Advances in the Inference of Chemical Reaction Networks from Time Series Data: A Systematic Survey

Preprint

JUSTIN NOAH KREIKEMEYER and ADELINDE M. UHRMACHER, University of Rostock, Germany

We provide an accessible introduction and overview of the machine-learning and system-identification problem of inferring chemical reaction networks (CRNs) from time series data. Specifically, given data on the temporal evolution of abundances or concentrations of entities, the task is to infer a set of reactions that could have produced this data, including their kinetic laws and corresponding parameters. Solving this problem is expected to significantly accelerate the prediction and understanding of population dynamics by automating the time-intensive modeling process. In a systematic survey of peer-reviewed and preprint articles published until December 2025, we identified 71 publications detailing 68 distinct methods motivated by chemical and biological applications. Based on this large sample of the literature, we provide a current perspective on the methodological developments, propose a three-dimensional taxonomy, and highlight promising future directions. Motivated by the absence of a recent state-of-the-art synthesis and scattered progress across the application domains of CRNs, we view this work as an important step toward further methodological progress.

CCS Concepts: • **General and reference** → **Surveys and overviews**; • **Computing methodologies** → *Artificial intelligence*; *Modeling and simulation*; • **Applied computing** → Systems biology; Chemistry.

Additional Key Words and Phrases: machine learning, model learning, system identification.

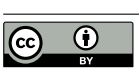
1 INTRODUCTION

The construction of simulation models is often referred to as a knowledge-intensive process [114], requiring knowledge about the domain and various computer science and mathematical methods. It involves the careful manual work of domain experts, modelers, and computer scientists to distill a compact, executable representation of hypothesized mechanisms that helps answer the questions of interest. Data is used as input for the simulation model and for its calibration or validation. The mechanisms that form a validated simulation model support reasoning about the system's behavior. Thus, the resulting model not only predicts behavior but is often also used to explain the behavior of the modeled system, thereby enabling a deeper understanding and effective manipulation.

A question that has appealed to researchers' interests for a long time [65], but recently received renewed attention [33, 84], is whether one might automatically infer mechanistic models, in particular of dynamic systems, from available time series data. This interest is motivated and enabled by the growing volume of high-quality data, driven by advances in sensor and experimental technologies. Learning *mechanistic* models of dynamical systems differs from training phenomenological models, like neural networks, which focus on prediction, but not necessarily explanation.

Here, we focus on a powerful class of mechanistic models, called chemical reaction networks (CRNs), which are ubiquitous in chemistry and biochemistry. They describe, in the form of reactions ($\mathcal{R}_0 : A + B \to C$, i.e., "$A$ and B react to C"), how individuals from homogeneous populations (here A and B) interact and are consumed to form products (here C). A reaction's speed is described by a

Authors' address: Justin Noah Kreikemeyer, justin.kreikemeyer@uni-rostock.de; Adelinde M. Uhrmacher, adelinde.uhrmacher@uni-rostock.de, University of Rostock, Rostock, Germany.



kinetic function f of the current system state and one or more model parameters $\mathbf{k}$. In the simplest case, the so-called law of mass action is assumed and $f_0(A, B, C, k_0) = k_0AB$.

CRNs were originally introduced to model chemical reactors [39]. Together with their descendant rule-based extensions [38], they also form the de facto standard formalism for modeling the dynamics of cells in biochemistry [61]. CRNs find application in chemical engineering [104], astrochemistry [146], ecology [102], and epidemiology [4] as well. The widespread application of CRNs likely stems from the intuitive modeling metaphor provided by chemical reactions and the possibility to assign one of several semantics to them, connecting them to various other popular model classes, most prominently ordinary differential equations (ODEs) and continuous-time Markov chains (CTMCs). They also closely relate to the stochastic or continuous Petri net and process algebra formalisms [53]. Modeling with CRNs puts interactions into focus rather than state variables and their changes, as in ODEs. Hence, although called "chemical" after their original field of application [39], they are not exclusive to this field and are sometimes also referred to as biochemical reaction network (BRN) [75], kinetic reaction network [93], (stochastic) population model [14, 68], or simply reaction model [143]. Since a set of reactions forms a network between reactants and products, the term network is often used interchangeably with model or system.

Nowadays, the problem of kinetic parameter estimation for known CRN structures (reactions) based on time series has many mature solutions [75, 86], ranging from simulation-based optimization [95, 98], over gradient matching procedures [76, 129, 130] to likelihood-free/simulation-based inference [31]. The new frontier is thus the inference of the CRN itself, including reactions, kinetic laws, and corresponding parameters. This development is not unique to CRNs. Research on identifying ODE models from time series data has become increasingly popular over the last ten years [84]. One driving force is probably also the SINDy method [16], which made sparse regression over a library of basis functions widely known and applicable. The domain of symbolic regression, which aims to infer mathematical equations from input-output examples, has seen increased interest [33]. In this context, the inference of CRNs constitutes a rather specific yet important and practical problem. It enables learning ODEs interpretable as higher-level reactions [84], offers semantics beyond ODEs, and provides insights into the learning of models expressed in domain-specific languages.

1.1 Motivation and Objectives

Two observations motivate this systematic survey. First, due to the broad range of applications and also theoretical interest in CRNs, work on their inference is scattered across chemistry, biology, computer science, and applied mathematics (Section 4). Different terminology complicates grasping the big picture or even discovering related work. As a result, it often seems that previous progress, especially comparisons to other methods, is considered rather sparingly [117]. For example, methods based on the sparse regression principle have been described over the course of twenty years [18, 29, 55, 63, 109, 119], but we observe no references to the earlier works with the exception of Srividhya et al. [119] citing Crampin et al. [29]. Second, there is no current source (cf. Section 8.1) that provides a comprehensive introduction to the problem, existing solutions, and open challenges.

The objectives of this survey are thus to provide an overview and classification of current and past methods for solving the CRN inference problem, to illuminate prevailing challenges, and promote progress in this field and related fields. After introducing CRNs and their semantics (Section 2), Section 2.5 defines the problem of CRN inference from time-series data. Section 3 provides a brief summary of our systematic survey process. Its quantitative results are presented in Section 4. In Section 5, we extensively summarize the current design space of CRN inference methods, leading to a taxonomy (Section 6). This reveals unresolved challenges (Section 7) that provide inspiration for future work. Finally, Section 8 summarizes the relations of CRN inference to other areas from

which insights may be gained. The supplemental material details our systematic survey process and contains all associated artifacts such as code and tables[1].

2 CHEMICAL REACTION NETWORKS AND THEIR INFERENCE

A chemical system may be modeled on several levels of abstraction [122, Fig. 2], weighing accuracy against efficiency of simulation. For relatively complex systems or relatively long timespans, a good tradeoff is achieved by assuming (1) homogeneous populations of species (or molecules, entities, etc.) abstracting from positions and momenta of atoms and (2) a "well-mixed" system with molecules distributed homogenously in a fixed volume [122, Ch. 4].

CRNs describe systems on this level of abstraction. To account for the abstracted information, like effects of small variations in entity positions, appropriate stochastic effects are introduced. Its mathematically exact interpretation is then a chemical master equation (CME). The CME comprises a set of coupled ordinary differential equations (ODEs) describing the time-evolution of the probability distribution over species' amounts [45, 122]. This stochastic interpretation is particularly important in systems with small populations, where the choice of a single entity to interact with one or another can significantly affect the state trajectory. However, the CME quickly becomes intractable, as it considers every possible system state [45]. Hence, one from a hierarchy of semantics is employed to approximate the species' evolution over time and for extensions beyond applications in chemistry. For a comprehensive overview of the formalisms and methods used in biology to represent and simulate CRNs, see the review by Loskot et al. [75].

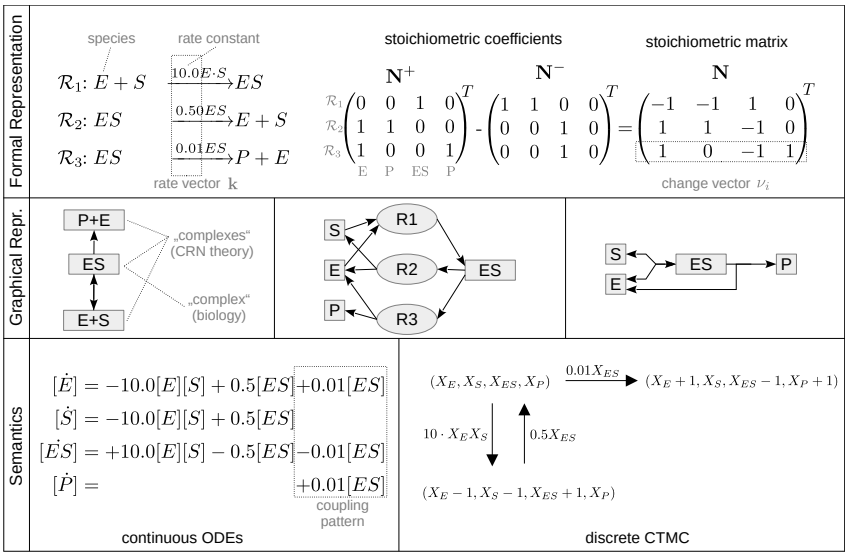


Fig. 1. The classical Michaelis-Menten model of an enzymatic process; the corresponding stoichiometry, graphical representations, and its interpretation as ODEs and CTMC. For didactic purposes, we disregard units and volumes. The ODEs show the characteristic coupling corresponding to the change vectors, in the given example, due to species E, ES, and P being involved in the last reaction. The reaction steps of this model may be aggregated in a single reaction following the Michaelis-Menten kinetic law (Section 2.3).

[1]In this preprint, the SI is attached at the end. Until publication, artifacts are available upon request.

2.1 Defining CRNs

We start by defining a finite set of m species S containing the names of all populations in the system. Based on S, a set of n reactions $\mathcal{R}$ is formed. A reaction consumes reactant species to generate product species. Commonly, definitions allow any "linear combination" [53, 3.4] or multiset [80] of species for reactants and products. For example, Martinelli et al. [80], define a CRN as a finite set of n reactions $\mathcal{R} = \{(R_i, P_i, f_i) | i = 1, \dots, n\}$, where R_i and P_i are multisets over S and f is a kinetic function (with a signature discussed later in this chapter). The reactants R_i can further be written as a linear combination: $R_i = r_{i,0}S_0 + r_{i,1}S_1 + \cdots + r_{i,m}S_m$, where the vector $\mathbf{r}_i$ contains the multiplicities in R_i for each species (likewise for products P_i). These sums are sometimes also referred to as *complexes*[2] [39, 40] and S_j is the name of one of $j = 1, \dots, m$ species in the model. Instead of the tuple (R_i, P_i, f_i), we can then conveniently write [53, 3.4]:

$$r_{i,0}S_0 + r_{i,1}S_1 + \cdots + r_{i,m}S_m \xrightarrow{f_i} p_{i,0}S_0 + p_{i,1}S_1 + \cdots + p_{i,m}S_m.$$

In this expression, we call $\mathbf{r_i}$ and $\mathbf{p_i}$ the *stoichiometric coefficients*. An example is provided in Figure 1. The situation where $\mathbf{r}_i = \mathbf{0}$ (or $\mathbf{p}_i = \mathbf{0}$) is often denoted as $\emptyset \rightarrow A$ (or $A \rightarrow \emptyset$).

Alternatively, CRN theory explicitly defines complexes that are related to form reactions [39]. Both approaches have a graphical representation. The definition by Feinberg [39] represents a CRN as a simple directed graph between complexes, sometimes referred to as the Feinberg-Horn-Jackson graph [126, Fig. 2], while the above definition defines either a bipartite directed graph between the species and reactions [106, Fig. 12] or a directed hypergraph between species [43, Fig. 1].

CRNs offer a flexible syntax with strong relations to other frameworks [53, 77], such as Petri nets and process algebras. For example, instead of the arrow notation $A \rightarrow B$, the bipartite relation between species and reactions inspires a direct representation as a Petri net. Each species is assigned a place, and transitions represent reactions. Arcs lead to (away from) a place if a reaction produces (consumes) the corresponding species. Extensions to the Petri net formalism have been developed to incorporate the stochastic or continuous semantics of CRNs. Both the classical and Petri net representations have been used in inferring CRNs from time series data (cf. Section 4).

From the above definition, several mathematical objects can be derived that aid the analysis and inference of CRNs. Inspired by chemistry, we can reason about the *stoichiometry* of reactions, i.e., the change in the counts of each species resulting from a reaction. The *change vector* of a reaction is $\nu_i = \mathbf{p}_i - \mathbf{r}_i$. Collecting all change vectors into a matrix leads to the *stoichiometric matrix* $\mathbf{N}$ of the CRN, containing for each reaction (column) and each species (row) the change in the system when the reaction is executed once. It is sometimes also convenient to talk about the stoichiometric coefficients themselves, which we can collect in two matrices $\mathbf{N}^- = (\mathbf{r}_1, \dots, \mathbf{r}_n)^T$ for reactants and $\mathbf{N}^+ = (\mathbf{p}_1, \dots, \mathbf{p}_n)^T$ for products, so that $\mathbf{N} = \mathbf{N}^+ - \mathbf{N}^-$. Given the kinetic functions, the matrices $\mathbf{N}^-$ and $\mathbf{N}^+$ can be used to fully specify a CRN. Thus, by determining these three, we can infer a CRN of fixed size (cf. Section 5.2.2). It is possible to work with rational stoichiometries [124], but we make the common assumption that the entries in $\mathbf{N}^+$ and $\mathbf{N}^-$ are non-negative natural numbers $\mathbb{N}_0$.

2.2 Semantics of CRNs

We now briefly present the most important interpretations of a CRN, namely as a CTMC and by fluid approximation. The latter is also referred to as the reaction rate equation semantics [45], or simply as the continuous "ODE" semantics. Whereas a CTMC directly implements the CME via the Forward Kolmogorov Equation, the fluid approximation arises in the limit of large populations, where the distribution over species counts can be faithfully approximated by just its mean [122]. Between these extremes lie moment equations, ODEs that approximate the CTMC's distribution

[2] "Complex" is herein used in its CRN theoretic meaning that is different from its use in biology/chemistry, cf. Figure 1.

up to n moments (typically $n = 2$, the mean and variance/covariance). They can be solved using moment closures or the linear noise approximation [116].

First, the state of the dynamical system needs to be defined. Here, it is a vector-valued function $\mathbf{X}(t)$ mapping from time t to either the abundances ($\mathbf{X}\colon \mathbb{R}_{\geq 0} \to \mathbb{N}_0^m$) or concentrations ($\mathbf{X}\colon \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}^m$) of all species m. In either case, $X_j(t)$ is the number (or concentration) of species S_j at a certain time t. Often, in an "abuse of notation", X_j and S_j are used interchangeably to mean either the species object or its current abundance in the system. Sometimes, $[S_j]$ is also used to refer to the concentration of species S_j. We can now also provide the signature of the kinetic functions $f_i : X \times \mathbb{R}_{\geq 0}^s \to \mathbb{R}_{\geq 0}$, where $X = \mathbb{R}_{\geq 0}^m$ (concentrations) or $X = \mathbb{N}_0^m$ (abundances). The second argument of f_i is a vector of kinetic parameters $\mathbf{k}_i$. The semantics of a CRN then only differ in how $\mathbf{X}(t)$ is obtained from $\mathcal{R}$ given $\mathbf{X}(0)$, i.e., how the system moves from one state to the next.

In the CTMC semantics, each reaction induces a transition in a Markov chain, moving the system from its current state to a new state as specified by the change vector (cf. Figure 1). For reaction i, a transition $(\mathbf{X}(t), \mathbf{X}(t + \Delta t))$ can occur in the chain, where $\mathbf{X}(t + \Delta t) = \mathbf{X}(t) - \mathbf{r}_i + \mathbf{p}_i = \mathbf{X}(t) + \nu_i$, iff $\mathbf{X}(t) \geq \mathbf{r}_i$. I.e., a reaction can occur only if the necessary quantities of reactants are available. The probability of each transition and the time step Δt are determined by the *propensity* $\alpha_i \sim \mathcal{E}(1/f_i(\mathbf{X}(t), \mathbf{k}_i))$, which is based on the kinetic function. Here, $\mathcal{E}(\lambda)$ is the exponential distribution with rate λ. The stochastic simulation algorithm (SSA) [45] and its extensions deliver exact sample trajectories through the CTMC, converging to the CME solution in the limit of infinite samples.

The ODE semantics describe the evolution of the mean state $\mathrm{d}\mathbf{X}(t)/\mathrm{d}t = \dot{\mathbf{X}}(t) = \mathbf{N}\mathbf{f}(\mathbf{X}(t), \mathbf{k})$, where $\mathbf{f}$ is the vector of all kinetic functions f_i. For a single species, we thus have $\dot{X}_j(t) = \sum_i \nu_{i,j} f_i(\mathbf{X}(t), \mathbf{k})$, i.e., each reaction contributes a change (such as 1 or -2) to the evolution of the species concentration X_j, multiplied by its kinetic function evaluated in the current state. The evolution of species X_j is thus given by a combination of all contributions by reactions that either consume or produce it. Note that the equation systems obtained in this manner form a subset of all ODE systems, where certain terms are shared between the derivatives. This is commonly referred to as a *coupled* ODE system, where the coupling is characteristic of CRNs. Consequently, this survey is restricted to methods that can infer ODE systems with the CRN coupling (cf. Section 3). We can now solve the initial-value problem to obtain the system's mean trajectory. In many practical cases, this is only feasible via numerical integration methods like the classical Euler method or using advanced numerical integration software such as LSODA [100] or CVODE [23].

2.3 Kinetics

The kinetic function controls how fast a reaction proceeds. A common assumption is that, in a closed volume, the velocity scales with the number of reactants available. The more reactants, the higher the concentration, and the more likely a collision leading to a reaction will be. This is summarized in the *law of mass action* [132], where, considering the probabilistic CTMC semantics:

$$f_i(\mathbf{X}(t), \mathbf{k}_i) = k_{i,0} \prod_j \binom{X_j(t)}{r_{i,j}} := k_{i,0} \prod_j \frac{X_j(t)!}{r_{i,j}!(X_j(t) - r_{i,j})!}.$$

For example, given the reaction $2A + B \xrightarrow{0.1} A$, $f_i(\mathbf{X}(t), \mathbf{k}_i) = 0.1 \cdot A(A-1)B/2$. Note that for mass-action reactions, $\mathbf{k}_i$ is scalar and we conveniently only write the *rate constant* $k_{i,0}$ over the reaction arrow. This equation accounts for the fact that after choosing one reactant, there is one less reactant available to choose from [45, 53]. For the continuous ODE semantics, where $X_j(t)$ is assumed large, $A(A-1) \approx A^2$ and this can be simplified to $f_i(\mathbf{X}(t), \mathbf{k}_i) = 0.05 \cdot A^2 B$. Note how the factorial is commonly absorbed into the rate constant.

Other and extended laws are also possible, which aggregate multiple elementary reactions (so-called *reaction steps*) into one (Michaelis-Menten, Figure 1) or model temperature dependency (Arrhenius), to give two prominent examples. We refer the reader to Loskot et al. [75, Sec. 3] for further information. Finally, we call a CRN system *linear* if f_i is linear in the state, e.g., mass-action systems with only one reactant ($A \longrightarrow B$) or no reactants ($\emptyset \longrightarrow B$).

2.4 Relation to Other Population Models

The CRN is one formalism among a range of descriptions of population dynamics, especially given its ODE semantics. In contrast to general ODE models, CRNs provide a specific view of the mechanics behind a system. The coupling introduced by reactions makes groups of ODE terms directly interpretable as interactions.

Further, S-Systems were developed to efficiently infer parameters and structure for biochemical systems [110]. These ODE models are based on power laws but allow only (non-coupled) one-to-one interactions and cannot enforce integer stoichiometries. Hence, they do not allow direct conclusions about the CRN structure, such as stoichiometry [144, p.2][113, p.328].

Another related formalism are gene regulatory networks (GRNs) [77]. However, they provide a high-level abstraction of the underlying system, often based on Boolean interpretations. Due to a plethora of available datasets, GRN inference is now supported by various tools and benchmarks [79]. The inference of CRN models is comparatively less researched; consequently, the learning of signaling and metabolic processes, which are commonly modeled using CRNs in systems biology [77], remains underexplored. For further information, we refer the reader to Machado et al. [77] and Loskot et al. [75].

2.5 The CRN Inference Problem

We define the problem of inferring a CRN from time-series data as follows (cf. Figure 2). Generally, it falls into the category of *system identification* and *inverse problems* (Section 8.6). Sometimes, it can also be decomposed into a supervised learning task (cf. Section 5.2.1). In some cases (Section 5.2.3), it thus becomes a special case of the *symbolic regression* machine learning problem [33], where parameters and structure of an expression are inferred from input-output data. What is regarded as input and output varies depending on the problem formulation (cf. Section 5.3).

As input, the algorithm is provided with measurements $D = \{(t_l, \tilde{\mathbf{X}}(t_l)) | l = 1, \ldots, L\}$ of all ($\tilde{\mathbf{X}}(t)$) or a subset ($\tilde{\mathbf{X}}(t)' \subset \mathbf{X}(t)$) of species at L discrete points in time $t_1, \ldots, t_L$. For convenience, we refer to a single such trajectory as $\{\tilde{\mathbf{X}}(t_l)\}_{l=1}^{L}$. Depending on the situation, multiple traces $D_1, D_2, \ldots$, e.g., with different initial conditions or realization of intrinsic randomness (replicates), are provided. Note that the Ds may not share the same set of discrete time points. The task of the algorithm is then to infer a CRN, i.e., a set of reactions $\mathcal{R}$, that (1) matches the dynamics of (all) D and (2) is, in the best case, also informative about the structure of the system that generated (all) D. In the case of extrinsic noise, the requirement (1) is to be adjusted to account for the fact that each measured D requires different kinetic parameters [28]. We assume that D stems from the transient/equilibration phase of the reference system including, e.g., oscillatory behavior.

It has long been known that an exact solution to this problem is inaccessible with commonly available datasets, even for linear systems [46]. For example, in biology, we may often only have a single trace of the system to work with. However, to ensure a linear CRN can be uniquely identified from time-series data, knock-out experiments are required [46], i.e., a set of trajectories where, respectively, one species is initially completely eliminated from the system (has concentration or abundance zero). The main reason for the inaccessible exact solution, even with high-quality data, is that one can generally identify *multiple* CRN consistent with the same time-series data [26, 37]. In chemistry, this problem is sometimes referred to as distinguishability [109, 121] or the

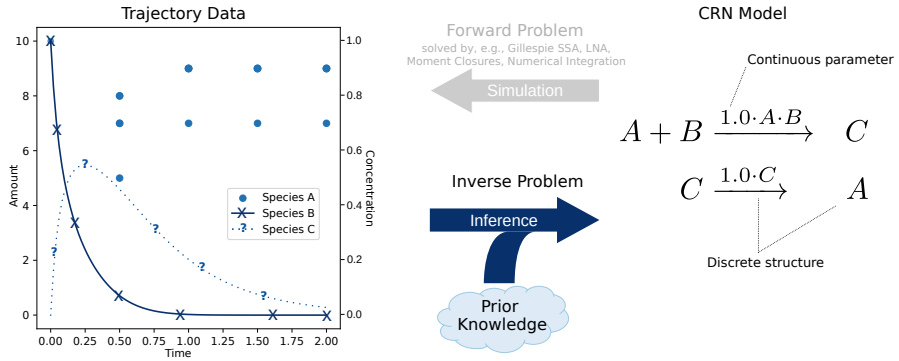


Fig. 2. Visual summary of the CRN inference problem. The *forward problem* simulates a given model to obtain deterministic (line) or stochastic (dots) trajectory data. The *inverse problem* discussed here faces the other way: Given trajectory data and knowledge introducing bias or constraints, an appropriate CRN (structure and parameters) is sought. Not all species trajectories may be measurable (cf. species C).

"fundamental dogma of chemical kinetics" [27, 120]. In system identification, such a system is termed *not identifiable*. Further, several formulations of the problem have been proven to be NP-hard [49, 91], so even given sufficient data, we expect no *exact* algorithm of less than exponential complexity exists. However, we can still hope to obtain reasonable *approximate* solutions, even for wild-type traces, by using relaxed problem formulations (Section 5.2) and applying inductive bias or additional constraints from domain knowledge (cf. Section 5.4).

3 LITERATURE SEARCH SUMMARY

Before continuing with the results, we briefly summarize the method employed for our systematic literature search and screening (cf. Figure 3). For the details, we refer the reader to the supplementary material[3]. We started with an initial list of 18 "seed publications", comprising those the authors had already identified as relevant in their previous work in this field and that also met the inclusion criteria (see supplement). Criteria included restrictions on input (time series data) and output (CRN, coupled ODEs), as well as distinctions from other methods, like those targeting GRNs. A structured list of keywords was extracted from the seed publications, pertaining to several aspects of the problem, such as automation ("autonomous") or the input data ("time series"). Large Language Artificial Intelligence Models (LLMs) were used to augment the list. Queries for five major literature databases were formulated and also refined using an LLM's perspective for inspiration. Through Scopus [36], seven preprint servers were included as well. Finally, we included the results of a previous bachelor thesis [128] that queried Google Scholar [48] and two "Deep Research" tools based on LLMs [47, 99]. We considered peer-reviewed and pre-print publications published until the cutoff date, 1st December 2025 (all queries were run or re-run after this date, so a few more recent entries may have been included as well). In total, 2393 publications were returned for our queries (2120 after automated deduplication). A prescreening based on only title and abstract was made, leaving 93 likely positive candidates. Based on their full texts, 48 were found to align with our selection criteria, covering 16 of the 18 seed publications. From each of these and the two missing seed publications, features were extracted into a large table. The table was augmented by a backward- and also automatic recommendation search, leading (after preselection and selection) to 21 new publications in the table. The final literature sample thus consists of 71 entries.

[3]Attached to the end of this document.

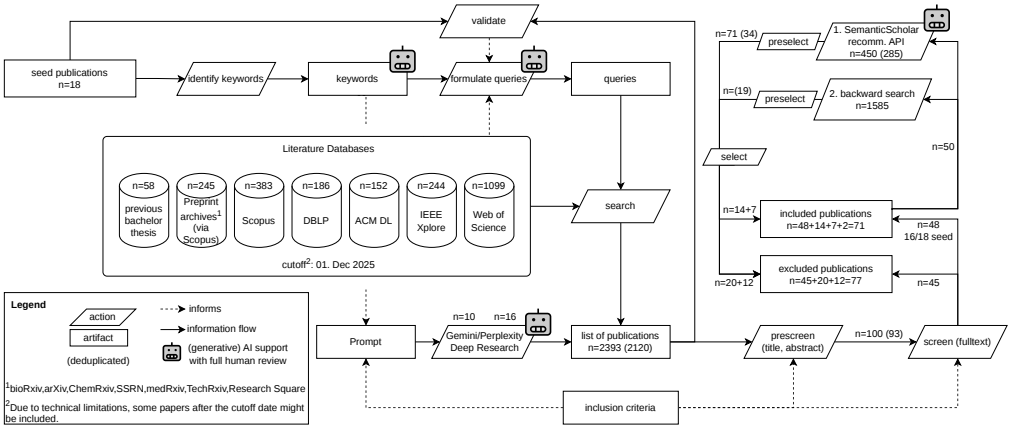


Fig. 3. Flowchart of our literature survey workflow. Previous bachelor thesis refers to van der Wall [128].

