Article

Parameter Estimation for a Kinetic Model of a Cellular System Using Model Order Reduction Method

Neveen Ali Eshtewy [1,*], Lena Scholz [2,*] and Andreas Kremling [3]

1   Department of Mathematics, Arish University, Arish 31111, Egypt
2   Department of Mathematics, Technische Universität Berlin, 10623 Berlin, Germany
3   School of Engineering and Design, Technische Universität München, 80333 Munich, Germany
*   Correspondence: eshtewy@mpi-magdeburg.mpg.de (N.A.E.); scholz@math.tu-berlin.de (L.S.)

Abstract: Order reduction methods are important tools for systems engineering and can be used, for example, for parameter estimation of kinetic models for systems biology applications. In particular, the Proper Orthogonal Decomposition (POD) method produces a reduced-order model of a system that is used for solving inverse problems (parameter estimation). POD is an intrusive model order reduction method that is aimed to obtain a lower-dimensional system for a high-dimensional system while preserving the main features of the original system. We use a singular value decomposition (SVD) to compute a reduced basis as it is usually numerically more robust to compute the singular values of the snapshot matrix instead of the eigenvalues of the corresponding correlation matrix. The reduced basis functions are then used to construct a data-fitting function that fits a known experimental data set of system substance concentrations. The method is applied to calibrate a kinetic model of carbon catabolite repression (CCR) in *Escherichia coli*, where the regulatory mechanisms on the molecular side are well understood and experimental data for a number of state variables is available. In particular, we show that the method can be used to estimate the uptake rate constants and other kinetic parameters of the CCR model.

Keywords: model order reduction; proper orthogonal decomposition; singular value decomposition; inverse problem; parameter estimation; kinetic model; Latin hypercube sampling

MSC: 9208; 65L09; 92C42

1. Introduction

In systems biology, mathematical models of cellular systems are essential for predicting and optimizing cells behavior in culture [1–3]. One category of these mathematical models is a continuous model, which describes the change of substance concentrations over time via a system of differential equations. Mathematical models are capable of understanding the full picture of biological systems, but finding the right parameter values (e.g., reaction rate constants) is a serious challenge in systems biology.

Parameter estimation aiming to determine the parameter values for a mathematical model so that the dynamical system evolves in a way such that the system characteristics obtained from experimental observations are properly matched, see [4–6]. Mathematically, the parameter estimation problem corresponds to an inverse problem [7], i.e., one seeks to find the unknown model inputs (e.g., parameter values) using the measured system states and other available information [8]. Different algorithms of computational optimization can be used for estimating unknown parameters of a dynamical system [9,10], e.g., Gauss–Newton Method [11], least square regression [12,13], or maximum likelihood estimation [14]. In addition, parameter estimation was recently discussed using machine learning algorithms, for instance, convolutional neural networks [15].

Proper orthogonal decomposition (POD) is a model order reduction (MOR) technique is aiming to project high-dimensional system onto a lower-dimensional subspace whilst

retaining the most important features of the original system [16,17]. The POD method, also known as Karhunen–Loève Expansion (KLE) [18,19], is considered an optimal linear method since it minimizes the squared distance between the original and reduced model. Recently, the POD has received much attention for analyzing complex physical systems [20–22]. Moreover, it has been applied to different dynamical systems in biology, e.g., the large-scale kinetic model of the metabolic network of *Escherichia coli* [3,23] and oscillating biological network models [24].

The POD method generates a reduced-order model that is a linear combination of basis elements. This basis elements are computed to capture the essential features of the system [25]. Thus, the POD method is well suited in optimal control and for solving inverse problem of dynamical systems, like the parameter estimation problem. Moreover, the POD method takes an optimally large set of parameter values and then reduces the set to a few estimates to be tested, so it is considered less computationally intense [24]. Often, the POD method is used for parameter estimation of partial differential equations (PDEs), e.g., in [26] it is used to compute a reduced-order model for the bidomain equations of cardiac electrophysiology which is utilized in an inverse problem solved using an evolutionary algorithm. Also in [27], the POD method has been applied to estimate scalar parameters in an elliptic PDE. In [24], the POD method is used for parameter estimation of ordinary differential equation (ODE) systems for stable oscillating biological networks. The study shows that the POD method is more accurate than the spline method for a stable oscillating network.

The aim of this study is to discuss parameter estimation using POD method for an ODE system representing the kinetic model of a cellular system. Specifically, the method is applied to the carbon catabolite repression (CCR) mechanism in Escherichia coli [28]. The POD method provides a reduced-order model that is a linear combination of reduced basis functions. We use a singular value decomposition (SVD) of the snapshot matrix to compute the reduced basis, while in [24,25,27], the reduced basis is computed via the eigenvalues and eigenvectors of the correlation matrix. The snapshot matrix is generated from numerical simulations of the kinetic model for different parameter values and with certain initial conditions. The reduced basis functions are then used for constructing a data-fitting function that fits measurement data sets of dynamical system substance concentrations. Non-linear least square fitting is considered the most common approach to estimate parameters for non-linear systems [9,10]. Therefore, we are aiming to perform parameter estimation for the CCR kinetic model using POD and the Least square method (LSQ) for comparison. As we will show in this study, the results of the CCR kinetic model with the obtained estimated parameter values using the POD method fit well with the experimental data.

The remainder of this paper is organized in the following manner. In Section 2, we discuss the parameter estimation procedure for non-linear ordinary differential equations using the POD method, singular value decomposition, and constructing a data-fitting function for the observable components. In Section 3, the approaches are applied to the carbon catabolite repression network of *E. coli*. Finally, we end with some concluding remarks in Section 4.

2. Parameter Estimation for ODEs Using the POD Method

We consider a system of ordinary differential equations given by

$$\dot{x}(t) = f(x(t), \Theta), \quad x(t_0) = x_0 \in \mathbb{R}^{n_m} \quad \text{on} \quad \mathbb{I} = [t_0, t_{end}], \tag{1}$$

where $x(t) \in \mathbb{R}^{n_m}$ is the state vector (in our case the vector of concentration of n_m species) and $\Theta \in \mathbb{R}^{n_\theta}$ is a vector of parameters that contains all unknown constants determining the dynamics, e.g., kinetic parameters. While the system may depend on a number of parameters $\Theta = [\theta_1, \ldots, \theta_{n_\theta}]$, at first, we estimate one parameter, $n_\Theta = 1$.

Assume that we have some measurement data from experiments that are given in the following matrix

$$Y = \begin{bmatrix} y_1(\tau_1) & \cdots & y_1(\tau_{n_\tau}) \\ & \vdots & \\ y_{n_d}(\tau_1) & \cdots & y_{n_d}(\tau_{n_\tau}) \end{bmatrix} \in \mathbb{R}^{n_d \times n_\tau},$$

containing measurements for n_d components at n_τ measurement time points τ_j in the interval $\mathbb{I}$. Our final goal is to fit the parameters of the mathematical model to the experimental data, we assume a suitable ordering of x for the sake of implementation simplicity i.e., $x_i(\tau_j) \approx y_i(\tau_j)$ for $i = 1, \ldots, n_d < n_m$ and $j = 1, \ldots, n_\tau$ with sufficient high accuracy.

