Abstract—It is well known that theoretical analysis based on the mathematical predictive models of middle-sized genetic regulatory networks are crucial to understand the life at system level for systems biology. Since there doesn’t exist an efficient mathematical model for the middle-sized or large-sized biological systems, it is also a challenge to theoretically investigate these networks. To overcome the limitations of the traditional continuous differential equations and discrete Boolean models, this paper aims to introduce a hybrid modelling approach for the general middle-sized genetic regulatory networks. This hybrid modelling method is able to quantitatively investigate the middle-sized or large-sized biological networks by merging the advantages of both continuous differential equation models and Boolean models.

I. INTRODUCTION

It is well known that system biology has received increasing attention in various disciplines over the last decade. Genetic regulatory networks, as an emerging research field, are composed by interactions between genes and genes. Some of these networks have been quantitatively investigated from the mathematical point of view. Generally speaking, there are two kinds of representative mathematical descriptions: Boolean models and differential equation models. In 1969, Kauffman [1] firstly proposed Boolean model to describe the metabolic stability and epigenesis in genetic circuits. After that, due to the relatively simplicity and parameter free of Boolean models, Boolean models had been extensively applied to investigate various biological systems. As we know now, Boolean models are very useful for the modelling of middle-sized biological systems, where tens of nodes are involved in the network. Such as the budding yeast cell cycle network [2], [3] and fission yeast cell cycle network [4]. However, Boolean models are not able to exactly characterize various biological processes. Differential equation models can elaborate reflect the evolution of molecular species in genetic circuits, especially for small-sized networks, such as the interleaved positive and negative feedback circuits [5], [6] and feed-forward loops [7]. But for middle-sized networks, for example, the yeast cell cycle network [8]-[12], differential equation models are very complicated. Moreover, how to determine the associated extensive parameters is often beyond the complexity of the original differential equations. Therefore, the modelling of these middle-sized or large-sized biological circuits based on differential equations is also a challenge problem today [13].

Considering the advantages and disadvantages of Boolean models and differential equation models, a natural idea is how to utilize the advantages of the above two modelling approaches and also simultaneously avoid the disadvantages of the above two modelling methods. The hybrid Boolean and differential equation model is a tradeoff. Note that the hybrid model was initiated by Glass and coworkers in 1973 and 1978 [14], [15], respectively. Until recently, Singhania et al. [16] have applied the above hybrid modelling approach to investigate the mammalian cell cycle network, where the cyclin-dependent kinase in the network are described by continuous linear differential equations, while the other nodes are modeled as Boolean dynamics, and the Boolean dynamic paths are determined by the attractive path in [2]. Compared with the differential equation models [10], the hybrid model is quite simple. Furthermore, the hybrid model can much better describe the detailed network dynamics than the differential equation models.

As we know now, one of the ultimate goals of bottom-up system biology is how to understand life at system level. Since the quantitative research of middle-sized or large-sized genetic regulatory networks [13]-[22] is an effective way to realize the above goal, the modelling of middle-sized networks has significant scientific meaning.

Motivated by the above various reasons, this paper aims to introduce a hybrid modelling approach for the general genetic regulatory networks, especially for the middle-sized networks. Hereafter, the central components or network motifs in the middle-sized networks are treated as continuous variables. However, the other elements are treated as discrete Boolean variables. As an effective attempt, the hybrid Boolean and continuous differential equations are introduced to model the general middle-sized genetic regulatory networks. Moreover, the dynamical behaviors of hybrid systems are also discussed.

The left paper is organized as follows. Section II briefly reviews some existing theoretical models. And the general hybrid models are constructed to characterize the general middle-sized genetic regulatory networks in Section III. Finally, some concluding remarks are given in Section IV.
II. MATHEMATICAL MODELS FOR GENETIC NETWORKS

For single component (node) biological systems, such as single gene auto-activation and single gene auto-repression, chemical master equations can be used to describe detailed molecular evolutions; For genetic regulatory networks with several nodes, one can establish ordinary differential equation models to quantitatively investigate deterministic dynamics in these systems, Michaelis-Menten equations or Hill equations are always used in modelling such highly nonlinear differential equation systems [18]. Single component or several components circuits are all simple regulatory circuits, and general format of ordinary differential equation models for these simple circuits are as follows:

\[ \frac{dx_i}{dt} = \sum_{j=1}^{M} s_{ij} f_j(x_1, x_2, ..., x_N). \]  

(1)

Where, \( x_i \) denotes protein concentration, \( s_{ij} \) represents the stoichiometric of species \( i \) in the \( j^{th} \) reaction [18]; \( f_j \) is the reaction rate of the \( j^{th} \) reaction, which is often highly nonlinear. In addition, for middle-sized genetic regulatory networks, Eq.(1) are always very complex differential algebraic equations.

