IET Systems Biology

Research Article

IET Journals
The Institution of Engineering and Technology

Identification of a time-varying intracellular signalling model through data clustering and parameter selection: application to NF-κB signalling pathway induced by LPS in the presence of BFA

ISSN 1751-8849
Received on 5th September 2018
Revised 7th February 2019
Accepted on 14th February 2019
E-First on 24th April 2019
doi: 10.1049/iet-syb.2018.5079
www.ietdl.org

Dongheon Lee[1,2], Arul Jayaraman[1,3], Joseph Sang-Il Kwon[1,2] ✉

[1]*Artie McFerrin Department of Chemical Engineering, Texas A&M University, College Station, TX 77843, USA*
[2]*Texas A&M Energy Institute, Texas A&M University, College Station, TX 77843, USA*
[3]*Department of Biomedical Engineering, Texas A&M University, College Station, TX 77843, USA*
✉ *E-mail: kwonx075@tamu.edu*

Abstract: Developing a model for a signalling pathway requires several iterations of experimentation and model refinement to obtain an accurate model. However, the implementation of such an approach to model a signalling pathway induced by a poorly-known stimulus can become labour intensive because only limited information on the pathway is available beforehand to formulate an initial model. Therefore, a large number of iterations are required since the initial model is likely to be erroneous. In this work, a numerical scheme is proposed to construct a time-varying model for a signalling pathway induced by a poorly-known stimulus when its nominal model is available in the literature. Here, the nominal model refers to one that describes the signalling dynamics under a well-characterised stimulus. First, global sensitivity analysis is implemented on the nominal model to identify the most important parameters, which are assumed to be piecewise constants. Second, measurement data are clustered to determine temporal subdomains where the parameters take different values. Finally, a least-squares problem is solved to estimate the parameter values in each temporal subdomain. The effectiveness of this approach is illustrated by developing a time-varying model for NF-κB signalling dynamics induced by lipopolysaccharide in the presence of brefeldin A.

1 Introduction

Through various intracellular signalling pathways, cells are able to modify their gene expression, metabolism, or other regulatory actions in order to sense and respond to perturbations in their environment [1]. Due to intrinsic complexity of the signalling pathways, a systems biology approach, which integrates experimental measurements and mathematical modelling, has become indispensable to gain a system-level understanding of the signalling dynamics [1–4]. Specifically, a priori knowledge about a signalling pathway is formulated into a mathematical model, usually in the form of a system of ordinary differential equations, and the model is subsequently calibrated by experimental observations. The resulting model is then used to analyse the underlying mechanisms and generate new hypotheses to be tested in new experiments. Previously, this systems biology approach has been implemented to gain new insights into various signalling pathways such as the nuclear factor κB (NF-κB) signalling pathway [5, 6], Janus family of kinases – signal transducer and activator of transcription signalling pathway [7], and mitogen-activated protein kinase signalling pathway [8].

Ideally, one would like to develop a comprehensive signalling pathway model that can predict the signalling dynamics under various conditions. However, this can be a difficult task as a single signalling pathway can be activated by many stimuli with different, usually unknown, corresponding reaction mechanisms. For example, out of around 100 different stimuli of the NF-κB signalling pathway [9], only a handful, such as tumour necrosis factor-α (TNFα) and lipopolysaccharide (LPS), and their reaction mechanisms are well characterised. Consequently, investigating and modelling a signalling pathway induced by a stimulus, which has not been well studied, is non-trivial. Specifically, as the pathway is only partially known beforehand, a number of different model structures need to be formulated and discriminated, which can become very challenging. In the literature, different approaches have been proposed and implemented for solving this problem. First, extensive experiments are performed to characterise as many reactions as possible between intracellular molecules, which in practice is nearly impossible due to the large number of interactions to be studied. Second, the aforementioned iterative approach between modelling and experiments can be implemented to improve the model gradually [2, 10, 11]. However, as the initial model is likely to be erroneous, this approach may require a large number of iterations between experiments and model refinement to reach a relatively satisfying model. Third, a number of different initial models, each of which corresponds to a different hypothesis on the signalling pathway structure, are synthesised from the beginning, and the best model structure is selected by solving an optimisation problem against experimental observations [12–14]. Although the optimisation-based approach is promising, it has several numerical and algorithmic challenges such as efficiently finding a global optimal solution for a large system [12–16].

Alternatively, when an accurate model for describing a signalling pathway under one stimulus is available, we can modify that model to describe the same signalling pathway under a lesser-known stimulus [17, 18]. Hereafter, we refer to the model constructed for the well-studied stimulus as the nominal model. The rationale for using the nominal model is two-fold. First, the nominal model already contains a number of important pathway components as well as their interactions, which are likely to be important under the lesser-known stimuli as well. Second, this approach avoids the lengthy model selection procedure, which requires a number of different candidate models to be synthesised, calibrated, and compared [17]. On the other hand, the structure of the nominal model is likely to be insufficient to describe the signalling dynamics under the lesser-known stimulus due to unincorporated and unknown reactions and components specific to this stimulus [17]. Therefore, a poorly characterised signalling pathway induced by a lesser-known stimulus needs to be described by a data-driven approach to complement the inaccuracy of the

nominal model. Here, we choose to introduce time-varying parameters to the nominal model, which is usually time-invariant, based on the available experimental measurements [18, 19]. Through this data-driven approach, a more accurate model for the lesser-known stimulus can be derived based on the nominal model and the available data.

Motivated by the above considerations, we propose a numerical scheme to construct a time-varying model to simulate an intracellular signalling pathway with a lesser-known stimulus based on a nominal model. First, global sensitivity analysis (SA) is performed on the nominal model to identify a set of parameters that are identifiable given the model structure and experimental observations, and only these parameters are assumed to vary with time. Next, the temporal profiles of the model parameters are partitioned into several temporal subdomains whose boundaries are determined by clustering the experimental observations. And the parameters determined by the SA have fixed values in each temporal subdomain. Finally, a least-squares problem is solved to estimate the values of the parameters in each temporal subdomain by minimising the difference between the model predictions and the experimental data.

The paper is organised as follows: first, the motivation for formulating a time-varying signalling model is presented. Second, the proposed methodology that consists of the optimal temporal clustering and the global SA to construct a time-varying model is presented in details. Finally, the proposed methodology is implemented to develop the model for a time-varying NF-κB signalling pathway induced by LPS in the presence of brefeldin A (BFA) to assess the efficiency and accuracy of the proposed scheme.

2 Background

2.1 System description

Consider an intracellular signalling pathway initiated by an external stimulus, u, which has been well characterised by the following model:

$$\begin{aligned} \dot{\boldsymbol{x}} &= \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{\theta}, u; t), \quad \boldsymbol{x}(0) = \boldsymbol{x}_0 \\ \boldsymbol{y} &= \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{\theta}, u; t) \end{aligned} \tag{1}$$

where $\boldsymbol{x} \in \mathbb{R}^{n_x}$ is the state vector, $\boldsymbol{\theta} \in \mathbb{R}^{n_\theta}$ is the parameter vector, $\boldsymbol{x}_0$ is the initial value of the state vector $\boldsymbol{x}$, and $\boldsymbol{y} \in \mathbb{R}^{n_y}$ is the output vector.

When a lesser-known stimulus, u_a, is added to a cell, the signalling dynamics deviate significantly from those predicted by (1). Due to the disparity in our understanding of the roles of u_a, u, and their interplay in the signalling dynamics, little information on the signalling dynamics is available a priori. Consequently, the construction of a high-fidelity model, which faithfully simulates the signalling dynamics initiated by u_a and u, requires iterative experimentation and model refinement, which can be an arduous and lengthy process [10, 20, 21].

A more viable alternative is to approximate the dynamics induced by u_a through introducing a time-varying model where $\boldsymbol{\theta}$ in (1) changes with time so that the well-defined model (1) can be used to describe the signalling dynamics under the two stimuli [19]. To this end, the temporal profile of $\boldsymbol{\theta}$ is described as piece-wise constant functions. Under this representation, the entire temporal domain is split into several temporal subdomains, each of which has its own parameter values. Consequently, the following modified form of (1) is used to describe the signalling dynamics under two stimuli:

$$\begin{aligned} \dot{\boldsymbol{x}} &= \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{\theta}_{\sigma(t)}, u; t), \quad \boldsymbol{x}(0) = \boldsymbol{x}_0 \\ \boldsymbol{y} &= \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{\theta}_{\sigma(t)}, u; t) \\ \sigma(t) &= i \ \text{ if } t \in \mathbb{T}_i, \quad i = \{1, \ldots, n_\sigma\} \end{aligned} \tag{2}$$

where $\boldsymbol{\theta}_i \in \mathbb{R}^{n_\theta}$, where $i = 1, \ldots, n_\sigma$, is the vector of parameter values used when the current time t belongs to the temporal subdomain, $\mathbb{T}_i$, n_σ is the number of temporal subdomains, and $\sigma(t)$ is the discrete variable to denote which $\boldsymbol{\theta}_i$ is used at the time t. Under this formulation, the overall temporal domain is partitioned into n_σ subdomains, where different values of $\boldsymbol{\theta}$ are used. From here on, u_a is neglected since the use of $\boldsymbol{\theta}_{\sigma(t)}$ implies the presence of u_a.

2.2 Experimental measurements

In order to train and validate (2), $\boldsymbol{y}$ is measured experimentally under different conditions. Here, n_u different values of u with a fixed value of u_a are used. Due to the technical and economic constraints in a biology experiment, $\boldsymbol{y}$ can be measured at only a few sampling time instants, t_l, $l = 1, \ldots, N_t$, where N_t is the number of sampling instants [22]. Also, it should be noted that commonly used biochemical assays such as Western blots, flow cytometry, or microarrays typically give qualitative or semi-quantitative datasets, which measure relative but not absolute concentrations of biomolecules [23]. In other words, the measured output is defined as

$$z_i^s(t_l) = c_i \cdot \hat{y}_i(u^s; t_l) + \nu_i \tag{3}$$

where z_i^s, $i = 1, \ldots, n_y$, is the relative measured output that is corrupted with measurement noise under the input u^s, $s = 1, \ldots, n_u$, $\hat{y}_i$ is the output in absolute concentration that is not directly measurable in the experiments, c_i is the proportional constant relating z_i^s and $\hat{y}_i$, and ν_i is the measurement noise. Here, it is assumed that the mean value of ν_i can be inferred during the equipment calibration procedure [24].

