Parameter Discovery for Stochastic Computational Models in Systems Biology Using Bayesian Model Checking

Faraz Hussain*, Christopher J. Langmead†, Qi Mi‡, Joyeeta Dutta-Moscato§, Yoram Vodovotz¶, Sumit K. Jha*

* *Dept. of Electrical Engineering and Computer Science, University of Central Florida. Email: {fhussain,jha}@eecs.ucf.edu*
† *School of Computer Science, Carnegie Mellon University. Email: cjl@cs.cmu.edu*
‡ *Dept. of Sports Medicine and Nutrition, University of Pittsburgh. Email: qim3@pitt.edu*
§ *Dept. of Biomedical Informatics, University of Pittsburgh. Email: jod30@pitt.edu*
¶ *Dept. of Surgery and McGowan Institute of Regenerative Medicine, University of Pittsburgh. Email: vodovotzy@upmc.edu*

***Abstract*—Parameterized probabilistic complex computational ($\mathbf{P^2C^2}$) models are being increasingly used in computational systems biology for analyzing biological systems. A key challenge is to build mechanistic $\mathbf{P^2C^2}$ models by combining prior knowledge and empirical data, given that certain system properties are unknown. These unknown components are incorporated into a model as *parameters* and determining their values has traditionally been a process of trial and error. We present a new algorithmic procedure for discovering parameters in agent-based models of biological systems against behavioral specifications mined from large data-sets. Our approach uses Bayesian model checking, sequential hypothesis testing, and stochastic optimization to synthesize parameters of $\mathbf{P^2C^2}$ models. We demonstrate our algorithm by discovering the amount and schedule of doses of bacterial lipopolysaccharide in a clinical agent-based model of the dynamics of acute inflammation that guarantee a set of desired clinical outcomes with high probability.**

I. Introduction

With rapid increase in the availability of high-performance computing, there has been a surge in interest in the construction of parameterized probabilistic complex computational (P^2C^2) models. Deterministic models cannot capture the unpredictability of natural phenomena or of multi-outcome man-made devices. *Stochastic models* address these problems by allowing a succinct representation of behavioral variability. However, for non-deterministic models, the process of model discovery and validation is extremely complicated.

In recent years, *agent-based modeling* has emerged as a popular method for the analyzing biological models [1]. An agent-based model (ABM) consists of autonomous, independently-acting agents with an internal state that includes time-varying spatial co-ordinates. Agent interaction is governed by a set of rules local to each type of agent. These rules themselves are probabilistic, enabling the agents to demonstrate behavioral variability in the face of environmental uncertainty. Formal verification of such ABMs is vital for users to have confidence in the predictions generated by them.

An important challenge faced by designers of an ABM is to find values of unknown parameters in a model that enable it to reproduce the behavior of a biological system. When the number of parameters is small, an exhaustive search of the state space is feasible. For high-dimensional models, however, such brute-force methods are prohibitively expensive. In this paper, we make the following new contributions:

- We present a *new algorithm for discovering parameters in stochastic agent-based models (ABMs)* that employs Bayesian statistical model checking to verify model properties and needs fewer samples than our earlier parameter synthesis algorithm that uses the Sequential Probability Ratio Test (SPRT) [3].
- We demonstrate the effectiveness of our approach by applying our algorithm to learn *twenty eight* parameters in a physiological model of the acute inflammatory response to endotoxin administration [2].

II. Case Study: Application to a complex physiological model

We applied our parameter learning algorithm to a physiological model of the acute inflammatory response (AIR) to administration of the endotoxin lipopolysaccharide (LPS) [2]. We wrote a tool in C++ that uses OpenMPI for parallelization and the GNU Scientific Library for mathematical computations. Simulations were performed using the SPARK agent-based modeling and simulation software [4].

Our aim was to discover the schedule and doses of LPS that make the model exhibit desired behavioral properties. We synthesized *twenty eight model parameters* (shown in Table I), for the following set of *four* specifications given to us by experts with extensive experience with the model:

1) A low dose of LPS causes inflammation which eventually resolves.
2) A high dose of LPS causes inflammation which never resolves.
3) For a certain time interval, when one administration of LPS is followed by a second administration of the same dose, the inflammatory response resulting from the second administration is *lesser* than that from the first.
4) For a certain time interval, when one administration of LPS is followed by a second administration of the same dose, the inflammatory response resulting from the second administration is *greater* than that from the first.

