

Stochastic properties of the plant circadian clock

Maria Luisa Guerriero¹, Alexandra Pokhilko², Aurora Piñas Fernández²,
Karen J. Halliday^{1,2}, Andrew J. Millar^{1,2} and Jane Hillston^{1,3}

¹ Centre for Systems Biology at Edinburgh, University of Edinburgh,
C.H. Waddington Building, King's Buildings Campus, Mayfield Road, EH9 3JD, Edinburgh, UK

² School of Biological Sciences, University of Edinburgh,
King's Buildings Campus, Mayfield Road, EH9 3JH, Edinburgh, UK

³ School of Informatics, University of Edinburgh,
Informatics Forum, 10 Crichton Street, EH8 9AB, Edinburgh, UK

Abstract

Circadian clocks are gene regulatory networks whose role is to help the organisms to cope with variations in environmental conditions such as the day/night cycle. In this work we explored the effects of molecular noise in single cells on the behaviour of the circadian clock in the plant model species *Arabidopsis thaliana*. The computational modelling language Bio-PEPA enabled us to give a stochastic interpretation to an existing deterministic model of the clock, and to easily compare results obtained via stochastic simulation and via numerical solution of the deterministic model.

First, the introduction of stochasticity in the model allowed us to estimate the unknown size of the system. Moreover, stochasticity improved the description of the available experimental data in several light conditions: noise-induced fluctuations yield a faster entrainment of the plant clock under certain photoperiods and are able to explain the experimentally observed dampening of the oscillations in plants under constant light conditions. The model predicts that the desynchronisation between noisy oscillations in single cells contributes to the observed damped oscillations at the level of the cell population. Analysis of the phase, period and amplitude distributions under various light conditions demonstrated robust entrainment of the plant clock to light/dark cycles, which closely matched available experimental data.

Keywords: circadian clock, *Arabidopsis thaliana*, discrete stochastic model, Bio-PEPA process algebra, oscillatory systems

1 Introduction and motivations

Circadian clocks are gene regulatory networks present in most organisms that help them to adapt to the 24-hour day/night cycle. They are composed of a small number of genes involved in interlocking transcriptional feedback loops and enable eukaryotic organisms to anticipate the daily changes in environmental conditions such as the duration of the light period. Circadian clocks have two essential features: (i) in the absence of external stimuli (e.g. constant light) the amounts of the involved mRNAs and proteins oscillate rhythmically with a period of approximately 24 hours (*circadian*); (ii) in the presence of external stimuli (e.g. light on/off) the oscillations *entrain* to the external stimuli, adjusting their rhythm to it. Stochasticity is known to play a major role in the behaviour of circadian clocks due to the small copy numbers in which most of the involved molecules are usually present, and a number

of studies on the effect of stochasticity and noise-induced fluctuations on circadian clocks in various organisms exist. For instance, a discrete stochastic model of the circadian clock in mammals has been presented by Forger and Peskin [10], while Gonze *et al.* developed and compared a deterministic and a discrete stochastic model of the *Drosophila* clock [16, 15, 17]; in [1] the effect of discreteness and stochasticity on the robustness of the clock in *Neurospora* was studied, and [2] presents a discrete stochastic model of the clock in the *Ostreococcus* alga, which is considered to be a naturally simplified version of the clock in higher plants such as *Arabidopsis*. In all these studies, important differences between discrete stochastic models and their continuous deterministic approximation were found, and in most cases stochasticity was shown to have a positive effect on the robustness of the free running (i.e. not entrained) clock.

Arabidopsis thaliana is the most studied model plant and it has been widely used for studying the plant circadian clock. Several deterministic models of the circadian clock in *Arabidopsis thaliana* have been developed previously [24, 23, 33, 26], and they are based on multiple data in wild type and mutant plants.

Here we explore the impact that intrinsic molecular noise has on the plant clock behaviour, by adding stochasticity to the behaviour of the recent deterministic model presented in [26]. The size of the system (i.e. the copy numbers of the molecules involved in the system) is unknown, but it is supposed to be small by analogy to other species and, consequently, the molecular noise is believed to have a significant impact on the behaviour of the system. This makes a discrete stochastic interpretation of the model essential in order to capture the noise-induced behaviour of the clock. We investigate the effect of the stochastic fluctuations in a single cell on the clock behaviour, the differences between continuous deterministic and the discrete stochastic interpretations of the model, and their relation to experimental data. We also apply measures commonly used for time-series analysis in signal processing in order to quantify the relevant features of the dynamic behaviour of the clock, such as the dampening of oscillations and their distribution of phase, period and amplitude.

2 The clock model

The principal scheme of the *Arabidopsis* clock introduced in [26] is shown in Figure 1. The key dawn genes *LATE ELONGATED HYPOCOTYL (LHY)* and *CIRCADIAN CLOCK ASSOCIATED 1 (CCA1)* are described in the model by a single variable, *LHY/CCA1*. They activate the expression of their own inhibitors, PSEUDO-RESPONSE REGULATOR 9 (PRR9), PRR7 and the night inhibitor NI (PRR5) inside the morning loop. LHY and CCA1 proteins also inhibit the expression of the evening genes *GIGANTEA (GI)*, *TIMING OF CAB EXPRESSION 1 (TOC1)* and its hypothetical activator gene *Y* in the morning, and thus provide a delay in their expression. GI promotes the inhibition of TOC1 protein function by positively regulating the F-box protein ZEITLUPE (ZTL). Evening genes form an evening loop of the clock, which feeds back to the morning loop by activating *LHY/CCA1* expression at the end of the night. The evening loop also inhibits *PRR9* expression at night and thus reduces night inhibition of *LHY/CCA1*, increasing the robustness of oscillations. The light activates LHY/CCA1, PRR9, PRR7, NI, GI and Y at the transcriptional and/or post-transcriptional levels, providing the day length sensing by both morning and evening loops.

Our model is based on the deterministic model introduced in [26], which formalises the scheme described above.

In order to model the light and the experimental settings, we applied a two-stage light function to describe an initial entrainment stage followed by an observation stage. Indeed, in laboratory experiments, plants are usually kept in alternating light/dark cycles for a few days, and then transferred to the desired experimental condition (e.g. constant light, dark, or different photoperiods). The light on/off switch can be modelled either as a discrete step function which switches instantly from 0 to 1 at dawn and from 1 to 0 at dusk, or as a

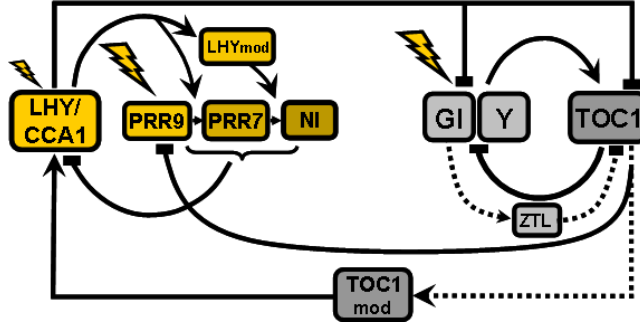


Figure 1: The scheme of the *Arabidopsis* circadian clock introduced in [26]. Morning and evening loops of the clock are shown in yellow and grey, respectively. Light inputs to gene transcription are shown by flashes.

smooth function in which the switch is rapid but not instantaneous. Though the discrete step function matches more closely the experimental conditions, the smooth function is closer to real life conditions and, moreover, has the advantage of being integrable. In our experiments we have not noticed significant differences in the model response to these two light functions. The results shown in the following have been obtained using the smooth function introduced in [27]. The formal definition of the light function is given, together with the full model, in the Electronic Supplementary Material (Appendix C).

We coded the clock model in a computational language called Bio-PEPA [5], which is a stochastic process algebra developed for modelling biochemical systems. Process algebras are formal languages commonly used in computer science areas such as performance modelling and concurrency theory. In addition to simulation, formal languages enable modellers to *verify* systems: for example, formal properties of models can be verified to uncover causal relations between events, reachability of specific states, or equivalences between different systems. Formal methods and process algebras have been used to model and analyse biochemical systems in several studies in recent years (see, for instance, [4, 18] for reviews on the approach). The Bio-PEPA language has been developed to represent biochemical systems in a formal and compositional way. It is equipped with both a discrete stochastic semantics defined in terms of a continuous-time Markov chain (CTMC), and a continuous deterministic semantics defined in terms of ordinary differential equations (ODEs). This allows us to analyse our model of the clock both by running stochastic simulations (using standard methods such as Gillespie’s [14] and Gibson-Bruck [13] algorithms) and by solving its (equivalent) underlying set of ODEs. The fact that the discrete stochastic and continuous deterministic models are formally derived from the same Bio-PEPA description ensures consistency between them, thus allowing us to compare the stochastic and the deterministic behaviour of the clock. A short introduction to Bio-PEPA is given in the Electronic Supplementary Material (Appendix B); for a more detailed presentation of the language, the reader is referred to [5].

