Article

# Parameter Estimation for Kinetic Models of Chemical Reaction Networks from Partial Experimental Data of Species' Concentrations 

Manvel Gasparyan 1,2,3,*(D) and Shodhan Rao ${ }^{2,3(\mathbb{D}}$<br>1 School of Environmental Engineering, University of Seoul, Seoul 02504, Republic of Korea<br>2 Center for Biosystems and Biotech Data Science, Ghent University Global Campus, Incheon 21985, Republic of Korea; shodhan.rao@ghent.ac.kr<br>3 Department of Data Analysis and Mathematical Modelling, Ghent University, 9000 Ghent, Belgium<br>* Correspondence: mgaspary23@office.uos.ac.kr

Citation: Gasparyan, M.; Rao, S. Parameter Estimation for Kinetic Models of Chemical Reaction Networks from Partial Experimental Data of Species' Concentrations. Bioengineering 2023, 10, 1056. https:/ /doi.org/10.3390/ bioengineering10091056

Academic Editors: Dipali Singh and Jianwen Ye

Received: 14 July 2023
Revised: 23 August 2023
Accepted: 31 August 2023
Published: 7 September 2023


Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/).


#### Abstract

The current manuscript addresses the problem of parameter estimation for kinetic models of chemical reaction networks from observed time series partial experimental data of species concentrations. It is demonstrated how the Kron reduction method of kinetic models, in conjunction with the (weighted) least squares optimization technique, can be used as a tool to solve the abovementioned ill-posed parameter estimation problem. First, a new trajectory-independent measure is introduced to quantify the dynamical difference between the original mathematical model and the corresponding Kron-reduced model. This measure is then crucially used to estimate the parameters contained in the kinetic model so that the corresponding values of the species' concentrations predicted by the model fit the available experimental data. The new parameter estimation method is tested on two real-life examples of chemical reaction networks: nicotinic acetylcholine receptors and Trypanosoma brucei trypanothione synthetase. Both weighted and unweighted least squares techniques, combined with Kron reduction, are used to find the best-fitting parameter values. The method of leave-one-out cross-validation is utilized to determine the preferred technique. For nicotinic receptors, the training errors due to the application of unweighted and weighted least squares are 3.22 and 3.61 respectively, while for Trypanosoma synthetase, the application of unweighted and weighted least squares result in training errors of 0.82 and 0.70 respectively. Furthermore, the problem of identifiability of dynamical systems, i.e., the possibility of uniquely determining the parameters from certain types of output, has also been addressed.


Keywords: systems biology; mathematical modeling; mass action kinetics; model reduction; least squares optimization; parameter identifiability

## 1. Introduction

An important step in understanding the dynamics of a chemical reaction network (CRN) is its mathematical modelling. The dynamics of a CRN are usually determined by a system of ordinary differential equations (ODEs) known as the kinetic model of the CRN. The set of parameters of such a system is usually partially or even entirely unknown and is often estimated from various types of observational data obtained from biochemical experiments. Typically, the experimental data available for estimating the parameters are time series, i.e., the data are collected at discrete points in time.

The bottom-up approach (see, e.g., [1,2]) is a modelling method widely utilized in various research domains, including systems biology. In this method, accessible experimental data are employed to construct a comprehensive mathematical model of a system. The bottom-up modelling technique in systems biology comprises four primary stages. The initial stage is draft reconstruction, which involves the compilation of appropriate data from biological experiments. The second stage involves manually curating the gathered data, for
example, by inserting absent values and removing irrelevant data. In the third stage, the knowledge concerning the biological interactions occurring in the CRN is transcribed into a mathematical expression. In the fourth stage, the parameters of this mathematical expression are numerically approximated from the observed experimental data, culminating in a comprehensive mathematical model.

The final stage of the bottom-up approach, namely parameter determination, can be executed using various approaches. The fundamental concept behind any method for parameter estimation is to compare the available experimental data with the corresponding values produced by the mathematical model. To guarantee a well-defined parameter estimation problem, it is essential to possess complete experimental data of external variables. This means having data corresponding to measurements of all external variables in the mathematical model. In the instance of CRNs, the choice of the most suitable technique generally relies on the qualities of the data amassed from biological experiments and the structure of the CRN being considered. Several mathematical techniques have been explored in the literature for parameter estimation of kinetic models of CRNs utilizing experimental data of species concentrations. A prevalent approach to solve the parameter estimation problem for scenarios where all species concentrations can be experimentally measured is the well-known (weighted) least squares technique (see, e.g., [3,4]). The application of this optimization approach minimises the (weighted) summation of squared residuals, i.e., the (weighted) summation of squared differences between the observed experimental concentration values and the corresponding values foreseen by the model. Some of the widely recognized methodologies such as maximum likelihood estimation, finite differences, quasi-linearization procedure, etc. have been deliberated in [5]. For an exhaustive overview of available mathematical techniques, consult [6-9]. In specific instances, experiments provide data corresponding to reaction rates (see, e.g., [10-13]). Recently, in [14], a method for parameter estimation from this form of experimental data for enzymatic CRNs was proposed. It is based on the approximation of the vector of species concentrations with parametric Bézier curves [15,16], which, in combination with the general least squares approach, leads to a complete mathematical model.

Bayesian-based parameter estimation techniques have also been widely developed for systems of ODEs (see, e.g., [17-22]). In such cases, the vector of parameters involved in the system of ODEs is usually treated as a random variable. A probability distribution (known as prior distribution) such as the normal distribution, the uniform distribution, the Poisson distribution, etc., (see, e.g., [23]) is therefore assigned to the vector of parameters. The technical core of the Bayesian parameter estimation approach is to construct the joint probability distribution (for the vector of parameters and the available data corresponding to the vector of dependent variables) and to perform computations to determine the posterior distribution (the conditional distribution of the parameters, given the experimental data corresponding to the dependent variable). Even though Bayesian-based approaches are useful techniques for parameter estimation and are extensively applied to CRNs (see, e.g., [24-28]), such approaches have many shortcomings. One of the main shortcomings is that most of these methods use only the available experimental data and do not consider, for the most part, the structure of the system of ODEs. Moreover, such an approach is mostly not straightforward to apply and may require huge computational efforts. Another shortcoming is the ambiguity in the construction of appropriate probability specifications since there is no explicit strategy for properly assigning a prior distribution. Depending on the characteristics of the available experimental data, it cannot always be assured that the model-predicted values corresponding to the obtained Bayesian estimates will be a satisfactory fit for the available experimental data. In such cases, a separate analysis of the fit of the model is required.

In most cases, not all concentrations can be measured experimentally. This incompleteness of the data makes the problem of parameter estimation more challenging, both mathematically and computationally, as we do not have a well-posed parameter estimation problem in this case. As expected, there is no general direct solution to this problem in the literature. Therefore, new mathematical techniques are necessary for estimating the
parameters included in mathematical models using this type of data. In this manuscript, we address the problem of parameter estimation for CRNs from observed time series partial experimental data of species' concentrations. We are mainly interested in CRNs governed by the mass action kinetic rate law (MAKRL), since this is the law that governs most real-life CRNs. We assume that measurements of concentrations are available for some of the species at discrete time points, which may not necessarily be equidistant. Direct estimation of the parameters of the mathematical model using the available partial experimental data is usually not feasible. This is because we do not have a well-posed parameter estimation problem, since not all concentrations involved in the mathematical model are measured experimentally. Therefore, we employ the Kron reduction method [29] to transform the ill-posed parameter estimation problem into a well-posed problem.

Before performing parameter estimation, it is important to understand whether the parameters are identifiable given a particular type of output, i.e., whether there is a unique parameter vector corresponding to the given output vector. We address the parameter identifiability issue of dynamical systems described by a system of ODEs. The uniqueness of the parameter vector depends on the structure of the system under consideration and the type of the available outputs. First, we recall two known definitions of parameter identifiability and then demonstrate the link between them

Our parameter estimation technique consists of three major steps. The first of these involves reduction of the known original model of the biochemical reaction network with unknown parameters using the technique of Kron reduction [29]. Of the several available techniques for model reduction of CRNs in the literature (see [30,31] for a thorough review of such techniques), we chose Kron reduction method for two particular reasons. The first and foremost is that Kron reduction preserves the kinetics of the original model. Thus, if the original network is governed by MAKRL, the reduced model obtained by Kron reduction also corresponds to another CRN, whose variables are concentrations of chemical species that are a subset of the species of the original network and moreover, the reduced CRN is also governed by MAKRL. This kinetics preservation property does not hold for several other known reduction techniques. The second reason is that by using Kron reduction method, it is possible to arrive at a reduced model of a CRN, whose dependent variables are exactly the set of concentrations of compounds whose time-series data are available. The (unknown) parameters of this reduced model are functions of the parameters of the original mathematical model. The availability of this reduced model with unknown parameters together with the time-series data of all the dependent variables of this model leads to a well-posed parameter estimation problem. The solution of this parameter estimation problem is the second step of our procedure. Because of its simplicity and computational feasibility, we apply the least squares optimization technique to deal with this estimation problem. In the final step, we solve an optimization problem in which the parameters of the original model are determined in such a way that the original mathematical model and the Kron reduced model have minimum difference between a key characteristic property associated with them. It is shown that this key characteristic property is related to the settling time of the corresponding CRN in case the given model is linear. The entire procedure has been automated and the corresponding MATLAB library is provided as Supplementary Material.

We apply our new techniques to two realistic examples of CRNs from the Biomodels database [32]. We consider a model of nicotinic acetylcholine receptors [33] and a model of Trypanosoma brucei trypanothione synthetase [34]. For each of these models, we first generate partial time series data of the species' concentrations using the parameter values given in the corresponding reference. We then explain how to derive the Kron-reduced mathematical model obtained by deleting a subset of complexes and determine the values for its parameters such that the dynamics of the Kron-reduced model in the least squares sense most closely approximate the available time series data. Using these estimated values for the parameters of the Kron-reduced model and their dependence on the parameters of
the original mathematical model, we finally determine the estimates for the parameters contained in the original model.

## 2. Background

In this section, we give a compact description of the mathematical tools that are necessary for demonstrating the main results of the current manuscript.

### 2.1. Notations

We introduce the notations that will be used throughout the rest of the manuscript. For a vector $v \in \mathbb{R}^{m}, v_{i}$ refers to its $i$ th entry, i.e., $v=\left[v_{i}\right]_{i=1}^{m}$. $\Gamma_{v}=\operatorname{diag}(v)$ denotes the $m \times m$ diagonal matrix, whose diagonal entries are the entries of the vector $v . M_{i j}$ is the entry of the matrix $M$ corresponding to the $i$ th row and the $j$ th column. The eigenvalues of the $m \times m$ square matrix $M$ are denoted by $\lambda_{1}(M), \ldots, \lambda_{m}(M)$. The spectrum of the square matrix $M$ is denoted by $\sigma(M)$, i.e., $\sigma(M)=\left\{\lambda_{i}(M) ; i=1, \ldots, m\right\} . \bar{u}$ is the complex conjugate of the complex number $u \in \mathbb{C} . \Re(u)$ and $\Im(u)$ denote the real and imaginary parts of the complex number $\mathbb{C}$, respectively. $|\mathbb{I}|$ is the cardinality of the set $\mathbb{I}$. Vectors of length $m$ composed entirely of ones and zeros are respectively denoted as $\mathbb{1}_{m}$ and $\mathbb{0}_{m}$. Furthermore, $\mathbb{O}_{m \times n}$ is the $m \times n$ matrix with all its entries set to zero.

### 2.2. Mathematical Models

We outline the process of deriving a mathematical formulation that describes the dynamics of a CRN. Let $\mathcal{X}_{i}, i=1, \ldots, s$, be the set of $s$ distinct chemical species of the considered CRN with $r$ unidirectional reactions $\mathcal{R}_{j}, j=1, \ldots, r$. The connection between the species and the reactions is established through an $s \times r$ stoichiometric matrix denoted as $S$. Its elements are determined as $S_{i j}=\beta_{j i}-\alpha_{j i}$, with $\alpha_{j i}$ representing the number of moles of the $i$ th species $\mathcal{X}_{i}$ within the substrate of the $j$ th reaction $\mathcal{R}_{j}$, and $\beta_{j i}$ indicating the same for the product of the reaction. Denote by $v \in \mathbb{R}_{+}^{r}$ the vector of unidirectional reaction rates. This vector is dependent on both the species' concentration vector $x$ and the parameter vector $\kappa \in \mathbb{R}_{+}^{q}$ inherent to the model. The fundamental framework that characterizes the evolution of the species concentration vector is given by the stoichiometric representation of the balance laws as:

$$
\begin{equation*}
\frac{d x}{d t}=S v(\kappa, x) \tag{1}
\end{equation*}
$$

Graphs are crucial tools for modelling various types of relations and processes in numerous scientific domains including systems biology. In chemical reaction network theory (CRNT), directed graphs are commonly used in the process of modelling CRNs to display the link between individual reactions. The complexes of a CRN are defined as the lefthand (substrate) and right-hand (product) sides of the reactions. Let $\mathcal{C}_{n}, n=1, \ldots, c$, denote the set of distinct complexes of the considered CRN. The complexes can be inherently linked with the vertices of a directed graph, where the directed edges align with the reactions present within the CRN. More precisely, if there is a reaction for which the complex $\mathcal{C}_{i}$ is the substrate and the complex $\mathcal{C}_{j}$ is the product, then in the corresponding graph of complexes there is a directed edge having the vertex associated with the complex $\mathcal{C}_{i}$ as the tail vertex and the vertex associated with the complex $\mathcal{C}_{j}$ as the head vertex. A linkage class of a CRN is a connected component of the corresponding graph of complex, i.e., a maximal set of complexes such that every complex in the set is connected by a directed edge to at least one other complex.