2.1. Creating the Snapshot Matrix

We perform n_s simulations of the model using different sets of parameter values $\Theta_1, \ldots \Theta_{n_s}$ equidistantly distributed in $[lb, ub]$, where lb and ub are the minimum and maximum for all parameter values. A Latin hypercube sampling (LHS) is used for generating the parameter space [29,30]. In each of the n_s simulations, we evaluate at n_t simulation time points. Note that n_t can be much larger than n_τ. In this way, we obtain the snapshot matrix

$$X = \begin{bmatrix} \tilde{x}(t_1; \Theta_1) & \cdots & \tilde{x}(t_{n_t}; \Theta_1) \\ \tilde{x}(t_1; \Theta_2) & \cdots & \tilde{x}(t_{n_t}; \Theta_2) \\ & \vdots & \\ \tilde{x}(t_1; \Theta_{n_s}) & \cdots & \tilde{x}(t_{n_t}; \Theta_{n_s}) \end{bmatrix} \in \mathbb{R}^{n_m \cdot n_s \times n_t},$$

where $\tilde{x}(t_i; \Theta_j)$ denotes a numerical approximation to the solution $x(t_i; \Theta_j)$ of (1) at the time point t_i for parameter value Θ_j.

We can regard $X = \begin{bmatrix} \bar{x}(t_1) & \cdots & \bar{x}(t_{n_t}) \end{bmatrix}$ as the snapshot matrix taken for the enlarged system

$$\dot{\bar{x}} = \bar{f}(\bar{x}), \quad \bar{x}(t_0) = \bar{x}_0 \in \mathbb{R}^{n_m \cdot n_s}, \tag{2}$$

where $\bar{x}(t) := \begin{bmatrix} x(t; \Theta_1) \\ x(t; \Theta_2) \\ \vdots \\ x(t; \Theta_{n_s}) \end{bmatrix} \in \mathbb{R}^{n_m \cdot n_s}$, with each $x(t; \Theta_i) \in \mathbb{R}^{n_m}$ is a solution of (1), and $\bar{f}(\bar{x}) = \begin{bmatrix} f(x, \Theta_1) \\ f(x, \Theta_2) \\ \vdots \\ f(x, \Theta_{n_s}) \end{bmatrix}$, as well as $\bar{x}_0 = \begin{bmatrix} x_0 \\ x_0 \\ \vdots \\ x_0 \end{bmatrix}$.

2.2. Computing the Reduced Basis

To obtain a reduced basis, one can use the eigenvalues of the correlation matrix $C = X^T X$ as in [24,25]. Since the eigenvalues λ_j of C, and the singular values σ_j of X are related as $\lambda_j = \sigma_j^2$, we use a singular value decomposition (SVD) to compute the singular values of X instead.

If the matrix X has the condition number $\kappa(X)$, then the condition number of C is $\kappa(C) = \kappa(X)^2$, such that for problems with high condition number the computation of eigenvalues is much less robust that the computation of singular values, see e.g., [31].

Let $d = rank(X) \leq min(n_m \cdot n_s, n_t)$. The SVD guarantees the existence of real numbers $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_d > 0$ and orthogonal matrices $V \in \mathbb{R}^{n_m \cdot n_s \times n_m \cdot n_s}$ with columns $\{v^i\}_{i=1}^{n_m \cdot n_s}$ and $W \in \mathbb{R}^{n_t \times n_t}$ with columns $\{w^i\}_{i=1}^{n_t}$, such that

$$V^T X W = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} =: \Sigma \in \mathbb{R}^{n_m \cdot n_s \times n_t} \tag{3}$$

where $D = diag(\sigma_1, \ldots, \sigma_d) \in \mathbb{R}^{d \times d}$ and the zeros denote matrices of appropriate dimensions. From (3), we obtain

$$X = V\Sigma W^T,$$

or similarly, the vectors $\{v^i\}_{i=1}^d$ and $\{w^i\}_{i=1}^d$ satisfy

$$Xw^i = \sigma_i v^i \quad \Leftrightarrow \quad v^i = \frac{1}{\sigma_i} Xw^i, \quad \text{for } i = 1, \ldots, d.$$

To obtain a reduced basis we can now cut off singular values with $\sigma_i < \varepsilon$, where ε is chosen such that a smaller number of basis elements $n_r << n_m \cdot n_s$ is sufficient to capture the main features of the solution of (2). By setting $V_r = [v^1, \ldots, v^{n_r}] \in \mathbb{R}^{n_m \cdot n_s \times n_r}$ (the so-called POD modesor POD basis vectors), then a reduced order model (ROM) for (2) is given by

$$\dot{x}_r = \bar{f}_r(t, \bar{x}_r), \quad \bar{x}_r(t_0) = V_r^T \bar{x}_0 \in \mathbb{R}^{n_r}, \tag{4}$$

where $\bar{x} \approx V_r \bar{x}_r$, i.e., $V_r \bar{x}_r$ serves as approximation for $\bar{x}$, and $\bar{f}_r(t, \bar{x}_r) := V_r^T \bar{f}(t, V_r \bar{x}_r)$.

It is well-known that $span\{v^1, \ldots, v^{n_r}\}$ is the best approximation of $Range(X)$ in the sense that minimizes the two-norm of the approximation error:

$$\sum_{j=1}^{n_t} \|\bar{x}(t_j) - V_r V_r^T \bar{x}(t_j)\|_2^2 = \sum_{i=n_r+1}^d \sigma_i^2,$$

see [32].

Usually, the goal is to choose n_r small enough while the relative information content [33] of the basis for the n_r-dimensional subspace, defined by

$$I(n_r) = \frac{\sum_{i=1}^{n_r} \sigma_i}{\sum_{i=1}^d \sigma_i},$$

is near to one. If the n_r-dimensional subspace should contain a percentage p of the information contained in the full dimensional space $\mathbb{R}^{n_m \cdot n_s}$, then one should choose n_r such that

$$n_r = \operatorname{argmin}\left\{ I(n_r) \mid I(n_r) \geq \frac{p}{100} \right\}.$$

Each data vector $\bar{x}^i \in \mathbb{R}^{n_m \cdot n_s}$, i.e., the i-th column of X, can be written as

$$\bar{x}^i = \sum_{j=1}^d b_{ij} v^j, \quad \text{with } b_{ij} = \langle v^j, \bar{x}^i \rangle = (v^j)^T \cdot \bar{x}^i, \tag{5}$$

where $\langle ., . \rangle$ denotes the canonical inner product in $\mathbb{R}^{n_m \cdot n_s}$.

