Mean field analysis of a spatial stochastic model of a gene regulatory network

M. Sturrock · **P. J. Murray** · **A. Matzavinos** · **M. A. J. Chaplain**

Abstract A gene regulatory network may be defined as a collection of DNA segments which interact with each other indirectly through their RNA and protein products. Such a network is said to contain a negative feedback loop if its products inhibit gene transcription, and a positive feedback loop if a gene product promotes its own production. Negative feedback loops can create oscillations in mRNA and protein levels while positive feedback loops are primarily responsible for signal amplification. It is often the case in real biological systems that both negative and positive feedback loops operate in parameter regimes that result in low copy numbers of gene products.

In this paper we investigate the spatio-temporal dynamics of a single feedback loop in a eukaryotic cell. We first develop a simplified spatial stochastic model of a canonical feedback system (either positive or negative). Using a Gillespie's algorithm, we compute sample trajectories and analyse their corresponding statistics. We then derive a system of equations that describe the spatio-temporal evolution of the stochastic means. Subsequently, we examine the spatially homogeneous case and compare the results of numerical simulations with the spatially explicit case. Finally, using a combination of steady-state analysis and data clustering techniques, we explore model behaviour across a subregion of the parameter space that is difficult to access experimentally and compare the parameter landscape of our spatio-temporal and spatially-homogeneous models.

Keywords gene regulatory network · feedback loop · spatial stochastic model · mean field · data clustering

1 Introduction

Gene regulatory networks (GRNs) lie at the core of intracellular signal transduction and can be defined to be a collection of DNA segments which interact with each other indirectly through their RNA and protein products. One of the key players in GRNs are a class of proteins called transcription factors. In response to a variety of biological signals, transcription factors modify the transcription rate of genes, allowing cells to produce the proteins they need at the appropriate times and in the appropriate quantities. It is now well established that GRNs contain a small set of recurring regulation patterns, which are commonly referred to as network motifs [Milo et al., 2002]. Network motifs can be thought of as recurring circuits of interactions from which complex GRNs are constructed. These motifs were first systematically defined in Escherichia coli, in which they were detected as patterns that occurred in the transcription network much more often than would be expected in random networks. The same motifs have since been found across a range of cell types, from bacteria and yeast to plants and animals [Eichenberger et al., 2004, Lee et al., 2002, Saddic et al., 2006, Boyer et al., 2005].

M. Sturrock (✉)
Mathematical Biosciences Institute, The Ohio State University, Columbus, OH 43210, USA
E-mail: sturrock.3@mbi.osu.edu

P. J. Murray · M. A. J. Chaplain
Division of Mathematics, University of Dundee, Dundee, DD1 4HN, UK

A. Matzavinos
Division of Applied Mathematics, Brown University, Providence, Rhode Island, 02912, USA

A GRN is said to contain a negative feedback loop if its products, either directly or indirectly, inhibit gene transcription. It is known that negative feedback loops can create oscillations in mRNA and protein levels [Kobayashi and Kageyama, 2011, Geva-Zatorsky et al., 2010, Nelson et al., 2004], keep mRNA levels at a constant level [Buratti and Baralle, 2011] or prevent run-away reactions, in the case where a negative feedback loop is coupled with a positive feedback loop [Harris and Levine, 2005]. Negative feedback loops are commonly found in a diverse range of biological processes, including inflammation, meiosis, apoptosis and the heat shock response [Alberts et al., 2008, Lahav et al., 2004, Fall et al., 2002]. On the other hand, when a gene product promotes its own production (directly or indirectly), it is known as a positive feedback loop. Positive feedback loops are primarily responsible for signal amplification, and are also known to create bistability in certain circumstances [Ingolia and Murray, 2007]. Examples of positive feedback loops in eukaryotic cells include the PtdInsP3- and Rho GTPase-mediated positive feedback and p42 mitogen-activated protein kinase (MAPK) is known to be involved in a positive feedback loop with the cell-division cycle protein kinase Cdc2 [Weiner et al., 2002, Xiong and Ferrell, 2003]. In this paper we are interested in understanding a single genetic feedback loop in a eukaryotic cell, be it a positive or negative.

Transcription factors and mRNA molecules are often found in low quantities in eukaryotic cells, thus raising the likelihood that variations in copy numbers can cause significant fluctuations in the dynamical evolution of gene regulatory networks [Shahrezaei and Swain, 2008, Barik et al., 2008, 2010]. Other sources of intracellular intrinsic noise include the binding and unbinding of transcription factors to promoter sites, production processes such as transcription and translation and degradation events [Wilkinson, 2009]. The spatial localisation of molecules is also a significant source of intrinsic noise in eukaryotic GRNs [Gamba et al., 2005]. For instance, many important cellular processes rely on collisions between reactants (for example, transcription factors and promoter sites) and occur in a highly localised subregions of the cell (e.g., cytoplasm, endoplasmic reticulum, Golgi apparatus, nucleus, mitochondrion, etc.) that often have different metabolic activities and are separated from the rest of the cell by thin lipid membranes.

The mathematical modelling of GRNs has a long and rich history. In the case of negative feedback loops, perhaps the earliest example comes from Goodwin, where a system of two ordinary differential equations (ODEs) was used to create a phenomenologically derived model of a self-repressing gene [Goodwin, 1965]. A Hill function was used to capture the feedback of protein on mRNA production, a feature which has been mimicked by many models since. The work of Goodwin was continued and extended by Griffith who showed that such a system could not exhibit oscillatory dynamics without the introduction of another model species [Griffith, 1968a]. Mahaffy and co-workers developed this work further by introducing spatial structure and also delays accounting for transcription and translation [Mahaffy and Pao, 1984, Busenberg and Mahaffy, 1985, Mahaffy, 1988]. Some years later, it was discovered that introducing delays to ODE models of negative feedback loops could produce sustained oscillatory dynamics [Tiana et al., 2002]. Jensen et al. found the invocation of an unknown third species (as Griffith had done) could be avoided via the introduction of delay terms to a model of the Hes1 GRN (representing the processes of transcription and translation) [Jensen et al., 2003]. The Hes1 GRN is a simple example of a GRN which possesses a single negative feedback loop and benefits from having been the subject of numerous biological experiments [Hirata et al., 2002, Kagemyama et al., 2007, Kobayashi et al., 2009, Kobayashi and Kageyama, 2010, 2011]. A delay differential equation (DDE) model of the Hes1 GRN was also studied in Monk [2003]. The effect of low particle numbers in Monk's DDE model of the Hes1 GRN was explored in Barrio et al. [2006]. A spatio-temporal model of the Hes1 GRN, using a partial differential equation (PDE) approach, was presented in Sturrock et al. [2011] and extensions of this model were considered in Sturrock et al. [2012]. A spatial stochastic model of the Hes1 GRN in embryonic stem cells was studied in Sturrock et al. [2013] – this model was derived using mass action kinetics and is analysed further in this paper. Wang et al. considered a spatially-homogeneous model of a self-repressing gene in Wang et al. [2012], and were able to identify certain regions of the parameter space which yielded 'stochastic oscillations'.

Positive feedback loops have also been extensively studied using mathematical models. In a similar manner to his negative feedback paper Griffith [1968a]), Griffith studied a model of a single genetic positive feedback loop Griffith [1968b] that was comprised of two ordinary differential equations. He was able to demonstrate multistability under certain parameter regimes, a phenomenon also demonstrated by Tyson and Othmer [1978]. Using a more detailed model of underlying biochemistry, Keller determined the number and stability of steady states of positive feedback systems with multiple transcription factors [Keller, 1994, 1995] and suggested that

the choice among multiple steady states of gene product concentrations could constitute a way of distinguishing between different cell types within a multicellular organism. Furthermore, it has been demonstrated that positive feedback loops are necessary for multistability, when using either the Boolean or continuous approaches [Snoussi and Thomas, 1993, Thomas, 1994, Thomas et al., 1995]. It has been shown that the response of a multistable gene network model to perturbations will depend on the state of the network when the perturbation is applied. However, in a DDE system modelling a positive feedback loop, the response to a given perturbation can be dramatically different. Smolen showed that in a single DDE model of a transcription factor activating its own transcription, bistability could be found in the solution of the model in certain regions of parameter space [Smolen et al., 1999a,b]. Walther et al. recently considered a stochastic model of cell polarisation which included a positive feedback driven interconversion of different forms of Rho GTPases [Walther et al., 2012].

Although there has been much work done in this area, there still remain some unanswered questions. In this paper we aim to answer the following questions:

– What is the mean field behaviour of a simple gene regulatory network consisting of a single feedback loop?
– Is it possible to accurately approximate an explicitly spatial model of a gene regulatory network with a spatially homogeneous one?
– How does the parameter landscape differ for the negative and positive feedback loop cases?

The format of this paper is as follows: we introduce a simplified stochastic model of a gene regulatory network consisting of a single feedback loop; we present the corresponding reaction diffusion master equation and, using Gillespie's algorithm, compute sample trajectories and their corresponding statistics. In the subsequent section a system of equations are derived that describe the spatio-temporal evolution of the stochastic means. Subsequently, we examine the spatially homogeneous case and compare the results of numerical simulations with the spatially explicit case. Finally, we explore, using data clustering techniques and steady-state analysis, a region of the parameter space that is difficult to access experimentally and compare solutions between spatially heterogeneous and homogeneous models.
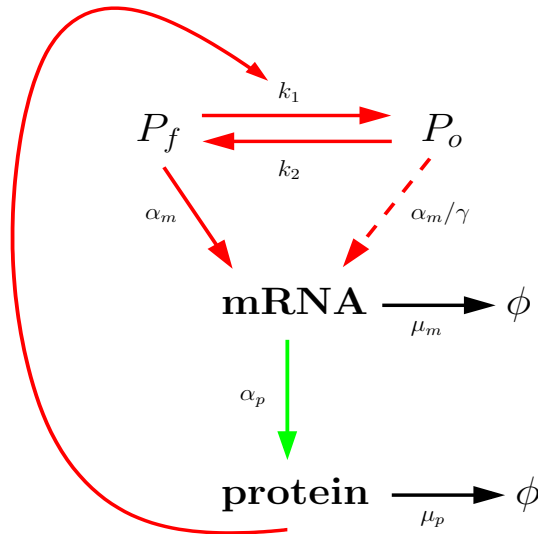

Fig. 1 A schematic illustration of a simple gene regulatory network consisting of a single feedback loop. If the parameter γ is less than or equal to 1 then the network has a single positive feedback loop, otherwise if γ is greater than 1 it has a single negative feedback loop. When the promoter site is free (P_f), mRNA is transcribed at a baseline rate α_m. mRNA can then diffuse to the cytoplasm where it produces protein via the process of translation. Newly created proteins (transcription factors) are able to then diffuse back into the nucleus where they can occupy the promoter (P_o) and promote or repress the transcription of their own mRNA (the baseline rate α_m is increased for a positive feedback loop and decreased for a negative feedback loop depending on the value of γ). Reaction arrows displayed in red only occur at the promoter site, while those in green occur only in the cytoplasm and those in black occur everywhere within the cell.
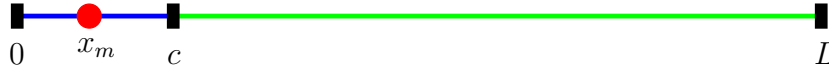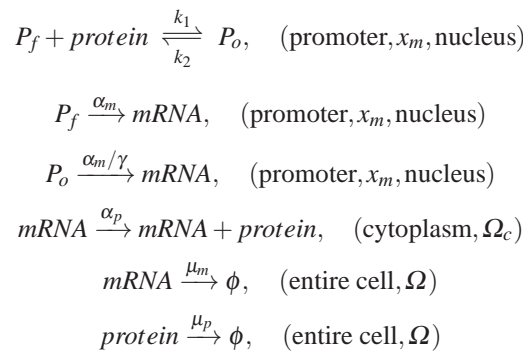
Fig. 2 The one-dimensional domain (Ω) of length L. The cytoplasm (Ω_c) is of length $L - c = \frac{4L}{5}$ and the nucleus (Ω_n) is of length $c = \frac{L}{5}$. The nuclear-cytoplasmic boundary is located at point c and the promoter/gene is located at the point $x_m = \frac{L}{10}$. Subdomains are coloured to match the reaction schematic in Figure 1.
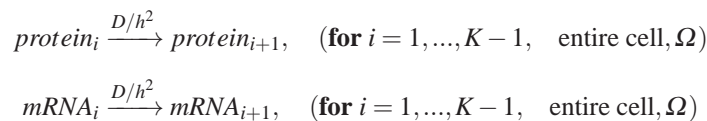
2 A spatial stochastic model of a gene regulatory network with a single feedback loop

The assumptions concerning the molecular reactions in our GRN model follow previous modelling efforts [Sturrock et al., 2013, Wang et al., 2012]. We assume *mRNA* is transcribed from a localised gene site within the cell nucleus and then diffuses from the nucleus across the nuclear membrane and into the cytoplasm where it is translated into *protein*. The *protein* we consider is assumed to be a transcription factor. Transcription factors are a special class of proteins able to promote or repress the recruitment of RNA polymerase, thereby promoting or repressing the production of mRNA [Alberts et al., 2008]. These transcription factors can work alone or with other proteins in a complex. We assume the transcription factor (*protein*) works by itself and after it is produced in the cytoplasm it is then able to move back into the nucleus via diffusion where it binds to the promoter site P in a reversible reaction. This means that the promoter site may be in one of two states – free or occupied, which we denote by P_f and P_o, respectively. When the promoter site is free (P_f), *mRNA* is transcribed at a baseline rate α_m and when the promoter site is occupied (P_o), *mRNA* is transcribed at a rate modified by a parameter γ, i.e., α_m/γ. If $\gamma < 1$ then the GRN has a single positive feedback loop, otherwise if $\gamma > 1$ it has a single negative feedback loop. In the case where $\gamma = 1$, the system can not be described as a feedback loop as the occupation of the promoter site would not make any difference to the evolution of the system. As a result of this, we do not consider the case where $\gamma = 1$, i.e., in this paper we are interested in feedback loops. Both the *mRNA* and *protein* are assumed to degrade at some rate μ_m and μ_p, respectively. The model considers explicitly the spatial distributions of the chemical species involved in the GRN and reactions are localised to appropriate compartments of the cell, as indicated by the colours of the arrows the schematic diagram shown in Figure 1. The model is defined on a one-dimensional spatial domain representing a single cell shown in Figure 2.
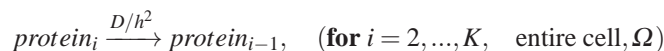
These assumptions are formalised by dividing the computational domain, Ω, into K compartments of length $h = \frac{L}{K}$ and denoting the number of molecules of chemical species X in the i-th compartment $[(i-1)h, ih)$ by X_i, $i = 1, ..., K$. The reactions, with their localisation stated in brackets, are defined as follows:

$$P_f + protein \underset{k_2}{\overset{k_1}{\rightleftharpoons}} P_o, \quad (\text{promoter}, x_m, \text{nucleus})$$

$$P_f \xrightarrow{\alpha_m} mRNA, \quad (\text{promoter}, x_m, \text{nucleus})$$

$$P_o \xrightarrow{\alpha_m/\gamma} mRNA, \quad (\text{promoter}, x_m, \text{nucleus})$$

$$mRNA \xrightarrow{\alpha_p} mRNA + protein, \quad (\text{cytoplasm}, \Omega_c)$$

$$mRNA \xrightarrow{\mu_m} \phi, \quad (\text{entire cell}, \Omega)$$

$$protein \xrightarrow{\mu_p} \phi, \quad (\text{entire cell}, \Omega)$$

where $k_1 = \frac{c_1}{n_A h}$, n_A is Avagadro's constant, c_1 is the rate of protein binding to the promoter in the corresponding well-mixed model and $\alpha_p = \frac{a_p}{L-c}$ where a_p is the rate of protein translation in the corresponding well-mixed model. The corresponding well-mixed model is explicitly defined in section 5. Diffusion events are defined as first order reactions in which a *protein* or *mRNA* molecule moves either to the right

$$protein_i \xrightarrow{D/h^2} protein_{i+1}, \quad (\textbf{for } i = 1, ..., K-1, \quad \text{entire cell}, \Omega)$$

$$mRNA_i \xrightarrow{D/h^2} mRNA_{i+1}, \quad (\textbf{for } i = 1, ..., K-1, \quad \text{entire cell}, \Omega)$$

or to the left

$$protein_i \xrightarrow{D/h^2} protein_{i-1}, \quad (\textbf{for } i = 2, ..., K, \quad \text{entire cell}, \Omega)$$

$$mRNA_i \xrightarrow{D/h^2} mRNA_{i-1}, \quad (\textbf{for } i = 2,...,K, \quad \text{entire cell}, \Omega)$$

Zero-flux boundary conditions are applied at 0 and L together with zero initial conditions with the exception of a single initial free promoter at the promoter site, which we denote by F_0.

2.1 The reaction diffusion master equation

We now describe the reaction diffusion master equation which corresponds to the system of chemical reactions presented in the previous section. More specifically, the GRN is formulated as a discrete-space continuous time Markov process.

