Tian, T. and Burrage, K. (2003) *Stochastic neural network models for gene regulatory networks.* In: CEC 2003: The 2003 Congress on Evolutionary Computation. Proceedings. Canberra, Australia, 8-12 December 2003. IEEE Computer Society, Piscataway, USA, pp. 162-169. ISBN 9780780378049

http://eprints.gla.ac.uk/13337/

Deposited on: 19 February 2010

Enlighten – Research publications by members of the University of Glasgow
http://eprints.gla.ac.uk

Stochastic Neural Network Models for Gene Regulatory Networks

Tianhai Tian and Kevin Burrage
Advanced Computational Modelling Centre
The University of Queensland
Brisbane, Qld 4072, Australia
tian@maths.uq.edu.au, kb@maths.uq.edu.au

Abstract-

Recent advances in gene-expression profiling technologies provide large amounts of gene expression data. This raises the possibility for a functional understanding of genome dynamics by means of mathematical modelling. As gene expression involves intrinsic noise, stochastic models are essential for better descriptions of gene regulatory networks. However, stochastic modelling for large scale gene expression data sets is still in the very early developmental stage. In this paper we present some stochastic models by introducing stochastic processes into neural network models that can describe intermediate regulation for large scale gene networks. Poisson random variables are used to represent chance events in the processes of synthesis and degradation. For expression data with normalized concentrations, exponential or normal random variables are used to realize fluctuations. Using a network with three genes, we show how to use stochastic simulations for studying robustness and stability properties of gene expression patterns under the influence of noise, and how to use stochastic models to predict statistical distributions of expression levels in a population of cells. The discussion suggests that stochastic neural network models can give better descriptions of gene regulatory networks and provide criteria for measuring the reasonableness of mathematical models.

1 Introduction

Recent advances in gene-expression profiling technologies have been providing simultaneous measurements of gene expression levels from different tissue types and different organisms at a rapid rate. The driving force behind this data collection effort is the hope that we might be able to reconstruct the underlying gene regulation networks. Progress in this field could have deep implications in bioengineering and therapeutic target discovery.

The existence of time-series data raises the possibility of determining regulatory interactions between genes. Recently, there have been many efforts for modelling gene regulatory networks using different classes of mathematical models ([Hasty01]; [de Jong02]). Mathematical models can be classified into fine-grained or coarse-grained approaches. Fine-grained approaches are based on detailed biochemical knowledge and complex networks of biochemical reactions. The complexity of these networks restricts applications of fine-grained approaches to very small systems such as stochastic mechanisms in gene expression ([McAdams97]) and bistability and genetic switching ([Hasty00]) in the gene network of Bacteriophage λ.

However, data availability usually cannot provide a comprehensive picture of biological regulation. It is essential to be able to construct coarse-grained descriptions of gene networks for studying large scale gene networks with uncertain properties. Instead of going down to the exact biochemical reactions, coarse-grained approaches analyse large gene networks at some intermediate levels by using macroscopic variables in a global fashion. Boolean network models and neural network models are two of the major coarse-grained approaches. For a very good survey of the application of network models to the problem of reverse engineering we refer to ([D'haeseleer00]). A more comprehensive comparative study of continuous network models can be found in ([Wessels01]).

A Boolean network model interprets gene interactions as connections between genes. The state of gene expression is simplified as being either completely ON or completely OFF. The Boolean expression state converges to a terminal state via a series of state transitions that is determined by the designed Boolean rules. Due to the nature of the finite transitive states, the terminal state is either a "point attractor" (steady state) if it is a single unchanged state or a "dynamic attractor" (limit cycle) if it is a series of states ([Liang98]). Recently noisy Boolean models ([Akutsu00]; [Shmulevich02]) have been proposed in order to cope with uncertainty in data and model selection.

In the case of multiple copies of the circuit, reactions in the system depend not only on the state of genes (ON/OFF) but also on the concentration of each compound controlling the process. In neural network models it is proposed that the expression of each gene is regulated by gene expression products in the network. Genes and regulatory interactions are represented in a network scheme by nodes and wirings between nodes, respectively. This network scheme is identical to that in a Boolean model in structure but is realized by a weight matrix that can interpret intermediate regulations.

There is considerable experimental evidence that indicates the presence of significant stochasticity in transcriptional regulation in both eukaryotes and prokaryotes

0-7803-7804-0 /03/$17.00 © 2003 IEEE

Authorized licensed use limited to: UNIVERSITY OF GLASGOW. Downloaded on February 19,2010 at 06:27:45 EST from IEEE Xplore. Restrictions apply.

([Kepler01]). In general, the amount of protein produced by a particular gene varies from cell to cell. Random fluctuation (noise) in biological systems can be classified into external noise due to the random variation of external control parameters and internal noise due to chance events in biochemical reactions ([Hasty00]). As living systems are optimized to function in the presence of stochastic fluctuations ([Thattai01]), it is expected that mathematical models that attempt to explain these systems should also be robust. Stochastic models can not only give more realistic simulations of biological systems but also be used as a criterion for measuring the robustness of mathematical models against stochastic noise.

