Nonlinear Response in Gaussian Process Models of Transcriptional Regulation

Neil D. Lawrence
work with Magnus Rattray (co-PI), Pei Gao, Michalis Titsias

School of Computer Science, University of Manchester, U.K.
Seminar at TIGEM, Naples

29th October 2009

Outline

Nonlinear Response Models

Consider the model of transcription,

$$\frac{\mathrm{d}x_j(t)}{\mathrm{d}t} = B_j + S_j g(f(t)) - D_j x_j(t),$$

where $g(\cdot)$ is a non-linear function. The differential equation can still be solved,

$$x_j(t) = \frac{B_j}{D_j} + S_j \int_0^t e^{-D_j(t-u)} g_j(f(u))\,\mathrm{d}u$$

Outline

Radiation Damage in the Cell

- Radiation can damages molecules including DNA.
- Most DNA damage is quickly repaired—single strand breaks, backbone break.
- Double strand breaks are more serious—a complete disconnect along the chromosome.
- Cell cycle stages:
  - G_1: Cell is not dividing.
  - G_2: Cell is preparing for meitosis, chromosomes have divided.
  - S: Cell is undergoing meitosis (DNA synthesis).
- Main problem is in G1. In G2 there are two copies of the chromosome. In G1 only one copy.

p53 "Guardian of the Cell"

- Responsible for Repairing DNA damage
- Activates DNA Repair proteins
- Pauses the Cell Cycle (prevents replication of damage DNA)
- Initiates *apoptosis* (cell death) in the case where damage can't be repaired.
- Large scale feeback loop with NF-κB.

p53 DNA Damage Repair

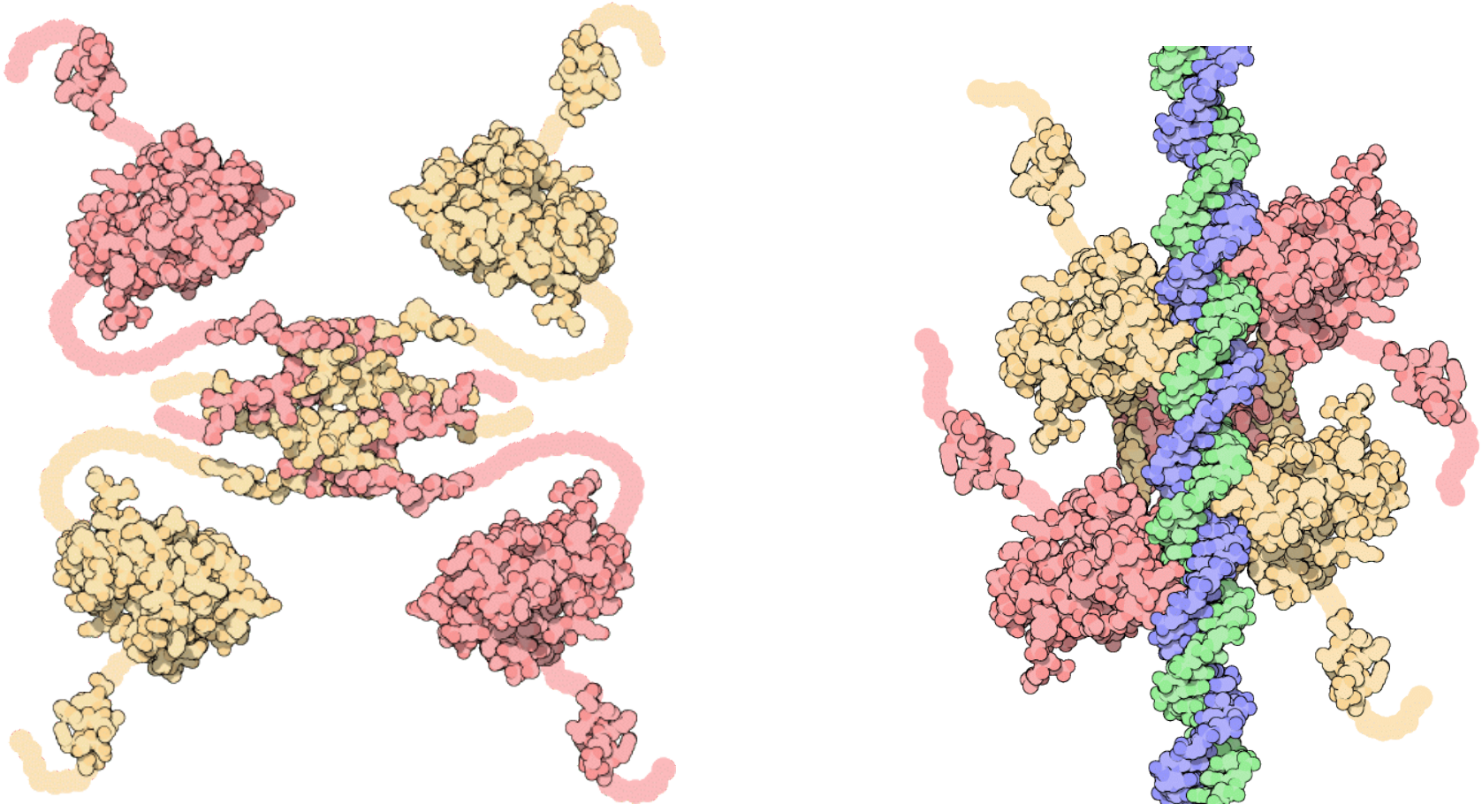


Figure: p53. *Left* unbound, *Right* bound to DNA. Images by David S. Goodsell from `http://www.rcsb.org/` (see the"Molecule of the Month" feature).

p53

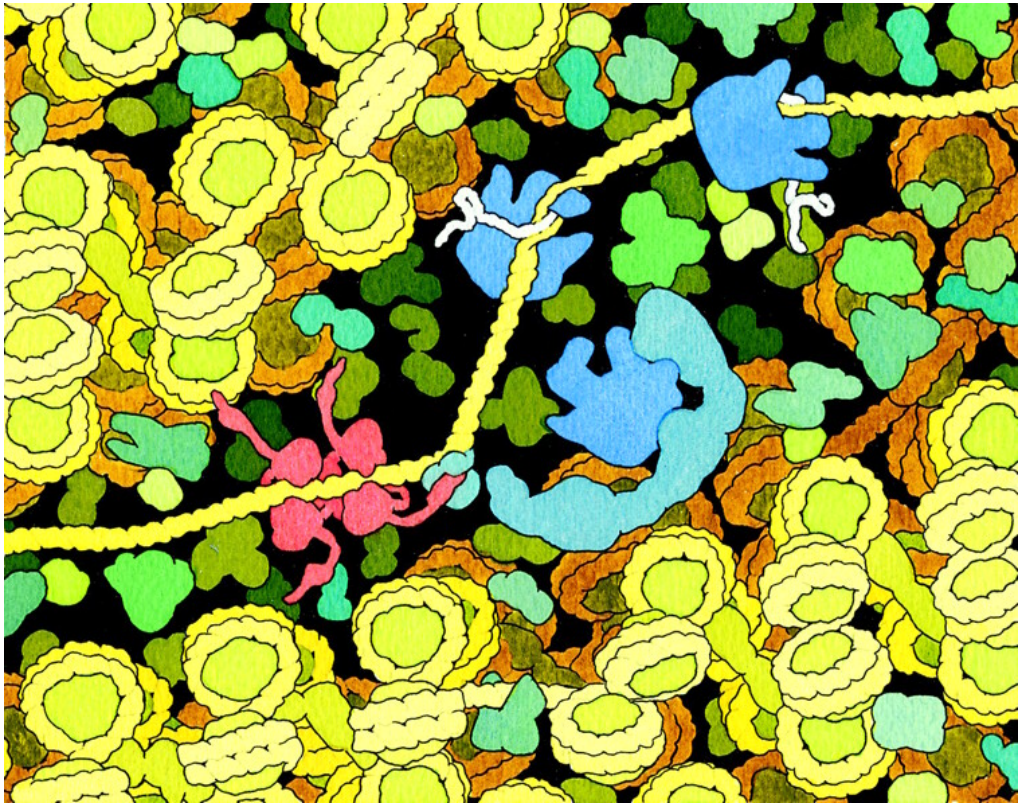


Figure: Repair of DNA damage by p53. Image from Goodsell (1999).

Some p53 Targets

DDB2 DNA Damage Specific DNA Binding Protein 2. (also governed by C/ EBP-beta, E2F1, E2F3,...).

p21 Cycline-dependent kinase inhibitor 1A (CDKN1A). A regulator of cell cycle progression. (also goverened by SREBP-1a, Sp1, Sp3,...).

hPA26/SESN1 sestrin 1 Cell Cycle arrest.

BIK BCL2-interacting killer. Induces cell death (apoptosis)

TNFRSF10b tumor necrosis factor receptor superfamily, member 10b. A transducer of apoptosis signals.

Modelling Assumption

- Assume p53 affects targets as a single input module network motif (SIM).

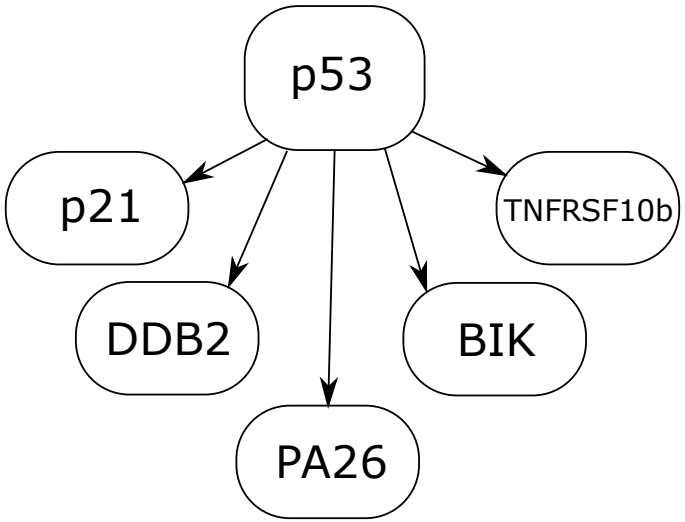


Figure: p53 SIM network motif as modelled by Barenco et al. 2006.

Mathematical Model

- Differential equation model of system.

$$\frac{\mathrm{d}x_j(t)}{\mathrm{d}t} = B_j + S_j f(t) - D_j x_j(t)$$

rate of mRNA transcription, baseline transcription rate, transcription factor activity, mRNA decay

Mathematical Model

- Differential equation model of system.

$$\frac{\mathrm{d}x_j(t)}{\mathrm{d}t} = B_j + S_j g(f(t)) - D_j x_j(t)$$

  rate of mRNA transcription, baseline transcription rate, transcription factor activity, mRNA decay

- $g(\cdot)$ is a nonlinear response to TF activity.

Respose to p53 ...

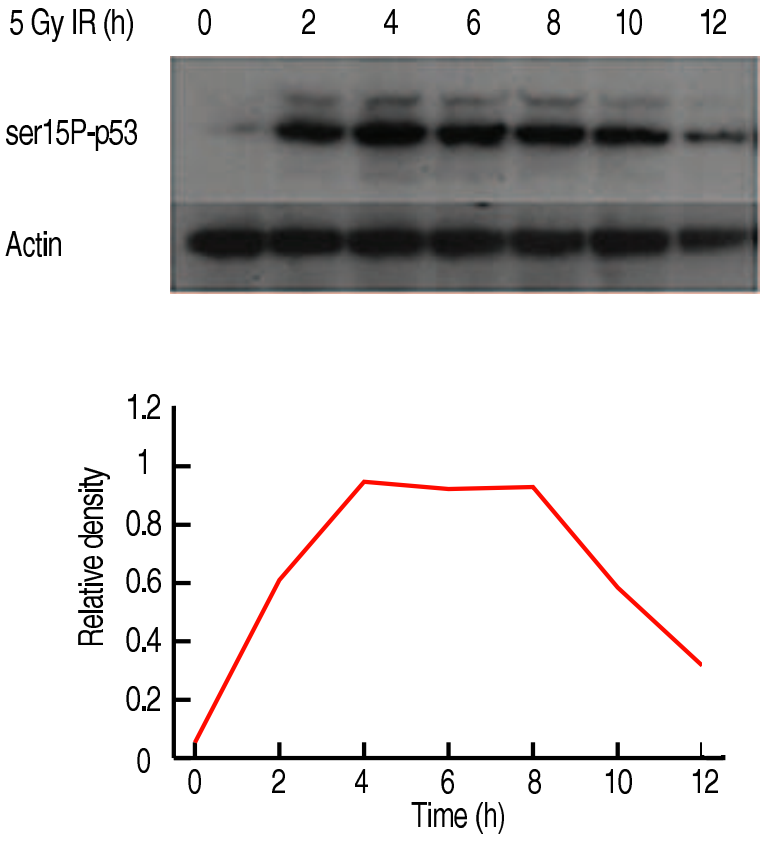


Figure: Results from Barenco et al. (2006). Activity profile of p53 was measured by Western blot to determine the levels of ser-15 phosphorylated p53 (ser15P-p53).

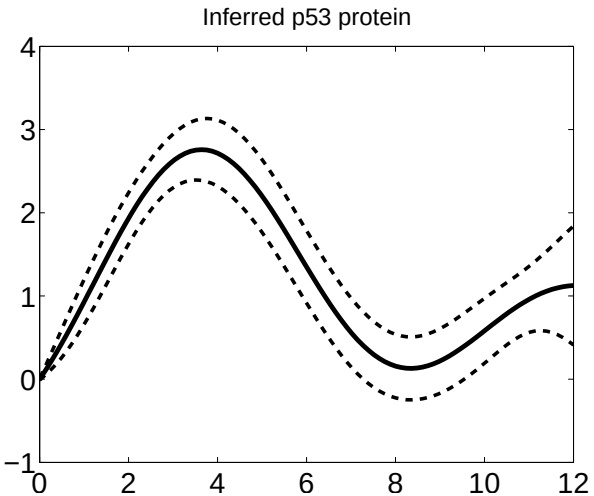
Michaelis-Menten Kinetics

Pei Gao

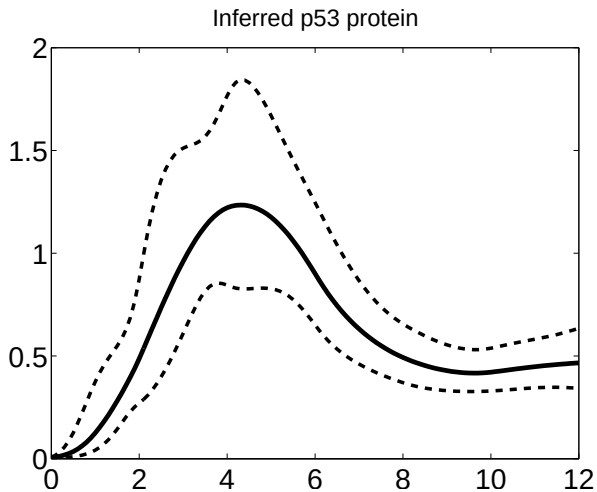
- The Michaelis-Menten activation model uses the following non-linearity

$$g_j\left(f\left(t\right)\right) = \frac{e^{f(t)}}{\gamma_j + e^{f(t)}},$$

where we are using a GP $f\left(t\right)$ to model the log of the TF activity.



(a) Linear Response



(b) Laplace Approximation Nonlinear

Valdiation of Laplace Approximation

Michalis Titsias

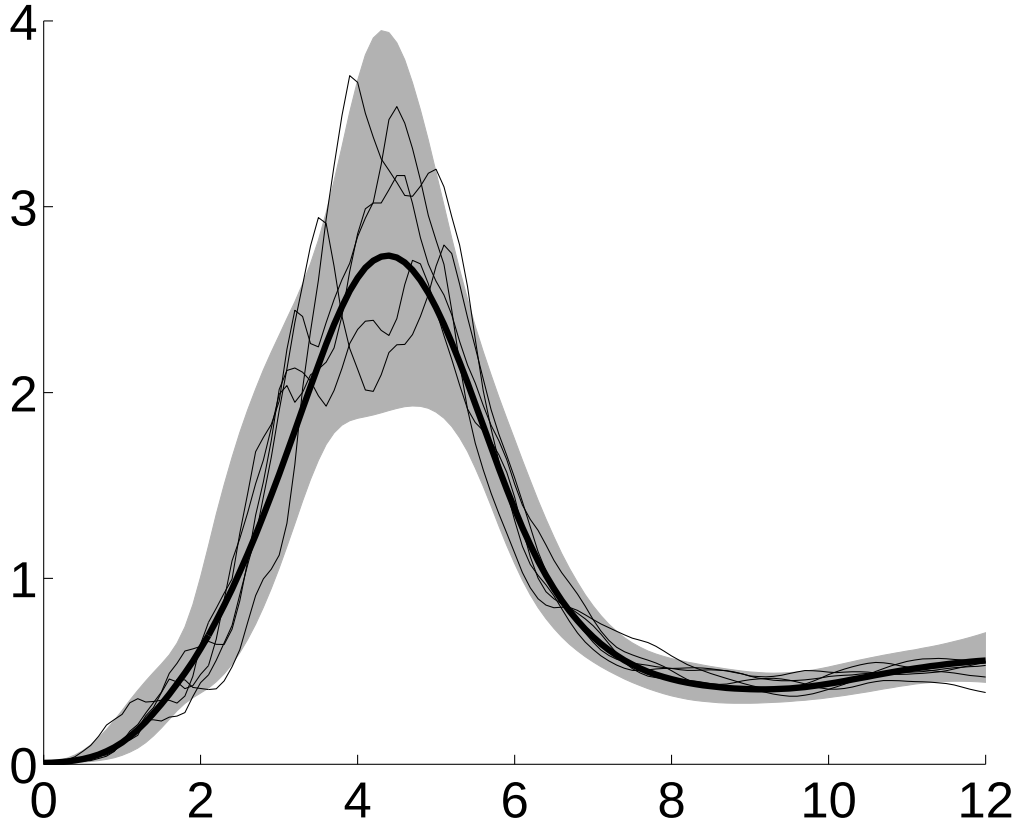


Figure: Laplace approximation error bars along with samples from the true posterior distribution.