Thus, the state vector can be approximated in the reduced basis as

$$\bar{x}^i \approx \sum_{j=1}^{n_r} b_{ij} v^j, \quad i = 1, \ldots, n_t,$$

and each n_m-dimensional component $\bar{x}_\ell^i$, $\ell = 1, \ldots, n_s$ of $\bar{x}^i$ (relating to (1)) can be represented in the reduced basis by

$$x(t_i; \Theta_\ell) = \bar{x}_\ell^i \approx \sum_{j=1}^{n_r} b_{ij} \hat{v}_\ell^j,$$

where v^j is decomposed into n_s components as in (2)

$$v^j = \begin{bmatrix} v^j_1 \\ \vdots \\ v^j_{n_m} \\ \hline v^j_{n_m+1} \\ \vdots \\ v^j_{2n_m} \\ \hline \vdots \\ \hline v^j_{(n_s-1)n_m+1} \\ \vdots \\ v^j_{n_m \cdot n_s} \end{bmatrix} =: \begin{bmatrix} \hat{v}^j_1 \\ \hat{v}^j_2 \\ \vdots \\ \hat{v}^j_{n_s} \end{bmatrix}.$$

Or analogously (in a continuous version of POD) as

$$x(t;\Theta_\ell) \approx \sum_{j=1}^{n_r} b_j(t)\hat{v}^j_\ell, \tag{6}$$

for coefficient functions $b_j(t) = \langle x(t;\Theta_\ell), \hat{v}^j \rangle$.

2.3. Construction of Data-Fitting Function

Next, we want to construct a data-fitting function $\Psi(t) \in \mathbb{R}^{n_m}$ that fits the known experimental data for the component $k = 1, \ldots, n_d$. We assume $\Psi(t)$ to be of the form

$$\Psi(t) := \sum_{\ell=1}^{n_s} c_\ell \bar{x}^k_\ell(t) = \sum_{\ell=1}^{n_s} c_\ell x(t;\Theta_\ell).$$

With (6), we obtain

$$\Psi(t) = \sum_{\ell=1}^{n_s} c_\ell \sum_{j=1}^{n_r} b_j(t)\hat{v}^j_\ell = \sum_{\ell=1}^{n_s} \sum_{j=1}^{n_r} c_\ell \langle x(t;\Theta_\ell), \hat{v}^j_\ell \rangle \hat{v}^j_\ell.$$

Usually only a few specific states of the dynamical system can be measured in experiments. Here, we assume that the first n_d components of the state vector can be measured and we want to determine the coefficients c_ℓ in such a way that the first n_d components of $\Psi = [\Psi_k]^{n_d}_{k=1}$ are an acceptable fit of the experimental data points, i.e., $\Psi_k(\tau_s) \approx y_k(\tau_s)$ for $s = 1, \ldots, n_\tau$ and $k = 1, \ldots, n_d$.

Thus, we construct a linear system to be solved for c_ℓ:

$$y_k(\tau_s) \overset{!}{=} \Psi_k(\tau_s) = \left[\sum_{\ell=1}^{n_s} \sum_{j=1}^{n_r} c_\ell b_j(\tau_s)\hat{v}^j_\ell \right]_k = \left[\sum_{\ell=1}^{n_s} \sum_{j=1}^{n_r} c_\ell \langle x(\tau_s;\Theta_\ell), \hat{v}^j_\ell \rangle \hat{v}^j_\ell \right]_k$$

for all $k = 1, \ldots, n_d$ and $s = 1, \ldots, n_\tau$.

We obtain $n_d \cdot n_\tau$ equations for the n_s unknowns c_ℓ, $\ell = 1, \ldots, n_s$. For each $k = 1, \ldots, n_d$, we obtain a linear system

$$A^k c = y^k$$

where $A^k = [a_{s\ell}^k]$, and $a_{s\ell}^k := \sum_{j=1}^{n_r} b_j(\tau_s) V((\ell-1)n_m + k, j)$, for $\ell = 1, \ldots, n_s$, $s = 1, \ldots, n_\tau$

and $y^k = \begin{bmatrix} y_k(\tau_1) \\ \vdots \\ y_k(\tau_{n_\tau}) \end{bmatrix}$. Summarizing all A^k and y^k into one large system yields $Ac = y$ with

$A \in \mathbb{R}^{n_d n_\tau \times n_s}$ and $y \in \mathbb{R}^{n_d n_\tau}$.

2.4. Estimating the Parameters

Once we have a representation for $\Psi(t)$ we can substitute Ψ and

$$\dot{\Psi}_i \approx \frac{\Psi(t_{i+1}) - \Psi(t_i)}{h_s}$$

as approximation for $\dot{\Psi}_i$ into the ODE (1) at the simulation time points t_i to obtain

$$\frac{\Psi(t_{i+1}) - \Psi(t_i)}{h_s} \approx f(t_i, \Psi(t_i), \theta), \quad i = 1, \ldots, n_t.$$

Here, Ψ can be assumed to be smooth since it is composed as linear combination of solution of an ODE. Inserting all information that we have gives

$$\frac{\Psi(t_{i+1}) - \Psi(t_i)}{h_s} = f(t_i, \Psi(t_i), \theta)$$

$$\frac{\sum_{\ell=1}^{n_s} c_\ell x(t_{i+1}; \Theta_\ell) - \sum_{\ell=1}^{n_s} c_\ell x(t_i; \Theta_\ell)}{h_s} = f(t_i, \sum_{\ell=1}^{n_s} c_\ell x(t_i; \Theta_\ell), \theta).$$

This system can be solved for the parameter θ by a nonlinear system solver (e.g., `fsolve` in Matlab). We can sum up all these steps in the following Algorithm 1.

Algorithm 1 Parameter estimation for a kinetic model.

---

1. Compute n_s solutions of a kinetic model given by (1) for different parameter values $\Theta_1, \ldots, \Theta_{n_s}$ and form the snapshot matrix X.
2. Compute the SVD of X. Truncate and remain only the most dominant singular values n_r to obtain the POD basis.
3. Express every vector in the snapshot matrix X as linear combination of the POD basis $\bar{x}^i = \sum_{j=1}^{n_r} b_{ij} v^j$ with $b_{ij} = (v^j)^T \cdot \bar{x}^i$ for $i = 1, \ldots, n_t$.
4. Construct a data-fitting function that fits the known experimental data ($y(\tau_s)$) as a linear combination of POD basis $\Psi(t) := \sum_{\ell=1}^{n_s} c_\ell \bar{x}_\ell^i(t)$.
5. Solve the linear system $\Psi(\tau_s) = y(\tau_s)$, $s = 1, \ldots, n_\tau$ to obtain the value of c_l, $l = 1, \ldots, n_s$.
6. Substitute the value of ψ into the original model $\frac{\Psi(t_{i+1}) - \Psi(t_i)}{h_s} \approx f(t_i, \Psi(t_i), \theta)$, $i = 1, \ldots, n_t$ and solve the system for θ parameter.

