CIRCADIAN RHYTHMS IN PLANTS: MACHINE LEARNING MODELS AND BIOINFORMATICS APPROACHES

Daniel Trejo Banos, Institute for Adaptive and Neural Computation, School of Informatics
Supervisors: Guido Sanguinetti, Andrew J. Millar and Christopher M. Bishop

d.trejo-banos@sms.ed.ac.uk

http://www.inf.ed.ac.uk/people/students/Daniel_Trejo_Banos.html

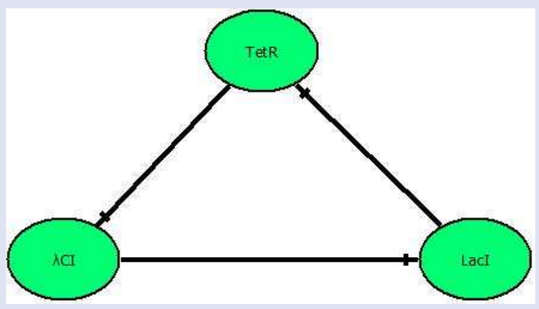
Introduction

Circadian Rhythms are characterized as endogenous and self sustaining oscillators with a period of approximately 24 hours (temperature compensated)[1].

The molecular basis of the circadian oscillator is currently described as many components interlocked in a set of negative and positive feedback loops[1]. They evolved to react and anticipate daily light-dark cycles, in order to control gene expression that regulates metabolic and behavioural processes. These oscillations are present in many kinds of living organisms (eukaryotes and cyanobacteria). To describe these oscillations, modelling tools such as ODE, differential equations with delays, Boolean networks and State Space models among others are used. *Ostreococcus Tauri* is a *pico alga*[3], a species of unicellular plant; it has been described as the smallest known free-living eukaryote. As such it has a compact genome, so is a good stepping stone into the study of plants. At the heart of its oscillator, two genes have been identified[4], but the exact mechanism and configuration is still under study[5]. Finally, there are non-transcriptional oscillations, which are present in the absence of transcription factor expression. The exact function and origin of these oscillations is still a great source of speculation[7].

Research questions and methods

One of the first models to be tested experimentally was the Repressillator, a three gene oscillator present in fruit flies. The figure shows a simplified representation of the Repressillator, the first repressor gene TetR inhibits the transcription of the second repressor gene which in turns inhibit the expression of the third gene. Finally the third gene inhibits TetR production, closing the feedback loop[2].



Repressillator model

The repressilator can be expressed in general terms using a system of nonlinear ODE.

$$\dot{m}_1 = v_1 \frac{k_{1,2}^n}{k_{1,2}^n + p_2^n} - k_1^m m \quad (1)$$

$$\dot{m}_2 = v_2 \frac{k_{2,3}^n}{k_{2,3}^n + p_3^n} - k_2^m m \quad (2)$$

$$\dot{m}_3 = v_3 \frac{k_{3,1}^n}{k_{3,1}^n + p_1^n} - k_3^m m \quad (3)$$

$$\dot{p}_1 = k_1 m_1 - k_1^p p \quad (4)$$

$$\dot{p}_2 = k_2 m_2 - k_2^p p \quad (5)$$

$$\dot{p}_3 = k_3 m_3 - k_3^p p \quad (6)$$

With v_i being the maximum rate of transcription of gene i, $k_{i,j}$ is the concentration of protein p_j at which gene i reaches half of it's maximum transcription rate, k_i^p and k_i^m are the degradation rates of protein and mRNA of gene i and k_i is the translational constant.

Results

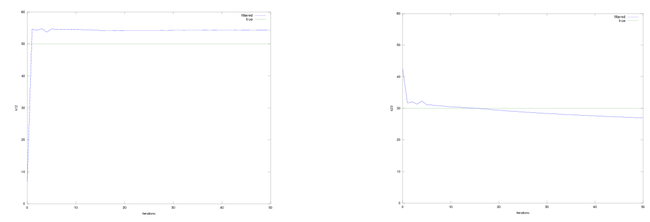
The research just started, the objective will consist in use computational and mathematical methodologies, in order to help biologists study and analyse data produced by circadian rhythm experiments involving *O. Tauri* and other plant organisms. Currently I am evaluating schemes for parameter estimation using Kalman filter methods for state space models. Some preliminary results include the application of a UKF filter to estimate some parameters over the classic Repressillator model presented in the previous section, using a slight modification to the algorithm presented in [6]. The UKF is a Kalman Filter variaton, its purpose is to estimate the continuous hidden state X_t of a state space model of the form:

$$\theta_t = \theta_{t-1} \quad (7)$$

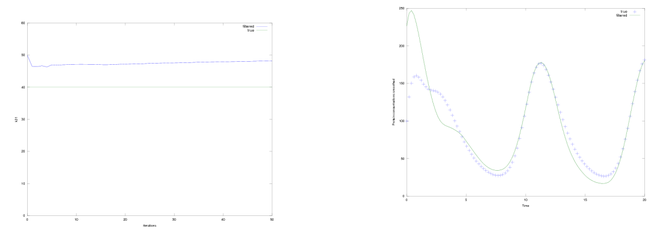
$$X_t = F\left(X_{t-1}; \theta_{t-1}\right) + \eta_t^x \quad (8)$$

$$Y_t = H\left(X_{t-1}; \theta_{t-1}\right) + \eta_t^y \quad (9)$$

by reducing the error between the predicted estimate $F(X_{t-1}; \theta_{t-1})$ and the observations Y_t; subject to process and observation noise . Because the studied model is non-linear, the predicted estimate is obtained using a method called *Unscented Transform*, that produces a set of points, called sigma points. These points are then propagated through the non-linear dynamics and used to calculate an approximation to the true mean and variance of the hidden state. The following figures show some estimations of constants $k_{i,j}$ and protein concentrations using a UKF filter-smoothing process with a joint "state-parameter" variable. It has shown to be a fast estimation process, but prone to find local minima.



$k_{1,2}$ $k_{2,3}$



$k_{3,1}$ *Protein 1 concentratios*

Future Work

Ongoing work includes the development of a tool to apply any model to the inference scheme. An SBML representation of the models will be used. SBML is a XML dialect focusing in modelling of biological and chemical reactions. Future perspectives include the application of machine learning models to study interactions between the components of the *O. Tauri* oscillator and to identify possible components of the non-transcriptional oscillator.

Sponsors

The research is being funded by Microsoft Research.

Microsoft® Research

References

[1] S. Brody.2011 Academic press Elsevier,

[2] M. Elowitz and S. Leibler.2000 Nature,403, 335-338.

[3] A. Monnier et al.2010 BMC Genomics, 11:192.

[4] P. Morant et al.2010 Chaos.

[5] W. Huang et al.2010 Science 336, 75-79

[6] M. Quach et al.2009 Bioinformatics 23, 3209-3216.

[7] ONeill et al. 2011.Nature 469, 7331: 554558.

School of informatics