Available online at www.sciencedirect.com

ScienceDirect

IFAC Papers Online CONFERENCE PAPER ARCHIVE

IFAC-PapersOnLine 49-26 (2016) 190–193

Analysis of Circadian Cellular Bioluminescence Recordings using a Kalman Smoother

Bharath Ananthasubramaniam *

* *Institute for Theoretical Biology and Charité Universitätsmedizin, Philippstr. 13 Haus 4, 10115 Berlin, Germany (e-mail: bharath.ananth@charite.de).*

Abstract: Circadian clocks permit mammals to adapt to their periodic natural environment. These clocks generate rhythms in the expression of genes and proteins in most tissues within the organism. Real-time bioluminescence recordings of gene or protein expression in an important tool to study the clock under different manipulations and conditions. Here, we present a tool to extract useful characteristics of the cellular rhythms, such as period and damping rate, using maximum-likelihood estimation in a stochastic damped oscillator model. The tool uses the expectation-maximization algorithm in combination with a Kalman Smoother to perform the joint state and parameter estimation. We apply this tool to quantify the differences between rhythms in the master circadian clock in the brains of two different knockout mice and compare the results to a standard autocovariance fitting-based approach for parameter estimation.

Keywords: Biomedical systems, Kalman filters, Parameter estimation, Smoothing filters, Maximum likelihood estimator

1. INTRODUCTION

Many organisms in nature have developed a biological clock in order to anticipate periodic changes in their environment due to the rotation of the earth (Dunlap et al., 2004). Therefore, the internal 'circadian' clocks adapt to the prevailing external day-night light environment, although the rhythms persist even in the absence of the external cues. In mammals, the basis for this clock is a transcriptional-translation feedback loop involving certain *clock genes*, whose protein products repress their own transcription (Reppert and Weaver, 2002). This results in a classic delay negative-feedback loop that generates rhythms. The master circadian pacemaker in mammals is situated in the suprachiasmatic nucleus (SCN) in the brain (Herzog, 2007). In fact, most tissues in mammals have the necessary components to oscillate and many exhibit rhythms.

One standard approach to real-time monitoring of oscillations in cells (Yamaguchi et al., 2003) involves expressing an enzyme called luciferase under the control of the promoter of the gene of interest or as a fusion-protein product of the gene. This enzyme catalyzes the oxidation of luciferin (a substrate) provided to the cellular medium in the presence of ATP to produce light by means of bioluminescence. This bioluminescence is recorded using a photo-multiplier tube to amplify the weak light emitted by single cells. The goal of real-time monitoring is to measure changes in the oscillation characteristics either between cells/tissue from different genetic backgrounds or the effects of pharmacological manipulation. Systematically quantifying parameters such as period, damping rate and system noise intensity is the focus of this work. Here we propose an expectation-maximization (EM) algorithm (Dempster et al., 1977)-based on Kalman smoothing to jointly estimate the system parameters and states.

2. MODEL OF OBSERVED RHYTHMS IN TISSUES

We are primarily interested in estimating the period and damping rate of the bioluminescence rhythms in the measured tissue. Thus, we adopt a simple stochastic damped oscillator model to represent rhythms in the tissue. Such a model was motivated from observing oscillations in the recordings that always damp out with time. This damping is a result of a combination of the loss of synchrony between oscillators in different cells in the tissue and the depletion of the luciferin in the medium:

$$\begin{aligned} \dot{x} &= -\lambda x + \omega y + \xi_x \\ \dot{y} &= -\omega x - \lambda y + \xi_y, \end{aligned} \tag{1}$$

where $\tau = 2\pi/\omega$ is the period and λ, the damping rate of the oscillations and ξ_x, ξ_y are independent $N(0, \sigma^2)$. The bioluminescence recording is modeled as noisy observation of the state variables (without loss of generality in this symmetric set-up):

$$z = x + y + \zeta, \tag{2}$$

where $\zeta \sim N(0, \sigma_z^2)$.

