

Experimental modeling of mitotic oscillator with reversibly binding inhibitor using a nonlinear state-space based approach

José Borges^{†,*}, Vincent Verdult[‡], Michel Verhaegen[‡] and Miguel Ayala Botto[†]

[†]Department of Mechanical Engineering
Instituto Superior Técnico

Av. Rovisco Pais 1096, Lisboa, Portugal
email: {jborges, ayalabotto}@ist.utl.pt

[‡]Delft Center for Systems and Control
Mekelweg 2, 2628 CD Delft, The Netherlands
email: {v.verdult, m.verhaegen}@dcsc.tudelft.nl

ABSTRACT: In this paper a state-space approach is used to model a mitotic oscillator with reversibly binder inhibitor. This biological system is a minimal description of the situation encountered in early amphibian embryos, and characterizes the mechanisms that drive the mitosis in cell division cycle. The system is modelled using a nonlinear type of model that results as a weighted combination of local linear state-space models. The procedure to estimate these models is briefly presented. State-space models are commonly used in engineering applications and are suitable to describe different types of systems. Its estimation is possible using either prediction error or subspace methods. The use of systematic tools that are common to several engineering fields on biological systems is promising. These tools are systematic and well tested, which promotes the development of new applications in life science, e.g. the modeling of complex biological systems using system identification tools. The results presented in this paper support this statement.

KEYWORDS: systems biology, cell biology, cell division cycle, nonlinear dynamics, modelling and identification, state-space models

INTRODUCTION

The modelling of a biological system is addressed in this work using techniques that are inherent to the engineering field denominated by system identification, which is an horizontal discipline that crosses the generality of engineering specialities. The biological system used as case study describes the mitosis part of the controlled cell division cycle (CDC) encountered in early amphibian embryos and we propose to use a nonlinear state-space based approach to model it.

Quite often it is desirable to describe a given organism in terms of a mathematical model. This model can be obtained by considering the constitutive biologic laws that regulate the organism, or it can be estimated using the observations of its *inputs* and *outputs*. Once this model is available it can be simulated in a computational environment and, e.g., used to help with the understanding of mechanisms that are intrinsic to a given biological behavior, or to investigate the efficiency of drug-delivery protocols. The direct translation of these terms into an engineering language reads as the analysis of system dynamics, or the system response to a given input. As pointed out by Kitano [1], there is a rising interest in computational systems biology, in particular the application of systematic procedures for the modelling of biological systems.

Several tools for system identification have been developed that are capable of treating a multitude of systems e.g. mechanical, electrical or chemical. These tools exist booth for linear or nonlinear systems and its use is normally systematic and well understood. The motivation for the present work is twofold: to emphasize the advantages of using such an engineering approach to model a biological system and to show the potentialities of a specific identification methodology that is presented hereafter.

The CDC model consists of two parts: the checkpoint free G_2/M phase transition, i.e. the mitotic oscillator, that was developed by Goldbeter [2] and the controller of cell division frequency, which was proposed by Gardener et al. [3].

The study of the cell division cycle has received an increasing attention in recent years and its main mechanisms have been extensively studied, see e.g. [4]. The main reason for such interest is the intrinsic relation between the CDC and

*This work was partially supported by the FCT grant - SFRH/BD/16370/2004, and by the "Programa de Financiamento Plurianual de Unidades de I&D (POCTI), do Quadro Comunitário de Apoio III, Fundação para a Ciência e a Tecnologia, Ministério da Ciência, Tecnologia e Ensino Superior, Portugal" .

the development of cancer [5]. Therefore, the development of mechanisms that artificially manipulate the CDC is highly important and can benefit from using accurate CDC models.

Due to the intrinsic nonlinear dynamics inherent to this type of system a nonlinear type of state-space models, which is denominated by composite local linear models [6], will be used. Similarly to [7], the identification of model parameters will result from using a prediction error method, where a combination of separable least squares with projected gradient method is used to improve the performance of the optimization process. The modelling of biological processes was addressed before, e.g., using neural networks [8] and fuzzy models [9].

