GP Applications

Oliver Stegle and Karsten Borgwardt

Machine Learning and
Computational Biology Research Group,
Max Planck Institute for Biological Cybernetics and
Max Planck Institute for Developmental Biology, Tübingen

Outline

Outline

Motivation

- Human heart rate is an important physiological trait.
- Measurement over long periods only viable with poor sensors.
- Motivation Gaussian process model for heart rate.

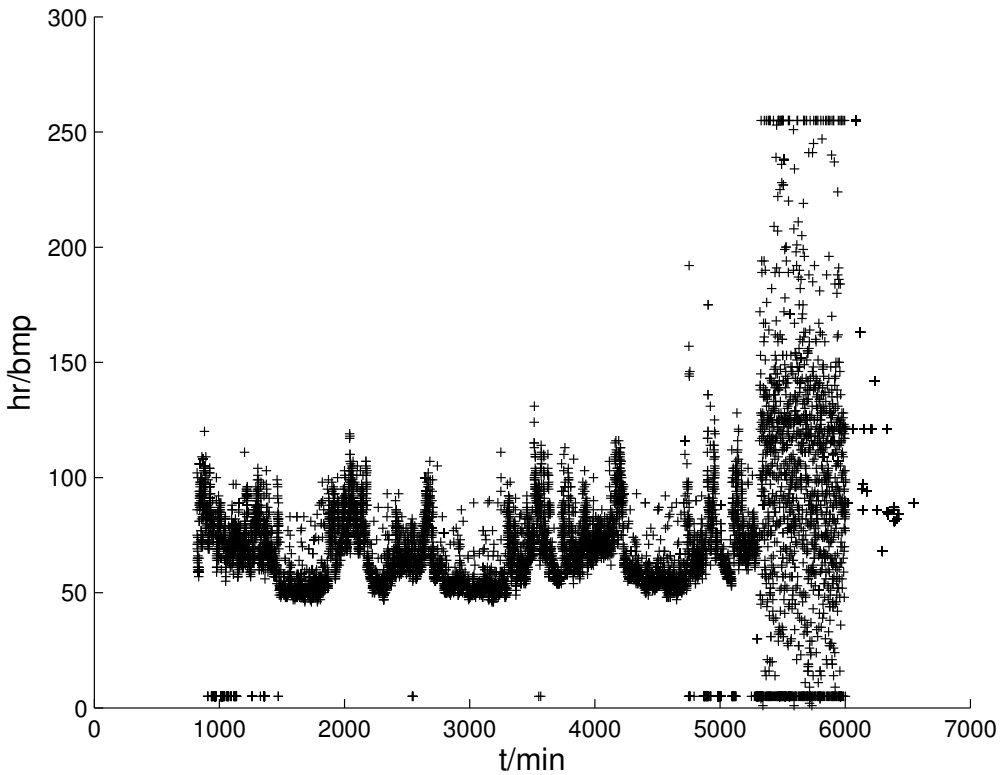
Motivation

- Human heart rate is an important physiological trait.
- Measurement over long periods only viable with poor sensors.
- Motivation Gaussian process model for heart rate.

The problem

Dataset

4 days of heart data



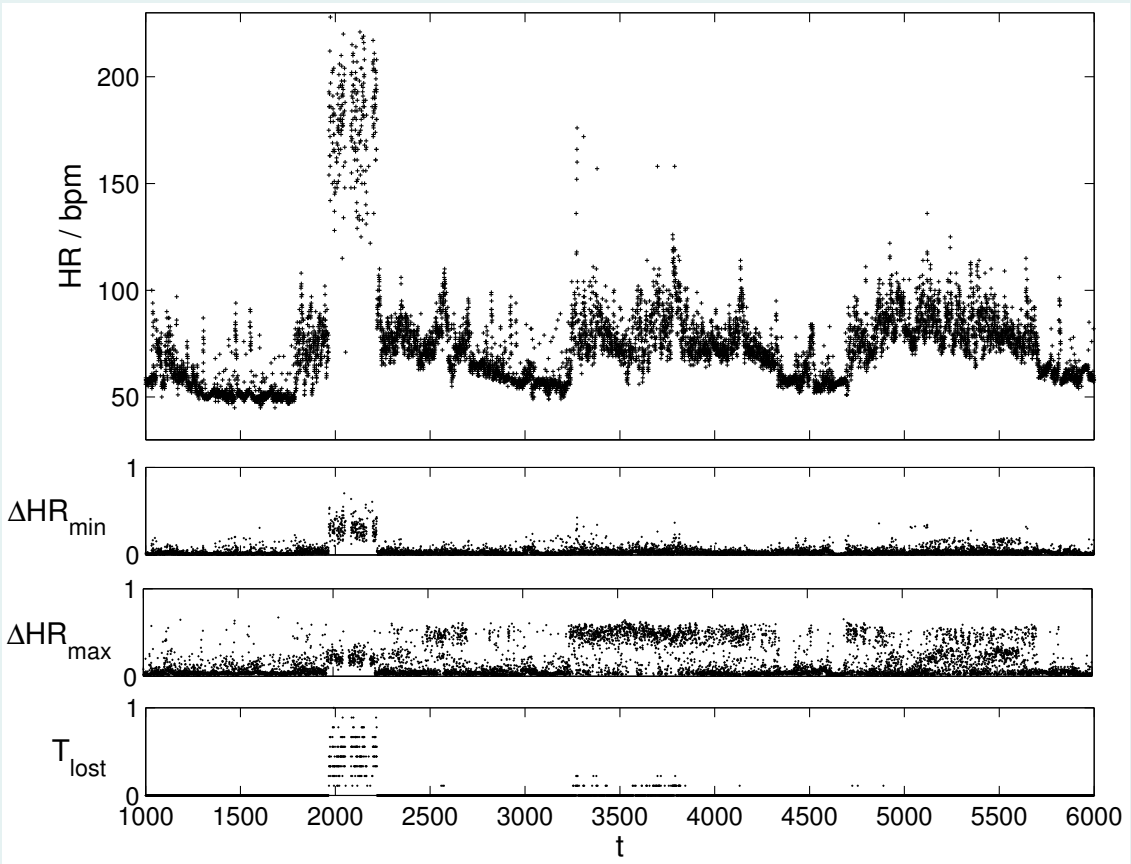
Features

- Different noise sources
- Two time scales, 24h rhythms
- Asymmetric around the mean
- Auxiliary variables indicative of noise

The problem

Dataset

4 days of heart data



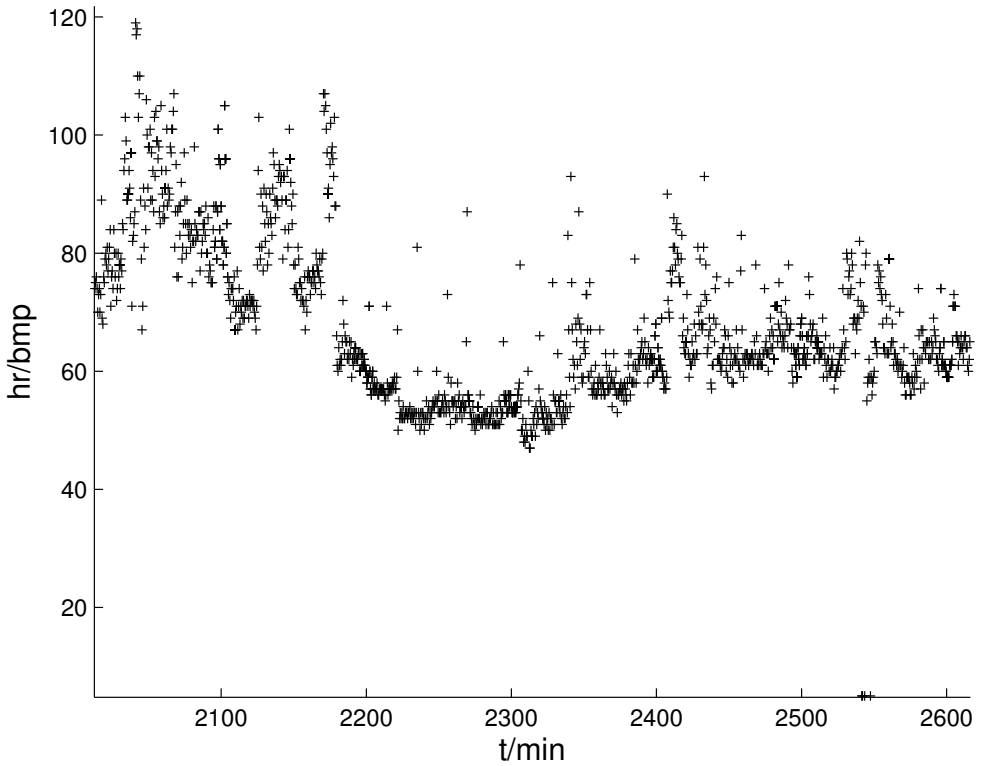
Features

- Different noise sources
- Two time scales, 24h rhythms
- Asymmetric around the mean
- Auxiliary variables indicative of noise

The problem

Dataset

zoomed view

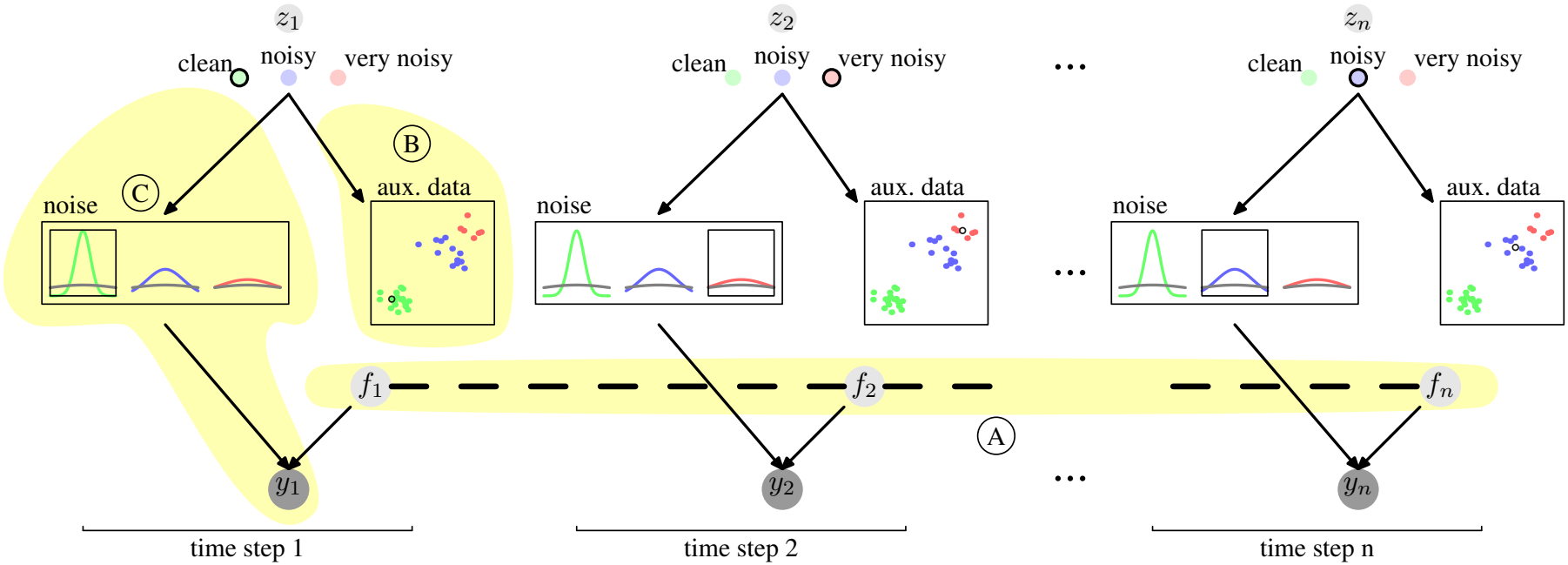


Features

- Different noise sources
- Two time scales, 24h rhythms
- Asymmetric around the mean
- Auxiliary variables indicative of noise

model

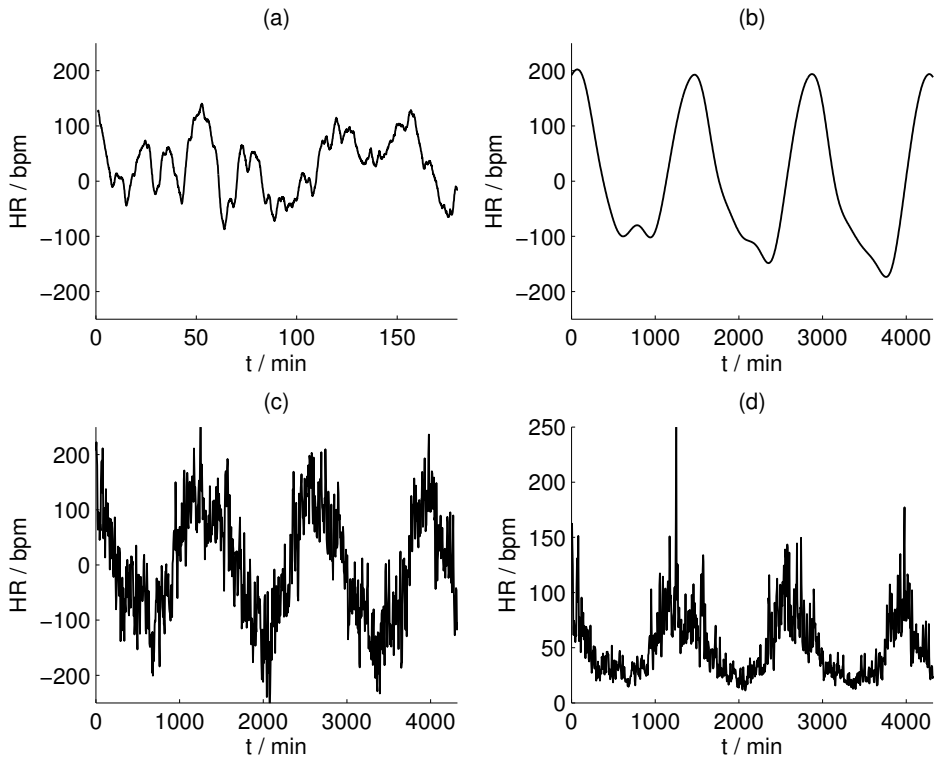
The Inference Model



- (a) Gaussian process prior on latent heart rate
- (b) Clustering of auxiliary data to extract noise classes
- (c) Heavy tailed noise model taking classes into account

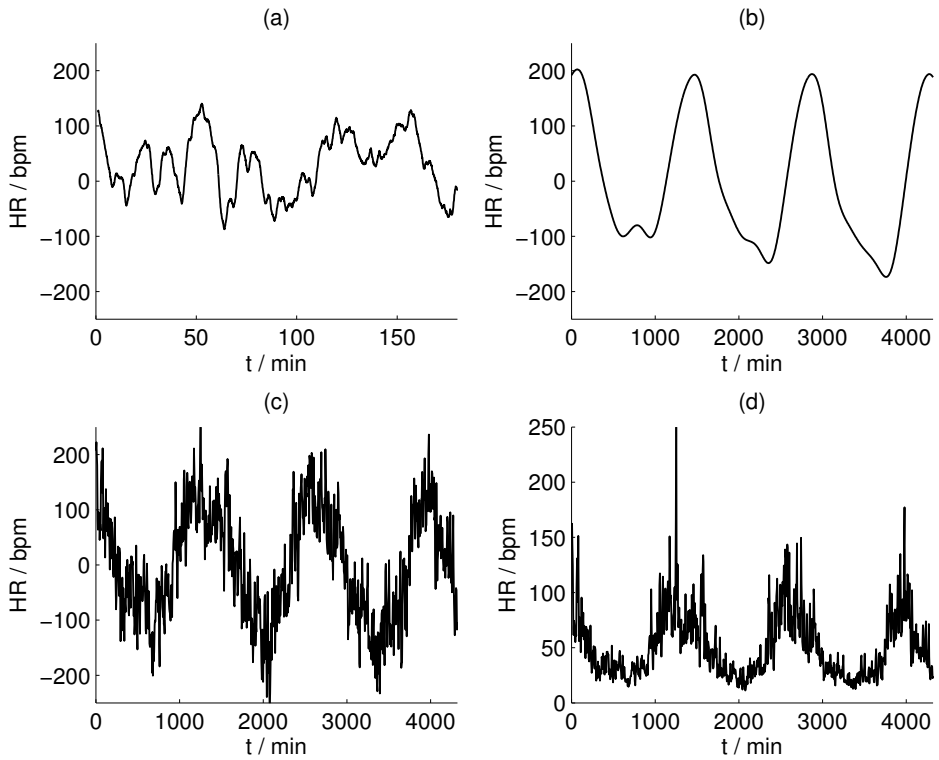
A GP prior for heart rate

- Covarianc function for short range fluctuations
- Long-range perdiodic signal
- Sum of (a) and (b)
- Log transformation (asymmetry)



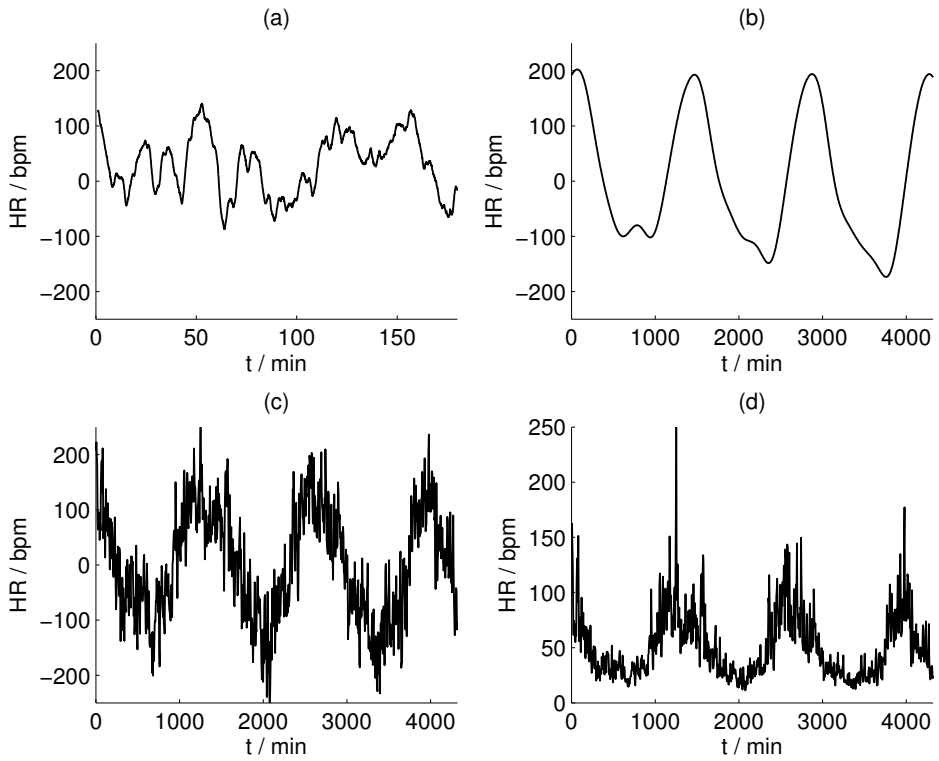
A GP prior for heart rate

- Covarianc function for short range fluctuations
- Long-range perdiodic signal
- Sum of (a) and (b)
- Log transformation (asymmetry)



