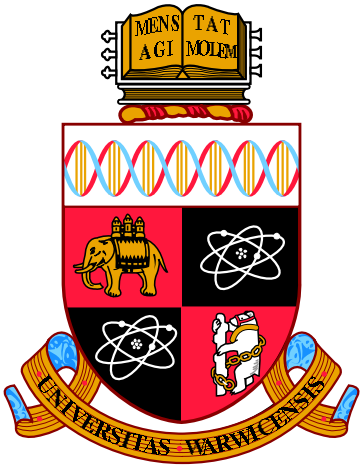

Spatio-temporal Inference for Circadian Gene Transcription in the Mammalian SCN

by

Måns Unosson

Thesis

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy

Department of Statistics, University of Warwick

June 2020

Contents

Acknowledgments

I thank my supervisors, Bärbel and Adam for their exceptional guidance during the last three and a half years, and for helpful suggestions in the preparation of this thesis. You always knew what questions to ask. It has been a pleasure working with you both.

I also wish to thank my parents for their unfailing support, and my brother for reminding me about the importance of having fun.

Thank you to the staff at the Department of Statistics in Uppsala for introducing me to statistics in the first place, and for providing a friendly environment during the strange times this spring.

I am grateful for having had the chance to meet some wonderful people along the way with whom to share the ups and downs of an otherwise solitary experience. For Joseph and Lea, and other members of "we are friends" for a wonderful community. For Anton and our shared misery, and Joe, Rachel and Giovanni for being reliable office neighbours. For Mariko and the music. For Jon, thank you for all the laughs.

Finally, I wish to thank you Toini, for your patience, love and support. You and I made this journey together.

Declarations

I hereby declare that this thesis contains my own original work, unless otherwise acknowledged and referenced. This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree. This work has been funded by the Department of Statistics at the University of Warwick and the Economic and Social Research Council, grant number 1791198. Analyses of previously unpublished data included in this work are performed on experimental data kindly provided by the Hastings lab at the MRC, Cambridge.

Abstract

Almost all life on earth exhibit circadian rhythms of behaviours that are tied to the natural day and night cycle. In mammals, the suprachiasmatic nucleus (SCN) is responsible for generating and communicating these rhythms to peripheral tissues. The neurons of the SCN function as noisy molecular clocks, expressing circadian genes in an oscillatory fashion over the course of 24 hours through a transcriptional/translational feedback loop (TTFL). The cells synchronise to form a robust clock, capable of exact timekeeping and entrainment to external stimuli, e.g. light, via intercellular signalling. This thesis investigates spatio-temporal inference for stochastic models of the TTFL, motivated by the availability of high-resolution bioimaging data of core circadian genes *Period* and *Cryptochrome* from mouse SCN.

We begin by introducing the mammalian clock and SCN bioimaging data. We then cover various methodologies for mechanistic and stochastic modelling of gene transcription, including chemical reaction networks, the chemical Langevin equation, and Markov chain Monte Carlo methods for Bayesian inference. We derive stability criteria for a model of the single-cell TTFL that describes transcriptional inhibition through a distributed delay. The model is fitted to imaging data of the gene Cry1, which allows us to infer the dynamics of circadian gene transcription and molecular population sizes.

A Bayesian hierarchical framework is developed to model spatial dependencies observed in the parameter estimates of the single-cell model. The methodology is applied to bioimaging data of the Cry1-gene and the analysis tools are developed further by deriving a Bayesian period estimator and an inhibition profile which allow us to study the spatial distribution of key properties of the TTFL across SCN tissue.

Finally, the methodology is extended to include an additional molecular species that captures transcriptional activation. This extension confers a mechanistic spatial interpretation to the model by describing the effect of intercellular signalling. By eliciting informative prior distributions for parameters of the circadian Per2 feedback loop, we are able to fit the model to simultaneous recordings of Per2 and calcium. The model fit represents a first step in obtaining a complete model of both single-cell and organ-wide dynamics with empirically estimated parameters.

Abbreviations

AVP Arginine vasopressin

CAR Conditional autoregression

CLE Chemical Langevin equation

CME Chemical master equation

CRE Calcium responsive element

CREB CRE binding protein

CT Circadian time

CTS Characteristic time scale

DDE Delay differential equation

EKBF Extended Kalman-Bucy filter

ESJD Expected squared jumped distance

ESS Effective sample size

HPDI Highest posterior density interval

IP Inhibition profile

LBM Lugsail batch means

LNA Linear noise approximation

MCMC Markov chain Monte Carlo

MJP Markov jump proccess

MRF Markov random field

mRNA Messenger ribonucleic acid

ODE Ordinary differential equation

PSD Power spectral density

RWM Random walk Metropolis

SCN Suprachiasmatic nucleus

SD Standard deviation

SDE Stochastic differential equation

SSA Stochastic simulation algorithm

TTFL Transcriptional/translational feedback loop

VIP Vasoactive intestinal peptide

WT Wild type

List of Tables

List of Figures

Chapter 1

Introduction

The work presented in this thesis revolves around statistical analyses of circadian bioimaging data. Biologists have developed methods to observe and record the gene-regulatory clockwork of circadian rhythms on a microscopic scale and the resulting data are often analysed using standard spectral analysis or other phenomenological models. While such tools may already be implemented in scientific software and require little computational power and customisation, they typically disregard potentially interesting aspects of the rich data, such as intrinsic stochasticity due to molecular noise and spatial correlations across imaged tissues. Throughout the thesis we develop and apply statistical methodology that allows for further scientific investigation of circadian gene transcription. This is achieved, wherever feasible, through mechanistic modelling of the underlying biological and experimental processes and Bayesian parameter inference, producing biologically interpretable results along with uncertainty quantification.

1.1 Outline

The thesis is structured into seven chapters. Chapters 3, 5 and the first half of Chapter 7 are focused on theory while Chapters 2, 4, 6 and second half of 7 revolve around applications to experimental data and are written with accessibility to a wider audience of anyone interested in modelling of gene transcription in mind.

The first and present chapter provides a motivating introduction to molecular oscillators, circadian gene expression and the mammalian clock. Competing modelling approaches are discussed and modelling that accommodates varying sources of stochasticity are motivated, both from a biological and more general, scientific perspective.

Chapter 2 introduces bioimaging data from mammalian SCN tissue, which moti-

vates a large part of the methodology developed throughout the thesis. An exploratory analysis of such data using the spectral bootstrap is presented, which provides a means to quantify the level of variation in periodicity, phase and amplitude of circadian gene expression.

Chapter 3 contains most of the mathematical techniques and theory utilized throughout the subsequent chapters. We move away from spectral methods and instead review and discuss mechanistic modelling approaches based on chemical reaction networks. Methodology that circumvents the inferential intractability of highly detailed models is covered, such as the Chemical Langevin Equation and likelihood approximation using Extended Kalman-Bucy filtering. Differential equations with distributed delays are reviewed along with stability theory for such equations.

A large portion of the chapter is devoted to Markov chain Monte Carlo methodology for Bayesian parameter inference. Adaptive and delayed acceptance random walk Metropolis algorithms are discussed in detail, building the foundation for their application in subsequent chapters. Chapter 3 is concluded with a novel measure of "robustness" of biological system that arises naturally in Bayesian inference and is consistent with general definitions robustness available in the systems biology literature.

In Chapter 4 a model for circadian mRNA transcription that describes negative auto-regulation through a feedback loop is reviewed and stability criteria for the macroscopic mean are derived. The methodology from the previous chapter is used to infer parameters of the transcription model using bioimaging data from organotypic SCN tissue where the circadian gene Cry1 is imaged. The methodology is geared towards individual time series representing the concentration of some molecular species over time, hence inference is repeated a large number of times at locations across the SCN tissue to obtain spatial distributions of parameter estimates, allowing investigation of functional variation between sub-regions of the SCN.

The spatially independent inference is extended in Chapter 5 to a spatio-temporal setting using a flexible Bayesian hierarchical model with parsimonious Markov Random Field prior distributions. The methodology is adapted to the circadian gene transcription model from the previous chapter and issues concerning the computational costs are discussed. We show how to implement an algorithm that explores the posterior distribution, while taking advantage of multiple processor cores for the likelihood approximation which dominates the computational cost. The parallel implementation reduces the computational cost approximately linearly with the number of available cores, making the inferential framework feasible on modern desktop PCs.

The sixth chapter provides an empirical application of the spatio-temporal method-

ology developed in Chapter 5. Three experimental replicates of Cry1-luc expression data are analysed in high resolution, which is made possible by the efficient and parallel algorithm for sampling the posterior distribution. The mechanistic model along with the parameter uncertainty quantification provided by the Bayesian methodology allows for scientific investigation of various properties of the SCN.

In the seventh and final chapter the modelling and inferential methodology is extended to a setting where a second transcription factor governs transcriptional regulation, in addition to the delayed auto-regulatory mechanism considered in the preceding chapters. The methodology is applied to simultaneous recordings of Per2:luc and GCaMP3-reported calcium concentration from mouse SCN.

1.2 The mammalian circadian clock

Most living organisms exhibit rhythmic behaviours tuned to the natural day-night cycle of their environment. These rhythms are self-sustained and entrainable to external cues such as light and temperature (Takahashi, 1995). In mammals, metabolic processes in most cells oscillate with a 24-hour cycle but rhythms of peripheral tissues are governed by what can be described as a "master clock" located in the suprachiasmatic nucleus (SCN) in the anterior hypothalamus. The SCN is a master clock in the sense that the constituting neurons exhibit an endogeneous rhythm of gene transcription, i.e. sustained oscillations that do not require an oscillating input. Furthermore, the SCN is able to coordinate the rhythms of other tissues (Meijer and Rietveld, 1989). Experiments have shown that cell cultures of peripheral tissues sustain oscillations for a few cycles before dampening while SCN neurons are able to maintain a 24-hour cycle for weeks without external input (Reppert and Weaver, 2002).

The biological mechanisms that constitute the clock were first studied in *Drosophila melanogaster* (Konopka and Benzer, 1971), *Pisum sativum* (Kloppstech, 1985) and a few other species accessible to early genetics. The current model organism is the mouse and homologues of several of the genes responsible for the clock in Drosophila are found in mammals (Reppert and Weaver, 2002). Studies of the circadian clock have over the years come to garner large attention and the 2017 Nobel prize in medicine was awarded to the three chronobiologists Jeffrey Hall, Michael Rosbash, and Michael Young for their respective discoveries of circadian mechanisms (Burki, 2017).

The working hypothesis is that oscillations in the SCN and thus behaviour aligned with the solar cycle is achieved through auto-regulatory transcriptional/translational feedback loops (TTFLs) where protein products repress the transcription of their re-

sponsible genes. Two interlocked TTFLs are thought to be the main mechanism for the stable oscillations in the mammal SCN (Hastings et al., 2008). In the primary loop heterodimers of transcription factors CLOCK and BMAL1 promote transcription of *Period* and *Cryptochrome* genes (Per1-3, Cry1 and Cry2 in mice) by binding to E-box motifs in the promoter-region of Pers and Crys. Negative autoregulation of transcription occurs when PER:CRY heterodimers translocate back to the cell nucleus and act on the CLOCK:BMAL1 complex, thus inhibiting further transcription (Shearman et al., 2000). The secondary loop consists of CLOCK:BMAL1 activating transcription of nuclear receptors Rev-erbα and Rorα, whose protein products respectively repress and activate *Bmal1*, producing oscillatory behaviour of *Bmal1* (Ko and Takahashi, 2006).

While the TTFL models oscillations on the cellular level, the neurons comprising the SCN belong to highly synchronised sub-populations (Yamaguchi et al., 2003). Transcription of Per2 typically exhibits a spatial wave emanating from the dorsomedial region and travelling to the ventrolateral (Evans et al., 2013; Pauls et al., 2014), see Figure 2.1 for a diagram of the SCN labeled by region. Taylor et al. (2017) find that the typical phase difference of peak Per2 expression between dorsomedial and ventrolateral neurons is 2.4 hours. Evidence suggests that not only is inter-neuronal connectivity important for robust oscillations, it also sets the SCN aside as the master clock (Mohawk and Takahashi, 2011). Furthermore, experimental protocols with changes in light-dark cycle show that retino-recipient parts of the SCN (ventrolateral) respond and adapt Per1 expression faster to such external stimuli and it is suggested that this is then relayed to dorsomedial parts of the SCN with fewer retinal fibers (Nakamura et al., 2005). Gu et al. (2017) desynchronize the ventrolateral and dorsomedial subpopulations using an artificial 22 hour light dark cycle (11:11) and show that with an artificial cycle length sufficiently different from the normal 24h the SCN cannot completely entrain to external cues. The light-information insensitive dorsomedial region keeps its intrinsic 24h period while the ventrolateral region adapts to information regarding new light schedules from the optic chiasma.

Cytoplasmic Ca^{2+} (calcium) is also highly circadian in SCN neurons and is believed to play an important role as input/output in both TTFLs of individual cells as well as in synchronisation of separate cellular oscillators (Hastings et al., 2014). The neuropeptide vasoactive intestinal peptide (VIP) is only present in the retino-recipient ventrolateral SCN neurons (Hastings et al., 2014) and is believed to carry a synchronising mechanism (Maywood et al., 2006). Specifically, VIP binding to the VPAC2 receptor activates Gq-signalling and leads to an increase in intracellular calcium which promotes transcription of Per1 and Per2 genes that carry a calcium responsive element (CRE)

(Hastings et al., 2014). Brancaccio et al. (2013) show experimentally that VIP knock-out SCN slices loose synchrony of oscillations of CREs. Furthermore, Irwin and Allen (2010) find that high levels of VIP suppress calcium and does so to a greater extent during daytime when calcium levels are generally high.

A better understanding of the machinery that governs circadian rhythms is important as there is evidence for a link between misalignment or disruption of the circadian rhythm and increased risk of various chronic diseases such as cancer (Innominato et al., 2014), metabolic syndrome (Shi et al., 2013) and psychiatric disorders (Coogan et al., 2013). Disruption of the circadian rhythm is classified as a probable carcinogen by the World Health Organization, and several studies have found links between Cry1-2 genes and tumour formation, where inactivation of these genes led to decreased survival rate in mice exposed to radiation (Chan and Lamia, 2020). In Chapters 4 and 6 of this thesis expression of the Cry1 gene is studied in detail. A research area of specific impact is the optimisation of medical interventions, both surgical and drug intake, with respect to the circadian rhythm known as *chronotherapy* (Lévi, 2001), which is shown to improve the prognosis of cancer patients (Lévi et al., 1999).

Chapter 2

Data & exploratory analysis

This chapter acts as a introductory text to organotypic SCN imaging data that we return to in several of the following chapters. Simultaneous recordings of Per2:luc and GCaMP3-reported calcium from Brancaccio et al. (2013) are investigated and spatial distributions of period, phase and amplitude of their circadian oscillations are estimated using spectral analysis. In addition, uncertainty quantification of the estimates is obtained using the spectral bootstrap method of Costa et al. (2013). For details on the experimental protocol used to obtain the recordings, and the two fusion constructs ":" and "-" we refer to Brancaccio et al. (2013), however a short description of the experimental procedure is given in this chapter (and expanded on in the first section of Chapter 4 to motivate some of the modelling assumptions made there).

Summarising the results, we find that circadian oscillations of Per2 and calcium are laterally symmetric and mirrored for the two halves of the SCN. While Per2 is primarily expressed in SCN tissue, calcium is found in adjacent tissues but only oscillates with the clock in SCN tissue. Calcium is likely to have less well defined oscillations than Per2 in the SCN as we find it has a lower signal-to-noise ratio, although the differences in experimental method for the two species could in part explain this difference. Circadian expression of Per2 exhibits a clear spatial wave-like trajectory that begins in the region next to the third ventricle (see Figure 2.1). Our findings regarding the phase delay between early and late Per2 peaks are consistent with those in the literature, with the addition of the finding that the earliest tissues in terms of Per2 expression tend to show non-circadian calcium expression (defined as a period length outside 18-30 hours). The spatial phase difference for calcium is less pronounced, although the uncertainty in the phase estimates is considerably higher for calcium. We find that both Per2 and calcium exhibit fairly exact 24 hour oscillations in SCN tissue, and the uncertainty in

the estimates, in terms of bootstrap confidence intervals, are similar despite the overall higher level of noise in the calcium signal.

In the final section of the chapter we discuss how the results from the exploratory analysis can be translated into insights that help in a more detailed, mechanistic modelling approach. Furthermore, limitations of the phenomenological modelling approach are discussed which serves as motivation for the methodology developed and presented in the rest of the thesis.

2.1 Exploratory analysis of Per2 and calcium expression in mammal SCN

This section aims to explore spatio-temporal features of Per2 and calcium expression across the SCN. Spectral analysis and bootstrap are used to determine period and phase of their respective reporter constructs. Uncovered features along with biological knowledge motivate the structure of models used in subsequent chapters, and for some of the modelling assumptions made to facilitate parameter inference. This section begins with a presentation of the available data and methods and further subsections are devoted to presenting the results of the analysis.

A schematic of a cross-sectional cut of the SCN labeled with the relevant regions is given in Figure 2.1. The organ consists of two laterally symmetric halves which in mice contain roughly 10 000 neurons each (Welsh et al., 2010). Each half can be anatomically subdivided into the dorsomedial (DM, or shell) and ventrolateral (VL, or core), which connects to and receives photic input from the optic chiasma (Abrahamson and Moore, 2001). In the following chapters we use shell/DM and core/VL interchangeably, and when reference is made to the interior of either of the two halves the term *central* SCN is used. A third axis, the rostral-caudal direction lies perpendicular to the two-dimensional slice of the SCN. In general we adhere to the assumption that experimental replicates originate from the approximate center of the third axis such that central locations of the slices are in the center of the three-dimensional organ.

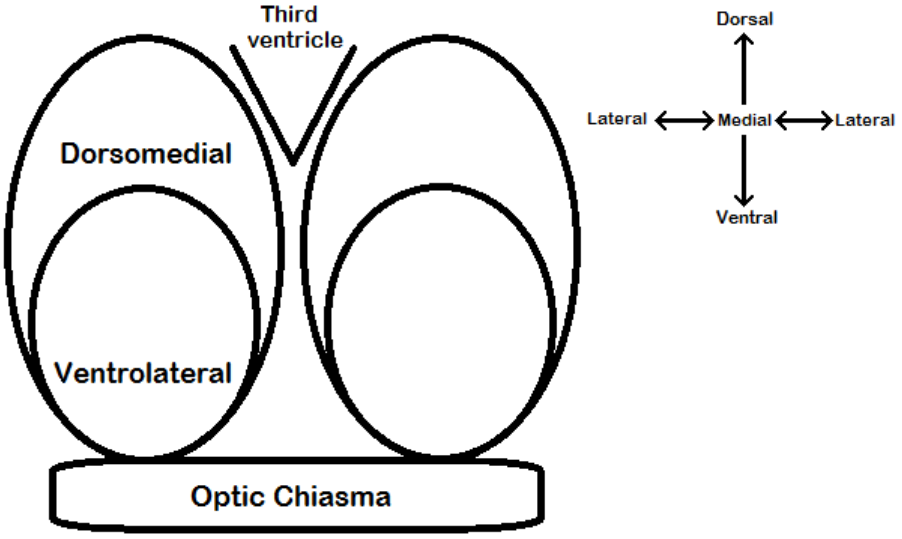

Figure 2.1: Diagram of the SCN, labeled with regions of interest. The anatomical subdivision of the two halves is laterally symmetric and each half is made up of approximately 10^4 neurons. The dorsomedial region partially envelops the ventrolateral SCN, hence is referred to as the SCN shell. The ventrolateral region is retinally innervated via the optic chiasma and receives photic input that is transmitted to the rest of the SCN via a combination of neuronal firing and neurotransmitter signalling.

2.1.1 Available data

Brancaccio et al. (2013) use viral transduction and flourescence/bioluminescence imaging to obtain simultaneous recordings of Per2 and calcium in mouse SCN. Calcium is reported by synthetic GCaMP3 which is activated without delay and Per2 is imaged using a genetically encoded luciferase contstruct Per2:luc (Yoo et al., 2004) where the reporter is delayed circa 1 hour, which accounts for transcription and translation (Brancaccio et al., 2013). Organotypic slices of the SCN were kept in medium while luminescence was recorded every 0.5h for 4-5 days. The experimental procedure yielded 183 frames of 423 by 453 pixels for the Per2 and calcium sequences. For experimental details and data see Brancaccio et al. (2013), Movie S3.

The extent of a typical neuron is roughly the size of an 8 by 8 block of pixels, although the neurons have highly irregular shapes. In order to reduce computational cost and obtain time series from tissue that more closely resembles the extent of neurons each frame is aggregated by taking the cross sectional arithmetic mean of blocks of 4

by 4 raw pixels. The aggregation can be assumed to preserve spatial features that lie on the scale of the regions of the SCN (see Figure 2.1) consisting of tens or hundreds of neurons, while leaving stochasticity at the approximate single cell level of interest.

A trend is present in most of the resulting aggregate pixels which is due to the experimental procedure: for Per2 the trend is negative as the luciferin substrate is consumed over time. The calcium reporter on the other hand exhibits a positive trend. Therefore the data are de-trended by fitting a linear trend using least squares to the logarithm of the data. As two different reporter constructs are used for calcium and Per2 there is no way of comparing the amplitude of oscillations of their respective underlying chemical species; hence, we do not require a transformation that preserves the relative amplitude of the two species.

Figure 2.2 gives a representative picture of the peak expression of Per2 (left) and peak calcium concentration (right). The image is calculated by taking the maximum for each pixel across the 91.5 hour recording. This is done to circumvent the fact that SCN neurons are not expressing circadian genes in complete synchrony but with a spatial phase delay on the order of hours. High peak expression of Per2 is contained within the two halves of the SCN whereas calcium is present in surrounding tissues in relatively higher concentrations.

Figures 2.3 and 2.4 contain pictures of the SCN with a suitably chosen 5 by 4 grid of pixels (left) and time series of these pixels (right). The blue time series refers to raw data and the red to the detrended series. The transformation works sufficiently well to remove any trend in most of the grid. It is worth noting that some pixels show plateaus during peaks (Figure 2.3), presumably because of overexposure in the recording. These truncated peaks may in turn cause issues with automated image analysis, however substantial truncation is only present in the first cycle and the problem is alleviated when data are log transformed. For Per2, pixels corresponding to the center of each half of the SCN have well defined peaks whereas pixels located along the edges have lower amplitude. Calcium has a visibly lower signal-to-noise ratio compared to Per2 but a similar spatial pattern with sharp peaks in central regions. Taken together with Figure 2.2, this suggests that while calcium is present in surrounding tissues it only oscillates with the 24 hour clock in the SCN. This finding is consistent with the fact that VIP is not present in surrounding tissues along with the findings of Irwin and Allen (2010) that VIP suppresses calcium levels during day time. Overall the data show clear signs of circadian oscillation with 4 distinct peaks over a circa 90 hour time frame.

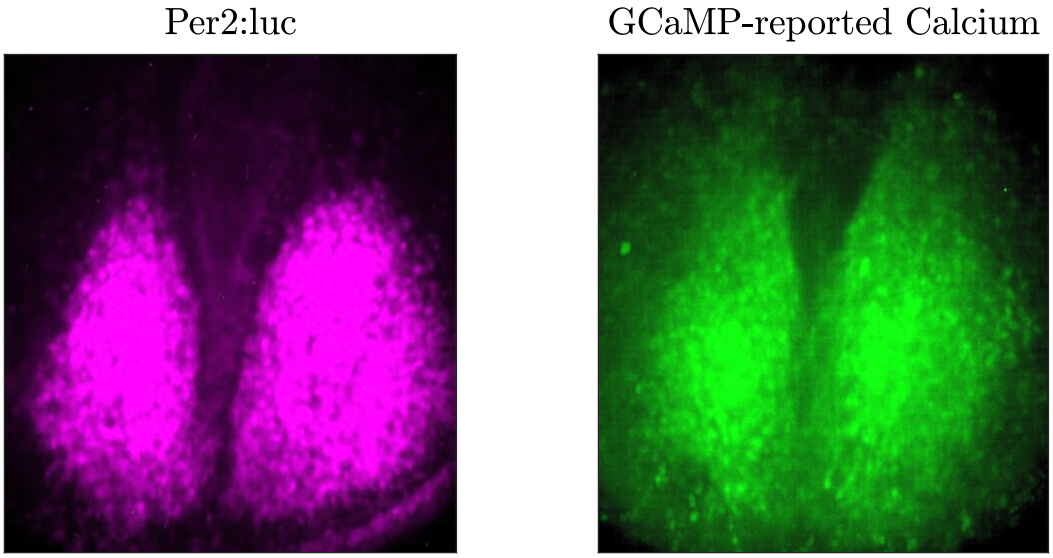

Figure 2.2: Representative image of Per2:luc (left) and GCaMP3-reported calcium (right) bioluminescence across the two halves of the SCN. The image is obtained by taking the pixel-wise maximum across the length of the experiment (approximately 4 days). Bioluminescence/flourescence of reporter constructs is used as a proxy for gene expression and calcium concentration. Per2 reporter has a circa 1 hour delay accounting for transcription and translation whereas the calcium reporter is instantaneous.

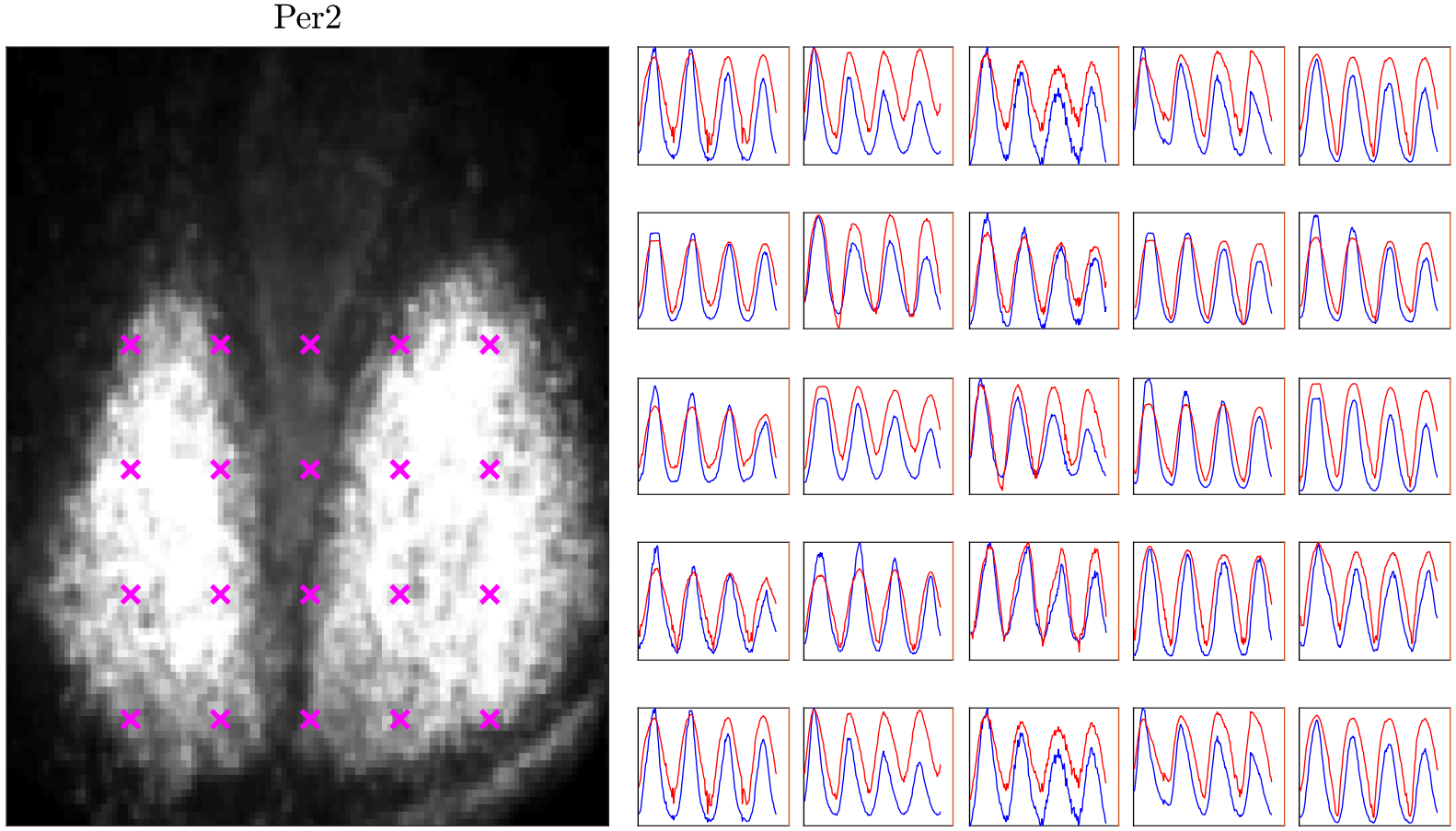

Figure 2.3: Left: Representative image of Per2 expression with selected pixels in magenta. Right: Raw (blue) and transformed (red) time series for the selected pixels over approximately four days. The time series are de-trended by fitting a linear least squares trend to the logarithm of data.

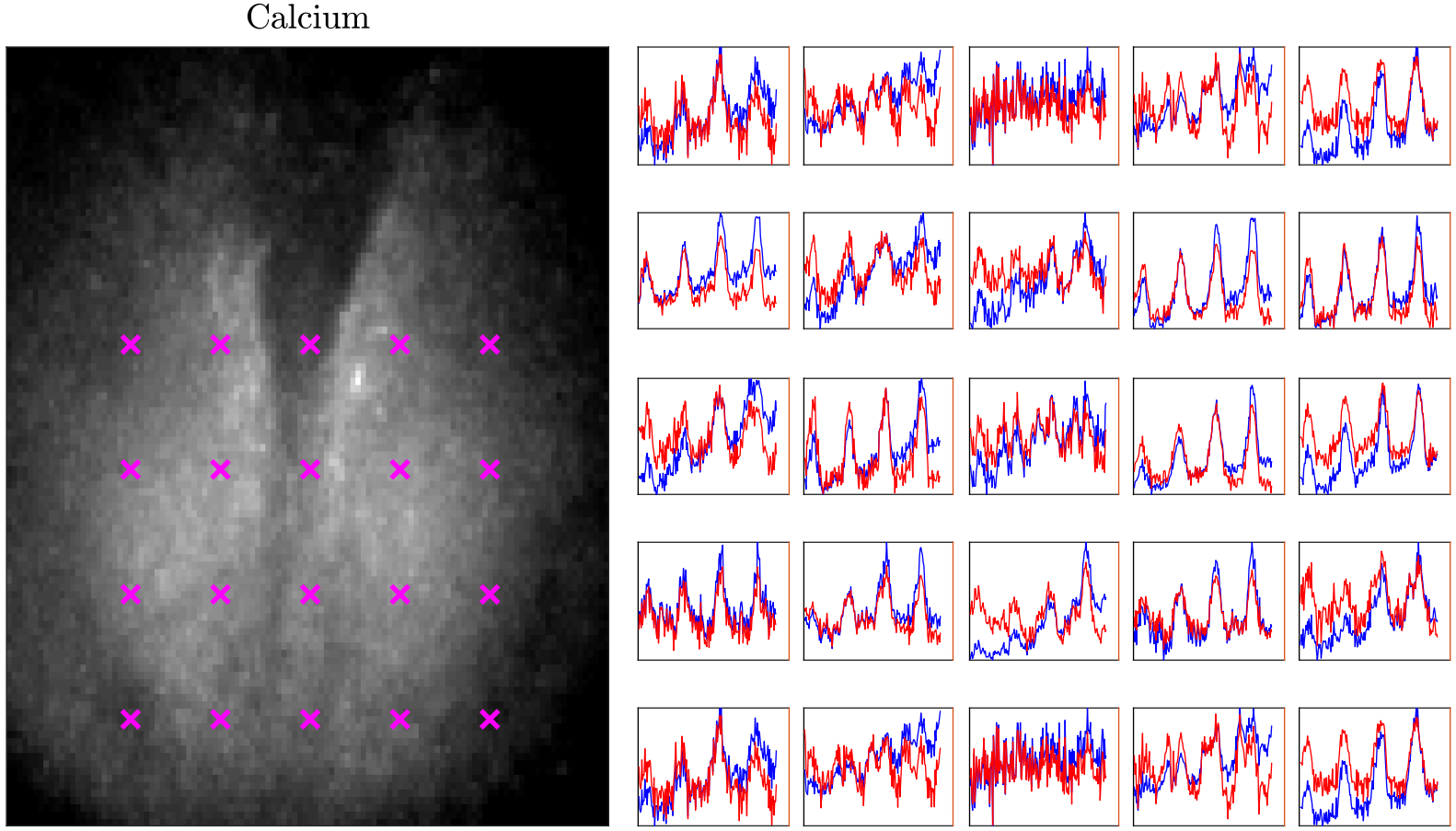

Figure 2.4: Left: Representative image of Syn-GCaMP3 reported calcium expression with selected pixels in magenta. Right: Raw (blue) and transformed (red) time series for the selected pixels over approximately four days. The time series are de-trended by fitting a linear least squares trend to the logarithm of data.

In an initial exploratory analysis of SCN imaging data we shall employ spectral analysis where each pixel-block is considered to be a time series. Spectral analysis relies on the fact that any signal can be represented as a combination of frequencies, where the Power spectral density (PSD) for the random signal $y(t)$, $t > 0$, sample size n and frequency $\omega > 0$, defined as

$$f(\omega) = \lim_{n\to\infty} \mathbb{E}\left[\frac{1}{n}\left|\sum_{t=1}^{n} y(t)e^{-i\omega t}\right|^2\right] \tag{2.1}$$

where $i = \sqrt{-1}$, determines the weight of component ω in the frequency domain representation (Stoica and Moses, 2005). The asymptotically unbiased, sample equivalent of Eq. (2.1), the *periodogram*, given by

$$\hat{f}_p(\omega) = \frac{1}{n}\left|\sum_{t=1}^{n} y(t)e^{-i\omega t}\right|^2 \tag{2.2}$$

is easy to compute using the fast Fourier tranform. It is however an inconsistent estimator of the true PSD. A better estimator, in terms of bias-variance trade off, is obtained by applying a window function to the observed signal (Percival and Walden, 1993). A variety of window functions for this purpose with slightly different properties are available in the literature, see e.g. (Stoica and Moses, 2005, Chapter 2) for an (incomplete) list.

Once a satisfactory estimate of the PSD is obtained, the period is simply the inverse of the frequency corresponding to the largest peak of the PSD. As the data in this chapter exhibit clear circadian oscillations we expect to find a large peak in the PSD around the frequency corresponding to 24 hours. However, the molecular oscillators are not exact 24 hour clocks, hence we report the spatial distribution of period estimates of the *dominant* frequency, i.e. the frequency at which the estimated PSD is maximised. Similarly, the *phase* of the dominant frequency is the angle, $\tan^{-1}$, of the complex Fourier coefficients whose squared magnitude are the maximum of the estimated PSD.

To quantify the uncertainty in period and phase estimates we resort to using the width of bootstrap confidence intervals. Costa et al. (2013) propose a spectrum bootstrap methodology for period estimation which uses the fact that the empirical spectrum can be approximately described by

$$I(\omega_k) = f(\omega_k)\epsilon_k, \;\; k = 1, ..., n/2, \tag{2.3}$$

where $\{\epsilon_k\}_{k=1}^{n/2}$ is a sequence of i.i.d. standard exponential random variables, $f(\cdot)$ is the true spectrum, $I(\cdot)$ is the empirical counterpart evaluated at frequencies ω_k and n is the sample size in terms of time. A consistent estimate of $f(\omega_k)$ is needed to resample the multiplicative residuals in Equation (2.3). First $\hat{f}_{b^\dagger}(\omega_k)$ is estimated and residuals calculated. Bootstrap samples of the periodogram are then calculated using another estimate of the spectrum $\hat{f}_{b^\ddagger}(\omega_k)$. The bootstrap sample of the spectrum is finally calculated by smoothing the periodogram sample to get $\hat{f}_b(\omega_k)$. The three different smoothing parameters used are, $b = cn^{-1/5}$ $b^\dagger = cn^{-1/4}$ $b^\ddagger = cn^{-1/6}$. The three smoothing parameters act to balance bias and variance of the estimator (Costa et al., 2013). We obtain phase estimates in a similar fashion and only analyse the phase and period of the dominant circadian frequency, i.e. the maximum of the spectral density estimate if that mode corresponds to a frequency between 18 and 30 hours. Matlab code provided as supplementary material to Costa et al. (2013) is used for the calculations.

To avoid that the resulting period estimates look discrete, zero-padding is used for the initial spectrum estimate. The effect of this is an interpolation of the peak of the spectrum when the temporal resolution is not sufficiently high (Percival and Walden, 1993, Chapter 3). The lengths of the series are 183 observations and these are padded with zeros such that the lengths of the padded series are 2048. The bootstrap algorithm is run for each block of 4 by 4 pixels, the bootstrap sample size is 200 and the percentiles that correspond to a 95 percent confidence interval are stored. The bootstrap sample size of 200 is motivated by the computational load, where three major reasons contribute: temporal resolution (padding), spatial resolution (large number of pixels) and bootstrap sample size. Spatial and temporal resolution are vital as dynamics along these axes are investigated.

For point estimates of period, phase and amplitude we resort to non-parametric spectral analysis. After detrending, applying a Hann window (Percival and Walden, 1993, Chapter 6), centering and zero-padding, each time series is decomposed using the Fast Fourier Transform. The frequency that maximizes the spectral density estimate is converted back to hours for a period estimate and the Fourier coefficients are used to calculate the phase of the dominant frequency, mapped to circadian time (CT0-24). Following Brancaccio et al. (2013), the phase mapping is done such that the peak of Per2:luc expression is at CT12 for the average pixel block. This utilizes biological knowledge of when Per2 expression typically peaks and the spatial variation of phase delay between calcium and Per2 can be investigated.

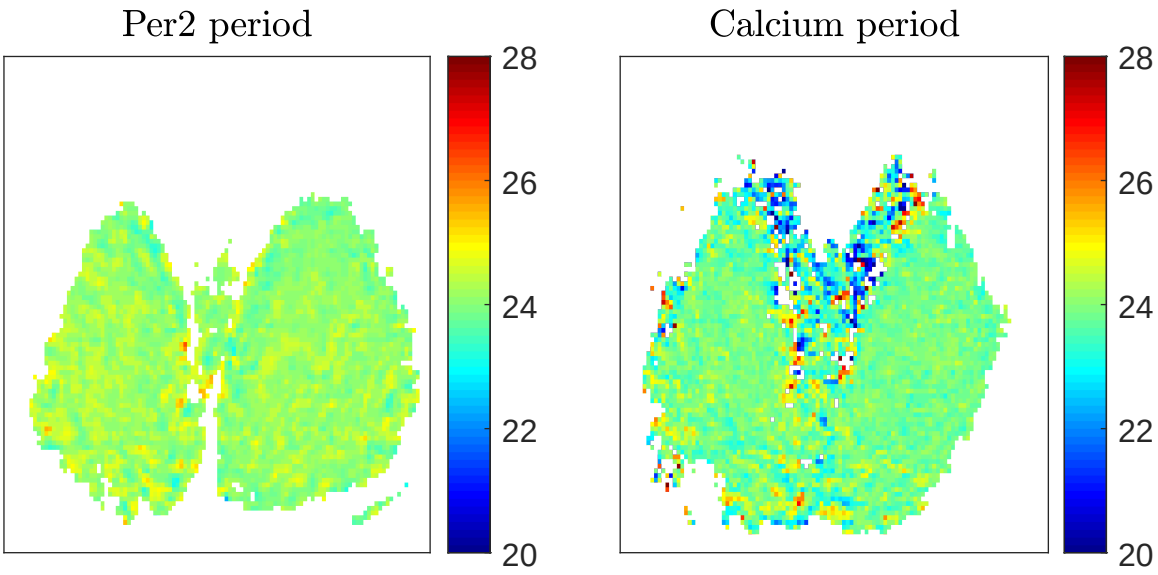

Figure 2.5: Heatmap of period of dominant frequency of Per2 (left) and calcium (right). Pixels with low luminosity (below average across locations and time points) or non-circadian expression (<18h or >30h) are omitted. Colours are mapped to period in hours.

Figure 2.5 shows point estimates of the period of the dominant frequency of circadian Per2 (left) and calcium (right) across the SCN, mapped to colour where blue represents shorter periods and red longer. The period of Per2 is spatially homogeneous and close to 24 hours over the whole set of pixels. The homogeneity of the perioid is to be expected as Per2 is a core clock gene and the SCN is the master clock. Nearly all examined sites exhibit a period between 22 and 26 hours. The mean period of Per2 for the shown sites is 24.16 hours. Calcium on the other hand has varying period length and some tissue exhibit non-circadian expression, especially along the third ventricle (between the two halves). In the dorsal and ventrolateral region the period length is however close to 24 hours with lower spatial variation locally. The mean period length for calcium across examined sites is 23.82 hours. This is consistent with the findings of Enoki et al. (2012) which showed that calcium has a homogeneous phase between the two subregions.

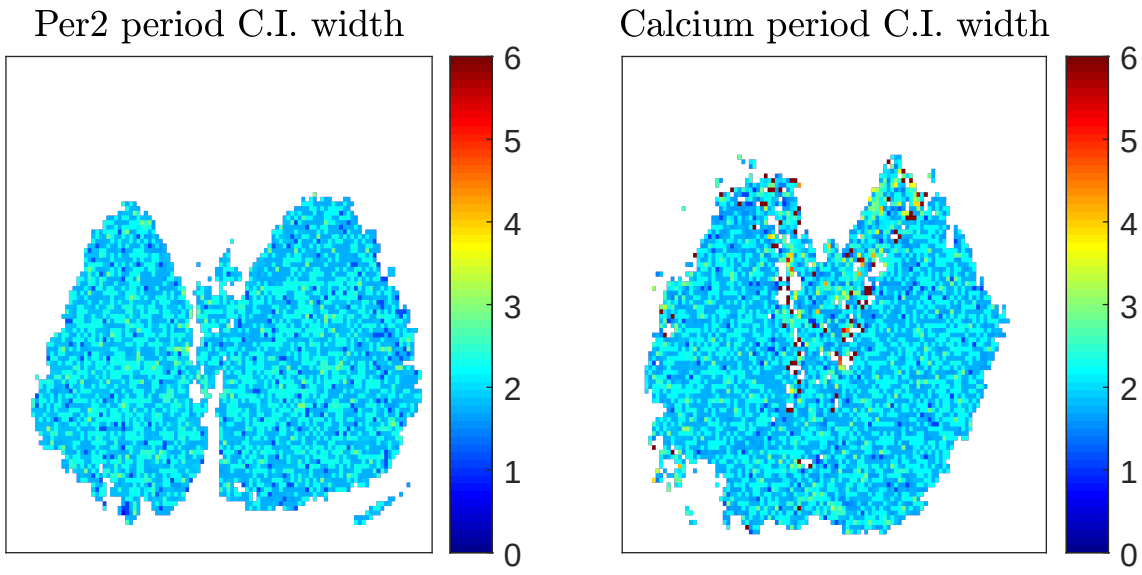

Figure 2.6: Heat map of width of 95 percent confidence interval of Per2 period (left) and calcium (right). Pixels with low luminosity (below average across locations and time points) or non-circadian expression (<18h or >30h) are omitted.

The confidence interval heat map in Figure 2.6 give estimates of the uncertainty in the period lengths. The width of the 95 per cent bootstrap confidence interval is mapped to color with blue representing a relatively narrow interval and red a wide interval (see colour bar for specific values). The width of the interval does not exhibit any clear spatial features for Per2 (left) and most intervals are narrow, less than 3 hours. This is consistent with the de-trended time series presented in Figure 2.3 showing well defined peaks and a high signal to noise ratio. Furthermore, it is consistent with the hypothesis that Per2 is a highly synchronised and exact component of the circadian clock across the SCN. Additionally, we find no differences in the period and phase between the left and right halves of the organ.

The homogeneity of the period, i.e. high synchrony of Per2 transcription across the SCN is important to note for our further analysis as any spatial variation of the phase of this dominant frequency cannot be attributed to differing periods nor their variability. This means that we are in fact investigating the properties of a fairly exact 24 hour clock. Similarly, due to the the stochastic and irregular behaviour of calcium along the third ventricle we do not expect any clear spatial features of the phase of

calcium in that region, as opposed to the smooth wave-like trajectory of Per2.

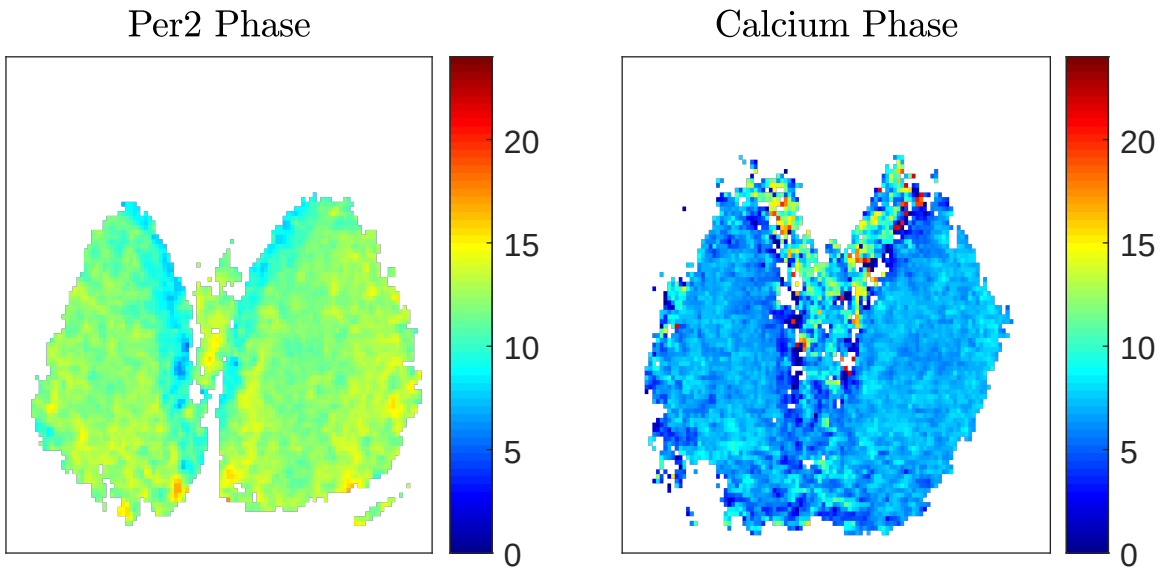

Figure 2.7: Heat map of phase of dominant frequency of Per2 (left) and calcium (right) mapped to 0-24 hours. Pixels with low luminosity (below average across locations and time points) or non-circadian expression (<18h or >30h) are omitted. Calcium is phase advanced compared to Per2, except for along bands along the dorsomedial region of the SCN. Circadian calcium has a homogeneous phase whereas Per2 shows clear spatial features.

The phase mapping in Figure 2.7 reveals clear spatial features of Per2 expression. The dorsomedial region is phase-advanced compared to the lateral and ventral regions. This is clear from the dominance of green areas in the top half of the organ compared to yellow in the bottom half and outer sides. This phase variation of Per2 transcription should however not be mistaken as a spatial flow of information between neurons, feeding in to the TTFL across the SCN network. Enoki et al. (2012) note that neuronal interactions are faster compared to the wave-like progression of Per2 transcription and instead the difference in phase is attributed to intrinsic differences between subregions such as differing periods. While the current data does not support differing period lengths between SCN subregions there is no evidence of a causal phase difference either.

As Per2:luc is mapped to peak at CT12 on average, Calcium peaks at CT6.73 which is consistent with Brancaccio et al. (2013) on the same data and similar to the 5.67h phase advance of calcium compared to Per2:luc found by Enoki et al. (2012) on a different recording using a different reporter for calcium. These findings reinforce the idea that calcium is an input in the TTFL, at least in certain regions of the organ. Calcium itself is spatially homogeneous with peak expression between CT4 and CT10, except for a region around the third ventricle that peaks considerably later. This region corresponds to the same area that shows erratic period estimates in Figure 2.5 and it is likely to not be systematic but a consequence of the noisy periodicity.

Furthermore, a narrow band along the third ventricle (the area between the two halves) of the dorsomedial SCN exhibits a marked phase advance of Per2:luc compared to the rest of the organ. This phase advance is more evident on the left half of the SCN. The earliest peak of this region is around CT7.7 and when taking into account the delay of the luciferace construct used for Per2 (circa 1h) compared to the near instantaneous reporter of calcium these results suggest that the most phase advanced regions begin transcription of Per2 at around the same time, or right after circadian calcium peaks in the organ as a whole. This complicates the picture of calcium as an input to the TTFL, suggesting that such a mechanism is not identical across the SCN circuit.

In addition to Figure 2.7, Figure 2.8 gives the same phase mapping of calcium and Per2:luc but only displays Per2:luc in areas where calcium is *non-circadian*, i.e. the period estimate is <18h or >30h. A clear pattern emerges of medially located neurons coinciding with the markedly phase advanced population in Figure 2.7. These neurons exhibit peak Per2:luc considerably earlier at CT9.67 compared to the rest of the organ. This suggests two distinct sub-populations of SCN neurons. The phase delay of these medially located sites is around 2.3 hours compared to the SCN as a whole which is similar to the 2.4 hour phase difference between dorsomedial and ventrolateral regions described in Taylor et al. (2017) on a different set of data.

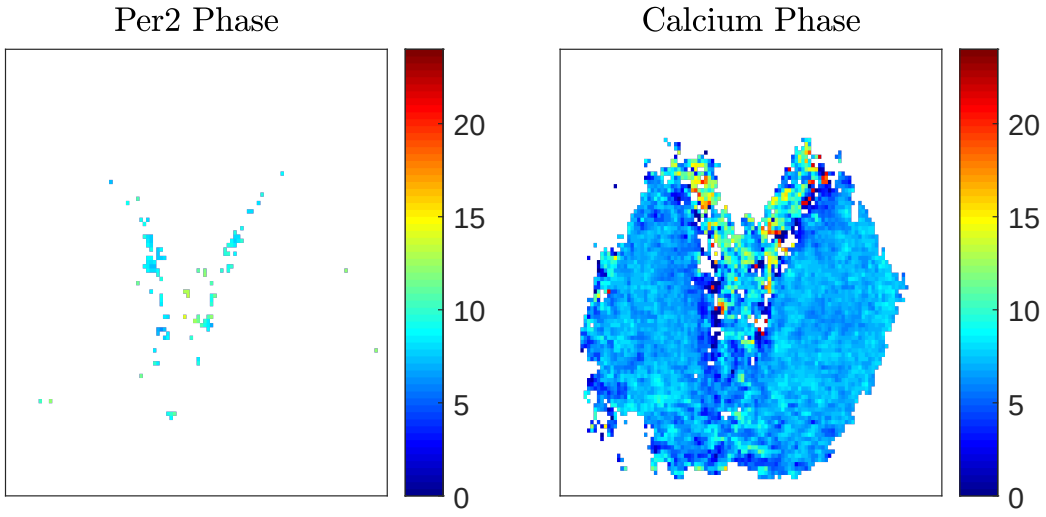

Figure 2.8: Heat map of phase of dominant frequency of Per2 (left) and calcium (right) mapped to 0-24 hours. Pixels with low luminosity (below average across locations and time points) or non-circadian expression (<18h or >30h) are omitted for calcium. Per2 pixels are only shown where calcium appears non-circadian.