---

3. Application of the Parameter Estimation Method to a Kinetic Model of CCR in *E. coli*

In this section, we apply the parameter estimation procedure introduced in the previous section to the kinetic model of carbon catabolite repression (CCR) in *E. coli* [28]. CCR is the main regulatory mechanism in *E. coli* for the control of carbohydrate uptake. The regulatory network is strongly hierarchical with a regulator Crp (cyclic AMP receptor protein) on the top. Crp serves as accelerator for gene expression in case of low C source availability and in this way coordinates different subsystems of the cell, responsible for the uptake of carbon sources, their breakdown for the production of energy and precursors, and the conversion of the latter to biomass. In the second level, substrate specific regulator proteins (main repressor proteins) are inactivated in case the specific substrate is available

in the growth medium. The interplay between regulators and the transcription apparatus results for example in the observation that only one substrate is taken up although two substrates are provided in the beginning. Two distinct growth phases are observed, and a sequential uptake of the substrates take place (diauxic growth).

3.1. The Kinetic Model of the Carbon Catabolite Repression

The main structure of the reaction scheme is outlined in Figure 1. Two extracellular substrates (S_1 and S_2) are taken up in enzyme catalysed reactions by their respective specific transporters (E_1 and E_2). The substrates have representatives inside the cell (X_1 and X_2). For example, in the case of glucose, X_1 represents the phosphorylated form glucose 6-phospahte while in the case of lactose, the substrate is not modified during the transport step. In general, transport systems for carbohydrate are inducible, that is, metabolites in the cell activate gene expression of the respective enzymes. In the model, these mechanisms are represented by a positive feedback from the X_i to the synthesis of the respective enzymes E_i (dashed blue lines). A special case for the interaction between two uptake systems in case of glucose and lactose is inducer exclusion (red dashed line); high levels of X_1 inhibit the activity of the second transporter for lactose uptake. The inhibition is so strong that the level of X_2 is minimal and, this way, the transporter E_2 cannot be synthesized. Only after they run out of glucose, the level of X_1 decreases and the inhibition is degraded. Both uptake branches converge in the central pathways, represented by metabolite M. From M, the synthesis of the main biomass compartment B' is described with a single reaction.
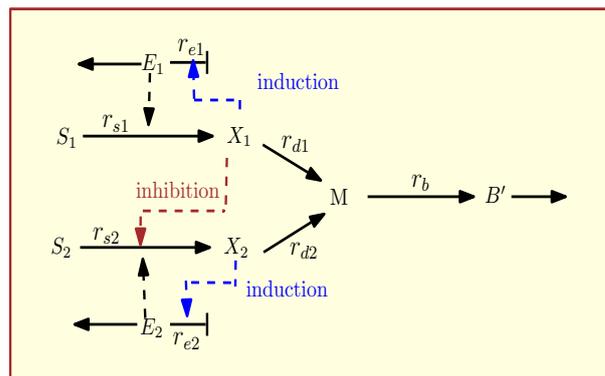
Figure 1. Reaction scheme for the diauxic growth network. S_1, S_2 are the substrates, X_1, X_2 are the intracellular metabolites, E_1, E_2 are the enzymes, M is the intermediate metabolite, and B' is the main Biomass. The reaction rates are indicated in (solid arrows), catalytic activities and regulatory interactions in (dashed arrows).

The system of differential equations associated with the chemical reactions network is given by

$$\dot{x}(t) = N \cdot r(x(t)),$$

where $x = [S_1, S_2, X_1, X_2, E_1, E_2, M, B']^T \in \mathbb{R}^8$ is the vector of metabolite concentrations, $r = [r_{s1}, r_{s2}, r_{d1}, r_{d2}, r_{e1}, r_{e2}, r_b]^T$ is the vector of reaction rates and N is the stoichiometric matrix given as follows:

$$N = \begin{array}{c} \\ S_1 \\ S_2 \\ X_1 \\ X_2 \\ E_1 \\ E_2 \\ M \\ B' \end{array} \begin{array}{ccccccc} r_{s1} & r_{s2} & r_{d1} & r_{d2} & r_{e1} & r_{e2} & r_b \\ \left[\begin{array}{ccccccc} -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right] \end{array} \in \mathbb{R}^{8 \times 7}.$$

The kinetic rate laws r_{s1}, r_{s2} are defined by Michaelis–Menten kinetics [34] and $r_{d1}, r_{d2}, r_{e1}, r_{e2}, r_b$ by mass action kinetics [35] as follows:

$$r_{s1} = k_{s1} E_1 \frac{S_1}{K_1 + S_1}, \qquad r_{s2} = k_{s2} E_2 \frac{S_2}{K_2 + S_2},$$
$$r_{e1} = k_{e1} f_1, \qquad r_{e2} = k_{e2} f_2,$$
$$r_{d1} = k_{x1} X_1, \qquad r_{d2} = k_{x2} X_2, \quad r_b = k_m M,$$

where k_{x1}, k_{x2}, and k_m are the constant rates all in $1/h$ unit, K_1, K_2 are the Michaelis–Menten constant in g/L unit and k_{e1}, k_{e2} are the constant rates in the mol/gDW.h. The turnover number of enzymes k_{s1}, k_{s2} is in the unit $1/h$.

The system of ordinary differential equations of the carbon catabolite repression is given by

$$\dot{S}_i = -r_{si} w_i B, \quad i = 1, 2$$
$$\dot{X}_i = r_{si} - r_{di},$$
$$\dot{E}_i = r_{ei} - \mu E_i,$$
$$\dot{M} = r_{d1} + r_{d2} - r_b,$$
$$\dot{B} = \mu B,$$
$$\dot{B}' = r_b - \mu B',$$

where w_i is the molecular weight of the substrates (g/mol). The main biomass compartment B' consists of macromolecular species like protein, DNA, RNA, lipid, etc., and variable B is the entire biomass, growing with the specific growth rate μ. The growth rate $\mu = Y_1 r_{s1} + Y_2 r_{s2}$ has the unit $1/h$ and Y_1, Y_2 are yield coefficients with unit gDW/mol. The process of induction is described with rates f_1, f_2 that are expressed in the following form

$$f_1 = \frac{X_1}{\beta_1 + X_1}, \qquad f_2 = \frac{X_2}{\beta_2 + X_2},$$

where β_1, β_2 are in the unit mol/gDW and k_1, k_2 are constant rates in mol/gDW.h unit. The intracellular components are assumed to be at the steady state i.e., $\dot{X}_1 = 0$, $\dot{X}_2 = 0$, and $\dot{M} = 0$. This case study considers inducer exclusion where metabolite $X1$ works as a regulatory metabolite and inhibits enzyme $E2$ which is responsible for substrate $S2$ uptake.