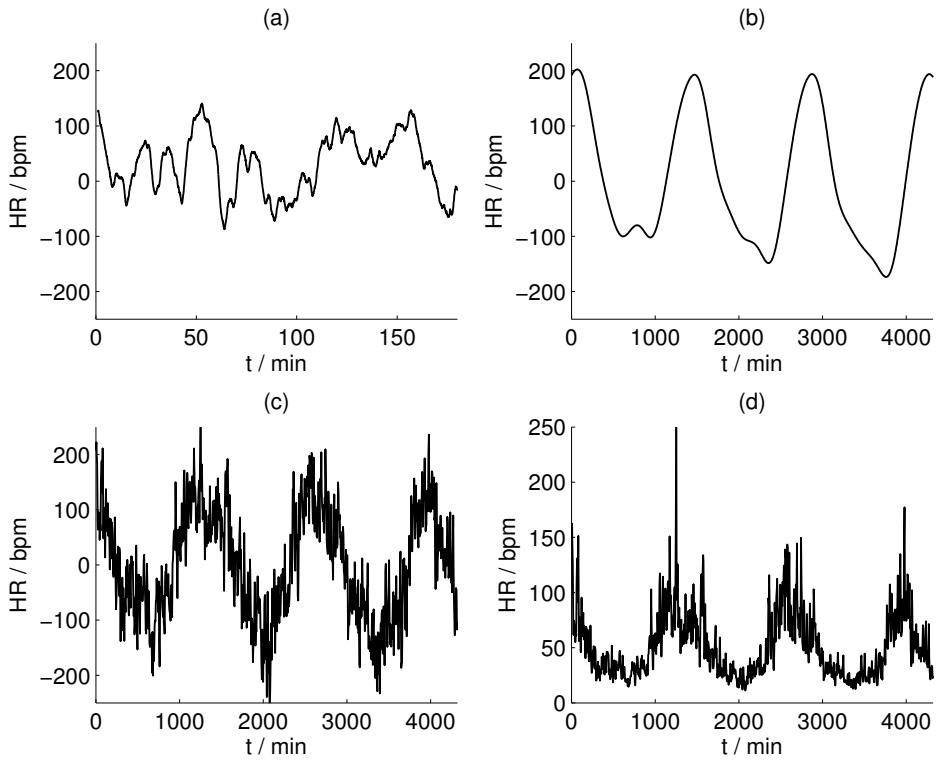
A GP prior for heart rate

- Covarianc function for short range fluctuations
- Long-range perdiodic signal
- Sum of (a) and (b)
- Log transformation (asymmetry)



A GP prior for heart rate

- Covarianc function for short range fluctuations
- Long-range perdiodic signal
- Sum of (a) and (b)
- Log transformation (asymmetry)



A GP prior for heart rate

- Short-range covariance

$$C_S(x, x') = C_1 \, \text{Matern}_{3/2}(x, x', \delta_S)$$

- Periodic covariance

$$C_L(x, x') = C_0 \, \exp\left[-\frac{(x - x')^2}{2\delta_L^2}\right] \exp\left[-2 \cdot \frac{\sin^2(\frac{2\pi}{p_L} \cdot (x - x'))}{A_L^2}\right]$$

- Total covariance

$$C(x, x') = C_L(x, x') + C_S(x, x')$$

- Non-linear transformation, reflecting strict positivity and asymmetry of heart rate

$$y_t^* = \log^\beta(y_t)$$

A GP prior for heart rate

- Short-range covariance

$$C_S(x, x') = C_1 \, \text{Matern}_{3/2}(x, x', \delta_S)$$

- Periodic covariance

$$C_L(x, x') = C_0 \, \exp\left[-\frac{(x - x')^2}{2\delta_L^2}\right] \exp\left[-2 \cdot \frac{\sin^2(\frac{2\pi}{p_L} \cdot (x - x'))}{A_L^2}\right]$$

- Total covariance

$$C(x, x') = C_L(x, x') + C_S(x, x')$$

- Non-linear transformation, reflecting strict positivity and asymmetry of heart rate

$$y_t^* = \log^\beta(y_t)$$

A GP prior for heart rate

- Short-range covariance

$$C_S(x, x') = C_1 \, \text{Matern}_{3/2}(x, x', \delta_S)$$

- Periodic covariance

$$C_L(x, x') = C_0 \, \exp\left[-\frac{(x - x')^2}{2\delta_L^2}\right] \exp\left[-2 \cdot \frac{\sin^2(\frac{2\pi}{p_L} \cdot (x - x'))}{A_L^2}\right]$$

- Total covariance

$$C(x, x') = C_L(x, x') + C_S(x, x')$$

- Non-linear transformation, reflecting strict positivity and asymmetry of heart rate

$$y_t^* = \log^\beta(y_t)$$

A GP prior for heart rate

- Short-range covariance

$$C_S(x, x') = C_1 \, \text{Matern}_{3/2}(x, x', \delta_S)$$

- Periodic covariance

$$C_L(x, x') = C_0 \, \exp\left[-\frac{(x - x')^2}{2\delta_L^2}\right] \exp\left[-2 \cdot \frac{\sin^2(\frac{2\pi}{p_L} \cdot (x - x'))}{A_L^2}\right]$$

- Total covariance

$$C(x, x') = C_L(x, x') + C_S(x, x')$$

- Non-linear transformation, reflecting strict positivity and asymmetry of heart rate

$$y_t^* = \log^\beta(y_t)$$

Clustering of auxiliary data

- Every datapoint can be member in one of K clusters.
- Use clustering approach to determine cluster membership

$$\boldsymbol{\pi}_n = \{\pi_{n,1}, \ldots, \pi_{n,K}\} \text{ where}$$

$$\sum_{k=1}^{K} \pi_{n,k} = 1$$

Clustering of auxiliary data

- Every datapoint can be member in one of K clusters.
- Use clustering approach to determine cluster membership

$$\boldsymbol{\pi}_n = \{\pi_{n,1}, \ldots, \pi_{n,K}\} \text{ where}$$

$$\sum_{k=1}^{K} \pi_{n,k} = 1$$

Robust noise model

- Remember: standard robust noise model, accounting for outliers

$$p(y_n \mid f_n) = \pi_{ok}\mathcal{N}\left(y_n \mid f_n, \sigma^2\right) + (1 - \pi_{ok})\mathcal{N}\left(y_n \mid f_n, \sigma^2_\infty\right)$$

- Here: clustering results (auxiliaries $\mathbf{S}$) encode useful information.

- Generalize likelihood K noise components, one for each cluster and use data-specific mixture probabilities $\boldsymbol{\pi}_n$

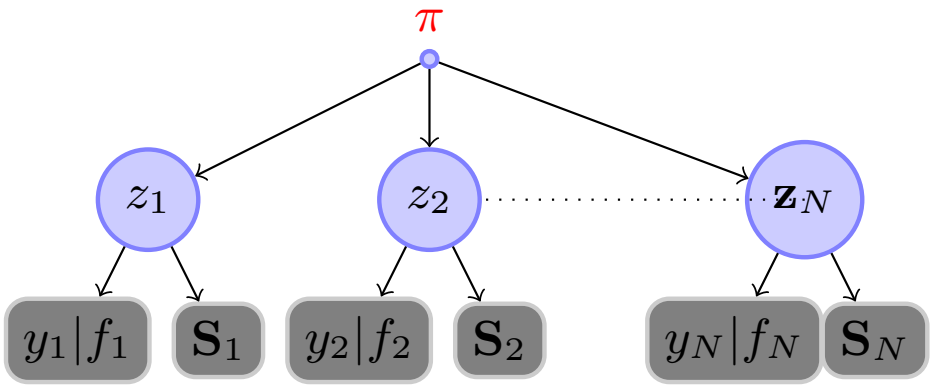
$$p(y_n \mid f_n) = \sum_{k=1}^{K} \pi_{k,n}\mathcal{N}\left(y_n \mid f_n, \sigma^2_k\right)$$

Robust noise model

- Remember: standard robust noise model, accounting for outliers

$$p(y_n \mid f_n) = \pi_{ok}\mathcal{N}\left(y_n \mid f_n, \sigma^2\right) + (1 - \pi_{ok})\mathcal{N}\left(y_n \mid f_n, \sigma^2_\infty\right)$$

- Here: clustering results (auxiliaries $\mathbf{S}$) encode useful information.



- Generalize likelihood K noise components, one for each cluster and use data-specific mixture probabilities $\boldsymbol{\pi}_n$

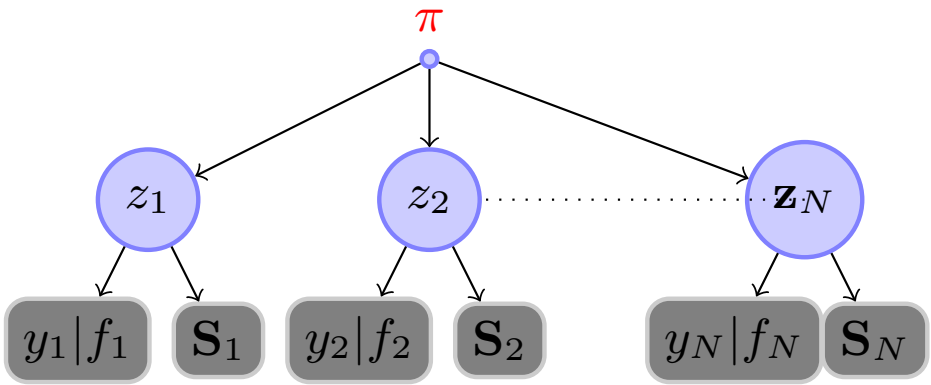
$$p(y_n \mid f_n) = \sum_{k=1}^{K} \pi_{k,n}\mathcal{N}\left(y_n \mid f_n, \sigma^2_k\right)$$

Robust noise model

- Remember: standard robust noise model, accounting for outliers

$$p(y_n \mid f_n) = \pi_{ok}\mathcal{N}\left(y_n \mid f_n, \sigma^2\right) + (1 - \pi_{ok})\mathcal{N}\left(y_n \mid f_n, \sigma^2_\infty\right)$$

- Here: clustering results (auxiliaries $\mathbf{S}$) encode useful information.

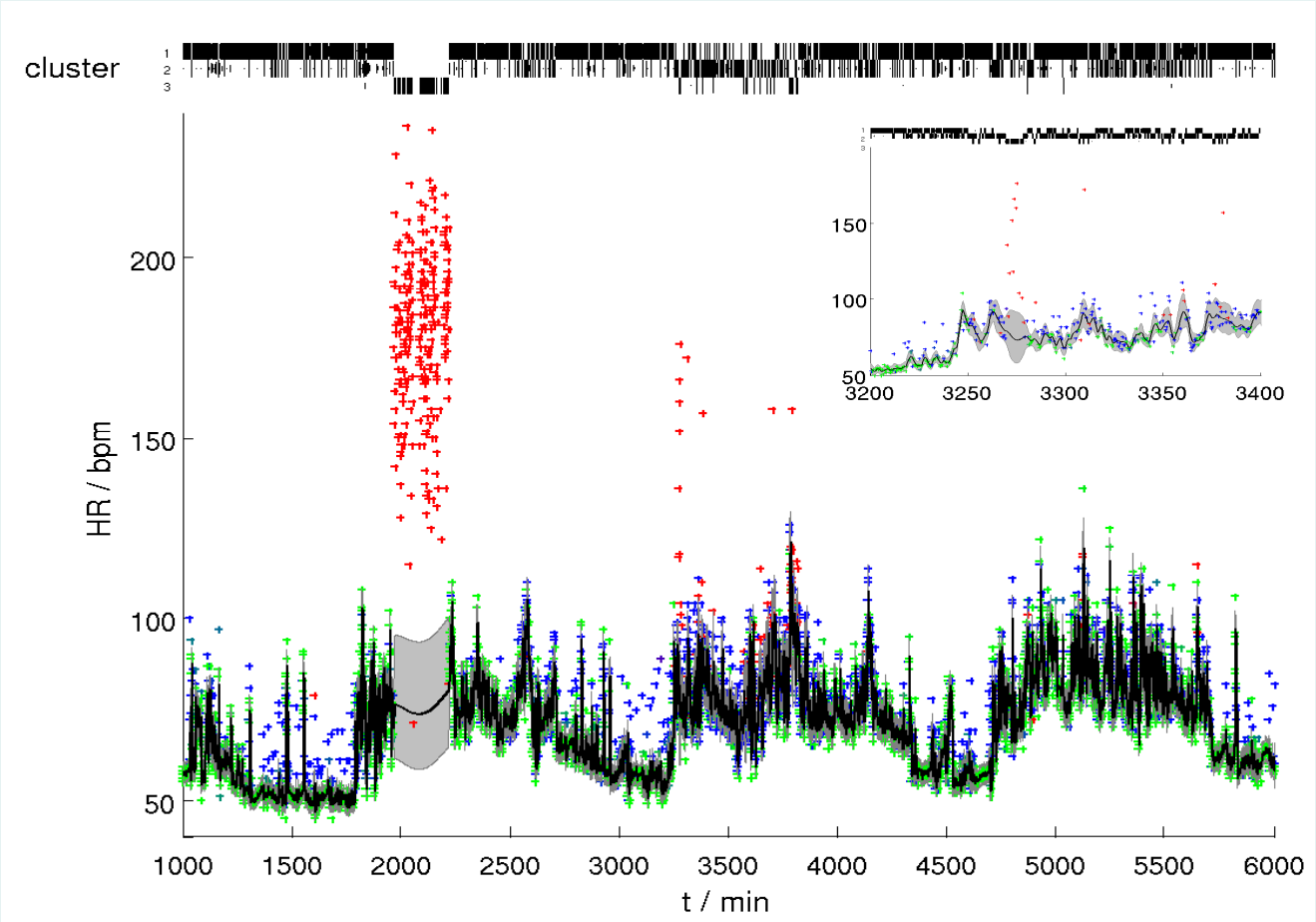


- Generalize likelihood K noise components, one for each cluster and use data-specific mixture probabilities $\boldsymbol{\pi}_n$

$$p(y_n \mid f_n) = \sum_{k=1}^{K} \pi_{k,n}\mathcal{N}\left(y_n \mid f_n, \sigma^2_k\right)$$

Regression for Heart Data

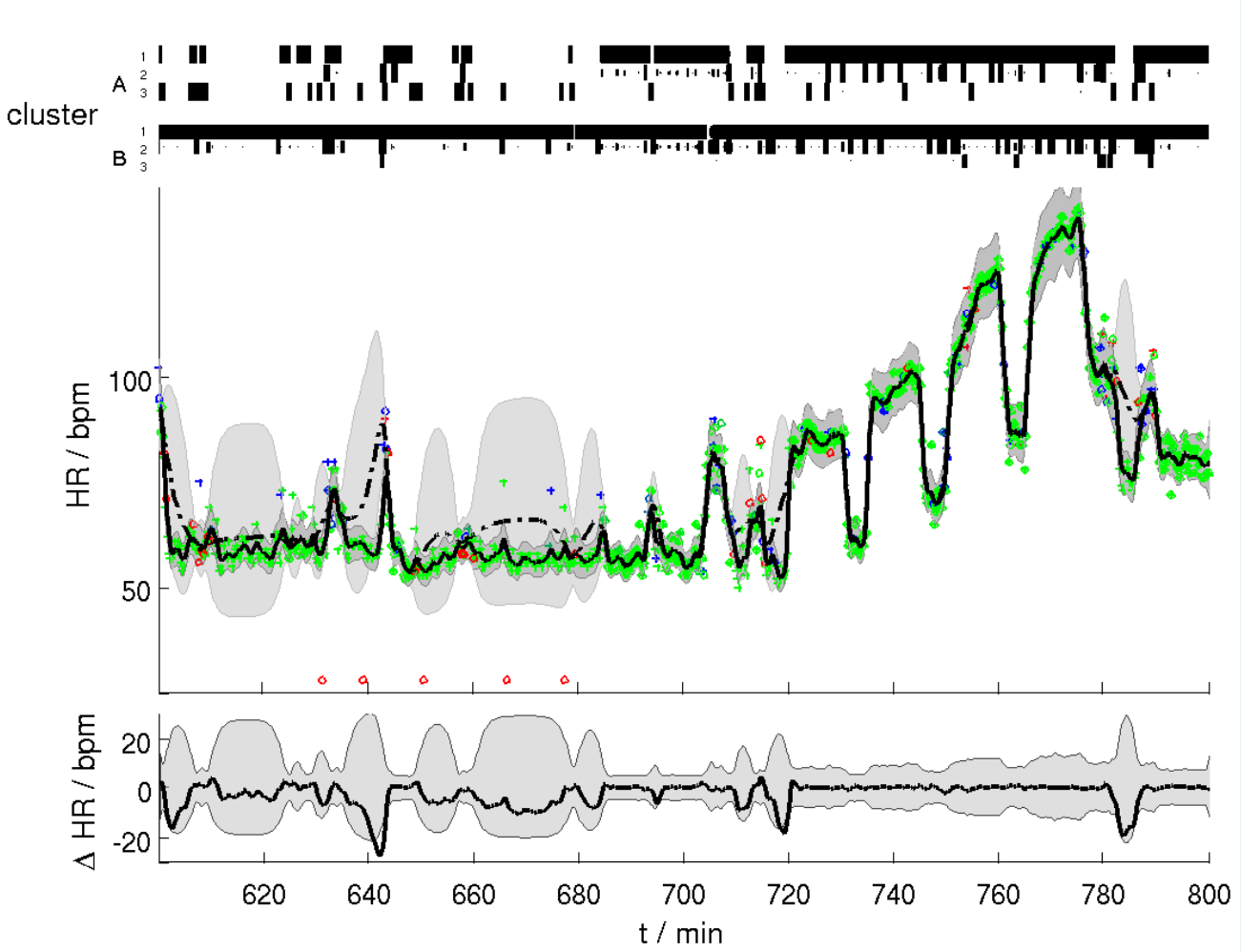
Single data set



- Clustering color-coded and Hinton diagrams
  - Three clusters - g,b,r-values for responsibilities
- Noise parameters $\{\sigma_c\}_{c=1}^{3}$ optimised along with other hyper parameters

Regression for Heart Data

Double data-set



- Two sensors for one heart
- GPs overlap well within error-bars
- Lower panel: difference plot and error bars

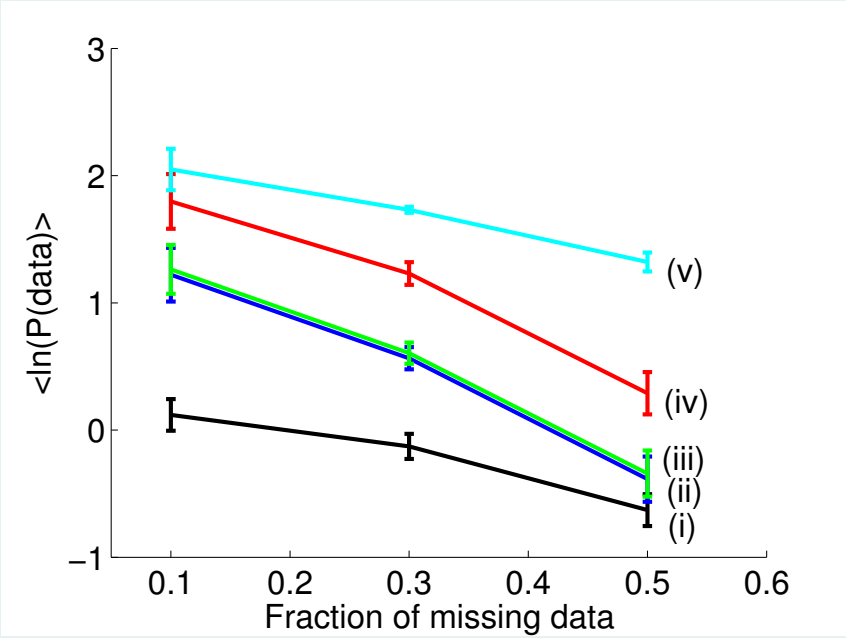
Fill-in test

Motivation

- Evaluate predictive performance to benchmark alternative models
- Fill-in test:
  - Model is trained on a subset of the data to predict the remainder.
  - Log probability as criteria rewards models with appropriately sized error bars.