The rest of the paper is structured as follows: we start by presenting the CDC model, followed by the description of the nonlinear state-space identification methodology and model. The conditions for the identification experiment and results are then presented. Finally, a brief discussion about the overall results is presented and some conclusions are drawn.

MITOTIC OSCILLATOR WITH BINDER INHIBITOR

Although the minimal model of Goldbeter [2] is less complex than the real cell cycle, it is still able to capture the intrinsic oscillatory dynamics of the molecular mechanism that drives the CDC oscillations.

The model reduces the CDC to three proteins: *cyclin*, *Cdc2 kinase* and *cyclin protease*. *Cyclin* is continually synthesized and, once it exceeds a threshold concentration, it activates *Cdc2 kinase*. Following a similar activation mechanism, *Cdc2 kinase* activates *cyclin protease*, which in turn degrades *cyclin*, closing in this way a negative feedback loop (see Figure 1).

Cycle oscillations are generated by threshold mechanisms for the activation of *Cdc2 kinase* and *cyclin protease*, and also the time lags associated with these threshold mechanisms. The minimal CDC model proposed by Goldbeter in [2] is given by:

$$\dot{C} = v_i - k_1 \frac{XC}{C + K_5} - k_d C, \quad (1)$$

$$\dot{M} = \frac{V_1(1 - M)}{(1 - M) + K_1} - \frac{V_2 M}{M + K_2}, \quad (2)$$

$$\dot{X} = \frac{V_3(1 - X)}{(1 - X) + K_3} - \frac{V_4 X}{X + K_4}, \quad (3)$$

where,

$$V_1 = \frac{C}{C + K_6} V_{1'}, \quad V_3 = M V_{3'}. \quad (4)$$

where C denotes the *cyclin* concentration; M and X denote the fraction of active *Cdc2 kinase* and *cyclin protease*, respectively; v_i is the rate of synthesis of *cyclin*; k_1 , k_d , and K_5 characterize the kinetics of *cyclin* degradation; the parameters V_i and K_i ($i = 1, \dots, 4$) characterize the kinetics of the enzymes involved in posttranslational modification of M and X . The constant K_6 characterizes the allosteric modulation of the enzyme corresponding to V_1 . The manipulation of CDC can be achieved either by introducing mutations into the genes that regulate the cycle, or by expressing proteins that reversibly bind and inactivates any of the proteins involved in the CDC. The model proposed by Gardner [3] fits in this last case, with the *cyclin* being the target of the inhibitor,

$$\dot{Y} = v_s - d_1 Y - a_1 C Y + (a_2 + \alpha k_d) Z, \quad (5)$$

$$\dot{Z} = a_1 C Y - (a_2 + \alpha k_d + \alpha d_1) Z. \quad (6)$$

where Y denotes the concentration of unbound inhibitor and Z denotes the concentration of inhibitor-target complex. The rate constants a_1 and a_2 determine the rate of binding and release, and the dissociation constant $K_d = a_2/a_1$. The rate of inhibitor synthesis, given by v_s , is balanced by d_1 , the basal rate of inhibitor degradation. Likewise, C is degraded at a basal rate given by k_d . The inhibitor and target proteins, when bound in a complex, also are degraded by proteolytic pathways, but at a fraction, $\alpha < 1$ of the rates of d_1 and k_d .

COMPOSITE LOCAL LINEAR MODELS

Similarly to [6], the composite local linear state-space models are defined as:

$$x_{k+1} = \sum_{i=1}^s p_i(\phi_k) (A_i x_k + B_i u_k + O_i), \quad (7)$$

$$y_k = C x_k + v_k, \quad (8)$$

where s is the number of local models, $x_k \in \mathbb{R}^n$ is the state vector, $u_k \in \mathbb{R}^m$ is the input, $y_k \in \mathbb{R}^\ell$ is the output, $v_k \in \mathbb{R}^\ell$ is a white-noise sequence and $p_i(\phi_k) \in \mathbb{R}^s$ are the weighting vectors. The weighting vectors $p_i(\phi_k) \in \mathbb{R}^s$ are parameterized using normalized radial basis functions:

$$p_i(\phi_k) = \frac{r_i(\phi_k; c_i, w_i)}{\sum_{j=1}^s r_j(\phi_k; c_j, w_j)}, \quad r_i = \exp\left(-(\phi_k - c_i)^T \text{diag}(w_i)^2 (\phi_k - c_i)\right).$$

where $\phi_k \in \mathbb{R}^q$ is the scheduling vector, c_i is the center and w_i the width of the i -th radial basis function.

