

# Bayesian inference for stochastic models of intracellular reaction networks

**Darren Wilkinson**

School of Mathematics & Statistics

and

Centre for Integrated Systems Biology of Ageing and Nutrition  
Newcastle University, UK

IMA Workshop:

Stochastic models for intracellular reaction networks

IMA, Minneapolis, U.S.A.

May 13, 2008

# Overview

- Markov process models of biochemical network dynamics
- Bayesian inference for model parameters
- The Chemical Langevin Equation (CLE)
- An efficient MCMC algorithm for diffusions
- Likelihood-free and emulation based approaches to Bayesian calibration of complex models
- If time permits:
  - Sparse VAR(1) models for HTP data
  - Linking VAR(1) models to the CLE

# Computational Systems Biology (CSB)

- Much of CSB is concerned with building models of complex biological pathways, then validating and analysing those models using a variety of methods, including time-course simulation
- Most CSB researchers work with continuous deterministic models (coupled ODE and DAE systems)
- There is increasing evidence that much intra-cellular behaviour (including gene expression) is **intrinsically stochastic**, and that systems cannot be properly understood unless stochastic effects are incorporated into the models
- **Stochastic models** are harder to build, estimate, validate, analyse and simulate than deterministic models...

# Stochastic Chemical Kinetics

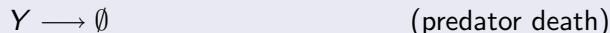
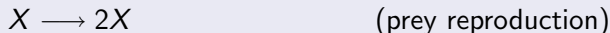
Stochastic molecular approach:

- Statistical mechanical arguments lead to a **Markov jump process** in continuous time whose instantaneous reaction rates are directly proportional to the number of molecules of each reacting species
- Such dynamics can be simulated (exactly) on a computer using standard **discrete-event simulation** techniques
- Standard implementation of this strategy is known as the “**Gillespie algorithm**” (just discrete event simulation), but there are several exact and approximate variants of this basic approach

# Lotka-Volterra system

Trivial (familiar) example from population dynamics (in reality, the “reactions” will be elementary biochemical reactions taking place inside a cell)

## Reactions



- $X$  – Prey,  $Y$  – Predator
- We can re-write this using matrix notation

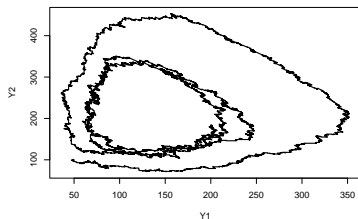
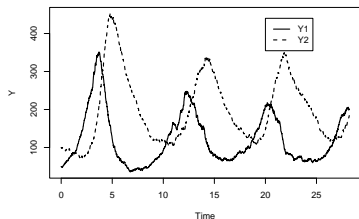
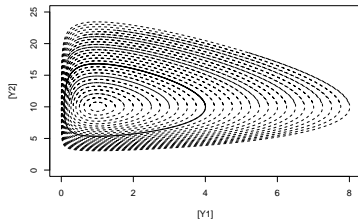
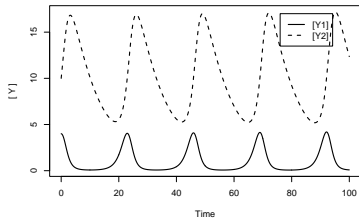
# Forming the matrix representation

## The L-V system in tabular form

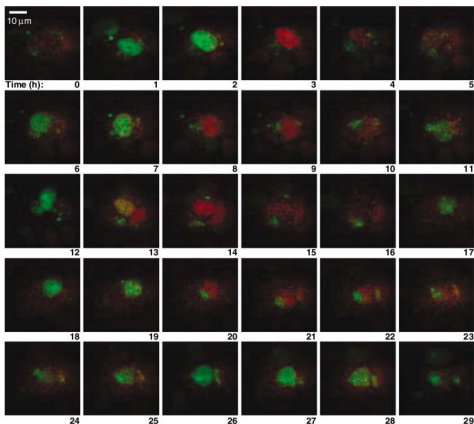
	Rate Law $h(\cdot, c)$	LHS		RHS		Net-effect	
		X	Y	X	Y	X	Y
$R_1$	$c_1x$	1	0	2	0	1	0
$R_2$	$c_2xy$	1	1	0	2	-1	1
$R_3$	$c_3y$	0	1	0	0	0	-1

Call the  $3 \times 2$  net-effect (or **reaction**) matrix  $N$ . The matrix  $S = N'$  is the **stoichiometry matrix** of the system. Typically both are **sparse**. The SVD of  $S$  (or  $N$ ) is of interest for structural analysis of the system dynamics...

