given in Fig. 1 (top). This problem formulation can be further extended in many ways, for example, by using regularization [36], a two-step procedure [17], or ensembling [12].

By requiring a specific coupling of the variables, SINDy can also be used to discover reaction networks [8]. The key idea is that Θ now consists of *vectors* of functions, reflecting the terms that will appear in each of the species' derivatives when a particular reaction is included in the model (cf. Fig. 1, center). In this formulation, the library Θ consists of the corresponding derivatives of all possible candidate reactions and there is only one coefficient ξ per reaction indicating its rate constant.

3 Related Work

As the automatic discovery of mechanistic models from data promises enormous benefits not only for predicting dynamics but also for furthering their understanding, it has a long-standing history under many names [19]. Nevertheless, there are still many unsolved challenges and the amount of available data steadily increases, making it an active research area. Here, we present some recent work closely related to our approach.

Discovering ODEs from data, with a focus on biochemical systems, is discussed in [10]. The authors automatically discover parsimonious models by systematically fitting models of increasing complexity. In [7], the seminal Sparse Identification of Nonlinear Dynamics (SINDy) is proposed as already introduced in the previous section. [12] presents an extension to SINDy in the form of E-SINDy, which allows for uncertainty quantification by subsampling either by time (bagging E-SINDy) or library terms (library bagging E-SINDy). As opposed to uncertainty quantification, a random subsampling of the library is not enough in our case (cf. Section 4), hence we perform a goal-driven evolution of our library of reactions. The authors of [8] introduce a method to generate reaction models with SINDy by considering the coupling of species. They test their method by building a reaction system surrogate for an agent-based model. We here extend this approach with a genetic algorithm and evaluate its ability to recover a ground truth reaction model from synthetic measurement data.

Manuscript

The latter is just one example of a method for the data-driven discovery of reaction systems, and the number of proposed methods is growing, particularly in the last decade. Recently, a deep learning architecture based on a variational autoencoder has been demonstrated to discover stochastic reaction models [3]. Preliminary results show the successful recovery of a single reaction with two species from data. This work is part of ongoing efforts in the field of relational inference using variational autoencoders to learn interaction graphs from data, such as [13]. Closely related to (coupled) SINDy, the "ReactioNet Lasso" [17] is a method to fit *stochastic* population models to heteroscedastic time series distribution snapshots. It works by considering the moment equations, an ODE system describing the evolution of stochastic moments derived from the chemical master equation (cf. Section 2). Another recent method is Reactmine [24], which builds models one reaction at a time by stochastically searching a model tree. By limiting the depth of the search tree, the discovery of parsimonious models is enforced. Here, we evaluate a library-based approach and limit model size using an evolutionary algorithm.

Combining sparse regression with genetic programming approaches is not a new idea. Most recently, [2] uses a combination of SINDy and genetic programming to allow the inclusion of rational functions in the library for the recovery of multibody physical systems. Similarly, [22] presents a method (not relying on genetic programming) to handle rational functions occurring in biological networks in the SINDy framework, deriving parsimonious models. A similar idea to our approach is presented in [23], where SINDy is integrated with symbolic regression via genetic programming in a framework called "Deep Symbolic Regression". The authors also find that, in their applications to orbital mechanics, genetic programming enables a limitation of the learned models' complexity. Here, we are interested in the special case of reaction systems and coupled SINDy.

4 Evolving Libraries

As introduced in Section 2, SINDy requires a library of all possible functions to be applied to the measured time series. The number of possible interactions and, thus, library terms increases superlinearly with the number of variables. When requiring a coupling of reactions as with coupled SINDy, this problem becomes significantly more pronounced (cf. Fig. 1), which can lead to a design matrix that poses major challenges for linear solvers in terms of problem size and collinearities (cf. Section 5.1).

To circumvent this issue, a straightforward idea would be to reduce the library size by subsampling, as also proposed in [12]. In the former, library subsampling is used to estimate the algorithm's uncertainty over the learned coefficients Ξ. However, for this to work, the samples must be large enough to ensure a high probability of including many "helpful" reactions. Typically only one or two library terms are omitted. Taking much smaller subsamples would result in a random search over models. Hence, this approach does not readily solve the problems that arise from large libraries. Thus, we take the idea one step further.

Manuscript

4.1 Searching the Sub-Library Space using a Genetic Algorithm

Instead of taking purely random subsamples, we perform a genetic search over possible subsamples of the library, assigning a higher fitness to libraries that lead to a better fit. Whereas coupled SINDy alone has to consider all possible models "at once", an evolutionary strategy can consider a small subset at a time. Limiting the number of reactions in the library also enforces a certain degree of parsimony, as observed in [23] for symbolic expressions. Further, combinations of reactions that lead to low regression accuracy are discouraged. A schematic overview can be found in Fig. 1 (bottom).

Our genetic algorithm (GA) follows a standard procedure [20] and starts with a randomly sampled initial population of n reaction sublibraries, each containing a fixed number of reactions. The samples are taken from an enumeration of all possible reaction structures up to configurable bounds for the number of reactants and products and a given number of species. Duplicate reactions are removed from a subsampled library.

In each step, coupled SINDy is applied to all library subsamples, yielding estimates of the rate constants. The R^2 score of each sublibrary, parametrized with the corresponding constants, is used as its *fitness value.* After sorting the population members by their fitness, the first $n' < n$ libraries are *selected* for reproduction, i.e., as the basis for generating the next generation's population. In our experiments, we typically use a relatively small value for n' between one and ten percent of n. From the n' selected libraries, n new libraries are obtained by (1) picking a library and randomly replacing one reaction with a newly sampled one (*mutation*) and (2) picking two libraries and exchanging several reactions among them (*crossover*). While mutation allows the exploration of unknown parts of the optimization space, crossover exploits good solutions under the assumption that, often, the combination of two fit solutions achieves a higher fitness than the original solutions alone. The latter is intuitive in the case of evolving libraries, as the presence of reactions explaining a certain behavior correlates with a high fitness of the sublibrary. In other words, the fitness of a sublibrary is expected to be correlated with the presence of certain "helpful" reactions. The above procedure is repeated over a given number of steps (generations) or until the current fittest sublibrary reaches a threshold for the R^2 score. Between steps, one elite population member (the currently fittest sublibrary) is preserved as-is for the next generation.

The coupled SINDy approach used to determine the rate constants and fitness values follows the description in [8] and was, together with all experiments, implemented[4] in the Python programming language. In preliminary experiments, we found that the non-negative least squares (nnls) algorithm described in [5] and, in our case, implemented in the `SciPy` Python module [31] produced the best results. Application of STLSQ [7] or the LASSO [36] often resulted in worse solutions in our coupled case. For numerical differentiation, a central differences method as implemented in `NumPy`'s [27] `gradient` function is used, as we found that forward differences performed consistently worse in our applications.

[4] https://zenodo.org/doi/10.5281/zenodo.11654439

Manuscript

4.2 Including Background Knowledge

In many cases, biochemical models are not built from the ground up but based on existing models and knowledge [34,1]. Considering the data-driven learning of models, this calls for a method to include such background knowledge.

The simplest form of this is the removal of certain reactions from the library. For example, if there is information about a compartment structure of the biochemical system, such as a division into cytosol and nucleus, one may want to exclude reactions between species located in different compartments. These can be removed from the reaction enumeration that underlies the initial population generation, as well as the mutation operator.

When some reactions are known, i.e., a model is extended, the coefficients of known reactions can be constrained to values greater than zero using a suitable regression algorithm [17]. Alternatively, if $\Theta = \Theta_1 \cup \Theta_2$ and a subset of reactions Θ_1 is known a priori along with their rate constants Ξ_1, the effects of these can be subtracted from the observations, and the problem from Equation 2 becomes:

$$\min_{\Xi_2} ||(Y' - \Theta_1(Y)\Xi_1) - \Theta_2(Y)\Xi_2||_2$$

This idea is also employed in [24] as part of a tree-based search over models.

An advantage of the GA is the possibility of including arbitrary (numerical) terms in the fitness function. This can be used to introduce an inductive bias, which encourages sublibraries with certain properties. In contrast to the "library-level" constraints already possible with SINDy, this allows the definition of constraints on the "model level", e.g., on the existence of a reaction that produces/consumes a certain species. For example, in the case of a compartment structure discussed above, this can be used to encourage the existence of reactions for shuttling between parts of the cell (cf. Section 5.3).

5 Case Studies

In this section, we apply the concept of evolving libraries in three benchmarks. First, we aim to recover a compartmental model of epidemic dynamics (SIR) [25] and a well-known model of predator-prey dynamics based on the works of Lotka and Volterra [28]. These models contain a rather small number of reactions and species (≤ 3). However, as noted before, even for these, the search space is quite large (around 30 to 100 candidate reactions or 2^{30} to 2^{100} models when allowing models of all sizes). As a final benchmark, we also study the recovery of a Wnt pathway model as presented in [34]. After flattening (cf. Section 2) and some reformulations, this model consists of 43 reactions over 19 species. For this model, we consider a search space of around 37 000 candidate reactions.

