

Modelling and Control of Gene Regulatory Networks for Perturbation Mitigation

Foo, M., Kim, J. & Bates, D.

Published PDF deposited in Coventry University's Repository

Original citation:

Foo, M, Kim, J & Bates, D 2019, 'Modelling and Control of Gene Regulatory Networks for Perturbation Mitigation' IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 16, no. 2, pp. 583-595

<https://dx.doi.org/10.1109/TCBB.2017.2771775>

DOI 10.1109/TCBB.2017.2771775

ISSN 1545-5963

ESSN 1557-9964

Publisher: Institute of Electrical and Electronics Engineers

This work is licensed under a Creative Commons Attribution 3.0 License. For more information, see <http://creativecommons.org/licenses/by/3.0/>

Copyright © and Moral Rights are retained by the author(s) and/ or other copyright owners. A copy can be downloaded for personal non-commercial research or study, without prior permission or charge. This item cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder(s). The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holders.

Modelling and Control of Gene Regulatory Networks for Perturbation Mitigation

Mathias Foo¹, Jongrae Kim², and Declan G. Bates¹

Abstract—Synthetic Biologists are increasingly interested in the idea of using synthetic feedback control circuits for the mitigation of perturbations to gene regulatory networks that may arise due to disease and/or environmental disturbances. Models employing Michaelis-Menten kinetics with Hill-type nonlinearities are typically used to represent the dynamics of gene regulatory networks. Here, we identify some fundamental problems with such models from the point of view of control system design, and argue that an alternative formalism, based on so-called S-System models, is more suitable. Using tools from system identification, we show how to build S-System models that capture the key dynamics of an example gene regulatory network, and design a genetic feedback controller with the objective of rejecting an external perturbation. Using a sine sweeping method, we show how the S-System model can be approximated by a linear transfer function and, based on this transfer function, we design our controller. Simulation results using the full nonlinear S-System model of the network show that the synthetic control circuit is able to mitigate the effect of external perturbations. Our study is the first to highlight the usefulness of the S-System modelling formalism for the design of synthetic control circuits for gene regulatory networks.

Index Terms—System identification, gene regulatory networks, feedback control systems, S-System model

1 INTRODUCTION

IN complex engineering networks such as transportation systems, power grids, irrigation networks, etc, the presence of external perturbations can have serious adverse effects on the functioning of the overall system. These undesirable effects include gridlock in the movement of vehicles, major power outages in residential and industrial areas, and unreliable water supply to farming areas. In view of this, the problem of developing a comprehensive theory of network control, particularly in the presence of perturbations, has recently been the subject of intensive studies that have provided many useful tools for the control of complex networks (see, e.g., [1], [2], [3], [4], [5]).

Due to advances in this area, synthetic biologists have recently began to investigate the application of the aforementioned tools to the control of biological networks and systems. Some notable examples can be found in [6], [7], [8], [9], [10], where strategies based on feedback control theory have been used to analyse the controllability, observability and stability of biological networks such that appropriate sets of control design rules can be developed.

In this paper, we focus our attention on the control of gene regulatory networks. The ability to control the dynamics of gene regulatory networks using feedback, especially

in the presence of perturbations, has many potential applications in the field of synthetic biology, where synthetic circuits can be developed to implement the proposed controllers and hence curb the effect of external perturbations due to disease or environmental changes. We investigate what types of network models are most appropriate to describe gene regulatory networks for the purposes of feedback controller design, and show how system identification techniques can be used to build such models based on available gene expression data. Using the identified models, we design a feedback controller that can be implemented genetically in order to mitigate the effect of perturbations that enter the network.

The paper is organised as follows. In Section 2, we present an example gene regulatory network for which we need to build a model for the purposes of control system design. In Section 3, we evaluate different types of possible models for gene regulatory networks from the perspective of controller design. Based on this analysis, in Section 4 we propose a system identification approach for building models of gene regulatory networks based on the so-called S-System modelling formalism. The corresponding controller design procedure for perturbation mitigation is described and closed-loop simulation results are provided in Section 5. Conclusions are given in Section 6. An early version of this work was presented in [11].

2 EXAMPLE GENE REGULATORY NETWORK

The DREAM *in silico* gene regulatory network challenge was established to serve as a benchmark to assess different proposed approaches to infer gene regulatory networks from given experimental data [12], [13], [14]. Typically, time-series data for each gene (or node) in the network are provided and the aim is to infer the underlying network,

- M. Foo and D. G. Bates are with the Warwick Integrative Synthetic Biology Centre, School of Engineering, University of Warwick, Coventry CV4 7AL, United Kingdom. E-mail: {M.Foo, D.Bates}@warwick.ac.uk.
- J. Kim is with the School of Mechanical Engineering, University of Leeds, Leeds LS2 9JT, United Kingdom. E-mail: menjkim@leeds.ac.uk.

Manuscript received 7 Dec. 2016; revised 4 July 2017; accepted 3 Nov. 2017. Date of publication 11 Jan. 2018; date of current version 29 Mar. 2019.

(Corresponding author: Mathias Foo.)

For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.

Digital Object Identifier no. 10.1109/TCBB.2017.2771775

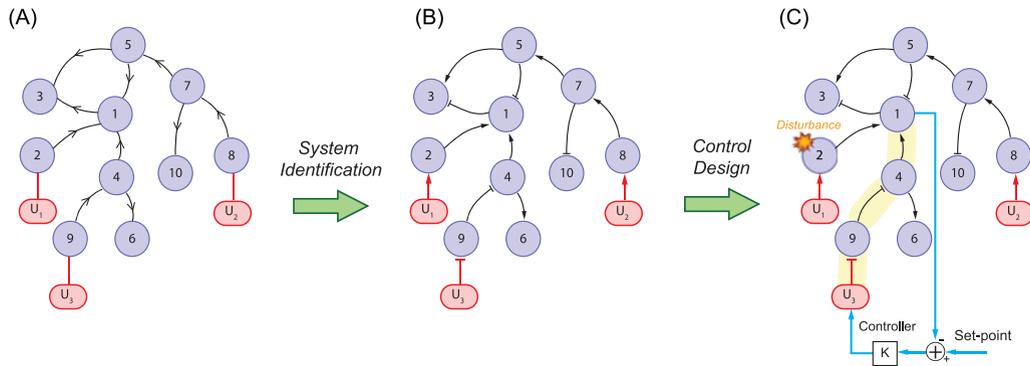


Fig. 1. (A) DREAM3 gene regulatory network. Purple circles represent genes and red rectangles represent external inputs. The arrow denotes the direction of the regulation. (B) Using system identification, the types of regulation in the network are identified. Arrow head indicates activation and Bar head indicates inhibition. (C) Proposed control design configuration for rejecting the effect of perturbation. The pathway highlighted in yellow indicates the series of regulations involved from the control action, U_3 to the output gene, N_1 .

i.e., identify interconnecting edges, the direction of information flow, etc. The provided gene regulatory networks are typically subsets of actual transcriptional networks in model organisms such as *E. coli* and *S. cerevisiae*, and hence are representative of real biological systems.

In this paper, we choose the DREAM3 Size 10 data set (hereafter we use the term DREAM3 to denote this network), which consists of mRNA temporal data on a network composed of 10 interconnecting genes that is a subset of a *S. cerevisiae* gene regulatory network. As the dataset does not include separate protein data, in the following, we make the following two assumptions: (i) the temporal evolution of the protein is similar to the mRNA and (ii) the protein is linearly translated from mRNA. Following these two assumptions, we can lump the protein dynamics into the transcription rate of the mRNA at steady state, and this results in a complete network that can be described solely using mRNA levels. In this DREAM3 data set, information regarding the interconnectivity between each gene is provided, while the regulation type (i.e., activatory or inhibitory) is unknown. The depiction of these interactions is shown in Fig. 1A. To facilitate the controller design procedure, a model describing the dynamics of the DREAM3 network is required, and in the following section, we discuss the selection of an appropriate modelling formalism for the DREAM3 gene regulatory network.

3 MODEL FORMALISMS FOR CONTROLLER DESIGN

3.1 Michaelis-Menten and Hill-Type Models

Model structures employing Michaelis-Menten and Hill-type nonlinearities are commonly used to describe the dynamics of gene regulatory networks. If the regulation type and the cooperative binding are known, the modeller can either specify

$$F_a = \frac{k_0 N_P^h}{K_M + N_P^h}, \quad (1)$$

for an activation type of regulation or

$$F_i = \frac{k_0}{K_M + N_P^h}, \quad (2)$$

for an inhibition type of regulation. In both Eqns. (1) and (2), N_P is the transcription factor, k_0 and K_M are associated with the Michaelis-Menten constants and h is the Hill coefficient.

In the context of network inference, this type of model structure can be used only if the type of regulation (activatory or inhibitory) between each gene in the network is known. In the event that the type of regulation is unknown, then this model structure is not suitable as the structure of an activation or an inhibition type of regulation is different and arbitrarily assigning them in the model building stage could thus lead to poor model accuracy.

A more fundamental problem in the context of synthetic biology is that models of this type are often not suitable for subsequent use in the design of synthetic controllers. For example, let us consider Eqn. (1) and assume that our control action (i.e., output of the controller) is given by N_P . If $N_P \gg K_M$, then $F_a \approx k_0 N_P^h / N_P^h = k_0$, which renders the control action ineffective. It is thus imperative that the value of K_M should be sufficiently large to ensure proper control, but as we will show below, obtaining a reliable estimate of K_M from time series data is often problematic.

To illustrate the problem, we consider a model of a simple gene regulatory network taken from [15], consisting of seven interconnecting genes, as shown in Fig. 2, based on a subset of an *E. coli* gene regulatory network. Assume that

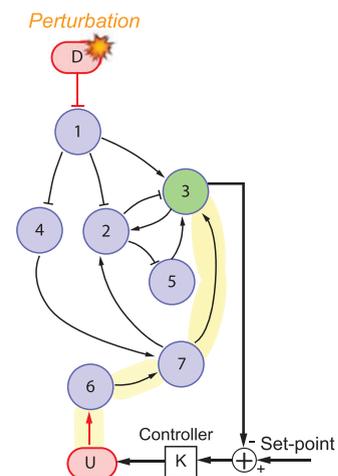


Fig. 2. Model of a gene regulatory network taken from [15] whose dynamics are represented using Michaelis-Menten kinetics and Hill-type nonlinearities. For this illustration, the controller, K is a simple proportional-integral (PI) controller with the controller gains, $K_p = 0.01$ and $K_I = 0.02$. The pathway highlighted in yellow indicates the series of regulations involved from the control action, U to the output gene, N_3 .