In a scheme, this is represented by an extension of the uptake rate r_{s2} to $r_{s2} \cdot \dfrac{\alpha}{\alpha + X_1}$, where α is a constant. Then the kinetic model of the above network is given as follows

$$
\begin{aligned}
\dot{S}_1 &= -\frac{k_{s1} \cdot E_1 \cdot S_1}{K_1 + S_1} \cdot w_1 \cdot B, \\
\dot{S}_2 &= -\frac{k_{s2} \cdot E_2 \cdot S_2}{K_2 + S_2} \cdot w_2 \cdot B \cdot f_{inhibit}, \\
\dot{E}_1 &= k_{e1} \frac{X_1}{\beta_1 + X_1} - \mu E_1, \\
\dot{E}_2 &= k_{e2} \frac{X_2}{\beta_2 + X_2} - \mu E_2, \\
\dot{B} &= \mu B, \\
\dot{B}' &= k_m \cdot M - \mu B',
\end{aligned}
\tag{7}
$$

where

$$
X_1 = \frac{k_{s1} \cdot E_1 \cdot S_1}{k_{x1}(K_1 + S_1)}, \quad X_2 = \frac{k_{s2} \cdot E_2 \cdot S2}{k_{x2}(K_2 + S_2)} \cdot f_{inhibit}
$$
$$
M = \frac{k_{x1} \cdot X_1}{k_m} + \frac{k_{x2} \cdot X_2}{k_m}, \quad f_{inhibit} = \frac{\alpha}{\alpha + X_1}.
$$

3.2. Numerical Results

In the following we present some examples in which we estimate certain parameter values of the kinetic model (7). All results presented in the following sections were computed with MATLAB (R2022a).

3.2.1. Estimation of the Parameter k_{s1}

In this first example we estimate one parameter value, namely, the uptake rate constant k_{s1}. First, we compute the snapshot matrix X from the simulation of the kinetic model (7) for different values of the parameter k_{s1} over the time interval $[0, 7]$ h. The different values of parameter k_{s1} are equidistantly distributed in the interval $[10^{-5}, 10^{-2}]$, which are the lower and upper bound of parameter value, respectively. In each of n_s simulation, we evaluate at n_t simulation time points. In this example, we set $n_s = 50$. The remaining constant reaction rates are given in Table 1 and the parameter value of $k_{s2} = 0.0033$ is assumed. We use the MATLAB function ode15s with tolerances $RTOL = ATOL = 10^{-6}$ to compute a numerical solution of the ODE system (7) with 50 different values for the parameter k_{s1}. The initial concentrations of extracellular and intracellular metabolites are assumed to be $(S_1, S_2, E_1, E_2, B, B') = (0.22, 1.18, 0.1, 0.1, 0.032, 0.5) \in \mathbb{R}^6$.

We used the MATLAB function svd to calculate the singular value decomposition of X. The singular values are shown in Figure 2.
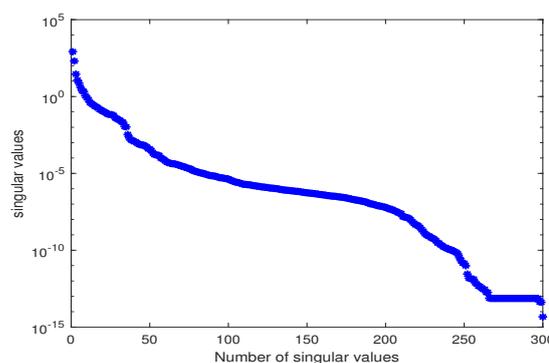
Figure 2. Singular values of snapshot matrix X for estimating parameter k_{s1}.

Table 1. The set of all parameters of the kinetic model that are taken from literature [28].

Constant Rates	Value	Unit
K_1	1×10^{-4}	g/L
K_2	0.2	g/L
k_{e1}	6	mol/gDWh
k_{e2}	8.69	mol/gDWh
k_{x1}	10	1/h
k_{x2}	10	1/h
k_m	10	1/h
β_1	1	mol/gDW
β_2	0.6	mol/gDW
α	11.05	mol/gDW
Y_1	90	gDW/mol
Y_2	102.6	gDW/mol
w_1	180	gDW/mol
w_2	342	gDW/mol

The behavior of the model order reduction method strongly depends on the decay of the singular values of the snapshot matrix. Figure 2 shows gradually decaying singular values with a strong decay for the smallest singular values indicating that ignoring these will not lead to any considerable loss of information. We truncate the most dominant singular values such that the relative information coefficient I is close to one. In this example, we use $n_r = 60$ where $I = 0.99$. The date-fitting function $\Psi(t)$ is constructed as in Section 2.3 to fit the experimental data set for the three measurement components S_1, S_2, and B at the measurement time points $n_\tau = 15$. The trajectories of the measurement data set and the data-fitting function are depicted in Figure 3.

(a) (b) (c)
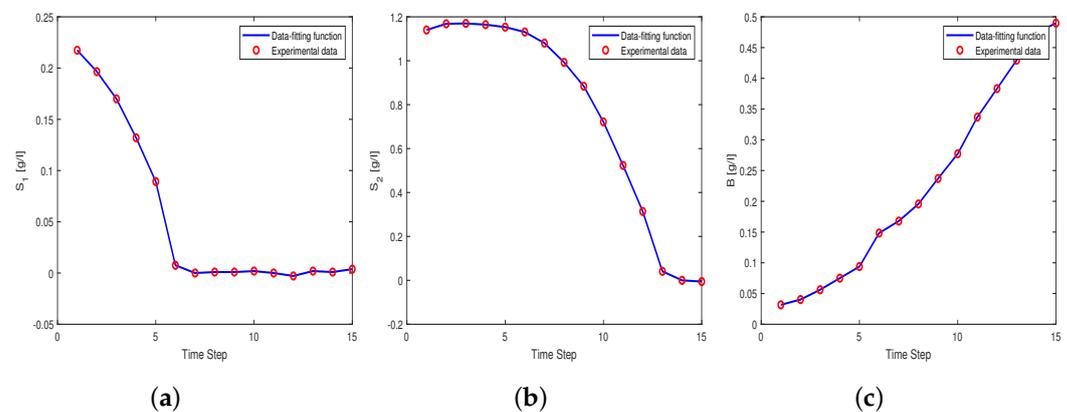
Figure 3. (**a**) trajectory of substrate S_1 using data-fitting function with a solid line and experimental data with a circle. (**b**) The trajectory of substrate S_2. (**c**) The trajectory of Biomass B.