Block size 1 minute

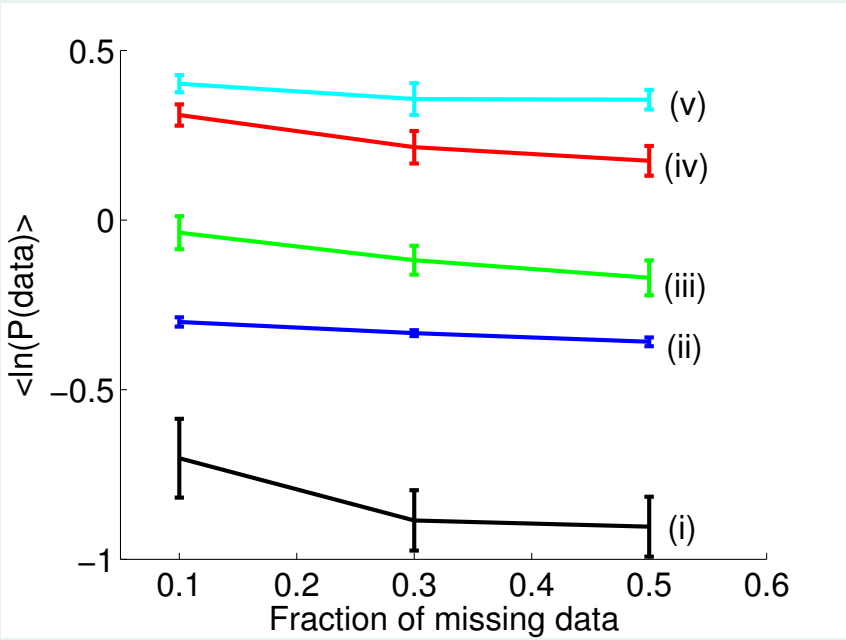
Block size 1 minute



- 5 different models compared
  - (i) baseline
  - (ii) GP with ks
  - (iii) GP with ks + kl
  - (iv) as (iii) but with Tlost threshold noise model
  - (v) as (iv) but with full robust noise model

Block size 60 minute

Block size 60 minutes



- 5 different models compared
  - (i) baseline
  - (ii) GP with ks
  - (iii) GP with ks + kl
  - (iv) as (iii) but with Tlost threshold noise model
  - (v) as (iv) but with full robust noise model
- Larger blocks removed rewards long range covariance

Outline

Outline

Response to External Stimuli

- Organisms such as plants respond to external stimuli:
  - Heat/Cold
  - Starvation
  - **Biotic stresses (fungus)**
  - ...
- Changes in observed gene expression levels.

Response to External Stimuli

- Organisms such as plants respond to external stimuli:
  - Heat/Cold
  - Starvation
  - Biotic stresses (fungus)
  - ...
- Changes in observed gene expression levels.

Response to External Stimuli

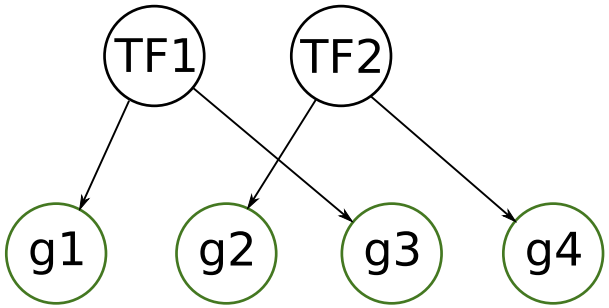
- Organisms such as plants respond to external stimuli:
  - Heat/Cold
  - Starvation
  - **Biotic stresses (fungus)**
  - ...
- Changes in observed gene expression levels.

Response to External Stimuli

- Organisms such as plants respond to external stimuli:
  - Heat/Cold
  - Starvation
  - **Biotic stresses (fungus)**
  - ...
- Changes in observed gene expression levels.

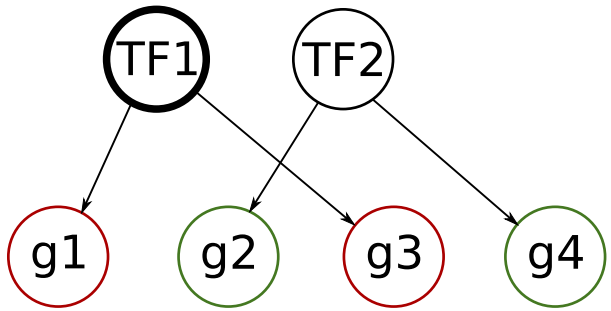
Differential Gene Expression

- An important goal is the identification of differentially expressed genes.
  - Identification of involved regulatory components.
  - Uncovering parts of the biological network.



Differential Gene Expression

- An important goal is the identification of differentially expressed genes.
  - Identification of involved regulatory components.
  - Uncovering parts of the biological network.



Differential Gene Expression in Time Series

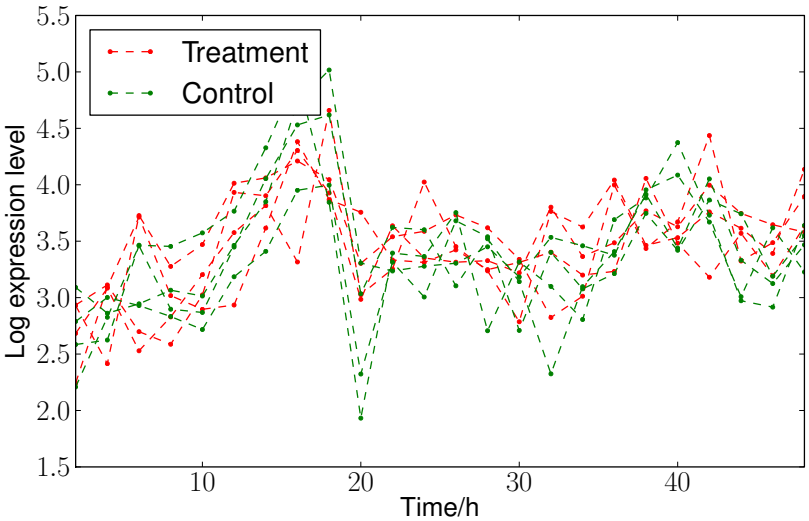
Time Series

- The response to external stimuli is a dynamic process.
- Hence the response should be studied as a function of time.

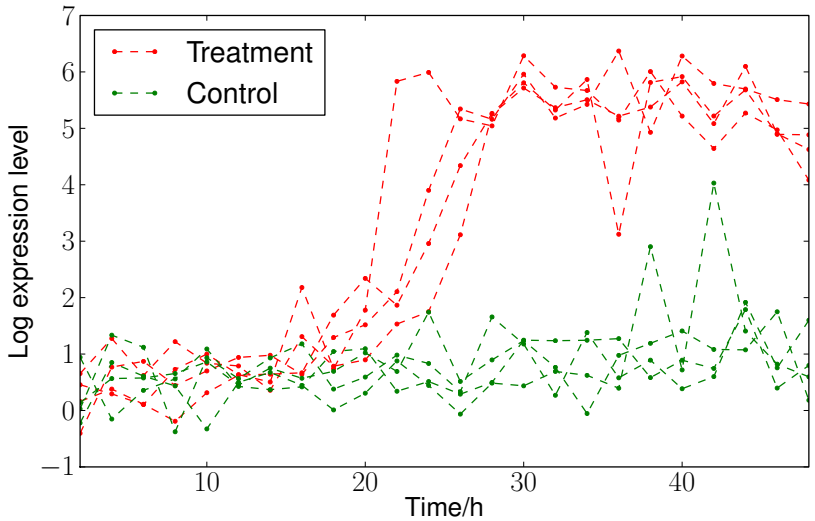
Differential Gene Expression in Time Series

Time Series

- The response to external stimuli is a dynamic process.
- Hence the response should be studied as a function of time.



Non-differentially expressed

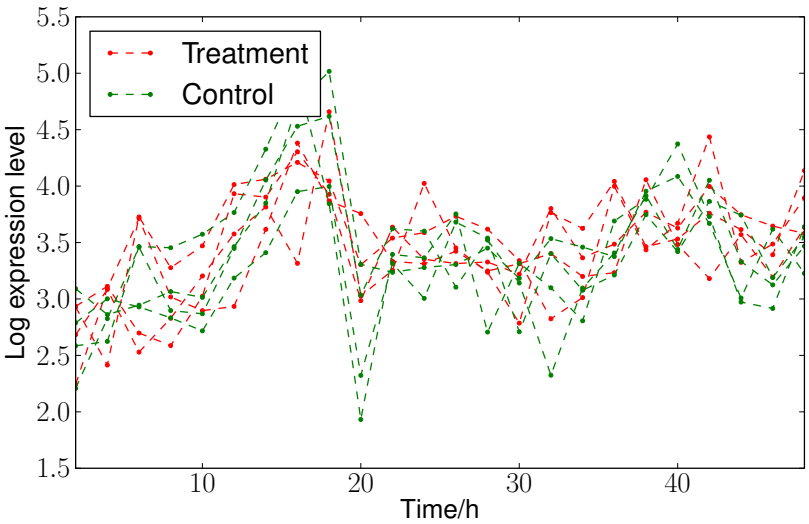


Differentially expressed

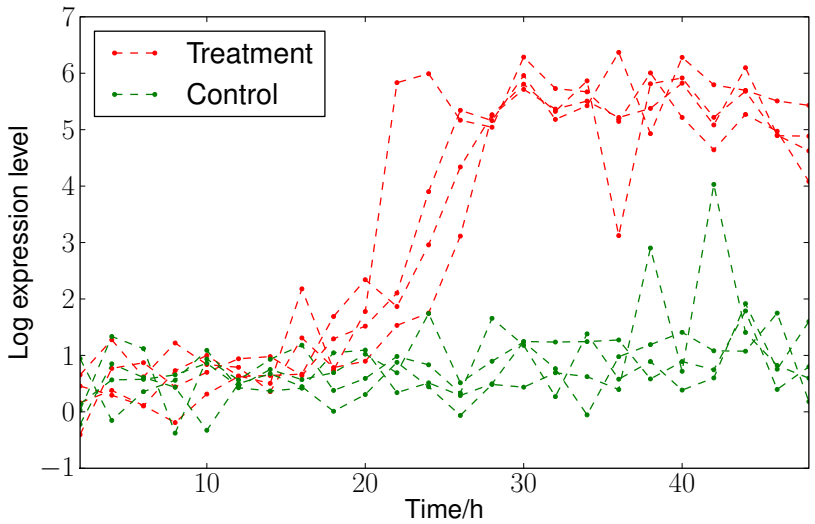
Differential Gene Expression in Time Series

Challenges

- Time series expression profiles vary smoothly over time.
- Noisy observations – *outliers*.
- Multiple replicates.
- Few observations.
- Temporal patterns (intervals) of differential gene expression.



Non-differentially expressed



Differentially expressed

Gaussian process Model

Model comparison

- The basic idea – a comparison of two models:
  - The *shared* model: Expression levels are explained by a single process.
  - The *independent* model: Expression levels are explained by two separate processes.

Gaussian process Model

Model comparison

- The basic idea – a comparison of two models:
  - The *shared* model: Expression levels are explained by a single process.
  - The *independent* model: Expression levels are explained by two separate processes.

Gaussian process Model

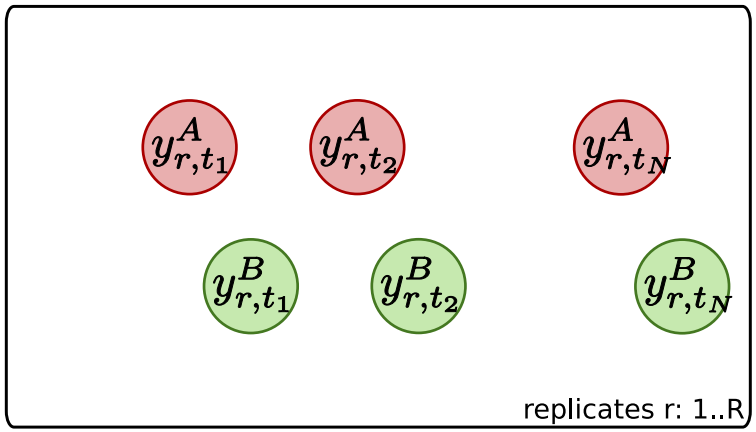
Model comparison

- The basic idea – a comparison of two models:
  - The *shared* model: Expression levels are explained by a single process.
  - The *independent* model: Expression levels are explained by two separate processes.

Gaussian process model

Bayesian Network: Shared Model

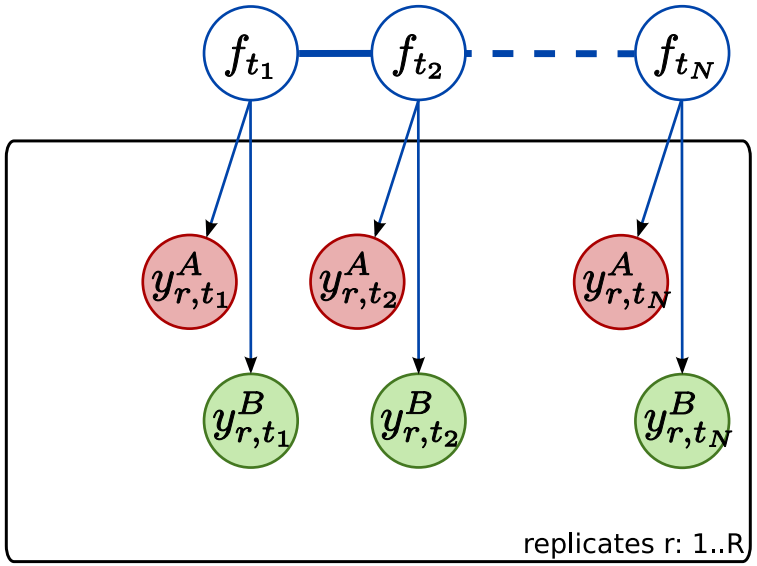
- Data in conditions A and B observed at N time points with R replicates.
- A Gaussian process prior incorporates beliefs about smoothness.
- Noise is is modeled separately per-replicate, $\sigma_r^{A/B}$.



Gaussian process model

Bayesian Network: Shared Model

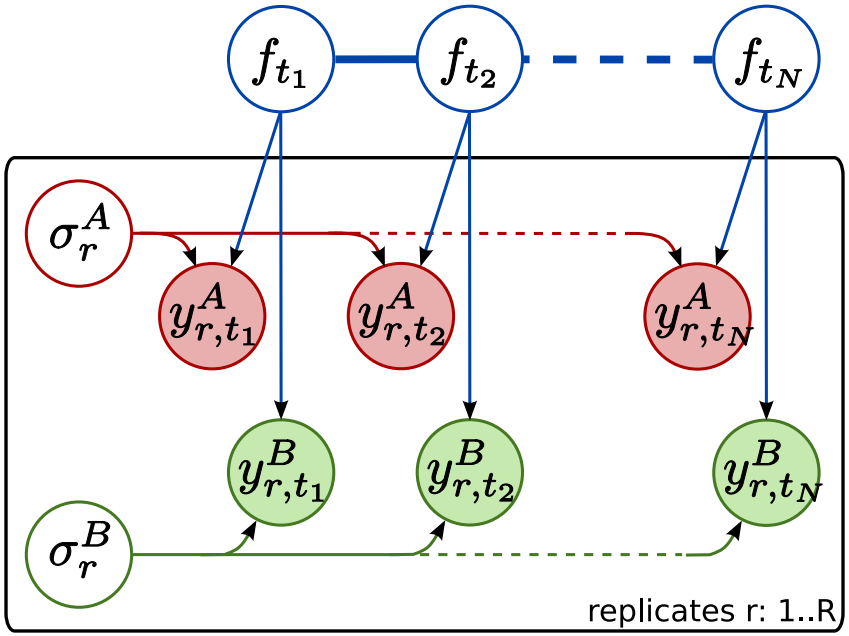
- Data in conditions A and B observed at N time points with R replicates.
- A Gaussian process prior incorporates beliefs about smoothness.
- Noise is is modeled separately per-replicate, $\sigma_r^{A/B}$.



Gaussian process model

Bayesian Network: Shared Model

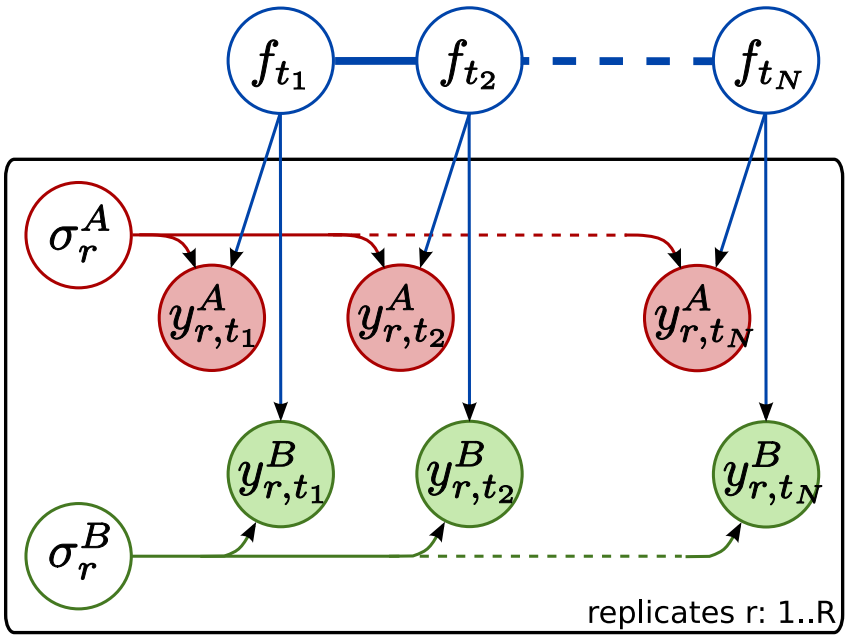
- Data in conditions A and B observed at N time points with R replicates.
- A Gaussian process prior incorporates beliefs about smoothness.
- Noise is is modeled separately per-replicate, $\sigma_r^{A/B}$.



