

Filtering and Inference for stochastic oscillators with distributed delays



Bärbel Finkenstädt, Department of Statistics, University of Warwick, UK

Joint with:

Silvia Calderazzo, Mans Unosson, Adam Johansen (Department of Statistics, Warwick, UK)

Marco Brancaccio, Mike Hastings (Medical Research Council Laboratory of Molecular Biology, Cambridge, UK)

Funding: MRC, EPSRC, ERC

WARWICK
THE UNIVERSITY OF WARWICK

EPSRC

Engineering and Physical Sciences
Research Council

MRC

Medical
Research
Council

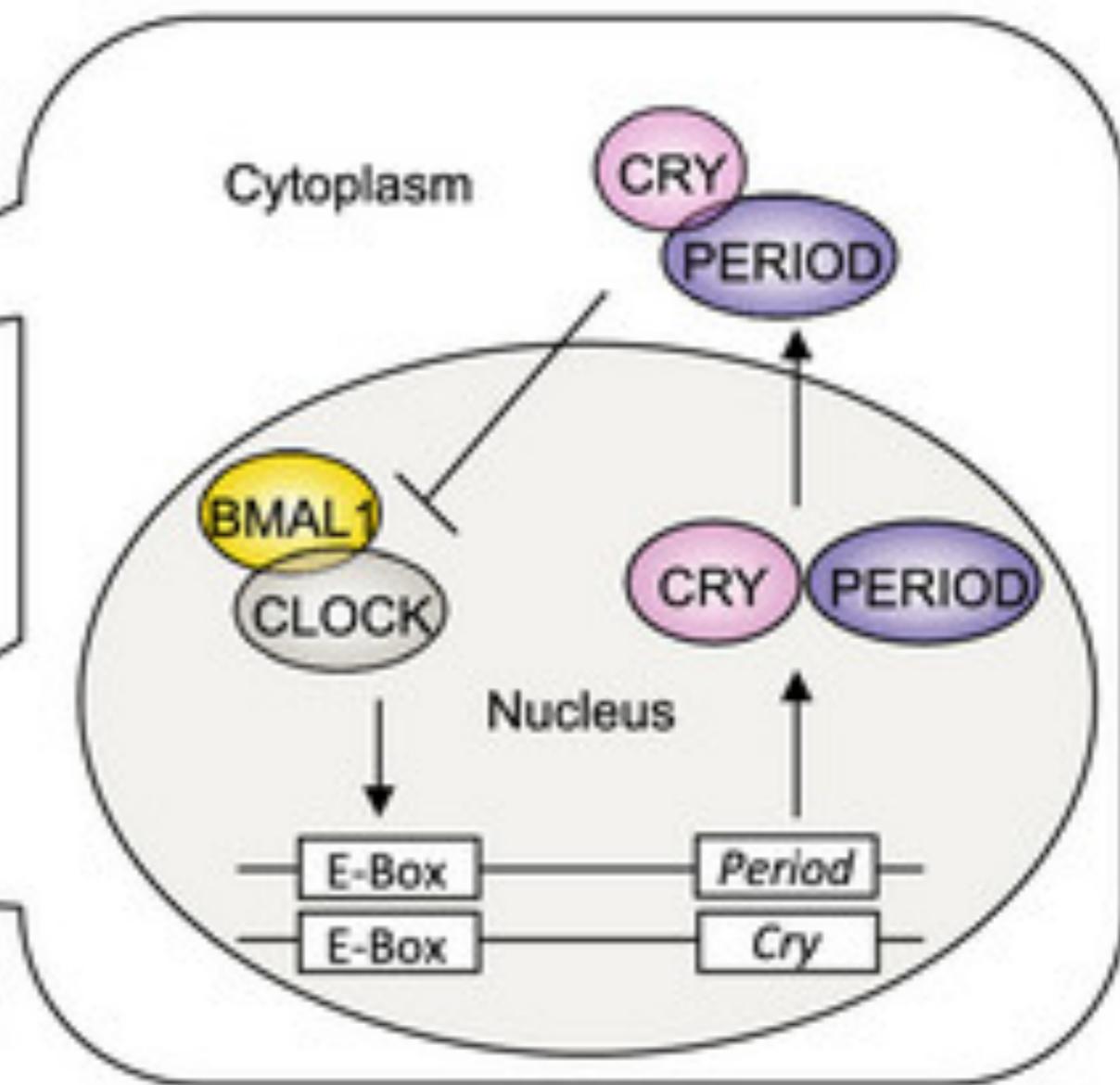
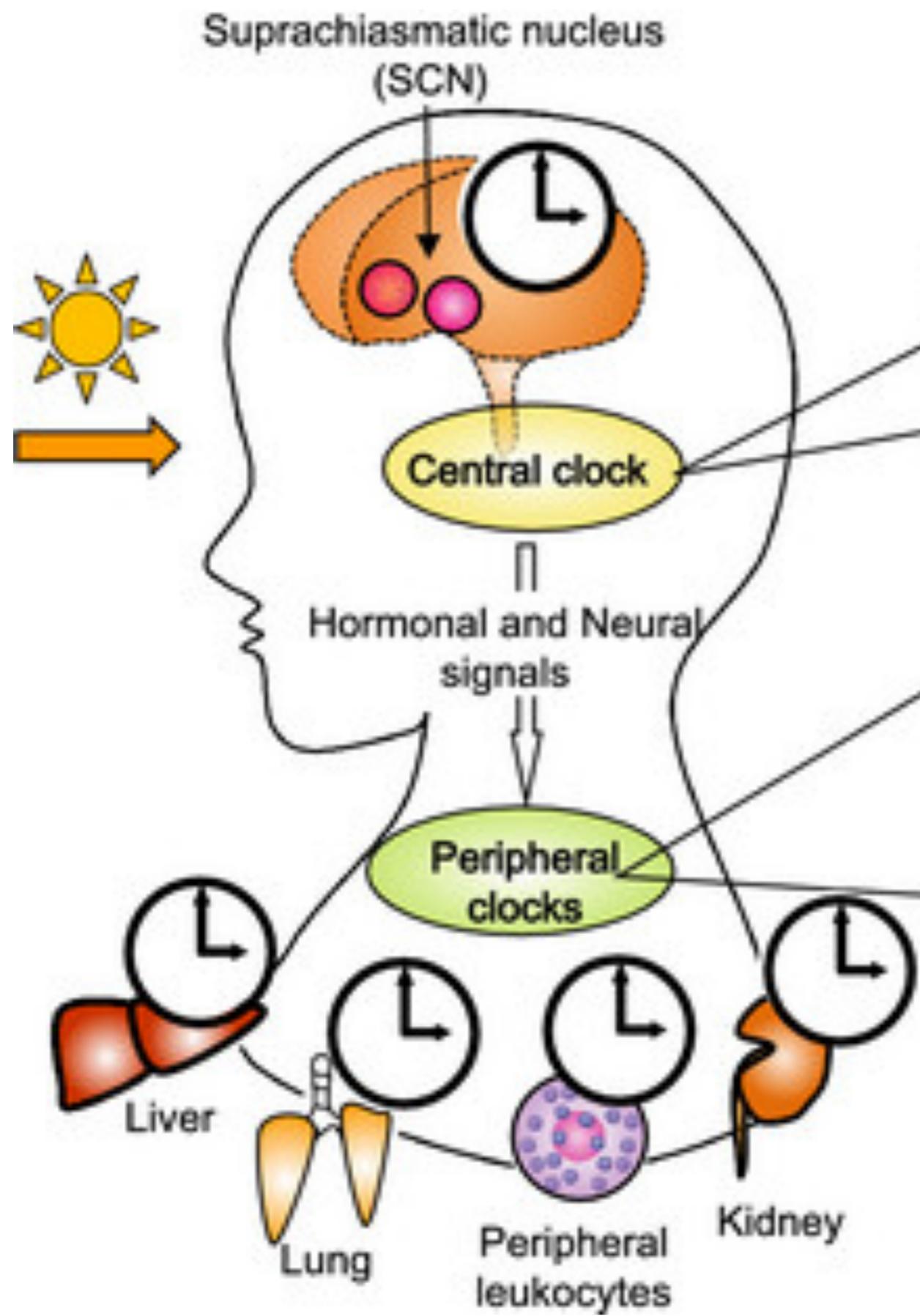
Circadian Timing System

Organisms have evolved an internal biological clock which allows them to temporally regulate and organize their physiological and behavioral responses to cope in an optimal way with the fundamentally periodic nature of the environment.

