Digital clocks – Boolean models of circadian systems

Ozgur E. Akman (O.E. Akman@exeter.ac.uk)

Introduction

The GRNs that comprise the circadian clock modulate biological function across a range of scales, from gene expression to performance and adaptive behaviour [1].

However, optimising the large parameter sets characteristic of these models places large demands on computational and experimental resources; this constrains the size and complexity of the models that can be constructed from data (the parameter explosion problem) [2].

In recent years, it has been shown that reduced models and evolutionary computing can dramatically reduce both the parametrisation and computational load, making the state and parameter spaces more tractable [3]. In particular, models based on **Boolean delay equations** (**BDEs**) fitted to time series data using evolutionary algorithms can accurately reproduce the behaviour of more biochemically detailed models (*e.g.* ordinary differential Equations (ODEs)) [2,4].

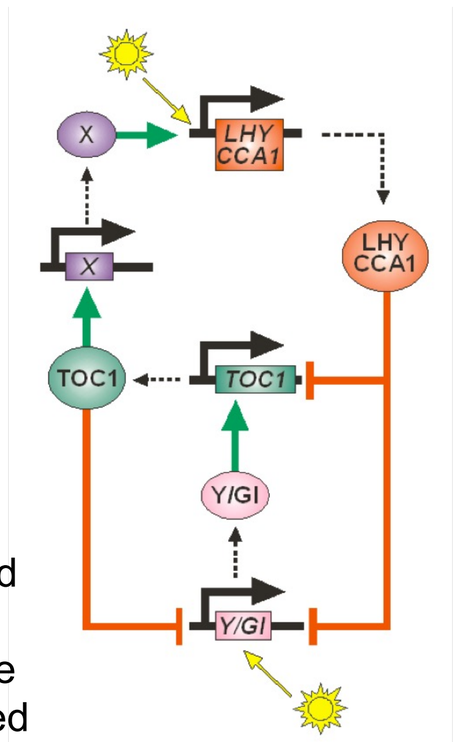

Fig.1. The circadian clock for *Arabidopsis thaliana* [5].

Boolean Delay Equations

In BDEs, the states $x_i(t)$ are represented by logical variables that are either ON (x_i=1) or OFF (x_i=0). A system of n BDEs is written as

$$x_i(t) = f_i\big(x_1(t-\tau_{i1}), x_2(t-\tau_{i2}), \dots, x_n(t-\tau_{in})\big);\ 1 \le i \le n,$$

where the *signalling delays* τ_{ij} prescribe the time it takes for x_j to affect x_i and the *logic gates* f_i specify how the interactions between $\{x_1,\dots, x_n\}$ determine the state of x_i [6,7]. In modelling GRNs, all the kinetic constants controlling the production and action of a transcription factor are thus telescoped into a single delay [2]. For example, the BDE formulation of the circuit in **Fig. 1** is the following:

$$\begin{aligned} LHY(t) &= X(t-\tau_3)L_1(t-\tau_7), \\ TOC1(t) &= \overline{LHY}(t-\tau_1)Y(t-\tau_6), \\ X(t) &= TOC1(t-\tau_2), \\ TOC1(t) &= \overline{LHY}(t-\tau_4)\overline{TOC1}(t-\tau_5)\big(L_2(t-\tau_8)+L_3(t-\tau_9)\big). \end{aligned} \quad \textbf{(1)}$$

This provides a significant reduction in complexity, compared with the ordinary differential equation model of the same circuit [2], as can be seen in **Fig. 2**. Indeed, the number of state variables is decreased from **thirteen** to **four**, whilst the number of parameters is reduced from **64** to **15**. This compression of the parameter space can help mitigate issues related to overfitting and identifiability that are common with ODE formulations, as well as accelerating the process of finding good fits to data.

