Long-time analytic approximation of large stochastic oscillators: simulation, analysis and inference

Giorgos Minas[1,2] and David A Rand[1,2,*]

[1] Mathematics Institute and

[2] Zeeman Institute for Systems Biology & Infectious Disease Epidemiology Research,

University of Warwick, Coventry CV4 7AL, UK*

Abstract

In order to analyse large complex stochastic dynamical models such as those studied in systems biology there is currently a great need for both analytical tools and also algorithms for accurate and fast simulation and estimation. We present a new stochastic approximation of biological oscillators that addresses these needs. Our method, called phase-corrected LNA (pcLNA) overcomes the main limitations of the standard Linear Noise Approximation (LNA) to remain uniformly accurate for long times, still maintaining the speed and analytically tractability of the LNA. As part of this, we develop analytical expressions for key probability distributions and associated quantities, such as the Fisher Information Matrix and Kullback-Leibler divergence and we introduce a new approach to system-global sensitivity analysis. We also present algorithms for statistical inference and for long-term simulation of oscillating systems that are shown to be as accurate but much faster than leaping algorithms and algorithms for integration of diffusion equations. Stochastic versions of published models of the circadian clock and NF-κB system are used to illustrate our results.

Keywords: stochastic models — oscillations — Linear Noise Approximation — biological systems — Fisher information — Kalman filter

Abbreviations: pcLNA, phase-corrected Linear Noise Approximation; KS, Kolmogorov-Smirnov; FIM, Fisher Information Matrix

*to whom correspondence may be addressed

1 Introduction

Dynamic cellular oscillating systems such as the cell cycle, circadian clock and other signaling and regulatory systems have complex structures, highly nonlinear dynamics and are subject to both intrinsic and extrinsic stochasticity. Moreover, current models of these systems have high-dimensional phase spaces and many parameters. Modelling and analysing them is therefore a challenge, particularly if one wants to both take account of stochasticity and develop an analytical approach enabling quantification of various aspects of the system in a more controlled way than is possible by simulation alone. The stochastic kinetics that arise due to random births, deaths and interactions of individual species give rise to Markov jump processes that, in principle, can be analyzed by means of master equations. However, these are rarely tractable and although an exact numerical simulation algorithm is available [1], for the large systems we are interested in, this is very slow.

It is therefore important to develop accurate approximation methods that enable a more analytical approach as well as offering faster simulation and better algorithms for data fitting and parameter estimation. A number of approximation methods aimed at accelerating simulation are currently available. This includes leaping algorithms [2, 3] and algorithms for integration of diffusion equation (or chemical Langevin equation (CLE)) [4] that provide faster simulation. However, these methods do not provide analytical tools for studying the dynamics of the system and they can also be extremely slow for data fitting and parameter estimation. One obvious candidate for overcoming these limitations is the Linear Noise Approximation (LNA). The LNA is based on a systematic approximation of the master equation by means of van Kampen's Ω-expansion [5] which uses the system size parameter Ω that controls the number of molecules present in the system. The large system size validity of the LNA has been shown in [6], in the sense that the distribution of the Markov jump process at a fixed finite time converges, as the system size Ω tends to ∞, to the LNA probability distribution. The latter distribution is analytically tractable allowing for fast estimation and simulation algorithms. However, the LNA has significant limitations, particularly in approximating long-term behaviour of oscillatory systems.

We show below that for the oscillatory systems that we study, the LNA approximation of the distribution $P_t = P(Y(t)|Y(0))$, of the state $Y(t)$ of the system at some time t becomes inaccurate when the time t is greater than a few periods of the oscillation. However, if we rather consider a similar system which in the $\Omega \to \infty$ limit instead of a limit cycle has an equilibrium point that is linearly stable, then the LNA approximation of P_t remains accurate for a much longer time-scale. For example, in Fig C in S3 Appendix we give an example where the LNA fails in a matter of a period or two for the oscillatory system, but for the corresponding equilibrium system it is very accurate for over a hundred times as long (and probably much longer). Similar behaviour is also observed in other systems and using different measures in [8].

The observation that non-degenerate limit cycles have such linearised stability in the directions transversal to the limit cycle suggests the way forward for oscillatory systems. Our approach exploits the fact that, because of this transversal linearised stability, the distributions

P_t for a general class of systems with a stable attracting limit cycle in the $\Omega \to \infty$ limit are, like the above fixed point systems, similarly well-behaved on long time-scales provided one conditions P_t on appropriate transversal sections to the limit cycle.

We introduce a modified LNA, called the phase-corrected LNA, or pcLNA, that exploits the above observations to overcome the most important shortcomings of the LNA and we develop methods for analysis, simulation and inference of oscillatory systems that are accurate for much larger times. We build on previous work of Boland et al. [9] which uses the 2-dimensional Brusselator system as an exemplar to investigate the failure of the LNA in approximating long-term behaviour of oscillatory systems and presents a method for computing power spectra and comparing exact simulations with LNA predictions of the same phase rather than time. Using various low-dimensional oscillatory systems for illustration, a related analysis has been employed to study the temporal variability of oscillatory systems in the tangental direction of the $\Omega \to \infty$ limit cycle [10] and/or the amplitude variability in the transversal direction of the limit cycle [11–13]. Other papers derive related descriptions of the asymptotic phase of stochastic oscillators [14, 15].

We extend these results in a number of ways including the following: (i) we develop a theory that treats the general case and provide analytical arguments that justify our approximations and enable computation of trajectory distributions, (ii) we show that the approach is practicable for larger nonlinear systems, (iii) we present a new powerful system-global sensitivity theory for such systems using measures such as the Fisher Information Matrix and the Kullback-Leibler divergence that are analytically computed, (iv) we present a simulation algorithm and show it is as accurate but faster than leaping and integration of diffusion equation algorithms, and (v) we derive the Kalman filter associated with the pcLNA in order to provide a practical way to accurately approximate the likelihood function thus facilitating estimation of system parameters θ and predictive algorithms. The approach in [9] uses transversal sections which are normal to the limit cycle. We follow this but in the supplementary information (S1 Sects. 8.2 & 8.3) we show that for most considerations one can use any transversal to the limit cycle, including those defined in [14, 15].

To illustrate and validate our approach we apply it to a relatively large published stochastic model of the *Drosophila* circadian clock due to Gonze et al. [16] (see S2 Sect. 1). This model involves 10 state variables and 30 reactions and its structure is discussed in S2. The large system limit is given by the differential equation system of 10 kinetic equations that are listed in the supplementary information (S2 Sect. 1) along with the reaction scheme of the system. The stochastic version of the Brusselator system and a stochastic version of a well-studied model of the NF-κB signalling system [17] are also used to illustrate our methods and the results can be found in S3 Appendix.

These systems are free-running oscillators in the sense that they correspond to a limit cycle of an autonomous differential equation in the the $\Omega \to \infty$ limit. However, our results also apply to the equally important classes of entrained forced oscillators and damped oscillations. We therefore consider two such systems in S2 and S3: the light-entrained *Drosophila* circadian clock model of [18] which is an example of a forced oscillator and the NF-κB system model [17].

The latter has the extra feature that the analysis is not concerned with a limit cycle but of a transient solution that converges to the limit cycle as time increases. This solution is the biologically interesting one that describes how the system responds to being stimulated by TNFα. The supplementary information S1 includes technical derivations and S2 and S3 contain further illustrative figures that we refer to in this paper.

2 Results

Stochastic models of cellular processes in signaling and regulatory systems are usually described in terms of reaction networks. A system of multiple different molecular subpopulations has state vector, $Y(t) = (Y_1(t), \ldots, Y_n(t))^T$ where $Y_i(t)$, $i = 1, \ldots, n$, denotes the number of molecules of each species at time t. These molecules undergo a number of possible reactions (e.g. transcription, translation, degradation) where the reaction of index j changes $Y(t)$ to $Y(t) + \nu_j$, $\nu_j \in \mathbb{R}^n$. The vectors ν_j are called stoichiometric vectors. Each reaction occurs randomly at a rate $w_j(Y(t))$ (often called the intensity of the reaction), which is a function of $Y(t)$ but also depends periodically on t when we are studying forced oscillators. Such systems can be exactly simulated using the Stochastic Simulation algorithm (SSA) [1].

It is common in studying stochastic systems to introduce a system size Ω which is a parameter that occurs in the intensities of the reactions $w_j(Y(t))$ and controls molecular numbers (see discussion in S1 Sect. 2). While having a system size parameter is not necessary to apply our methods, it allows one to study the dependence of stochastic fluctuations upon system size and to calculate the deterministic equations that describe the evolution of the concentration vector $X(t) = Y(t)/\Omega$ in the limit of $\Omega \to \infty$ (see S1 Sect. 3). Although more general conditions can be used, a condition that will be sufficient for our purposes is that the rates $w_j(Y(t))$ depend upon Ω as $w_j(Y) = \Omega u_j(Y/\Omega)$ (cf. [5–7]).

In the limit $\Omega \to \infty$ the time dependence of $X(t)$ is given by the solution $x(t)$ of the differential equation

$$\dot{x} = F(x), \quad F(x) = \sum_j \nu_j u_j(x(t)), \tag{1}$$

with the appropriate initial condition (see S1 Sect. 3). For free-running oscillators the differential equation (1) is autonomous, whereas for forced oscillators F also depends periodically on t.

Throughout we will be interested in the case where the solution $x(t)$ of interest is a stable limit cycle γ of minimal period $\tau > 0$ given by $x = g(t)$, $0 \leq t \leq \tau$. We shall also always assume the generic situation for stable limit cycles of autonomous systems in which one of the characteristic exponents of the limit cycle is equal to zero and the rest have negative real part ([19], [20, Sect. 1.5] and S1 Sect. 1). For an entrained forced oscillator all the characteristic exponents of the limit cycle have negative real parts.

Fig 1(A) displays a stochastic trajectory of the concentrations $X(t) = Y(t)/\Omega$ of two of the species of the circadian clock system obtained using exact SSA simulation over a period of time $t \in [0, 8.5\tau]$ where $\tau \approx 26.98$ hours is the period of the limit cycle γ. Here the system size is

$\Omega = 300$, imposing moderate to high levels of stochasticity (see also Table C in S2 Appendix). Results for system sizes $\Omega = 200$, 500 and 1000 are also reported in Fig B in S2 Appendix. Fig 1(B) shows realizations of the key probability distributions

$$P(X(t)|X(t_0) = x_0), \qquad t > t_0, \tag{2}$$

which describe the state of the system at some time $t > t_0$. It is very rare that accurate analytical approximations for such probability distributions can be derived from the exact Markov-jump process when t is large. Furthermore, as we can see, the SSA samples of $P(X(t)|X(t_0) = x_0)$ spread along the curved periodic solution, $x = g(t)$, of the limiting ($\Omega \to \infty$) deterministic system, implying that for large t this distribution is far from being normal and has a complex structure.

However, we will show that there are important related distributions that can be well approximated even for large times t. For example, for each point x on the $\Omega \to \infty$ limit cycle γ, consider the $(n-1)$-dimensional hyperplane $\mathcal{S}_x$ normal to γ at $x \in \gamma$ (i.e. orthogonal to the tangent vector $F(x)$ at x). We will show that the distribution of the intersection points of stochastic trajectories $X(t)$ with $\mathcal{S}_x$ can be well-approximated by a multivariate normal (MVN) distribution that can be relatively easily calculated.

We need to define more precisely what we mean by intersection points. Consider the mapping G_N defined by $G_N(X) = x \in \gamma$ if $X \in \mathcal{S}_x$ (see Fig 1(C)). Suppose that the limit cycle is given by $x = g(t)$ and extend $g(t)$ to all $t \in \mathbb{R}$ by periodicity. Now consider a stochastic trajectory $X(t_i)$, $i = 0, 1, 2, \ldots$. Suppose that $G_N(X(t_{i-1})) = g(s_{i-1})$, $i > 0$. If $G_N(X(t_i)) = g(s)$ then it equals $g(s + q\tau)$ for all integers q. Choose q so that $s_i = s + q\tau$ satisfies $s_{i-1} - \tau/2 \le s_i < s_{i-1} + \tau/2$. Note that the difference between s_i and s_{i-1} can be both positive and negative and the choice of s_i minimises $|s_i - s_{i-1}|$. We define the relative phase $\delta\varphi_X(t_i)$ to be $2\pi(s_i - s_{i-1})/\tau$. With this definition the lifted phase $\bar{\varphi}_X$ is defined to be the function

$$\bar{\varphi}_X(t_0) = 0 \quad \text{and} \quad \bar{\varphi}_X(t_i) = \sum_{k=1}^{i} \delta\varphi_X(t_i) = 2\pi(s_i - s_0)/\tau \quad \text{if} \quad i \ge 1.$$

An example of $\bar{\varphi}_X(t)$ is shown in Fig 1(C). Now, as shown in that figure, if the system size is not too small, although there will be some reversals, the stochastic process $G_N(X(t))$, $t > 0$, will move around γ in the direction of the deterministic flow given by (1) so that $\bar{\varphi}_X(t)$ increases at a definite positive rate because phase advances exceed retreats. Our approximations (S1 Sect. 6) give that the long term rate is $2\pi/\tau$ and that the variance of the fluctuations $\bar{\varphi}_X(t) - 2\pi t/\tau$ grows linearly with a growth rate that is $\mathrm{O}(t/\Omega)$.

Now consider stochastic trajectories $X(t_i)$ with $X(t_0) \in \mathcal{S}_{g(T_0)}$ and consider how this trajectory passes through $\mathcal{S}_{g(T_1)}$. We can assume that $T_0 < T_1 \le T_0 + \tau$ by the periodicity of g. The rth pass will occur between the first time t_i when

$$\bar{\varphi}_X(t_i) \le 2\pi(r-1) + 2\pi(T_1 - T_0)/\tau < \bar{\varphi}_X(t_{i+1})$$

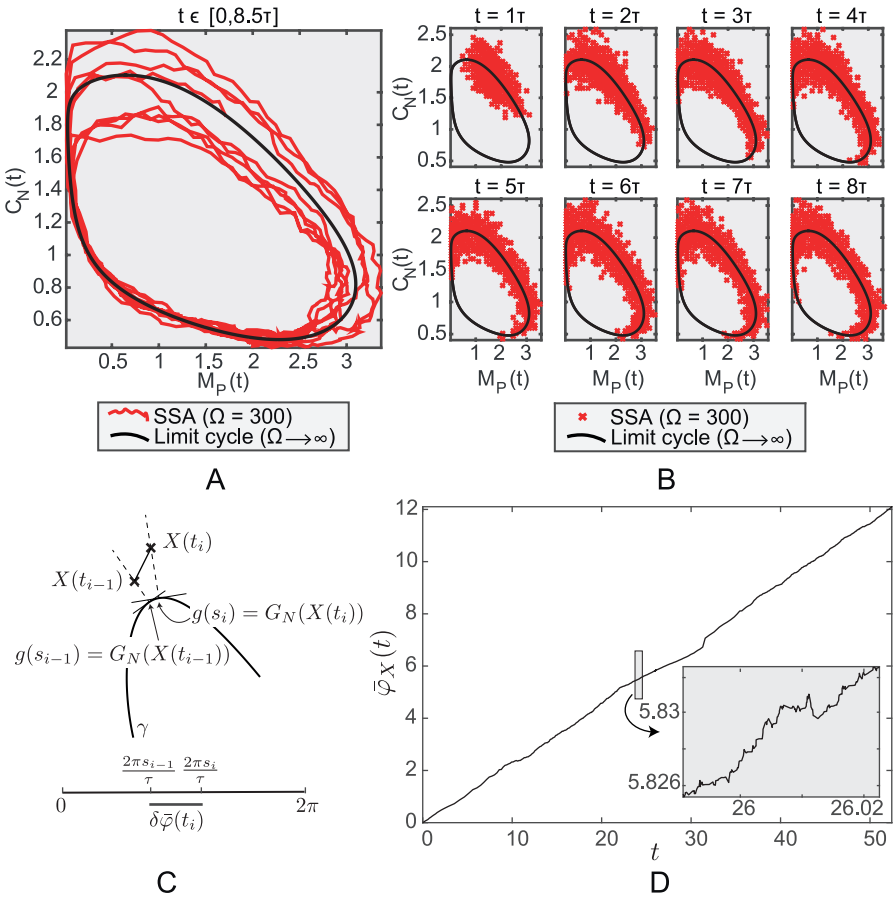

Fig 1. Exact stochastic simulation of the *Drosophila* circadian clock system. (A) A stochastic trajectory obtained by running the SSA over the time-interval $t \in [0, 8.5\tau]$ and (B) SSA samples ($R = 3000$) at times $t = \tau, 2\tau, \ldots, 8\tau$. The concentrations, $X(t) = Y(t)/\Omega$, of two (out of 10) of the species are displayed (*per* mRNA M_P (x-axis) and nuclear PER-TIM complex C_N (y-axis)). For this model the system size Ω is Avogadro's number in units of nM^{-1} multiplied by the cell volume in litres and the concentrations are nanomolar. The value of Ω used here is 300 in units of nM which assumes a cell volume of approximately 0.5×10^{-12} litres. The number of M_P molecules ranges from 0 to 1200 and the number of C_N molecules from 100 to 900. It is only through Ω that the cell volume appears in the equations. The black solid curve is the $\Omega \to \infty$, limit cycle solution. (C) Schematic diagram illustrating the mapping $G_N(\cdot)$ and the relative phase $\delta\varphi_X(\cdot)$. (D) A plot of the lifted phase function $\bar{\varphi}_X(t)$ for a trajectory of this system. Note the long-term linear increase combined with frequent reversals.

and t_{i+1}. Therefore, we define the *first intersection point of the rth pass* to be

$$Q_{x_1}^{(r)} = Q_{x_1}^{(r,X)} = X(t_i)$$

since X does not change between t_i and t_{i+1}.

These points of intersection $Q_x^{(r)}$ describe the stochasticity of the system around a particular phase x of the system. The above ideas work equally well with any transversal to γ. For example, the time at which the i-th variable is maximal in the deterministic system is given by the transversal submanifold Σ defined by $\dot{x}_i = 0, \ddot{x}_i < 0$ so the intersection of the stochastic trajectory with Σ can be regarded as giving the statistics of the maxima of the stochastic trajectory (see Fig 2(A)). Close to the limit cycle, Σ is well-approximated by its tangent space at the point of intersection with the limit cycle. Therefore, these transversal distributions are useful for analysing various aspects of the system.

Our first observation is that the empirical transversal distributions

$$P_{x_1}^{(r)} = P(Q_{x_1}^{(r,X)} | X(t_0) = x_0)$$

obtained by exact simulation are approximately multivariate normal (Fig 2(B)). Moreover, as r increases $P_{x_1}^{(r)}$ and $P_{x_1}^{(r+1)}$ are hardly distinguishable and appear to converge to a fixed, approximately normal transversal distribution as r increases (Fig 2(C)). Similar results hold if we use a different family of transversal sections to γ as explained in S1 Sect. 8.2 & 8.3.

A natural question that arises is whether one obtains a different distribution when instead of taking the first intersection point of the rth pass one takes a later intersection point of the same round. This is addressed in S1 Sect. 12 where we show that in exact simulations there appears to be no difference as would be expected from the LNA approximation.

The Linear Noise Approximation (LNA)

The convergence of the transversal distributions to approximately normal distributions naturally raises the question of whether asymptotic approximation methods such as the LNA, which provide multivariate normally distributed approximation of the stochastic system, can be used to accurately approximate these transversal distributions.

The LNA as formulated by [6] is derived directly from the underlying Markov jump process and is valid for any time interval of finite fixed length. It is based on the ansatz

$$X(t) = \frac{Y(t)}{\Omega} = x(t) + \frac{\xi(t)}{\sqrt{\Omega}} \tag{3}$$

where $x(t)$ is a solution of the limiting ($\Omega \to \infty$) deterministic system (1) and $\xi(t)/\sqrt{\Omega}$ describes the stochastic variations. In our case we always take $x(t)$ to be the periodic solution $g(t)$. A key aspect of this ansatz is that $\xi(t)$ satisfies a linear stochastic differential equation that is independent of Ω, with drift and diffusion matrices that are functions of the deterministic solution $g(t)$. Details are given in S1 Sect. 4 & 5.

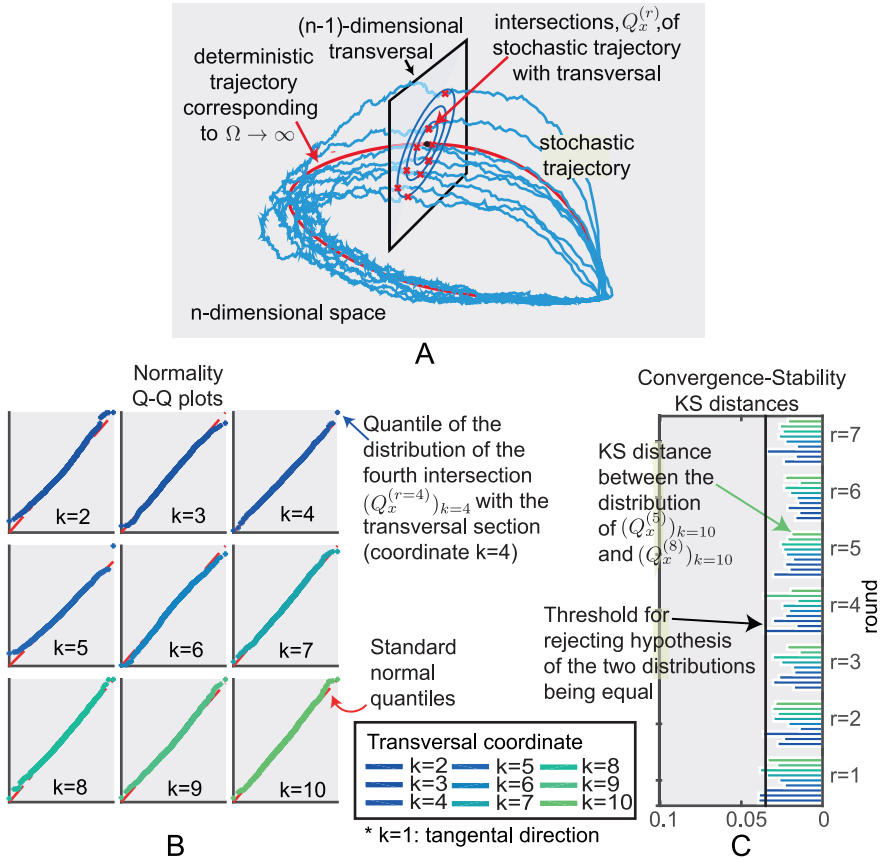

Fig 2. Intersections of the stochastic trajectories with a transversal section $\mathcal{S}_{g(t_0)}$. (A) Schematic representation of the intersections. (B) We consider an adapted coordinate system $(x_1, \ldots, x_{10})$ (see S1 Sect. 1) so that $(x_2, \ldots, x_{10})$ are orthogonal coordinates on $\mathcal{S}_{g(t_0)}$ and then consider the distribution $P^{k,r}$ of the values of x_k for $k = 2, \ldots, 10$ at the intersection points $Q^{(r)}_{g(t_0)}$. Quantile-Quantile (Q-Q) plots of these distributions for the fourth pass, $Q^{(r=4)}_{g(t_0)}$, show that the distributions are very close to being normal, (see Fig C in S2 Appendix for similar plots for $r = 1, 2, \ldots, 8$). (C) KS distances between the above distributions for the first $r = 1, 2, \ldots, 7$ rounds of the stochastic trajectory and the distribution in the 8th round ($r = 8$), $k = 2, 3, \ldots, 10$.

Given an initial time t_0 and an initial condition $\xi(t_0)$ for ξ, the LNA determines the distribution, of $\xi(t)$, $t > t_0$, and hence $X(t) = g(t) + \xi(t)/\sqrt{\Omega}$ that we respectively denote by $P_{\text{LNA}}(\xi(t)|t_0, \xi(t_0))$ and $P_{\text{LNA}}(X(t)|t_0, \xi(t_0))$. If $\xi(t_0)$ is only known up to a multivariate normal (MVN) distribution P_0 then we denote these distributions, respectively, by $P_{\text{LNA}}(\xi(t)|t_0, \xi(t_0) \sim P_0)$ and $P_{\text{LNA}}(X(t)|t_0, \xi(t_0) \sim P_0)$. Details of how to calculate these distributions are given in S1 Sect. 4. Each of the above distributions is MVN enabling analytical approaches, for example in analysing the stochastic sensitivities of the system.

If we fix $t > t_0$ then as $\Omega \to \infty$ the true distribution of ξ converges to the distribution $P_{\text{LNA}}(\xi(t)|t_0, \xi(t_0))$ (see e.g. [5]). However, one most certainly cannot reverse the limits i.e. for a fixed Ω one cannot expect the approximation to hold for large time $t \to \infty$. As we now show, this is certainly the case for oscillators and we aim to overcome this limitation by developing methods that remain accurate for much larger times than the LNA.

We first consider the distribution $P(X(t)|X(0) = x_0)$ and compare this for SSA simulated samples and the LNA at a sequence of times $t = \tau, 2\tau, \ldots, 8\tau$ and for an arbitrary (fixed) initial state $x_0 \in \gamma$. As we can see in Fig 3, the LNA fits the SSA simulations relatively well in the short run ($t = \tau$), but as time progresses the Kolmogorov-Smirnov (KS) distance between the two distributions for each state variable for the LNA and the SSA increases substantially beyond the threshold level (see Fig 3(B)). The LNA predictions spread along the tangental direction and therefore fail to accurately reflect the SSA samples that have instead spread along the curved limit cycle.

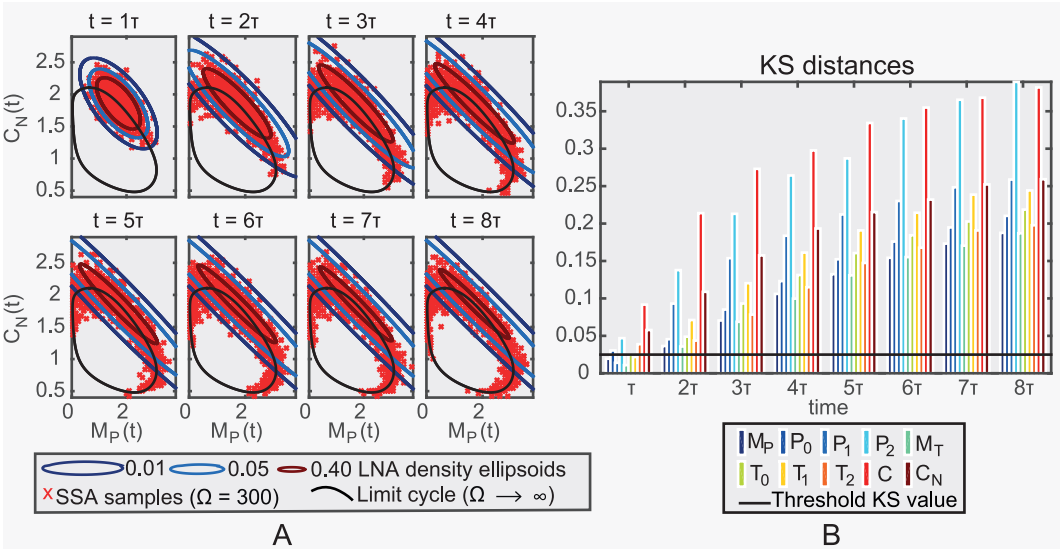

Fig 3. Comparison between LNA and exact simulations. (a) Samples (in nanomolar concentrations) obtained from the SSA simulation algorithm (red crosses) and $0.01, 0.05, 0.40$ contours of the LNA probability density (black ellipsoids) at fixed times, $t = \tau, 2\tau, \ldots, 8\tau$ (τ: minimal period), for the circadian clock system (M_P *per* mRNA; C_N nuclear PER:TIM complex P_0 & T_0 PER & TIM protein; P_1 & T_1 phosphorylated PER & TIM protein; P_2 & T_2 twice phosphorylated PER & TIM protein; C cytoplasmic PER:TIM complex, units as in Fig 1). The limit cycle ODE solution is also displayed (black solid line). (b) KS distance between the empirical distribution of SSA samples and the LNA distribution of each species (different colors, see legend) at the fixed times. The threshold level is also displayed (black solid line). The system size is $\Omega = 300$.

On the other hand, as we saw earlier, the transversal distributions $P_{x_1}^{(r)} = P(Q_{x_1}^{(r)}|X(t_0) = x_0)$

of the *Drosophila* circadian clock system are approximately normal (Fig 3(B)). We next derive an approximation of $P_{x_1}^{(r)}$ under the LNA and show that it accurately approximates $P_{x_1}^{(r)}$ for the *Drosophila* circadian clock, Brusselator and NF-κB systems.

Calculating transversal distributions

We now generalise slightly and consider the set of stochastic trajectories X where the initial conditions $X(t_0)$ have a MVN distribution P_0 that is supported on the normal transversal section $\mathcal{S}_{g(t_0)}$ (denoted $X(t_0) \sim P_0$). We consider how to approximate the distribution $P_{x_1}^{(r,P_0)}$ of the intersection points $Q_{g(t)}^{(r,X)}$ of these trajectories with the normal transversal section $\mathcal{S}_{g(t_1)}$, $t_1 > t_0$. As an approximation we take the conditional distribution

$$P_{\mathrm{LNA},t_1}^{(r)} = P_{\mathrm{LNA}}(X(t_1^{(r)})|X(t_1^{(r)}) \in \mathcal{S}_{g(t_1)}, \xi(t_0) \sim P_0), \quad t_1^{(r)} = t_1 + (r-1)\tau \tag{4}$$

given by conditioning $P_{\mathrm{LNA},t_1}^{(r,\mathrm{free})} = P_{\mathrm{LNA}}(X(t_1^{(r)})|\xi(t_0) \sim P_0)$ on $X(t_1^{(r)}) \in \mathcal{S}_{g(t_1)}$. It gives a MVN distribution supported on $\mathcal{S}_{g(t_1)}$.

In S1 we show that, although, for free-running oscillators, $P_{\mathrm{LNA},t_1}^{(r,\mathrm{free})}$ diverges as $r \to \infty$, the mean and covariance of the MVN transversal distribution $P_{\mathrm{LNA},x_1}^{(r)}$ converge exponentially fast to those of a MVN distribution $P_{\mathrm{LNA},x_1}^{(\infty)}$ (S1 Sect. 8, cf. Fig 4). The distribution $P_{\mathrm{LNA},x_1}^{(\infty)}$ is a fixed point in the sense that if the distribution of $X(t_1+\tau)$ is given by the LNA using as initial condition $\xi(t_1) \sim P_{\mathrm{LNA},x_1}^{(\infty)}$ then conditioning on $X(t_1+\tau) \in \mathcal{S}_{g(t_1)}$ gives

$$(X(t_1+\tau)|X(t_1+\tau) \in \mathcal{S}_{g(t_1)}) \sim P_{\mathrm{LNA},x_1}^{(\infty)}.$$

Using this fact enables us to calculate $P_{\mathrm{LNA},x_1}^{(\infty)}$ directly because we show in S1 that its mean and covariance matrix satisfy a simple fixed point equation that is easily solved numerically (S1 Sect. 9.1).

The reader will note that in (4) we approximate by conditioning on $X(t_1^{(r)}) \in \mathcal{S}_{g(t_1)}$ whereas we should have conditioned on $X(t) \in \mathcal{S}_{g(t_1)}$ for arbitrary t corresponding to the rth round. In S1 Sect. 7 we argue that the error in the mean and variance of the distribution due to taking $t = t_1^{(r)}$ is $\mathrm{O}(\Omega^{-1})$.

