Parameter estimation for models of chemical reaction networks from experimental data of reaction rates

Manvel Gasparyan[a,b], Arnout Van Messem[a,c,d] and Shodhan Rao[a,b]

[a]Center for Biosystems and Biotech Data Science, Ghent University Global Campus, Incheon, Republic of Korea; [b]Department of Data Analysis and Mathematical Modelling, Ghent University, Belgium; [c]Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Belgium. [d]Department of Mathematics, University of Liège, Liège, Belgium.

ABSTRACT

Mathematical models of chemical reaction networks are typically nonlinear ordinary differential equations that represent the dynamical behavior of the species participating in the corresponding reactions. For the purpose of precise mathematical modelling of chemical reaction networks, useful techniques for estimating their parameters from experimental data are necessary. In this manuscript, we propose a new parameter estimation method for enzymatic chemical reaction networks from time-series experimental data of reaction rates. The main idea is based on retrieving time-series data of the species' concentrations from the available experimental data of reaction rates by making use of parametric Bézier curves. The least squares method is applied to these retrieved data in order to determine the best-fitting values of the parameters in the corresponding mathematical model. Subsequently, we demonstrate the applicability of our parameter estimation method on three examples of enzymatic chemical reaction networks including a model of ryanodine receptor adaptation and a model of protein kinase cascades. We also address the issue of identifiability of chemical reaction network models from reaction rates.

KEYWORDS

Systems biology, Bottom-up modelling approach, System identification, Bézier curves, Least squares method.

1. Introduction

One of the most natural problems in systems biology is the parameter estimation problem of chemical reaction networks (CRNs) using observed data available from biological experiments. Typically, a mathematical model of a CRN contains a system of ordinary differential equations (ODEs) that describes the dynamics of the concentrations of its constituent species. The aforementioned dynamics strictly depend on the parameters involved in the corresponding mathematical model. These parameters are most of the time partially or even fully unknown and are often estimated from experimental data. Usually, the available experimental data for estimating the parameters are time-series, i.e., the data are collected at discrete time instants.

The bottom-up approach (see e.g. Edwards & Thiele, 2013; Mock, Chiblak, & Herold-Mende, 2014) is a modelling procedure widely used in a variety of research fields including systems biology. This approach uses the available experimental data to retrieve a complete mathematical model of a system. The bottom-up modelling approach in systems biology consists of four major steps. The first step is draft reconstruction, which is based on collecting time-series data from biological experiments. The second step is the manual curation of the collected data by, for example, inserting missing values and eliminating irrelevant data. In the third step, the knowledge about biological interactions occurring in the CRN is translated into a mathematical formulation. In the final step, the parameters involved in this mathematical formulation are numerically estimated from experimental time-series data leading to a complete mathematical model.

The last step of the bottom-up approach, i.e., the parameter estimation, can be carried out by various methods, and the choice of the most suitable technique usually depends on the type of the observed

CONTACT Shodhan Rao. Email: Shodhan.Rao@ghent.ac.kr

time-series data available from experiments. Numerous parameter estimation techniques from time-series observed data of the species' concentrations are available in the literature. In case all the species' concentrations can be measured in an experimental setup, a common approach for identifying the parameters of the corresponding mathematical model is the least squares method (see e.g. Himmelblau & Riggs, 2012; Ljung, 1999). This optimization technique minimizes the sum of squared residuals, i.e., the sum of the squared differences between the observed experimental concentration values and the corresponding model-predicted values. A very efficient identification technique for non-linear mathematical models of CRNs is the quasi-linearization procedure described in Donnelly and Quon (1970), which is a second-order iterative procedure that considers the non-linear initial value problem (IVP) as the limit of a sequence of linear IVPs. For a comprehensive review of available techniques we refer to Crampin, Schnell, and McSharry (2004); Loskot, Atitey, and Mihaylova (2019); Ross, Schreiber, and Vlad (2005). Some of the well-known approaches for genome-scale bio-CRNs, such as maximum likelihood estimation, finite differences, etc., have been addressed in Fröhlich, Kaltenbacher, Theis, and Hasenauer (2017).

As we have already discussed above, various mathematical techniques have addressed the parameter estimation problem for CRNs from time-series observed data of the species' concentrations. However, in most of the cases not all concentrations can be measured experimentally. This data incompleteness makes the parameter estimation problem more challenging, mathematically as well as computationally. In reality, the output of a biological experiment is generally a combination of partial measurements corresponding to the species' concentrations and reaction rates, meaning that only some of the concentrations and some of the reaction rates can be measured in an experimental setup. The parameter estimation problem for CRNs from this kind of experimental data is an important subject that has not yet been considered.

There are various methods of measuring reaction rates known in the literature, such as the classical metabolic flux analysis without tracer experiments (Antoniewicz, 2015), flux balance analysis (Orth, Thiele, & Palsson, 2010), an electrochemical method proposed in Hodges and Chatelier (2002), a computational method called maximum metabolic flexibility (Megchelenbrink, Rossell, Huynen, Notebaart, & Marchiori, 2015), and a number of constraint-based reconstruction and analysis methods (Heirendt et al., 2019). However, most of these methods are indirect and they require a few key assumptions (e.g., the intracellular and the extracellular metabolite fluxes of a cell balance each other) and model simplifications (e.g., removal of futile metabolic cycles). The state-of-the-art method for precisely measuring reaction rates in biological systems is the ^{13}C-metabolic flux analysis (^{13}C-MFA), see for example Antoniewicz (2018); Guo, Sheng, and Feng (2016); Long and Antoniewicz (2019); Sauer (2006); Wiechert (2001). It consists of six major steps. The first step is the identification of optimal isotopic tracers that provide suitable resolution of fluxes. In the second step, isotopic labeling measurements are collected when isotopic steady state is reached. In the third step, both isotopic labeling and external rates are measured from the data collected from tracer experiments. It can be done by, for example, gas chromatography-mass spectrometry (Sparkman, Penton, & Kitson, 2011), which is the most widely used technique. The fourth step is metabolic model construction, which can be performed by using information from different databases. In the fifth step, minimization of the differences between the observed and simulated measurements results in the measurement of fluxes. Finally, a statistical analysis is performed to validate the measurements and determine the corresponding confidence intervals.

In this manuscript, we consider the parameter estimation problem for CRNs with non-autocatalytic reactions from observed experimental time-series data of the reaction rates corresponding to the CRN. To the best of our knowledge, there exists no general solution to this problem in the literature. We are mainly interested in CRNs governed by enzyme kinetic rate laws, which are the governing laws of most real-life bio-CRNs. We assume that certain measurements of the reaction rates are available at discrete time instants, which are not necessarily equidistant. Given the structure of the considered CRN, we first derive a new system of ODEs with dependent variables being the corresponding reaction rates. We use Bézier curves to approximate the vector of species' concentrations in the newly obtained system of ODEs. This approximation combined with the available experimental time-series data of reaction rates allows us to retrieve time-series data of the vector of species' concentrations. These data are then used to estimate the system parameters. Because of its simplicity and computational feasibility, we apply the least squares method to determine the best-fitting parameter values. The entire procedure has been automated and the corresponding Matlab library is provided as supplementary material.

Before parameter estimation is carried out, it is important to understand whether the parameters are identifiable given a certain type of output, i.e., there is a unique parameter vector corresponding to the given output. The uniqueness of the parameter vector depends on the structure of the mathematical model of the considered CRN and the type of the given output. We consider the identifiability of mathematical models of enzymatic CRNs given linear combinations of reaction rates as output.

We demonstrate our parameter estimation procedure for three examples of CRNs, which includes two real-life examples retrieved from the BioModels database (Malik-Sheriff et al., 2020), namely a model of ryanodine receptor adaptation (Keizer & Levine, 1996) and a model of protein kinase cascades (Markevich, Hoek, & Kholodenko, 2004) and an artificial example of a CRN that is identifiable given the reaction rates as output. For each of these examples, we show how to estimate the parameters contained in the corresponding mathematical model from time-series experimental data of overall outgoing reaction rates in the forward direction.

2. Preliminaries

In this section, we give a compact description of the mathematical techniques that are necessary for our parameter estimation method. We first explain a framework for deriving a system of ODEs that describes the dynamics of the concentrations of constituent species of a given CRN. Finally, we recall the Bézier curve defined by a given set of control points, which can be used to effectively approximate any continuous function.

2.1. Notations and useful identities

We commence by introducing the notations that are used throughout the manuscript. For any vector $v \in \mathbb{R}^m$, v_i denotes its i^{th} element, i.e., $v = [v_i]_{i=1}^m$. Denote by $\Gamma_v = \text{diag}(v)$ the $m \times m$ diagonal matrix, whose diagonal entries are the elements of the vector v. M_{ij} denotes the entry of the matrix M corresponding to the i^{th} row and the j^{th} column. Denote by I_m the $m \times m$ identity matrix. Define the vector-valued function $\text{Exp} : \mathbb{R}^m \to \mathbb{R}^m_+$ as $[x_i]_{i=1}^m \mapsto [e^{x_i}]_{i=1}^m$, and the vector-valued function $\text{Ln} : \mathbb{R}^m_+ \to \mathbb{R}^m$ as $[x_i]_{i=1}^m \mapsto [\ln(x_i)]_{i=1}^m$. For a vector-valued differentiable function $f : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^p$ given as $(x, y) \mapsto f(x, y)$, denote the $p \times m$ partial Jacobian matrix $\mathbf{J}_x(f)$ with respect to its first variable x with $\mathbf{J}_x(f) = \left[\frac{\partial f_i}{\partial x_j}\right]_{ij}$. Similarly, denote the partial Jacobian matrix of f with respect to its second variable y with $\mathbf{J}_y(f) = \left[\frac{\partial f_i}{\partial y_j}\right]_{ij}$.

We continue with two straightforward identities that will be used to prove the main proposition of this manuscript.

Lemma 2.1. *Let $x, y \in \mathbb{R}^m$ be two vector valued functions of a single variable t. We then have the following identities:*

(1) $Ln\,(\Gamma_x y) = Ln(x) + Ln(y)$, *(2)* $\frac{d}{dt} Ln(x) = \Gamma_x^{-1} \frac{dx}{dt}$.

Proof. Simplification of the left-hand side of the first identity leads to:

$$\text{Ln}\,(\Gamma_x y) \;=\; \text{Ln}\left(\begin{bmatrix} x_1 & 0 & \cdots & 0 \\ 0 & x_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & x_m \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}\right) = \text{Ln}\left(\begin{bmatrix} x_1 y_1 \\ x_2 y_2 \\ \vdots \\ x_m y_m \end{bmatrix}\right) = \text{Ln}(x) + \text{Ln}(y).$$

For the second identity we have:

$$\frac{d}{dt}\mathrm{Ln}(x) = \begin{bmatrix} \frac{d}{dt}\ln(x_1) \\ \frac{d}{dt}\ln(x_2) \\ \vdots \\ \frac{d}{dt}\ln(x_m) \end{bmatrix} = \begin{bmatrix} \frac{1}{x_1}\frac{dx_1}{dt} \\ \frac{1}{x_2}\frac{dx_2}{dt} \\ \vdots \\ \frac{1}{x_m}\frac{dx_m}{dt} \end{bmatrix} = \begin{bmatrix} \frac{1}{x_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{x_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{x_m} \end{bmatrix} \frac{dx}{dt} = \Gamma_x^{-1}\frac{dx}{dt}.$$

Hence, the lemma is proved.

