

Reducing the Complexity of Mathematical Models for the Plant Circadian Clock by Distributed Delays

Isao T. Tokuda^{a,*}, Ozgur E. Akman^b, James C. W. Locke^{c,*}

^a*Graduate School of Science and Engineering, Ritsumeikan University, Noji-higashi, Kusatsu, Shiga 525-8577, Japan*

^b*College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter, EX4 4QD, UK*

^c*Sainsbury Laboratory, University of Cambridge, Bateman Street, Cambridge, CB2 1LR, UK*

Abstract

A major bottleneck in the modelling of biological networks is the parameter explosion problem – the exponential increase in the number of parameters that need to be optimised to data as the size of the model increases. Here, we address this problem in the context of the plant circadian clock by applying the method of distributed delays. We show that using this approach, the system architecture can be simplified efficiently – reducing the number of parameters – whilst still preserving the core mechanistic dynamics of the gene regulatory network. Compared to models with discrete time-delays, which are governed by functional differential equations, the distributed delay models can be converted into sets of equivalent ordinary differential equations, enabling the use of standard methods for numerical integration, and for stability and bifurcation analyses. We demonstrate the efficiency of our modelling approach by applying it to three exemplar mathematical models of the *Arabidopsis* circadian clock of varying complexity, obtaining significant reductions in complexity in each case. Moreover, we revise one of the most up-to-date *Arabidopsis* models, updating the regulation of the *PRR9* and *PRR7* genes by *LHY* in accordance with recent experimental data. The revised model more accurately reproduces the *LHY*-induction experiments

*Corresponding authors.

Email addresses: isao@fc.ritsumei.ac.jp (Isao T. Tokuda),
O.E.Akman@exeter.ac.uk (Ozgur E. Akman), James.Locke@slcu.cam.ac.uk (James C. W. Locke)

of core clock genes, compared with the original model. Our work thus shows that the method of distributed delays facilitates the optimisation and reformulation of genetic network models.

Keywords: Plant circadian clock, *Arabidopsis thaliana*, Distributed delays, Parameter optimisation, Systems biology, computational modelling

1. Introduction

Almost all living organisms co-ordinate their activities with daily environmental changes. Endogenous circadian clocks with approximately 24 h rhythmicity have been found in most eukaryotes as well as in some prokaryotes, *e.g.*, cyanobacteria (Dunlap et al., 2004). Circadian rhythms control the timing of diverse biological processes throughout the day/night cycle, ranging from flowering, gene expression, leaf movement, photosynthesis, and plant growth (Dodd et al., 2005; McClung, 2006; Atamian et al., 2016) to the anticipation of herbivory (Goodspeed et al., 2012).

Circadian clocks involve feedback loops that generate rhythmic expression of core clock genes. Molecular genetic studies in the higher plant *Arabidopsis thaliana* have revealed a complex clock network. For systematic understanding of clock function, mathematical modelling of the gene regulatory network has become one of the central approaches used (Locke et al., 2005a,b, 2006; Pokhilko et al., 2010, 2012; Akman et al., 2012; Bujdoso and Davis, 2013; Fogelmark and Troein, 2014; Foo et al., 2016; De Caluwé et al., 2016). Despite a large collection of advanced molecular experiments (Flis et al., 2015), many biochemical details – such as the values of chemical reaction constants – are still unknown. Parameter optimisation is therefore indispensable in reproducing quantitative features of the experimental data (Locke et al., 2005a; Pokhilko et al., 2010; Akman et al., 2008, 2010, 2012; Foo et al., 2016; De Caluwé et al., 2016; Doherty et al., 2017). The optimisation procedure, however, rapidly becomes computationally intractable as the network architecture gets more complex, due to the concomitant increase in the number of unknown parameters.

One approach to mitigating this parameter explosion problem is to reduce the number of unknown parameters, *e.g.*, by simplifying the network architecture (Akman et al., 2012). In this work, we simplify the plant clock architecture by introducing distributed time delays (Mittler et al., 1998; MacDonald, 2008; Akman et al., 2008, 2010) to represent protein translation and

transport. The idea of time-delayed feedback is conceptually much simpler than biochemical reaction models and allows a potentially large number of parameters related to protein production and modification to be dropped.

Time-delay models have been used previously in computational biological rhythms research (Mackey and Glass, 1977), including circadian biology (olde Scheper et al., 1999; Lema et al., 2000; Smolen et al., 2001, 2002; Sriram and Gopinathan, 2004; Akman et al., 2008; Troein et al., 2009; Akman et al., 2010, 2012; Korenčič et al., 2012, 2014; Ananthasubramaniam et al., 2014). These models are mostly based on discrete (delta-function) delays, in which only the states at specific past times are considered. One drawback of the discrete delay models is that they are governed by delay-differential equations that belong to a class of functional differential equations (Hale, 1977; Hale and Lunel, 2013). Thus, more advanced numerical techniques are required for their simulation and analysis (*e.g.* for determining stability properties and bifurcations). Distributed delay models, on the other hand, can be cast into equivalent sets of ordinary differential equations by the linear chain trick (MacDonald, 2008), enabling standard stability and bifurcation analysis methods to be used. The distributed delay approach has been applied to understand the viral dynamics of HIV-1 infection (Mittler et al., 1998) and the design principles of the fungal circadian clock (Akman et al., 2008, 2010), but not extensively to other systems.

The aim of the present paper is to investigate the utility of the distributed delay technique in reducing the complexity of plant circadian clock models. To this end, three established models of the *Arabidopsis* circadian clock of varying complexity are simplified using the distributed delay framework: (i) the single feedback loop model of Locke et al. (2005a); (ii) the Kernel model of Foo et al. (2016); and (iii) the compact model of De Caluwé et al. (2016). We show that, in the three models, this leads to a marked reduction in the number of parameters (38%, 13% and 21%, respectively). Despite this simplification, essential properties of each clock model – such as the free-running period and the phases of entrainment – are preserved. Furthermore, we exploit the advantages conferred by the reduction in complexity to efficiently revise one of the most up-to-date *Arabidopsis* models, the compact model of De Caluwé et al. (2016). Specifically, we replace the activation of the *PRR9* and *PRR7* genes by *LHY* with repression, in line with recent experimental and computational work (Fogelmark and Troein, 2014; Adams et al., 2015). The revised compact model reproduces the experimental data more precisely than the original one, demonstrating the power of our approach.

The paper is organized as follows. In Section 2, the distributed delay method is introduced. In Section 3, the method is applied to each of the three chosen *Arabidopsis* circadian clock models (Locke et al., 2005a; De Caluwé et al., 2016; Foo et al., 2016). Simulations generated with the revised compact model are then compared with the *LHY*-induction experiment of Adams et al. (2015). The final section is devoted to discussion and future work.

2. Methods & Results

2.1. Distributed delay modelling of protein dynamics

Let us denote the mRNA concentration of the i th gene in a gene regulatory network by $c_i^{(m)}(t)$. In the distributed delay framework (MacDonald, 2008), the corresponding protein level $c_i^{(d_{n_i})}(t)$ in the nucleus is modelled by the following integral of the delayed mRNA level (Akman et al., 2008, 2010):

$$c_i^{(d_{n_i})}(t) = \frac{1}{f_i} \int_{-\infty}^t G_{f_i}^{n_i-1}(t-s) c_i^{(m)}(s) ds. \quad (1)$$

Here, the delay kernel $G_{f_i}^{n_i-1}(t) = \frac{t^{n_i-1}}{(n_i-1)!} f_i^{n_i} e^{-f_i t}$ specifies the distribution in delays that are introduced by translation and protein modification processes, which govern the action of $c_i^{(d_{n_i})}(t)$ as a transcription factor (Akman et al., 2008, 2010). $G_{f_i}^{n_i-1}(t)$ is a gamma distribution, parametrised by an integer-valued shape parameter n_i and a real-valued rate parameter f_i , yielding a mean delay of n_i/f_i and a variance of n_i/f_i^2 . Using the linear chain trick (MacDonald, 2008), the delay term (1) can be computed by solving the equivalent set of ordinary differential equations

$$\begin{aligned} \frac{dc_i^{(d_1)}}{dt} &= c_i^{(m)}(t) - f_i c_i^{(d_1)}(t), \\ \frac{dc_i^{(d_2)}}{dt} &= f_i c_i^{(d_1)}(t) - f_i c_i^{(d_2)}(t), \\ &\vdots \\ \frac{dc_i^{(d_{n_i})}}{dt} &= f_i c_i^{(d_{n_i-1})}(t) - f_i c_i^{(d_{n_i})}(t). \end{aligned} \quad (2)$$

The shape parameter n_i thus specifies the number $(n_i - 1)$ of intermediate species $\{c_i^{(d_j)} : j = 1, 2, \dots, n_i - 1\}$ in the chain, whilst the rate parameter f_i specifies the rate at which each intermediate species $c_i^{(d_j)}$ is converted into the subsequent species $c_i^{(d_{j+1})}$ in the chain.

2.2. Distributed delay modelling of the plant circadian clock

In plant circadian clocks, there are a variety of protein-related biochemical processes that yield oscillation-generating delays – these include translation of mRNA, transport between cytoplasmic and nuclear proteins and protein phosphorylation (Kurosawa et al., 2002). Since experimental values are mostly unavailable for the rate constants of these processes, the corresponding parameters are fitted to experimental data. The distributed delay method therefore simplifies this optimisation procedure by reducing the number of fitted parameters in each protein production term to only the shape parameters n_i and the rate constants f_i (Akman et al., 2008, 2010).