In order to have a correct stochastic interpretation of the model, its variables must represent discrete molecule counts rather than continuous concentrations. Hence, the model variables and kinetic laws described in [26] are rescaled by a parameter Ω which represents the size of the system (see Section 3.1 for a discussion of Ω). Consequently, the set of ODEs underlying the Bio-PEPA model is the same set of ODEs of the model presented in [26] except for the scaling factor Ω . In our model the species amounts represent molecule counts, where species that reached peak values close to 1 in the relative units of [26] now peak close to Ω . The stochastic semantics of our Bio-PEPA model follows the standard stochastic chemical kinetics approach, where reaction rates represent propensity functions which define the probability of occurrence of each reaction from a given state, governed by the chemical

master equation (CME), and where stochastic simulation methods (such as the Gillespie’s method [14]) are commonly adopted for model analysis. It is worth pointing out that our model is not a continuous stochastic model described by stochastic differential equations (in which a Gaussian noise term is added to ordinary differential equations to represent stochasticity), but rather a discrete stochastic model in which stochasticity emerges by the randomness of the behaviour defined by the reaction propensity functions.

For model analysis, we used a tool called the Bio-PEPA Eclipse Plug-in [3]. Among the various simulators and solvers available within this tool, we used the Gibson-Bruck stochastic simulator [13] and a Dormand-Prince adaptive step-size deterministic solver [6].

The Bio-PEPA model of the clock is described in the Electronic Supplementary Material (Appendix C), and the full model is available as a Bio-PEPA file (Appendix F) and as an equivalent SBML model exported from Bio-PEPA (Appendix G). The SBML model is also available from the Biomodels database [21] and the Plant Systems-biology Modelling database [30].

3 Results

We discuss here some of the simulation results we have obtained, comparing them with experimental results reported elsewhere (for example in [24, 23, 19, 28, 26]). In the following we will use the notations LL and DD for constant light and constant dark, respectively, and LD $h_l:h_d$ for alternating cycles of h_l hours of light and h_d hours of dark. The initial conditions are the same for all the simulation results reported below, both for the deterministic and the stochastic simulations: namely the initial values are taken from the limit cycle reached by the deterministic model in LD 12:12, as in [26] (rescaled to obtain molecule counts, as described in detail in the following section), and dawn is assumed to be at $t = 0$.

We will use the classic notions of phase, period and amplitude to characterise the oscillatory behaviour of the system; their most commonly adopted definitions are the following.

- **Peak:** highest value obtained by the variable in one oscillation.
- **Trough:** lowest value obtained by the variable in one oscillation.
- **Phase:** time of the day at which peak occurs, relative to dawn.
- **Period:** peak-to-peak time difference between two cycles.
- **Amplitude:** difference between peak and trough values.

Further details on these notions and on the method we used to compute them for time-series data obtained from stochastic simulation are in the Electronic Supplementary Material (Appendix D).

3.1 The system size and its effect on molecular noise

The continuous approximation underlying ODE models of biochemical systems is valid when the copy numbers of the involved molecules are large. As for many genetic networks, this assumption is generally not valid for circadian clocks: their system size (i.e. the average copy numbers of the involved molecular components) is believed to be small and, therefore, stochastic noise generally has a major effect on the system’s behaviour (see, for instance, [15, 1, 2]).

The system size of the clock network in *Arabidopsis* is unknown. The variables of the original deterministic model are given in relative units and are normalised to peak at 1. In order to correctly quantify stochastic fluctuations the species amounts must represent numbers of molecules per cell. Hence, we introduced the characteristic size of the system as an additional parameter Ω which is used to scale the species amounts and kinetic laws

accordingly (see Appendix C in the Electronic Supplementary Material). Note that Ω is not the maximum number of molecules of any or all species. Rather, Ω scales the variables of the ODEs such that a concentration of 1 in the ODEs becomes a molecule count close to Ω .

To compare the individual variations in the copy numbers of the clock components, we analysed the data underlying relative qPCR measurements, which demonstrated similar peak expression levels of the clock genes, with at most a ten-fold range among the transcripts: this justifies our use of a single scaling factor (shown in Supplementary Table 1 (Appendix A)). The average copy number of the clock components at the peaks was estimated to be around several hundreds of molecules per cell (see the Electronic Supplementary Material (Appendix A) and Supplementary Figure 1 for details).

Quantifying the fluctuations due to stochastic variations provides another method to estimate the unknown system size. Therefore, we used stochastic simulation to verify the above estimations: by observing the effect of the parameter Ω on the behaviour of the system (such as the ability of protein and mRNA amounts to entrain to the light stimulus) we could identify the value of Ω for which the behaviour is closest to the experimental data. Figure 2 reports a comparison of the behaviour of the clock for $\Omega = 10$ and $\Omega = 100$ (which represent 10 and 100 molecules, respectively). These results show that, in the entrained clock (LD 12:12, Figure 2(a–d)), whilst for $\Omega = 10$ the protein and mRNA amounts exhibit poor entrainment in a single simulation (i.e. representing a single cell, Figure 2(a)), for $\Omega = 100$, instead, the entrainment is good (Figure 2(c)); moreover, the mean behaviour for $\Omega = 100$ (Figure 2(d)) is closer to the experimental data (Figure 2(e)) compared to the one for $\Omega = 10$ (Figure 2(b)): in particular, the asymmetry in the peak visible for $\Omega = 10$ is not observed in the experimental data, and the standard deviation for $\Omega = 10$ is very high. It is worth noting that experimental techniques introduce additional extrinsic noise into the data, which is not considered in our stochastic simulations. In particular, qPCR measurements provide more precise information about the shape of profiles compared to luciferase bioluminescence traces, because the qPCR measures native RNA directly, whereas LUC luminescence is an indirect measure of transcription [9]. However, individual qPCR assays are more variable than LUC image analysis (Figure 2(e,f)). Our simulations in the free running clock (LL, Figure 2(g–j)) also show that a very small system size ($\Omega = 10$, results in low amplitude, fast-damping oscillations (Figure 2(h)), due to the high noise, while $\Omega = 100$ (Figure 2(j)) gives a better match with the experimental data (Figure 2(k)). Similar results (not shown) have been obtained for all the other considered photoperiods. Intermediate values for Ω such as 50 (results not shown) also resulted in worse agreement to the experimental results compared to $\Omega = 100$. On the other hand, higher values for Ω such as 1000 (results not shown) lead to a further decrease in noise compared to $\Omega = 100$ that makes the stochastic effects become almost insignificant and results in worse agreement to the experimental data in constant light compared to $\Omega = 100$. The conclusion we can draw from these results is that the size of the real system must be in the order of a few hundreds of molecules per cell. Further confidence in the correctness of these model estimations of the system size is provided by the fact that they are in the same order of magnitude of the estimations, based on the experimental data, which are discussed above and in the Electronic Supplementary Material (Appendix A). All the results reported in the rest of the paper are obtained using the estimated value $\Omega = 100$ as the system size.

3.2 Individual versus mean behaviour

Each individual stochastic simulation run is one possible evolution of the model and hence, given that the model represents the behaviour of a cell, it can be seen as the behaviour of one single cell. The experimental data suggest that individual oscillators in the plant cells are essentially independent over timescales of several days [29]. One report quantifies the coupling across the surface of detached leaves, and finds that it is very weak, with most potential significance over long timescales [11]. The weak coupling between cells allows us