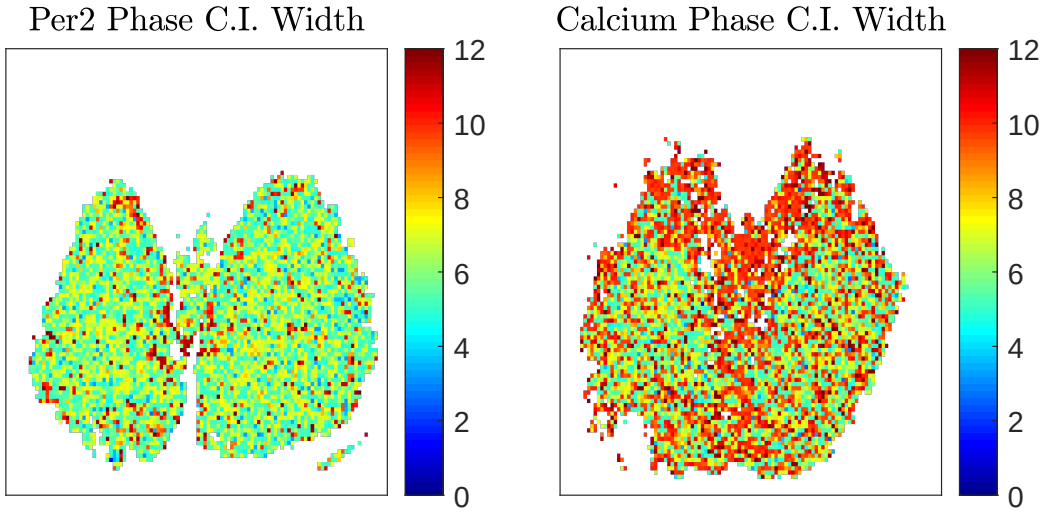

Figure 2.9: Heat map of width of 95 percent confidence interval of phase. Pixels with low luminosity (below average across locations and time points) or non-circadian expression (<18h or >30h) are omitted.

The width of a 95 per cent bootstrap confidence interval for phase is given in Figure 2.9. There are no clear spatial features for the Per2:luc interval (left), except possibly a slight clustering of wide (>10 hours) intervals along the third ventricle. Most of the intervals are however between 4 and 8 hours. Calcium on the other hand exhibits wider intervals, especially towards the edges of the two halves. This reflects the higher noise in the calcium series visible in Figure 2.4 and the irregularity along the third ventricle visible in Figure 2.7.

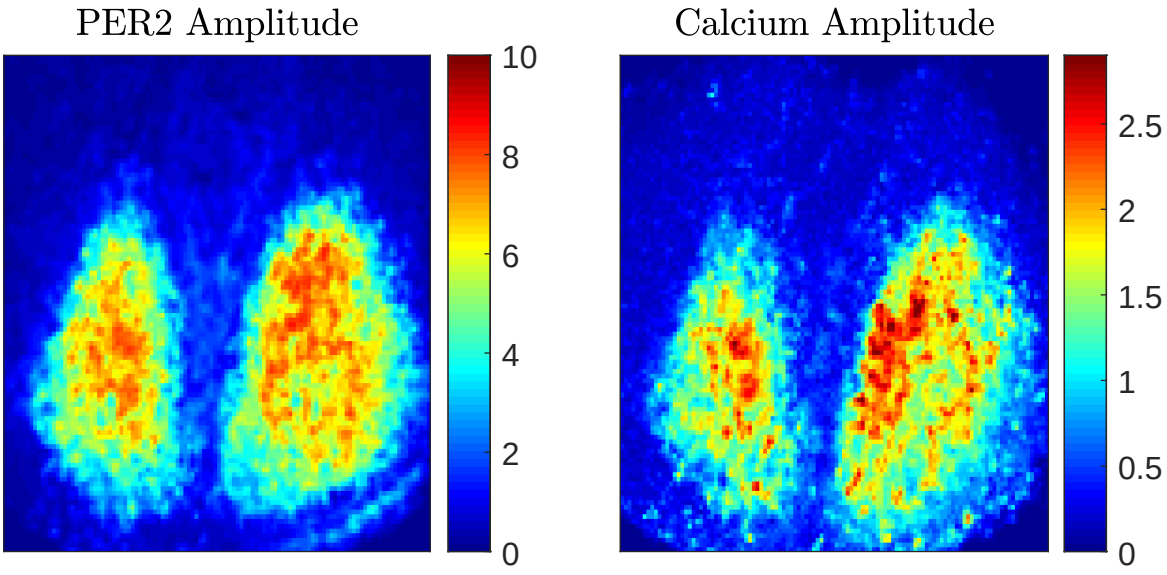

Figure 2.10: Heat map of amplitude of dominant frequency of Per2 (left) and calcium (right) mapped to 0-24 hours. Note that specific values of the amplitude cannot be compared between Per2 and calcium as the measurement processes are different.

The heatmap of amplitude (Figure 2.10) reveals oval-shaped regions in the centre of each half of the SCN with high amplitude and lower amplitudes along the edges. This is true for both Per2 and calcium but high amplitude regions for calcium are located more medially. The Per2 findings are consistent with Taylor et al. (2017) which reports essentially identical spatial organization of amplitude on another set of data. It has been hypothesized that cellular oscillators can be divided into weak and strong types, where weak oscillators are in fact superior synchronisers with faster phase response (Webb

et al., 2012). In Figure 2.10 the ventral regions of each half have a relatively low Per2 amplitude compared to the rest of the organ. A possible hypothesis is that photic input to these regions is communicated to the rest of the organ and these neurons have to be weakly oscillating in order to respond to and transmit external stimuli.

2.2 Discussion

The spatial and temporal properties of Per2 and GCaMP3-reported calcium across a mouse SCN slice in vitro were investigated using spectral analysis and spectral bootstrap. The results of the analysis were in line with previous empirical results regarding the mammalian clock, both on the same data (Brancaccio et al., 2013) and other data with slightly different experimental procedures and reporter proteins (Enoki et al., 2012; Taylor et al., 2017).

In general, the two halves of the SCN show very similar properties of Per2 and calcium oscillations. We find that calcium is present, albeit in lower concentrations, in adjacent non-SCN tissue but only oscillates with the clock in the SCN. We find a phase delay between early and late Per2 peaks that is well-documented in the literature (Evans et al., 2013; Pauls et al., 2014). The additional finding that the earliest tissues in terms of Per2 expression tend to exhibit non-circadian calcium expression (defined as a period length outside 18-30 hours) may be an intriguing phenomenon for further study.

Limitations of the spectral analysis approach employed in this chapter is that we cannot differentiate between phenomena that arise through the use of two different imaging techniques and differences in the dynamics of the underlying molecular species. For example, calcium is likely to have less well defined oscillations compared to Per2 in the SCN tissue as we find it has a lower signal-to-noise ratio, although this difference could in principle be caused by differences between the synthetic GCamp3 reporter and genetically encoded luciferase construct used to image calcium and Per2, respectively. Additionally, spectral methods rely on high frequency observations in order to provide results with high temporal resolution, while bioimaging data are typically available at sampling rates of 1 or 2 frames per hour due to the low levels of light emitted by the reporter protein. To overcome the limitations of phenomenological modelling on the level of observed data we aim to adopt and further develop approaches where the mechanism of gene expression *and* experimental procedure used to obtain data are explicit parts of the model formulation.

The analysis in Section 2.1 along with previous findings available in the literature on the network properties of the SCN make it clear that the SCN is not spatially

homogeneous and modelling on the level of a single cell oscillator cannot capture the full dynamics of the neuronal circuit which makes the master clock. Spatial properties should be considered in a realistic model, but it is not obvious what the proper approach is. There is evidence in the literature for distinct sub-populations of SCN neurons with slightly different functionality, e.g. strong and weak oscillators and local variation phase and amplitude (Taylor et al., 2017). A possible approach is to model pre-defined regions of the SCN, e.g. the core and shell separately. On the other hand, there is evidence for differences even within these putative sub-populations and there is the additional problem of defining a border between regions in experimental data. A less parsimonious model, capable of smooth spatial variation is likely to be required to accurately model local variation, such as the wave-like trajectory exhibited by Per2:luc transcription. Such fine-grained spatial variation could for example be captured with a hierarchical Bayesian approach where means of parameters are themselves modelled with a parametrised spatial variation.

The availability of simultaneous circadian gene and calcium recordings along with model formulations of a two-component clock (Calderazzo, 2016, Chapter 5), means that in principle a more complicated yet realistic description of the SCN TTFL can be investigated. A major obstacle in inference for detailed, realistic models of the clock is identifiability of parameters, due to the unavailability of data on the different molecular species involved in the cellular clock. Although the data set analysed in this chapter contains thousands of time series, interest primarily lies in differences in the dynamics those time series exhibit. In addition the time series are relatively short, less than 200 observations in this case, due to the low light intensity of reporter proteins requiring long exposure times to generate single frames and the difficulties in sustaining organotypic tissue in the required experimental settings for long periods of time. Bayesian inference provides a principled way to combine existing knowledge with data through the use of prior distributions. By developing informative prior distributions for a subset of the parameters of a mechanistic model it may be possible to fit more detailed models of the TTFL that capture the both the dynamics of individual SCN neurons and synchronising mechanisms between them.

Chapter 3

Reaction networks, inference & stability

This chapter introduces several key ideas and methodological components in a general setting with some motivating examples based on bioimaging data of gene expression. The methods covered here are subsequently used in an empirical setting for modelling and parameter inference in Chapters 4 and 6. This chapter begins with an introduction to chemical reaction networks along with the chemical Langevin equation and stochastic distributed delay models. Methodology for parameter inference is covered in the following two sections: likelihood approximation using the Extended Kalman-Bucy filter for delay models in Section 3.4 and Markov chain Monte Carlo for Bayesian inference in Section 3.5. The chapter is concluded by Section 3.6 where we develop a Bayesian robustness measure for oscillating biological systems.

3.1 Chemical reaction networks

This section begins by describing reaction networks and simulation strategies. Some approximate and deterministic modelling approaches are discussed and stochastic extensions are motivated from biochemical and statistical perspectives.

Chemical reaction network simply describe the interactions of a finite set of *species* through a finite set of *reactions* (Feinberg, 2019) and thus provide a tool to model chemical systems. An example of a simple model which incorporates inhibition, transcription, translation, and degradation of mRNA and protein is given by the following set of reac-

tions:

$$
\begin{aligned}
&G + P \rightleftharpoons H \\
&G \longrightarrow G + R \\
&R \longrightarrow R + P \\
&R \longrightarrow \emptyset \\
&P \longrightarrow \emptyset.
\end{aligned}
\tag{3.1}
$$

The reaction $G + P \rightleftharpoons H$ describes an auto-repressive mechanism where protein P binds to the regulatory region of gene G (creating a blocked region H). The binding reaction inhibits transcription by disabling $G \longrightarrow G + R$: the transcription of gene G resulting in mRNA molecule R. Translation of mRNA to protein is captured by $R \longrightarrow R + P$. Two-sided arrow indicate reversibility, i.e. that the reaction can occur in the opposite direction. In this case a blocked regulatory region can become unblocked, freeing up a protein molecule P and a regulatory region G. Finally, the reactions $R \longrightarrow \emptyset$ and $P \longrightarrow \emptyset$ describe degradation of mRNA and protein molecules, respectively. Note that this models degradation as an irreversible reaction where the molecules disappear from the system without leaving any waste products.

The classical, deterministic approach in chemical kinetics is to consider the variables continuous real-valued functions and analysing the system of reactions in Eq. (3.1) by invoking the *law of mass action*, i.e. that the rate of each reaction is proportional to the concentrations of each reactant (see e.g. Wilkinson (2011)). Let $P(t)$ denote the concentration of protein P at time t, and similarly for the other species G, H and R. By first specifying reaction rates c_i for each of the six reactions in Eq. (3.1) we can write a set of ODEs for the system where each equation describes the time-evolution of a reacting species. The corresponding ODEs for Eq. (3.1) are given by

$$
\begin{aligned}
\frac{dP(t)}{dt} &= -c_1 G(t) + c_2 H(t) + c_4 R(t) - c_6 P(t) \\
\frac{dG(t)}{dt} &= c_2 H(t) - c_1 P(t) \\
\frac{dH(t)}{dt} &= c_1 G(t) P(t) - c_2 H(t) \\
\frac{dR(t)}{dt} &= c_3 G(t) - c_5 R(t).
\end{aligned}
\tag{3.2}
$$

The equations in Eq. (3.2) provide a deterministic description of the system in Eq. (3.1). That is, subject to some initial condition $\phi_0 = (P_0 \ G_0 \ H_0 \ R_0)$ and reaction rates

c_i, $i = 1, ..., 6$, the concentration at any time $t > 0$ is exactly determined. Note that mass-action allows for a scale-free description and the real-valued functions in Eq. (3.2) are obtained by an implicit unit conversion from molecular population to *concentration*, e.g. by dividing each population by a common system size parameter Ω.

McQuarrie (1967) argues that such a deterministic representation is inadequate when the molecular population sizes are small, e.g. in biological cells, and introduces what has come to be known as the Chemical master equation (CME). The CME describes the time evolution of the probability distributions of the system and the possible states are the natural numbers. The CME hence provides a stochastic and exact description of the number of each type of molecule. Stochasticity arises from the fact that reactions are caused by molecular collisions and the initial condition only specifies population sizes of each reactant species, not their individual positions and momenta.

Gillespie (1992) provides general solutions of the CME for systems interacting through first-order reactions and a handful of higher order systems. The exact stochastic description is obtained by considering N species interacting through M reactions $R_1, ..., R_M$ inside a container of fixed volume Ω and a random state vector $\mathbf{X}(t) = (X_1(t), ..., X_N(t))$ where $X_i(t)$ is the number of molecules of the i^{th} species at time t. In general and unless otherwise stated, capital letters indicate random variables and lower case letters their realisations. Bold face indicates non-scalar quantities, i.e. vectors or matrices. Now define a set of propensity functions c_j, $j = 1, ..., M$ such that

$$\begin{aligned} c_j(\boldsymbol{x})dt \coloneqq &\text{ Probability that a reaction} \\ &R_j \text{ occurs in the time interval } [t,\, t + dt). \end{aligned} \tag{3.3}$$

Here we require dt such that the probability of more than one reaction is negligible compared to $c_j(\boldsymbol{x})dt$. Additionally define a M by N stoichiometry matrix $\boldsymbol{S}$ such that the element $s_{j,i}$ is the change in the i^{th} species due to the j^{th} reaction. We can now analyse conditional probability statements on the form

$$\mathbb{P}(\boldsymbol{X}(t) = \boldsymbol{x} | \boldsymbol{X}(0) = \boldsymbol{x}_0), \tag{3.4}$$

by the recursive relationship

$$\begin{aligned}\mathbb{P}(\boldsymbol{X}(t+dt)|\boldsymbol{x}_0) = \mathbb{P}(\boldsymbol{X}(t)|\boldsymbol{x}_0)\left(1-\sum_{j=1}^{M} c_j(\boldsymbol{x})dt\right) \\ +\sum_{j=1}^{M}\left(\mathbb{P}(\boldsymbol{X}(t)-\boldsymbol{s}_j|\boldsymbol{x}_0)c_j(\boldsymbol{x}(t)-\boldsymbol{s}_j)dt\right)+o(dt),\end{aligned} \tag{3.5}$$

which after rearranging and taking the limit $dt \to 0$ gives the CME

$$\frac{d}{dt}\mathbb{P}(\boldsymbol{X}(t)|\boldsymbol{x}_0) = \sum_{j=1}^{M}\left(c_j(\boldsymbol{x}-\boldsymbol{s}_j)\mathbb{P}(\boldsymbol{X}(t)-\boldsymbol{s}_j|\boldsymbol{x}_0) - c_j(\boldsymbol{x})\mathbb{P}(\boldsymbol{X}(t)|\boldsymbol{x}_0)\right). \tag{3.6}$$

The subtle non-linearity in Eq. (3.1), introduced by the reaction $G+P \longrightarrow H$ is however enough to make solutions to the CME cumbersome and it should be noted that from an applied perspective Eq. (3.1) is most likely a much too simplified model of gene regulation to provide any interesting biological insight. In other words, there is a gap between the systems we *want* to study and the systems we can study exactly.

Darvey and Staff (1966) prove that for general systems of N species interacting through first-order reactions, the CME provides expected values that coincide with the deterministic description. For systems with higher order reactions however the ODE representation and CME may predict different population sizes given the same initial conditions. In these higher order systems, the difference between the solutions provided by the deterministic set of equations and expected value of the stochastic CME is larger for small population sizes. As the population sizes grow the deterministic dynamics begin to dominate and the differences vanish. Hence the deterministic approach is considered a "macroscopic" rate equation, obtained when the container size of the system is scaled up sufficiently.

As the CME proves difficult to work with for reasonably realistic models, several approximations have been developed that retain the influence of noise but do not correctly capture all the moments of the probability distribution of the system states. These approximate methods include the the linear noise approximation (Kurtz, 1972; van Kampen, 2006) and the chemical Langevin equation (Gillespie, 2000) which we shall cover in detail in Section 3.4. Chemical kinetics is typically concerned with empirically determining the rate constants c_i in systems such as that in Eq. (3.2) and a less discussed motivation for a stochastic approach, approximate or exact, is that in a well-specified stochastic model the noise can aid in estimation of the rate parameters. The inclusion of

a diffusion term restricts the model by an additional equation that describes the expected size of random fluctuations about the mean that may enable structural identification. A model where the diffusion term is parametrised by the same rate parameters as the mean function poses, in a statistical sense, an easier inferential problem compared to inference based solely on the mean.

3.1.1 Stochastic simulation algorithm

While realistic systems typically are intractable in the sense that closed form expressions of the probability distributions of states do not exist, i.e. solutions to the CME do not exist, it is possible to draw samples from the distribution of possible states at any time $t > 0$ using the stochastic simulation algorithm (SSA) (Gillespie, 1977). By assuming that the reactions take place in a container of fixed volume and uniform mixture, i.e. *well-stirred*, and treating the time evolution of the system as a stochastic process, specifically a continuous time Markov process with discrete state space, stochastic simulation is straightforward. Let M denote the number of different reactions and $h_j(\mathbf{x}, c_i)$ the hazard for reaction j where $\mathbf{x}$ is the vector of molecule counts of reactants, c_j the rate constants for reaction j and $j = 1, ..., M$. The total hazard, h_0, for a reaction occurring is then given by

$$h_0(\mathbf{x}, \mathbf{c}) \coloneqq \sum_{j=1}^{M} h_j(\mathbf{x}, c_j). \tag{3.7}$$

To generate realisations from a stochastic reaction system for *all* time points up to time T_{max}, the SSA is executed as follows:

Algorithm 1 Stochastic simulation algorithm to simulate exact states of reaction network

Initialise $t = 0$ and states $\mathbf{x} = [x_1, \ ... \ , x_N]$
while $t < T_{\max}$ **do**
  for reaction $j = 1, \ ... \ , M$ **do**
    Calculate hazard $h_j(\mathbf{x}, c_j)$
  end for
  Calculate total hazard $h_0(\mathbf{x}, \mathbf{c}) = \sum_{j=1}^{M} h_j(\mathbf{x}, c_j)$
  Draw time to next reaction $t \coloneqq t + \Delta t$, $\Delta t \sim \text{Exp}(h_0(\mathbf{x}, \mathbf{c}))$
  Draw reaction index $i \in \{1, \ ... \ , M\}$ with probability $P(i = j) = h_j(\mathbf{x}, c_j)/h_0(\mathbf{x}, \mathbf{c})$
  Update $\mathbf{x}$ according to reaction i
  Save $\mathbf{x}$ and t
end while

where $y \sim \text{Exp}(\lambda)$ denotes that y is exponentially distributed with rate parameter

λ.

Using the SSA one may simulate exact realisations, however the inverse problem of obtaining the likelihood of a specific realisation, given a set of parameters (e.g. rate constants) is intractable. To this end, exactness may be traded for tractability by employing approximations, such as diffusion processes describing the population sizes of the chemical species.

3.2 Diffusion approximation through the chemical Langevin equation

A typical approach to modelling transcriptional dynamics in the context of circadian rhythms is through a system of ODEs such as the three component Goodwin oscillator (Goodwin, 1965) or larger, more intricate systems of ODEs consisting of tens of equations describing the molecular oscillator that constitutes the circadian clock (Gonze et al., 2005; Relógio et al., 2011). While the deterministic approach provides a flexible framework to capture interactions between a large number of reactions and molecular species, it is well known that such interactions at the cellular level are intrinsically stochastic due to the low number of molecules (McAdams and Arkin, 1997).

To account for stochasticity, the time evolution of N chemical species $\mathbf{X}(t) = [X_1(t),\, X_2(t), \ldots, X_N(t)]$ in a system of M reactions can be described by the continuous time, discrete state space Markov Jump Process (MJP) resulting from Eq. (3.3). Under that description, conditional on the system being in state $\mathbf{x}$ at time t, the j^{th} reaction has an instantaneous hazard rate $h_j(\mathbf{x})$ and the time to next reaction is exponentially distributed with hazard $h_0(\mathbf{x}) = \sum_{j=1}^{M} h_j(\mathbf{x})$. The number of reactions during a given time interval is well approximated by a Poisson random variable if the interval is short enough such that the change in species do not change the reaction probabilities appreciably. Specifically, the change in the state vector during an infinitesimal time interval dt is given by SdR_t, where S is the matrix of stoichiometries that summarize the consumption and production of species arising from the j^{th} reaction, and dR_t is a M-vector of Poisson$(h_j(\mathbf{x})dt)$ random variables (Wilkinson, 2011, Chapter 6). Stochastic simulation from such a model is straightforward using the Stochastic Simulation Algorithm (Gillespie, 1977), outlined in Section 3.1, but statistical inference is typically infeasible for descriptions with that level of detail, at least in large scale due to computational complexity.

The chemical Langevin equation (CLE) provides a continuous state-space (diffu-

sion) approximation to the dynamics of the MJP. Instead of counting the occurrences of each reaction with a Poisson-process, a Gaussian approximation that matches the mean and variance of the Poisson-process represents the infinitesimal change in the state vector. The approximation is valid under the conditions on the existance of a macroscopic infinitesimal time interval, i.e. the time step dt is small enough that the hazards do not change appreciably and big enough such that one should expect to see a sufficiently large number of each reaction during dt (Wilkinson, 2011). We shall follow the derivation of Gillespie (2000) here.

Suppose that the state of the system at time t is $\boldsymbol{X}(t) = \boldsymbol{x}_t$ and let the random variable $K_j(\boldsymbol{x}_t, \tau)$ be the number of reactions of the j^{th} kind that occur in the time interval $(t,\ t+\tau]$ for some fixed $\tau > 0$. We then have that the number of molecules of the i^{th} species is

$$X_i(t+\tau) = x_{t,i} + \sum_{j=1}^{M} s_{j,i} K_j(\mathbf{x}_t, \tau),\ i = 1, ..., N. \tag{3.8}$$

We shall now make use of two conditions on the time scale τ, firstly we need that τ is small enough such that

$$c_j(\mathbf{X}(t+\epsilon)) \approx c_j(\mathbf{x}_t), \text{ for all } \epsilon \in (0,\ \tau] \text{ and } j = 1, ..., M. \tag{3.9}$$

That is, we want *few* reactions, such that none of the reaction probabilities change appreciably. If this condition holds $K_j(\mathbf{x}_t, \tau),\ j = 1, ..., M$ are independent Poisson random variables with mean and variance $c_j(\mathbf{x}_t)\tau$. The second condition is that we need τ *large* enough such that we expect to see a large number of each reaction type, i.e.

$$\mathbb{E}[K_j(\mathbf{x}_t, \tau)] \gg 1, \text{ for all } j = 1, ..., M. \tag{3.10}$$

The second condition allows for a Gaussian approximation of the Poisson random variables, i.e. $K_j(\mathbf{x}_t, \tau) \dot{\sim} \mathcal{N}(c_j(\mathbf{x}_t)\tau, c_j(\mathbf{x}_t)\tau)$ and hence we can write

$$X_i(t+\tau) \approx x_{t,i} + \sum_{j=1}^{M} s_{j,i} Y_j,\ i = 1, ..., N \tag{3.11}$$

where $Y_j = c_j(\mathbf{x}_t)\tau + \sqrt{c_j(\mathbf{x}_t)\tau} W_j,\ W_j \sim \mathcal{N}(0, 1)$ are independent (due to τ small) Gaussian random variables. Now let $\tau = dt$ be a "macroscopically infinitesimal" time scale that satisfies both conditions. By removing the implicit conditioning on $\mathbf{X}(t) = \mathbf{x}_t$

and subtracting $\mathbf{X}(t)$ from both sides, letting $d\mathbf{X}(t) = \mathbf{X}(t + dt) - \mathbf{X}(t)$ we obtain an equation of the form

$$d\mathbf{X}(t) = \mathbf{S}c(\mathbf{X}(t))dt + \sqrt{\mathbf{S}\,\mathrm{diag}\{c(\mathbf{X}(t))\}\mathbf{S}^{\intercal}}dW(t), \tag{3.12}$$

where $W(t)$ tends to an N-dimensional Wiener process as $dt \to 0$, $c(\cdot)$ is the M-dimensional vector $[c_1(\cdot), ..., c_M(\cdot)]$, $\mathrm{diag}\{c(\mathbf{X}(t))\}$ is a matrix with $c(\mathbf{X}(t))$ along the main diagonal, and $\mathbf{S}$ is the M by N matrix of stoichiometries.

At first glance it seems heroic to assume the existence of a time-scale that contains both a small *and* large number of reactions. However, the first condition is only indirectly an assumption on the number of reactions and may be rephrased as *relatively small* changes in the population sizes of each species, which is clearly helped by large population sizes. The second assumption requires large absolute numbers of reactions and similarly relies on large enough population sizes. Hence, one can expect the CLE to be a good approximation of systems where population sizes are large enough for the two assumptions to be mutually realistic and useful when population sizes are small enough to disobey laws of mass action.

Another limitation of the CLE is that, depending on the functional form of the diffusion coefficient, the variance of the j^{th} component is not necessarily zero for $X_j(t) = 0$, meaning that there is some positive probability for $X_j(t) < 0$. This requires some caution as the CLE is almost always used to model non-negative quantities. In fact, it can be shown that solutions break down in finite time (Schnoerr et al., 2014). Nonetheless, the CLE provides a description that is easier to work with, however, analytical solutions and closed form transition densities are only available for a small class of diffusion processes, meaning inference is again non-trivial.

An approach to obtaining closed form transition densities is provided by the so-called Linear noise approximation (LNA). Wallace (2010) provides a simplified derivation the LNA, originally derived by Kurtz (1971) from the forward equation of the MJP, but with the novel interpretation as an approximation to the CLE. The key feature of the LNA is that the non-linear rate functions are approximated by their first order Taylor expansion about the deterministic limit of the process. This yields Gaussian transition densities as the approximation is linear in the state variable and the diffusion term does not depend on the current state. This was further developed and exploited for Bayesian inference by Komorowski et al. (2009), Finkenstädt et al. (2013), Stathopoulos and Girolami (2013) and Fearnhead et al. (2014). Additionally, experimental data consist of observations at discrete time points, often observed with some measurement error. In

Section 3.4 we shall review how the typically intractable likelihood for processes on the form of Eq. (3.12) may be approximated under these conditions.

3.3 Distributed delay differential equations

Introducing delays provides a parsimonious way to model biochemical processes that depend on a large number of reactions and species. A discrete delay involves letting the rate at time t depend on the state at a single previous time point $t - \tau$, while a distributed delay formulation incorporates an entire past path, either on $(-\infty, t]$ or up to a maximum delay time $[t - \tau_m, t]$. Dependence on the past, not just on a discrete time point but the entire trajectory is intuitively more realistic and this "memory" of the past is typically introduced as a weighted average where the weighting is determined by a density function, hence the name distributed delay. Monk (2003) motivates delay models by the fact that when incorporating negative feedback, they can exhibit behaviours that match the oscillatory expression of genes without requiring additional explicit modelling of intermediate steps such as transcription and protein synthesis. Korenčič 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. However, distributed delays, although deterministic in formulation, arise naturally in systems with intrinsic stochasticity. Smith (2011) describes the "linear chain trick" with which a distributed delay equation arises from a set of ODEs by the following theorem:

Theorem 1 (Linear Chain Trick, Smith (2011))**.** *Let $\frac{dx}{dt} = \dot{x}$ and consider the $N+1$-dimensional system of ODEs given for all $t > 0$ by*

$$\begin{aligned} \dot{x}_0(t) &= f(x_N(t)) \\ \dot{x}_j(t) &= c[x_{j-1}(t) - x_j(t)], \;\; j = 1, ..., N, \end{aligned} \tag{3.13}$$

with possibly nonlinear, continuously differentiable f and initial value function $\phi : (-\infty, 0] \to \mathbb{R}$,

$$\begin{aligned} x_0(t) &= \phi(t) \\ x_j(t) &= \int_t^\infty \phi(-s) g_{j,c}(s) ds, \end{aligned} \tag{3.14}$$

for $t < 0$, where

$$g_{N,c}(u) = \frac{c^N u^{N-1} e^{-cu}}{(N-1)!}, \; u \geq 0 \tag{3.15}$$

is the density of a gamma distributed random variable with shape N and rate c. We have that

$$\dot{x}_0(t) = f\left(\int_0^\infty x_0(t-s)g_{N,c}(s)ds\right) = f\left(\int_{-\infty}^t x_0(s)g_{N,c}(t-s)ds\right). \tag{3.16}$$

Proof (Smith, 2011). Note that the gamma density satisfies $g_{1,c}(0) = c$, and for $j \geq 2$ $g_{j,c}(0) = 0$, along with derivatives $g'_{1,c}(u) = -cg_{1,c}(u)$ and $g'_{j,c}(u) = c(g_{j-1,c}(u) - g_{j,c}(u))$. Now define

$$x_j(t) = \int_{-\infty}^t x_0(s)g_{j,c}(t-s)ds,\ j = 1, 2, ..., N. \tag{3.17}$$

For $x_1(t)$ we have that

$$\dot{x}_1(t) = x_0(t)g_{1,c}(0) + \int_{-\infty}^t x_0(s)\frac{d}{dt}g_{1,c}(t-s)ds = c[x_0(t) - x_1(t)] \tag{3.18}$$

and similarly for $j = 2, 3, ..., N$

$$\begin{aligned}\dot{x}_j(t) &= \int_{-\infty}^t x_0(s)\frac{d}{dt}g_{1,c}(t-s)ds \\ &= \int_{-\infty}^t x_0(s)c[g_{j-1,c}(t-s) - g_{j,c}(t-s)]ds \\ &= c[x_j(t) - x_{j-1}(t)].\end{aligned} \tag{3.19}$$

□

From a modelling perspective, Theorem 1 allows for examining the dynamics of x_0 without explicit modelling of the intermediate steps in Eq. (3.13), i.e. $j = 1, ..., N$. The final product x_0 is instead considered to be available after a random delay with mean c/N and variance c/N^2. In the context of biochemical processes with low molecule numbers, this interpretation contains a link to the stochastic nature of molecular interactions as the time at which a molecule has progressed through the intermediate steps is random. That is, we may transform the system specified by a single non-linear reaction and N first order reactions,

$$\begin{aligned} X_0 &\xrightarrow{f(X_N)} X_1 \\ X_1 &\xrightarrow{c} X_2 \\ &\vdots \\ X_{N-1} &\xrightarrow{c} X_N \end{aligned} \tag{3.20}$$

to the lower dimensional system

$$\emptyset \xrightarrow{f(\int_0^\infty X_0(t-s)g_{N,c}(s)ds)} X_0. \tag{3.21}$$

The new, delayed reaction assumes mass action laws for the intermediate species that enter the distributed delay, i.e. that their dynamics are well-approximated by their deterministic macroscopic rate equations. The key assumption is that the observable species of interest, X_0, is responsible for the molecular noise and its dynamics are accurately modelled through dependence on the deterministic mean trajectory of the intermediate species. If in addition the usual assumption of the existence of a macroscopically infinitesimal time scale holds, the reduced CRN in Eq. (3.21) may be approximated using the CLE.

3.4 Likelihood approximation for SDEs

In this section we shall review a likelihood approximation for CLE's that are non-linear and non-Markovian due to the incorporation of a distributed delay. Additionally, the process may be observed indirectly with Gaussian measurement errors.

3.4.1 Extended Kalman-Bucy filter for CLE's with distributed delay

Bayesian parameter inference requires computing the posterior distribution of parameters given data. For discrete-time Gaussian state space models the likelihood is available using Kalman filtering (Kalman, 1960), and in a continuous time setting using the Kalman-Bucy filter (Kalman and Bucy, 1961). Computing the likelihood of the CLE is challenging as the transition densities are typically not available in closed form. Calderazzo et al. (2018) recently derived a filtering approach to approximate the likelihood for the CLE of models that accommodate non-linear dynamics and distributed delays using the LNA. Consider a CLE with distributed delay, for computational purposes truncated at some maximum delay time τ_{m}, given by

$$\begin{aligned} dX(t) =& f\left(\int_{t-\tau_{\mathrm{m}}}^{t} X(s)g_{p,a}(t-s)ds, X(t)\right)dt \\ &+ \sqrt{A\left(\int_{t-\tau_{\mathrm{m}}}^{t} X(s)g_{p,a}(t-s)ds, X(t)\right)}dW(t). \end{aligned} \tag{3.22}$$

Here $f = S\,h(\cdot)$ and $A = S\,\mathrm{diag}\{h(\cdot)\}S^\intercal$ which reduces the dimensionality of the system from P reactions to N species, as in Eq. (3.12).

Replacing f and A by their first order Taylor expansions about the deterministic mean $\rho(t)$ and assuming that $X(0) \sim \mathcal{N}(\rho(0), P(0))$ the following, approximate Gaussian transition densities are obtained,

$$X(t) \dot{\sim} \mathcal{N}(\rho(t), P(t)), \tag{3.23}$$

where the mean and covariance are given by

$$\begin{aligned} d\rho(t) =& f\left(\int_{t-\tau_m}^{t} \rho(s) g_{p,a}(t-s)ds, \rho(t)\right) dt \\ dP(t) =& J_f\left(\int_{t-\tau_m}^{t} \rho(s) g_{p,a}(t-s)ds, \rho(t)\right) P(t)dt \\ &+ P(t)^\intercal J_f\left(\int_{t-\tau_m}^{t} \rho(s) g_{p,a}(t-s)ds, \rho(t)\right)^\intercal dt \\ &+ A\left(\int_{t-\tau_m}^{t} \rho(s) g_{p,a}(t-s)ds, \rho(t)\right) dt, \end{aligned} \tag{3.24}$$

and J_f is the Jacobian of f (Calderazzo et al., 2018). Approximate solutions to the LNA mean and covariance equations can be obtained numerically using e.g. the Euler method. With a suitable choice of time-step δt and by approximating the integral of the delay with a discrete sum, the numerical solution of the mean equation ρ_t is propagated with

$$\rho_{t+\delta t} = \rho_t + \delta t f\left(\sum_{s=\delta t, 2\delta t \ldots}^{\tau_m} \rho_{t-s} g_{p,a}(s)\delta t\right), \tag{3.25}$$

and similarly the covariance.

Inferring parameters of a model describing the dynamics of a set of chemical species from observations at discrete times, typically measured with noise, requires an additional measurement model. Assume that $X(t)$ is the molecular concentration of a species of interest and is proportional to the light signal generated by a fluorescent reporter protein, perturbed by additive Gaussian noise generated by signal amplification in the measuring equipment. We assume that this measurement process can be expressed as

$$Y_t = \kappa \int_{t-\Delta t}^{t} X(s)ds + \epsilon_t, \;\; \epsilon_t \sim \mathcal{N}(0, \Sigma_\epsilon), \tag{3.26}$$

where $\kappa > 0$ is a scaling parameter between light intensity and molecular concentration

and Δt is the time between observations, as well as the time during which a camera is recording photons from the reporter to make the measurement Y_t.

We shall now make the assumption that the continuous-time process can be suitably approximated by some discretization at small intervals such that

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

The discrete-time approximation is crucial for computational purposes in the empirical applications presented in the following chapters and thus, whenever we refer to the stochastic distributed delay differential equation a suitable discrete-time approximation is invoked. In the final section of the present chapter some aspects of the choice of discretization time-step are investigated. We typically have that $\delta t < \Delta t$, i.e. the frequency at which observations are received is too coarse to directly approximate the transition density as being normally distributed on that time scale. While Δt is often predetermined by experimental procedures the choice of δt is flexible in the sense that we only require $\Delta t/\delta t$ to be an integer. The measurement process in Eq. (3.26) can then be approximated by

$$Y_t = \kappa \tilde{F} X_{(t-\Delta t+\delta t):t} + \epsilon_t,\ \ \epsilon_t \sim \mathcal{N}(0, \Sigma_\epsilon), \tag{3.27}$$

where $\tilde{F}$ is a $\Delta t/\delta t$ by N matrix averaging the unobserved states over Δt. For ease of notation, let κ be absorbed into $\tilde{F}$ and define $\kappa\tilde{F} = F$ while still treating $\tilde{F}$ as fixed and κ as part of the parameter set.

Propagating Eq. (3.24) gives an estimate of the mean and variance of the conditional distribution of the unobserved states given observations and parameters θ until time t, $\pi(X_{(t+\Delta t-\tau_m):(t+\Delta t)}|Y_{0:t}, \theta)$, which is subsequently updated by conditioning on $Y_{t+\Delta t}$ using the Kalman update

$$\begin{aligned} \rho^*_{(t+\Delta t-\tau_m):(t+\Delta t)} &= \rho_{(t+\Delta t-\tau_m):(t+\Delta t)} + C(y_{t+\Delta t} - F\rho_{t+\Delta t}) \\ P^*_{(t+\Delta t-\tau_m):(t+\Delta t)} &= P_{(t+\Delta t-\tau_m):(t+\Delta t)} - CFP_{t+\Delta t,(t+\Delta t-\tau_m):(t+\Delta t)} \end{aligned} \tag{3.28}$$

where

$$C = P_{(t+\Delta t-\tau_m):(t+\Delta t),t+\Delta t} F^\intercal (FP_{t+\Delta t}F^\intercal + \Sigma_\epsilon)^{-1}. \tag{3.29}$$

Note that an initial estimate is required of $\rho_{\Delta t:\tau_m}$ and $P_{\Delta t:\tau_m}$, which can either be obtained using some initial data model that does not contain a delay or by starting to propagate Eq. (3.24) at $t = \tau_m + \Delta t$ and thus using the observations during $t = [0, \tau_m]$

as initialisation.

For parameter estimation we are interested in evaluating the marginal likelihood of parameters given observations. Let θ denote the set of parameters of interest. Then $\mathcal{L}(\theta; Y_{1:T}) = \pi(Y_{1:T}|\theta) = \prod_{t=1}^{T} \pi(Y_t|Y_{1:t-1}, \theta)$, known as the *prediction error decomposition* (first given by Schweppe (1965), see Durbin and Koopman (2012, Chapter 7) for a modern treatment). We have Gaussian errors and $\mathbb{E}(Y_t|Y_{1:t-1}, \theta) = F\rho_{(t-\Delta t+\delta t):t}$, and $\mathrm{Var}(Y_t|Y_{1:(t-1)}, \theta) = FP_{(t-\Delta t+\delta t):t}F^\intercal + \Sigma_\epsilon$. Letting $e_t = Y_t - F\rho_{(t-\Delta t+\delta t):t}$ denote the prediction error we can write the marginal log-likelihood, up to an additive constant, as

$$\log \mathcal{L}(\theta; Y_{1:T}) = -\frac{1}{2}\sum_{t=1}^{T}[\log |FP_{(t-\Delta t+\delta t):t}F^\intercal + \Sigma_\epsilon| + e_t^\intercal (FP_{(t-\Delta t+\delta t):t}F^\intercal + \Sigma_\epsilon)^{-1} e_t] + c, \tag{3.30}$$

where the required terms are calculated at each iteration of the filtering procedure, and the resulting sum is interpreted as a function of the parameters given the fixed observations.

3.5 Bayesian Inference

In this section we present a short review of the general ideas underlying Bayesian statistics, which will be the main mode of inference throughout the rest of the thesis. Lesaffre and Lawson (2012) provide a comprehensive introductory text on the subject from a biostatistics perspective and for the initiated statistician Gelman et al. (2013) is highly recommended and followed in this exposition.

Bayesian statistics provides a principled approach to incorporate expert knowledge regarding parts of the analysis through the use of prior distributions. Inference begins by specifying a full probability model for the problem being studied. The word full refers here to the fact that to proceed, a joint probability distribution is required, not just for data, $\mathbf{Y}$, but for the parameters, $\boldsymbol{\theta}$, one wishes to infer. The probability model for the unknown parameters are referred to as prior distributions, which we shall denote $\pi(\boldsymbol{\theta})$, and the goal is to compute a *posterior distribution* $\pi(\boldsymbol{\theta}|\mathbf{Y})$ by combining the assumed probability model with observational data. To this end, Bayes' theorem is applied, which states that

$$\pi(\boldsymbol{\theta}|\mathbf{Y}) = \frac{\pi(\mathbf{Y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int_{\boldsymbol{\Theta}} \pi(\mathbf{Y}, \boldsymbol{\theta}) d\boldsymbol{\theta}}. \tag{3.31}$$

The numerator contains the full probability model, factored as the probability of observing data $\mathbf{Y}$ given parameters $\boldsymbol{\theta}$, and the prior distribution on the parameters. Typically

$\pi(\mathbf{Y}|\boldsymbol{\theta})$ is replaced with $\mathcal{L}(\boldsymbol{\theta};\mathbf{Y})$ to emphasize that the likelihood is considered a function of the parameters for some fixed observed data, although the difference is semantic.

The denominator contains the marginal distribution of data, obtained by integrating the joint probability distribution over the support of the parameters. As the parameters are integrated out, the denominator is a fixed normalising constant. The notation here uses boldface to highlight that data *and* parameters are often high dimensional, rendering the integral in Eq. (3.31) non-trivial unless specific care is taken in the choice of probability model. If analytical tractability is required it may be reasonable to use *conjugate* prior distributions which ensure that the posterior distribution is from a known distributional family. For example a binomial likelihood and a beta distribution prior on the probability parameter of the binomial model produces a beta posterior distribution. However, the list of conjugate probability distributions is short compared to the list of useful probability models, hence methods to handle intractable posterior distributions are necessary. Markov chain Monte Carlo is one such approach and is covered in the next section.

3.6 Markov chain Monte Carlo

Markov chain Monte Carlo (MCMC) algorithms provide a method to draw samples from intractable probability distributions. A rapid increase in computational power available in the average workstation coupled with large scale development of the theory behind MCMC algorithms has lead to a massive increase in their practical relevance, often being the tool-of-choice for the Bayesian statistician to explore intractable posterior distributions.

The idea is that it is in many cases easier to construct a Markov chain with a given stationary (target) distribution compared to directly sampling from that target distribution. Theoretical justification for sampling using MCMC was initially developed by Tierney (1994) who provided a Markov chain central limit theorem (CLT) and Chan and Geyer (1994) who relaxed the required conditions. A good review of the field is given by Jones (2004). Sampling using MCMC relies on being able to generate a Markov chain, $X^{(t)}$ with stationary distribution f. We require a law of large numbers, such that for a function h with finite variance,

$$\lim_{t\to\infty} \bar{h}_t = \lim_{t\to\infty} \frac{1}{t}\sum_{i=1}^{t} h(X^{(i)}) = \int_{\mathcal{X}} h(X)f(X)dX = \mathbb{E}_f[h(X)], \tag{3.32}$$

and a CLT such that

$$\sqrt{t}(\bar{h}_t - \mathbb{E}_f[h(X)]) \xrightarrow{d} \mathcal{N}(0, \sigma_h^2), \tag{3.33}$$

where the variance of the limiting distribution is given by

$$\sigma_h^2 = \mathrm{Var}_f(X^{(0)}) + 2\sum_{i=1}^{\infty} \mathrm{Cov}_f[h(X^{(0)}), h(X^{(i)})]. \tag{3.34}$$

The law of large numbers lets us use $\bar{h}_t$ to estimate the expectation $\mathbb{E}_f[h(X)]$ and the variance of the CLT tells us how good such an estimate is expected to be. Estimating the variance σ_h^2 is not straightforward, but in Section 3.6.4 one approach is reviewed in detail.

The result obtained by Chan and Geyer (1994) establishes a CLT for *geometrically ergodic* Markov chains, i.e. those where the total variation distance to the stationary distribution f can be bounded by a geometric series. Formally, let $\mathcal{F}$ denote the collection of events for which the target distribution assigns probabilities and let $f^{(t)}(x, A) = \mathbb{P}(X^{(t)} \in A | X^{(0)} = x)$ denote the distribution of the chain at time t. The chain is geometrically ergodic if there exists a nonnegative real-valued function M and a constant $0 < r < 1$ such that

$$||f^{(t)}(x, \cdot) - f(\cdot)||_{\mathrm{TV}} \coloneqq \sup_{A \in \mathcal{F}} |f^{(t)}(x, A) - f(A)| < M(x) r^t \tag{3.35}$$

for all x (Jones, 2004).

In the following sections we shall cover the Metropolis-Hastings algorithm, along with two techniques that accelerate exploration of the target distribution: adaptive MCMC which under fairly mild conditions allows for varying the proposal variance "on the go", often decreasing the number of iterations required in practice and reducing the amount of tuning required for the algorithm to work adequately. We also cover delayed acceptance MCMC, which achieves a speed-up by carefully choosing when to evaluate an expensive target density and when to approximate it.

3.6.1 Metropolis Hastings

Originally described by Metropolis et al. (1953) and generalized into the well known Metropolis Hastings (MH) algorithm by Hastings (1970) has by now become a rich class of algorithms where a thorough treatment is given by (Robert and Casella, 2004, Chapter 7). For notational simplicity, this section uses X and Y for random variables of interest, not for data. The reader is urged to keep in mind that the algorithms described here

are often used in Bayesian inference to explore a posterior distribution, in which case the the random variable of interest is the parameters, θ, of the probability model. With abuse of notation in mind, let X denote a random variable of interest with realizations x and probability density $f(x)$, and let Y denote *proposals* with density $q(y)$. For the target density f and proposal density q such that the ratio $f(y)/q(y|x)$ is known up to a constant free of x, the goal is to generate an ergodic Markov chain $x^{(t)}$ that has stationary distribution f. A general statement of the algorithm is given in Algorithm (2).

Algorithm 2 Metropolis-Hastings algorithm

Initialise x_1 and $t = 1$
while $t < T_{\max}$ **do**
Generate a proposal $Y_t \sim q(y|x_t)$
Take

$$x_{t+1} = \begin{cases} Y_t, & \text{with probability } \alpha(x_t, Y_t) \\ x_t, & \text{with probability } 1 - \alpha(x_t, Y_t) \end{cases}$$

Set $t := t + 1$
end while

The acceptance probability is given by $\alpha(x_t, Y_t) = \min\{\frac{f(y)}{f(x)}\frac{q(x|y)}{q(y|x)}, 1\}$, which, if the proposal distribution is taken to be symmetric around x_t, reduces to $\alpha(x_t, Y_t) = \min\{\frac{f(y)}{f(x)}, 1\}$. A common symmetric proposal density is the Gaussian, centred on the current state of the chain. I.e. proposals are given by $Y_t \sim \mathcal{N}(x_t, \sigma^2)$. This algorithm is referred to as Random Walk Metropolis (RWM) (Robert and Casella, 2004) and in the next section we shall discuss choice of variance for the RWM proposal kernel.

3.6.2 Adaptive MCMC

While the proposal q is close to arbitrary in theory and the mathematical justification of MH algorithms rely on asymptotics, in practice significant efficiency gains (in terms of mixing and convergence) can be achieved by a careful choice of the variance of the proposal distribution. In the case of RWM, if the variance is too low most proposals will be accepted but exploring the target distribution takes a long time, i.e. the *mixing* is slow. On the other hand, if the variance is too large very few proposals will be accepted and the exploration of the target distribution is again slow. These observations hint at the fact that the asymptotic acceptance rate of the algorithm is a useful measure of

its efficiency, which can be estimated by the proportion of accepted proposals while the algorithm is running (Roberts and Rosenthal, 2001).

Most of the early results are with regards to choice of σ for proposal distributions of the form $\mathcal{N}(x^{(t)}, \sigma^2 I_d)$, where d is the dimension of the target distribution, i.e. RWM algorithms. Under the assumption of a target density of the form $f(x_1, x_2, ..., x_d) = g(x_1)g(x_2)\dots g(x_d)$ Roberts et al. (1997) proved that the optimal acceptance rate as $d \to \infty$ is 0.234, attained by scaling the variance of the target by 2.38^2. These tangible numbers have since proved to be robust to various relaxations: the class and dimension of the target distribution, where 0.234 is close to optimal in dimensions as low as $d = 5$ and various families of distributions; non-spherical target densities where the optimal proposal distribution is given by $\mathcal{N}(x^{(t)}, 2.38^2\Sigma)$, i.e. matching the target density covariance and applying the same scaling constant to achieve an asymptotic acceptance rate of 0.234. However, if the target distribution is one dimensional, e.g. in the case where a Metropolis-within-Gibbs algorithm samples individual components conditionally on the rest of the target distributions components, the optimal acceptance rate is instead closer to 0.44 (Roberts and Rosenthal, 2001).

These theoretical optimality results have given rise to a large body of literature on adaptive MCMC, which is already a broad class of algorithms which share the common feature that the proposal variance is tuned automatically by some empirical criterion of optimality, often the acceptance rate (Haario et al., 2001). It is important to note that letting the proposal distribution depend on the past trajectory of the chain by definition destroys the Markov property, however, various theoretical results on ergodicity of adaptive MCMC have been developed (Haario et al., 2001; Andrieu and Moulines, 2006; Roberts and Rosenthal, 2007) that give conditions around which adaptive algorithms can be constructed in practice. Roberts and Rosenthal (2007) proved that asymptotic convergence holds under the conditions of *Diminishing Adaption* (DA) and *Containment*, and Rosenthal (2011) notes that as the latter holds for almost all adaptive schemes in very general settings it is DA that in practice should be the main concern. The DA condition requires that the total variation distance for successive proposal kernels goes to zero over the entire support, formally that

$$\lim_{t\to\infty} \sup_{x\in\mathcal{X}} ||P_{\Gamma_{t+1}}(x, \cdot) - P_{\Gamma_t}(x, \cdot)||_{\mathrm{TV}} = 0 \quad \text{in probability,} \tag{3.36}$$

where P_{Γ_t} is the transition kernel of the chain with adaptive parameters Γ_t at iteration t and $\mathcal{X}$ is the support of x. This condition is satisfied if the magnitude of adaption done at iteration t goes to zero as $t \to \infty$. In practice this is easily incorporated by letting the

adaption amount depend inversely on t or simply stopping the adaption at some point in the chain (Rosenthal, 2011).

