Statistical analysis of nonlinear dynamical systems using differential geometric sampling methods

Ben Calderhead* and Mark Girolami

*Department of Statistical Science, University College London, Gower Street,
London WC1E 6BT, UK*

Mechanistic models based on systems of nonlinear differential equations can help provide a quantitative understanding of complex physical or biological phenomena. The use of such models to describe nonlinear interactions in molecular biology has a long history; however, it is only recently that advances in computing have allowed these models to be set within a statistical framework, further increasing their usefulness and binding modelling and experimental approaches more tightly together. A probabilistic approach to modelling allows us to quantify uncertainty in both the model parameters and the model predictions, as well as in the model hypotheses themselves. In this paper, the Bayesian approach to statistical inference is adopted and we examine the significant challenges that arise when performing inference over nonlinear ordinary differential equation models describing cell signalling pathways and enzymatic circadian control; in particular, we address the difficulties arising owing to strong nonlinear correlation structures, high dimensionality and non-identifiability of parameters. We demonstrate how recently introduced differential geometric Markov chain Monte Carlo methodology alleviates many of these issues by making proposals based on local sensitivity information, which ultimately allows us to perform effective statistical analysis. Along the way, we highlight the deep link between the sensitivity analysis of such dynamic system models and the underlying Riemannian geometry of the induced posterior probability distributions.

Keywords: nonlinear dynamic systems; statistical inference; Markov chain Monte Carlo; Riemann manifold sampling methods; Bayesian analysis; parameter estimation

1. INTRODUCTION

The use of mathematical modelling has long played an important role in describing and predicting the behaviour of natural processes [1–3]. It is only more recently, however, that the use of such models within a statistical framework has allowed modelling and experimental approaches to be more tightly bound together forming an iterative, more symbiotic relationship [4,5]. Mathematical models are abstract representations of reality that are useful for making testable predictions with varying levels of detail. Deterministic nonlinear ordinary differential equations (ODEs), for example, may be most appropriate when describing the average concentration of a protein within a population of cells, in which the stochastic effects of individual molecules do not greatly affect the overall dynamics; this is estimated to be for populations above the range 10^2 – 10^3 molecules per chemical species [6]. For smaller numbers of molecules, the use of stochastic models may be more appropriate [7]. Similarly, when modelling more global physiology, at a tissue or organ

level, for example, it is no longer feasible to describe the dynamics with such molecular detail and a more course-grained approach or multi-scale techniques are often necessary, e.g. [8,9]. The current availability of cellular transcript and proteomic assay techniques [10], which measure average concentrations in cell populations leads us to focus our attention in this paper on continuous deterministic ODE models, which are applicable not just in biology, but also in a large number of other areas of science and engineering, e.g. [5].

The biochemical mechanisms by which even relatively well-known enzymatic networks operate are not always clear [4]. There are often multiple plausible network topologies that are consistent with the known underlying biology, and in this context, the idea of modularity within biology is important as it provides a natural means of iteratively building up a model description alongside successive biological experiments, with each half of the process informing the other. Model hypotheses can thus be constructed and compared with one another by incorporating new components into an existing model and then assessing the model based on the data [11]. Recent advances in understanding the molecular origin of circadian rhythms in *Arabidopsis*

*Author for correspondence (ben@stats.ucl.ac.uk).

One contribution of 9 to a Theme Issue 'Inference in complex systems'.

thaliana [12,13] offer an excellent example of this type of approach, whereby differential equation models have been able to offer predictions that have subsequently been tested experimentally [14].

Mechanistic modelling is a first step towards characterizing the aetiology of many different types of diseases that are thought to be the result of disruptions to the normal functioning of signalling pathways and biochemical networks. Obtaining a detailed *mechanistic* understanding of such enzymatic control processes could have major implications in the study and potential treatment of disease at a molecular level. The circadian clock, for example, appears to play a central role in the physiology of plants and mammals, controlling many important cellular functions and influencing pathways implicated with a variety of diseases [15]; indeed the timing of drug administration relative to circadian rhythms appears to strongly affect the efficacy of certain anti-cancer agents [16]. Signalling pathways are also strongly implicated in the origin of many cancers [17] and understanding the intricate networks of nonlinear interactions is key to being able to predict the impact of biochemical changes within the system, whether natural or artificially induced with the use of drugs [18].

Current investigations into biochemical networks are characterized by complex nonlinear dynamics, as well as by measurement and model uncertainty. When studying the possible structure of such systems, working hypotheses can be encoded as statistical models based on systems of nonlinear differential equations that capture all assumptions regarding the likely mechanisms of interaction. As models increase in size and complexity, so does the need for sophisticated statistical methodology that can consistently evaluate and update the evidence in favour of each model, as new data become available (e.g. [4,19]). Given the limited and variable experimental data that are often available, a probabilistic approach based on Bayesian statistics offers a natural way of dealing with such parameter and model uncertainty. Rather than finding an optimal working set of parameters, as is the case in the frequentist setting [20], the Bayesian paradigm advocates averaging over all possible parameter sets with respect to their individual probabilities. This marginalization procedure automatically provides a compromise between fitting the data and penalizing model complexity, such that we may find the simplest model that is, however, still *complex enough* to accurately describe the observed dynamic behaviour.

Initial proof of concept investigations has shown that a Bayesian approach to model ranking can be very successful [4,11]; however, it is acknowledged that performing Bayesian inference over ODE models is extremely challenging [21]. The procedure is equivalent to evaluating integrals involving a highly nonlinear function over a high-dimensional space. In more than three or four dimensions deterministic approaches are no longer feasible and we must resort to stochastic integration using simulation-based Monte Carlo techniques. A common approach is to construct a Markov process that converges to the target posterior distribution [22]; however, this is not straightforward in practice owing to the following three main issues.

- *Global convergence.* Moving from a random starting position in parameter space to the true stationary distribution may be difficult, owing to the possibility of multiple modes. We require sampling methods that can explore all modes globally with the correct frequency without becoming stuck in local maxima of negligible probability mass.
- *Local mixing.* Once the Markov chain reaches a region of high probability mass, we want it to fully explore and ideally obtain near-uncorrelated samples. This is challenging for many types of differential equation-based models, particularly those that are high dimensional and whose nonlinear dynamics induce very strong nonlinear correlation structures in the posterior distribution.
- *Computational cost.* The likelihood of these models can be expensive to evaluate, since it involves approximately solving the system of ODEs with a numerical integration scheme for each set of proposed parameters. Although generally unavoidable, this cost can be minimized through efficient exploration of the parameter space, which can be measured in terms of effective sample size (ESS), normalized by the overall computational time [23].

The issue of identifiability is very important and has a direct impact on our ability to perform efficient frequentist or Bayesian statistical analysis. It has been noted many times that ODE models of biochemical networks generally exhibit widely varying parameter sensitivities [24–28]; investigation of second-order sensitivities of these models, evaluated at the maximum likelihood, often reveals a wide eigenvalue spectrum that itself may change depending on the point in parameter space at which it is calculated. In settings with such varying parameter scalings, standard Markov chain Monte Carlo (MCMC) samplers generally have very poor mixing properties and produce highly correlated samples [23], resulting in estimates of the required Bayesian quantities with large Monte Carlo errors. This is often a result of structural unidentifiability of the model [20], such that parameters cannot be estimated with low variance. For example, if the output of a model depends strongly only on the *ratio* of two parameters θ_1/θ_2 , then the conditional distributions $p(\theta_1|\theta_2)$ and $p(\theta_2|\theta_1)$ may be well constrained, but their marginal distributions $p(\theta_1)$ and $p(\theta_2)$ might not be.