4 THE FIELD IN NUMBERS

Figure 4, Figure 5, and Figure 6 give a birds-eye view of the most important statistics characterizing our publication list and the methods described therein. Percentages given there and in the following text are rounded to one significant digit. To keep this section readable, sometimes only a few representative examples are given in the citations, and we refer the reader to the complete table provided in the supplementary material (cf. the end of Section 1). We use n to refer to the absolute number of publications and % to refer to the fraction of publications from our sample of size $n = 71$.

4.1 Bibliographical Data

We observe that interest in the topic fluctuates over the years (Figure 4), with a slowly increasing trend in the number of publications per five-year period. Particularly in 2019 and during the last three years, we observed increased interest.

Interest in the problem is motivated by four major directions: biology ($n = 17$, [62, 69, 140], including computational biology, systems biology, biochemistry, synthetic biology), chemistry ($n = 19$, [58, 71, 142], including computational chemistry, chemical engineering, chemical computing), computer science ($n = 24$, [14, 40, 67], including machine learning, evolutionary optimization), and applied mathematics ($n = 11$, [21, 25, 74], including CRN theory). These categories were determined from the full text based on keywords (e.g., “biochemistry”, “chemistry”, “systems biology”), case studies used, and the corresponding publication venues. Each publication was assigned three major motivating domains in increasing order of specificity (“biology” vs. “systems biology”). Figure 4 shows a word cloud based on the respective weighted (by a factor of three to one) counts.

4.2 Input and Output

Time Series. Many approaches can handle multiple time series. This includes stochastic replicates with the same initial conditions to infer models in the stochastic setting [62, 147], or trajectories with different initial conditions for additional information during the inference when inferring deterministic models [55]. For example, when inferring temperature-dependent mechanisms, time series originating from different temperatures need to be provided [71].

Prior Knowledge. The integration of prior knowledge beyond the time series data D plays a major role and is supported (or required) and actively discussed by around three-quarters (76%) of publications. Prior knowledge may include anything from constraints on parameter values or

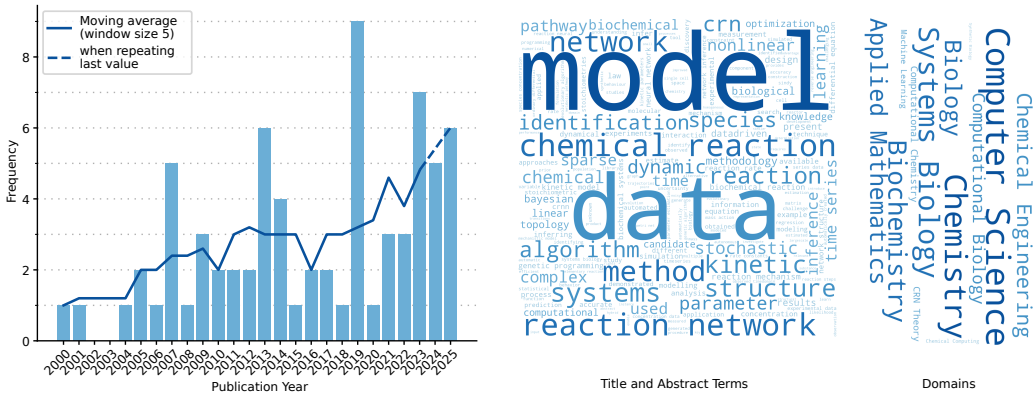


Fig. 4. Bibliographical results: Number of publications per year, word clouds over titles/abstracts and domains.

model properties/structural constraints [69, 136] over reaction candidates [55, 59] and candidate modules [20, 107, 139] to known partial models [62, 69].

Syntax and Semantics. Most of the works (82%) focus on the deterministic ODE semantics ($n = 58$, e.g., [6, 40, 55, 69]), but some (additionally) consider the more exact stochastic semantics ($n = 13$,[13, 14, 21, 22, 62, 94, 112, 147]) through CTMC [21, 22, 112, 147], approximation with moment closures [62], or linear noise approximation [14]. Most publications refer to the standard syntax of a CRN as reaction equations, but several also use a stochastic Petri net representation [96, 105, 139, 140] and some infer a (subset of) the P Systems formalism [20, 107], which can be regarded as a superset of CRNs including compartments.

Specialized kinetics. With respect to kinetic functions, the majority of approaches consider only the mass action law ($n = 52$, 73.2%). Some approaches go beyond this and also consider Arrhenius ($n = 6$, [34, 57, 58, 71, 73, 124]), Hill ($n = 4$, [3, 80–82]), and Michaelis-Menten ($n = 5$, [43, 80–82, 97]) or even allow arbitrary predefined ($n = 6$, [8, 15, 42, 125, 137, 142]) kinetic laws to be inferred.

Output. Many of the methods discussed provide point estimates, i.e., one CRN is considered the best explanation of the input data ($n = 41$, 58%). Due to the non-identifiability, it is often not possible to claim that one model is the exact explanation sought. To this end, several publications also consider alternative explanations ($n = 21$, e.g., [6, 51, 80, 126, 142]) and uncertainty over parameters ($n = 1$, [22]), structure ($n = 3$, [13, 14, 40]), or both ($n = 5$, [43, 59, 73, 74, 92]).

4.3 Testing and Benchmarking

Case Studies. The number of case studies used to test an approach's capabilities is rather small. Almost all publications use three or fewer test cases (the median is two, cf., Figure 6). The distribution we obtained is also in line with that presented by Skrzypczak [117], which is based on a different selection of the literature, more focused on benchmarking. We treat each distinct ground truth model or experimental dataset as a single test case, and treat variations in the data (e.g., different noise levels) produced by the same ground truth system as a single case. Following Skrzypczak [117], we distinguish testing on data from simulations of synthetic CRNs ($n = 40$), CRNs taken from realistic applications, ($n = 44$), or even experimental data ($n = 14$). The gold standard of testing across all three categories is achieved by only a handful ($n = 5$) of publications.

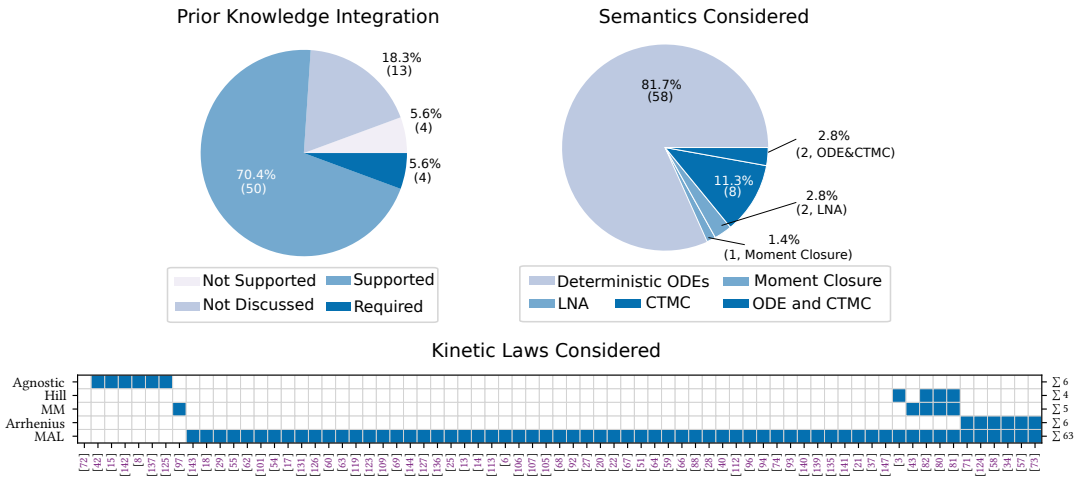


Fig. 5. Slected quantitative results on inputs and outputs: integration of piror knowledge, consideration of CRN semantics, and kinetic laws inferred. Most methods rely only on deterministic semantics. The inclusion of prior knowledge is a central topic. The majority of methods exclusively support the mass action law.

Noisy and Missing Data. To imitate practical scenarios, synthetic data generated by simulating a ground truth model is often corrupted with noise ($n = 33$, 46.5% of publications). A further 19.7% of publications ($n = 14$) consider experimental (noisy) datasets. In total, 59.2% ($n = 42$) of publications consider measurement noise in some way ($n = 5$ publications consider both, synthetic and experimental, noise). The way noise is applied in our literature sample is almost evenly distributed between relative and absolute noise. The predominant distribution is additive zero-mean Gaussian noise ($n = 26$, 62% of the $n = 42$ publications considering noise). In the relative case, the variance may be set relative to each concentration measurement [137], to the maximum observed concentration [144], to the standard deviation of the time series trajectories [8], or to the range (difference between maximum and minimum) in the trajectories [40, 113]. It can either be assumed to be homogenous over species [144] or be different for each species [8, 40]. Absolute noise assumes a constant noise level that does not depend on the trajectories [27, 125, 127], and the variance may, like with relative noise, vary between species or not. Few publications consider other kinds of noise, such as log-normal ($n = 1$, [59]), white ($n = 2$, [8, 136]), binomial ([62]), or real experimental data ($n = 14$). 31% of publications also consider the practically relevant case where the time series of one or more species is not available, e.g., as it could not be observed.

Runtime. The runtime performance of algorithms is discussed rather infrequently. Only around one-third ($n = 24$) of publications discuss in some way performance results regarding some time measurements ($n = 19$), computational complexity ($n = 3$), or optimization steps ($n = 2$, text only, figures not counted) of the inference process.

Source Availability. A final relevant statistic is that only $n = 17$ (around a quarter) of the publications made the source code or data supporting them readily available ($n = 45$ unavailable, $n = 6$ not accessible anymore, $n = 1$ not yet available, $n = 2$ upon request, and $n = 17$ available). Of these 17, only five (7% of all publications) were uploaded to a permanent archive such as Zenodo or included code in the appendix/supplemental material (pseudocode not counted). So most of the method's

implementations are inaccessible or are at risk to become inaccessible in the future[4]. Further, the available implementation is sometimes specialized to the examples in the respective publication, making its reuse difficult.

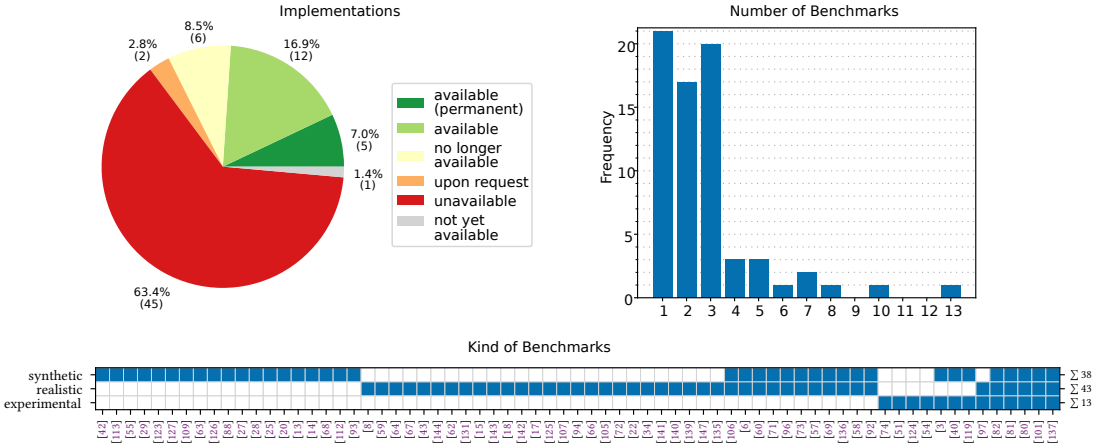


Fig. 6. Quantitative results on evaluation practices: implementation availability, number, and kind of benchmarks. Implementations are often unavailable, and only a few benchmarks are considered. Benchmarking rarely considers experimental data and mostly relies on synthetic or realistic ground-truth CRNs.

5 CRN INFERENCE BUILDING BLOCKS

We now outline the extent of the current design space for CRN inference methods, summarized in Figure 7 and 8. In our literature sample, we can distinguish several categories of building blocks. Methods are formed by combining these basic components. Note that the presented results are based on current publications. Further building blocks are likely to emerge in the future.

As the exact solution of the CRN inference problem is inaccessible, approaches employ relaxed search spaces (Section 5.2) and inductive bias (Section 5.4). An important distinction also regards the error (loss) formulation (Section 5.3). Several preprocessing steps (Section 5.1) may be employed to prepare and aid the former three. This leads to different problem formulations providing a tradeoff between fidelity to the original problem and tractability.

For example, an exhaustive search (Section 5.2.4) within certain bounds is highly accurate, but requires vast computational resources [42, 142]. On the other hand, optimizing a template stoichiometric matrix (Section 5.2.2) is more efficient, but the combined discrete-continuous optimization landscape leads to less accurate solutions [57, 68]. Approximations are also motivated by practical constraints, like data availability. For example, formulating the problem in derivative space (Section 5.3.1) may result in a convex optimization landscape [55], whereas fitting to state trajectories (Section 5.3.2) requires considerating the non-linear relationship between the CRN and its behavior over time. However, derivatives are often not directly measurable and have to be approximated (Section 5.1.1), decreasing accuracy.

In this way, the existence of a "best" (class of) approach that is optimal in all scenarios is unlikely. CRN inference is also affected by the famous No Free Lunch (NFL) theorems [138]. The NFL emphasizes that good performance requires matching the algorithm to the structure of the problem. They indicate "the importance of incorporating problem-specific knowledge into the behavior of

[4]Note that, using tools like the Internet Archive at web.archive.org, some can be recovered, but we found this is not guaranteed in our literature sample ($n = 5$ of the six URLs now offline were not archived there).

the [optimization] algorithm" [138, p.77] (Section 5.4). This is acknowledged by the vast majority of the methods, which employ inductive biases (87%), e.g., based on prior knowledge (76%).

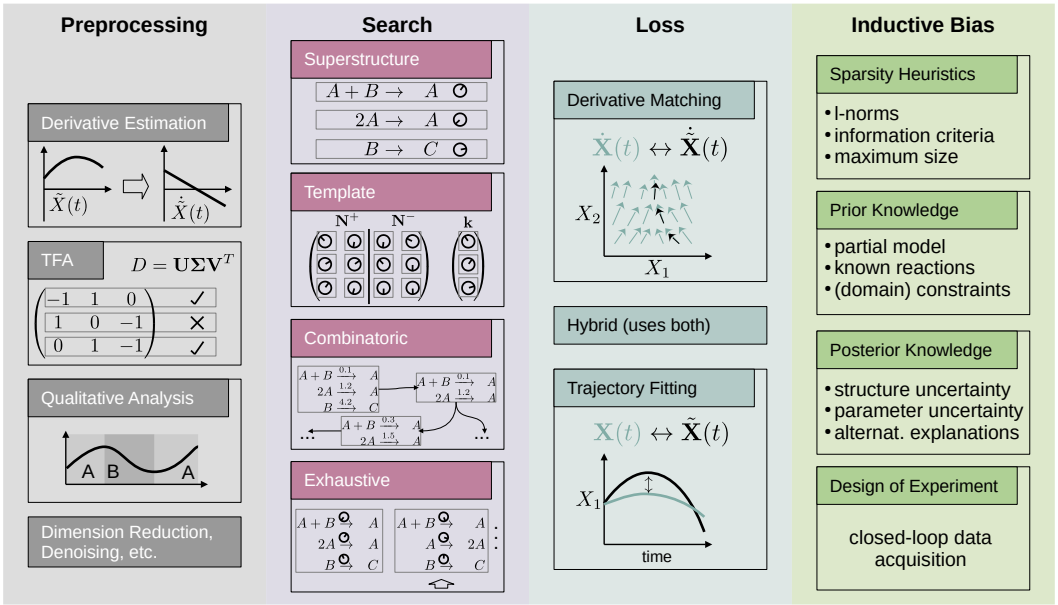


Fig. 7. We identified four major types of building blocks that make up CRN inference algorithms: search space (Section 5.2), loss formulation (Section 5.3), and techniques to introduce inductive bias (Section 5.4). Also, a variety preprocessing steps (Section 5.1) are used to support the former.

5.1 Data Preprocessing: Making the most of the Data

Before starting the automatic inference process, the measurement data D can be transformed and scrutinized to derive useful quantities or even draw first conclusions about the underlying CRN.

5.1.1 Derivative Estimation and Resampling. Most importantly, when using the derivative-matching approach (Section 5.3), the derivative must be obtained from measurements. It can be trivially estimated with finite-difference methods, like central differences:

$$\dot{\mathbf{X}}(t_l) \approx \dot{\tilde{\mathbf{X}}}(t_l) = \frac{\tilde{\mathbf{X}}(t_{l+1}) - \tilde{\mathbf{X}}(t_{l-1})}{t_{l+1} - t_{l-1}}.$$

The disadvantage of this approach is that large measurement intervals lead to bias and noise in the data being amplified. Consider the case where $\tilde{\mathbf{X}}(t_{l+1}) \sim \mathcal{N}(\mathbf{X}(t_{l+1}), \sigma^2)$ and $\tilde{\mathbf{X}}(t_{l-1}) \sim \mathcal{N}(\mathbf{X}(t_{l-1}), \sigma^2)$ under the i.i.d. assumption. Then the variance of $\dot{\tilde{\mathbf{X}}}(t_l)$ will be $(2\sigma^2)/(t_{l+1} - t_{l-1})^2$. Additionally, a larger time step $t_{k+1} - t_{k-1}$ increases the bias wrt. the true derivative $\dot{\mathbf{X}}(t)$.

To avoid problems with noise and measurement intervals, derivative estimation may rely on probabilistic filtering [50], e.g., via Gaussian processes [76], smoothing splines [32], polynomial models [54], or surrogate methods like Automated Learning of Algebraic Models for Optimization (ALAMO) [24, 137] and the Dynamic Response Surface Methodology (DRSM) [42, 142–144]. All of these may alleviate the problems above, but only at the cost of increased bias. The latter two surrogate models, once fitted, allow differentiation at any point [137]. Fitting a model to the data may thus also be used beyond differentiation for resampling at different discrete time points, i.e., obtaining measurements in between the original sample times by interpolation.

5.1.2 Noise reduction. Even when not relying on derivative estimation, reducing measurement noise is beneficial. It may rely on the methods discussed earlier, such as probabilistic filtering and smoothing. Further, if there is an estimate of the measurement noise or a known noise model, it can be used to denoise a time series much more efficiently [21, 76]. If no estimate on the noise is available, prior knowledge such as atomic balances can be used to reduce noise [15, Sec. 3.1.3].

5.1.3 Qualitative Analysis. An interesting approach that has not yet been extensively applied in CRN inference is qualitative analysis [111, 131]. With this approach, the first step is to extract episodes (patterns) from the time series data based on first- and second-order derivatives. Episodes may include classes like "upward", "downward", "constant", but also "oscillation". Reactions can then be inferred from episodes alone [131]. For example, if species A decreases, and species B increases during the same episode, the reaction $A \longrightarrow B$ could be postulated. Note, however, that the definition of meaningful episodes may require first- and even second-order derivatives, which are difficult to approximate from noisy time-series data (Section 5.1.1).

5.1.4 Target Factor Analysis. The stoichiometric matrix $\mathbf{N}$ provides valuable information about the net changes induced by the reactions underlying the time series data. Factor analysis can be used to obtain an abstract stoichiometric matrix $\mathbf{N}_a$ of linearly independent stoichiometric vectors that span the stoichiometric subspace of the CRN [12]. This is done by singular value decomposition $D = \mathbf{U}\Sigma\mathbf{N}_a$. The stoichiometry of any reaction, i.e., any column of $\mathbf{N}$, is then expected to be a linear combination of these stoichiometric basis vectors. Further, the dimensionality of the basis corresponds to the expected number of linearly independent reactions in the CRN.

The method of *Target Factor Analysis* (TFA) can be employed to test, for a given stoichiometric vector v_i, whether it is compatible with the abstract stoichiometric matrix $\mathbf{N}_a$ [12]. TFA may thus be used to prune the search space by excluding any reactions with an incompatible stoichiometry, i.e., that are not a linear combination of the basis stoichiometries [15, 42, 142–144]. Bonvin and Rippin [12] also show how noisy data and prior knowledge can be accounted for.

5.1.5 Others. Generally, other dimension-reduction methods can also be applied to analyze D a priori and obtain useful information for the inference process. However, we found that TFA was the only method used in our literature sample. Additionally, the time series may be analyzed for correlations between species trajectories. These contain information about the species' interaction graph. For example, the method Correlation Metric Construction (CMC) uses this fact to reconstruct an underlying relation graph between species [108]. To this end, relational inference methods may also be employed (Section 8.7).

5.2 Search: How and What to Optimize

After preprocessing the data, a search space needs to be defined and a method devised to traverse it. In the literature, several main approaches can be distinguished.

5.2.1 Superstructure. Superstructure approaches ($n = 26$ publications/25 methods) start with a large, overspecified model and subsequently reduce it to a smaller, well-fitting model. This approach can also be regarded as "general-to-specific" [29, 119], or "backward-elimination" [41]. The superstructure is typically built from a library of possible reactions, which can be obtained from chemical reaction generators [44], full enumeration [18], or prior knowledge [55]. Note that enumerating all or a large fraction of all possible reactions drastically increases computation time and difficulty [69, 109, 119]. The library may thus be pruned using, e.g., TFA [144] (Section 5.1.4) or evolutionary methods [69].

In the continuous semantics, such a superstructure induces a large coupled ODE model containing the (coupled) terms of all possible reactions. Several options are available for reducing the superstructure. The first one uses stepwise heuristics, such as a greedy selection [29, 30, 71, 119]. Also, a likelihood-based selection may be used [25, 27, 28, 74]. The second is to employ integer optimization like mixed-integer (non-)linear programming (MINLP) for the selection of reactions [125, 137, 144], where integer coefficients steer inclusion (1) and exclusion (0) of reactions. Finally, the problem can be relaxed into a continuous optimization problem by exploiting the kinetic law's constants. In this case, any rate constant $k \approx 0$ leads to the exclusion of a reaction and any other value to inclusion with the constant's value [18, 55, 62, 68, 109]. However, it is important to consider the multiscale nature of CRNs to determine how close to zero the constants should be to warrant exclusion (Section 7.6). Especially in the continuous approach, the problem needs to be constrained to solve for non-negative parameters.

The latter two approaches are often referred to as *sparse regression*. The objective is generally formulated as:

$$\arg\min_{\mathbf{k}} ||\dot{\tilde{\mathbf{X}}}(t) - \Theta(\tilde{\mathbf{X}}(t))\mathbf{k}||_p + \Lambda(\mathbf{k}), p \in \mathbb{N},$$

where Θ is the library of reactions applied to the measured time series data, Λ is a regularizing function like $||\mathbf{k}||_2$, and $\mathbf{k}$ are (differing from Section 2.2) the coefficients of each reaction, assuming an appropriate kinetic law where each k_i enters as a simple factor (e.g., mass-action, Section 2.3). Compared to the standard SINDy formulation [16], the library matrix has an additional dimension, and the unknowns are a vector (of rate constants) instead of a matrix. This enables enforcement of the coupling (cf. Figure 1 and also [69, Fig. 1]). The linear relation between $\dot{\tilde{\mathbf{X}}}(t)$ and $\mathbf{k}$ leads to a very simple convex optimization problem. However, derivatives approximation (Section 5.1.1) and linear dependencies between library terms pose challenges. Further, the non-negativity of $\mathbf{k}$ needs to be ensured. It may thus be solved with, e.g., $p = 2$ and non-negative least squares algorithms [18, 55, 69]. For further details on how to implement the sparse regression approach, the reader is referred to Métayer et al. [84, Algorithm 2], which presents it in the context of general ODE inference. Kinetics beyond those linear in $\mathbf{k}$ can also be supported, including multi-parameter formulations such as the Arrhenius law. However, then the problem may no longer be convex.

The loss formulation above is sometimes also called *gradient matching*[5] [62, 97, 129][101, SI4] and has long been applied for parameter estimation [129, 130], but also the superstructure approach [16, 30, 55]. We can also compare states instead of derivatives (Section 5.3.2) if we simulate candidate CRNs. In the case of the continuous semantics, this amounts to (numerical) integration [59, 71]. In the case of the discrete semantics, the SSA [45] is employed [147].

Especially with the superstructure approach, a problem arises from linear dependencies in the postulated reaction library. For example, if one stoichiometric vector (e.g, $B \longrightarrow A + B$, $\nu = (0, 1)^T$) is just a scaled version of another (e.g., $B \longrightarrow 2A + B$, $\nu = (0, 2)^T$), it is not possible to distinguish between them without additional information. In the previous example, setting the rate of the second reaction to twice that of the first will have the same impact on the deterministic trajectories. By considering stochastic semantics (and data), we may still distinguish between them, e.g., as one may produce a higher variance [37, Fig. 1].

5.2.2 Template Optimization. The second class of search ($n = 17$ publications/16 methods) uses a specific kind of superstructure directly related to CRNs, namely the two stoichiometric constant matrices $\mathbf{N}^-, \mathbf{N}^+$. They optimize the stoichiometric constants in $\mathbf{N}^-, \mathbf{N}^+$ and kinetic law parameters

[5]This name might be misleading at first, as $\dot{\mathbf{X}}(t)$ is a tangent (derivative) vector and not formally a gradient vector. It probably stems from physics, where the matched ODE drift is often the gradient of a potential.

$\mathbf{k}$ of f directly. This assumes a system with a size limited by $\mathbf{N}$, although smaller CRNs can be inferred by either setting the columns in $\mathbf{N}^-, \mathbf{N}^+$ or the respective rate constant $\mathbf{k}$ to zero.

Generally, this formulation results in a combination of discrete ($\mathbf{N}^-, \mathbf{N}^+$) and continuous ($\mathbf{k}$) optimization parameters. However, the discrete part may be relaxed to continuous values [58, 68]. Another option is to recognize that, given a maximum stoichiometry, the constants may be treated as categorical data and thus one-hot encoded [14]. This introduces some new artificial dimensions that may benefit optimization. In the end, any numerical optimizer may be used, such as gradient descent [13, 14, 57, 58, 88] or evolutionary algorithms [112].

A relatively recent and popular approach is the chemical reaction neural network (CRNN) [58] and its extensions [34, 73, 92]. In this approach, $\mathbf{N}^-$, $\mathbf{N}^+$, and the mass-action and Arrhenius laws are encoded as continuous weights of a specialized neural network. Pruning and rounding the weight matrices allows a direct interpretation of the trained CRNN as parametrized CRN model.

In the template formulation, the relation between the optimized parameters (stoichiometric constants) and the objective is relatively complicated. The effect of changing a stoichiometric coefficient on the derivatives and trajectories produced by the CRN can be rather drastic. Consider, for example, $A \longrightarrow B$ with mass-action kinetics kA. Increasing ν_0 by 1 leads to $2A \longrightarrow B$ with kinetics kA^2. Further, many undesired local minima exist. These problems have been tackled with partial parameter freezing [57] or random restarts [68].

Finally, instead of the stoichiometric matrix, some publications optimize the adjacency matrix $\mathbf{K}$ of the directed graph over complexes (Section 2.1) in a CRN theoretical interpretation [3, 60, 126, 127]. This is done by decomposing $\mathbf{N} = \mathbf{Y}\mathbf{A}_\kappa$ with $\mathbf{A}_\kappa = \mathbf{K} - diag(\mathbf{K}^T\mathbf{e})$ [3].

