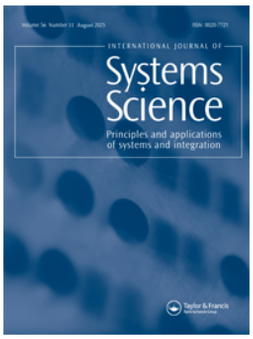

International Journal of Systems Science

ISSN: 0020-7721 (Print) 1464-5319 (Online) Journal homepage: www.tandfonline.com/journals/tsys20

Sparse identification and mathematical framework for analyzing metabolic-regulatory networks

Neveen Ali Eshtewy, Ali Forootani, Shumaila Noreen & Mohammad Khosravi

To cite this article: Neveen Ali Eshtewy, Ali Forootani, Shumaila Noreen & Mohammad Khosravi (19 Jun 2025): Sparse identification and mathematical framework for analyzing metabolic-regulatory networks, International Journal of Systems Science, DOI: 10.1080/00207721.2025.2520353

To link to this article: https://doi.org/10.1080/00207721.2025.2520353

Published online: 19 Jun 2025.

INTERNATIONAL JOURNAL OF SYSTEMS SCIENCE
https://doi.org/10.1080/00207721.2025.2520353

Taylor & Francis
Taylor & Francis Group

Sparse identification and mathematical framework for analyzing metabolic-regulatory networks

Neveen Ali Eshtewy[a], Ali Forootani[b], Shumaila Noreen[c] and Mohammad Khosravi[d]

[a]Faculty of Sciences, Arish University, Al-Arish, North Sinai, Egypt; [b]Department of Bio-Energy, Helmholtz Center for Environmental Research-UFZ, Leipzig, Germany; [c]College of Arts and Sciences, University of Nizwa, Nizwa, Oman; [d]Delft Center for Systems and Control, Delft, The Netherlands

ABSTRACT

We present a continuous modelling framework for simulating the dynamics of metabolic-regulatory networks (MRNs), designed to overcome the scalability limitations of traditional hybrid models. Hybrid approaches, often based on Boolean logic to represent regulatory interactions, become computationally intractable as the number of regulatory proteins increases, due to an exponential growth in discrete modes and transitions. To address this, our framework replaces discrete logic with smooth Hill functions, enabling the approximation of switch-like regulatory behaviour without introducing combinatorial complexity. This continuous formulation maintains the biological interpretability of hybrid models while greatly enhancing computational efficiency. Parameter estimation, a common bottleneck in continuous models, is simplified in our approach by requiring fewer kinetic parameters than typical hybrid models. We further employ sparse-based system identification, a data-driven technique that efficiently infers network dynamics by selecting a minimal set of nonlinear terms. This method avoids exhaustive search procedures and yields interpretable kinetic models. Applied to MRNs, our framework demonstrates the ability to capture essential regulatory mechanisms with reduced complexity and improved scalability.

ARTICLE HISTORY
Received 21 December 2024
Accepted 8 June 2025

KEYWORDS
Continuous model; hybrid system; kinetic model; metabolic-regulatory network; sparse identification

1. Introduction

Computational techniques in systems biology provide powerful methodologies for uncovering the fundamental mechanisms that govern cellular regulation and metabolic processes. Despite significant progress, the integration of metabolic and regulatory systems remains highly complex and only partially understood (Taou et al., 2018). Mathematical modelling plays a crucial role in bridging this gap by enabling the simultaneous study of metabolic pathways, dynamic enzyme availability, and regulatory control processes (Politano et al., 2014).

One widely used approach is *regulatory flux balance analysis* (rFBA) (Covert et al., 2001), which combines traditional flux balance analysis (FBA) with regulatory Boolean logic and ordinary differential equations (ODEs). This hybrid framework allows for the simulation of dynamic cellular behaviour by integrating metabolic flux predictions with the influence of regulatory and signalling networks, particularly in model organisms such as *Escherichia coli* (Covert et al., 2008).

To address the limitations of parameter-intensive kinetic models, Marmiesse et al. (2015) introduced *FlexFlux*, a computational tool designed to integrate metabolic and regulatory network analyses at the genome scale. Notably, FlexFlux does not require kinetic parameters, making it especially suitable for large-scale models. It employs multi-state qualitative logic to identify regulatory steady states, which are then imposed as constraints in the FBA framework to refine metabolic flux predictions.

In addition to rFBA-based approaches, hybrid systems modelling provides a versatile paradigm for capturing the interactions between discrete regulatory events and continuous metabolic dynamics. Originally developed for embedded systems analysis, the *Hybrid Automaton* framework has been successfully adapted to biological networks (Bortolussi & Policriti, 2008). In particular, in Liu and Bockmayr (2020)

CONTACT Neveen Ali Eshtewy neveen@unizwa.edu.om Faculty of Sciences, Arish University, Al-Arish 45111, North Sinai, Egypt; Ali Forootani aliforootani@ieee.org, ali.forootani@ufz.de College of Arts and Sciences, University of Nizwa, Nizwa 616, Oman

proposed a formal hybrid automaton model that integrates continuous-time metabolic processes with discrete gene regulatory controls. This enables more accurate and analyzable simulations of metabolic-regulatory networks (MRNs), accommodating both time-dependent transitions and complex control logic.

In the case of *Escherichia coli*, MRNs represent a tightly coupled system where transcriptional, translational, and allosteric regulation layers collectively coordinate metabolic activity. These regulatory mechanisms govern gene expression, enzyme synthesis, post-translational modifications, and metabolite concentrations to facilitate robust adaptation to environmental fluctuations and nutrient availability (Martínez-Antonio & Collado-Vides, 2003). Understanding such integrated control is essential for advancing metabolic engineering, synthetic biology, and precision medicine.

Continuous models, typically formulated using systems of ordinary differential equations (ODEs), have been extensively employed to represent the dynamic interplay between metabolic pathways and regulatory mechanisms. Such models offer a mechanistic and time-resolved understanding of system behaviour, enabling the simulation of complex cellular responses under varying environmental or genetic conditions. For instance, the continuous modelling approaches presented in Ali Eshtewy and Scholz (2020) and N. A. S. Eshtewy (2020) are applied to metabolic-genetic networks, providing a detailed framework analogous to regulatory flux balance analysis (rFBA) for capturing both metabolic and regulatory dynamics. The study in Kremling et al. (2018) introduces a model ensemble approach to explore multiple hypotheses concerning the mechanisms underlying carbon catabolite repression (CCR). This ensemble-based framework facilitates a deeper understanding of the diverse regulatory strategies potentially involved in CCR by comparing alternative models against experimental data.

A key challenge in continuous modelling remains the estimation of kinetic parameters, many of which are difficult to measure directly. To address this, a variety of parameter estimation techniques have been developed for biological systems. These methods rely on computational optimisation algorithms to infer unknown parameters by minimising the discrepancy between model predictions and experimental observations (Altman & Krzywinski, 2015; Björck, 2024; Deuflhard, 2011; Enders, 2005; N. A. Eshtewy et al., 2023; Kreutz, 2018; Raue et al., 2013). Examples of such techniques include the Gauss–Newton method (Deuflhard, 2011), which iteratively refines parameter estimates based on local curvature information, least squares regression (Björck, 2024), which minimises squared prediction errors, and maximum likelihood estimation (Enders, 2005), which seeks parameter values that maximise the probability of observing the data given the model.

In N. A. Eshtewy et al. (2023), a model order reduction strategy is integrated with parameter estimation to effectively calibrate a kinetic model for CCR in *Escherichia coli*. This approach enables the accurate inference of uptake rate constants and other kinetic parameters while reducing computational burden, thus improving the tractability and interpretability of large-scale biological models.

Deriving governing equations directly from data remains a central and unresolved challenge in the modelling of complex dynamical systems. This issue is systematically investigated in Brunton et al. (2016a), where sparsity-promoting techniques are combined with machine learning to infer nonlinear dynamical equations from time-series data. By applying sparse regression, the method isolates a minimal set of relevant terms from a large library of candidate nonlinear functions, thereby producing compact models that offer both predictive accuracy and interpretability. This parsimonious modelling approach enables the identification of fundamental system dynamics, even when the data is noisy or partially observed.

An extension of this methodology is proposed in Brunton et al. (2016b), where sparse regression is adapted to identify nonlinear systems subject to external inputs. This formulation allows the discovery of control-influenced dynamics and enhances the applicability of sparse models to a broader class of input-driven systems.

In the domain of biological networks, numerous mathematical techniques have been developed to reconstruct the complex regulatory and metabolic interactions from experimental data (Cakir & Khatibipour, 2014; Chasman et al., 2016; Marbach et al., 2012). These methods typically focus on uncovering the structure and function of gene regulatory networks, metabolic pathways, and signalling cascades. For instance, Mangan et al. (2016) introduces a sparsity-promoting optimisation framework tailored to biological systems, which identifies a key

subset of nonlinear interactions that capture the essential features of metabolic-regulatory dynamics. This approach enables accurate model inference while avoiding the overfitting and redundancy issues common in high-dimensional biological datasets.

1.1. Contributions

In this paper, we introduce a continuous mathematical model for metabolic-regulatory networks (MRNs), inspired by the hybrid automaton framework presented in Liu and Bockmayr (2020). While the hybrid model effectively captures the interplay between discrete regulatory logic and continuous metabolic dynamics, its scalability is limited. As the number of regulatory proteins increases, the number of discrete modes and transition events grows exponentially, scaling as 2^n, where n denotes the number of regulatory rules. This exponential complexity presents significant computational challenges in the modelling and analysis of large-scale MRNs.

To address this limitation, our proposed continuous model replaces Boolean logic with nonlinear Hill functions to describe regulatory interactions. Hill functions offer a smooth and flexible alternative, capable of approximating switch-like behaviours while maintaining continuity and differentiability – essential properties for numerical simulation and optimisation. This formulation enables a more tractable and scalable framework, particularly advantageous when dealing with high regulatory complexity.

Furthermore, we incorporate the implicit-sparse system identification method from Mannan et al. (2015) to analyze and refine our MRN model. This data-driven technique promotes parsimony by selecting a minimal set of governing terms from a large candidate library of nonlinear functions. It effectively captures system dynamics while avoiding the combinatorial explosion associated with explicit rule enumeration. The implicit formulation enables the discovery of algebraic relationships between variables and their derivatives without requiring explicit state reconstruction.

Our continuous model combines Hill-based nonlinearities with polynomial interaction terms to compactly represent regulatory and metabolic dynamics. Although parameter estimation is a well-known challenge in continuous models, our framework reduces this burden by requiring fewer kinetic parameters than

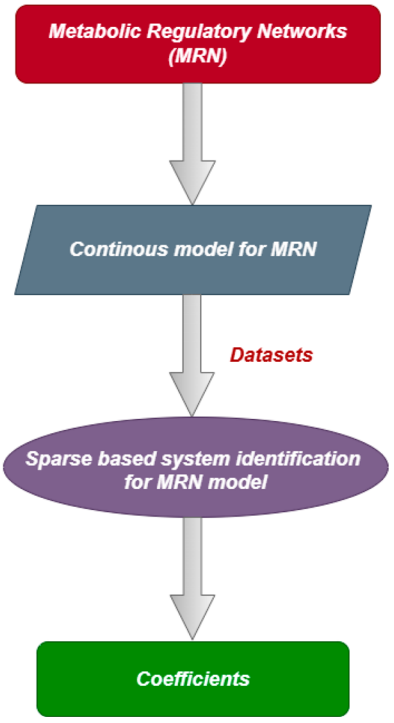

Figure 1. Illustrating the modelling and system identification of the MRNs.

hybrid models. This results in enhanced model interpretability and reduced calibration effort.

MRNs in *Escherichia coli* coordinate gene expression and enzymatic activities in response to internal metabolic states and external environmental cues. These regulatory processes often involve nonlinear, cooperative, and threshold-dependent mechanisms, which are effectively modelled using Hill-type functions (Martínez-Antonio & Collado-Vides, 2003). By applying implicit-sparse system identification to our kinetic model of the MRN in *Escherichia coli*, we efficiently infer the governing equations for the majority of state variables. While most interactions are successfully captured, a small subset of equations remains unidentified due to the complexity and coupling of higher-order nonlinear interactions. Flowchart 1 illustrates the process of identifying coefficients in MRNs. The process begins with the continuous modelling of MRNs, followed by the integration of datasets and the application of sparse-based system identification techniques, ultimately leading to the determination of the model's coefficients.