For middle-sized genetic regulatory networks, another widely used kind of models is Boolean model, where each node has two possible states, “on” or “off”. Here, state update of nodes follows certain rules, from Ref. [2], [3], [4], general update rules are as follows:

\[ S_i(t + 1) = \begin{cases} 1, & \sum_j a_{ij} S_j(t) > \theta_i, \\ 0, & \sum_j a_{ij} S_j(t) < \theta_i, \\ S_i(t), & \sum_j a_{ij} S_j(t) = \theta_i. \end{cases} \]  

(2)

Here, \( S_i(t) \) denotes state of protein \( i \) at time \( t \), which can be 0 or 1, \( \theta_i \) denotes the activation threshold for protein \( i \); \( a_{ij} \) represents network topology, if protein \( j \) can activate the expression of protein \( i \), then \( a_{ij} = 1 \); if the regulation is repression, then \( a_{ij} = -1 \); if there is no interaction between \( j \) and \( i \), then \( a_{ij} = 0 \); \( a_{ij} \) can be \(-1, 1, 0\), depending on whether there is auto-degradation, auto-activation or no self-regulation.

Beyond the above two kinds of models, there are also other models, such as stochastic differential equation models, partial differential equation models [19], since they have the similar problems as ordinary differential equation models, here, we omit the discussion. As it has been mentioned in the introduction, both Boolean models and ordinary differential equation models have their defects in modelling middle-sized genetic networks, therefore, it is crucial to develop general effective models for middle-sized networks.

III. HYBRID MATHEMATICAL MODELLING FOR MIDDLE-SIZED GENETIC REGULATORY NETWORKS

A main idea of hybrid modelling for middle-sized genetic regulatory networks is that, nodes that involved in central modules are taken as continuous variables, while other not so important nodes are modeled as Boolean variables. A natural arising question is how to determine whether a node is important or not, we note that, there has been many methods to measure the relative importance of a node, such as node degree, betweenness centrality, closeness centrality [20].

As an example, we consider a general genetic network as shown in Fig.1. The above panel of Fig.1 shows an abstract middle-sized genetic regulatory interaction network, the lower panel shows part of regulatory interactions in the above panel, and also can be seen as a middle-sized network, which contains a repressilator module [21], namely \( X_1 \dashv X_2 \dashv X_3 \dashv X_1 \). Taking the lower panel of Fig.1 as an example, we consider hybrid modelling of such middle-sized regulatory network. The repressilator module may resort to oscillation in the network, if we want to investigate oscillation problems in such systems, then the repressilator module becomes important, and we can model gene \( X_1, X_2, X_3 \) as continuous variables and other nodes treated as discrete Boolean variables. \( X_1, X_2, X_3 \) can be modeled as:

\[ \frac{dx_j}{dt} = \frac{\sum_i \alpha_{ji} a_{ji}^t (x_i / K_i)^{n_i}}{1 + \sum_i \alpha_{ji}^t (x_i / K_i)^{m_i}} - d_j x_j, \quad j = 1, 2, 3; \]  

(3)

Where \( x_j \) denotes concentration of protein \( j \) form gene \( X_j \), \( K_i \) denotes activation or repression coefficient for gene \( X_i \); \( \alpha_{ji} \) is the maximum contribution of the expression of gene \( X_i \) to gene \( X_j \); \( d_j \) represents disassociation rate; \( n_i, m_i \) are...
Hill coefficients, for an activator, one has $n_i = m_i$, while for a repressor, $n_i = 0, m_i > 0$. $a_{ji}$ reflects the interactions in the network, if there is an activate interaction from gene $X_j$ to gene $X_i$, then $a_{ji} = 1$; If the interaction is a repression, then $a_{ji} = -1$; otherwise $a_{ji} = 0$. For example, in the lower panel of Fig.1, when $j = 1$, $a_{ji}$ is nonzero only for $i \in \{3, 5, 6, 8\}$.