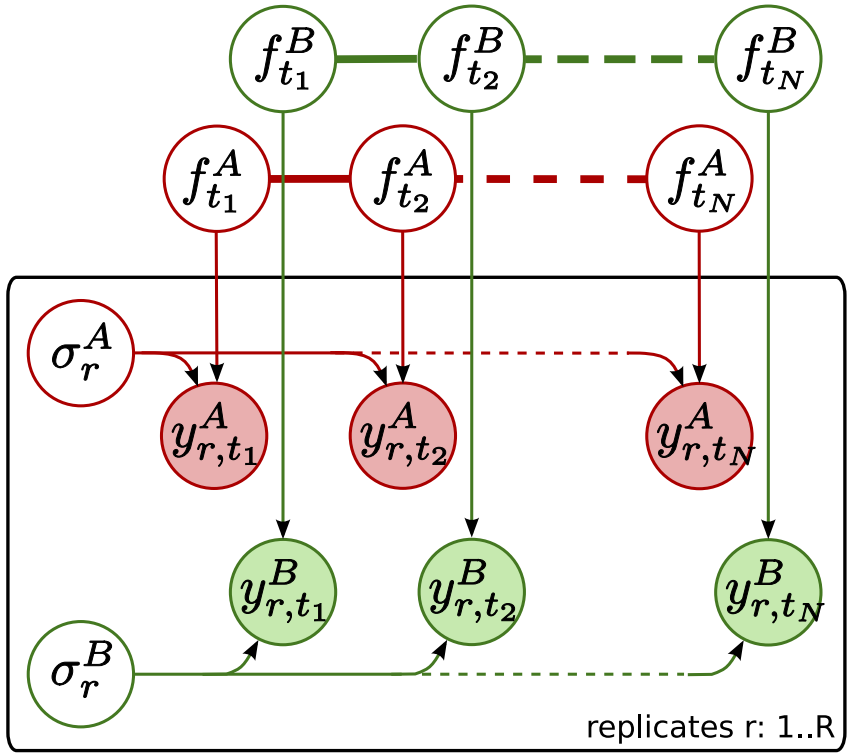
Gaussian process model

Bayesian Network: Both Models

- The independent model follows in an analogous manner.



Shared model $\mathcal{H}_S$



Independent model $\mathcal{H}_I$

Gaussian Process Model

Inference

- Models are compared using the Bayes factor

$$\text{Score} = \log \frac{\overbrace{P(\mathcal{D}_A, \mathcal{D}_B \,|\, \mathcal{H}_\mathsf{I})}^{\textit{Independent}\text{ model}}}{\underbrace{P(\mathcal{D}_A, \mathcal{D}_B \,|\, \mathcal{H}_\mathsf{S})}_{\textit{Shared}\text{ model}}}.$$

- Writing out the GP models explicitly leads to

$$\text{Score} = \log \frac{P(\mathbf{Y}^A \,|\, \mathcal{H}_{\mathsf{GP}}, \mathbf{T}^A,) P(\mathbf{Y}^B \,|\, \mathcal{H}_{\mathsf{GP}}, \mathbf{T}^B,)}{P(\mathbf{Y}^A \cup \mathbf{Y}^B \,|\, \mathcal{H}_{\mathsf{GP}}, \mathbf{T}^A \cup \mathbf{T}^B,)}.$$

($\mathbf{Y}^{A/B}$: expression levels in conditions A and B; $T^{A/B}$: observation time points)

Gaussian Process Model

Inference

- Models are compared using the Bayes factor

$$\text{Score} = \log \frac{\overbrace{P(\mathcal{D}_A, \mathcal{D}_B \,|\, \mathcal{H}_\mathsf{I})}^{\textit{Independent}\text{ model}}}{\underbrace{P(\mathcal{D}_A, \mathcal{D}_B \,|\, \mathcal{H}_\mathsf{S})}_{\textit{Shared}\text{ model}}}.$$

- Writing out the GP models explicitly leads to

$$\text{Score} = \log \frac{P(\mathbf{Y}^A \,|\, \mathcal{H}_\mathsf{GP}, \mathbf{T}^A, \boldsymbol{\theta}_\mathsf{I}) P(\mathbf{Y}^B \,|\, \mathcal{H}_\mathsf{GP}, \mathbf{T}^B, \boldsymbol{\theta}_\mathsf{I})}{P(\mathbf{Y}^A \cup \mathbf{Y}^B \,|\, \mathcal{H}_\mathsf{GP}, \mathbf{T}^A \cup \mathbf{T}^B, \boldsymbol{\theta}_\mathsf{S})}.$$

($\mathbf{Y}^{A/B}$: expression levels in conditions A and B; $T^{A/B}$: observation time points)

Gaussian Process Model

Inference

- Models are compared using the Bayes factor

$$\text{Score} = \log \frac{\overbrace{P(\mathcal{D}_A, \mathcal{D}_B \,|\, \mathcal{H}_\mathsf{I})}^{\textit{Independent}\text{ model}}}{\underbrace{P(\mathcal{D}_A, \mathcal{D}_B \,|\, \mathcal{H}_\mathsf{S})}_{\textit{Shared}\text{ model}}}.$$

- Writing out the GP models explicitly leads to

$$\text{Score} = \log \frac{P(\mathbf{Y}^A \,|\, \mathcal{H}_\mathsf{GP}, \mathbf{T}^A, \boldsymbol{\theta}_\mathsf{I}) P(\mathbf{Y}^B \,|\, \mathcal{H}_\mathsf{GP}, \mathbf{T}^B, \boldsymbol{\theta}_\mathsf{I})}{P(\mathbf{Y}^A \cup \mathbf{Y}^B \,|\, \mathcal{H}_\mathsf{GP}, \mathbf{T}^A \cup \mathbf{T}^B, \boldsymbol{\theta}_\mathsf{S})}.$$

($\mathbf{Y}^{A/B}$: expression levels in conditions A and B; $T^{A/B}$: observation time points)

Gaussian Process Model

Shared Model

- Given observed data from both conditions $\mathcal{D} = \{\mathbf{Y}^{A/B}, \mathbf{T}^{A/B}\}$ the posterior distribution over latent function values $\mathbf{f}$ is

$$P(\mathbf{f} \mid \mathbf{Y}, \mathbf{T}, \boldsymbol{\theta}_{\mathsf{K}}, \boldsymbol{\theta}_{\mathsf{L}}) \propto \mathcal{N}\left(\mathbf{f} \mid \mathbf{0}, \mathbf{K}_{\mathbf{T}}(\boldsymbol{\theta}_{\mathsf{K}})\right) \times \prod_{c \in \{A,B\}} \prod_{r=1}^{R} \prod_{n=1}^{N} p_{\mathsf{L}}(y^{c}_{r,t_n} \mid f_{t_n}, \boldsymbol{\theta}_{\mathsf{L}}),$$

  - Covariance funciton (kernel)
  - Noise model
  - Hyperparameters $\boldsymbol{\theta}_{\mathsf{S}} = \{\boldsymbol{\theta}_{\mathsf{K}}, \boldsymbol{\theta}_{\mathsf{L}}\}$ (length scale, noise levels)

- For Gaussian noise, $p_{\mathsf{L}}(y^{c}_{r,t} \mid f^{c}_{r,t}, \boldsymbol{\theta}_{\mathsf{L}}) = \mathcal{N}\left(y^{c}_{r,t} \,\middle|\, f^{c}_{r,t}, (\sigma^{c}_{r})^2\right)$, the model is tractable in closed form.

Gaussian Process Model

Shared Model

- Given observed data from both conditions $\mathcal{D} = \{\mathbf{Y}^{A/B}, \mathbf{T}^{A/B}\}$ the posterior distribution over latent function values $\mathbf{f}$ is

$$P(\mathbf{f} \mid \mathbf{Y}, \mathbf{T}, \boldsymbol{\theta}_{\mathsf{K}}, \boldsymbol{\theta}_{\mathsf{L}}) \propto \mathcal{N}\left(\mathbf{f} \,\middle|\, \mathbf{0}, \mathbf{K}_{\mathbf{T}}(\boldsymbol{\theta}_{\mathsf{K}})\right) \times \prod_{c \in \{A,B\}} \prod_{r=1}^{R} \prod_{n=1}^{N} p_{\mathsf{L}}(y^c_{r,t_n} \mid f_{t_n}, \boldsymbol{\theta}_{\mathsf{L}}),$$

  - Covariance funciton (kernel)
  - Noise model
  - Hyperparameters $\boldsymbol{\theta}_{\mathsf{S}} = \{\boldsymbol{\theta}_{\mathsf{K}}, \boldsymbol{\theta}_{\mathsf{L}}\}$ (length scale, noise levels)

- For Gaussian noise, $p_{\mathsf{L}}(y^c_{r,t} \mid f^c_{r,t}, \boldsymbol{\theta}_{\mathsf{L}}) = \mathcal{N}\left(y^c_{r,t} \,\middle|\, f^c_{r,t}, (\sigma^c_r)^2\right)$, the model is tractable in closed form.

Gaussian Process Model

Shared Model

- Given observed data from both conditions $\mathcal{D} = \{\mathbf{Y}^{A/B}, \mathbf{T}^{A/B}\}$ the posterior distribution over latent function values $\mathbf{f}$ is

$$P(\mathbf{f} \mid \mathbf{Y}, \mathbf{T}, \boldsymbol{\theta}_\mathsf{K}, \boldsymbol{\theta}_\mathsf{L}) \propto \mathcal{N}\left(\mathbf{f} \,\middle|\, \mathbf{0}, \mathbf{K}_\mathbf{T}(\boldsymbol{\theta}_\mathsf{K})\right) \times \prod_{c \in \{A,B\}} \prod_{r=1}^{R} \prod_{n=1}^{N} p_\mathsf{L}(y^c_{r,t_n} \mid f_{t_n}, \boldsymbol{\theta}_\mathsf{L}),$$

  - Covariance funciton (kernel)
  - Noise model
  - Hyperparameters $\boldsymbol{\theta}_\mathsf{S} = \{\boldsymbol{\theta}_\mathsf{K}, \boldsymbol{\theta}_\mathsf{L}\}$ (length scale, noise levels)

- For Gaussian noise, $p_\mathsf{L}(y^c_{r,t} \mid f^c_{r,t}, \boldsymbol{\theta}_\mathsf{L}) = \mathcal{N}\left(y^c_{r,t} \,\middle|\, f^c_{r,t}, (\sigma^c_r)^2\right)$, the model is tractable in closed form.

Gaussian Process Model

Shared Model

- Given observed data from both conditions $\mathcal{D} = \{\mathbf{Y}^{A/B}, \mathbf{T}^{A/B}\}$ the posterior distribution over latent function values $\mathbf{f}$ is

$$P(\mathbf{f} \mid \mathbf{Y}, \mathbf{T}, \boldsymbol{\theta}_\mathsf{K}, \boldsymbol{\theta}_\mathsf{L}) \propto \mathcal{N}\left(\mathbf{f} \,\middle|\, \mathbf{0}, \mathbf{K}_\mathbf{T}(\boldsymbol{\theta}_\mathsf{K})\right) \times \prod_{c \in \{A,B\}} \prod_{r=1}^{R} \prod_{n=1}^{N} p_\mathsf{L}(y^c_{r,t_n} \mid f_{t_n}, \boldsymbol{\theta}_\mathsf{L}),$$

  - Covariance funciton (kernel)
  - Noise model
  - Hyperparameters $\boldsymbol{\theta}_\mathsf{S} = \{\boldsymbol{\theta}_\mathsf{K}, \boldsymbol{\theta}_\mathsf{L}\}$ (length scale, noise levels)

- For Gaussian noise, $p_\mathsf{L}(y^c_{r,t} \mid f^c_{r,t}, \boldsymbol{\theta}_\mathsf{L}) = \mathcal{N}\left(y^c_{r,t} \,\middle|\, f^c_{r,t}, (\sigma^c_r)^2\right)$, the model is tractable in closed form.

Gaussian Process Model

Shared Model

- Given observed data from both conditions $\mathcal{D} = \{\mathbf{Y}^{A/B}, \mathbf{T}^{A/B}\}$ the posterior distribution over latent function values $\mathbf{f}$ is

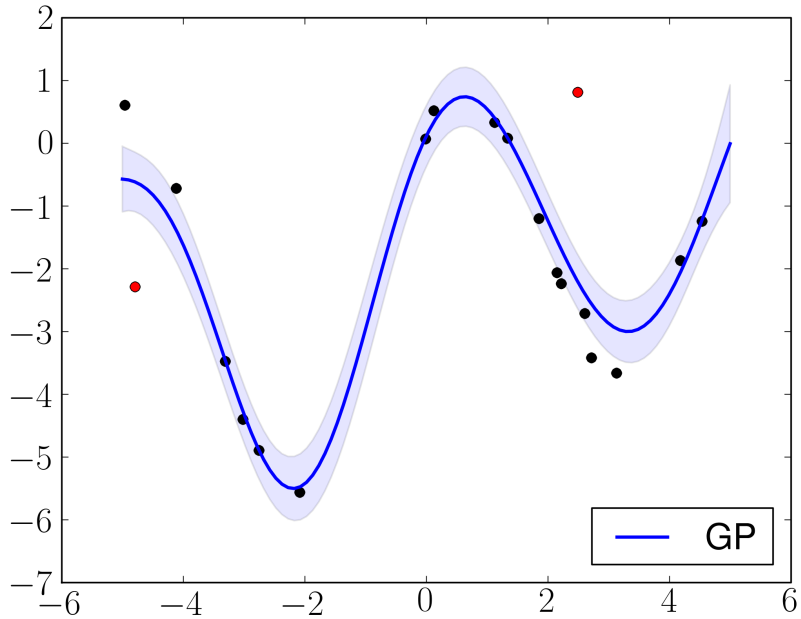
$$P(\mathbf{f} \mid \mathbf{Y}, \mathbf{T}, \boldsymbol{\theta}_\mathsf{K}, \boldsymbol{\theta}_\mathsf{L}) \propto \mathcal{N}\left(\mathbf{f} \,\middle|\, \mathbf{0}, \mathbf{K}_\mathbf{T}(\boldsymbol{\theta}_\mathsf{K})\right) \times \prod_{c \in \{A,B\}} \prod_{r=1}^{R} \prod_{n=1}^{N} p_\mathsf{L}(y^c_{r,t_n} \mid f_{t_n}, \boldsymbol{\theta}_\mathsf{L}),$$

  - Covariance funciton (kernel)
  - Noise model
  - Hyperparameters $\boldsymbol{\theta}_\mathsf{S} = \{\boldsymbol{\theta}_\mathsf{K}, \boldsymbol{\theta}_\mathsf{L}\}$ (length scale, noise levels)

- For Gaussian noise, $p_\mathsf{L}(y^c_{r,t} \mid f^c_{r,t}, \boldsymbol{\theta}_\mathsf{L}) = \mathcal{N}\left(y^c_{r,t} \,\middle|\, f^c_{r,t}, (\sigma^c_r)^2\right)$, the model is tractable in closed form.

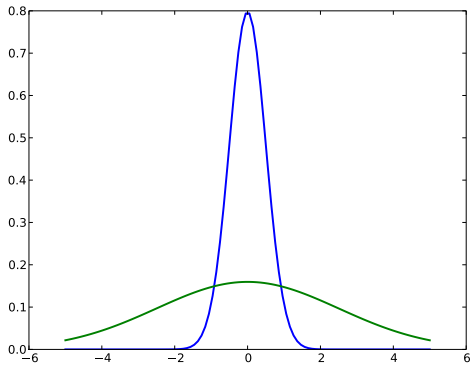
Robustness With Respect to Outliers

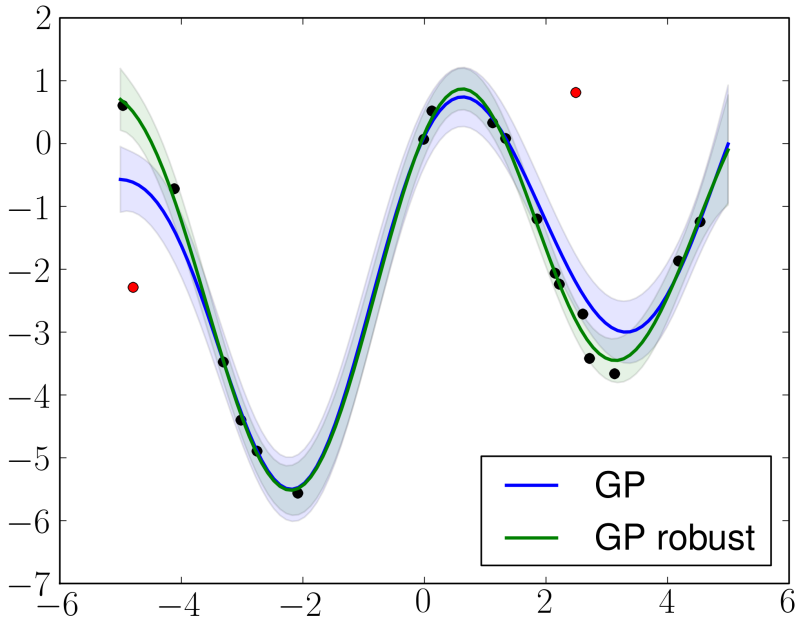
- *Outliers* in the expression profile can obscure the regression results.
- A mixture noise-model accounts for outliers.
- Inference in this model is done using Expectation Propagation.



Robustness With Respect to Outliers

- *Outliers* in the expression profile can obscure the regression results.
- A mixture noise-model accounts for outliers.

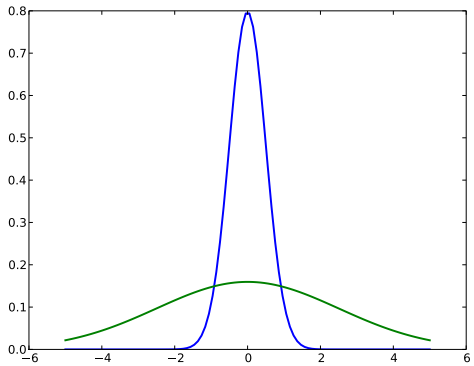


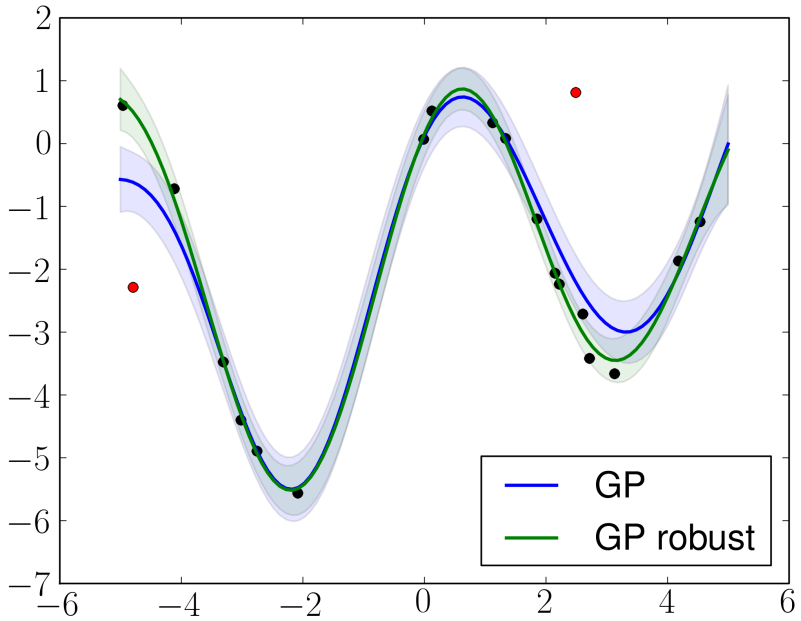


- Inference in this model is done using Expectation Propagation.

Robustness With Respect to Outliers

- *Outliers* in the expression profile can obscure the regression results.
- A mixture noise-model accounts for outliers.





- Inference in this model is done using Expectation Propagation.

Illustration of the Model Comparison

A Differentially Expressed Gene

- *Shared* model.