TABLE 1
Parameters for the Network Model Shown in Fig. 2 Using Michaelis-Menten with the Hill-Type Nonlinearities Model Structure

Gene	Parameter Values
N_1	$k_{0,1} = 0.0362, K_{M,1} = 0.1259, \gamma_1 = -0.4060,$
N_2	$k_{0,2} = 1.0106, K_{M,2} = 1.7937, k_{0,3} = 0.3550,$ $K_{M,3} = 1.2069, k_{0,4} = 0.7472, K_{M,4} = 1.2858,$ $\gamma_2 = -2.1362$
N_3	$k_{0,5} = 2.4007, K_{M,5} = 0.8218, k_{0,6} = 0.8511,$ $K_{M,6} = 1.7099, k_{0,7} = 2.8247, K_{M,7} = 1.6656,$ $k_{0,8} = 0.6081, K_{M,8} = 0.0202, \gamma_3 = -3.8740,$
N_4	$k_{0,9} = 0.0903, K_{M,9} = 0.0699, \gamma_4 = -0.7256$
N_5	$k_{0,10} = 0.5264, K_{M,10} = 0.9600, \gamma_5 = -0.7466$
N_6	$k_{0,11} = 0.6541, K_{M,11} = 1.0891, \gamma_6 = -0.4525$
N_7	$k_{0,12} = 0.0090, K_{M,12} = 0.5191, k_{0,13} = 1.1236$ $K_{M,13} = 0.4986, \gamma_{33} = -0.9473$

an external perturbation enters the network through gene 1, its effect on gene 3 is measured, and fed back to a controller that regulates gene 6 through the input U . Using the standard modelling framework employing Michaelis-Menten kinetics and Hill-type nonlinearities, the associated Ordinary Differential Equations (ODEs) describing Fig. 2 are given as follows:

$$\begin{aligned}
 \frac{dN_1}{dt} &= \frac{k_{0,1}}{(K_{M,1} + D^h)} + \gamma_1 N_1 \\
 \frac{dN_2}{dt} &= \frac{k_{0,2}}{(K_{M,2} + N_1^h)} + \frac{k_{0,3} N_3^h}{(K_{M,3} + N_3^h)} + \frac{k_{0,4} N_7^h}{(K_{M,4} + N_7^h)} + \gamma_2 N_2 \\
 \frac{dN_3}{dt} &= \frac{k_{0,5} N_1^h}{(K_{M,5} + N_1^h)} + \frac{k_{0,6}}{K_{M,6} + N_2^h} + \frac{k_{0,7} N_5^h}{(K_{M,7} + N_5^h)} \\
 &\quad + \frac{k_{0,8} N_7^h}{(K_{M,8} + N_7^h)} + \gamma_3 N_3 \\
 \frac{dN_4}{dt} &= \frac{k_{0,9}}{(K_{M,9} + N_1^h)} + \gamma_4 N_4 \\
 \frac{dN_5}{dt} &= \frac{k_{0,10}}{(K_{M,10} + N_2^h)} + \gamma_5 N_5 \\
 \frac{dN_6}{dt} &= \frac{k_{0,11} U^h}{(K_{M,11} + U^h)} + \gamma_6 N_6 \\
 \frac{dN_7}{dt} &= \frac{k_{0,12} N_4^h}{K_{M,12} + N_4^h} + \frac{k_{0,13} N_6^h}{K_{M,13} + N_6^h} + \gamma_7 N_7,
 \end{aligned} \tag{3}$$

where $k_{0,j}$, $K_{M,j}$ with $j = 1, 2 \dots$ and h are the parameters associated with the Michaelis-Menten coefficients and Hill-type nonlinearities, and γ is associated with the degradation term. Without loss of generality, for the purposes of illustration, we choose $h = 1$. The rest of the parameters describing Eqn. (3) are shown in Table 1. These parameters are estimated from available experimental data in [15], where one data set is used for parameter estimation and an independent data set is used for model validation. The parameters are estimated using the prediction error method with quadratic criterion, i.e.,

$$\hat{\Theta} = \arg \min_{\Theta} \frac{1}{L} \sum_{i=1}^{T-1} \sum_{t=1}^L [N_i(t) - \hat{N}_i(t, \Theta)]^2, \tag{4}$$

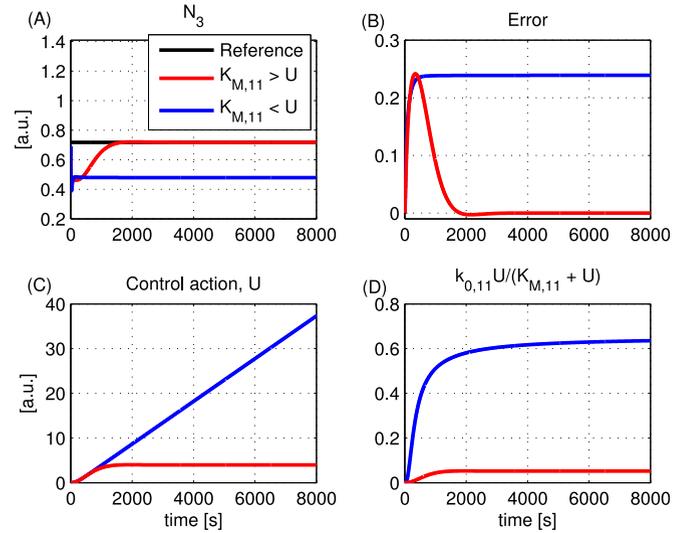


Fig. 3. Feedback control response when perturbation enters the gene regulatory network shown in Fig. 2. (A) Comparison with the output and reference values. (B) Error signal between the reference and output values. (C) Control action, U . (D) The time series of $k_{0,11}U/(K_{M,11} + U)$.

where T is the number of genes, L is the length of the data, $\Theta = \{k_{0,j}, K_{M,j}, \gamma_j\}$ with j denotes the appropriate index describing the parameters in Eqn. (3). N_i and \hat{N}_i represent the real experimental data and simulated data from Eqn. (3) respectively. Eqn. (4) is solved using MATLAB function *fminsearch*, which uses the Nelder-Mead simplex algorithm. For the controller, we choose a standard proportional-integral (PI) controller with the proportional gain, $K_P = 0.01$ and the integral gain $K_I = 0.02$, where these parameters can be selected using standard rules, such as the Ziegler-Nichols tuning rules (see, e.g., [16]).

In our simulation, shown by the solid blue line in Fig. 3, when the perturbation enters the network at time 0s it causes the expression level of N_3 to drop from its intended reference value of 0.718 (Fig. 3A). Upon sensing this drop in the expression level, the controller asserts appropriate control action, U (Fig. 3C) in its attempt to bring the expression level of N_3 back to 0.718. However, as shown in Fig. 3A, a full recovery of the output to its intended reference value is not achievable. This is because in the controller's attempt to perform the needed recovery, the exerted control action U becomes larger than $K_{M,11}$, thus the term $k_{0,11}U/(K_{M,11} + U) \approx k_{0,11} = 0.6541$, which is shown in Fig. 3D. This implies no appropriate control action can be given to the network to counter the effect of the perturbation, resulting in a large error between the output and reference value (Fig. 3B). In reality, however, this may not necessarily be the case - the apparent limitation is due to the estimated value of $K_{M,11}$ from the available experimental data. If the value of $K_{M,11}$ is sufficiently larger than U , the 'saturation' issue is avoided. In addition, a closer look at the series of regulation along the pathway highlighted in yellow shown in Fig. 2 indicates that the values of $K_{M,8}$ and $K_{M,13}$ also need to be sufficiently large in order to achieve a proper control action and recover the levels of N_3 .

The problems identified above are due to the values of $K_{M,11}$, $K_{M,8}$ and $K_{M,13}$ that are estimated from the available experimental data. These estimated values are relatively small when compared to the necessary control action,

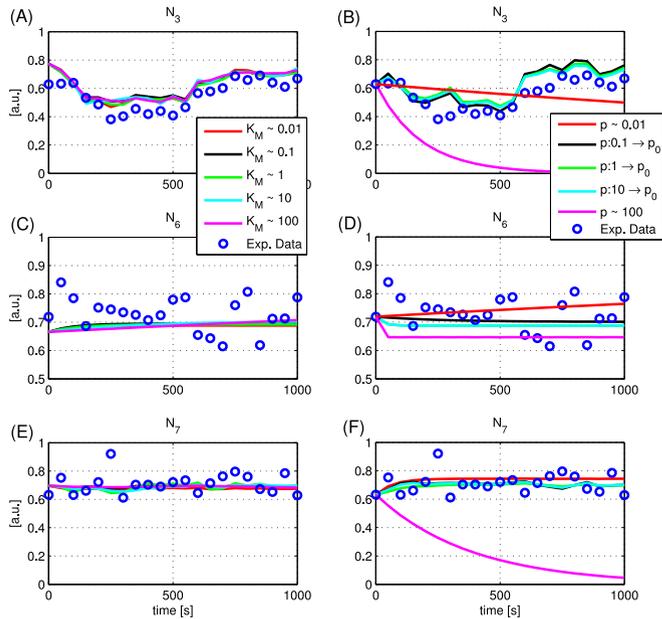


Fig. 4. Comparison of model and experimental data for different sets of estimated parameter given different initial values for optimisation. The initial values used for optimisation are 0.01, 0.1, 1, 10, and 100. Only genes in the highlighted pathway in Fig. 2 are shown. The experimental data shown here is an independent data set that is not used for parameter estimation. Left panel: Subfigures (A), (C), and (E) show the plots using Michaelis-Menten with Hill-type nonlinearities model structure for genes 3, 6, and 7, respectively. Here, the estimated values of K_M are close to the initial set of parameters used for optimisation. Right panel: Subfigures (B), (D), and (F) show the plots using S-System model structure for genes 3, 6, and 7, respectively. The notation p_0 denotes the parameter set obtained when initial value of 1 is used for the optimisation (shown in Table 1). The notation $p : 0.1, 1, 10 \rightarrow p_0$ indicates the estimated parameters using initial values of 0.1, 1 and 10 are similar to p_0 .

leading to saturated responses and large errors. Thus, a natural question arises as to whether or not these values (shown in Table 1) represent reliable estimates of these parameters. For the network shown in Fig. 2, the estimated values of $K_{M,11}$, $K_{M,8}$ and $K_{M,13}$ shown in Table 1 are the result of using 1 as the initial values for the parameters in the optimisation problem defined in Eqn (4). If a different set of initial values is used for the optimisation, do we obtain similar parameter values to those shown in Table 1 particularly for $K_{M,11}$, $K_{M,8}$ and $K_{M,13}$? To investigate this, we repeated the parameter estimation using 0.01, 0.1, 10 and 100 as initial values for the optimisation, and the results are shown in Figs. 4A, C and E.

The plots show that the estimated parameter values are very different to the ones shown in Table 1. Using terminology from the field of system identification, there is no *consistent* estimate of the model parameters, as given different initial values for the optimisation, the optimiser can find different sets of parameters (see Table 2) that are equally well able to reproduce the experimental data, as shown in Figs. 4A, C and E.

