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

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

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

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

*to whom correspondence may be addressed

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 Ω. Its large volume ($\Omega \to \infty$) validity has been shown in [6], in the sense that the distribution of the Markov jump process at a fixed finite time converges, as the volume 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, as we show below, 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. Linearly stable means that the eigenvalues of the system linearised at the equilibrium point have negative real parts. For example, in Fig. 3 of **S**3 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 behavior is also observed in other systems and using different measures in [7].

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 if 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 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. [8] which uses the 2-dimensional Brusselator system as an exemplar to investigate the failure of the LNA in approximating long-term behavior 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. Other papers also employ a similar analysis to study the temporal variability of oscillatory systems in the tangental direction of the $\Omega \to \infty$ limit cycle [9] and/or the amplitude variability in the transversal direction of the limit cycle [10, 11, 12], using various low-dimensional oscillatory systems for illustration.

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 also for prediction. The approach in [8] uses transversal sections which are normal to the limit cycle. We follow this but in the supplementary information (Sect. 9 **S**1) we show that for most considerations one can use any transversal to the limit cycle.

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. [13] (see Sect. 1 **S**2). This model involves 10 state variables and 30 reactions. 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 is transported into the cytosol where it is degraded and translated into protein (P_0 and T_0). These proteins are multiply phosphorylated (PER: $P_0 \to P_1 \to P_2$; TIM: $T_0 \to T_1 \to T_2$) and these modifications can be reversed by a phosphatase. The fully phosphorylated form of the proteins is targeted for degradation and forms a complex, which is transported into the nucleus in a reversible manner where the nuclear form of the PER–TIM complex represses the transcription of *per* and *tim* genes. The large system limit is given by the differential equation system of 10 kinetic equations that are listed in the supplementary information (see Sect. 1 **S**2) 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 [14] are also used to illustrate our methods and the results can be found in **S**3. The supplementary information **S**1 contains technical derivations and **S**2, **S**3 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 $Y(t)+\nu_j$ with the vectors $\nu_j \in \mathbb{R}^n$ 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 $Y(t)$. Such systems can be exactly simulated using the Stochastic Simulation algorithm (SSA) [1].

The classical deterministic equations that describe the evolution of the concentration vector $x(t)$ are derived by explicitly taking into account the volume Ω of the reaction system and assuming that the rates $w_j(Y(t))$ depend upon Ω as $w_j(Y) = u_j(Y/\Omega, \Omega)$. Here the dependence of u_j on Ω is of canonical form in the sense of [5]. The limiting system as $\Omega \to \infty$ is given by a differential equation of the form

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

where $u_j(x(t))$ the limiting ($\Omega \to \infty$) deterministic rates (see Sect. 3 **S**1). 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 in which all but one of the characteristic exponents of the limit cycle have negative real part [15] [16, Sect. 1.5].

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 $t \in [0, 8.5\tau]$ ($\tau \approx 26.98$ hours). Here $\Omega = 300$ is the volume of the system, imposing moderate to high levels of stochasticity. Results for system volumes $\Omega = 200$, 500 and 1000 are also reported in Fig. 2 **S**2. Figure 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 $N(x)$ normal to the limit cycle γ at $x \in \gamma$ (i.e. orthogonal to the tangent vector $F(x)$ at x) and the mapping G_N defined by $G_N(X) = x \in \gamma$

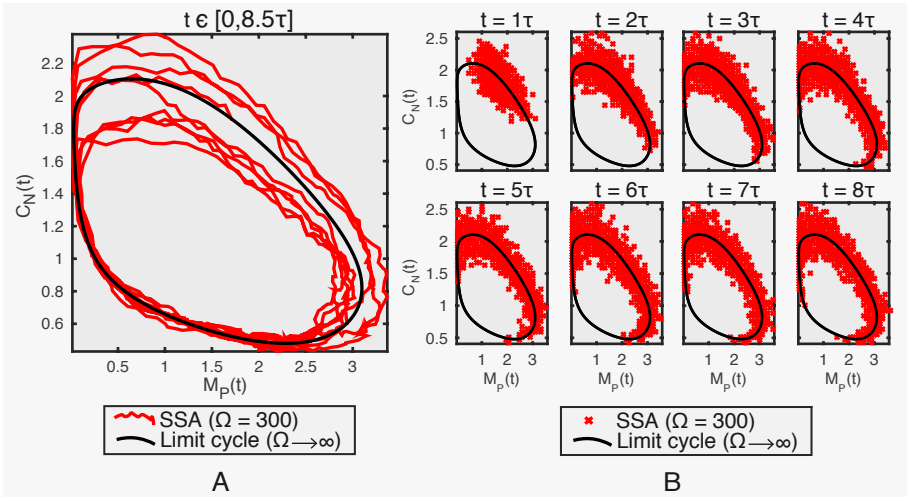

Figure 1: **Exact stochastic simulation of the drosophila circadian clock system**. (A) A stochastic trajectory obtained by running the SSA over the time-interval $t \in [0, 8.5\tau]$ and (B) SSA samples ($R = 3000$) at times $t = \tau, 2\tau, \ldots, 8\tau$. Two (out of 10) of the species are displayed (*per* mRNA (x-axis) and nuclear PER-TIM complex (y-axis)). The volume size is $\Omega = 300$. The black solid curve is the large volume, $\Omega \to \infty$, limit cycle solution.

if $X \in N(x)$. 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) and, as t grows, $G_N(X(t))$ will repeatedly pass by the value x every time the trajectory $X(t)$ passes through the hyperplane $N(X)$. In exact simulations, these intersection points, $Q_x^{(r)}$, of $X(t)$ with $N(x)$, in round $r = 1, 2, \ldots$ are derived by interpolation between the last state before and the first state after the r-th intersection (see Sect. 2 **S**1). These points of intersection describe the stochasticity of the system around a particular phase x of the system. 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 this 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 distributions are useful for analysing various aspects of the system.