Recently developed differential geometric MCMC methods, on the other hand, seem particularly well suited to this type of application, since they exploit the natural representation of the model parameter space as a Riemannian manifold [29], which is induced using the expected Fisher information as a metric tensor [23]. The Fisher information, therefore, *defines* a local distance measure and effectively allows the MCMC proposals to be based on the curvature of the manifold, which is directly defined by the parameter sensitivities of the underlying model describing the dynamical system, see §10 in Girolami & Calderhead [23].

We examine the application of manifold MCMC methodology for performing Bayesian inference over ODE models of biologically relevant size with complex dynamics and partially unidentifiable structures. We

perform posterior inference on biologically realistic examples of enzymatic control processes that have recently been studied in the systems biology literature. We consider a model describing the main circadian clock components in *A. thaliana* [12], building up inference over different sets of model parameters to gain insight into the challenges of a statistical analysis, and also a model describing a cell signalling pathway [30]. We perform inference using datasets generated from each model and infer the posterior distributions describing the probable parameter values.

2. METHODS: MODELS AND INFERENCE

2.1. Differential equation models

We consider statistical models based on systems of ODEs of the form $\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), \theta, \mathbf{x}_0)$, which defines a relationship between the time derivatives of the states and, often nonlinear, functions of the states $\mathbf{x} = (x_1, \dots, x_N)^T \in \mathbb{R}_+^N$ and parameters $\theta = (\theta_1, \dots, \theta_M)^T \in \mathbb{R}_+^M$. In all but the very simplest cases, such systems are analytically intractable and must be solved numerically given suitable initial state conditions $\mathbf{x}_0$ at $t = 0$. If we wish to employ sampling methods based on differential geometric information, we must be able to calculate the first, and in some methods second-order sensitivities of the model outputs with respect to the input parameters. One method of obtaining these required sensitivities at all time points is to approximate them using finite differences, however this method may not be very accurate, particularly for higher order sensitivities. A much more accurate approach for ODE models is to calculate them via a forward simulation of auxiliary equations obtained by differentiating the original system with respect to each parameter.¹ The general form of the first-order sensitivities therefore follows as

$$\begin{aligned} \dot{S}_{t,n}^i &= \frac{d}{dt} \frac{\partial x_n(t)}{\partial \theta_i} = \frac{\partial f_n(\mathbf{x}(t), \theta, \mathbf{x}_0)}{\partial \theta_i} \\ &= \sum_{l=1}^N \frac{\partial f_{t,n}}{\partial x_l} S_{t,l}^i + \frac{df_{t,n}}{d\theta_i}, \end{aligned} \quad (2.1)$$

where we have taken the total derivative of $\mathbf{f}(\mathbf{x}(t), \theta, \mathbf{x}_0)$ with respect to each θ_i , since the states $\mathbf{x}$ also depend implicitly on the parameter values. The sensitivity of x_n at time t with respect to a change in the parameter θ_i is therefore given by $S_{t,n}^i$. This gives us a set of linear differential equations to augment the original system.

ODE models for most biological systems can be constructed from a set of common components describing various mechanisms of molecular interaction, such as Michaelis–Menten type enzyme-mediated interactions [6], and their proposed structure can be informed from current knowledge of the underlying biology. The systems we consider illustrate the typical challenges associated with performing statistical inference over realistic biochemical models.

¹For models with a large number of parameters and relatively few states, adjoint differentiation may be used to calculate the sensitivities much more efficiently [31].

2.1.1. Circadian control model

Arabidopsis thaliana is widely accepted in plant biology as a model plant worthy of close and careful study [32]. Although its genetic networks are much simpler than higher organisms, it is still complex enough to offer much insight into a large variety of biologically interesting mechanisms of biochemical interaction, such as negative feedback loops that induce nonlinear oscillatory behaviour. Overviews of the recent advances in understanding this reasonably complex biological system are given in McClung [33] and Hubbard *et al.* [34].

We employ a model based on the core circadian network in *A. thaliana* comprising transcription factors LHY and CCA, and the pseudo-response regulator TOC1. As a simplification, both LHY and CCA are modelled as one component, since they have qualitatively similar behaviour. Michaelis–Menten kinetics are used to describe enzyme-driven protein degradation, with Hill functions describing transcriptional activation of mRNA for LHY/CCA and TOC1. The model is from Locke *et al.* [12] and is the minimal description of the network describing circadian behaviour of *Arabidopsis*, which we simulate in constant darkness. It consists of six nonlinear differential equations and a total of 24 parameters. The concentrations of the species are represented by $[TOC1]$ and $[LHY]$ with subscripts m , c and n denoting mRNA, protein within the cytoplasm and protein within the nucleus, respectively. The Hill coefficients a and b are set to 1 and 2 respectively, based on evidence from the literature [12]. There are therefore 22 free parameters to be inferred from the experimental data, where (n_i, g_i) are transcription rates, (m_i, k_i) are degradation rates, (p_i) is translation rate and (r_i) is rate-defining transport between the nucleus and cytoplasm. The structure of this biochemical network is represented in figure 1 and the equations are given in appendix A.1.

2.1.2. A cell signalling model

The modelling of cell surface receptors is an important means of furthering our understanding of intra-cellular signalling. Such mechanisms allow extra-cellular cues to drive activation and dynamic behaviour within each cell [35]. In particular, changes in the environment are often encoded at a molecular level in terms of changes in the concentration of ligands outside the cell; such ligands may then bind to the surface of cells inducing internal changes brought about by phosphorylation mechanisms. A recent example of such a model is given in Becker *et al.* [30], which describes the nonlinear dynamic behaviour of the *erythropoietin* (Epo) receptor in response to changes in the ligand Epo concentrations outside the cell. Encoded signals sent in this manner have significant biological knock-on effects, invoking cellular responses that activate a number of signalling networks further downstream. The model we employ consists of six nonlinear differential equations, based on mass-action kinetics, with eight parameters, and details of the equations and parameter values are given in appendix A.2. We reparametrize the model such that we have parameters measured in $\log_{10}$ space, allowing us to more

Figure 1. This is a diagrammatic representation of the model developed in Locke *et al.* [12], which we employ to model the main circadian oscillator network in *A. thaliana*, consisting of a negative feedback loop. TOC1 mRNA produces TOC1 protein in the cytoplasm, which is transported into the nucleus and increases the production of LHY/CCA mRNA. This in turn produces LHY/CCA protein in the cytoplasm, which is transported into the nucleus and inhibits the production of TOC1 mRNA. This negative feedback loop is capable of producing oscillatory and highly nonlinear dynamical behaviour, providing a challenge for any optimization or MCMC algorithm. Solid line end with circles, production; solid line end with triangles, transport; solid line end with vertical line, inhibition.

easily consider a wide range of possible parameter values that vary by orders of magnitude. This model is complicated further by the addition of an observation model, since the biochemical species in this network are not all observable individually; in particular, only the total concentration of $[Epo]$ and $[dEpo_e]$ is biologically observable, along with $[Epo_EpoR]$, such that our observations take the form $y_1 = [Epo] + [dEpo_e]$ and $y_2 = [Epo_EpoR]$.

2.2. Parameter identifiability

A model parameter can be considered only weakly identifiable if changes in its value have very little effect on the output of the model. There are two main causes of unidentifiability [20]; the first comes from measurement uncertainty in the data, and the second is a result of the model structure, in particular the mathematical relationship between the parameters and the states. Structural identifiability issues can occur, for example, when the parameters of a statistical model appear as a ratio; if the output of the model depends only on this ratio, then the numerator and denominator may effectively take on an infinite number of values. This effect has been termed parameter evaporation [27] since these unconstrained parameters can tend to infinity. This poses a particular

problem for methods that make use of second-order geometric information, as the unconstrained parameters can result in an ill-conditioned Hessian or Fisher information matrix [36], and so the point-wise variance estimates of such parameters can be very poor.

