BioMed Central

Research article

Bifurcation analysis informs Bayesian inference in the Hes1 feedback loop

Catherine F Higham

Address: Faculty of Biomedical and Life Sciences, University of Glasgow, Glasgow, Scotland, UK

Email: Catherine F Higham - c.higham.1@research.gla.ac.uk

Abstract

Background: Ordinary differential equations (ODEs) are an important tool for describing the dynamics of biological systems. However, for ODE models to be useful, their parameters must first be calibrated. Parameter estimation, that is, finding parameter values given experimental data, is an inference problem that can be treated systematically through a Bayesian framework.

A Markov chain Monte Carlo approach can then be used to sample from the appropriate posterior probability distributions, provided that suitable prior distributions can be found for the unknown parameter values. Choosing these priors is therefore a vital first step in the inference process. We study here a negative feedback loop in gene regulation where an ODE incorporating a time delay has been proposed as a realistic model and where experimental data is available. Our aim is to show that *a priori* mathematical analysis can be exploited in the choice of priors.

Results: By focussing on the onset of oscillatory behaviour through a Hopf Bifurcation, we derive a range of analytical expressions and constraints that link the model parameters to the observed dynamics of the system. Computational tests on both simulated and experimental data emphasise the usefulness of this analysis.

Conclusion: Mathematical analysis not only gives insights into the possible dynamical behaviour of gene expression models, but can also be used to inform the choice of priors when parameters are inferred from experimental data in a Bayesian setting.

Background

Aims

Mathematical models can help biologists understand the mechanisms and dynamics behind their experimental observations (Tomlin et al. [1]). The most widely used approach to modelling the dynamics of a genetic network is to employ systems of ordinary differential equations (ODEs) (Voit [2] and de Jong [3]). These models have biological parameters, some of which can be measured experimentally and some of which cannot. Parameter estimation, that is, recovering unknown parameters from experimental data, is an important step towards obtaining a good model that can not only explain observed results but can also be used for prediction and "what if" scenarios.

The inference of parameters, from real biological data, within a Bayesian framework is a relatively new, albeit currently very active area. Bayesian inference and Markov chain Monte Carlo (MCMC) methods have been recently advocated for the estimation of model parameters from biological systems described by ODEs (Rogers et al. [4]).

This paper aims to show that a general technique, using dynamical systems analysis to inform the choice of priors, can add value. The negative feedback loop studied here is a key feature of many complex biological networks and is referred to as a motif by Alon [5]. Hence our work will be of interest whenever larger networks are dealt with by a modular 'divide and conquer' approach. A Bayesian setting links the quantity that we are interested in, the probability that our parameters take certain values given the data, to two quantities that we can assign, the probability that we would have observed the measured data if the parameters took those values and our prior biological knowledge or ignorance about these parameters (Sivia [6]). Whereas traditional parameter estimation methods are deterministic and point valued (for example, COPASI [7]), these new methods use Bayes' Theorem to assign probabilities to parameter values and can handle noise inherently.

Using a Bayesian approach to parameter estimation for nonlinear dynamical systems throws up several challenges, and these typically increase when time delays are included. Key issues are

• dimensionality: models may involve several undetermined parameters,

• identification: different parameter combinations may produce similar dynamics, for example, increasing a production rate may be almost equivalent to decreasing a decay rate,

• local maxima: the likelihood function may have many locally optimal values.

MCMC methods [8] can go some way to addressing these difficult issues. In this work we show that *a priori* mathematical analysis can also be an effective tool. Our focus is on oscillatory time series, and we use the standard and widely applicable tools of Hopf bifurcation analysis and Lindstedt's method [9] in order to inform the choice of priors, thereby simplifying and focussing the parameter estimation process in order to find a particular parameter regime. More precisely, we follow the approach of Verdugo and Rand [10] in studying the case where oscillations arise via bifurcation through the delay parameter. Verdugo and Rand show that this leads to biologically realistic parameter values. We show here that inference with real data further supports this viewpoint.

The setting for our work is the biologically important instance where the expression of a gene is down regulated by its protein product. This arises, for example, with the p53 tumor suppressor protein whose intracellular activity is regulated through a feedback loop involving its transcriptional target [11]. We focus here on the case of the delayed Hes1 feedback loop featuring hes1 mRNA and Hes1 protein, where both mathematical models and quantitative experimental data are available in [12] and [13].

Biological Data
It has been observed that mRNAs for Notch signalling molecules such as the bHLH factor Hes1 oscillate with 2-hour cycles during somite segmentation, although the molecular mechanism of such oscillations remains to be fully determined. Hirata et al. [13] investigated the oscillations of mRNAs for Notch signalling molecules by examining the time course of hes1 mRNA and its Hes1 protein in detail. They observed that a single serum treatment induced a 2-hour cycle oscillation of hes1 mRNA in a variety of cultured cells and that Hes1 protein also oscillated in a 2-hour cycle after the single serum treatment with a delay of about 15 minutes relative to the hes1 mRNA oscillation. The half lives of hes1 mRNA and Hes1 protein were measured and the proteases for Hes1 protein degradation were identified.

Hirata et al. confirmed experimentally that degradation of Hes1 protein is required for hes1 mRNA increase and that *de novo* production of the protein is required for reduction of hes1 mRNA. These facts together support the theory that Hes1 is an essential component of a two hour cycle clock. They showed that the same mechanism applies to hes1 mRNA oscillation in the presomitic mesoderm.

The Hirata data comprises scaled hes1 mRNA expression levels every 30 minutes over a 12 hour period; see the discrete points in Figure 1.

A system featuring negative feedback and described by only two ODEs (one for each species, in this case hes1 mRNA and Hes1 protein), can be shown never to generate sustained oscillations [14]. Hirata et al. proposed that the observed oscillations could be modelled by introducing a Hes1 interacting factor as a third molecular species, but there is no experimental evidence for such an interacting factor. Monk [12] showed that observed oscillatory behaviour can be explained by incorporating a time delay.

Model for Hes1 Feedback Loop
Monk [12] suggests that the observed oscillations are due to the non-instantaneous nature of transcriptional and translational delays, and proposes a mathematical model where a delay is introduced to account for this feature. Monk's model was able to explain, via numerical simulations, the oscillation of hes1 mRNA and Hes1 protein in cultured cells observed by Hirata et al. Letting $m(t)$ and $p(t)$ denote the concentration of hes1 mRNA and Hes1 protein at time t, respectively, the model proposed by Monk for the Hes1 feedback loop takes the form
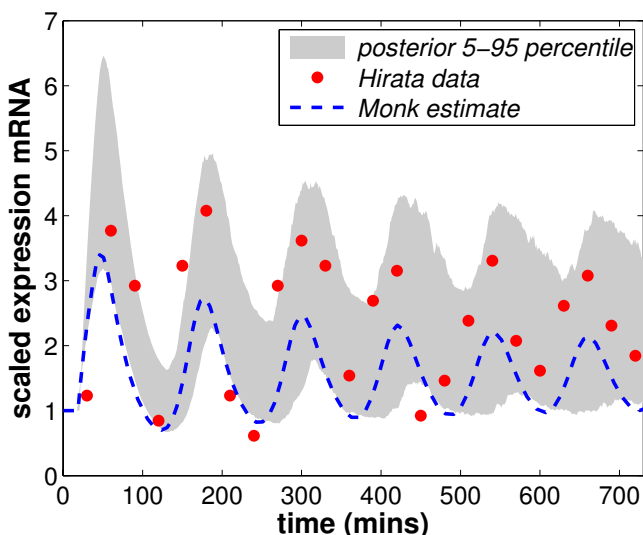
Figure 1
Experiment 2: Hirata data comprises 25 mRNA values taken every 30 minutes over a 12 hour period. Initial conditions are assumed to hold for $0 < t < \tau$.

$$\dot{m} = \frac{1}{1+(p(t-\tau)/p_0)^n} - \mu_m m, \qquad (1)$$

$$\dot{p} = m - \mu_p p, \qquad (2)$$

where μ_m and μ_p are the rates of degradation of mRNA and protein, respectively, p_0 is the normalised repression threshold, n is a hill coefficient and τ is a constant delay caused by transcription and translation. In this system, the Hes1 gene transcribes hes1 mRNA which passes from the nucleus to the cytoplasm. There it is translated into Hes1 protein. An unusual feature is that the Hes1 protein binds to the gene promoter and represses the transcription of hes1 mRNA. Following Monk, the equations have been rewritten so that the delay appears in the equation describing the regulation of transcription and not in the equation describing protein translation or synthesis. Throughout the rest of this paper, we will refer to τ as a transcriptional delay. The equations have also been rescaled so that the dynamics of the model are determined by five parameters. Using an ad hoc parameter fitting procedure, Monk showed that this model reproduced the broad features of the oscillation of hes1 mRNA and Hes1 protein in cultured cells observed by Hirata et al. [13].