Compared with recent progress in fine-grained approaches, stochastic modelling via a coarse-grained approach is still in the very early developmental stage. In this paper we introduce stochastic neural network models for studying noise in gene expression. Our aim is to provide a platform for exploring whether currently assumed mechanisms can meet the experimental requirements for robustness against stochastic noise. Random variables are introduced to model synthesis and degradation processes. By using a network with three genes we measure the robustness and stability of gene expression patterns and predict distributions of expression levels.

2 Neural network models

Neural network architecture is uniquely determined by the number of nodes and wirings between nodes. Normally the number of nodes is defined as the number of genes although it may also be defined by other factors involved in the regulatory network. Let a N-dimensional vector $\mathbf{u}(t)$ be the expression state of a gene network containing N genes. Element $u_j(t)$ is the expression state of gene j at time t. The wirings define regulatory interactions between genes, which are represented by a weight matrix $\mathbf{w}$. A wiring from gene j to gene i means a non-zero weight w_{ij}. A positive weight implies a stimulating effect (positive feedback) while a negative weight implies repression (negative feedback). A zero weight w_{ij} means no regulatory interaction. The control strength is the multiplication of weight w_{ij} and state value u_j. The total regulatory input to gene i is the sum of regulatory strengths of all genes in the network, namely

$$r_i(t) = \sum_{j=1}^{N} w_{ij}u_j(t) + \alpha_i, \quad i = 1, 2, \ldots, N, \qquad (1)$$

where α_i is a parameter to represent the influence of external inputs ([Wessels01]) or reaction delay ([Vohradsky01]). The notation $i = 1, \ldots, N$ will be omitted hereafter.

A "squashing" function then transfers the regulatory input $r_i(t)$ into a normalized transcriptional response. The commonly used squashing function is the sigmoidal transfer function, with the form

$$g_i(t) = \frac{1}{1 + e^{-r_i(t)}}. \qquad (2)$$

Other squashing functions, such as the arctan function, have also been used. This transcriptional response is a value between 0 and 1, where 0 represents complete repression and 1 represents maximal expression. By multiplying by a maximal expression rate s_i, we get the "real" expression output $s_i g_i(t)$.

Neural network models are constructed by the accumulation of gene expression products that is the difference between synthesis and degradation. It can be either in differential equation form or in difference equation form ([Wessels01]). Based on the accumulation rate, a differential equation model takes the form

$$\frac{du_i}{dt} = s_i g_i(t) - d_i u_i, \qquad (3)$$

where d_i is the degradation rate of gene i. The basal transcription is included in the parameter α_i.

Neural network models can also be represented in difference equation form based on the accumulation of expression products. The updated gene expression levels are given by

$$u_{i(n+1)} = u_{in} + h_n[s_i g_i(t_n) - d_i u_{in}], \qquad (4)$$

where $u_{in} = u_i(t_n)$ and $h_n = t_{n+1} - t_n$ is the stepsize. This discrete model is convenient for us to study noise in biological systems, especially for internal noise. Model (3) can be obtained from (4) if $h_n \to 0$ and model (4) in fact is the Euler approximation of (3).

Gene network reconstruction requires the determination of the weight matrix from experimental data. Rate parameters s_i and d_i may be measured or estimated from experimental data directly ([Vohradsky01]) or determined in network reconstruction by a reverse engineering approach ([Wahde01]).

More detailed models can also be constructed by including both transcription and translation if expression levels and protein concentrations are available. Transcription and translation are considered as being controlled in two separated stages. The expression of mRNAs with concentrations $\mathbf{r} = (r_1, \ldots, r_N)$ is controlled by regulatory proteins with concentrations $\mathbf{p} = (p_1, \ldots, p_N)$. The rate of protein formation in the process of translation is controlled by mRNA concentrations. Based on the same assumption of model (3), the rates of mRNA and protein accumulation are given by

$$\begin{aligned} \frac{dr_i}{dt} &= s_{1i} f_{1i}(\mathbf{p}, \mathbf{w}_1, b_1) - d_{1i} r_i \\ \frac{dp_i}{dt} &= s_{2i} f_{2i}(\mathbf{r}, \mathbf{w}_2, b_2) - d_{2i} p_i, \end{aligned} \qquad (5)$$

Authorized licensed use limited to: UNIVERSITY OF GLASGOW. Downloaded on February 19,2010 at 06:27:45 EST from IEEE Xplore. Restrictions apply.

where s_{1i} and s_{2i} are maximal synthesis rates, d_{1i} and d_{2i} are degradation rates, functions f_1 and f_2 are sigmoidal transfer functions 'squashing' the influence of proteins and mRNAs using weight matrices $\mathbf{w_1}$ and $\mathbf{w_2}$, respectively; while b_1 and b_2 are delay vectors.