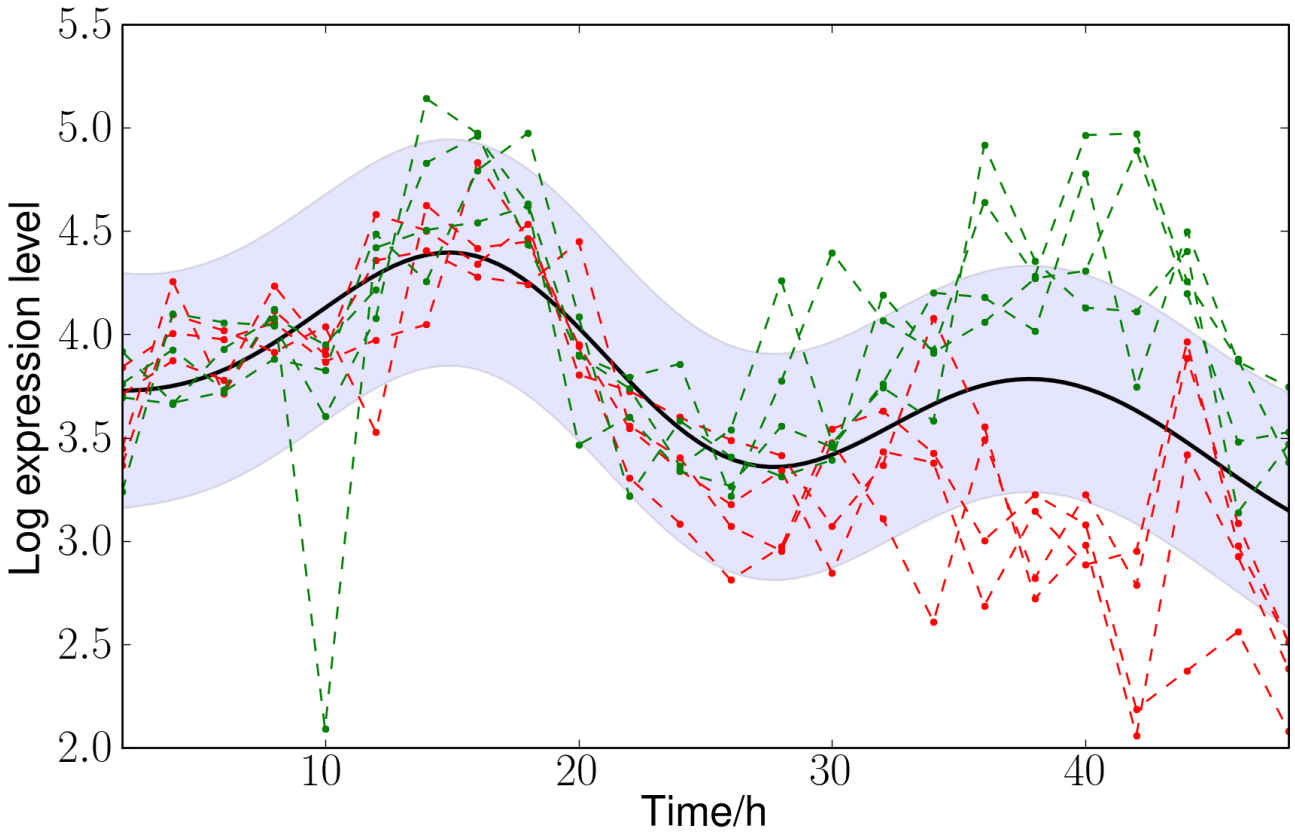


Illustration of the Model Comparison

A Differentially Expressed Gene

- *Independent* model.

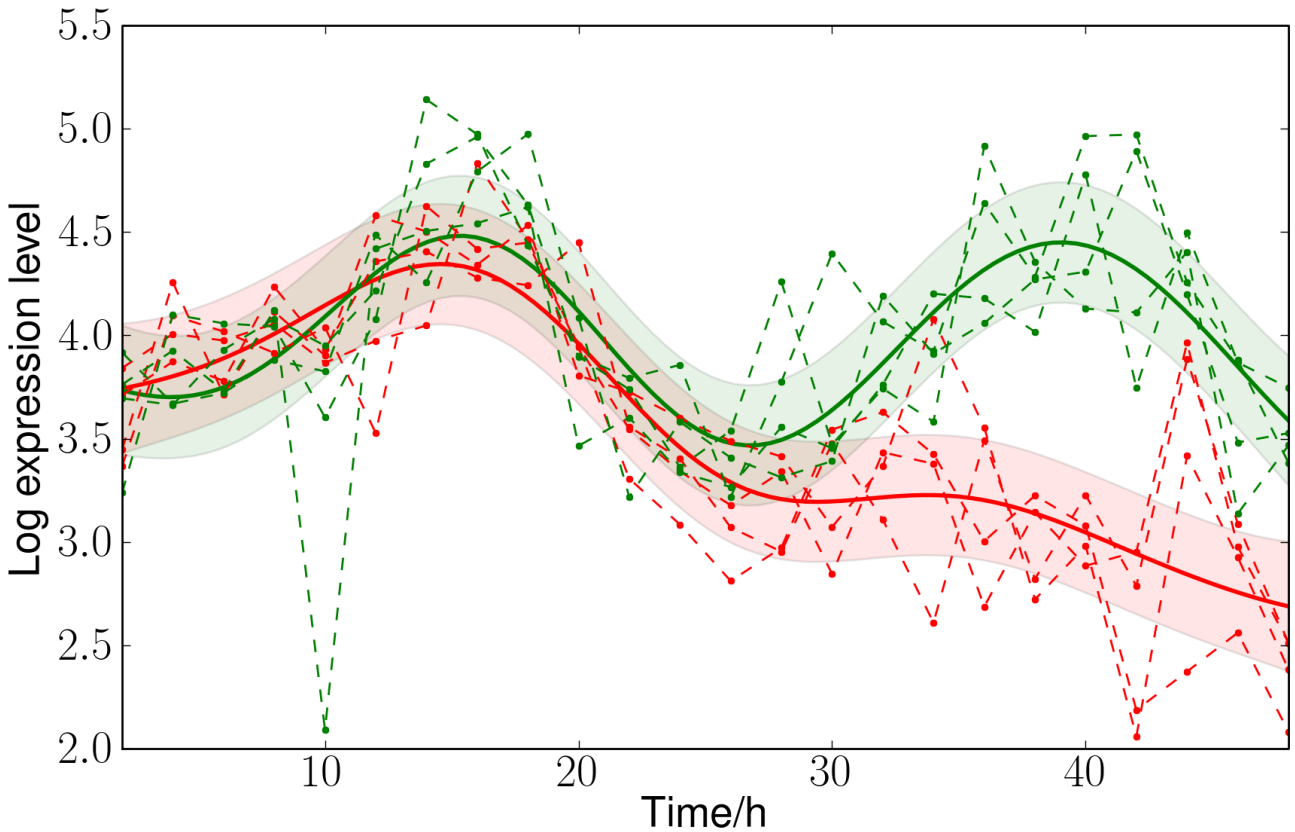
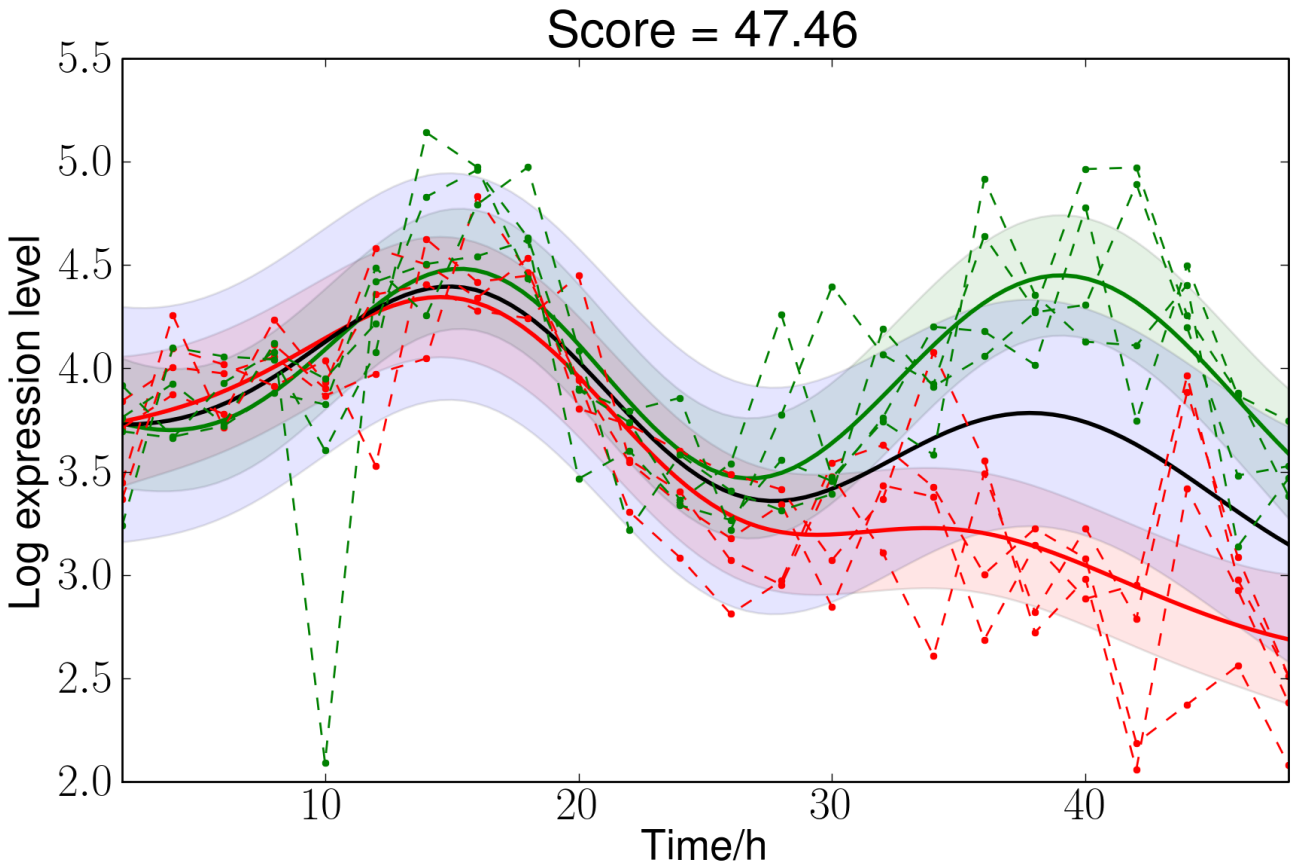


Illustration of the Model Comparison

A Differentially Expressed Gene

- Model comparison.



Predictive Performance (RECOMB09)

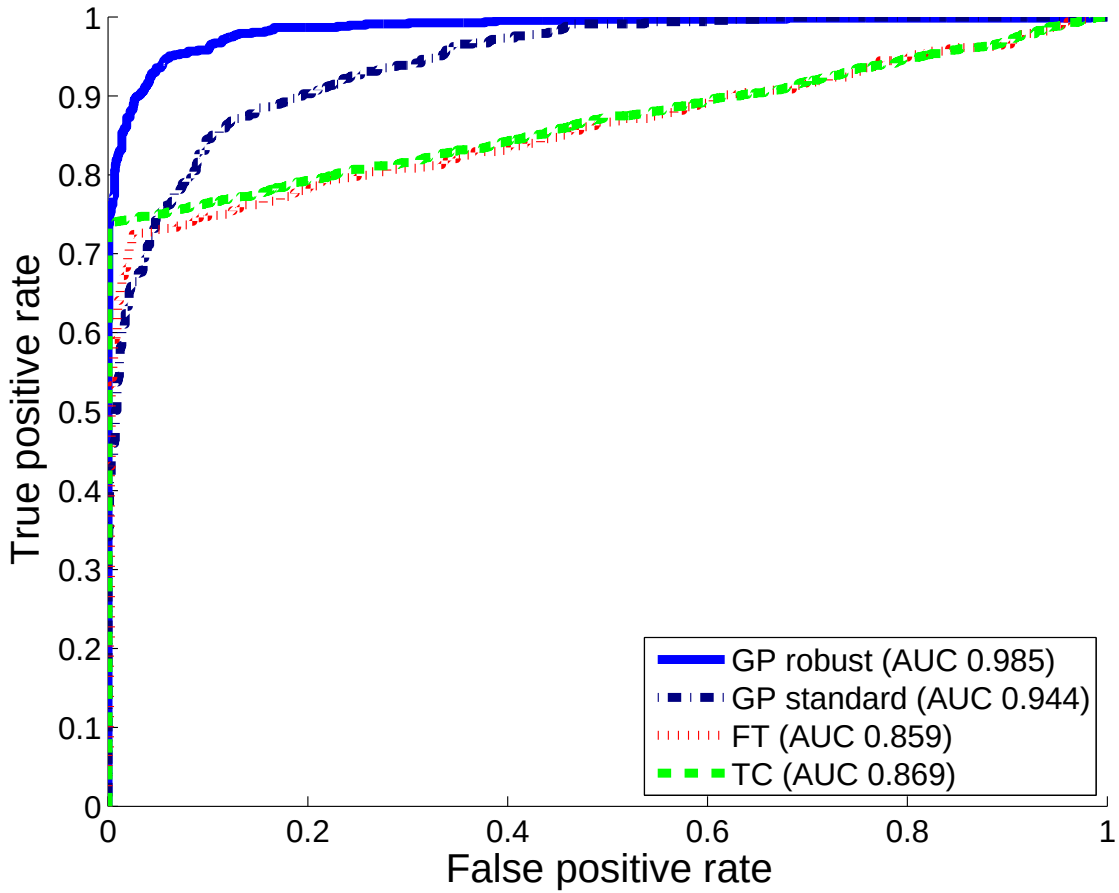
- Data:
  - 30,336 *Arabidopsis thaliana* gene probes
  - Biotic stress: fungus infection
  - 24 time points, 4 biological replicates
- Evaluation of alternative methods on 2000 randomly chosen human-labeled probes:
  - GP no robust
  - GP robust
  - F-Test (FT) (MAANOVA package)
  - *Timecourse* (TC) (Tai and Speed)

Predictive Performance (RECOMB09)

- Data:
  - 30,336 *Arabidopsis thaliana* gene probes
  - Biotic stress: fungus infection
  - 24 time points, 4 biological replicates
- Evaluation of alternative methods on 2000 randomly chosen human-labeled probes:
  - GP no robust
  - GP robust
  - F-Test (FT) (MAANOVA package)
  - *Timecourse* (TC) (Tai and Speed)

Predictive Performance (RECOMB09)

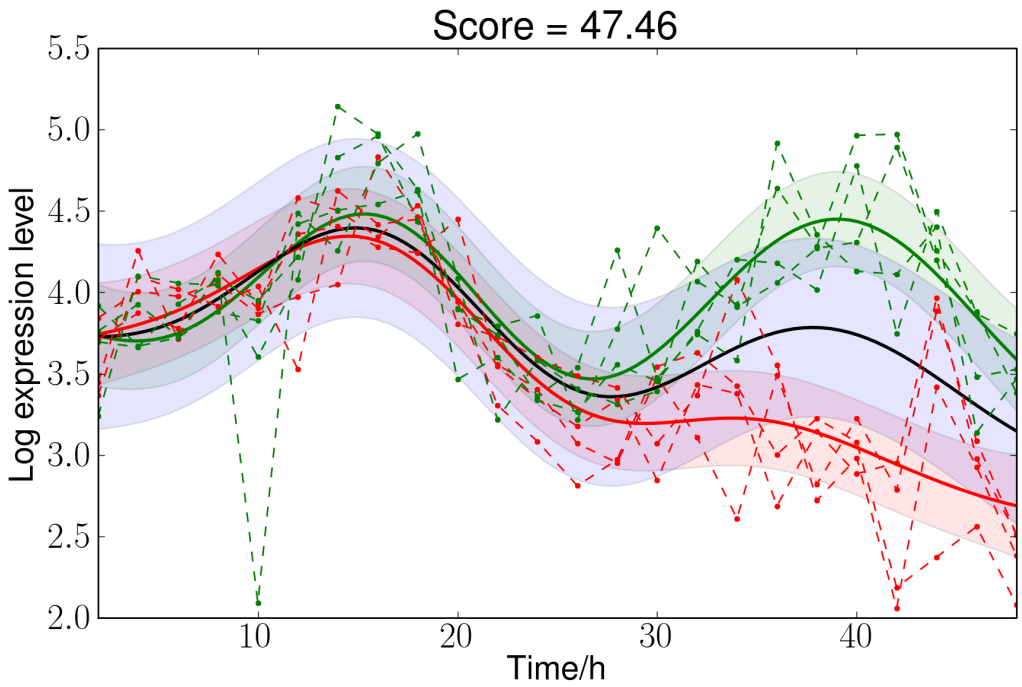
ROC Curves



Time-local GPTwoSample

Motivation

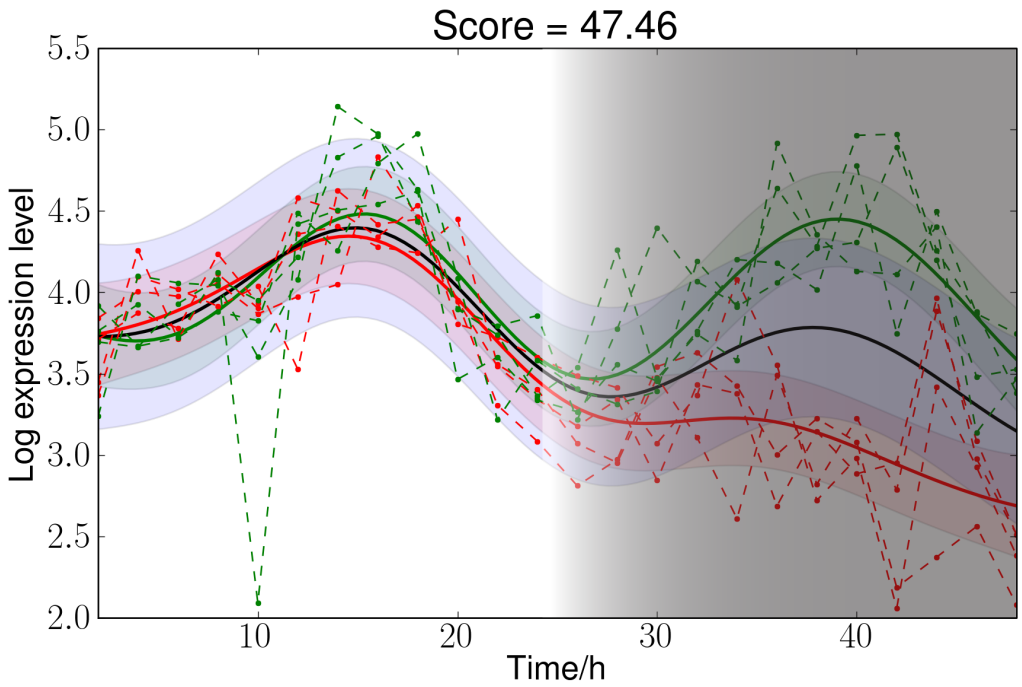
- Differential expression is **not** static over time.
  - The response develops over time.
  - Different regulators may be active at distinct times and trigger each other.