Thus Monk [12] argues that the observed oscillatory behaviour is best accounted for by the introduction of a delay parameter that acts as a proxy for the many subprocesses that make up transcription and translation, rather than by introducing an unknown third agent [13].

By performing mathematical analysis of the model (1)–(2), we will show that the convincing but ad hoc parameter fitting exercise in [12] can be extended to a fully Bayesian setting.

Mathematical Analysis
Verdugo and Rand [10] gave some mathematical analysis of Monk's model [12] for the Hes1 feedback loop. In particular, they derived closed form approximations for the amplitude and frequency of oscillation, where oscillatory behaviour is assumed to arise through Hopf bifurcation in the delay parameter. The analysis in [10] applies to the case where the decay rates of hes1 mRNA and Hes1 protein, key components of the feedback, are equal, that is, $\mu_m = \mu_p$ in (1)–(2). In this work, we study the more realistic case where the decay rates are allowed to be different, also focussing on oscillatory behaviour. By allowing these degradation rates to differ, we expose some interesting features of the nonlinear system under consideration; namely that oscillations of this type only occur when the difference between degradation rates is small. The novelty of this work lies in using bifurcation tools to aid what turns out to be a very difficult inference problem. Details of the mathematical analysis can be found in the methods section. Our results are of interest in their own right as a means to understand how the system dynamics are driven by the model parameters, but our main aim here is to inform the corresponding Bayesian inference problem.

Experiments
In order to illustrate the use of mathematically informed priors when inferring model parameters we conducted two experiments. Experiment 1 uses synthetic data, computer generated from the mathematical equations describing the Monk model with known parameters. Experiment 2 looks at the published data of Hirata et al. Data is noisy and we have noted the error bars in the Hirata data when making assumptions about this type of extrinsic noise. Bayesian methods are very useful when dealing with this type of experimental noise because we can include noise as a hyper parameter, treat it as a nuisance parameter and integrate it out. Intrinsic noise will also come into play at the molecular level [15] and there is recent work on the subject of inferring parameters in stochastic models [16]. However, we do not believe that there is enough data to make it feasible to fit a stochastic model that would take into account intrinsic noise. We therefore use the ODE-based model in line with other authors [12] and [15]. Consequently, we are dealing with extrinsic and experimental noise. To make the synthetic data in Experiment 1 more realistic, Gaussian noise is added, based on the absolute size of the error bars of Hirata with a mean of zero and a standard deviation of 0.2. In our example, we have used larger initial conditions so the relative noise is lower. However, investigations with a higher level of noise

to match the relative values from Hirata give broadly similar conclusions. The justification for adding Gaussian noise to make the data more realistic lies with the *Central Limit Theorem*, which shows that under appropriate conditions the overall effect of extrinsic noise sources will be normal [4]. An advantage of using synthetic data is that the parameters to be inferred are known and so the accuracy of the method to infer them can be evaluated. For Experiment 2, two of the parameters values, μ_m and μ_p, have been measured independently. So mathematically informed priors are complemented with biological priors where available.

In Experiment 2 (Hirata Data) there is data for both species m and p and so in Experiment 1 (Synthetic Data) we generate data for both species m and p as well. We will show that some aspects of the data for p inform the mathematical analysis, however the actual inference of parameters for both m and p is based on the data for m only. This is because protein expression levels are harder to measure accurately in practice so it is useful to establish what can be inferred from the more accurately measured species alone. With the Hirata data, there is also the added complication that the protein levels measured have been scaled in an undisclosed manner [15].

The location in time of the sampling points is clearly an important issue. For species that oscillate, it may be difficult to know in advance which time points will provide most information. In this work the real data is already available and hence our initial test was chosen to match those details. Generally, for oscillatory behaviour, we would expect that the key constraint on the sample times is that the sampling frequency should be sufficiently large.

Methodology
We have some data and a proposed model. We use a Bayesian framework to infer the parameters underlying the model. In a Bayesian framework, the posterior pdfs are given by the prior pdfs modified by the likelihood of the data given those parameters. More precisely the posterior pdfs are proportional to the product of the likelihood and the priors. The posterior pdf for a particular parameter depends on the values of all the other parameters and are shown for individual parameters by integrating over the other parameter values. Hence the posterior pdfs that we show are marginal distributions for each parameter and not distributions for fixed values of the remaining parameters. The resulting equations for the posterior pdfs are intractable so we use MCMC methods to sample from the posterior distributions. This section spells out further details of this methodology which differs slightly for Experiment 1 and Experiment 2.

Bayesian Inference
Bayesian methods have several advantages over other approaches to parameter estimation [17]: use of back-

ground information, the ability to include uncertainty in all parameter values and the ease of making inferences about some parameters irrespective of the values of others. The development of high-speed computing means that the potential of Bayesian methods is being realised, but there are still many computational difficulties. In this section, we summarise the key features of Bayesian inference and highlight the technical decisions which have to be made. For more information about Bayesian analysis we refer readers to [6] and reviews such as [17].

Our initial state of knowledge (or ignorance) about the model parameters is encapsulated in the *prior probability*. The priors are then modified by the experimental measurements through the *likelihood function* to give us the *posterior probability*, which represents our new state of knowledge about the value of the parameters. A likelihood function needs to be chosen and after using a normal distribution for Experiment 1, we chose to base the likelihood function for Experiment 2 on the chi-squared distribution. The chi-squared distribution, χ^2, has longer tails than the normal distribution and so encourages the inclusion of outliers in the data, which is important for the estimation of model parameters which influence in particular the amplitude of the oscillations. This proved to be an important consideration in Experiment 2, possibly because of experimental noise levels.

For Experiment 1, we therefore suppose that the probability of the kth datum having a value x_k is given by $N(f_k, \sigma)$ where $f_k = f(\theta, t_k)$ is the true value of the function of parameters $\theta = \{p_0, n, \mu_m, \mu_p, \tau\}$ of interest, and the variance, σ, accounts for the error in its measurement. In our case, f_k is the mRNA level $m(p_0, n, \mu_m, \mu_p, \tau)$ at time t_k from the delay differential equation model (1)–(2).

For Experiment 2, we suppose that the probability of the kth datum having a value x_k is given by $\chi^2(f_k, \nu)$ where the degree of freedom ν, accounts for the error in its measurement. Given that the Hirata data has been scaled, it will also be necessary to infer an additional scaling parameter, which we refer to as k_s, so $\theta = \{p_0, n, \mu_m, \mu_p, \tau, k_s\}$.

For both experiments, we assume that the value of σ and ν is known. Our inference about the value of θ is expressed, therefore, by the posterior pdf $p(\theta \mid \{x_k\}, I)$ where I denotes all other background information; to help us calculate it, we use Bayes' Theorem [6]

$$p(\theta \mid \{xk\}, \nu, I) \propto p(\{x_k\} \mid \theta, \nu, I) \times p(\theta, \nu, I). \quad (3)$$

The relation in (3) is expressed using proportionality, because the term $p(data|I)$ has been omitted from the denominator in the right hand side. This is fine for data analysis problems, such as this one, involving *parameter estimation*, since the missing denominator is simply a nor-

malisation constant not depending explicitly on the parameters. However if we were to consider *model selection*, this term would play a crucial role, and is thus given the special name of *evidence*.