As is commonly done, we focus on a synthetic environment for model learning to highlight the raw capabilities of our approach. For each benchmark model, we generate time-series measurements by simulating the ground truth model over a certain time period with the LSODA integrator [29]. We found this variable step-size method to consistently yield the best results among the methods

Manuscript

we tested (explicit Runge-Kutta methods of orders three, five, eight; Backward Differentiation Formulas). The likely reason is its automatic adaptation to the observed stiffness of the system, which can vary widely between learned models. Measurements are taken at fixed intervals or at each (variable-size) integration step. The latter can yield better results than fixed steps, as regions with larger derivatives, higher frequencies, and turning points are sampled more densely. Generally, measurement data is rarely well-formed like this but may be obtained by an appropriate experiment design. Even then, it will exhibit noise depending on the measurement technique, which we disregard here for simplicity.

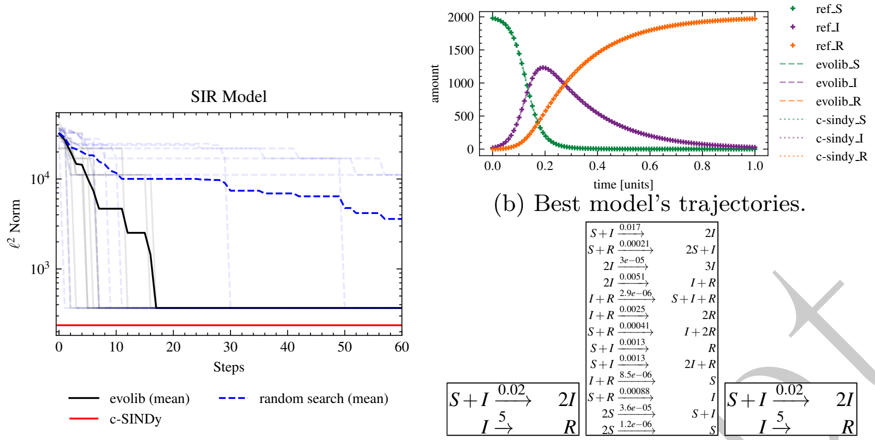
We compare the results of coupled SINDy (c-SINDy) to the results obtained with the evolving libraries (evolib) approach. Additionally, we compare a random search that follows the same procedure as the GA but samples a new population at each step, reporting the fitness of the best-found solution until the current step. For each problem, the "complete" library consists of all reactions between the respective number of species with up to two reactants and up to three (SIR, Predator-Prey) or two (Wnt) products. In every case we limit the maximum change $\nu_{i,j}$ per species i (cf. Section 2) to one, excluding some cases where the underlying derivatives are collinear, e.g., $A \xrightarrow{1} 3A$ ($dA/dt = 2A$) and $A \xrightarrow{1} 2A$ ($dA/dt = A$) where only the latter will be part of the library. As the results of evolib and random search are stochastic, we perform ten replications of the complete optimization, additionally reporting the mean. An overview of the hyperparameters used in our experiments can be found in Appendix A.

5.1 Susceptible-Infected-Recovered Model

The susceptible-infected-recovered (SIR) model describes the dynamics of an infectious disease spreading in a population. Susceptible individuals are infected upon interaction with an infected individual ($S + I \rightarrow 2I$). Infected individuals recover at a certain rate ($I \rightarrow R$). Here, the rate constants of the ground truth model are set to 0.02 for the infection and 5 for the recovery reaction; the initial population consists of 1980 susceptible, 20 infected, and 0 recovered individuals. The data comprises 100 points measured at equal distances over a period of one time unit. The GA worked with 100 sublibraries, i.e., each generation of the GA contains 100 individuals. Each sublibrary contains 2 reactions, and the total number of candidate reactions to select from is 97.

The results shown in Fig. 2 indicate that, on this dataset, the solver used with c-SINDy and the GA in evolib converge to a solution yielding a good fit to the data. In this rather small search space of choosing two out of 97 reactions, the random search also eventually converges to the same solution as evolib, but at a much slower rate. Compared to c-SINDy, the evolib approach results in a more parsimonious model, in this case, recovering the ground truth model. eikeThe resulting model of the c-SINDy approach requires many more reactions with lower rate constants to achieve its fit, learning "wrong" mechanisms. This disadvantage can be attributed to the design matrix's properties. Whereas evolib's smaller matrices often have full rank, the complete matrix representing 97 reactions only has rank 18, posing a rather ill-conditioned problem.

Manuscript



$$\begin{array}{rl} S+I \xrightarrow{0.02} & 2I \\ I \xrightarrow{5} & R \end{array}$$

$$\begin{array}{rl} S+I \xrightarrow{0.017} & 2I \\ S+R \xrightarrow{0.00021} & 2S+I \\ 2I \xrightarrow{3e-05} & 3I \\ 2I \xrightarrow{0.0051} & I+R \\ I+R \xrightarrow{2.9e-06} & S+I+R \\ I+R \xrightarrow{0.0025} & 2R \\ S+R \xrightarrow{0.00041} & I+2R \\ S+I \xrightarrow{0.0013} & R \\ S+I \xrightarrow{0.0013} & 2I+R \\ I+R \xrightarrow{8.5e-06} & S \\ S+R \xrightarrow{0.00088} & I \\ 2S \xrightarrow{3.6e-05} & S+I \\ 2S \xrightarrow{1.2e-06} & S \end{array}$$

$$\begin{array}{rl} S+I \xrightarrow{0.02} & 2I \\ I \xrightarrow{5} & R \end{array}$$

(a) Convergence of optimization.

(b) Best model's trajectories.

(c) Learned models (ground truth, coupled SINDy, evolving libraries).

Fig. 2: Results for recovering the SIR model from synthetic measurement data. Even though the solver used with coupled SINDy is an iterative procedure, the steps are not comparable to the GA steps and we only indicate the final achieved loss for coupled SINDy by a straight line in all figures.

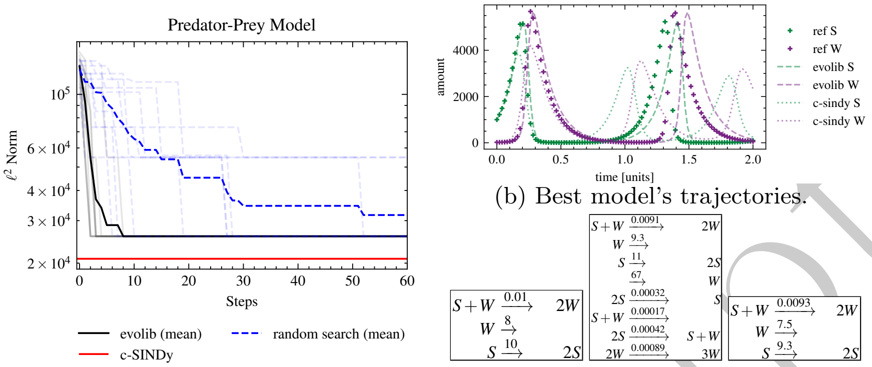
5.2 Lotka-Volterra Model

Our second example is a model of predator-prey dynamics. The predator (here W/wolf) interacts with the prey (here S/sheep), increasing the predator population ($S + W \to 2W$). The predator population decays over time ($W \to \emptyset$), whereas the prey population multiplies ($S \to 2S$). The rate constants and initial population for the ground truth model were set to 0.01, 8, 10.0 for predation, predator decay, and prey multiplication, respectively, and the initial population is given by $(S, W) = (1000, 20)$. The data comprises 100 equally-spaced measurements over two time units. Analogous to the SIR benchmark, the GA worked with 100 sublibraries of size 3 (total number of candidate reactions 28).

We see that the results shown in Fig. 3 are very similar to the observations when learning the SIR model. In contrast to these results, for this ground truth parametrization featuring an oscillation between the predator and prey species, we found that 100 measurements were not enough for an accurate fit, although very similar dynamics producing an oscillation of different frequencies were still discovered. Notably, c-SINDy achieves a lower Loss, but the model learned by evolib provides a better fit to the reference trajectory. This is because the solution quality is measured by the goodness of fit to the *numerical derivatives* (cf. Y' in Eq. 2 in Section 2), which in this case are not reliable as there is an insufficient number of samples at the population peaks. Thus, whereas c-SINDy can overfit the (misleading) derivatives, evolib is forced to generalize, abiding

Manuscript

by the configured maximum model size. Using the LASSO with c-SINDy and different regularization weights also yielded similar results.