Let $P(\mathbf{p},\mathbf{m},\mathbf{f},t)dt$ be the joint probability in the time interval $(t,t+\Delta t)$ that $protein_i(t) = p_i$, for $i = 1,...,K$ (where $\mathbf{p} = [p_1,p_2,...,p_K]$), $mRNA_i(t) = m_i$, for $i = 1,...,K$ (where $\mathbf{m} = [m_1,m_2,...,m_K]$), and $P_{f_{i_m}}(t) = f_{i_m}$, for $i = 1,...,K$ (where $\mathbf{f} = 1$ or 0 in compartment K_{i_m} and zero for all other compartments). Note that we do not have to define a probability for the occupied promoter state (P_o) as the number of occupied promoters at the promoter site can be written as $F_0 - f_{i_m}$. Let us define operators $R_i, L_i : \mathbb{N}^K \to \mathbb{N}^K$ by

$$R_i : [n_1,...,n_i,n_{i+1},...,n_K] \to [n_1,...,n_i+1,n_{i+1}-1,...n_K], \quad i = 1,...,K-1,$$

$$L_i : [n_1,...,n_{i-1},n_i,...,n_K] \to [n_1,...,n_{i-1}-1,n_i+1,...n_K], \quad i = 2,...,K.$$

Then the reaction diffusion master equation for $protein$, $mRNA$ and P_f may be written as:

$$\begin{aligned}
\frac{\partial P(\mathbf{p},\mathbf{m},\mathbf{f},t)}{\partial t} = &\frac{D}{h^2}\sum_{i=1}^{K-1}\left[(p_i+1)P(R_i\mathbf{p},\mathbf{m},\mathbf{f},t) - p_iP(\mathbf{p},\mathbf{m},\mathbf{f},t)\right] \\
&+\frac{D}{h^2}\sum_{i=2}^{K}\left[(p_i+1)P(L_i\mathbf{p},\mathbf{m},\mathbf{f},t) - p_iP(\mathbf{p},\mathbf{m},\mathbf{f},t)\right] \\
&+\alpha_p\sum_{i=c}^{K}\left[(m_i)P(p_{i_c},...p_i-1,...,p_K,\mathbf{m},\mathbf{f},t) - m_iP(\mathbf{p},\mathbf{m},\mathbf{f},t)\right] \\
&+\mu_p\sum_{i=1}^{K}\left[(p_i+1)P(p_1,...p_i+1,...,p_K,\mathbf{m},\mathbf{f},t) - p_iP(\mathbf{p},\mathbf{m},\mathbf{f},t)\right] \\
&+\frac{D}{h^2}\sum_{i=1}^{K-1}\left[(m_i+1)P(\mathbf{p},R_i\mathbf{m},\mathbf{f},t) - m_iP(\mathbf{p},\mathbf{m},\mathbf{f},t)\right] \\
&+\frac{D}{h^2}\sum_{i=2}^{K}\left[(m_i+1)P(\mathbf{p},L_i\mathbf{m},\mathbf{f},t) - m_iP(\mathbf{p},\mathbf{m},\mathbf{f},t)\right] \\
&+\alpha_m\left[f_{i_m}P(\mathbf{p},m_1,...m_{i_m}-1,...,m_K,\mathbf{f},t) - f_{i_m}P(\mathbf{p},\mathbf{m},\mathbf{f},t)\right] \\
&+\frac{\alpha_m}{\gamma}\left[(F_0-f_{i_m})P(\mathbf{p},m_1,...m_{i_m}-1,...,m_K,\mathbf{f},t) - (F_0-f_{i_m})P(\mathbf{p},\mathbf{m},\mathbf{f},t)\right] \\
&+\mu_m\sum_{i=1}^{K}\left[(m_i+1)P(\mathbf{p},m_1,...m_i+1,...,m_K,\mathbf{f},t) - m_iP(\mathbf{p},\mathbf{m},\mathbf{f},t)\right] \\
&+k_1(f_{i_m}+1)p_{i_m}P(\mathbf{p},\mathbf{m},0,...,f_{i_m}+1,...,0,t) - k_1f_{i_m}p_{i_m}P(\mathbf{p},\mathbf{m},\mathbf{f},t) \\
&+k_2((F_0-f_{i_m})+1)P(\mathbf{p},\mathbf{m},0,...,f_{i_m}-1,...,0,t) - k_2(F_0-f_{i_m})P(\mathbf{p},\mathbf{m},\mathbf{f},t).
\end{aligned}$$

2.2 Parameter values

The baseline parameter set (see Table 1) is based on physiologically realistic estimates that have been previously used for modelling similar GRN models [Sturrock et al., 2013, Monk, 2003, Terry et al., 2011]. For simplicity, we do not use different diffusion coefficients for mRNA or protein species, nor do we use different diffusion coefficients in the nucleus and cytoplasm. All simulations are run for times corresponding to 1200 minutes, which

is motivated by experimental measurements of a simple GRN [Kobayashi et al., 2009]. The parameters in the model that are most difficult to measure and for which there exists the least data to support are k_1 and k_2. Indeed, the estimation of k_1, k_2 values has relied on theoretical models [Tafvizi et al., 2011]. Another parameter for which their exists much uncertainty is the diffusion coefficient, D. Experimentalists have found the diffusion coefficient of soluble proteins in aqueous buffers to be in the range $6 \times 10^{-12} \text{m}^2\text{min}^{-1}$ to $6 \times 10^{-11} \text{m}^2\text{min}^{-1}$ [Matsuda et al., 2008, Seksek et al., 1997] but this value may depend on factors such as the size of the protein and the cell type. We note that the results of Klonis et al. [2002] show that diffusion rates of molecules in the cytoplasm and nucleus are up to 100-times slower than in aqueous buffers, which brings the values of Matsuda et al. [2008] and Seksek et al. [1997] into the range $6 \times 10^{-14} \text{m}^2\text{min}^{-1}$ to $6 \times 10^{-13} \text{m}^2\text{min}^{-1}$. Hence, we use the value $6 \times 10^{-13} \text{m}^2\text{min}^{-1}$ for our baseline parameter set. In order to shed light on the influence of these parameter values, in section 6 we will use the developed framework to explore the effects of changing parameters k_1, k_2 and D.

Parameter	Description	Values used in simulations
α_m	Basal transcription rate of mRNA	3.00min^{-1}
γ	Scale of transcriptional repression (enhancement)	$100.00\ (0.1)$
a_p	Translation rate of protein	$2.4 \times 10^{-5} \text{min}^{-1}$
μ_m	Degradation rate of mRNA	0.03min^{-1}
μ_p	Degradation rate of protein	0.12min^{-1}
c_1	Rate of protein binding to promoter	$1.00 \times 10^{15} \text{min}^{-1}$
k_2	Rate of protein unbinding from promoter	0.1min^{-1}
F_0	Number of promoter sites	1
D	Diffusion coefficient	$0.60 \mu\text{m}^2\text{min}^{-1}$
x_m	Position of promoter site	$1.00 \mu\text{m}$
c	Position of beginning of cytoplasm	$2.00 \mu\text{m}$
L	Length of cell	$10.00 \mu\text{m}$
K	Number of compartments which subdivide the domain	100

Table 1 Description of parameters in the GRN model and values used in sections 3, 4, and 5. The upper partition of the table corresponds to chemical kinetics parameters and the lower partition corresponds to spatial parameters.

3 Numerical simulations of the spatial stochastic model

In this section, we present trajectories of the spatial stochastic model which are computed using Gillespie's algorithm. The parameter values used are presented in Table 1. We consider two different parameter sets: in the first, $\gamma = 100$ which corresponds to a negative feedback loop and in the second $\gamma = 0.1$ which corresponds to a positive feedback loop (all other parameters remain the same). Computational results from both single trajectories and the mean of 1000 trajectories are presented.

3.1 Negative feedback loop ($\gamma = 100$)

The first row of Figure 3 shows the temporal evolution (for all model species) of a single trajectory for the negative feedback case. These plots were obtained by summing up the total number of molecules in each compartment of the spatial domain. Protein levels (displayed in blue, first plot, first row) clearly evolve in a noisy oscillatory manner, a phenomenon that is also reflected in the full spatio-temporal evolution of protein molecules presented in the first plot in the second row. One can observe the existence of "stripes" which are unevenly spaced with respect to time, complementing the noisy oscillations observed in the temporal evolution. Furthermore, these stripes are not surprising if one considers the interactions involved in a negative feedback loop (in 1 spatial dimension). Proteins are created in the cytoplasm via the process of translation and proceed to diffuse rapidly to every part of the one-dimensional domain. When sufficiently many proteins accumulate in the promoter region of the nucleus, the promoter becomes occupied, resulting in decreased mRNA production which in turn results in fewer proteins in the spatio-temporal solution. Eventually, protein levels decrease sufficiently so that the promoter site becomes free to once again produce mRNA at the maximum rate. The mRNA levels (displayed in red, second plot, first row) show similar behaviour to the proteins but in much smaller numbers. This is consistent with biological knowledge regarding the relative abundance of proteins and their mRNA [Kar et al., 2009, Fusco et al., 2003]. Due to the effect of low copy number, the spatio-temporal dynamics of mRNA are much noisier than those of the protein. Finally, the temporal dynamics of the promoter occupancy (displayed in light green, third plot, first row) display constant switching between the free and occupied states. The promoter site location is shown in the full spatio-temporal evolution, as it flashes between occupied and free states.

Upon averaging over 1000 trajectories of the master equation, the oscillatory behaviour of the individual trajectories is lost for all model species (see plots in the third row of Figure 3). It is known that when out-of-phase oscillatory time series are summed together, the oscillations collapse and a steady state time series results. In terms of our GRN, these plots are the equivalent of taking the average gene expression data of 1000 distinct individual cells and finding overall constant levels of expression. It is worth noting that it is possible to produce synchronised oscillatory behaviour in a population of cells when cell-to-cell communication is intact [Terry et al., 2011]. Interestingly, the mean plot for the free promoter species reveals that for the particular parameter set presented in Table 1, the promoter is occupied (on average) 50% of the time. The average level of the free promoter appears to be a good indicator of whether or not noisy oscillations are observed in single trajectories, i.e., if on average the free promoter is not occupied 100% or 0% of the time but somewhere in between, then we know that there is some switching between the free and occupied states. The amount of switching is at its maximum if the mean free promoter steady state is 0.5. The spatio-temporal averages displayed in the bottom row, highlight where on average the levels of species are found in the highest abundance. For proteins, it is clear that this is the part of the cytoplasm closest to the nucleus – which is consistent with intuition, as (on average) mRNA molecules won't reach the cell membrane without being degraded. Hence the protein level is not as high near the cell membrane. Also, there is no protein production in the nucleus, hence this is the region with the lowest abundance of protein. The level of mRNA molecules is found on average to be highest in the nucleus, and lowest near the cell membrane which is again consistent with intuition.
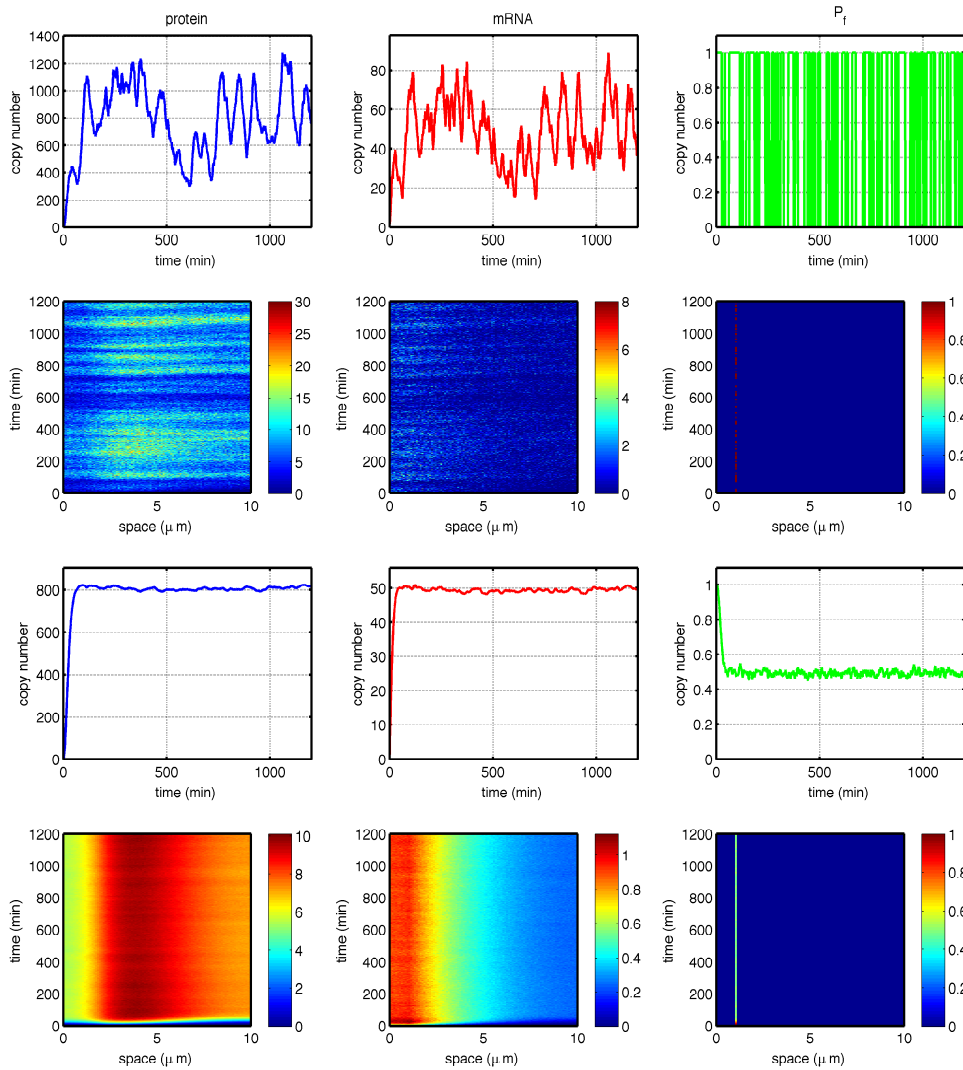
Fig. 3 Plots showing the spatio-temporal evolution of the species in the GRN with a negative feedback loop. Parameter values are as per Table 1 with $\gamma = 100$. The first row shows the temporal evolution of a single trajectory for the protein, mRNA and free promoter species respectively, which were obtained by summing over the individual compartments comprising the domain. The second row shows the full spatio-temporal evolution of the model species (with the ordering the same as in the first row) for a single sample trajectory. The colour bar indicates the number of molecules in a single compartment. The third row shows the temporal evolution of the mean of 1000 trajectories (ordering of species the same as previous rows). Finally, the fourth row shows the full spatio-temporal evolution of the mean of 1000 trajectories for the model species (ordering of species same as previous rows). The colour bar indicates the number of molecules in a single compartment (which may now be non-integer because the means are plotted).

3.2 Positive feedback loop ($\gamma = 0.1$)

Figure 4 provides insight into the spatio-temporal evolution of all model species in the GRN with a positive feedback loop. The format of this Figure is precisely the same as Figure 3, the only difference being that we are now considering the case $\gamma = 0.1$ (all other parameters are as per Table 1). Protein levels no longer evolve in an oscillatory manner for the single trajectory solution presented in the first row. Instead the levels are much higher and appear to approach a steady state of approximately 15000 molecules. This is reflected in the plot showing the spatio-temporal evolution of protein molecules which reveals particularly high levels throughout the cytoplasm (first plot, second row). The mRNA levels display similar behaviour to the proteins but in smaller numbers, as was found for the negative feedback case – though we note that the mRNA levels are much higher than in the negative feedback case. The spatio-temporal plot of mRNA copy number reveals a relatively high number of

mRNA molecules in the nucleus and a noisy distribution of mRNA in the cytoplasm. The plot of the temporal evolution of the free promoter reveals that, in the positive feedback case, there are very few switches between the free and occupied states. Indeed, the promoter site is barely distinguishable in the spatio-temporal solution (shown in the third plot of the second row) because it is occupied almost the entire duration of the simulation.