The goal is to determine, from a finite number of measurements of the input u_k and output y_k , the matrices A_i , B_i , O_i , C , and the centers c_i and widths w_i that describe the radial basis functions. Let the vectors c and w consist of the centers c_i and w_i , respectively, and θ parameterize the entries of matrices A_i , B_i , O_i and C , with $i = [1, \dots, s]$.

The estimation of θ , c and w , is based on the minimization of cost function $J_N(\theta, c, w) = E_N^T E_N$, where $E_N = Y_N - \hat{Y}_N$ denotes the error vector, with Y_N a vector containing N samples of the measured outputs and \hat{Y}_N a vector containing the outputs of the estimated model. This optimization is performed using an iterative Levenberg-Marquardt gradient based algorithm.

System identification using separable least squares and projected gradient

In this section a brief summary of the combined approach consisting of separable least squares and projected gradient method is presented. The complete description is presented in [7]. Consider the vector of estimated outputs defined as follows,

$$\hat{Y}_N = \Phi(\eta) \theta_\ell, \quad (9)$$

where η parameterizes the entries of matrices A_i , with $i = [1, \dots, s]$, C and the parameters c and w of radial basis functions, and θ_ℓ parameterizes the entries of matrices B_i and O_i , with $i = [1, \dots, s]$. Also $\Phi(\eta)$ is a matrix whose block-rows are defined as,

$$\Phi(k, \eta) := \sum_{j=0}^{k-1} (p(\phi_j) \otimes \tilde{u}_j)^T \otimes C \left(\prod_{h=j+1}^{k-1} \mathcal{A}_h \right), \quad (10)$$

where \otimes represents the Kronecker product, $\mathcal{A}_k := \sum_{i=1}^s A_i p_i(\phi_k)$ and $\tilde{u}_j := [u_j^T \ 1]^T$. Using equation (9), E_N can be written as $E_N = Y_N - \Phi(\eta) \theta_\ell$. Given a fixed η , the minimization of the norm of E_N , with respect to θ_ℓ yields that,

$$\hat{\theta}_\ell(\eta) = \Phi(\eta)^\dagger Y_N, \quad (11)$$

with $\Phi^\dagger = (\Phi^T \Phi)^{-1} \Phi^T$. Therefore, by replacing $\theta_\ell = \hat{\theta}_\ell(\eta)$, it is possible to define the error vector as $\tilde{E}_N(\eta) := E_N(\theta_\ell, \eta)|_{\theta_\ell = \hat{\theta}_\ell(\eta)} = (I - P(\eta)) Y_N$, where $P(\eta) = \Phi(\eta) \Phi(\eta)^\dagger$. The principle of SLS proposed by Golub and Pereyra [10] states that the estimation of $(\hat{\eta}, \hat{\theta}_\ell) = (\hat{\theta}, \hat{w}, \hat{c})$ using cost function $J_N(\theta, c, w)$ is equivalent to, first, estimate $\hat{\eta}$ by minimizing $\tilde{J}_N(\eta) := \tilde{E}_N^T \tilde{E}_N$, which is independent of θ_ℓ , and then estimate $\hat{\theta}_\ell$ using the linear least squares optimization (11).

Full parameterized A_i and C matrices are used in the optimization process. A major drawback of using full parameterized matrices in the identification of composite local linear state-space models, is the redundancy that is introduced in the optimization procedure. This redundancy problem results from the nonuniqueness of the cost function with respect to similarity transformations, i.e. $T^{-1} A_i T$ and CT , with T nonsingular, and can be efficiently solved by using projected gradient search, as described in [7].

