Approximate Inference for partly observed Continuous Time Markov Processes

Manfred Opper

Computer Science

collaboration with:

Andreas Ruttor, Florian Stimberg (TU Berlin)

Dan Cornford, Yuan Shen, Remi Barillec, Michail Vrettas (Aston U)

John Shawe–Taylor, Cédric Archambeau (UCL)

Guido Sanguinetti (Edinburgh U)

Overview

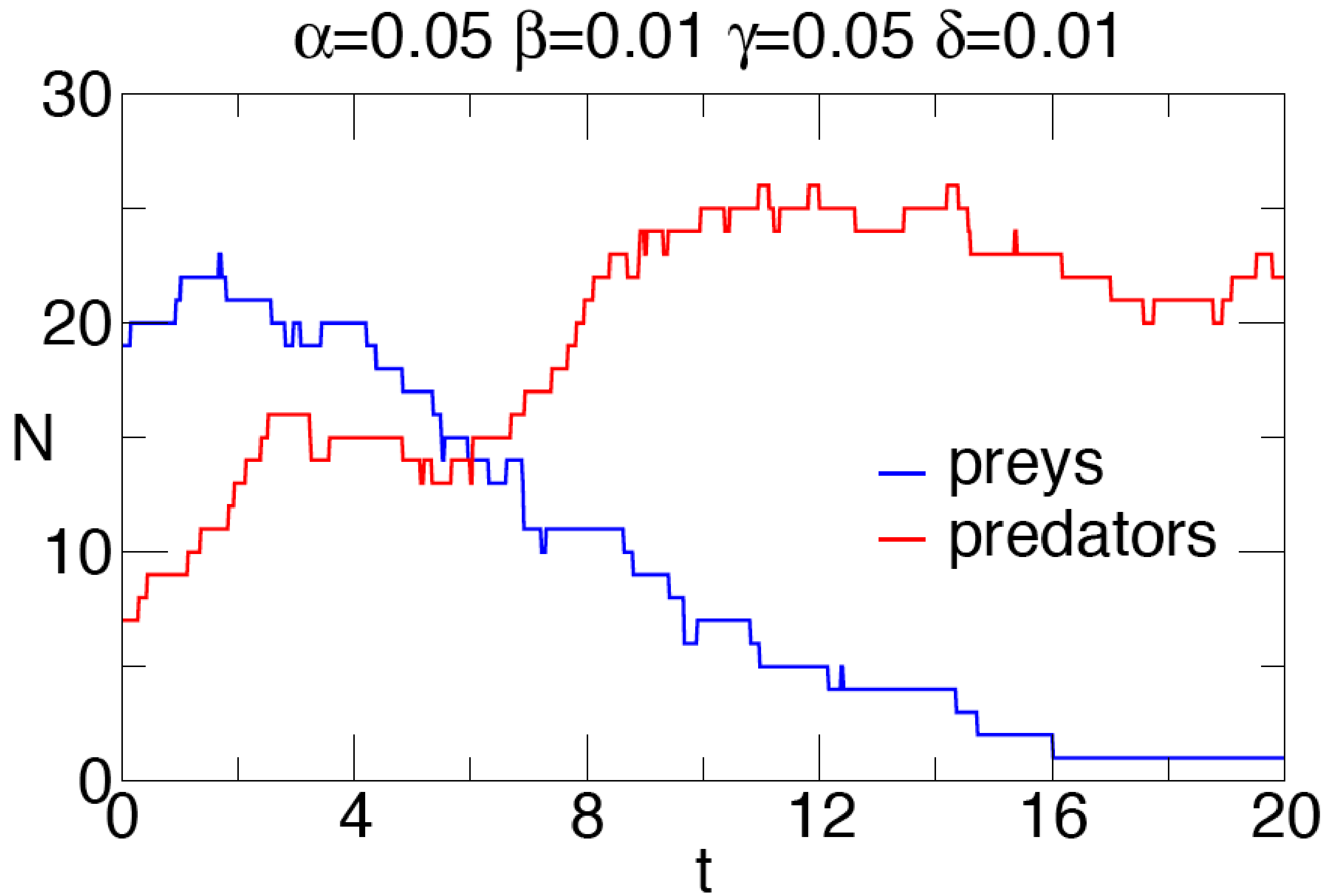
- Examples of partly observed continuous time Markov processes
- Reminder of Markov processes
- Monte Carlo, exact and linear noise approximations
- Variational approaches to inference
- Applications to a model of transcriptional regulation
- Outlook

Stochastic Lotka Volterra Model

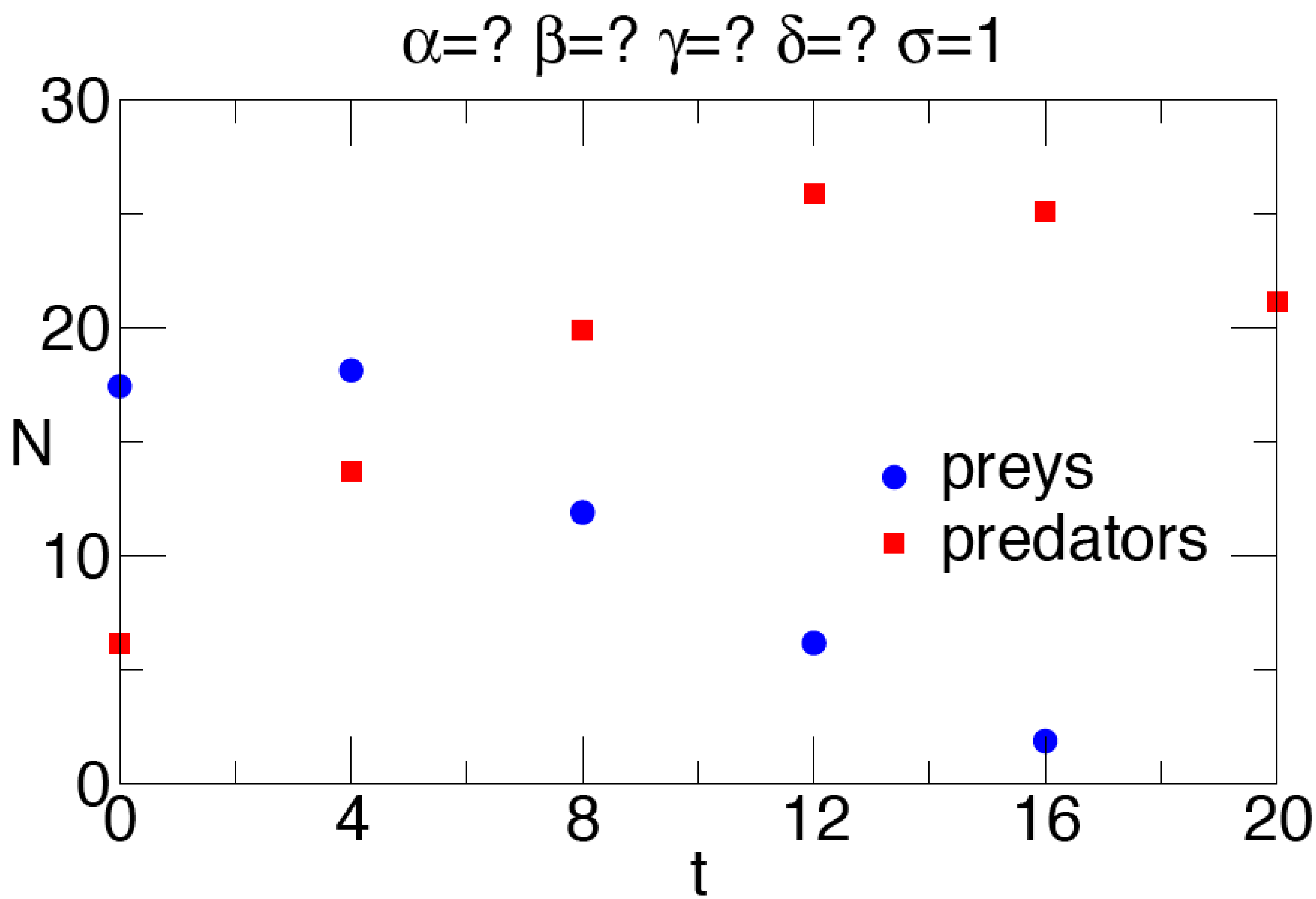
2 interacting species (animals): predators and prey.

Numbers of animals: X_{Prey} and X_{Pred}.

$$
\begin{aligned}
X_{\text{Prey}} &\rightarrow X_{\text{Prey}} + 1 \quad \text{with Rate} \quad \alpha X_{\text{Prey}} \\
X_{\text{Prey}} &\rightarrow X_{\text{Prey}} - 1 \quad \text{with Rate} \quad \beta X_{\text{Prey}} X_{\text{Pred}} \\
X_{\text{Pred}} &\rightarrow X_{\text{Pred}} + 1 \quad \text{with Rate} \quad \delta X_{\text{Prey}} X_{\text{Pred}} \\
X_{\text{Pred}} &\rightarrow X_{\text{Pred}} - 1 \quad \text{with Rate} \quad \gamma X_{\text{Pred}}
\end{aligned}
$$

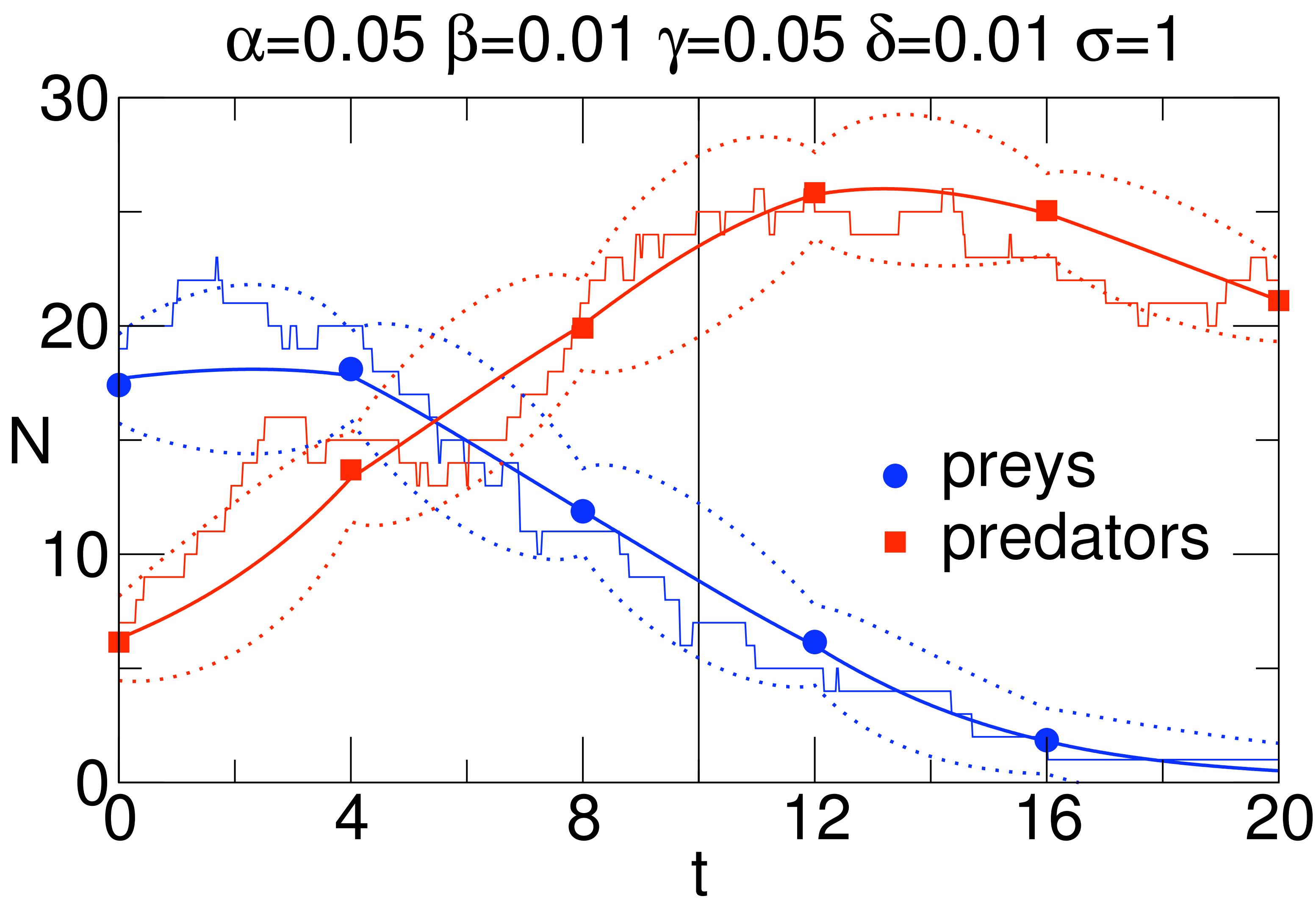


A realisation of the stochastic process and the reaction constants

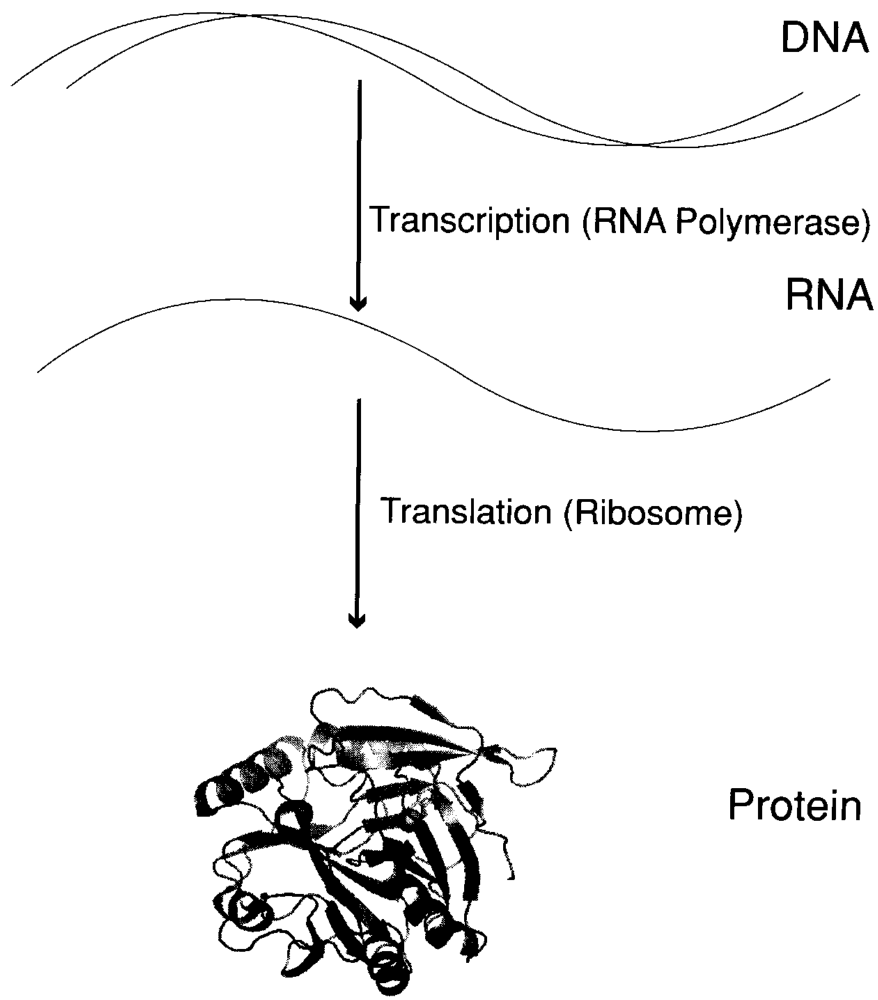


Discrete noisy observations from the continuous time series

What we would like to predict ...



Gene expression



Simple model of autoregulatory network

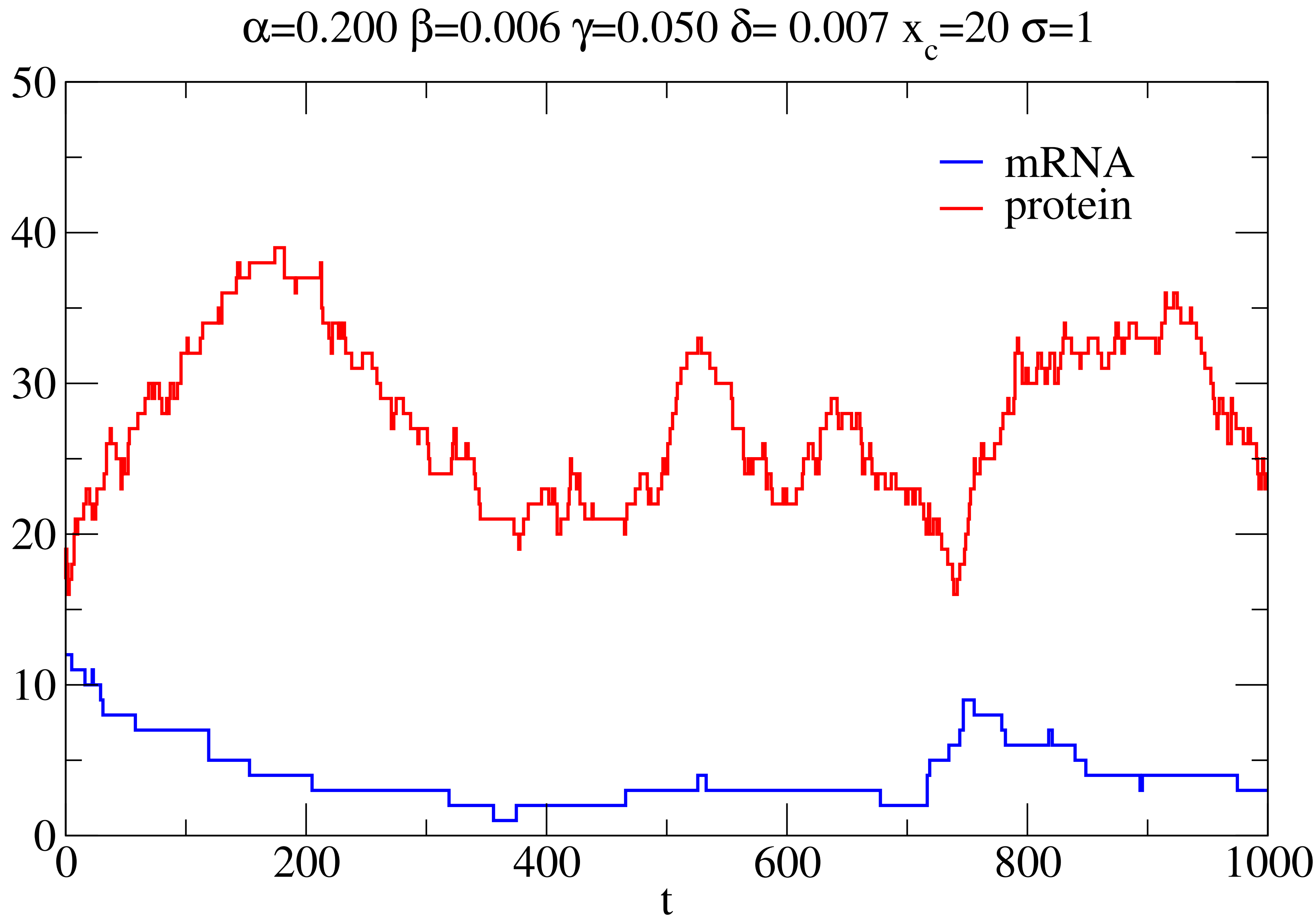
2 interacting molecules: mRNA and a Protein