From Table 2, we note that there is one set of parameters that includes large values of $K_{M,11}$, $K_{M,8}$ and $K_{M,13}$. Using these larger values of $K_{M,11} = 120.4219$, $K_{M,8} = 145.0575$ and $K_{M,13} = 99.4842$, we repeat the simulation of the feedback controller shown in Fig. 2. As shown by the solid red line in Fig. 3A, the same controller is now able to exert a proper control action to mitigate the effect of the perturbation, as the

TABLE 2
Estimated Parameters Given Different Initial Values for Optimisation as Shown in Fig. 4A, C, and E

Initial Value	Gene	Parameter Values
0.01	N_3	$k_{0,5} = 1.6657, K_{M,5} = 0.7744, k_{0,6} = 1.2067, K_{M,6} = 2.1379, k_{0,7} = 1.0197, K_{M,7} = 0.4737, k_{0,8} = 0.8413, K_{M,8} = 1.4508, \gamma_3 = -2.9516,$
	N_6	$k_{0,11} = 0.3914, K_{M,11} = 0.0118, \gamma_6 = -0.5621$
	N_7	$k_{0,12} = 0.0093, K_{M,12} = 0.5206, k_{0,13} = 0.8078, K_{M,13} = 0.0540, \gamma_{33} = -1.1136$
0.1	N_3	$k_{0,5} = 1.8356, K_{M,5} = 0.7843, k_{0,6} = 1.1053, K_{M,6} = 1.5175, k_{0,7} = 1.7524, K_{M,7} = 2.1010, k_{0,8} = 0.8168, K_{M,8} = 0.1019, \gamma_3 = -3.5506,$
	N_6	$k_{0,11} = 0.4247, K_{M,11} = 0.1192, \gamma_6 = -0.5462$
	N_7	$k_{0,12} = 0.0091, K_{M,12} = 0.5330, k_{0,13} = 0.8627, K_{M,13} = 0.1068, \gamma_{33} = -1.0819$
1	N_3	$k_{0,5} = 2.4007, K_{M,5} = 0.8218, k_{0,6} = 0.8511, K_{M,6} = 1.7099, k_{0,7} = 2.8247, K_{M,7} = 1.6656, k_{0,8} = 0.6081, K_{M,8} = 0.0202, \gamma_3 = -3.8740,$
	N_6	$k_{0,11} = 0.6541, K_{M,11} = 1.0891, \gamma_6 = -0.4525$
	N_7	$k_{0,12} = 0.0090, K_{M,12} = 0.5191, k_{0,13} = 1.1236, K_{M,13} = 0.4986, \gamma_{33} = -0.9473$
10	N_3	$k_{0,5} = 2.5208, K_{M,5} = 0.9741, k_{0,6} = 1.7396, K_{M,6} = 0.7365, k_{0,7} = 1.7937, K_{M,7} = 2.5356, k_{0,8} = 0.1980, K_{M,8} = 15.2691, \gamma_3 = -4.0848,$
	N_6	$k_{0,11} = 1.0025, K_{M,11} = 9.0799, \gamma_6 = -0.1412$
	N_7	$k_{0,12} = 0.0049, K_{M,12} = 0.6385, k_{0,13} = 1.5460, K_{M,13} = 10.0345, \gamma_{33} = -0.1460$
100	N_3	$k_{0,5} = 1.0820, K_{M,5} = 0.7472, k_{0,6} = 1.4625, K_{M,6} = 1.4727, k_{0,7} = 0.2059, K_{M,7} = 1.5799, k_{0,8} = 0.8413, K_{M,8} = 145.0575, \gamma_3 = -1.9053,$
	N_6	$k_{0,11} = 1.0059, K_{M,11} = 120.4219, \gamma_6 = -0.0090$
	N_7	$k_{0,12} = 0.0104, K_{M,12} = 0.6691, k_{0,13} = 1.5496, K_{M,13} = 99.4842, \gamma_{33} = -0.0211$

value of $K_{M,11}$ is now larger than the control action, U (Fig. 3C) and no issues with saturation are observed (Fig. 3D).

The results shown here suggest that for this typical experimental data set and network structure, the estimated values of the model parameters, in particular $K_{M,11}$, $K_{M,8}$ and $K_{M,13}$, are not consistent. This clearly poses a significant problem when designing a controller to mitigate the effects of perturbations on this network, since different estimated values of $K_{M,11}$, $K_{M,8}$ and $K_{M,13}$ lead to very different closed-loop behaviour of the control system. In light of this, coupled with the previously mentioned need for *a priori* knowledge of regulation type to use the Michaelis-Menten with Hill-type nonlinearities model structure, an alternative modelling formalism is clearly required in order to allow for the rational design of feedback controllers. The alternate model formalism needs to have a general structure that can accommodate both activatory and inhibitory regulations, and more importantly, the estimated model parameters from experimental data should be consistent, so that it can be reliably used for controller design.

3.2 S-System Models

The so-called S-System modelling formalism has been proposed as an alternative approach to describe the dynamics of gene regulatory networks. The S-System modelling framework was originally developed from the field of biochemical system theory (see, e.g., [17], [18]), and when it has been

TABLE 3
Parameters for the Network Model Shown in Fig. 2
Using the S-System Model Structure

Gene	Parameter Values
N_1	$b_1 = -0.3789, c_1 = -0.2488, d_1 = 0.2724,$
N_2	$a_2 = 0.4729, p_{2,1} = -0.0490, p_{2,2} = 0.0015,$ $p_{2,3} = 0.0360, b_2 = -1.2252$
N_3	$a_3 = 5.6808, p_{3,1} = 0.2232, p_{3,2} = -0.0568,$ $p_{3,3} = 0.0210, p_{3,4} = 0.3906, b_3 = -6.4230,$
N_4	$a_4 = 0.0695, p_{4,1} = -0.8931, b_4 = -0.6381$
N_5	$a_5 = 0.2552, p_{5,1} = -0.1822, b_5 = -0.5814$
N_6	$b_6 = -1.8949, c_6 = 1.3030$
N_7	$a_7 = 0.5916, p_{7,1} = 0.0001, p_{7,2} = 0.4048$ $b_7 = -0.7338$

used to describe the dynamics of gene regulation (see, e.g., [19], [20]), it has been shown to be as accurate as Michaelis-Menten with Hill-type nonlinearity models (see [21]). In particular, the authors in [21] rigorously analysed the ‘validity’ range of the concentrations produced by both S-System and Michaelis-Menten models to determine which model differs most from the ‘true’ concentration obtained via experiment. It was found that, not only were S-System models as accurate as Michaelis-Menten type models within the same concentration range, but the S-System models were more accurate over a wider range of concentrations. Based on this and other analyses, the authors suggested that the S-System model formalism better represents the actual biochemical system.

The S-System models we consider in this work have the following form

$$\frac{dN_i}{dt} = a_i \prod_{j=1}^{M_1} N_j^{p_{i,j}} + b_i \prod_{j=1}^{M_2} N_j^{q_{i,j}} + \sum_{j=1}^{M_3} c_{i,j} U_j, \quad (5)$$

where i denotes the number of biochemical component, $a_i > 0, b_i < 0$ and $c_{i,j} \in (-\infty, +\infty)$ are constants, N_i represents the biochemical component, M_1 and M_2 are the total number of components involved in the interaction, U_j is the external input and M_3 is the number of input. The power exponent terms, $p_{i,j}$ and $q_{i,j}$ are associated with the production and degradation terms respectively. For simplicity, we assume $q_{i,j} = 1$ throughout this paper, so that a positive value for the parameter $p_{i,j}$ represents activation while a negative value represents inhibition. Thus, the S-System model has a general structure that can accommodate either an activation or inhibition type of regulation via the sign of $p_{i,j}$, and no prior knowledge of the type of regulation at each node in the network is required in the model building process.

The S-System model describing the gene regulatory network shown in Fig. 2 is given as follows

$$\begin{aligned} \frac{dN_1}{dt} &= b_1 N_1 + c_1 D + d_1 \\ \frac{dN_2}{dt} &= a_2 N_1^{p_{2,1}} N_3^{p_{2,2}} N_7^{p_{2,3}} + b_2 N_2 \\ \frac{dN_3}{dt} &= a_3 N_1^{p_{3,1}} N_2^{p_{3,2}} N_5^{p_{3,3}} N_7^{p_{3,4}} + b_3 N_3 \\ \frac{dN_4}{dt} &= a_4 N_1^{p_{4,1}} + b_4 N_4 \end{aligned}$$

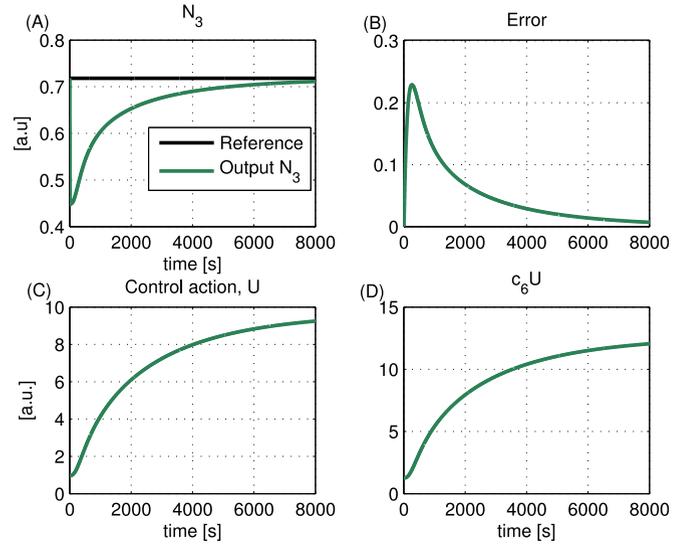


Fig. 5. Feedback control response when a perturbation enters the gene regulatory network that is modelled using the S-System formalism. (A) Output and reference values. (B) Error signal between the reference and output values. (C) Control action, U . (D) The time series of $c_6 U$.

$$\begin{aligned} \frac{dN_5}{dt} &= a_5 N_2^{p_{5,1}} + b_5 N_5 \\ \frac{dN_6}{dt} &= b_6 N_6 + c_6 U \\ \frac{dN_7}{dt} &= a_7 N_4^{p_{7,1}} N_6^{p_{7,2}} + b_7 N_7, \end{aligned} \quad (6)$$

Note that for dN_1/dt , a constant value denoted by d_1 is added to the model to ensure the overall mRNA level stays positive since D is negatively correlated with N_1 and b_1 is negative due to the degradation term. Like before, we use one set of experimental data for parameter estimation and an independent set of data for model validation. The parameters are estimated using the prediction error method with quadratic criterion (Eqn. (4)) with $\Theta = \{a_i, b_i, c_i, d_1, p_{i,j}\}$ where i and j denote the appropriate indices in Eqn. (6). The estimated parameters, using 1 as the initial value for all parameters in the optimisation, are given in Table 3.

We repeat the feedback control design using the same configuration shown in Fig. 2. The feedback control response when a perturbation enters the gene regulatory network is shown in Fig. 5. When the S-System model is used, the controller is able to produce an appropriate control action to attenuate the effect of the disturbance. There is no saturation issue observed, unlike in the scenario where the Michaelis-Menten with Hill-type nonlinearities model structure is used.