□

2.2. Mathematical models of chemical reaction networks

We describe how to obtain the stoichiometric representation of the balance laws corresponding to a given CRN. Assume that s chemical species X_i, $i = 1, \ldots, s$, are participating in r unidirectional reactions $\mathcal{R}_j$, $j = 1, \ldots, r$. The left-hand and the right-hand sides of the reaction $\mathcal{R}_j$ are called its *substrate* and *product* complexes, respectively. The relation between the species and the reactions is given by an $s \times r$ *stoichiometric* matrix S defined as follows. Let α_{ij} and β_{ij} be the number of moles of species X_i in the substrate and the product complexes of the reaction $\mathcal{R}_j$, respectively. The entries of the stoichiometric matrix are then defined as $S_{ij} = \beta_{ij} - \alpha_{ij}$.

Example 2.1. *In order to clearly illustrate the modelling procedure we demonstrate it for the following example of a CRN.*

$$\mathcal{R}_{1,2}:\ 3X_1 + X_2 \rightleftharpoons 2X_3, \qquad\qquad \mathcal{R}_3:\ X_3 + X_4 \longrightarrow X_5. \tag{1}$$

There are five chemical species participating in three unidirectional reactions. Note that the second reaction can be considered as the reverse reaction of the first one. Also note that in our modelling procedure, we split each reversible reaction into two unidirectional reactions. Therefore, the 5×3 *stoichiometric matrix of the CRN* (1) *is:*

$$S = \begin{bmatrix} -3 & 3 & 0 \\ -1 & 1 & 0 \\ 2 & -2 & -1 \\ 0 & 0 & -1 \\ 0 & 0 & 1 \end{bmatrix}.$$

Let $x \in \mathbb{R}^s_+$ be the vector of concentrations of the species participating in the CRN. Denote by $\nu \in \mathbb{R}^r$ the vector of reaction rates, which is a function of x and depends on the governing laws of the CRN. The basic structure describing the dynamics of the species' concentration vector is given by the stoichiometric representation of the *balance laws* as:

$$\frac{dx}{dt} = S\nu. \tag{2}$$

Example 2.1 (Continued)**.** *The ODEs corresponding to the CRN* (1) *can then be written as:*

$$\frac{dx_1}{dt} = -3\nu_1 + 3\nu_2, \qquad \frac{dx_3}{dt} = 2\nu_1 - 2\nu_2 - \nu_3, \qquad \frac{dx_5}{dt} = \nu_3.$$
$$\frac{dx_2}{dt} = -\nu_1 + \nu_2, \qquad \frac{dx_4}{dt} = -\nu_3,$$

2.3. Enzyme kinetics rate laws

Any CRN can be uniquely described by the ODE (2), independent of its governing laws. We are interested in CRNs that are governed by an *enzyme kinetics rate law* (EKRL). As explained in Gasparyan, Van Messem, and Rao (2018, 2020); Rao, van der Schaft, van Eunen, Bakker, and Jayawardhana (2014), according to this rate law, the reaction rate of the j^{th} reaction, $j = 1, \ldots, r$, is given in the following form:

$$\nu_j = k_j g_j(x, K) \prod_{i=1}^{s} x_i^{\alpha_{ij}}, \tag{3}$$

where k_j is the rate constant of the j^{th} reaction, $K \in \mathbb{R}_+^q$ is the vector of *Michaelis constants* of the species, $g_j : \mathbb{R}_+^s \times \mathbb{R}_+^q \to \mathbb{R}_+$ is a rational function of its arguments, and α_{ij} is the number of moles of the species X_i in the substrate of the j^{th} reaction. In this case, there are two different types of parameters, namely the rate constants $k \in \mathbb{R}_+^r$ of the reactions and the Michaelis constants $K \in \mathbb{R}_+^q$ of the species.

Example 2.1 (Continued)**.** *For the CRN given in* (1)*, a possibility of reaction rates is:*

$$\nu_1 = \frac{k_1 x_1^3 x_2}{1 + \frac{x_1}{K_1} + \frac{x_2}{K_2}}, \qquad \nu_2 = \frac{k_2 x_3^2}{1 + \frac{x_1}{K_1} + \frac{x_2}{K_2}}, \qquad \nu_3 = \frac{k_3 x_3 x_4}{1 + \frac{x_3}{K_3} + \frac{x_4}{K_4}}.$$

In this case, the vector of reaction rates can be written in the form (3) *by denoting the rational terms in the expressions of the reaction rates as:*

$$g_1(x, K) = \frac{1}{1 + \frac{x_1}{K_1} + \frac{x_2}{K_2}}, \qquad g_2(x, K) = \frac{1}{1 + \frac{x_1}{K_1} + \frac{x_2}{K_2}}, \qquad g_3(x, K) = \frac{1}{1 + \frac{x_3}{K_3} + \frac{x_4}{K_4}}.$$

Note that, if the considered CRN (governed by EKRL) involves competitive, uncompetitive and non-competitive modifiers, the corresponding reaction rates can still be written in the form (3). This is because the terms involving the concentrations of modifiers appear in the rational terms corresponding to the reaction rates. In the supplementary material, we provide a detailed explanation of some well known EKRLs that can be written in the form (3).

A special case of an EKRL is the *mass action kinetics rate law* (MAKRL), which assumes that the rate of a reaction is proportional to the concentrations of the species participating in its substrate. In other words, if some of the reactions are governed by MAKRL, their reaction rates can be written in the form (3) with corresponding rational functions being identically equal to one. In this case, the only parameters contained in the corresponding mathematical model are the rate constants.

Example 2.1 (Continued)**.** *If all the reactions of the CRN given in* (1) *are governed by MAKRL, according to this law the corresponding reaction rates are given by:*

$$\nu_1 = k_1 x_1^3 x_2, \qquad \nu_2 = k_2 x_3^2, \qquad \nu_3 = k_3 x_3 x_4.$$

Next, we obtain an expression for the vector of reaction rates $\nu \in \mathbb{R}^r$ using equation (3) in terms of matrix multiplication. We consider CRNs with non-autocatalytic reactions meaning that the substrate and the product complexes of the same reaction do not share a common species. Define the $s \times r$ *substrate composition* matrix Δ of the CRN as follows:

$$\Delta_{ij} = \begin{cases} -S_{ij}, \text{ if } S_{ij} \leq 0 \\ 0, \text{ otherwise} \end{cases}. \tag{4}$$

For $\Gamma_k = \operatorname{diag}(k) \in \mathbb{R}^{r \times r}$ and $\Gamma_g = \operatorname{diag}(g) \in \mathbb{R}^{r \times r}$, define the *substrate expression* function $\varphi : \mathbb{R}^s \to \mathbb{R}^r$ as $x \mapsto \operatorname{Exp}(\Delta^\top \operatorname{Ln}(x))$. Then observe that

$$\nu = \Gamma_k \Gamma_g \varphi(x). \tag{5}$$

In this case, the ODEs given in (2) can be rewritten as (see Gasparyan et al., 2020; Rao et al., 2014):

$$\frac{dx}{dt} = S\Gamma_k\Gamma_g \mathrm{Exp}\left(\Delta^\top \mathrm{Ln}(x)\right).$$

Note that, if the considered CRN is governed by MAKRL, then $\Gamma_g \equiv I_r$.

Example 2.1 (Continued)**.** *With reference to the CRN* (1)*, the vector of reaction rates can be written in the form* (5)*, with the substrate composition matrix* $\Delta \in \mathbb{R}^{5\times 3}$ *and the complex expression function* $\varphi : \mathbb{R}^5 \to \mathbb{R}^3$ *given by:*

$$\Delta = \begin{bmatrix} 3 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \qquad \varphi(x) = \begin{bmatrix} x_1^3 x_2 \\ x_3^2 \\ x_3 x_4 \end{bmatrix}.$$

2.4. Bézier curves

A *Bézier curve* (Bézier & Sioussiou, 1983; Prautzsch, Boehm, & Paluszny, 2002) is a parametric curve defined by a set of given control points. It plays a crucial role in a number of research fields, such as computer graphics, user interface design, animation, and vector graphics (see e.g. Bézier, 1986; Farin, 1993). A Bézier curve is described as a linear combination of *Bernstein basis polynomials* (Bernštein, 1912) as follows. For a positive integer N, let $F = \{F_m | m = 0, \ldots, N\}$ be a set of given distinct (control) points. The Bézier curve defined by these points is uniquely expressed as:

$$\beta(t|F) = \sum_{m=0}^{N} b_{N,m}(t) F_m, \quad 0 \le t \le 1,$$

where $b_{N,m}(t) = \binom{N}{m} t^m (1-t)^{N-m}$, $m = 0, \ldots, N$, are the Bernstein basis polynomials of degree N. Note that the first and the last control points F_0 and F_N are the endpoints of the Bézier curve, i.e., $\beta(0|F) = F_0$ and $\beta(1|F) = F_N$. However, the curve generally does not pass through the intermediate points F_m, $m = 1, \ldots, N-1$.

In mathematics, Bézier curves are used to uniformly approximate any continuous function. As explained in Berry and Patterson (1997), the control points are uniquely defined for every Bézier curve.

Note that the domain of a Bézier curve is restricted to $[0, 1]$. Since in our case the dependent variable t represents time, we need to re-define the curve over the interval $[0, T]$, with $T > 0$. We therefore define the *contracted Bézier curve* $\bar{\beta}(t|F)$ by a set of control points F over the interval $[0, T]$ as:

$$\bar{\beta}(t|F) = \beta\left(\frac{t}{T}\middle| F\right), \quad 0 \le t \le T, \tag{6}$$

Similar to the regular Bézier curve, the first and the last control points F_0 and F_N correspond to the endpoints of the contracted Bézier curve, i.e., $\bar{\beta}(0|F) = F_0$ and $\bar{\beta}(T|F) = F_N$. An example of a contracted Bézier curve defined by a set of seven given control points over a pre-specified interval is included in the supplementary material. Note that simplification of the right-hand side of (6) leads to:

$$\bar{\beta}(t|F) = \frac{1}{T^N} \sum_{m=0}^{N} \bar{b}_{N,m}(t) F_m, \quad 0 \le t \le T,$$

with coefficient polynomials being defined as $\bar{b}_{N,m}(t) = \binom{N}{m} t^m (T-t)^{N-m}$, $m = 0, \ldots, N$.

3. Automated parameter estimation procedure

In this section, we describe the main contribution of this manuscript, i.e., a novel method for estimating the parameters contained in the mathematical model (2) from observed time-series experimental data of reaction rates.

3.1. Problem statement: available experimental data

Typically, biological experiments provide measurements of species' concentrations or reaction rates. In this manuscript, we assume that in an experimental setup reaction rates are measured. Suppose that the output of a biological experiment is of the form:

$$y = H\nu, \tag{7}$$

where $H \in \mathbb{R}^{l\times r}$ is a constant matrix. In general, H can have any arbitrary structure. However, for our parameter estimation method we assume that in an experimental setup overall rates of reversible reactions (in a certain direction) and the rates of irreversible reactions are measured. Thus, the matrix H has the following structure:

- each row of H corresponding to an irreversible reaction has only one non-zero element, which is equal to one and is placed in the position corresponding to the particular reaction;
- each row corresponding to a reversible reaction has exactly two non-zero elements, which are one and negative one and are placed in the positions corresponding to their particular reactions.