Our first observation is that the empirical transversal distributions

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

obtained by exact simulation are approximately 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 Sect. 10 **S**1.

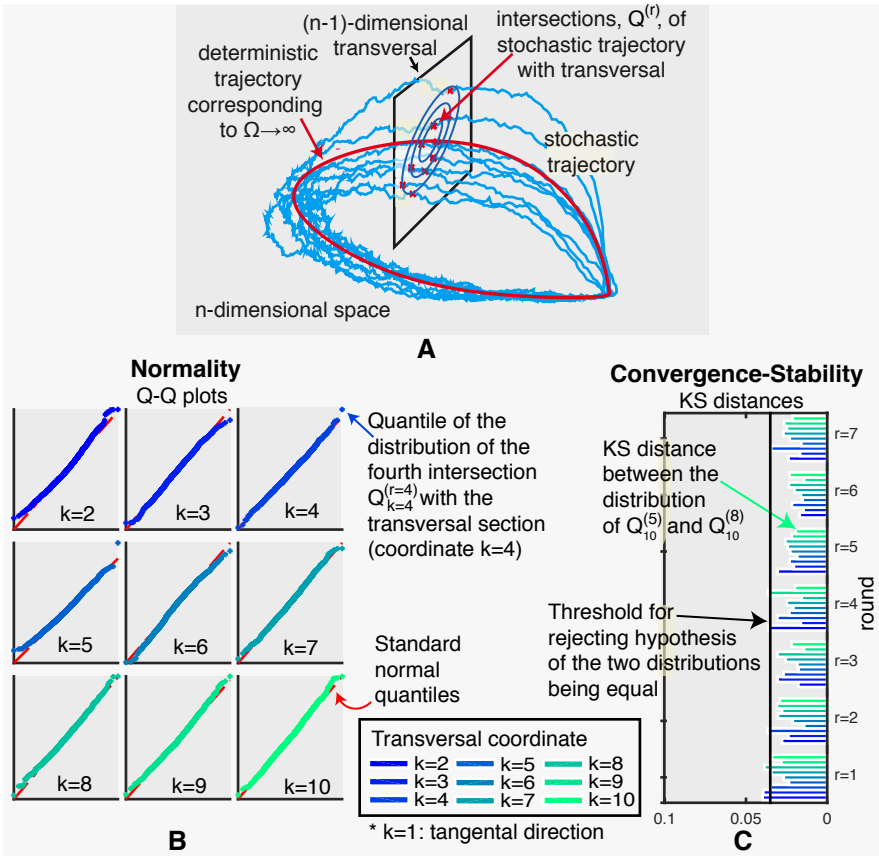

Figure 2: **Intersections of the stochastic trajectories with a transversal section.** (A) Schematic representation of the intersections. (B) Quantile-Quantile (Q-Q) plots of the distribution of the fourth intersection, $Q_k^{(r=4)}$, in transversal coordinates $k = 2, 3, \ldots, 10$ (see Fig. 3 **S**2 for $r = 1, 2, \ldots, 8$) . (C) KS distances between the empirical distribution of the intersections $Q_k^{(r)}$ for the first $r = 1, 2, \ldots, 7$ rounds of the stochastic trajectory and the distribution $Q_k^{(r=8)}$ in the final round ($r = 8$), $k = 2, 3, \ldots, 10$.

2.1 The Linear Noise Approximation (LNA)

This 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). 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 (SDE) that is independent of Ω, with drift and diffusion matrices that are functions of the deterministic solution $g(t)$. Details are given in Sect. 4 **S**1.

Given an initial time t_0 and an initial condition for $\xi(t_0)$ the LNA determines the distributions, 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 $P_{\text{LNA}}(\xi(t)|t_0, \xi_0 \sim P_0)$ and $P_{\text{LNA}}(X(t)|t_0, \xi_0 \sim P_0)$ (see Fig 4). Details of how to calculate these are given in Sect. 4 **S**1.

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 $t \to \infty$. This is certainly the case when the LNA is used with oscillators and we aim to overcome this 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 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 spread along the curved limit cycle.

On the other hand, as we saw earlier, the transversal distributions $P_{x_1}^{(r)} = P(Q_{x_1}^{(r)}|X(t_0) = x_0)$ are approximately normal. We next derive an approximation of $P_{x_1}^{(r)}$ under the LNA and show that it accurately approximates $P_{x_1}^{(r)}$ in the Drosophila circadian clock, Brusselator and NF-κB systems. Importantly, this pcLNA approximation is MVN and therefore it enables the analytical study of various aspects of the system.