We proceed further to check whether the estimated parameters for the S-System model are consistent or not. As before, we choose the initial parameter values for the optimisation to be 0.01, 0.1, 10 and 100. The resulting estimated parameters are given in Table 4. The results shown in Figs. 4B, D and F indicate that, using this model structure, the estimated parameters are now consistent. Denoting p_0 as the estimated parameter set obtained when 1 is used as the initial value for optimisation, we observe that when initial values of 0.1 and 10 are used, the estimated parameters are close to p_0 (see Table 4). When initial values of 0.01 and

TABLE 4
Estimated Parameters Given Different Initial Values
for the Optimisation as Shown in Figs. 4B, D, and F

Initial Value	Gene	Parameter Values
0.01	N_3	$a_3 = 0.0008, p_{3,1} = 0.0137, p_{3,2} = -0.0098,$ $p_{3,3} = 0.0101, p_{3,4} = 0.0111, b_3 = -0.0129,$
	N_6	$b_6 = -0.0102, c_6 = 0.0099$
	N_7	$a_7 = 0.5620, p_{7,1} = 0.0002, p_{7,2} = 0.0107$ $b_7 = -0.7522$
0.1	N_3	$a_3 = 2.0744, p_{3,1} = 0.1772, p_{3,2} = -0.0950,$ $p_{3,3} = 0.0278, p_{3,4} = 0.3888, b_3 = -2.5312,$
	N_6	$b_6 = -1.8702, c_6 = 1.3104$
	N_7	$a_7 = 0.2737, p_{7,1} = 0.0003, p_{7,2} = 0.2681$ $b_7 = -0.3564$
1	N_3	$a_3 = 5.6808, p_{3,1} = 0.2232, p_{3,2} = -0.0568,$ $p_{3,3} = 0.0210, p_{3,4} = 0.3906, b_3 = -6.4230,$
	N_6	$b_6 = -1.8949, c_6 = 1.3030$
	N_7	$a_7 = 0.5916, p_{7,1} = 0.0001, p_{7,2} = 0.4048$ $b_7 = -0.7338$
10	N_3	$a_3 = 5.1614, p_{3,1} = 0.1791, p_{3,2} = -0.0745,$ $p_{3,3} = 0.0259, p_{3,4} = 0.3456, b_3 = -6.3618,$
	N_6	$b_6 = -1.9653, c_6 = 1.3513$
	N_7	$a_7 = 0.5153, p_{7,1} = 0.0003, p_{7,2} = 0.3256$ $b_7 = -0.6566$
100	N_3	$a_3 = 8.1321, p_{3,1} = 0.2214, p_{3,2} = -0.0610,$ $p_{3,3} = 0.0249, p_{3,4} = 117.9354, b_3 = -0.2774,$
	N_6	$b_6 = -119.4802, c_6 = 82.1233$
	N_7	$a_7 = 0.8009, p_{7,1} = 0.0004, p_{7,2} = 106.1156$ $b_7 = -0.1311$

100 are used, the estimated parameters are not close to p_0 , but in this case the model responses do not reproduce the experimental data.

Taken altogether, these results suggest that we are able to obtain consistent estimates of the model parameters from experimental data when using the S-System model structure, making this modelling formalism much more suitable for use in the design of feedback controllers for perturbation mitigation.

4 IDENTIFICATION OF AN DREAM3 NETWORK USING S-SYSTEM MODEL

In the previous section, we have illustrated why the S-System model formalism is a more appropriate way to model gene regulatory networks for the purposes of control system design. We now proceed to use the S-System model structure to identify, model, and design a biologically implementable perturbation mitigation controller for the DREAM3 network. Fig. 1A shows the interconnection between the genes in the DREAM3 gene regulatory network. In contrast to the network shown in Fig. 2, here no information is provided regarding the type of regulation between the interconnecting genes, and therefore we use system identification techniques (see, e.g., [22]) to infer the type of regulation within the network. Note that, since no information regarding the type of regulation between the interconnecting genes is available, the Michaelis-Menten with Hill-type nonlinearities model structure cannot be used in this case.

System identification techniques have been used to build models of gene regulatory networks in several previous studies, including [23], [24], [25], where linear *black box* network models were considered and the directions and the types of regulation were identified based on available data on gene expression profiles. In this paper, we consider a nonlinear *grey box* S-System model, given that we have prior knowledge about the network interconnections, and focus our attention on the identification of the type of regulation between the interconnecting genes. We use one data set for parameter estimation and another independent data set for model validation. Note that both the estimation and validation data sets used are the provided temporal profiles from the DREAM3 gene regulatory network challenge. The S-System model for the DREAM3 gene regulatory network following Fig. 1A is given by

$$\begin{aligned}
 \frac{dN_1}{dt} &= a_1 N_2^{p_{1,1}} N_4^{p_{1,2}} N_5^{p_{1,3}} + b_1 N_1 \\
 \frac{dN_2}{dt} &= b_2 N_2 + c_2 U_1 \\
 \frac{dN_3}{dt} &= a_3 N_1^{p_{3,1}} N_5^{p_{3,2}} + b_3 N_3 \\
 \frac{dN_4}{dt} &= a_4 N_9^{p_{4,1}} + b_4 N_4 \\
 \frac{dN_5}{dt} &= a_5 N_7^{p_{5,1}} + b_5 N_5 \\
 \frac{dN_6}{dt} &= a_6 N_4^{p_{6,1}} + b_6 N_6 \\
 \frac{dN_7}{dt} &= a_7 N_8^{p_{7,1}} + b_7 N_7 \\
 \frac{dN_8}{dt} &= b_8 N_8 + c_8 U_2 \\
 \frac{dN_9}{dt} &= b_9 N_9 + c_9 U_3 + d_9 \\
 \frac{dN_{10}}{dt} &= a_{10} N_7^{p_{10,1}} + b_{10} N_{10}.
 \end{aligned} \tag{7}$$

Again note that for dN_9/dt , a constant value denoted by d_9 is added to the model to ensure that the overall mRNA level stays positive since U_3 is negatively correlated with N_9 and b_9 is negative due to the degradation term. The parameters are estimated using Eqn. (4) with $\Theta = \{a_i, b_i, c_i, d_9, p_{i,j}\}$ and $T = 10$.

Using 1 as the initial value for all parameter in the optimisation, the estimated parameters of Eqn. (7) are given in Table 5. Fig. 6 shows the comparison between the S-System model and the real data on the validation data set. The initial conditions for solving the ODEs are the first data points of each gene taken from the experimental data set.

From the estimated parameters shown in Table 5, we are able to determine the type of regulation in the network, where a positive value of the power term denotes activation while a negative value of the power term denotes inhibition. Reassuringly, all the *a priori* known degradation terms were identified to have negative values, in accordance with current biological data on the network.

The comparison between the S-System model and the real data on the validation data set shows good agreement, suggesting a good level of accuracy of the model. To quantify this, we calculate the Mean Square Error (MSE) for each

TABLE 5
Estimated Parameters for the DREAM3 S-System Model

Gene	Parameter Values
N_1	$a_1 = 0.2757, p_{1,1} = 0.3502, p_{1,2} = 0.0559,$ $p_{1,3} = -0.2789, b_1 = -0.4023$
N_2	$b_2 = -0.1875, c_2 = 0.0946$
N_3	$a_3 = 0.1478, p_{3,1} = -0.0021, p_{3,2} = 0.1393,$ $b_3 = -0.1481$
N_4	$a_4 = 0.0023, p_{4,1} = -5.1622, b_4 = -0.3555$
N_5	$a_5 = 0.1199, p_{5,1} = 0.0760, b_5 = -0.2057$
N_6	$a_6 = 0.2567, p_{6,1} = -0.0120, b_6 = -0.3035$
N_7	$a_7 = 0.0607, p_{7,1} = 0.1104, b_7 = -0.1237$
N_8	$b_8 = -0.0298, c_8 = 0.0108$
N_9	$b_9 = -0.1793, c_9 = -0.0268, d_9 = 0.1733$
N_{10}	$a_{10} = 0.0139, p_{10,1} = -1.5609, b_{10} = -0.0480$

gene between the S-System model and the real data. The MSE is computed using,

$$\text{MSE} = \frac{1}{L} \sum_{t=1}^L [N_i(t) - \hat{N}_i(t, \theta)]^2, \quad (8)$$

where L is the length of the data, N_i and \hat{N}_i respectively represent the experimental and the simulated data and $i = 1, 2, \dots, 10$. Table 6 shows the computed MSE for both the estimation and validation data sets.

The total MSE, MSE_T , is obtained by summing all the individual MSE from each genes. In general, the MSE values are small and similar between the two data sets. With the regulation types in the DREAM3 network as identified, the network interactions are as shown in Fig. 1B.

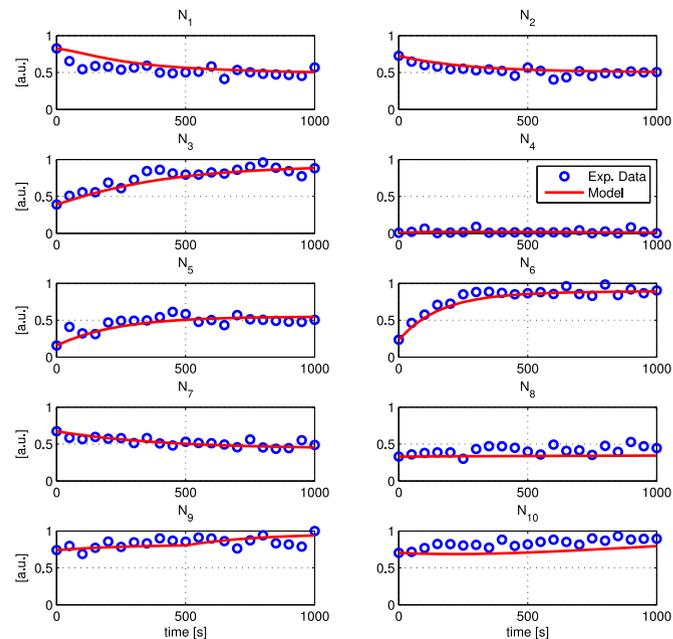


Fig. 6. Comparison between S-System model and DREAM3 data on the validation data set that is not used for parameter estimation.