Almost all cells of our body display circadian rhythms which are endogenous in that they can persist with high precision for many weeks in isolation of any external cues but can be adjusted (entrained) to the local environment by external cues such as the solar state.

This network is coordinated by a hypothalamic pacemaker, the suprachiasmatic nucleus (SCN), the principal circadian clock in the brain of mammals which is entrained by visual afferents, input from other brain and peripheral oscillators.

About 20K neurons within the SCN where each neuron sustains oscillations in the expression of clock genes and electric firing.



The synchrony of an organism with both its external and internal environments is critical to the organism's well-being and survival

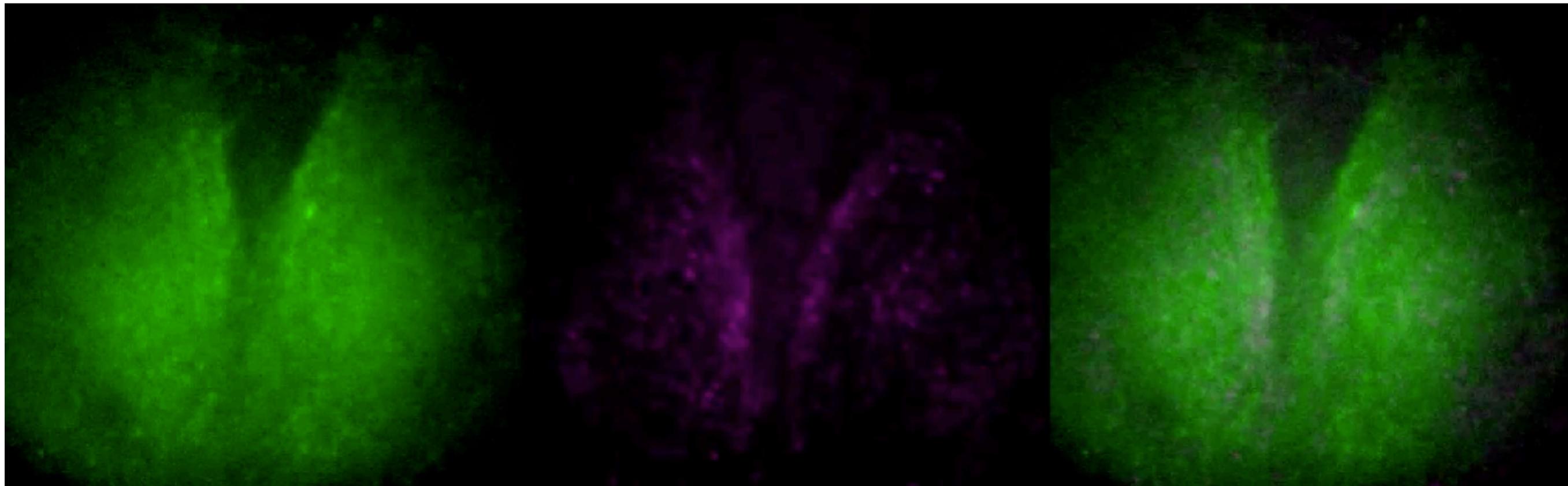
Mounting evidence supports a link between circadian misalignment and increased risk:

- chronic diseases including cancer, metabolic syndrome (obesity, hypertension, arteriosclerosis, diabetes,)
- psychiatric disorders (depression, bipolar, schizophrenia, attention deficit).

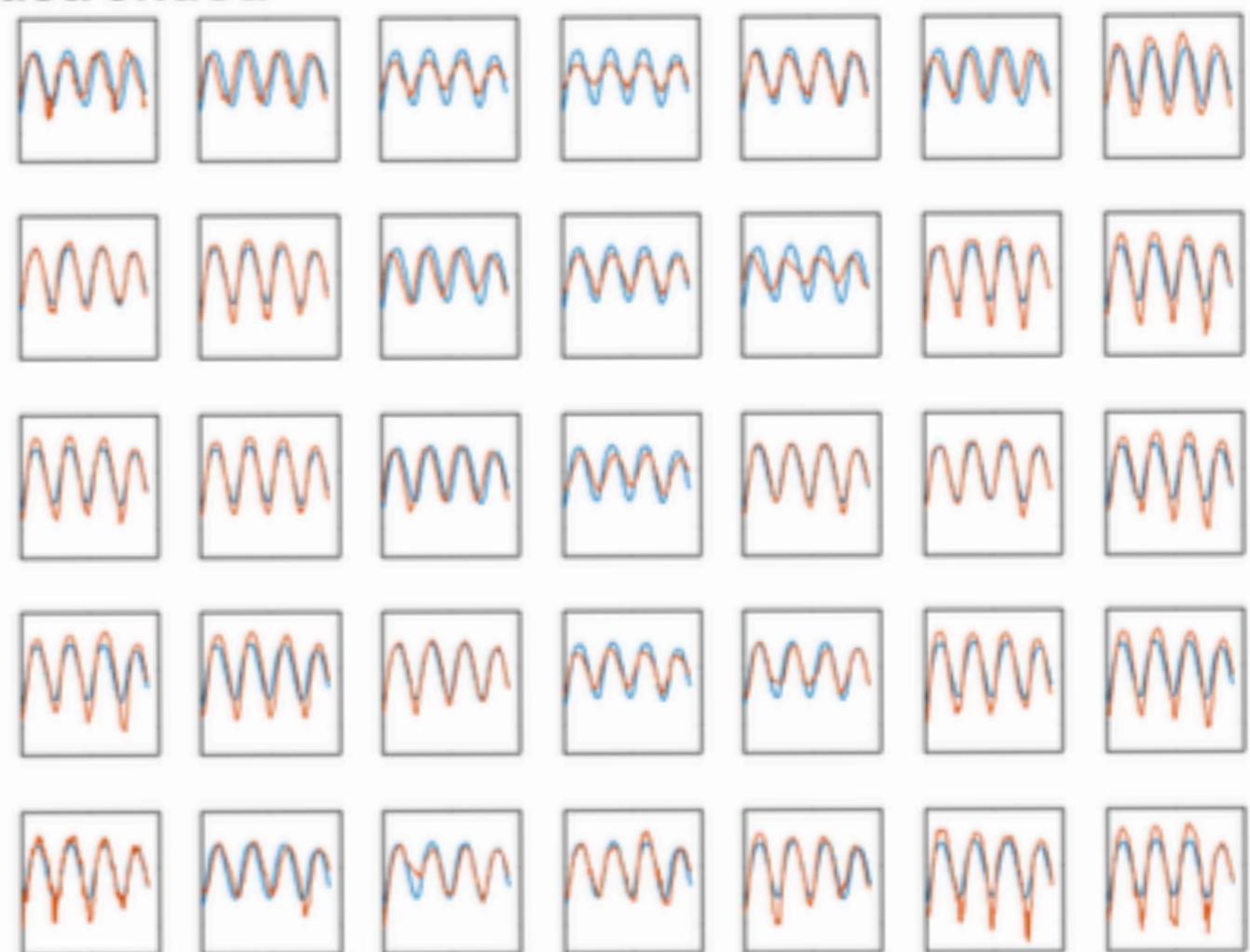
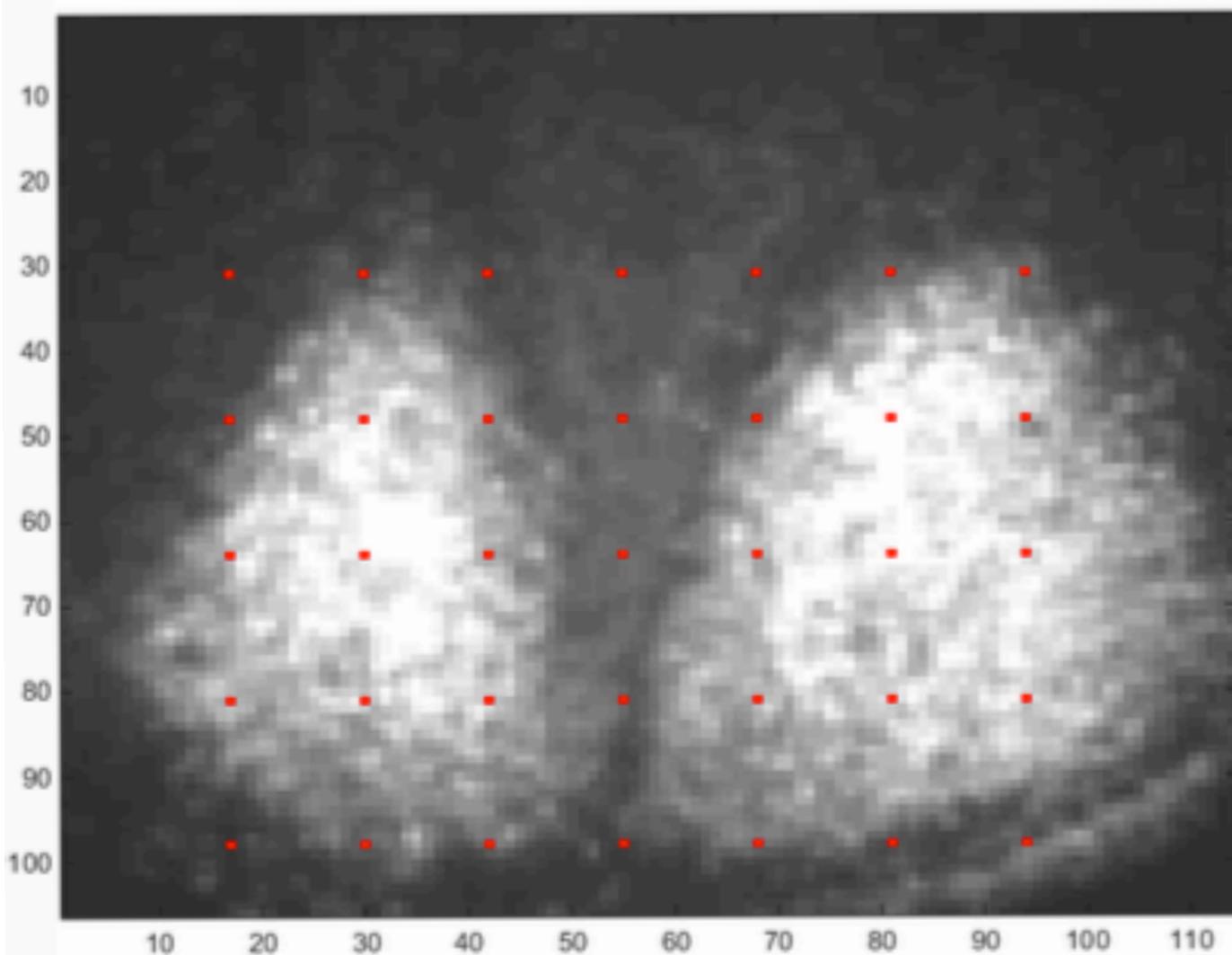
Circadian rhythm alterations have been consistently associated with poor quality of life and poor survival, especially in large cohorts of cancer patients.