Since the values of $\boldsymbol{c}$ are not usually known beforehand, an alternative quantity is computed to facilitate the comparison between the model and the experimental measurements. Specifically, fold changes in the measurements are calculated as follows [21]:

$$\bar{y}_i^s(t_l) = \frac{z_i^s(t_l) - \bar{\nu}_i}{z_i^s(t_1) - \bar{\nu}_i} = \frac{\hat{y}_i(u^s; t_l)}{\hat{y}_i(u^s; t_1)} \tag{4}$$

where $\bar{y}_i^s(t_l)$ is fold change of z_i^s at the time instant at t_l, t_1 is the first sampling instant (usually $t_1 = 0$), and $\bar{\nu}_i$ is the average measurement noise that can be obtained by performing a negative control measurement without reagents.

2.3 Problem statement

In this study, we seek to construct a time-varying model (2) by estimating the temporal dynamics of $\boldsymbol{\theta}$, and this can be achieved by addressing the following two problems:

Problem 1: Given the model (2) and the experimental measurements (3) and (4), determine the number of temporal subdomains, n_σ, as well as the temporal subdomains, $\mathbb{T}_i$, $\forall i = 1, \ldots, n_\sigma$.

Since n_σ and $\mathbb{T}_i$ are not known a priori, the experimental measurements are clustered to estimate the value of n_σ and the temporal subdomains, $\mathbb{T}_i$.

Problem 2: Given the model (2), the experimental measurements (3), and the temporal subdomains of the parameters $(\mathbb{T}_1, \ldots, \mathbb{T}_{n_\sigma})$, estimate the values of parameters, $\boldsymbol{\theta}_i$, $i = 1, \ldots, n_\sigma$, in each temporal subdomain.

For many intracellular signalling pathways, only a small subset of $\boldsymbol{\theta}$ is identifiable from the experimental measurements [22, 25]. As a result, the additional parameters introduced in the time-varying model (2), which increases the size of the parameter space by n_σ-fold, are likely to be even more unidentifiable. Hence, a sequential parameter selection methodology is implemented to identify the most important parameters in $\boldsymbol{\theta}$, and the values of these

parameters in each temporal subdomain are estimated. The resultant model then can be used to investigate the system dynamics and design the optimal experiments for future studies to advance our understanding.

3 Temporal clustering

Since the intracellular signalling dynamics are described by the time-varying model (2) with the piecewise constant $\boldsymbol{\theta}$, the value of n_σ and all the temporal subdomains, $\mathbb{T}_i$, need to be determined. In this work, they are inferred by clustering the experimental measurements into several temporal subdomains in a way that the data points contained in each subdomain exhibit similar temporal behaviours [26]. This inference assumes that the time-invariant parameters in one temporal subdomain, $\mathbb{T}_i$, result in the relatively uniform dynamics in $\boldsymbol{y}$.

For the given experimental measurements $\boldsymbol{D} \in \mathbb{R}^{N_c \times N_t}$, where $N_c = n_u \cdot n_y$,

$$\boldsymbol{D} = \begin{bmatrix} \bar{y}_1^1(t_1) & \cdots & \bar{y}_1^1(t_{N_t}) \\ \vdots & \ddots & \vdots \\ \bar{y}_1^{n_u}(t_1) & \cdots & \bar{y}_1^{n_u}(t_{N_t}) \\ \bar{y}_2^1(t_1) & \cdots & \bar{y}_2^1(t_{N_t}) \\ \vdots & \ddots & \vdots \\ \bar{y}_2^{n_u}(t_1) & \cdots & \bar{y}_2^{n_u}(t_{N_t}) \\ \vdots & \ddots & \vdots \\ \bar{y}_{n_y}^{n_u}(t_1) & \cdots & \bar{y}_{n_y}^{n_u}(t_{N_t}) \end{bmatrix} \tag{5}$$

a clustering algorithm will assign N_t column vectors of $\boldsymbol{D}$ into n_σ different temporal subdomains by minimising the distance between vectors in a subdomain and the centre of the subdomain, which is measured by the following intra-cluster error sum [26]:

$$\Lambda = \sum_{i=1}^{N_t} \sum_{k=1}^{n_\sigma} z_{ik} \| \boldsymbol{D}_i - \boldsymbol{c}_k \|_2^2 \tag{6}$$

where $\boldsymbol{D}_i$ is the i^{th} column of $\boldsymbol{D}$, z_{ik} is a binary variable indicating whether $\boldsymbol{D}_i$ is in the k^{th} subdomain, and $\boldsymbol{c}_k \in \mathbb{R}^{N_c}$ is the centre of the k^{th} cluster.

Since a value of n_σ is not known a priori, a clustering method is implemented with all possible number of subdomains $(1, \ldots, N_t)$ to find an optimal n_σ by computing and comparing the values of Λ as well as the inter-cluster error sum, Γ, which is defined as follows [26]:

$$\Gamma = \sum_{k=1}^{n_\sigma} \| \boldsymbol{c}^* - \boldsymbol{c}_k \|_2^2 \tag{7}$$

where $\boldsymbol{c}^* \in \mathbb{R}^{N_c}$ is the global cluster centre, which is defined as

$$c_j^* = \frac{1}{N_t} \sum_{i=1}^{N_t} \boldsymbol{D}_{ji} \tag{8}$$

where c_j^* is the j^{th} element of $\boldsymbol{c}^*$. When an optimal clustering configuration is achieved, the value of Λ is minimised while the value of Γ is maximised to achieve the maximum intra-cluster similarity and inter-cluster dissimilarity [26, 27]. Mathematically, this is quantified by the clustering balance, ϵ, which was proposed in [27], as follows:

$$\epsilon = 0.5\Gamma + 0.5\Lambda \tag{9}$$

where 0.5 in front of Γ and Λ is a weight coefficient, which can be adjusted based on the problem [28]. As Λ and Γ are expected to decrease and increase, respectively, with the increase in the number of subdomains, a turning point in the value of ϵ determines the optimal value of n_σ [26]. Once the value of n_σ is determined, all the $\mathbb{T}_i$ can also be determined by clustering $\boldsymbol{D}$ into n_σ temporal subdomains.

4 Parameter estimation

The aim of the parameter estimation is to quantitatively calibrate a model so that it can make an accurate and robust prediction of the system, which then can be used to analyse underlying mechanisms and design optimal experiments [4, 20]. We can formulate a parameter estimation for (2) as a least-squares problem to minimise the difference between model predictions and measurements as follows:

$$\min_{\boldsymbol{c}, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{n_\sigma}} \quad \sum_{s=1}^{n_u} \sum_{i=1}^{n_y} \sum_{l=1}^{N_t} \left(y_i(u^s; t_l) - \bar{y}_i^s(t_l) \right)^2 \tag{10a}$$

$$\text{s.t.} \quad \dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{\theta}_{\sigma(t_l)}, u^s; t_l), \quad \boldsymbol{x}(0) = \boldsymbol{x}_0 \tag{10b}$$

$$\boldsymbol{y} = \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{\theta}_{\sigma(t_l)}, u^s; t_l) \tag{10c}$$

$$\sigma(t_l) = k \quad \text{if } t_l \in \mathbb{T}_k, \quad k = \{1, \ldots, n_\sigma\} \tag{10d}$$

$$\boldsymbol{\theta}^{\text{lb}} \le \boldsymbol{\theta}_k \le \boldsymbol{\theta}^{\text{ub}} \tag{10e}$$

where $\boldsymbol{\theta}^{\text{lb}}$ and $\boldsymbol{\theta}^{\text{ub}}$ are lower and upper bounds for the values of the model parameters, respectively.

It should be noted that the parameter estimation (10) is often ill-conditioned and results in a non-unique solution [29]. This is especially problematic for calibrating biological models since biological systems are often partially observable and over-parameterised (i.e. $n_y \ll n_\theta$) [25]. As the time-dependency of the model parameters is introduced, the issue of the non-uniqueness in the parameter estimation exacerbates since the number of parameters increases by n_σ-fold. In order to handle this issue, we assume that only identifiable parameters, which is a subset of $\boldsymbol{\theta}$, vary with time while the remaining parameters are time-invariant and fixed at their nominal values. Consequently, this study carries out the parameter selection methodology before the parameter estimation to determine the identifiable parameters and estimate their values in each $\mathbb{T}_i$ by solving the least-squares problem.

4.1 Parameter selection

The objective of the parameter selection procedure is to determine the identifiable parameters that will be estimated in the parameter estimation step. In this study, two global SA techniques are implemented to determine which parameters are identifiable.

4.1.1 Sensitivity analysis: In the literature, several analytical methods have been proposed to determine the parameter identifiability, including Taylor series expansion [30], differential algebra [31], or similarity transformation [32]. But these methods require symbolic manipulation and thus only applicable to a relatively small system (for $n_\theta + n_y \le 10$) due to the computational requirements of these methods [33].

Alternatively, the parameter identifiability can be assessed by SA, which evaluates the importance of the model parameters by quantifying changes in model outputs due to changes in model parameters. A common method is the local SA method that is based on the direct differentiation of a system model with respect to its parameters. However, the evaluation of the system model as well as its derivatives with respect to its parameters depends on the values of the model parameters, which are unknown before the parameter estimation. Therefore, a result of the local SA method is local in nature and likely to be unreliable, particularly when the parameter values are largely uncertain [34, 35].

In this study, two global SA methods, Morris method [36] and Sobol' method [37], are implemented sequentially to determine the most important parameters. Even though the parameters take different values in each temporal subdomain, the model structure remains the same. Hence, the results of global SA on the time-invariant model will be valid for the time-varying one because the global SA computes the importance of model parameters over the entire parametric domain. Therefore, all the analysis in the following sections is conducted based on the time-invariant model (1).