TABLE 6
MSE for Both Estimation and Validation Data Sets

Gene	MSE (Estimation)	MSE (Validation)
N_1	0.0029	0.0054
N_2	0.0013	0.0021
N_3	0.0014	0.0031
N_4	0.0009	0.0010
N_5	0.0010	0.0037
N_6	0.0017	0.0036
N_7	0.0019	0.0016
N_8	0.0012	0.0088
N_9	0.0033	0.0050
N_{10}	0.0017	0.0128
MSE_T	0.0171	0.0470

4.1 Modelling of DREAM3 with Michaelis-Menten with Hill-Type Nonlinearities

Now that the regulation types between each node (activation or inhibition) have been identified, we can also use Michaelis-Menten with Hill-type nonlinearities to model the DREAM3 network, as follows:

$$\begin{aligned} \frac{dN_1}{dt} &= \frac{k_{0,1}N_2^h}{K_{M,1} + N_2^h} + \frac{k_{0,2}N_4^h}{K_{M,2} + N_4^h} + \frac{k_{0,3}}{K_{M,3} + N_5^h} + \gamma_1 N_1 \\ \frac{dN_2}{dt} &= \frac{k_{0,4}U_1^h}{K_{M,4} + U_1^h} + \gamma_2 N_2 \\ \frac{dN_3}{dt} &= \frac{k_{0,5}}{K_{M,5} + N_1^h} + \frac{k_{0,6}N_5^h}{K_{M,6} + N_5^h} + \gamma_3 N_3 \\ \frac{dN_4}{dt} &= \frac{k_{0,7}}{K_{M,7} + N_9^h} + \gamma_4 N_4 \\ \frac{dN_5}{dt} &= \frac{k_{0,8}N_7^h}{K_{M,8} + N_7^h} + \gamma_5 N_5 \\ \frac{dN_6}{dt} &= \frac{k_{0,9}N_4^h}{K_{M,9} + N_4^h} + \gamma_6 N_6 \\ \frac{dN_7}{dt} &= \frac{k_{0,10}N_8^h}{K_{M,10} + N_8^h} + \gamma_7 N_7 \\ \frac{dN_8}{dt} &= \frac{k_{0,11}U_2^h}{K_{M,11} + U_2^h} + \gamma_8 N_8 \\ \frac{dN_9}{dt} &= \frac{k_{0,12}}{K_{M,12} + U_3^h} + \gamma_9 N_9 \\ \frac{dN_{10}}{dt} &= \frac{k_{0,13}}{K_{M,13} + N_7^h} + \gamma_{10} N_{10}. \end{aligned} \quad (9)$$

We want to investigate whether the Michaelis-Menten with Hill-type nonlinearities model would encounter the same problem of inconsistent parameter estimates as highlighted in Section 3.1. For the purposes of illustration, we focus only on the highlighted pathway that involves the series of regulation from the control action to the output gene (see Fig. 1C) and as before set $h = 1$.

We repeat the parameter estimation exercise (i.e., using Eqn. (4)) where we choose 0.01, 0.1, 10 and 100 as the initial values for the optimisation for both Michaelis-Menten with Hill-type nonlinearities and S-System model structures, focusing only on genes 1, 4 and 9. The results are shown in Fig. 7 and the estimated model parameters are given in Tables 7 and 8.

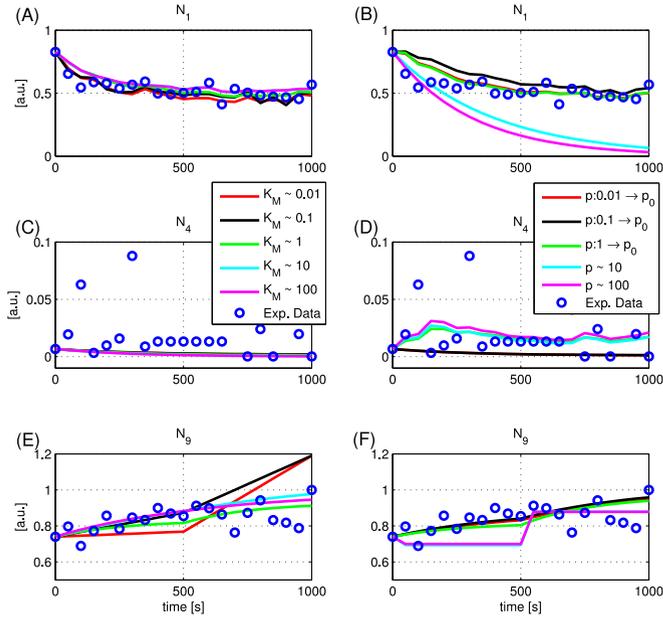


Fig. 7. Comparison of model and experimental data for different sets of estimated parameter given different initial values for optimisation. The initial values used for optimisation are 0.01, 0.1, 1, 10, and 100. Only genes in the highlighted pathway shown in Fig. 1C are shown. The experimental data shown here is an independent data set that is not used for parameter estimation. Left panel: Subfigures (A), (C), and (E) show the plots using Michaelis-Menten with Hill-type nonlinearities model structure for genes 1, 4, and 9, respectively. Here, the estimated values of K_M are close to the initial values for optimisation. Right panel: Subfigures (B), (D), and (F) show the plots using S-System model structure for genes 1, 4, and 9, respectively. The notations p_0 and $p : 0.01, 0.1, 1, 10, 100 \rightarrow p_0$ follow the same interpretation given in previous section.

As shown in Fig. 7, the estimated parameters using the Michaelis-Menten with Hill-type nonlinearities model are not consistent, as different sets of parameter are able to reproduce the dynamics of the experimental data equally well. For the S-System model, however, we obtain consistent estimates of the model parameters for genes 1 and 9 when the initial values used for optimisation are 0.01, 0.1 and 1, while for initial values of 10 and 100, the resulting parameters cannot reproduce the experimental data. For gene 4, we obtain consistent estimates of the model parameters when the initial values used for optimisation are 1, 10 and 100, while for initial values of 0.01 and 0.1 there is again poor agreement between model responses and experimental data.

4.2 Discussion on the Parameter Estimates of the Model Structures

Through our analysis of different modelling formalisms for the gene regulatory networks considered here, we have illustrated the inconsistent estimates of the model parameters obtained when using Michaelis-Menten with Hill-type nonlinearities model. This means that these model parameters are not identifiable from the available experimental data. One reason for this could be that these experimental data do not excite the relevant dynamics (in particular the saturation region) thus making the data not informative enough to obtain a consistent estimate. This inconsistent estimate is related to the notion of ‘practical parameter identifiability’ (see, e.g., [26], [27]) where the available experimental data is

TABLE 7
Estimated Parameters Given Different Initial Values for the Optimisation as Shown in Figs. 7A, C, and E

Initial Value	Gene	Parameter Values
0.01	N_1	$k_{0,1} = 0.6428, K_{M,1} = 1.6590, k_{0,2} = 0.1856, K_{M,2} = 0.0127, k_{0,3} = 0.2160, K_{M,3} = 2.7946, \gamma_1 = -0.6389,$
	N_4	$k_{0,7} = 0.0002, K_{M,7} = 0.0124, \gamma_4 = -0.1675$
	N_9	$k_{0,12} = 0.0038, K_{M,12} = 0.0725, \gamma_9 = -0.000001$
0.1	N_1	$k_{0,1} = 0.7106, K_{M,1} = 1.6914, k_{0,2} = 0.2425, K_{M,2} = 0.1409, k_{0,3} = 0.2087, K_{M,3} = 1.9633, \gamma_1 = -0.5968,$
	N_4	$k_{0,7} = 0.0003, K_{M,7} = 0.1005, \gamma_4 = -0.1671$
	N_9	$k_{0,12} = 0.0231, K_{M,12} = 0.7250, \gamma_9 = -0.0001$
1	N_1	$k_{0,1} = 0.6868, K_{M,1} = 1.5259, k_{0,2} = 0.4995, K_{M,2} = 1.0623, k_{0,3} = 0.1982, K_{M,3} = 2.7770, \gamma_1 = -0.4799,$
	N_4	$k_{0,7} = 0.0004, K_{M,7} = 1.5874, \gamma_4 = -0.1657$
	N_9	$k_{0,12} = 1.4335, K_{M,12} = 8.3217, \gamma_9 = -0.1852$
10	N_1	$k_{0,1} = 0.7539, K_{M,1} = 0.9300, k_{0,2} = 0.5454, K_{M,2} = 9.3423, k_{0,3} = 0.2165, K_{M,3} = 2.1919, \gamma_1 = -0.6471,$
	N_4	$k_{0,7} = 0.0005, K_{M,7} = 9.2461, \gamma_4 = -0.1643$
	N_9	$k_{0,12} = 1.5362, K_{M,12} = 20.3145, \gamma_9 = -0.0707$
100	N_1	$k_{0,1} = 0.7995, K_{M,1} = 1.1288, k_{0,2} = 0.5369, K_{M,2} = 83.2009, k_{0,3} = 0.2240, K_{M,3} = 1.6028, \gamma_1 = -0.6607,$
	N_4	$k_{0,7} = 0.0004, K_{M,7} = 99.3634, \gamma_4 = -0.1643$
	N_9	$k_{0,12} = 9.3453, K_{M,12} = 105.3270, \gamma_9 = -0.0897$

unable to excite the relevant dynamics to provide consistent estimate for a given model structure, as observed here. The problem of inconsistent parameter estimates is also observed

TABLE 8
Estimated Parameters Given Different Initial Values for the Optimisation as Shown in Figs. 7B, D, and F

Initial Value	Gene	Parameter Values
0.01	N_1	$a_1 = 0.2585, p_{1,1} = 0.3542, p_{1,2} = 0.0392, p_{1,3} = -0.2917, b_1 = -0.4143,$
	N_4	$a_4 = 0.0001, p_{4,1} = -0.0055, b_4 = -0.1656$
	N_9	$b_9 = -0.1043, c_9 = -0.0132, d_9 = 0.1051$
0.1	N_1	$a_1 = 0.2589, p_{1,1} = 0.4126, p_{1,2} = 0.0590, p_{1,3} = -0.2674, b_1 = -0.3307,$
	N_4	$a_4 = 0.0002, p_{4,1} = -0.0864, b_4 = -0.1657$
	N_9	$b_9 = -0.1043, c_9 = -0.0132, d_9 = 0.1067$
1	N_1	$a_1 = 0.2757, p_{1,1} = 0.3502, p_{1,2} = 0.0559, p_{1,3} = -0.2789, b_1 = -0.4023,$
	N_4	$a_4 = 0.0023, p_{4,1} = -5.1622, b_4 = -0.3555$
	N_9	$b_9 = -0.1793, c_9 = -0.0268, d_9 = 0.1733$
10	N_1	$a_1 = 0.3429, p_{1,1} = 0.4194, p_{1,2} = 9.9089, p_{1,3} = -0.3139, b_1 = -0.1267,$
	N_4	$a_4 = 0.0025, p_{4,1} = -5.6379, b_4 = -0.4444$
	N_9	$b_9 = -13.9910, c_9 = -2.6026, d_9 = 12.3031$
100	N_1	$a_1 = 0.3180, p_{1,1} = 0.3997, p_{1,2} = 112.6809, p_{1,3} = -0.2893, b_1 = -0.1622,$
	N_4	$a_4 = 0.0035, p_{4,1} = -5.1136, b_4 = -0.4614$
	N_9	$b_9 = -140.4127, c_9 = -25.0935, d_9 = 123.3048$

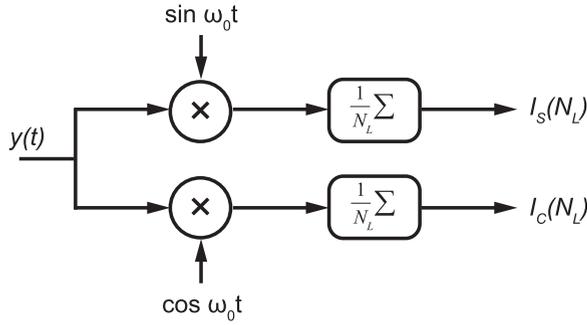


Fig. 8. Correlation method.

in [28], where the authors attempt to build a comprehensive network model for the plant circadian system, and the interactions between genes are modelled using the Michaelis-Menten with Hill-type nonlinearities model structure. The model parameters are estimated from experimental data, which are the temporal profiles of the circadian genes and proteins and a total of eight different parameter sets are found to be able to reproduce the experimental data. The estimated values of the Michaelis-Menten coefficients (K_M) from these eight sets of parameters cover a large range of possible values (from 0.01 to 490).