For example, with reference to the CRN (1) we have:

$$H = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Further assume that we have n different sets of time-series data of y collected from biological experiments on the same CRN. For $i = 1, \ldots, l$; $j = 1, \ldots, n$ and $m = 0, \ldots, N_j$ let $y_{i,m}^{(j)}$ be the observed value of the i^{th} output at time instant $t_m^{(j)}$ corresponding to the j^{th} set of data. For compactness we consider the following observed experimental time-series *data sets*:

$$\Lambda^{(j)} = \left\{ \left(t_m^{(j)}, y_{i,m}^{(j)} \right) \mid i = 1, \ldots, l,\ m = 0, \ldots, N_j \right\},\ j = 1, \ldots, n. \tag{8}$$

We aim to identify the best-fitting parameters corresponding to the mathematical model (2) of the considered CRN from observed time-series data collected from experiments. Recall that if the considered CRN is governed by an EKRL which is not MAKRL, then there are two types of parameters contained in the corresponding mathematical model, namely the rate constants $k \in \mathbb{R}_+^r$ and the Michaelis constants $K \in \mathbb{R}_+^q$. However, if the governing law of the given CRN is MAKRL, then the only parameters of the corresponding mathematical model are the rate constants. We therefore consider the parameter estimation procedure for each governing law separately. For each of these cases, we develop an automated process which uses the inputs provided by the user to determine the best-fitting parameter values in a fully automated manner.

3.1.0.1. Inputs: The list of inputs required for our parameter estimation procedure is:

- the $s \times r$ stoichiometric matrix S of the considered CRN, where s and r are the number of the species and the reactions, respectively;
- the vector-valued function $g : \mathbb{R}_+^s \times \mathbb{R}_+^q \mapsto \mathbb{R}_+^r$, which is the vector of the rational terms in the expressions of the reaction rates;
- the observed time-series experimental data $\Lambda^{(j)}$, $j = 1, \ldots, n$, of reaction rates given in (8).

3.1.0.2. Output: The output of our automated parameter estimation method are the best-fitting parameter values corresponding to the experimental time-series' observed data (8).

3.2. New mathematical formulation

We now present an intermediate result that is crucial in our parameter estimation method: we show how to derive an IVP with reaction rates as dependent variables.

Proposition 3.1. *Suppose that a chemical reaction network is governed by an enzyme kinetic rate law, i.e., the corresponding mathematical model is given by* (2) *with the vector of unidirectional reaction rates* (5)*. Then the vector of unidirectional reaction rates satisfies the following initial value problem:*

$$\begin{cases} \dfrac{d\nu}{dt} = \Gamma_\nu \Gamma_g^{-1} \mathbf{J}_x(g) S\nu + \Gamma_\nu \Delta^\top \Gamma_x^{-1} S\nu \\ \nu(0) = \nu_0 \end{cases}, \tag{9}$$

where $\nu_0 \in \mathbb{R}^r_+$ is the vector of initial unidirectional reaction rates.

Proof. We first find a formula for the derivative of $g : \mathbb{R}^s \times \mathbb{R}^q \to \mathbb{R}^r$ with respect to t:

$$\begin{aligned} \frac{dg(x,K)}{dt} &= \left[\frac{dg_i(x,K)}{dt}\right]_{i=1}^{r} = \left[\sum_{j=1}^{s} \frac{\partial g_i(x,K)}{\partial x_j}\frac{dx_j}{dt}\right]_{i=1}^{r} \\ &= \mathbf{J}_x(g)\frac{dx}{dt} = \mathbf{J}_x(g) S\nu, \end{aligned}$$

where we used equation (2). Now applying the logarithmic operator to (5) and using the first identity of Lemma 2.1, we get:

$$\mathrm{Ln}(\nu) = \mathrm{Ln}(k) + \mathrm{Ln}(g(x,K)) + \Delta^\top \mathrm{Ln}(x).$$

Subsequently, we differentiate both sides with respect to t and apply the second identity of Lemma 2.1:

$$\Gamma_\nu^{-1}\frac{d\nu}{dt} = \Gamma_g^{-1}\mathbf{J}_x(g)S\nu + \Delta^\top \Gamma_x^{-1}\frac{dx}{dt}.$$

Using (2) leads to:

$$\frac{d\nu}{dt} = \Gamma_\nu \Gamma_g^{-1}\mathbf{J}_x(g)S\nu + \Gamma_\nu \Delta^\top \Gamma_x^{-1} S\nu,$$

which completes the proof. □

One of the main advantages of the IVP (9) is that it does not explicitly depend on the rate constants $k \in \mathbb{R}^r_+$. Therefore, the only parameters involved in it are the Michaelis constants $K \in \mathbb{R}^q_+$ of the species. This is because the Michaelis constants appear in the vector $g \in \mathbb{R}^r$ of the rational terms in the expressions of the reaction rates.

Corollary 3.1. *Suppose that a chemical reaction network is governed by mass action kinetic rate law, i.e., the corresponding mathematical model is given by* (2) *where the vector of unidirectional reaction rates is given by* (5) *with $\Gamma_g \equiv I_r$. Then the vector of unidirectional reaction rates satisfies the following initial value problem:*

$$\begin{cases} \dfrac{d\nu}{dt} = \Gamma_\nu \Delta^\top \Gamma_x^{-1} S\nu \\ \nu(0) = \nu_0 \end{cases}, \tag{10}$$

where $\nu_0 \in \mathbb{R}^r_+$ is the vector of initial unidirectional reaction rates.

3.3. Parameter estimation for mass action kinetics rate law

In this part of the manuscript, we consider the parameter estimation problem from time-series experimental data sets (8) for CRNs governed by MAKRL.

For $i = 1, \ldots, s$; $j = 1, \ldots, n$ and $m = 0, \ldots, N_j$, let $x_{i,m}^{(j)} \in \mathbb{R}_+$ be the value of the i^{th} concentration $x_i \in \mathbb{R}_+$ at time instant $t_m^{(j)}$ corresponding to the j^{th} experiment. Since these values are unknown we treat them as parameters and the goal now becomes finding estimates for them from experimental time-series data $\Lambda^{(j)}$. We define the following parameter matrix:

$$\theta^{(j)} := \begin{bmatrix} x_{1,0}^{(j)} & x_{1,1}^{(j)} & \cdots & x_{1,N_j}^{(j)} \\ x_{2,0}^{(j)} & x_{2,1}^{(j)} & \cdots & x_{2,N_j}^{(j)} \\ \vdots & \vdots & \ddots & \vdots \\ x_{s,0}^{(j)} & x_{s,1}^{(j)} & \cdots & x_{s,N_j}^{(j)} \end{bmatrix} \in \mathbb{R}^{s \times (N_j+1)}. \tag{11}$$

Let $\theta_i^{(j)}$, $i = 1, \ldots, s$, denote the i^{th} row of the matrix $\theta^{(j)}$, which is the time-series (parametric) data of the concentration of the i^{th} species. Denote by $\gamma_i^{(j)}$ the contracted Bézier curve defined by the control points $\theta_i^{(j)}$ over the time-interval $[0, t_{N_j}^{(j)}]$, i.e., $\gamma_i^{(j)} = \bar{\beta}\left(t|\theta_i^{(j)}\right)$. We consider the vector $\gamma^{(j)} \in \mathbb{R}^s$ of the contracted Bézier curves corresponding to the species concentrations, i.e., $\gamma^{(j)} = \left[\gamma_i^{(j)}\right]_{i=1}^{s}$.

We use the aforementioned contracted Bézier curves to approximate the species' concentrations in the IVP (10) described in Corollary 3.1. More precisely, we consider the following IVP:

$$\begin{cases} \dfrac{d\nu}{dt} = \Gamma_\nu \Delta^\top \Gamma_{\gamma^{(j)}}^{-1} S\nu \\ \nu(0) = \nu_0^{(j)} \end{cases}. \tag{12}$$

Note that even though the vector $\nu_0^{(j)}$ of initial unidirectional reaction rates is not provided, it still can be assigned by solving the system of linear equations

$$H\nu_0^{(j)} = y_0^{(j)} \tag{13}$$

numerically for $\nu_0^{(j)}$. Also note that the solution to the above-mentioned system of linear equations is not unique if H does not have full column rank. In this case, we show later with the help of some examples that in general our procedure does not lead to a unique set of parameters, since there is an infinite number of possible initial reaction rates satisfying (13).

The only parameters involved in the system are the parameter values of the species' concentrations at distinct time instants given in (11). For every $j = 1, \ldots, n$, we estimate the values of these parameters, such that the data set $\Lambda^{(j)}$ fits the IVP (12). The aforementioned estimation problem can be treated in various ways. Because of its simplicity and computational feasibility, we use the well-known least squares method in this manuscript. Denote by $y\left(t, \theta^{(j)}\right) \in \mathbb{R}^l$ the model-predicted value of the output (7) corresponding to the IVP (12), i.e., the output of the system (12) as a function of $\theta^{(j)}$, where $t \in \left[0, t_{N_j}^{(j)}\right]$. Define the least squares objective function $\varepsilon\left(\theta^{(j)}\right)$ as follows:

$$\varepsilon\left(\theta^{(j)}\right) = \sum_{i=1}^{l} \sum_{m=1}^{N_j} \left(y_i\left(t_m^{(j)}, \theta^{(j)}\right) - y_{i,m}^{(j)}\right)^2.$$

The mathematical problem now becomes finding the values of $\theta^{(j)}$ for which $\varepsilon\left(\theta^{(j)}\right)$ is minimal. In other words, we consider the optimization problem:

$$\min_{\theta^{(j)}} \varepsilon\left(\theta^{(j)}\right).$$

Denote the solution to the above-mentioned optimization problem by $\widehat{\theta}^{(j)} \in \mathbb{R}^{s\times(N_j+1)}$. Finally, we determine the parameter values $k \in \mathbb{R}^r$, for which our retrieved time-series data $\widehat{\theta}^{(j)}$ fits the mathematical model (2). In other words, we have a parameter estimation problem from time-series data of species' concentrations, for which we may again use the least squares method. First, for $j = 1, \ldots, n$, we consider the system of ODEs (2) with the vector of initial concentrations being $x_0^{(j)} = \left[\widehat{x}_{i,0}^{(j)}\right]_{i=1}^{s}$, which is the first column of the retrieved data-matrix $\widehat{\theta}^{(j)}$. For $i = 1, \ldots, s$ and $m = 1, \ldots, N_j$, let $x_i^{(j)}(t, k)$, $t \in \left[0, t_{N_j}^{(j)}\right]$, be the i^{th} element of the solution-vector, as a function of k, of the aforementioned IVP corresponding to the j^{th} experiment. We define the least squares objective function as:

$$\rho(k) = \sum_{j=1}^{n}\sum_{i=1}^{s}\sum_{m=1}^{N_j}\left(x_i^{(j)}\left(t_m^{(j)}, k\right) - \widehat{x}_{i,m}^{(j)}\right)^2.$$

We obtain the best-fitting parameter values $\widehat{k} \in \mathbb{R}_+^r$ for which the least squares objective function $\rho(k)$ has a minimum value.

