

Model Based Target Identification from Gene Expression with Gaussian Processes

Neil D. Lawrence

work with Magnus Rattray (co-PI)

Post-docs: Pei Gao, Guido Sanguinetti

Visitor: Antti Honkela

Sheffield Institute of Translational Neuroscience and Department of Computer
Science, University of Sheffield, U.K.

ABCD 2011, Ravenna, Italy

10th September 2011

Outline

Motivation

Cascade Differential Equations

Discussion and Future Work

Outline

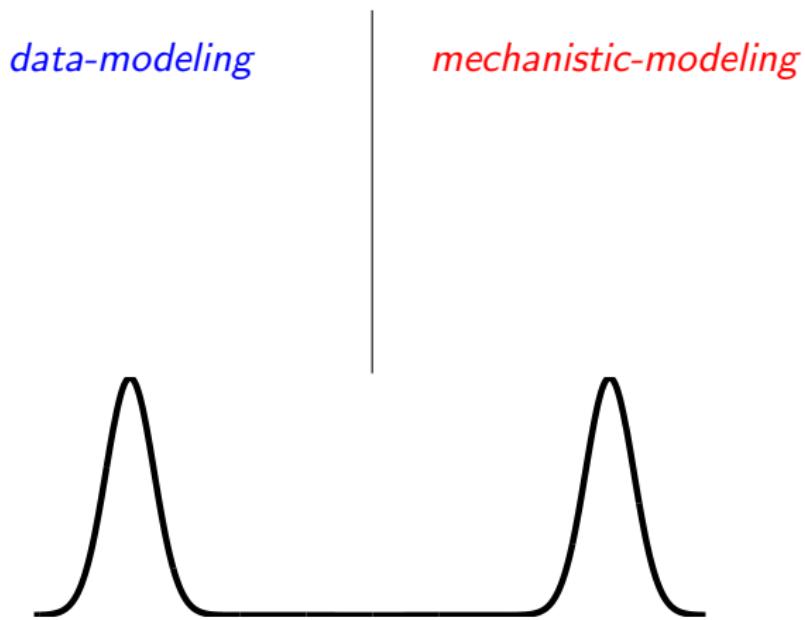
Motivation

Cascade Differential Equations

Discussion and Future Work

Computational Biology vs Computational Systems Biology

Broadly Speaking: Two approaches to modeling



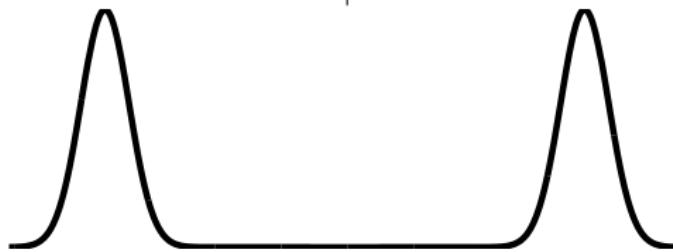
Computational Biology vs Computational Systems Biology

Broadly Speaking: Two approaches to modeling

data-modeling

let the data “speak”

mechanistic-modeling



Computational Biology vs Computational Systems Biology

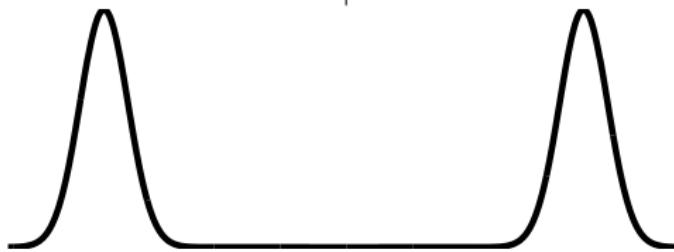
Broadly Speaking: Two approaches to modeling

data-modeling

let the data “speak”

mechanistic-modeling

impose physical laws



Computational Biology vs Computational Systems Biology

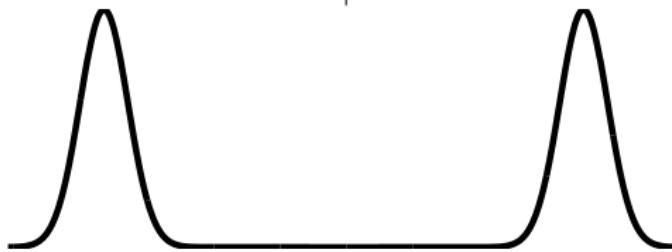
Broadly Speaking: Two approaches to modeling

data-modeling

let the data “speak”
computational models

mechanistic-modeling

impose physical laws



Computational Biology vs Computational Systems Biology

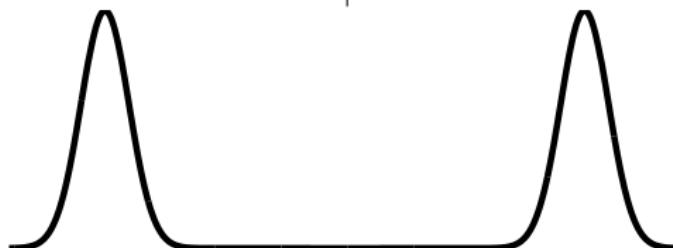
Broadly Speaking: Two approaches to modeling

data-modeling

let the data “speak”
computational models

mechanistic-modeling

impose physical laws
systems models



Computational Biology vs Computational Systems Biology

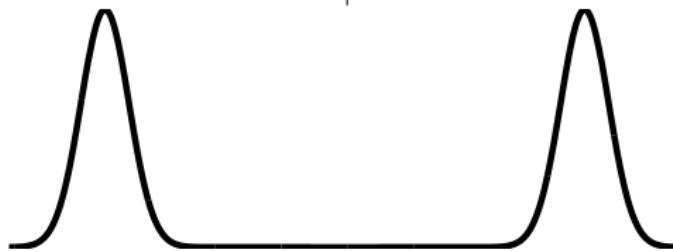
Broadly Speaking: Two approaches to modeling

data-modeling

let the data “speak”
computational models
adaptive models

mechanistic-modeling

impose physical laws
systems models



Computational Biology vs Computational Systems Biology

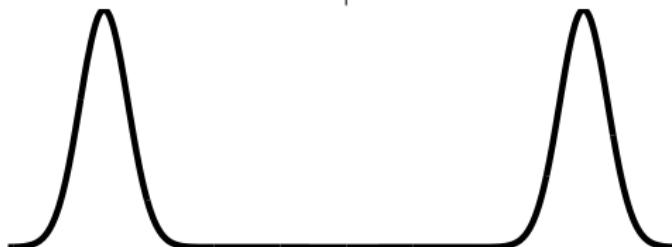
Broadly Speaking: Two approaches to modeling

data-modeling

let the data “speak”
computational models
adaptive models

mechanistic-modeling

impose physical laws
systems models
differential equations



Computational Biology vs Computational Systems Biology

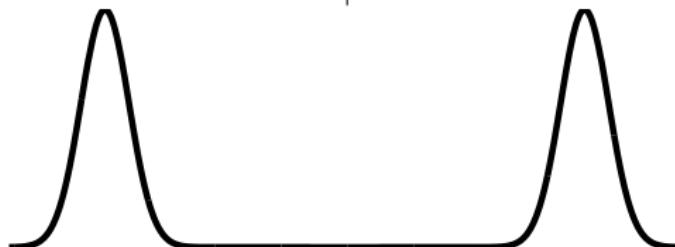
Broadly Speaking: Two approaches to modeling

data-modeling

let the data “speak”
computational models
adaptive models
PCA, clustering

mechanistic-modeling

impose physical laws
systems models
differential equations



Computational Biology vs Computational Systems Biology

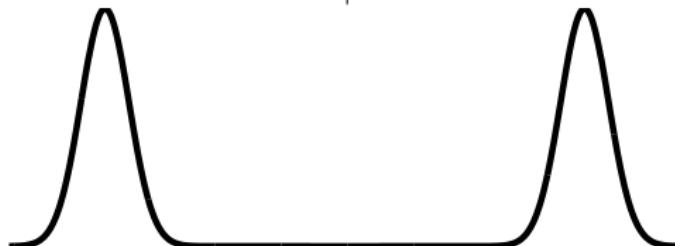
Broadly Speaking: Two approaches to modeling

data-modeling

let the data “speak”
computational models
adaptive models
PCA, clustering

mechanistic-modeling

impose physical laws
systems models
differential equations
SDE, ODE models



Computational Biology vs Computational Systems Biology

Broadly Speaking: Two approaches to modeling

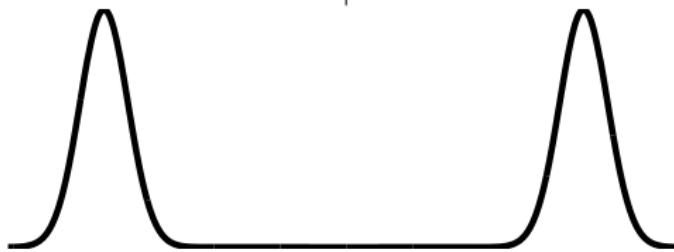
data-modeling

let the data "speak"
computational models
adaptive models
ICA, clustering

Weakly Mechanistic

mechanistic-modeling

impose physical laws
systems models
differential equations
SDE, ODE models



Computational Biology vs Computational Systems Biology

Broadly Speaking: Two approaches to modeling

data-modeling

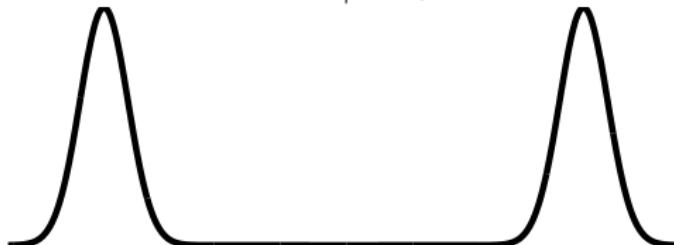
let the data "speak"
computational models
adaptive models
ICA, clustering

Weakly Mechanistic

mechanistic-modeling

impose physical laws
systems models
differential equations
SDE, ODE models

Strongly Mechanistic



A Hybrid Approach

Introduce aspects of systems biology to computational models

- ▶ We advocate an approach *between* systems and computational biology.
- ▶ Introduce aspects of systems biology to the computational approach.
 - ▶ There is a computational penalty, but it may be worth paying.
 - ▶ Ideally there should be a smooth transition from pure computational (PCA, clustering, SVM classification) to systems (non-linear (stochastic) differential equations).

Standard Approach

Clustering of Gene Expression Profiles

- ▶ Assume that coregulated genes will cluster in the same groups.
- ▶ Perform clustering, and look for clusters containing target genes.
- ▶ These are candidates, look for confirmation in the literature etc.

Mathematical Model

- ▶ Differential equation model of system.

$$\frac{dm_j(t)}{dt} = b_j + s_j p(t) - d_j m_j(t)$$

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

- ▶ We have observations of $m_j(t)$ from gene expression.

Mathematical Model

- ▶ Differential equation model of system.

$$\frac{dm_j(t)}{dt} = b_j + s_j p(t) - d_j m_j(t)$$
$$d_j m_j(t) + \frac{dm_j(t)}{dt} = b_j + s_j p(t)$$

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

- ▶ We have observations of $m_j(t)$ from gene expression.

Mathematical Model

- ▶ Clustering model is equivalent to assuming d_j , b_j , and s_j are v. large.

$$\frac{dm_j(t)}{dt} = b_j + s_j p(t) - d_j m_j(t)$$

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

- ▶ We have observations of $m_j(t)$ from gene expression.

Mathematical Model

- ▶ Clustering model is equivalent to assuming d_j , b_j , and s_j are v. large.

$$\frac{dm_j(t)}{dt} = b_j + s_j p(t) - d_j m_j(t)$$
$$d_j m_j(t) \approx b_j + s_j p(t)$$

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

- ▶ We have observations of $m_j(t)$ from gene expression.
- ▶ Reorder differential equation and ignore gradient term.

Mathematical Model

- ▶ Clustering model is equivalent to assuming d_j , b_j , and s_j are v. large.

$$\frac{dm_j(t)}{dt} = b_j + s_j p(t) - d_j m_j(t)$$
$$d_j m_j(t) \approx b_j + s_j p(t)$$

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

- ▶ We have observations of $m_j(t)$ from gene expression.
- ▶ Reorder differential equation and ignore gradient term.
- ▶ This suggests genes are scaled and offset versions of the TF.

Mathematical Model

- ▶ Clustering model is equivalent to assuming d_j , b_j , and s_j are v. large.

$$\frac{dm_j(t)}{dt} = b_j + s_j p(t) - d_j m_j(t)$$
$$d_j m_j(t) \approx b_j + s_j p(t)$$

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

- ▶ We have observations of $m_j(t)$ from gene expression.
- ▶ Reorder differential equation and ignore gradient term.
- ▶ This suggests genes are scaled and offset versions of the TF.
- ▶ By normalizing data and clustering we hope to find those TFs.

Mathematical Model

Method

Open Access

Ranked prediction of p53 targets using hidden variable dynamic modeling

Martino Barenco^{*†}, Daniela Tomescu^{*}, Daniel Brewer^{*†}, Robin Callard^{*†}, Jaroslav Stark^{†‡} and Michael Hubank^{*†}

Addresses: ^{*}Institute of Child Health, University College London, Guilford Street, London WC1N 1EH, UK. [†]CoMPLEX (Centre for Mathematics and Physics in the Life Sciences and Experimental Biology), University College London, Stephenson Way, London, NW1 2HE, UK. [‡]Department of Mathematics, Imperial College London, London SW7 2AZ, UK.

Correspondence: Michael Hubank. Email: m.hubank@ich.ucl.ac.uk

Published: 31 March 2006

Genome Biology 2006, **7**:R25 (doi:10.1186/gb-2006-7-3-r25)

Received: 24 November 2005

Revised: 30 January 2006

Accepted: 21 February 2006

Example: Transcriptional Regulation

- ▶ First Order Differential Equation

$$\frac{dm_j(t)}{dt} = b_j + s_j p(t) - d_j m_j(t)$$

- ▶ We can use a probabilistic process model for $p(t)$.
- ▶ This implies a probabilistic process over $p(t)$ and all its targets: $m_1(t), m_2(t), \dots$ etc.
- ▶ This incorporates the assumptions in the differential equation.
- ▶ This gives us a *probabilistic* model for transcriptional regulation.

Example: Transcriptional Regulation

- ▶ First Order Differential Equation

$$\frac{dm_j(t)}{dt} = b_j + s_j p(t) - d_j m_j(t)$$

- ▶ We can use a probabilistic process model for $p(t)$.
- ▶ This implies a probabilistic process over $p(t)$ and all its targets: $m_1(t), m_2(t), \dots$ etc.
- ▶ This incorporates the assumptions in the differential equation.
- ▶ This gives us a *probabilistic* model for transcriptional regulation.

Example: Transcriptional Regulation

- ▶ First Order Differential Equation

$$\frac{dm_j(t)}{dt} = b_j + s_j p(t) - d_j m_j(t)$$

- ▶ We can use a probabilistic process model for $p(t)$.
- ▶ This implies a probabilistic process over $p(t)$ and all its targets: $m_1(t), m_2(t), \dots$ etc.
- ▶ This incorporates the assumptions in the differential equation.
- ▶ This gives us a *probabilistic* model for transcriptional regulation.

Example: Transcriptional Regulation

- ▶ First Order Differential Equation

$$\frac{dm_j(t)}{dt} = b_j + s_j p(t) - d_j m_j(t)$$

- ▶ We can use a probabilistic process model for $p(t)$.
- ▶ This implies a probabilistic process over $p(t)$ and all its targets: $m_1(t), m_2(t), \dots$ etc.
- ▶ This incorporates the assumptions in the differential equation.
- ▶ This gives us a *probabilistic* model for transcriptional regulation.

Example: Transcriptional Regulation

- ▶ First Order Differential Equation

$$\frac{dm_j(t)}{dt} = b_j + s_j p(t) - d_j m_j(t)$$

- ▶ We can use a probabilistic process model for $p(t)$.
- ▶ This implies a probabilistic process over $p(t)$ and all its targets: $m_1(t), m_2(t), \dots$ etc.
- ▶ This incorporates the assumptions in the differential equation.
- ▶ This gives us a *probabilistic* model for transcriptional regulation.