Although its relevance from the point of view of control system design has not to-date been considered, the problem of obtaining consistent estimates of parameters in the Michaelis-Menten model structure has been previously investigated (see the review paper [29] and references therein). In [30] and [31], different methods for fitting the Michaelis-Menten equation were analysed, and both studies concluded that different fitting methods will give different estimates of the parameters unless the experimental data is free from error (which in biological reality it never is). Different approaches to estimate the Michaelis-Menten coefficients have also been studied in [32], [33] and [34], and those studies concluded that it is difficult to obtain a consistent estimate of the Michaelis-Menten coefficients unless particular design considerations are taken into account.

On the other hand, for the parameters of the S-System model, our two illustrative examples indicate that these parameters are *locally identifiable* [35], as we are able to obtain consistent parameter estimate when different initial values are used for the optimisation. The identifiability of model parameters using a power law type of model structure (that includes the S-System model) has been investigated in [36]. Their analyses show that while in general it is practically challenging to obtain consistent estimate for all the parameters in the model, one can obtain consistent estimates of the model parameters under certain conditions. Recent work by [37] also shows that with an appropriate choice of optimiser, one can obtain consistent parameter estimates using the S-System model structure.

5 DESIGN OF A FEEDBACK CONTROLLER FOR PERTURBATION MITIGATION

Here, we show how the S-System model of the considered gene regulatory network can be used to design a controller for perturbation mitigation. To achieve an implementable design, a genetic-based controller is required, and there are

frameworks available for such designs (see, e.g., [38], [39]). In this paper, we employ a frequency domain control design methodology, motivated by the design framework proposed in [39]. In order to design controllers in the frequency domain, a linear model is required. As the S-System is a nonlinear model, we linearise it to obtain a transfer function model using the sine sweeping method (see, e.g., [22], [40]).

5.1 Sine Sweeping Method

In the sine sweeping method, sinusoidal input signals over the frequency range of interest are given as the inputs to the system. The output responses within the frequency range are then analysed in terms of their magnitude and phase relative to the input signal. By collecting these magnitude and phase values, the frequency response and transfer function model of the system can be easily obtained. Here, we summarise the procedure for obtaining a transfer function model using the sine sweeping method and refer readers to [22], [40] for complete details.

Consider a sinusoidal input $u(t) = A \sin(\omega_0 t)$, where A is the amplitude and ω_0 is the frequency. For any linear time invariant system, the output would be also sinusoidal with the same frequency but with scaled amplitude and a phase shift. In practice, the output response is subject to transient effects, as well as the effects of nonlinearities and disturbances $d(t)$, yielding,

$$y(t) = B \sin(\omega_0 t + \phi) + d(t) + \text{transient} + \text{nonlinearities}, \quad (10)$$

where $B = A|G(j\omega_0)|$, $\phi = \angle G(j\omega_0) = \tan^{-1} \frac{\text{Im}\{G(j\omega_0)\}}{\text{Re}\{G(j\omega_0)\}}$ and $G(j\omega_0)$ is the transfer function relating the input and output with j denotes the imaginary number.

The effects of transients and nonlinearities can be removed by neglecting the initial part of the data and assuming that the linear dynamics make the dominant contribution to the overall response. To reduce the effect of $d(t)$ on $y(t)$, one can use a correlation method [22], where the idea is to correlate y with a sine and cosine of the same frequency and average it over the length of the data N_L (see Fig. 8).

From Fig. 8, we obtain,

$$\begin{aligned} I_S(N_L) &= \frac{1}{N_L} \sum_{t=1}^{N_L} y(t) \sin(\omega_0 t) \\ I_C(N_L) &= \frac{1}{N_L} \sum_{t=1}^{N_L} y(t) \cos(\omega_0 t). \end{aligned} \quad (11)$$

Substituting Eqn. (10) into (11), and after some algebraic manipulation, we arrive at

$$\begin{aligned} I_S(N_L) &= \frac{A}{2} |G(j\omega_0)| \cos \phi - \frac{A}{2} |G(j\omega_0)| \frac{1}{N_L} \sum_{t=1}^{N_L} \cos(2\omega_0 t) \\ &\quad + \phi + \frac{1}{N_L} \sum_{t=1}^{N_L} d(t) \sin(\omega_0 t) \\ I_C(N_L) &= \frac{A}{2} |G(j\omega_0)| \sin \phi - \frac{A}{2} |G(j\omega_0)| \frac{1}{N_L} \sum_{t=1}^{N_L} \sin(2\omega_0 t) \\ &\quad + \phi + \frac{1}{N_L} \sum_{t=1}^{N_L} d(t) \cos(\omega_0 t). \end{aligned} \quad (12)$$

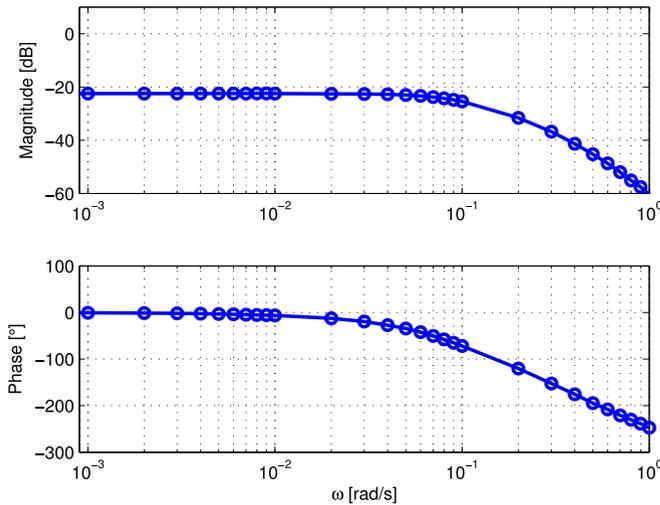


Fig. 9. Bode plot of DREAM3 network from input U_3 to output N_1 .

From Eqn. (12), the second term for both $I_S(N_L)$ and $I_C(N_L)$ will go to zero as $N_L \rightarrow \infty$. Assuming $d(t)$ is a stationary stochastic process with zero mean value and covariance function $R_d(t)$ such that $\sum_{l=0}^{\infty} l |R_d(t)| < \infty$, the third term for both $I_S(N_L)$ and $I_C(N_L)$ will be zero as $N_L \rightarrow \infty$, since the variance of the third term decays at a rate of $1/N_L$ (see [22] for details). From the remaining terms of Eqn. (12), the magnitude, $|G(j\omega_0)|$ and the phase, $\angle G(j\omega_0)$ can be estimated using the following equations, i.e.,

$$\begin{aligned} |G(j\omega_0)| &= \frac{2\sqrt{I_S^2(N_L) + I_C^2(N_L)}}{A} \\ \angle G(j\omega_0) &= \tan^{-1} \frac{I_C(N_L)}{I_S(N_L)}. \end{aligned} \quad (13)$$

For the DREAM3 network, we assume that the input to the network is through U_3 and the output of interest is the expression of gene N_1 . We apply sinusoidal signals of the form $3 \sin(\omega t) + 3$ with the frequency ω ranging from 0.001 rad/s to 1.000 rad/s. Despite using a nonlinear model, we note that the output sinusoidal responses have the same frequency as the input and no subharmonics are apparent, indicating a dominant linearity of the model. By computing the magnitude and phase values using Eqn. (13), the Bode plot of the DREAM3 network from input U_3 to output N_1 is obtained and shown in Fig. 9.

From the Bode plot, we note the following: (i) At low frequency, the magnitude of the system is about -22.5dB. (ii) The corner frequency is 0.11 rad/s. (iii) At the corner frequency, the slope is close to -40dB/dec and the phase is approximately -90° , suggesting a second order system with repeating poles. Thus, the transfer function relating input U_3 to output N_1 can be approximated by

$$\frac{N_1(s)}{U_3(s)} = \frac{0.075}{(1 + \frac{s}{0.11})^2} = \frac{0.0009}{s^2 + 0.22s + 0.012}. \quad (14)$$

From the sine sweeping method, the linear transfer function of the gene regulatory from U_3 to N_1 is given by Eqn. (14). We compare the accuracy of the linear model with the nonlinear S-System model through a step response comparison, as shown in Fig. 10. Since the base signal level

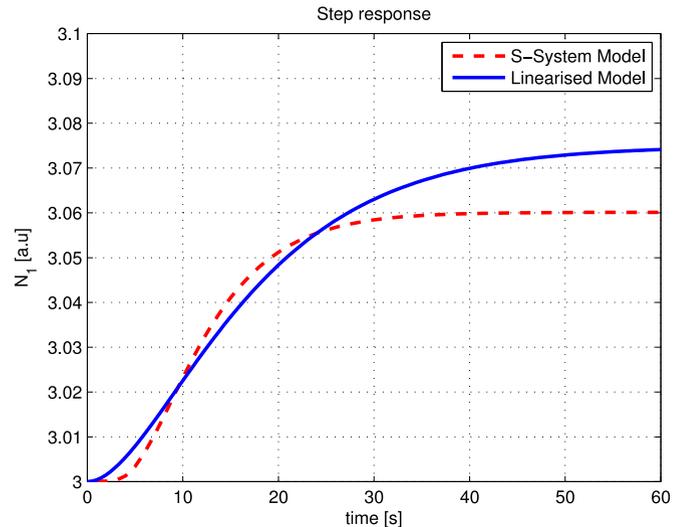


Fig. 10. Step response comparison between the linear model obtained through sine sweeping method and the full nonlinear S-System model.

used in the sine sweeping method is 3, the input is stepped from 3 to 4.

From Fig. 10, we observe similar performance between the two models in terms of their transient responses, i.e., similar rise time and settling time. On the other hand, the steady state levels between the two models are different with the linear model having a higher steady state level compared to the nonlinear model. Nevertheless, the difference between these two steady state level is relatively small, indicating acceptable accuracy of the linear model in approximating the nonlinear S-System model relating input U_3 to output N_1 .