3.4. Parameter estimation for enzyme kinetics rate laws

We now consider the parameter estimation problem for CRNs governed by an EKRL, which is not MAKRL. Similar to the case of MAKRL, we want to estimate the vector of species' concentrations at different time instants using the time-series observed data $\Lambda^{(j)}$, $j = 1, \ldots, n$. In this case, we use the contracted Bézier curves to approximate the species' concentrations in the IVP (9). Recall that for a CRN governed by EKRL the corresponding mathematical model (2) contains two different types of parameters, namely, the rate constants of the reactions and the Michaelis constants.

One of the main differences with MAKRL is that the Michaelis constants are still involved in the IVP (9) obtained after extracting the rate constants from the mathematical model (2). The approximation of the vector of the species' concentrations in (9) with corresponding contracted Bézier curves results in the parameter sets K and θ, where the latter is defined as in (11). Hence, unlike the case of MAKRL, here we consider a global constrained least squares problem. We find the values of K and $\theta^{(j)}$ for which the datasets $\Lambda^{(j)}$ fit the IVP (9), in the sense of least-squares, with the additional constraint being the uniqueness of K, i.e., $K^{(j)} = K^{(j')}$ for every $j, j' \in \{1, \ldots, n\}$. Finally, using the obtained Michaelis constants and the retrieved time series data of species' concentrations we determine the best-fitting values of rate constants $\widehat{k} \in \mathbb{R}_+^r$ for the system of ODEs (2).

4. On identifiability of mathematical models

Consider a mathematical model (2) of a CRN, which is not necessarily governed by EKRL. Recall that the vector of reaction rates $\nu \in \mathbb{R}^r$ strictly depends on the vector of species' concentrations $x \in \mathbb{R}_+^s$ and the vector of parameters $\kappa \in \mathbb{R}_+^p$. We begin with the definition of identifiability of systems corresponding to a mathematical model of a CRN. With the help of simple examples we further demonstrate the non-identifiability of such systems, if the output corresponds to the overall reaction rates, in certain directions. In the last part of the section, we show that, under certain conditions, if a system with the species' concentrations as output is identifiable, so is the corresponding system with linear combinations of reaction rates as output.

For any given $\kappa \in \mathbb{R}_+^p$, let $\mathcal{Y}_\kappa$ denote the set of admissible output-trajectories corresponding to the system:

$$\begin{cases} \dfrac{dx}{dt} = S\nu \\ y = \psi(\kappa, x) \end{cases}, \tag{14}$$

where $\psi : \mathbb{R}_+^p \times \mathbb{R}_+^s \to \mathbb{R}^l$ is the *measurement function*. In the context of this manuscript the term *identifiability of the mathematical model* refers to the structural identifiability of the parameters κ

involved in it, i.e., the possibility of the parameters κ to be determined uniquely given $\mathcal{Y}_\kappa$. We define identifiability as follows:

Definition 4.1. *The mathematical model* (14) *is identifiable, if for all* $\kappa_1 \neq \kappa_2 \in \mathbb{R}^p_+$, *we have*

$$\mathcal{Y}_{\kappa_1} \neq \mathcal{Y}_{\kappa_2}. \tag{15}$$

Note that the above-mentioned definition of the concept of an identifiable system correlates with the one introduced in Bellman and Åström (1970). Different mathematical techniques have been used to address the identifiability of nonlinear dynamical systems (see for example Bellman & Åström, 1970; Bernard & Bastin, 2005; Cobelli & DiStefano 3rd, 1980; Craciun & Pantea, 2008; Farina et al., 2006; Miao, Xia, Perelson, & Wu, 2011; Stanhope, Rubin, & Swigon, 2014). These methods are also applicable for mathematical models of CRNs. In this manuscript, we are interested in studying the identifiability of the mathematical model (2) of a CRN given the overall outgoing reaction rates (in a certain direction) as the output.

4.1. Non-identifiability of mathematical models corresponding to overall reaction rates

Let $\mathcal{X}_k$ denote the set of admissible concentration trajectories corresponding to the system:

$$\begin{cases} \dfrac{dx}{dt} = S\nu \\ y = x \end{cases}. \tag{16}$$

There are various (non) identifiability results for the system (16) known in the literature. In this manuscript, we assume that the output of the mathematical model (2) is a linear combination of reaction rates. We therefore consider the identifiability of the system:

$$\begin{cases} \dfrac{dx}{dt} = S\nu \\ y = H\nu \end{cases}, \tag{17}$$

where $H \in \mathbb{R}^{l\times r}$ is the measurement matrix. With the following simple examples of CRNs we show that if there is a reversible reaction in the CRN for which only the overall reaction rate (in a certain direction) can be measured, then the system (17) is not necessarily identifiable. We conjecture that, in general, the system (17) is not identifiable if the matrix H does not have full column rank.

Example 4.1. *Consider the following reversible reaction occurring among two distinct chemical species* X_1 *and* X_2*:*

$$X_1 \underset{\nu_r}{\overset{\nu_f}{\rightleftharpoons}} X_2,$$

where ν_f *and* ν_r *are the rate functions of the forward and reverse reactions, respectively. For* $i = 1, 2$, *let* x_i *denote the concentration of* X_i. *Assume that the reactions are governed by MAKRL, i.e., the corresponding rate functions are given by*

$$\nu_f = k_f x_1, \qquad \nu_r = k_r x_2, \tag{18}$$

where k_f *and* k_r *are the rate constants of the forward and reverse reactions, respectively. Further assume that the output of the mathematical model* (2) *is the overall outgoing reaction rate in the forward direction:*

$$y = \nu_f - \nu_r = H\nu, \tag{19}$$

where $H = \begin{bmatrix} 1 & -1 \end{bmatrix}$ *and* $\nu = \begin{bmatrix} \nu_f & \nu_r \end{bmatrix}^\top$. *In this case, the ODEs* (2) *can be written as:*

$$\frac{dx_1}{dt} = -y, \qquad \frac{dx_2}{dt} = y. \tag{20}$$

The goal of this example is to show that the system (17) *is not identifiable, since the matrix* H *does not have full column rank. For this purpose, first differentiate* (19) *with respect to* t *and then substitute* (18) *and* (20)*, which results in:*

$$\frac{dy}{dt} + (k_f + k_r)y = 0. \tag{21}$$

Clearly, the parameters k_f *and* k_r *contained in* (21) *cannot be uniquely determined given the set of admissible output-trajectories of* (17)*, since for any two distinct vectors* $k^{(1)} = \begin{bmatrix} k_f^{(1)} & k_r^{(1)} \end{bmatrix}^\top$ *and* $k^{(2)} = \begin{bmatrix} k_f^{(2)} & k_r^{(2)} \end{bmatrix}^\top$*, such that*

$$k_f^{(1)} + k_r^{(1)} = k_f^{(2)} + k_r^{(2)},$$

the corresponding solutions to (21) *are identical, which also means that* $\mathcal{Y}_{k^{(1)}} = \mathcal{Y}_{k^{(2)}}$.

Example 4.2. *Suppose that four distinct chemical species* X_i, $i = 1, \ldots, 4$, *are participating in the following reversible reaction:*

$$X_1 + X_2 \underset{\nu_r}{\overset{\nu_f}{\rightleftharpoons}} X_3 + X_4,$$

where ν_f *and* ν_r *are the rate functions of the forward and reverse reactions, respectively. For* $i = 1, \ldots, 4$, *let* x_i *denote the concentration of* X_i. *Assume that the reactions are governed by MAKRL, i.e., the corresponding rate functions are given by*

$$\nu_f = k_f x_1 x_2, \qquad \nu_r = k_r x_3 x_4, \tag{22}$$

where k_f *and* k_r *are the rate constants of the forward and reverse reactions, respectively. Further assume that, similar to the previous example, the output of the mathematical model* (2) *is the overall outgoing reaction rate in the forward direction given in* (19). *In this case, the ODEs* (2) *can be written as:*

$$\frac{dx_1}{dt} = \frac{dx_2}{dt} = -y, \qquad \frac{dx_3}{dt} = \frac{dx_4}{dt} = y. \tag{23}$$

By differentiating (19) *with respect to* t *and then substituting* (22) *and* (23)*, we obtain:*

$$\frac{1}{y}\frac{dy}{dt} = -k_f(x_1 + x_2) - k_r(x_3 + x_4). \tag{24}$$

Differentiating (24) *with respect to* t *and substituting* (23) *results in:*

$$y\frac{d^2y}{dt^2} - \left[\frac{dy}{dt}\right]^2 - 2(k_f - k_r)y^3 = 0. \tag{25}$$

Similar to the previous example, the parameters k_f *and* k_r *contained in* (25) *cannot be uniquely determined from the admissible set* $\mathcal{Y}_k$ *of output-trajectories, since for any two distinct vectors* $k^{(1)} = \begin{bmatrix} k_f^{(1)} & k_r^{(1)} \end{bmatrix}^\top$ *and* $k^{(2)} = \begin{bmatrix} k_f^{(2)} & k_r^{(2)} \end{bmatrix}^\top$*, such that*

$$k_f^{(1)} - k_r^{(1)} = k_f^{(2)} - k_r^{(2)},$$

the corresponding solutions to (21) *are identical, i.e.,* $\mathcal{Y}_{k^{(1)}} = \mathcal{Y}_{k^{(2)}}$.

4.2. An equivalence result on identifiability

With the following theorem we show that, under certain conditions on the structure of the considered CRN and the measurement matrix of the system (17), the identifiability of the system (16) is equivalent to the identifiability of the system (17).

Theorem 4.1. *With reference to the systems* (16) *and* (17) *the following two statements hold.*

(1) Assume that H has full column rank. If the system (16) *is identifiable, then the system* (17) *is also identifiable.*
(2) Assume that S has full column rank. If the system (16) *is not identifiable, then the system* (17) *is also not identifiable.*

Proof.

(1) For any $k \in \mathbb{R}^p_+$, let $\mathcal{V}_\kappa$ denote the set of admissible reaction rate trajectories of (2). With reference to (17), assume that for any $\kappa_1, \kappa_2 \in \mathbb{R}^p_+$, we have $\mathcal{Y}_{\kappa_1} = \mathcal{Y}_{\kappa_2}$. Since H has full column rank, we obtain $\mathcal{V}_{\kappa_1} = \mathcal{V}_{\kappa_2}$. We therefore deduce from (2) that $\mathcal{X}_{\kappa_1} = \mathcal{X}_{\kappa_2}$. Since the system (16) is identifiable, we obtain $\kappa_1 = \kappa_2$. We conclude that the system (17) is identifiable.
(2) For any $k \in \mathbb{R}^p_+$, let $\mathcal{X}'_\kappa$ denote the admissible set of trajectories of the derivative of concentrations corresponding to (2). Suppose that the system (16) is not identifiable. Then there exist $\kappa_1 \neq \kappa_2 \in \mathbb{R}^p_+$, such that $\mathcal{X}_{\kappa_1} = \mathcal{X}_{\kappa_2}$. This implies that $\mathcal{X}'_{\kappa_1} = \mathcal{X}'_{\kappa_2}$. Since S has full column rank, we obtain from (2) that $\mathcal{V}_{\kappa_1} = \mathcal{V}_{\kappa_2}$. Therefore, the system (17) is not identifiable.