# The Lotka-Volterra model



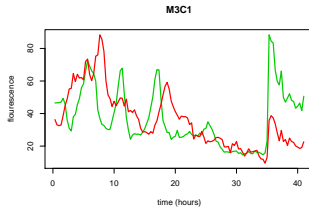
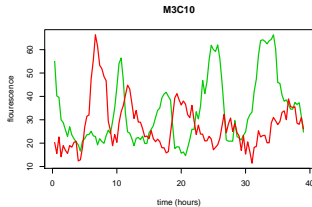
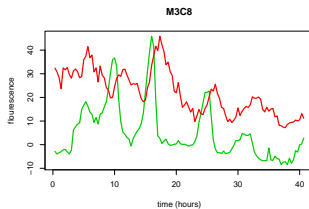
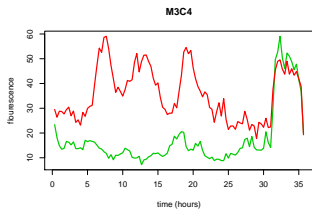
# Single cell fluorescence microscopy



p53-CFP and Mdm2-YFP

p53/Mdm2 oscillations subsequent to gamma irradiation

# Single cell time course data

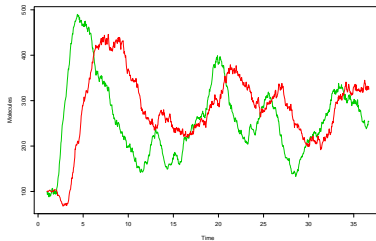
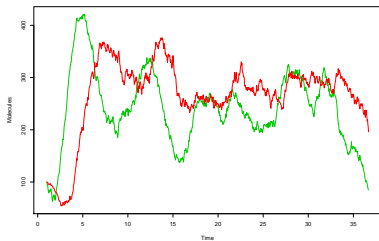
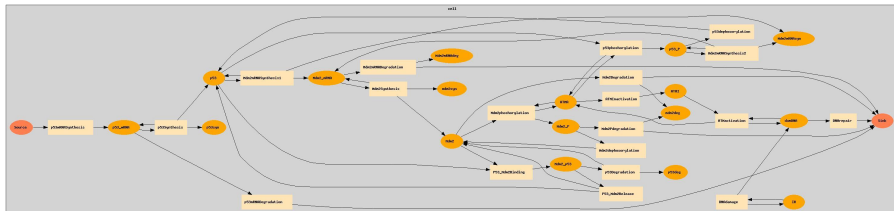


Geva-Zatorsky et al (2006), *Mol. Sys. Bio.* [Uri Alon's lab]

# Stochastic kinetic model

- Stochastic kinetic model developed at Newcastle (by C. J. Proctor) for the key biomolecular interactions between p53, Mdm2 and their response to DNA damage induced by irradiation
- More complex than the simple Lotka-Volterra system (17 species and 20 reactions), but essentially the same regulatory feedback mechanism (Mdm2 synthesis depends on the level of free p53, and Mdm2 encourages degradation of p53)
- Some information about most kinetic parameters, but considerable uncertainty for several — ideal for a Bayesian analysis

# Model structure and sample output



# Bayesian inference

Tuning model parameters so that output from the model “better matches” experimental data is a standard optimisation problem, but is problematic and unsatisfactory for a number of reasons:

- Defining an appropriate “objective function” is not straightforward if the model is stochastic or the measurement error has a complex structure (not IID Gaussian)
- The statistical concept of **likelihood** provides the “correct” way of measuring the evidence in favour of a set of model parameters, but typically requires computationally intensive Monte Carlo procedures for evaluation in complex settings
- Simple optimisation of the likelihood (the **maximum likelihood** approach) is also unsatisfactory, as there are typically many parameter combinations with very similar likelihoods (and the likelihood surface is typically multi-modal, making global optimisation difficult)

# Markov chain Monte Carlo (MCMC)

- Additionally, likelihood ignores any existing information known about likely parameter values *a priori*, which can be very useful for regularising the inference problem — better to base inference on the **posterior distribution**
- MCMC algorithms can be used to explore plausible regions of parameter space in accordance with the posterior distribution — these provide rich information
- eg. rather than simple point estimates for parameter values, can get **plausible ranges** of values, together with information on parameter **identifiability** and **confounding**
- MCMC algorithms are computationally intensive, but given that evaluation of the likelihood is typically computationally intensive anyway, nothing to lose and everything to gain by doing a Bayesian analysis

## Fully Bayesian inference

- In principle it is possible to carry out rigorous Bayesian statistical inference for the parameters of stochastic kinetic models
- Fairly detailed experimental data are required — eg. **quantitative single-cell time-course data** derived from live-cell imaging
- The standard procedure uses GFP labelling of key reporter proteins together with time-lapse confocal microscopy, but other approaches are also possible
- Techniques for **exact inference** for the **true discrete model** (Boys, W, Kirkwood 2008) do not scale well to problems of realistic size and complexity, due to the difficulty of efficiently exploring large complex integer lattice state spaces

## Rate parameter inference from complete data

- Observe process  $\mathbf{x} = \{x(t) : t \in [0, T]\}$
- Time and type of  $i$ th reaction event is  $(t_i, \nu_i)$ ,  $i = 1, \dots, n$ .  
Also define  $t_0 = 0, t_{n+1} = T$
- Let  $r_i$  be the number of type  $i$  events (so  $n = \sum_{i=1}^n r_i$ )
- The complete-data likelihood for the observed sample path is

$$L(c; \mathbf{x}) = \left\{ \prod_{i=1}^n h_{\nu_i}(x(t_{i-1}), c_i) \right\} \exp \left\{ - \int_0^T h_0(x(t), c) dt \right\}$$