Circadian Rhythm in SCN

Calcium (green) and Per2 (purple) (*Data/video: Hasting's lab, Cambridge*)



Per2:luc detrended



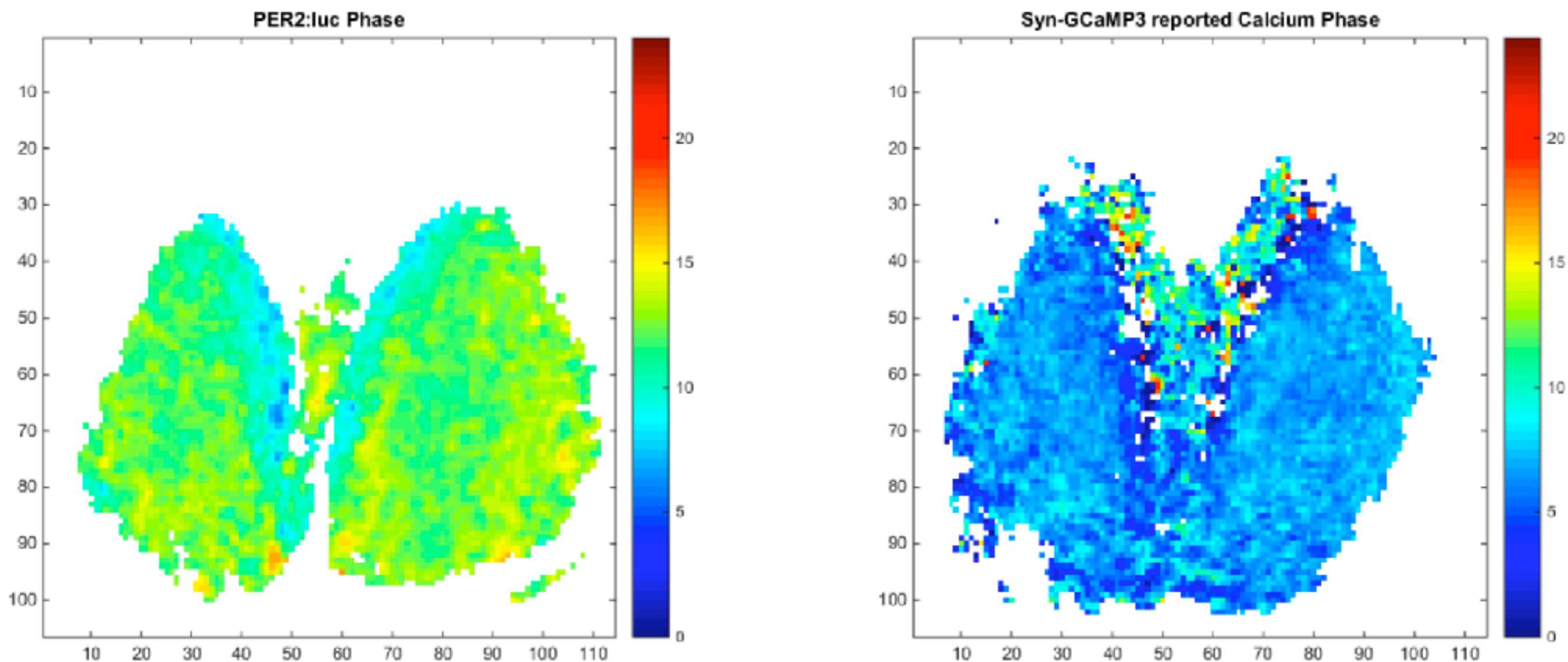


Figure 9: Heatmap of phase of dominant frequency of Per2:luc (left) and calcium (right) mapped to 0-24 hours. Pixels with low luminosity (below average across time and space) or non-circadian expression ($<18\text{h}$ or $>30\text{h}$) are omitted. Calcium is phase advanced compared to Per2, except for along bands along the dorsomedial region of the SCN. Circadian calcium has a homogeneous phase whereas Per2 shows clear spatial features.