□

Note that in Theorem 4.1 we do not provide a general identifiability result for the system (17) that directly follows from its structure. The main purpose of the aforementioned theorem is to be able to infer about the identifiability of (17) from the identifiability of (16). Thus, in order to tackle the problem of identifiability of the system (17), we can consider the identifiability of system (16), for which, as mentioned earlier, there are many different results known in the literature (see for example Bellman & Åström, 1970; Bernard & Bastin, 2005; Cobelli & DiStefano 3rd, 1980; Craciun & Pantea, 2008; Farina et al., 2006; Miao et al., 2011; Stanhope et al., 2014).

5. Application to examples

In this section, we apply our automatic parameter estimation algorithm to three examples of CRNs. We first consider two real-life computational models of biological processes retrieved from the BioModels database (Malik-Sheriff et al., 2020). Finally, we consider an artificial example of a CRN for which the corresponding system (17) is identifiable. For each of these models we first generate data corresponding to the reaction rates using the values of parameters provided in the corresponding reference. Next, we perturb these generated data with a white Gaussian noise with zero mean and sufficiently small standard deviation. We then apply our estimation technique to determine the best-fitting values of parameters in a fully automated manner. The corresponding Matlab library is provided as supplementary material.

5.1. Ryanodine receptor adaptation

We consider a simple example of a realistic mass action CRN to illustrate the concepts described in Section 3, namely a model of ryanodine receptor adaptation (RyRA) developed in Keizer and Levine (1996). A schematic representation of the RyRA model is illustrated in the left-hand panel of Figure 1. For a detailed explanation of the corresponding mathematical model we refer to Keizer and Levine (1996). In spite of its simplicity, the model is good enough to demonstrate all the important elements of our parameter estimation method.

There are four chemical species participating in three reversible mass action reactions. Recall that, in our modelling procedure, we split each reversible reaction into two unidirectional reactions. The list

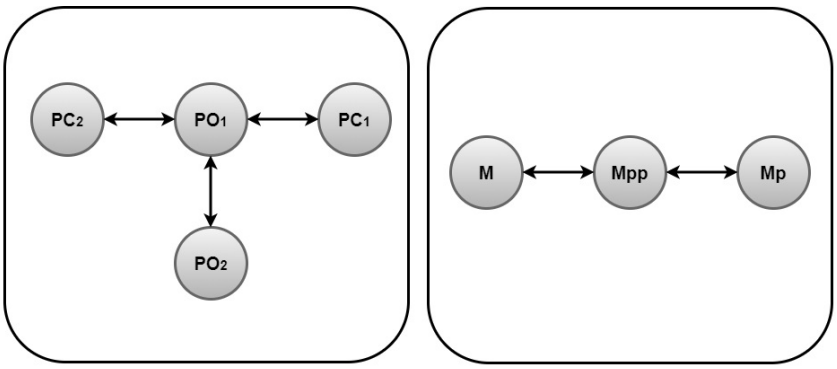

Figure 1.: Schematic representation of the model of ryanodine receptor adaptation (left-hand panel) and the model of protein kinase cascades (right-hand panel).

of the reversible reactions is given below.

$$\mathcal{R}_{1,2} : X_1 \underset{\nu_2}{\overset{\nu_1}{\rightleftharpoons}} X_3, \qquad \mathcal{R}_{3,4} : X_1 \underset{\nu_4}{\overset{\nu_3}{\rightleftharpoons}} X_2, \qquad \mathcal{R}_{5,6} : X_1 \underset{\nu_6}{\overset{\nu_5}{\rightleftharpoons}} X_4,$$

where ν_j, $j = 1, \ldots, 6$, is the rate function of the j^{th} unidirectional reaction. For $i = 1, \ldots, 4$, denote by x_i the concentration of the species X_i. Recall that since the reactions are governed by MAKRL, we can use (3) to rewrite the reaction rates in the following form:

$$\begin{aligned} \nu_1 &= k_1 x_1, & \nu_4 &= k_4 x_2, \\ \nu_2 &= k_2 x_3, & \nu_5 &= k_5 x_1, \\ \nu_3 &= k_3 x_1, & \nu_6 &= k_6 x_4. \end{aligned}$$

We suppose that the outputs are the overall reaction rates described by the following matrix:

$$H = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix}.$$

To show the applicability of our estimation method in the case when there are datasets of different length, we generate two distinct datasets for y from the IVP (2) using the values of the parameters k_i, $i = 1, \ldots, 6$, (see Table 1) provided in Keizer and Levine (1996). These datasets contain six and nine different time instants, respectively. Note that the time intervals of interest have been chosen to be of different length. Additionally, the obtained data have been perturbed with white Gaussian noise of zero mean and standard deviation 0.02 and 0.005, respectively. These data are illustrated in Figure 2 (highlighted in red). To demonstrate the applicability of our proposed method, we estimate the parameters contained in the model of RyRA using these datasets.

According to our parameter estimation procedure, we first determine the IVP (10) of reaction rates. The ODEs involved in the above-mentioned IVP corresponding to the RyRA model can be extended as:

$$\begin{aligned} \frac{d\nu_1}{dt} &= -\frac{\nu_1}{x_1}(\nu_1 - \nu_2 + \nu_3 - \nu_4 + \nu_5 - \nu_6), & \frac{d\nu_4}{dt} &= \frac{\nu_4}{x_2}(\nu_3 - \nu_4), \\ \frac{d\nu_2}{dt} &= \frac{\nu_2}{x_3}(\nu_1 - \nu_2), & \frac{d\nu_5}{dt} &= -\frac{\nu_5}{x_1}(\nu_1 - \nu_2 + \nu_3 - \nu_4 + \nu_5 - \nu_6), \\ \frac{d\nu_3}{dt} &= -\frac{\nu_3}{x_1}(\nu_1 - \nu_2 + \nu_3 - \nu_4 + \nu_5 - \nu_6), & \frac{d\nu_6}{dt} &= \frac{\nu_6}{x_4}(\nu_5 - \nu_6). \end{aligned} \tag{26}$$

Note that the manual derivation of the ODEs (26) is not straightforward. However, the usage of Matlab symbolic variables allows us to derive these ODEs in a fully automated manner. The corresponding Matlab file is also included in the Matlab library provided as supplementary material.

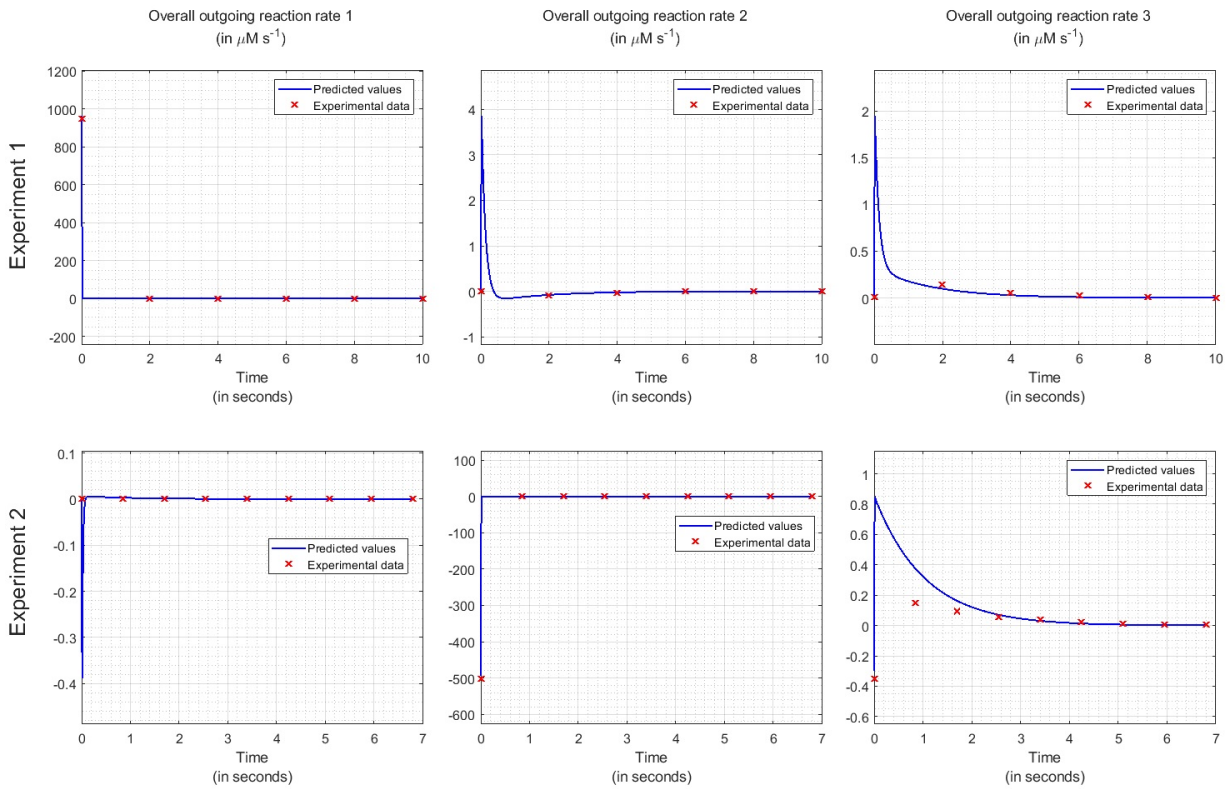

Figure 2.: Time-series data of reaction rates (in red) of the model of ryanodine receptor adaptation and the corresponding predicted values of reaction rates (in blue) of the model with parameters estimated by our method.

In the case of the first experiment, the contracted Bézier curve corresponding to each species is of degree five since we collect data at six different time instants. On the other hand, since there are four species participating in the CRN, the corresponding vector of parametric contracted Bézier curves is of length four. We therefore have 24 parameters contained in the system of ODEs (12). Similarly, in the case of the second experiment, since we collect data at nine different time instants, there are 36 parameters contained in the system of ODEs (12). For each of these experiments, the application of the least squares method determines the best-fitting values of the control points of Bézier curves, which are the values of the species' concentrations (11) at different time instants. These data are used to determine the best-fitting values of parameters (see Table 1). We calculate 95% confidence intervals (see Table 1) for the estimated parameters using the Matlab package provided in Hedengren (Retrieved July 12, 2021). This Matlab package is specifically developed to simulate a differential equation model and optimize its parameters based on measurements. It also provides the corresponding confidence intervals. The comparison of the time-series data of reaction rates and the corresponding values obtained from (2) and (5) with estimated parameter values $\widehat{k}$ can be seen in Figure 2.

Note from Table 1 that the estimated values of some of the parameters (e.g. k_3 and k_4) differ by a large percentage from their corresponding values provided in Keizer and Levine (1996). However, the mathematical model with the estimated values of parameters is a good fit (as can be seen in Figure 1) for the generated time-series data of overall outgoing reaction rates. This means that the parameters involved in the considered mathematical model of RyRA cannot be uniquely determined from the output (7). We are of the opinion that the reason for this is the fact that the multiplication matrix H does not have full column rank.