The question remains as to how well these distributions capture the exact simulation transversal distribution $P_{x_1}^{(r)}$. This is addressed in Fig 5 where it is shown that the fit is excellent even for Ω as low as 300. The fit is even better for higher system sizes (Fig E in S2 Appendix). In S3 we also show similar low Ω results for the Brusselator (Fig B in S3 Appendix) and the NF-κB system (Fig E in S3 Appendix). The result is also true for the light-entrained *Drosophila* circadian clock system (see Fig J in S2 Appendix) and the transient oscillations of the NF-κB system (see Fig H in S3 Appendix). Thus we note that although the LNA cannot be used directly to accurately compute $P(X(t)|X(0))$ for a fixed Ω and increasing t, using it to compute the transversal distributions provides accurate estimates of $P_{x_1}^{(r)}$ for much larger times $t_1 + r\tau$.

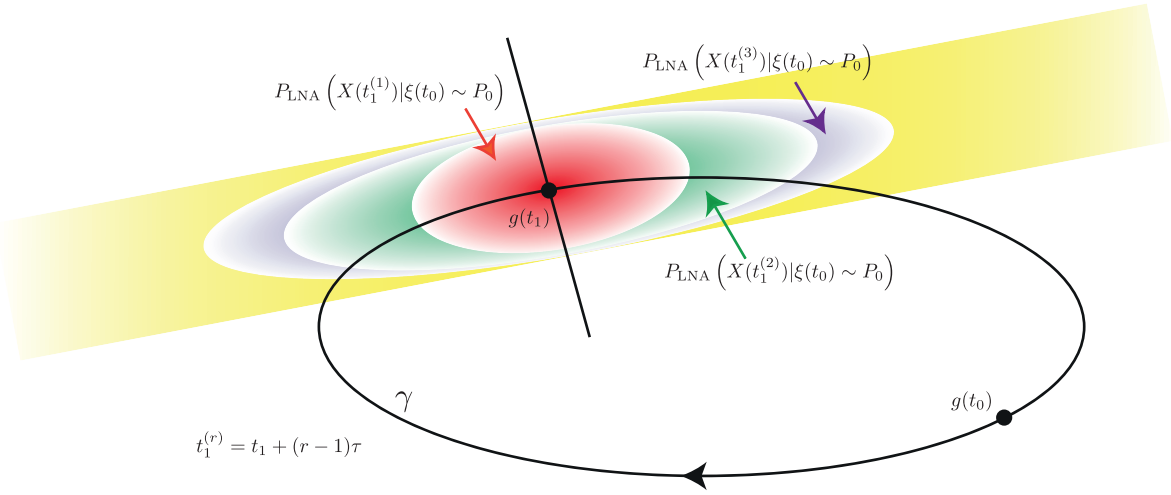

Fig 4. Schematic diagram of the distributions $P_{\mathbf{LNA}}(X(t_1^{(r)})|\xi(t_0)\sim P_0)$ as r increases. The yellow distribution represents the limit as $r\to\infty$. Although, for free oscillators the latter distribution diverges, the corresponding conditional distributions on the transversal converge.

Moreover, in S1 Sects. 8.2 & 8.3 we also explain why the convergence of the distribution on normal hyperplanes implies convergence on other transversal sections to γ.

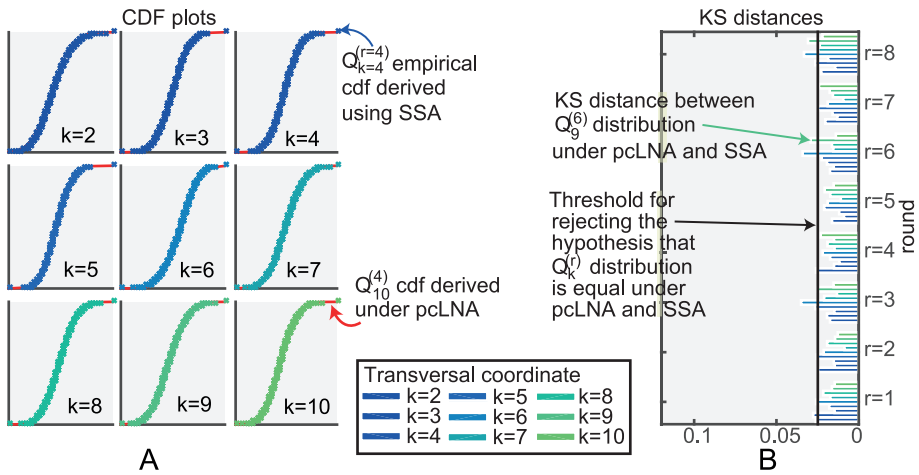

Fig 5. Comparison of pcLNA and exact transversal distributions for the *Drosophila* circadian clock ($\Omega = 300$). (A) CDF plots of the distributions $P^{k,r}$ as in Fig 2 for the pcLNA (red line) and the SSA (empirical CDF, crosses) for $k = 2, 3, \ldots, 10$ and round $r = 4$ (see Fig D in S2 Appendix for $r = 1, 2, \ldots, 8$). (B) KS distances between the corresponding distributions under pcLNA and SSA, $r = 1, 2, \ldots, 8$, $k = 2, 3, \ldots, 10$.

In S1 Sect. 8.4. we explain that in contradistinction to free-running oscillators, for entrained

forced oscillators, $P_r = P_{\text{LNA}}(X((r-1)\tau + t_1)|x_0, \xi_0 \sim P_0)$ converges as $r \to \infty$ so that, under the LNA, the phase fluctuation have a variance that is bounded independently of r. The corresponding conditional distribution is therefore a correspondingly good approximation to the transversal distribution $P_{x_1}^{(r,P_0)}$ (see Fig J in S2 Appendix). However, it does not mean that $P_{\text{LNA},t_1}^{(r,\text{free})}$ is a good approximation to the corresponding distribution $P(X(t_1^{(r)})|X(0))$ for an exact simulation. In fact, we show in Fig I in S2 Appendix that $P_{\text{LNA},t_1}^{(r,\text{free})}$ is a poor approximation of the empirical distribution $P(X(t_1^{(r)})|X(0))$ derived from exact simulations for the light-entrained *Drosophila* circadian clock ($\Omega = 300$). The bounded variance of the phase fluctuations as $r \to \infty$ for forced oscillators is the basic mechanism behind the population-level entrainment of stochastic oscillators introduced in [21].

Stochastic fluctuations in periods and timing

We now consider the fluctuations δt in the time taken for the lifted phase of a stochastic trajectory to go from a given phase φ_1 to a greater one φ_2. If $\varphi_2 - \varphi_1 = 2r\pi$ then this corresponds to the time taken to perform r cycles.

In Fig 6 we give an example using the *Drosophila* circadian clock model where we take φ_1 to be 0 (with a fixed initial condition $x_0 \in \gamma$) and $\varphi_2 = 2\pi r$ for $r = 1, \ldots, 8$. The distributions of δt appear to be very close to normal and the variance appears to grow linearly with r. We also consider the case where $\varphi_1 = 2(r-1)\pi$ and $\varphi_2 = 2r\pi$ for $r = 1, \ldots, 8$. Again the distributions are approximately normal but the variances are approximately constant (Fig 6(C)). Because for a given r the trajectory has done $r-1$ cycles before reaching the lifted phase φ_1 the distribution of the state at this phase is changing with r. We expect that this distribution is converging with increasing r and this result (Fig 6(C)) is in accordance with this.

In S1 Sect. 6 we approximate the statistics of δt using the LNA and show that as a random variable δt is approximately normal with mean that is $\mathrm{O}(\Omega^{-3/2})$ and we also calculate its variance up to terms that are $\mathrm{O}(\Omega^{-3/2})$ and the extent of its divergence from normality.

All points with a given lifted phase φ_1 lie in a particular transversal $\mathcal{S}_{g(t_1)}$ with $0 \leq t_1 < \tau$. If $t_2 = t_1 + (\varphi_2 - \varphi_1)\tau/2\pi$, then, the mean and variance of δt can be calculated in terms of t_1 and t_2. If the initial conditions $\xi(t_1)$ are MVN distributed on $\mathcal{S}_{g(t_1)}$ with mean 0 and covariance V_1, this variance is $(1/\alpha^2\Omega)\,\check{V}_{11} + \mathrm{O}(\Omega^{-3/2})$ where $\check{V}_{11} = \check{V}(t_1, t_2)_{11}$,

$$\check{V}(t_1, t_2) = C(t_1, t_2) V_1 C(t_1, t_2)^T + V(t_1, t_2).$$

written in adapted coordinates at $g(t_2)$ (see S1 Sect. 1). All terms on the right hand side of this equation are defined in S1 Sect. 4. The above exact simulations of the *Drosophila* circadian clock agree with these theoretical predictions. It is easy to see (cf. S1 Sect. 8) that $\check{V}_{11}$ grows roughly linearly with $t_2 - t_1$.

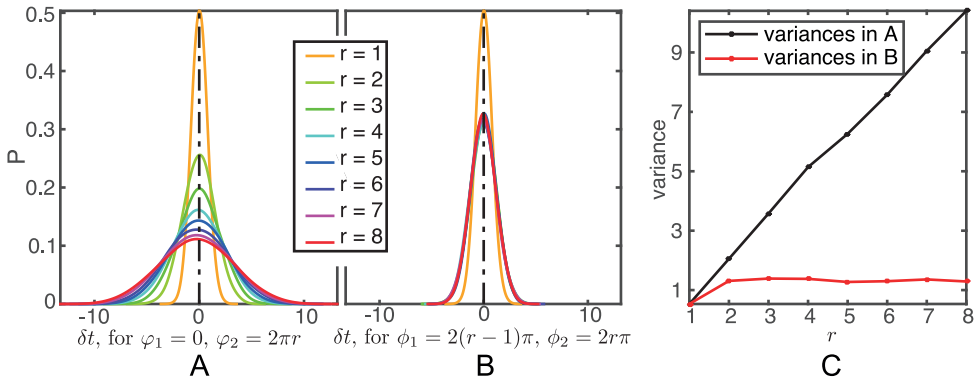

Fig 6. Exact empirical distribution of the fluctuations δt in *Drosophila* circadian clock system size $\Omega = 300$. (A) The (smoothed) empirical density function of δt in the time taken for the lifted phase of a stochastic trajectory to go from $\varphi_1 = 0$ to $\varphi_2 = 2r\pi$, $r = 1, 2, \ldots, 8$. (B) The (smoothed) empirical density function of the fluctuations in the time taken for the lifted phase of a stochastic trajectory to go from $\varphi_1 = 2(r-1)\pi$ to $\varphi_2 = 2r\pi$, $r = 1, 2, \ldots, 8$. (C) The variance of the distributions in (A) and (B).

pcLNA Simulation Algorithm

Given the ability to accurately approximate the transversal distributions and the results in [9] we realised it should be possible to use this to construct a rapid simulation algorithm. The linear increase of the variance of the deviations δt, or equivalently, the linear growth in the variance of the deviation of the lifted phase $\bar{\varphi}_X(t_i)$ from $2\pi t_i/\tau$, indicates the reasons for the long-time failure of the standard LNA. It is unable to cope with the increasing phase deviations. This motivates the phase correction approach used in the simulation algorithm we now define.

The approach is to amend the LNA Ansatz $X(t) = g(t) + \Omega^{-1/2}\xi(t)$ to $X(t) = g(s) + \Omega^{-1/2}\kappa(s)$ where $g(s) = G_N(X(t))$ and to use resetting of t to s to cope with the growth in the variance of $\bar{\varphi}_X(t_i) - 2\pi t_i/\tau$ keeping the LNA fluctuation $\kappa(s)$ normal to γ. While for free-running oscillators the variance of $\xi(t)$ grows without bound as t increases, $\kappa(s)$ has uniformly bounded variance.

The pcLNA simulation algorithm iteratively uses standard LNA steps of length $\Delta\tau$ to move from a state $X(s_{i-1})$ to a new state $X(s_{i-1} + \Delta\tau) = X_i$, $i = 1, 2, \ldots$. After each LNA step, the phase of the system is reset or "corrected" such that $g(s_i) = G_N(X_i)$ and the (global) fluctuations $\xi(s_{i-1} + \Delta\tau) = \Omega^{-1/2}(X_i - g(s_{i-1} + \Delta\tau))$ are replaced by the normally transversal fluctuation $\kappa(s_i) = \Omega^{-1/2}(X_i - g(s_i))$ which are MVN distributed and, as we showed in the previous section, approximate well the transversal fluctuations under the exact Markov Jump process.

The steps of the pcLNA simulation algorithm are described next in more detail (see also Fig 7).

1. Choose a time-step size $\Delta\tau > 0$.

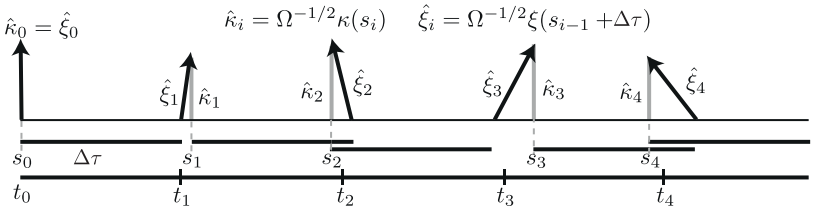

Fig 7. The main step in the pcLNA algorithm. The solid horizontal bars below the horizontal axis are all of length $\Delta\tau$, the basic time step of the algorithm. The black arrows show $\hat{\xi}_i = \Omega^{-1/2}\xi(s_{i-1} + \Delta\tau)$ and the grey arrows $\hat{\kappa}_i = \Omega^{-1/2}\kappa(s_i)$. Having calculated $\hat{\kappa}(s_{i-1})$ one uses $\kappa(s_{i-1})$ as the initial state and updates it using the LNA and a time-step $\Delta\tau$ to obtain ξ at $s_{i-1} + \Delta t$. Then $\xi(s_{i-1} + \Delta\tau)$ is replaced by $\kappa(s_i)$ so that $g(s_{i-1} + \Delta\tau) + \Omega^{-1/2}\xi(s_{i-1} + \Delta\tau) = g(s_i) + \Omega^{-1/2}\kappa(s_i)$ where $\kappa(s_i)$ is normal to the limit cycle. Therefore, s_i gives the phase of $\kappa(s_i)$ and the corresponding time is $t_i = t_0 + i\,\Delta\tau$.

2. Input **initial conditions** $\kappa(s_0)$ and $X_0 = g(s_0) + \Omega^{-1/2}\kappa(s_0)$.
3. **For iteration** $i = 1, 2, \ldots$
   (a) sample $\xi(s_{i-1} + \Delta\tau)$ from $\mathrm{MVN}(C_i\kappa(s_{i-1}), V_i)$;
   (b) compute $X_i = g(s_{i-1} + \Delta\tau) + \Omega^{-1/2}\xi(s_{i-1} + \Delta\tau)$;
   (c) set s_i to be such that $G_N(X_i) = g(s_i)$ and $\kappa_i = \Omega^{1/2}(X_i - g(s_i))$.

In the for loop $C_i = C(s_{i-1}, s_{i-1} + \Delta\tau)$ and $V_i = V(s_{i-1}, s_{i-1} + \Delta\tau)$ are the drift and diffusion matrices in the linear SDE describing the evolution of the noise process $\xi(t)$ under the LNA (see S1 Sect. 4).

The simulated sample X_i corresponds to time $t_i = t_0 + i\Delta\tau$, $i = 1, 2, \ldots$, where t_0 is the initial time. The time t_i is not necessarily equal to the phase s_i, defined by the relation $g(s_i) = G_N(X_i)$, which is stochastic and has variance linearly increasing with the time step $\Delta\tau$.

If one wants to record simulated trajectories at a finer time-scale than $\Delta\tau$ then one can run the algorithm with $\Delta\tau$ replaced by $\Delta\tau/M$ for some integer $M > 1$ and only carry out the phase correction in step 3(c) every Mth step and at all the other steps just proceeding as in the standard LNA (ignoring step 3(c)). This gives the same distribution as if the intermediate points had not been calculated because of the transitive nature of the LNA i.e. the distribution $P_{\mathrm{LNA}}(X(s+t)|X(0))$ is equal to the distribution $P_{\mathrm{LNA}}(X(t)|X(s) \sim P_{\mathrm{LNA}}(X(s)|X(0)))$. In the simulation results described below the time-step $\Delta\tau = 6$ hours and $M = 3$ so that there are $\tau/6 \approx 4.5$ corrections in every round of the limit cycle. The effect of less frequent correction is studied in S2 Sect. 5.

Comparisons to other simulation algorithms

We compared the pcLNA simulation algorithm in terms of both CPU time and precision of the approximation with some of the most important alternatives, the tau-leap method and integration of the chemical Langevin equation (CLE) method. Exact simulations produced by

the SSA are also used as a reference for the comparison.

For the tau-leap simulation we use the algorithm described in [3], which is a refinement of the original tau-leap method first proposed in [2]. In [3] the authors suggest an optimal method to compute the largest possible time step such that the leap condition is still approximately satisfied. The leap condition ensures that the state change in any time step is small enough so that no rate function will experience a macroscopically significant change in its value. The error of this approximation is controlled by a parameter ϵ. For integrating the CLE described in [4], we use the classical Euler method (see [32]) with a fixed time step Δt. The integration of the CLE can be done using methods that include higher-order terms in the integration and this has been shown to improve the speed of implementation in low-dimensional systems, albeit with a cost in the complexity of the algorithm (see [33, 34]). However, we are not aware of any implementations of these methods for high-dimensional systems such as those considered here, or of comparisons of such methods to the Euler method for such problems.

For both the tau-leap and the CLE approximation we explored different values of their parameters ϵ and Δt to attain a good balance of precision and CPU time. Here we present the results for the largest values of both ϵ and Δt, hence smallest CPU time, which attain similar performance with the pcLNA algorithm in terms of precision. If little improvement could be achieved in terms of precision by lowering either ϵ or Δt, the larger values are preferred.

The algorithms are implemented for a fixed time-interval (8.5 times the period of the limit cycle of the system) and the comparison is made at 8 fixed time points using the KS distances between the empirical distribution of each algorithm and the empirical distribution derived using the SSA simulations. Note that for all approximation algorithms considered here, the probability of generating negative populations is non-zero and there are a number of methods for dealing with this. Our simple approach is described in S1 Sect. 13.

Figure 8 displays the median CPU times for a single trajectory simulation in $t \in [0, 8.5\tau]$, under the competing approaches for different system sizes, $\Omega = 300$, 1000, 3000, along with a comparison in terms of precision for $\Omega = 300$ (see Figs G & H in S2 Appendix for $\Omega = 1000$ & 3000). A sample of size $R = 2000$ is produced for each algorithm. We see that the precision of all approximation methods is fairly similar, with their empirical distributions almost indistinguishable to exact simulations. In terms of CPU times, we see that the SSA is much slower than the other algorithms particularly for large system sizes. The tau-leap offers some improvement to the CPU times but this is relatively small compared to the CLE approximation and especially the pcLNA algorithm. One reason for this relatively small improvement for the tau-leap algorithm is the stiffness of the considered system, a property that is however very common in biological systems and it is known to slow down the tau-leap method by requiring small values of the ϵ parameter to ensure that the leap condition is satisfied and hence the approximation is fairly precise. Note that for similar reasons, a small Δt was necessary to achieve good precision with the CLE approximation.

As we can see in Fig 8, the pcLNA algorithm is about 24 times faster than the CLE approximation, tens to hundreds of times faster than the tau-leap and hundreds to thousands of times faster than the SSA. In our simulations, it took 0.4sec to derive this long-time trajectory,

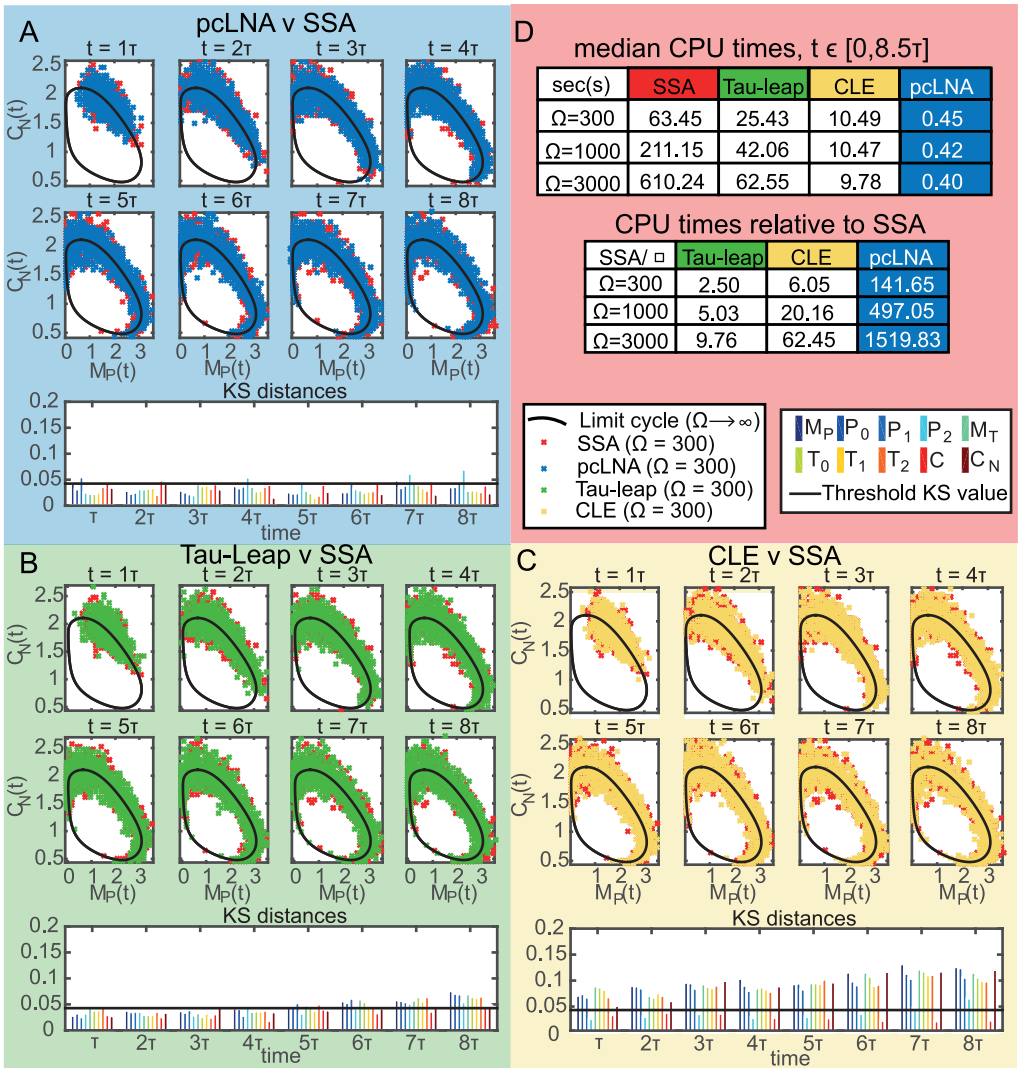

Fig 8. Comparison between pcLNA, tau-leap and CLE simulation algorithms for the *Drosophila* circadian clock. Panels (A), (B) and (C) contain the samples (in concentrations, units as in Fig 1) produced by respectively the pcLNA, tau-leap ($\epsilon = 0.002$) and CLE ($\Delta t = 0.002$) algorithms and the exact simulation (SSA) at time-points $t = \tau, 2\tau, \ldots, 8\tau$, for $\Omega = 300$, along with the KS distances between the empirical distributions of each approximation and the SSA for each system variable (coloured bars; variable names as in Fig 3). Panel (D) provides the median CPU times for a single trajectory run in the time-interval $[0, 8.5\tau]$ under the different simulation algorithms along with the ratio of the median CPU time under SSA and each approximation algorithm.

which means that in about 7 minutes one can generate more than 1000 trajectories of this large system over a long-time compared to about 2.7 hours with CLE approximation and much

longer times for the other methods. Therefore, the pcLNA offers a substantial improvement in CPU times compared to standard approaches in simulating oscillatory systems without compromising the precision of the simulation substantially.

Perhaps more importantly, this pcLNA simulation algorithm has the advantage of being based on an analytical framework that allows calculation of some key distributions. Therefore, it enables more rigorous methods for assessing accuracy and robustness.

Analysis and inference of oscillatory systems using pcLNA

The derivation of analytical expressions of the transversal distributions allows us to analyse various aspects of the stochastic behaviour of these systems that can possibly involve a large number of variables and parameters. Here we illustrate the use of pcLNA transversal distributions to perform such an analysis. We begin by describing the pcLNA joint distribution of multiple intersections to possibly different transversal sections on the limit cycle and then discuss Fisher information, sensitivity analysis and estimation by Kalman filtering.

Joint distribution on multiple transversals

Consider q phase states of the limit cycle $x_i = g(t_i)$, $i = 1, \ldots, q$, on γ where $0 \leq t_1 < t_2 < \ldots < t_q < \tau$. If $X(t)$ is a stochastic trajectory, we consider how it meets the transversal sections at the x_i as t increases using the lifted phase function $\bar{\varphi}_X$. We can talk of the times when $G_N(X(t))$ first takes the phase x_i during the r-th revolution of $G_N(X(t))$ around γ. Using $\bar{\varphi}_X$ to define the points $Q^{(r)}_{g(t)}$ we have that it first meets $\mathcal{S}_{x_i}$ in $Q^{(1)}_{x_i}$ for $i = 1, \ldots, q$. If $i < q$ then we let $Q^{(k)}_{x_{i+1}}$ denote the first point in $\mathcal{S}_{x_{i+1}}$ that X meets after it leaves $Q^{(k)}_{x_i}$. If $i = q$ then the next transversal it meets is $\mathcal{S}_{x_0}$ and the intersection point is $Q^{(k+1)}_{x_1}$. In this way (see Fig 9) we derive a sequence of intersection points $\underline{Q} = Q^{(1)}_{x_1}, \ldots, Q^{(1)}_{x_q}, Q^{(2)}_{x_1}, \ldots, Q^{(2)}_{x_q}, \ldots, Q^{(m)}_{x_1}, \ldots, Q^{(m)}_{x_q}$.

We shall be interested in the distribution

$$P(\underline{Q}|X(t_0)) = P(Q^{(1)}_{x_1}, \ldots, Q^{(m)}_{x_q}|X(t_0)). \tag{5}$$

Remarkably, in our approximation, this distribution is MVN with a covariance matrix whose inverse has a simple tridiagonal form in terms of the drift and diffusion matrices coming from the LNA (S1 Sects. 9.2 & 9.3).

The fact that the above transversal distributions are MVN allows us to analytically compute the Fisher Information matrix and associated quantities that can be used to perform a stochastic sensitivity analysis of oscillatory systems.

Fisher Information

Fisher Information quantifies the information that an observable random variable carries about an unknown parameter θ. If $P(X, \theta)$ is a probability distribution depending on parameters θ,

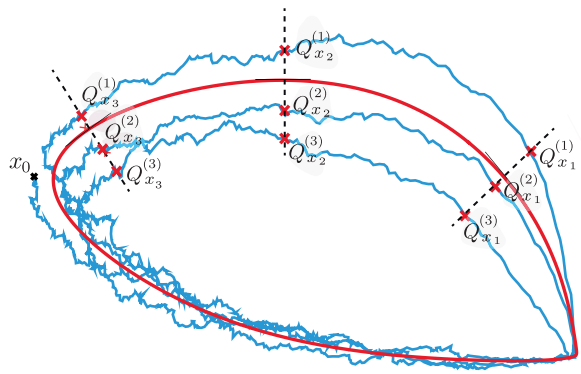

Fig 9. The sequence Q in two-dimensions. The stochastic trajectory $X(t)$ (blue line), initiated from x_0, intersects each of the transversal sections $\mathcal{S}_{x_1}, \mathcal{S}_{x_2}$ and $\mathcal{S}_{x_3}$ (dashed lines) of the limit cycle (red solid line) three times, following a path $Q^{(1)}_{x_1} \to Q^{(1)}_{x_2} \to Q^{(1)}_{x_3} \to Q^{(2)}_{x_1} \to Q^{(2)}_{x_2} \to Q^{(2)}_{x_3} \to Q^{(3)}_{x_1} \to Q^{(3)}_{x_2} \to Q^{(3)}_{x_3}$.

the Fisher Information Matrix (FIM) $I = I_P$ has entries

$$I_{ij} = E\left[\frac{\partial \ell}{\partial \theta_i}\frac{\partial \ell}{\partial \theta_j}\right] = -E\left[\frac{\partial^2 \ell(\theta; X)}{\partial \theta_i \partial \theta_j}\right], \tag{6}$$

where $\ell = \log P$, and θ_i and θ_j are the ith and jth components of the parameter θ. If P is MVN with mean and covariance $\mu = \mu(\theta)$ and $\Sigma = \Sigma(\theta)$ then

$$I_{ij} = \frac{\partial \mu}{\partial \theta_i}^T \Sigma^{-1} \frac{\partial \mu}{\partial \theta_j} + \frac{1}{2}\mathrm{tr}\left(\Sigma^{-1}\frac{\partial \Sigma}{\partial \theta_i}\Sigma^{-1}\frac{\partial \Sigma}{\partial \theta_j}\right). \tag{7}$$

The FIM measures the sensitivity of P to a change in parameters in the sense that

$$D_{KL}(P(\cdot, \theta + \delta\theta), P(\cdot, \theta)) = \frac{1}{2}\delta\theta^T I_P \delta\theta + \mathrm{O}(||\delta\theta||^3)$$

where D_{KL} is the Kullback-Leibler divergence. The significance of the FIM for sensitivity and experimental design follows from its role in (6) as an approximation to the Hessian of the log-likelihood function at a maximum. Assuming non-degeneracy, if θ^* is a parameter value of maximum likelihood there is a $s \times s$ orthogonal matrix V such that, in the new parameters $\theta' = V \cdot (\theta - \theta^*)$,

$$\ell(\theta) \approx \ell(\theta^*) - \sum_i \sigma_i^2 \, {\theta'_i}^2 .$$

for θ near θ^*. From these facts it follows that the σ_i^2 are the eigenvalues of the FIM and that the matrix V diagonalises it. If we assume that the σ_i are ordered so that $\sigma_1^2 \geq \cdots \geq \sigma_s^2$ then it follows that near the maximum the likelihood is most sensitive when θ'_1 is varied and least sensitive when θ'_s is. Moreover, σ_i is a measure of this sensitivity.

The theory of optimal experimental design is based on the idea of trying to make the σ_i decrease as slowly as possible so that the likelihood is as peaked as possible around the maximum, thus maximising the information content of the experimental sampling methods. Various criteria for experimental design have been proposed including D-optimality that maximises the determinant of the FIM and A-optimality that minimise the trace of the inverse of the FIM [6]. Diagonal elements of the inverse of FIM constitute a lower-bound for variance of any unbiased estimator of elements of θ (Cramer-Rao inequality). However, for the systems we consider here the σ_i typically decrease very fast and there are many of them. Thus, in general, criteria based on a single number are more likely to be of less use than consideration of the set of σ_i as a whole.