Number of mRNA and Protein molecules: X_{mRNA}, X_{Prot}

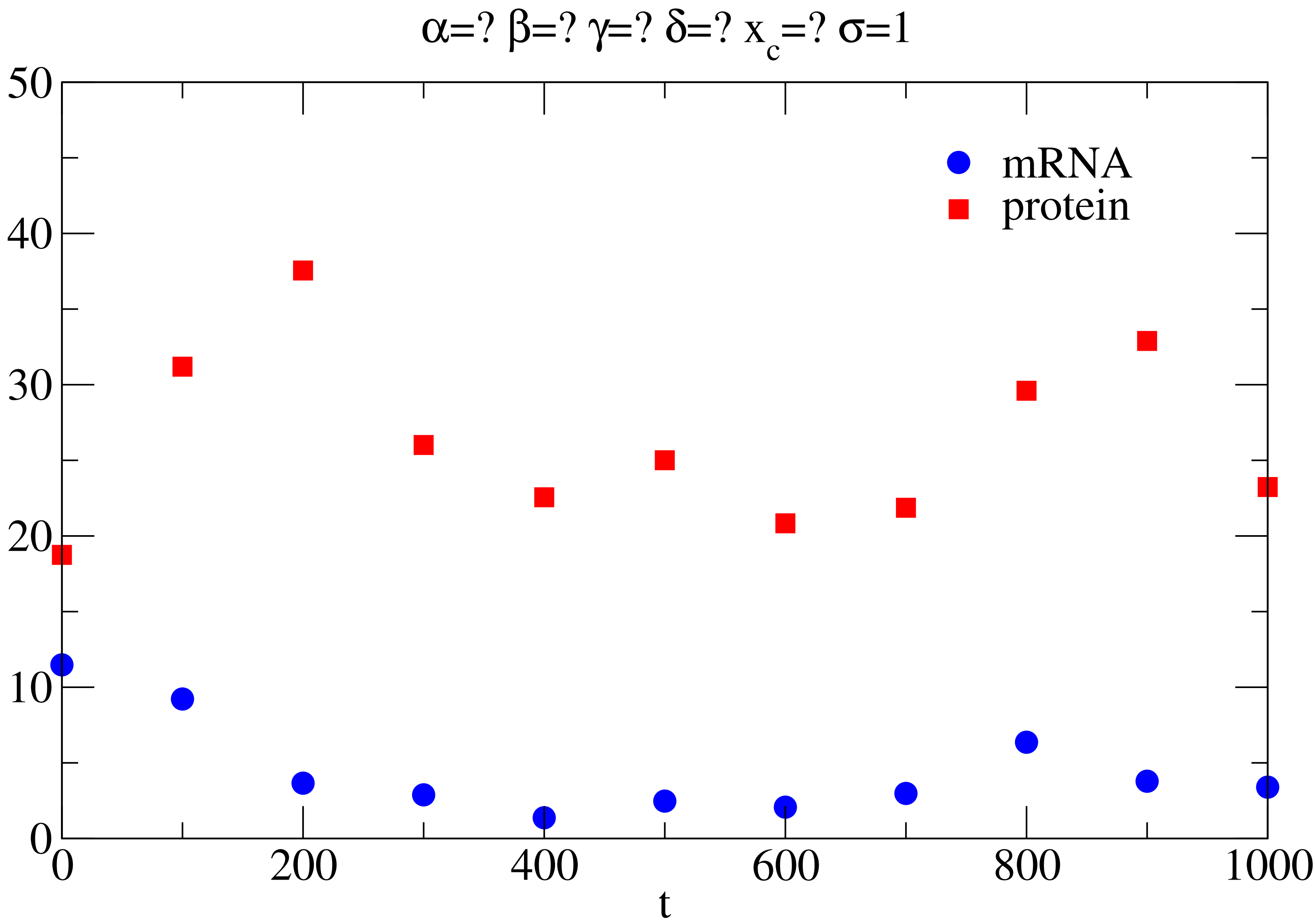
$\mathbf{X} = (X_{\text{mRNA}}, X_{\text{Prot}})$

$$
\begin{array}{rcll}
X_{\text{Prot}} & \rightarrow & X_{\text{Prot}} + 1 & : \text{ with Rate } \quad \gamma X_{\text{mRNA}} \,, \\
X_{\text{Prot}} & \rightarrow & X_{\text{Prot}} - 1 & : \text{ with Rate } \quad \delta X_{\text{Prot}} \,, \\
X_{\text{mRNA}} & \rightarrow & X_{\text{mRNA}} - 1 & : \text{ with Rate } \quad \beta X_{\text{mRNA}} \,, \\
X_{\text{mRNA}} & \rightarrow & X_{\text{mRNA}} + 1 & : \text{ with Rate } \quad \alpha(1 - \alpha_c \Theta(X_{\text{Prot}} - \theta_c)) \,,
\end{array}
$$

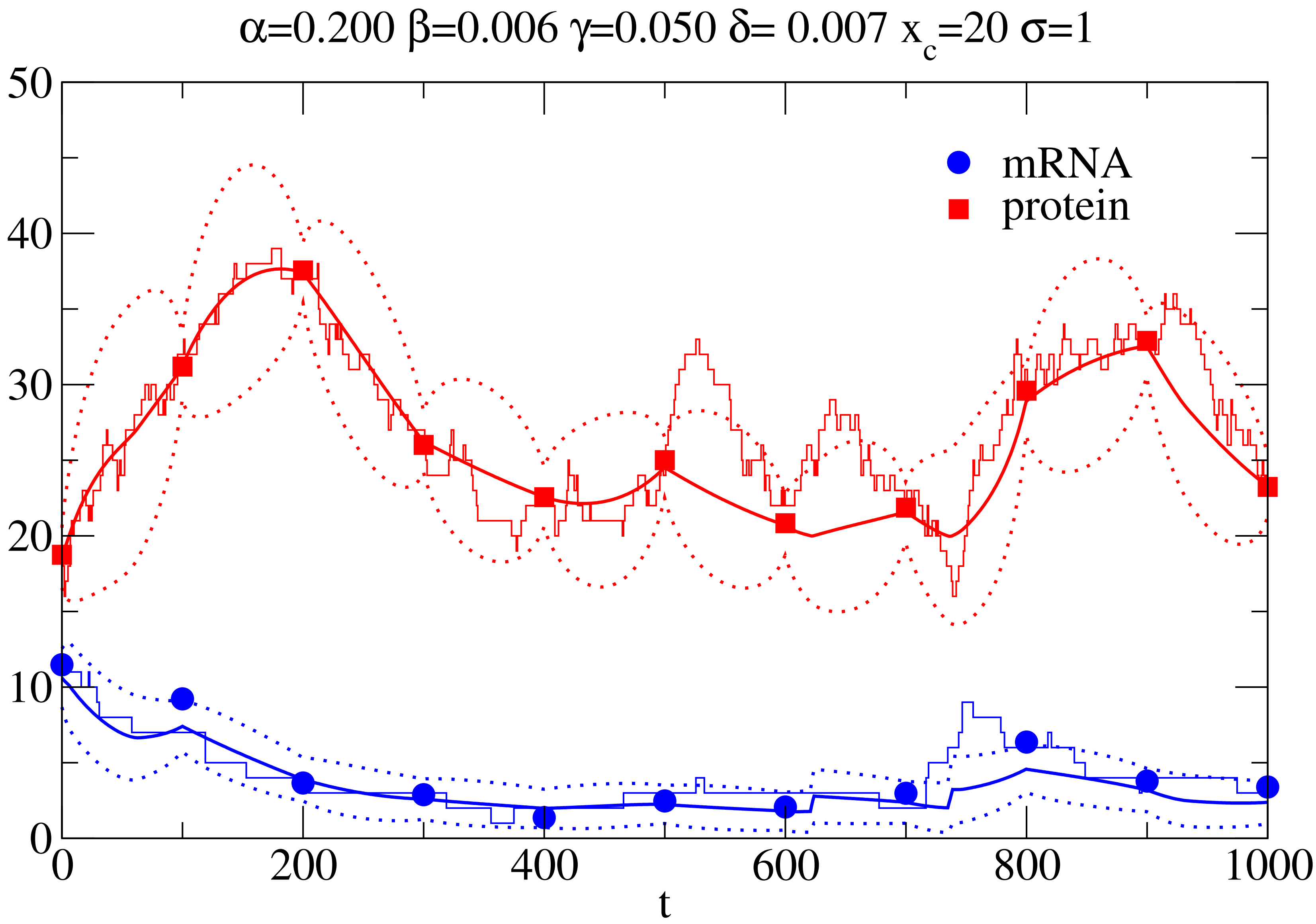
where $\Theta(x) = 1$ if $x \geq 0$ and $\Theta(x) = 0$ for $x < 0$.



Simulation of process



(Noisy) observations at discrete times.



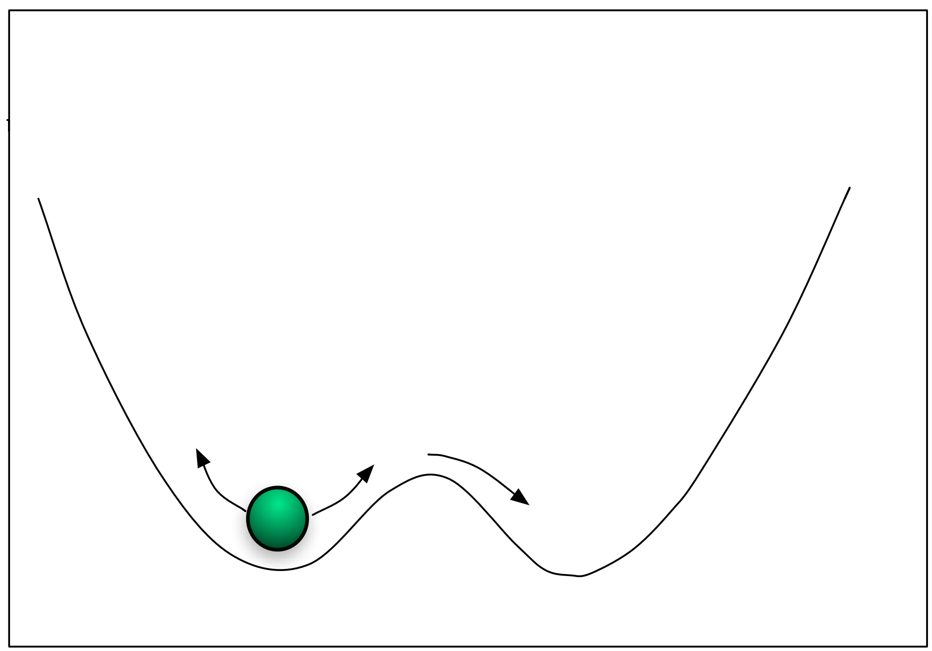
Inference of state variable.

Motion in double-well potential

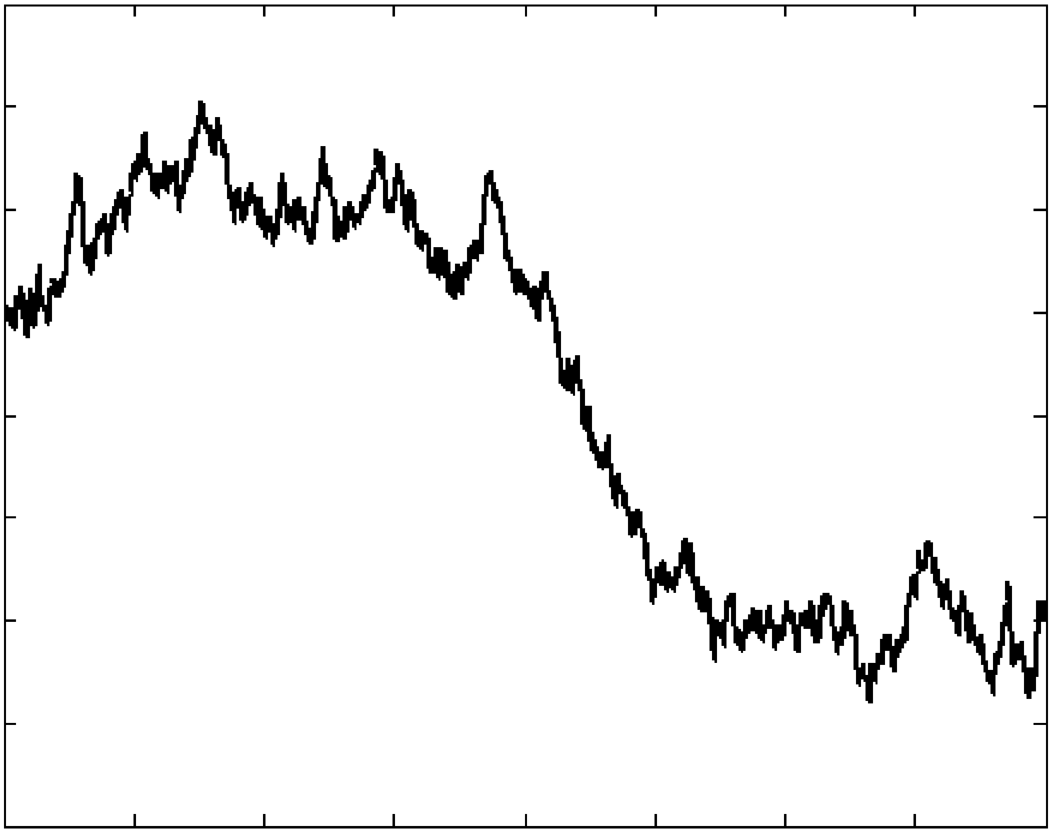
$$dX = f(X)dt + \sigma^2 dW.$$

with $f(x) = -\frac{dV(x)}{dx}$

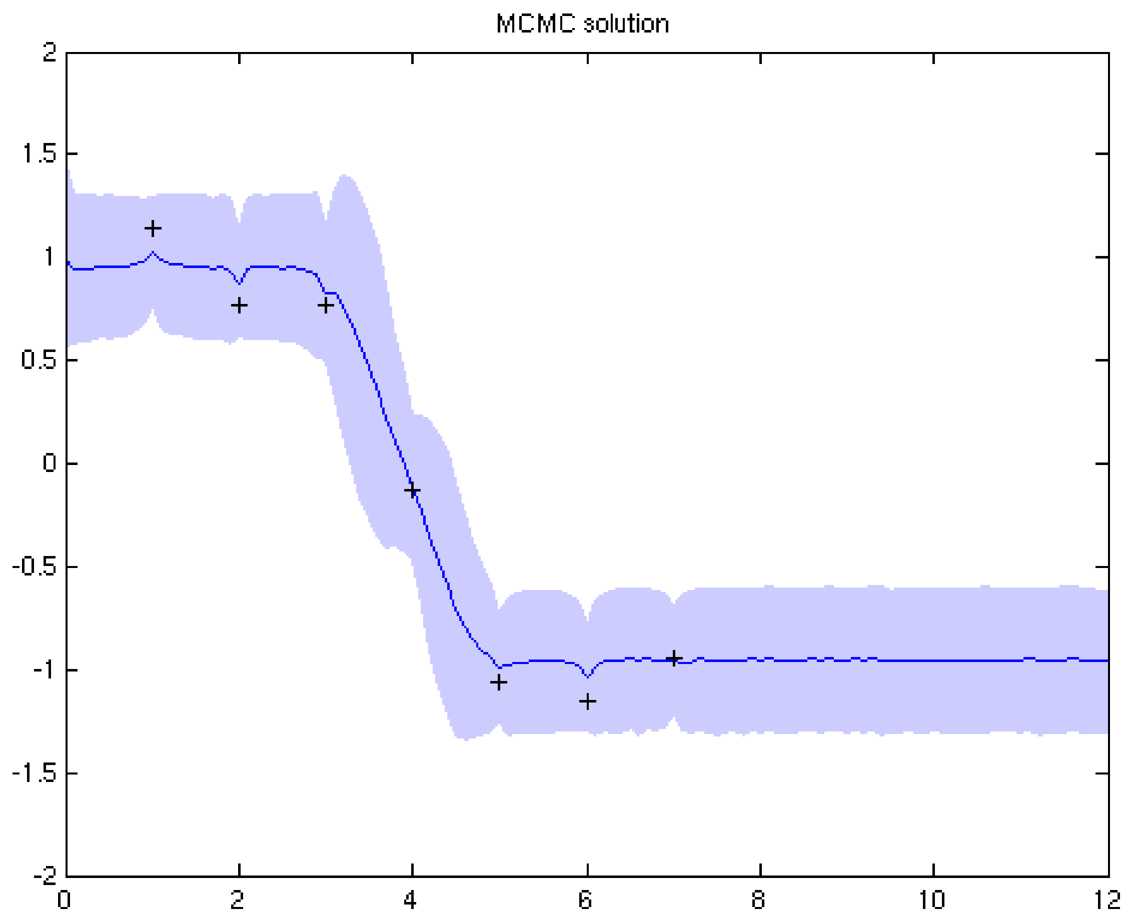
and $V(x)$ is a double well potential



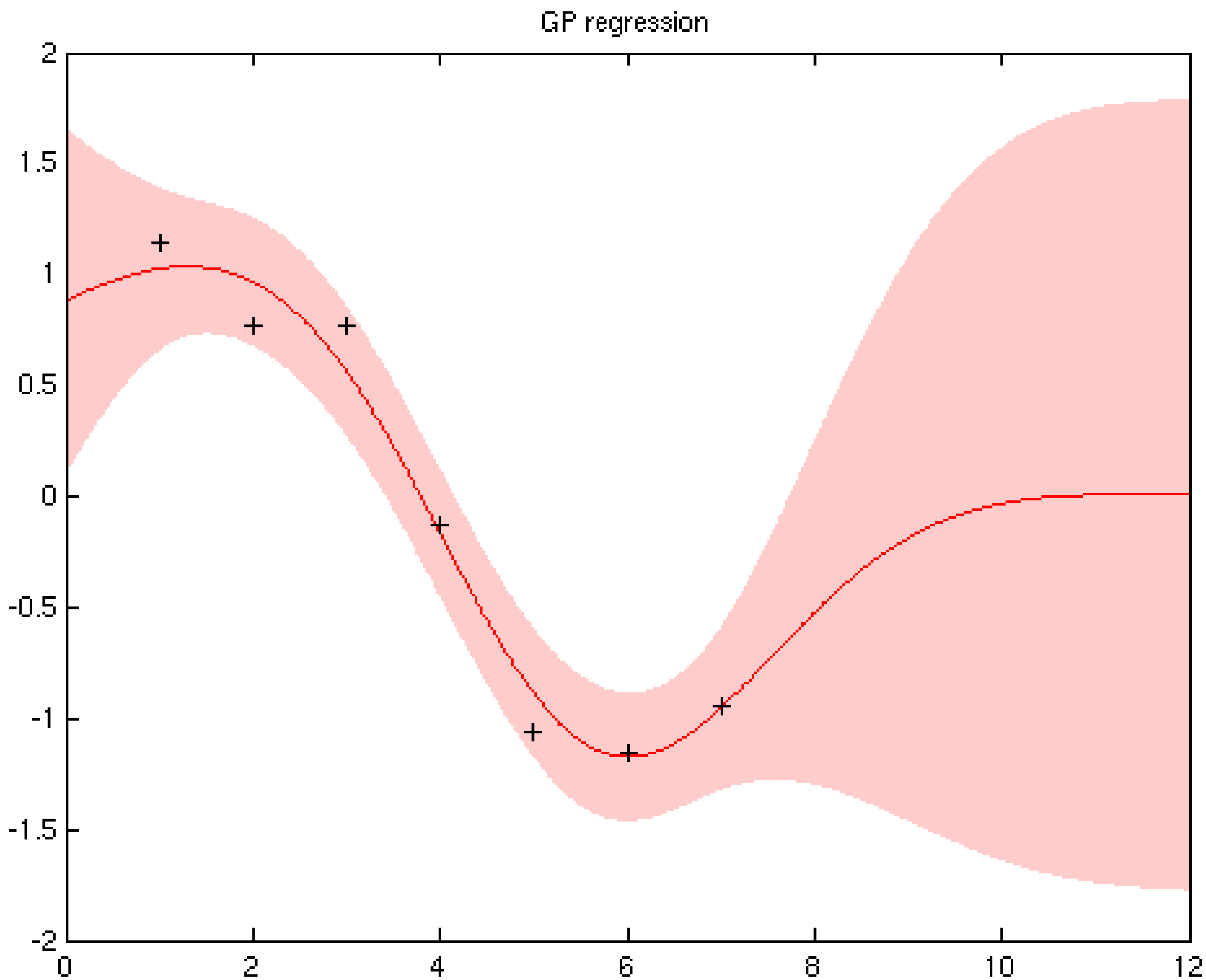
A sample path might look like this



Observations & optimal prediction



If we forget about the model and use a Bayesian regression ...