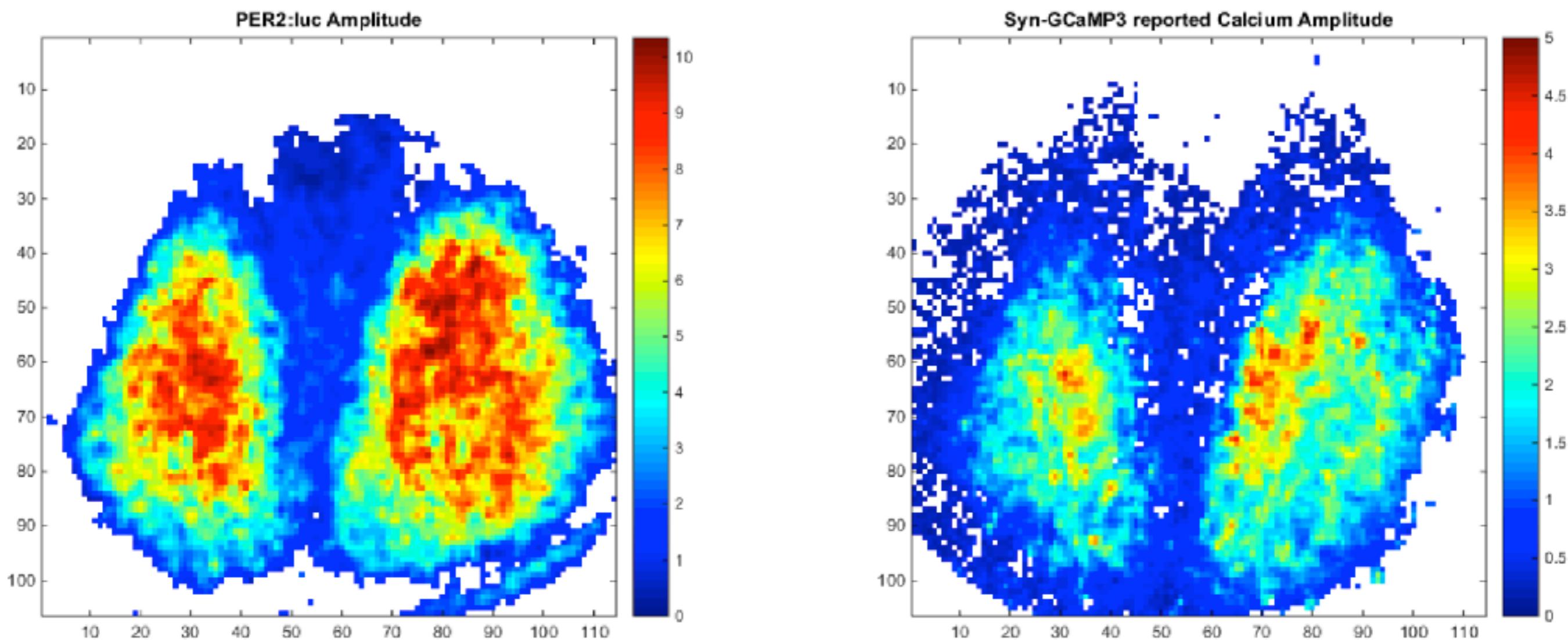


Figure 12: Heatmap of amplitude of dominant frequency of Per2:luc (left) and calcium (right) mapped to 0-24 hours. Pixels with low luminosity (below 0.3x average across time and space) or non-circadian expression ($<18\text{h}$ or $>30\text{h}$) are omitted. Note that specific values of the amplitude cannot be compared between Per2:luc and calcium as the measurement processes are different.

Individual SCN neurons are individual circadian clocks each possessing an intrinsic TTFL oscillator

Clock and Bmal1 activate expression of Period (Per) and Cryptochrome (Cry) via E-box regulatory sequences. Cry and Per proteins suppress E-box activation, creating a negative feedback loop as activation of the E-box can only recommence when these negative regulators are cleared.

Mathematical model of the mammalian clock (Relógio et al. 2011) which comprises 20 ordinary differential equations modeling 20 molecular species including 5 genes to model the main negative Per/Cry feedback loop along with the secondary ROR/Bmal/REV-ERB loop which is achieving some fine-tuning of Bmal thus adding robustness to the system.

A significant simplification can be obtained by reducing the TTFL to negative feedback loop with a delay in a single clock gene Per that is transcribed and then inhibits its own transcription after a delay.

TTFL with Delays

- Introducing **delays** provides a parsimonious way to model biochemical processes that depend on a large number of reactions and species.
- Can exhibit the oscillatory expression of genes (Monk, 2003)
- Korencic et al. (2012) show that a single negative feedback loop with discrete delay can give rise to sustained oscillations in a model of the circadian gene Per2.
- **Distributed delays**, although deterministic in formulation (“linear chain trick” Smith, 2011), arise naturally in systems with intrinsic stochasticity.

Chemical Langevin Equation

$$d\mathbf{X}(t) = S h(\mathbf{X}(t)) dt + \sqrt{S \operatorname{diag}\{h(\mathbf{X}(t))\} S^{\top}} dW(t)$$

where

$W(t)$ is a N -dimensional Wiener process,

$h(\cdot)$ is the P -dimensional vector of hazards,

S is the $N \times P$ matrix of stoichiometries.

Computing the likelihood of the CLE is challenging as the transition densities are typically not available in closed form

Consider a CLE with distributed delay truncated at some maximum delay time

$$dX(t) = f \left(\int_{t-\tau_m}^t X(s) g_{p,a}(t-s) ds, X(t) \right) dt \\ + \sqrt{A \left(\int_{t-\tau_m}^t X(s) g_{p,a}(t-s) ds, X(t) \right)} dW(t),$$

where

$$f = S h(\cdot)$$

$$A = S \text{diag}\{h(\cdot)\} S^\top$$

$$g_{p,a}(u) = \frac{a^p u^{p-1} e^{-au}}{(p-1)!}, \quad u > 0$$

Replacing f and A by their first order Taylor expansions about $\rho(\cdot)$
 assuming that $X(0) \sim \mathcal{N}(\rho(0), P(0))$

$$X(t) \sim \mathcal{N}(\rho(t), P(t)),$$

where the mean and covariance are propagated by

$$\begin{aligned} d\rho(t) &= f \left(\int_{t-\tau_m}^t \rho(s) g_{p,a}(t-s) ds, \rho(t) \right) dt \\ dP(t) &= J_f \left(\int_{t-\tau_m}^t \rho(s) g_{p,a}(t-s) ds, \rho(t) \right) P(t) dt + P(t)^\top J_f \left(\int_{t-\tau_m}^t \rho(s) g_{p,a}(t-s) ds, \rho(t) \right)^\top dt \\ &\quad + A \left(\int_{t-\tau_m}^t \rho(s) g_{p,a}(t-s) ds, \rho(t) \right) dt, \end{aligned}$$

J_f is the Jacobian of f

measurement process

$$Y_t = \kappa \int_{t-\Delta t}^t X(t)dt + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, \Sigma_\epsilon),$$

where

κ is a scaling parameter between light intensity
and molecular concentration

Δt is the time between observations.

Assume some discretization for the process $X(t)$

$$X_{0:T} = [X_0, X_{\delta t}, X_{2\delta t}, \dots, X_T]$$

Note that we typically have that $\delta t < \Delta t$

We can then write the measurement

$$Y_t = \kappa F X_{t-\Delta t+\delta t:t} + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, \Sigma_\epsilon),$$

where F is a $\Delta t/\delta t$ by N matrix averaging the unobserved states over Δt .

Kalman update (extended Kalman- Bucy filter (EKBF))

$$\rho_{t+\Delta t-\tau_m:t+\Delta t}^* = \rho_{t+\Delta t-\tau_m:t+\Delta t} + C(y_{t+\Delta t} - F\rho_{t+\Delta t})$$

$$P_{t+\Delta t-\tau_m:t+\Delta t}^* = P_{t+\Delta t-\tau_m:t+\Delta t} - CFP_{t+\Delta t,t+\Delta t-\tau_m:t+\Delta t}$$

where

$$C = P_{t+\Delta t-\tau_m:t+\Delta t,t+\Delta t}F^\top (FP_{t+\Delta t}F^\top + \Sigma_\epsilon)^{-1}$$

Note that an initial estimate is required of $\rho_{\Delta t:\tau_m}$ and $P_{\Delta t:\tau_m}$

Auto-repressive circadian gene regulation

$$dX(t) = \left[\nu \left(\int_{t-\tau_m}^t X(s) g_{p,a}(t-s) ds \right) - \mu X(t) \right] dt + \sqrt{\nu \left(\int_{t-\tau_m}^t X(s) g_{p,a}(t-s) ds \right) + \mu X(t)} dW(t)$$

transcription function $\nu(\cdot)$

$$\nu \left(\int_{t-\tau_m}^t X(s) g_{p,a}(t-s) ds \right) = \frac{R_0}{1 + \left(\frac{\int_{t-\tau_m}^t X(s) g_{p,a}(t-s) ds}{K_d} \right)^n}$$

R_0 is the maximum transcription rate

n the Hill coefficient

K_d the dissociation constant

μ represents the degradation rate

Application to Cry1 imaging data

- Cry1 is a key gene in the TTFL
- Data from video recording of an organotypic slice of mouse SCN (Brancaccio et al. 2013)
- Luciferase construct Cry1::Luc to image Cry1 expression
- 6 days, exposure time of 0.5 hours
- 288 frames of 414 by 217 pixels
- Use 4 by 4 aggregation

Parameter Inference

- Random Walk Metropolis with combination of univariate and block proposals
- Adaptive scheme during the first 4K iterations
- Delayed acceptance algorithm (Christen & Fox 2005):

