Parameter estimation and model selection in computational biology

Gabriele Lillacci[1], and Mustafa Khammash[1]

Short Abstract — **Parameter estimation is a key issue in computational and systems biology, as it represents the crucial step to obtaining quantitative information from dynamical models of biological systems. We introduce a new approach to this problem built on the combination of Kalman filtering and optimization techniques, which is fast, easy to implement, can be applied to large parameter spaces, can handle sparse and noisy data, and provides an evaluation of the statistical significance of the computed estimates. Furthermore, we show how these techniques can be also used to address the problem of model selection, in which one has to pick the most plausible model for a given process among a list of possible candidates.**

Keywords — **Dynamic modeling, Ordinary differential equations, Kalman filter, Identifiability, Heat shock, Gene expression.**

I. INTRODUCTION

Many biological processes are modeled using ordinary differential equations (ODEs) that describe the evolution over time of certain quantities of interest, such as concentrations or number of molecules of proteins and mRNA. Once the pathway structure is known, the corresponding equations are relatively easy to write down using widely accepted kinetic laws, such as the law of mass action or the Michaelis-Menten law.

In general the equations will depend on several unknown parameters. While sometimes it is feasible to measure them experimentally, in many cases this is very hard, expensive, time consuming, or even impossible. However, it is usually possible to measure some of the other variables involved in the models (such as abundance of chemical species) using PCR, immunoblotting assays, fluorescent markers, and the like. For these reasons, the problem of *parameter estimation* represents a key issue in computational and systems biology.

II. METHODS

Parameter estimation has been addressed with several methods, among which optimization [1], Bayesian methods [2], and state observers [3].

We extend the results on parameter estimation by state observers by introducing a constrained hybrid extended Kalman filtering algorithm, together with a measure of accuracy of the estimation process based on a statistical variance test. Our procedure is fast, easy to implement, can be applied to large parameter spaces, can handle sparse and noisy data, and provides an evaluation of the statistical significance of the computed estimates that can also be used for model selection. We demonstrate our method on a model of the heat shock response in *E. coli* [4].

III. RESULTS

Our technique is capable of producing a parameter set that is consistent with the experimental observations. As shown in the Figure, the solution of the identified model (blue) is very close to the true solution (red), even if the data points used for the estimation (green squares) are sparse and very noisy. In this simulation the variance of the noise is of the order of 500% of the mean of the measurement signal.
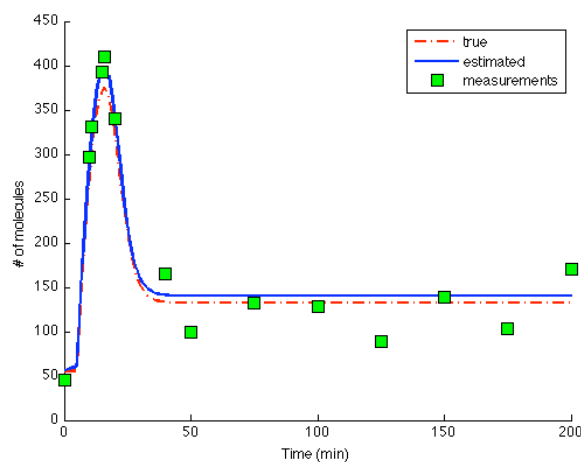

Figure. Comparison between the model solution for the true parameter values and the estimated parameter values.

In the same situation, classical parameter estimation methods, such as least-squares fitting, failed in giving answers that met the same statistical accuracy criteria.

Our method was also successful in discriminating between two candidate models for the heat shock process, one with and one without a feedback loop.

REFERENCES

[1] Moles CG, Mendes P, Banga JR (2003) Parameter estimation in biochemical pathways: A comparison of global optimization methods. *Gen Res* **13**, 2467-2474.
[2] Wilkinson DJ (2007) Bayesian methods in bioinformatics and computational systems biology. *Brief Bioinform* **8**, 109-116.
[3] Dochain D (2003) State and parameter estimation in chemical and biochemical processes: a tutorial. *J Process Control* **13,** 801-818.
[4] El-Samad H et al. (2006) Advanced methods and algorithms for biological networks analysis. *Proc IEEE* **94**, 832-853.

[1]Center for Control, Dynamical Systems and Computation. University of California at Santa Barbara, Santa Barbara, CA 93106.
E-mails: gabriele@engr.ucsb.edu, khammash@engr.ucsb.edu.