5.2.3 Combinatorial. Combinatorial approaches ($n = 24$ publications/19 methods) move more freely through the space of possible parametrized CRN models. Instead of being limited to steps of reactions or stoichiometric constants, they can employ any kind of step in the parameter domain, structure domain, or both [43]. This category includes approaches that construct a CRN from the bottom up [80–82], also called "specific-to-general" [29, 119] or "forward selection" [41].

A prominent and large subcategory is *hybrid evolutionary algorithms* [105, 141]. Hybrid here refers to a combination of two optimization algorithms, one optimizing structure and one optimizing parameters [69, 72, 93, 94, 105, 106, 135, 140, 141]. Typically, the outer layer is an evolutionary algorithm that evolves structures. As one cannot easily judge the fit of a structure to the data without determining optimal parameters, the parameter optimization is wrapped inside its fitness function. So, with this approach, steps in the model domain (structure) are separated from those in the parameter domain. Sometimes, the (re-)optimization of parameters is only done every couple of structure steps (generations), e.g., every 250 generations [141].

It is also not strictly necessary to split the optimization of structure and parameters in this way: the evolutionary algorithm may evolve the continuous parameters as well [64, 66, 67, 72], possibly allowing steps in the search space not possible with the hybrid approach. However, for the parameter optimization of CRNs, more effective methods are accessible than for their structure. Choices for the parameter optimization in our sample include simulated annealing [105, 106, 141], particle swarm optimization [93, 94], genetic algorithms [107] or derivative matching [6, 69]. For more details on how the hybrid evolutionary approach may be implemented, the reader may also refer to Métayer et al. [84, Algorithm 1], which presents it in the context of general ODE inference.

Metaheuristics, such as simulated annealing, can also be used directly to traverse the structure search space [139]. A further method is to use a Markov Chain Monte Carlo (MCMC) approach to take steps in the structure, parameter, or both domains [43]. This has the advantage that the full posterior probability distribution over parametrized CRNs is approximated (Section 5.4.3).

Finally, a sequential selection of reactions can be used to construct CRNs bottom-up using greedy heuristics [30, 81, 82, 119] or tree search [80].

5.2.4 Exhaustive Search. The final approach to CRN inference we could distinguish is exhaustive search ($n = 8$ methods) [21, 42, 68, 101, 124, 131, 142, 143]. On first consideration, this may seem trivial and without practical relevance—and the latter is certainly the case for large problem instances involving many species. However, through appropriate preprocessing (e.g., [131]), it can yield new insights on practically relevant problems, as demonstrated by Taylor et al. [124].

In an exhaustive search, the CRN inference problem is relaxed to a parameter-estimation problem, much like in some superstructure formulations. However, instead of reactions, all their viable combinations are enumerated, together with kinetic laws. The model candidates' parameters are then optimized to assess their fitness with respect to the data. According to some criteria (discussed in the following two subsections), the best model is selected. It is also possible to construct a pareto front over possible models [21]. In any case, pruning of the possible reactions is crucial to ensure the tractability of exhaustive search approaches. Most commonly, this is done using chemical knowledge [124, 143], TFA [142] or also limiting the set of possible reactions based on the data [21].

Although this brings the problem of simulation model inference close to simple model selection and exploratory modeling [87], this automated exhaustive search can easily accommodate many situations that previous approaches have difficulty with. For example, unmeasured species may be easily postulated, and their missing trajectories may even be inferred from the existing data [142].

5.3 Error Function: Matching Derivatives vs. Fitting States

A central design decision are the inputs to error function used to assess the quality of solutions in the search space. It influences robustness to noise and further characterizes the search space, e.g., the "ruggedness" of the optimization landscape. To implement the error function, we need a way to compare a CRN candidate model with the provided data D.

A (deterministic) dynamic system is fully characterized by providing the "direction and speed", a velocity, for every one of its possible states. This is also called its characteristic *vector field*. However, concrete trajectories through the system's states form an integral curve or, when measured at discrete times, consist of a sequence of states traversed, e.g., $\{\tilde{\mathbf{X}}(t_l)\}_{l=1}^{L}$. When inferring a dynamical system, we can either compare these observed integral curves $\{\tilde{\mathbf{X}}(t)\}_{l=1}^{L}$ or observed velocities $\{\dot{\tilde{\mathbf{X}}}(t)\}_{l=1}^{L}$ to their counterparts induced by a candidate CRN (Figure 7). Comparison of the model's dynamics to the measured data D requires equal initial conditions and measurement times. Interpolation (Section 5.1.1) may be used to equalize the sampling points between data and inference.

5.3.1 Derivative Matching. The relation between model formulation and the resulting derivative space is more direct and often much simpler (e.g., linear) than its relation to produced trajectories. At any point in state space, each reaction $\mathcal{R}_i$ corresponds to a characteristic direction, given by its change vector ν_i, and its speed is given by its kinetic function f_i. So the vector field of a CRN is just the sum of the characteristic velocities of its reactions (Section 2.2), leading to a very simple relationship between reactions (or their parameters, cf. Section 5.2.1) and $\dot{\mathbf{X}}(t)$. In general, the derivative space loss can be formulated with the objective function:

$$G\left(\{\dot{\tilde{\mathbf{X}}}(t_l)\}_{l=1}^{L}, \{\dot{\mathbf{X}}(t_l)\}_{l=1}^{L}\right),$$

where G is a metric of choice, such as the mean squared error $G(\tilde{x}, x) = L^{-1} \sum_{l=1}^{L} (\tilde{x}_l - x_l)^2$, with L measurements. Although the term gradient matching is often used in this context (cf. Section 5.2.1), we herein refer to this approach as *derivative matching*, since $\dot{\mathbf{X}}(t)$ and $\dot{\tilde{\mathbf{X}}}(t)$ are *tangents* (and

not gradients) along the system's trajectory containing the derivative of each state variable wrt. time. To consider stochastic semantics, moment closures [62] or matching the transition rates of a CTMC [37] may be used. A disadvantage of derivative matching is the amplification of noise and/or bias in D due to approximation (Section 5.1.1). So, especially when data is scarce, the derivative matching approach may not yield the desired (accurate) result.

5.3.2 Trajectory Fitting. Fitting in state space is generally more difficult, as the relation between the model and its possible state trajectories is, most of the time, highly non-linear. Further, obtaining $\{\mathbf{X}(t)\}_{l=1}^{L}$ also generally requires simulation, which is time-consuming. The reason this approach remains desirable is that it avoids the practical disadvantages of derivative fitting. The objective function can be formulated as

$$G\left(\{\tilde{\mathbf{X}}(t_l)\}_{l=1}^{L}, \{\mathbf{X}(t_l)\}_{l=1}^{L}\right),$$

Here, the noise encountered is just that of the raw measurements, making it easier to account for. For this approach, we introduce the more accessible term *trajectory fitting*, referring to the inferred model being adjusted to reproduce the measured trajectory or trajectories for a given set of initial conditions. It can readily be applied to all semantics. However, for CTMC semantics, typically distributions over states are compared [112] or the likelihood of all replicates in D is maximized [147].

5.3.3 Hybrid. Both approaches can be combined in several ways, possibly exploiting the strengths of each. One way is to use the derivative matching approach for fast inference and then use trajectory fitting for model selection [6, 113], parameter optimization [17], or parameter refinement [54]. It is also possible to constrain trajectory fitting to producing the desired derviatives [101].

5.4 Inductive Bias: Dealing with the Non-Identifiability

A CRN is generally not uniquely identifiable from data [26, 37]. In system identification terms, the problem is ill-posed. However, additional information and assumptions can be used to distinguish similar models. Thus, the final building block(s) to choose are ways to introduce an *inductive bias*, a preference for one model over other equivalent or similar ones.

There are several angles from which this issue may be approached, which may also be combined: (1) The use of heuristics like sparsity and information criteria to introduce an implicit preference for certain (less complex) solutions; (2) The use of prior knowledge to constrain the search space and thus limit the possible solutions; (3) Providing a list of alternative models or a sense of uncertainty in estimation to inform an a-posteriori choice; (4) A closed-loop experimentation with an (automated) design of experiments (DoE) to optimally adjust the dataset. Note that, especially in (3) and (4), the bias is generally applied by a human after the inference. Each of these four approaches will be discussed in the following subsections. All of them have been discussed in the literature, for example, by Nobile et al. [93, Sec. V], who identified (2)-(4) in their future work statement.

5.4.1 Heuristics. Heuristics can be incorporated before (e.g., by setting a maximum number of reactions [58, 69, 80]) or during the inference (e.g., using a penalty term [40, 55]). When a selection of alternative models is available after the inference (Section 5.4.3), they can also be applied in the end to select a final candidate [6]. The following will outline some common heuristics.

Encouraging Sparsity. A common assumption about the mechanics underlying a system is that they comprise only some of the possible mechanisms. In the philosophy of science, this heuristic is called the principle of Occam's Razor [11]. For CRNs, simple means that a model comprises far fewer reactions than all possible reactions, a suitable assumption in many biological, chemical, and other contexts. Hence, parsimonious CRNs reproducing the time series data with the *least*

amount of reactions are sought. This has been investigated in CRN theory as well: A CRN can have different realizations (*dense* and *sparse*), which lead to the same ODEs [120] and, thus, behavior or data. Besides reducing the results to simple models, sparsity will also avoid overfitting (explaining measurement noise instead of the signal) to the supplied data by (artificially) limiting the degrees of freedom of the learning algorithm.

In different formulations of the search space, enforcing sparsity is more or less straightforward. For example, sparse regression (Section 5.2.1) cannot guarantee a maximum number of reactions in the continuous (non-MINLP) case, whereas template optimization (Section 5.2.2) does so by its very nature. However, from the perspective of computational theory, exactly enforcing sparsity is proven as NP-hard [91], even in the case of ℓ_1 and higher norms [49]. So enforcing sparsity constraints during inference, while narrowing the search space, also results in more difficult problems [68]. It may alternatively be supported by estimating the number of expected reactions from the data a priori, e.g., by singular decomposition/target factor analysis [8, 12, 17].

Information Criteria. While often a suitable assumption, sparsity is a heuristic based solely on model size. However, there is an important tradeoff between explanatory power and model size: Choosing a slightly larger model might be warranted if it provides a drastically better explanation of the data [6]. This is often achieved by introducing a weighted sum between the residuals and sparsity (e.g., using LASSO or Ridge methods [55]). The weight is a hyperparameter to be determined. Information criteria provide a rigorous way to quantify this trade-off, aiding automated decision-making. Some choices are the Akaike information criterion (AIC), Bayesian information criterion (BIC; also called Schwarz information criterion, SIC), and Hannan–Quinn information criterion (HQIC) [85]. They can also be used to select between alternatives produced by different strengths of regularization (e.g., with ℓ-norms as discussed in the previous subsection). Especially for noisy input data, future work should investigate the use of information criteria, which are currently underrepresented (Figure 8). Note, however, that their applicability depends on the context. For example, BIC is only theoretically justified when working with the *likelihood*. Hence, the popular mean-squared error may only be used under the assumption of normally-distributed residuals, where it is proportional to the likelihood.

5.4.2 Constraints and Prior Knowledge. Aside from sparsity heuristics, we can also employ arbitrary constraints based on prior knowledge to limit the search space. Examples include starting from a partial CRN that is to be completed [62, 69] or limiting reactions to a certain maximum stoichiometry (Section 5.5.2). In a Bayesian setting, prior knowledge is naturally incorporated through Bayes' law [43, 59]. The level at which such constraints may be applied is generally limited by the problem formulation [69, 84]. These levels include properties of the whole model [69], properties of single reactions [55] or stoichiometries [15, 143], or properties of the parameters [20]. The available prior knowledge also depends on the application domain, necessitating domain-specific approaches that specialize general CRN inference methods. Automated inclusion of prior knowledge closes the loop to the manual modeling process outlined at the beginning of Section 1. In fact, related work approaches the problem from a knowledge-first perspective (Section 8.3).

5.4.3 Uncertainty Estimation and Alternative Models. Instead of aiming for a single, unique model as the best explanation for the data, it is also possible to embrace the uncertainty over possible explanations and enumerate multiple candidates. It is then up to a separate algorithm or a human to determine which model to choose (which may, again, be based on the above criteria). For example, Tuza and Stan [126] demonstrates a framework to first infer a CRN using sparse regression and then enumerate all possible linearly conjugate networks that could also fit the data. In Foo et al. [40], a structural posterior is calculated. It is also possible to enumerate all realizations of a given

CRN regarding its ODE semantics using the concept of sparse and dense realizations from CRN theory [120]. In sparse regression, different values of the sparsity regularization strength also lead to alternatives [62]. Relatively few ($n = 21$) methods from our literature sample propose alternative models to cope with the non-identifiability issue. However, it is a very promising direction as, even with good heuristics and many constraints, the existing information is not enough to determine a definite explanation. Future work should thus combine good heuristics with alternative creation. Already, there is a trend of extending existing methods to include such uncertainty considerations, e.g., reactive SINDy [55], which has been extended by Jiang et al. [59] or the CRNN, which has been extended by Li et al. [73], Nieves et al. [92]. Finally, uncertainty quantification has the added benefit of reflecting uncertainty due to measurement noise.

5.4.4 Design of Experiments. If input data and bias are insufficient, it may be possible to discriminate alternative models with further experiments, by improving the input data in a systematic way [125, 137]. This leads to an automatic inference that is deeply integrated with (automatic) experimentation. However, this is not always practical, as conducting the necessary experiments on the real system and taking measurements can be slow and difficult. This is especially true for biological systems where the conditions of the "reactors" (cells) are difficult to control (Section 5.5.3).

5.5 Further Considerations and Practical Concerns

5.5.1 Handling Unmeasured Species. Many practical systems elude observability/measurability of all species. This is, for example, the case for intermediate products in chemistry [58, 71] or in-vitro experiments in cell biology [84]. Generally both formulations, derivative-matching and trajectory-fitting, can handle missing species trajectories by omitting them from $\mathbf{X}(t)$ or $\dot{\mathbf{X}}(t)$ in the loss calculation G, relying solely on the remaining error signal(s). This way, if enough effects are recorded in the data of the measured species and initial conditions are available for the unmeasured species (or assumed, e.g., as 0 for intermediates), an additional species may be postulated a priori and included in the model during inference. For example, in the template search, the dimensions of $\mathbf{N}^-, \mathbf{N}^+$ may be increased to include additional species [58]. In an exhaustive search, it is easy to accomodate this situation, and it may even be possible to "backcalculate" the missing species trajectory based on the enumerated candidates when given their initial concentrations [142]. If at least the stoichiometries of reactions are already available (e.g., from TFA, Section 5.1.4), this is possible as well [15]. The hybrid evolutionary approach can also easily accommodate partial observability [6]. On the other hand, the sparse regression approach requires all species to be measured to calculate $\Theta(\tilde{\mathbf{X}}(t))$. Missing measurements can still be handled by using simulation and trajectory fitting [59, 71]. Finally, when using derivative matching to determine rate constants for reactions involving missing species, these rate constants may be guessed randomly and iteratively refined based on simulation and trajectory fitting [54].

5.5.2 Common Assumptions. We observed that some common assumptions are currently made in order to focus on the core algorithms, but also with practical applications in mind. The former are to be incrementally relaxed in the future.

Probably the most common assumption (made in 72.5% of publications) is that reactions follow the well-established mass-action law kinetics [132]. Approaches considering other kinetic laws are rare. Especially in methods with envisioned application in chemistry, isothermality is often assumed. That is, the temperature inside the reaction volume stays the same. Only some publications (Section 4) specifically consider the Arrhenius kinetic law and include a temperature dependence. In biology, on the other hand, the isothermal assumption may be warranted for homiotherm organisms. Still, cell-biological CRNs can benefit from rational kinetics, such as Michaelis-Menten and Hill,

that aggregate multiple reaction steps into a single "overall" reaction, especially in the case of unmeasured intermediates and to increase human comprehension of models (Section 7).

Further, it is often assumed that the reactions are at most bimolecular, that is, they have a maximum of two reactants. This is motivated by applications in which reactions involving more than two reactants occur in separate steps. Physically, it can be claimed that the probability of a reactive collision between more than three species is rather low. Hence, when considering elementary reaction steps, this assumption is warranted. Similarly, some publications constrain not the complexes, but stoichiometric coefficients ($\mathbf{N}$) or constants ($\mathbf{N}^-, \mathbf{N}^+$) directly. Often, integer values are assumed for those, with some exceptions [124].

Many synthetic case studies also implement some imitation of measurement noise. In these cases, the vast majority of publications assume additive Gaussian noise (Section 4), which may not adequately reflect the kinds of noise arising in practice [62].

5.5.3 Differences Between Application Domains. Depending on the envisioned application of an inference approach, there are some differences. For example, as mentioned in the previous subsection, support for the Arrhenius law is important in many chemical applications to capture thermal variance, whereas in Biology, Michaelis-Menten kinetics are useful to abstract from reaction steps. As expected, controlled chemical reactions are easier to measure than biochemical reactors (e.g., cells), generally yielding datasets with more desirable properties in terms of data quantity and noise. Both domains struggle with the partial observability of (intermediate) species. In chemistry, models are typically sought for reactors like batch reactors, which are closed systems without inflow or outflow, and continuously stirred tank reactors (CSTR), in which the inflow and outflow are often assumed constant [71, 113, 136, 137][143, Fig. 1]. When considering a biological system like a cell, there is less control over the species actually perturbing the (sub-)system of interest. In biological applications, compartments play a large role in distinguishing between different levels, such as the cell, cytosol, and nucleus [69]. The only approaches in our literature sample that could directly support compartments are those that rely on the P Systems formalism [20, 107]. Due to the difficulty of measurement and observation, support for various types of noise plays a significant role in biologically motivated approaches [62]. Methods developed mainly from a mathematical or theoretical standpoint often assume near-ideal input data [25, 27, 28, 147].

6 A TAXONOMY OF CRN INFERENCE METHODS

Based on the building blocks identified in the previous Section 5, below we span the principal axes in decreasing order of importance and classify all methods in our literature sample accordingly. In comparison to the building blocks identified in Figure 7 we omitted preprocessing steps from our taxonomy as they are less informative about methods.

6.1 Dimension 1: Loss

The most discriminative feature distinguishes the space a method formulates its loss function in (cf. Section 5.3). This distinction has been made at least since 1970 [7, Sec. 1] and was also observed by other authors, e.g., recently Foo et al. [40]. We add a third, hybrid, option based on our findings.

Derivative Matching (Section 5.3.1)	Goodness of fit is based on the difference between model's derivatives and measurement-derived derivatives, e.g., Hoffmann et al. [55].
Trajectory Fitting (Section 5.3.2)	Goodness of fit is based on the difference between the model's state trajectories and the measured state trajectories, e.g., Jiang et al. [59].
Hybrid (Section 5.3.3)	Combines both, e.g., Banos et al. [6], Searson et al. [113].

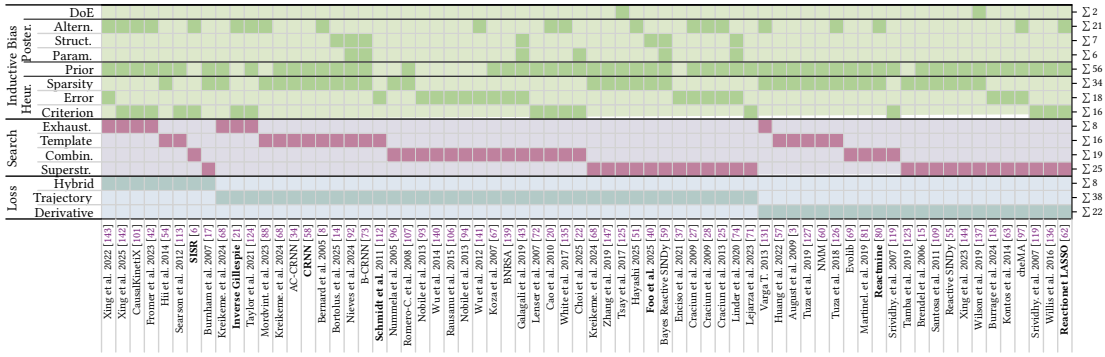


Fig. 8. Visual classification of the 68 methods into the dimensions: loss formulation, search space, and inductive bias. The latter dimension is further subdivided into finer categories, and multiple categories may be specified per method. Publications of special interest (Section 6.4) are bolded. Note that some of the 71 publications detail multiple methods according to our taxonomy, while others pertain to the same method (providing more details or additional benchmarks). These have either been split or unified keeping the latest.

6.2 Dimension 2: Search Space

The second dimension concerns how the structure of the search space is defined. Our findings here loosely align with how the related work is presented by Foo et al. [40], who distinguish sparse regularization, stoichiometric optimization, model search, and exhaustive enumeration. Values of this axis are ordered roughly according to the size of the considered search space, from smallest to largest. A fully exhaustive search, by definition, considers the largest space possible.

Superstructure (Section 5.2.1)	Select terms from a superstructure or library of basis reactions, e.g., Crampin et al. [29], Hoffmann et al. [55]. This includes sparse regression methods. The search is performed top down, systematically reducing a large structure.
Template (Section 5.2.2)	Directly determine stoichiometric coeff. $\mathbf{N}^-, \mathbf{N}^+$ and kinetic parameters $\mathbf{k}$ for a template CRN of fixed maximum size, e.g., Huang et al. [57], Ji and Deng [58]. The optimization happens on a lower level (coefficients vs. reactions).
Combinatorial (Section 5.2.3)	Freely explore the space of all CRNs by taking steps in the structure and/or parameter domain, e.g., using genetic programming [6] or Markov-Chain Monte Carlo methods [43]. Steps may have different granularities, like changes in species, reactions, or even model components [20]. CRN(s) can freely grow and shrink between steps, specifically including bottom-up approaches [29, 80].
Exhaustive (Section 5.2.4)	Enumerate and test all possible models within certain (domain-informed) bounds, e.g., Fromer et al. [42], Xing et al. [142]. The enumeration typically relies on combining all realizable reactions and kinetic laws for a given dataset.

Although the distinction is not perfect, we consider the classes defined here helpful for discussing current CRN inference approaches.

6.3 Dimension 3: Inductive Bias

This axis concerns addressing non-identifiability. To differentiate among all models that could fit the data, bias must be applied at some point in the inference process. The inductive biases on this

axis are ordered according to when in the process they are applied. Here, an approach may fall into more than one category, e.g., support prior knowledge and use the AIC:

Heuristics (Section 5.4.1)	Encouraging or enforcing sparsity in the number of reactions, information criteria soundly trading off sparsity with accuracy, and other statistical criteria. They can be applied automatically during (or also after) inference. Some methods also apply no heuristic and only rely on the error/likelihood.
Prior Knowledge (Section 5.4.2)	Incorporation of existing models, known reactions, and other constraints that are provided by domain experts or knowledge databases.
A Posteriori (Section 5.4.3)	The decision is (partially) deferred to *after* the inference. Approaches that estimate uncertainty in structure or parameters, e.g., by structural posterior [40] or by proposing similar alternative candidates [80].
DoE (Section 5.4.4)	Automatic design of experiments closes the loop by performing new (maximal-informative) real experiments based on the previous inference.

6.4 Recent Approaches and Spotlight

Finally, we make some recommendations for papers of special interest (in no particular order, bolded in Figure 8), also with respect to the challenges defined in the next section. Recently, Foo et al. [40] described how a posterior distribution over CRNs may be obtained and visualized. They further compare the effects of different ways to encourage sparsity. The recent publication by Banos et al. [6] employs the popular hybrid evolutionary approach and provides a strong evaluation and consideration of unobserved species trajectories. The tree search approach "Reactmine" by Martinelli et al. [80] is an interesting original combinatorial method. The method by Schmidt and Lipson [112] considers how distributions may be matched when inferring CRNs with the stochastic CTMC semantics in mind. For the latter, Chattopadhyay et al. [21] propose the very original "Inverse Gillespie" approach to infer CRNs with CTMC semantics from stochastic trajectories. Klimovskaia et al. [62] propose a method based on moment closures and consider binomial noise to model capture efficiency of cytometric measurements. Ji and Deng [58] developed the highly-cited CRNN approach, which has since been applied to practical problems as well.

7 OPEN CHALLENGES AND FUTURE DIRECTIONS IN CRN INFERENCE

After thoroughly analysing the extent and the currently well-covered areas of the design space for CRN inference algorithms, we highlight several selected challenges and promising gaps that future research can and should follow up on. Some challenges are shared with general dynamical system identification [84], but others are specific to CRNs.

7.1 Going beyond Sparsity

As outlined already in Section 2.5, identifiability is probably the major challenge for CRN inference and also any inductive modeling approach in general. Aside from prior knowledge, there is a range of methods that can be applied in CRN inference (cf. Section 5.4), the most common of which is the encouragement or enforcement of sparsity in the inferred models (Figure 8). It is also popular for mitigating overfitting and improving human comprehension of the inferred models. We argue that heuristics soundly trading off sparsity and accuracy, especially in combination with producing alternative possible models (explanations), will be expedient in future work. The main reason for this belief is that no simple "universal" inductive bias (e.g., sparsity) can perfectly distinguish all CRNs while performing well across every domain and data scenario. As Santosa and Weitz [109]

conclude, even very sparse systems are often not unique. An alternative to the posterior strategy is to strongly build on prior knowledge (cf. Section 5.4.2, Section 8.3). Already, prior knowledge is supported or sometimes required by the majority of methods (Figure 8). Still, systematically integrating a variety of knowledge types (qualitative, graphs, etc.) is another future research avenue.

7.2 Accounting for Measurement Noise

Identifiability is a frequently mentioned concern in publications on the topic (cf. Figure 4). Besides theoretical issues, practical concerns must be considered for methods to be applied under realistic conditions. The most prominent is probably the presence of noise in the data (mentioned in around 59% of publications). Noise is typically assumed to follow an additive zero-mean Gaussian with standard deviation σ: $\mathbf{X}(t)+\eta(t)$ with $\eta(t) \sim \mathcal{N}(0,\sigma)$. Measurement noise is a problem in derivative matching, as it interferes with the ability to determine faithful derivatives from the concentration data. To alleviate this problem, in addition to preprocessing (Section 5.1), weak formulations in sparse regression are an interesting approach that is underexplored in CRN inference [83]. Future work should also consider noise levels informed by practical applications beyond the normal assumption [142, p.11], e.g., binomial noise [62] and also experimental datasets. Some publications also consider intrinsic noise arising from low species abundances, e.g., by producing synthetic data with the SSA [55, 112, 147] or extrinsic noise resulting from differences between cell populations [25]. To our knowledge, so far no approach considers both.

7.3 Dealing with Partial Observability

Solutions to partial observability were discussed in around 31% of the publications. How easily this situation can be supported depends on the class of approach (Section 5.5.1). In particular, the very efficient sparse regression requires all species to be measured [55]. On the other hand, this situation is easily supported by a computationally intensive exhaustive search [142]. Thus, one challenge is to develop efficient inference methods that also account for unobserved species trajectories.

7.4 Kinetic Laws beyond Mass-Action

Most methods only consider mass-action kinetics, which can be applied in many scenarios. However, especially when moving towards practical applications, it is important to consider a wide range of kinetic laws depending on the context. For example, in chemistry, the addition of the Arrhenius law for temperature dependence is of special interest [58]. Some of these laws also allow inference of higher-level descriptions, in which multiple reaction steps are combined into a single reaction with specialized kinetics. This is important in cases where intermediate species are unmeasured and, instead of trying to infer them (Section 5.5.1), we want to construct an "overall" model [147, Rem. 3.6.1]. Such abstraction can also increase human comprehension and explanative power of the inferred CRN. In biology, these include Hill (ligand-macromolecule binding) and Michaelis-Menten (enzyme kinetics) laws [82]. The challenge in inferring such laws, and the reason why mass action is more popular (Section 4), is that they are often *rational functions* (fractions of two polynomials). As such, they take away beneficial properties of some problem formulations, such as the linear relation between the derivative and model structure (Section 5.2.1). Beyond well-established kinetics, advanced kinetic functions [137, Table 2] may benefit expressivity. To this end, some works already consider kinetic law inference for known reactions [9, 118] and may be applicable.

