



Higham, C. F., and Husmeier, D. (2015) Inference of circadian regulatory pathways based on delay differential equations. In: Ortuno, F. and Ignacio, R. (eds.) *Bioinformatics and Biomedical Engineering*. Series: *Lecture Notes in Computer Science*, 9044 (9044). Springer, pp. 468-478. ISBN 9783319164793

Copyright © 2015 Springer International Publishing Switzerland.

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

The content must not be changed in any way or reproduced in any format or medium without the formal permission of the copyright holder(s)

When referring to this work, full bibliographic details must be given

<http://eprints.gla.ac.uk/107139/>

Deposited on: 9 June 2015

Enlighten – Research publications by members of the University  
of Glasgow <http://eprints.gla.ac.uk>

# Inference of Circadian Regulatory Pathways based on Delay Differential Equations

Catherine F. Higham and Dirk Husmeier

School of Mathematics and Statistics, College of Science and Engineering,  
University of Glasgow, Glasgow G12 8QQ, Scotland, UK  
{Catherine.Higham,Dirk.Husmeier}@glasgow.ac.uk

**Abstract.** Inference of circadian regulatory network models is highly challenging due to the number of biological species and non-linear interactions. In addition, statistical methods that require the numerical integration of the data model are computationally expensive.

Using state-of-the-art adaptive gradient matching methods which model the data with Gaussian processes, we address these issues through two novel steps. First, we exploit the fact that, when considering gradients, the interacting biological species can be decoupled into sub-models which contain fewer parameters and are individually quicker to run. Second, we substantially reduce the complexity of the network by introducing time delays to simplify the modelling of the intermediate protein dynamics.

A Metropolis-Hastings scheme is used to draw samples from the posterior distribution in a Bayesian framework. Using a recent delay differential equation model describing circadian regulation affecting physiology in the mouse liver, we investigate the extent to which deviance information criterion can distinguish between under-specified, correct and over-specified models.

**Keywords:** Bayesian Inference, Gaussian Processes, Adaptive Gradient Matching, Circadian Regulation, Delay Differential Equations

## 1 Introduction

### 1.1 Biological Background/Motivation

The circadian clock is a molecular mechanism, involving interlocked, transcriptional feedback loops, that synchronises biological processes with the day/night cycle and is found in many organisms, see [19]. Mathematical models are being developed to describe the dynamics of the clock transcriptional network and its downstream regulation, for *Arabidopsis*, see [11], [12], and for the mouse liver and adrenal gland, see [8], [9]. The field is now sufficiently mature for us to consider validating, comparing and extending these models in the presence of experimental data.

## 1.2 Network inference

Statistical pathway inference techniques aim to make inference in a network where the vertices are molecular components such as genes or gene products and the edges represent regulatory interactions between these components. Statistical models for exploring large spaces are typically linear for reasons of speed but come at the cost of over-simplifying the non-linear features of the network. When the network is known, differential equations (DEs) are widely used to model biochemical dynamics and capture a wealth of detail about the network. Fitting approaches which directly solve the DEs, see [17], are currently infeasible for large systems especially when model comparison is required. Hence, recent work, to select between network models focuses on an intermediate approach, incorporating prior knowledge about the structure of biochemical DE models into an inference framework, see [10].

Differential equation models are used extensively in science and engineering. A common requirement is to estimate model parameters by fitting them to observed data collected over time. This involves repeatedly finding a solution to the DEs which, because these systems are typically non-linear, involves numerical approximation. Numerical integration is computationally expensive, in all but the simplest cases, and hence, there is much interest in methods that avoid this step. One alternative approach focuses on gradient matching with Gaussian Processes (GPs). This is currently a very active area (*e.g.* [2], [4], [18], [7]).