In the context of the circadian biochemical models above, the commonly used Michaelis–Menten terms are of particular interest [27]. This is a model to describe an irreversible enzymatic reaction and relates the rate of the reaction to the concentration of a substrate, $[S]$. Let us consider the Michaelis–Menten equation

$$r = \frac{r_{\max}[S]}{K + [S]}, \quad (2.2)$$

where r is the current reaction rate $r_{\max}$ is the maximum reaction rate and K is the Michaelis–Menten rate, which is inversely related to the enzyme's affinity for the substrate S . Firstly, let us note that

$$K \ll [S], \quad r \approx r_{\max}, \quad \text{since } \frac{[S]}{K + [S]} \approx 1. \quad (2.3)$$

In this case, when the rate constant K is much less than the concentration level of S , the overall reaction rate is determined by $r_{\max}$. K is, therefore, unidentifiable as it has almost no effect on the model output. Similarly we see that

$$K \gg [S], \quad r \approx \frac{r_{\max}[S]}{K}, \quad (2.4)$$

and so it is the ratio of $r_{\max}$ to K that is important in determining model behaviour. In this case, both $r_{\max}$ and K are potentially unidentifiable, depending on the values of $[S]$. This unidentifiability results from our decision to employ Michaelis–Menten kinetics to describe our system; the fact that parameter values could blow up to infinity or evaporate to zero may not have any bearing on the biological reality, but are more likely to be artefacts of our model approximation of the underlying system. Within a Bayesian setting, the use of priors may be used to enforce weak identifiability and improve numerical conditioning [36].

2.3. The choice of priors

Priors are used in Bayesian statistics as a means of incorporating existing knowledge regarding likely parameter values. In this setting, such knowledge commonly arises from the consideration of the underlying biology and experimental observations, however it may also arise by considering the model itself. In previous work examining oscillatory behaviour of biochemical networks [37], bifurcation analysis of a relatively simple delay differential equation model was employed to find the regions of parameter space that permitted oscillatory output, and this information then informed the choice of priors. This approach, however, is obviously dependent on the model being amenable to such analysis; bifurcation analysis on much larger, more complex models quickly becomes unfeasible. The idea of informing the range of admissible parameter values based on a mathematical analysis of the model has also been suggested in Transtrum *et al.* [27], in particular for the case of examining

Michaelis–Menten kinetics, where the aim was to avoid numerical issues and parameter evaporation. Using such an approach for the circadian rhythm models considered in this paper, we can enforce weak identifiability without restricting the range of behaviour our model can produce. In this case, we want K to be neither too small nor too big, but rather roughly the same order of magnitude as we would expect $[S]$ to be.

In the clock model, we employ gamma priors, with shape parameter $k = 1.5$ and scale parameter $\theta = 2$, for the model parameters. These priors have positive support, reflecting the fact that parameters correspond to physical rate constants, and they discourage parameter values approaching zero, since we assume that all terms in each model play a role in the overall biochemical process. Such priors are also suitable for preventing the rate parameters getting too large, since we assume that all chemical processes are slow enough to be observed. The prior thus serves to constrain the parameter values, and this has the added numerical benefit of regularizing the Fisher information, when it might otherwise be near singular with a very high condition number [36]; the overall metric tensor is therefore formed by taking the expectation of the posterior distribution, as discussed in §4.2 in Girolami & Calderhead [23]. We employ gamma priors, with shape parameter $k = 1.5$ and scale parameter $\theta = 10$, for the initial conditions of the clock model; these were chosen to allow the initial values to be roughly the same order of magnitude as the expected concentration levels of the chemical species, as measured experimentally [12], and they cover the support of the likely initial conditions for each observed species.

In the signalling model, we employ Gaussian priors with a mean of -2 and a standard deviation of 2; since we reparametrize the model in $\log_{10}$ space, we no longer require our prior to have only positive support, and such priors cover parameter values of many orders of magnitude.

2.4. Differential geometric Markov chain Monte Carlo for ordinary differential equation models

Sensitivities of parameters can change depending on the dynamical behaviour of the model, and thus can vary dramatically in different parts of parameter space. This has implications for the successful application of MCMC methods, since the transition kernel must therefore be tuned in different parts of the space, particularly during the transient phase in which the chain may be traversing large regions of space to reach the stationary distribution (see figure 9). In this situation, standard Metropolis–Hastings sampling methods can fail or perform poorly at best. Adaptive MCMC methods [38] also address this issue by attempting to estimate the correlation structure of the target distribution during sampling; however, this is not an easy task particularly as the dimensionality increases. In this section, we introduce MCMC approaches that calculate this local structure *directly* via the Fisher information.

Strong correlations and widely differing variances make it challenging to propose large moves that will

have a high probability of acceptance. In addition, the expense of calculating the likelihood in this setting underlines the importance of obtaining effective proposal mechanisms. Some parameter values may be perturbed by large amounts and yet still have very little effect on the model output; in contrast, other parameters may be extremely sensitive in terms of their effect on model output and must therefore be perturbed by very small amounts. To complicate matters, sometimes the model output may be sensitive to correlated changes in subsets of parameters. However, recent advances in MCMC methodology provide a systematic means of overcoming these difficulties by the design of proposal mechanisms that make use of the local sensitivity information of the system [23].

A statistical model can be thought of as a family of parametrized probability distributions, as defined through the likelihood function, which gives the probability of a particular dataset given a set of parameter values. Work in the 1940s by Jeffreys [39] and Rao [40] sought to define distance between probability distributions. The elegant solution recast probability theory in the language of differential geometry—in particular they found that a natural measure of distance may be defined between probability distributions using the expected Fisher information as a local metric, and the task is equivalent to calculating shortest paths over a Riemannian manifold [29], where the expected second-order sensitivities at a set of parameter values define a local distance measure. We can consider two sets of perturbed parameter values to be similar if they result in similar posterior probabilities.

We begin with an analogy. Let us consider measuring distance on a map of a mountainous region. On the flat two-dimensional map, one can define a distance purely in terms of its Euclidean coordinate system; in real life, however, the actual measure of distance must also take into account changes in the height of the terrain. The greater the variability in the height of the terrain, the less accurate the two-dimensional approximation. Similarly, in our statistical model, any measure of distance based purely on the parameter values is likely to be very misleading; a much more accurate measure takes into account rates of change in the probability of the statistical model, and it is exactly this information that is given in a local sensitivity analysis via the expected Fisher information.

From this geometrical point of view, we see why standard MCMC approaches may fare badly, since they generally employ a Euclidean measure of distance when proposing moves in parameter space. Yet, a small change in each parameter value may in actual fact have widely varying effects on the output, and hence probability, of the statistical model. The Fisher information provides a local measure of the expected second-order sensitivities of a system evaluated at a particular set of parameter values. An eigenvalue decomposition of the Fisher information can elucidate the parameter dependencies within the model; most biochemical networks have a relatively small number of parameter combinations that strongly affect the model output, with the remaining combinations having little effect [20,24,25,41].

The differential geometric MCMC methods developed in Girolami & Calderhead [23] effectively make use of local sensitivity information to propose superior moves in parameter space. In this work, we embed these sampling schemes within a population MCMC framework to help escape local maxima and fully explore the parameter space [21,42,43].