METHODS

As proposed by Gardener et al. [3], the control of the CDC can be achieved either by changing the rate of inhibitor synthesis v_s , or by introducing modifications into the binding/release rates, respectively a_1 and a_2 . In this work we model the system

considering the first approach, since it is more intuitively related with the idea of input of a dynamical system. Similarly to the authors, we choose the *Cyclin* concentration, which is identified by C in equations (1)–(6), as the output of the model.

A Monte-Carlo simulation consisting of 100 identification experiments is used to evaluate the performance of the methodology in the identification of the controlled CDC model. Each iteration simulates 40 min of the cell division cycle, which generates 40000 data-points of input/output pairs. At each iteration a new input sequence is produced and used to drive the system to generate the estimation data set. All the computational work is implemented using the *Matlab* environment. The simulations of the differential equations that describe the controlled CDC model were performed with the *ode23* solver and using the implemented default options. The values for the fixed parameters of the controlled CDC model are collected from the reference [3]. The concentrations are dimensionless and rate parameters have dimensions min^{-1} . The values used for the binding/release rates are fixed as $a_1 = 0.75 \text{ min}^{-1}$ and $a_2 = 0.75 \text{ min}^{-1}$, therefore satisfying the constraint on the dissociation constant, i.e. $K_d = 1$.

Gardener et al. [3] suggest that the level of expression of inhibitor can be controlled dynamically with a suitably designed protein promoter. Therefore, assuming that such promoter is available, and also that the binding dynamics of inhibitor/promotor is neglectable with respect to the CDC dynamics, we have designed an input profile that is suitable to reveal the nonlinear dynamics with respect to a given value of inhibitor synthesis. Therefore, at each experiment of the Monte-Carlo simulations the input v_s is obtained as a sum $v_s = \bar{v}_s + \delta_{v_s}$ in the following way: a fixed component \bar{v}_s whose amplitude is randomly generated at each Monte-Carlo experiment following a uniform distribution on the interval $(0.0, 0.7)$. This interval was selected according to Figure 2 of [3] with $a_1 + a_2 = 1.5 \text{ min}^{-1}$, such that the CDC model will produce limit cycle oscillations in this range; a stepwise-changing component, δ_{v_s} , with step size of 1 min and amplitude randomly generated at each step using a normal distribution with variance 0.05.

At each experiment the input/output data is collected and downsampled by a factor of 100. Then it is divided into two sets: the initial set (*IdData*) is used both for the identification and validation of the nonlinear model, while the second set (*ValData*) is used only for validation purposes. The variance accounted for (VAF) is used as qualitative performance criteria,

$$\text{VAF} := \max \left\{ 1 - \frac{\text{var}(y_k - \hat{y}_k)}{\text{var}(y_k)}, 0 \right\} \times 100\%. \quad (12)$$

We use the PO-MOESP subspace method of Verhaegen [11] to provide initial estimates that are used for the startup of the optimization algorithm and, at the same time, to provide an indication for the order of the system. We schedule the local model using the input, i.e. $\phi_k = u_k$. The number of local models s is determined by trial and error according to a criteria of best data fitting.

RESULTS

The pre-analysis of the identification problem suggested that the best trade-off between model complexity and performance results from using 4 local linear models with order $n = 5$.

Figures 2(a) – 2(c) show the simulated outputs using both the linear and nonlinear models obtained at one of the Monte-Carlo experiments and with inputs from both data sets *IdData* and *ValData*. As it is already expected, the feedback controller of Gardener et al. [3] improves the linearity of the CDC model locally for a given value of inhibitor synthesis v_s . The interesting fact depicted in this set of figures is that, when the dynamics of the controlled CDC model are still predominantly nonlinear, naturally the nonlinear model is better capable of capturing this dynamic behavior.