2.2 Calculating transversal distributions

We firstly consider how to approximate the transversal distribution

$$P_{x_1}^{(r)} = P(Q_{x_1}^{(r)}|t_0, \xi(t_0) \sim P_0)$$

where $x_1 = g(t_1)$, $t_1 > t_0$, and P_0 is a MVN distribution supported on the normal $N(g(t_0))$. This is the distribution of intersection points of stochastic trajectories with the section $N(g(t_1))$. To approximate $P_{x_1}^{(1)}$ we take $P_{\text{LNA}}(X(t_1)|t_0, \xi(t_0) \sim P_0)$ and calculate the conditional distribution $P_{\text{LNA},x_1}^{(1)}$ given by conditioning on $X(t_1) \in N(g(t_1))$. Since $P_{\text{LNA}}(X(t_1)|t_0, \xi(t_0) \sim P_0)$ is

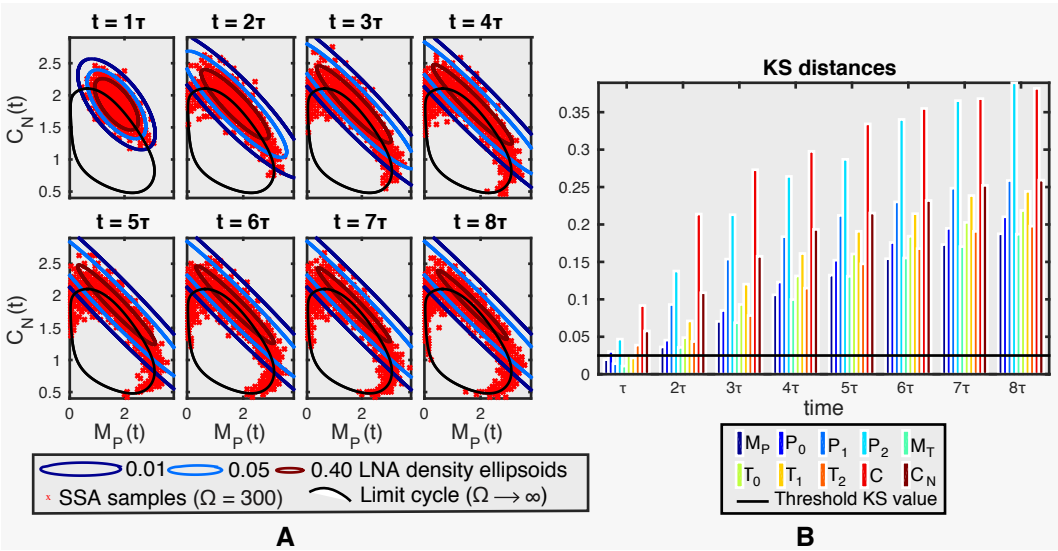

Figure 3: **Comparison between LNA and exact simulations.** (a) Samples obtained from the SSA simulation algorithm (red crosses) and $0.01, 0.05, 0.40$ contours of the LNA probability density (black ellipsoids) at fixed times, $t = \tau, 2\tau, \ldots, 8\tau$ (τ: minimal period), for the circadian clock system. The limit cycle ODE solution is also displayed (black solid line). (b) KS distance between the empirical distribution of SSA samples and 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 $\Omega = 300$.

MVN this conditioning is relatively simple and done using the Schur complement (Sect. 5-6 **S**1). It gives a MVN distribution supported on $N(x_1)$. Now suppose that we have calculated $P^{(r-1)}_{\text{LNA},x_1}$. Then, $P^{(r)}_{\text{LNA},x_1}$ is $P_{\text{LNA}}(X(t_1+\tau)|t_1, \xi(t_1) \sim P^{(r-1)}_{\text{LNA},x_1})$ conditioned on $X(t_1+\tau) \in N(x_1)$ where τ is the period of γ.

Note that $P^{(r)}_{\text{LNA},x_1}$ is also given by conditioning $P_{\text{LNA}}(X((r-1)\tau + t_1)|t_0, \xi(t_0) \sim P_0)$ on $X((r-1)\tau + t_1) \in N(g(t_1))$ because the LNA is transitive i.e. for all $t_1, t_2 > 0$,

$$P_{\text{LNA}}(X(t_1+t_2)|t_0, \xi(t_0) \sim P_0) \\ = P_{\text{LNA}}\Big(X(t_2)\big|t_1, \xi(t_1) \sim P_{\text{LNA}}(\xi(t_1)|t_0, \xi(t_0) \sim P_0)\Big).$$

In **S**1 we show that, although $P_{\text{LNA}}(X((r-1)\tau + t_1)|x_0, \xi_0 \sim P_0)$ diverges as $r \to \infty$, $P^{(r)}_{\text{LNA},x_1}$ converges exponentially fast to a distribution $P^{(\infty)}_{\text{LNA},x_1}$ (Sect. 7 **S**1, cf. Fig. 4). The distribution $P^{(\infty)}_{\text{LNA},x_1}$ is a fixed point in the sense that $P_{\text{LNA}}(X(t_1+\tau)|t_1, \xi(t_1) \sim P^{(\infty)}_{\text{LNA},x_1})$ conditioned on $X(t_1+\tau) \in N(g(t_1))$ is $P^{(\infty)}_{\text{LNA},t_1}$. Using this fact enables us to calculate $P^{(\infty)}_{\text{LNA},x_1}$ directly because, as we show in the SI, its mean and covariance matrix satisfy a simple fixed point equation that is easily solved numerically (Sect. 8 **S**1).