We assume normally distributed additive errors in the experimental data, in which case the expected Fisher information takes the form of the weighted inner product of the sensitivities at each observed time-point such that $\mathbf{G}(\theta)_{ij} = \sum_{n=1}^N \mathbf{S}_{:,n}^{i^T} \Sigma_n^{-1} \mathbf{S}_{:,n}^j$, where $\mathbf{G}(\theta)_{ij}$ denotes the (i,j) th entry in the Fisher information matrix obtained at the parameter values θ . The vector $\mathbf{S}_{:,n}^j$ has elements that correspond to the sensitivity of the n th species with respect to the j th parameter at each of the sampled time points. The error covariance for the n th species at all sampled time points is denoted as Σ_n . The metric tensor of the Riemannian manifold induced by the statistical model is then $\mathbf{G}(\theta) + \Phi(\theta)$, where $\Phi(\theta)$ is the negative Hessian of the log-prior [23], and we employ this in the sampling schemes throughout this paper.

Geodesics are the equivalent of straight lines on a Riemannian manifold, and are formally paths along which the tangent vector is constant. In the case of the Riemannian connection used in the following MCMC algorithms, they also locally define the shortest path between two points on the manifold, allowing them to generate samples very efficiently [23]. The manifold MCMC methods we employ are generalizations of the Langevin and Hamiltonian Monte Carlo methods [44,45]. The first method, the manifold Metropolis-adjusted Langevin algorithm (mMALA) makes proposals by simulating a diffusion process across the manifold. The proposal is a Gaussian distribution, whose mean is given by following the curvature of the space along the direction of steepest natural gradient, and whose covariance is given by a scaled inverse Fisher information matrix. A simplified version of mMALA may also be used by assuming a locally constant metric at each point when making proposals, whose mean then involves a cheaper to compute approximation of the path defined in mMALA. The second method, Riemannian manifold Hamiltonian Monte Carlo (RMHMC) proposes moves by randomly following a geodesic from the current point on the manifold. This is more expensive to compute, as the geodesic paths must be obtained by symplectically integrating a non-separable Hamiltonian system. The reader is directed to Girolami & Calderhead [23] for a full exposition, and accompanying Matlab code, of the sampling methods employed.

2.5. Implementation

One practical challenge of using the Fisher information matrix is that it can become ill-conditioned for parameters that are highly insensitive [36]. The use of a prior helps mediate this problem somewhat and does make the overall metric tensor $\mathbf{G}(\theta) + \Phi(\theta)$ better conditioned, however we will see that this ill-conditioning still negatively impacts on sampling efficiency. The parameters in these models of biochemical networks can be

naturally partitioned according to their functional role, and in the case of the Michaelis–Menten parameters, it is possible to predict where the issue of unidentifiability may arise. We can use these natural parameter groupings to design a blocked MCMC algorithm [46], such that each set of parameters is sampled separately using a manifold MCMC method conditioned on all other parameters; we iterate through exploring the submanifolds induced by our chosen groups of parameters. This blocking strategy is designed to capture a large part of the correlation structure in the system, while improving the numerical stability by using smaller Fisher information matrices. We shall examine the effect of this strategy in §4.

Synthetic datasets are very useful for evaluating the performance of new methodology, and with complex ODE models, it is convenient to have a set of parameter values that generate the data as a benchmark. We investigate the models using a variety of noise levels, which we detail along with the results in §3; ultimately we wish to test our methodology using data with biologically realistic numbers of observations and levels of noise.

For the circadian model, we generate data by sampling 10 data points for each species evenly over a time period of 48 h, using parameter values given in the literature [12]. Using a Gaussian error model, we add zero mean noise to the data points generated for species $LHY_{[m,c,n]}$ and $TOC1_{[m,c,n]}$, with standard deviation proportional to the amplitude of the concentration of each species. The circadian model we investigate has the difficulty that the Fisher information over all parameters is numerically ill-conditioned, as often happens when parameters are weakly identifiable [36]. We therefore employ blocking strategies that are informed by analytical consideration of the model and partition the parameters into subgroups, each containing potentially strong correlation structure. Instead of using one large ill-conditioned Fisher information matrix to inform our sampling proposals, we employ multiple smaller, but better-conditioned, Fisher information matrices.

We embed the sampler within a population framework to help escape local maxima and fully explore the parameter space, noting that the samples obtained could be further used for calculating Bayes factors via thermodynamics integration [21,42,43]. Within this framework, we explore multiple tempered distributions simultaneously, each defined by a power posterior [43]. These form a smooth family of distributions between the prior and posterior, and exchange moves between these distributions allow faster convergence to the global mode of interest. We employ the same temperature schedule as detailed in Calderhead & Girolami [21].

A burn-in phase of between 10 000 and 20 000 samples was found to be sufficiently long for the Markov chains to converge to the stationary distributions defined by each of the power posteriors used. During this time, the step sizes of the parameters were adjusted every 100 iterations to achieve an acceptance ratio of between 20 and 50 per cent for standard Metropolis and between 30 and 70 per cent for manifold sampling using simplified mMALA. After the burn-in period, step sizes were fixed to ensure samples were drawn from the stationary distribution.

The cell signalling model is made identifiable by fixing two parameters, the values of which may be obtained

Figure 2. We compare samples of the linear parameters in the circadian model, which are obtained using a variety of MCMC sampling algorithms; Metropolis–Hastings, simplified mMALA, mMALA and RMHMC. The true parameter values are represented by the grey line. The woeful performance of the Metropolis–Hastings sampler is apparent from the slowly moving, highly correlated samples it produces. By contrast, the samplers that make proposals using local geometric information, as defined by the sensitivities of the ODE system, do far better. RMHMC produces nearly uncorrelated samples; however, this is at great computational expense since the second-order sensitivities must be calculated multiple times over the geodesic proposal path. Similarly mMALA requires the second-order sensitivities to be computed. Simplified mMALA offers performance approaching that of mMALA, at a much reduced computational cost requiring only the first-order sensitivities to be calculated.

experimentally [30] (see appendix A.2). The initial condition of *EpoR* may also be found in the experiment of Becker *et al.* [30]. The initial condition of *Epo* is assumed to be known, and the other four initial conditions are set to zero. The initial conditions of the observed species *Epo* + *dEpo_e* and *Epo_{EpoR}* are, therefore, completely defined. The Fisher information for the remaining six parameters is well conditioned and we may therefore infer these together, without using a blocking approach. For this model, a burn-in phase of 5000 samples was found to be sufficiently long for the Markov chains to converge.

We obtained the auxiliary sensitivity equations for our ODE models using the Symbolic Math Toolbox in Matlab. This automated process is extremely fast and helps prevent human mathematical errors. We solve them by making use of the SBToolbox2 [47] for Matlab, which provides an interface to a C implementation of the Sundials solver [31] and is up to two orders of magnitude faster than using the built-in solvers in Matlab.

3. RESULTS

3.1. Circadian model

We first consider three subsets of the parameters individually; the linear parameters ($\mathbf{p}_{1:2}$, $\mathbf{r}_{1:4}$), the transcription parameters ($\mathbf{n}_{1:2}$, $\mathbf{g}_{1:2}$) and the Michaelis–Menten parameters ($\mathbf{m}_{1:6}$, $\mathbf{k}_{1:6}$). This allows us to see the potential challenges of inferring parameters in these three different types of functions. We perform inference over each set of parameters with the other sets of parameters fixed at their true values.

3.1.1. Linear parameters

For now, we run a single chain initialized on the true parameter values for the purpose of evaluating the

sampling properties at the mode, and we fix the initial conditions.