The rate of protein accumulation may be determined only by the amount of mRNA and then $f_{2i} \equiv r_i$. In this case the time series of the amount of proteins are similar to those of mRNAs ([Vohradsky01]). In fact the one-stage model (3) uses this assumption and then the regulation of proteins is realized by that of the expression products. The difference between the time series of mRNA and protein is included in the weight matrix. However, as the rate of protein accumulation may be determined by one or more processes in translation, experimental evidence has indicated that the time series of mRNA and proteins can be different ([Vohradsky01]).

3 Stochastic neural network models

Recently researchers have investigated the effect of noise in gene regulation by experiments ([Ozbudak02]; [Elowitz02]), theoretical research ([Thattai01]; [Kepler01]) and numerical simulations ([Hasty00]; [Arkin98]). The stochastic simulation algorithm ([Gillespie77]) and stochastic differential equations are two major approaches for studying stochastic phenomena numerically in gene regulatory networks. The former is an exact method for simulating biochemical reaction systems, especially for well stirred systems with small numbers of reactants. Both approaches are based on detailed biochemical knowledge and rich data sources.

In this paper we introduce stochastic neural network models in the framework of a coarse-grained approach. We will concentrate on stochastic models based on one-stage models (3) or (4). Two-stage stochastic models can be derived similarly by introducing random variables into two-stage models (5). For experimental data with expression levels, we use Poisson processes to describe the synthesis and degradation of expression products. Corresponding to the difference model (4), stochastic models based on Poisson random variables take the form

$$u_{i(n+1)} = u_{in} + P(s_i g_i(t_n) h_n) - P(d_i u_{in} h_n), \quad (6)$$

where $P(\lambda)$ is a Poisson random variable with mean λ, whose distribution is

$$\text{Prob}\{P(\lambda) = m\} = \frac{\lambda^m}{m!} e^{-\lambda}, \; m = 0, 1, \ldots.$$

For expression data obtained by clustering techniques, gene expression levels are normalized within a range of unity in order to represent expression levels of a group of genes ([Wen98], [Wahde01]), and so gene expression and degradation should also be within that range. To this end exponential random variables are used for realizing fluctuations. Stochastic models with exponential random variables are given by

$$u_{i(n+1)} = u_{in} + \overline{E}(s_i g_i(t_n) h_n) - \overline{E}(d_i u_{in} h_n), \quad (7)$$

where $\overline{E}(\lambda)$ is an exponential random variable with mean λ. Using $\lambda_1 = 1/\lambda$, the distribution of $\overline{E}(\lambda)$ is

$$\text{Prob}\{\overline{E}(\lambda) < x\} = \int_0^x \lambda_1 e^{-\lambda_1 x} dx, \; x > 0.$$

Similar to Poisson random variables, the distribution of an exponential random variable is determined by the mean. However, the variance of these random variables are different as $\text{var}[P(\lambda)] = \lambda$ while $\text{var}[\overline{E}(\lambda)] = \lambda^2$. In order to match the variance, stochastic models can also be constructed with normal random variables, given by

$$u_{i(n+1)} = u_{in} + s_i g_i(t_n)(h_n + N_{i1}) - d_i u_{in}(h_n + N_{i2}), \quad (8)$$

where $N_{ik} \sim N(0, h_n \sigma_{ik}^2)$. The advantage of model (8) is that σ_{ik}^2 is an adjustable parameter. It can also be written in the differential equation form

$$du_i = [s_i g_i(t) - d_i u_i]dt + \sigma_{i1} s_i g_i(t) dW_{i1} - \sigma_{i2} d_i u_i dW_{i2}, \quad (9)$$

where the W_{ik} are Wiener processes whose increments $\Delta W_{ik} = W_{ik}(t + \Delta t) - W_{ik}(t)$ are normal random variables $N(0, \Delta t)$. Model (9) can be regarded as a stochastic analogue of the continuous model (3), however, it is used just in the normalized concentration case.

For stochastic models introduced in this section, the following aspects are discussed for validating the approach.

(1) The Poisson random variable has been used for representing chance events of gene expression in prokaryotes in a fine-grained approach ([Thattai01]). It has also been used in the τ-leap method ([Gillespie01]), a recent approach for approximating the stochastic simulation algorithm with greater efficiency. For a chemical reaction system with N molecular species and M reaction channels, a propensity function $a_j(X(t))$ is defined for the j-th reaction channel at current state X and time t. According to the τ-leap method, the number of times that the j-th reaction channel occurs in the time interval $[t, t + \tau)$ is a Poisson $P(a_j(X(t))\tau)$.

Motivated by these successful applications, we assume that gene expression is a process of two major reactions: synthesis and degradation, and that the numbers of gene products in both synthesis and degradation are Poisson in each time interval. This is the basis of model (6).

(2) A Poisson random variable with large mean can be approximated by a normal random variable, namely

Authorized licensed use limited to: UNIVERSITY OF GLASGOW. Downloaded on February 19,2010 at 06:27:45 EST from IEEE Xplore. Restrictions apply.