A reminder of Markov Processes

- Discrete time (Markov chain)

$$P(X_{n+1} \in A | X_n = x, X_{n-1} = x_{n-1}, \ldots, X_0 = x_0)$$

$$= P(X_{n+1} \in A | X_n = x)$$

- Continuous time

$$P(X_{t+dt} \in A | X_{0:t} = x_{0:t}) = P(X_{t+dt} \in A | X_t = x_t)$$

where $x_{0:t} = \{x(s) | 0 \leq s \leq t\}$ denotes the entire path.

Jump Processes

Assume that X_t jumps between discrete states.

- Short time behaviour of transition kernel defined by **transition rate** f:

$$P(y, t + \Delta t | x, t) \simeq \delta_{xy} + f(y|x, t)\Delta t$$

  for $\Delta t \to 0$. Note that $f(x|x, t) = -\sum_{z \neq x} f(z|x, t)$.

- Density for time t to next event (homogeneous process)

$$p(t) = -f(x|x)\, e^{f(x,x)t}$$

Discrete event simulation

1. initialise X_0.

2. Let $X_t = x$. Simulate waiting time t' to next event with density $p(t') = -f(x|x)\, e^{f(x|x)t'}$.

3. Put $t := t + t'$

4. Draw new state x' with prob $-\frac{f(x'|x)}{f(x|x)}$

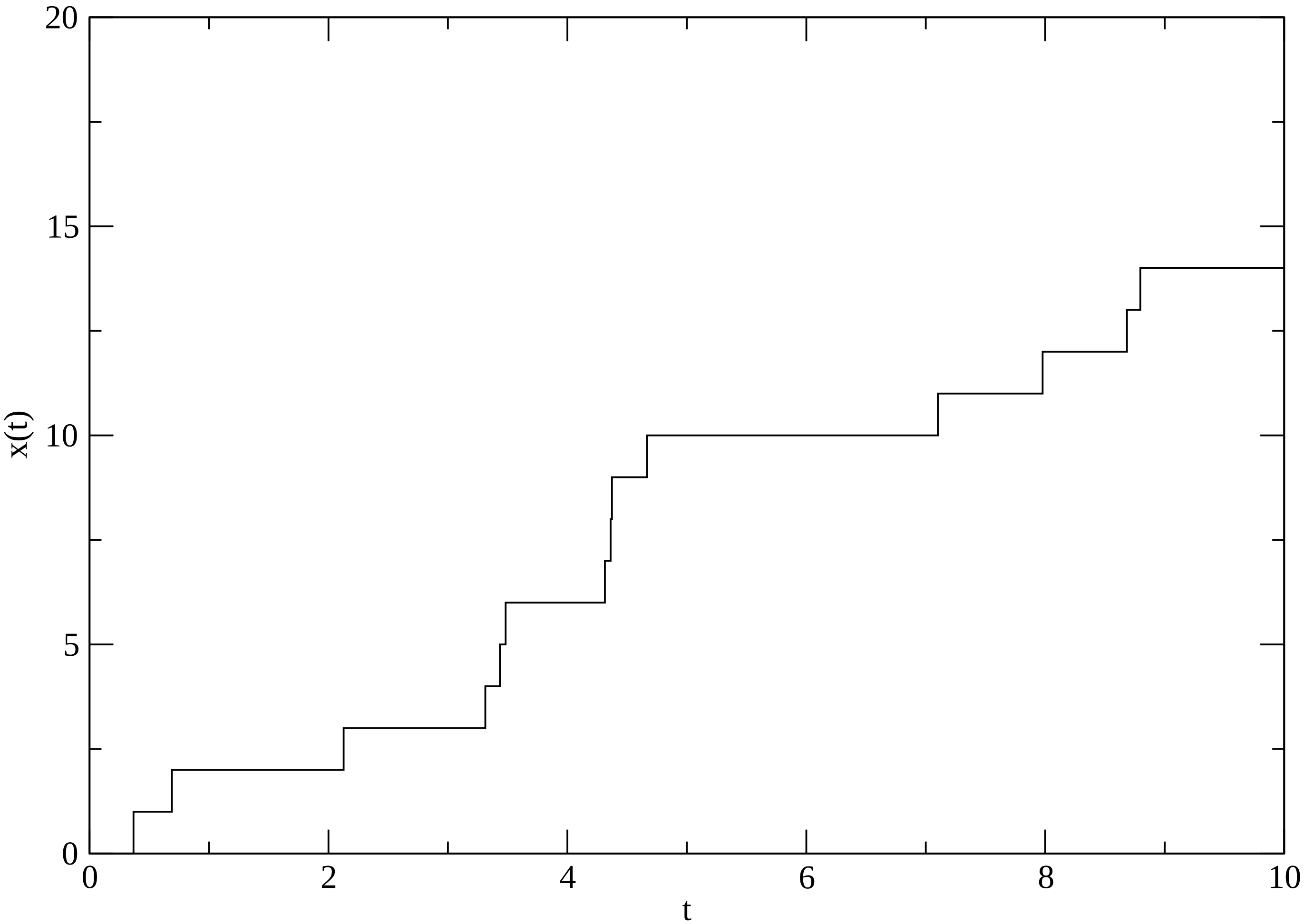
5. Output $X_t = x'$

6. If $t < T_{max}$ return to step 2.

This is the basis for the *Gillespie algorithm*.

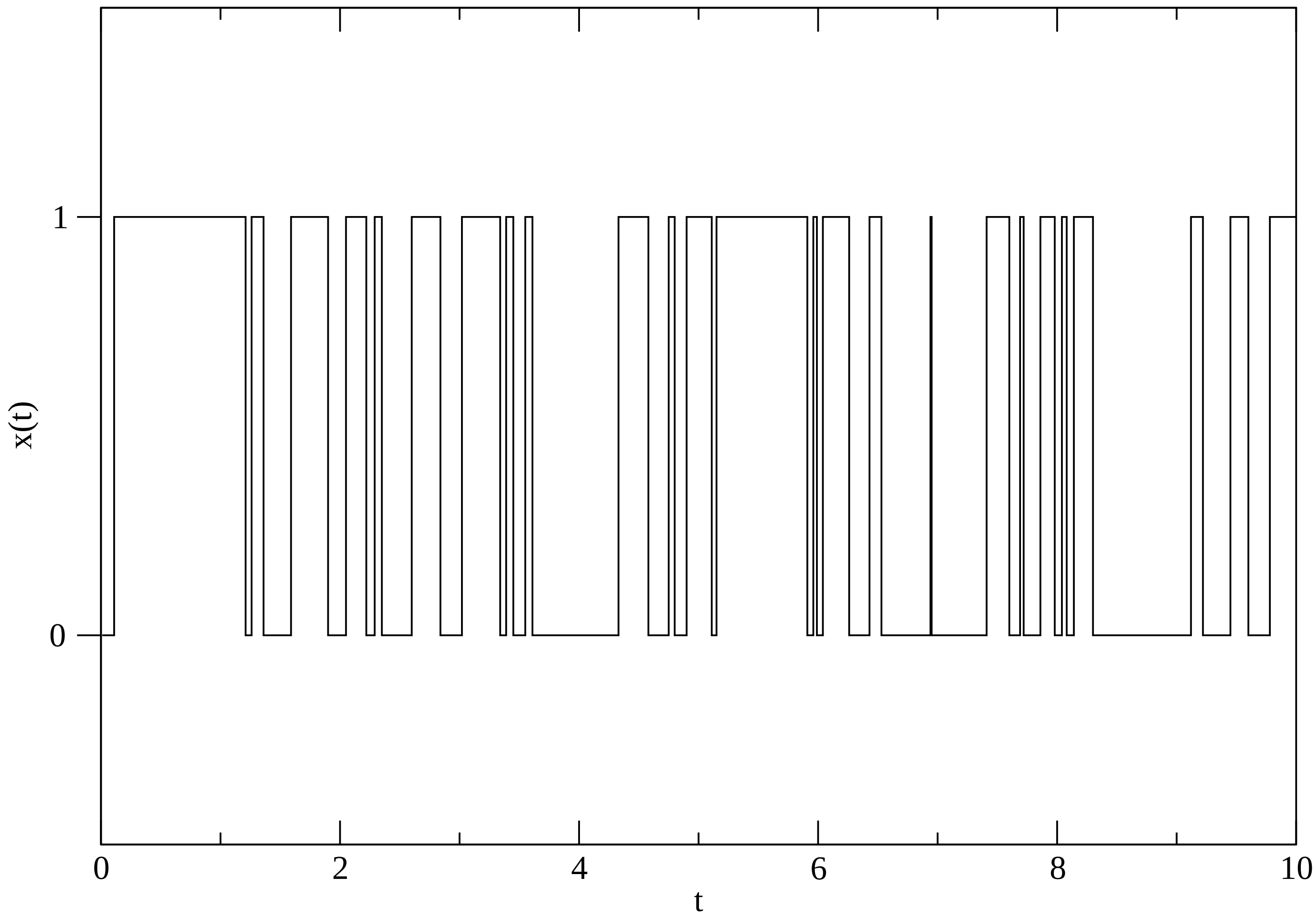
Examples

Poisson Process: $X_t \in N_0$ and $f(x'|x) = \lambda \delta_{x',x+1}$



Telegraph Process: $X_t \in \{0, 1\}$

$$f(x'|x) = \lambda \delta_{x',1-x}$$



Forward and Backward equations

- Forward (Master) equation

$$\frac{dP\left(x,t|x_0,t_0\right)}{dt}=\sum_{x'\neq x}\left[P\left(x',t|x_0,t_0\right)f\left(x|x'\right)-P\left(x,t|x_0,t_0\right)f\left(x'|x\right)\right].$$

- Kolmogorov Backward equation

$$\frac{dP(x',t'|x,t)}{dt}=\sum_{y\neq x}f(y|x)\left\{P(x',t'|x,t)-P(x',t'|y,t)\right\}$$

Diffusion approximation

Assume that jumps $\Delta x \doteq x' - x << x$. Try approximation by a continuous state Markov process

Introduce **drift** (1. Jump moment)

$$f(x) = \lim \frac{1}{\Delta t} E\left[X_{t+\Delta t} - X_t | X_t = x\right] = \sum_{x' \neq x} f(x'|x)(x' - x)\,,$$

and **diffusion matrix** (2. jump moment)

$$D(x) = \lim \frac{1}{\Delta t} E\left[(X_{t+\Delta t} - X_t)(X_{t+\Delta t} - X_t)^\top | X_t = x\right] = \sum_{x' \neq x} (x' - x) f(x'|x)(x' - x)^\top\,.$$

Drift & diffusion for Lotka - Volterra

$$
\begin{aligned}
&h(X_1 \to X_1 + 1) = \alpha X_1\,, &&h(X_1 \to X_1 - 1) = \beta X_1 X_2\,,\\
&h(X_2 \to X_2 + 1) = \delta X_1 X_2\,, &&h(X_2 \to X_2 - 1) = \gamma X_2\,,
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{f}(\mathbf{x}) &= \begin{pmatrix} \alpha x_1 - \beta x_1 x_2 \\ \delta x_1 x_2 - \gamma x_2 \end{pmatrix}\\
D(\mathbf{x}) &= \begin{pmatrix} \alpha x_1 + \beta x_1 x_2 & 0 \\ 0 & \delta x_1 x_2 + \gamma x_2 \end{pmatrix}
\end{aligned}
$$

Deterministic limit

Neglect fluctuations (diffusion), i.e. assume

$$X_{t+\Delta t} - X_t = f(X_t)\Delta t$$

Hence, $X_t \approx m(t) \doteq E[X_t]$. This yields 'classical' rate equation

$$\frac{dm}{dt} \approx f(m(t))$$

Diffusion approximation

$$X_{t+\Delta t} - X_t = f(X_t)\Delta t + D^{1/2}(X_t)\sqrt{\Delta t}\,\epsilon_k \ .$$

with ϵ_k i.i.d. $\sim \mathcal{N}(0, I)$, has the same 2 conditional moments $f(x)$ and $D(x)$. Jump sizes will go $\to 0$ and process becomes continuous!

We will write this (in the limit) as an **(Ito) stochastic differential equation**

$$dX_t = f(X_t)dt + D^{1/2}(X_t)dW_t$$

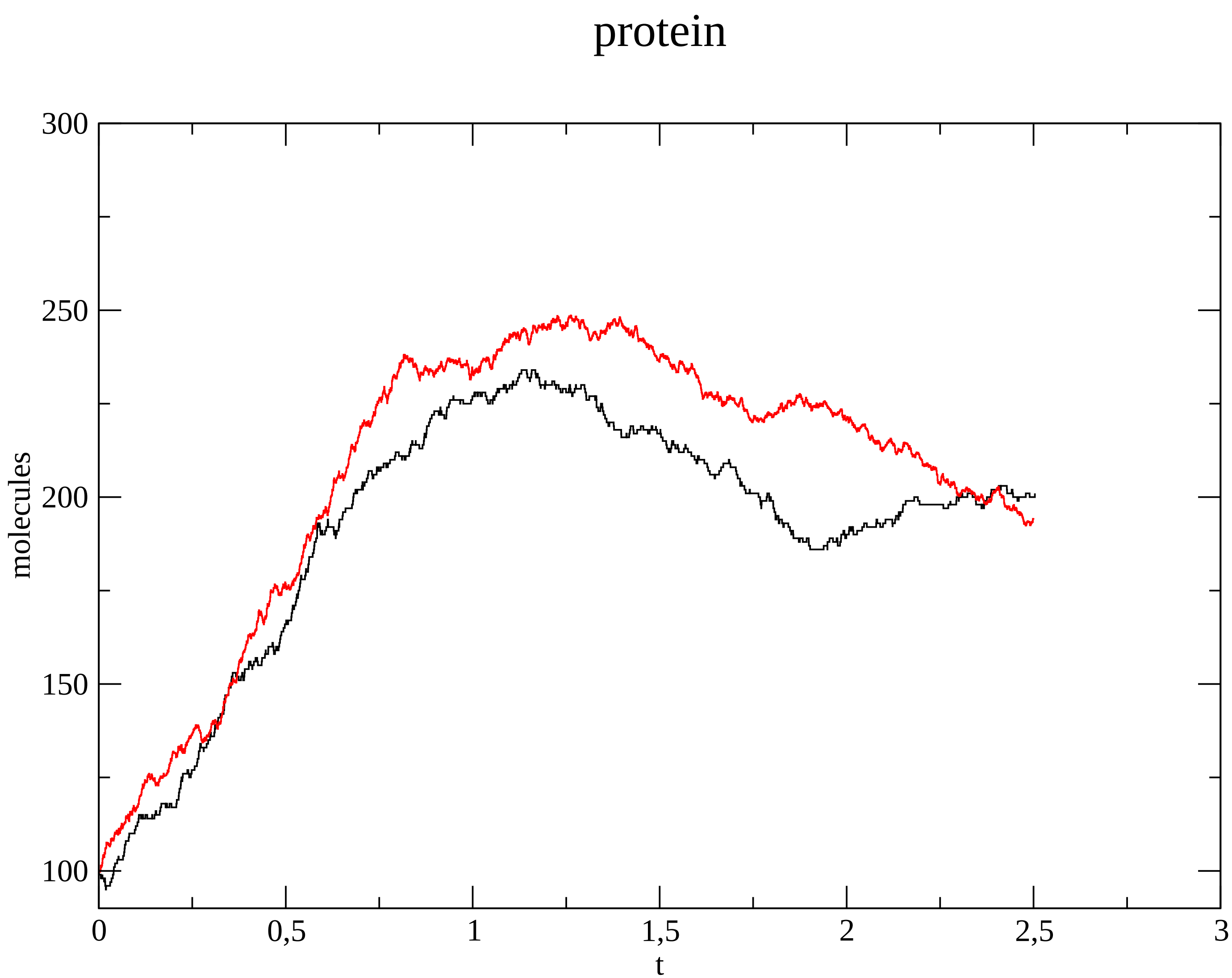
Short time behaviour of transition kernel

for Ito SDE

$$p\left(x', t + \Delta t | x, t\right) =$$

$$= \frac{1}{|2\pi D(x)|} \exp\left[-\frac{1}{2\Delta t}\left((x' - x - f(x)\Delta t\right)^\top D^{-1}(x)\left(x' - x - f(x)\Delta t\right)\right]$$

Simulations of Jump and corresponding diffusion processes



Forward and backward equations for diffusion processes

Let $x \doteq (x_1, \dots x_d)$ and $f(x) = (f_1(x), \dots, f_d(x))$

- Forward (Fokker - Planck) equation

$$\frac{\partial p(x,t|x_0,t_0)}{\partial t} =$$

$$-\sum_i \frac{\partial}{\partial x_i}(f_i(x)p(x,t|x_0,t_0)) \;\; + \;\; \frac{1}{2}\sum_{ij}\frac{\partial^2}{\partial x_i \partial x_j}\Big(D_{ij}(x)p(x,t|x_0,t_0)\Big)$$

- Kolmogorov Backward equation

$$\frac{\partial p(x',t'|x,t)}{\partial t} = -\sum_i f_i(x)\frac{\partial}{\partial x_i}p(x',t'|x,t) - \frac{1}{2}\sum_{ij} D_{ij}(x)\frac{\partial^2}{\partial x_i \partial x_j}p(x',t'|x,t)$$

Short summary

- **Markov jump processes:** Discrete states X

$$p\left(X', t + \Delta t | X, t\right) \simeq \delta_{X'X} + \Delta t \; f\left(X' | X\right)$$

for $\Delta t \to 0$

- **Diffusion processes:** (Ito) stochastic differential equation for state $X(t) \in R^d$

$$dX(t) = \underbrace{f(X(t))}_{\text{Drift}} dt + \underbrace{D^{1/2}(X(t))}_{\text{Diffusion}} \times \underbrace{dW(t)}_{\text{Wiener process}}$$

Limit of discrete time process X_k

$$\Delta X_k \equiv X_{k+1} - X_k = f(X_k)\Delta t + D^{1/2}(X_k)\sqrt{\Delta t}\, \epsilon_k \ .$$

ϵ_k i.i.d. Gaussian.

Inference Problems

Given **noisy observations** $D \equiv (y_1, \ldots, y_N)$ of **hidden process** $X(t_i)$ at times t_i for $i = 1, \ldots, N$.

- Estimate $X(t)$ for $0 \leq t \leq T$ (**smoothing**).

- Estimate system parameters θ in drift f and diffusion D or in rates $f(X'|X)$

Can't we just discretise ....

and run an MCMC (Gibbs) sampler ?

- **Alternate** between sampling paths $p(x_{0:T}|\theta)$ and parameters $p(\theta|x_{0:T})$.

- **Path sampling**: Paths are described by a large vectors $x_{0:T} = (x_1, \dots x_K)$.