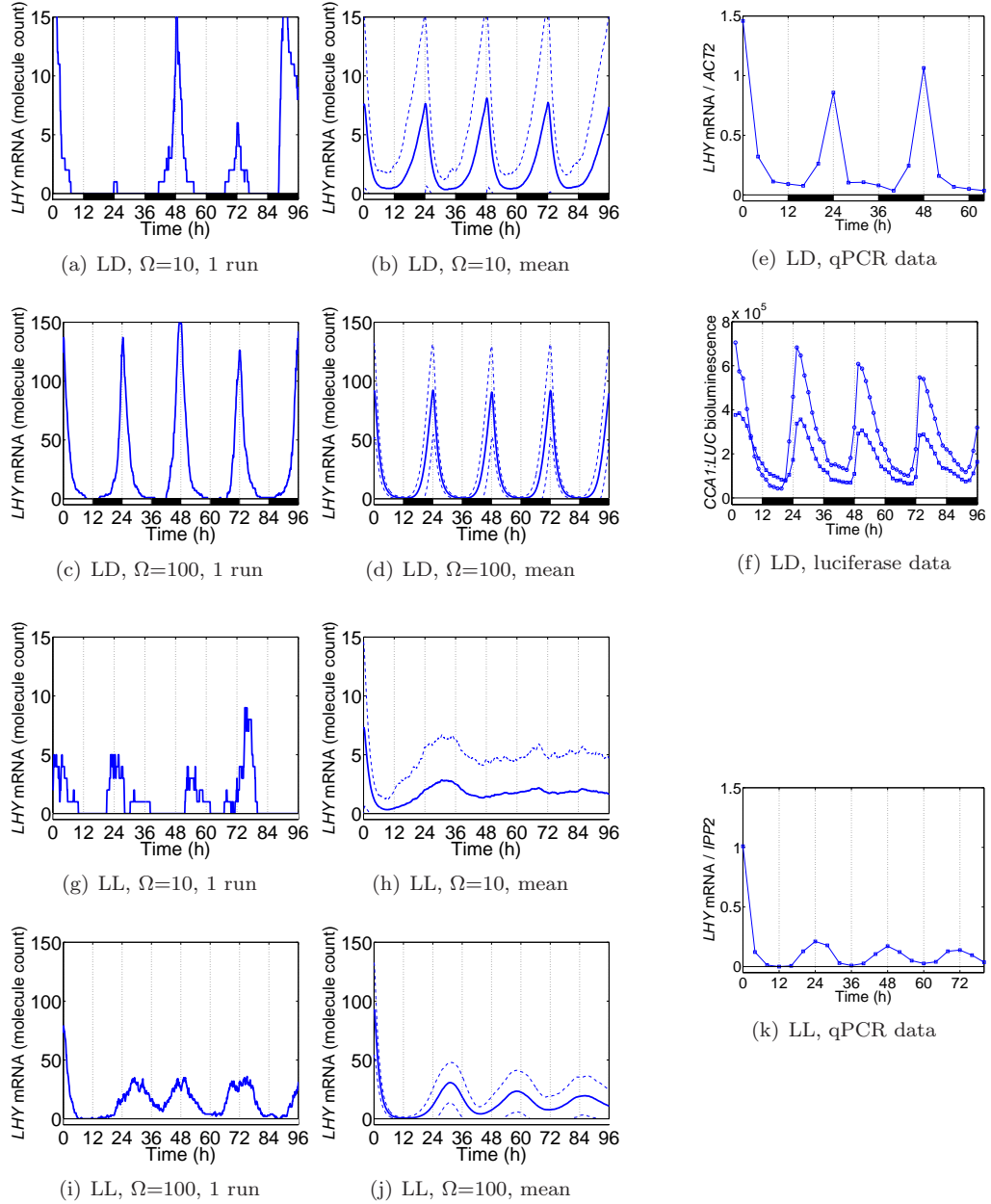


Figure 2: Effect of the system size on behaviour in LD 12:12 (a–f) and LL (g–k). Panels (a–d) and (g–j) report the amount of LHY mRNA obtained from stochastic simulation. We compare two hypothesised system sizes of around 10 and 100 peak molecule copy numbers, represented by $\Omega = 10$ and $\Omega = 100$, respectively. We report the amount obtained in random individual runs, and also the average value μ (thick line) and the standard deviation ω (thin lines, $\mu \pm \omega$) obtained from 1000 runs. Time $t = 0$ in panels (a–d) is the time of dawn of day 30 after entrainment to light cycles, while in panels (g–j) is the time of transition into constant light. The experimental qPCR measurement of LHY expression is shown in panel (e) for LD [19] and in panel (k) for LL [8]; panel (f) reports the non-normalised bioluminescence of CCA1:LUC measured as described in [7] for two individual seedlings in LD. As the units in our simulations are molecule counts, the vertical scales are different for different values of Ω . For comparison, the qPCR data must be scaled by the same factor. As discussed in the main text, the scale of the luminescence data cannot be directly compared with the one of qPCR data and simulations, as they are an indirect measure of mRNA.

to consider the mean behaviour of a stochastic model (obtained by averaging over multiple simulation runs) and the deterministic behaviour (which also, implicitly, is a mean behaviour) as representative of the mean behaviour of a uniform population of independent cells, at least over several days. This is the typical timescale for the data available in plants, and the main focus of our work: the suggested intercellular coupling is too weak to have much effect on this timescale. We also simulate longer-term behaviour under the same simplifying assumption. If the results of [11] are representative of cells in the intact plant, then intercellular coupling cannot prevent desynchronisation but may limit its extent.

Generally, experimental data represent a mean behaviour of a population of cells (standard currently available experimental tools report average measures over 10 thousands of cells taken from whole seedlings or even multiple plants). However, learning how single cells behave is essential in order to understand how a whole organism behaves and, therefore, simulation can be a powerful tool, enabling us to hypothesise the behaviour of a single cell which would be impossible to observe experimentally. On the other hand, experimental data can be used for model validation by comparing them to the mean simulated behaviour.

In the following we consider constant light and light/dark conditions, and compare the deterministic behaviour with the stochastic one. The qualitative behaviour observed in the deterministic model for both the entrained (LD 12:12) and free running clocks (LL and DD) is a perfectly periodic permanent oscillation (see [26]). The differences in these three settings lie in the period, phase and amplitude of oscillations; for instance, the period is 24 hours in LD 12:12, around 24.5 hours in LL, and around 27 hours in DD.

The behaviour observed in a single evolution of the stochastic model in LD 12:12 is similar to the regular oscillatory behaviour of the deterministic model [26], with the addition of noise-induced stochastic fluctuations (e.g. Figure 2(c)): the noise causes the oscillations to be less regular in terms of amplitude, but the light entrainment forces the oscillations to synchronise with the light and, consequently, only allows for a very small variation in period.

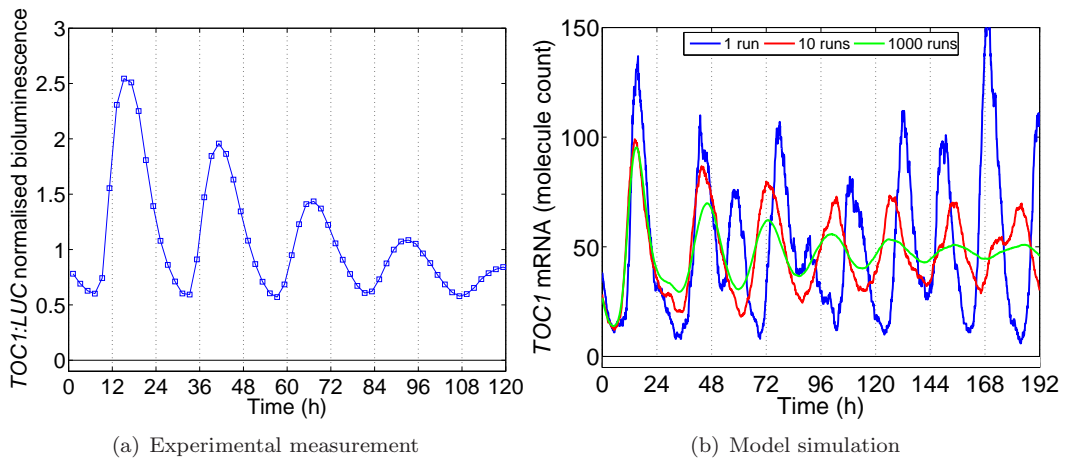


Figure 3: Stochastic behaviour and experimental data in LL. Panel (a) reports the experimentally measured amount of TOC1 mRNA (TOC1:LUC); bioluminescence was normalised to the mean level and averaged over 20 7-days-old plants as described in [23]. Panel (b) illustrates the most likely explanation for the experimentally observed damping of oscillations: the persistent oscillations of single stochastic simulation runs are lost when observing the system at the population level because of their asynchrony. Time $t = 0$ is the time of transition into constant light. See also Figure 6 and Supplementary Figure 2 (Appendix E).

Contrary to the entrained system, the effect of discreteness and stochasticity on the behaviour of the clock is clearly visible in the free running clock: while the deterministic

behaviour is a persistent oscillation, the mean stochastic behaviour is a damped oscillation, matching well with the experimental population data (Figure 3 shows the LL system).