Calculation of the FIM for stochastic systems using the LNA has been carried out in [22] but only for small systems and short times where the LNA is accurate. It is notable that the pcLNA enables one to do such sensitivity analysis for large systems and large times. As an example, we analyse the stochastic behaviour of the *Drosophila* circadian clock based on the limit distribution $P(\underline{Q}|Q_0)$ when

$$\underline{Q} = Q^{(1)}_{x_0}, Q^{(1)}_{x_1}, Q^{(2)}_{x_0}, Q^{(2)}_{x_1}, \ldots, Q^{(m)}_{x_0}, Q^{(m)}_{x_1}$$

where $x_0 = g(t_0)$ and $x_1 = g(t_1)$ are chosen so that t_0 is the time of the peak of *per* mRNA M_P, and t_1 is the peak of the nuclear complex of PER and TIM proteins C_N. We compute the Fisher Information of the distribution $P(\underline{Q}|Q_0)$ using the closed form expression (S1 Sect. 9.3) for this distribution. As we can see in Fig 10(A) the eigenvalues of the Fisher Information matrix decay exponentially, with a sharp decline followed by a slower decrease. This reveals that the influential directions in the parameter space of the system are much less than its total dimension and that only a few parameters appear to be most influential. The eigenvectors associated with the two largest eigenvalues of Fisher Information matrix (see Fig. 10(B)) have large entries only for the parameters k_{dn} (PER-TIM complex nuclear degradation), k_d (*per* mRNA linear degradation), k_2 (PER-TIM complex transportation to cytosol), v_{st} (*tim* mRNA transcription), k_{ip} (*per* mRNA Hill coefficient) and k_{it} (*tim* mRNA Hill coefficient).

The exponential decrease of the eigenvalues is typical of tightly coupled deterministic systems [25–31], but has to our knowledge not been demonstrated before for stochastic systems. It has important consequences. For example, it tells us that only a few parameters can be estimated efficiently from time-series data unless the system is perturbed in some way to get complementary data and that there will be identifiability problems that can be analysed using the FIM. It can also be used to design experiments by considering the FIM of a combination of models including one for the proposed new experiment, choosing the new experiment so as to optimally alleviate the decline of the eigenvalues.

Sensitivity analysis for stochastic systems

The fact that we can calculate the Fisher Information allows a new approach to sensitivity analysis for stochastic systems. Anderson [23] and Srivastava et al. [24] also perform sensitivity

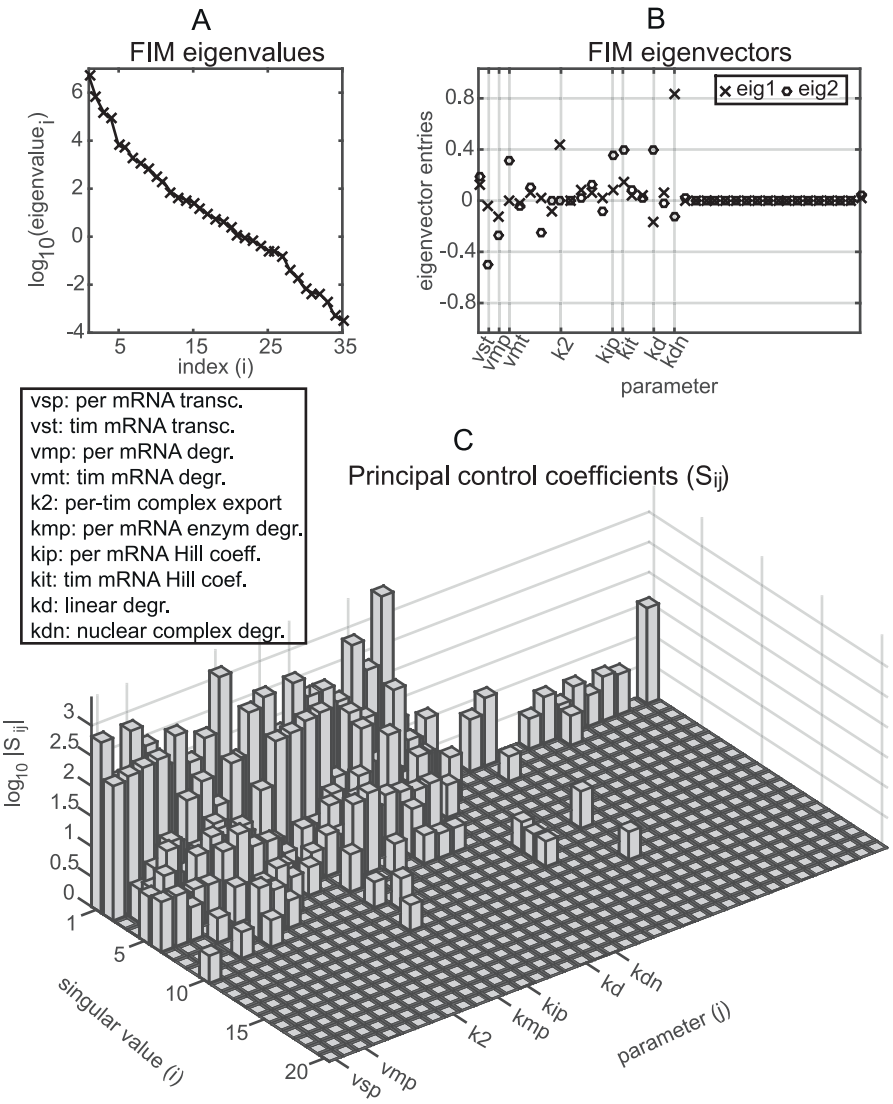

Fig 10. Fisher Information & Stochastic sensitivity analysis of the transversal distribution of the *Drosophila* circadian clock system at the times of the peaks of *per* mRNA and the nuclear complex of PER and TIM proteins. (A) The logarithm of the eigenvalues of the Fisher Information Matrix (FIM). (B) The entries/weights of the eigenvectors corresponding to the 2 largest eigenvalues of FIM (C) The largest principal control coefficients S_{ij}. Small values are reduced to 0.

analysis for small stochastic systems (up to 4 species and 8 reactions) in which they calculate the dependence of certain summary functions or statistics at one or more times to individual parameters. Our approach is different in that we use the fact that our distributions of interest are MVN and measure the change in the distribution of the system state at any given set of

phases without recourse to any summary function and, moreover, this change is calculated for any combination of parameter variations. A major difference is our use of SVD below to find a basis of mean-covariance space using the principal components that enables us to decompose these changes into different orthogonal directions that pick out the important and unimportant directions. The approach can also be formulated for the wider class of exponential families i.e. distributions that admit a representation of the form

$$P(x) = \exp\left\{C(x) + \sum_i \theta_i F_i(x) - \varphi(\theta)\right\}$$

in terms of functions $C, F_1, \ldots F_m$ of the state variable x and a function φ of the parameters θ.

We consider a family of probability distributions $P(X, \theta)$ which we assume are MVN with mean $\mu(\theta)$ and covariance matrix $\Sigma(\theta)$ depending on the parameters θ. We show that there is a natural matrix of sensitivities S_{ij} associated with such a system. These are system-global in that they look at how all components of the systems change with parameters. They also have an intimate relationship with Fisher information. Note that these results are not restricted to the transversal distributions derived in previous sections but apply more generally to any MVN distribution with mean $\mu = \mu(\theta)$ and covariance matrix $\Sigma = \Sigma(\theta)$ parameterised by a s-dimensional vector θ, $s \geq 1$.

As is well-known in Information Geometry, the set of multivariate normal distributions MVN^n on $\mathbb{R}^n$ can be given the structure of a Riemannian manifold in which the Riemannian metric is given by the line element

$$ds^2 = d\mu^T \Sigma^{-1}\, d\mu + (1/2)\mathrm{tr}\{(\Sigma^{-1} d\Sigma)^2\}.$$

Points in MVN^n are denoted by $\Theta = (\mu, \Sigma)$ where μ is the mean and Σ the covariance matrix. The corresponding inner product in the tangent space at $\Theta_0 = (\mu, \Sigma)$ is given by

$$\langle \delta\Theta_1, \delta\Theta_2 \rangle_{\Theta_0} = \delta\mu_1 \Sigma^{-1} \delta\mu_2 + \frac{1}{2}\mathrm{tr}\left(\Sigma^{-1}\delta\Sigma_1 \Sigma^{-1}\delta\Sigma_2\right) \tag{8}$$

where $\delta\Theta_j = (\delta\mu_j, \delta\Sigma_j)$, $j = 1, 2$.

In calculating the FIM we have to determine the partial derivatives $\partial\mu/\partial\theta_i$ and $\partial\Sigma/\partial\theta_i$. The derivative M of the mapping $\theta \to (\mu(\theta), \Sigma(\theta))$ at a parameter value θ_0 is given by

$$M \cdot \delta\theta = \left(\sum_i \frac{\partial\mu}{\partial\theta_i}\delta\theta_i, \sum_i \frac{\partial\Sigma}{\partial\theta_i}\delta\theta_i\right)$$

where the derivatives are calculated at θ_0.

We can regard M as a linear mapping between the parameter space $\mathbb{R}^s$ and MVN^n with the inner product given in equation (8). We can then prove (S1 Sect. 11) that we can find s

orthonormal vectors V_i spanning the parameter space $\mathbb{R}^s$, s orthonormal vectors U_i in the space MVN^n and numbers $\sigma_1 \geq \cdots \geq \sigma_s \geq 0$ such that

$$MV_i = \sigma_i U_i, \qquad i = 1, \ldots, s. \tag{9}$$

Note that the orthonormality of the U_i is with respect to the inner product $\langle \cdot, \cdot \rangle_{\Theta_0}$. The eigenvalues of the FIM F are the squares of the σ_i because with respect to the standard inner product on θ-space and $\langle \cdot, \cdot \rangle_{\Theta_0}$ on MVN^n the adjoint M^* satisfies $M^*M = F$ (S1 Sect. 11).

If we let $U_i = (U_i^\mu, U_i^\Sigma)$ denote the decomposition of U_i into μ and Σ components, then the following key property follows from (9): if $\delta\theta$ is any change of parameters, the change in μ and Σ is given by

$$\begin{aligned} \delta\mu &= \sum_i U_i^\mu \left(\sum_j S_{ij} \delta\theta_j \right) + \mathrm{O}(||\delta\theta||^2) \\ \delta\Sigma &= \sum_i U_i^\Sigma \left(\sum_j S_{ij} \delta\theta_j \right) + \mathrm{O}(||\delta\theta||^2) \end{aligned} \tag{10}$$

where $S_{ij} = \sigma_i V_{ji}$.

One can define other sensitivities in a similar way but using a different orthogonal basis of MVN^n, but the above S_{ij} satisfy an important optimality condition explained in S1 Sect. 11 which asserts that the basis U_i and the corresponding sensitivities S_{ij} are optimal for capturing as much sensitivity as possible in the low order principal components U_i.

In view of this we call the S_{ij} the *principal control coefficients*. Note that the role of the S_{ij} as sensitivities is seen from the following relation which follows from (10),

$$\|\delta\Theta\| = \|S \cdot \delta\theta\| + O(\|\delta\theta\|^2). \tag{11}$$

These sensitivities are relatively easy to calculate using the information in S1 Sect. 11. In Fig 10(C) we show the S_{ij} for the transversal distribution of the *Drosophila* circadian clock at the times of the peak of *per* mRNA and the peak of the nuclear complex of PER and TIM proteins. As we can see, because $S_{ij} = \sigma_i V_{ji}$ the coefficients rapidly decrease with the singular values σ_i, while a few parameters, similar to those with large eigenvector entries, have high coefficients.

Calculating likelihoods via a pcLNA Kalman Filter

The likelihood function of a set of time-series observations of a system can be used for parameter estimation, hypothesis testing and other forms of statistical inference. For example, one may wishes to use the likelihood function to estimate parameters of a biological system. Although there is no elegant formula for

$$P(\underline{X}|X(t_0)) = P(X(t_1), \ldots, X(t_m)|X(t_0))$$

similar to that for $P(\underline{Q}|X(t_0))$ above, we can efficiently calculate it. To do this we derive a Kalman Filter for the pcLNA that is a modification of the Kalman Filter associated with the LNA [35]. This can be used to compute the likelihood function $L(\theta; \underline{\hat{X}})$ of the system parameters θ with respect to observations $\underline{\hat{X}} = \left(\hat{X}(t_0), \hat{X}(t_1), \ldots, \hat{X}(t_N)\right)$ recorded at N times $t_0, t_1, \ldots, t_N$ that are noisy linear functions of the species concentrations. This is slightly more general than just calculating $P(\underline{X}|X(t_0))$ because we allow for a measurement equation. The Kalman filter can also be used for forward prediction.

We assume the measurement equation,

$$\hat{X}(t) = BX(t) + \epsilon, \tag{12}$$

relating the observations $\hat{X}(t)$ to the state variables, $X(t)$. Here B is a transformation matrix (often simply removing unobserved species or introducing unknown scalings) and $\epsilon = (\epsilon_1, \ldots, \epsilon_n) \sim MVN(0, \Sigma_\epsilon)$ the observational error. The pcLNA likelihood can be decomposed as the product

$$L(\theta; \underline{\hat{X}}) = P(\hat{X}(t_0); \theta) \prod_{i=1}^{n} P(\hat{X}(t_i)|\hat{X}(t_{i-1}); \theta).$$

The pcLNA Kalman Filter algorithm, which we describe in more detail in S1 Sect. 15, uses a recursive algorithm for computing the terms in $L(\theta; \underline{\hat{X}})$. The algorithm proceeds by iteration $i = 1, 2, \ldots$ and uses Bayes rule to derive the posterior distributions $(X(t_{i-1})|\hat{X}(t_{i-1}))$ and a phase correction to obtain $g(s_{i-1}) = G_N(\mu^*(t_{i-1}))$ and the corrected noise distribution $(\kappa(s_{i-1})|\hat{X}(t_{i-1}))$. The LNA transition equation (S1 Eq. (4.2)) is then used to derive the distribution of $(\xi(t_i)|\hat{X}(t_{i-1}))$ and the LNA ansantz equation (3) to obtain $(X(t_i)|\hat{X}(t_{i-1}))$. The measurement equation (12) is finally used to obtain the $(i+1)$th term of the likelihood function $P(\hat{X}(t_i)|\hat{X}(t_{i-1}))$ before proceeding to the next iteration. All the distributions obtained in this way are MVN with easily computable parameter values.

If the observations are recorded in short time intervals, the phase correction can be omitted in some steps, in which case the algorithm proceeds as in [35]. Computational methods such as those described in [32] and [35] can then be used to perform likelihood-based statistical inference.

Methods

All computations have been carried out using MATLAB Release 2016b, The MathWorks, Inc., Natick, MA, USA. In particular, the empirical CDF plots, (q-q) plots, histograms, smooth probability density functions and KS distances are derived using the ecdf, qqplot, histogram, ksdensity, kstest functions of MATLAB and Statistics Toolbox. The computations for the SSA, tau-leap, integration of diffusion and pcLNA simulation algorithms, and the computation of Fisher Information and principal control coefficients for the sensitivity analysis were performed using the PeTTSy software which is discussed in S1 Sect. 14 and is freely

available at `http://www2.warwick.ac.uk/fac/sci/systemsbiology/research/software/`. Further details concerning methods are given in S1 Sect. 16.

3 Discussion

We present a comprehensive treatment of stochastic modelling for large stochastic oscillatory systems. Practical algorithms for fast long-term simulation and likelihood-based statistical inference are provided along with the essential tools for a more analytical study of such systems.

There is considerable scope for future work in various directions. We expect that these results can be extended to a broader class of systems including those that are chaotic in the $\Omega \to \infty$ limit. Our approach should provide the opportunity to develop new methodology for parameter estimation, likelihood-based inference and experimental design in such systems. Finally, there is currently much interest in information transfer and decision-making in signaling systems and our methods provide new tools with which to tackle problems in this area.

If system biologists are to reliably use complex stochastic models to provide robust understanding it is crucial that there are analytical tools to enable a rigorous assessment of the quality and selection of these models and their fit to current biological knowledge and data. Our aim in this paper is to contribute to that but the results should be of much broader interest.

Supporting Information

S1 Appendix. Technical Details. In this note we give further details about the mathematical underpinnings of the pcLNA methods discussed in the main paper.

S2 Appendix. *Drosophila* circadian clock system. In this note we give details about the *Drosophila* circadian clock and use this system to illustrate further the accuracy of distributions and simulations discussed in the main paper.

S3 Appendix. Brusselator and NF-κB systems. In this note we give details about the deterministic and stochastic models of the Brusselator and NF-κB systems and use them to illustrate further the results described in the main paper.

Acknowledgments

This research was funded by the BBSRC Grant BB/K003097/1 (Systems Biology Analysis of Biological Timers and Inflammation). DAR was also supported by funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 305564.

References

[1] Gillespie DT. Exact stochastic simulation of coupled chemical reactions. The Journal of Physical Chemistry. 1977 Dec;81(25):2340–2361.

[2] Gillespie DT. Approximate accelerated stochastic simulation of chemically reacting systems. The Journal of Chemical Physics. 2001 Jul;115(4):1716–1733.

[3] Gillespie DT, Petzold LR. Improved leap-size selection for accelerated stochastic simulation. The Journal of Chemical Physics. 2003 Oct;119(16):8229–8234.

[4] Gillespie DT. The chemical Langevin equation. The Journal of Chemical Physics. 2000 Jul;113(1):297–306.

[5] van Kampen NG. Stochastic Processes in Physics and Chemistry, Third Edition. Amsterdam: Elsevier. Boston and London: Elsevier; 2007.

[6] Kurtz TG. Limit Theorems for Sequences of Jump Markov Processes Approximating Ordinary Differential Processes. Journal of Applied Probability. 1971 Jun;8(2), 344–356.

[7] Kurtz TG. Approximation of Population Processes. Regional Conference Series in Applied Mathematics vol. 36. Society for Industrial and Applied Mathematics. 1981.

[8] Hayot F, Jayaprakash C. The linear noise approximation for molecular fluctuations within cells. Physical Biology. 2004 Dec;1(3-4):205-10.

[9] Boland RP, Galla T, McKane AJ. How limit cycles and quasi-cycles are related in systems with intrinsic noise. Journal of Statistical Mechanics: Theory and Experiment. 2008 Sep;P09001.

[10] Koeppl H, Hafner M, Ganguly A, Mehrotra A. Deterministic characterization of phase noise in biomolecular oscillators. Physical Biology. 2011 Oct;8(5):055008.

[11] Tomita K, Ohta T, Tomita, H. Irreversible Circulation and Orbital Revolution: Hard Mode Instability in Far-from-Equilibrium Situation. Progress of Theoretical Physics. 1974 Dec;52(6):1744–1765.

[12] Scott M, Ingalls B, Kærn M (2006). Estimations of intrinsic and extrinsic noise in models of nonlinear genetic networks. Chaos. 2006 Jun;16(2):026107.

[13] Ito Y, Uchida K. Formulas for intrinsic noise evaluation in oscillatory genetic networks. Journal of Theoretical Biology. 2010 Aug;267(2):223–234.

[14] Schwabedal JTC, Pikovsky A. Phase Description of Stochastic Oscillations. Physical Review Letters. 2013 May;110(20):204102.

[15] Thomas PJ and Lindner B. Asymptotic Phase for Stochastic Oscillators. Physical Review Letters. 2014 Dec;113(25):254101.

[16] Gonze D, Halloy J, Leloup JC, Goldbeter A. Stochastic model for circadian rhythms: effect of molecular noise on periodic and chaotic behaviour. Comptes Rendus Biologies. 2003 Apr;326(2):189–203.

[17] Ashall L, Horton CA, Nelson DE, Paszek P, Harper CV, Sillitoe K, Ryan S, Spiller DG, Unitt JF, Broomhead DS, Kell DB, Rand DA, Sée V, White MRH. Pulsatile stimulation determines timing and specificity of NF-kappa B-dependent transcription. Science. 2009 Apr;324(5924);242–6.

[18] Leloup JC, Goldbeter A. A model for circadian rhythms in Drosophila incorporating the formation of a complex between the PER and TIM proteins. J Biol Rhythms 1998 Feb;13(1):70-87.

[19] Hartman P. Ordinary differential equations. New York: John Wiley & Sons, Inc. London and Sydney: Wiley; 1964.

[20] Guckenheimer J, Holmes PJ. Nonlinear Oscillations, Dynamical Systems and Bifurcations of Vectorfields. New York: Springer-Verlag, Inc. Berlin, Heidelberg and Tokyo: Springer-Verlag; 1983

[21] Gupta A, Hepp B, Khammash M. Noise Induces the Population-Level Entrainment of Incoherent, Uncoupled Intracellular Oscillators. Cell Systems. 2016 3(6):521—531.e13.

[22] Komorowski M, Costa MJ, Rand DA Stumpf MPH. Sensitivity, robustness and identifiability in stochastic chemical kinetics models. Proceedings of the National Academy of Sciences of the United States of America. 2011 May;108(21):8645–8650.

[23] Anderson DF. An Efficient Finite Difference Method for Parameter Sensitivities of Continuous Time Markov Chains. ArXiv e-prints. 2011 Sep.

[24] Srivastava R, Anderson DF, Rawlings JB. Comparison of finite difference based methods to obtain sensitivities of stochastic chemical kinetic models. The Journal of Chemical Physics. 2013;138(7):074110.

[25] Brown KS, Sethna JP. Statistical mechanical approaches to models with many poorly known parameters. Phys. Rev. E. 2003 Aug;68(2 Pt 1):021904.

[26] Brown KS, Hill CC, Calero GA, Myers CR, Lee KH, Sethna JP, Cerione RA. The statistical mechanics of complex signaling networks: nerve growth factor signaling. Physical biology. 2003 Dec;1(3):184–195.

[27] Gutenkunst RN, Waterfall JJ, Casey FP, Brown KS, Myers CR, Sethna JP. Universally sloppy parameter sensitivities in systems biology models. PLoS Computational Biology. 2007 Oct;3(10):1871–78.

[28] Rand DA, Shulgin BV, Salazar D, Millar AJ. Design principles underlying circadian clocks. Journal of The Royal Society Interface. 2004 Nov;1(1):119–30.

[29] Rand DA, Shulgin BV, Salazar JD, Millar AJ. Uncovering the design principles of circadian clocks: Mathematical analysis of flexibility and evolutionary goals. Journal of Theoretical Biology. 2006 Feb;238(3):616–35.

[30] Rand DA. Mapping the global sensitivity of cellular network dynamics: Sensitivity heat maps and a global summation law. Journal of The Royal Society Interface. 2008 Aug;5 Suppl 1:S59–69.

[31] Waterfall JJ, Casey FP, Gutenkunst RN, Brown KS, Myers CR, Brouwer PW, Elser V, Sethna JP. Sloppy-Model Universality Class and the Vandermonde Matrix. Phys. Rev. Lett. 2006 Oct;97(15):150601.

[32] Wilkinson DJ. Stochastic modelling for systems biology. Boca Raton: CRC Press/Taylor & Francis; 2012.

[33] Kloeden PE, Platen E, Schurz H. Numerical solution of SDE through computer experiments/ Diskette. Berlin: Springer-Verlag; 1997.

[34] Iacus SM. Simulation and Inference for Stochastic Differential Equations: With R Examples. New York, NY: Springer New York:Springer e-books; 2008

[35] Finkenstädt B, Woodcock DJ, Komorowski M, Harper CV, Davis JRE, White MRH, Rand DA. Quantifying intrinsic and extrinsic noise in gene transcription using the linear noise approximation: An application to single cell data. The Annals of Applied Statistics. 2013 Dec;7(4):1960–1982.

Supplementary information S1 for the paper Long-time analytic approximation of large stochastic oscillators: simulation, analysis and inference

Giorgos Minas[1,2] & David A Rand[1,2]

[1] Zeeman Institute for Systems Biology & Infectious Disease Epidemiology Research & [2]Mathematics Institute,

University of Warwick, Coventry CV4 7AL, UK

July 22, 2017

Contents

1 Mathematical details **1**

2 The system size parameter Ω **3**

3 Derivation of the classical deterministic equation **3**

4 The Linear Noise Approximation (LNA) **4**

5 The LNA transversal distribution at a fixed time **5**

6 Distribution of phase fluctuations δt and $g(t) - G_N(X(t))$ **6**

7 The approximation of the LNA transversal distributions **8**

8 Proof of convergence for LNA transversal distributions **9**

8.1 Free-running oscillators. 9

8.2 Extending this result to arbitrary transversals. 11

8.3 Alternative geometrical proof that convergence on $\mathcal{S}_N$ implies convergence on other transversals $\mathcal{S}$. 11

8.4 Entrained forced oscillators. 11

9 Analytic expression for pcLNA transversal distributions **12**

9.1 Fixed point distribution. 13

9.2 The joint distribution of multiple transversals 13

9.3 Precision matrix for $P(\underline{Q}|X(t_0))$. 14

10 Other transversal sections **14**

11 Details behind the sensitivity analysis. **15**

11.1 Deducing equation (S11.4) from SVD. 17

12 Multiple crossings in each rotation **17**

13 Negative populations **21**

14 PeTTSy **21**

15 The pcLNA Kalman Filter **23**

16 Details related to various computations **24**

Abstract

We refer to the paper "Long-time analytic approximation of large stochastic oscillators: simulation, analysis and inference" by **I**. In this note we give further details about the mathematical underpinnings of the pcLNA methods discussed in **I**.

1 Mathematical details

When we refer to a function, mapping or vectorfield as being smooth we mean that it has C^2 dependence on the relevant variables. We are assuming throughout that the vector field F introduced in **I** is C^2 and that the limit cycle γ given by $g(t)$ is a stable elementary limit cycle in the sense we describe now. Here $x \in \mathbb{R}^n$.

If $g(t)$ is the periodic solution of interest (with period τ), consider the linear differential equation $dC/dt = J(t)\,C$ where $J(t)$ is the $n \times n$ Jacobian matrix whose ijth entry is $\partial F_i/\partial x_j$ evaluated at $g(t)$. Let $C(s,t)$ be the solution of this equation with initial condition $C(s,s) = I_n$ the identity $n \times n$ matrix. If $C(t) = C(s,t)$ then $C(t)$ has a representation of the form $C(t) = Z(t)e^{R(t-s)}$ where $Z(t+\tau) = Z(t)e^{R\tau}$ (Theorem 6.1 of Chapter IV [5]). If $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of R then $e^{t\lambda_1}, \ldots, e^{t\lambda_1}$ are the eigenvalues of e^{Rt}. As such they are only determined modulo $2\pi i/\tau$. Modulo this indeterminacy, they are called the *characteristic exponents* of the limit cycle. The characteristic exponents are independent of the choice of the initial time s.

We assume that all limit cycles γ considered in this paper are *elementary and stable* in the following sense. In the case of a limit cycle γ of a free-running oscillator (i.e. when F does not depend directly upon t), we say that γ is *elementary* if $n-1$ of the characteristic exponents have non-zero real parts and the remaining characteristic exponent is zero (modulo $2\pi i/\tau$). Such limit cycles are often called *hyperbolic.* If in addition the $n-1$ non-zero characteristic exponents have negative real part then we say it is *stable and elementary.*

In the case of periodically forced oscillators the limiting differential equation is $\dot{x} = F(t,x)$ and, in this case γ is elementary if all the characteristic exponents have non-zero real parts and it is stable if all have negative real parts.

A number of desirable properties flow from these conditions. Firstly, for elementary and stable limit cycles there is a neighbourhood of γ in $\mathbb{R}^n$ such that any solution of $\dot{x} = F(t,x)$ with its initial condition in the neighbourhood converges exponentially fast to γ (Theorem 11.1. Chap. IX [5]). Secondly, suppose F depends smoothly on parameters θ so that $F = F_\theta$. Then if $g_{\theta_0}(t)$ is a elementary limit cycle for $F = F_{\theta_0}$, there is a unique family g_θ defined for θ near θ_0 which depends smoothly upon θ such that g_θ is an elementary limit cycle of $F = F_\theta$ (Theorem 2.3, Chap. XII [5]).

Almost all the results described in the main paper that are for normal transversals $N(x)$ hold for arbitrary transversals $\mathcal{S}_x$. By a transversal section through $x \in \gamma$ we mean a $(n-1)$-dimensional linear hyperplane $\mathcal{S}_x$ containing x and transversal to the tangent vector, $F(x)$, to γ at x. A particular example is the hyperplane normal to γ at x. A transversal system is a family $\mathcal{S}_{g(t)}$ of transversal sections which vary smoothly with t in the sense that the unit normal vector to $\mathcal{S}_{g(t)}$ varies smoothly with t. Throughout this document we consider this more general situation.

A transversal system defines a mapping G of a neighborhood of γ onto γ where if $X \in \mathcal{S}_x$ then $G(X) = x \in \gamma$. In cases where $X(s)$ lies in more than one transversal sections, $\mathcal{S}_{x(t')}$, $t' = t_1, t_2, \ldots$, then $G(X(s)) = x(t)$ with $t = \min_i |t_i - s|$ the closest time to s. We denote this mapping for the normal transversal system by G_N.

An adapted coordinate system $\mathcal{C}_{g(t)}$ at a point $g(t)$ on γ is one determined by a set of orthonormal basis vectors $e_1(t), \ldots, e_n(t)$ with $e_1(t)$ the unit normal vector to $\mathcal{S}_{g(t)}$ and the vectors $e_2(t), \ldots, e_n(t)$ forming an orthonormal basis of $\mathcal{S}_{g(t)}$. If these are defined for t in some interval in $\mathbb{R}$ then we always assume that the $e_i(t)$ have smooth (i.e. C^2) dependence upon t. It is important that the coordinates are defined by an orthonormal basis in the original coordinates because this effectively preserves the covariance matrix in the sense that a covariance matrix V in the adapted coordinates is UVU^T in the original coordinates with U a real orthogonal

matrix. In particular, the eigenvalues are preserved.

2 The system size parameter Ω

It is very common to consider stochastic systems involving a system size parameter Ω which is a parameter that occurs in the intensities of the reactions $w_j(Y(t))$ [8, 9]. The precise description of this parameter depends on the system and it governs the size of the fluctuations and therefore the size of the jumps. In population models it might be considered to be of the same order of magnitude as the total population size while in chemical systems a natural choice is to use molar concentrations and therefore regard Ω as Avogadro's number in the appropriate molar units (e.g. nM^{-1}) multiplied by the volume of the reacting solution in appropriate units (e.g. in litres (L)). In the *Drosophila* circadian clock model we consider it has units L/nM.

It can be interpreted as being the same order of magnitude as the mean size of the molecular populations. When the system and the systems size parameter satisfy the LNA Ansatz $Y(t)/\Omega = x(t) + \xi(t)/\sqrt{\Omega}$ (see Sect. S4) then $Y(t) = \mathrm{O}(\Omega)$ and the fluctuations are $\mathrm{O}(\sqrt{\Omega})$.

In our examples, the rates $w_j(Y(t))$ depend upon Ω as $w_j(Y) = \Omega u_j(Y/\Omega)$ as used in the proof by Kurtz for the convergence of the LNA as $\Omega \to \infty$ [8]. For example, Hill functions in our rate equations have the form $w_H = c\Omega N^h/(N^h + (k\Omega)^h)$ where N stands for the relevant population numbers and c and k are parameters. Therefore, if $n = N/\Omega$, $w_H = c\Omega n^h/(n^h + k^h)$. Similarly, degradation and binding reaction rates have the form $w_{\mathrm{deg}} = cN$ and $w_{\mathrm{bind}} = cN_1N_2/\Omega$ so that $w_{\mathrm{deg}} = c\Omega n$ and $w_{\mathrm{bind}} = c\Omega n_1 n_2$.

In a certain sense the system size parameter is just a mathematical convenience to enable the study of the dependence of stochastic fluctuations upon system size. Indeed, the methods we develop in this paper can and should be applied to systems that do not involve a system size parameter but then it will be necessary to ensure that the population sizes achieved in the given system are large enough.

3 Derivation of the classical deterministic equation

We provide a derivation of the macroscopic law of mass action that is similar to [1].