In the following sections, we demonstrate the efficacy of our distributed delay modelling approach by applying it to each of the three *Arabidopsis* circadian clock models in turn (Locke et al., 2005a; De Caluwé et al., 2016; Foo et al., 2016). We note that for simplicity, in each model considered, the shape parameters n_i were set equal to the same value (and hence counted as a single parameter). All simulations were carried out using MATLAB (R2018a, MathWorks, Inc.). Numerical integration of ODEs was performed using the *ode23tb* solver. Bifurcation diagrams were generated by combining the *fsolve* and *eig* solvers.

2.3. Distributed delay model of the single feedback loop system

Among various mathematical models developed for plant circadian clocks (Bujdoso and Davis, 2013), we first selected the simplest single feedback loop model comprising one repressor and one activator (Locke et al., 2005a). The repressor combines two clock genes, *LATE ELONGATED HYPOCOTYL (LHY)* and *CIRCADIAN CLOCK ASSOCIATED 1 (CCA1)*, whereas the activator represents *TIMING OF CAB EXPRESSION 1 (TOC1)* (Note that at the time that the model was constructed, TOC1 was believed to be an activator, rather than a repressor (Huang et al., 2012)). Despite its simplicity, this architecture captures some core features of the clock gene circuit and quantitatively reproduces experimental results. As described in Appendix A.1, the original model – which we refer to as L2005A – consists of 7 state variables: mRNA and cytoplasmic/nuclear protein levels for the two genes and the level of a light-sensitive protein. By replacing the protein concentration equations with distributed delay terms of the form (1), the original

system can be reduced to the following 3 equations:¹

$$\begin{aligned}
\frac{dc_L^{(m)}}{dt} &= q_1 c_P L(t) + \frac{h_1 c_T^{(d_{n_T})^a}}{g_1^a + c_T^{(d_{n_T})^a}} - \frac{m_1 c_L^{(m)}}{k_1 + c_L^{(m)}}, \\
\frac{dc_T^{(m)}}{dt} &= \frac{h_2 g_2^b}{g_2^b + c_L^{(d_{n_L})^b}} - \frac{m_4 c_T^{(m)}}{k_4 + c_T^{(m)}}, \\
\frac{dc_P}{dt} &= (1 - L(t)) p_3 - \frac{m_7 c_P}{k_7 + c_P} - q_2 L(t) c_P.
\end{aligned} \tag{3}$$

Here, $c_i^{(m)}(t)$ and $c_i^{(d_{n_i})}(t)$ represent mRNA concentration and the delayed product of the i th gene, respectively (*i.e.* the final (n_i th) species in the chain), whilst $i = L$ and T correspond to the *LHY/CCA1* and *TOC1* genes, respectively. The rate constants denote transcription (h_l, g_l), degradation (m_l, k_l), translation (p_l), transport between the nucleus and cytoplasm (r_l) and the light input (q_l). (a, b) are the Hill coefficients. The concentration level c_P of the protein P is regulated by the light input $L(t)$ ($L = 1$ and $L = 0$, when light is on and off, respectively). Note that in order to facilitate comparison between the original, fully parametrised models and their distributed delay formulations, we preserve the parameter names used for describing equivalent processes (*e.g.*, in the single feedback loop model, k_4 denotes the Michaelis constant for *TOC1* mRNA degradation in both formulations). This convention is applied to all three models considered here.

To restrict the parameter search, the Hill coefficients were fixed to the same values as in the original model ($a = 1, b = 2$) (Locke et al., 2005a). For a fixed choice of shape parameters (n_L, n_T), the remaining parameters,

$$\Phi = (f_L, f_T, h_1, h_2, g_1, g_2, m_1, m_4, m_7, k_1, k_4, k_7, p_3, q_1, q_2),$$

were optimised by minimizing the following cost function:

$$E(\Phi) = \frac{1}{\sigma^2} \sum_{i=L,T} \left[\min_{\alpha_{i,1}, \alpha_{i,2}} \sum_{j=1}^N \{ \alpha_{i1} c_i^{(m)}(t_j) + \alpha_{i2} - \tilde{c}_i^{(m)}(t_j) \}^2 \right]. \tag{4}$$

In the above, $\tilde{c}_i^{(m)}(t_j)$ represents the mRNA concentration of the i th gene generated from the original single feedback loop model. For each parameter

¹Here and thereafter, for brevity we do not show the equations associated with the distributed delay terms, as these are functionally homogenous. For this model, these yield a further 10 differential equations.

combination Φ , the scaling constants $\alpha_{i,1}$ and $\alpha_{i,2}$ were obtained by minimizing the least-squares-error between the time traces $\tilde{c}_i^{(m)}(t_j)$ and $c_i^{(m)}(t_j)$. The normalisation constant is given by $\sigma^2 = \sum_{i=L,T} \sum_{j=1}^N \{\tilde{c}_i^{(m)}(t_j) - \bar{c}\}^2$, where $\bar{c} = \frac{1}{2N} \sum_{i=L,T} \sum_{j=1}^N \tilde{c}_i^{(m)}(t_j)$. A normalized error $E(\Phi)$ of 1 therefore implies that the difference between the timeseries of the distributed delay and original models is comparable to the variation in the trajectory of the original model, whilst an error close to 0 indicates that the two model timeseries are very similar to each other. In the computation of (4), the delay model and the original model were simulated under light-dark cycles (12h:12h) for 5 days, under constant light (LL) for 3 days, under light-dark cycles (12h:12h) again for 5 days, and then under constant darkness (DD) for 3 days. After removing the first 3 days of each LD phase as transient dynamics, the remaining 10 days of data (sampling interval of 1 h) were used for evaluating the cost function.

Using the Sobol algorithm, 10^5 quasi-random points were generated in the parameter space $\Phi \subset [0.1, 10]^{15}$. From the Sobol points, the cost function was calculated and sorted in descending order. The 100 points with the lowest cost function score obtained from the Sobol sampling were then passed to a simulated annealing routine for further minimisation (Locke et al., 2005a,b, 2006; Akman et al., 2008, 2010).

Following (Akman et al., 2008, 2010), the shape parameters (n_L, n_T) were initially both set to 2 in order to minimize the number of equations in each chain (*cf.* equations (2)); however, this yielded poor fits. Consequently, (n_L, n_T) were incremented in further optimization rounds, until sufficiently good solutions were obtained with $(n_L, n_T) = (5, 5)$. For this choice of shape parameters, the average cost of annealed solutions was 0.2729 ± 0.0923 , and 22 % of solutions exhibited limit cycle oscillations under constant DD and LL conditions. Fig. 1 shows simulations of the distributed delay model for the solution yielding the minimum cost ($E = 0.107$) amongst the 22 sets of autonomously oscillating solutions (the corresponding parameter values are given in Table 1). Figs. 1(a-b) compare the time traces of *LHY/CCA1* and *TOC1* mRNA concentration for the distributed delay model (dashed lines) and the original L2005A model (solid lines); these were obtained by simulating in 12h:12h LD cycles for the first 3 days, in LL for the next 3 days, and in DD for the last 3 days. It can be seen that quantitatively similar waveforms were obtained for both genes.

We note that whilst this optimal parameter set simulated the acute light-

induced activation characteristic of the original model to some degree, more pronounced acute responses were observed with other parameter sets, although these were not the best in terms of cost. The free-running period of the distributed delay model was 24.0 h (in both LL and DD), which is close to the free-running period of 24.6 h (in both LL and DD) observed in the original L2005A model. Under LL, the phase difference between *LHY/CCA1* and *TOC1* mRNA levels was 6.27 h for the distributed delay model and 5.53 h for the original model, agreeing with each other reasonably well.

In Figs. 1 (c-d), peak phases of mRNA expression are plotted as a function of the photoperiod. For *LHY/CCA1* (red circles) and *TOC1* (green triangles), the peak phase remains almost the same as photoperiod varies. These properties discernible in the original model (d) are well reproduced in the distributed delay model (c).

In Figs. 1 (e-f), bifurcation diagrams are plotted for both the distributed delay model and the original model under DD. The *LHY/CCA1* transcription rate (h_1) was used as the bifurcation parameter, normalized to its nominal value in the case of the original model and its optimized value in the case of the distributed delay formulation (in each case, a normalized parameter value \bar{h}_1 of 1 thus gives rise to the periodic oscillations of the simulated plant circadian clock under DD – these are shown as green circles in the figure). In both diagrams, it can be seen that a stable equilibrium (red points) bifurcates into a limit cycle oscillation (green points) through a super-critical Hopf bifurcation, at which a complex conjugate pair of eigenvalues of the Jacobian matrix at the equilibrium become purely imaginary (Guckenheimer, 1983). Although the bifurcation point \bar{h}_H was located more closely to the reference value $\bar{h}_1 = 1$ in the original L2005A model than the distributed delay model, their bifurcation structures are qualitatively similar.

Concerning model complexity, the distributed delay formulation has 18 parameters, which is a significant reduction from the 29 parameters of the original model. Thus, the model complexity has been efficiently reduced in the distributed delay model, whilst preserving the quantitative dynamics.