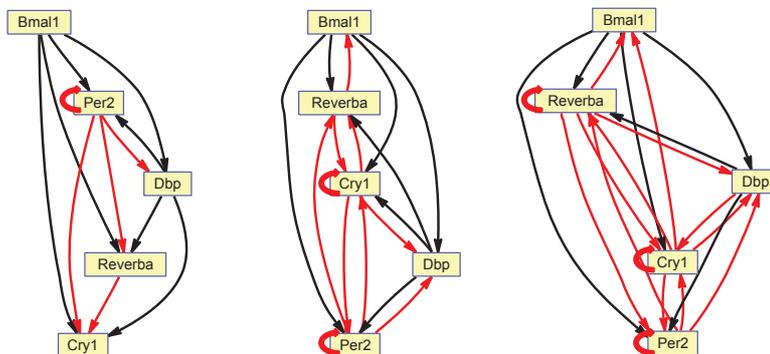
Gradient matching uses an alternative model of the data, an interpolant, and matches the derivative of this interpolant with the DE outputs, thus avoiding explicit numerical integration, see [13]. GPs are a natural choice for the alternative data model, and in particular they admit exact derivative expressions. A GP may be fitted to the data and the DE parameters found by matching the GP derivative, for which there is an analytical expression, to the DE derivative. However the accuracy of the original method proposed in [2], was limited by the lack of regularisation from the DE parameter inference to the GP inference. In work by [4], all parameters are consistently inferred in the context of the whole model, rather than in a piecewise heuristic manner. This introduces, in effect, a coupling mechanism between the Gradient Process and the DEs which enhances the learning of the parameters associated with the DEs.

Here, we contribute to this field by developing novel methodological advances and illustrating them on a core clock model for mouse liver and adrenal gland developed by [9]. This model comprises five clock genes and is based on expression and experimentally verified circadian *cis*-regulatory sites, see figure 1. The intermediate protein dynamics are modelled using time delays, vastly simplifying the network complexity. The expression of each clock gene is described by a delay differential equation with a production term that depends on the concentrations of core clock regulatory components and a decay term. The adaptive gradient matching (AGM) statistical model developed in [4] is our framework for Bayesian inference. We introduce modularisation by exploiting the fact that when gradient matching, the system reduces to five equations, one for each species, which

are no longer coupled (as the right-hand side does not require any of the other left-hand side values).

A Metropolis-Hastings scheme is devised to sample from the posterior probability densities for the model parameters (including the parameters for the DEs and the hyper-parameters for the GPs).

We generate data from the full model ( $M_0$ ), see [9], using the parameters described by the authors. The aim of this work is to investigate whether deviation information criterion (DIC) which considers both the measure of fit and the measure of complexity based on the posterior samples, can distinguish between under-specified, correct and over-specified models for a range of alternative hypothesis.



**Fig. 1.** Regulatory network Model  $M_0$  (*middle*) containing five core clock genes (*boxes*). Direction of activation (*bold black line*) or inhibition (*regular red line*) is indicated by the (*arrowhead*). Regulatory network Model  $M_r$  with reduced number of edges (*left*). Regulatory network Model  $M_a$  with added number of edges (*right*).

## 2 Methods

We describe here our approach to Bayesian model fitting and selection for systems of differential equations (ordinary and time delay) using adaptive gradient matching with a Gaussian process.

### 2.1 Reaction Graph

Consider a regulatory network of genes which may activate or inhibit transcription directly or indirectly. This network can be considered as a reaction graph where the nodes are the molecular components such as genes or gene products and the edges are the regulatory interactions between these components. Changes over time in the gene and gene product states can be modelled using

differential equations. To simplify the modelling of the intermediate protein dynamics we introduce time-delayed variables. This reduces the network to gene transcripts and is convenient for parameter estimation and model selection since this data is more readily available than protein data. Proteins are sometimes treated as missing data and modelled using latent variables; employing time-delayed variables is an alternative approach [8].