  Sample from multivariate density

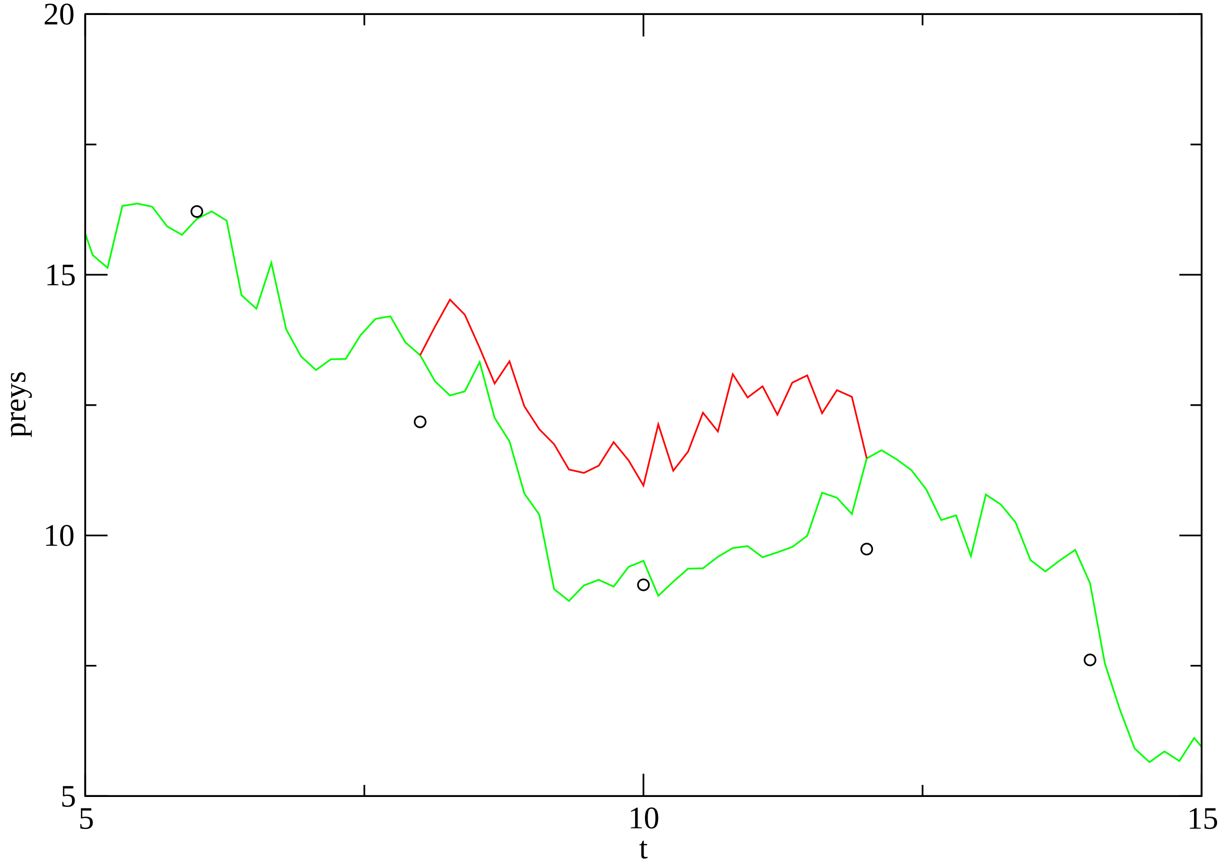
$$q(x_{0:T}) \propto e^{-V(x_{0:T})}$$

  with

$$V(x_{0:T}) = \frac{1}{2\Delta t} \sum_i \Big(x_{i+1} - x_i - f(x_i)\Delta t\Big)^\top D^{-1}(x_i) \Big(x_{i+1} - x_i - f(x_i)\Delta t\Big)$$

$$- \sum_{\text{obs}} \ln p(y_{\text{obs}}|x_{\text{obs}})$$

- Block updates using bridge proposals within Metropolis - Hastings algorithm (Golightly, Wilkinson 2008).



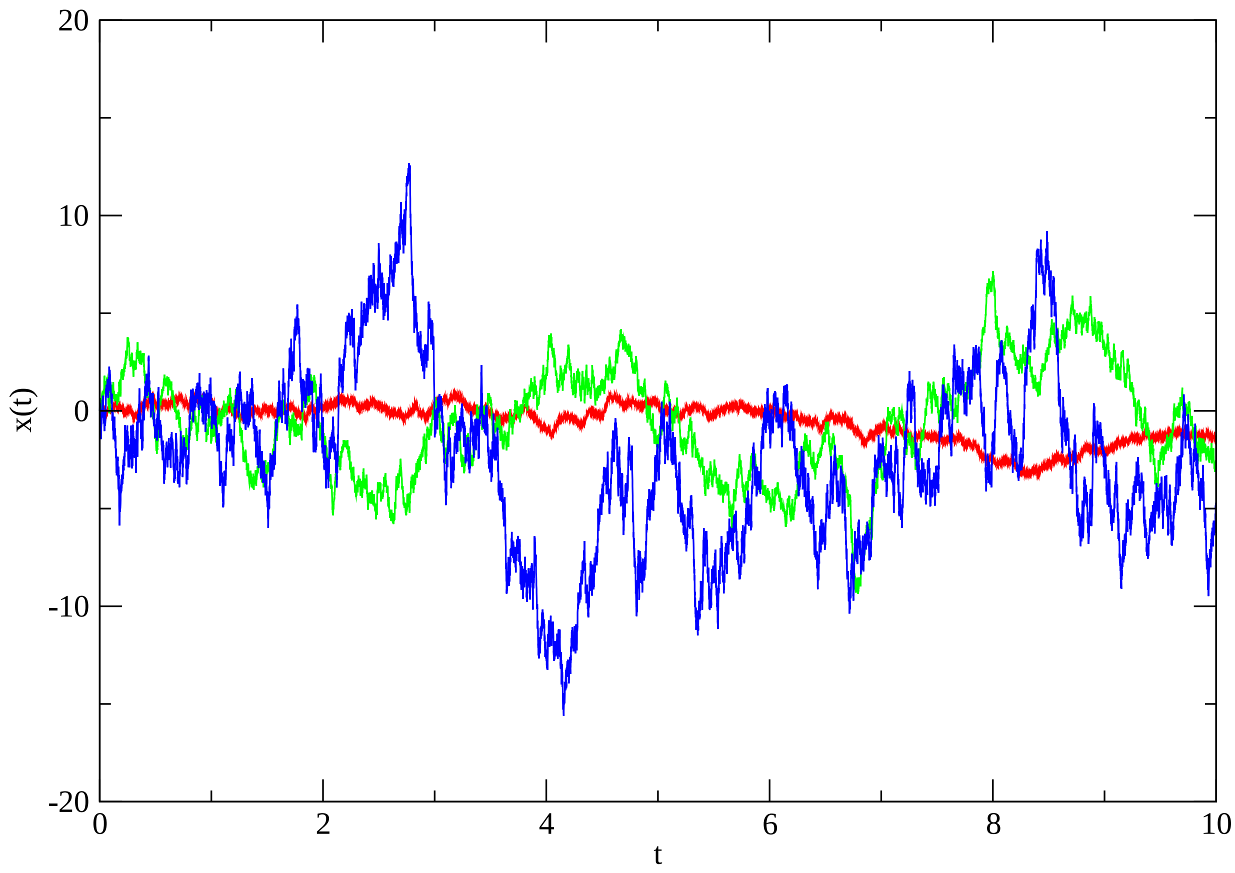
- Alternative: Hybrid Monte Carlo (Alexander et al 2005)

Problem with Gibbs samplers

- Problem: For $\Delta t \to 0$ (complete SDE path) diffusion D and $x_{0:T}$ are highly dependent:

$$\sum_t \left(X_{t+\Delta t} - X_t\right)^2 \to \int_0^T D(X_t)dt$$

- A Gibbs sampler will get stuck in the same (diffusion) parameters for small Δt.



Ornstein Uhlenbeck - processes for $\sigma = 1$ (red), $\sigma = 5$ (green), $\sigma = 10$ (blue).

Isn't this just some sort of hidden Markov model ?

- Yes, but we don't know the transition probabilities.

- Can't we still use a Forward - Backward Algorithm ?

Forward - Backward Solution

- Marginal posterior at time t

$$p_t(x|\{y_i\}_{i=1}^N) \propto \underbrace{p(x_t|\{y_i\}_{t_i<t})}_{\doteq p_F(x,t)} \times \underbrace{p(\{y_i\}_{t_i\geq t}|X_t=x)}_{\doteq \psi(x,t)}$$

- Fokker - Planck equation (between observations)

$$\left\{\frac{\partial}{\partial t} + \nabla f - \frac{1}{2}\mathrm{Tr}(\nabla\nabla^T D)\right\} p_F(x,t) = 0$$

- Backward equation

$$\left\{\frac{\partial}{\partial t} + f^\top\nabla + \frac{1}{2}\mathrm{Tr}(D\nabla^\top\nabla)\right\} \psi(x,t) = 0$$

- & a set of jump conditions at observations.

KSP equations (Kushner '62, Stratonovich '60 & Pardoux '82).

The backward equation is important !

- Posterior process is Markovian and fulfils SDE

$$dX(t) = g(X(t), t)dt + D^{1/2}(X(t))\, dW(t)$$

with an effective time dependent drift $g(X(t), t)$

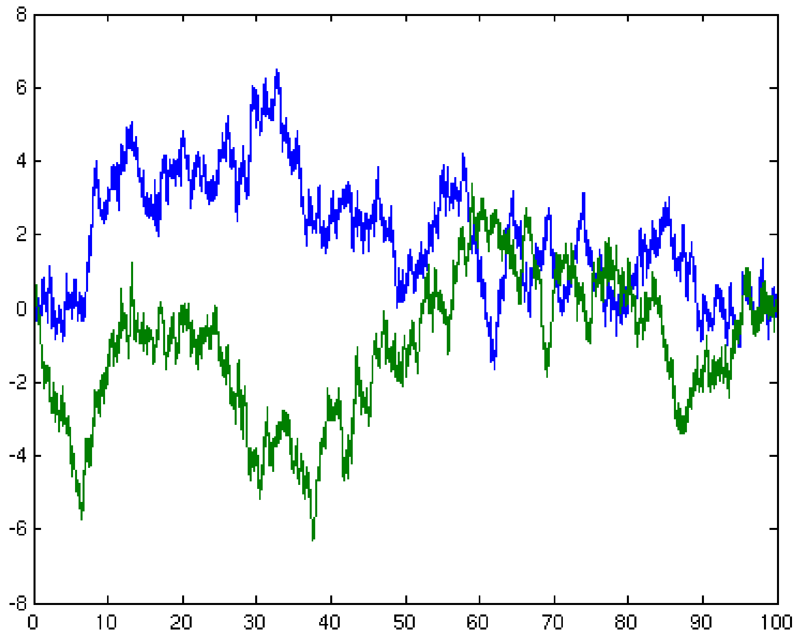
$$g(x, t) = f(x) + D(x, t)\nabla \ln \psi(x, t)$$

and ψ is the solution to the backward equation.

- **Likelihood of observations** $P(\mathbf{y}_1, \ldots, \mathbf{y}_n | X_0 = x) = \psi(x, 0)$.

Example

Wiener process with single, noise free observation $y = x(t = T) = 0$



Posterior drift $g(x,t) = -\frac{x}{T-t}$ for $0 < t < T$.

The linear noise approximation

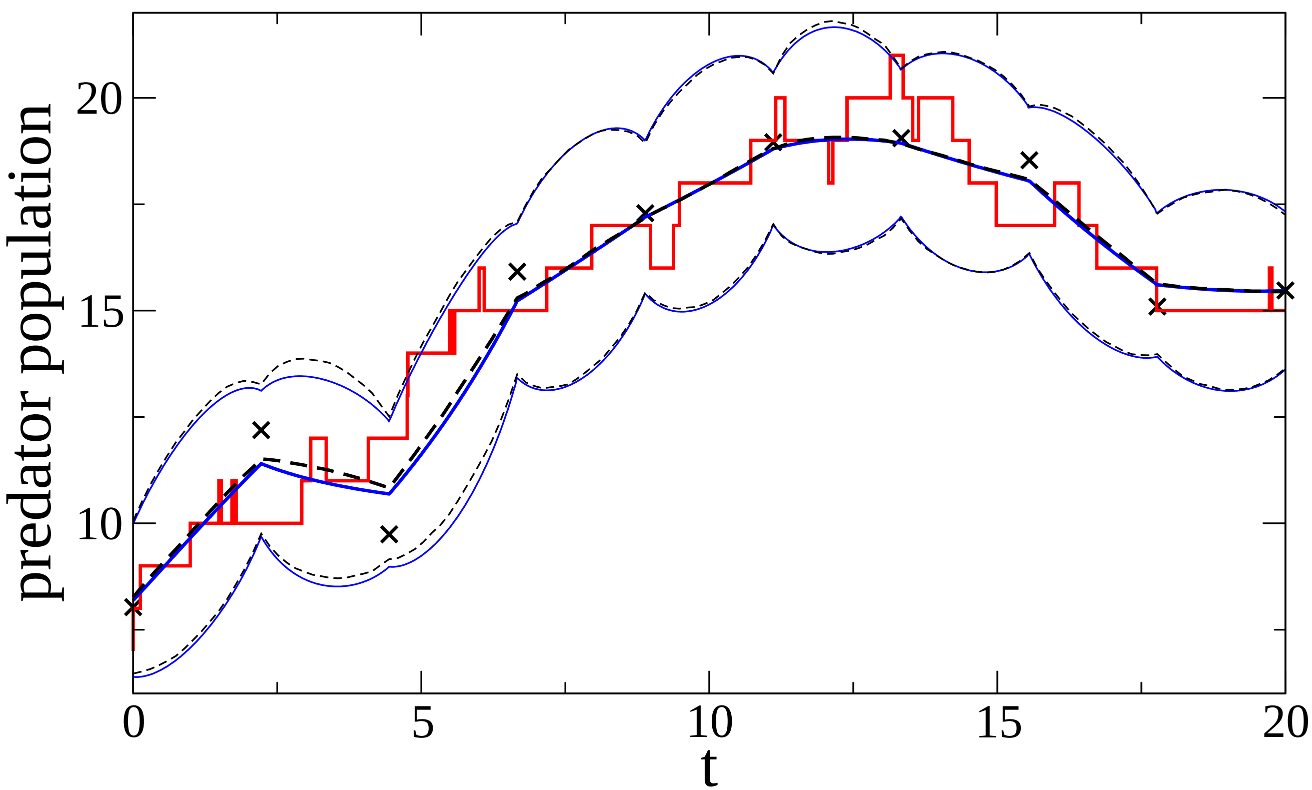
- Similar to extended Kalman filter approach.

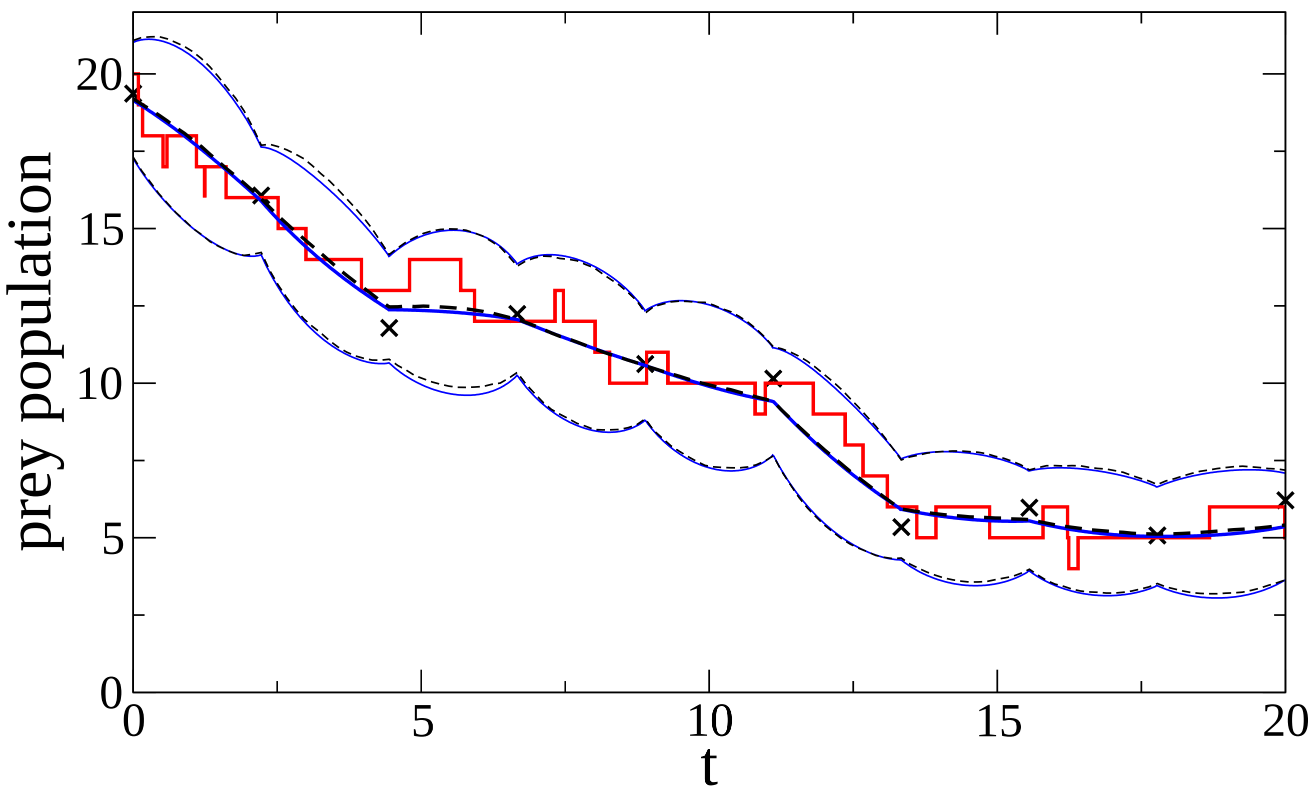
- Expansion in small parameter like 1/Number of molecules (Van Kampen).

- leads to **linear** SDE: X_t becomes approximated by a Gaussian process !

- Can be applied to backward equation $\rightarrow$ ODEs instead of PDEs !

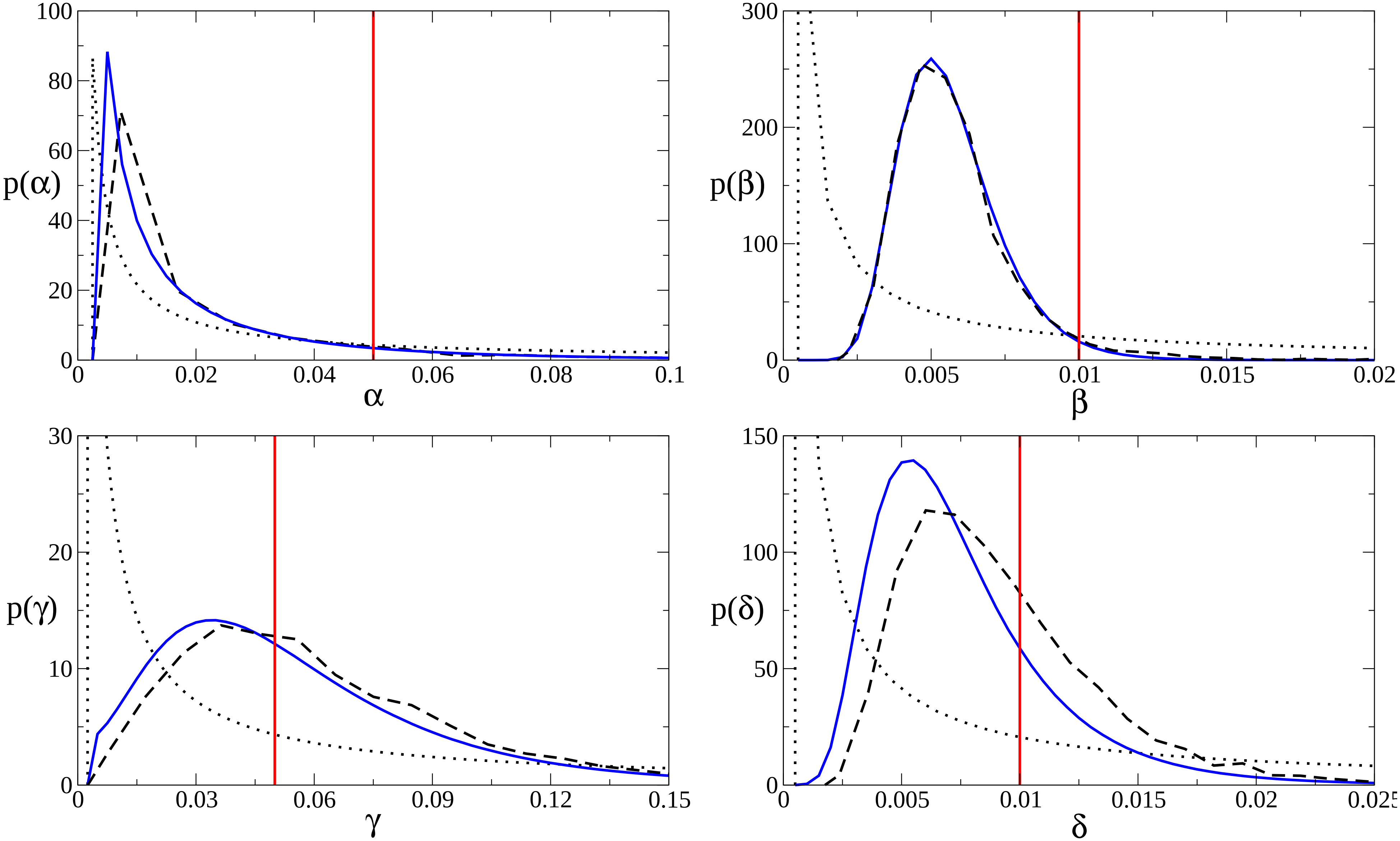
Comparison of linear noise with MCMC





prey population
t
0
5
10
15
20
0
5
10
15
20

Posterior for parameters



Comments

- Good approximation for small fluctuations

- Does not work well for multistable systems

- Gaussian approximation not optimised !