With this transfer function identified, we can proceed with the design of the controller using a frequency domain approach.

5.2 Design of a Genetic Phase Lag Controller

Here, we illustrate the design of the genetic phase lag controller. A phase lag controller is chosen, as this type of controller is typically used to improve disturbance rejection and reduce steady state errors. The phase lag controller has the following form

$$\begin{aligned} K(s) &= \frac{K_1}{s + a_P} + K_2 \\ &= \frac{K_2 \left(s + a_P + \frac{K_1}{K_2} \right)}{s + a_P}, \end{aligned} \quad (15)$$

where the zero of the controller $z = -(a_P + (K_1/K_2))$ and the pole of the controller $p = -a_P$, with the gain of the controller being K_2 . As both the gain and phase margins of the system obtained from the Bode plot are infinite, our primary focus is on improving the transient dynamics of the disturbance rejection and reducing the steady state error.

The transfer function given in Eqn. (14) is a type 0 system, and with the use of a phase lag controller, there is no integrator in the open loop gain to eliminate the steady state error. As such, when choosing the pole of the phase lag controller, we try to place the pole, a_P as close as possible to the origin. Likewise, the static error constant, $K_p = 0.0027K_2$

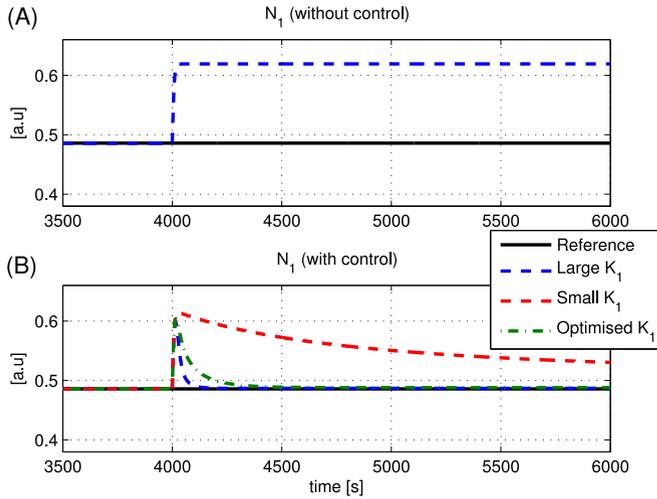


Fig. 11. (A) N_1 set-point regulation (without control). (B) N_1 set-point regulation (with control). Black solid line: Set-point. Red dotted line: N_1 response to small K_1 . Blue dashed line: N_1 response to large K_1 . Green dash-dotted line: N_1 response to optimised K_1 .

should be chosen as large as possible to reduce the steady state error. The choice of the design parameters are constrained by the achievable biological values and following the range of allowable values given in [39]; the following allowable parameter ranges are adhered to: $0.0002 \leq a_P \leq 0.0040$, $K_1 < 2.3$ and $K_2 < 1.8$.

5.3 Simulation Results

While the design of the controller is carried out using the linear model, for implementation, we carried out our simulation using the nonlinear S-System network model. In most gene regulatory network perturbation mitigation problems, we are interested in maintaining the steady state level of a particular gene of interest in the presence of a perturbation. Biologically, this can be interpreted as maintaining the level of expression of a gene of interest to ensure optimal biological function. Thus, in this simulation example, we are interested in maintaining the steady state level of N_1 at its desired reference value in the presence of a perturbation. Here, we assume that the perturbation enters the network through U_1 and our control action is provided by U_3 as depicted in Fig. 1C.

In the absence of a perturbation, the steady state level of N_1 is 0.486, thus, our control objective is to maintain the steady state level of N_1 close to 0.486 in the presence of a perturbation. In our simulation, a perturbation in the form of a step response with amplitude of 2 enters the network at time 4000s. As can be seen in Fig. 11A, without control, the steady state level of N_1 increase to 0.63 and is unable to return to its desired value.

In the design of the phase lag controller, the following values are chosen. To have the pole close to the origin, we choose $a_P = 0.0002$. To have the static error constant as large as possible, we choose $K_2 = 1.7$. For K_1 , we initially consider two cases, i.e., $K_1 = 0.04$ (controller's zero close to origin) and $K_1 = 2$ (controller's zero far from the origin). The simulation results are shown in Fig. 11B. For a small value of K_1 , we see that the performance of the system is slow and at time 6000s, there is still a noticeable steady state error, i.e., 0.044. On the other hand, for a large value of K_1 ,

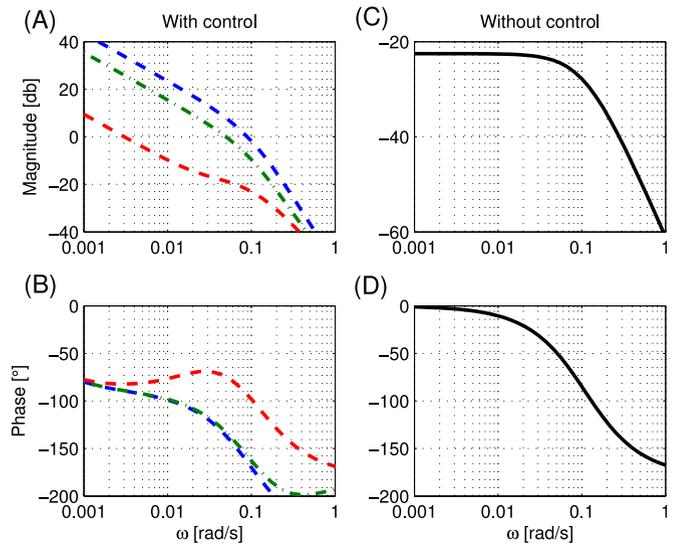


Fig. 12. (A) & (B) Gain and phase plots of system with control. Red dotted line: Small K_1 , Blue dashed line: Large K_1 . Green dash-dotted line: Optimised K_1 . (C) & (D) Gain and phase plots of system without control.

we see a significant improvement in the performance, where we get a faster response and an almost zero steady state error, i.e., 0.0008.

The Bode plots of the system with and without control are shown in Fig. 12. For a small value of K_1 , we note that the phase margin of the system is 97° . On the other hand, for a large value of K_1 , despite the good performance, we note that the phase margin of the system reduces from 97° to 10° , which is less than typically specified values. Thus, a compromise between the transient performance and overall stability robustness needs to be performed when designing the controller, and this trade-off can be effectively managed through the choice of the controller parameter K_1 . According to standard specifications, the phase margin is typically required to be between 45° to 60° (see, e.g., [16]) to achieve satisfactory performance. To find the 'optimal' value of K_1 that can achieve fast response, small steady state error and achieve a phase margin in the aforementioned range, we proceed as follows.

The transfer functions of the process and the lag compensator are given by Eqns. (14) and (15) respectively. Rewriting them here together with the substitution of $a_P = 0.0002$ and $K_2 = 1.7$, as well as defining $G_{OL}(s)$ as the open loop gain transfer function, we have the following expression.

$$G_{OL}(s) = \left[\frac{0.075}{\left(\frac{s}{0.11} + 1\right)^2} \right] \left[\left(\frac{1.7(s + 0.0002 + \frac{K_1}{1.7})}{s + 0.0002} \right) \right]. \quad (16)$$

Replacing $s = j\omega$, and after some algebraic manipulation we have

$$G_{OL}(j\omega) = \frac{Q(T_1 j\omega + 1)}{(T_2 j\omega + 1)^2 (T_3 j\omega + 1)}, \quad (17)$$

where $Q = (0.1275 + 375K_1)$, $T_1 = 1/(0.0002 + \frac{K_1}{1.7})$, $T_2 = 1/0.11$ and $T_3 = 1/0.0002$.

The magnitude and phase of $G_{OL}(j\omega)$ can be computed as follows,

$$\begin{aligned}
|G_{OL}(j\omega)| &= 20\log_{10}Q + 20\log_{10}|T_1j\omega + 1| \\
&\quad - 40\log_{10}|T_2j\omega + 1| + 20\log_{10}|T_3j\omega + 1| \\
\angle G_{OL}(j\omega) &= \tan^{-1}(T_1\omega) - 2\tan^{-1}(T_2\omega) \\
&\quad + \tan^{-1}(T_3\omega),
\end{aligned} \tag{18}$$

and we are now left with the task to find K_1 and ω to achieve our desired phase margin.

From the Bode plot in Fig. 12A, we observe that to achieve the desired phase margin would require the gain cross over frequency of $G_{OL}(j\omega)$ to be around the frequency 0.05 rad/s. With $\omega = 0.05$, solving K_1 such that $|G_{OL}(j\omega)| = 0$ and $45^\circ \leq \angle G_{OL}(j\omega) + 180 \leq 60^\circ$ are satisfied, we obtain the optimal $K_1 = 0.8$.

As shown by the green dash-dotted line in Fig. 11B, with $K_1 = 0.8$, the magnitude plot has shifted to the left. This left shift in magnitude changes the gain cross over frequency from 0.1 rad/s to the one we specified, i.e., 0.05 rad/s. On the other hand, the phase plot is similar to the case when using large K_1 . Nevertheless, more importantly, the Bode plot shown in Fig. 12A and B shows that the new phase margin is 47.4° when using $K_1 = 0.8$, which is within the preferred range and a significant improvement compared to using large K_1 .

6 CONCLUSIONS

Although several modelling formalisms are now available for the representation of gene regulatory networks, the question of their suitability for the design of synthetic feedback control systems has so far received little attention in the literature. In this paper, we show that standard modelling approaches employing Michaelis-Menten models with Hill-type nonlinearities are not appropriate for use in the design of synthetic controllers, for two reasons. First, such models require the type of regulation between interacting genes in the network to be known *a priori*, which is highly unlikely to be the case in general. Even more problematically, the values of the particular parameters in such models on which the controller design depends cannot in general be reliably identified from standard time-series data.

As an alternative approach, we propose the use of the S-System modelling formalism. While the use of the S-System modelling formalism for describing the dynamics of gene regulatory networks is well established, its usefulness for the purposes of control design has not so far been investigated. Here, we showed that using this modelling formalism combined with standard system identification procedures allows us to establish the type of regulation between each gene, obtain consistent estimates of model parameters, and hence derive a model that is suitable for the design of a synthetic genetic feedback controller. Given that the design of the considered genetic feedback controller is carried out in frequency domain, we showed that the nonlinear S-System model can be approximated by a second order linear transfer function using the sine sweeping method. Based on this transfer function model, we designed a genetic phase lag feedback controller, whose structure and parameter values can be readily implemented biologically. Simulation results show satisfactory performance of the controller in mitigating external network perturbations. The proposed modelling

and control system design approach considered here has been tailored to the problem of mitigating external perturbations in gene regulatory network. However, the proposed approach can be readily extended to address other control problems (e.g., reference tracking) and should have wide potential application to network control problems throughout the field of synthetic biology.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the financial support EPSRC and BBSRC via research grants BB/M017982/1 and from the School of Engineering of the University of Warwick. The authors would also like to thank Prof. Michael Chappell from the School of Engineering, University of Warwick for his useful discussions on parameter identifiability.