7.5 Considering Stochastic Semantics

Only very few methods explicitly consider the stochastic semantics of CRNs. This is partly to be expected, as simplicity, comparatively fast simulation, and analyzability make the deterministic ODE interpretation attractive. Also, it can rely on a large body of related work regarding the discovery

of dynamical (ODE) systems [84]. However, stochastic considerations are important when species counts are small. In this case, the system's intrinsic stochasticity plays a major role in its dynamics. Thus, an important goal is the development of methods considering the CTMC semantics (or approximations thereof) in their inference [74, Sec. 2.1][21]. Considering intrinstic noise may lead to different sets of parameters and reactions, i.e., CRN, and can distinguish systems indistinguishable under deterministic semantics [37]. Many approaches build on transitions being observed [22, 37, 147], but this requirement can be drastically relaxed in certain situations [21]. Further, for practical applications, approaches relying on simulation to obtain state trajectories for comparison [112], and approaches relying on moment closures [62] and linear noise approximations [14] seem attractive. We have classified the methods as reported in the publications (Section 4), but some may, with small adaptations, be applied to infer CTMCs as well, e.g., Wu et al. [141].

7.6 Considering the Multi-Scale Nature of CRNs

In many applications, reaction velocities span several orders of magnitude (cf. Figure 1). In order to account for this property, the data needs to be sampled according to Nyquist's theorem or even more frequently when accounting for measurement noise [71, p.8]. This property also makes the search space more difficult, as relatively small steps in rate constants or stoichiometric constants may be significant in some dimensions, but not others [51, 57, 69, 95][147, Remark 3.5]. Considering the ODE semantics, this further results in so-called "stiff" systems [71, p.6], which are difficult to integrate numerically. Solutions include reshaping the search space by a logarithmic transformation [51, 68, 95], rescaling [147], or obtaining a bootstrap estimate of the parameter's magnitudes [62]. This has been exploited for parameter inference but is not widely used in current CRN inference literature. We postulate that this is a result of the limited, typically small case studies currently used for testing, and that it will become more important as benchmarks become more challenging.

7.7 Search Space Analysis

For inference to succeed, the relation between steps in the search space and the corresponding steps toward the target behavior needs to be as direct as possible. A convex, continuous space is much easier to navigate than a nonconvex, combined discrete-continuous space. A major decision involved when designing an approach to CRN inference is thus how to formulate the loss. In Section 5.3, we discussed the advantages and disadvantages of fitting derivatives versus states. It has been argued by other authors that the search space induced by derivative matching is suited better, being, e.g., smoother [40, p.13], but measurement noise leads to problems. Filtering (Section 5.1.1) and, in sparse regression, weak formulations [83] may be used to alleviate this downside. Beyond this, the structural definition of the search space according to the second dimension of our taxonomy also greatly influences the above relation. Future work should further investigate the search spaces induced by the currently available methods, their influence on the problem solution, and the corresponding tradeoffs made with respect to the exact solution of the CRN inference problem.

7.8 Testing Generalization

In machine learning, it is common to split available data into training, testing, and sometimes validation sets to test a model's generalization capability. For CRNs, this means testing whether the inferred reactions can also explain the system's trajectories beyond those supplied to the inference process. While not an issue when a ground truth model is available for comparison, this kind of testing will become more important as we move to experimental data without a known ground truth. So far, there are only very few publications considering generalizability [6, 101].

7.9 Common Benchmarks and Availability of Implementations

It is evident from the data that methods are evaluated on only a very limited number of case studies (median 2; cf. Figure 6). Further, while there are some common benchmarking scenarios, there is not enough overlap between publications to make meaningful comparisons. Often, methods are not even directly compared. This is a major blocker of progress in the field, as, beyond very early efforts [133], no quantitative data exists to identify promising regions with respect to certain capabilities in the taxonomy Figure 8. Another major concern is that the source code is publicly available for only about a quarter of the methods. Future publications should take measures to make their artifacts *permanently* available by uploading them to an archive specialized in research artifacts (e.g., Zenodo[6]) or the publication's appendix. Only 7% of methods in our sample are made available in such a permanent way. While this paper spans the design space of possible methods and locates existing publications within it, a future, complementary paper will need to provide a quantitative comparison of current methods based on this qualitative survey. This can be based on current research on automatically generating realistic CRNs for benchmarking [145].

7.10 Evaluation Metrics

For such benchmarking, metrics play a major role. It is comparatively easy to evaluate the accuracy of a deterministic model with respect to D, as supported by many methods for comparing time series. Comparing stochastic outcomes [37] is more difficult and may also require comparing distributions that are only characterized by (few) samples [115]. Judging whether the learned structure is close to a ground truth structure is even more difficult. For benchmarking new and existing approaches, the consideration of which metrics to apply will hence be central.

7.11 Time Performance

Finally, it is of course also a concern to arrive at solutions as quickly as possible. However, one must consider that manual modeling for real-world problem instances may take tens to hundreds of days. So even an algorithm that requires hours or days to converge can achieve real practical utility, e.g., as demonstrated in the collaboration with AstraZeneca by Taylor et al. [124]. Still, a time-consuming, exhaustive search approach might only be the optimal choice for very specific, smaller problem instances.

It is clear that the time performance of algorithms is not currently a focus of the field (it is only mentioned, mostly briefly, in around one-third of publications). We postulate that it will become increasingly important as successes in addressing the other challenges increase. Generally, sparse regression can be very fast due to the limited, comparatively well-behaved search space and the linearity of the problem, but a trade-off must be made between convergence speed and fidelity to the original problem in each situation.

8 RELATED WORK

To complete the picture, we summarize how other fields relate to and can benefit the solution of the CRN inference problem.

8.1 Other Reviews

We start with some adjacent reviews. Our structured search (Section 3) did not yield any recent reviews on learning CRNs from time-series data. The related reviews identified in our search are either older than 20 years [30] or discuss an adjacent topic or subset of our presentation [41, 84].

[6] https://zenodo.org/

In the context of continuous ODE semantics, CRN inference can be viewed as a special, constrained case of (dynamical) system identification. Most notably, a recent review (preprint) discusses the application of ODE model identification to learn biological models from time-series data [84]. It uncovers many important challenges across both biological and methodological aspects, which align with our findings. Although the authors only briefly discuss CRN inference, they emphasize it as the most important form of "structural prior" in biological digital twin discovery. Another review from 2008 discusses several approaches to general equation discovery in systems biology [35]. An early review also briefly discussed a sparse regression approach to the inference of CRN-constrained ODE models [30, 5.2.2]. We here focus on an up-to-date overview of the constrained (coupled) case, also beyond superstructures and the continuous CRN semantics. Marku and Pancaldi [79] presents a recent review on GRN inference. Generally, CRN inference methods may be employed for GRN inference as well, but not the other way around (Section 2.4).

In their book chapter, Fröhlich et al. [41] primarily outline parameter estimation methods and also discuss scalable inference for (non-coupled) ODE models in biochemistry. They present several angles to approach it: by forward-selection, backward-elimination, and sparse regularization. The former two are similar to the simple-to-general and general-to-specific approaches, respectively, discussed by Crampin et al. [29], Srividhya et al. [119], and also Martinelli et al. [82].

Puliyanda et al. [104] views the topic of the present work from a chemical standpoint as "Automated data-driven approaches for reaction or kinetics monitoring in real time" [104]. They also name a few of the methods presented here, such as reactive SINDy [55] and the CRNN [58], but focus on the chemical perspective and knowledge-driven inference. The inference of CRN from time-series data is discussed only briefly as a secondary focus. Also, there have been many developments since 2022 (cf. Figure 4) [13, 40, 80]. Similarly, Wen et al. [134] discusse system identification for CRNs but focus on data-driven phenomenological models that assume a known model structure, such as physics-informed neural networks. Here, we provided a much broader perspective.

An early review from 2004 discusses all aspects of kinetic model identification in chemistry and biochemistry at the time, but its main focus is on data-analysis and manual construction steps rather than automated algorithms [78]. The review by Crampin et al. [30] mentioned earlier summarizes the field's state in 2004 as well. Of particular interest is Section 5 therein. It proposes a general framework for ODE inference (based on polynomials or also reactions) very similar to SINDy [16].

8.2 CRN Theory

CRN theory is concerned with the relation between qualitative (structural) CRN properties and the quantitative behavior of the induced dynamic system, commonly assuming the mass-action law [39, 56]. It studies these questions *without* considering kinetics, i.e., values for the rate constants in mass-action networks. One of the main results is the "deficiency theorems", which make conclusions about which CRN structures may or may not admit certain equilibria in their simulation. Hence, besides providing the theoretical foundation for CRN, results from this field are highly relevant for solving the CRN inference problem [37, Sec. 5]. Nevertheless, only a few of the identified publications consider these theoretical results explicitly [3, 21, 25, 27, 28, 37, 74, 123, 126]. For example, they may be used to constrain the search space by excluding structures known to be unable to produce certain time series data (e.g., that include multiple steady states) or to enumerate alternative realizations that could produce the same time-series data [123, 126]. There are already some works in this mathematical field that pertain to the inference of CRN from data, e.g., the works by Craciun [25, 27, 30]. However, they often consider more exact solutions with the tradeoff that they are only applicable in certain cases where specific kinds of data are available (i.e., measurements of all CTMC transitions [22, 147]). It would be interesting for future work to merge more of the results from CRN theory with more practically driven methods for CRN inference.

8.3 CRN Inference from Knowledge

It is clear from the previous sections that even in a data-driven method, prior knowledge is vital to its success when identifiability issues are at play (see also Santosa and Weitz [109]). In fact, we can tackle the problem not only from the angle of analyzing new data, but from the angle of combining our existing knowledge in novel ways. Such methods typically rely on databases of known substances, reactions, and their properties, but also on automated literature analysis. In chemistry, a large body of work concerns the construction of networks from known reactions, e.g., to discover how to synthesize a certain product (retrosynthesis) [104]. In biological applications, several methods try to construct models that satisfy certain conditions from known parts [1, 19].

8.4 CRN Inference from Natural Language

Knowledge is often only available in non-formal, written form. To make use of it, also for the methods discussed in the previous section, we need to extract information into more formal representations. To this end, so-called *reading engines* have been developed [5]. They can be used to extract formal representations, such as reactions, from, e.g., research papers. The results of this extraction can be refined by existing formal knowledge, like from reaction databases, a process often referred to as *grounding*. CRN inference from natural language is also briefly discussed by Wen et al. [134]. More recently, large language models (LLMs) revolutionized the field of natural language processing and have since been used to extract CRN models from textual descriptions [2, 70].

8.5 Generating CRNs with desired behavior

A closely related problem to CRN inference from time-series data is the construction of a CRN with a specific behavior. This goes beyond fitting measured time series data from chemically or biologically plausible CRN to obtaining CRN that implement, e.g., mathematical functions [72], or the mapping from bits to seven-segment display activations [88]. One motivation for this work is to synthesize or "program" cells or chemical reactors to execute specific functions, or even perform computation and learning [52, 89, 90]. If they allow setting transient behavior as target, these methods can of course be applied to the problem defined in the present publication [72, 88]. However, especially for computation, the target is often a certain *steady* state [52].

8.6 System Identification

System identification pursues a very similar and more general goal of obtaining a mathematical formulation describing the input-output relation of a dynamical system [103]. Depending on the nature of the available data and information, white-box, grey-box, and black-box tasks are distinguished, with the first assuming full knowledge of the model structure and the last assuming no prior structural knowledge. As opposed to the inference task discussed here, *classical system identification* works by selecting between a set of possible structures, e.g., using AIC or BIC [85], by fitting their parameters. This can be regarded as a more elaborate *model selection* rather than direct inference of structure as proposed here. Hence, only in the case of the sparse regression approach to CRN inference can a close connection be made. More recently, the term system identification has been used in a broader sense to encompass a wide range of methods for identifying dynamical systems [16]. To this end, Pillonetto et al. [103, Fig.1 (right)] present a more modern formulation where a large and possibly infinite space of models is searched. Central in this modern view is the identification of non-linear dynamical systems [10]. A relatively recent large impact in this field was made by Brunton et al. [16] with the SINDy framework. Similar approaches have, however, been discussed already much earlier, also in the context of CRNs [29, 30].

8.7 Network Inference

In the problem of CRN inference, obtaining a description of the network without kinetics can be regarded as a preliminary step. Network inference aims to construct a graph structure from (time series) data. It has mainly been used to infer GRNs in modeling genetic networks based on positive and negative influences between species [79]. For CRN inference, also hypergraph or bipartite structures need to be accounted for (Section 2.1).

8.8 Symbolic Regression

We conclude with a field that subsumes most reverse-engineering efforts in science. In symbolic regression, the task is to choose one expression from the space of all possible symbolic expressions (generally, mathematical equations) to fit input-output examples [33]. The earliest approaches here also relied on genetic programming [65], which is still explored to date, also for CRN inference [6, 93]. Thus, CRN inference can be regarded as a subproblem of symbolic regression, where input-output examples are considered by derivative matching or can be generated via simulation.

9 CONCLUSION

This survey presents an overview and taxonomy of the current state of the art in inferring CRNs from time series data on the concentrations or abundances of the involved species. Research on this topic is scattered among different disciplines, including chemistry, biology, computer science, and applied mathematics. We unify the terminology and provide an accessible starting point for understanding the landscape by briefly introducing CRNs and the problem of learning them from time series data. We quantify the current trends, such as consideration of different CRN semantics, using prior knowledge to constrain the search space, kinetic laws, and uncertainty quantification in the inferred CRN to cope with indistinguishability. Here, we already see that the vast majority support or sometimes require additional knowledge, and that deterministic semantics and mass-action kinetics prevail. To complete the picture, we also provide insights into evaluation practices and reproducibility. Typically, only a few selected benchmarks based on synthetic or realistic ground truth models are used. Rarely, experimental data is considered. Only a minority of publications make their source code available, even fewer in a permanent manner.

Subsequently, we identify four fundamental categories of building blocks that make up existing methods, pertaining to data preprocessing, how the search is conducted, how the loss is formulated, and the kinds of inductive bias applied to cope with unidentifiability. For each category and its building blocks, we present a brief introduction and discuss its use in the literature. Also, we characterize the tradeoffs made by choosing among them in terms of inference speed, fidelity to the original problem, and ease of search. This leads to the proposal of a three-dimensional taxonomy consisting of the dimensions loss space, search, and inductive bias. We classify each discovered method in this design space and give suggestions for (recent) publications of high interest.

Finally, we offer guidance for future research by highlighting selected important and unsolved challenges. Some are specific to CRN, like consideration of their multiscale nature and support for stochastic semantics; others are shared with system identification in general, such as measurement noise and partial observability. To this end, we also summarize the relation to other fields.

We hope that, by providing an accessible entry point and a novel bird's-eye perspective, our synthesis of the state of the art will enable readers to effectively select and pursue promising avenues for research in the increasingly important field of learning CRNs from time-series data.

ACKNOWLEDGMENTS

We acknowledge the previous work of van der Wall [128] and Skrzypczak [117], whose bachelor's and, respectively, master's theses provided a starting point for the research presented here. We thank Kevin Burrage, Fiete Haack, and Glenn Skrzypczak for their helpful comments on the first draft.

REFERENCES

[1] Yasmine Ahmed, Cheryl A. Telmer, Gaoxiang Zhou, and Natasa Miskov-Zivanov. 2024. Context-Aware Knowledge Selection and Reliable Model Recommendation with ACCORDION. *Frontiers in Systems Biology* 4 (April 2024), 1308292. https://doi.org/10.3389/fsysb.2024.1308292

[2] Niloofar Arazkhani, Haomiao Luo, Difei Tang, Brent Cochran, and Natasa Miskov-Zivanov. 2025. GBM model refinement with literature curation, rule-based NLP, and LLMs. bioRxiv. https://doi.org/10.1101/2025.03.27.645730

[3] Elias August and Antonis Papachristodoulou. 2009. Efficient, Sparse Biological Network Determination. *BMC Syst. Biol.* 3 (2009), 0–13. https://doi.org/10.1186/1752-0509-3-25

[4] Florin Avram, Rim Adenane, and Mircea Neagu. 2024. Advancing mathematical epidemiology and chemical reaction network theory via synergies between them. *Entropy* 26, 11 (2024), 936. https://doi.org/10.3390/e26110936

[5] John A. Bachman, Benjamin M. Gyori, and Peter K. Sorger. 2023. Automated Assembly of Molecular Mechanisms at Scale from Text Mining and Curated Databases. *Molecular Systems Biology* 19, 5 (May 2023), e11325. https://doi.org/10.15252/msb.202211325

[6] Manuel Palma Banos, Joel D. Kress, Rigoberto Hernandez, and Galen T. Craven. 2025. Stoichiometrically-Informed Symbolic Regression for Extracting Chemical Reaction Mechanisms from Data. arXiv:2510.20655

[7] André Bardow and Wolfgang Marquardt. 2004. Incremental and Simultaneous Identification of Reaction Kinetics: Methods and Comparison. *Chem. Eng. Sci.* 59, 13 (July 2004), 2673–2684. https://doi.org/10.1016/j.ces.2004.03.023

[8] Olivier Bernard and Georges Bastin. 2005. Identification of Reaction Networks for Bioprocesses: Determination of a Partially Unknown Pseudo-Stoichiometric Matrix. *Bioprocess Biosyst. Eng.* 27, 5 (Aug. 2005), 293–301. https://doi.org/10.1007/s00449-005-0407-3

[9] Nirav Bhatt, Nimet Kerimoglu, Michael Amrhein, Wolfgang Marquardt, and Dominique Bonvin. 2012. Incremental Identification of Reaction Systems—a Comparison between Rate-Based and Extent-Based Approaches. *Chem. Eng. Sci.* 83 (Dec. 2012), 24–38. https://doi.org/10.1016/j.ces.2012.05.040

[10] Amirmohammad Ziaei Bideh, Aleksandra Georgievska, and Jonathan Gryak. 2026. MDBench: Benchmarking Data-Driven Methods for Model Discovery. *Proceedings of the AAAI Conference on Artificial Intelligence* 40, 24 (Mar. 2026), 19746–19754. https://doi.org/10.1609/aaai.v40i24.39056

[11] Anselm Blumer, Andrzej Ehrenfeucht, David Haussler, and Manfred K. Warmuth. 1987. Occam's Razor. *Inform. Process. Lett.* 24, 6 (April 1987), 377–380. https://doi.org/10.1016/0020-0190(87)90114-1

[12] D. Bonvin and D.W.T. Rippin. 1990. Target Factor Analysis for the Identification of Stoichiometric Models. *Chem. Eng. Sci.* 45, 12 (1990), 3417–3426. https://doi.org/10.1016/0009-2509(90)87147-K

[13] Luca Bortolussi, Francesca Cairoli, Julia Klein, and Tatjana Petrov. 2023. Data-Driven Inference of Chemical Reaction Networks via Graph-Based Variational Autoencoders. In *Quantitative Evaluation of Systems*, Nils Jansen and Mirco Tribastone (Eds.). Vol. 14287. Springer Nature Switzerland, Cham, 143–147. https://doi.org/10.1007/978-3-031-43835-6_10

[14] Luca Bortolussi, Francesca Cairoli, Julia Klein, and Tatjana Petrov. 2025. Neuro-Symbolic Discovery of Markov Population Processes. In *Proceedings of the International Conference on Neuro-symbolic Systems*. PMLR, Philadelphia, PA, USA, 396–408. https://proceedings.mlr.press/v288/bortolussi25a.html

[15] Marc Brendel, Dominique Bonvin, and Wolfgang Marquardt. 2006. Incremental Identification of Kinetic Models for Homogeneous Reaction Systems. *Chem. Eng. Sci.* 61, 16 (Aug. 2006), 5404–5420. https://doi.org/10.1016/j.ces.2006.04.028

[16] Steven L. Brunton, Joshua L. Proctor, and J. Nathan Kutz. 2016. Discovering Governing Equations from Data by Sparse Identification of Nonlinear Dynamical Systems. *Proceedings of the National Academy of Sciences* 113, 15 (April 2016), 3932–3937. https://doi.org/10.1073/pnas.1517384113

[17] S.C. Burnham, M.J. Willis, and A.R. Wright. 2007. Identifying Chemical Reaction Network Models. *IFAC Proc. Vol.* 40, 5 (2007), 225–230. https://doi.org/10.3182/20070606-3-MX-2915.00085

[18] Pamela M. Burrage, Hasitha N. Weerasinghe, and Kevin Burrage. 2024. Using a Library of Chemical Reactions to Fit Systems of Ordinary Differential Equations to Agent-Based Models: A Machine Learning Approach. *Numerical Algorithms* 96, 3 (July 2024), 1063–1077. https://doi.org/10.1007/s11075-023-01737-0

[19] Adam A. Butchy, Niloofar Arazkhani, Cheryl A. Telmer, and Natasa Miskov-Zivanov. 2025. Automating Knowledge-Driven Model Recommendation: Methodology, Evaluation, and Key Challenges. *IEEE Transactions on Computational Biology and Bioinformatics* 22, 2 (2025), 493–502. https://doi.org/10.1109/TCBBIO.2024.3507781

[20] Hongqing Cao, Francisco J. Romero-Campero, Stephan Heeb, Miguel Cámara, and Natalio Krasnogor. 2010. Evolving Cell Models for Systems and Synthetic Biology. *Syst. Synth. Biol.* 4, 1 (March 2010), 55–84. https://doi.org/10.1007/s11693-009-9050-7

[21] Ishanu Chattopadhyay, Anna Kuchina, Gürol M. Süel, and Hod Lipson. 2013. Inverse Gillespie for Inferring Stochastic Reaction Mechanisms from Intermittent Samples. *Proc. Natl. Acad. Sci.* 110, 32 (Aug. 2013), 12991–12995. https://doi.org/10.1073/pnas.1214559110

[22] Boseung Choi, Hey-Won Kang, and Grzegorz A. Rempala. 2025. Inference for Stochastic Reaction Networks via Logistic Regression. arXiv:2507.19979

[23] Scott D Cohen, Alan C Hindmarsh, and Paul F Dubois. 1996. CVODE, a stiff/nonstiff ODE solver in C. *Computers in physics* 10, 2 (1996), 138–143. https://doi.org/10.1063/1.4822377

[24] Alison Cozad, Nikolaos V. Sahinidis, and David C. Miller. 2014. Learning surrogate models for simulation-based optimization. *AIChE Journal* 60, 6 (2014), 2211–2227. https://doi.org/10.1002/aic.14418

[25] Gheorghe Craciun, Jaejik Kim, Casian Pantea, and Grzegorz A. Rempala. 2013. Statistical Model for Biochemical Network Inference. *Commun. Stat.- Simul. Comput.* 42, 1 (Jan. 2013), 121–137. https://doi.org/10.1080/03610918.2011.633200

[26] Gheorghe Craciun and Casian Pantea. 2008. Identifiability of Chemical Reaction Networks. *Journal of Mathematical Chemistry* 44, 1 (July 2008), 244–259. https://doi.org/10.1007/s10910-007-9307-x

[27] Gheorghe Craciun, Casian Pantea, and Grzegorz A. Rempala. 2009. Algebraic Methods for Inferring Biochemical Networks: A Maximum Likelihood Approach. *Comput. Biol. Chem.* 33, 5 (Oct. 2009), 361–367. https://doi.org/10.1016/j.compbiolchem.2009.07.014

[28] Gheorghe Craciun, Casian Pantea, and Grzegorz A. Rempala. 2009. A Dimension Reduction Method for Inferring Biochemical Networks. arXiv:0902.4417

[29] Edmund J. Crampin, Patrick E. McSharry, and Santiago Schnell. 2004. Extracting Biochemical Reaction Kinetics from Time Series Data. In *Knowledge-Based Intelligent Information and Engineering Systems*. Vol. 3214. Springer Berlin Heidelberg, Berlin, Heidelberg, 329–336. https://doi.org/10.1007/978-3-540-30133-2_42

[30] E. J. Crampin, S. Schnell, and P. E. McSharry. 2004. Mathematical and Computational Techniques to Deduce Complex Biochemical Reaction Mechanisms. *Prog. Biophys. Mol. Biol.* 86, 1 (Sept. 2004), 77–112. https://doi.org/10.1016/j.pbiomolbio.2004.04.002

[31] Kyle Cranmer, Johann Brehmer, and Gilles Louppe. 2020. The Frontier of Simulation-Based Inference. *Proceedings of the National Academy of Sciences* 117, 48 (Dec. 2020), 30055–30062. https://doi.org/10.1073/pnas.1912789117

[32] Carl De Boor and Carl De Boor. 1978. *A practical guide to splines*. Vol. 27. Springer, New York.

[33] Junlan Dong and Jinghui Zhong. 2025. Recent Advances in Symbolic Regression. *ACM Comput. Surv.* 57, 11 (June 2025), 290:1–290:37. https://doi.org/10.1145/3735634

[34] Felix A. Döppel and Martin Votsmeier. 2024. Robust Mechanism Discovery with Atom Conserving Chemical Reaction Neural Networks. *Proc. Combust. Inst.* 40, 1-4 (2024), 105507. https://doi.org/10.1016/j.proci.2024.105507

[35] Sašo Džeroski and Ljupčo Todorovski. 2008. Equation Discovery for Systems Biology: Finding the Structure and Dynamics of Biological Networks from Time Course Data. *Curr. Opin. Biotechnol.* 19, 4 (Aug. 2008), 360–368. https://doi.org/10.1016/j.copbio.2008.07.002

[36] Elsevier B.V. 2026. Scopus. https://www.scopus.com Last Accessed: 2026-01-29.

[37] German Enciso, Radek Erban, and Jinsu Kim. 2021. Identifiability of Stochastically Modelled Reaction Networks. *Eur. J. Appl. Math.* 32, 5 (Oct. 2021), 865–887. https://doi.org/10.1017/S0956792520000492

[38] James R Faeder, Michael L Blinov, and William S Hlavacek. 2009. Rule-based modeling of biochemical systems with BioNetGen. In *Systems biology*. Springer, Cham, 113–167. https://doi.org/10.1007/978-1-59745-525-1_5

[39] Martin Feinberg. 1980. Lectures on Chemical Reaction Networks - M. Feinberg, University of Wisconsin, 1979. (1980).