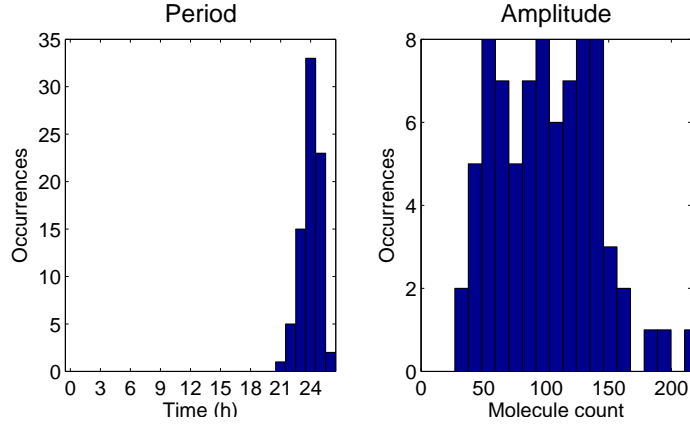
Our stochastic simulation results show that the damping of oscillations in constant light conditions is most likely the result of averaging over multiple noisy oscillators in different cells: despite the stable oscillations in the single run, the mean stochastic behaviour damps fast (Figure 3(b)) because of the lack of synchronisation in the oscillations in different runs.

Additionally to asynchrony, there are other potential explanations for the experimentally observed damping. First, the depletion of the luciferin substrate could contribute to the damping by reducing the signal progressively over time. Indeed, the data suggest that a decaying exponential with a time constant > 3 days might reflect this substrate depletion, as recently observed in algal cultures [31]. The observed faster damping of the amplitude of oscillations suggests the existence of another major mechanism of damping unrelated to experimental variation (Supplementary Figure 2 (Appendix E)). The oscillations at the level of the whole organism might also subside because of the damping of oscillations in each individual cell. Though we cannot completely rule out this hypothesis, the recent deterministic model of the plant clock, which was based on multiple experimental data, suggests that oscillations in single cells are stable [26]. The damping of the oscillations in a population of persistent cellular clocks was also observed in fibroblast cultures and in peripheral explants from mammalian tissues [25, 22]. Oscillation damping could also be reproduced by a population of deterministic oscillators with different periods [20]. However, because of the low molecule numbers involved in this (and other) biological clocks, the presence of stochastic effects is certain, with or without additional period variation. Thus we conclude that noise-induced asynchrony among different cells is the most likely mechanism of the population-level damping of oscillations in constant light conditions.

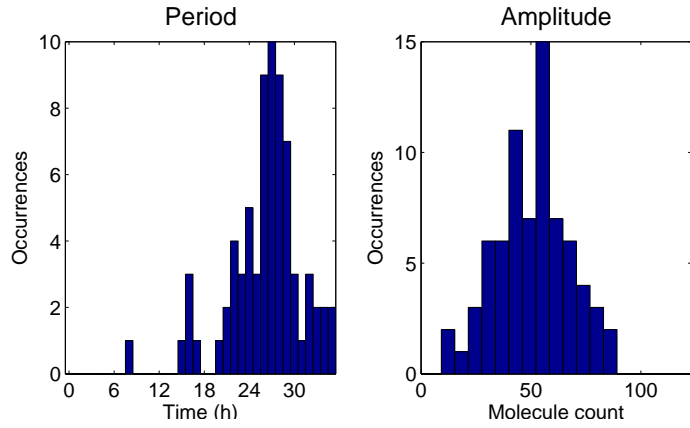
Additional insight on the model behaviour and on its variability is obtained by computing statistic measures such as distribution of phase, period and amplitude of oscillations. We compute these distributions over individual time-series both for the deterministic and the stochastic model; for the latter we also compute time-dependant distributions at a specific time over a number of different simulation runs. Details on the computation of these distributions can be found in the Electronic Supplementary Material (Appendix D).

Figure 4(a) shows the distribution of period and amplitude of the LHY mRNA level in a single 80-day-long stochastic simulation run of the LD 12:12 system: this confirms that the period is distributed quite tightly around 24 hours, while the amplitude is more variable. Comparing these distributions with the ones for the free running clock in LL and DD (Figure 4(b) and Supplementary Figure 5 (Appendix E)), we can observe how the light entrainment makes the period more stable (the distribution is less wide), whereas the variation in amplitude is smaller in LL than in LD 12:12. This agrees with experimental observation [24, 28]. The additional noise in the amplitude in LD 12:12 is probably related to the noise brought by acute activation of LHY, PRR9 and GI expression by light.

The distributions of period and amplitude shown in Figure 4 illustrate the variability of the system in an individual run, i.e. how regular is the behaviour over time. Figure 5 and Supplementary Figure 6 (Appendix E), instead, illustrate the variability over multiple runs, i.e. how similarly different runs behave: they report the time-dependent distribution of phase, period and amplitude of the oscillations over 1000 independent simulation runs at specific times (on the second and on the 80th day of observation phase) for each of the three light conditions. Figure 5(a) confirms that in LD 12:12 the distribution of period is more narrow than the one of amplitude, shows that the entrainment is quite quick (i.e. the difference between the distributions at day 2 and those at day 80 are not significant), and that the period of LHY mRNA oscillations is always very close to 24 hours and the phase is near dawn. In the free running systems, instead, the variability of period length results over time in desynchronisation of runs so that at day 80 the distribution of phase spans the entire cycle (Figure 5(b) and Supplementary Figure 6 (Appendix E)). The fact that, in the free running clock, the period and amplitude distributions are quite similar at early and later times while



(a) LD 12:12



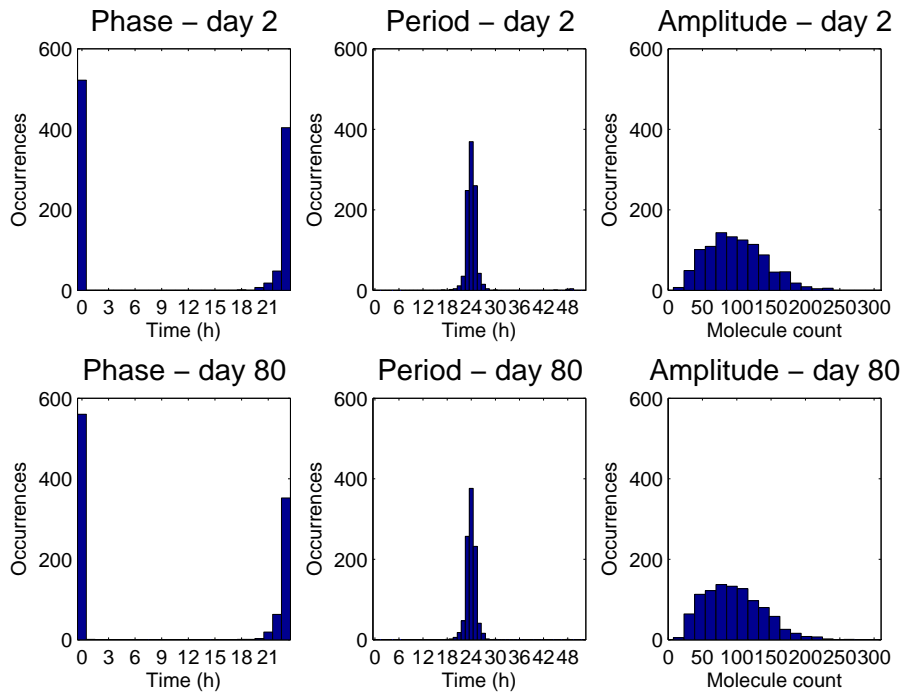
(b) LL

Figure 4: Distribution of period and amplitude of LHY mRNA oscillations over 80 days in a single stochastic run: (a) LD 12:12 and (b) LL. In order to get rid of transient effects, the first 20 simulation days are discarded in both LD and LL, so the plots refer to values obtained between day 20 and day 100.