An adaptive RWM algorithm, inspired by Atchadé and Rosenthal (2005) and Roberts and Rosenthal (2009) is given in Algorithm (3).

Algorithm 3 Random walk Metropolis algorithm with adaptive proposal variance

Initialise x_1, γ_1, and $t = 1$
while $t < T_{\max}$ **do**
  Draw $Y_t \sim \mathcal{N}(x_{t-1}, \gamma_t \Sigma)$
  Draw $U_t \sim \text{Unif}(0, 1)$
  if $u_t < \min(f(y_t)/f(x_{t-1}), 1)$ **then**
    Set $x_t = Y_t$
  else
    Set $x_t = x_{t-1}$
  end if
  if $t \mod n = 0$ and $t < S$ **then**
    Calculate proportion p_t of accepted proposals in the last n iterations
    if $p_t > 0.234$ **then**
      Set $\gamma_{t+1} = \exp\{\gamma_t + \min(c, t^{-1})\}$
    else
      Set $\gamma_{t+1} = \exp\{\gamma_t - \min(c, t^{-1})\}$
    end if
  end if
  Set $t := t + 1$
end while

Here the adaption targets the scaling, γ, of the proposal covariance matrix by the exponential of a small constant c, is stopped completely after S iterations and is diminishing before that. Adaption is done after examining the acceptance rate of batches of n proposals.

3.6.3 Delayed acceptance MCMC

In the case where $f(y)$ is expensive to evaluate but an approximation $f_x^*(y)$ is cheap, Christen and Fox (2005) propose a modified algorithm that avoids evaluating $f(y)$ for proposals that are likely to be rejected. By introducing a nested step, proposals y are

first promoted for possible acceptance with probability

$$\alpha_1(x, y) = \min\left(\frac{q(x|y)}{q(y|x)}\frac{f_x^*(y)}{f_x^*(x)}, 1\right) \tag{3.37}$$

and rejected with probability $1 - \alpha_1(x, y)$. For promoted proposals a second acceptance probability is calculated as

$$\alpha_2(x, y) = \min\left(\frac{f(y)}{f(x)}\frac{f_x^*(x)}{f_x^*(y)}, 1\right). \tag{3.38}$$

The resulting algorithm accepts proposals with probability $\alpha_1\alpha_2$ but only calculates the expensive $f(y)$ for proposals that are likely to be accepted. Christen and Fox (2005) note that while introducing a delayed acceptance step *reduces* statistical efficiency compared to a standard Metropolis-Hastings algorithm it can increase the efficiency for a given CPU time. In the ideal case the delayed acceptance gives a speed-up corresponding to the inverse of the acceptance rate of the algorithm. This occurs when the approximation is good such that $\alpha_2 \approx 1$ and the computational cost of evaluating f_x^* is negligible compared to f.

Delayed acceptance algorithms have been used successfully for inference in stochastic kinetic models by Golightly et al. (2015) who use a particle MCMC method with the LNA as a cheap likelihood approximation in a nested step and (Calderazzo et al., 2018) who increase the discretisation coarseness of the CLE in a nested step.

3.6.4 Gelman-Rubin convergence diagnostic & effective sample size

A widely used method to determine if a MCMC sampler has converged to the target distribution is the Gelman-Rubin (GR) diagnostic (Gelman and Rubin, 1992). Let f denote the target distribution with mean μ and variance σ^2. The diagnostic relies on generating multiple independent chains and using two different estimators of the target variance that are strongly consistent, i.e. converge almost surely to σ^2 as $T \to \infty$ where T is the sample size, in this case the length of the chain. In finite samples the estimators are biased, one from above and the other from below, and as the sample size increases GR diagnostic formed by the square root of their ratio converges to unity. We now give a definition of the univariate GR diagnostic. Let m denote the number of chains and let $X_{i,t}$ denote the state of the i^{th} chain at iteration t. Consider the quantities

$$\bar{X}_i = \frac{1}{T}\sum_{t=1}^{T} X_{i,t}, \quad \hat{\mu} = \frac{1}{m}\sum_{i=1}^{m} \bar{X}_i$$

the within-chain mean and overall mean respectively, and similarly

$$s_i^2 = \frac{1}{T-1}\sum_{t=1}^{T}(X_{i,t} - \bar{X}_i)^2, \quad s^2 = \frac{1}{m}\sum_{i=1}^{m} s_i^2$$

the within-chain and overall sample variance. As MCMC samples typically exhibit positive autocorrelation, s_i^2 (and s^2), while asymptotically consistent, underestimate the variance of the target distribution in finite samples. The sample variance of sample means can be estimated from the m chains by

$$\frac{\tilde{S}}{T} = \frac{1}{m-1}\sum_{i=1}^{m}(\bar{X}_i - \hat{\mu})^2$$

which in turn yields an estimator for σ^2 given by

$$\hat{\sigma}^2 = \frac{T-1}{T}s^2 + \frac{\tilde{S}}{T},$$

that is biased from above if the chains have over-dispersed starting conditions relative to the target density.

The GR diagnostic is constructed as

$$\hat{R} = \sqrt{\frac{\hat{\sigma}^2}{s^2}}. \tag{3.39}$$

For observed $\hat{R}$ close to 1, as a rule of thumb $\hat{R} < 1.1$, the sampler is terminated and the output used for scientific investigation.

With the rise of high-dimensional inferential problems, e.g. Bayesian hierarchical modelling of big spatio-temporal data sets, the computational cost of running multiple chains is increasingly prohibitive as the time spent in the transient phase typically grows with the dimension of the target density. This problem is exacerbated by the fact that the GR diagnostic requires over-dispersed starting conditions which means that to calculate it requires intentionally initialising the sampler in regions of low probability. As the iterations the chains spend in the transient phase have to be treated as burn-in and discarded, one would rather devote a given computational budget to a single chain such that a higher proportion of the iterations will be spent at stationarity compared to the transient phase.

Vats and Knudson (2018) propose an alternative construction of the GR diagnostic. By replacing the original estimator $\tilde{S}$ by the lugsail batch means (LBM) estimator

of Vats and Flegal (2018) the computational requirements of the original GR diagnostic are alleviated. Let $T = a \cdot b$ where a is the number of batches of size b and again assume m independent chains. Define the k^{th} batch mean for the i^{th} chain as

$$\bar{X}_{i,k} = \frac{1}{b} \sum_{t=(k-1)b+1}^{kb} X_{i,t},\ k = 1, ..., a.$$

For the i^{th} chain the batch means (BM) estimator of the CLT variance, $\tau_T^2 = T \cdot \text{Var}(\bar{X}_i)$ is given by

$$\hat{\tau}_{b,i}^2 = \frac{b}{a-1} \sum_{k=1}^{a} (\bar{X}_{i,k} - \bar{X}_i)^2$$

and the LBM estimator by

$$\hat{\tau}_{i,L}^2 = 2\hat{\tau}_{b,i}^2 - \hat{\tau}_{b/3,i}^2,$$

where $\hat{\tau}_{b/3,i}^2$ is the BM estimator using batch size $b/3$. The LBM estimates for each chain are combined to form an estimator for τ_T^2 by taking their simple averages,

$$\hat{\tau}_L^2 = \frac{1}{m} \sum_{i}^{m} \hat{\tau}_{i,L}^2.$$

A biased-from-above estimator of σ^2 is given by

$$\hat{\sigma}_L^2 = \frac{T-1}{T} s^2 + \frac{\hat{\tau}_L^2}{T}$$

and an alternative GR diagnostic by

$$\hat{R}_L = \sqrt{\frac{\hat{\sigma}_L^2}{s^2}}. \tag{3.40}$$

Vats and Knudson (2018) note that the large sample variance of the LBM estimator $\hat{\tau}_L^2$ is free of m. This means that if a is taken to increase with T we are in principle free to take $m = 1$. In addition, the estimator is biased-from-above in finite samples by construction meaning that over-dispersed initialisation is not required. They further propose a direct multivariate generalization of the alternative GR diagnostic in Eq. (3.40). The multivariate diagnostic for a p-dimensional Markov chain is constructed by

replacing all variance estimators by their covariance counterparts and calculating

$$\hat{R}^{(p)} = \sqrt{\frac{T-1}{T} + \frac{\det(\hat{S}^{-1}\hat{\boldsymbol{\tau}}_L)^{(1/p)}}{T}} \tag{3.41}$$

where $\hat{S}$ is formed by calculating the sample mean of sample covariances of the m chains and $\hat{\boldsymbol{\tau}}_L$ is the LBM estimator of $T \cdot \mathrm{Cov}(\bar{\mathbf{X}}_i)$.

While a converged Markov chain gives samples from the target density, those samples often have non-negligible positive autocorrelation by construction. The effective sample size (ESS) is a widely used measure to determine the sample size of an i.i.d. sample with the same dispersion as the serially correlated Monte Carlo sample. Several definitions and estimators of ESS are available in the literature. Gong and Flegal (2016) consider a univariate construction of ESS given by

$$\mathrm{ESS} = T\frac{\sigma^2}{\tau_\infty^2} \tag{3.42}$$

which can be estimated consistently by replacing σ^2 and τ_∞^2 by their respective consistent estimators. Vats et al. (2019) consider a multi-chain generalization given by

$$\mathrm{ESS}^{(p)} = mT\left(\frac{\det(\Sigma)}{\det(\boldsymbol{\tau}_\infty)}\right)^{1/p}, \tag{3.43}$$

and Vats and Knudson (2018) link the ESS to the lugsail GR diagnostic by considering the lugsail ESS estimator, given by

$$\widehat{\mathrm{ESS}}_L^{(p)} = mT\left(\frac{\det(\hat{S})}{\det(\hat{\boldsymbol{\tau}}_L)}\right)^{1/p}. \tag{3.44}$$

If we consider the case $m = 1$ we can write

$$\hat{R}_L \approx \sqrt{\frac{\widehat{\mathrm{ESS}}_L + 1}{\widehat{\mathrm{ESS}}_L}}, \tag{3.45}$$

which is a very good approximation for any resonable chain length T.

It is worth noting that the estimator of ESS in Eq. (3.44), and similarly the alternative GR diagnostic in Eq. (3.45) depends heavily on the choice of batch size. The estimate can vary dramatically even within the span of commonly used $\lfloor T^{1/3} \rfloor$ to $\lfloor T^{1/2} \rfloor$. A conservative choice is to calculate $\hat{R}_L$ or $\widehat{\mathrm{ESS}}_L^{(p)}$ for several batch sizes and report the

worst case scenario ESS.

3.7 Stability of dynamical systems

Stability analysis is concerned with the behaviour of a system of ODEs *in equilibrium* when subjected to some small perturbation. A steady-state solution x^*, or a fixed point is a solution to the system $\dot{x}_t = f(x_t)$ such that $f(x^*) = 0$. Strogatz (2018) gives a treatment of the principle of linearised stability analysis about a fixed point that we follow here. Let $x_t = x^* + \epsilon_t$ be a solution to the nonlinear ODE

$$\dot{x}_t = f(x_t). \tag{3.46}$$

We can examine the behaviour of the perturbation $\epsilon(t)$ by noting that because x^* is a constant $\dot{\epsilon} = \dot{x}$. Taylor expanding f about x^* gives

$$f(x^* + \epsilon) = f(x^*) + \epsilon f'(x^*) + O(\epsilon^2). \tag{3.47}$$

Because x^* is a fixed point $f(x^*) = 0$ by definition. Now if ϵ_t is sufficiently small we have

$$\dot{\epsilon}_t \approx \epsilon_t f'(x^*) \tag{3.48}$$

which is a linear ODE for the (small) perturbation ϵ_t to the ODE in Eq. (3.46) close to its equilibrium. There are three cases for the dynamics of ϵ_t. If $f'(x^*) < 0$ the perturbations decay, if $f'(x^*) > 0$ the perturbations grow, or if $f'(x^*) = 0$ the nonlinear terms dominate and linearised stability analysis cannot provide any classification of the dynamics.

For a system of ODEs the linearized stability analysis generalises to examining the eigenvalues λ of the Jacobian of the system, evaluated at the equilibrium solution of interest. Higher dimensional systems may exhibit chaotic behaviour, however, by the Poincaré-Bendixson theorem, continuous systems in $\mathbb{R}^2$ do not (see e.g. Teschl (2012)). For these systems, complex eigenvalues correspond to oscillations with period given by $2\pi/\text{Imag}(\lambda)$ and the real part of the eigenvalues determine stability of the oscillations. Smith (2011) gives a thorough review of stability results for systems of ODEs, particularly the case $\max \text{Re}(\lambda) < 0$, where λ are the eigenvalues of the Jacobian, evaluated at the fixed point x^*, then x^* is *locally asymptotically stable*, meaning that near x^* there is a $b > 0$ such that $|\phi - x^*| < b$ implies that $\lim_{t\to\infty} x_t(\phi) = x^*$. That is, if the initial value ϕ is sufficiently close to the equilibrium any solution with initial value ϕ will

converge to the equilibrium (the omega limit point is the equilibrium solutions). Global asymptotic stability is a stronger result in which case the omega limit set consists of the (unique) equilibrium and convergence is guaranteed for any initial value function, not just those sufficiently close to equilibrium. In some cases, e.g. linear time-invariant systems, the rate of convergence can be bounded by $|x_t(\phi) - x^*| \leq K|\phi - x^*|e^{\max \mathrm{Re}(\lambda)t/2}$, $t \geq 0$, which implies *exponential stability*. The bounded convergence rate allows for interpreting $|\max \mathrm{Re}(\lambda)|^{-1}$ as the *characteristic timescale* (CTS), i.e. the order of magnitude of the time for perturbations to decay and thus a measure of the sensitivity to noise. If $\max \mathrm{Re}(\lambda) > 0$, the equilibrium x^* is unstable when subjected to a small perturbation, which for complex eigenvalues correspond to solutions consisting of either a set of periodic orbits or exploding oscillations.

3.7.1 Robustness of biological systems

While robustness is a widely studied concept in the context of biological systems, several definitions are available in the literature. Kitano (2007) uses the following definition: "robustness is a property that allows a system to maintain its functions against internal and external perturbations" and proposes a mathematical formalisation of the robustness, R, of system s , with regard to some property a of the system against a set of perturbations P. This robustness is defined as

$$R^s_{a,P} = \int_P \pi_P(p) D^s_a(p) dp \tag{3.49}$$

where $\pi_P(p)$ is the probability distribution of perturbation p and $D^s_a(p)$ is a function that evaluates to what degree perturbation p impairs property a. Depending on the specific system and property, the evaluation function can be boolean, such that $D^s_a(p) = 1$ when property a is present under perturbation p and $D^s_a(p) = 0$ when perturbation p abolishes property a. Alternatively, D can return a ratio given by $f_a(p)/f_a(0)$ where $f_a(0)$ is the system output under unperturbed conditions. In either case, Eq. (3.49) is the expected degree with which the system remains intact with respect to property a when its parameters are subjected to perturbations characterized by π_P. A desirable property of the robustness definition is that it takes values $0 \leq R^s_{a,P} \leq 1$, regardless of system or property. It is thus not specific to π, s or a and in principle allows for comparing the relative robustness of two or more separate systems with different properties and perturbation mechanisms.

A word of caution is required here however. In this context the term *robustness*

should not be interpreted in the statistical sense, i.e. an estimator that is insensitive to outliers or deviations from distributional assumptions (Huber, 2004). We emphasise that the definition provided in Eq. (3.49) instead quantifies some output of a model when the parameters are varied.

3.7.2 A Bayesian measure of biological robustness

Woods et al. (2016) propose a measure of robustness for evaluating the effect of noise on transcriptional oscillators in the context of model design. By combining a MCMC sampler that explores model space with the formalisation of robustness by Kitano (2007), where the perturbation distribution is taken to be a Bayesian prior distribution, the authors compare relative robustness between a large number of models using the Bayesian model evidence.

We propose a variation of Eq. (3.49), still within the framework of Kitano, obtained when considering a single model and the distribution of perturbations is instead taken to be the posterior distribution of parameters given observed data. Our choice of $\pi_P = \pi(\theta|\mathbf{Y})$ is motivated by the fact that Bayesian inference produces a distribution of statistically plausible values of θ and thus incorporates parameter uncertainty in a formal manner while conditioning on empirical data.

To study the robustness of a dynamical system with respect to oscillations we propose an evaluation function I_ω that takes the value 1 when the parameters θ give rise to a limit cycle and 0 when the system exhibits damped oscillations for θ, i.e. when the long-run solutions converge to an equilibrium. Formally,

$$I_\omega(\theta) = \begin{cases} 1 & \omega \text{ is a periodic orbit} \\ 0 & \omega = x^* \end{cases} \tag{3.50}$$

where ω is the omega limit set and x^* the equilibrium solution. With this choice of evaluation function we define the robustness of oscillations by

$$P_{\text{lim. cyc}} = \int_\Theta \pi(\theta|\mathbf{Y}) I_\omega(\theta) d\theta, \tag{3.51}$$

where $\pi(\theta|\mathbf{Y})$ is the posterior distribution of parameters θ given the observed data $\mathbf{Y}$. The definition in Eq. (3.51) coincides with the Bayesian posterior probability of a limit cycle and in Chapters 4 and 6 we shall use this statistic to infer the robustness of oscillations for a distributed delay, negative feedback model of circadian gene expression. Estimates of $P_{\text{lim. cyc}}$ can be obtained straightforwardly by evaluating the indicator func-

tion and storing the results at each iteration of an MCMC algorithm targeting the model parameters. The resulting chain can then be averaged to produce a posterior mean estimate $\hat{P}_{\text{lim. cyc}}$.

The resulting estimates can either be used for model validation, i.e. under the assumption that data are generated by a limit cycle oscillator, we may wish to verify that the model is consistent with that behaviour. The model validation view is similar to the ideas of Woods et al. (2016), the only difference being conditioning on observed data through Eq. (3.51). On the other hand, one may, under the premise that the model is the true data generating process, make inferences regarding the dynamics of the system of study, obtaining e.g. credible intervals for the probabilities for some type of dynamic behaviour.

3.8 Discussion

In this chapter we have reviewed most of the mathematical and statistical tools and theory that are used throughout the remaining chapters. We show how to approximate a detailed chemical reaction network involving several chemical species that follow evolution laws given by the CME by a CLE of lower dimension but with a distributed delay. Furthermore we review methodology developed by Calderazzo et al. (2018) that allows approximating the likelihood of such a distributed delay CLE when observations are discrete and corrupted with additive Gaussian noise.

MCMC methodology, specifically the random walk Metropolis algorithm, adaptive MCMC and the delayed acceptance algorithm of Christen and Fox (2005) were reviewed in Section 3.6. These methodologies shall be used in the following chapters to sample intractable distributions in the context of Bayesian inference. Random walk-type proposals typically give rise to samples with positive autocorrelation and hence underestimate the variance of the target distribution. This fact motivates the review of the Gelman-Rubin diagnostic and effective sample size. As we shall see in the upcoming chapters, inference for spatio-temporal imaging data presents a computationally expensive problem. The lugsail batch means estimator allows for obtaining an estimate of the effective sample size without running multiple MCMC chains which would be prohibitively expensive in our applications.

We end the chapter with a review of stability analysis. The concept of classifying the behaviour of a dynamical system as either stable limit cycle or damped (noise driven) oscillations is connected to a formalisation of biological robustness of Kitano (2007). Using that connection, we propose a novel, empirical measure of robustness of oscillations,

corresponding to the Bayesian posterior probability of limit cycle oscillations.

Chapter 4

A spatially independent model for Cry1-luc data

In this chapter we apply methodology developed in Chapter 3 to imaging data of the circadian gene expression of Cry1 from mouse SCN. A recently proposed model for auto-regulatory dynamics of circadian genes is derived and its validity for Cry1 is motivated from biological and mathematical perspectives. Some limitations of the model are discussed and stability criteria for the corresponding deterministic reaction rate equation are derived.

In an empirical setting, we show how to elicit an informative prior distribution for the measurement noise variance in imaging data when reference "dark" recordings from the camera are not available. Furthermore, an *ad hoc* two-stage prior distribution is proposed for the light-to-molecule scaling parameter which mitigates some difficulties associated with identification of molecular population sizes. Parameter estimation for a large number of locations across the experimental replicate is achieved by combining the MCMC methodology and Extended Kalman-Bucy filter from Chapter 3 with the informative prior distributions.

The results are interpreted using (i) a clustering algorithm that classifies locations based on the posterior means of parameters associated with mRNA dynamics and (ii) the measure of oscillatory robustness proposed in the previous chapter. Both methods are able to distinguish SCN neurons from adjacent tissues but the latter approach provides the added benefit of parameter uncertainty quantification and a clear biological interpretation. The chapter is concluded with a discussion of the limitations of the methodology and findings, along with suggestions for further reasearch.

4.1 Available Cry1-luc data

Cry1 is a key gene in the circadian TTFL in mammalian cells and its transcriptional promotion is driven by CLOCK/BMAL1 binding to E-box motifs. Unlike Per1 and 2 it lacks a calcium responsive element, hence the model employed in this chapter relies on a purely auto-regulatory feedback loop. Available data consist of a video recording of organotypic cross sections of mouse SCN, where Brancaccio et al. (2013) use a genetically encoded luciferase construct (Cry1-luc) to image Cry1 expression. Tissue samples of mouse SCN were kept in medium for 6 days while luminescence was recorded using an EM-CCD camera with an exposure time of 0.5 hours (Brancaccio et al., 2013). Resulting data consist of of 288 frames each around 350 by 200 pixels in size. The resolution is such that a neuron is approximately the size of an 8 by 8 block of pixels and in the analysis we consider super-pixels where 4 by 4 blocks are aggregated by taking their within-block arithmetic mean for each frame. The purpose of aggregating pixels is to achieve a level of stochasticity that approximately corresponds to that of single cell measurements, such that the implied molecular population sizes are realistic.

The experimental procedure uses a luciferin substrate that is gradually consumed throughout the experiment. This results in an approximately linear trend which is removed by estimating the least squares linear trend for each super-pixel. Furthermore, at some locations early peaks of luminosity are overexposed, resulting in a truncation of data. This phenomenon is typically only seen in the first cycle and the effect on the likelihood and subsequently the parameter estimates are assumed to be mitigated by using the first 24 hours of data to initialise the model.

4.2 A stochastic distributed delay model for autorepressive circadian gene regulation

We now combine the general filtering approach for delayed SDEs with the stochastic transcription model proposed by Calderazzo et al. (2018). Let X_0 denote the molecular count of Cry1 mRNA, transcribed according to rate function $\hat{\nu}(X_p)$ and degrading with rate μ. The argument X_p is the copy number of a protein product that has undergone transport between the cell nucleus and cytoplasm, dimerization with Per protein and is available in the nucleus again to repress transcription of Cry1. Furthermore, let X_i, $i = 1, ..., p-1$ denote copy numbers of different species in the chain between mRNA and transcription inhibitor. We assume that the intermediate reactions can be modelled

with a common reaction rate α, resulting in the following set of reactions,

$$\begin{aligned} \emptyset &\xrightarrow{\hat{\nu}(X_p)} X_0 \\ X_0 &\xrightarrow{\beta X_0} \emptyset \\ X_{i-1} &\xrightarrow{aX_{i-1}} X_{i-1} + X_i \\ X_i &\xrightarrow{aX_i} \emptyset, \end{aligned} \tag{4.1}$$

for $i = 1, ..., p$, where the transcriptional inhibition is modelled using the Hill function

$$\hat{\nu}(X_p) = \frac{R_0}{1 + \left(\frac{X_p}{K}\right)^n}, \tag{4.2}$$

where R_0 is the maximum transcription rate in molecules per hour. The dissociation constant K represents the number of molecules at which 50 percent inhibition is achieved, and the Hill coefficient, n, represents the number of binding sites for the inhibitory complex under strict assumptions, and more generally the degree of inhibitory cooperativity. The Hill-function is widely used in modelling circadian oscillators (see e.g. Gonze and Abou-Jaoudé (2013)).

The birth/death process above involving multiple intermediate species can be represented as a univariate process using a distributed delay over the unobserved intermediate species through the linear chain trick given in (Section 3.3, Theorem 1). This provides a reduction of the number of reactions to two, namely

$$\begin{aligned} \emptyset &\xrightarrow{\nu(X)} X \\ X &\xrightarrow{\beta X} \emptyset \end{aligned} \tag{4.3}$$

where $X = X_0$,

$$\nu(X) = \frac{R_0}{1 + \left(\frac{\int_{-\infty}^{t} X(s) g_{p,a}(t-s) ds}{K}\right)^n}, \tag{4.4}$$

and $g_{p,a}$ is a gamma density with shape parameter p and rate a.

While the biochemical process of interest involves a multitude of reactions and species, the distributed delay formulation captures the auto-repressive behaviour of Cry1 on its own transcription without explicitly modelling the intermediate reactions such as protein synthesis and transport. The approach is useful for empirical purposes as usually measurements can be made of only a single species at any given time.

The lack of CREs in the promoter region of Cry1 underpin the simplification that transcriptional dynamics are driven by a purely auto-regulatory loop of delayed Cry1 mRNA. Additionally, transcriptional induction by CLOCK/BMAL1 is assumed approximately constant throughout the ciradian cycle, and is modelled by the maximum transcription rate R_0.

The corresponding CLE obtained by a Gaussian approximation that matches the mean and variance of the Poisson-processes counting occurrences of the reactions in Eq. (4.3) is

$$\begin{aligned} dX(t) = &\left[\frac{R_0}{1 + \left(\frac{\int_{-\infty}^{t} X(s) g_{p,\alpha}(t-s) ds}{K} \right)^n} - \beta X(t) \right] dt \\ &+ \sqrt{\frac{R_0}{1 + \left(\frac{\int_{-\infty}^{t} X(s) g_{p,\alpha}(t-s) ds}{K} \right)^n} + \beta X(t)} \, dW(t) \end{aligned} \tag{4.5}$$

where $X(t)$ is approximate mRNA molecule count at time t as it is no longer integer-valued and $W(t)$ is a standard Wiener process.

We shall here review some common criticisms of the CLE that are applicable to the model in Eq. (4.5). As the drift term scales with population size and the dispersion term with square root of population size, exploding solutions can be ruled out with the argument that as population size grows, the monotonically decreasing drift term dominates the dynamics and forces the process towards the deterministic equilibrium, given by the solution to $\nu(x) - \beta x = 0$. In fact, large population sizes are exactly the condition that allows for approximating the dynamics of the reaction network using the deterministic reaction rate equation, which corresponds to the drift in Eq. (4.5). However, a problem arises for population sizes close to zero. Due to the fact that the dispersion term does not go to zero with population size, there is some positive probability of reaching *negative* concentrations, which are of course unrealistic and an undesirable property of the model. What is worse is that for sufficiently negative $X(t)$, solutions may break down altogether due to the problem of evaluating non-integer powers of negative quantities. In fact, there are no general existence or uniqueness results for solutions of CLEs where such breakdowns occur (Leite and Williams, 2019). In the case of Eq. (4.5) the dispersion term vanishes for $\nu(X) = -\beta X(t)$ which at first glance seems to prevent a breakdown, but as the process may reach negative values, $\int_{-\infty}^{t} X(s) g_{p,a}(t-s) ds$ can take negative values such that breakdown occurs.

Several modifications to the CLE to obtain existence and uniqueness of solutions have been proposed. These include extending the domain to the complex numbers (Schnoerr et al., 2014) and reflective boundaries that prevent the process from reaching negative states (Leite and Williams, 2019). We argue however that for our purposes the CLE as stated in Eq. (4.5) is adequate with the argument that it is a *local* approximation to the much more cumbersome reaction network, in the sense that the transition densities of the CLE are close to those of the CME for observed data. I.e. we are using the CLE as a stepping stone to obtain an approximate likelihood for the well-posed set of reactions in Eq. (4.1) and data that are generated under the laws of nature.

4.2.1 Stability analysis of the distributed delay single-cell model

The ideas of stability analysis, covered in Chapter 3, Section 3.7, have been extended to e.g. n^{th}-moment stability and stability in probability for stochastic dynamical systems (see Khasminskii (2011) for a review). The general idea is to classify solutions as $t \to \infty$, and, as discussed in the previous subsection, solutions to Eq. (4.5) may not be available after the first time the process reaches negative values. We thus restrict our attention to the stability of the macroscopic rate equation, in a sense studying the dynamics as $t \to \infty$ *and* the reaction container volume $\Omega \to \infty$, or big enough for negligible influence of intrinsic noise.

The macroscopic rate equation (and LNA mean equation) corresponding to the reactions in Eq. (4.3) is given by

$$\dot{\rho}(t) = f\left(\int_{t-\tau_m}^{t} \rho(s) g_{p,a}(t-s) ds, \rho(t)\right) \tag{4.6}$$

where

$$f\left(\int_{t-\tau_m}^{t} \rho(s) g_{p,a}(t-s) ds, \rho(t)\right) = \frac{R_0}{1 + \left(\frac{\int_{t-\tau_m}^{t} \rho(s) g_{p,a}(t-s) ds}{K}\right)^n} - \beta\rho(t). \tag{4.7}$$

Due to the distributed delay, the differential equation in (4.6) is infinite dimensional. Dependence on an entire portion of the past trajectory complicates the usual method of studying the behaviour of its solutions, i.e. the approach described in Chapter 3, Section 3.7 where the linearised dynamics around a special, equilibrium solution ρ^* are examined and classified. It is easy to verify that the required derivative $f'(\rho^*) = d\dot{\rho}(t)/d\rho^*$ is in fact an infinite dimensional gradient for any continuous solution ρ.

Various methods of studying the stability of distributed delay differential equa-

tions are suggested by Hale and Lunel (2013) and reviewed by Bani-Yaghoub (2017) in the context of biology and medicine. These include the extension of Lyapunov functions to Lyapunov functionals, the method of characteristics using the Routh-Hurwitz criterion, and the method of reduction to ODEs. In our case, reducing the equation to a system of ODEs is straightforward as the distributed delay initially arose from a high (but finite) dimensional system of ODEs through application of the linear chain trick (Section 3.3).

By applying the linear chain trick "in reverse", one may expand the macroscopic rate equation into a system of ODEs, for $0 \leq t < T$

$$\begin{aligned} \dot{\rho}_0(t) &= f_0(\rho_p(t), \rho_0(t)) \\ \dot{\rho}_i(t) &= f_i(\rho_i(t), \rho_{i-1}(t)), \;\; i = 1, ..., p. \end{aligned} \tag{4.8}$$

where

$$f_0(\rho_p(t), \rho_0(t)) = \frac{R_0}{1 + (\rho_p(t)/K)^n} - \beta\rho_0(t), \tag{4.9}$$

and

$$f_i(\rho_i(t), \rho_{i-1}(t)) = a[\rho_{i-1}(t) - \rho_i(t)], \;\; i = 1, ..., p. \tag{4.10}$$

The resulting equations represent a $p+1$-dimensional monotone cyclic feedback system (Mallet-Paret and Smith, 1990), meaning that flows through the system are unidirectional, i.e. for some $\delta_i \in \{-1, 1\}$, $\delta_i \frac{df_i(\rho_i, \rho_{i-1})}{d\rho_{i-1}} > 0$ for all i and $\rho_i \in \mathbb{R}^+$. Furthermore, the sign of $\prod_{i=0}^{p} \delta_i$ characterizes whether the system as a whole exhibits positive or negative feedback. Mallet-Paret and Smith (1990) prove that omega-limit sets of bounded orbits of such systems can be embedded in $\mathbb{R}^2$ and thus admit a Poincaré-Bendixson theorem ensuring global asymptotic stability. If additionally

$$\left(\prod_{i=0}^{p} \delta_i\right) \det(-\mathbf{J}) < 0 \tag{4.11}$$

where $\mathbf{J}$ is the Jacobian of Eq. (4.9)-(4.10), omega-limit sets are either (i) the unique fixed point or (ii) unique periodic orbits.

To show that Eq. (4.11) holds, we first note that in our case it is equivalent to $\det(-\mathbf{J}) > 0$ as for any $p \in \mathbb{N}^+$ the system in Eq. (4.9)-(4.10) exhibits negative feedback, i.e. $\prod_{i=0}^{p} \delta_i = -1$. One may then note that the Jacobian of the system is a $p+1$ by

$p+1$ matrix with a specific structure, given by

$$\mathbf{J}(\rho^*) = \begin{pmatrix} -\beta & 0 & \dots & 0 & f_0'(\rho^*) \\ a & -a & 0 & \dots & 0 \\ 0 & a & -a & 0 & 0 \\ \vdots & 0 & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & a & -a \end{pmatrix}, \tag{4.12}$$

where

$$f_0'(\rho^*) = -\frac{nR_0\left(\frac{\rho^*}{K}\right)^n}{\rho^*\left(1+\left(\frac{\rho^*}{K}\right)^n\right)^2}. \tag{4.13}$$

Let A be the bidiagonal p by p sub-matrix obtained by deleting the first row and column of $-\mathbf{J}(\rho^*)$. The determinant of A is trivially a^p and additionally we have that β, $a > 0$ and $f_0'(\rho^*) < 0$. We thus conclude that

$$\det(-\mathbf{J}(\rho^*)) = \beta\det(A) - (-1^p)f_0'(\rho^*)\det(-A^\intercal) = (\beta - f_0'(\rho^*))a^p > 0 \tag{4.14}$$

holds uniformly in ρ^*, i.e. Eq. (4.11) holds for any fixed point and as $t \to \infty$ solutions are either the unique positive equilibrium or unique periodic orbits. A necessary and sufficient condition for the latter is given by Mallet-Paret and Smith (1990, Theorem 4.1) which in our case corresponds to $\mathbf{J}$ having at least two eigenvalues with strictly positive real part.

The existence of a periodic orbit, or limit cycle, is largely determined by the value of the Hill coefficient, n and delay dispersion $\sigma_\Gamma^2 = p/\alpha^2$. Letting $n_{\min}$ be the minimum n for a periodic orbit and letting the delay distribution converge to a point mass at the mean $\mu_\Gamma = p/a$ we have that $n_{\min} \downarrow 1$ if the degradation rates of all the involved species are the same, i.e. $\beta = a$. In our case the degradation rate of the first component is different from that of the intermediate states and thus $n_{\min} \downarrow c$ where c depends on some unknown function of the other parameters (Tyson, 2002). Generally, a limit cycle may be created by increasing the Hill-coefficient or decreasing the delay dispersion. With this in mind, we conclude that distributed delay systems exhibit weaker oscillations, in the sense of a limit cycle, for a given Hill coefficient. We shall return to properties of these two parameters in Chapter 6, where we explore *entrainability*, i.e. to what degree a given perturbation phase shifts the system away from a limit cycle, and discuss how varying degrees of entrainability may be modelled by a delay distribution and nonlinearity.

4.2.2 Measurement model and likelihood approximation

The model in Eq. (4.5) may be used to describe the evolution of the underlying mRNA concentration but available data are measurements of light intensity emitted by the reporter protein. In addition, a single measurement is the average light intensity during a time interval, Δt, corresponding to the camera exposure time required to record a single frame. For the available Cry1-luc data, Y_t, we have $\Delta t = 0.5$ hours and continue under the assumption that the measurement process is adequately modelled by discretising the underlying states in representative tissues $X(t)$ and parametrising the molecule count-to-light by a multiplicative parameter κ. A Gaussian measurement error model produces the following measurement equation (Calderazzo et al., 2018),

$$Y_t = \kappa F X_{(t-\Delta t+\delta t):t} + \eta_t, \;\; \eta_t \sim \mathcal{N}(0, \sigma_\eta^2), \tag{4.15}$$

where F is a vector that averages X over $t = t-\Delta t+\delta t, ..., t-\delta t, t$ and κ is a parameter that scales the underlying molecular population size to light intensity.

Truncation of the delay distribution of the model in Eq. (4.5) at some $\tau_{\max}$ is required to approximate the likelihood using the EKBF detailed in Section 3.4.1. We argue that $\tau_{\max} = 24$ hours is a suitable assumption, containing the negative feedback to one circadian cycle. This is supported by the fact that previous estimates of the delay distribution parameters on a single time series of Cry1-luc data produced a mean and SD of 9.67 and 3.56, respectively (Calderazzo et al., 2018), placing 99.87 percent of the probability mass below 24 hours.

Using the EKBF methodology in Section 3.4.1 to approximate the likelihood of the model in Eq. (4.5) together with (4.15) additionally requires a choice of time grid coarseness δt. While the process is univariate, the entire past path entering the distributed delay has to be updated at each iteration of the filter. This is equivalent in computational cost to filtering a process with dimension equal to the number of evaluation points of the distributed delay integral, given by $\tau_{\max}/\delta t$. Specifically, evaluating the covariance matrix of the past path is expensive as it involves multiplication of a square matrix with a vector of length $\tau_{\max}/\delta t$. The computational cost of the filter thus scales with $(\tau_{\max}/\delta t)^2$, meaning that for computational purposes we want $\tau_{\max}$ small and δt large.

To investigate the trade-off between accuracy of the likelihood approximation and computation time, data from the full multivariate birth/death process obtained from the reactions in Eq. (4.1) are simulated using the stochastic simulation algorithm (see Section 3.1.1) with biologically plausible parameter values, given in the caption

of Figure 4.1. Counts of the species of interest are stored at the reaction times and transformed into synthetic measurements by multiplying the counts by a pre-fixed value of κ and, furthermore, averaging over a time grid equivalent to the camera exposure time for each frame (0.5h) in available data. These synthetic measurements are subsequently corrupted with additive Gaussian noise with variance approximately deduced from data to represent a somewhat realistic level (see Section 4.3.1). The likelihoods of each of the eight parameters are then evaluated using the EKBF while holding the other parameters fixed at their true simulation values. This procedure is repeated at two time-discretisation levels, $\delta t = 0.5$ and $\delta t = 0.1$ to determine if there is a considerable difference in the bias between the two levels. The results are reported in Figure 4.1.

The findings suggest that the bias is low for both levels of δt as the maximum of the likelihood functions coincide well with the simulation values (given by the black lines) for all parameters except the measurement error SD. The measurement error SD exhibits slightly less bias for the coarser approximation $\delta t = 0.5$, possibly due to the two consecutive approximations (reaction network to CLE, and CLE to discrete-time process) introducing bias in opposite directions. The difference in bias is however small relative to the curvature width and the measurement noise process can in fact be estimated accurately outside of the state-space model (see Section 4.3.1). The shapes of the likelihood functions are similar for the remaining parameters across the two choices of δt, suggesting that the coarser grid of $\delta t = 0.5$ may be adequate in this application and will be used in later chapters for comparison. Here, we resort to using a combination discretisation schemes in a delayed acceptance step in the MCMC algorithm which effectively results in the finer approximation at a reduced computational cost. It is worth noting that the computational cost for the likelihood evaluation with $\delta t = 0.1$ is roughly 25 times that of $\delta t = 0.5$.

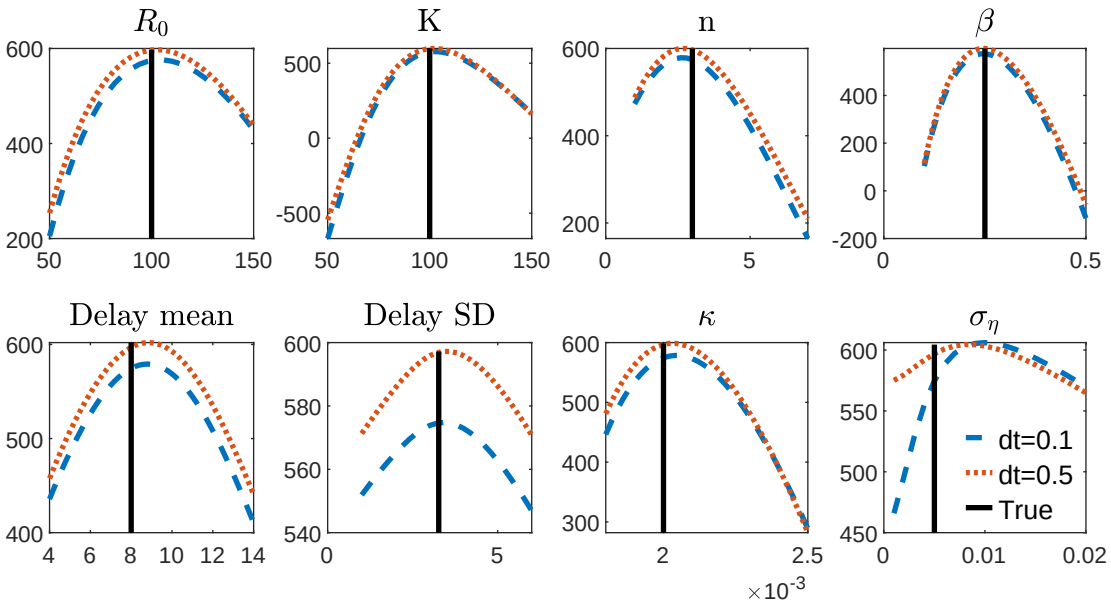

Figure 4.1: Likelihood approximation with varying discretisation coarseness of the unobserved process $X(t)$ and true simulation parameter values (black line). Parameter values are set to $R_0 = 100$, $K = 100$, $\mu_\Gamma = 8$, $\sigma_\Gamma = 3.27$, $n = 3$, $\beta = 0.25$, $\kappa = 2 \times 10^{-3}$, $\sigma_\eta = 5 \times 10^{-3}$.

4.3 Prior distributions

We generally use diffuse prior distributions, specified for the mean and SD of the delay distribution μ_Γ, and σ_Γ, and the logarithms of the maximum transcription rate R_0, dissociation constant K and Hill coefficient n. The prior distribution for the logarithm of the degradation rate is elicited from Yamaguchi et al. (2003) who infer the half-life of the luciferase reporter construct. The same distribution for the degradation rate is used by Calderazzo et al. (2018) where the variance is assumed to account for the reporter process not being explicitly modelled. These prior distributions are summarised in Table 4.1.

Table 4.1: Prior distributions for parameters of the transcriptional feedback loop and degradation.

μ_Γ	σ_Γ	$\log R_0$	$\log K$	$\log n$	$\log \beta$
$\text{Unif}(0, 24)$	$\text{Unif}(0, 20)$	$\mathcal{N}(0, 10)$	$\mathcal{N}(0, 10)$	$\mathcal{N}(0, 10)$	$\mathcal{N}(-0.54, 0.25)$

We additionally suggest two approaches to elicit informative prior distributions for the measurement error dispersion $\pi(\sigma_\eta)$ and light scaling parameter $\pi(\kappa)$. The method used to obtain $\pi(\sigma_\eta)$, described in Section 4.3.1, is principled in the sense that we do not require the data itself, but a reference recording from the experimental procedure. However, in the absence of a reference recording we demonstrate how to substitute it with the data itself, using a transformation that renders the data free of information, other than that of the measurement error distribution.

The method to obtain an informative prior for the light scaling parameter $\pi(\kappa)$ is not principled in the sense of a judgement *a priori* analysing data. Instead we will argue that it can be seen as an approximation to a hierarchical Bayesian model that allows sharing statistical strength between the pixel-wise time series, which are assumed independent within the modelling framework.

4.3.1 Measurement error dispersion

An informative prior distribution for the measurement error variance, σ_η^2, can be elicited from any replicate of the same experimental procedure. The CCD camera used for measuring Cry1 concentration relies on collecting photoelectrons at discrete collection sites called potential wells. The digital image is created when an amplifier transforms the electric charge at each potential well into measurable voltage. The amplifier generates so-called read noise with zero mean, independent of the number of recorded electrons, which at low signal levels dominates other sources of measurement noise such as shot noise that arises from the discrete nature of recorded electrons (Healey and Kondepudy, 1994). The video data is stored as a sequence of multichannel (red-green-blue) images, resulting in two separate measurements of Cry1 concentration. The red and blue channel thus contain the same signal and the differences in the two channels can be assumed to

be dominated by realizations of camera read noise process. Consider

$$\begin{aligned} Y_t^R &= \kappa \int_{t-\Delta t}^{t} X(s)ds + \eta_t^R \\ Y_t^B &= \kappa \int_{t-\Delta t}^{t} X(s)ds + \eta_t^B \end{aligned} \tag{4.16}$$

where X_t is the concentration of Cry1 mRNA at time t, Y_t^R and Y_t^B are the recorded light intensities of the red and blue channel respectively and η_t^R, η_t^B is the measurement noise in the respective channel. Further assume that $V(\eta_t^R) = V(\eta_t^B) = \sigma_\eta^2$ and $\text{Cov}(\eta_t^R, \eta_t^B) = 0$ for all t. Then

$$V[(Y_t^R - Y_t^B)/\sqrt{2}] = \sigma_\eta^2. \tag{4.17}$$

Let $Y_t^D = Y_t^R - Y_t^B$ denote the difference between the two channels. The log SD is estimated at each super-pixel using

$$\ln \sqrt{T^{-1} \sum_{t=1}^{T} (Y_t^D - \bar{Y}_t^D)^2} = \ln \hat{\sigma}_\eta, \tag{4.18}$$

where in fact $T^{-1} \sum_{t=1}^{T} Y_t^D = \bar{Y}_t^D \approx 0$ at all locations, why the scaling constant κ can be assumed to be identical for the two channels. If the assumption of a common variance of the two channels is violated, one may instead opt use the sum of the two channels for analysis and obtain estimates of the measurement error variance in the composite signal using the fact that the variance of the difference is equal to the variance of the sum if the covariance is equal to zero.

The estimated SD is assumed to be spatially uniform, consistent with read noise, and verified by visual examination. Therefore the prior distribution is taken to be identical across the whole biological sample instead of location specific. This gives an empirical distribution of log SDs to which a Gaussian distribution is fitted as $\mathcal{N}(-5.3, 0.17^2)$. The empirical distribution along with a fitted normal distribution is given in Figure 4.2. The fitted distribution is then used as a prior for σ_η^2 in the measurement equation.

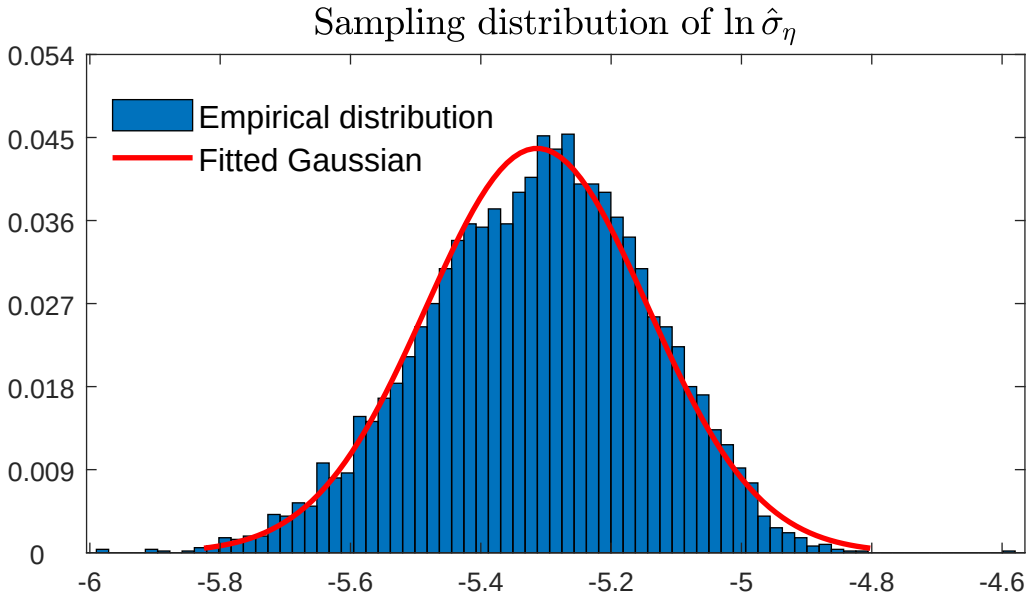

Figure 4.2: Blue histogram: Empirical distribution of the logarithm of the estimated measurement error SD, $\ln \hat{\sigma}_\eta$. Red curve: A Gaussian distribution (mean $= -5.3$, variance $= 0.17^2$) fitted to the empirical to be used as prior distribution.

4.3.2 Light-scaling constant

The parameter κ, which relates measured light intensity to concentration of Cry1 mRNA could be expected to be spatially homogeneous as both Cry1-protein and luciferase are produced in similar proportions throughout the SCN. However, findings from preliminary inference using a weakly informative $\mathcal{N}(0, 10^2)$ prior show considerable spatial structure, in that κ is estimated to be greater in central compared to peripheral locations in the sample, meaning that the center of the sample shines brighter than the edges. There are two presumptive causes: firstly an edge effect caused by light scattering through the tissue, and secondly, the geometry of the sample as the center is raised and thus photons have a longer distance to travel to reach the camera if originating from the edges (M. Brancaccio 2018, personal communication, 31 May).

The spatial structure of κ is not a fundamental issue as the parameter is intended to parametrize varying luminosity from a given concentration of molecules. Since the cause of heterogeneity is most likely due to the experimental procedure (as opposed to biological) we take an approach where "strength is borrowed" between nearby locations

when inferring κ. The results from an initial run on a sparse grid using a weakly informative Gaussian prior distribution on $\log \kappa$ with mean zero and variance 10 are smoothed using a Gaussian kernel with bandwidth selected using leave-one-out cross validation. This produces a surface across all locations of the tissue. The prior distribution on κ for the further inference is then taken to be Gaussian with mean equal to the smoothed surface at that location and variance equal to the squared distance between the initial estimate and the surface. In the case where no initial point is available the nearest (in terms of Euclidean distance) point and surface are used for the variance. Generally, the prior variance is low due to the spatial smoothness of initial estimates. In Figure 4.3 the smoothed surface along with the initial estimates are shown.

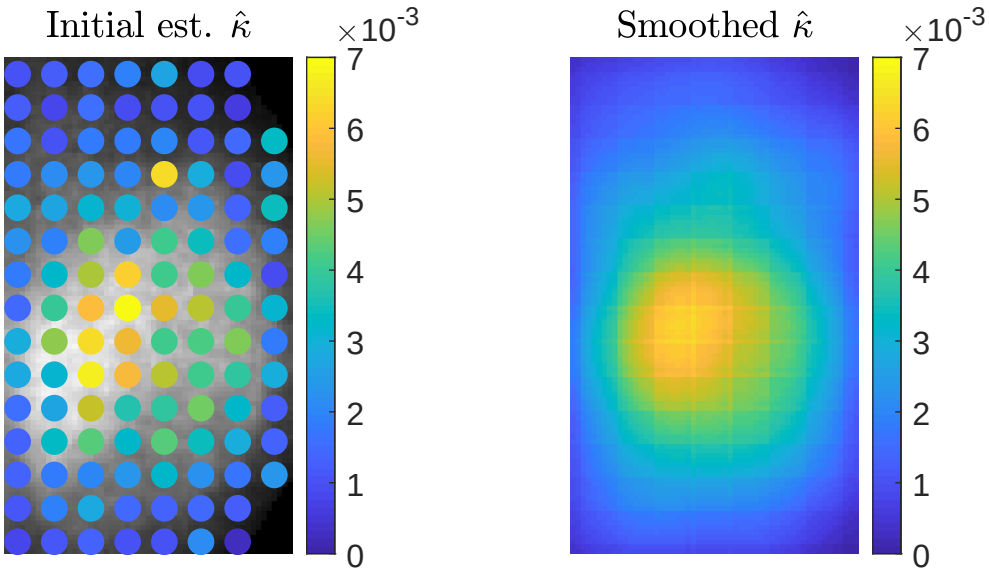

Figure 4.3: **Left:** Initial estimates of κ on a sparse grid across the sample. **Right:** Surface of κ-values using a Gaussian smoothing kernel. Smoothing bandwidth chosen by leave-one-out cross validation and minimising the sum-of-squared distances between initial points and surface.