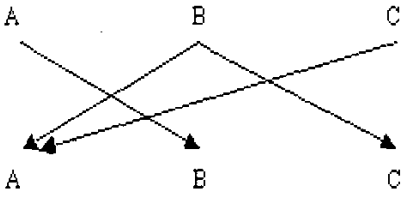

Figure 1: A network with three genes

$P(\lambda) \approx N(\lambda, \lambda)$. Thus it is appropriate to use a Poisson random variable with large mean to represent the synthesis process that is supposed to be normally distributed ([Hume00]). On the other hand, for a Poisson random variable with small mean, the generated numbers have two possible values 0 or 1 in most cases. This is a unique property of the Poisson random variable that is critical to biochemical reactions with small numbers of reactants and can be used to present the binary expression states for transcriptional reactions with small probability. In this way we can simulate the "all-or-none" phenomena stochastically.

(3) The stepsize is a key element in numerical simulations of these stochastic models. A restriction on the stepsize comes not only from the requirement of accuracy but also from biological reactions and stability. For differential equation models with large noise components, implicit methods such as the implicit Taylor methods should be used to ensure stability ([Tian01]).

4 Results

In this section we use a gene network to demonstrate the impact of noise on expression patterns. The network in Figure 1 has been studied by the Boolean model ([Somogyi96]) and the neural network model ([Vohradsky01]). In this network, the product of gene A controls the expression of gene B, which initiates the expression of gene C. Gene B induces the expression of gene A forming a positive feedback. Gene C in turn negatively controls the expression of gene A, forming a negative feedback. The regulation in this network is characterized by the weight matrix ([Vohradsky01])

$$\mathbf{w} = \begin{bmatrix} 0 & 10 & -10 \\ 10 & 0 & 0 \\ 0 & 10 & 0 \end{bmatrix}. \qquad (10)$$

According to the analysis based on the Boolean model, this network can predict two expression patterns: a point attractor (steady state) and a limit cycle (oscillation), which are determined by different initial states ([Somogyi96]). Vohradsky ([Vohradsky01]) has realized these patterns by using the neural network model with the weight matrix (10)

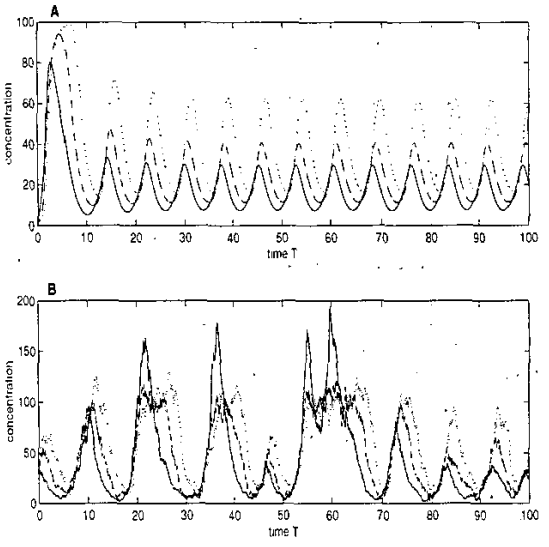

Figure 2: Simulations of the expression level case with data (11) and (12). (A) deterministic model, (B) stochastic model with Poisson random variables. The deterministic model gives sustainable oscillatory simulations. For the stochastic model, oscillations can also be observed but with periods and amplitudes that fluctuate widely in time. (u_1, solid; u_2, dotted; u_3, dashed).

and different other model parameters. It seems that oscillatory simulations can be realized only by limited parameter sets while it is quite easy to simulate steady state patterns.

We first discuss the oscillatory expression pattern. In order to use the stochastic model with Poisson random variables (6), the weight matrix

$$\mathbf{w} = \begin{bmatrix} 0 & 0.01 & -0.01 \\ 0.01 & 0 & 0 \\ 0 & 0.01 & 0 \end{bmatrix}, \qquad (11)$$

and parameters

$$\mathbf{s} = (450, 100, 100),\ \mathbf{d} = (0.6, 1, 1),\ \mathbf{b} = -(3, 3, 3) \qquad (12)$$

are used to simulate the gene network with expression levels. Simulations of the deterministic and stochastic models are presented in Figure 2. For the deterministic model (4), different initial conditions will always lead simulations to the oscillatory state. Stochastic simulations suggest that this network is not robust enough to sustain oscillatory expression patterns under the influence of internal noise.

The discussion above may be questionable because of very small elements in the weight matrix. Due to the limited parameter sets for realizing oscillatory patterns, elements in the weight matrix have to be small in order to use

Authorized licensed use limited to: UNIVERSITY OF GLASGOW. Downloaded on February 19,2010 at 06:27:45 EST from IEEE Xplore. Restrictions apply.

large synthesis rates. However, the discussion based on the weight matrix (10) in the normalized concentration case will confirm that the oscillatory pattern is not robust to internal noise. In this case, both deterministic and stochastic models are based on the weight matrix (10) and parameters

$$\mathbf{s} = (4.5, 1, 1),\ \mathbf{d} = (0.6, 1, 1),\ \mathbf{b} = -(3, 3, 3). \quad (13)$$

Simulations are presented in Figure 3. Two deterministic simulations in Figure 2 (A) and 3 (A) have similar oscillatory pattern as values of squashing functions are nearly the same. The main difference is synthesis rates. The deterministic simulation in Figure 3 (A) is similar to that presented by ([Vohradsky01]) in Figure 3 (b).