4.1.2 Morris method: The Morris method computes the average sensitivity of a model parameter by calculating the average change in model outputs due to changes in its value. Specifically, the value of a parameter $\theta_j \in \boldsymbol{\theta}$, $j = 1, \ldots, n_\theta$, is perturbed by Δ_j to compute its effect on an output y_i, which is quantified as follows [38]:

$$d_{ij}(u^s; t_l) = \frac{y_i(\boldsymbol{x}, \theta_1, \ldots, \theta_j + \Delta_j, \ldots, \theta_{n_\theta}, u^s; t_l) - y_i(\boldsymbol{x}, \boldsymbol{\theta}, u^s; t_l)}{\Delta_j} \quad (11)$$

where $d_{ij}(u^s; t_l)$ is called the elementary effect of θ_j on y_i at a time instant t_l, $l = 1, \ldots, N_t$. By calculating N_m different d_{ij} with N_m different values of Δ_j, the average sensitivity measure of the parameter θ_j, which is denoted as s_{ij}, is computed as follows:

$$\begin{aligned} s_{ij}(u^s; t_l) &= \frac{1}{N_m} \frac{\theta_j}{y_i(\boldsymbol{x}, \boldsymbol{\theta}, u^s; t_l)} \sum_{k=1}^{N_m} \left| d_{ij}^{(k)}(u^s; t_l) \right| \\ &= \frac{1}{N_m} \frac{\theta_j}{y_i(\boldsymbol{x}, \boldsymbol{\theta}, u^s; t_l)} \\ &\quad \times \sum_{k=1}^{N_m} \left| \frac{y_i(\boldsymbol{x}, \theta_1, \ldots, \theta_j + \Delta_j^{(k)}, \ldots, \theta_{n_\theta}, u^s; t_l) - y_i(\boldsymbol{x}, \boldsymbol{\theta}, u^s; t_l)}{\Delta_j^{(k)}} \right| \end{aligned} \quad (12)$$

where $\Delta_j^{(k)}$ is the kth perturbation applied to the parameter θ_j. Here, the term $d_{ij}(u^s; t_l)$ is normalised by $\theta_j / y_i(\boldsymbol{x}, \boldsymbol{\theta}, u^s; t_l)$ to eliminate possible scaling effects [38]. And, the suggested value for N_m is $r(n_\theta + 1)$, where r is usually around six [39].

Then, the final scaled sensitivity of all the model outputs with respect to a parameter across all the time instants is defined as follows:

$$S_j = \frac{1}{n_y} \sum_{i=1}^{n_y} \left\| \begin{bmatrix} s_{ij}(u^1; t_1) & \cdots & s_{ij}(u^1; t_{N_t}) & \cdots & s_{ij}(u^{n_u}; t_{N_t}) \end{bmatrix} \right\|_2 \quad (13)$$

where $s_{ij}(u^s; t_l)$ is the average sensitivity computed under an input u_s, $s = 1, \ldots, n_u$, at a time instant t_l.

Although the Morris method is conceptually simple and easy to be implemented, it has a limited capability in capturing the non-linear output behaviour and the dependency among parameters [38]. Therefore, this study utilises the Morris method as a screening tool to reduce the number of parameters to be analysed by the Sobol' method, which overcomes the problems of the Morris method but is computationally more expensive.

4.1.3 Sobol' method: Once the Morris method screens out less important parameters from $\boldsymbol{\theta}$, the importance of the remaining parameters, which are denoted as $\bar{\boldsymbol{\theta}} \in \mathbb{R}^{n_p}$, $n_p < n_\theta$, is analysed via the Sobol' method. Different from the local SA method or the Morris method, the Sobol' method is a variance-based method. Specifically, the sensitivity of a parameter is computed by quantifying how much each parameter contributes to the output variance. A brief overview on the Sobol' SA method is presented below, and further details can be found elsewhere [37, 39, 40].

The main idea of the Sobol' method is the decomposition of the model output into summands of increasing dimensionality. Specifically, a model output y can be decomposed as follows:

$$\begin{aligned} y(\bar{\boldsymbol{\theta}}) = y_0 &+ \sum_{i=1}^{n_p} y_i(\bar{\theta}_i) + \sum_{i=1}^{n_p - 1} \sum_{j>i}^{n_p} y_{i,j}(\bar{\theta}_i, \bar{\theta}_j) \\ &+ \sum_{i=1}^{n_p - s + 1} \cdots \sum_{j > i+s-2}^{n_p} y_{i,\ldots,j}(\bar{\theta}_i, \ldots, \bar{\theta}_j) + y_{1,\ldots,n_p}(\bar{\boldsymbol{\theta}}) \end{aligned} \quad (14)$$

where s, where $3 \le s \le n_p$, is the number of parameters involved in a summand $y_{i,\ldots,j}(\bar{\theta}_i, \ldots, \bar{\theta}_j)$, and y_0 is defined as follows:

$$y_0 = \int y(\bar{\boldsymbol{\theta}}) \mathrm{d}\bar{\boldsymbol{\theta}} \quad (15)$$

Here, we assume y_0 is a constant and the integrals of every summand over any of its variables are zero, i.e.

$$\begin{aligned} &\int y_{i_1,\ldots,i_r}(\bar{\theta}_{i_1}, \ldots, \bar{\theta}_{i_r}) \mathrm{d}\bar{\theta}_k = 0 \\ &\forall k = i_1, \ldots, i_r,\ 1 \le i_1 < \cdots < i_r \le n_p \end{aligned} \quad (16)$$

in order for the decomposition of y as (14) to hold [22, 37].

Then, by assuming that y is square integrable, its variance can be expressed as follows:

$$\begin{aligned} V &= \int \left(y^2(\bar{\boldsymbol{\theta}}) - y_0^2 \right) \mathrm{d}\bar{\boldsymbol{\theta}} \\ &= \int \left(\sum_{i=1}^{n_p} y_i^2(\bar{\theta}_i) + \sum_{i=1}^{n_p - 1} \sum_{j>i}^{n_p} y_{i,j}^2(\bar{\theta}_i, \bar{\theta}_j) + \cdots + y_{1,\ldots,n_p}^2(\bar{\boldsymbol{\theta}}) \right) \mathrm{d}\bar{\boldsymbol{\theta}} \\ &= \sum_{i=1}^{n_p} V_i + \sum_{i=1}^{n_p - 1} \sum_{j>i}^{n_p} V_{i,j} + \sum_{i=1}^{n_p - s + 1} \cdots \sum_{j > i+s-2}^{n_p} V_{i,\ldots,j} + V_{1,\ldots,n_p} \end{aligned} \quad (17)$$

where V is the total variance of a model output y, and $V_{i,\ldots,j}$ is the partial variance of the output due to the parameters $\bar{\theta}_i, \ldots, \bar{\theta}_j$.

Based on the V and $V_{i,\ldots,j}$, the importance of a parameter, $\bar{\theta}_j$, can be quantified by the first order and total sensitivity indices, which are defined as follows [41]:

$$\begin{aligned} S_{S_j} &= \frac{V_j}{V} \\ S_{T_j} &= \frac{1}{V} \left(V_j + \sum_{k \ne j}^{n_p} V_{j,k} + \cdots + V_{1,\ldots,n_p} \right) = \frac{1}{V} (V - V_{\sim j}) \end{aligned} \quad (18)$$

where S_{S_j} and S_{T_j} are the first-order and total sensitivity indices, respectively, of the model parameter $\bar{\theta}_j$, V_j is the partial variance of a model output due to $\bar{\theta}_j$, and $V_{\sim j}$ is the partial variance of a model output due to join effects of the model parameters $\bar{\boldsymbol{\theta}}$ except $\bar{\theta}_j$. Here, S_{S_j} refers to the main effect of the parameter $\bar{\theta}_j$, and S_{T_j} measures the importance of a parameter $\bar{\theta}_j$ by taking into account the direct effect $\left(S_{S_j}\right)$ as well as its joint effects with other parameters. It should be noted that the difference between S_{T_j} and S_{S_j} indicates how much $\bar{\theta}_j$ is involved in interactions with other parameters in terms of changing the model output [40].

In this study, a Monte Carlo method proposed by Homma and Saltelli [41] is implemented to estimate the total sensitivity indices. First of all, two matrices ($\boldsymbol{A}$ and $\boldsymbol{B} \in \mathbb{R}^{N_{\text{int}} \times n_p}$) are generated randomly from the parameter space via a Sobol' sequence to produce parameter samples without overlapping [37, 42]. Here, N_{int} is the sample size for the Monte Carlo estimation, which is typically around a few hundreds to thousands [39]. Then, another set of matrices $\boldsymbol{C}_j \in \mathbb{R}^{N_{\text{int}} \times n_p}$, $\forall j = 1, \ldots, n_p$, can be defined for

every parameter, $\bar{\theta}_j$, by replacing the jth column of $\boldsymbol{B}$ with the jth column of $\boldsymbol{A}$. Next, the model outputs can be computed for all the sampled parameter values in the matrices $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{C}_j$. Finally, the first-order and total sensitivity indices in (18) can be approximated as follows:

$$\begin{aligned} S_{S_j}(u^s;t_l) &\simeq \frac{1/N_{\text{int}}\sum_{i=1}^{N_{\text{int}}} y(u^s,\boldsymbol{a}^{(i)};t_l)\cdot y(u^s,\boldsymbol{c}_j^{(i)};t_l) - f_0^2(u^s;t_l)}{1/N_{\text{int}}\sum_{i=1}^{N_{\text{int}}} (y(u^s,\boldsymbol{a}^{(i)};t_l)^2 - f_0^2(u^s,t_l)} \\ S_{T_j}(u^s;t_l) &\simeq 1 - \frac{1/N_{\text{int}}\sum_{i=1}^{N_{\text{int}}} y(u^s,\boldsymbol{b}^{(i)};t_l)\cdot y(u^s,\boldsymbol{c}_j^{(i)};t_l) - f_0^2(u^s;t_l)}{1/N_{\text{int}}\sum_{i=1}^{N_{\text{int}}} (y(u^s,\boldsymbol{a}^{(i)};t_l)^2 - f_0^2(u^s,t_l)} \end{aligned} \tag{19}$$

where $\boldsymbol{a}^{(i)}$, $\boldsymbol{b}^{(i)}$, and $\boldsymbol{c}_j^{(i)}$ are ith rows of $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{C}_j$, respectively, and $f_0^2(u^s,t_l)$ is defined as follows:

$$f_0^2(u^s,t_l) = \left(\frac{1}{N_{\text{int}}}\sum_{i=1}^{N_{\text{int}}} y_i(u^s,\boldsymbol{a}^{(i)};t_l)\right)^2 \tag{20}$$