The state of the system at some time t is

$$Y(t) = Y(0) + \sum_j \nu_j Z_j \left(\int_0^t w_j(Y(s))ds \right)$$

with Z_j independent unit Poisson processes corresponding to the j-th reaction channel. The

latter equation can be written in terms of $X = Y/\Omega$ as

$$X(t) = X(0) + \sum_j \nu_j \Omega^{-1} Z_j \left(\int_0^t \Omega u_j(X(s)) ds \right),$$

since $w_j(Y) = \Omega u_j(X)$ (see Section S2). Using the law of large numbers, as $\Omega \to \infty$, we have that

$$x(t) = x(0) + \sum_j \nu_j \int_0^t u_j(x(s)) ds,$$

where here $x(t)$ is the limit of $X(t)$ for $\Omega \to \infty$. Consequently, $x(t)$ satisfies the ordinary differential equation (ODE)

$$\dot{x} = \sum_j \nu_j u_j(x). \tag{S3.1}$$

4 The Linear Noise Approximation (LNA)

The LNA is based on the Ansatz

$$X(t) = \frac{Y(t)}{\Omega} = x(t) + \frac{\xi(t)}{\sqrt{\Omega}} \tag{S4.1}$$

where $x(t)$ is a solution of the limiting $\Omega \to \infty$ system in Eq. (S3.1) which we assume is a stable limit cycle of minimal period $\tau > 0$ given by $x = g(t)$, $0 \leq t \leq \tau$.

Recall the definition of the $n \times n$ matrices $C(s,t)$ introduced in Section S1. In the LNA, the stochastic variable ξ in (S4.1) satisfies

$$\xi(t) = C(t_0, t)\xi(t_0) + \eta(t_0, t), \qquad t_0 < t, \tag{S4.2}$$

where $\eta(t_0, t) \sim \mathrm{MVN}(0, V(t_0, t))$ is multivariate normal with mean 0 and covariance matrix

$$V(t_0, t) = \int_{t_0}^t C(s,t)E(s)E(s)^T C(s,t)^T ds \tag{S4.3}$$

and $E(s)$ the matrix product of the stoichiometry matrix (i.e. the matrix whose columns are the vectors ν_j introduced in **I**) and the square root of the diagonal matrix with main diagonal the reaction rates $u_j(x(s))$. The noise $\xi(t)$ also satisfies the Stochastic Differential Equation,

$$d\xi(t) = J(g(t))\xi(t)dt + E(s)dW(t),$$

where $W(t)$ is a K-dimensional Brownian motion (K: the number of possible reactions). Equation (S4.3) follows from this by using Itô's change of variable formula and the Itô Isometry.

The proof of convergence as $\Omega \to \infty$ of the true distribution of ξ to a multivariate normal distribution with mean and covariance matrix given by (S4.2) and (S4.3) is contained in Chap. 8 of [8].

Henceforth, we write $\zeta \sim \mathrm{MVN}(m, S)$ to mean that a random variable ζ is multivariate normal with mean m and covariance matrix S.

5 The LNA transversal distribution at a fixed time

We now suppose that our initial condition $\xi(t_0)$ is drawn from a MVN distribution P_0 which is supported on the transversal section $S_{g(t_0)}$. This is denoted $\xi(t_0) \sim P_0$. For $t_0 < t_1$, we consider the distribution, $P_{\mathrm{LNA}}(X(t_1)|X(t_1) \in \mathcal{S}_{g(t_1)}, \xi(t_0) \sim P_0)$, of $X(t_1)$ conditional on $X(t_1)$ lying on the transversal sections $\mathcal{S}_{g(t_1)}$ and $\xi(t_0) \sim P_0$.

We have adapted coordinate systems $(y_1, \mathbf{y}_2)$ at $x_1 = g(t_1)$ so that $\mathcal{S}_{x_1}$ is given by $y_1 = 0$. We write the matrices $C(t_0, t_1)$ and $V(t_0, t_1)$ in the latter coordinate system as

$$C(t_0, t_1) = \begin{bmatrix} C_{11} & C_{12} \\ \\ C_{21} & C_{22} \end{bmatrix} \qquad \text{and} \qquad V(t_0, t_1) = \begin{bmatrix} V_{11} & V_{12} \\ \\ V_{21} & V_{22} \end{bmatrix}$$

where $V_{12} = V_{21}^T$. By $X(t_0) \in \mathcal{S}_{x_0}$ we have that $\xi(t_0) = \kappa_0$. Then, under the LNA,

$$\xi(t_1) = (\xi_1(t_1), \xi_{\mathbf{2}}(t_1))^T = C(t_0, t_1) \cdot \xi(t_0) + \eta = (C_{1\cdot}\kappa_0, C_{2\cdot}\kappa_0)^T + \eta$$

where $C_{1\cdot} = [C_{11} \;\; C_{12}]$, $C_{2\cdot} = [C_{21} \;\; C_{22}]$ and $\eta \sim \mathrm{MVN}(0, V(t_0, t_1))$. Consequently, $\xi_{\mathbf{2}}(t_1)$ conditional on $\xi_1(t_1) = 0$ is MVN with mean $\check{C}(t_0, t_1)\kappa_0$ and covariance matrix $\check{V}(t_0, t)$, where $\check{C}(t_0, t_1) = C_{2\cdot} - V_{21}C_{1\cdot}/V_{11}$ and $\check{V}(t_0, t_1) = V_{22} - V_{21}V_{12}/V_{11}$. Thus if $\kappa_0 \sim \mathrm{MVN}(m_0, S_0)$ and $\kappa(t_1)$ denotes $\xi_{\mathbf{2}}(t_1)$ conditional on $\xi_1(t_1) = 0$, then $\kappa(t_1)$ is MVN with mean and covariance matrix respectively given by

$$m_1 = \mathrm{E}(\kappa(t_1)) = \check{C}(t_0, t_1) \cdot m_0, \quad S_1 = \mathrm{Var}(\kappa(t_1)) = \check{C}(t_0, t_1) S_0 \check{C}(t_0, t_1)^T + \check{V}(t_0, t_1) \quad \text{(S5.1)}$$

and $P_{\mathrm{LNA}}(X(t_1)|X(t_1) \in \mathcal{S}_{x_1}, \xi(t_0) \sim MVN(m_0, S_0))$ is $MVN(\mu_1, \Sigma_1)$ with $\mu_1 = \Omega^{-1/2} m_1$ and $\Sigma_1 = \Omega^{-1} S_1$.

6 Distribution of phase fluctuations δt and $g(t) - G_N(X(t))$

In Sect. "Stochastic fluctuations in periods and timing" of **I**, we consider the distribution of the fluctuation δt in the time taken for the lifted phase of a stochastic trajectory to go from a given phase ϕ_1 to a greater one ϕ_2. We showed that for moves from $\phi_1 = 0$ and $\phi_2 = 2\pi q$ for $q = 1, \ldots, 8$ the distribution of the fluctuations is approximately normal with variance increasing with q and that for $\phi_1 = 2(q-1)\pi$ and $\phi_2 = 2q\pi$ for $q = 1, \ldots, 8$ the distributions are again approximately normal with approximately constant variance. We now consider an approximation of the distribution of the fluctuations under the LNA.

In this and the next section we will be considering how to approximate the distribution of first intersections with $\mathcal{S}_{g(t_1)}$ of stochastic trajectories $X(t)$ that start in $\mathcal{S}_{g(t_0)}$. To do this we will consider LNA trajectories with initial condition a given MVN distribution P_0 on $\mathcal{S}_{g(t_0)}$ and their first intersection at time $t = t_1 + \delta t$ with $\mathcal{S}_{g(t_1)}$. In this section we show that the distribution of the variation in time, δt, is approximately normal and calculate approximations of the mean and variance. Thus, we consider the distribution under the LNA of $\xi(t)$ such that $X(t) = g(t) + \Omega^{-1/2}\xi(t) \in \mathcal{S}_{g(t_1)}$ for $t_0 \leq t < t_1 + \tau/2$. We denote this distribution by $P_{\text{LNA}}(\xi(t)|X(t) \in \mathcal{S}_{g(t_1)})$.

To calculate the statistics of the variation in time, δt, for the distribution of the rth intersection with $\mathcal{S}_{g(t_1)}$ of stochastic trajectories $X(t)$ that start in $\mathcal{S}_{g(t_0)}$ one decomposes the trajectory into the part from $\mathcal{S}_{g(t_0)}$ to its first intersection with $\mathcal{S}_{g(t_1)}$ and then the $(r-1)$ passes from $\mathcal{S}_{g(t_1)}$ to itself and uses the fact that the total variation is the sum of the individual ones which are independent random variables. Since we show below that the individual variation are approximately normal it follows that the total is. Therefore in this section we now focus just on first intersections.

We assume that the initial conditions $\xi(t_0)$ have a MVN distribution P_0 on $\mathcal{S}_{g(t_0)}$ with mean 0 and covariance matrix V_0. We fix adapted coordinate systems at each $g(t)$. Then $\xi(t_1 + \delta t)$ has mean 0 and covariance matrix $\check{V}(t_0, t_1 + \delta t) = C(t_0, t_1 + \delta t)V_0 C(t_0, t_1 + \delta t)^T + V(t_0, t_1 + \delta t)$ where $V(t_0, t_1 + \delta t)$ is given by (S4.3).

If $\xi(t_1 + \delta t)$ is such that $X(t_1 + \delta t) \in \mathcal{S}_{g(t_1)}$ then $G(X(t_1 + \delta t)) = g(t_1)$ and therefore $\kappa(t_1 + \delta t) = \xi(t_1 + \delta t) + \sqrt{\Omega}(g(t_1 + \delta t) - g(t_1))$. If $\underline{n}(t)$ denotes the unit vector $F(g(t))/||F(g(t))||$, since $\kappa(t_1 + \delta t)$ is orthogonal to $\underline{n}(t_1)$, we have that

$$\underline{n}(t_1) \cdot \xi(t_1 + \delta t) + \sqrt{\Omega}\, \underline{n}(t_1) \cdot (g(t_1 + \delta t) - g(t_1)) = 0 \tag{S6.1}$$

But since $\dot{g}(t) = F(g(t))$ and the time derivative of $F(g(t))$ is $J(g(t)) \cdot F(g(t))$ it follows that,

up to terms that are $\mathrm{O}(|\delta t|^3)$,

$$g(t_1+\delta t)-g(t_1)=F(t_1)\delta t+(1/2)J(g(t_1))\cdot F(g(t_1))(\delta t)^2. \tag{S6.2}$$

Now choose adapted coordinates $(x_1,\ldots,x_n)$ at $g(t_1)$ so that $F(g(t_1))=(\alpha,0,\ldots,0)$ and $\underline{n}(t_1)=(1,0,\ldots,0)$. Then, $\underline{n}(t_1)\cdot(J(g(t_1))\cdot F(g(t_1)))=\alpha J_{11}$ where J_{11} is the first entry of $J(g(t_1))$. Thus using (S6.1),

$$\frac{\underline{n}(t_1)\cdot\xi(t_1+\delta t)}{\sqrt{\Omega}}+\alpha\delta t+\frac{1}{2}\alpha J_{11}\delta t^2=\mathrm{O}(|\delta t|^3).$$

Since all terms but one are independent of Ω and since $\xi(t_1+\delta t)$ has finite variance, it follows that the standard deviation of δt is $\mathrm{O}(\Omega^{-1/2})$. Thus the rhs η is a random variable depending stochastically only on δt which is independent of Ω and has a standard deviation and mean that is $\mathrm{O}(\Omega^{-3/2})$. Letting $\kappa=\alpha J_{11}/2$ and $\Xi=(\underline{n}(t_1)\cdot\xi(t_1+\delta t))/\sqrt{\Omega}$, completing the square of the lhs gives $\delta t+\alpha/2\kappa=\pm\sqrt{1-4\kappa(\Xi-\eta)/\alpha^2}$ where η is the above $\mathrm{O}(|\delta t|^3)$ term. Using $\sqrt{1+x}=1+x/2-x^2/8+\mathrm{O}(|x|^3)$ and ignoring the negative root gives,

$$\delta t=-\frac{1}{\alpha}\Xi-\frac{J_{11}}{2\alpha^2}\Xi^2. \tag{S6.3}$$

Note that $\underline{n}(t_1)\cdot\xi(t_1+\delta t)$ differs from $\underline{n}(t_1)\cdot\xi(t_1)$ by an amount that is $\mathrm{O}(|\delta t|^2)$ (since the base point moves a distance that is $\mathrm{O}(|\delta t|)$ around the limit cycle) and the mean of $\xi(t_1+\delta t)$ given δt is $C(t_0,t_1+\delta t)\,E_{P_0}(\xi(t_0))$ which is 0 if the mean $E_{P_0}(\xi(t_0))$ of P_0 is 0. Thus, the mean of Ξ is 0 and, up to an error that is $\mathrm{O}(\Omega^{-3/2})$, the mean of the last term in (S6.3) is $(J_{11}/2\alpha^2)\mathrm{Var}(\Xi)$ which is $\mathrm{O}(\Omega^{-1})$ and the variance of this term is $\mathrm{O}(\Omega^{-2})$.

The variance of δt is,

$$\begin{aligned}\mathrm{Var}(\delta t)&=(1/\alpha^2)\mathrm{Var}(\Xi)+\mathrm{O}(\Omega^{-2})\\&=(1/\alpha^2\Omega)\,\mathrm{Var}(\underline{n}(t_1)\cdot\xi(t_1+\delta t))+\mathrm{O}(\Omega^{-2})\\&=(1/\alpha^2\Omega)\,\underline{n}(t_1)^T\check{V}(t_0,t_1+\delta t)\underline{n}(t_1)+\mathrm{O}(\Omega^{-2})\\&\overset{(1)}{=}(1/\alpha^2\Omega)\,\underline{n}(t_1)^T\check{V}(t_0,t_1)\underline{n}(t_1)+\mathrm{O}(\Omega^{-3/2})\\&=(1/\alpha^2\Omega)\,\check{V}_{11}+\mathrm{O}(\Omega^{-3/2})\end{aligned} \tag{S6.4}$$

where $\check{V}_{11}$ is the upper left entry of the covariance matrix $\check{V}(t_0,t_1)$ written in the given adapted coordinates. The equality marked $\overset{(1)}{=}$ follows from the fact that $\check{V}(t_0,t_1+\delta t)=\check{V}(t_0,t_1)+\mathrm{O}(|\delta t|)$.

Since Ξ in (S6.3) has a MVN distribution, it follows that δt is approximately normal but is

biased by the $(J_{11}/2\alpha^2)\Xi^2$ term which is $\mathrm{O}(\Omega^{-1})$ and hence δt has mean $\mathrm{O}(\Omega^{-1})$ and variance $(1/\alpha^2\Omega)\,V_{11} + \mathrm{O}(\Omega^{-3/2})$. For the *Drosophila* clock and system sizes $\Omega = 300$, 500 and 1000, the standard deviation of δt for completing one round of the limit cycle under the above LNA formula is respectively 0.7218, 0.5591 and 0.3954 whereas the empirical standard deviation of δt derived from the SSA is 0.7333, 0.5490 and 0.3862.

7 The approximation of the LNA transversal distributions

Recall the definition of $P_{\mathrm{LNA}}(\xi(t)|X(t) \in \mathcal{S}_{g(t_1)})$ in Sect. S6. In this section we consider how well this is approximated by $P_{\mathrm{LNA}}(\xi(t_1)|X(t_1) \in \mathcal{S}_{g(t_1)})$ i.e. by constraining time t to be the fixed time t_1. We show that the variance $\mathrm{Var}_{\delta t}(X(t_1+\delta t))$ of $X(t) = X(t_1+\delta t)$ conditional on $X(t) \in \mathcal{S}_{g(t_1)}$ equals the variance of $X(t_1)$ conditional on $X(t_1) \in \mathcal{S}_{g(t_1)}$ up to a term that is $\mathrm{O}(\Omega^{-1})$. The latter conditional variance is relatively easily calculated using the results in Sect. S5.

Throughout this section for a given $\xi(t)$, $X(t)$ denotes $g(t) + \Omega^{-1/2}\xi(t) \in \mathcal{S}_{g(t_1)}$. We first consider the variance of $X(t_1+\delta t)$ conditional on a fixed value of δt and $X(t_1+\delta t) \in \mathcal{S}_{g(t_1)}$. We omit the conditioning on $X(t_1+\delta t) \in \mathcal{S}_{g(t_1)}$ from the notation in the rest of this section.

The corresponding covariance matrix $\mathrm{Var}(X(t_1+\delta t)|\delta t)$ has a Taylor expansion

$$\mathrm{Var}(X(t_1+\delta t)|\delta t) = \mathrm{Var}(X(t_1)|\delta t = 0) + \delta t W + \mathrm{O}(\delta t^2)$$

where the symmetric matrix W is easily calculated.

Note that $E_{\delta t}\,[\delta t W] = \mathrm{O}(\Omega^{-1})$ since the mean of δt has this order (see Sect. S6). Consequently,

$$E_{\delta t}\,[\mathrm{Var}(X(t_1+\delta t)|\delta t)] = \mathrm{Var}(X(t_1)) + \mathrm{O}(\Omega^{-1}).$$

We now consider $\mathrm{Var}_{\delta t}(E(X(t_1+\delta t)|\delta t))$. The mean $E(X(t_1+\delta t)|\delta t) = g(t_1+\delta t)$ and therefore from (S6.2) and (S6.4), $\mathrm{Var}_{\delta t}(E(X(t_1+\delta t)|\delta t)) = O(\Omega^{-1})$. Hence we have the result because, by the Law of Total Variation,

$$\mathrm{Var}(X(t_1+\delta t)) = E(\mathrm{Var}(X(t_1+\delta t)|\delta t) + \mathrm{Var}(E(X(t_1+\delta t)|\delta t)) = \mathrm{Var}(X(t_1)) + \mathrm{O}(\Omega^{-1}).$$

8 Proof of convergence for LNA transversal distributions

8.1 Free-running oscillators.

We show that for any transversal $\mathcal{S} = \mathcal{S}_{g(t)}$ to γ there is a unique MVN distribution $P_{\mathcal{S}}^{(\mathrm{fp})}$ on $\mathcal{S}$ with the property that under the pcLNA, $P_{\mathrm{LNA}}(Q_{x_i}^{(2)}|Q_{x_i}^{(1)} \sim P_{\mathcal{S}}^{(\mathrm{fp})}) = P_{\mathcal{S}}^{(\mathrm{fp})}$. To do this we study the limit of $P_{\mathrm{LNA}}(Q_{x_i}^{(N)}|Q_{x_i}^{(1)})$ as $N \to \infty$ because this is such a fixed point.

We start by considering the case where $\mathcal{S}$ is the normal hypersection and fix adapted coordinates $(x_1, \mathbf{x}_2)$ in which the transversal section $\mathcal{S}$ is given by $x_1 = 0$.

By periodicity the covariance matrix of $P_{\mathrm{LNA}}(\xi(t+n\tau)|X(0) = g(t)) = P_{\mathrm{LNA}}(\xi(0)|X(t-n\tau) = g(t))$ is $V(t-N\tau, t)$ (see (S4.3)). We have,

$$\begin{aligned} V(t-N\tau,t) &= \int_{t-N\tau}^{t} C(s,t)E(s)E(s)^T C(s,t)^T\, ds \\ &= \sum_{k=0}^{N-1} C(t-k\tau,t)V(t,t+\tau)C(t-k\tau,t)^T \\ &= \sum_{k=0}^{N-1} C^k V (C^k)^T = \sum_{k=0}^{N-1} (C^k L)(C^k L)^T \end{aligned} \tag{S8.1}$$

where $C = C(t, t+\tau)$, $V = V(t, t+\tau) = \int_t^{t+\tau} C(s,t)E(s)E(s)^T C(s,t)^T\, ds$ and $V = LL^T$ is the Cholesky decomposition of the positive definite matrix V.

In what follows we utilise the vec operator and the Kronecker product $\otimes$. All the facts about these that are used here can be found in [7] Chapter 4. A basic equality is that $\mathrm{vec}(AXB) = (B^T \otimes A)\mathrm{vec}(X)$. Therefore, since $\mathrm{vec}(C^k V (C^k)^T) = (C^k \otimes C^k)\mathrm{vec}(V) = (C \otimes C)^k \mathrm{vec}(V)$,

$$\mathrm{vec}(V(t-N\tau,t)) = \sum_{k=0}^{N-1} (C \otimes C)^k \mathrm{vec}(V). \tag{S8.2}$$

Our adapted coordinates $(x_1, \mathbf{x}_2) = (x_1, x_1, \ldots, x_n)$ at $g(t)$ are determined by the the unit vectors $e_i(t)$ (see Sect. S5) where $e_1(t) = F(g(t))/||F(g(t))||$ tangent to the limit cycle. Since $C\, e_1(t) = e_1(t)$ it follows that

$$C^k = \begin{bmatrix} 1 & \alpha^T \\ 0 & A \end{bmatrix}^k = \begin{bmatrix} 1 & b^{(k)} \\ 0 & A^k \end{bmatrix}$$

where A is an $(n-1)\times(n-1)$ matrix, α^T is a $n \times 1$ vector and $b^{(k)} = \alpha^T + \alpha^T A \cdots + \alpha^T A^{k-1}$.

Since γ is elementary and stable, 1 is a simple eigenvalue of C, and therefore all eigenvalues λ of A have $|\lambda| < 1$ and therefore $b^{(k)}$ converges to $b^{(\infty)} = \alpha^T (I - A)^{-1}$. Moreover, all eigenvalues $\lambda \neq 1$ of C^k (i.e. all eigenvalues of A^k) have modulus $|\lambda| < c\nu^k$ where $0 < \nu < 1$ and $c > 0$ are

constants independent of k.

It follows that $\mathbf{C}^{(k)} = (C \otimes C)^k = C^k \otimes C^k$ has the form

$$\mathbf{C}^{(k)} = \begin{bmatrix} 1 & \mathbf{b}^{(k)} \\ \underline{0} & \mathbf{A}^{(k)} \end{bmatrix}$$

where $\mathbf{b}^{(k)}$ is $1 \times (n^2-1)$ and $\mathbf{A}^{(k)}$ is $(n^2-1) \times (n^2-1)$. The eigenvalues of $C^k \otimes C^k$ are all of the form $\lambda\mu$ where λ and μ are eigenvalues of C^k. Thus 1 is a simple eigenvalue and all other eigenvalues λ have $|\lambda| < c\nu^k$ where $0 < \nu < 1$ and c are constants independent of k. Consequently, all eigenvalues of $\mathbf{A}^{(k)}$ have the latter property.

Moreover, if $b^{(k)} = (b_1^{(k)}, \ldots, b_{n-1}^{(k)})$ then the top row of $\mathbf{C}^{(k)}$ is obtained by concatenating the row vectors $1 \otimes [1\ b^{(k)}]$ and $b^{(k)} \otimes [1\ b^{(k)}]$ for $i = 1, \ldots, n-1$, and therefore the entries in this row converge exponentially fast to finite limits. Consequently, if we write $\mathrm{vec}(V)$ as $\alpha_1 E_1 + Z$ where $E_1 = (1, 0, \ldots, 0)^T$ and $Z = (0, \underline{z})^T$ is a vector perpendicular to E_1,

$$\mathbf{C}^{(k)}\mathrm{vec}(V) = \begin{bmatrix} \alpha_1 + \mathbf{b}^{(k)}\underline{z} \\ \mathbf{A}^{(k)}\underline{z} \end{bmatrix}$$

and the top entry converges exponentially fast to a limit $\alpha_1 + \mathbf{b}^{(\infty)}\underline{z} > 0$ i.e. $|\mathbf{b}^{(k)} - \mathbf{b}^{(\infty)}| \leq c\nu^k$ where $0 < \nu < 1$ and c are constants independent of k.

It follows immediately by (S8.2) that if we write $V(t - N\tau, t)$ in terms of the above coordinates $(x_1, \mathbf{x}_2) = (x_1, x_1, \ldots, x_n)$ as

$$V(t - N\tau, t) = \begin{bmatrix} V_{11} & V_{21}^T \\ \\ V_{21} & V_{22} \end{bmatrix}$$

then, as $N \to \infty$, while V_{21} and V_{22} have finite limits, $V_{11} \to +\infty$. Using the Schur complement, the precision matrix of $P^{(N)} = P_{\mathrm{LNA}}(\xi(0)|X(t - N\tau) = g(t))$ is

$$V(t - N\tau, t)^{-1} = \begin{bmatrix} (V_{11} - V_{21}^T V_{22}^{-1} V_{21})^{-1} & -V_{11}^{-1} V_{21}^T (V_{22} - V_{21} V_{11}^{-1} V_{21}^T)^{-1} \\ \\ -(V_{22} - V_{21} V_{11}^{-1} V_{21}^T)^{-1} V_{21} V_{11}^{-1} & (V_{22} - V_{21} V_{11}^{-1} V_{21}^T)^{-1} \end{bmatrix}.$$

This has a well-defined limit as $N \to \infty$ where, since $V_{11} \to \infty$, all entries but $(V(t - N\tau, t)^{-1})_{22}$ converge to zero and $(V(t - N\tau, t)^{-1})_{22}$ has a finite non-zero limit equal to $(\lim_{n\to\infty} V_{22})^{-1}$. Thus, if $\check{P}^{(N)}$ is the distribution given by conditioning $P^{(N)}$ on $x_1 = 0$ then the precision matrix of $\check{P}^{(N)}$ is $(V_{22} - V_{21} V_{11}^{-1} V_{21}^T)^{-1}$. Consequently, the covariance matrix of $\check{P}^{(N)}$ is $V_{22} - V_{21} V_{11}^{-1} V_{21}^T$

which in the limit $N \to \infty$ equals $V_{22}^{(\infty)} = \lim_{n\to\infty} V_{22}$. We thus deduce that the conditional distribution converges as $n \to \infty$ to a MVN distribution P with this covariance matrix. Since P is derived via the limiting process as $N \to \infty$, it is clear that $P_{\text{LNA}}(Q_{x_i}^{(2)}|Q_{x_i}^{(1)} \sim P) = P$ so $P = P_{\mathcal{S}}^{(\text{fp})}$.

8.2 Extending this result to arbitrary transversals.

Any transversal hyperplane $\mathcal{S}$ is of the form $W\mathcal{S}_N$ where $\mathcal{S}_N$ is the normal hyperplane and W is a $n \times n$ orthogonal matrix with $W_{11} > 0$. This latter condition ensures that $\mathcal{S}$ is transversal to the circle. Thus if $\mathbf{z} = (z_1, \mathbf{z}_2) = W\mathbf{y}$ then in the $\mathbf{z}$ coordinate system $\mathcal{S}$ is given by $z_1 = 0$. In this $\mathbf{z}$ coordinate system, the precision matrix of $P = P_{\mathcal{S}}^{(\text{fp})}$ is given by $WV(t - n\tau, t)^{-1}W^T$. However, from (8.1)

$$WV(t - n\tau, t)^{-1}W^T \to \begin{bmatrix} W_{12}(V_{22}^{(\infty)})^{-1}W_{12}^T & W_{12}(V_{22}^{(\infty)})^{-1}W_{22}^T \\ \\ W_{22}(V_{22}^{(\infty)})^{-1}W_{12}^T & W_{22}(V_{22}^{(\infty)})^{-1}W_{22}^T \end{bmatrix}.$$

Therefore, in the limit, $\mathbf{z}_2|(z_1 = 0)$ has covariance matrix $W_{22}V_{22}^{(\infty)}W_{22}^T$.

8.3 Alternative geometrical proof that convergence on $\mathcal{S}_N$ implies convergence on other transversals $\mathcal{S}$.

Let $\mathcal{S}_N$ be given by $y_1 = 0$ in the $\mathbf{y}$ coordinate system. Consider the ellipsoid $\mathcal{E}$ given by $\mathbf{y}V_n^{-1}\mathbf{y} = \mathbf{y}_0V_n^{-1}\mathbf{y}_0$ where $\mathbf{y}_0 \in \mathcal{S}_N$. If $p \in \mathcal{S}_N$ consider $p + \delta e_1$ where $\delta \in \mathbb{R}$ and e_1 is the vector $(1, 0, \ldots, 0)$ in the $\mathbf{y}$ system. Clearly there is a unique δ such that $p + \delta e_1 \in \mathcal{S}$. Denote this point by $\phi(p)$. Then ϕ is linear. Consider $Q_n(p) = \phi^{-1}(p)^TV_n^{-1}\phi^{-1}(p)$. By the above result for $\mathcal{S}_N$, on $\mathcal{S}_N$, Q_n converges to a $Q_{\infty,N}$. In the region $|y_1| < r$, the ellipsoids given by $Q_n(p) = \varepsilon > 0$ converge to cylinders of the form $\{(y_1, \mathbf{y}_2) : Q_{\infty,N}(\mathbf{y}_2) = \varepsilon\}$. Therefore, $Q_n(p) \to Q_\infty(p)$ where $Q_\infty(p) = Q_{\infty,N}(\mathbf{y}_2)$ if p has coordinate $(y_1, \mathbf{y}_2)$.

Extension to $P(Q_{x_j}^{(N)}|Q_{x_i}^{(1)})$, $x_j = g(t_j), x_i = g(t_i)$, $t_j \geq t_i$. The result is easily derived using that $V(t_i, t_j + N\tau) = C(t_i, t_j)V(t_i, t_i + N\tau)C(t_i, t_j)^T + V(t_i, t_j)$ which implies existence of the limit of $V^{-1}(t_i, t_j + N\tau)$, as $N \to \infty$, if $\lim_{N\to\infty} V^{-1}(t_i, t_i + N\tau)$ exists.

8.4 Entrained forced oscillators.

The above proof of Section S8.1 has to be slightly modified as follows. In this case all eigenvalues of $C(s, t)$ have $|\lambda| < 1$ and therefore all eigenvalues λ of $C^k \otimes C^k$ have $|\lambda| < c\nu^k$ where $0 < \nu < 1$

and c are constants independent of k. Consequently, using (S8.2), all entries of $V(t - N\tau, t)$ converge to a finite limit as $N \to \infty$.

This shows that in contradistinction to free-running oscillators, for entrained forced oscillators, $P_{\text{LNA}}(X(N\tau)|X(0))$ converges as $N \to \infty$. However, it does not mean that this is a good approximation to $P(X(N\tau)|X(0))$ for an exact simulation. As we show in the simulation study displayed in S2 Fig. 9, these two distributions appear to remain relatively close to each other as N increases. In these simulations, $P(X(N\tau)|X(0))$ tends to wrap itself around the limit cycle while $P_{\text{LNA}}(X(N\tau)|X(0))$ remains a bounded MVN. On the other hand, the conditional distributions $P_{\text{LNA}}(Q_{x_i}^{(N)}|Q_{x_i}^{(1)})$ do appear to remain accurate as $N \to \infty$.

9 Analytic expression for pcLNA transversal distributions

To calculate an analytic expression for

$$P_{\text{LNA}}(\underline{Q}|X(t_0)) = P_{\text{LNA}}(Q_{x_0}^{(1)}, \ldots, Q_{x_q}^{(m)}|X(t_0)) \tag{S9.1}$$

we need to introduce some new matrices. At each of the points x_k we fix a coordinate system $\mathcal{C}_{x_k}$ defined by a set of orthonormal vectors $e_1(x_k), \ldots, e_n(x_k)$ where $e_1(x_k)$ is normal to the hyperplane $\mathcal{S}_{x_k}$. We write coordinates in $\mathcal{C}_{x_k}$ in the form $(y_1, \boldsymbol{y}_2)$ where $y_1 \in \mathbb{R}$ and $\boldsymbol{y}_2 \in \mathbb{R}^{n-1}$ and $\boldsymbol{y}_2$ are coordinates on $\mathcal{S}_{x_k}$. The matrices $C(s,t)$ and $V(s,t)$ in this coordinate systems are

$$C(s,t) = \begin{bmatrix} C_{11} & C_{12} \\ \\ C_{21} & C_{22} \end{bmatrix} \quad \text{and} \quad V(s,t) = \begin{bmatrix} V_{11} & V_{12} \\ \\ V_{21} & V_{22} \end{bmatrix},$$

where $V_{12} = V_{21}^T$. Let $\check{C}(s,t) = C_{2\cdot} - V_{21}V_{11}^{-1}C_{1\cdot}$ and $\check{V}(s,t) = V_{22} - V_{21}V_{11}^{-1}V_{12}$. Note that $\check{V}(s,t)$ is the Schur complement of V_{22} in $V(s,t)$ so if $(y_1, \boldsymbol{y}_2)$ is MVN with covariance $V(s,t)$ then the distribution of $\boldsymbol{y}_2$ conditional on y_1 having a specific value has covariance $\check{V}(s,t)$.