As in [4], we assume that the data are independent, so that the measurement of one datum does not influence what we can infer about the outcome of another (when given the values of θ). The *likelihood function* is then given by the product of the probabilities of obtaining the *N* individual data:

$$p(\{x_k\} \mid \theta, I) = \prod_{k=1}^{N} p(x_k \mid \theta, I). \qquad (4)$$

We assign a Gaussian pdf as prior for each parameter θ_j, with mean μ_j and variance σ_j, where $j = 1{:}5$ in Experiment 1 and $j = 1{:}6$ in Experiment 2:

$$p(\theta_j \mid \sigma, I) = p(\theta_j \mid I) = \frac{1}{\sigma_j \sqrt{2\pi}} \exp\left(-\frac{(\theta_j - \mu_j)^2}{2\sigma_j^2}\right). \qquad (5)$$

According to (3), if we multiply the prior stated in (5) by the likelihood function (4), we obtain the posterior pdf. Our approach here will be to deal with the posterior pdf numerically, using MCMC.

MCMC Method

MCMC techniques are used to sample from distributions and more generally to solve optimisation and integration problems in large dimensional spaces (Andrieu et al. [8]). Here, we use MCMC to sample from the posterior pdf defined in (3). The approach is, essentially, to compute candidate values for our unknown parameters by constructing a Markov chain with an appropriate invariant measure. The candidate values are chosen randomly using a *proposal function* and are either accepted or rejected

depending on an *acceptance probability*. The frequency at which a particular parameter is successfully sampled is indicative of its *posterior probability*. Once we have enough samples from the solution space, they can be binned, smoothed and normalised to indicate the relative posterior likelihood of a given parameter. Sampling can start from any particular parameter value. The *proposal function* is chosen so that samples are drawn widely and performance can be monitored by the *acceptance rate*. With Markov chains, after a *burn in period*, the memory of the initial starting point is lost. At this point sampling will be exclusively from the required distribution. We need to establish at which point this has occurred. Gelman et al. [18] propose setting off parallel sampling chains and monitoring their convergence with a statistic $\hat{R}$. We have followed this convention. More details can be found in the Methods section and Tables 1 (Experiment 1) and 2 (Experiment 2).

Results and discussion
Mathematical Analysis

We summarise the main analytical results here. Further details about the derivations of the formulae can be found in the Methods section.

Following [10] we can derive an expression for the equilibrium state of *m* and *p* (*m** and *p**, respectively, see equations (21)–(20)) of model (1)–(2) and show that by increasing the delay the system moves from a stable state to an unstable state. At a critical value τ_{cr} of the time delay a Hopf Bifurcation occurs. For values of delay τ close to τ_{cr}, the nonlinear system is expected to exhibit a periodic solution which can be expressed for both the mRNA and protein in terms of their amplitudes, *A* and *B*, respectively, their oscillatory frequency ω (at τ_{cr}) and their phase difference φ. Algebraic manipulation of these expressions results in explicit formulae linking the model's biological parameters which generalise those in [10]. Specifically, at the Hopf bifurcation we have

$$\tan \phi = \frac{\omega}{\mu_p}, \qquad (6)$$

$$\frac{B}{A} = \sqrt{\omega^2 + \mu_p^2}, \qquad (7)$$

$$w^2 = -\frac{(\mu_m^2 + \mu_p^2) \pm \sqrt{(\mu_m^2 + \mu_p^2)^2 - 4(\mu_m^2 \mu_p^2 - K^2)}}{2} \qquad (8)$$

Table 1: Experiment 1 (Synthetic Data).

θ	sp1	sp2	sp3	σ	acc rate	$\hat{R}$
P_0	85	95	91	0.5	26	1.0009
n	4	5	5.3	0.7	34	1.0001
μ_m	0.024	0.028	0.022	0.1	43	1.0004
μ_p	0.036	0.034	0.033	0.1	33	1.0000
τ	19	19.2	18.7	0.01	20	1.0000

This table shows inputs and convergence information for 3 illustrative MCMC chains. For each parameter, θ we show the starting point (sp) of the Markov chains (columns 2–4), variance σ of the Gaussian proposal function (column 5), the acceptance rate (column 6) and statistic $\hat{R}$ (column 7).

Table 2: Experiment 2 (Hirata Data).

θ	sp1	sp2	sp3	σ	acc rate	$\hat{R}$
P_0	99	103	103	0.55	29	1.0000
n	5	4.8	5.3	0.7	34	1.0003
μ_m	0.028	0.03	0.03	0.3	28	1.0001
μ_p	0.028	0.03	0.031	0.5	31	1.0002
τ	18	19	18.5	0.5	27	1.0001
k_s	2.5	2.2	2	0.2	27	1.0000

This table shows inputs and convergence information for 3 illustrative MCMC chains. For each parameter, θ we show the starting point (sp) of the Markov chains (columns 2–4), variance σ of the Gaussian proposal function (column 5), the acceptance rate (column 6) and statistic $\hat{R}$ (column 7).

$$\tau_{cr} = \frac{1}{\omega}\arctan\left(\frac{\omega\mu_p + \omega\mu_m}{\omega^2 - \mu_m\mu_p}\right), \qquad (9)$$

$$\sin\omega\tau_{cr} = \frac{\omega\mu_p + \omega\mu_m}{K}, \qquad (10)$$

$$\cos\omega\tau_{cr} = \frac{\omega^2 - \mu_m\mu_p}{K}, \qquad (11)$$

where

$$K = \frac{n\beta}{p^*(1+\beta)^2}, \qquad (12)$$

and

$$\beta = \left(\frac{p^*}{p_0}\right)^n. \qquad (13)$$

The result (7) characterises the ratio between mRNA and protein amplitudes, but not their absolute values. To proceed we use Lindstedt's method which is a technique for uniformly approximating periodic solutions to ODEs when regular perturbation approaches fail. We regard the frequency as unknown in advance and solve by demanding that an appropriate series expansion contains no secular terms. The result is closed form approximate expressions for the amplitude and frequency of oscillation expressed in terms of the model parameters. For the derivation of these formulae and more details, see the Methods section. In summary, our approximation to A^2 and the observed frequency – may be written in the form

$$A^2 = f_A(n, p_0, \mu_m, \mu_p)\Delta, \qquad (14)$$

$$\Omega = \omega - f_\Omega(n, p_0, \mu_m, \mu_p)\Delta, \qquad (15)$$

and substituting (14) into (7) gives an approximation for B^2

$$B^2 = (\omega^2 + \mu_p^2)f_A(n, p_0, \mu_m, \mu_p)\Delta, \qquad (16)$$

where f_A and f_Ω are positive functions defined in terms of the model parameters and $\Delta = \tau - \tau_{cr}$. As Δ increases we would expect A and B to increase proportionally to $\sqrt{\Delta}$ and Ω to decrease.

The frequency of oscillation, ω, takes a maximum that can be explicitly defined in terms of the model parameters. We may show that for any $\mu_m\mu_p = c$, a unique maximum value of ω occurs when $\mu_m = \mu_p = \sqrt{c}$. We may then find the value of c and hence μ_m and μ_p for which the maximum of ω occurs. We can use this information combined with equation (15) to deduce that $\omega > \Omega$. This result allows us to make an informed judgement about ω given Ω, which can be used in the setting of priors.

Biology meaningful solutions must also have $\omega > 0$ and $\tau_{cr} > 0$. Imposing these conditions in equation (9) gives

$$\omega^2 > \mu_m\mu_p, \qquad (17)$$

$$\omega\mu_m + \omega\mu_p < K, \qquad (18)$$

from which we deduce that

$$\mu_m\mu_p < \frac{1}{p_0\left(\frac{2}{n-2}\right)^{1/n}\left(\frac{n}{n-2}\right)}. \qquad (19)$$

This is a key result because we have defined a computable, finite region of the μ_m - μ_p plane in which oscillations can occur and in which we can search for oscillatory solutions; see Figure 2.