(b) Best model's trajectories.

(a) Convergence of optimization.

(c) Learned models (ground truth, coupled SINDy, evolving libraries).

Fig. 3: Results for recovering the Predator-Prey model from synthetic measurement data.

5.3 Wnt Pathway

Finally, we also study a model of the Wnt signaling pathway as a very large application with significant relevance to systems biology. We use the model given in [34], an extension of both, the seminal Wnt model by Lee et al. [21], and an adaptation of Haack et al. [14]. The goal of the study [34] was to explain the cellular response of osteoblasts to increased oxidative stress induced by placing them on a micropillar structure found on a titanium implant. The original model, written in ML-Rules [15], uses attributes and compartments. For this study, we transformed the model into a flattened version so it presents a simple reaction system (cf. Section 2). Further, without changing the model behavior, the Sox17 species was replaced by respectively increasing the degradation rate of the TCF/β-catenin complex (R14). The Ros synthesis (R1), for which the rate depends on a counting species P (H1) was rewritten with mass-action kinetics. The final model consists of 43 reactions between 19 species (cf. App. C), resulting in a highly non-trivial learning task.

Hence, we make some additional assumptions. First, as we are mostly limited to the capabilities of SINDy, we can only readily consider the case where all species are measured. This is rarely the case in systems biology, despite continuous advances in measurement techniques, such as single-cell mass-cytometry [33]. In [34], solely measurements of Axin, β-catenin, Sox17, and ICAT at a small number of time points were available in addition to the basis models. Secondly,

Manuscript

we here use the steps proposed by the numeric integration with LSODA as measurement points as described at the beginning of this section. We additionally truncated the time series to start at the 16-minute mark and end at 500 minutes, as we found that the extremely high derivatives at the beginning, which are a result of a long and fast chain of reactions, hindered a fast convergence of the optimization. In fact, the ground truth model was fitted with the primary goal of reproducing the steady-state behavior observed in the wet lab experiments, disregarding the initial transient response phase. Note, however, that this may generally discard important information on the model dynamics, and we leave an in-depth analysis of alternative ways to deal with this situation inside the SINDy framework as future work.

Similar to [17], we consider two cases: (1) learning of the complete model "from scratch" and (2) extending a base model (representing the known reactions from [21,14]) with the six rules that have been added and fitted in [34] (amounting to nine reactions when flattened). For both cases, we also tested the inclusion of background knowledge about the compartment structure of the system. This entails reducing the library only to include reactions where species within the same compartment (nucleus or cytosol) interact and further the adjustment of the fitness function to penalize libraries that do not include shuttling of species, encouraging the existence of at least one migration into/out of the nucleus, respectively. We implement this by subtracting 1 from the R^2 score fitness if there is no shuttling into the nucleus (similarly for shuttling out of the nucleus). Given that the R^2 score quickly approaches its maximum value of one during optimization, this is a large enough penalty to downrank sublibraries when sorting them by their fitness. As it enables unwanted constructs, we also penalize the use of the counting species P with 0.01. We choose a low penalty as two reactions in the ground truth model still depend on it.

With this in mind, the complete library size for c-SINDy is 37 018. When extending the model, the 35 fixed reactions are removed from this library. In the case where we constrain reactions to occur only between reactants within the same compartment, the library size is significantly reduced to 10 462 reactions.

Fig. 4 shows the convergence of the four scenarios described above. As a first result, we see that, as expected, the evolutionary algorithm significantly outperforms the random search, which we only performed for the unconstrained cases. In all cases, evolib yields a higher loss than c-SINDy. Looking at the resulting models reveals that c-SINDy's solutions hinge on a huge amount of around 500 reactions, some of which exhibit enormous rates. These cannot reasonably be depicted here. The models learned with evolib are shown in Appendix C. Thus, we conclude that, similar to our smaller case studies, evolib trades off a higher loss value for a more comprehensible model.

In contrast to the smaller case studies, we observed that there is no significant overlap between the ground truth model and the learned models. Some ground truth reactions are recovered, but often their effects are scattered over multiple different reactions in the learned model so that a single reaction is expressed by multiple ones.

Manuscript

For example, the model learned by evolib from scratch shown in Appendix C Fig. 6a includes the Axin-induced degradation of β-catenin, which is a central part of the Wnt pathway. In the constrained case Fig. 7a, the shuttling of β-catenin is recovered. The small overlap with the ground truth model is a hint that our chosen constraints are not enough to enforce biologically viable models, which underlines the problem of model identifiability (cf. Section 6). This is further underpinned by the fit to the reference trajectory shown in Appendix B. In the extension cases, a visually accurate fit is achieved when simulating the learned models, despite them being different from the ground truth. Judging from Fig. 4, the (excessive) models learned by c-SINDy seem to produce much more accurate fits but were not amenable to integration with LSODA, which resulted in errors regarding numerical precision (cf. Section 6).

Note, that at this time, the implementation in Python is not optimized for performance, so we can only reliably estimate theoretically the performance benefits of evolib (cf. Section 6). With our current setup and hyperparameters, we found that performing ten replications of evolib in parallel, also parallelizing the fitness evaluation, took around five times as long as a single run of c-SINDy for the Wnt experiments (1h vs. 5h), but this result depends heavily on the population size and number of steps performed. For example, increasing the population size allows for decreasing the number of steps, where the computations on the former can be fully parallelized, leading to much fewer sequential steps and faster convergence. Note also that we limited the number of iterations for non-negative least squares to 10^8 for both approaches.

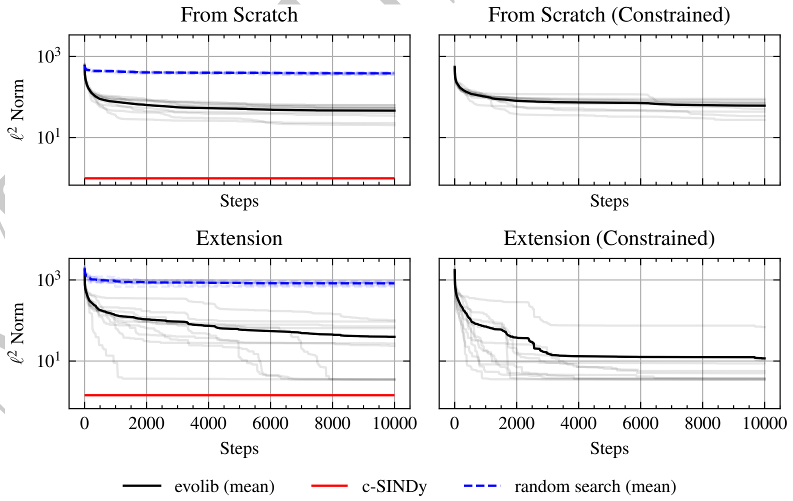


Fig. 4: Convergence of evolving libraries for learning the Wnt pathway and loss of the solution found by coupled SINDy. In the constrained case, coupled SINDy is not applicable and we deemed a random search uninformative.

6 Discussion

Our results demonstrate that the genetic evolution of libraries can provide benefits in terms of model parsimony and fit. In the smaller case studies, evolib enables the recovery of the ground truth model, which was not possible with coupled SINDy alone. However, evolib was not able to recover a biologically viable structure for the complex Wnt pathway, even though achieving a visually accurate fit to the reference trajectory in some scenarios. Partially, the missing ground truth structures can be explained by our truncation of the time-series data to a smaller interval. Note also that we here set the hyperparameters of the genetic search, e.g., population size, to values we determined manually (cf. Appendix A). We leave a systematic (and costly) exploration of hyperparameters, which may be able to achieve slightly better results, as future work. Generally, however, we attribute the mismatch between the learned and the ground truth model to the fact that, in most cases, there is no single model that exclusively explains a given time series. Rather, it is known that reaction systems are not *identifiable* [9], i.e., multiple models fit the same data, necessitating an informed selection. This is why background knowledge, e.g., about the possible interactions, known reactions, and model-level constraints, plays an important role in ensuring that learned models are grounded in the (known) laws of science, and are coherent with the current knowledge about mechanisms. To this end, evolib enables the inclusion of model-level constraints, in addition to constraints on the reaction level possible with (c-)SINDy. Here, we explored how both constraints can be combined to limit reactions to occur within compartments while encouraging the inclusion of shuttling reactions between them. Future work may adjust the fitness function to encourage libraries where the species/reactions match an ontology or are aligned with information gathered from biological databases. In fact, some methods, such as [1], approach the problem from the perspective of knowledge rather than data by automatically constructing models from facts found in literature databases.