Any CRN can be uniquely described by a system of ODEs given in (1), independent of its governing laws. In this manuscript, we consider only mass action kinetics rate law, since it is the governing law of a wide range of real-life CRNs. According to this rate law, the rate of a reaction is directly proportional to the concentration of each species involved in the substrate of the reaction, raised to a power equal to the number of its moles in
the expression of this substrate. More precisely, the reaction rate $v_{j}$ of the $j$ th reaction, $j=1, \ldots, r$, is given in the following form

$$
\begin{equation*}
v_{j}=k_{j} \prod_{i=1}^{s} x_{i}^{\alpha_{i j}} \tag{2}
\end{equation*}
$$

where, as earlier, $k_{j}, j=1, \ldots, r$, is the rate constant of the $j$ th reaction and $\alpha_{i j}, i=1, \ldots, s$, is the number of moles of the species $\mathcal{X}_{i}$ in the substrate of the $j$ th reaction. Note that in this case, the only parameters contained in the mathematical model are the rate constants of the reactions, i.e., $\kappa=k$ and the number of unknown parameters is $q=r$. Next, we obtain an expression for the vector of reaction rates $v \in \mathbb{R}_{+}^{r}$ given in terms of matrix multiplication, which is a useful approach for automated modelling purposes. Define the $s \times r$ the substrate composition matrix $\Omega$ and the substrate expression function $\varphi: \mathbb{R}_{+}^{s} \rightarrow \mathbb{R}_{+}^{r}$ of the CRN as:

$$
\Omega_{i j}=\left\{\begin{array}{rl}
-S_{i j}, & \text { if } S_{i j}<0 \\
0, & \text { otherwise }
\end{array} \quad \quad \varphi(x)=\left[\prod_{i=1}^{s} x_{i}^{\Omega_{i j}}\right]_{j=1}^{r} .\right.
$$

The conductance matrix $\Gamma_{k}$ of the CRN is a $r \times r$ diagonal matrix whose $i$ th diagonal entry is the rate constant of the $i$ th reaction $\mathcal{R}_{i}$, i.e., if $k \in \mathbb{R}_{+}^{r}$ is the vector of rate constants, then $\Gamma_{k}=\operatorname{diag}(k)$. Then observe that the vector of reaction rates can be expressed in the following matrix multiplication form:

$$
\begin{equation*}
v(k, x)=\Gamma_{k} \varphi(x) . \tag{3}
\end{equation*}
$$

Example 1. To elucidate the outlined modelling process, we apply it to the subsequent example of a CRN. Consider a scenario where five chemical species, denoted as $\mathcal{X}_{i}$, for $i=1, \ldots, 5$, are engaged in three distinct unidirectional reactions, given as follows:

$$
\begin{equation*}
\mathcal{X}_{1}+\mathcal{X}_{2} \rightleftharpoons \mathcal{X}_{3} \longrightarrow \mathcal{X}_{4}+\mathcal{X}_{5} \tag{4}
\end{equation*}
$$

For $i=1, \ldots, 3$, let $k_{i}$ be the rate constant of the $i$ th reaction. Observe that the second reaction can be interpreted as the reverse of the first reaction. In our modelling approach, we treat each reversible reaction as a pair of distinct unidirectional reactions. There are three distinct complexes $\mathcal{C}_{i}$, $i=1, \ldots, 3$, involved in the $C R N$, which are given as $\mathcal{C}_{1}=\mathcal{X}_{1}+\mathcal{X}_{2}, \mathcal{C}_{2}=\mathcal{X}_{3}$, and $\mathcal{C}_{3}=\mathcal{X}_{4}+\mathcal{X}_{5}$. For the first reaction, since the complex $\mathcal{C}_{1}$ is the substrate and the complex $\mathcal{C}_{2}$ is the product, in the graph of complexes corresponding to the CRN shown in (4) there is a directed edge having $\mathcal{C}_{1}$ as the tail vertex and $\mathcal{C}_{2}$ as the head vertex. Similarly, we can construct the edges of the graph of complex corresponding to the other reactions. The resulting graph of complexes is thus $\mathcal{C}_{1} \rightleftharpoons \mathcal{C}_{2} \rightarrow \mathcal{C}_{3}$. Note that this graph of complexes consists of only a single linkage class.

Given the assumption that the reactions (4) are governed by MAKRL, the reaction rates can be computed by Equation (2) as follows:

$$
v_{1}=k_{1} x_{1} x_{2}, \quad v_{2}=k_{2} x_{3}, \quad v_{3}=k_{3} x_{3}
$$

The vector of reaction rates can be written in the matrix multiplication form (3) with the substrate composition matrix $\Omega$ and the substrate expression function $\varphi$ given by

$$
\Omega=\left[\begin{array}{ccc}
1 & 0 & 0 \\
1 & 0 & 0 \\
0 & 1 & 1 \\
0 & 0 & 0 \\
0 & 0 & 0
\end{array}\right], \quad \varphi(x)=\left[\begin{array}{c}
x_{1} x_{2} \\
x_{3} \\
x_{3}
\end{array}\right]
$$

Thus, in this case the balance laws (1) can be written as:

$$
\begin{align*}
& \frac{d x_{1}}{d t}=\frac{d x_{2}}{d t}=-k_{1} x_{1} x_{2}+k_{2} x_{3} \\
& \frac{d x_{3}}{d t}=k_{1} x_{1} x_{2}-k_{2} x_{3}-k_{3} x_{3}  \tag{5}\\
& \frac{d x_{4}}{d t}=\frac{d x_{5}}{d t}=k_{3} x_{3}
\end{align*}
$$

### 2.3. The Weighted Directed Laplacian Matrix

In CRNT, the (weighted directed) Laplacian matrix is a matrix representation of the reactions occurring between the different complexes. Here we explain how to construct the Laplacian matrix of the CRN using the (weighted directed) adjacency matrix corresponding to its graph of complexes. The adjacency matrix $A$ is a $c \times c$ matrix, with $c$ being the number of complexes of the CRN, such that its entry $A_{i j}$ is equal to $k$ if there is a reaction having the $j$ th complex of the network as substrate and $i$ th complex of the network as a product with $k$ being the rate constant of the reaction. The (weighted directed) degree matrix $D$ of the graph of complexes corresponding to a CRN is a $c \times c$ diagonal matrix such that its $i$ th diagonal entry is equal to the sum of the elements of the $i$ th column of the weighted adjacency matrix $A$. The $c \times c$ Laplacian matrix associated with the graph of complexes of the considered CRN is defined as follows:

$$
L=D-A .
$$

If there is a reaction for which the complex $\mathcal{C}_{j}$ is the substrate and the complex $\mathcal{C}_{i}$ is the product, then the off-diagonal element $L_{i j}$ is equal to the rate constant of the respective reaction taken with the negative sign. For useful properties of Laplacian matrices, we refer to [35]. Any directed graph is defined by an incidence matrix [36], which represents the connections between its vertices and edges. In the case of CRNs, the $c \times r$ incidence matrix $B$ of the graph of complexes is defined as follows:

$$
B_{i j}=\left\{\begin{align*}
-1, & \text { if the complex } \mathcal{C}_{i} \text { is the substrate of the reaction } \mathcal{R}_{j}  \tag{6}\\
1, & \text { if the complex } \mathcal{C}_{i} \text { is the product of the reaction } \mathcal{R}_{j} \\
0, & \text { otherwise. }
\end{align*}\right.
$$

Define the $c \times r$ outgoing matrix $\Delta$ of the considered CRN as follows:

$$
\Delta_{i j}= \begin{cases}0, & \text { if } B_{i j}=1 \\ B_{i j}, & \text { otherwise }\end{cases}
$$

It can be shown that the Laplacian is given in matrix multiplication form as follows:

$$
\begin{equation*}
L=B \Gamma_{k} \Delta^{\top} . \tag{7}
\end{equation*}
$$

For automatic modelling purposes, it is convenient to construct the Laplacian matrix using the simple matrix multiplication form given in (7).

Next, we show how to represent the balance laws (1) in terms of the weighted directed Laplacian matrix $L$. The $c$ complexes of the considered CRN are described by an $s \times c$ complex composition matrix $Z$, whose columns express the complexes of the CRN in terms of their species. More precisely, the element $Z_{i j}$ of the complex composition matrix $Z$ is the number of moles of the $i$ th species $\mathcal{X}_{i}$ in the expression of $j$ th complex $\mathcal{C}_{j}$. As explained in [29,37], it can be shown that the balance laws of a mass action CRN can be rewritten as follows:

$$
\begin{equation*}
\frac{d x}{d t}=-Z L \psi(x) \tag{8}
\end{equation*}
$$

where $\psi: \mathbb{R}_{+}^{s} \rightarrow \mathbb{R}_{+}^{c}$ is the complex expression function defined as:

$$
\psi(x)=\left[\prod_{i=1}^{s} x_{i} Z_{i j}\right]_{j=1}^{c}
$$

Example 2. With reference to the CRN example (4), the $3 \times 3$ weighted adjacency matrix $A$ and the $3 \times 3$ weighted degree matrix $D$ are:

$$
A=\left[\begin{array}{ccc}
0 & k_{2} & 0 \\
k_{1} & 0 & 0 \\
0 & k_{3} & 0
\end{array}\right], \quad D=\left[\begin{array}{ccc}
k_{1} & 0 & 0 \\
0 & k_{2}+k_{3} & 0 \\
0 & 0 & 0
\end{array}\right]
$$

The $3 \times 3$ weighted directed Laplacian L of the CRN (4) is therefore given by

$$
L=\left[\begin{array}{rcc}
k_{1} & -k_{2} & 0 \\
-k_{1} & k_{2}+k_{3} & 0 \\
0 & -k_{3} & 0
\end{array}\right]
$$

On the other hand, the Laplacian matrix can be computed using Equation (7) with the incidence matrix $B$ and the outgoing matrix $\Delta$ given by:

$$
B=\left[\begin{array}{rrr}
-1 & 1 & 0 \\
1 & -1 & -1 \\
0 & 0 & 1
\end{array}\right], \quad \Delta=\left[\begin{array}{rrr}
-1 & 0 & 0 \\
0 & -1 & -1 \\
0 & 0 & 0
\end{array}\right]
$$

The balance laws (5) can be rewritten as Equation (8) with the $5 \times 3$ complex composition matrix $Z$ and the complex expression function $\psi$ given by:

$$
Z=\left[\begin{array}{lll}
1 & 0 & 0 \\
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1 \\
0 & 0 & 1
\end{array}\right], \quad \psi(x)=\left[\begin{array}{c}
x_{1} x_{2} \\
x_{3} \\
x_{4} x_{5}
\end{array}\right]
$$

In the following theorem we recall certain important spectral properties (see, e.g., [38]) of the weighted directed Laplacian matrix associated with the graph of complexes corresponding to a CRN governed by MAKRL.

Theorem 1 (Spectrum of the weighted directed Laplacian matrix). If the graph of complexes of a CRN governed by MAKRL has a single linkage class, then the eigenvalues $\lambda_{i}, i=1, \ldots, c$, of the weighted directed Laplacian matrix L associated with the CRN can be ordered in such a way that:

$$
\begin{equation*}
0=\lambda_{1}<\Re\left(\lambda_{2}\right) \leq \Re\left(\lambda_{3}\right) \leq \ldots \leq \Re\left(\lambda_{c}\right) \tag{9}
\end{equation*}
$$

First note that $0 \in \sigma(L)$. This is simply because of the fact that the sum of each column of $L$ is equal to zero according to the definition of the Laplacian matrix, i.e., $L^{\top} \mathbb{1}_{c}=\mathbb{D}_{c}$. Moreover, from the generalization of the matrix-tree theorem it follows that the multiplicity of the zero eigenvalue is equal to the number of connected components, which is $c-\operatorname{rank}(L)$. Using Greshgorin's circle theorem [39], it can be shown that the real parts of non-zero complex eigenvalues of $L$ are strictly positive, i.e., if $\lambda \in \sigma(L)$ and $\lambda \neq 0$, then $\Re(\lambda)>0$. For a detailed explanation of the proof of Theorem 1 we refer to [40].

### 2.4. Kron Reduction of Chemical Reaction Networks

The Kron reduction for mathematical models of CRNs [29,37,41] is performed by assuming that certain intermediate complexes are complex balanced and is carried out by computing the Schur complement of the weighted Laplacian matrix associated with
the corresponding graph of complexes. We therefore first recall the definition of Schur complements (see, e.g., [42]) of a given square matrix.