The data can tell us something about the parameters directly. Given sufficient data points from complete periodic cycles, the model parameter μ_p can be approximated by the ratio of the average of m to the average of p. Further details can be found in the Methods section; see equation (70).

We now show how our results can be used to inform the choice of priors in our two inference experiments.
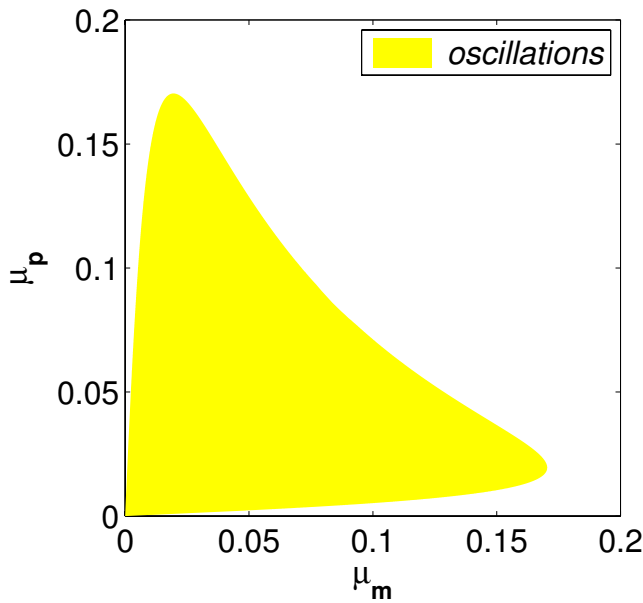
Figure 2
Constrained region of the μ_m-μ_p plane where the inequality (19) predicts that oscillations may occur for arbitary τ. (In this picture $n = 5$ and $p_0 = 90$).
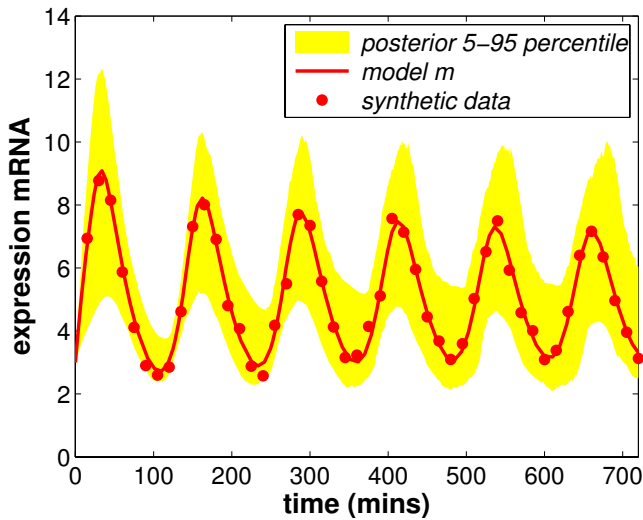

Figure 3
Experiment 1: Synthetic data for the inference experiment comprising 49 mRNA values and 49 protein values (not shown here as they are not used directly in the inference problem). Data comes from the model with the following parameter values: $p_0 = 90$, $n = 5$, μ_m = 0.025, μ_p = 0.035 and τ = 19.5. Initial conditions are 3 for the mRNA and 100 for the protein. Independent Gaussian noise (mean = 0 and standard deviation = 0.2) is then added to make the data more realistic.

Experiment 1 (synthetic data)
The synthetic data in Figure 3 comprises 49 mRNA values. Details of the parameter values used with the Monk equations and the added noise can be found in the caption of Figure 3. The corresponding 49 protein values were also generated and feature in Figure 4 where they are considered for the μ_p prior. The protein values, as for Experiment 2, are not used in the inference problem. The following section describes how the mathematical analysis is used to find priors for the model parameters.

We can estimate μ_p as $\frac{m_{ave}}{p_{ave}}$ from equation (70). In computing m_{ave} and p_{ave}, care needs to be taken to discard the transient data and to select data points from a complete number of cycles. In our example, as the oscillatory period appears to be about 2 hours, we take the last 20 data points which form approximately 5 cycles, see Figure 4. This gives an estimate for μ_p of 0.0344. We estimate the ratio of the amplitudes $\frac{B}{A}$ using $B \approx m_{max} - m_{min}$ and $A \approx p_{max} - p_{min}$, see Figure 4. Substituting $\mu_p \approx 0.0344$ and $\frac{B}{A} \approx 0.0656$ into (7) gives an estimate for ω of 0.0559. We now use the fact that A, B and ω are functions of the unknowns p_0, n and μ_m in order to solve numerically (7), (8), (14)
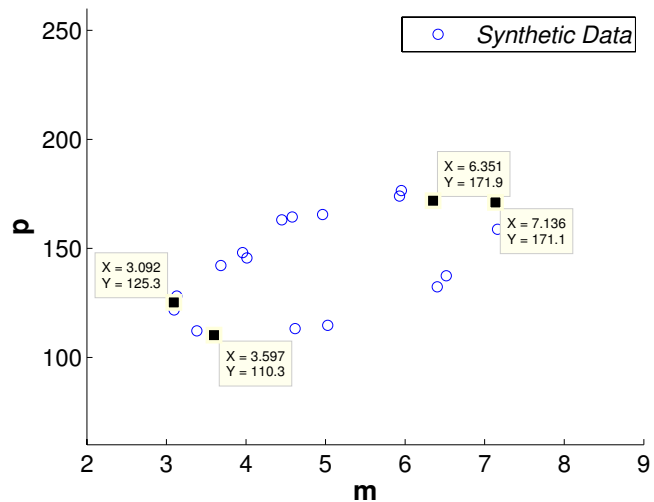
$$\mu_p = \frac{m_{ave}}{p_{ave}} = \frac{4.9783}{144.8241} = 0.0344$$

Experiment 1: We use the observed data to estimate
$\mu_p = \frac{m_{ave}}{p_{ave}} = \frac{4.9783}{144.8241} = 0.0344$. We use the data points representing m_{max} = 7.136, m_{min} = 3.092, p_{max} = 171.9 and p_{min} = 110.3 to calculate $\frac{B}{A} = \frac{7.136-3.092}{171.9-110.3} = 0.0656$.

and (15). This gives $p_0 \approx 90$, $n \approx 5$ and $\mu_p \approx 0.026$. We also require a prior for τ. Lindstedt's method tells us that the closeness in value of τ and τ_{cr} depends on the difference between ω and the observed frequency Ω. Our observed period is about 2 hours giving an observed frequency of 0.052. Substituting our estimates into (9) and (15) we obtain $\tau_{cr} \approx 17.6567$ and $\tau \approx 19.9508$. See Table 3 for a summary of our priors for the synthetic experiment. The variances for the priors are based, in the case of μ_m and μ_p, on the decay rate measurements taken by Hirata et al., and in the case of the other variables on biological assumptions combined with the mathematical analysis.