(the integral is just a finite sum)

## Factorisation of the complete-data likelihood

- For rate laws of the form  $h_i(x, c_i) = c_i g(x)$  (true for mass-action stochastic kinetic models), the complete-data likelihood factorises as

$$L(c; \mathbf{x}) = \prod_{j=1}^{\nu} L_j(c_j; \mathbf{x})$$

where

$$L_j(c_j; \mathbf{x}) = c_j^{r_j} \exp \left\{ -c_j \int_0^T g_j(x(t)) dt \right\}, \quad j = 1, \dots, \nu$$

# Parameter estimation

- Factorisation leads to MLEs of the form

$$\hat{c}_j = r_j / \int_0^T g_j(x(t)) dt$$

- or can be combined with priors of the form  $c_j \sim \Gamma(a_j, b_j)$  to get full-conditionals

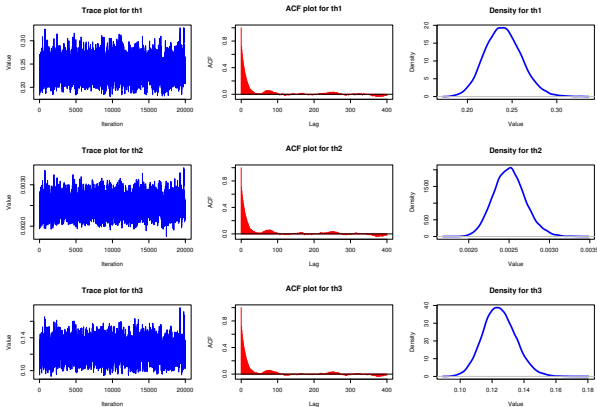
$$c_j | \mathbf{x} \sim \Gamma \left( a_j + r_j, b_j + \int_0^T g_j(x(t)) dt \right)$$

## Discrete-time observation

- Given discrete-time observations, the process breaks up into a collection of independent bridge processes that appear not to be analytically tractable
- Can use MCMC to explore sample paths consistent with the end-points
- Need to explore  $r_t$  consistent with  $x_{t+1} - x_t = Sr_t$
- Both reversible jump and block-updating strategies are possible — tedious details omitted!

# Inference for the Lotka-Volterra model

Data:  $z = \{x(t), y(t) : t = 0, 1, 2, \dots, 49\}$



Model remains identifiable when only the prey are observed...

## General reaction systems

- In the case of perfect observation of the state of all species at discrete times, there is free software, `stochInf`, which does this (approximately)
- Also possible to extend to problems in which not all (bio-)chemical species involved in the reaction system are observed and species are observed with error (idea — update intervals in pairs, relaxing constraints appropriately at the centre-point)
- Computational difficulties with large systems, large numbers of molecules, partial observation and measurement error...

# The Chemical Langevin Equation (CLE)

- The CLE is just a **diffusion approximation** for the Markov Jump Process (MJP) derived from stochastic chemical kinetic theory
- It is the Itô **stochastic differential equation** (SDE) model which “most closely matches” the dynamics of the MJP
- Formally, it is constructed by first considering a second-order approximation to the Kolmogorov forward equations for the process — known in this context as the **Chemical Master Equation** (CME)
- This second-order approximation is known as the **Fokker-Planck equation**, and is the Kolmogorov forward equation associated with the CLE

## Constructing the CLE

- Informally, we can easily construct the CLE as the SDE with the same infinitesimal mean and variance as the MJP
- In a time increment,  $dt$ , the change in state,  $dX_t$ , is given by  $dX_t = SdR_t$ , where the  $i$ th element of  $dR_t$  is a  $Po(h_i(X_t, c_i)dt)$  random quantity (independently of the other elements)
- Matching the mean and variance we put

$$dR_t \simeq h(X_t, c)dt + \text{diag} \left\{ \sqrt{h(X_t, c)} \right\} dW_t$$

where  $dW_t$  is the increment of a  $v$ -dimensional Wiener process.

- Then

$$dX_t = Sh(X_t, c)dt + S \text{diag} \left\{ \sqrt{h(X_t, c)} \right\} dW_t$$

## Constructing the CLE (ctd.)

- It is unnecessary (and sometimes inconvenient) to have the SDE being driven by a Brownian motion of higher dimension than the system state
- For this reason, the CLE is often written differently
- Since it is clear that  $\text{Var}(dX_t) = S \text{diag}\{h(X_t, c)\} S' dt$ , the CLE can be written

### The CLE

$$dX_t = Sh(X_t, c)dt + \sqrt{S \text{diag}\{h(X_t, c)\} S'} dW_t$$

where  $dW_t$  is now the increment of a  $u$ -dimensional Wiener process

## Good properties of the CLE

The CLE inherits many of the desirable properties of the MJP it approximates

### CLE properties

- Realisations of the CLE preserve conservation laws in the reaction network (associated with rank-degeneracy of the stoichiometry matrix,  $S$ )
- The CLE conserves matter (for models of a closed system) — no random creation or destruction of matter
- The CLE describes a non-negative stochastic process

## Inference for the CLE

Inference for a fairly general non-linear multivariate diffusion process, observed partially and discretely (and with error)

- Idea: Use an MCMC algorithm which “fills-in” the missing diffusion bridges between successive observations
- Use an Euler approximation to the true diffusion, but on a much **finer scale than the data**
- There are pathological mixing/convergence problems for regular MCMC schemes as the discretisation gets finer (essentially, there is an **infinite amount of information about the parameters** in the augmentation)
- It is nevertheless possible to develop effective MCMC algorithms...