For stochastic models with normal random variables, even a quite small noise resource ($\sigma_{ik}^2 = 0.01$) will alter the oscillatory pattern significantly (Figure 3 (C)). If the variance is larger, for example $\sigma_{ik}^2 = 0.04$, it is difficult to see any oscillatory phenomena.

For the gene network with steady expression states, Figure 4 gives simulations of the deterministic model and the stochastic model with Poisson random variables based on the weight matrix (10) and parameters

$$\mathbf{s} = (3.2, 8, 6),\ \mathbf{d} = (0.2, 0.2, 0.2),\ \mathbf{b} = -(1, 1, 1). \quad (14)$$

The deterministic simulation approaches the steady state $\mathbf{u} = (16, 40, 30)$. The steady state of u_1 is determined by the balance between the positive feedback from u_2 and the negative feedback from u_3. This state then determines the state of u_2 and subsequently that of u_3 by positive feedbacks. Stochastic simulations in Figure 3 (B) show that any fluctuation in u_2 or u_3 will cause a corresponding fluctuation in u_1, which then adjust the system to the steady state.

We calculate means and variances of this stochastic simulation based on expression levels when $t \in [10, 100]$, given by

$$\overline{\mathbf{u}} = (13.4, 40.6, 29.7),\ \sigma^2(\mathbf{u}) = (11.4, 33.5, 11.4).$$

The Fano factors of this simulation are

$$\nu_i = \frac{\sigma^2(u_i)}{\overline{u}_i} < 1, \quad i = 1, 2, 3.$$

As the Fano factor of the Poisson random variable is always 1, the Fano factor is a sensitive measure of stochastic systems compared with the Poisson behaviour ([Thattai01]). Analysis indicates that the stochastic model (6) can give quite robust simulations for steady expression states with the given data.

Based on 5000 simulations we calculate numerical distributions of expression levels at $t = 25$. Numerical distributions of u_1, u_2 and u_3 can be approximated by distributions of normal random variables $N(14, 20.25)$, $N(40, 40)$ and $N(30, 30)$, respectively. It should be noticed that distributions of u_2 and u_3 are also Poisson, namely $N(40, 40) \approx P(40)$ and $N(30, 30) \approx P(30)$. Figure 5 (A) indicates that numerical distributions of u_2 can be approximated by the distribution of $N(40, 40)$ quite well. The distribution of $P(40)$ is not shown as it is very close to that of $N(40, 40)$.

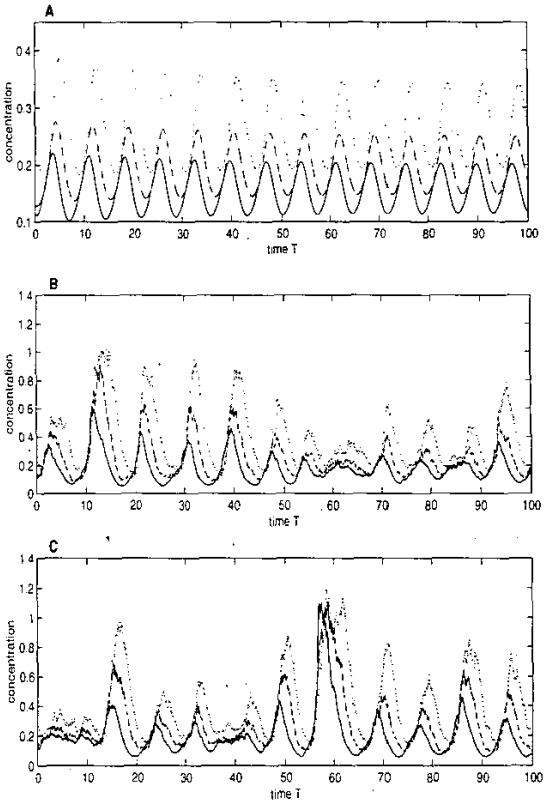

Figure 3: Simulations of the normalized concentration case with data (10) and (13). (A) deterministic model, (B) stochastic model with exponential random variables, (C) stochastic model with normal random variables ($\sigma_{ik}^2 = 0.01$). The deterministic model gives sustainable oscillatory simulations. Oscillations can be observed in these stochastic simulations but again with periods and amplitudes that fluctuate widely in time. (u_1, solid; u_2, dotted; u_3, dashed).