Figures 3(a) – 3(c) show a series of histograms with the values of VAF collected along the 100 Monte-Carlo experiments. The main facts depicted in these figures are that: the nonlinear model achieves better performance than the linear model for the most number of runs; the values of VAF are zero for some runs when considering the nonlinear system, this fact can be justified with numerical problems that are intrinsic to the nonlinear optimization method used to estimate the nonlinear model; the values of VAF obtained using the nonlinear model with the *ValData* set, with 86 runs in the 90%–100% interval, further states the degree of goodness of the estimated model obtained for each value of \bar{v}_s .

CONCLUSIONS

The work described in this paper treats the identification of the controlled CDC model, as proposed in [3], using the nonlinear state-space type of model proposed in [6] and tools from system identification, as described in [7]. The quality of the results obtained along the 100 experiments of the Monte-Carlo simulations allows to conclude that this task was successfully achieved. The nonlinear models obtained at each run of the Monte-Carlo experiment characterize the dynamical behavior of the controlled CDC for a given inhibitor synthesis v_s , with \bar{v}_s fixed and δ_{v_s} time-varying with a given variance.

An input sequence v_s was designed and applied at each run to uncover the nonlinear dynamics of the system. Since the experiments were carried out having an engineering perspective, the feasibility of such input sequence from the life science

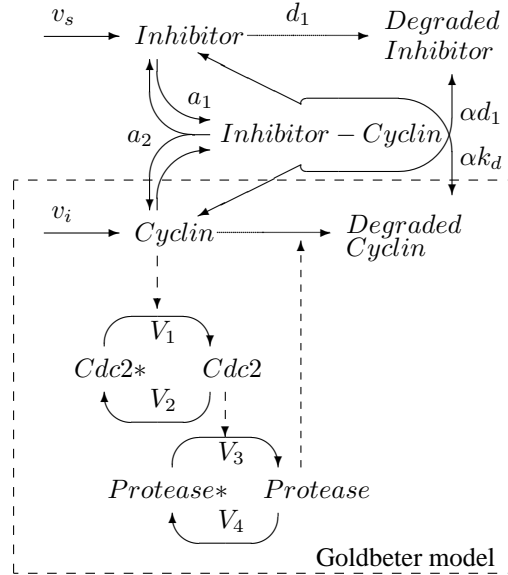
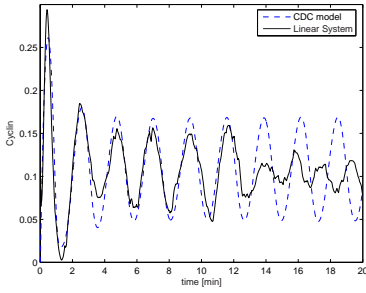
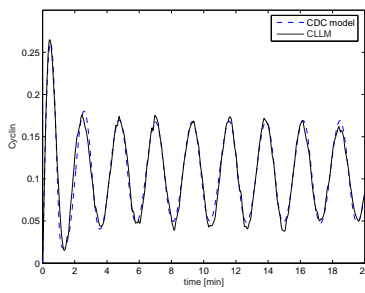


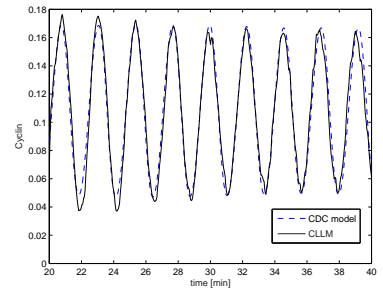
Figure 1: Control of the Goldbeter model [2] using a reversibly binder inhibitor, as proposed by Gardner et al. [3].



(a) Output obtained by simulation of estimated linear model using the input from *IdData* (VAF=62.17%).

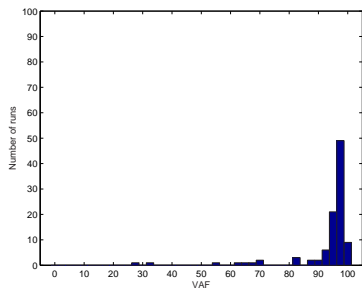


(b) Output obtained by simulation of estimated nonlinear model using the input from *IdData* (VAF=97.53%).