## Likelihood concepts

- Putting  $\mu(x, c) = Sh(x, c)$  and  $\beta(x, c) = S \text{diag}\{h(x, c)\}S'$ :

$$dX_t = \mu(X_t, c)dt + \sqrt{\beta(X_t, c)}dW_t$$

- If we choose a small enough  $\Delta t$ , we get the Euler-Maruyama approximation

$$X_{t+\Delta t} | X_t, c \sim N(X_t + \mu(X_t, c)\Delta t, \beta(X_t, c)\Delta t)$$

- Perfect observation of the system state on this time grid leads to the “complete”-data likelihood

$$L(c; x) \propto \left\{ \prod_{i=0}^{n-1} |\beta(x_{i\Delta t}, c)|^{-1/2} \right\} \times \exp \left\{ -\frac{1}{2} \sum_{i=0}^{n-1} \left( \frac{\Delta x_{i\Delta t}}{\Delta t} - \mu(x_{i\Delta t}, c) \right)' \beta(x_{i\Delta t}, c)^{-1} \left( \frac{\Delta x_{i\Delta t}}{\Delta t} - \mu(x_{i\Delta t}, c) \right) \Delta t \right\}$$

## Likelihood problems

- Unfortunately the likelihood has no limit as  $\Delta t \rightarrow 0$
- **If** the diffusion term  $\beta(x, c)$  were independent of  $c$ , then it **would** be possible to discard some terms and then get a nice limit (exponential of the sum of a regular integral and an Itô stochastic integral), but it isn't...
- This is at the root of all of the computational problems concerning inference for diffusions
- More formally, the issue is whether or not the relevant Radon-Nikodym derivatives exist...

# Global MCMC algorithm

- Basic block-updating algorithm (perfect observation)
  - 1 Initialise parameters,  $c$ , and sample path,  $\mathbf{x}$
  - 2 For  $t = 2, \dots, n$ , propose a new sample path for  $\mathbf{x}_t$  using the MDB and accept/reject with a M-H step
  - 3 Conditional on  $\mathbf{x}$ , propose a new  $c$  and accept/reject with a M-H step
  - 4 Output state and return to step 2
- In the case of partial and/or noisy measurements, step 2 can be replaced with
  - 2 For  $t = 2, \dots, n - 1$ , propose a new sample path for the interval pair  $(\mathbf{x}_t, \mathbf{x}_{t+1})$  using the MDB, and accept/reject with a M-H step

together with special updating steps for the first and last intervals

## Modified diffusion bridge (MDB)

- Need a tractable process  $q(\mathbf{x}_{t+1}^* | c^*, x_t^*, d_{t+1})$  that is locally equivalent to  $\pi(\mathbf{x}_{t+1}^* | c^*, x_t^*, d_{t+1})$
- Diffusion  $dX_t = \mu(X_t)dt + \beta(X_t)^{\frac{1}{2}}dW_t$
- The nonlinear diffusion bridge

$$dX_t = \frac{x_1 - X_t}{1 - t} dt + \beta(X_t)^{\frac{1}{2}} dW_t$$

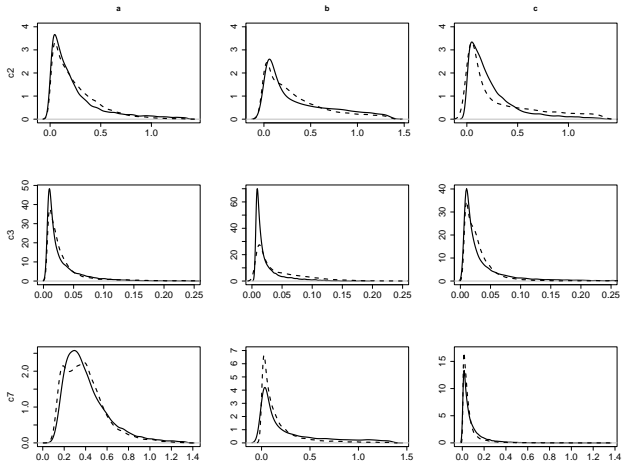
hits  $x_1$  at  $t = 1$ , yet is locally equivalent to the true diffusion as it has the same diffusion coefficient

- This forms the basis of an efficient proposal; see Durham & Gallant (2002), Chib, Pitt & Shephard (2004), Delyon & Hu (2006), and Stramer & Yan (2007) for technical details

# Irreducibility

- The above algorithm will give an effective irreducible MCMC algorithm provided that the diffusion term of the SDE does not depend on any model parameters
- On the other hand, if the diffusion term does depend on model parameters, then the algorithm will be reducible, due to the fact that there is an infinite amount of information in the augmented sample path  $\mathbf{x}$
- In practice we work with finite discretisations, so the information isn't "infinite", but sufficient to make the algorithm intolerably bad
- There is a solution to this based on a reparametrisation of the process — technical details in Golightly & Wilkinson (2008)