The structure of this paper is as follows: Section 2 offers an overview of hybrid automata and its key elements as discussed in Liu and Bockmayr (2020). In Section 3, we construct a mathematical model for

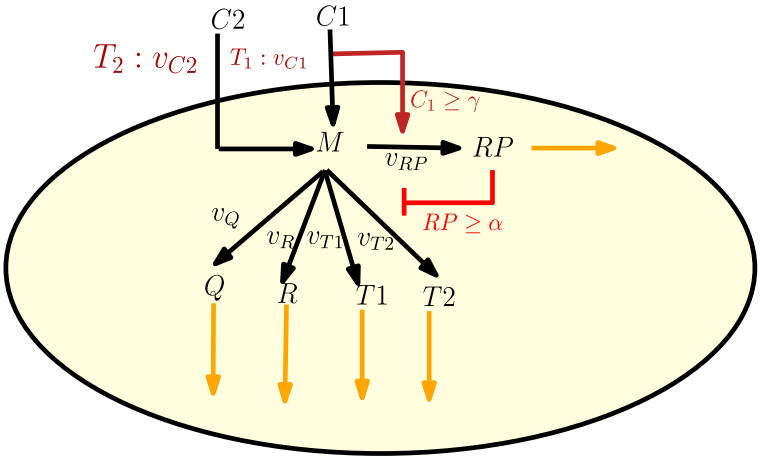

Figure 2. A metabolic network governed by two regulatory principles. C_1 and C_2 represent the two carbon sources. The enzymes T_1 and T_2 catalyse the conversion of carbon sources into the precursor metabolite *M*. The symbol *Q* denotes a non-catalytic macro-molecule and *RP* is a regulatory protein, where *R* denotes the ribosome catalyzing the creation of the protein.

the MRN network, aligning it with the hybrid model. Section 4 details the application of the sparse based system identification algorithm to the MRN network, displaying the inferred model. Section 5 presents our numerical findings to simulate the proposed continuous model and its sparse identified model. Finally, we conclude the paper with some closing remarks and future directions in Section 6 and provide the code availability statement in Section 7.

2. Introduction to hybrid automata and key elements relevant to the MRN

Hybrid systems are characterised by the coexistence of both continuous and discrete dynamics over time. In this study, we developed a continuous model to represent the hybrid metabolic-regulatory network (MRN) that governs the diauxic shift, as introduced in Liu and Bockmayr (2020) and illustrated in Figure 2.

The MRN involves two carbon sources, denoted as C_1 and C_2, which are supplied to the cell via the external medium and subsequently converted into a precursor metabolite, *M*. This metabolite undergoes further biochemical transformations within the cell. Two key regulatory proteins, *RP* and T_2, are governed by Boolean variables $\overline{RP}$ and $\overline{T_2}$, respectively. The regulatory logic controlling the network behaviour is defined by the following conditions:

- If the concentration of C_1 exceeds a threshold value γ, the gene responsible for producing *RP* is activated.
- If the concentration of *RP* exceeds a specified threshold α, it represses the gene encoding the regulatory protein T_2.

These regulatory dependencies are depicted in Figure 2, where brown and red arrows indicate activation and repression, respectively.

2.1. Elements of the hybrid MRN system

The hybrid MRN system (Figure 2) integrates multiple components that capture both the continuous and discrete dynamics underlying the regulatory-metabolic processes. These components are defined as follows:

- *Continuous variables (Y)*: The system's continuous state variables include the concentrations of metabolites and regulatory components: C_1, C_2, *M*, *Q*, *R*, T_1, T_2, and *RP*. Each of these variables evolves over time according to a set of ordinary differential equations (ODEs) that represent synthesis, degradation, uptake, and biochemical transformation rates within the system.
- *Regulatory modes*: The discrete regulatory states of proteins *RP* and T_2 define four distinct modes in the hybrid automaton: (on, on), (on, off), (off, off), and (off, on). These modes represent all possible activation/inhibition combinations of the associated genes, driven by threshold comparisons involving the concentrations of C_1 and *RP*, respectively. The transitions between these modes reflect changes in the regulatory logic, thereby modulating the biochemical dynamics of the system.

- *Event-driven transitions*: Discrete mode transitions are triggered by threshold-crossing events. Specifically, when the concentration of C_1 exceeds the threshold γ, the gene encoding *RP* is turned on. Similarly, when the concentration of *RP* surpasses the threshold α, it represses the gene encoding T_2. Each mode is associated with a unique set of differential equations and invariant conditions, which constrain the evolution of continuous variables while the system remains in that mode.
- *Kinetics and degradation*: The hybrid model incorporates enzyme-substrate dynamics using Michaelis-Menten kinetics (Michaelis & Menten, 1913), which provide a nonlinear representation of reaction rates dependent on substrate concentration and enzyme affinity. Additionally, the degradation of molecules such as metabolites, enzymes, and regulatory proteins is modelled using mass action kinetics (Guldberg & Waage, 1867), where degradation rates are proportional to the concentration of the corresponding species.

2.2. Dynamic equations for carbon sources and precursor molecule

The continuous dynamics of the carbon sources C_1, C_2, and the metabolic precursor M are governed by the following system of differential equations:

$$\dot{C_1} = -v_{C_1}, \tag{1}$$

$$\dot{C_2} = -v_{C_2}, \tag{2}$$

$$\dot{M} = v_{C_1} + v_{C_2} - v_M, \tag{3}$$

where the reaction rates are given by:

$$v_{C_1} = \frac{k_{\text{cat1}} \cdot T_1 \cdot C_1}{K_T + C_1}, \tag{4}$$

$$v_{C_2} = \frac{k_{\text{cat2}} \cdot T_2 \cdot C_2}{K_T + C_2}, \tag{5}$$

$$v_M = \frac{k_r \cdot R \cdot M}{K_r + M}. \tag{6}$$

Here, k_r and $k_{\text{cat}i}$ (for $i = 1, 2$) are turnover constants indicating the reaction rates, K_T represents the Michaelis-Menten constant, and K_r is the half-saturation constant for enzyme-substrate interactions (see Table 1).

Table 1. Summary of the kinetic parameters, threshold values, and enzyme sequence lengths.

Reaction rate constants	Value	unit
k_{cat1}	3000	min^{-1}
k_{cat2}	2000	min^{-1}
k_r	7	min^{-1}
Michaels-Menten constants		
K_T	1000	mmol
K_r	1260	mmol
Constants of degradation		
kd_e	0.01	min^{-1}
kd_{RP}	0.2	min^{-1}
kd_{T1}	0.05	min^{-1}
kd_{T2}	0.05	min^{-1}
Length of ribosome, enzyme		
n_Q, n_{RP}	300	
n_{T1}	400	
n_{T2}	1500	
n_R	7459	
Thresholds		
γ	20	mmol
α	0.03	mmol
Molar weight		
w	100	mg mmol^{-1}

2.3. Synthesis and degradation of macromolecules

The synthesis rates v_p of macromolecules $p \in \{Q, R, T_1, T_2, RP\}$ are adapted from the work reported in Faizi et al. (2018). For instance, the differential equation for Q is given by

$$\dot{Q} = v_p - k_d \cdot Q, \tag{7}$$

where k_d represents the degradation constant. Reaction rates v_p are expressed as:

$$v_p = \frac{\beta_p}{n_p} \cdot v_M, \tag{8}$$

$$v_M = \frac{k_r \cdot M \cdot R}{K_r + M}, \quad p \in \{Q, R, T_1, T_2, RP\}, \tag{9}$$

here, β_p represents the fraction of cellular resources allocated to the synthesis of protein p, subject to the constraint $\sum_p \beta_p = 1$, and n_p denotes the length of protein p. For additional details, see Faizi et al. (2018). At each node of the hybrid automaton (Figure 2), the values of β_p are adjusted according to the regulatory rules, dynamically shifting among the values 1/5, 1/4, and 1/3 to maintain balanced resource allocation across different proteins.

2.4. Biomass equation

The total biomass of the cell, defined as the sum of molecular masses, is expressed by

$$\text{Biomass}(t) = w \cdot M(t) + \sum_{p \in \{Q,R,T_1,T_2,RP\}} w \cdot n_p \cdot p(t), \tag{10}$$

where w denotes the average molar weight of the precursor M.

In next section, we aim to develop a continuous model that replicates the hybrid model of the MRN shown in Figure 2, regulatory proteins and metabolite concentrations are depicted as continuous variables.

3. Mathematical modelling of the MRN network in a continuous framework

Metabolic networks require sophisticated modelling frameworks to accurately capture their regulatory dynamics and biochemical complexity. Traditional hybrid models, which rely on binary (on/off) representations of regulatory states, exhibit inherent limitations in capturing the gradual transitions and graded responses that are commonly observed in biological systems. To overcome these constraints, we propose a continuous modelling approach that utilises Hill functions (Hill, 1909) to describe protein regulation and activity.

The Hill function is a well-established mathematical expression used to model cooperative binding and nonlinear regulatory effects in biochemical systems. It relates a system's output response (e.g. gene expression or enzymatic activity) to the concentration of a regulatory molecule (e.g. a ligand or transcription factor). This formulation enables the modelling of sigmoidal dose-response relationships and threshold-dependent activation or inhibition (Weiss, 1997), making it especially suitable for capturing switch-like behaviours in biological regulation.

Our continuous model retains several key features of the hybrid modelling framework. Specifically, it employs mass action kinetics (Guldberg & Waage, 1867) for modelling molecular degradation and Michaelis-Menten kinetics (Michaelis & Menten, 1913) for enzymatic synthesis and substrate consumption. However, the model departs fundamentally from the hybrid framework in its treatment of regulatory protein dynamics. Instead of Boolean functions that restrict regulatory protein states to binary values (0 or 1), we implement Hill functions to represent gene expression and protein activation on a continuous scale.

This modification allows the model to capture the full spectrum of protein activity levels, reflecting the graded and often cooperative nature of biological regulation. As a result, the continuous model offers a more realistic and mechanistically faithful representation of metabolic-regulatory interactions, enabling more accurate simulation and analysis of system behaviour under dynamic and heterogeneous cellular conditions.

At the heart of our continuous MRN model lies the *Hill function*, which transforms discrete protein activities into continuous expressions. This mathematical framework captures the subtle gradations in cellular responses to concentration changes, superseding the limitations of binary state representations. Formally, the Hill function is defined as a mapping $f : \mathbb{R}^+ \to [0, 1]$ from positive real numbers to the unit interval.

The network dynamics are governed by a system of ODEs:

$$\dot{x} = f(x(t), \theta) = S \cdot v(x(t), \theta), \quad x(t_0) = x_0, \tag{11}$$

where:

- $x \in \mathbb{R}^{n_m}$ denotes the species concentration vector, encompassing n_m distinct species
- $\theta \in \mathbb{R}^{n_\theta}$ represents the kinetic parameter vector
- $S \in \mathbb{R}^{n_m \times n_r}$ is the stoichiometric matrix
- $v \in \mathbb{R}^{n_r}$ describes the reaction rate vector for n_r reactions,

The reaction rates in vector v can incorporate various kinetic models, including Michaelis-Menten kinetics, mass action, and Hill functions, providing flexibility in describing different types of biochemical interactions.

The metabolic regulatory network features two key regulatory proteins, RP and T_2, operating under thresholds γ and α, respectively. In traditional hybrid models, these proteins follow binary behaviour:

- RP activates (value $= 1$) when concentration $C_1 \geq \gamma$
- T_2 deactivates (value $= 0$) when concentration $RP \geq \alpha$

Our continuous model replaces these discrete transitions with Hill functions, RPC_1 and RP_{RP}, defined as:

- For regulatory protein RP:

$$RPC_1 = \frac{C_1^h}{\gamma^h + C_1^h}, \tag{12}$$

where $h = 2$ is the Hill coefficient. This function exhibits saturating behaviour, approaching unity at high C_1 concentrations while decreasing smoothly as C_1 levels fall.

- For regulatory protein T_2:

$$RP_{RP} = \frac{\alpha^h}{\alpha^h + RP^h}, \tag{13}$$

This function displays inverse regulation, approaching zero at high RP concentrations and increasing gradually as RP levels decline.

The regulatory proteins are incorporated into the system through differential equations. For instance, the evolution of T_2 is described by:

$$\dot{T}_2 = v_M \cdot RP_{RP} - kd_{T2} \cdot T_2, \tag{14}$$

where kd_{T2} represents the degradation constant and RP_{RP} is the previously defined Hill function.

The reaction rates in our system are defined by:

$$v_{C1} = \frac{k_{cat1} \cdot T_1 \cdot C_1}{K_T + C_1}, \tag{15}$$

$$v_{C2} = \frac{k_{cat2} \cdot T_2 \cdot C_2}{K_T + C_2}, \tag{16}$$

$$v_M = \frac{k_r \cdot R \cdot M}{K_r + M}, \tag{17}$$

$$RPC_1 = \frac{C_1^2}{\gamma^2 + C_1^2}, \tag{18}$$

$$RP_{RP} = \frac{\alpha^2}{\alpha^2 + RP^2}. \tag{19}$$

3.1. Complete system dynamics

The MRN is given by the following system of ODEs:

$$\dot{C}_1 = -v_{C1}, \tag{20}$$

$$\dot{C}_2 = -v_{C2}, \tag{21}$$

$$\dot{M} = v_{C1} + v_{C2} - v_M, \tag{22}$$

$$\dot{Q} = \frac{\beta_p}{n_Q} \cdot v_M - k_de \cdot Q, \tag{23}$$

$$\dot{R} = \frac{\beta_p}{n_R} \cdot v_M - k_de \cdot R, \tag{24}$$

$$\dot{T}_1 = \frac{\beta_p}{n_{T1}} \cdot v_M - k_{dT1} \cdot T_1, \tag{25}$$

$$\dot{T}_2 = \frac{\beta_p}{n_{T2}} \cdot v_M \cdot RP_{RP} - k_{dT2} \cdot T_2, \tag{26}$$

$$\dot{RP} = \frac{\beta_p}{n_{RP}} \cdot v_M \cdot RPC_1 - k_{dRP} \cdot RP. \tag{27}$$

The total biomass of the system can be calculated using:

$$\text{Biomass}(t) = w \cdot M(t) + \sum_{p \in \{Q,R,T_1,T_2,RP\}} w \cdot n_p \cdot p(t), \tag{28}$$

where all species quantities (Q, R, T_1, T_2, RP) are measured in *mmol* and can be converted to *mg* using the factor $w \cdot n_p \cdot p(t)$. Parameter values and initial conditions are taken from Liu and Bockmayr (2020), with specific reaction constants listed in Table 1.