Joint Sampling of $f(t)$ and $x(t)$

► simSample

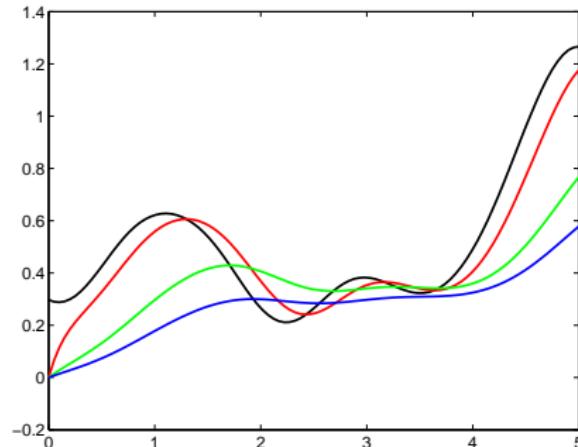


Figure: Joint samples from the ODE covariance, *black*: $p(t)$, *red*: $m_1(t)$ (high decay/sensitivity), *green*: $m_2(t)$ (medium decay/sensitivity) and *blue*: $m_3(t)$ (low decay/sensitivity).

Joint Sampling of $f(t)$ and $x(t)$

► simSample

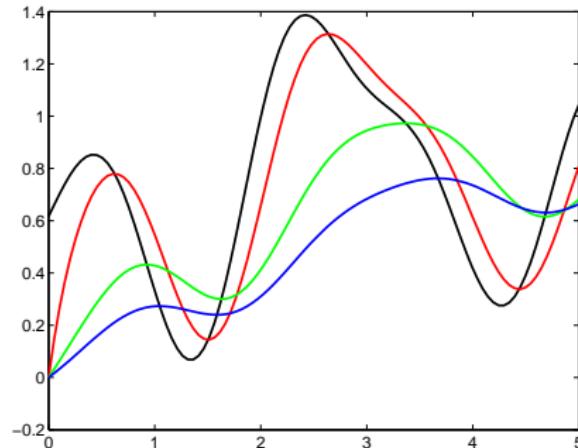


Figure: Joint samples from the ODE covariance, *black*: $p(t)$, *red*: $m_1(t)$ (high decay/sensitivity), *green*: $m_2(t)$ (medium decay/sensitivity) and *blue*: $m_3(t)$ (low decay/sensitivity).

Joint Sampling of $f(t)$ and $x(t)$

► simSample

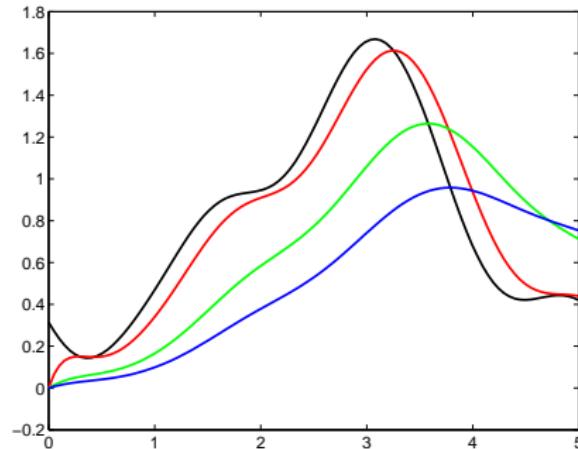


Figure: Joint samples from the ODE covariance, *black*: $p(t)$, *red*: $m_1(t)$ (high decay/sensitivity), *green*: $m_2(t)$ (medium decay/sensitivity) and *blue*: $m_3(t)$ (low decay/sensitivity).

Joint Sampling of $f(t)$ and $x(t)$

► simSample

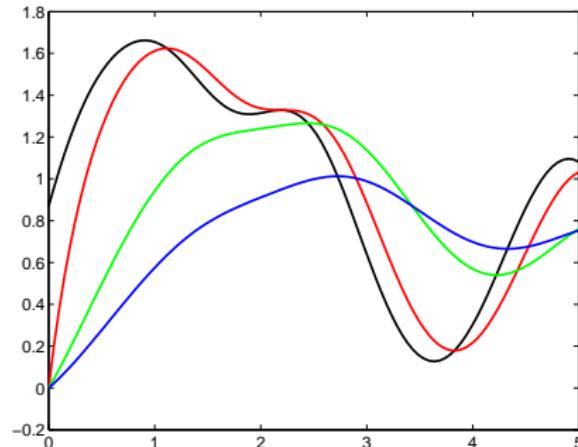
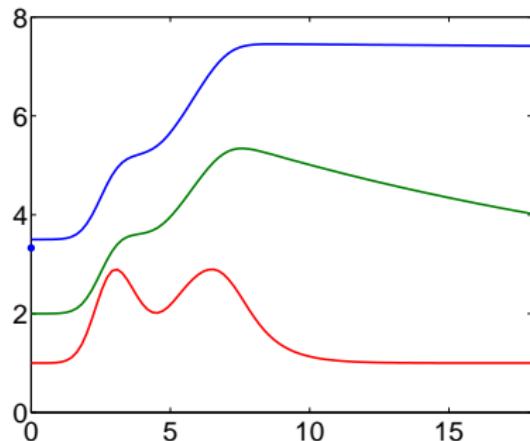


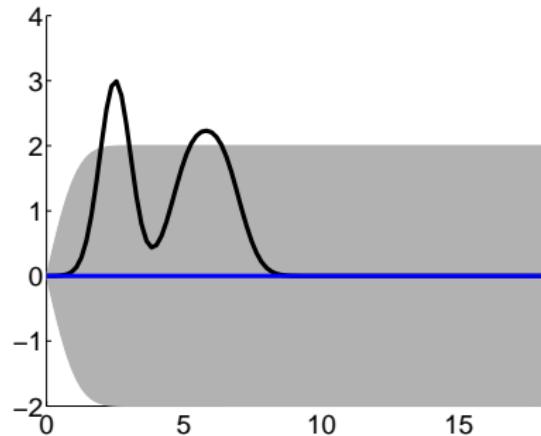
Figure: Joint samples from the ODE covariance, *black*: $p(t)$, *red*: $m_1(t)$ (high decay/sensitivity), *green*: $m_2(t)$ (medium decay/sensitivity) and *blue*: $m_3(t)$ (low decay/sensitivity).

Artificial Example: Inferring $p(t)$

Inferring TF activity from artificially sampled genes.



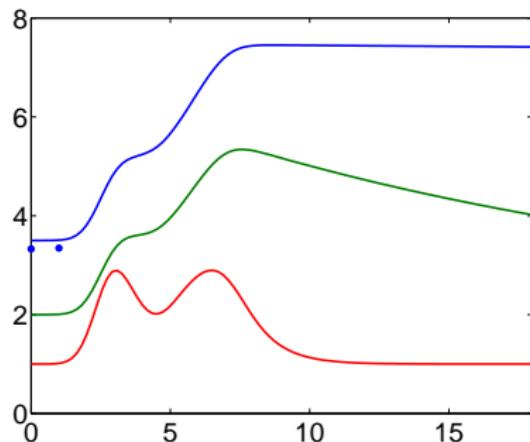
True “gene profiles” and noisy observations.



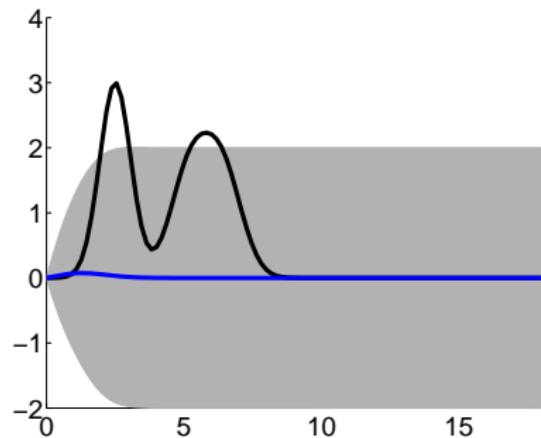
Inferred transcription factor activity.

Artificial Example: Inferring $p(t)$

Inferring TF activity from artificially sampled genes.



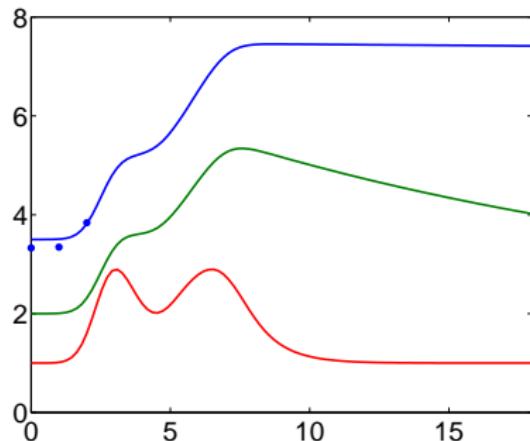
True “gene profiles” and noisy observations.



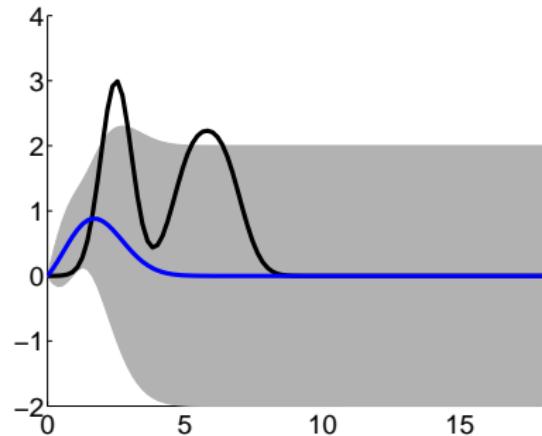
Inferred transcription factor activity.

Artificial Example: Inferring $p(t)$

Inferring TF activity from artificially sampled genes.



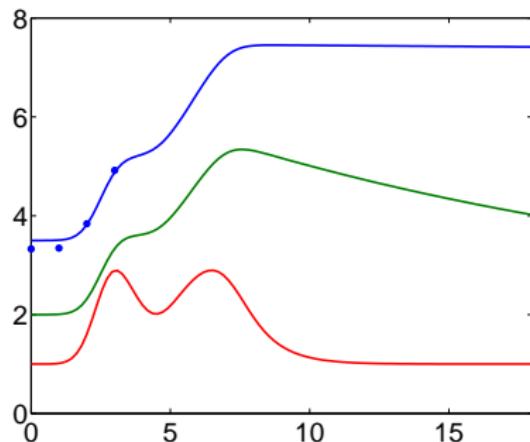
True “gene profiles” and noisy observations.



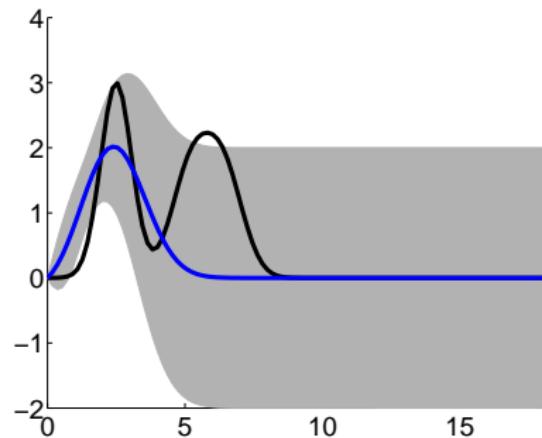
Inferred transcription factor activity.

Artificial Example: Inferring $p(t)$

Inferring TF activity from artificially sampled genes.



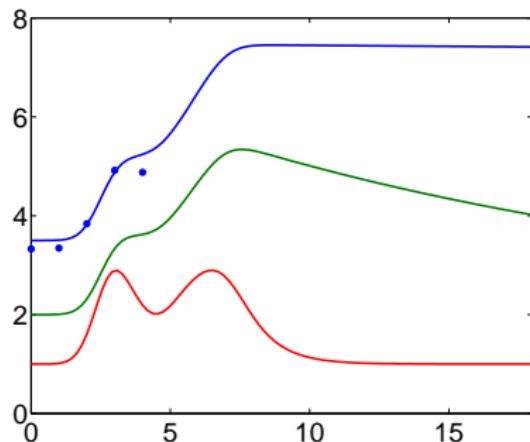
True “gene profiles” and noisy observations.



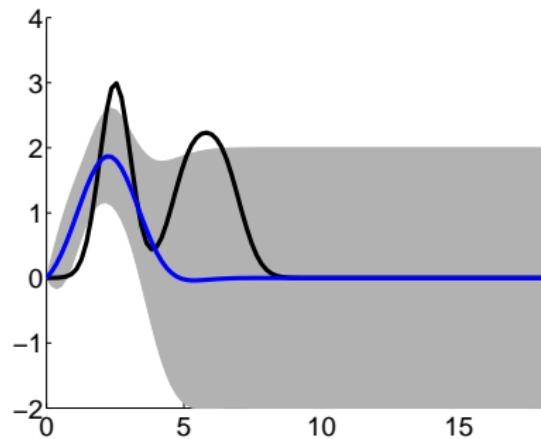
Inferred transcription factor activity.

Artificial Example: Inferring $p(t)$

Inferring TF activity from artificially sampled genes.



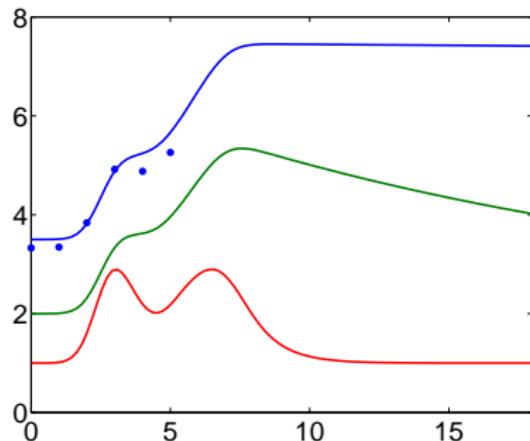
True “gene profiles” and noisy observations.



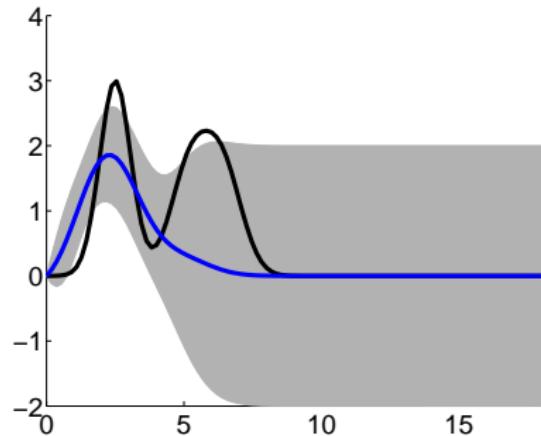
Inferred transcription factor activity.

Artificial Example: Inferring $p(t)$

Inferring TF activity from artificially sampled genes.



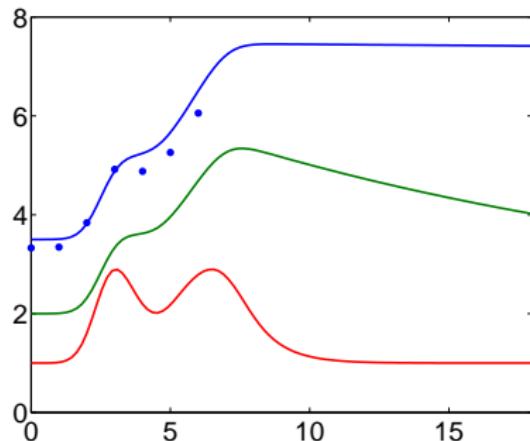
True “gene profiles” and noisy observations.



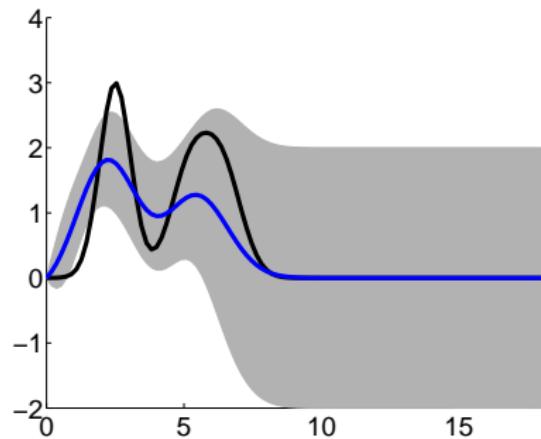
Inferred transcription factor activity.