Whereas a regression considering all possible reactions at once often leads to very long model descriptions, the GA produced much more sparse solutions, as it allows enforcing a certain maximum number of reactions. This leads to a tradeoff between model complexity and goodness of fit. Limiting model complexity results in humanly comprehensible models, but typically less accurate predictions, which, however, may be beneficial in cases with low-quality data (cf. the predator-prey case study in Section 5). Further, in contrast to the much shorter model descriptions learned by evolib in our Wnt case study, the models learned by c-SINDy could not be readily integrated with LSODA, yielding errors regarding numerical precision. These may be attributed to the presence of singularities, extreme oscillations or due to the fact that LSODA is not switching effectively between stiff and non-stiff regimes.

In our evaluation, we used relatively ideal conditions to test the principal capabilities of the approach. However, in a real-world scenario, measurement data would be available only for a subset of species and be subject to noise. Handling these scenarios is a major limitation of SINDy and, thus, also a limitation of

Manuscript

our approach. Due to its relevance, there have been attempts to also include unmeasured variables, e.g., by using dynamic mode decomposition (DMD) to identify those variables before employing SINDy [6] or implicitly in a black-box manner by training neural networks to act as correction terms [30].

In terms of execution time, we did not observe a speedup using the GA on smaller libraries as opposed to c-SINDy on a very large library. Theoretically, assuming coupled SINDy takes $f(n_a)$ steps (n_a is the library size), its wrapping in a GA results in $cf(n_b)$ steps for some $n_b << n_a$ (n_b is the sublibrary size). The constant c is determined by the number of generations and individuals in each generation. Note that the fitness calculations are easily parallelizable, reducing the factor c. While not changing the complexity class, evolib may still provide a speedup in practice when n_b and c can be chosen small enough. Here, we work with a first prototypical implementation and leave a comprehensive empirical evaluation of the execution time performance as future work.

7 Conclusion

We demonstrated the use of a genetic algorithm together with an extension of SINDy to the case of coupled differential equations, which allows learning compact biochemical reaction models (with many species) from time-series data. Our results show the benefits of the approach for two small and one very large reaction system to be learned. In the smaller cases, the approach can recover the ground truth model. For the case of identifying a very large Wnt model, where coupled SINDy failed to learn models amenable to numerical integration, it learned well-fitting parsimonious models, but they lack biological meaning. Thus, we believe that including background knowledge (inductive bias) is essential to learning a model that can serve as a useful theory to explain the mechanisms at work. Our approach offers the ability to include background knowledge at the level of the model as well as on the level of individual reactions. Although our first and rather simple restrictions that enforced a compartment structure and encouraged shuttling reactions between compartments did not suffice to learn a biologically meaningful Wnt signaling model, we will further pursue this line of research as particularly promising. In addition, we plan to evaluate the method in comparison to other methods for learning reactions, such as the Reactionet LASSO [17] (e.g., by only considering the mean) or Reactmine [24]. For that purpose, identifying a set of suitable benchmark models/data sets to test different capabilities will be essential. In addition, to make these approaches widely applicable in systems biology, the possibility of inferring (time series for) unmeasured species needs further research. This may, for example, be done by including a pre-processing step [6].

Acknowledgments. JNK and AU acknowledge the funding of the DFG Project GrEASE (grant number 320435134). KB acknowledges the funding of the ARC Centre of Excellence for Plant Success in Nature and Agriculture CE 200100015. The authors thank Fiete Haack for many helpful discussions, particularly over his model of the Wnt pathway, and his feedback on the evaluation.

Manuscript

References

1. Ahmed, Y., Telmer, C., Miskov-Zivanov, N.: Accordion: Clustering and selecting relevant data for guided network extension and query answering. arXiv preprint arXiv:2002.05748 (2020). https://doi.org/10.48550/arXiv.2002.05748
2. Askari, E., Crevecoeur, G.: Evolutionary sparse data-driven discovery of multibody system dynamics. Multibody System Dynamics **58**, 197–226 (6 2023). https://doi.org/10.1007/s11044-023-09901-z
3. Bortolussi, L., Cairoli, F., Klein, J., Petrov, T.: Data-Driven Inference of Chemical Reaction Networks via Graph-Based Variational Autoencoders, pp. 143–147. Springer Nature Switzerland (9 2023). https://doi.org/10.1007/978-3-031-43835-6_10
4. Boutillier, P., Maasha, M., Li, X., Medina-Abarca, H.F., Krivine, J., Feret, J., Cristescu, I., Forbes, A.G., Fontana, W.: The kappa platform for rule-based modeling. Bioinformatics **34**(13), i583–i592 (2018). https://doi.org/10.1093/bioinformatics/bty272
5. Bro, R., De Jong, S.: A fast non-negativity-constrained least squares algorithm. Journal of Chemometrics **11**(5), 393–401 (1997). https://doi.org/10.1002/(SICI)1099-128X(199709/10)11:5<393::AID-CEM483>3.0.CO;2-L
6. Brummer, A.B., Xella, A., Woodall, R., Adhikarla, V., Cho, H., Gutova, M., Brown, C.E., Rockne, R.C.: Data driven model discovery and interpretation for car t-cell killing using sparse identification and latent variables. Frontiers in Immunology **14** (2023). https://doi.org/10.3389/fimmu.2023.1115536
7. Brunton, S.L., Proctor, J.L., Kutz, J.N.: Discovering governing equations from data by sparse identification of nonlinear dynamical systems. Proceedings of the National Academy of Sciences **113**, 3932–3937 (4 2016). https://doi.org/10.1073/pnas.1517384113
8. Burrage, P.M., Weerasinghe, H.N., Burrage, K.: Using a library of chemical reactions to fit systems of ordinary differential equations to agent-based models: a machine learning approach. Numerical Algorithms (1 2024). https://doi.org/10.1007/s11075-023-01737-0
9. Craciun, G., Pantea, C.: Identifiability of chemical reaction networks. Journal of Mathematical Chemistry **44**(1), 244–259 (2008). https://doi.org/10.1007/s10910-007-9307-x
10. Daniels, B.C., Nemenman, I.: Automated adaptive inference of phenomenological dynamical models. Nature Communications **6** (8 2015). https://doi.org/10.1038/ncomms9133
11. Faeder, J.R., Blinov, M.L., Hlavacek, W.S.: Rule-based modeling of biochemical systems with bionetgen. In: Systems Biology, pp. 113–167. Springer (2009). https://doi.org/10.1007/978-1-59745-525-1_5
12. Fasel, U., Kutz, J.N., Brunton, B.W., Brunton, S.L.: Ensemble-sindy: Robust sparse model discovery in the low-data, high-noise limit, with active learning and control. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences **478** (4 2022). https://doi.org/10.1098/rspa.2021.0904
13. Großmann, G., Zimmerlin, J., Backenköhler, M., Wolf, V.: Unsupervised relational inference using masked reconstruction. Applied Network Science **8**(1), 18 (2023). https://doi.org/10.1007/s41109-023-00542-x
14. Haack, F., Lemcke, H., Ewald, R., Rharass, T., Uhrmacher, A.M.: Spatio-temporal model of endogenous ros and raft-dependent wnt/beta-catenin signaling driving cell fate commitment in human neural progenitor cells. PLOS Computational Biology **11**(3), 1–28 (03 2015). https://doi.org/10.1371/journal.pcbi.1004106