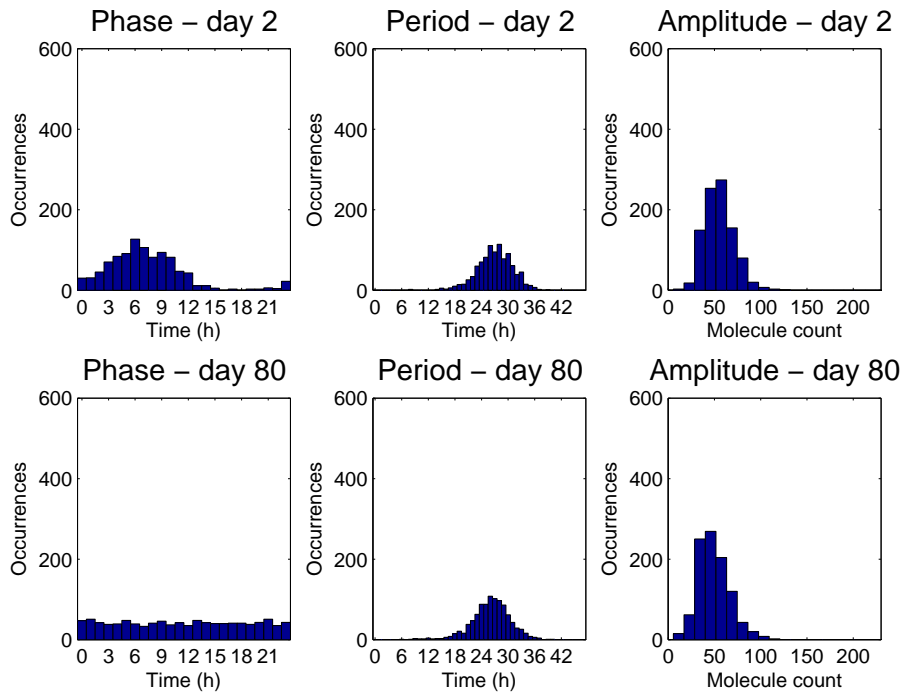
the phase distributions are wider at later times confirms the visual observation that the oscillations in each individual run are rather regular but they are not synchronised (i.e. the peak occurs at different times of the day), thus causing the dampening of oscillations observed under multiple runs and in experimental population data (Figure 3 and [26]). Supplementary Figure 3 (Appendix E) shows the distribution of period and amplitude of the LHY mRNA level over 80 days in the deterministic solution of the model: in the free running LL system the period is slightly longer (24.5 hours instead of 24) and the amplitude is smaller.

Summing up the comparison between the distributions of phase, period and amplitude in ODEs and stochastic simulations, (i) in the entrained system they agree (although the stochastic simulation results show quite a lot of variation in amplitude) and the distributions at early times are similar to the distributions at later times (i.e. there is little or no transient effect); (ii) in the free running system they show substantial differences: the ODEs show stable oscillations with a shift in phase caused by the period longer than 24 hours, the single stochastic simulation also shows persistent (noisy) oscillations, but the mean stochastic behaviour dampens quickly due to the variation in phase in different runs.

The lack of synchronisation between different cells is further illustrated in Figure 6(b). The five random simulation runs plotted clearly show how the oscillations quickly become



(a) LD 12:12



(b) LL

Figure 5: Time-dependent distribution of phase, period and amplitude of LHY mRNA oscillations in 1000 simulation runs: (a) LD 12:12 and (b) LL. In both LD and LL we assume that at day 0 the model is in an entrained state. The uniform phase distribution after 80 days in LL reflects the assumed absence of coupling among cells.

out of phase in LL; this makes the probability distribution of values so spread that the average over multiple runs dampens fast. This demonstrates why we are unable to catch the persistent oscillatory behaviour of the cellular system from the population data, neither the experimental ones (Figure 3(a)) nor those obtained by averaging multiple stochastic runs (Figures 3(b),6(b)). In contrast, the individual oscillations of the entrained clock are synchronised, although with variable amplitude. The distribution of values is quite tight around the mean (Figure 6(a)), so the population data (Figures 2(d),2(e),2(f)) resemble the single cell data (Figure 2(c)). Figure 6(a) also shows that in LD the distribution of LHY mRNA levels is broader during the rising phase than during the falling phase. Our model suggests an explanation for this asymmetry. Because different runs are not perfectly synchronised, the rising phase of LHY mRNA during the night can begin at slightly different times in different runs, hence causing some variability of the levels at any specified time in the rising phase. The faster light-dependent degradation of LHY mRNA, however, causes LHY mRNA levels to quickly decrease at dawn in all runs. The regulated mRNA degradation derives directly from experimental results [32, 7]. Consequently, we observe less variability in mRNA levels at a particular time in the falling phase, and hence better synchronisation among runs. A secondary effect is that the LHY mRNA peak is lower in runs in which the rising phase started later. This explains both the asymmetry in the broadness of the distribution of LHY mRNA level in LD and also the higher variability in peak height compared to LL, in which the rising phase is not abruptly interrupted by dawn.

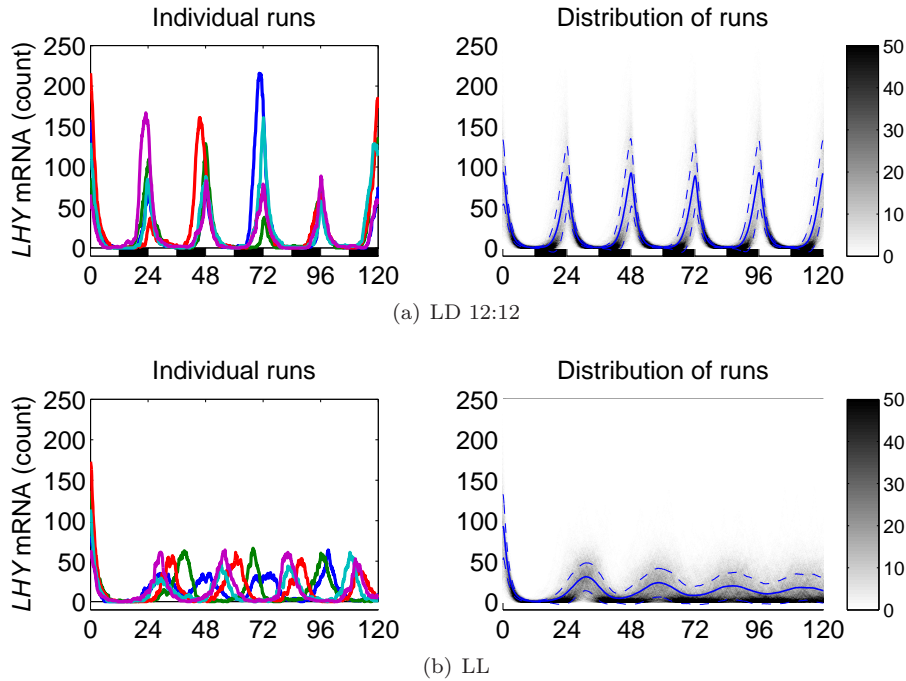
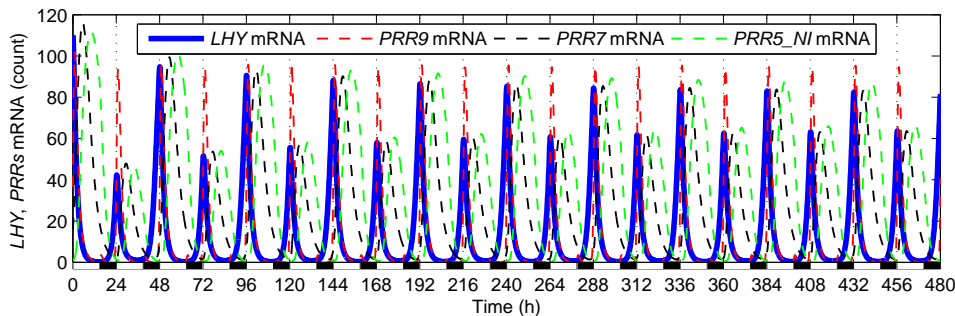


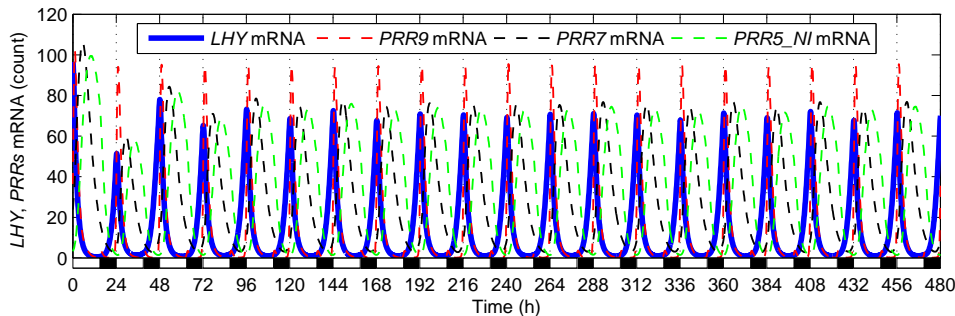
Figure 6: Variation in LHY mRNA time-series observed in independent stochastic simulations: (a) LD 12:12 and (b) LL. For each light condition five independent simulation runs (left) and the probability distribution of LHY mRNA level from 1000 runs (right) are reported. In the heatmaps representing the probability distributions, a darker colour corresponds to a higher probability for LHY mRNA to be at that level, while blue lines show the average value μ (thick line) and the standard deviation ω (dashed thin lines, $\mu \pm \omega$). Similar changes in distribution with time were obtained for TOC1 protein.