Consider a set of  $T$  arbitrary time points  $t_1 < \dots < t_T$  and  $K$  gene transcription states. We define  $\mathbf{x}_k \equiv \mathbf{x}_k(t) \equiv [x_k(t_1), x_k(t_2), \dots, x_k(t_T)]^\top$  as the transcription state sequence for the  $k$ th state.

## 2.2 The Dynamical Model

Let  $K$  denote the number of genes in the network and  $i$ , where  $i = 1 \dots N_k$ , the associated set of  $N_k$  regulatory genes. The time series for the  $k$ th gene transcription state is represented by a set of  $K$  differential equations of the form

$$\dot{\mathbf{x}}_k(t) \equiv \frac{d\mathbf{x}_k(t)}{dt} = \mathbf{f}_k(\mathbf{x}_k(t), \mathbf{x}_i(t), \boldsymbol{\theta}).$$

Here  $\boldsymbol{\theta} = \{\boldsymbol{\theta}_1 \dots \boldsymbol{\theta}_K\}$  is our general notation for the parameters of the differential equations. Specifically, we have

$$\mathbf{f}_k(\mathbf{x}_k(t), \mathbf{x}_i(t), \boldsymbol{\theta}_k) = \prod_{i=1 \dots N_k} \left( \frac{1 + av_i \mathbf{x}_i(t_{\tau_i})/a_i}{1 + \mathbf{x}_i(t_{\tau_i})/a_i} \right)^{p_i} - d_k \mathbf{x}_k(t). \quad (1)$$

In this notation,  $p_i$  is the number of clock-controlled elements and  $(1/a_i)^{p_i}$  represents the basal production rate of species  $i$ . When species  $i$  is an activator,  $av_i$  scales the activation and when species  $i$  is a repressor,  $av_i = 0$ . We use  $t_{\tau_i}$  to indicate the value  $t - \tau_i$ , thus accounting for the delay including translation, post-translational modifications, complex formation and nuclear translocation. The parameter  $d_k$  is the degradation rate of species  $k$ .

## 2.3 The Observation Model

Let  $\mathbf{y}_k(t) = \mathbf{x}_k(t) + \epsilon(t)$  be noisy observations of this process, where  $\epsilon(t)$  is assumed to have a zero-mean Gaussian distribution with variance  $\sigma_k^2$  for each of the model states. Assuming independence over the observation times, we have an observation model

$$p(\mathbf{y}_k | \mathbf{x}_k) = \prod_t p(\mathbf{y}_k(t) | \mathbf{x}_k(t), \sigma_k^2) = \prod_t \mathcal{N}(\mathbf{x}_k(t), \sigma_k^2 \mathbf{I}). \quad (2)$$

Here  $\mathcal{N}(\mathbf{x}_k(t), \sigma_k^2 \mathbf{I})$  denotes the probability density function for a Gaussian random variable with mean  $\mathbf{x}_k(t)$  and variance  $\sigma_k^2 \mathbf{I}$ .

## 2.4 Adaptive Gradient Matching

Instead of obtaining  $\mathbf{x}_k$  by solving the dynamical system, recent gradient matching approaches in [2] and [4], put a GP prior on  $\mathbf{x}_k$  and match the GP derivatives with the derivatives arising from the differential equations. This approach leads naturally to a decoupling of  $\mathbf{x}_k$  and a reduction in complexity for parameter estimation.