The mean behaviour for the positive feedback case is perhaps less surprising than the negative feedback case. There is not a discernible difference between the first and third rows of Figure 4 for the mRNA and protein species. Taking the mean of 1000 trajectories appears to simply smooth out the noise in the time series. However, for the free promoter species, examining the mean behaviour makes it very clear that the promoter site is occupied for almost 95% of the time – a feat which is unsurprising given the abundance of protein in the cell domain. The spatio-temporal solution for the mean of 1000 trajectories show a similar but 'noise-filtered' version of the single trajectory. For this parameter set, we were not able to find bistable switching – which can exist for positive feedback loops subject to delays [Smolen et al., 1999a,b]. The mean full spatio-temporal solution appears quite similar for the positive and negative feedback cases (row 4, Figures 3 and 4). However, the number of molecules is much higher in the positive feedback case and the free promoter is not nearly so visible (apart from at the initial time). The findings here corroborate what we found for the negative feedback case, i.e., if the promoter is on average occupied close to 100% or 0% of the time, it is not possible to find oscillatory dynamics. In other words, the state of the free promoter drives the dynamical behaviour.
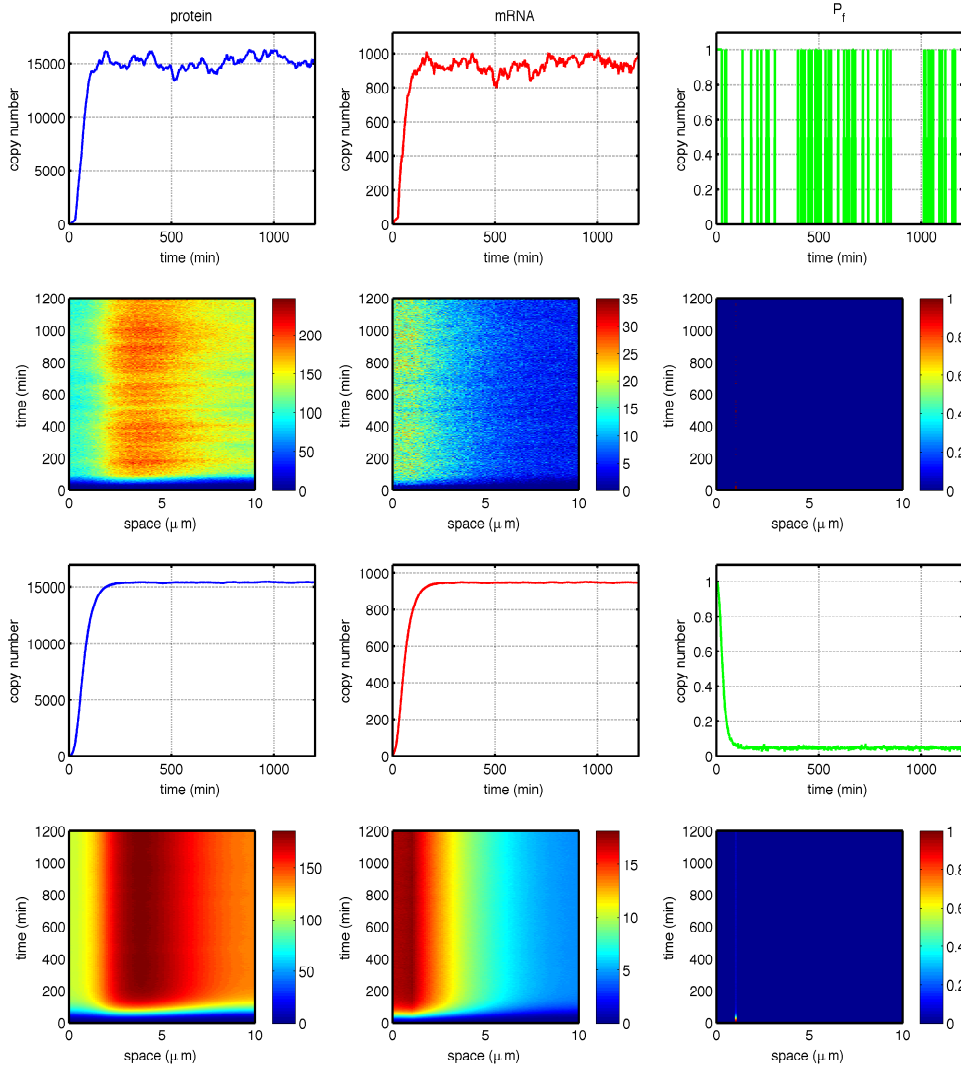
Fig. 4 Plots showing the spatio-temporal evolution of the species in the GRN with a positive feedback loop. Parameters are as per Table 1 with $\gamma = 0.1$. The first row shows the temporal evolution of a single trajectory for the protein, mRNA and free promoter species respectively, which were obtained by summing over the individual compartments comprising the domain. The second row shows the full spatio-temporal evolution of the model species (with the ordering the same as in the first row) for a single trajectory. The colour bar indicates the number of molecules in a single compartment. The third row shows the temporal evolution of the mean of 1000 trajectories (ordering of species the same as previous rows). Finally, the fourth row shows the full spatio-temporal evolution of the mean of 1000 trajectories for the model species (ordering of species same as previous rows). The colour bar indicates the number of molecules in a single compartment (which may now be non-integer because the means are plotted).

4 Stochastic mean vectors

In this section we derive systems of equations for the stochastic mean vectors for variables **p**, **m**, and **f** which are defined as:

$$\langle p_i \rangle(t) = \sum_{\mathbf{p}} \sum_{\mathbf{m}} \sum_{\mathbf{f}} p_i P(\mathbf{p}, \mathbf{m}, \mathbf{f}, t),$$

$$\langle m_i \rangle(t) = \sum_{\mathbf{p}} \sum_{\mathbf{m}} \sum_{\mathbf{f}} m_i P(\mathbf{p}, \mathbf{m}, \mathbf{f}, t),$$

$$\langle f_i \rangle(t) = \sum_{\mathbf{p}} \sum_{\mathbf{m}} \sum_{\mathbf{f}} f_i P(\mathbf{p}, \mathbf{m}, \mathbf{f}, t).$$

By multiplying the reaction diffusion master equation by p_i, summing over **p**, **m**, and **f** and performing some standard manipulations (see [Erban et al., 2007]), a system of equations is obtained for the evolution of the protein

stochastic mean vector:

$$\frac{\partial \langle p_1 \rangle(t)}{\partial t} = \frac{D}{h^2}\Big[\langle p_2 \rangle(t) - \langle p_1 \rangle(t)\Big] - \mu_p \langle p_1 \rangle(t),$$

$$\frac{\partial \langle p_i \rangle(t)}{\partial t} = \frac{D}{h^2}\Big[\langle p_{i+1} \rangle(t) - 2\langle p_i \rangle(t) + \langle p_{i-1} \rangle(t)\Big] - \mu_p \langle p_i \rangle(t), \qquad i = 2, ..., i_c,$$

$$\frac{\partial \langle p_i \rangle(t)}{\partial t} = \frac{D}{h^2}\Big[\langle p_{i+1} \rangle(t) - 2\langle p_i \rangle(t) + \langle p_{i-1} \rangle(t)\Big] + \alpha_p \langle m_i \rangle(t) - \mu_p \langle p_i \rangle(t), \qquad i = i_c + 1, ..., K - 1,$$

$$\frac{\partial \langle p_K \rangle(t)}{\partial t} = \frac{D}{h^2}\Big[\langle p_{K-1} \rangle(t) - \langle p_K \rangle(t)\Big] + \alpha_p \langle m_K \rangle(t) - \mu_p \langle p_K \rangle(t),$$

which corresponds to a discretised version of the reaction-diffusion equation given by

$$\frac{\partial \langle \mathbf{p} \rangle(x,t)}{\partial t} = D\nabla^2 \langle \mathbf{p} \rangle(x,t) + \alpha_p \langle \mathbf{m} \rangle(x,t)\chi_{[c,L]} - \mu_p \langle \mathbf{p} \rangle(x,t),$$

with zero-flux boundary conditions

$$\frac{\partial \langle \mathbf{p} \rangle(0,t)}{\partial x} = \frac{\partial \langle \mathbf{p} \rangle(L,t)}{\partial x} = 0.$$

Here, $\chi_{[c,L]}$ is the characteristic function of the interval $[c,L]$. Similarly, multiplying the reaction diffusion master equation by m_i and summing over $\mathbf{p}$, $\mathbf{m}$, and $\mathbf{f}$ a system of equations is obtained for the evolution of the mRNA stochastic mean vector:

$$\frac{\partial \langle m_1 \rangle(t)}{\partial t} = \frac{D}{h^2}\Big[\langle m_2 \rangle(t) - \langle m_1 \rangle(t)\Big] - \mu_m \langle m_1 \rangle(t),$$

$$\frac{\partial \langle m_i \rangle(t)}{\partial t} = \frac{D}{h^2}\Big[\langle m_{i+1} \rangle(t) - 2\langle m_i \rangle(t) + \langle m_{i-1} \rangle(t)\Big] - \mu_m \langle m_i \rangle(t), \qquad i = 2, ..., i_m - 1,$$

$$\frac{\partial \langle m_{i_m} \rangle(t)}{\partial t} = \frac{D}{h^2}\Big[\langle m_{i_m+1} \rangle(t) - 2\langle m_{i_m} \rangle(t) + \langle m_{i_m-1} \rangle(t)\Big] + \alpha_m\Big[\langle f_{i_m} \rangle(t) + \frac{(F_0 - \langle f_{i_m} \rangle(t))}{\gamma}\Big] - \mu_m \langle m_{i_m} \rangle(t),$$

$$\frac{\partial \langle m_i \rangle(t)}{\partial t} = \frac{D}{h^2}\Big[\langle m_{i+1} \rangle(t) - 2\langle m_i \rangle(t) + \langle m_{i-1} \rangle(t)\Big] - \mu_m \langle m_i \rangle(t), \qquad i = i_m + 1, ..., K - 1,$$

$$\frac{\partial \langle m_K \rangle(t)}{\partial t} = \frac{D}{h^2}\Big[\langle m_{K-1} \rangle(t) - \langle m_K \rangle(t)\Big] - \mu_m \langle m_K \rangle(t),$$

which corresponds to a discretised version of the reaction-diffusion equation given by

$$\frac{\partial \langle \mathbf{m} \rangle(x,t)}{\partial t} = D\nabla^2 \langle \mathbf{m} \rangle(x,t) + \alpha_m\Big[\langle \mathbf{f} \rangle(x,t) + \frac{F_0 - \langle \mathbf{f} \rangle(x,t)}{\gamma}\Big]\delta_{x_m}(x) - \mu_m \langle \mathbf{m} \rangle(x,t),$$

with zero-flux boundary conditions

$$\frac{\partial \langle \mathbf{m} \rangle(0,t)}{\partial x} = \frac{\partial \langle \mathbf{m} \rangle(L,t)}{\partial x} = 0.$$

Here, $\delta_{x_m}(x)$ is the Dirac-delta function localised at the point x_m. Finally, by multiplying the reaction diffusion master equation by f_i and summing over $\mathbf{p}$, $\mathbf{m}$, and $\mathbf{f}$, a system of equations is obtained for the evolution of the free promoter stochastic mean vector:

$$\frac{\partial \langle f_{i_m} \rangle(t)}{\partial t} = k_2(F_0 - \langle f_{i_m} \rangle(t)) - k_1 \langle p_{i_m} f_{i_m} \rangle(t),$$

which corresponds to a discretised version of the PDE

$$\frac{\partial \langle \mathbf{f} \rangle(x,t)}{\partial t} = \Big[k_2(F_0 - \langle \mathbf{f} \rangle(x,t)) - k_1 \langle \mathbf{pf} \rangle(x,t)\Big]\delta_{x_m}(x).$$

Due to the nonlinear term, namely $\langle p_{i_m} f_{i_m} \rangle(t)$, the equations describing the first moments are not closed. The simplest approximation, in terms of the resulting equations, is that the variables $\langle p_{i_m} \rangle(t)$ and $\langle f_{i_m} \rangle(t)$ are uncorrelated, i.e.,

$$\langle p_{i_m} f_{i_m} \rangle(t) \sim \langle p_{i_m} \rangle(t)\langle f_{i_m} \rangle(t).$$

This approximation appears to be largely accurate for both the negative and positive feedback cases, and we demonstrate this in Figure 5. The first and second row correspond to the negative feedback case and the third and

fourth correspond to the positive feedback case. The first two columns display the temporal and spatio-temporal evolution of $\langle pf \rangle(t)$ and $\langle p \rangle(t)\langle f \rangle(t)$ respectively. In the third column we present the error. In particular, the negative feedback case appears to approximated well for the sample set of parameters presented in Table 1. There is a noticeable error near the beginning of the time series for the positive feedback case but it quickly fades away.
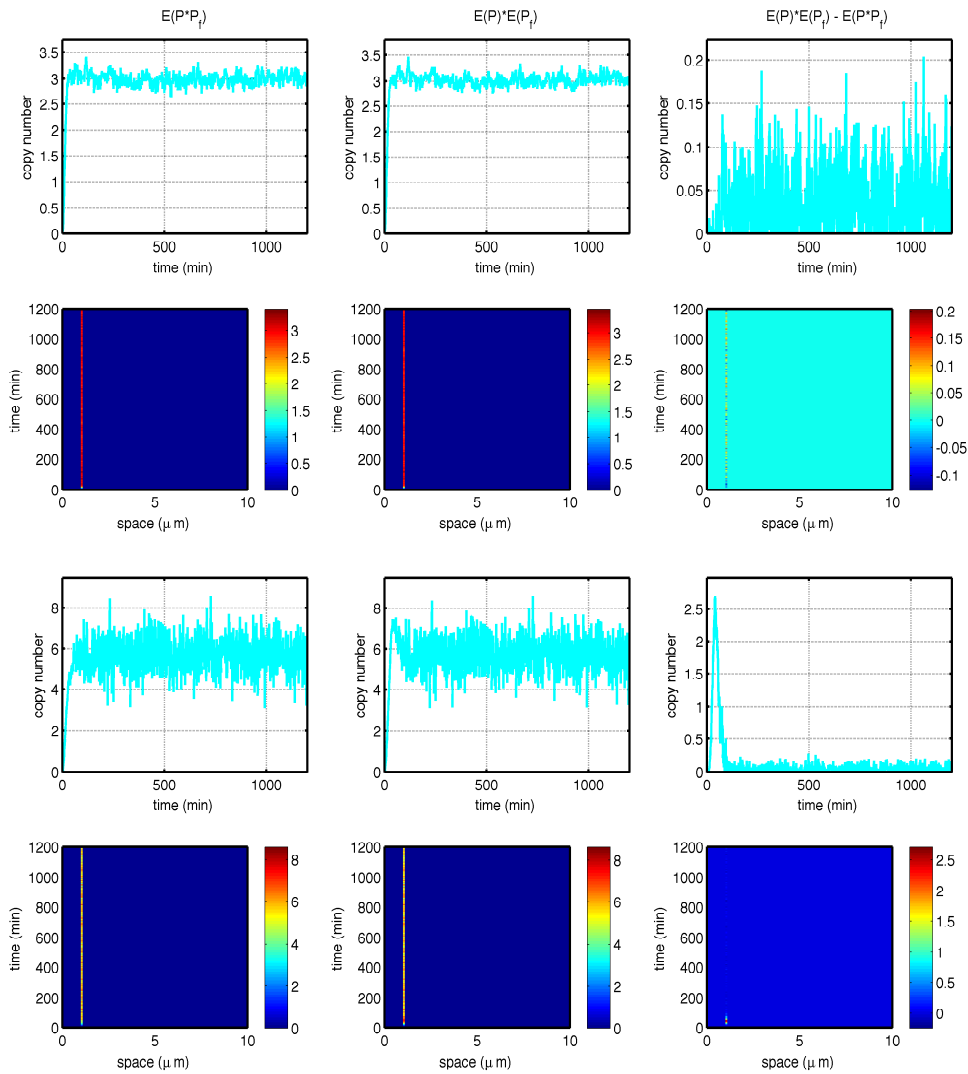

Fig. 5 Plots visualising the accuracy of the first order moment closure approximation. Parameters are stated in Table 1 with $\gamma = 100$ for rows one and two (representing the negative feedback case) and $\gamma = 0.1$ for rows three and four (representing the positive feedback case). The first column displays plots of the expected value of protein and the free promoter species for 1000 realisations of the spatial stochastic model. The first row shows the temporal evolution which is calculated by summing over individual compartments which comprise the domain. The second row shows the spatio-temporal evolution. In the second column, the same plots are shown but this time for the expected value of protein multiplied by the expected value of the free promoter (i.e., the approximation we use). The third column shows the error in making this moment closure approximation, both temporally (first or third row) and spatially (second or fourth row).

Hence, using the above first order moment closure approximation the closed system of equations approximating the stochastic means can be written as:

$$\frac{\partial \langle \mathbf{p} \rangle (x,t)}{\partial t} = D\nabla^2 \langle \mathbf{p} \rangle (x,t) + \alpha_p \langle \mathbf{m} \rangle (x,t) \chi_{[c,L]} - \mu_p \langle \mathbf{p} \rangle (x,t), \tag{1}$$

$$\frac{\partial \langle \mathbf{m} \rangle (x,t)}{\partial t} = D\nabla^2 \langle \mathbf{m} \rangle (x,t) + \alpha_m \left[\langle \mathbf{f} \rangle (x,t) + \frac{F_0 - \langle \mathbf{f} \rangle (x,t)}{\gamma} \right] \delta_{x_m}(x) - \mu_m \langle \mathbf{m} \rangle (x,t), \tag{2}$$

$$\frac{\partial \langle \mathbf{f} \rangle (x,t)}{\partial t} = \left[k_2 (F_0 - \langle \mathbf{f} \rangle (x,t)) - k_1 \langle \mathbf{p} \rangle (x,t) \langle \mathbf{f} \rangle (x,t) \right] \delta_{x_m}(x), \tag{3}$$

subject to zero-flux boundary conditions and zero initial conditions (except $\langle f_{x_m} \rangle (0) = F_0$).

It is notable that upon making the approximation that the promoter occupancy switching dynamics occur on a much faster time scale than diffusive transport and kinetics of protein and mRNA species, we can make a pseudo-steady state approximation and obtain

$$\langle \mathbf{f} \rangle (x,t) \sim \frac{F_0}{1 + \frac{\langle \mathbf{p} \rangle (x,t)}{p_0}}, \tag{4}$$

where $p_0 = k_2/k_1$. Upon substitution for $\langle \mathbf{f} \rangle (x,t)$ in equation (2)

$$\frac{\partial \langle \mathbf{p} \rangle (x,t)}{\partial t} = D\nabla^2 \langle \mathbf{p} \rangle (x,t) + \alpha_p \langle \mathbf{m} \rangle (x,t) \chi_{[c,L]} - \mu_p \langle \mathbf{p} \rangle (x,t), \tag{5}$$

$$\frac{\partial \langle \mathbf{m} \rangle (x,t)}{\partial t} = D\nabla^2 \langle \mathbf{m} \rangle (x,t) + \alpha_m F_0 \left[\frac{1 + \frac{\langle \mathbf{p} \rangle (x,t)}{\gamma p_0}}{1 + \frac{\langle \mathbf{p} \rangle (x,t)}{p_0}} \right] \delta_{x_m}(x) - \mu_m \langle \mathbf{m} \rangle (x,t), \tag{6}$$

It is worth noting that when γ is sufficiently large (i.e. the case of strong negative feedback), this system of PDEs is equivalent, up to the value of a Hill exponent, to the phenomenologically derived system of equations considered in Sturrock et al. [2011, 2012].