Definition 1 (Schur complement). Let $A_{1} \in \mathbb{R}^{n \times n}, A_{2} \in \mathbb{R}^{n \times m}, A_{3} \in \mathbb{R}^{m \times n}$, and $A_{4} \in \mathbb{R}^{m \times m}$ be constant matrices such that the latter is invertible. Consider the following $(n+m) \times(n+m)$ block matrix:

$$
L=\left[\begin{array}{ll}
A_{1} & A_{2} \\
A_{3} & A_{4}
\end{array}\right]
$$

The Schur complement of the block matrix $A_{4}$ is the $n \times n$ matrix $\widehat{L}$ defined as:

$$
\widehat{L}=A_{1}-A_{2} A_{4}^{-1} A_{3} .
$$

Let $\mathbb{I}$ be the set of indices corresponding to the complexes of the $C R N$, i.e., $\mathbb{I}=\{1, \cdots, c\}$. Suppose our objective involves removing the complexes associated with the subset of indices denoted as $\overline{\mathbb{I}}$. Note that it should be ensured that $|\overline{\bar{I}}|<c$. The removal of complexes is accomplished through the computation of the Schur complement $\widehat{L} \in \mathbb{R}^{|\hat{I}| \times|\hat{\mathbb{I}}|}$ of the block matrix of the Laplacian matrix $L$ corresponding to the set of indices $\overline{\mathbb{I}}$. Here, $\widehat{\mathbb{I}}=\mathbb{I} \backslash \overline{\mathbb{I}}$ is the set of indices corresponding to the complexes remaining in the reduced graph of complexes. $\widehat{L}$ is again a Laplacian matrix since it satisfies the properties of Laplacian matrices ([29], Proposition 1). Furthermore, it has been proven that the equation

$$
\frac{d y}{d t}=-\widehat{Z} \widehat{L} \widehat{\psi}(y)
$$

describes the dynamics of a CRN governed by MAKRL, with a smaller number of complexes. Here, $y \in \mathbb{R}_{+}^{s}$ is the vector of species' concentrations in the reduced mathematical model (which contains a subset of the elements of $x$.), $\widehat{Z}$ is the complex composition matrix of the reduced CRN, and

$$
\widehat{\psi}(y)=\left[\prod_{i=1}^{s} y_{i}^{\widehat{Z}_{i j}}\right]_{j=1}^{|\widehat{\mathbb{I}}|} .
$$

As explained in $[29,37]$, a well-chosen $\overline{\mathbb{I}}$ will result in a reduction of dependent variables within the corresponding mathematical model. Note that the parameters contained in the Kron-reduced mathematical model can be represented as a function of the parameters involved in the original model. More precisely, if $p \in \mathbb{R}_{+}^{\widehat{r}}$ denotes the vector of parameters of the reduced model with $\widehat{r}$ being the number of reactions in it, then there is a function $f: \mathbb{R}_{+}^{r} \rightarrow \mathbb{R}_{+}^{\widehat{r}}$ such that $p=f(k)$. In general, the manual derivation of the explicit form of the function $f$ is not straightforward. However, we use MATLAB symbolic variables to derive the explicit form of $f$ in a fully automated fashion. We refer to the function $f$ as the parameter dependence function, since it specifies the dependence of the vector of parameters $p$ of the reduced model on the vector of parameters $k$ of the original model.

In order to determine the structure of the reduced CRN, we need to find the incidence matrix and the complex composition matrix of the reduced network. This can be done according to the automated procedure described in [37]. The incidence matrix $\widehat{B}$ is determined by making use of its Laplacian matrix $\widehat{L}$. According to this procedure, if $\widehat{L}_{i j}<0$, $i \neq j$, then in the reduced graph of complexes there is a reaction for which the $j$ th complex of the reduced CRN is the substrate and $i$ th complex of the reduced CRN is the product complex. Therefore, the entries of the incidence matrix $\widehat{B}$ of the reduced graph of complexes are defined according to (6). The complex composition matrix $\widehat{Z}$ of the reduced CRN is obtained by simply removing the columns of the incidence matrix $Z$ of the original CRN that correspond to the set of indices $\overline{\mathbb{I}}$. As mentioned earlier, the incidence matrix describes the reactions occurring between the complexes and the complex composition matrix gives the expression of complexes in terms of the species. We therefore use $\widehat{B}$ and $\widehat{Z}$ to determine the reactions corresponding to the reduced graph of complexes.

Example 3. To illustrate the Kron reduction method for CRNs, we demonstrate it for the example given in (4). Assume that we want to delete the complex $\mathcal{C}_{2}$ from the graph of complexes by applying the Kron reduction method. In other words, $\overline{\mathbb{I}}=\{2\}$ and $\widehat{\mathbb{I}}=\{1,3\}$. The elimination of $\mathcal{C}_{2}$ is carried out by computing the Schur complement of the Laplacian matrix corresponding to the set of indices $\overline{\mathbb{I}}$, which results in the following $2 \times 2$ weighted directed Laplacian matrix associated with the reduced graph of complexes:

$$
\widehat{L}=\left[\begin{array}{cc}
\frac{k_{1} k_{3}}{k_{2}+k_{3}} & 0 \\
-\frac{k_{1} k_{3}}{k_{2}+k_{3}} & 0
\end{array}\right]
$$

As explained in [37], the complex composition matrix $\widehat{Z}$ of the reduced $C R N$ is obtained by eliminating the second column of the complex composition matrix Z of the original CRN, i.e.,

$$
\widehat{Z}=\left[\begin{array}{ll}
1 & 0 \\
1 & 0 \\
0 & 0 \\
0 & 1 \\
0 & 1
\end{array}\right]
$$

Thus the balance laws of the Kron-reduced model are as follows:

$$
\begin{aligned}
& \frac{d y_{1}}{d t}=\frac{d y_{2}}{d t}=-\frac{k_{1} k_{3}}{k_{2}+k_{3}} y_{1} y_{2} \\
& \frac{d y_{3}}{d t}=0 \\
& \frac{d y_{4}}{d t}=\frac{d y_{5}}{d t}=\frac{k_{1} k_{3}}{k_{2}+k_{3}} y_{1} y_{2} .
\end{aligned}
$$

Using the Laplacian matrix $\widehat{L}$ of the Kron-reduced model we obtain the incidence matrix $\widehat{B}$ of the reduced complex graph:

$$
\widehat{B}=\left[\begin{array}{r}
-1 \\
1
\end{array}\right] .
$$

Taking into account $\widehat{Z}$ and $\widehat{B}$ we derive the reactions of the reduced $C R N$ :

$$
\mathcal{X}_{1}+\mathcal{X}_{2} \xrightarrow{p} \mathcal{X}_{4}+\mathcal{X}_{5}
$$

where the parameter $p$ is given in terms of the parameters of the original model as $p=\frac{k_{1} k_{3}}{k_{2}+k_{3}}$, i.e., in this case for the explicit form of the function $f$ we have $f(k)=\frac{k_{1} k_{3}}{k_{2}+k_{3}}$. Note that after deleting the complex $\mathcal{C}_{2}$ from the graph of complexes by the Kron reduction approach, the species $\mathcal{X}_{3}$ is not involved in the resulting reduced $C R N$, since its concentration $x_{3}$ is conserved in time.

In $[29,37]$, the optimal combination of complexes for deletion is selected by making use of an error integral, which quantifies the difference between the dynamical behaviors of the original model and the corresponding reduced model. This error integral is measure that is based on a particular trajectory corresponding to the original model.

### 2.5. The Least Squares Optimization Method

We explain how to apply the (weighted) least squares optimization technique to estimate the parameters $\kappa \in \mathbb{R}_{+}^{p}$ involved in the kinetic model (1) describing the dynamics of the given CRN. For the general least squares optimization method we refer to, for example, $[3,4]$. This optimization method plays a crucial role in our parameter estimation method.

Assume that biological experiments provide complete experimental data of species' concentrations, i.e., measurements that correspond to all of the species' concentrations. For $j=1, \ldots, n$, let $\widehat{x}_{i, m}^{(j)}, i=1, \ldots, s ; m=0, \ldots, N_{j}$, be the observed value of the $i$ th concentration at time instant $t_{m}^{(j)}$, which is the $m$ th time point corresponding to the $j$ th experiment. We aim to identify the best-fitting parameter values of the mathematical model (1) corresponding to the above-mentioned observed time-series experimental data of species' concentrations. For every $j=1, \ldots, n$, consider the following initial value problem (IVP):

$$
\left\{\begin{array}{l}
\frac{d x}{d t}=S v(\kappa, x)  \tag{10}\\
x\left(t_{0}^{(j)}\right)=\widehat{x}_{i, 0}^{(j)}
\end{array}\right.
$$

Since the available experimental data corresponds to the measurements of all the species' concentrations, for every parameter vector $\kappa \in \mathbb{R}_{+}^{p}$ the IVP given in (10) generates data for the species' concentrations. More precisely, the IVP (10) can be numerically solved with respect to time. However, this numerical integration is not always possible if the available experimental data corresponds to only some of the concentrations.

For every $j=1, \ldots, n$, let $x_{i}\left(t_{m}^{(j)}, \kappa\right), i=1, \ldots, s ; m=0, \ldots, N_{j}$, denote the modelpredicted value of the $i$ th concentration obtained by numerically solving the IVP (10). In this case, the least squares error is defined as the sum of squared residuals, which are the differences between the observed experimental values of concentrations and the corresponding model-predicted values provided from the IVP given in (10):

$$
\begin{equation*}
\varepsilon(\kappa)=\sum_{j=1}^{n} \sum_{i=1}^{s} w_{i}^{(j)} \sum_{m=0}^{N_{j}}\left(x_{i}\left(t_{m}^{(j)}, \kappa\right)-\widehat{x}_{i, m}^{(j)}\right)^{2} . \tag{11}
\end{equation*}
$$

Here, for $i=1, \ldots, s$ and $j=1, \ldots, n, w_{i}^{(j)}$ denotes the weight of the corresponding measurement. In this case, it is assumed that the measurements have different uncertainties. Each weight can be taken, for example, equal to the reciprocal of the variance of the measurement:

$$
w_{i}^{(j)}=\frac{1}{\left[\sigma_{i}^{(j)}\right]^{2}}, \quad i=1, \ldots, s ; j=1, \ldots, n
$$

We will refer to this approach as the weighted least squares (WLS). If the measurements have equal variance, then the weights can be taken equal to one. We will refer to the corresponding approach as the unweighted least squares (UWLS). The (weighted) least squares optimization technique finds the optimal parameter values by minimizing the error (11). This minimization can be done, for example, by the standard Levenberg-Marquardt algorithm [43,44], or the modified Levenberg-Marquardt algorithm [45]. We denote by $\widehat{\kappa} \in \mathbb{R}_{+}^{p}$ the solution to the aforementioned optimization problem, i.e., $\widehat{\kappa}=\arg \min _{\kappa} \varepsilon(\kappa)$.

## 3. Parameter Identifiability

In this section, we first recall the definitions of least squares parameter identifiability and parameter identifiability, and in addition demonstrate the link between these two identifiability concepts. Consider a dynamical system described by a system of ODEs that is given in following the form:

$$
\left\{\begin{array}{l}
\frac{d x}{d t}=f(\kappa, x)  \tag{12}\\
y=g(\kappa, x)
\end{array}\right.
$$

where $f$ is an $s$-dimensional vector-valued function depending on the structure of the system and $g$ is $n$-dimensional. Here, $\kappa \in \mathbb{R}_{+}^{p}$ is the parameter vector, $x: \mathbb{R}_{+} \rightarrow \mathbb{R}_{+}^{s}$ is the vector of states of the system and $y$ is the $n$-dimensional output vector. For a given vector of initial states $x_{0} \in \mathbb{R}_{+}^{s}$, let $y\left(t \mid \kappa, x_{0}\right)$ denote the output trajectory of the system (12) corresponding to the parameter vector $\kappa \in \mathbb{R}_{+}^{p}$ and initial states $x_{0} \in \mathbb{R}_{+}^{s}$.

Assume that, in an experimental setup, the output has been continuously measured over the time interval $[0, T]$, for some pre-specified $T>0$. Let $\hat{y}: \mathbb{R}_{+} \rightarrow \mathbb{R}^{n}$ be the resulting measured output. Consider the cost function $\varepsilon_{x_{0}}: \mathbb{R}_{+}^{p} \rightarrow \mathbb{R}_{+}$defined as:

$$
\begin{equation*}
\varepsilon_{x_{0}}(\kappa)=\sum_{i=1}^{n} \int_{0}^{T}\left(y_{i}\left(t \mid \kappa, x_{0}\right)-\widehat{y}_{i}(t)\right)^{2} d t \tag{13}
\end{equation*}
$$

We recall the definition of least squares parameter identifiability of dynamical systems given in the form (12), which was first introduced in [46].

Definition 2 (Least squares parameter identifiability). The dynamical system (12) is least squares parameter identifiable, if for every given vector of initial states $x_{0}$ and for every given measurement function $\widehat{y}$, the cost function (13) admits a unique minimum.

If there exists at least one vector of initial states $x_{0}$ and a measurement function $\widehat{y}$ such that the cost function (13) has multiple minima, then the dynamical system (12) is least squares parameter nonidentifiable.

Each vector of parameters $\kappa \in \mathbb{R}_{+}^{p}$ determines a set $\mathcal{Y}_{\kappa}$ of admissible output trajectories of the system (12). We recall the definition of parameter identifiability of dynamical systems provided in [14].

Definition 3 (Parameter identifiability). The dynamical system (12) is parameter identifiable, if for parameter vectors $\kappa, \bar{\kappa} \in \mathbb{R}_{+}^{p}$ such that $\kappa \neq \bar{\kappa}$, we have $\mathcal{Y}_{\kappa} \neq \mathcal{Y}_{\bar{\kappa}}$.

Equivalently, the dynamical system (12) is parameter identifiable, if $\mathcal{Y}_{\kappa}=\mathcal{Y}_{\bar{\kappa}}$ for two parameter vectors $\kappa, \bar{\kappa} \in \mathbb{R}_{+}^{p}$, implies $\kappa=\bar{\kappa}$. If there are two distinct parameter vectors $\kappa \neq \bar{\kappa}$ for which $\mathcal{Y}_{\kappa}=\mathcal{Y}_{\bar{\kappa}}$, then the dynamical system (12) is parameter unidentifiable.

We finally turn our attention to the main contribution of this section. In the following theorem we specify a link between the two identifiability concepts given above.