[40] Yong See Foo, Adriana Zanca, Jennifer A. Flegg, and Ivo Siekmann. 2025. Quantifying Structural Uncertainty in Chemical Reaction Network Inference. arXiv:2505.15653

[41] Fabian Fröhlich, Carolin Loos, and Jan Hasenauer. 2018. Scalable Inference of Ordinary Differential Equation Models of Biochemical Processes. In *Methods in Molecular Biology*. Springer Nature, Cham, 385–422. https://doi.org/10.1007/978-1-4939-8882-2_16

[42] Jenna Fromer, Christos Georgakis, and Jason Mustakis. 2023. Toward the Identification of Stoichiometric Models for Complex Reaction Mixtures. *Ind. Eng. Chem. Res.* 62, 5 (Feb. 2023), 2225–2239. https://doi.org/10.1021/acs.iecr.2c01231

[43] Nikhil Galagali and Youssef M. Marzouk. 2019. Exploiting Network Topology for Large-Scale Inference of Nonlinear Reaction Models. *J. R. Soc. Interface* 16, 152 (March 2019), 20180766. https://doi.org/10.1098/rsif.2018.0766

[44] Connie W. Gao, Joshua W. Allen, William H. Green, and Richard H. West. 2016. Reaction Mechanism Generator: Automatic construction of chemical kinetic mechanisms. *Computer Physics Communications* 203 (2016), 212–225. https://doi.org/10.1016/j.cpc.2016.02.013

[45] Daniel T Gillespie. 1976. A General Method for Numerically Simulating the Stochastic Time Evolution of Coupled Chemical Reactions. *J. Comput. Phys.* 22, 4 (Dec. 1976), 403–434. https://doi.org/10.1016/0021-9991(76)90041-3

[46] Jorge Goncalves and Sean Warnick. 2008. Necessary and Sufficient Conditions for Dynamical Structure Reconstruction of LTI Networks. *IEEE Trans. Automat. Control* 53, 7 (Aug. 2008), 1670–1674. https://doi.org/10.1109/TAC.2008.928114

[47] Google, Inc. 2026. Google Gemini. https://gemini.google.com Last Accessed: 2026-01-29.

[48] Google, Inc. 2026. Google Scholar. https://scholar.google.com/ Last Accessed: 2026-01-29.

[49] Aparna Gupte and Vinod Vaikuntanathan. 2021. The Fine-Grained Hardness of Sparse Linear Regression. arXiv:2106.03131

[50] Hugo Hadfield. 2024. *Kalmangrad*. GitHub. https://github.com/hugohadfield/kalmangrad Last Accessed 2026-02-20.

[51] Shun Hayashi. 2025. Sparse Identification of Chemical Reaction Mechanisms from Limited Concentration Profiles. *Digital Discovery* 4, 11 (2025), 3092–3097. https://doi.org/10.1039/D5DD00293A

[52] Mathieu Hemery and François Fages. 2022. Algebraic Biochemistry: A Framework for Analog Online Computation in Cells. In *Computational Methods in Systems Biology*, Ion Petre and Andrei Păun (Eds.). Springer International Publishing, Cham, 3–20. https://doi.org/10.1007/978-3-031-15034-0_1

[53] Thomas Henzinger, Barbara Jobstmann, and Verena Wolf. 2011. Formalisms for Specifying Markovian Population Models. *Int. J. Found. Comput. Sci.* 22, 4 (June 2011), 823–841. https://doi.org/10.1142/S0129054111008441

[54] Charles J. K. Hii, Allen R. Wright, and Mark J. Willis. 2014. Utilizing a Genetic Algorithm to Elucidate Chemical Reaction Networks: An Experimental Case Study. *Int. J. Chem. Eng. Appl.* 5, 6 (Dec. 2014), 516–520. https://doi.org/10.7763/IJCEA.2014.V5.439

[55] Moritz Hoffmann, Christoph Fröhner, and Frank Noé. 2019. Reactive SINDy: Discovering Governing Reactions from Concentration Data. *J. Chem. Phys.* 150, 2 (Jan. 2019), 25101. https://doi.org/10.1063/1.5066099

[56] F. Horn and R. Jackson. 1972. General Mass Action Kinetics. *Archive for Rational Mechanics and Analysis* 47, 2 (Jan. 1972), 81–116. https://doi.org/10.1007/BF00251225

[57] Juntao Huang, Yizhou Zhou, and Wen-An Yong. 2022. Data-Driven Discovery of Multiscale Chemical Reactions Governed by the Law of Mass Action. *J. Comput. Phys.* 448 (Jan. 2022), 110743. https://doi.org/10.1016/j.jcp.2021.110743

[58] Weiqi Ji and Sili Deng. 2021. Autonomous Discovery of Unknown Reaction Pathways from Data by Chemical Reaction Neural Network. *J. Phys. Chem. A* 125, 4 (Feb. 2021), 1082–1092. https://doi.org/10.1021/acs.jpca.0c09316

[59] Richard Jiang, Prashant Singh, Fredrik Wrede, Andreas Hellander, and Linda Petzold. 2022. Identification of Dynamic Mass-Action Biochemical Reaction Networks Using Sparse Bayesian Methods. *PLOS Comput. Biol.* 18, 1 (Jan. 2022), e1009830. https://doi.org/10.1371/journal.pcbi.1009830

[60] Alexey V. Karnaukhov, Elena V. Karnaukhova, and James R. Williamson. 2007. Numerical Matrices Method for Nonlinear System Identification and Description of Dynamics of Biochemical Reaction Networks. *Biophys. J.* 92, 10 (May 2007), 3459–3473. https://doi.org/10.1529/biophysj.106.093344

[61] Sarah M Keating, Dagmar Waltemath, Matthias König, Fengkai Zhang, Andreas Dräger, Claudine Chaouiya, Frank T Bergmann, Andrew Finney, Colin S Gillespie, Tomáš Helikar, et al. 2020. SBML Level 3: an extensible format for the exchange and reuse of biological models. *Molecular systems biology* 16, 8 (2020), e9110. https://doi.org/10.15252/msb.20199110

[62] Anna Klimovskaia, Stefan Ganscha, and Manfred Claassen. 2016. Sparse Regression Based Structure Learning of Stochastic Reaction Networks from Single Cell Snapshot Time Series. *PLOS Comput. Biol.* 12, 12 (Dec. 2016), e1005234. https://doi.org/10.1371/journal.pcbi.1005234

[63] János Kontos, Lászó Richárd Tóth, and Tamás Varga. 2014. Development of a Reaction Structure Identification Algorithm. *Hung. J. Ind. Chem.* 42, 1 (Aug. 2014), 51–56. https://doi.org/10.1515/363

[64] J. Koza, William Mydlowec, Guido Lanza, Jessen Yu, and M. A. Keane. 2001. Automatic Synthesis of Both the Topology and Sizing of Metabolic Pathways Using Genetic Programming. In *GECCO'01: Proceedings of the 3rd Annual Conference on Genetic and Evolutionary Computation*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 57–65. https://gpbib.cs.ucl.ac.uk/gecco2001/koza_gecco2001.pdf

[65] John R. Koza. 1993. *Genetic Programming: On the programming of computers by means of natural selection* (3rd print ed.). MIT Press, Cambridge, MA, USA; London, UK.

[66] John R. Koza, William Mydlowec, Guido Lanza, Jessen Yu, and Martin A. Keane. 2000. Reverse Engineering of Metabolic Pathways from Observed Data Using Genetic Programming. In *Proceedings of the Pacific Symposium*. WORLD SCIENTIFIC, Mauna Lani, Hawaii, 434–445. https://doi.org/10.1142/9789814447362_0043

[67] John R. Koza, William Mydlowec, Guido Lanza, Jessen Yu, and Martin A. Keane. 2007. Automatic Computational Discovery of Chemical Reaction Networks Using Genetic Programming. In *Computational Discovery of Scientific Knowledge*, Sašo Džeroski and Ljupčo Todorovski (Eds.). Vol. 4660. Springer Berlin Heidelberg, Berlin, Heidelberg,

205–227. https://doi.org/10.1007/978-3-540-73920-3_10

[68] Justin Noah Kreikemeyer, Philipp Andelfinger, and Adelinde M. Uhrmacher. 2024. Towards Learning Stochastic Population Models by Gradient Descent. In *Proceedings of the 38th ACM SIGSIM Conference on Principles of Advanced Discrete Simulation*. ACM, Atlanta GA USA, 88–92. https://doi.org/10.1145/3615979.3656058

[69] Justin N. Kreikemeyer, Kevin Burrage, and Adelinde M. Uhrmacher. 2024. Discovering Biochemical Reaction Models by Evolving Libraries. In *Computational Methods in Systems Biology*, Roberta Gori, Paolo Milazzo, and Mirco Tribastone (Eds.). Vol. 14971. Springer Nature Switzerland, Cham, 117–136. https://doi.org/10.1007/978-3-031-71671-3_10

[70] Justin Noah Kreikemeyer, Miłosz Jankowski, Pia Wilsdorf, and Adelinde M. Uhrmacher. 2025. Using (Not-so) Large Language Models to Generate Simulation Models in a Formal DSL: A Study on Reaction Networks. *ACM Trans. Model. Comput. Simul.* 35, 4, Article 31 (Sept. 2025), 27 pages. https://doi.org/10.1145/3733719

[71] Fernando Lejarza, Elsa Koninckx, Linda J. Broadbelt, and Michael Baldea. 2023. A Dynamic Nonlinear Optimization Framework for Learning Data-Driven Reduced-Order Microkinetic Models. *Chem. Eng. J.* 462 (April 2023), 142089. https://doi.org/10.1016/j.cej.2023.142089

[72] Thorsten Lenser, Thomas Hinze, Bashar Ibrahim, and Peter Dittrich. 2007. Towards Evolutionary Network Reconstruction Tools for Systems Biology. In *Evolutionary Computation,Machine Learning and Data Mining in Bioinformatics*, Elena Marchiori, Jason H. Moore, and Jagath C. Rajapakse (Eds.). Vol. 4447. Springer Berlin Heidelberg, Berlin, Heidelberg, 132–142. https://doi.org/10.1007/978-3-540-71783-6_13

[73] Qiaofeng Li, Huaibo Chen, Benjamin C. Koenig, and Sili Deng. 2023. Bayesian Chemical Reaction Neural Network for Autonomous Kinetic Uncertainty Quantification. *Phys. Chem. Chem. Phys.* 25, 5 (2023), 3707–3717. https://doi.org/10.1039/D2CP05083H

[74] Daniel F. Linder and Grzegorz A. Rempała. 2020. Synthetic Likelihood Method for Reaction Network Inference. *Math. Methods Appl. Sci.* 43, 18 (Dec. 2020), 10547–10568. https://doi.org/10.1002/mma.6631

[75] Pavel Loskot, Komlan Atitey, and Lyudmila Mihaylova. 2019. Comprehensive Review of Models and Methods for Inferences in Bio-Chemical Reaction Networks. *Front. Genet.* 10, 549 (June 2019). https://doi.org/10.3389/fgene.2019.00549

[76] Benn Macdonald, Mu Niu, Simon Rogers, Maurizio Filippone, and Dirk Husmeier. 2016. Approximate parameter inference in systems biology using gradient matching: a comparative evaluation. *Biomedical Engineering Online* 15, Suppl 1 (2016), 80. https://doi.org/10.1186/s12938-016-0186-x

[77] Daniel Machado, Rafael S Costa, Miguel Rocha, Eugénio C Ferreira, Bruce Tidor, and Isabel Rocha. 2011. Modeling Formalisms in Systems Biology. *AMB Express* 1, 1 (2011), 45. https://doi.org/10.1186/2191-0855-1-45

[78] G Maria. 2004. A Review of Algorithms and Trends in Kinetic Model Identification for Chemical and Biochemical Systems. *Chem. Biochem. Eng. Q.* 18, 3 (2004), 195–222. http://silverstripe.fkit.hr/cabeq/assets/Uploads/CABEQ-18-3-1.pdf

[79] Malvina Marku and Vera Pancaldi. 2023. From Time-Series Transcriptomics to Gene Regulatory Networks: A Review on Inference Methods. *PLOS Comput. Biol.* 19, 8 (Aug. 2023), e1011254. https://doi.org/10.1371/journal.pcbi.1011254

[80] Julien Martinelli, Jeremy Grignard, Sylvain Soliman, Annabelle Ballesta, and François Fages. 2023. Reactmine: A Statistical Search Algorithm for Inferring Chemical Reactions from Time Series Data. arXiv:2209.03185

[81] Julien Martinelli, Jeremy Grignard, Sylvain Soliman, and François Fages. 2019. On Inferring Reactions from Data Time Series by a Statistical Learning Greedy Heuristics. In *Computational Methods in Systems Biology*, Luca Bortolussi and Guido Sanguinetti (Eds.), Vol. 11773. Springer International Publishing, Cham, 352–355. https://doi.org/10.1007/978-3-030-31304-3_25

[82] Julien Martinelli, Jeremy Grignard, Sylvain Soliman, and François Fages. 2019. A Statistical Unsupervised Learning Algorithm for Inferring Reaction Networks from Time Series Data. In *ICML - Workshop on Computational Biology*. Long Beach, CA, USA, n.a. https://inria.hal.science/hal-02163862v1

[83] Daniel A. Messenger and David M. Bortz. 2021. Weak SINDy: Galerkin-Based Data-Driven Model Selection. *Multiscale Modeling & Simulation* 19, 3 (2021), 1474–1497. https://doi.org/10.1137/20M1343166

[84] Clémence Métayer, Annabelle Ballesta, and Julien Martinelli. 2025. Data-Driven Discovery of Digital Twins in Biomedical Research. arXiv:2508.21484

[85] Jeffrey A. Mills and Kislaya Prasad. 1992. A comparison of model selection criteria. *Econometric Reviews* 11, 2 (1992), 201–234. https://doi.org/10.1080/07474939208800232

[86] Eshan D. Mitra and William S. Hlavacek. 2019. Parameter estimation and uncertainty quantification for systems biology models. *Current Opinion in Systems Biology* 18 (2019), 9–18. https://doi.org/10.1016/j.coisb.2019.10.006

[87] Enayat A Moallemi, Jan Kwakkel, Fjalar J de Haan, and Brett A Bryan. 2020. Exploratory modeling for analyzing coupled human-natural systems under uncertainty. *Global Environmental Change* 65 (2020), 102186. https://doi.org/10.1016/j.gloenvcha.2020.102186

[88] Alexander Mordvintsev, Ettore Randazzo, and Eyvind Niklasson. 2023. Differentiable Programming of Chemical Reaction Networks. arXiv:2302.02714

[89] Rajiv Teja Nagipogu and John H. Reif. 2025. Neural CRNs: A Natural Implementation of Learning in Chemical Reaction Networks. *ACS Synthetic Biology* 14, 10 (2025), 3899–3912. https://doi.org/10.1021/acssynbio.5c00099 arXiv:https://doi.org/10.1021/acssynbio.5c00099

[90] Nils E Napp and Ryan P Adams. 2013. Message Passing Inference with Chemical Reaction Networks. In *Advances in Neural Information Processing Systems* (Lake Tahoe, Nevada), C.J. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K.Q. Weinberger (Eds.), Vol. 26. Curran Associates, Inc., Red Hook, NY, 2247–2255. https://proceedings.neurips.cc/paper_files/paper/2013/file/e5e63da79fcd2bebbd7cb8bf1c1d0274-Paper.pdf

[91] B. K. Natarajan. 1995. Sparse Approximate Solutions to Linear Systems. *SIAM J. Comput.* 24, 2 (April 1995), 227–234. https://doi.org/10.1137/S0097539792240406

[92] Emily Nieves, Raj Dandekar, and Chris Rackauckas. 2024. Uncertainty Quantified Discovery of Chemical Reaction Systems via Bayesian Scientific Machine Learning. *Front. Syst. Biol.* 4 (March 2024), 1338518. https://doi.org/10.3389/fsysb.2024.1338518

[93] Marco S. Nobile, Daniela Besozzi, Paolo Cazzaniga, Dario Pescini, and Giancarlo Mauri. 2013. Reverse Engineering of Kinetic Reaction Networks by Means of Cartesian Genetic Programming and Particle Swarm Optimization. In *2013 IEEE Congress on Evolutionary Computation*. IEEE, Cancun, Mexico, 1594–1601. https://doi.org/10.1109/CEC.2013.6557752

[94] Marco S. Nobile and Giancarlo Mauri. 2013. Evolutionary Inference of Biochemical Reaction Networks Accelerated on Graphics Processing Units. In *2013 International Conference on High Performance Computing & Simulation (HPCS)*. IEEE, Helsinki, Finland, 668–670. https://doi.org/10.1109/HPCSim.2013.6641490

[95] Marco S. Nobile, Daniele M. Papetti, Simone Spolaor, Paolo Cazzaniga, and Luca Manzoni. 2022. Shaping and Dilating the Fitness Landscape for Parameter Estimation in Stochastic Biochemical Models. *Applied Sciences* 12, 13 (July 2022), 6671. https://doi.org/10.3390/app12136671

[96] Jeremiah Nummela and Bryant A. Julstrom. 2005. Evolving Petri Nets to Represent Metabolic Pathways. In *GECCO05: Genetic and Evolutionary Computation Conference*. ACM, Washington DC USA, 2133–2139. https://doi.org/10.1145/1068009.1068361

[97] Chris J. Oates, Frank Dondelinger, Nora Bayani, James Korkola, Joe W. Gray, and Sach Mukherjee. 2014. Causal Network Inference Using Biochemical Kinetics. *Bioinformatics* 30, 17 (Sept. 2014), i468–i474. https://doi.org/10.1093/bioinformatics/btu452

[98] Kaan Öcal, Ramon Grima, and Guido Sanguinetti. 2020. Parameter Estimation for Biochemical Reaction Networks Using Wasserstein Distances. *Journal of Physics A: Mathematical and Theoretical* 53, 3 (Jan. 2020), 034002. https://doi.org/10.1088/1751-8121/ab5877

[99] Perplexity AI, Inc. 2026. Perplexity Search. https://www.perplexity.ai/ Last Accessed: 2026-01-29.

[100] Linda Petzold. 1983. Automatic Selection of Methods for Solving Stiff and Nonstiff Systems of Ordinary Differential Equations. *SIAM J. Sci. Statist. Comput.* 4, 1 (1983), 136–148. https://doi.org/10.1137/0904010

[101] Niklas Pfister, Stefan Bauer, and Jonas Peters. 2019. Learning Stable and Predictive Structures in Kinetic Systems. *Proc. Natl. Acad. Sci.* 116, 51 (Dec. 2019), 25405–25411. https://doi.org/10.1073/pnas.1905688116

[102] Maria E. Pierce, Tom Warnke, Uwe Krumme, Tobias Helms, Cornelius Hammer, and Adelinde M. Uhrmacher. 2017. Developing and validating a multi-level ecological model of eastern Baltic cod (Gadus morhua) in the Bornholm Basin - a case for domain-specific languages. *Ecological Modeling* 361 (October 2017), 49–65. https://doi.org/10.1016/j.ecolmodel.2017.07.012

[103] Gianluigi Pillonetto, Aleksandr Aravkin, Daniel Gedon, Lennart Ljung, Antônio H. Ribeiro, and Thomas B. Schön. 2025. Deep Networks for System Identification: A Survey. *Automatica* 171 (Jan. 2025), 111907. https://doi.org/10.1016/j.automatica.2024.111907

[104] Anjana Puliyanda, Karthik Srinivasan, Kaushik Sivaramakrishnan, and Vinay Prasad. 2022. A Review of Automated and Data-Driven Approaches for Pathway Determination and Reaction Monitoring in Complex Chemical Systems. *Digital Chem. Eng.* 2 (March 2022), 100009. https://doi.org/10.1016/j.dche.2021.100009

[105] Silvia Rausanu, Crina Grosan, Zujian Wu, Ovidiu Parvu, and David Gilbert. 2013. Evolving Biochemical Systems. In *2013 IEEE Congress on Evolutionary Computation*. IEEE, Cancun, Mexico, 1602–1609. https://doi.org/10.1109/CEC.2013.6557753

[106] Silvia Rausanu, Crina Grosan, Zujian Wu, Ovidiu Parvu, Ramona Stoica, and David Gilbert. 2015. Computational Models for Inferring Biochemical Networks. *Neural Comput. Appl.* 26, 2 (Feb. 2015), 299–311. https://doi.org/10.1007/s00521-014-1617-x

[107] Francisco J. Romero-Campero, Hongqing Cao, Miguel Camara, and Natalio Krasnogor. 2008. Structure and Parameter Estimation for Cell Systems Biology Models. In *GECCO08: Genetic and Evolutionary Computation Conference*. ACM, Atlanta GA USA, 331–338. https://doi.org/10.1145/1389095.1389153

[108] John Ross. 2008. Determination of Complex Reaction Mechanisms. Analysis of Chemical, Biological and Genetic Networks. *J. Phys. Chem. A* 112, 11 (March 2008), 2134–2143. https://doi.org/10.1021/jp711313e

[109] Fadil Santosa and Benjamin Weitz. 2011. An Inverse Problem in Reaction Kinetics. *J. Math. Chem.* 49, 8 (Sept. 2011), 1507–1520. https://doi.org/10.1007/s10910-011-9835-2

[110] Michael A. Savageau. 1988. Introduction to S-systems and the Underlying Power-Law Formalism. *Mathematical and Computer Modelling* 11 (Jan. 1988), 546–551. https://doi.org/10.1016/0895-7177(88)90553-5

[111] David Schaich, Ralf Becker, and Rudibert King. 2001. Qualitative Modelling for Automatic Identification of Mathematical Models of Chemical Reaction Systems. *Control Eng. Pract.* 9, 12 (2001), 1373–1381. https://doi.org/10.1016/S0967-0661(01)00080-6

[112] Michael Douglas Schmidt and Hod Lipson. 2011. Automated Modeling of Stochastic Reactions with Large Measurement Time-Gaps. In *Proceedings of the 13th Annual Conference on Genetic and Evolutionary Computation*. Association for Computing Machinery, Dublin, Ireland, 307–314. https://doi.org/10.1145/2001576.2001619

[113] Dominic P. Searson, Mark J. Willis, and Allen Wright. 2012. Reverse Engineering Chemical Reaction Networks from Time Series Data. In *Statistical Modelling of Molecular Descriptors in QSAR/QSPR* (1 ed.), Matthias Dehmer, Kurt Varmuza, and Danail Bonchev (Eds.). Wiley, 327–348. https://doi.org/10.1002/9783527645121.ch12

[114] Robert E. Shannon. 1998. Introduction to the art and science of simulation. In *Proceedings of the 30th Conference on Winter Simulation* (Washington, D.C., USA) *(WSC '98)*. IEEE Computer Society Press, Washington, DC, USA, 7–14. https://doi.org/10.1109/WSC.1998.744892

[115] Brock D Sherlock, Marko AA Boon, Maria Vlasiou, and Adelle CF Coster. 2024. The distance between: an algorithmic approach to comparing stochastic models to time-series data. *Bulletin of mathematical biology* 86, 9 (2024), 111. https://doi.org/10.1007/s11538-024-01331-y

[116] Abhyudai Singh and Ramon Grima. 2017. The linear-noise approximation and moment-closure approximations for stochastic chemical kinetics. arXiv:1711.07383

[117] Glenn Skrzypczak. 2025. *Benchmarking von Methoden zum Lernen von Chemischen Reaktionsnetzwerken aus Zeitreihendaten.* Master's thesis. University of Rostock. http://eprints.mosi.informatik.uni-rostock.de/846/

[118] Sriniketh Srinivasan, Julien Billeter, and Dominique Bonvin. 2019. Sequential Model Identification of Reaction Systems—the Missing Path between the Incremental and Simultaneous Approaches. *AIChE J.* 65, 4 (April 2019), 1211–1221. https://doi.org/10.1002/aic.16530

[119] Jeyaraman Srividhya, Edmund J. Crampin, Patrick E. McSharry, and Santiago Schnell. 2007. Reconstructing Biochemical Pathways from Time Course Data. *Proteomics* 7, 6 (March 2007), 828–838. https://doi.org/10.1002/pmic.200600428

[120] Gábor Szederkényi. 2010. Computing Sparse and Dense Realizations of Reaction Kinetic Systems. *Journal of Mathematical Chemistry* 47, 2 (Feb. 2010), 551–568. https://doi.org/10.1007/s10910-009-9525-5

[121] Gábor Szederkényi, Julio R Banga, and Antonio A Alonso. 2011. Inference of complex biological networks: distinguishability issues and optimization-based solutions. *BMC systems biology* 5, 1 (2011), 177. https://doi.org/10.1186/1752-0509-5-177

[122] Tamás Székely and Kevin Burrage. 2014. Stochastic Simulation in Systems Biology. *Computational and Structural Biotechnology Journal* 12, 20 (Nov. 2014), 14–25. https://doi.org/10.1016/j.csbj.2014.10.003

[123] Tua A. Tamba and Yul Yunazwin Nazaruddin. 2019. Data-Driven Construction of Chemical Reaction Network Graph Using Constrained LASSO. In *Proceedings of the 2019 6th International Conference on Instrumentation, Control, and Automation*. Institute of Electrical and Electronics Engineers Inc., Bandung, Indonesia, 226–230. https://doi.org/10.1109/ICA.2019.8916720

[124] Connor J. Taylor, Hikaru Seki, Friederike M. Dannheim, Mark J. Willis, Graeme Clemens, Brian A. Taylor, Thomas W. Chamberlain, and Richard A. Bourne. 2021. An Automated Computational Approach to Kinetic Model Discrimination and Parameter Estimation. *React. Chem. Eng.* 6, 8 (2021), 1404–1411. https://doi.org/10.1039/D1RE00098E

[125] Calvin Tsay, Richard C. Pattison, Michael Baldea, Ben Weinstein, Steven J. Hodson, and Robert D. Johnson. 2017. A Superstructure-Based Design of Experiments Framework for Simultaneous Domain-Restricted Model Identification and Parameter Estimation. *Comput. Chem. Eng.* 107 (Dec. 2017), 408–426. https://doi.org/10.1016/j.compchemeng.2017.02.014

[126] Zoltan A. Tuza and Guy-Bart Stan. 2018. Characterization of Biologically Relevant Network Structures Form Time-Series Data. In *2018 IEEE Conference on Decision and Control (CDC) (Proceedings of the IEEE Conference on Decision and Control, Vol. 2018-December)*. IEEE, Miami Beach, FL, 1089–1095. https://doi.org/10.1109/CDC.2018.8619360

[127] Zoltan A. Tuza and Guy-Bart Stan. 2019. An Automatic Sparse Model Estimation Method Guided by Constraints That Encode System Properties. In *2019 18th European Control Conference (ECC)*. IEEE, Naples, Italy, 2171–2176. https://doi.org/10.23919/ECC.2019.8795919

[128] Robin van der Wall. 2025. Vergleich von Methoden zur kombinierten Struktur- und Parameterinferenz von Reaktionsmodellen. Bachelor's Thesis. http://eprints.mosi.informatik.uni-rostock.de/829/

[129] Yulan B van Oppen and Andreas Milias-Argeitis. 2025. Gradient matching accelerates mixed-effects inference for biochemical networks. *Bioinformatics* 41, 4 (04 2025), btaf154. https://doi.org/10.1093/bioinformatics/btaf154

[130] J. M. Varah. 1982. A Spline Least Squares Method for Numerical Parameter Estimation in Differential Equations. *SIAM J. Sci. Statist. Comput.* 3, 1 (1982), 28–46. https://doi.org/10.1137/0903003

[131] T. Varga. 2013. Qualitative Analysis Based Reaction Mechanism Identification. *Chem. Eng. Trans.* 35 (Sept. 2013), 769–774. https://doi.org/10.3303/CET1335128

[132] Eberhard O. Voit, Harald A. Martens, and Stig W. Omholt. 2015. 150 Years of the Mass Action Law. *PLOS Computational Biology* 11, 1 (01 2015), 1–7. https://doi.org/10.1371/journal.pcbi.1004012