4.1 Numerical simulations of the partial differential equation system

We used a forward time, centred space finite difference numerical approximation of the PDE system given by Equations (1) - (3). This method of numerical approximation is the natural choice to make given the form the stochastic mean evolution equations take (see derivation of equations in previous section). Numerical simulations of this system of equations are presented in Figure 6. Parameter values are as stated in Table 1 with $\gamma = 100$ for rows one and two (negative feedback) and $\gamma = 0.1$ for rows three and four (positive feedback). The various protein solutions are displayed in the first column, mRNA solutions in the second column and the free promoter solutions in the third column. The temporal plot for the PDE approximation system is obtained by first integrating the numerical solution over the domain (black line) and the true solution (dashed coloured line) is obtained by averaging 1000 trajectories of the spatial stochastic model. The PDE system provides excellent agreement with the true mean behaviour of the system for both the positive and negative feedback cases (both spatially and temporally). The only noticeable minor disagreement is in the case of the mRNA temporal evolution for the negative feedback case, where there appears to be a small discrepancy. However, it is worth highlighting that we only took the average of 1000 realisations of the spatial stochastic model, and any minor discrepancies may disappear if we were to take more. Notice also that the free promoter evolution is particularly well approximated. As we mentioned previously, it is the free promoter evolution that drives the interesting behaviour in the individual trajectories and when we explore the parameter landscape (section 6) we will focus solely on the free promoter temporal evolution produced by the PDE system.
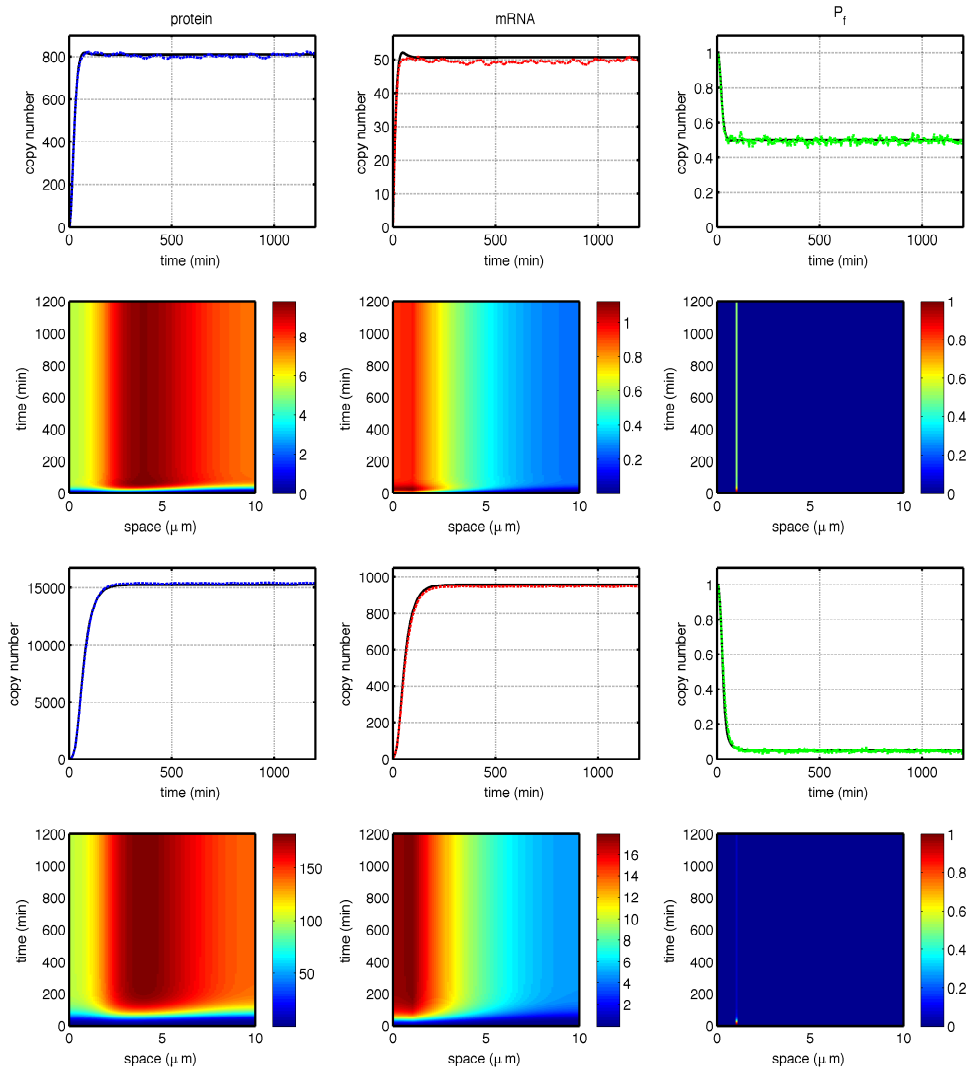
Fig. 6 Plots showing numerical solutions of the PDE system derived to approximate the stochastic means (Equations (1) - (3)). Parameters are stated in Table 1 with $\gamma = 100$ for rows one and two (negative feedback) and $\gamma = 0.1$ for rows three and four (positive feedback). Protein is displayed in the first column, mRNA in the second and the free promoter in the third. The temporal evolution is shown in the first and third rows and the spatio-temporal solution in the second and fourth rows. The dashed line in the temporal plots corresponds to spatial stochastic model trajectories obtained by summing over the domain after obtaining the mean from 1000 trajectories and the black line corresponds to the approximation obtained by integrating over the PDE solution. The spatio-temporal solution revealed in the second row can be compared with the fourth row of Figure 3 and the fourth row can be compared with the fourth row of Figure 4.

5 Comparison with spatially homogeneous case

In this section we consider the spatially homogeneous version of our GRN model. It is common practice when formulating mathematical models of GRNs to make this assumption [Ciliberto et al., 2005, Nguyen and Kulasiri, 2009]. We consider this version of our model so that we can examine whether or not it is a good approximation for modelling a GRN with a single feedback loop. It can be argued also from a biological perspective, that the spatially homogeneous model corresponds to a GRN in a prokaryotic cell. In prokaryotic cells, reactions are less compartmentalised and the processes of transcription and translation are coupled; that is, translation begins while the mRNA is still being transcribed. The chemical master equation for the spatially homogeneous case can be written simply by omitting the diffusion terms and no longer accounting for the localisation of certain reactions, i.e.,

$$\frac{\partial P(p,m,f,t)}{\partial t} = \alpha_p \Big[mP(p-1,m,f,t) - mP(p,m,f,t) \Big]$$
$$+ \mu_p \Big[(p+1)P(p+1,m,f,t) - pP(p,m,f,t) \Big]$$
$$+ \alpha_m \Big[fP(p,m-1,f,t) - fP(p,m,f,t) \Big]$$
$$+ \frac{\alpha_m}{\gamma} \Big[(F_0-f)P(p,m-1,f,t) - (F_0-f)P(p,m,f,t) \Big]$$
$$+ \mu_m \Big[(m+1)P(p,m+1,f,t) - mP(p,m,f,t) \Big]$$
$$+ k_1(f+1)pP(p,m,f+1,t) - k_1 f pP(p,m,f,t)$$
$$+ k_2((F_0-f)+1)P(p,m,f-1,t) - k_2(F_0-f)P(p,m,f,t).$$

In order to compare the behaviour of this model with the spatially inhomogeneous case, we must first derive the corresponding stochastic mean evolution equations. The stochastic means are defined by

$$\langle p \rangle(t) = \sum_p \sum_m \sum_f pP(p,m,f,t)$$

$$\langle m \rangle(t) = \sum_p \sum_m \sum_f mP(p,m,f,t)$$

$$\langle f \rangle(t) = \sum_p \sum_m \sum_f fP(p,m,f,t)$$

In a similar manner to section 4, we derive the following system of ODEs describing the time evolution of the mean behaviour:

$$\frac{d\langle p \rangle(t)}{dt} = \alpha_p \langle m \rangle(t) - \mu_p \langle p \rangle(t), \tag{7}$$

$$\frac{d\langle m \rangle(t)}{dt} = \alpha_m \Big[\langle f \rangle(t) + \frac{F_0 - \langle f \rangle(t)}{\gamma} \Big] - \mu_m \langle m \rangle(t), \tag{8}$$

$$\frac{d\langle f \rangle(t)}{dt} = k_2(F_0 - \langle f \rangle(t)) - k_1 \langle p \rangle(t)\langle f \rangle(t). \tag{9}$$

These equations are subject to zero initial conditions (except $\langle f \rangle(0) = F_0$). Note that we have used the same first order moment closure approximation that we used in the spatially explicit case. As in the spatially explicit case, we present plots which demonstrate how accurate this approximation is in Figure 7. For the spatially homogeneous case, the moment closure approximation is poorer. The negative feedback case has persistent spiking errors throughout the time series whereas the positive feedback case has a large transient error at the beginning that fades away. However, these errors do not seem to influence the overall solution much, as we will show in the next section.

5.1 Numerical simulations of ordinary differential equation system

Equations (7) - (9) were solved numerically using the Matlab stiff ODE solver 'ode15s'. We present numerical simulations of the ODE system and individual trajectories from the spatially homogeneous stochastic case in Figure 8. The same parameters that were used in the spatially dependent case were also used here. It is worth noting that the parameter k_1 now takes the form $k_1 = \frac{c_1}{n_A L}$ because we are no longer localising the promoter occupation reaction to a single compartment and similarly, $\alpha_p = \frac{a_p}{L}$. As in the other Figures, we present the protein plots in the first column, mRNA in the second column and the free promoter in the third column. As in Figure 6, the first two rows correspond to the negative feedback case ($\gamma = 100$) and the third and fourth rows correspond to the positive feedback case ($\gamma = 0.1$). The simulation results are in remarkably good agreement with the spatially dependent case (both qualitatively and quantitatively). This is a feature of the spatial stochastic model which is in contrast to previously studied PDE models presented in Sturrock et al. [2011], where it was found that the spatially
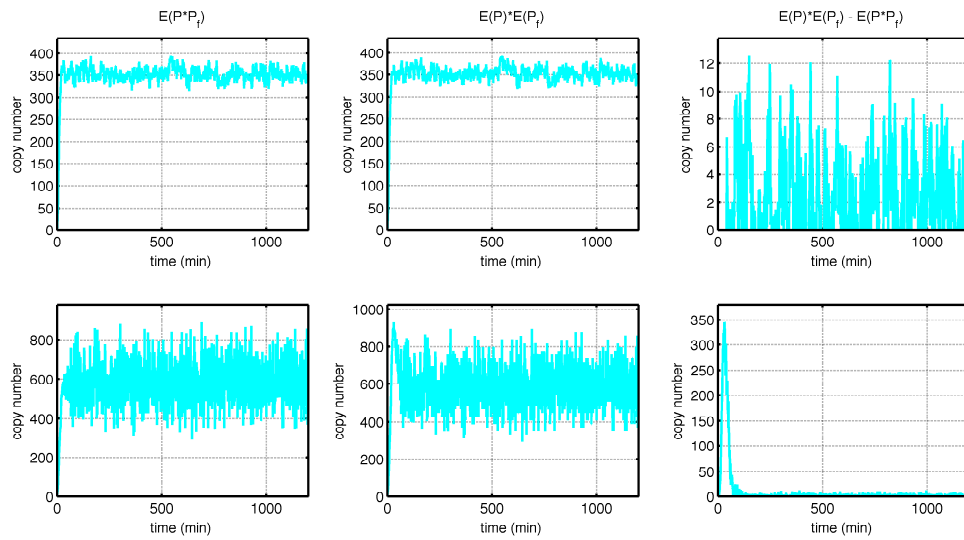
Fig. 7 Plots visualising the accuracy of the first order moment closure approximation for the spatially homogeneous case. Parameters are stated in Table 1 with $\gamma = 100$ for the first row and $\gamma = 0.1$ for the second row. The first column displays plots of the expected value of protein and the free promoter species for 1000 realisations of the spatially homogeneous stochastic model. In the second column, the same plots are shown but this time for the expected value of protein multiplied by the expected value of the free promoter (i.e., the approximation we use). The third column shows the error in making this moment closure approximation.

homogeneous case yielded qualitatively different results. If we scrutinise the time series presented in Figure 8 and compare them with those presented in Figures 3 and 4 (rows two and four) we can find minor quantitative differences. For example, in the negative feedback case we find the free promoter is occupied more often and in the positive feedback case we find a higher quantity of protein. Both these differences can be attributed to the lack of a delay in the spatially homogeneous case. In other words, it does not take time for the protein to reach the promoter to occupy it, so it is occupied more often and also it does not take time for mRNA to reach the cytoplasm to produce protein. Hence, one might expect the differences between the spatially homogeneous case and the spatially dependent cases to be amplified if the diffusion coefficient were smaller.
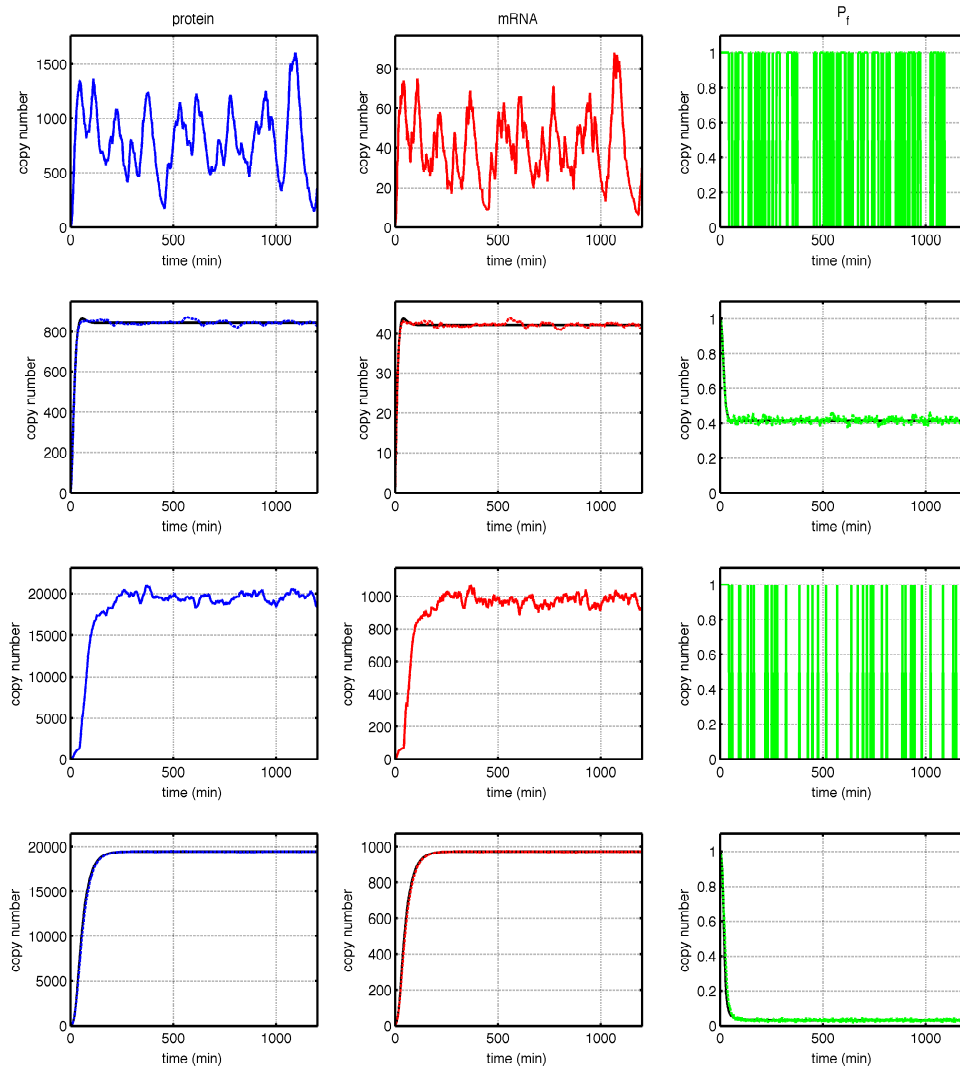
Fig. 8 Plots showing the temporal evolution of the spatially homogeneous genetic feedback loop model. Parameters are stated in Table 1 with $\gamma = 100$ for rows one and two (negative feedback) and $\gamma = 0.1$ for rows three and four (positive feedback). Protein is displayed in the first column, mRNA in the second and the free promoter in the third. The first and third rows display a single trajectory of the spatially homogeneous stochastic model. The second and fourth rows show the mean of 1000 trajectories of the spatially homogeneous stochastic model plotted along with the numerical solution of the ODEs (Equations (7) - (9)) which approximate the stochastic means. The dashed line corresponds to the spatially homogeneous stochastic model trajectories obtained by summing over the domain after obtaining the mean from 1000 trajectories and the black line corresponds to the approximation obtained from the ordinary differential equation solution. In order to compare the spatially homogeneous and spatially explicit cases, these plots can be compared with rows 1 and 3 of Figures 3 and 4.