4. Application of the sparse based system identification approach to infer kinetic models

The continuous model introduced in the previous section provides a flexible and biologically relevant framework to capture the nonlinear regulatory dynamics of metabolic networks through differential equations incorporating Hill functions and other kinetic laws. However, as the complexity of such models increases, so does the challenge of inferring accurate and parsimonious representations of their governing equations directly from experimental or simulation data. To address this, we employ a sparse-based system identification approach, which complements the continuous model by enabling automated discovery of the underlying dynamical structure. By leveraging the sparsity inherent in biological interactions – where only a subset of all possible reactions and regulatory effects are active at any given time – this method efficiently identifies the most significant terms in the continuous model.

The sparse based system identification algorithm is an invaluable tool for system identification, particularly in understanding the complex dynamics of metabolic regulatory networks. Utilizing sparse regression, sparse based system identification helps derive the underlying equations governing intricate systems directly from data, thus enabling the extraction of significant dynamical interactions without extensive assumptions about the system's structure. By applying sparse based system identification to our

continuous model, we aim to pinpoint critical interactions and regulatory processes that drive cellular metabolism. This approach not only deepens our insights into metabolic dynamics but also provides a structured method for refining model parameters, ultimately enhancing predictive capabilities in biological systems.

The process of identifying dynamical systems from empirical data has been a cornerstone of mathematical physics for decades. Traditionally, this endeavour has relied on precise measurements, domain-specific knowledge, and expert intuition to uncover the governing equations underlying observed phenomena. However, recent advancements in computational capabilities, along with the increasing availability of large-scale data, have catalysed a paradigm shift toward the *automated* discovery of governing equations for dynamical systems. This emergent approach falls within the field of *system identification*, where modern statistical techniques and machine learning algorithms are utilised to infer dynamical models directly from observed data.

System identification involves formulating models that explain how system states evolve over time, typically by fitting data to predefined dynamical structures while imposing constraints that reflect physical principles or prior knowledge. The choice of constraints and assumptions differentiates various methodologies in system identification. For instance, *Dynamic Mode Decomposition* (DMD) Kutz et al. (2016) provides a robust framework for constructing optimal linear models by decomposing spatiotemporal data into modes and corresponding eigenvalues. These linear models capture dominant patterns in the data, making DMD an effective tool for high-dimensional systems where linear approximations are sufficient.

In scenarios where nonlinear dynamics dominate, alternative methodologies are required to extend system identification to more complex cases. A prominent example of such a method is the use of the *Koopman operator* (Khosravi, 2023; Koopman, 1931), an infinite-dimensional linear operator that acts on the observables of the system, rather than on the states themselves. The Koopman framework enables the representation of nonlinear systems through linear dynamics in a higher-dimensional space, thereby facilitating advanced spectral analysis of dynamical systems (Budišić et al., 2012; Mezić, 2005; Mezic, 2013). This approach has proven particularly effective for systems exhibiting quasi-periodic or chaotic behaviours, where traditional nonlinear techniques often struggle to provide a clear analytical framework.

Recent advances in this area have focussed on integrating physical laws and constraints directly into system identification frameworks. For example, the enforcement of conservation laws, such as energy conservation, or other physical invariants, can result in models that not only capture the system dynamics accurately but also respect the underlying physical principles. This integration enhances the interpretability of the models and improves their predictive capabilities (Majda & Harlim, 2012). Physics-informed models, which combine data-driven discovery with fundamental theoretical principles, bridge the gap between empirical analysis and traditional mechanistic modelling, offering a more holistic approach to system identification. This approach is particularly valuable in fields where domain-specific knowledge about the governing physical laws can be leveraged to guide model construction and improve accuracy.

Furthermore, symbolic regression techniques, such as those employed in *genetic programming*, have demonstrated remarkable success in identifying governing equations and conservation laws directly from data. These methods iteratively evolve candidate solutions by optimising an objective function, allowing for the discovery of compact and interpretable equations. Landmark studies have showcased the ability of genetic programming to recover equations of motion, conservation laws, and even novel dynamical insights with minimal prior assumptions (Bongard & Lipson, 2007; Forootani, Goyal, et al., 2025; Schmidt & Lipson, 2009). The flexibility and generality of such approaches make them particularly appealing for diverse applications across physics, biology, and engineering.

In summary, the field of system identification has evolved from reliance on domain expertise and manual derivation to a data-driven and algorithmically sophisticated discipline. Modern methods, ranging from linear models like DMD to advanced frameworks like the Koopman operator and genetic programming, have expanded the scope of systems that can be effectively modelled and understood. These advancements promise transformative impacts across scientific domains, enabling the discovery of dynamical systems that were previously beyond reach.

This section examines *sparse-based system identification*, a cutting-edge technique that leverages advancements in machine learning, sparse regression, and computational optimisation to identify nonlinear dynamical systems directly from data (Brunton et al., 2016a; Forootani & Benner, 2024; Forootani, Kapadia, et al., 2024; Peterson et al., 2025). Sparse-based system identification builds upon earlier approaches, such as compressed sensing (Baraniuk, 2007; Candès, 2006; Donoho, 2006), which utilised sparsity-promoting techniques to recover signals from limited or incomplete data (Wang et al., 2011). However, sparse regression (Forootani & Benner, 2024; Tibshirani, 1996) introduces enhanced stability and robustness in handling noisy or overdetermined datasets, making it particularly well-suited for identifying dynamical systems.

The core principle of sparse-based system identification is rooted in the observation that many dynamical systems, even those capable of generating complex or chaotic behaviours, are often governed by equations that exhibit sparsity in an appropriate functional basis. Dynamical systems can be generically represented as:

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}), \tag{29}$$

where $\mathbf{x} \in \mathbb{R}^n$ is the state vector, $\dot{\mathbf{x}}$ is its time derivative, and $\mathbf{f}(\mathbf{x})$ is a potentially nonlinear function dictating the system's dynamics. Sparse-based system identification aims to discover the essential terms within $\mathbf{f}(\mathbf{x})$, thereby providing interpretable models that capture the underlying dynamics.

A classic example that illustrates the concept of sparsity in dynamics is the Lorenz system. Despite its ability to exhibit highly complex and chaotic behaviour, the Lorenz system is governed by a relatively small number of terms in its governing equations. This characteristic makes it an ideal candidate for sparse regression techniques, which aim to identify the most significant terms driving the system's dynamics while discarding less relevant contributions. One such technique is ℓ_1-regularised regression, which effectively identifies these critical terms by penalising the inclusion of nonessential ones, thereby promoting sparsity in the model.

This principle of sparsity is central to methods like *Sparse Identification of Nonlinear Dynamics* (SINDy), which has emerged as a cornerstone of the field of system identification (Brunton et al., 2016a). SINDy utilises sparse regression to infer the governing equations of nonlinear dynamical systems directly from data, selecting only the most influential terms from a large library of candidate functions. The ability to recover compact, interpretable models from high-dimensional data is one of the primary advantages of SINDy, making it a powerful tool for understanding complex systems, particularly those exhibiting chaotic or nonlinear behaviour.

Advancements in scientific machine learning, such as Reinforcement Learning (RL), have focussed on combining efficient model-based techniques with interpretable frameworks for dynamic systems (Rackauckas et al., 2020). For instance, kernel-based methods have been applied in sensor scheduling (Forootani, Iervolino, et al., 2024), least-squares Monte Carlo simulation methods have been used for optimal stopping time, sensor scheduling or resource allocation problems (Forootani et al., 2023; Forootani, Iervolino, et al., 2025; Forootani, Liuzza, et al., 2021; Forootani, Tipaldi, et al., 2021; Forootani et al., 2020), demonstrating the effectiveness of these techniques in decision-making and optimisation problems. Building on these advancements, methods like SINDy-RL (Zolman et al., 2024) combine the sparse identification framework with RL to develop interpretable, efficient, model-based RL algorithms.

Figure 3 illustrates the overall methodology of sparse-based system identification.

The algorithmic process begins by collecting time-series data according to Equation (29), forming a data matrix:

$$\mathbf{X} = [\mathbf{x}(t_1) \quad \mathbf{x}(t_2) \quad \cdots \quad \mathbf{x}(t_m)]^T, \tag{30}$$

where ^{T} indicates the transpose. The matrix $\mathbf{X}$ has dimensions $m \times n$, where n is the dimensionality of the state vector $\mathbf{x} \in \mathbb{R}^n$, and m represents the number of time samples. For example, in biological systems, $\mathbf{x}$ may correspond to concentrations of enzymes, metabolites, or transcription factors.

To estimate the system's dynamics, the derivative matrix is computed or assembled from $\mathbf{X}$:

$$\dot{\mathbf{X}} = [\dot{\mathbf{x}}(t_1) \quad \dot{\mathbf{x}}(t_2) \quad \cdots \quad \dot{\mathbf{x}}(t_m)]^T, \tag{31}$$

where robust numerical methods, such as total-variation regularisation (Rudin et al., 1992), are employed to estimate derivatives accurately, even in the presence of noise (Chartrand, 2011).

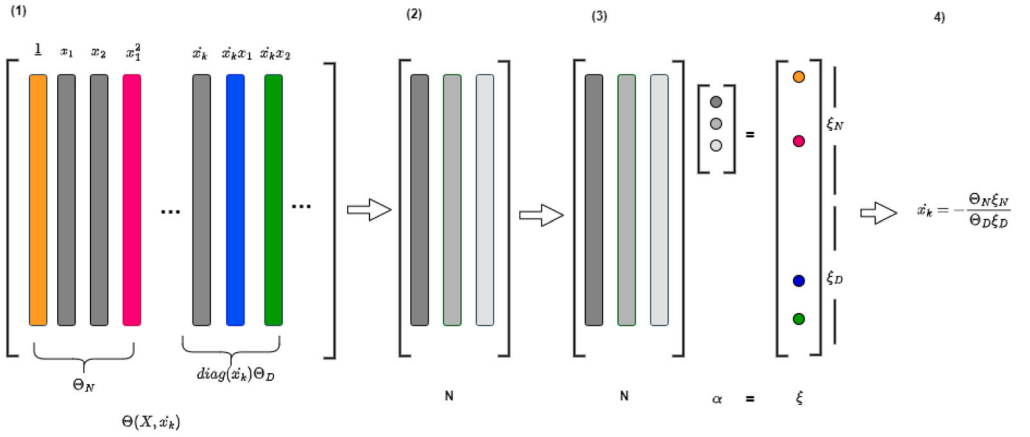

Figure 3. The sparse based system identification algorithm for identifying system dynamics follows these steps: (1) Construct a matrix $\boldsymbol{\Theta}(\mathbf{X}, \dot{x}_k)$, with each column representing a nonlinear function applied to the time-series data $x_1, x_2, x_3, \ldots$ and the derivative $\dot{x}_k(\mathbf{t})$ of a particular species; (2) Calculate $\mathbf{N}$, an orthonormal basis for the null space of $\boldsymbol{\Theta}$; (3) Use the alternating directions method Qu et al. (2014) to identify a sparse vector $\boldsymbol{\xi}$ within the null space, ensuring that $\boldsymbol{\xi}$ satisfies $\boldsymbol{\Theta}\boldsymbol{\xi} = 0$; (4) Leverage the nonzero entries in $\boldsymbol{\xi}$ together with the function library $\boldsymbol{\Theta}$ to build the inferred model. This procedure is repeated for each species' derivative $\dot{x}_k$.

The next step involves constructing a comprehensive library of potential candidate functions to model $\mathbf{f}(\mathbf{x})$. This library, represented as:

$$\boldsymbol{\Theta}(\mathbf{X}) = [\mathbf{1} \quad \mathbf{X} \quad \mathbf{X}^2 \quad \cdots \quad \mathbf{X}^d \quad \cdots \quad \sin(\mathbf{X}) \cdots], \tag{32}$$

contains various linear, polynomial, and nonlinear functions of the state vector $\mathbf{x}$. Here, $\mathbf{X}^d$ represents matrices containing all possible polynomial terms up to degree d. For instance, in a two-dimensional system where $\mathbf{x} = [x_1 \;\; x_2]^T$, the second-order terms $\mathbf{X}^2$ would include entries like $[x_1^2(t) \;\; x_1x_2(t) \;\; x_2^2(t)]$, capturing interactions and nonlinearities within the system.

Sparse regression is then applied to identify the most relevant terms from the library $\boldsymbol{\Theta}(\mathbf{X})$ that contribute to $\mathbf{f}(\mathbf{x})$. This involves solving the optimisation problem:

$$\min_{\boldsymbol{\xi}} \|\dot{\mathbf{X}} - \boldsymbol{\Theta}(\mathbf{X})\boldsymbol{\xi}\|_2^2 + \lambda\|\boldsymbol{\xi}\|_1, \tag{33}$$

where $\boldsymbol{\xi}$ contains the coefficients of the terms in the library, $\|\cdot\|_2$ is the least-squares norm, and $\lambda\|\cdot\|_1$ imposes sparsity via ℓ_1-regularisation. This process ensures that only the most essential terms are retained in the identified model, enabling both interpretability and predictive accuracy.