4.4 MCMC algorithm

The posterior distribution associated with the EKBF likelihood approximation to the model in Eq. (4.15), and prior distributions from the previous section is intractable. We

thus have to resort to simulation-based inference and draw samples from the posterior until we can describe its moments sufficiently well. To this end, we design a random walk Metropolis algorithm based on the techniques outlined in Chapter 3, Section 3.6.

We begin by using a coarse likelihood discretisation with $\delta t = 0.5$ and univariate proposal distributions, updating the parameters one-at-a-time in a fixed scan Gibbs fashion with all parameters except delay distribution mean and SD sampled on the log-scale. The log-transformation makes the posterior exhibit similar scales in different directions which is important for MH-algorithms to work well. Additionally, we avoid proposing negative values, which would always be rejected, for positive real-valued parameters. During the initial 4k iterations of univariate proposals the proposal variance is tuned to achieve an empirical acceptance rate of $\alpha = 0.44$, which is the acceptance rate found to minimize first-order autocorrelation when exploring a one-dimensional Gaussian target (Roberts and Rosenthal, 2001). After the initial 4k iterations we partition the parameters into two blocks, where the first block consists of $b_1 = \{\mu_\Gamma,\, \sigma_\Gamma,\, \log R_0,\, \log K,\, \log n,\, \log \kappa\}$ and the second block consists of $b_2 = \{\log \beta,\, \log \sigma_\eta\}$. The within-block covariance matrices are estimated after discarding the first 1000 iterations and used to form block proposal distributions given by $\mathcal{N}(\boldsymbol{\theta}_j^{(i-1)}, \gamma_{b_j}^{(i)} \hat{\Sigma}_{b_j})$ where $\boldsymbol{\theta}_j^{(i-1)}$ is the state of the MCMC chain of the j^{th} block at iteration $i-1$ and $\gamma_{b_i}^{(i)}$ is tuned to produce an empirical acceptance rate of $\alpha_1^{(\text{opt.})} = 0.234$ (Roberts and Rosenthal, 2001) for b_1 and $\alpha_2^{(\text{opt.})} = 0.33$ for b_2. The value 0.33 is chosen by rough approximation, i.e. in between the univariate and high dimensional optimal rates as the dimension of b_2 is 2. The covariance scaling is tuned by adding/subtracting a small constant $c^{(i)} = c^{(i-1)} - c^{(1)}/10^4$ and $c^{(1)} = 0.02$ to the logarithm of $\gamma_{b_j}^{(i)}$ if the acceptance rate over the last 50 iterations is above/below optimal. Formally, let $\{a_j\}$ denote a chain where $a_j^{(i)} = 1$ if the i^{th} proposal of the j^{th} block is accepted and zero otherwise. The covariance scaling is then implemented by

$$\log \gamma_{b_j}^{(i)} = \begin{cases} \log(\gamma_{b_j}^{(i-1)} + c^{(i)}), \text{ if } \frac{1}{50}\sum_{k=0}^{49} a_j^{(i-k)} > \alpha_j^{(\text{opt.})} \\ \log(\gamma_{b_j}^{(i-1)} - c^{(i)}), \text{ if } \frac{1}{50}\sum_{k=0}^{49} a_j^{(i-k)} < \alpha_j^{(\text{opt.})}. \end{cases} \tag{4.19}$$

To decrease the computational cost a delayed acceptance step (see Section 3.6.3) is introduced after the initial 4k iterations, where the likelihood is first calculated in a nested step using $\delta t = 0.5$ and proposals that are accepted in the nested step are subsequently evaluated using a expensive likelihood evaluation where $\delta t = 0.1$ (see e.g. Calderazzo et al. (2018)). As suggested by the computational complexity of the likelihood scaling with $(\tau_{\max}/\delta t)^2$, the evaluation using $\delta t = 0.5$ is approximately 25 times faster than that with $\delta t = 0.1$. The computational speed up by the delayed acceptance implementation

thus amounts to a speed up close to the inverse of the acceptance rate.

The MCMC algorithm is run for 10k iterations after the initial 4k iterations of coarse likelihood evaluations and univariate proposals, for a total of 14k iteration. The first 5k iterations are discarded as burn-in, as determined by visual examination of the likelihood trace plots. The whole procedure is repeated for a grid of 216 evenly spaced pixels across the experimental replicate, which can easily be implemented in parallel for additional speed-up.

4.5 Results

Before interpreting the resulting parameter estimates various diagnostics of the MCMC algorithm and residual time series are presented. Residuals are tested for normality and periodicity to verify the model fit, while the autocorrelation of the MCMC chains is inspected using effective sample size. Parameter estimates are presented in this chapter in terms of their posterior means. Their spatial distribution is explored by a clustering approach that classifies locations based on the inferred posterior mean of the biological parameters. The resulting clusters are compared to results obtained by calculating equilibrium population sizes of Cry1 mRNA across the biological tissue sample and evaluating the proposed measure of oscillatory robustness from the previous chapter.

4.5.1 Effective sample size

The MCMC chains are evaluated using univariate lugsail batch means ESS, reviewed in Section 3.6.4. The estimates are conservative in the sense that several batch sizes ranging between $N^{1/3}$ and $N^{1/2}$, where N is the nominal length of the chains after burn-in is discarded, are calculated and the lowest ESS estimate is reported. The procedure is repeated for the 216 pixels and 8 parameters and results reported in Figure 4.4. The effective sample sizes are typically between 50 and 100, i.e. about 1% of the nominal chain length, apart from the measurement error SD which exhibits a considerably higher ESS, between 150 and 400.

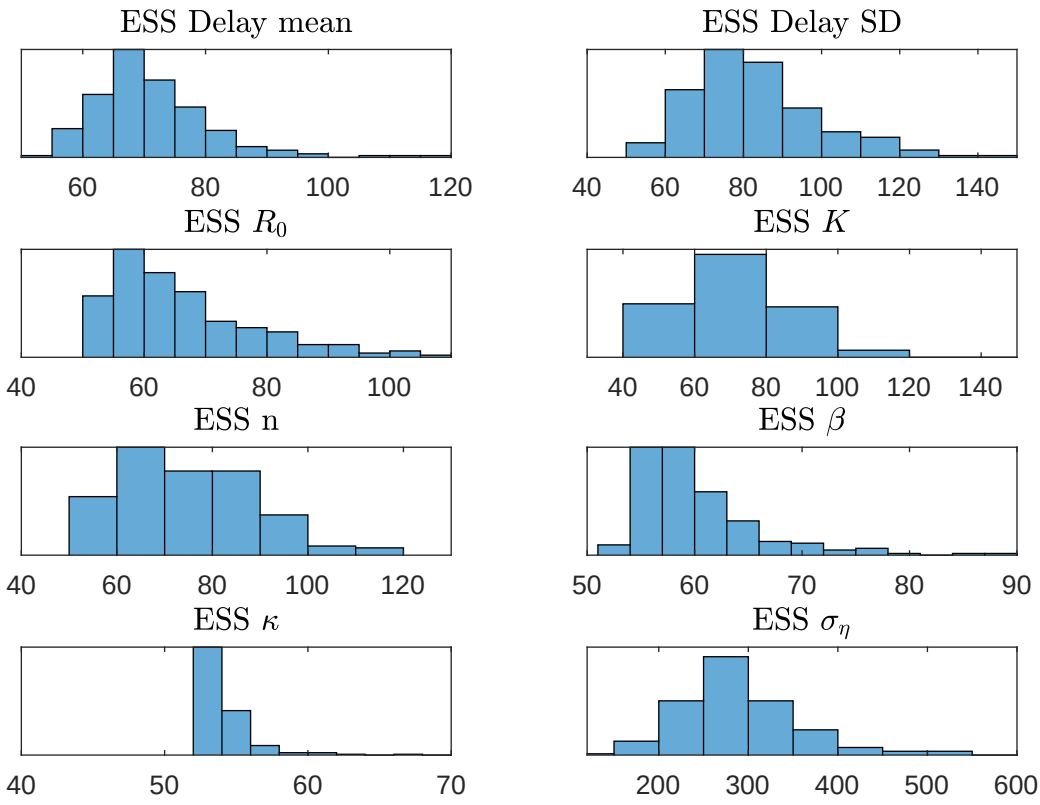

Figure 4.4: Distribution of conservative univariate effective sample sizes, calculated using lugsail batch means estimator with batch size between $N^{1/3}$ and $N^{1/2}$.

4.5.2 Residual diagnostics

To examine the model fit, residual time series are calculated as the difference between observed data and the filter mean, evaluated at the mean of the posterior distribution at each spatial location. While calculating residuals from point estimates of the parameters ignores parameter uncertainty in the model fit evaluation, the primary motivation behind our residual analysis is the examine whether the model can capture the approximately 24 hour oscillations observed in data, and to examine the spatial distribution of inconsistencies between data and model. Hence, the residual time series here may be seen as approximate representative single draws from an implied posterior residual

distribution at each spatial location.

To test whether the residual time series are approximately normally distributed the one-sample Kolmogorov-Smirnov test (Massey Jr, 1951) is employed with the null hypothesis that residuals come from a normal distribution. Letting $\hat{F}$ denote the empirical cumulative distribution function (cdf) of the residual time series and G the cdf of the normal distribution, the test statistic is given by

$$D = \max(|\hat{F}(x) - G(x)|) \tag{4.20}$$

and the implementation in Matlab R2019a (MATLAB, 2019) computes the critical value by interpolation of a table of critical values.

To investigate how well the model captures the oscillatory behaviour of data the residuals are tested for periodicity. Because the first cycle is used as initialisation the residuals are calculated from the second cycle and onwards. The periodogram of the residual time series is calculated at each location and tested for significant spectral component using Fishers G-statistic (Percival and Walden, 1993), given by

$$g = \frac{\max_{1 \leq k \leq m} \hat{S}^{(p)}(f_k)}{\sum_{k=1}^{m} \hat{S}^{(p)}(f_k)}, \tag{4.21}$$

with distribution under the null given by

$$\mathbb{P}(g > g_0) = \sum_{j=1}^{M} (-1)^{j-1} \binom{m}{j} (1 - j g_0)^{m-1}, \tag{4.22}$$

where $\hat{S}^{(p)}(f_k)$ are the periodogram terms, $m = \lfloor T/2 \rfloor$ and M is the largest integer satisfying $M < 1/g_0$ and $M \leq m$.

As the hypothesis tests for normality and periodicity are repeated at 216 locations the 5 percent significance level is adjusted using Bonferroni correction. The results of the hypothesis tests are shown in Figure 4.5, where locations with residual time series that significantly deviate from normality are shown in red and those where normality cannot be rejected in green. A majority of pixels can be assumed to have normally distributed residuals and non-normality is predominantly found along edges of the biological sample corresponding to tissue that is adjacent to the SCN cells. In the right panel of Figure 4.5, pixels that have a significant spectral component in the residuals are displayed along with the period of the component.

The significant residual periodicities correpond to either 12 or 24 hours oscil-

lations, i.e. circadian periodicity or harmonics thereof. While residual circadian (24 hour) periodicity is rare (3.7 percent of locations) and scattered across the tissue sample with some tendency towards the third ventricle, the 12 hour periodicity is frequent (20.4 percent of locations) and concentrated to SCN tissue. Presumptive causes of the 12-hour periodicity are cell-autonomous clock mechanisms that oscillate with 12 hours as found in liver tissue by Zhu et al. (2017). Westermark and Herzel (2013) show that circadian transcription factors binding to separate binding sites in the same promoter region can produce 12 hour rhythms and since our approach only models a single binding site and purely autoregulatory dynamics the residuals are likely to capture rhythms due to additional transcription factors and binding sites. Additionally, low amplitude anti-phasic oscillations of Cry1 in astrocytes (Brancaccio et al., 2017), a type of glial cell that outnumber neurons five to one (Sofroniew and Vinters, 2010), may produce residual periodicity if not explicitly modelled. The reasoning is that additional signalling, while not part of the intracellular TTFL, produces a slight distortion of the waveform of observed Cry1-luc in each pixel, i.e. an additional periodicity, that the model cannot accommodate.

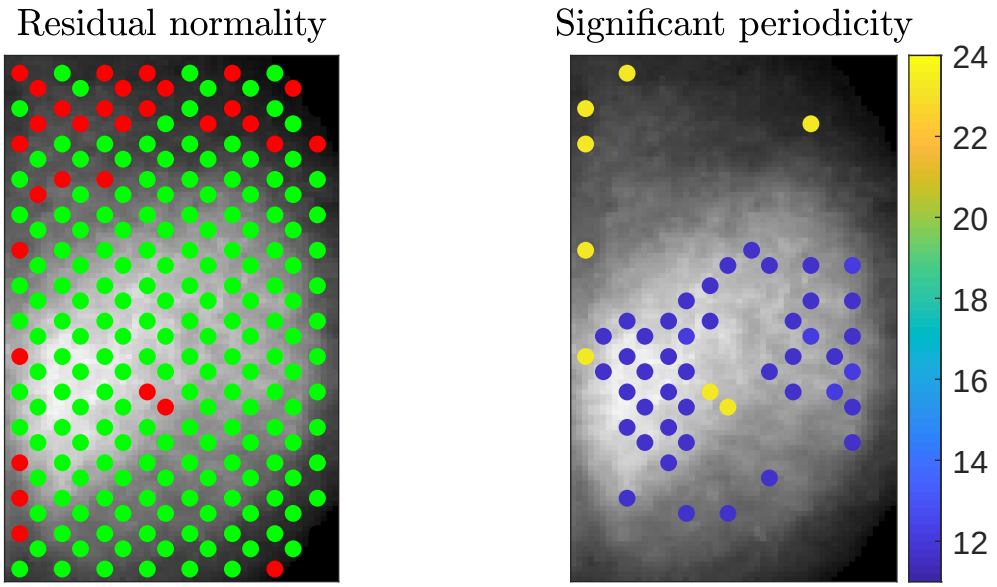

Figure 4.5: **Left:** Test of normality of residual. Green indicates compliance with normality while red indicates rejection of the null hypothesis that residuals are normally distributed ($\alpha = 0.05$, adjusted for multiple testing using Bonferroni-correction). **Right:** Locations with significant residual periodicity, using Fishers g-statistic. The periodicity is indicated by colour. A few locations show significant 24 hour periodicity. Locations with significant 12 hour periodicity ($\alpha = 0.05$, adjusted for multiple testing using Bonferroni-correction) are clustered in the SCN.

4.5.3 Parameter estimates and clustering

Posterior means of the parameters are reported in Figure 4.6. In general, we find relatively low spatial variation across SCN locations in parameter estimates of the delay mean μ_Γ, maximum transcription rate R_0, dissociation coefficient K and measurement error SD σ_η. This finding is expected as the period of oscillations is homogeneous across the biological sample and is mainly driven by the delay mean. Similarly we expect similar population sizes of Cry1 mRNA between neurons as copy numbers are mainly driven by R_0 and K. High values of R_0 (> 140) and K (> 200) are found along the left edge of the tissue sample and these locations are typically associated with a relatively high posterior dispersion for at least one of the two parameters. A presumptive cause is that the spatial smoothing used to obtain an informative prior on κ (Section 4.3.2), which is highly correlated with the two population size parameters, may create an asymmetric "edge

effect" as additional tissue is present outside the cropped frame on the left side (but not right). The measurement error SD is assumed to be spatially homogeneous through an informative prior as specified in Section 4.3.1 and the low spatial variation in posterior means of the parameter is indicative that we are able to separate the measurement noise source from stochasticity in the underlying process.

Systematic spatial structure is found for the delay dispersion σ_Γ, Hill coefficient n, the light scaling parameter κ and, to lesser degree, the degradation rate β. The light scaling parameter closely follows the structure imposed by the informative prior distribution specification and is arguably caused by the experimental process, specifically the light scattering in tissue combined with the geometry of the biological sample. The other three parameters with posterior means that follow a clear spatial distribution are known to be important for the dynamics of the process (Tyson, 2002) and also exhibit clear differences between SCN and surrounding tissue.

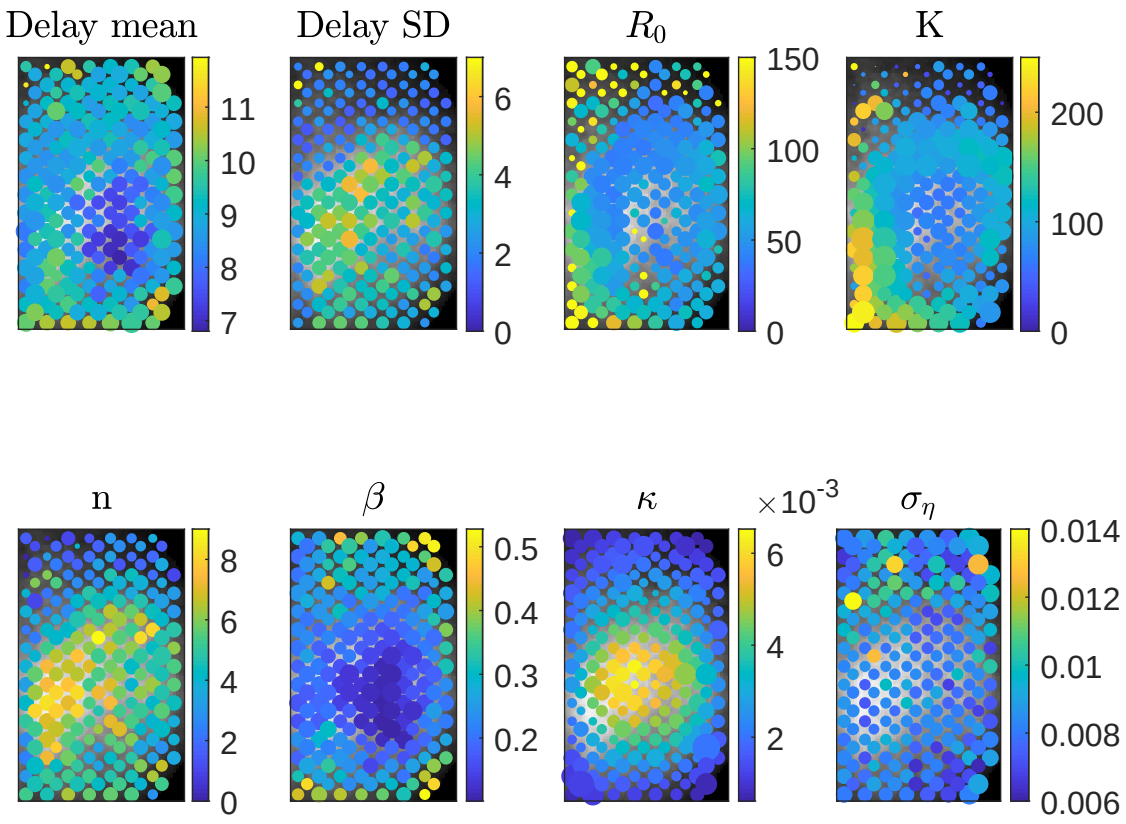

Figure 4.6: Posterior means of the eight parameters. Size of dots are inversely proportional to the coefficient of variation such that a small dot indicates large uncertainty (large posterior dispersion compared to mean).

An exploratory approach to attempt identification of subregions of the biological sample and SCN is cluster analysis, where locations are assigned to a given cluster based on the associated posterior means. In the clustering procedure we disregard parameters that are associated with the measurement equation (κ and σ_η) and focus on those involved in the transcriptional feedback loop. The 216 estimates of the parameter vector are treated as a sample and standardized to have zero mean and unit variance. Standardization is motivated by the fact that the posterior mean with the highest spatial variance otherwise dominates cluster assignment. A k-medoids clustering algorithm is then used to classify locations into either two or three different clusters. The choice of number of clusters is based on the fact that the sample contains SCN neurons and surrounding

tissues, and the SCN neurons are either VIP-ergic or AVP-ergic. The VIP-ergic neurons are predominantly found in the core while AVP neurons are found in the shell of the SCN. We also used k-means clustering and Principal Components Analysis (PCA), as well as kurtosis maximization Independent Components Analysis (ICA) (Moore, 2018) to verify the robustness of the findings. The algorithm used to assign cluster membership is given by Hastie et al. (2009) as follows.

1. For each of the k cluster centres $\mathbf{c}_k$, identify the points that is closer to $\mathbf{c}_k$ than any other centre $\mathbf{c}_j$, $j \neq k$.

2. Compute the mean (median) for each variable for the points in each cluster. Set the cluster centre to the mean (median) vector.

The resulting classification for $k = 2$ is given in the left panel of Figure 4.7 and gives a clear spatial boundary where locations corresponding to SCN tissue are classified to belong to the same cluster while locations from adjacent tissue form another cluster. For $k = 3$ a similar pattern is produced, albeit with more overlap between the three clusters. The classification in this case roughly finds one cluster in the dorsal shell, one in the ventral SCN and a third corresponding to non-SCN tissue. However there is considerable overlap between cluster assignment of ventral SCN and adjacent tissue locations. With $k = 3$ the proportion of location in each cluster is 32.9 percent in the non-SCN cluster, 36.6 percent in the dorsal cluster and 30.5 in the ventral cluster (Figure 4.7, right panel).

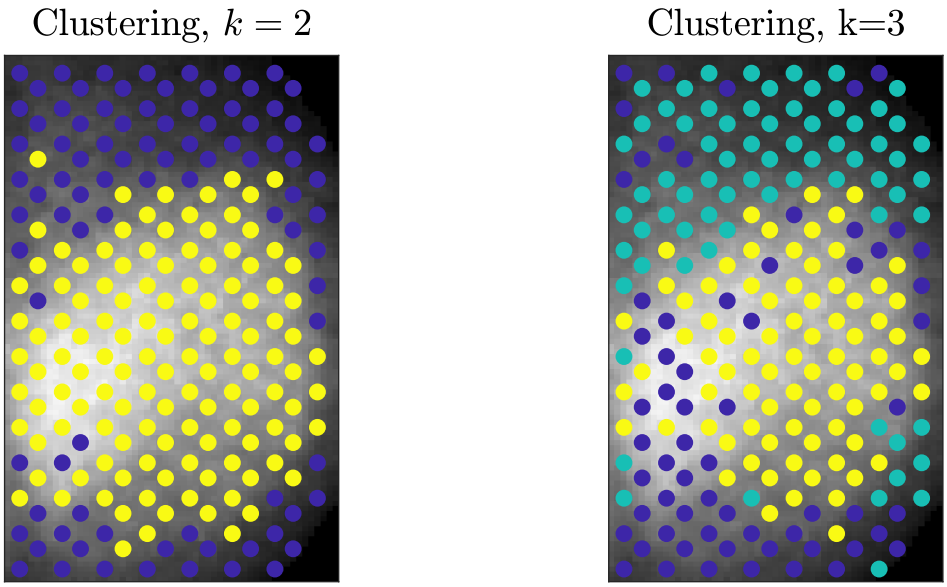

Figure 4.7: **Left:** The output of k-medoids clustering of the parameter estimates with $k = 2$. Locations are classified as either SCN tissue or adjacent tissue with little overlap. **Right:** The output of k-medoids clustering with $k = 3$. The locations are largely classified as ventral (blue), dorsal (yellow), or non-SCN (green). There is substantial overlap between the classification of ventral and non-SCN tissues.

4.5.4 Robustness of oscillations

Using the stability criteria derived in Section 4.2.1 combined with the posterior distribution of parameters obtained from the MCMC chains enables calculating, at each pixel, the robustness of the oscillations, evaluated by the posterior probability of a limit cycle $P_{\text{lim. cyc.}}$ (see Eq. (3.51)). The calculation involves constructing the Jacobian matrix in Eq. (4.12) for each sample of the posterior distribution and calculating its eigenvalues where the real part of the largest eigenvalue is the quantity of interest, along with whether or not the eigenvalues have non-zero imaginary part. In practice this approach can be prohibitively expensive, in which case the eigenvalues of the Jacobians for thinned MCMC chains may be evaluated instead.

A limitation of the stability criteria obtained by transforming the DDE to a system of ODEs using the linear chain trick is that it is only valid for integer-valued shape parameters of the associated Gamma distribution, which does not hold in practice unless specifically enforced by design in the model and inference. We deem such a

design unfavourable for two reasons, firstly due to the difficulties in constructing good MCMC samplers for distributions where some coordinates are integer-valued and others real-valued. Secondly, the closed form distributed delay arises by assuming that the intermediate processes have the same rate parameter and additionally restricting the number of intermediate processes to be integer-valued may well yield a delay distribution that is a worse approximation to the true data generating process when the assumption on the rate parameters is violated.

To evaluate the required Jacobians we construct a rounding scheme of the shape parameter p that preserves the dispersion of the delay distribution. Let μ_Γ and σ_Γ^2 denote the mean and variance of the delay distribution. The shape and rate parameters of the distribution are given by $p = \mu_\Gamma^2/\sigma_\Gamma^2$ and $a = \mu_\Gamma/\sigma_\Gamma^2$, respectively, which are rounded using

$$\begin{aligned} p^* &= \lfloor p \rceil \\ a^* &= \sqrt{a^2 \frac{p^*}{p}} \end{aligned} \tag{4.23}$$

where $\lfloor \cdot \rceil$ denotes rounding to nearest integer. To assess the influence of rounding the calculations are repeated for $p^* = \lfloor p \rfloor$ and $p^* = \lceil p \rceil$ which we found to have negligible effect on $P_{\text{lim. cyc.}}$. Furthermore, we note that it is possible to bound the uncertainty in the final estimates induced by rounding as the propensity for a limit cycle is monotonically increasing/decreasing with decreasing/increasing delay dispersion, other parameters held fixed. Hence, an alternative rounding scheme that preserves the mean of the delay distribution can be used and calculating two different $P_{\text{lim. cyc.}}$ using the floor and ceiling functions to obtain integer p give the widest possible interval due to rounding.

In addition to the posterior probability of limit cycle dynamics, the equilibrium solution for each sample of the posterior distribution is recorded at every pixel and averaged to produce an estimate of the posterior mRNA population size. The two statistics are reported in Figure 4.8, where from the left panel it is evident that $P_{\text{lim. cyc.}}$ accurately distinguishes SCN neurons from surrounding tissues by their stronger oscillatory dynamics. Most of the SCN has a $P_{\text{lim. cyc.}}$ greater than 0.5 and the dorsal SCN typically has a $P_{\text{lim. cyc.}}$ close to 1. Immediately neighbouring non-SCN tissues have values of $P_{\text{lim. cyc.}}$ close to 0 despite data from a majority of such locations exhibiting clear 24 hour oscillations of Cry1:luc. Our approach based on mechanistic modelling of the TTFL successfully distinguishes between SCN neurons that can exhibit such oscillations persistently as opposed to noisy oscillators in the surrounding tissue. Interestingly, a

small cluster of relatively low values of $P_{\text{lim. cyc.}}$ is also found in the ventral and central SCN which is associated with different neuropeptides (VIP) compared to the dorsal organ (AVP).

In the right panel of Figure 4.8 the posterior means of mRNA population sizes are given for the 216 locations. The variation of mRNA copy numbers is found to be low with typical values for SCN tissues in the range $50-150$. Larger estimated copy numbers along the left side of the tissue sample may be due to the smoothing used for κ which results in a restrictive prior distribution and gives values of κ that are too low towards the left edge of the sample, resulting in over-estimation of population size parameters R_0 and K. While empirical results on mRNA copy numbers for circadian genes in the SCN are unavailable in the literature our findings here are consistent with approximate Per mRNA counts from Abel et al. (2015) who forward-simulate a stochastic model of a set of coupled circadian oscillators and find oscillations in the range $0-300$ molecules. Our population size estimates suggest that the approximations we employ in modelling and data transformations produces a level of stochasticity that is typical of single cells and that the analysed pixels are representative of tissue corresponding to single cells.

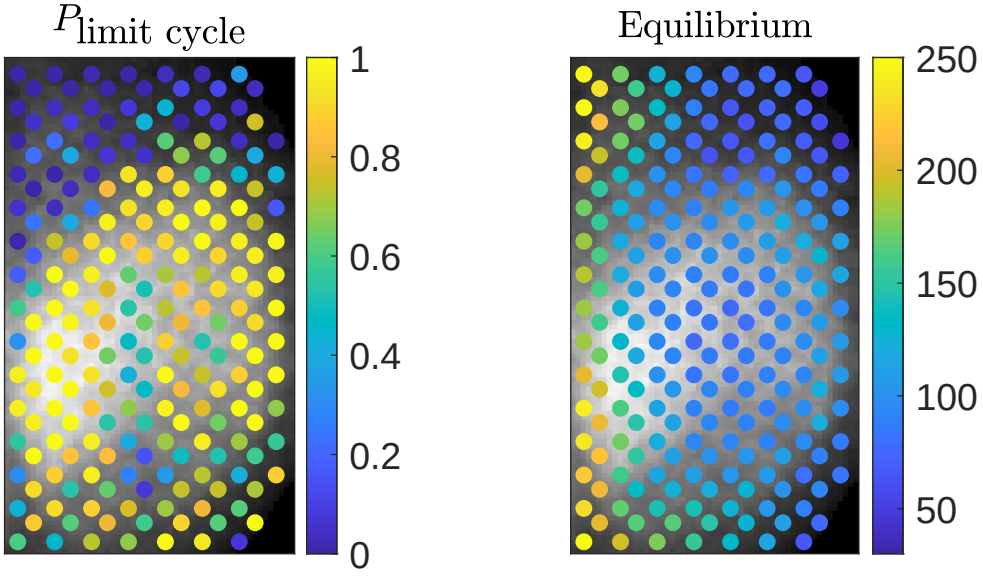

Figure 4.8: **Left:** Posterior probability of limit cycle dynamics (oscillatory robustness). The statistic accurately identifies SCN neurons from surrounding tissues by the propensity of cell-autonomous oscillations. **Right:** Posterior mean of equilibrium solution, corresponding to typical population sizes of mRNA in representative tissues.

4.6 Discussion

In this chapter we combined a recently proposed stochastic model of circadian gene regulation in single cells with the inferential methodology from the previous chapter to model the circadian rhythm of Cry1-luc in spatio-temporal bioimaging data from mouse SCN. Stability criteria of the deterministic rate equation were derived which allow calculation of the proposed measure of oscillatory robustness. A method to obtain informative prior distributions was developed for the measurement error dispersion by a transformation of data that isolates realisations of the noise process such that the variance can be estimated using a maximum likelihood estimator and the variance of the sampling distribution informs the prior dispersion.

For the light scaling parameter, a prior was designed in two steps where statistical strength is shared by spatial smoothing of a set of initial estimates and may be viewed as a crude approximation to a Bayesian hierarchical model. The informative prior on the light scaling parameter substantially improved inference for other parameters related to molecular copy numbers, i.e. R_0 and K which were difficult to identify without an informative prior. By calculating the posterior mean of the equilibrium solution we obtained a measure of approximate mRNA copy numbers which, for SCN neurons, were previously not available in the literature. Population sizes are identifiable only when explicitly modelling the intrinsic stochasticity separately from the measurement process and we find that typical population sizes of Cry1 mRNA molecules are in the range $50 - 250$. These population sizes are consistent with simulation-based results obtained by Abel et al. (2015) who make the assumption that population sizes are similar for SCN neurons and cells in peripheral tissues, an assumption that we do not make here. The population sizes were estimated somewhat higher along the left edge of the biological sample which may be an indication that over-smoothing of κ causes over-estimation of population size parameters. In the next chapter we shall extend the idea by introducing an explicit spatial model in an hierarchical setting which will prevent such issues.

To estimate the parameters of the model we designed a random walk Metropolis algorithm with an adaptive proposal variance and a delayed acceptance step. Residuals were examined for normality and periodicity and autocorrelation of MCMC chains was evaluated using the lugsail batch means estimator of effective sample size. Residual time series from SCN tissue are typically consistent with the normality assumption but exhibit low amplitude 12 hour rhythms in central locations of the SCN. These residual

periodicities may be caused by additional processes that are not included in the TTFL model, such as anti-phasic Cry1 oscillations in astrocytes or non-constant influence of additional transcription factors and binding sites.

Parameters were interpreted in terms of their posterior means and coefficient of variation. The posterior dispersion is homogeneous for all parameters except R_0 and K, which relate the dynamics to mRNA copy numbers. These parameters exhibit higher posterior variance in tissue that is adjacent to the SCN. The variation in dynamics between sub-regions of the SCN is driven mainly by the Hill coefficient n and delay dispersion σ_Γ which are estimated higher in SCN tissue compared to surrounding locations, and higher in the central SCN compared to the SCN shell.

Cluster analysis of the posterior means of parameters governing transcription and degradation rates classified tissue as SCN or adjacent tissue. The addition of a third cluster hinted at a distinction between ventral and dorsal SCN but did not clearly reveal two separate regions of the SCN tissue as hypothesised from the different neuropeptide types documented in dorsal and ventral SCN. The analysis was repeated using clustering of principle and independent components obtained by PCA and ICA with highly similar results.

The posterior probability of a limit cycle, i.e. robustness of oscillations, allowed for a clear distinction between SCN and adjacent tissue as the adjacent tissue exhibits $P_{\text{lim. cyc.}}$ close to 0 for all locations and the SCN tissue values greater than 0. We emphasize that as data are oscillatory, methods such as spectral analysis that are often used to analyse circadian data (see Chapter 2) cannot detect such a distinction. The findings support the hypothesis that SCN neurons exhibit strong self-sustained oscillations. Furthermore, the SCN tissues may be divided into a dorsal shell region with $P_{\text{lim. cyc.}}$ close to 1 and a ventral core region with $P_{\text{lim. cyc.}}$ around 0.5. As data are generated by molecular oscillators in an intact circuit, we cannot differentiate between dynamics being robust due to inter-cellular signalling or if this is a property of the single-cell feedback loop. While our proposed framework allows to answer questions regarding the capability of SCN neurons to generate sustained oscillations, a more precise answer regarding the role of signalling requires further analysis of data from dissociated cells or a model specification that is extended to explicitly capture signalling. Both of these issues are addressed in Chapter 7.

Chapter 5

A Hierarchical model for spatio-temporal bioimaging data with CAR priors

Throughout the earlier chapters the statistical techniques have been focused on inference for single time series. These univariate approaches have been repeated for large numbers of time series that are themselves associated with a spatial coordinate in order to build empirical spatial distributions over biological tissues. Such an approach is motivated by lower model complexity and computational cost as spatial coordinates are a priori disregarded and calculations can be trivially parallellised. In this chapter the univariate approach from the previous chapters is extended to a spatial setting, allowing analyses of bioimaging video data that do not disregard the spatial coordinate associated with each time series when performing inference. To this end, Bayesian hierarchical modelling is employed to combine single cell and spatial modelling in a computationally feasible manner.

The chapter begins with an introduction to Bayesian hierarchical models for spatial data. The conditional autoregressive model is covered in detail and implemented as a spatial prior distribution for a subset of the parameters in the single-cell model of circadian gene regulation from Chapter 4. An MCMC algorithm is designed to sample the posterior distribution of the hierarchical model and some aspects of updating strategies of the spatial components are investigated. The chapter is concluded by a simulation study where synthetic circadian gene expression data is generated from the single-cell TTFL model and the methodology proposed in this chapter used to re-estimate the parameters.

5.1 Bayesian hierarchical modelling for spatial data

Hierarchical models (HM) provide a principled approach to letting multiple sources of uncertainty propagate to the resulting inference. The HM decomposes the full model into nested conditional probability models, or layers. Berliner (1996) expresses a general hierarchical model using the following three layers:

1. Data model: P[data|process, parameters]
2. Process model: P[process|parameters]
3. Parameter model: P[parameters].

Here each stage represents a conditional probability model whose product is a model for the data along with all unknown quantities. If the final stage represents a prior distribution of the unknown quantities one has specified a Bayesian Hierarchical Model (BHM). The final stage can also be a point estimate of the parameters based on observed data. This case is referred to as an Empirical Hierarchical Model (EHM) or empirical-Bayesian model (Cressie and Wikle, 2015, Chapter 3). The key idea is that each layer contributes uncertainty to the resulting inference. The hierarchical structure provides tractability as the product of the conditional models is far too involved to be studied without the decomposition into conditional probabilities. See e.g. Cressie and Wikle (2015) for a treatment in the context of spatial and spatio-temporal modelling.

The statistical methodologies for spatial data, i.e. data where every observation is associated with a spatial coordinate, stem from applications in mining engineering (geostatistics, kriging etc.), agriculture and forestry. The massive growth of the field since the mid 90's has been driven in large part by advances in high speed computing as the models are rarely tractable and hence require computationally intensive simulation based inferential procedures such as MCMC.

Hierarchical modelling is appealing in the context of spatial data as spatial dependencies observed in data can be modelled in either parameter or process models instead of explicitly in the data model. This layer of abstraction is often convenient from a practical point of view, e.g. the mechanism by which spatial dependencies arise is too complicated to be modelled explicitly. Alternatively, the mechanism of spatial dependency can arise at a different layer than the actual observations. This is the case when modelling mRNA concentration using recorded luminosity from a reporter protein. The data are subject to measurement noise that can be assumed to be homogeneous and spatially i.i.d. Meanwhile, non-linear functions of the data are highly spatially correlated

as, for example, nearby cells tend to oscillate with similar period and phase. This spatial dependence can be captured in the process model or the parameter model through location specific random effects.

In a spatial setting random effects are typically specified as non i.i.d. with a dependence structure that obeys the first law of geography: everything is related to everything else, but near things are more related than distant things (Tobler, 1970). Fitting of hierarchical models is typically done using MCMC methods but even with the substantial gains in computing power over the last 30 years, inference for large scale spatial data sets remain infeasible as the time complexity typically grows with the cube of the number of spatial locations. Solutions to this problem come in three categories: approximations of the spatial process on a lower dimensional subspace, spectral domain approaches where the likelihood is approximated in terms of spectral densities, or replacing/approximating the spatial process by a Markov random field (MRF) (see Banerjee et al. (2008) for a review). In the following two chapters, we shall direct our attention to MRFs for solving the issues of computational complexity in spatial modelling.

5.1.1 Conditional autoregressive (CAR) models

Introduced by Besag (1974), the conditional autoregressive (CAR) model is a convenient and computationally tractable specification of an MRF to model spatial correlation using a set of conditional probability models. The exposition here follows that of Carlin et al. (2014) and Cressie and Wikle (2015) and, although not essential for the main ideas, we restrict attention to the Gaussian case. Let ϵ_i, $i = 1, ..., L$ be a finite collection of random variables, each associated with a spatial location on a grid and let negative sub/superscript denote the set of variables excluding the index, i.e. $-i = \{1, ..., i-1, i+1, ..., L\}$. Now assume a neighbourhood structure defined by a proximity matrix W, such that random variables ϵ_i and ϵ_j are *neighbours* if they share a common border, encoded by the $(i, j)^{th}$ element of W: $w_{i,j} = 1$. We shall denote this by $j \in N_i$, i.e. j^{th} location is in the neighbourhood of the i^{th} location. If locations i and j are not neighbours then $w_{i,j} = 0$ and $j \notin N_i$. Next we impose a distribution on each ϵ_i conditional on the other locations of the grid

$$\epsilon_i | \epsilon_{-i} \sim \mathcal{N}\left(\sum_{j \neq i} \tilde{w}_{i,j}\epsilon_j, \tau_i\right), \quad i, j = 1, ..., L, \tag{5.1}$$

such that the conditional mean of the i^{th} component is an average of its neighbours, weighted by $\tilde{w}_{i,j}$ and the conditional variance τ_i is location specific. In practical appli-

cations it is typically assumed that $\tilde{w}_{i,j} \geq 0$ as negative spatial autocorrelation is rarely observed (Griffith and Arbia, 2010). The full set of conditional distributions imply that for $\boldsymbol{\epsilon} = [\epsilon_1, \epsilon_2, ..., \epsilon_L]$

$$\pi(\boldsymbol{\epsilon}) \propto \exp\left(-\frac{1}{2}\boldsymbol{\epsilon}^\intercal D^{-1}(I - \tilde{W})\boldsymbol{\epsilon}\right) \tag{5.2}$$

hinting at a multivariate normal distribution with (sparse) precision matrix $\Sigma^{-1} = D^{-1}(I - \tilde{W})$ where the matrix $\tilde{W}$ is composed of the weights $\tilde{w}_{i,j}$ and $D_{i,i} = \tau_i$ is a diagonal matrix. To ensure symmetry of the precision matrix we need

$$\frac{\tilde{w}_{i,j}}{\tau_i} = \frac{\tilde{w}_{j,i}}{\tau_j}. \tag{5.3}$$

One possible solution to obtain a symmetric matrix is to set $\tilde{w}_{i,j} = w_{i,j}/w_{i,+}$ and $\tau_i = \tau/w_{i,+}$, where $w_{i,+}$ is the column sum of the i^{th} row of W, i.e. the number of members of N_i. This parametrisation, known as the intrinsic CAR model has several appealing features: a single parameter, τ, parametrizes the variances of the L locations in such a way that locations with large neighbourhoods have a lower variance than those with small neighbourhoods. Furthermore, the contribution to location i is shared evenly between the population of its neighbourhood, N_i, such that the weights in Eq. 5.1 are equal and sum to one. However, this parsimony comes with a drawback. Because the rows of $\tilde{W}$ sum to one, the rows of $(I - \tilde{W})$ sum to zero, hence the precision matrix is not invertible. This means that the full set of conditional distributions imply a joint distribution which is not proper and the intrinsic CAR model cannot be used to model (or generate) data. Instead, it is typically used as a prior distribution (Cressie and Wikle, 2015).

Several variations of the CAR model that resolve the issue of impropriety are available in the literature. Lee (2011) gives a review of four such models in the context of Bayesian disease mapping. The most parsimonious variation is referred to as the *Cressie model* (Stern and Cressie, 2000) where a propriety parameter, ρ, controls the degree of spatial correlation. The Cressie model is given by

$$\epsilon_i | \epsilon_{-i} \sim \mathcal{N}\left(\rho \sum_{j \neq i} \tilde{w}_{i,j}\epsilon_j, \tau_i\right), \ i, \ j = 1, ..., L \tag{5.4}$$

which has the intrinsic CAR model as a special case when $\rho = 1$ and implies a proper joint distribution for $\boldsymbol{\epsilon}$ if $|\rho| < 1$, given by

$$\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, (I - \rho\tilde{W})^{-1}D). \tag{5.5}$$

The added flexibility attained by parametrizing the degree of correlation through ρ means that this model can also capture weaker correlation structures than the intrinsic CAR model which imposes $\rho = 1$. It is worth noting that while $\rho = 0$ implies zero spatial correlation, the variance of the i^{th} component is still given by $\tau/w_{i,+}$, a function of the number of neighbours of i, which is incompatible with the notion of spatial independence. We shall for the most part make use of the intrinsic specification as a prior distribution in this chapter and the empirical application in Chapter 6. However, for the simulation study in Section 5.3 we require the ability to simulate data and thus make use of the Cressie model as a data generating process.

5.1.2 Incorporating spatial prior distributions for the single-cell transcription model

To incorporate spatial dependence for the parameters of the transcription function of the Cry1 model in Eq. (4.5), let subscript denote the index $i = 1, 2, ..., L$ of pixels and superscript denote the parameter index $p = 1, ..., P$. We assume a spatial prior on five parameters, namely the mean and SD of the delay distribution μ_Γ and σ_Γ, the maximum transcription rate R_0, the dissociation coefficient K and the Hill coefficient n, thus $P = 5$. For the neighbourhood structure we assume that every pixel is a priori correlated with all its surrounding pixels (8-connected) as opposed to only the pixel with which it shares a border (4-connected) as this gives longer range correlations which appear to more closely resemble the spatial structures seen in the parameter estimates obtained using a spatially independent model in Chapter 4.

The logarithm of each of the spatially dependent parameters $\Theta = [\Theta^{(1)}, ..., \Theta^{(5)}]$ is modelled by random effects model

$$\log \Theta^{(p)} = \log \mu_{\Theta^{(p)}} + \boldsymbol{\epsilon}^{(p)}, \tag{5.6}$$

with a global location parameter $\mu_{\Theta^{(p)}}$ and location-specific random effects $\boldsymbol{\epsilon}^{(p)}$, amounting to a multiplicative random effects model for the untransformed parameters, where

$$\epsilon_i^{(p)} | \epsilon_{-i}^{(p)}, \tau_i^{(p)} \sim \mathcal{N}\left(\sum_{j \in N_i} \tilde{w}_{i,j} \epsilon_j^{(p)}, \tau_i^{(p)} \right), \; p \in \{1, ..., P\}, \; i \in \{1, ..., L\}. \tag{5.7}$$

and $\Theta^{(p)}$ is assumed independent of $\Theta^{(-p)}$. To ensure that the precision matrix of the full set of conditionals is symmetric we take assume $\tilde{w}_{i,j} = w_{i,j}/w_{i,+}$ and $\tau_i^{(p)} = \tau^{(p)}/w_{i,+}$ which further reduces the number of unknown parameters as $\boldsymbol{\tau} = [\tau^{(1)}, ..., \tau^{(5)}]$ along

with the fixed neighbourhood structure parametrizes the variance of all 5 parameters and L locations. From Eq. 5.6-5.7 it is evident that value of $\log \Theta_i^{(p)}$ is modelled as deviating from the common mean $\log \mu_{\Theta^{(p)}}$ with deviations obeying a first order MRF process.

Furthermore, the three remaining parameters, namely the degradation rate μ, light scaling parameter κ and measurement error SD σ_η are a priori spatially i.i.d. Let $\Psi_i = [\mu, \kappa, \sigma_\eta]_i$, modelled by assuming $\pi(\Psi_i) = \pi(\Psi_j),\ j, i = 1, ..., L$, and components of Ψ_i a priori uncorrelated.

The full Bayesian hierarchical model is then specified for data Y, unobserved process X, locations $i = 1, ..., L$, spatially modelled parameters $\Theta^{(p)}$, $p = 1, ..., 5$, and a priori spatially i.i.d. parameters Ψ as

Observation model:

$$Y_{i,t} = \kappa \int_{t-\Delta t}^{t} X_i(s)ds + \eta_{i,t},\ \ \eta_{i,t} \sim \mathcal{N}(0, \sigma_\eta^2)$$

Process model:

$$\begin{aligned} dX_i(t) = &\left[\frac{R_0}{1 + \left(\frac{\int_{-\infty}^{t} X_i(s) g_{p,\alpha}(t-s)ds}{K} \right)^n} - \beta X_i(t) \right] dt \\ &+ \sqrt{\frac{R_0}{1 + \left(\frac{\int_{-\infty}^{t} X_i(s) g_{p,\alpha}(t-s)ds}{K} \right)^n} + \beta X_i(t)}\, dW_i(t) \end{aligned} \tag{5.8}$$

Spatial parameter model:

$$\Theta_i^{(p)} = \mu_{\Theta^{(p)}} \cdot \exp \epsilon_i^{(p)}$$

Prior distributions:

$$\epsilon_i^{(p)} | \epsilon_{-i}^{(p)}, \tau^{(p)} \sim \mathcal{N}\left(\sum_{j \in N_i} \tilde{w}_{i,j} \epsilon_j^{(p)}, w_{i,+}^{-1} \tau^{(p)} \right)$$

$$\log \tau^{(p)} \sim \mathcal{N}(0, \sigma_\tau^2)$$

$$\log \Psi_i \sim \mathcal{N}(0, \Sigma_\Psi)$$

$$\log \mu_{\Theta^{(p)}} \sim \mathcal{N}(0, \sigma_{\Theta^{(p)}}^2)$$

5.2 Inference for hierarchical Bayesian models with CAR priors

To infer the parameters of the full model we resort to an adaptive random walk Metropolis (RWM) as this has proven to work well in practice, both for the gene transcription model in Eq. 4.5 (Calderazzo et al., 2018) as well as for spatial models, e.g. in ecology (Wikle, 2003) and disease mapping (Waller et al., 1997).

Let $\mathbf{Y}$ denote the full data and let $\mathbf{Y}_i = [Y_{i,\Delta t}, ..., Y_{i,T}]$ denote time series data associated with location $i = 1, ..., L$. Conditional independence between data and hyper-variances $\boldsymbol{\tau}$ given Θ gives $\pi(\mathbf{y}|\Theta, \Psi, \boldsymbol{\tau}) = \pi(\mathbf{y}|\Theta, \Psi)$. Furthermore, dependence between $\mathbf{Y}_i$ and $\mathbf{Y}_j$ is captured by the dependence between Θ_i and Θ_j. Hence we have "spatial factoring" of the likelihood $\pi(\mathbf{y}|\Theta, \Psi) = \prod_{i=1}^{L} \pi(\mathbf{y}_i|\Theta_i, \Psi_i)$ where each term in the product can be evaluated in parallel. The prior independence of $\Theta^{(p)}$ and $\Theta^{(q)}$, $p \neq q$ gives "horizontal" factoring of the hierarchical prior $\pi(\Theta|\boldsymbol{\tau}) = \prod_{p=1}^{P} \pi(\Theta^{(p)}|\tau^{(p)})$.

The unknown parameters are the five parameters of the transcription function, each associated with L random effects, the degradation rate, light scaling and measurement error SD at each of the L locations and the hyper-variances, $\tau^{(p)}$, $p = 1, ..., 5$ of the CAR prior. The total number is given by $5(L + 2) + 3L$. Neal et al. (2006) explore blocking strategies to speed up convergence of RWM algorithms and note that there is often computational overhead associated with updates of the full target if it is high dimensional. For the spatial prior specification considered here, partitioning the pixels and updating only a subset of the random effects at a time for a given parameter requires some care as no update block can contain any neighbours. The possible partitions vary between 8-connected and 4-connected grids as an 8-connected grid requires at least four blocks while a 4-connected grid only requires two.

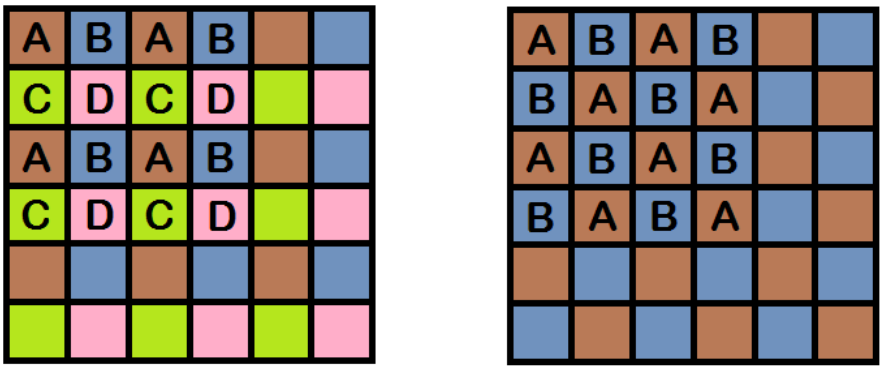

Figure 5.1: The grid on the left shows a configuration where no block (A-D) contains any neighbours. To fill an 8-connected grid with blocks where no neighbours share block label requires four or more groups. The grid on the right is divided in a similar minimal configuration for a 4-connected grid where it can be done with only two labels.