A popular technique that is often used during parameter estimation with small datasets is cross-validation (CV, see Arlot & Celisse, 2010; Bergmeir & Benítez, 2012; Stone, 1974). The idea is to avoid overfitting and obtain more accurate estimates for the parameters by splitting the dataset into different folds and using each fold subsequently as a validation set while the other folds are used for training the model (i.e., determining the optimal parameter values). In order to apply 5-fold CV, we first generate a dataset containing 25 data points for y using the parameter values provided in Keizer

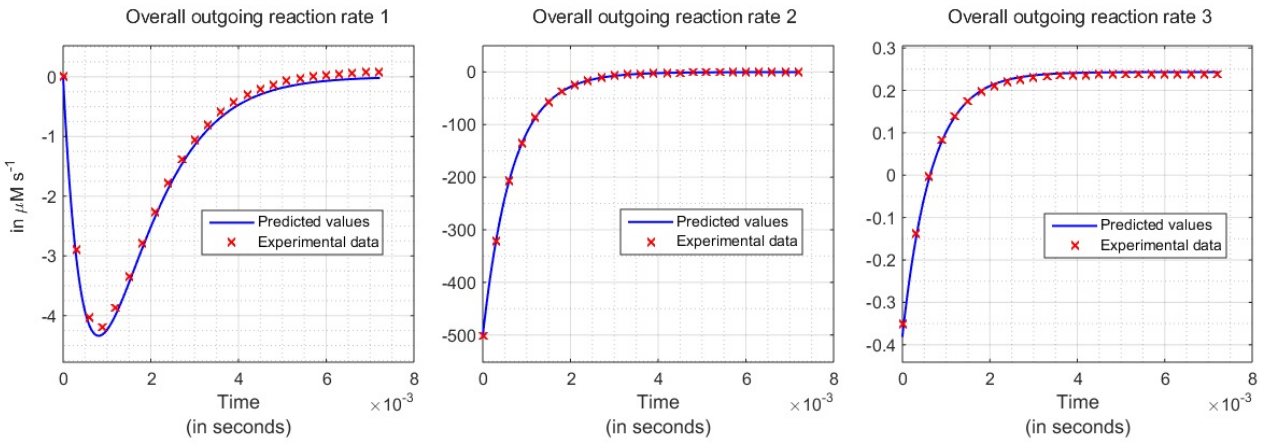

Figure 3.: Time-series data of reaction rates (in red) of the model of ryanodine receptor adaptation and the corresponding predicted values of reaction rates (in blue) of the model with parameters obtained by applying cross-validation to our parameter estimation method.

and Levine (1996). We then randomly split the generated dataset into five subsets of five data points each. Using CV, we obtain 5 estimates for each parameter, which we then average to produce a single estimate of the parameters involved in the mathematical model of RyRA. The resulting estimates are given in Table 1. Note that, unlike our main estimation procedure, the CV technique results in estimates for the parameters k_3 and k_4 that are closer to their corresponding provided values. One of the reasons behind this is the bigger size of the dataset used for CV. Figure 3 illustrates the comparison of the dataset of 25 data points and the corresponding model predicted values using the estimated values of parameters obtained by CV.

	k_1	k_2	k_3	k_4	k_5	k_6
Provided values	28.8000	984.1500	1093.5000	385.9000	1.7500	0.1000
Estimated values	27.5623	997.4123	4.8534	1.7213	2.3802	0.1492
Confidence intervals	[25.3574,31.7672]	[993.2074,1001.6172]	[0.6485,9.0583]	[0.0000,5.9262]	[0.0000,6.5851]	[0.0000,4.3541]
CV estimates	29.1995	1000.1721	1093.5171	385.9031	1.8514	0.1082

Table 1.: The values of the rate constants provided in Keizer and Levine (1996), the estimated values obtained using our method with their corresponding confidence intervals, and the estimates obtained by applying cross validation to our estimation method. The unit of each parameter is s^{-1}.

5.2. Protein kinase cascades

Subsequently, we apply our parameter estimation method to a model of protein kinase cascades (PKC), which is a model of a CRN governed by EKRL. The network is schematically represented in the right-hand panel of Figure 1. The detailed explanation of the complete mathematical model of PKC can be found in Markevich et al. (2004).

The reaction network consists of three chemical species participating in four unidirectional reactions governed by Michaelis-Menten kinetics. These reactions are given below.

$$\mathcal{R}_{1,2} : X_1 \underset{\nu_2}{\overset{\nu_1}{\rightleftharpoons}} X_2, \qquad \mathcal{R}_{3,4} : X_2 \underset{\nu_4}{\overset{\nu_3}{\rightleftharpoons}} X_3,$$

where, for $j = 1, \ldots, 4$, ν_j is the rate function of the j^{th} unidirectional reaction. For $j = 1, \ldots, 4$ and $i = 1, 2, 3$, let k_j denote the rate constant of the j^{th} reaction and let x_i denote the concentration of

the i^{th} species. The reaction rates are given in the form (3):

$$\nu_1 = \frac{k_1 x_1}{1 + \frac{x_1}{K_1} + \frac{x_2}{K_2}}, \qquad \nu_3 = \frac{k_3 x_2}{1 + \frac{x_1}{K_1} + \frac{x_2}{K_2}},$$
$$\nu_2 = \frac{k_2 x_2}{1 + \frac{x_1}{K_5} + \frac{x_2}{K_4} + \frac{x_3}{K_3}}, \qquad \nu_4 = \frac{k_4 x_3}{1 + \frac{x_1}{K_5} + \frac{x_2}{K_4} + \frac{x_3}{K_3}}.$$

Assume that we have collected measurements of overall outgoing reaction rates in respective directions specified by the following matrix:

$$H = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix}.$$

We generate data for y using the values of initial concentrations and parameters provided in Markevich et al. (2004) (see Table 2). These generated data have been perturbed with Gaussian white noise of zero mean and standard deviation 0.02.

In this case, the ODEs involved in the IVP (9) can be rewritten as:

$$\begin{aligned}
\frac{d\nu_1}{dt} &= \frac{\nu_1\,\nu_2}{x_1} - \frac{{\nu_1}^2}{x_1} - \frac{\nu_1\,\sigma_2(\nu,K)}{\sigma_4(x,K)}, \\
\frac{d\nu_2}{dt} &= \frac{\nu_1\,\nu_2}{x_2} - \frac{{\nu_2}^2}{x_2} - \frac{\nu_2\,\nu_3}{x_2} + \frac{\nu_2\,\nu_4}{x_2} + \frac{v_2\,\sigma_1(\nu,K)}{\sigma_3(x,K)}, \\
\frac{d\nu_3}{dt} &= \frac{v_1\,v_3}{x_2} - \frac{{v_3}^2}{x_2} - \frac{v_2\,v_3}{x_2} + \frac{v_3\,v_4}{x_2} - \frac{v_3\,\sigma_2(\nu,K)}{\sigma_4(x,K)}, \\
\frac{d\nu_4}{dt} &= \frac{\nu_3\,\nu_4}{x_3} - \frac{{\nu_4}^2}{x_3} + \frac{\nu_4\,\sigma_1(\nu,K)}{\sigma_3(x,K)},
\end{aligned} \tag{27}$$

where σ_i, $i = 1, \ldots, 4$, are given as:

$$\begin{aligned}
\sigma_1(\nu,K) &= K_3\,K_4\,\nu_1 - K_3\,K_4\,\nu_2 - K_3\,K_5\,\nu_1 + K_3\,K_5\,\nu_2 \\
&+ K_3\,K_5\,\nu_3 - K_3\,K_5\,\nu_4 - K_4\,K_5\,\nu_3 + K_4\,K_5\,\nu_4, \\
\sigma_2(\nu,K) &= K_1\,\nu_1 - K_1\,\nu_2 - K_2\,\nu_1 - K_1\,\nu_3 + K_2\,\nu_2 + K_1\,\nu_4, \\
\sigma_3(x,K) &= K_3\,K_4\,K_5 + K_3\,K_4\,x_1 + K_3\,K_5\,x_2 + K_4\,K_5\,x_3, \\
\sigma_4(x,K) &= K_1\,K_2 + K_1\,x_2 + K_2\,x_1.
\end{aligned}$$

Similar to the case of the model of RyRA, usage of Matlab symbolic variables allows us to derive these ODEs in a fully automated manner. Since we generate data at six different time instants, the contracted Bézier curve corresponding to each species is of degree five. On top of that, since there are three species participating in the CRN, the corresponding vector of parametric contracted Bézier curves is of length three. Adding the five Michaelis constants of the species, our system of ODEs (9) contains 23 parameters. The application of the least squares method determines the best-fitting values of the control points of the contracted Bézier curves (which are the values of species' concentrations (11) at six different time instants) and the Michaelis constants of the species (see Table 2). These data of species concentrations' and the values of the Michaelis constants are used to determine the best-fitting values of parameters (see Table 2). Similar to the model of RyRA, we also provide the 95% confidence interval of each parameter as calculated by Matlab in Table 2. The comparison of the generated data and the model-predicted values of reaction rates corresponding to the above-mentioned best-fitting parameter values is illustrated in Figure 4.

Note that, similar to the case of RyRA, for some of the parameters (e.g. k_3) there is a substantial difference between the values provided in Markevich et al. (2004) and their corresponding estimated values. We can therefore infer that the parameters contained in the system (17) are not identifiable

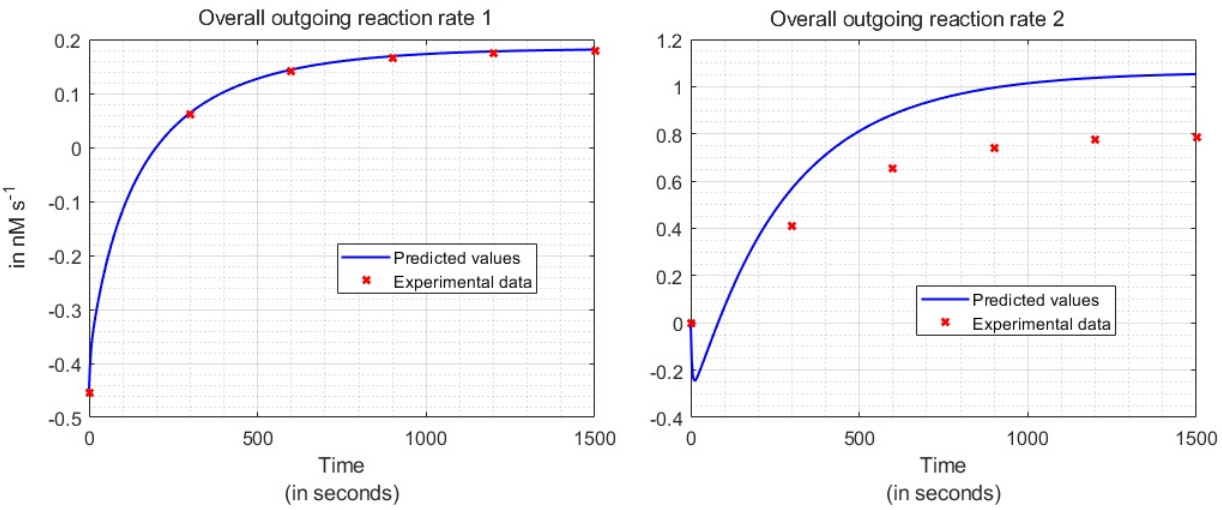

Figure 4.: Time-series data of reaction rates (in red) of the model of protein kinase cascades and the corresponding predicted values of reaction rates (in blue) of the model with parameters estimated by our method.

from the output (7). As mentioned earlier, we are of the opinion that the reason behind the non-identifiability of the parameters is the fact that the multiplication matrix H appearing in the output does not have full column rank.

	k_1	k_2	k_3	k_4	K_1	K_2	K_3	K_4	K_5
Provided values	0.0100	0.3333	1.5000	0.3820	50.000	500.000	22.000	18.000	78.000
Estimated values	0.0101	0.3540	1.9148	0.4836	50.186	501.193	21.938	18.067	76.084
Confidence intervals	[0.0000,7.1063]	[0.0000,7.4502]	[0.0000,9.0110]	[0.0000,7.5798]	[43.0898,57.2822]	[494.0968,508.2892]	[14.8418,29.0342]	[10.9708,25.1632]	[68.9878,83.1802]

Table 2.: The values of the rate constants and the Michaelis constants provided in Markevich et al. (2004), the estimated values using our parameter estimation method and the corresponding confidence intervals. The units of rate constants and Michaelis constants are s^{-1} and nM, respectively.