## 2.5 Gaussian Process Modelling

A Gaussian Process (GP) is a stochastic process governing the properties of a function. A GP is defined by a mean and correlation function called a kernel. In this work we use the most commonly used kernel, the "squared-exponential" kernel. With a GP prior on  $\mathbf{x}_k$ ,  $p(\mathbf{x}_k|\phi_k) = N(\mathbf{x}_k|\mu_k, \mathbf{C}_{\phi_k})$  where  $\mu_k$  is the data mean,  $\mathbf{C}_{\phi_k}$  is the correlation function and  $\phi_k$  are the hyper-parameters of the GP. The derivative of a GP is also a GP and the conditional distribution for the state derivatives is  $\mathcal{N}(\mathbf{m}_k, \mathbf{A}_k)$  where  $\mathbf{m}_k = {}'\mathbf{C}_{\phi_k} \mathbf{C}_{\phi_k}^{-1} (\mathbf{x}_k - \mu_k)$  and  $\mathbf{A}_k = \mathbf{C}_{\phi_k}'' - {}'\mathbf{C}_{\phi_k} \mathbf{C}_{\phi_k}^{-1} \mathbf{C}_{\phi_k}'$ , respectively, see [15]. Here, the matrix  $\mathbf{C}_{\phi_k}''$  denotes the auto-covariance for each state derivative, and the matrices  $\mathbf{C}_{\phi_k}'$  and  ${}'\mathbf{C}_{\phi_k}$  denote the cross-variances between the  $k$ th state and its derivative. For further details concerning the derivation and analytical form of these expressions, see [4].

## 2.6 Statistical Model

Bayes' theorem links what we would like to know, the posterior probability distributions for the unknown parameters, to the likelihood of seeing the observations given the model and its parameters and the prior information about the parameters. Following [4] we propose the following adaptive gradient matching (AGM) model over states  $\mathbf{x}_k$ , their derivatives  $\dot{\mathbf{x}}_k$ , observations  $\mathbf{y}_k$  and parameters associated with the DE,  $\boldsymbol{\theta}_k$ , with the observational noise,  $\sigma_k^2$ , and with the GP,  $\phi_k$ ,

$$p(\mathbf{y}_k, \mathbf{x}_k, \boldsymbol{\theta}_k, \phi_k, \gamma_k^2, \sigma_k^2) = p(\mathbf{y}_k|\mathbf{x}_k, \sigma_k^2)p(\mathbf{x}_k|\boldsymbol{\theta}_k, \phi_k, \gamma_k^2)p(\boldsymbol{\theta}_k, \phi_k, \gamma_k^2, \sigma_k^2). \quad (3)$$

The term  $p(\mathbf{x}_k|\boldsymbol{\theta}_k, \phi_k, \gamma_k^2)$  combines the DE gradient with the GP gradient in a compatibility function and arises from a products of experts approach described in [4]

$$p(\mathbf{x}_k|\boldsymbol{\theta}_k, \phi_k, \gamma_k^2) \propto \frac{\exp[-1/2(\mathbf{f}_k - \mathbf{m}_k)^\top (\mathbf{A}_k + \gamma_k^2 \mathbf{I})^{-1} (\mathbf{f}_k - \mathbf{m}_k)]}{(2\pi)^{n/2} |\mathbf{A}_k + \gamma_k^2 \mathbf{I}|^{1/2}}. \quad (4)$$

The function  $\mathbf{f}_k$  is defined in equation (1),  $\mathbf{A}_k$  and  $\mathbf{m}_k$  are defined in section 2.5,  $n$  is the number of time points (also equal to the number of rows in  $\mathbf{A}_k$ ),  $\gamma_k$  is the slack parameter controlling the coupling of GP and DE, and  $\mathbf{I}$  is the identity matrix.

## 2.7 Sampling

We use a Metropolis-Hastings scheme to draw samples from the posterior distribution. Denoting  $q_1(\Theta_k)$  and  $q_2(x_k)$ , where  $\Theta_k = \{\theta_k, \phi_k, \gamma_k^2, \sigma_k^2\}$ , as the proposal distributions for the parameters,  $\Theta_k$ , and the states,  $\mathbf{x}_k$ , the proposal moves are accepted or rejected according to the standard Metropolis-Hastings criteria

$$P_{accept} = \min \left\{ 1, \frac{\pi(\mathbf{y}_k, \mathbf{x}'_k, \Theta'_k)}{\pi(\mathbf{y}_k, \mathbf{x}_k, \Theta_k)} \right\} \quad (5)$$

where  $\pi = \frac{p(\mathbf{y}_k, \mathbf{x}_k, \theta_k, \phi_k, \gamma_k^2, \sigma_k^2)}{q_1(\Theta_k)q_2(\mathbf{x}_k)}$  and the numerator is defined in equation (3).