We first compare posterior sampling of the linear parameters using Metropolis, simplified mMALA, MALA and RMHMC, and a comparison of the sample traces is given in figure 2. We employ a component-wise Metropolis sampler, whereby each parameter is updated sequentially conditioned on all other parameters using a Gaussian proposal, and we note that the component-wise Metropolis–Hastings sampler, therefore, requires the ODE model to be solved once for each parameter. The proposal variances are automatically tuned to achieve acceptance rates of between 20 and 50 per cent, and then fixed before drawing posterior samples. This standard sampler performs very poorly and is unable to sample efficiently from the strongly correlated posterior distribution. The manifold methods, however, perform far better, making use of the local geometry information to propose better moves through parameter space. RMHMC makes proposals that follow geodesic paths across the manifold; however, in our implementation, it is computationally much more expensive,² requiring an average of five evaluations of the second-order sensitivity equations and an average of 25 evaluations of the first-order sensitivity equations. mMALA makes moves based on a diffusion process across the manifold, and also performs much better than Metropolis–Hastings, although again this is computationally expensive because of the need to calculate second-order sensitivities of the differential equation model; it requires two evaluations of the second-order sensitivity equations per iteration. Simplified mMALA makes valid MCMC proposals based on

²This performance could be substantially improved by using adjoint differentiation methods and programming using a compiled language, rather than the interpreted language Matlab that we use here.

Figure 3. Scatter plots and pairwise density estimates of the posterior samples demonstrate the strong correlation structure present in this nonlinear system for the linear parameters, with all other parameters fixed at their true values. The samples were obtained using the simplified mMALA sampler with a $\text{Gamma}(1.5, 2)$ prior, and artificially generated data with zero mean Gaussian noise, whose standard deviation was equal to 1% of the amplitude of the oscillations in each species.

an approximate diffusion across the manifold. It gives similar results to mMALA and is computationally much more efficient as it only requires two evaluations of the first-order sensitivities per iteration. In practice, we find that solving the extended first-order ODE system twice for simplified mMALA is only three to four times slower than solving the original system for each parameter in a component-wise Metropolis-Hastings sampler, yet it produces samples that are visibly far better (figure 2). We refer the interested reader to Girolami & Calderhead [23] for a more detailed comparison of the time-normalized efficiency of manifold sampling methods, noting that the extended first-order system consists only of additional linear equations that may be solved very efficiently [31]. We therefore employ simplified mMALA as our sampler of choice for all subsequent experiments in this paper.

Figure 3 shows scatter plots of the samples obtained from simplified mMALA for each parameter combination. Even with these linear parameters, there are some very strong correlations and severe scaling differences, illustrating the difficulty of sampling from statistical models of this type. As an example, let us consider parameters r_1 and r_2 , which define transport between the nucleus and cytoplasm for LHY/CCA. There is a very strong correlation structure between these parameters; as the concentration of the protein in the nucleus increases, the concentration of protein in the cytoplasm decreases and vice versa. There is a similar strong correlation structure between parameters r_3 and r_4 describing nuclear/cytoplasmic transport for TOC1. Although the other parameters do not exhibit such strong dependencies, their differing scales are evident, which further compounds the difficulties.

We also compared the effect that changing the noise level has on parameter inference, and in particular on the practical identifiability of the model. Even with very low levels of noise, we observed wide variability in the parameter estimates; however, the model was still able to make predictions with high certainty, as has been observed in biological models examined in previous work [27]; dealing with large uncertainty in parameter values is an unavoidable part of the modelling process. For many models, the experimental accuracy required to obtain tight bounds on individual parameters is simply unachievable, and indeed often undesirable if we are mainly interested in the model predictions. The careful use of priors can therefore help to constrain parameter values to biologically meaningful ranges.

3.1.2. Transcription parameters

We now consider inference over the four transcription parameters with Hill coefficients, and fix all other parameters at their true values. Again, there are strong correlations between these parameters, particularly between parameters appearing together within the same algebraic term. This correlation structure is shown in figure 4 and was seen to remain even with increasing levels of error variance in the data.

3.1.3. Michaelis–Menten parameters

Fixing all parameters apart from those appearing in the Michaelis–Menten terms, we investigate the effect of a blocking strategy on sampling. Informed by the analysis of the Michaelis–Menten terms in §2.2, we may group the 12 parameters into pairs; each pair appears together

Figure 4. Scatter plots and pairwise density estimates of the sampled parameters associated with Hill coefficients, n_1 , n_2 , g_1 , g_2 . The samples were obtained using simplified mMALA with a $\text{Gamma}(1.5, 2)$ prior. This time the Gaussian noise used to generate the data that had standard deviation equal to 10% of the amplitude of the oscillations. Even with this higher level of uncertainty in the data, the samples still exhibit very strong correlation structure, reinforcing the need for more advanced sampling methodology.

Figure 5. The 12 Michaelis–Menten parameters from the circadian model were sampled using simplified mMALA, with and without a blocking strategy. The red trace plots show the parameter samples obtained by grouping (m_i, k_i) and sampling using a simplified mMALA within Gibbs scheme. The black trace plots show the parameter samples obtained by sampling all parameters simultaneously using simplified mMALA. In particular, parameters (m_1, k_1) are only weakly identifiable and this results in an ill-conditioned Fisher information that negatively impacts on the sampling efficiency. Blocking the pairs of parameters in each Michaelis–Menten term allows for smaller, better conditioned FI matrices; this numerical stability allows for more efficient sampling using a blocking approach.

in a Michaelis–Menten term and in certain cases the system may respond to changes in the ratios of these pairs. Figure 5 shows the traces of Michaelis–Menten parameter samples obtained with and without the use of blocking. Without blocking, the Fisher information becomes badly conditioned (condition number greater than 10^8) and the resulting numerical inaccuracy results in very small step sizes and highly correlated samples being drawn. The use of blocking parameters is seen to mitigate this problem.

Scatter plots of the parameter pairs (m_i, k_i) are also shown in figure 6. Interestingly, we see that the pair of Michaelis–Menten parameters with the strongest correlation (m_1, k_1) is indeed for a species with low concentration, such that $k_1 \gg [LHY]_m$. We see that this type of severe correlation structure was premeditated in §1, where we considered the case of the ratio of m_1 to k_1 driving the dynamics. Clearly, when

performing inference using MCMC, we wish to make moves in parameter space that take into account these strong correlations. There was seen to be less correlation between (m_i, k_i) and (m_j, k_j) , when $i \neq j$, which we might also expect from the structure of the equations since each Michaelis–Menten term contains only the chemical species of the equation in which it appear.

3.1.4. Inference with full and partial observations

Until now we have employed a single chain initialized at the correct parameter values to examine the local mixing properties of manifold MCMC methodology. We now consider the effect of using random starting values for our parameters. Performing inference on just the linear parameters, keeping all others fixed, we find that there appears to be a single mode, which our Markov chain can reach regardless of the starting parameters. However, if we infer both the linear parameters and the

Figure 6. Scatter plots and density estimates of the pairs of Michaelis–Menten parameter samples (m_i, k_i) obtained using simplified mMALA with a $\text{Gamma}(1.5, 2)$ prior and Gaussian noise with standard deviation equal to 10% of the amplitudes of oscillation. The possibility of such strong correlation structure can be predicted by an analytical consideration of the underlying equations.

Figure 7. The use of a population-based MCMC scheme is vital for ensuring convergence to the correct mode [21]. As an illustration, we first employ a single chain initialized at a chosen set of parameter values to obtain samples from the correct posterior mode using a manifold MCMC method. The predictive model output based on these samples is shown in red. We then initialized a single chain at a random set of parameter values sampled from a $\text{Gamma}(1.5, 2)$ prior distribution, and ran it until apparent convergence. This time the predictive model output converges to a local maximum, shown in blue, with the predictions cutting halfway through the oscillations in the data. Such local modes may also occur as the model output moves in and out of phase with oscillatory data.

transcription parameters, there is another mode to which our Markov chain may converge. This alternative set of parameters gives a non-oscillatory output that roughly corresponds to the average of the true model output, as shown in figure 7. Systems with complex nonlinear dynamics are particularly susceptible to this problem,

since local maxima may occur when the model moves in and out of phase with different parts of the data. Any MCMC methodology used for this problem must therefore not only have good *local* mixing properties, but it must also be capable of making more *global* steps, allowing it to escape from local modes of negligible

Figure 8. Comparison of posterior outputs from the circadian model obtained by performing Bayesian inference over all parameters using a blocked, population sampling scheme with simplified mMALA. With a fully observed system, the posterior model predictions are all made with low uncertainty. In contrast, we obtain much more vague predictions for unobserved species, and the uncertainty in the predictions increases as the number of observed species decreases.

probability mass. Using simplified mMALA to explore a population of tempered distributions [21] allows us to resolve this issue.