In Figure 5.1 example configurations with the smallest number of subsets are given for the two neighbourhood structures. Note however that it is possible in both cases to update *all* L random effects in a single update, even in the case of specifying the prior distribution with a non-invertible precision matrix as the joint density over all random effects exists but is non-integrable. That is, there is no strict need to devise a blocking scheme in order to evaluate the prior distribution imposed on the random effects as the expression in Eq. 5.2 exists and is sufficient.

To evaluate blocking strategies in sampling the random effects, for each parameter, the L pixels are divided into B blocks that are arranged such that their components are a priori independent, i.e. no neighbours belong to the same block. This is important as it allows for sampling components of a block by a single Metropolis update and evaluation of the associated prior distributions and likelihoods can be performed in parallel. The partitioning is done such that the blocks are balanced, each containing $\frac{L}{B}$ random effects. If we assume that updating an entire block takes unit time, meaning that we have $\frac{L}{B}$ processor cores available (fully parallelizable assumption), we want the blocks to be big as this corresponds to exploring the target distribution in more directions with a single update. However, increasing the block size (decreasing the number of blocks, B) results in smaller jumps in each coordinate of the target density. We thus want to find a suitable trade-off between the speed with which we are exploring the conditional distributions and the number of conditional distributions we are exploring in a single

update.

Let $\theta_i^{(k)}$ denote the state of the Markov chain at location i and iteration k. The expected squared jumped distance (ESJD), defined as

$$\mathbb{E}||\theta^{(k)} - \theta^{(k-1)}||^2 \tag{5.9}$$

is used by Pasarica and Gelman (2010) to adaptively scale the proposal distribution to improve mixing and convergence. With the fully parallelizable assumption, the quantity

$$\widehat{\text{ESJD}} = \frac{1}{B}\sum_{i=1}^{L}\left(\frac{1}{k_{\max}}\sum_{k=1}^{k_{\max}}||\theta_i^{(k)} - \theta_i^{(k-1)}||^2\right) \tag{5.10}$$

uses a sample estimate of Eq. 5.9 to measure the average distance explored in unit time. Taking B small (blocks big) increases the number of directions with which we explore the target density with a single proposal but decreases (the expected value of) the size of jumps. Conversely, taking B big (blocks small) decreases the number of directions but increases the (expected value of) the jump sizes. Theory on optimal scaling of random walk Metropolis proposal distributions suggests that the quantity in Eq. 5.10 remains fairly constant for varying B. The optimal proposal variance, for a spherical d-dimensional Gaussian target, is given by $\sigma_d^2 I_d$ where σ_d^2 is $O(d^{-1})$ (Roberts and Rosenthal, 2001) and hence exactly cancels what is gained by exploring the target in more directions simultaneously.

5.2.1 Random walk Metropolis algorithm for hierarchical spatial random effects model

In this section an MCMC algorithm for sampling the posterior of parameters, random effects and hyper-variances of the hierarchical model is presented. The steps are given in Algorithm (4).

All proposals are of random walk-type, i.e. Gaussian, centred on the state of the chain at iteration k and variances $\gamma_{\cdot,k}$ are adaptively scaled using the scheme in Section 3.6.2. To implement a blocking scheme for the random effects, the step involving proposing random effects can be expanded to several nested steps where in each sub-step a subset of the random effects are updated for non-neighbouring locations, e.g. according to the partitioning given in Figure 5.1.

Algorithm 4 Random walk Metropolis algorithm to sample posterior distribution of Hierarchical TTFL-model with spatial random effects

while $k \leq$ max iterations **do**
  Set $\mu_{\Theta,k} = \mu_{\Theta,k-1}$, $\boldsymbol{\tau}_k = \boldsymbol{\tau}_{k-1}$, $\boldsymbol{\epsilon}_k = \boldsymbol{\epsilon}_{k-1}$ and $\boldsymbol{\Psi}_k = \boldsymbol{\Psi}_{k-1}$
  Propose global location parameters $\hat{\mu}_\Theta \sim \mathcal{N}(\mu_{\Theta,k-1}, \gamma_{\mu,k}\hat{\Sigma}_{\mu_\Theta})$ and set $\mu_{\Theta,k} = \hat{\mu}_\Theta$ w.p. $\min\left(1, \frac{\prod_{i=1}^{L} \pi(\mathbf{Y}_i|\hat{\mu}_\Theta, \epsilon_i^p, \Theta_i^{-p})\pi(\hat{\mu}_\theta^p)}{\prod_{i=1}^{L} \pi(\mathbf{Y}_i|\mu_\Theta, \epsilon_i^p, \Theta_i^{-p})\pi(\mu_\theta^p)}\right)$
  for parameter $p = 1, ..., P$ **do**
    Propose hyper-variance $\hat{\tau}^{(p)} \sim \mathcal{N}(\tau_{k-1}^{(p)}, \gamma_{\tau,k})$ and set $\tau_k^{(p)} = \hat{\tau}^{(p)}$ w.p. $\min\left(1, \frac{\pi(\boldsymbol{\epsilon}^{(p)}|\hat{\tau}^{(p)})\pi(\hat{\tau}^{(p)})}{\pi(\boldsymbol{\epsilon}^{(p)}|\tau^{(p)})\pi(\tau^{(p)})}\right)$
    Propose random effects $\hat{\boldsymbol{\epsilon}}^{(p)} \sim \mathcal{N}(\boldsymbol{\epsilon}_{k-1}^{(p)}, \text{diag}(\gamma_{\boldsymbol{\epsilon},k}))$ and set $\boldsymbol{\epsilon}_k^{(p)} = \hat{\boldsymbol{\epsilon}}^{(p)}$ w.p. $\min\left(1, \frac{\prod_{i=1}^{L} \pi(\mathbf{Y}_i|\mu_\Theta, \hat{\epsilon}_i^p, \Theta_i^{-p})\pi(\hat{\epsilon}_i^p|\epsilon_{N_i}^p, \tau^p)}{\prod_{i=1}^{L} \pi(\mathbf{Y}_i|\mu_\Theta, \epsilon_i^p, \Theta_i^{-p})\pi(\epsilon_i^p|\epsilon_{N_i}^p, \tau^p)}\right)$
  end for
  for location $i = 1, ..., L$ **do**
    Propose parameters $\hat{\Psi}_i \sim \mathcal{N}(\Psi_{i,k-1}, \gamma_{i,\Psi,k}\hat{\Sigma}_\Psi)$ and set $\Psi_{i,k} = \hat{\Psi}_i$ w.p. $\min\left(1, \frac{\pi(\mathbf{Y}_i|\Theta_i, \hat{\Psi}_i)\pi(\hat{\Psi}_i)}{\pi(\mathbf{Y}_i|\Theta_i, \Psi_i)\pi(\Psi_i)}\right)$
  end for
  Set $k \coloneqq k + 1$
end while

5.3 Simulation study

In this section a simulation study is conducted to evaluate the performance of the algorithm proposed to infer parameters and random effects of the hierarchical CAR model outlined in this chapter. Simulation parameter values are set to represent realistic levels based on the results of Chapter 4. As it is not possible to simulate data from the prior distribution specification of the random effects we resort to using the Cressie model as data generating process with propriety parameter set to $\rho = 0.99$ and an 8-connected neighbourhood configuration. The parameter means and variances used for simulating data are summarized in Table 5.1.

Table 5.1: Parameter values used to generate synthetic data.

Parameter	μ_Γ	σ_Γ	R_0	K	n	β	κ	σ_η
Global location parameter	9	2	60	120	3.5	0.3	5×10^{-3}	5×10^{-3}
Random effect variance (τ)	0.05	0.05	0.2	0.2	0.05	-	-	-

5.3.1 Evaluation of blocking strategies

Synthetic Cry1 concentrations are generated by Euler-Maruyama approximations (Kloeden and Platen, 1992) to the model in Eq. 4.5, which are subsequently averaged over 0.5 hour intervals and corrupted with Gaussian measurement noise to replicate the true measurement process. This is repeated for $L = 400$ locations on a 20 by 20 grid.

Inference is carried out using Algorithm (4) outlined in Section 5.2.1. All parameters are sampled on a log-scale to ensure positivity and constrict the variation along the different dimensions to a similar scale. Prior distributions used in the simulation study are all Gaussian and centred at the true simulation value except for the delay distribution parameters where a uniform prior is used. The priors for the degradation rate, light scaling and measurement error are informative while the priors for the transcription function parameters and hyper-variances are vague.

For the five parameters subject to spatial random effects only the product $\mu_\Theta^{(p)} \exp(\epsilon_i^{(p)})$ is identifiable. Hence a sum-to-zero constraint on each full set of random effects $\boldsymbol{\epsilon}^{(p)}$ is required for identifiability of the overall location parameter and random effects separately. The constraint is implemented by centering $\boldsymbol{\epsilon}^{(p)}$ and adding the mean to the corresponding $\log \mu_{\theta^{(p)}}$ at every iteration of the algorithm after updating all random effects but before updating $\mu_{\theta^{(p)}}$.

To investigate blocking strategies for the high dimensional random effects updates the ESJD for various block sizes are examined. In Table 5.2 the ESJD and wall-clock times for various blocking strategies are reported. The ESJD appears similar for all strategies apart from that where 4 locations at a time are updated which has a considerably lower ESJD and hence worse performance. The wall clock times of complete updates of parameters and random effects are given for the various block sizes. There is a significant cost associated with choosing smaller block sizes that are updated sequentially as a complete update of the 400 locations takes more than twice the time in the strategy where random effects are updated in blocks of 4 compared to a joint full update. Part of this difference is likely due to inefficient queueing between processor cores because when updating only four locations at a time three cores are idle while the slowest likelihood evaluation finishes. The results suggest that there is no computational or statistical gain from a blocking strategy for the random effects which instead should be updated simultaneously, at least for dimensions corresponding to the application considered here.

Table 5.2: Wall-clock time and estimated average ESJD for single full update of all parameters and random effects of 400 locations using four Intel Core i7-6700 processor cores at 3.40 GHz. Average time over 100 iterations.

Random effect block size	4	16	25	100	400
Time (sec)	33.72	21.28	20.09	16.76	15.62
Average $\widehat{\text{ESJD}}$	1.4×10^{-4}	2.5×10^{-4}	2.8×10^{-4}	2.8×10^{-4}	2.8×10^{-4}

5.3.2 Inference validation using synthetic data

To investigate to what extent the methodology developed in this chapter is able to recover the spatial distribution of parameters, data are generated on an 8 by 8 grid using the model in Eq. (5.8) and the parameters (values given in Table 5.1) are re-estimated using the MCMC algorithm given in Section 5.2.1.

For the study to represent a realistic scenario, the prior distributions are highly disperse apart from those of the measurement error SD and degradation rate for which we have informative priors. The prior distributions for hyper-variances are uninformative, uniform on [0, 10]. For the logarithm of global location parameters of R_0, K and n, and $\log \kappa$ we use Gaussian priors with zero mean and variance 10. For the global location of delay mean and SD we use uniform priors on [0, 24] and [0, 20] respectively. For the degradation rate $\log \beta$ and measurement error SD $\log \sigma_\eta$ the Gaussian priors are centered on the true simulation value and have variances 0.25 and 0.17^2, respectively, representing realistic prior knowledge about these parameters as obtained from previous studies of degradation rates of reporter proteins (Yamaguchi et al., 2003) and estimates of the measurement error variance obtainable from reference recordings (Section 4.3.1).

The random effects are initialised by a draw from Eq. (5.5) with $\rho = 0.99$, i.e. the Cressie model which imposes a propriety parameter on the CAR model to obtain a valid data generating process. The value of ρ is chosen to produce a correlation structure that is close to the intrinsic CAR specification, which has an implied value of $\rho = 1$. The hyper-variances are initialised at $\tau = 1$ and the global location parameters and parameters modelled without random effects in a region of high posterior density, ($\pm 20\%$ away from the true simulation value). The algorithm is run for 25k iterations with the initial 10k iterations discarded as burn-in. In Figures 5.2 and 5.3 we report the spatial distributions of posterior means of the transformed parameters, along with the spatial distribution of the true simulation parameters. Additionally, in Figure 5.4 we report the posterior probability of a limit cycle and the equilibrium population size,

along with the dynamics and population size under the true parameters used to generate data.

For the five spatially modelled parameters shown in Figure 5.2 we find that the spatial distribution of posterior means closely resembles that of the true parameter values, both in terms of values of individual pixels and the typical variability across the grid. Some over-smoothing is observed for the delay SD (second row) where the posterior means exhibit a smoother spatial distribution than the true parameter values. We note however that in this scenario, i.e. the first order CAR model being the data generating process, the scale of spatial variability is likely to be much smaller than that of available resolutions of circadian bio-imaging data. In other words, the 8 by 8 grid poses an easier problem in the sense of dimensionality of the posterior, while recovering the spatial variability of the first order CAR process is likely to pose a harder problem than inference for circadian bio-imaging data due to the small-scale structures observed here in the distributions of true values compared to the fairly smooth variability displayed by the estimates from the real data in Chapter 4, Figure 4.6.

True values of the degradation rate β and parameters associated with the experimental procedure, κ and σ_η, are identical across the grid. The posterior means are on average very close to the true value and the spatial variability is small, especially for κ and σ_η. The posterior means of β hint at spatial dependency despite true values being spatially homogeneous. This may be due to dependency structures of the posterior, e.g. similarity can be seen between overestimation of K and the structure of estimated β.

In Figure 5.4 we compare the dynamics, in terms of limit cycle propensity, and population sizes, associated with the true and estimated parameter values. The posterior probability of a limit cycle largely recovers the spatial distribution of the indicator function for sustained oscillations of the macroscopic mean. Some locations exhibit substantial uncertainty ($\hat{P}_{\text{lim. cyc.}} \approx 0.5$). Typically, such locations neighbour both strong and weak oscillators and hence we assert that there is some degree of spatial smoothing of the dynamics. The posterior means of the equilibrium population size are generally very similar to the true values with large differences generally occurring along edges and in corners of the grid, i.e. locations with few neighbours.

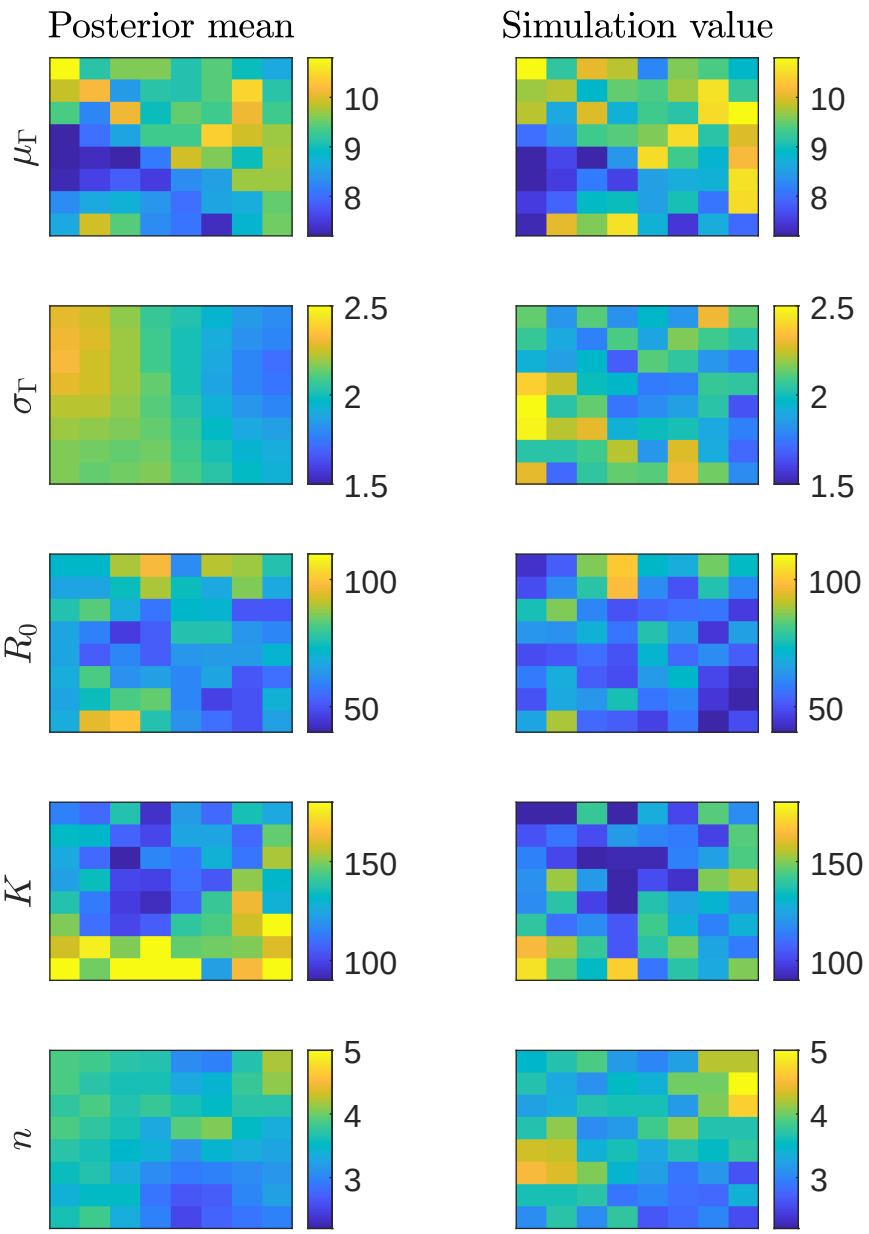

Figure 5.2: Left column: Spatial distribution of posterior means of parameters equipped with spatial prior after 25k MCMC iterations with 10k iterations discarded as burn-in. Right column: True values used to generate data. Values are drawn from the prior distribution CAR model.

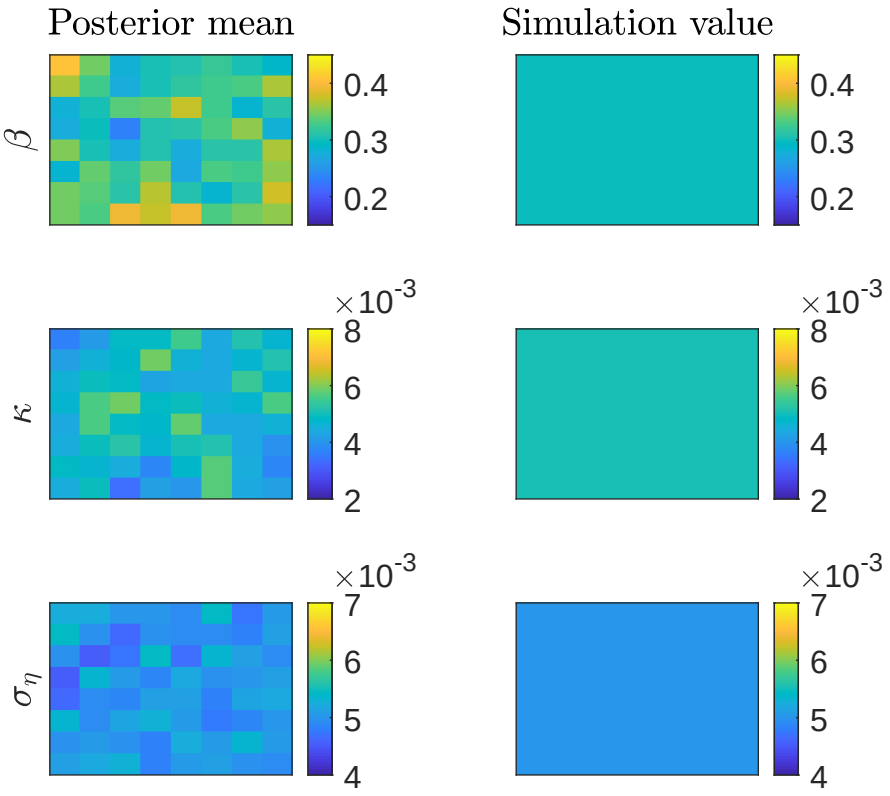

Figure 5.3: Left column: Spatial distribution of posterior means of parameters without spatial prior after 25k MCMC iterations with 10k iterations discarded as burn-in. Right column: True value used to generate data.

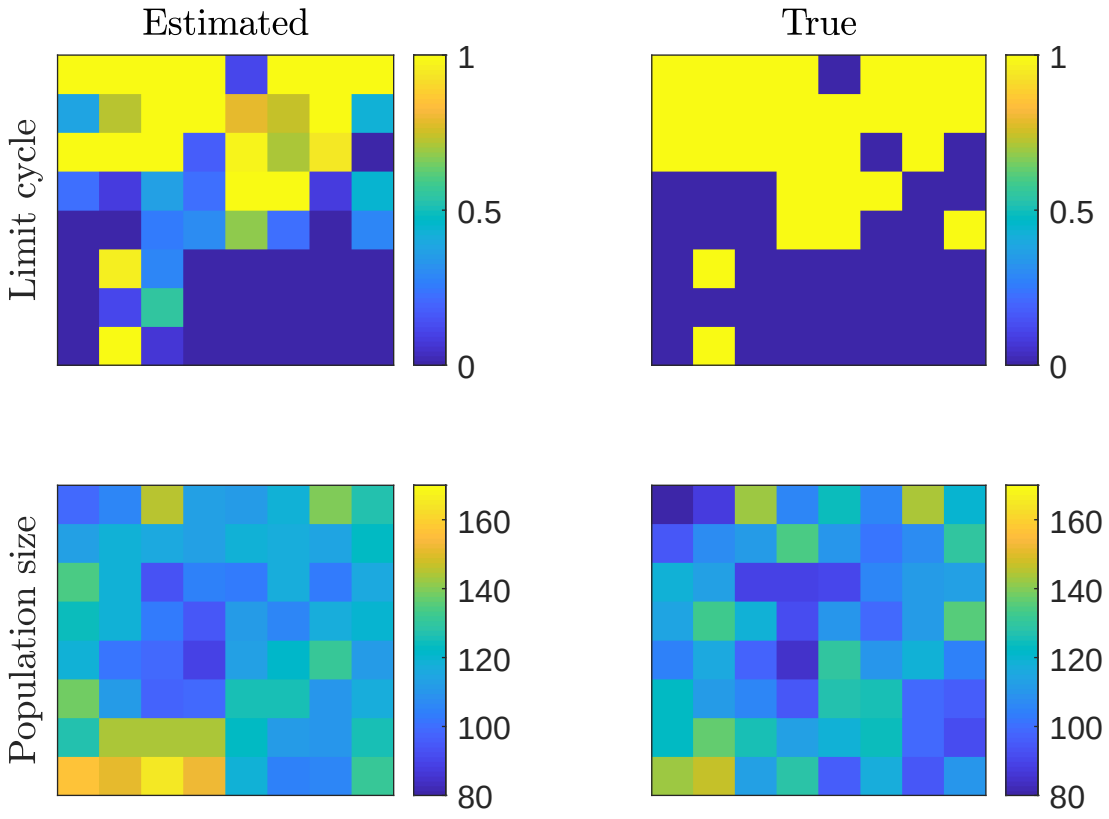

Figure 5.4: Left column: Spatial distribution of posterior probability of a limit cycle and equilibrium population size after 25k MCMC iterations with 10k iterations discarded as burn-in. Right column: Values under data generating parameters.

5.4 Discussion

In this chapter we have considered Bayesian hierarchical modelling, and two variations of the CAR model. The TTFL and measurement model from the previous chapter were extended to a spatial setting using hierarchical modelling by equipping parameters associated with the transcriptional birth-process with a spatial random effects model. By placing a CAR prior distribution on the random effects and designing an MCMC algorithm to sample the otherwise intractable posterior we develop a methodology to

simultaneously infer parameters for a large number of locations, or pixels, of circadian bioimaging data.

Due to the high-dimensional nature of the posterior distribution we directed some attention to efficiency of blocking strategies for sampling the posterior distribution of the random effects. Here we find that, in our setting, no gains, in terms of ESJD or wall-clock time are achieved by sampling the random effects in smaller blocks. The finding regarding the ESJD is consistent with theoretical results for optimal scaling of the proposal variance for exploring a d-dimensional Gaussian distributions by Roberts and Rosenthal (2001), which suggest that the gain achieved by simultaneously exploring more dimensions is exactly cancelled by the smaller jump-sizes required to optimally explore those dimensions. Instead the optimal exploration of the target distribution, in terms of both sample ESJD and wall-clock time, is achieved when simultaneously updating the whole set of random effects.

To validate the methodology we re-estimated parameters of data simulated from the model. We find that we are able to recover slightly smoothed spatial distributions of parameters, along with the dynamical behaviour of the macroscopic rate equation, i.e. limit cycle behaviour and population sizes, as generated by the CAR model. This finding is a preparation to applying the methodology to circadian bio-imaging data of Cry1-luc in the following Chapter 6. The data are subject to spatial variability, as seen in the resulting parameter estimates of Chapter 4, on a typically coarser scale compared to the synthetic data generated here by the CAR model. I.e. our interest primarily lies in spatial structures on the order of tens of pixels, while we are able to recover the dynamics of individual pixels with a large degree of accuracy.

Chapter 6

Application of spatio-temporal model to Cry1-luc data

In this chapter the Bayesian hierarchical framework developed in Chapter 5 is applied to three experimental replicates of the Cry1-luc imaging data introduced in Chapter 4. Parameters of the underlying stochastic single-cell model are estimated using random walk Metropolis along with the random effects and hyper-variances of their respective spatial parameter models, in addition to the experimentally related parameters that govern light-scaling and measurement error variance. Analyses of the data are done at a higher spatial resolution than before which is made possible by the fact that the spatial parameter model allows sharing statistical strength between the large number of time series present in each experimental replicate. The spatial distribution of parameter estimates are interpreted in terms of the robustness measure from Section 3.7.2, where we find that it accurately classifies tissue as either SCN or non-SCN through the posterior probability of limit cycle dynamics of the macroscopic rate equation.

We show how to obtain a posterior distribution of the main oscillation period from the MCMC output using model based stability criteria. We find that the credible intervals are narrower compared to the confidence intervals obtained through spectral bootstrap. The reason for this may be that period estimation using our method, while computationally expensive, provides a higher temporal resolution and precision as the measurement noise is explicitly modelled, in contrast to the spectral bootstrap used in Chapter 2 as an exploratory "black-box" tool, which is limited by the frequency at which data are observed.

Further analysis of the model is carried out by deriving an inhibition profile for the single-cell model, defined as the derivative of the transcription rate with respect to the

delayed inhibitory species. A simulation study is designed to examine how the shape of the inhibition profile affects synchrony and the ability to phase shift from a given periodic orbit (entrainability) of ensembles of molecular oscillators. By evaluating the inhibition profiles at the posterior means of parameters across the SCN locations we are able to investigate the spatial distribution of qualitative properties of the TTFL, in particular differential responses to perturbations in molecular concentration. We generally find that such perturbations cause a smaller phase dispersion in the central organ compared to the edges. This phenomenon can be given a mechanistic interpretation in terms of the inhibition profile and is related to a well documented trade-off between synchrony and entrainability of molecular oscillators (Hasegawa and Arita, 2014).

6.1 Spatial modelling of Cry1-luc

Available data consist of three experimental replicates of Cry1-luc imaging data from the right half of the SCN, denoted Replicate 1-3. To model the spatial dependence observed in the parameters of the single-cell TTFL-model in Chapter 4, the hierarchical model with CAR random effects introduced the previous chapter is assumed. While both parameters and non-linear functions of the data, such as period and amplitude of oscillations, are highly correlated between neighbouring locations, modelling spatial correlation in the parameter layer instead of the process layer of the hierarchical model is motivated by the fact that while temporal filtering provides estimates of the underlying states, i.e. molecular concentration per pixel, we cannot infer the spatial extent of neurons based on pixel observations. Such a procedure would be the first step to fitting a more detailed model where connections between individual neurons are parametrized and mapped. While the network topology of the SCN is an area of ongoing study, feasible approaches suffer some limitations due to the complexity of the problem. Vasalou et al. (2009) simulate and compare various connectivities in a network of highly detailed but synthetic circadian oscillators, Abel et al. (2016) employ a model-free approach, inferring functional connections from experimental bio-imaging data using the maximal information coefficient and McBride and Petzold (2018) construct a deterministic model for the network topology and fit parameters to experimental data by assuming a simplistic phase-amplitude oscillator for the network nodes. The mechanistic single cell transcription model we use here is highly parametrized and additional coupling parameters describing the synchronization between cells in a mechanistic fashion would not be identifiable from the available data. We opt instead to model spatial dependency by parametrizing the correlation of biologically interesting model parameters of nearby

pixels.

The hierarchical approach serves two purposes: firstly, a joint posterior distribution for the entire biological tissue sample is obtained. This is a more principled, albeit computationally expensive, way of inference compared to the approach taken in Chapter 4 where inference was done independently at each spatial location. Secondly, we are able to effectively borrow statistical strength across the entire biological sample which reduces the variance of the resulting estimates while still providing information regarding the spatial variation of parameters which is the property of primary interest. Additionally, the Markov random field specification of the prior distribution of the random effects is specified for points on a regular grid, and while pixels form a regular grid the cells do not. The necessary speed up of the inferential algorithm achieved by assuming a MRF for the spatial process thus relies on adequately capturing spatial dependence on the scale of pixels instead of the cells whose locations are only partially observed.

6.1.1 Data processing

In this chapter, two further Cry1-luc recordings are available, denoted Replicate 1 and 3, in addition to Replicate 2 which was analysed in Chapter 4. In Figure 6.1 the recordings are summarised through their average luminosity at each pixel. Recall that the resolution of the raw data is such that a neuron is roughly the size of an 8 by 8 square of raw pixels. Data preprocessing involves averaging across 4 by 4 raw pixels to produce pixels with a level of stochasticity that was found in Chapter 4 to be realistic for the chemical Langevin approximation. The aggregated time series are de-trended using the same method as in Chapter 4, i.e. subtracting a linear trend estimated using least squares. We use a sub-sample of the data by discarding even-numbered pixel rows and columns. The sub-sampling significantly reduces the computational cost of the analysis. As the parallel implementation of the joint likelihood scales linearly with the number of pixels, the cost of evaluating the likelihood of the sub-sampled data is one quarter of that using the full set of pixels. Furthermore, the exploration of the posterior distribution, specifically for the random effects, is faster when the dimension is reduced, in the sense that the MCMC algorithm requires fewer iterations to adequately explore a lower dimensional space. The sub-sampled data are deemed to have approximate correspondence between the three slices regarding the number of pixels and number of neurons, given that the two halves of the SCN are approximately spherical and symmetric and consist of around 10^4 neurons each, and therefore we can reasonably assume that the sub-sampling retains enough spatial resolution to uncover local structures in the organ.

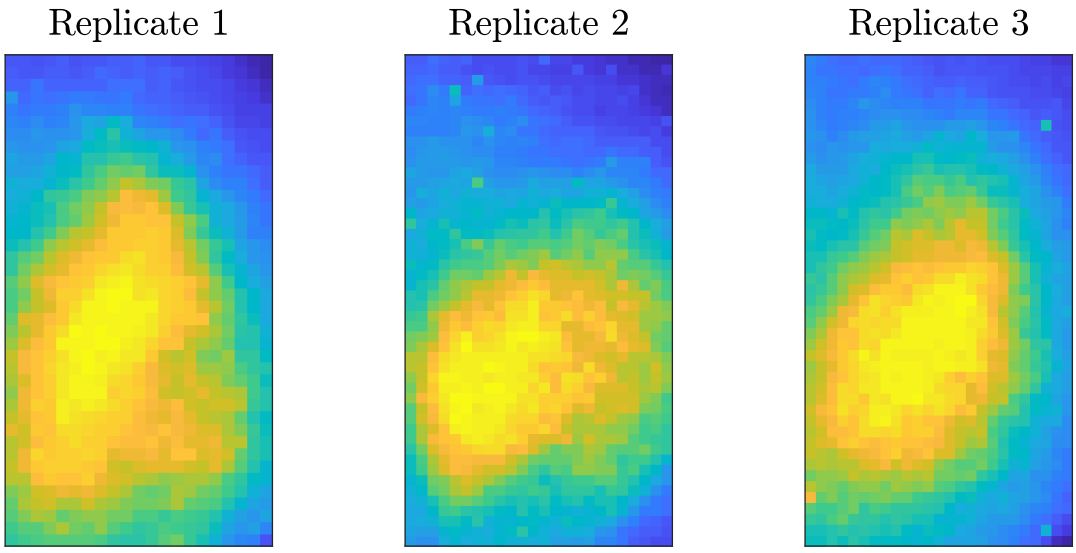

Figure 6.1: Average luminosity of three replicates of Cry1-luc imaging data after preprocessing and sub-sampling. Images cover the right side of the organ along with adjacent tissues in the upper half of the frames due to the drop-shaped geometry of the organ.

6.1.2 Spatial parameter model and prior distributions

The aim is to fit the single-cell distributed delay transcription model from Chapter 4, Eq. (4.5) to the data obtained at each of the pixels of the three experimental replicates. Parameters of the transcription function, μ_Γ, σ_Γ, R_0, K and n are each equipped with the spatial parameter model introduced in Chapter 5, i.e.

$$\theta_i = \mu_\theta \exp(\epsilon_i) \tag{6.1}$$

where θ_i is the parameter value at the i^{th} pixel, μ_θ is a global location parameter that is common for all pixels of each experimental replicate and ϵ_i is a multiplicative random effect that is location and parameter specific. The the prior distribution for the L random effects for the p^{th} parameter, $p = 1, ..., 5$, conditioned on all pixels except i is given by

$$\epsilon_i^{(p)} | \epsilon_{-i}^{(p)}, \tau_i^{(p)} \sim \mathcal{N}\left(\sum_{j \in N_i} w_i \epsilon_j^{(p)}, \tau_i^{(p)} \right), \; p \in \{1, ..., 5\}, \; i, \; j \in \{1, ..., L\}, \tag{6.2}$$

where N_i is the neighbourhood of the i^{th} pixel, here defined as the pixels that share a border or corner with pixel i and w_i is the reciprocal of the size of the neighbourhood, which accounts for pixels at edges and corners having fewer neighbours. Similarly, the variances are parametrised as $\tau_i^{(p)} = \tau^{(p)} w_i$, where $\tau^{(p)}$ is a constant, i.e. random effects at pixels with fewer neighbours have a higher a priori variance. The prior distributions of the hyper variances and global location parameters are taken to be highly dispersed $\log \tau^{(p)} \sim \mathcal{N}(0, 10^2)$ and $\log \mu_{\theta^{(p)}} \sim \mathcal{N}(0, 10^2)$ for $p = 1, ..., 5$.

The prior for the log degradation rate, $\log \beta$, is again taken to be $\mathcal{N}(-0.55, 0.25^2)$, as elicited from Yamaguchi et al. (2003) who report the mean estimates along with standard errors of the functional half-life of luciferase in slice cultures from a 20 neuron experiment. For the logarithm of the measurement error standard deviation $\log \sigma_\eta$ we invoke the procedure outlined in Section 4.3.1 where the variances of the scaled difference of two simultaneous recordings are estimated using maximum likelihood under the assumption of Gaussian measurement noise. We find that this estimation technique produces identical sampling distributions for the three replicates, fitted as $\mathcal{N}(-5.3, 0.17^2)$. Finally, preliminary simulations suggest that the spatial parameter model, specifically for population-size parameters R_0 and K brings enough statistical strength to place a disperse prior on the light-scaling parameter κ ($\log \kappa \sim \mathcal{N}(0, 10^2)$), which was previously found difficult to estimate from single time series.

6.1.3 MCMC algorithm

To explore the posterior distribution we again resort to an adaptive random walk Metropolis algorithm where the likelihood is approximated using the EKBF in Section 3.4.1 with time step $dt = 0.5$h. The algorithm is explicitly described in the previous chapter, Section 5.2.1 and again briefly summarised here. For each experimental replicate, the parameters are divided into blocks consisting of the hyper-variances, $\boldsymbol{\tau}$, random effects for each parameter of the transcription function $\boldsymbol{\epsilon}_\theta$, global location parameters $\boldsymbol{\mu}_{\boldsymbol{\theta}}$ and parameters that are modelled without random effects $\Psi = [\beta,\ \kappa,\ \sigma_\eta]$. The dimension of the update of each block is given in Table 6.1, where L is the total number of pixels for each experimental replicate ($L_{\text{Rep.1}} = 840$, $L_{\text{Rep.2}} = 1152$, $L_{\text{Rep.3}} = 1125$). Note that as β, κ and σ_η are a priori spatially independent the likelihood evaluation in the accept-reject step can be fully parallelised as no computational gains are made by jointly updating these parameters for multiple locations. The proposal variance is scaled adaptively to achieve an empirical acceptance rate of 0.234 (Roberts et al., 1997) using a diminishing adaptation scheme. For the L-dimensional updates the proposal kernel has

a spherical covariance due to the computational cost arising from matrix multiplication involved in generating high-dimensional Gaussians with dependence structure, while for the other blocks proposals are made using the empirical covariance matrix estimated from the MCMC chain. For further details of the algorithm we refer to Section 5.2.1.

Table 6.1: Subdivision of parameters and dimension of updates.

Block	$\boldsymbol{\tau}$	$\boldsymbol{\epsilon}_{\mu_\Gamma}$	$\boldsymbol{\epsilon}_{\sigma_\Gamma}$	$\boldsymbol{\epsilon}_{R_0}$	$\boldsymbol{\epsilon}_K$	$\boldsymbol{\epsilon}_n$	$\boldsymbol{\mu_\theta}$	Ψ
Dimension	5	L	L	L	L	L	5	3

To further decrease the computational cost we aim too initialise chains in dense regions of the parameter space in order to achieve faster convergence. The transcription function parameters and random effects are initialised at values obtained in the independent analysis in Chapter 4, suitably transformed to give preliminary estimates of the random effects and global location parameter of the transcription function parameters that can serve as starting values. Previous results are only available for a grid of pixels and linear interpolation is used to obtain initial values between these points for the full subset of L pixels.

Let $\hat{\boldsymbol{\theta}}^{(p)} = [\hat{\theta}_1^{(p)}, ..., \hat{\theta}_{L'}^{(p)}]$ be the preliminary estimates of the p^{th} parameter of the transcription function, where $L' < L$, i.e. there are only estimates for a subset of the total number of pixels. Now assume that $\log \hat{\boldsymbol{\theta}}^{(p)} = \log \hat{\mu}_\theta^{(p)} + \hat{\boldsymbol{\epsilon}}^{(\boldsymbol{p})}$ where $\hat{\boldsymbol{\epsilon}}^{(\boldsymbol{p})}$ sum to zero by construction. The global location parameters are then initialised at $\frac{1}{L'} \sum_{i=1}^{L'} \log \hat{\boldsymbol{\theta}}^{(p)} = \hat{\mu}^{(p)}$ and the random effects, where available, at $\log \hat{\theta}_i^{(p)} - \log \hat{\mu}^{(p)}$. Similarly, the parameters modelled without a spatial prior distribution (degradation rate β, light scaling κ and measurement error SD σ_η) are initialised everywhere at the spatial mean of posterior means from Chapter 4.

For each of the three replicates, the algorithm is run for 60k iterations and the initial 30k iterations are considered as burn-in and discarded. The number of iterations is taken to produce effective sample sizes of > 100 while the length of the burn-in is chosen by visually examining a subset of trace-plots and discarding iterations until the chains look approximately stationary. The effective sample size of the remaining samples are evaluated using the lugsail batch means ESS estimator of Vats and Knudson (2018), reviewed in Chapter 3, Section 3.6.4.

6.1.4 Effective sample size

In Table 6.2 the minimum, across pixels, conservative univariate lugsail ESS of the resulting chains are calculated and reported. Conservative estimates are obtained by varying the batch size between $N^{1/3}$ and $N^{1/2}$ where N is the nominal chain length after burn-in is discarded and the batch size resulting in the lowest ESS is chosen. The parameters typically have an ESS around .3 percent of the nominal length of the chain, indicative of a high degree of autocorrelation and slow mixing.

Table 6.2: Minimum lugsail batch means effective sample sizes.

Parameter	μ_Γ	σ_Γ	R_0	K	n	β	κ	σ_η
Rep. 1 ESS	106	106	106	106	106	101	103	104
Rep. 2 ESS	105	105	105	105	105	113	111	111
Rep. 3 ESS	106	106	105	105	105	109	109	110

6.1.5 Residual analysis

Residuals time series are obtained at each pixel by comparing the filtering mean output evaluated at the posterior mean of the parameters with the observed data. To verify approximate normality of the residuals a one-sample Kolmogorov-Smirnov test is performed with the null hypothesis that the residuals come from a normal distribution. The test is repeated at each pixel and the resulting p-values are adjusted using Bonferroni correction to account for multiple testing. The test does not reject the null hypothesis at any of the locations apart from three in Replicate 3.

To assess the goodness of fit the residual time series are tested for significant periodicity. The spectral densities, specifically the bootstrap distributions of the period of the residual series are estimated using the spectral resampling methodology developed by Costa et al. (2013) and reviewed in Chapter 2. If the 99 percent bootstrap confidence interval has endpoints in the range $0-30$ hours, the mean period is recorded and reported in Figure 6.2. Generally, we can locate low amplitude 12 hour periodicity in central SCN tissue and oscillations with period close to 0h, consistent with noise, are found in adjacent tissues. The majority of locations are free of 24 hour periodicity in the residuals, indicating that the model and estimated parameters essentially capture the circadian dynamics. We note that the findings are consistent with those in Chapter 4 but bring a higher spatial resolution to the distribution of 12 hour periodicity, highlighting that the

generating mechanism is limited to the SCN tissue, though not present everywhere.

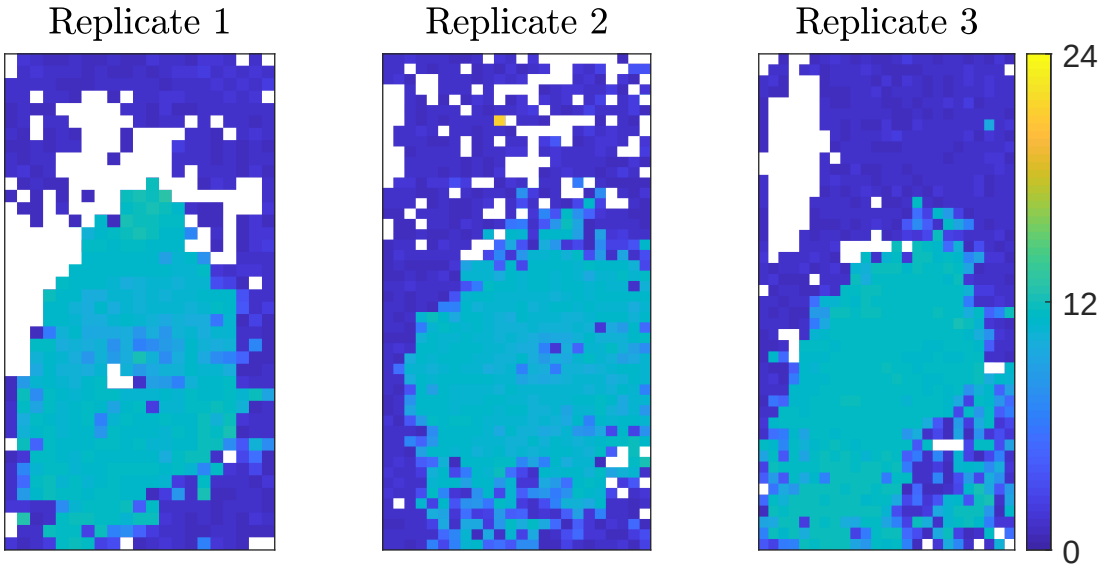

Figure 6.2: Dominant periodicity in residual time series, if period is in the range $0-30$h, obtained by spectral resampling.

6.1.6 Parameter estimates

The posterior means of the spatially modelled parameters are calculated by recombining the chains for the spatial parameter means and random effects using the identity in Eq. (6.1) after discarding the first 30k iterations. The resulting parameter estimates for the three replicates are reported in Figures 6.3-6.5 while numerical values are given in Appendix A, Tables A.1-A.3.

The estimates of the delay distribution mean are consistent between the three experimental replicates with values between 8.5 and 9.5 hours. There is some spatial structure in the estimates that varies between the three replicates but the SCN neurons generally oscillate with a period very close to 24 hours and as the mean of the delay distribution strongly influences the period of the oscillations a low variation in the estimates is expected. Our estimates are highly consistent with those of Ananthasubramaniam et al. (2014) where a delay of 9.5 hours produced circadian oscillations with

a period of 23.7 hours in a deterministic, discrete delay model of autorepressive *Per* transcription mediated by VIP.

The dispersion of the delay distribution dispersion is a key parameter in tuning the strength of oscillatory dynamics, along with the Hill coefficient and degradation rate. All three replicates exhibit higher delay dispersion estimates for central SCN locations compared to shell locations. The estimates are consistent across the three replicates with values ranging between 2 and 4.5 hours and somewhat lower for Replicate 3. Both the maximum transcription rate, R_0 and the dissociation coefficient K show no consistent spatial structure between the three replicates. Values for the two parameters are more similar for Replicates 1 and 2 ($45 < \hat{R}_0 < 65$ and $100 < \hat{K} < 150$) while somewhat higher for Replicate 3 ($70 < \hat{R}_0 < 90$ and $150 < \hat{K} < 210$). Both parameters are related to the underlying molecular population, R_0 being the maximum transcription rate (molecules/hour) and K the population of delayed Cry1 required for a 50 percent decline in transcription. In Figure 6.4 we report approximate molecular counts associated with the equilibrium solution at each location across the three replicates and find that higher estimates of R_0 and K combined with lower estimates of κ in fact suggest a higher average population size. The estimated molecular population sizes are consistent with those obtained assuming the spatially independent approach in Chapter 4. It should be noted that the estimates reflect a temporal mean at each location and do not convey the magnitude of the oscillations about that mean. For typical amplitudes we refer to Figure 6.11 in which paths of the stochastic model are simulated for parameter values typical of Replicate 1 and 2, suggesting counts ranging between 50 and 200 molecules across the circadian cycle. These estimates are likely to be very sensitive to slight variations in the experimental procedure and violations of the model assumptions and specific counts should therefore only be viewed as approximate indicators of molecular population sizes involved. Single neuron measurements of Cry1 mRNA population sizes are not available in the literature, however, Abel et al. (2015) simulate data using a stochastic model of Per mRNA and find population sizes in the range $0 - 400$, consistent with our findings.

In addition to the delay dispersion, the Hill coefficient, n, is also a decisive parameter in tuning the degree with which solutions oscillate and whether a limit cycle exists. Interestingly, estimates of n are systematically higher in central locations of the SCN compared to the shell and this is consistent across all three replicates. The spatial distribution is similar to that of the delay dispersion, suggesting an interaction between the two parameters which we will explore further in Section 6.2.

The estimates of parameters modelled without a random effect component, namely the degradation rate β, the light scaling κ and the measurement error standard deviation

σ_η, are reported in Figure 6.5. Neither the degradation rate nor standard deviation of the measurement error show any noticeable spatial structure in Replicate 3, however in Replicates 1 and 2 the degradation rate is estimated to be lower in the central SCN. Heterogeneity of β is unexpected, in fact the parameter is modelled without a priori spatial dependence. The spatial mean of the estimates are consistent between the three replicates and the posterior distributions are more concentrated than the informative priors, ($0.1 < \hat{\beta} < 0.3$ and $2 \times 10^{-3} < \hat{\sigma}_\eta < 10^{-2}$). The light scaling parameter has a clear structure, similar to that of the smoothed estimates used to elicit prior means in Chapter 4. The structure is again likely due to the geometry of the biological sample where the centre is raised and closer to the camera (see Chapter 4, Section 4.3.2). The estimated values are more similar for Replicates 1 and 2, while for Replicate 3 the estimated values are slightly lower, which is consistent with a higher average population size.

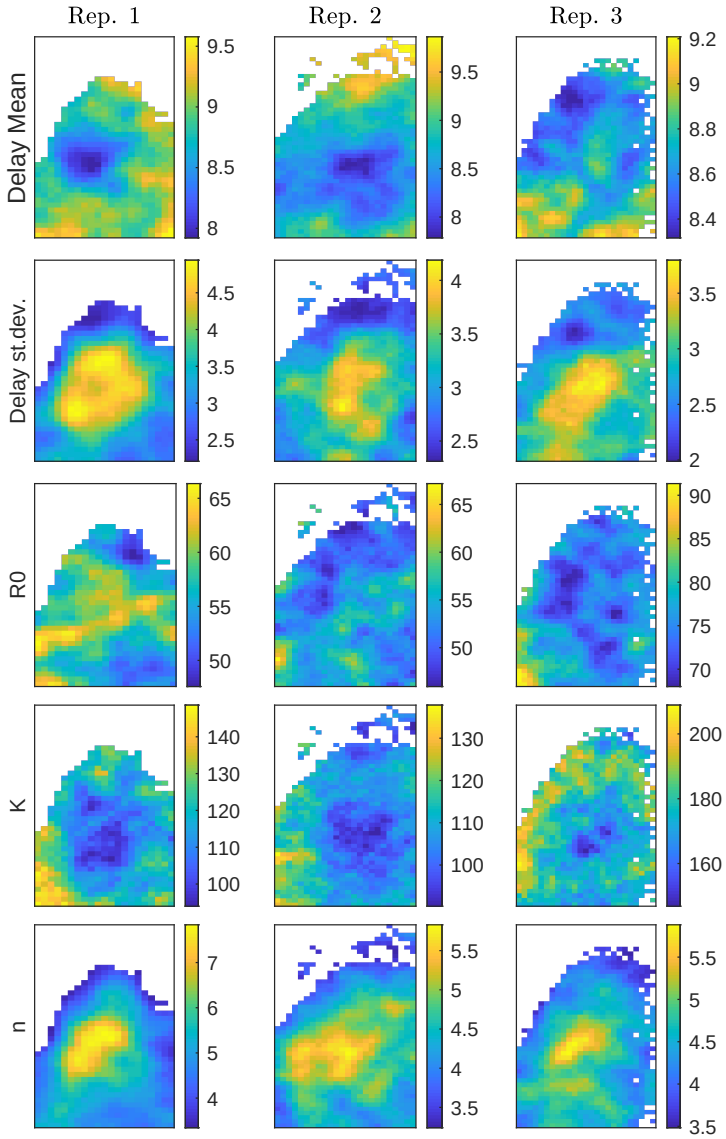

Figure 6.3: Posterior means for delay distribution mean μ_Γ (top row), delay distribution standard deviation σ_Γ (second row), maximum transcription rate R_0 (third row), dissociation coefficient K (second row from bottom) and Hill coefficient n (bottom row) for three replicates of Cry1-luc imaging data. Colour-scaling varied between the three replicates to illustrate similarity of spatial distribution.

