HAL
open science

Comparing different ODE modelling approaches for gene regulatory networks

A. Polynikis, S.J. Hogan, M. Di Bernardo

To cite this version:

A. Polynikis, S.J. Hogan, M. Di Bernardo. Comparing different ODE modelling approaches for gene regulatory networks. Journal of Theoretical Biology, 2009, 261 (4), pp.511. ⟨10.1016/j.jtbi.2009.07.040⟩. ⟨hal-00554639⟩

HAL Id: hal-00554639

https://hal.science/hal-00554639v1

Submitted on 11 Jan 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Autorisation HAL

HAL Authorization

Author's Accepted Manuscript

www.elsevier.com/locate/yjtbi

Comparing different ODE modelling approaches for gene regulatory networks

A. Polynikis, S.J. Hogan, M. di Bernardo

PII: S0022-5193(09)00352-X
DOI: doi:10.1016/j.jtbi.2009.07.040
Reference: YJTBI 5657

To appear in: *Journal of Theoretical Biology*

Received date: 30 October 2008
Revised date: 8 July 2009
Accepted date: 30 July 2009

Cite this article as: A. Polynikis, S.J. Hogan and M. di Bernardo, Comparing different ODE modelling approaches for gene regulatory networks, *Journal of Theoretical Biology*, doi:10.1016/j.jtbi.2009.07.040

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Comparing different ODE modelling approaches for gene regulatory networks

A.Polynikis*[,a], S.J. Hogan[a], M. di Bernardo[a,b]

[a] *Department of Engineering Mathematics, University of Bristol, Queen's Building, University Walk, Bristol BS8 1TR, UK*
[b] *Department of Systems and Computer Science, University of Naples Federico II, Via Claudio 21, 80125 Naples, Italy*

Abstract

A fundamental step in synthetic biology and systems biology is to derive appropriate mathematical models for the purposes of analysis and design. For example, to synthesize a gene regulatory network, the derivation of a mathematical model is important in order to carry out in silico investigations of the network dynamics and to investigate parameter variations and robustness issues. Different mathematical frameworks have been proposed to derive such models. In particular, the use of sets of nonlinear ordinary differential equations (ODEs) has been proposed to model the dynamics of the concentrations of mRNAs and proteins. These models are usually characterized by the presence of highly nonlinear Hill function terms. A typical simplification is to reduce the number of equations by means of a quasi-steady-state assumption on the mRNA concentrations. This yields a class of simplified ODE models. A radically different approach is to replace the Hill functions by piecewise-linear approximations [5]. A further modelling approach is the use of discrete-time maps [7] where the evolution of the system is modelled in discrete, rather than continuous, time. The aim of this paper is to discuss and compare these different modelling approaches, using a representative gene regulatory network. We will show that different models often lead to conflicting conclusions concerning the existence and stability of equilibria and stable oscillatory behaviours. Moreover, we shall discuss, where possible, the viability of making certain modelling approximations (e.g. quasi-steady-state mRNA dynamics or piecewise-linear approximations of Hill functions) and their effects on the overall system dynamics.

Key words: Transcription, Hill coefficient, Hopf bifurcation

Accepted manuscript

1. Introduction

A number of gene regulatory networks has been proposed to perform certain desired functions. Examples include genetic switches [18], robust genetic oscillators [16] and the many entries submitted every year to the international IGEM competition on synthetic biology (see [49] for more details). A key step in the design and analysis of synthetic biological networks is the possibility of in silico testing of their behaviour, evaluation of the possible design options and validation of their performance and viability. The availability of a realistic mathematical model of the network of interest is of the utmost importance to carry out such testing.

The recent development of advanced experimental techniques in molecular biology has increased the amount of available experimental data on gene regulation which has led to a rapidly growing interest in mathematical modelling methods for the study and analysis of gene regulation [9, 15, 27, 29, 40, 47]. One of the very first mathematical approaches is the framework of boolean networks [30, 32, 42, 46] which is based on three assumptions: (i) the state of each gene can be either ON or OFF, (ii) the regulatory control of gene expression can be approximated by Boolean logical rules and (iii) all genes update their ON and OFF

*Corresponding author
Email address: Th.Polynikis@bristol.ac.uk (A.Polynikis)

ACCEPTED MANUSCRIPT

state synchronously [40]. Some recent studies deal with the comparison of Boolean models with ordinary differential equations models by considering specific biological networks [6, 8] . Specifically, in [8] they demonstrate how a Boolean model can be derived in terms of a mathematically well defined coarse-grained limit of an underlying ODE model.

Instead of taking a continuous deterministic approach, some authors have proposed using discrete stochastic models of gene regulation. Two approaches widely used to model stochastic events in gene regulatory networks are the chemical master equation and the stochastic simulation algorithm [2, 20, 21, 33, 38, 39].

This paper focuses on mathematical models based on ordinary differential equations (ODEs). This kind of model is arguably the most widespread formalism for modelling gene regulatory networks. These models are best analyzed using tools developed for nonlinear systems, in order to investigate bifurcation behaviour, locate limit cycles or analyze network dynamics. In the extensive literature on different ODE modelling approaches, several options are available, such as the number of equations to be used, the functional form of the kinetic laws and parameter values. The aim of this paper is to focus attention on the importance of these choices. The goal is to study and compare the different dynamics predicted by each model, emphasizing advantages and disadvantages. We will see that the choice of modelling framework and the assumptions made can determine the nature and quality of the expected behaviour during the in silico testing and validation phases.

For the sake of clarity and simplicity, we will illustrate our findings by means of a widely used representative example: a two-gene activator-inhibitor network (see Figure 4). Such an example, despite its simplicity, is well suited to emphasize the major dynamic consequences of the various modelling options being explored. It is worth mentioning here that different versions of the activator-inhibitor network have been often studied in previous work, e.g. in [48] where no self-regulation is considered, and also in [11, 13, 26] where self-regulation of one or both genes is considered. In our case we shall not consider any self-regulation.

Here, we use this network to explore the impact of some key assumptions commonly made when modelling gene networks. Specifically, we study:

1. the effects of quasi-steady-state hypothesis for the mRNA dynamics;
2. the effects of variation of the Hill coefficients;
3. the effects of taking the limit of Hill coefficient to infinity, namely the approximation of Hill functions with piecewise-linear (PWL) functions;
4. the effects of discretizing the continuous-time ODE models.

We study all of the above cases by expounding in a new framework some key results presented in the literature and by extending and integrating them with novel analytical tools. We wish to emphasize that the results presented in this paper can have implications when larger and more complex synthetic networks are studied.

The outline of the paper is as follow. In Section 2 we briefly give an overview of gene regulatory networks. In Section 3 we put the problem of modelling gene regulatory networks into the framework of ordinary differential equations. We present the different ODE models studied in this paper. We also derive a general discrete-time model, as a discretized version of the continuous time model. In Section 4, we write down the explicit equations of each model, for the representative example of an activation-inhibition network. Sections 5, 6 and 7 present the mathematical analysis of the various models. Specifically, in section 5 we perform stability and bifurcation analysis of the nonlinear models and reveals the effects of: (i) the steady-state mRNA assumption, (ii) the selection of Hill coefficient values. Section 6 deals with the effects of the piecewise linear approximation of the Hill function and section 7 shows the effects of the discretization of the continuous-time models. In the final section we present our conclusions.

2. Gene regulatory networks: an overview

The central dogma defines the paradigm of molecular biology. Genes are perpetuated as sequences of nucleic acid, but function by being expressed in the form of proteins [35]. *Transcription* and *translation* are responsible for their conversion from one form to the other. Transcription generates a messenger RNA

Accepted manuscript

ACCEPTED MANUSCRIPT

Table 1: Notation

a, b	: genes		
R_a, R_b	: transcribed mRNAs,	r_a, r_b	: concentration of transcribed mRNAs,
P_a, P_b	: translated proteins,	p_a, p_b	: concentration of translated proteins,
m_a, m_b	: maximal transcription rates,	γ_a, γ_b	: mRNA degradation rates,
k_a, k_b	: translation rates,	δ_a, δ_b	: protein degradation rates,
θ_a, θ_b	: expression thresholds,	n_a, n_b	: Hill coefficients,
$h^+(\cdot)$	: Hill function for activation,	$h^-(\cdot)$	: Hill function for inhibition,
$s^+(\cdot)$	: PWL function for activation,	$s^-(\cdot)$	: PWL function for inhibition.

(mRNA) which provides an intermediate that carries the copy of a DNA sequence that represents a protein. It is a single-stranded RNA identical in sequence with one of the strands of the duplex DNA. In protein-coding genes, translation will convert the nucleotide sequence of mRNA into the sequence of amino acids comprising a protein [35]. This two-stage process is called *gene expression*.

Each protein produced by the genes, has its own role in the cell. Some proteins are structural and will accumulate at the cell-wall or within the cell to give it particular properties. Other proteins can be enzymes that catalyse certain reactions. A large group of proteins have an important role in the regulation of the genes, known as *transcription factors*. Gene regulation by transcription factors can be negative or positive. In negative regulation, an inhibitor protein binds the operator to prevent a gene from being expressed. In positive regulation, a transcription factor is required to bind at the promoter in order to enable RNA polymerase to initiate transcription [35].

Several other steps in the gene expression process may be modulated [35]. Apart from DNA transcription regulation, the expression of a gene may be controlled during RNA processing and transport (in eukaryotes), RNA translation, and the post-translational modification of proteins [9]. The degradation of gene products can also be regulated in the cell. Hence, a gene regulatory network is a collection of DNA, RNA, proteins, and other molecules which interact with each other. These interactions control the rates at which genes in the network are transcribed into mRNA, the rates at which the mRNA are translated into proteins and in general control the cell behavior. Gene regulation gives the cell control over its structure and function, like the response of cells to environmental signals, the differentiation of cells and groups of cells in the unfolding of developmental programs, and the replication of the genome preceding cell division [9].

3. Modelling gene regulatory networks

Gene regulatory networks can be modelled from first principles using Michaelis-Menten enzymatic kinetics, together with the usual rules of reaction kinetics [1]. The resulting models, when spatial effects are neglected, are given in terms of ordinary differential equations describing the rate of change of the concentrations of gene products and proteins. A key component of all these models is the Hill function [28], used to describe the transcription phase. The presence of this highly nonlinear function, whilst accurately modelling the network, inevitably leads to restrictions on the analytical tools available to understand and predict the dynamics. It was proposed that the resulting equations can be simplified by considering piecewise-linear approximations of these Hill functions [5]. Another possibility [7] is to discretize the continuous-time ODEs to obtain a discrete-time system. In what follows, we briefly outline the main features of each of these modelling approaches.

3.1. Ordinary differential equations

When ordinary differential equations are used, the cellular concentration of proteins, mRNAs and other molecules are represented by continuous time variables with the constraint that a concentration can not be negative. For a typical transcription-translation process, the ODEs modelling approach associates two ODEs with any given gene i; one modelling the rate of change of the concentration of the transcribed mRNA, say

Accepted manuscript

ACCEPTED MANUSCRIPT

r_i, and the other describing the rate of change of the concentration of its corresponding translated protein, say p_i. Thus for a network with N genes we have:

$$\textit{Transcription:} \qquad \frac{dr_i}{dt} = F(f_i^R(p_1), f_i^R(p_2), ..., f_i^R(p_n)) - \gamma_i r_i, \tag{1}$$

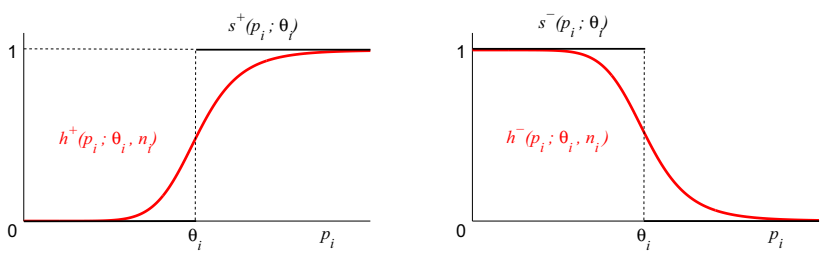
$$\textit{Translation:} \qquad \frac{dp_i}{dt} = f_i^P(r_i) - \delta_i p_i, \tag{2}$$

where $i = 1, \ldots, N$. The functions $f_i^R(p_j) : \mathbb{R} \to \mathbb{R}$ are usually nonlinear. They describe the dependence of mRNA concentration on protein concentration p_j. If protein p_j has no effect on mRNA r_i, then $f_i^R(p_j)$ is set to zero. The functional $F(\cdot)$ in (1) is typically defined in terms of sums and products of functions f_i^R. For example, if two proteins p_l and p_m are both needed to regulate mRNA r_i, then a candidate functional F might be $F(f_i^R(p_l), f_i^R(p_m)) = f_i^R(p_l) f_i^R(p_m)$. Equation (1) states that the rate of change in the concentration of mRNA r_i is the difference between the *synthesis* term $F(f_i^R(p_1), f_i^R(p_2), ..., f_i^R(p_n))$ and the *degradation* term $\gamma_i r_i$. Function $f_i^P(r_i)$ in (2) describes the translation of the mRNA r_i into a protein p_i. Parameters γ_i, δ_i ($i = [1, .., N]$), represent the degradation parameters of the mRNAs and proteins produced by gene i. As is common in many models, we shall assume that the degradation of proteins or mRNAs is not regulated, namely that it does not depend on the concentrations of other molecules in the cell.

Transcription functions, $f_i^R(\cdot)$, are derived from chemical first principles (e.g. the law of mass action) or simple "second principles" (e.g. Michaelis-Menten enzymatic kinetics). Experimental evidence suggests a monotonic sigmoidal-shaped function [50, 51] which increases when p_i is an activator and decreases when p_i is an inhibitor. A useful function satisfying this property is the *Hill function*. The Hill function for activation, $h^+(p_i; \theta_i, n_i) : \mathbb{R}_{\geq 0} \times \mathbb{R}^2_{>0} \to \mathbb{R}_{\geq 0}$, is increasing and has two real parameters, θ_i and n_i:

$$h^+(p_i; \theta_i, n_i) = \frac{p_i^{n_i}}{p_i^{n_i} + \theta_i^{n_i}}. \tag{3}$$

It describes a curve that rises from zero and approaches unity as shown in Figure 1(a). The parameter θ_i is the *expression threshold*, and has units of concentration. It is the threshold of protein concentration, p_i, needed to produce a significant increase in the mRNA regulated by p_i. The parameter n_i is called *Hill coefficient* (or *cooperativity coefficient*) and it controls the steepness of the Hill function. The larger n_i, the more step-like is the Hill function. Biologically, the Hill coefficient is related to the molecular binding mechanism. In simple cases n is the number of protein monomers required for saturation of binding to the DNA [48]. The Hill function for inhibition, $h^-(p_i; \theta_i, n_i) : \mathbb{R}_{\geq 0} \times \mathbb{R}^2_{>0} \to \mathbb{R}_{\geq 0}$, is defined in a similar way



(a) Activation function (b) Inhibition function

Figure 1: Transcription functions for activation and inhibition. Hill functions are plotted in red, PWL functions in black.