# Comparison of results



## Calibration of large simulators

CaliBayes — Integration of GRID-based post-genomic data resources through Bayesian calibration of biological simulators

<http://www.calibayes.ncl.ac.uk/>

- Bayesian model calibration is concerned with the problem of parameter estimation, model validation, design and analysis based only on the ability to **forward simulate** from the model
- It is particularly appropriate for slow and/or complex models and/or data, where likelihood-based methods are computationally infeasible

# MCMC-based fully Bayesian inference for *fast* computer models

- Before worrying about the issues associated with **slow** simulators, it is worth thinking about the issues involved in calibrating **fast deterministic** and **stochastic** simulators, based only on the ability to **forward-simulate** from the model
- In this case it is often possible to construct MCMC algorithms for fully Bayesian inference using the ideas of **likelihood-free MCMC** (Marjoram et al 2003)
- Here an MCMC scheme is developed exploiting forward simulation from the model, and this causes problematic likelihood terms to drop out of the M-H acceptance probabilities

## Generic problem

- Model parameters:  $c$
- (Stochastic) model output:  $\mathbf{x}$
- (Noisy and/or partial) data:  $\mathcal{D}$
- For simplicity suppose that  $c \perp\!\!\!\perp \mathcal{D} | \mathbf{x}$  (but can be relaxed)
- We wish to treat the model as a “black box”, which can only be forward-simulated
- We are thinking about data relating to a single realisation of the model (so no need to explicitly treat initial conditions), but replicate runs and multiple conditions can be handled sequentially (as will become clear)

# MCMC-based Bayesian inference

- Target:  $\pi(c|\mathcal{D})$
- Specify a “measurement error model”,  $\pi(\mathcal{D}|\mathbf{x})$  — eg. just a product of Gaussian or t densities
- Generic MCMC scheme:
  - Propose  $c^* \sim f(c^*|c)$
  - Accept with probability  $\min\{1, A\}$ , where

$$A = \frac{\pi(c^*)}{\pi(c)} \times \frac{f(c|c^*)}{f(c^*|c)} \times \frac{\pi(\mathcal{D}|c^*)}{\pi(\mathcal{D}|c)}$$

- $\pi(\mathcal{D}|c)$  is the “marginal likelihood” (or “observed data likelihood”, or...)

## Special case: deterministic model

- Deterministic function  $g(\cdot)$  such that  $\mathbf{x} = g(c)$
- Then

$$\begin{aligned}\pi(\mathcal{D}|c) &= \pi(\mathcal{D}|c, g(c)) \\ &= \pi(\mathcal{D}|c, \mathbf{x}) \\ &= \pi(\mathcal{D}|\mathbf{x})\end{aligned}$$

- Here  $\pi(\mathcal{D}|\mathbf{x})$  is just the “measurement error model” — eg. simple product of Gaussian or  $t$  densities
- This setup is somewhat simplistic for the deterministic case, but we are really more concerned with the stochastic case...

# Stochastic model

- Can't get at the marginal likelihood directly, so make the target  $\pi(c, \mathbf{x}|\mathcal{D})$ , where  $\mathbf{x}$  is the “true” simulator output which led to the observed data...
- Clear that we can marginalise out  $\mathbf{x}$  if necessary, but typically of inferential interest anyway
- Use ideas from “likelihood-free MCMC” (Marjoram et al, 2003)
- Propose  $(c^*, \mathbf{x}^*) \sim f(c^*|c)\pi(\mathbf{x}^*|c^*)$ , so that  $\mathbf{x}^*$  is a forward simulation from the (stochastic) model based on the proposed new  $c^*$

$$A = \frac{\pi(c^*)}{\pi(c)} \times \frac{f(c|c^*)}{f(c^*|c)} \times \frac{\pi(\mathcal{D}|\mathbf{x}^*)}{\pi(\mathcal{D}|\mathbf{x})}$$

# “Likelihood-free” MCMC

- Again  $\pi(\mathcal{D}|\mathbf{x})$  is a simple measurement error model...
- Crucially, because the proposal exploits a forward simulation, the acceptance probability does not depend on the likelihood of the simulator output — important for complex stochastic models
- This scheme is completely general, and works very well provided that  $|\mathcal{D}|$  is small
- **Problem:** If  $|\mathcal{D}|$  is large, the MCMC scheme will mix very poorly (very low acceptance rates)
- **Solution:** Exploit the Markovian structure of the process, and adopt a sequential approach, updating one observation at a time...