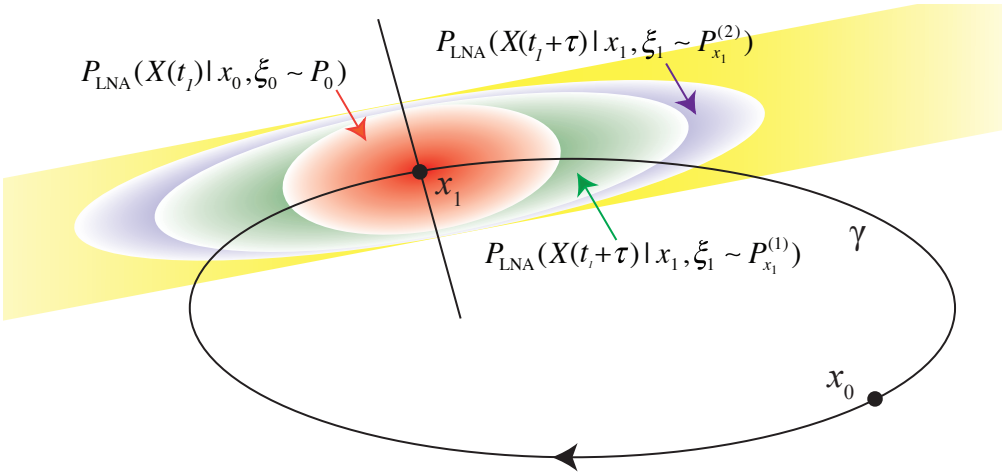

Figure 4: **Schematic diagram of the distributions $P_{\mathbf{LNA}}(X(t_1+\tau)|x_1,\xi(t_1) \sim P^{(r)}_{x_1})$ as r increases.** The yellow distribution represents the limit as $r \to \infty$. Although, for free oscillators, the latter distribution diverges the corresponding conditional distributions converge.

The question remains as to how well these distributions capture the exact simulation transversal distribution $P^{(r)}_{x_1}$. 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. 5 **S**2). In **S**3 we also show similar low Ω results for the Brusselator (Fig. 2 **S**3) and the NF-κB system (Fig. 5 **S**3). 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 $P^{(r)}_{x_1}$ for much larger times $t_1 + r\tau$. Moreover, in Sect. 7 **S**1 we also explain why the convergence of the distribution on normal hyperplanes implies convergence on other transversal sections to γ. However, this result only applies to free-running oscillators and does not apply to forced oscillators as is explained there.

The covariance structure of $P^{(r)}_{\mathrm{LNA},x_1}$ is very interesting and accurately represents the covariance structure of the distribution $P^{(r)}_{x_1}$ obtained from exact simulations. The variances in the direction of the principle axes rapidly decrease. On the other hand, the variances of the individual state variables at these same points on γ are much more uniform in their sizes (Fig 1).

Next, we describe the pcLNA joint distribution of multiple intersections to possibly different transversal sections on the limit cycle. This is MVN and has parameters in a relatively simple form.

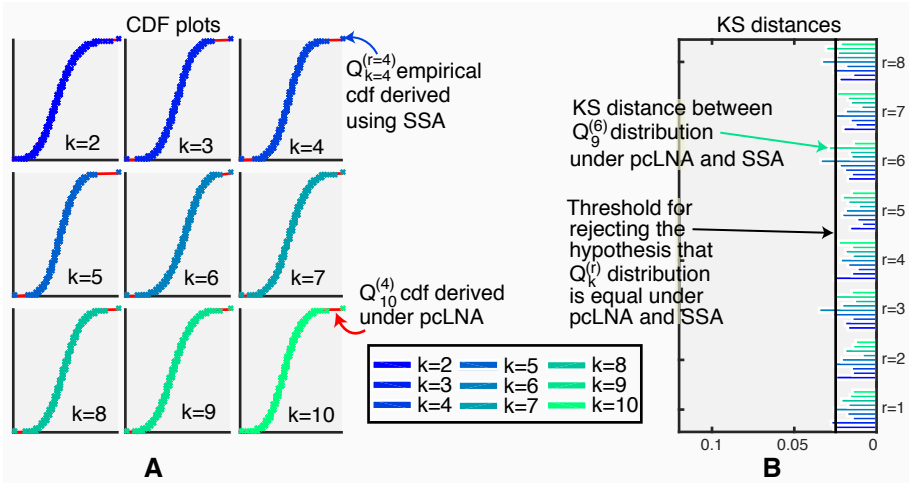

Figure 5: **Comparison of pcLNA and exact transversal distributions.** (A) CDF plots of the transversal distributions $Q_k^{(r)}$ under the pcLNA (red line) and the SSA (empirical cdf, crosses) for transversal coordinates $k = 2, 3, \ldots, 10$ and round $r = 4$ (see Fig. 4 **S2** for $r = 1, 2, \ldots, 8$). (B) KS distances between the $Q_k^{(r)}$ distributions under pcLNA and SSA, $r = 1, 2, \ldots, 8$, $k = 2, 3, \ldots, 10$.