We can observe that the data-fitting function fits well with the measurement data. Once the data-fitting function $\Psi(t)$ is obtained, we can substitute the function $\Psi(t)$ in the kinetic model (7) and solve the system for the parameter k_{s1}. We use the MATLAB function `fsolve` with function tolerance $TOL = 10^{-6}$ using the Levenberg–Marquardt algorithm with initial guess $k_{s1} = 0.0167$. The solution gives an estimated parameter value of $k_{best} = 0.018$. For the purpose of comparison, we also provide results from the least square method (LSQ). We perform the LSQ method using MATLAB function `lsqnonlin` and `lhsdesign` with the same assumptions (e.g., number of fits n_s and lower and upper bounds) as in the POD method. The solution gives an estimated value $k_{lsq} = 0.01$ of the parameter k_{s1}. The simulation results of the kinetic model with k_{best} in comparison with the

simulation results of the kinetic model with k_{lsq}, and the measurement data are depicted in Figure 4.
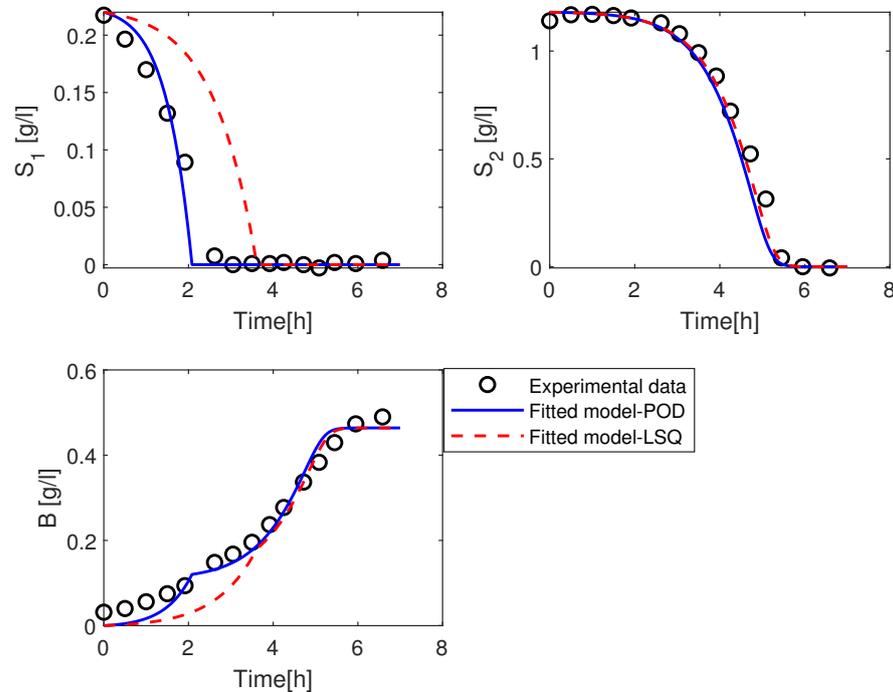
Figure 4. Simulation of the solution of the kinetic model with the parameter value k_{s1} using POD method in blue (solid line), LSQ method in red (dashed line), and the experimental data in black (open circles).

We can observe that the dynamical behavior of the kinetic model states with parameter value k_{best} matches well with the experimental data for the substrates S_1, S_2, and the Biomass B. While the simulation of the solution of the kinetic model with parameter value k_{lsq} for the substrate S_1 is unable to match well with the data. The substrate S_1 spends more time to consume compared to the measured data. In contrast, the trajectories of S_2 and Biomass still fit good with the measured data. From the above figure, the POD method gives better results than the LSQ method. Note, that with a change of the lower and upper bounds in the LSQ method, we could obtain a better result. Both methods are sensitive to assumptions, which means that the lower and upper bound of parameter value lb, ub, the mesh size M, and the reduced basis n_r can be adjusted to obtain a good fitting.

3.2.2. Estimation of the Parameters k_{s1} and k_{s2}

In this example, we consider the two parameters for the uptake rate constants k_{s1} and k_{s2}. The snapshot matrices X is computed from $n_s = 40$ simulation of the kinetic model equations (7) over the time interval. We use a Latin hypercube sampling (LHS) to obtain different values of the parameters via the MATLAB function `lhsdesign`. To generate parameter sampling for the parameters k_{s1} and k_{s2}, we set the lower and upper bound for the parameters as $lb = 10^{-5}$, $ub = 10^{-2}$. We use the MATLAB function `svd` to calculate the singular value decomposition of X. The singular values are depicted in Figure 5.
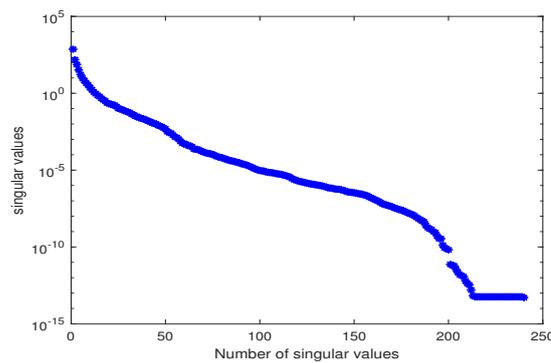
Figure 5. Singular values of the snapshot matrix X for estimating Parameters k_{s1} and k_{s2}.

Here, we truncate at $n_r = 60$, with relative information coefficient $I = 0.9$. Substituting ψ in the kinetic model and solving the system for the two parameters using MATLAB function `fsolve` with function tolerance 10^{-6} and initial guesses $k_{s1} = 0.0167, k_{s2} = 0.0033$, we obtain the estimated values of the parameters $k_{best} = [0.0194, 0.0032]$. In this example, we perform the LSQ method with different assumptions of bounds. We set the lower and upper bounds in the LSQ method to be $lb = 10^{-3}$, $ub = 10^{-1}$, we obtain the estimated values $k_{lsq} = [0.0163, 0.0029]$.

From Figure 6, we can observe that the trajectories for the experimental data and the simulation results of kinetic model with the estimated parameter values using POD method match good for the substrates trajectories S_1, S_2 and Biomass curve. While the simulation of the kinetic model with parameters value using the LSQ method is slightly different from the measured data.
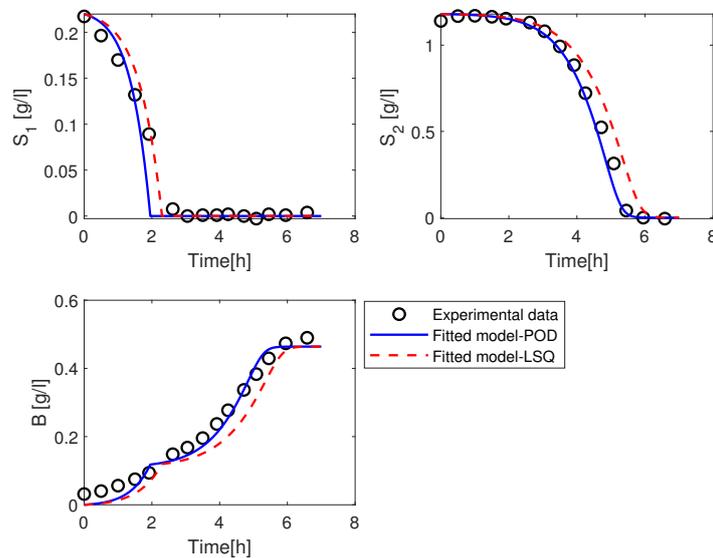
Figure 6. Simulation of the solution of the kinetic model with the parameters value k_{s1} and k_{s2} using POD method in blue (solid line), LSQ method in red (dashed line), and the experimental data in black (open circles).