Artificial Example: Inferring $p(t)$

Inferring TF activity from artificially sampled genes.



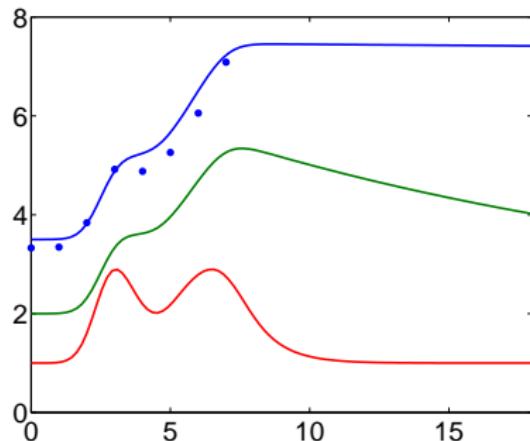
True “gene profiles” and noisy observations.



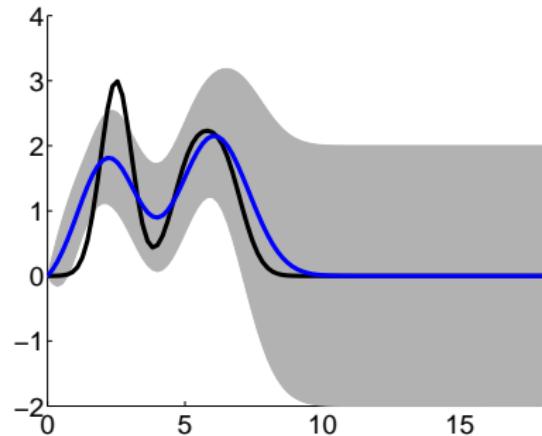
Inferred transcription factor activity.

Artificial Example: Inferring $p(t)$

Inferring TF activity from artificially sampled genes.



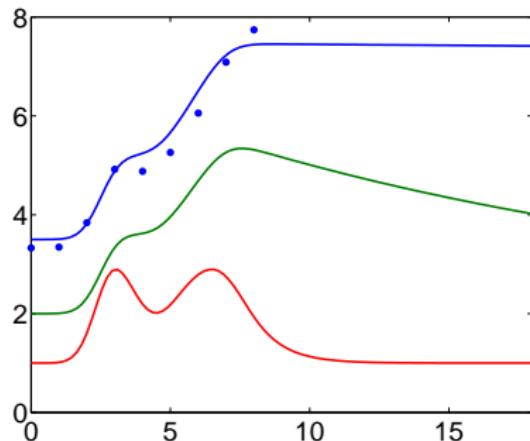
True “gene profiles” and noisy observations.



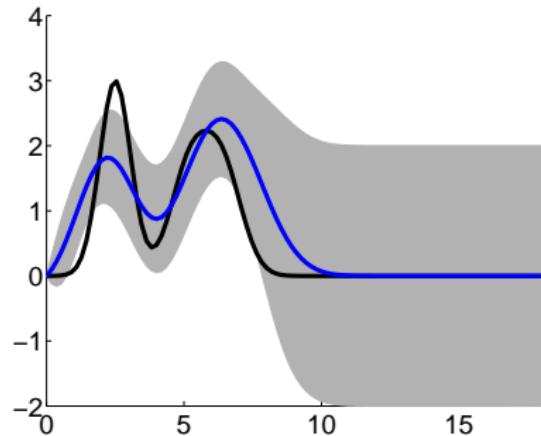
Inferred transcription factor activity.

Artificial Example: Inferring $p(t)$

Inferring TF activity from artificially sampled genes.



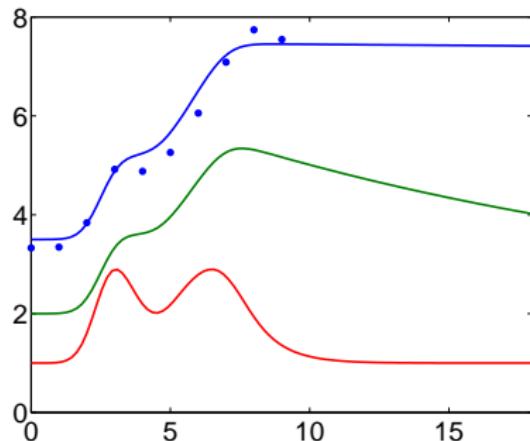
True “gene profiles” and noisy observations.



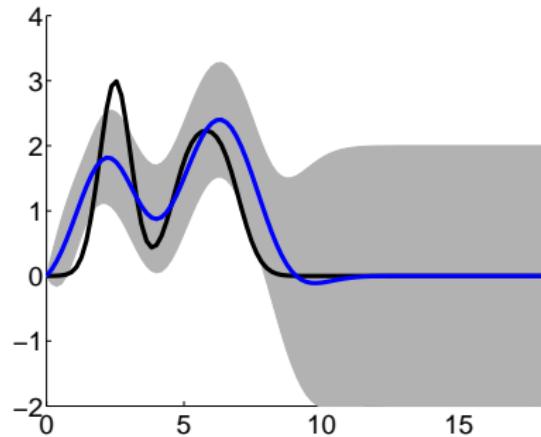
Inferred transcription factor activity.

Artificial Example: Inferring $p(t)$

Inferring TF activity from artificially sampled genes.



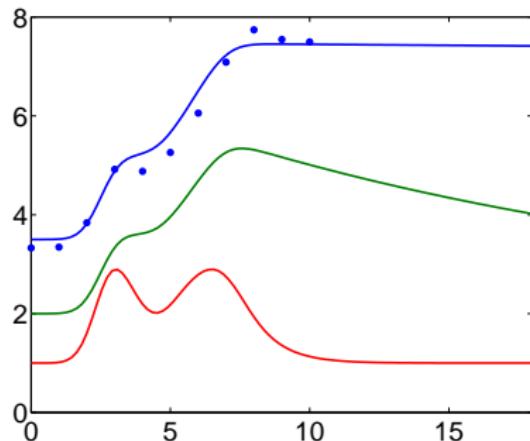
True “gene profiles” and noisy observations.



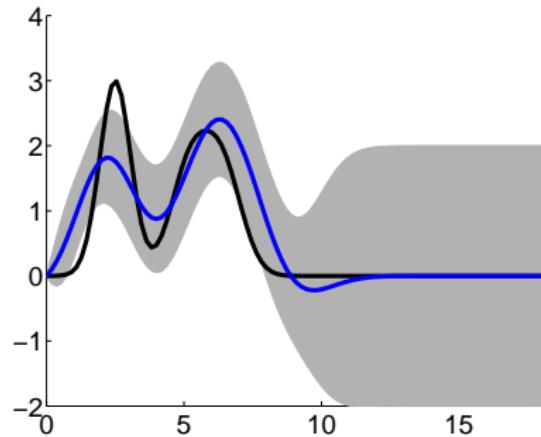
Inferred transcription factor activity.

Artificial Example: Inferring $p(t)$

Inferring TF activity from artificially sampled genes.



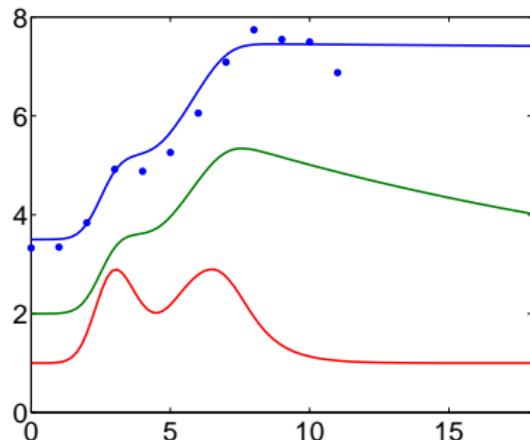
True “gene profiles” and noisy observations.



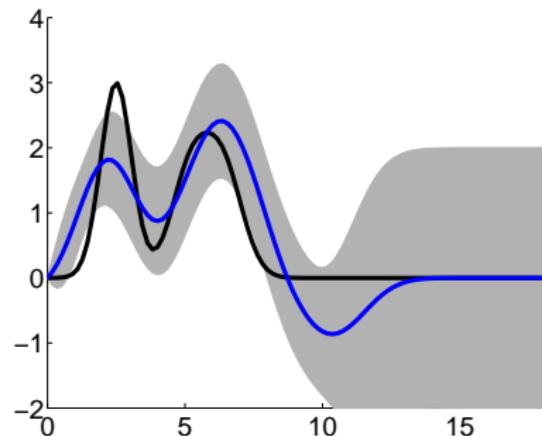
Inferred transcription factor activity.

Artificial Example: Inferring $p(t)$

Inferring TF activity from artificially sampled genes.



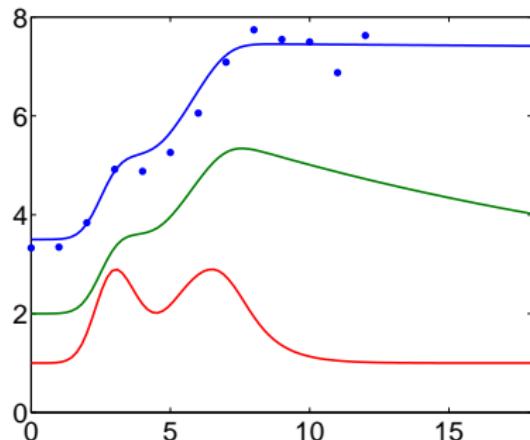
True “gene profiles” and noisy observations.



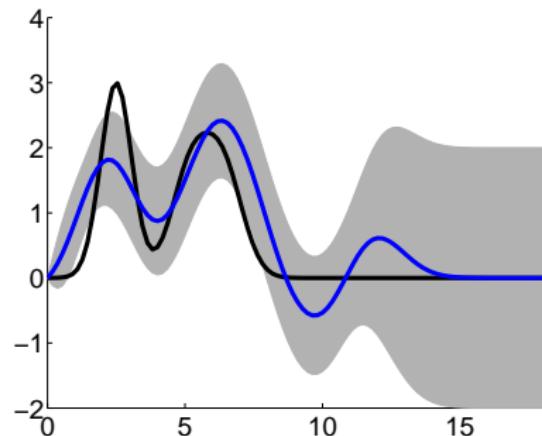
Inferred transcription factor activity.

Artificial Example: Inferring $p(t)$

Inferring TF activity from artificially sampled genes.



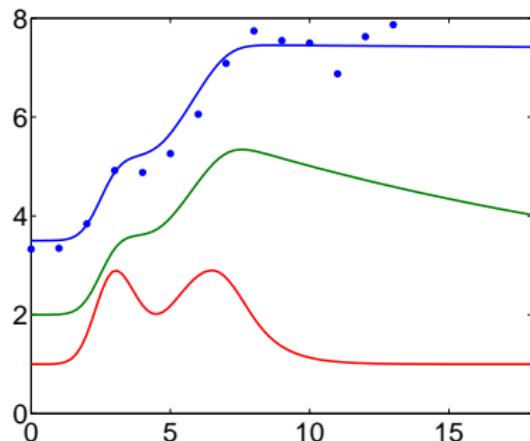
True “gene profiles” and noisy observations.



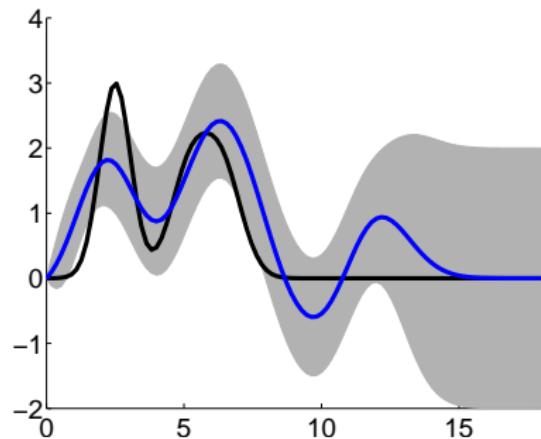
Inferred transcription factor activity.

Artificial Example: Inferring $p(t)$

Inferring TF activity from artificially sampled genes.



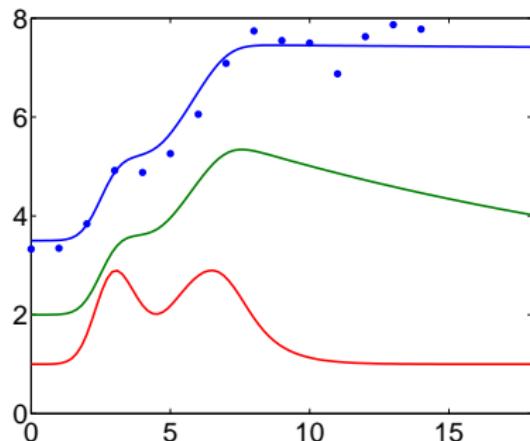
True “gene profiles” and noisy observations.



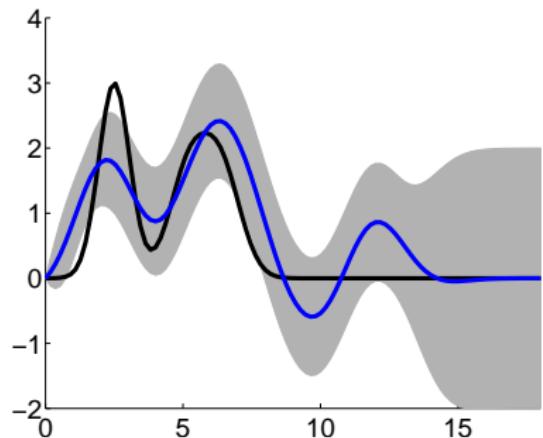
Inferred transcription factor activity.

Artificial Example: Inferring $p(t)$

Inferring TF activity from artificially sampled genes.



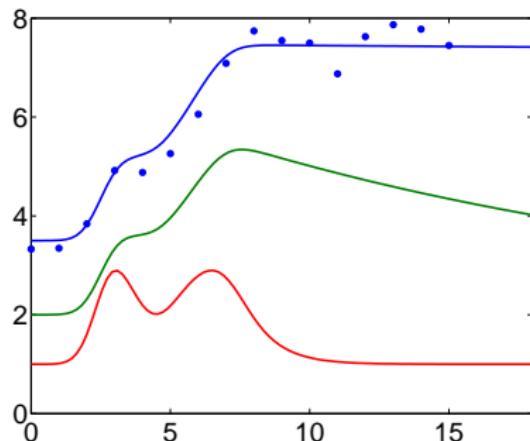
True “gene profiles” and noisy observations.



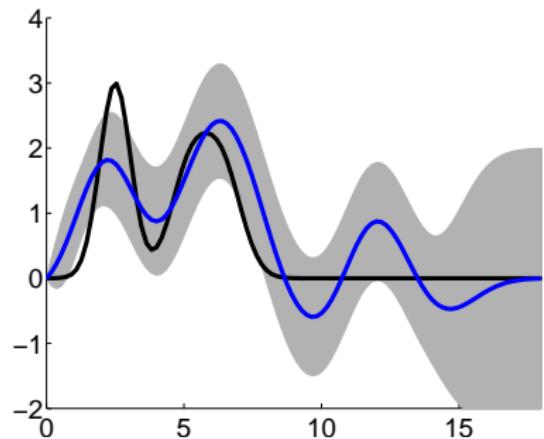
Inferred transcription factor activity.

Artificial Example: Inferring $p(t)$

Inferring TF activity from artificially sampled genes.



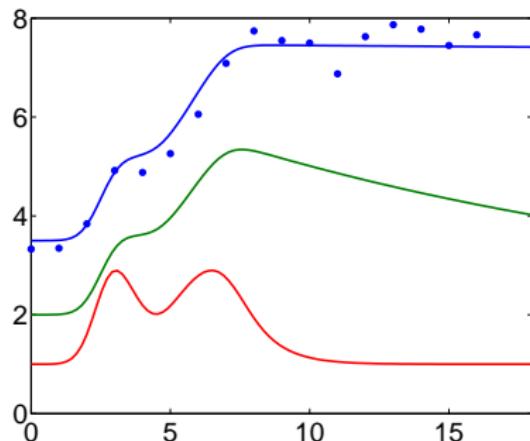
True “gene profiles” and noisy observations.