3.3 Moderate noise in a multicellular environment accelerates light entrainment

The mean stochastic behaviour gives a good description of the experimental data and, particularly in certain experiments, exhibits a faster entrainment to the light when compared to the deterministic behaviour of the system. Figure 7 shows, for instance, the behaviour of the model after the transition from short days (9h light) to long days (15h light). In the deterministic model (Figure 7(a)) the amplitude of LHY mRNA peak value exhibits a transient *doubling of period*, that is an alternation of a higher and a lower peak, which is still observed after 50 days (Figure 7(a) shows the first 20 days). The stochastic model does not show this effect and instead predicts that LHY mRNA is quickly entrained to a regular value. A similar behaviour is observed for other proteins and mRNAs (results not shown).



(a) Deterministic behaviour



(b) Mean stochastic behaviour

Figure 7: Behaviour of the system immediately following the transition from short days (LD 9:15) into long days (LD 15:9). Comparison of deterministic and mean (over 1000 runs) stochastic behaviour. The deterministic model shows a transient “doubling of period” for LHY mRNA which is not observed experimentally. The simulations consisted in 40 days of LD 9:15 followed by the transition into LD 15:9 (time $t = 0$ in the plots is the time of transition into LD 15:9).

The doubling of period in the current deterministic model is a consequence of superposition of the mutual regulation of LHY and PRRs genes and strong acute light induction of PRR wave. This results in an alternation of higher and lower amplitudes of LHY: a higher LHY amplitude results in high PRR7/5 expression, which in turn reduces the expression of LHY the following morning, while strong acute light induction of PRRs cuts LHY amplitude at a lower level immediately after dawn on the next day. Such a strong regulation greatly improves the entrainment and robustness of the clock [26]. However it can cause some transient doubling of period in the deterministic model. The stochastic simulations explain why

the doubling of period was never observed experimentally: the noise, together with the averaging over multiple runs, reduces the effect of variations in amplitude which cause doubling of period in the deterministic solution of the model, consequently improving the speed of entrainment.

3.4 Stochasticity improves the entrainment to “skeleton” artificial light patterns

To investigate the effects of the dawn and dusk light separately, some of the experiments were carried out in special entrainment conditions where the 12 hours of light have been interrupted by 6 hours of dark. This creates a second dawn in the evening and, since LHY is responsive to light, it has the effect of introducing a secondary peak in addition to the morning peak (Figure 8(d)).

The evening peak is not reproduced by the deterministic model (Figure 8(a)), but is instead clearly present in the mean stochastic behaviour (Figure 8(b)). The better match to the data obtained using the stochastic model is related to the low LHY peaks during the first light pulse, which sometimes occur in single runs. The first (dawn) light pulse occasionally results in a lower than normal LHY peak (see Figure 8(c) at 72 h). This causes lower PRR7/5 expression while TOC1 trough levels are elevated. This in turn increases the LHY response to the second (dusk) light pulse (see Figure 8(c) at 81 h). Thus, the addition of noise-induced behaviour improves the description of experimental data on the “skeleton” photoperiods too.

3.5 Autocorrelation of time-series data: a quantification of phase distribution

Most of the above considerations have been drawn by visual comparison of time-series data obtained from simulations with time-series data obtained from laboratory experiments. Using this approach qualitative features such as “oscillations dampen” or “oscillations are persistent” can be easily observed, but it is hard to provide a precise quantitative measure of the stochastic effects. As discussed in the Electronic Supplementary Material (Appendix D), when considering noisy data such as experimental data or those coming from stochastic simulation, performing data analysis (e.g. measuring phase, period and amplitude of oscillations) can be hard. Time-series analysis techniques originating in signal processing can be used in order to compute summary measures of the quality of the circadian clock related to the regularity of the oscillations, such as the distribution of phase.

The autocorrelation function of a time-series measures the similarity of the time-series with a shifted version of itself, and is commonly used in signal processing to detect periodicity in noisy signals. For perfectly periodic signals, the autocorrelation function oscillates regularly between $+1$ and -1 , with the same period as the signal, and with the highest value at time $t = 0$. For noisy periodic signals, the autocorrelation function is also oscillating; if the noise affects the phase of the oscillations of the signal, then the oscillations of the autocorrelation are dampened, the envelope of the function decreases exponentially, and the speed of dampening can be used as a measure to quantify the effect of noise on signal periodicity. For signals which are only random noise the autocorrelation function immediately reaches values very close to 0.

Generally, a signal is considered to be statistically different from random noise if its autocorrelation function is outside the 95% confidence band (which can be computed approximately as $\left[-\frac{2}{\sqrt{N}}, +\frac{2}{\sqrt{N}}\right]$, where N is the number of samples in the time-series). The *half-life* of the autocorrelation, corresponding to its 50% decrease (i.e. the time at which the envelope of the autocorrelation function is smaller than half its maximum), can be used as a measure of the robustness of periodic signals with respect to noise. An application of autocorrelation to biochemical oscillators can be found in [12].

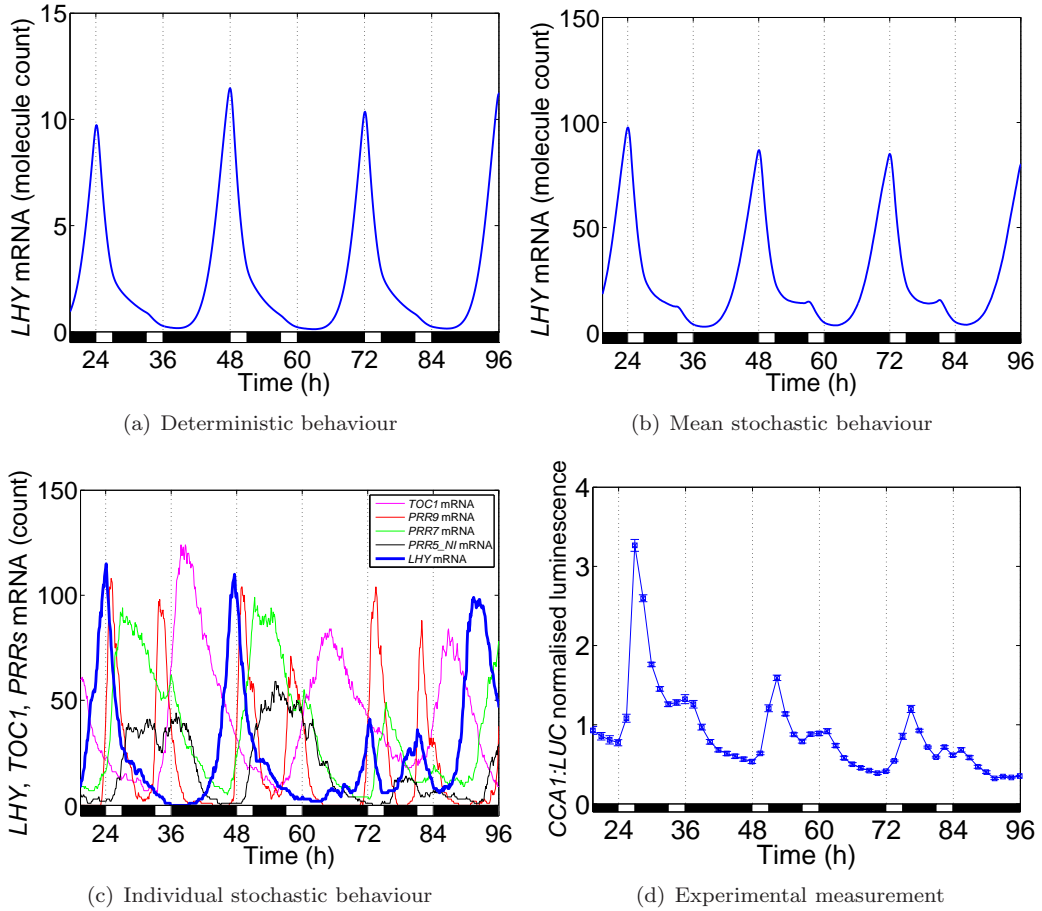


Figure 8: Comparison of the deterministic behaviour of LHY mRNA, its mean (over 1000 runs) and individual stochastic behaviour, and the experimentally measured amount (CCA1:LUC) in the skeleton light pattern; the experimental data represent normalised bioluminescence, averaged over 9 6-days-old plants [26]. The rare occurrence of high LHY peaks at dusk is visible in the individual stochastic run (c), and is the cause of the emergence of the secondary evening peak in the mean behaviour (b) and in experimental data (d). The simulations consisted in 40 days of LD 12:12 followed by the transition into the skeleton light pattern (time $t = 0$ in the plots is the time of transition).