2.4. Distributed delay representation of the Kernel model

As our second example, the Kernel model (Foo et al., 2016) was utilized. From a comprehensive model (MF2015) of the genetic circuitry in the *Arabidopsis* circadian clock, essential interactions were extracted to obtain the kernel model (MF2015K). By dropping 6 components that have no feedback to the main circadian loop, the MF2015K model can be further reduced to

a set of 18 ordinary differential equations (see Appendix A.2). Introducing delay terms of the form (1) to represent protein or protein complex formation in this reduced model yields the following 10 equations:²

$$\begin{aligned}
\frac{dc_L^{(m)}}{dt} &= \frac{\theta_4}{\theta_5^2 + \theta_6 \left(\theta_{54} c_{P9}^{(d_{nP9})} + \theta_{55} c_{P7}^{(d_{nP7})} + \theta_{56} c_{P5}^{(d_{nP5})} \right)^2} + \theta_7 c_L^{(m)}, \\
\frac{dc_{P9}^{(m)}}{dt} &= \theta_{10} c_P L(t) + \theta_{11} + \frac{\theta_{12} c_L^{(d_{nL})^2}}{\theta_{13}^2 + c_L^{(d_{nL})^2}} + \theta_{14} c_{P9}^{(m)}, \\
\frac{dc_{P7}^{(m)}}{dt} &= \frac{\theta_{17}}{(\theta_{18} + c_{EC}) \left(\theta_{19}^2 + c_T^{(d_{nT})^2} \right)} + \theta_{20} c_{P7}^{(m)}, \\
\frac{dc_{P5}^{(m)}}{dt} &= \frac{\theta_{23}}{\left(\theta_{24}^2 + c_T^{(d_{nT})^2} \right) \left(\theta_{25}^2 + c_L^{(d_{nL})^2} \right)} + \theta_{26} c_{P5}^{(m)}, \\
\frac{dc_T^{(m)}}{dt} &= \frac{\theta_{29}}{(\theta_{30} + c_{EC}) \left(\theta_{31}^2 + c_L^{(d_{nL})^2} \right)} + \theta_{32} c_T^{(m)}, \\
\frac{dc_{E3}^{(m)}}{dt} &= \frac{\theta_{37}}{\theta_{38}^2 + c_L^{(d_n)^2}} + \theta_{39} c_{E3}^{(m)}, \\
\frac{dc_{E4}^{(m)}}{dt} &= \frac{\theta_{42}}{(\theta_{43} + c_{EC}) \left(\theta_{44}^2 + c_L^{(d_{nL})^2} \right)} + \theta_{45} c_{E4}^{(m)}, \\
\frac{dc_{LUX}^{(m)}}{dt} &= \frac{\theta_{48}}{(\theta_{49} + c_{EC}) \left(\theta_{50}^2 + c_L^{(d_{nL})^2} \right)} + \theta_{51} c_{LUX}^{(m)}, \\
\frac{dc_P}{dt} &= \theta_1 L(t) + \theta_2 c_P + \theta_3, \\
\frac{dc_{EC}}{dt} &= \theta_{35} c_{E3}^{(d_{nE3})} c_{E4}^{(d_{nE4})} c_{LUX}^{(d_{nLUX})} + \theta_{36} c_{EC}. \tag{5}
\end{aligned}$$

²Incorporation of the homogenous delay equations yields a further 16 equations.

Here, $c_i^{(m)}(t)$ and $c_i^{(d_{n_i})}(t)$ represent the concentrations of mRNA and the final element of the corresponding protein modification chain, respectively. The subscripts $i = L, P9, P7, P5, T, E3, E4$ and LUX denote LHY (as a proxy for $LHY/CCA1$), $PRR9, PRR7, PRR5, TOC1, ELF3$ (*EARLY FLOWERING 3*), $ELF4$ and LUX (*LUX ARRHYTHMO*), respectively. Finally, c_P and c_{EC} represent the concentrations of proteins P and EC (*EVENING COMPLEX*), respectively. The variables for the concentration levels are dimensionless with an arbitrary scale for each component. The θ_i s parametrize transcription, translation and degradation processes.

In our modelling, the shape parameters n_i were all set equal to 2 on the basis of preliminary parameter sweeps, which established that larger values were not required to obtain good fits. Values for the rate parameters $\Phi = (f_L, f_{P9}, \dots, f_{LUX})$ were optimized by minimizing the cost function defined in (4). All other parameters (*i.e.*, the θ_i s) were fixed at their nominal values in the original model, listed in Table A.2.

Using the same procedure as that performed in the previous subsection, the Sobol algorithm was used to generate 5×10^4 quasi-random points within the hypercube $\Phi \subset [0.1, 10]^8$. Among them, 100 points with the lowest cost function scores were passed to the simulated annealing routine. The average cost of the 100 annealed solutions was 0.4942 ± 0.1235 . 87 % of these solutions exhibited limit cycle oscillations under constant DD and LL conditions.

The optimised parameter values that gave the minimum cost ($E = 0.147$) amongst the 87 parameter sets yielding self-sustained oscillations are shown in Table 2. Fig. 2 plots the corresponding simulation results. Time traces for the mRNA concentrations of LHY (a), $TOC1$ (b), $PRR9$ (c), $ELF3$ (d) and LUX (e) are compared between the original Kernel model (solid lines) and its formulation with distributed delays (dashed lines). The simulations were carried out under 12h:12h LD cycles for the first 3 days, under LL for the next 3 days and under DD for the last 3 days. It can be seen that quantitatively similar waveforms were again reproduced. The free running period of the distributed delay model was 25.0 h (in both LL and DD), which is very close to the original period of 25.8 h (in both LL and DD).

In Figs. 2 (f,g), peak phases of mRNA expression are plotted as a function of the photoperiod in LD cycles. In the original Kernel model (panel (f)), clock genes attain peak phase in the following order: $PRR9$ (green triangles), LUX (magenta stars), $TOC1$ (blue squares), $ELF3$ (black diamonds) and LHY (red circles). The same ordering is preserved in the distributed delay

formulation (panel (g)). In both models, the peak phases are insensitive to the change in photoperiod. We finally note that the distributed delay model has 49 parameters, compared with the 56 parameters of the original model.

2.5. Distributed delay representation of the compact model

For our final example, we chose the compact model introduced by De Caluwé et al. (2016), which will be referred to hereafter as DC2016. The original model consists of 9 ordinary differential equations (see Appendix A.3). 8 equations describe the temporal evolution of the mRNA and protein levels of the main clock genes, grouped into four sets of lumped pairs. These pairs are represented by the following labels: CL ($CCA1$ and LHY), $P97$ ($PRR9$ and $PRR7$), $P51$ ($PRR5$ and $TOC1$) and EL ($ELF4$ and LUX). The 9th equation is for light-sensitive protein P ($PIF4$ and $PIF5$).

To obtain the reduced model considered here, two modifications were made. First, the protein levels were replaced by delay terms of the form (1). Second, in the original model (De Caluwé et al., 2016), the $P97$ variable ($PRR9$ and $PRR7$) is assumed to be activated by CL ($CCA1$ and LHY), based on the experiments reported in (Farré and Kay, 2007). More recent work (Fogelmark and Troein, 2014; Adams et al., 2015), however, has revealed that the LHY gene acts as a repressor of all other clock components, including $PRR9$ and $PRR7$. We therefore revised the compact model accordingly, replacing the activation term for the $CL \rightarrow P97$ connection with a repression term. The revised model is governed by the following differential equations:³

$$\begin{aligned} \frac{dc_{CL}^{(m)}}{dt} &= (v_1 + v_{1L}L(t)c_P) \cdot \frac{1}{1 + \left(\frac{\binom{d_n P97}{c_{P97}}}{K_1}\right)^2 + \left(\frac{\binom{d_n P51}{c_{P51}}}{K_2}\right)^2 + \left(\frac{\binom{d_n CL}{c_{CL}}}{K_0}\right)^2} \\ &\quad - (k_{1L}L(t) + k_{1D}D(t)) c_{CL}^{(m)}, \\ \frac{dc_{P97}^{(m)}}{dt} &= (v_{2L}L(t)c_P + v_{2A}) \cdot \frac{1}{1 + \left(\frac{\binom{d_n P51}{c_{P51}}}{K_4}\right)^2 + \left(\frac{\binom{d_n EL}{c_{EL}}}{K_5}\right)^2 + \left(\frac{\binom{d_n CL}{c_{CL}}}{K_3}\right)^2} \\ &\quad - k_2 c_{P97}^{(m)}, \end{aligned}$$

³Incorporation of the homogenous delay equations yields a further 8 equations.

$$\begin{aligned}
\frac{dc_{P51}^{(m)}}{dt} &= v_3 \frac{1}{1 + \left(\frac{c_{CL}^{(d_{n_{CL}})}}{K_6}\right)^2 + \left(\frac{c_{P51}^{(d_{n_{P51}})}}{K_7}\right)^2} - k_3 c_{P51}^{(m)}, \\
\frac{dc_{EL}^{(m)}}{dt} &= L(t) v_4 \frac{1}{1 + \left(\frac{c_{CL}^{(d_{n_{CL}})}}{K_8}\right)^2 + \left(\frac{c_{P51}^{(d_{n_{P51}})}}{K_9}\right)^2 + \left(\frac{c_{EL}^{(d_{n_{EL}})}}{K_{10}}\right)^2} - k_4 c_{EL}^{(m)}, \\
\frac{dc_P}{dt} &= 0.3(1 - c_P)(1 - L(t)) - c_P L(t). \tag{6}
\end{aligned}$$

As before, $c_i^{(m)}(t)$ and $c_i^{(d_n)}(t)$ represent the concentrations of mRNA and the corresponding final delayed protein element, respectively ($i = CL, P97, P51$ and EL). c_P represents the concentration of the light-activated protein P . In our modelling, the shape parameters n_i were all set to 2 on the basis of initial parameter sweeps that indicated larger values were not necessary for accurate data-fitting. The following rate parameters

$$\Phi = (f_{CL}, f_{P97}, f_{P51}, f_{EL}, K_0, K_3, K_4, K_5),$$

specifying the delays and thresholds for CL repression, were optimised by minimizing the cost function defined in Appendix B (since the original compact model has been revised here, the cost function (4) used to fit distributed delay versions of L2005A and MF2015K was not utilized in this case). The remaining parameters were fixed at the values reported in the original paper of De Caluwé et al. (2016) (*cf.* Table A.3). Sobol search and simulated annealing were combined as before to optimise Φ within the range $\Phi \subset [0.1, 10]^8$. The average cost of the annealed solutions, which all exhibited self-sustained oscillations under constant DD and LL conditions, was 3.99 ± 0.38 .