REFERENCES

- [1] M. Cantoni, E. Weyer, Y. Li, S. K. Ooi, I. Mareels, and M. Ryan, "Control of large-scale irrigation networks," *Proc. IEEE*, vol. 95, no. 1, pp. 75–91, Jan. 2007.
- [2] R. F. Arritt and R. C. Dugan, "Distribution system analysis and the future smart grid," *IEEE Trans. Ind. Appl.*, vol. 47, no. 6, pp. 2343–2350, Nov./Dec. 2011.
- [3] S. P. Cornelius, W. L. Kath, and A. E. Motter, "Realistic control of network dynamics," *Nature Commun.*, vol. 4, no. 1942, p. 9, 2012.
- [4] D. Schwabenberg, B. P. J. Becker, and M. Xu, "Journal of hydroinformatics," *Open Real-Time Control (RTC)-Tools Softw. Framework Model. RTC Water Resources Systems*, vol. 17, no. 1, pp. 130–148, 2015.
- [5] M. Hajjahmadi, B. De Schutter, and H. Hellendoorn, "Robust H_∞ switching control techniques for switched nonlinear systems with application to urban traffic control," *Int. J. Robust Nonlinear Control*, vol. 26, pp. 1286–1306, 2016.
- [6] Y.-Y. Liu, J. J. Slotine, and A.-L. Barabasi, "Controllability of complex networks," *Nature*, vol. 7346, pp. 167–173, 2011.
- [7] Y. Tang, G. Huijin, D. Wei, L. Jianquan, A. V. Vasilakos, and J. Kurths, "Robust multiobjective controllability of complex neuronal networks," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 13, no. 4, pp. 778–791, Jul./Aug. 2015.
- [8] Y.-Y. Liu and A.-L. Barabasi, "Control principles of complex systems," *Rev. Modern Phys.*, vol. 88, no. 3, 2016, Art. no. 035006.
- [9] C. Nowzari, V. M. Preciado, and G. J. Pappas, "Analysis and control of epidemics: A survey of spreading processes on complex networks," *IEEE Control Syst.*, vol. 36, no. 1, pp. 26–46, Feb. 2016.
- [10] A. Vinayagam, et al., "Controllability analysis of the directed human protein interaction network identifies disease genes and drug targets," *Proc. Nat. Acad. Sci. USA*, vol. 113, no. 18, pp. 4976–4981, 2016.
- [11] M. Foo, J. Kim, and D. G. Bates, "System identification of gene regulatory networks for perturbation mitigation via feedback control," in *Proc. IEEE Int. Conf. Netw. Sensing Control*, 2017, pp. 216–221.
- [12] D. Marbach, T. Schaffter, T. Mattiussi, and D. Floreano, "Generating realistic in silico gene networks for performance assessment of reverse engineering methods," *J. Comput. Biol.*, vol. 16, no. 2, pp. 229–239, 2009.
- [13] G. Stolovitzky, D. Monroe, and A. Califano, "Dialogue on reverse-engineering assessment and methods: The DREAM of high-throughput pathway inference," *Ann. New York Acad. Sci.*, vol. 1115, pp. 1–22, 2007.
- [14] G. Stolovitzky, R. J. Prill, and A. Califano, "Lessons from the DREAM2 challenges," *Ann. New York Acad. Sci.*, vol. 1158, pp. 159–195, 2009.
- [15] M. Foo, I. Gherman, K. J. Denby, and D. G. Bates, "Control strategies for mitigating the effect of external perturbations on gene regulatory networks," *Proc. IFAC World Congress*, vol. 50, pp. 12647–12652, 2017.
- [16] K. Ogata, *Modern Control Engineering*. Upper Saddle River, NJ, USA: Prentice Hall, 2010.
- [17] M. Savageau, *Biochemical Systems Analysis: A Study of Function and Design in Molecular Biology*. Reading MA, USA: Addison-Wesley, 1976.

- [18] E. O. Voit, *Canonical Nonlinear Modeling. S-System Approach to Understanding Complexity*. New York, NY, USA: Van Nostrand Reinhold, 1991.
- [19] S. Kikuchi, D. Tominaga, M. Arita, K. Takahashi, and M. Tomita, "Dynamic modeling of genetic networks using genetic algorithm and S-system," *Bioinf.*, vol. 19, no. 5, pp. 643–650, 2003.
- [20] S. Kimura, et al., "Inference S-system models of genetic networks using cooperative coevolutionary algorithm," *Bioinf.*, vol. 21, no. 7, pp. 1154–1163, 2005.
- [21] E. O. Voit and M. Savageau, "Accuracy of alternative representations for integrated biochemical systems," *Biochemistry*, vol. 26, no. 21, pp. 6869–6880, 1987.
- [22] L. Ljung, *System Identification: Theory for the User*. Upper Saddle River, NJ, USA: Prentice Hall, 1999.
- [23] T. S. Gardner, D. di Bernardo, D. Lorenz, and J. J. Collins, "Inferring genetic networks and identifying compound mode of action via expression profiling," *Sci.*, vol. 301, no. 5629, pp. 102–105, 2003.
- [24] D. di Bernardo, et al., "Chemogenomic profiling on a genome-wide scale using reverse-engineered gene networks," *Nature Biotechnol.*, vol. 23, no. 3, pp. 377–383, 2005.
- [25] M. Bansal, V. Belcastro, A. Ambesi-Impiomato, and D. di Bernardo, "How to infer gene networks from expression profiles," *Molecular Syst. Biol.*, vol. 78, p. 10, 2007.
- [26] A. Holmberg, "On the practical identifiability of microbial growth models incorporating Michaelis-Menten type nonlinearities," *Math. Biosci.*, vol. 1, pp. 23–43, 1982.
- [27] D. L. I. Janzen, et al., "Parameter identifiability of fundamental pharmacodynamic models," *Frontiers Physiology*, vol. 7, 2016, Art. no. 590.
- [28] K. Fogelmark and C. Troein, "Rethinking transcriptional activation in the arabidopsis circadian clock," *PLoS Comput. Biol.*, vol. 10, no. 7, 2014, Art. no. e1003705.
- [29] S. Schnell and P. K. Maini, "A century of enzyme kinetics: Reliability of the K_M and v_{max} estimates," *Comments Theoretical Biol.*, vol. 8, pp. 169–187, 2003.
- [30] G. L. Atkins and I. A. Nimmo, "A comparison of seven methods for fitting the Michaelis-Menten equation," *Biochemical J.*, vol. 149, pp. 775–777, 1975.
- [31] D. J. Currie, "Estimating Michaelis-Menten parameters: Bias, variance and experimental design," *Biometrics*, vol. 38, pp. 907–919, 1982.
- [32] A. Cornish-Bowden, "Weighting of linear plots in enzyme kinetics," *J. Molecular Sci.*, vol. 2, no. 4, pp. 107–112, 1982.
- [33] T. L. Toulas and C. P. Kitsos, "Fitting the Michaelis-Menten model," *J. Comput. Appl. Math.*, vol. 296, pp. 303–319, 2016.
- [34] T. L. Toulas and C. P. Kitsos, "Estimation aspects of the Michaelis-Menten model," *REVSTAT Statistical J.*, vol. 14, no. 2, pp. 101–118, 2017.
- [35] K. R. Godfrey, "The identifiability of parameters of models used in biomedicine," *Math. Modelling*, vol. 7, pp. 1195–1214, 1986.
- [36] S. Srinath and R. Gunawan, "Parameter identifiability of power-law biochemical systems models," *J. Biotechnol.*, vol. 149, pp. 132–140, 2010.
- [37] M. Iwata, K. Sriyudthsak, M. Y. Hirai, and F. Shiraiishi, "Estimation of kinetic parameters in an S-System equation model for a metabolic reaction system using the Newton-Raphson method," *Math. Biosci.*, vol. 248, pp. 11–21, 2014.
- [38] J. Ang, S. Bagh, B. P. Ingalls, and D. R. McMillen, "Considerations for using integral feedback control to construct perfectly adapting synthetic gene network," *J. Theoretical Biol.*, vol. 266, no. 4, pp. 723–738, 2010.
- [39] A. W. K. Harris, J. A. Dolan, C. L. Kelly, J. Anderson, and A. Papachristodoulou, "Designing genetic feedback controllers," *IEEE Trans. Biomed. Circuits Syst.*, vol. 9, no. 4, pp. 475–484, Aug. 2015.
- [40] T. Soderstrom and P. Stoica, *System Identification*. Prentice-Hall, NJ, USA: Englewood Cliffs, 1988.



Mathias Foo received the BEng (Hons.) and MEngSc degrees in electronic engineering from Multimedia University, Malaysia, in 2002 and 2004, respectively, and the PhD degree from the University of Melbourne, Melbourne, Australia, in 2012. He was a post-doctoral research fellow with the Asia Pacific Center for Theoretical Physics (APCTP), Pohang, South Korea, from 2012–2015. He is currently a research fellow in the Warwick Integrative Synthetic Biology Centre, School of Engineering, University of Warwick, United Kingdom. His research interests include dynamical system modelling, application of control system, and control theory for synthetic biology.



Jongrae Kim received the PhD degree in aerospace engineering from Texas A&M University, College Station, Texas, in 2002. Currently, he is an associate professor in the Institute of Design, Robotics & Optimisation (iDRO) and aerospace systems engineering in the School of Mechanical Engineering, the University of Leeds, Leeds, United Kingdom. He was a postdoctoral researcher with the University of California, Santa Barbara, California, in 2002 and 2003, and a research associate with the University of Leicester, Leicester, United Kingdom from 2004–2007. He was a lecturer in biomedical engineering/aerospace sciences, University of Glasgow, Glasgow, United Kingdom from 2007–2014. His main research interests include the area of robustness analysis, optimal control and estimation, large-scale network analysis, system identification, dynamics, robotics, systems biology, synthetic biology, and neuroscience.



Declan G. Bates received the BEng degree in electronic engineering and the PhD degree in robust control theory from the School of Electronic Engineering, Dublin City University, Ireland, in 1992 and 1996, respectively. On completing his PhD, he joined the Control and Instrumentation Research Group led by Prof. Ian Postlethwaite in the Department of Engineering, Leicester University, where he worked as a post-doctoral research associate, lecturer, and senior lecturer, before being appointed to a personal chair in control engineering. In 2010, he was appointed professor of biological systems engineering in the College of Engineering, Mathematics, and Physical Sciences of the University of Exeter and in 2013 he moved to the University of Warwick as a professor of bioengineering. His research is focussed on the modelling, analysis, design, and control of complex biological and medical systems. He is co-director of the EPSRC/BBSRC Warwick Integrative Synthetic Biology Centre (WISB), and co-director at the Warwick EPSRC/BBSRC Centre for Doctoral Training in Synthetic Biology.

▷ For more information on this or any other computing topic, please visit our Digital Library at www.computer.org/publications/dlib.