(see Figure 1(b)). It is a decreasing function given by:

$$h^-(p_i; \theta_i, n_i) = 1 - h^+(p_i; \theta_i, n_i) = \frac{\theta_i^{n_i}}{p_i^{n_i} + \theta_i^{n_i}}. \tag{4}$$

ACCEPTED MANUSCRIPT

Because of the nonlinearity of the Hill functions, the solutions of a system of ordinary differential equations of a network of many genes cannot generally be determined by analytical means.

Several authors have proposed to approximate the Hill functions by piecewise-linear (PWL) functions [5, 22, 23, 31, 44, 45]. This approximation is based on the switch-like character displayed by some genes whose expression is regulated by steep sigmoid curves. Below (above) a certain concentration, the activator (inhibitor) protein has little influence, whereas above (below) this concentration, the influence of the protein rapidly reaches a maximum level (normalized to unity). From the mathematical point of view, a piecewise-linear function can be seen as the limit of the Hill function as the Hill coefficient n_i tends to infinity.

These piecewise-linear approximations are step functions, $s^-(p_i;\theta_i)$ and $s^+(p_i;\theta_i)$, given by:

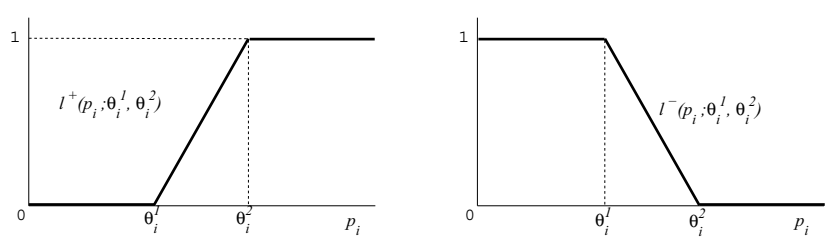
$$s^+(p_i;\theta_i) = \begin{cases} 0, & p_i < \theta_i, \\ 1 & p_i > \theta_i, \end{cases} \qquad s^-(p_i;\theta_i) = 1 - s^+(p_i;\theta_i). \tag{5}$$

These are shown in Figures 1(a) and 1(b) in black. These functions are not defined for $p_i = \theta_i$. Later we will show that this limitation has important consequences for this modelling approach.

To avoid this problem, one can include a third section between full activation (inhibition) and no activation (inhibition), where the function increases (decreases) linearly with p_i. In particular, in the work originated by Plahte et al [36], and also in [3, 4, 19] the following PWL function for activation is used:

$$l^+(p_i;\theta_i^1,\theta_i^2) = \begin{cases} 0, & p_i < \theta_i^1 \\ \mu p_i + \nu, & \theta_i^1 < p_i < \theta_i^2 \\ 1, & p_i > \theta_i^2 \end{cases} \tag{6}$$

which uses two threshold parameters θ_i^1, θ_i^2 to define a saturation interval, and two real parameters, $\mu > 0$ and $\nu < 0$ which define the slope of the function between these two thresholds. Similarly the inhibition function $l^-(p_i;\theta_i^1,\theta_i^2) = 1 - l^+(p_i;\theta_i^1,\theta_i^2)$.



(a) Activation function (b) Inhibition function

Figure 2: PWL functions with saturation interval for activation and inhibition.

While these functions resolve the problem of discontinuity at the threshold, the extra linear region gives rise to other problems, namely the need to identify the two threshold parameters θ_i^1, θ_i^2. Also, these functions become locally nonlinear if multiplication of transcription functions are allowed.

The translation phase is modelled by (2). The function $f_i^P(r_i)$ is usually taken to be a linear term proportional to the concentration of mRNA r_i, resulting in a linear differential equation.

3.2. Assumption of quasi-steady-state mRNA concentrations

Many models in the literature make an important simplifying assumption that the control of gene expression resides in the regulation of gene transcription. This assumption is based on the fact that, in some

ACCEPTED MANUSCRIPT

gene regulatory networks, the mRNA dynamics are much faster than the protein dynamics, leading to the mRNA concentrations reaching their equilibrium much faster than the protein concentrations.

From the mathematical point of view, this assumption is equivalent to taking $\dot{r}_i \approx 0$ in (1) leading to the static equation:

$$r_i = \frac{1}{\gamma_i} F(f_i^R(p_1), f_i^R(p_2), ..., f_i^R(p_n)). \tag{7}$$

Substituting this into (2) gives a reduced order model, involving only the protein concentrations of each gene, of the form:

$$\dot{p}_i = f_i^P(\frac{1}{\gamma_i} F(f_i^R(p_1), f_i^R(p_2), ..., f_i^R(p_n))) - \delta_i p_i. \tag{8}$$

This assumption is common in the literature, since many of the proposed models silently adopt this simplification and use equations only for the protein concentrations. However, as we will see later, in some cases this assumption can have effects on the predicted dynamics of a gene regulatory network.

3.3. Discrete-time modelling

As originally proposed by [23] and more recently in [7], discrete-time models can be used to study gene regulatory networks. The idea is to derive a difference-equation model describing the change of the gene product concentrations at discrete time intervals. It is suggested that this may be appropriate to (coarse-grain) model gene regulation where local complex chemical reactions have to be integrated over short time scales in order to produce interactions affecting expression levels on larger time scales [7]. The discrete-time model in [7], is based on the quasi-steady-state mRNA assumption. Hence the model has a single state variable (either mRNA or protein) for each gene. The protein concentrations evolve according to combined interactions from other genes in the network. The interactions are given by step functions which assume that a gene acts on another gene, or becomes inactive, only when its product concentration exceeds a threshold. As will be shown later in this paper, it is possible to consider the model in [7] as a specific instance of a wider class of models obtained by discretizing the ODE model of the network of interest. Whilst greatly simplifying computations, we will show that spurious dynamics are introduced which can severely hinder understanding of the network under consideration.

3.4. A summarizing scheme

The models under investigation in this paper are summarized in Figure 3. The *complete nonlinear* model (CNM) considers different variables for the concentrations of mRNAs and proteins. Transcription is modelled by a nonlinear Hill function and the translation of mRNAs to proteins is modelled by simple linear functions.

If we replace the Hill functions in the CNM by step functions $s^-(p_i;\theta_i)$ and $s^+(p_i;\theta_i)$, we have the *complete piecewise linear* model (CPWLM). This model retains different variables for the concentrations of mRNAs and proteins. If we make the quasi-steady-state mRNA assumption, then from the complete nonlinear model (CNM), we derive the *simplified nonlinear* model, (SNM). If we replace the Hill functions in the SNM by step functions $s^-(p_i;\theta_i)$ and $s^+(p_i;\theta_i)$, we have the *simplified piecewise linear* model (SPWLM). Later we will show a connection between the SPWLM and the discrete-time model proposed by [7]. (A higher-dimensional discrete-time model could also be obtained by discretizing the CPWLM but this would present the same problems discussed later as the one derived from the SPWLM. For the sake of brevity, this model was therefore left out from this paper.)

4. A representative example

To illustrate the advantages and disadvantages of the various models, we use the **activation-inhibition** two-gene network as a representative example (see Figure 4). In doing so, we will integrate and expand analysis presented in [48] for a class of two-gene networks.

In our network, the DNA is assumed to carry two genes, gene a and gene b. Gene a has a binding site in the promoter region for an activator (protein P_b) and gene b has a binding site for an inhibitor (protein P_a).

Accepted manuscript

ACCEPTED MANUSCRIPT

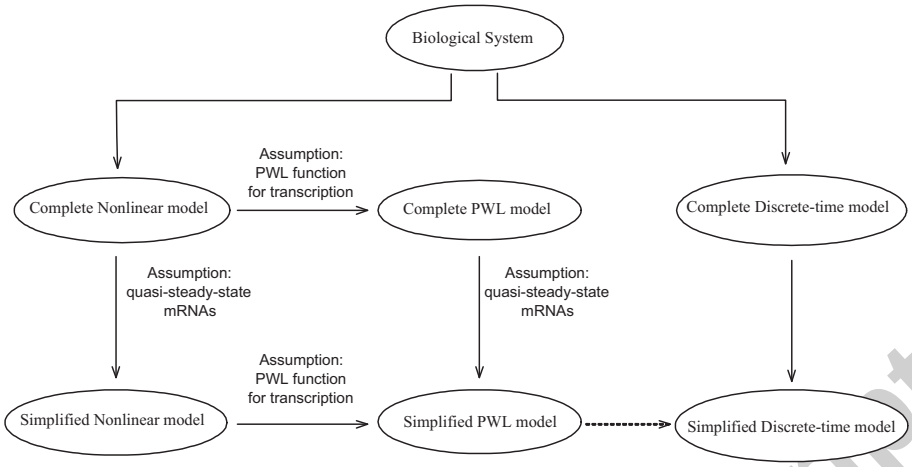


Figure 3: The relationships between the biological system and its mathematical models.

Binding of the proteins is assumed to occur fast compared to transcription and translation, and accordingly the equilibrium assumption is valid [48]. We shall not consider self-regulation; the protein produced by a gene does not affect the expression of the gene itself. The notation $p_i \rightarrow r_j$ means that, protein p_i activates

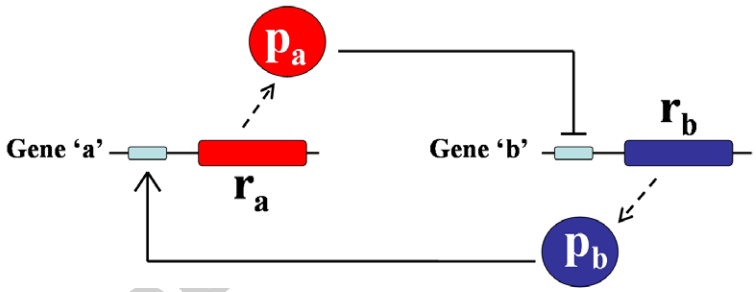


Figure 4: An example of a genetic regulatory network consisting of two genes a and b. It consists of four molecular species; proteins P_a, P_b and mRNAs R_a, R_b. Their concentrations are represented by the continuous variables p_a, p_b, r_a, r_b respectively. Protein P_b acts as an activator on gene a; it increases the production of mRNA R_a. Protein P_a acts as an inhibitor on gene b, reducing the production of mRNA R_b.

gene j, resulting in maximum transcription of mRNA r_j; whereas $p_i \dashv r_j$ means that protein p_i inhibits the gene expression of gene i.

4.1. The complete nonlinear model (CNM)

We start with the complete nonlinear model (CNM) of the network of two genes shown in Figure 4. Such a model uses four state variables. The concentration of mRNA produced by gene i is denoted by r_i while the corresponding protein concentration is denoted by p_i, for $i = a, b$. The activation of gene a by protein P_b is modelled by the Hill function for activation $h^+(p_b; \theta_b, n_b)$. The inhibition of gene b by protein P_a is modelled by the Hill function for inhibition, $h^-(p_a; \theta_a, n_a)$. The translation of mRNA and the degradation of mRNA and protein are all modelled by linear functions. Based on the above, the ordinary differential equations describing the reaction kinetics are:

$$\begin{aligned} \dot{r_a} &= m_a h^+(p_b; \theta_b, n_b) - \gamma_a r_a, \\ \dot{r_b} &= m_b h^-(p_a; \theta_a, n_a) - \gamma_b r_b, \end{aligned} \tag{9}$$

ACCEPTED MANUSCRIPT

$$\begin{aligned} \dot{p_a} &= k_a r_a - \delta_a p_a, \\ \dot{p_b} &= k_b r_b - \delta_b p_b. \end{aligned} \tag{10}$$

Other quantities in (9), (10) are defined in Table 1.

4.2. The simplified nonlinear model (SNM)

An alternative model can be obtained by assuming that the mRNA dynamics are extremely fast when compared to the protein dynamics and hence reach their equilibrium instantly. Assuming quasi-steady-state mRNA concentrations for the activation-inhibition network of Figure 4, the dynamics can be described by just two variables, say p_a and p_b. Figure 5 shows the network corresponding to the simplified nonlinear model (SNM). More precisely, if we assume that $\dot{r_a} \approx 0$ and $\dot{r_b} \approx 0$ then (9) yields:

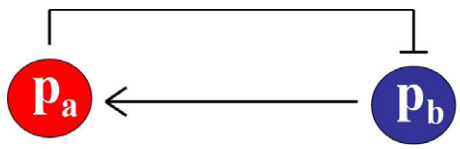


Figure 5: The network for the simplified nonlinear model.

$$\begin{aligned} r_a &= \frac{m_a}{\gamma_a} h^+(p_b; \theta_b, n_b), \\ r_b &= \frac{m_b}{\gamma_b} h^-(p_a; \theta_a, n_a). \end{aligned} \tag{11}$$

Substituting (11) into (10), the equations for the protein concentrations p_a and p_b become:

$$\begin{aligned} \dot{p_a} &= k_a^{'} h^+(p_b; \theta_b, n_b) - \delta_a p_a, \\ \dot{p_b} &= k_b^{'} h^-(p_a; \theta_a, n_a) - \delta_b p_b; \end{aligned} \tag{12}$$

where:

$$k_a^{'} = \frac{m_a}{\gamma_a} k_a \qquad , \; k_b^{'} \; = \frac{m_b}{\gamma_b} k_b. \tag{13}$$

4.3. The complete piecewise linear model (CPWLM) and the simplified piecewise linear model (SPWLM)

If we approximate the transcription stages of the CNM with the PWL functions $s^+(p_i; \theta_i)$ and $s^-(p_i; \theta_i)$ as proposed in [5], then we obtain the equations of the *complete piecewise linear model* (CPWLM), as follows:

$$\begin{aligned} \dot{r_a} &= m_a s^+(p_b; \theta_b) - \gamma_a r_a \\ \dot{r_b} &= m_b s^-(p_a; \theta_a) - \gamma_b r_b \\ \dot{p_a} &= k_a r_a - \delta_a p_a \\ \dot{p_b} &= k_b r_b - \delta_b p_b \end{aligned} \tag{14}$$

To further simplify the CPWLM, we can make the quasi-steady-state mRNA assumption to give the *simplified piecewise linear model* (SPWLM):

$$\begin{aligned} \dot{p_a} &= k_a^{'} s^+(p_b; \theta_b) - \delta_a p_a \\ \dot{p_b} &= k_b^{'} s^-(p_a; \theta_a) - \delta_b p_b \end{aligned} \tag{15}$$

where $k_a^{'}, k_b^{'}$ are given in (13).