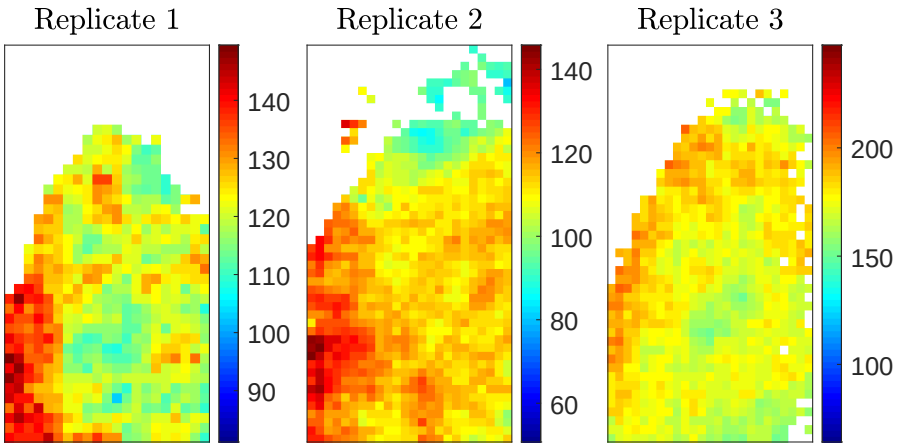

Figure 6.4: Estimated molecular population sizes for three experimental replicates of Cry1-luc. The molecule counts are obtained by calculating the unique positive equilibrium solution of the macroscopic rate equation.

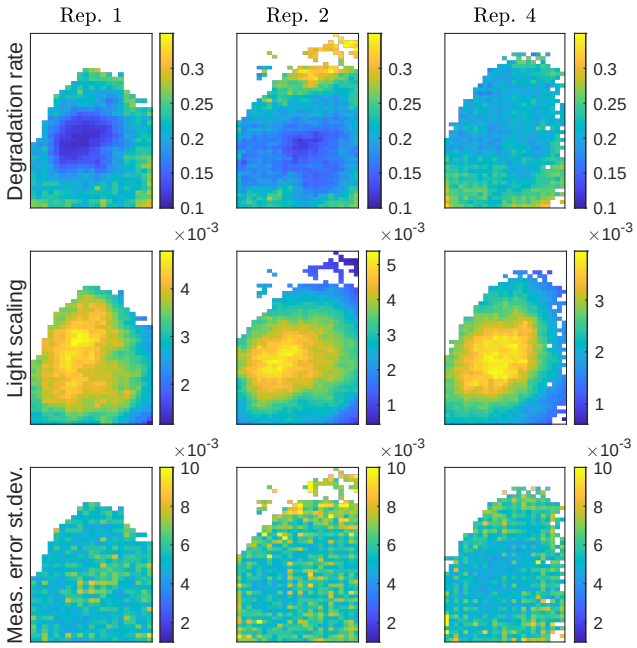

Figure 6.5: Posterior means for degradation rate β (top row), light-scaling κ (second row) and measurement error standard deviation σ_η (third row) for three replicates of Cry1-luc imaging data.

Replicate 2 was analysed in Chapter 4 on a sparser grid assuming a spatially independent univariate time series methodology and a finer likelihood discretisation. The results obtained in this chapter are similar for the parameter estimates corresponding to the SCN tissue but now provide a higher spatial resolution. A compelling argument in favour of the random effects formulation used in this analysis is that it results in an improved identification of the light scaling parameter. The reason is that R_0, K and κ are all related to population size, R_0 and K of the underlying molecular population and κ of the light signal from a given population. If the noise level is low due to large molecule counts the likelihood surface of the three parameters has very little curvature. With the approach taken here the global location parameters of R_0 and K become easier to estimate as more data can be used to evaluate the likelihood. This, combined with the sum-to-zero constraint on the random effects results in an improved identification of κ despite its marginal likelihood using the same amount of data as the spatially independent inferential procedure and the prior being diffuse.

6.1.7 Robustness of oscillations

In Figure 6.6 the posterior probability of a limit cycle, $\hat{P}_{\text{lim. cyc.}}$ is reported for the three replicates (see Section 3.7.2, specifically Eq. (3.51)). This quantity is estimated by calculating the proportion of the MCMC chain for which the parameters imply a positive real part of the principal eigenvalue of the Jacobian. The rounding scheme in Eq. (4.23) is used to construct an approximate integer-size Jacobian.

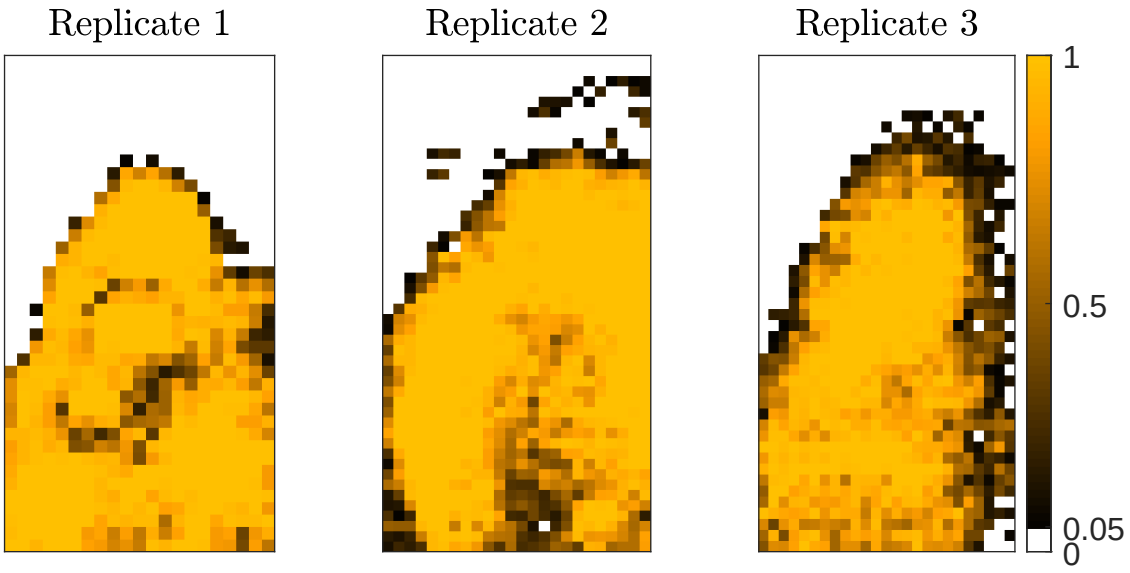

Figure 6.6: Posterior probability of a limit cycle for three replicates of Cry1-luc imaging data.

The resulting estimates are very similar to those obtained previously using spatially independent prior distributions in Chapter 4 but offer a higher spatial resolution. The SCN tissue locations typically have a $\hat{P}_{\text{lim. cyc.}}$ close to 1 with a few locations in Replicate 1 and 2 exhibiting low values in the central and ventral SCN.

A high value of $\hat{P}_{\text{lim. cyc.}}$ is associated mainly with high values of the Hill coefficient, n, as well as an increase in σ_Γ. Recall from Section 4.2.1 that, as the delay distribution variance goes to zero, the Hill coefficient required for limit cycle decreases. Thus, holding all other parameters fixed, a large value of the delay distribution dispersion σ_Γ is associated with *less* robust oscillators in the sense of Eq. (3.51), but this

effect here is compensated by an increased Hill coefficient n. The relationship between the two parameters is examined in Figure 6.9 where estimates of σ_Γ and n are plotted for the three experimental replicates and coloured coded with respect to $\hat{P}_{\text{lim. cyc.}}$. A clear positive relationship between the two parameters is evident and generally, weak oscillators tend to have low values of both parameters while the strong oscillators are associated with higher Hill coefficient and a higher variability in delay dispersion. This relationship and its effect on the robustness of circadian dynamics in the SCN is explored in further detail through a simulation study in Section 6.2 in the end of this chapter.

6.1.8 Bayesian Period estimation

Estimation of the period of oscillations is typically a key step in the analysis of circadian data. For example, several different techniques for period estimation are used and implemented in BioDare2, a data repository and data analysis platform available at biodare2.ed.ac.uk (Zielinski et al., 2014). The techniques provided in BioDare are typically based on spectral analysis and are generally computationally fast but lack uncertainty quantification, with the exception of the spectrum resampling method by Costa et al. (2013) (see Chapter 2). Here, we shall discuss and apply a model-based Bayesian approach for period estimation and uncertainty quantification and compare the results to the spectrum resampling method for the Cry1-luc data.

Recall from Section 3.7 that dynamical systems with oscillating solutions have Jacobian matrices with complex eigenvalues, and that the period of the oscillatory solutions are given by $P = 2\pi/\text{Imag}(\lambda)$. With this in mind, it is straight forward to obtain e.g. the posterior mean of the period P from the MCMC output on the parameter estimates, in a similar fashion to how we calculate the posterior probability of a limit cycle. Formally, the quantity we want to estimate is

$$\mathbb{E}[P|\mathbf{Y}] = \int_\Theta \frac{2\pi}{\text{Imag}(\lambda_\theta)} \pi(\theta|\mathbf{Y}) d\theta, \tag{6.3}$$

where λ_θ depends on the parameters θ through the Jacobian in Eq. (4.12). The posterior mean period can in practice be estimated by constructing the Jacobian of the system for each sample of the posterior distribution and recording its eigenvalues, which are subsequently used to construct samples of the posterior distribution of the period of oscillations using $P^{(i)} = 2\pi/\text{Imag}(\lambda_{\theta^{(i)}})$. Similarly, one may construct credible intervals by calculating the relevant percentiles of the samples of the period posterior distribution.

The Bayesian approach to period estimation allows for, as in our case, separately

modelling the measurement process and underlying process of interest. Hence parameter uncertainty is propagated to the period uncertainty while the measurement noise is not. Additionally, while spectral methods rely on frequently sampled data to give high temporal resolution in period estimates, the Bayesian approach only relies indirectly on high frequency observations as it typically improves parameter inference. However, the Bayesian approach relies on the assumed model and may thus be sensitive to model misspecification. Additionally, in our case the eigenvalues used to construct the period estimator are those of the macroscopic rate equation, hence the period estimates ignore any effects of the noise term on period length.

We note that the identity $P = 2\pi/\text{Imag}(\lambda)$ is undefined for strictly real eigenvalues and thus we are in a sense conditioning on the existence of oscillations to guarantee existence of Eq. (6.3). In the case of posterior mass lying in a region where $\text{Imag}(\lambda) = 0$ an alternative definition is obtained by replacing $\pi(\theta|\mathbf{Y})$ with $\pi(\theta|\mathbf{Y}, \text{Imag}(\lambda) > 0)$ and integrating over the parts of the parameter space for which $\text{Imag}(\lambda) > 0$, producing a posterior mean of the period conditional on oscillating solutions. In our case however all posterior mass lies in a region where the eigenvalues are strictly complex, i.e. the posterior probability of non-oscillating solutions is zero.

In Figure 6.7 we report the spatial distribution of period estimates obtained using spectral resampling along with the width of 95 percent bootstrap confidence interval for the three experimental replicates. The period estimates are close to 24 hours across the entire SCN and the widths of the confidence intervals are 1.5-2.5 hours. In Figure 6.8 the posterior mean of the period is reported, along with the width of 95 percent credible intervals. The color-scaling is kept fixed between the two methods, revealing an increased temporal resolution in period estimates obtained using the Bayesian approach. Furthermore, the widths of the credible intervals are between 0.5 and 1.5h. Note however that widths of the bootstrap confidence intervals and Bayesian credible intervals are not directly comparable as their interpretations differ: the bootstrap confidence intervals capture uncertainty in period estimates due to measurement *and* molecular noise, while the credible intervals capture the uncertainty due to unknown parameter values of an approximating deterministic process.

Application of the model-based Bayesian period estimator reveals evidence for spatial variation of the periodicity of the molecular oscillators. Specifically Replicate 1 and 2 exhibit a lower period (< 23h) in central regions of the SCN compared to the edges of the organ. Such spatial variation is not found using the spectral resampling methodology, which produces period estimates close to 24 hours across all SCN tissue.

Li et al. (2020) recently found experimental evidence that intrinsic noise, i.e.

noise caused by the discrete nature of molecular interactions and here modelled by the CLE, may lengthen the period of circadian oscillations. By treating single cells with a transient transcriptional noise enhancer the authors found that the period of circadian oscillations was significantly increased by approximately 1.5 hours. For the single-cell TTFL model, we find that intrinsic noise can both increase and decrease the period length when simulated trajectories of the stochastic model are compared to numerical solutions of the macroscopic rate equation. The mean difference in period depends largely on the assumed parameter values and for parameters corresponding to those inferred from the Cry1-luc data the difference between the deterministic and stochastic oscillator and is in the range -2 to 2 hours. Our empirical findings, taken together with those of Li et al. (2020) suggest that intrinsic noise may play an important role in circadian functionality and may have different impact in different regions of the SCN.

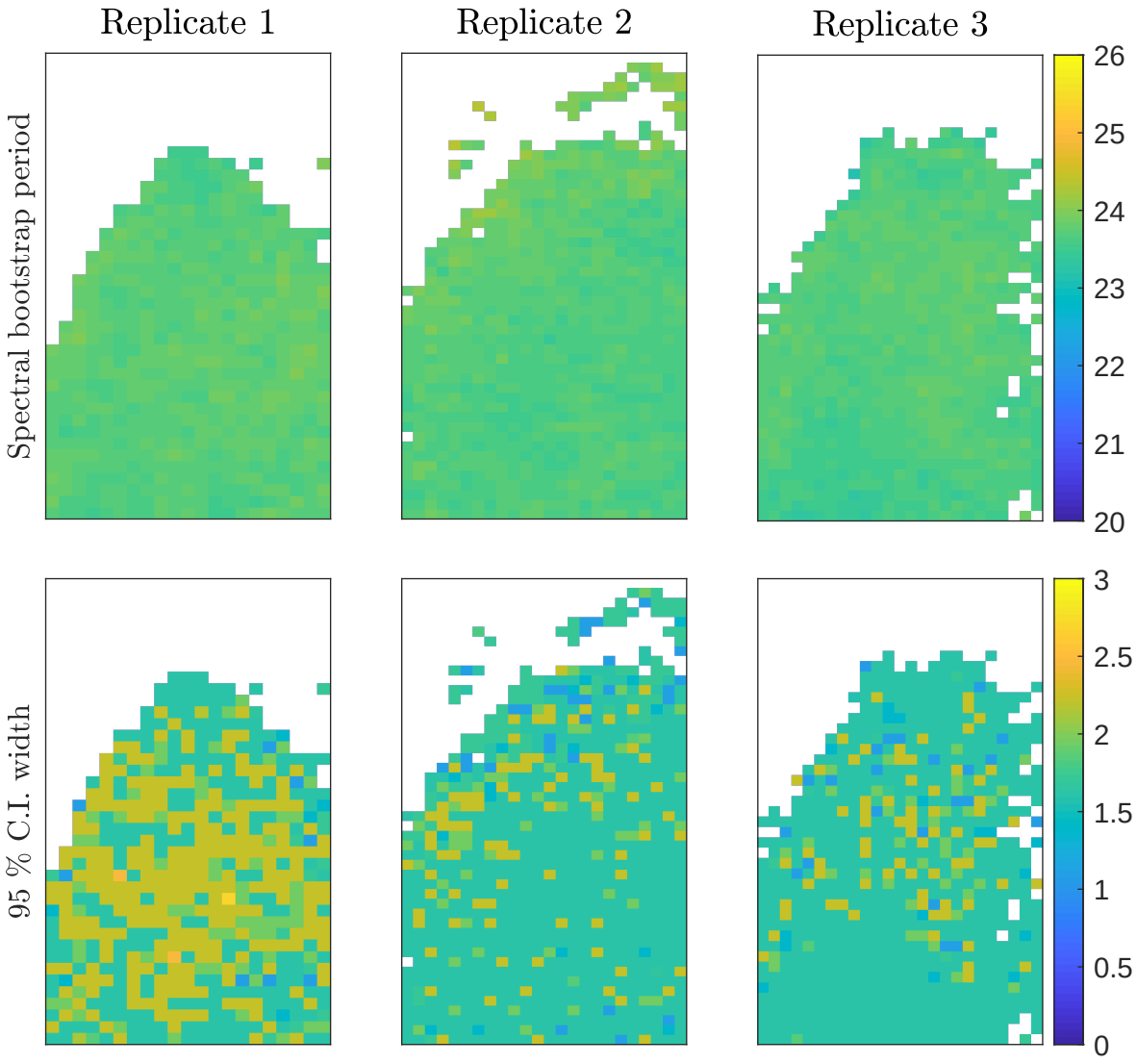

Figure 6.7: Mean of bootstrap estimates of period corresponding to the dominant peak of the spectral density and widths of 95 percent bootstrap confidence intervals.

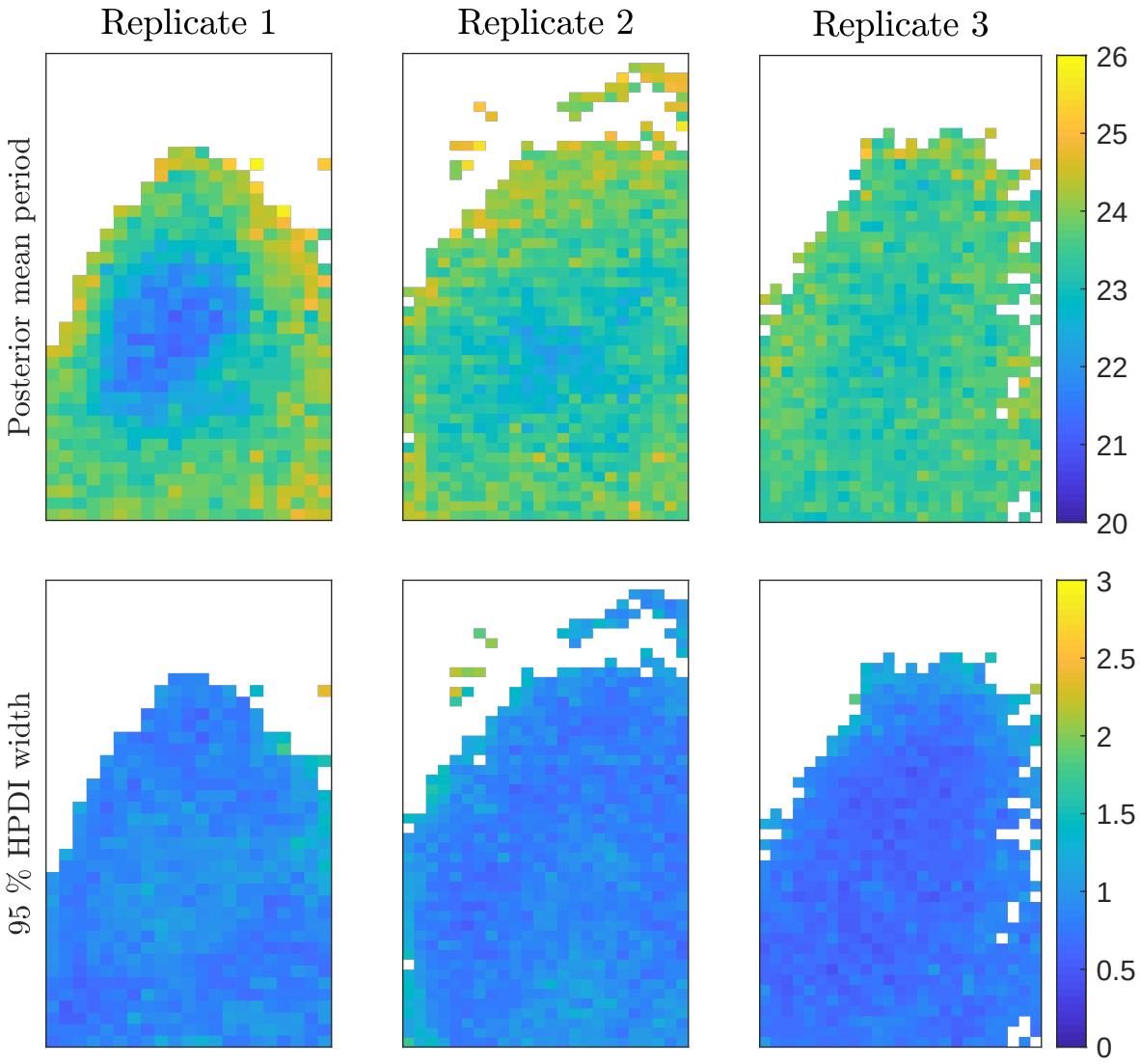

Figure 6.8: Posterior mean of period and widths of 95 percent HPDIs calculated using Bayesian model-based method.

6.2 Inhibition profile & Entrainment of molecular oscillators

We find that the propensity for limit cycle dynamics is mainly driven by the Hill coefficient, n, consistent with the theory on the Hill function discussed in Section 4.2.1 and Tyson (2002). That is, for other parameters held fixed, limit cycle oscillations can be induced by increasing n to some critical value which depends on the other parameters and for the special case of a fixed delay is known to be exactly $n_{\min} = 1$. The delay dispersion and Hill coefficient can thus trade-off in a way such that an noise induced oscillator, i.e. where $\mathrm{Re}(\lambda) < 0$ and $\mathrm{Imag}(\lambda) \neq 0$, may have a relatively High hill coefficient due to a high delay dispersion, and vice versa. This trade-off can be inspected in Figure 6.9, where only locations corresponding to SCN tissue are plotted with the estimated values of the Hill coefficient on the y-axis and estimated values of the delay dispersion (SD) on the x-axis for the three replicates. Additionally, the points are coloured according to their posterior probability of a limit cycle.

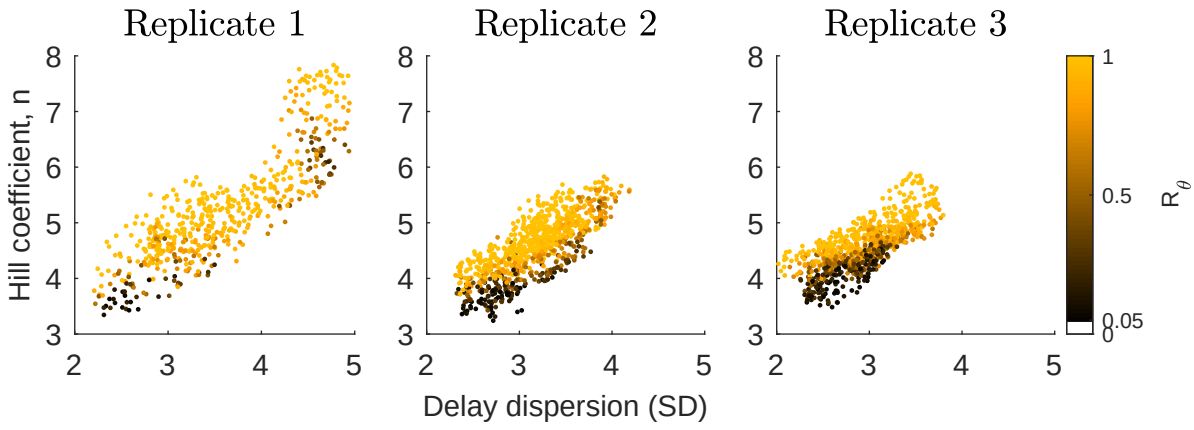

Figure 6.9: Scatter plot of estimated delay distribution standard deviation and Hill coefficient for the three experimental replicates, color coded by probability of limit cycle. Oscillators lie on a ridge in the space formed by the Hill coefficient and delay SD. Weak oscillators are found primarily at the lower end of the relationship and typically have lower estimates of both parameters.

We find that the oscillators lie on a ridge in the parameter space such that there

is a positive correlation between the Hill coefficient and delay SD for all three replicates. The relatively weak oscillators typically occupy the lower part of this relationship with estimates of n, and interestingly of the delay dispersion as well. As lowering the delay dispersion theoretically increases the propensity for limit cycle dynamics, and a wide range of estimates for the delay dispersion are found for the relatively strong oscillators, we hypothesize that the shape of the delay distribution may have other qualitative effects on the behaviour of the oscillators.

To study putative dynamical differences induced by moving along the (n, σ_Γ)-ridge we define the *inhibition profile* (IP) of the transcription function as the derivative of the transcription rate with respect to the past levels of mRNA, evaluated at and divided by the equilibrium level x^* to produce a normalised quantity that may be compared between different oscillators. As the IP can only be evaluated at a finite number of points between 0 and the maximum delay time $\tau_{\max}$ we state it here in discretised form, where g_j are the weights given by the gamma delay distribution,

$$\text{Inhibition}(t) = (x^*)^{-1} \frac{d}{dx_t} \frac{R_0}{1 + \left(\frac{\sum_{j=0}^{\tau_{\max}} x_{\tau_{\max} - j} g_j}{K} \right)^n}. \tag{6.4}$$

The magnitude of the IP evaluated at time $t \in (0, \tau_{\max})$ corresponds to the change in transcription rate due to a small increase in inhibiting species at t. The function allows us to examine how perturbations in mRNA affect the transcriptional activity over the subsequent 24 hours through the delay distribution. In Figure 6.10 the IPs of various oscillators along the (n, σ_Γ)-ridge are visualised.

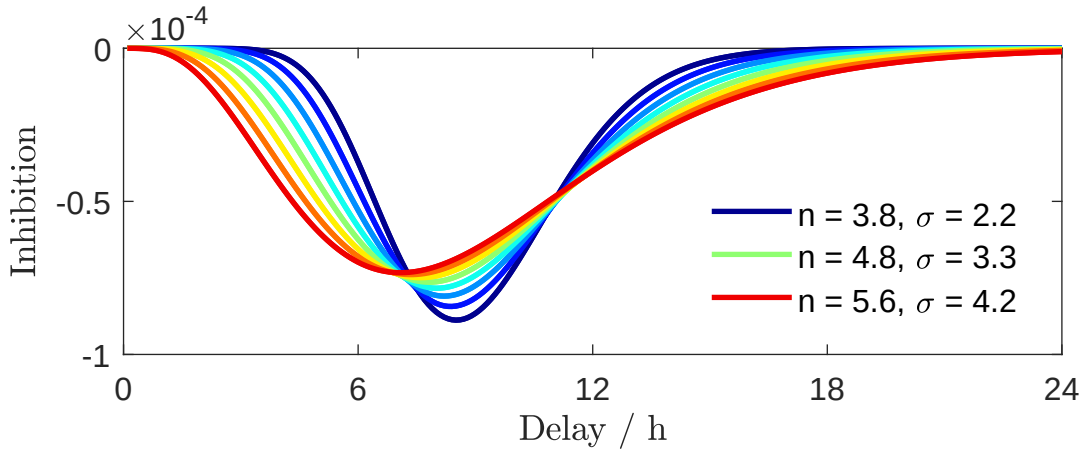

Figure 6.10: Inhibition profile for various estimated combinations of Hill coefficient and delay dispersion. Low delay dispersion gives a peaked inhibition profile while a high delay dispersion produces a flatter curve. Colour spectrum (blue to red) of curves corresponds to moving along the (n, σ_Γ)-ridge (low to high).

The IP is more peaked for oscillators with low delay dispersion, meaning that the inhibitory action is concentrated to a shorter time period, and the onset of the inhibition is not as gradual compared to an oscillator with wide IP. Oscillators with a low delay dispersion are driven by inhibitory dynamics that are closer to a step function, i.e. an "on-off" switch for the inhibition due to a certain mRNA fluctuation. In contrast, a wide IP distributes the inhibitory action over a longer time period with earlier and less sudden onset and a lower maximum effect.

As circadian timekeeping is an emergent property of the aggregate of synchronised, noisy cellular clocks, we turn our attention to the ensemble properties of the single cell model and investigate the sensitivity of the synchronisation of the circadian rhythms to perturbations. While our model lacks a mechanism for inter-cellular signalling and zeitgeber (e.g. light) input, which are the de-facto conferrers of synchrony and entrainment in intact SCN tissue (Herzog et al., 2017), we achieve synchronised oscillations of the ensemble through a common initial condition and approximate a zeitgeber input by a shock that is inconsistent with the oscillators' natural limit cycle.

Synchrony and entrainability intuitively trade-off as the former is the ability of an ensemble of oscillators to remain on or close to a given periodic orbit and the latter the ability to phase-shift to a new orbit (see e.g. Hasegawa and Arita (2014) where our definition of synchrony is referred to as *regularity*). We stress that in the absence of a

coupling mechanism, defining synchrony as the phase coherence of ensembles of oscillators is reasonable. The trade-off between synchrony and entrainability is experimentally studied in SCN neurons by An et al. (2013) who find that faster entrainment is achieved when high levels of VIP desynchronise individual neurons. Our working hypothesis is that molecular oscillators with wide inhibition profiles exhibit greater resilience to desynchronisation in terms of periodicity and phase when exposed to exogenous shocks in mRNA counts, as a given shock acts over a longer time interval. To investigate this hypothesis, a simulation study is designed where ensembles of two prototypical oscillators on either end of the $(n,\ \sigma_\Gamma)$-ridge are forward simulated from a common initial data function containing a positive burst of mRNA with varying duration (0-4h) and amplitude corresponding to a typical mRNA peak. The distribution of the phases for the varying shock scenarios are recorded and compared to the distribution of phases of the unperturbed ensembles.

For each oscillator type 2000 realisations are simulated using an Euler-Maruyama approximation (Kloeden and Platen, 1992) with step size $dt = 0.1$. The Hill coefficients are set to 5.6 and 3.82 and delay SD to 4.2 and 2.2 for the two oscillator types, respectively. To achieve oscillators with identical period the delay mean is set to 9.4 and 9.1h, respectively, for the two types. All other parameters are set identical for the two oscillators, and are based on estimated values for the three experimental replicates: $R_0 = 50$, $K = 100$, $\beta = 0.25$. The phase distributions are obtained by estimating the phase of each simulation replicate based on five circadian cycles following the initial data using spectral analysis.

The results of the simulation study are summarised in Figure 6.11, where the initial data along with the first cycle of oscillations are shown for a randomly selected subset of the replicates to avoid over-plotting. Trajectories of the oscillator type with wide IP are given in red and narrow IP in blue. The panels in the top row of Figure 6.11 reveal that in the unperturbed scenario the two oscillator types have similar dynamics in terms of phase, amplitude and period. As the duration of the shock increases the oscillators with narrow IP tend to oscillate with a small peak directly following the shock, followed by a larger oscillation which is phase-shifted by a larger value than the red ones. The temporal smoothing of the wide IP mitigates a sudden transcriptional response, instead delaying the first oscillation post-shock which appears less stochastic.

The polar histograms reveal that in the unperturbed scenario both oscillator types have similar phase dispersion. As the duration of the shock is increased the synchrony of both oscillator types is gradually weakened, in that the phase dispersion is increased. The loss of synchrony is greater for the narrow IP oscillator, along with total shift in

ensemble mean phase. Interestingly, the longest (4 hour) shock duration produces exact time keeping with a low phase dispersion compared to shorter shocks.

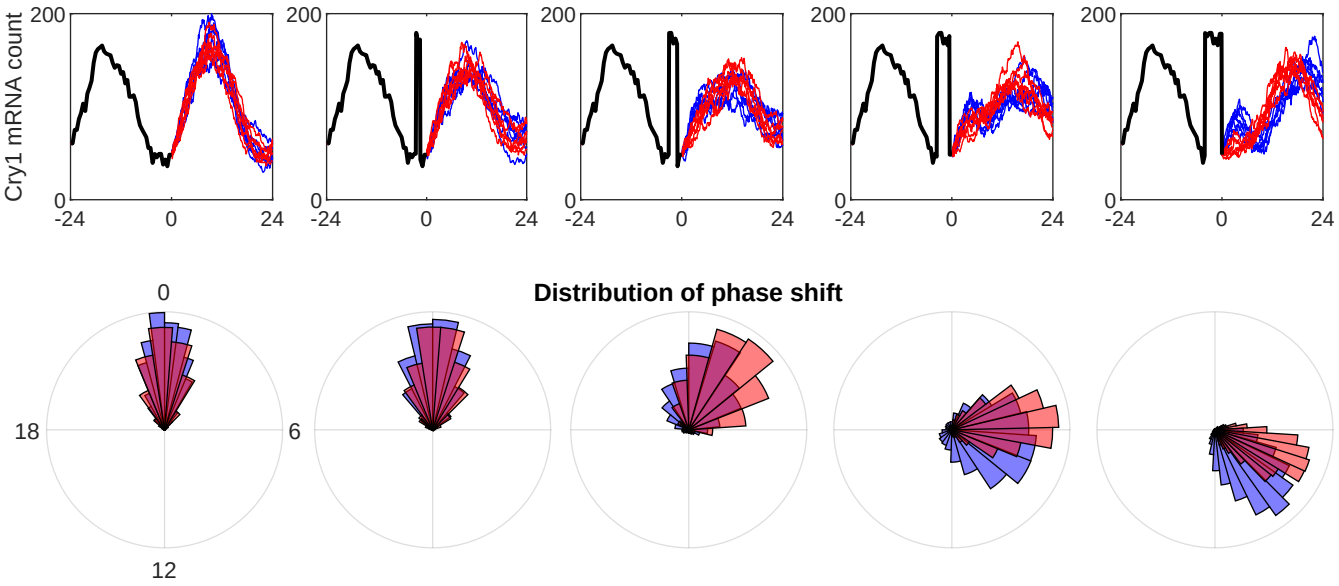

Figure 6.11: Simulations of trajectories with large perturbation of varying length (0-4 hours). Red lines and histograms have parameter combinations with large delay distribution SD ($\sigma_\Gamma = 4.2$) and Hill coefficient ($n = 5.6$) while blue have low values for both $\sigma_\Gamma = 2.2$, $n = 3.82$. Polar histograms in bottom panels show deviation between average peak times of the non-perturbed and perturbed cases. The "wide IP oscillators" (red) exhibit smaller phase dispersion and phase shift compared to the "narrow IP oscillators" (blue).

To investigate whether we can construct spatial heterogeneity of phase-resetting and dispersion in different regions of the SCN the simulation study is conducted for the 4 hour shock scenario but this time using the spatially distributed posterior means of the parameters, evaluated at each pixel for each of the three experimental replicates. Due to the relatively large number of pixels the number of replicates is reduced to 200 for each pixel. The results, summarised by the standard deviation of the phase distribution, are reported in Figure 6.12. We find consistently across all three replicates that centrally located oscillators have a lower phase dispersion compared to the shell oscillators following a 4 hour peak-sized shock in mRNA levels. The spatial distribution

of the phase dispersion is similar to that found for the delay dispersion parameter (see Figure 6.3), and hence the shape of the inhibition profile.

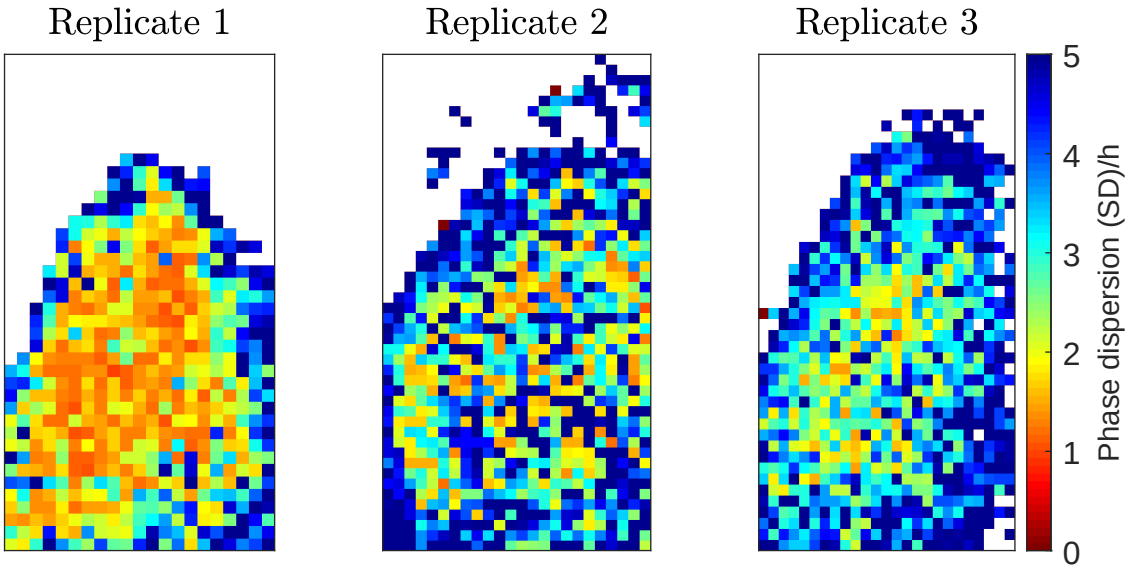

Figure 6.12: Spatial distribution of phase dispersion (SD) following a 4 hour peak-size shock in initial data during trough. For each pixel we generate 200 realisations of the model using the location specific parameter estimates and record the phases to obtain a phase distribution with an associated standard deviation. The positive shock in mRNA count causes a stochastically induced de-synchronisation of the ensemble timekeeping that increases the standard deviation of the phase distribution to 1-5 hours. The phase dispersion is typically lower for centrally located oscillators due to their wider inhibition profiles, suggesting that the ensemble timekeeping in the central SCN is less sensitive to exogenous shocks compared to the shell.

The relative insensitivity to large shocks in the central SCN can be understood by the delay distribution acting as a regulator of delayed mRNA. A highly dispersed delay distribution (high σ_Γ) gives a wide IP, meaning that time with which Cry1 mRNA is transcribed and the resulting protein products undergo transport and dimerization is highly variable. This in turn diminishes the effect of the perturbation by distributing the repressive input over a wider time interval. The decreased oscillatory capability associated with an increase in σ_Γ can be counteracted by an increased responsiveness to a given repressive input through a higher Hill coefficient.

The results suggest that ensembles of core-type oscillators remain relatively synchronised when subjected to large shocks and phase shift to a new circadian regime in a predictable fashion. The shell-type oscillators on the other hand may exhibit a similar degree of synchronisation when unperturbed, but we find that large enough perturbations desynchronise the individual oscillators such that circadian timekeeping of the ensemble is lost. The difference in phase shift between the two types of oscillators after a 4 hour perturbation suggests that while core-ensembles phase shift in unison, that predictability comes at a cost: the phase of the core-ensemble is less adjusted to the new circadian regime compared to the shell-ensemble. Additionally, the $(n,\ \sigma_\Gamma)$-ridge suggests that there are not two distinct sub-populations with regards to synchrony and entrainability but SCN neurons instead lie on a spectrum with a higher density of sensitive neurons in the shell and insensitive neurons in the central SCN.

6.3 Discussion

In this chapter we estimated the parameters of a spatio-temporal Bayesian hierarchical model, describing auto-repressive Cry1 transcription at the process layer. Inference was achieved using an MCMC algorithm that utilizes parallelisation for the likelihood approximation which is the dominating step of the algorithm in terms of the total computational cost. While still computationally costly, the algorithm is executable on a standard modern PC, taking approximately 300 hours to achieve an effective sample size of > 100 for all parameters for a dataset consisting of ≈ 1000 pixels and 280 time points on an Intel Core i7-6700 processor running at 3.40 GHz.

The model fit was evaluated by examining residual normality and periodicity. Locations associated with weak oscillators, i.e. adjacent tissues generally show good model fit while locations associated with strong oscillations in SCN tissue exhibit low amplitude 12 hour residual periodicity, these findings are consistent with those of Chapter 4, but the higher spatial resolution provided by the spatio-temporal model provides additional evidence that SCN tissues oscillate with 12 hour rhythms that are not consistent with a simple autoregulatory feedback loop. A likely explanation is that low amplitude, oscillating expression of Cry1 mRNA has recently been observed in astrocytes which is antiphasic to the SCN neurons, as described by Brancaccio et al. (2017).

The output allowed us to study spatial variation in parameter estimates and conclude that Cry1 mRNA population sizes are relatively uniform across the SCN. Our measure of oscillatory robustness clearly identifies SCN neurons from surrounding tissues and the additional spatial resolution provided here gives evidence that the entire (intact)

circuit of SCN neurons should be modelled as limit cycle oscillators, in the sense of the dynamics of the macroscopic mean of Cry1 levels.

We show how to obtain the posterior distribution of the period of oscillations from the MCMC output when stability properties of the model are known. While a computationally costly approach to period estimation, it comes with the benefit of uncertainty quantification while allowing separate modelling of the measurement error process. The Bayesian period estimation approach is compared to spectrum resampling (Costa et al., 2013) which also offers uncertainty quantification through bootstrap confidence intervals. Spectral methods rely on high frequency sampling to provide a high temporal resolution while the our approach does not. This is relevant for circadian data which is typically available at sampling frequencies of 1 or 2 observations per hour. We identify spatial differences in periodicity across SCN tissues, specifically that the central neurons oscillate with a lower period (< 24 hours) compared to the shell (> 24 hours). As the Bayesian approach estimates the period of the macroscopic rate equation, this finding suggests that intrinsic noise, which is found to have a period lengthening effect by Li et al. (2020), may play a differential role in the generation of rhythms across the SCN circuit.

A novel finding is also that centrally located neurons have a higher delay dispersion, i.e. the timing of the feedback loop, from transcription of mRNA to binding and inhibition of the regulatory dimer has a higher variability. We show through a simulation study that the increased stochasticity of intermediate species may in certain situations produce oscillators with a higher degree of ensemble synchrony. This phenomenon is given a mechanistic interpretation in terms of the inhibition profile.

A simulation study was designed to investigate ensemble properties of model output. Trajectories of two prototypical oscillators from either end of the (n, σ_Γ)-ridge were forward-simulated for seven circadian cycles. As the model does not describe intercellular communication, synchronised synthetic data was obtained by a common path of initial values along with common perturbations of varying duration and synchrony was defined as the ensemble regularity of phases. The results give evidence of additional functional differences between the shell and core SCN: perturbations disrupted synchronisation of the shell-type oscillators while central oscillators maintained relatively synchronised circadian time-keeping.

The findings were consistent across three experimental replicates, and with previously obtained estimates from Chapter 4 using a spatially independent model, suggesting that the results are characteristic of intact organotypic slice cultures of SCN neurons, and thereby constitute empirical input parameters of a model of auto-regulated Cry1

transcription in the SCN, capable of generating realistic synthetic mRNA data. In the next chapter we shall apply a model incorporating calcium as a transcription factor, in addition to autoregulatory inhibition of delayed mRNA. This additional transcription factor will allow an investigation of synchronisation in a mechanistic fashion by means of calcium signalling. Studying the effect of an additional transcription factor is possible due to strong priors on the inhibition dynamics of the auto-regulatory loop and simultaneous recordings of calcium and Per2, which contains a calcium responsive element (CRE) in addition to the E-box motif common between Cry and Per.

Chapter 7

Synchronisation through a second transcription factor

This chapter extends the statistical methodology from Chapters 3 and 4 to a more general model of circadian gene transcription where activation is regulated via a transcription factor in addition to the auto-regulatory TTFL.

In Section 7.1 we present a Goodwin-type model that incorporates both auto-repression and transcriptional regulation through a second transcription factor. In the subsequent sections we adapt the statistical methodologies from Chapter 3 to the extended transcription model and propose a phenomenological state-space model for a partially observed secondary transcription factor. This allows for matching the discrete-time approximation frequency of the circadian gene and transcription factor processes, required for the likelihood approximation. An MCMC algorithm is proposed for sampling the posterior distribution of the model.

In the final section of this chapter we estimate the parameters associated with the secondary transcription factor from experimental data of simultaneously recorded Per2:luc and GCaMP-reported calcium from wild-type (WT) mouse SCN. This is achieved by equipping parameters associated with the idiosyncratic core TTFL Per2 loop with informative prior distributions. These prior distributions are obtained by fitting a simpler model, reviewed in Chapter 4, to Per2 expression in VIP-null SCN tissue, made available by Maywood et al. (2011). As VIP signalling acts through the same pathway as calcium (see Section 7.4.1), imaging data from VIP-null tissue allows us to obtain estimates for a subset of the model parameters describing the core TTFL mechanism in the absence of the VIP, calcium and CRE signalling cascade. The remaining parameters, describing activation of Per2 transcription by calcium responsive element binding

protein (CREB) are subsequently estimated from the simultaneous recording of Per2:luc and GCaMP-reported calcium.

7.1 A model of the TTFL with a secondary activation mechanism

Calderazzo (2016) derives a distributed delay SDE that assumes Hill-type dynamics for calcium-induced activation and auto-repressive regulation of circadian gene transcription. In this section we follow that derivation in a general setting of delayed auto-repression and activation through a secondary exogenous transcription factor. Our contribution consists of a rescaling of the deterministic dynamics from mRNA counts to "promoter saturation", effectively decoupling the regulatory dynamics from molecular population sizes (Section 7.1.2).

7.1.1 Macroscopic deterministic model

Let $X(t)$ denote population of mRNA of circadian gene, e.g. Per, and let $X_i(t)$ denote concentrations of intermediate species of the TTFL. Additionally, let $Z(t)$ denote the concentration of the activating transcription factor at time t and define delayed $Z(t)$ as

$$Z_\tau(t) = \int_{-\infty}^{t} Z(s) g_{a_{TF}, p_{TF}}(t-s) ds, \tag{7.1}$$

where $g_{a,p}$ denotes the density of a gamma distributed random variable with rate a and shape p. The delay distribution for the transcription factor relies on the linear chain trick (Smith, 2011) and provides flexibility for settings where $Z(t)$ does not directly activate transcription but instead represents an observed species that mediates activation through a series of reactions. These unobserved reactions and species are modelled by

$$\frac{dZ_j(t)}{dt} = a_{TF}[Z_{j-1}^{(TF)}(t) - Z_j(t)],\ j = 1, ..., p_{TF}, \tag{7.2}$$

where $Z_{p_{TF}}$ is the end-product activating species, while we wish to model activation by the "earlier" species Z_1 in the chain of reactions.

We assume that the intermediate species in the circadian mRNA TTFL can be

modelled by the following system of ODEs

$$\begin{aligned} \frac{dX(t)}{dt} &= \nu(X_p(t), Z_\tau(t)) - \beta X(t) \\ \frac{dX_i(t)}{dt} &= a(X_{i-1}(t) - X_i(t)),\ i = 1, ...p, \end{aligned} \tag{7.3}$$

where $X_0(t) \coloneqq X(t)$ and $\nu(\cdot, \cdot)$ and β are transcription and degradation rate, respectively, of the species of interest. This is identical to the description in Section 4.2, except we now let the birth-rate ν depend on the additional variable $Z_\tau(t)$. Note that $Z_\tau(t)$ is an exogenous (but time varying) variable and hence Eq. (7.3) is in fact a system of non-delayed differential equations. As before in Chapter 4, we apply the linear chain trick and write the system in Eq. (7.3) as a distributed delay differential equation of the form

$$\frac{dX(t)}{dt} = \nu(X_\tau, Z_\tau(t)) - \beta X(t), \tag{7.4}$$

where

$$X_\tau(t) = \int_{-\infty}^{t} X(s) g_{a,p}(t-s) ds. \tag{7.5}$$

The delay distribution for mRNA accounts for the same intermediate processes as in the purely auto-regulatory model used in previous chapters, i.e. translation, transport, dimerisation and nuclear import/export.

Assuming Hill-function dynamics, transcriptional activation by the transcription factor $Z_\tau(t)$ and an interaction effect between inhibition due to delayed mRNA and activation by $Z_\tau(t)$, Calderazzo (2016) derives a macroscopic rate equation for circadian mRNA transcription, given by

$$\frac{dX(t)}{dt} = \frac{R_0 + R_{TF}\left(\frac{Z_\tau(t)}{K_{TF}}\right)^{n_{TF}} - R_{int.}\left(\frac{Z_\tau(t)}{K_{TF}}\right)^{n_{TF}}\left(\frac{X_\tau(t)}{K}\right)^{n}}{1 + \left(\frac{X_\tau(t)}{K}\right)^{n} + \left(\frac{Z_\tau(t)}{K_{TF}}\right)^{n_{TF}} + \left(\frac{X_\tau(t)}{K}\right)^{n}\left(\frac{Z_\tau(t)}{K_{TF}}\right)^{n_{TF}}} - \beta X(t). \tag{7.6}$$

It is easy to see that for $Z_\tau(t) \coloneqq 0$ we recover the single transcription factor model used in the previous chapters. Hence, R_0 molecules per unit time is the maximum transcription rate of mRNA in the absence of activation through $Z_\tau(t)$ and K is the count of delayed mRNA at which that transcription rate is at 50 percent of the maximum rate. Letting $X_\tau(t) \coloneqq 0$ and $Z_\tau(t) \to \infty$ reveals that R_{TF} is the "partial" maximum activation due to $Z_\tau(t)$ and for $Z_\tau(t) = K_{TF}$ activation is at 50 percent. In addition, the model captures a non-linear interaction effect of transcriptional activation/inhibition due to the delayed species through $R_{int.}$. Letting $Z_\tau(t) \to \infty$ for some fixed $X_\tau(t)$ reveals that the maximum

activation due to $Z_\tau(t)$ is given by $R_{TF} - R_{int.}(X_\tau(t)/K)^n$. Idiosyncratic non-linearity is captured by factor specific Hill coefficients n_{TF} and n that parametrise the cooperativity of the binding sites for the activating and inhibiting species, respectively.

7.1.2 Rate rescaling and decoupling of population size

Parameters R_0 and K are specified in terms of circadian gene mRNA per unit time while K_{TF} is specified in terms of $Z_\tau(t)$ molecules. Unless molecular population sizes of the exogenous transcription factor are known, which is typically not the case, K_{TF} is not directly identifiable. However, we note that $Z_\tau(t)$ only enters the transcription function through the fraction $Z_\tau(t)/K_{TF}$. Under the reasonable assumption that $cZ_\tau(t)$ is known for some unknown constant c and unknown population size $Z_\tau(t)$, e.g. the case where $Z_\tau(t)$ is measured by a proportional light signal such as in bioluminescence imaging techniques, it is helpful to rewrite the relevant fractions as

$$\left(\frac{Z_\tau(t)}{K_{TF}}\right) = \left(\frac{cZ_\tau(t)}{cK_{TF}}\right) = \left(\frac{cZ_\tau(t)}{K^*_{TF}}\right). \tag{7.7}$$

The interpretation of K^*_{TF} is the level, in arbitrary units, of $cZ_\tau(t)$ at which the exogenous activation of circadian gene mRNA transcription is at 50 percent of maximum.

We shall now consider an additional rescaling of the deterministic dynamics given in Eq. (7.6) from molecular population size to "promoter saturation", which decouples the dynamics from population size and hence reduces the number of parameters. By defining

$$X'(t) = \frac{X(t)}{R_0}, \tag{7.8}$$

we obtain the following delay differential equation for the evolution of X',

$$\frac{dX'(t)}{dt} = \frac{1 + R'_{TF}\left(\frac{cZ_\tau(t)}{K^*_{TF}}\right)^{n_{TF}} - R'_{int.}\left(\frac{cZ_\tau(t)}{K^*_{TF}}\right)^{n_{TF}}\left(\frac{X'_\tau}{K'}\right)^n}{1 + \left(\frac{X'_\tau}{K'}\right)^n + \left(\frac{cZ_\tau(t)}{K^*_{TF}}\right)^{n_{TF}} + \left(\frac{X'_\tau}{K'}\right)^n\left(\frac{cZ_\tau(t)}{K^*_{TF}}\right)^{n_{TF}}} - \beta X'(t). \tag{7.9}$$

where $R'_{TF} = R_{TF}/R_0$ and $R'_{int.} = R_{int.}/R_0$ have the interpretation of exogenous activation and interaction relative to the baseline maximum, respectively. Similarly, $K' = K/R_0$ is not expressed in terms of molecular population but instead is interpreted as the level at which repression is 50 percent, relative to the non-inhibited transcription rate. Furthermore, Hill coefficients n and n_{TF}, degradation rate β and delay distribution parameters are unaffected by the rescaling and hence independent of population sizes in both Eq. (7.6) and (7.9).