We then express the time derivatives in $\dot{\mathbf{X}}$ in terms of the candidate nonlinear functions in $\boldsymbol{\Theta}(\mathbf{X})$ using the relation:

$$\dot{\mathbf{X}} = \boldsymbol{\Theta}(\mathbf{X})\boldsymbol{\Xi}, \tag{34}$$

where $\boldsymbol{\Xi}$ is a sparse coefficient matrix. Each column $\boldsymbol{\xi}_k$ of $\boldsymbol{\Xi}$ represents a coefficient vector corresponding to the kth equation in the dynamical system described by Equation (29). The goal is to identify the active terms in $\boldsymbol{\Theta}(\mathbf{X})$ that contribute to the dynamics of the system while promoting sparsity in $\boldsymbol{\Xi}$.

To achieve this, sparse regression techniques are employed to estimate each column $\boldsymbol{\xi}_k$. Specifically, $\boldsymbol{\xi}_k$ is obtained by solving the following optimisation problem:

$$\boldsymbol{\xi}_k = \text{argmin}_{\boldsymbol{\xi}_k'} \|\dot{\mathbf{X}}_k - \boldsymbol{\Theta}(\mathbf{X})\boldsymbol{\xi}_k'\|_2^2 + \lambda\|\boldsymbol{\xi}_k'\|_1, \tag{35}$$

where $\dot{\mathbf{X}}_k$ denotes the kth column of $\dot{\mathbf{X}}$, $\|\cdot\|_2$ represents the least-squares norm, and $\|\cdot\|_1$ is the ℓ_1 norm that enforces sparsity. The parameter $\lambda > 0$ is a regularisation weight that controls the trade-off between sparsity and accuracy. This formulation is commonly referred to as LASSO (Least Absolute Shrinkage and Selection Operator) regression (Tibshirani, 1996), a well-known method for sparse signal reconstruction and model selection.

The sparse coefficient vectors $\boldsymbol{\xi}_k$, once estimated, allow us to reconstruct the dynamics of the nonlinear system in a compact form:

$$\dot{\mathbf{x}} = \boldsymbol{\Xi}^T \left(\boldsymbol{\Theta}(\mathbf{x}^T)\right)^T. \tag{36}$$

This model effectively captures the essential dynamical relationships in the system, while ignoring spurious or redundant terms, thereby ensuring both interpretability and efficiency.

Identifying the active terms in the dynamics from the candidate library $\boldsymbol{\Theta}(\mathbf{X})$ via sparse regression presents a convex optimisation challenge. A brute-force approach would involve searching across all possible subsets of terms in $\boldsymbol{\Theta}(\mathbf{X})$ and selecting the subset that optimally balances model accuracy and sparsity. However, this exhaustive search becomes computationally infeasible due to the combinatorial explosion in the number of potential subsets, particularly for high-dimensional systems or large candidate libraries. As the dimensionality of the system or the size of the library increases, the number of possible combinations grows exponentially, making such an approach impractical for most real-world applications.

To overcome this challenge, sparse regression methods, such as LASSO (Tibshirani, 1996), sequential thresholding (Brunton et al., 2016a), or other sparsity-promoting algorithms (Candès, 2006; Donoho, 2006), offer computationally efficient alternatives. These methods iteratively refine the solution by penalising the inclusion of unnecessary terms, effectively reducing the size of the active set of coefficients. Recent advancements in optimisation techniques, including coordinate descent and proximal gradient methods, have further improved the scalability and robustness of sparse-based system identification (Line et al., 2011).

An alternative to explicit sparsity-promoting algorithms is the *implicit sparse-based system identification* method, which leverages regularisation techniques combined with heuristic or adaptive selection rules to achieve sparsity without explicitly solving a combinatorial problem. Implicit methods aim to minimise the computational cost while preserving the accuracy and interpretability of the identified model. By dynamically adjusting the threshold for retaining terms in $\boldsymbol{\Xi}$, implicit approaches strike an effective balance between model complexity and performance.

In summary, sparse regression techniques form the backbone of system identification by enabling the selection of sparse models from large candidate libraries in a computationally feasible way. By leveraging advanced optimisation methods and sparsity-promoting algorithms, these techniques provide interpretable and accurate representations of nonlinear dynamical systems, facilitating their application across a wide range of scientific and engineering domains (Brunton et al., 2016a; Candès, 2006; Tibshirani, 1996).

The polynomial and trigonometric terms included in Equation (32) generally provide sufficient flexibility to model a broad array of dynamical systems. For instance, incorporating all polynomials up to degree n implies that the system's dynamics can be represented by mass-action kinetics with terms up to the nth degree. This approach offers a robust framework to effectively capture the complex behaviour typical of biological networks and similar systems.

The sparse based system identification method excels with high-dimensional datasets, where many traditional modelling approaches may become inefficient or overly complex. By leveraging sparse regression, sparse based system identification pinpoints the key terms driving the dynamics, creating a simplified yet accurate representation of the system. In real-world applications, the success of sparse based system identification depends significantly on carefully choosing candidate functions and optimising the regularisation parameter λ, as these choices directly impact the method's accuracy and interpretability.

Recent advancements have broadened the scope of sparse based system identification to tackle more intricate systems, such as those involving time delays (Leylaz et al., 2022) and systems with additive noise (Brunton & Kutz, 2019). These modifications enhance the robustness and general applicability of sparse based system identification, enabling it to address a wider variety of real-world dynamical systems.

In summary, sparse based system identification is a robust and efficient method for identifying dynamical systems, utilising sparse regression techniques to uncover the governing equations of complex systems from empirical data. Its ability to handle high-dimensional and noisy data makes it an invaluable tool in modern system identification.

4.1. Inferring nonlinear dynamical systems with rational functions for metabolic regulatory networks

In this section, we extend the approach introduced in Mangan et al. (2016) by applying it to kinetic models of metabolic regulatory networks (MRNs). Our objective is to uncover the intricate dynamics of these networks, building upon the foundational principles laid out in the original work. We will describe the specific methods employed in this study, illustrating how

we adapt the sparse-based system identification framework to capture the nonlinear interactions that characterise metabolic regulation. This extension aims to deepen our understanding of the fundamental mechanisms driving metabolic processes, providing valuable insights for both theoretical research and practical applications in systems biology, such as metabolic engineering and synthetic biology.

Many real-world dynamical systems, including biological networks, are characterised by interactions governed by rational functions. This highlights the need to adapt the sparse-based system identification algorithm to accommodate a broader class of nonlinearities, extending beyond polynomial and trigonometric terms. The ability to handle these more complex terms is crucial for accurately modelling systems that exhibit intricate interactions, particularly those involving feedback loops, enzymatic kinetics, and cooperative effects.

Integrating rational functions into the sparse-based system identification framework presents a significant challenge. Unlike polynomials or trigonometric terms, rational functions cannot be easily represented as sparse linear combinations of a fixed set of functions. To address this, the sparse regression approach must be adapted to account for the more complex structure of rational functions. This involves modifying the underlying optimisation problem, which now needs to include rational terms both in the numerator and denominator of the functions. This adaptation ensures that the sparse-based algorithm can identify the sparsest ordinary differential equation (ODE) that accurately captures the dynamics of systems with rational function-based interactions, such as Michaelis-Menten kinetics in metabolic reactions.

By extending the sparse-based system identification framework to handle rational functions, we can uncover more accurate and parsimonious models of complex systems. This modification allows for a more flexible representation of the system's dynamics, facilitating the discovery of the essential terms that govern the system while avoiding overfitting or unnecessary complexity.

Consider a dynamical system described by Equation (29), where the evolution of each variable $k = 1, 2, \ldots, n$ may involve interactions governed by rational functions. These dynamics can be expressed as:

$$\dot{x}_k = \frac{f_N(\mathbf{x})}{f_D(\mathbf{x})}, \tag{37}$$

where $f_N(\mathbf{x})$ and $f_D(\mathbf{x})$ are polynomials in the state vector $\mathbf{x}$. In this formulation, $f_N(\mathbf{x})$ represents the numerator polynomial, capturing the driving forces of the system, while $f_D(\mathbf{x})$ is the denominator polynomial, which introduces nonlinear dependencies on the state variables.

To handle the rational form of these dynamics in the context of system identification, we eliminate the fraction by multiplying both sides of Equation (37) by the denominator polynomial $f_D(\mathbf{x})$. This reformulates the dynamics as:

$$f_N(\mathbf{x}) - f_D(\mathbf{x})\dot{x}_k = 0, \tag{38}$$

which represents the system in an implicit form. This transformation ensures that the dynamics can now be expressed entirely as a sum of polynomial terms involving both the state variables $\mathbf{x}$ and their derivatives $\dot{x}_k$.

This implicit representation necessitates a generalisation of the function library $\boldsymbol{\Theta}$ from Equation (32), extending it to accommodate the new structure involving both $\mathbf{x}$ and $\dot{x}_k$. The augmented library can be written as:

$$\boldsymbol{\Theta}(\mathbf{X}, \dot{x}_k(\mathbf{t})) = [\boldsymbol{\Theta}_N(\mathbf{X}) \quad \text{diag}\,(\dot{x}_k(\mathbf{t}))\,\boldsymbol{\Theta}_D(\mathbf{X})], \tag{39}$$

where:

(1) $\boldsymbol{\Theta}_N(\mathbf{X})$ is the library of monomials representing the numerator terms,
(2) $\boldsymbol{\Theta}_D(\mathbf{X})$ is the library of monomials for the denominator terms,
(3) $\text{diag}\,(\dot{x}_k(\mathbf{t}))$ is a diagonal matrix that scales the denominator terms by the time derivative $\dot{x}_k$.

For a single state variable x_k, the second term $\text{diag}(\dot{x}_k(\mathbf{t}))\boldsymbol{\Theta}_D(\mathbf{X})$ expands as follows:

$$\text{diag}(\dot{x}_k(\mathbf{t}))\boldsymbol{\Theta}_D(\mathbf{X}) = [\dot{x}_k(\mathbf{t}) \quad \dot{x}_k x_k(\mathbf{t}) \quad \dot{x}_k x_k^2(\mathbf{t}) \cdots]. \tag{40}$$

In practice, we typically use the same polynomial degree for both the numerator and denominator libraries, such that $\boldsymbol{\Theta}_N(\mathbf{X}) = \boldsymbol{\Theta}_D(\mathbf{X})$. Consequently, the augmented library $\boldsymbol{\Theta}(\mathbf{X}, \dot{x}_k(\mathbf{t}))$ effectively

doubles the size of the original library described in Equation (32), making it more expressive while also increasing the computational complexity.

The dynamics of the system in Equation (38) can now be expressed using this augmented library as:

$$\boldsymbol{\Theta}(\mathbf{X}, \dot{x}_k(\mathbf{t}))\boldsymbol{\xi}_k = \mathbf{0}, \tag{41}$$

where $\boldsymbol{\xi}_k$ is the sparse coefficient vector representing the contributions of each term to the dynamics. Non-zero entries in $\boldsymbol{\xi}_k$ indicate the active terms in the dynamics, providing a compact representation of the system.

Unlike the standard sparse regression approach used in linear or polynomial systems (Brunton et al., 2016a; Tibshirani, 1996), applying sparse regression directly to Equation (41) may lead to trivial solutions where all entries in $\boldsymbol{\xi}_k$ are zero. This issue arises because the equality constraint in Equation (41) enforces a strict balance between terms, which cannot be achieved when sparsity is overly penalised.

To address this challenge, alternative formulations are required. One approach involves relaxing the constraint in Equation (41) by introducing a residual tolerance that allows small approximation errors. Another strategy leverages iterative refinement, where the active terms are first identified approximately and then refined by solving a sequence of constrained optimisation problems (Line et al., 2011).

In summary, by reformulating rational dynamics in terms of an augmented function library, we enable the application of sparse regression techniques for system identification. This approach captures the essential interactions in the system while accommodating the complex interplay between state variables and their derivatives, paving the way for more expressive and accurate models of nonlinear dynamical systems.

To identify the sparsest non-zero vector $\boldsymbol{\xi}_k$ that fulfils Equation (41), we recognise that such a vector resides in the null space of $\boldsymbol{\Theta}$. Once the null space is identified, the task is to find the sparsest vector within it. Although this is a non-convex optimisation problem, methods like the alternating directions method (ADM) introduced by Qu et al. (2014) can be applied to efficiently identify the sparsest solution in this subspace.

4.2. Algorithm for sparse identification of rational functions

The procedure for identifying the sparse coefficient vector $\boldsymbol{\xi}_k$ for the dynamical system is structured into four main steps, as outlined below:

(1) *Construct the Augmented Function Library*: Begin by building the extended library $\boldsymbol{\Theta}(\mathbf{X}, \dot{x}_k(\mathbf{t}))$, which incorporates terms involving both the state variables $\mathbf{X}$ and their time derivatives $\dot{x}_k(\mathbf{t})$. The construction of this library combines polynomial terms for both the numerator and denominator dynamics, as described in Equation (39). This ensures that the rational form of the dynamics is adequately captured, enabling the discovery of both linear and nonlinear interactions.