Since the proposed model (2) has n_y outputs obtained under n_u values for u sampled at N_t time instants, a lumped sensitivity metric for the total sensitivity index is defined to ease the parameter selection process:

$$\text{ST}_{ij} = \frac{1}{N_c}\sum_{s=1}^{n_u}\sum_{l=1}^{N_t} S_{i,T_j}(u^s;t_l) \tag{21}$$

where ST_{ij} is the sensitivity of y_i with respect to $\bar{\theta}_j$, which will be used to select the most influential parameters, and S_{i,T_j} is the total sensitivity index, S_{T_j}, computed for an output, y_i, $i = 1,\ldots,n_y$. Similarly, a lumped sensitivity metric for the first-order sensitivity index is defined as follows:

$$\text{SS}_{ij} = \frac{1}{N_c}\sum_{s=1}^{n_u}\sum_{l=1}^{N_t} S_{i,S_j}(u^s;t_l) \tag{22}$$

where SS_{ij} is the sensitivity of y_i with respect to $\bar{\theta}_j$, which will be used to select the most influential parameters, and S_{i,S_j} is the first-order sensitivity index, S_{S_j}, computed for an output, y_i, $i = 1,\ldots,n_y$.

Based on the values of ST_{ij} and SS_{ij}, a set of identifiable parameters, $\boldsymbol{\Theta} \in \mathbb{R}^{n_{p_s}}$, where $n_{p_s} \le n_p$, can be identified from $\bar{\boldsymbol{\theta}}$. And the final parameter estimation problem (4) is reformulated as follows:

$$\min_{\boldsymbol{\Theta}_1,\ldots,\boldsymbol{\Theta}_{n_\sigma}} \quad \sum_{s=1}^{n_u}\sum_{i=1}^{n_y}\sum_{l=1}^{N_t}\left(y_i(u^s;t_l) - \bar{y}_i^s(t_l)\right)^2 \tag{23a}$$

$$\text{s.t.} \quad \dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x},\boldsymbol{\Theta}_{\sigma(t_l)},u^s;t_l), \quad \boldsymbol{x}(0) = \boldsymbol{x}_0 \tag{23b}$$

$$\boldsymbol{y} = \boldsymbol{g}(\boldsymbol{x},\boldsymbol{\Theta}_{\sigma(t_l)},u^s;t_l) \tag{23c}$$

$$\sigma(t_l) = k \quad \text{if } t_l \in \mathbb{T}_k,\ k = \{1,\ldots,n_\sigma\} \tag{23d}$$

$$\boldsymbol{\theta}^{\text{lb}} \le \boldsymbol{\Theta}_{\sigma(t_l)} \le \boldsymbol{\theta}^{\text{ub}} \tag{23e}$$

The computational time required for the SA depends on the sample size, the number of parameters, and the time for running a model. For the Morris and Sobol' methods, the number of simulations required to compute the sensitivity indices are $n_\theta \times N_m \times n_u$ and $(n_p+2)N_{\text{int}} \times n_u$, respectively, which shows that the computational cost will increase linearly. Moreover, the computational cost of solving (23) depends on the time required for running a model, the number of the model parameters, the number of the temporal subdomains, and the number of different initial guesses to solve (23).

5 Application to NF-κB signalling

In this section, we applied the proposed methodology to model the NF-κB signalling dynamics in RAW murine macrophages induced by LPS in the presence of BFA.

5.1 NF-κB signalling pathway

NF-κB is an important regulator of inflammation and immune responses in various immune cells such as macrophages [43]. Under homeostatic conditions, the activity of NF-κB is minimal because it is sequestered by isomers of IκB (inhibitors of κB) proteins such as IκB-α, -β and -ϵ [43]. In the classical NF-κB activation pathway, an external stimulus (e.g. LPS) activates IκB kinase (IKK), which leads to degradation of IκB and thus activates NF-κB [43]. Then, the derepressed NF-κB protein translocates to the nucleus and upregulates the expression of various target genes such as IκB, and pro-inflammatory cytokines such as TNFα, which propagates the inflammatory signals to adjacent cells and tissues [44, 45].

As a component in gram-negative bacteria's outer membranes, LPS is a potent activator of the NF-κB signalling pathway in macrophages through Toll-like receptor 4 (TLR4) [46]. By forming a complex with LPS, TLR4 and its accessory molecules activate NF-κB signalling through the classical activation pathway as described earlier. In contrast, BFA activates NF-κB through an alternate signalling pathway [47, 48]. Since exposure to BFA leads to the Golgi apparatus fusing with the endoplasmic reticulum (ER), normal intracellular trafficking is disrupted, which leads to the accumulation of proteins in the ER. This, in turn, initiates the ER-stress pathway and leads to the activation of NF-κB [21, 47, 49].

Although several mechanisms have been proposed to explain how NF-κB activity is induced by the ER-stress pathway, mechanistic details have not been fully elucidated yet due to the complexity of the ER-stress signalling pathway [50, 51]. Furthermore, recent studies demonstrated that interactions between the ER-stress and NF-κB signalling pathways are bidirectional, which further complicates the system analysis (see [48] and references therein). To unravel the complexity of the ER-stress signalling pathway, several computational models [50–52] have been proposed; however, they have not been validated thoroughly under various physiological conditions, whereas the NF-κB signalling pathway model has been continuously tested and improved since the early 2000s [53–55]. Furthermore, few studies have attempted to model the crosstalk between the ER-stress and NF-κB signalling pathways. Consequently, this study chose to use the time-varying model to represent the LPS-induced NF-κB signalling dynamics in the presence of BFA because the detailed model structure is still not known fully. The proposed model can be used to design future experiments that can help elucidate the underlying molecular interactions in future studies.

Motivated by the above considerations, we considered the LPS-induced NF-κB signalling model as the well-characterised model (1) while the model for the NF-κB signalling dynamics induced by LPS in the presence of BFA is considered as the unknown high-fidelity model, which would be approximated by the LPS-induced signalling model with time-varying parameters.

5.2 Dynamic model of LPS-induced NF-κB signalling

The schematic diagram for the NF-κB signalling pathway and the TNF-α production induced by LPS in the presence of BFA is shown in Fig. 1.

The starting point of the model is the LPS-induced NF-κB signalling model developed by Hoffmann *et al.* [56, 58], where the LPS-NF-κB signalling pathway model was adopted from Caldwell *et al.* [58], and a model describing the regulation of the TNF-α production by internalised LPS-TLR4 complexes was adopted from Junkin *et al.* [56]. Lee *et al.* [21] further updated the model by incorporating a new role for A20 protein as an inhibitor of LPS-induced signalling. Also, the well-known effect of a BFA addition on the collapse of Golgi apparatus was taken into account by introducing time-dependent decays in rate constants associated

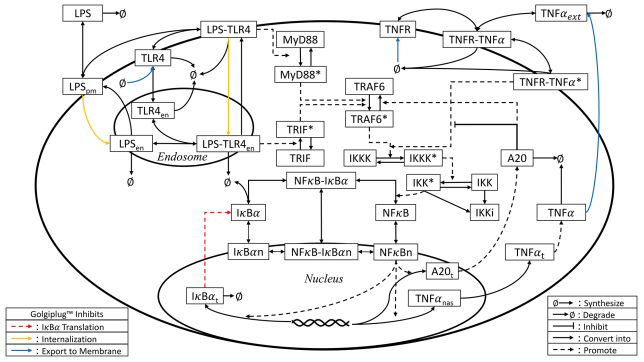

Fig. 1 *Schematic diagram for the LPS-induced NFκB signalling pathway (adapted from [21]). Due to space limitation, TRIF-dependent regulation of TNF-α production [56] and the feedback regulation between NF-κB and Iκ B-β and -ϵ are not illustrated. Also, the NF-κB activation induced by TNF-α-TNFR is not shown in details due to the limited space (see [57] for details). Coloured arrows indicate the processes affected by the addition of BFA (see text for details)*

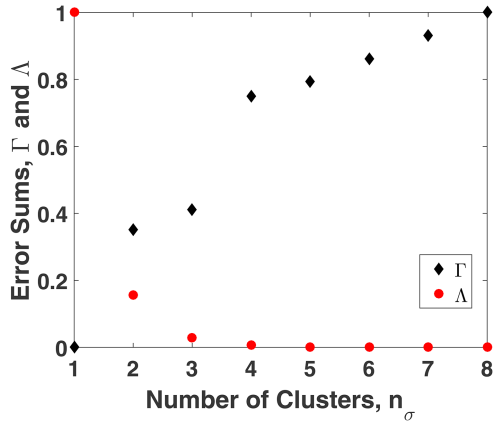

Fig. 2 *Values of the intra-cluster error sum (Λ) and inter-cluster error sum (Γ) with different number of subdomains (n_σ). Each error sum is normalised by its maximum value*

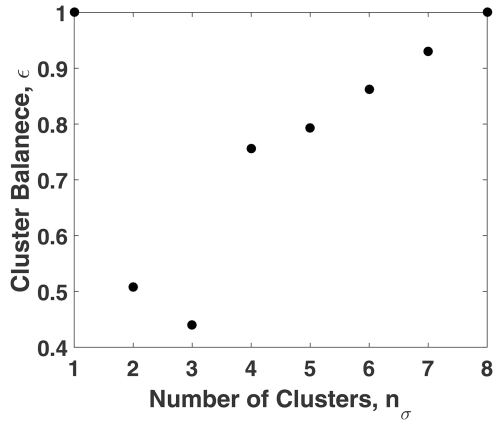

Fig. 3 *Change in the cluster balance value (ϵ) with the change in the number of subdomains (n_σ). The cluster balance value is normalised by its maximum value*

with protein secretion and protein translocation to the membrane [21]. The model outputs are the dynamics of IκBα protein and intracellular TNF α-protein (i.e. $n_y = 2$), and the updated model contains 49 states and 146 parameters (i.e. $n_x = 49$ and $n_\theta = 146$, and see [21] for the details on the model).