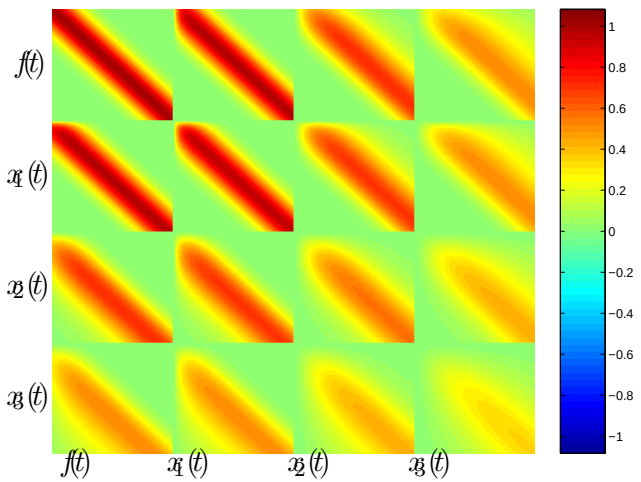
Covariance for Transcription Model

RBF covariance function for $f(t)$

$$x_i(t) = \frac{B_i}{D_i} + S_i \exp(-D_i t) \int_0^t f(u) \exp(D_i u) \, \mathrm{d}u.$$

- Joint distribution for $x_1(t)$, $x_2(t)$, $x_3(t)$, and $f(t)$.
- Here:

D_1	S_1	D_2	S_2	D_3	S_3
5	5	1	1	0.5	0.5



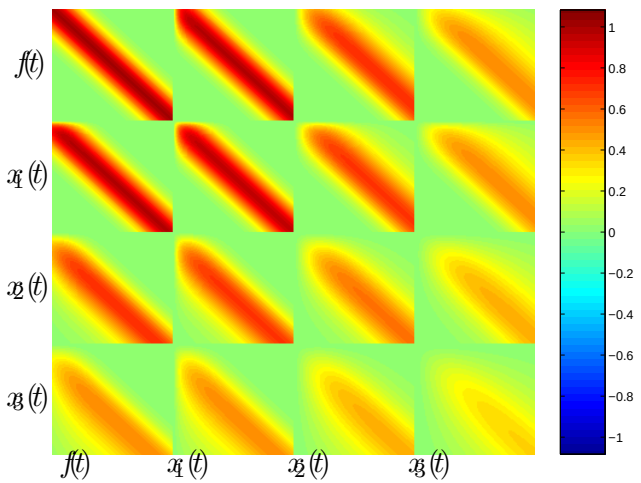
Covariance for Transcription Model

RBF covariance function for $f(t)$

$$x = b/d + \sum_i \mathbf{e}_i^\top \mathbf{f} \quad \mathbf{f} \sim \mathcal{N}(\mathbf{0}, \Sigma_i) \rightarrow x \sim \mathcal{N}\left(b/d, \sum_i \mathbf{e}_i^\top \Sigma_i \mathbf{e}_i\right)$$

- Joint distribution for $x_1(t)$, $x_2(t)$, $x_3(t)$, and $f(t)$.
- Here:

D_1	S_1	D_2	S_2	D_3	S_3
5	5	1	1	0.5	0.5



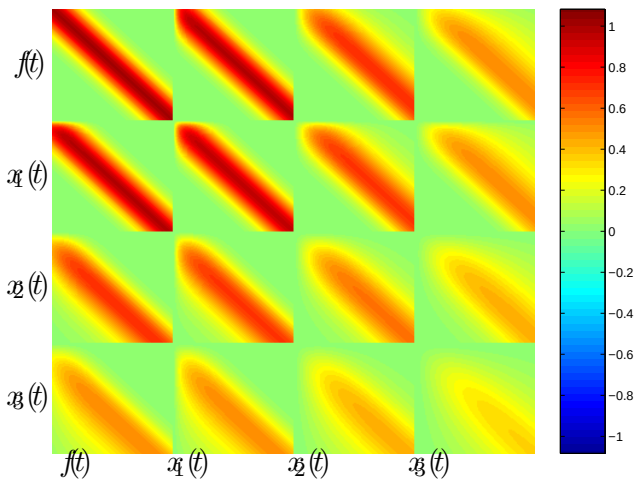
Covariance for Transcription Model

RBF covariance function for $f(t)$

$$x_i(t) = \frac{B_i}{D_i} + S_i \exp(-D_i t) \int_0^t g(f(u)) \exp(D_i u) \, \mathrm{d}u.$$

- Joint distribution for $x_1(t)$, $x_2(t)$, $x_3(t)$, and $f(t)$.
- Here:

D_1	S_1	D_2	S_2	D_3	S_3
5	5	1	1	0.5	0.5



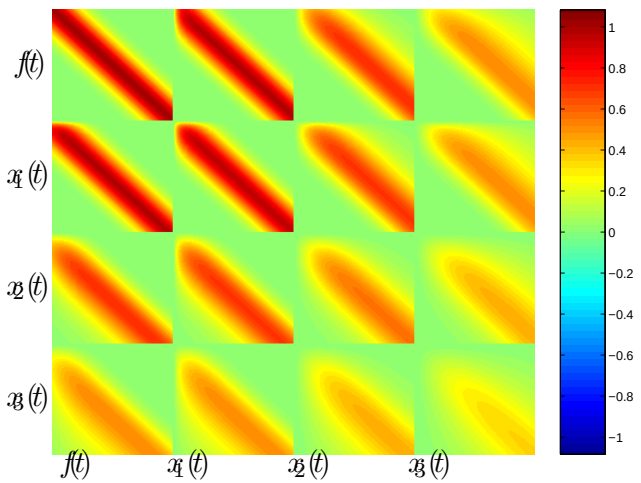
Covariance for Transcription Model

RBF covariance function for $f(t)$

$$x = b/d + \sum_i \mathbf{e}_i^\top g(\mathbf{f}) \quad \mathbf{f} \sim \mathcal{N}(\mathbf{0}, \Sigma_i) \rightarrow x \sim \mathcal{N}\left(b/d, \sum_i \mathbf{e}_i^\top \Sigma_i \mathbf{e}_i\right)$$

- Joint distribution for $x_1(t)$, $x_2(t)$, $x_3(t)$, and $f(t)$.
- Here:

D_1	S_1	D_2	S_2	D_3	S_3
5	5	1	1	0.5	0.5



Outline

MAP-Laplace Approximation

Laplace's method: approximate posterior mode as Gaussian

$$p(\mathbf{f} \mid \mathbf{x}) = N\left(\hat{\mathbf{f}}, \mathbf{A}^{-1}\right) \propto \exp\left(-\frac{1}{2}\left(\mathbf{f}-\hat{\mathbf{f}}\right)^{T}\mathbf{A}\left(\mathbf{f}-\hat{\mathbf{f}}\right)\right)$$

where $\hat{\mathbf{f}} = \operatorname{argmax} p(\mathbf{f} \mid \mathbf{x})$ and $\mathbf{A} = -\nabla\nabla \log p(\mathbf{f} \mid \mathbf{y}) \mid_{\mathbf{f}=\hat{\mathbf{f}}}$ is the Hessian of the negative posterior at that point. To obtain $\hat{\mathbf{f}}$ and $\mathbf{A}$,

we define the following function $\psi(\mathbf{f})$ as:

$$\log p(\mathbf{f}|\mathbf{x}) \propto \psi(\mathbf{f}) = \log p(\mathbf{x} \mid \mathbf{f}) + \log p(\mathbf{f})$$

MAP-Laplace Approximation

Assigning a GP prior distribution to $f(t)$, it then follows that

$$\log p(\mathbf{f}) = -\frac{1}{2}\mathbf{f}^T\mathbf{K}^{-1}\mathbf{f} - \frac{1}{2}\log|\mathbf{K}| - \frac{n}{2}\log 2\pi$$

where $\mathbf{K}$ is the covariance matrix of $f(t)$. Hence,

$$\nabla\psi(\mathbf{f}) = \nabla \log p(\mathbf{x}|\mathbf{f}) - \mathbf{K}^{-1}\mathbf{f}$$
$$\nabla\nabla\psi(\mathbf{f}) = \nabla\nabla \log p(\mathbf{x}|\mathbf{f}) - \mathbf{K}^{-1} = -\mathbf{W} - \mathbf{K}^{-1}$$

Estimation of $\psi(\mathbf{f})$

Newton's method is applied to find the maximum of $\psi(\mathbf{f})$ as

$$\begin{aligned}\mathbf{f}^{new} &= \mathbf{f} - (\nabla\nabla\psi(\mathbf{f}))^{-1}\nabla\psi(\mathbf{f}) \\ &= (\mathbf{W} + \mathbf{K}^{-1})^{-1}(\mathbf{W}\mathbf{f} - \nabla \log p(\mathbf{x}|\mathbf{f}))\end{aligned}$$

In addition, $\mathbf{A} = -\nabla\nabla\psi(\hat{f}) = \mathbf{W} + \mathbf{K}^{-1}$ where $\mathbf{W}$ is the negative Hessian matrix. Hence, the Laplace approximation to the posterior is a Gaussian with mean $\hat{\mathbf{f}}$ and covariance matrix $\mathbf{A}^{-1}$as

$$p(\mathbf{f} \mid \mathbf{x}) \simeq N(\hat{\mathbf{f}}, \mathbf{A}^{-1}) = N(\hat{\mathbf{f}}, (\mathbf{W} + \mathbf{K}^{-1})^{-1})$$

Model Parameter Estimation

The marginal likelihood is useful for estimating the model parameters θ and covariance parameters l

$$p\left(\mathbf{x}|\boldsymbol{\theta},\boldsymbol{\phi}\right)=\int p\left(\mathbf{x}|\mathbf{f},\boldsymbol{\theta}\right)p\left(\mathbf{f}|\boldsymbol{\phi}\right)df=\int \exp\left(\psi\left(\mathbf{f}\right)\right)df$$

Using Taylor expansion of $\psi(\mathbf{f})$,

$$\log p(\mathbf{x}|\boldsymbol{\theta},\boldsymbol{\phi})=\log p\left(\mathbf{x}|\hat{\mathbf{f}},\boldsymbol{\theta},\boldsymbol{\phi}\right)-\frac{1}{2}\mathbf{f}^{T}\mathbf{K}^{-1}\mathbf{f}-\frac{1}{2}\log\left|\mathbf{I}+\mathbf{KW}\right|$$

The parameters $\boldsymbol{\eta}=\{\boldsymbol{\theta},\boldsymbol{\phi}\}$ can be then estimated by using

$$\frac{\partial\log p\left(\mathbf{x}|\boldsymbol{\eta}\right)}{\partial\boldsymbol{\eta}}=\frac{\partial\log p\left(\mathbf{x}|\boldsymbol{\eta}\right)}{\partial\boldsymbol{\eta}}\,|_{\text{explicit}}+\frac{\partial\log p\left(\mathbf{x}|\boldsymbol{\eta}\right)}{\partial\hat{\mathbf{f}}}\frac{\partial\hat{\mathbf{f}}}{\partial\boldsymbol{\eta}}$$

Outline

SOS Response

- DNA damage in bacteria may occur as a result of activity of antibiotics.
- LexA is bound to the genome preventing transcription of the SOS genes.
- RecA protein is stimulated by single stranded DNA, inactivates the LexA repessor.
- This allows several of the LexA targets to transcribe.
- The SOS pathway may be essential in antibiotic resistance Cirz et al. (2005).
- Aim is to target these proteins to produce drugs to increase efficacy of antibiotics Lee et al. (2005).

LexA Experimental Description

- Data from Courcelle et al. (2001)
- UV irradiation of *E. coli.* in both wild-type cells and lexA1 mutants, which are unable to induce genes under LexA control.
- Response measured with two color hybridization to cDNA arrays.

Khanin et al. Model

Given measurements of gene expression at N time points $(t_0, t_1, \ldots, t_{N-1})$, the temporal profile of a gene i, $x_i(t)$, that solves the ODE in Eq. 1 can be approximated by

$$x_i(t) = x_i^0 e^{-D_i t} + \frac{B_i}{D_i} + S_i e^{-D_i t} \int_0^t g(f(u)) e^{D_i u} \mathrm{d}u.$$

$$x_i(t) = x_i^0 e^{-D_i t} + \frac{B_i}{D_i} + S_i e^{-D_i t} \frac{1}{t_{j+1} - t_j} \sum_{j=0}^{N-2} g(\bar{f}_j) \left(e^{D_i t_{j+1}} - e^{D_i t_j} \right)$$

where $\bar{f}_j = \frac{(f(t_j) + f(t_{j+1}))}{2}$ on each subinterval $(t_j, t_j + 1)$, $j = 0, \ldots, N-2$. This is under the simplifying assumption that $f(t)$ is a piece-wise constant function on each subinterval $(t_j, t_j + 1)$. Repression model: $g(f(t)) = \frac{1}{\gamma + e^{f(t)}}$.

Khanin et al. Results

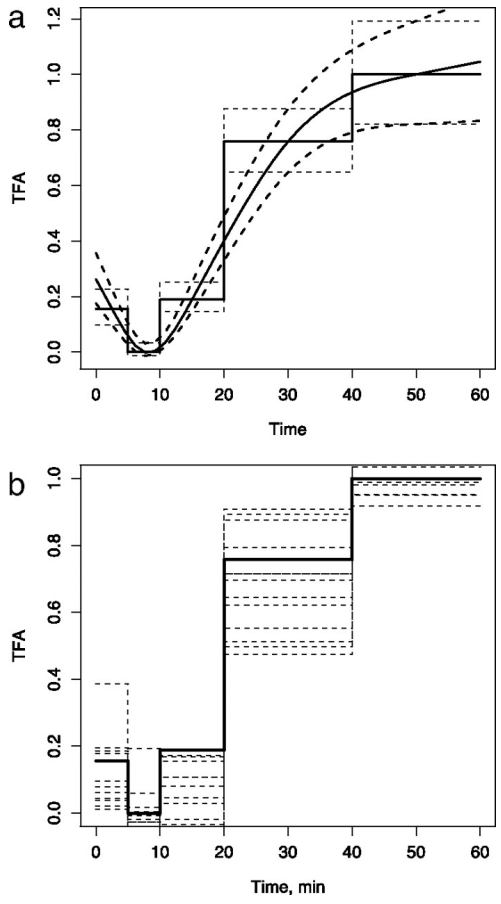


Figure: Fig. 2 from Khanin et al. (2006): Reconstructed activity level of master repressor LexA, following a UV dose of 40 J/m2.

Khanin et al. Results

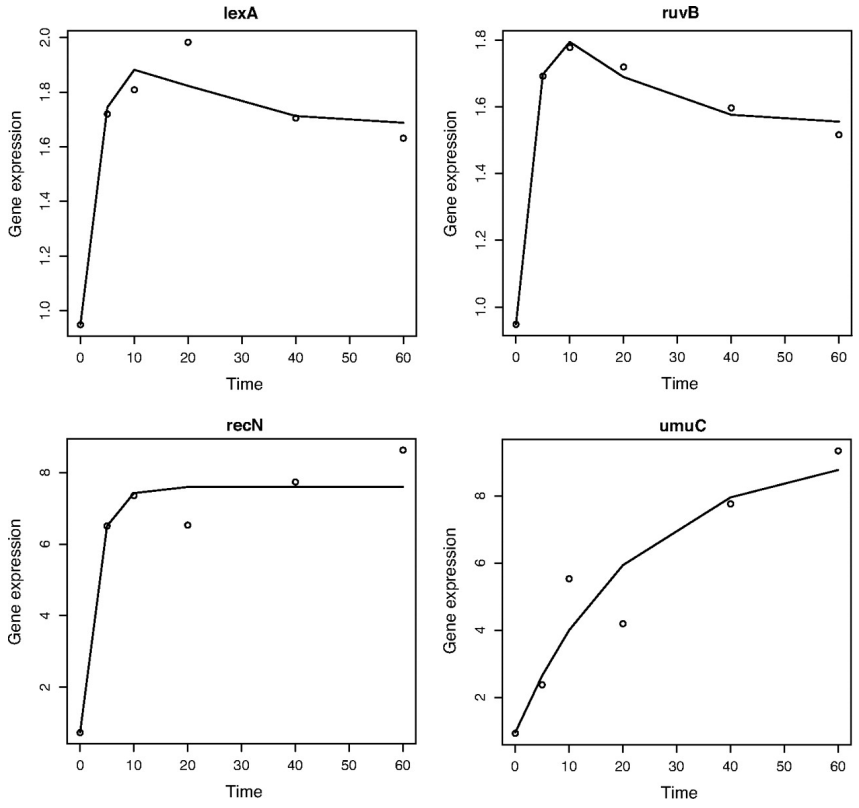


Figure: Fig. 3 from Khanin et al. (2006): Reconstructed profiles for four genes in the LexA SIM.

Repression Model

- We can use the same model of repression,

$$g_j\left(f\left(t\right)\right) = \frac{1}{\gamma_j + e^{f(t)}}$$

In the case of repression we have to include the transient term,

$$x_j\left(t\right) = \alpha_j e^{-D_j t} + \frac{B_j}{D_j} + S_j \int_0^t e^{-D_j(t-u)} g_j(f\left(u\right))\mathrm{d}u$$

Results for the repressor LexA

Pei Gao

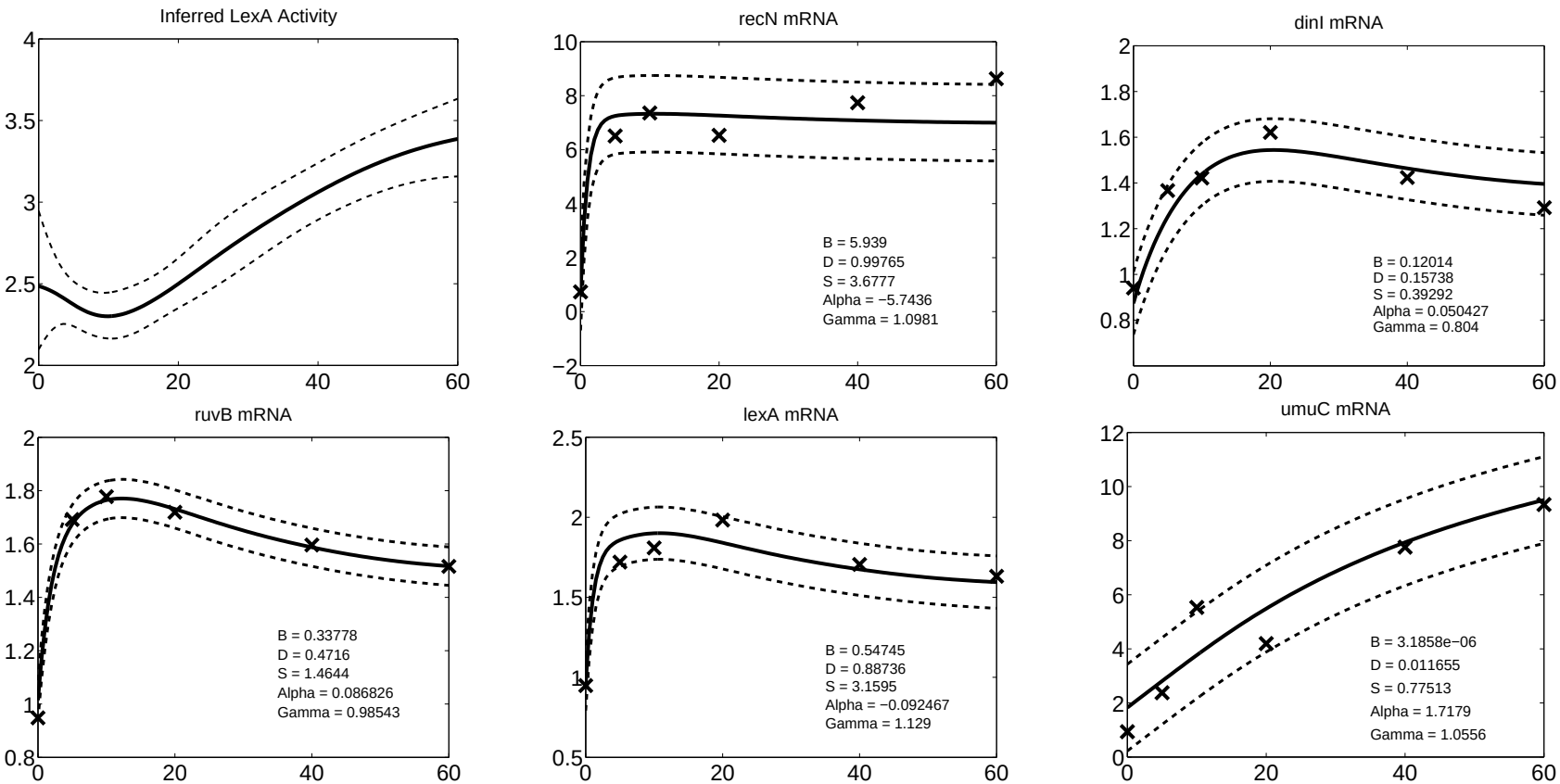


Figure: Our results using an MLP kernel. From Gao et al. (2008).

Outline

Use Samples to Represent Posterior

Michalis Titsias

- Sample in Gaussian processes

$$p(\mathbf{f}|\mathbf{x}) \propto p(\mathbf{x}|\mathbf{f})\, p(\mathbf{f})$$

- Likelihood relates GP to data through

$$x_j(t) = \alpha_j e^{-D_j t} + \frac{B_j}{D_j} + S_j \int_0^t e^{-D_j(t-u)} g_j(f(u)) \mathrm{d}u$$

- We use *control points* for fast sampling.

MCMC for Non Linear Response

The Metropolis-Hastings algorithm

- Initialize $\mathbf{f}^{(0)}$
- Form a Markov chain. Use a proposal distribution $Q(\mathbf{f}^{(t+1)}|\mathbf{f}^{(t)})$ and accept with the M-H step

$$\min\left(1, \frac{p(\mathbf{x}|\mathbf{f}^{(t+1)})p(\mathbf{f}^{(t+1)})}{p(\mathbf{x}|\mathbf{f}^{(t)})p(\mathbf{f}^{(t)})}\frac{Q(\mathbf{f}^{(t)}|\mathbf{f}^{(t+1)})}{Q(\mathbf{f}^{(t+1)}|\mathbf{f}^{(t)})}\right)$$

- $\mathbf{f}$ can be very *high dimensional* (hundreds of points)
- How do we choose the proposal $Q(\mathbf{f}^{(t+1)}|\mathbf{f}^{(t)})$?
  - Can we use the GP prior $p(\mathbf{f})$ as the proposal?