Firstly, we consider $P_{\text{LNA}}(Q_{x_j}^{(l)}|Q_{x_i}^{(k)})$ where either $k < l$ or $k = l$ and $i < j$. In Section S7 we showed that this is very well approximated by a MVN distribution with mean $\check{C}_j Q_{x_i}^{(k)}$ and covariance $\check{V}_j$ where

$$\check{C}_j = \check{C}(t_j, t_j - t_i + (l-k)\tau), \; \check{V}_j = \check{V}(t_j, t_j - t_i + (l-k)\tau).$$

We also proved in Section 8 that as $l - k \to \infty$, $P_{\text{LNA}}(Q_{x_j}^{(l)}|Q_{x_i}^{(k)})$ converges to a limit P_{x_j} which is independent of $Q_{x_i}^{(k)}$ and does this exponentially fast in the sense that the mean and covariance

of $P_{\text{LNA}}(Q_{x_j}^{(l)}|Q_{x_i}^{(k)})$ converge exponentially fast to their limiting values.

9.1 Fixed point distribution.

The special case where the transitions are returns to the same transversal sections ($i = j$) is especially interesting. Clearly, $P_{\text{LNA}}(Q_{x_i}^{(k)}|Q_{x_i}^{(1)})$ converges to the same limit P_{x_i} as above and this distribution is a fixed point in the sense that the distribution $P_{\text{LNA}}(Q_{x_i}^{(k+1)}|Q_{x_i}^{(k)} \sim P_{x_i}) = P_{x_i}$. We therefore denote it $P_{x_i}^{(\text{fp})}$. Because of the aforementioned convergence the transversal distribution has the following important property

$$\lim_{k\to\infty} P_{\text{LNA}}(Q_{x_i}^{(k)}|X(t_0)) = P_{x_i}^{(\text{fp})}$$

for any initial condition $X(t_0)$. That is, the distribution of the system at any transversal section converges to a limiting multivariate normal distribution.

This distribution can be easily calculated numerically. To derive its mean and covariance matrix one solves the equations $m = \check{C}_i m$ and $S = \check{C}_i S \check{C}_i^T + \check{V}_i$, respectively. The latter can be solved by using the fact that $\text{vec}(\check{C}_i S \check{C}_i^T) = (\check{C}_i \otimes \check{C}_i)\text{vec}(S)$ which implies that

$$\text{vec}(S) = (I - \check{C}_i \otimes \check{C}_i)^{-1}\text{vec}(\check{V}_i). \tag{S9.2}$$

Here $\text{vec}(S)$ is the vector obtained by stacking all the columns of S on top of each other and $\otimes$ is the Kronecker or tensor product (see e.g. Chapter 4 of [7]).

For the circadian clock the convergence is fast: the L^2-norm of the difference between the covariance matrices of the limiting distribution, $P_{x_0}^{(\text{fp})}$ and $P_{\text{LNA}}(Q_{x_0}^{(r)}|X(t_0))$ for $r = 1, 2, \ldots, 5$ is respectively $(1250, 70, 3.9, 0.2, 0.1) \cdot 10^{-3}$.

9.2 The joint distribution of multiple transversals

We now consider the distribution of the sequence of transitions between transversal sections of different phases $P_{\text{LNA}}(\underline{Q}|X(t_0))$. We relabel $Q_{x_0}^{(1)}, \ldots, Q_{x_q}^{(1)}, \ldots, Q_{x_0}^{(m)}, \ldots, Q_{x_q}^{(m)}$ as $Q_1, \ldots, Q_N$ where $N = m(q+1)$. To each $Q_{x_i}^{(k)}$ there is a corresponding time $t_i + (k-1)\tau$. We label these times in increasing order as $T_1, \ldots T_N$ so that T_n corresponds to Q_n. Let $T_0 = t_0$. With this notation if follows from the above that $P_{\text{LNA}}(Q_{i+1}|Q_i)$ is approximately MVN with mean $\check{C}_i Q_i$ and covariance $\check{V}_i$ where $\check{C}_i = \check{C}(T_i, T_{i+1})$ and $\check{V}_i = \check{V}(T_i, T_{i+1})$. Consequently, if Q_i has covariance matrix $\check{S}_i$ then Q_{i+1} has covariance matrix given by $S_{i+1} = \check{V}_i + \check{C}_i S_i \check{C}_i^T$. As

explained in the next section, it follows that $P_{\text{LNA}}(\underline{Q}|Q_0)$ is MVN with mean

$$\mu = (\check{C}_0 Q_0, \ldots, \check{C}_{N-1} \ldots \check{C}_0 Q_0)$$

and covariance Σ where the precision matrix Σ^{-1} is a tridiagonal matrix with only non-zero entries, the main diagonal $\Sigma_{i,i} = \check{C}_i^T \check{V}_i^{-1} \check{C}_i + \check{V}_{i-1}^{-1}$ for $i = 1, 2, \ldots, N-1$ and $\Sigma_{N,N} = \check{V}_{N-1}^{-1}$, upper diagonal $-\check{C}_i^T \check{V}_i^{-1}$ and lower diagonal $-\check{V}_i^{-1}\check{C}_i$, $i = 1, \ldots, N-1$.

9.3 Precision matrix for $P(\underline{Q}|X(t_0))$.

It is clear from the above that the mean of $P_{\text{LNA}}(Q_{m+1}|Q_m)$ is $\check{C}_m Q_m$ and that $\check{S}_{i+1} = \check{V}_i + \check{C}_i S_i \check{C}_i^T$ which imply

$$\begin{aligned}
&(Q_{m+1} - \check{C}_m Q_m)^T V_{m+1}^{-1} (Q_{m+1} - \check{C}_m Q_m) \\
&\quad = \begin{bmatrix} Q_m^T & Q_{m+1}^T \end{bmatrix} \begin{bmatrix} \check{C}_m^T V_{m+1}^{-1} \check{C}_m & -\check{C}_m^T V_{m+1}^{-1} \\ -V_{m+1}^{-1} \check{C}_m & V_{m+1}^{-1} \end{bmatrix} \begin{bmatrix} Q_m \\ Q_{m+1} \end{bmatrix}
\end{aligned}$$

and the formula for the precision matrix for $P(\underline{Q}|X(t_0))$ follows from this using $P_{\text{LNA}}(\underline{Q}|X(t_0)) = P_{\text{LNA}}(Q_1|X(t_0)) \prod_{m=1}^{N-1} P_{\text{LNA}}(Q_{m+1}|Q_m)$.

10 Other transversal sections

For each state variable x_i two interesting transversals to γ are given by the submanifolds on which $\dot{x}_i(t) = 0$, i.e. where the maximum or minimum of $x_i(t)$ occurs at time t. These are both given by $F_i(x) = 0$ and their tangent spaces consist of those vectors v such that $\partial F_i / \partial x(g(t)) \cdot v = 0$. A continuous time trajectory intersects these transversal sections at times when the ith variable is at its maxima and minima. Thus $\underline{Q}$ gives the state of the system at their (large volume limit) maxima and minima.

Another very interesting transversal system is that given by the isochrons of γ. These are defined as follows. Let $\Phi(t, x)$ be the flow associated to F. For $x \in \gamma$ the stable manifold $W^s(x)$ of x is the set of points y such that $\Phi(t, y) - \Phi(t, x) \to 0$ as $t \to \infty$. If γ is an elementary and stable limit cycle as defined in Sect. S1, then, in a neighbourhood of γ, these stable manifolds are codimension one submanifolds that fill the neighbourhood. This follows from the Stable Manifold Theorem [6] and is explained in [4]. They are called isochrons because two points in the same isochron asymptotically have the same phase. They are invariant manifolds in the sense that $W^s(\Phi(t, x)) = \Phi(W^s(x), t)$ i.e. at time t the flow caries the stable manifold of x

exactly onto the stable manifold at $\Phi(t,x)$. In particular, if τ is the period then $\Phi(\tau,\cdot)$ maps $W^s(x)$ to $W^s(x)$ and therefore, using the notation above, if $x = g(t)$, the linear map given by $C(t,t+\tau)$ maps the $(n-1)$-dimensional tangent space V_x of $W^s(x)$ to itself. By using the fact that V_x is transverse to γ, or directly from the Invariant Manifold Theorem, it follows that V_x is spanned by the generalised eigenspace of $C(t,t+\tau)$ corresponding to the eigenvalues λ with $|\lambda| < 1$. Consequently, a very interesting transversal system is that where $\mathcal{S}_{g(t)} = V_{g(t)}$ is the tangent space to the isochron through $g(t)$. By the invariance we have $C_{12} = 0$ so that $\check{C} = C_{22}$.

A final interesting transversal system is based on the V matrices. For an arbitrary point on the limit cycle $x_t = g(t)$, the matrix $V = V(t,t+\tau)$ is symmetric positive definite and therefore its orthonormal eigenvectors $\tilde{e}_i$, $i = 1,\ldots,n$ span $\mathbb{R}^N$. The eigenvector corresponding to the largest eigenvalue of V is approximately equal to the unit tangent. We can define the transversal section as the hyperplane spanned by $\tilde{e}_2,\ldots,\tilde{e}_n$. The advantage of this transversal system is that the largest variability direction is eliminated in the transversal distributions.

11 Details behind the sensitivity analysis.

If P is MVN with mean and covariance $\mu = \mu(\theta)$ and $\Sigma = \Sigma(\theta)$ then the entries of the FIM are given by

$$I_{ij} = \frac{\partial \mu}{\partial \theta_i}^T \Sigma^{-1} \frac{\partial \mu}{\partial \theta_j} + \frac{1}{2}\mathrm{tr}\left(\Sigma^{-1}\frac{\partial \Sigma}{\partial \theta_i}\Sigma^{-1}\frac{\partial \Sigma}{\partial \theta_j}\right). \tag{S11.1}$$

As above we consider a family of probability distributions $P(X,\theta)$ depending on parameters in θ. We assume these are MVN with mean $\mu(\theta)$ and covariance matrix $\Sigma(\theta)$. In calculating the FIM we have to determine the partial derivatives $\partial\mu/\partial\theta_i$ and $\partial\Sigma/\partial\theta_i$. The derivative M of the mapping $\theta \to (\mu(\theta),\Sigma(\theta))$ at a parameter value θ_0 is given by

$$M \cdot \delta\theta = \left(\sum_i \frac{\partial \mu}{\partial \theta_i}\delta\theta_i, \sum_i \frac{\partial \Sigma}{\partial \theta_i}\delta\theta_i\right)$$

where the derivatives are calculated at θ_0.

As is well-known in Information Geometry the set of multivariate normal distributions MVN^n on $\mathbb{R}^n$ can be given the structure of a Riemannian manifold in which the Riemannian metric is given the line element $ds^2 = d\mu^T\Sigma^{-1}\,d\mu + (1/2)\mathrm{tr}\{(\Sigma^{-1}d\Sigma)^2\}$. Points in this are denoted by $\Theta = (\mu,\Sigma)$ where μ is the mean and Σ the covariance matrix. The inner product

in the tangent space at $\Theta_0 = (\mu(\theta_0), \Sigma(\theta_0))$ is given by

$$\langle \delta\Theta_1, \delta\Theta_2 \rangle_{\Theta_0} = \delta\mu_1 \Sigma^{-1} \delta\mu_2 + \frac{1}{2} \mathrm{tr}\left(\Sigma^{-1}\delta\Sigma_1\Sigma^{-1}\delta\Sigma_2\right) \tag{S11.2}$$

where $\Theta_0 = (\mu, \Sigma)$ and $\delta\Theta_j = (\delta\mu_j, \delta\Sigma_j)$, $j = 1, 2$.

The derivative M above at θ_0 is a linear map of the parameter space into MVN^n and therefore (taking the standard inner product $\langle \delta\theta_1, \delta\theta_2 \rangle = \delta\theta_1^T \cdot \delta\theta_2$ on the parameter space) has an adjoint M^* such that $\langle M\delta\theta_1, M\delta\theta_2 \rangle_{\mathrm{IG}} = \delta\theta_1^T M^* M \delta\theta_2$. In fact, the adjoint operator M^* to M is given by

$$M^* U = (\eta_1, \ldots, \eta_s)$$

where

$$\eta_j = \left\langle \frac{\partial\Theta}{\partial\theta_j}, U \right\rangle_{\Theta_0}.$$

By (S11.1) and (S11.2), $\delta\theta_1^T M^* M \delta\theta_2 = \langle M\delta\theta_1, M\delta\theta_2 \rangle_{\Theta_0} = \delta\theta_1^T F \delta\theta_2$. and therefore it follows that

$$M^* M = F. \tag{S11.3}$$

By applying Singular Value Decomposition appropriately (see below) we can find orthonormal vectors V_i spanning the parameter space $\mathbb{R}^s$, s orthonormal vectors U_i in the space MVN^n and numbers $\sigma_1 \geq \cdots \geq \sigma_s$ such that

$$MV_i = \sigma_i U_i, \qquad i = 1, \ldots, s. \tag{S11.4}$$

Note that the orthonormality of the U_i is with respect to the inner product $\langle \cdot, \cdot \rangle_{\Theta_0}$ It follows from (S11.4) that $M^* M V_i \cdot V_j = \sigma_i \sigma_j \delta_{ij}$ and therefore that $F V_i = M^* M V_i = \sigma_i^2 V_i$.

We regard a $s \times s$ matrix S'_{ij} as a sensitivity matrix if there is a corresponding basis for MVN^n $U'_i = (U_i^{\mu\prime}, U_i^{\Sigma\prime})$ such that for any change of parameters $\delta\theta$, the change in μ and Σ is given by

$$\begin{aligned} \delta\mu &= \sum_i U_i^{\mu\prime} \left(\sum_j S'_{ij} \delta\theta_j \right) + \mathrm{O}(||\delta\theta||^2) \\ \delta\Sigma &= \sum_i U_i^{\Sigma\prime} \left(\sum_j S'_{ij} \delta\theta_j \right) + \mathrm{O}(||\delta\theta||^2) \end{aligned} \tag{S11.5}$$

Note that the role of the S'_{ij} as sensitivities is seen from the following relation which follows

from (S11.5),

$$\|\delta\Theta\| = \|S \cdot \delta\theta\| + O(\|\delta\theta\|^2). \tag{S11.6}$$

With this definition $S_{ij} = \sigma_i V_{ji}$ is a sensitivity matrix (with $U_i' = U_i$) which has the following optimality property: for all $k = 1, \ldots, s$

$$\sum_{i \le k} \sum_j S_{ij}^2 \ge \sum_{i \le k} \sum_j S_{ij}'^2 \qquad \text{and} \qquad \sum_{i > k} \sum_j S_{ij}^2 \le \sum_{i > k} \sum_j S_{ij}'^2. \tag{S11.7}$$

This tells us that the basis U_i and the corresponding sensitivities S_{ij} are optimal for capturing as much sensitivity as possible in the low order principal components. Note also that (i) the rows $S_i^{(r)}$ of S_{ij} are orthogonal, and (ii) $S_i^{(r)} \cdot V_j = \sigma_i \delta_{ij}$.

11.1 Deducing equation (S11.4) from SVD.

We can write the points of MVN^n in Cartesian coordinates as $(\mu, \mathrm{vec}(\Sigma))$ and a straightforward calculation shows that in this representation the inner product $\langle (\mu_1, \Sigma_1), (\mu_2, \Sigma_2) \rangle_{(\mu,\Sigma)}$ is given by

$$\mu_1^T \Sigma^{-1} \mu_2 + \frac{1}{2} vec(\Sigma_1)(I \otimes \Sigma^{-1})^T (I \otimes \Sigma^{-1}) vec(\Sigma_2)$$

which can be written as $(\mu_1, vec(\Sigma_1))\mathcal{F}(\mu_1, vec(\Sigma_2))$ where $\mathcal{F}$ the $(n + n^2) \times (n + n^2)$ square matrix

$$\mathcal{F} = \begin{bmatrix} \Sigma^{-1} & \mathbf{0} \\ \mathbf{0} & (I \otimes \Sigma^{-1})^T (I \otimes \Sigma^{-1}) \end{bmatrix}. \tag{S11.8}$$

Let $R^T R = \mathcal{F}$ be the Cholesky decomposition of $\mathcal{F}$. Then, if $y_i = R \cdot (\mu_i, \mathrm{vec}(\Sigma_i))^T$, $i = 1, 2$, the above inner product equals $y_1^T \cdot y_2$. Thus, we let WDV^T be the classical SVD of RM where D is a diagonal matrix with entries $\sigma_1 \ge \cdots \ge \sigma_s$. Then the columns V_i of V and the vectors $U_i = R^{-1} W_i$ (W_i the ith column of W) satisfy (S11.4). This also gives a practical algorithm to calculate the V_i, U_i and σ_i.

12 Multiple crossings in each rotation

The exact stochastic trajectories considered in this article are generated by the SSA that simulates from the Markov jump process with rates dependent on the state of the system as described in **I**. In this Markov jump process, at almost every state of the system, $X(t)$, with phase $G(X(t)) = g(s)$, $s > 0$, there is a possibility for the next point, $X(t+\delta)$, where δ the time of the next jump, to have phase $G(X(t + \delta)) = g(s')$ with $s' < s$. That is, the probability of backward (in terms of phase) moves is positive in almost every state, with only few exceptions

(e.g. absorbing states).

If we fix a phase and consider the crossings of the exact stochastic trajectory to the corresponding transversal section, the probability of more than one crossings in each rotation is again almost always positive. There are a few related measures that are worth considering here. First, the number of crossings in each rotation that we know it is finite as the Markov jump process can only perform a finite number of jumps in a finite time interval. An interesting question is how fast the probability of the number of crossings declines with increasing number of crossings. Secondly, how much time does the stochastic trajectory spend in this forward and backward moves and how does this compare with the period of the limit cycle. Thirdly, whether the distribution of the first crossings differ from the distribution of the second crossings and henceforth.

To address these questions we performed a simulation study using the *Drosophila* circadian clock system for system size $\Omega = 300$. We computed 1000 simulated trajectories by letting the SSA run for a time interval of 4.5 periods of the limit cycle solution. Note that this simulation study requires all jumps of the SSA to be recorded and therefore it requires a substantially greater computer memory space than the other simulation studies where each trajectory is thinned.

As we can see in Figure A, the probability of the number of crossings declines exponentially fast, with only $\approx 20\%$ of the trajectories giving more than 5 crossings. Furthermore, as we can see in Figure B, the distribution of the time between the first and the last crossing is also declining exponentially fast with only $\approx 5\%$ time intervals being larger than 0.014 which is $O(10^{-4})$ compared to the period of the limit cycle. Finally, as we can see in Figure C, the distribution of the points at the different crossings of the same transversal section does not differ substantially in any rotation or transversal coordinate. The deviations are expected to decline with smaller variability levels (larger Ω) as well as the number of crossings and the time spend at each phase.

This suggests that the transversal points derived using the thinned stochastic trajectories of the SSA are good representation of all crossings, at least for the variability levels considered in this article.

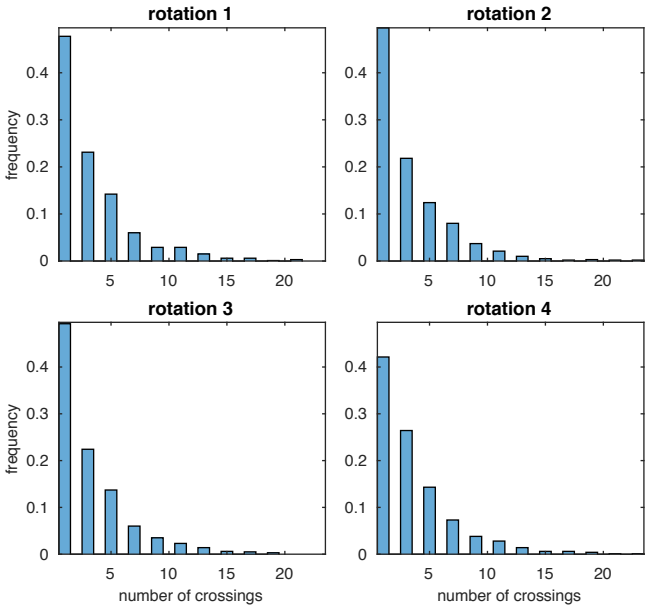

Figure A: Histograms of the empirical distribution of the number of crossings from a transversal section in each of the four rotations of the stochastic trajectories of the *Drosophila* circadian clock simulated using the SSA at system size $\Omega = 300$.

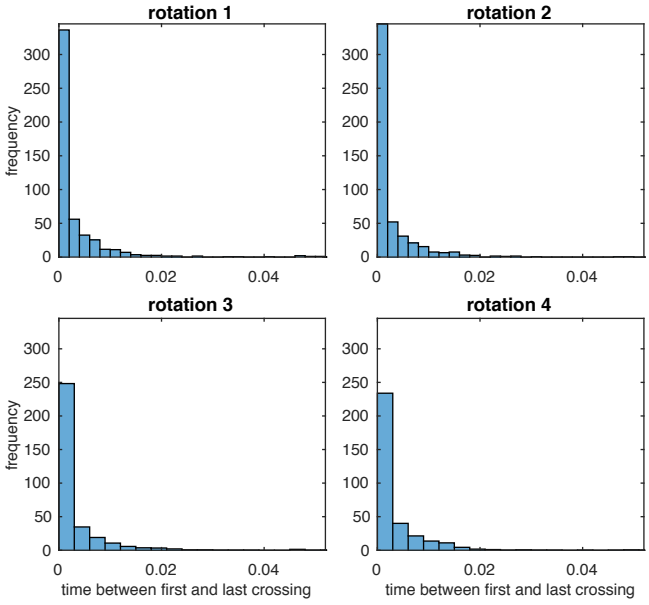

Figure B: Histograms of the empirical distribution of the time between the first and the last crossing from a transversal section in each of the four rotations of the stochastic trajectories of the *Drosophila* circadian clock simulated using the SSA at system size $\Omega = 300$.

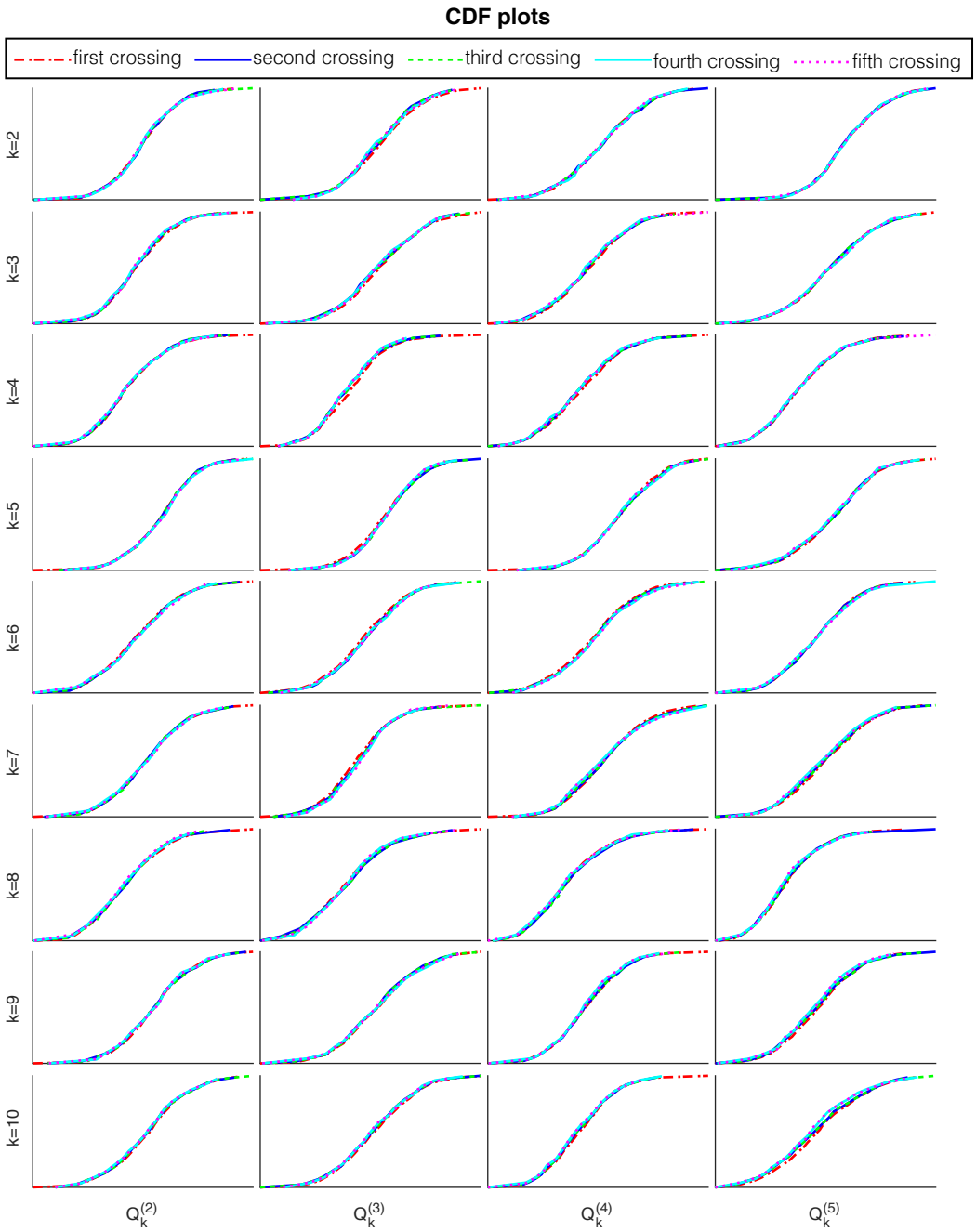

Figure C: Empirical CDF plots of the distribution of the points, $Q_k^{(r)}$, on the transversal section at the first, second, ..., fifth crossings (see legend) in each of the four rotations, $r = 1, 2, 3, 4$ and for transversal coordinates $k = 2, 3, \ldots, 10$ of the stochastic trajectories of the *Drosophila* circadian clock simulated using the SSA at system size $\Omega = 300$.

It is also worth clarifying here that for the pcLNA transversal distributions, as we discuss in Section S6 the distributions of the points in all crossings is considered and then approximated by the distribution of the crossing point at a specific time.

13 Negative populations

For the tau-leap, integration of CLE and pcLNA approximation algorithms considered here, the probability of generating negative populations is non-zero. To make the comparison fair to the SSA exact algorithm, we slightly modify these approximation algorithms to ensure that no negative species populations are derived. We use the same simple approach for all approximation algorithms to ensure the comparison between the approximations is fair. This is simply to replace any steps that result in negative populations by an SSA simulation run for the same time interval and then continue with the approximation algorithm. This increases CPU times as SSA is slower than the simulation algorithms, but also improves precision by introducing some bias towards SSA results. However, from our experience, the effect of this change in our comparisons is minimal in terms of precision and substantially affects CPU times only in rare cases. In particular, for large Ω $(= 1000, 3000)$ the SSA simulation is very rarely employed. For pcLNA, we also take advantage of the property, which also applies to standard LNA, of being able to analytically compute the distribution at the next time-point which is MVN and, before using the SSA step to replace negative populations, we sample from a truncated MVN distribution. For the latter we use a Matlab program[1] that is much faster than SSA simulation but in the challenging circumstances of the illustrations used in this article, it often fails to generate samples. In that case, an SSA simulation is used as in the other algorithms.

14 PeTTSy

PeTTSy is freely available at `http://www2.warwick.ac.uk/fac/sci/systemsbiology/research/software/`.

In PeTTSy the drift and diffusion matrices, $C(t)$ and $V(t)$, respectively, are computed for a large set of time-points. We next provide the description of the computation method for the drift matrix $C(s,t)$.

The deterministic solution is divided into a number of blocks, $tb_i, tb_{i+1} \ldots, tb_N$. By default, $N = 70$ blocks. The model Jacobian $J(t)$ is integrated over each of these time blocks in both directions to give forward and reverse solutions for each time block. Forward solution, $C(s,t)$ where s is tb_i, t runs from tb_i to tb_{i+1} is a solution of

$$\frac{dC(s,t)}{dt} = J(t) \cdot C(s,t)$$

[1] http://uk.mathworks.com /matlabcentral/fileexchange/34402-trunca ted-multivariate-normal

with initial condition $C(s,s) = I$ and I the identity matrix. Reverse solution, $C(s,t)$, where t is tb_i, s runs from tb_i to tb_{i-1} is a solution of the adjoint equation

$$\frac{dC(s,t)^T}{ds} = -J(s)^T \cdot C(s,t)^T$$

with initial condition $C(t,t) = I$. This is equivalent to solving,

$$\frac{dC(t-u,t)^T}{du} = J(t-u)^T \cdot C(t-u,t)^T \tag{S14.1}$$

for $t = tb_i$, $i = 1, 2, \ldots, N$, and where u runs from 0 to $tb_i - tb_{i-1}$. The results of this is that the forward and backward solutions allow us to construct any solution, $C(a,b)$, where if $a < tb_i$ and $b > tbi$, we can write the following

$$C(a,b) = C(tb_i, b) \cdot C(a, tb_i)$$

where $C(a, tbi)$ comes from the reverse solutions and $C(tb_i, b)$ comes from the forward solution. However, when a and b are inside the same interval, $[tb_i, tb_{i+1}]$ then

$$C(a,b) = C(tb_i, b) \cdot C(a, tb_i)^{-1}$$

or,

$$C(a,b) = C(tb_{i+1}, b)^{-1} C(a, tb_{i+1})$$

namely we have to invert either the reverse solution or the forward solution in order to calculate $C(a,b)$. PeTTSy provides a condition number that is equal to the minimum of one over the eigenvalue of the forward and backward solutions and if it is too large the matrix may be close to singular and then the inverse inaccurate. This means that forward and reverse solutions were not calculated finely enough, i.e., there were too few time blocks, and this will affect the accuracy of the results. One can check the accuracy of the $C(s,t)$ matrices, after forward and reverse solutions have been calculated, by a plot of the condition numbers. As a guide, if condition numbers are above $1/eps$, where eps is the floating accuracy of Matlab, the inverse is not accurate. The user is then advised to increase the value of N.

In a similar fashion, forward and reverse solutions are computed for the diffusion matrix V, using the equation (S4.3). Alternatively, one may solve the ODE

$$\frac{dV}{dt} = JV + VJ^T + EE^T,$$

where $E = E(s)$ the matrix described in equation (S4.3), with initial condition $V(t_0, t_0) = 0$. The reverse solutions are derived in a similar fashion to above by using the ODE

$$\frac{dV(t-u,t)^T}{du} = J(t-u)V(t-u,t) + V(t-u,t)J(t-u)^T + E(t-u)E(t-u)^T, \qquad \text{(S14.2)}$$

with initial condition $V(t_0, t_0) = 0$, for $t = tb_i$, $i = 1, 2, \ldots, N$, and where u runs from 0 to $tb_i - tb_{i-1}$.