The datasets obtained through flow cytometry in our previous study [21] were used to perform the temporal clustering as well as the parameter estimation required in the proposed methodology. As discussed earlier, the datasets obtained through flow cytometry are relative data, which will not give the measurements in absolute concentrations, so the fold changes were computed based on (4). Therefore, the model output functions $\boldsymbol{y} = \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{\theta}_{\sigma(l)}, u; t)$ are also defined as the fold change of the two states with respect to their initial conditions as follows:

$$y_1(t_l) = \frac{\text{I}\kappa\text{B}\alpha_{\text{total}}(t_l)}{\text{I}\kappa\text{B}\alpha_{\text{total}}(t_1)}$$
$$y_2(t_l) = \frac{\text{TNF}\alpha(t_l)}{\text{TNF}\alpha(t_1)} \quad (24)$$

$y_1(t_l)$ and $y_2(t_l)$ are the predicted fold changes of IκBα and intracellular TNFα concentrations, respectively, at the time t_l, and TNFα(t_l) and IκBα$_{\text{total}}(t_l)$ are the predicted IκBα and intracellular TNF-α concentrations, respectively, by the model.

5.3 Temporal clustering

In the flow cytometry experiments described in our previous study [21], the sampling time instants were 0, 10, 20, 30, 60, 120, 240, 360 min (i.e. $N_t = 8$) after LPS and BFA were added to the cell culture. Two LPS concentrations (10 and 250 ng/ml) and one concentration of BFA (1 μg/ml) were used to obtain the experimental data sets (i.e. $n_u = 2$). Then, the temporal clustering methodology described in the preceding section was implemented to partition the measurement data sets to determine the optimal value of n_σ as well as the corresponding temporal subdomains, $\mathbb{T}_i$, $\forall i = 1, \ldots, n_\sigma$. In this work, the k-means clustering algorithm was used via *kmeans* function available in MATLAB, and multiple initial conditions were used to initialise the k-means clustering for each number of subdomains.

Since the value of N_t is eight, the maximum number of possible subdomains is eight in this work. Fig. 2 shows the changes in the intra- and inter-cluster error sums (Λ and Γ, respectively) for all possible number of subdomains. As expected, the value of Λ decreases with the number of subdomains, while the value of Γ increases. Based on these two values, the cluster balance (ϵ) defined in (9) can be computed for each number of subdomains and plotted in Fig. 3. As described earlier, a turning point in Fig. 3 is used to determine the optimal value of n_σ, which is found to be three.

Based on $n_\sigma = 3$, each temporal subdomain can be determined by clustering the experimental datasets into three temporal subdomains, which are shown in Fig. 4. Specifically, the first, second and third subdomains contain the data points spanning from 0 to 60 min, 120 min, and 240 to 360 min, respectively. Each temporal subdomain can be interpreted to represent a different phase of the NFκB signalling pathway induced by LPS in the presence of BFA. The first subdomain shows the early phase of the NF-κB signalling, where IκBα is quickly degraded while TNFα has not been synthesised. The second subdomain corresponds to the transition from the late phase of the LPS-induced NF-κB signalling pathway, where the rate of TNF-α synthesis accelerates and the IκB α is being re-synthesised, to the BFA-dominated signalling. The last subdomain can be seen as BFA-induced NF-κB dynamics, where the IκBα concentration is sustained at a low level due to the inhibition of its translation by the BFA [21].

5.4 SA result

The Morris and Sobol' sensitivity methods were implemented as described above, and all the sensitivity computation was performed in parallel in the ADA supercomputing cluster at Texas A&M University. The result of the SA via the Morris method is shown in Table 1. For each parameter, six different values were randomly sampled from its parameter domain ranging from 10 to 1000% of its nominal value, and the average sensitivity of each parameter with respect to the two outputs was computed (13). Table 1 only lists the parameters whose sensitivity measures were at least 1% of that of the most important parameter. Interestingly enough, parameters whose normalised S_j values are at least 0.1 are the ones directly involved in the TNF-α dynamics such as synthesis rate

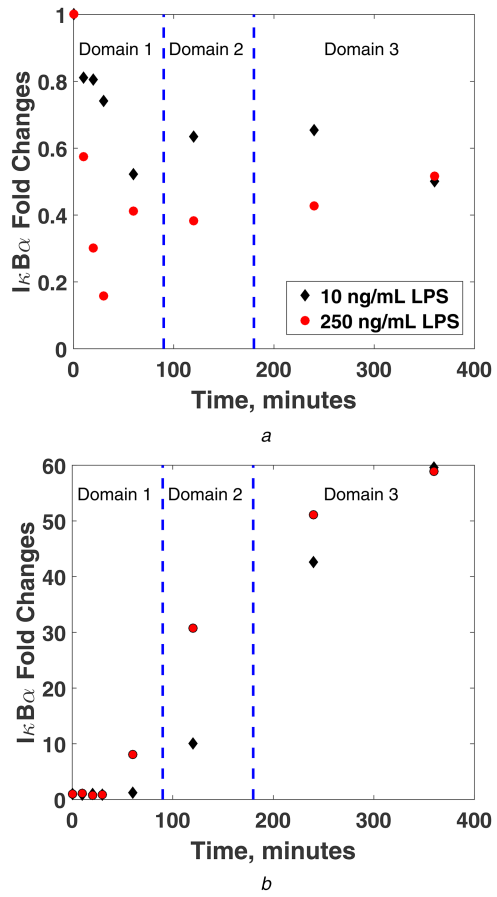

Fig. 4 *Temporal subdomains of the measured (a) IκBα and (b) TNF-α dynamics from [21]. The data points in black diamonds and red circles were measured under 10 and 250 ng/ml of LPS, respectively, in the presence of 1 μg/ml of BFA, and three different temporal subdomains are separated by blue dash lines*

Table 1 Result of Morris SA

Rank	Parameter	Normalised S_j
1	Hill coefficient for TNF-α transcription	1.00
2	constant for TRIF*-induced TNF-α production enhancement (K_{a0})	0.99
3	constant for TRIF*-induced TNF-α production enhancement (K_a)	0.29
4	TNF-α protein synthesis rate constant	0.28
5	TNF-α nascent mRNA processing rate constant	0.16
6	TNF-α protein degradation rate constant	0.12
7	maximum degradation rate constant for TNF-α transcript	0.089
8	IKK*-mediated degradation rate constant for IκBα in NF-κB-IκBα	0.081
9	IKKK*-mediated IKK activation rate constant	0.053
10	constitutive IKKK activation rate constant	0.032
11	IκBα transcript degradation rate constant	0.032
12	IκBα translation rate constant	0.025
13	IκBα degradation rate in nucleus	0.023
14	IκBα degradation rate in cytoplasm	0.023
15	EC50 constant for TNF-α transcription	0.022
16	constitutive IκBα transcription rate constant	0.021
17	hill coefficient for IκBα transcription	0.021
18	NF-κB-induced TNF-α transcription rate constant	0.015
19	constitutive IKKK deactivation rate constant	0.012
20	rate constant for IκBα and NF-κB association in nucleus	0.011
21	constitutive rate constant for IKK inactivation (IKK→IKKi)	0.011
22	constitutive rate constant for IKK activation	0.010

constants and degradation rate constants of TNF-α transcripts and proteins. On the other hand, the parameters involved in the IκBα dynamics appear to be less important. This is probably because of the intrinsic property of the NF-κB reaction network. Specifically, the existence of the feedback loop formed between NF-κB and IκB proteins, where the activation of one will be inhibited by the other [5], ensures that an abrupt change in a reaction rate relevant to IκB α is less likely to result in an abrupt change in the overall IκBα dynamics.

Out of 22 parameters selected from the Morris method, the first 11 parameters were further analysed by the subsequent SA through the Sobol' method [38]. Here, the parameters after the 11th parameters were not further analysed as their significance became negligible since the cumulative sum of the normalised S_j value does not increase >1% after the eleventh parameter [34]. Same as the Morris method, the values of these eleven parameters were varied from 10 to 1000% of their nominal values, and the sample size for the integration approximation (N_{int}) was 5000. It should be noted that two different sensitivity indices were computed for each parameter with respect to each output separately through the Sobol' method, and parameters that are important to at least one output were selected for the subsequent parameter estimation.

Based on the values of ST_{ij} and SS_{ij} computed with respect to the IκBα dynamics (21) and (22), the top four parameters were selected for the subsequent estimation since the values of both sensitivity measures were one order of magnitude larger than the remaining ones (Table 2). With respect to the TNF-α dynamics, the ST_{ij} values are relatively large while the SS_{ij} values are quite small for all the parameters. This means that the parameters are highly dependent on each other in terms of making changes in the TNF-α dynamics while one parameter has a little effect in changing the TNF-α dynamics. Therefore, only one parameter, the Hill coefficient for TNF-α transcription, was selected for the parameter estimation as it has the highest sensitivity measures. In summary, five parameters were selected to vary with time, and their values in each temporal subdomain were determined in the following parameter estimation step (Table 3).

5.5 Parameter estimation

With the results from the temporal clustering and SA, the parameter estimation problem (23) was solved to obtain the values of these parameters in each temporal subdomain (Table 3). Here, the model evaluation and the parameter estimation were performed via MATLAB built-in functions, *ode15s* and *fmincon*, and the *multistart* function available in MATLAB was used to solve the optimisation problem multiple times with different initial values.