We infer all sets of model parameters together using a blocking strategy and population scheme with 20 tempered distributions, as in Calderhead & Girolami [21]. Figure 8 shows the differences in the predictive model outputs for differing numbers of observed species. This is important as it is usually not possible to obtain measurements for all components in a biological system, owing to either financial or technical constraints.

3.2. Cell signalling model

We infer parameters over the cell signalling model described in §2.1.2. in which the observations are a linear combination of the underlying modelled species, demonstrating that manifold sampling methodology extends straightforwardly to incorporate observation models. It also demonstrates how this methodology may be used to infer parameters of models based on commonly used mass action kinetics. The model is parametrized in $\log_{10}$ space allowing exploration over several orders of magnitude. Despite having only six parameters, this cell signalling model still exhibits reasonably complex dynamics and sensitivities that change markedly throughout the parameter space. This is shown in figure 9, where we observe the path taken by a Markov chain during the burn-in period

Figure 9. The path taken during the burn-in phase of a Markov chain moving through three of the six dimensions of parameter space of the cell signalling example (§8) is shown, with the colour representing the posterior probability of the current set of parameters; the chain starts with very low probability (blue) moves to higher probability region (red) and finally finds the posterior mode (brown). All proposal steps were made using local sensitivity information via simplified mMALA sampling. Interestingly, the sensitivities of the parameters change dramatically as the chain moves through the parameter space. At the beginning of the path (shown by the blue points), the chain moves in all three of these parameter directions, until it reaches a plateau along which two of the parameters become tightly constrained. At the end of the plateau, the sensitivities of two of the parameters switch with the chain now moving upwards, with little movement along the previously preferred axis.

Figure 10. Scatter plots and density estimates of the parameter samples from the erythropoietin cell signalling model obtained using simplified mMALA with a $\mathcal{N}(-2, 2)$ prior and Gaussian noise with standard deviation equal to 10% of the amplitudes of the outputs. We note the orders of magnitude difference in the scaling of each of the parameters.

Figure 11. The top row shows the posterior model predictions for each of the biochemical species in the erythropoietin cell signalling model, and the bottom row shows the predictive model outputs using the additional observation model, with the dataset shown in black. The tightest predictions are made for *Epo*, *dEpo_e* and *Epo_EpoR*, and indeed the observations are a linear combination of these species. Inference on this constrained model produces reasonably tight predictions for the unobserved species too, although this is a result of constraints on the model to ensure identifiability [30].

guided by the sensitivities of the model parameters, which dramatically change as the chain explores new regions of the space. Despite the very wide priors covering orders of magnitude, the population-based simplified mMALA sampler is still able to converge to the correct mode. In addition, we find that such a model based on mass action kinetics also induces a highly correlated, asymmetric posterior distribution that is challenging to sample from, as shown in figure 10, particularly as the parameter values are of differing orders of magnitude. Finally, model predictions

are made with very low uncertainty (figure 11); however, it has been noted [30] that a careful initial analysis of the model is necessary to achieve this structural and practical identifiability.

4. DISCUSSION AND CONCLUSIONS

Statistical analysis of mechanistic models describing biological systems plays a vital role in furthering our understanding of the underlying mechanisms driving observed behaviour [4, 19, 30, 48]. Both frequentist and

Bayesian approaches to statistical inference face the challenge of searching through high-dimensional parameter space; the former relying on optimization, the latter on MCMC sampling.

In this paper, we have focused on the Bayesian approach and find standard sampling methods are inadequate for this task. We require our sampler to take into account the locally changing correlation structure in order to propose moves that are accepted with high probability and adequately and efficiently explore the space. One approach is to employ adaptive MCMC methods [38] that estimate the covariance using previous samples from the Markov chain; however, the rate of diminishing adaptation must be carefully controlled to ensure convergence and it may also take a while to obtain reasonable estimates of what is essentially the global covariance structure. In contrast, we obtain the local covariance structure directly using the induced Riemannian geometry of the model. Manifold MCMC methods [23] make moves in parameter space based directly on the sensitivities of the underlying differential equations with respect to each parameter, resulting in efficient convergence to the correct posterior distribution.

A further inferential challenge stems from the fact that many complex ODE models may induce multiple modes, which can occur, for example, when a non-oscillatory steady-state solution is found that may be fitted to the average values of the observed data as we have illustrated in figure 7. Population MCMC approaches offer an effective solution to this problem and further allow for accurate estimation of marginal likelihoods [43].

The RMHMC sampler clearly offers the most effective scheme in terms of generating nearly independent posterior samples; however, solving the equations to calculate the geodesic paths on the induced manifold can be computationally expensive. Ideas for mitigating this issue include the adoption of adjoint differentiation methods for efficiently calculating sensitivities of large numbers of parameters, the use of appropriate guiding Hamiltonians in the proposal mechanism [44] and implementation in computationally more efficient programming languages. Further work will bring these costs down and make RMHMC available to a larger class of problems. In the meantime, the simplified mMALA sampler offers the benefits of a geometric approach to MCMC, while remaining computationally feasible for large models.

Many realistic biological models have parameters that are unidentifiable and this issue has recently received a lot of attention [20]. Such problems can be mitigated by considering a reduced model or, in a Bayesian setting, employing biological and mathematical analysis to inform the choice of priors and enforce weak identifiability [27]. Although ODE models may be applied to model a wide range of important natural phenomena, other modelling formalisms, such as stochastic and partial differential equations [7,49], are also important in other settings. Recent work developing calculations of the Fisher information for SDEs

[50] suggest that such systems may also be considered within a manifold framework.

The manifold samplers we employed use the local sensitivities of the model to construct an efficient Markov transition kernel, allowing us to sample from complex posterior densities, which in turn provide the desired global assessment of sensitivity over a range of dynamic responses of the system and not just at a single operating point. In this way, manifold samplers provide an efficient route from local to global sensitivity analysis.

We thank the four anonymous referees for constructive comments that have improved this paper. This work was funded by the following research grants: The Molecular Nose ‘EPSRC EP/E032745/1, Advanced Research Fellowship’ EPSRC EP/E052029/1, Inference-based Modelling in Population and Systems Biology ‘BBSRC BB/G006997/1, The Silicon Trypanosome’ BBSRC BB/1004599/1, Analysing and Striking the Sensitivities of Embryonal Tumours’ EU FP7, Bayesian Inference in Systems Biology: Modelling Organ Specificity of Circadian Control in Plants’ Microsoft Research, Research Fellowship—2020 Science—UCL, University of Oxford, Microsoft Research.