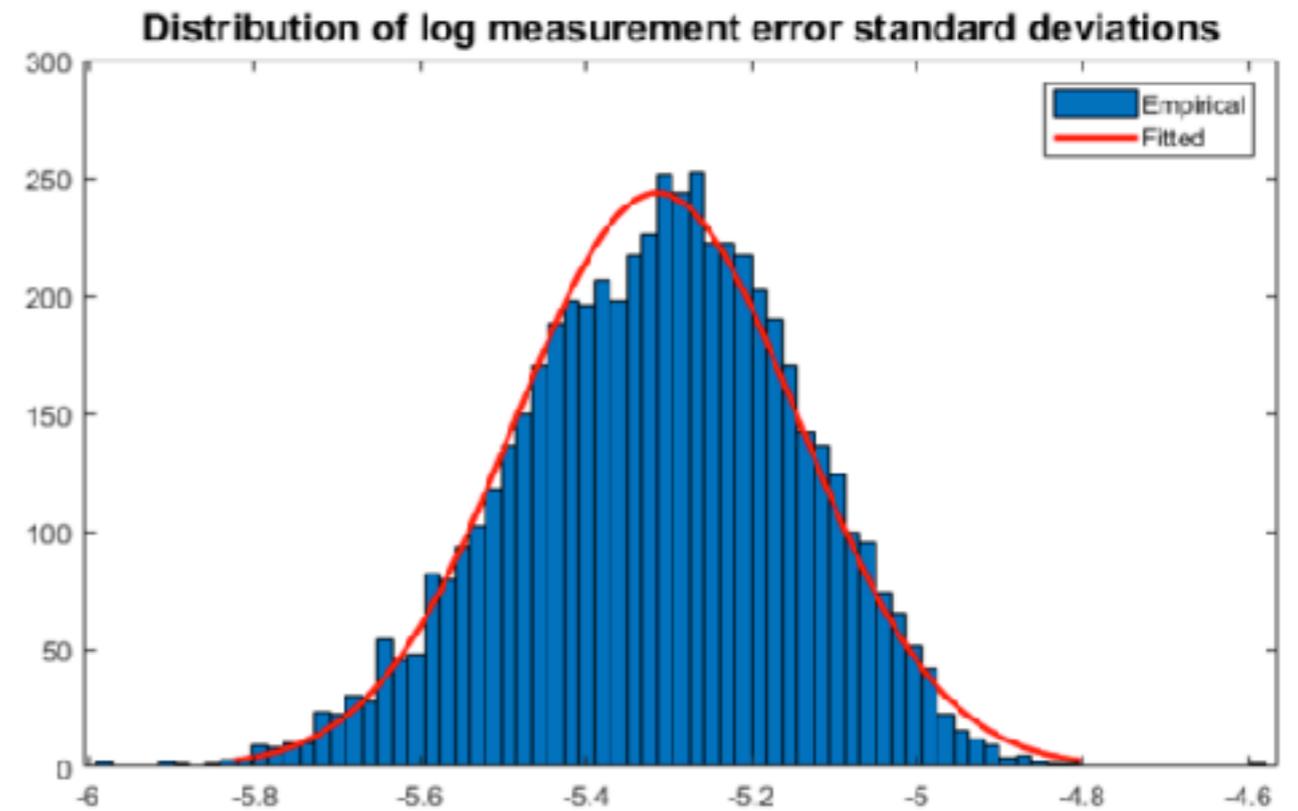
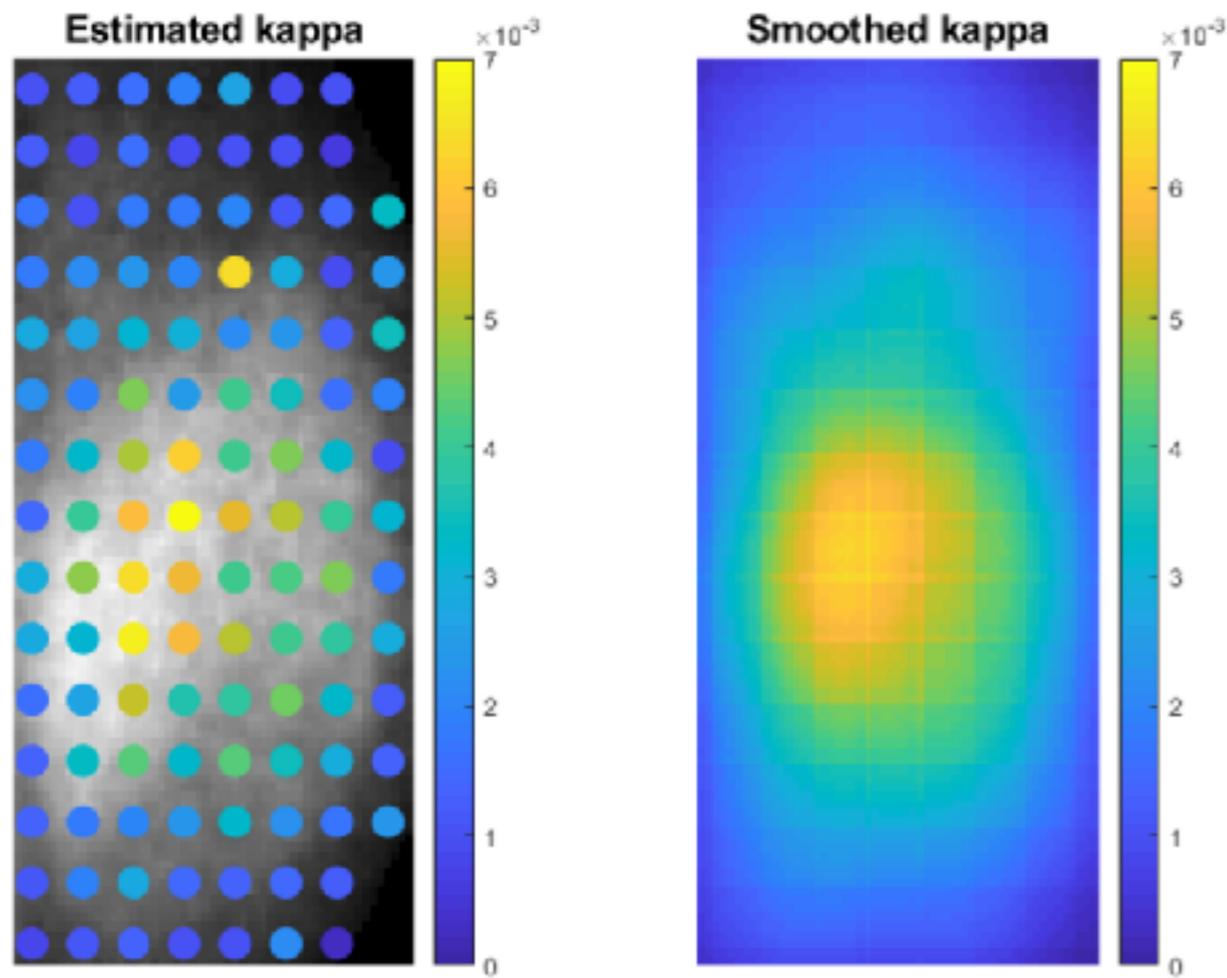
$$a_{fast}(\theta^{i+1}, \theta^i) = \min \left\{ 1, \frac{\mathcal{L}_{0.5}(\theta^{i+1} | \mathbf{Y}) \pi(\theta^{i+1})}{\mathcal{L}_{0.5}(\theta^i | \mathbf{Y}) \pi(\theta^i)} \right\}$$

$$a_{DA}(\theta^{i+1}, \theta^i) = \min \left\{ 1, \frac{\mathcal{L}_{0.1}(\theta^{i+1} | \mathbf{Y}) \mathcal{L}_{0.5}(\theta^i | \mathbf{Y})}{\mathcal{L}_{0.1}(\theta^i | \mathbf{Y}) \mathcal{L}_{0.5}(\theta^{i+1} | \mathbf{Y})} \right\}$$

Priors

Table 1: Priors used for the main run of the inferential algorithm.