The parameters  $\Theta_k$  are proposed simultaneously from a multivariate Gaussian using the efficient adaptive MCMC algorithm described by [5], adapted for our non standard posterior distribution. The priors for all parameters are informed gamma priors. Here,  $\mathbf{x}_k$  and  $\phi_k$  are initialised using a GP regression fit with maximum likelihood to the data  $\mathbf{y}_k$ , see [14]. The proposal function for  $\mathbf{x}_k$  is  $\mathcal{N}(\boldsymbol{\mu}_{\mathbf{x}|\mathbf{y}}, \boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{y}})$ , dropping the subscript  $k$  for convenience, and where  $\boldsymbol{\mu}_{\mathbf{x}|\mathbf{y}} = (\mathbf{C}_\phi^{-1} + (\sigma^2 I)^{-1})^{-1}(\sigma^2 I)^{-1}\mathbf{y}$ ,  $\boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{y}} = (\mathbf{C}_\phi^{-1} + (\sigma^2 I)^{-1})^{-1}$ .

## 2.8 Model Selection using Deviance Information Criterion

For competing parametric statistical models, the deviance information criterion (DIC), as described in [16], considers both the measure of fit and the measure of complexity. The deviance (associated with the observed likelihood  $p(\mathbf{y}|\theta)$ ) is defined by

$$D(\theta) = -2 \log p(\mathbf{y}|\theta) + 2 \log h(\mathbf{y}),$$

where  $h(\mathbf{y})$  depends only on the data. The model complexity or effective dimension,  $p_D$ , is defined as

$$p_D = \bar{D}(\theta) - D(\bar{\theta}),$$

and

$$\begin{aligned} DIC &= \bar{D}(\theta) + p_D, \\ &= 2\bar{D}(\theta) - D(\bar{\theta}) \end{aligned}$$

where  $\bar{D}(\theta)$  is the expected value of  $D(\theta)$  and  $\bar{\theta}$  is the expected value of  $\theta$ . For model comparison, we set  $h(\mathbf{y}) = 1$  for all models so that  $D(\theta) = -2 \log p(\mathbf{y}|\theta)$ . For  $D(\theta)$  available in closed form,  $\bar{D}(\theta)$  can be approximated from the MCMC run by taking the sample mean of the simulated values of  $D(\theta)$ .

In our case, as  $\mathbf{y}$  is conditioned on  $\theta$  and  $\mathbf{X} = \mathbf{x}_1, \dots, \mathbf{x}_k$ , we use the complete DIC suggested in [3]

$$DIC(\mathbf{y}, \mathbf{X}) = -4\mathbb{E}[\log p(\mathbf{y}, \mathbf{X}|\theta) | \mathbf{y}, \mathbf{X}] + 2 \log p(\mathbf{y}, \mathbf{X} | \mathbb{E}[\theta | \mathbf{y}, \mathbf{X}]).$$

The intuitive idea is that models with a smaller DIC score are preferred to models with a larger DIC score. Models are penalised by the value of  $\bar{D}$  but also (in common with other information criteria) by the effective number of parameters  $p_D$ . Since  $\bar{D}$  will decrease as the number of parameters in a model increases, the  $p_D$  term compensates for this effect by favouring models with a smaller number of parameters.

An advantage of DIC, over other criteria such as Bayes factors, is that DIC is easily calculated from the samples generated by a Markov chain Monte Carlo simulation.