Recordings are typically made by integrating the emitted bioluminescence over a period of time ΔT (in photo-multiplier tubes) to compensate for the light intensities

⋆ BA acknowledges the support of BMBF (US-German Computational Neuroscience Grant, 01GQ1503), DFG (SPP InKomBio) and Bernstein Center for Computational Neuroscience Berlin (grant 01GQ1001C)

10.1016/j.ifacol.2016.12.124

much lower than in fluorescent reporters and hence low signal-to-noise ratios. It is thus convenient to use the discrete equivalents of (1) and (2), i.e.,

$$\overbrace{\begin{bmatrix} x_{n+1} \\ y_{n+1} \end{bmatrix}}^{\boldsymbol{x}_{n+1}} = \overbrace{e^{-\lambda \Delta T} \begin{bmatrix} \cos(\omega \Delta T) & -\sin(\omega \Delta T) \\ \sin(\omega \Delta T) & \cos(\omega \Delta T) \end{bmatrix}}^{\boldsymbol{F}} \overbrace{\begin{bmatrix} x_n \\ y_n \end{bmatrix}}^{\boldsymbol{x}_n} + \begin{bmatrix} \xi_x \\ \xi_y \end{bmatrix}$$

$$z_{n+1} = \overbrace{\begin{bmatrix} 1 & 1 \end{bmatrix}}^{\boldsymbol{H}} \begin{bmatrix} x_{n+1} \\ y_{n+1} \end{bmatrix} + \zeta, \tag{3}$$

where the new noise variances (without loss of generality) are $\xi_x, \xi_y \sim N(0, \sigma^2)$ and $\zeta \sim N(d, \sigma_z^2)$. However we have often observed that bioluminescence recordings have a decaying baseline due to some changes in the cell viability and the loss of chemical reactants, such as ATP and luciferin, in the medium (see Fig. 1, for example). We augment the system (3) with a simple AR(1) model representing the mean:

$$u_{n+1} = e^{-\alpha \Delta T} u_n + \epsilon,$$

with $\epsilon \sim N(0, \sigma_u^2)$ and α is the decay rate of the baseline. The updated observation model is thus

$$z_{n+1} = x_{n+1} + y_{n+1} + u_{n+1} + \zeta.$$

The initial state and its covariance are $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, respectively.

3. ESTIMATION OF MODEL PARAMETERS

We adopt an EM algorithm to estimate the model parameters $\Theta = (\omega, \lambda, \sigma^2, \sigma_z^2, \alpha, \sigma_u^2, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ (following Shumway and Stoffer (2013)). The complete model log-likelihood is

$$\begin{aligned} \log L = &-\frac{1}{2}\log|\Sigma| - \frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) - \frac{N}{2}\log((\sigma^2)^2) \\ &- \frac{1}{2\sigma^2}\sum_{k=1}^{N}(\boldsymbol{x}_k - \boldsymbol{F}\boldsymbol{x}_{k-1})^{\mathrm{T}}(\boldsymbol{x}_k - \boldsymbol{F}\boldsymbol{x}_{k-1}) \\ &- \frac{N}{2}\log\sigma_u^2 - \frac{1}{2\sigma_u^2}\sum_{k=1}^{N}(u_k - e^{-\alpha\Delta T}u_{k-1})^2 \\ &- \frac{N}{2}\log\sigma_z^2 - \frac{1}{2\sigma_z^2}\sum_{k=1}^{N}(z_k - \boldsymbol{H}\boldsymbol{x}_k - u_k)^2. \end{aligned}$$

The expected log-likelihood (from the E-step) is

$$\begin{aligned} Q(\Theta) = E(\log L) = &-\frac{1}{2}\log|\Sigma| - \frac{1}{2}\mathrm{tr}\left\{\Sigma^{-1}(\boldsymbol{P}_{0|N} \right. \\ &\left. + (\hat{\boldsymbol{x}}_{0|N} - \boldsymbol{\mu})(\hat{\boldsymbol{x}}_{0|N} - \boldsymbol{\mu})^{\mathrm{T}})\right\} - \frac{N}{2}\log((\sigma^2)^2) \\ &- \frac{1}{2\sigma^2}\mathrm{tr}\left\{\boldsymbol{E}_{k,k} - 2\boldsymbol{F}\boldsymbol{E}_{k-1,k} + e^{-2\lambda\Delta T}\boldsymbol{E}_{k-1,k-1}\right\} \\ &- \frac{1}{2\sigma_u^2}\mathrm{tr}\left\{E^u_{k,k} - 2e^{-2\alpha\Delta T}E^u_{k-1,k} + e^{-2\alpha\Delta T}E^u_{k-1,k-1}\right\} \\ &- \frac{N}{2}\log\sigma_u^2 - \frac{N}{2}\log\sigma_z^2 \\ &- \frac{1}{2\sigma_z^2}\sum_{k=1}^{N}(z_k - \boldsymbol{H}\hat{\boldsymbol{x}}_{k|N} - \hat{u}_{k|N})^2 + \boldsymbol{H}\boldsymbol{P}_{k|N}\boldsymbol{H}^{\mathrm{T}}, \end{aligned}$$