The optimised parameter values that gave the minimum cost ($E = 3.435$) for the distributed delay formulation are listed in Table 3. This cost value implies that most of the criteria implemented in the cost function detailed in Appendix B are satisfied, with the exception of a few free-running periods or phases that deviate from the desired ones by about 0.1 h. The corresponding simulation results are shown in Fig. 3, which compares the timeseries generated by the distributed delay model (dashed lines) and the original compact model (solid lines) for $CCA1/LHY$ (a), $PRR5/TOC1$ (b), $PRR9/PRR7$ (c), and $ELF4/LUX$ (d). Simulations were carried out under 12h:12h LD cycles

for the first 3 days, under LL for the next 3 days and under DD for the last 3 days. The simulation plots demonstrate that quantitatively similar waveforms were reproduced by the reduced model. The free running periods of the distributed delay model are 23.6 h (LL) and 29.0 h (DD). Compared to the free running periods of 23.5 h (LL) and 26.3 h (DD) observed in the original model, the period deviates in DD but is still very close to the original period in LL.

Figs. 3 (e,f) plot the peak phases of mRNA expression as a function of photoperiod. In both the original compact model and its distributed delay formulation, the clock genes peak in the following order: *PRR9/PRR7* (green triangles), *ELF4/LUX* (black diamonds), *PRR5/TOC1* (blue squares) and *CCA1/LHY* (red circles). In both models, the phases of the genes, especially *ELF4/LUX*, get progressively delayed as the photoperiod is increased. It should be noted that the distributed delay model has 27 parameters, a reduction from the 34 parameters of the original model.

Finally, in order to assess the consequences of our modification to the clock gene circuit, we simulated the experiment of Adams et al. (2015), which revealed that the *LHY* gene represses the expression of other clock components. The experimental protocol was reproduced as follows. First, the model was entrained to a 12h:12h LD cycle. Then, at $t = 0$, the simulation condition was changed to LL. Next, at $t = t_0$, the *LHY* gene ($i = CL$) was over-expressed by adding a constant term of 2.5 nM/h to the right-hand-side of the *CL* equation in (6), for different values of t_0 spaced at 4 h intervals over the circadian cycle.

Fig. 4 compares expression levels of the clock genes (filled back bars), which were measured 2 h after the *LHY* induction, with those of the control (white bars), under which *LHY* was not induced. The time axis indicates when the gene expression was measured (*i.e.*, $t_0 + 2$). As expected, the *LHY* gene (a,b) was clearly expressed at a higher level in both models, compared with the control. In the distributed delay model, both the *PRR5/TOC1* (d) and *PRR9/PRR7* (g) genes were repressed, agreeing quite well with the experiment (f,i). In the original compact model, on the other hand, *PRR9/PRR7* (h) was not repressed, and although *PRR5/TOC1* (e) was repressed, the temporal expression level pattern did not fit the data as well as the distributed delay formulation (d). This is due to the inherent architecture of the original compact model, in which *LHY* activates *PRR9/PRR7*. The repression from $CL \rightarrow P97$ introduced in our distributed delay model therefore improved the predictive capacity of the compact model, in addition

to reducing the number of parameters to be optimised.

3. Discussion

3.1. Distributed delays – a simple approach to reducing model complexity

In the present study, the method of distributed delays has been introduced to mathematical modelling of the plant circadian clock for the first time. The distributed delay yields a minimal model of protein translation, transport and modification in the underlying gene regulatory network, which reduces the number of parameters required to represent these intermediate processes (MacDonald, 2008; Akman et al., 2008, 2010). The key advantages this confers over more detailed biochemical network models are that: (1) the reduction in the number of model parameters controlling protein synthesis greatly lowers the computational cost of the parameter optimisation process, thereby facilitating data-fitting; and (2) the distributed delay framework provides a simplified representation of the core mechanism of the gene network, facilitating the interpretation of experimental results.

Here, by constructing distributed delay versions of three established mathematical models of the *Arabidopsis* circadian clock, and fitting these to synthetic and experimental data, the efficiency and accuracy of this modelling approach has been comprehensively demonstrated. Notably, the number of parameters was reduced from 29 to 18 in the single feedback loop model L2005A (Locke et al., 2005a), from 56 to 49 in the Kernel model MF2015K (Foo et al., 2016) and from 34 to 27 in the compact model DC2016 (De Caluwé et al., 2016). Despite this simplification, the essential properties of the original models – such as the endogenous period, the peak phases of gene expression and the dependence of peak phase on photoperiod – were accurately reproduced. Indeed, for the compact model – which exhibits photoperiod sensitivity – although our reduced formulation was optimised to data simulated in 12h:12h LD cycles only, we obtained a good match across all photoperiods, conferring a degree of validation. Furthermore, our newly introduced repression term from *LHY* to *PRR9* and *PRR7* revised the compact model (De Caluwé et al., 2016), which is now capable of reproducing the *LHY* induction experiment of (Adams et al., 2015) more faithfully.

This type of approach to reducing computational complexity could be of significant use in constructing quantitative models of more complex circadian systems, including mammalian clocks (Korenčič et al., 2012, 2014; Anantha-

subramaniam et al., 2014), the dynamics of clock output pathways (Martins et al., 2016), as well as other biochemical networks.

One disadvantage of the distributed delay models is that when the integrodifferential equation terms are converted into sets of equivalent ordinary differential equations using the linear chain trick (MacDonald, 2008), the number of resulting differential equations to be solved can be increased significantly, depending upon the values of the shape parameter n_i in each such term. For a network of N clock genes, $\sum_{i=1}^N n_i$ equations of the form (2) should be added to compute the auxiliary variables $\{c_i^{(d_1)}, c_i^{(d_2)}, \dots, c_i^{(d_{n_i})}\}$ ($i = 1, 2, \dots, N$). However, the homogenous nature of these equations means that provided $\sum_{i=1}^N n_i$ is not too large, the increased computational cost that may be incurred by numerically integrating the auxiliary equations is still offset by the significant reduction in the number of model parameters, and the subsequent acceleration in the computationally intensive task of searching over highly-parametrised cost landscapes (Akman et al., 2012; Doherty et al., 2017).

It should also be noted that in this study, all shape parameters n_i for a given model were set to the same value and hence counted as a single optimisation parameter. Assigning different n_i values to different genes would therefore increase the number of parameters to be optimised. Consequently, for clock models such as MF2015K (Foo et al., 2016), where protein production is described with only two rate constants (translation and degradation), the distributed delay approach may not significantly reduce the size of the parameter space. However, compared with such simplified representations of translation and protein modification processes, varying the shape parameters provides a more flexible framework for data-fitting by allowing a broader family of delay kernels to be explored by the optimisation process (Akman et al., 2008, 2010).

3.2. Future work

To further extend the approach presented here, the dependence of model behaviour on the shape parameters n_i should be investigated in more detail. For each gene, n_i determines the form of the distributed delay kernel $G_{f_i}^{n_i}(t)$. In particular, in the limit $n_i, f_i \rightarrow \infty$ with $n_i/f_i \rightarrow \tau_i$, $G_{f_i}^{n_i}(t) \rightarrow \delta(t - \tau_i)$ and so the distributed delay term (1) converges to the discrete delay term

$$c_i^{(d_{n_i})}(t) = c_i^{(m)}(t - \tau_i). \quad (7)$$

Consequently, optimisation algorithms have the potential to drive n_i and f_i to very large values. However, as mentioned above, to keep the system size (and hence the computational cost) tractable, it is necessary to keep the n_i s within reasonable ranges. A systematic criterion for constraining n_i and f_i is therefore desired to develop distributed delay models of optimal complexity.

Another potential extension of this approach is to cellular network modelling of the plant system. A variety of spatiotemporal dynamics – such as phase wave propagation, spiral waves and stripe waves – have been observed in plant circadian systems (Fukuda et al., 2007, 2012; Wenden et al., 2012; Gould et al., 2018). These dynamics are due to cell-to-cell interactions, which can be very strong, *e.g.*, in shoot apices (Takahashi et al., 2015). Although the coupling agents have yet to be clearly identified by experimental studies, it will be straightforward to develop a reduced cellular network model for key clock components using distributed delays, which can then be employed to explore the effect of coupling different cells together. Future work will deal with the modelling of coupled plant cells to gain an insight into potential candidates for the coupling agents, and to identify viable coupling schemes.

Acknowledgements

The authors are grateful to Mr. Mark Greenwood for stimulating discussions. The experimental data for Fig. 4 (c,f,i) were provided by Dr. Isabelle Carre. This work was partially supported by the Grant-in-Aid for Scientific Research (No. 17H06313, No. 16H05011, No. 16K00343) from the Japan Society for the Promotion of Science (JSPS), and by the Engineering and Physical Sciences Research Council (grant number EP/N017846/1). The work in the J.C.W.L. laboratory is supported by a fellowship from the Gatsby Foundation (GAT3272/GLC).