### 3 Results

#### 3.1 Mouse Liver Model

The model selection framework is applied to the five component clock model for mouse liver and adrenal gland developed in [9]. Data is generated from this delay differential equation model using numerical integration over an interval of 24 hours with the published parameter set. Clean data were then sampled in 2 hour intervals and corrupted with additive Gaussian noise which corresponds to a signal-to-noise ratio of 10, see figure 2. In this experiment, we fit a GP to each of the five time series in order to provide an initial estimate for the GP hyper-parameters,  $\phi_k = \{l_k, sf_k\}$ . Here  $l_k$  is the length scale parameter and  $sf_k$  is the vertical scale parameter. The GP fitting, illustrated in figure 2, also provides an initial estimate for  $\sigma_k^2$ .

The AGM framework combined with the Metropolis-Hastings scheme, outlined in section 2.7, is used to obtain posterior samples for the parameters of the statistical model comprising DE parameters  $\theta_k = \{ap_i, av_i, d_k, \tau_i\}$ , see equation (1) for more details, GP parameters,  $\phi_k$ , the noise parameter  $\sigma_k^2$  and the slack parameter,  $\gamma_k^2$ , associated with the AGM framework, see equation (4). In total, for five network species (Bmal1, Rev-erba, Per2, Cry1 and Dbp), see figure 2, there are 54 parameters to be learnt. Note that  $p_i$ , the number of clock controlled elements in the  $x_k$  regulatory region, are taken as given and not inferred in these experiments. Details of the priors used for each parameter type are given in section 2.7. Two MCMC chains were run for  $2 \times 10^6$  iterations and convergence was monitored using the potential scale reduction factor (PSRF) discussed in [1].

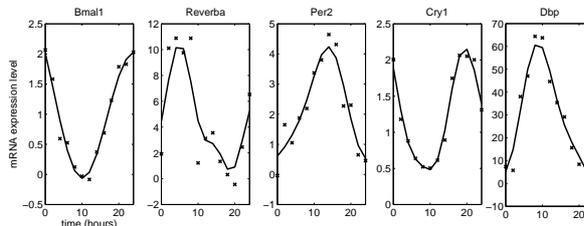
#### 3.2 Model Selection Experiment

To illustrate these new tools in the context of model selection, we propose two alternative models  $M_r$  (one edge per species is *removed*) and  $M_a$  (one edge per species is *added* up to a maximum of five edges) to the true model,  $M_0$ , and then apply the inference framework to each model using the dataset described above. The specific details as to which edges were removed, added or changed are described in table 1. The DIC score for each model was estimated using 10,000 posterior samples taken from the end of the MCMC chains. Each entry

in the table required approximately 2 hours of CPU time on a HPC cluster. Models with a smaller DIC are preferred to models with a larger DIC. For comparison between models, the scores for the alternative models are adjusted by subtracting the value arising from the fitting of the true model  $M_0$ . A positive adjusted score indicates that the true model is preferred to the alternative model and a negative adjusted score indicates that the alternative model is preferred to the true model. For this experiment, the DIC differences for  $M_r$  were positive for all species (Reverba (adj DIC=0.1), Per2 (adj DIC=0.9), Cry1 (adj DIC=8.6) and Dbp (adj DIC=21.4)) indicating that the true model is preferred to the alternative model,  $M_r$  for all species, strongly for Cry1 and Dbp, and weakly for Reverba and Per2. Bmal1 was not included in this experiment as it only has one interaction in the true model. Over all 5 species, the total adjusted DIC score is 31 suggesting that the true model is preferred to an alternative model with edges missing as outlined for  $M_r$ , see table 1. For  $M_a$ , the DIC differences are negative for Bmal1 but positive for Reverba, Per2, Cry1 and Dbp (adjusted DICs -1.8, 0.6, 18.7 and 25.4 respectively), see table 1. The total adjusted DIC score is 44.7 suggesting that the true model is preferred to the specified model  $M_a$ .