2.2.1 Joint distribution of 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. As mentioned above, if the system size is not too small, although there will be some reversals, $G_N(X(t))$ will move around γ in the direction of the deterministic flow given by (1) and so 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. Suppose it first meets $\mathcal{S}_{x_i}$ in $Q_{x_i}^{(1)}$ for $i = 1, \ldots, q$. If $i < q$ then we let $Q_{x_{i+1}}^{(k)}$ denote the first point in $\mathcal{S}_{x_{i+1}}$ that X meets after it leaves $Q_{x_i}^{(k)}$. If $i = q$ then the next transversal it meets is $\mathcal{S}_{x_0}$ and the intersection point is $Q_{x_1}^{(k+1)}$. In this way we derive a sequence of points $\underline{Q} = Q_{x_1}^{(1)}, \ldots, Q_{x_q}^{(1)}, Q_{x_1}^{(2)}, \ldots, Q_{x_q}^{(2)}, \ldots, Q_{x_1}^{(m)}, \ldots, Q_{x_q}^{(m)}$.

We shall be interested in the distribution

$$P(\underline{Q}|X(t_0)) = P(Q_{x_1}^{(1)}, \ldots, Q_{x_q}^{(m)}|X(t_0)). \tag{4}$$

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 (Sect. 8 **S1**).

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.

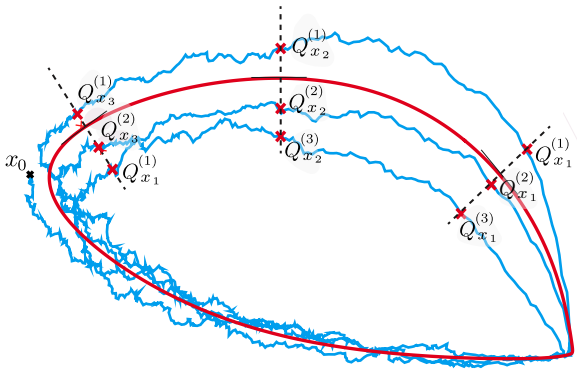

Figure 6: **The sequence Q in two-dimensions.** The stochastic trajectory $X(t)$ (blue line), initiated from x_0, intersects each of the transversal sections $\mathcal{S}_{x_1}, \mathcal{S}_{x_2}$ and $\mathcal{S}_{x_3}$ (dashed lines) of the limit cycle (red solid line) three times, following a path $Q^{(1)}_{x_1} \to Q^{(1)}_{x_2} \to Q^{(1)}_{x_3} \to Q^{(2)}_{x_1} \to Q^{(2)}_{x_2} \to Q^{(2)}_{x_3} \to Q^{(3)}_{x_1} \to Q^{(3)}_{x_2} \to Q^{(3)}_{x_3}$.

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

$$I_{ij} = E\left[\frac{\partial \ell}{\partial \theta_i}\frac{\partial \ell}{\partial \theta_i}\right] = -E\left[\frac{\partial^2 \ell(\theta; X)}{\partial \theta_i \theta_j}\right], \tag{5}$$

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

$$I_{ij} = \frac{\partial \mu}{\partial \theta_i}^T \Sigma^{-1} \frac{\partial \mu}{\partial \theta_j} + \frac{1}{2}\mathrm{tr}\left(\Sigma^{-1}\frac{\partial \Sigma}{\partial \theta_i}\Sigma^{-1}\frac{\partial \Sigma}{\partial \theta_j}\right). \tag{6}$$

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

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

where D_{KL} is the Kullback-Leibler divergence. The significance of the FIM for sensitivity and experimental design follows from its role in (5) 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 [17] but only for small systems and short times where the LNA is accurate. It is notable that the pcLNA enables one to do this for large systems and large times. As an example, we analyse the stochastic behavior of the circadian clock based on the limit distribution $P(\underline{Q}|Q_0)$ when

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

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 the peak of the nuclear complex of PER and TIM proteins C_N. We compute the Fisher Information of the latter distribution using the closed form expression (Sect. 7.3 **S**1) for this distribution. As we can see in Fig. 7(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. Furthermore, a few parameters appear to be most influential. The eigenvectors associated with the two largest eigenvalues of Fisher Information matrix (see Fig. 7(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 [18, 19, 20, 21, 22, 23, 24], 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 also can be used to design experiments that will give data so that the FIM of the combined models (old and new experimental data) will alleviate the decline of the eigenvalues.

2.4 Sensitivity analysis for stochastic systems

The fact that we can calculate the Fisher Information allows a new approach to sensitivity analysis for stochastic systems. As above 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.

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

$$\langle \delta\Theta_1, \delta\Theta_2 \rangle_{\Theta_0} = \delta\mu_1 \Sigma^{-1} \delta\mu_2 + \frac{1}{2}\mathrm{tr}\left(\Sigma^{-1}\delta\Sigma_1 \Sigma^{-1}\delta\Sigma_2\right) \tag{7}$$