Time-local GPTwoSample

Motivation

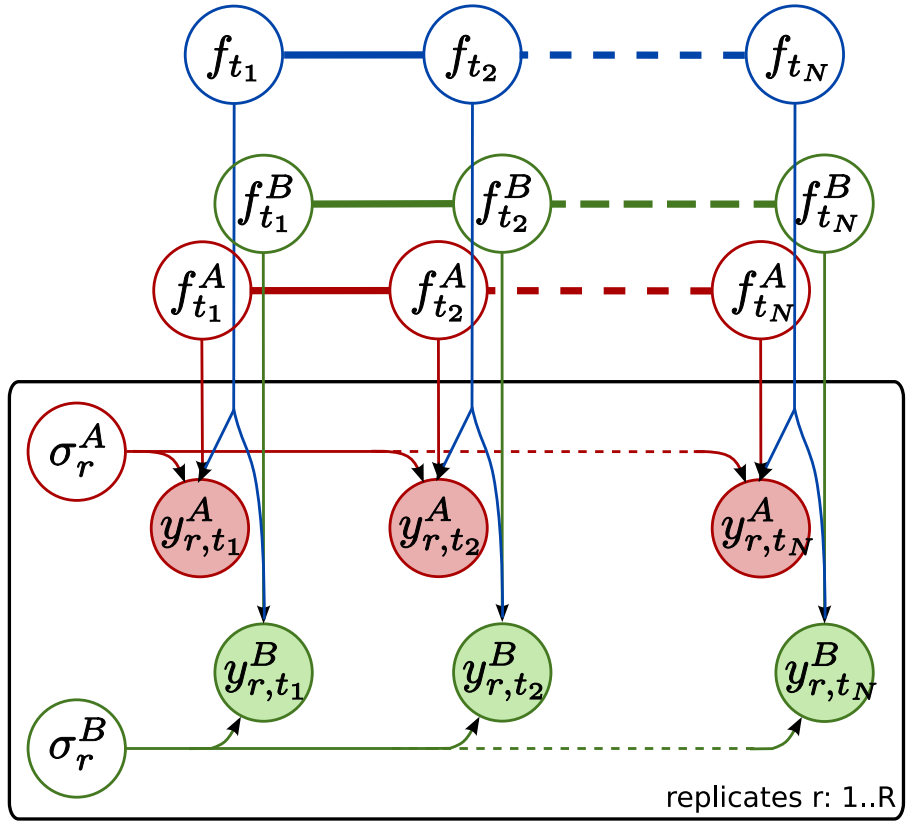
- Differential expression is **not** static over time.
  - The response develops over time.
  - Different regulators may be active at distinct times and trigger each other.



Time-local GPTwoSample

Model

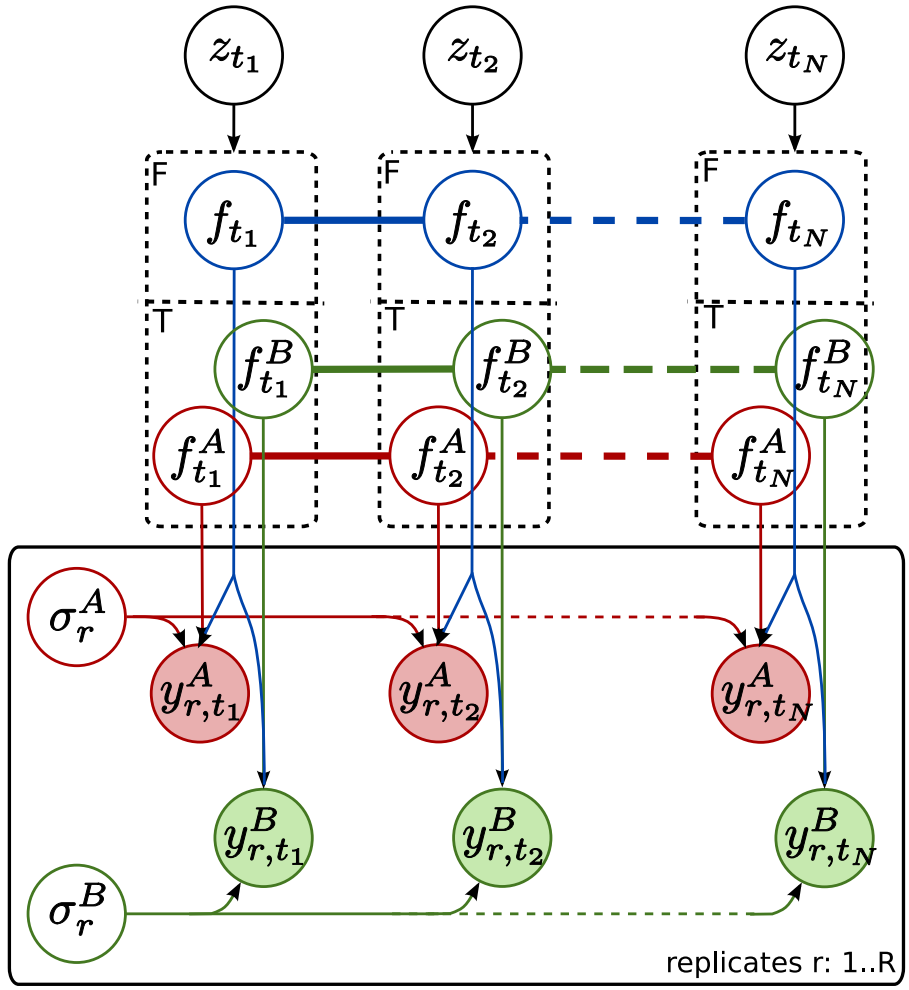
- The *shared* and *independent* model are interleaved over time.
- Indicator variables z_{t_n} switch between both models.
- Inference is done using Gibbs sampling (later).



Time-local GPTwoSample

Model

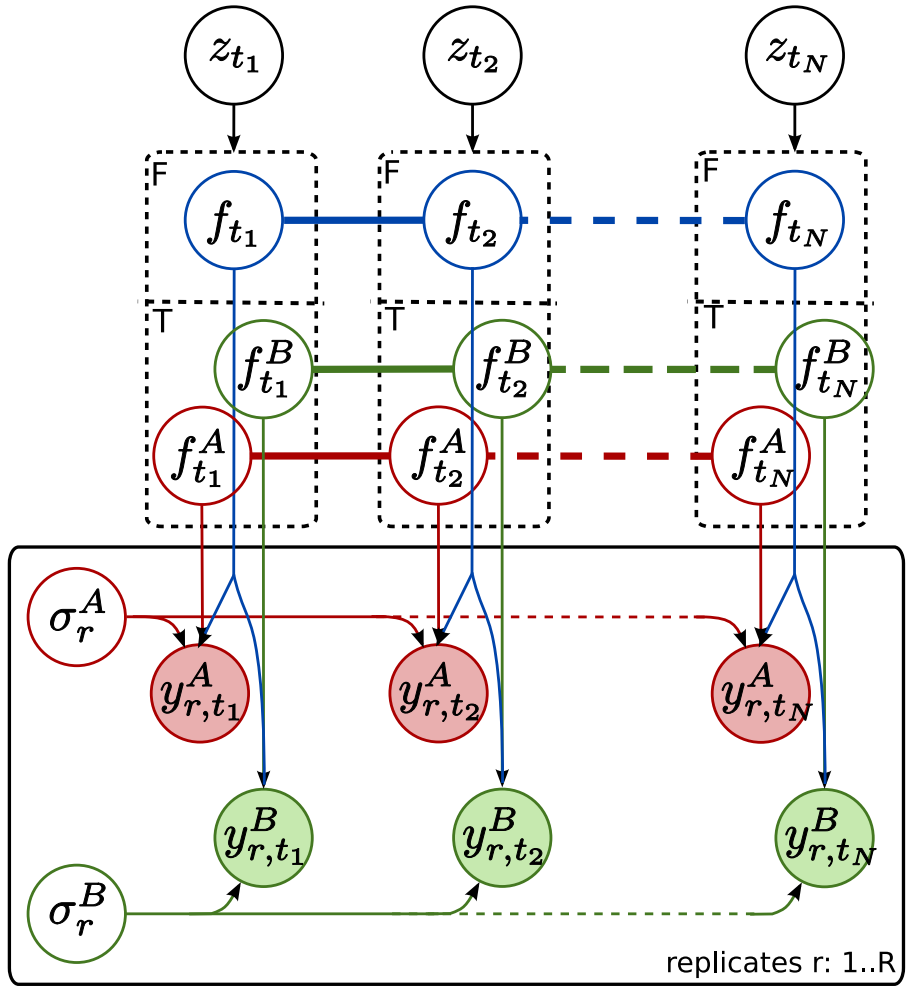
- The *shared* and *independent* model are interleaved over time.
- Indicator variables z_{t_n} switch between both models.
- Inference is done using Gibbs sampling (later).



Time-local GPTwoSample

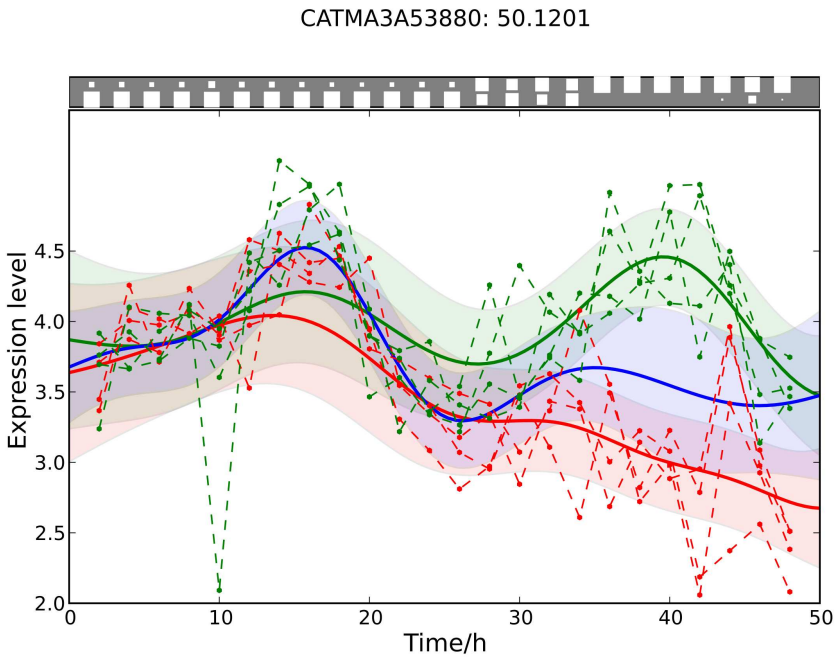
Model

- The *shared* and *independent* model are interleaved over time.
- Indicator variables z_{t_n} switch between both models.
- Inference is done using Gibbs sampling (later).



Time-local GPTwoSample

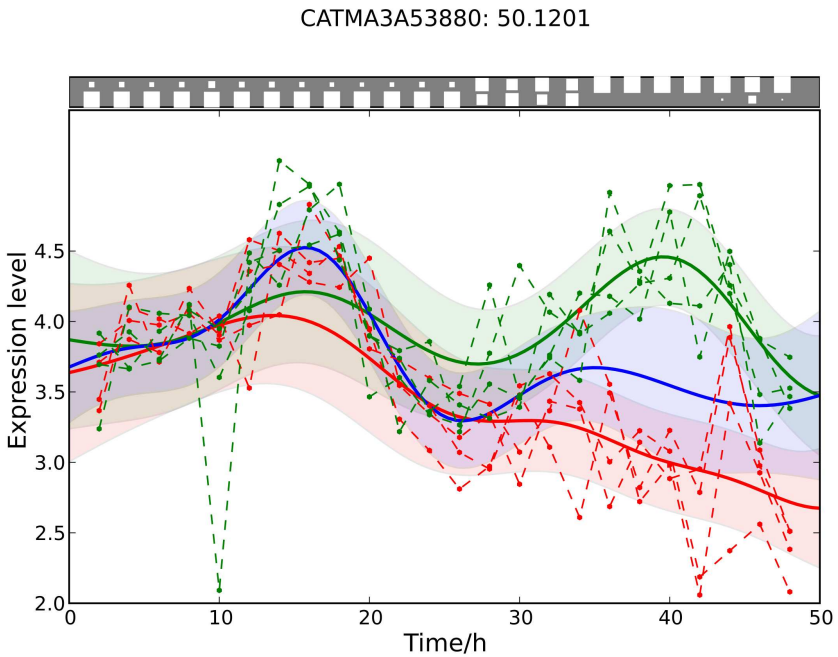
Example Results



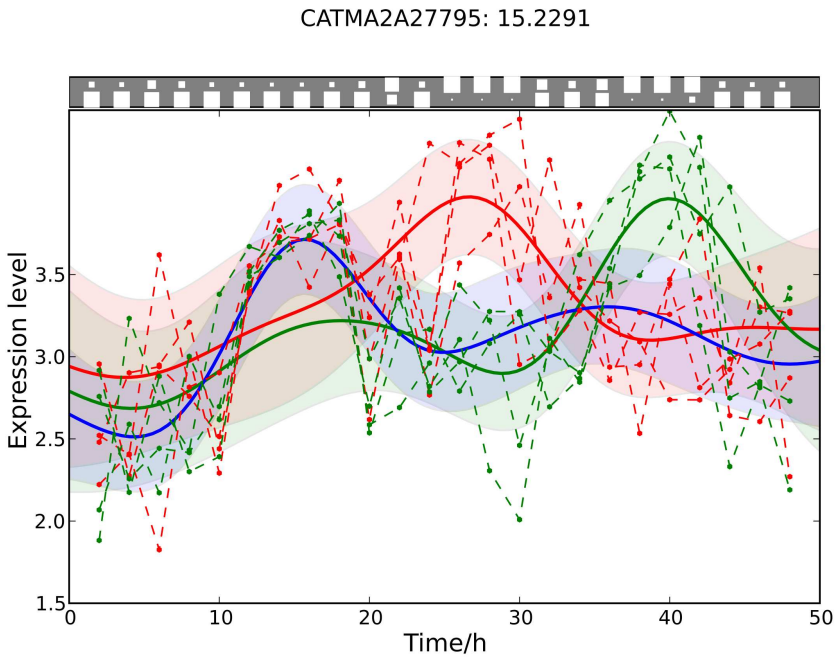
The example from before

Time-local GPTwoSample

Example Results



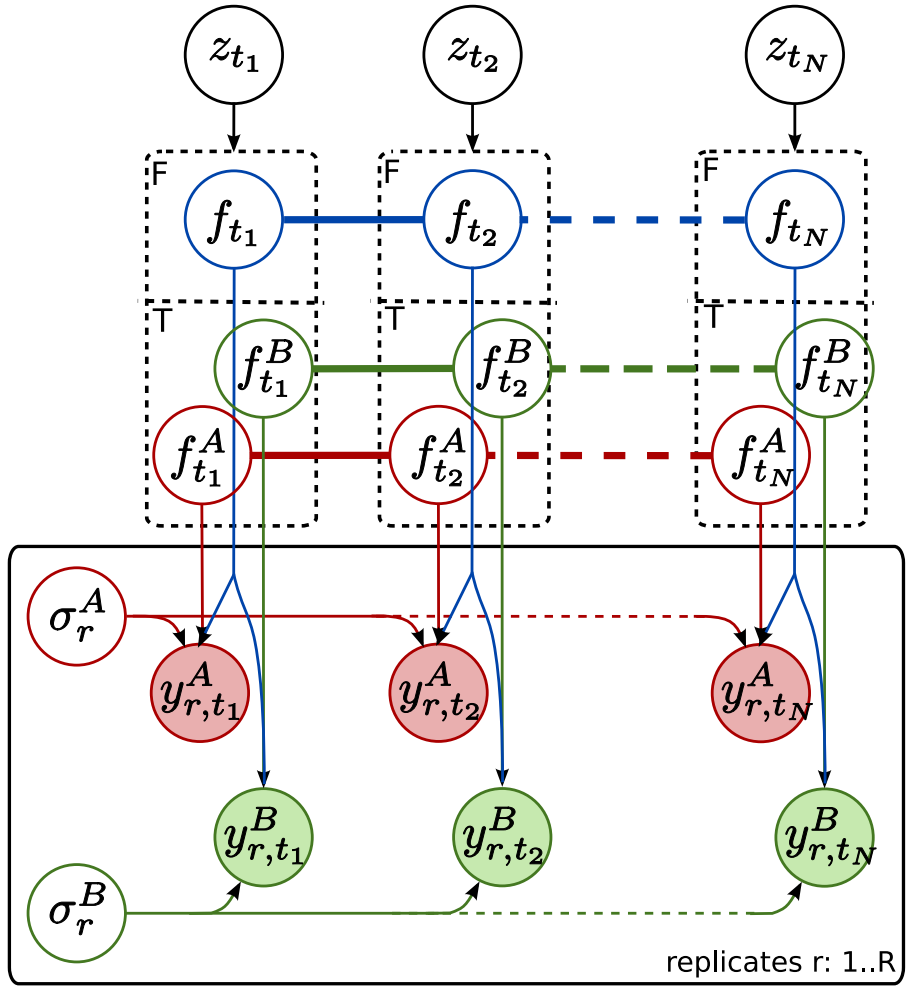
The example from before



Periodic differential expression

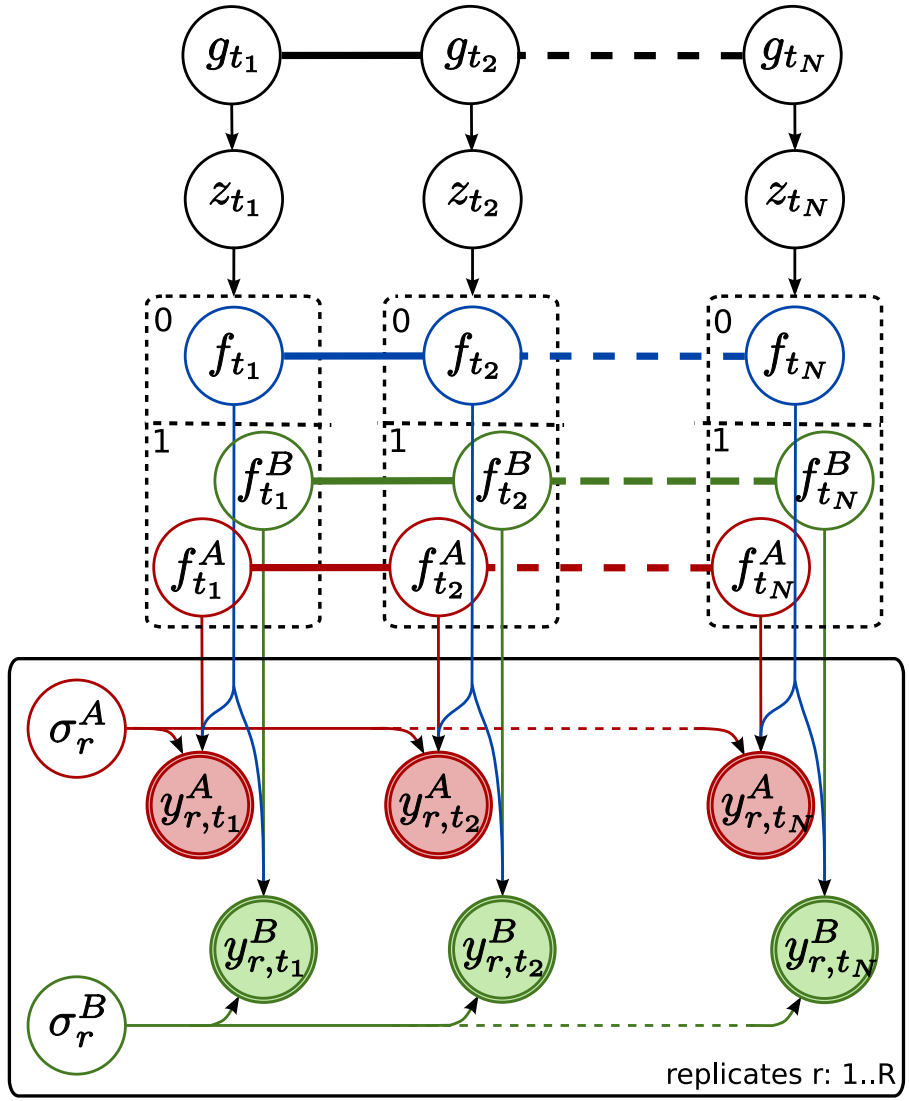
Smooth Time-local GPTwoSample

- Transitions between non-differential and differential expression can occur at unobserved time points and are smooth.
- Extending the time-local model with a GP prior as gating network on the switch variables.
- Inference using Gibbs sampling.