Manuscript

15. Helms, T., Warnke, T., Maus, C., Uhrmacher, A.M.: Semantics and efficient simulation algorithms of an expressive multilevel modeling language. ACM Transactions on Modeling and Computer Simulation (TOMACS) **27**(2), 1–25 (2017). https://doi.org/https://doi.org/10.1145/2998499
16. Keating, S.M., Waltemath, D., König, M., Zhang, F., Dräger, A., Chaouiya, C., Bergmann, F.T., Finney, A., Gillespie, C.S., Helikar, T., et al.: Sbml level 3: an extensible format for the exchange and reuse of biological models. Molecular systems biology **16**(8), e9110 (2020). https://doi.org/10.15252/msb.20199110
17. Klimovskaia, A., Ganscha, S., Claassen, M.: Sparse regression based structure learning of stochastic reaction networks from single cell snapshot time series. PLOS Computational Biology **12**, e1005234 (12 2016). https://doi.org/10.1371/journal.pcbi.1005234
18. Koza, J.R., Mydlowec, W., Lanza, G., Yu, J., Keane, M.A.: Reverse Engineering of Metabolic Pathways From Observed Data Using Genetic Programming, pp. 434–445. World Scientific (2000). https://doi.org/10.1142/9789814447362_0043
19. Kozin, F., Natke, H.: System identification techniques. Structural safety **3**(3-4), 269–316 (1986). https://doi.org/10.1016/0167-4730(86)90006-8
20. Kramer, O.: Genetic Algorithms, pp. 11–19. Springer International Publishing, Cham (2017). https://doi.org/10.1007/978-3-319-52156-5_2
21. Lee, E., Salic, A., Krüger, R., Heinrich, R., Kirschner, M.W.: The roles of apc and axin derived from experimental and theoretical analysis of the wnt pathway. PLoS biology **1**(1), e10 (2003). https://doi.org/10.1371/journal.pbio.0000010
22. Mangan, N.M., Brunton, S.L., Proctor, J.L., Kutz, J.N.: Inferring biological networks by sparse identification of nonlinear dynamics. IEEE Transactions on Molecular, Biological and Multi-Scale Communications **2**, 52–63 (6 2016). https://doi.org/10.1109/tmbmc.2016.2633265
23. Manzi, M., Vasile, M.: Orbital anomaly reconstruction using deep symbolic regression (October 2020), 71st International Astronautical Congress, IAC 2020
24. Martinelli, J., Grignard, J., Soliman, S., Ballesta, A., Fages, F.: Reactmine: a statistical search algorithm for inferring chemical reactions from time series data. arXiv preprint arXiv:2209.03185v2 (Sep 2022), http://arxiv.org/abs/2209.03185v2
25. Milgroom, M.G.: Epidemiology and SIR Models, pp. 253–268. Springer International Publishing, Cham (2023). https://doi.org/10.1007/978-3-031-38941-2_16
26. Nobile, M.S., Besozzi, D., Cazzaniga, P., Pescini, D., Mauri, G.: Reverse engineering of kinetic reaction networks by means of cartesian genetic programming and particle swarm optimization. In: 2013 IEEE Congress on Evolutionary Computation (CEC). IEEE (6 2013). https://doi.org/10.1109/cec.2013.6557752
27. NumPy team an contributors: Numpy. Version 1.24.3 (2023-02-19), https://numpy.org/
28. Parker, M., Kamenev, A.: Extinction in the lotka-volterra model. Phys. Rev. E **80**, 021129 (Aug 2009). https://doi.org/10.1103/PhysRevE.80.021129
29. Petzold, L.: Automatic selection of methods for solving stiff and nonstiff systems of ordinary differential equations. SIAM Journal on Scientific and Statistical Computing **4**(1), 136–148 (1983). https://doi.org/10.1137/0904010
30. Rackauckas, C., Ma, Y., Martensen, J., Warner, C., Zubov, K., Supekar, R., Skinner, D., Ramadhan, A., Edelman, A.: Universal differential equations for scientific machine learning. arXiv preprint arXiv:2001.04385v4 (Jan 2020). https://doi.org/10.48550/arXiv.2001.04385
31. SciPy team and contributors: Scipy. Verison 1.10.1 (2023-02-19), https://scipy.org/

Manuscript

32. Soliman, S., Heiner, M.: A unique transformation from ordinary differential equations to reaction networks. PloS one **5**(12), e14284 (2010). https://doi.org/10.1371/journal.pone.0014284
33. Spitzer, M.H., Nolan, G.P.: Mass cytometry: Single cells, many features. Cell **165**(4), 780–791 (2016). https://doi.org/10.1016/j.cell.2016.04.019
34. Staehlke, S., Haack, F., Waldner, A.C., Koczan, D., Moerke, C., Mueller, P., Uhrmacher, A.M., Nebe, J.B.: Ros dependent wnt/β-catenin pathway and its regulation on defined micro-pillars—a combined in vitro and in silico study. Cells **9**(8) (2020). https://doi.org/10.3390/cells9081784
35. Székely, T., Burrage, K.: Stochastic simulation in systems biology. Computational and Structural Biotechnology Journal **12**(20), 14–25 (2014). https://doi.org/https://doi.org/10.1016/j.csbj.2014.10.003
36. Tibshirani, R.: Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. Series B (Methodological) **58**(1), 267–288 (1996), http://www.jstor.org/stable/2346178

Manuscript

A Complete List of Experiment (Hyper-)parameters

parameter↓/model→	SIR	Predator-Prey	Wnt	Wnt-X	Wnt (cstr.)	Wnt-X (cstr.)
replications	10					
measurement points	100		78			
measurement interval	0.0 – 1.0	0.0 – 2.0	16.0 – 500.0			
measurement steps	equidistant		determined by LSODA			
sublibrary size	2	3	60	20	60	20
population size	100		200			
max steps	100		10000			
crossover probability	0.8					
mutation probability	0.2					
crossover points	1		5			
num parents	10					
max. number of reactants	2					
max. number of products	3		2			

Table 1: Hyperparameters used for evolib. The random search and c-SINDy use the same parameters where applicable. In particular, for the non-negative least squares, the maximum number of iterations was limited to 10^8. Abbreviations: X extended, cstr. constrained.

B Learned Model's Trajectories for the Wnt Pathway

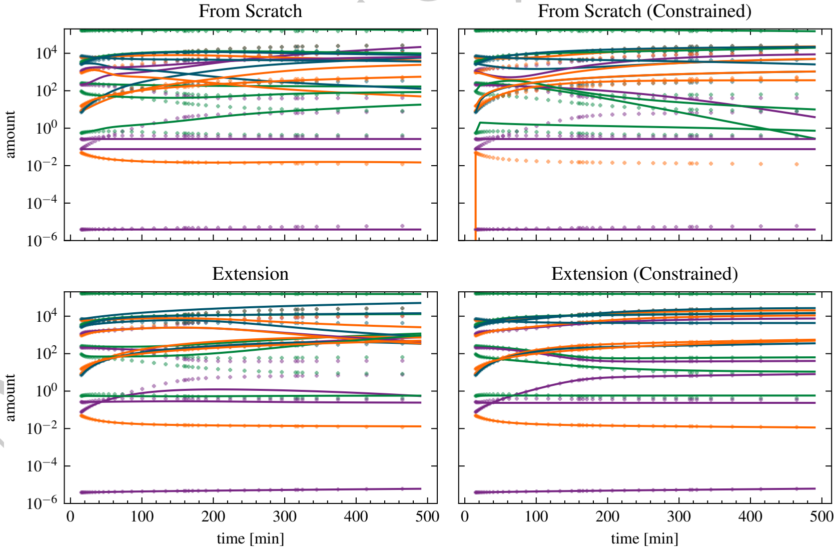


Fig. 5: Results from simulating the models learned with evolib for the Wnt pathway. The + symbols mark measurement points, and the lines are the trajectories simulated for the 19 species. Note that, for clarity, we omit the labeling of species, and only every second measurement point is shown. The integration of the (large) models produced by c-SINDy resulted in errors due to numerical problems.