6 Classifying model behaviour in different regions of the parameter space

We present two complementary methods of classifying model behaviour in different regions of the parameter space: k-means clustering and a steady-state analysis.

6.1 k-means clustering

Data clustering can be described as the unsupervised classification of patterns (observations, data items, or time series) into groups (clusters). The desired outcome is that the objects within a group should be similar (or related) to one another and different from (or unrelated to) the objects in other groups. The greater the similarity (or homogeneity) within a group and the greater the difference between groups, the 'better' or more distinct the clustering.

In this section, we apply data clustering methods to classify the time series data of the state of the free promoter generated by solving the mean field models presented in the previous sections for a range of parameter values. We wish to highlight that without our systems of equations for the stochastic means, this problem would be computationally intractable. This would be particularly true when exploring regions of the parameter space with fast diffusion coefficients – where Gillespie's algorithm would be markedly slower. We focus on the data concerning the state of the free promoter, since the state of the promoter clearly drives the dynamical evolution of the system (see Figures 3, 4 and 8). As the parameter space is 11-dimensional (cf. Table 1), it would be too computationally expensive to explore the complete parameter space unless a very coarse sampling was used and presenting our results would be nontrivial (but there is no reason why one could not do it). Instead, we focus on a subset of the parameter space. As we stated before, certain parameters in our model are easier to measure than others. For our GRN model, the production rates (transcription and translation), degradation rates, position of the cytoplasm or promoter and size of the cell can all be measured, but the parameters about which we have the least information, are the binding and unbinding rates of the transcription factor to the promoter [Sturrock et al., 2013]. Furthermore, it has been highlighted in various works how important the rate of switching between promoter states is to the overall behaviour of gene regulatory networks [Kepler and Elston, 2001, Lipniacki et al., 2006]. Therefore, as an exemplar of how powerful a tool data clustering can be in the classification of model time series, we explore the k_1, k_2 parameter space. In the PDE model we also explore the effect of changing the parameter, D, the diffusion coefficient – we do this in an effort to assess whether or not one can make the 'well-mixed assumption' with confidence in its accuracy. For simplicity, we use k-means clustering and specify the number of clusters to be 3. There exist methods to estimate the 'correct' number of clusters (see R.Tibshirani et al. [2001] for example). However in order to compare the clustering between different model scenarios (positive vs. negative feedback, spatially dependent vs. spatially homogeneous) we wish to keep the number of clusters constant, and choose 3 clusters for each scenario.

We introduce here some of the basic notions underlying the classical k-means algorithm Kogan [2007], Liu et al. [2013]. In what follows, we consider a set of data

$$\mathscr{D} = \{x_1, x_2, \ldots, x_n\} \subset \mathbb{R}^m.$$

embedded in a Euclidean space. The output of a data clustering algorithm is a partition:

$$\Pi = \{\pi_1, \pi_2, \ldots, \pi_k\}, \tag{10}$$

where $k \leq n$ and each π_i is a nonempty subset of $\mathscr{D}$. Π is a partition of $\mathscr{D}$ in the sense that

$$\bigcup_{i \leq k} \pi_i = \mathscr{D} \text{ and } \pi_i \cap \pi_j = \varnothing \text{ for all } i \neq j. \tag{11}$$

In this context, the elements of Π are usually referred to as clusters. In practice, one is interested in partitions of $\mathscr{D}$ that satisfy specific requirements, usually expressed in terms of a distance function $d(\cdot, \cdot)$ that is defined on the background Euclidean space.

The classical k-means algorithm is based on reducing the notion of a cluster π_i to that of a cluster representative or centroid $c(\pi_i)$ according to the relation

$$c(\pi_i) = \arg\min_{y \in \mathbb{R}^m} \sum_{x \in \pi_i} d(x, y). \tag{12}$$

In its simplest form, k-means consists of initializing a random partition of $\mathscr{D}$ and subsequently updating iteratively the partition Π and the centroids $\{c(\pi_i)\}_{i \leq k}$ through the following two steps Kogan [2007]:

(a) Given $\{\pi_i\}_{i \leq k}$, update $\{c(\pi_i)\}_{i \leq k}$ according to (12).
(b) Given $\{c(\pi_i)\}_{i \leq k}$, update $\{\pi_i\}_{i \leq k}$ according to centroid proximity, i.e., for each $i \leq k$,

$$\pi_i = \{x \in \mathscr{D} \mid d(c_i, x) \leq d(c_j, x) \text{ for each } j \leq k\}$$

Step (a) in the above iteration requires to solve a constrained optimization problem, the difficulty of which depends on the distance function $d(\cdot, \cdot)$. In this paper, the data set $\mathscr{D}$ consists of simulated time series. The distance function adopted is the usual Euclidean metric in $\mathbb{R}^m$, where m is the size of each time series in $\mathscr{D}$.

The result of clustering the raw time series of the PDE model (Equations (1) - (3)) is presented in Figure 9 for the negative feedback case (i.e., $\gamma = 100$) and in Figure 10 for the positive feedback case (i.e., $\gamma = 0.1$). We sampled 100,000 time series generated from the k_1, k_2, D parameter space with the values shown on the axes of the plots in the first row of Figures 9 and 10. All other parameters are as stated in Table 1. Figure 9 reveals that the parameter space is occupied mostly by the blue cluster, which corresponds to those free promoter time series which are close to centroid 1 (where the free promoter reaches a steady state which is occupied approximately 30% of the time). The red cluster corresponds to those time series which are close to centroid 3 which contains time series where the free promoter is occupied approximately 70% of the time (this appears to be the smallest cluster). The green cluster consists of those free promoter time series which are close to the second centroid, which is completely free for the 1200 minute time period. Consistent with intuition the promoter is completely free for time series corresponding to very low diffusion coefficients (this is made clear in the second row where we show slices of the parameter space corresponding to different values of D). As the diffusion coefficient is increased, fewer and fewer time series are clustered into the second (green) cluster. In the third row we present plots of the centroids as computed by the k-means algorithm. In the final row, we present plots of single trajectories from the spatial stochastic model for parameter sets corresponding to the different clusters. Our main finding from clustering the data in the negative feedback case is that we have identified regions of the parameter space where we find oscillatory dynamics in individual trajectories of the spatial stochastic model (the red and blue clusters).

Figure 10 reveals that the parameter landscape for the positive feedback case takes a different form when compared to the negative feedback case, though there are some similarities. We note that two of the centroids have changed. The first centroid corresponds to those free promoter time series which reach steady states that are occupied approximately 50% of the time, the second centroid is the same as in the negative feedback case (100% free) and the third centroid corresponds to time series which are occupied approximately 90% of the time. The parameter landscape is dominated by the red cluster which is in stark contrast to the negative feedback case. It appears that the structure of the positive feedback loop favours an occupied promoter steady state whereas the negative feedback loop favours a partially occupied steady state. As we noted for the negative feedback case, we find that as the diffusion coefficient is increased, the size of the green cluster shrinks (see second row of plots). In fact, for large values of D, the green cluster is even smaller for the positive feedback case. The blue cluster highlights a region of the parameter space which we did not expect to find for the positive feedback case, i.e., parameter sets which yield time series which are occupied approximately 50% of the time and which we would expect to yield oscillatory dynamics in the mRNA and protein species. Indeed, we show sample trajectories from the spatial stochastic model in the final row of Figure 10 and one can clearly see a large amount of switching between free and occupied states for an example trajectory from the blue cluster. Moreover, we found noisy oscillatory dynamics in the corresponding mRNA and protein time series, which are similar to the oscillations found in the negative feedback case but the mean copy number is much higher for both mRNA and protein species. Interestingly, the blue clusters for the negative and positive feedback cases overlap which suggests that the same parameter sets could yield oscillatory dynamics in the negative and positive feedback cases. To test this, for Figure 13 we chose a parameter set from a region of the parameter space which was clustered into the blue cluster for the positive and negative feedback cases, specifically $k_1 = 1.75 \times 10^{-2} (\text{min}^{-1})$, $k_2 = 0.95$ (min^{-1}) and $D = 7.29 \times 10^{-12} (\text{m}^2 \text{min}^{-1})$. Rows one and two correspond to the negative and positive feedback cases respectively. For this choice of parameters, it is clear from the sample trajectories of the spatial stochastic model that all model species exhibit oscillatory dynamics for both negative and positive feedback cases.

Figures 11 and 12 show the results for the negative and positive feedback cases respectively of clustering the time series produced by the ODE system (Equations (7) - (9)). We choose equivalent values of k_1 and exactly the same values of k_2, and once again these values are shown on the axes of the plots. The results of the spatially homogeneous case are in remarkably good agreement with the spatially dependent case when the diffusion coefficient is fast (compare the third column, second row of Figures 9 and 10 with the first row of Figures 11 and 12). However, if the diffusion coefficient is one order of magnitude slower, then the parameter landscape appears to change significantly and no longer agrees well with the spatially homogeneous case. Hence, the accuracy of the
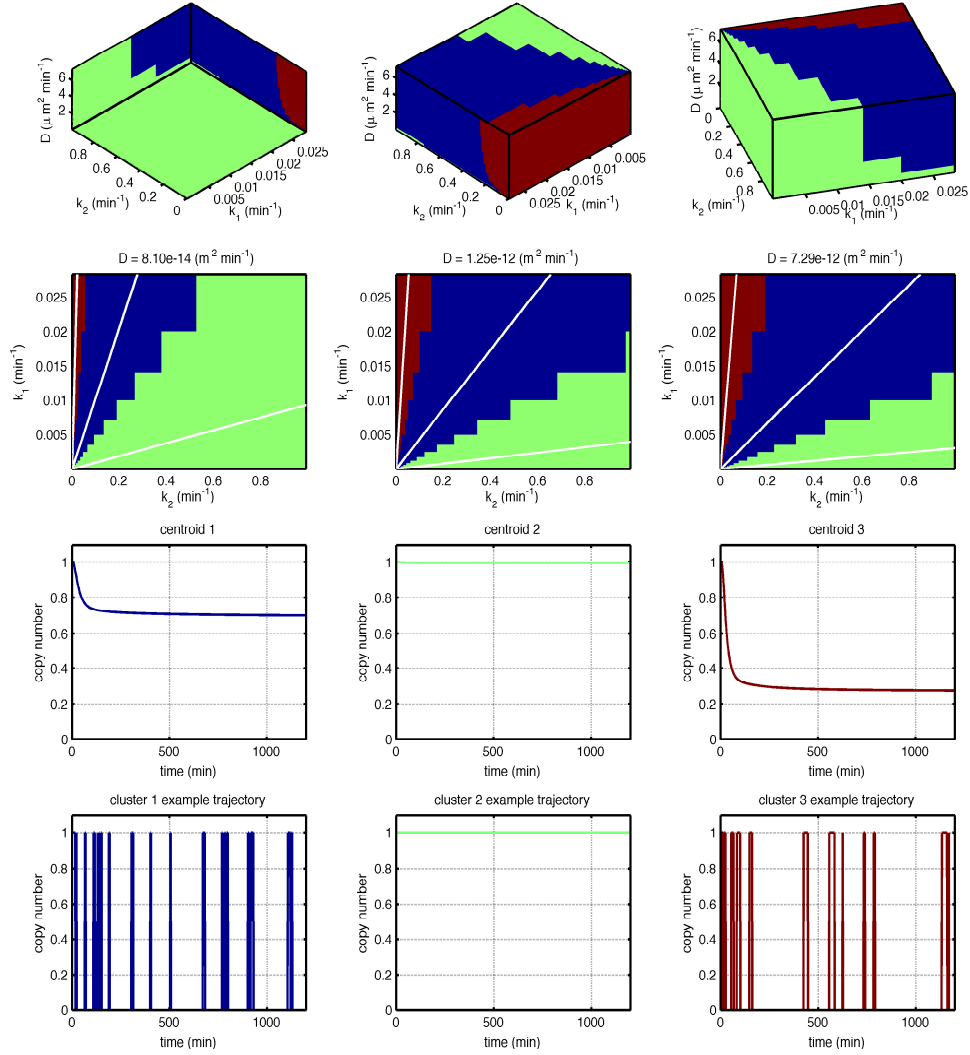
Fig. 9 Plots showing the result of k-means clustering of time series produced by the partial differential equation model for the negative feedback case. The values of D we use are sampled such that for $u = 1$ to 10, $D = 1.00 \times 10^{-15} + (u-1)^3 \times 10^{-14}$ (m^{2}min^{-1}), while values for k_1 were sampled such that for $v = 1$ to 100, $k_1 = \frac{1.42 \times 10^{v}}{n_A h}$ (min^{-1}), and values of k_2 were sampled such that for $w = 1$ to 100, $k_2 = w(w-1) \times 10^{-4}$ (min^{-1}). All other parameters are as per Table 1 with $\gamma = 100$. The first row shows the full three dimensional parameter space (from 3 different angles) and the result of the k-means algorithm dividing it into 3 clusters (depicted in red, green and blue). The second row shows slices of the parameter space corresponding to different values of D (with the values of D indicated in the title). The white lines correspond to the parameter values that produce the cluster centroids as revealed by a steady state analysis, see equation (17). The third row shows the time series corresponding to the cluster centroids which are coloured to match the different clusters shown in the first two rows. The final row shows example trajectories of the spatial stochastic model for parameters corresponding to the 3 different clusters. The example trajectory from cluster 1 uses the parameters $D = 2.71 \times 10^{-13}$ (m^{2}min^{-1}), $k_1 = 0.02$ (min^{-1}), and $k_2 = 0.4$ (min^{-1}). The example trajectory from cluster 2 uses the parameters $D = 1.00 \times 10^{-15}$ (m^{2}min^{-1}), $k_1 = 1.67 \times 10^{-2}$ (min^{-1}), and $k_2 = 0.1$ (min^{-1}). The example trajectory from cluster 3 uses the parameters $D = 7.29 \times 10^{-12}$ (m^{2}min^{-1}), $k_1 = 0.025$ (min^{-1}), and $k_2 = 0.01$ (min^{-1}).

spatially homogeneous approximation is dependent on the value of the diffusion coefficient of mRNA and protein molecules. If it is known that the diffusion coefficients of mRNA and protein molecules for a particular GRN are greater than or equal to 7.29×10^{-12}m^{2}min^{-1}, then our results suggest that the spatially homogeneous approximation is an accurate representation of a one dimensional spatial model of a cell of length $10\mu m$. We highlight the fact that the cell size may considerably vary (see Alberts et al. [2008]), and further work would have to done to see whether the spatially homogeneous approximation is accurate for larger 3-dimensional cells. Unlike similar GRN models based on PDEs which use a Hill-function representation of the negative feedback (see Sturrock et al. [2011]), the spatially homogeneous model presented in this paper is in good qualitative agreement with the
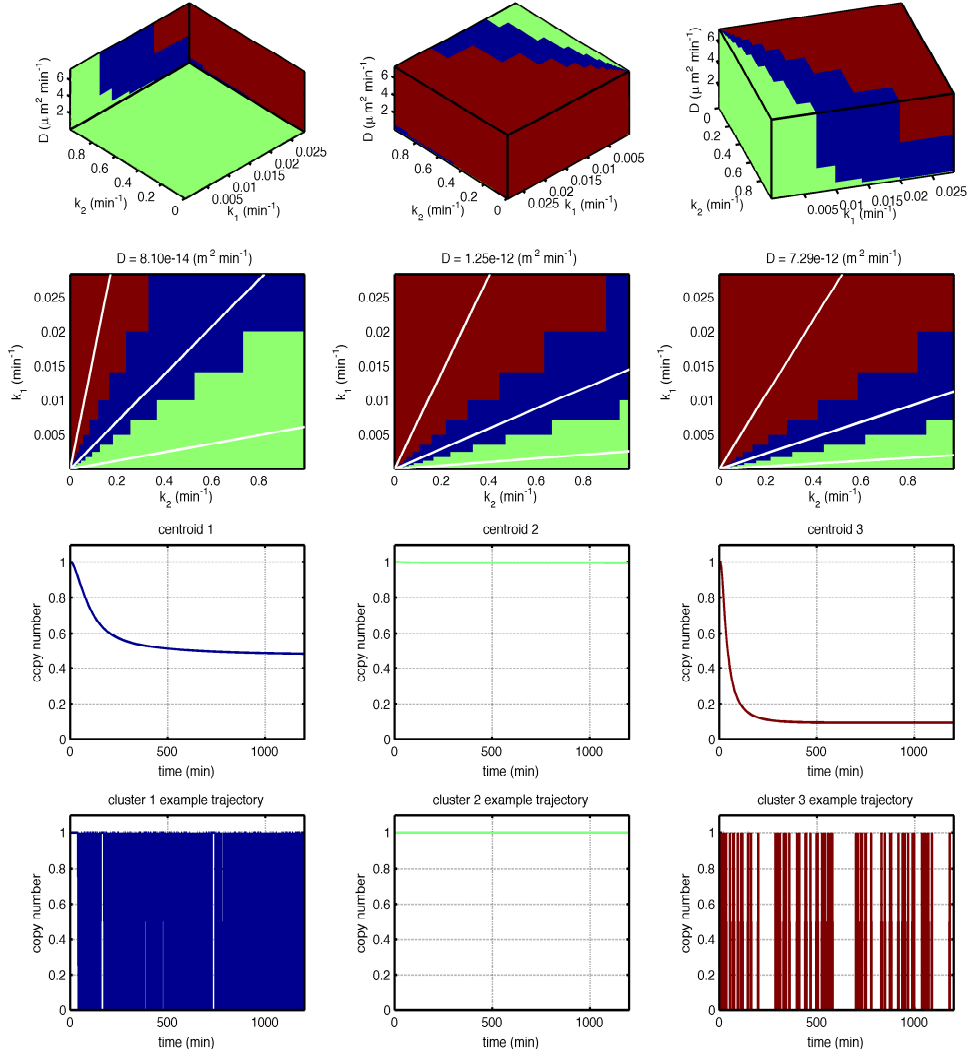
Fig. 10 Plots showing the result of k-means clustering of time series produced by the partial differential equation model for the positive feedback case. The values of D we use are sampled such that for $u = 1$ to 10, $D = 1.00 \times 10^{-15} + (u-1)^3 \times 10^{-14}$ ($m^2 min^{-1}$), while values for k_1 were sampled such that for $v = 1$ to 100, $k_1 = \frac{1.42 \times 10^v}{n_A h}$ (min^{-1}), and values of k_2 were sampled such that for $w = 1$ to 100, $k_2 = w(w-1) \times 10^{-4}$ (min^{-1}). All other parameters are as per Table 1 with $\gamma = 0.1$. The first row shows the full three dimensional parameter space (from 3 different angles) and the result of the k-means algorithm dividing it into 3 clusters (depicted in red, green and blue). The second row shows slices of the parameter space corresponding to different values of D (with the values of D indicated in the title). The white lines correspond to the parameter values that produce the cluster centroids as revealed by a steady state analysis, see equation (17). The third row shows the time series corresponding to the cluster centroids which are coloured to match the different clusters shown in the first two rows. The final row shows example trajectories of the spatial stochastic model for parameters corresponding to the 3 different clusters. The example trajectory from cluster 1 uses the parameters $D = 2.71 \times 10^{-13}$ ($m^2 min^{-1}$), $k_1 = 0.5$ (min^{-1}), and $k_2 = 0.01$ (min^{-1}). The example trajectory from cluster 2 uses the parameters $D = 1.00 \times 10^{-15}$ ($m^2 min^{-1}$), $k_1 = 1.67 \times 10^{-2}$ (min^{-1}), and $k_2 = 0.1$ (min^{-1}). The example trajectory from cluster 3 uses the parameters $D = 7.29 \times 10^{-12}$ ($m^2 min^{-1}$), $k_1 = 0.02$ (min^{-1}), and $k_2 = 0.2$ (min^{-1}).