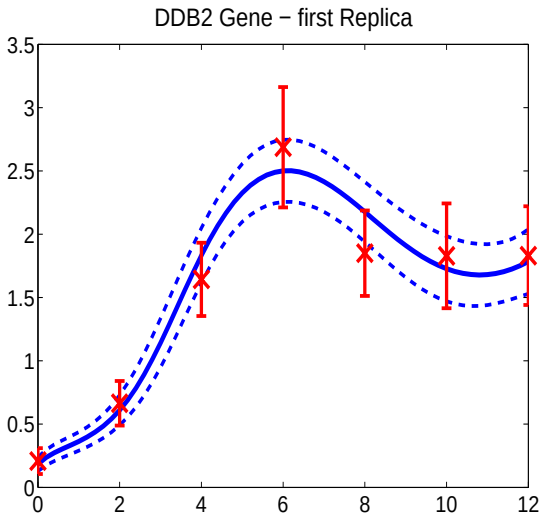
p53 System Again

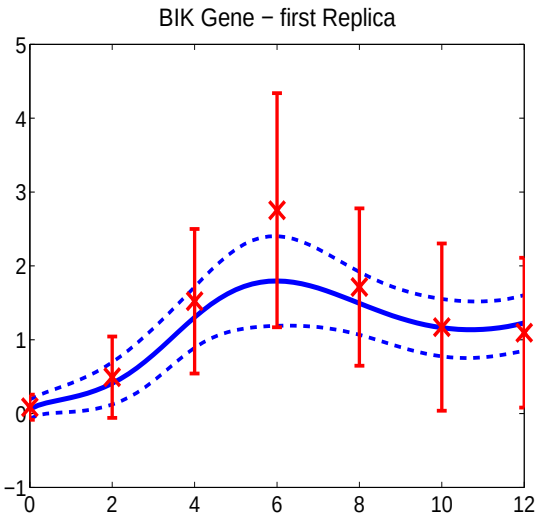
- One transcription factor (p53) that acts as an activator. We consider the Michaelis-Menten kinetic equation

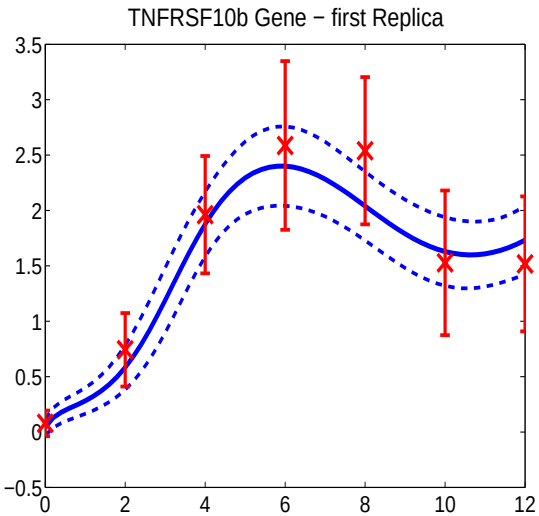
$$\frac{\mathrm{d}x_j(t)}{\mathrm{d}t} = B_j + S_j \frac{\exp(f(t))}{\exp(f(t)) + \gamma_j} - D_j x_j(t)$$

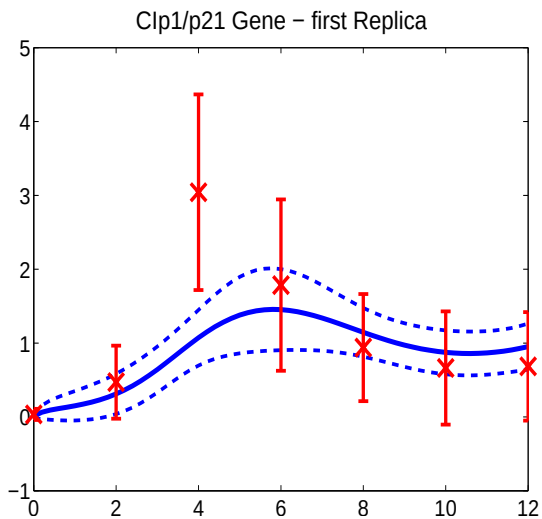
- We have 5 genes
- Gene expressions are available for $T = 7$ times and there are 3 replicas of the time series data
- TF ($\mathbf{f}$) is discretized using 121 points
- MCMC details:
  - 7 control points are used (placed in a equally spaced grid)
  - Running time 4/5 hours for 2 million sampling iterations plus burn in
  - Acceptance rate for $\mathbf{f}$ after burn in was between $15\% - 25\%$

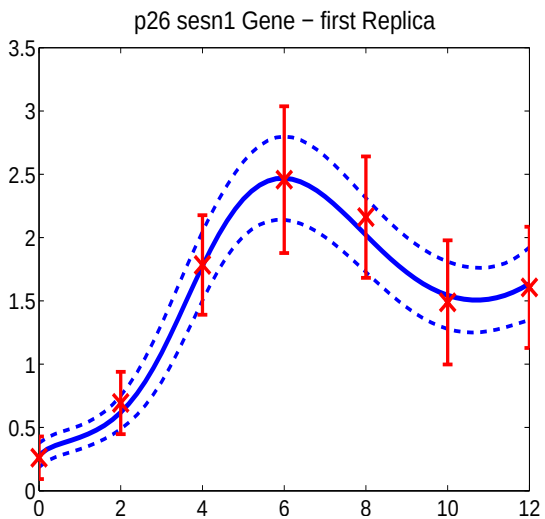
Data used by Barenco et al. (2006): Predicted gene expressions for the 1st replica



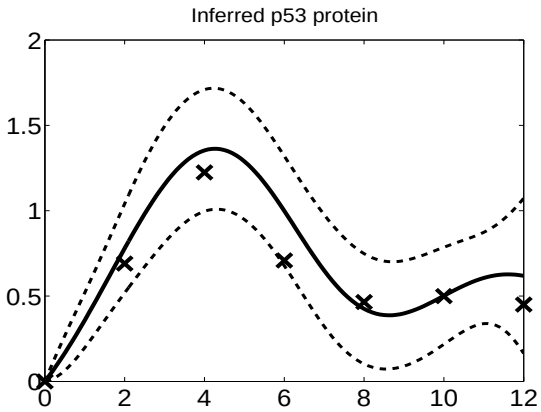


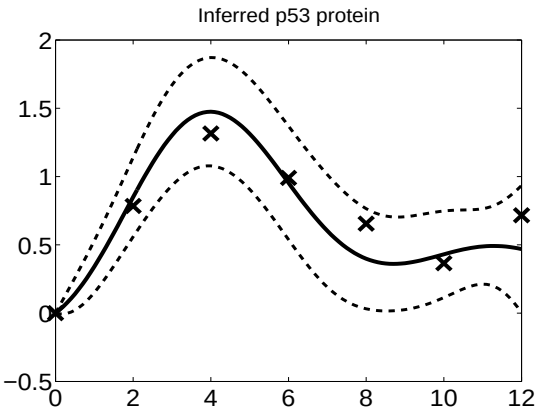


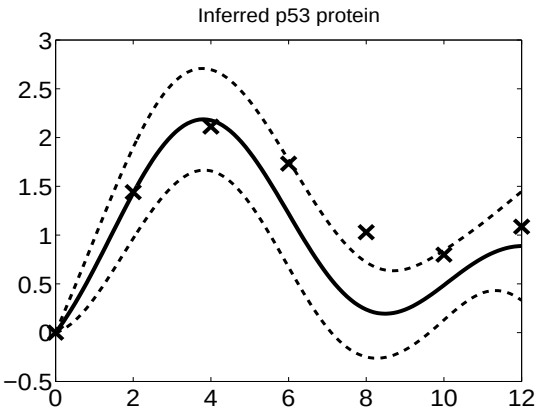




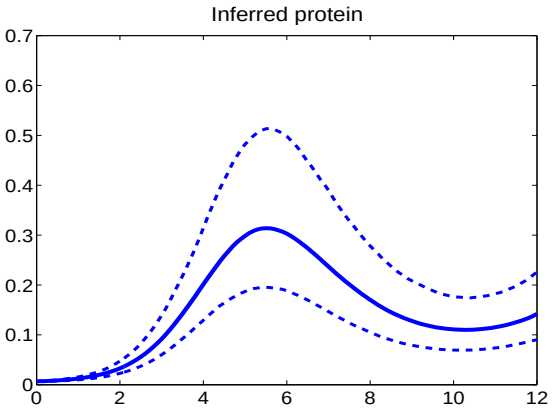
Data used by Barenco et al. (2006): Protein concentrations

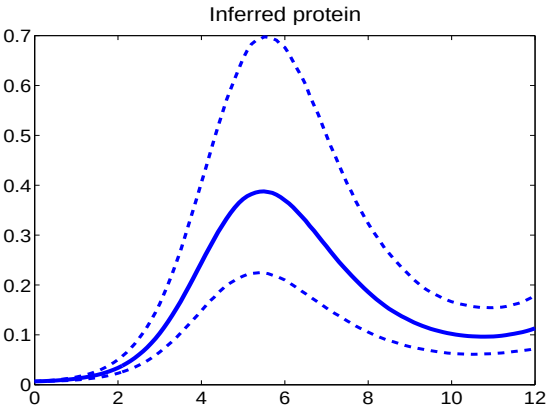


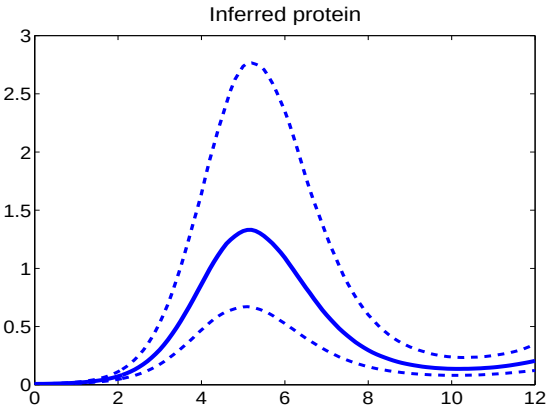




Linear model (Barenco et al. predictions are shown as crosses)

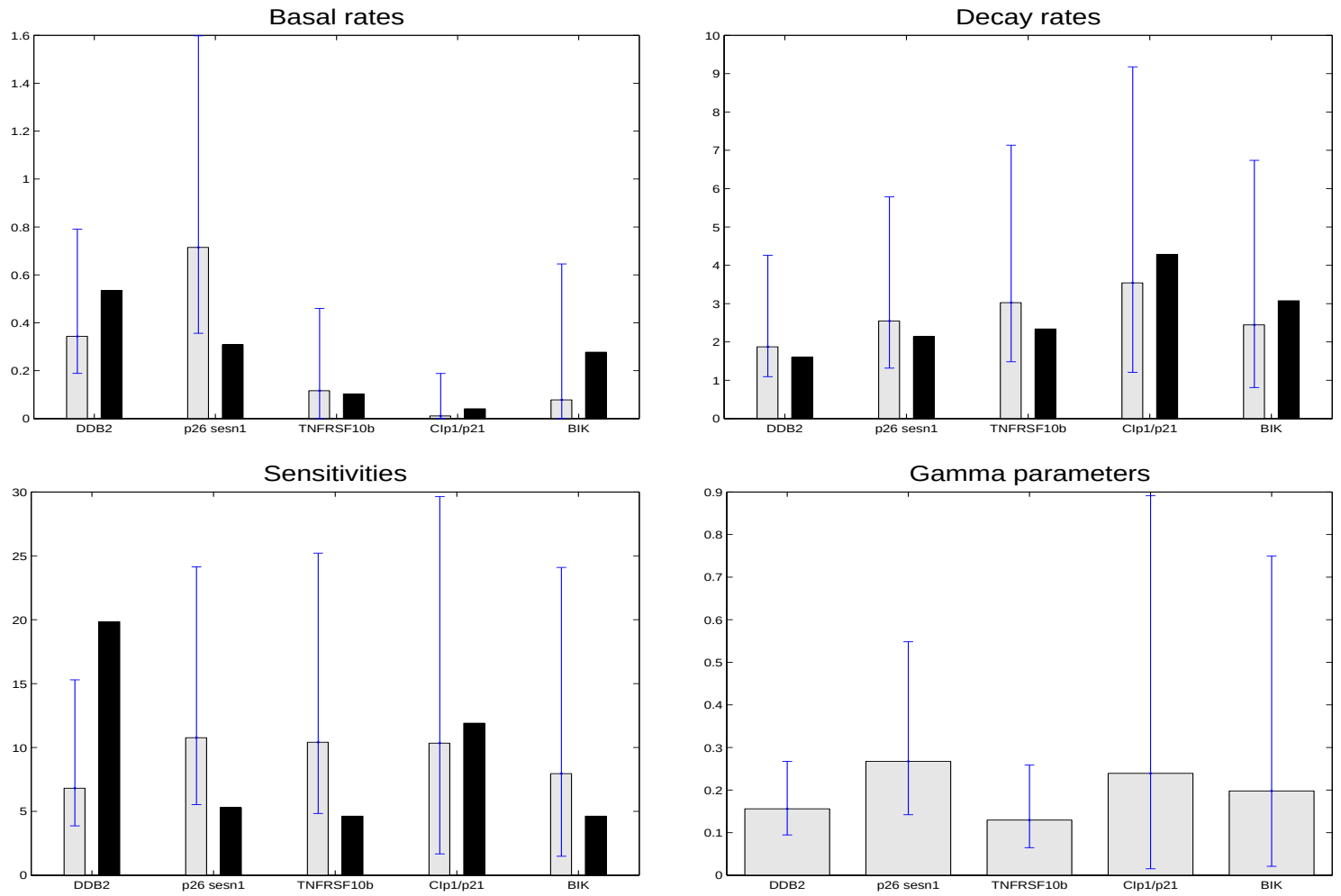






Nonlinear (Michaelis-Menten kinetic equation)

p53 Data Kinetic parameters



Our results (grey) compared with Barenco et al. (2006) (black).
Note that Barenco et al. use a linear model

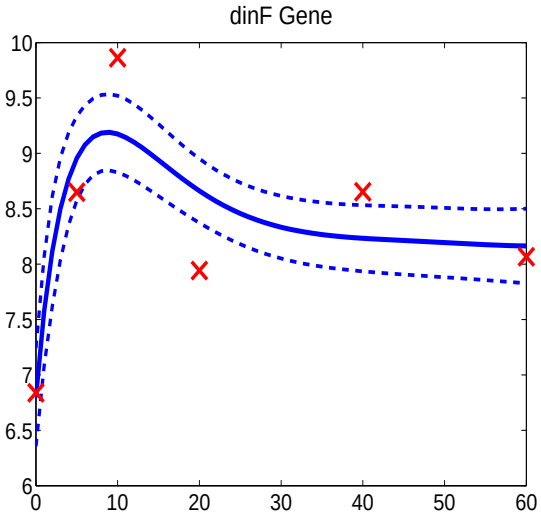
Results on SOS System

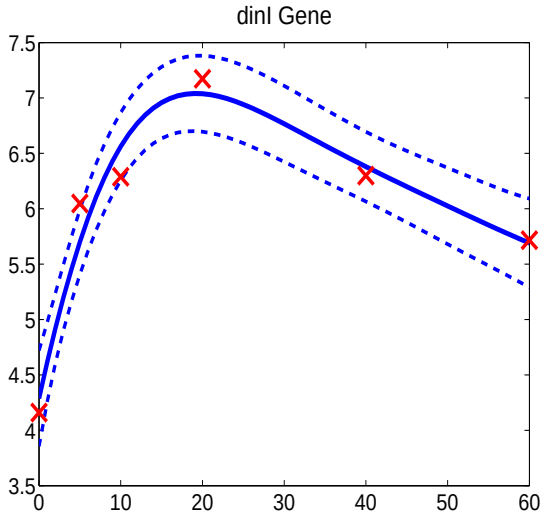
- Again consider the Michaelis-Menten kinetic equation

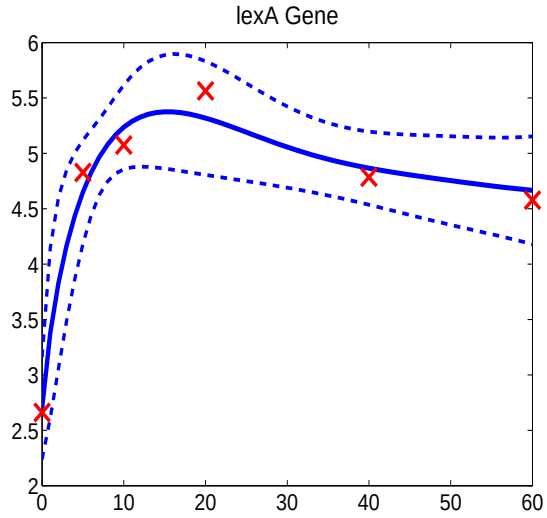
$$\frac{\mathrm{d}x_j(t)}{\mathrm{d}t} = B_j + S_j \frac{1}{\exp(f(t)) + \gamma_j} - D_j x_j(t)$$

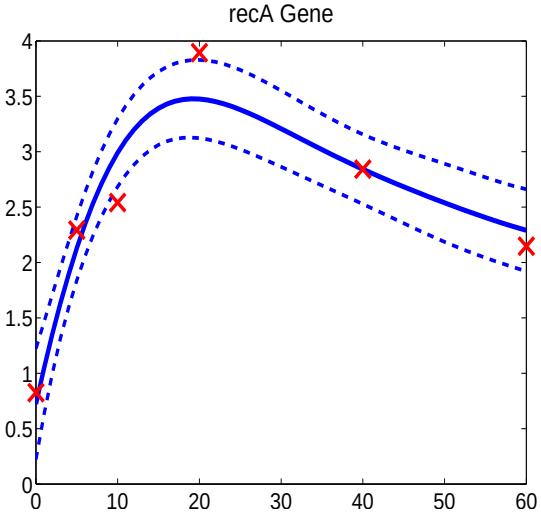
- We have 14 genes (5 kinetic parameters each)
- Gene expressions are available for $T = 6$ time slots
- TF ($\mathbf{f}$) is discretized using 121 points
- MCMC details:
  - 6 control points are used (placed in a equally spaced grid)
  - Running time was 5 hours for 2 million sampling iterations plus burn in
  - Acceptance rate for $\mathbf{f}$ after burn in was between $15\% - 25\%$

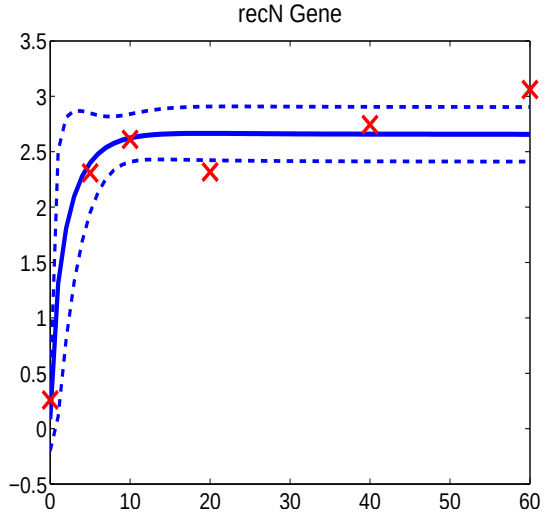
Results in E.coli data: Predicted gene expressions