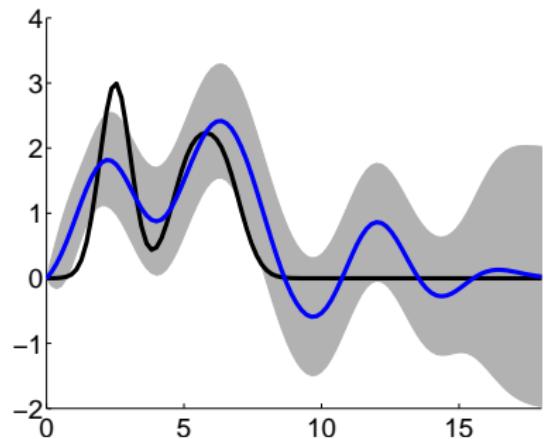
Inferred transcription factor activity.

Artificial Example: Inferring $p(t)$

Inferring TF activity from artificially sampled genes.



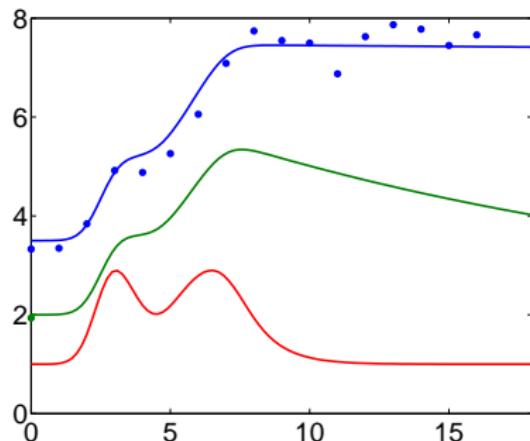
True “gene profiles” and noisy observations.



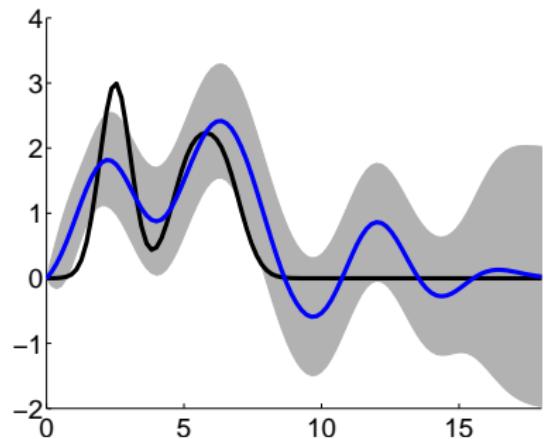
Inferred transcription factor activity.

Artificial Example: Inferring $p(t)$

Inferring TF activity from artificially sampled genes.



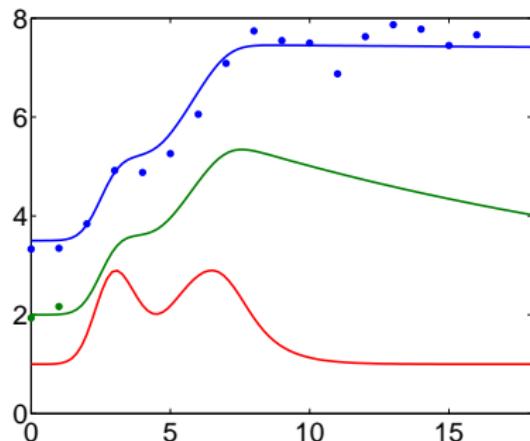
True “gene profiles” and noisy observations.



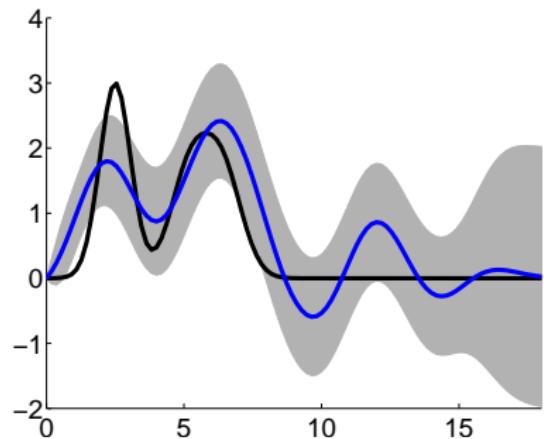
Inferred transcription factor activity.

Artificial Example: Inferring $p(t)$

Inferring TF activity from artificially sampled genes.



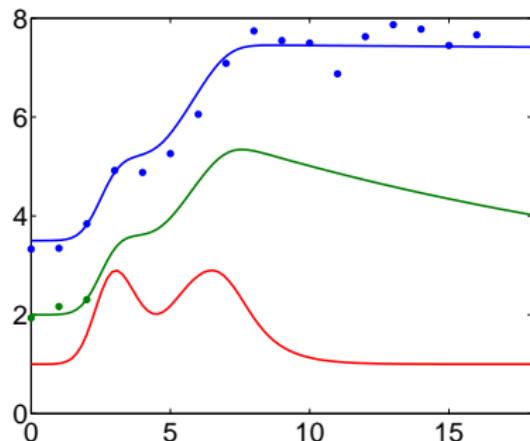
True “gene profiles” and noisy observations.



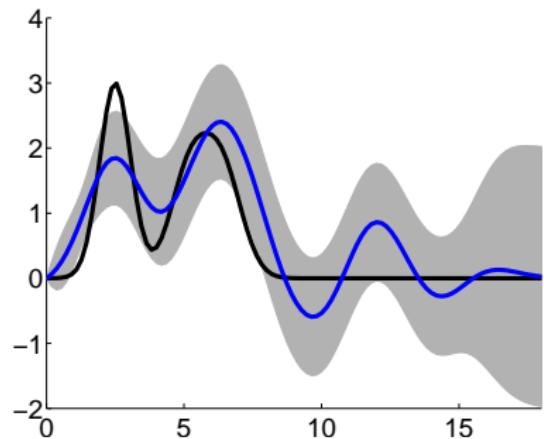
Inferred transcription factor activity.

Artificial Example: Inferring $p(t)$

Inferring TF activity from artificially sampled genes.



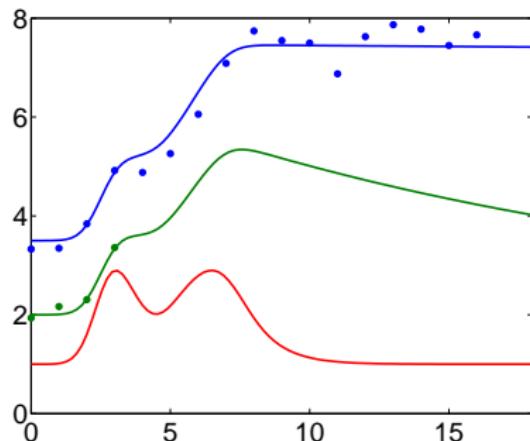
True “gene profiles” and noisy observations.



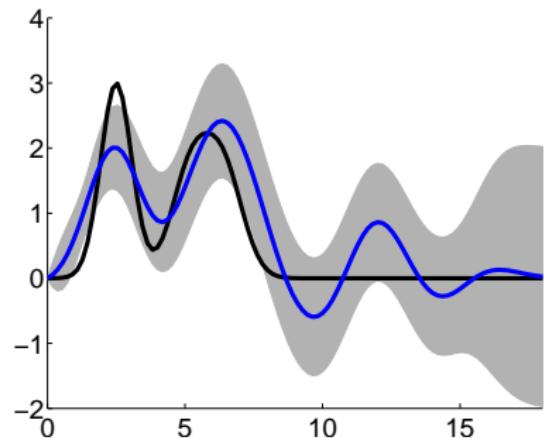
Inferred transcription factor activity.

Artificial Example: Inferring $p(t)$

Inferring TF activity from artificially sampled genes.



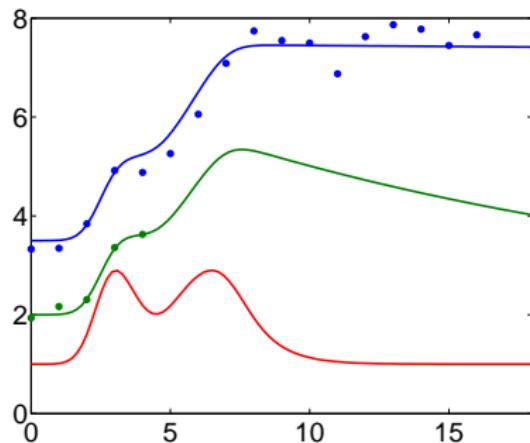
True “gene profiles” and noisy observations.



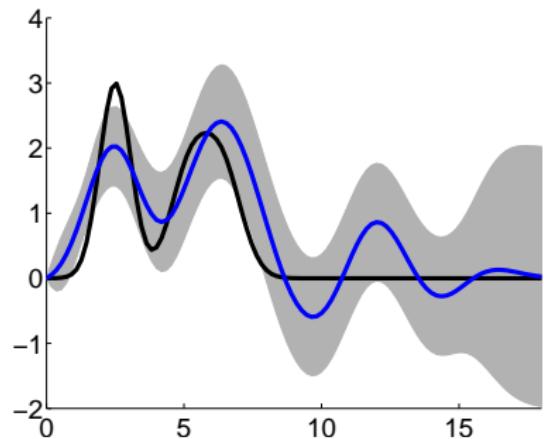
Inferred transcription factor activity.

Artificial Example: Inferring $p(t)$

Inferring TF activity from artificially sampled genes.



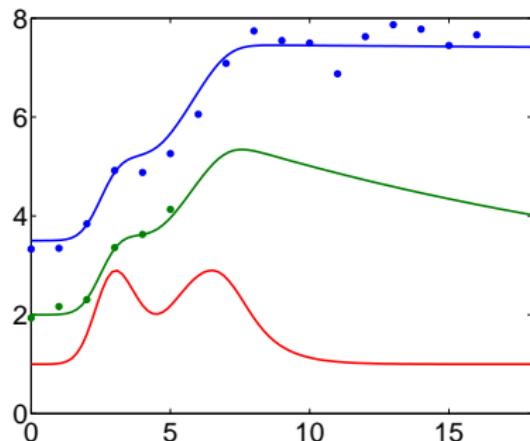
True “gene profiles” and noisy observations.



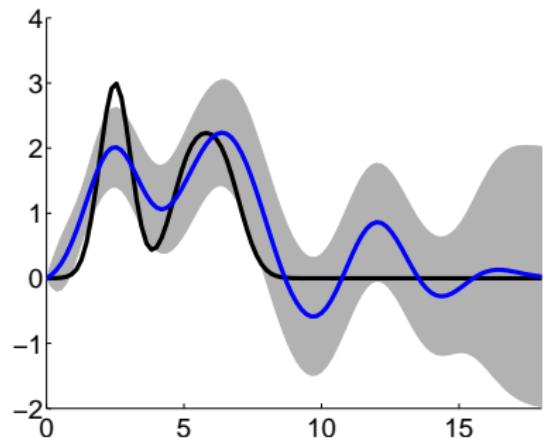
Inferred transcription factor activity.

Artificial Example: Inferring $p(t)$

Inferring TF activity from artificially sampled genes.



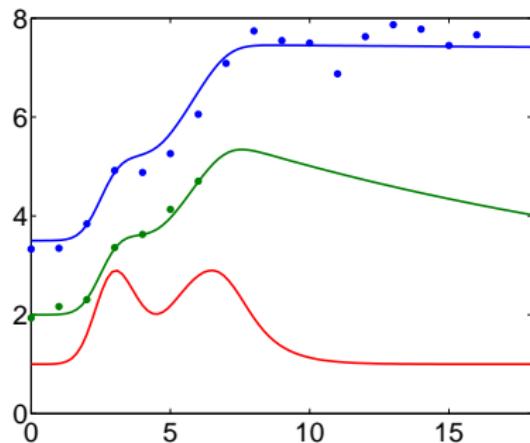
True “gene profiles” and noisy observations.



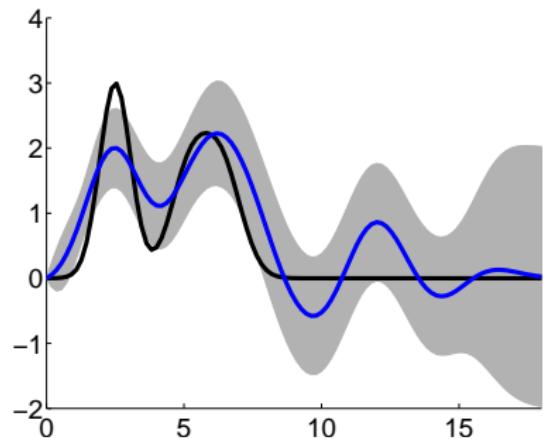
Inferred transcription factor activity.

Artificial Example: Inferring $p(t)$

Inferring TF activity from artificially sampled genes.



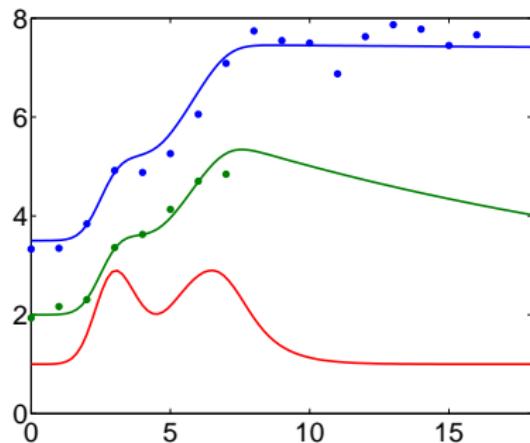
True “gene profiles” and noisy observations.



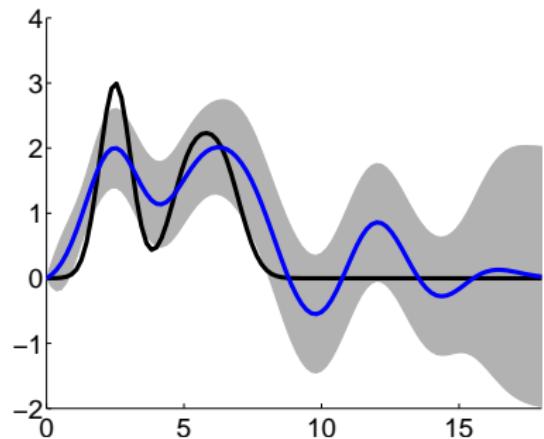
Inferred transcription factor activity.

Artificial Example: Inferring $p(t)$

Inferring TF activity from artificially sampled genes.



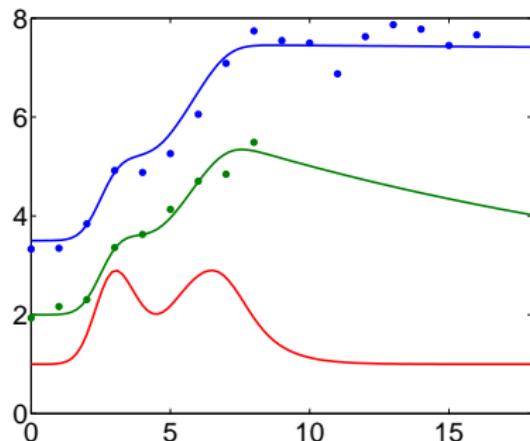
True “gene profiles” and noisy observations.



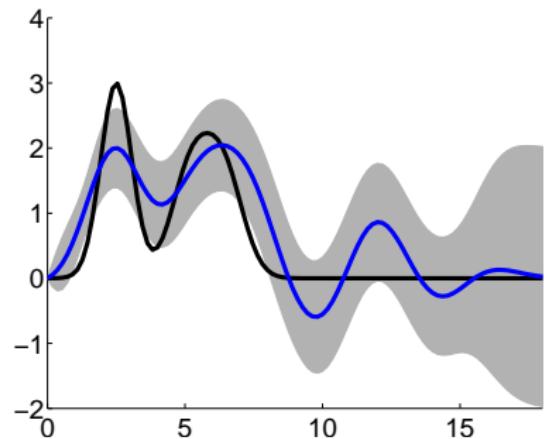
Inferred transcription factor activity.