where we define the following matrices based on the Kalman smoother (Rauch-Tung-Striebel) (Rauch et al., 1965) estimates $(\hat{\boldsymbol{x}}_{k|N}, \boldsymbol{P}_{k|N}, \boldsymbol{P}_{k-1,k|N}, \hat{u}_{k|N}, P^u_{k|N}, P^u_{k-1,k|N})$ at the current parameter estimate $\theta^{(n)}$:

$$\begin{aligned} \boldsymbol{E}_{k,k} &= \sum_{k=1}^{N}\boldsymbol{P}_{k|N} + \hat{\boldsymbol{x}}_{k|N}\hat{\boldsymbol{x}}_{k|N}^{\mathrm{T}} \\ \boldsymbol{E}_{k-1,k-1} &= \sum_{k=1}^{N}\boldsymbol{P}_{k-1|N} + \hat{\boldsymbol{x}}_{k-1|N}\hat{\boldsymbol{x}}_{k-1|N}^{\mathrm{T}} \\ \boldsymbol{E}_{k-1,k} &= \sum_{k=1}^{N}\boldsymbol{P}_{k-1,k|N} + \hat{\boldsymbol{x}}_{k-1|N}\hat{\boldsymbol{x}}_{k|N}^{\mathrm{T}}, \\ E^u_{k,k} &= \sum_{k=1}^{N}P^u_{k|N} + \hat{u}^2_{k|N} \\ E^u_{k-1,k-1} &= \sum_{k=1}^{N}P^u_{k-1|N} + \hat{u}^2_{k-1|N} \\ E^u_{k-1,k} &= \sum_{k=1}^{N}P^u_{k-1,k|N} + \hat{u}_{k-1|N}\hat{u}_{k|N}. \end{aligned}$$

Setting the derivative to zero to maximize expected log-likelihood (the M-step) yields the following updates to the model parameters (i.e., $\theta^{(n+1)}$ from $\theta^{(n)}$):

$$\boldsymbol{\mu}^{(n+1)} = \hat{\boldsymbol{x}}_{0|N} \tag{4}$$

$$\boldsymbol{\Sigma}^{(n+1)} = \boldsymbol{P}_{0|N} \tag{5}$$

$$\lambda^{(n+1)} = -\frac{1}{\Delta T}\log\frac{\mathrm{tr}(\boldsymbol{S})}{\mathrm{tr}(\boldsymbol{E}_{k-1,k-1})} \tag{6}$$

$$\sigma^{2^{(n+1)}} = \frac{1}{2N}\left(\mathrm{tr}(\boldsymbol{E}_{k,k}) - \frac{\mathrm{tr}(\boldsymbol{S})^2}{\mathrm{tr}(\boldsymbol{E}_{k-1,k-1})}\right) \tag{7}$$

$$\alpha^{(n+1)} = -\frac{1}{\Delta T}\log\frac{E^u_{k-1,k}}{E^u_{k-1,k-1}} \tag{8}$$

$$\sigma_u^{2^{(n+1)}} = \frac{1}{N}\left(E^u_{k,k} - \frac{(E^u_{k-1,k})^2}{E^u_{k-1,k-1}}\right) \tag{9}$$

$$\begin{aligned} \sigma_z^{2^{(n+1)}} = &\frac{1}{N}\sum_{k=1}^{N}(z_k - \boldsymbol{H}\hat{\boldsymbol{x}}_{k|N} - \hat{u}_{k|N})^2 + \boldsymbol{H}\boldsymbol{P}_{k|N}\boldsymbol{H}^{\mathrm{T}} \\ &+ P^u_{k|N} \end{aligned} \tag{10}$$

$$\boldsymbol{F}^{(n+1)} = \frac{\mathrm{tr}(\boldsymbol{S})}{\mathrm{tr}(\boldsymbol{E}_{k-1,k-1})}VU^{\mathrm{T}}$$

$$\omega^{(n+1)} = \tan^{-1}\frac{\boldsymbol{F}_{12}^{(n+1)}}{\boldsymbol{F}_{11}^{(n+1)}} \tag{11}$$

where $\boldsymbol{E}_{k-1,k}^{\mathrm{T}} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^{\mathrm{T}}$.

4. RESULTS

4.1 Performance of the proposed method

We applied our analysis approach to recordings of SCN slice cultures from mice that expressed luciferace as a fusion protein with the clock protein PER2 and were heterozygote or homozygote knockouts for a sub-unit of the G-protein signaling receptor G_o (Lauffenburger and Linderman, 1995). We had data for eight SCNs each for the homozygous and heterozygous genotype, which were of differing durations, but we retained only the first six days for all SCNs for uniformity of analysis.

The magnitude, amplitude and the baseline of the recordings were very variable between SCNs even from the same