3.2.3. Estimation of the Parameters k_{s1}, k_{e1}, k_{x1} and k_{x2}

In this example, we apply the method to estimate four parameters in the kinetic model (7). We use the MATLAB function `lhsdesign` to generate parameter sampling for the parameters k_{s1}, k_{s2}, k_{e2}, and β_2, where the lower and upper bound for the parameters are $[10^{-2}, 10^{+1}]$. We set the number of simulation to $n_s = 20$. The snapshot matrix X is computed and the SVD is applied; Figure 7 shows the singular values. We choose $n_r = 60$

and the system is solved for the four parameters using MATLAB function `lsqnonlin` with options algorithm `levenberg-marquardt` and initial guess of parameters $k_{s1} = 0.0167$, $k_{e1} = 6$, $k_{x1} = 10$, and $k_{x2} = 10$. The estimated values of the parameters are *kbest* = $[0.01, 125.7, 103.6, 9.97]$. We perform the LSQ method using the same assumptions as in the POD method, the solution gives an estimated values $k_{lsq} = [0.059, 0.084, 0.333, 9.99]$.
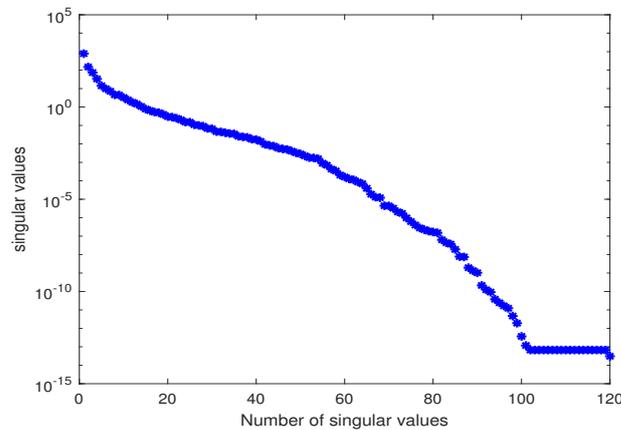
Figure 7. Singular values of the snapshot matrix X for estimating parameters k_{s1}, k_{e1}, k_{x1} and k_{x2}.

At the beginning of the simulation, we can observe from Figure 8 that the trajectory of S_1 using the estimated values from the POD method consumes slowly compared to the trajectories by the LSQ method and the experimental data. Around the time 2 h, it is exhausted a bit earlier than other curves. The trajectories of S_2 and B from the POD method are qualitatively similar to the measured data while their trajectories of the LSQ method are slightly different.
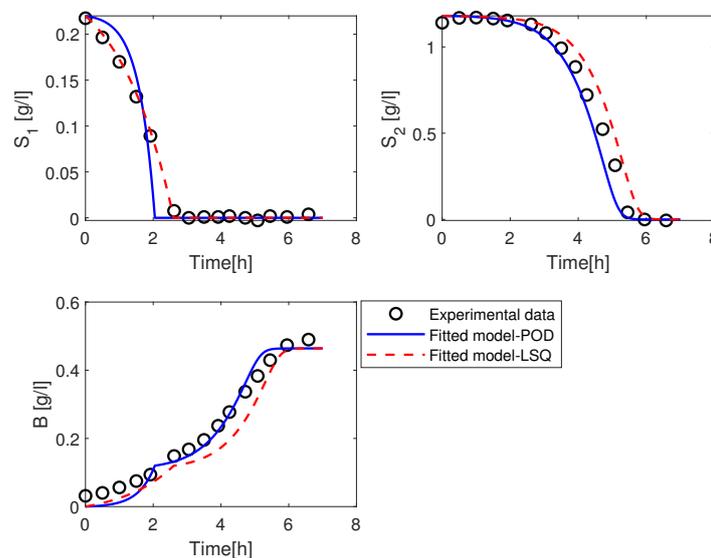
Figure 8. Simulation of the solution of the kinetic model with the parameters value k_{s1}, k_{e1}, k_{x1} and k_{x2} using POD method in blue (solid line), LSQ method in red (dashed line), and the experimental data in black (open circles).

4. Conclusions and Discussion

In this paper, we have discussed the parameter estimation procedure using model order reduction by the POD method for a kinetic model of a cellular system. The process of finding parameter values is considered to be a main challenge in mathematical modelling

of biological system, so the possibility to use the POD method in the parameter estimation problem is of crucial importance. We have applied the POD method and the singular value decomposition to obtain the reduced basis elements, which are used for constructing a function that fits the trajectory of the kinetic model to the observable data set of the biological system. The parameter estimation using the POD method is capable to estimate the uptake rate constants and other parameters of the CCR kinetic model. The solution of the kinetic model with the estimated parameter values matches good with the experimental data set. We performed the least square method to estimate parameter values of the CRR kinetic model for the sake of comparison with the POD method and show to what extent the POD method could estimate the values of parameters of the CRR kinetic model. We could observe that the number of reduced basis, the lower and upper bounds, and the number of simulations play an important role in the parameter estimation problem. For every parameter estimation case, these numbers can be adjusted until the desired results are obtained. In addition, the presented algorithm of parameter estimation can be used as a reference for different kinetic models of biological systems. The presented approach has some limitations as it strongly depends on the behavior of the singular value spectrum. If there is only a small number of dominant eigenmodes, then the singular values will rapidly decay and only a small number of them is enough to capture the characteristic behavior of the system. However, in some ill-posed cases, this might not be the case and the heuristic approach presented in the algorithm might not lead to satisfactory results. The choice of a suitable threshold for discarding certain singular values also depends on the quality of the data, in particular, rounding and approximation errors. It has been shown in [36] that the truncated SVD is a suitable method for the regularization of an ill-posed problem when the coefficient matrix is ill-conditioned with a well-determined numerical rank. In this case, the solution obtained by the truncated SVD with a truncation threshold equal to the numerical rank of the matrix is guaranteed to be similar to the regularized solution where the regularization parameter is chosen near its intuitive optimum value.

Author Contributions: N.A.E. and L.S. designed the study. N.A.E. performed the simulation studies and wrote the first draft of the paper. A.K. provided a kinetic model for the CCR network. All authors discussed the results. All authors have read and agreed to the published version of the manuscript.