Tables

Parameter	Value
n_L	5
n_T	
f_L	8.5029 1/h
f_T	1.8297 1/h
h_1	2.7545 nM/h
h_2	2.0592 nM/h
g_1	2.5540 nM
g_2	1.8043 nM
m_1	3.5510 nM/h
m_4	0.6280 nM/h
m_7	9.4567 nM/h
k_1	8.1412 nM
k_4	0.1502 nM
k_7	5.0711 nM
p_3	6.8702 1/h
q_1	0.4013 1/h
q_2	8.2053 1/h

Table 1: Optimal parameter values for the distributed delay formulation of the single loop feedback model L2005A.

Parameter	Value
n_L	2
n_{P9}	
n_{P7}	
n_{P5}	
n_T	
n_{E3}	
n_{E4}	
n_{LUX}	
f_L	
f_{P9}	9.3862 1/h
f_{P7}	1.0027 1/h
f_{P5}	0.5900 1/h
f_T	0.5183 1/h
f_{E3}	2.5701 1/h
f_{E4}	9.5609 1/h
f_{LUX}	0.1181 1/h

Table 2: Optimal parameter values for the distributed delay formulation of the kernel model MF2015K.

Parameter	Value
n_{CL}	2
n_{P97}	
n_{P51}	
n_{EL}	
f_{CL}	2.0236 1/h
f_{P97}	0.8375 1/h
f_{51}	0.6977 1/h
f_{EL}	6.0013 1/h
K_0	8.4945 nM
K_3	1.9591 nM
K_4	0.3940 nM
K_5	4.7937 nM

Table 3: Optimal parameter values for the distributed delay formulation of the compact model DC2016.

Figures

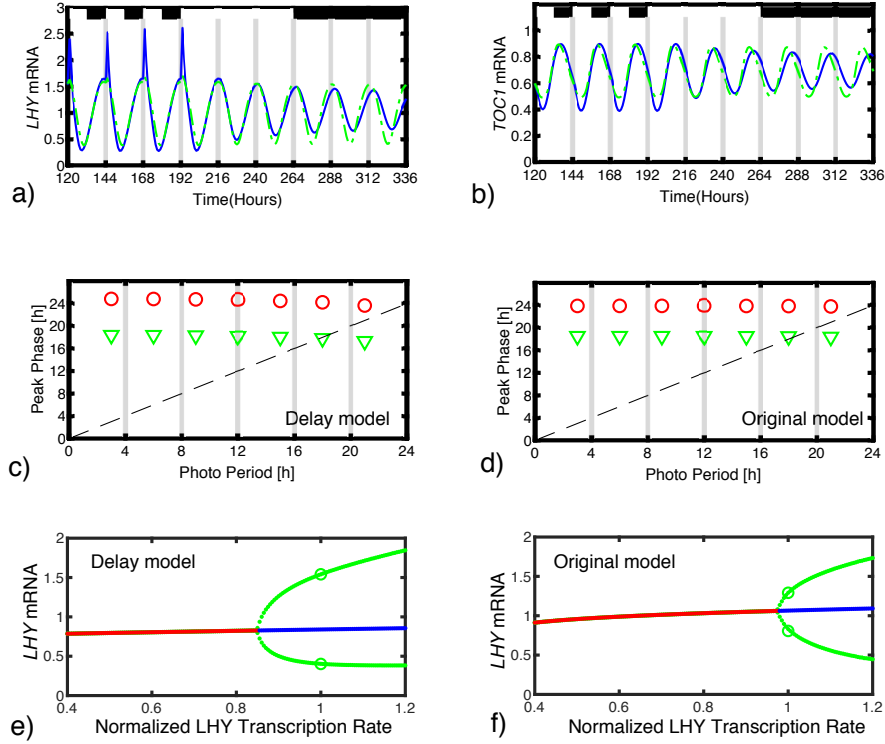


Figure 1: (a,b) Time traces of mRNA concentration $c_i^{(m)}(t)$ in the single feedback loop model L2005A (Locke et al., 2005a) (blue solid lines) and its distributed delay implementation (green dashed lines). The models were simulated under 12h:12h LD cycles for 3 days, under LL for the next 3 days and then finally under DD for 3 days. *LHY/CCA1* ($i = L$) and *TOC1* ($i = T$) genes are plotted in (a) and (b), respectively. (c,d) Peak phases of mRNA expression in LD cycles as a function of the photoperiod (*LHY/CCA1*: red circles; *TOC1*: green triangles). Results for the distributed delay model and the original model are plotted in (c) and (d), respectively. Dotted lines denote dusk. (e,f) Bifurcation diagrams for the distributed delay model (e) and the original model (f). Stable equilibria (red lines), unstable equilibria (blue lines), and limit cycles (green lines) are plotted as a function of the *LHY/CCA1* transcription rate (h_1), normalized so that a value of 1 (green circles) corresponds to limit cycle oscillations of the plant circadian clock under DD.

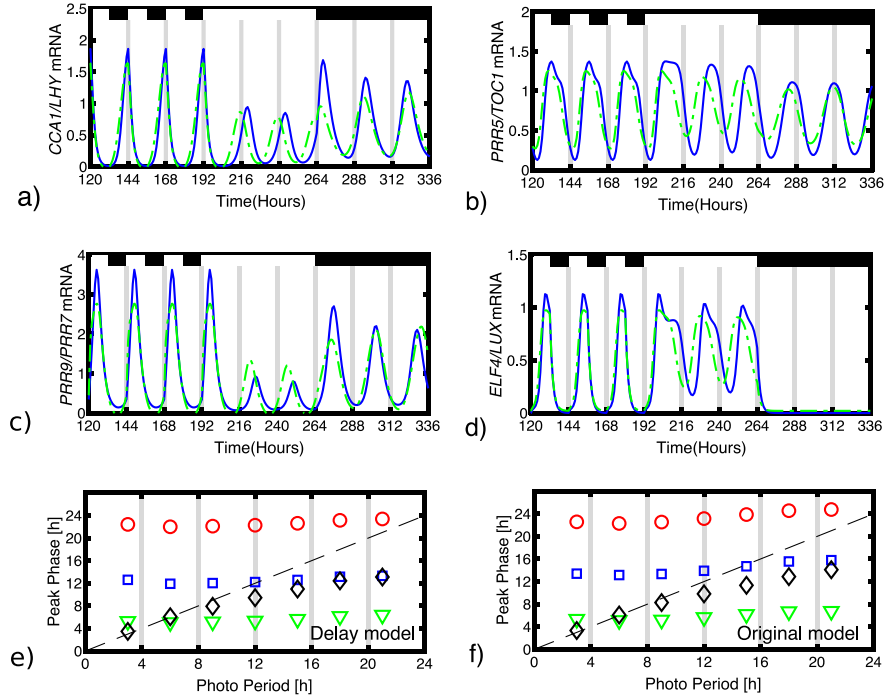


Figure 3: (a-d) Time traces of mRNA concentration $c_i^{(m)}(t)$ in the compact model DC2016 (De Caluwé et al., 2016) (blue solid lines) and its distributed delay formulation (green dashed lines). The models were simulated under 12h:12h LD cycles for 3 days, under LL for the next 3 days and then finally under DD for 3 days. $CCA1/LHY$ ($i = CL$), $PRR5/TOC1$ ($i = P51$), $PRR9/PRR7$ ($i = P97$), and $ELF4/LUX$ ($i = EL$) are plotted in panels (a) to (d), respectively. (e,f) Peak phases of mRNA expression as a function of photoperiod ($CCA1/LHY$: red circles; $PRR9/PRR7$; green triangles, $PRR5/TOC1$; blue squares, $ELF4/LUX$; black diamonds). Results for the distributed delay model and the original model are plotted in panels (e) and (f), respectively. Dotted lines denote dusk.

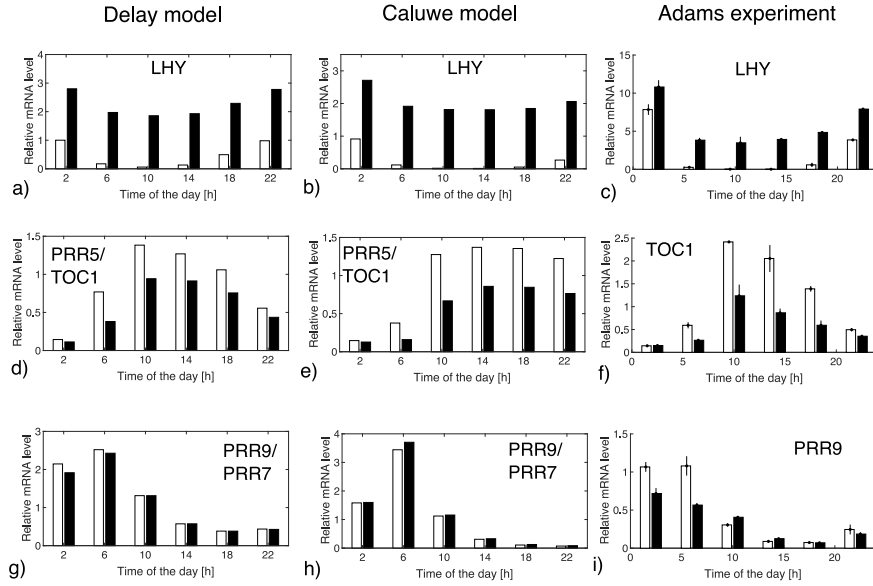


Figure 4: Simulating over-expression of LHY using the distributed delay formulation of the compact model DC2016, in which LHY represses the expression of the other clock components. After entrainment to 12h:12h LD cycles, the simulation environment is changed to LL at $t = 0$. At $t = t_0$, the *LHY* gene ($i = CL$) was over-expressed, for the following values of t_0 : 0 h, 4 h, 8 h, 12 h, 16 h, 20 h. In each case, mRNA expression levels (filled back bars) for *LHY* (a–c), *PRR5/TOC1* (d–f), and *PRR9/PRR7* (g–i) were compared 2 h post-perturbation with those of the control (white bars), under which *LHY* was not overexpressed. The times shown indicate when gene expression was measured (*i.e.* $t_0 + 2$). Panels (a,d,g) plot simulations generated by the distributed delay model, (b,e,h) plot the corresponding simulations generated by the original model (De Caluwé et al., 2016) and panels (c,f,i) show the experimental results of (Adams et al., 2015).