[133] C. Ward, E. Yeung, T. Brown, B. Durtschi, S. Weyerman, R. Howes, Jorge Goncalves, and S. Warnick. 2009. A Comparison of Network Reconstruction Methods for Chemical Reaction Networks. In *The proceedings of the Third International Conference on Foundations of Systems Biology in Engineering (FOSBE 2009)* (Denver, Colorado, United States). 197–200. https://hdl.handle.net/10993/20416

[134] Mingjian Wen, Evan Walter Clark Spotte-Smith, Samuel M. Blau, Matthew J. McDermott, Aditi S. Krishnapriyan, and Kristin A. Persson. 2023. Chemical Reaction Networks and Opportunities for Machine Learning. *Nat Comput Sci* 3, 1 (Jan. 2023), 12–24. https://doi.org/10.1038/s43588-022-00369-z

[135] Jason R. White and Ranjan Srivastava. 2017. Integration of Reaction Kinetics Theory and Gene Expression Programming to Infer Reaction Mechanism. In *Applications of Evolutionary Computation*, Giovanni Squillero and Kevin Sim (Eds.). Vol. 10199. Springer International Publishing, Cham, 53–66. https://doi.org/10.1007/978-3-319-55849-3_4

[136] Mark J. Willis and Moritz Von Stosch. 2016. Inference of Chemical Reaction Networks Using Mixed Integer Linear Programming. *Comput. Chem. Eng.* 90 (July 2016), 31–43. https://doi.org/10.1016/j.compchemeng.2016.04.019

[137] Zachary T. Wilson and Nikolaos V. Sahinidis. 2019. Automated Learning of Chemical Reaction Networks. *Comput. Chem. Eng.* 127 (Aug. 2019), 88–98. https://doi.org/10.1016/j.compchemeng.2019.05.020

[138] D.H. Wolpert and W.G. Macready. 1997. No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation* 1, 1 (1997), 67–82. https://doi.org/10.1109/4235.585893

[139] Zujian Wu, Qian Gao, and David Gilbert. 2010. Target Driven Biochemical Network Reconstruction Based on Petri Nets and Simulated Annealing. In *CMSB '10: 8th Conference on Computational Methods in Systems Biology*. ACM, Trento Italy, 33–42. https://doi.org/10.1145/1839764.1839770

[140] Zujian Wu, Crina Grosan, and David Gilbert. 2014. Empirical Study of Computational Intelligence Strategies for Biochemical Systems Modelling. In *Nature Inspired Cooperative Strategies for Optimization (NICSO 2013): Learning, Optimization and Interdisciplinary Applications*, German Terrazas, Fernando E. B. Otero, and Antonio D. Masegosa (Eds.). Springer International Publishing, Cham, 245–260. https://doi.org/10.1007/978-3-319-01692-4_19

[141] Zujian Wu, Shengxiang Yang, and David Gilbert. 2012. A Hybrid Approach to Piecewise Modelling of Biochemical Systems. In *Parallel Problem Solving from Nature - PPSN XII*. Vol. 7491. Springer Berlin Heidelberg, Berlin, Heidelberg, 519–528. https://doi.org/10.1007/978-3-642-32937-1_52

[142] Yafeng Xing, Yachao Dong, Christos Georgakis, and Aaron Gould. 2025. Stoichiometry Model Identification for Homogeneous Reaction Mixture: High-Dimension and Missing Measurement Case Studies. *AIChE J.* 72, 3 (Dec. 2025), e70183. https://doi.org/10.1002/aic.70183

[143] Yafeng Xing, Yachao Dong, Christos Goergakis, Yu Zhuang, Lei Zhang, Jian Du, and Qingwei Meng. 2022. Automatic Data-Driven Stoichiometry Identification and Kinetic Modeling Framework for Homogeneous Organic Reactions. *AIChE J.* 68, 7 (July 2022), e17713. https://doi.org/10.1002/aic.17713

[144] Yafeng Xing, Yachao Dong, Wenjin Zhou, Jian Du, and Qingwei Meng. 2023. Optimization-Based Simultaneous Modelling of Stoichiometries and Kinetics in Complex Organic Reaction System. *Chem. Eng. Sci.* 276 (July 2023), 118758. https://doi.org/10.1016/j.ces.2023.118758

[145] Jin Xu, H. Steven Wiley, and Herbert M. Sauro. 2024. Generating Synthetic Signaling Networks for in Silico Modeling Studies. *Journal of Theoretical Biology* 593 (Oct. 2024), 111901. https://doi.org/10.1016/j.jtbi.2024.111901

[146] Jeehyun Yang and Renyu Hu. 2024. Automated Chemical Reaction Network Generation and Its Application to Exoplanet Atmospheres. *The Astrophysical Journal* 966, 2 (May 2024), 189. arXiv:2402.14784

[147] Wei Zhang, Stefan Klus, Tim Conrad, and Christof Schütte. 2019. Learning Chemical Reaction Networks from Trajectory Data. *SIAM J. Appl. Dyn. Syst.* 18, 4 (Jan. 2019), 2000–2046. https://doi.org/10.1137/19M1265880

Supplemental Information for the Paper "Advances in the Inference of Chemical Reaction Networks from Time Series Data: A Systematic Survey"

Preprint

JUSTIN NOAH KREIKEMEYER and ADELINDE M. UHRMACHER, University of Rostock, Germany

This document contains a detailed description and reporting of the survey process used in the paper "Advances in the Inference of Chemical Reaction Networks from Time Series Data: A Systematic Review".

CCS Concepts: • **General and reference** → **Surveys and overviews**; • **Computing methodologies** → *Artificial intelligence*; *Modeling and simulation*; • **Applied computing** → Systems biology; Chemistry.

Additional Key Words and Phrases: machine learning, model learning, system identification.

1 SYSTEMATIC LITERATURE SURVEY METHODS

Here we summarize our methodological approach to building a representative sample of the literature on CRN inference from time series data to date (1st December 2025). For readers familiar with them, we also report the key facts of our survey process according to the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) [81] in Appendix F. They were originally established for reviews of health interventions, but are now widely known and employed. In particular, we provide a PRISMA flow chart, as well the checklist according to the PRISMA-Search extension [90] relevant to this survey.

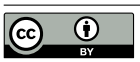
1.1 Search Methodology

To identify the relevant literature, we followed the plan shown in Figure 1. Whereas we argue that current Artifial Intelligence (AI) models cannot replace a manual literature review in the sense presented here [87], we still pay special attention to benefitting from recent developments and integrate AI (not only of the generative kind) as a support in multiple places: augmenting the keyword list and queries, as a query agent for literature, and recommendations based on positive and negative examples. We also experimented with using generative AI (large language models, LLM) for feature extraction from publication PDFs, but obtained results were too unreliable. All AI-generated content and decisions that entered the process were critically and fully reviewed by a human.

1.2 Scope and Selection Criteria

We considered *peer-reviewed* and *pre-print* publications until a cutoff date of 1st December 2025, written in the *English* language. We excluded PhD theses, poster abstracts, publications not online, publications inaccessible due to paywalls, complete conference proceedings or books, and editorials. Note that, for every excluded PhD thesis, our final corpus includes at least one publication from the same author (with the exception of Ronceray [91], but this entry does not meet the criteria below).

Authors' address: Justin Noah Kreikemeyer, justin.kreikemeyer@uni-rostock.de; Adelinde M. Uhrmacher, adelinde.uhrmacher@uni-rostock.de, University of Rostock, Rostock, Germany.



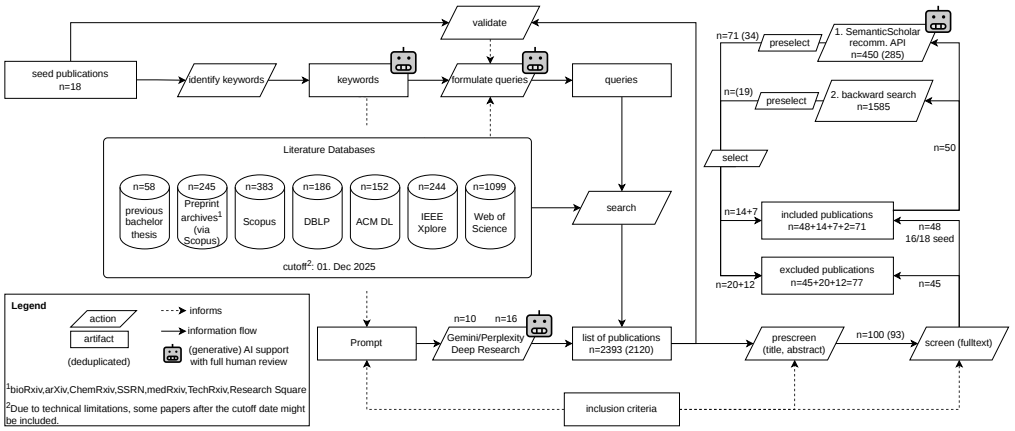


Fig. 1. Flowchart of our literature survey workflow. Previous bachelor thesis refers to van der Wall [108].

Also, only very few publications were excluded due to paywall ($n = 4$) or unavailable online fulltext ($n = 3$). We make the full corpus of excluded works available as well (Appendix E). The criteria for inclusion were as follows:

- Describes a *novel and general method* or *significant combination* or *extension* of existing methods for CRN inference.
- The main input to the method is a single or multiple *transient time series of concentrations or amounts* of the involved species as specified at the end of Section 2.6 in the main text.
- The output of the method should be a CRN, that is, *reactions and kinetic parameters* or be trivially transformable into such a representation (e.g., coupled ODEs or special petri net representations). Methods that just infer the structure are considered only if they specifically target CRNs, leaving general network inference methods out of scope for this survey. Their relation is briefly discussed in Section 8.7 of the main text.
- We allow methods that consider any kind of kinetic law, such as mass-action, Michaelis-Menten (biochemistry), or Arrhenius laws (chemistry).
- The method should allow infering *non-linear* (as well as linear) CRNs.
- The method should search the entire (combinatorial) space of possible models, as opposed to selecting from a limited set of candidate models. However, methods generating models from a large list of candidate *reactions* and/or using prior knowledge to constrain the search space are considered.

For example, we explicitly exclude the following:

- Works that apply an existing method to a specific scientific problem or develop a methodology to solve such a specific problem.
- Works on inferring gene-regulatory networks specifically (e.g., boolean networks), S-Systems, and other targets that are less expressive/general than CRNs (cf. Section 2.5 of the main text).
- Model selection/exploratory modeling methods.
- Parameter estimation methods which require an existing CRN.
- Methods relying on steady-state data only or requiring steady-state data.
- Symbolic regression approaches inferring the kinetic laws/functions for given reactions.

- Approaches that do not produce an explicit form of the CRN, e.g., lack the characteristic ODE coupling or produce a phenomenological model.

These criteria are chosen to give this survey a sharp focus on the problem at hand. However, we acknowledge that many of the excluded works are (marginally) relevant to CRN inference as well, and hence these topics, their relation to our problem, and what may be learned from them, are discussed in Section 8 in the main text. Also, in chemical engineering, many methods that focus on specific applications like combustion or hydrolysis exist, but these are out of scope as this survey focuses on the general problem of CRN inference independent of the domain.

1.3 Seed Papers and Validation Set

We started with a selection of 18 "seed publications" (Appendix A), comprising those the authors already identified as relevant in their previous work in this field and which also match the selection criteria (Section 1.2). Of these, 3 are preprint articles, and 15 are peer-reviewed publications. According to the Scopus database [34], they mainly fall into the fields of Mathematics (12, [12, 13, 19, 29, 48, 55, 60, 65, 71, 74, 76, 93]), Computer Science (8, [12, 13, 48, 55, 59, 60, 71, 76]), Chemistry (3, [47, 51, 93]), Engineering (2, [13, 65]), and Physics (2, [47, 48]).

1.4 Keywords

The seed publications were analyzed to collect a structured list of keywords pertaining to the aspect of automation (e.g., "autonomous"), the target of learning (e.g., "chemical reaction network"), the input to the algorithm (e.g., "time series"), the task to solve (e.g., "inference"), and other relevant keywords (e.g., "systems biology"). We used a large language model to moderately expand our initial list of keywords. The full categorized list of keywords can be found in Appendix B.

1.5 Databases and Queries

Based on the keywords, we formulated queries for different literature databases (cf. Appendix C). We include peer-reviewed publications (as indexed by Scopus [34], Web of Science [24], ACM DL [5], IEEE Xplore [49], and DBLP [30]) as well as preprints (uploaded to bioRxiv [79], arXiv [28], ChemRxiv [20], SSRN [35], medRxiv [80], TechRxiv [50], and Research Square [89]) to capture the latest trends. The preprint archives were queried using the preprint feature in Scopus (includes publications from 2017 and older), enabling direct reuse of our query, which is not possible with each individual archive. As this was a feature in development, no direct export of the results was possible. We instead obtained a machine-readable form by saving the resulting webpage's text 200 entries at a time and subsequently extracted and merged the information. Additionally, we used the literature database built by a previous bachelor's thesis on this topic [108], which queried Google Scholar [40] with different keywords, considering the first 10 pages (100 entries) returned for each query. These are treated just like all other raw results from the literature databases above, undergoing further preselection as well. Finally, we employ an automated "Deep Research" based on Google's Gemini [39] large language model and the Perplexity AI [86] search tool to supplement our results. As there is no agreed upon standard tool, we chose these two based on our existing access to the premium subscription (as opposed to their free tier). These and other novel tools aim to fully automate the discovery and synthesis of literature. However, in our experience, they can currently only act as an additional tool in a classical (systematic) review, particularly in terms of accuracy, specificity, and comprehensiveness. We experienced frequent hallucinations of publication titles or URLs (often still based on existing publications that could subsequently be identified using, e.g., Google Scholar [40]; see Appendix D). Nevertheless, their narrative-style reviews might provide a good starting point for newcomers in a specific field. On the other hand, the purpose of this survey

is to build a taxonomy and identify the current and major future research directions, requiring a more rigorous treatment and background knowledge.

To ensure our queries achieve a high coverage of the relevant literature, we employed the seed publications as a validation set: the formulation of the search queries was fine-tuned (e.g., adding wildcards like "dicovery"→"discov*") to yield a high number of the seed papers, ensuring the best possible coverage. This tuning was done independently for each database queried and we avoided overfitting by only increasing the amount of results in this process. The whole process described in this section is summarized on the left side of Figure 1.

1.6 Query Results and Preselection

In total, we obtained 2393 results at this stage (2120 after automatic deduplication based on title and year). For all of them, we performed a prescreening based on title and abstract, according to the previously defined criteria. If title and abstract were not informative enough or missing, the full text was used as well in a few cases to support the decision. The selection was made by a single reviewer, classifying each entry as "included" (in scope), "marginal" (not fitting the criteria, but deemed relevant for the survey in general) or "excluded" (deemed out of scope). This process yielded a total of 100 candidates for further consideration. Seven of those were identified as duplicates that escaped the previous automatic check and removed.

1.7 Feature Extraction

For each paper in the 93 final candidates, the following features were extracted. We built on our previous experience in the field and the bachelor's/master's theses by van der Wall [108] (on the comparison of CRN inference methods) and Skrzypczak [101] (on benchmarking CRN inference methods) to come up with the initial list of features to extract:

- *Bibliographic information*: Title, Authors, Year, DOI/URL
- *Brief summary*
- *Domain of origin*/motivation (Systems Biology, Chemical Engineering, etc.)
- *Implementation available?* (available at, upon request, inaccessible, not available)
- *General approach category*: Genetic Programming, SINDy-like, …
- *Input*: single time series, multiple time series, distribution snapshots, …
- *Output with uncertainty?* (point estimate, alternative models, etc.)
- *Degree of reconstruction*[1]: deterministic ODEs, CTMC approximation, CTMC
- *Allows knowledge inclusion or constraints* (yes/no)
- *Kind of knowledge/constraints*: known reactions, constraints on parameters, …
- kind of *heuristic* (sparsity, information criterion, knowledge-based, ...)
- combined with *DoE*? (yes/no)
- *Number of benchmark cases*: 1, 2, …
- *Kind of benchmark cases*: synthetic, simulated real CRN, experimental data
- Any results on *complexity/execution time*
- *Additional notes/Reason for exclusion*

It was further refined by several entries during the extraction process and once we finalized our taxonomy, giving finally: **Bibliographic** information; **Summary**; Method **name**, **acronym**, and possible **group** (e.g., in case of journal extensions); Motivating **domain(s)**; **Implementation** availability; **Input** data; **Prior knowledge**; **Design of experiments**; Measurement **noise** considered; Kind of **benchmarks** considered; Number of benchmarks considered; Partial **observability**; CRN **Semantics** considered; Consideration of **uncertainty in inference**; **Kinetic laws** inferrable;

[1]This term was coined by van der Wall [108]

Runtime information; Taxonomy dimension 1: **loss space**; Taxonomy dimension 2: **search**; Taxonomy dimension 3: **inductive bias**. The feature table and complete column descriptions are part of the artifacts archived on Zenodo[2].

During the extraction, if it became clear a paper did not align with the inclusion criteria, it was consequently excluded. In this way, 45 further publications were excluded, yielding 48 final candidates. The excluded candidates and the reason for exclusion are listed in Appendix E for reference and transparency. Among the included publications, 16 out of the 18 seed publications are containted ($\approx$ 89%), leading to a reasonable coverage.

1.8 Augmentation of Results

The final candidates were further augmented using the recommendations API of Semantic Scholar [3], a (non-generative) bibliographic AI tool. First, we queried the internal IDs of our 48 final candidates and the two previously undiscoevered seed publications using the title search API. These were then used as positive examples for a single recommendation query each using the endpoint http://api.semanticscholar.org/recommendations/v1/papers/forpaper. We set a limit of 10 recommendations for each positive id and searched the `all-cs` list. After deduplicaton, this resulted in 34 more candidates, which were screened and from which we extracted the features using the two-step process just like before. This augmented the final candidates list by 14 new results.

For all 48 + 2 final candidates from the initial search, we also conducted a backward search, collecting references from each work, yielding 1585 publications. The additional publications from the backward search underwent the same two step process of prescreening and screening/feature extraction. Prescreening by title and abstract, yielded 19 more publications to consider for extraction. The final number after considering the fulltext is 7.

Overall, the final literature sample considered in the results of this survey includes 48+14+7 = 69, plus two undiscovered seed papers, leading to 71 publications. In the extraction step, we identified which publications discuss the same method (e.g. publications extending short- or conference papers providing additional benchmarks, like Bortolussi et al. [12] and Bortolussi et al. [13]) and also some publications that detailed multiple search methods according to our taxonomy, e.g., Kreikemeyer et al. [59], who discuss a superstructure, an exhaustive, and a template method. By unifying the former and splitting the latter, we arrived at 68 distinct methods in our sample. However, in the main text, we primarily report statistics on the 71 *publications* for simplicity, with the exception of the taxonomy figure, where *methods* are listed.

1.9 Technical Setup

For managing the query results (and also querying in the case of dblp [30] and Semantic Scholar [3]), we employed a set of custom Python scripts (included in the online artifacts) to: transform all result data to a canonical comma-separated value (csv) format (including just the minimal information including id, title, year, authors, and abstract), interactively prescreen entries based on title and abstract, merge results, and deduplicate datasets. Entries in our csv format were hashed based on title and year, which we found sufficiently discriminative. In some cases, we also did some manual cleaning, e.g., to merge entries where the year had been reported differently by two search engines or where a peer-reviewed publication was published later. As the basis for bibliographical data management of the preselected candidates, we used the open-source software Zotero[3] in version `7.0.23` with collections for every step and subcollections for included and excluded entries at the

[2]Available upon request until publication
[3]https://www.zotero.org/

respective step. The Linter for Zotero plugin[4] in version `2.3.0` and manual search were used in addition to Zotero's internal tools for enriching entries with appropriate metadata. The meta-data enriched databases as built with Zotero are included as plain-text BibTeX files with the online artifacts. The second stage screening and feature extraction were done using spreadsheets in the open-source tool LibreOffice Calc[5].

1.10 Sources of Bias

The range of literature databases considered was very broad, consisting of five (six, if counting Google Scholar [40]) common databases, of which two (three) focus on general literature (including biology and chemistry) [24, 34, 40], two focus on computer science [5, 30], and one focuses on electrical engineering, computer science and related fields [49]. In addition, seven preprint servers targeted at biology [79]; chemistry [20]; health sciences [80]; physics, mathematics, computer science, quantitative biology, quantitative finance, statistics, electrical engineering and systems science, and economics [28]; and engineering, computer science, and related fields [50]; as well as two general archives [35, 89], were considered. Bias may arise from the use of Scopus' preprint feature, which is limited to preprints published after 2017. However, this is a plausible restriction as eight years is, in most cases, enough for any preprint to be published (or disregarded).

The statistical significance is also impacted by the selection of keywords from and the fine-tuning of queries to the validation set. Here, an additional test set or even cross-validation would yield more significant results, but the impact of such a time-expensive measure was deemed small given the purpose of this survey. We applied an augmentation of the keyword list by an LLM as a measure against bias in the keyword selection induced by the validation set.

Around 89% (16 out of 18) of the publications in our validation was recovered by the initial search. An investigation showed that the reason for missing Mordvintsev et al. [72] is twofold. First, this publication is only available on arXiv [28]. Thus, the only way it would have ended up in the candidate list is via the Scopus preprint search. As this publication focuses on inferring CRNs from any kind of data on its expected behavior, its title and abstract were missing some important keywords pertaining to the automation and input categories (e.g., "data"). The second candidate (Huang et al. [48]) was likely missed due to only mentioning "chemical reactions" in the title and abstract. We had to eliminate this keyword from Scopus and Web of Science queries, as it lead to too many false positive results on these cross-domain databases and this publication was not indexed in any other database. Both situations could potentially be repaired by adjusting the queries, but only at the cost of an unproportionate number of false positive results. Overall, we consider the obtained sample to be sufficient with respect to the goals of this survey.

A small bias may also be introduced by disregarding works that are not accessible online (e.g., as they are too old) or paywalled. However, this only affected a relatively small number ($n = 4$ no access due to paywall, $n = 3$ not online) of results (146 total considered after prescreening, Appendix E).

The backward and recommendation search may increase the number of publications of the authors already in the corpus, e.g., if they cite their own previous work. We can observe this in our literature sample (based only on last name and first initial): 28 of the 132 authors from our initial publication list before augmentation appeared on at least one additional publication in the final list after augmentation. However, to some extent, this just reflects the community around CRN inference and it is not uncommon for authors to publish or supervise several PhD students

[4] https://github.com/northword/zotero-format-metadata
[5] https://de.libreoffice.org/

on similar topics. Most importantly, 42 new authors (appearing on publications 47 times) were discovered in the augmentation as well.

Finally, the authors of this work have a background in computer science and are most familiar with the terminology in systems biology, so one could expect a bias towards methods developed from this angle in the prescreening and screening processes. However, our final literature sample has a roughly uniform distribtion between these three domains and includes fewer publications motivated by biological rather than chemical applications (17 biology, 19 chemistry, 24 computer science, 11 applied mathematics).

1.11 Remarks on the Helpfulness of AI in our Literature Survey Process

With respect to the use of generative AI, we perceived it as generally helpful in our review process, and employed it as search engine, for augmenting keyword lists, and refining search queries. Generally, the deep research resulted in highly relevant publications (only one exclusion out of 26 during prescreening). However, all except two were also identified via traditional databse queries in our search. We observed some problems with more advanced tasks. Aside from what is reported above, we also attempted to employ Gemini 3 [39] for feature extraction from PDF's during screening. However, we found the results to be unhelpful, as human review of every entry was necessary anyway. When dealing with a large amount of data, hallucinations became frequent and the content of several papers would otherwise have been drastically misrepresented. The same issues as in the deep research were observed: titles were altered and sometimes content was altered. Some of these issues might be alleviated with more advanced prompting, but so far we perceived only a limited capacity of current LLMs to extract features from papers or faithfully synthesize them. Established (non-generative) AI tools, such as Semantic Scholar's recommendations API had a very positive impact on this survey by providing several relevant candidates beyond our initial sample.

A SEED PAPERS

Title	Year	Reference
An inverse problem in reaction kinetics	2011	Santosa and Weitz [93]
Autonomous discovery of unknown reaction pathways from data by chemical reaction neural network	2021	Ji and Deng [51]
Data-driven discovery of multiscale chemical reactions governed by the law of mass action	2022	Huang et al. [48]
Data-driven inference of chemical reaction networks via graph-based variational autoencoders	2023	Bortolussi et al. [12]
Differentiable Programming of Chemical Reaction Networks	2023	Mordvintsev et al. [72]
Discovering biochemical reaction models by evolving libraries	2024	Kreikemeyer et al. [60]
Neuro-symbolic discovery of markov population processes	2025	Bortolussi et al. [13]
On inferring reactions from data time series by a statistical learning greedy heuristics	2019	Martinelli et al. [71]
Quantifying structural uncertainty in chemical reaction network inference	2025	Foo et al. [37]

Title	Year	Reference
Reactive SINDy: discovering governing reactions from concentration data	2019	Hoffmann et al. [47]
Reactmine: a statistical search algorithm for inferring chemical reactions from time series data	2023	Martinelli et al. [70]
Reverse engineering of kinetic reaction networks by means of cartesian genetic programming and particle swarm optimization	2013	Nobile et al. [76]
Sparse regression based structure learning of stochastic reaction networks from single cell snapshot time series	2016	Klimovskaia et al. [55]
Statistical model for biochemical network inference	2013	Craciun et al. [29]
Synthetic likelihood method for reaction network inference	2020	Linder and Rempała [65]
Towards learning stochastic population models by gradient descent	2024	Kreikemeyer et al. [59]
Uncertainty quantified discovery of chemical reaction systems via bayesian scientific machine learning	2024	Nieves et al. [74]
Using a library of chemical reactions to fit systems of ordinary differential equations to agent-based models: a machine learning approach	2024	Burrage et al. [19]

B KEYWORDS

This is the list of keywords based on the seed publications (Appendix A). Keywords are ordered from general to specific.