Theorem 2. If the dynamical system (12) is parameter unidentifiable, then it is also least squares parameter unidentifiable.

Proof. We assume that the dynamical system (12) is parameter unidentifiable and we prove that it is also least squares parameter unidentifiable. Since the dynamical system (12) is parameter unidentifiable, there are two parameter vectors $\kappa, \bar{\kappa} \in \mathbb{R}_{+}^{p}$, such that $\kappa \neq \bar{\kappa}$ and $\mathcal{Y}_{\kappa}=\mathcal{Y}_{\bar{\kappa}}$. This means that there is at least one vector of initial states $x_{0} \in \mathbb{R}_{+}^{s}$ for which $y\left(t \mid \kappa, x_{0}\right)=y\left(t \mid \bar{\kappa}, x_{0}\right), t \in[0, T]$. Choose a measurement function $\widehat{y}: \mathbb{R}_{+} \rightarrow \mathbb{R}^{n}$ as:

$$
\widehat{y}(t):=y\left(t \mid \kappa, x_{0}\right)=y\left(t \mid \bar{\kappa}, x_{0}\right), \quad t \in[0, T] .
$$

Note that this choice results in $\varepsilon_{x_{0}}(\kappa)=\varepsilon_{x_{0}}(\bar{\kappa})=0$. For two parameter vectors $\kappa \neq \bar{\kappa}$ there is a vector of initial states $x_{0} \in \mathbb{R}_{+}^{s}$ and a measurement function $\widehat{y}: \mathbb{R}_{+} \rightarrow \mathbb{R}^{n}$ such that $\varepsilon_{x_{0}}(\kappa)=\varepsilon_{x_{0}}(\bar{\kappa})=0$. Note that zero is the minimum value of $\varepsilon_{x_{0}}$ since it is a non-negative function. We conclude that the minimum of $\varepsilon_{x_{0}}$ is not unique in this case and thus the dynamical system (12) is least squares parameter unidentifiable.

## 4. Parameter Estimation Procedure

In this section, we describe the main contribution of the current manuscript. We show how to use the Kron reduction method for kinetic models described in the previous section,
as a tool for estimating the parameters involved in the corresponding mathematical model from time-series partial experimental data of species' concentrations.

### 4.1. 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 some of the species' concentrations are measured. Suppose that the output of a biological experiment corresponding to the mathematical model (8) is of the form:

$$
z=H x
$$

where $H \in \mathbb{R}^{n \times s}$ is a constant matrix known as the measurement matrix. In general, $H$ can have an arbitrary structure. However, in general only some of the species' concentrations are measured experimentally. Thus, we assume that each row of $H$ has only one non-zero element, which is equal to one and is placed in the position corresponding to the particular species. For instance, with reference to the CRN (4), if the species that are measured experimentally are $\mathcal{X}_{1}, \mathcal{X}_{2}, \mathcal{X}_{4}$ and $\mathcal{X}_{5}$, then

$$
H=\left[\begin{array}{lllll}
1 & 0 & 0 & 0 & 0  \tag{14}\\
0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1
\end{array}\right]
$$

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

$$
\begin{equation*}
\Lambda^{(j)}=\left\{\left(t_{m}^{(j)}, \widehat{z}_{i, m}^{(j)}\right) \mid i=1, \ldots, n ; m=0, \ldots, N_{j}\right\}, \quad j=1, \ldots, l . \tag{15}
\end{equation*}
$$

We aim to identify the best-fitting parameters corresponding to the mathematical model (8) of the considered CRN from observed time-series data $\Lambda^{(j)}, j=1, \ldots, l$, collected from biological experiments.

### 4.2. Trajectory-Independent Error

We propose a trajectory-independent alternative method to the computation of the error integral defined in [29] for comparing the dynamics of the Kron-reduced model to the one of the original mathematical model. In our parameter estimation procedure, this error will be used as an objective function for minimization. It is based on matching the so called relaxation constant of the Kron-reduced mathematical model to the one of the original mathematical model. We define the relaxation constant $\tau(L)$ of a CRN as the smallest non-zero real part of the eigenvalues of its Laplacian matrix, i.e.,

$$
\tau(L)=\min _{\substack{\lambda \in \sigma(L) \\ \lambda \neq 0}} \Re(\lambda)
$$

The quantity $\tau(L)$ illustrates a key characteristic property of the CRN that is represented by the Laplacian matrix $L$. In the case when each complex is a single-species complex, i.e., $\mathrm{Z}=I_{s}$, it can be shown that relaxation constant $\tau(L)$ is an exact indicator of the settlingtime of the CRN, i.e., the time instant after which the concentrations of all the species fall and remain within some specified percentage of their corresponding steady-state values. This case is encountered commonly as in one of our demonstrative real-life examples
considered in the next section. In this case, the solution to the balance laws (8) is of the form:

$$
\begin{equation*}
x(t)=\sum_{i=1}^{q} \sum_{j=0}^{p_{i}-1} t^{j} w^{(i j)}(t) e^{-\Re\left(\lambda_{i}\right) t} \tag{16}
\end{equation*}
$$

where $p_{i}$ is the multiplicity of the eigenvalue $\lambda_{i}$ and

$$
w^{(i j)}=b^{(i j)} \cos \left(\Im\left(\lambda_{i}\right) t\right)+c^{(i j)} \sin \left(\Im\left(\lambda_{i}\right) t\right) \quad \text { with } \quad a, b^{(i j)}, c^{(i j)} \in \mathbb{R}^{s}
$$

Since the real parts of eigenvalues of the Laplacian matrix $L$ are all nonnegative, from (16) it follows that the settling-time of the CRN is determined by the slowest decaying term within the summation sign in the right hand side of the equation. Observe that for every $i=1, \ldots, q$ the terms $t^{j} w^{(i j)}(t) e^{-\Re\left(\lambda_{i}\right) t}, j=1, \ldots, p_{i}-1$, decay faster than the term $w^{(i 0)}(t) e^{-\Re\left(\lambda_{i}\right) t}$ whose time constant is equal to $\Re\left(\lambda_{i}\right)^{-1}$. Among the terms $w^{(i 0)}(t) e^{-\Re\left(\lambda_{i}\right) t}$, the slowest decaying term is the one with the biggest time constant, i.e., the one with the smallest value of $\Re\left(\lambda_{i}\right)$. Thus, for comparing the settling-time of the Kron-reduced model with the one of the original model it is reasonable to compare the relaxation constant $\tau(\widehat{L})$ of the Kron-reduced model with the relaxation constant $\tau(L)$ of the original mathematical model. In general, irrespective of whether the complex composition matrix $Z$ is the identity matrix or not, we quantify the difference between the dynamics of the original mathematical model and the one of the Kron-reduced model using the trajectory-independent spectral based error $\delta: \mathbb{R}_{+}^{r} \rightarrow \mathbb{R}_{+}$defined as:

$$
\begin{equation*}
\delta(k)=|\tau(L)-\tau(\widehat{L})| . \tag{17}
\end{equation*}
$$

Note that the spectral based error $\delta$ defined in (17) is a function that only depends on the vector of parameters $k$ contained in the original model. This is due to the fact that the entries as well as the eigenvalues of both the original Laplacian matrix $L$ and the Kron-reduced Laplacian matrix $\widehat{L}$ are functions of the parameter vector $k$.

Remark 1. In the physical sciences, the term "relaxation" usually refers to the return of a system to equilibrium. Each relaxation process can be categorized by a relaxation time (see, e.g., [47-49]). Since the quantity $\tau(L)$ is an indicator of the settling time of the CRN, i.e., its relaxation, we refer to the constant $\tau(L)$ as the relaxation constant of the CRN described by the Laplacian matrix $L$.

### 4.3. Estimation Procedure

We turn our attention to the parameter estimation procedure. We assume that in an experimental setup certain measurements $\Lambda^{(j)}, j=1, \ldots, l$, of the form given in (15) are collected. The principal goal is to find estimates for the parameters $k$ of the original model (8) such that the corresponding model-predicted values fit the available time-series partial experimental data (15).

As explained in Section 2.5, since we do not have complete data on species' concentrations, direct estimation of the parameters involved in the balance laws (8) is not possible. In order to convert the problem to a well-posed parameter estimation problem, we first derive the Kron-reduced mathematical model obtained after deleting a certain subset of complexes from the corresponding graph of complexes. For this purpose, we identify this subset of complexes by making use of the complex composition matrix $Z$ and the measurement matrix $H$. The latter specifies the set of indices $\mathbb{J}$ corresponding to the species that are measured experimentally. For instance, if the measurement matrix $H$ is the one given in (14), then the above-mentioned set of indices is simply $\mathbb{J}=\{1,2,4,5\}$, i.e., only the species $\mathcal{X}_{i}$, $i=1,2,4,5$, are measured experimentally. Furthermore, the complex composition matrix $Z$ identifies the set of complexes $\overline{\mathbb{I}}$ that have at least one unmeasured species. We now delete the complexes corresponding to the set of indices $\overline{\mathbb{I}}$ using the Kron reduction method. The $\widehat{s} \times \widehat{c}$ complex composition matrix $\widehat{Z}$ and the $\widehat{c} \times \widehat{r}$ incidence matrix $\widehat{B}$ are computed using the procedure explained earlier.

Subsequently, we determine the parameter dependence function $f: \mathbb{R}_{+}^{r} \rightarrow \mathbb{R}_{+}^{\hat{r}}$. In order to express the vector of parameters $p \in \mathbb{R}_{+}^{\hat{r}}$ of the Kron-reduced model as a function of the vector of parameters $k \in \mathbb{R}_{+}^{r}$ of the original model, i.e $p=f(k)$. In general, the explicit form of this function is complicated and its manual derivation is not straightforward. However, the usage of MATLAB symbolic variables allows us to derive the explicit form of the parameter dependence function $f$. Using the properties of the Kron reduction method for CRNs governed by mass action kinetics rate law, we conclude that the dependence function is a vector-valued rational function of its argument.

We now consider the parameter estimation problem for the Kron-reduced mathematical model from the available observed time-series experimental data of species' concentrations (15). It is a well-posed parameter estimation problem since the available experimental data corresponds to all the concentrations involved in the Kron-reduced mathematical model. We apply the least squares optimisation technique to determine the best-fitting values of parameters $\hat{p} \in \mathbb{R}^{\widehat{\gamma}}$, i.e., the parameter values for which the available observed time-series experimental data $\Lambda^{(j)}, j=1, \ldots, l$ given in (15) fits the Kron-reduced model.

Finally, we determine the values of parameters $\widehat{k} \in \mathbb{R}^{r}$ for which the trajectoryindependent spectral-based error function $\delta$ given in (17) admits its minimum value subject to the constraint

$$
\begin{equation*}
f(k)=\widehat{p} \tag{18}
\end{equation*}
$$

In other words, we have the following well-posed constrained optimisation problem

$$
\min _{k \in \mathbb{R}_{+}^{r}} \delta(k) \text { subject to } \quad f(k)=\widehat{p} .
$$

We use the method of Lagrange multipliers to solve the above-mentioned problem using MATLAB Optimization Toolbox. We developed a MATLAB library for the automation of our estimation procedure. The inputs required for our parameter estimation procedure are the $s \times c$ complex composition matrix $Z$, the $c \times r$ incidence matrix $B$, the $n \times s$ measurement matrix $H$, and the observed time-series experimental datasets $\Lambda^{(j)}, j=1, \ldots, l$. The output of the parameter estimation method is the vector $\widehat{k} \in \mathbb{R}_{+}^{r}$ of the best-fitting parameter values corresponding to the experimental time-series observed data given in (15).

## 5. Application to Real-Life Examples

We demonstrate the applicability of our automated parameter estimation algorithm from observed time-series partial experimental data of species' concentrations on two real-life computational models of biological processes retrieved from the BioModels database [32]. We consider a model of nicotinic acetylcholine receptors [33] and a model of Trypanosoma brucei trypanothione synthetase [34]. For each of these models, we first generate partial time-series data corresponding to the species' concentrations using the values of parameters provided in the corresponding reference. Next, we perturb these generated data with 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. Nicotinic Acetylcholine Receptors

We consider a model of nicotinic acetylcholine receptors (NAR) developed in [33]. Nicotinic receptors are receptor polypeptides (short chains of amino acids) that respond to the neurotransmitter acetylcholine (a signalling molecule secreted by a neuron) as well as to drugs such as the agonist nicotine. They are found in the central and peripheral nervous systems, muscles, and several other tissues of different organisms. A schematic representation of the network corresponding to the mathematical model of NAR is provided in the left-hand panel of Figure 1. A detailed description of the mathematical model can be found in [33].



Figure 1. Schematic representation of the original model (left-hand panel) of nicotinic acetylcholine receptors and the corresponding Kron-reduced model (right-hand panel) obtained by deleting the single-species complexes BL, AL, IL, DL, BLL, ALL, ILL, and DLL from the graph of complexes.

### 5.1.1. The Considered Model and the Available Data

In the considered model of NAR, there are 12 chemical species participating in 17 reversible mass action reactions. All complexes here are single-species complexes. There are thus 12 complexes in the corresponding graph of complexes. As mentioned earlier, for the purpose of automated modelling purposes we regard each reversible reaction as two separate unidirectional reactions. Table 1 provides an overview of the primary compounds participating in the reaction system.