(2) *Compute the Null Space Representation*: Solve for the null space of the augmented function library $\boldsymbol{\Theta}(\mathbf{X}, \dot{x}_k(\mathbf{t}))$. This results in a matrix $\mathbf{N}$, where each column represents a vector in the null space of $\boldsymbol{\Theta}$. The null space approach ensures that we identify the coefficients $\boldsymbol{\xi}_k$ that satisfy the implicit formulation of the system dynamics, as given in Equation (41). By leveraging this null space structure, the problem is reformulated to capture the interactions between numerator and denominator terms effectively.

(3) *Extract Sparse Coefficients via Optimization*: Identify sparse linear combinations of the null space basis vectors by applying a sparsity-promoting optimisation method. Specifically, we employ the Alternating Directions Method (ADM) proposed by Qu et al. (2014), which iteratively solves for a sparse vector $\boldsymbol{\xi}_k$ by minimising a regularised cost function. This optimisation process balances the trade-off between fitting the observed data and enforcing sparsity in the coefficients. The sparsity is controlled using a regularisation parameter λ, which penalises the inclusion of non-essential terms in the model.

(4) *Reconstruct the Dynamical Model*: Using the sparse vector $\boldsymbol{\xi}_k$, reconstruct the governing equations for the system. This involves selecting the active terms in the function library $\boldsymbol{\Theta}(\mathbf{X}, \dot{x}_k(\mathbf{t}))$ based on the non-zero entries of $\boldsymbol{\xi}_k$. These active terms define the final inferred model, capturing the dominant interactions and dependencies in the dynamics.

To ensure the robustness of the identified model, the selection of the sparsity regularisation parameter λ is performed through an iterative process:

- *Regularization Parameter Tuning*: Since the optimal value of λ is not known a priori, we explore a range of λ values. For each value of λ, steps 3 and 4 are repeated to derive a model. Increasing λ encourages greater sparsity in the coefficient vector $\boldsymbol{\xi}_k$, resulting in simpler models with fewer active terms.
- *Model Selection Using Pareto Analysis*: The trade-off between model sparsity and accuracy is assessed using a Pareto front. This front represents the set of models obtained for different λ values, plotting sparsity (number of active terms) against accuracy (model error). The final model is selected as the one that achieves the maximum sparsity without significantly compromising accuracy, striking an optimal balance between interpretability and predictive performance.

This systematic approach allows for the identification of key rational functions and their coefficients, yielding a sparse and interpretable representation of the underlying dynamical system.

4.3. A comprehensive framework for identifying implicit dynamical systems

The method described earlier can be extended to identify more general implicit ODEs, allowing for a broader class of nonlinearities beyond just rational functions. The library $\boldsymbol{\Theta}(\mathbf{X}, \dot{x}_k(\mathbf{t}))$ is a subset of a larger library $\boldsymbol{\Theta}([\mathbf{X}\ \dot{\mathbf{X}}])$, which incorporates both nonlinear functions of the state variables $\mathbf{x}$ and their derivatives $\dot{\mathbf{x}}$. This extended library enhances the flexibility of the modelling framework, enabling the identification of more complex dynamical systems that cannot be captured by rational function nonlinearities alone.

By identifying the sparsest vector in the null space of $\boldsymbol{\Theta}([\mathbf{X}\ \dot{\mathbf{X}}])$, we can capture intricate nonlinear equations, including those with mixed terms that involve powers of both the state variables and their derivatives. For example, the following equation:

$$\dot{x}^2 x^2 - x\dot{x}^2 - x^4 = 0, \tag{42}$$

which can be expressed as a sparse vector within the null space of $\boldsymbol{\Theta}([\mathbf{X}\ \dot{\mathbf{X}}])$.

Extending this framework to higher-order derivatives is straightforward by augmenting the feature set in the $\boldsymbol{\Theta}$ library. For instance, second-order dynamical systems can be modelled by including terms that involve second derivatives, leading to a more comprehensive library capable of capturing such dynamics:

$$\boldsymbol{\Theta}([\mathbf{X}\quad \dot{\mathbf{X}}\quad \ddot{\mathbf{X}}]). \tag{43}$$

This enhanced approach expands the potential for system identification, enabling the detection of a wider range of dynamics. By incorporating higher-order derivatives or mixed state-derivative interactions, this methodology allows for the discovery of complex systems that extend beyond the capabilities of traditional rational function-based models.

5. Numerical experiments and simulation results

The numerical simulations presented in this section are organised into two main parts. In the first part, we investigate the behaviour of the continuous model under various settings of the weighting parameter β_p. The model's dynamic responses are analysed through simulations that reflect different metabolic and regulatory scenarios. In the second part, we focus on sparse system identification applied to the continuous model. This approach aims to infer the underlying mathematical structure and parameters of the MRN using data-driven techniques. Together, these two parts provide a comprehensive view of both the simulated dynamics and the inferred model structure.

5.1. Numerical results of the continuous model

In the hybrid model, the parameter β_p transitions as follows: from $\beta_p = \frac{1}{5}$ in the (on, on) mode, to $\beta_p = \frac{1}{4}$ in the (on, off) mode, then to $\beta_p = \frac{1}{3}$ in the (off, off) mode, and finally back to $\beta_p = \frac{1}{4}$ in the (off, on) mode. To ensure a proper representation in the continuous model, the parameter β_p will be adjusted to values of $\frac{1}{5}$, $\frac{1}{4}$, or $\frac{1}{3}$. The simulation outcomes of the continuous model will be analysed as follows.

The continuous model simulations were generated in `Python` using the `scipy` library (Virtanen et al., 2020). The simulations were run with tolerances of $rtol = atol = 10^{-6}$, and the initial conditions were set as $C_1 = 500$, $C_2 = 1000$, $M = 20$, $Q = 0.1$, $R = 0.01$, $T_1 = 0$, $T_2 = 0$, and $RP = 0$, all in units of

millimoles (mmol). Figures 4, 5, and 6 illustrate the outcomes of three distinct continuous models, each characterised by a different value of β. The trajectories of metabolites, macromolecules, and biomass in these models are shown in panels (a), (b), and (c), respectively. Figure 5 presents the behaviour of the state variables in the mathematical model for the case $\beta = 1/4$. This case aligns closely with the hybrid model described in Liu and Bockmayr (2020). We observe that, at the start of the simulation, the levels of regulatory proteins RP and T_2 rise, while the metabolites C_1 and C_2 are consumed at a slow rate. During the 0–20 minute period, it becomes evident that protein RP continues to increase, while T_2 remains relatively stable without significant production. During this interval, C_2 remains inactive, and the cell consumes C_1 at an exponential rate. Around 20 minutes, the concentration of C_1 is almost depleted, prompting the cell to begin utilising the carbon source C_2. Furthermore, around 30 minutes, the production of regulatory protein RP ceases entirely, reducing its level to zero, while the concentration of T_2 continues to increase.

In cases where $\beta = 1/5$ and $\beta = 1/3$, the behaviour of biomass differs significantly from that of the hybrid model. Specifically, while the biomass trajectory follows a similar qualitative trend, its levels are comparatively lower or higher, respectively, than in the hybrid model. In Figure 4, the trajectory of C_2 shows exhaustion after 60 minutes, whereas, in Figure 6, C_2 is depleted slightly quick.

In Figure 5, the behaviour of macromolecules T_1 and T_2 aligns qualitatively with that of the hybrid model. However, in Figures 4 and 6, the trajectories of T_1 and T_2 appear lower with a minor delay of 56 minutes and higher with a slight lead, respectively, compared to the pattern observed in Figure 5.

The results above indicate that the case $\beta = 1/4$ is the most accurate representation of weight changes in the continuous model. This finding may appear to conflict with the constraint $\sum_p \beta_p = 1$. However, an alternative interpretation resolves this issue by showing no conflict with $\sum_p \beta_p = 1$. Initially, T_2 is inactive in the presence of C_1, resulting in an absence of T_2 and indicating that four macromolecules are involved, with a total combined weight of 1. While C_1 is consumed, RP is absent, and T_2 becomes active, maintaining the total weight at 1. Consequently, in the continuous model, this weight distribution is represented by $\beta_p = 1/4$.

5.1.1. Generalizing a continuous model for MRN dynamics

In this section, our objective is to derive a general formula for a continuous model that describes the hybrid metabolic-genetic network. As discussed earlier, the dynamics of the regulatory proteins are controlled by Hill functions. We observe that the weight of each mode is impacted by the states of these regulatory proteins. To extend this weighting approach, we define it as a function dependent on the regulatory proteins. In our specific case study, we focus on five macromolecules, with two regulatory proteins, RP and T_2, characterised by Hill functions, namely RP_{Rp} and RPC_1. The weights can then be expressed as

$$\beta_p = \frac{1}{(5-2) + RP_{Rp} + RPC_1}. \tag{44}$$

By substituting this expression into the ODE system, We obtain the simulation results of the continuous model, as depicted in Figure 7.

The continuous model exhibits qualitatively similar macromolecule trajectories to those in the hybrid model. The general formula of the weight can be given by:

$$\beta_p = \frac{1}{(n-ng) + Reg_{ng}}, \tag{45}$$

Here, n denotes the total count of macromolecules, ng specifies the number of regulatory proteins, and Reg refers to the Hill functions associated with these regulatory proteins.

5.2. Results of sparse based system identification for discovering of MRN network

In this section, we apply the system identification method to the continuous model of the metabolic regulatory network (MRN) system introduced earlier. The network consists of eight equations, each involving a mix of rational functions and polynomials. Notably, three of these equations include multiple rational terms, particularly those associated with M, T_2, and RP. For simplicity in our analysis, we set $\beta = 1$ and relabel the variables as follows:

$$\begin{aligned} C_1 = x_1, \quad C_2 = x_2, \quad M = x_3, \\ Q = x_4, \quad R = x_5, \quad T_1 = x_6, \\ T_2 = x_7, \quad RP = x_8. \end{aligned}$$

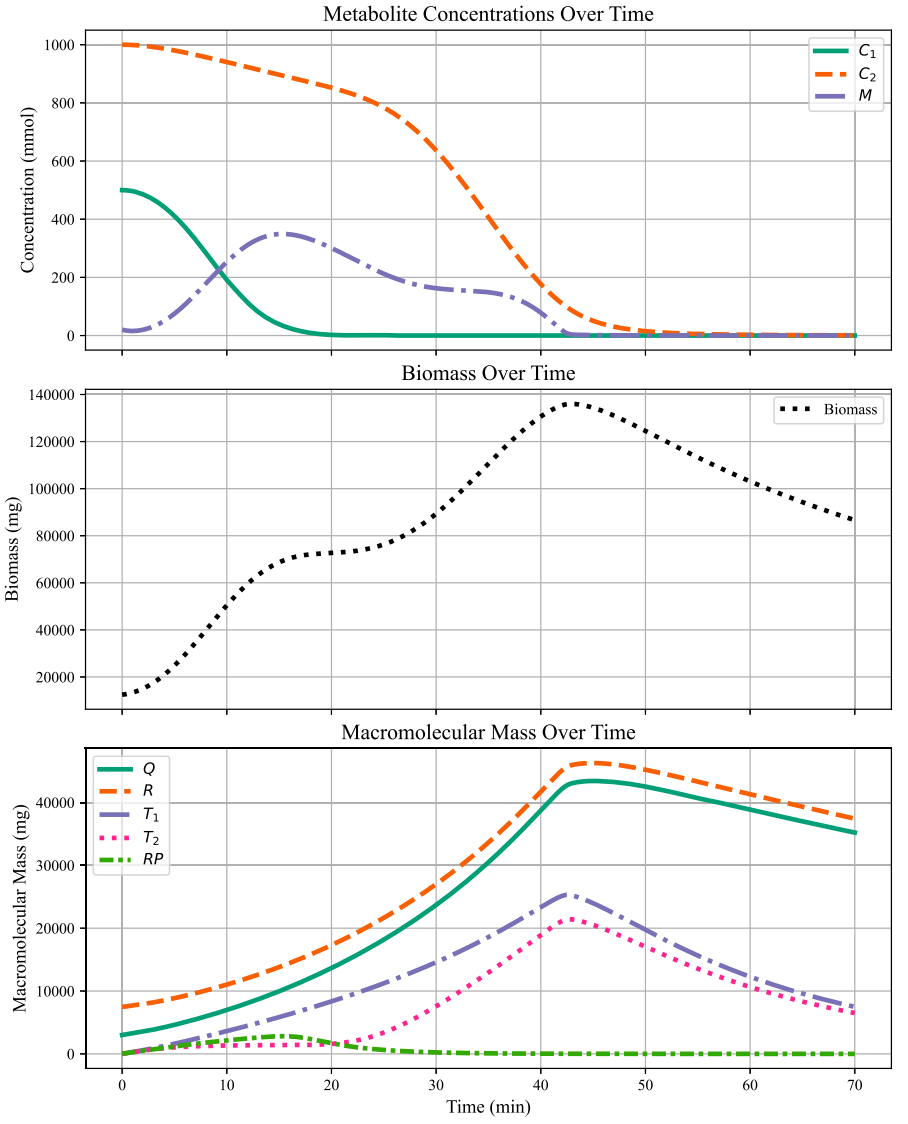

Figure 4. The simulation results of metabolites C_1, C_2, M are shown in (a), the macromolecules Q, R, T_1, T_2, RP depicted in (b), while the biomass is shown in (c). The trajectories of metabolites and biomass in a continuous model are generated using the value of the weight parameter $\beta_p = 1/3$.