A. DIFFERENTIAL EQUATION MODELS

A.1. Circadian control model

$$\begin{aligned}\frac{d[LHY]_m}{dt} &= \frac{n_1[TOC1]_n^a}{g_1^a + [TOC1]_n^a} - \frac{m_1[LHY]_m}{k_1 + [LHY]_m}, \\ \frac{d[LHY]_c}{dt} &= p_1[LHY]_m - r_1[LHY]_c \\ &\quad + r_2[LHY]_n - \frac{m_2[LHY]_c}{k_2 + [LHY]_c}, \\ \frac{d[LHY]_n}{dt} &= r_1[LHY]_c - r_2[LHY]_n - \frac{m_3[LHY]_n}{k_3 + [LHY]_n}, \\ \frac{d[TOC1]_m}{dt} &= \frac{n_2 g_2^b}{g_2^b + [LHY]_n^b} - \frac{m_4[TOC1]_m}{k_4 + [TOC1]_m}, \\ \frac{d[TOC1]_c}{dt} &= p_2[TOC1]_m - r_3[TOC1]_c \\ &\quad + r_4[TOC1]_n - \frac{m_5[TOC1]_c}{k_5 + [TOC1]_c}, \\ \text{and } \frac{d[TOC1]_n}{dt} &= r_3[TOC1]_c - r_4[TOC1]_n \\ &\quad - \frac{m_6[TOC1]_n}{k_6 + [TOC1]_n}.\end{aligned}$$

We used the following parameter values from the literature [12]: $p_1 = 9.0002$, $p_2 = 3.6414$, $r_1 = 5.6429$, $r_2 = 8.2453$, $r_3 = 1.2789$, $r_4 = 5.3527$, $n_1 = 0.6187$, $n_2 = 7.7768$, $g_1 = 3.7051$, $g_2 = 9.7142$, $k_1 = 7.8618$, $m_1 = 7.3892$, $k_2 = 3.2829$, $m_2 = 0.4716$, $k_3 = 6.3907$, $m_3 = 4.1307$, $k_4 = 1.0631$, $m_4 = 5.7775$, $k_5 = 0.9271$, $m_5 = 4.4555$, $k_6 = 5.0376$, $m_6 = 7.6121$. The Hill coefficients were fixed at $a = 1$ and $b = 2$, and the initial conditions used for generating the synthetic data were $cLm(0) = 0.1290$, $cLc(0) = 13.6937$, $cLn(0) = 9.1584$, $cTm(0) = 1.9919$, $cTc(0) = 5.9266$, $cTn(0) = 1.1007$.

A.2. Signalling pathway model

$$\begin{aligned}\frac{d[Epo]}{dt} &= -k_{on}[Epo][EpoR] + k_{on}k_D[Epo_EpoR] + k_{ex}[Epo_EpoR_i], \\ \frac{d[EpoR]}{dt} &= -k_{on}[Epo][EpoR] + k_{on}k_D[Epo_EpoR] \\ &\quad + k_t B_{max} - k_t[EpoR] + k_{ex}[Epo_EpoR_i], \\ \frac{d[Epo_EpoR]}{dt} &= k_{on}[Epo][EpoR] - k_{on}k_D[Epo_EpoR] - k_e[Epo_EpoR], \\ \frac{d[Epo_EpoR_i]}{dt} &= k_e[Epo_EpoR] - k_{ex}[Epo_EpoR_i] \\ &\quad - k_{di}[Epo_EpoR_i] - k_{de}[Epo_EpoR_i], \\ \frac{d[dEpo_i]}{dt} &= k_{di}[Epo_EpoR_i] \\ \text{and } \frac{d[dEpo_e]}{dt} &= k_{de}[Epo_EpoR_i].\end{aligned}$$

The parameter values (in $\log_{10}$ scale) were determined from the literature to be: $k_{on} = -4.091$, $k_{ex} = -2.447$, $k_t = -1.758$, $k_e = -1.177$, $k_{di} = -2.730$ and $k_{de} = -1.884$. The remaining parameters were fixed to enforce identifiability: $k_D = 2.583$, $B_{max} = 2.821$. The initial conditions were fixed at: $Epo(0) = 2098.9$, $EpoR(0) = 662.22$, $Epo_EpoR(0) = 0$, $Epo_EpoR_i(0) = 0$, $dEpo_i(0) = 0$ and $dEpo_e(0) = 0$.

REFERENCES

- Gershenfeld, N. 1999 *The nature of mathematical modeling*. Cambridge, UK: Cambridge University Press.
- Kolch, W., Calder, M. & Gilbert, D. 2005 When kinases meet mathematics: the systems biology of MAPK signaling. *FEBS Lett.* **579**, 1891–1895. (doi:10.1016/j.febslet.2005.02.002)
- Tomlin, C. J. & Axelrod, J. D. 2007 Biology by numbers: mathematical modelling in developmental biology. *Nat. Rev. Genet.* **8**, 331–340. (doi:10.1038/nrg2098)
- Xu, T. *et al.* 2010 Inferring signaling pathway topologies from multiple perturbation measurements of specific biochemical species. *Sci. Signal.* **3**, ra20. (doi:10.1126/scisignal.2000517)
- Mohamed, L., Calderhead, B., Filippone, M., Christie, M. & Girolami, M. 2011 Population MCMC methods for history matching and uncertainty quantification. *Comput. Geosci.* **15**. (doi:10.1007/S10569-011-9232-8)
- Chen, W. W., Niepel, M. & Sorger, P. K. 2010 Classic and contemporary approaches to modeling biochemical reactions. *Genes Dev.* **24**, 1861–1875. (doi:10.1101/gad.1945410)
- Wilkinson, D. J. 2006 *Stochastic modelling for systems biology*. London, UK: Chapman & Hall/CRC.
- Macklin, P., McDougall, S., Anderson, A., Chaplain, M., Cristini, V. & Lowengrub, J. 2009 Multiscale modelling and nonlinear simulation of vascular tumour growth. *J. Math. Biol.* **58**, 765–798. (doi:10.1007/s00285-008-0216-9)
- Konukoglu, E. *et al.* In press. Efficient probabilistic model personalization integrating uncertainty on data and parameters: application to eikonal-diffusion models in cardiac electrophysiology. *Prog. Biophys. Mol. Biol.* (doi:10.1016/j.pbiomolbio.2011.07.002)
- Nolan, T., Hands, R. E. & Bustin, S. A. 2006 Quantification of mRNA using real-time RT-PCR. *Nat. Protocols* **1**, 1559–1582. (doi:10.1038/nprot.2006.236)
- Vyshemirsky, V. & Girolami, M. 2008 Bayesian ranking of biochemical system models. *Bioinformatics* **24**, 833–839. (doi:10.1093/bioinformatics/btm607)
- Locke, J. C. W., Millar, A. & Turner, M. 2005 Modelling genetic networks with noisy and varied experimental data: the circadian clock in *Arabidopsis thaliana*. *J. Theor. Biol.* **234**, 383–393. (doi:10.1016/j.jtbi.2004.11.038)
- Locke, J. C. W., Southern, M. M., Kozma-Bognar, L., Hibberd, V., Brown, P. E., Turner, M. S. & Millar, A. 2005 Extension of a genetic network model by iterative experimentation and mathematical analysis. *Mol. Syst. Biol.* **1**. (doi:10.1038/msb4100018)
- Locke, J. C. W., Kozma-Bognar, L., Gould, P. D., Feher, B., Kevei, E., Nagy, F., Turner, M. S., Hall, A. & Millar, A. 2006 Experimental validation of a predicted feedback loop in the multi-oscillator clock of *Arabidopsis thaliana*. *Mol. Syst. Biol.* **2**, 59. (doi:10.1038/msb4100102)
- Takahashi, J. S., Hong, H., Ko, C. H. & McDearmon, E. L. 2008 The genetics of mammalian circadian order and disorder: implications for physiology and disease. *Nat. Rev. Genet.* **9**, 764–775. (doi:10.1038/nrg2430)
- Lévi, F., Okyar, A., Dulong, S., Innominato, P. F. & Clairambault, J. 2010 Circadian timing in cancer treatments. *Annu. Rev. Pharmacol. Toxicol.* **50**, 377–421. (doi:10.1146/annurev.pharmtox.48.113006.094626)
- Sawyers, C. 2004 Targeted cancer therapy. *Nature* **432**, 294–297. (doi:10.1038/nature03095)
- Butcher, E. C., Berg, E. L. & Kunkel, E. J. 2004 Systems biology in drug discovery. *Nat. Biotechnol.* **22**, 1253–1259. (doi:10.1038/nbt1017)
- Toni, T., Welch, D., Strelkowa, N., Ipsen, A. & Stumpf, M. P. H. 2009 Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *J. R. Soc. Interface* **6**, 187–202. (doi:10.1098/rsif.2008.0172)
- Raue, A., Kreutz, C., Maiwald, T., Bachmann, J., Schilling, M., Klingmüller, U. & Timmer, J. 2009 Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics* **25**, 1923–1929. (doi:10.1093/bioinformatics/btp358)
- Calderhead, B. & Girolami, M. 2009 Estimating Bayes factors via thermodynamic integration and population MCMC. *Comput. Stat. Data Anal.* **53**, 4028–4045. (doi:10.1016/j.csda.2009.07.025)
- Robert, C. P. & Casella, G. 2004 *Monte Carlo statistical methods*. New York, NY: Springer.
- Girolami, M. & Calderhead, B. 2011 Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* **73**, 123–214. (doi:10.1111/j.1467-9868.2010.00765.x)
- Brown, K. & Sethna, J. P. 2003 Statistical mechanical approaches to models with many poorly known parameters. *Phys. Rev. E* **68**, 021904. (doi:10.1103/PhysRevE.68.021904)
- Gutenkunst, R., Waterfall, J., Casey, F. P., Brown, K. S., Myers, C. R. & Sethna, J. P. 2007 Universally sloppy parameter sensitivities in systems biology models. *PLoS Comput. Biol.* **3**, e189. (doi:10.1371/journal.pcbi.0030189)
- Daniels, B., Chen, Y., Sethna, J. P., Gutenkunst, R. N. & Myers, C. R. 2008 Sloppiness, robustness, and evolvability in systems biology. *Curr. Opin. Biotechnol.* **19**, 389–395. (doi:10.1016/j.copbio.2008.06.008)