Table 1. An overview of the primary compounds participating in the considered model of the nicotinic acetylcholine receptors.

| Species Notation | Species Name | Species ID |
| :---: | :---: | :---: |
| A | State of higher affinities with <br> open channel | Active |
| B | Activatable resting closed state | Basal |
| I | Inactivatable state of higher <br> affinities with closed channel | Inactivatable |
| D | Desensitised state of higher <br> affinities with closed channel | Desensitised |

The rest of the compounds are intermediate enzyme complexes participating in the reaction network. The network consists of the following reactions, all of which are reversible.
B $\underset{k_{2}}{\stackrel{k_{1}}{\rightleftharpoons}}$ BL,
BL $\underset{k_{4}}{\stackrel{k_{3}}{\rightleftharpoons}}$ BLL,
B $\underset{k_{12}}{\stackrel{k_{11}}{\rightleftharpoons}} \mathrm{~A}$,
BLL $\underset{k_{6}}{\stackrel{k_{4}}{k_{5}}}$ ALL,
BL $\underset{k_{14}}{\stackrel{k_{13}}{\rightleftharpoons}} \mathrm{AL}$,
$\mathrm{A} \underset{k_{8}}{\stackrel{k_{7}}{\rightleftharpoons}} \mathrm{AL}$,
$\mathrm{AL} \underset{k_{10}}{\stackrel{k_{9}}{\rightleftharpoons}}$ ALL,
I $\underset{k_{16}}{\stackrel{k_{15}}{\rightleftharpoons}}$ IL,
IL $\underset{k_{18}}{\stackrel{k_{17}}{\rightleftharpoons}}$ ILL,
A $\underset{k_{20}}{\stackrel{k_{19}}{\rightleftharpoons}} \mathrm{I}$,
AL $\underset{k_{22}}{\stackrel{k_{21}}{\rightleftharpoons}}$ IL,
ALL $\underset{k_{24}}{\stackrel{k_{23}}{\rightleftharpoons}}$ ILL,
$\mathrm{I} \underset{k_{30}}{\stackrel{k_{29}}{\rightleftharpoons}} \mathrm{D}$,
IL $\underset{k_{32}}{\stackrel{k_{31}}{\rightleftharpoons}} \mathrm{DL}$,
D $\underset{k_{26}}{\stackrel{k_{25}}{\rightleftharpoons}}$ DL,
ILL $\underset{k_{34}}{\stackrel{k_{33}}{\rightleftharpoons}}$ DLL.
DL $\underset{k_{28}}{\stackrel{k_{27}}{\rightleftharpoons}}$ DLL,

The values of the parameters $k_{i}, i=1, \ldots, 34$, provided in [33] are given in Table 2. We assume that the species that are measured in an experimental setup are the ones having a greater impact on the reactions, which are the activatable resting closed state B (the Basal state), the state of higher affinities with open channel A (the Active state), the states of higher affinities with closed channels I (the Inactivatable state) and D (the Desensitized state). In other words, the corresponding measurement matrix is:

$$
H=\left[\begin{array}{ll}
I_{4} & \mathbb{O}_{4 \times 8}
\end{array}\right] .
$$

We generate data for these species using the values of parameters provided in [33]. Additionally, we perturb the obtained data with white Gaussian noise of zero mean and standard deviation 0.0001. We explain how to estimate the parameters involved in the corresponding mathematical model from the above-mentioned time-series partial data of concentrations using our parameter estimation method.

### 5.1.2. Proving Parameter Unidentifiability

Before we perform the parameter estimation, we show that the parameters of the mathematical model under consideration cannot be uniquely determined from the available partial time series data of the species' concentrations. The reason is that we are able to provide two different vectors of parameters, both leading to the same system of ODEs corresponding to the measured concentrations of the species. Consider two vectors of parameters $k, \bar{k} \in \mathbb{R}_{+}^{34}$ with the following properties:

$$
\begin{array}{rll}
k_{i} & =\bar{k}_{i,} & \\
k_{i} & =\bar{k}_{i}=0, & \\
k_{i}=2,8,11,12,15,19,20,25,29,30, \\
k_{i} & \neq \bar{k}_{i}, & \\
\text { otherwise. }
\end{array}
$$

Any two vectors of parameters $k$ and $\bar{k}$ with these properties lead to the same ODE system corresponding to the measured concentrations of the species, and thus to the same sets of admissible output trajectories. Note that it may be cumbersome to derive these properties manually. Therefore, we have used MATLAB symbolic variables to derive these properties in an automatic way. We conclude that the corresponding mathematical model is parameter unidentifiable. From Theorem 2 it follows that the considered mathematical model is also least squares parameter unidentifiable.

### 5.1.3. Parameter Estimation Procedure

According to our procedure, we first determine the Kron reduced mathematical model obtained by deleting the complexes involving at least one species that is not measured experimentally. By making use of the complex composition matrix and the measurement matrix we identify the single-species complexes BL, AL, IL, DL, BLL, ALL, ILL, and DLL to be the ones that should be deleted from the original mathematical model. These are the precisely the species whose concentrations are not measured. We consequently obtain a new mathematical model in which only the concentrations of the measured species are involved. The reactions corresponding to the Kron reduced mathematical model are given below:
B
$\mathrm{A} \underset{p_{4}}{\stackrel{p_{1}}{\stackrel{p_{3}}{\rightleftharpoons}} \mathrm{~A}} \mathrm{I}$,
$I \underset{p_{6}}{\stackrel{p_{5}}{\rightleftharpoons}} \mathrm{D}$,
B
$\mathrm{B} \underset{p_{12}}{\stackrel{p_{9}}{p_{11}}} \mathrm{D}$,

Table 2. The values of the rate constants of the model of the nicotinic acetylcholine receptors provided in [33], the estimated values obtained using our method (for both WLS and UWLS), and the corresponding confidence intervals calculated using bootstrapping.

| Parameters | Provided Values | Estimated Values (UWLS) | Estimated Values (WLS) | Confidence Intervals |
| :---: | :---: | :---: | :---: | :---: |
| $k_{1}$ | 3000.0000 | 2985.3982 | 2999.1102 | [2975.0001, 3004.0154] |
| $k_{2}$ | 8000.0000 | 7997.6479 | 7992.4287 | [7658.0406, 8106.2202] |
| $k_{3}$ | 1500.0000 | 1499.4586 | 1486.9596 | [1400.0112, 1587.4735] |
| $k_{4}$ | 16,000.00 | 15,995.28 | 15,978.15 | [15,116.77, 16,007.37] |
| $k_{5}$ | 30,000.0000 | 29,991.0951 | 29,956.5732 | [2821.5981, 3033.6088] |
| $k_{6}$ | 700.0000 | 700.4901 | 722.1788 | [675.6376, 896.1618] |
| $k_{7}$ | 3000.0000 | 3011.9891 | 3038.551 | [2999.0560, 3066.3532] |
| $k_{8}$ | 8.6400 | 2.0077 | 36.7439 | [1.4659, 237.2756] |
| $k_{9}$ | 1500.0000 | 1499.3846 | 1496.7745 | [1424.2376, 1608.9460] |
| $k_{10}$ | 17.2800 | 18.0771 | 5.4371 | [2.3910, 26.9894] |
| $k_{11}$ | 0.5400 | 0.7522 | 0.0934 | [2.82 $\left.\times 10^{-5}, 0.7812\right]$ |
| $k_{12}$ | 10,800.00 | 10,809.95 | 10,822.44 | [10,800.01, 10,833.63] |
| $k_{13}$ | 130.0000 | 129.8977 | 0.01295 | [0.0002, 145.5059] |
| $k_{14}$ | 2740.0000 | 2739.3315 | 2736.4242 | [2546.3118, 2843.7101] |
| $k_{15}$ | 3000.0000 | 3000.5636 | 2848.1724 | [2772.8053, 3000.0009] |
| $k_{16}$ | 4.0000 | 3.2436 | 162.0834 | [1.1588, 211.1424] |
| $k_{17}$ | 1500.0000 | 1500.0776 | 1491.0929 | [1488.1059, 1513.4998] |
| $k_{18}$ | 8.0000 | 11.4275 | 378.6929 | [6.9902, 503.0242] |
| $k_{19}$ | 19.7000 | 19.6393 | 0.6956 | [0.1307, 12.2759] |
| $k_{20}$ | 3.7400 | 4.0754 | 5.5581 | [1.8582, 27.4541] |
| $k_{21}$ | 19.8500 | 18.9582 | 104.6638 | [17.6869, 190.2117] |
| $k_{22}$ | 1.7400 | 0.2558 | 66.9278 | [0.0327, 77.2369] |
| $k_{23}$ | 20.0000 | 17.8055 | 0.0246 | [0.0014, 23.0991] |
| $k_{24}$ | 0.8100 | 0.0019 | 0.01142 | [0.8588, 2.6275] |
| $k_{25}$ | 3000.0000 | 2999.178 | 2998.2315 | [2880.6353, 3012.3044] |
| $k_{26}$ | 4.0000 | 4.0298 | 4.4320 | [1.4111, 5.6222] |
| $k_{27}$ | 1500.0000 | 1499.6145 | 1500.2207 | [960.5602, 1739.9138] |
| $k_{28}$ | 8.0000 | 5.7722 | 14.2218 | [3.6051, 18.2161] |
| $k_{29}$ | 0.0500 | 0.0041 | 0.0566 | [3.93 $\left.\times 10^{-6}, 0.0567\right]$ |
| $k_{30}$ | 0.0010 | 0.3051 | 0.4618 | [0.0002, 6.9397] |
| $k_{31}$ | 0.0500 | $4.99 \times 10^{-6}$ | 0.0015 | [5.15 $\left.\times 10^{-7}, 68.9054\right]$ |
| $k_{32}$ | 0.0010 | 0.0006 | 0.0577 | [0.0001, 1.8794] |
| $k_{33}$ | 0.0500 | $6.78 \times 10^{-8}$ | 0.0028 | [1.3 $\left.\times 10^{-8}, 0.5321\right]$ |
| $k_{34}$ | 0.0010 | $1.60 \times 10^{-6}$ | 0.0009 | [ $\left.0.85 \times 10^{-6}, 0.0021\right]$ |

The Kron reduced network is schematically illustrated in the right-hand panel of Figure 1. Recall that the vector of parameters $p$ is a function of the vector of parameters $k$. We now possess a complete time-series data of species' concentrations corresponding to the Kron reduced mathematical model. In other words, we have a well-posed parameter estimation problem from species' concentrations, which is solved using the well known least squares method. In order to test the performance of our parameter estimation method, we use both UWLS and WLS to determine the corresponding values of the parameters for which the Kron reduced mathematical model fits the available data of species' concentrations. These estimates are given in Table 3.

Table 3. The best-fitting parameter values (for both WLS and UWLS approaches) of the Kron reduced mathematical model of the nicotinic acetylcholine receptors corresponding to the available time-series experimental data.

| Parameters | Estimated Values (UWLS) | Estimated Values (WLS) |
| :---: | :---: | :---: |
| $p_{1}$ | 2060.8744 | 0.2430 |
| $p_{2}$ | 947.0466 | 0.6338 |
| $p_{3}$ | 1385.3214 | $1.73 \times 10^{-5}$ |
| $p_{4}$ | $13,732.9647$ | $13,783.3992$ |
| $p_{5}$ | 1061.5483 | 53.9256 |
| $p_{6}$ | 8.6512 | 1.1069 |
| $p_{7}$ | 2860.8880 | 2762.7458 |
| $p_{8}$ | 313.8923 | 16.0378 |
| $p_{9}$ | 0.0192 | 0.2593 |
| $p_{10}$ | 1617.0009 | 340.8659 |
| $p_{11}$ | 2769.5029 | 3.8327 |
| $p_{12}$ | 1.3036 | 88.8853 |

Figure 2 represents the comparison of the available time-series data of species' concentrations of the model of NAR and the corresponding predicted values (for both WLS and UWLS approaches) obtained from the Kron reduced mathematical model with estimated parameters provided in Table 3.


Figure 2. The available time-series data of species' concentrations of the model of nicotinic acetylcholine receptors and the corresponding predicted values (using WLS and UWLS) obtained from the Kron reduced mathematical model with estimated parameters provided in Table 3.

Note that, while Kron reduction method effectively reduces the complexity of the original mathematical model, it may not fully capture all relevant dynamical properties. As can be seen from Figure 2, the Kron reduced mathematical model is not a good fit for the available time-series data. Finally, we determine the values of parameters $\widehat{k}$ (for both WLS and UWLS) that minimize the eigenvalue-based error $\delta$ defined in (17) subject to the constraint defined in (18). These best-fitting values of parameters are provided in Table 2. The comparison of the available time-series data of species' concentrations and the corresponding model predicted values obtained from (8) with estimated parameter values $\widehat{k}$ (for both WLS and UWLS approaches) is given in Figure 3.


Figure 3. The available time-series data of species' concentrations of the model of nicotinic acetylcholine receptors and the corresponding model predicted values (for both WLS and UWLS approaches) with parameters estimated by our method. These estimated parameters are given in Table 2.

We observe that our parameter estimation method is able to derive a complete mathematical model that is able to make accurate predictions about the dynamics of the CRN. Note from Table 2 that the estimated values of some of the parameters, e.g., $k_{i}, i=21, \ldots, 24$ and $i=31, \ldots, 34$, differ by a large percentage from their corresponding values provided in [33]. However, the mathematical model with the estimated values of parameters is a reasonably good fit (as can be seen in Figure 3) for the generated time-series data of species' concentrations. The reason behind this, as discussed earlier, is the fact that the parameters involved in the considered mathematical model of NAR are not least squares identifiable from the output corresponding to the compounds B, A, I, and D, as proved in Section 5.1.2.