The variational approach in machine learning

Let $P(X|D) = \frac{P(D|X)P(X)}{P(D)}$. Then we have

$$P(\cdot|D) = \arg\min_Q \mathcal{F}[Q]$$

where the *variational free energy* is defined as

$$\mathcal{F}[Q] = KL(Q|P_{prior}) - E_Q[\ln P(D|\cdot)]$$

with the Kullback - Leibler divergence

$$KL(Q|P) = E_Q[\ln \frac{Q}{P}] \ .$$

and

$$\min_Q \mathcal{F}[Q] = P(D)$$

Parameter inference $P(D) \to p(D|\theta)$: Approximate posterior of parameters (Lappalainen, 2000)

$$q(\theta|D) \approx \frac{e^{-\mathcal{F}_\theta[q]}\ p(\theta)}{\int e^{-\mathcal{F}_\theta[q]}\ p(\theta)\ d\theta}.$$

The posterior measure

Conditional (posterior) distribution over **paths** $X_{0:T}$ given the data $D = (y_1, \ldots, y_n)$

$$P(X_{0:T}|D) = \frac{P_{prior}(X_{0:T})}{P(D)} \times \prod_{n=1}^{N} p(y_n|X(t_n)),$$

The Kullback - Leibler divergence for paths

Discretize !

Prior and posterior over paths are Markovian with transition probabilities $P(x', t'|x, t_k)$ and $Q(x', t'|x, t_k)$.

$$
\begin{aligned}
KL\,[Q\|P] &= \int dx_{0:T}\, Q(x_{0:T}) \ln \frac{Q(x_{0:T})}{P(x_{0:T})} \\
&= \sum_{k=0}^{K-1} \int dx\, Q(x, t_k) \int dx'\, Q(x', t_k + \Delta t|x, t_k) \ln \frac{Q(x', t_k + \Delta t|x, t_k)}{P(x', t_k + \Delta t|x, t_k)} \\
&= \sum_{k=0}^{K-1} \int dx\, Q(x, t_k) KL\,[Q(\cdot, t_k + \Delta t|x, t_k)\|P(\cdot, t_k + \Delta t|x, t_k)]
\end{aligned}
$$

... use short time transition probability

For diffusions

$$p\left(X', t + \Delta t | X, t\right) \propto \exp\left[-\frac{1}{2\Delta t}\left\|\Delta X - f(X)\Delta t\right\|_D^2\right]$$

as $\Delta t \to 0$,

with $\|F\|_D^2 = F^\top D^{-1} F$.

and take the limit $\Delta t \to 0$

Let Q and P be measures over paths for SDEs with drifts $g(X,t)$ and $f(X,t)$ with **same diffusion** $D(X)$. Then

$$KL\,[Q\|P] = \frac{1}{2}\int_0^T dt\left\{\int dx\; q(x,t)\;\; \|g(x,t) - f(x,t)\|_D^2\right\}$$

$q(x,t)$ is the density of $X(t)$.

The variational problem (exact inference !)

Minimise variational free energy

$$\mathcal{F}_\theta(Q) = \frac{1}{2}\int_0^T \int q(x,t)\ \|g(x,t) - f(x)\|^2_D\ dx\ dt\ - \sum_i E_Q[\ln p(y_i|X(t_i))]$$

The marginal density q fulfils the Fokker - Planck equation

$$\frac{\partial q}{\partial t} = \left\{-\nabla g + \frac{1}{2}\mathrm{Tr}(\nabla\nabla^T)D\right\} q \equiv L_g q$$

Variation of Lagrange function

$$
\begin{aligned}
L \;=\; & \frac{1}{2}\int_0^T dt \int dx \, \|g(x,t) - f(x,t)\|_D^2 \; q(x,t) \\
& - \sum_i \int dx \; q(x,t_i) \ln p(y_i|x)] \\
& - \int_0^T dt \int dx \; \lambda(x,t) \; \left(\frac{\partial q}{\partial t} - L_g \, q\right)
\end{aligned}
$$

with respect to q and g, setting $\lambda(x,t) = -\ln \psi(x,t)$ yields

$$
g(x,t) = f(x) + D(x,t)\nabla \ln \psi(x,t)
$$

$$
\left\{\frac{\partial}{\partial t} + f^\top \nabla + \frac{1}{2}\mathrm{Tr}(D\nabla^\top \nabla)\right\}\psi(x,t) = 0
$$

+ jump conditions at observations.

Relation to optimal control

(B Kappen, E Todorov)

Consider SDE

$$dX(t) = (u(X(t),t) + f(X(t)))dt + D^{1/2}(X(t))\, dW(t)$$

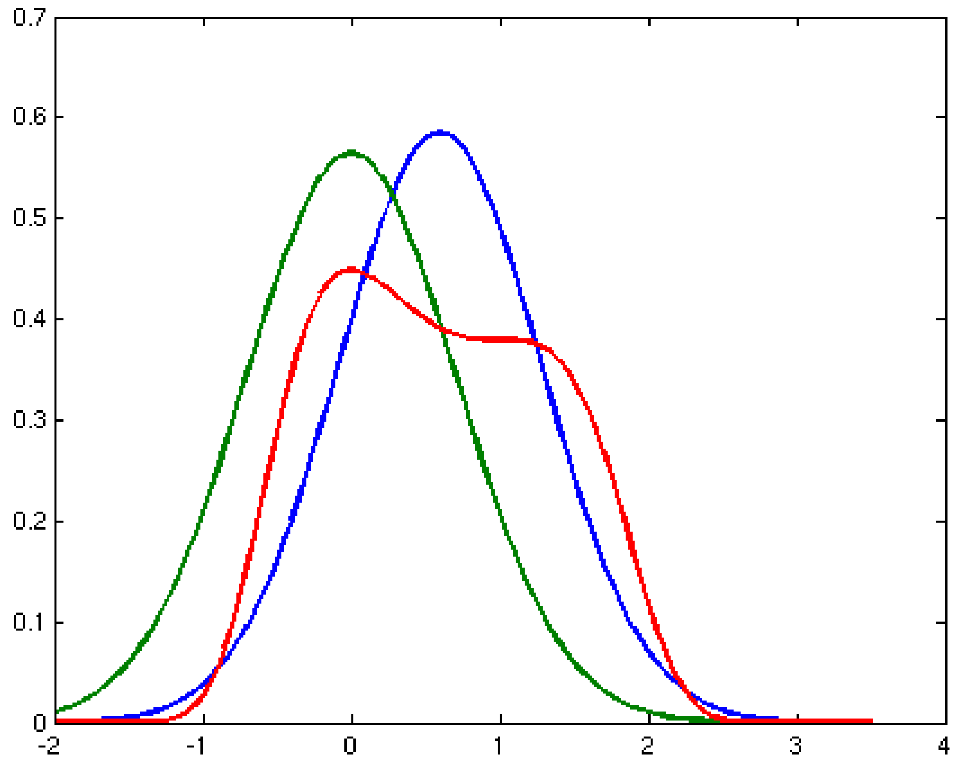
Adjust **control** $u(X_t, t)$ so that

$$V = \int_0^T E^u \left[L_t(X_t, u_t)\right]$$

is minimal, where

$$L_t(x,u) = \frac{1}{2}\,\|u\|_D^2 + U(x(t),t)$$

The Gaussian Variational Approximation



For previous applications in machine learning (see e.g. Barber & Bishop (1998), Seeger (2000), Honkela & Valpola (2005)).

The Gaussian Variational Approximation for SDEs

(Archambeau, Cornford, Opper & Shawe - Taylor, 2007)

- Diffusion must be independent of X.

- Gaussian measure over paths $X_{0:T}$ induced by linear (approximate) posterior SDE:

$$dX(t) = \{-A(t)X + b(t)\}\, dt + D^{1/2} dW$$

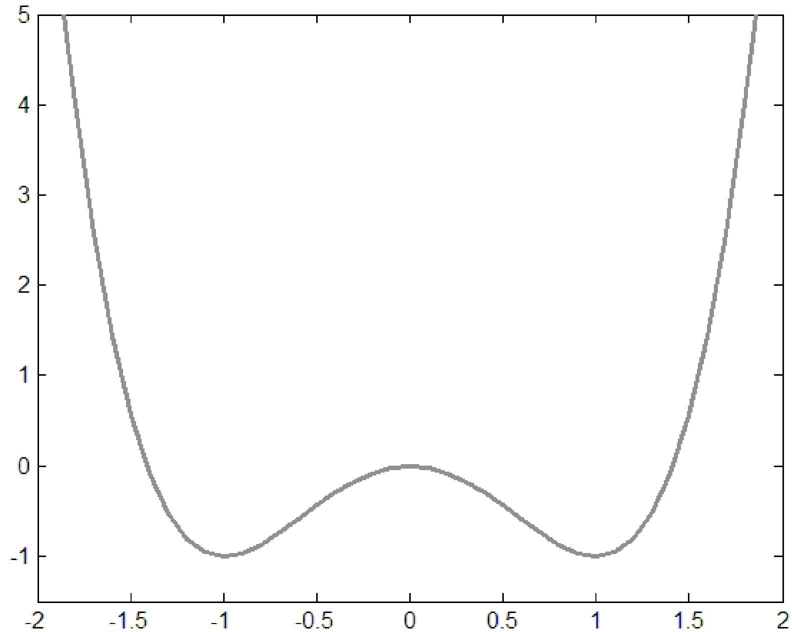
Constraints are evolution eqs. for marginal **mean** and **covariance**

$$\begin{aligned} \frac{dm}{dt} &= -Am + b(t) \\ \frac{dS}{dt} &= -AS - SA^\top + D. \end{aligned}$$

$\rightarrow$ **nonlinear ODEs** instead of PDEs !

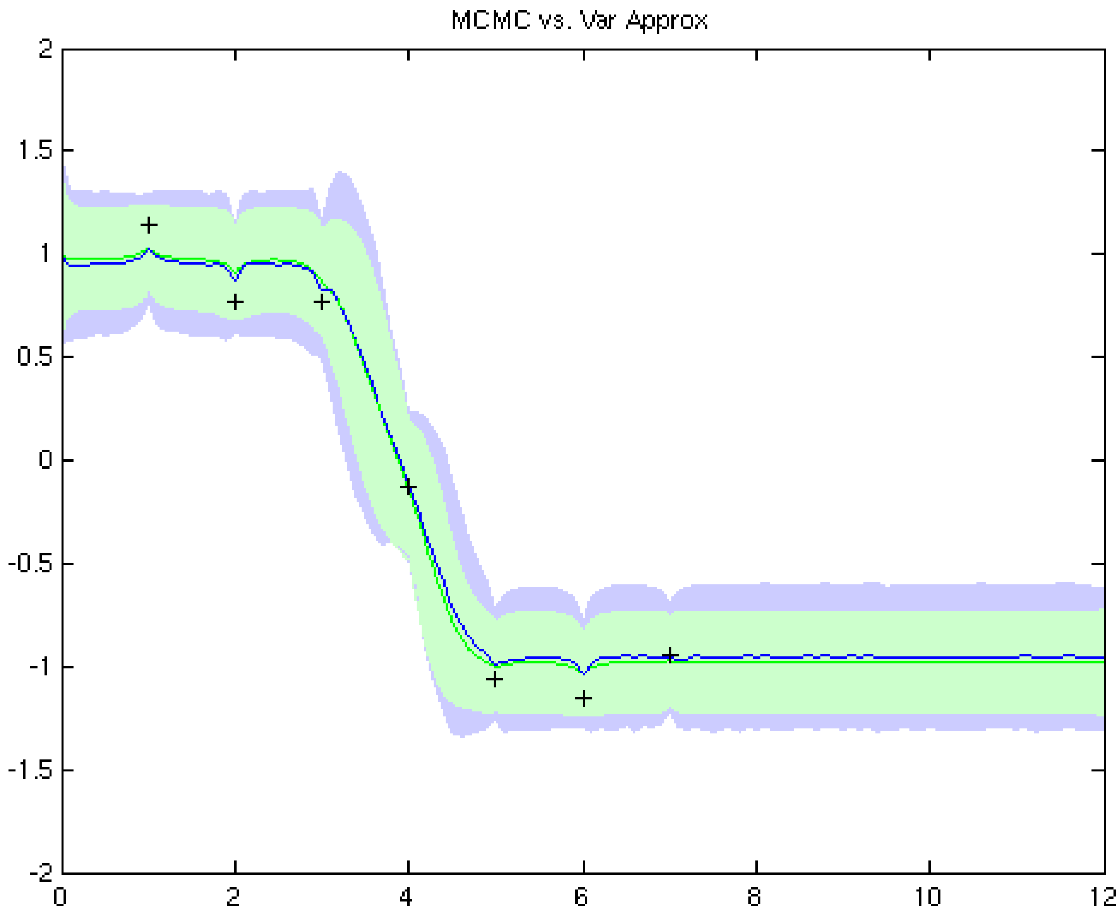
Motion in double-well potential

$$dX = 4X(1 - X^2)dt + \sigma^2 dW.$$



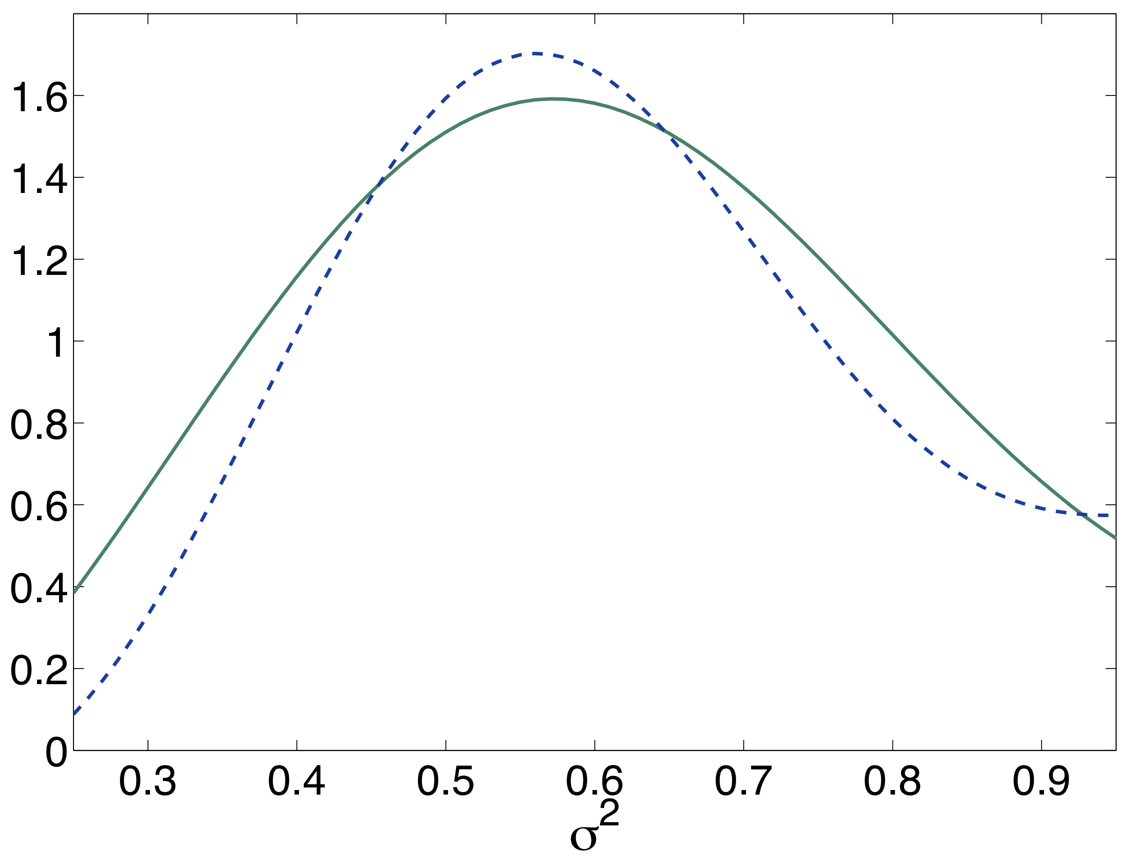
Variational result and comparison to MCMC

(Archambeau, Opper, Shen, Cornford, Shawe-Taylor, 2008)