The dynamics for x_1, x_2, x_4, x_5, and x_6 are identified using Latin hypercube sampling (McKay et al., 2000), to sample the initial conditions for the concentrations of the species. In this study, we use the following bounds for the states:

- Lower bounds: $(400, 900, 18, 0.001, 0.001, 0.001, 0, 0.01)$
- Upper bounds: $(600, 1100, 22, 0.2, 0.2, 0.2, 0.5, 0.1)$.

We begin our simulation with the initial conditions $(500, 500, 1000, 20, 0.1, 0.01, 0.01, 0.01)$. To identify the vector ξ and sparsely select the active terms for the equation involving x_6, we apply the implicit sparse based system identification method (Mangan et al., 2016). The vector ξ identifies five active terms from the function library, constructing a rational function from Θ and ξ. The method successfully selects five terms: three in the numerator and two in the denominator. When rearranged, the coefficients of these terms match those of the original system.

5.3. Construction of inferred model for the x_6 equation

Here, we take the x_6 equation as an example to explain the method and its results.

$$0.35\, x_6 - 3.15\, x_3 x_5 + 0.05\, x_3 x_6 + (7 + x_3)\dot{x_6} = 0, \tag{46}$$

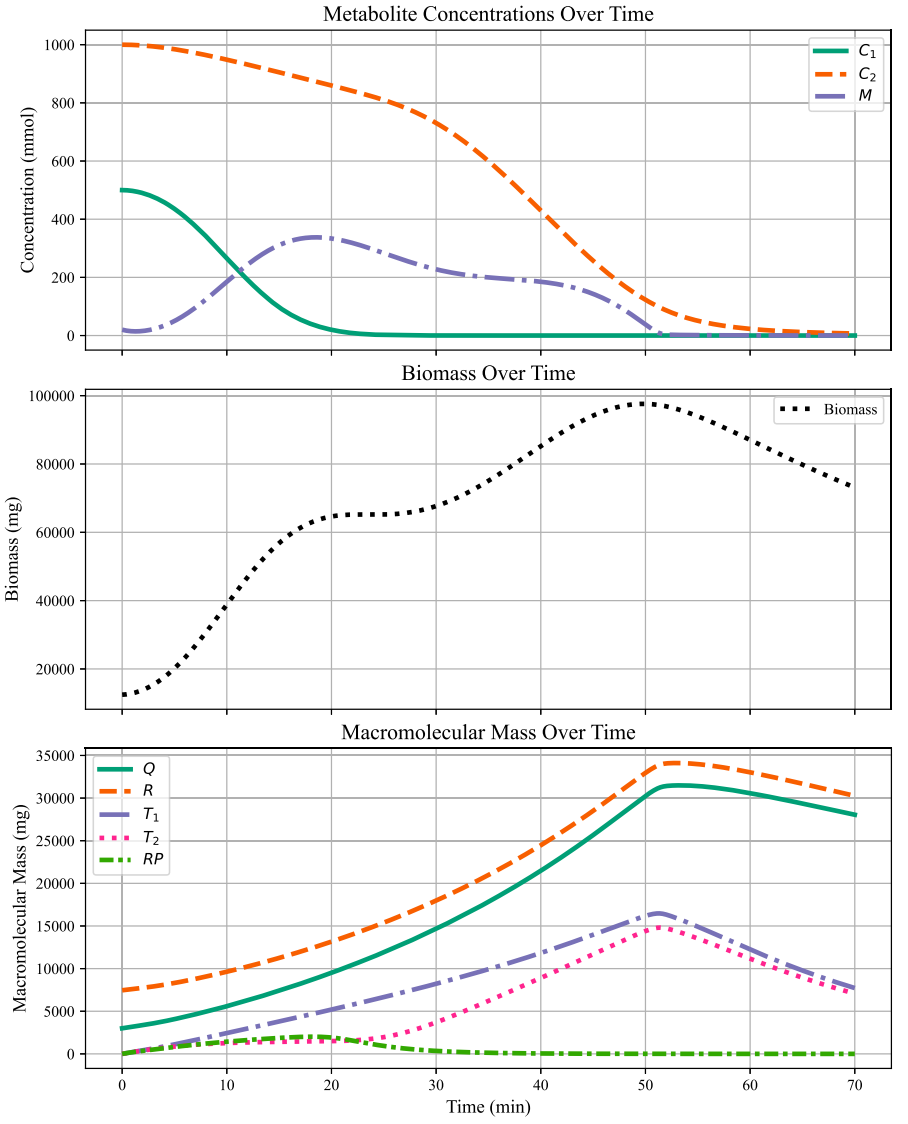

Figure 5. The simulation results of metabolites C_1, C_2, M are shown in (a), the macromolecules Q, R, T_1, T_2, RP depicted in (b), while the biomass is shown in (c). The trajectories of metabolites and biomass in a continuous model are generated using the value of the weight parameter $\beta_p = 1/4$.

$$\dot{x_6} = \frac{-0.35\, x_6 - 0.05\, x_3 x_6 + 3.15\, x_3 x_5}{7 + x_3}, \quad (47)$$

$$\dot{x_6} = \frac{-0.05\, x_6 (7 + x_3)}{(7 + x_3)} + \frac{3.15\, x_3 x_5}{(7 + x_3)}, \quad (48)$$

$$\dot{x_6} = -0.05\, x_6 + \frac{3.15\, x_3 x_5}{7 + x_3}. \quad (49)$$

Figure 9 illustrates the comparison between the concentration over time and its derivative for various state variables in the MRN model. The figure displays the evolution of the concentrations along with their time derivatives, providing insights into the dynamic behaviour and rate of change of these species across different time points. This comparison aids in understanding the temporal dynamics of the system's components, essential for capturing the regulatory interactions and metabolic processes.

The implicit sparse based system identification method successfully infers the network structure and coefficients for the system. The dynamics of five equations – specifically x_1, x_2, x_4, x_5, and x_6 – are identified accurately. However, the algorithm fails to identify sparse dynamics for the variables x_3, x_7, and x_8, as their dynamics involve more than just rational functions. Additionally, the identification results are influenced by the specified lower and upper bounds

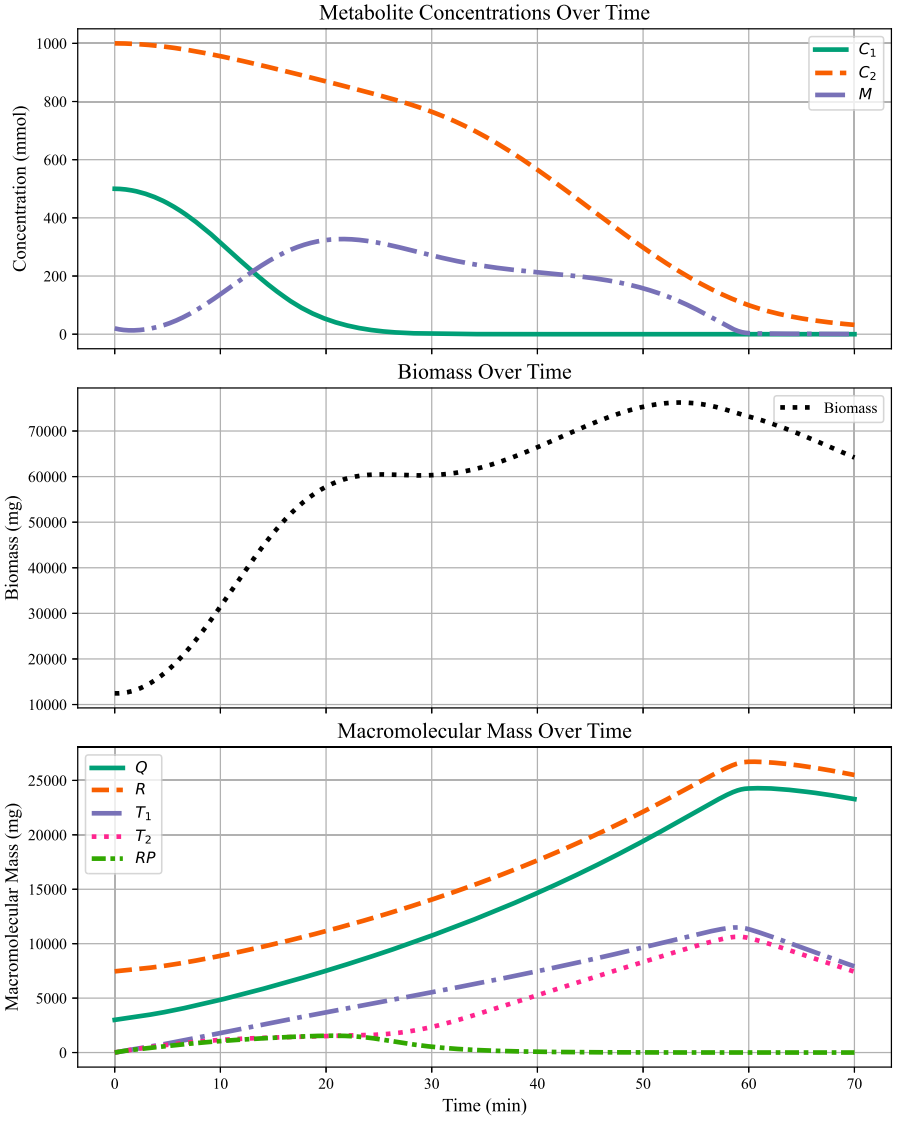

Figure 6. The simulation results of metabolites C_1, C_2, M are shown in (a), the macromolecules Q, R, T_1, T_2, RP depicted in (b), while the biomass is shown in (c). The trajectories of metabolites and biomass in a continuous model are generated using the value of the weight parameter $\beta_p = 1/5$.

for the species, as well as the tolerance and lambda parameters.

$$\dot{x_1} = -\frac{a_1 x_1 x_6}{a_2 + x_1}$$
$$\dot{x_2} = -\frac{b_1 x_2 x_7}{b_2 + x_2},$$
$$\dot{x_4} = \frac{c_1 x_3 x_5}{c_2 + x_3} - c_3 x_4,$$
$$\dot{x_5} = \frac{d_1 x_3 x_5}{d_2 + x_3} - d_3 x_5,$$
$$\dot{x_6} = \frac{e_1 x_3 x_5}{e_2 + x_3} - e_3 x_6,$$

The Pareto front, as shown in Figure 10 for the state x_6, highlights the trade-off between the approximation error and the number of non-zero terms in the sparse vector derived from our algorithm. This analysis is central to identifying optimal sparse solutions that balance model accuracy and simplicity. The error, plotted on a logarithmic scale, represents the norm of the residuals (i.e. the difference between the predicted and actual values). The x-axis indicates the number of non-zero terms, which corresponds to the model complexity. Points along the Pareto front represent solutions where further reduction in error would require a significant increase in the number of terms, while solutions to the left of the curve offer lower model

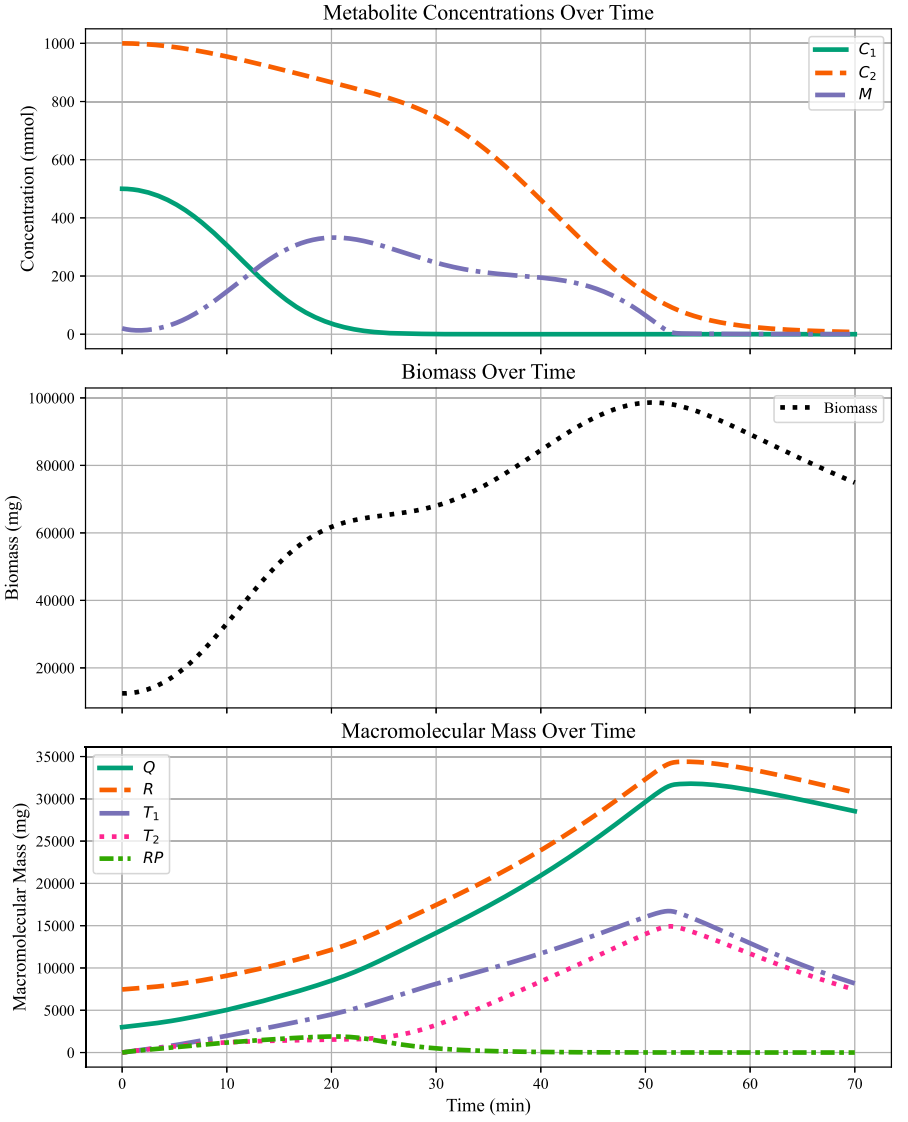

Figure 7. The simulation results of metabolites C_1, C_2, M are shown in (a), the macromolecules Q, R, T_1, T_2, RP depicted in (b), while the biomass is shown in (c). The trajectories are generated using the weight value given by $\beta_p = 1/(3 + RP_{Rp} + RPC_1)$.

complexity but higher error. These results are critical in selecting a sparse model that effectively captures the dynamics of the system with minimal complexity.