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 (7). We can then prove (Sect. 10 **S**1) 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 $\sigma_1 \geq \cdots \geq \sigma_s$ such that

$$MV_i = \sigma_i U_i, \qquad i = 1, \ldots, s. \tag{8}$$

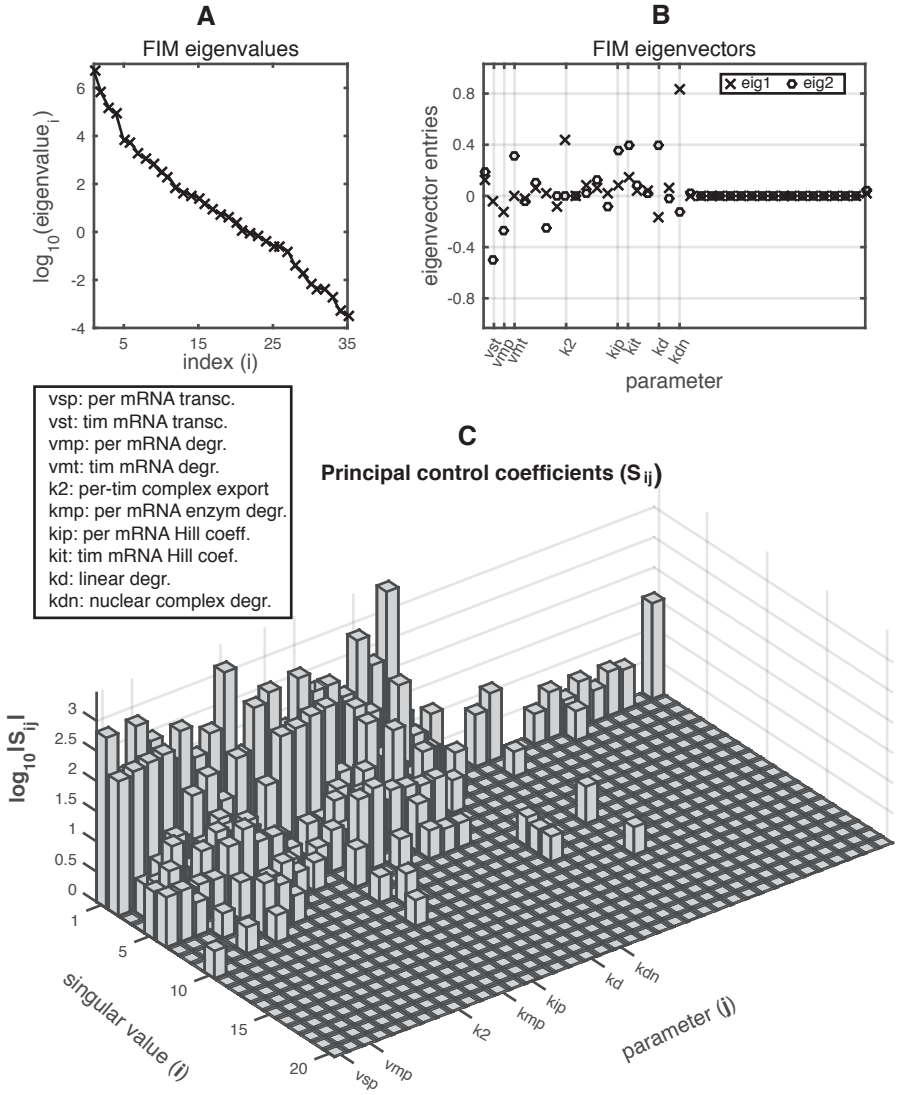

Figure 7: **Stochastic sensitivity analysis of the transversal distribution of the drosophila circadian clock system at the times of the peaks of *per* mRNA and the nuclear complex of PER and TIM proteins.** (A) The logarithm of the eigenvalues of the Fisher Information Matrix (FIM). (B) The entries/weights 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.

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 $M^*M = F$ (Sect. 10 **S1**).

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 (8): if $\delta\theta$ is any change of parameters, the change in μ and Σ is given by

$$\delta\mu = \sum_i U_i^{\mu} \left(\sum_j S_{ij} \delta\theta_j \right) + \mathrm{O}(||\delta\theta||^2)$$
$$\delta\Sigma = \sum_i U_i^{\Sigma} \left(\sum_j S_{ij} \delta\theta_j \right) + \mathrm{O}(||\delta\theta||^2) \tag{9}$$

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 Sect. 10 **S**1 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 (9),

$$\|\delta\Theta\| = \|S \cdot \delta\theta\| + O(\|\delta\theta\|^2). \tag{10}$$

These sensitivities are relatively easy to calculate using the information in Sect. 10 **S**1. In Fig. 7(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.

2.5 pcLNA Simulation Algorithm

Given the ability to accurately approximate the transversal distributions and the results in [8] we realised it should be possible to use this to construct a rapid simulation algorithm. The approach is to use resetting of $g(t)$ to $G_N(X(t))$ to cope with the growth in the variance of $g(t) - G_N(X(t))$ and keep the LNA fluctuation in the transversal direction. The fact that $g(t) - G_N(X(t))$ is expected to be approximately normal (see Sect. 11 **S**1) enables us to do this in a well-defined way.

The pcLNA simulation algorithm is described next. Its main step is illustrated in Fig. 8.

1. Choose a time-step size $\delta\tau > 0$.
2. Input **initial condition** $\kappa(t_0)$ and $X(t_0) = g(t_0) + \Omega^{-1/2}\kappa(t_0)$.
3. **For iteration** $i = 1, 2, \ldots$
   (a) sample $\xi(t_{i-1} + \delta\tau)$ from $\mathrm{MVN}(C_i\kappa(t_{i-1}), V_i)$;
   (b) compute $X_i = g(t_{i-1} + \delta\tau) + \Omega^{-1/2}\xi(t_{i-1} + \delta\tau)$;
   (c) set t_i to be such that $G_N(X_i) = g(t_i)$ and $\kappa(t_i) = \Omega^{1/2}(X_i - g(t_i))$.