Figure 9 shows the autocorrelation function for both the entrained and the free running system computed over LHY mRNA time-series data obtained from stochastic simulation. Supplementary Figure 4 (Appendix E) reports the autocorrelation of the deterministic behaviour. In the LD 12:12 system, the autocorrelation of the deterministic time-series is periodic and regular, from +1 to -0.5 (Supplementary Figure 4(a)). The mean autocorrelation of multiple stochastic runs is similar to the deterministic one: it is also regular but with a smaller amplitude, due to the variation in the oscillations (Figure 9(a)). The autocorrelation of a single stochastic run is also regular in periodicity, but not in amplitude since the amplitude of the time-series is not regular (Figure 9(b)).

In LL, the autocorrelation of the deterministic time-series is similar to the one in LD 12:12: it is periodic and regular, from +1 to -0.8 (Supplementary Figure 4(b)). The mean autocorrelation of multiple stochastic runs instead quickly dampens (Figure 9(c)), and its half-life is around 31 hours. The autocorrelation of a single stochastic run, despite being noisy, persistently oscillates outside the 95% confidence band (Figure 9(d)), thus clearly showing the presence of a regular circadian rhythm. In Figure 9(c) we can also observe that the mean autocorrelation function remains outside the 95% confidence band for 8–10 days. Hence, in addition to confirming the persistence of rhythmic oscillations in the individual cell behaviour, the above analysis also suggests that phase synchrony is totally lost only after 8–10 days.

The results for DD are similar to those for LL (Supplementary Figure 7 (Appendix E)): the autocorrelation of the deterministic time-series is periodic and regular, from +1 to -0.7 , the autocorrelation of the mean stochastic behaviour dampens, and the autocorrelation of an individual stochastic run is noisy and constantly outside the 95% confidence band. Compared to the LL system, in LL the autocorrelation half-life is longer, around 55 hours, and the phase distribution is slower, with the mean autocorrelation function remaining outside the 95% confidence band for 18–20 days.

Being normalised, the autocorrelation amplitude and its half life are good measures to compare the effect of noise on phase distribution and its relation with system size. The autocorrelation for the stochastic model with $\Omega = 10$ (Supplementary Figure 8 (Appendix E)) are qualitatively similar to the ones for $\Omega = 100$ but show important quantitative differences, and provide a further measure to quantify the noise-induced variations and their relation with the system size. Specifically, in the LD system, we can observe that the autocorrelation amplitude is considerably smaller for $\Omega = 10$ because of the larger variation in amplitude. In the LL system we notice that the mean autocorrelation of multiple stochastic runs immediately drops to zero, with a half life shorter than a single cycle, around 20 hours, showing that no synchrony is observed among the oscillations of individual cells; the autocorrelation of a single stochastic run, in fact, does not exhibit oscillations with a constant period due to the huge noise-induced fluctuations observed in the single cell behaviour.

4 Conclusions

The importance of molecular noise in the behaviour of genetic networks such as circadian clocks has been shown previously in various systems (e.g. [15, 1, 2]). In this work we have investigated the noise-induced behaviour in the circadian clock of the model organism *Arabidopsis thaliana*. The model we have presented here is obtained by extending an existing model of the clock [26] with the addition of stochasticity.

The simulation results obtained by analysing the discrete stochastic representation of the system allowed us to estimate the unknown size of the system to around 100 copies of proteins per cell. We also showed how moderate noise is able to accelerate the entrainment of the plant clock under certain light conditions and that the stochastic interpretation of the clock model showed better agreement with the available experimental data compared to the deterministic solution. The analysis of individual simulation runs, representing the

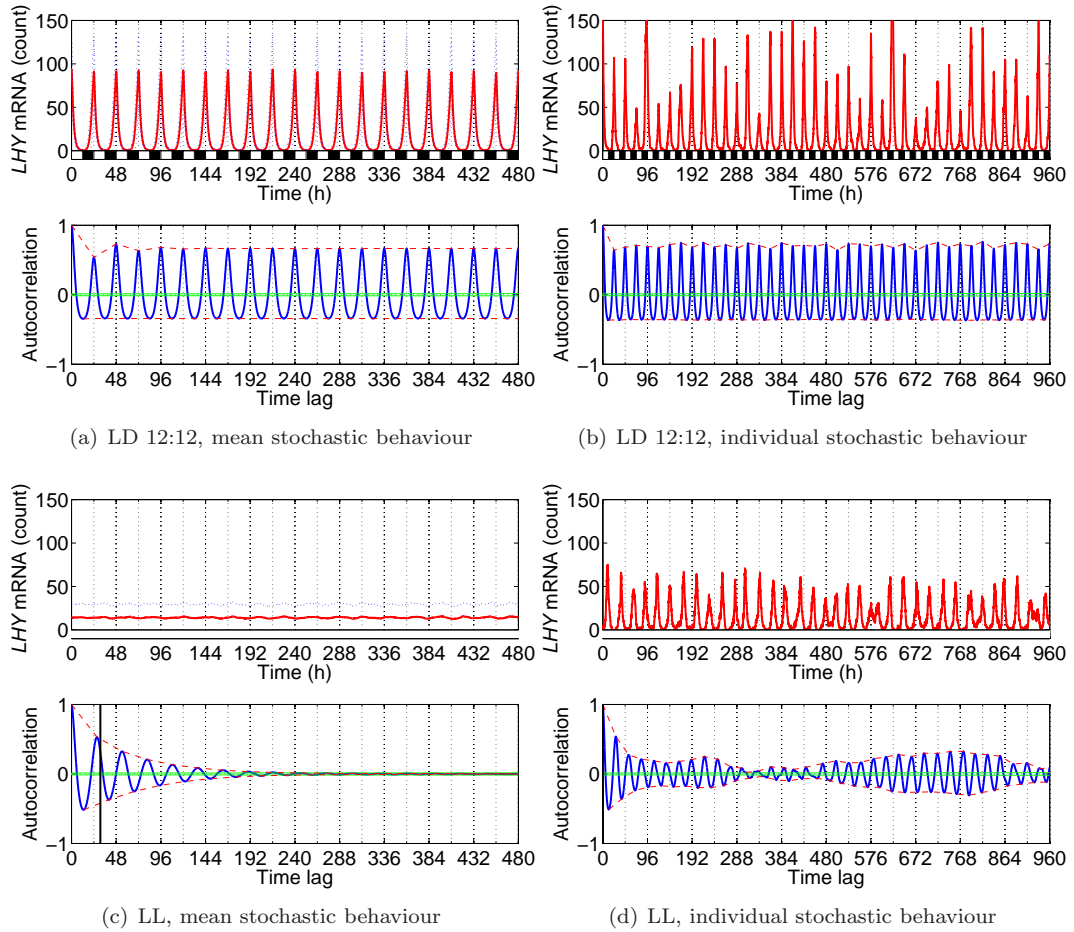


Figure 9: LHY mRNA time-series data and its autocorrelation function: (a–b) LD 12:12 and (c–d) LL. Comparison of the autocorrelation of a single stochastic run and of a mean behaviour obtained over 1000 stochastic simulation runs. The first 15 simulation days were discarded when computing the autocorrelation function in order to get rid of transient effects. In the autocorrelation plots, the green horizontal lines show the 95% confidence band, while the black vertical line is the half-life of the mean autocorrelation function.

behaviour of individual independent cells, helped us to explain the experimentally observed dampening of oscillations in constant light, which was not captured by the deterministic model: the oscillations in single cells, although persistent, are not synchronised in different cells, thus causing phase diffusion. The distribution of phase, period and amplitude computed over multiple simulation runs closely matched available experimental data, exhibiting robust entrainment of the plant clock to light/dark cycles with some fluctuations of the amplitudes. The half-life of the autocorrelation of simulation time-series data was used as a measure to quantify the dampening of oscillations in the free running system.

The suitability of stochastic models for understanding single cell behaviour in the presence of noise is well known. Interestingly, our stochastic model has also proven to give better account of population data compared to the deterministic model: the mean stochastic dynamics captures the damping of oscillations experimentally observed in constant light (Section 3.2) and shows a better entrainment to light in several light patterns, such as a faster entrainment to changes in photoperiod (Section 3.3) and the occurrence of secondary evening peaks in “skeleton” photoperiods (Section 3.4).