Accepted manuscript

ACCEPTED MANUSCRIPT

4.4. Discrete-time model

A different way to model the network is to discretize its dynamics. We show below that the model presented in [7] can be obtained by appropriately sampling the state of the SPWLM presented earlier. Equations (15) can be recast in matrix form as:

$$\dot{\mathbf{p}} = A\mathbf{p} + B\mathbf{u}, \tag{16}$$

where

$$\mathbf{p} = \begin{pmatrix} p_a \\ p_b \end{pmatrix} \ , \ A = \begin{pmatrix} -\delta_a & 0 \\ 0 & -\delta_b \end{pmatrix} \ , \ B = \begin{pmatrix} k_a^{'} & 0 \\ 0 & k_b^{'} \end{pmatrix} \ , \ \mathbf{u} = \begin{pmatrix} s^+(p_b;\theta_b) \\ s^-(p_a;\theta_a) \end{pmatrix}. \tag{17}$$

Integrating (16), we have:

$$\mathbf{p}(t) = e^{At}\mathbf{p}_0 + \Big(e^{At} - \mathbf{I}\Big)A^{-1}B\mathbf{u}, \tag{18}$$

where

$$\begin{pmatrix} p_a(0) \\ p_b(0) \end{pmatrix} = \begin{pmatrix} p_{a_0} \\ p_{b_0} \end{pmatrix}. \tag{19}$$

Over a sufficiently small time step T we have:

$$\mathbf{p}(t+T) = \begin{pmatrix} e^{-\delta_a(t+T)} & 0 \\ 0 & e^{-\delta_b(t+T)} \end{pmatrix} \begin{pmatrix} p_{a_0} \\ p_{b_0} \end{pmatrix} + \begin{pmatrix} -\frac{k_a^{'}}{\delta_a}(e^{-\delta_a(t+T)} - 1) & 0 \\ 0 & -\frac{k_b^{'}}{\delta_b}(e^{-\delta_b(t+T)} - 1) \end{pmatrix} \begin{pmatrix} s^+(p_{b_0};\theta_b) \\ s^-(p_{a_0};\theta_a) \end{pmatrix}. \tag{20}$$

Note that T must be chosen small enough so that the discretized dynamics approximate the continuous dynamics. Typically T must be significantly smaller than the time constants associated to the linear part of the continuous-time model. Hence we take:

$$T = \frac{1}{10}\max\{\frac{1}{\delta_a}, \frac{1}{\delta_b}\}. \tag{21}$$

Then for $t = 0$, we have

$$\begin{aligned} p_a(T) &= e^{-\delta_a T}p_{a_0} + \frac{k_a^{'}}{\delta_a}\Big(1 - e^{-\delta_a T}\Big)s^+(p_{b_0};\theta_b), \\ p_b(T) &= e^{-\delta_b T}p_{b_0} + \frac{k_b^{'}}{\delta_b}\Big(1 - e^{-\delta_b T}\Big)s^-(p_{a_0};\theta_a). \end{aligned} \tag{22}$$

Let us now rescale time such that $T = 1$. Then if we set $p_{a_0} = p_a(n)$ and $p_a(T) = p_a(n+1)$ (similarly $p_{b_0} = p_b(n)$ and $p_b(T) = p_b(n+1)$), then we have the discretized form of equations (15):

$$\begin{aligned} p_a(n+1) &= e^{-\delta_a}p_a(n) + \frac{k_a^{'}}{\delta_a}\Big(1 - e^{-\delta_a}\Big)s^+(p_b(n);\theta_b), \\ p_b(n+1) &= e^{-\delta_b}p_b(n) + \frac{k_b^{'}}{\delta_b}\Big(1 - e^{-\delta_b}\Big)s^-(p_a(n);\theta_a). \end{aligned} \tag{23}$$

For the case $\delta_a = k_a^{'} = \delta_b = k_b^{'}$, equations (23) become

$$\begin{aligned} p_a(n+1) &= \alpha p_a(n) + (1-\alpha)s^+(p_b(n);\theta_b), \\ p_b(n+1) &= \alpha p_b(n) + (1-\alpha)s^-(p_a(n);\theta_a), \end{aligned} \tag{24}$$

where

$$\alpha = e^{-\delta_a} = e^{-\delta_b}. \tag{25}$$

This corresponds to the model given by Coutinho *et al.* in [7]. Parameter α represents the degradation of the gene and is always between $[0, 1]$.

We move now to the analysis of the dynamics predicted by each model.

Accepted manuscript

ACCEPTED MANUSCRIPT

5. Analysis

We will now present a systematic analysis of the models described in the previous section. After deriving the equilibria of both the CNM and SNM, we discuss their stability. We look for the presence of persistent oscillations (limit cycles) by studying the occurrence of Hopf bifurcations. We will show that the presence of this bifurcation phenomenon is dependent on the modelling. We also investigate the effect of varying the Hill coefficients, confirming and extending the results of [48]. Finally we show the effects of taking PWL approximations of the nonlinear Hill kinetics and discuss how discretization introduces spurious dynamics that can lead to incorrect predictions. In what follows, for the sake of simplicity, we assume (with a slight abuse of notation) that $\theta_a \equiv \theta_a^{n_a}$ and $\theta_b \equiv \theta_b^{n_b}$.

5.1. Existence of equilibria

We start with the equilibria of the CNM and SNM. We set:

$$\dot{r_a} = \dot{r_b} = \dot{p_a} = \dot{p_b} = 0, \tag{26}$$

in (9) and (10). We will label $\tilde{r}_a, \tilde{r}_b, \tilde{p}_a, \tilde{p}_b$ as the steady-state values of mRNA and protein concentrations respectively. Note that from (10) we have that:

$$\tilde{r}_a = \frac{\delta_a}{k_a}\tilde{p}_a, \quad \tilde{r}_b = \frac{\delta_b}{k_b}\tilde{p}_b. \tag{27}$$

After some algebraic manipulation, we find that $(\tilde{p}_a, \tilde{p}_b)$ are given by:

$$\theta_b\Big(\sum_{k=0}^{n_b} \tilde{p}_a^{n_a(n_b-k)+1}\binom{n_b}{k}\theta_a^k\Big) + (\tilde{p}_a - \phi_a)(\phi_b\theta_a)^{n_b} = 0, \tag{28}$$

$$\tilde{p}_b = \frac{\phi_b\theta_a}{\theta_a + \tilde{p}_a^{n_a}}, \tag{29}$$

where

$$\phi_a = \frac{m_a k_a}{\gamma_a \delta_a}\ ,\ \phi_b = \frac{m_b k_b}{\gamma_b \delta_b}. \tag{30}$$

Solutions of equation (28) are possible equilibrium concentrations $\tilde{p}_a$ (equilibrium concentrations $\tilde{p}_b$, $\tilde{r}_a$, $\tilde{r}_b$ are then easily obtained, using (26), (29)). Equation (28) is a polynomial of degree $n_a n_b + 1$. Hence, for large Hill coefficients n_a and n_b, it is difficult (or even impossible) to obtain analytical forms of all the allowed equilibrium concentrations.

Recall that the equations for the SNM were derived using the steady-state mRNA assumption, ($\dot{r_a} = \dot{r_b} = 0$). Therefore, the protein equilibrium concentrations for the SNM will also be given by equation (28). In Figure 6 we can see that for a given parameter region both models eventually approach the same fixed point. However trajectories of the SNM approach equilibrium much faster and in a less oscillatory manner than those of the CNM.

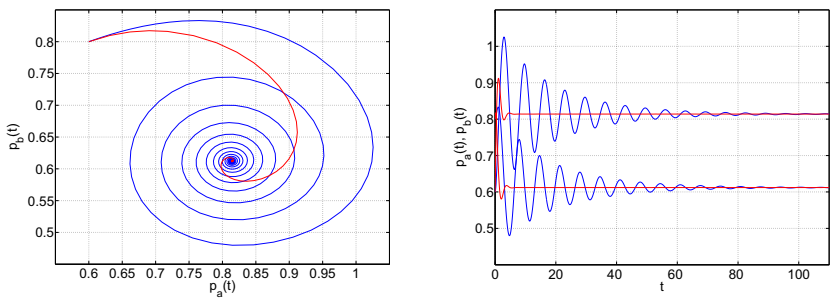
5.2. Stability and bifurcations

We now study the stability of the predicted equilibria of the CNM and the SNM. Let A_{CNM} be the Jacobian of equations (9), (10) of the CNM. For the vector $\mathbf{x} = (x_1, x_2, x_3, x_4)^T = (r_a, r_b, p_a, p_b)^T$, we have

$$A_{CNM} = \Big\{a_{ij} = \frac{\partial \dot{x_i}}{\partial x_j}\Big\} = \begin{pmatrix} -\gamma_a & 0 & 0 & m_a\frac{\partial h^+(p_b;\theta_b,n_b)}{\partial p_b} \\ 0 & -\gamma_b & m_b\frac{\partial h^-(p_a;\theta_a,n_a)}{\partial p_a} & 0 \\ k_a & 0 & -\delta_a & 0 \\ 0 & k_b & 0 & -\delta_b \end{pmatrix} \tag{31}$$

Accepted manuscript

ACCEPTED MANUSCRIPT



(a) (b)

Figure 6: (a) Projection of trajectories in the protein subspace (p_a,p_b) of the CNM (blue) and SNM (red) for the same parameters and initial conditions; $m_a = m_b = 1.8$, $\theta_a = \theta_b = 0.28$, and $k_a = k_b = \gamma_a = \gamma_b = \delta_a = \delta_b = 1$. (b) Time evolution of the protein concentrations p_a and p_b. Blue corresponds to the CNM and red corresponds to the SNM.

The characteristic equation is then [48]

$$(\lambda + \gamma_a)(\lambda + \gamma_b)(\lambda + \delta_a)(\lambda + \delta_b) + D_{CNM} = 0 \tag{32}$$

where

$$D_{CNM} = m_a m_b k_a k_b \theta_a \theta_b \frac{n_a n_b \tilde{p}_a^{(n_a-1)} \tilde{p}_b^{(n_b-1)}}{(\theta_a + \tilde{p}_a^{n_a})^2 (\theta_b + \tilde{p}_b^{n_b})^2}. \tag{33}$$

Solving the characteristic equation (32) for different values of D_{CNM}, we find the different possible dynamical behaviours of our system. Figure 7(a) depicts the four eigenvalues of equation (32) as a function of D_{CNM}. Note that D_{CNM} is always positive (for the sake of completeness, we plot the eigenvalues also for $D_{CNM} < 0$). For a certain value of $D_{CNM} = D_{Hopf}$, the real part of one of the eigenvalues crosses zero, indicating a loss of stability through a Hopf bifurcation. Widder et al. [48] calculated this value explicitly as:

$$D_{Hopf} = \frac{(\gamma_a + \gamma_b)(\gamma_a + \delta_a)(\gamma_a + \delta_b)(\gamma_b + \delta_a)(\gamma_b + \delta_b)(\delta_a + \delta_b)}{(\gamma_a + \gamma_b + \delta_a + \delta_b)^2} \tag{34}$$

for the case when the Hill coefficients n_a, n_b are equal and greater than two[1]. An example of oscillatory behaviour predicted by CNM is shown in Figures 8(a) and 8(b).

We shall now show that, under the mRNA quasi-steady-state assumption, such limit cycles are not possible in the SNM. The corresponding Jacobian matrix A_{SNM} of SNM given by equations (12) is:

$$A_{SNM} = \begin{pmatrix} -\delta_a & k_a' \frac{\partial h^+(p_b;\theta_b,n_b)}{\partial p_b} \\ k_b' \frac{\partial h^-(p_a;\theta_a,n_a)}{\partial p_a} & -\delta_b \end{pmatrix} \tag{35}$$

and the characteristic equation is then

$$(\delta_a + \lambda)(\delta_b + \lambda) + D_{SNM} = 0, \tag{36}$$

where

$$D_{SNM} = \frac{D_{CNM}}{\gamma_a \gamma_b}. \tag{37}$$

[1]The proof is based on the criterion by Lienard-Chipart (see [17], p.221)

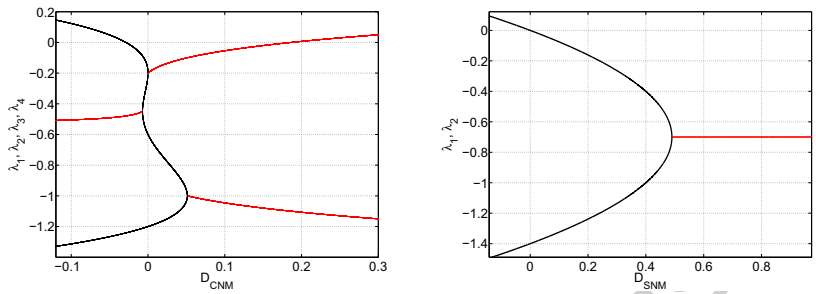
Accepted manuscript

ACCEPTED MANUSCRIPT

Equation (36) is quadratic and so the two eigenvalues $\lambda_{1,2}$ are given by

$$\lambda_{1,2} = \frac{-(\delta_a + \delta_b) \pm \sqrt{(\delta_a + \delta_b)^2 - 4D_{SNM}}}{2}. \tag{38}$$

So for $\lambda_{1,2}$ complex, their real part will be always equal to $-\frac{1}{2}(\delta_a + \delta_b)$. Since the protein degradation rates δ_a, δ_b are biologically meaningful only when they are positive, a Hopf bifurcation will never be possible in the SNM. In Figure 7(b), $\lambda_{1,2}$ are plotted as a function of D_{SNM}. We can see that the real part of the



(a) (b)

Figure 7: Eigenvalues of the Jacobian matrices of the CNM and the SNM, plotted as functions of D_{CNM} or D_{SNM}. Real eigenvalues are drawn in black and the real parts of complex conjugate pairs of eigenvalues are drawn in red.

complex eigenvalues remains constant and negative as a function of D_{SNM}.