Smooth Time-local GPTwoSample

- Transitions between non-differential and differential expression can occur at unobserved time points and are smooth.
- Extending the time-local model with a GP prior as gating network on the switch variables.
- Inference using Gibbs sampling.



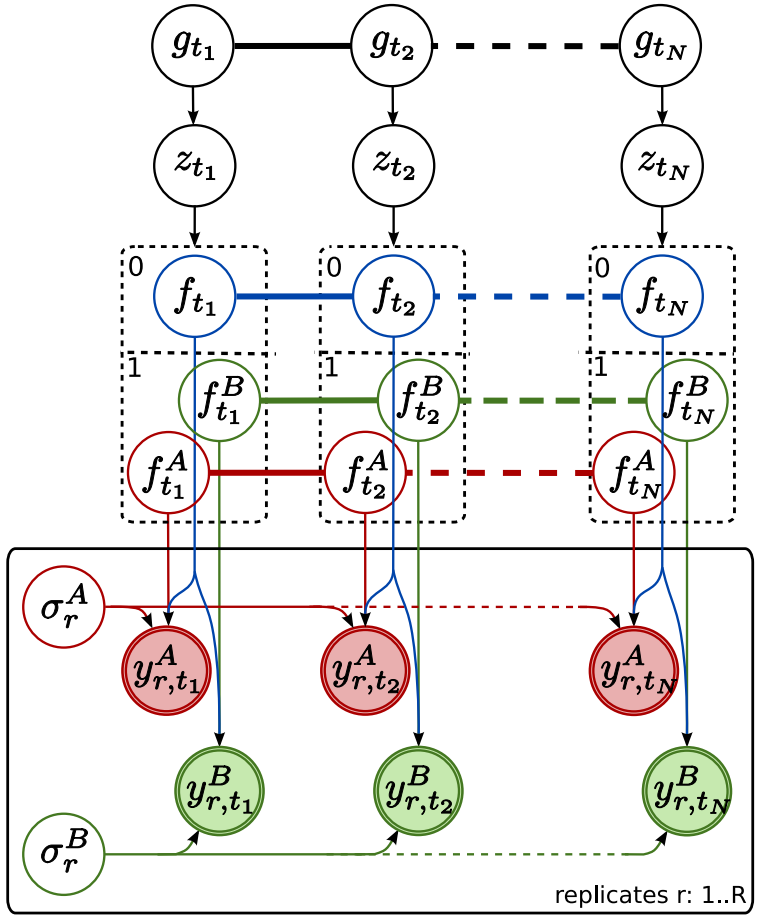
Smooth Time-local GPTwoSample

Gibbs Sampling

- Gibbs sampling exploits tractable conditional distributions.
- Individual indicators z_{t_i} are resampled in turn

$$P\left(z_{t_i} = s \,|\, \mathbf{z}^{\setminus t_i}, \mathbf{T}, \mathbf{Y}, \theta_S, \theta_I, \theta_G\right)$$
$$\sim P\left(\mathbf{Y} \,|\, z_{t_i} = s, \mathbf{z}^{\setminus t_i}, \mathbf{T}, \theta_I, \theta_S\right)$$
$$\times P\left(z_{t_i} = s \,|\, \mathbf{z}^{\setminus t_i}, \theta_G\right).$$

- Data likelihood from GP experts
- Predictive distribution from gating network



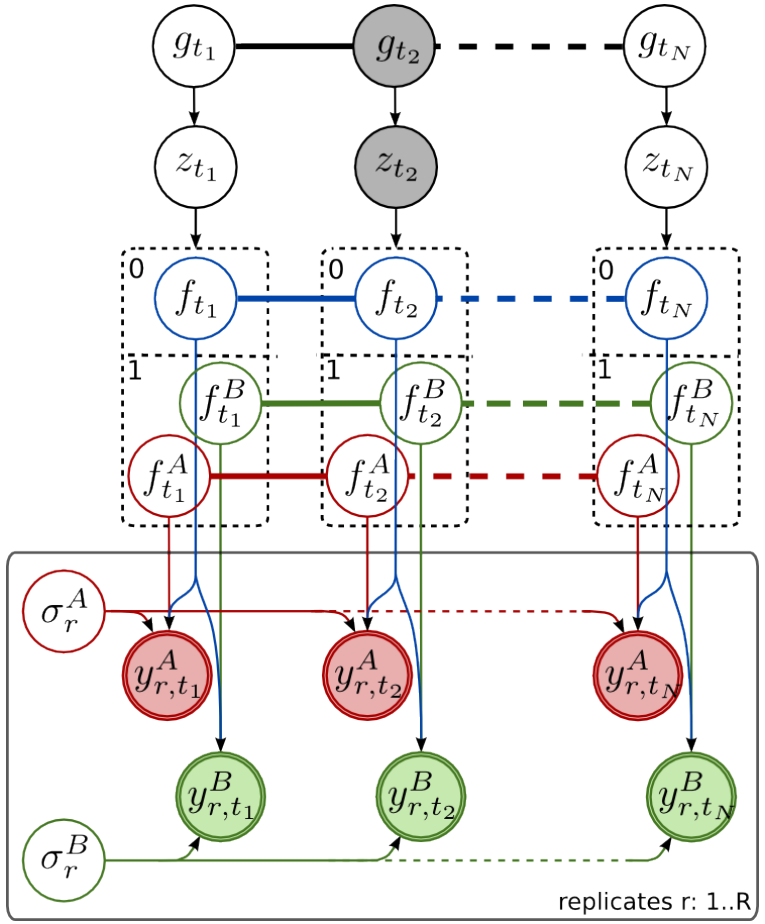
Smooth Time-local GPTwoSample

Gibbs Sampling

- Gibbs sampling exploits tractable conditional distributions.
- Individual indicators z_{t_i} are resampled in turn

$$
\begin{aligned}
&P\left(z_{t_i} = s \,|\, \mathbf{z}^{\setminus t_i}, \mathbf{T}, \mathbf{Y}, \boldsymbol{\theta}_\mathsf{S}, \boldsymbol{\theta}_\mathsf{I}, \boldsymbol{\theta}_\mathsf{G}\right) \\
&\quad \sim P\left(\mathbf{Y} \,|\, z_{t_i} = s, \mathbf{z}^{\setminus t_i}, \mathbf{T}, \boldsymbol{\theta}_\mathsf{I}, \boldsymbol{\theta}_\mathsf{S}\right) \\
&\quad \times P\left(z_{t_i} = s \,|\, \mathbf{z}^{\setminus t_i}, \boldsymbol{\theta}_\mathsf{G}\right).
\end{aligned}
$$

- Data likelihood from GP experts
- Predictive distribution from gating network



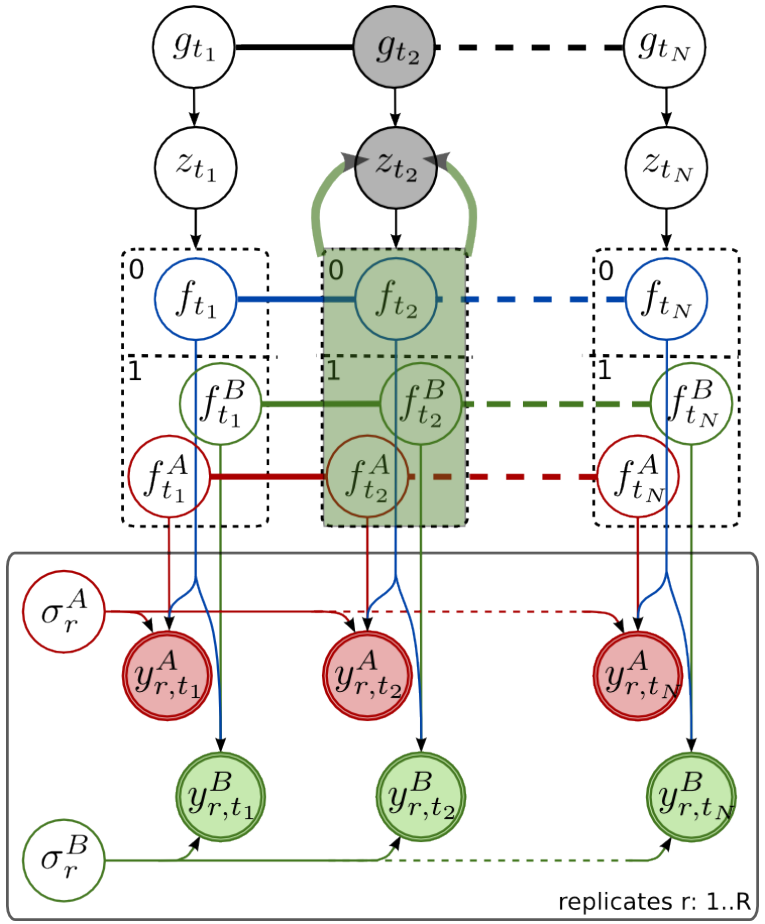
Smooth Time-local GPTwoSample

Gibbs Sampling

- Gibbs sampling exploits tractable conditional distributions.
- Individual indicators z_{t_i} are resampled in turn

$$
\begin{aligned}
& P\left(z_{t_i} = s \,|\, \mathbf{z}^{\setminus t_i}, \mathbf{T}, \mathbf{Y}, \boldsymbol{\theta}_\mathsf{S}, \boldsymbol{\theta}_\mathsf{I}, \boldsymbol{\theta}_\mathsf{G}\right) \\
& \sim P\left(\mathbf{Y} \,|\, z_{t_i} = s, \mathbf{z}^{\setminus t_i}, \mathbf{T}, \boldsymbol{\theta}_\mathsf{I}, \boldsymbol{\theta}_\mathsf{S}\right) \\
& \times P\left(z_{t_i} = s \,|\, \mathbf{z}^{\setminus t_i}, \boldsymbol{\theta}_\mathsf{G}\right).
\end{aligned}
$$

- Data likelihood from GP experts
- Predictive distribution from gating network



Smooth Time-local GPTwoSample

Gibbs Sampling

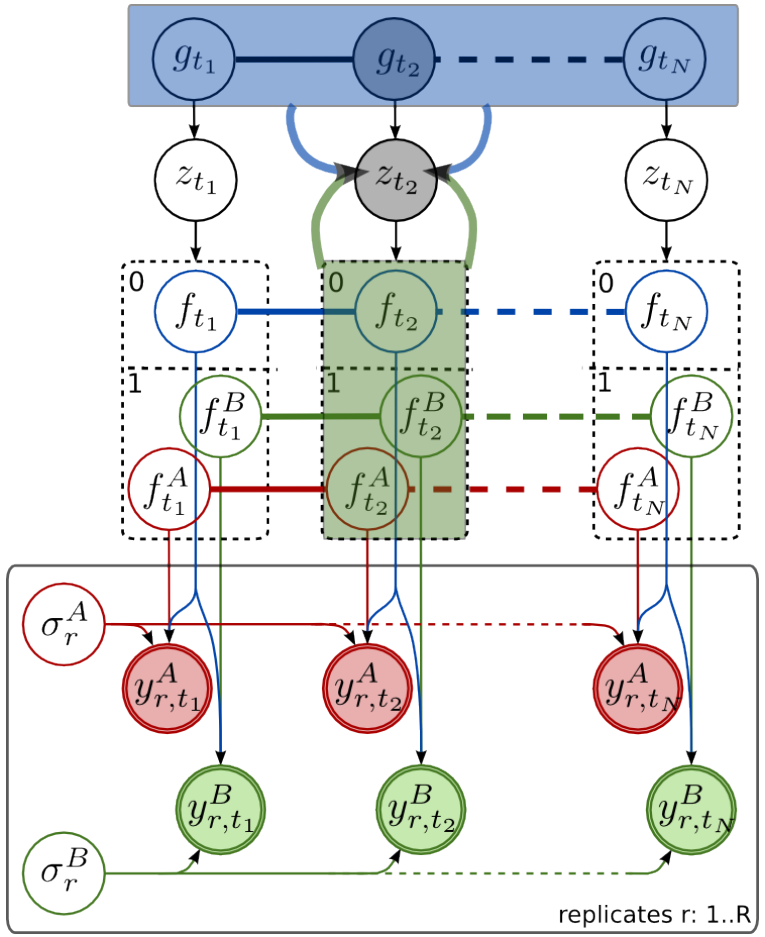
- Gibbs sampling exploits tractable conditional distributions.
- Individual indicators z_{t_i} are resampled in turn

$$P\left(z_{t_i} = s \,|\, \mathbf{z}^{\setminus t_i}, \mathbf{T}, \mathbf{Y}, \boldsymbol{\theta}_\mathsf{S}, \boldsymbol{\theta}_\mathsf{I}, \boldsymbol{\theta}_\mathsf{G}\right)$$

$$\sim P\left(\mathbf{Y} \,|\, z_{t_i} = s, \mathbf{z}^{\setminus t_i}, \mathbf{T}, \boldsymbol{\theta}_\mathsf{I}, \boldsymbol{\theta}_\mathsf{S}\right)$$

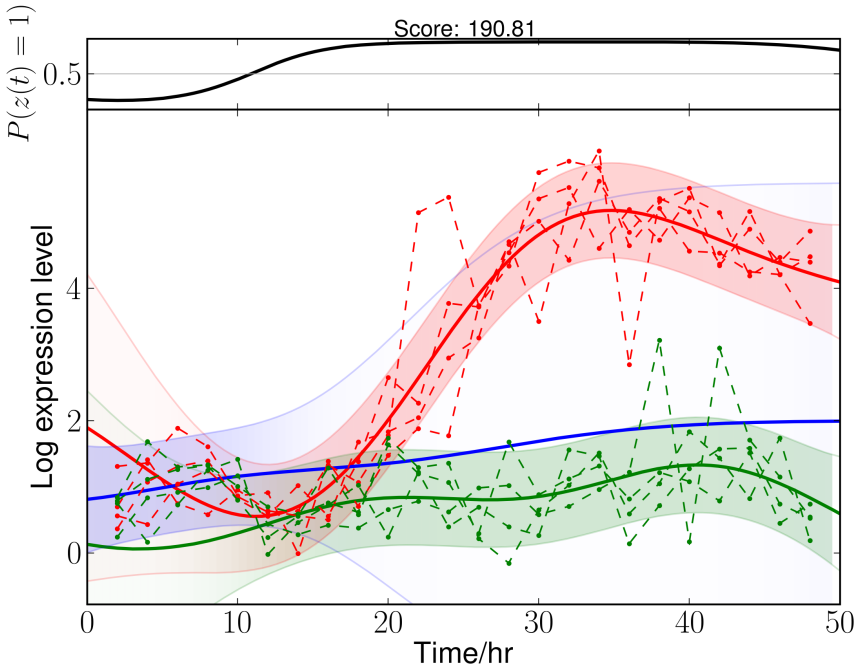
$$\times P\left(z_{t_i} = s \,|\, \mathbf{z}^{\setminus t_i}, \boldsymbol{\theta}_\mathsf{G}\right).$$

- Data likelihood from GP experts
- Predictive distribution from gating network

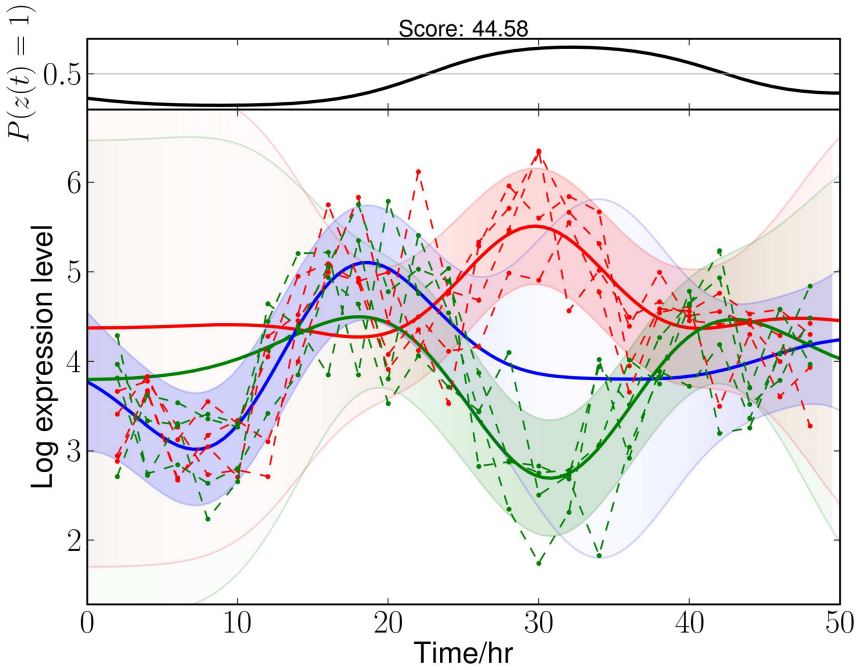


Smooth Time-local GPTwoSample

Example Results



Early differential expression

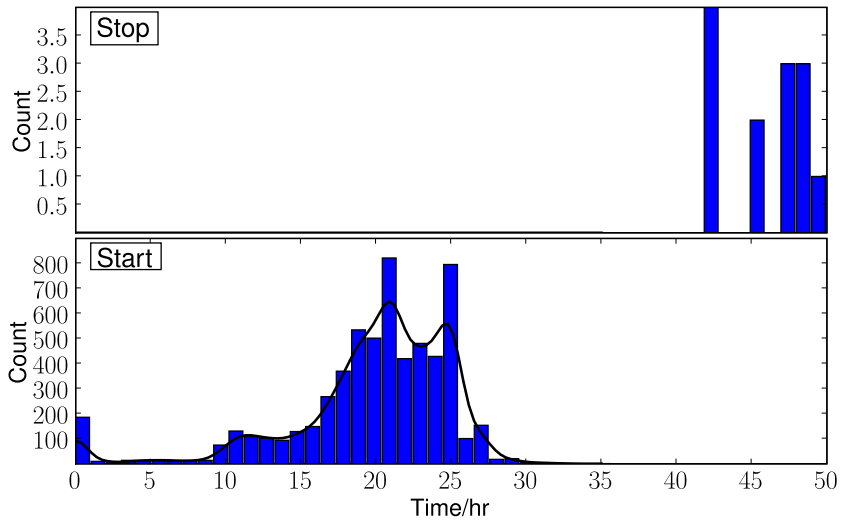


Transient differential expression

Detecting Transition Points in *Arabidopsis* Microarray Time Series

Start/Stop Times

- Studying the temporal distribution of differential expression across genes.
- Considered were the top 6000 differentially expressed genes.
- Differential expression appears to occur in two waves with start times at 20h an 25h after infection.
- Only very few genes stop differential behavior within the measured time interval.