- 27 Transtrum, M., Machta, B. B. & Sethna, J. P. 2011 Geometry of nonlinear least squares with applications to sloppy models and optimization. *Phys. Rev. E* **83**, 036701. (doi:10.1103/PhysRevE.83.036701)
- 28 Erguler, K. & Stumpf, M. P. H. 2011 Practical limits for reverse engineering of dynamical systems: a statistical analysis of sensitivity and parameter inferability in systems biology models. *Mol. BioSyst.* **7**, 1593–1602. (doi:10.1039/c0mb00107d)
- 29 do Carmo, M. P. 1992 *Riemannian geometry*. Boston, MA: Birkhauser.
- 30 Becker, V., Schilling, M., Bachmann, J., Baumann, U., Raue, A., Maiwald, T., Timmer, J. & Klingmuller, U. 2010 Covering a broad dynamic range: information processing at the erythropoietin receptor. *Science* **328**, 1404–1408. (doi:10.1126/science.1184913)
- 31 Hindmarsh, A. C., Brown, P. N., Grant, K. E., Lee, S. L., Serban, R., Shumaker, D. E. & Woodward, C. S. 2005 SUNDIALS: Suite of nonlinear and differential/algebraic equation solvers. *ACM Trans. Math. Softw. (TOMS)* **31**, 363–396. (doi:10.1145/1089014.1089020)
- 32 Koornneef, M. & Meinke, D. 2010 The development of *Arabidopsis* as a model plant. *Plant J.* **61**, 909–921. (doi:10.1111/j.1365-3113X.2009.04086.x)
- 33 McClung, C. 2006 Plant circadian rhythms. *Plant Cell* **18**, 792–803. (doi:10.1105/tpc.106.040980)
- 34 Hubbard, K. E., Robertson, F. C., Dalchau, N. & Webb, A. A. R. 2009 Systems analyses of circadian networks. *Mol. BioSyst.* **5**, 1502–1511. (doi:10.1039/b907714f)
- 35 Shankaran, H., Resat, H. & Wiley, H. S. 2007 Cell surface receptors for signal transduction and ligand transport: a design principles study. *PLoS Comput. Biol.* **3**, e101. (doi:10.1371/journal.pcbi.0030101)
- 36 Vallisneri, M. 2008 Use and abuse of the Fisher information matrix in the assessment of gravitational-wave parameter-estimation prospects. *Phys. Rev. D* **77**, 2001. (doi:10.1103/PhysRevD.77.042001)
- 37 Higham, C. 2009 Bifurcation analysis informs Bayesian inference in the Hes1 feedback loop. *BMC Syst. Biol.* **3**, 12. (doi:10.1186/1752-0509-3-12)
- 38 Haario, H., Laine, M., Mira, A. & Saksman, E. 2006 DRAM: efficient adaptive MCMC. *Stat. Comput.* **16**, 339–354. (doi:10.1007/s11222-006-9438-0)
- 39 Jeffreys, H. 1939 *Theory of probability*, 1st edn. Oxford, UK: The Clarendon Press.
- 40 Rao, C. R. 1945 Information and the accuracy attainable in the estimation of several parameters. *Calcutta Math. Bull.* **37**, 81–91.
- 41 Brown, K., Hill, C., Calero, G., Myers, C. R., Lee, K. H., Sethna, J. P. & Cerione, R. A. 2004 The statistical mechanics of complex signaling networks: nerve growth factor signaling. *Phys. Biol.* **1**, 184–195. (doi:10.1088/1478-3967/1/3/006)
- 42 Lartillot, N. & Philippe, H. 2006 Computing Bayes factors using thermodynamic integration. *Syst. Biol.* **55**, 195–207. (doi:10.1080/10635150500433722)
- 43 Friel, N. & Pettitt, A. N. 2008 Marginal likelihood estimation via power posteriors. *J. R. Stat. Soc. Ser. B* **70**, 589–607. (doi:10.1111/j.1467-9868.2007.00650.x)
- 44 Duane, S., Kennedy, A. D., Pendleton, B. J. & Roweth, D. 1987 Hybrid Monte Carlo. *Phys. Lett. B* **195**, 216–222. (doi:10.1016/0370-2693(87)91197-X)
- 45 Roberts, G. O. & Tweedie, R. L. 1996 Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli* **2**, 341–363. (doi:10.2307/3318418)
- 46 Chib, S. & Ramamurthy, S. 2010 Tailored randomized block MCMC methods with application to DSGE models. *J. Econom.* **155**, 19–38. (doi:10.1016/j.jeconom.2009.08.003)
- 47 Schmidt, H. & Jirstrand, M. 2006 Systems biology toolbox for MATLAB: a computational platform for research in systems biology. *Bioinformatics* **22**, 514–515. (doi:10.1093/bioinformatics/bti799)
- 48 Swameye, I., Muller, T., Timmer, J., Sandra, O. & Klingmuller, U. 2003 Identification of nucleocytoplasmic cycling as a remote sensor in cellular signaling by databased modeling. *Proc. Natl Acad. Sci. USA* **100**, 1028–1033. (doi:10.1073/pnas.0237333100)
- 49 Müller, T. & Timmer, J. 2002 Fitting parameters in partial differential equations from partially observed noisy data. *Physica D Nonlinear Phenom.* **171**, 1–7. (doi:10.1016/S0167-2789(02)00546-8)
- 50 Komorowski, M., Costa, M., Rand, D. A. & Stumpf, M. 2011 Sensitivity, robustness and identifiability in stochastic chemical kinetics models. *Proc. Natl Acad. Sci. USA* **108**, 8645–8650. (doi:10.1073/pnas.1015814108)