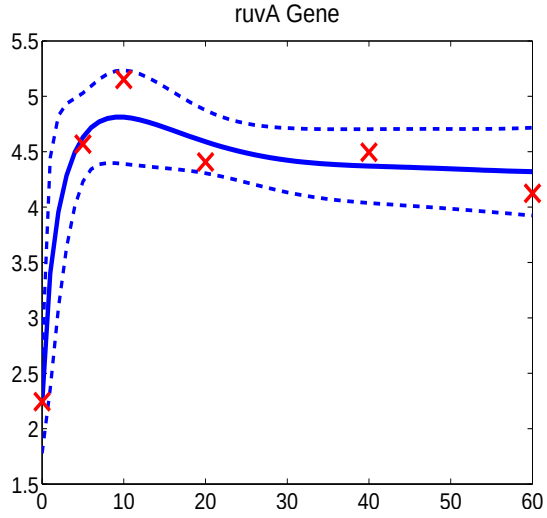




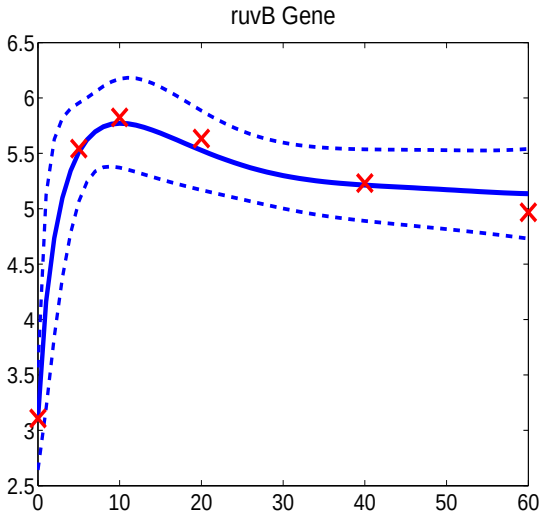


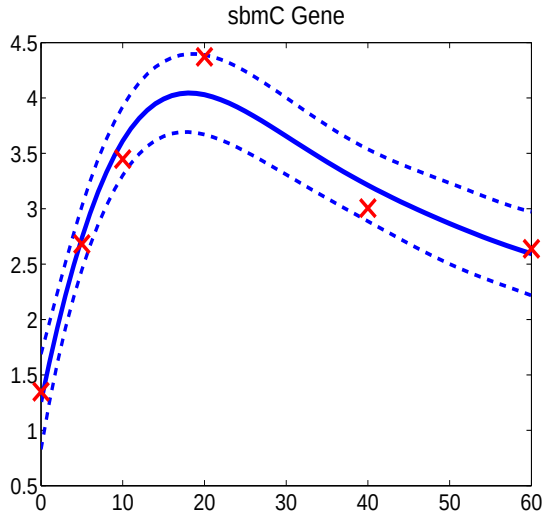


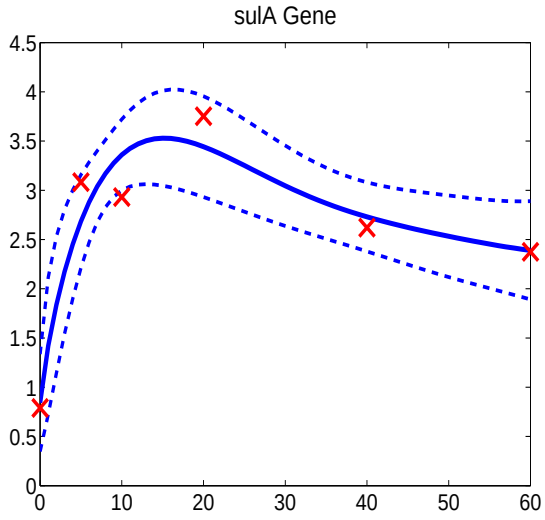


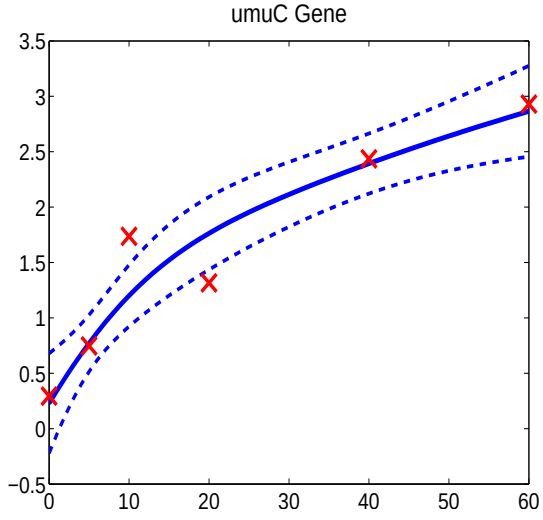


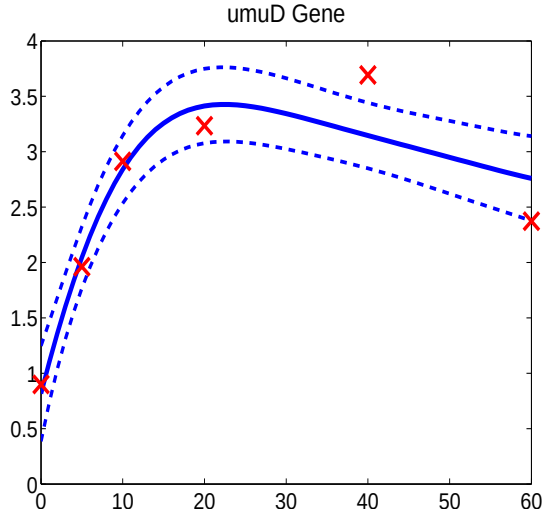
Results in E.coli data: Predicted gene expressions

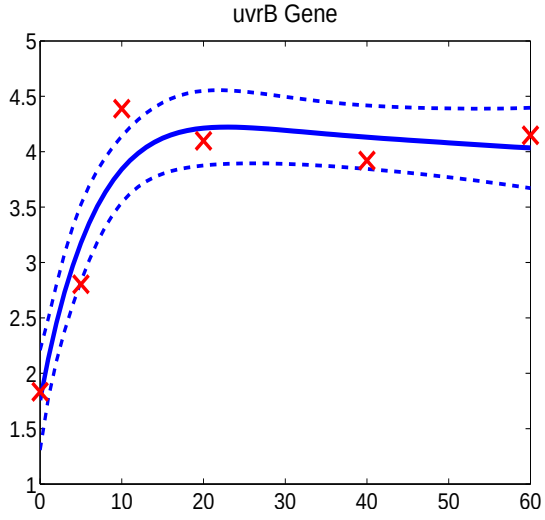




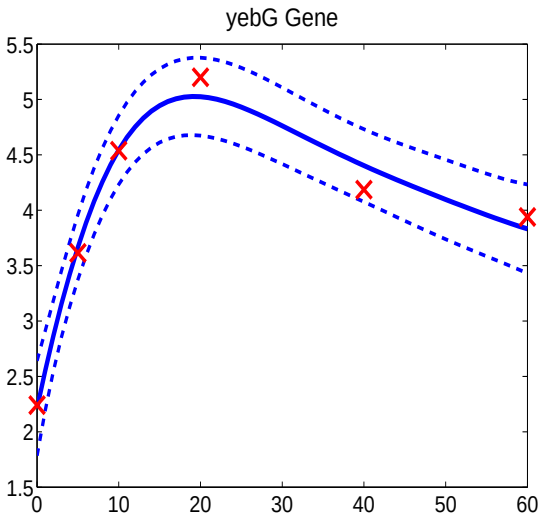


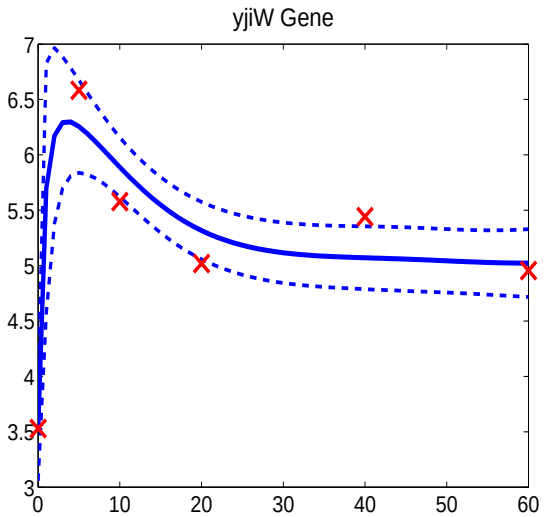




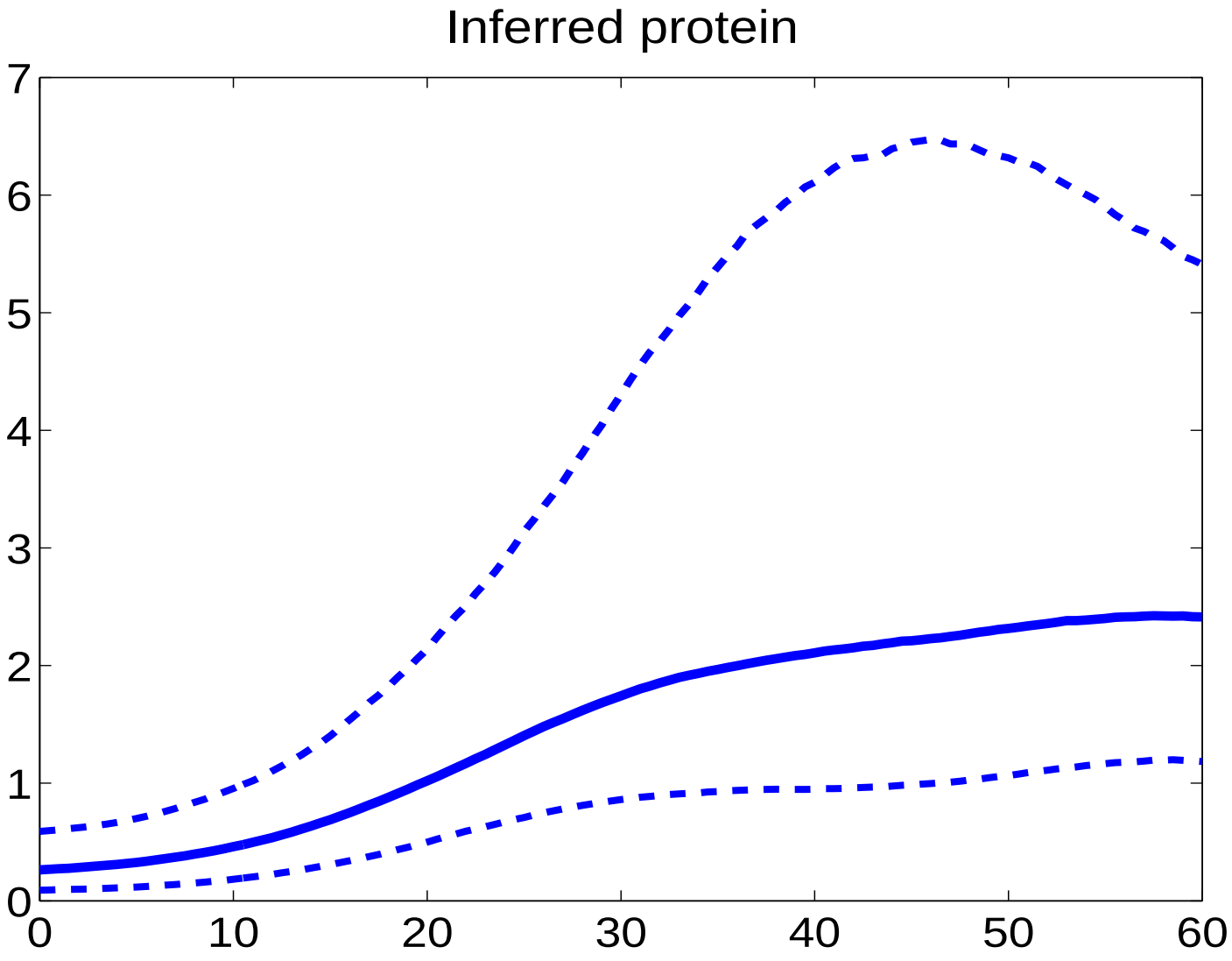


Results in E.coli data: Predicted gene expressions

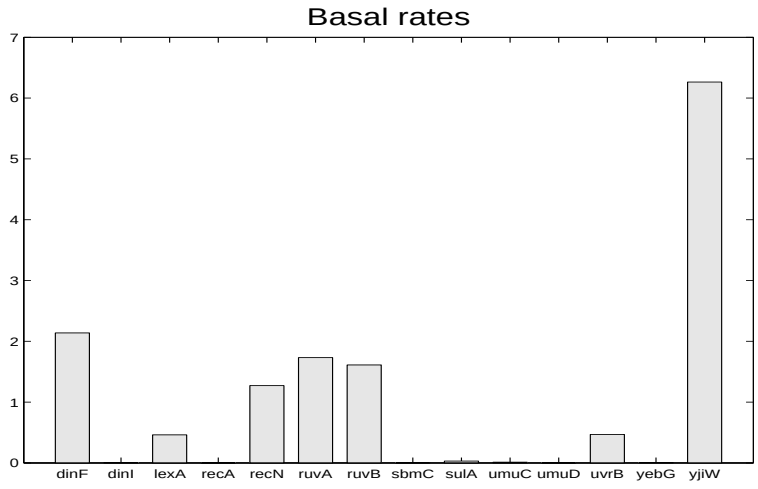


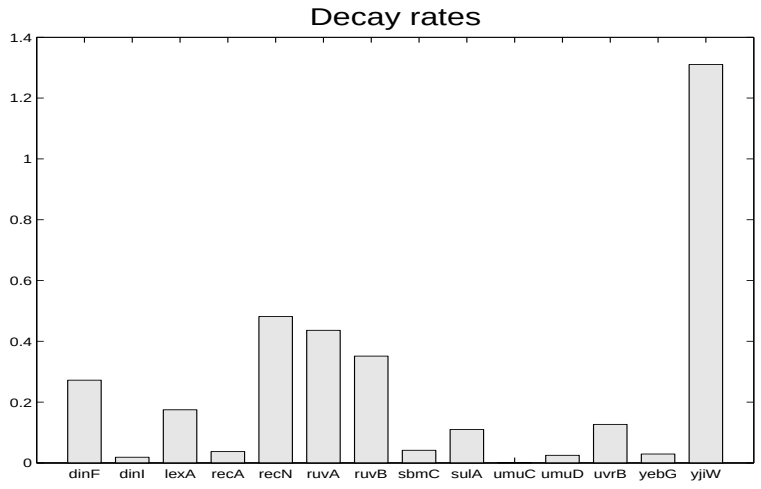


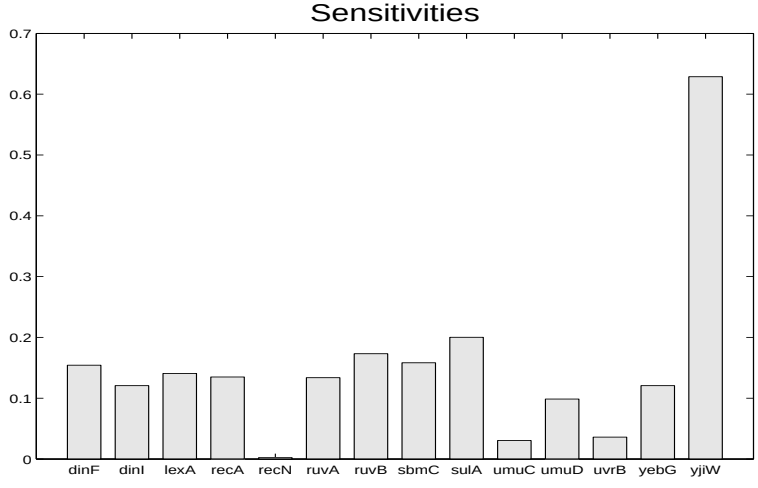
Results in E.coli data: Protein concentration