Figures 7(a) and 7(b) illustrate an important qualitative difference between the two nonlinear models. The equilibria predicted by the SNM are always stable whereas the same equilibria predicted by the CNM are liable to lose their stability under parameter variation. The mRNA quasi-steady-state assumption results in an over-simplification of the dynamics, with the loss of the Hopf bifurcation. For example, Figures 8(c) and 8(d) for the SNM, with the same parameters as Figures 8(a) and 8(b) for the CNM, have only stable equilibria.

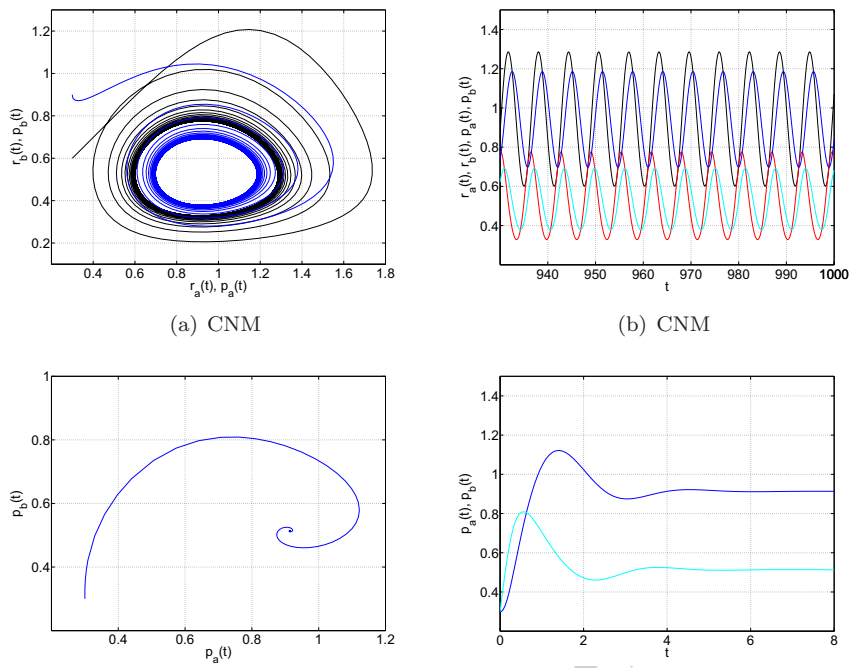
5.3. The quasi-steady-state mRNA assumption

If the mRNA concentrations reach their steady state values on a time scale much quicker than the concentrations of the proteins, then we are able to make the quasi-steady-state mRNA assumption. For the system to behave in this way, any transients in the mRNA concentrations have to be damped out quickly. In other words, the two eigenvalues associated with the mRNA subspace of the four dimensional CNM state space have to be in the left-hand side of the complex plane and have much larger real parts in modulus than the two eigenvalues associated with protein subspace.

The four eigenvalues of the CNM state space are given by the roots of equation (32). Whilst an exact solution of this equation is unwieldy, we can see that in the case of $D_{CNM} = 0$, the four eigenvalues are given exactly by $\lambda_{1,2,3,4} = -\gamma_a, -\gamma_b, -\delta_a, -\delta_b$. Similarly for D_{CNM} small, these values are approximately correct. So it is natural to think of $\gamma_a^{-1}, \gamma_b^{-1}$ as time scales for the mRNA subspace and $\delta_a^{-1}, \delta_b^{-1}$ as time scales for the protein subspace. Now let γ_a/δ_a and γ_b/δ_b be the ratios of the time scales between the mRNA and protein dynamics for gene a and b respectively.

In order to make the quasi-steady-state mRNA assumption, we need to take both ratios large. This is in line with our intuition since in this case the damping ($\gamma_{a,b}$) associated with the mRNA dynamics is much greater than the damping ($\delta_{a,b}$) associated with the protein dynamics, which in turn means that the mRNA transients die out more quickly.

ACCEPTED MANUSCRIPT



(a) CNM (b) CNM

(c) SNM (d) SNM

Figure 8: Plots for the CNM and SNM of the network of activation-inhibition of Figure 4. For both models, the corresponding parameters have the same values; $m_a = m_b = 2.35$, $\theta_a = \theta_b = 0.21$, $n_a = n_b = 3$ and $k_a = k_b = \delta_a = \delta_b = \gamma_a = \gamma_b = 1$. For the left hand pictures, the projection of the trajectory onto the mRNA subspace, $(r_a(t), r_b(t))$, is shown in black and the projection onto the protein subspace, $(p_a(t), p_b(t))$ in blue. For the right hand pictures, black denotes $r_a(t)$, red $r_b(t)$, blue $p_a(t)$ and turquoise $p_b(t)$.

To gain a deeper insight, following [11], we now consider the CNM equations in the form:

$$\begin{aligned} \dot{r_a} &= m_a h^+(p_b; \theta_b, n_b) - \frac{\gamma_a}{\varepsilon} r_a, \\ \dot{r_b} &= m_b h^-(p_a; \theta_a, n_a) - \frac{\gamma_b}{\varepsilon} r_b, \\ \dot{p_a} &= \frac{k_a}{\varepsilon} r_a - \delta_a p_a, \\ \dot{p_b} &= \frac{k_b}{\varepsilon} r_b - \delta_b p_b. \end{aligned} \tag{39}$$

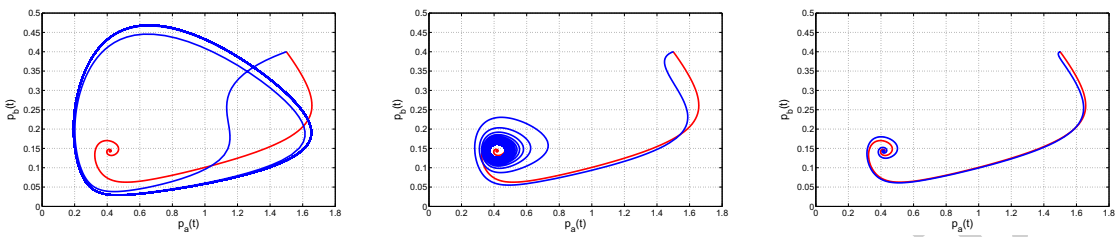
For $\varepsilon = 1$, equations (39) are the exact equations of the CNM. Looking at the above equations, we can see that the time constants for the mRNA concentrations r_a, r_b are $\tau_{r_a} = \frac{\varepsilon}{\gamma_a}, \tau_{r_b} = \frac{\varepsilon}{\gamma_b}$ respectively. Also, for the protein dynamics of p_a, p_b the time constants are $\tau_{p_a} = 1/\delta_a, \tau_{p_b} = 1/\delta_b$. Therefore, the ratio between the time scales of mRNA dynamics and protein dynamics will be given by (for example for gene a):

$$\frac{\tau_{r_a}}{\tau_{p_a}} = \varepsilon \frac{\delta_a}{\gamma_a}. \tag{40}$$

As depicted in Figure 9, the predictions of the SNM and the CNM become significantly different as the time scales between the mRNA and protein dynamics are varied. Specifically, when the two time scales are comparable, the SNM just predicts the average concentrations of the oscillations given by the CNM. However, as the separation of time scales between mRNAs and proteins becomes larger (ε large), then

ACCEPTED MANUSCRIPT

the amplitude of the oscillations predicted by the CNM gets smaller and smaller. In that case although qualitatively the SNM still predicts a stable equilibrium, quantitatively it will be very close to the predictions of the CNM. For example, Figure 9c shows that if the mRNA degradation rate is 50 times faster than the degradation rate of the proteins, then the SNM will be able to give very similar quantitative predictions to the CNM.



(a) $\varepsilon = 1$ (b) $\varepsilon = 10$ (c) $\varepsilon = 50$

Figure 9: Effects of the different time scales between the mRNA dynamics and protein dynamics. Parameter ε regulates this time scales ratio. CNM predictions are plotted in blue colour and SNM predictions in red. Parameter values are: $\gamma_a = \gamma_b = \delta_a = \delta_b = k_a = k_b = 1$, $\theta_a = 0.21, \theta_b = 0.21, m_a = 2.35, m_b = 2.35$, $n_a = n_b = 4$.

To show that the Hopf bifurcation disappears in the SNM, we now recast system (39) as a *slow-fast* system by setting $\tilde{r}_{a,b} = r_{a,b}/\epsilon$. Dropping the tildes, we have

$$\begin{aligned}
\frac{1}{\epsilon}\dot{r_a} &= m_a h^+(p_b;\theta_b,n_b) - \gamma_a r_a,\\
\frac{1}{\epsilon}\dot{r_b} &= m_b h^-(p_a;\theta_a,n_a) - \gamma_b r_b,\\
\dot{p_a} &= k_a r_a - \delta_a p_a,\\
\dot{p_b} &= k_b r_b - \delta_b p_b.
\end{aligned} \tag{41}$$

For $\epsilon = 1$, equations (41) are equations (9), (10) for the CNM. The limiting case $\epsilon \to \infty$ corresponds to the quasi steady-state mRNA assumption, since:

$$\lim_{\epsilon\to\infty} \frac{1}{\epsilon}\dot{r_a} = \lim_{\epsilon\to\infty} \frac{1}{\epsilon}\dot{r_b} = 0. \tag{42}$$

We want to study how the stability of equilibrium solutions to equations (41) varies in the limit $\epsilon \to \infty$. The Jacobian, $A_{SF}(\epsilon)$, derived from the slow-fast model (41) is:

$$A_{SF}(\epsilon) = \begin{pmatrix} -\epsilon\gamma_a & 0 & 0 & \epsilon\frac{m_a \partial h^+(p_b;\theta_b,n_b)}{\partial p_b} \\ 0 & -\epsilon\gamma_b & \epsilon\frac{m_b \partial h^-(p_a;\theta_a,n_a)}{\partial p_a} & 0 \\ k_a & 0 & -\delta_a & 0 \\ 0 & k_b & 0 & -\delta_b \end{pmatrix}. \tag{43}$$

and the characteristic equation is

$$(\lambda + \epsilon\gamma_a)(\lambda + \epsilon\gamma_b)(\lambda + \delta_a)(\lambda + \delta_b) + D_{SF}(\epsilon) = 0 \tag{44}$$

where

$$D_{SF}(\epsilon) = \epsilon^2 D_{CNM} = \epsilon^2 \gamma_a \gamma_b D_{SNM} \tag{45}$$

with D_{CNM} being given by (33) and having used equation (37).

Accepted manuscript

ACCEPTED MANUSCRIPT

Dividing (44) throughout by ϵ^2 gives

$$(\frac{\lambda}{\epsilon}+\gamma_a)(\frac{\lambda}{\epsilon}+\gamma_b)(\lambda+\delta_a)(\lambda+\delta_b)+\gamma_a\gamma_b D_{SNM}=0. \tag{46}$$

In the limit $\epsilon \to \infty$, equation (46) becomes

$$(\lambda+\delta_a)(\lambda+\delta_b)+D_{SNM}=0, \tag{47}$$

which is precisely the characteristic equation (36) of the SNM.

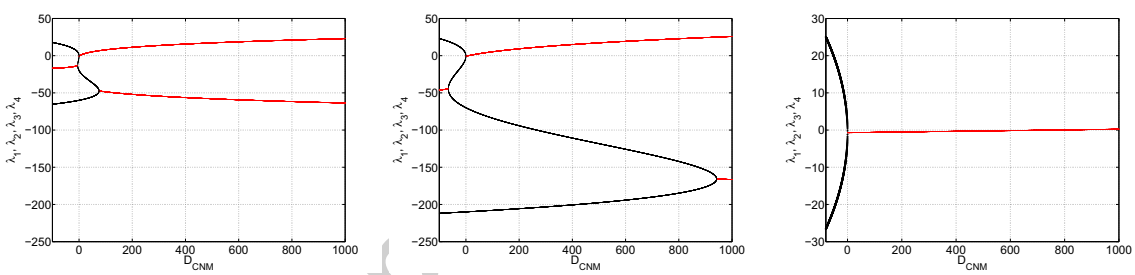
We can find the analytical form of $D_{Hopf}(\epsilon)$ as a function of ϵ, where $D_{Hopf}(\epsilon)$ is the value of $D_{SF}(\epsilon)$ at which the system (41) can undergo a Hopf bifurcation. After a lengthy calculation, we find

$$D_{Hopf}(\epsilon)=\frac{(\epsilon\gamma_a+\epsilon\gamma_b)(\epsilon\gamma_a+\delta_a)(\epsilon\gamma_a+\delta_b)(\epsilon\gamma_b+\delta_a)(\epsilon\gamma_b+\delta_b)(\delta_a+\delta_b)}{(\epsilon\gamma_a+\epsilon\gamma_b+\delta_a+\delta_b)^2}. \tag{48}$$

In the limit as $\epsilon \to \infty$, we have

$$\lim_{\epsilon\to\infty} D_{Hopf}(\epsilon)=\lim_{\epsilon\to\infty}\frac{\epsilon^3\gamma_a^2\gamma_b^2(\delta_a+\delta_b)}{(\gamma_a+\gamma_b)}=\infty. \tag{49}$$

Hence we can see that the SNM never undergoes a Hopf bifurcation and so will be unable to sustain any form of limit cycle.



(a) ϵ=100 (b) ϵ=350 (c) ϵ=30000

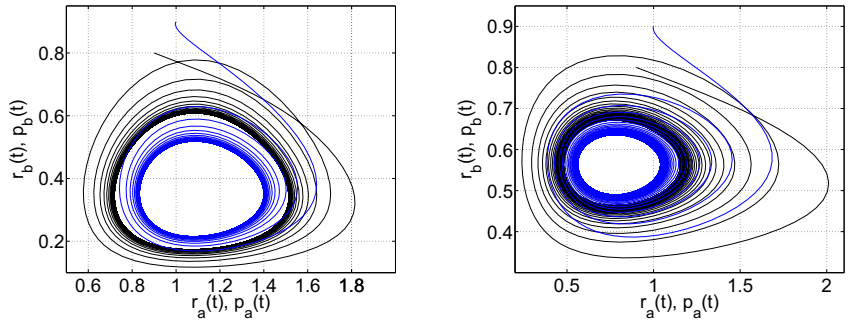
Figure 10: Eigenvalues of the Jacobian for the slow-fast model (41), plotted as functions of D_{CNM} for various values of ϵ. Real eigenvalues are shown in black and the real parts of complex conjugate pairs of eigenvalues in red.