The next crucial question is how to determine state update rules for other discrete variables. Just similar to Ref. [2], we set states update rules for $X_4, ..., X_{15}$ as follows:

$$x_i(t+1) = \begin{cases} 1, & \sum_j a_{ij}x_j(t) > \theta_i, \\ 0, & \sum_j a_{ij}x_j(t) < \theta_i, \\ x_i(t), & \sum_j a_{ij}x_j(t) = \theta_i. \end{cases} \quad i \in \{4, ..., 15\}$$

Here, $x_i(t)$ denotes state of protein $i$ at time $t$, which can be 0 or 1, represents inactivation and activation states, respectively. $\theta_i$ denotes activation threshold for protein $i$.

There are only 17 reaction rate coefficients, 7 Hill coefficients, and 12 threshold values $\theta_i$; Hill coefficients can be set as 1, 2 or 4, and threshold value $\theta_i$ can be modulated by experience. Therefore, one can only determine 17 rate constants, compared with network size $N = 15$, amounts of parameters have been greatly reduced.

Since there are no experimental data to estimate these 17 parameters, and the network in Fig.1 is far from real biological systems, one sets $K_{ji} = 1$, $d_i = 0.5$, $n_i = 2$, $\alpha_{13} = 216$, $\alpha_{15} = 0.2$, $\alpha_{16} = 10$, $\alpha_{18} = 20$, $\alpha_{213} = 10$, $\alpha_{221} = 216$, $\alpha_{32} = 216$. Following, we numerically simulate the system, for simplicity, one suppose the update steps for Boolean variables are the same as discretized continuous variables; For discrete Boolean variables, there are about $2^{12} = 4096$ possible initial values, we randomly choose a set of initial values for each simulation. Initial values for continuous variables are randomly taken as non-negative real values.

Chosen different activation threshold $\theta_i$, for example, one chooses $\theta_1 = 0.5, 8$ for all $x_i$ ($i = 4, ..., 15$), then time evolutions of continuous variables $x_1, x_2, x_3$ are shown in Fig.2, and states update for $x_4$ under $\theta_4 = 8$ are shown in Fig.3. From Fig.2, one sees that the oscillation behaviors in $x_1, x_2, x_3$ can be regulated by $\theta_i$, with the increasing of $\theta_i$, the system all can display damped oscillations, and oscillation amplitudes become bigger and bigger, $x_4$ can realize switch between “on” and “off”.

One can also choose different threshold value $\theta_i$ for different protein $i$, which may be more practical. For example, one takes $\theta_4 = 10, \theta_5 = 10, \theta_6 = 0.8, \theta_7 = 6, \theta_8 = 8, \theta_9 = 3, \theta_{10} = 4, \theta_{11} = 8, \theta_{12} = 5, \theta_{13} = 4, \theta_{14} = 6, \theta_{15} = 2$, then time evolutions of continuous variables $x_1, x_2, x_3$ and Boolean variable $x_4$ are shown in Fig.4; The system can also display damped oscillation, and initially, Boolean variable $x_4$ can switch between “off” and “on”, but with the passage of time, $x_4$ can only rest on “off” state, which can be seen as damped switch.

The above example shows that hybrid model can well reflect dynamics in the network, and hybrid model can overcome the defects of continuous models and Boolean models.

IV. Conclusion

This paper has further discussed hybrid modelling of middle-sized genetic regulatory networks. The main idea for hybrid modelling is that, genes in central modules or key regulatory components are treated as continuous variables, while other nodes are treated as discrete Boolean variables. Hybrid models can greatly reduce parameter numbers compared with ordinary differential equation models, and at the same time, they are more detailed than Boolean models; A
recent investigation on cell cycle, which modeled the mammalian cell cycle system as hybrid piecewise linear differential equations and Boolean models, revealed that hybrid models can well reflect the evolution of biological processes as compared with experimental data. Moreover, we note that, though our investigations are about genetic regulatory networks, the general modelling method can also be applied in protein-protein interaction networks, metabolic networks and so on.

The bottom-up system biology aims to understand life at the system level, mathematical models can provide predictive tools for quantitative investigations, and hybrid modelling is promising for middle-sized or large-sized genetic networks. Our investigations provide a possible way of hybrid modelling for general middle-sized genetic regulatory networks, and may have its real-world applications in the near future.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China under Grants 61025017, 11072254, 61203148.

REFERENCES