Figs. 5 and 6 show the predicted dynamics of TNF-α and IκBα after the parameter estimation. In order to show the improvement of the prediction accuracy, the predicted dynamics after the parameter estimation were compared with the experimental measurements [21] and those predicted before the estimation. The prediction accuracy for the dynamics of the proteins was significantly improved. In particular, the model was able to track the TNF-α dynamics very accurately under both conditions (Fig. 5). Although there was some discrepancy between the model prediction and the experimental measurements for the IκBα dynamics under 10 ng/ml LPS, the overall prediction was improved.

In order to further validate the resulted model, the prediction accuracy of the resulted model was assessed with the experimental dataset, which was not used to train the model. In Fig. 7, the TNF-α and IκBα dynamics predicted by the model under 50 ng/ml LPS in the presence of BFA were plotted and compared with the corresponding experimental dataset. As shown in Fig. 7, the resultant model was able to accurately predict the TNF-α and IκBα

Table 2 Result of SA by the Sobol' method

Rank	Parameter	ST_{ij}	SS_{ij}
With respect to IκBα			
1	IKK*-mediated degradation rate constant for IκBα in NFκB-IκBα	0.49	0.30
2	IKKK*-mediated IKK activation rate constant	0.36	0.19
3	constitutive IKKK activation rate constant	0.31	0.10
4	IκBα transcript degradation rate constant	0.30	0.15
5	K_{a0}	0.075	0.02
6	TNF-α protein degradation rate constant	0.075	0.02
7	maximum degradation rate constant for TNF-α transcript	0.075	0.02
8	TNF-α nascent mRNA processing rate constant	0.075	0.02
9	TNF-α protein synthesis rate constant	0.075	0.02
10	Hill coefficient for TNF-α transcription	0.075	0.02
11	K_a	0.075	0.02
With respect to TNF-α			
1	Hill coefficient for TNF-α transcription	0.95	0.08
2	K_a	0.85	0.03
3	TNF-α nascent mRNA processing rate constant	0.79	0.00
4	IKK*-mediated degradation rate constant for IκBα in NFκB-IκBα complexes	0.78	0.00
5	K_{a0}	0.76	0.01
6	maximum degradation rate constant for TNF-α transcript	0.46	0.00
7	TNF-α protein degradation rate constant	0.43	0.01
8	Constitutive IKKK activation rate constant	0.41	0.00
9	IκBα transcript degradation rate constant	0.25	0.00
10	IKKK*-mediated IKK activation rate constant	0.22	0.00
11	TNF-α protein synthesis rate constant	0.21	0.00

Table 3 Result of the parameter estimation

Parameter	Parameter values in each temporal subdomain		
	$\mathbb{T}_1$	$\mathbb{T}_2$	$\mathbb{T}_3$
IKK*-mediated degradation rate constant for IκBα in NFκB-IκBα, /(μM·min)	2.59	0.23	0.04
IKKK*-mediated IKK activation rate constant, /(μM·min)	5200	52	4230
constitutive IKKK activation rate constant, /min	5×10^{-6}	1.3×10^{-7}	4.9×10^{-6}
IκBα transcript degradation rate constant, /min	0.33	0.18	0.12
Hill coefficient for TNF-α transcription	3.73	1.96	2.02

dynamics reasonably well even though the 50 ng/ml dataset was not used in the model calibration, which demonstrated the robustness of the calibrated model and thus validates the proposed methodology.

Lastly, the resulted model was compared with our previous model, which partially incorporated the ER-stress signalling pathway through further experimentation and literature survey. Figs. 8 and 9 compare the model performance of these two models by comparing their temporal dynamics under two different LPS concentrations. In general, both models were similar in terms of reproducing the IκBα dynamics while the model developed in this work predicted slightly more accurately in terms of root-mean-squares (RMS) of the parameter estimation: the normalised RMS of the parameter estimation for the presented model is 1.75 while the RMS value for the previous model is 2.29, which showed the improved model accuracy by implementing the proposed approach. It should be noted that the development of the previous model went

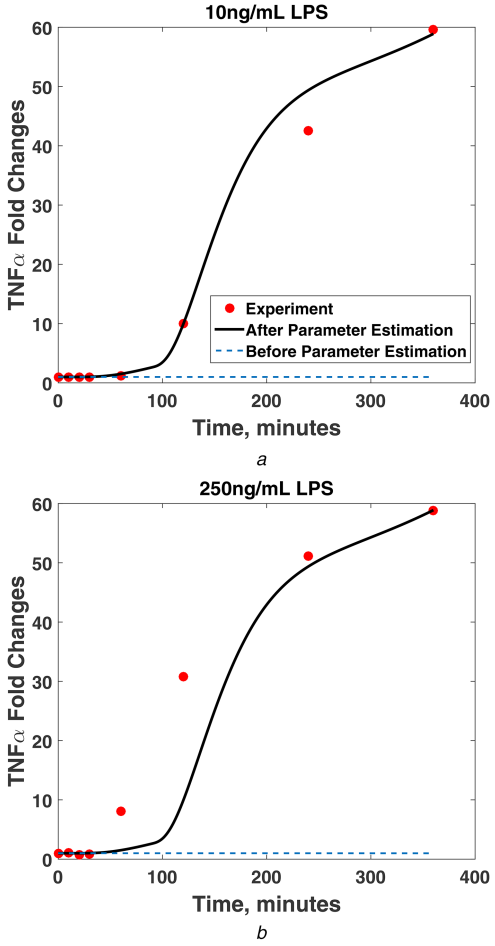

Fig. 5 *Result of parameter estimation. The predicted dynamics of TNFα before (dash line) and after (solid line) the parameter estimation were compared with the experimental observations under*
(a) 10 ng/ml and, ***(b)*** 250 ng/ml of LPS in the presence of BFA

through the iterative implementation of experiments and modelling, which can be time-consuming. However, through the proposed approach, one can get a model with a reasonable prediction accuracy in a shorter amount of time.

Since intracellular signalling pathways regulate various cellular behaviours, their dynamics and outcomes bear great importance for studying and predicting the tissue-level responses in vivo. One important factor dictating the signalling pathways is different stimuli that initiate the pathways. As discussed in the paper, there can be multiple stimuli for one signalling pathway, and the number of stimuli is likely to be higher for those with highly complex network structures. For example, it has been found that there are around 100 stimuli that can trigger the NF-κB signalling pathway. Moreover, the dynamics of one signalling pathway induced by different stimuli can be very different since these stimuli activate the intracellular signalling pathway through different mechanisms. Again, with the NF-κB signalling pathway as an example, TNF-α and LPS, two well-known stimuli of the NF-κB signalling pathway, activate the signalling pathway through two different molecules (TNF-α receptor and TLR4, respectively), resulting in the distinctive signalling dynamics.

Therefore, the comprehensive characterisation of an intracellular signalling pathway is non-trivial since each stimulus of the signalling pathway has its own distinct activation mechanism and corresponding dynamics. Under this circumstance, a model-based approach can be implemented to facilitate the study. However, this model-based approach is often feasible only for a handful of well-characterised stimuli such as TNF-α and LPS for the NF-κB signalling pathway since the underlying signalling mechanisms induced by these stimuli are relatively well studied. Motivated by the above considerations, the current study proposes a methodology to construct a data-driven mechanistic model for

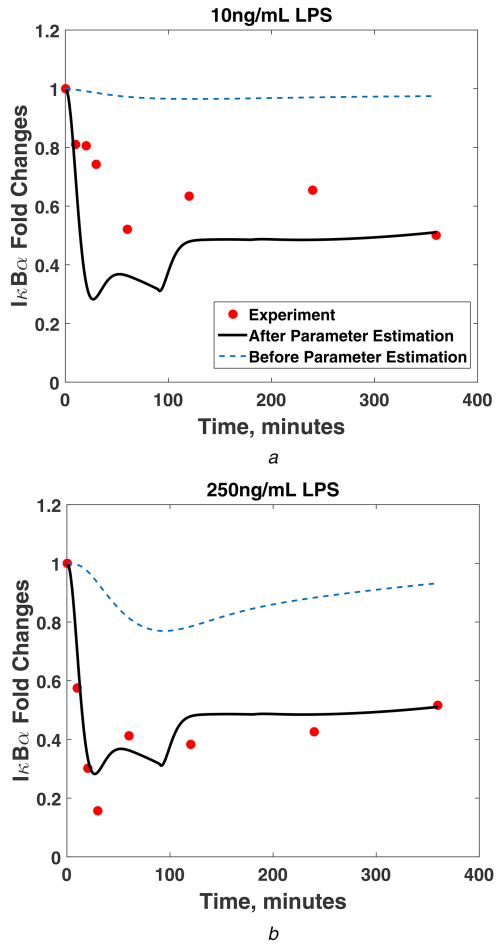

Fig. 6 *Result of parameter estimation. The predicted dynamics of IκBα before (dash line) and after (solid line) the parameter estimation were compared with the experimental observations under*
(a) 10 ng/ml and ***(b)*** 250 ng/ml of LPS in the presence of BFA

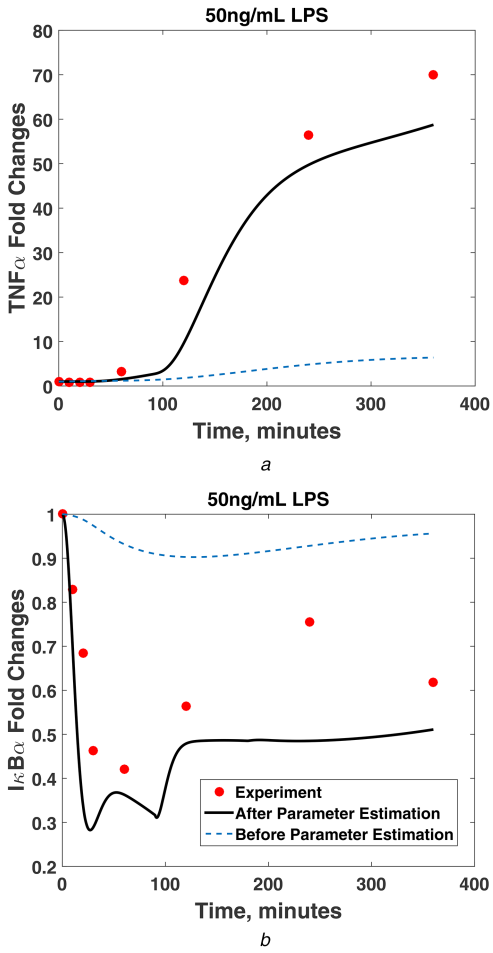

Fig. 7 *Validation of the parameter estimation results with an independent data set, which was not used in the model calibration. The predicted dynamics of*
(a) TNF*α* and, ***(b)*** IκB*α* before (dash line) and after (solid line) the parameter estimation were compared with the experimental observations under 50 ng/ml of LPS in the presence of BFA

those less-studied stimuli, whose corresponding signalling dynamics are less characterised. This is feasible since the mechanisms of the signalling pathway induced by different stimuli overlap with each other. For example, the NF-*κ*B signalling pathway network induced by TNF-*α* and LPS will converge at the IKK level, which frees the NF-*κ*B proteins from their inhibitors. Therefore, for a new or less-studied stimulus of an intracellular signalling pathway, its corresponding signalling dynamics can be described by modifying the nominal model into a time-varying one as discussed in the paper. Then, the constructed model can be used for the optimal experimental design to enhance our understanding.