μ_Γ	σ_Γ	$\log R_0$	$\log K$	$\log n$	$\log \mu$	$\log \kappa$	$\log \sigma_c$
$U(0, 23)$	$U(0, 20)$	$\mathcal{N}(0, 10^2)$	$\mathcal{N}(0, 10^2)$	$\mathcal{N}(0, 10^2)$	$\mathcal{N}(-0.55, 0.25^2)$	$\mathcal{N}(\hat{\mu}_\kappa^{(i,j)}, \hat{\sigma}_\kappa^{(i,j)})$	$\mathcal{N}(-5.3, 0.17^2)$



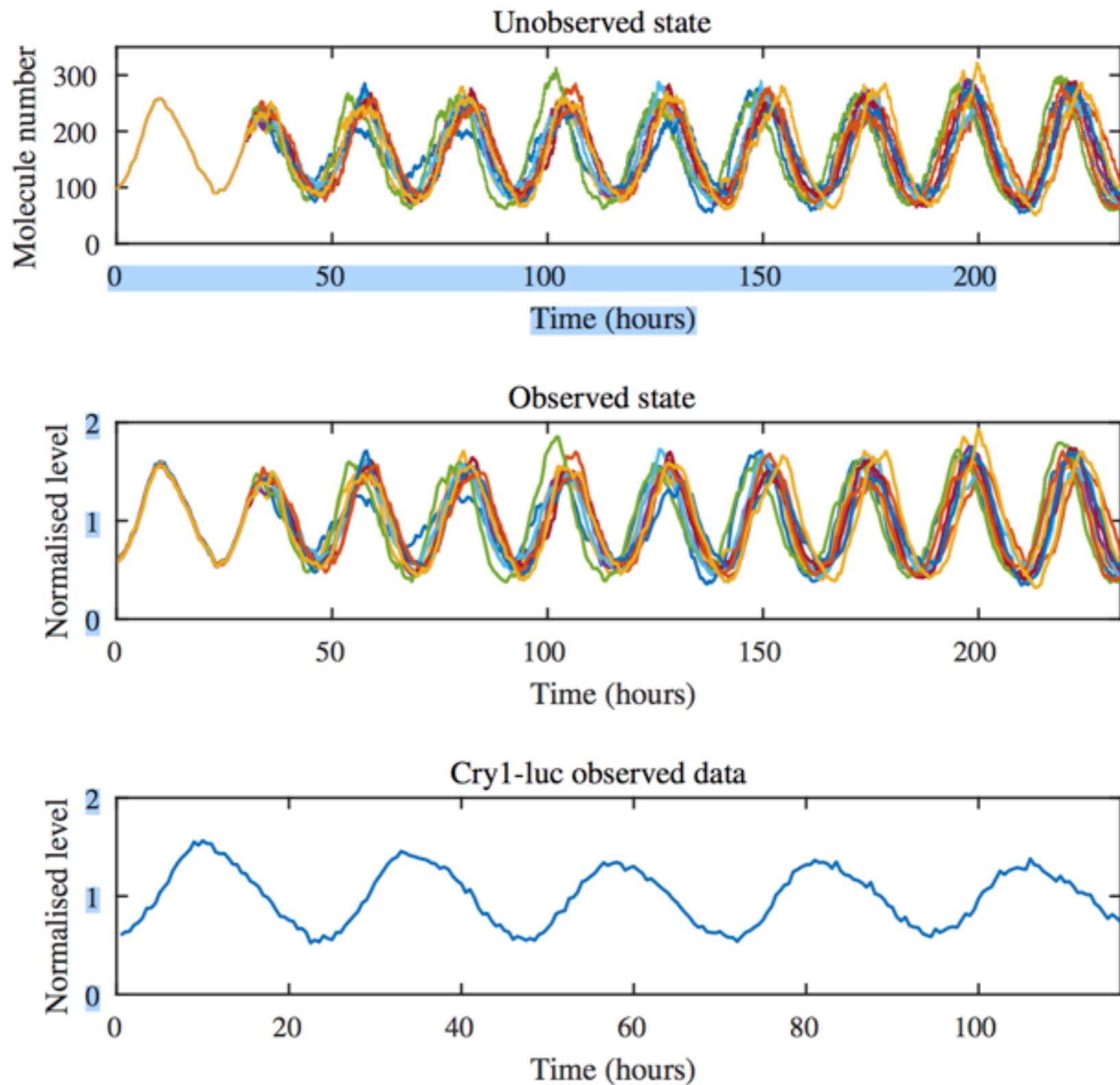


Fig. 1. Top: SSA-type simulations for the reactions in (8) and (9). **Center:** molecule counts are rescaled by their mean level, integrated over 0.5 hours and corrupted with measurement error. The assumed levels of signal to noise ratio is 100. **Bottom:** experimental *Cry1-luc* imaging time series, aggregated, de-trended and normalised.

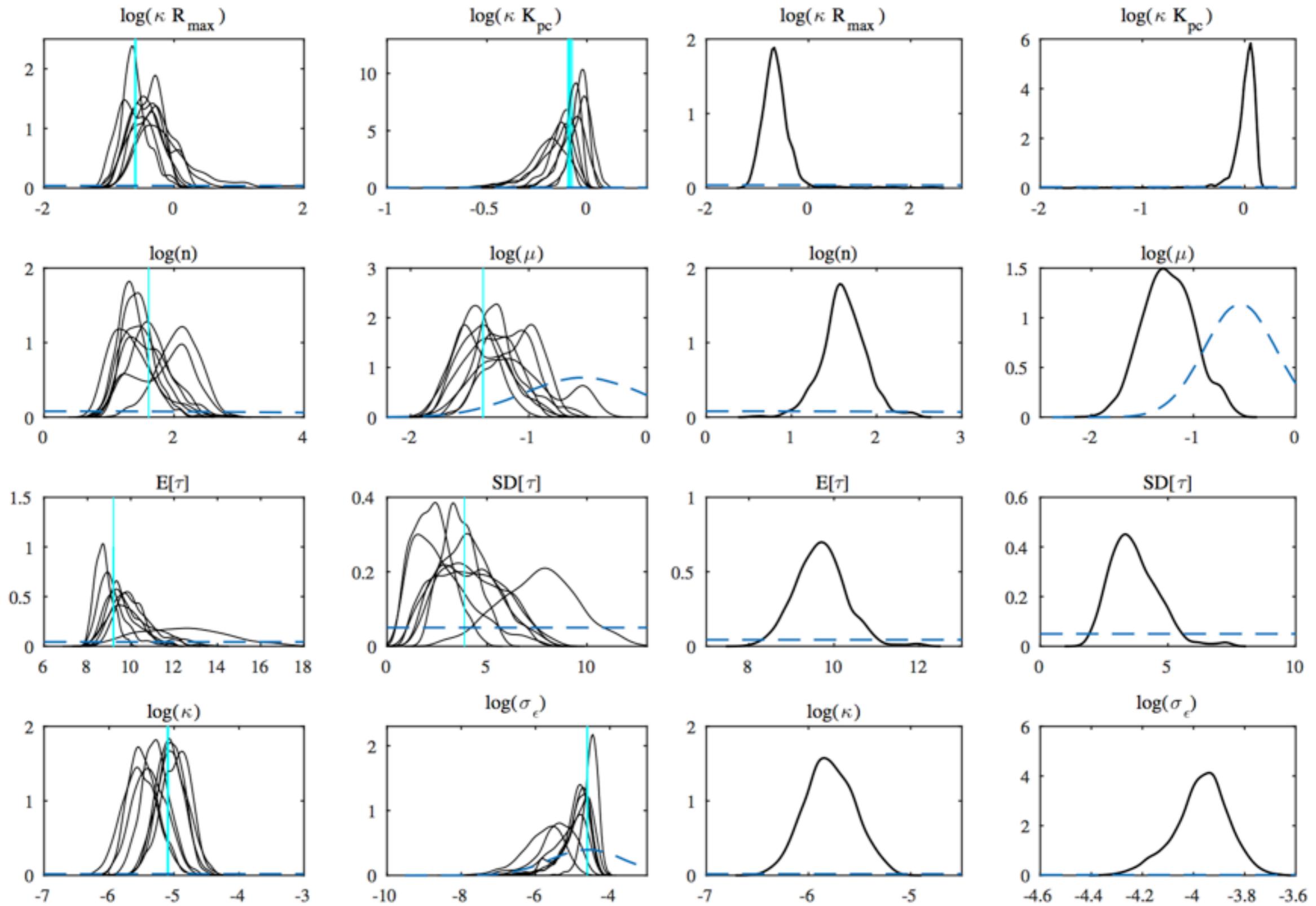


Fig. 2. Left two columns: Results for simulation study. Kernel densities estimates of the model parameters posterior densities, excluding the parameters of the initial condition. $E[\tau]$ and $SD[\tau]$ denote the mean and standard deviation of the delay density K . The prior density is shown as a dashed line while the vertical line marks the true value. Results for the last five cycles of the simulated data shown in the central panel of Figure 1 (two chains are excluded due to non-convergence). **Right two columns:** Results for observed *CryI-luc* shown in bottom panel of Figure 1. Same notations and definitions as in panels on left.