In Figure 10, we plot the eigenvalues of the slow-fast system (41) as a function of ϵ. We note that as ϵ increases, two of the eigenvalues become large and negative, indicating that their transients quickly die away and so justifying the quasi-steady-state mRNA assumption. Note that for ϵ=30000, we are only able to plot two of the four eigenvalues, since the other two have extremely large and negative values.

5.4. Variation of the Hill coefficients

As we saw in the previous section, the SNM can never support a limit cycle for the activation-inhibition network, no matter what the value of the Hill coefficients, n_a and n_b. However, for the CNM, a Hopf bifurcation is possible depending on the value of the system parameters. Widder et al. [48] have shown that if the two Hill coefficients are equal, namely $n_a = n_b = n$, then the two-node activation-inhibition network can undergo a Hopf bifurcation for $n > 2$. In this section, we will extend this result to the case when $n_a \neq n_b$ for the activation-inhibition network, a general situation closer to biological applications. Additionally, we will consider non-integer values for the Hill coefficients since this reflects the fact that such a function might fit well the experimental data [28].

ACCEPTED MANUSCRIPT



(a) $n_a = 4$, $n_b = 2$ (b) $n_a = 1$, $n_b = 6$

Figure 11: Limit cycles of the CNM for different pairs of Hill coefficients (n_a, n_b). The other parameters of the system are $m_a = m_b = x$, $\theta_a = \theta_b = \theta$, where $\theta = 0.1786$ in plot a) and $\theta = 0.1282$ in plot b), $k_a = k_b = \delta_a = \delta_b = \gamma_a = \gamma_b = 1$. The projection of the trajectory onto the mRNA subspace, $(r_a(t), r_b(t))$, is shown in black and onto the protein subspace, $(p_a(t), p_b(t))$, in blue.

Following Widder et al [48], we consider a Hopf bifurcation along the one-dimensional manifold defined by

$$m_a = \chi_a s \ , \ m_b = \chi_b s \ , \ \theta_a = \frac{\lambda_a}{s}, \ \theta_b = \frac{\lambda_b}{s}. \tag{50}$$

From equations (50) and (30) it follows that $\phi_i = v_i s$, for $i = a, b$ where

$$v_a = \frac{k_a}{\gamma_a \delta_a}\chi_a \ , \ v_b = \frac{k_b}{\gamma_b \delta_b}\chi_b. \tag{51}$$

Substituting equations (50) into (28) and expanding we have

$$\begin{aligned}&\lambda_b\Big(\tilde{p}_a^{n_a n_b+1} + n_b\frac{\lambda_a}{s}\tilde{p}_a^{n_a(n_b-1)+1} + ... + n_b\Big(\frac{\lambda_a}{s}\Big)^{n_b-1}\tilde{p}_a^{n_a+1} + \Big(\frac{\lambda_a}{s}\Big)^{n_b}\tilde{p}_a\Big)\\ &+ (\tilde{p}_a - v_a s)(v_b\lambda_a)^{n_b} s = 0\end{aligned} \tag{52}$$

We want to take the limit of this expression when the auxiliary variable $s \gg 1$. So we consider the equilibrium point $\tilde{p}_a$ to be of the form

$$\tilde{p}_a = a_1 s^m + O(s^{m-1}). \tag{53}$$

Substituting (53) into equation (52) and neglecting high-order terms $O(s^{m-1})$, it can be shown that, for very large s,

$$\lambda_b \tilde{p}_a^{n_a n_b+1} s^{n_b} - v_a(v_b\lambda_a)^{n_b} s^{n_b+2} = 0 \tag{54}$$

and hence

$$\tilde{p}_a = a_1 s^{\frac{2}{n_a n_b+1}}, \tag{55}$$

where

$$a_1 = \Big(\frac{v_a(v_b\lambda_a)^{n_b}}{\lambda_b}\Big)^{\frac{1}{n_a n_b+1}}. \tag{56}$$

We find $\tilde{p}_b$ from equation (29):

$$\tilde{p}_b = a_2 s^{-\frac{2n_a}{n_a n_b+1}}, \tag{57}$$

where

$$a_2 = \Big(\frac{\lambda_b(v_b\lambda_a)^{\frac{1}{n_a}}}{v_a}\Big)^{\frac{n_a}{n_a n_b+1}}. \tag{58}$$

Accepted manuscript

ACCEPTED MANUSCRIPT

We now substitute the expressions for $\tilde{p}_a$ and $\tilde{p}_b$, given by (55) and (57) respectively, into equation (33) in order to obtain D_{lim}, the limit of D_{CNM} for s very large, to find

$$D_{lim} = n_a n_b \gamma_a \gamma_b \delta_a \delta_b. \tag{59}$$

This must be compared with the value D_{Hopf} which is required for a Hopf bifurcation. As in [48], we consider the function:

$$\begin{aligned} H(\gamma_a, \gamma_b, \delta_a, \delta_b, n_a, n_b) &= \frac{D_{lim}}{D_{Hopf}} \\ &= \frac{n_a n_b \gamma_a \gamma_b \delta_a \delta_b (\gamma_a + \gamma_b + \delta_a + \delta_b)^2}{(\gamma_a + \gamma_b)(\gamma_a + \delta_a)(\gamma_a + \delta_b)(\gamma_b + \delta_a)(\gamma_b + \delta_b)(\delta_a + \delta_b)} \end{aligned} \tag{60}$$

A value of $H > 1$ indicates that a limit cycle exists for sufficiently large values of s, for the activation-inhibition network. The maximum of H is computed by partial differentiation with respect to the degradation rate constants $\gamma_a, \gamma_b, \delta_a, \delta_b$. Because (60) is symmetric with respect to all four degradation parameters, all four partial derivatives will have identical analytical expressions. For example ([48]):

$$\begin{aligned} &\left(\frac{\partial H}{\partial \gamma_a}\right) = 0 \Rightarrow \\ &(\gamma_a)^3(\gamma_b + \delta_a + \delta_b) + (\gamma_a)^2\Big((\gamma_b)^2 + (\delta_a)^2 + (\delta_b)^2\Big) - 3\gamma_a(\gamma_b \delta_a \delta_b) - \gamma_b \delta_a \delta_b (\gamma_b + \delta_a + \delta_b) = 0. \end{aligned} \tag{61}$$

If we consider $\gamma_b = \delta_a = \delta_b = \gamma$ then it can be shown that equation (61) will give $\gamma_a = \gamma$ [48] . Hence, from (60) we have that:

$$H(\gamma, \gamma, \gamma, \gamma, n_a, n_b) = \frac{n_a n_b}{4}. \tag{62}$$

Through numerical simulations it was observed that other values of the degradation parameters lead to smaller values of the maximum of the function as observed in [48] for $n_a = n_b$. Hence systems with $n_a n_b > 4$ can exhibit oscillatory behaviour in certain regions of parameter space. In Figure 12 we plot regions in

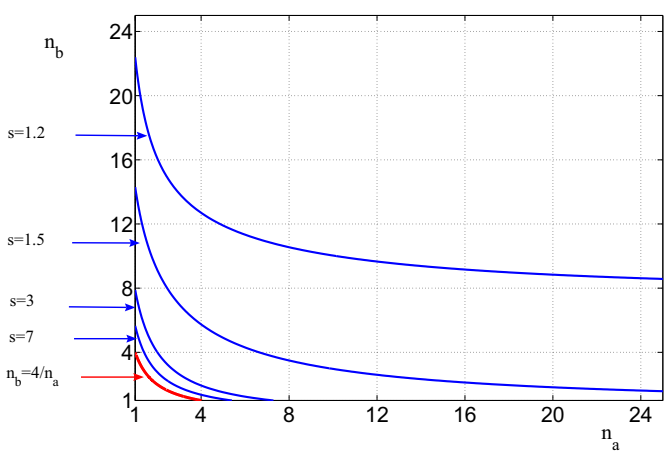


Figure 12: Each blue line represents critical pairs of Hill coefficients (n_a, n_b) for which a Hopf bifurcation can occur for a certain value of the auxiliary variable s. Each of these lines was obtained using numerical continuation. The red line is the hyperbola $n_a n_b = 4$. Parameter values: $m_a = m_b = s$, $\theta_a^{n_a} = \theta_b^{n_b} = \frac{1}{s}$, $k_a = k_b = \gamma_a = \gamma_b = \delta_a = \delta_b = 1$.

parameter space (n_a, n_b) which have oscillatory behaviour.

ACCEPTED MANUSCRIPT

6. Approximating the transcription Hill function

6.1. From the CPWLM to the SPWLM

When the Hill function of transcription in the CNM is approximated by means of piecewise linear (PWL) functions, we obtain the CPWLM. An important issue is to establish how this can be simplified, by means of the quasi-steady-state mRNA assumption, to give the SPWLM. To understand the advantages and limitations of this approach, let us rewrite equations (14) of the CPWLM in matrix form:

$$\dot{\mathbf{x}} = R\mathbf{x} + S\mathbf{u} \tag{63}$$

where

$$\mathbf{x} = \begin{pmatrix} r_a \\ r_b \\ p_a \\ p_b \end{pmatrix} \ , R = \begin{pmatrix} -\gamma_a & 0 & 0 & 0 \\ 0 & -\gamma_b & 0 & 0 \\ k_a & 0 & -\delta_a & 0 \\ 0 & k_b & 0 & -\delta_b \end{pmatrix} \ , S = \begin{pmatrix} m_a & 0 & 0 & 0 \\ 0 & m_b & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \ , \mathbf{u} = \begin{pmatrix} s^+(p_b;\theta_b) \\ s^-(p_a;\theta_a) \\ 0 \\ 0 \end{pmatrix}. \tag{64}$$

Matrix R is lower-triangular and so its eigenvalues lie along the diagonal:

$$\lambda_1 = -\gamma_a \ , \lambda_2 = -\gamma_b \ , \lambda_3 = -\delta_a \ , \lambda_4 = -\delta_b. \tag{65}$$

The eigenvalues corresponding to gene a are λ_1 and λ_3. The SPWLM will be biologically valid if the ratio of the two eigenvalues, say $\rho_a = \frac{\lambda_1}{\lambda_3} = \frac{\gamma_a}{\delta_b}$ is large enough.

Hence, if the degradation of mRNA is sufficiently faster (say, at least ten times) than the degradation of the corresponding protein, then the stationarity approximation is biologically justified. Note that, under certain conditions, it can be assumed that the stationarity approximation is only valid for some genes of the network. In that case, only those genes for which the assumption is valid can be modelled by a single equation for the protein concentration, while the rest will be associated to two equations, one for the mRNA and one for protein concentration.

6.2. Dynamics of the SPWLM

We have already seen that the SNM has qualitatively different dynamics from the CNM. It is therefore reasonable to expect that the use of the PWL approximation does not change this conclusion.

In [36] Plahte et al described two problems when the PWL step function are being used to model gene regulatory networks. One problem is to define a continuous solution across the threshold hyperplanes when the model includes self-regulation. The second is to prove that the solution of equations with step functions is close to the solution of the same equations with steep (large Hill coefficients) sigmoid functions.

As shown in [25] and [14] , under certain conditions the SPWLM cannot predict the existence of limit cycles contained in the CNM. [2]

The equations of the SPWLM are given in (15) [3]. These equations split the (p_a, p_b) state space into four subregions, given by

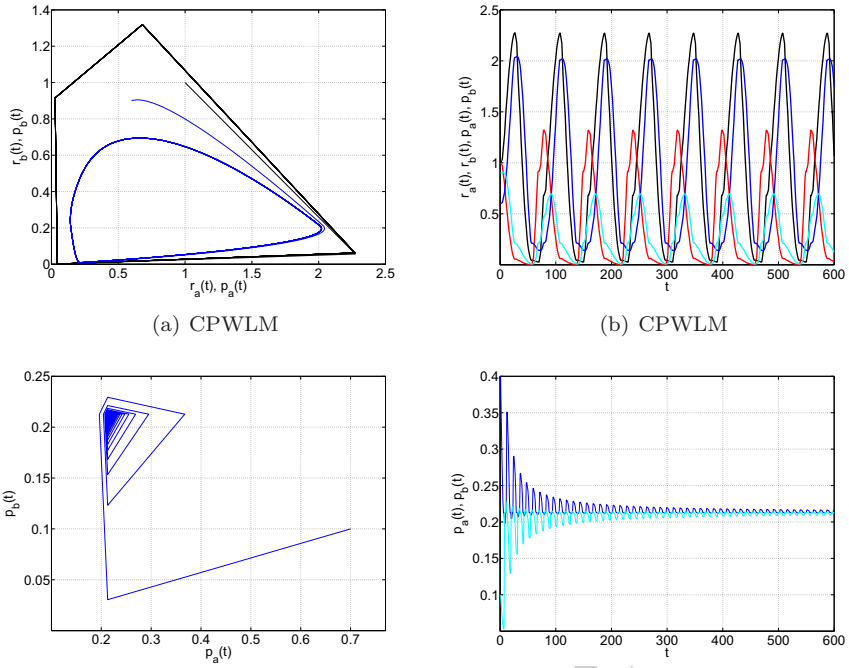
$$\begin{array}{llll} 1. & 0 < p_a < \theta_a & \text{and } p_b > \theta_b & \text{(subregion I)} \\ 2. & p_a > \theta_a & \text{and } p_b > \theta_b & \text{(subregion II)} \\ 3. & p_a > \theta_a & \text{and } 0 < p_b < \theta_b & \text{(subregion III)} \\ 4. & 0 < p_a < \theta_a & \text{and } 0 < p_b < \theta_b & \text{(subregion IV)} \end{array} \tag{66}$$

[2] In [25], it was shown that the simplified PWL model of the activation-inhibition network converges toward a unique stable equilibrium point. In the case of networks with three genes or more with a negative feedback loop, it was shown that the SPWLM predicts a unique stable periodic orbit [25]). Their proof is an extension of a theorem presented by Snoussi et al in [41].

[3] For convenience in this section, we set $k_a^{'} = k_a$ and $k_b^{'} = k_b$

Accepted manuscript

ACCEPTED MANUSCRIPT



(a) CPWLM (b) CPWLM

(c) SPWLM (d) SPWLM

Figure 13: The two PWL models of the activation-inhibition network of Figure 4. For all the plots, the parameters have the same values as in Figure 8. For the left hand plots, the projection of the trajectory onto the mRNA subspace, $(r_a(t), r_b(t))$, is shown in blue and onto the protein subspace, $(p_a(t), p_b(t))$, in turqoise. For the right hand plots, blue refers to $r_a(t)$, red to $r_b(t)$, turquoise to $p_a(t)$ and yellow to $p_b(t)$.