It should be noted that the proposed methodology is a semi-data-driven approach, where the model construction is guided by both the available experimental data and the mechanistic model. Specifically, based on the experimental data, the temporal profiles of the model parameters are inferred to complement the model mismatch due to the use of a nominal model. As a result, the resultant model can provide relatively accurate predictions in spite of the incomplete knowledge of the underlying system. At the same time, the use of the mechanistic model allows the resultant model to be used in the detailed analysis of the underlying mechanisms, which is difficult to be performed through a data-driven model.

Additionally, the proposed time-varying model was able to robustly predict the dynamics of IκB*α* and TNF-*α* proteins, which are the core components in the NF-*κ*B signalling pathway, under the various conditions although the detailed ER-stress signalling mechanisms were not incorporated into the model. Due to the accuracy and robustness of the model, it can be used in future studies to design optimal experiments to enhance our understandings on how BFA can activate the NF-*κ*B signalling pathway.

Although the proposed methodology can be used to obtain a more accurate and predictive model as described above, it has the following limitations. First, the increase in the number of parameters to be estimated due to the temporal partitioning the parameters may exacerbate the unidentifiability issue in the model calibration. This can be a severe issue since a signalling pathway model is often over-parameterised while the available experimental measurements are limited. Second, the identified model may not reflect the true mechanisms associated with the less-studied stimulus. Specifically, the proposed method relies on the global SA to identify which parameters are time-varying, but it does not consider any biological significance while selecting the parameters. Therefore, the identified temporal profiles of the parameters may not have the biological relevance, which will constrain the process analysis based on the resultant model. It should be noted that this limitation can be mitigated by adding additional constraints into the minimisation problem (23) so that the resultant parameters retain their biological significance.

6 Conclusion

In this work, we presented a methodology for constructing a time-varying model for an intracellular signalling pathway when its reaction network is not fully known a priori. First, experimental data were clustered through the *k*-mean clustering algorithm to determine the temporal subdomains for the model parameters, where the parameters have different values in each temporal subdomain. Next, the global SA, which uses the Morris and Sobol' methods in sequence, was carried out to identify the most important parameters with respect to the model outputs. And only

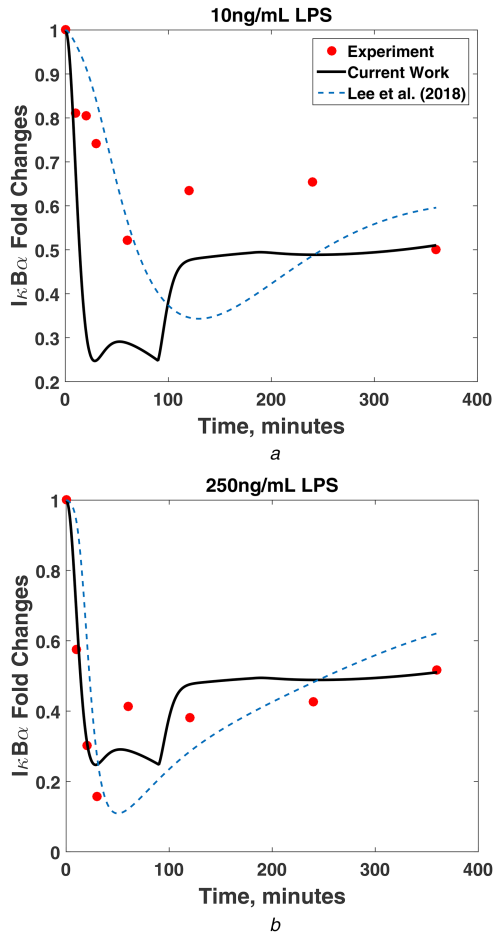

Fig. 8 *Comparison between the model developed in this study and the model developed by Lee et al. [21]. The predicted dynamics of IκBα were compared with the experimental observations under*
(a) 10 ng/ml and, ***(b)*** 250 ng/ml LPS concentration in the presence of BFA

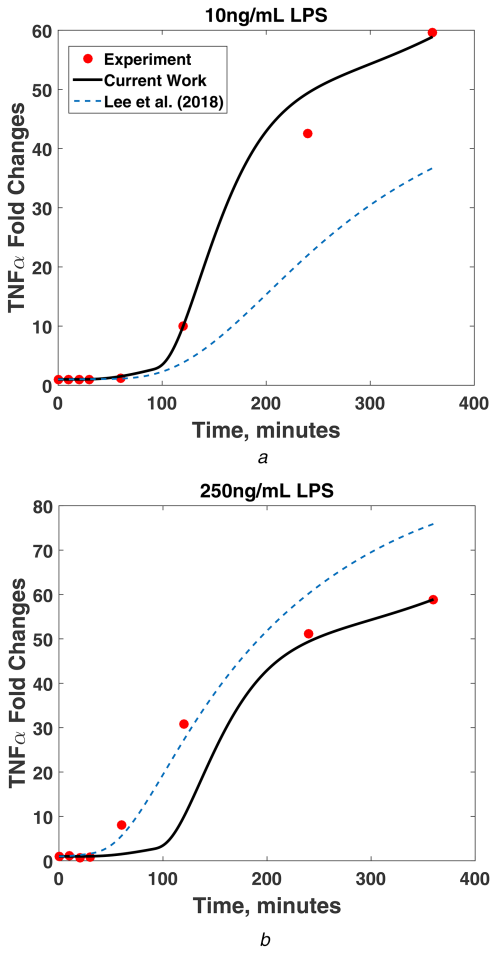

Fig. 9 *Comparison between the model developed in this study and the model developed by Lee et al. [21]. The predicted dynamics of TNFα were compared with the experimental observations under*
(a) 10 ng/ml and, ***(b)*** 250 ng/ml LPS concentration in the presence of BFA

these parameters were determined to be time-varying while the remaining parameters were fixed at their nominal values. Finally, the least-squares problem was solved to estimate the values of five parameters in each temporal subdomain to construct an accurate time-varying model. The proposed methodology was implemented to model the NF-κB signalling pathway induced by LPS in the presence of BFA to predict the dynamics of IκBα and TNF-α proteins. The prediction accuracy of the resulted model was comparable to that of a more detailed model proposed by Lee *et al.* [21], which demonstrated the performance of the proposed methodology. In summary, the proposed methodology speeds up the overall model development process without losing the prediction accuracy by avoiding the time-consuming procedure of experimentation and literature survey for developing a high-fidelity model.

7 Acknowledgments

The authors gratefully acknowledge financial support from Artie McFerrin Department of Chemical Engineering and the Texas A&M Energy Institute and the supercomputer resources from the Texas A&M University High Performance Computer Center. Also, support from the Nesbitt Chair Endowment to A.J. and flow cytometry data provided by Dr. Yufang Ding (Texas A&M University) are gratefully acknowledged.

8 References

[1] Klipp, E., Liebermeister, W.: 'Mathematical modeling of intracellular signaling pathways', *BMC Neurosci.*, 2006, **7**, p. S10
[2] Kitano, H.: 'System biology: a brief overview', *Science*, 2002, **295**, pp. 1662–1664
[3] Singh, A., Jayaraman, A., Hahn, J.: 'Modeling regulatory mechanisms in IL-6 signal transduction in hepatocytes', *Biotechnol. Bioeng.*, 2006, **95**, pp. 850–862
[4] Moya, C., Huang, Z., Cheng, P., *et al.*: 'Investigation of IL-6 and IL-10 signalling via mathematical modelling', *IET Syst. Biol.*, 2011, **5**, pp. 15–26
[5] Hoffmann, A., Levchenko, A., Scott, M.L., *et al.*: 'The IκB-NF-κB signaling module: temporal control and selective gene activation', *Science*, 2002, **298**, pp. 1241–1245
[6] Lipniacki, T., Paszek, P., Brasier, A.R., *et al.*: 'Mathematical model of NF-κB regulatory module', *J. Theor. Biol.*, 2004, **228**, pp. 195–215
[7] Swameye, I., Müller, T.G., Timmer, J., *et al.*: 'Identification of nucleocytoplasmic cycling as a remote sensor in cellular signaling by databased modeling', *Proc. Natl. Acad. Sci.*, 2003, **100**, pp. 1028–1033
[8] Schoeberl, B., Eichler-Jonsson, C., Gilles, E.D., *et al.*: 'Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors', *Nat. Biotechnol.*, 2002, **20**, pp. 370–375
[9] Pahl, H.L.: 'Activators and target genes of Rel/NF-κB transcription factors', *Oncogene*, 1999, **18**, pp. 6853–6866
[10] Gadkar, K.G., Gunawan, R., Doyle, F.J.: 'Iterative approach to model identification of biological networks', *BMC Bioinform.*, 2005, **6**, p. 155
[11] Balsa-Canto, E., Alonso, A.A., Banga, J.R.: 'An iterative identification procedure for dynamic modeling of biochemical networks', *BMC Syst. Biol.*, 2010, **4**, p. 11
[12] Rodriguez-Fernandez, M., Rehberg, M., Kremling, A., *et al.*: 'Simultaneous model discrimination and parameter estimation in dynamic models of cellular systems', *BMC Syst. Biol.*, 2013, **7**, p. 76
[13] Penas, D.R., Henriques, D., Gonzalez, P., *et al.*: 'A parallel metaheuristic for large mixed-integer dynamic optimization problems, with applications in computational biology', *PLoS ONE*, 2017, **12**, p. e0182186
[14] Maurya, M.R., Bornheimer, S.J., Venkatasubramanian, V., *et al.*: 'Mixed-integer nonlinear optimisation approach to coarse-graining biochemical networks', *IET Syst. Biol.*, 2008, **3**, pp. 24–39
[15] Verheijen, P.J.T.: 'Model selection: an overview of practices in chemical engineering', *Comput. Aided Chem. Eng.*, 2003, **16**, pp. 85–104