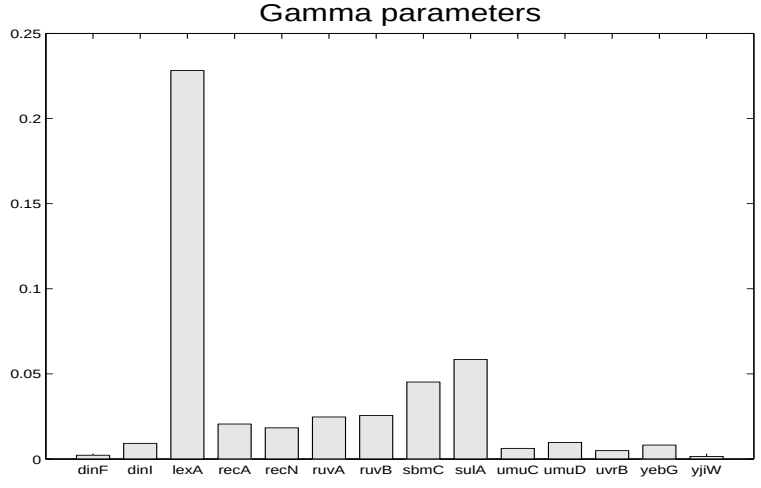


Results in E.coli data: Kinetic parameters

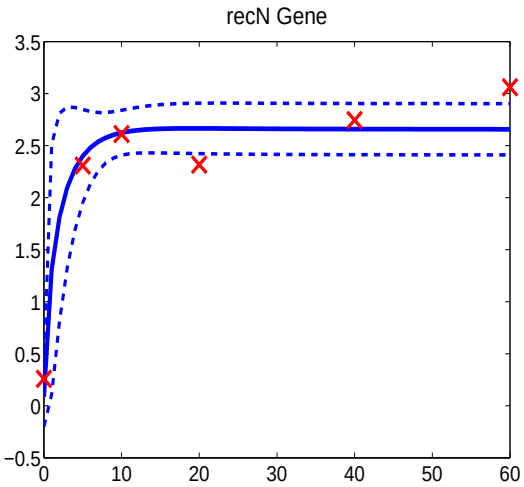


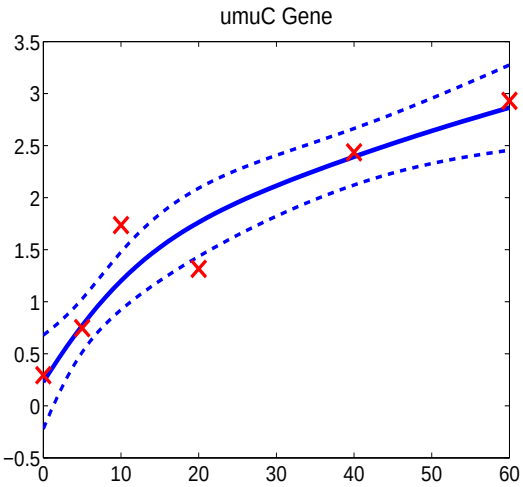


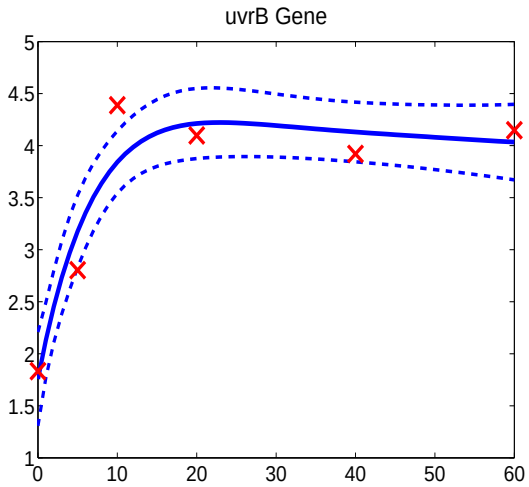


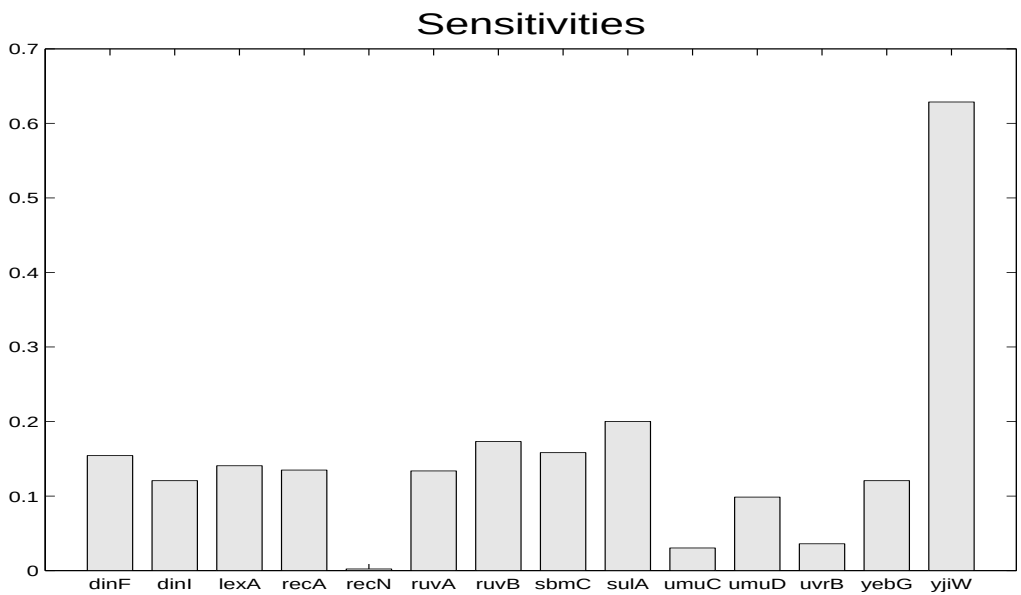


Results in E.coli data: Genes with low sensitivity value

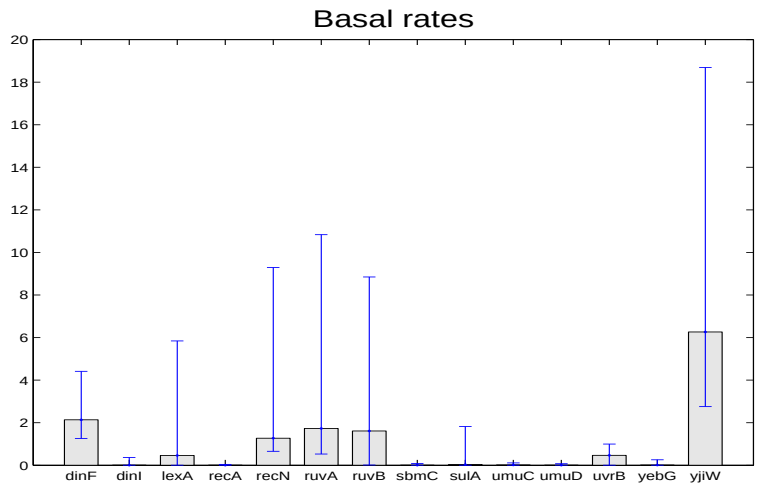


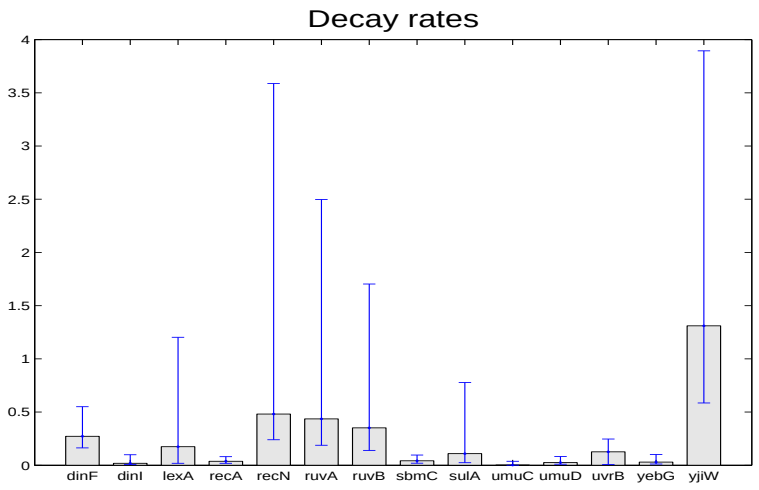


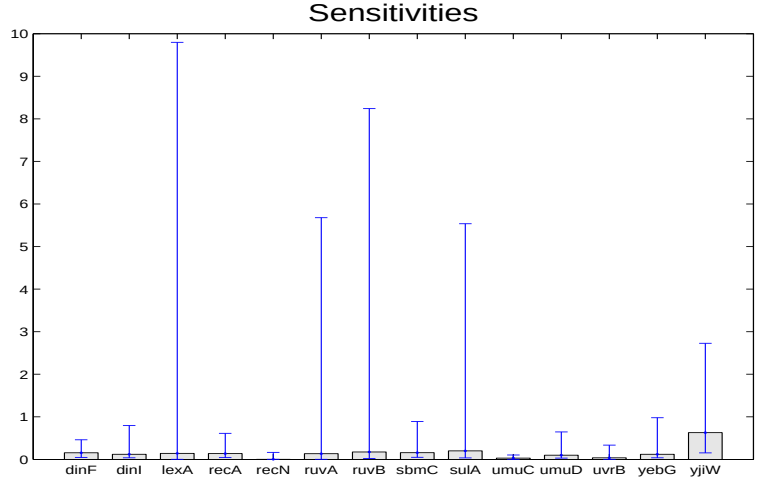


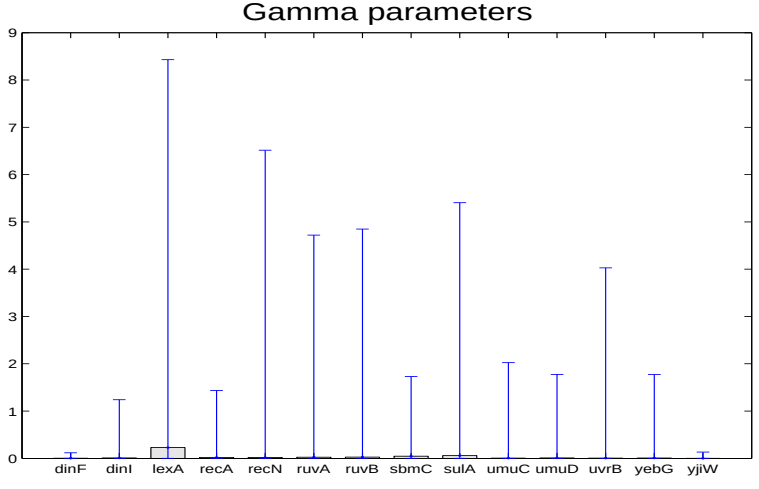


Results in E.coli data: Confidence intervals for the kinetic parameters









Outline

Multiple TFs

- We can generalize the Gaussian process sampling framework to estimate from gene expression data multiple and possibly interacting TFs.
- For linear response, this is tractable, but for nonlinear response (in general) we use sampling.

Learning multiple TFs

- General form of the multiple TF model

$$\frac{\mathrm{d}x_j(t)}{\mathrm{d}t} = B_j + S_j g(f_1(t), \ldots, f_l(t); \mathbf{w}_j) - D_j x_j(t), \quad (1)$$

where the l-dimensional vector $\mathbf{w}_j$ stores the interaction weights between the jth gene and the l TFs. There may be also some bias weight w_{0j} for each gene.

Sigmoid model

- Choose the joint activation function $g(u)$ to be the sigmoid (Mjolsness et al., 1991)

$$h_j = \sum_{i=1}^{I} w_{ji} f_i(t) + w_{j0},$$

$$g(h_j) = \frac{1}{1 + \exp(-h_j)}.$$

- For single TF the above activation function gives rise to Michaelis-Menten when we fix $w_j = 1$.
- For the repressor case we set $w_j = -1$, which however doesn't give rise to the exact Michaelis-Menten repressor equation

Bayesian model

- Likelihood:

$$\prod_{j=1}^{N}\prod_{t=1}^{T} p(x_{jt}|\{\mathbf{f}_i(1 \leq p \leq P_t)\}_{i=1}^{I}, \{A_j, B_j, D_j, S_j\}, \mathbf{w}_j, \sigma_j^2), \tag{2}$$

  where these terms are Gaussians and σ_j^2 is gene-specific variance

- Prior
  - Kinetics $\{A_j, B_j, D_j, S_j\}$ are positive and are represented in the log space: Gaussian priors are used
  - $\{\mathbf{f}\}_{i=1}^{I}$ are the log of the TFs: GP rbf priors with separate timescales
  - $\{\mathbf{w}_j\}$ take real values: Gaussian priors are used
  - Noise variances and GP lengthscales $\{\sigma_j^2, \ell_j^2\}$: Gamma priors

MCMC

Component-wise M-H algorithm. Iteratively sample from conditional posteriors:

1. For $i = 1, \ldots, I$ sample $\mathbf{f}_i$ from the conditional posterior based on the approach of Titsias et. al [2009]
2. For $j = 1, \ldots, N$ sample the kinetic parameters $\{A_j, B_j, D_j, S_j\}$
3. For $j = 1, \ldots, N$ sample the interaction weights $\mathbf{w}_j$
4. For $j = 1, \ldots, N$ sample the gene-specific noise variance σ_j^2.
5. For $i = 1, \ldots, I$ sample the lengthscale ℓ_i^2 of the rbf kernel function.

Side Information

Learning the real TFs that produced the gene expression is not easy because of identifiability problems in parameter space and limited amount of data. Side information obtained from ChiP data can be useful.

- Side information involves the weights W that represent the interactions between genes and TFs. W is $N \times I$ matrix where N the number of genes and I the number of TFs.
- Side information can be expressed as a binary $N \times I$ matrix X. When $x_{ji} = 0$, there is no interaction between the j gene and the i TF, thus $w_{ji} = 0$. When $x_{ji} = 1$, the value w_{ji} can take a positive or negative value which must be inferred by MCMC.
- This scheme can be generalized to probabilistically expressed side information where each x_{ij} is drawn from some probability π_{ji} that expresses our prior belief that the j gene has been regulated by the i TF.

Artificial data

- We consider a toy example with two TFs, that can regulate 20 genes.
- We assume that we have deterministic side information for 8 out of 20 genes. i.e. we know which weights w_{j1} and w_{j2} are zero for these 8 genes, say $j = 1, \ldots, 8$.
- We also assume that the initial conditions in the differential equations are all zero and also that we know the initial (at $t = 0$) activation of the TFs. The number of non-zero elements in the 20×2 matrix W is 25.

Artificial data

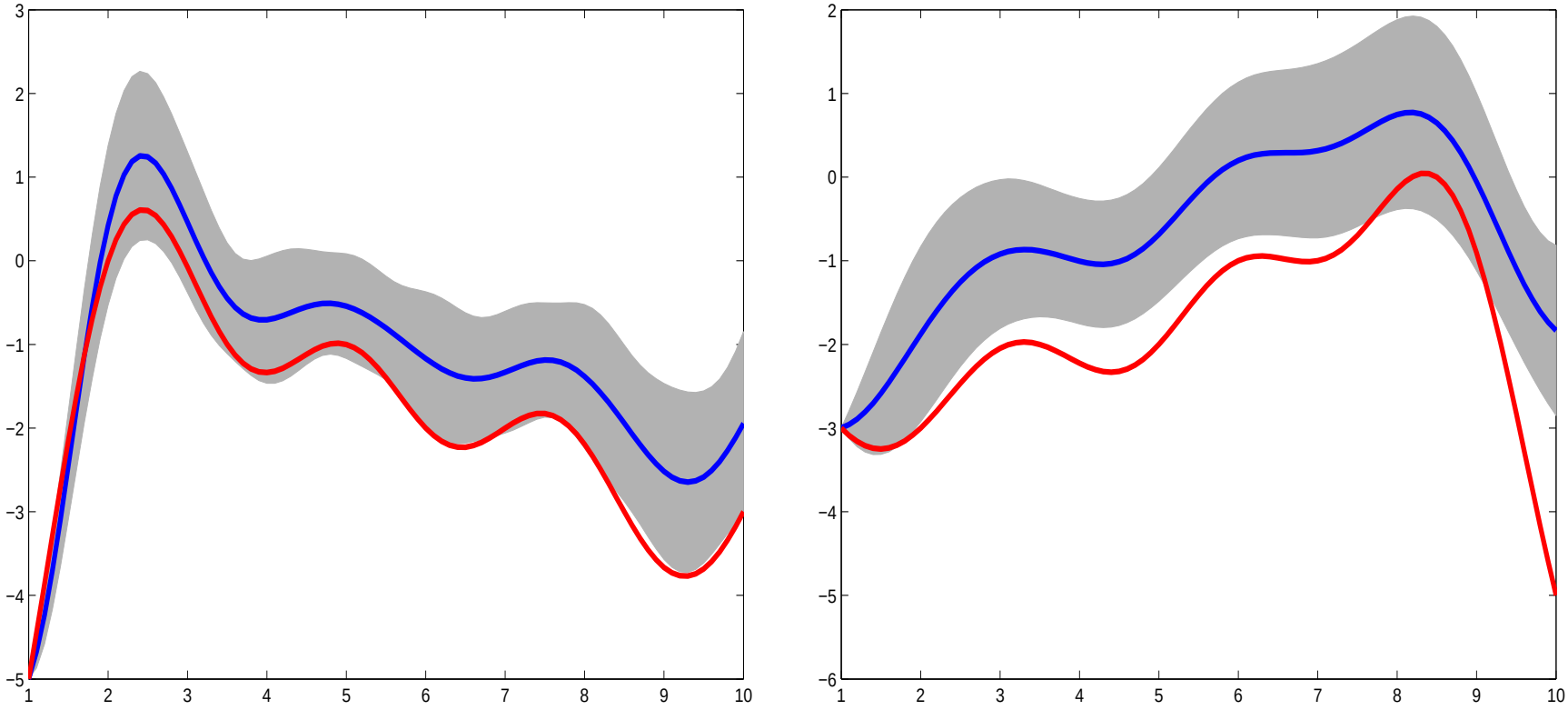


Figure: The inferred profiles of the two TFs (in the log space). With red solid lines are the ground-truth TFs used to generate the toy data. With blue lines shaded error bars are the inferred TF profiles.

Artificial data

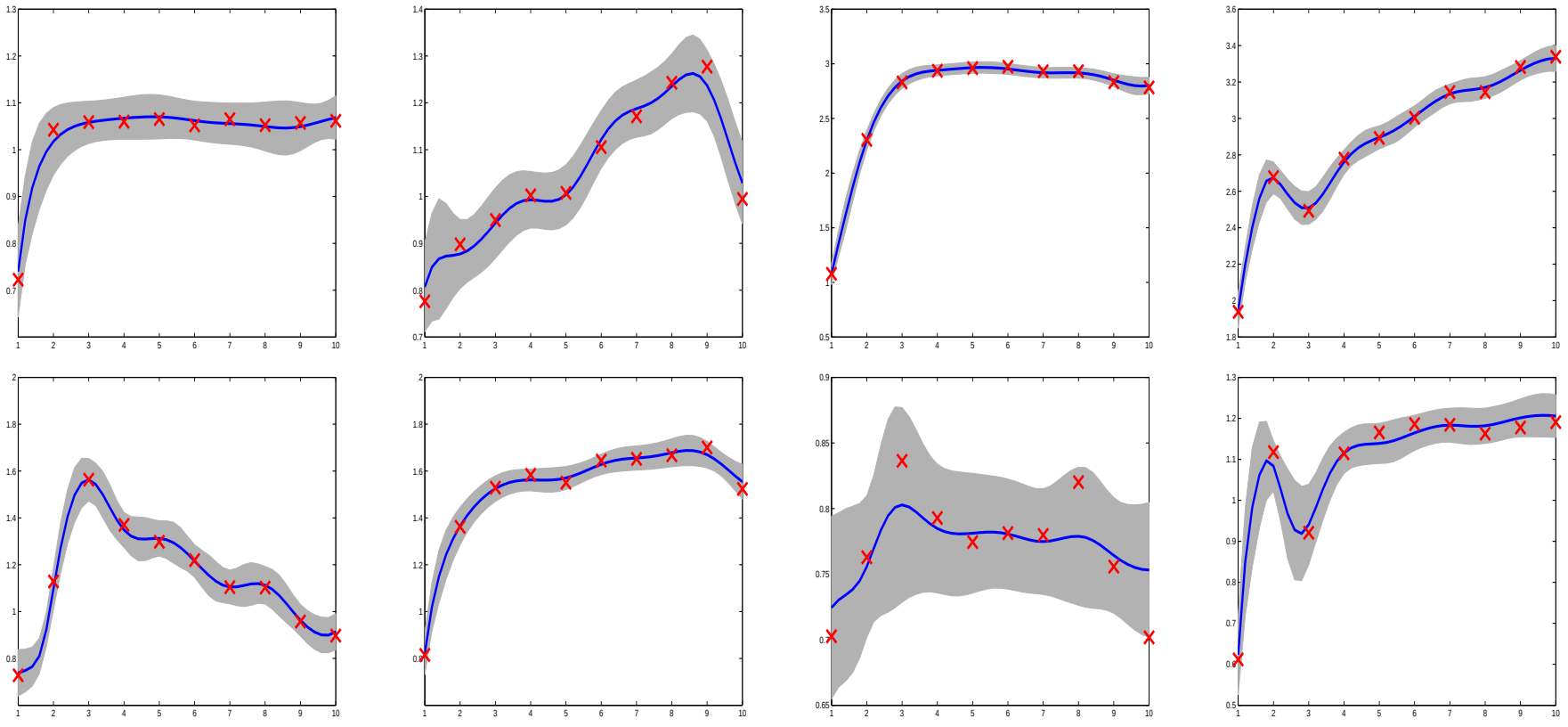


Figure: The predicted gene expressions. Red crosses represent the actual gene expression and the blue line with shaded error bars are the prediction found by MCMC.

Artificial data

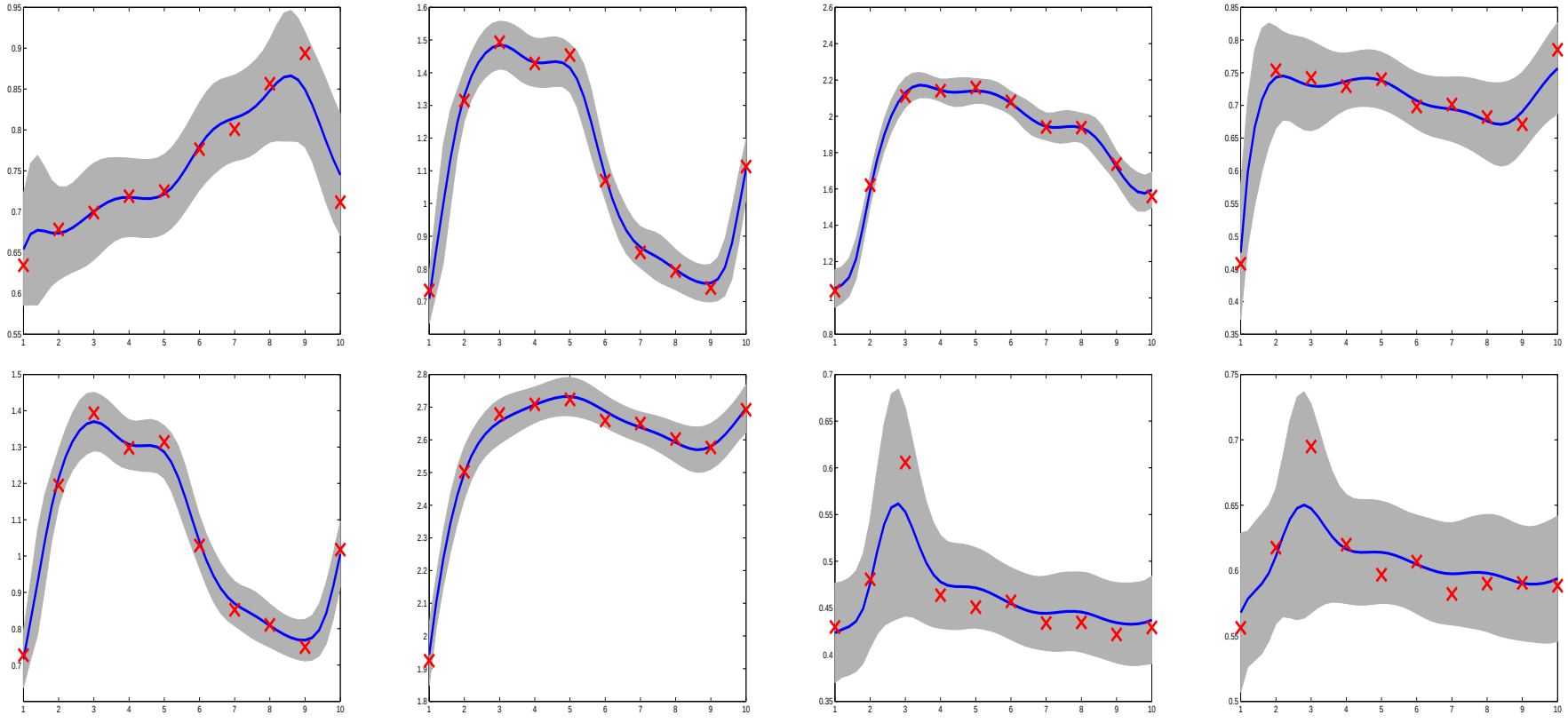


Figure: The predicted gene expressions. Red crosses represent the actual gene expression and the blue line with shaded error bars are the prediction found by MCMC.