C Learned Models for the Wnt Pathway

$$\begin{array}{rcl}
\mathit{Bcata}+\mathit{NucTCF} & \xrightarrow{0.00022} & \mathit{NucBcata}+\mathit{NucTCF}\\
\mathit{NrxnO}+\mathit{NucTCF} & \xrightarrow{1.4} & \mathit{NucBcatc}+\mathit{NucTCF}\\
\mathit{Axinp}+\mathit{Bcati} & \xrightarrow{8.4e-05} & 2\mathit{Bcati}\\
\mathbf{Axinp}+\mathbf{Bcati} & \xrightarrow{0.00025} & \mathbf{Axinp}\\
\mathit{Axinp}+\mathit{Bcata} & \xrightarrow{1.3e-05} & \mathit{Axinp}+\mathit{NrxO}\\
\mathit{NucBcati}+\mathit{Counter} & \xrightarrow{7.4e-06} & \mathit{DvlAxinu}+\mathit{Counter}\\
\mathit{ICAT}+\mathit{NucBcati} & \xrightarrow{0.0055} & \mathit{NrxnO}+\mathit{NucBcati}\\
\mathit{Axinu}+\mathit{Bcata} & \xrightarrow{8.3e-05} & \mathit{Axinu}\\
\mathit{NucBcatc} & \xrightarrow{0.023} & \mathit{Dvli}+\mathit{NrxO}\\
\mathit{NrxO}+\mathit{ICAT} & \xrightarrow{0.00065} & \mathit{Bcati}+\mathit{DvlAxinp}\\
\mathit{Axinp}+\mathit{NrxO} & \xrightarrow{1.5e-05} & \mathit{Axinp}+\mathit{ICAT}\\
\mathit{Bcata}+\mathit{DvlAxinu} & \xrightarrow{2.3e-06} & \mathit{Bcati}+\mathit{NucBcata}\\
\mathit{Dvli}+\mathit{NucTCF} & \xrightarrow{8.6e-06} & 2\mathit{NucTCF}\\
\mathit{Dvli}+\mathit{NrxnO} & \xrightarrow{1.4} & \mathit{Bcata}+\mathit{Dvli}\\
\mathit{NrxnO}+\mathit{NucTCF} & \xrightarrow{3.6} & \mathit{NucBcati}+\mathit{NucBcatc}\\
\mathit{ICAT}+\mathit{NucBcatc} & \xrightarrow{0.0019} & \mathit{Bcati}+\mathit{Dvli}\\
\mathit{Dvla} & \xrightarrow{7} & \mathit{Bcata}+\mathit{Dvla}\\
\mathit{NrxnO}+\mathit{DvlAxinp} & \xrightarrow{0.9} & \mathit{NrxnO}+\mathit{DvlAxinu}\\
\mathit{NrxnO}+\mathit{ICAT} & \xrightarrow{4.5e+02} & \mathit{Axinp}+\mathit{NucBcatc}\\
\mathit{NrxnO}+\mathit{DvlNrx} & \xrightarrow{0.0026} & \mathit{NucBcata}+\mathit{NucTCF}\\
\mathit{DvlAxinu}+\mathit{NucBcati} & \xrightarrow{2.8e-06} & \mathit{Bcati}+\mathit{NrxO}\\
\mathit{Dvli}+\mathit{DvlAxinp} & \xrightarrow{2.6e-06} & \mathit{DvlAxinp}+\mathit{NucBcata}\\
\mathit{Bcati} & \xrightarrow{9.3e-05} & \mathit{Axinu}+\mathit{Bcata}\\
\mathit{Axinu}+\mathit{NrxnO} & \xrightarrow{0.89} & \mathit{Axinu}+\mathit{NrxO}\\
\mathit{ICAT} & \xrightarrow{1.1e+02} & \mathit{NrxO}\\
\mathit{Bcati}+\mathit{NrxnO} & \xrightarrow{0.71} & \mathit{NrxnO}+\mathit{NucBcata}\\
\mathit{Dvli}+\mathit{NrxnO} & \xrightarrow{1} & \mathit{Bcati}+\mathit{Dvli}\\
\mathit{Bcata}+\mathit{NucTCF} & \xrightarrow{7.5e-05} & \mathit{NrxnO}+\mathit{NucTCF}\\
\mathit{Axinu}+\mathit{NucBcata} & \xrightarrow{0.00015} & \mathit{Axinu}+\mathit{NucBcati}\\
\mathit{Bcati}+\mathit{DvlAxinu} & \xrightarrow{5.4e-06} & \mathit{Dvli}+\mathit{DvlAxinu}\\
\mathit{Axinp}+\mathit{NucBcati} & \xrightarrow{0.0002} & 2\mathit{Axinp}\\
\mathit{NrxO}+\mathit{ICAT} & \xrightarrow{0.0027} & \mathit{DvlNrx}+\mathit{NucBcata}\\
\mathit{Dvla}+\mathit{ICAT} & \xrightarrow{6.4} & \mathit{Dvla}+\mathit{NucBcati}\\
\mathit{NrxnO}+\mathit{NrxO} & \xrightarrow{0.61} & \mathit{ICAT}+\mathit{NucBcatc}\\
\mathit{NucBcata}+\mathit{NucTCF} & \xrightarrow{0.00088} & \mathit{DvlNrx}+\mathit{NucTCF}\\
\mathit{DvlNrx}+\mathit{ICAT} & \xrightarrow{0.00037} & \mathit{Bcata}+\mathit{ICAT}\\
\mathit{DvlNrx} & \xrightarrow{6.4e-06} & \mathit{Counter}\\
\mathit{Axinu}+\mathit{Bcati} & \xrightarrow{1.6e-06} & \mathit{Bcati}+\mathit{Dvli}\\
\mathit{NrxO} & \xrightarrow{0.0041} & \mathit{NucBcati}\\
\mathit{DvlAxinu}+\mathit{Counter} & \xrightarrow{3.7e-05} & \mathit{Bcati}+\mathit{Counter}\\
\mathit{Bcata}+\mathit{NucTCF} & \xrightarrow{2.7e-05} & \mathit{Bcati}+\mathit{Dvli}\\
\mathit{NucBcati} & \xrightarrow{0.0043} & \mathit{Bcata}+\mathit{NrxO}\\
\mathit{DvlNrx} & \xrightarrow{0.0031} & \mathit{NrxnO}+\mathit{DvlNrx}\\
\mathit{Dvli} & \xrightarrow{0.0075} & \mathit{DvlNrx}
\end{array}$$

(+ 9 more)

(a) Evolib from scratch

$$\begin{array}{rcl}
 & \xrightarrow{6e+02} & \mathbf{Bcata}\\
\mathbf{NucBcata} & \xrightarrow{0.14} & \mathbf{Bcata}\\
\mathbf{Axinp}+\mathbf{Bcati} & \xrightarrow{0.00021} & \mathbf{Axinp}\\
\mathbf{Bcati} & \xrightarrow{0.00011} & \\
\mathbf{Axinu} & \xrightarrow{0.03} & \mathbf{Axinp}\\
\mathbf{NrxnO}+\mathbf{Rosa} & \xrightarrow{5e+02} & \mathbf{NrxO}\\
\mathbf{NrxO} & \xrightarrow{0.02} & \mathbf{NrxnO}\\
\mathbf{Dvli} & \xrightarrow{0.0005} & \mathbf{Dvla}\\
\mathbf{DvlAxinu} & \xrightarrow{0.068} & \mathbf{Dvla}+\mathbf{Axinu}\\
\mathbf{Counter} & \xrightarrow{0.2} & \mathbf{Counter}+\mathbf{Rosa}\\
 & \xrightarrow{1} & \mathbf{Counter}\\
\mathbf{Dvla}+\mathbf{Axinp} & \xrightarrow{0.075} & \mathbf{DvlAxinp}\\
\mathbf{Axinp}+\mathbf{Bcata} & \xrightarrow{0.00021} & \mathbf{Axinp}\\
\mathbf{NucBcati} & \xrightarrow{0.00011} & \\
\mathbf{Dvla}+\mathbf{NrxnO} & \xrightarrow{22} & \mathbf{DvlNrx}\\
\mathbf{NucBcata}+\mathbf{NucTCF} & \xrightarrow{0.002} & \mathbf{NucBcatc}\\
\mathbf{Axinp} & \xrightarrow{0.03} & \mathbf{Axinu}\\
\mathbf{Axinp} & \xrightarrow{0.0045} & \\
\mathbf{NucBcatc} & \xrightarrow{0.00011} & \\
\mathbf{Bcata} & \xrightarrow{0.055} & \mathbf{NucBcata}\\
\mathbf{DvlNrx}+\mathbf{Rosa} & \xrightarrow{3.2e+02} & \mathbf{Dvli}+\mathbf{NrxO}\\
\mathbf{Dvla}+\mathbf{Axinu} & \xrightarrow{0.075} & \mathbf{DvlAxinu}\\
\mathbf{Bcati} & \xrightarrow{0.055} & \mathbf{NucBcati}\\
\mathbf{NucBcatc} & \xrightarrow{0.014} & \mathbf{NucBcata}+\mathbf{NucTCF}\\
\mathbf{DvlNrx} & \xrightarrow{0.023} & \mathbf{Dvli}+\mathbf{NrxnO}\\
\mathbf{NucBcatc} & \xrightarrow{0.0004} & \mathbf{NucBcatc}+\mathbf{Axinu}\\
\mathbf{Dvla} & \xrightarrow{0.5} & \mathbf{Dvli}\\
\mathbf{NucBcata} & \xrightarrow{0.00011} & \\
\mathbf{Bcata} & \xrightarrow{0.00011} & \\
\mathbf{Axinu} & \xrightarrow{0.0045} & \\
\mathbf{ICAT} & \xrightarrow{0.055} & \mathbf{NucICAT}\\
\mathbf{NucICAT} & \xrightarrow{0.14} & \mathbf{ICAT}\\
\mathbf{DvlAxinp} & \xrightarrow{0.068} & \mathbf{Dvla}+\mathbf{Axinp}\\
\mathbf{Dvli}+\mathbf{NrxnO} & \xrightarrow{22} & \mathbf{DvlNrx}\\
\mathbf{NucBcati} & \xrightarrow{0.14} & \mathbf{Bcati}\\
\hline
\mathit{NrxnO}+\mathit{NucBcati} & \xrightarrow{0.0025} & \mathit{DvlNrx}+\mathit{NucBcatc}\\
 & \xrightarrow{4.1} & \mathit{Bcati}+\mathit{Rosa}\\
\mathit{Bcata}+\mathit{ICAT} & \xrightarrow{0.0017} & \mathit{ICAT}\\
\mathit{Bcati}+\mathit{ICAT} & \xrightarrow{2.9e-06} & \mathit{NrxnO}+\mathit{NucICAT}\\
 & \xrightarrow{0.18} & \mathit{Bcata}+\mathit{DvlNrx}\\
\mathit{NucBcati} & \xrightarrow{0.00067} & \mathit{Bcata}+\mathit{Bcati}\\
\mathit{Bcati} & \xrightarrow{0.032} & \mathit{Bcata}\\
\mathit{Bcata}+\mathit{ICAT} & \xrightarrow{0.098} & \mathit{Bcati}+\mathit{ICAT}\\
\mathit{NucBcati} & \xrightarrow{0.00069} & 2\mathit{NucBcati}\\
\mathit{DvlNrx}+\mathit{NucICAT} & \xrightarrow{6.6e-06} & \mathit{Bcata}+\mathit{NucICAT}\\
 & \xrightarrow{1e+02} & \mathit{Rosa}+\mathit{NucTCF}\\
\mathit{DvlAxinp}+\mathit{NucTCF} & \xrightarrow{2.7e-05} & \mathit{Dvli}+\mathit{NrxnO}\\
 & \xrightarrow{96} & \mathbf{Rosa}
\end{array}$$