6. Conclusion and future directions

In this paper, we presented a continuous model of a metabolic-regulatory network (MRN), providing an effective and scalable framework for analyzing the interactions and dynamics within such systems. The proposed model proved particularly useful in scenarios involving a large number of regulatory rules, allowing for a detailed investigation of regulatory protein behaviour while avoiding the combinatorial complexity inherent in hybrid modelling approaches.

In comparison, the hybrid model, although effective in certain contexts, became increasingly impractical as the number of regulatory rules increased. Specifically, we observed four transition graphs representing the interactions between two regulatory proteins, consistent with the growth pattern 2^2 for two proteins. More generally, the number of modes increased exponentially as 2^n, where n denoted the number of regulatory rules. This exponential growth significantly limited the comprehensibility and applicability of hybrid models in large-scale systems.

To conclude our analysis, we applied the implicit sparse-based system identification method for model inference. This approach successfully identified the governing dynamics of five out of the eight equations

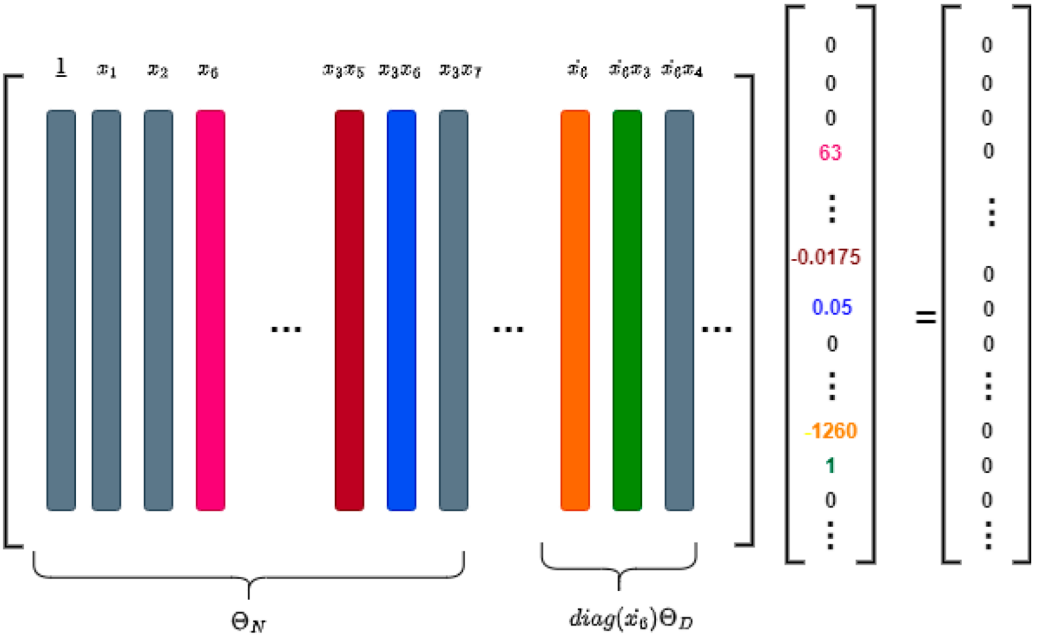

Figure 8. The algorithm is applied to the MRN of Escherichia coli. The figure presents the function library Θ along with the corresponding coefficient vector ξ for x_6. In this case, five functions are active, with three terms contributing to the numerator and two to the denominator.

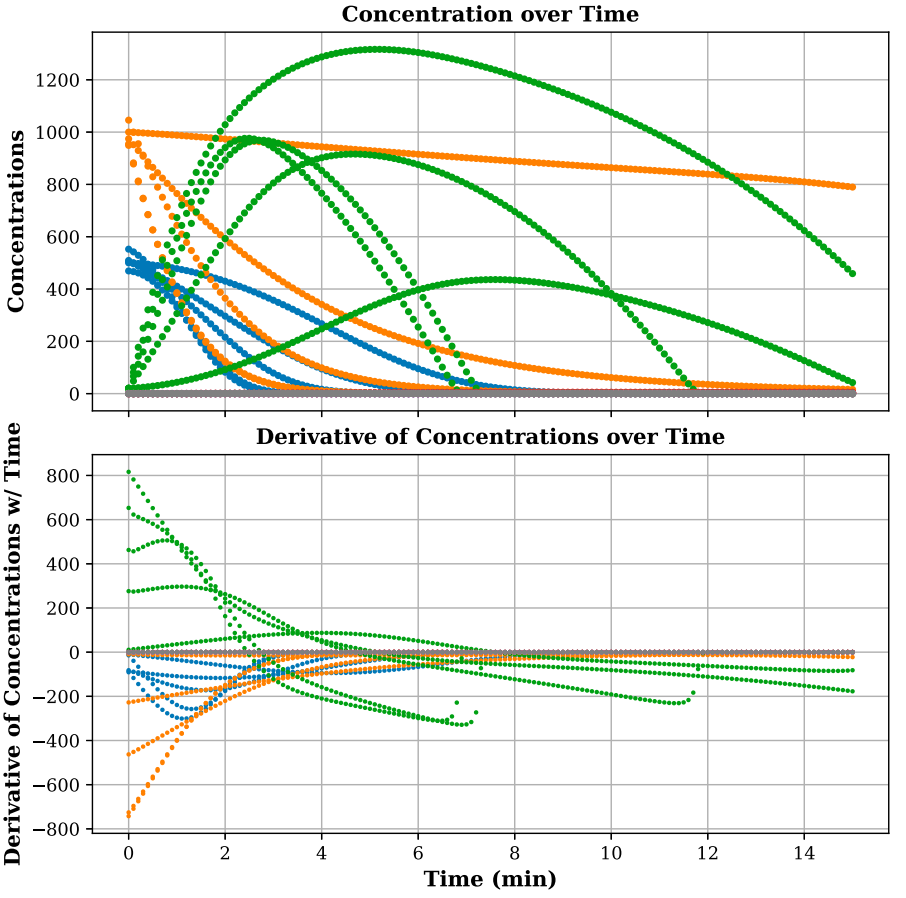

Figure 9. Comparison of Concentration over Time and its Derivative.

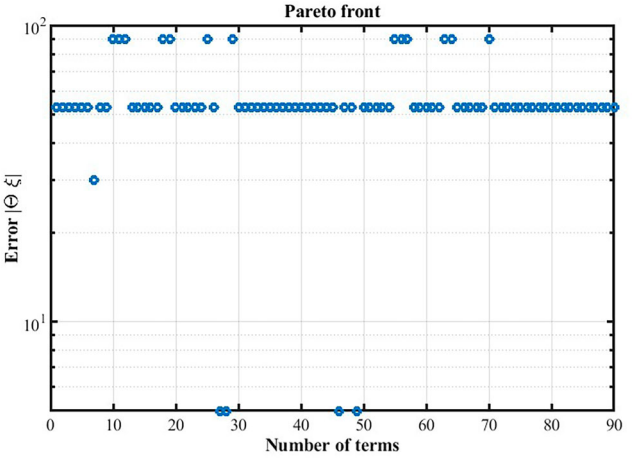

Figure 10. Pareto front for the state x_6.

Table 2. Identified Parameters for the Kinetic Model of the MRN.

Parameter	Unit	*TrueValue*	*EstimatedValues*
a_1	min^{-1}	3000	3000
a_2	mmol	1000	1000
b_1	min^{-1}	2000	2000
b_2	mmol	1000	1000
c_1	min^{-1}	4.2	4.2
c_2	mmol	7	7
c_3	min^{-1}	0.01	0.01
d_1	min^{-1}	0.1689	0.1689
d_2	mmol	7	7
d_3	min^{-1}	0.01	0.01
e_1	min^{-1}	3.15	3.15
e_2	mmol	7	7
e_3	min^{-1}	0.05	0.05

in the network, corresponding to the state variables x_1, x_2, x_4, x_5, and x_6. The method accurately reconstructed both the structure and coefficients of these equations, demonstrating strong agreement with the ground truth (see Table 2 for details).

However, the method did not accurately recover the dynamics of x_3, x_7, and x_8, likely due to the complexity of their underlying nonlinear interactions, which were not sufficiently captured by the sparse regression model used in this study.

Future Directions. In recent years, the application of deep learning to predict circRNA-protein binding sites has garnered significant attention. CircRNAs, known for their unique circular structure and roles in gene expression regulation, interact with various proteins, including RNA-binding proteins (RBPs), influencing cellular processes and disease progression. The study reported in Shen et al. (2025) provides a comprehensive review of deep learning models, such as CRIP, CircSLNN, and CSCRSites, which employ sophisticated architectures like convolutional neural networks (CNNs), bidirectional long short-term memory (BiLSTM) networks, and attention mechanisms to effectively identify RBP binding sites within circRNA sequences. These models use diverse encoding strategies, such as stacked codon-based encoding and one-hot encoding, to represent circRNA sequences and extract relevant features, with sequence labelling techniques and ensemble learning methods enhancing prediction accuracy. Deep learning-based models outperform traditional machine learning approaches, offering faster, more cost-efficient alternatives to biological experimental methods like ChIP-seq and PAR-CLIP. The survey also discusses future directions for improving these models, including the need for more complex computational architectures and incorporating additional biological data. Looking ahead, the integration of deep learning models for circRNA-protein binding prediction offers a promising future direction for MRNs. These methods, which excel at capturing complex biological interactions and dependencies from large datasets, could be adapted to MRN modelling, enhancing the prediction and analysis of regulatory and metabolic dynamics. By leveraging deep learning's ability to handle high-dimensional, heterogeneous data, we can uncover new insights into the relationships between metabolic pathways, regulatory elements, and gene expression, advancing the development of more accurate and scalable models for systems biology. Incorporating multimodal data, such as omics-level data, protein interactions, and gene expression profiles, can further improve MRN models, driving advancements in metabolic engineering, synthetic biology, and precision medicine applications.

7. Code and data availability statement

The `MATLAB` code developed for implementing the continuous model and performing sparse-based system identification for the analysis of metabolic-regulatory networks (MRNs) is publicly available on GitHub at https://github.com/Ali-Forootani/Metabolic_regulatory_network, and has been permanently archived on Zenodo with the following DOI: 10.5281/zenodo.14540008.

All data used in the simulations and analyses presented in this study are included in the GitHub repository and mirrored in the Zenodo archive.

The repository provides comprehensive documentation, including step-by-step instructions and example scripts, to facilitate full reproducibility of the results. Users can run the provided code to replicate the model behaviour, perform parameter sweeps, and apply the sparse identification procedure to their own datasets.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

Ali Eshtewy, N., & Scholz, L. (2020). Model reduction for kinetic models of biological systems. *Symmetry*, *12*(5), 863. https://doi.org/10.3390/sym12050863

Altman, N., & Krzywinski, M. (2015). *Points of significance: Simple linear regression*. Nature Publishing Group.

Baraniuk, R. G. (2007). Compressive sensing. *IEEE Signal Processing Magazine*, *24*(4), 118–121. https://doi.org/10.1109/MSP.2007.4286571

Björck, Å. (2024). *Numerical methods for least squares problems*. SIAM.

Bongard, J., & Lipson, H. (2007). Automated reverse engineering of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, *104*(24), 9943–9948. https://doi.org/10.1073/pnas.0609476104

Bortolussi, L., & Policriti, A. (2008). Hybrid systems and biology: Continuous and discrete modeling for systems biology. In *Formal Methods for Computational Systems Biology: 8th International School on Formal Methods for the Design of Computer, Communication, and Software Systems, SFM 2008 Bertinoro, Italy, June 2–7, 2008 Advanced Lectures 8* (pp. 424–448). Springer.

Brunton, S. L., & Kutz, J. N. (2019). *Data-driven science and engineering: Machine learning, dynamical systems, and control*. Cambridge University Press.