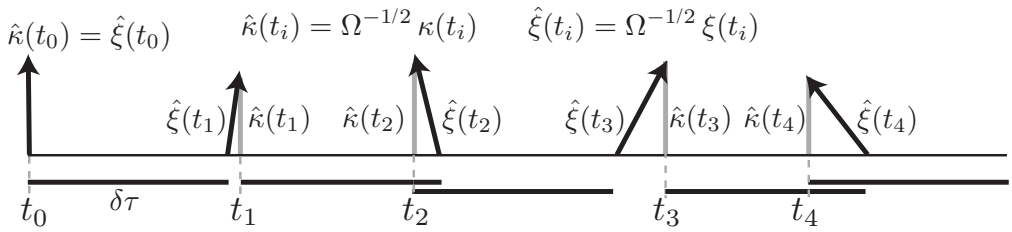

Figure 8: **The main step in the pcLNA algorithm.** The solid horizontal bars below the horizontal axis are all of length $\delta\tau$, the basic timestep of the algorithm. The black arrows show $\hat{\xi}(t_i) = \Omega^{-1/2}\xi(t_i)$ and the grey arrows $\hat{\kappa}(t_i) = \Omega^{-1/2}\kappa(t_i)$. Having calculated $\hat{\kappa}(t_i)$ one uses $\kappa(t_i)$ as the initial state and updates it using the LNA and a time-step δt to obtain ξ at $t_i + \delta t$. Then ξ is replaced by κ so that $g(t_i + \delta t) + \Omega^{-1/2}\xi = g(t') + \Omega^{-1/2}\kappa$ where κ is normal to the limit cycle. Lastly, $t_{i+1} = t'$ and $\hat{\kappa}(t_{i+1}) = \Omega^{-1/2}\kappa$.

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

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. $P_{\text{LNA}}(X(s + t)|X(0)) = P_{\text{LNA}}(X(t)|X(s) \sim P_{\text{LNA}}(X(s)|X(0)))$. 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 Sect. 4.1 **S**2.

The pcLNA algorithm is fast because (i) it can make relatively large steps from one transversal distribution to another maintaining accurate temporal correlation, and (ii) because, although it has to integrate the variational equation of the system (Sect. 4 **S**1), there is an efficient method to compute and store a subset of the drift $C(s,t)$ and diffusion $V(s,t)$ matrices for the system and the given parameters once and for all so that all required variational solutions can be computed extremely quickly from this subset. This subset is computed up-front. A Matlab routine to do this is contained with the package PeTSSy which is freely available (Sect. 12 **S**1). Thus, the only extra computational cost compared to the standard LNA is due to the calculation of $G_N(X)$ which can be practically performed through an optimization procedure (e.g. the Newton-Raphson algorithm).

2.5.1 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 the 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 [25]) with a fixed time step δt.

For both the tau-leap and the CLE approximation we explored different values of their parameters ϵ and δt, respectively, 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, 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 Sect. 13 **S**1.

Figure 9 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 Fig. 7, 8 **S**2 for $\Omega = 1000, 3000$, respectively). 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.

The pcLNA algorithm is about 24 times faster than the CLE approximation and much faster than the tau-leap and the SSA. In our simulations, it took 0.4sec to derive this long-time trajectory, 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,