5.3. An example of an identifiable system

In this part of the manuscript, we consider an example of a CRN governed by MAKRL, for which the corresponding system (17) is identifiable. Assume that 19 distinct chemical species X_i, $i = 1, \ldots, 19$, are participating in the following six irreversible reactions:

$$\begin{aligned} &\mathcal{R}_1 : X_1 + X_2 \longrightarrow X_3 + X_4, && \mathcal{R}_4 : X_{10} + X_{11} \longrightarrow X_{12} + X_{13}, \\ &\mathcal{R}_2 : X_4 + X_5 \longrightarrow X_6 + X_7, && \mathcal{R}_5 : X_{13} + X_{14} \longrightarrow X_{15} + X_{16}, \\ &\mathcal{R}_3 : X_7 + X_8 \longrightarrow X_9 + X_{10}, && \mathcal{R}_6 : X_{16} + X_{17} \longrightarrow X_{18} + X_{19}. \end{aligned} \tag{28}$$

Further assume that the reactions (28) are governed by MAKRL, i.e., the corresponding reaction rates are given as:

$$\begin{aligned} \nu_1 &= k_1 x_1 x_2, & \nu_4 &= k_4 x_{10} x_{11}, \\ \nu_2 &= k_2 x_4 x_5, & \nu_5 &= k_5 x_{13} x_{14}, \\ \nu_3 &= k_3 x_7 x_8, & \nu_6 &= k_6 x_{16} x_{17}, \end{aligned}$$

where, as earlier, k_i, $i = 1, \ldots, 6$, denotes the rate constant of the i^{th} reaction and x_i, $i = 1, \ldots, 19$, denotes the concentration of the i^{th} species X_i. First note that the system (17) corresponding to this example of CRN is identifiable, i.e., the rate constants k_i, $i = 1, \ldots, 6$, can be uniquely determined if measurements of reaction rates ν_i, $i = 1, \ldots, 6$, are available. The proof of the identifiablity is included in the supplementary material. We generate time-series data of reaction rates of the six irreversible reactions (28) using randomly chosen parameter values (see Table 3).

In this case, the ODEs involved in the IVP (9) can be rewritten as:

$$\begin{aligned}
\frac{d\nu_1}{dt} &= -\frac{\nu_1^2}{x_1} - \frac{\nu_1^2}{x_2}, & \frac{d\nu_4}{dt} &= \frac{\nu_3\nu_4}{x_{10}} - \frac{\nu_4^2}{x_{10}} - \frac{\nu_4^2}{x_{11}},\\
\frac{d\nu_2}{dt} &= \frac{\nu_1\nu_2}{x_4} - \frac{\nu_2^2}{x_4} - \frac{\nu_2^2}{x_5}, & \frac{d\nu_5}{dt} &= \frac{\nu_4\nu_5}{x_{13}} - \frac{\nu_5^2}{x_{13}} - \frac{\nu_5^2}{x_{14}},\\
\frac{d\nu_3}{dt} &= \frac{\nu_2\nu_3}{x_7} - \frac{\nu_3^2}{x_7} - \frac{\nu_3^2}{x_8}, & \frac{d\nu_6}{dt} &= \frac{\nu_5\nu_6}{x_{16}} - \frac{\nu_6^2}{x_{16}} - \frac{\nu_6^3}{x_{17}}.
\end{aligned} \tag{29}$$

Since there are 19 species participating in the CRN, the corresponding vector of parametric contracted Bézier curves is of length 19. On the other hand, since our dataset corresponds to seven distinct time instants, the system of ODEs contains 133 parameters. The application of the least squares method determines the values of species' concentrations (11) at seven different time instants. These data of species concentrations' are used to determine the best-fitting values of parameters (see Table 3). The 95% confidence intervals as calculated by Matlab are included in Table 3. The comparison of the generated data and the model-predicted values of reaction rates corresponding to the above-mentioned best-fitting parameter values is illustrated in Figure 5.

Note from Table 3 that the estimated values of the parameters are close to their corresponding parameter values that we used to generate data for the reaction rates. The reason behind this is the fact that the mathematical model (17) corresponding to the considered CRN is identifiable. This means that the parameters involved in the considered mathematical model can be uniquely determined from the output (7). Therefore, the mathematical model with the estimated values of parameters is a good fit (as expressed in Figure 5) for the available time-series data of overall outgoing reaction rates.

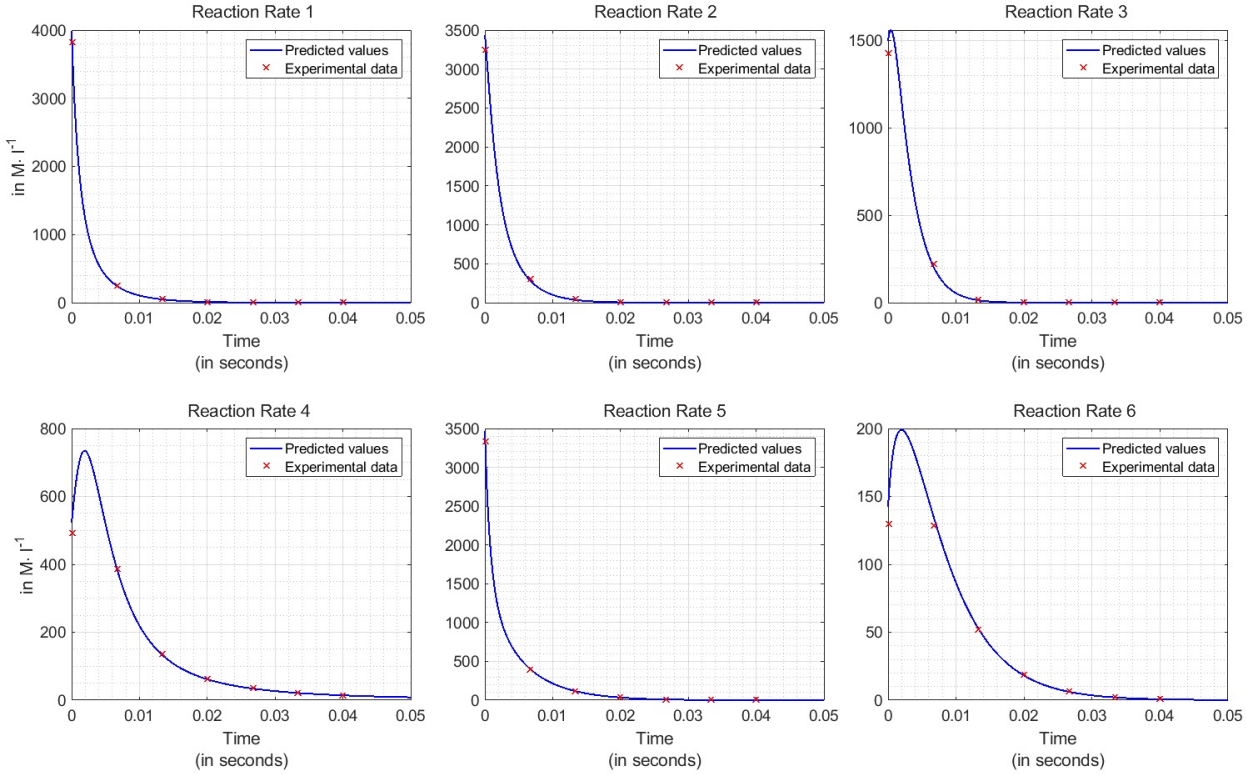

Figure 5.: Time-series data of reaction rates (in red) of the model of the CRN illustrated in (28) and the corresponding predicted values of reaction rates (in blue) of the model with parameters estimated using our method.

	k_1	k_2	k_3	k_4	k_5	k_6
Provided values	34.1000	72.3000	59.5000	35.1000	61.8000	13.0000
Estimated values	34.6805	72.9589	58.9197	35.6440	61.3884	13.4073
Confidence intervals	[25.8329,43.5281]	[64.1113,81.8065]	[50.0721,67.7673]	[26.7964,44.4916]	[52.5408,70.2360]	[4.5597,22.2549]

Table 3.: The values of the rate constants, their estimated values using our parameter estimation method, and the corresponding confidence intervals.

6. Conclusion and discussion

In this manuscript, we presented a parameter estimation method for enzymatic CRNs from observed time-series experimental data of reaction rates. We consider enzymatic CRNs, since it is known that the reaction rates of a bio-CRN can be precisely determined in an experimental setup using ^{13}C metabolic flux analysis. Currently, to the best of our knowledge there is no direct method for estimating the parameters involved in a mathematical model from experimental time-series data of reaction rates alone. We filled the gap by proposing an algorithmic approach for the above-mentioned problem. The principle behind our method is to retrieve time-series data of species' concentrations by making use of parametric Bézier curves. We first used the specific behavior of an enzymatic CRN to derive an IVP, independent of the rate constants, with the corresponding dependent variable being the vector of reaction rates. We then used parametric Bézier curves to approximate the species' concentrations in the newly obtained system. As a result, we converted the main problem to the parameter estimation problem of a CRN from time-series data of species' concentrations. There are a number of techniques available in the literature to deal with such a parameter estimation problem. Because of its simplicity and low computational intensity, we have used the least squares method to identify the best-fitting values of the parameters involved in the mathematical model of a CRN. We implemented an automated procedure for our parameter estimation method, creating a Matlab library that contains all the necessary files related to it. It can be used to estimate the parameters from time-series' experimental data of reaction rates in a fully automated manner. This Matlab library is provided as supplementary material. Finally, we used the aforementioned Matlab library to successfully apply our parameter estimation method on two real-life examples of CRNs retrieved from the Biomodels database (Malik-Sheriff et al., 2020) and an artificial example for which the corresponding system (17) is identifiable. For each of these models, we observe that our parameter estimation method is able to derive a complete mathematical model that is able to make accurate predictions about the dynamical behavior of the CRN. We have also calculated 95% confidence intervals for the estimated parameters using the Matlab package Hedengren (Retrieved July 12, 2021). It is a useful tool that is specially developed to simulate a mathematical model of ODEs, estimate the parameters involved in it from experimental data, and also calculate corresponding confidence intervals. Note that, for certain parameters, the confidence intervals are not symmetric. This is because of the fact that the parameters cannot admit negative values, so the negative parts of the obtained intervals are excluded. Also note that, in the case of our three demonstrative examples, the associated confidence intervals for certain parameters seem to be large. The reasons behind this are the factors affecting the width of the confidence interval (e.g., the size of the available experimental data). Larger experimental data will tend to produce a better estimate of the parameters and a broader confidence interval. For our parameter estimation method, other techniques could also be used in order to compute confidence intervals of the estimated parameters. However which of these techniques is the best suited for our method is a separate topic that can be regarded as future work.