Brunton, S. L., Proctor, J. L., & Kutz, J. N. (2016a). Discovering governing equations from data: Sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, *113*(15), 3932–3937. https://doi.org/10.1073/pnas.1517384113

Brunton, S. L., Proctor, J. L., & Kutz, J. N. (2016b). Sparse identification of nonlinear dynamics with control (SINDYc). *IFAC-PapersOnLine*, *49*(18), 710–715. https://doi.org/10.1016/j.ifacol.2016.10.249

Budišić, M., Mohr, R., & Mezić, I. (2012). Applied Koopmanism. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *22*(4), 47510. https://doi.org/10.1063/1.4772195

Cakir, T., & Khatibipour, M. J. (2014). Metabolic network discovery by topdown and bottom-up approaches and paths for reconciliation. *Frontiers in Bioengineering and Biotechnology*, *2*, 62.

Candès, E. J. (2006). Compressive sensing. *Proceedings of the International Congress of Mathematicians*, *3*, 1433–1452.

Chartrand, R. (2011). Numerical differentiation of noisy, nonsmooth data. *International Scholarly Research Notices, 2011*(1), 164564.

Chasman, D., Siahpirani, A. F., & Roy, S. (2016). Network-based approaches for analysis of complex biological systems. *Current Opinion in Biotechnology*, *39*, 157–166. https://doi.org/10.1016/j.copbio.2016.04.007

Covert, M. W., Schilling, C. H., & Palsson, B. (2001). Regulation of gene expression in flux balance models of metabolism. *Journal of Theoretical Biology*, *213*(1), 73–88. https://doi.org/10.1006/jtbi.2001.2405

Covert, M. W., Xiao, N., Chen, T. J., & J. R. Karr (2008). Integrating metabolic, transcriptional regulatory and signal transduction models in Escherichia coli. *Bioinformatics*, *24*(18), 2044–2050. https://doi.org/10.1093/bioinformatics/btn352

Deuflhard, P. (2011). *Newton methods for nonlinear problems: Affine invariance and adaptive algorithms*, Vol. 35. Springer Science & Business Media.

Donoho, D. L. (2006). Compressed sensing. *IEEE Transactions on Information Theory*, *52*(4), 1289–1306. https://doi.org/10.1109/TIT.2006.871582

Enders, C. K. (2005). Direct maximum likelihood estimation. In *Encyclopedia of statistics in behavioral science*. Wiley Online Library.

Eshtewy, N. A. S. (2020). *Mathematical modeling of metabolic-genetic networks*. Freie Universitaet Berlin.

Eshtewy, N. A., Scholz, L., & Kremling, A. (2023). Parameter estimation for a kinetic model of a cellular system using model order reduction method. *Mathematics*, *11*(3), 699. https://doi.org/10.3390/math11030699

Faizi, M., Zavřel, T., Loureiro, C., Červenỳ, J., & Steuer, R. (2018). A model of optimal protein allocation during phototrophic growth. *Biosystems*, *166*, 26–36. https://doi.org/10.1016/j.biosystems.2018.02.004

Forootani, A., & Benner, P. (2024). GN-SINDy: Greedy sampling neural network in sparse identification of nonlinear partial differential equations, arXiv preprint arXiv:2405.08613.

Forootani, A., Forootani, S., Iervolino, R., & Zarch, M. G. (2023). Stochastic dynamic programming solution to transmission scheduling: Multi sensor-multi process with wireless noisy channel. *Computers and Electrical Engineering*, *106*, 108573. https://doi.org/10.1016/j.compeleceng.2022.108573

Forootani, A., Goyal, P., & Benner, P. (2025). A robust sparse identification of nonlinear dynamics approach by combining neural networks and an integral form. *Engineering Applications of Artificial Intelligence*, *149*, 110360. https://doi.org/10.1016/j.engappai.2025.110360

Forootani, A., Iervolino, R., Tipaldi, M., & Baccari, S. (2024). A kernel-based approximate dynamic programming approach: Theory and application. *Automatica*, *162*, 111517. https://doi.org/10.1016/j.automatica.2024.111517

Forootani, A., Iervolino, R., Tipaldi, M., & Khosravi, M. (2025). Off-policy temporal difference learning for perturbed markov decision processes. *IEEE Control Systems Letters*, *8*, 3488–3493. https://doi.org/10.1109/LCSYS.2025.3547629

Forootani, A., Kapadia, H., Chellappa, S., Goyal, P., & Benner, P. (2024). GS-PINN: Greedy sampling for parameter estimation in partial differential equations, arXiv preprint arXiv:2405.08537.

Forootani, A., Liuzza, D., Tipaldi, M., & Glielmo, L. (2021). Allocating resources via price management systems: A dynamic programming-based approach. *International Journal of Control*, *94*(8), 2123–2143. https://doi.org/10.1080/00207179.2019.1694178

Forootani, A., Tipaldi, M., Iervolino, R., & Dey, S. (2021). Enhanced exploration least-squares methods for optimal stopping problems. *IEEE Control Systems Letters*, *6*, 271–276. https://doi.org/10.1109/LCSYS.2021.3069708

Forootani, A., Tipaldi, M., Zarch, M. G., Liuzza, D., & Glielmo, L. (2020). A least-squares temporal difference based method for solving resource allocation problems. *IFAC Journal of Systems and Control*, *13*, 100106. https://doi.org/10.1016/j.ifacsc.2020.100106

Guldberg, C. M., & Waage, P. (1867). *Etudes sur les affinités chimiques*. Brøgger & Christie.

Hill, A. V. (1909). The mode of action of nicotine and curari, determined by the form of the contraction curve and the method of temperature coefficients. *The Journal of Physiology*, *39*(5), 361–373. https://doi.org/10.1113/jphysiol.1909.sp001344

Khosravi, M. (2023). Representer theorem for learning Koopman operators. *IEEE Transactions on Automatic Control*, *68*(5), 2995–3010. https://doi.org/10.1109/TAC.2023.3242325

Koopman, B. O. (1931). Hamiltonian systems and transformation in Hilbert space. *Proceedings of the National Academy of Sciences*, *17*(5), 315–318. https://doi.org/10.1073/pnas.17.5.315

Kremling, A., Geiselmann, J., Ropers, D., & de Jong, H. (2018). An ensemble of mathematical models showing diauxic growth behaviour. *BMC Systems Biology*, *12*(1), 1–16. https://doi.org/10.1186/s12918-018-0604-8

Kreutz, C. (2018). An easy and efficient approach for testing identifiability. *Bioinformatics*, *34*(11), 1913–1921. https://doi.org/10.1093/bioinformatics/bty035

Kutz, J. N., Brunton, S. L., Brunton, B. W., & Proctor, J. L. (2016). *Dynamic mode decomposition: Data-driven characterization of complex systems*. SIAM.

Leylaz, G., Wang, S., & Sun, J.-Q. (2022). Identification of nonlinear dynamical systems with time delay. *International Journal of Dynamics and Control*, *10*(1), 13–24. https://doi.org/10.1007/s40435-021-00783-7

Line, C., Hastie, T., Witten, D., & Ersbøll, B. (2011). Sparse discriminant analysis. *Technometrics*, *53*(4), 406–413. https://doi.org/10.1198/TECH.2011.08118

Liu, L., & Bockmayr, A. (2020). Formalizing metabolic-regulatory networks by hybrid automata. *Acta Biotheoretica*, *68*(1), 73–85. https://doi.org/10.1007/s10441-019-09354-y

Majda, A. J., & Harlim, J. (2012). Physics constrained nonlinear regression models for time series. *Nonlinearity*, *26*(1), 201–217. https://doi.org/10.1088/0951-7715/26/1/201

Mangan, N. M., Brunton, S. L., Proctor, J. L., & Kutz, J. N. (2016). Inferring biological networks by sparse identification of nonlinear dynamics. *IEEE Transactions on Molecular, Biological and Multi-Scale Communications*, *2*(1), 52–63. https://doi.org/10.1109/TMBMC.2016.2633265

Mannan, A. A., Toya, Y., Shimizu, K., McFadden, J., Kierzek, A. M., & Rocco, A. (2015). Integrating kinetic model of E. coli with genome scale metabolic fluxes overcomes its open system problem and reveals bistability in central metabolism. *PloS One*, *10*(10), e0139507. https://doi.org/10.1371/journal.pone.0139507

Marbach, D., Costello, J. C, Küffner, R., Vega, N. M, Prill, R. J, Camacho, D. M, Allison, K. R, Kellis, M., Collins, J. J, & Stolovitzky, G. (2012). Wisdom of crowds for robust gene network inference. *Nature Methods*, *9*(8), 796–804. https://doi.org/10.1038/nmeth.2016

Marmiesse, L., Peyraud, R., & Cottret, L. (2015). FlexFlux: Combining metabolic flux and regulatory network analyses. *BMC Systems Biology*, *9*(1), 1–13. https://doi.org/10.1186/s12918-015-0238-z

Martínez-Antonio, A., & Collado-Vides, J. (2003). Identifying global regulators in transcriptional regulatory networks in bacteria. *Current Opinion in Microbiology*, *6*(5), 482–489. https://doi.org/10.1016/j.mib.2003.09.002https://doi.org/10.1016/j.mib.2003.09.002 .

McKay, M. D., Beckman, R. J., & Conover, W. J. (2000). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, *42*(1), 55–61. https://doi.org/10.1080/00401706.2000.10485979

Mezić, I. (2005). Spectral properties of dynamical systems, model reduction and decompositions. *Nonlinear Dynamics*, *41*(1–3), 309–325. https://doi.org/10.1007/s11071-005-2824-x

Mezic, I. (2013). Analysis of fluid flows via spectral properties of the Koopman operator. *Annual Review of Fluid Mechanics*, *45*(1), 357–378. https://doi.org/10.1146/fluid.2013.45.issue-1

Michaelis, L., & Menten, M. L. (1913). Die Kinetik der Invertinwirkung. *Biochemische Zeitschrift*, *49*(333-369), 352.

Peterson, L., Forootani, A., Medina, E. I. S., Gosea, I. V., Sundmacher, K., & Benner, P. (2025). Towards digital twins for power-to-X processes: Comparing surrogate models for a catalytic CO2 methanation reactor. *IEEE Transactions on Automation Science and Engineering*. https://doi.org/10.1109/TASE.2025.3564637

Politano, G., Savino, A., Benso, A., Di Carlo, S., Rehman, H. U., & Vasciaveo, A. (2014). Using Boolean networks to model post-transcriptional regulation in gene regulatory networks. *Journal of Computational Science*, *5*(3), 332–344. https://doi.org/10.1016/j.jocs.2013.10.005

Qu, Q., Sun, J., & Wright, J. (2014). Finding a sparse vector in a subspace: Linear sparsity using alternating directions. *Advances in Neural Information Processing Systems 27*, 3401–3409. https://doi.org/10.1109/TIT.2016.2601599

Rackauckas, C., Ma, Y., Martensen, J., Warner, C., Zubov, K., Supekar, R., Skinner, D., Ramadhan, A., & Edelman,

A. (2020). Universal differential equations for scientific machine learning, arXiv preprint arXiv:2001.04385.

Raue, A., Schilling, M., Bachmann, J., Matteson, A., Schelke, M., Kaschek, D., Hug, S., Kreutz, C., Harms, B. D., Theis, F. J., Klingmüller, U., Timmer, J., & Hernandez-Lemus, E. (2013). Lessons learned from quantitative dynamical modeling in systems biology. *PLoS ONE*, *8*(9), e74335. https://doi.org/10.1371/journal.pone.0074335https://doi.org/10.1371/journal.pone.0074335 .

Rudin, L. I., Osher, S., & Fatemi, E. (1992). Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, *60*(1–4), 259–268. https://doi.org/10.1016/0167-2789(92)90242-F

Schmidt, M., & Lipson, H. (2009). Distilling free-form natural laws from experimental data. *Science*, *324*(5923), 81–85. https://doi.org/10.1126/science.1165893

Shen, Z., Yuan, L., Bao, W., Wang, S., Zhang, Q., & Huang, D.-S. (2025). A brief survey of deep learning-based models for CircRNA-protein binding sites prediction. *Neurocomputing*, *628*, 129637. https://doi.org/10.1016/j.neucom.2025.129637

Taou, N. S., Corne, D. W., & Lones, M. A. (2018). Investigating the use of Boolean networks for the control of gene regulatory networks. *Journal of Computational Science*, *26*, 147–156. https://doi.org/10.1016/j.jocs.2018.04.012

Tibshirani, R. (1996). Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *58*(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

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, Brett, M., Wilson, J., Jarrod Millman, K., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., …Vàzquez-Baeza, Y. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*(3), 261–272. https://doi.org/10.1038/s41592-019-0686-2

Wang, W. X., Yang, R., Lai, Y. C., Kovanis, V., & Grebogi, C. (2011). Predicting catastrophes in nonlinear dynamical systems by compressive sensing. *Physical Review Letters*, *106*(15), 1–4. https://doi.org/10.1103/PhysRevLett.106.154101

Weiss, J. N. (1997). The Hill equation revisited: uses and misuses. *The FASEB Journal*, *11*(11), 835–841. https://doi.org/10.1096/fsb2.v11.11

Zolman, N., Fasel, U., Kutz, J. N., & Brunton, S. L. (2024). SINDy-RL: interpretable and efficient model-based reinforcement learning, arXiv preprint arXiv:2403.09110.