Artificial data

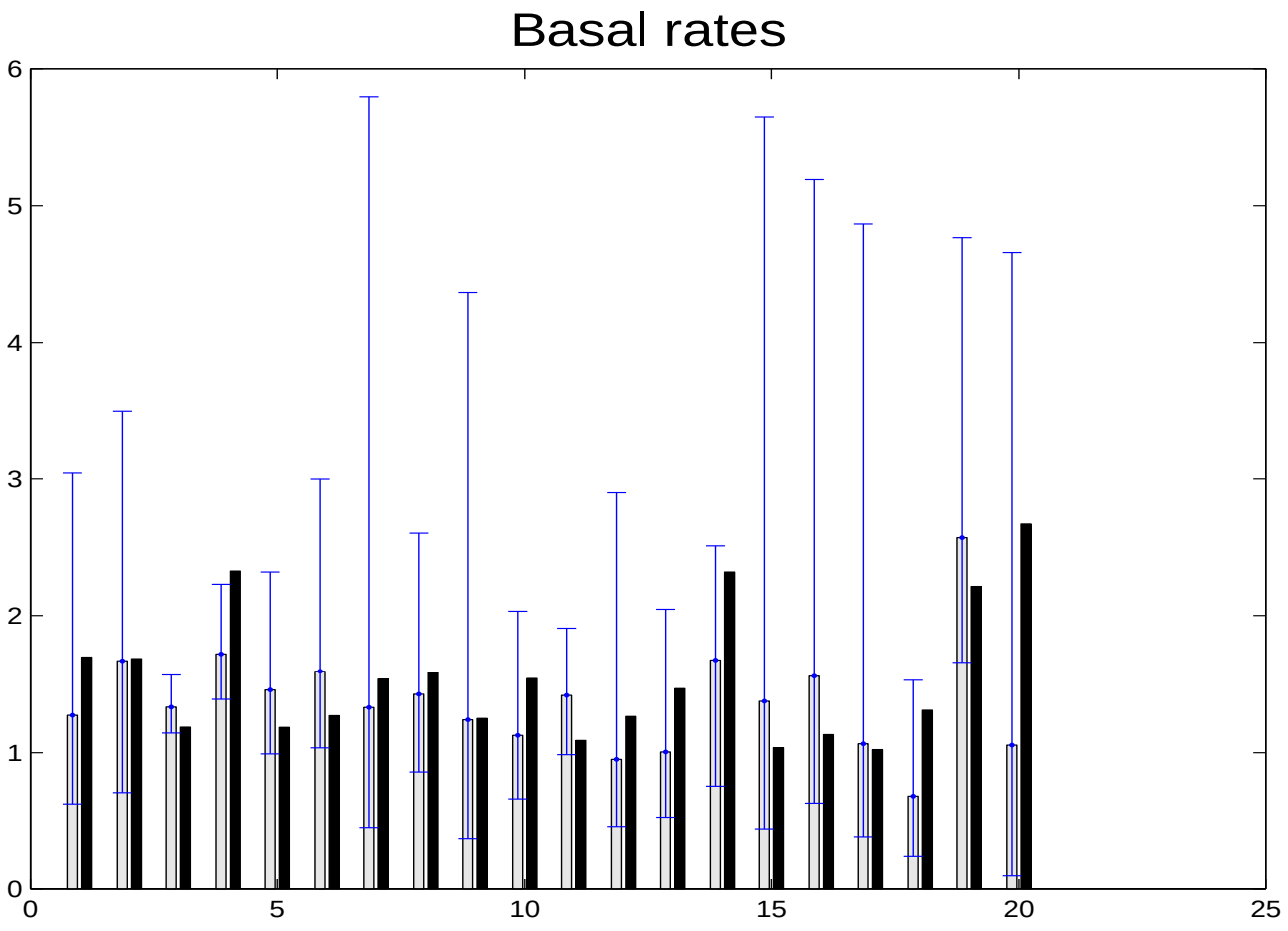


Figure: The inferred basal rates for the 20 genes.

Artificial data

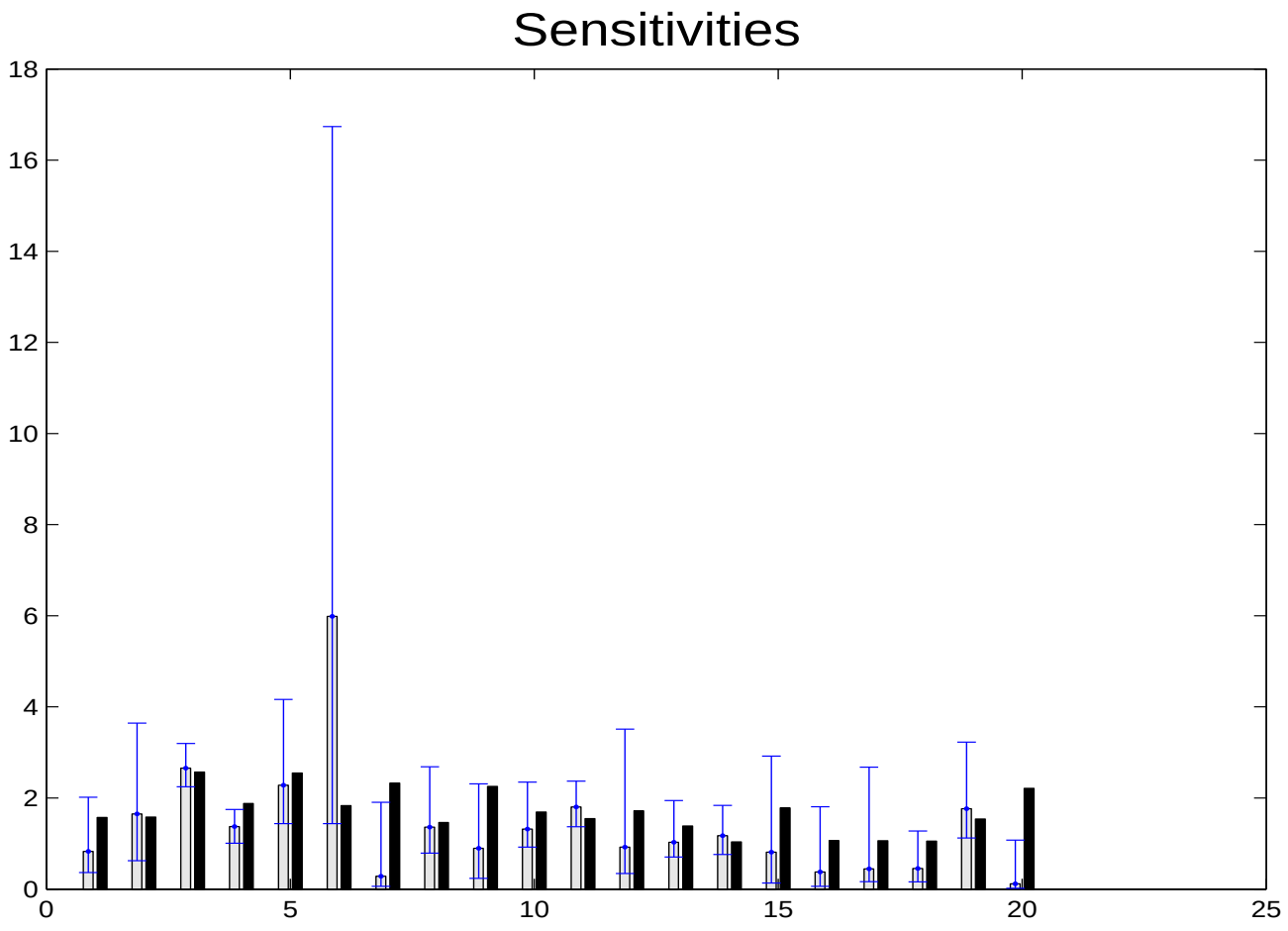


Figure: The inferred sensitivities for the 20 genes.

Artificial data

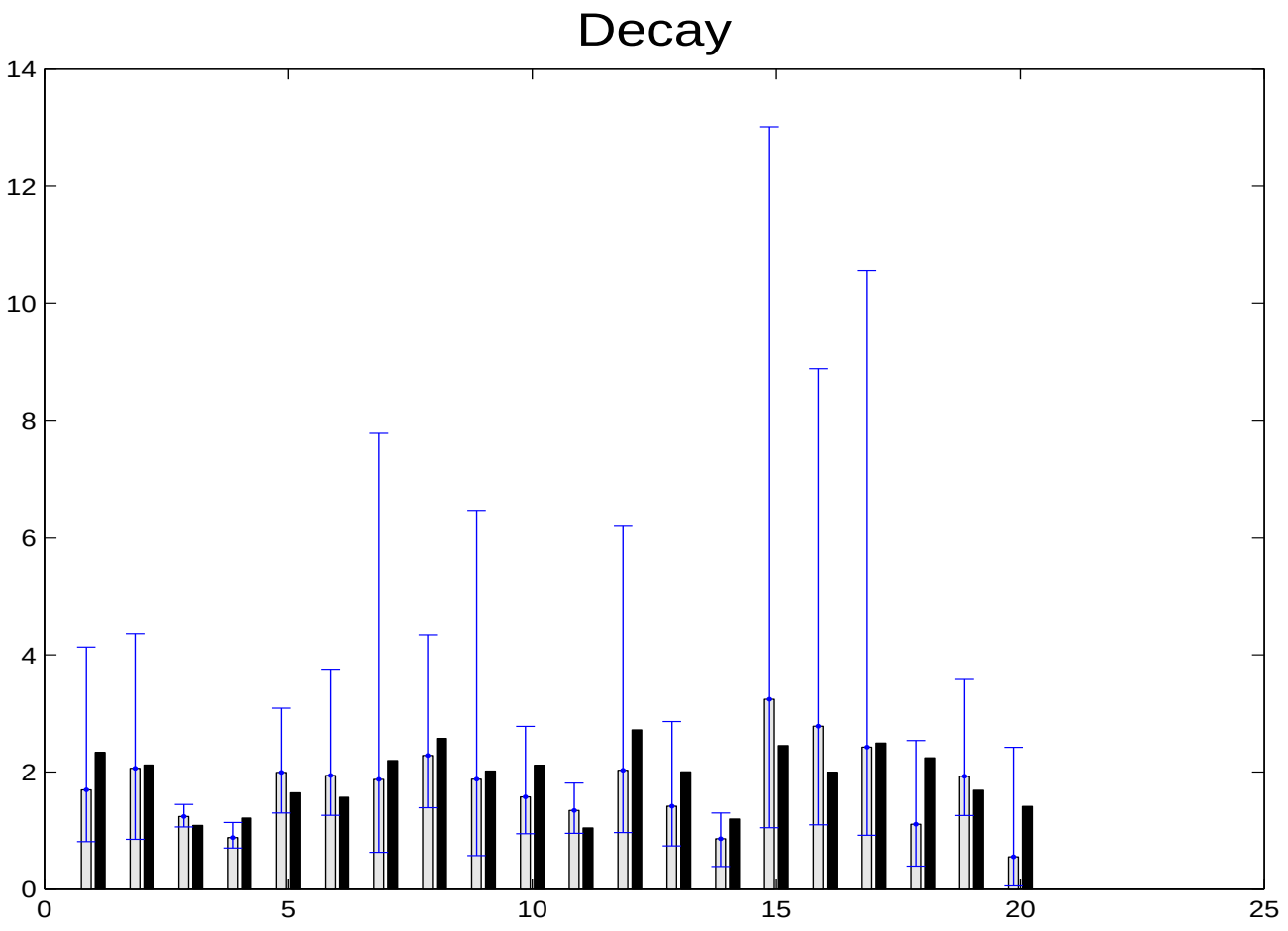


Figure: The inferred decays for the 20 genes.

Artificial data

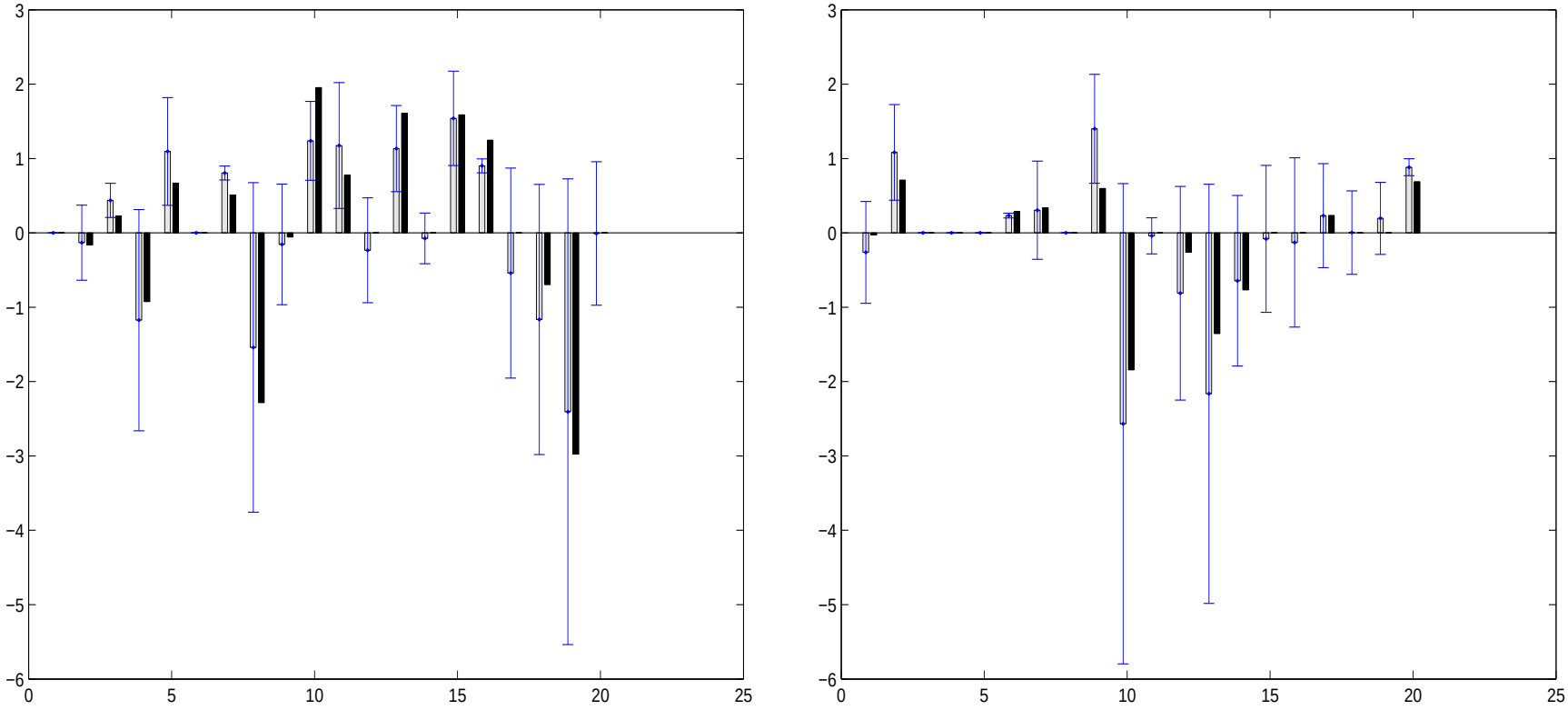


Figure: The inferred interaction weights W. (left) show the interaction weights between the first TF and the 20 genes. (right) show the corresponding weights for the second TF.

Spelman et. al. yeast data

We selected 30 genes regulated by 3 TFs. The 3 TFs are MBP1, FKH2 and STE12. The selection was done based on the ChiP data avalaible so that only the genes that are regulated exclusively by at least one of these 3 TFs were selected.

Yeast data

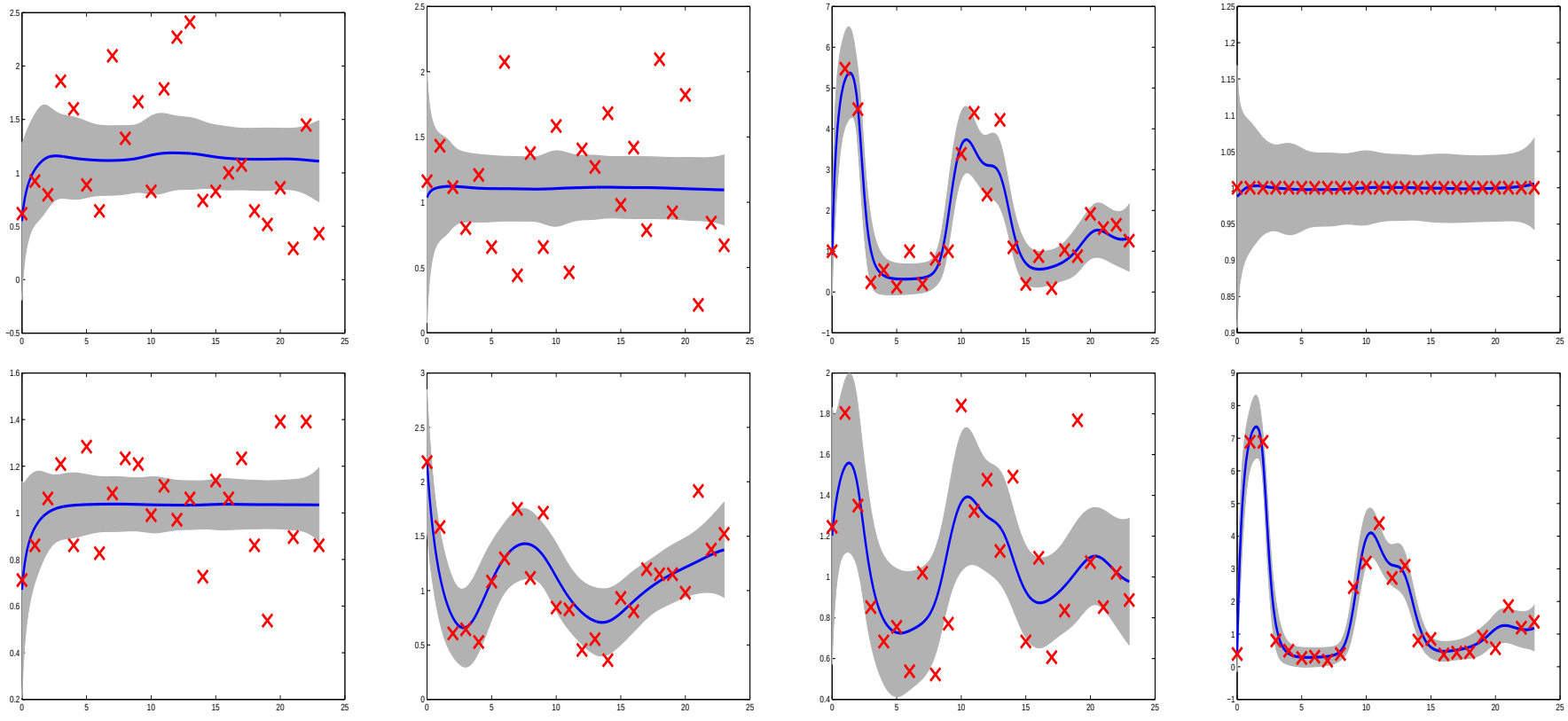


Figure: The predicted gene expressions. Red crosses represent the actual gene expression and the blue line with shaded error bars are the prediction found by MCMC.