Artificial Example: Inferring $p(t)$

Inferring TF activity from artificially sampled genes.



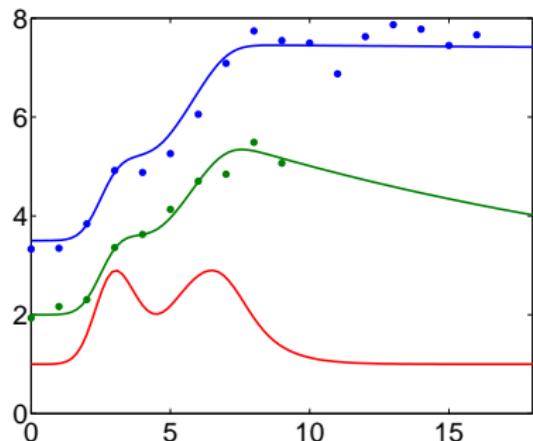
True “gene profiles” and noisy observations.



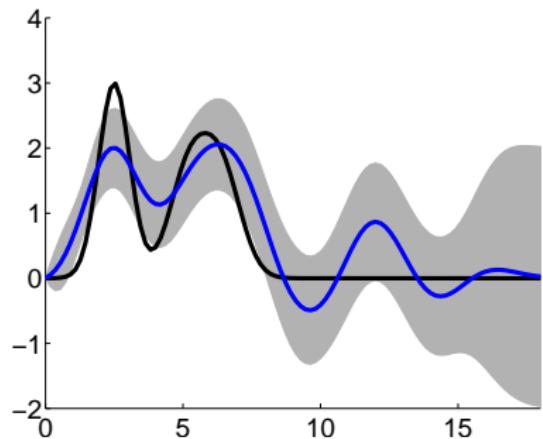
Inferred transcription factor activity.

Artificial Example: Inferring $p(t)$

Inferring TF activity from artificially sampled genes.



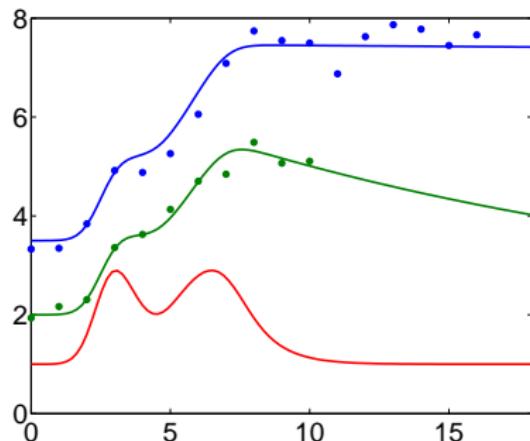
True “gene profiles” and noisy observations.



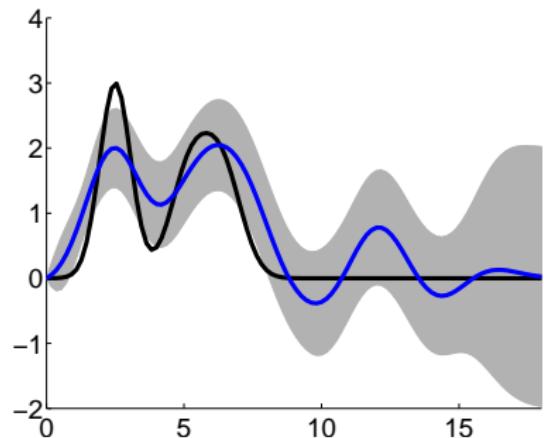
Inferred transcription factor activity.

Artificial Example: Inferring $p(t)$

Inferring TF activity from artificially sampled genes.



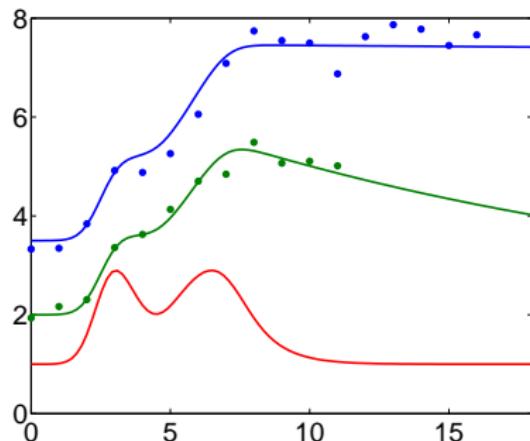
True “gene profiles” and noisy observations.



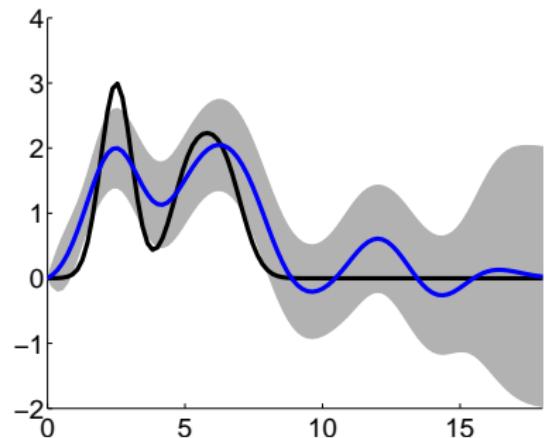
Inferred transcription factor activity.

Artificial Example: Inferring $p(t)$

Inferring TF activity from artificially sampled genes.



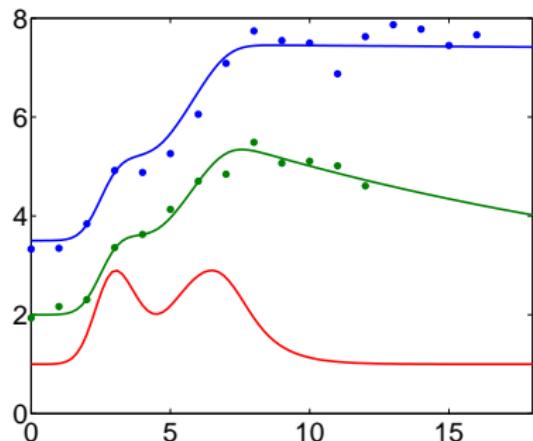
True “gene profiles” and noisy observations.



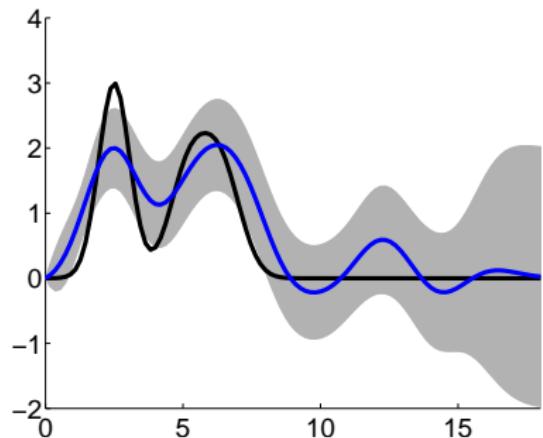
Inferred transcription factor activity.

Artificial Example: Inferring $p(t)$

Inferring TF activity from artificially sampled genes.



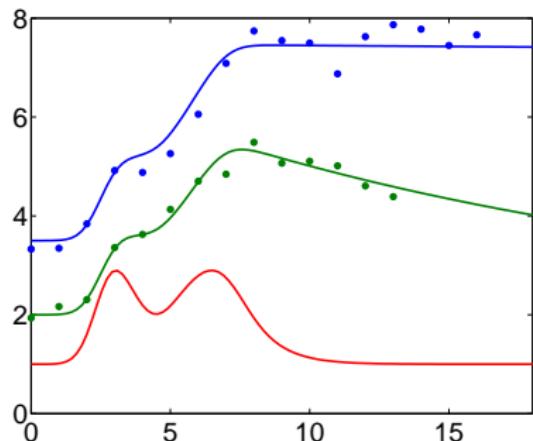
True “gene profiles” and noisy observations.



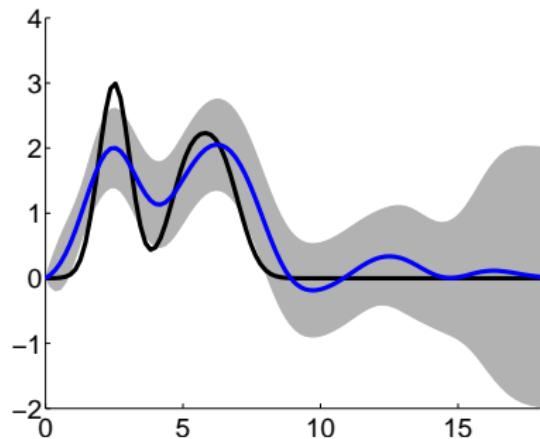
Inferred transcription factor activity.

Artificial Example: Inferring $p(t)$

Inferring TF activity from artificially sampled genes.



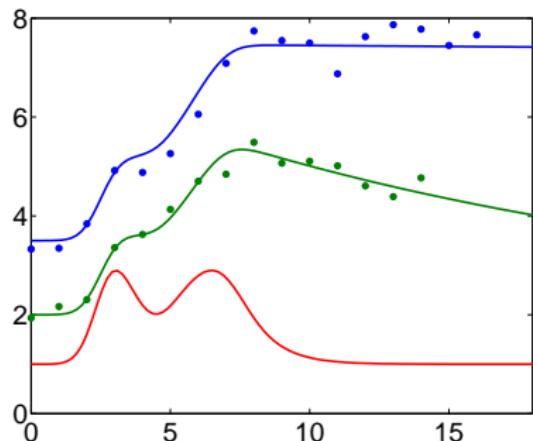
True “gene profiles” and noisy observations.



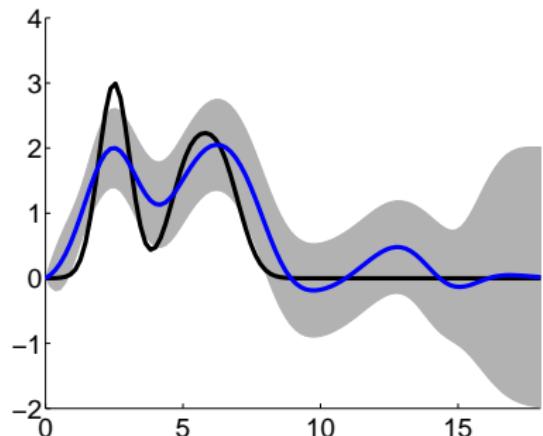
Inferred transcription factor activity.

Artificial Example: Inferring $p(t)$

Inferring TF activity from artificially sampled genes.



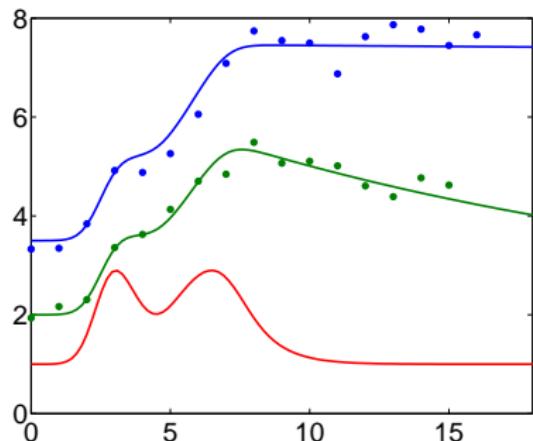
True “gene profiles” and noisy observations.



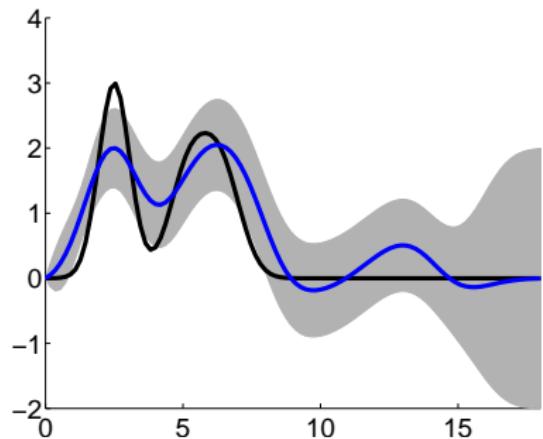
Inferred transcription factor activity.

Artificial Example: Inferring $p(t)$

Inferring TF activity from artificially sampled genes.



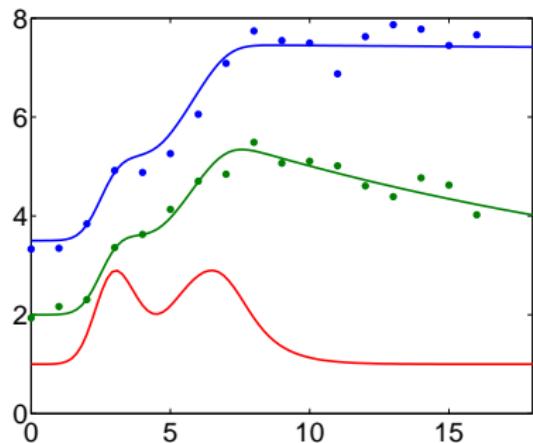
True “gene profiles” and noisy observations.



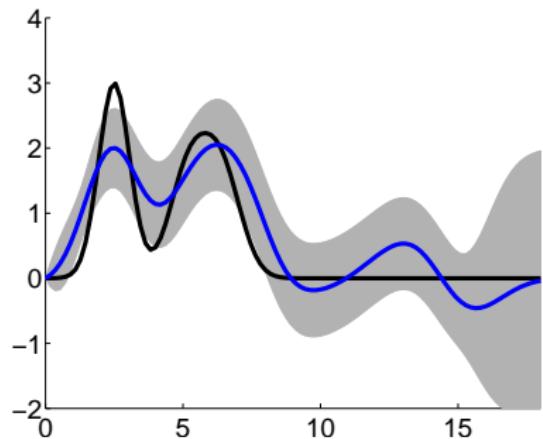
Inferred transcription factor activity.

Artificial Example: Inferring $p(t)$

Inferring TF activity from artificially sampled genes.



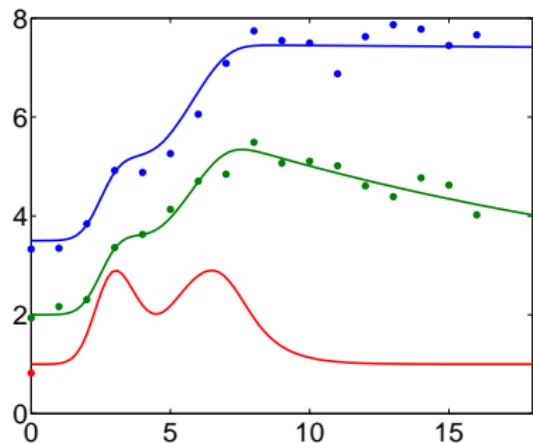
True “gene profiles” and noisy observations.



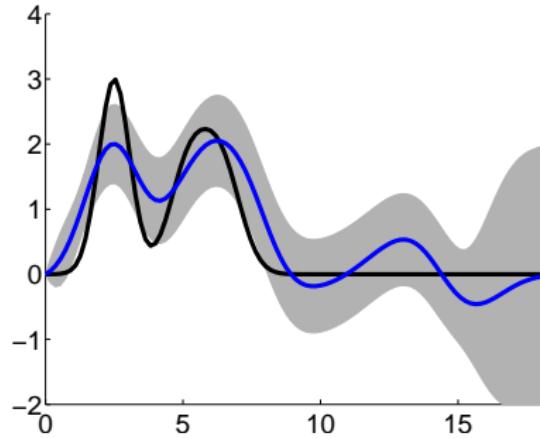
Inferred transcription factor activity.

Artificial Example: Inferring $p(t)$

Inferring TF activity from artificially sampled genes.