B.1 Automation

data-driven, autonomous, automatic, algorithm

B.2 Target

kinetic, chemical reactions, chemical reaction network, reaction network, reaction model, reaction system, kinetic model, biochemical reaction network, coupled differential equations, reaction kinetics, population model

B.3 Input

data, time-series data, concentration trajectories, observational data, experimental data

B.4 Task

learning, machine learning, reverse engineering, discovery, inferring, inference, inverse modeling, model synthesis, model generation, identification, system identification, structural calibration, structure learning, empirical modeling, nonlinear regression, sparse regression, symbolic regression

B.5 Others

systems biology, chemistry, in silico, automated science, AI for science, computational biology, scientific machine learning, dynamic modeling

C QUERIES

C.1 Scopus

Last search: 22nd December 2025, 18:00 (preprints and peer-reviewed databases).

```
TITLE-ABS-KEY (
    (
        ( "reaction network" OR "reaction model" OR "kinetic model" OR "reaction
        ↪  system" OR "stochastic population model" OR "reaction pathway" )
        W/70
        ( infer* OR reconstruct* OR recover* OR identif* OR gener* OR discov* OR
        ↪  "inverse problem" OR "reverse engineering" )
    )
    AND ( "automat*" OR "autonomous" OR "data driven" OR "data based" OR
    ↪  "extract*" OR "machine learning" OR "reverse engineering" OR
    ↪  "algorithm*" OR "regression" )
    AND ( "time series" OR data OR timeseries OR "experimental data" OR "observ*
    ↪  data" OR ("time" W/60 "concentration" W/60 "data"))
)
AND NOT TITLE({parameter inference} OR {parameter estimation} OR {rate
↪  constants} OR "boolean network")
AND NOT ABS("boolean network")
AND
(LIMIT-TO (LANGUAGE, "English"))
AND
(
    LIMIT-TO ( SUBJAREA,"COMP" ) OR
    LIMIT-TO ( SUBJAREA,"MATH" )
)
```

C.2 ACM DL

Last search: 22nd December 2025 18:00.

```
[[All: data-driven] OR [All: autonomous] OR [All: automatic]]
AND
[[All: "kinetic"] OR [All: "chemical reaction"] OR [All: "chemical reaction
↪  network"] OR [All: "reaction system"] OR [All: "biochemical reaction
↪  network"] OR [All: "coupled differential equations"] OR [All: "population
↪  model"]]
AND
[[Abstract: "time-series data"] OR [Abstract: "concentration trajectories"] OR
↪  [Abstract: "observational data"] OR [Abstract: "experimental data"]]
AND
[[All: "machine learning"] OR [All: "reverse engineering"] OR [All: discovery]
↪  OR [All: inferring] OR [All: inference] OR [All: inverse modeling] OR [All:
↪  model synthesis] OR [All: model generation] OR [All: system identification]
↪  OR [All: structural calibration] OR [All: structure learning] OR [All:
↪  empirical modeling] OR [All: nonlinear regression] OR [All: "sparse
↪  regression"] OR [All: "symbolic regression"]]
```

```
AND
[E-Publication Date: (* TO 12/22/2025)]
```

C.3 IEEE Xplore

Last search: 22nd December 2025 18:00. Used in conjunction with filter to date range 01/01/1884 - 11/30/2025).

```
("All Metadata":data-driven OR "All Metadata":autonomous OR "All
↪  Metadata":automatic OR "All Metadata":learning)
AND
("All Metadata":"kinetic" OR "All Metadata":"chemical reaction" OR "All
↪  Metadata":"chemical reactions" OR "All Metadata":"chemical reaction
↪  network" OR "All Metadata":"chemical reaction networks" OR "All
↪  Metadata":"reaction system"OR "reaction systems" OR "All
↪  Metadata":"biochemical reaction network" OR "All Metadata":"biochemical
↪  reaction networks" OR "All Metadata":"coupled differential equations" OR
↪  "All Metadata":"coupled differential equation" OR "All
↪  Metadata":"population model" OR "All Metadata":"population models")
AND
("All Metadata":"time-series" OR "All Metadata":"concentration trajectories" OR
↪  "All Metadata":"observational data" OR "All Metadata":"experimental data")
AND
("All Metadata":"machine learning" OR "All Metadata":"reverse engineering" OR
↪  "All Metadata":discovery OR "All Metadata":inferring OR "All
↪  Metadata":inference OR "All Metadata":inverse OR "All Metadata":"model
↪  synthesis" OR "All Metadata":generation OR "All Metadata":"system
↪  identification" OR "All Metadata":"structural calibration" OR "All
↪  Metadata":"structur* learning" OR "All Metadata":"empirical modeling" OR
↪  "All Metadata":"nonlinear regression" OR "All Metadata":"sparse regression"
↪  OR "All Metadata":"symbolic regression")
AND
("All Metadata":"kinetic" OR "All Metadata":"chemical reaction" OR "All
↪  Metadata":"chemical reactions" OR "All Metadata":"chemical reaction
↪  network" OR "All Metadata":"chemical reaction networks" OR "All
↪  Metadata":"reaction system"OR "reaction systems" OR "All
↪  Metadata":"reaction model" OR "All Metadata":"reaction models" OR "All
↪  Metadata":"biochemical reaction network" OR "All Metadata":"biochemical
↪  reaction networks" OR "All Metadata":"coupled differential equations" OR
↪  "All Metadata":"coupled differential equation" OR "All
↪  Metadata":"population model" OR "All Metadata":"population models")
```

```
AND
("All Metadata":"kinetic" OR "All Metadata":"chemical reaction" OR "All
↪  Metadata":"chemical reactions" OR "All Metadata":"chemical reaction
↪  network" OR "All Metadata":"chemical reaction networks" OR "All
↪  Metadata":"reaction system"OR "reaction systems" OR "All
↪  Metadata":"reaction model" OR "All Metadata":"reaction models" OR "All
↪  Metadata":"biochemical reaction network" OR "All Metadata":"biochemical
↪  reaction networks" OR "All Metadata":"coupled differential equations" OR
↪  "All Metadata":"coupled differential equation" OR "All
↪  Metadata":"population model" OR "All Metadata":"population models")
```

C.4 DBLP

The following cross-product of keyword sets was queried: {"reaction network", "reaction model", "stochastic population model", "kinetic model"}×{"inference", "learning", "reverse engineering", "identification", "symbolic regression", "discovery"}. Last search: 22nd December 2025 18:00.

C.5 Web of Science

We additionally set a filter to only include the document types Article, Other, Unspecified, Review, and Early Access. Last search: 22nd December 2025 18:00.

```
TS=(
   (
     (
       ("reaction network" OR "reaction model" OR "kinetic model" OR
        "reaction system" OR "stochastic population model" OR
        "reaction pathway")
       NEAR/70
       (infer* OR reconstruct* OR recover* OR identif* OR gener* OR discov* OR
        "inverse problem" OR "reverse engineering")
     )
     AND
     ("automat*" OR "autonomous" OR "data driven" OR "data based" OR extract* OR
      "machine learning" OR "reverse engineering" OR algorithm* OR regression)
     AND
     ("time series" OR data OR timeseries OR "experimental data" OR
      "observ* data" OR (time NEAR/60 concentration NEAR/60 data))
   )
 )
AND LA=(English)
NOT TI=("parameter inference" OR "parameter estimation" OR "rate constants" OR
        "boolean network" OR "retrosynthe*")
NOT AB=("boolean network" OR "retrosynthe*")
AND PY=(1900-2026)
```

D DEEP RESEARCH PROMPT

The prompt used for querying deep research generative AI tools was as follows. Both tools queried, Gemini (last query 2nd January 2026 20:00) and Preplexity (last query 3rd January 2026 12:00), did

not adhere to the final statement, writing a multi-page review and multiple tables, respectively. Also, both routinely output wrong citations, changing the title or using a DOI that does not refer to any or another work. However, in many cases the intended references could be identified manually.

"*The review is about methods for inferring Chemical Reaction Networks (CRNs) from time-series data on the involved species. It should include methods that can infer a list of reactions including a kinetic law (e.g., mass action) and corresponding rate constants that give rise to the observed dynamics. The kinetic law may be fixed a priori, for example to mass-action kinetics. Selected papers (true positives) are characterized as follows:*

- *describes a novel method or a significant combination of existing methods.*
- *input to the algorithm is a time-series of species concentrations or amounts (or distributions thereof over time); and possibly some prior information as supplement.*
- *output of the algorithm should be a CRN, consisting of reactions and rate constants/kinetics. Methods that just infer the structure are considered only if they specifically target CRNs, leaving general network inference methods out of scope for this review.*
- *The kinetic law should also be inferred along with the reactions. However, we also allow methods that consider a (set of) fixed kinetic law(s), such as mass-action, Michaelis-Menten, or Arrhenius laws.*
- *the algorithm should be able to discover the whole spectrum of CRNs, e.g., linear as well as non-linear CRNs*
- *the algorithm should perform a search in the (possibly complete) model space and not just perform model selection on a small set of possible candidate CRNs. However, methods generating models from a large list of candidate reactions and/or using prior knowledge to constrain the search space are considered.*

These are some things that often end up as false positives to be aware of:

- *Papers targeting specific applications in chemistry or biology only (not describing a general method), e.g., approaches specifically targeting combustion*
- *Papers pertaining to, e.g., boolean models of gene regulatory networks or S-Systems as target only*
- *pure parameter estimation methods (which may still often be called "CRN inference")*
- *model selection methods (selecting from a large list of possible reactions, however, is fine)*
- *symbolic regression methods that only infer the kinetic laws for a fixed reaction structure*
- *Approaches that don't produce an explicit/interpretable form of the CRN (e.g., standard SINDy, Neural ODE).*

VERY IMPORTANT: The output should be a SINGLE TABLE of true positives and NOTHING ELSE. The table should include as header: title,year,authors,doi/url. Do not provide an executive summary or similar. Do not provide a mini literature review or similar."

E CANDIDATES EXCLUDED DURING SCREENING

The reasons for exclusion were grouped into:

- **Inference process**
  - **parameter estimation:** ($n = 3$) only estimates parameters and not structure.
  - **model selection:** ($n = 4$) only selects between a limited set of alternative models.
  - **CRN known:** ($n = 6$) if a known CRN structure is required as input.
  - **not fully automated** ($n = 3$) when some part of the inference requires manual intervention/deduction.
- **Input**
  - **not time series:** ($n = 11$) the input is not time series data (e.g., steady state data).

- **Target**
  - **target not CRN:** ($n = 11$) the target is not a CRN, e.g., S-System or gene-regulatory network (GRN).
  - **no automatic coupling:** ($n = 21$) if the output is a dynamical system (e.g., ODE system) supposed to model a CRN, but the characteristic reaction coupling is not algorithmically enforced.
  - **only linear:** ($n = 2$) the learned CRN may only be linear.
  - **hybrid:** ($n = 2$) describes a hybrid phenomenological/mechanistic approach, e.g., kinetic laws are approximated by neural networks.
- **Bibliographical and Other**:
  - **pr available:** ($n = 1$) if there was a **p**eer-**r**eviewed version available, the preprint was excluded in favor of the former.
  - **phd thesis:** ($n = 6$) whether the entry is a phd theses.
  - **not online:** ($n = 3$) was not available online.
  - **no access:** ($n = 4$) could not access due to paywall.

The following table provides a complete overview over all 77 publications excluded during screening and the (main) reason for exclusion:

Title	Year	Ref.	Reason for exclusion
A novel strategy for dynamic modelling of genome-scale interaction networks	2022	[14]	parameter estimation
Active learning of chemical reaction networks via probabilistic graphical models and boolean reaction circuits	2023	[26]	not time series
Algorithm to generate reaction pathways for computer-assisted elucidation	1992	[106]	not time series
Applying intelligent computing techniques to modeling biological networks from expression data	2008	[63]	target not CRN
Automated kinetic model discovery – a methodological framework	2023	[32]	no automatic coupling
automated kinetic model identification via cloud services using model-based design of experiments	2024	[2]	model selection
Automatic kinetic model generation and selection based on concentration versus time curves	2020	[73]	model selection
Autonomous kinetic model identification using optimal experimental design and retrospective data analysis: methane complete oxidation as a case study	2023	[85]	model selection
Constraint-guided symbolic regression for data-efficient kinetic model discovery	2025	[100]	target not CRN
Data-driven discovery of chemical reaction mechanisms from limited concentration profiles	2025	[45]	pr available
Data-driven discovery of reaction kinetic models in dynamic plug flow reactors using symbolic regression	2024	[25]	target not CRN

Title	Year	Ref.	Reason for exclusion
Derivative-Free Domain-Informed Data-Driven Discovery of Sparse Kinetic Models	2025	[88]	no automatic coupling
DoE-SINDy: an automated framework for model generation and selection in kinetic studies	2025	[66]	no automatic coupling
Estimation and discrimination of stochastic biochemical circuits from time-lapse microscopy data	2012	[105]	CRN known
Evolving biochemical reaction networks with stochastic attributes	2009	[53]	not time series
Experimental design for the identification of hybrid reaction models from transient data	2008	[15]	hybrid
Identifiability and reconstruction of biochemical reaction networks from population snapshot data	2018	[23]	only linear
Identification of a metabolic reaction network from time-series data of metabolite concentrations	2013	[103]	target not CRN
Identification of chemical reaction mechanism from batch process data	2006	[98]	no automatic coupling
Identification of nonlinear state-space systems from heterogeneous datasets	2018	[84]	no automatic coupling
Identifying biochemical reaction networks from heterogeneous datasets	2015	[83]	no automatic coupling
Inference of chemical reaction networks	2008	[18]	no automatic coupling
Inference of chemical reaction networks based on concentration profiles using an optimization framework	2019	[61]	not time series
Inference of chemical reaction networks using hybrid S-system models	2007	[99]	no automatic coupling
Inference of differential equations for modeling chemical reactions	2009	[113]	no automatic coupling
Inferring biochemical reaction pathways: the case of the gemcitabine pharmacokinetics	2012	[62]	only linear
Inferring biological networks by sparse identification of nonlinear dynamics	2016	[68]	no automatic coupling
Inferring reaction networks using perturbation data	2018	[22]	not time series
Mass conservation and inference of metabolic networks from high-throughput mass spectrometry data	2011	[6]	not time series
Model-based evaluation and data requirements for parallel kinetic experimentation and data-driven reaction identification and optimization	2023	[52]	CRN known

Title	Year	Ref.	Reason for exclusion
note concerning the online kinetics identification by using parallel short-cut recursive exact estimators and a knowledge kinetic model-data bank - exemplification for a catalytic batch process	1995	[69]	not online
On sparse identification of complex dynamical systems: a study on discovering influential reactions in chemical reaction networks	2020	[44]	CRN known
Qualitative modelling and simulation of chemical reaction systems	1999	[95]	CRN known
Qualitative modelling for automatic identification of mathematical models of chemical reaction systems	2001	[94]	model selection
Realizing reduced and sparse biochemical reaction networks from dynamics	2025	[36]	no automatic coupling
Reconstructing chemical reaction networks: data mining meets system identification	2008	[21]	not time series
Reconstruction of arbitrary biochemical reaction networks: a compressive sensing approach	2012	[82]	target not CRN
Reverse engineering of biochemical reaction networks using Co-evolution with eng-genes	2013	[41]	no automatic coupling
Rule-based ab initio kinetic model for alkyl sulfide pyrolysis	2015	[107]	not time series
Semi-supervised machine learning approach for reaction stoichiometry and kinetic model identification using spectral data from flow reactors	2024	[109]	not time series
Simultaneous parameter identification and discrimination of the nonparametric structure of hybrid semiparametric models	2017	[111]	no automatic coupling
SINDy-CRN: sparse identification of chemical reaction networks from data	2023	[10]	no automatic coupling
Sparse identification of chemical reaction networks from concentration-time series	2024	[9]	no automatic coupling
The automated discovery of kinetic rate models – methodological frameworks	2024	[31]	no automatic coupling
A linear-in-parameters genetic programming method for chemical kinetics system identification	2019	[114]	no automatic coupling
An integrated qualitative and quantitative biochemical model learning framework using evolutionary strategy and simulated annealing	2015	[112]	not time series

Title	Year	Ref.	Reason for exclusion
automated probing and inference of analytical models for metabolic network dynamics	2010	[110]	no automatic coupling
Automated refinement and inference of analytical models for metabolic networks	2011	[97]	no automatic coupling
Bayesian inference for protein signalling networks	2013	[78]	phd thesis
Bayesian inference of chemical reaction networks	2016	[38]	phd thesis
Chemical kinetics bayesian inference toolbox (CKBIT)	2021	[27]	parameter estimation
Elucidation of chemical reaction networks through genetic algorithm	2017	[46]	phd thesis
Evolutionary inference of biological systems accelerated on graphics processing units	2015	[75]	phd thesis
GPU-powered evolutionary design of mass-action-based models of gene regulation	2016	[77]	target not CRN
Identifying metabolic pathways and gene regulation networks with evolutionary algorithms	2003	[54]	no access
Kolmogorov-arnold chemical reaction neural networks for learning pressure-dependent kinetic rate laws	2025	[56]	CRN known
Learning dynamical models from stochastic trajectories	2024	[91]	phd thesis
Machine science: automated modeling of deterministic and stochastic dynamical systems	2011	[96]	phd thesis
Modeling the dynamics of biological networks from time course data	2010	[33]	no access
reverse engineering and automatic synthesis of metabolic pathways from observed data using genetic programming	2001	[57]	no access
reverse engineering by means of genetic programming of metabolic pathways from observed data	2004	[58]	not online
Solving dynamical inverse problems by means of metabolic P systems	2012	[67]	target not CRN
towards the automated deduction of chemical reaction mechanism	2006	[17]	not online
A superstructure based approach to chemical reactor network synthesis	1990	[1]	target not CRN
An incremental approach for the identification of reaction kinetics	2004	[16]	hybrid
Discovering governing equations via moving horizon learning: the case of reacting systems	2022	[64]	no automatic coupling

Title	Year	Ref.	Reason for exclusion
Identification of regulatory structure and kinetic parameters of biochemical networks via mixed-integer dynamic optimization	2013	[42]	target not CRN
Incremental and simultaneous identification of reaction kinetics: methods and comparison	2004	[7]	parameter estimation
On the deduction of chemical reaction pathways from measurements of time series of concentrations	2001	[92]	not fully automated
Reverse engineering of biochemical equations from time-course data by means of genetic programming	2005	[104]	no automatic coupling
Sequential model identification of reaction systems—the missing path between the incremental and simultaneous approaches	2019	[102]	CRN known
Statistical construction of chemical reaction mechanisms from measured time-series	1995	[4]	target not CRN
Structural identification of nonlinear mathematical models for bioprocesses	1998	[8]	not fully automated
Target factor analysis for the identification of stoichiometric models	1990	[11]	target not CRN

F PRISMA REPORTING

This appendix contains reporting documents according to the PRISMA guidelines, in particular the extension PRISMA Search [90] relevant to this paper as well as a flow chart [43].

F.1 PRISMA Flowchart

We present a standard PRISMA flowchart in Figure 2. It was initially generated using the tooling developed by Haddaway et al. [43] available at https://estech.shinyapps.io/prisma_flowdiagram and subsequently edited further to better fit this survey.

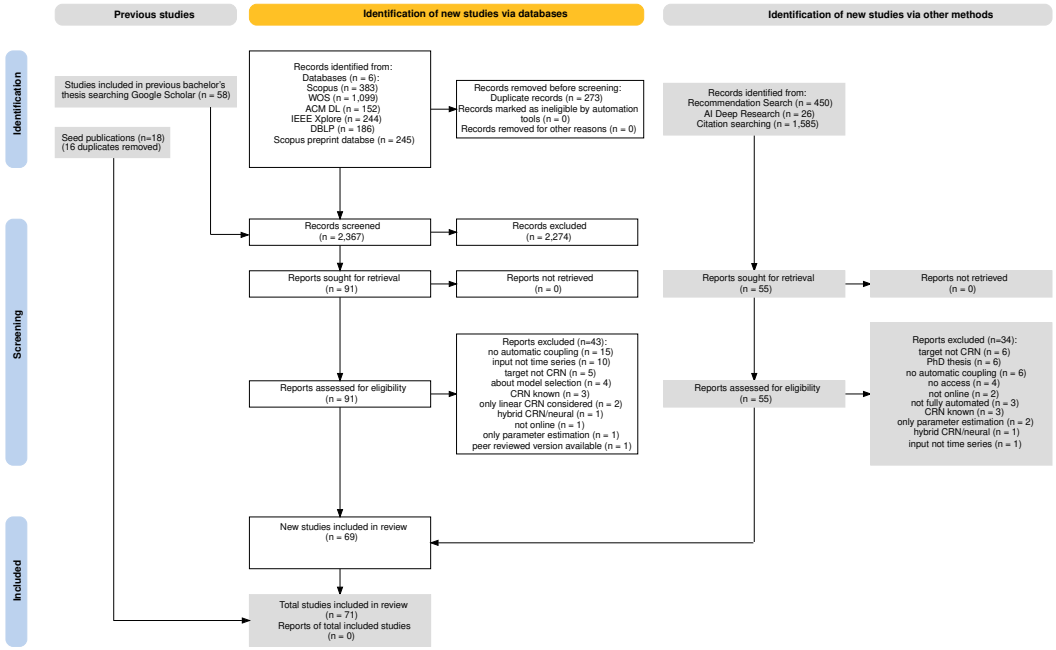


Fig. 2. PRISMA Flowchart. Note that the numbers slightly differ from our own presentation, as the previous bachelor thesis and deep research results were included together in the first step while the recommendation and citation search were separated. Here the previous bachelor thesis is included under previous studies and the deep research is included under other methods.

F.2 PRISMA Search Checklist

Section/topic	#	Checklist item	Location(s) Reported
INFORMATION SOURCES AND METHODS			
Database name	1	Name each individual database searched, stating the platform for each.	Supp. 1.5, Supp. Fig. 1, 2
Multi-database searching	2	If databases were searched simultaneously on a single platform, state the name of the platform, listing all of the databases searched.	Supp. 1.5, Supp. Fig. 1, 2
Study registries	3	List any study registries searched.	n/a
Online resources and browsing	4	Describe any online or print source purposefully searched or browsed (e.g., tables of contents, print conference proceedings, web sites), and how this was done.	n/a
Citation searching	5	Indicate whether cited references or citing references were examined, and describe any methods used for locating cited/citing references (e.g., browsing reference lists, using a citation index, setting up email alerts for references citing included studies).	Supp. 1.8
Contacts	6	Indicate whether additional studies or data were sought by contacting authors, experts, manufacturers, or others.	n/a
Other methods	7	Describe any additional information sources or search methods used.	Supp. 1.5, 1.8
SEARCH STRATEGIES			
Full search strategies	8	Include the search strategies for each database and information source, copied and pasted exactly as run.	Supp. C, 1.5, 1.8
Limits and restrictions	9	Specify that no limits were used, or describe any limits or restrictions applied to a search (e.g., date or time period, language, study design) and provide justification for their use.	Supp. C, 1.5, 1.8
Search filters	10	Indicate whether published search filters were used (as originally designed or modified), and if so, cite the filter(s) used.	n/a
Prior work	11	Indicate when search strategies from other literature reviews were adapted or reused for a substantive part or all of the search, citing the previous review(s).	n/a
Updates	12	Report the methods used to update the search(es) (e.g., rerunning searches, email alerts).	n/a
Dates of searches	13	For each search strategy, provide the date when the last search occurred.	Supp. C
PEER REVIEW			
Peer review	14	Describe any search peer review process.	n/a
MANAGING RECORDS			
Total Records	15	Document the total number of records identified from each database and other information sources.	Supp. Fig. 1, 2
Deduplication	16	Describe the processes and any software used to deduplicate records from multiple database searches and other information sources.	Supp. 9

PRISMA-S: An Extension to the PRISMA Statement for Reporting Literature Searches in Systematic Reviews
Rethlefsen ML, Kirtley S, Waffenschmidt S, Ayala AP, Moher D, Page MJ, Koffel JB, PRISMA-S Group.
Last updated February 27, 2020.

REFERENCES

[1] L.K.E. Achenie and L.T. Biegler. 1990. A Superstructure Based Approach to Chemical Reactor Network Synthesis. *Comput. Chem. Eng.* 14, 1 (Jan. 1990), 23–40. https://doi.org/10.1016/0098-1354(90)87003-8

[2] Emmanuel Agunloye, Panagiotis Petsagkourakis, Muhammad Yusuf, Ricardo Labes, Thomas Chamberlain, Frans L. Muller, Richard A. Bourne, and Federico Galvanin. 2024. Automated Kinetic Model Identification via Cloud Services Using Model-Based Design of Experiments. *React. Chem. Eng.* 9, 7 (2024), 1859–1876. https://doi.org/10.1039/D4RE00047A

[3] AI2. 2026. SemanticScholar. A free, AI-powered research tool for scientific literature. https://www.semanticscholar.org/ Last Accessed: 2026-01-29.

[4] Adam Arkin and John Ross. 1995. Statistical Construction of Chemical Reaction Mechanisms from Measured Time-Series. *J. Phys. Chem.* 99, 3 (Jan. 1995), 970–979. https://doi.org/10.1021/j100003a020

[5] Association for Computing Machinery (ACM). 2026. ACM Digital Library. https://dl.acm.org/ Last Accessed: 2026-01-29.

[6] Pradeep Bandaru, Mukesh Bansal, and Ilya M. Nemenman. 2011. Mass Conservation and Inference of Metabolic Networks from High-Throughput Mass Spectrometry Data. *J. Comput. Biol.* 18, 2 (2011), 147–154. https://doi.org/10.1089/cmb.2010.0222

[7] André Bardow and Wolfgang Marquardt. 2004. Incremental and Simultaneous Identification of Reaction Kinetics: Methods and Comparison. *Chem. Eng. Sci.* 59, 13 (July 2004), 2673–2684. https://doi.org/10.1016/j.ces.2004.03.023

[8] Olivier Bernard and Georges Bastin. 1998. Structural Identification of Nonlinear Mathematical Models for Bioprocesses. In *Proceedings of the Nonlinear Control Systems Symposium*. Elsevier, Enschede, The Netherlands, 449–454.