Appendix A. Plant clock models

Appendix A.1. The single feedback loop model – L2005A

The single feedback loop model (Locke et al., 2005a) consists of the *LHY* and *TOC1* genes, acting as repressor and activator, respectively. The dynamics is described by the following seven ordinary differential equations:

$$\begin{aligned}
 \frac{dc_L^{(m)}}{dt} &= q_1 c_P^{(c)} L(t) + \frac{h_1 c_T^{(n)a}}{g_1^a + c_T^{(n)a}} - \frac{m_1 c_L^{(m)}}{k_1 + c_L^{(m)}}, \\
 \frac{dc_L^{(c)}}{dt} &= p_1 c_L^{(m)} - r_1 c_L^{(c)} + r_2 c_L^{(n)} - \frac{m_2 c_L^{(c)}}{k_2 + c_L^{(c)}}, \\
 \frac{dc_L^{(n)}}{dt} &= r_1 c_L^{(c)} - r_2 c_L^{(n)} - \frac{m_3 c_L^{(n)}}{k_3 + c_L^{(n)}}, \\
 \frac{dc_T^{(m)}}{dt} &= \frac{h_2 g_2^b}{g_2^b + c_L^{(n)b}} - \frac{m_4 c_T^{(m)}}{k_4 + c_T^{(m)}}, \\
 \frac{dc_T^{(c)}}{dt} &= p_2 c_T^{(m)} - r_3 c_T^{(c)} + r_4 c_T^{(n)} - \frac{m_5 c_T^{(c)}}{k_5 + c_T^{(c)}}, \\
 \frac{dc_T^{(n)}}{dt} &= r_3 c_T^{(c)} - r_4 c_T^{(n)} - \frac{m_6 c_T^{(n)}}{k_6 + c_T^{(n)}}, \\
 \frac{dc_P^{(c)}}{dt} &= (1 - L(t))p_3 - \frac{m_7 c_P^{(c)}}{k_7 + c_P^{(c)}} - q_2 L(t)c_P^{(c)}. \tag{A.1}
 \end{aligned}$$

In the above, $c_i^{(j)}(t)$ represents the cellular concentration of the i th gene/protein, where $i = L$ and T correspond to the *LHY* and *TOC1* genes, respectively, $i = P$ denotes light-sensitive protein P and the labels $j = m, c, n$ denote mRNA and protein in the cytoplasm and nucleus, respectively. Note that *CCA1* and *LHY* are considered as one component (labelled *LHY*) (Locke et al., 2005a). $L(t)$ represents the input light signal ($L(t) = 0$: light off; $L(t) = 1$: light on). The rate constants denote transcription (h_l, g_l), degradation (m_l, k_l), translation (p_l), transport between the nucleus and cytoplasm (r_l) and light input (q_l). (a, b) are the Hill coefficients. The model has 29 parameters in all. The optimised parameter values reported in (Locke et al., 2005a) that were used to generate target timeseries data for this study are listed in Table A.1.

Parameter	Value
a	1
b	2
h_1	7.5038 nM/h
h_2	0.6801 nM/h
g_1	1.4992 nM
g_2	3.0412 nM
m_1	10.0982 nM/h
m_2	1.9685 nM/h
m_3	3.7511 nM/h
m_4	2.3422 nM/h
m_5	7.2482 nM/h
m_6	1.8981 nM/h
m_7	1.20 nM/h
k_1	3.8045 nM
k_2	5.3087 nM
k_3	4.1946 nM
k_4	2.5356 nM
k_5	1.442 nM
k_6	4.8600 nM
k_7	1.20 nM
p_1	2.1994 1/h
p_2	9.444 1/h
p_3	0.5 1/h
r_1	0.2817 1/h
r_2	0.7676 1/h
r_3	0.4364 1/h
r_4	7.3021 1/h
q_1	4.5703 1/h
q_2	1.0 1/h

Table A.1: Nominal parameter values for the single feedback loop model L2005A (Locke et al., 2005a).

Appendix A.2. Kernel model – MF2015K

Here, we consider the system obtained from the kernel model (Foo et al., 2016) of the *Arabidopsis* circadian clock by removing the components with

no feedback to the main circadian loop (namely the *RVE8* and *GI* genes, their proteins and COP1 and ZTL proteins). This results in a set of 18 ordinary differential equations for the genes *LHY* (as a proxy for *LHY/CCA1*), *PRR9*, *PRR7*, *PRR5*, *TOC1*, *ELF3*, *ELF4* and *LUX*, together with the light-sensitive protein *P* and the evening complex *EC*:

$$\begin{aligned} \frac{dc_L^{(m)}}{dt} &= \frac{\theta_4}{\theta_5^2 + \theta_6 \left(\theta_{54}c_{P9}^{(p)} + \theta_{55}c_{P7}^{(p)} + \theta_{56}c_{P5}^{(p)} \right)^2} + \theta_7c_L^{(m)}, \\ \frac{dc_L^{(p)}}{dt} &= \theta_8c_L^{(m)} + \theta_9c_L^{(p)}, \\ \frac{dc_{P9}^{(m)}}{dt} &= \theta_{10}c_P^{(p)}L(t) + \theta_{11} + \frac{\theta_{12}c_L^{(p)2}}{\theta_{13}^2 + c_L^{(p)2}} + \theta_{14}c_{P9}^{(m)}, \\ \frac{dc_{P9}^{(p)}}{dt} &= \theta_{15}c_{P9}^{(m)} + \theta_{16}c_{P9}^{(p)}, \\ \frac{dc_{P7}^{(m)}}{dt} &= \frac{\theta_{17}}{\left(\theta_{18} + c_{EC}^{(p)} \right) \left(\theta_{19}^2 + c_T^{(p)2} \right)} + \theta_{20}c_{P7}^{(m)}, \\ \frac{dc_{P7}^{(p)}}{dt} &= \theta_{21}c_{P7}^{(m)} + \theta_{22}c_{P7}^{(p)}, \\ \frac{dc_{P5}^{(m)}}{dt} &= \frac{\theta_{23}}{\left(\theta_{24}^2 + c_T^{(p)2} \right) \left(\theta_{25}^2 + c_L^{(p)2} \right)} + \theta_{26}c_{P5}^{(m)}, \\ \frac{dc_{P5}^{(p)}}{dt} &= \theta_{27}c_{P5}^{(m)} + \theta_{28}c_{P5}^{(p)}, \\ \frac{dc_T^{(m)}}{dt} &= \frac{\theta_{29}}{\left(\theta_{30} + c_{EC}^{(p)} \right) \left(\theta_{31}^2 + c_L^{(p)2} \right)} + \theta_{32}c_T^{(m)}, \\ \frac{dc_T^{(p)}}{dt} &= \theta_{33}c_T^{(m)} + \theta_{34}c_T^{(p)}, \\ \frac{dc_{E3}^{(m)}}{dt} &= \frac{\theta_{37}}{\theta_{38}^2 + c_L^{(p)2}} + \theta_{39}c_{E3}^{(m)}, \\ \frac{dc_{E3}^{(p)}}{dt} &= \theta_{40}c_{E3}^{(m)} + \theta_{41}c_{E3}^{(p)}, \end{aligned}$$

$$\begin{aligned}
\frac{dc_{E4}^{(m)}}{dt} &= \frac{\theta_{42}}{\left(\theta_{43} + c_{EC}^{(p)}\right) \left(\theta_{44}^2 + c_L^{(p)2}\right)} + \theta_{45}c_{E4}^{(m)}, \\
\frac{dc_{E4}^{(p)}}{dt} &= \theta_{46}c_{E4}^{(m)} + \theta_{47}c_{E4}^{(p)}, \\
\frac{dc_{LUX}^{(m)}}{dt} &= \frac{\theta_{48}}{\left(\theta_{49} + c_{EC}^{(p)}\right) \left(\theta_{50}^2 + c_L^{(p)2}\right)} + \theta_{51}c_{LUX}^{(m)}, \\
\frac{dc_{LUX}^{(p)}}{dt} &= \theta_{52}c_{LUX}^{(m)} + \theta_{53}c_{LUX}^{(p)}, \\
\frac{dc_P^{(p)}}{dt} &= \theta_1L(t) + \theta_2c_P^{(p)} + \theta_3, \\
\frac{dc_{EC}^{(p)}}{dt} &= \theta_{35}c_{E3}^{(p)}c_{E4}^{(p)}c_{LUX}^{(p)} + \theta_{36}c_{EC}^{(p)}.
\end{aligned} \tag{A.2}$$

Here, $c_i^{(m)}$ (resp. $c_i^{(p)}$) represents the i th mRNA (resp. protein or protein complex) concentration, and has a dimensionless value with arbitrary scale for each component; thus, values within the same component – but not between different components – can be meaningfully compared. The subscripts $i = P, L, P9, P7, P5, T, EC, E3, E4$ and LUX denote protein P , LHY (as a proxy for $LHY/CCA1$), $PRR9$, $PRR7$, $PRR5$, $TOC1$, EC , $ELF3$, $ELF4$ and LUX , respectively. $L(t)$ represents the input light signal ($L(t) = 0$: light off; $L(t) = 1$: light on). The model has a total of 56 parameters. The optimised parameter values presented in (Foo et al., 2016) that were used to generate synthetic data for this study are listed in Table A.2.