Figure 6 gives simulations for the normalized concentration case with the weight matrix (10) and parameters

$$\mathbf{s} = \mathbf{d} = (1, 1, 1), \quad \mathbf{b} = -\mathbf{s}. \quad (15)$$

The deterministic simulation is similar to that presented by ([Vohradsky01]) in Figure 3 (c) and reaches the steady state $u = (0.0012, 0.27, 0.85)$. For stochastic simulations,

Authorized licensed use limited to: UNIVERSITY OF GLASGOW. Downloaded on February 19,2010 at 06:27:45 EST from IEEE Xplore. Restrictions apply.

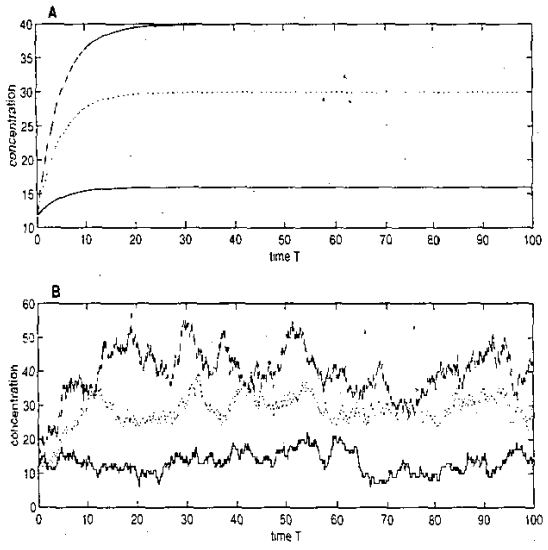

Figure 4: Simulations of the expression level case with data (11) and (14). (A) deterministic model, (B) stochastic model with Poisson random variables. (u_1, solid; u_2, dotted; u_3, dashed)

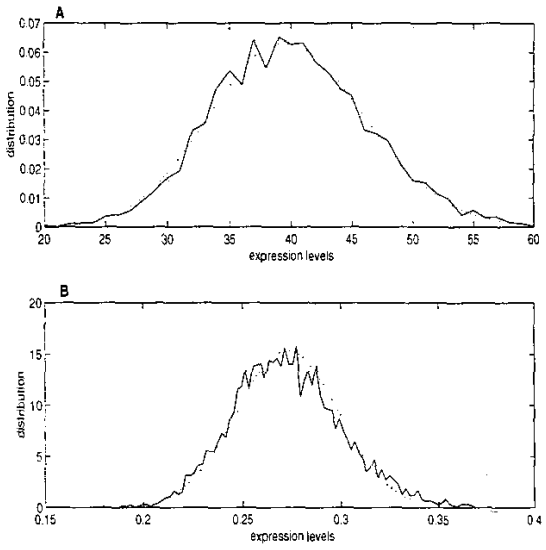

Figure 5: Distributions of numerical simulations and random variables. (A) For the stochastic model with Poisson random variables (6), numerical distribution of u_2 and distribution of normal random variable $N(40,40)$, (B) for the stochastic model with exponential random variables (7), numerical distribution of u_2 and distribution of normal random variable $N(0.27, 6.7E-4)$. (numerical, solid; random variable, dashed).

means and variances are calculated based on numerical concentrations when $t \in [10, 100]$, given by

$\overline{u}_E$=(0.0016, 0.27, 0.84),
σ^2_E=(5.1E-7, 8.2E-4, 5.4E-3),
$\overline{u}_N$=(0.0043, 0.29, 0.84),
σ^2_N=(1.6E-5, 3.6E-3, 0.021),

where subscripts E and N represent simulation results from the stochastic model with exponential and normal random variables, respectively. Similar to the Fano factor, other factors can also be defined to measure the sensitivity of stochastic systems based on exponential or normal behaviour. For example, for the exponential random variable $\overline{E}(\lambda)$ with mean λ, we have that

$$E^2[\overline{E}(\lambda)] = \text{var}[\overline{E}(\lambda)] = \lambda^2,$$

and so the following factor

$$\overline{\nu}_i = \sigma^2_E(u_i)/\overline{u}^2_E(u_i), \quad i = 1, 2, 3$$

can be defined for stochastic models with exponential random variables. Analysis indicates that both stochastic models give robust simulations with the given data.

We also use the stochastic model with exponential random variables to calculate numerical distributions of concentrations at $t = 25$ based on 5000 simulations. Numerical distributions of u_1, u_2 and u_3 can be approximated by the distributions of normal random variables $N(0.0018$, 4.9E-7), $N(0.27$, 6.7E-4) and $N(0.84, 0.0072)$, respectively. Specifically, Figure 5 (B) gives distributions of u_2 and $N(0.27, 6.7E-4)$. Similar numerical distribution results can also be obtained from the stochastic model with normal random variables.

5 Discussion

A Boolean network model gives a description of the "all-or-none" phenomena in gene expression process. However, over-simplification when using just two states cannot describe the intermediate regulation of gene products. Based on the same regulation structure, neural network models solve this problem by using a weight matrix that can represent intermediate regulation. However, the continuous nature of neural network models is an obstacle for describing chance events in biochemical reactions with small numbers of reactants. Motivated by recent success in fine-grained approaches, Poisson random variables are introduced to neural network models in the synthesis and degradation processes. For biochemical reactions with small probabilities, the generated numbers of Poisson random variables may be 0 or 1 in most cases. Thus stochastic neural network model can

Authorized licensed use limited to: UNIVERSITY OF GLASGOW. Downloaded on February 19,2010 at 06:27:45 EST from IEEE Xplore. Restrictions apply.