Acknowledgements

The Centre for Systems Biology at Edinburgh is a Centre for Integrative Systems Biology (CISB) funded by BBSRC and EPSRC, reference BB/D019621/1. AP was supported by BBSRC award E015263 to AJM, and by European Commission FP7 Collaborative Project TiMet (award 245143) to AJM and others. APF was supported by the ROBuST project, award F005237 from BBSRC and EPSRC to KJH and others. We are grateful to Dr Hirofumi Ishihara and Ronan Sulpice of the TiMet project team, Max Planck Institute of Molecular Plant Physiology, Potsdam-Golm, Germany, for the information about the total number of cells in *Arabidopsis* tissues. We gratefully acknowledge the expert technical assistance of Gavin Steel.

References

- [1] O.E. Akman, F. Ciocchetta, A. Degasperi, and M.L. Guerriero. Modelling Biological Clocks with Bio-PEPA: Stochasticity and Robustness for the *Neurospora Crassa* Circadian Network. In *Proc. of CMSB'09*, volume 5688 of *LNCS*, pages 52–67. Springer, 2009.
- [2] O.E. Akman, M.L. Guerriero, L. Loewe, and C. Troein. Complementary approaches to understanding the plant circadian clock. In *Proc. of FBTC, EPTCS*, pages 1–19, 2010.
- [3] Bio-PEPA Home Page. <http://www.biopepa.org/>.
- [4] F. Ciocchetta and J. Hillston. Process Algebras in Systems Biology. In *Formal Methods for Computational Systems Biology (SFM'08)*, volume 5016 of *LNCS*, pages 265–312. Springer-Verlag, 2008.
- [5] F. Ciocchetta and J. Hillston. Bio-PEPA: a Framework for the Modelling and Analysis of Biological Systems. *Theoretical Computer Science*, 410(33–34):3065–3084, 2009.
- [6] J.R. Dormand and P.J. Prince. A family of embedded Runge-Kutta formulae. *Journal of Computational and Applied Mathematics*, 6(1):19–26, 1980.
- [7] K.D. Edwards, O.E. Akman, K. Knox, P.J. Lumsden, A.W. Thomson, P.E. Brown, A. Pokhilko, L. Kozma-Bognar, F. Nagy, D.A. Rand, and A.J. Millar. Quantitative analysis of regulatory flexibility under changing environmental conditions. *Mol Syst Biol*, 6:424, 2010.

- [8] E.M. Farré, S.L. Harmer, F.G. Harmon, M.J. Yanovsky, and S.A. Kay. Overlapping and Distinct Roles of PRR7 and PRR9 in the Arabidopsis Circadian Clock. *Current Biology*, 15(1):47–54, 2005.
- [9] B. Finkenstädt, E.A. Heron, M. Komorowski, K. Edwards, S. Tang, C.V. Harper, J.R.E. Davis, M.R.H. White, A.J. Millar, and D.A. Rand. Reconstruction of transcriptional dynamics from gene reporter data using differential equations. *Bioinformatics*, 24(24):2901–2907, 2008.
- [10] D.B. Forger and C.S. Peskin. Stochastic simulation of the mammalian circadian clock. *PNAS*, 102(2):321–324, 2005.
- [11] H. Fukuda, N. Nakamichi, M. Hisatsune, H. Murase, and T. Mizuno. Synchronization of Plant Circadian Oscillators with a Phase Delay Effect of the Vein Network. *Physical Review Letters*, 99(9):098102, 2007.
- [12] P. Gaspard. The Correlation Time of Mesoscopic Chemical Clocks. *J. Chem Phys.*, 117:8905–8916, 2002.
- [13] M.A. Gibson and J. Bruck. Efficient Exact Stochastic Simulation of Chemical Systems with Many Species and Many Channels. *The Journal of Chemical Physics*, 104:1876–1889, 2000.
- [14] D.T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, 81(25):2340–2361, 1977.
- [15] D. Gonze, J. Halloy, and A. Goldbeter. Deterministic versus Stochastic Models for Circadian Rhythms. *Journal of Biological Physics*, 28:637–653, 2002.
- [16] D. Gonze, J. Halloy, and A. Goldbeter. Robustness of circadian rhythms with respect to molecular noise. *PNAS*, 99(2):673–678, 2002.
- [17] D. Gonze, J. Halloy, J.-C. Leloup, and A. Goldbeter. Stochastic models for circadian rhythms: effect of molecular noise on periodic and chaotic behaviour. *Comptes Rendus Biologies*, 326(2):189–203, 2003.
- [18] M.L. Guerriero and J.K. Heath. Computational Modeling of Biological Pathways by Executable Biology. In Michael L. Johnson and Ludwig Brand, editors, *Computer Methods, Part C*, volume 487 of *Methods in Enzymology*, pages 217–251. Academic Press, 2011.
- [19] J.Y. Kim, H.R. Song, B.L. Taylor, and I.A. Carré. Light-regulated translation mediates gated induction of the Arabidopsis clock protein LHY. *EMBO J.*, 22(4):935–944, 2003.
- [20] N. Komin, A.C. Murza, E. Hernández-García, and R. Toral. Synchronization and entrainment of coupled circadian oscillators. *Interface Focus*, 1:167–176, 2011.
- [21] N. Le Novère, B. Bornstein, A. Broicher, M. Courtot, M. Donizelli, H. Dharuri, L. Li, H. Sauro, M. Schilstra, B. Shapiro, J.L. Snoep, and M. Hucka. BioModels Database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. *Nucleic Acids Research*, 34(Database issue):D689–D691, 2006.
- [22] A.C. Liu, D.K. Welsh, C.H. Ko, H.G. Tran, E.E. Zhang, A.A. Priest, E.D. Buhr, O. Singer, K. Meeker, I.M. Verma, F.J. Doyle III, J.S. Takahashi, and S.A. Kay. Intercellular Coupling Confers Robustness against Mutations in the SCN Circadian Clock Network. *Cell*, 129(3):605–616, 2007.

- [23] J.C.W. Locke, L. Kozma-Bognar, P.D. Gould, B. Fehér, E. Kevei, F. Nagy, M.S. Turner, A. Hall, and A.J. Millar. Experimental validation of a predicted feedback loop in the multi-oscillator clock of *Arabidopsis thaliana*. *Mol. Syst. Biol.*, 2:59, 2006.
- [24] J.C.W. Locke, M.M. Southern, L. Kozma-Bognar, V. Hibberd, P.E. Brown, M.S. Turner, and A.J. Millar. Extension of a genetic network model by iterative experimentation and mathematical analysis. *Mol. Syst. Biol.*, 1:2005.0013, 2005.
- [25] E. Nagoshi, C. Saini, C. Bauer, T. Laroche, F. Naef, and U. Schibler. Circadian Gene Expression in Individual Fibroblasts: Cell-Autonomous and Self-Sustained Oscillators Pass Time to Daughter Cells. *Cell*, 119(5):693–705, 2004.
- [26] A. Pokhilko, S.K. Hodge, K. Stratford, K. Knox, K.D. Edwards, A.W. Thomson, T. Mizuno, and A.J. Millar. Data assimilation constrains new connections and components in a complex, eukaryotic circadian clock model. *Mol. Syst. Biol.*, 6:416, 2010.
- [27] J.D. Salazar, T. Saithong, P.E. Brown, J. Foreman, J.C.W. Locke, K.J. Halliday, I.A. Carré, D.A. Rand, and A.J. Millar. Prediction of Photoperiodic Regulators from Quantitative Gene Circuit Models. *Cell*, 139:1170–1179, 2009.
- [28] P.A. Salomé and C.R. McClung. PSEUDO-RESPONSE REGULATOR 7 and 9 Are Partially Redundant Genes Essential for the Temperature Responsiveness of the Arabidopsis Circadian Clock. *Plant Cell*, 17:791–803, 2005.
- [29] S.C. Thain, A. Hall, and A.J. Millar. Functional independence of circadian clocks that regulate plant gene expression. *Current Biology*, 10(16):951–956, 2000.
- [30] The Plant Systems-biology Modelling Database. <http://www.plasmo.ed.ac.uk>.
- [31] C. Troein, F. Corellou, L.E. Dixon, G. van Ooijen, J.S. O'Neill, F.-Y. Bouget, and A.J. Millar. Multiple light inputs to a simple clock circuit allow complex biological rhythms. *The Plant Journal*, 66(2):375–385, 2011.
- [32] E. Yakir, D. Hilman, M. Hassidim, and R.M. Green. CIRCADIAN CLOCK ASSOCIATED1 Transcript Stability and the Entrainment of the Circadian Clock in Arabidopsis. *Plant Physiology*, 145(3):925–932, 2007.
- [33] M.N. Zeilinger, E.M. Farré, S.R. Taylor, S.A. Kay, and F.J. Doyle III. A novel computational model of the circadian clock in *Arabidopsis* that incorporates PRR7 and PRR9. *Mol. Syst. Biol.*, 2:58, 2006.