Parameter	Value	Parameter	Value
θ_1	-0.43301	θ_2	-0.4949
θ_3	0.4358	θ_4	1.9601
θ_5	0.7412	θ_6	0.3821
θ_7	-2.7804	θ_8	1.9601
θ_9	0.7412	θ_{10}	0.3821
θ_{11}	-2.7804	θ_{12}	0.83604
θ_{13}	-0.64719	θ_{14}	0.73369
θ_{15}	6.4976×10^{-5}	θ_{16}	0.75976
θ_{17}	1.3352	θ_{18}	-0.59706
θ_{19}	0.50077	θ_{20}	-0.22337
θ_{21}	0.0099479	θ_{22}	0.034706
θ_{23}	0.87726	θ_{24}	-0.23934
θ_{25}	0.23744	θ_{26}	-0.18571
θ_{27}	0.026728	θ_{28}	0.13201
θ_{29}	0.031588	θ_{30}	-40.9795
θ_{31}	4.0753	θ_{32}	-2.5872
θ_{33}	0.01501	θ_{34}	0.036926
θ_{35}	0.011307	θ_{36}	-12.2192
θ_{37}	1.4074	θ_{38}	-1.5159
θ_{39}	0.32848	θ_{40}	-0.14877
θ_{41}	0.014607	θ_{42}	0.23392
θ_{43}	-0.27265	θ_{44}	14.0123
θ_{45}	-17.3569	θ_{46}	0.014384
θ_{47}	0.032117	θ_{48}	0.051737
θ_{49}	-11.007	θ_{50}	0.67365
θ_{51}	-0.39718	θ_{52}	0.028907
θ_{53}	0.01727	θ_{54}	1.275
θ_{55}	1.5983	θ_{56}	-0.44539

Table A.2: Nominal parameter values for the kernel model MF2015K (Foo et al., 2016).

Appendix A.3. Compact model - DC2016

The compact model of De Caluwé et al. (2016) is based on 4 pairs of clock genes, *CL* (*CCA1* and *LHY*), *P97* (*PRR9* and *PRR7*), *P51* (*PRR5*

and *TOC1*) and *EL* (*ELF4* and *LUX*), together with the light-sensitive protein *P* (*PIF4* and *PIF5*) (De Caluwé et al., 2016). The system dynamics are modelled by the following 9 ordinary differential equations:

$$\begin{aligned}
\frac{dc_{CL}^{(m)}}{dt} &= \left(v_1 + v_{1L}L(t)c_P^{(p)} \right) \cdot \frac{1}{1 + \left(\frac{c_{P97}^{(p)}}{K_1} \right)^2 + \left(\frac{c_{P51}^{(p)}}{K_2} \right)^2} \\
&\quad - (k_{1L}L(t) + k_{1D}D(t)) c_{CL}^{(m)}, \\
\frac{dc_{CL}^{(p)}}{dt} &= (p_1 + p_{1L}L(t)) c_{CL}^{(m)} - d_1 c_{CL}^{(p)}, \\
\frac{dc_{P97}^{(m)}}{dt} &= \left(v_{2L}L(t)c_P^{(p)} + v_{2A} + v_{2B} + \frac{c_{CL}^{(p)2}}{K_3^2 + c_{CL}^{(p)2}} \right) \cdot \frac{1}{1 + \left(\frac{c_{P51}^{(p)}}{K_4} \right)^2 + \left(\frac{c_{EL}^{(p)}}{K_5} \right)^2} \\
&\quad - k_2 c_{P97}^{(m)}, \\
\frac{dc_{P97}^{(p)}}{dt} &= p_2 c_{P97}^{(m)} - (d_{2D}D(t) + d_{2L}L(t)) c_{P97}^{(p)}, \\
\frac{dc_{P51}^{(m)}}{dt} &= v_3 \frac{1}{1 + \left(\frac{c_{CL}^{(p)}}{K_6} \right)^2 + \left(\frac{c_{P51}^{(p)}}{K_7} \right)^2} - k_3 c_{P51}^{(m)}, \\
\frac{dc_{P51}^{(p)}}{dt} &= p_3 c_{P51}^{(m)} - (d_{3D}D(t) + d_{3L}L(t)) c_{P51}^{(p)}, \\
\frac{dc_{EL}^{(m)}}{dt} &= L(t) v_4 \frac{1}{1 + \left(\frac{c_{CL}^{(p)}}{K_8} \right)^2 + \left(\frac{c_{P51}^{(p)}}{K_9} \right)^2 + \left(\frac{c_{EL}^{(p)}}{K_{10}} \right)^2} - k_4 c_{EL}^{(m)}, \\
\frac{dc_{EL}^{(p)}}{dt} &= p_4 c_{EL}^{(m)} - (d_{4D}D(t) + d_{4L}L(t)) c_{EL}^{(p)}, \\
\frac{dc_P^{(p)}}{dt} &= 0.3 \left(1 - c_P^{(p)} \right) D(t) - c_P^{(p)} L(t). \tag{A.3}
\end{aligned}$$

In the above, $c_i^{(m)}$ (resp. $c_i^{(p)}$) represents the concentration of the i th mRNA (resp. protein or protein complex), for $i = CL, P97, P51, EL$ and P . $L(t)$ represents the input light signal ($L(t) = 0$: light off; $L(t) = 1$: light on) and $D(t)$ denotes a corresponding input dark signal, given by $D(t) = 1 - L(t)$. The model contains 34 parameters, the values of which were obtained through

automated optimisation in (De Caluwé et al., 2016). These values are listed in Table A.3. They were used here to generate synthetic data for constructing the distributed delay version of the modified DC2016 clock circuit in which CL is a universal repressor.

Appendix B. Optimising the distributed delay representation of the compact model

This section provides details of the protocol used to optimise the parameters of equations (6). To compute the cost function, the model was simulated under 12h:12h LD conditions for a total of 25 days and then released into constant light (LL) for 15 days, followed by constant dark (DD) for 15 days. In each light condition, the first 10 days were discarded as transient dynamics.

To ensure detectable rhythmicity under LL/DD conditions, all variables were required to have a minimum value of 0.1, as well as a minimum difference of 10 % between their minimum and maximum values. Any solution that did not meet these criteria was considered to be arrhythmic. Arrhythmic solutions were penalised with an arbitrarily large score. For rhythmic solutions, the free-running period was calculated using the chi-square periodogram (Sokolove and Bushell, 1978) of the LHY gene at a significance level of 1%. A score of 0 was assigned to a solution having a free-running period between 24 and 25 h under LL and between 25 and 28 h under DD. Solutions with free-running periods outside these ranges were allocated the following scores: $(\tau_{LL} - 24.5)^2 / (0.1 \cdot 24.5)^2$ (τ_{LL} : free-running period under LL) and $(\tau_{DD} - 26.5)^2 / (0.1 \cdot 26.5)^2$ (τ_{DD} : free-running period under DD).

For simulations under 12h:12h LD cycles, solutions that were not entrained to the LD cycles were penalised with an arbitrarily large score. Entrained solutions were given a score of 0 for each gene that attained peak expression within ± 1 h of the expected ZT, which were as follows: $CCA1/LHY$ (ZT24), $PRR5/TOC1$ (ZT12), $PRR9/PRR7$ (ZT6) and $ELF4/LUX$ (ZT9) (De Caluwé et al., 2016). Expression peaks lying outside these intervals were scored as follows: $(ZT_{CL} - 24)^2 / (0.1 \cdot 24.5)^2$, $(ZT_{P51} - 12)^2 / (0.1 \cdot 24.5)^2$, $(ZT_{P97} - 6)^2 / (0.1 \cdot 24.5)^2$ and $(ZT_{EL} - 9)^2 / (0.1 \cdot 24.5)^2$, where ZT_i denotes the target Zeitgeber time of the i th gene’s peak expression).

Parameter	Value
v_1	4.58 nM/h
v_{1L}	3.0 nM/h
v_{2A}	1.27 nM/h
v_{2B}	1.48 nM/h
v_{2L}	5.0 nM/h
v_3	1.0 nM/h
v_4	1.47 nM/h
k_{1L}	0.53 1/h
k_{1D}	0.21 1/h
k_2	0.35 1/h
k_3	0.56 1/h
k_4	0.57 1/h
p_1	0.76 1/h
p_{1L}	0.42 1/h
p_2	1.01 1/h
p_3	0.64 1/h
p_4	1.01 1/h
d_1	0.68 1/h
d_{2D}	0.5 1/h
d_{2L}	0.29 1/h
d_{3D}	0.48 1/h
d_{3L}	0.78 1/h
d_{4D}	1.21 1/h
d_{4L}	0.38 1/h
K_1	0.16 nM
K_2	1.18 nM
K_3	0.24 nM
K_4	0.23 nM
K_5	0.3 nM
K_6	0.46 nM
K_7	2.0 nM
K_8	0.36 nM
K_9	1.9 nM
K_{10}	1.9 nM

Table A.3: Nominal parameter values for the compact model DC2016 (De Caluwé et al., 2016).