Yeast data

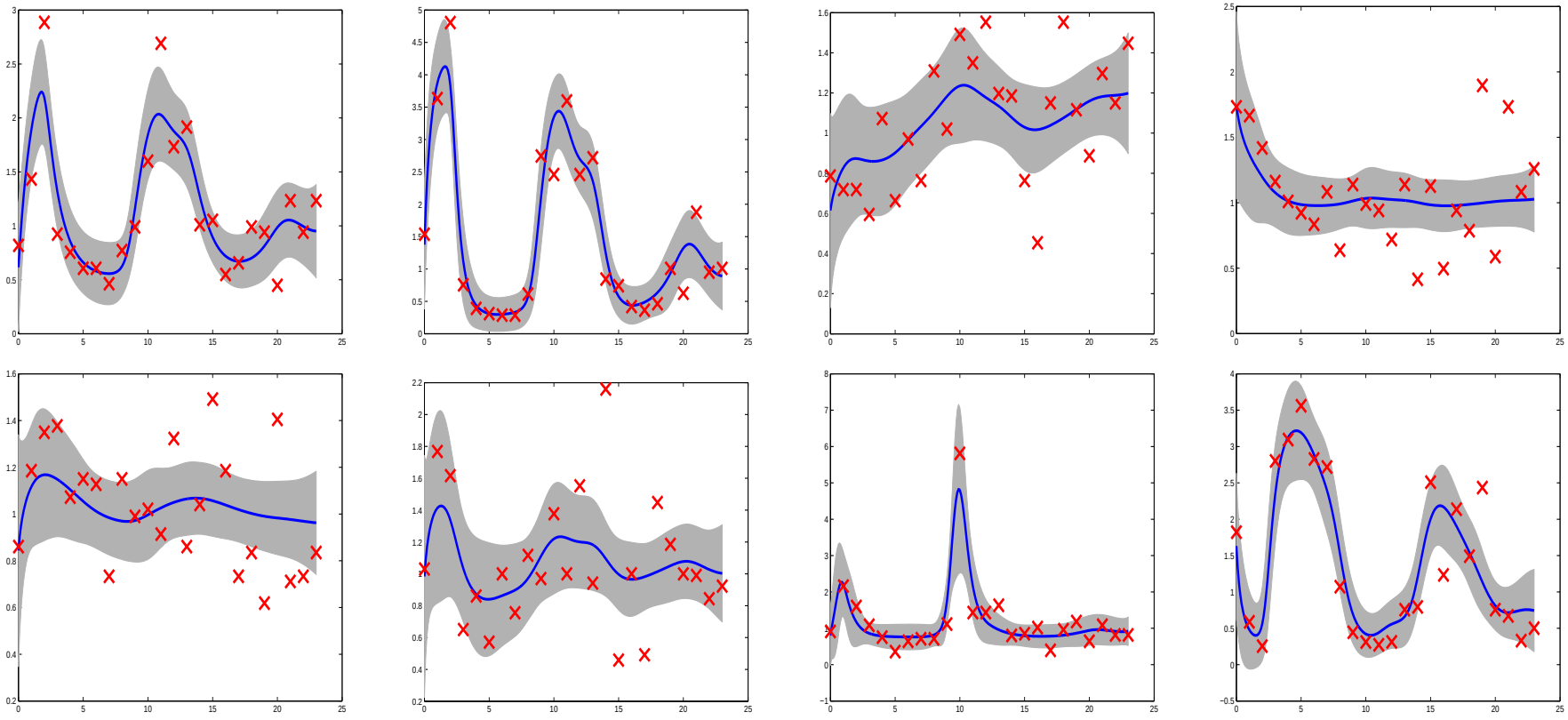


Figure: The predicted gene expressions. Red crosses represent the actual gene expression and the blue line with shaded error bars are the prediction found by MCMC.

Yeast data

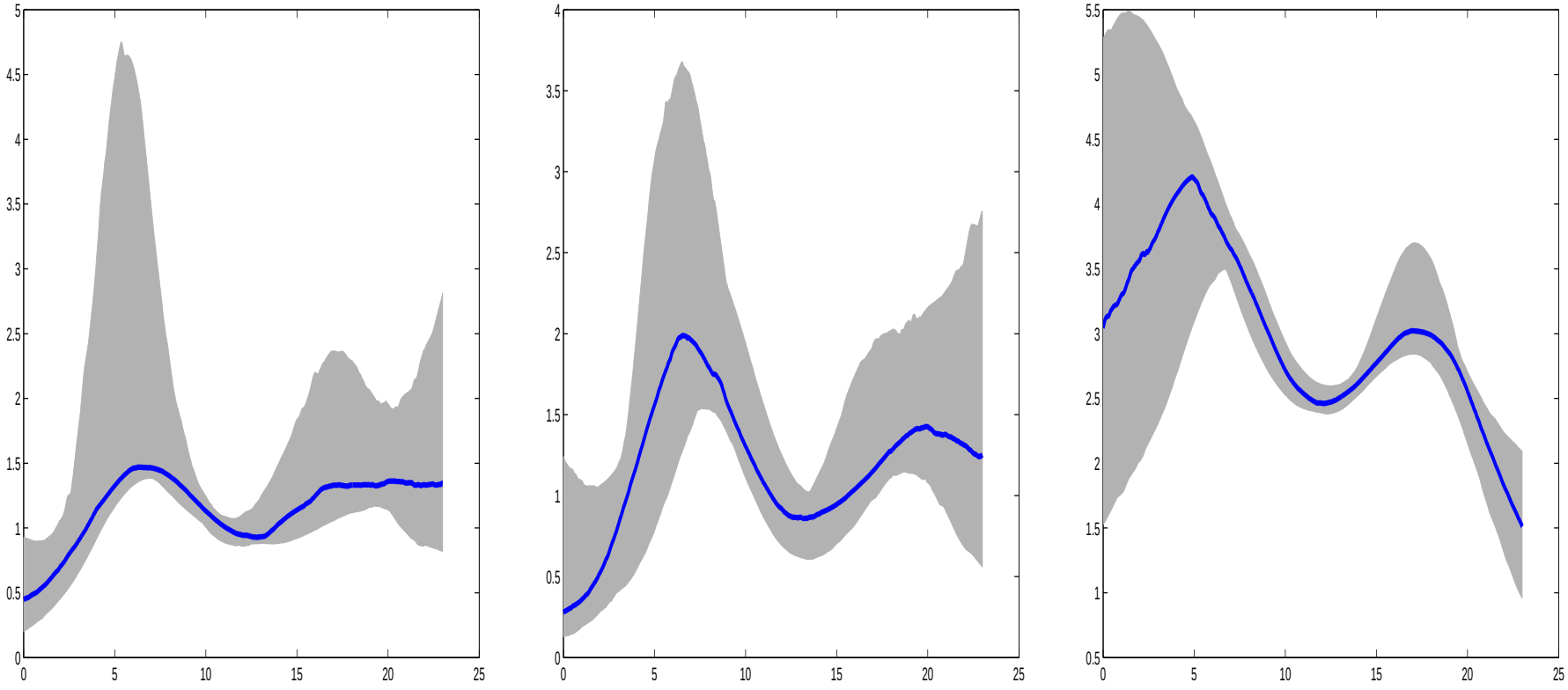


Figure: TF profiles

Sigmoid model

- The sigmoid model is perhaps less biologically plausible. Particularly it assumes that all TFs (activators and repressors) are combined by multiplication

$$\text{sigmoid} = \frac{1}{1 + \prod_{p=1} [\exp(f_p(t))]^{-w_{jp}} \exp(-w_{j0})}$$

  recall that $\exp(f_p(t))$ is the TF.
- This does not look so intuitive.
- Can we define activation functions where the combination is done by addition?
- Saturation and the ability of repressors to turn off the gene expression must be incorporated.
- Next we discuss such a model which can be viewed as a generalization of the Michaelis-Menten model for the single TF case.

Michaelis-Menten multiple TF model

$$\frac{\mathrm{d}x_j(t)}{\mathrm{d}t} = B_j + S_j g(f_1(t), \ldots, f_I(t); \mathbf{w}_j) - D_j x_j(t), \qquad (3)$$

- Let $\mathcal{P} = \{1, \ldots, P\}$ be the set of all TFs
- A_j be the set of TFs that are activators for jth gene and R_j the set of repressors.
- $A_j \cup R_j \subseteq \mathcal{P}$. That is some of the TFs may not regulate the jth gene
- The activation function takes the form

$$g = \frac{\sum_{i \in R_j} w_{ji} + \sum_{i \in A_j} w_{jp} \exp(f_i(t))}{1 + \sum_{i \in R_j} w_{ji} \exp(f_i(t)) + \sum_{i \in A_j} w_{ji} \exp(f_i(t))}$$

where w_{ji} are now non-negative and can be thought as relative sensitivities

Michaelis-Menten multiple TF model

$$g(f_1(t), \ldots, f_I(t); \mathbf{w}_j) = \frac{\sum_{i \in R_j} w_{ji} + \sum_{i \in A_j} w_{ji} \exp(f_i(t))}{1 + \sum_{i \in R_j} w_{ji} \exp(f_i(t)) + \sum_{i \in A_j} w_{ji} \exp(f_i(t))}$$

- Michaelis-Menten equation for a single TF can be obtained as a special case
  - Activation: $A_j = \{1\}$, $R_j = \emptyset$,

$$g(f_1(t); \mathbf{w}_j) = \frac{w_{j1} f_1(t)}{1 + w_{j1} f_1(t)} = \frac{f_1(t)}{\gamma_j + f_1(t)}$$

  - Repression: $A_j = \emptyset$, $R_j = \{1\}$

$$g(f_1(t); \mathbf{w}_j) = \frac{w_{j1}}{1 + w_{j1} f_1(t)} = \frac{1}{\gamma_j + f_1(t)}$$

    where $\gamma_j = \frac{1}{w_{j1}}$

MCMC

- Similar to the sigmoid model. But the set of the activators A_j and the set of repressors R_j are sampled based on Gibbs sampling by taking all possible combinations.

Artificial data

- We consider a set of 30 genes regulated by 3 TFs.
- Side information: We assume we know which TFs regulate each gene, but we do not know whether a TF activates or represses a certain gene
- We wish to estimate the TF profiles, kinetic parameters, etc
  - and to predict which TFs are activators and which are repressors for each gene

Artificial data

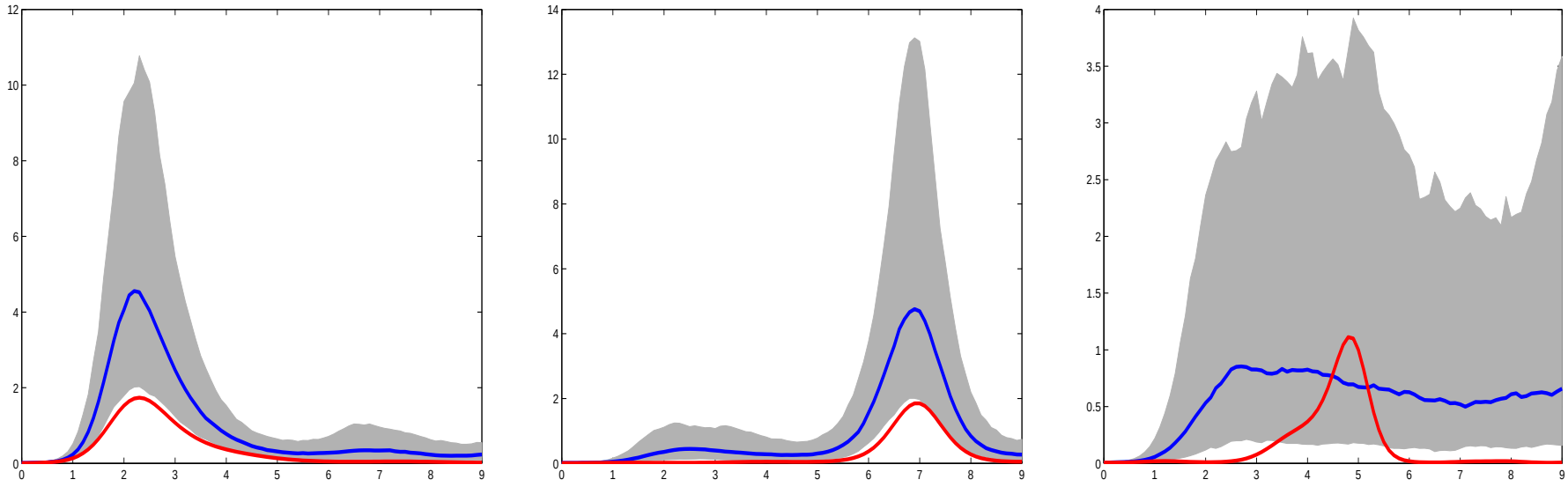


Figure: The inferred profiles of the three TFs. With red solid lines are the ground-truth TFs used to generate the toy data. With blue lines shaded error bars are the inferred TF profiles.

Artificial data

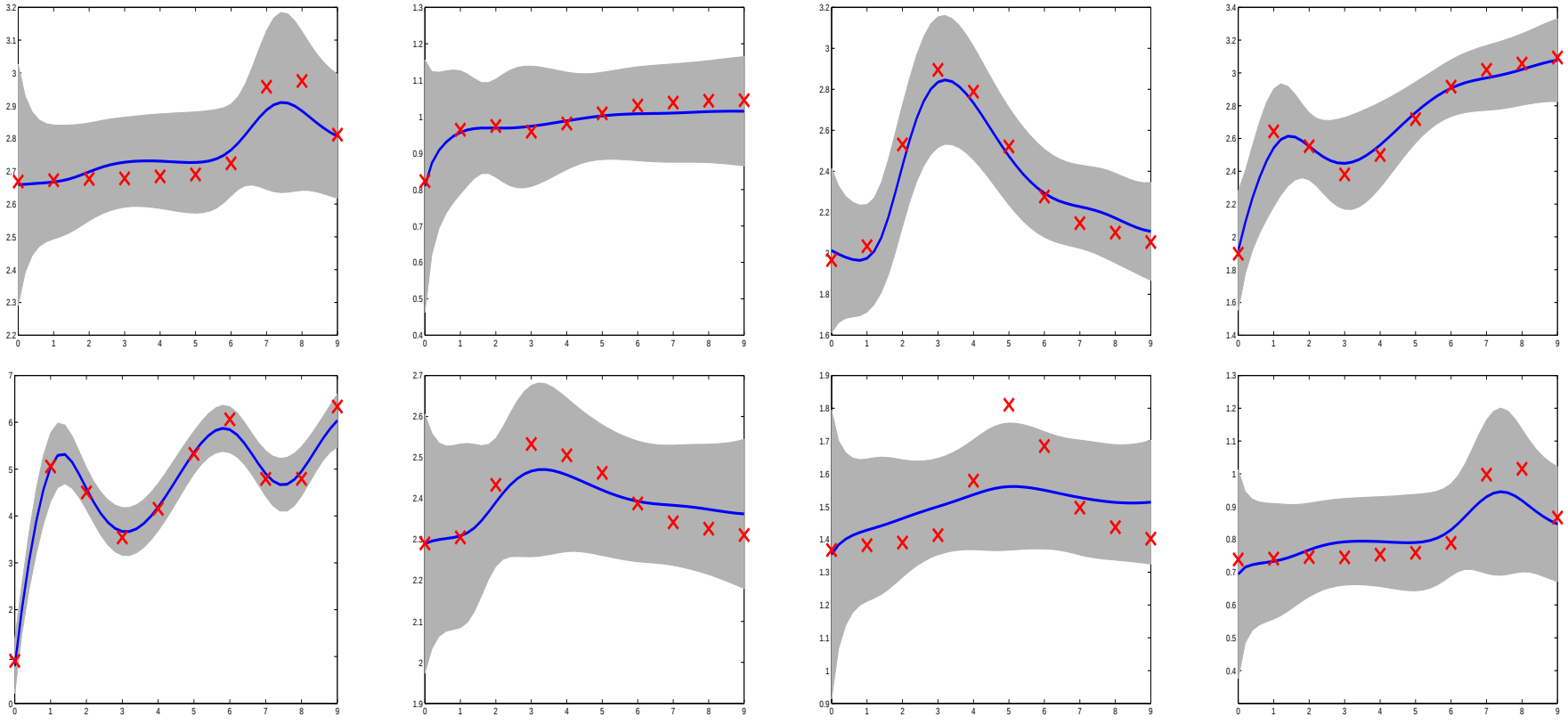


Figure: The predicted gene expressions. Red crosses represent the actual gene expression and the blue line with shaded error bars are the prediction found by MCMC.

Artificial data

Total classfication error regarding which TFs are activators and which are repressors for each gene

$$0.2447 \pm 0.0617$$

Outline

Discussion and Future Work

- Nonlinear response makes standard GP approach intractable.
- Approximate solutions include:
  - Laplace's approximation.
  - Sampling
- Sampling approach allows for multiple TFs.
- As system complexity increases identifiability drecreses.
- In principle this isn't a problem for Bayesian approaches
- In practice it can effect the mixing of the sampler.

Acknowledgements

- Investigators: Neil Lawrence and Magnus Rattray
- Researchers: **Pei Gao, Michalis Titsias**
- Martino Barenco and Mike Hubank at the Institute of Child Health in UCL (p53 pathway).
- Raya Khanin and Ernst Wit for assistance with SOS data..

Funded by the BBSRC award "Improved Processing of microarray data using probabilistic models" and EPSRC award "Gaussian Processes for Systems Identification with applications in Systems Biology"

References I

M. Barenco, D. Tomescu, D. Brewer, R. Callard, J. Stark, and M. Hubank. Ranked prediction of p53 targets using hidden variable dynamic modeling. *Genome Biology*, 7(3):R25, 2006.

R. T. Cirz, J. K. Chin, D. R. Andes, V. de CrÃ©cy-Lagard, W. A. Craig, and F. E. Romesberg. Inhibition of mutation and combating the evolution of antibiotic resistance. *PLoS Biology*, 3(6), 2005.

J. Courcelle, A. Khodursky, B. Peter, P. O. Brown, , and P. C. Hanawalt. Comparative gene expression profiles following UV exposure in wild-type and SOS-deficient *Escherichia coli*. *Genetics*, 158:41–64, 2001.

P. Gao, A. Honkela, M. Rattray, and N. D. Lawrence. Gaussian process modelling of latent chemical species: Applications to inferring transcription factor activities. *Bioinformatics*, 24:i70–i75, 2008. [PDF]. [DOI].

D. S. Goodsell. The molecular perspective: p53 tumor suppressor. *The Oncologist, Vol. 4, No. 2, 138-139, April 1999*, 4(2):138–139, 1999.

R. Khanin, V. Viciotti, and E. Wit. Reconstructing repressor protein levels from expression of gene targets in *E. Coli*. *Proc. Natl. Acad. Sci. USA*, 103(49):18592–18596, 2006. [DOI].

A. M. Lee, C. T. Ross, B.-B. Zeng, , and S. F. Singleton. A molecular target for suppression of the evolution of antibiotic resistance: Inhibition of the *Escherichia coli* RecA protein by N6-(1-Naphthyl)-ADP. *J. Med. Chem.*, 48(17), 2005.

E. Mjolsness, D. H. Sharp, and J. Reinitz. A connectionist model of development. *Journal of Theoretical Biology*, 152(4):429–453, 1991.