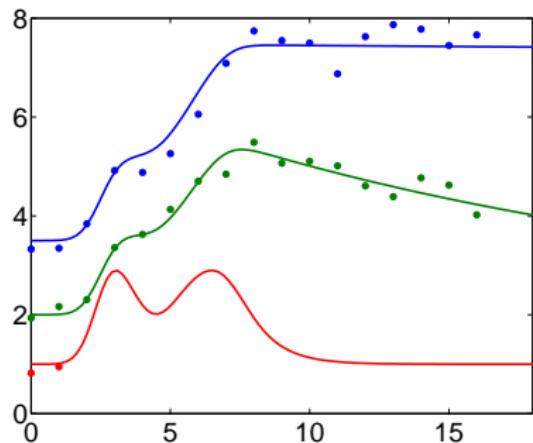
True “gene profiles” and noisy observations.



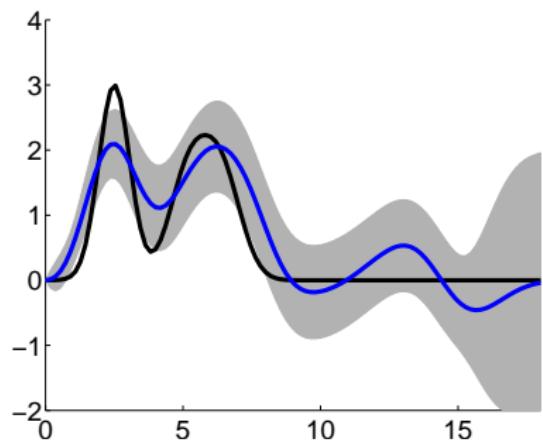
Inferred transcription factor activity.

Artificial Example: Inferring $p(t)$

Inferring TF activity from artificially sampled genes.



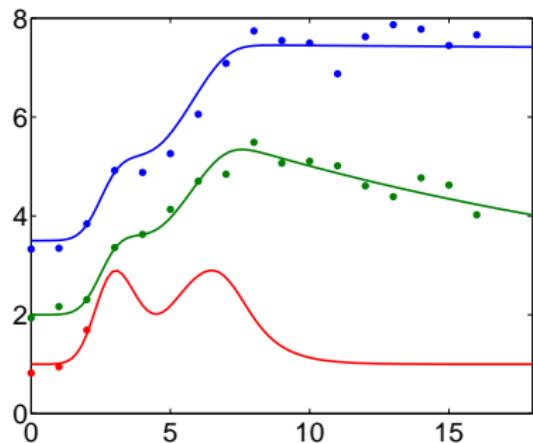
True “gene profiles” and noisy observations.



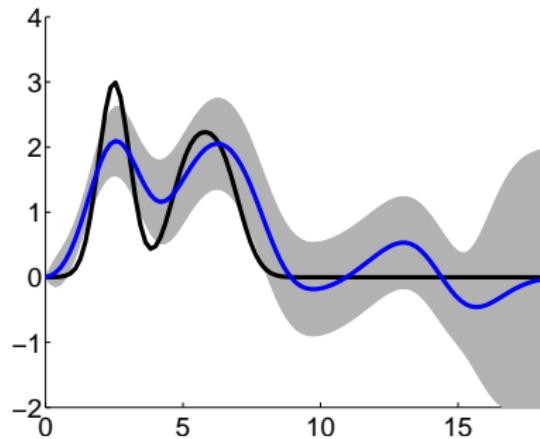
Inferred transcription factor activity.

Artificial Example: Inferring $p(t)$

Inferring TF activity from artificially sampled genes.



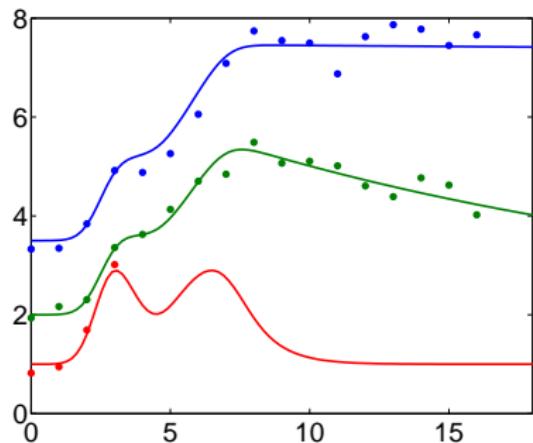
True “gene profiles” and noisy observations.



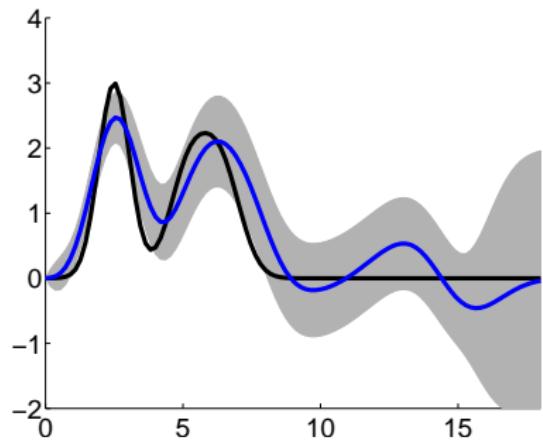
Inferred transcription factor activity.

Artificial Example: Inferring $p(t)$

Inferring TF activity from artificially sampled genes.



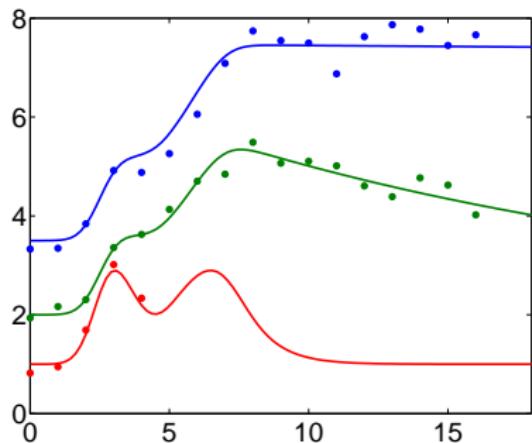
True “gene profiles” and noisy observations.



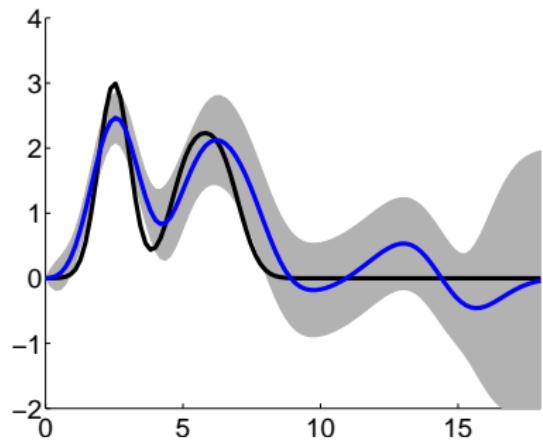
Inferred transcription factor activity.

Artificial Example: Inferring $p(t)$

Inferring TF activity from artificially sampled genes.



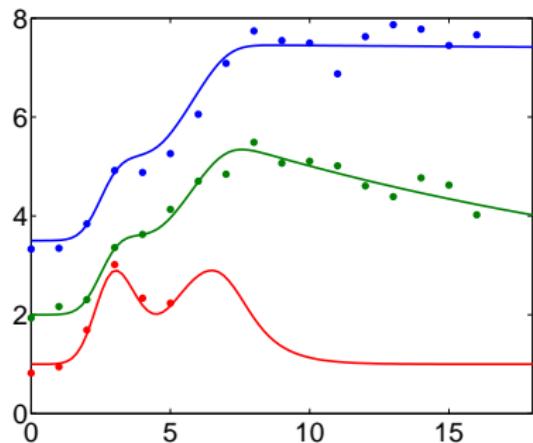
True “gene profiles” and noisy observations.



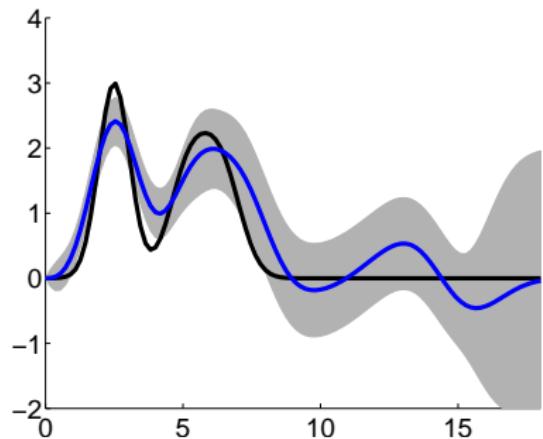
Inferred transcription factor activity.

Artificial Example: Inferring $p(t)$

Inferring TF activity from artificially sampled genes.



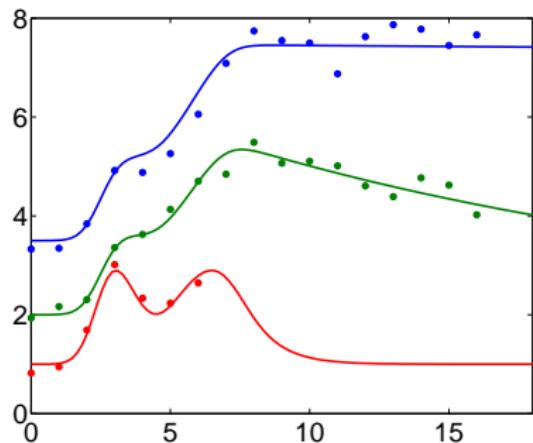
True “gene profiles” and noisy observations.



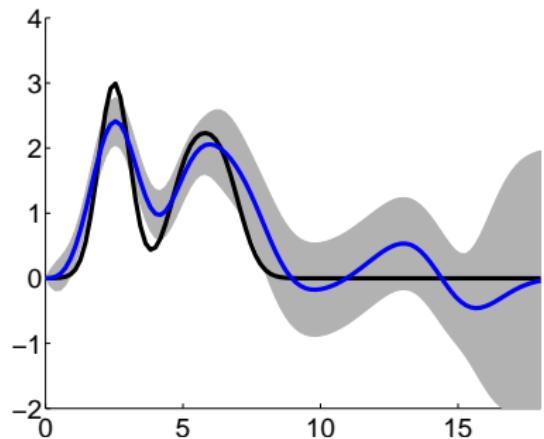
Inferred transcription factor activity.

Artificial Example: Inferring $p(t)$

Inferring TF activity from artificially sampled genes.



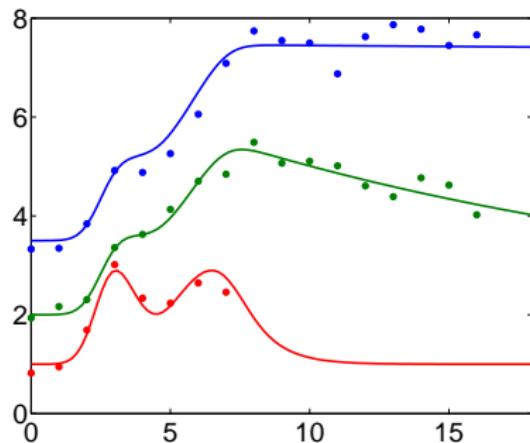
True “gene profiles” and noisy observations.



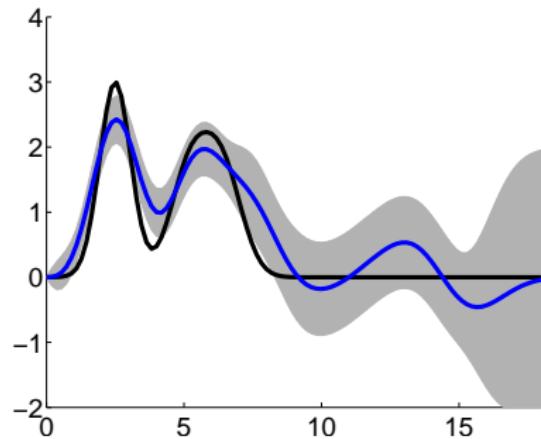
Inferred transcription factor activity.

Artificial Example: Inferring $p(t)$

Inferring TF activity from artificially sampled genes.



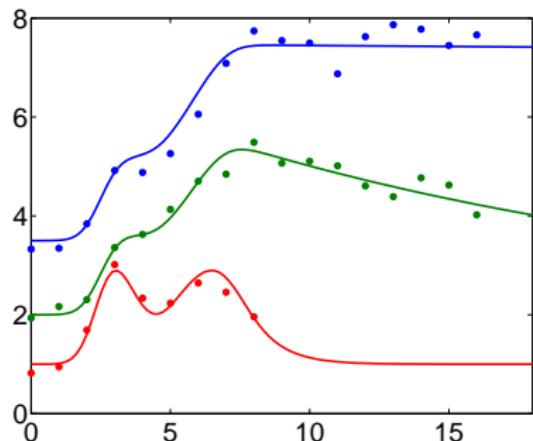
True “gene profiles” and noisy observations.



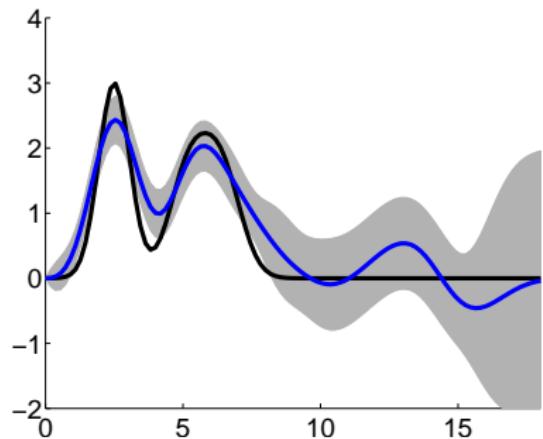
Inferred transcription factor activity.

Artificial Example: Inferring $p(t)$

Inferring TF activity from artificially sampled genes.



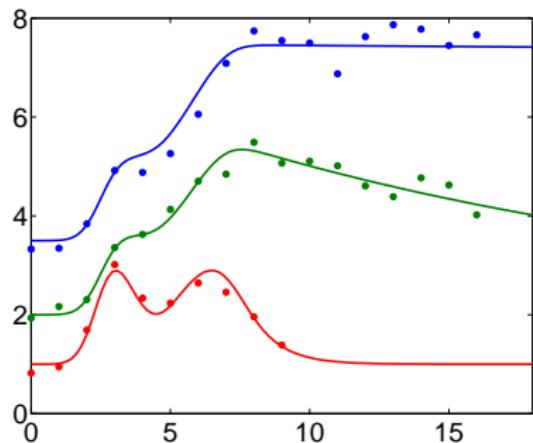
True “gene profiles” and noisy observations.



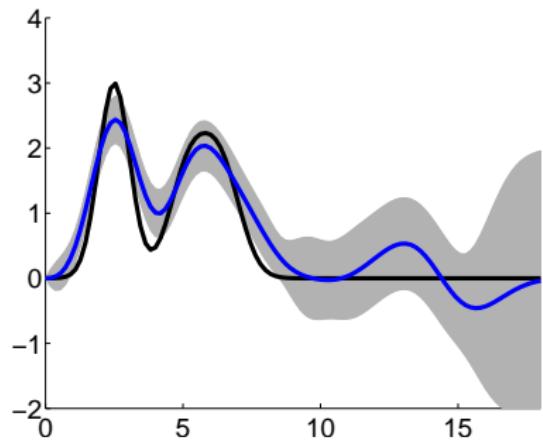
Inferred transcription factor activity.

Artificial Example: Inferring $p(t)$

Inferring TF activity from artificially sampled genes.



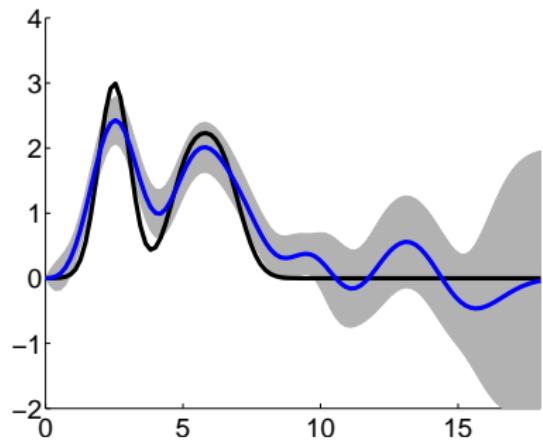
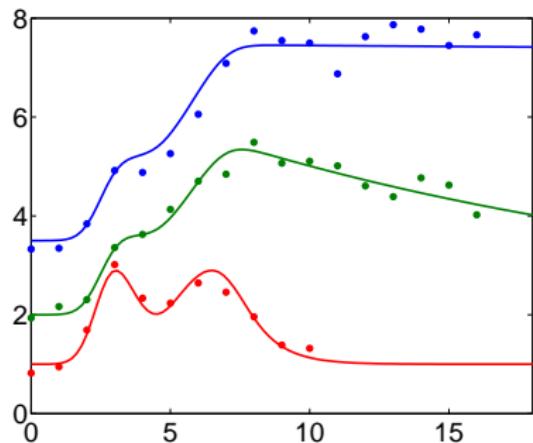
True “gene profiles” and noisy observations.