Each subregion has different governing equations and hence different equilibria. The value of the equilibria $(\tilde{p}_a, \tilde{p}_b)$ in each subregion is given by

$$\begin{aligned} \text{I:} \quad & (\frac{k_a}{\delta_a}, \frac{k_b}{\delta_b}) \\ \text{II:} \quad & (\frac{k_a}{\delta_a}, 0) \\ \text{III:} \quad & (0, 0) \\ \text{IV:} \quad & (0, \frac{k_b}{\delta_b}) \end{aligned} \tag{67}$$

Immediately it can be seen that the equilibria for subregions II and III do not lie in their respective subregions. They are *inaccessible* from those subregions. Other equilibria will be accessible or inaccessible, depending on parameter values.

The key to understanding the dynamics of the SPWLM is that, depending on system parameters, at most one subregion equilibrium is accessible. In these cases the system tends to this equilibrium and no limit cycle is possible. But, in certain regions of parameter space, all four subregion equilibria are inaccessible. However, a limit cycle is not possible in this case either ([14, 25]) and the system tends to a *pseudo-equilibrium*, a point that is not an equilibrium of any subregion.

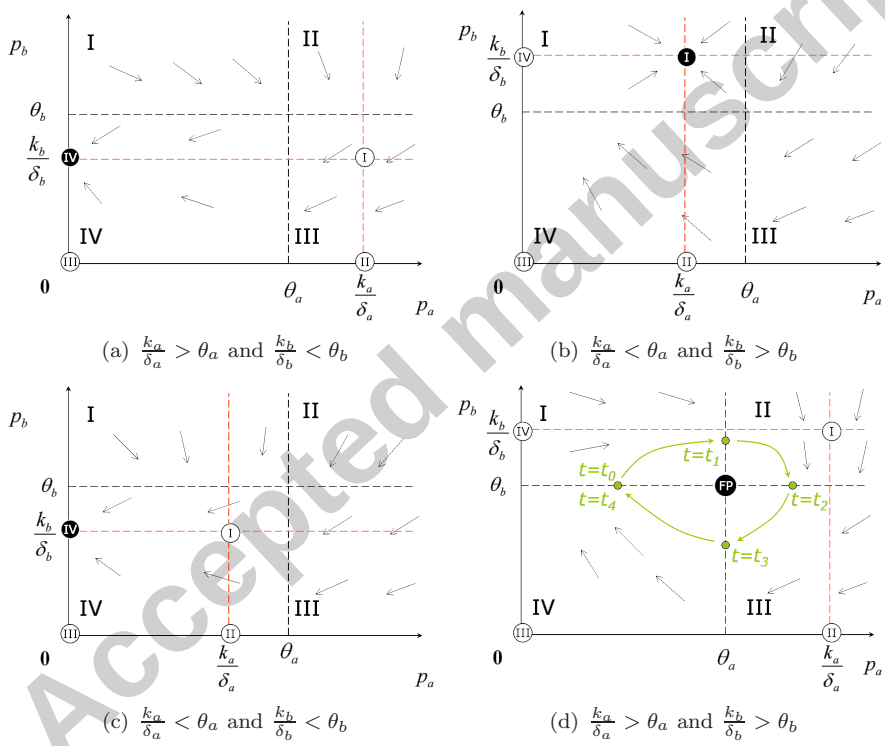
A simple analysis shows that, depending on the parameters k_a, k_b, δ_a, and δ_b, we have four different cases

ACCEPTED MANUSCRIPT

to consider, namely

$$
\begin{aligned}
&1. \quad \frac{k_a}{\delta_a} > \theta_a \text{ and } \frac{k_b}{\delta_b} < \theta_b, \\
&2. \quad \frac{k_a}{\delta_a} < \theta_a \text{ and } \frac{k_b}{\delta_b} > \theta_b, \\
&3. \quad \frac{k_a}{\delta_a} < \theta_a \text{ and } \frac{k_b}{\delta_b} < \theta_b, \\
&4. \quad \frac{k_a}{\delta_a} > \theta_a \text{ and } \frac{k_b}{\delta_b} > \theta_b.
\end{aligned}
\tag{68}
$$

The first three cases, shown in Figure 14, are very similar. No matter what set of initial conditions are chosen, the solution trajectory always arrives at the accessible equilibrium (the arrows in each of these Figures denote sample trajectories). All four subregion equilibria are shown in each Figure, with open circles denoting inaccessible equilibria and closed circles accessible equilibria. In Figures 14(a) and 14(c),



(a) $\frac{k_a}{\delta_a} > \theta_a$ and $\frac{k_b}{\delta_b} < \theta_b$ (b) $\frac{k_a}{\delta_a} < \theta_a$ and $\frac{k_b}{\delta_b} > \theta_b$

(c) $\frac{k_a}{\delta_a} < \theta_a$ and $\frac{k_b}{\delta_b} < \theta_b$ (d) $\frac{k_a}{\delta_a} > \theta_a$ and $\frac{k_b}{\delta_b} > \theta_b$

Figure 14: Vector fields for each of the different cases described in (68). The arrows in each of these subfigures denote sample trajectories. In each figure, the equilibrium of each of the four subregions is denoted by the number of that region inside a circle. Open circles denote inaccessible equilibria and closed circles accessible equilibria. FP stands for *focus point*

.

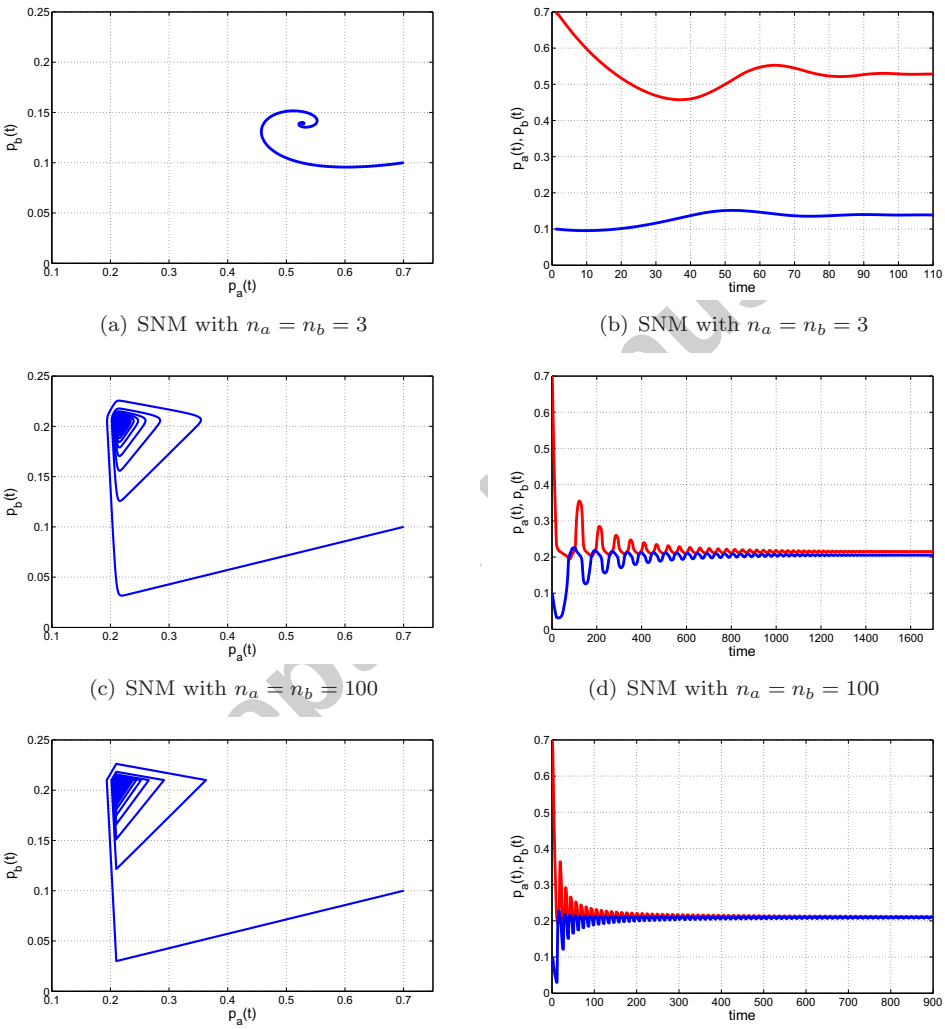
the accessible equilibrium is $(0, \frac{k_b}{\delta_b})$ and in Figure 14(b) it is $(\frac{k_a}{\delta_a}, \frac{k_b}{\delta_b})$. The fourth case is shown in Figure 14(d), together with a possible limit cycle. This limit cycle can only trivially exist. Namely, the only possible periodic solution are the trivial $p_a = \theta_a$, $p_b = \theta_b$, what we call a focus-like point. Starting from any initial condition, the system eventually converges to this point. Due to the nature of the vector field, all trajectories are forced to follow the sequence of switching regions, $I \to II \to III \to IV \to I$, which finally leads the trajectories to the intersection of the two thresholds θ_a, θ_b. For example, for the parameter values considered in Figure 13(d) we have $\frac{k_a}{\delta_a} = \frac{k_b}{\delta_b} = 2.35$ and $\theta_a = \theta_b = 0.21$, which is case 3 in 68 illustrated in Figure 14(d). As shown in Figure 13(d), the system eventually approaches the focus-like point (θ_a, θ_b). The

Accepted manuscript

ACCEPTED MANUSCRIPT

stability of equilibria in switching domains, was presented in [5], where the work of Gouze et al [24] and de Jong et al [10] was extended using the framework of differential inclusions and Filippov solutions.

Similarly to the SNM, the SPWLM does not predict oscillations for the activation-inhibition network. Therefore, the mRNA quasi-steady-state assumption has also important effects when the SPWLM is derived from the CPWLM. Additionally, comparing the two simplified models, namely SNM and SPWLM, we see a consistent difference caused by the approximation of the continuous Hill functions by the step functions. In particular, the SPWLM for some range of parameters, predicts a focus-like equilibrium point which corresponds to trajectories not always close to those of the SNM.



(a) SNM with $n_a = n_b = 3$

(b) SNM with $n_a = n_b = 3$

(c) SNM with $n_a = n_b = 100$

(d) SNM with $n_a = n_b = 100$

(e) SPWLM

(f) SPWLM

Figure 15: Effects of taking the limit of the Hill coefficients n_a, n_b to infinity. Other parameter values are: $\gamma_a = \gamma_b = \delta_a = \delta_b = k_a = k_b = 1$, $\theta_a = 0.21, \theta_b = 0.21, m_a = 2.35, m_b = 2.35$. Protein $p_a(t)$ is shown in red and protein $p_b(t)$ in blue. We can see that the SNM with very large Hill coefficients $n_a = n_b = 100$ (c,d), behaves similarly with the SPWLM (e,f). However, the difference between the predictions of SNM with small Hill coefficients $n_a = n_b = 3$ (a,b), is different than the predictions of SPWLM (e,f).

As shown in Figure 15, the behaviour of the SPWLM and the SNM converge in the limit of large Hill coefficients. It is only when such coefficients are sufficiently high that the SNM solutions behave as the focus-

ACCEPTED MANUSCRIPT

like equilibrium point in the SPWLM. For example, for $n_a = n_b = 100$ (Figure 15c,d), the predictions of the SNM are extremely close to those of the SPWLM (Figure 15e,f) yet they differ when $n_a = n_b = 3$ (Figure 15a,b). Specifically, Figures 15a,b and Figures 15e,f, show that the SNM for Hill coefficients $n_a = n_b = 3$ predicts protein equilibrium values with protein p_a up to 5 times larger than protein p_b while, in the SPWLM, they both converge towards the same value at the expression thresholds θ_a, θ_b. Hence, the PWL models will give sufficient quantitative predictions when compared to the nonlinear models only if the Hill function describing the transcription has large Hill coefficients.

To quantify the mismatch between the predictions of the two models, we plot in Figure 16, the trajectories of the SNM and those of the SPWLM as the Hill coefficients increase and the relative percentage differences between their predicted values. In Figure 16a,b, we see that, for some parameter values, the predictions of the SPWLM are close to those of the SNM when the Hill coefficients $n_a = n_b = \tilde{n} > 20$. For different sets of parameters, the threshold can become significantly lower as shown in Figure 16c,d, where $\tilde{n} \approx 6$.

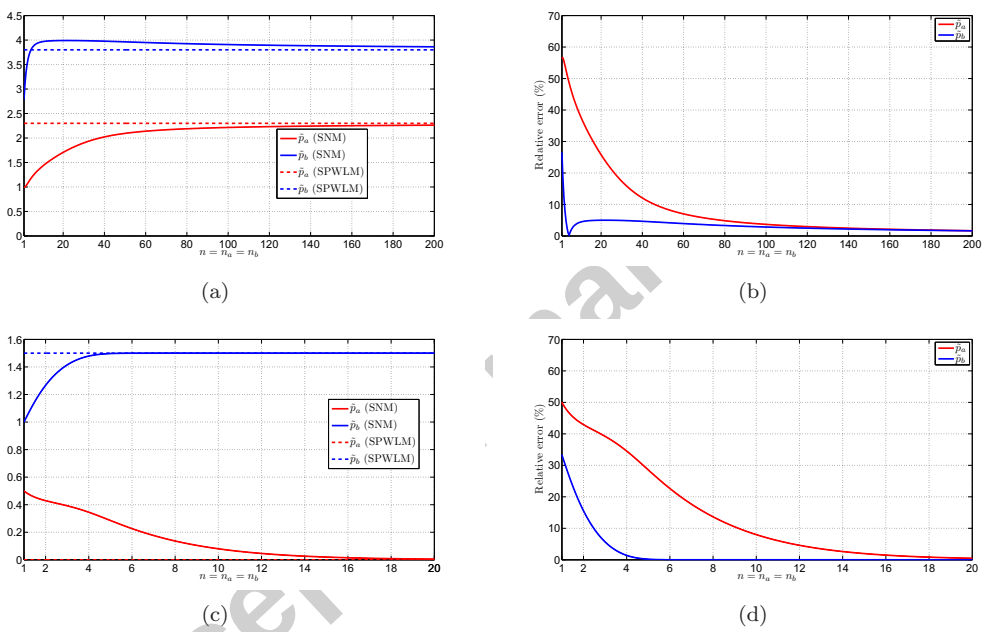


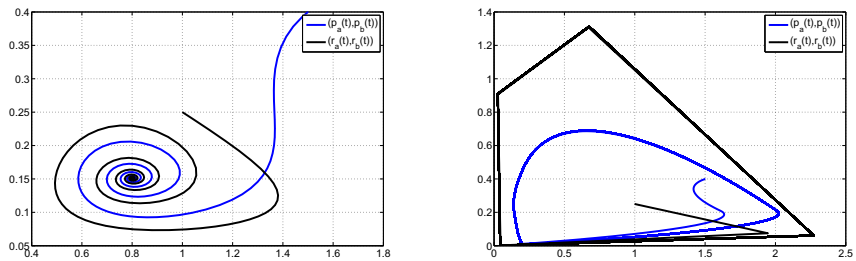
Figure 16: Comparing the predictions of the SPWLM with the predictions of the SNM as the Hill coefficients increase. The left plots show the equilibrium protein concentrations $\tilde{p}_a$ (red) and $\tilde{p}_b$ (blue) for both models (SPWLM predictions as dotted lines and SNM predictions as solid lines). The right plots give the relative error between the two models. a,b) Other parameter values are: $\gamma_a = \gamma_b = \delta_a = \delta_b = k_a = k_b = 1$, $\theta_a = 2.3, \theta_b = 3.8, m_a = 2.35, m_b = 4$. c,d) Other parameter values are: $\gamma_a = \gamma_b = \delta_a = \delta_b = k_a = k_b = 1$, $\theta_a = 1, \theta_b = 2, m_a = 1.5, m_b = 1.5$.