15 The pcLNA Kalman Filter

The pcLNA Kalman Filter algorithm uses the following recursive algorithm for computing the terms in $L\left(\theta; \underline{\hat{X}}\right)$. Note that (μ^*, Σ^*) denote posterior estimates of (μ, Σ) conditional on the observed measurement at the current time.

1. Input $\left(\hat{X}(t_0), \hat{X}(t_1), \ldots, \hat{X}(t_N)\right)$, $\mu(t_0)$, $\Sigma(t_0)$, B and Σ_ϵ.
2. Compute $P\left(\hat{X}(t_0); \theta\right)$ from $MVN(\hat{\mu}(t_0), \hat{\Sigma}(t_0))$.
3. **For iteration** $i = 1, 2, \ldots$
   (a) Derive $(X(t_{i-1})|\hat{X}(t_{i-1})) \sim MVN(\mu^*(t_{i-1}), \Sigma^*(t_{i-1}))$;
   (b) Compute $g(s_{i-1}) = G(\mu^*(t_{i-1}))$ and set $(\kappa(s_{i-1})|\hat{X}(t_{i-1})) \sim MVN(m^*(s_{i-1}), S^*(s_{i-1}))$;
   (c) Derive $(X(t_i)|\hat{X}(t_{i-1})) \sim MVN(\mu(t_i), \Sigma(t_i))$;
   (d) Compute $P\left(\hat{X}(t_i)|\hat{X}(t_{i-1}); \theta\right)$ from $MVN(\hat{\mu}(t_i), \hat{\Sigma}(t_i))$.

In the for loop $\mu^*(t_{i-1}) = \mu(t_{i-1}) + \Sigma(t_{i-1})B^T\hat{\Sigma}(t_{i-1})^{-1}\hat{\epsilon}(t_{i-1})$, $\Sigma^*(t_{i-1}) = \Sigma(t_{i-1}) - \Sigma(t_{i-1})B^T\hat{\Sigma}(t_{i-1})^{-1}B\Sigma(t_{i-1})$, $m^*(s_{i-1}) = \Omega^{1/2}(\mu^*(t_{i-1}) - g(s_{i-1}))$, with $\hat{\epsilon}(t_i) = \hat{X}(t_i) - \hat{\mu}(t_i)$ and

$$S^*(s_{i-1}) = E_2(s_{i-1})\left(S^*_{22}(s_{i-1}) - \frac{S^*_{21}(s_{i-1})S^*_{12}(s_{i-1})}{S^*_{11}(s_{i-1})}\right)E_2(s_{i-1})^T$$

where the matrix $E_2(t) = [e_2(t) \ldots e_n(t)]$ has columns orthonormal vectors which are also orthogonal to the unit tangent vector $e_1(t) = F(x(t))/\|F(x(t))\|$ and

$$S^*_{11}(t) = e_1(t)^T S^*(t) e_1(t), \quad S^*_{12}(t) = e_1(t)^T S^*(t) E_2(t),$$

$$S^*_{21}(t) = S^*_{12}(t)^T, \quad S^*_{22}(t) = E_2(t)^T S^*(t) E_2(t).$$

The measurement equation (Eq. 12 in **I**) is used to derive the predictive probabilities $P(\hat{X}(t_0); \theta)$ and $P(\hat{X}(t_i)|\hat{X}(t_{i-1}); \theta)$ in steps 2 and 3(d), respectively. The posterior distribution $(X(t)|\hat{X}(t))$ in step 3(a) is derived by Bayes rule, while the posterior distribution of the corrected noise process $(\kappa(s)|\hat{X}(t))$ in step 3(b) is derived by restricting the poste-

rior distribution of the noise process on the transversal section $\mathcal{S}_{g(s)}$ normal to $e_1(s)$ using the Schur complement as described earlier. The cartesian coordinate system is used here. If this correction is omitted, step 3(b) of the above algorithm is replaced by the standard LNA step, $(\xi(t_{i-1})|\hat{X}(t_{i-1})) \sim MVN(\hat{m}(t_{i-1}), \hat{S}(t_{i-1}))$ where $\hat{m}(t_{i-1}) = \Omega^{1/2}\hat{\mu}(t_{i-1}) - g(t_{i-1})$, $\hat{S}(t_{i-1}) = \Omega\hat{\Sigma}(t_{i-1})$. The LNA Ansantz (Eq. S4.1) and transition (Eq. S4.2) are used to derive the prior distribution $(X(t_i)|\hat{X}(t_{i-1}))$ in step 3(c).

16 Details related to various computations

All computations for the illustrations in **I** and SI including:

1. SSA, tau-leap, integration of diffusion and pcLNA simulation algorithms described in Sections "pcLNA Simulation algorithm" and "Comparisons to other simulation algorithms" (see also the note in S1 Sect. 13),

2. the computation of Fisher Information and principal control coefficients for the sensitivity discussed in the Section "Sensitivity analysis for stochastic systems" and S1 Sect. 11

are done using the PeTTSy software which is discussed in S1 Sect. 14 and it is freely available at `http://www2.warwick.ac.uk/fac/sci/systemsbiology/research/software/`.

The phase correction that involves identifying the normal mapping $G(X) = x \in \gamma$ of a given state $X = X(t)$ of the trajectory is performed by Newton's optimisation minimising the square of the function

$$F(x(s)) \cdot (X - g(x(s))), \quad s > 0.$$

Note that the minimum is achieved when $s > 0$ is such that $X - g(x(s))$ is normal to the tangent vector $F(x(s))$. Multiple repetitions of the above optimisation are performed, using different initial conditions $s > 0$, to ensure that all local minima, $s' = s_1, s_2, \ldots$, are derived and then, as explained in S1 Sect. 1, $G(X(t)) = x(s)$ with $s = \min_i |s_i - t|$ the closest time to t. Note that other methods of optimisation could be equally applicable.

FIM computation We use the results in [5] to derive the necessary derivatives for the computation of the FIM. For the readers' convenience some of the derivations in chapter V in [5] are reproduced here. Let

$$\dot{x} = F(t, x, \theta), \quad x(t_0) = x_0$$

be the initial value problem with parameter vector θ which has solution $x = g(t, t_0, x_0, \theta)$. The Jacobian matrix $J(t) = J(t, x, \theta)$ at $x = g(t, t_0, x_0, \theta)$ has entries

$$\left(\frac{\partial F}{\partial x}\right)_{k,k'} = \frac{\partial F_k}{\partial x_{k'}}, \quad k, k' = 1, 2, \ldots, n$$

and the Hessian matrix $H(t) = H(t, x, \theta)$ at $x = g(t, t_0, x_0, \theta)$ has entries

$$\left(\frac{\partial^2 F}{\partial x^2}\right) = (H(F_1), \ldots, H(F_n))', \qquad (H(F_k))_{ss'} = \left(\frac{\partial^2 F_k}{\partial x_s \partial x_{s'}}\right).$$

Define

$$w_k = \frac{\partial g}{\partial x_{0,k}}, \qquad v_j = \frac{\partial g}{\partial \theta_j} \quad k = 1, 2, \ldots, n, \; j = 1, 2, \ldots, m.$$

Then, w_k is the solution of

$$\dot{x} = J(t)x, \quad x(t_0) = e_k$$

and e_k a vector with all zero entries except the kth entry which is equal to 1.

Note that the principal fundamental matrix C is the solution of the latter initial value problem. The matrix $C(t_0, t)$ has columns $w_k(t)$, $k = 1, 2, \ldots, n$. Furthermore, v_j is the solution of

$$\dot{x} = J(t)x + h_j(t), \quad x(t_0) = 0, \tag{S16.1}$$

where $h_j(t) = \partial F / \partial \theta_j$.

The $mn \times n$ matrix $(dC/d\theta_1 \ldots dC/d\theta_m)^T$ is the solution of

$$\dot{Z} = \begin{bmatrix} \mathrm{Diag}\left({w_1}^T\right) H v_1 & \ldots & \mathrm{Diag}\left({w_n}^T\right) H v_1 \\ \vdots & & \vdots \\ \mathrm{Diag}\left({w_1}^T\right) H v_m & \ldots & \mathrm{Diag}\left({w_n}^T\right) H v_m \end{bmatrix} + \mathrm{Diag}(J) Z + \begin{bmatrix} \frac{\partial J}{\partial \theta_1} w_1 & \ldots & \frac{\partial J}{\partial \theta_1} w_n \\ \vdots & & \vdots \\ \frac{\partial J}{\partial \theta_m} w_1 & \ldots & \frac{\partial J}{\partial \theta_m} w_n, \end{bmatrix} \tag{S16.2}$$

where $\mathrm{Diag}(J)$ is the $mn \times mn$ block matrix with main diagonal block entries the $n \times n$ matrix J. The $mn \times n$ matrix $(dV/d\theta_1, \ldots, dV/d\theta_m)^T$ has blocks $dV/d\theta_j$, $j = 1, 2, \ldots, m$ which are the solutions of

$$\dot{Z}_j = J Z_j + Z_j J^T + V \left(H^T \mathrm{Diag}\left(v_j\right) + \frac{\partial J^T}{\partial \theta_j}\right) + \left(\mathrm{Diag}\left({v_j}^T\right) H + \frac{\partial J}{\partial \theta_j}\right) V + B \frac{dR}{d\theta_j} B^T. \tag{S16.3}$$

Here B is the stoichiometry matrix of the system and $dR/d\theta_j$ the diagonal matrix with main

diagonal entries the derivatives of the reaction rates

$$\frac{du_i(x)}{d\theta_j} = \frac{\partial u_i(x)}{\partial x}^T \frac{\partial x}{\partial \theta_j} + \frac{\partial u_i(x)}{\partial \theta_j}, \quad \text{at } x = g(t, t_0, x_0, \theta).$$

In Section "Fisher Information" of **I** we discuss properties of the FIM of the distribution

$$\left(Q_{x_0}, Q_{x_1} | Q_0 = Q_{x_0}^{\infty}\right) \sim N(\mu, \Sigma)$$

where $Q_{x_0}^{\infty}$ the fixed point pcLNA distribution at $x_0 = g(t_0)$. The parameters of this distribution are

$$\mu = (x_0, x_1)^T, \quad \Sigma = S/\Omega,$$

and $S = A^{-1}$ with

$$A = \begin{bmatrix} (\check{C}_{01})^T (\check{V}_{01})^{-1} \check{C}_{01} + (S_0^{\infty})^{-1} & -(\check{C}_{01})^T (\check{V}_{01})^{-1} \\ \\ -(\check{V}_{01})^{-1} \check{C}_{01} & (\check{V}_{01})^{-1} \end{bmatrix}.$$

Here

$$\check{V}_{01} = P_1 V_{01}, \quad \check{C}_{01} = P_1 C_{01}, \quad P_1 = I - \frac{V_{01} \dot{x}_1 \dot{x}_1^T}{\dot{x}_1^T V_{01} \dot{x}_1}, \quad C_{01} = C(t_0, t_1),\ V_{01} = V(t_0, t_1) \quad \text{(S16.4)}$$

and S_0^{∞} the solution of the contraction equation

$$\text{vec}(S_0^{\infty}) = (I - \check{C} \otimes \check{C})^{-1} \text{vec}(\check{V})$$

where

$$\check{C} = \check{C}_{10} \check{C}_{01}, \quad \check{V} = \check{C}_{10} \check{V}_{01} \check{C}_{10}^T + \check{V}_{10}.$$

The matrices $\check{V}_{01}$ and $\check{C}_{01}$ are derived in (S16.4) and

$$\check{V}_{10} = P_0 V_{10}, \quad \check{C}_{10} = P_0 C_{10}, \quad P_0 = I - \frac{V_{10} \dot{x}_0 \dot{x}_0^T}{\dot{x}_0^T V_{10} \dot{x}_0}, \quad C_{10} = C(t_1, \tau + t_0),\ V_{10} = V(t_1, \tau + t_0)$$

and τ the period of the system. Thus,

$$\frac{d\mu}{d\theta_j} = \left(\frac{dx_0}{d\theta_j}, \frac{dx_1}{d\theta_j}\right)^T,$$

and

$$\frac{d\Sigma}{d\theta_j} = \Omega \frac{dA^{-1}}{d\theta_j} = \Omega A^{-1} \frac{dA}{d\theta_j} A^{-1}, \quad \frac{d\Sigma^{-1}}{d\theta_j} = \Omega \frac{dA}{d\theta_j}.$$

The derivative $dA/d\theta_j$ is a block matrix with blocks

$$\begin{aligned}
\frac{dA_{(1,1)}}{d\theta_j} &= \frac{d\check{C}_{01}}{d\theta_j}^T (\check{V}_{01})^{-1} \check{C}_{01} - \check{C}_{01}^T \check{V}_{01}^{-1} \frac{d\check{V}_{01}}{d\theta_j} \check{V}_{01}^{-1} \check{C}_{01} + \check{C}_{01}^T \check{V}_{01}^{-1} \frac{d\check{C}_{01}}{d\theta_j} - S_0^{\infty -1} \frac{dS_0^{\infty}}{d\theta_j} S_0^{\infty -1}, \\
\frac{dA_{(1,2)}}{d\theta_j} &= \check{C}_{01}^T \check{V}_{01}^{-1} \frac{d\check{V}_{01}}{d\theta_j} \check{V}_{01}^{-1} - \frac{d\check{C}_{01}}{d\theta_j}^T \check{V}_{01}^{-1}, \\
\frac{dA_{(2,1)}}{d\theta_j} &= \check{V}_{01}^{-1} \frac{d\check{V}_{01}}{d\theta_j} \check{V}_{01}^{-1} \check{C}_{01} - \check{V}_{01}^{-1} \frac{d\check{C}_{01}}{d\theta_j}, \\
\frac{dA_{(2,2)}}{d\theta_j} &= -\check{V}_{01}^{-1} \frac{d\check{V}_{01}}{d\theta_j} \check{V}_{01}^{-1}.
\end{aligned}$$

The necessary derivatives for computing the entries

$$I_{(jj')}(\theta) = \Omega \left(\frac{dx_0}{d\theta_j}^T, \frac{dx_1}{d\theta_j}^T \right) A \left(\frac{dx_0^T}{d\theta_j}, \frac{dx_1^T}{d\theta_{j'}} \right)^T + \frac{1}{2} tr \left(\frac{dA}{d\theta_j} A^{-1} \frac{dA}{d\theta_{j'}} A^{-1} \right)$$

of the FIM are derived as described above. In particular, $v_j = dx_i/d\theta_j$ is the solution of (S16.1) and the derivative $d\dot{x}_i/d\theta_j$ used to derive $dP_i/d\theta_j$ is equal to $J(t_i)v_j + g_j(t_i)$. Furthermore, the derivatives $dC/d\theta_j$ and $dV/d\theta_j$ are derived as solutions of the initial value problems in (S16.2) and (S16.3), respectively.

References

[1] Anderson DF & Kurtz TG. Continuous Time Markov Chain Models for Chemical Reaction Networks. In: Koeppl H, Setti G, di Bernardo M, Densmore D, editors. Design and Analysis of Biomolecular Circuits: Engineering Approaches to Systems and Synthetic Biology. New York: Springer.; 2011.

[2] Freidlin, MI, Wentzell, AD, & Szücs, J. . Random Perturbations of Dynamical Systems. Berlin: Springer Berlin; 2014.

[3] Greene JM, Kim JS The calculation of Lyapunov spectra. Physica D: Nonlinear Phenomena. 1987 Aug;24(1-3):213–225.

[4] Guckenheimer J. Isochrons and phaseless sets. Journal of Mathematical Biology. 1975 Sep;1(3):259–273.

[5] Hartman P. Ordinary differential equations. New York: John Wiley & Sons, Inc. London and Sydney: Wiley.; 1964.

[6] Hirsch MW, Pugh CC & Shub M. Invariant manifolds. Springer-Verlag; 1977.

[7] Horn, RA & Johnson CR. Matrix analysis. Cambridge University Press. Cambridge, New York; 1991.

[8] Kurtz TG. Approximation of Population Processes. Regional Conference Series in Applied Mathematics vol. 36. Society for Industrial and Applied Mathematics. 1981.

[9] Van Kampen, NG. Stochastic processes in physics and chemistry. Amsterdam: North Holland; 1997.

Supplementary information S2 for the paper Long-time analytic approximation of large stochastic oscillators: simulation, analysis and inference

Giorgos Minas[1,2] & David A Rand[1,2]

[1]Zeeman Institute for Systems Biology & Infectious Disease Epidemiology Research & [2]Mathematics Institute,

University of Warwick, Coventry CV4 7AL, UK

July 22, 2017

Abstract

We refer to the paper "Long-time analytic approximation of large stochastic oscillators: simulation, analysis and inference" by **I**. In this note we give details about the *Drosophila* circadian clock system and use it to illustrate further the accuracy of distributions and simulations discussed in **I**.

Contents

1 *D*rosophila circadian clock system **1**

2 Exact simulations **5**

3 Exact transversal distributions: normality **6**

4 pcLNA and exact transversal distributions **7**

5 Less frequent correction **9**

6 Comparison between different simulation algorithms **10**

7 Light entrained *D*rosophila Circadian Clock **11**

1 *Drosophila* circadian clock system

A schematic representation of the *Drosophila* circadian clock system, as provided in [1], is displayed in Figure A. The oscillations are driven by the negative feedback exerted on the *per* and *tim* genes by the complex formed from PER and TIM proteins following phosphorylation. The *per* and *tim* mRNA, M_P and M_T, respectively, are transported into the cytosol where they are degraded and translated into protein (P_0 and T_0). These proteins are multiply phosphorylated (PER: $P_0 \rightarrow P_1 \rightarrow P_2$; TIM: $T_0 \rightarrow T_1 \rightarrow T_2$) and these modifications can be reversed by a phosphatase. The fully phosphorylated form of the proteins is targeted for degradation and forms a complex, C, which is transported into the nucleus in a reversible manner where the nuclear form of the PER–TIM complex, C_N, represses the transcription of *per* and *tim* genes.

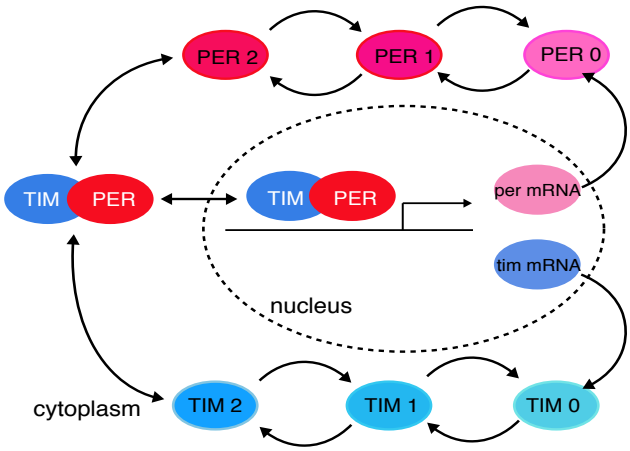

Figure A: Schematic representation of *Drosophila* circadian clock system [1]

The variables of the model along with the initial conditions (in nanomolar concentrations) used in our implementation are provided in Table A.

variable	description	initial condition
M_P	PER mRNA	3.0975
P_0	PER protein 0	1.2547
P_1	PER protein 1	1.2302
P_2	PER protein 2	1.7997
M_T	TIM mRNA	3.0975
T_0	TIM protein 0	1.2346
T_1	TIM protein 1	1.0577
T_2	TIM protein 2	0.3593
C	PER-TIM cytosolic complex	0.6230
C_N	PER-TIM nuclear complex	0.8178

Table A: The variables of *Drosophila* circadian clock system and the initial conditions (in nanomolar concentrations) used to derive their ODE solution.

The parameter values used to derive the ODE solution of the system are provided in Table B.

parameter	description	value	measurement unit
v_{sP}	M_P transcription	1.10	nMh^{-1}
v_{sT}	M_T transcription	1.00	nMh^{-1}
v_{mP}	M_P degradation	1.00	nMh^{-1}
v_{mT}	M_T degradation	0.70	nMh^{-1}
v_{dP}	P_2 degradation	2.20	nMh^{-1}
k_{sP}	M_P translation	0.90	h^{-1}
k_{sT}	M_T translation	0.90	h^{-1}
k_1	$C \rightarrow C_N$	0.80	h^{-1}
k_2	$C_N \rightarrow C$	0.20	h^{-1}
k_3	$P_2 + T_2 \rightarrow C$	1.20	h^{-1}
k_4	$C \rightarrow P_2 + T_2$	0.60	h^{-1}
K_{mP}	M_P enzymatic degradation	0.20	h^{-1}
K_{mT}	M_T enzymatic degradation	0.20	h^{-1}
K_{IP}	M_P Hill coefficient	1.00	h^{-1}
K_{IT}	M_T Hill coefficient	1.00	h^{-1}
K_{dP}	P_2 enzymatic degradation	0.20	h^{-1}
K_{dT}	T_2 enzymatic degradation	0.20	h^{-1}
k_d	linear degradation	0.01	h^{-1}
k_{dC}	C degradation	0.01	h^{-1}
k_{dN}	C_N degradation	0.01	h^{-1}
v_{dT}	T_2 degradation	3.00	nMh^{-1}
K_{1P}	$P_0 \rightarrow P_1$ enzymatic	2.00	h^{-1}
K_{1T}	$T_0 \rightarrow T_1$ enzymatic	2.00	h^{-1}
K_{2P}	$P_1 \rightarrow P_0$ enzymatic	2.00	h^{-1}
K_{2T}	$T_1 \rightarrow T_0$ enzymatic	2.00	h^{-1}
K_{3P}	$P_1 \rightarrow P_2$ enzymatic	2.00	h^{-1}
K_{3T}	$T_1 \rightarrow T_2$ enzymatic	2.00	h^{-1}
K_{4P}	$P_2 \rightarrow P_1$ enzymatic	2.00	h^{-1}
K_{4T}	$T_2 \rightarrow T_1$ enzymatic	2.00	h^{-1}
V_{1P}	$P_0 \rightarrow P_1$	8.00	nMh^{-1}
V_{1T}	$T_0 \rightarrow T_1$	8.00	nMh^{-1}
V_{2P}	$P_1 \rightarrow P_0$	1.00	nMh^{-1}
V_{2T}	$T_1 \rightarrow T_0$	1.00	nMh^{-1}
V_{3P}	$P_1 \rightarrow P_2$	8.00	nMh^{-1}
V_{3T}	$T_1 \rightarrow T_2$	8.00	nMh^{-1}
V_{4P}	$P_2 \rightarrow P_1$	1.00	nMh^{-1}
V_{4T}	$T_2 \rightarrow T_1$	1.00	nMh^{-1}
n	Hill power	4.00	NA

Table B: The parameters of *Drosophila* circadian clock system and the values used to derive their ODE solution.

The ODE system for the *Drosophila* circadian clock is

$$
\begin{aligned}
\dot{M}_P &= v_{sP}\frac{K_{IP}^n}{K_{IP}^n + C_N^n} - v_{mP}\frac{M_P}{K_{mP} + M_P} - k_d M_P \\
\dot{P}_0 &= k_{sP} M_P - V_{1P}\frac{P_0}{K_{1P} + P_0} + V_{2P}\frac{P_1}{K_{2P} + P_1} - k_d P_0 \\
\dot{P}_1 &= V_{1P}\frac{P_0}{K_{1P} + P_0} - V_{2P}\frac{P_1}{K_{2P} + P_1} - V_{3P}\frac{P_1}{K_{3P} + P_1} + V_{4P}\frac{P_2}{K_{4P} + P_2} - k_d P_1 \\
\dot{P}_2 &= V_{3P}\frac{P_1}{K_{3P} + P_1} - V_{4P}\frac{P_2}{K_{4P} + P_2} - k_3 P_2 T_2 + k_4 C - v_{dP}\frac{P_2}{K_{dP} + P_2} - k_d P_2 \\
\dot{M}_T &= v_{sT}\frac{K_{IT}^n}{K_{IT}^n + C_N^n} - v_{mT}\frac{M_T}{K_{mT} + M_T} - k_d M_T \\
\dot{T}_0 &= k_{sT} M_T - V_{1T}\frac{T_0}{K_{1T} + T_0} + V_{2T}\frac{T_1}{K_{2T} + T_1} - k_d T_0 \\
\dot{T}_1 &= V_{1T}\frac{T_0}{K_{1T} + T_0} - V_{2T}\frac{T_1}{K_{2T} + T_1} - V_{3T}\frac{T_1}{K_{3T} + T_1} + V_{4T}\frac{T_2}{K_{4T} + T_2} - k_d T_1 \\
\dot{T}_2 &= V_{3T}\frac{T_1}{K_{3T} + T_1} - V_{4T}\frac{T_2}{K_{4T} + T_2} - k_3 P_2 T_2 + k_4 C - v_{dT}\frac{T_2}{K_{dT} + T_2} - k_d T_2 \\
\dot{C} &= k_3 P_2 T_2 - k_4 C - k_1 C + k_2 C_N - k_{dC} C \\
\dot{C}_N &= k_1 C - k_2 C_N - k_{dN} C_N.
\end{aligned}
$$

The SSA trajectories are derived using the rates provided in Table 1 of [1].

In the simulations of the *Drosophila* circadian clock provided in the next section and in **I**, we consider five values of the system size $\Omega = 200$, 300, 500, 1000 and 3000. In Table C, we provide the approximate ranges of the molecular numbers for each species observed in our simulations. Notice the increase in the maximum number of molecules with increasing Ω, but also the decrease in the length of the range in terms of concentrations, $X(t) = Y(t)/\Omega$, that reflects the lower levels of stochasticity for increasing Ω.

variable	$\Omega = 200$	$\Omega = 300$	$\Omega = 500$	$\Omega = 1000$	$\Omega = 3000$
M_P	$(0, 800)$	$(0, 1200)$	$(0, 1900)$	$(0, 3500)$	$(40, 10000)$
P_0	$(0, 500)$	$(0, 600)$	$(0, 900)$	$(0, 1600)$	$(10, 4500)$
P_1	$(0, 500)$	$(0, 700)$	$(0, 1000)$	$(0, 1700)$	$(10, 4700)$
P_2	$(0, 1400)$	$(0, 1700)$	$(0, 2500)$	$(0, 4300)$	$(30, 10600)$
M_T	$(0, 800)$	$(0, 1200)$	$(0, 1900)$	$(0, 3500)$	$(40, 10000)$
T_0	$(0, 400)$	$(0, 600)$	$(0, 900)$	$(0, 1600)$	$(10, 4000)$
T_1	$(0, 400)$	$(0, 500)$	$(0, 800)$	$(0, 1400)$	$(10, 3900)$
T_2	$(0, 200)$	$(0, 300)$	$(0, 400)$	$(0, 600)$	$(10, 1400)$
C	$(6, 400)$	$(10, 500)$	$(30, 800)$	$(75, 1400)$	$(280, 3700)$
C_N	$(60, 700)$	$(100, 900)$	$(200, 1400)$	$(400, 2600)$	$(1300, 7000)$

Table C: The approximate ranges of the molecular numbers, $Y(t)$, of each of the species of *Drosophila* circadian clock system in simulations derived using SSA for various system sizes Ω.

For more details considering the *Drosophila* circadian clock system see [1].

2 Exact simulations

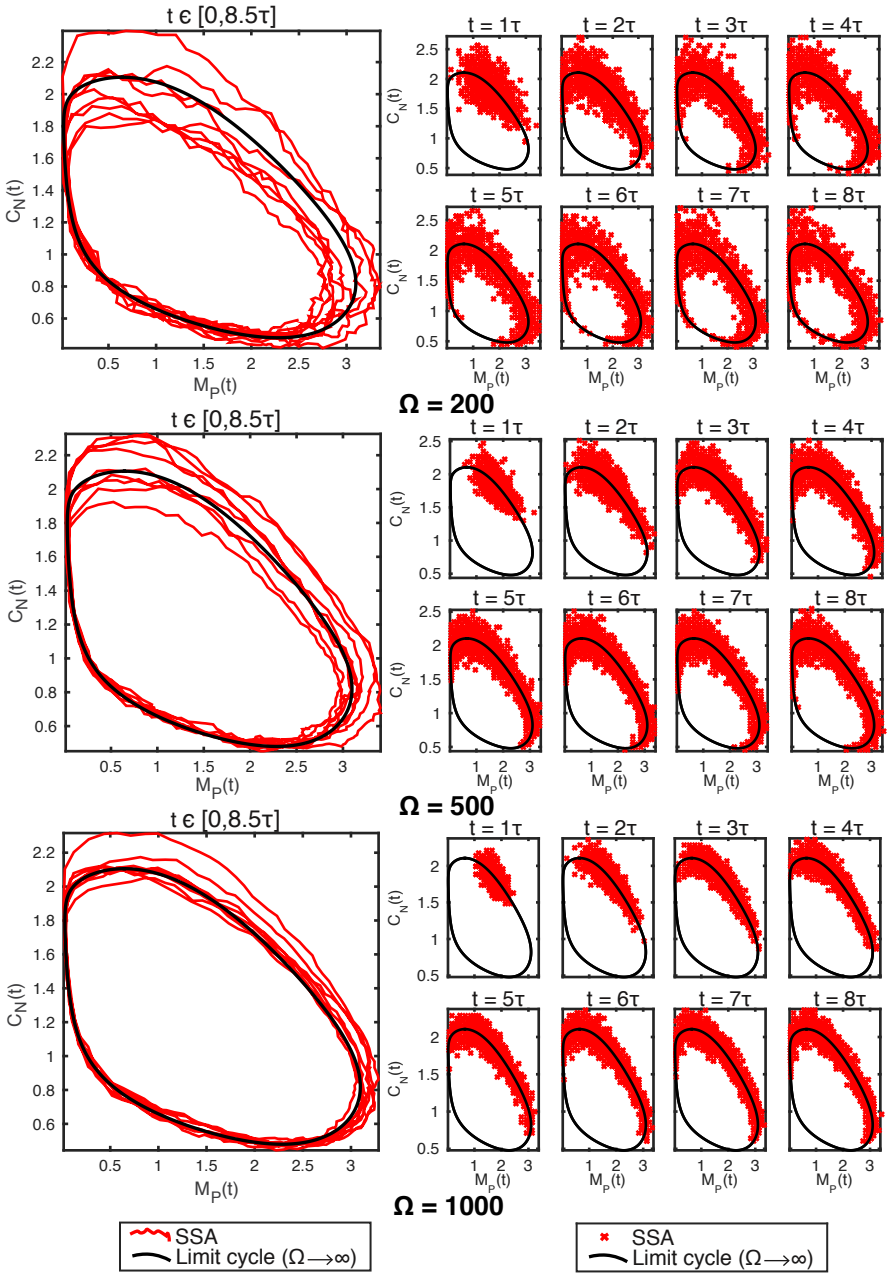

Figure B: Exact stochastic simulation of the *Drosophila* circadian clock system in three different system sizes, $\Omega = 200$ (top panel), $\Omega = 500$ (middle panel), $\Omega = 1000$ (bottom panel). A stochastic trajectory obtained by running the SSA over the time-interval $t \in [0, 8.5\tau]$ is displayed on the left panels and SSA samples ($R = 3000$) at times $t = \tau, 2\tau, \ldots, 8\tau$ are displayed on the right panels. Two (out of 10) of the species are displayed (*per* mRNA M_P (x-axis) and nuclear PER-TIM complex C_N (y-axis)). The black solid curve is the large volume, $\Omega \to \infty$, limit cycle solution.