spatially dependent model. The spatially homogeneous equivalent of the phenomenologically derived PDE model was not able to produce oscillatory dynamics under any parameter regime and in fact corresponds to the model first derived in Goodwin [1965]. We show examples of the oscillatory dynamics that the spatially homogeneous model can produce in rows three and four of Figure 13. We chose the same values as we did for the spatially explicit model, i.e., the values which produce oscillatory dynamics for both the negative and positive feedback cases, which for the spatially homogeneous model are: $k_1 = 1.49 \times 10^{-4}$ (min^{-1}), $k_2 = 0.95$ (min^{-1}). By comparing rows one and three and two and four of Figure 13, it is easy to see good qualitative agreement between the trajectories of the spatial stochastic model and the spatially homogeneous stochastic model.
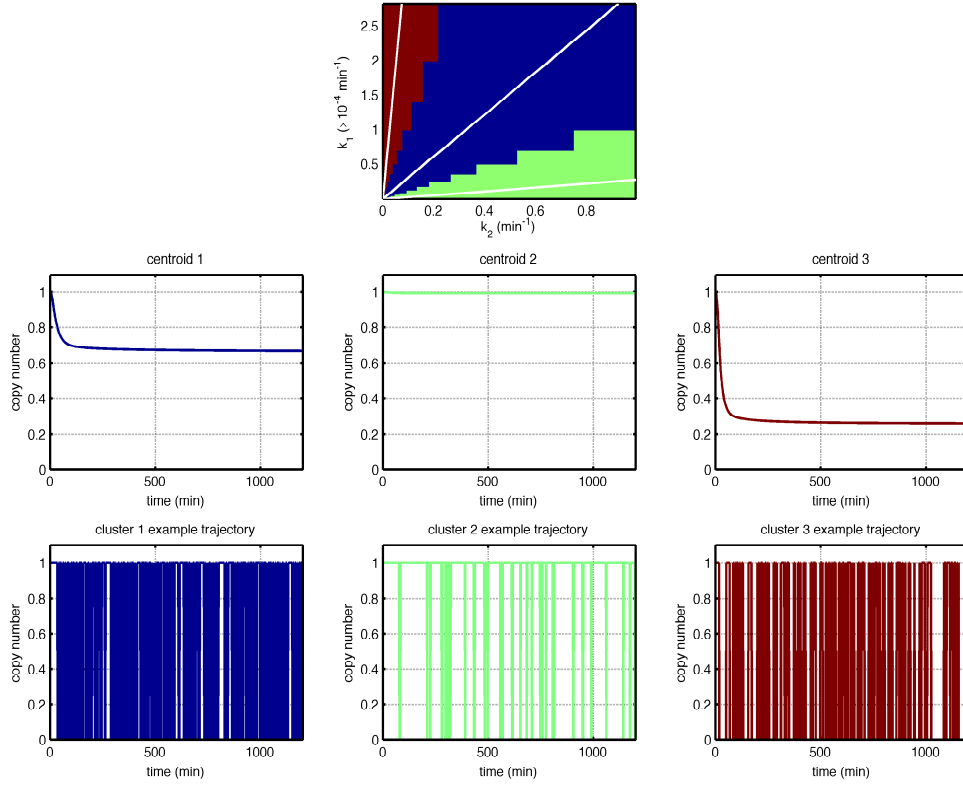
Fig. 11 Plots showing the result of k-means clustering of time series produced by the ordinary differential equation model for the negative feedback case. The values for k_1 were sampled such that for $v = 1$ to 100, $k_1 = \frac{1.42 \times 10^v}{n_A h}$ (min^{-1}) and values of k_2 were sampled such that for $k = 1$ to 100, $k_2 = w(w-1) \times 10^{-4}$ (min^{-1}). All other parameters are as per Table 1 with $\gamma = 100$. The first row shows the full parameter space and the result of the k-means algorithm dividing it into 3 clusters (depicted in red, green and blue). The white lines correspond to the parameter values that produce the cluster centroids as revealed by a steady state analysis, see equation (18). The second row shows the time series corresponding to the cluster centroids which are coloured to match the different clusters shown in the first row. The third row shows example trajectories computed using Gillespie's algorithm for parameters corresponding to the 3 different clusters. The example trajectory from cluster 1 uses the parameters $k_1 = 2.00 \times 10^{-4}(\text{min}^{-1})$ and $k_2 = 0.5$ (min^{-1}). The example trajectory from cluster 2 uses the parameters $k_1 = 0.20 \times 10^{-4}(\text{min}^{-1})$, and $k_2 = 0.8$ (min^{-1}). The example trajectory from cluster 3 uses the parameters $k_1 = 2.50 \times 10^{-4}(\text{min}^{-1})$ and $k_2 = 0.1$ (min^{-1}).

6.2 Steady-state analysis

Motivated by the results from the k-means clustering analysis, we sought to derive an expression that could relate the mean occupancy of the promoter with the given model parameters. Firstly, we partitioned the domain $[0, L]$ into three subdomains as follows: Ω_1, $(0 < x < x_m)$, Ω_2, $(x_m < x < c)$, and Ω_3, $(c < x < L)$. In Ω_1, we considered the steady-state solutions to Equations (1) - (3), for which $\langle m \rangle$ satisfies

$$D\nabla^2 \langle m \rangle - \mu_m \langle m \rangle = 0, \tag{13}$$

and has solution, assuming a no-flux boundary condition at $x = 0$, given by

$$\langle m \rangle = B \cosh(k_m x), \tag{14}$$

where $k_m = \sqrt{\mu_m/D}$ and B is an integration constant. Using similar arguments, we construct solutions for $\langle m \rangle$ in $\Omega_2 \bigcup \Omega_3$ and $\langle p \rangle$ in $\Omega_1 \bigcup \Omega_2$ and Ω_3 and obtain

$$\langle m \rangle = \begin{cases} B \cosh(k_m x); & 0 < x < x_m, \\ C \cosh(k_m(x-L)); & x_m < x < L, \end{cases} \tag{15}$$

$$\langle p \rangle = \begin{cases} A \cosh(k_p x); & 0 < x < c, \\ E \cosh(k_p(x-L)) + \frac{\alpha_p}{(\mu_p - \mu_m)} C \cosh(k_m(x-L)); & c < x < L, \end{cases} \tag{16}$$
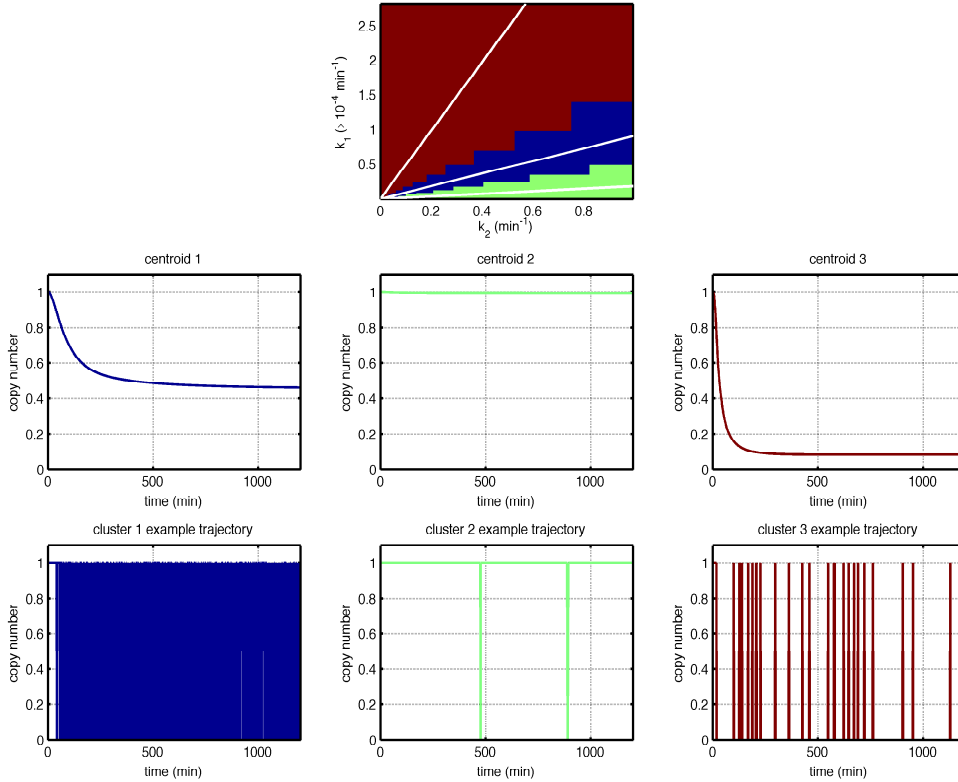
Fig. 12 Plots showing the result of k-means clustering of time series produced by the ordinary differential equation model for the positive feedback case. The values for k_1 were sampled such that for $j = 1$ to 100, $k_1 = \frac{1.42 \times 10^{\nu}}{n_A h}$ (min^{-1}) and values of k_2 were sampled such that for $k = 1$ to 100, $k_2 = w(w-1) \times 10^{-4}$ (min^{-1}). All other parameters are as per Table 1 with $\gamma = 0.1$. The first row shows the full parameter space and the result of the k-means algorithm dividing it into 3 clusters (depicted in red, green and blue). The second row shows the time series corresponding to the cluster centroids which are coloured to match the different clusters shown in the first row. The third row shows example trajectories computed using Gillespie's algorithm for parameters corresponding to the 3 different clusters. The example trajectory from cluster 1 uses the parameters $k_1 = 0.90 \times 10^{-4}(\text{min}^{-1})$ and $k_2 = 0.9$ (min^{-1}). The example trajectory from cluster 2 uses the parameters $k_1 = 0.01 \times 10^{-4}(\text{min}^{-1})$, and $k_2 = 0.9$ (min^{-1}). The example trajectory from cluster 3 uses the parameters $k_1 = 2.50 \times 10^{-4}(\text{min}^{-1})$ and $k_2 = 0.1$ (min^{-1}).

where B, C, A and E are integration constants and $k_p = \sqrt{\mu_p/D}$.

Upon enforcing continuity of $\langle p \rangle$ and $\langle p \rangle'$ at $x = c$, continuity of $\langle m \rangle$ at $x = x_m$ and conservation of mRNA copy number at $x = x_m$, we determine the integrations constants (see Appendix for details) and hence obtain the steady-state solution to Equations (1) - (3). In order to relate the promoter occupancy at steady-state to model parameters we define β to be the occupancy of the promoter site at steady-state, such that $\beta = \langle f \rangle$, and upon rearranging the constructed solutions find the relationship

$$
k_2 = \frac{\beta}{1 - \beta} \frac{k_1 h \alpha_m \alpha_p F_0 (\beta(\gamma - 1) + 1)}{\gamma(k_p^2 - k_m^2) k_m D^2} \frac{\cosh(k_m(c - L)) \cosh(k_p x_m)}{\cosh(k_p c) \cosh(k_m(x_m - L))} \times
$$
$$
\frac{\frac{k_m}{k_p} \tanh(k_m(c - L)) - \tanh(k_p(c - L))}{(\tanh(k_p c) - \tanh(k_p(c - L)))(\tanh(k_m(x_m - L)) - \tanh(k_m x_m) + k_m h)}, \tag{17}
$$

which specifies the value the parameter k_2 must attain, in terms of all the other model parameters, such that the promoter occupancy at steady-state is β. In Figures 9 and 10 we plot Equation (17) as a function of k_1 and D (using a white line), choosing β values consistent with the centroids, and demonstrate excellent agreement with the k-means clustering algorithm results. Within each cluster, the closer the parameters are chosen to the white line, the closer the time series produced by these parameters will be to the centroids presented in the third row. The
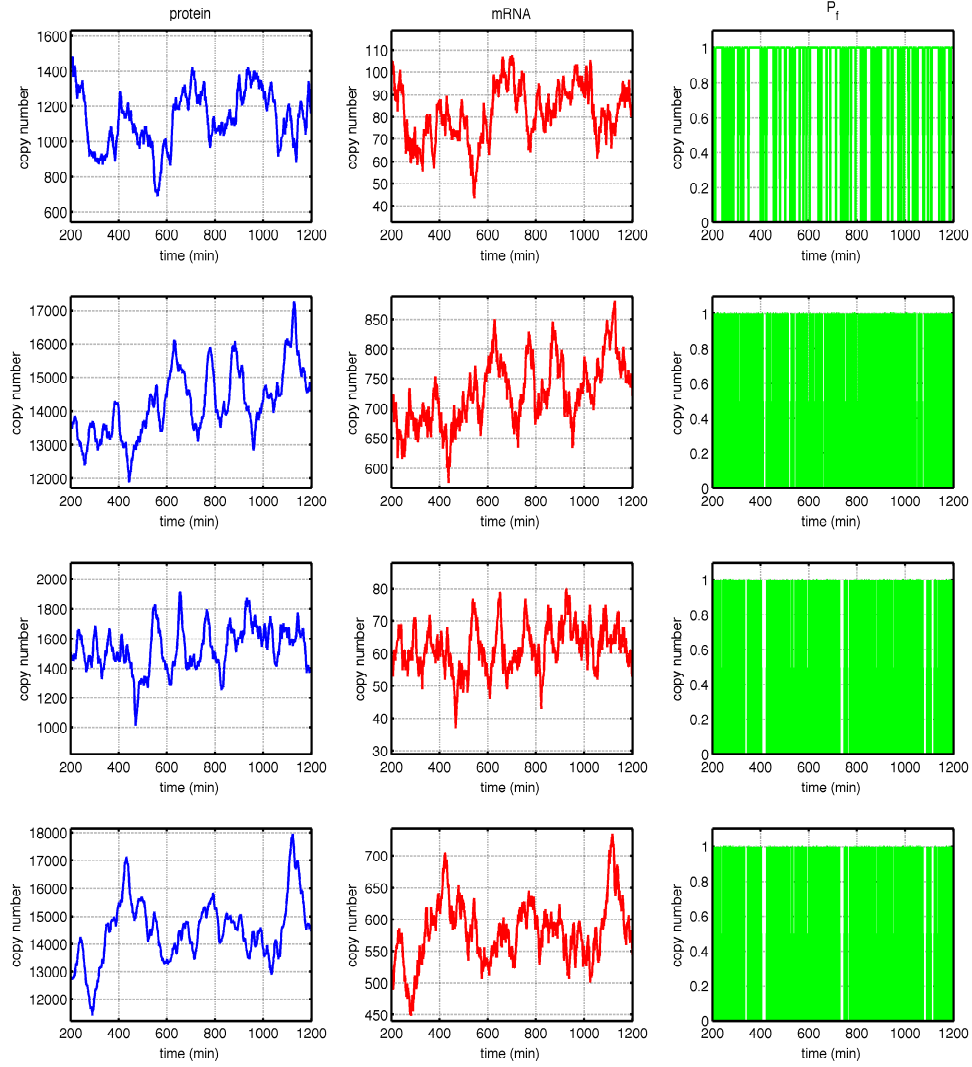
Fig. 13 Plots showing sample trajectories from the region of the parameter space which produces oscillations for the negative and positive feedback loops for both the spatially homogeneous and spatially explicit cases (as was discovered by k-means clustering). For rows one and two, parameters are as per Table 1 with $k_1 = 1.75 \times 10^{-2}(\mathrm{min}^{-1})$, $k_2 = 0.95 \ (\mathrm{min}^{-1})$, $D = 7.29 \times 10^{-12}(\mathrm{m}^2\mathrm{min}^{-1})$. We set $\gamma = 100$ for row one and $\gamma = 0.1$ for row two. The first row shows the temporal evolution of a single trajectory from the spatial stochastic model for the protein, mRNA and free promoter species respectively, which were obtained by summing over the individual compartments comprising the domain. The second row shows a single trajectory from the spatial stochastic model for the positive feedback case. For rows three and four, parameters are as per Table 1 with $k_1 = 1.49 \times 10^{-4}(\mathrm{min}^{-1})$ and $k_2 = 0.95 \ (\mathrm{min}^{-1})$. We set $\gamma = 100$ for row three and $\gamma = 0.1$ for row four. The third row shows the temporal evolution of a single trajectory from the spatially homogeneous stochastic model for the protein, mRNA and free promoter species respectively. Finally, the fourth row shows a single trajectory from the spatially homogeneous stochastic model for the positive feedback case.

spatially homogeneous case can be examined upon taking the limit of $D \to \infty$ in equation (17) and we find that

$$k_2 = \frac{\beta}{1-\beta} \frac{k_1 \alpha_m \alpha_p F_0(\beta(\gamma-1)+1)}{\gamma \mu_m \mu_p}. \tag{18}$$

In Figures 11 and 12 we plot Equation (18) as a function of k_1 (using a white line) and, again, demonstrate excellent agreement with the k-means clustering algorithm results. Again, the values of β used to produce the white lines are consistent with the cluster centroids (presented in the second row) and hence, within each cluster, the closer the parameters are chosen to the white line, the closer the time series produced by these parameters will be to the centroids.