References

1. Mannan, A.A.; Toya, Y.; Shimizu, K.; McFadden, J.; Kierzek, A.M.; Rocco, A. Integrating kinetic model of *E. coli* with genome scale metabolic fluxes overcomes its open system problem and reveals bistability in central metabolism. *PLoS ONE* **2015**, *10*, e0139507. [CrossRef] [PubMed]
2. Lima, A.P.; Baixinho, V.; Machado, D.; Rocha, I. A comparative analysis of dynamic models of the central carbon metabolism of Escherichia coli. *IFAC-PapersOnLine* **2016**, *49*, 270–276. [CrossRef]
3. Ali Eshtewy, N. *Mathematical Modeling of Metabolic-Genetic Networks*; Freie Universität Berlin: Berlin, Germany, 2020.
4. Aster, R.C.; Borchers, B.; Thurber, C.H. *Parameter Estimation and Inverse Problems*; Elsevier: Amsterdam, The Netherlands, 2018.
5. Moles, C.G.; Mendes, P.; Banga, J.R. Parameter estimation in biochemical pathways: A comparison of global optimization methods. *Genome Res.* **2013**, *13*, 2467–2474. [CrossRef] [PubMed]
6. Tarantola, A. *Inverse Problem Theory and Methods for Model Parameter Estimation*; SIAM: Philadelphia, PA, USA, 2005; Volume 89.
7. Zhdanov, M.S. *Inverse Theory and Applications in Geophysics*; Elsevier: Amsterdam, The Netherlands, 2015; Volume 36.
8. Parker, R.L. Understanding inverse theory. *Annu. Rev. Earth Planet. Sci.* **1977**, *5*, 35–64. [CrossRef]
9. Kreutz, C. An easy and efficient approach for testing identifiability. *Bioinformatics* **2018**, *34*, 1913–1921. [CrossRef]

10. Raue, A.; Schilling, M.; Bachmann, J.; Matteson, A.; Schelke, M.; Kaschek, D.; Hug, S.; Kreutz, C.; Harms, B.D.; Theis, F.J.; et al. Lessons Learned from Quantitative Dynamical Modeling in Systems Biology. *PLoS ONE* **2013**, *8*, e74335. [CrossRef]

11. Deuflhard, P. *Newton Methods for Nonlinear Problems: Affine Invariance and Adaptive Algorithms*; Springer: Berlin/Heidelberg, Germany, 2011; Volume 35.

12. Altman, N.; Krzywinski, M. *Points of Significance: Simple Linear Regression*; Nature Publishing Group: London, UK, 2015.

13. Björck, Å. *Numerical Methods for Least Squares Problems*; Society for Industrial and Applied Mathematics: Philadelphia, PA, USA, 1996.

14. Enders, C.K. Maximum likelihood estimation. In *Encyclopedia of Statistics in Behavioral Science*; Wiley: Hoboken, NJ, USA, 2005.

15. Rudi, J.; Bessac, J.; Lenzi, A. Parameter Estimation with Dense and Convolutional Neural Networks Applied to the FitzHugh—Nagumo ODE. *Math. Sci. Mach. Learn.* **2022**, 781–808. [CrossRef]

16. Kerschen, G.; Golinval, J.; Vakakis, A.F.; Bergman, L.A. The method of proper orthogonal decomposition for dynamical characterization and order reduction of mechanical systems: An overview. *Nonlinear Dyn.* **2005**, *41*, 147–169. [CrossRef]

17. Volkwein, S. Proper orthogonal decomposition: Theory and reduced-order modelling. *Lect. Notes Univ. Konstanz* **2013**, *4*, 1–29.

18. Karhunen, K. Über lineare Methoden in der Wahrscheinlichkeitsrechnung. *Ann. Acad. Sci. Fenn. Ser. A Math. Phys.* **1947**, *47*, 1–79.

19. Loeve, M. *Elementary Probability Theory*; Springer: Berlin/Heidelberg, Germany, 1977; pp. 1–52.

20. Benner, P.; Goyal, P.; Heiland, J.; Duff, I.P. Operator inference and physics-informed learning of low-dimensional models for incompressible flows. *arXiv* **2020**, arXiv:2010.06701.

21. Cazemier, W.; Verstappen, R.W.C.P.; Veldman, A.E.P. Proper orthogonal decomposition and low-dimensional models for driven cavity flows. *Phys. Fluids* **1998**, *10*, 1685–1699. [CrossRef]

22. Luo, Z.; Chen, J.; Navon, I.M.; Yang, X. Mixed finite element formulation and error estimates based on proper orthogonal decomposition for the nonstationary Navier–Stokes equations. *SIAM J. Numer. Anal.* **2009**, *47*, 1–19. [CrossRef]

23. Ali Eshtewy, N.; Scholz, L. Model Reduction for Kinetic Models of Biological Systems. *Symmetry* **2020**, *12*, 863. [CrossRef]

24. Rehm, A.M.; Scribner, E.Y.; Fathallah-Shaykh, H. Proper orthogonal decomposition for parameter estimation in oscillating biological networks. *J. Comput. Appl. Math.* **2014**, *258*, 135–150. [CrossRef]

25. Ly, H.V.; Tran, H.T. Proper orthogonal decomposition for flow calculations and optimal control in a horizontal CVD reactor. *Q. Appl. Math.* **2002**, *60*, 631–656. [CrossRef]

26. Boulakia, M.; Schenone, E.; Gerbeau, J.-F. Reduced-order modeling for cardiac electrophysiology. Application to parameter identification. *Int. J. Numer. Methods Biomed. Eng.* **2012**, *28*, 727–744. [CrossRef]

27. Kahlbacher, M.; Volkwein, S. Estimation of regularization parameters in elliptic optimal control problems by POD model reduction. In Proceedings of the IFIP Conference on System Modeling and Optimization, Cracow, Poland, 23–27 July 2007; Springer: Berlin/Heidelberg, Germany, 2007; pp. 307–318.

28. Kremling, A.; Geiselmann, J.; Ropers, D.; De Jong, H. An ensemble of mathematical models showing diauxic growth behaviour. *BMC Syst. Biol.* **2018**, *12*, 82. [CrossRef]

29. Helton, J.C.; Davis, F.J. Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. *Reliab. Eng. Syst. Saf.* **2003**, *81*, 23–69. [CrossRef]

30. McKay, M.D.; Beckman, R.J.; Conover, W.J. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* **2000**, *42*, 55–61. [CrossRef]

31. Golub, G.H.; Van Loan, C.F. *Matrix Computations*, 3rd ed.; Johns Hopkins University Press: Baltimore, MD, USA, 1996.

32. Hinze, M.; Volkwein, S. Model Order Reduction by Proper Orthogonal Decomposition. *Konstanz. Schriften Math.* **2019**, 47–96. [CrossRef]

33. Afanasiev, K.; Hinze, M. Adaptive Control of a Wake Flow Using Proper Orthogonal Decomposition. *Lect. Notes Pure Appl. Math* **2001**, *216*, 317–332.

34. Michaelis, L.; Menten, M.L. Die Kinetik der Invertinwirkung. *Biochem. Z.* **1913**, *49*, 333–369.

35. Guldberg, C.M.; Waage, P. *Etudes Sur Les Affinités Chimiques*; Brøgger & Christie: Christiania, Denmark, 1867.

36. Per Christian, H. The truncatedSVD as a method for regularization. *Bit Numer. Math.* **1987**, *27*, 534–553.