Results

Figure 5 shows the prior pdfs and the corresponding posterior pdfs for each of the parameters inferred in Experiment 1. The parameter values used in the model to simulate the data are shown on each x-axis with a dot. The peak of each curve indicates which parameter values are most probable, and a sharper peak indicates more certainty. For all parameters in Experiment 1 the peak lies over the input value to within visual accuracy. We are also interested in the extent to which the posterior pdfs, which express the probability of each parameter given the data, modify our prior beliefs. Here we used Gaussian priors with mean based on the point estimates outlined above and small variance. The posterior pdfs for p_0, n, μ_m and μ_p provide only slight modifications of the prior pdfs suggesting that the *a priori* mathematical analysis has allowed us to select these priors very effectively. The fact that these mathematically informed priors, which are based on broad features of the data, are reasonably close to the posterior pdfs suggests that the bifurcation analysis is relevant and consistent with the data. The parameter τ shows the most change from prior to posterior. We can show from (16) that the amplitude of mRNA is sensitive to τ, and this experiment (and Experiment 2 below) suggests that the likelihood function is also strongly affected by τ. In Figure 6 we rerun the same experiment using uniform priors over the point estimates. We see more modification of the posterior pdfs, relative to the prior pdfs but, as shown by the

Table 3: Experiment 1 (Synthetic Data).

| θ	mean	std dev
P_0	90	10
n	5	1
μ_m	0.025	0.001
μ_p	0.035	0.001
τ	18.5	1

This table details the Gaussian priors used for inferring from the synthetic data.
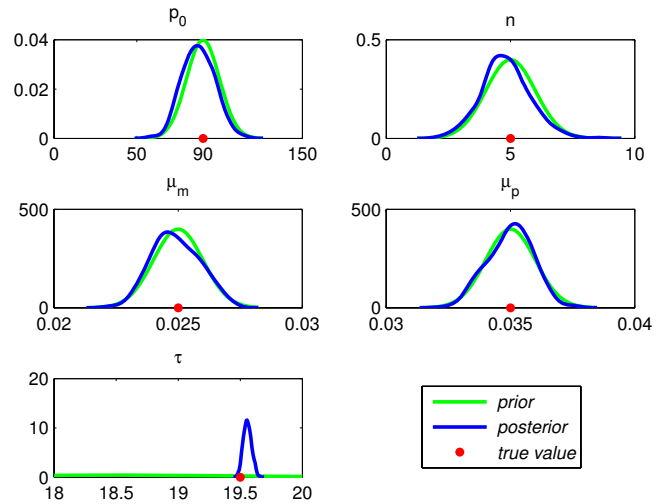
Figure 5
Experiment 1: Prior pdfs, posterior pdfs and parameter value used in model.

flatness of the pdf curves, there is increased uncertainty in the prediction of parameter values. We ran several more experiments with different priors (not shown here) and conclude that when we apply more general priors we run into trouble with this hard inference problem and get stuck in suboptimal posteriors. This highlights the importance of exploiting whatever information can be gleamed from the structure of the model.

MCMC provides samples and in order to recover the posterior pdfs these samples are binned, smoothed and nor-
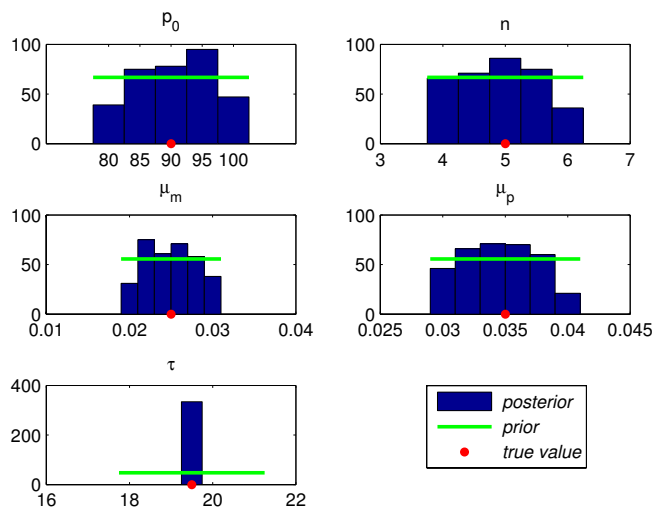
Figure 6
Experiment 1: Binned posterior samples compared to mean ie uniform prior.

malised. We can also simulate the dynamic activity associated with each of these samples using the associated parameter values. The variance of this dynamic activity at each time point indicates the likely range of activity that could have produced the data. The shaded area in Figure 3 shows this range (5–95 percentile) for Experiment 1. The fit of this area in terms of position and width shows how well we have recovered the model parameters from the data.

Experiment 2 (biological data)

Hirata et al. measured the half lives of hes1 mRNA and Hes1 protein and estimated that hes1 mRNA is degraded with a half life of 24.1 ± 1.7 minutes and that Hes1 protein is degraded with a half life of 22.3 ± 3.1 minutes. These half lives are close but not identical. Expressed as decay rates, the rate for hes1 mRNA is 0.0288 ± 0.002 and the rate for Hes1 protein is 0.0311 ± 0.004. We used these rates to give biologically informed priors for the parameters μ_m and μ_p. Using Lindstedt's method we obtained an expression for the observed oscillatory frequency Ω from equation (15). As $k_2\Delta < 0$ we have $\omega > \Omega$. In other words, we have a lower bound for ω. Hirata et al. observed an oscillatory period of around 2 hours which corresponds to an oscillatory frequency (in mins) of $\frac{2\pi}{120}$. Using this observation, we can argue that $\omega > 0.0524$. Substituting $\omega = 0.0524$, $\mu_m = 0.0288$ and $\mu_p = 0.0311$ into equation (8) gives $K > 0.036$. The value of K in equation (12) can be shown, via numerical computation, to be sensitive to the value of n, whereas the values of m^* and p^* are more sensitive to p_0 than n. Given m^* (or p^* as $m^* = \mu_p p^*$ from equation (21)) and the estimate for K, we can obtain point estimates from equations (21) and (12) of $n \approx 5$ and $p_0 \approx 100$. We also require a prior for τ. Substituting our estimates for ω, μ_m and μ_p into equation (9) gives $\tau_{cr} < 19.8$ and we use this estimate of τ_{cr}, which is close to τ, for choosing our prior for τ. We also need a scaling factor k_s

Table 4: Experiment 2 (Hirata Data).

θ	mean	std dev
P_0	100	10
n	5	1
μ_m	0.0288	0.002
μ_p	0.0311	0.004
τ	19	1
k_s	2.2	0.1

This table details the Gaussian priors used for inferring the Hirata data. The priors for μ_m and μ_p are based on biological measurements.

because the data has been scaled by a undisclosed constant, and this will also affect the initial condition. We set k_s between 2 and 3, the initial condition of p to be 100 and m to be k_s. See Table 4 for a summary of the priors.

Results

Figure 7 shows the prior pdfs and the corresponding posterior pdfs for each of the parameters inferred in Experiment 2. Once again, the customised priors are very close to the final pdfs. In this experiment, we chose a χ^2 distribution rather than a normal distribution for the likelihood, in order to encourage inclusion of any outlying data points.

The shaded area in Figure 1 shows the fit of the parameter estimates within the 5–95 percentile for Experiment 2. The early data points are covered by the prediction. Compared to Monk's results, there is broad agreement, however we predict a slightly longer delay (around 18.9 compared to Monk's 18.5) and slightly different values for μ_m and μ_p (around 0.0285 and 0.0305, respectively compared to Monk's 0.03 for both). We also inferred a different scaling factor which contributed significantly to the improved match to the Hirata et al. data that can be seen in Figure 1.

Conclusion

We have shown that mathematical analysis can give insights into the possible dynamical behaviour of gene expression models. We derived a range of analytical expressions and constraints that link the model parameters to the observed dynamics of the system. The basis of the analysis is the assumption that oscillatory behaviour arises through Hopf bifurcation in the delay parameter. It
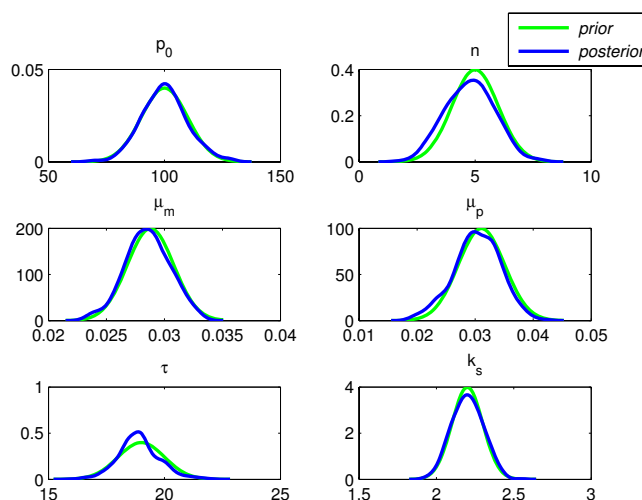
Figure 7
Experiment 2: Prior and posterior pdfs.

is then possible to say for which range of parameters and to what extent this behaviour will occur. For the problem of inferring parameter values from data, we can use this analysis to inform the choice of priors. For these types of problems, where different parameter values can result in similar activity, identification of the parameters through the data can be ambiguous. MCMC methods can go some way towards resolving these difficulties but we showed here that mathematically informed priors can add value too. On the experimental data from Hirata et al., a full Bayesian MCMC computation improved on the parameter fit published by Monk, predicting non-equal values for mRNA and protein decay rates and a longer time delay for transcription.