## Sequential likelihood-free algorithm

- Data  $\mathcal{D}_t = \{d_1, \dots, d_t\}$ ,  $\mathcal{D} \equiv \mathcal{D}_n$ . Sample paths  $\mathbf{x}_t \equiv \{x_s | t-1 < s \leq t\}$ ,  $t = 2, 3, \dots, n$ , so that  $\mathbf{x} \equiv \{\mathbf{x}_2, \dots, \mathbf{x}_n\}$ .
- ① Assume at time  $t$  we have a (large) sample from  $\pi(c, \mathbf{x}_t | \mathcal{D}_t)$  (for time 0, initialise with sample from prior)
- ② Run an MCMC algorithm which constructs a proposal in two stages:
  - ① First sample  $(c^*, \mathbf{x}_t^*) \sim \pi(c, \mathbf{x}_t | \mathcal{D}_t)$  by picking at random and perturbing slightly (sampling from the kernel density estimate)
  - ② Next sample  $\mathbf{x}_{t+1}^*$  by forward simulation from  $\pi(\mathbf{x}_{t+1}^* | c^*, \mathbf{x}_t^*)$
  - ③ Accept/reject  $(c^*, \mathbf{x}_{t+1}^*)$  with

$$A = \frac{\pi(d_{t+1} | \mathbf{x}_{t+1}^*)}{\pi(d_{t+1} | \mathbf{x}_{t+1})}$$

- ③ Output state, put  $t := t + 1$ , return to step 2.

## Advantages of the sequential algorithm

- In the presence of measurement error, the sequential likelihood-free scheme is effective, and is **much** simpler than a more efficient MCMC approach
- The likelihood-free approach is easier to tailor to non-standard models and data
- The essential problem is that of **calibration** of complex stochastic computer models
- For **slow** stochastic models, there is considerable interest in developing fast **emulators** and embedding these into MCMC algorithms

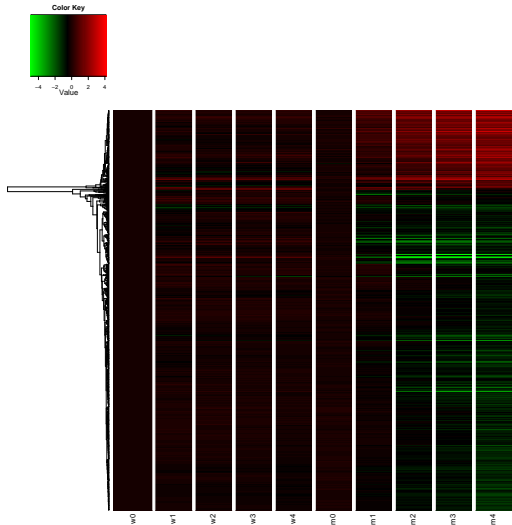
## Building emulators for *slow* simulators

- Use **Gaussian process regression** to build an emulator of a slow deterministic simulator
- Obtain runs on a carefully constructed set of design points (eg. a Latin hypercube) — easy to exploit parallel computing hardware here
- For a stochastic simulator, many approaches are possible
  - (Mixtures of) Dirichlet processes (and related constructs) are potentially quite flexible
  - Can also model output parametrically (say, Gaussian), with parameters modelled by (independent) Gaussian processes
  - Will typically want more than one run per design point, in order to be able to estimate distribution

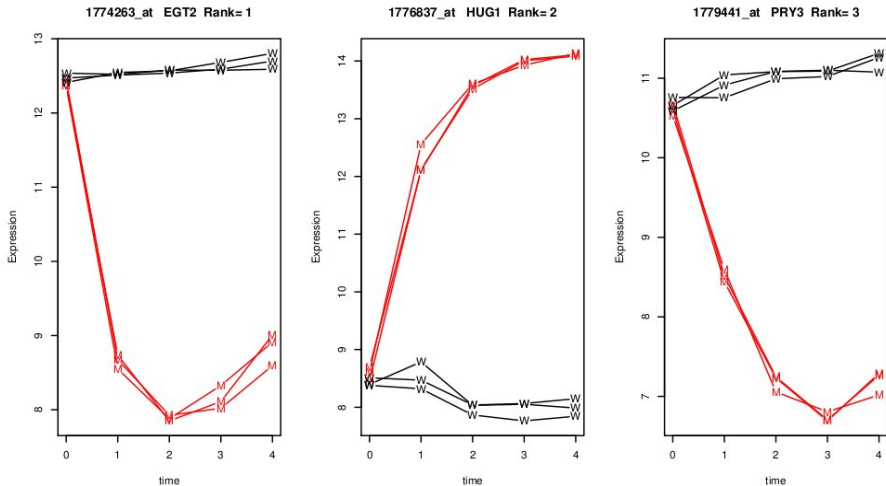
## High throughput data

- Although we would prefer to use high-resolution single-cell time course data for all of our statistical modelling, such data is difficult to obtain in a high throughput (HTP) fashion for large numbers of proteins
- We therefore wish to integrate HTP data into our modelling approach. Such data is usually of **lower resolution** and possessing relative poor **dynamic range**, but provides (simultaneous) measurement of very large numbers of biological features
- Time course HTP data is most useful for making inferences about dynamics
- **Time course microarray data** is a typical example