Outline

Sampling using control points

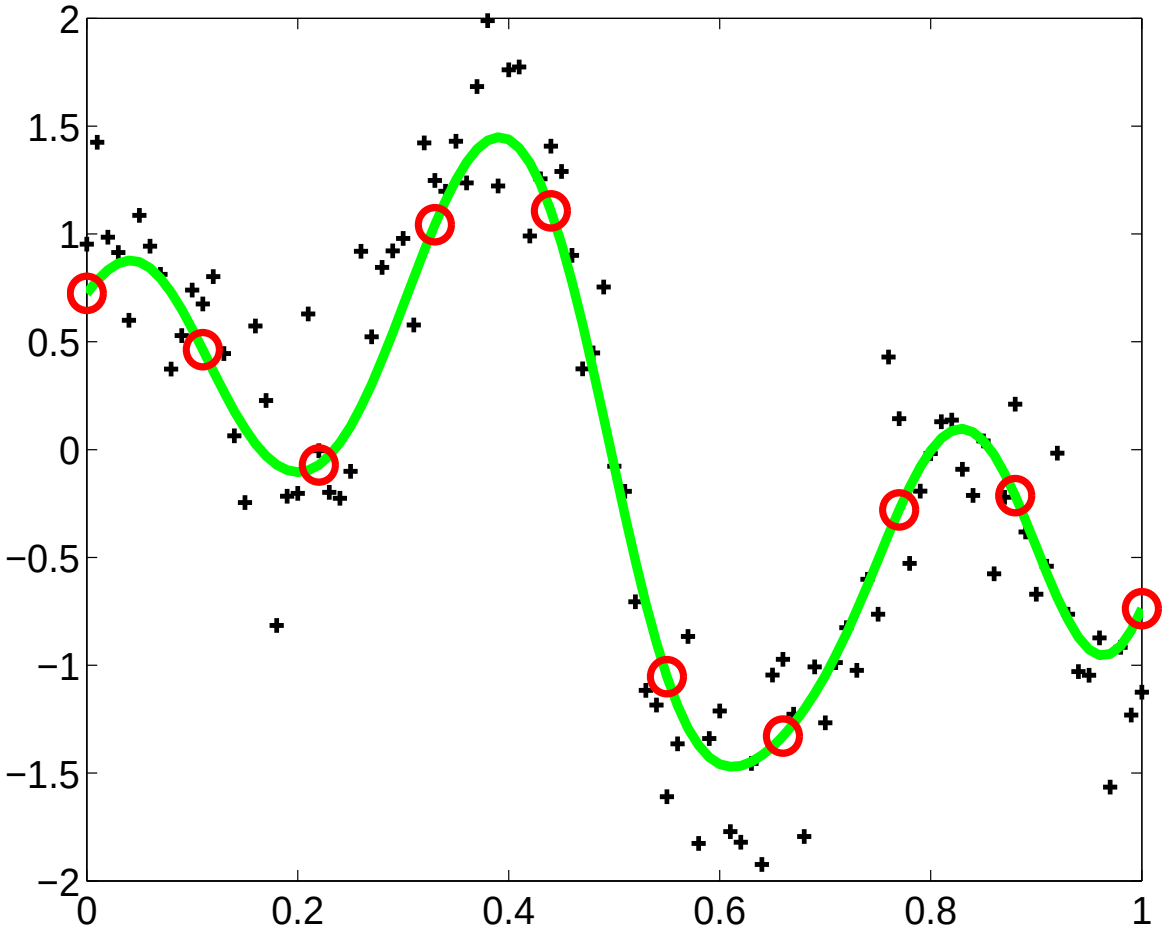
- Separate the points in $\mathbf{f}$ into two groups:
  - few control points $\mathbf{f}_c$
  - and the large majority of the remaining points $\mathbf{f}_\rho = \mathbf{f} \setminus \mathbf{f}_c$
- Sample the control points $\mathbf{f}_c$ using a proposal $q\left(\mathbf{f}_c^{(t+1)}|\mathbf{f}_c^{(t)}\right)$
- Sample the remaining points $\mathbf{f}_\rho$ using the conditional GP prior $p\left(\mathbf{f}_\rho^{(t+1)}|\mathbf{f}_c^{(t+1)}\right)$
- The whole proposal is

$$Q\left(\mathbf{f}^{(t+1)}|\mathbf{f}^{(t)}\right) = p\left(\mathbf{f}_\rho^{(t+1)}|\mathbf{f}_c^{(t+1)}\right) q\left(\mathbf{f}_c^{(t+1)}|\mathbf{f}_c^{(t)}\right)$$

- Its like sampling from the prior $p(\mathbf{f})$ but imposing random walk behaviour through the control points

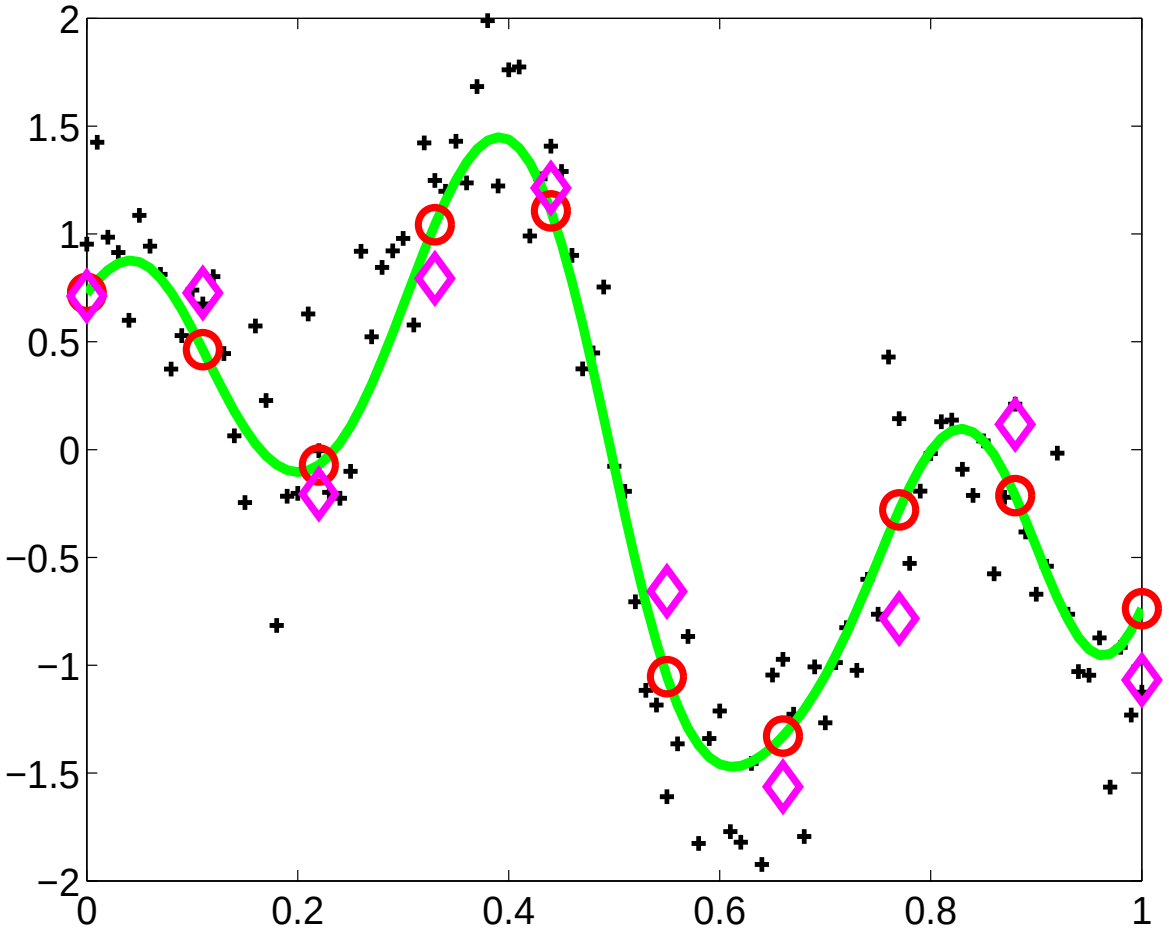
Sampling using control points: Regression-Examples

Sample 121 points using 10 control points



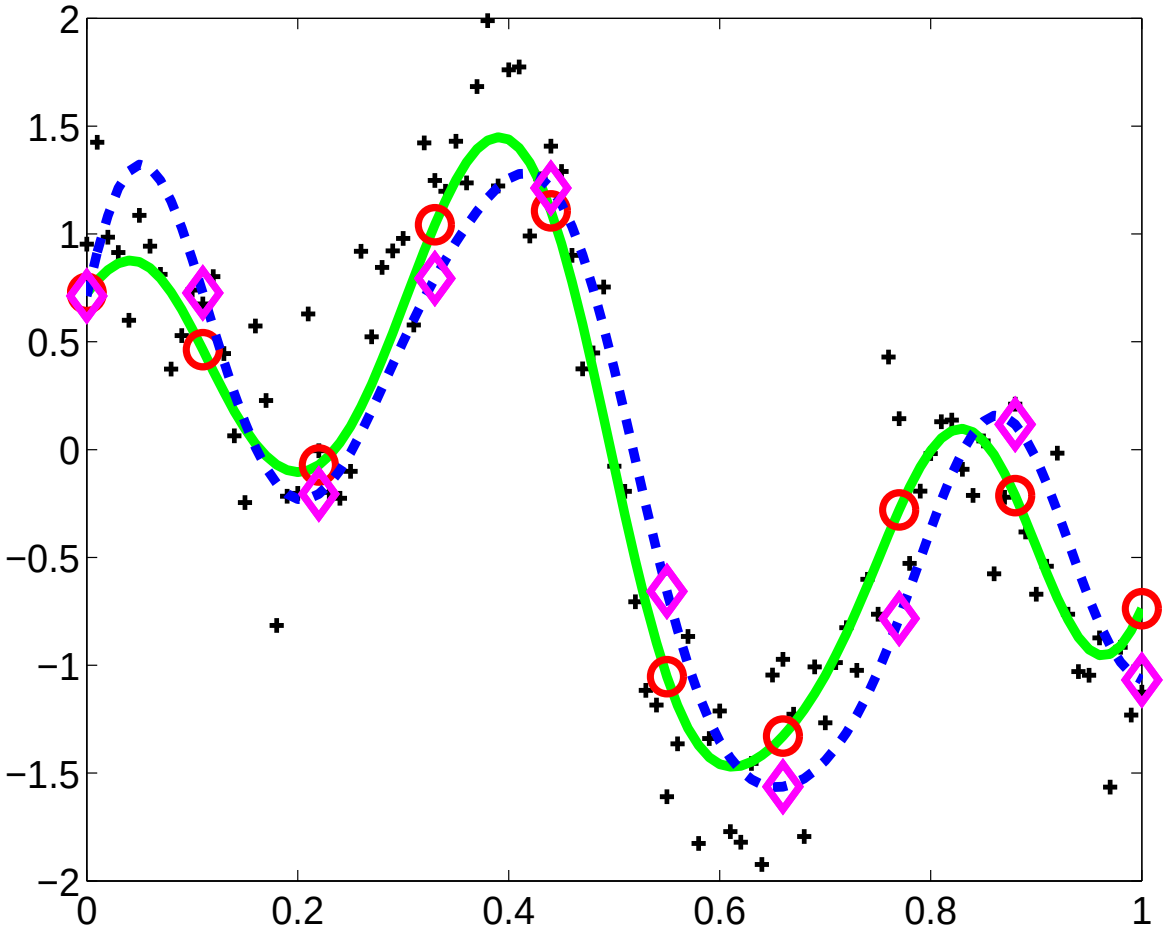
Sampling using control points: Regression-Examples

Sample 121 points using 10 control points



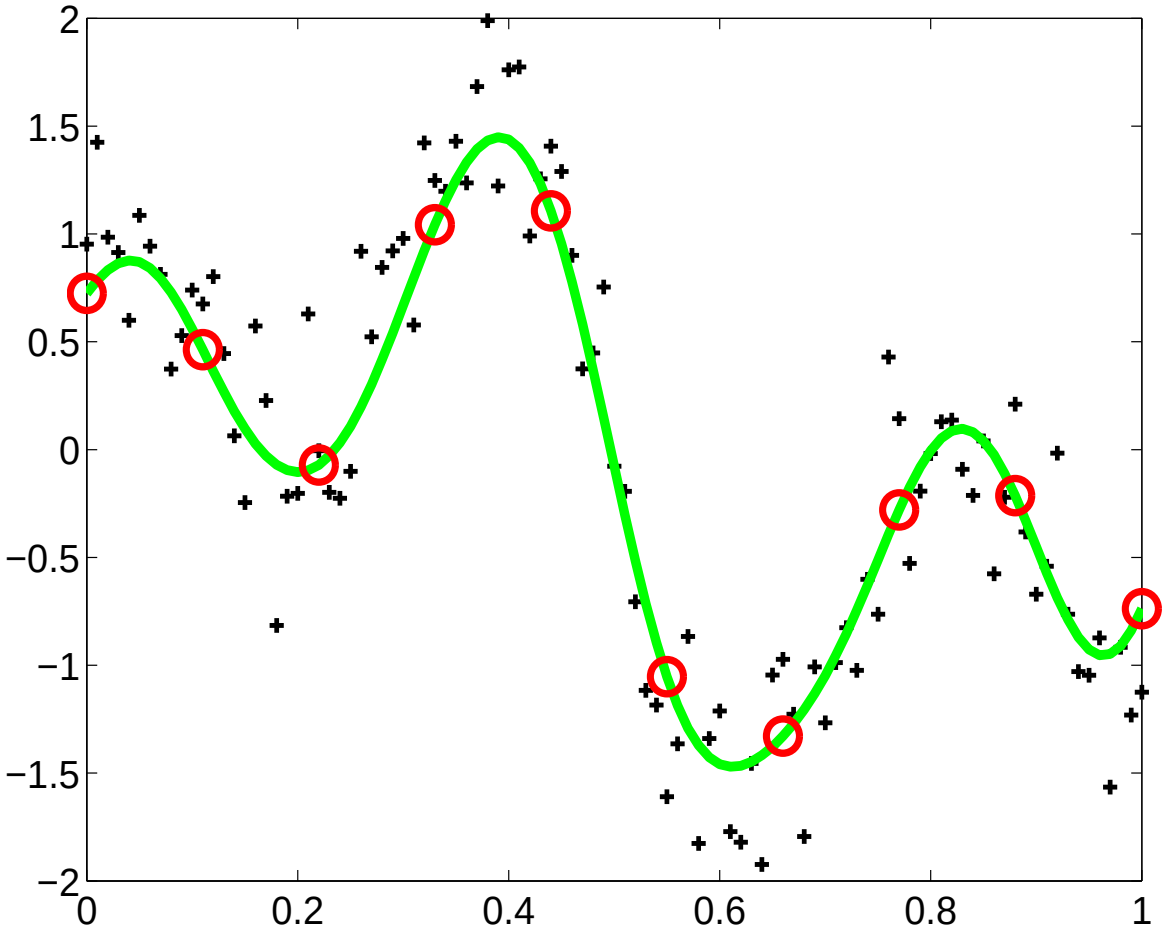
Sampling using control points: Regression-Examples

Sample 121 points using 10 control points



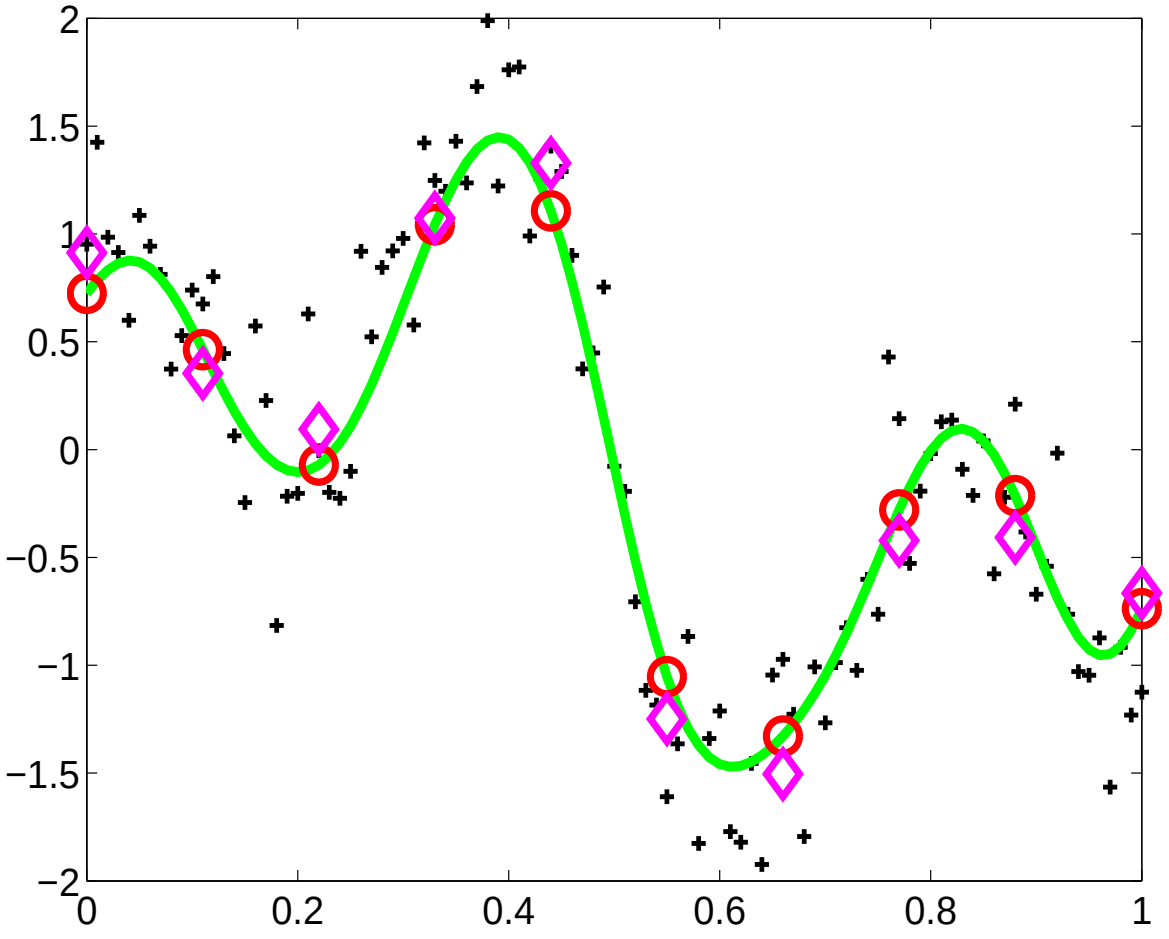
Sampling using control points: Regression-Examples

Sample 121 points using 10 control points



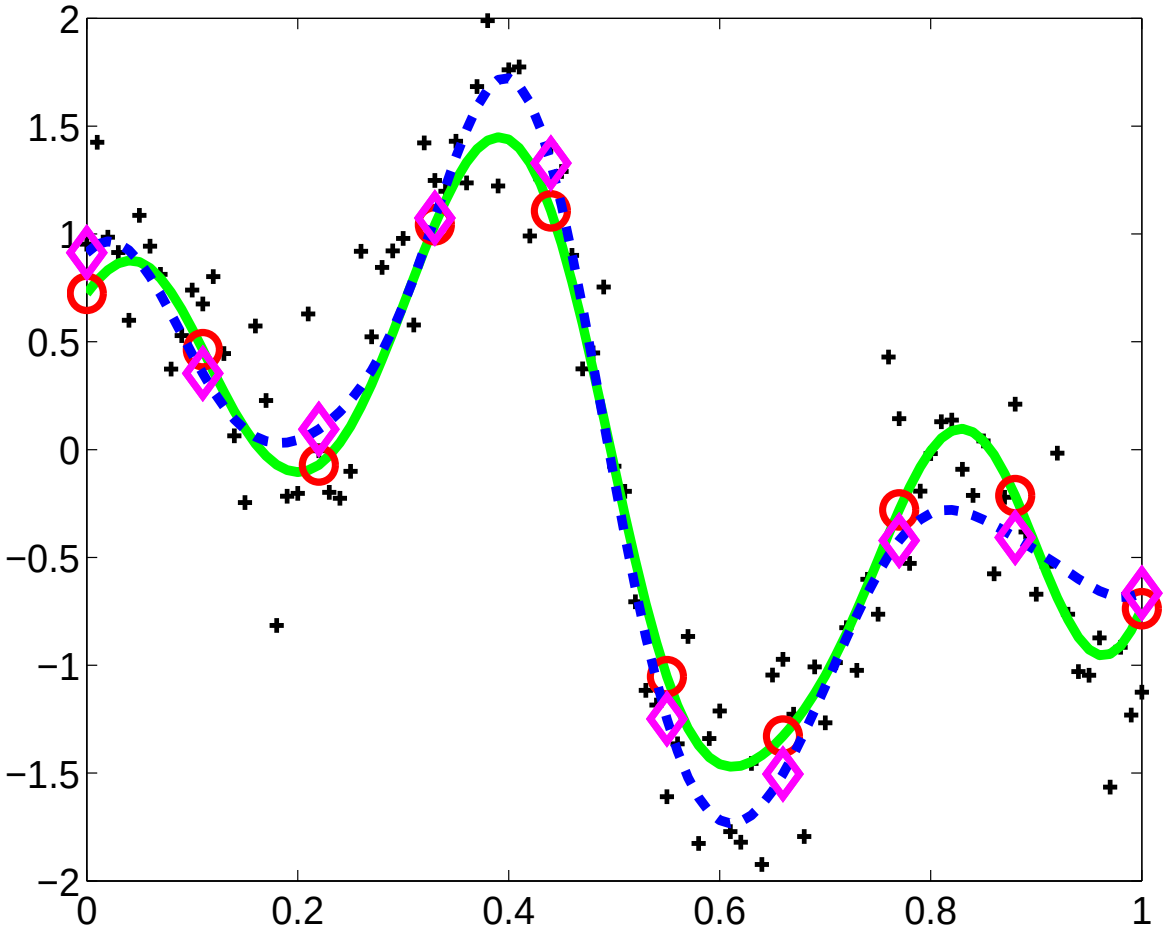
Sampling using control points: Regression-Examples

Sample 121 points using 10 control points



Sampling using control points: Regression-Examples

Sample 121 points using 10 control points



Sampling using control points

Few samples drawn during MCMC