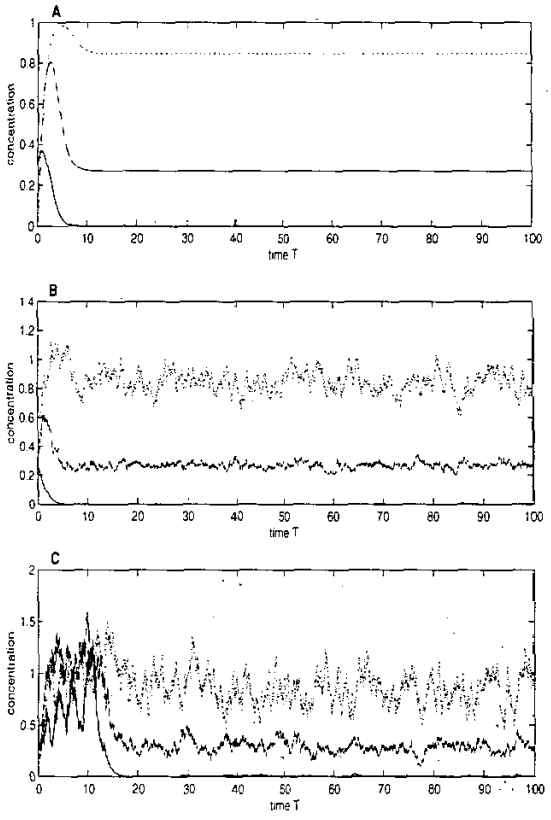

Figure 6: Simulations of the normalized concentration case with data (10) and (13). (A) deterministic model, (B) stochastic model with exponential random variables, (C) stochastic model with normal random variables ($\sigma_{ik}^2 = 0.04$). (u_1, solid; u_2, dotted; u_3, dashed).

represent not only intermediate regulation but also chance events in gene expression.

Robustness and stability are important properties of mathematical models. For gene network models, these properties have been discussed for deterministic simulations and for noise in gene expression data ([Wessels01]). The sensitivity of a model to experimental errors can be tested by simulations with perturbed input data. Here we emphasize that the discussion of these properties should include both internal noise and external noise by means of stochastic simulations. A full test should study a series of similar stochastic models using perturbed input data within the known standard derivation for each measurement.

Using a network with three genes we have discussed these properties for internal noise. For two expression patterns predicted by the deterministic model, the steady expression state is robust while the oscillatory pattern is not stable. Noise alters the oscillatory expression pattern significantly. This raises doubt for the possible "limit cycle" state predicted by Boolean models. This oscillatory pattern may be determined by the gene network or by the finite state nature of Boolean models. In the latter case the oscillatory pattern may not exist in the gene network. A similar question for plausible oscillatory expression profiles in Boolean models can be found in ([Wessels01]), and it has been observed that negative feedback in Boolean models will always cause oscillations rather than increase stability ([D'haeseleer00]). Thus it is suggested that oscillation phenomena observed in Boolean models should be restudied.

A similar study occurs for circadian oscillations where oscillatory simulations are altered by internal noise. In order to realize robust circadian rhythms, a hysteresis-based oscillation mechanism has been used, and it has been proposed that the ability to resist biochemical noise imposes strict constraints on the oscillation mechanism ([Barkai99]). For the gene network discussed in this paper, it is clear that other regulatory mechanisms are needed for generating robust oscillatory expression patterns.

Another feature of stochastic models is that they give probabilistic distributions of expression levels. This property can be used for describing the variation of expression products from cell to cell. Numerical distributions generated from stochastic models can be approximated by distributions of Poisson or normal random variables. These distribution properties give another possible means to investigate the so-called "random error" in gene expression data, especially for probes with low intensity levels ([Nadon02]). The resource of random error includes the internal noise discussed in this paper. The reliability of data mining results can be assessed by comparing random errors estimated from observed data in replication with the variation predicted by stochastic models.

Theoretical analysis and numerical simulations have indicated that stochastic models can provide better descriptions of biological systems and can be used to measure the reasonableness of mathematical models. Reliable functioning of the cell may require genetic networks that are robust to fluctuations ([Elowitz02]; [Alon99]), and so this property should be realized by mathematical models that are robust to both internal noise and external noise. This principle is useful for the reverse engineering approach for predicting gene regulatory structures. Robustness and stability can be additional criteria in order to select candidate weight matrices obtained by optimization methods such as genetic algorithms. We can filter out candidate models if they generate unstable simulations. However, this assumption will be justified by applications of stochastic neural network models to experimental gene expression profiles. This is certainly a direction for future work.

Authorized licensed use limited to: UNIVERSITY OF GLASGOW. Downloaded on February 19,2010 at 06:27:45 EST from IEEE Xplore. Restrictions apply.