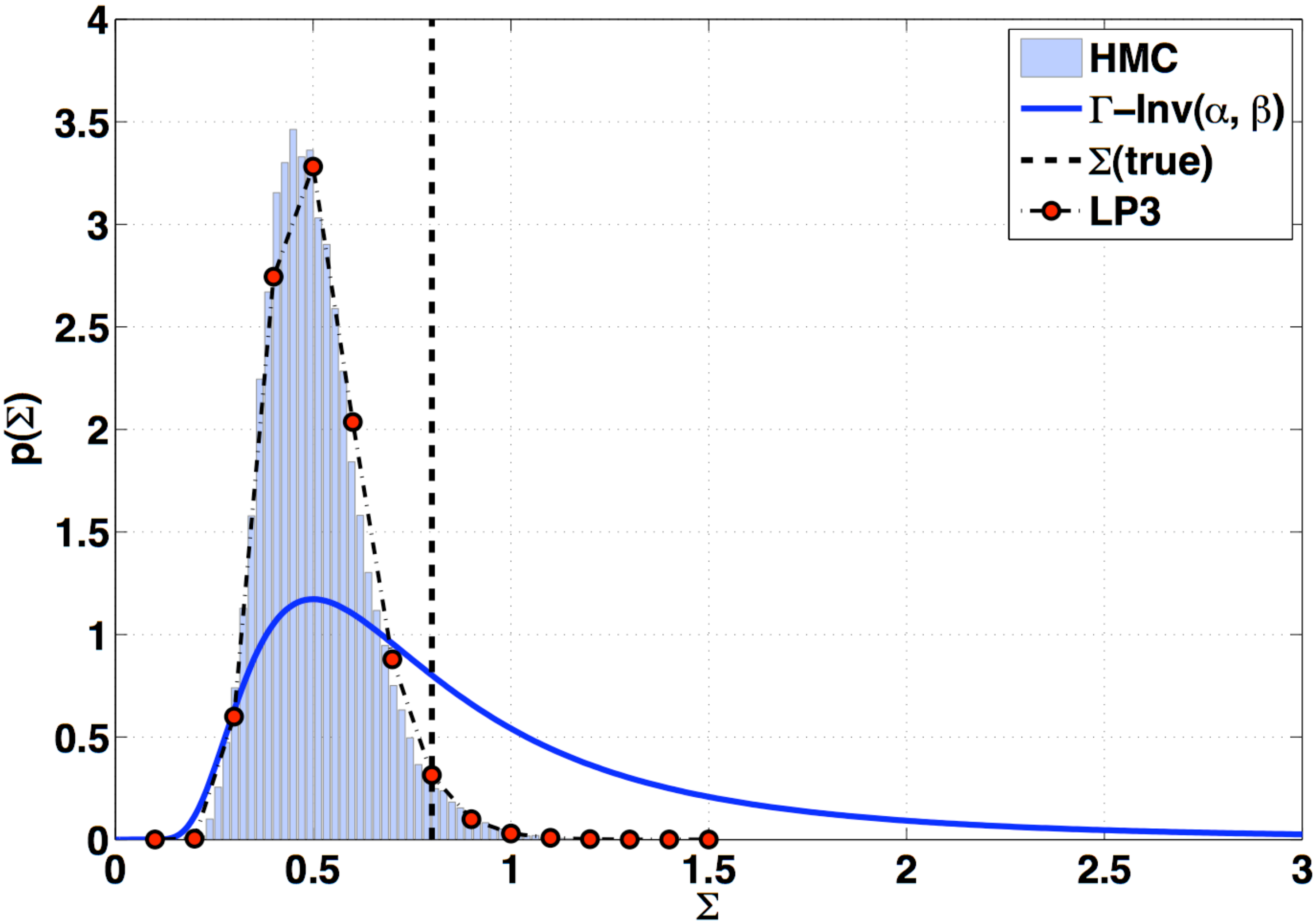


Estimation of Diffusion constant σ^2

$$q(\theta|D) \approx \frac{e^{-\mathcal{F}_\theta(Q)}\ p(\theta)}{\int e^{-\mathcal{F}_\theta(Q)}\ p(\theta)\ d\theta}$$



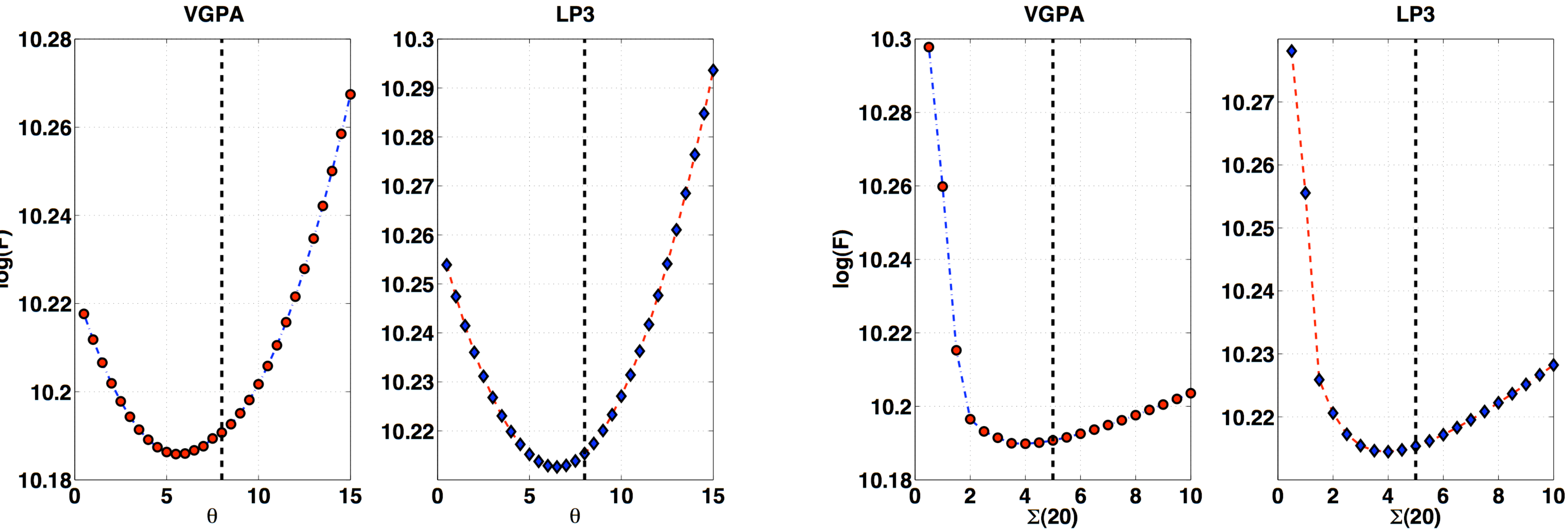
With more observations



More dimensions

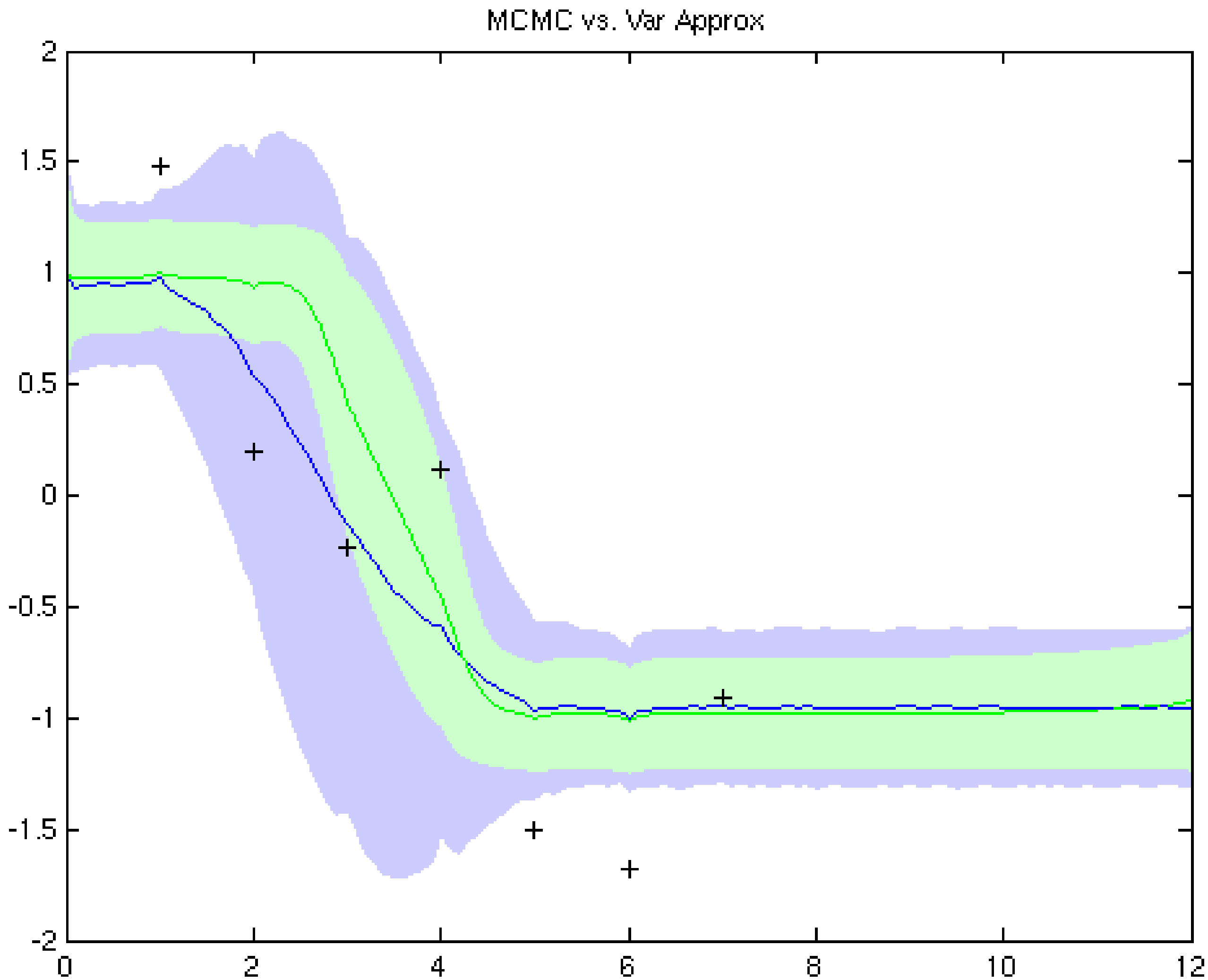
Toy weather model (Lorenz and Emanuel, 1998): $x = (x^1, \ldots, x^{40})$ with drift

$$f_i(x_t) = \left(x_t^{i+1} - x_t^{i-2}\right) x_t^{i-1} - x_t^i + \theta$$



(Vrettas, Cornford and Opper, submitted)

Large observation noise



Double well with observation noise $\sigma_o = 0.6$

KL divergence for Markov jump processes

Assume transition rates $g(X'|X,t)$ and $f(X'|X,t)$

$$KL\,[Q||P] =$$

$$\int_0^T dt\ \sum_x q(x,t) \sum_{x':x'\neq x} \left\{ g(x'|x,t) \ln \frac{g(x'|x,t)}{f(x'|x,t)} + f(x'|x,t) - g(x'|x,t) \right\}$$

Mean field approximation

Multivariate states $X = (X(1), \ldots, X(d))$

Exact inference: Linear ODEs in S^d variables

Variational approximation: Optimise in family of factorising measures, i.e. of the type

$Q[X_{0:T}] = \prod_{i=1}^{d} Q_i[X_{0:T}(i)]$

Linear ODEs in Sd variables.

(Sanguinetti & Opper, 2008)

Simple model of autoregulatory network

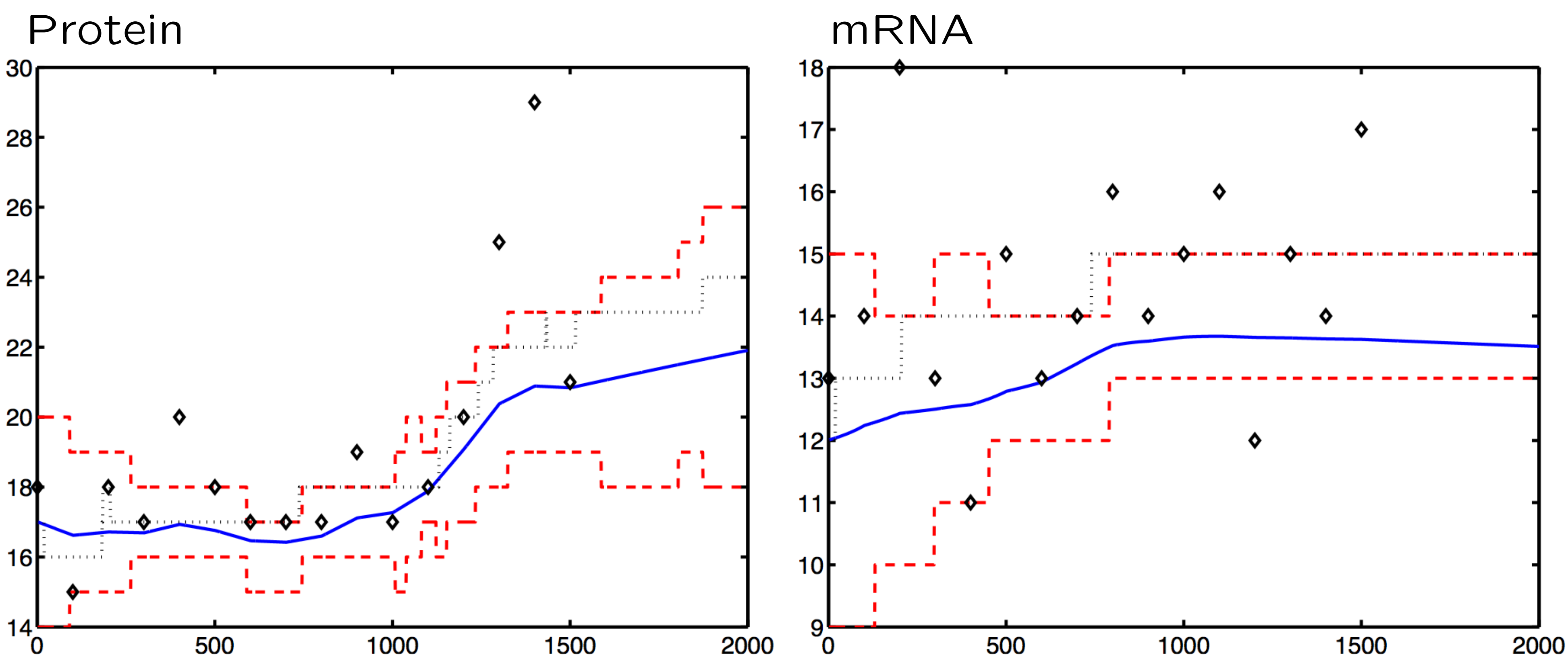
2 interacting molecules: mRNA and a Protein

Number of mRNA and Protein molecules: X_{mRNA}, X_{Prot}

$\mathbf{X} = (X_{\text{mRNA}}, X_{\text{Prot}})$

$$
\begin{array}{rcllll}
X_{\text{Prot}} & \rightarrow & X_{\text{Prot}} + 1 & : & \text{with Rate} & \gamma X_{\text{mRNA}}\,, \\
X_{\text{Prot}} & \rightarrow & X_{\text{Prot}} - 1 & : & \text{with Rate} & \delta X_{\text{Prot}}\,, \\
X_{\text{mRNA}} & \rightarrow & X_{\text{mRNA}} - 1 & : & \text{with Rate} & \beta X_{\text{mRNA}}\,, \\
X_{\text{mRNA}} & \rightarrow & X_{\text{mRNA}} + 1 & : & \text{with Rate} & \alpha(1 - \alpha_c \Theta(X_{\text{Prot}} - \theta_c))\,,
\end{array}
$$

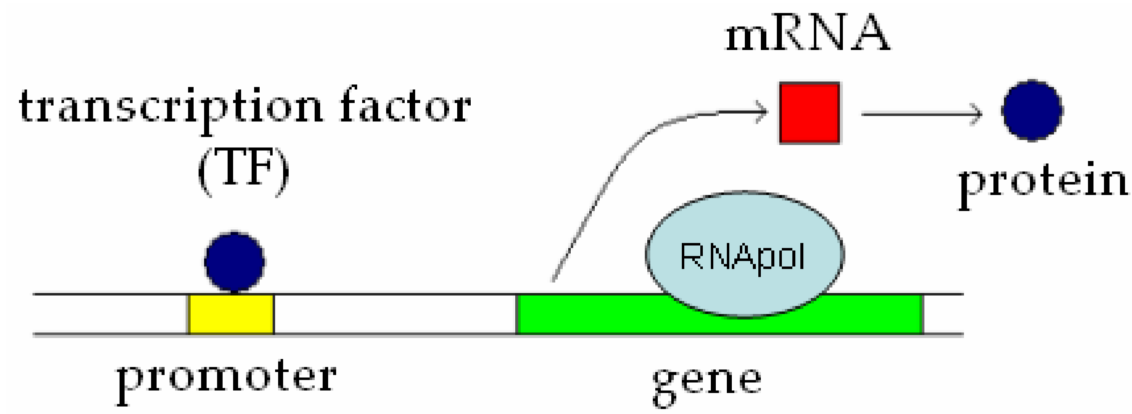
where $\Theta(x) = 1$ if $x \geq 0$ and $\Theta(x) = 0$ for $x < 0$.



(M. Opper & G. Sanguinetti, NIPS07)

Critical parameter $\theta_c = 20$ (exact) and $\theta_c = 19$ (ML estimate).

Hybrid models: Inference of transcriptional regulation



- Transcription factors regulate genes by binding to specific sites.

- Hard to measure transcription factor activity directly. Inference must be based on measurement of mRNA concentration of target genes.

- Big networks: Use Clustering. Groups of genes with similar expression profiles likely regulated by same TF.

- Big networks: Latent variable models (Factor analysis).

- Small subnetworks: more detailed dynamical model (Barenco et al:)

  $$\frac{dx_i}{dt} = -\lambda_i x_i(t) + A_i f(t)$$

  takes sensitivity and degradation into account.

  Try predictions on TF activity $f(t)$ and learn parameters using measurements of mRNA concentration of target genes:

  $$y_{ik} = x_i(t_k) + \text{noise}$$

  .

A priori knowledge of transcription factor activity $f(t)$

- Discretize and assume $f(t)$ independent for different times t (Barenco et al, 2006).

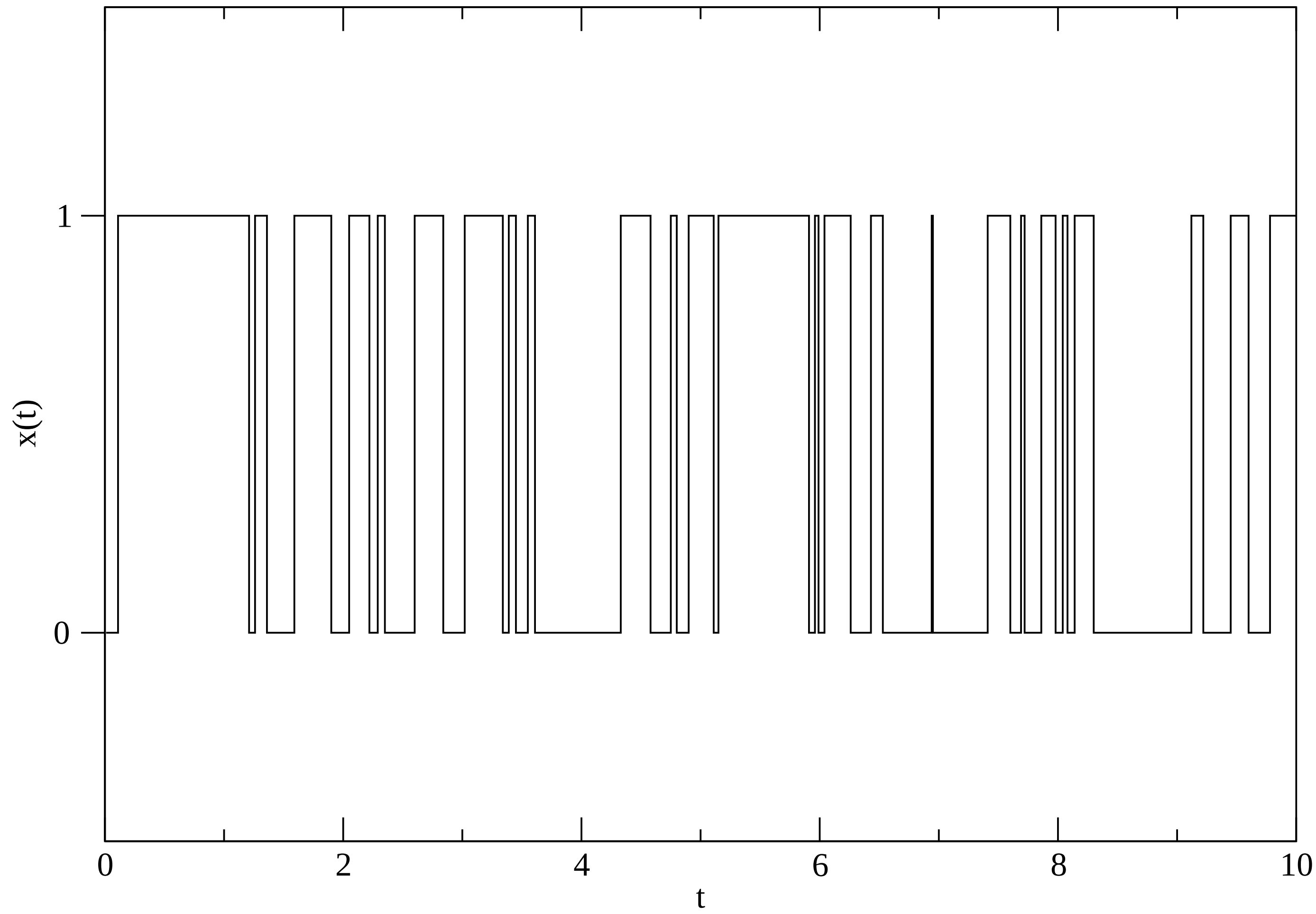
- $f(t)$ smooth & modelled by a Gaussian process (Lawrence, Sanguinetti & Rattray, 2006)