Residual Analysis

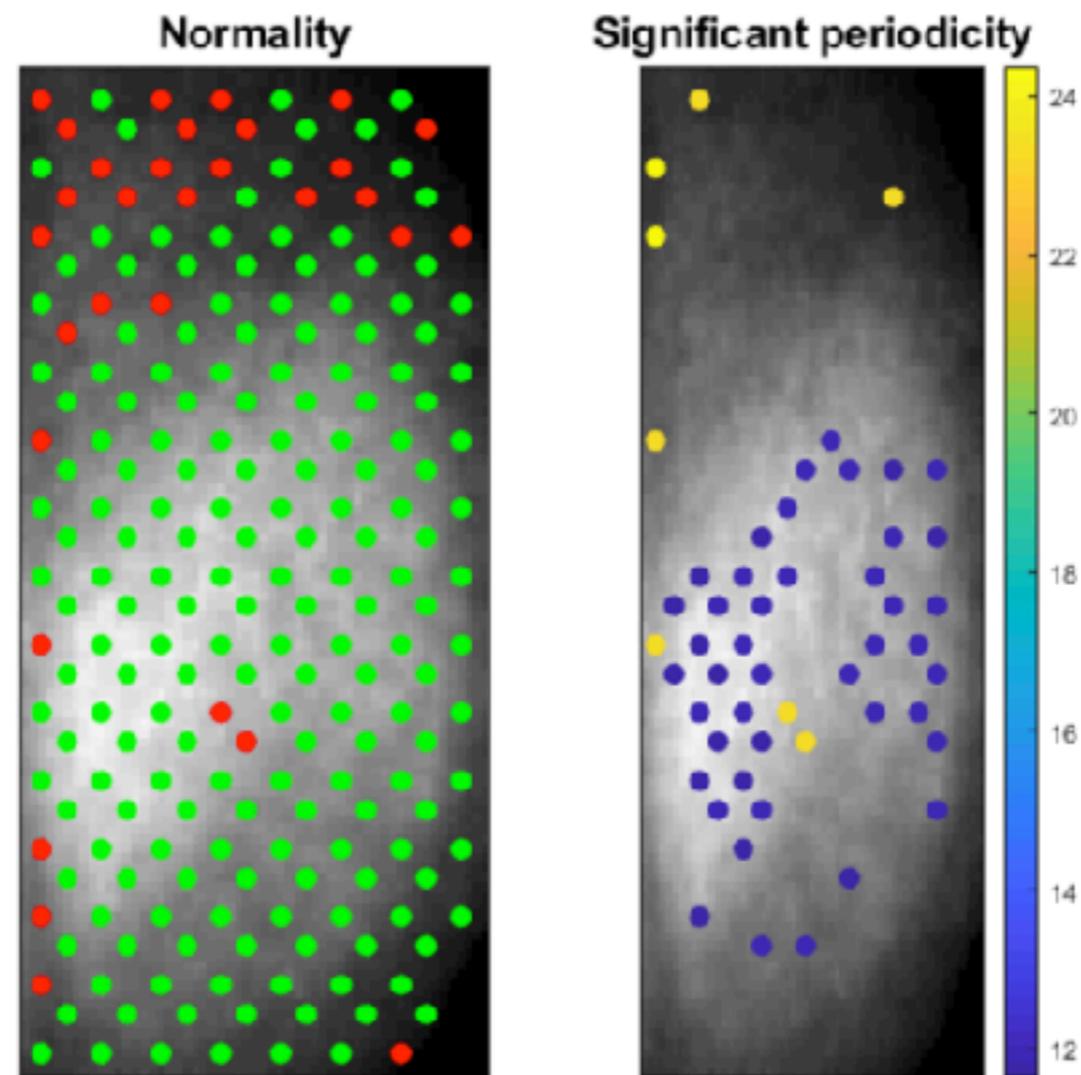


Figure 4: **Left:** Red indicates significant deviation from normality using one-sample Kolmogorov-Smirnov test, green indicates no significant deviation from normality. **Right:** Locations with significant residual periodicity, using Fishers g-statistic. The periodicity is indicated by colour. A few locations show significant 24 hour periodicity (yellow). Locations with significant 12 hour periodicity (blue) are clustered in the core of the SCN.

Clustering of estimated parameters

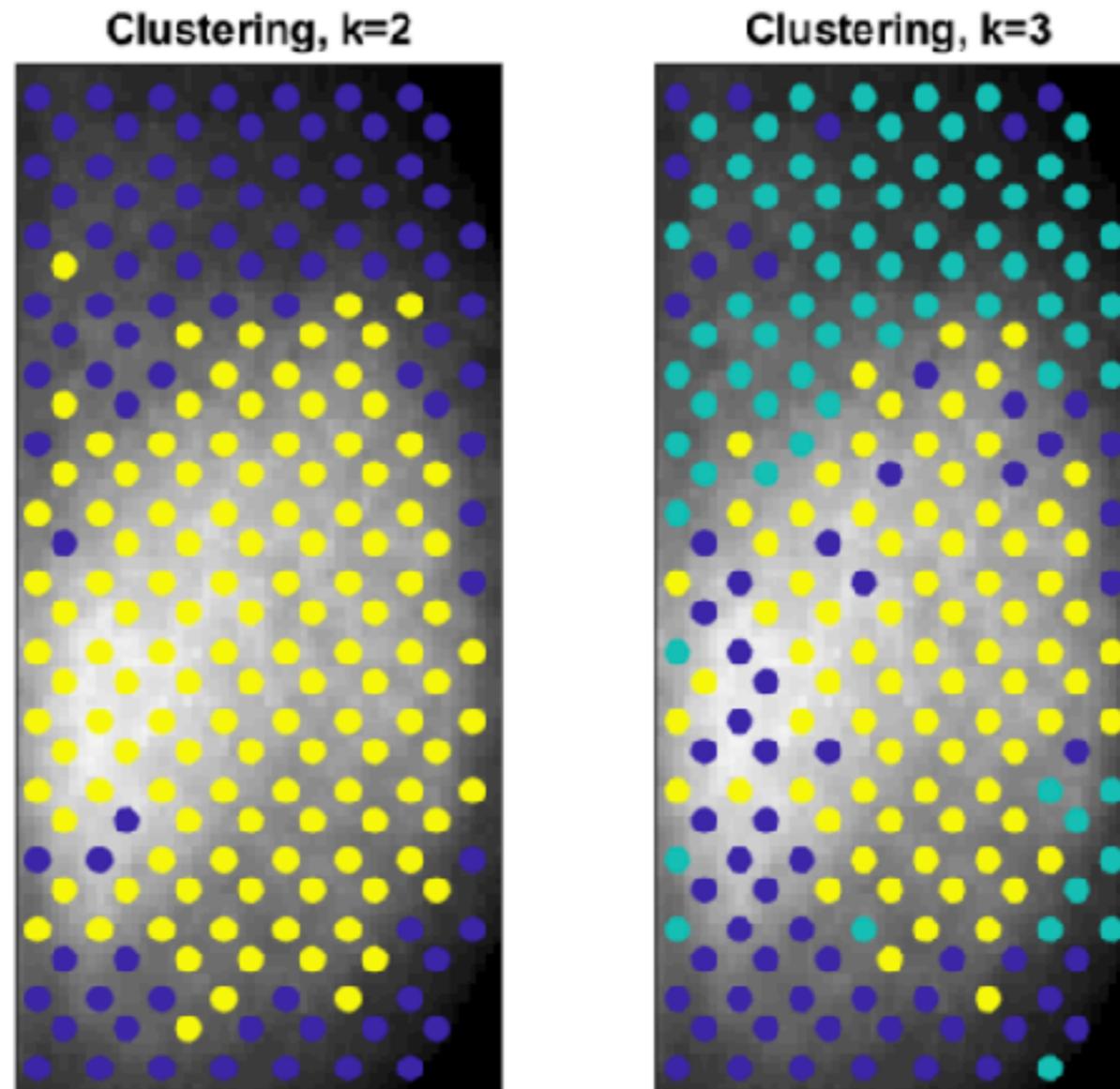
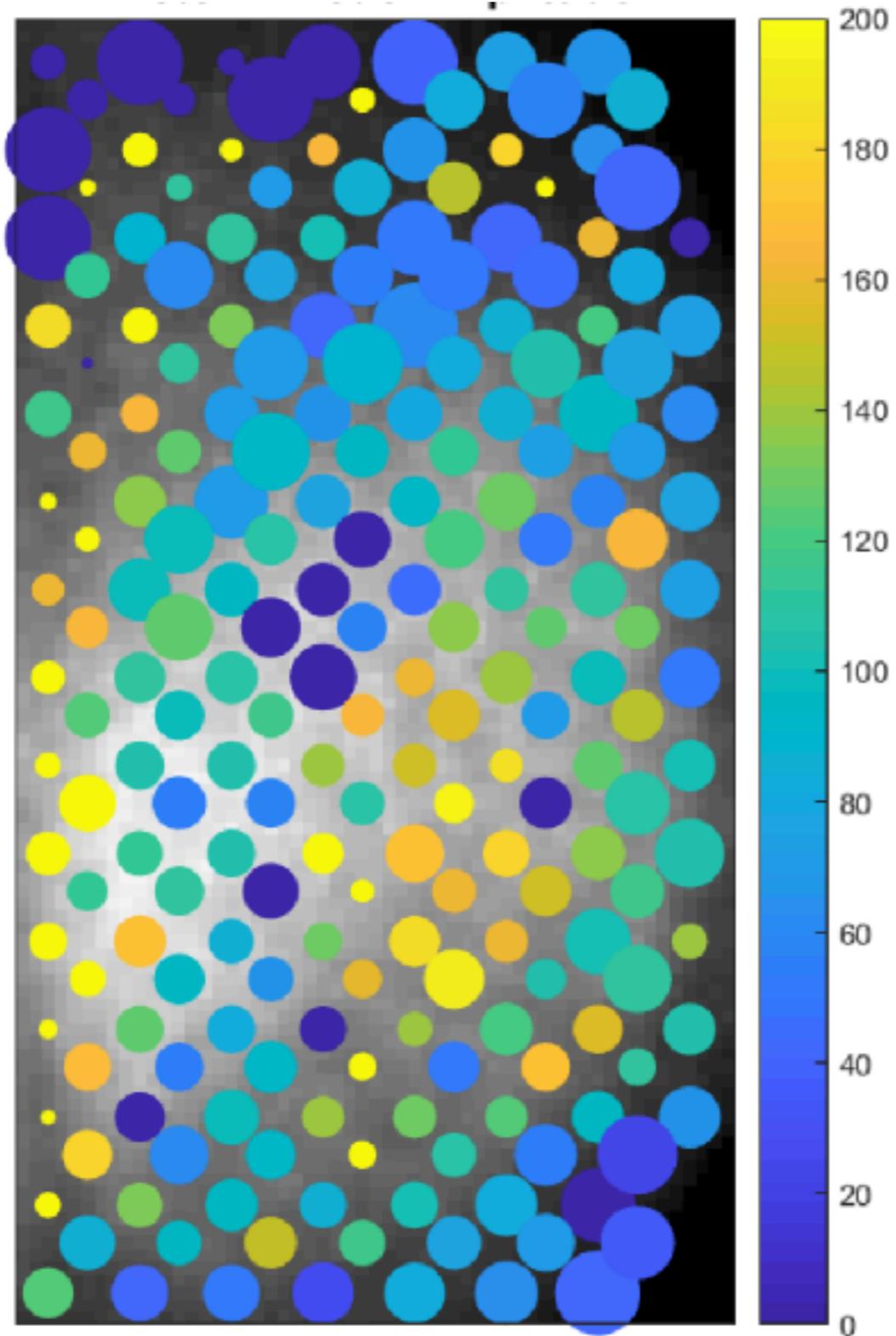