### 3.3 Parameter Estimation

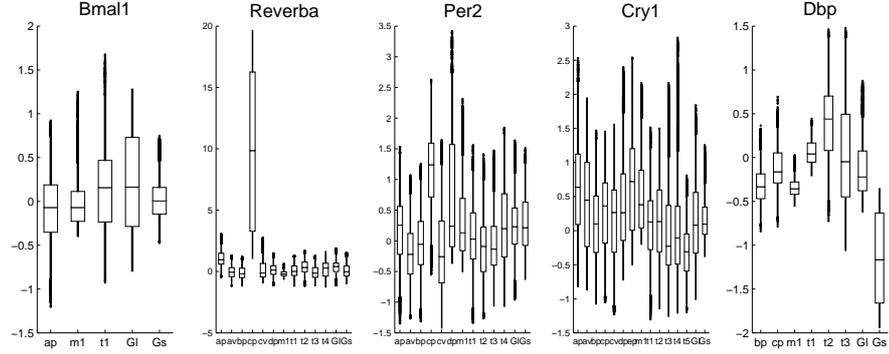
Comparison of the posterior densities for the model parameters between the species, see figure 3, suggests that uncertainty increases with the number of parameters. Generally recovery of the true value (comparison possible when using synthetic data) is good with the distributions lying over the true value. The occasional parameter is very different. In the case of parameter  $cp$  for Reverba, this is explained by over-fitting to the noisy data. The method allows for further investigation of individual differences or deviations from the true value.



**Fig. 2.** Synthetic expression level time series data *crosses* generated for five species over 24 hours and the initial GP fit *line*.

## 4 Conclusion

We present here new tools for model selection in gene regulatory networks, with an emphasis on improving computational efficiency for large-scale simulations.



**Fig. 3.** Posterior probability densities for model parameters shown as box plot distributions. The values have been shifted so that "0" signifies the true parameter.

**Table 1.** Adjusted DIC scores for the true model  $M_0$  and alternative models  $M_r$  and  $M_a$ . The production term (listed below) summarises for each species, the direction of the edges acting on that species. For example act1\*rep3\*rep4 is short-hand for the the activation of Bmal1(1), and the repression of Per2(3) and Cry1(4) on Dbp,  $M_0$

Species	$M_0$	$M_r$	$M_a$
Bmal1	0	n/a	-1.8
Reverba	0	0.1	1.8
Per2	0	0.9	0.6
Cry1	0	8.6	18.7
Dbp	0	21.4	25.4
Total	0	31.0	44.7

Production terms	$M_0$	$M_r$	$M_a$
Bmal1 (1)	rep2	n/a	rep1*rep4
Reverba (2)	act1*rep3*act5*rep4	act1*rep3*act5	act1*rep3*act5*rep4*rep2
Per2 (3)	act1*rep3*act5*rep4	act1*rep3*act5	act1*rep3*act5*rep4*rep2
Cry1 (4)	act1*rep3*act5*rep2*rep4	act1*rep3*act5*rep2	act1*rep3*rep5*rep2*rep4
Dbp (5)	act1*rep3*rep4	act1*rep3	act1*rep3*rep4*rep2

Model selection for complex networks is a demanding topic that is challenging state-of-the-art methodology. There is an on-going requirement from experimentalists to revise models as more data becomes available and to choose between alternative hypothesis. For example linking clock mechanics to down-stream activities such as metabolism.

For large networks, methods requiring numerical integration are infeasible. Here we use a state-of-the-art gradient matching approach which substantially reduces the computational expense [2]. The relative cost of AGM to numerical integration of the DE, in the examples considered here is conservatively ten

times faster. This improvement makes model selection between several models a realistic objective. Here we looked in detail at one true model and two alternative models. Future work could accommodate many more alternatives.

In terms of model selection, we show that it is possible to distinguish between the true model and a under- or over- specified model where the number of edges have been reduced or added.

Future work will compare the DIC measure with other statistics for model selection and investigate the sensitivity of model selection to factors such as the prior information.