3 Exact transversal distributions: normality

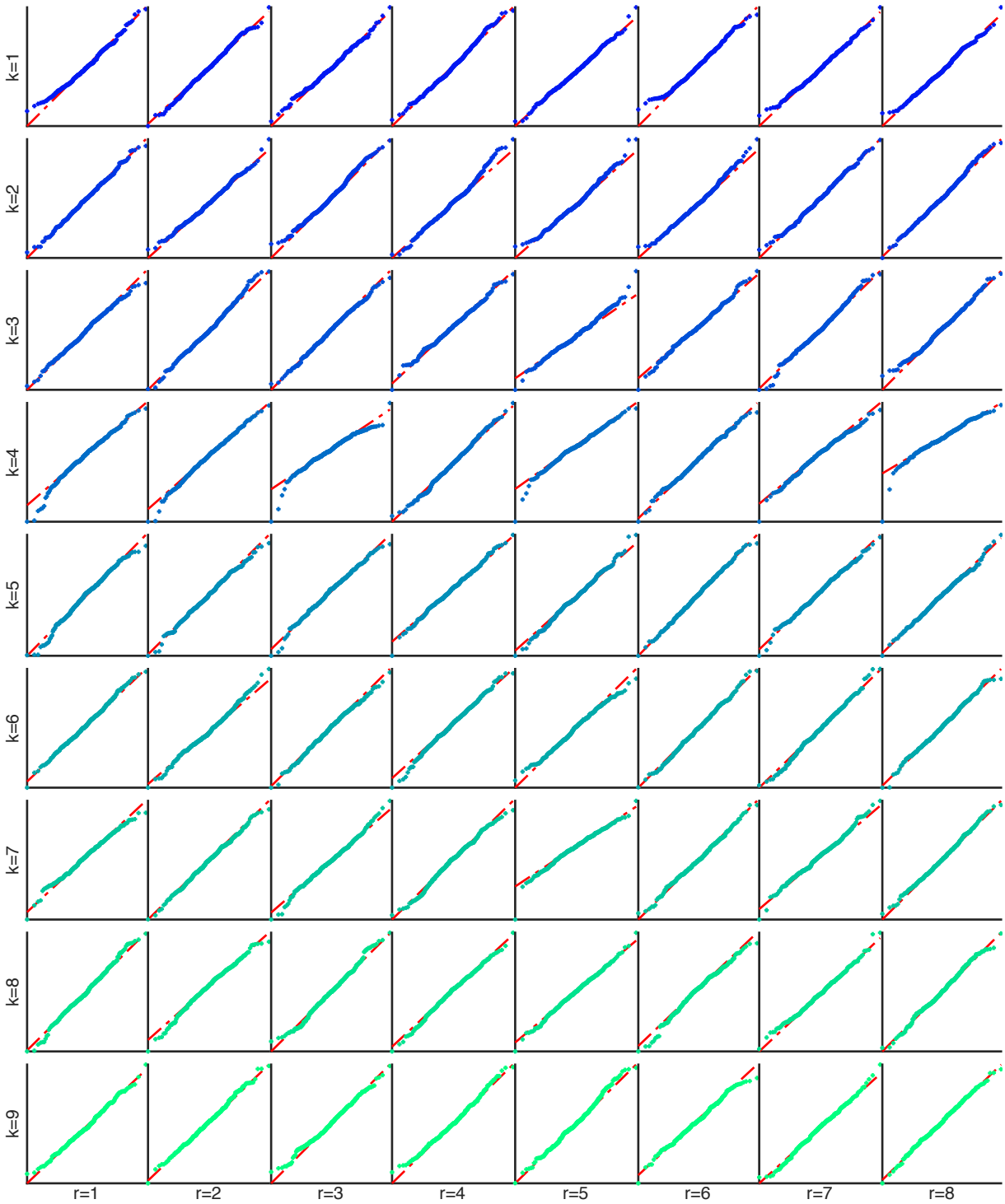

Figure C: Quantile-Quantile (Q-Q) plots of the distribution of the first intersection in the r-th pass, $Q_k^{(r)}$, in transversal coordinates $k = 2, 3, \ldots, 10$ for $r = 1, 2, \ldots, 8$.

4 pcLNA and exact transversal distributions

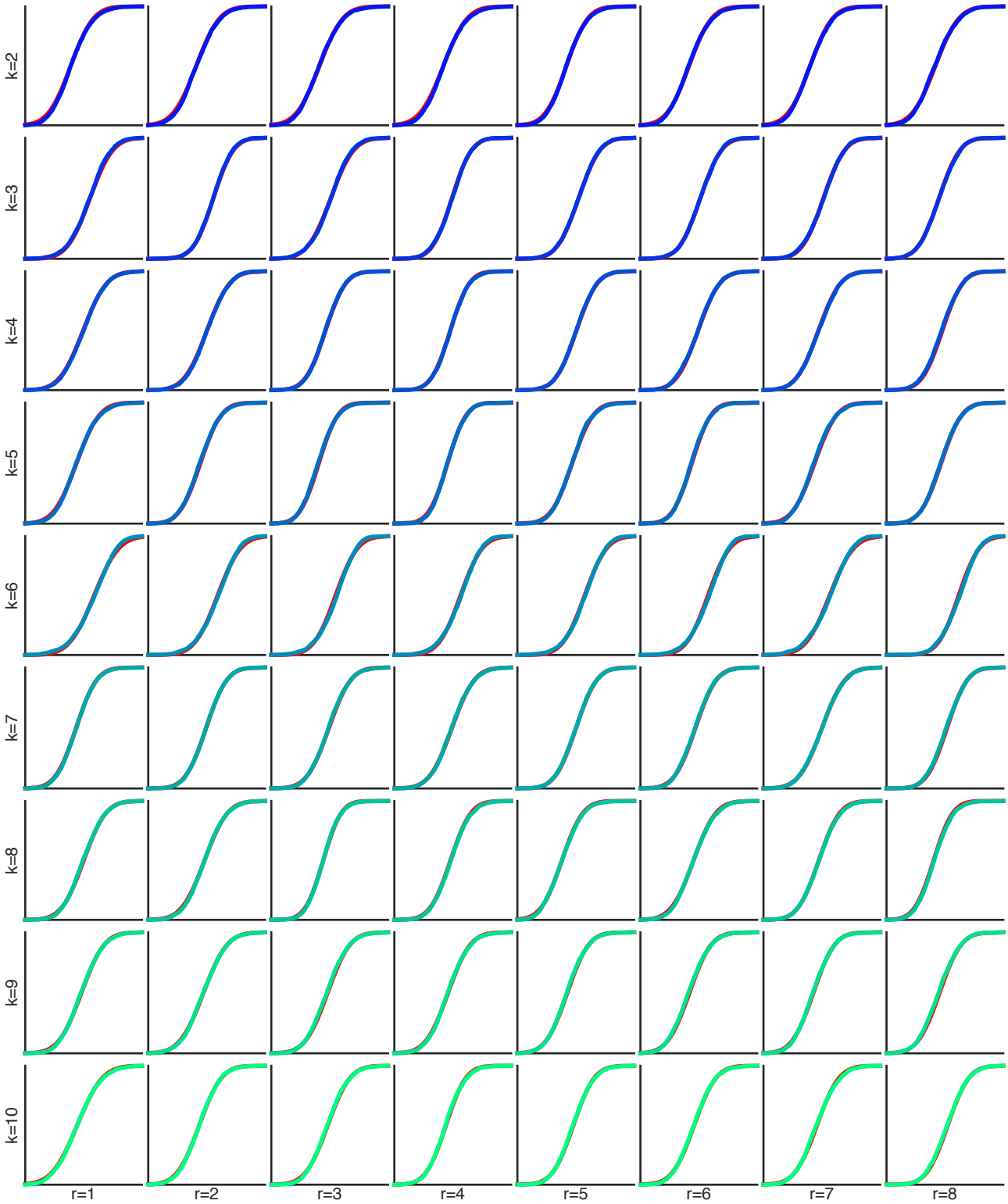

Figure D: CDF plots of the transversal distributions $Q_k^{(r)}$ under the pcLNA (red line) and the SSA (empirical CDF, crosses) for transversal coordinates $k = 2, 3, \ldots, 10$ and round $r = 1, 2, \ldots, 8$.

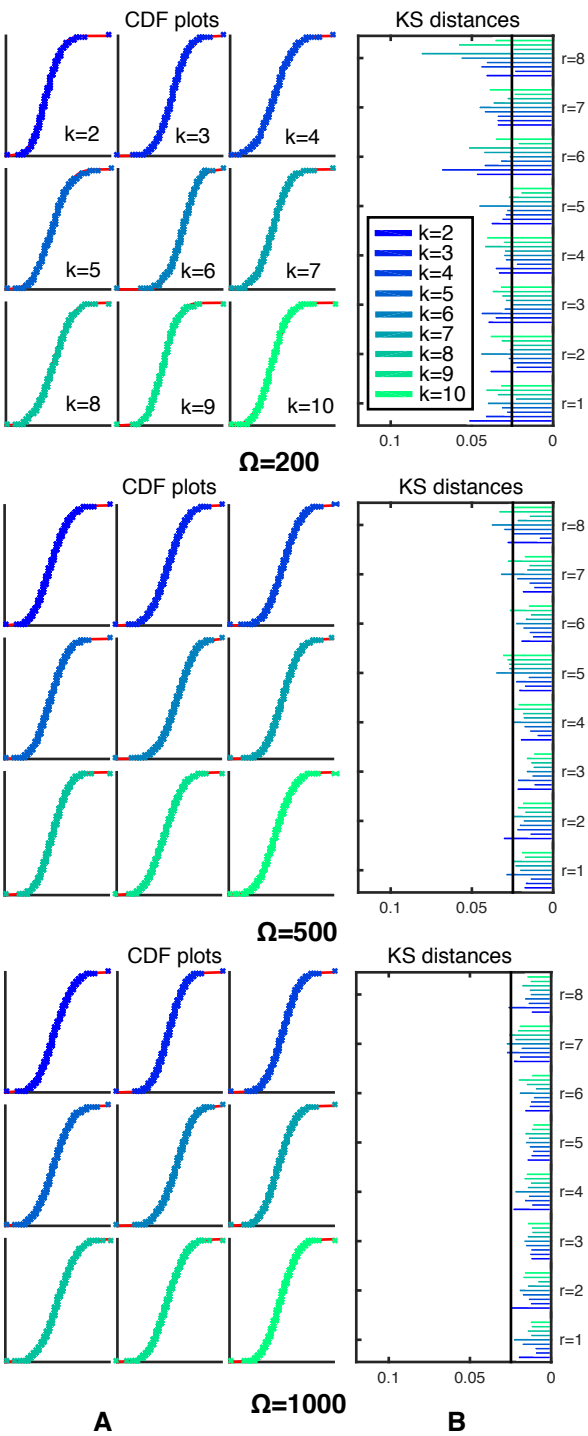

Figure E: Comparison of pcLNA and exact transversal distributions in systems sizes $\Omega = 200$ (top panel), $\Omega = 500$ (middle panel), $\Omega = 1000$ (bottom panel).(A) CDF plots under the pcLNA (red line) and the SSA (empirical CDF, crosses) of the first intersection in the r-th pass, $Q_k^{(r)}$, in transversal coordinates $k = 2, 3, \ldots, 10$ for $r = 4$. (B) KS distances between the $Q_k^{(r)}$ distributions under pcLNA and SSA, $r = 1, 2, \ldots, 8$, $k = 2, 3, \ldots, 10$.

5 Less frequent correction

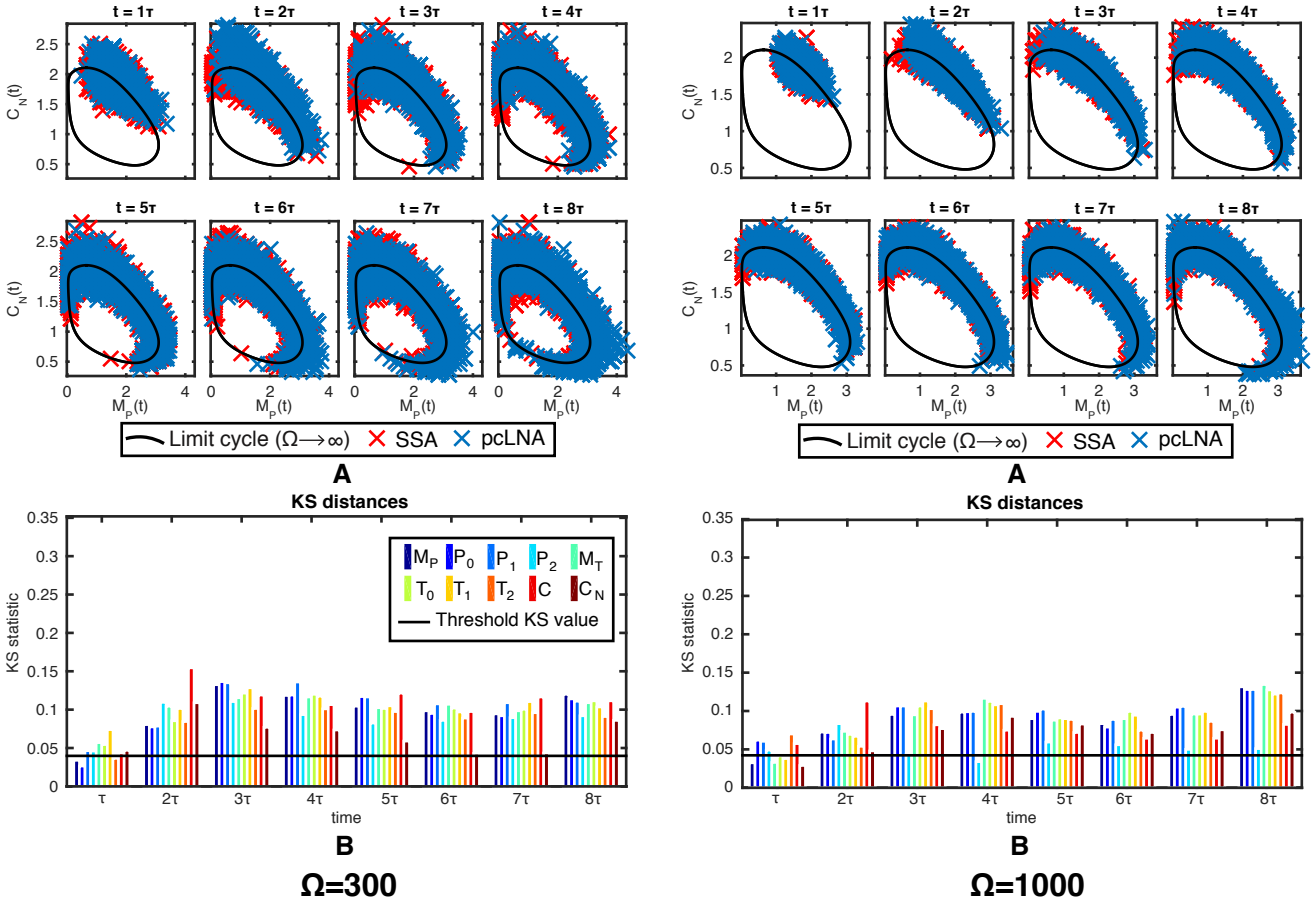

Figure F: Comparison between pcLNA and SSA simulations for correction frequency $24h$. This is 3 times less frequent than the correction used in the simulations presented in Fig. 9 of **I** and the figures in the next section. Panels (A) contain the samples produced by the pcLNA and the exact simulation (SSA) at time-points, $t = 1\tau, 2\tau, \ldots, 8\tau$, for $\Omega = 300$ (left panel) and $\Omega = 1000$ (right panel), and panels (B) the KS distances between the empirical distributions of pcLNA and the SSA for each system variable (colored bars) for two system sizes $\Omega = 300$ (left panel) and $\Omega = 1000$ (right panel).

Less frequent phase correction in pcLNA algorithm has substantial impact on pcLNA simulations. However, the precision of the approximation seems to stabilise after the third round with KS distances much smaller compared to standard LNA. The median CPU times under this correction frequency are $0.45secs$ for $\Omega = 300$ and $0.22secs$ for $\Omega = 1000$, compared to $0.45secs$ and $0.42secs$, respectively, for correction frequency $6h$. The reason for the absence of speed improvement for $\Omega = 300$ is because at this system size and with less corrections negative populations appear more frequently and therefore SSA simulation are used more frequently, as explained in Sec. 13 in **S**1, and therefore any improvements due to applying correction less frequently are counterbalanced. For $\Omega = 1000$, where negative populations are much less likely the speed improvement, is almost twofold.

6 Comparison between different simulation algorithms

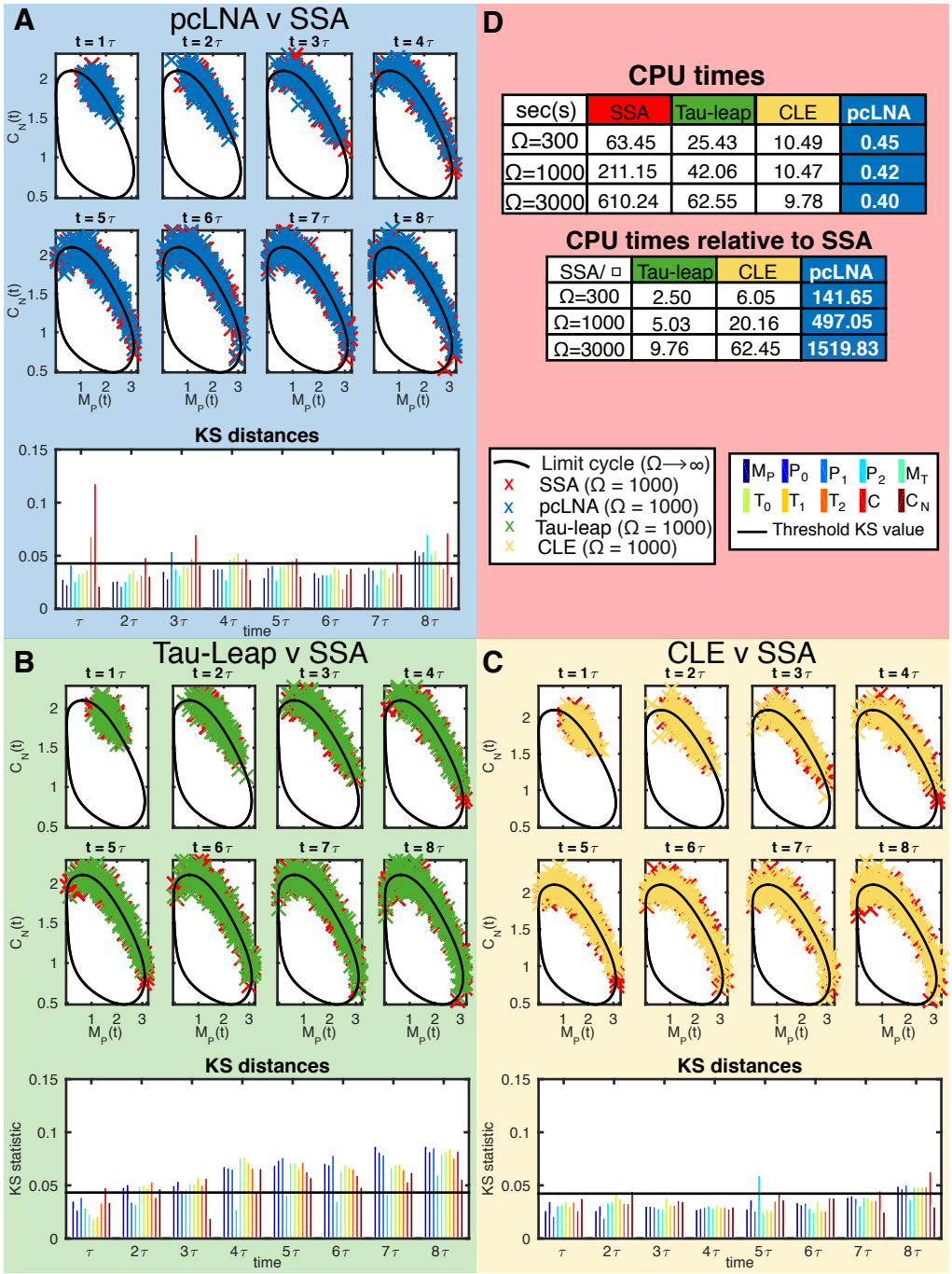

Figure G: Comparison between pcLNA, tau-leap and CLE simulation algorithms for the *Drosophila* circadian clock. This figure has the same form as Fig. 9 of **I**, but (A), (B) and (C) panels contain results for $\Omega = 1000$. As in Fig. 9 of **I**, the parameter values $\epsilon = 0.002$ and $\Delta t = 0.002$ for the tau-leap and the CLE approximation are the largest values to achieve precision similar to the pcLNA simulation, and the results displayed here are simulated using these values.

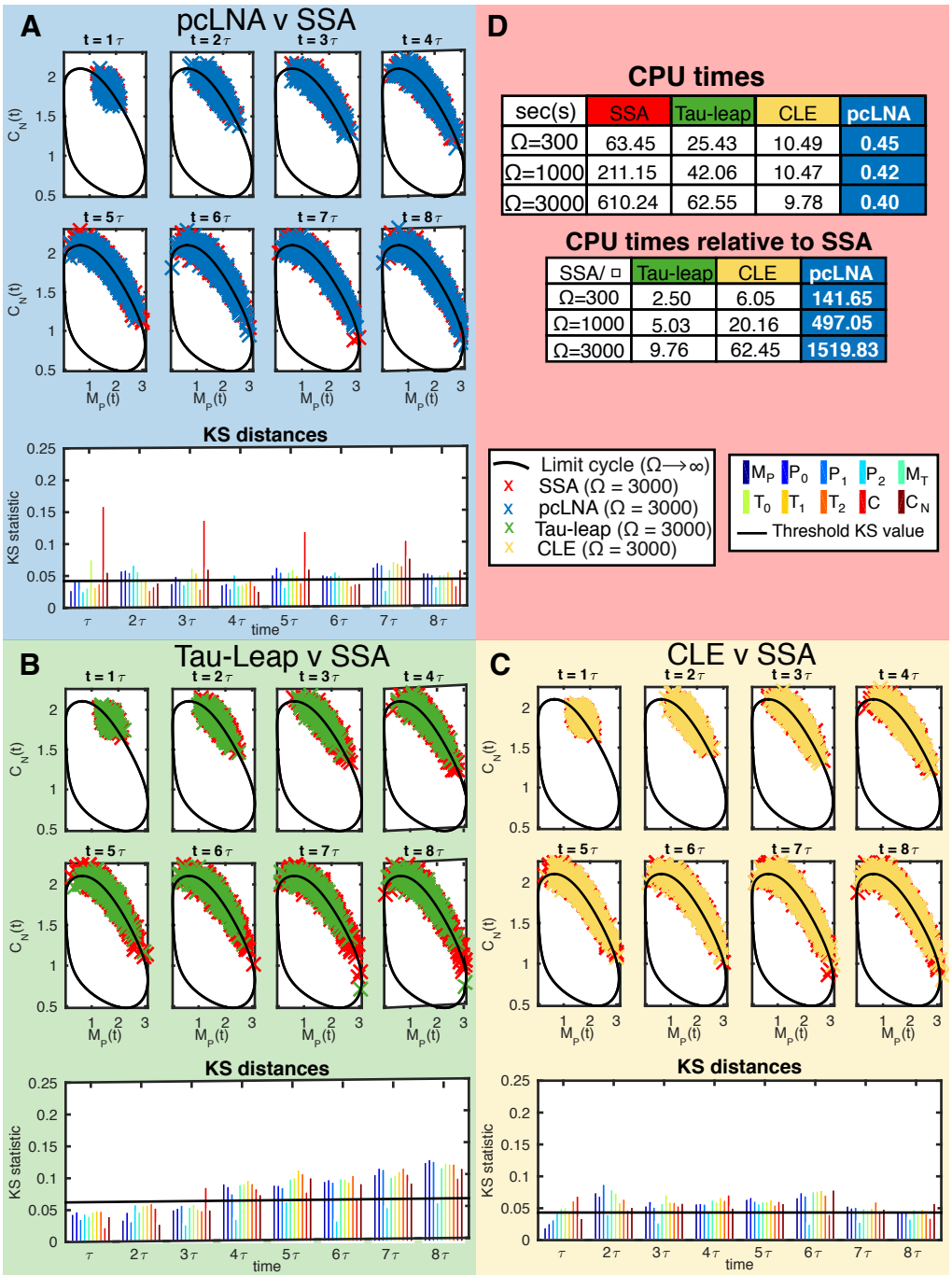

Figure H: Comparison between pcLNA, tau-leap and CLE simulation algorithms for the *Drosophila* circadian clock. This figure has the same form as Fig. 9 of **I**, but (A), (B) and (C) panels contain results for $\Omega = 3000$. As in Fig. 9 of **I**, the parameter values $\epsilon = 0.002$ and $\Delta t = 0.002$ for the tau-leap and the CLE approximation are the largest values to achieve precision similar to the pcLNA simulation, and the results displayed here are simulated using these values.

7 Light entrained *D*rosophila Circadian Clock

The Light entrained *Drosophila* circadian clock system first proposed in [2] is exactly the same with the *Drosophila* circadian clock without entrainment, apart from the degradation rate, v_{dT}, of the phosphorylated TIM protein, T_2, which instead of being equal to 3.4, it is equal to 4 during the day-time (6-18h) and equal to 2 during night-time (18h-6h).

We first compare the LNA distributions at fixed times with the empirical distributions derived using SSA simulations. As we can see in Fig. I, despite that the empirical distributions of the exact stochastic model do not spread along the curved limit cycle as much as the corresponding distributions for the system without entrainment (see Fig. 3 of **I**) and that the spread of the distributions in the light entrained system appears to stabilise even after the second cycle, the standard LNA distributions still do not approximate well these distributions.

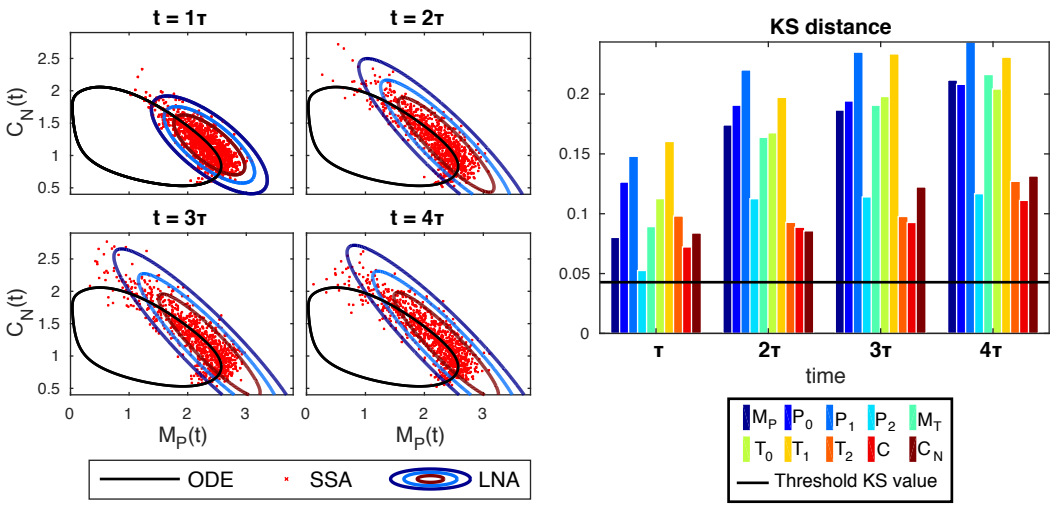

Figure I: Comparison between LNA and exact simulations in the light entrained *Drosophila* circadian clock. (a) Samples (in nanomolar concentrations) obtained from the SSA simulation algorithm (red crosses) and $0.01, 0.05, 0.40$ contours of the LNA probability density (black ellipsoids) at fixed times, $t = \tau, 2\tau, 3\tau, 4\tau$ (τ: minimal period). The limit cycle ODE solution is also displayed (black solid line). (b) KS distance between the empirical distribution of SSA samples and the LNA distribution of each species (different colors, see legend) at the fixed times. The threshold level is also displayed (black solid line). The system size is $\Omega = 300$.

On contrary, as we can see in Fig. J, the pcLNA distributions appear to approximate well the empirical transversal distributions derived under the SSA.

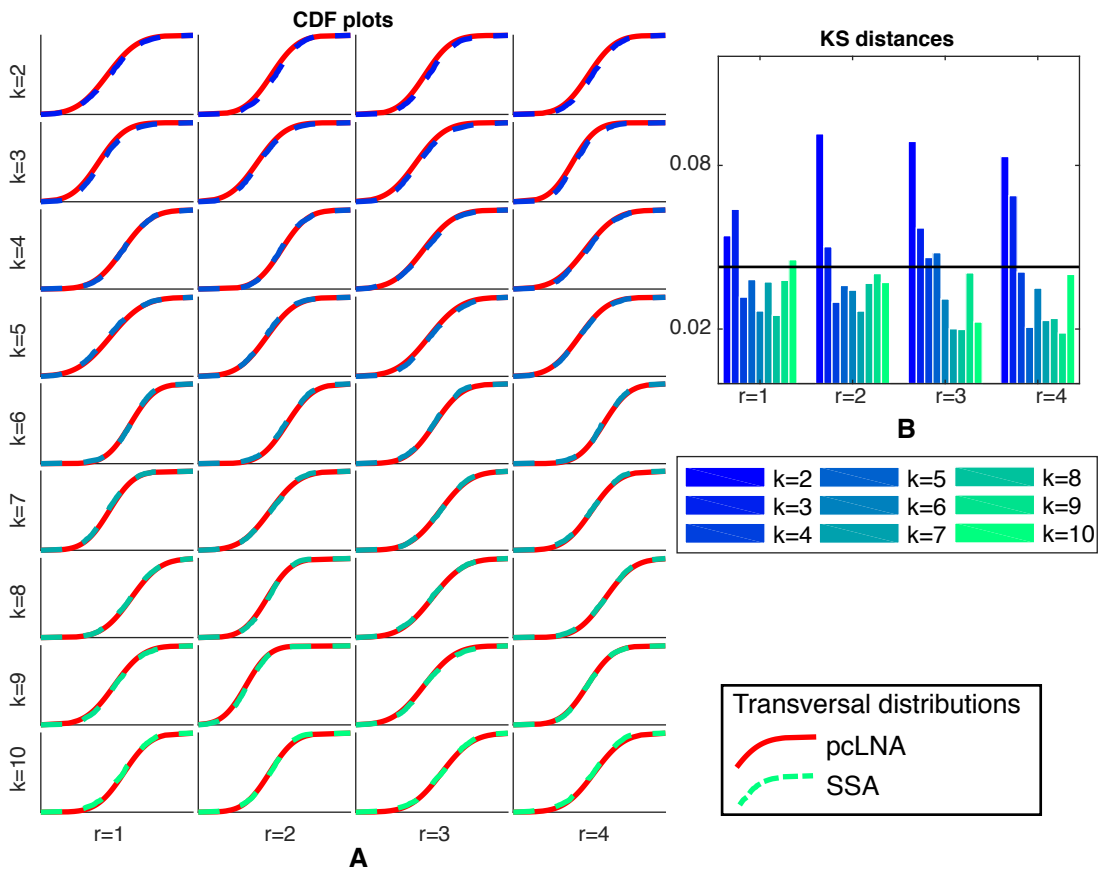

Figure J: Comparison of pcLNA and exact empirical transversal distributions of the light entrained *Drosophila* circadian clock. (A) CDF plots of the transversal distributions $Q_k^{(r)}$ under the pcLNA (red line) and the SSA (empirical CDF, colored dashed line, see legend) in transversal coordinates $k = 2, 3, \ldots, 10$ and round $r = 1, 2, 3, 4$. (B) KS distances between the $Q_k^{(r)}$ distributions under the pcLNA and the SSA, $k = 2, 3, \ldots, 10$, $r = 1, 2, 3, 4$. The system size is $\Omega = 300$.