Inferred transcription factor activity.

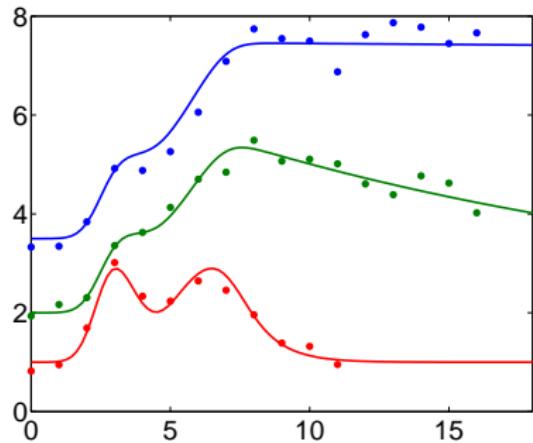
Artificial Example: Inferring $p(t)$

Inferring TF activity from artificially sampled genes.

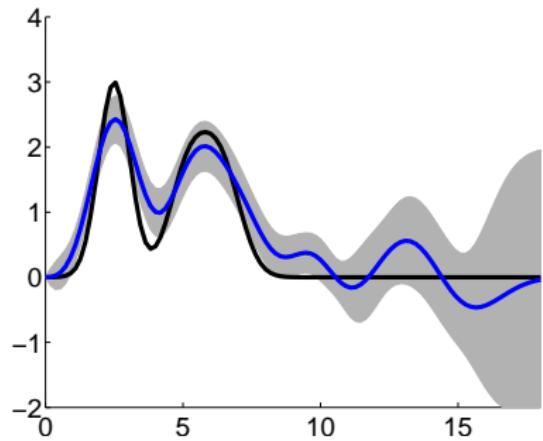


Artificial Example: Inferring $p(t)$

Inferring TF activity from artificially sampled genes.



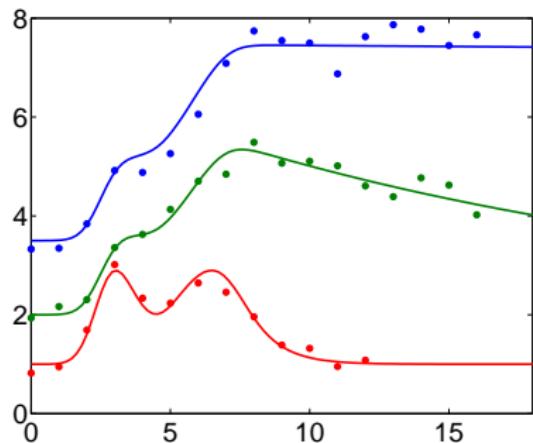
True “gene profiles” and noisy observations.



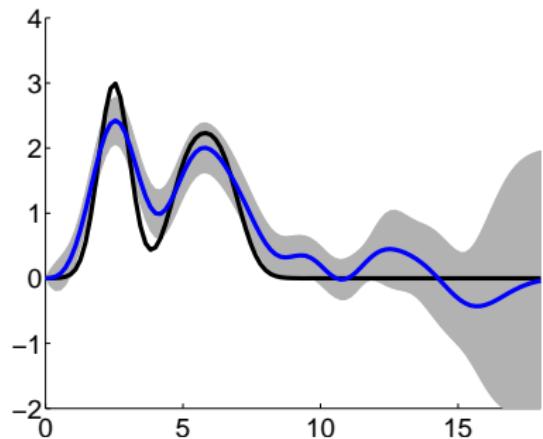
Inferred transcription factor activity.

Artificial Example: Inferring $p(t)$

Inferring TF activity from artificially sampled genes.



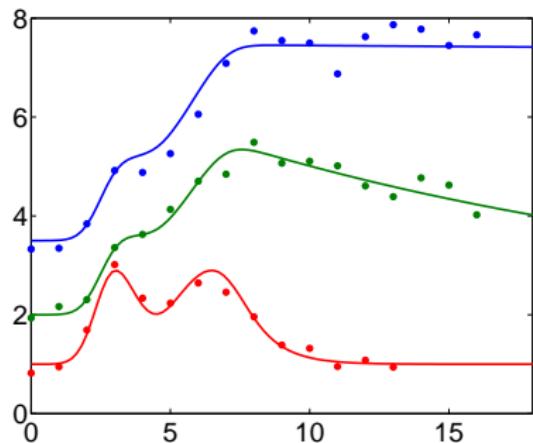
True “gene profiles” and noisy observations.



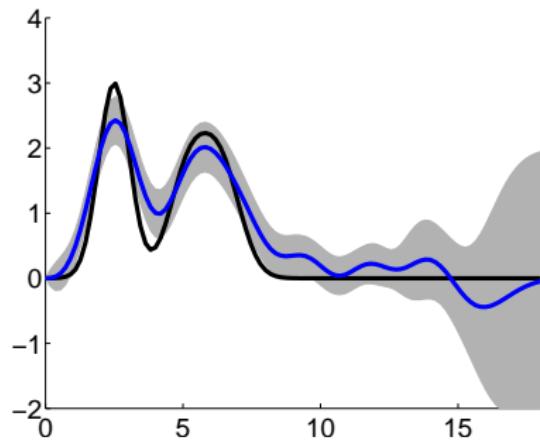
Inferred transcription factor activity.

Artificial Example: Inferring $p(t)$

Inferring TF activity from artificially sampled genes.



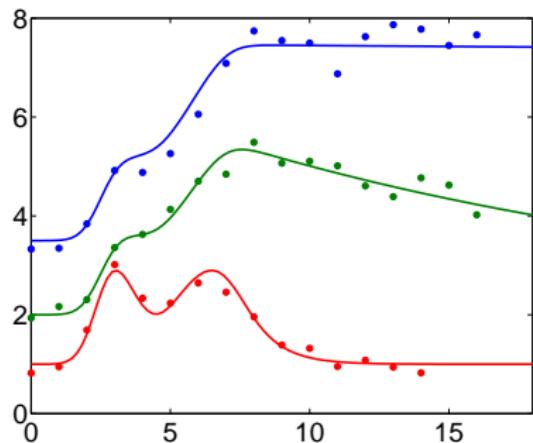
True “gene profiles” and noisy observations.



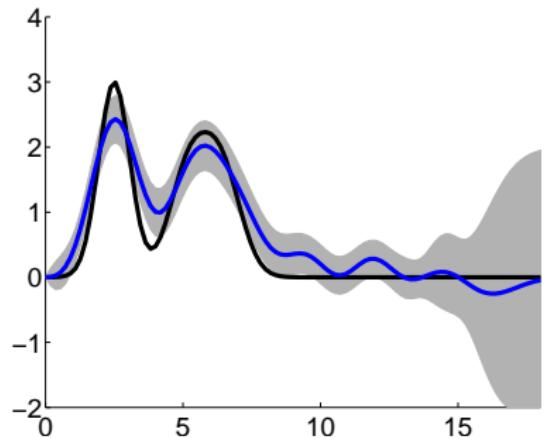
Inferred transcription factor activity.

Artificial Example: Inferring $p(t)$

Inferring TF activity from artificially sampled genes.



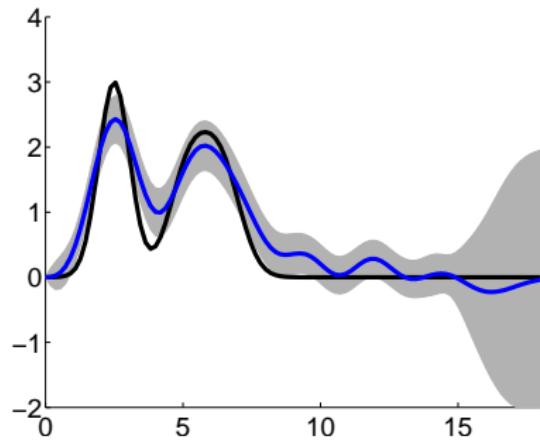
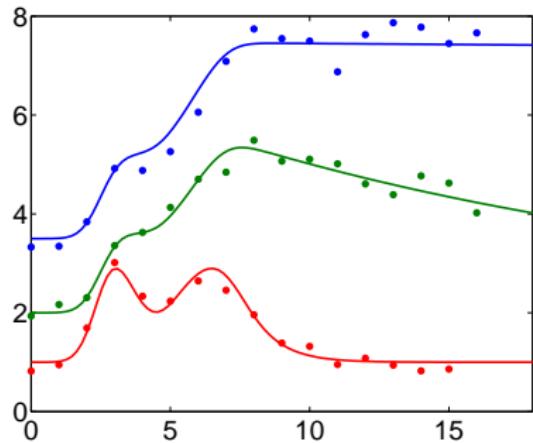
True “gene profiles” and noisy observations.



Inferred transcription factor activity.

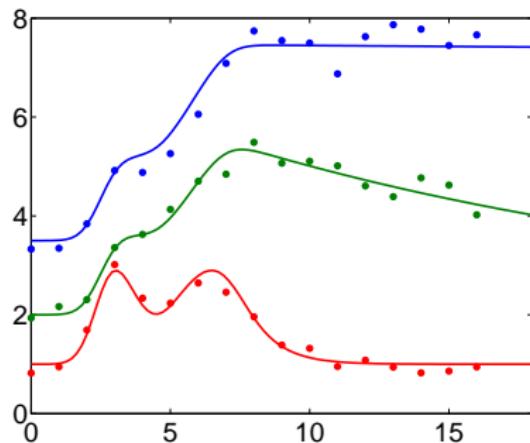
Artificial Example: Inferring $p(t)$

Inferring TF activity from artificially sampled genes.

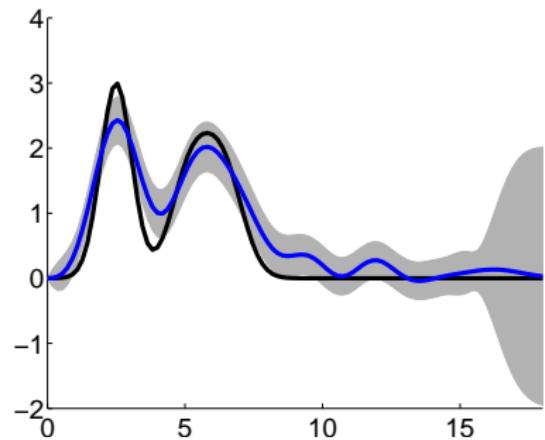


Artificial Example: Inferring $p(t)$

Inferring TF activity from artificially sampled genes.



True “gene profiles” and noisy observations.



Inferred transcription factor activity.

Outline

Motivation

Cascade Differential Equations

Discussion and Future Work

Model-based method for transcription factor target identification with limited data

Antti Honkela^{a,1}, Charles Girardot^b, E. Hilary Gustafson^b, Ya-Hsin Liu^b, Eileen E. M. Furlong^b, Neil D. Lawrence^{c,1}, and Magnus Rattray^{c,1}

^aDepartment of Information and Computer Science, Aalto University School of Science and Technology, Helsinki, Finland; ^bGenome Biology Unit, European Molecular Biology Laboratory, Heidelberg, Germany; and ^cSchool of Computer Science, University of Manchester, Manchester, United Kingdom

Edited by David Baker, University of Washington, Seattle, WA, and approved March 3, 2010 (received for review December 10, 2009)

We present a computational method for identifying potential targets of a transcription factor (TF) using wild-type gene expression time series data. For each putative target gene we fit a simple differential equation model of transcriptional regulation, and the

used for genome-wide scoring of putative target genes. A key consideration is required to apply our method is wild-type time series data that are collected over a period where TF activity is changing. Our method allows for complementary evidence from expression

Cascaded Differential Equations

(Honkela et al., 2010)

- ▶ Transcription factor protein also has governing mRNA.
- ▶ This mRNA can be measured.
- ▶ In signalling systems this measurement can be misleading because it is activated (phosphorylated) transcription factor that counts.
- ▶ In development phosphorylation plays less of a role.

Twist Results

- ▶ Use mRNA of Twist as driving input.
- ▶ For each gene build a cascade model that forces Twist to be the only TF.
- ▶ Compare fit of this model to a baseline (e.g. similar model but sensitivity zero).
- ▶ Rank according to the likelihood above the baseline.
- ▶ Compare with correlation, knockouts and time series network identification (TSNI) (Della Gatta et al., 2008).

Results for Twi using the Cascade model

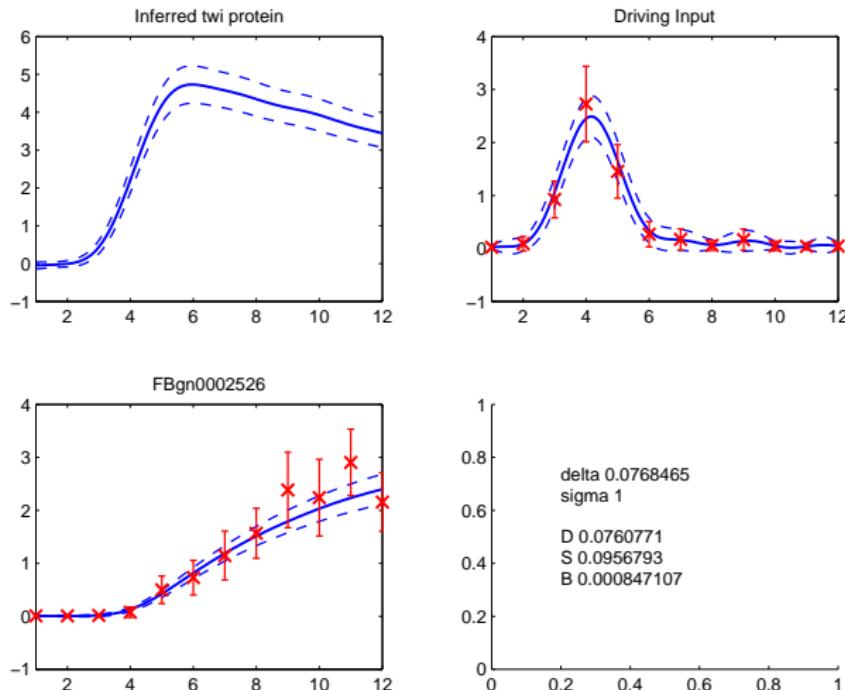


Figure: Model for flybase gene identity FBgn0002526.

Results for Twi using the Cascade model

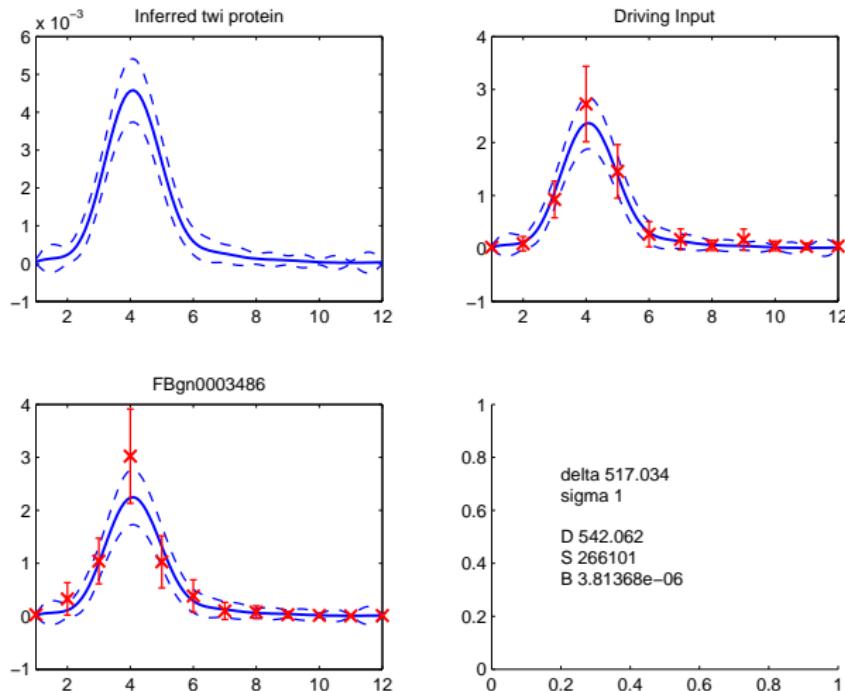


Figure: Model for flybase gene identity FBgn0003486.