# Time course microarray data



# Time course microarray data



Small number of very high-dimensional short time courses

# Top-down modelling

- Using high-throughput 'omics data to fit statistical models and uncover networks of interacting bio-molecules
- The models are often **static**, but there is increasing interest in **dynamic** models, fitted to **time course data**
- Many approaches, including **Dynamic Bayesian Networks** (DBNs) for discretised data and sparse **Dynamic Linear Models** (DLMs) for (normalised) continuous data
- A special case of the DLM is the sparse vector auto-regressive model of order 1, known as the **sparse VAR(1) model**, and this appears to be a particularly effective model for uncovering dynamic network interactions (Opgen-Rhein & Strimmer, 2007)

# Sparse VAR(1) model

- Observe a  $p$ -dimensional vector  $X_t$ , at each of  $n$  time points,  $t = 1, \dots, n$  (with  $p \gg n$ )

$$X_{t+1} = \mu + A(X_t - \mu) + \epsilon_t, \quad \epsilon_t \sim N(0, V)$$

- The  $p \times p$  matrix  $A$  is assumed to be **sparse** (ie. most elements are expected to be exactly zero)
- Sparsity can be modelled in many ways. Simplest:

$$\Pr(a_{ij} \neq 0) = \pi, \quad \forall i, j, \quad a_{ij} | a_{ij} \neq 0 \sim N(0, \sigma^2), \quad \forall i, j$$

- The non-zero structure of  $A$  can be associated with a **graph** (network) of **dynamic interactions** (non-zero  $a_{ij}$  implies arc from node  $j$  to node  $i$ )

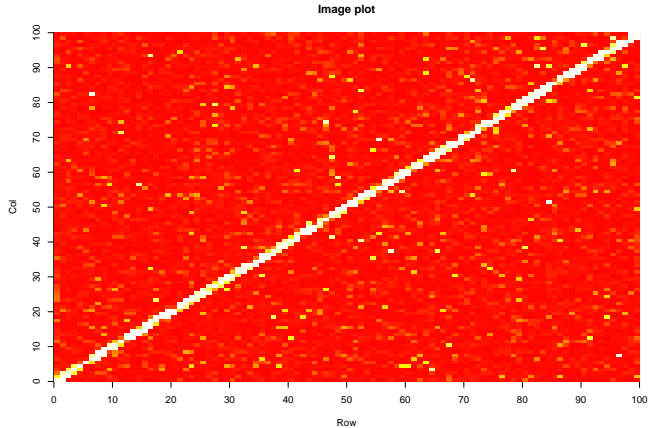
## Inference for model parameters and structure from data

- Can get a **point estimate** for the network structure by computing a **shrinkage estimate** of  $A$  and then thresholding (Opgein-Rhein & Strimmer, 2007)
- Can also use **Bayesian MCMC methods** to explore the space of plausible interaction graphs
- MCMC methods allow computation of useful quantities such as  $\Pr(a_{ij} \neq 0 | \mathcal{D})$
- Inference for graphs is a hard problem...

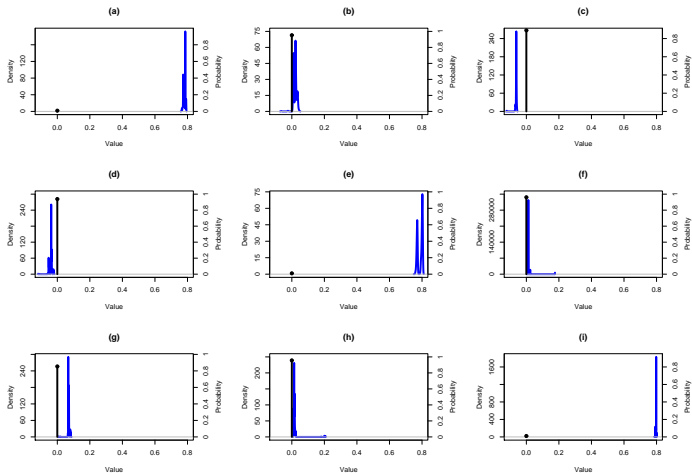
# MCMC for sparse VAR(1) models

- **RJ-MCMC algorithm** to explore both graphical structure and model parameters (auto-regressive coefficients, mean vector, variance components) — routine to develop and implement, but exhibits very poor mixing in high-dimensional settings
- Conditional on the graphical structure, possible (but messy) to develop a **variational algorithm** which gives an approximate marginal log-likelihood for the model after a few iterations — can embed this in a very simple MCMC algorithm to explore just the graphical structure
- Even this algorithm **mixes poorly** for large  $p$  (say,  $p > 200$ ), but there are  $2^{p^2}$  graphs, after all...
- Could probably get reasonable speed-up by using (parallel) **sparse matrix algorithms**