7 Discussion

Gene regulatory networks (GRNs) play an important role in many key intracellular signalling processes and transcription factors (proteins which control the expression of genes) are key components of GRNs. The GRNs of many well studied organisms appear to be made up of a small set of recurring regulation patterns, called network motifs. Two well known and ubiquitous examples of network motifs are negative feedback loops and positive feedback loops, which serve as building blocks for more complex GRNs and interactions between multiple transcription factors. The mathematical modelling of single positive and negative feedback loops has up until now been dominated by phenomenologically derived ODE models. However, in reality, GRNs involving transcription factors are inherently stochastic and spatial systems, which has led us to develop and investigate a spatial stochastic model of a generic single feedback loop GRN in this paper.

We derived a system of equations which govern the stochastic means for our model species. Using a statistical independence assumption for the only nonlinear reaction, we were able to close the system of equations. We highlighted the fact that our mass action derived model does not resemble previously studied genetic feedback loop models as it does not rely on a Hill function to capture the feedback. Also, given the structure of our model, by changing a single parameter, the model can represent either a negative feedback loop ($\gamma < 1$) or a positive feedback loop ($\gamma > 1$). Numerical simulations were presented for individual trajectories of the spatial stochastic model as well as for the mean behaviour of the different molecular species. For our sample parameter set (presented in Table 1), we found that the mean behaviour for the negative feedback case was qualitatively different from the individual trajectories. However, for the positive feedback case we found good agreement between the individual trajectories and the mean behaviour. We found that the PDE system governing the stochastic means approximated the true means (computed by taking the mean of 1000 trajectories from the spatial stochastic model) very closely for both the positive and negative feedback cases.

A common assumption in formulating mathematical models of GRNs is that mRNA and protein molecules diffuse sufficiently quickly so that they exist in a state of spatial homogeneity. Under such an assumption, the localisation of reactions is not important, nor is it necessary to account for diffusion (or other spatial transport). Hence, in order to examine more carefully the accuracy of this modelling assumption, we derived a spatially homogeneous version of our GRN model. As for our spatially explicit model, we derived equations which governed the evolution of the stochastic means of the model species. In order to close the resultant system of ODEs, we made the same moment closure approximation that we used for the PDE system. For the sample parameter set, we found qualitative agreement between the spatially homogeneous and spatially dependent cases. We noted some quantitative differences in the time series produced by two models which we attributed to the delay caused by the diffusion of molecules between distinct reaction sites (e.g. the time it takes protein to diffuse to the promoter site and occupy it).

The application of data clustering for the classification of time series is usually reserved for gene expression data [Pirim et al., 2012, Liao, 2005]. In this paper, we showed how useful clustering analysis can be as an exploratory tool of the parameter space of dynamical systems. In particular, we explored a region of the parameter space for which experimental measurements are lacking – the k_1, k_2, D parameter space. We focused on clustering time series of the free promoter species (produced by our mean field models), as we found that the evolution of the mRNA and protein species was entirely dependent on the state of the free promoter species. We found the parameter landscape to take a different form for the positive and negative feedback cases, though we noted that there were some intersecting regions. In general, the parameter landscape of the positive feedback case seemed to favour an almost 100% occupied promoter state. Furthermore, we were able to identify regions of the parameter space which produced noisy oscillatory dynamics for the negative feedback case (a known biological phenomenon – see Sturrock et al. [2013]). This was not too surprising, and we expected to find this. However, we also found a region of the parameter space for which oscillations could be found in the positive feedback case. This is not something that we have found reported previously in the literature, i.e., a single positive feedback loop GRN exhibiting oscillatory dynamics. We also compared the parameter spaces for the spatially homogeneous and spatially dependent cases. For both the positive and negative feedback cases, we found that as the diffusion coefficient was increased, the parameter landscapes for the spatially dependent model was in better agreement with the spatially homogeneous model. Hence, our results suggest that if the diffusion coefficient for mRNA and protein species

are known prior to constructing a model, then one can use the spatially homogeneous approximation with some confidence if the diffusion coefficient is greater than $7.29 \times 10^{-12} \mathrm{m}^2 \mathrm{min}^{-1}$. One advantage of this is that an ODE model may prove less computationally expensive to simulate than the corresponding PDE model.

An alternative approach to the use of data clustering is to calculate and use the steady-state solutions of Equations (1) - (3) to identify regions of the parameter space in which the promoter occupancy attains particular values. The advantage of this technique over k-means clustering is that it provides a global picture of how model behaviour varies in different regions of the parameter space; in contrast k-means clustering is reliant on sampling finite regions of the parameter space. However, in general closed form solutions to similar (perhaps nonlinear) problems will not be available but numerical techniques, such as k-means clustering, will still be an employable tool.

In future work we will consider investigating higher order moments in order to quantify the variance for each model species. It is often the case that transcription factors can not function in monomer form (as we have assumed in this paper), instead they must form dimers before they can function. Other mathematical models have accounted for this (see [Smolen et al., 2000, Sturrock et al., 2014]), and we expect that by accounting for this in our model, we may find a different form of equation (6) that is more consistent with previous phenomenological efforts. In order to keep our model as simple and general as possible, we neglected additional alleles and promoter binding sites. However, we may investigate in future the influence of additional spatially distinct alleles and promoter binding sites such as was considered for the Hes1 gene regulatory network in Zeiser et al. [2007]. In addition, there is experimental evidence that molecular movement within a cell can be subdiffusive [Weiss et al., 2004, Wachsmuth et al., 2000, Cabal et al., 2006], which is something we plan to investigate in future models. It may be that accounting for subdiffusion makes the spatially homogeneous approximation less accurate. It has been shown that cells containing negative feedback loops can synchronise their periods with neighbouring cells by using cell-cell communication [Masamizu et al., 2006, Shimojo et al., 2013]. We may also consider cell-cell communication in future work to see what influence this has on our model species. It may be that in a model with cell-cell communication, the mean field behaviour would produce oscillations as opposed to steady state behaviour. Finally, we plan to investigate the interplay between intrinsic stochasticity and extrinsic noise (both temporal and spatio-temporal) as is discussed in Shahrezaei et al. [2008], Caravagna et al. [2013], de Franciscis and d'Onofrio [2013].

Acknowledgements

Appendix: Steady-state solutions

In this appendix we calculate the steady-state solutions of Equations (1) - (3) in the main text and use them to identify regions in parameter space at which the promoter occupancy attains the value β (where $0 < \beta < 1$). We proceed by considering the steady-state solution of Equations (1) - (3) and partition $[0, L]$ into three subdomains as follows: Ω_1, $(0 < x < x_m)$, Ω_2, $(x_m < x < c)$, and Ω_3, $(c < x < L)$. On the boundary between Ω_2 and Ω_3 (i.e., $x = c$) we impose continuity and conservation of protein, such that

$$\frac{\partial \langle \mathbf{p} \rangle}{\partial x}\bigg|_{x=c^-} = \frac{\partial \langle \mathbf{p} \rangle}{\partial x}\bigg|_{x=c^+}, \quad \langle \mathbf{p} \rangle (x = c^+) = \langle \mathbf{p} \rangle (x = c^-). \tag{19}$$

At the boundary between Ω_1 and Ω_2 (i.e. $x = x_m$) we impose continuity of mRNA, such that

$$\langle \mathbf{m} \rangle (x = x_m^+) = \langle \mathbf{m} \rangle (x = x_m^-). \tag{20}$$

At steady-state, Equation (3) can be rearranged to yield the expression

$$\langle f \rangle = \frac{F_0}{\frac{\langle p \rangle}{p_0} + 1}, \tag{21}$$

where $p_0 = k_2/k_1$. Upon substitution for $\langle f \rangle$, the source term in Equation (2) takes the form

$$\alpha_m F_0 \frac{1 + \frac{\langle p \rangle}{\gamma p_0}}{1 + \frac{\langle p \rangle}{p_0}}. \tag{22}$$

Application of conservation of mRNA molecule number at $x = x_m$ yields the boundary condition

$$D \left(\frac{d \langle m \rangle}{dx}\bigg|_{x=x_m^+} - \frac{d \langle m \rangle}{dx}\bigg|_{x=x_m^-} \right) + h \mu_m \langle m \rangle (x_m) = h \left(\alpha_m F_0 \frac{1 + \frac{\langle p \rangle (x_m)}{\gamma p_0}}{1 + \frac{\langle p \rangle (x_m)}{p_0}} \right), \tag{23}$$

where h is defined to be the volume of the region in which the mRNA production reaction occurs.

In Ω_1, $\langle m \rangle$ satisfies (13) and has solution, assuming the zero-flux boundary condition at the origin, given by (14). In $\Omega_2 \bigcup \Omega_3$, $\langle m \rangle$ satisfies (13) and has solution, assuming the zero-flux boundary condition at $x = L$, given by

$$\langle m \rangle = C \cosh(k_m(x - L)), \tag{24}$$

where C is an integration constant. Similarly, in $\Omega_1 \bigcup \Omega_2$ $\langle p \rangle$ satisfies

$$D \nabla^2 \langle p \rangle - \mu_p \langle p \rangle = 0, \tag{25}$$

and has solution, assuming the zero-flux boundary condition at the origin, given by

$$\langle p \rangle = A \cosh(k_p x), \tag{26}$$

where $k_p = \sqrt{\mu_p/D}$ and A is an integration constant. In Ω_3, $\langle p \rangle$ satisfies

$$D \nabla^2 \langle p \rangle - \mu_p \langle p \rangle + \alpha_p \langle m \rangle = 0. \tag{27}$$

Substituting for $\langle m \rangle$ from Equation (24), we obtain

$$D \nabla^2 \langle p \rangle - \mu_p \langle p \rangle = -\alpha_p C(\cosh(k_m(x - L))). \tag{28}$$

This second order, non-homogeneous, ordinary differential equation can be solved using the method of variable coefficients and we find that in Ω_3

$$\langle p \rangle = E \cosh(k_p(x - L)) + rC \cosh(k_m(x - L)), \tag{29}$$

where $r = \alpha_p/(\mu_p - \mu_m)$, E is an integration constant and a zero-flux boundary condition has been imposed at $x = L$.

At $x = c$ we impose boundary condition (19) and obtain, after some manipulations, expressions for the integration constants A and E given by

$$A = rC\frac{\cosh(k_m(c-L))}{\cosh(k_pc)}\frac{\frac{k_m}{k_p}\tanh(k_m(c-L)) - \tanh(k_p(c-L))}{\tanh(k_pc) - \tanh(k_p(c-L))}, \tag{30}$$

$$E = rC\frac{\cosh(k_m(c-L))}{\cosh(k_p(c-L))}\frac{\frac{k_m}{k_p}\tanh(k_m(c-L)) - \tanh(k_pc)}{\tanh(k_pc) - \tanh(k_p(c-L))}. \tag{31}$$

Similarly, at $x = x_m$ we impose boundary condition (20) and obtain that

$$B = C\frac{\cosh(k_m(x_m - L))}{\cosh(k_m x_m)}. \tag{32}$$

Finally, in order to determine the remaining integration constant, C, we impose boundary condition (23). Summing terms on the left-hand side yields

$$Ck_mD\cosh(k_m(x_m - L))\left(\tanh(k_m(x_m - L)) - \tanh(k_m x_m) + hk_m\right) = Cg_1, \tag{33}$$

where, for convenience, we define g_1, a function of model constants, to be

$$g_1 = k_mD\cosh(k_m(x_m - L))\left(\tanh(k_m(x_m - L)) - \tanh(k_m x_m) + k_m h\right). \tag{34}$$

Using equations (26) and (31) to evaluate $\langle p \rangle$ at $x = x_m$

$$\langle p \rangle(x_m) = C\frac{r\cosh(k_m(c-L))\cosh(k_p x_m)}{\cosh k_p c}\frac{\frac{k_m}{k_p}\tanh(k_m(c-L)) - \tanh(k_p(c-L))}{\tanh(k_pc) - \tanh(k_p(c-L))} = Cg_2, \tag{35}$$

with g_2, a function of model parameters, defined for notational convenience to be

$$g_2 = \frac{r\cosh(k_m(c-L))\cosh(k_p x_m)}{\cosh k_p c}\frac{\frac{k_m}{k_p}\tanh(k_m(c-L)) - \tanh(k_p(c-L))}{\tanh(k_pc) - \tanh(k_p(c-L))}. \tag{36}$$

Note that to ensure that we obtain positive steady-state solutions, it is necessary that $g_2 > 0$.

Equating terms in the boundary condition (23) yields, after rearranging, a quadratic equation in the unknown C that is given by

$$C^2 g_2 g_1 + C\left(g_1 p_0 - \frac{hg_2\alpha_m F_0}{\gamma}\right) - h\alpha_m F_0 p_0 = 0. \tag{37}$$

Given the constraints that g_2 and g_1 are positive, equation (37) has a single positive solution given by

$$C = \frac{1}{2g_2g_1}\left(\frac{g_2\alpha_m hF_0}{\gamma} - g_1 p_0 + \sqrt{g_1^2 p_0^2 + 2\alpha_m hF_0 p_0 g_2 g_1\left(2 - \frac{1}{\gamma}\right) + \frac{h^2 g_2^2 \alpha_m^2 F_0^2}{\gamma^2}}\right). \tag{38}$$

Using the relationships derived above, expressions can be obtained for the other integration constants and the steady-state solution can be found for variables $\langle m \rangle$ and $\langle p \rangle$ in each of the subdomains.

In order to describe the parameter space results presented in the main text, we define the promoter occupancy at steady state, $\beta = \langle f \rangle/F_0$, e.g. with $\beta = 1/2$ the promoter is occupied on average half of the time. Upon rearranging Equation (21),

$$\langle p \rangle(x_m) = p_0\left(\frac{1}{\beta} - 1\right). \tag{39}$$

However, using the calculated solution we also find that

$$\langle p \rangle(x_m) = A\cosh(k_p x_m) = rC\frac{\cosh(k_m(c-L))}{\cosh(k_pc)}\frac{\frac{k_m}{k_p}\tanh(k_m(c-L)) - \tanh(k_p(c-L))}{\tanh(k_pc) - \tanh(k_p(c-L))}\cosh(k_p x_m) = Cg_2. \tag{40}$$
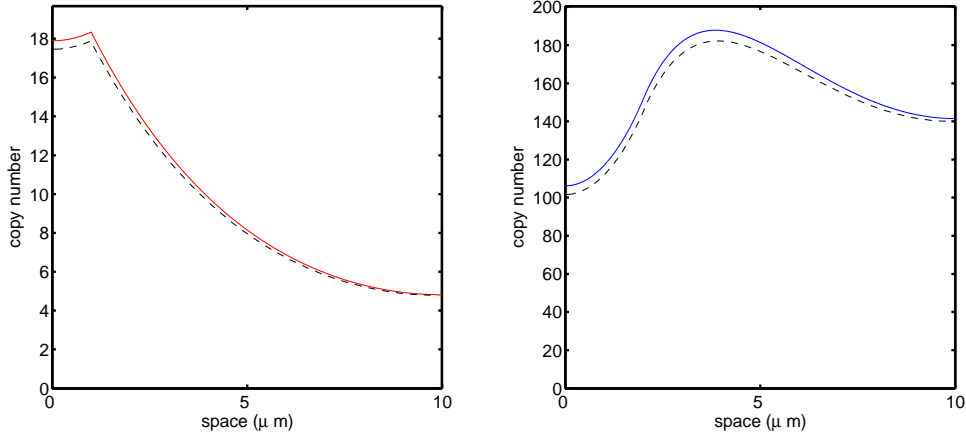
Fig. 14 The steady-state solutions defined by Equations (14), (24), (26) and (29) are compared with the numerical solutions of equations (1)–(3) (dashed lines) evaluated as $t \to \infty$. Left: $\langle p \rangle$ is plotted against x. Right: $\langle m \rangle$ is plotted against x.

