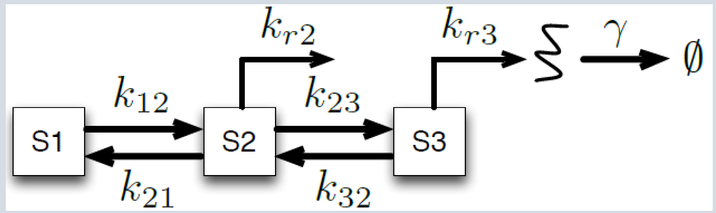
Identification of Gene Regulation Models from Single-Cell Data

Lisa Weber[1], Will Raymond[1,2], Brian Munsky[1,2]

[1]Department of Chemical & Biological Engineering, Colorado State University [2]School of Biomedical Engineering, Colorado State University

Introduction

We define a three-state generalization of the bursting gene expression model [1,2]. We extend this model to allow for a time-dependent input signal that controls the state transition reactions: k_{12}, k_{23}, k_{21} or k_{32}.



Three-state bursting gene expression model.

We fit these model hypotheses to a finite set of simulated single-cell data, and we attempt to identify the model mechanisms and parameters. We use multiple different analyses (e.g., deterministic and stochastic) for the same model and same data, and we explore how uncertainty in parameter space varies with respect to the chosen analysis approach or specific experiment design.

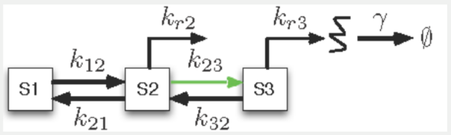
The approach to be taken is based upon previous experimental and computational investigations undertaken to explore signal-activated gene expression models in yeast [3] and human cells [4].

Approaches

1. **Deterministic Analysis of Averaged mRNA Expression**
   - We compute the likelihood that the average sample data comes from the model's deterministic ordinary differential equation (using the chi-squared likelihood function).
2. **Finite State Projection (FSP) Analysis of Full mRNA Distributions**
   - We compute the likelihood that the entire data histograms come from the full probability distributions.
3. **Metropolis-Hastings Algorithm (MHA)**
   - We use a Markov Chain Monte Carlo analysis to estimate parameter uncertainties for each model and each likelihood function (i.e., the ODE-based chi-squared function or the FSP likelihood function).

Problem Description

Our goal is to identify the mechanism of action (i.e., determine which k_{ij} depends upon the input) and find the model parameters.

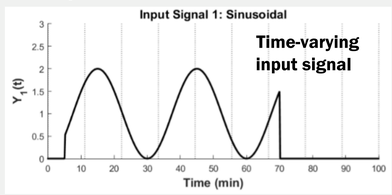


The input-dependent transition rates can be one of:

M1: $k_{12}(t)$; M2: $k_{23}(t)$;
M3: $k_{21}(t)$; M4: $k_{32}(t)$.

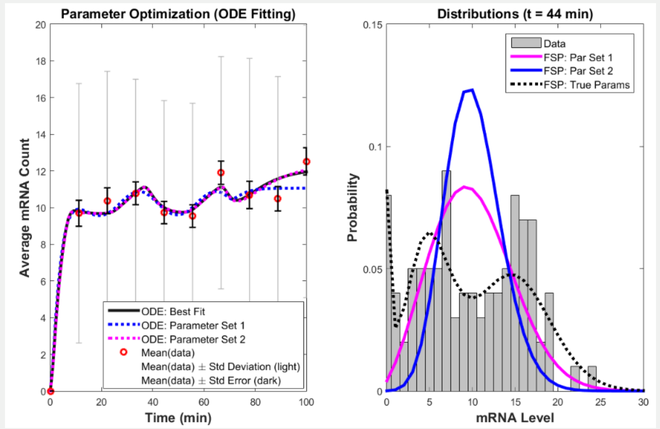
Input - We consider a known, deterministic input of the form:

$$Y_1(t) = \begin{cases} 0 & \text{for } (t \le 5) (t \ge 70) \\ 1 - \cos\left(\frac{2\pi t}{30}\right) & \text{for } t \in (5, 70) \end{cases}$$



Data - We simulate 100 single-cell measurements for each of 10 equally spaced time points.

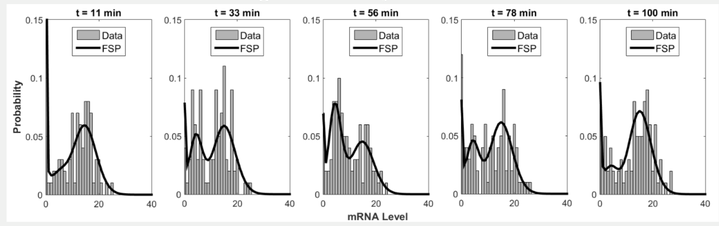
Results



(Left) Mean gene expression for Model 2 for two parameter sets (Λ_1 and Λ_2) near the maximum of the chi-squared likelihood function (ODE fit).

(Right) Full distributions at t = 44 min for Λ_1 and Λ_2 compared to the data and the true distributions. Both parameter sets from the ODE fit completely fail to capture the bimodal behavior.