Recall that in our parameter estimation method, we make use of the least squares optimization technique. As mentioned above, we used both WLS and UWLS to determine the best-fitting values of parameters corresponding to each of these approaches (see Table 2). We perform LOOCV in order to understand which one of these approaches is a preferable technique for data-fitting in our proposed parameter estimation procedure. We first give a short summary of LOOCV. It involves splitting the data set into two parts: a single datapoint, that is used for validation (excluded datapoint); and the remaining datapoints (training dataset), that make up the set used for parameter estimation. The parameter estimation procedure is applied using the training dataset, and a model-predicted value is computed for the excluded datapoint. We then compute the mean squared error between the excluded datapoint and its corresponding model-predicted values. Since the excluded datapoint is not used in the fitting process, this error provides an unbiased estimate for the test error. On the other hand, it is a poor estimate for the test error since it is based upon a single datapoint. We can repeat the procedure by excluding each datapoint of the given data exactly once and compute the mean squared error between the excluded datapoint and its corresponding model-predicted values. The LOOCV training error is the average of these test error estimates.

We used WLS in our parameter estimation procedure and computed the LOOCV training error. Subsequently, we used UWLS in our parameter estimation procedure and again computed the LOOCV training error. These training errors are provided in Table 4.

Table 4. Leave-one-out cross-validation training errors for the mathematical model of the nicotinic acetylcholine receptors corresponding to weighted and unweighted least squares optimization methods.

|  | Unweighted Least Squares | Weighted Least Squares |
| :---: | :---: | :---: |
| Training error | 3.2164 | 3.6108 |

As we can see from this table, using UWLS in our parameter estimation procedure results in a smaller LOOCV training error than using WLS. We thus conclude that, for this particular dataset, UWLS is a preferable approach of data-fitting in our proposed parameter estimation procedure.

We calculated the $95 \%$ confidence intervals for the estimated values of parameters corresponding to UWLS by performing bootstrapping (see, e.g., [50-53]) for our parameter estimation method. The procedure for this calculation is as follows. We first construct 2000 resampled datasets (of the same length as the available data) by randomly choosing datapoints from the available data. Note that the same datapoint can be chosen multiple times. Secondly, we use our parameter estimation method to determine the best-fitting values of parameters corresponding to each of the resampled dataset. As a result, for each of the parameters we obtain a set of 2000 estimates. The $95 \%$ bootstrap confidence intervals were constructed by choosing $2.5 \%$ and $97.5 \%$ percentiles of the corresponding bootstrap estimates. These confidence intervals are included in Table 2.

### 5.2. Trypanosoma Brucei Trypanothione Synthetase

Subsequently, we demonstrate the applicability of our new parameter estimation method on a very different type of a CRN. We consider a kinetic model of Trypanosoma brucei trypanothione synthetase (TBTS). This mathematical model of TBTS was developed in [34] and describes the entire kinetic profile. Trypanosoma brucei is a species of parasitic kinetoplastid (an organism whose cells contain a cell nucleus and that is not an animal, plant, or fungus). Unlike other parasites (normally infecting blood and tissue cells) it is exclusively extracellular and inhabits the blood plasma as well as body fluids. It causes deadly diseases in humans such as African trypanosomiasis or sleeping sickness. A trypanothione synthetase is a catalytic enzyme. A schematic representation of the model of TBTS is shown on the left-hand side of Figure 4.


Figure 4. Schematic representation of the original model (left-hand panel) of trypanosoma brucei trypanothione synthetase and the corresponding Kron-reduced model (right-hand panel) obtained by deleting the single-species complexes EA, EB, EC, EQ, ER, E_Q, EAB, EAC, EAQ, EBC, EBQ, EABQ, EABC from the graph of complexes.

### 5.2.1. The Considered Model and the Available Data

In the considered mathematical model of TBTS, there are 22 species participating in 59 unidirectional reactions in terms of 24 distinct complexes. Table 5 provides an overview of the primary compounds participating in the reaction system.

Table 5. An overview of the primary compounds participating in the considered model of the Trypanosoma brucei trypanothione synthetase.

| Species Notation | Species Name | Species ID |
| :---: | :---: | :---: |
| E | Enzyme | E |
| P | Adenosine diphosphate | ADP |
| Q | Glutathionylspermidine | GSP |
| C | Bis(glutathionyl)spermine | T(SH) 2 |
| B | Spermidine | Spd |
| A | Glutathione | GSH |
| $X$ | Adenosine triphosphate | ATP |
|  | Glutathione spermidine Enzyme | E.GS_P |

The remaining compounds occurring in the network are intermediate complexes of the enzymes E and X. The reactions occurring in the network are given as:
$\mathrm{E} \underset{k_{2}}{\stackrel{k_{1}}{\rightleftharpoons}} \mathrm{EA}$,
EA $\underset{k_{4}}{\stackrel{k_{3}}{\rightleftharpoons}}$ EAB,
$\mathrm{XQ} \underset{k_{31}}{\stackrel{k_{30}}{\rightleftharpoons}} \mathrm{XQB}$,
$\mathrm{E} \underset{k_{6}}{\stackrel{k_{5}}{\rightleftharpoons}} \mathrm{~EB}$,
EB $\underset{k_{8}}{\stackrel{k_{7}}{7}} \mathrm{EAB}$,
$\mathrm{EAB} \underset{k_{10}}{\stackrel{k_{9}}{\rightleftharpoons}} \mathrm{EABQ}$,
EA $\underset{k_{12}}{\stackrel{k_{11}}{\rightleftharpoons}}$ EAQ,
$\mathrm{EB} \underset{k_{14}}{\stackrel{k_{13}}{\rightleftharpoons}} \mathrm{EBQ}$,
$\mathrm{E} \underset{k_{16}}{\stackrel{k_{15}}{\rightleftharpoons}} \mathrm{EQ}$,
$\mathrm{EQ} \underset{k_{18}}{\stackrel{k_{17}}{\rightleftharpoons}} \mathrm{EAQ}$,
EAQ $\underset{k_{20}}{\stackrel{k_{19}}{\rightleftharpoons}}$ EABQ,
EQ $\underset{k_{22}}{\stackrel{k_{21}}{\rightleftharpoons}}$ EBQ,
EBQ $\underset{k_{24}}{\stackrel{k_{23}}{\rightleftharpoons}} \mathrm{EABQ}$,
$\mathrm{EABQ} \xrightarrow{k_{25}} \mathrm{P}+\mathrm{XQ}$,
$\mathrm{XQ} \underset{k_{27}}{\stackrel{k_{26}}{2}} \mathrm{ER}$,
$\mathrm{ER} \underset{k_{29}}{\stackrel{k_{28}}{\rightleftharpoons}} \mathrm{E}$,
$X Q \underset{k_{33}}{\stackrel{k_{32}}{\rightleftharpoons}} X Q R$,
$\mathrm{E} \underset{k_{35}}{\stackrel{k_{34}}{\rightleftharpoons}} \mathrm{EC}$,
$\mathrm{EC} \underset{k_{37}}{\stackrel{k_{36}}{\tau}}$ EAC,
EAC $\underset{k_{39}}{\stackrel{k_{38}}{\rightleftharpoons}}$ EABC,
$\mathrm{EC} \underset{k_{41}}{\stackrel{k_{40}}{\rightleftharpoons}} \mathrm{EBC}$,
$\mathrm{EBC} \underset{k_{43}}{\stackrel{k_{42}}{\rightleftharpoons}}$ EABC,
EA $\underset{k_{45}}{\stackrel{k_{44}}{\rightleftharpoons}}$ EAC,
EB $\underset{k_{47}}{\stackrel{k_{46}}{\rightleftharpoons}}$ EBC,
EAB $\underset{k_{49}}{\stackrel{k_{48}}{\rightleftharpoons}}$ EABC,
$\mathrm{EABC} \xrightarrow{k_{50}} \mathrm{P}+\mathrm{XC}$,
$X C \underset{k_{52}}{\stackrel{k_{51}}{\rightleftharpoons}}$ E_Q,
E_Q $\underset{k_{54}}{\stackrel{k_{53}}{c}} \mathrm{E}$,
$\mathrm{XC} \underset{k_{56}}{\stackrel{k_{55}}{\rightleftharpoons}}$ XCB,
$X C \underset{k_{58}}{\stackrel{k_{57}}{\rightleftharpoons}}$ XCR,

$$
\mathrm{P} \xrightarrow{k_{59}} \mathrm{~S}
$$

where $k_{i}$ is the rate constant of the $i$ th reaction. The values of these parameters provided in [34] are listed in Table 6.

Table 6. The values of the rate constants of the model of the trypanosoma brucei trypanothione synthetase provided in [34], the estimated values obtained using our method (for both WLS and UWLS), and the corresponding confidence intervals calculated using bootstrapping.

| Parameters | Provided Values | Estimated Values (UWLS) | Estimated Values (WLS) | Confidence Intervals |
| :---: | :---: | :---: | :---: | :---: |
| $k_{1}$ | 53.4178 | 44.1002 | 42.8180 | [36.3628, 53.1468] |
| $k_{2}$ | 9.06800 | 16.7531 | 12.4590 | [8.4905, 15.9902] |
| $k_{3}$ | 1.1917 | 8.7110 | 3.9547 | [1.6893, 7.7521] |
| $k_{4}$ | 3.6300 | 16.4578 | 4.8968 | [3.7351, 7.9581] |
| $k_{5}$ | 67.8700 | 58.5546 | 68.7404 | [51.6995, 69.8990] |
| $k_{6}$ | 9.5200 | 62.8531 | 0.3219 | [0.0410, 10.7708] |
| $k_{7}$ | 7.9700 | 13.8148 | 0.2034 | [0.0662, 6.8344] |
| $k_{8}$ | 4.5000 | 52.2816 | 5.1613 | [4.6057, 10.0443] |
| $k_{9}$ | 1.7600 | 1.8273 | 5.2016 | [0.0798, 7.7487] |
| $k_{10}$ | 5.5000 | 11.7959 | 11.5207 | [5.9256, 13.4458] |
| $k_{11}$ | 5.0734 | 85.8926 | 0.5021 | [0.0919, 8.6912] |
| $k_{12}$ | 5.0396 | 54.9415 | 11.2865 | [6.5235, 11.8915] |
| $k_{13}$ | 92.9616 | 86.3749 | 22.6051 | [16.9820, 91.2826] |
| $k_{14}$ | 69.6700 | 74.2373 | 0.2118 | [0.0664, 67.9548] |
| $k_{15}$ | 58.2800 | 51.5885 | 48.5051 | [45.9896, 70.1724] |
| $k_{16}$ | 81.5400 | 89.7101 | 76.2469 | [68.4823, 95.0394] |
| $k_{17}$ | 7.9000 | 85.8912 | 0.2689 | [0.0213, 7.9164] |
| $k_{18}$ | 8.8900 | 101.0203 | 14.0437 | [9.2273, 16.0937] |
| $k_{19}$ | 13.0500 | 2.9113 | 8.5301 | [0.3245, 12.0491] |
| $k_{20}$ | 6.5400 | 93.7566 | 9.0591 | [4.4476, 12.4538] |
| $k_{21}$ | 6.2615 | 52.6466 | 0.2456 | [0.0223, 7.9409] |
| $k_{22}$ | 9.5536 | 106.5830 | 15.6348 | [5.1120, 16.2136] |
| $k_{23}$ | 2.7479 | 29.7198 | 0.7268 | [0.0109, 5.7346] |
| $k_{24}$ | 7.9500 | 53.8979 | 13.1923 | [8.4006, 19.5529] |
| $k_{25}$ | 8.1300 | 67.2479 | 4.1898 | [1.9219, 12.9717] |
| $k_{26}$ | 2.7800 | 0.0056 | 0.1197 | [0.2465, 4.5403] |
| $k_{27}$ | 4.8100 | 24.7439 | 9.0857 | [5.8291, 13.3738] |
| $k_{28}$ | 0.0900 | 100.8293 | 3.4949 | [0.0506, 1.8452] |
| $k_{29}$ | 5.4700 | 18.1014 | 0.3826 | [0.1967, 1.7861] |
| $k_{30}$ | 5.5200 | 1.7397 | 0.5090 | [0.0.4621, 1.7866] |
| $k_{31}$ | 7.8669 | 4.1334 | 0.8026 | [0.7527, 3.0838] |
| $k_{32}$ | 8.6056 | 5.0141 | 0.8054 | [0.7705, 3.3345] |
| $k_{33}$ | 2.6729 | 1.1504 | 0.2569 | [0.2016, 0.9534$]$ |
| $k_{34}$ | 6.6400 | 56.1957 | 0.6883 | [0.6682, 6.4683] |

Table 6. Cont.