**Acknowledgements.** This work was funded by the EU under the FP7 scheme as part of the TiMeT project. TiMet - Linking the Clock to Metabolism is a Collaborative Project (Grant Agreement 245143) funded by the European Commission FP7, in response to call FP7-KBBE-2009-3.

## References

1. Brooks, S. P. and A. Gelman (1998): General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, 7, 434.
2. Calderhead, B., M. Girolami, and N. D. Lawrence (2009): Accelerating Bayesian inference over nonlinear differential equations with Gaussian processes. in D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou, eds., *Advances in Neural Information Processing Systems 21*, 217224.
3. Celeux, G., F. Forbes, C. P. Robert, and M. Titterton (2003): Deviance information criteria for missing data models.
4. Dondelinger, F., D. Husmeier, S. Rogers, and M. Filippone (2013): ODE parameter inference using adaptive gradient matching with Gaussian processes. *Proceedings of the Sixteenth International Conference on Artificial Intelligence and Statistics*, 216–228.
5. Haario, H., M. Laine, A. Mira, and E. Saksman (2006): DRAM: Efficient adaptive MCMC. *Statistics and Computing*, 16, 339–354.
6. Higham, C. F. and D. Husmeier (2013): A Bayesian approach for parameter estimation in the extended clock gene circuit of *Arabidopsis thaliana*. *BMC Bioinformatics*, 14, S3.
7. Holsclaw, T., B. Sansó, H. K. H. Lee, K. Heitmann, S. Habib, D. Higdon, and U. Alam (2012): Gaussian process modeling of derivative curves. *Technometrics*, 55, 57–67.
8. Korenčič, A., G. Bordyugov, R. Košir, D. Rozman, M. Goličnik, and H. Herzel (2012): The interplay of cis-regulatory elements rules circadian rhythms in mouse liver. *PLoS ONE*, 7, e46835.
9. Korenčič, A., R. Košir, G. Bordyugov, R. Lehmann, D. Rozman, and H. Herzel (2014): Timing of circadian genes in mammalian tissues. *Scientific Reports*, 4.
10. Oates, C. J., F. Dondelinger, N. Bayani, J. Korola, J. W. Gray, and S. Mukherjee (2014): Causal network inference using biochemical kinetics. *arXiv:1406.0063 [stat]*.
11. Pokhilko, A., A. P. Fernández, K. D. Edwards, M. M. Southern, K. J. Halliday, and A. J. Millar (2012): The clock gene circuit in *Arabidopsis* includes a repressilator with additional feedback loops. *Molecular Systems Biology*, 8, 574.

12. Pokhilko, A., P. Mas, and A. J. Millar (2013): Modelling the widespread effects of TOC1 signalling on the plant circadian clock and its outputs. *BMC Systems Biology*, 7, 23, PMID: 23506153.
13. Ramsay, J. O., G. Hooker, D. Campbell, and J. Cao (2007): Parameter estimation for differential equations: a generalized smoothing approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69, 741–796.
14. Rasmussen, C. E. and H. Nickish (2010): Gaussian processes for machine learning (GPML) toolbox. *Journal of Machine Learning Research*, 11, 3011–3015.
15. Solak, E., R. Murray-Smith, W. Leithead, C. Rasmussen, and D. Leith (2003): Derivative observations in gaussian process models of dynamic systems. in *Advances in Neural Information Processing Systems (NIPS)*, 1033–1040.
16. Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. Van Der Linde (2002): Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64, 583–639.
17. Vyshemirsky, V. and M. A. Girolami (2008): Bayesian ranking of biochemical system models. *Bioinformatics*, 24, 833–839.
18. Wang, Y. and D. Barber (2014): Gaussian processes for Bayesian estimation in ordinary differential equations. *JMLR: W&CP*, volume 32.
19. Zhang, E. E. and S. A. Kay (2010): Clocks not winding down: unravelling circadian networks. *Nature Reviews. Molecular Cell Biology*, 11, 764–776.