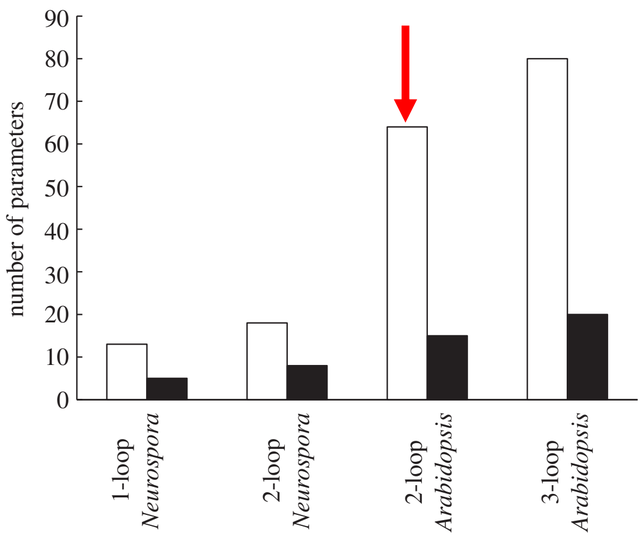

Fig. 2. The number of parameters required for ODE (white) and BDE (black) models of different circadian networks [2].

Fitting BDEs to data

In fitting Boolean variables $x_i(t)$ to continuous data $D_i(t)$, it is necessary to introduce *discretisation thresholds* T_i [2,4]. The goodness-of-fit of a BDE model for a given combination of delays and discretisation thresholds is then quantified by the cost function

$$C(\tau, T) = \frac{1}{nT}\sum_{i=1}^{n}\int_0^T |\hat{x}_i(t;\tau_1,\dots,\tau_N) - D_i(t;T_i)|dt \quad \textbf{(2)}$$

that computes the Hamming distance between the predictions $\hat{x}_i(t;\tau)$ and discretised data $D_i(t;T_i)$ for each model variable. This natural costing method contrasts with ODEs, for which there are number of possible cost functions (*e.g.* least squares), each of which defines a different optimisation problem [4]. Moreover, it avoids the arbitrary cost terms that are often necessary in continuous models to capture qualitative features of the target dynamics [8].