Finally, notice that quantitative differences also occur between the CNM and the CPWLM. An additional difference here, is that for some parameter regions, we have also qualitative differences. There are regions of parameter space where the CNM predicts a stable equilibrium, whereas the CPWLM predicts oscillations. This is expected, because as we showed in section 5.4, a Hopf bifurcation is possible for our network if $n_a \cdot n_b \geq 4$ (a condition which is effectively always satisfied by the CPWLM). The difference between CNM and CPWLM is illustrated in Figure 17.

Also, in contrast with the CNM and SNM, the predictions of the CPWLM do not approach the SPWLM as the separation of time scales between the mRNA dynamics and protein dynamics becomes larger (ϵ large). In Figure 18, we plot numerical simulations of the CPWLM and CNM for $\epsilon = 10$ together with numerical simulations of the SPWLM. The other parameters are the same for all the three models. One can see the significant qualitative difference between the CPWLM and the other two models. The reason is

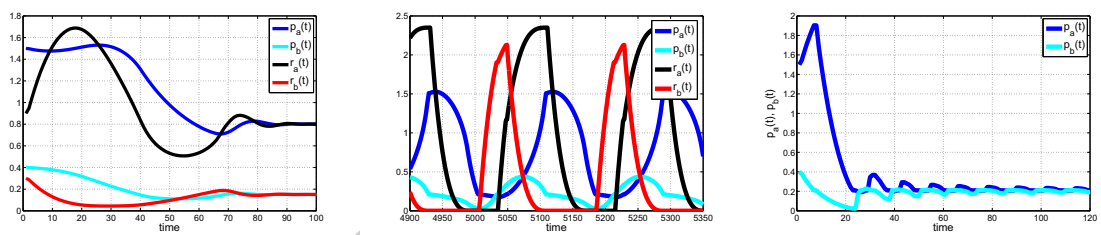
ACCEPTED MANUSCRIPT

the occurrence of sliding motion [4]. Figure 18(b) illustrates how sliding motion is present in the CPWLM. Namely the mRNA concentrations $r_a(t), r_b(t)$ are sliding. The CPWLM predicts oscillations where the CNM and SPWLM both predict a stable equilibrium (with the difference that the SPWLM predicts the focus-like equilibrium)).



(a) CNM (b) CPWLM

Figure 17: Comparison between the complete models with Hill function or PWL function for the transcription. The projection of the trajectories in the mRNA subspace $(r_a(t),\ r_b(t))$ are shown in black and the projection of the trajectories onto the protein subspace are shown in turqoise. Other parameter values are: $\gamma_a = \gamma_b = \delta_a = \delta_b = k_a = k_b = 1$, $\theta_a = 0.21, \theta_b = 0.21, m_a = 2.35, m_b = 2.35$, $n_a = n_b = 2$.



(a) CNM with $n_a = n_b = 2$ (b) CPWLM (c) SPWLM

Figure 18: Numerical simulations of the time evolution of the CNM (with $\epsilon = 10$), CPWLM(with $\epsilon = 10$) and SPWLM. $r_a(t)$(black), $r_b(t)$(red), $p_a(t)$(blue), $p_b(t)$ (turqoise). Other parameter values are: $\gamma_a = \gamma_b = \delta_a = \delta_b = k_a = k_b = 1$, $\theta_a = 0.21, \theta_b = 0.21, m_a = 2.35, m_b = 2.35$.

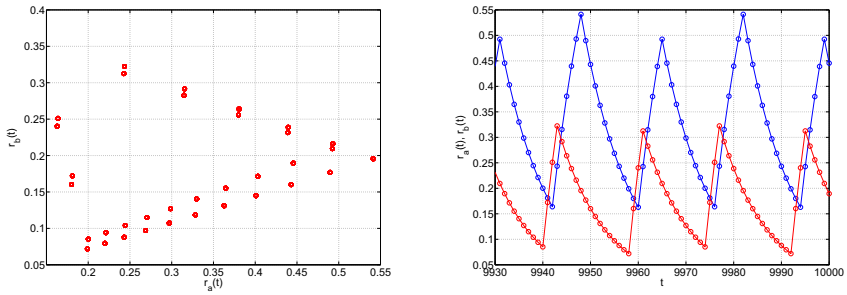
7. Effects of the discretization

To investigate the dynamics of the discrete-time model in (24) obtained by discretizing the SPWLM, we compute the value of the parameter α corresponding to the value of δ_a and δ_b used to obtain Figure 13. Specifically, we set $\alpha = 0.9048$. Figure 19, shows the evolution of the network predicted by this model. We observe a periodic solution which does not match the evolution predicted either by the CPWLM or the SPWLM (see Figure 13).

Moreover, when the model parameters are varied, we observe the onset of more complex behaviour as summarized in Figure 20 where a bifurcation diagram is shown, obtained by varying the parameters α and θ. Each colour corresponds to a different periodic solution that exists for a pair of values of the parameters α, θ. As it is apparent from the figure, the discrete-time model exhibits a large variety of periodic solutions with

[4] In the context of piecewise-smooth systems, sliding refers to the high-frequency (theoretically infinite) switching of the model between its possible configurations. For more details about sliding motion in piecewise linear systems see [12].

ACCEPTED MANUSCRIPT



(a) (b)

Figure 19: a-b) Simulation of the discrete-time model with the corresponding parameters values the same as in Figures 8 and 13. a)The projection of the trajectory onto the protein subspace, $(r_a(t), r_b(t))$. b) The concentrations $r_a(t)$ and $r_b(t)$ plotted against the time t. Blue colour refers to $r_a(t)$, and red to $r_b(t)$.

different periodicity that are not necessarily associated to realistic dynamics of the gene network. Some of these periodic orbits, were also confirmed analytically as shown in Appendix B and independently in [7]. The important difference between the discretized model with the continuous time models, is that it predicts only oscillatory behavior. Namely, for any range of parameters the discrete model always predicts oscillations of the protein concentrations, in contradiction with all the continuous models, where we saw that oscillations exist only in some parameter regions (CNM, CPWLM) or they are totally absent (SNM, SPWLM). Finally, note that the predictions of the discretized model are highly affected by the time-step chosen and the parameter values being set. Therefore, our analysis indicates that such an approach can be unviable to capture experimentally observed behaviour.

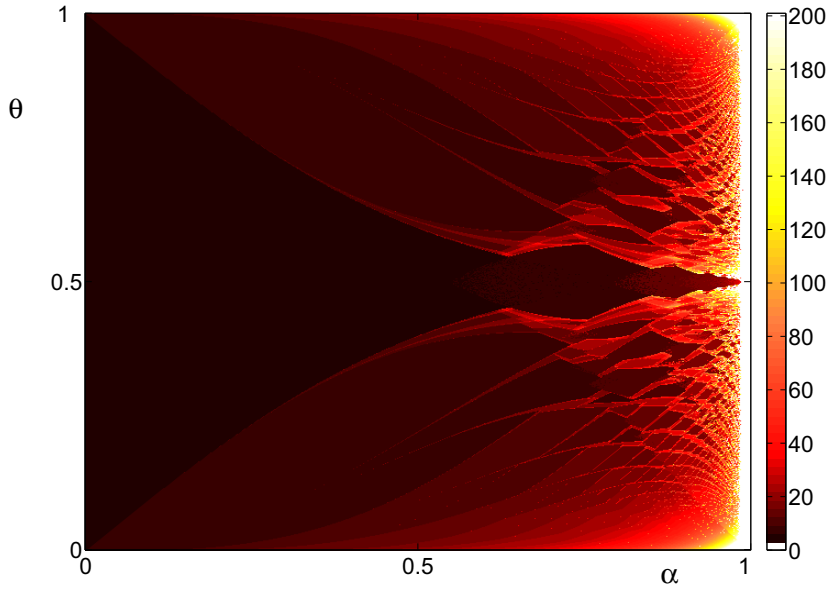


Figure 20: This picture shows the regions of existence of different periodic solutions of the two-gene network. Different colour implies different periodicity of the network.

ACCEPTED MANUSCRIPT

8. Discussion

The results presented in the paper suggest that while some qualitative behaviour is preserved when making different assumptions, the quantitative predictions of different models can be surprisingly different. We expect this to be more the case when larger networks are considered.

An important issue is then whether qualitative or quantitative predictions are needed. In some cases, using more abstract models than PWL models can be an acceptable option. For example, in [8] and [6] Boolean network models are used to describe the yeast cell-cycle control network and the Drosophila patterning network respectively. It is shown that, under certain circumstances, the predicted behaviour is qualitatively similar to that obtained by using an ODE model of the network. The problem is when quantitative predictions are needed. For instance, Mochizuki [34] shows that the predictions of Boolean models can become unrealistic or too complex for larger networks when compared to those of the corresponding ODE models.

Unfortunately, there is not yet a unifying mathematical framework to decide what the best modeling approach to use is and what assumptions can be safely made to simplify the network of interest. The big challenge is how to keep the model simple, without risking missing important features of the real system. This paper offers some guidelines, highlighting the unwanted effects of some of the most commonly made assumptions when modeling biological networks.

We wish to emphasize that our findings apply to other network structures and larger networks. For example, in large networks one can use the mRNA quasi-steady-state assumption in order to simplify the model. As this paper showed, this assumption is only possible for those genes with significantly different time scales for their corresponding mRNA and protein. Also, in the case of the PWL approximation, CPWLMs of larger networks will be large-scale extended piecewise linear systems whose dynamics is bound to be affected by the presence of sliding motion which can cause the predictions of the models to be further away from realistic expectations. We believe that the PWL approximation can only be made in combination with the mRNA quasi-steady-state assumption. In that case, if the transcription dynamics are step-like, then the Hill function might be replaced by a PWL function.

A full understanding of the impact of various modelling assumptions on generic network structures is a pressing open problem that remain to be addressed.

9. Conclusions

We discussed the modelling of gene regulatory networks using different approaches by means of a representative two-gene network. We looked at the effects on the dynamics of some key assumptions often made in the literature on modeling gene networks.

After deriving a complete nonlinear ODE model describing both mRNA and protein concentrations, we considered a simplified model obtained by considering a quasi-steady-state assumption on the mRNA dynamics. We then studied the existence and stability of equilibria in both the complete and simplified nonlinear models. We proved that, while the complete nonlinear model shows the occurrence of a Hopf bifurcation leading to persistent oscillatory behaviour, when the simplified nonlinear model is considered this phenomenon disappears. We then investigated in greater detail the effect on the model behaviour of taking the quasi-steady state assumption on the mRNAs. By considering an appropriate slow-fast model, we showed that the predictions of the SNM and the CNM become significantly different as the time scales of the mRNA and proteins are varied with the Hopf bifurcation point disappearing at infinity when the quasi-steady-state approximation is made.

Another important issue we looked at is the choice of the Hill coefficients which has two important consequences on the model framework used and its predictions. Specifically, we proved that, under certain conditions, oscillatory behaviour is exhibited by the network model if the Hill coefficients are sufficiently high. In particular, we found that in the CNM a Hopf bifurcation in only possible if the Hill coefficients values are above a certain threshold (in our example $n_a n_b \geq 4$). Moreover, if the Hill coefficients are large enough then it is possible to approximate Hill kinetics terms with step functions.

Accepted manuscript

ACCEPTED MANUSCRIPT

This approximation gives rise to PWL models of the network that we further discussed and analyzed. In particular, after presenting the complete PWL model of the network of interest, we discussed the ensuing dynamics showing the presence of solutions such as a high-frequency switching behaviour which is not always close to the predictions of the nonlinear models. Indeed, we found that the PWL and smooth models give the same qualitative and quantitative predictions if the Hill coefficients are chosen to be above a certain threshold value dependent on the parameter region of interest.

Finally, we investigated discrete-time models recently presented in the literature. We showed that such models can be obtained by discretizing the continuous-time ones. The resulting model, though, were shown to predict spurious dynamics, often unrealistic for the network of interest.

Our analysis suggests that particular care must be taken when modelling gebe regulatory networks. In particular, special care must be taken in considering the assumptions discussed in this paper.

References

[1] Alon, U., 2006. An introduction to systems biology: Design principles of biological circuits. Chapman & Hall.

[2] Arkin, A., Ross, J., McAdams, H.A., 1998. Stochastic kinetic analysis of developmental pathway bifurcation in phage λ-infected Escherichia coli cells. Genetics 149, 16331648.

[3] Batt, G., Belta, C., Weiss, R., 2008. Temporal Logic Analysis of Gene Networks under Parameter Uncertainty, IEEE Trans. Aut. Con. 53, 215-229

[4] Belta, C., Esposito, J.M., Kim, J., Kumar, V., 2005. International Journal of Robotics Research 24, 219.

[5] Casey, R., de Jong, H., Gouzé, J.-L., 2006. Piecewise-linear models of genetic regulatory networks: Equilibria and their stability. J. Math. Biol. 52(1), 27-56.

[6] Chaves, M., Sontag, E.D., Albert, R., 2006. IEE Proceedings - Systems Biology 153, 154-167.