Methods

We describe here in more detail the mathematical and computational aspects of this work. We generalise the approach of Verdugo and Rand to the more biologically realistic case where the degradation rates of mRNA and protein are not assumed to be equal. The analysis reveals that we would expect to see oscillations when the difference between μ_m and μ_p is small which justifies, generally, the assumption that the degradation rates are almost equal.

Stability of Equilibrium

Following the approach of Verdugo and Rand [10], equilibrium points for the system (1) and (2) are found by setting $\dot{m} = 0$ and $\dot{p} = 0$. After elimination and substitution, we obtain two expressions for p^* and m^*:

$$(p^*)^{n+1} + p_0^n p^* - \frac{p_0^n}{\mu_m \mu_p} = 0, \qquad (20)$$

$$m^* = \mu_p p^*. \qquad (21)$$

To find out whether m^* and p^* are stable, we linearize about these points and define ζ and η to be deviations from the equilibrium: $\zeta(t) = m(t) - m^*$, $\eta(t) = p(t) - p^*$, and $\eta_d = \eta(t - \tau)$. This results in the linear system:

$$\dot{\zeta} = -\mu_m \zeta - K\eta_d, \qquad (22)$$

$$\dot{\eta} = \zeta - \mu_p \eta, \qquad (23)$$

where K and β are given in (12) and (13). Equations (22) and (23) can be written as the second order DDE:

$$\ddot{\eta} + (\mu_m + \mu_p)\dot{\eta} + \mu_m\mu_p\eta + K\eta_d = 0. \qquad (24)$$

We now look for a solution of the form $\eta = e^{\lambda t}$, which leads to

$$\lambda^2 + (\mu_m + \mu_p)\lambda + \mu_m\mu_p + Ke^{-\lambda\tau} = 0. \qquad (25)$$

First, we focus on the non-delay case, $\tau = 0$. Here we have

$$\lambda = \frac{-(\mu_m+\mu_p) \pm \sqrt{(\mu_m+\mu_p)^2 - 4(\mu_m\mu_p+K)}}{2}. \qquad (26)$$

Examination of equation (26) reveals that since $\mu_m + \mu_p > 0$ and $\sqrt{(\mu_m+\mu_p)^2 - 4(\mu_m\mu_p+K)} < \mu_m + \mu_p$, the real part of λ will always be negative, implying stability. For

$$4(\mu_m\mu_p + K) > (\mu_m + \mu_p)^2 \qquad (27)$$

λ has a non-zero imaginary part as well as a negative real part and the equilibrium point (m^*, p^*) will be a stable spiral. When $\mu_m = \mu_p$, the inequality (27) becomes $K > 0$ which always holds. However when $\mu_m \neq \mu_p$ this inequality becomes $4K > (\mu_m - \mu_p)^2$. In other words, we only have a stable spiral when $|\mu_m - \mu_p|$ is sufficiently small. For larger values of $|\mu_m - \mu_p|$ we will have two negative real roots for λ, resulting in a fixed stable point with no oscillations. We conclude that, overall, the non-delay model always has a linearly stable fixed point.

Following the approach in [10], we now assume, by continuity, that as τ increases the roots λ will cross the imaginary axis (this will not be at the origin because $\mu_m > 0$ and $\mu_p > 0$ and therefore the real and imaginary parts of λ cannot both be zero) at some critical value $\tau = \tau_{cr}$, and for $\tau > \tau_{cr}$ the steady state (m^*, p^*) will lose stability giving rise to a Hopf bifurcation. We thus assume that for $\tau = \tau_{cr}$ the system (22)–(23) will exhibit a pair of pure imaginary eigenvalues $\pm \omega i$ corresponding to the solution

$$\zeta(t) = B \cos(\omega t + \varphi), \qquad (28)$$

$$\eta(t) = A \cos \omega t, \qquad (29)$$

where A and B are the amplitudes of the $\eta(t)$ and $\zeta(t)$ oscillations, and where φ is a phase angle. More generally, for values of delay τ close to τ_{cr}, the nonlinear system is expected to relax to a periodic solution which can be written in the approximate form of equations (28) and (29).

Substituting equations (28) and (29) into equations (22) and (23) and matching time dependent terms results in the explicit formulae (6)–(11), which generalise those in [10].

Lindstedt's method

Changing the first order system (22)–(23) into a second order DDE results in

$$\ddot{\eta} + (\mu_m + \mu_p)\dot{\eta} + \mu_m\mu_p\eta = -K\eta_d + H_2\eta_d^2 + H_3\eta_d^3 + \ldots$$

(30)

where the coefficients K, H_2 and H_3 are obtained by Taylor series expansion of the nonlinear term.

Assuming that the true solution is periodic, our goal is to find an analytical approximation that is valid for all t. The key idea is to regard the frequency ω as unknown in advance, and to solve for it by demanding that η contains no secular terms. We introduce a small parameter , and let $\Delta = \varepsilon^2 \delta$ with

$$\eta = \varepsilon u,$$

(31)

$$\tau = \tau_{cr} + \Delta = \tau_{cr} + \varepsilon^2 \delta.$$

(32)

We then stretch time by replacing the independent variable t by σ, where

$$\sigma = \Omega t.$$

(33)

We then have

$$\Omega^2 \frac{d^2u}{d\sigma^2} + (\mu_m + \mu_p)\Omega \frac{du}{d\sigma} + \mu_m\mu_p u = -Ku_d + {}^2H_2u_d^2 + {}^2{}^2H_3u_d^3 + \ldots,$$

(34)

where $u_d = u\ (\sigma - \Omega\tau)$. Expanding Ω in a power series in , omitting the $O\ ()$ term for convenience, since it turns out to be zero, we have

$$\Omega = \omega + \varepsilon^2 k_2 + \ldots$$

(35)

Next we expand the delay term u_d

$$u_d = u\ (\sigma - \omega\tau_{cr}) - \varepsilon^2\ (k_2\ \tau_{cr} + \omega\delta)\ u'\ (\sigma - \omega\tau_{cr}) + O(\varepsilon^3)$$
(36)

Finally we expand $u\ (t)$ in a power series in to obtain

$$u\ (\sigma) = u_0\ (\sigma) + \varepsilon u_1\ (\sigma) + \varepsilon^2 u_2\ (\sigma) + \ldots$$

(37)

Substituting and collecting terms we find

$$\omega^2 \frac{d^2u_0}{d\sigma^2} + (\mu_m + \mu_p)\omega \frac{du_0}{d\sigma} + \mu_m\mu_p u_0 = -Ku_0(\sigma - \omega\tau_{cr})$$

(38)

$$\omega^2 \frac{d^2u_1}{d\sigma^2} + (\mu_m + \mu_p)\omega \frac{du_1}{d\sigma} + \mu_m\mu_p u_1 = -Ku_1(\sigma - \omega\tau_{cr}) + H_2u_0^2(\sigma - \omega\tau_{cr})$$