- This work: Fast switching activity $\mu(t) \in \{0, 1\}$ and $\mu \to 1 - \mu$ with rates $\gamma_\pm$. (Sanguinetti, Ruttor, Archambeau, Opper 2009)

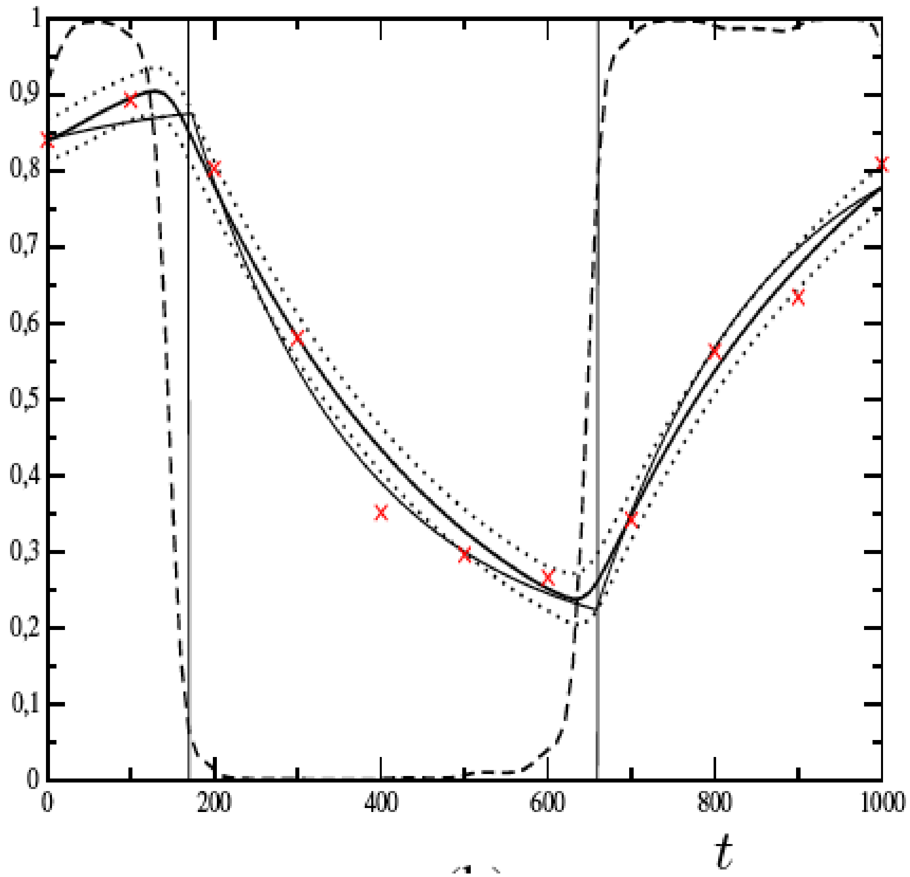
- The latter model is

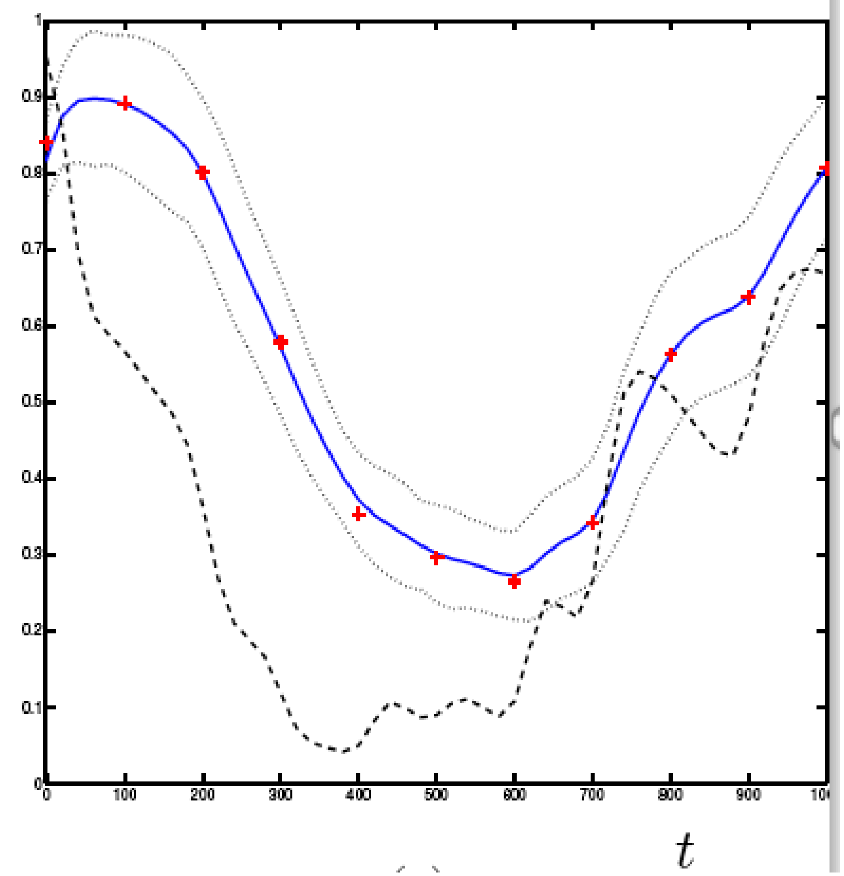
$$\frac{dx_i}{dt} = -\lambda_i x_i(t) + A_i \mu(t) + b_i$$

Telegraph Process: $X_t \in \{0, 1\}$



The role of the prior: telegraph vs GP





Variational approximation for Telegraph process

- We need the *variational free energy*

$$\mathcal{F}[Q] = KL(Q|P_{prior}) - E_Q[\ln p(D|\cdot)]$$

where Q is a Markov process over binary paths $\mu_{0:T}$.

- $D = (y_1, \ldots, y_n)$ are noisy observations of $x(t_i)$

- where $x(t)$ is a solution to

$$\frac{dx_i}{dt} = -\lambda_i x_i(t) + A_i \mu(t) + b_i$$

- Hence

$$p(D|\mu_{0:T}) \propto \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x(t_i) - y_i)^2\right]$$

- the linear ODE is solved by

$$x(t) = \exp(-\lambda t)\left\{x(0) + \int_0^t \exp(\lambda s)\left[A\mu(s) + b\right] ds\right\}$$

- Let $f_\pm$ be the prior rates and $g_\pm(t)$ the posterior rates. Then

$$\begin{aligned} KL\left[Q\|P_{prior}\right] = & \int_0^T dt\; q_1(t)\left[g_-(t)\ln\frac{g_-(t)}{f_-(t)} + f_-(t) - g_-(t)\right] + \\ & \int_0^T dt\;\left[1 - q_1(t)\right]\left[g_+(t)\ln\frac{g_+(t)}{f_+(t)} + f_+(t) - g_+(t)\right]. \end{aligned}$$

The total free energy

$$\mathcal{F}[q] = \int_0^T dt\; q_1(t) \left[g_-(t) \ln \frac{g_-(t)}{f_-(t)} + f_-(t) - g_-(t) \right] +$$

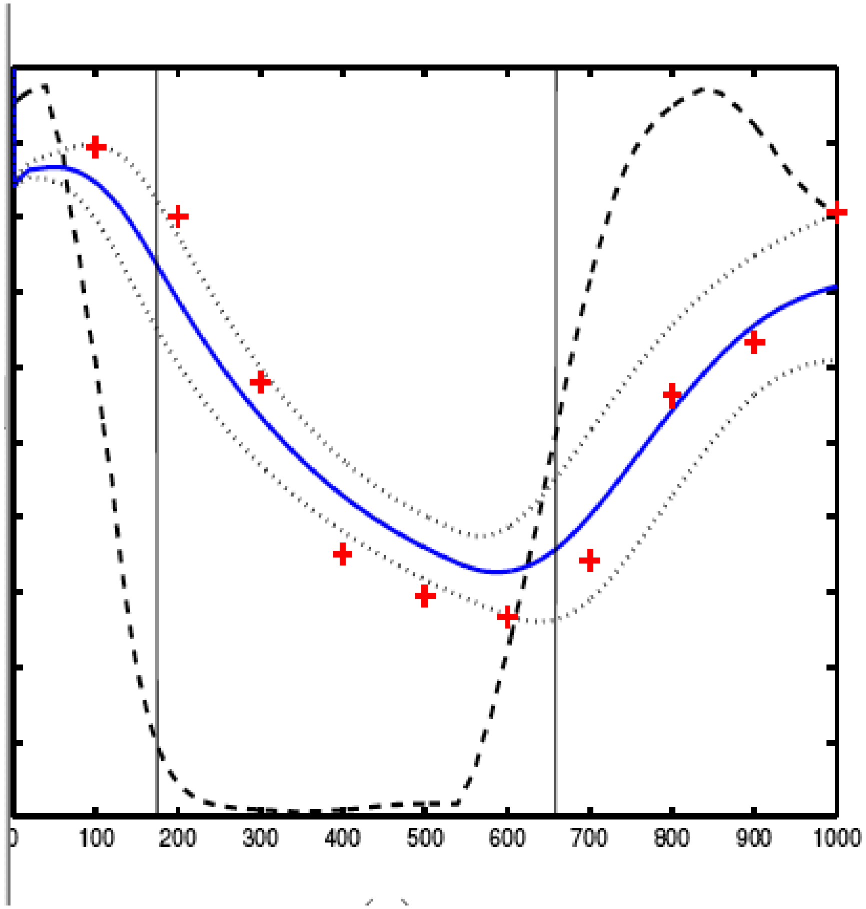
$$\int_0^T dt\; [1 - q_1(t)] \left[g_+(t) \ln \frac{g_+(t)}{f_+(t)} + f_+(t) - g_+(t) \right] +$$

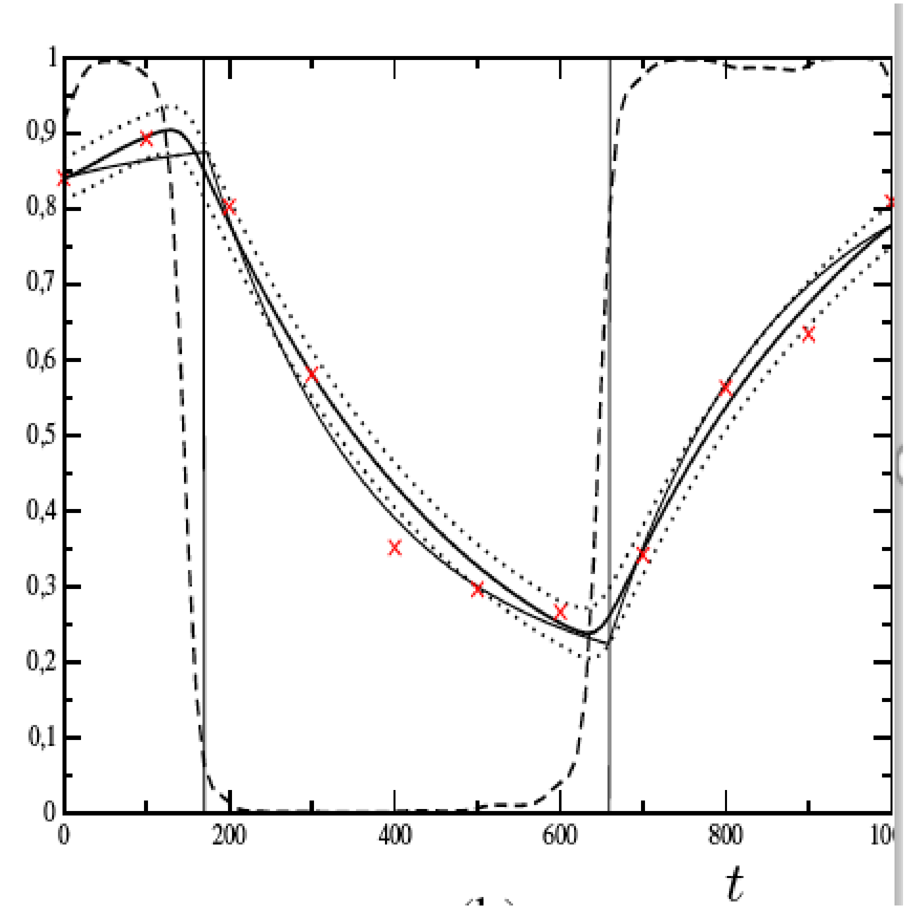
$$\frac{1}{2\sigma^2} \sum_{i=1}^{n} E_q \left(\int_0^t \exp\left(-\lambda(t-s)\left[A\mu\left(s\right) + b\right] ds - y_i \right)^2$$

has to be minimised wrt $g_\pm(t)$ and $q_1(t)$ under the condition that

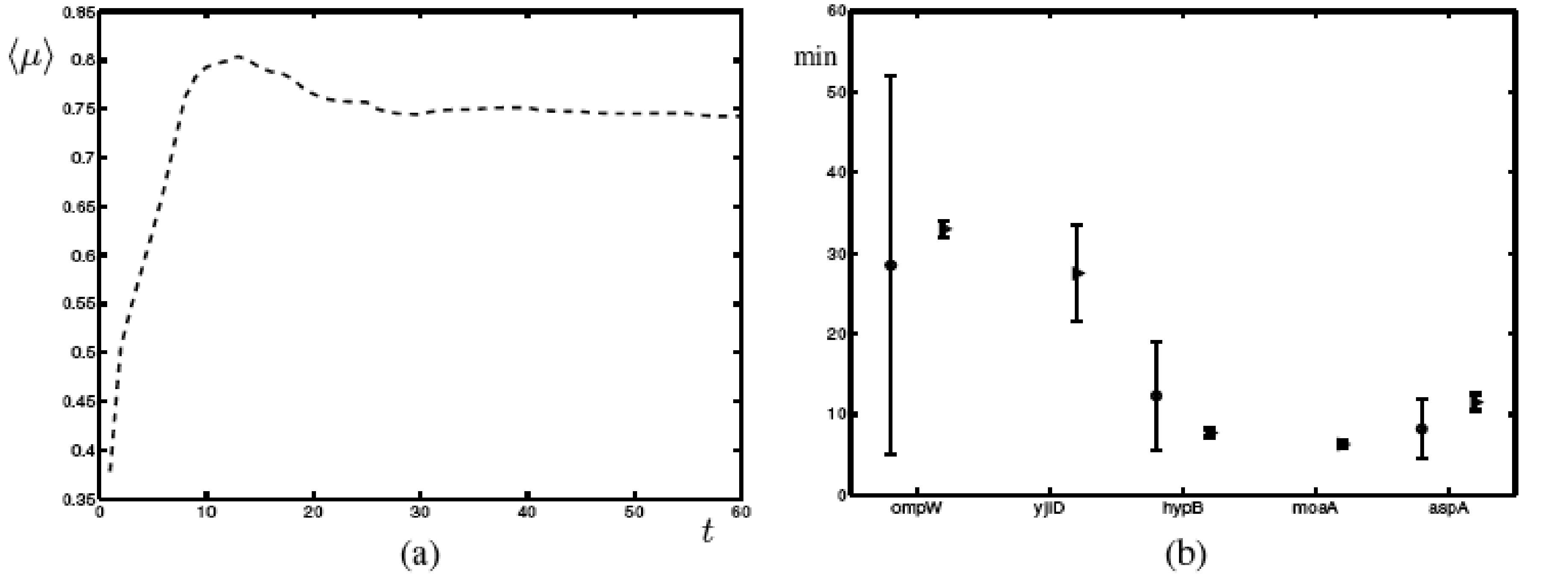
$$\frac{dq_1\left(t\right)}{dt} = -(g_+ + g_-)q_1\left(t\right) + g_+\left(t\right)$$

Approximate vs exact inference





Predictions for *E. coli* deprived of oxygen at $t = 0$

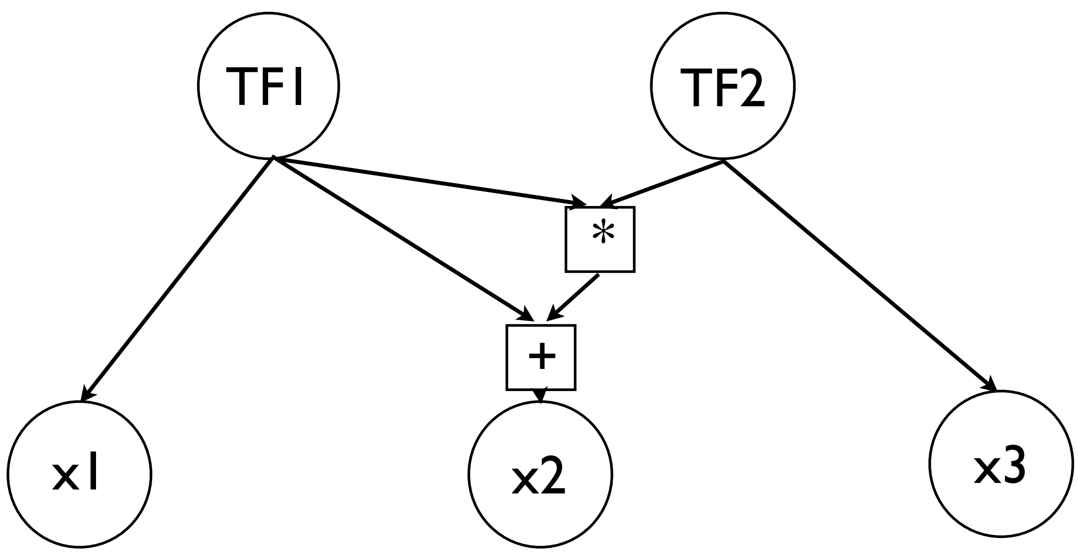


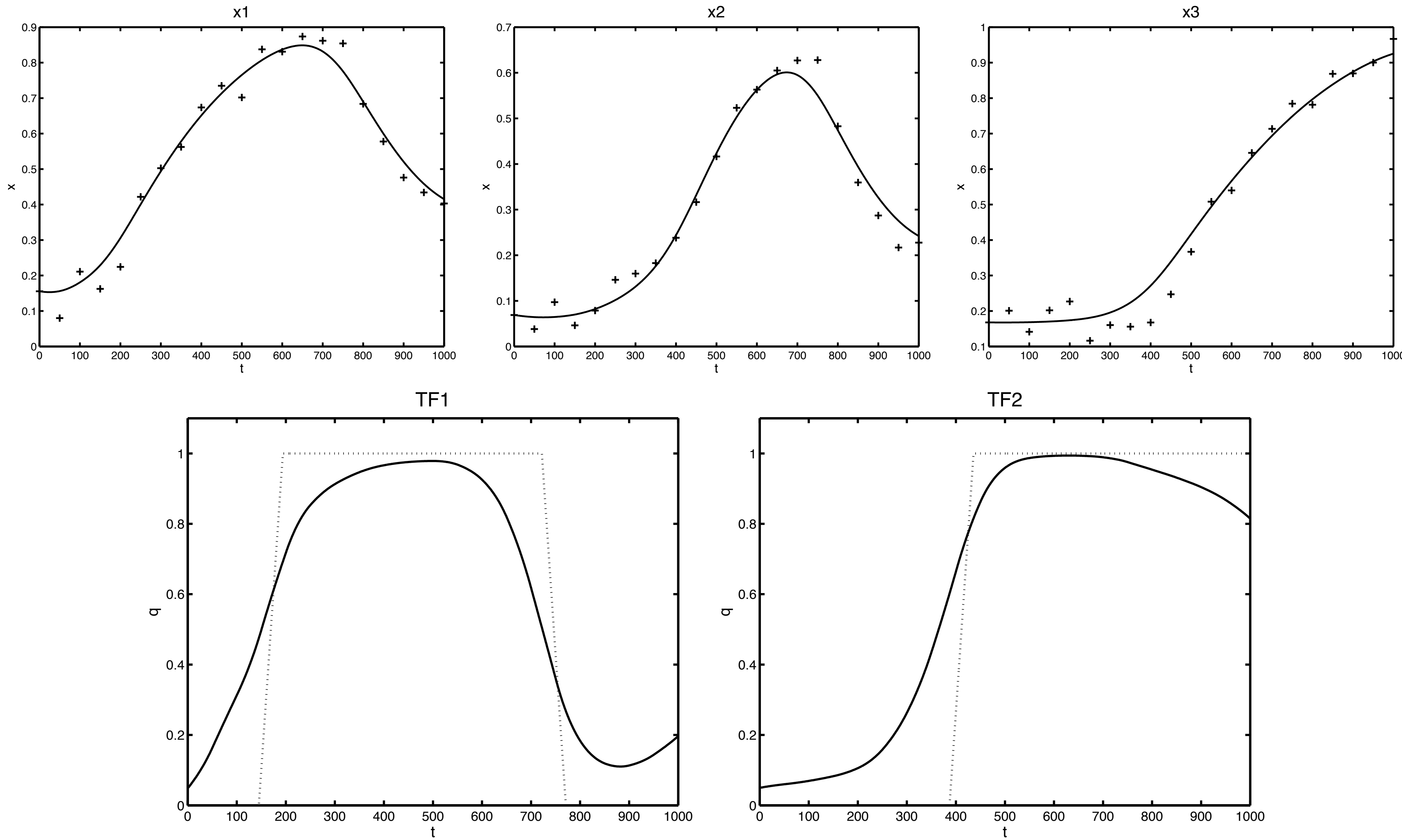
based on measurements on 5 target genes at 4 time points.