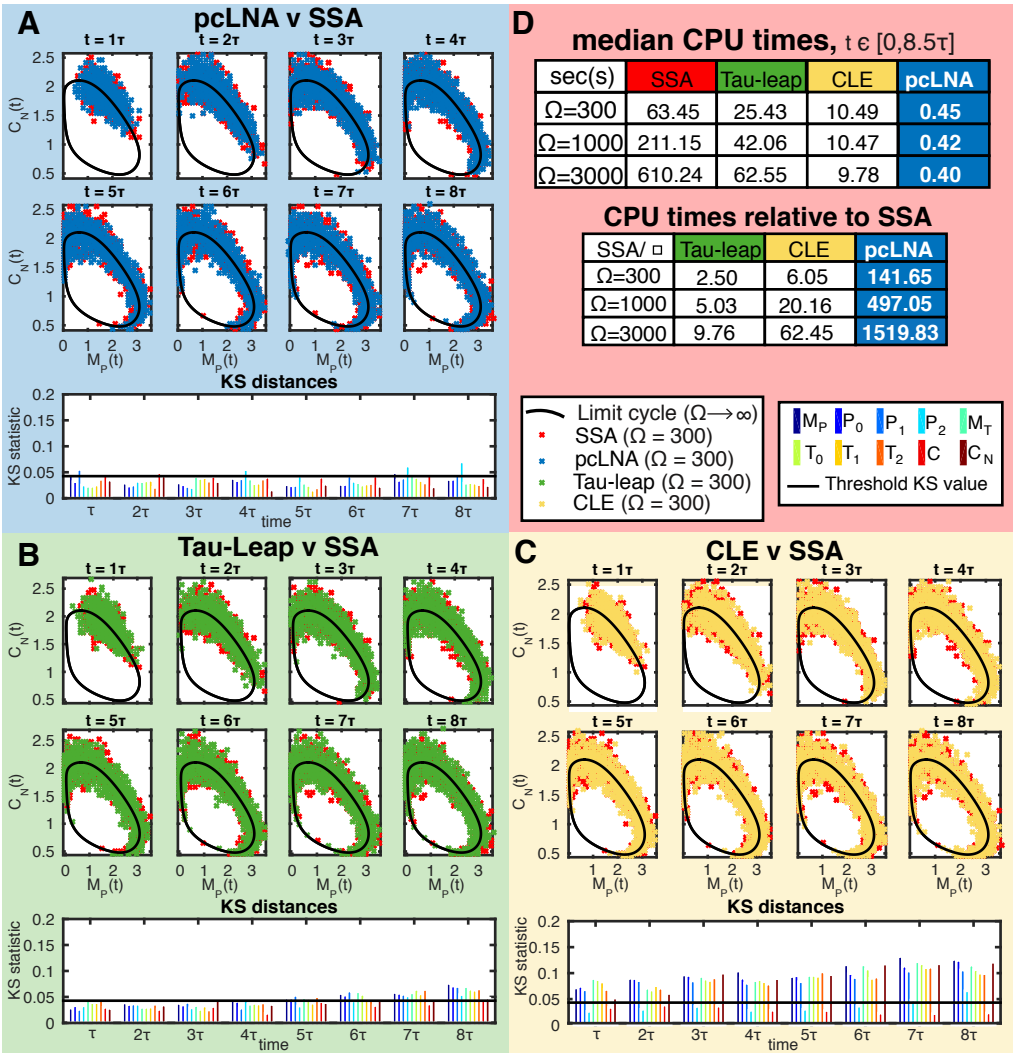

Figure 9: **Comparison between pcLNA, tau-leap and CLE simulation algorithms for the drosophila circadian clock.** Panels (A), (B) and (C) contain the samples produced by respectively the pcLNA, tau-leap ($\epsilon = 0.002$) and CLE ($\delta t = 0.002$) algorithms and the exact simulation (SSA) at time-points, $t = 1\tau, 2\tau, \ldots, 8\tau$, for $\Omega = 300$, along with the KS distances between the empirical distributions of each approximation and the SSA for each system variable (colored bars). 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.

it is based on a framework which one can also use to analyse oscillatory systems and perform statistical inference. For the latter the following Kalman filter is particularly relevant.

2.6 Calculating likelihoods via a pcLNA Kalman Filter

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 $P(\underline{Q}|X(t_0))$ 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 [26]. 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,

$$\hat{X}(t) = BX(t) + \epsilon, \tag{11}$$

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

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

Let $m(t)$ and $S(t)$ be the mean and covariance matrix of the stochastic process $\xi(t) = \sqrt{\Omega}(X(t) - g(t))$ under the LNA. Then $\mu(t) = g(t) + \Omega^{-1/2}m(t)$ and $\Sigma(t) = \Omega^{-1}S(t)$ are the mean and covariance matrix of $X(t)$ and $\hat{\mu}(t) = B\mu(t)$, $\hat{\Sigma}(t) = B\Sigma(t)B^T + \Sigma_\epsilon$ the mean and covariance of $\hat{X}(t)$. Denote $\hat{\epsilon}(t) = \hat{X}(t) - \hat{\mu}(t)$ the prediction error at time t.

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 $g(t'_{i-1}) = G(\mu^*(t_{i-1}))$ and set $(\kappa(t'_{i-1})|\hat{X}(t_{i-1})) \sim MVN(m^*(t'_{i-1}), S^*(t'_{i-1}))$;
   (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 $\mu^*(t_{i-1}) = \mu(t_{i-1}) + \Sigma(t_{i-1})B^T\hat{\Sigma}(t_{i-1})^{-1}\hat{\epsilon}(t_{i-1})$, $\Sigma^*(t_{i-1}) = \Sigma(t_{i-1}) - \Sigma(t_{i-1})B^T\hat{\Sigma}(t_{i-1})^{-1}B$ $m^*(t'_{i-1}) = \Omega^{1/2}(\mu^*(t_{i-1}) - g(t'_{i-1}))$, and

$$S^*(t'_{i-1}) = E_2(t'_{i-1})\left(S^*_{22}(t'_{i-1}) - \frac{S^*_{21}(t'_{i-1})S^*_{12}(t'_{i-1})}{S^*_{11}(t'_{i-1})}\right)E_2(t'_{i-1})^T$$

where the matrix $E_2(t) = [e_2(t) \dots e_n(t)]$ 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 (11) 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 $(\kappa(t')|\hat{X}(t))$ in step 3(b) is derived by restricting the posterior distribution of the noise process on the transversal section $\mathcal{S}_{g(t')}$ normal to $e_1(t')$ 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. 3) and transition (Eq. 4.3 **S**1) are used to derive the prior distribution $(X(t_i)|\hat{X}(t_{i-1}))$ in step 3(c).

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

[8] 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.

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

[10] 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.

[11] 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.

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

[13] 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.

[14] 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.

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

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

[17] 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.

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

[19] 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.

[20] 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.

[21] 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.

[22] 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.

[23] 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.

[24] 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.

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

[26] 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.