Results for Twi using the Cascade model

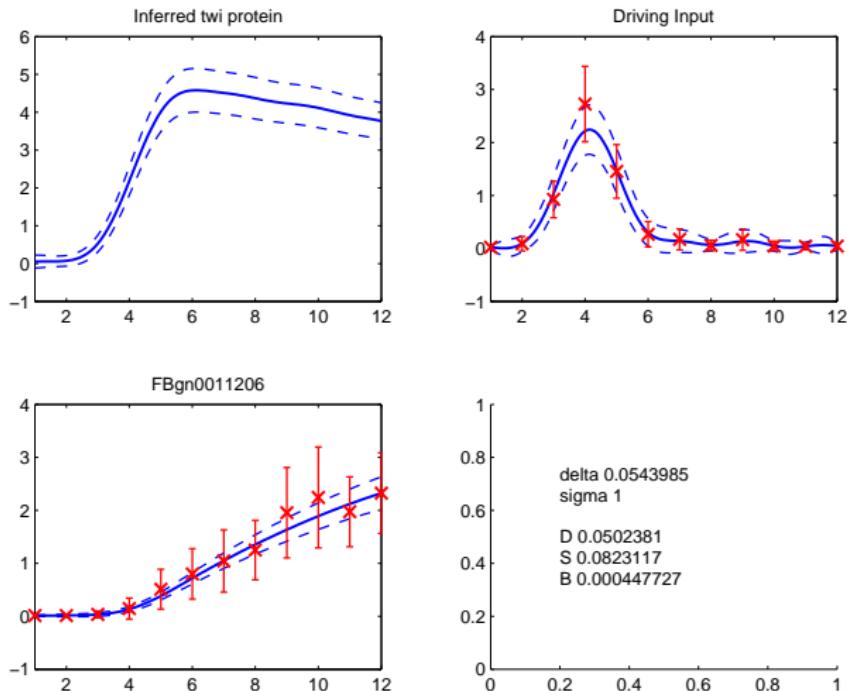


Figure: Model for flybase gene identity FBgn0011206.

Results for Twi using the Cascade model

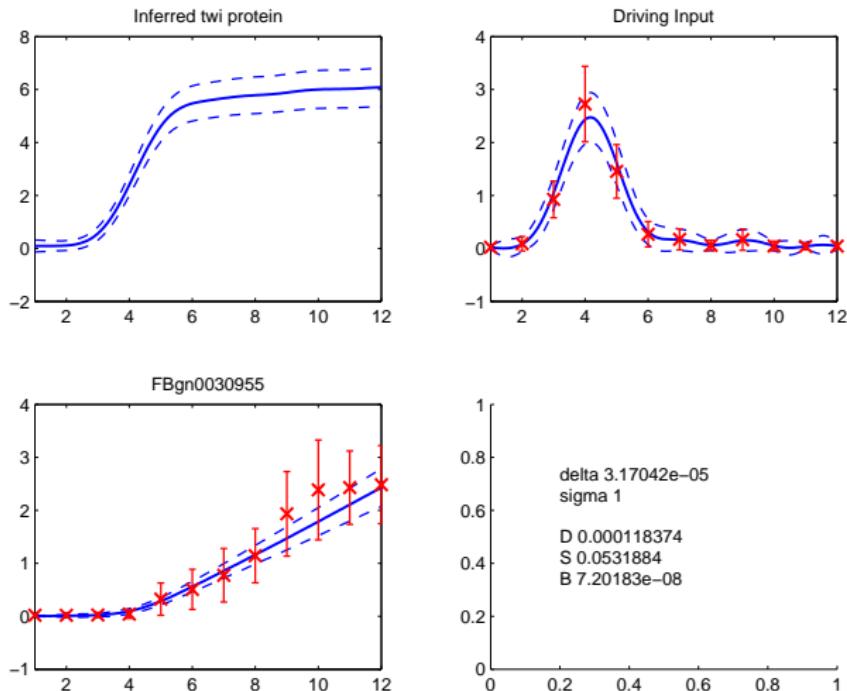


Figure: Model for flybase gene identity FBgn0030955.

Results for Twi using the Cascade model

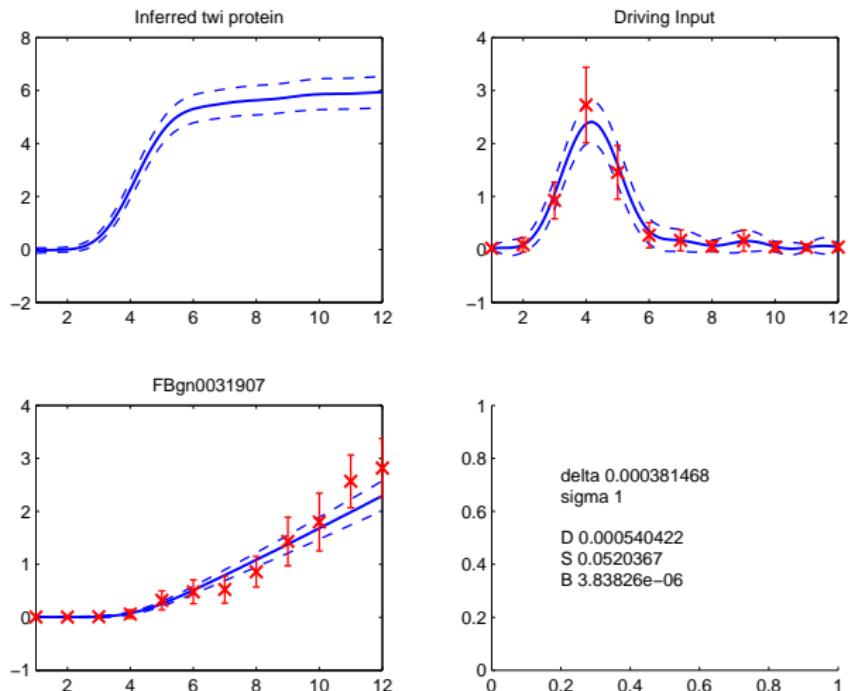


Figure: Model for flybase gene identity FBgn0031907.

Results for Twi using the Cascade model

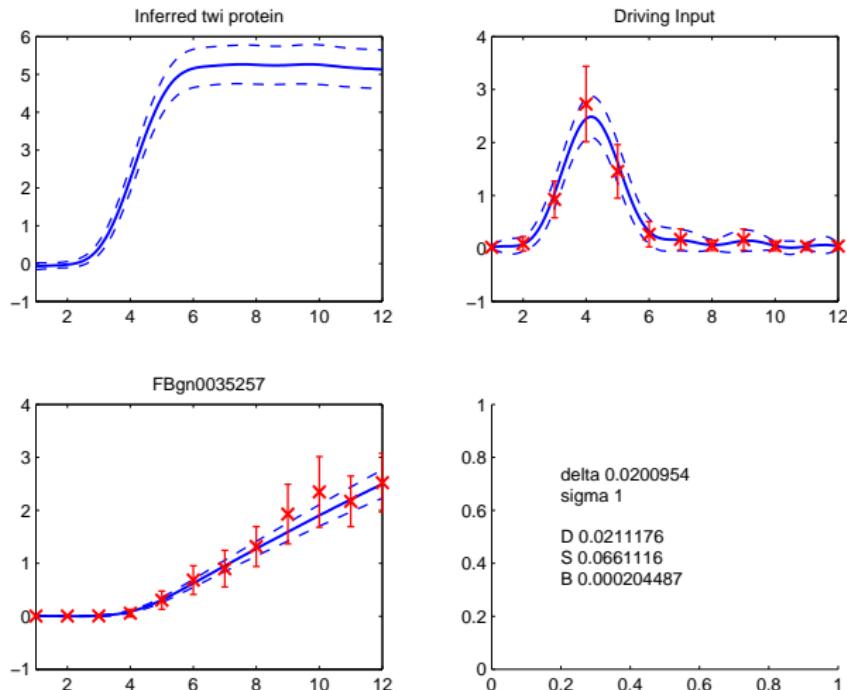


Figure: Model for flybase gene identity FBgn0035257.

Results for Twi using the Cascade model

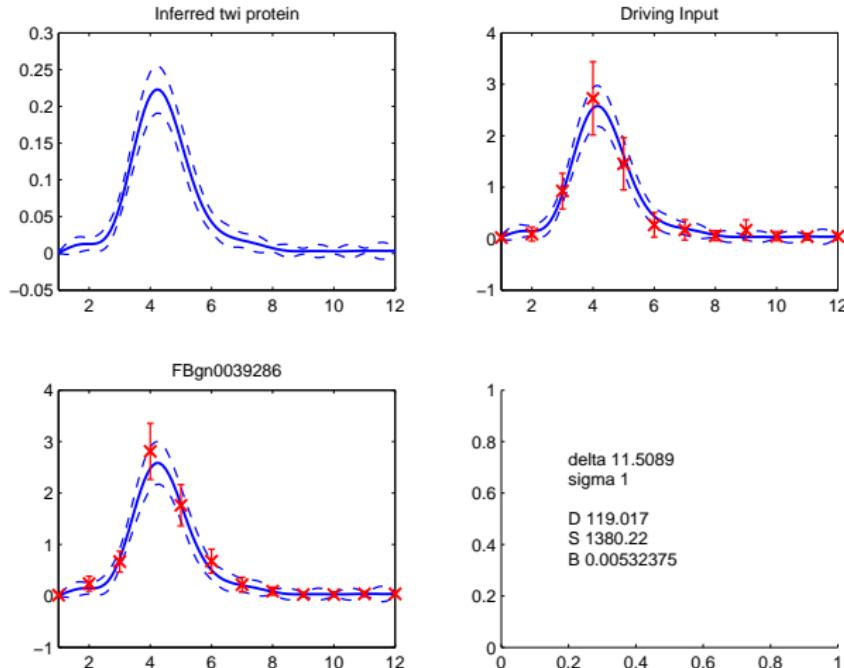
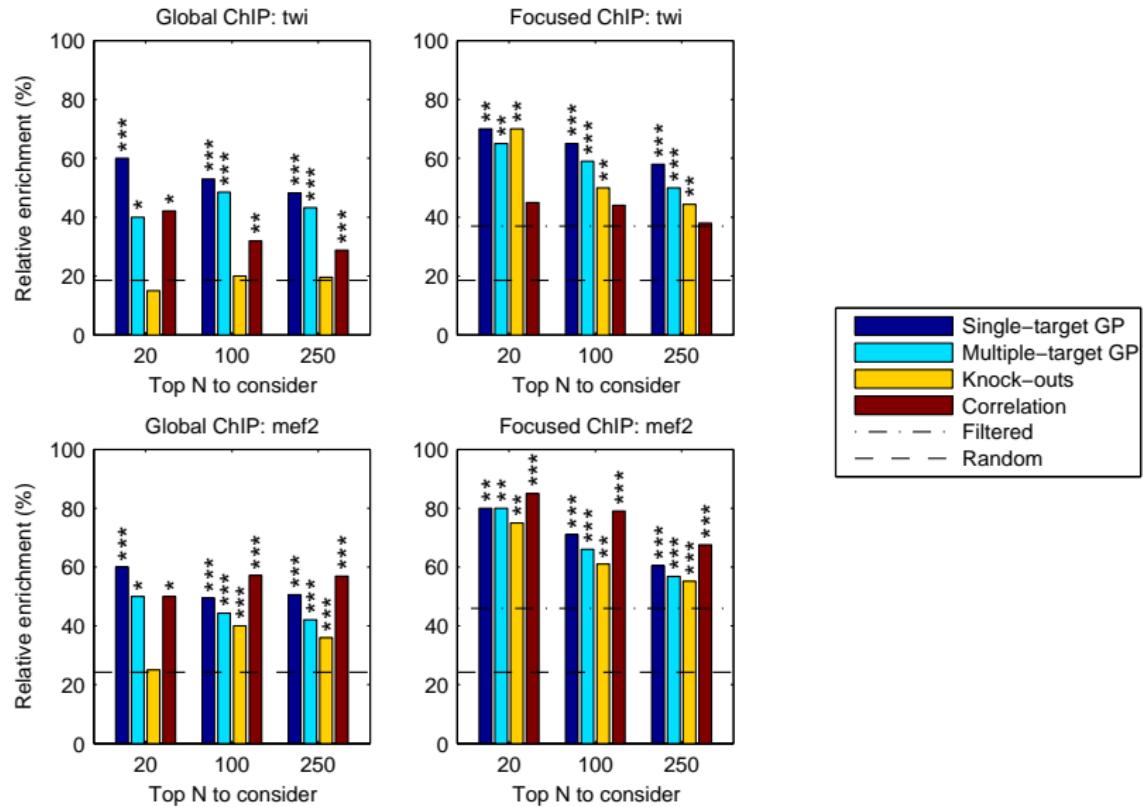


Figure: Model for flybase gene identity FBgn0039286.

Evaluation methods

- ▶ Evaluate the ranking methods by taking a number of top-ranked targets and record the number of "positives" (Zinzen et al., 2009):
 - ▶ targets with ChIP-chip binding sites within 2 kb of gene
 - ▶ (targets differentially expressed in TF knock-outs)
- ▶ Compare against
 - ▶ Ranking by correlation of expression profiles
 - ▶ Ranking by q -value of differential expression in knock-outs
- ▶ Optionally focus on genes with annotated expression in tissues of interest

Results



****: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$

Summary

- ▶ Cascade models allow genomewide analysis of potential targets given only expression data.
- ▶ Once a set of potential candidate targets have been identified, they can be modelled in a more complex manner.
- ▶ We don't have ground truth, but evidence indicates that the approach *can* perform as well as knockouts.

Outline

Motivation

Cascade Differential Equations

Discussion and Future Work

Discussion and Future Work

- ▶ Integration of probabilistic inference with mechanistic models.
- ▶ Software available through bioconductor (TIGRE Package)
<http://bioconductor.org/packages/2.6/bioc/html/tigre.html>.
- ▶ Applications in modeling gene expression.
- ▶ Ongoing/other work:
 - ▶ Non linear response and non linear differential equations.
 - ▶ Improving computational complexity.
 - ▶ Stochastic differential equations.
 - ▶ Cascade model introduces model of translation.

Acknowledgements

- ▶ Investigators: Neil Lawrence and Magnus Rattray
- ▶ Researchers: Pei Gao, Antti Honkela, Guido Sanguinetti, Michalis Titsias.
- ▶ Martino Barenco and Mike Hubank at the Institute of Child Health in UCL (p53 pathway).
- ▶ Charles Girardot and Eileen Furlong of EMBL in Heidelberg (mesoderm development in *D. Melanogaster*).

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” Academy of Finland and FP7 PASCALII NoE

References I

G. Della Gatta, M. Bansal, A. Ambesi-Impiombato, D. Antonini, C. Missero, and D. di Bernardo. Direct targets of the trp63 transcription factor revealed by a combination of gene expression profiling and reverse engineering. *Genome Research*, 18(6):939–948, Jun 2008. [\[URL\]](#). [\[DOI\]](#).

A. Honkela, C. Girardot, E. H. Gustafson, Y.-H. Liu, E. E. M. Furlong, N. D. Lawrence, and M. Rattray. Model-based method for transcription factor target identification with limited data. *Proc. Natl. Acad. Sci. USA*, 107(17):7793–7798, Apr 2010. [\[DOI\]](#).

R. P. Zinzen, C. Girardot, J. Gagneur, M. Braun, and E. E. M. Furlong. Combinatorial binding predicts spatio-temporal cis-regulatory activity. *Nature*, 462(7269):65–70, Nov 2009. [\[URL\]](#). [\[DOI\]](#).