References

[1] Gonze D, Halloy J, Leloup JC, Goldbeter A. Stochastic model for circadian rhythms: effect of molecular noise on periodic and chaotic behaviour. Comptes Rendus Biologies 2003 Apr;326(2):189–203.

[2] Leloup JC & Goldbeter A. A Model for Circadian Rhythms in Drosophila Incorporating the Formation of a Complex between the PER and TIM Proteins. Journal of Biological Rhythms 1998 Feb; 13(1): 70-87.

Supplementary information S3 for the paper Long-time analytic approximation of large stochastic oscillators: simulation, analysis and inference

Giorgos Minas[1,2] & David A Rand[1,2]

[1] Zeeman Institute for Systems Biology & Infectious Disease Epidemiology Research & [2]Mathematics Institute,

University of Warwick, Coventry CV4 7AL, UK

July 22, 2017

Contents

1 Brusselator system **1**

1.1 Fixed point system 2

2 NF-κB system **3**

2.1 The NF-κB response to TNFα activation 9

Abstract

We refer to the paper "Long-time analytic approximation of large stochastic oscillators: simulation, analysis and inference" by **I**. In this note we give further illustrations for the Brusselator and the NF-κB system.

1 Brusselator system

The ODE system of the Brusselator is

$$\begin{aligned}\dot{A} &= 1 - A(1 + b - cAB),\\ \dot{B} &= A(b - cAB).\end{aligned}$$

For the limit cycle ODE solution, we use the initial conditions $A(t_0) = 0.75$, $B(t_0) = 2.00$, and the parameter values $b = 2.20$, $c = 1.00$. We use the SSA to exactly simulate the system

and produce $R = 3000$ samples of stochastic trajectories (see Figure A) for a time-length of $8.5 \times \tau$, where $\tau \approx 6.37$ is the period of the periodic solution of the ODE. The rates of the reactions used for the SSA of the Brusselator are provided in Table A.

reaction	rate
$\emptyset \xrightarrow{1} A$	Ω
$A \xrightarrow{1} \emptyset$	A
$A \xrightarrow{b} B$	bA
$B \xrightarrow{c} A$	cA^2B/Ω^2

Table A: Reactions of the Brusselator system and their rates

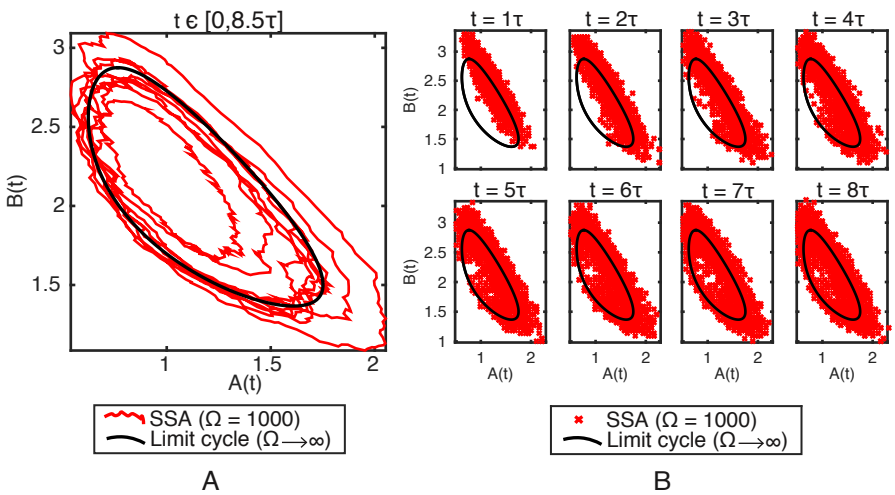

Figure A: Exact stochastic simulation of the Brusselator system. (A) A stochastic trajectory of $X(t) = Y(t)/\Omega$ obtained by running the SSA over the time-interval $t \in [0, 8.5\tau]$ and (B) SSA samples ($R = 3000$) at times $t = \tau, 2\tau, \ldots, 8\tau$. The volume size is $\Omega = 1000$. The black solid curve is the large volume, $\Omega \to \infty$, limit cycle solution.

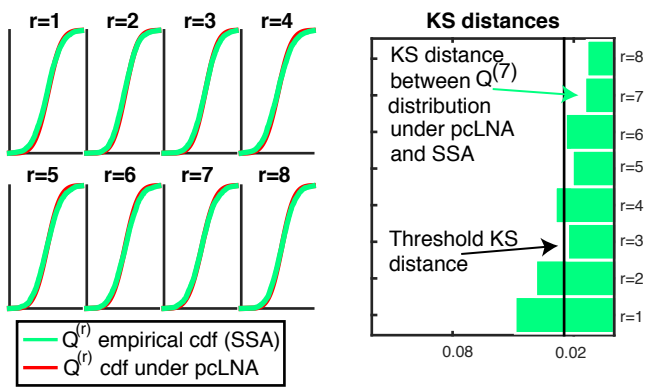

Figure B: Comparison of pcLNA and exact empirical transversal distributions. (A) CDF plots of the (one-dim) transversal distributions $Q^{(r)}$ under the pcLNA (red line) and the SSA (empirical CDF, crosses) in round $r = 1, 2, \ldots, 8$. (B) KS distances between the $Q_k^{(r)}$ distributions under pcLNA and SSA, $r = 1, 2, \ldots, 8$.

1.1 Fixed point system

For comparisons, we also consider a set of parameter values that give a limiting system ($\Omega \to \infty$) with an equilibrium fixed point instead of a limit cycle. Here, the initial conditions are $A(t_0) = 1.156$, $B(t_0) = 1.461$, and the parameter values $b = 0.25$, $c = 0.25$, which give an equilibrium point $A_{eq} = B_{eq} = 1$ with Jacobian matrix of the system (referred as linearisation matrix in **I**) with both eigenvalues equal to -0.5.

We use the SSA to exactly simulate this system and produce $R = 2000$ samples of stochastic trajectories (see Fig. CA) for a time-length of 700, which is somewhat larger than 100τ, where τ the period of the limit cycle ODE solution considered above. In Fig. CB we provide the empirical CDF plots of the SSA at time $637 \approx 100\tau$ and compare them with the LNA distributions at the same time.

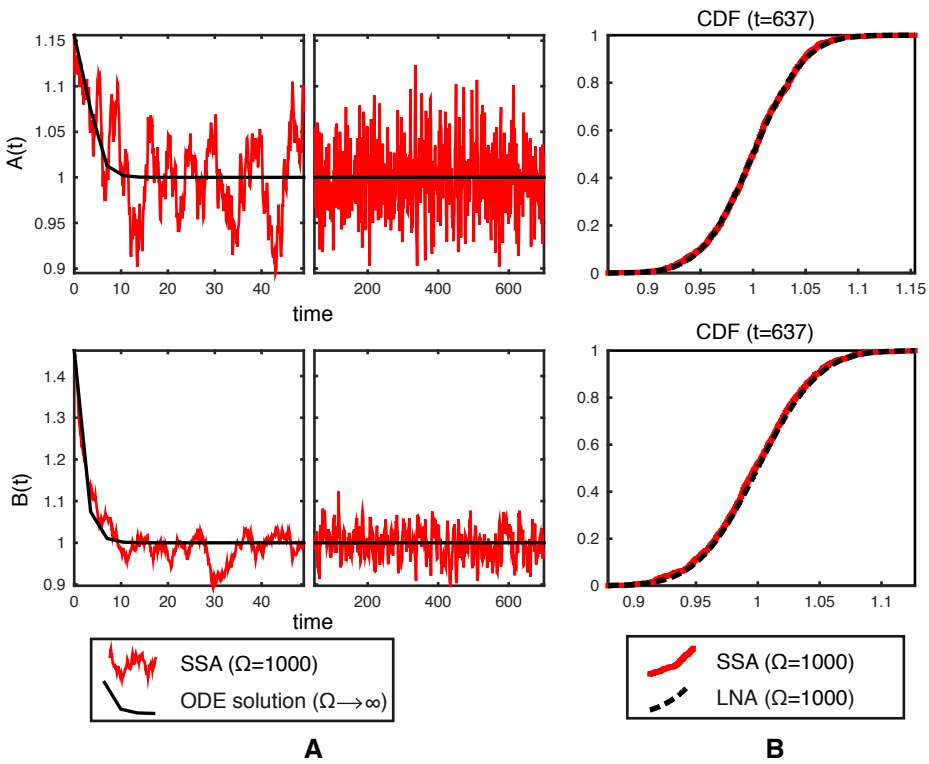

Figure C: Exact stochastic simulation and comparison to the LNA for the fixed point Brusselator system. The parameter values are $b = c = 0.25$ and the system size $\Omega = 1000$. (A) A stochastic trajectory of $X(t) = Y(t)/\Omega$ for the two variables A (top panel) and B (bottom panel) of the Brusselator system obtained by running the SSA over the time-interval $t \in [0, 700]$ and its large volume, $\Omega \to \infty$, solution. (B) Empirical CDF plots of the SSA samples ($R = 2000$) and CDF plots of the LNA at time $t = 637$ for variables A (top panel) and B (bottom panel).

2 NF-κB system

The NF-κB system model used in this SI describes the oscillatory response of the system following stimulation by tumor necrosis factor alpha (TNFα). With no stimulation the system

remains in a stable equilbrium state. Following TNFα stimulation that is constant in time the system responds by a transient pulse followed by relaxation to a stable limit cycle.

We first consider the case where the system starts in the stable limit cycle solution, and in the next section consider the biologically interesting situation where the signal is received while the system is inactive and the initial oscillation is transitory. The species of the system and the initial conditions (in micromolar concentrations) used to derive the limit cycle deterministic solution are provided in Table B.

	name	description	initial value
1	N_c	Free Cytoplasmic NFkB	0.0056
2	I_c	Free Cytoplasmic IkBa	0.0298
3	NI_c	Cytoplasmic NFkB-IkBa	0.0723
4	N_n	Free nuclear NFkB	0.0011
5	I_n	Free nuclear IkBa	0.0075
6	NI_n	Nuclear NFkB-IkBa	0.0013
7	I_m	IkBa transcription	0.0004
8	K_n	Kinase IKKn	0.0020
9	K_a	Kinase IKKa	0.0024
10	A_m	A20 transcription	0.0001
11	A	A20	0.0076

Table B: The variables of NF-κB system and the initial conditions (in concentrations) used to derive the limit cycle ODE solution.

The system model considered here is a slight modification of the system model in [1]. In particular, some variables are omitted because the sum of the concentrations of all forms of NF-κB and IKK are both fixed and the level of phosphorylated IkBa is an end state that plays no further role in the model. Therefore, the redundant species IKKi, phosphorylated IkBα and phosphorylated NFκB-IκBα are removed to avoid rank-deficiencies in the covariance matrices of the LNA.

We also write the system in a form where concentrations are all written in terms of the same volume i.e. the total cell volume. The original system in [1] is written in cytoplasmic concentrations for all species except, N_n, I_n and NI_n that are written in nuclear concentrations and are set to be 3.3 larger than cytoplasmic concentrations. The nuclear and cyctoplasmic concentrations can easily be recovered from our concentration if one knows the ratio of nuclear to cytoplasmic volume.

We used the same parameter values as in [1] to derive the ODE solution. These are provided in Table C.

parameter	description	value	measurement unit
k_v	C:N ratio	3.300000	–
k_p	IKKn production	1.3030e-04	s^{-1}
k_a	Activation caused by TNFa	0.003	s^{-1}
k_i	Spontaneous IKK activation	0.002	s^{-1}
k_{a1a}	NFkB-IkBa association	0.5	$\mu M^{-1}s^{-1}$
k_{d1a}	NFkB-IkBa dissociation	0.0005	s^{-1}
k_{c1a}	Catalysis of IKK-IkBa dimer	0.037	s^{-1}
k_{c2a}	Catalysis of IKK-IkBa-NFkB trimer	0.185	s^{-1}
k_{t1a}	degradation of IkBa (IKK dependent from dimer)	0.025	s^{-1}
k_{t2a}	degradation of IkBa (IKK dependent from trimer)	0.025	s^{-1}
c_{4a}	Free IkBa degradation	0.00019	s^{-1}
c_{5a}	NFkB complexed IkBa degradation	0.000022	s^{-1}
k_{i1}	NFkB nuclear import	0.003	s^{-1}
k_{e1}	NFkB nuclear export	6.0000e-05	s^{-1}
k_{e2a}	NFkB-IkBa nuclear export	0.012	s^{-1}
k_{i3a}	IkBa nuclear import	0.001	s^{-1}
k_{e3a}	IkBa nuclear export	0.0005	s^{-1}
h	Order of hill function	2	–
k	Hill constant	0.0430	–
c_{1a}	IkBa mRNA synthesis	1.4200e-07	$\mu M^{-1}s^{-1}$
c_{2a}	IkBa translation rate	0.5	s^{-1}
c_{3a}	IkBa mRNA degradation	0.00048	s^{-1}
c_1	IkBa mRNA synthesis	1.4200e-07	$\mu M^{-1}s^{-1}$
c_2	A20 mRNA translation	0.5	s^{-1}
c_3	A20 mRNA degradation	0.00048	s^{-1}
c_4	A20 degradation	0.0045	s^{-1}
k_{bA20}	Half-max A20 inhibition concentration	0.0018	–
TNFα	Tumor necrosis factor alpha level	100	ng/ml
TNF-κB	total NF-κB concentration	0.08	μM
$TIKK$	total IKK concentration	0.08	μM

Table C: The parameters of NF-κB system and the values used to derive their ODE solution.

The ODE system for the NF-κB system considered in this SI is:

$$\dot{N}_c = k_{d1a}NI_c - k_{a1a}N_cI_c - k_{i1}N_c + c_{5a}NI_c + k_vk_{e1}N_n + k_{t2a} \times (TNFKB - N_c - NI_c - N_n + NI_n)$$

$$\dot{I}_c = k_{d1a}NI_c - k_{a1a}N_cI_c - k_{i3a}I_c + k_vk_{e3a}I_n - c_{4a}I_c + c_{2a}I_m - k_{c1a}K_aI_c$$

$$\dot{NI}_c = k_{a1a}N_cI_c - k_{d1a}NI_c + k_vk_{e2a}NI_n - c_{5a}NI_c - k_{c2a}K_aNI_c$$

$$\dot{N}_n = k_{d1a}NI_n - k_vk_{a1a}N_nI_n + k_{i1}N_c - k_vk_{e1}N_n$$

$$\dot{I}_n = k_{d1a}NI_n - k_vk_{a1a}N_nI_n + k_{i3a}I_c - k_vk_{e3a}I_n - c_{4a}I_n$$

$$\dot{NI}_n = k_vk_{a1a}NI_n - k_{d1a}NI_n - k_vk_{e2a}NI_n$$

$$\dot{I}_m = c_{1a}(N_n^h/(N_n^h + (k/k_v)^h)) - c_{3a}I_m$$

$$\dot{K}_n = k_p(TIKK - K_n - K_a)(k_{bA20}/(k_{bA20} + A \times TNF\alpha)) - k_aTNF\alpha K_n$$

$$\dot{K}_a = k_aTNF\alpha K_n - k_iK_a$$

$$\dot{A}_m = c_1(N_n^h/(N_n^h + (k/k_v)^h)) - c_3A_m$$

$$\dot{A} = c_2A_m - c_4A$$

The reaction rates used for the SSA are provided in Table D. The values of the parameters are the same as in Table C.

reaction	rate
$N_c + I_c \xrightarrow{k_{a1a}} NI_c$	$k_{a1a} \times I_c \times N_c/\Omega$
$NI_c \xrightarrow{k_{d1a}} N_c + I_c$	$k_{d1a} \times NI_c$
$N_n + I_n \xrightarrow{k_{a1a}} NI_n$	$k_v k_{a1a} \times I_n \times N_n/\Omega$
$NI_n \xrightarrow{k_{d1a}} N_n + I_n$	$k_{d1an} \times NI_n$
$K_a + I_c \xrightarrow{k_{c1a}} K_a + I_p$	$k_{c1a} \times K_a \times I_c/\Omega$
$K_a + NI_c \xrightarrow{k_{c2a}} K_a + NI_p$	$k_{c2a} \times K_a \times NI_c/\Omega$
$NI_p \xrightarrow{k_{t2a}} N_c$	$k_{t2a} \times NI_p$
$N_c \xrightarrow{k_{i1}} N_n$	$k_{i1} \times N_c$
$N_n \xrightarrow{k_{e1}} N_c$	$k_{e1} \times k_v \times N_n$
$NI_n \xrightarrow{k_{e2a}} NI_c$	$k_{e2a} \times k_v \times NI_n$
$I_c \xrightarrow{k_{i3a}} I_n$	$k_{i3a} \times I_c$
$I_n \xrightarrow{k_{e3a}} I_c$	$k_{e3a} \times k_v \times I_c$
$\varnothing \xrightarrow{H_I} I_m$	$(c_{1a}\Omega)\frac{N_n^h}{N_n^h+(k\Omega/k_v)^h}$
$I_m \xrightarrow{c_{2a}} I_m + I_c$	$c_{2a} \times I_m$
$I_m \xrightarrow{c_{3a}} \varnothing$	$c_{3a} \times I_m$
$I_c \xrightarrow{c_{4a}} \varnothing$	$c_{4a} \times I_c$
$I_n \xrightarrow{c_{4a}} \varnothing$	$c_{4a} \times I_n$
$NI_c \xrightarrow{c_{5a}} N_c$	$c_{5a} \times NI_c$
$\varnothing \xrightarrow{H_A} A_m$	$(c_1\Omega)\frac{N_n^h}{N_n^h+(k\Omega/k_v)^h}$
$A_m \xrightarrow{c_2} A_m + A$ v $A_m \xrightarrow{c_3} \varnothing$	$c_2 \times A_m$ $c_3 \times A_m$
$A \xrightarrow{c_4} \varnothing$	$c_4 \times A$
$K_i \xrightarrow{M_A} K_n$	$k_p(TIKK - K_n - K_a)\frac{k_{bA20}\times\Omega}{(k_{bA20}\Omega)+A\times TNF\alpha}$
$K_n \xrightarrow{TNF\alpha\times k_a} K_a$	$TNF\alpha \times k_a \times K_n$
$K_a \xrightarrow{k_i} K_i$	$k_i \times K_a$

Table D: Reactions of the NF-κB system and their rates

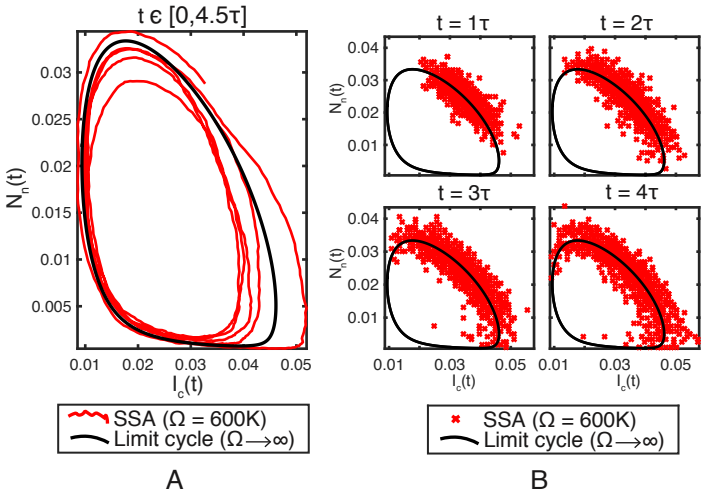

Figure D: Exact stochastic simulation of the NF-κB system. (A) A stochastic trajectory (in concentrations) obtained by running the SSA over the time-interval $t \in [0, 4.5\tau]$ and (B) SSA samples ($R = 3000$) at times $t = \tau, 2\tau, 3\tau, 4\tau$. Two (out of 11) of the species are displayed (IκBα I_C (x-axis) and nuclear NF-κB N_n (y-axis)). The volume size is $\Omega = 0.6M$. This is substantially smaller than the system size used in [1] ($\Omega = 1.25M$) to provide higher noise levels. The black solid curve is the large volume, $\Omega \to \infty$, limit cycle solution.

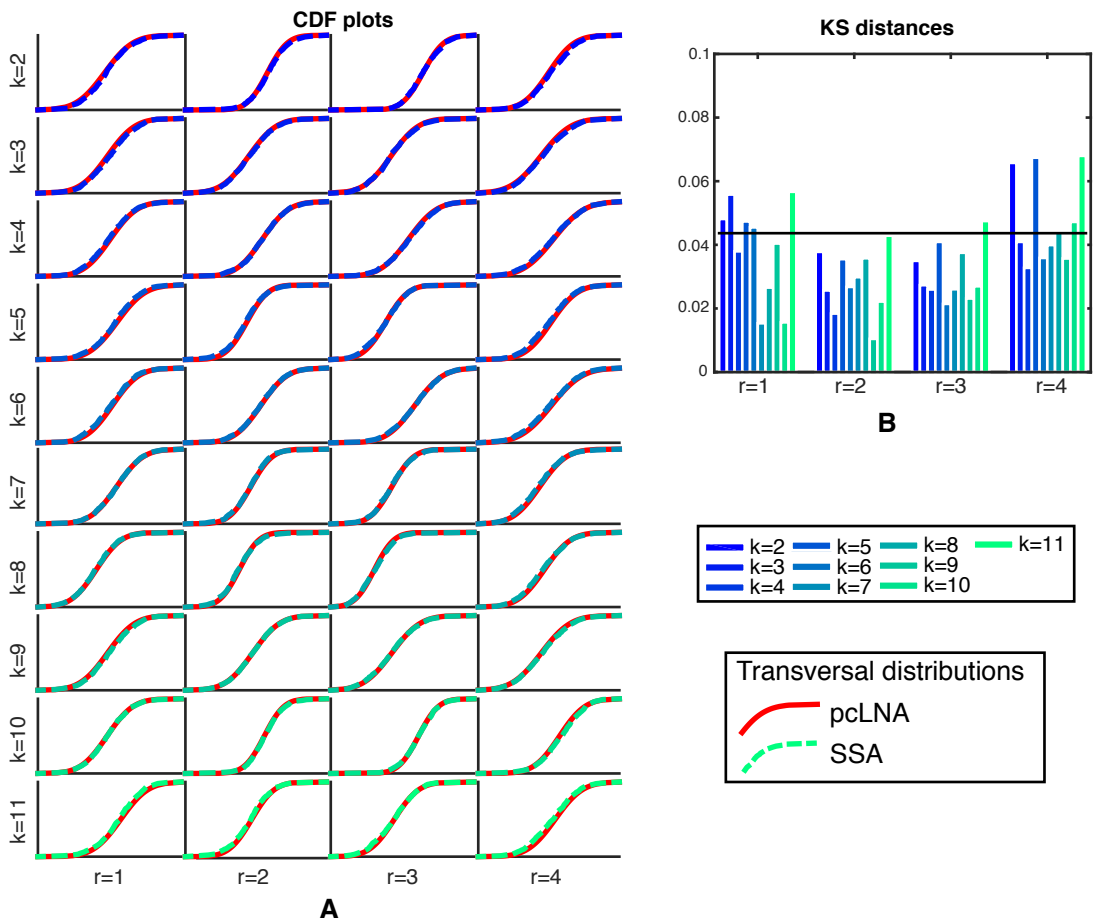

Figure E: Comparison of pcLNA and exact transversal distributions. (A) CDF plots of the transversal distributions $Q_k^{(r)}$ under the pcLNA (red line) and the SSA (empirical CDF, colored dashed line, see legend) in transversal coordinates $k = 2, 3, \ldots, 11$ and round $r = 1, 2, 3, 4$. (B) KS distances between the $Q_k^{(r)}$ distributions under pcLNA and SSA, $k = 2, 3, \ldots, 11$, $r = 1, 2, 3, 4$. The system size is $\Omega = 0.6M$.

2.1 The NF-κB response to TNFα activation

Here we study the response of the *NF-κB* system to a continuous TNFα signal received while the system being in the equilibrium fixed point state. This equilibrium state which is given in Table E is therefore the initial conditions used to derive the deterministic solution provided in Fig. F.

	name	initial value
1	N_c	0.0036
2	I_c	0.0160
3	NI_c	0.0719
4	N_n	0.0042
5	I_n	0.0013
6	NI_n	0.0003
7	I_m	0.0000
8	K_n	0.0800
9	K_a	0
10	A_m	0.0000
11	A	0.0014

Table E: The initial conditions (in concentrations) used to derive the ODE solution in Fig 6.

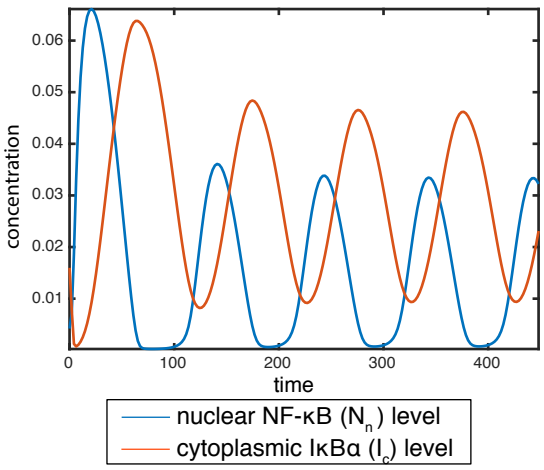

Figure F: The limiting ($\Omega \to \infty$) deterministic solution of the NF-κB system. The initial conditions are provided in Table E. Two species of the system are displayed.

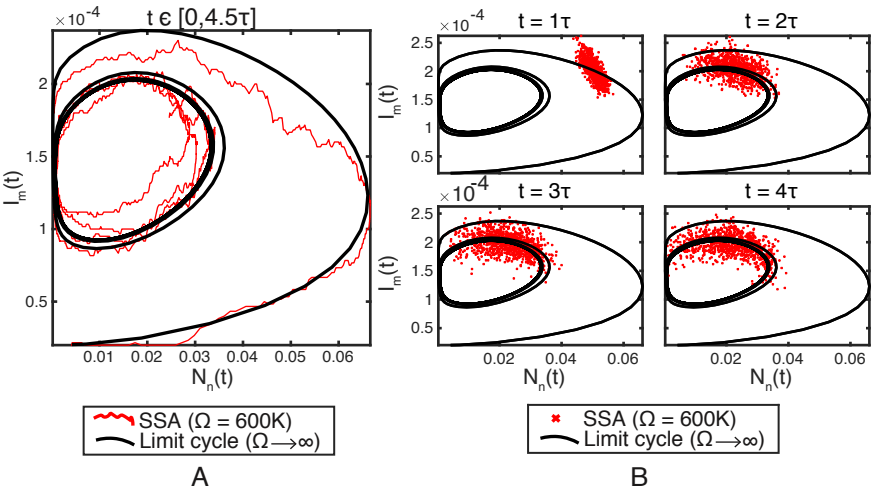

Figure G: Exact stochastic simulation of the NF-κB system. (A) A stochastic trajectory (in concentrations) obtained by running the SSA over the time-interval $t \in [0, 4.5\tau]$ and (B) SSA samples ($R = 3000$) at times $t = \tau, 2\tau, 3\tau, 4\tau$. Two (out of 11) of the species are displayed (nuclear NF-κB (x-axis) and IκBα mRNA (y-axis)). The volume size is $\Omega = 0.6M$. The black solid curve is the large volume, $\Omega \to \infty$, limit cycle solution.

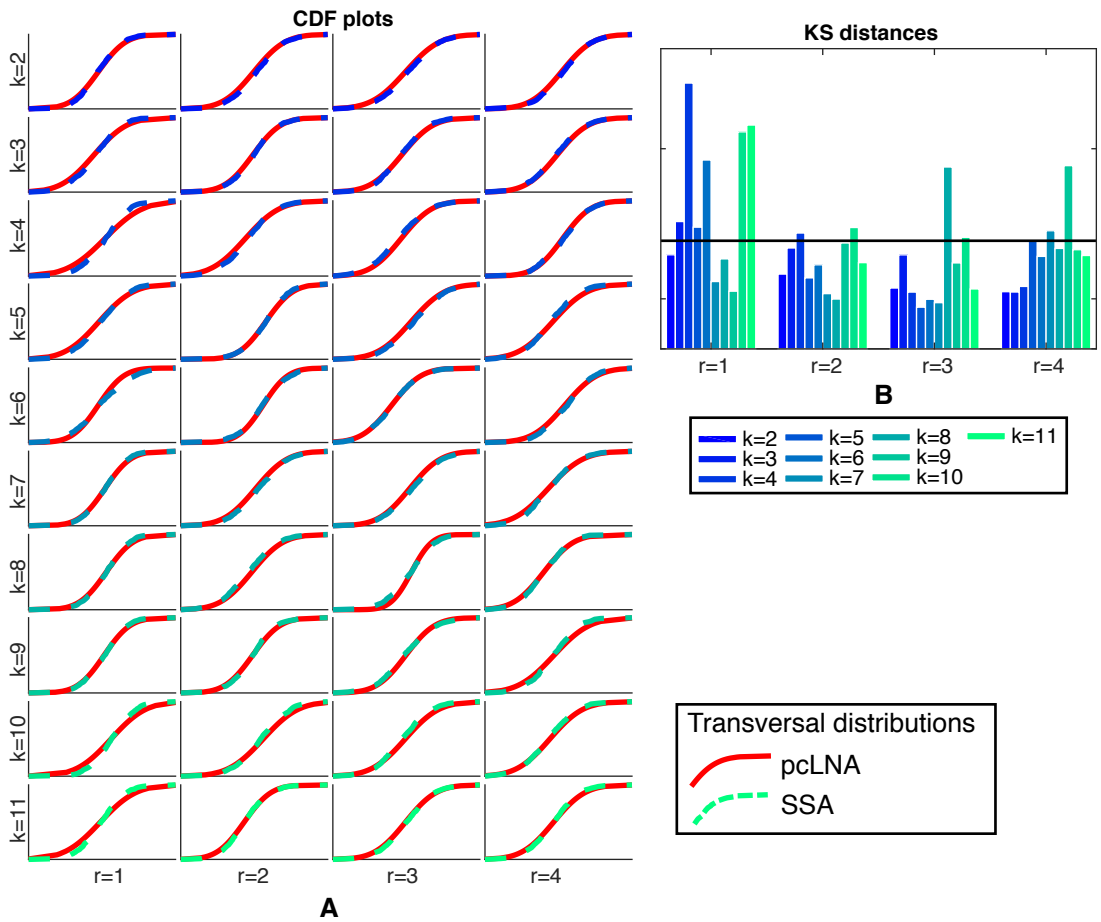

Figure H: Comparison of pcLNA and exact transversal distributions. (A) CDF plots of the transversal distributions $Q_k^{(r)}$ under the pcLNA (red line) and the SSA (empirical CDF, colored dashed line, see legend) in transversal coordinates $k = 2, 3, \ldots, 11$ and round $r = 1, 2, 3, 4$. (B) KS distances between the $Q_k^{(r)}$ distributions under pcLNA and SSA, $k = 2, 3, \ldots, 11$, $r = 1, 2, 3, 4$. The system size is $\Omega = 0.6M$.

References

[1] Ashall L, Horton CA, Nelson DE, Paszek P, Harper CV, Sillitoe K, Ryan S, Spiller DG, Unitt JF, Broomhead DS, Kell DB, Rand DA, Sée V, White MRH. Pulsatile stimulation determines timing and specificity of NF-kappa B-dependent transcription. Science. 2009 Apr;324(5924);242–6.