(Sanguinetti et al, 2009)

Multiple (2) transcription factors (Opper & Sanguinetti, 2010):

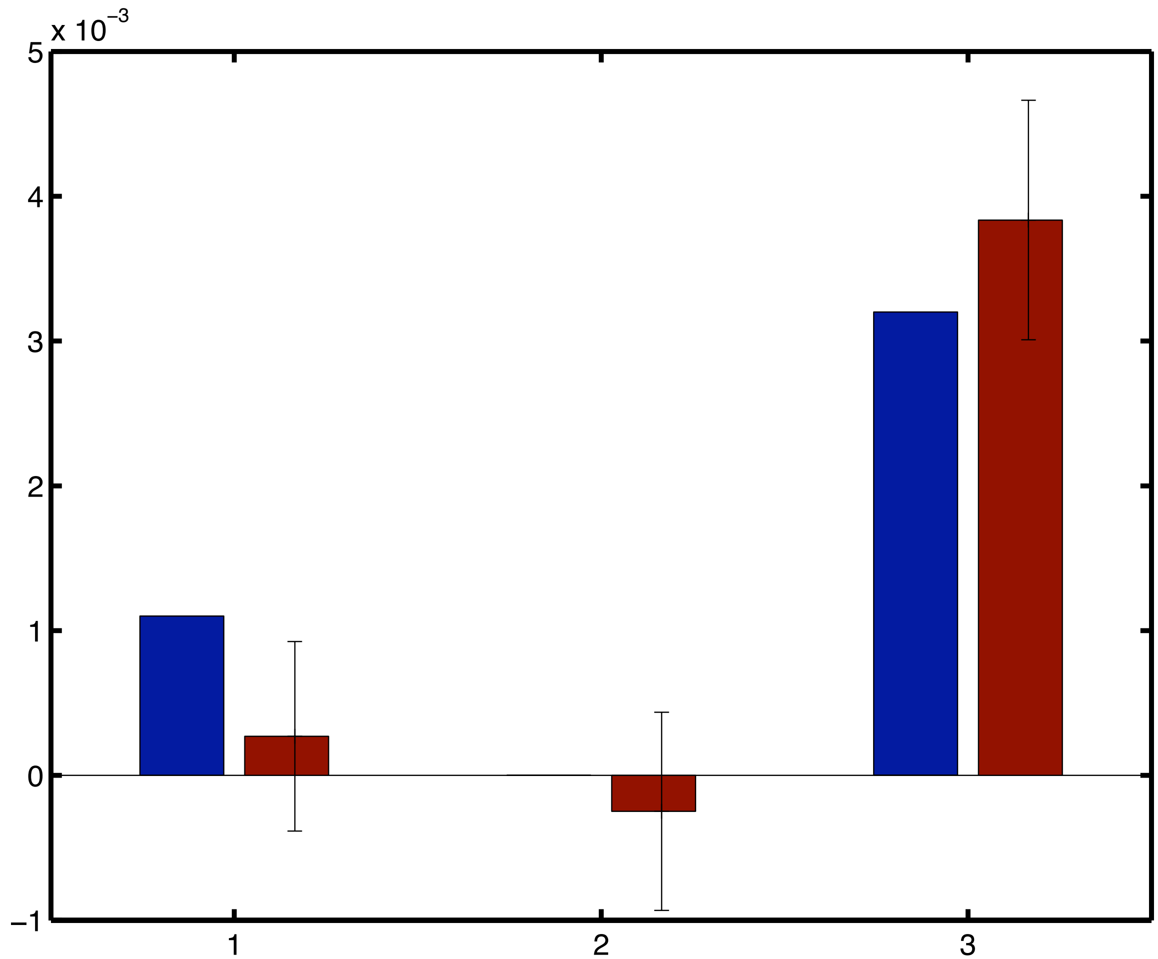
$$\frac{dx_i}{dt} = -\lambda_i x_i(t) + A^i \mu_1(t) + A^i \mu_1(t) + A^i_{12} \mu_1(t)\mu_2(t) + b_i$$



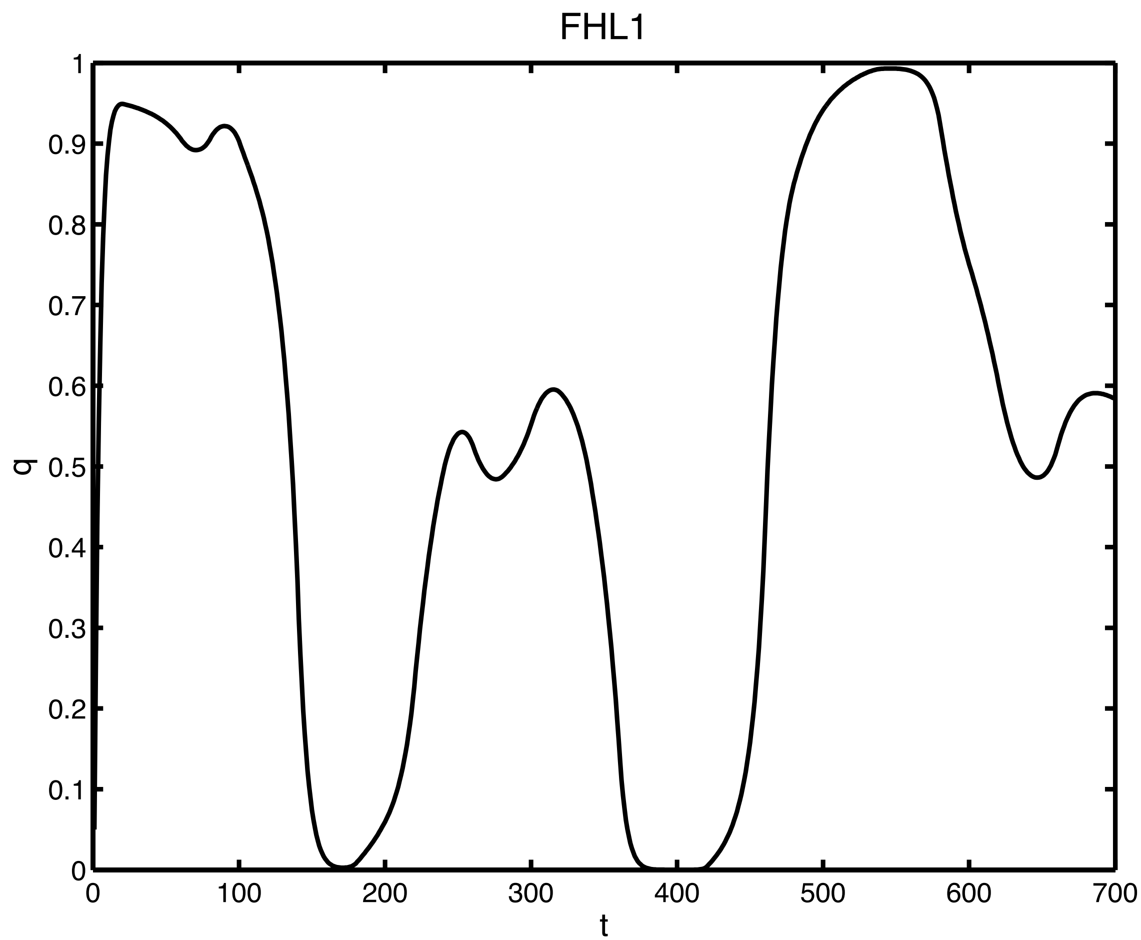


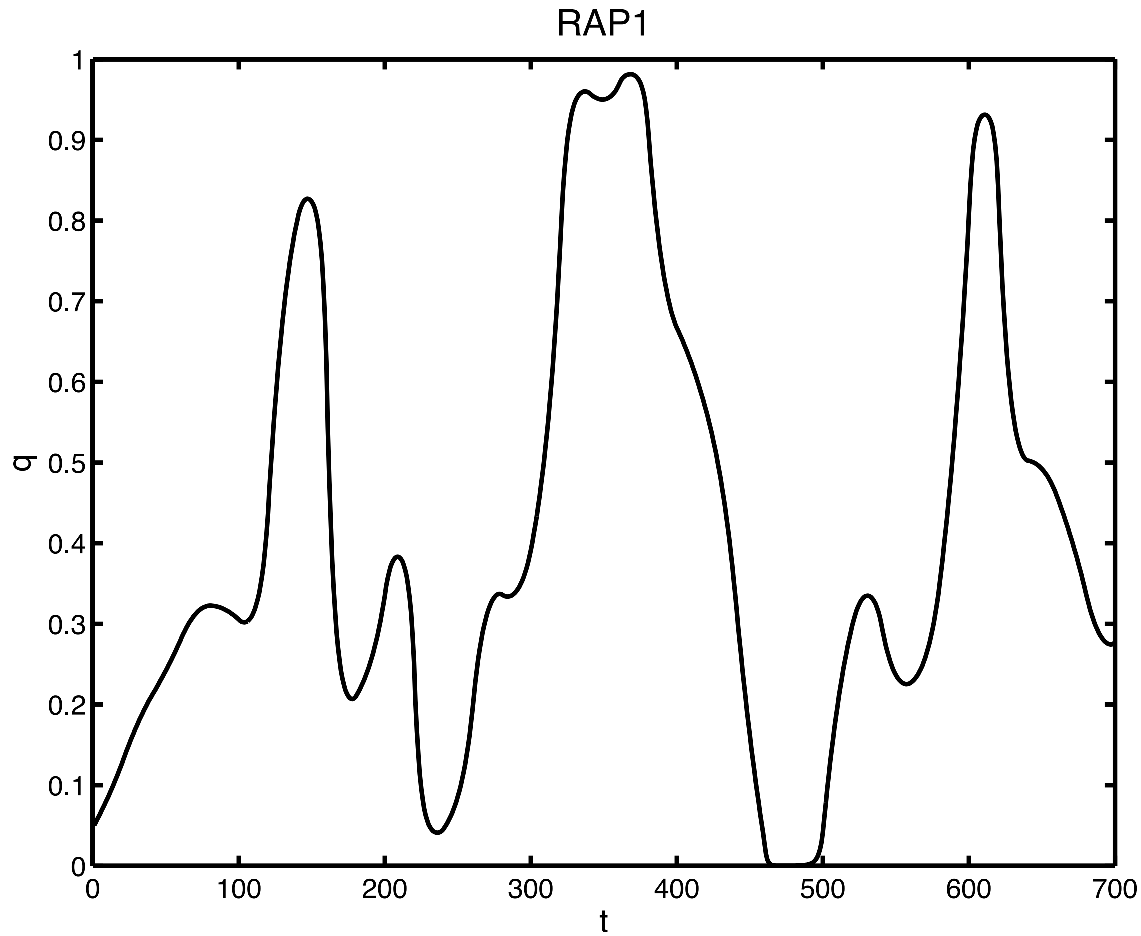
x1
x2
x3
TF1
TF2
x
t
q

Parameter inference for A_1^2 (blue), A_2^2 (green) and A_{12}^2 (red).

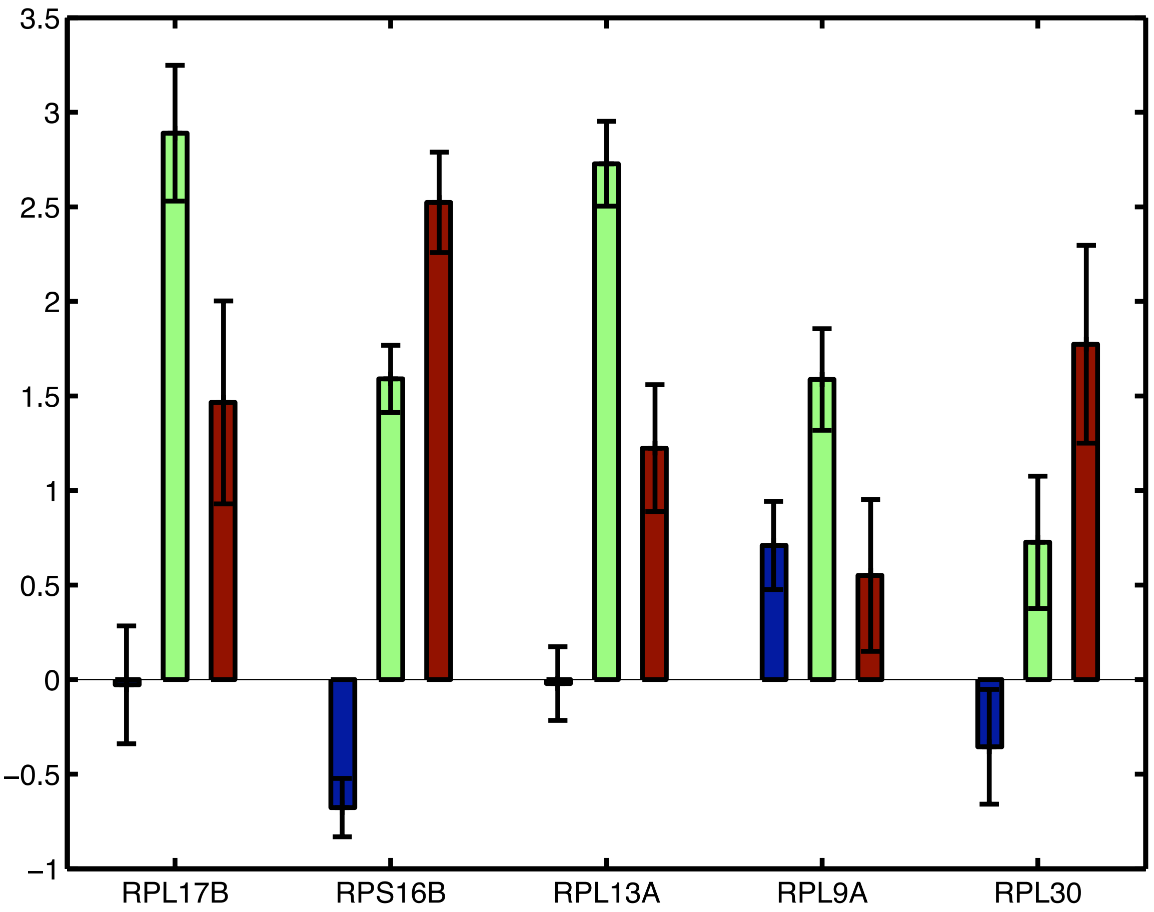


Prediction of activity of transcription factors **FHL1** and **RAP1** (Microarray data from yeast metabolic cycle)




RAP1
q
t
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
100
200
300
400
500
600
700

A_1^i, A_2^i and A_{12}^i for 5 target genes.



Comments

- Method scales linearly in number of genes, number of observations & number of TFs.

- We can calculate upper and lower bounds on free energy.

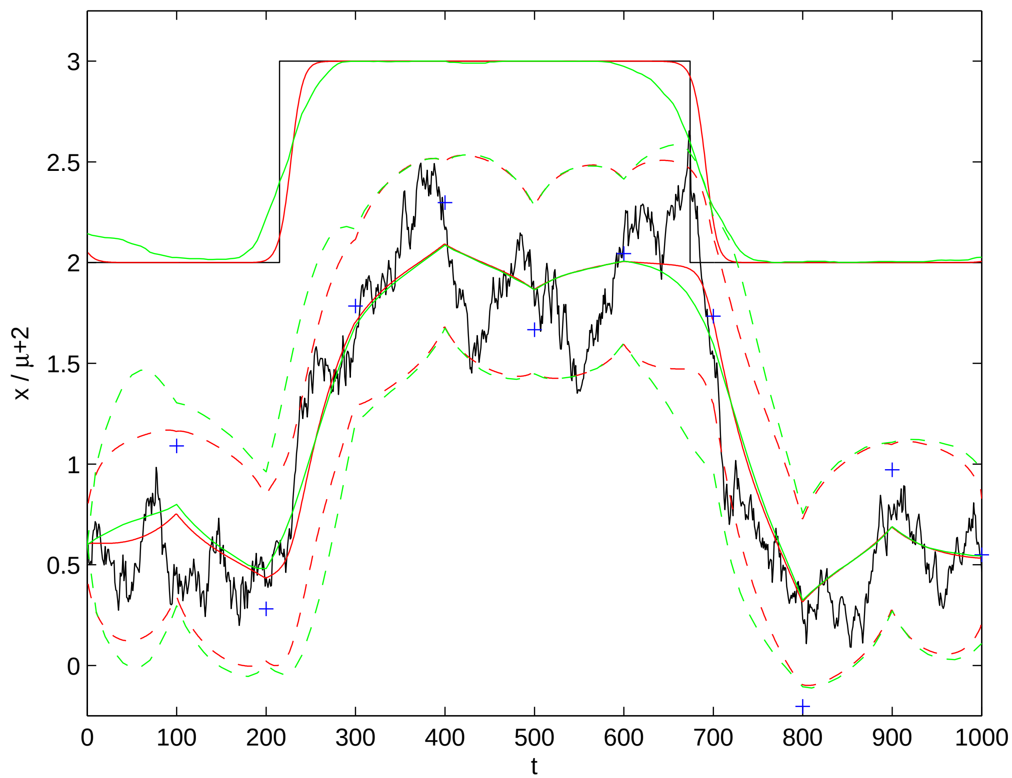
- Can be generalised to more complex architectures (forward loops, autoregulatory networks).

- Can be generalised to additional white noise input (linear SDEs).

Inclusion of white noise (for single cell fluorescence measurements)

$$dx_i = (-\lambda_i x_i(t) + A_i \mu(t) + b_i)\, dt + \sigma dW$$

Toy data



Opper, Ruttor and Sanguinetti, NIPS2010

Present & Future work

- Variational path densities as proposal for MCMC

- Perturbative corrections, PAC Bayesian confidence bounds

- Variational approaches to multiplicative noise ?

- Realistic higher dimensional systems: Need good suboptimal variational parametrisations.