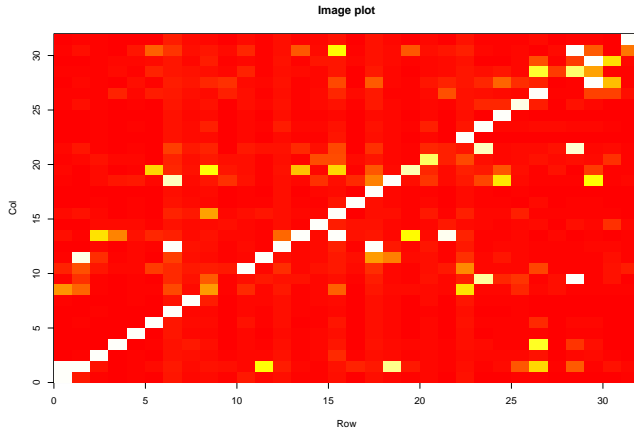
# Sparse VAR(1) sparsity structure



# Sparse VAR(1) autoregressive parameter posteriors



# Application to some EEG data



# Linking VAR(1) models to the CLE

- In the CLE, expand  $h(x_t, c)$  about a “typical” value,  $\bar{x}$

$$h(x_t, c) \simeq h(\bar{x}, c) + \left. \frac{\partial h}{\partial x} \right|_{\bar{x}} (x_t - \bar{x})$$

- Plug into CLE (first-order for the drift and zeroth-order for diffusion)

$$dX_t = S \left[ h(\bar{x}, c) + \left. \frac{\partial h}{\partial x} \right|_{\bar{x}} (X_t - \bar{x}) \right] dt + \sqrt{S \operatorname{diag}\{h(\bar{x}_t, c)\} S'} dW_t$$

- Gives a multivariate Gaussian **Ornstein-Uhlenbeck** (OU) process

# Linear Gaussian OU process

- Linearisation of the CLE gives a Gaussian OU process

$$dX_t = -H(X_t - \mu)dt + \Lambda dW_t$$

where

$$H = -S \frac{\partial h}{\partial x} \Big|_{\bar{x}}$$
$$\mu = -\bar{x} - H^{-1} S h(\bar{x}, c)$$
$$\Lambda = \sqrt{S \operatorname{diag}\{h(\bar{x}_t, c)\} S'}$$

- the auto-regressive matrix  $H$  is sparse (due to the sparsity of  $S$  and the Jacobian)

# Time discretisation

- The Gaussian OU process is analytically tractable
- In particular, it can be time-discretised **exactly**, giving a VAR(1) model with auto-regressive matrix

$$A = \exp\{-H\Delta t\}$$

- Although  $H$  will be very sparse,  $A$  typically won't be as sparse, due to “fill-in”
- Two possibilities:
  - 1 Although  $A = \exp\{-H\Delta t\}$  isn't so sparse, the first order approx  $A \simeq I - H\Delta t$  is (could get same result by discretising time first with an Euler approximation and then linearising)
  - 2 Don't fit VAR(1) models with sparse  $A$ , but models with sparse  $\log(A)$

## Issues






- $H$  doesn't have the same sparsity structure as  $S$  in general, so there's still work to do to get at sparsity of  $S$ , even if we can infer the sparsity of  $H$
- Population averaged versus single-cell data — complicates argument slightly (the mean of the CLE is typically **not** the RRE), but doesn't change the essential conclusions
- Consistency of sparsity structures under a **log transformation**
- Consistency of sparsity structures under **marginalisation** — can't measure everything of interest to a bottom-up modeller in a top-down high-throughput experiment (rules from graphical modelling theory can be applied)

# Conclusions

- The CLE sits at the interface between many different but related bottom-up modelling paradigms
- The Gaussian OU process formed by linearising the CLE links naturally to a class of sparse discrete time linear models used as part of a top-down modelling approach for high-dimensional HTP data
- Direct fully Bayesian inference for the CLE based on discrete time course data is also possible, and is especially useful for inferring quantitative dynamics
- Likelihood-free methods for inference can sometimes work well in this context, and are very simple and general in terms of their applicability

# Acknowledgements

- **Stochastic kinetic models:** Richard Boys, Tom Kirkwood, Colin Gillespie, Conor Lawless, Pete Milner, Daryl Shanley, Carole Proctor, Wan Ng (funding from BBSRC, EPSRC, MRC, DTI, Unilever)
- **Likelihood-free modelling:** Richard Boys, Daniel Henderson, Eryk Wolski, Jake Wu (funding from BBSRC)
- **Inference for diffusions:** Andrew Golightly (studentship funding from EPSRC)
- **Sparse VAR(1) modelling:** Richard Boys, Colin Gillespie, Adrian Houghton, Guiyuan Lei (funding from BBSRC and studentship funding from EPSRC)

-  Boys, R. J., D. J. Wilkinson and T. B. L. Kirkwood (2008). Bayesian inference for a discretely observed stochastic kinetic model. *Statistics and Computing*, 18(2), 125–135.
-  Golightly, A. and D. J. Wilkinson (2006). Bayesian sequential inference for stochastic kinetic biochemical network models. *Journal of Computational Biology*. 13(3), 838–851.
-  Golightly, A. and D. J. Wilkinson (2008). Bayesian inference for nonlinear multivariate diffusion models observed with error. *Computational Statistics and Data Analysis*. 52(3), 1674–1693.
-  Henderson, D. A., Boys, R. J., Krishnan, K. J., Lawless, C., Wilkinson, D. J. (2008). Bayesian emulation and calibration of a stochastic computer model of mitochondrial DNA deletions in substantia nigra neurons. *Journal of the American Statistical Association*, to appear.
-  Wilkinson, D. J. (2006). *Stochastic Modelling for Systems Biology*. Chapman & Hall/CRC Press.