Bibliography

[Akutsu00] Akutsu, T., Miyano, S. and Kuhara, S. (2000) Inferring qualitative relations in genetic networks and metabolic pathways. Bioinformatics 16, 727-734.

[Alon99] Alon, U., Surette, M.G., Barkai, N. and Leibler, S. (1999) Robustness in bacterial chemotaxis. Nature 397, 168-171.

[Arkin98] Arkin, A., Ross, J. and McAdams, H.H. (1998) Stochastic kinetic analysis of developmental pathway bifurcation in phage λ-infected *Escherichia coli* cells. Genetics 149, 1633-1648.

[Barkai99] Barkai, N. and Leibler, S. (1999) Biological rhythms - Circadian clocks limited by noise. Nature 403, 267-268.

[de Jong02] de Jong, H. (2002) Modeling and simulation of genetic regulatory systems: A literature review. J. Comput. Bio. 9, 67-103.

[D'haeseleer00] D'haeseleer, P., Liang, S., and Somogyi, R. (2000) Genetic network inference: from co-expression clustering to reverse engineering. Bioinformatics, 16 707-726.

[Elowitz02] Elowitz, M.B., Levine, A.J., Siggia, E.D. and Swain, P.S. (2002) Stochastic gene expression in a single cell. Science 297, 1183-1186.

[Gillespie01] Approximate accelerated stochastic simulation of chemically reacting systems. J. Chem. Phys. 115, 1716-1733.

[Gillespie77] Gillespie, D.T. (1977) Exact stochastic simulation of coupled chemical reactions. J. Phys. Chem. 81, 2340-2361.

[Hasty00] Hasty, J., Pradines, J., Dolnik, M. and Collins, J. J. (2000) Noise-based switches and amplifiers for gene expression. Proc. Natl. Acad. Sci. 97, 2075-2080.

[Hasty01] Hasty, J., McMillen, D., Isaacs, F. and Collins, J. J. (2001) Computational studies of gene regulatory networks: in numero molecular biology. Nat. Rev. Genet. 2, 268-279

[Hume00] Hume, D.A. (2000) Probability in transcriptional regulation and its implications for leukocyte differentiation and inducible gene expression. Blood 96, 2323-2328.

[Kepler01] Kepler, T.B. and Elston, T.C. (2001) Stochasticity in transcriptional regulation: Origins, consequences, and mathematical representations. Biophys. J. 81, 3116-3136.

[Liang98] Liang, S. Fuhrman, S. and Somogyi, R. (1998) REVEAL, a general reverse engineering algorithm for inference of genetic network architectures. Pacific Symp. Biocomput 3, 18-29.

[McAdams97] McAdams, H.H. and Arkin, A. (1997) Stochastic mechanisms in gene expression. Proc. Natl. Acad. Sci. 94, 814-819.

[Nadon02] Nadon, R. and Shoemaker, J. (2002) Statistical issues with microarrays: processing and analysis. Trends, Genet 18, 265-271.

[Ozbudak02] Ozbudak, E.M., Thattai, M., Kurtser, I., Grossman, A.D. and van Oudenaarden, A. (2002) Regulation of noise in the expression of a single gene. Nature Genet. 31, 69-73.

[Shmulevich02] Shmulevich, I., Dougherty, E.R., Kim, S. and Zhang, W. (2002) Probabilistic Boolean networks: a rule-based uncertainty model for gene regulatory networks. Bioinformatics 18, 261-274.

[Somogyi96] Somogyi, R. and Sniegoski, C.A. (1996) Modeling the complexity of genetic networks: understanding multigenic and pleiotropic regulation. Complexity 1, 45-63.

[Thattai01] Thattai, M. and van Oudenaarden, A. (2001) Intrinsic noise in gene regulatory networks. Proc. Natl. Acad. Sci. 98, 8614-8619.

[Tian01] Tian, T. and Burrage, K. (2001) Implicit Taylor methods for stiff stochastic differential equations. Appl. Numer. Math. 38, 167-185.

[Vohradsky01] Vohradsky, J. (2001) Neural network model of gene expression. FASEB J. 15, 846-854.

[Wahde01] Wahde, M and Hertz, J. (2001) Modeling genetic regulatory dynamics in neural development. J. Comput. Bio. 8, 429-442.

[Wen98] Wen, X., Fuhrman, S., Michaels, G.S., Carr, D.B., Smith, S., Barker, J.L. and Somogyi, R. (1998) Large-scale temporal gene expression mapping of central nervous system development. Proc. Natl. Acad. Sci. 95, 334-339.

[Wessels01] Wessels, L.F.A., Van Someren, E.P. and Reinders, M.J.T. (2001) A comparison of genetic network models. Pacific Symp. Biocomput. 6, 508-519.

Authorized licensed use limited to: UNIVERSITY OF GLASGOW. Downloaded on February 19,2010 at 06:27:45 EST from IEEE Xplore. Restrictions apply.