(+ 7 more)

(b) Evolib extension

$$\begin{array}{rcl}
 & \xrightarrow{6e+02} & \mathit{Bcata}\\
\mathit{NucBcata} & \xrightarrow{0.14} & \mathit{Bcata}\\
\mathit{Axinp}+\mathit{Bcati} & \xrightarrow{0.00021} & \mathit{Axinp}\\
\mathit{Bcati} & \xrightarrow{0.00011} & \\
\mathit{NucICAT}+\mathit{NucBcata} & \xrightarrow{0.1} & \mathit{NucBcati}\\
\mathit{Axinu} & \xrightarrow{0.03} & \mathit{Axinp}\\
\mathit{NrxnO}+\mathit{Rosa} & \xrightarrow{5e+02} & \mathit{NrxO}\\
\mathit{NrxO} & \xrightarrow{0.02} & \mathit{NrxnO}\\
\mathit{Dvli} & \xrightarrow{0.0005} & \mathit{Dvla}\\
\mathit{DvlAxinu} & \xrightarrow{0.068} & \mathit{Dvla}+\mathit{Axinu}\\
\mathit{Counter} & \xrightarrow{0.2} & \mathit{Counter}+\mathit{Rosa}\\
 & \xrightarrow{1} & \mathit{Counter}\\
\mathit{Dvla}+\mathit{Axinp} & \xrightarrow{0.075} & \mathit{DvlAxinp}\\
\mathit{Axinp}+\mathit{Bcata} & \xrightarrow{0.00021} & \mathit{Axinp}\\
\mathit{NucBcati} & \xrightarrow{0.00011} & \\
\mathit{Dvla}+\mathit{NrxnO} & \xrightarrow{22} & \mathit{DvlNrx}\\
 & \xrightarrow{2.5e+02} & \mathit{ICAT}\\
 & \xrightarrow{1e+02} & \mathit{NucTCF}\\
\mathit{NucBcata}+\mathit{NucTCF} & \xrightarrow{0.002} & \mathit{NucBcatc}\\
\mathit{Axinp} & \xrightarrow{0.03} & \mathit{Axinu}\\
\mathit{NucBcatc} & \xrightarrow{0.023} & \\
\mathit{Axinp} & \xrightarrow{0.0045} & \\
\mathit{NucBcatc} & \xrightarrow{0.00011} & \\
\mathit{Bcata} & \xrightarrow{0.055} & \mathit{NucBcata}\\
\mathit{Bcati} & \xrightarrow{0.032} & \mathit{ICAT}+\mathit{Bcata}\\
\mathit{DvlNrx}+\mathit{Rosa} & \xrightarrow{3.2e+02} & \mathit{Dvli}+\mathit{NrxO}\\
\mathit{Dvla}+\mathit{Axinu} & \xrightarrow{0.075} & \mathit{DvlAxinu}\\
\mathit{Bcati} & \xrightarrow{0.055} & \mathit{NucBcati}\\
\mathit{NucBcatc} & \xrightarrow{0.014} & \mathit{NucBcata}+\mathit{NucTCF}\\
\mathit{DvlNrx} & \xrightarrow{0.023} & \mathit{Dvli}+\mathit{NrxnO}\\
\mathit{NucBcatc} & \xrightarrow{0.0004} & \mathit{NucBcatc}+\mathit{Axinu}\\
\mathit{NucBcati} & \xrightarrow{0.032} & \mathit{NucICAT}+\mathit{NucBcata}\\
\mathit{ICAT}+\mathit{Bcata} & \xrightarrow{0.1} & \mathit{Bcati}\\
\mathit{Dvla} & \xrightarrow{0.5} & \mathit{Dvli}\\
\mathit{NucBcata} & \xrightarrow{0.00011} & \\
\mathit{Bcata} & \xrightarrow{0.00011} & \\
\mathit{Axinu} & \xrightarrow{0.0045} & \\
\mathit{ICAT} & \xrightarrow{0.055} & \mathit{NucICAT}\\
\mathit{NucICAT} & \xrightarrow{0.14} & \mathit{ICAT}\\
 & \xrightarrow{2e+02} & \mathit{Rosa}\\
\mathit{DvlAxinp} & \xrightarrow{0.068} & \mathit{Dvla}+\mathit{Axinp}\\
\mathit{Dvli}+\mathit{NrxnO} & \xrightarrow{22} & \mathit{DvlNrx}\\
\mathit{NucBcati} & \xrightarrow{0.14} & \mathit{Bcati}
\end{array}$$

(c) Ground truth

Fig. 6: The models inferred in the Wnt case study (unconstrained) compared to the ground truth model. For the extension, the reactions above the line are fixed. Only reactions with a rate above 10^{-6} are shown, and if applicable the number of excluded reactions is shown in the lower left. Bolded reactions indicate an overlap with the ground truth reactions. Note in particular how in (a) the Axin-induced degradation of β-catenin and in (b) the synthesis of Ros was recovered, which are both central components of the ground truth model of the Wnt pathway.

$$
\begin{array}{rcl}
DvlNrx + ICAT & \xrightarrow{1.1e-05} & DvlAxinp + Counter \\
Bcata + NrxnO & \xrightarrow{0.068} & NucBcata + NucBcati \\
Axinp + Bcata & \xrightarrow{8.3e-06} & Dvli + NrxO \\
Bcata + NrxnO & \xrightarrow{0.11} & Bcati + DvlNrx \\
DvlAxinu & \xrightarrow{0.64} & Bcata + DvlAxinu \\
NucBcatc & \xrightarrow{0.01} & Dvli + ICAT \\
Axinp + DvlNrx & \xrightarrow{2.8e-06} & Axinp + NrxO \\
NucBcata & \xrightarrow{0.0058} & 2NucBcata \\
Axinu + NrxO & \xrightarrow{1.4e-06} & Bcata + DvlAxinu \\
Dvli + ICAT & \xrightarrow{0.008} & Bcata + ICAT \\
Bcata + Rosa & \xrightarrow{2.9e+03} & DvlNrx + Rosa \\
ICAT & \xrightarrow{34} & Dvli + DvlNrx \\
NucBcati + NucTCF & \xrightarrow{0.00028} & Dvli + DvlNrx \\
NucBcata + NucTCF & \xrightarrow{0.00071} & NucTCF \\
Dvla + ICAT & \xrightarrow{4.8} & Dvla + NrxO \\
Bcati + ICAT & \xrightarrow{0.00061} & NucBcati \\
Axinu + Bcata & \xrightarrow{7.5e-05} & DvlNrx \\
Bcata + DvlAxinp & \xrightarrow{3.5e-06} & Dvli + DvlAxinp \\
Axinu + Bcata & \xrightarrow{7.5e-05} & 2Axinu \\
NucBcati + NucTCF & \xrightarrow{0.00021} & Bcati + DvlNrx \\
ICAT & \xrightarrow{76} & NucTCF \\
NrxnO + DvlNrx & \xrightarrow{0.019} & 2NrxnO \\
Axinu + DvlNrx & \xrightarrow{3.3e-06} & Axinu + Bcata \\
DvlNrx + Counter & \xrightarrow{3.8e-06} & Counter \\
 & \xrightarrow{2.5e+02} & DvlNrx \\
NucBcatc & \xrightarrow{0.0041} & Bcata + NrxO \\
DvlNrx + ICAT & \xrightarrow{0.00047} & Bcata + ICAT \\
Axinp + NrxO & \xrightarrow{2.2e-05} & Axinp + DvlNrx \\
NrxnO & \xrightarrow{3.5e+02} & NucBcata + NucTCF \\
ICAT & \xrightarrow{79} & 2ICAT \\
Axinp + Bcata & \xrightarrow{7.7e-06} & 2Axinp \\
NrxnO + ICAT & \xrightarrow{1.9e+03} & Bcata + Dvli \\
NrxnO & \xrightarrow{7.8e+02} & Bcata + NrxO \\
\mathbf{NucBcata} & \xrightarrow{0.12} & \mathbf{Bcata} \\
Axinu + Bcati & \xrightarrow{0.00011} & Axinu \\
Dvla & \xrightarrow{7.7} & Dvla + DvlNrx \\
NucBcata + NucTCF & \xrightarrow{0.00035} & NucBcati + NucBcatc \\
\mathbf{Bcata} & \xrightarrow{0.047} & \mathbf{NucBcata}
\end{array}
$$