(39)

$$\omega^2 \frac{d^2u_2}{d\sigma^2} + (\mu_m + \mu_p)\omega \frac{du_2}{d\sigma} + \mu_m\mu_p u_2 = -Ku_2(\sigma - \omega\tau_{cr}) + H_3u_0^3(\sigma - \omega\tau_{cr})$$
$$+ 2H_2u_0(\tau - \omega\tau_{cr})u_1(\tau - \omega\tau_{cr})$$
$$+ K(k_2\tau_{cr} + \omega\delta)u_0'(\tau - \omega\tau_{cr}).$$

(40)

We take the solution of the u_0 and u_1 equations to have the form

$$u_0(\sigma) = A\cos\sigma,$$

(41)

$$u_1(\sigma) = m_1 \sin 2\sigma + m_2 \sin 2\sigma + m_3,$$

(42)

where $A = \hat{A}\varepsilon$ from (29) and (31). Next we substitute (41) and (42) into (40) to give

$$\begin{aligned}
\omega^2(-4m_1 \sin 2\sigma - 4m_2 \cos 2\sigma) + \\
\omega(\mu_m + \mu_p)(2m_1 \cos 2\sigma - 2m_2 \sin 2\sigma) + \\
Km_1(\sin 2\sigma \cos 2\omega\tau_{cr} - \cos 2\sigma \sin 2\omega\tau_{cr}) + \\
Km_2(\cos 2\sigma \cos 2\omega\tau_{cr} - \sin 2\sigma \sin 2\omega\tau_{cr}) + \\
Km_3 + \mu_m\mu_p(m_1 \sin 2\sigma + m_2 \cos 2\sigma)
\end{aligned} = \begin{aligned}
\frac{1}{2}H_2A^2 \cos 2\sigma \cos 2\omega\tau_{cr} + \\
\frac{1}{2}H_2A^2 \sin 2\sigma \sin 2\omega\tau_{cr} + \frac{1}{2}H_2A^2.
\end{aligned}$$

Collecting $\sin 2\sigma$ terms gives

$$\begin{aligned}
-4\omega^2 m_1 - 2(\mu_m + \mu_p)\omega m_2 + \\
Km_1 \cos 2\omega\tau_{cr} + Km_2 \sin 2\omega\tau_{cr} + \mu_m\mu_p m_1
\end{aligned} = \frac{1}{2}H_2A^2 \sin 2\omega\tau_{cr}.$$

Collecting $\cos 2\sigma$ terms gives

$$\begin{aligned}
-4\omega^2 m_2 - 2(\mu_m + \mu_p)\omega m_1 + \\
Km_1 - \sin 2\omega\tau_{cr} + Km_2 \cos 2\omega\tau_{cr} + \mu_m\mu_p m_2
\end{aligned} = \frac{1}{2}H_2A^2 \cos 2\omega\tau_{cr}.$$

Collecting other terms gives

$$Km_3 + \mu_m\mu_p m_3 = \frac{1}{2}H_2\hat{A}^2.$$

Letting

$$\alpha = -4\omega^2 + K\cos 2\omega\tau_{cr} + \mu_m\mu_p,$$

(43)

$$\beta = -2\omega(\mu_m + \mu_p) + K\sin 2\omega\tau_{cr},$$

(44)

$$\hat{H}_1 = \frac{1}{2} H_2 \hat{A}^2 \sin 2\omega\tau_{cr}, \tag{45}$$

$$\hat{H}_2 = \frac{1}{2} H_2 \hat{A}^2 \cos 2\omega\tau_{cr}, \tag{46}$$

$$\hat{H}_3 = \frac{1}{2} H_2 \hat{A}^2, \tag{47}$$

we find that

$$m_1 = \frac{\alpha\hat{H}_1 - \beta\hat{H}_2}{\alpha^2 + \beta^2}, \tag{48}$$

$$m_2 = \frac{\alpha\hat{H}_1 + \beta\hat{H}_2}{\alpha^2 - \beta^2}, \tag{49}$$

$$m_3 = \frac{\hat{H}_3}{K + \mu_m \mu_p}. \tag{50}$$

Now we substitute (41) and (42) into (38) and equate to zero the coefficients of the resonant terms $\sin\sigma$ and $\cos\sigma$. This gives

$$\begin{aligned} k_2(-\mu_m - \mu_p + K\tau_{cr}\cos\omega\tau_{cr}) & \\ + \omega\delta K\cos\omega\tau_{cr} = & \frac{3}{4}H_3 A^2 \sin\omega\tau_{cr} + \\ & H_2^2 A^2(m_1\cos\omega\tau_{cr} + m_2\sin\omega\tau_{cr} + 2m_3\sin\omega\tau_{cr}) \end{aligned}$$

and

$$\begin{aligned} k_2(-2\omega - K\tau_{cr}\sin\omega\tau_{cr}) & \\ - \omega\delta K\sin\omega\tau_{cr} = & \frac{3}{4}H_3 A^2 \cos\omega\tau_{cr} + \\ & H_2^2 A^2(-m_1\sin\omega\tau_{cr} + m_2\cos\omega\tau_{cr} + 2m_3\cos\omega\tau_{cr}). \end{aligned}$$

Let

$$c_1 = -\mu_m - \mu_p + K\tau_{cr}\cos\omega\tau_{cr}, \tag{51}$$

$$c_2 = -2\omega - K\tau_{cr}\sin\omega\tau_{cr}, \tag{52}$$

$$d_1 = \frac{3}{4}H_3\sin\omega\tau_{cr} + H_2^2(m_1\cos\omega\tau_{cr} + m_2\sin\omega\tau_{cr} + 2m_3\sin\omega\tau_{cr}), \tag{53}$$

$$d_2 = \frac{3}{4}H_3\cos\omega\tau_{cr} + H_2^2(-m_1\sin\omega\tau_{cr} + m_2\cos\omega\tau_{cr} + 2m_3\cos\omega\tau_{cr}). \tag{54}$$

Then we have

$$A^2 = \frac{(c_2\cos\omega\tau_{cr} + c_1\sin\omega\tau_{cr})}{d_1 c_2 - c_1 d_2}\omega\Delta K, \tag{55}$$

$$k_2 = -\frac{d_2\cos\omega\tau_{cr} + d_1\sin\omega\tau_{cr}}{c_1 d_2 - c_2 d_1}\omega\delta K, \tag{56}$$

Multiplying (56) by ^{2} and substituting back into (35) gives

$$\Omega = \omega + k_2\Delta, \tag{57}$$

which has the form of equation (15) (noting that k_2 is negative).

Maximum ω

We have seen that K takes the same value for any $\mu_m\mu_p = c$ where c is a constant. We are interested in investigating ω along $\mu_m\mu_p = c$. We will show that ω has a unique maximum along this line and the maximum value of ω occurs when $\mu_m = \mu_p = \sqrt{c}$. Rearranging equation (8) gives

$$\omega^4 + (\mu_m^2 + \mu_p^2)\omega^2 + \mu_m^2\mu_p^2 - K^2 = 0. \tag{58}$$

Substituting $\mu_m\mu_p = c$ and $K = c_K$ into the above equation gives:

$$\omega^4 + \omega^2\left(\frac{c}{\mu_p^2} + \mu_p^2\right)\omega^2 + c^2 - c_K^2 = 0. \tag{59}$$

Differentiating w.r.t. μ_p gives:

$$4\omega^3\frac{d\omega}{d\mu_p} + 2\omega\frac{d\omega}{d\mu_p}\left(\frac{c^2}{\mu_p^2} + \mu_p^2\right) + \omega^2\left(-\frac{2c^2}{\mu_p^3} + 2\mu_p\right) = 0, \tag{60}$$

So $\frac{d\omega}{d\mu_p} = 0$ when

$$\omega^2\left(\frac{-2c^2}{\mu_p^3} + 2\mu_p\right) = 0 \tag{61}$$

$$\frac{-2c^2}{\mu_p^3} + 2\mu_p = 0$$

giving

$$\mu = \sqrt{c}. \tag{62}$$

As $\mu_m\mu_p = c$, we have $\mu_m = \mu_p = \sqrt{c}$. The maximum value of ω therefore occurs when $\mu_m = \mu_p$ and has the form

$$\omega = \sqrt{K - \mu^2}.$$ (63)