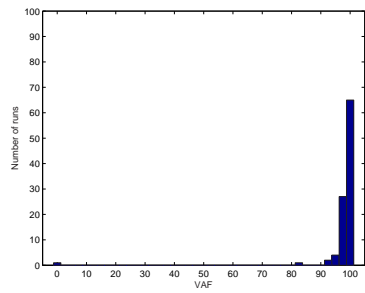


(c) Output obtained by simulation of estimated nonlinear model using the input from *ValData* (VAF=96.96%).

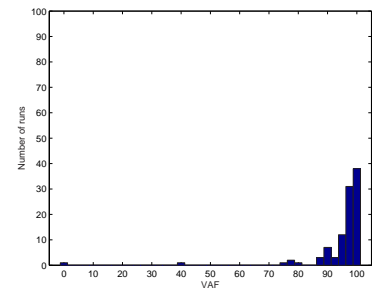
Figure 2: Simulations using the models estimated at one of the Monte-Carlo experiments ($\bar{v}_s = 0.164 \text{ min}^{-1}$).



(a) Linear model using the identification data-set.



(b) Nonlinear model using the identification data-set.



(c) Nonlinear model using the validation data-set.

Figure 3: Histograms with the VAF considering the complete 100 Monte-Carlo experiments.

perspective is a point of discussion. Actually, the input design for the identification of engineering applications is complex and normally requires a detailed level of knowledge about the system and conditions to perform the experiments.

The experiments were carried out within a simulation environment, where all the conditions and means are controlled by the user. A more challenging task is to design an *in vitro* experiment, with definition of protocols and collection of data, similarly to what is the current practice in engineering applications.

Rather than a final solution, this work suggests that more meaningful results are likely to be achieved if both the engineering perspective, with its systematic methods for system identification, and the life science, with its knowledge of biological systems, come together to deal with the complexity of this field and take benefit of each-other's expertise.

References

- [1] H. Kitano. Computational systems biology. *Nature*, 420:206–210, November 2002.
- [2] A. Goldbeter. A theory for controlling cell cycle dynamics using a reversibly binding inhibitor. *Proceedings National Academy Science USA*, 88(20):9107–9111, October 1991.
- [3] T.S. Gardner, Dolnik M., and J.J. Collins. A theory for controlling cell cycle dynamics using a reversibly binding inhibitor. *Proceedings National Academy Science USA*, 95(24):November, 14190–14195 1998.
- [4] J. Lukas and J. Bartek. Cell division: The heart of the cycle. *Nature*, 432:564–567, December 2004.
- [5] L. H. Hartwell and M.B. Kastan. Cell cycle control and cancer. *Science*, 266:1821–1828, December 1994.
- [6] V. Verdult, L. Ljung, and M. Verhaegen. Identification of composite local linear state-space models using a projected gradient search. *International Journal of Control*, 75(16/17):1385–1398, 2002.
- [7] J. Borges, V. Verdult, M. Verhaegen, and M. Ayala Botto. Separable least squares for projected gradient identification of composite local linear state-space models. In *Proceedings of the 16th International Symposium on Mathematical Theory of Networks and Systems*, Leuven, Belgium, July 2004.
- [8] H.J.L. van Can, H.A.B. te Braake, A. Bijman, C. Hellinga, K.Ch.A.M. Luyben, and HeijnenJ.J. An efficient model development strategy for bioprocesses based on neural networks in macroscopic balances: Part II. *Biotechnology and Bioengineering*, 62(6):666–681, 1999.
- [9] J.A. Roubos, P. Krabben, W.T.A.M. de Laat, R. Babuska, and J.J. Heijnen. Clavulanic acid degradation in streptomyces clavuligerus fed-batch cultivations. *Biotechnology Progress*, 18(3):451–457, 2002.
- [10] G. H. Golub and V. Pereyra. The differentiation of pseudo-inverse and nonlinear least squares problems whose variables separate. *SIAM J. of Numerical Analysis*, 10(2):413–432, April 1973.
- [11] M. Verhaegen. Identification of the deterministic part of MIMO state space models given in innovations form from input-output data. *Automatica*, 30(1):61–74, 1994.