(+ 9 more)

(a) Evolib from scratch

$$
\begin{array}{rcl}
 & \xrightarrow{6e+02} & \mathbf{Bcata} \\
\mathbf{NucBcata} & \xrightarrow{0.14} & \mathbf{Bcata} \\
\mathbf{Axinp + Bcati} & \xrightarrow{0.00021} & \mathbf{Axinp} \\
\mathbf{Bcati} & \xrightarrow{0.00011} & \\
\mathbf{Axinu} & \xrightarrow{0.03} & \mathbf{Axinp} \\
\mathbf{NrxnO + Rosa} & \xrightarrow{5e+02} & \mathbf{NrxO} \\
\mathbf{NrxO} & \xrightarrow{0.02} & \mathbf{NrxnO} \\
\mathbf{Dvli} & \xrightarrow{0.0005} & \mathbf{Dvla} \\
\mathbf{DvlAxinu} & \xrightarrow{0.068} & \mathbf{Dvla + Axinu} \\
\mathbf{Counter} & \xrightarrow{0.2} & \mathbf{Counter + Rosa} \\
 & \xrightarrow{1} & \mathbf{Counter} \\
\mathbf{Dvla + Axinp} & \xrightarrow{0.075} & \mathbf{DvlAxinp} \\
\mathbf{Axinp + Bcata} & \xrightarrow{0.00021} & \mathbf{Axinp} \\
\mathbf{NucBcati} & \xrightarrow{0.00011} & \\
\mathbf{Dvla + NrxnO} & \xrightarrow{22} & \mathbf{DvlNrx} \\
\mathbf{NucBcata + NucTCF} & \xrightarrow{0.002} & \mathbf{NucBcatc} \\
\mathbf{Axinp} & \xrightarrow{0.03} & \mathbf{Axinu} \\
\mathbf{Axinp} & \xrightarrow{0.0045} & \\
\mathbf{NucBcatc} & \xrightarrow{0.00011} & \\
\mathbf{Bcata} & \xrightarrow{0.055} & \mathbf{NucBcata} \\
\mathbf{DvlNrx + Rosa} & \xrightarrow{3.2e+02} & \mathbf{Dvli + NrxO} \\
\mathbf{Dvla + Axinu} & \xrightarrow{0.075} & \mathbf{DvlAxinu} \\
\mathbf{Bcati} & \xrightarrow{0.055} & \mathbf{NucBcati} \\
\mathbf{NucBcatc} & \xrightarrow{0.014} & \mathbf{NucBcata + NucTCF} \\
\mathbf{DvlNrx} & \xrightarrow{0.023} & \mathbf{Dvli + NrxnO} \\
\mathbf{NucBcatc} & \xrightarrow{0.0004} & \mathbf{NucBcatc + Axinu} \\
\mathbf{Dvla} & \xrightarrow{0.5} & \mathbf{Dvli} \\
\mathbf{NucBcata} & \xrightarrow{0.00011} & \\
\mathbf{Bcata} & \xrightarrow{0.00011} & \\
\mathbf{Axinu} & \xrightarrow{0.0045} & \\
\mathbf{ICAT} & \xrightarrow{0.055} & \mathbf{NucICAT} \\
\mathbf{NucICAT} & \xrightarrow{0.14} & \mathbf{ICAT} \\
\mathbf{DvlAxinp} & \xrightarrow{0.068} & \mathbf{Dvla + Axinp} \\
\mathbf{Dvli + NrxnO} & \xrightarrow{22} & \mathbf{DvlNrx} \\
\mathbf{NucBcati} & \xrightarrow{0.14} & \mathbf{Bcati} \\
\hline
Dvla + NrxnO & \xrightarrow{7} & Dvla + DvlNrx \\
NrxO & \xrightarrow{0.0094} & NrxO + Rosa \\
Bcati & \xrightarrow{0.032} & Bcata + Bcati \\
NrxnO + Rosa & \xrightarrow{1.7e+09} & NrxnO \\
 & \xrightarrow{2.2e+02} & Bcati \\
 & \xrightarrow{26} & Bcati + Rosa \\
Dvli + NrxnO & \xrightarrow{0.0011} & Bcata + DvlNrx \\
ICAT & \xrightarrow{3.7} & Rosa + ICAT \\
Bcata + ICAT & \xrightarrow{0.1} & ICAT \\
 & \xrightarrow{1e+02} & \mathbf{NucTCF} \\
NucBcatc & \xrightarrow{0.023} & Rosa \\
NrxO + Rosa & \xrightarrow{14} & Dvli + NrxO \\
NrxnO & \xrightarrow{6.2e+03} & NrxnO + Rosa \\
Dvli + Dvla & \xrightarrow{9.8e-06} & Dvla + Rosa \\
Rosa + DvlAxinp & \xrightarrow{6.9e+03} & DvlAxinp
\end{array}
$$

(+ 3 more)

(b) Evolib extension

$$
\begin{array}{rcl}
 & \xrightarrow{6e+02} & Bcata \\
NucBcata & \xrightarrow{0.14} & Bcata \\
Axinp + Bcati & \xrightarrow{0.00021} & Axinp \\
Bcati & \xrightarrow{0.00011} & \\
NucICAT + NucBcata & \xrightarrow{0.1} & NucBcati \\
Axinu & \xrightarrow{0.03} & Axinp \\
NrxnO + Rosa & \xrightarrow{5e+02} & NrxO \\
NrxO & \xrightarrow{0.02} & NrxnO \\
Dvli & \xrightarrow{0.0005} & Dvla \\
DvlAxinu & \xrightarrow{0.068} & Dvla + Axinu \\
Counter & \xrightarrow{0.2} & Counter + Rosa \\
 & \xrightarrow{1} & Counter \\
Dvla + Axinp & \xrightarrow{0.075} & DvlAxinp \\
Axinp + Bcata & \xrightarrow{0.00021} & Axinp \\
NucBcati & \xrightarrow{0.00011} & \\
Dvla + NrxnO & \xrightarrow{22} & DvlNrx \\
 & \xrightarrow{2.5e+02} & ICAT \\
 & \xrightarrow{1e+02} & NucTCF \\
NucBcata + NucTCF & \xrightarrow{0.002} & NucBcatc \\
Axinp & \xrightarrow{0.03} & Axinu \\
NucBcatc & \xrightarrow{0.023} & \\
Axinp & \xrightarrow{0.0045} & \\
NucBcatc & \xrightarrow{0.00011} & \\
Bcata & \xrightarrow{0.055} & NucBcata \\
Bcati & \xrightarrow{0.032} & ICAT + Bcata \\
DvlNrx + Rosa & \xrightarrow{3.2e+02} & Dvli + NrxO \\
Dvla + Axinu & \xrightarrow{0.075} & DvlAxinu \\
Bcati & \xrightarrow{0.055} & NucBcati \\
NucBcatc & \xrightarrow{0.014} & NucBcata + NucTCF \\
DvlNrx & \xrightarrow{0.023} & Dvli + NrxnO \\
NucBcatc & \xrightarrow{0.0004} & NucBcatc + Axinu \\
NucBcati & \xrightarrow{0.032} & NucICAT + NucBcata \\
ICAT + Bcata & \xrightarrow{0.1} & Bcati \\
Dvla & \xrightarrow{0.5} & Dvli \\
NucBcata & \xrightarrow{0.00011} & \\
Bcata & \xrightarrow{0.00011} & \\
Axinu & \xrightarrow{0.0045} & \\
ICAT & \xrightarrow{0.055} & NucICAT \\
NucICAT & \xrightarrow{0.14} & ICAT \\
 & \xrightarrow{2e+02} & Rosa \\
DvlAxinp & \xrightarrow{0.068} & Dvla + Axinp \\
Dvli + NrxnO & \xrightarrow{22} & DvlNrx \\
NucBcati & \xrightarrow{0.14} & Bcati
\end{array}
$$

(c) Ground truth

Fig. 7: The models inferred in the Wnt case study (constrained) compared to the ground truth model. For the extension, the reactions above the line are fixed. Only reactions with a rate above 10^{-6} are shown, and if applicable, the number of excluded reactions is shown in the lower left. Bolded reactions indicate an overlap with the ground truth reactions. Note in particular how in (a) the shuttling of β-catenin in and out of the nucleus and in (b) the production of TCF was recovered.