7.1.3 CLE approximation

To account for stochasticity due to low numbers of mRNA molecules we assume a birth-death process as in Calderazzo (2016). The delay differential equation in Eq. (7.6) implies the following set of reactions

$$\begin{aligned} \emptyset &\xrightarrow{\nu(X_\tau, Z_\tau)} X \\ X &\xrightarrow{\quad\beta\quad} \emptyset. \end{aligned} \tag{7.10}$$

In the first reaction circadian mRNA is synthesized from abundant reactants with rate given by the first term of the right hand side of Eq. (7.6). The second reaction describes linear degradation of mRNA with rate β. As discussed in Chapter 3, and by additionally conditioning on the exogenous transcription factor, i.e. treating $Z_\tau(t)$ as a time-varying *parameter* of the system, we can approximate the associated Markov Jump process by the CLE to obtain an equation on the form

$$dX(t) = (\nu(X_\tau(t), Z_\tau(t)) - \beta X(t))\, dt + \sqrt{\nu(X_\tau(t), Z_\tau(t)) + \beta X(t)} dW(t), \tag{7.11}$$

where $W(t)$ is a standard Wiener process. By applying the rate rescaling from Section 7.1.2 we can decouple the drift function from population size and obtain the following delayed SDE

$$\begin{aligned} dX'(t) = &\left(\frac{1 + R'_{TF} \left(\frac{cZ_\tau(t)}{K^*_{TF}}\right)^{n_{TF}} - R'_{int.} \left(\frac{cZ_\tau(t)}{K^*_{TF}}\right)^{n_{TF}} \left(\frac{X'_\tau}{K'}\right)^{n}}{1 + \left(\frac{X'_\tau}{K'}\right)^{n} + \left(\frac{cZ_\tau(t)}{K^*_{TF}}\right)^{n_{TF}} + \left(\frac{X'_\tau}{K'}\right)^{n} \left(\frac{cZ_\tau(t)}{K^*_{TF}}\right)^{n_{TF}}} - \beta X'(t) \right) dt \\ &+ \sqrt{\gamma \left(\frac{1 + R'_{TF} \left(\frac{cZ_\tau(t)}{K^*_{C}}\right)^{n_{TF}} - R'_{int.} \left(\frac{cZ_\tau(t)}{K^*_{TF}}\right)^{n_{TF}} \left(\frac{X'_\tau}{K'}\right)^{n}}{1 + \left(\frac{X'_\tau}{K'}\right)^{n} + \left(\frac{cZ_\tau(t)}{K^*_{TF}}\right)^{n_{TF}} + \left(\frac{X'_\tau}{K'}\right)^{n} \left(\frac{cZ_\tau(t)}{K^*_{TF}}\right)^{n_{TF}}} + \beta X'(t) \right)} dW(t) \end{aligned} \tag{7.12}$$

where $\gamma = R_0^{-1}$ is a dispersion parameter that links the level of stochasticity at the population level of mRNA to the level of promoter saturation $X'(t)$.

7.1.4 Measurement model and Extended Kalman-Bucy filter for the TTFL model with secondary activation

In addition to the model in Eq. (7.12) we need a model that accounts for measurement noise and time-aggregation of the underlying concentrations. As before (see Section

3.4.1), we assume that the observations, available at $t = \Delta t, ..., T$, are formed by scaled temporal aggregates of underlying states, corrupted with additive Gaussian measurement noise η_t with time-homogeneous variance σ_η^2 to represent camera exposure time and measurement noise, i.e.

$$Y_t = \kappa \int_{t-\Delta t}^{t} X'(s)ds + \eta_t. \tag{7.13}$$

The EKBF for delayed systems, developed by Calderazzo et al. (2018) and reviewed in Chapter 3, Section 3.4.1 can be adapted to approximate the likelihood of the extended transcription model in Eq. (7.12) along with measurement equation in Eq. (7.13). We treat the second transcription factor as exogenous and hence the likelihood is given by the probability of observed data under a given value of the parameter set $\Psi = [R_0, R_{TF}, ..., \beta]$ *and* a fixed sequence of the activating transcription factor $Z(t), t \in [0, T]$, interpreted as a function of the parameters, i.e. $\mathcal{L}(\Psi; Y_{\Delta t:T}, Z_{0:T}) = P(Y_{\Delta t:T}|\Psi,\ Z_{0:T})$. In the next section we propose a stochastic model to describe the time evolution and measurement process of the second transcription factor, but for now the likelihood is conditioned on an available sequence of measurements of the underlying concentration of the activating transcription factor, up to a multiplicative constant as shown earlier by the rate rescaling.

To approximate the likelihood, we assume a common discretisation of the underlying circadian mRNA $X_{0:T} = [X_{\delta t},\ X_{2\delta t},\ ...\ , X_T]$ and activating species $Z_{0:T} = [Z_{\delta t},\ Z_{2\delta t},\ ...\ , Z_T]$. The delay distributions for both of the species are truncated at a maximum delay time $\tau_{max} = 24$, which implies that the effects of the delayed species are fully exhausted within a full circadian cycle. The mean and variance of the discretised CLE are linearised by Taylor expanding around a deterministic mean ρ_t. These quantities are propagated forward by δt using Eq. (3.24) and subsequently updated using the Kalman update in Eq. (3.28). The steps are given explicitly in Appendix B. A Gaussian likelihood follows from assuming $X_0 \sim \mathcal{N}(\rho_0, P_0)$, approximate linearity of the process on subintervals of length δt, linearity of the measurement equation and additive Gaussian measurement errors.

7.2 A harmonic dynamic linear model for a partially observed exogenous transcription factor

In practice, it is often the case that recordings of an additional transcription factor consist of time-averaged measurements corrupted by some measurement noise process. If the aim is parameter estimation, we first require some methodology to infer the underlying

states of the transcription factor, given the noisy measurements. The reason for state inference is twofold: firstly, the measurement process of the transcription factor can cause biases in the estimation of the transcription model if not appropriately modelled. For example, conditioning inference on a single realisation of noisy measurements will result in underestimation of the posterior dispersion. Secondly, data for the transcription factor and the circadian gene expression might be available at discrete time points with interval length dictated by the experiment, and the EKBF likelihood approximation requires evaluation of the circadian mRNA and transcription factor processes at some discrete time points δt, $t \in [1, T]$ with common interval length δ, where δt is smaller than the time interval Δt, at which measurements are taken. As the primary interest in the exogenous transcription factor lies in reconstructing the states we develop a harmonic dynamic linear model for $Z(t)$ which can be solved by Kalman filtering and smoothing.

Petris et al. (2009) gives a thorough treatment of Fourier form dynamic linear models, which is followed here. Assume some arbitrary time discretisation of the underlying concentration $Z(t)$, represented by some periodic function

$$Z_t = \sum_{j=1}^{s/2} S_j(t), \tag{7.14}$$

where s is the period. The j^{th} harmonic and conjugate harmonic are given by

$$\begin{aligned} S_j(t) &= a_j \cos(t\omega_j) + b_j \sin(t\omega_j) \\ S_j^*(t) &= -a_j \cos(t\omega_j) + b_j \sin(t\omega_j) \end{aligned} \tag{7.15}$$

for Fourier frequencies

$$\omega_j = \frac{2\pi j}{s}, \;\; j = 1, ...s/2. \tag{7.16}$$

Eq. (7.15) can be shown to have the following iterative form

$$\begin{aligned} S_j(t+1) &= S_j(t)\cos(\omega_j) + S_j^*(t)\sin(\omega_j) \\ S_j^*(t+1) &= -S_j(t)\cos(\omega_j) + S_j^*(t)\sin(\omega_j) \end{aligned} \tag{7.17}$$

with

$$S_j(0) = a_j, \quad S_j^*(0) = b_j \tag{7.18}$$

which can be written in matrix form as

$$\begin{bmatrix} S_j(t+1) \\ S_j^*(t+1) \end{bmatrix} = \begin{bmatrix} \cos(\omega_j) & \sin(\omega_j) \\ -\sin(\omega_j) & \cos(\omega_j) \end{bmatrix} \begin{bmatrix} S_j(t) \\ S_j^*(t) \end{bmatrix}, \ t = 0, ..., T, \ j = 1, ..., \frac{s}{2}. \tag{7.19}$$

For modelling purposes it is generally the case that fewer than $s/2$ harmonic frequencies are needed to give an adequate summary of the signal. Let n denote the number of Fourier frequencies to retain, corresponding to a truncation of the sum in Eq. (7.14), so that terms $j = 1, ..., n$ are included. To model this collection of harmonics we collect the components representing the state variables in a $2n$-dimensional vector θ_t,

$$\theta_t = \begin{pmatrix} S_1(t) \\ S_1^*(t) \\ \vdots \\ S_n(t) \\ S_n^*(t) \end{pmatrix}. \tag{7.20}$$

The evolution of the n harmonic components is then given by

$$\theta_{t+1} = H\theta_t + \epsilon_t, \ \epsilon_t \sim \mathcal{N}(0, Q) \tag{7.21}$$

where H is a $2n$ by $2n$ block diagonal matrix reproducing Eq. (7.19) for each of the frequencies included and ϵ_t is a Gaussian noise term.

As the transcription factor may be measured in a similar fashion to circadian mRNA, i.e. using a reporter protein whose emitted light is recorded over a time interval, measurements constitute time-aggregates of the underlying states. To model the experimental procedure we construct a measurement equation for observed signal at time t. Let $k = \Delta/\delta$ denote the number of states of the discretised underlying concentration per observation time interval, and

$$Z_t^{(\text{obs})} = \frac{1}{k} \sum_{i=t-k+1}^{t} Z_i + \upsilon_t = F\Theta_t + \upsilon_t, \tag{7.22}$$

where $\upsilon_t \sim \mathcal{N}(0, R)$ represents measurement noise,

$$\Theta_t = \begin{pmatrix} \theta_t \\ \theta_{t-1} \\ \vdots \\ \theta_{t-k+1} \end{pmatrix} \tag{7.23}$$

and

$$F = \begin{pmatrix} \frac{1}{k} & 0 & \frac{1}{k} & \dots & \frac{1}{k} & 0 & \frac{1}{k} \end{pmatrix}. \tag{7.24}$$

Eq. (7.21-7.22) can be combined to obtain the following state space model

$$\begin{aligned} Z_t^{(\text{obs})} &= F\Theta_t + \upsilon_t, \;\; \upsilon_t \sim \mathcal{N}(0, R) \\ \Theta_{t+1} &= \mathbf{H}\Theta_t + \boldsymbol{\epsilon}_t, \;\; \boldsymbol{\epsilon}_t \sim \mathcal{N}(0, \boldsymbol{Q}) \end{aligned} \tag{7.25}$$

where $\boldsymbol{\epsilon}_t = [\epsilon_t \, 0 \, ... \, 0]^\intercal$ and the matrix $\mathbf{H}$ is equal to H in the first $2n$ rows and columns and subsequent entries make up a shifted identity matrix which reproduces $\theta_j = \theta_j$ for $j = t-1, ..., t-k+1$. Hence, the distribution of $\boldsymbol{\epsilon}_t$ is degenerate with $\boldsymbol{Q}$ having at most $2n$ non-zero diagonal entries. Note that including $\boldsymbol{\epsilon}_t$ (and ϵ_t) means that the states are not strictly periodic as Eq. (7.14) does not hold for all t. As the state space model in Eq. (7.25) is linear and Gaussian it is straightforward to use a Kalman filter and smoother to perform inference for and generate draws from the distribution of the states given the observations (see e.g. Durbin and Koopman (2012)). An explicit algorithm is given in Appendix B.

7.3 An MCMC algorithm for sampling the posterior distribution of the extended transcription model

For the extended TTFL model we again resort to designing an adaptive random walk Metropolis algorithm, given in Algorithm (5), to generate samples from the posterior distribution of parameters Ψ, governing the dynamics of mRNA transcription. To account for partial observations of the activating species we propose an empirical hierarchical approach, where the parameters of the harmonic dynamic linear model describing the activating species, i.e. the number of harmonics, their coefficients a_j and b_j and the variance of the measurement noise process R may be treated as fixed hyper-parameters and estimated a priori using e.g. maximum likelihood. Inference is subsequently conditioned on the resulting mean estimates, $\mathbb{E}(\theta_{\delta t:T} | Z^{obs}_{\Delta t:T})$, calculated using the Kalman

smoother in Appendix B, at the time points dictated by the choice of discretisation used for the likelihood approximation of the transcription model. It should be noted that the likelihood evaluation requires a partial reconstruction of the unobserved species using Eq. (7.20), and Eq. (7.14) truncated at the appropriate number of harmonics.

Algorithm 5 Random walk Metropolis algorithm to sample posterior distribution of TTFL model with additional activating TF

Initialise iteration counter $k = 1$, parameters $\Psi^{(1)}$, proposal variance scaling $\gamma^{(1)}$ and transcription factor $\tilde{\theta}_{\delta t:T} = \mathbb{E}(\theta_{\delta t:T}|Z^{obs}_{\Delta t:T})$
while $k <$ max iteration **do**
  Set $\Psi_k = \Psi_{k-1}$
  Draw $\hat{\Psi}^{(k+1)} \sim \mathcal{N}(\Psi^{(k)}, \gamma_{\Psi}^{(k)}\hat{\Sigma}_{\Psi})$ and let $\Psi^{(k+1)} = \hat{\Psi}^{(k+1)}$ w.p. $\min\left\{1, \frac{\mathcal{L}(\hat{\Psi}^{(k+1)};Y_{\Delta t:T},\tilde{\theta}_{\delta t:T})\pi(\hat{\Psi}^{(k+1)})}{\mathcal{L}(\Psi^{(k)};Y_{\Delta t:T},\tilde{\theta}_{\delta t:T})\pi(\Psi^{(k)})}\right\}$
  if The proportion of accepted proposals in iterations $1:k$ is > 0.234 **then**
    Set $\gamma^{(k+1)} = \exp\{\log\gamma^{(k)} + \min(c, k^{-1})\}$
  else
    Set $\gamma^{(k+1)} = \exp\{\log\gamma^{(k)} - \min(c, k^{-1})\}$
  end if
  Set $k \coloneqq k + 1$
end while

7.4 Application to Per2:luc and GCaMP-reported Calcium imaging data

We shall now return to the data introduced in Chapter 2, obtained by Brancaccio et al. (2013) using simultaneous recording of Per2:luc and GCaMP-reported calcium in organotypic wild type (WT) SCN tissue. A cropped and sub-sampled frame of recorded Per2:luc expression is shown in Figure 7.1, along with 15 locations across the right half of the SCN at which we shall fit the extended TTFL model. The small number of locations is motivated by balancing the computational cost of the analysis and providing a proof of concept of the methodology. For representative time series of Per2:luc and calcium the reader is referred to Figures 2.3 and 2.4 in Chapter 2 where the data was initially introduced and analysed.

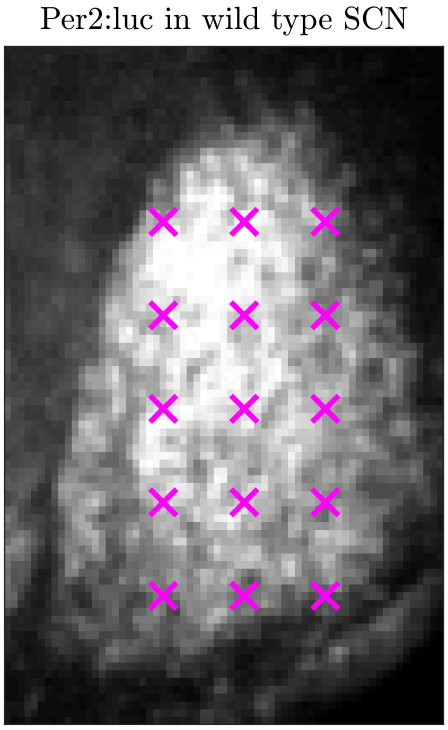

Figure 7.1: Per2:luc expression in organotypic, wild type SCN with simultaneous GCaMP-calcium recording (Brancaccio et al., 2013). Inference is done at 15 locations, each consisting of 2 by 2 pixel blocks across the right half of the organ and marked in magenta.

7.4.1 Calcium induced transcriptional activation of Per2

Circadian expression of Per2 is only partly driven by the intercellular TTFL. Rhythmic firing of neurons, which peaks during the subjective day (Cutler et al., 2003), also drive intercellular coupling and entrainment to photic stimuli entering the dorsal SCN via the optic chiasma (Herzog et al., 2017). The neuropeptide VIP is only expressed by these ventral SCN cells but binds to the VPAC2 receptor present in all SCN neurons (An et al., 2012). This binding activates Gq-signalling, increasing calcium levels and subsequently promoting Per2 transcription via calcium responsive element binding protein (CREB) (Hastings et al., 2014). Brancaccio et al. (2013) show experimentally that VIP knock-out SCN cultures lack synchrony of oscillations of Calcium and CREB, and blocking calcium

influx abolishes rhythmic expression of Per2:luc (Lundkvist et al., 2005). Maywood et al. (2011) make available recordings of genetically modified SCN cultures that do not express VIP (VIP-null) where the Per2::luc reporter exhibits damped and desynchronised expression.

Availability of Per2::luc expression from VIP-null SCN cultures allows us to observe the core circadian, auto-regulatory feedback loop of Per2. To formulate informative prior distributions for the extended model we shall fit the simpler model from Chapter 4 to the Per2::luc expression data using the hierarchical framework developed in Chapter 6. The resulting estimates, along with the formulation of prior distributions for the extended model are summarised in Section 7.4.3

The role of calcium signalling as a mediator of inter-cellular communication confers a spatial interpretation of the model. While we do not parametrise or estimate the network topology, calcium-induced transcription may be viewed as a coupling mechanism between individual oscillators for some assumed topology. A thorough investigation of such an interpretation is outside the scope of this thesis, however some further discussion on the topic is presented in the Further research section of the Conclusions.

7.4.2 Simplifying assumptions and parameter restrictions

While the informative prior distributions derived from the analysis of Per2 expression in VIP-null SCN tissue are necessary, we shall impose some additional restrictions to simplify inference. We choose to incorporate calcium by treating the observed GCaMP-reported calcium as the underlying transcription factor in the model. In practice, this amounts to omitting maximum likelihood estimation of the parameters of the model in Eq. (7.25) and instead evaluating the likelihood approximation at observed calcium. The effects of this additional approximation are likely to be small as the activating species in Eq. (7.12) enters through a convolution with a delay distribution, producing a weighted average of past levels of the species. We deem that the temporal smoothing of the signal due to the delay distribution sufficiently mitigates both the effects of measurement noise and temporal smoothing due to the camera exposure time on parameter inference. The proposed empirical hierarchical model provides flexibility for different experimental protocols and choices of discretisation, however, in our application the data consist of simultaneous 0.5h measurements of the two species and a finer discretisation than 0.5h makes the likelihood approximation prohibitively expensive. Hence the interpolating mechanism of the state-space model for the additional transcription factor is not required.

Another parameter restriction related to the measurement process is imposed on the measurement error dispersion. As the Per2:luc signal is available in two of the three RGB channels realisations of the measurement error process are available as the scaled difference of the two channels. The SD is held fixed at its pixel-wise ML estimate (see Section 4.3.1 for formal procedure to obtain these estimates). For the delay distribution which accounts for time delays between CREB and observed calcium, which serves as a proxy, we assume an exponential distribution (which is a special case of the Gamma distribution, obtained by restricting the shape parameter to 1).

The dissociation coefficient for delayed Per2, K, is held fixed at a level corresponding to the spatial mean across VIP-null SCN tisue. This restriction is motivated by the difficulty in inferring molecular population sizes. In the analysis of Cry1-luc in Chapters 4 and 6, this was circumvented by running very long MCMC chains and pooling statistical strength across tissues using the hierarchical model from Chapter 5 to explore the posterior distribution. Inference for additional parameters associated with a second transcription factor significantly exacerbates the problem hence the restriction on K is deemed necessary.

Finally, a restriction on the local variation of parameters is imposed. This involves inferring the parameters for a 2 by 2 pixel block simultaneously and restricting the parameters to be identical within the block. The motivation is that in previous analyses of Cry1-luc the spatial distribution of parameter estimates was found to be fairly smooth, apart from at the border between the SCN and adjacent tissues. In other words, we expect spatial parameter variation to lie on a scale that is far greater than individual pixels (and neurons if approximate correspondence between pixels and neurons is achieved) and neighbouring pixels to have very similar parameter values. This assumption adds curvature to the likelihood as the amount of data is quadrupled and neighbouring pixels have very similar signal for both Per2:luc and calcium. While the likelihood evaluation is computationally expensive, a parallel implementation of the joint likelihood over the 2 by 2 block of pixels is straightforward as the assumption of independence of data conditioned on the model parameters allows for pixel-wise factoring of the likelihood, i.e. the joint likelihood of the pixel block is equal to the product of the likelihood evaluated at the individual pixels.

7.4.3 Prior distributions

As parameter identification is a considerable challenge in inference for the TTFL model with an additional transcription factor in Eq. (7.12), investigation of the transcriptional

activation of intracellular calcium on Per2 requires the formulation of informative prior distributions on a subset of the parameters.

The aim here is to be able to infer parameters governing the calcium induced activation of Per2 transcription by placing informative prior distributions on the parameters of the auto-regulatory Per2 feedback loop while placing diffuse priors on the unknown parameters related to transcriptional activation by calcium. To elicit informative prior distributions on the core TTFL, the hierarchical spatio-temporal framework developed in Chapter 5 is used to estimate the parameters of a model containing just the simple auto-regulation, i.e. the model in Eq. (4.5) on bio-imaging data of Per2::luc expression in VIP-null SCN tissue obtained by Maywood et al. (2011). Parameter estimation is achieved using the RWM algorithm described in Section 5.2.1 with 80k iterations and the same de-trending and prior distributions as for the Cry1-luc expression data in Chapter 6. For the VIP-null data no pixel aggregation is done as the resolution of raw pixels is considerably lower compared to the Cry1-luc data, however the video is sub-sampled such that only every fourth pixel column and row are analysed. Additionally, the video is cropped to include only the right half of the SCN and shortened by excluding the first three days of recording due to overexposure that causes a truncation of the data. The preprocessing results in 31 by 17 pixels and 146 frames corresponding to approximately 6 days (see Figure 7.2).

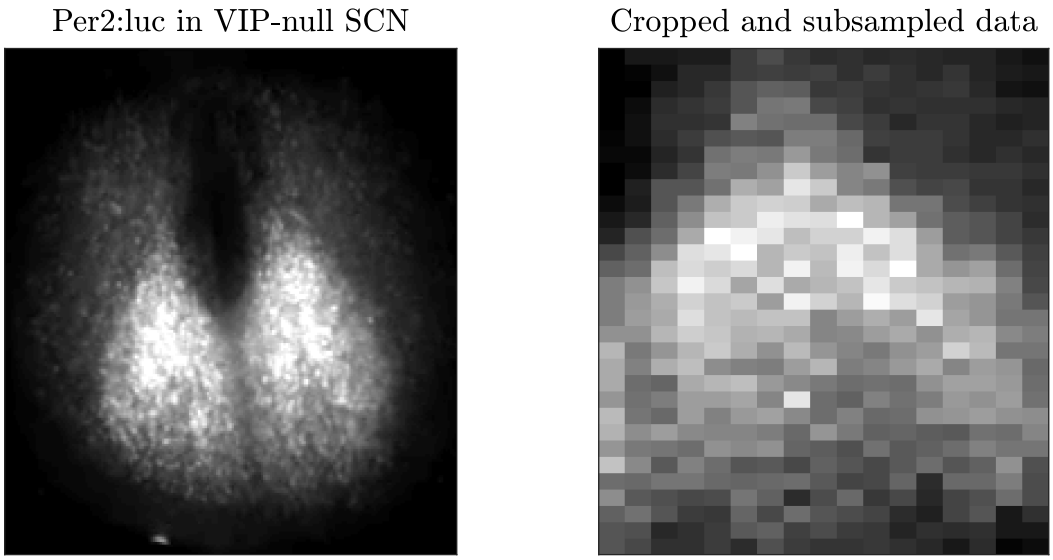

Figure 7.2: Raw and cropped frame of Per2::luc expression in VIP-null SCN obtained by Maywood et al. (2011). Right image contains full extent of the cross-section of right half of the SCN.

The cropped video contains mostly SCN tissue and captures the full extent of a section across the right half of the SCN. The informative prior distributions for parameters governing the auto-regulatory inhibition of Per2 and the degradation rate, i.e. R_0, K_p, n, μ_Γ, σ_Γ and β, are obtained by fitting a Gaussian distribution to the posterior of the logarithm of each of the parameters, i.e. across the MCMC chains *and* pixels. The resulting distributions are summarised in Table 7.1 and presented in Figure 7.3, while the diffuse prior distributions associated with calcium induced transcriptional activation are summarised in Table 7.2.

Table 7.1: Mean and standard deviation of Gaussian prior distributions for the auto-regulatory Per2 feedback loop and light scaling parameter. Informative prior distributions are obtained by fitting a Gaussian to the posterior parameter estimates obtained from VIP-null Per2::luc data. Only the mean of the fitted distribution for $\log K$ is used for further inference while the SD is set to 0, motivated by the difficulty associated with inferring molecular population sizes.

Parameter	$\log \mu_{\Gamma^{(p)}}$	$\log \sigma_{\Gamma^{(p)}}$	$\log R_0$	$\log K$	$\log n$	$\log \beta$	$\log \kappa$
Prior mean	2.23	0.97	4.50	4.78	0.69	-1.31	0
Prior SD	0.07	0.28	0.41	0.37*	0.41	0.28	10

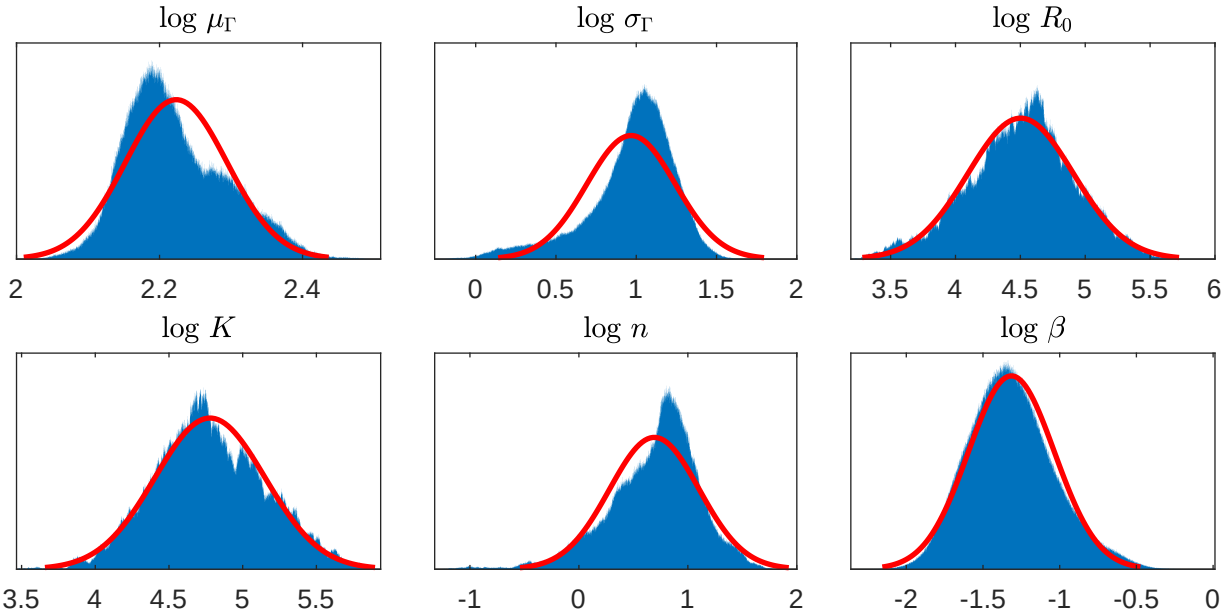

Figure 7.3: Empirical distributions of Hill-function parameters and degradation rate from VIP-null SCN, along with fitted Gaussian distributions. Estimates are obtained from posterior distributions obtained by fitting a model of simple auto-repressive gene regulation to Per2::luc expression from organotypic VIP-null SCN tissue.

Table 7.2: Mean and standard deviation of diffuse Gaussian (lower and upper bound of uniform) prior distributions for parameters governing the calcium induced transcriptional activation.

Parameter	$\log \mu_{\Gamma^{(c)}}$	$\log R_{TF}$	$\log K_{TF}$	$\log n_{TF}$	$\log R_{int}$
Prior mean (or lower bound)	(0)	0	0	0	0
Prior SD (or upper bound)	(12)	6	3	5	6

7.4.4 Results for the full model with an exogenous activation mechanism

The MCMC algorithm described in Section 7.3 is run for 10k iterations at each of the marked locations in Figure 7.1. The chains are initialised at the prior means for parameters with informative prior distributions and at regions of high posterior density for the remaining parameters, obtained by pilot runs of the algorithm at a single location in the middle of the biological sample. The initial 500 iterations are discarded as burn-in and the remaining 9500 iterations used for analysis.

We first examine the parameter trace plots and posterior distributions of log-parameters compared to prior distributions from a randomly chosen location corresponding to the second row, first column of marked locations in Figure 7.1. Results for the parameters of the auto-regulatory Hill-function (partial maximum transcription rate R_0, Hill-coefficient n, and mean and standard deviation of the delay distribution approximating the Per2 loop, $\mu_\Gamma^{(p)}$ and $\sigma_\Gamma^{(p)}$) are given in Figure 7.4. Results for parameters corresponding to calcium activation (activation rate R_{TF}, dissociation coefficient and Hill coefficient of the approximated transcription factor K_{TF} and n_{TF}, mean of the exponential delay distribution $\mu_\Gamma^{(TF)}$ accounting for delay between GCaMP-reported calcium and CRE binding, and interaction rate $R_{int.}$) are presented in Figure 7.5 while light scaling κ and degradation rate β are treated in Figure 7.6.

For the parameters of the Per2 feedback loop, equipped with informative prior distributions, there is still considerable additional concentration of the posterior. As the prior distributions encompass spatial variation of parameters across the VIP-null tissue while the posterior distributions are obtained by analysing a 2 by 2 pixel block in the WT SCN, some such concentration can be expected. For the maximum transcription rate R_0, concentration is likely, at least in part, due to fixing the dissociation coefficient K at the spatial mean estimate obtained from the VIP-null data. The posterior distribution

of the Hill coefficient n also shows considerable additional concentration compared to the prior and the posterior mean of $\log n$ is 1.29, higher than the prior mean of 0.69, suggesting a greater degree of non-linearity of the auto-regulatory loop in the WT SCN compared to the VIP-null. The mean of the auto-regulatory delay distribution exhibits concentration around levels higher than the prior mean while the standard deviation of the delay distribution closely follows the prior distribution. The chains, as examined by the trace plots exhibit relatively rapid convergence and good mixing.

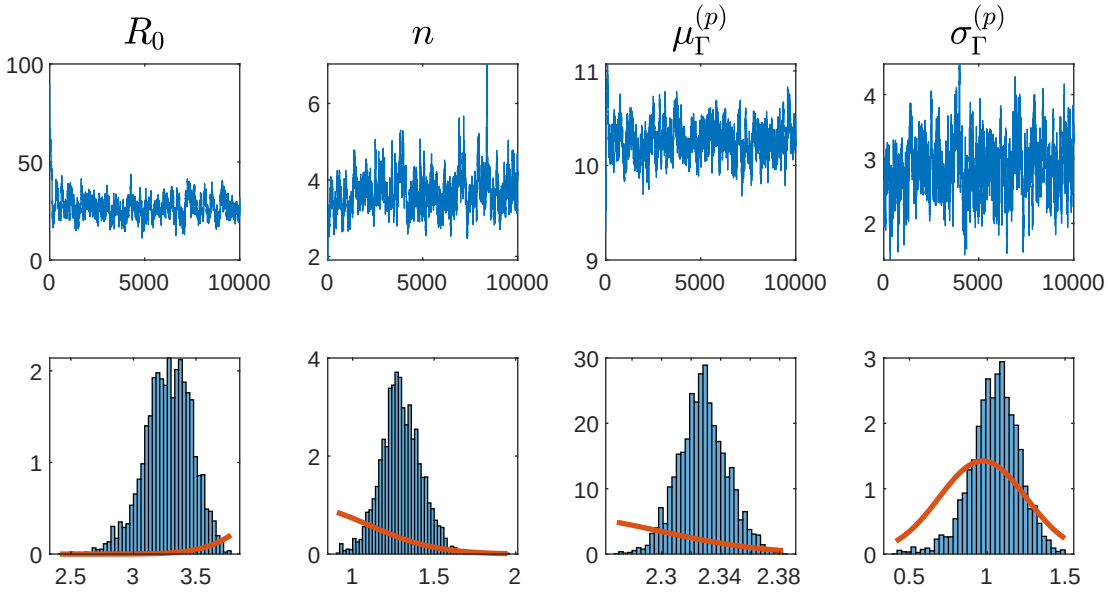

Figure 7.4: Results for one location corresponding to the second row, first column of marked locations in Figure 7.1. **Top row:** MCMC chains of parameters associated with the auto-regulatory Per2 loop for 10k MCMC iterations. The trace plots suggest rapid convergence and good mixing (low autocorrelation). **Bottom row:** Posterior distributions of log parameters and prior distribution over the support of the posterior. The posterior distributions are unimodal and exhibit concentration compared to the prior distributions.

The chains of parameters associated with calcium activation also show rapid convergence and good mixing. Interestingly, the posterior distribution of the interaction rate between delayed Per2 and calcium activation is consistent with $R_{int.} = 0$, i.e. that the degree of activation associated with a given calcium concentration is constant for varying levels of delayed Per2. The dissociation coefficient of the additional transcription factor, K_{TF}

has a posterior mean of 0.42, which coincides with the temporal mean of the GCaMP-calcium signal. This is indicative of a "symmetry" in transcriptional induction, as the degree of transcriptional activation is at 50 percent of maximum at the temporal mean level of calcium.

The posterior of the degradation rate β and light scaling constant κ have means of 0.29 and 3.1×10^{-3}, respectively, which is consistent with previous estimates of Per2:luc in VIP-null SCN and Cry1-luc in WT SCN from Chapters 4 and 6.

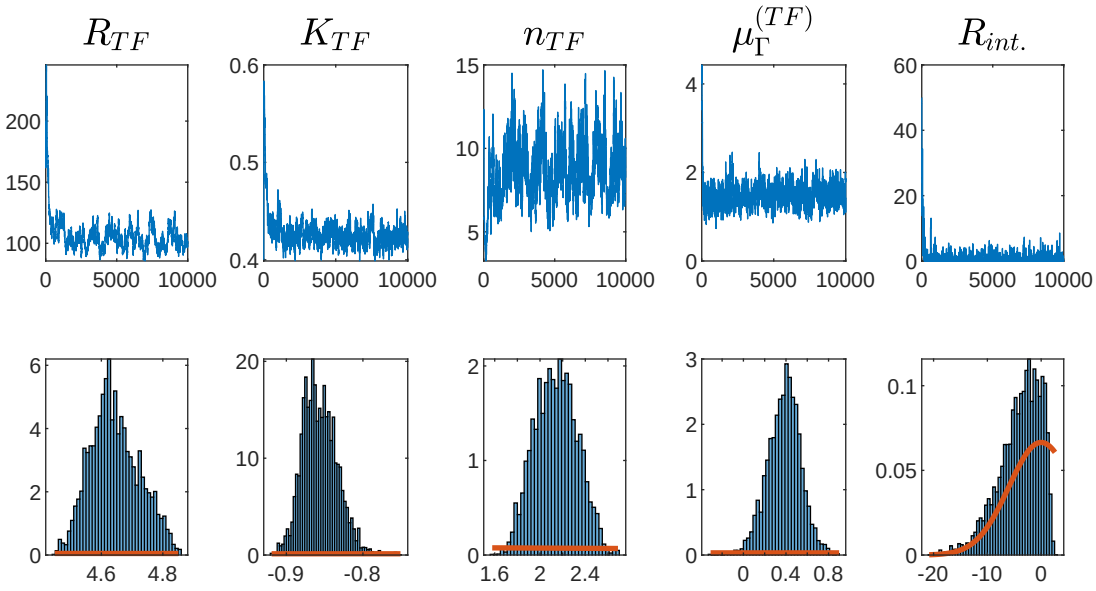

Figure 7.5: Results for one location corresponding to the second row, first column of marked locations in Figure 7.1. **Top row:** MCMC chains of parameters associated with calcium-induced activation of Per2 transcription over 10k MCMC iterations. The trace plots suggest rapid convergence and good mixing (low autocorrelation). **Bottom row:** Posterior distributions of log parameters and prior distribution over the support of the posterior.

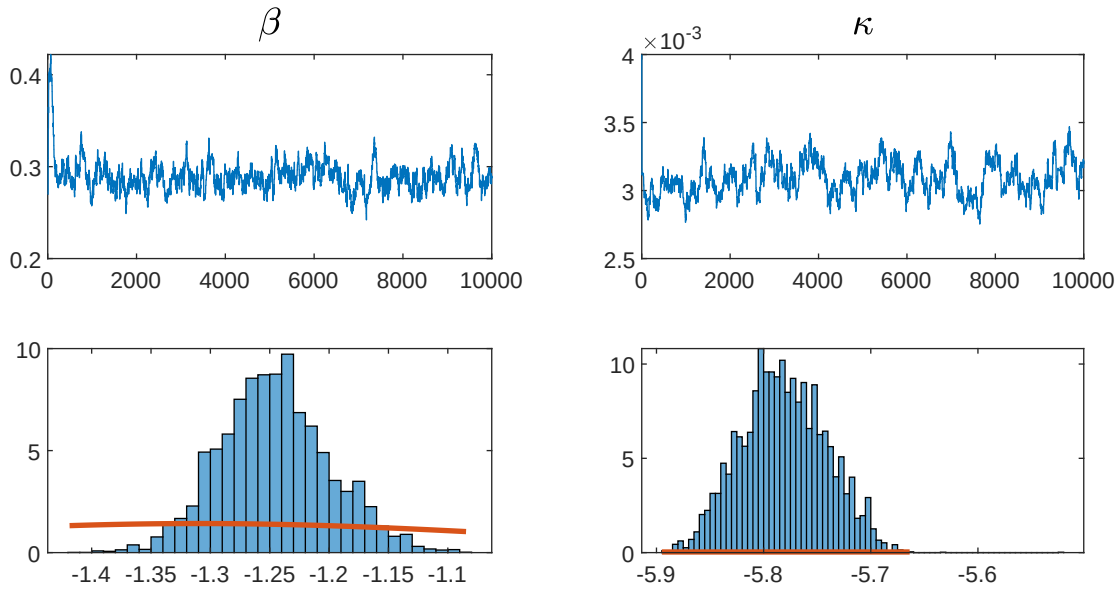

Figure 7.6: Results for one location corresponding to the second row, first column of marked locations in Figure 7.1. **Top row:** MCMC chains for degradation rate β and light scaling κ over 10k MCMC iterations. The trace plots suggest rapid convergence and good mixing (low autocorrelation). **Bottom row:** Posterior distributions of log parameters and prior distribution over the support of the posterior.

7.4.5 Parameter estimates across SCN tissue

The analysis is repeated at 15 locations, each consisting of a 2 by 2 pixel-block, across the SCN and the resulting posterior means are plotted in Figures 7.7 and 7.8 where the size of each marker is inversely proportional to the square root of the coefficient of variation, thus giving an indication of the posterior dispersion. Note that since the proportionality constant is different for each parameter, the posterior dispersion is only comparable between locations, not parameters.

The results suggest that the parameters of the Per2 loop, along with degradation rate β and light scaling κ have low spatial variation of the posterior dispersion. Furthermore, there is little evidence of systematic variation of the parameters across the biological tissue, apart from κ which has higher posterior means in the top left locations, corresponding to the dorsal SCN.

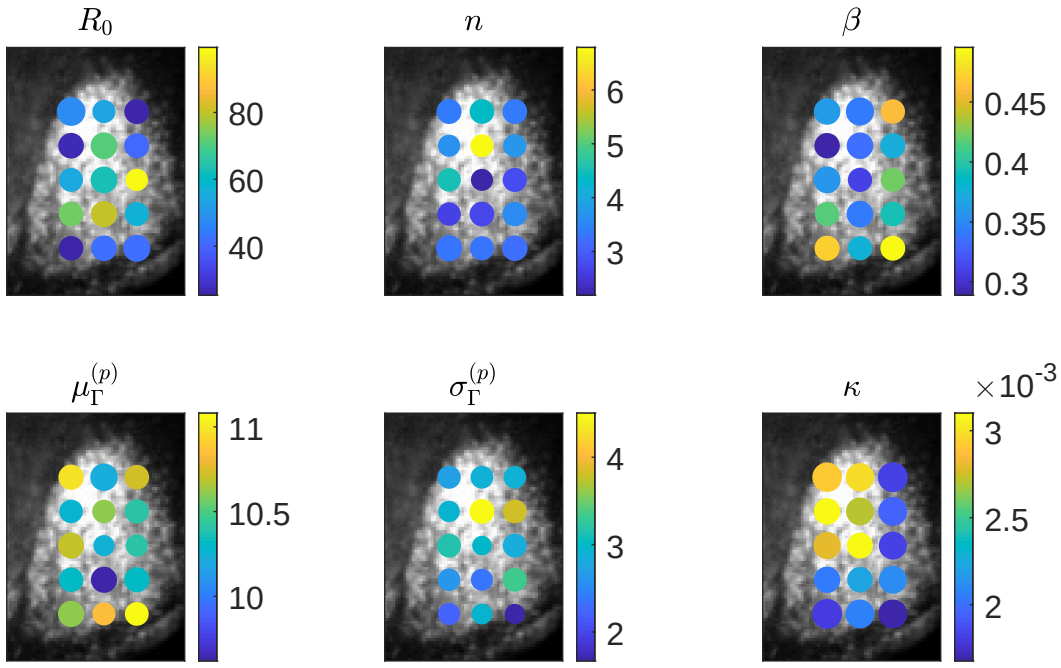

Figure 7.7: Posterior means of parameters associated with Per2 loop, along with degradation rate β and light scaling κ across WT SCN tissue. Size of markers is inversely proportional to the square root coefficient of variation of the posterior distribution.

The parameters governing calcium activation in Figure 7.8 have higher spatial variation of the posterior dispersion, which is expected due to their disperse prior distributions. Here, the transformed parameter R_{TF}/R_0 is reported, which has a posterior mean in the range $2-8$. A value above 1 indicates that the activating effect of CREB binding is greater than the activating effect of CLOCK/BMAL which is captured by R_0 and assumed constant in the model. Furthermore, a majority of locations have estimates of $R_{int.}$ that are consistent with no (or very low) interaction effect between delayed Per2 and calcium.

The mean of the delay distribution of calcium has a posterior mean of $<$ 8h for all locations except one. The "mean of means", across the 15 locations is 3.82h which is slightly lower than the phase difference between calcium and Per2 found in Chapter 2 (5.27h) for the same data. However, the average phase difference relates the peak timing of the two species while the delay mean relates peak timing of calcium to maximum

transcription rate of Per2, which is achieved before peak concentrations.

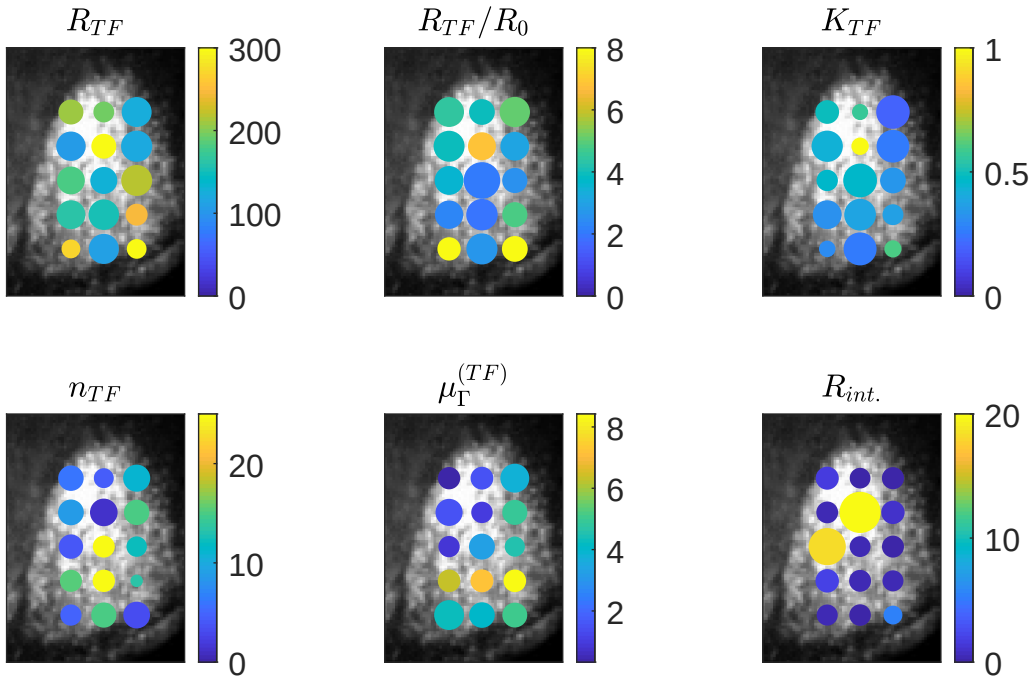

Figure 7.8: Posterior means of parameters associated with calcium induced activation of Per2 transcription across WT SCN tissue. Size of markers is inversely proportional to the square root coefficient of variation of the posterior distribution.

The model fit is evaluated by examining residuals obtained by subtracting the filter mean, evaluated at the posterior mean of the parameters, from the observed data. The residuals are tested for normality using the one-sample Kolmogorov-Smirnov test with Bonferroni correction and Fishers g-statistic (Section 4.5.2). The tests generally show compliance with normality as the null hypothesis of normality is rejected in only 10 percent of analysed pixels (6 out of 60). Residual periodicity is only significant and in the circadian range (18-30 hours) for a single full 2 by 2 pixel block (fourth row, third column). A handful of other pixels show significant residual period, possibly due to the restriction that all parameters are equal within each pixel block. However, it is worth noting that the residuals do not show such systematic 12 hour periodicity as found for the Cry1-luc data and TTFL model with simple auto-regulation in Chapters 4 and 6.

7.4.6 In silico CREB knock-out experiment

The transcription model with purely auto-regulatory dynamics treated in Chapter 4 is a special case of the TTFL model with a secondary activation mechanism in Eq. (7.12). The functional form of the two models is identical if the exogenous transcription factor, in this case calcium, is set equal to zero. Hence the stability criteria derived in Chapter 4 can also be used to analyse the macroscopic rate equation in Eq. (7.6) in the absence of the activating species. With this in mind, we construct an experiment to examine the effects of calcium-induced activation by evaluating the posterior probability of a limit cycle, developed as a measure of oscillatory robustness in Sections 3.7.2 and 4.5.4. We shall also examine the effects of the activating mechanism on the period of the circadian oscillations by applying the period estimation technique outlined in Section 6.1.8 for the parameter estimates obtained for Per2 and the extended transcription model.

The interpretation of the limit cycle and period estimates is the effects of an artificial removal of calcium/CREB from the circuit, which we hypothesize results in dynamics that are similar to those observed for Per2 in VIP-null SCN tissue, i.e. damped oscillations. For the procedure to obtain these estimates from the MCMC output we refer to Chapter 4, Section 4.5.4 and Chapter 6, Section 6.1.8 for the posterior probability of a limit cycle and period estimation, respectively, and present here only the results in Figure 7.9.

Despite the fact that the Hill coefficient associated with delayed Per2 is estimated to be higher for the WT SCN compared to the VIP-null SCN from which the prior distribution was elicited, the posterior probability of limit cycle dynamics is close to zero for all 15 locations apart from one where it is > 0.9. The posterior mean population size is around 100 for most locations and the period is typically greater than 24 (mean across locations is 26.72 hours) and greater than the estimates obtained using spectral analysis on the same data (see Chapter 2, Figure 2.5). Interestingly, observational studies have found evidence that the intrinsic period of circadian rhythms in mammals, which emerges in the absence of external stimuli, is in fact longer than 24 hours (see e.g. Campbell (2000) and Czeisler et al. (2000)).

In summary, our findings suggest that removal of calcium signalling induces damped oscillations of the macroscopic mean, and a lengthened period that, in the absence of molecular noise, reach an equilibrium at around 100 mRNA molecules.

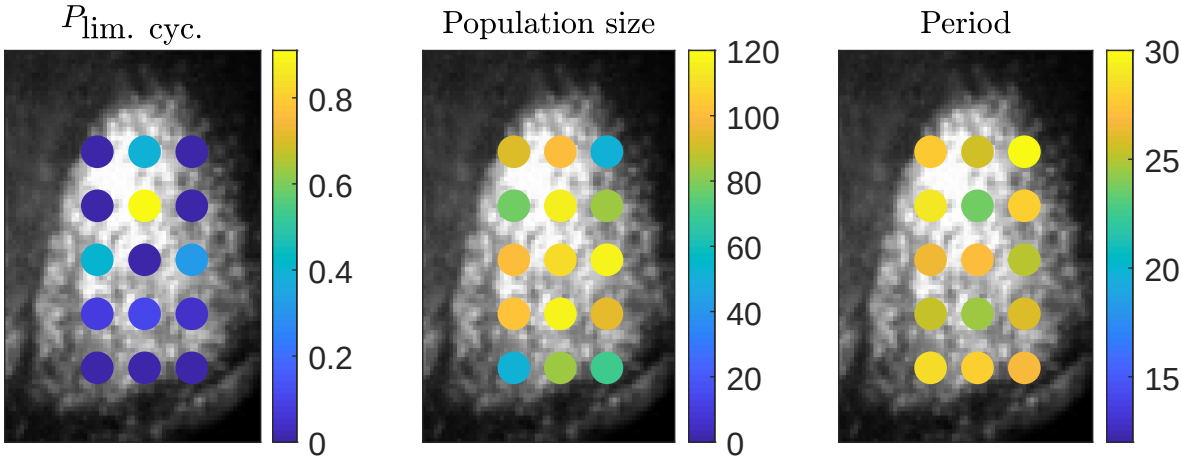

Figure 7.9: Posterior probability of limit cycle dynamics of the macroscopic mean, posterior mean population size and posterior mean period of oscillations of Per2 mRNA and synthetic calcium removal, for 15 pixel blocks across WT SCN tissue.