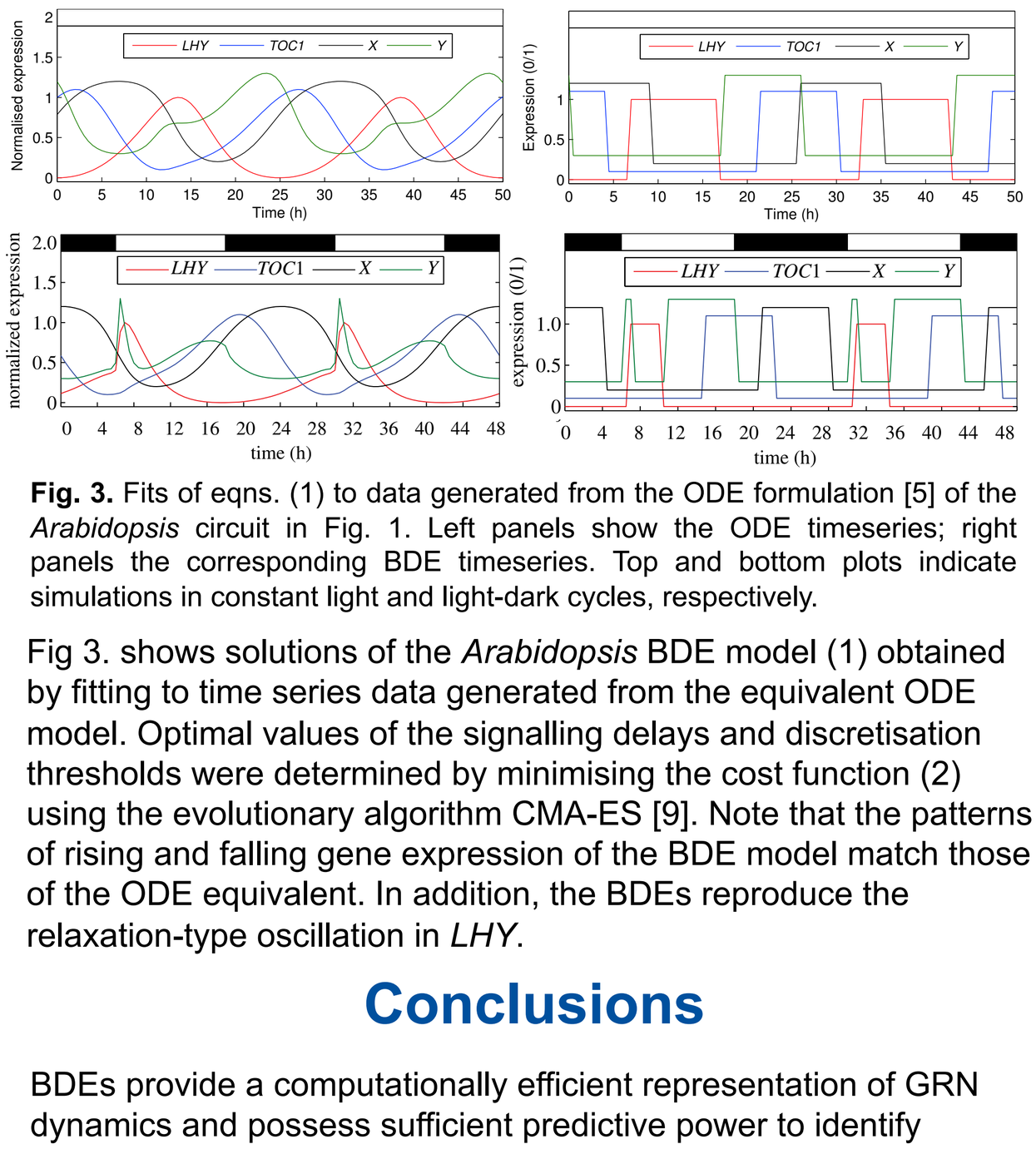

Fig. 3. Fits of eqns. (1) to data generated from the ODE formulation [5] of the *Arabidopsis* circuit in Fig. 1. Left panels show the ODE timeseries; right panels the corresponding BDE timeseries. Top and bottom plots indicate simulations in constant light and light-dark cycles, respectively.

Fig 3. shows solutions of the *Arabidopsis* BDE model (1) obtained by fitting to time series data generated from the equivalent ODE model. Optimal values of the signalling delays and discretisation thresholds were determined by minimising the cost function (2) using the evolutionary algorithm CMA-ES [9]. Note that the patterns of rising and falling gene expression of the BDE model match those of the ODE equivalent. In addition, the BDEs reproduce the relaxation-type oscillation in *LHY*.

Conclusions

BDEs provide a computationally efficient representation of GRN dynamics and possess sufficient predictive power to identify optimal regulatory structures from experimental data [4,10]. Advances in statistical methods, such as cost landscape analysis [11] and emulation could further expand the scope of BDE modelling to multi-scale models that integrate regulatory and metabolic pathways. Towards this aim, a MATLAB package for numerically solving BDEs **(BDEtools)** was recently released [12].

References

[1] E.E. Zhang & S. Kay. *Nat Rev Mol Cell Biol*, 11: 764–776 (2010). [2] O.E Akman *et al*. *J Roy Soc Interface,* 9: 2365–2382 (2012). [3] M. Foo, D.G. Bates & O.E. Akman. *PLoS Comput Biol*, 16: e1007671 (2020). [4] O.E. Akman & J.E. Fieldsend. *Proc BICOB 2020*, 70: 149–162 (2020). [5] J.C.W. Locke *et al*. *Mol Syst Biol*, 1:2005.0013 (2005). [6] D. Dee & M. Ghil. *SIAM J Appl Math*, 44: 111–126 (1984). [7] M. Ghil, I. Zaliapin & B. Coluzzi. *Physica D,* 237: 2967–2986 (2008). [8] O.E. Akman *et al*. *BMC Syst Biol*, 4: 88 (2010). [9] N. Hansen & A. Ostermeier. *Evol Comput*, 9: 159–195 (2001). [10] K. Doherty *et al*. *Proc GECCO 2017*, 1644–1651 (2017). [11] K. Alyahya *et al*. *Metaheuristics for Finding Multiple Solutions*, 229–258 (2021). [12] Akman. https://github.com/oeakman/BDEtools

UNIVERSITY OF EXETER
College of Engineering, Mathematics and Physical Sciences