An important tool in our parameter estimation method is the parametric Bézier curve that is used to approximate the concentrations of the species participating in the considered CRN. An alternative to the Bézier curve may be the B-spline, which is a generalization of the Bézier curve. A B-spline is a spline function (a function defined piece-wise by polynomials) defined by its order and a pre-specified set of given knots. As a matter of fact, a Bézier curve is a B-spline with no interior knots. For a detailed description of B-splines we refer to Prautzsch et al. (2002). However, even though B-splines are more powerful and flexible curves than Bézier curves, the theory behind such curves is more complicated. Moreover, the application of B-splines is computationally more expensive. Another powerful technique in approximating continuous curves is polynomial interpolation, which is the interpolation of a given

dataset by the polynomial of lowest possible degree that passes through the points of the dataset. For any given dataset there is a unique interpolating polynomial, known as the Lagrange interpolating polynomial (Waring, 1779). Even though approximating the species' concentrations with Lagrange interpolating polynomials sounds logical, it has a serious shortcoming compared to Bézier curves. In our case the considered datasets correspond to time-series data of species' concentrations, which are always non-negative. It is therefore assured that the Bézier curves are also non-negative. This is due to the fact that any Bézier curve is contained in the convex hull of the corresponding Bézier polygon, which is the polygon formed by connecting the control points with lines, starting at the first control point and finishing with the last control point. In general, it is not assured that the Lagrange interpolating polynomial corresponding to the available dataset will also be non-negative. As such, it is preferred to use Bézier curves for approximating the species' concentrations. In the supplementary material we include the definition of Lagrange interpolating polynomials and also provide an example of a dataset consisting of non-negative data points, for which the corresponding Lagrange interpolating polynomial is negative over some time interval. The associated Matlab function is included in our Matlab library that is provided as supplementary material.

We considered the structural identifiability of the parameters involved in the balance laws (2) of a given CRN with the output being a linear combination of unidirectional reaction rates ν. In other words, we examined the possibility to uniquely determine the parameters assuming that the output was the admissible set of output-trajectories of the form $y = H\nu$, for some constant matrix H. In the case when H has full column rank we proved that, if the parameters involved in the balance laws were structurally identifiable from the admissible set of concentration trajectories, then they were also structurally identifiable from the admissible set of output-trajectories corresponding to y. On top of that, we conjectured the impossibility of the parameters to be structurally identifiable in the case when H is not of a full column rank. Note that in this manuscript we did not provide a general result on this question. However, with the help of two simple examples we demonstrated that if the considered CRN contained only reversible reactions and that if the output was the corresponding vector of overall outgoing reaction rates (in a certain direction), then the aforementioned system was not necessarily identifiable. As a future work, we intend to give an answer to the general problem of identifiability described above.

References

Antoniewicz, M. R. (2015). Methods and advances in metabolic flux analysis: a mini-review. *Journal of Industrial Microbiology & Biotechnology*, *42*(3), 317–325.

Antoniewicz, M. R. (2018). A guide to ^{13}C-metabolic flux analysis for the cancer biologist. *Experimental & Molecular Medicine*, *50*(4), 1–13.

Arlot, S., & Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys*, *4*, 40–79.

Bellman, R., & Åström, K. J. (1970). On structural identifiability. *Mathematical Biosciences*, *7*, 329–339.

Bergmeir, C., & Benítez, J. M. (2012). On the use of cross-validation for time series predictor evaluation. *Information Sciences*, *191*, 192–213.

Bernard, O., & Bastin, G. (2005). Identification of reaction networks for bioprocesses: determination of a partially unknown pseudo-stoichiometric matrix. *Bioprocess and Biosystems Engineering*, *27*(5), 293–301.

Bernštein, S. (1912). Démonstration du théoreme de Weierstrass fondée sur le calcul des probabilities. *Communications of the Kharkov Mathematical Society*, *13*, 1–2.

Berry, T. G., & Patterson, R. R. (1997). The uniqueness of Bézier control points. *Computer Aided Geometric Design*, *14*, 877–879.

Bézier, P. E. (1986). *The Mathematical Basis of the UNISURF CAD System.* Newton, MA: Butterworth-Heinemann.

Bézier, P. E., & Sioussiou, S. (1983). Semi-automatic system for defining free-form curves and surfaces. *Computer-Aided Design*, *15*(2), 65–72.

Cobelli, C., & DiStefano 3rd, J. J. (1980). Parameter and structural identifiability concepts and ambiguities: a critical review and analysis. *American Journal of Physiology. Regulatory, Integrative and Comparative Physiology*, *239*(1), R7–R24.

Craciun, G., & Pantea, C. (2008). Identifiability of chemical reaction networks. *Journal of Mathematical Chemistry*, *44*, 244–259.

Crampin, E. J., Schnell, S., & McSharry, P. E. (2004). Mathematical and computational techniques to deduce complex biochemical reaction mechanisms. *Progress in Biophysics and Molecular Biology*, *86*(1), 77–112.

Donnelly, J. K., & Quon, D. (1970). Identification of parameters in systems of ordinary differential equations using quasilinearization and data perturbation. *The Canadian Journal of Chemical Engineering*, *48*(1), 114–119.

Edwards, L. M., & Thiele, I. (2013). Applying systems biology methods to the study of human physiology in extreme environments. *Extreme Physiology & Medicine*, *2*, 8.

Farin, G. (1993). *Curves and Surfaces for Computer-Aided Geometric Design: A Practical Guide* (3rd ed.). San Diego, CA: Academic Press, Inc.

Farina, M., Findeisen, R., Bullinger, E., Bittanti, S., Allgower, F., & Wellstead, P. (2006). Results towards identifiability properties of biochemical reaction networks. In *Proceedings of the 45th IEEE Conference on Decision & Control* (pp. 2104–2109).

Fröhlich, F., Kaltenbacher, B., Theis, F. J., & Hasenauer, J. (2017). Scalable parameter estimation for genome-scale biochemical reaction networks. *PLOS Computational Biology*, *13*(1), e1005331.

Gasparyan, M., Van Messem, A., & Rao, S. (2018). A novel technique for model reduction of biochemical reaction networks. *IFAC-PapersOnLine*, *51*(19), 28–31.

Gasparyan, M., Van Messem, A., & Rao, S. (2020). An automated model reduction method for biochemical reaction networks. *Symmetry-Basel*, *12*(8), 1321.

Guo, W., Sheng, J., & Feng, X. (2016). ^{13}C-metabolic flux analysis: an accurate approach to demystify microbial metabolism for biochemical production. *Bioengineering-Basel*, *3*(1), 3.

Hedengren, J. (Retrieved July 12, 2021). Dynamic parameter estimation and confidence intervals. *MATLAB Central File Exchange*. (`https://www.mathworks.com/matlabcentral/fileexchange/45496-dynamic-parameter-estimation-and-confidence-intervals`)

Heirendt, L., Arreckx, S., Pfau, T., Mendoza, S. N., Richelle, A., Heinken, A., ... Fleming, R. M. (2019). Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. *Nature Protocols*, *14*(3), 639–702.

Himmelblau, D. M., & Riggs, J. B. (2012). *Basic Principles and Calculations in Chemical Engineering* (8th ed.). Upper Saddle River, NJ: Prentice Hall.

Hodges, A., & Chatelier, R. (2002). *Electrochemical method for measuring chemical reaction rates.* Google Patents. (US Patent 6,444,115)

Keizer, J., & Levine, L. (1996). Ryanodine receptor adaptation and Ca^{2+}-induced Ca^{2+} release-dependent Ca^{2+} oscillations. *Biophysical Journal*, *71*(6), 3477–3487.

Ljung, L. (1999). *System Identification: Theory for the User* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.

Long, C. P., & Antoniewicz, M. R. (2019). High-resolution ^{13}C metabolic flux analysis. *Nature Protocols*, *14*(10), 2856–2877.

Loskot, P., Atitey, K., & Mihaylova, L. (2019). Comprehensive review of models and methods for inferences in bio-chemical reaction networks. *Frontiers in Genetics*, *10*, 549.

Malik-Sheriff, R. S., Glont, M., Nguyen, T. V. N., Tiwari, K., Roberts, M. G., Xavier, A., ... Hermjakob, H. (2020). BioModels – 15 years of sharing computational models in life science. *Nucleic Acids Research*, *48*(D1), D407–D415.

Markevich, N. I., Hoek, J. B., & Kholodenko, B. N. (2004). Signaling switches and bistability arising from multisite phosphorylation in protein kinase cascades. *The Journal of Cell Biology*, *164*(3), 353–359.

Megchelenbrink, W., Rossell, S., Huynen, M. A., Notebaart, R. A., & Marchiori, E. (2015). Estimating metabolic fluxes using a maximum network flexibility paradigm. *PLOS ONE*, *10*(10), e0139665.

Miao, H., Xia, X., Perelson, A. S., & Wu, H. (2011). On identifiability of nonlinear ODE models and applications in viral dynamics. *SIAM Review*, *53*(1), 3–39.

Mock, A., Chiblak, S., & Herold-Mende, C. (2014). Lessons we learned from high-throughput and top-down systems biology analyses about glioma stem cells. *Current Pharmaceutical Design*, *20*(1), 66–72.

Orth, J. D., Thiele, I., & Palsson, B. P. (2010). What is flux balance analysis? *Nature Biotechnology*, *28*(3), 245–248.

Prautzsch, H., Boehm, W., & Paluszny, M. (2002). *Bézier and B-spline Techniques.* Berlin: Springer-Verlag.

Rao, S., van der Schaft, A., van Eunen, K., Bakker, B. M., & Jayawardhana, B. (2014). A model reduction method for biochemical reaction networks. *BMC Systems Biology*, *8*, 52.

Ross, J., Schreiber, I., & Vlad, M. O. (2005). *Determination of Complex Reaction Mechanisms: Analysis of Chemical, Biological, and Genetic Networks.* Oxford: Oxford University Press.

Sauer, U. (2006). Metabolic networks in motion: ^{13}C-based flux analysis. *Molecular Systems Biology*, *2*, 62.

Sparkman, O. D., Penton, Z., & Kitson, F. G. (2011). *Gas Chromatography and Mass Spectrometry: A Practical Guide* (2nd ed.). San Diego, CA: Academic Press.

Stanhope, S., Rubin, J. E., & Swigon, D. (2014). Identifiability of linear and linear-in-parameters dynamical systems from a single trajectory. *SIAM Journal on Applied Dynamical Systems*, *13*(4), 1792–1815.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal*

Statistical Society: Series B (Methodological), *36*(2), 111–147.
Waring, E. (1779). Problems concerning interpolations. *Philosophical Transactions of the Royal Society of London*, *69*, 59–67.
Wiechert, W. (2001). ^{13}C-metabolic flux analysis. *Metabolic Engineering*, *3*(3), 195–206.