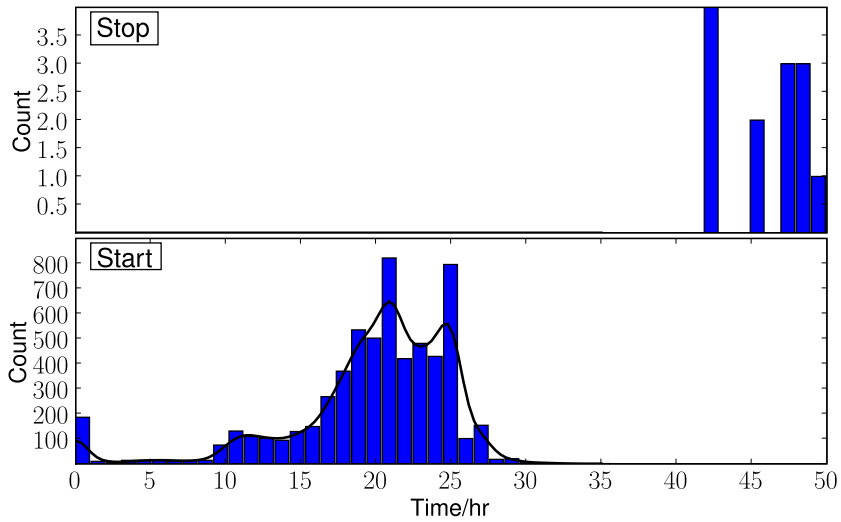


Distribution of differential start/stop time

Detecting Transition Points in *Arabidopsis* Microarray Time Series

Start/Stop Times

- Studying the temporal distribution of differential expression across genes.
- Considered were the top 6000 differentially expressed genes.
- Differential expression appears to occur in two waves with start times at 20h an 25h after infection.
- Only very few genes stop differential behavior within the measured time interval.

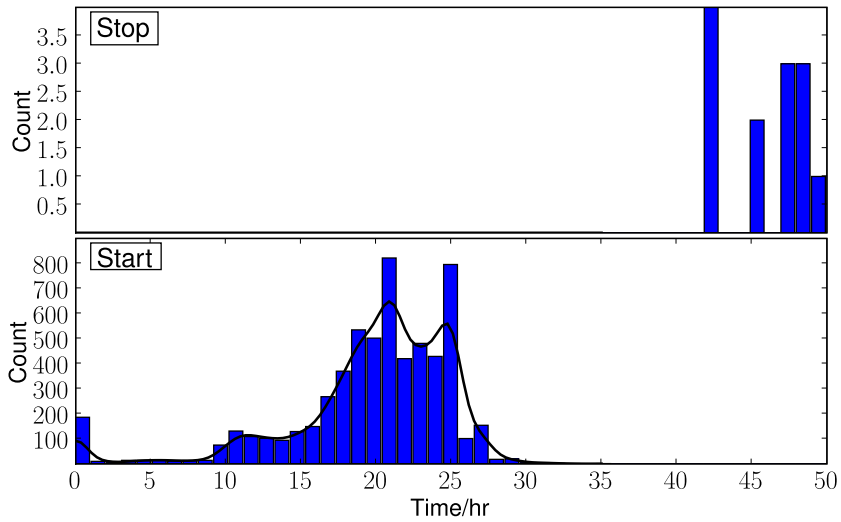


Distribution of differential start/stop time

Detecting Transition Points in *Arabidopsis* Microarray Time Series

Start/Stop Times

- Studying the temporal distribution of differential expression across genes.
- Considered were the top 6000 differentially expressed genes.
- Differential expression appears to occur in two waves with start times at 20h an 25h after infection.
- Only very few genes stop differential behavior within the measured time interval.

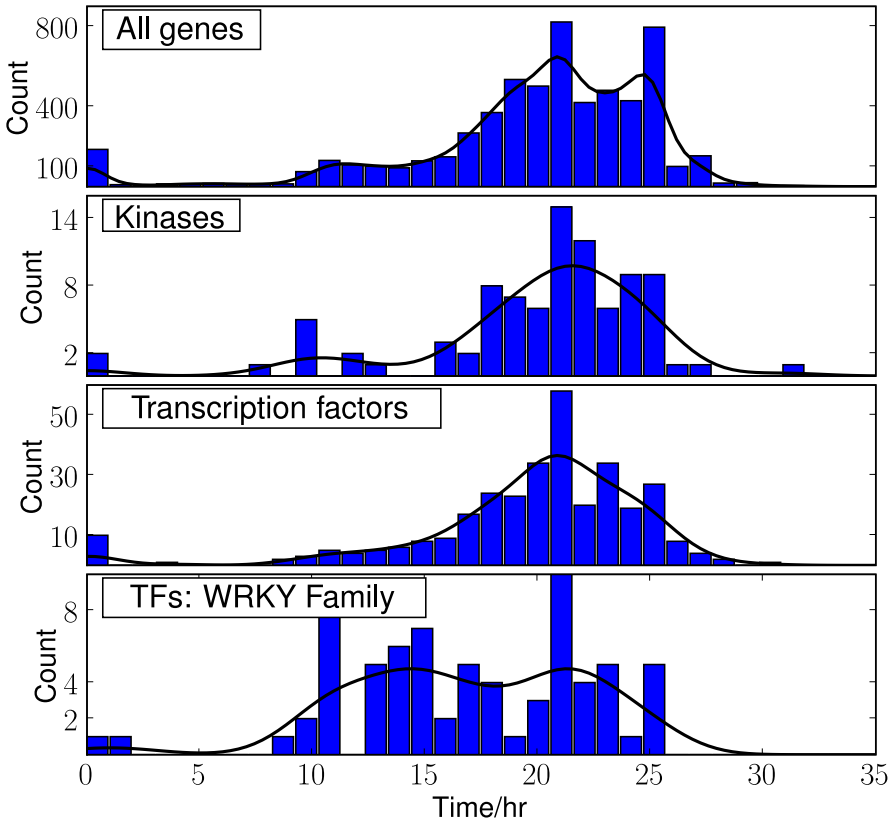


Distribution of differential start/stop time

Detecting Transition Points in *Arabidopsis* Microarray Time Series

Start Times for Gene Categories

- This distribution can be broke down into gene categories.
- WRKY Family of transcription factors is known to be involved in stress response.

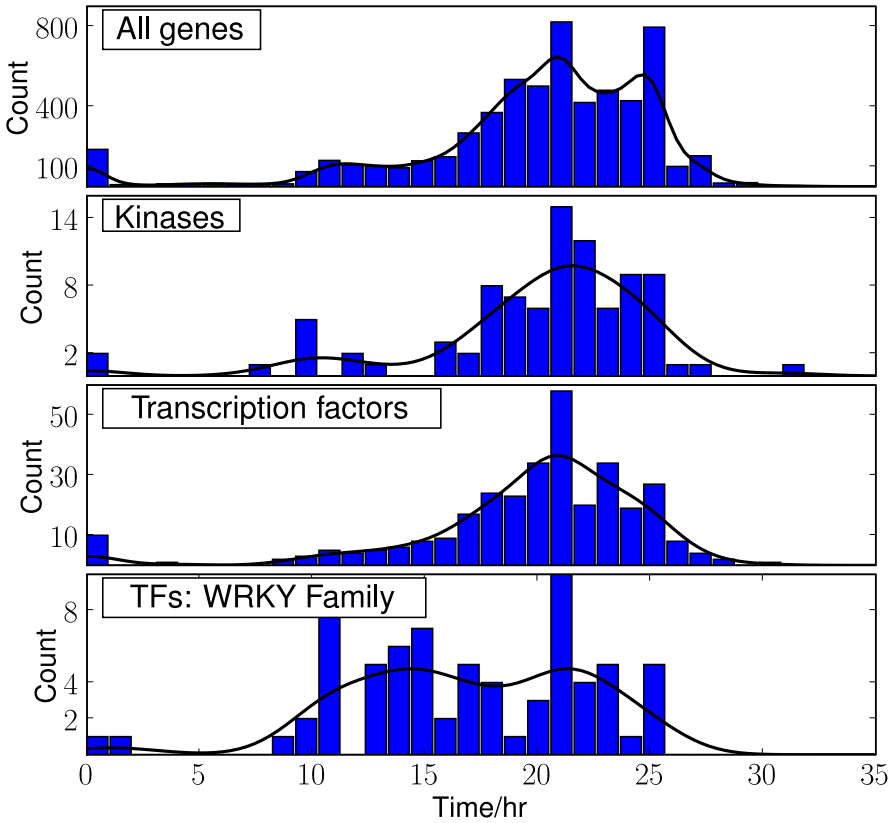


Distribution of differential start time

Detecting Transition Points in *Arabidopsis* Microarray Time Series

Start Times for Gene Categories

- This distribution can be broke down into gene categories.
- WRKY Family of transcription factors is known to be involved in stress response.



Distribution of differential start time

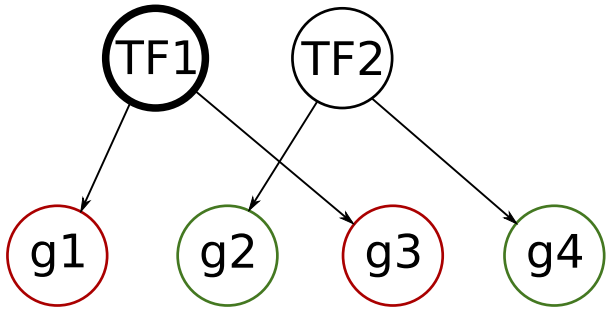
Outline

Motivation

- This part of the course is inspired and based on results of a publication by Neil D. Lawrence et al.
  Modelling transcriptional regulation using Gaussian processes
  `ftp://ftp.dcs.shef.ac.uk/home/neil/gpsim.pdf`.

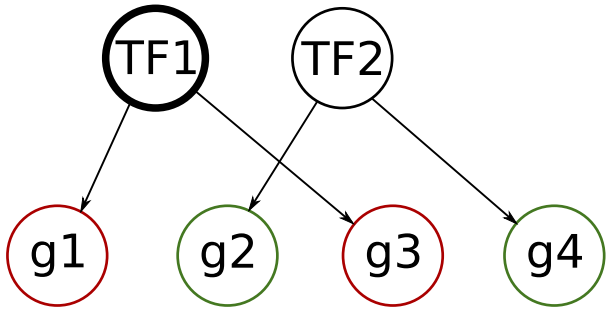
Motivation

- Microarray technologies allow to measure mRNA levels.
- The functional proteins and their concentration levels remain unobserved.
- Motivation: Infer the hidden protein concentrations?



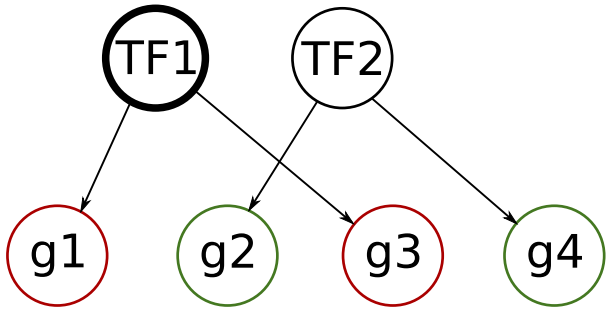
Motivation

- Microarray technologies allow to measure mRNA levels.
- The functional proteins and their concentration levels remain unobserved.
- Motivation: Infer the hidden protein concentrations?



Motivation

- Microarray technologies allow to measure mRNA levels.
- The functional proteins and their concentration levels remain unobserved.
- Motivation: Infer the hidden protein concentrations?



A single gene model

- The change in gene expression abundance y_i for a gene i is approximately described by a differential equation model of the form

$$\frac{\mathrm{d}y_i(t)}{\mathrm{d}t} = B_i + S_i f(t) - D_i y_i(t)$$

  - $f(t)$ regulatory transcription factor.
  - B_i basal transcription rate.
  - S_i sensitivity of the gene to the transcription factor.
  - D_i decay rate of the mRNA.
- Goal: infer the unobserved activation $f(t)$ from mRNA measurements of multiple target genes.

A single gene model

- The change in gene expression abundance y_i for a gene i is approximately described by a differential equation model of the form

$$\frac{\mathrm{d}y_i(t)}{\mathrm{d}t} = B_i + S_i f(t) - D_i y_i(t)$$

  - $f(t)$ regulatory transcription factor.
  - B_i basal transcription rate.
  - S_i sensitivity of the gene to the transcription factor.
  - D_i decay rate of the mRNA.
- Goal: infer the unobserved activation $f(t)$ from mRNA measurements of multiple target genes.

Derivative observations

- The key to solving this problem are derivative observations.
- Given knowledge about the derivative of a function f we would like to infer its function values:

$$dy = \frac{\partial f(t)}{\partial t}$$

- We wish to find the joint probability of function values and function derivatives

$$\text{cov}(dy_i, y_j) = \frac{\partial}{\partial t_i}\text{cov}(y_i, y_j)$$

$$\text{cov}(dy_i, dy_j) = \frac{\partial^2}{\partial t_i \partial t_j}\text{cov}(y_i, y_j)$$

- Using these covariance functions we can combine function observations and derivatives as training data in GP regression.

Derivative observations

- The key to solving this problem are derivative observations.
- Given knowledge about the derivative of a function f we would like to infer its function values:

$$dy = \frac{\partial f(t)}{\partial t}$$

- We wish to find the joint probability of function values and function derivatives

$$\text{cov}(dy_i, y_j) = \frac{\partial}{\partial t_i}\text{cov}(y_i, y_j)$$

$$\text{cov}(dy_i, dy_j) = \frac{\partial^2}{\partial t_i \partial t_j}\text{cov}(y_i, y_j)$$

- Using these covariance functions we can combine function observations and derivatives as training data in GP regression.

Derivative observations

Squared exponential kernel

- For the squared exponential kernel we obtain:

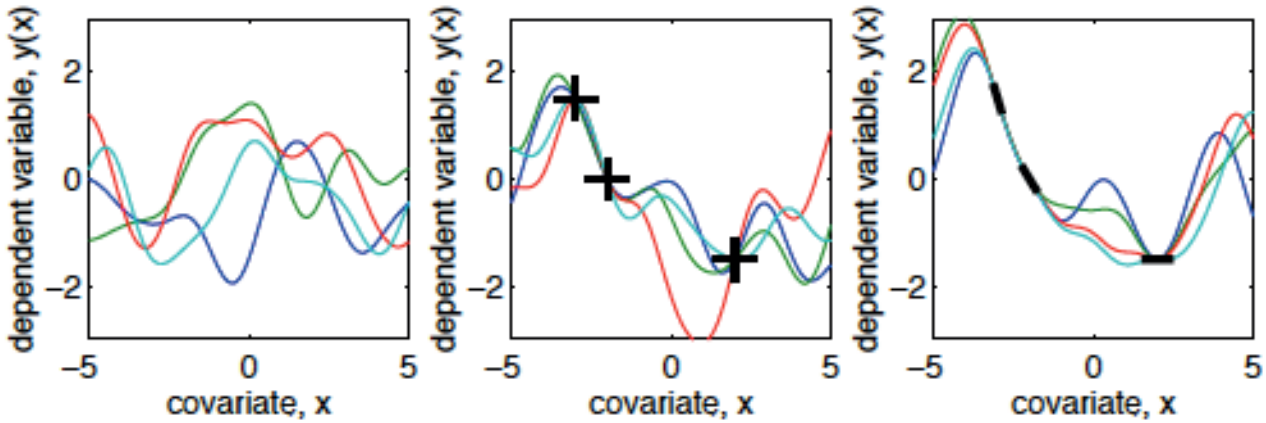
$$\mathrm{cov}(y_i, y_j) = k(t_i, t_j) = A^2 e^{-0.5 \frac{(t_i - t_j)^2}{L^2}}$$

$$\mathrm{cov}(dy_i, y_j) = -\mathrm{cov}(y_i, y_j) \frac{(t_i - t_j)}{L^2}$$

$$\mathrm{cov}(dy_i, dy_j) = \mathrm{cov}(y_i, y_j) \frac{1}{L^2} \left[\delta_{i,j} - \frac{1}{L^2} (t_i - t_j)^2 \right]$$

Derivative observations

Example



(From E. Solak et al.

Derivative observations in Gaussian Process models of dynamic systems)

Back to the ODE model for gene regulation

$$\frac{\mathrm{d}y_i(t)}{\mathrm{d}t} = B_i + S_i f(t) - D_i y_i(t)$$

- An explicit solution of the ODE system can be derived (standard ODE techniques)

$$y_i(t) = \frac{B_i}{D_i} + k_i e^{-D_i t} + S_i e^{-D_i t} \int_0^t f(u) \exp(D_i u) \mathrm{d}u$$

$$y_i(t) = \frac{B_i}{D_i} + L_i[f](t)$$

- Realizing that L_i is a linear operator (like taking the derivative), we can again evaluate the covariance between $f(t)$ and $L_i[f](t)$.

ODE model for gene regulation

Inference results

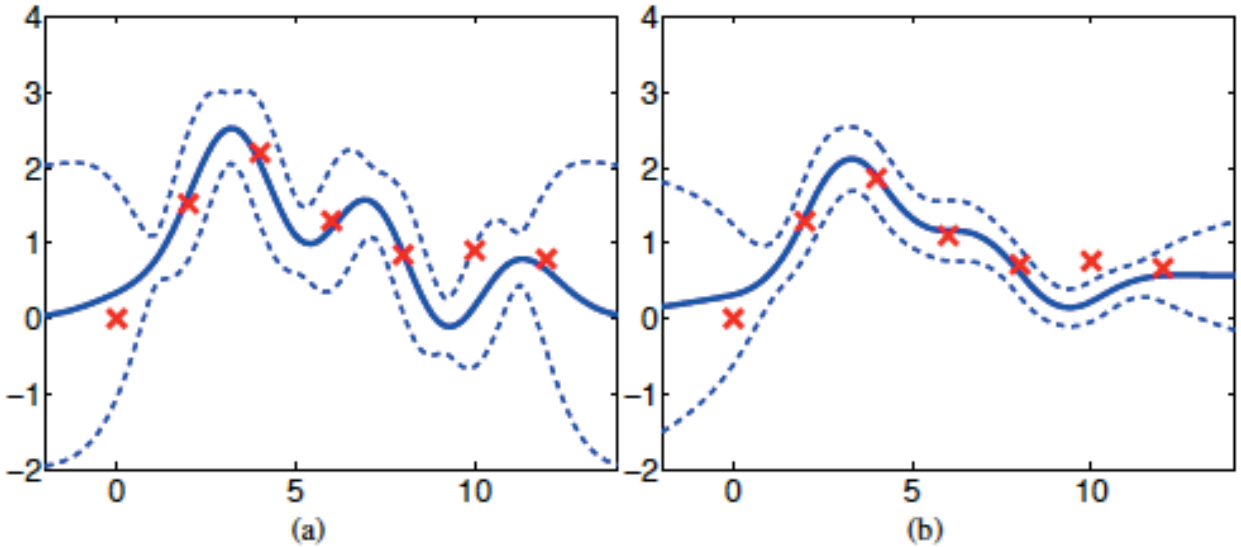


Figure 1: Predicted protein concentration for p53 using a linear response model: (a) squared exponential prior on f; (b) MLP prior on f. Solid line is mean prediction, dashed lines are 95% credibility intervals. The prediction of Barenco *et al.* was pointwise and is shown as crosses.

(From N. D. Lawrence et al.

Modelling transcriptional regulation using Gaussian processes)

Outline

Summary

- The design and choice of covariance functions allows for flexible modeling tasks.
  - Prior on heart rate
  - Derivative observations, ODE systems
- Model comparison using Gaussian processes.
  - Testing for differential gene expression.