| Parameters | Provided Values | Estimated Values (UWLS) | Estimated Values (WLS) | Confidence Intervals |
| :---: | :---: | :---: | :---: | :---: |
| $k_{35}$ | 8.3500 | 21.4817 | 10.9232 | [8.8105, 12.3105] |
| $k_{36}$ | 2.6000 | 60.6397 | 3.9687 | [2.0922, 7.0713] |
| $k_{37}$ | 6.0900 | 68.2011 | 10.5792 | [7.4107, 13.1953] |
| $k_{38}$ | 5.9800 | 71.0124 | 5.3738 | [3.3981, 30.2118] |
| $k_{39}$ | 7.0800 | 90.1295 | 9.6122 | [7.1826, 11.6251] |
| $k_{40}$ | 4.2300 | 97.8241 | 5.0378 | [3.7236, 6.6922$]$ |
| $k_{41}$ | 6.9012 | 77.2770 | 9.0691 | [7.0122, 10.3685] |
| $k_{42}$ | 5.1458 | 55.6608 | 4.9221 | [3.0766, 9.6034] |
| $k_{43}$ | 9.8363 | 94.6713 | 12.5293 | [9.9403, 20.2133] |
| $k_{44}$ | 8.0100 | 55.9309 | 11.7194 | [8.6101, 12.8384] |
| $k_{45}$ | 1.7000 | 5.8116 | 5.0751 | [1.6858, 6.1590] |
| $k_{46}$ | 2.0900 | 10.3061 | 0.2928 | [0.1207, 19.6999] |
| $k_{47}$ | 6.2700 | 87.7272 | 7.2096 | [6.2626, 33.9641] |
| $k_{48}$ | 8.4300 | 48.0844 | 8.3244 | [8.5120, 11.2713] |
| $k_{49}$ | 4.4900 | 84.9190 | 6.1626 | [4.3403, 9.6653] |
| $k_{50}$ | 0.9400 | 1.7720 | 0.2076 | [0.0253, 0.8675$]$ |
| $k_{51}$ | 5.2317 | 51.0507 | 0.8873 | [0.8291, 1.6575] |
| $k_{52}$ | 2.8599 | 62.8507 | 5.3629 | [0.7383, 6.4126] |
| $k_{53}$ | 3.2270 | 2.2573 | 10.4618 | [3.2739, 15.3034] |
| $k_{54}$ | 6.4700 | 1.6055 | 0.1016 | [0.0154, 0.7756$]$ |
| $k_{55}$ | 3.2400 | 1.3503 | 0.6410 | [0.2601, 1.5042] |
| $k_{56}$ | 4.9500 | 2.4355 | 0.8050 | [0.7399, 2.1214] |
| $k_{57}$ | 8.9600 | 5.6437 | 0.8007 | [0.0290, 3.5065] |
| $k_{58}$ | 1.2500 | 0.4722 | 0.3622 | [0.0869, 0.4206$]$ |
| $k_{59}$ | 2.3100 | 0.9681 | 0.8058 | [0.7560, 1.1369] |

Note that, unlike the previous example, not every complex is a single-species complex. We assume that the species that are measured experimentally are $\mathrm{P}, \mathrm{S}, \mathrm{XQ}, \mathrm{XC}, \mathrm{XQB}, \mathrm{XQR}$, $X C B, X C R$ and the main enzyme $E$. We generate data corresponding to the concentrations of the above-mentioned species using the balance laws (8) and the parameter values provided in [34]. These parameters are given in Table 6. Additionally, the obtained data have been perturbed with white Gaussian noise of zero mean and standard deviation 0.02.

### 5.2.2. Proving Parameter Unidentifiability

We show that the parameters of the considered mathematical model cannot be uniquely determined from the available partial time series data of the species' concentrations. Similar to the case of the mathematical model of NAR, we provide two different vectors of parameters, both leading to the same system of ODEs corresponding to the measured concentrations of the species. Consider two vectors of parameters $k, \bar{k} \in \mathbb{R}_{+}^{59}$ with the following properties:

$$
\begin{array}{rll}
k_{i} & =\bar{k}_{i,} & i=1,2,5,6,15,16,19,25, \ldots, 35,50, \ldots, 59, \\
k_{i} & \neq \bar{k}_{i,} & \\
\text { otherwise. }
\end{array}
$$

Any two vectors of parameters $k$ and $\bar{k}$ with these properties lead to the same ODE system corresponding to the measured species' concentrations, and thus to the same sets of admissible output trajectories. These properties have been derived in an automatic way using MATLAB symbolic variables. We conclude that the corresponding mathematical model is parameter unidentifiable. From Theorem 2 it follows that the considered mathematical model is also lest squares parameter unidentifiable.

### 5.2.3. Parameter Estimation Procedure

Using the complex composition matrix and the measurement matrix our procedure selects the single-species enzymatic complexes EA, EB, EC, EQ, ER, E_Q, EAB, EAC, EAQ, $\mathrm{EBC}, \mathrm{EBQ}, \mathrm{EABC}, \mathrm{EABQ}$ to form the set of complexes that should be deleted from the graph of complexes by applying the Kron reduction method. The reactions corresponding to the resulting Kron-reduced model are given as:

$$
\begin{aligned}
& \mathrm{E} \xrightarrow{p_{1}} \mathrm{P}+\mathrm{XQ}, \\
& \mathrm{E} \xrightarrow{\stackrel{p_{2}}{\longrightarrow} \mathrm{P}+\mathrm{XC},} \\
& \mathrm{XQ} \underset{p_{4}}{\stackrel{p_{3}}{\rightleftharpoons}} \mathrm{E}, \\
& \mathrm{XQ} \underset{p_{6}}{\stackrel{p_{5}}{5}} \mathrm{XQB}, \\
& \mathrm{XQ} \underset{p_{8}}{\stackrel{p_{7}}{\rightleftharpoons}} \mathrm{XQR},
\end{aligned}
$$

$$
X C \underset{\substack{p_{10} \\ n_{10}} \stackrel{p_{9}}{\rightleftharpoons}}{ } \mathrm{E},
$$

$$
\mathrm{XC} \underset{p_{12}}{\stackrel{p_{11}}{\rightleftharpoons}} \mathrm{XCB},
$$

$$
\mathrm{XC} \underset{p_{14}}{\stackrel{p_{13}}{\rightleftharpoons}} \mathrm{XCR}
$$

$$
\mathrm{P} \xrightarrow{p_{15}} \mathrm{~S} .
$$

Here, as usual, for every $i=1, \ldots, 15, p_{i}$ denotes the rate constant of the $i$ th reaction corresponding to the Kron-reduced model. The schematic representation of the Kron reduced network is illustrated in the right-hand panel of Figure 4.

Since we have complete data of the concentrations of the species involved in the Kron reduced model, we may apply the least squares method to find the best-fitting values of the parameters $\hat{p} \in \mathbb{R}_{+}^{15}$. Similar to the case of the mathematical model of NAR, we use both WLS and UWLS to determine the values of the parameters for which the Kron reduced mathematical model fits the available data of species' concentrations. These estimates are given in Table 7. Figure 5 visualizes the comparison of the available time-series data and the corresponding model predicted values (for both WLS and UWLS approaches) obtained from the Kron reduced model with estimated values of parameters $\widehat{p}$. Note that, in this case, both WLS and UWLS result in similar model-predicted values for the output concentrations. Also note that the Kron reduced mathematical model is not a good fit for the available time-series data.

In the final step, we determine the values of parameters $\widehat{k} \in \mathbb{R}_{+}^{59}$ (for both WLS and UWLS approaches) of the original model that minimize the eigenvalue-based error $\delta$ defined in (17) subject to the constraint defined in (18). These values of parameters are given in Table 6. The comparison of the available time-series data and the corresponding model predicted values obtained from the balance laws (8) with the estimated values of parameters $\widehat{k}$ (for both WLS and UWLS approaches) be seen in Figure 6. Note that our parameter estimation method resulted in a complete mathematical model that is able to make accurate predictions about the dynamical behavior of the CRN.

Observe from Table 6 that for some of the parameters, e.g., $k_{i}, i=7, \ldots, 14$ and $i=20, \ldots, 24$, there is a substantial difference between the values provided in [54] and their corresponding estimated values. The reason behind this is the fact that the parameters contained in the mathematical model are least squares unidentifiable from the available time-series partial data of species' concentrations, as proved in Section 5.2.2.

Table 7. The best-fitting parameter values (for both WLS and UWLS approaches) of the Kron reduced mathematical model of the Trypanosoma brucei trypanothione synthetase corresponding to the available time-series experimental data.

| Parameters | Estimated Values (UWLS) | Estimated Values (WLS) |
| :---: | :---: | :---: |
| $p_{1}$ | 8.8681 | 0.3572 |
| $p_{2}$ | 2.3387 | 0.5386 |
| $p_{3}$ | 0.0409 | 0.0893 |
| $p_{4}$ | 5.4580 | 0.9955 |
| $p_{5}$ | 4.6200 | 3.0531 |
| $p_{6}$ | 7.8525 | 4.9069 |
| $p_{7}$ | 8.5119 | 1.9480 |
| $p_{8}$ | 2.9776 | 0.6329 |
| $p_{9}$ | 2.5213 | 7.1061 |
| $p_{10}$ | 3.1172 | 2.7720 |
| $p_{11}$ | 3.2643 | 6.1818 |
| $p_{12}$ | 4.9430 | 3.6137 |
| $p_{13}$ | 9.0230 | 3.2929 |
| $p_{14}$ | 0.8362 | 0.5099 |
| $p_{15}$ | 2.2947 | 1.5590 |

Table 8. Leave-one-out cross-validation training errors for mathematical model of the Trypanosoma brucei trypanothione synthetase corresponding to unweighted and weighted least squares optimization methods.

|  | Unweighted Least Squares | Weighted Least Squares |
| :---: | :---: | :---: |
| Training error | 0.8233 | 0.6977 |



Figure 5. The available time-series data of species' concentrations of the model of trypanosoma brucei trypanothione synthetase and the corresponding predicted values obtained from the Kron-reduced mathematical model with estimated parameters provided in Table 7.


Figure 6. The available time-series data of species' concentrations of the model of trypanosoma brucei trypanothione synthetase and the corresponding model predicted values with parameters estimated by our method. These estimated parameters are given in Table 6.

Similar to the case of the mathematical model of NAR, we perform LOOCV to understand which one of WLS and UWLS is a more preferable approach for data-fitting in our proposed parameter estimation procedure. The corresponding LOOCV training errors are given in Table 8. As we can see from this table, unlike the case of NAR, using WLS in our parameter estimation procedure results in a smaller LOOCV training error than using UWLS. We thus conclude that, for this particular dataset, WLS is a more preferable approach of data-fitting in our proposed parameter estimation procedure. We calculate the $95 \%$ confidence intervals for the estimated values of parameters corresponding to WLS by performing bootstrapping for our parameter estimation method. These confidence intervals are included in Table 6.

## 6. Discussion

The Kron-reduced mathematical model with the best-fitting values of parameters (in the sense of least squares), as we can see from Figures 2 and 5, is generally not an appropriate approximation, meaning that the corresponding model predicted values are far from being good fits for the available time-series data. This is because of the fact that, in general depending on the number of complexes deleted from the graph of complexes, it is not assured that the Kron-reduced model is a reasonable approximation for the original mathematical model.

The choice of the Kron reduction technique as a tool for reducing mathematical models in our parameter estimation method is based on several advantages of this reduction technique that are particularly appropriate for the problem. First of all, it does not impose any restrictions on the choice of complexes to be deleted. Thus, we can delete all the complexes containing at least a single unmeasured species. A second advantage is that Kron reduction method preserves the kinetics of the CRN, i.e., if MAKRL governs the given CRN, then the corresponding Kron-reduced model is also governed by this rate law. A third advantage of the Kron reduction method is that we are able to compare the dynamics of the original model to the one of the reduced model using the Laplacian matrix of the original model and the Laplacian matrix of the reduced model. To the best of our knowledge, there is no other reduction technique that offers all these aforementioned advantages.

The suggested parameter estimation method is only applicable to a mass action CRN with a constant Laplacian. This is because of the fact that in the estimation procedure, the eigenvalues of the Laplacian matrix are used. For a general enzymatic CRN, the corresponding Laplacian matrix is not constant since it depends on the vector of species' concentrations. In such cases, it is not straightforward how to use a similar technique for parameter estimation purposes. The parameter estimation of enzyme kinetic reaction networks from partial data of species' concentrations is still an open problem that will be considered in future work.

As explained in $[37,41]$, a linkage class of a CRN is a connected component of its graph of complexes. It is also stated in these papers that if a network has a linkage class with only one reaction, then the removal of a complex involved in such a reaction by Kron reduction leads to the removal of the reaction. In the case, where the intermediate Kron reduction phase of our parameter estimation procedure leads to the removal of some of the reactions of the original model, we would expect that the estimated parameters of the original model associated with the removed reactions would have larger confidence intervals compared with those of the other parameters that are associated with the remaining reactions of the network.

## 7. Conclusions

In this paper, we have introduced an innovative parameter estimation approach for mathematical models of mass action CRNs using observed time-series incomplete experimental data of species' concentrations. As far as we know, there exists no direct technique for deducing the parameters in a mathematical model from this sort of experimental data. We have addressed this problem by devising an algorithmic strategy, which involves the application of Kron reduction technique for kinetic models as an intermediate step in the overall parameter estimation approach. The complexes that should be deleted using Kron reduction are chosen in such a way that in the reduced model only the concentrations of the measured species are involved. Since all the species' concentrations involved in the Kron-reduced model are measured we now have a well-posed parameter estimation problem. We estimate the parameters involved in the Kron-reduced model using the least squares method to identify the best-fitting values of the parameters involved in the Kron-reduced model. To estimate the parameters contained in the original mathematical model, we have devised a new trajectory-independent measure to quantify the difference between the dynamics of the original model and the corresponding Kron-reduced model. It is based on comparing the smallest non-zero real part of the eigenvalues of the original Laplacian matrix with the one of the Kron-reduced Laplacian matrix. The reason behind the choice of measure is the fact that the smallest non-zero real part of the Laplacian matrix is related to the settling time of the CRN that is characterized by the Laplacian matrix. This measure can be regarded as a function of the parameter vector of the original mathematical model since the eigenvalues of both the original Laplacian matrix as well as the Kron-reduced Laplacian matrix depend only on this vector of parameters. Finally, we find the estimates of the parameters for which the above-mentioned spectral-based measure admits its smallest value.