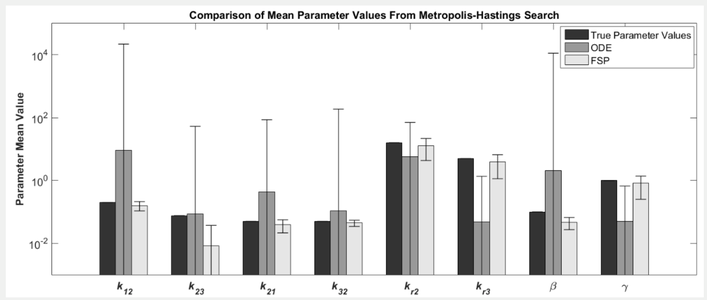
Maximum likelihood fits using the FSP analysis.

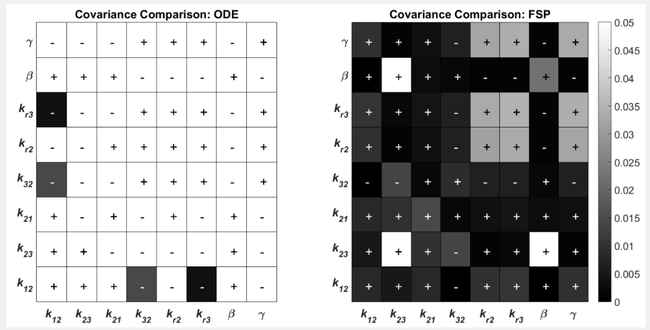


In contrast to the ODE approach, the FSP quantitatively captures the bimodal behavior of the data at all time points.

Using the MHA, we find that the FSP fit comes much closer to the true parameter values.

Furthermore, the FSP gives much tighter bounds on the parameter uncertainties.

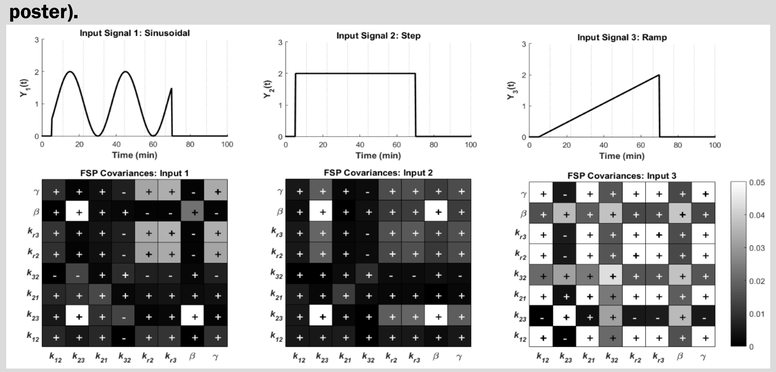




The covariances for parameter combinations are much larger for the ODE compared to the FSP. The (+) indicates a positive covariance and (-) indicates a negative covariance.

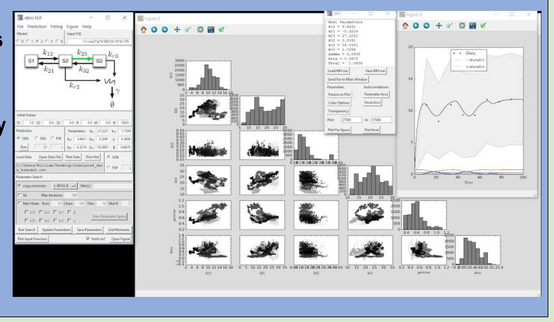
Importance of Experiment Design

We simulated data from three different potential inputs: the original sinusoidal function, a step function, and a ramp function. Each input results in a different amount of parameter uncertainty after running the MHA. The step and sinusoidal inputs reduce uncertainty far more than does the ramp input (see also Fox/Munsky poster).



GUI

The CH30-GUI provides a user-friendly means to generate or import simulated data, specify input signals, choose different models, and perform all analyses described here (ODEs, FSP and MHA).



Conclusions

- Fitting average behavior with ODE analyses can lead to poor and highly uncertain identification of parameters.
- Fitting single-cell distributions using an FSP likelihood function can substantially improve identification results.
- Certain single-cell experiments provide more information than others.
- The methods demonstrated here can be applied to a wide range of gene regulation models for parameter identification and to gain valuable insight into gene regulatory dynamics.

References and Acknowledgements

[1] B. MUNSKY, G. NEUERT, AND A. VAN OUDENAARDEN, *Using Gene Expression Noise to Understand Gene Regulation*, *Science*, 336 (2012), pp. 183–187.
[2] J. PECCOUD AND B. YCART, *Markovian Modeling of Gene-Product Synthesis*, Theoretical Population Biology, 48 (1995), pp. 222–234.
[3] G. NEUERT, B. MUNSKY, R. Z. TAN , L. TEYTELMAN , M. KHAMMASH , AND A. VAN OUDENAARDEN, *Systematic Identification of Signal-Activated Stochastic Gene Regulation*, Science, 339 (2013),pp. 584–587.
[4] A. SENECAL , B. MUNSKY, F. PROUX , N. LY, F. E. BRAYE , C. ZIMMER , F. MUELLER , AND X. DARZACQ, *Transcription factors modulate c-Fos transcriptional bursts*, Cell Reports, 8 (2014), pp. 75–83.
[5] J. F. APGAR , J. E. TOETTCHER , D. ENDY, F. M. WHITE , AND B. TIDOR, *Stimulus Design for Model Selection and Validation in Cell Signaling*, PLoS Computational Biology, 4 (2008), p. e30.
[6] B. MÉLYKÚTI , E. AUGUST, A. PAPACHRISTODOULOU , AND H. EL -SAMAD, *Discriminating between rival biochemical network models: three approaches to optimal experiment design*, BMC Systems Biology, 4 (2010), p. 1.

W.M. KECK FOUNDATION
NIGMS