References

- Adams, S., Manfield, I., Stockley, P., Carré, I. A., 2015. Revised morning loops of the Arabidopsis circadian clock based on analyses of direct regulatory interactions. *PLoS One* 10 (12), e0143943.
- Akman, O. E., Locke, J. C. W., Tang, S., Carré, I., Millar, A. J., Rand, D. A., 2008. Isoform switching facilitates period control in the *Neurospora crassa* circadian clock. *Mol. Syst. Biol.* 4, 64.
- Akman, O. E., Rand, D. A., Brown, P. E., Millar, A. J., 2010. Robustness from flexibility in the fungal circadian clock. *BMC Syst. Biol.* 4 (1), 88.
- Akman, O. E., Watterson, S., Parton, A., Binns, N., Millar, A. J., Ghazal, P., 2012. Digital clocks: simple Boolean models can quantitatively describe circadian systems. *J. Roy. Soc. Interface* 9 (74), 2365–2382.
- Ananthasubramaniam, B., Herzog, E. D., Herzog, H., 2014. Timing of neuropeptide coupling determines synchrony and entrainment in the mammalian circadian clock. *PLoS Comput. Biol.* 10 (4), e1003565.
- Atamian, H. S., Creux, N. M., Brown, E. A., Garner, A. G., Blackman, B. K., Harmer, S. L., 2016. Circadian regulation of sunflower heliotropism, floral orientation, and pollinator visits. *Science* 353 (6299), 587–590.
- Bujdoso, N., Davis, S. J., 2013. Mathematical modelling of an oscillating gene circuit to unravel the circadian clock network of *Arabidopsis thaliana*. *Front. Plant Sci.* 4, 3.
- De Caluwé, J., Xiao, Q., Hermans, C., Verbruggen, N., Leloup, J.-C., Gonze, D., 2016. A compact model for the complex plant circadian clock. *Front. Plant Sci.* 7, 74.
- Dodd, A. N., Salathia, N., Hall, A., Kévei, E., Tóth, R., Nagy, F., Hibberd, J. M., Millar, A. J., Webb, A. A., 2005. Plant circadian clocks increase photosynthesis, growth, survival, and competitive advantage. *Science* 309 (5734), 630–633.
- Doherty, K., Alyahya, K., Akman, O. E., Fieldsend, J. F., 2017. Optimisation and landscape analysis of computational biology models: A case study. In: *Proc. GECCO 2017*, 1644–1651.

- Dunlap, J. C., Loros, J. J., DeCoursey, P. J., 2004. *Chronobiology: Biological Timekeeping*. Sinauer Assoc., Sunderland, Massachusetts, USA.
- Farré, E. M., Kay, S. A., 2007. PRR7 protein levels are regulated by light and the circadian clock in *Arabidopsis*. *Plant J.* 52 (3), 548–560.
- Flis, A., Fernández, A. P., Zielinski, T., Mengin, V., Sulpice, R., Stratford, K., Hume, A., Pokhilko, A., Southern, M. M., Seaton, D. D., et al., 2015. Defining the robust behaviour of the plant clock gene circuit with absolute RNA timeseries and open infrastructure. *Open Biol.* 5 (10), 150042.
- Fogelmark, K., Troein, C., 2014. Rethinking transcriptional activation in the *Arabidopsis* circadian clock. *PLoS Comput. Biol.* 10 (7), e1003705.
- Foo, M., Somers, D. E., Kim, P.-J., 2016. Kernel Architecture of the Genetic Circuitry of the *Arabidopsis* Circadian System. *PLoS Comput. Biol.* 12 (2), e1004748.
- Fukuda, H., Nakamichi, N., Hisatsune, M., Murase, H., Mizuno, T., 2007. Synchronization of plant circadian oscillators with a phase delay effect of the vein network. *Phys. Rev. Lett.* 99 (9), 098102.
- Fukuda, H., Ukai, K., Oyama, T., 2012. Self-arrangement of cellular circadian rhythms through phase-resetting in plant roots. *Phys. Rev. E* 86 (4), 041917.
- Goodspeed, D., Chehab, E. W., Min-Venditti, A., Braam, J., Covington, M. F., 2012. *Arabidopsis* synchronizes jasmonate-mediated defense with insect circadian behavior. *Proc. Natl. Acad. Sci. USA* 109 (12), 4674–4677.
- Gould, P., Domijan, M., Greenwood, M., Tokuda, I., Rees, H., Kozma-Bognar, L., Hall, A., Locke, J., 2018. Coordination of robust single cell rhythms in the *Arabidopsis* circadian clock via spatial waves of gene expression. *Elife* 7, e31700.
- Guckenheimer, J., Holmes, P., 1983. *Nonlinear Oscillations, Dynamical Systems and Bifurcations of Vector Fields*. Springer Verlag, New York.
- Hale, J. K., 1977. Retarded functional differential equations: basic theory. In: *Theory of Functional Differential Equations*. Springer, pp. 36–56.

- Hale, J. K., Lunel, S. M. V., 2013. Introduction to Functional Differential Equations. Vol. 99. Springer Science & Business Media.
- Huang, W., Pérez-García, P., Pokhilko, A., Millar, A. J., Antoshechkin, I., Riechmann, J. L., Mas, P., 2012. Mapping the core of the Arabidopsis circadian clock defines the network structure of the oscillator. *Science* 336 (6077), 75–79.
- Korenčič, A., Bordyugov, G., Lehmann, R., Rozman, D., Herzel, H., et al., 2014. Timing of circadian genes in mammalian tissues. *Sci. Rep.* 4, 5782.
- Korenčič, A., Bordyugov, G., Rozman, D., Goličnik, M., Herzel, H., et al., 2012. The interplay of cis-regulatory elements rules circadian rhythms in mouse liver. *PLoS ONE* 7 (11), e46835.
- Kurosawa, G., Mochizuki, A., Iwasa, Y., 2002. Comparative study of circadian clock models, in search of processes promoting oscillation. *J. Theor. Biol.* 216 (2), 193–208.
- Lema, M. A., Golombek, D. A., Echave, J., 2000. Delay model of the circadian pacemaker. *J. Theor. Biol.* 204 (4), 565–573.
- Locke, J., Millar, A., Turner, M., 2005a. Modelling genetic networks with noisy and varied experimental data: the circadian clock in *Arabidopsis thaliana*. *J. Theor. Biol.* 234 (3), 383–393.
- Locke, J. C. W., Kozma-Bognar, L., Gould, P. D., Fehér, B., Kevei, E., Nagy, F., Turner, M. S., Hall, A., Millar, A. J., 2006. Experimental validation of a predicted feedback loop in the multi-oscillator clock of *Arabidopsis thaliana*. *Mol. Syst. Biol.* 2, 59.
- Locke, J. C. W., Southern, M. M., Kozma-Bognar, L., Hibberd, V., Brown, P. E., Turner, M. S., Millar, A. J., 2005b. Extension of a genetic network model by iterative experimentation and mathematical analysis. *Mol. Syst. Biol.* 1, 2005.0013.
- MacDonald, N., 2008. Biological delay systems: Linear stability theory. Vol. 9. Cambridge University Press, Cambridge.
- Mackey, M. C., Glass, L., 1977. Oscillation and chaos in physiological control systems. *Science* 197 (4300), 287–289.

- Martins, B., Das, A., Antunes, L., Locke, J., 2016. Frequency doubling in the cyanobacterial circadian clock. *Mol. Syst. Biol.* 12 (12).
- McClung, C. R., 2006. Plant circadian rhythms. *Plant Cell* 18 (4), 792–803.
- Mittler, J. E., Sulzer, B., Neumann, A. U., Perelson, A. S., 1998. Influence of delayed viral production on viral dynamics in HIV-1 infected patients. *Math. Biosci.* 152 (2), 143–163.
- olde Scheper, T., Klinkenberg, D., Pennartz, C., Van Pelt, J., 1999. A mathematical model for the intracellular circadian rhythm generator. *J. Neurosci.* 19 (1), 40–47.
- Pokhilko, A., Fernández, A. P., Edwards, K. D., Southern, M. M., Halliday, K. J., Millar, A. J., 2012. The clock gene circuit in *Arabidopsis* includes a repressilator with additional feedback loops. *Mol. Syst. Biol.* 8 (1), 574.
- Pokhilko, A., Hodge, S. K., Stratford, K., Knox, K., Edwards, K. D., Thomson, A. W., Mizuno, T., Millar, A. J., 2010. Data assimilation constrains new connections and components in a complex, eukaryotic circadian clock model. *Mol. Syst. Biol.* 6 (1), 416.
- Smolen, P., Baxter, D. A., Byrne, J. H., 2001. modelling circadian oscillations with interlocking positive and negative feedback loops. *J. Neurosci.* 21 (17), 6644–6656.
- Smolen, P., Baxter, D. A., Byrne, J. H., 2002. A reduced model clarifies the role of feedback loops and time delays in the *Drosophila* circadian oscillator. *Biophys. J.* 83 (5), 2349–2359.
- Sokolove, P. G., Bushell, W. N., May 1978. The chi square periodogram: Its utility for analysis of circadian rhythms. *J. Theor. Biol.* 72 (1), 131–160.
- Sriram, K., Gopinathan, M., 2004. A two variable delay model for the circadian rhythm of *Neurospora crassa*. *J. Theor. Biol.* 231 (1), 23–38.
- Takahashi, N., Hirata, Y., Aihara, K., Mas, P., 2015. A hierarchical multi-oscillator network orchestrates the *Arabidopsis* circadian system. *Cell* 163 (1), 148–159.

- Troein, C., Locke, J. C., Turner, M. S., Millar, A. J., 2009. Weather and seasons together demand complex biological clocks. *Curr. Biol.* 19 (22), 1961–4.
- Wenden, B., Toner, D. L., Hodge, S. K., Grima, R., Millar, A. J., 2012. Spontaneous spatiotemporal waves of gene expression from biological clocks in the leaf. *Proc. Natl. Acad. Sci. USA* 109 (17), 6757–6762.