7.5 Discussion

In this chapter we have reviewed a stochastic model of the TTFL governing circadian gene transcription developed by Calderazzo (2016) that accommodates a time-varying, exogenous activation mechanism in addition to delayed self-inhibition. The functional form of the model is motivated by the fact that the promoter region of the Per2 gene exhibits CREs and transcription of the gene is thus up-regulated by CREB, which is a circadian species downstream of calcium in the VIP signalling cascade. The model incorporates a distributed delay for both the auto-regulatory mechanism and the exogenous species, allowing for the use of an upstream proxy of the true activation mechanism.

We derived the explicit dependence between parameters and population sizes of the two molecular species which yielded a decoupling of the drift term/macroscopic rate (but not noise term) from molecular counts. A harmonic dynamic linear model was developed to account for temporal smoothing and measurement errors in observations of the exogenous circadian species. Finally, an MCMC algorithm was proposed to jointly sample the posterior of parameters of the extended TTFL model and unobserved states of the exogenous transcription factor.

In Section 7.4 we applied the methodology to simultaneous recordings of Per2:luc and GCaMP-reported calcium from Brancaccio et al. (2013). As the full set of param-

eters was not identifiable using the available data, we elicited informative prior distributions for the parameters of the auto-regulatory loop of Per2 by fitting the purely autoregulatory TTFL model from previous chapters to Per2 expression data from VIP-null SCN obtained by Maywood et al. (2011). These prior distributions, in combination with a series of simplifying assumptions allowed us to estimate the parameters governing the CREB-induced transcriptional activation of Per2.

The results suggest that transcriptional activation by CREB, and by extension VIP-signalling is typically 2-8 times that of the basal transcription rate R_0 which captures a time-constant CLOCK/BMAL1-induced activation. Additionally, we find evidence of a low or non-existent interaction effect between CREB activation and delayed Per2 inhibition, i.e. transcriptional activation through CREB binding is not found to be dependent on the concentration of the inhibitory Per2 heterodimer, modelled here by delayed mRNA. It should however be noted that while we model the calcium cascade as an exogenous input, Maywood et al. (2011) suggest that calcium is not only an input but its rhythmic expression is tuned by the TTFL in an interdependent fashion, which limits conclusions regarding the causal relationship of Per2 and calcium.

While the analysis only covers a small subset of locations across the right half of the SCN, there are indications of spatial heterogeneity of parameters between the dorsomedial and ventrolateral SCN. Specifically regarding the mean delay between observed calcium and Per2 and the Hill coefficient of unobserved CREB, which were both estimated to be higher in the VIP-ergic ventrolateral SCN. The Hill coefficient of CREB is estimated in the range $10 - 30$ for several locations across the tissue, which is likely to be an unrealistic degree of cooperativity. This finding may indicate that transcriptional activation by the VIP-signalling cascade is accurately modelled by a fast acting transcriptional on-off switch.

As the autoregulatory TTFL model in Chapters 4 and 6 is a special case of the extended model assumed in this chapter, the stability criteria derived in Section 4.2.1 may be applied to determine the dynamics in the absence of the activating mechanism. This is exploited to design an in silico CREB-knockout experiment, where the posterior probabilities of a limit cycle, equilibrium population sizes, and periods of oscillations are calculated from the MCMC output. We find that in the absence of CREB, the macroscopic mean of the cellular clocks typically exhibit damped oscillations of Per2 ($P_{\text{lim. cyc.}} < 0.5$ in 14 out of 15 locations). The equilibrium population sizes are around 100 mRNA molecules, similar to those inferred in wild-type Cry1-luc data. Interestingly, removal of CREB was found to cause lengthening of the circadian period to $24-30$ hours.

Conclusions

In this work we have studied methodology motivated by its use in analysing stochastically driven circadian gene transcription and regulation. Methodologies have been developed and extended in two main directions: inferential approaches motivated by the detailed form of mechanistic models and the availability of high-dimensional spatio-temporal imaging data from biological tissue, and analysis approaches that facilitate biological insight from the intricate structure of the resulting model fits.

In Chapters 1 and 2 the reader is introduced to the molecular machinery that underlies the generation of circadian rhythms in the SCN. Bioimaging data of Per2 expression is investigated using spectral methods which are commonly used as a phenomenological analysis tool for circadian data. Future research directions are outlined and some limitations of spectral methods are discussed. While the spectral resampling method of Costa et al. (2013) used in Chapter 2 provides uncertainty quantification, it is not possible to separate different sources of noise, e.g. intrinsic noise, due to the discrete nature of molecular interactions, and measurement noise generated by the experimental procedure. Additionally, spectral methods rely on high frequency observations to provide results with high temporal resolution and bioimaging data typically requires long exposure times (0.5-1 hours) due to the low light intensity of the involved reporter proteins. We find that the phenomenological approach is limited in the type of inferences that can be made, motivating approaches where transcriptional regulation is modelled in a mechanistic fashion.

From Chapter 3 and onwards, we review, develop and apply mechanistic modelling approaches for transcriptional regulation of circadian genes. Chapter 3 serves as an introduction to most of the methodology used in the subsequent chapters. Chemical reaction networks, along with approximations that lead to the Chemical Langevin Equation are explained, with the motivation that molecular noise may be an important aspect in the generation of biological rhythms. We discuss Bayesian inference and likelihood approximation using the extended Kalman-Bucy filter. Markov chain Monte

Carlo methodology, which we rely on to sample from intractable posterior distributions in Chapters 5-7, is reviewed. Finally, we review stability analysis and propose a novel measure of Bayesian biological robustness, given by the posterior distribution of some system output. Our definition is consistent with a more general definition of biological robustness by Kitano (2007) and can be applied to study e.g. to what extent SCN neurons exhibit sustained oscillations.

Generally, the expression of circadian genes in SCN neurons has previously been described using either limit cycle oscillators or damped oscillators, and often, the dynamics are deterministically modelled. The question whether single SCN neurons should be modelled as deterministically sustained oscillators or damped/noise-induced oscillators has previously been studied by Westermark et al. (2009). An advantage of the model fitting approach employed in Chapters 4, 6 and 7 is that various aspects of the dynamic behaviour of the molecular oscillators may be empirically inferred from observed data, and we propose a coherent methodology to answer questions regarding the long-run dynamics of molecular oscillators. In Chapter 4 we derive a stochastic distributed delay model for circadian gene expression, originally proposed by Calderazzo et al. (2018). Our methodological contribution consists of deriving stability criteria for the macroscopic mean, corresponding to the drift term of the Chemical Langevin equation, enabling estimation of our proposed measure of robustness.

An empirical application is provided where the model is fitted to bioimaging data of Cry1-luc expression obtained by Brancaccio et al. (2013). The posterior distribution is sampled using an adaptive, delayed acceptance random walk Metropolis algorithm and as the data is spatio-temporal, the inferential procedure is repeated across a large number of locations to build a spatial distribution of parameter estimates. We subsequently use the stability criteria to construct a spatial distribution of posterior probabilities of a limit cycle across the imaged tissue and find that in intact tissue, the SCN single-cell oscillator exhibits sustained oscillations whereas adjacent tissue does not. We are thus able to accurately classify SCN tissue from surrounding tissue using the novel robustness statistic. Estimates of the involved molecular population sizes are consistent with simulation-based results obtained under strict assumptions by Abel et al. (2015). Studying the residuals of the model fit reveals 12 hour periodicities in central locations of the organ that are not adequately modelled by the auto-regulatory feedback loop. These additional periodicities may be generated by astrocytic circadian gene expression that is antiphasic to the neuronal oscillators, described by Brancaccio et al. (2017).

The spatial aspect of imaging data is tackled in Chapter 5. A Bayesian hierarchical model that captures spatial dependencies at the parameter level is proposed,

specifically a spatial random effects model where a conditional autoregressive prior distribution is placed on the random effects. This approach provides a parsimonious model, where inference across a large number of pixels can be done in parallel. Spatial hierarchical models have been used successfully in e.g. rainfall modelling (Cooley and Sain, 2010) and disease mapping (Best et al., 2005) but modelling spatial dependence between pixels in bioimaging data is, to the best of our knowledge, a novel application. A random walk Metropolis algorithm is developed to sample the posterior distribution of parameters and random effects and a simulation study is designed to investigate optimal blocking strategies. Finally, we validate the methodology by re-estimating parameters of synthetic circadian bioimaging data and find that we are able to recover spatial structures of parameters and dynamics to a satisfactory degree.

In Chapter 6 we present an empirical application of the methodology developed in the previous chapter. The Cry1-luc imaging data from Chapter 4 is re-analysed using the Bayesian hierarchical model, along with two additional experimental replicates of the same gene. We find that using the hierarchical model enables estimation of population sizes without an informative prior on the light scaling parameter κ, due to parameter restrictions imposed by the spatial model, along with the increased amount of data entering the likelihood. Biologically interpretable findings were found to be largely similar to those obtained using the spatially independent approach in Chapter 4, but we obtain results with much higher spatial resolution for additional data sets.

We further develop methodology to analyse various aspects of the model under the estimated parameters. Specifically, we propose a Bayesian period estimator which, unlike spectral methods, does not rely on high frequency observations to produce period estimates with high temporal resolution. Our estimator reveals spatial heterogeneity of the period of the macroscopic mean for the three experimental replicates, which we are unable to detect using spectral methods. This finding suggests that either the sampling rate of data is inadequate to produce high temporal resolution of period estimates obtained by spectral analysis, or that period homogeneity is driven by intrinsic noise.

An inhibition profile for the distributed delay model is derived, which yields a mechanistic interpretation to an observed relationship between the Hill coefficient, n, and dispersion of the delay distribution, σ_Γ. In a simulation study, the shape of the inhibition profile is linked to a well-known trade-off between two key features of SCN neurons: synchrony and entrainability (Hasegawa and Arita, 2014). Repeating the simulation study in a spatial setting using the empirical spatial distribution of parameter estimates provides evidence that central SCN neurons exhibit a smaller phase dispersion, or loss of synchrony, compared to those along the edges of the organ.

In the final chapter of the thesis we go beyond transcriptional auto-regulation and incorporate a second, exogenous transcription factor that represents signalling between cells and hence a synchronising mechanism. A model, originally proposed by Calderazzo (2016), that captures both transcriptional inhibition and activation through distributed delays is derived. Our contribution consists of a rate rescaling that clarifies parameter interpretation in the case where the two molecular species involved are specified in arbitrary units. The rescaling decouples the drift term from molecular counts and places a dispersion parameter in the noise term that reflects population size. The EKBF for distributed delay models is adapted to the extended transcription function involving an additional exogenous species and a harmonic dynamic linear model is developed for a partially observed, circadian transcription factor.

An empirical application is provided using simultaneously recorded Per2 and calcium expression bioimaging data from organotypic SCN tissue, obtained by Brancaccio et al. (2013). Informative prior distributions for parameters of the auto-regulatory loop are developed by applying the hierarchical model from the previous chapter to bioimaging data of Per2 expression from VIP-null SCN tissue developed by Maywood et al. (2011). These prior distributions allow for fitting the extended TTFL model to Per2 and calcium, which serves as a proxy for CREB, from wild-type SCN tissue. The approach allows us to make inferences regarding the relative importance of the modelled transcription factors and the parameter estimates provide input values for a realistic, synthetic model of the SCN. Additionally, inference for the extended TTFL model constitutes a first step in developing a mechanistic spatial model for ensembles of SCN neurons as rhythmic CREB is the output of a signalling cascade that begins with light-information reaching retinally innervated ventral SCN neurons. The chapter is concluded with an *in silico* experiment where we exploit the fact that the extended TTFL model contains the purely auto-regulatory TTFL model from Chapter 4 as a special case. In the experiment we study the effect on the dynamics of Per2 expression by an artificial removal CREB and find that such an intervention leads to damped oscillations with an increased period.

Throughout the thesis we show that it is indeed possible to estimate parameters of mechanistic and stochastic models through principled statistical analysis of circadian bioimaging data. The models, while simpler than those typically employed to study the SCN, have rich dynamics, such that the results elucidate mechanisms with which biological rhythms are created and maintained. The findings constitute input parameters for an empirically tuned, synthetic model that may be used by experimentalists in an iterative manner to construct novel biological experiments.

Further research

Most of the empirical applications in this thesis are computationally costly, with algorithms taking days or weeks of wall-clock time to produce useful output. This is likely to be prohibitively expensive for practitioners that aim to implement the methodology developed in this thesis. A direction for future research is thus techniques that allows for faster likelihood approximation or faster exploration of posterior distributions. Significant speed-up is likely to be achieved if quantification of parameter uncertainty is abandoned and an optimisation approach is used instead. Such a view would however restrict the type of inferences that may be made as e.g. our robustness measure relies on being able to quantify some behaviour over a range of parameter values.

Inference for the extended TTFL model in Chapter 7 constitutes a first step in developing a mechanistic spatial model for the SCN circuit as the additional transcription factor is an output of a inter-cellular signalling cascade. The network topology of the SCN is studied by e.g. Vasalou et al. (2009) using detailed deterministic models but no parameter estimation and Abel et al. (2016) using principled inference for the coupling strength of simplistic phase-amplitude oscillators. Combining our inference for the extended TTFL model with an additional model describing the circuit connectivity, e.g. in an hierarchical framework, would yield a "complete" model of both single-cell and organ-wide dynamics.

Appendix A

Numerical values of parameter estimates for hierarchical model fit to Cry1-luc data

Table A.1: Spatial distribution of posterior mean estimates, Replicate 1.

Parameter	R_0	K	μ_Γ	σ_Γ	n	β	κ	σ_η
10^{th} percentile	52.23	103.16	8.34	2.69	4.08	0.132	0.0028	0.0047
Spatial mean	57.38	117.13	8.84	3.58	5.24	0.197	0.0036	0.0055
90^{th} percentile	61.88	131.85	9.18	4.63	6.85	0.246	0.0043	0.0066

Table A.2: Spatial distribution of posterior mean estimates, Replicate 2.

Parameter	R_0	K	μ_Γ	σ_Γ	n	β	κ	σ_η
10^{th} percentile	49.18	97.23	8.18	2.60	3.82	0.154	0.0020	0.0049
Spatial mean	53.13	107.20	8.67	3.17	4.56	0.203	0.0034	0.0062
90^{th} percentile	57.36	118.01	9.22	3.72	5.29	0.269	0.0046	0.0076

Table A.3: Spatial distribution of posterior mean estimates, Replicate 3.

Parameter	R_0	K	μ_Γ	σ_Γ	n	β	κ	σ_η
10^{th} percentile	71.19	165.50	8.48	2.43	3.99	0.193	0.0017	0.0046
Spatial mean	75.55	177.84	8.70	2.93	4.55	0.219	0.0026	0.0057
90^{th} percentile	79.9	191.07	8.95	3.46	5.10	0.253	0.0036	0.0069

Appendix B

Likelihood approximation of TTFL model with exogenous activation using Extended Kalman-Bucy filter

The discrete-time approximation of the CLE may be written

$$X_{t+\delta t} = X_t + f(X_\tau, X_t) + \sqrt{\delta t}\sqrt{A(X_\tau, X_t)}W_t, \tag{A.1}$$

where dependence on the additional transcription factor Z is suppressed, as it is treated as an observed constant. For the extended TTFL model the explicit form is

$$\begin{aligned} X_{t+\delta t} = X_t &+ \left(\frac{R_0 + R_{TF}\left(\frac{Z_\tau}{K_{TF}}\right)^{n_{TF}} - R_{int.}\left(\frac{Z_\tau}{K_{TF}}\right)^{n_{TF}}\left(\frac{X_\tau}{K}\right)^{n}}{1 + \left(\frac{X_\tau}{K}\right)^{n} + \left(\frac{Z_\tau}{K_{TF}}\right)^{n_{TF}} + \left(\frac{Z_\tau}{K_{TF}}\right)^{n_{TF}}\left(\frac{X_\tau}{K}\right)^{n}} - \beta X_t \right)\delta t \\ &+ \sqrt{\delta t}\sqrt{\left(\frac{R_0 + R_{TF}\left(\frac{Z_\tau}{K_{TF}}\right)^{n_{TF}} - R_{int.}\left(\frac{Z_\tau}{K_{TF}}\right)^{n_{TF}}\left(\frac{X_\tau}{K}\right)^{n}}{1 + \left(\frac{X_\tau}{K}\right)^{n} + \left(\frac{Z_\tau}{K_{TF}}\right)^{n_{TF}} + \left(\frac{Z_\tau}{K_{TF}}\right)^{n_{TF}}\left(\frac{X_\tau}{K}\right)^{n}} + \beta X_t \right)} W_t \end{aligned} \tag{A.2}$$

where

$$\begin{aligned} Z_\tau &= \sum_{s=\delta t}^{\tau_{\max}} Z_{t-s} g'_{a_{TF}, p_{TF}}(s) \\ X_\tau &= \sum_{s=\delta t}^{\tau_{\max}} X_{t-s} g'_{a,p}(s), \end{aligned} \tag{A.3}$$

and $g'(s)$, $s = \delta t,\ 2\delta t,\ \ldots, 24$ are weights from the truncated and normalised gamma delay distribution and $W_t \sim \mathcal{N}(0,1)$. Similarly, the time integration in the measurement equation may be approximated in discrete time by

$$Y_t = \kappa \tilde{F} X_{(t-\Delta t):t} + \eta_t = F X_{(t-\Delta t):t} + \eta_t,\ \eta_t \sim \mathcal{N}(0, \sigma_\eta^2). \tag{A.4}$$

Replacing f and A by their first order Taylor approximation about the deterministic mean ρ and assuming $X_0 \sim \mathcal{N}(\rho_0, P_0)$, we obtain Gaussian transition densities through the LNA equations, which are propagated in discrete time until the next observation by

$$\begin{aligned}\rho_{t+\delta t} &= \rho_t^* + f(\rho_\tau^*, \rho_t^*)\delta t \\ P_{t+\delta t} &= J_f(\rho_\tau^*, \rho_t^*)P_t^*\delta t + P_t^{*\intercal} J_f(\rho_\tau^*, \rho_t^*)^\intercal \delta t + A(\rho_\tau^*, \rho_t^*)\delta t,\end{aligned} \tag{A.5}$$

where again, subscripted τ denotes the discrete-time convolution with the delay distribution weights

$$\rho_\tau^* = \sum_{s=\delta t}^{\tau_{\max}} \rho_{t-s}^* g'_{a,p}(s), \tag{A.6}$$

and

$$\begin{aligned}J_f = \Bigg(&\frac{nR_{int}\left(\frac{Z_\tau}{K_{TF}}\right)^{n_{TF}}\left(\frac{\rho_\tau^*}{K}\right)^n}{K\left(1+\left(\frac{\rho_\tau^*}{K}\right)^n + \left(\frac{Z_\tau}{K_{TF}}\right)^{n_{TF}} + \left(\frac{\rho_\tau^*}{K}\right)^n\left(\frac{Z_\tau}{K_{TF}}\right)^{n_{TF}}\right)} \\ &- \frac{\left(n\left(\frac{Z_\tau}{K_{TF}}\right)^{n_{TF}}\left(\frac{\rho_\tau^*}{K}\right)^{n-1} + n\left(\frac{\rho_\tau^*}{K}\right)^{n-1}\right)\left(R_0 + R_{TF}\left(\frac{Z_\tau}{K_{TF}}\right)^{n_{TF}} + R_{int.}\left(\frac{Z_\tau}{K_{TF}}\right)^{n_{TF}}\left(\frac{\rho_\tau^*}{K}\right)^n\right)}{K\left(1+\left(\frac{\rho_\tau^*}{K}\right)^n + \left(\frac{Z_\tau}{K_{TF}}\right)^{n_{TF}} + \left(\frac{\rho_\tau^*}{K}\right)^n\right)\left(\frac{Z_\tau}{K_{TF}}\right)^{n_{TF}}}\Bigg) \\ &\cdot [g'_{p,a}(\delta t),\ g'_{p,a}(2\delta t),\ \ldots, g'_{p,a}(\tau_{max})]\end{aligned} \tag{A.7}$$

is the Jacobian of f, w.r.t. $X_{t:(t-\tau_{max})}$, evaluated at the updated deterministic mean $\rho^*_{t:(t-\tau_{max})}$. While the predicted mean and covariance are obtained by propagating Eq. (A.5), the updated quantities ρ^* and P^* are calculated using the Kalman update

$$\begin{aligned}\rho^*_{(t+\Delta t-\tau_{max}):(t+\Delta t)} &= \rho_{(t+\Delta t-\tau_{max}):(t+\Delta t)} + C(Y_{t+\Delta t} - F\rho_{t+\Delta t}) \\ P^*_{(t+\Delta t-\tau_{max}):(t+\Delta t)} &= P_{(t+\Delta t-\tau_{max}):(t+\Delta t)} - CFP_{t+\Delta t,(t+\Delta t-\tau_{max}):(t+\Delta t)} \\ C &= P_{(t+\Delta t-\tau_{max}):(t+\Delta t),t+\Delta t}F^\intercal(FP_{t+\Delta t}F^\intercal + \sigma_\eta^2)^{-1}.\end{aligned} \tag{A.8}$$

Finally, the likelihood is calculated using the prediction errors $e_t = Y_t - F\rho_{(t-\Delta t+\delta t):t}$ and

$$\log \mathcal{L}(\Psi; Y_{\Delta t:T}) = -\frac{1}{2}\sum_{t=\Delta t}^{T}[\log|FP_{(t-\Delta t+\delta t):t}F^\intercal + \sigma_\eta^2| + e_t^\intercal(FP_{(t-\Delta t+\delta t):t}F^\intercal + \sigma_\eta^2)^{-1}e_t] + c, \tag{A.9}$$

where, as discussed in the main text, normality follows from linearity and additive Gaussian measurement errors.

Kalman Filter and simulation smoother for dynamic harmonic linear model of exogenous transcription factor

We shall here describe a Kalman filter and simulation smoother for the state-space representation of the model describing the exogenous activating species, given by

$$\begin{aligned} Z_t^{(\text{obs})} &= F\Theta_t + \upsilon_t, \;\; \upsilon_t \sim \mathcal{N}(0, R) \\ \Theta_{t+1} &= \mathbf{H}\Theta_t + \boldsymbol{\epsilon}_t, \;\; \boldsymbol{\epsilon}_t \sim \mathcal{N}(0, \boldsymbol{Q}). \end{aligned} \tag{A.10}$$

Let $\Theta_{t|t-1} = \mathbb{E}[\Theta_t | Z_{1:t-1}^{(obs)}]$ and $P_{t|t-1} = V[\Theta_t | Z_{1:t-1}^{(obs)}]$. The Kalman filter (forward) recursion is then given by

$$\begin{aligned} \Theta_{t|t-1} &= \mathbf{H}\Theta_{t-1|t-1} \\ P_{t|t-1} &= \mathbf{H}\Theta_{t|t-1}\mathbf{H}^\intercal + \mathbf{Q} \\ \hat{\upsilon}_t &= Z_t^{(obs)} - F\Theta_{t|t-1} \\ V_t &= FP_{t|t-1}F^\intercal + R \\ K_t &= P_{t|t-1}F^\intercal V_t^{-1} \\ \Theta_{t|t} &= \Theta_{t|t-1} + K_t^\intercal \hat{\upsilon}_t \\ P_{t|t} &= P_{t|t-1} - P_{t|t-1}F^\intercal V_t^{-1} F P_{t|t-1}. \end{aligned} \tag{A.11}$$

By linearity and normality, the log-likelihood is given by

$$\log \mathcal{L}(\theta_{1:T}, Q, R; Z_{1:T}^{(obs)}) = c - \sum_{t=1}^{T} \left(\frac{\hat{\upsilon}_t^2}{V_t} + \log |V_t| \right). \tag{A.12}$$

While the filter (forward) recursion gives the distribution of $\Theta_t | Z_{1:t}^{(obs)}$ the smoothing (backward) recursion gives the expected value of states conditioned on the entire set of

observations, i.e. $\hat{\Theta}_t = \mathbb{E}[\Theta_t | Z_{1:T}^{(obs)}]$. Let $r_T = 0$, then

$$\begin{aligned} L_{t-1} &= \mathbf{H} - \mathbf{H}K_{t-1}F \\ r_{t-1} &= FV_t^{-1}\hat{v}_t + L_t' r_t \\ \tilde{\nu}_{t-1} &= RV_{t-1}^{-1}\hat{v}_{t-1} - R\mathbf{H}K_{t-1}' r_{t-1} \\ \tilde{\boldsymbol{\epsilon}}_{t-1} &= \mathbf{Q}r_{t-1} \end{aligned} \tag{A.13}$$

and forward

$$\hat{\Theta}_t = \mathbf{H}\Theta_{t|t} + \mathbf{Q}r_t.$$

Bibliography

John H Abel, Lukas A Widmer, Peter C St John, Jörg Stelling, and Francis J Doyle. A Coupled Stochastic Model Explains Differences in *Cry* Knockout Behavior. *IEEE Life Sciences Letters*, 1(1):3–6, 2015.

John H Abel, Kirsten Meeker, Daniel Granados-Fuentes, Peter C St John, Thomas J Wang, Benjamin B Bales, Francis J Doyle, Erik D Herzog, and Linda R Petzold. Functional network inference of the suprachiasmatic nucleus. *Proceedings of the National Academy of Sciences*, 113(16):4512–4517, 2016.

Eric E Abrahamson and Robert Y Moore. Suprachiasmatic nucleus in the mouse: retinal innervation, intrinsic organization and efferent projections. *Brain Research*, 916(1-2): 172–191, 2001.

Sungwon An, Connie Tsai, Julie Ronecker, Alison Bayly, and Erik D Herzog. Spatiotemporal distribution of vasoactive intestinal polypeptide receptor 2 in mouse suprachiasmatic nucleus. *Journal of Comparative Neurology*, 520(12):2730–2741, 2012.

Sungwon An, Rich Harang, Kirsten Meeker, Daniel Granados-Fuentes, Connie A Tsai, Cristina Mazuski, Jihee Kim, Francis J Doyle, Linda R Petzold, and Erik D Herzog. A neuropeptide speeds circadian entrainment by reducing intercellular synchrony. *Proceedings of the National Academy of Sciences*, 110(46):E4355–E4361, 2013.

Bharath Ananthasubramaniam, Erik D Herzog, and Hanspeter Herzel. Timing of neuropeptide coupling determines synchrony and entrainment in the mammalian circadian clock. *PLoS Computational Biology*, 10(4):e1003565, 2014.

Christophe Andrieu and Éric Moulines. On the ergodicity properties of some adaptive MCMC algorithms. *The Annals of Applied Probability*, 16(3):1462–1505, 2006.

Yves F Atchadé and Jeffrey S Rosenthal. On adaptive Markov chain Monte Carlo algorithms. *Bernoulli*, 11(5):815–828, 2005.

Sudipto Banerjee, Alan E Gelfand, Andrew O Finley, and Huiyan Sang. Gaussian predictive process models for large spatial data sets. *Journal of the Royal Statistical Society: Series B*, 70(4):825–848, 2008.

Majid Bani-Yaghoub. Analysis and Applications of Delay Differential Equations in Biology and Medicine. *arXiv preprint arXiv:1701.04173*, 2017.

L Mark Berliner. Hierarchical Bayesian Time Series Models. In *Maximum Entropy and Bayesian Methods*, pages 15–22. Springer, 1996.

Julian Besag. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B*, 36(2):192–225, 1974.

Nicky Best, Sylvia Richardson, and Andrew Thomson. A comparison of Bayesian spatial models for disease mapping. *Statistical Methods in Medical Research*, 14(1):35–59, 2005.

Marco Brancaccio, Elizabeth S Maywood, Johanna E Chesham, Andrew SI Loudon, and Michael H Hastings. A Gq-Ca 2+ axis controls circuit-level encoding of circadian time in the suprachiasmatic nucleus. *Neuron*, 78(4):714–728, 2013.

Marco Brancaccio, Andrew P Patton, Johanna E Chesham, Elizabeth S Maywood, and Michael H Hastings. Astrocytes control circadian timekeeping in the suprachiasmatic nucleus via glutamatergic signaling. *Neuron*, 93(6):1420–1435, 2017.

Talha Burki. Nobel Prize awarded for discoveries in circadian rhythm. *The Lancet*, 390 (10104):e25, 2017.

Silvia Calderazzo. *Stochastic modelling of transcriptional regulation with applications to circadian genes*. PhD thesis, University of Warwick, 2016.

Silvia Calderazzo, Marco Brancaccio, and Bärbel Finkenstädt. Filtering and inference for stochastic oscillators with distributed delays. *Bioinformatics*, 35(8):1380–1387, 2018.

Scott Campbell. Is there an intrinsic period of the circadian clock? *Science*, 288(5469): 1174–1175, 2000.

Bradley P Carlin, Alan E Gelfand, and Sudipto Banerjee. *Hierarchical Modeling and Analysis for Spatial Data*. Chapman and Hall/CRC, 2014.

Alanna B Chan and Katja A Lamia. Cancer, hear my battle CRY. *Journal of Pineal Research*, page 00:e12658, 2020.

Kung Sik Chan and Charles J Geyer. Discussion: Markov chains for exploring posterior distributions. *The Annals of Statistics*, 22(4):1747–1758, 1994.

J Andrés Christen and Colin Fox. Markov chain Monte Carlo using an approximation. *Journal of Computational and Graphical statistics*, 14(4):795–810, 2005.

Andrew N Coogan, Barbora Schutová, Susanne Husung, Karolina Furczyk, Bernhard T Baune, Peter Kropp, Frank Häßler, and Johannes Thome. The circadian system in Alzheimers disease: disturbances, mechanisms, and opportunities. *Biological psychiatry*, 74(5):333–339, 2013.

Daniel Cooley and Stephan R Sain. Spatial hierarchical modeling of precipitation extremes from a regional climate model. *Journal of Agricultural, Biological, and Environmental Statistics*, 15(3):381–402, 2010.

Maria J Costa, Bärbel Finkenstädt, Véronique Roche, Francis Lévi, Peter D Gould, Julia Foreman, Karen Halliday, Anthony Hall, and David A Rand. Inference on periodicity of circadian time series. *Biostatistics*, 14(4):792–806, 2013.

Noel Cressie and Christopher K Wikle. *Statistics for Spatio-Temporal Data*. John Wiley & Sons, 2015.

David J Cutler, Mai Haraura, Helen E Reed, Sanbing Shen, W John Sheward, Christine F Morrison, Hugh M Marston, Anthony J Harmar, and Hugh D Piggins. The mouse VPAC2 receptor confers suprachiasmatic nuclei cellular rhythmicity and responsiveness to vasoactive intestinal polypeptide in vitro. *European Journal of Neuroscience*, 17(2):197–204, 2003.

CA Czeisler, DJ Dijk, RE Kronauer, EN Brown, JF Duffy, JS Allan, TL Shanahan, DW Rimmer, JM Ronda, and RF Mitchell. Is there an intrinsic period of the circadian clock? response. *Science*, 288(5469):1174–1175, 2000.

IG Darvey and PJ Staff. Stochastic Approach to First-Order Chemical Reaction Kinetics. *The Journal of Chemical Physics*, 44(3):990–997, 1966.

James Durbin and Siem Jan Koopman. A simple and efficient simulation smoother for state space time series analysis. *Biometrika*, 89(3):603–616, 2002.

James Durbin and Siem Jan Koopman. *Time Series Analysis by State Space Methods*. OUP Oxford, 2012.

Ryosuke Enoki, Shigeru Kuroda, Daisuke Ono, Mazahir T Hasan, Tetsuo Ueda, Sato Honma, and Ken-ichi Honma. Topological specificity and hierarchical network of the circadian calcium rhythm in the suprachiasmatic nucleus. *Proceedings of the National Academy of Sciences*, 109(52):21498–21503, 2012.

Jennifer A Evans, Tanya L Leise, Oscar Castanon-Cervantes, and Alec J Davidson. Dynamic interactions mediated by nonredundant signaling mechanisms couple circadian clock neurons. *Neuron*, 80(4):973–983, 2013.

Paul Fearnhead, Vasilieos Giagos, and Chris Sherlock. Inference for reaction networks using the linear noise approximation. *Biometrics*, 70(2):457–466, 2014.

Martin Feinberg. *Foundations of Chemical Reaction Network Theory*. Springer, 2019.

Bärbel Finkenstädt, Dan J Woodcock, Michal Komorowski, Claire V Harper, Julian RE Davis, Mike RH White, and David A Rand. Quantifying intrinsic and extrinsic noise in gene transcription using the linear noise approximation: An application to single cell data. *The Annals of Applied Statistics*, 7(4):1960–1982, 2013.

Andrew Gelman and Donald B Rubin. Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4):457–472, 1992.

Andrew Gelman, John B Carlin, Hal S Stern, David B Dunson, Aki Vehtari, and Donald B Rubin. *Bayesian Data Analysis*. CRC press, 2013.

Daniel T Gillespie. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81(25):2340–2361, 1977.

Daniel T Gillespie. A rigorous derivation of the chemical master equation. *Physica A: Statistical Mechanics and its Applications*, 188(1-3):404–425, 1992.

Daniel T Gillespie. The chemical Langevin equation. *The Journal of Chemical Physics*, 113(1):297–306, 2000.

Andrew Golightly, Daniel A Henderson, and Chris Sherlock. Delayed acceptance particle MCMC for exact inference in stochastic kinetic models. *Statistics and Computing*, 25 (5):1039–1055, 2015.

Lei Gong and James M Flegal. A practical sequential stopping rule for high-dimensional Markov chain Monte Carlo. *Journal of Computational and Graphical Statistics*, 25(3): 684–700, 2016.

Didier Gonze and Wassim Abou-Jaoudé. The Goodwin model: behind the Hill function. *PloS One*, 8(8), 2013.

Didier Gonze, Samuel Bernard, Christian Waltermann, Achim Kramer, and Hanspeter Herzel. Spontaneous synchronization of coupled circadian oscillators. *Biophysical Journal*, 89(1):120–129, 2005.

Brian C Goodwin. Oscillatory behavior in enzymatic control processes. *Advances in Enzyme Regulation*, 3:425–437, 1965.

Daniel A Griffith and Giuseppe Arbia. Detecting negative spatial autocorrelation in georeferenced random variables. *International Journal of Geographical Information Science*, 24(3):417–437, 2010.

Changgui Gu, Huijie Yang, and Jos HT Rohling. Dissociation between two subgroups of the suprachiasmatic nucleus affected by the number of damped oscillated neurons. *Physical Review E*, 95(3):032302, 2017.

Heikki Haario, Eero Saksman, and Johanna Tamminen. An adaptive Metropolis algorithm. *Bernoulli*, 7(2):223–242, 2001.

Jack K Hale and Sjoerd M Verduyn Lunel. *Introduction to Functional Differential Equations*. Springer, 2013.

Yoshihiko Hasegawa and Masanori Arita. Circadian clocks optimally adapt to sunlight for reliable synchronization. *Journal of The Royal Society Interface*, 11(92):20131018, 2014.

Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, 2009.

MH Hastings, M Brancaccio, and ES Maywood. Circadian pacemaking in cells and circuits of the suprachiasmatic nucleus. *Journal of Neuroendocrinology*, 26(1):2–10, 2014.

Michael H Hastings, Elizabeth S Maywood, and John S O'Neill. Cellular circadian pacemaking and the role of cytosolic rhythms. *Current Biology*, 18(17):R805–R815, 2008.

W Keith Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57:97–109, 1970.

Glenn E Healey and Raghava Kondepudy. Radiometric CCD camera calibration and noise estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(3):267–276, 1994.

Erik D Herzog, Tracey Hermanstyne, Nicola J Smyllie, and Michael H Hastings. Regulating the suprachiasmatic nucleus (scn) circadian clockwork: interplay between cell-autonomous and circuit-level mechanisms. *Cold Spring Harbor Perspectives in Biology*, 9(1):a027706, 2017.

Peter J Huber. *Robust Statistics*. John Wiley & Sons, 2004.

Pasquale F Innominato, Véronique P Roche, Oxana G Palesh, Ayhan Ulusakarya, David Spiegel, and Francis A Lévi. The circadian timing system in clinical oncology. *Annals of Medicine*, 46(4):191–207, 2014.

Robert P Irwin and Charles N Allen. Neuropeptide-mediated calcium signaling in the suprachiasmatic nucleus network. *European Journal of Neuroscience*, 32(9):1497–1506, 2010.

Galin L Jones. On the Markov chain central limit theorem. *Probability Surveys*, 1 (299-320):5–1, 2004.

Rudolph E Kalman. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1):35–45, 1960.

Rudolph E Kalman and Richard S Bucy. New results in linear filtering and prediction theory. *Journal of Basic Engineering*, 83(1):95–108, 1961.

Rafail Khasminskii. *Stochastic Stability of Differential Equations*. Springer, 2011.

Hiroaki Kitano. Towards a theory of biological robustness. *Molecular Systems Biology*, 3(1):137, 2007.

Peter E Kloeden and Eckhard Platen. *Numerical Solution of Stochastic Differential Equations*. Springer, 1992.

K Kloppstech. Diurnal and circadian rhythmicity in the expression of light-induced plant nuclear messenger rnas. *Planta*, 165(4):502–506, 1985.

Caroline H Ko and Joseph S Takahashi. Molecular components of the mammalian circadian clock. *Human Molecular Genetics*, 15(suppl 2):R271–R277, 2006.

Michał Komorowski, Bärbel Finkenstädt, Claire V Harper, and David A Rand. Bayesian inference of biochemical kinetic parameters using the linear noise approximation. *BMC Bioinformatics*, 10(1):343, 2009.

Ronald J Konopka and Seymour Benzer. Clock mutants of Drosophila melanogaster. *Proceedings of the National Academy of Sciences*, 68(9):2112–2116, 1971.

Anja Korenčič, Grigory Bordyugov, Damjana Rozman, Marko Goličnik, and Hanspeter Herzel. The interplay of cis-regulatory elements rules circadian rhythms in mouse liver. *PLoS One*, 7(11):e46835, 2012.

Thomas G Kurtz. Limit theorems for sequences of jump Markov processes. *Journal of Applied Probability*, 8(2):344–356, 1971.

Thomas G Kurtz. The relationship between stochastic and deterministic models for chemical reactions. *The Journal of Chemical Physics*, 57(7):2976–2978, 1972.

Duncan Lee. A comparison of conditional autoregressive models used in Bayesian disease mapping. *Spatial and Spatio-temporal Epidemiology*, 2(2):79–89, 2011.

Saul C Leite and Ruth J Williams. A constrained Langevin approximation for chemical reaction networks. *The Annals of Applied Probability*, 29(3):1541–1608, 2019.

Emmanuel Lesaffre and Andrew B Lawson. *Bayesian Biostatistics*. John Wiley & Sons, 2012.

Francis Lévi. Circadian chronotherapy for human cancers. *The Lancet Oncology*, 2(5): 307–315, 2001.

Francis Lévi, Rachid Zidani, Silvano Brienza, Luigi Dogliotti, Bruno Perpoint, Mathieu Rotarski, Yves Letourneau, Jean-François Llory, Philippe Chollet, Annick Le Rol, et al. A multicenter evaluation of intensified, ambulatory, chronomodulated chemotherapy with oxaliplatin, 5-fluorouracil, and leucovorin as initial treatment of patients with metastatic colorectal carcinoma. *Cancer*, 85(12):2532–2540, 1999.

Yan Li, Yongli Shan, Ravi V Desai, Kimberly H Cox, Leor S Weinberger, and Joseph S Takahashi. Noise-driven cellular heterogeneity in circadian periodicity. *Proceedings of the National Academy of Sciences*, 2020.

Gabriella B Lundkvist, Yongho Kwak, Erin K Davis, Hajime Tei, and Gene D Block. A calcium flux is required for circadian rhythm generation in mammalian pacemaker neurons. *Journal of Neuroscience*, 25(33):7682–7686, 2005.

John Mallet-Paret and Hal L Smith. The Poincaré-Bendixson theorem for monotone cyclic feedback systems. *Journal of Dynamics and Differential Equations*, 2(4):367–421, 1990.

Frank J Massey Jr. The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American Statistical Association*, 46(253):68–78, 1951.

MATLAB. *version 9.6.0.1135713 (R2019a)*. The MathWorks Inc., Natick, Massachusetts, 2019.

Elizabeth S Maywood, Akhilesh B Reddy, Gabriel KY Wong, John S O'Neill, John A O'Brien, Douglas G McMahon, Anthony J Harmar, Hitoshi Okamura, and Michael H Hastings. Synchronization and maintenance of timekeeping in suprachiasmatic circadian clock cells by neuropeptidergic signaling. *Current Biology*, 16(6):599–605, 2006.

Elizabeth S Maywood, Johanna E Chesham, John A O'Brien, and Michael H Hastings. A diversity of paracrine signals sustains molecular circadian cycling in suprachiasmatic nucleus circuits. *Proceedings of the National Academy of Sciences*, 108(34):14306–14311, 2011.

Harley H McAdams and Adam Arkin. Stochastic mechanisms in gene expression. *Proceedings of the National Academy of Sciences*, 94(3):814–819, 1997.

David McBride and Linda Petzold. Model-based inference of a directed network of circadian neurons. *Journal of Biological Rhythms*, 33(5):515–522, 2018.

Donald A McQuarrie. Stochastic approach to chemical kinetics. *Journal of Applied Probability*, 4(3):413–478, 1967.

JH Meijer and WJ Rietveld. Neurophysiology of the suprachiasmatic circadian pacemaker in rodents. *Physiological Reviews*, 69(3):671–707, 1989.

Nicholas Metropolis, Arianna W Rosenbluth, Marshall N Rosenbluth, Augusta H Teller, and Edward Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 1953.

Jennifer A Mohawk and Joseph S Takahashi. Cell autonomy and synchrony of suprachiasmatic nucleus circadian oscillators. *Trends in Neurosciences*, 34(7):349–358, 2011.

Nicholas AM Monk. Oscillatory expression of Hes1, p53, and NF-κB driven by transcriptional time delays. *Current Biology*, 13(16):1409–1413, 2003.

Brian Moore. PCA and ICA Package v.2.2.0.0, 2018. URL `https://uk.mathworks.com/matlabcentral/fileexchange/38300-pca-and-ica-package`. Online; accessed 31 August 2018.

Wataru Nakamura, Shin Yamazaki, Nana N Takasu, Kazuo Mishima, and Gene D Block. Differential response of Period 1 expression within the suprachiasmatic nucleus. *Journal of Neuroscience*, 25(23):5481–5487, 2005.

Peter Neal, Gareth Roberts, et al. Optimal scaling for partially updating MCMC algorithms. *The Annals of Applied Probability*, 16(2):475–515, 2006.

Cristian Pasarica and Andrew Gelman. Adaptively scaling the Metropolis algorithm using expected squared jumped distance. *Statistica Sinica*, pages 343–364, 2010.

S Pauls, NC Foley, DK Foley, J LeSauter, MH Hastings, ES Maywood, and R Silver. Differential contributions of intra-cellular and inter-cellular mechanisms to the spatial and temporal architecture of the suprachiasmatic nucleus circadian circuitry in wild-type, cryptochrome-null and vasoactive intestinal peptide receptor 2-null mutant mice. *European Journal of Neuroscience*, 40(3):2528–2540, 2014.

Donald B Percival and Andrew T Walden. *Spectral Analysis for Physical Applications*. Cambridge University Press, 1993.

Giovanni Petris, Sonia Petrone, and Patrizia Campagnoli. *Dynamic Linear Models with R*. Springer, 2009.

Angela Relógio, Pal O Westermark, Thomas Wallach, Katja Schellenberg, Achim Kramer, and Hanspeter Herzel. Tuning the mammalian circadian clock: robust synergy of two loops. *PLoS Computational Biology*, 7(12):e1002309, 2011.

Steven M Reppert and David R Weaver. Coordination of circadian timing in mammals. *Nature*, 418(6901):935–941, 2002.

Christian Robert and George Casella. *Monte Carlo Statistical Methods*. Springer, 2004.

Gareth O Roberts and Jeffrey S Rosenthal. Optimal scaling for various Metropolis-Hastings algorithms. *Statistical Science*, 16(4):351–367, 2001.

Gareth O Roberts and Jeffrey S Rosenthal. Coupling and ergodicity of adaptive Markov chain Monte Carlo algorithms. *Journal of Applied Probability*, 44(2):458–475, 2007.

Gareth O Roberts and Jeffrey S Rosenthal. Examples of adaptive MCMC. *Journal of Computational and Graphical Statistics*, 18(2):349–367, 2009.

Gareth O Roberts, Andrew Gelman, and Walter R Gilks. Weak convergence and optimal scaling of random walk Metropolis algorithms. *The Annals of Applied Probability*, 7 (1):110–120, 1997.

Jeffrey S Rosenthal. Optimal proposal distributions and adaptive MCMC. *Handbook of Markov Chain Monte Carlo*, 4(10.1201), 2011.

David Schnoerr, Guido Sanguinetti, and Ramon Grima. The complex chemical Langevin equation. *The Journal of Chemical Physics*, 141(2):07B606_1, 2014.

Fred Schweppe. Evaluation of likelihood functions for Gaussian signals. *IEEE Transactions on Information Theory*, 11(1):61–70, 1965.

Lauren P Shearman, Sathyanarayanan Sriram, David R Weaver, Elizabeth S Maywood, Inẽs Chaves, Binhai Zheng, Kazuhiko Kume, Cheng Chi Lee, Michael H Hastings, and Steven M Reppert. Interacting molecular loops in the mammalian circadian clock. *Science*, 288(5468):1013–1019, 2000.

Shu-qun Shi, Tasneem S Ansari, Owen P McGuinness, David H Wasserman, and Carl Hirschie Johnson. Circadian disruption leads to insulin resistance and obesity. *Current Biology*, 23(5):372–381, 2013.

Hal L Smith. *An Introduction to Delay Differential Equations with Applications to the Life Sciences*, volume 57. Springer, 2011.

Michael V Sofroniew and Harry V Vinters. Astrocytes: biology and pathology. *Acta Neuropathologica*, 119(1):7–35, 2010.

Vassilios Stathopoulos and Mark A Girolami. Markov chain Monte Carlo inference for Markov jump processes via the linear noise approximation. *Philosophical Transactions of the Royal Society A*, 371(1984):20110541, 2013.

Hal S Stern and Noel Cressie. Posterior predictive model checks for disease mapping models. *Statistics in Medicine*, 19(17-18):2377–2397, 2000.

Petre Stoica and Randolph L Moses. *Spectral Analysis of Signals*. Pearson Prentice Hall Upper Saddle River, NJ, 2005.

Steven H Strogatz. *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering*. CRC Press, 2018.

Joseph S Takahashi. Molecular neurobiology and genetics of circadian rhythms in mammals. *Annual Review of Neuroscience*, 18(1):531–553, 1995.

Stephanie R Taylor, Thomas J Wang, Daniel Granados-Fuentes, and Erik D Herzog. Resynchronization Dynamics Reveal that the Ventral Entrains the Dorsal Suprachiasmatic Nucleus. *Journal of Biological Rhythms*, 32(1):35–47, 2017.

Gerald Teschl. *Ordinary Differential Equations and Dynamical Systems*, volume 140. American Mathematical Soc., 2012.

Luke Tierney. Markov chains for exploring posterior distributions. *The Annals of Statistics*, pages 1701–1728, 1994.

Waldo R Tobler. A computer movie simulating urban growth in the Detroit region. *Economic Geography*, 46(sup1):234–240, 1970.

John J Tyson. Biochemical oscillations. In *Computational Cell Biology*, pages 230–260. Springer, 2002.

N.G. van Kampen. *Stochastic Processes in Physics and Chemistry*. North Holland, 2006.

Christina Vasalou, Erik D Herzog, and Michael A Henson. Small-world network models of intercellular coupling predict enhanced synchronization in the suprachiasmatic nucleus. *Journal of Biological Rhythms*, 24(3):243–254, 2009.

Dootika Vats and James M Flegal. Lugsail lag windows and their application to MCMC. *arXiv preprint arXiv:1809.04541*, 2018.

Dootika Vats and Christina Knudson. Revisiting the Gelman-Rubin Diagnostic. *arXiv preprint arXiv:1812.09384*, 2018.

Dootika Vats, James M Flegal, and Galin L Jones. Multivariate output analysis for Markov chain Monte Carlo. *Biometrika*, 106(2):321–337, 2019.

Edward WJ Wallace. A simplified derivation of the linear noise approximation. *arXiv preprint arXiv:1004.4280*, 2010.

Lance A Waller, Bradley P Carlin, Hong Xia, and Alan E Gelfand. Hierarchical Spatio-Temporal Mapping of Disease Rates. *Journal of the American Statistical Association*, 92(438):607–617, 1997.

Alexis B Webb, Stephanie R Taylor, Kurt A Thoroughman, Francis J Doyle III, and Erik D Herzog. Weakly circadian cells improve resynchrony. *PLoS Computational Biology*, 8(11):e1002787, 2012.

David K Welsh, Joseph S Takahashi, and Steve A Kay. Suprachiasmatic nucleus: cell autonomy and network properties. *Annual Review of Physiology*, 72:551–577, 2010.

Pål O Westermark and Hanspeter Herzel. Mechanism for 12 hr rhythm generation by the circadian clock. *Cell Reports*, 3(4):1228–1238, 2013.

Pål O Westermark, David K Welsh, Hitoshi Okamura, and Hanspeter Herzel. Quantification of circadian rhythms in single cells. *PLoS Computational Biology*, 5(11), 2009.

Christopher K Wikle. Hierarchical Bayesian models for predicting the spread of ecological processes. *Ecology*, 84(6):1382–1394, 2003.

Darren J Wilkinson. *Stochastic Modelling for Systems Biology*. CRC press, 2011.

Mae L Woods, Miriam Leon, Ruben Perez-Carrasco, and Chris P Barnes. A statistical approach reveals designs for the most robust stochastic gene oscillators. *ACS Synthetic Biology*, 5(6):459–470, 2016.

Shun Yamaguchi, Hiromi Isejima, Takuya Matsuo, Ryusuke Okura, Kazuhiro Yagita, Masaki Kobayashi, and Hitoshi Okamura. Synchronization of cellular clocks in the suprachiasmatic nucleus. *Science*, 302(5649):1408–1412, 2003.

Seung-Hee Yoo, Shin Yamazaki, Phillip L Lowrey, Kazuhiro Shimomura, Caroline H Ko, Ethan D Buhr, Sandra M Siepka, Hee-Kyung Hong, Won Jun Oh, and Ook Joon Yoo. PERIOD2::LUCIFERASE real-time reporting of circadian dynamics reveals persistent circadian oscillations in mouse peripheral tissues. *Proceedings of the National Academy of Sciences*, 101(15):5339–5346, 2004.

Bokai Zhu, Qiang Zhang, Yinghong Pan, Emily M Mace, Brian York, Athanasios C Antoulas, Clifford C Dacso, and Bert W OMalley. A Cell-Autonomous Mammalian 12 hr Clock Coordinates Metabolic and Stress Rhythms. *Cell Metabolism*, 25(6): 1305–1319, 2017.

Tomasz Zielinski, Anne M Moore, Eilidh Troup, Karen J Halliday, and Andrew J Millar. Strengths and limitations of period estimation methods for circadian data. *PloS One*, 9(5), 2014.