Hence,

$$Cg_2 = p_0 \left(\frac{1}{\beta} - 1 \right). \tag{41}$$

Equations (37) and (41) represent a pair of simultaneous equations in the unknown constant C that can be solved to yield the expression

$$p_0 = \frac{g_2}{g_1} \frac{\beta}{1 - \beta} \frac{h\alpha_m F_0(\beta(\gamma - 1) + 1)}{\gamma}. \tag{42}$$

Substituting for p_0, g_1 and g_2 we obtain equation (17). Hence, regions of parameter space that yield a promoter occupancy β at steady-state can be identified. In order to derive a similar expression for the case of the spatially homogeneous model, given by equation (18), we consider the limit of $D \to \infty$ in equation (17). Given the inverse dependence of the parameters k_m and k_p on D, the trigonometric functions in equation (17) can be Taylor expanded about zero, and, after cancellation, leading order terms yield equation (18).

References

B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter. *Molecular Biology of the Cell*. Garland Science, Taylor and Francis Group Ltd, Oxford, fifth edition, 2008.

D. Barik, M. R. Paul, W. T. Baumann, Y. Cao, and J. J. Tyson. Stochastic simulation of enzyme-catalyzed reactions with disparate timescales. *Biophys. J.*, 95:3563–3574, 2008.

D. Barik, W. T. Baumann, M. R. Paul, B. Novak, and J. J. Tyson. A model of yeast cell-cycle regulation based on multisite phosphorylation. *Mol. Sys. Biol.*, 6: 405, 2010.

M. Barrio, K. Burrage, A. Leier, and T. Tian. Oscillatory regulation of Hes1: Discrete stochastic delay modelling and simulation. *PLoS ONE*, 2:e117, 2006.

L. Boyer, T. Lee, M. Cole, S. Johnstone, S. Levine, J. Zucker, M. Guenther, R. Kumar, H. Murray, R. Jenner, D. Gifford, D. Melton, R. Jaenisch, and R. Young. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell*, 122:947–956, 2005.

E. Buratti and F.E. Baralle. New aspects of autoregulation mechanisms in RNA binding proteins and their connection with human disease. *FEBS J.*, 278:3530–3538, 2011.

S. Busenberg and J. M. Mahaffy. Interaction of spatial diffusion and delays in models of genetic control by repression. *J. Math. Biol.*, 22:313–333, 1985.

G. G. Cabal, A. Genovesio, S. Rodriguez-Navarro, C. Zimmer, O. Gadal, A. Lesne, H. Buc, F. Feuerbach-Fournier, J. Olivo-Marin, E. C. Hurt, and U. Nehrbass. SAGA interacting factors confine sub-diffusion of transcribed genes to the nuclear envelope. *Nature*, 441:770–773, 2006.

G. Caravagna, G. Mauri, and A. d'Onofrio. The interplay of intrinsic and extrinsic bounded noises in bimolecular networks. *PLoS ONE*, 8:e51174, 2013.

A. Ciliberto, B. Novak, and J. J. Tyson. Steady states and oscillations in the p53/Mdm2 network. *Cell Cycle*, 4: 488–493, 2005.

S. de Franciscis and A. d'Onofrio. Cellular polarization: Interaction between extrinsic noises and the wave-pinning mechanism. *Phys. Rev. E*, 88:032709, 2013.

P. Eichenberger, M. Fujita, S. Jensen, E. Conlon, D. Rudner, S. Wang, C. Ferguson, K. Haga, T. Sato, J. Liu, and R. Losick. The program of gene transcription for a single differentiating cell type during sporulation in bacillus subtilis. *PLoS Biol.*, 2:e328, 2004.

R. Erban, J. Chapman, and P. Maini. A practical guide to stochastic simulations of reaction-diffusion processes. *arXiv preprint*, arXiv:0704.1908, 2007.

C. P. Fall, E. S. Marland, J. M. Wagner, and J. J. Tyson. *Computational Cell Biology*. Springer, New York, fifth edition, 2002.

D. Fusco, N. Accornero, B. Lavoie, S. M. Shenoy, J. Blanchard, R. H. Singer, and E. Bertrand. Single mRNA molecules demonstrate probabilistic movement in living mammalian cells. *Curr. Biol.*, 13:161–167, 2003.

A. Gamba, A. Candia, S. D. Talia, A. Coniglio, F. Bussolino, and G. Serini. Diffusion-limited phase separation in eukarytoic chemotaxis. *Proc. Natl. Acad. Sci. USA*, 47:16927–16932, 2005.

N. Geva-Zatorsky, E. Dekel, E. Batchelor, G. Lahav, and U. Alon. Fourier analysis and systems identification of the p53 feedback loop. *Proc. Natl. Acad. Sci. USA*, 107:13550–13555, 2010.

B. C. Goodwin. Oscillatory behavior in enzymatic control processes. *Adv. Enzyme Regul.*, 3:425–428, 1965.

J. S. Griffith. Mathematics of cellular control processes. I. negative feedback to one gene. *J. Theor. Biol.*, 20: 202–208, 1968a.

J. S. Griffith. Mathematics of cellular control processes. II. positive feedback to one gene. *J. Theor. Biol.*, 20: 209–216, 1968b.

S. Harris and A. Levine. The p53 pathway: positive and negative feedback loops. *Oncogene*, 24:2899–2908, 2005.

H. Hirata, S. Yoshiura, T. Ohtsuka, Y. Bessho, T. Harada, K.Yoshikawa, and R. Kageyama. Oscillatory expression of the bHLH factor Hes1 regulated by a negative feedback loop. *Science*, 298:840–843, 2002.

N.T. Ingolia and A.W. Murray. Positive-feedback loops as a flexible biological module. *Curr. Biol.*, 17:668–677, 2007.

M. H. Jensen, J. Sneppen, and G. Tiana. Sustained oscillations and time delays in gene expression of protein hes1. *FEBS Lett.*, 541:176–177, 2003.

R. Kagemyama, T. Ohtsuka, and T. Kobayashi. The hes1 gene family: repressors and oscillators that orchestrate embryogenesis. *Development*, 134:1243–1251, 2007.

S. Kar, W. T. Baumann, M. R. Paul, and J. J. Tyson. Exploring the roles of noise in the eukaryotic cell cycle. *Proc. Natl. Acad. Sci. USA*, 106:6471–6476, 2009.

A. D. Keller. Specifying epigenetic states with autoregulatory transcription factors. *J. Theor. Biol.*, 170:175–181, 1994.

A. D. Keller. Model genetic circuits encoding autoregulatory transcription factors. *J. Theor. Biol.*, 172:169–185, 1995.

T. B. Kepler and T. C. Elston. Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations. *Biophys. J.*, 81:3116–3136, 2001.

N. Klonis, M. Rug, I. Harper, M. Wickham, A. Cowman, and L. Tilley. Fluorescence photobleaching analysis for the study of cellular dynamics. *Eur. Biophys. J.*, 31:36–51, 2002.

T. Kobayashi and R. Kageyama. Hes1 regulates embryonic stem cell differentiation by suppressing notch signaling. *Genes to Cells*, 15:689–698, 2010. ISSN 1356-9597. doi: 10.1111/j.1365-2443.2010.01413.x.

T. Kobayashi and R. Kageyama. Hes1 oscillations contribute to heterogeneous differentiation responses in embryonic stem cells. *Genes*, 2:219–228, 2011. ISSN 2073-4425. doi: 10.3390/genes2010219.

T. Kobayashi, H. Mizuno, I. Imayoshi, C. Furusawa, K. Shirahige, and R. Kageyama. The cyclic gene hes1 contributes to diverse differentiation responses of embryonic stem cells. *Genes & Development*, 23:1870–1875, 2009.

Jacob Kogan. *Introduction to Clustering Large and High-Dimensional Data*. Cambridge University Press, New York, 2007.

G. Lahav, N. Rosenfeld, A. Sigal, N. Geva-Zatorsky, A. J. Levine, M. B. Elowitz, and U. Alon. Dynamics of the p53-Mdm2 feedback loop in individual cells. *Nature Genet.*, 36:147–150, 2004.

T. Lee, N. Rinaldi, F. Robert, D. Odom, Z. Bar-Joseph, G. Gerber, N. Hannett, C. Harbison, C. Thompson, I. Simon, J. Zeitlinger, E. Jennings, H. Murray, B. Gordon, B. Ren, J. Wyrick, J. Tagne, T. Volkert, E. Fraenkel, D. Gifford, and R. Young. Transcriptional regulatory networks in saccharomyces cerevisiae. *Science*, 298:799–804, 2002.

T. Warren Liao. Clustering of time series data – a survey. *Pattern Recogn.*, 38:1857–1874, 2005.

T. Lipniacki, P. Paszek, A. Marciniak-Czochra, A. R. Brasier, and M. Kimmel. Transcriptional stochasticity in gene expression. *J. Theor. Biol.*, 21:348–367, 2006.

Sijia Liu, Anastasios Matzavinos, and Sunder Sethuraman. Random walk distances in data clustering and applications. *Advances in Data Analysis and Classification*, 7:83–108, 2013.

J. M. Mahaffy. Genetic control models with diffusion and delays. *Math. Biosci.*, 90:519–533, 1988.

J. M. Mahaffy and C. V. Pao. Models of genetic control by repression with time delays and spatial effects. *J. Math. Biol.*, 20:39–57, 1984.

Y. Masamizu, T. Ohtsuka, Y. Takashima, H. Nagahara, Y. Takenaka, K. Yoshikawa, H. Okamura, and R. Kageyama. Real-time imaging of the somite segmentation clock: revelation of unstable oscillators in the individual presomitic mesoderm cells. *Proc. Natl. Acad. Sci. USA*, 103:1313–1318, 2006.

T. Matsuda, A. Miyawaki, and T. Nagai. Direct measurement of protein dynamics inside cells using a rationally designed photoconvertible protein. *Nature Meth.*, 5:339–345, 2008.

R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. Network motifs: simple building blocks of complex networks. *Science*, 298:824–827, 2002.

N. A. M. Monk. Oscillatory expression of Hes1, p53, and NF-κB driven by transcriptional time delays. *Curr. Biol.*, 13:1409–1413, 2003.

D. E. Nelson, A. E. C. Ihekwaba, M. Elliott, J. R. Johnson, C. A. Gibney, B. E. Foreman, G. Nelson, V. See, C. A. Horton, D. G. Spiller, S. W. Edwards, H. P. McDowell, J. F. Unitt, E. Sullivan, R. Grimley, N. Benson, D. Broomhead, D. B. Kell, and M. R. H. White. Oscillations in NF-κB signaling control the dynamics of gene expression. *Science*, 306:704–708, 2004.

L. K. Nguyen and D. Kulasiri. On the functional diversity of dynamical behaviour in genetic and metabolic feedback systems. *BMC Syst. Biol.*, 3:51, 2009.

H. Pirim, B. Ekşioğlu, A. Perkins, and Ç. Yüceer. Clustering of high throughput gene expression data. *Comput. Oper. Res.*, 39:3046–3061, 2012.

R.Tibshirani, G. Walther, and T. Hastie. Estimating the number of clusters in a data set via the gap statistic. *J. R. Stat. Soc.*, 63:411–423, 2001.

L. Saddic, B. Huvermann, S. Bezhani, Y. Su, C. Winter, C. Kwon, R. Collum, and D. Wagner. The LEAFY target LMI1 is a meristem identity regulator and acts together with LEAFY to regulate expression of CAULIFLOWER. *Development*, 133:1673–1682, 2006.

O. Seksek, J. Biwersi, and A. S. Verkman. Translational diffusion of macromolecule-sized solutes in cytoplasm and nucleus. *J. Cell. Biol.*, 138:131–142, 1997.

V. Shahrezaei and P. S. Swain. The stochastic nature of biochemical networks. *Curr. Opin. Biotechnol.*, 19: 369–374, 2008.

V. Shahrezaei, J. F. Ollivier, and P. S. Swain. Colored extrinsic fluctuations and stochastic gene expression. *Mol. Syst. Biol.*, 4:1–9, 2008.

H. Shimojo, Y. Maeda, T. Ohtsuka, and R. Kageyama. *Dynamic Notch Signaling in Neural Progenitor Cells*. Springer Japan, 2013.

P. Smolen, D. A. Baxter, and J. H. Byrne. Effects of macromolecular transport and stochastic fluctuations on the dynamics of genetic regulatory systems. *Am. J. Physiol.*, 277:C777–C790, 1999a.

P. Smolen, D. A. Baxter, and J. H. Byrne. Modeling clarifies the role of delays and feedback in circadian oscillators. *Soc. Neurosci. Abstr.*, 25:867, 1999b.

P. Smolen, D. A. Baxter, and J. H. Byrne. Modeling transcriptional control in gene networks – methods, recent results and future directions. *Bull. Math. Biol.*, 62:247–292, 2000.

E. H. Snoussi and R. Thomas. Logical identification of all steady states: the concept of feedback loop characteristic states. *Bull. Math. Biol.*, 55:973–991, 1993.

M. Sturrock, A. J. Terry, D. P. Xirodimas, A. M. Thompson, and M. A. J. Chaplain. Spatio-temporal modelling of the Hes1 and p53-Mdm2 intracellular signalling pathways. *J. Theor. Biol.*, 273:15–31, 2011.

M. Sturrock, A. J. Terry, D. P. Xirodimas, A. M. Thompson, and M. A. J. Chaplain. Influence of the nuclear membrane, active transport and cell shape on the Hes1 and p53–Mdm2 pathways: Insights from spatio-temporal modelling. *Bull. Math. Biol.*, 74:1531–1579, 2012.

M. Sturrock, A. Hellander, A. Matzavinos, and M. A. J. Chaplain. Spatial stochastic modelling of the hes1 gene regulatory network: intrinsic noise can explain heterogeneity in embryonic stem cell differentiation. *J. R. Soc. Interface*, 10:20120988, 2013.

M. Sturrock, A. Hellander, S. Aldakheel, L. Petzold, and M. A. J. Chaplain. The role of dimerisation and nuclear transport in the Hes1 gene regulatory network. *Bull. Math. Biol.*, 76:766–798, 2014.

A. Tafvizi, L. A. Mirny, and A. M. V. Oijen. Dancing on DNA: kinetic aspects of search processes on DNA. *Chem. Phys. Chem.*, 12:1481–1489, 2011.

A. J. Terry, M. Sturrock, J. K. Dale, M. Maroto, and M. A. J. Chaplain. A spatio-temporal model of Notch signalling in the zebrafish segmentation clock: Conditions for synchronised oscillatory dynamics. *PLoS ONE*, 6:e16980, 2011.

R. Thomas. The role of feedback circuits: positive feedback circuits are a necessary condition for positive real eigenvalues of the jacobian matrix. *Ber. Busenges. Phys. Chem.*, 98:1148–1151, 1994.

R. Thomas, D. Thieffry, and M. Kauffman. Dynamical behaviour of biological regulatory networks – I. biological role of feedback loops and practical use of the concept of the loop-characteristic state. *Bull. Math. Biol.*, 57: 247–276, 1995.

G. Tiana, M. H. Jensen, and K. Sneppen. Time delay as a key to apoptosis induction in the p53 network. *Eur. Phys. J. B*, 29:135–140, 2002.

J. J. Tyson and H. G. Othmer. The dynamics of feedback control circuits in biochemical pathways. *Prog. Theor. Biol.*, 5:1–62, 1978.

M. Wachsmuth, W. Waldeck, and J. Langowski. Anomalous diffusion of fluorescent probes inside living cell nuclei investigated by spatially-resolved fluorescence correlation spectroscopy. *J. Mol. Biol.*, 298:677–689, 2000.

G. R. Walther, A. F. M. Marée, L. Edelstein-Keshet, and V. A. Grieneisen. Deterministic versus stochastic cell polarisation through wave-pinning. *Bull. Math. Biol.*, 74:2570–2599, 2012.

J. Wang, M. Lefranc, and Q.Thommen. Stochastic oscillations induced by intrinsic fluctuations in a self-repressing gene: a deterministic approach. 04 2012.

O. D. Weiner, P. O. Neilsen, G. D. Prestwich, M. W. Kirschner, L. C. Cantley, and H. R. Bourne. A PtdinsP3- and Rho GTPase-mediated positive feedback loop regulates neutrophil polarity. *Nat. Cell Biol.*, 4:509–513, 2002.

M. Weiss, H. Hashimoto, and T. Nilsson. Anomalous protein diffusion is a measure for cytoplasmic crowding in living cells. *Biophys. J.*, 87:3518–3524, 2004.

D. Wilkinson. Stochastic modelling for quantitative description of heterogeneous biological systems. *Nat. Rev. Genet.*, 10:122–133, 2009.

W. Xiong and J. E. Ferrell. A positive-feedback-based bistable 'memory module' that governs a cell fate decision. *Nature*, 426:460–465, 2003.

S. Zeiser, J. Muller, and V. Liebscher. Modeling the Hes1 oscillator. *J. Comp. Biol.*, 14:984–1000, 2007.