[7] Coutinho, R., Fernandez, B., Lima, R., Meyroneinc, A., 2006. Discrete time piecewise affine models of genetic regulatory networks. J. Math. Biol. 52, 524-570.

[8] Davidich, M., Bornholdt, S., 2008. The transition from differential equations to Boolean networks: A case study. J. Theor. Biol. 255, 269-277. in simplifyingaregulatorynetworkmodel

[9] de Jong., H., 2002. Modeling and simulation of genetic regulatory systems: A literature review. J. Comput. Biol. 9 (1), 69105.

[10] de Jong, H., Gouzé, J-L., Hernandez, C., Page, M., Sari, T., Geiselmann., J.,2004. Qualitative simulation of genetic regulatory networks using piecewise-linear models. Bull. Math. Biol. 6 (2), 301340.

[11] Del Vecchio, D., 2007. Design and analysis of an activator-repressor clock in *E.Coli*. Proc. of American Control Conference, July 2007, New York.

[12] Di Bernardo, M., Budd, C.J., Champneys, A.R., Kowalczyk, P., 2007. Piecewise-smooth dynamical systems: theory and applications. Springer-Verlag (Applied Mathematics series no. 163).

[13] Edelstein-Keshet, Leah, 1988. Mathematical Models in Biology SIAM Rev. Volume 30, Issue 4, pp. 296-299.

[14] Edwards, R., 2000/ Analysis of continuous-time switching networks. Physica D 146, 165-199.

[15] Endy, D., Brent, R., 2001. Modelling cellular behaviour. Nature 409 391395.

[16] Elowitz, M.B., Leibler, S., 2000. A synthetic oscillatory network of transcriptional regulators. Nature 403, 335338.

[17] Gantmacher, F.R. 1998. The Theory of Matrices, vol. 2. AMS Chelsea Publishing, Providence, RI. Reprint of the 1959 Translation from Russian by K.A. Hirsch.

[18] Gardner, T.S., Cantor, C.R., Collins, J.J., 2000. Construction of a genetic toggle switch in *Escherichia coli*. Nature 403, 339342.

[19] Gebert, J., Radde, N., Weber,G.-W., 2007. Modeling gene regulatory networks with piecewise linear differential equations. Eur. J. Oper. Res. 181, 1148-1165.

[20] Gibson, M.A., Bruck, J., 2000. Efficient exact stochastic simulation of chemical systems with many species and many channels. J. Phys. Chem. A 104, 18761889.

[21] Gillespie, D.T., 1977. Exact stochastic simulation of coupled chemical reactions. J. Phys. Chem. 81(25), 2340-2361.

[22] Glass, L. 1975. Classification of biological networks by their qualitative dynamics. J. Theo. Biol. 54, 85-107.

[23] Glass, L., Kauffman, S.A., 1973. The logical analysis of continuous non-linear biochemical control networks. J. Theo. Biol. 39, 103-129.

[24] Gouze, J.-L., Sari, T., 2002. A class of piecewise linear differential equations arising in biological models. Dyn. Syst. 17(4), 299-316.

[25] Farcot, E., Gouze, J.-L., 2006. Periodic solutions of piecewise affine gene network models: the case of a negative feedback loop. INRIA report.

[26] Guantes, R., Poyatos J., 2006. Dynamical principles of two-component genetic oscillators. PLOS Computational Biology 2, 188-197.

[27] Hasty, J., McMillen, D., Isaacs, F., Collins, J.J., 2001. Computational studies of gene regulatory networks: in numero molecular biology. Nat. Rev. Genet. 2, 268279.

[28] Hill, A.V., 1910. The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves. J. Physiol. 40 (Suppl.),iv-vii.

Accepted manuscript

ACCEPTED MANUSCRIPT

[29] Karlebach, G., Shamir R., 2008. Modelling and analysis of gene regulatory networks. Nat. Rev. Mol. Cell Biol. 9, 770-780.
[30] Kauffman, S.A., 1969. Homeostasis and differentiation in random genetic control networks. Nature 224, 177178.
[31] Kauffman, S.A., 1969. Metabolic stability and epigenesis in randomly constructed genetic nets. J. Theo. Biol. 22, 437-467.
[32] Kauffman, S.A., 1993. The Origins of Order: Self-Organization and Selection in Evolution. Oxford University Press, New York.
[33] McAdams, H.M., Arkin, A., 1997. Stochastic mechanisms in gene expression. Proc. Natl. Acad. Sci. 94, 814819.
[34] Mochizuki, A., 2005. An analytical study of the number of steady states in gene regulatory networks. J. Theor. Biol. 236, 291-310.
[35] Lewin, B., 2007. Genes IX. Jones and Bartlett.
[36] Plahte, E., Kjoglum, S., 2005. Analysis and generic properties of gene regulatory networks with graded response functions. Physica D 201, 150-176.
[37] Pomerening, J.R , Kim, S.Y., Ferrell, J.E., 2005. Systems-level dissection of the cell-cycle oscillator: Bypassing positive feedback produces damped oscillations. Cell 122, 565578.
[38] Ribeiro, Andre S., Zhu, R., Kauffman, S.A., 2006. A General Modeling Strategy for Gene Regulatory Networks with Stochastic Dynamics, J. Comp. Biol. 13(9), 1630-1639.
[39] Samad, H.E., Khammash, M., Petzold, L., Gillespie, D., 2005. Stochastic modelling of gene regulatory networks. Int. J. Robust Nonlinear Control 15, 691711.
[40] Smolen, P., Baxter, D.A., Byrne, J.H., 2000. Modeling transcriptional control in gene networksmethods, recent results, and future directions. Bull. Math. Biol. 62, 247292.
[41] Snoussi, E.H., 1989. Qualitative dynamics of piecewise-linear differential equations: a discrete mapping approach. Dyn. Stabil. Syst. 4, 565-583.
[42] Somogyi, R., Sniegoski, C.A., 1996. Modeling the complexity of genetic networks: Understanding multigenic and pleiotropic regulation. Complexity 1(6), 4563.
[43] Strogatz, S. H., 2001. Nonlinear Dynamics and Chaos: With Applications in Physics, Biology, Chemistry, and Engineering. Perseus Books Group, ISBN 0-7382-0453-6.
[44] Sugita, M., 1961. Functional analysis of chemical systems in vivo using a logical circuit equivalent. J. Theo. Biol. 1, 179-192.
[45] Sugita, M., 1963. Functional analysis of chemical systems in vivo using a logical circuit equivalent: II. The idea of a molecular automaton. J. Theor. Biol. 4, 179-192.
[46] Thomas, R., 1973. Boolean formalization of genetic control circuits. J. Theor. Biol. 42, 563-585.
[47] Tyson, Othmer 1978. The dynamics of feedback control circuits in biochemical pathways, Prog.Theor.Biol., 5, 2-62.
[48] Widder, S., Schicho, J., Schuster, P., 2007. Dynamic patterns of gene regulation I: Simple two-gene systems. J. Theo. Biol. 246, 395-419.
[49] www.igem.org
[50] Yagil, G., Yagil, E., 1971. On the relationship between effector concentration and the rate of induced enzyme synthesis. Biophys. J. 11, 11-27.
[51] Yagil, G., 1975. Quantitative aspects of protein induction. Curr. Top. Cell Regul. 9, 183-235.

Accepted manuscript

A. A generalization in $\mathbb{R}^2$

The lack of persistent oscillations in the SNM can be generalised to a wider class of two-gene networks. In particular, we prove that any model of a two-gene network, under the steady-state mRNA assumption and no self-regulation[5], cannot exhibit a limit cycle associated with persistent oscillations of the gene products and protein concentrations. The key is the so-called Bendixson criterion [43], for planar nonlinear systems (see also [47] and [13] for a review of previous results).

Consider an ODE model of a two-gene network, with neither of the two genes being self-inhibited or self-activated, of the form

$$\begin{aligned} \dot{p_a} &= k_a f_a(p_b) - \delta_a p_a, \\ \dot{p_b} &= k_b f_b(p_a) - \delta_b p_b. \end{aligned} \tag{69}$$

In order to apply the Bendixson's criterion, we must calculate the divergence of the vector field $\dot{\mathbf{p}} = (\dot{p_a}, \dot{p_b})$. Since gene a is not self-regulated, the function f_a depends only upon p_b, and not on p_a. Hence the partial derivative $\partial f_a(p_b)/\partial p_a$ will be zero. The same applies to the partial derivative $\partial f_b(p_a)/\partial p_b$. Therefore:

$$\nabla \cdot \dot{\mathbf{p}} = -(\delta_a + \delta_b) \tag{70}$$

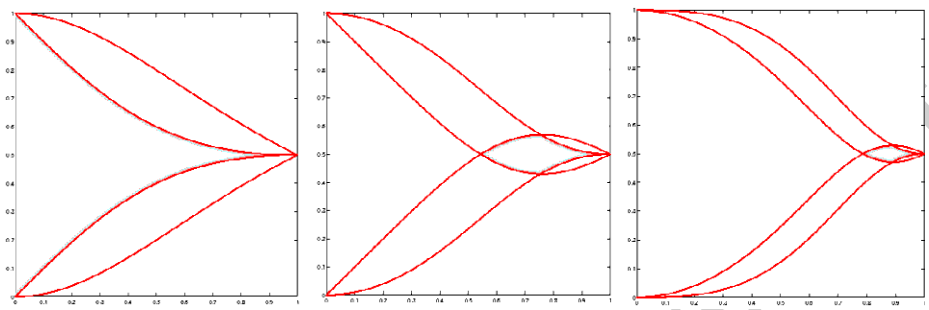
Hence the divergence of $\dot{\mathbf{p}}$ is always negative. Immediate application of Bendixson's criterion then shows that no limit cycles are possible for systems of the type described by equations (69).

[5] There is no self-loop in the network; no proteins can regulate the gene that encoded them.

ACCEPTED MANUSCRIPT

B. Periodic orbits in the discrete-time model

To illustrate the analytical procedure needed to prove the existence of periodic solutions for the activation-inhibition network, we look now at a representative periodic solution. We assume that $\theta_a = \theta_b = \theta$ and we focus in the case of *balanced* periodic orbits, which are periodic orbits which have equal number of iterations in each of the four regions determined by the inequalities $(p_a > \theta)$ and $(p_b > \theta)$, $(p_a > \theta)$ and $(p_b < \theta)$, $(p_a < \theta)$ and $(p_b > \theta)$,$(p_a < \theta)$ and $(p_b < \theta)$. Specifically it is possible to prove the following statement which was also independently presented in [7].



(a) $n = 1$ (b) $n = 2$ (c) $n = 3$

Figure 21: Shaded regions correspond to the regions of existence and stability of the balanced period-$4n$.

Proposition 1. *For $\alpha \in (0,1)$ and $\theta \in (0,1)$, a balanced period $4n$ exists if and only if:*

$$\max\left\{\frac{\alpha^n}{1+\alpha^{2n}}, \frac{1-\alpha^{n-1}+\alpha^{2n}}{1+\alpha^{2n}}\right\} \leq \theta < \min\left\{\frac{1-\alpha^n+\alpha^{2n}}{1+\alpha^{2n}}, \frac{\alpha^{n-1}}{1+\alpha^{2n}}\right\}. \tag{71}$$

Moreover, the periodic solution is always stable in its interval of existence.

Proof. Let $p_1, p_2, ...p_n, ..., p_{2n}, ...p_{3n}, ...p_{4n}$ the iterates of the orbit of interest. Clearly this period $4n$ exists if and only if there exists a fixed point sequence:

$$p_1 < \theta \text{ ,..., } p_{2n} < \theta \text{ , } p_{2n+1} \geq \theta \text{ , ... ,} p_{2n} \geq \theta \tag{72}$$

of the map (24), where:

$$p_j = \alpha p_{j-1} \qquad \text{for all} \qquad 2 \leq j \leq n+1 \tag{73}$$

$$p_i = \alpha p_{i-1} + 1 - \alpha \qquad \text{for all} \qquad n+2 \leq i \leq 3n+1 \tag{74}$$

$$p_k = \alpha p_{k-1} \qquad \text{for all} \qquad 3n+2 \leq k \leq 4n \tag{75}$$

$$p_1 = \alpha p_{4n}. \tag{76}$$

By back-and forward- substitution of equations (73), (74, (75, (76) we can find p_1:

$$p_1 = \frac{\alpha^n}{1+\alpha^{2n}}. \tag{77}$$

Because $0 \leq \alpha \leq 1$ equations (73) imply that $p_1 \geq p_j$ for all $2 \leq j \leq n+1$. Additionally, equations (74) imply that $p_{2n} \geq p_i$ for all $n+2 \leq i \leq 2n-1$. In other words, if $p_1 \leq \theta$ and simultaneously $p_{2n} \leq \theta$, then $p_t \leq \theta$ for all $1 \leq t \leq 2n$. Similarly, it can be shown that if $p_{4n} \geq \theta$ and simultaneously p_{2n+1} then $p_r \geq \theta$ for all $2n+1 \leq r \leq 4n$. Hence, compatibility conditions will hold if and only if:

$$p_1 \leq \theta \text{ and } p_{2n} \leq \theta, \tag{78}$$

$$p_{2n+1} \geq \theta \text{ and } p_{4n} \geq \theta \tag{79}$$

Accepted manuscript

ACCEPTED MANUSCRIPT

From (77) we can easily derive the solution of each p_i, $\forall 1 \leq i \leq 4n$. Specifically, for p_{2n}, p_{2n+1}, p_{4n} we have:

$$p_{2n} = \frac{1-\alpha^{n-1}+\alpha^{2n}}{1+\alpha^{2n}} \ , \ p_{2n+1} = \frac{1-\alpha^{n}+\alpha^{2n}}{1+\alpha^{2n}} \ , \ p_{4n} = \frac{\alpha^{n-1}}{1+\alpha^{2n}} \tag{80}$$

Let a_1 the root of equation $p_{4n} - p_{2n} = 0$ or $1 - 2\alpha^{n-1} + \alpha^{2n} = 0$. Then it can be shown that for $\alpha \geq \alpha_1$ there are real values of α that satisfy equation (71). Now let, a_2 the root of equation $1-\alpha^{n-1}-\alpha^{n}+\alpha^{2n} = 0$ (which is equivalent to both equations $p_{2n} - p_1 = 0$ and $p_{4n} - p_{2n+1}$). It can be easily shown that for $0 \leq \alpha \leq \alpha_1$ we have $p_{2n} \geq p_1$ and $p_{4n} \leq p_{2n+1}$ where for $\alpha \geq \alpha_2$ the opposite inequalities hold respectively.

Accepted manuscript