We have devised an automatic process for our parameter estimation approach, crafting a MATLAB library that can be employed to deduce the parameters from experimental data of species' concentrations automatically. This MATLAB library is given as Supplementary Material. We utilized this MATLAB library to effectively employ our parameter estimation approach on two real-life CRNs taken from the Biomodels database [32]. For each of these models, we observed that our parameter estimation method resulted in a complete mathematical model that could make accurate predictions about the dynamics of the CRN. While we have exclusively applied and tested our parameter estimation method on a limited scale, involving only two real-life instances of CRNs, its applicability extends to all networks regulated by MAKRL. It should be noted that the method places no restriction on the size or scale of the model as well as the biological purpose of the associated network as
long as its reactions are governed by MAKRL. Thus the method is applicable for models of core metabolism (for instance E.coli central carbon metabolism models reviewed in [55]) as well as models of regulatory networks. It can also be applied for small models like the one of NAR considered in this paper as well as genome-scaled models as in [56].

Supplementary Materials: The following supporting information can be downloaded at: https:/ / www.mdpi.com/article/10.3390/bioengineering10091056/s1, The MATLAB library corresponding to the parameter estimation method described in the main manuscript.
Author Contributions: S.R. proposed the problem addressed in the manuscript, while M.G. formulated and further developed the problem; M.G. initially developed the solution methodology, which was further refined and enhanced by S.R.; M.G. conducted the computational experiments associated with the research; M.G. took the lead in writing the manuscript, and S.R. provided critical feedback, reviewed the content, and contributed to its improvement. Both authors actively participated in discussions and revisions to ensure the accuracy and clarity of the final manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.
Institutional Review Board Statement: Not applicable.
Informed Consent Statement: Not applicable.
Data Availability Statement: Not applicable.
Acknowledgments: This work has been supported by the Research and Development Program of Ghent University Global Campus (GUGC), South Korea. We would like to express our sincere gratitude to Samvel B. Gasparyan (AstraZeneca, Gothenburg, Sweden) and Joris Vankerschaver (University of Ghent, Ghent, Belgium) for their invaluable assistance with both the numerical experiments and statistical analysis. Their contributions were instrumental in the successful completion of this research.

Conflicts of Interest: The authors declare no conflict of interest.

| Abbreviations |  |
| :--- | :--- |
| The following abbreviations are used in this manuscript: |  |
| CRN | Chemical Reaction Network |
| CRNT | Chemical Reaction Network Theory |
| IVP | Initial Value Problem |
| LOOCV | Leave-One-Out Cross-Validation |
| MAKRL | Mass Action Kinetics Rate Law |
| NAR | Nicotinic Acetylcholine Receptors |
| ODE | Ordinary Differential Equation |
| TBTS | Trypanosoma Brucei Trypanothione Synthetase |
| UWLS | Unweighted Least Squares |
| WLS | Weighted Least Squares |

## References

1. Edwards, L.M.; Thiele, I. Applying systems biology methods to the study of human physiology in extreme environments. Extrem. Physiol. Med. 2013, 2, 8. [CrossRef]
2. Mock, A.; Chiblak, S.; Herold-Mende, C. Lessons we learned from high-throughput and top-down systems biology analyses about glioma stem cells. Curr. Pharm. Des. 2014, 20, 66-72. [CrossRef] [PubMed]
3. Himmelblau, D.M.; Riggs, J.B. Basic Principles and Calculations in Chemical Engineering, 8th ed.; Prentice Hall: Upper Saddle River, NJ, USA, 2012.
4. Ljung, L. System Identification: Theory for the User, 2nd ed.; Prentice Hall: Upper Saddle River, NJ, USA, 1999.
5. Fröhlich, F.; Kaltenbacher, B.; Theis, F.J.; Hasenauer, J. Scalable parameter estimation for genome-scale biochemical reaction networks. PLoS Comput. Biol. 2017, 13, e1005331. [CrossRef] [PubMed]
6. Crampin, E.J.; Schnell, S.; McSharry, P.E. Mathematical and computational techniques to deduce complex biochemical reaction mechanisms. Prog. Biophys. Mol. Biol. 2004, 86, 77-112. [CrossRef] [PubMed]
7. Donnelly, J.K.; Quon, D. Identification of parameters in systems of ordinary differential equations using quasilinearization and data perturbation. Can. J. Chem. Eng. 1970, 48, 114-119. [CrossRef]
8. Loskot, P.; Atitey, K.; Mihaylova, L. Comprehensive review of models and methods for inferences in bio-chemical reaction networks. Front. Genet. 2019, 10, 549. [CrossRef] [PubMed]
9. Ross, J.; Schreiber, I.; Vlad, M.O. Determination of Complex Reaction Mechanisms: Analysis of Chemical, Biological, and Genetic Networks; Oxford University Press: Oxford, UK, 2005.
10. Antoniewicz, M.R. Methods and advances in metabolic flux analysis: A mini-review. J. Ind. Microbiol. Biotechnol. 2015, 42, 317-325. [CrossRef] [PubMed]
11. Hodges, A.; Chatelier, R. Electrochemical Method for Measuring Chemical Reaction Rates. US Patent 6,444,115, 3 September 2002.
12. Megchelenbrink, W.; Rossell, S.; Huynen, M.A.; Notebaart, R.A.; Marchiori, E. Estimating metabolic fluxes using a maximum network flexibility paradigm. PLoS ONE 2015, 10, e0139665. [CrossRef]
13. Orth, J.D.; Thiele, I.; Palsson, B.Ø. What is flux balance analysis? Nat. Biotechnol. 2010, 28, 245-248. [CrossRef]
14. Gasparyan, M.; Van Messem, A.; Rao, S. Parameter estimation for models of chemical reaction networks from experimental data of reaction rates. Int. J. Control 2021, 96, 392-407. [CrossRef]
15. Bézier, P.E. The Mathematical Basis of the UNISURF CAD System; Butterworth-Heinemann: Newton, MA, USA, 1986.
16. Bézier, P.E.; Sioussiou, S. Semi-automatic system for defining free-form curves and surfaces. Comput.-Aided Des. 1983, 15, 65-72. [CrossRef]
17. Gelman, A.; Bois, F.; Jiang, J. Physiological pharmacokinetic analysis using population modeling and informative prior distributions. J. Am. Stat. Assoc. 1996, 91, 1400-1412. [CrossRef]
18. Girolami, M. Bayesian inference for differential equations. Theor. Comput. Sci. 2008, 408, 4-16. [CrossRef]
19. Barber, D.; Wang, Y. Gaussian Processes for Bayesian Estimation in Ordinary Differential Equations. In Proceedings of the 31st International Conference on Machine Learning, Beijing, China, $22-24$ June 2014; Volume 32, pp. 1485-1493.
20. Chkrebtii, O.A.; Campbell, D.A.; Calderhead, B.; Girolami, M.A. Bayesian solution uncertainty quantification for differential equations. Bayesian Anal. 2016, 11, 1239-1267. [CrossRef]
21. Liu, B.; Wang, L.; Cao, J. Bayesian estimation of ordinary differential equation models when the likelihood has multiple local modes. Monte Carlo Methods Appl. 2018, 24, 117-127. [CrossRef]
22. Huang, H.; Handel, A.; Song, X. A Bayesian approach to estimate parameters of ordinary differential equation. Comput. Stat. 2020, 35, 1481-1499. [CrossRef]
23. Wasserman, L. All of Statistics: A Concise Course in Statistical Inference; Springer: New York, NY, USA, 2004.
24. Murzin, D.Y.; Wärnå, J.; Haario, H.; Salmi, T. Parameter estimation in kinetic models of complex heterogeneous catalytic reactions using Bayesian statistics. React. Kinet. Mech. Catal. 2021, 133, 1-15. [CrossRef]
25. Rathinam, M.; Yu, M. State and parameter estimation from exact partial state observation in stochastic reaction networks. J. Chem. Phys. 2021, 154, 034103. [CrossRef]
26. Savara, A.; Walker, E.A. CheKiPEUQ Intro 1: Bayesian parameter estimation considering uncertainty or error from both experiments and theory. ChemCatChem 2020, 12, 5385-5400. [CrossRef]
27. Galagali, N.; Marzouk, Y.M. Bayesian inference of chemical kinetic models from proposed reactions. Chem. Eng. Sci. 2015, 123, 170-190. [CrossRef]
28. Cohen, M.; Vlachos, D.G. Chemical kinetics Bayesian inference toolbox (CKBIT). Comput. Phys. Commun. 2021, 265, 107989. [CrossRef]
29. Rao, S.; van der Schaft, A.; van Eunen, K.; Bakker, B.M.; Jayawardhana, B. A model reduction method for biochemical reaction networks. BMC Syst. Biol. 2014, 8, 52. [CrossRef] [PubMed]
30. Radulescu, O.; Gorban, A.N.; Zinovyev, A.; Noel, V. Reduction of dynamical biochemical reactions networks in computational biology. Front. Genet. 2012, 3, 131. [CrossRef] [PubMed]
31. Snowden, T.J.; van der Graaf, P.H.; Tindall, M.J. Methods of model reduction for large-scale biological systems: A survey of current methods and trends. Bull. Math. Biol. 2017, 79, 1449-1486. [CrossRef]
32. Malik-Sheriff, R.S.; Glont, M.; Nguyen, T.V.N.; Tiwari, K.; Roberts, M.G.; Xavier, A.; Vu, M.T.; Men, J.; Maire, M.; Kananathan, S.; et al. BioModels-15 years of sharing computational models in life science. Nucleic Acids Res. 2020, 48, D407-D415. [CrossRef] [PubMed]
33. Edelstein, S.J.; Schaad, O.; Henry, E.; Bertrand, D.; Changeux, J.P. A kinetic mechanism for nicotinic acetylcholine receptors based on multiple allosteric transitions. Biol. Cybern. 1996, 75, 361-379. [CrossRef] [PubMed]
34. Leroux, A.E.; Haanstra, J.R.; Bakker, B.M.; Krauth-Siegel, R.L. Dissecting the catalytic mechanism of Trypanosoma brucei trypanothione synthetase by kinetic analysis and computational modeling. J. Biol. Chem. 2013, 288, 23751-23764. [CrossRef]
35. Bapat, R.B. Graphs and Matrices; Springer: Berlin/Heidelberg, Germany, 2010; Volume 27.
36. Bollobás, B. Modern Graph Theory: Graduate Texts in Mathematics; Springer: Berlin/Heidelberg, Germany, 1998; Volume 184.
37. Gasparyan, M.; Van Messem, A.; Rao, S. An automated model reduction method for biochemical reaction networks. Symmetry 2020, 12, 1321. [CrossRef]
38. Agaev, R.; Chebotarev, P. On the spectra of nonsymmetric Laplacian matrices. Linear Algebra Its Appl. 2005, 399, 157-168. [CrossRef]
39. Gershgorin, S.A. Über die abgrenzung der eigenwerte einer matrix. Bull. l'Acad. Sci. l'URSS 1931, 6, 749-754.
40. Mesbahi, M.; Egerstedt, M. Graph Theoretic Methods in Multiagent Networks; Princeton University Press: Princeton, NJ, USA, 2010.
41. Gasparyan, M.; Van Messem, A.; Rao, S. A novel technique for model reduction of biochemical reaction networks. IFACPapersOnLine 2018, 51, 28-31. [CrossRef]
42. Zhang, F. The Schur Complement and Its Applications; Springer Science \& Business Media: Berlin/Heidelberg, Germany, 2006; Volume 4.
43. Levenberg, K. A method for the solution of certain non-linear problems in least squares. Q. Appl. Math. 1944, 2, 164-168. [CrossRef]
44. Marquardt, D.W. An algorithm for least-squares estimation of nonlinear parameters. J. Soc. Ind. Appl. Math. 1963, 11, $431-441$. [CrossRef]
45. Fletcher, R. A Modified Marquardt Subroutine for Non-Linear Least Squares; Theoretical Physics Division, Atomic Energy Research Establishment: Harwell, UK, 1971.
46. Bellman, R.; Åström, K.J. On structural identifiability. Math. Biosci. 1970, 7, 329-339. [CrossRef]
47. Caravelli, F.; Traversa, F.L.; Di Ventra, M. Complex dynamics of memristive circuits: Analytical results and universal slow relaxation. Phys. Rev. E 2017, 95, 022140. [CrossRef] [PubMed]
48. Atkins, P.; de Paula, J. Atkins' Physical Chemistry; Oxford University Press: Oxford, UK, 2014.
49. Yau, M.K.; Rogers, R.R. A Short Course in Cloud Physics; Elsevier: Amsterdam, The Netherlands, 1996.
50. Efron, B.; LePage, R. Introduction to Bootstrap; Wiley \& Sons, New York, NY, USA, 1992.
51. Efron, B.; Tibshirani, R.J. An Introduction to the Bootstrap; CRC Press: Boca Raton, FL, USA, 1994.
52. Johnson, R.W. An introduction to the bootstrap. Teach. Stat. 2001, 23, 49-54. [CrossRef]
53. Varian, H. Bootstrap tutorial. Math. J. 2005, 9, 768-775.
54. Clancy, C.E.; Rudy, Y. Cellular consequences of HERG mutations in the long QT syndrome: precursors to sudden cardiac death. Cardiovasc. Res. 2001, 50, 301-313. [CrossRef]
55. Shepelin, D.; Machado, D.; Nielsen, L.K.; Herrgård, M.J. Benchmarking kinetic models of Escherichia coli metabolism. bioRxiv 2020. [CrossRef]
56. Smallbone, K.; Simeonidis, E.; Swainston, N.; Mendes, P. Towards a genome-scale kinetic model of cellular metabolism. BMC Syst. Biol. 2010, 4, 1-9. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