[16] Boukouvala, F., Misener, R., Floudas, C.A.: 'Global optimization advances in mixed-integer nonlinear programming, MINLP, and constrained derivative-free optimization, CDFO', *Eur. J. Oper. Res.*, 2016, **252**, pp. 701–727

[17] Engelhardt, B., Fröhlich, H., Kschischo, M.: 'Learning (from) the errors of a systems biology approach', *Sci. Rep.*, 2016, **6**, p. 20772

[18] Lin, Z., Beck, M.B.: 'On the identification of model structure in hydrological and environmental systems', *Water Resour. Res.*, 2007, **43**, p. W02402

[19] Tiemann, C.A., Vanlier, J., Oosterveer, M.H.*, et al.*: 'Parameter trajectory analysis to identify treatment effects of pharmacological interventions', *PLoS Comput. Biol.*, 2013, **9**, p. e1003166

[20] Balsa-Canto, E., Alonso, A.A., Banga, J.R.: 'Computational procedures for optimal experimental design in biological systems', *IET Syst. Biol.*, 2007, **2**, pp. 163–172

[21] Lee, D., Ding, Y., Jayaraman, A.*, et al.*: 'Mathematical modeling and parameter estimation of intracellular signaling pathway: application to LPS-induced NFκB activation and TNFα production in macrophages', *Processes*, 2018, **6**, p. 21

[22] Dobre, S., Bastogne, T., Profeta, C.*, et al.*: 'Limits of variance-based sensitivity analysis for non-identifiability testing in high dimensional dynamic models', *Automatica*, 2012, **48**, pp. 2740–2749

[23] Hasenauer, J., Waldherr, S., Doszczak, M.*, et al.*: 'Identification of models of heterogeneous cell populations from population snapshot data', *BMC Bioinform.*, 2011, **12**, p. 125

[24] Lillacci, G., Khammash, M.: 'A distribution-matching method for parameter estimation and model selection in computational biology', *Int. J. Robust Nonlinear Control*, 2012, **22**, pp. 1065–1081

[25] Raue, A., Kreutz, C., Maiwald, T.*, et al.*: 'Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood', *Bioinformatics*, 2009, **25**, pp. 1923–1929

[26] Narasingam, A., Siddhamshetty, P., Kwon, J.S.: 'Temporal clustering for order reduction of nonlinear parabolic PDE systems with time-dependent spatial domains: application to a hydraulic fracturing process', *AIChE J.*, 2017, **63**, pp. 3818–3831

[27] Jung, Y., Park, H., Du, D.*, et al.*: 'A decision criterion for the optimal number of clusters in hierarchical clustering', *J. Glob. Optim.*, 2003, **25**, pp. 91–111

[28] Tan, M.P., Broach, J.R., Floudas, C.A.: 'A novel clustering approach and prediction of optimal number of clusters: global optimum search with enhanced positioning', *J. Glob. Optim.*, 2007, **39**, pp. 323–346

[29] Kravaris, C., Hahn, J., Chu, Y.: 'Advances and selected recent developments in state and parameter estimation', *Comput. Chem. Eng.*, 2013, **51**, pp. 111–123

[30] Pohjanpalo, H.: 'System identifiability based on the power series expansion of the solution', *Math. Biosci.*, 1978, **41**, pp. 21–33

[31] Ljung, L., Glad, T.: 'On global identifiability for arbitrary model parametrizations', *Automatica*, 1994, **30**, pp. 265–276

[32] Vajda, S., Godfrey, K.R., Rabitz, H.: 'Similarity transformation approach to identifiability analysis of nonlinear compartmental models', *Math. Biosci.*, 1989, **93**, pp. 217–246

[33] Kiparissides, A., Koutinas, M., Kontoravdi, C.*, et al.*: 'Closing the loop' in biological systems modeling-from the in silico to the in vitro', *Automatica*, 2011, **47**, pp. 1147–1155

[34] Chu, Y., Hahn, J.: 'Parameter set selection for estimation of nonlinear dynamic systems', *AIChE J.*, 2007, **53**, pp. 2858–2870

[35] Lee, D., Singla, A., Wu, H.*, et al.*: 'An integrated numerical and experimental framework for modeling of CTB and GD1b ganglioside binding kinetics', *AIChE J.*, 2018, **64**, pp. 3882–3893

[36] Morris, M.D.: 'Factorial sampling plans for preliminary computational experiments', *Technometrics*, 1991, **33**, pp. 161–174

[37] Sobol, I.M.: 'Global sensitivity indices for nonlinear mathematical models and their Monte-Carlo estimates', *Math. Comput. Simul.*, 2001, **55**, pp. 271–280

[38] Chu, Y., Jayaraman, A., Hahn, J.: 'Parameter sensitivity analysis of IL-6 signalling pathways', *IET Syst. Biol.*, 2007, **1**, pp. 342–352

[39] Saltelli, A., Ratto, M., Tarantola, S.*, et al.*: 'Sensitivity analysis for chemical models', *Chem. Rev.*, 2005, **105**, pp. 2811–2827

[40] Saltelli, A., Gatelli, D., Campolongo, F.*, et al.*: '*Global sensitivity analysis: the primer*' (John Wiley & Sons, 2008)

[41] Homma, T., Saltelli, A.: 'Importance measures in global sensitivity analysis of nonlinear models', *Reliab. Eng. Syst. Saf.*, 1996, **52**, pp. 1–17

[42] Kontoravdi, C., Asprey, S.P., Pistikopoulos, E.N.*, et al.*: 'Application of global sensitivity analysis to determine goals for design of experiments: an example study on antibody-producing cell cultures', *Biotechnol. Prog.*, 2005, **21**, pp. 1128–1135

[43] Ghosh, G., Wang, V.Y., Huang, D.*, et al.*: 'NF-κB regulation: lessons from structures', *Immunol. Rev.*, 2012, **246**, pp. 36–58

[44] Parameswaran, N., Patial, S.: 'Tumor necrosis factor-α signaling in macrophages', *Crit. Rev. Eukaryot. Gene Expr.*, 2010, **20**, pp. 87–103

[45] Maiti, S., Dai, W., Alaniz, R.C.*, et al.*: 'Mathematical modeling of pro- and anti-inflammatory signaling in macrophages', *Processes*, 2015, **3**, pp. 1–18

[46] Kawai, T., Akira, S.: 'The role of pattern-recognition receptors in innate immunity: update on toll-like receptors', *Nat. Immunol.*, 2010, **11**, pp. 373–384

[47] Pahl, H.L., Baeuerle, P.A.: 'A novel signal transduction pathway from the endoplasmic reticulum to the nucleus is mediated by transcription factor NF-κB', *EMBO J.*, 1995, **14**, pp. 2580–2588

[48] Smith, J.A.: 'Regulation of cytokine production by the unfolded protein response; implications for infection and autoimmunity', *Front. Immunol.*, 2018, **9**, p. 422

[49] Chardin, P., McCormick, F.: 'Brefeldin A: the advantage of being uncompetitive', *Cell*, 1999, **97**, pp. 153–155

[50] Erguler, K., Pieri, M., Deltas, C.: 'A mathematical model of the unfolded protein stress response reveals the decision mechanism for recovery, adaptation and apoptosis', *BMC Syst. Biol.*, 2013, **7**, p. 16

[51] Diedrichs, D.R., Gomez, J.A., Huang, C.*, et al.*: 'A data entrained computational model for testing the regulatory logic of the vertebrate unfolded protein response', *Mol. Biol. Cell*, 2018, **29**, pp. 1502–1517

[52] Cho, H., Wu, M., Zhang, L.*, et al.*: 'Signaling dynamics of palmitate-induced ER stress responses mediated by ATF4 in HepG2 cells', *BMC Syst. Biol.*, 2013, **7**, p. 9

[53] Lipniacki, T., Kimmel, M.: 'Deterministic and stochastic models of NF-κB pathway', *Cardiovasc. Toxicol.*, 2007, **7**, pp. 215–234

[54] Cheong, R., Hoffmann, A., Levchenko, A.: 'Understanding NF-κB signaling via mathematical modeling', *Mol. Syst. Biol.*, 2008, **4**, p. 192

[55] Williams, R.A., Timmis, J., Qwarnstrom, E.E.: 'Computational models of the NF-κB signalling pathway', *Computation*, 2014, **2**, (4), pp. 131–158

[56] Junkin, M., Kaestli, A.J., Cheng, Z.*, et al.*: 'High-content quantification of single-cell immune dynamics', *Cell. Rep.*, 2016, **15**, pp. 411–422

[57] Werner, S.L., Kearns, J.D., Zadorozhnaya, V.*, et al.*: 'Encoding NF-κB temporal control in response to TNF: distinct roles for the negative regulators IκBα and A20', *Genes Dev.*, 2008, **22**, pp. 2093–2101

[58] Caldwell, A.B., Cheng, Z., Vargas, J.D.*, et al.*: 'Network dynamics determine the autocrine and paracrine signaling fucntions of TNF', *Genes Dev.*, 2014, **28**, pp. 2120–2133