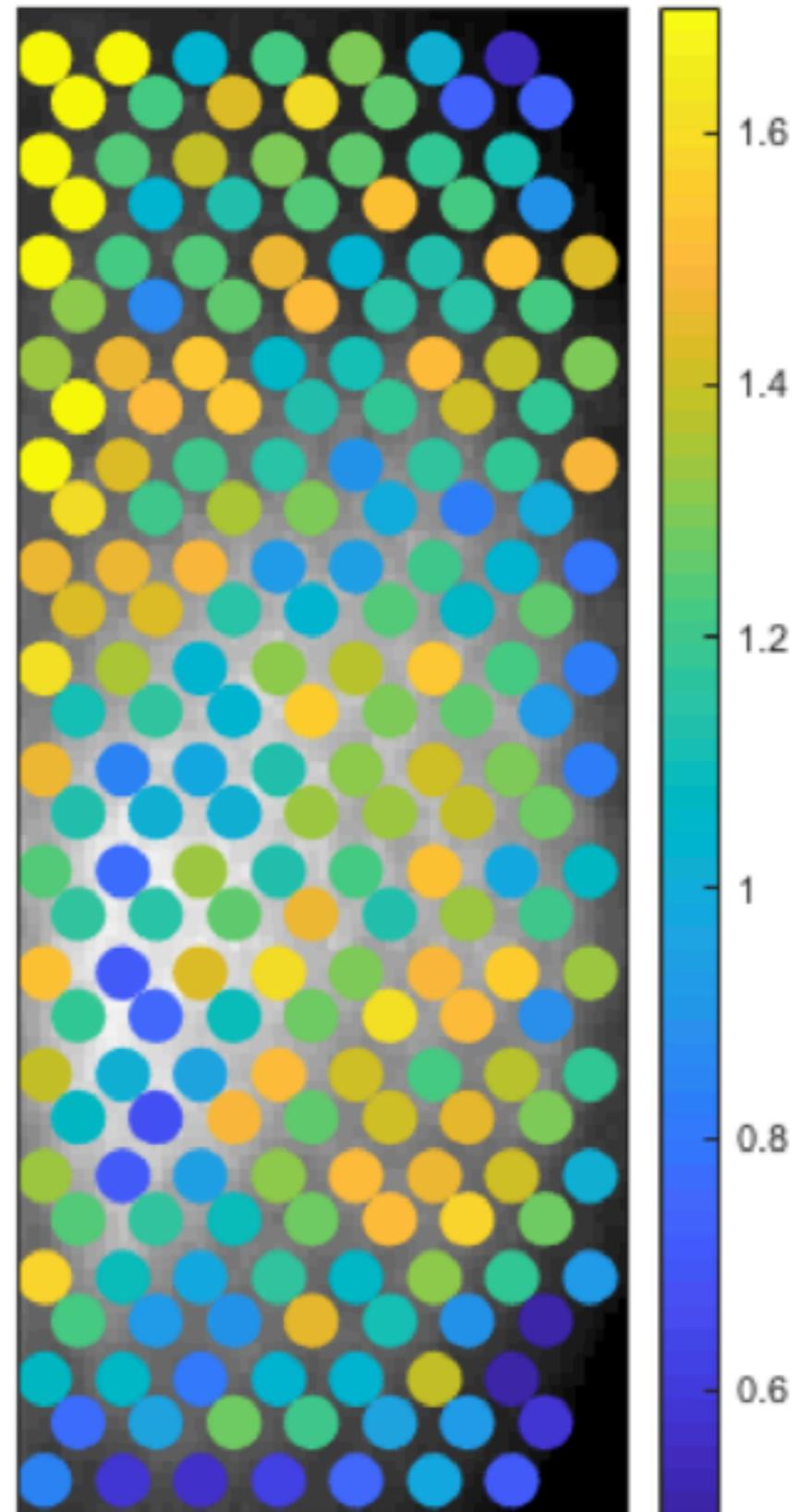
Figure 5: **Left:** The output of k -medoids clustering of the parameter estimates with $k = 2$. Locations are classified as either core or shell with little overlap. **Right:** The output of k -medoids clustering with $k = 3$. The locations are largely classified as dorsomedial (green), core (yellow), or ventral (blue). There is some spatial overlap between the three clusters.

Analysis of Model

Noise characteristic time scale



Amplitude



Summary

- Stochastic Model for circadian TTFL involving a distributed Delay term
- Measurement process
- Adaptive MCMC algorithm
- Cry1: known to cause spatial variation across SCN
- Spatial distribution of clusters are very similar to recent experimental results produced by Taylor et al. 2017 who studied phase restoration after washout of tetrodotoxin treatment during which intercellular communication is abolished. Ventral entrains dorsal SCN.
- Future work: Incorporate TTFL and Signaling in a 2-TF- model will allow us to study synchronisation in the SCN

Systems biology

Filtering and inference for stochastic oscillators with distributed delays

Silvia Calderazzo^{1,2,*}, Marco Brancaccio³ and Bärbel Finkenstädt^{1,*}

¹Department of Statistics, University of Warwick, Coventry CV4 7AL, UK, ²Division of Biostatistics, German Cancer Research Center, 69120 Heidelberg, Germany and ³Division of Neurobiology, Medical Research Council Laboratory of Molecular Biology, Cambridge CB2 0QH, UK