System composability and behavior prediction of genetic networks from characterized models of subsystems

Vipul Singhal*
vs@alumni.caltech.edu
Genome Institute of Singapore
Singapore, Singapore

1 INTRODUCTION

Composition of computational models of systems, such that the combined models are predictive of real world systems, is an important paradigm across traditional engineering disciplines. In the field of synthetic biology, which aims to bring engineering ideas to the design of biological systems, such composability has remained elusive, partly because of the phenomenon of structural non-identifiability of the parameters involved. We demonstrate tools for identifying joint distributions of system parameters, and for composing subsystem models in the presence of non-identifiability, such that the resulting composed models are predictive of the corresponding systems' real-world behaviour.

2 RESULTS

As a first step towards predictive composability, we describe a set of tools for *consensus* inference of parameters associated with genetic 'circuits', which are widely used in the field of synthetic biology [1]. Consensus inference refers to the scenario where parameters may be informed by more than one experiment (Figure 1). Our tools use efficient affine invariant ensemble Markov chain Monte Carlo samplers to perform inference of parameters [2] in high dimensional spaces. This results in the joint distribution of parameters, and gives users information on which parameters are non-identifiable.

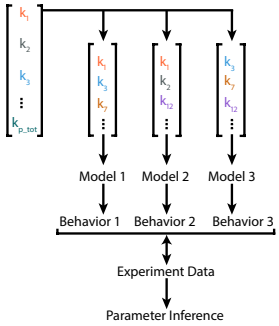

Figure 1: Consensus parameter inference: a common set of parameters is informed by multiple experiments.

We describe two scenarios where subsystems may be composed into larger systems. First, we demonstrate the characterization of parameters associated with subsystems of the well-known incoherent feed-forward loop (IFFL) genetic circuit motif, and compose these subsystems into the full IFFL, whose predicted behaviour we validate with experimental data [4].

*This work was performed while VS was with the CNS department at Caltech.

As an example of parameter inference involved in this problem, we inferred transcription and translation rate parameters involved in expression of fluorescent proteins using data from three types of experiments (Figure 2). Details can be found in [4].

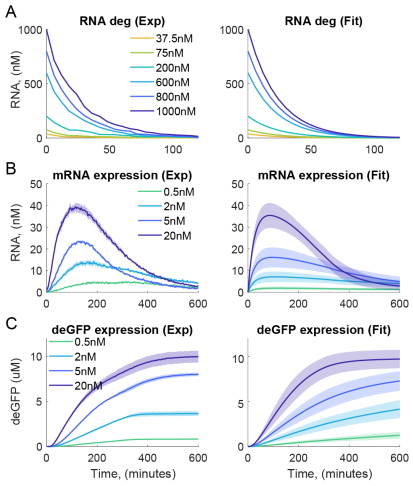

Figure 2: Left column: experimental data. Right column: simulations generated using inferred parameter distributions. (A) Degradation of spiked-in fluorescent RNA. (B, C) Expression time course of fluorescent aptamer tagged mRNA (B) and GFP protein (C) for different initial DNA concentrations.

The second scenario where we used the composability paradigm involved correcting the variability in a system's behavior across different environments [3]. Here, we calibrated the parametric characteristics of different environments using our MCMC inference tools, and used these to 'correct' data across these environments. We also derived a set of *consistency conditions* under which parameter non-identifiability does not hinder the batch correction procedure.

REFERENCES

[1] Michael B. Elowitz and Stanislas Leibler. 2000. A synthetic oscillatory network of transcriptional regulators. *Nature* 403, 6767 (Jan. 2000), 335–338. https://doi.org/10.1038/35002125 Number: 6767 Publisher: Nature Publishing Group.

[2] Jonathan Goodman and Jonathan Weare. 2010. Ensemble samplers with affine invariance. *Communications in Applied Mathematics and Computational Science* 5, 1 (Jan. 2010), 65–80. https://doi.org/10.2140/camcos.2010.5.65

[3] V. Singhal and R. M. Murray. 2019. Transforming Data Across Environments Despite Structural Non-Identifiability. In *2019 American Control Conference (ACC)*. 5639–5646.

[4] Vipul Singhal, Zoltan A Tuza, Zachary Z Sun, and Richard M Murray. 2021. A MATLAB toolbox for modeling genetic circuits in cell-free systems†. *Synthetic Biology* 6, 1 (Oct. 2021), ysab007. https://doi.org/10.1093/synbio/ysab007