Writing $K - \mu_2$ in terms of β gives

$$K - \mu^2 = \frac{n\beta - (1+\beta)}{p_0 \beta^{1/n}(1+\beta)^2}.$$ (64)

Differentiating with respect to β gives

$$\frac{\beta^{1/n}(1+\beta)^2(n-1) - (n\beta - \beta - 1)\left(\frac{1}{n}\beta^{1/n-1}(1+\beta)^2 + 2\beta^{1/n}(1+\beta)\right)}{p_0\left(\beta^{1/n}(1+\beta)^2\right)^2}.$$ (65)

Setting this expression to zero and dividing by $\beta^{1/n}(1+\beta)/p_0$ gives

$$n + n\beta - \beta - 1 - (n-1)(1/n(1+\beta) + 2\beta) + \frac{1}{n\beta}(1+\beta) + 2 = 0,$$ (66)

which simplifies to

$$(n^2 - 1)\ \beta^2 - (n^2 + 2)\ \beta - 1 = 0.$$ (67)

Can the data tell us anything about the parameters directly?

Integrating equation (2) gives

$$\int_0^T \dot{p}(t)dt = \int_0^T m(t)dt - \mu_p \int_0^T p(t)dt.$$ (68)

From the Fundamental Theorem of Calculus, it follows that

$$p(T) - p(0) = \int_0^T m(t)dt - \mu_p \int_0^T p(t)dt.$$ (69)

If p is periodic and T is a multiple of the period then the left-hand side of this equation vanishes and the right-hand side may be written $m_{ave} - \mu_p p_{ave}$. This gives us the result

$$\mu_p = \frac{m_{ave}}{p_{ave}}.$$ (70)

This allows us, from appropriate data sets, to infer the parameter μ_p by averaging data from m and p. This, of course, presupposes that the period T can be estimated accurately and that there is enough data to form good approximations to the averages.

We remark that μ_m cannot be approximated this way because of the nonlinearity in (1).

MCMC (Metropolis Hastings) Method

Our approach was to choose candidate values for our unknown parameters, θ_j^*, ($j = 1 \ldots 5$ in Experiment 1 and $j = 1 \ldots 6$ in Experiment 2) in logspace, based on the current value of $\theta_j^{(i)}$, using a Gaussian *proposal distribution*, $q\left(\theta_j^* \mid \theta_j^{(i)}\right)$ given by

$$q\left(\theta_j^* \mid \theta_j^{(i)}\right) = \frac{1}{\sigma_j\sqrt{2\pi}}\exp\left(-\frac{\left(\theta_j^* - \theta_j^{(i)}\right)^2}{2\sigma_j^2}\right).$$ (71)

σ_j values for the proposal rates are chosen in order that the acceptance rates are between 20% and 35%.

The candidate values are accepted or rejected with an *acceptance probability*, $\mathcal{A}\ (\theta^{(i)}, \theta^*)$, given by

$$\mathcal{A}\left(\theta_j^{(i)}, \theta_j^*\right) = \min\left\{1, \frac{p\left(\theta_j^*\right)q\left(\theta_j^{(i)} \mid \theta_j^*\right)}{p\left(\theta_j^{(i)}\right)q\left(\theta_j^* \mid \theta_j^{(i)}\right)}\right\},$$ (72)

where

$$\hat{p} = p(\mathcal{D} \mid \theta_j)p(\theta_j),$$ (73)

$$p(\mathcal{D} \mid \theta_j) = p(\{x_k\} \mid \theta_j, \sigma, I),$$ (74)

and

$$p(\theta_j) = \mathcal{N}(\mu_j, \sigma_{pj}).$$ (75)

A pseudo-code outline for the MCMC Metropolis Hastings algorithm is as follows:

step 1 Initialise $\theta_j^{(0)}$

step 2 For $i = 0$ to N:

Sample $r \sim \mathcal{N}\ (0, 1)$

Sample $\theta_j^* \sim q\left(\theta_j^* \mid \theta_j^{(i)}\right)$ (choosing j at random

$$\text{If } r < \mathcal{A}\left(q_j^{(i)}, q_j^*\right) = \min\left\{1, \frac{p\left(q_j^*\right)q\left(q_j^{(i)}|q_j^*\right)}{p\left(q_j^{(i)}\right)q\left(q_j^*|q_j^{(i)}\right)}\right\}:$$

$$\theta_j^{(0)} = \theta_j^*$$

else

$$\theta_j^{(0)} = \theta_j^{(i)}$$

Acknowledgements

References

1. Tomlin C, Axelrod J: **Biology by numbers: mathematical modelling in developmental biology.** *Nature Reviews Genetics* 2007, **8:**331-340.
2. Voit E: *Computational Analysis of Biochemical Systems* Cambridge University Press; 2000.
3. de Jong H: **Modeling and simulation of genetic regulatory systems: a literature review.** *J Comput Biol* 2002, **9:**67-103.
4. Rogers S, Khanin R, Girolami M: **Bayesian model-based inference of transcription factor activity.** *BMC Bioinformatics* 2006, **8 Suppl 2:**S2.
5. Alon U: *An Introduction to Systems Biology: Design Principles of Biological Circuits* Chapman and Hall/CRC; 2007.
6. Sivia D: *Data Analysis: A Bayesian Tutorial* 2nd edition. Oxford University Press; 2006.
7. Hoops S, Sahle S, Gauges R, Lee C, Pahle J, Simus N, Singhal M, Xu L, Mendes P, Kummer U: **COPASI a COmplex PAthway SImulator.** *Bioinformatics* 2006, **22:**3067-74.
8. Andrieu C, de Freitas N, Doucet A, Jordan MI: **An introduction to MCMC for machine learning.** *Machine Learning* 2003, **50:**5-43.
9. Strogatz SH: *Nonlinear Dynamics and Chaos* Addison Wesley; 1994.
10. Verdugo A, Rand R: **Hopf bifurcation in a DDE model of gene expression.** *Communications in Nonlinear Science and Numerical Simulation* 2008, **13:**235-242.
11. Bar-Or R, Maya R, Segel L, Alon U, Levine A, Oren M: **Generation of oscillations by the p53-Mdm2 feedback loop: a theoretical and experimental study.** *Proceedings of the National Academy of Sciences USA* 2000, **97:**11250-11255.
12. Monk NA: **Oscillatory Expression of Hes1, p53, and NF-kB Driven by Transcriptional Time Delays.** *Current Biology* 2003, **13:**1409-1413.
13. Hirata H, Yoshiura S, Ohtsuka T, Bessho Y, Harada T, Yoshikawa K, Kageyama R: **Oscillatory Expression of the bHLH Factor Hes1 Regulated by a Negative Feedback Loop.** *Science* 2002, **298:**.
14. Bernard S, Жajavec B, Pujo-Menjouet L, Mackey MC, Herzel H: **Modeling transcriptional feedback loops: the role of Gro/TLE1 in Hes1 oscillations.** *Philos Transact A Math Phys Eng Sci* 2006, **364(1842):**1155-70.
15. Barrio M, Burrage K, Leier A, Tian T: **Oscillatory Regulation of Hes1: Discrete Stochastic Delay Modelling and Simulation.** *PLoS Computational biology* 2006, **2:**1017-1030.
16. Wilkinson DJ: *Stochastic Modelling for Systems Biology* Chapman and Hall/CRC; 2006.
17. Beaumont MA, Rannala B: **The Bayesian revolution in genetics.** *Nature Reviews Genetics* 2004, **5:**251-261.
18. Gelman A, Carlin JB, Stern HS, Rubin DB: **.** In *Bayesian Data Analysis* Chapman and Hall CRC; 2004.