[9] Nirav Bhatt and Bayu Jayawardhana. 2024. Sparse Identification of Chemical Reaction Networks from Concentration-Time Series. In *26th International Symposium on Mathematical Theory of Networks and Systems*, Vol. 58. IFAC, Cambridge, UK, 0–5. https://hdl.handle.net/11370/c5724385-4c24-4803-ab5f-f9a19df31ece

[10] Nirav Bhatt, Bayu Jayawardhana, and Santiago Sánchez-Escalonilla Plaza. 2023. SINDy-CRN: Sparse Identification of Chemical Reaction Networks from Data. In *2023 62nd IEEE Conference on Decision and Control (CDC)*. IEEE, Singapore,

Singapore, 3512–3518. https://doi.org/10.1109/CDC49753.2023.10384032
[11] D. Bonvin and D.W.T. Rippin. 1990. Target Factor Analysis for the Identification of Stoichiometric Models. *Chem. Eng. Sci.* 45, 12 (1990), 3417–3426. https://doi.org/10.1016/0009-2509(90)87147-K
[12] Luca Bortolussi, Francesca Cairoli, Julia Klein, and Tatjana Petrov. 2023. Data-Driven Inference of Chemical Reaction Networks via Graph-Based Variational Autoencoders. In *Quantitative Evaluation of Systems*, Nils Jansen and Mirco Tribastone (Eds.). Vol. 14287. Springer Nature Switzerland, Cham, 143–147. https://doi.org/10.1007/978-3-031-43835-6_10
[13] Luca Bortolussi, Francesca Cairoli, Julia Klein, and Tatjana Petrov. 2025. Neuro-Symbolic Discovery of Markov Population Processes. In *Proceedings of the International Conference on Neuro-symbolic Systems*. PMLR, Philadelphia, PA, USA, 396–408. https://proceedings.mlr.press/v288/bortolussi25a.html
[14] Pooya Borzou, Jafar Ghaisari, Iman Izadi, Yasin Eshraghi, and Yousof Gheisari. 2022. A Novel Strategy for Dynamic Modelling of Genome-Scale Interaction Networks. https://doi.org/10.1101/2022.05.20.491854
[15] Marc Brendel and Wolfgang Marquardt. 2008. Experimental Design for the Identification of Hybrid Reaction Models from Transient Data. *Chem. Eng. J.* 141, 1-3 (July 2008), 264–277. https://doi.org/10.1016/j.cej.2007.12.027
[16] Marc Brendel, Adel Mhamdi, Dominique Bonvin, and Wolfgang Marquardt. 2004. An Incremental Approach for the Identification of Reaction Kinetics. *IFAC Proc. Vol.* 37, 1 (Jan. 2004), 173–178. https://doi.org/10.1016/S1474-6670(17)38727-X
[17] Burnham, Searson, Willis, and Wright. 2006. Towards the Automated Deduction of Chemical Reaction Mechanism. In *Proceedings of the 17th International Congress of Chemical and Process Engineering*. Prague, Czech Republic.
[18] Samantha C. Burnham, Dominic P. Searson, Mark J. Willis, and Allen R. Wright. 2008. Inference of Chemical Reaction Networks. *Chem. Eng. Sci.* 63, 4 (Feb. 2008), 862–873. https://doi.org/10.1016/j.ces.2007.10.010
[19] Pamela M. Burrage, Hasitha N. Weerasinghe, and Kevin Burrage. 2024. Using a Library of Chemical Reactions to Fit Systems of Ordinary Differential Equations to Agent-Based Models: A Machine Learning Approach. *Numerical Algorithms* 96, 3 (July 2024), 1063–1077. https://doi.org/10.1007/s11075-023-01737-0
[20] ChemRxiv co-owners. 2026. ChemRxiv: chemistry preprints. https://chemrxiv.org/ Last Accessed: 2026-01-29.
[21] Yong Ju Cho, Naren Ramakrishnan, and Yang Cao. 2008. Reconstructing Chemical Reaction Networks: Data Mining Meets System Identification. In *KDD08: The 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining)*. ACM, Las Vegas Nevada USA, 142–150. https://doi.org/10.1145/1401890.1401912
[22] Kiri Choi, Joesph Hellerstein, H. Steven Wiley, and Herbert M. Sauro. 2018. Inferring Reaction Networks Using Perturbation Data. , 351767 pages. https://doi.org/10.1101/351767
[23] Eugenio Cinquemani. 2018. Identifiability and Reconstruction of Biochemical Reaction Networks from Population Snapshot Data. *Processes* 6, 9 (Aug. 2018), 136–160. https://doi.org/10.3390/pr6090136
[24] Clarivate. 2026. Web of Science. https://www.webofscience.com/ Last Accessed: 2026-01-29.
[25] Benjamin G. Cohen, Burcu Beykal, and George M. Bollas. 2024. Data-Driven Discovery of Reaction Kinetic Models in Dynamic Plug Flow Reactors Using Symbolic Regression. *Comput. Aided Chem. Eng.* 53 (2024), 2947–2952. https://doi.org/10.1016/B978-0-443-28824-1.50492-0
[26] Maximilian Cohen, Tejas Goculdas, and Dionisios G. Vlachos. 2023. Active Learning of Chemical Reaction Networks *via* Probabilistic Graphical Models and Boolean Reaction Circuits. *React. Chem. Eng.* 8, 4 (2023), 824–837. https://doi.org/10.1039/D2RE00315E
[27] Maximilian Cohen and Dionisios G. Vlachos. 2021. Chemical Kinetics Bayesian Inference Toolbox (CKBIT). *Comput. Phys. Commun.* 265 (Aug. 2021), 107989. https://doi.org/10.1016/j.cpc.2021.107989
[28] Cornell University. 2026. arXiv.org e-Print archive. https://arxiv.org/ Last Accessed: 2026-01-29.
[29] Gheorghe Craciun, Jaejik Kim, Casian Pantea, and Grzegorz A. Rempala. 2013. Statistical Model for Biochemical Network Inference. *Commun. Stat.- Simul. Comput.* 42, 1 (Jan. 2013), 121–137. https://doi.org/10.1080/03610918.2011.633200
[30] dblp team, Schloss Dagstuhl, and Universität Trier. 2026. dblp computer science bibliography. https://dblp.org/ Last Accessed: 2026-01-29.
[31] Miguel Ángel De Carvalho Servia, Ilya Orson Sandoval, King Kuok (Mimi) Hii, Klaus Hellgardt, Dongda Zhang, and Ehecatl Antonio Del Rio Chanona. 2024. The Automated Discovery of Kinetic Rate Models – Methodological Frameworks. *Digital Discovery* 3, 5 (2024), 954–968. https://doi.org/10.1039/D3DD00212H
[32] Miguel Angel de Carvalho Servia, Ilya Orson Sandoval, Dongda Zhang, Klaus Hellgardt, King Kuok Mimi Hii, and Ehecatl Antonio Del Rio-Chanona. 2023. Automated Kinetic Model Discovery – a Methodological Framework. *Comput. Aided Chem. Eng.* 52 (2023), 33–38. https://doi.org/10.1016/B978-0-443-15274-0.50006-8
[33] Sašo Džeroski and Ljupčo Todorovski. 2010. Modeling the Dynamics of Biological Networks from Time Course Data. In *Systems Biology for Signaling Networks*, Sangdun Choi and Sangdun Choi (Eds.). Springer New York, New York, NY, 275–294. https://doi.org/10.1007/978-1-4419-5797-9_11

[34] Elsevier B.V. 2026. Scopus. https://www.scopus.com Last Accessed: 2026-01-29.
[35] Elsevier, Inc. 2026. SSRN Home Page. https://www.ssrn.com/ Last Accessed: 2026-01-29.
[36] Maurice Filo and Mustafa Khammash. 2025. Realizing Reduced and Sparse Biochemical Reaction Networks from Dynamics. arXiv:2508.18096
[37] Yong See Foo, Adriana Zanca, Jennifer A. Flegg, and Ivo Siekmann. 2025. Quantifying Structural Uncertainty in Chemical Reaction Network Inference. arXiv:2505.15653
[38] Nikhil Galagali. 2016. *Bayesian Inference of Chemical Reaction Networks*. Thesis. Massachusetts Institute of Technology.
[39] Google, Inc. 2026. Google Gemini. https://gemini.google.com Last Accessed: 2026-01-29.
[40] Google, Inc. 2026. Google Scholar. https://scholar.google.com/ Last Accessed: 2026-01-29.
[41] Padhraig Gormley, Kang Li, Olaf Wolkenhauer, George W. Irwin, and Dajun Du. 2013. Reverse Engineering of Biochemical Reaction Networks Using Co-evolution with Eng-Genes. *Cognit. Comput.* 5, 1 (March 2013), 106–118. https://doi.org/10.1007/s12559-012-9159-y
[42] Gonzalo Guillén-Gosálbez, Antoni Miró, Rui Alves, Albert Sorribas, and Laureano Jiménez. 2013. Identification of Regulatory Structure and Kinetic Parameters of Biochemical Networks via Mixed-Integer Dynamic Optimization. *BMC Syst. Biol.* 7, 1 (2013), 113–124. https://doi.org/10.1186/1752-0509-7-113
[43] Neal R. Haddaway, Matthew J. Page, Chris C. Pritchard, and Luke A. McGuinness. 2022. PRISMA2020: An R package and Shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimised digital transparency and Open Synthesis. *Campbell Systematic Reviews* 18, 2 (2022), cl2.1230. https://doi.org/10.1002/cl2.1230
[44] Farshad Harirchi, Doohyun Kim, Omar Khalil, Sijia Liu, Paolo Elvati, Mayank Baranwal, Alfred Hero, and Angela Violi. 2020. On Sparse Identification of Complex Dynamical Systems: A Study on Discovering Influential Reactions in Chemical Reaction Networks. *Fuel* 279 (Nov. 2020), 118204. https://doi.org/10.1016/j.fuel.2020.118204
[45] Shun Hayashi. 2025. Data-Driven Discovery of Chemical Reaction Mechanisms from Limited Concentration Profiles. https://doi.org/10.26434/chemrxiv-2025-2cw0z
[46] Charles Jun Khiong Hii. 2017. *Elucidation of Chemical Reaction Networks through Genetic Algorithm*. Thesis. Newcastle University.
[47] Moritz Hoffmann, Christoph Fröhner, and Frank Noé. 2019. Reactive SINDy: Discovering Governing Reactions from Concentration Data. *J. Chem. Phys.* 150, 2 (Jan. 2019), 25101. https://doi.org/10.1063/1.5066099
[48] Juntao Huang, Yizhou Zhou, and Wen-An Yong. 2022. Data-Driven Discovery of Multiscale Chemical Reactions Governed by the Law of Mass Action. *J. Comput. Phys.* 448 (Jan. 2022), 110743. https://doi.org/10.1016/j.jcp.2021.110743
[49] Institute of Electrical and Electronics Engineers (IEEE). 2026. IEEE Xplore. https://ieeexplore.ieee.org Last Accessed: 2026-01-29.
[50] Institute of Electrical and Electronics Engineers (IEEE). 2026. TexRxiv. https://www.techrxiv.org/ Last Accessed: 2026-01-29.
[51] Weiqi Ji and Sili Deng. 2021. Autonomous Discovery of Unknown Reaction Pathways from Data by Chemical Reaction Neural Network. *J. Phys. Chem. A* 125, 4 (Feb. 2021), 1082–1092. https://doi.org/10.1021/acs.jpca.0c09316
[52] Nathan Jiscoot, Evgeny A. Uslamin, and Evgeny A. Pidko. 2023. Model-Based Evaluation and Data Requirements for Parallel Kinetic Experimentation and Data-Driven Reaction Identification and Optimization. *Digital Discovery* 2, 4 (2023), 994–1005. https://doi.org/10.1039/D3DD00016H
[53] Thomas R. Kiehl. 2009. Evolving Biochemical Reaction Networks with Stochastic Attributes. In *GECCO09: Genetic and Evolutionary Computation Conference*. ACM, Montreal Québec Canada, 2065–2070. https://doi.org/10.1145/1570256.1570277
[54] Junji Kitagawa and Hitoshi Iba. 2003. Identifying Metabolic Pathways and Gene Regulation Networks with Evolutionary Algorithms. In *Evolutionary Computation in Bioinformatics*. Elsevier, 255–278. https://doi.org/10.1016/B978-155860797-2/50014-7
[55] Anna Klimovskaia, Stefan Ganscha, and Manfred Claassen. 2016. Sparse Regression Based Structure Learning of Stochastic Reaction Networks from Single Cell Snapshot Time Series. *PLOS Comput. Biol.* 12, 12 (Dec. 2016), e1005234. https://doi.org/10.1371/journal.pcbi.1005234
[56] Benjamin C. Koenig and Sili Deng. 2025. Kolmogorov-Arnold Chemical Reaction Neural Networks for Learning Pressure-Dependent Kinetic Rate Laws. https://doi.org/10.48550/ARXIV.2511.07686
[57] J. Koza, William Mydlowec, Guido Lanza, Yu Jessen, and Martin A. Keane. 2001. Reverse Engineering and Automatic Synthesis of Metabolic Pathways from Observed Data Using Genetic Programming. *Biocomputing 2001* (2001), 434–445. https://doi.org/10.1142/9789814447362_0043
[58] J. Koza, William Mydlowec, Guido Lanza, Jessen Yu, and M. A. Keane. 2004. Reverse Engineering by Means of Genetic Programming of Metabolic Pathways from Observed Data. In *International Conference on Systems Biology*, Vol. n.a. Oral presentation.
[59] Justin Noah Kreikemeyer, Philipp Andelfinger, and Adelinde M. Uhrmacher. 2024. Towards Learning Stochastic Population Models by Gradient Descent. In *Proceedings of the 38th ACM SIGSIM Conference on Principles of Advanced*

Discrete Simulation. ACM, Atlanta GA USA, 88–92. https://doi.org/10.1145/3615979.3656058

[60] Justin N. Kreikemeyer, Kevin Burrage, and Adelinde M. Uhrmacher. 2024. Discovering Biochemical Reaction Models by Evolving Libraries. In *Computational Methods in Systems Biology*, Roberta Gori, Paolo Milazzo, and Mirco Tribastone (Eds.). Vol. 14971. Springer Nature Switzerland, Cham, 117–136. https://doi.org/10.1007/978-3-031-71671-3_10

[61] Damoun Langary and Zoran Nikoloski. 2019. Inference of Chemical Reaction Networks Based on Concentration Profiles Using an Optimization Framework. *Chaos: Interdiscip. J. Nonlinear Sci.* 29, 11 (Nov. 2019), 113121. https://doi.org/10.1063/1.5120598

[62] Paola Lecca, Daniele Morpurgo, Gianluca Fantaccini, Alessandro Casagrande, and Corrado Priami. 2012. Inferring Biochemical Reaction Pathways: The Case of the Gemcitabine Pharmacokinetics. *BMC Syst. Biol.* 6 (2012), 0–21. https://doi.org/10.1186/1752-0509-6-51

[63] Wei-Po Lee and Kung-Cheng Yang. 2008. Applying Intelligent Computing Techniques to Modeling Biological Networks from Expression Data. *Genomics Proteomics Bioinformatics* 6, 2 (June 2008), 111–120. https://doi.org/10.1016/S1672-0229(08)60026-1

[64] Fernando Lejarza and Michael Baldea. 2022. Discovering Governing Equations via Moving Horizon Learning: The Case of Reacting Systems. *AIChE J.* 68, 6 (June 2022), e17567. https://doi.org/10.1002/aic.17567

[65] Daniel F. Linder and Grzegorz A. Rempała. 2020. Synthetic Likelihood Method for Reaction Network Inference. *Math. Methods Appl. Sci.* 43, 18 (Dec. 2020), 10547–10568. https://doi.org/10.1002/mma.6631

[66] Wenyao Lyu and Federico Galvanin. 2025. DoE-SINDy: An Automated Framework for Model Generation and Selection in Kinetic Studies. *Comput. Chem. Eng.* 202 (2025), 0–14. https://doi.org/10.1016/j.compchemeng.2025.109265

[67] V. Manca and L. Marchetti. 2012. Solving Dynamical Inverse Problems by Means of Metabolic P Systems. *Biosystems* 109, 1 (July 2012), 78–86. https://doi.org/10.1016/j.biosystems.2011.12.006

[68] Niall M. Mangan, Steven L. Brunton, Joshua L. Proctor, and J. Nathan Kutz. 2016. Inferring Biological Networks by Sparse Identification of Nonlinear Dynamics. *IEEE Trans. Mol. Biol. Multi-Scale Commun.* 2, 1 (June 2016), 52–63. https://doi.org/10.1109/TMBMC.2016.2633265

[69] G Maria and Dwt Rippin. 1995. Note Concerning the Online Kinetics Identification by Using Parallel Short-Cut Recursive Exact Estimators and a Knowledge Kinetic Model-Data Bank - Exemplification for a Catalytic Batch Process. , n.a. pages.

[70] Julien Martinelli, Jeremy Grignard, Sylvain Soliman, Annabelle Ballesta, and François Fages. 2023. Reactmine: A Statistical Search Algorithm for Inferring Chemical Reactions from Time Series Data. arXiv:2209.03185

[71] Julien Martinelli, Jeremy Grignard, Sylvain Soliman, and François Fages. 2019. On Inferring Reactions from Data Time Series by a Statistical Learning Greedy Heuristics. In *Computational Methods in Systems Biology*, Luca Bortolussi and Guido Sanguinetti (Eds.), Vol. 11773. Springer International Publishing, Cham, 352–355. https://doi.org/10.1007/978-3-030-31304-3_25

[72] Alexander Mordvintsev, Ettore Randazzo, and Eyvind Niklasson. 2023. Differentiable Programming of Chemical Reaction Networks. arXiv:2302.02714

[73] Tibor Nagy, János Tóth, and Tamás Ladics. 2020. Automatic Kinetic Model Generation and Selection Based on Concentration versus Time Curves. *Int. J. Chem. Kinet.* 52, 2 (Feb. 2020), 109–123. https://doi.org/10.1002/kin.21335

[74] Emily Nieves, Raj Dandekar, and Chris Rackauckas. 2024. Uncertainty Quantified Discovery of Chemical Reaction Systems via Bayesian Scientific Machine Learning. *Front. Syst. Biol.* 4 (March 2024), 1338518. https://doi.org/10.3389/fsysb.2024.1338518

[75] M. S. Nobile. 2014. *Evolutionary Inference of Biological Systems Accelerated on Graphics Processing Units*. Ph. D. Dissertation. Italy, Milano.

[76] Marco S. Nobile, Daniela Besozzi, Paolo Cazzaniga, Dario Pescini, and Giancarlo Mauri. 2013. Reverse Engineering of Kinetic Reaction Networks by Means of Cartesian Genetic Programming and Particle Swarm Optimization. In *2013 IEEE Congress on Evolutionary Computation*. IEEE, Cancun, Mexico, 1594–1601. https://doi.org/10.1109/CEC.2013.6557752

[77] Marco S. Nobile, Davide Cipolla, Paolo Cazzaniga, and Daniela Besozzi. 2016. GPU-powered Evolutionary Design of Mass-Action-Based Models of Gene Regulation. In *Evolutionary Computation in Gene Regulatory Network Research* (1 ed.), Hitoshi Iba and Nasimul Noman (Eds.). Wiley, Hoboken, NJ, USA, 118–150. https://doi.org/10.1002/9781119079453.ch6

[78] Chris J. Oates. 2013. *Bayesian Inference for Protein Signalling Networks*. Ph. D. Dissertation. University of Warwick.

[79] openRxiv. 2026. bioRxiv.org - the preprint server for Biology. https://www.biorxiv.org/ Last Accessed: 2026-01-29.

[80] openRxiv. 2026. medRxiv.org - the preprint server for Health Sciences. https://www.medrxiv.org/ Last Accessed: 2026-01-29.

[81] Matthew J Page, Joanne E McKenzie, Patrick M Bossuyt, Isabelle Boutron, Tammy C Hoffmann, Cynthia D Mulrow, Larissa Shamseer, Jennifer M Tetzlaff, Elie A Akl, Sue E Brennan, et al. 2021. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Systematic Reviews* 10, Article 89 (2021), 11 pages. https://doi.org/10.1186/s13643-021-01626-4

[82] Wei Pan, Ye Yuan, Jorge Goncalves, and Guy-Bart Stan. 2012. Reconstruction of Arbitrary Biochemical Reaction Networks: A Compressive Sensing Approach. In *2012 IEEE 51st Annual Conference on Decision and Control (CDC) (Proceedings of the IEEE Conference on Decision and Control)*. IEEE, Maui, HI, USA, 2334–2339. https://doi.org/10.1109/CDC.2012.6426216

[83] Wei Pan, Ye Yuan, Lennart Ljung, Jorge Goncalves, and Guy-Bart Stan. 2015. Identifying Biochemical Reaction Networks from Heterogeneous Datasets. In *2015 54th IEEE Conference on Decision and Control (CDC)*. IEEE, Osaka, 2525–2530. https://doi.org/10.1109/CDC.2015.7402596

[84] Wei Pan, Ye Yuan, Lennart Ljung, Jorge Goncalves, and Guy-Bart Stan. 2018. Identification of Nonlinear State-Space Systems from Heterogeneous Datasets. *IEEE Trans. Control Netw. Syst.* 5, 2 (June 2018), 737–747. https://doi.org/10.1109/TCNS.2017.2758966

[85] Arun Pankajakshan, Solomon Gajere Bawa, Asterios Gavriilidis, and Federico Galvanin. 2023. Autonomous Kinetic Model Identification Using Optimal Experimental Design and Retrospective Data Analysis: Methane Complete Oxidation as a Case Study. *React. Chem. Eng.* 8, 12 (2023), 3000–3017. https://doi.org/10.1039/D3RE00156C

[86] Perplexity AI, Inc. 2026. Perplexity Search. https://www.perplexity.ai/ Last Accessed: 2026-01-29.

[87] Christopher Polzak, Alejandro Lozano, Min Woo Sun, James Burgess, Yuhui Zhang, Kevin Wu, and Serena Yeung-Levy. 2025. Can Large Language Models Match the Conclusions of Systematic Reviews? arXiv:2505.22787

[88] Siddharth Prabhu, Nick Kosir, Mayuresh V. Kothare, and Srinivas Rangarajan. 2025. Derivative-Free Domain-Informed Data-Driven Discovery of Sparse Kinetic Models. *Ind. Eng. Chem. Res.* 64, 5 (Feb. 2025), 2601–2615. https://doi.org/10.1021/acs.iecr.4c02981

[89] Research Square Company. 2026. Research Square. https://www.researchsquare.com/ Last Accessed: 2026-01-29.

[90] Melissa L Rethlefsen, Shona Kirtley, Siw Waffenschmidt, Ana Patricia Ayala, David Moher, Matthew J Page, and Jonathan B Koffel. 2021. PRISMA-S: an extension to the PRISMA statement for reporting literature searches in systematic reviews. *Systematic reviews* 10, 1 (2021), 39. https://doi.org/10.1186/s13643-020-01542-z

[91] Pierre Ronceray. 2024. *Learning Dynamical Models from Stochastic Trajectories*. Ph. D. Dissertation. Aix-Marseille Université, Marseille, France.

[92] Michael Samoilov, Adam Arkin, and John Ross. 2001. On the Deduction of Chemical Reaction Pathways from Measurements of Time Series of Concentrations. *Chaos: Interdiscip. J. Nonlinear Sci.* 11, 1 (March 2001), 108–114. https://doi.org/10.1063/1.1336499

[93] Fadil Santosa and Benjamin Weitz. 2011. An Inverse Problem in Reaction Kinetics. *J. Math. Chem.* 49, 8 (Sept. 2011), 1507–1520. https://doi.org/10.1007/s10910-011-9835-2

[94] David Schaich, Ralf Becker, and Rudibert King. 2001. Qualitative Modelling for Automatic Identification of Mathematical Models of Chemical Reaction Systems. *Control Eng. Pract.* 9, 12 (2001), 1373–1381. https://doi.org/10.1016/S0967-0661(01)00080-6

[95] David Schaich and Rudibert King. 1999. Qualitative Modelling and Simulation of Chemical Reaction Systems. *Comput. Chem. Eng.* 23, SUPPL. 1 (1999), S415 – S418. https://doi.org/10.1016/S0098-1354(99)80102-1

[96] Michael Schmidt. 2011. *Machine Science: Automated Modeling of Deterministic and Stochastic Dynamical Systems*. Ph. D. Dissertation. Cornell University, Ithaca, New York, USA. hdl:1813/33574

[97] Michael D Schmidt, Ravishankar R Vallabhajosyula, Jerry W Jenkins, Jonathan E Hood, Abhishek S Soni, John P Wikswo, and Hod Lipson. 2011. Automated Refinement and Inference of Analytical Models for Metabolic Networks. *Phys. Biol.* 8, 5 (Oct. 2011), 55011. https://doi.org/10.1088/1478-3975/8/5/055011

[98] Dominic P. Searson, Samantha C. Burnham, Mark J. Willis, and Allen R. Wright. 2006. Identification of Chemical Reaction Mechanism from Batch Process Data. In *Proceedings of the 17th IASTED International Conference on Modelling and Simulation (MS'06)*. ACTA Press, USA, 511–516.

[99] Dominic P Searson, Mark J Willis, Simon J Horne, and Allen R Wright. 2007. Inference of Chemical Reaction Networks Using Hybrid S-system Models. *Chem. Prod. Process Model.* 2, 1 (March 2007), 0–23. https://doi.org/10.2202/1934-2659.1029

[100] Miguel Ángel de Carvalho Servia, Ilya Orson Sandoval, King Kuok Mimi Hii, Klaus Hellgardt, Dongda Zhang, and Ehecatl Antonio del Rio Chanona. 2025. Constraint-Guided Symbolic Regression for Data-Efficient Kinetic Model Discovery. arXiv:2507.02730

[101] Glenn Skrzypczak. 2025. *Benchmarking von Methoden zum Lernen von Chemischen Reaktionsnetzwerken aus Zeitreihendaten*. Master's thesis. University of Rostock. http://eprints.mosi.informatik.uni-rostock.de/846/

[102] Sriniketh Srinivasan, Julien Billeter, and Dominique Bonvin. 2019. Sequential Model Identification of Reaction Systems—the Missing Path between the Incremental and Simultaneous Approaches. *AIChE J.* 65, 4 (April 2019), 1211–1221. https://doi.org/10.1002/aic.16530

[103] Kansuporn Sriyudthsak, Fumihide Shiraishi, and Masami Yokota Hirai. 2013. Identification of a Metabolic Reaction Network from Time-Series Data of Metabolite Concentrations. *PLOS One* 8, 1 (Jan. 2013), e51212. https://doi.org/10.1371/journal.pone.0051212

[104] Masahiro Sugimoto, Shinichi Kikuchi, and Masaru Tomita. 2005. Reverse Engineering of Biochemical Equations from Time-Course Data by Means of Genetic Programming. *Biosystems* 80, 2 (May 2005), 155–164. https://doi.org/10.1016/j.biosystems.2004.11.003

[105] David Thorsley and Eric Klavins. 2012. Estimation and Discrimination of Stochastic Biochemical Circuits from Time-Lapse Microscopy Data. *PLOS One* 7, 11 (Nov. 2012), e47151. https://doi.org/10.1371/journal.pone.0047151

[106] Raúl E. Valdés-Pérez. 1992. Algorithm to Generate Reaction Pathways for Computer-Assisted Elucidation. *J. Comput. Chem.* 13, 9 (Nov. 1992), 1079–1088. https://doi.org/10.1002/jcc.540130906

[107] Ruben Van De Vijver, Nick M. Vandewiele, Aäron G. Vandeputte, Kevin M. Van Geem, Marie-Françoise Reyniers, William H. Green, and Guy B. Marin. 2015. Rule-Based Ab Initio Kinetic Model for Alkyl Sulfide Pyrolysis. *Chem. Eng. J.* 278 (Oct. 2015), 385–393. https://doi.org/10.1016/j.cej.2014.10.067

[108] Robin van der Wall. 2025. Vergleich von Methoden zur kombinierten Struktur- und Parameterinferenz von Reaktionsmodellen. Bachelor's Thesis. http://eprints.mosi.informatik.uni-rostock.de/829/

[109] Manokaran Veeramani, Sreeja Shanmuga Doss, Sridharakumar Narasimhan, and Nirav Bhatt. 2024. Semi-Supervised Machine Learning Approach for Reaction Stoichiometry and Kinetic Model Identification Using Spectral Data from Flow Reactors. *React. Chem. Eng.* 9, 2 (2024), 355–368. https://doi.org/10.1039/D3RE00334E

[110] J. Wikswo, Michael D. Schmidt, J. Jenkins, J. D. Hood, and H. Lipson. 2010. Automated Probing and Inference of Analytical Models for Metabolic Network Dynamics. In *APS March Meeting 2010*, Vol. 55. American Physical Society, Portland, OR, USA, BAPS.2010.MAR.P10.12.

[111] Mark J. Willis and Moritz von Stosch. 2017. Simultaneous Parameter Identification and Discrimination of the Nonparametric Structure of Hybrid Semi-Parametric Models. *Comput. Chem. Eng.* 104 (2017), 366–376. https://doi.org/10.1016/j.compchemeng.2017.05.005

[112] Zujian Wu, Wei Pang, and George M. Coghill. 2015. An Integrated Qualitative and Quantitative Biochemical Model Learning Framework Using Evolutionary Strategy and Simulated Annealing. *Cognit. Comput.* 7, 6 (Dec. 2015), 637–651. https://doi.org/10.1007/s12559-015-9328-x

[113] Bin Yang, Yuehui Chen, and Qingfang Meng. 2009. Inference of Differential Equations for Modeling Chemical Reactions. In *Advances in Neural Networks – ISNN 2009*, Wen Yu, Haibo He, and Nian Zhang (Eds.). Vol. 5551. Springer Berlin Heidelberg, Berlin, Heidelberg, 1014–1023. https://doi.org/10.1007/978-3-642-01507-6_114

[114] Zhengfeng Zhang and Yan Zhou. 2019. A Linear-in-Parameters Genetic Programming Method for Chemical Kinetics System Identification. In *2019 2nd International Conference on Information Systems and Computer Aided Education (ICISCAE)*. IEEE, Dalian, China, 169–173. https://doi.org/10.1109/ICISCAE48440.2019.221611