978-1-4673-1321-6/12/$31.00 ©2014 IEEE

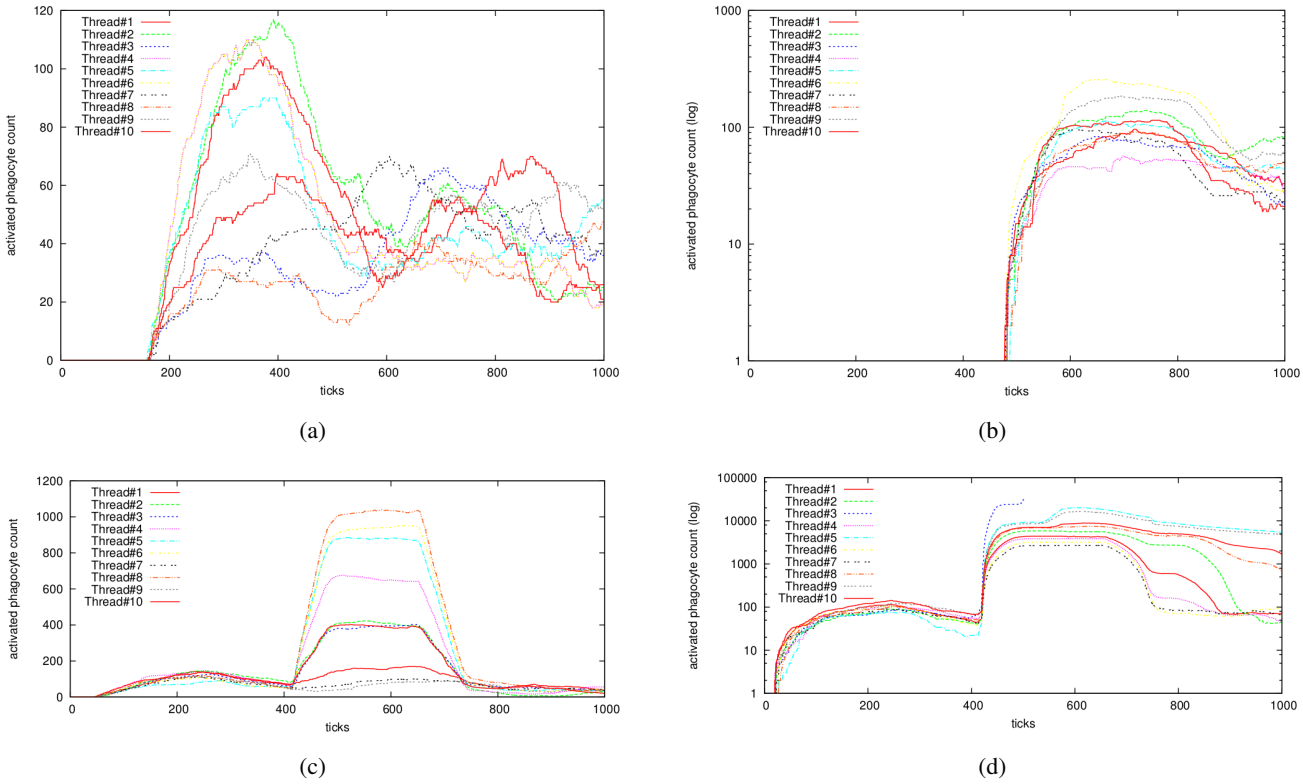

Fig. 1: Parallel simulations using MPI show SPARK output for 10 threads. Figures (a), (b), (c) and (d) show traces for the activated phagocyte count over time on invocation of the ABM simulator. One can visually verify that the ABM parameterized with the values in Table I satisfies all four specifications.

TABLE I: Parameters of the Acute Inflammatory Response model synthesized by our algorithm.

LPS-evap	0.932575	exp1-dose-amount	3352.54
mac-act-LPS	0.661416	exp2-dose-time	467.262
mac-act-pro	0.326682	exp2-dose-duration	458.451
mac-regen	12.655	exp2-dose-amount	896067
mac-age	60.5967	exp3-1st-dose-time	33.3838
mac-act-dam	0.3916	exp3-1st-dose-duration	41.6759
max-pro-dam	18.5986	exp3-2nd-dose-time	407.352
pro-dam-thresh	0.51023	exp3-2nd-dose-duration	41.6759
damage-evap	0.276594	exp3-doses-amount	2628.97
anti-heal-thresh	7.92487	exp4-1st-dose-time	8.24293
mac-anti	0.442621	exp4-1st-dose-duration	4.40842
anti-evap	0.623503	exp4-2nd-dose-time	411.959
pro-evap	0.142298	exp4-2nd-dose-duration	4.40842
mac-prop	8.39519	exp4-doses-amount	4494.65
exp1-dose-time	149.574	exp1-dose-duration	4.80575

Figure 1 depicts model simulation traces when parameterized at the synthesized parameter set from Table I and shows the satisfaction of all four specifications.

III. Conclusion and Future Work

Future work includes automating the process of *learning specifications* from time-series data and designing parameter discovery algorithms capable of handling high-dimensional models (of the order of hundreds of parameters) by exploiting topological characteristics of the biological system under consideration.

Acknowledgment

We acknowledge support from the Air Force Research Lab under contract #CA0116UCF2013 (SKJ), the Oak Ridge National Lab under contract #4000126570 (SKJ), from NIH grants P41 GM103712 (CJL) and P50-GM-53789 (YV), from the National Institute on Disability Rehabilitation Research (NIDRR) under grant #H133E070024 (YV), and from the University of Central Florida with a Graduate Research Excellence Fellowship (FH).

References

[1] G. An, Q. Mi, J. Dutta-Moscato, and Y. Vodovotz. Agent-based models in translational systems biology. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, 1(2):159–171, 2009.

[2] J. Day, J. Rubin, Y. Vodovotz, C. C. Chow, A. Reynolds, and G. Clermont. A reduced mathematical model of the acute inflammatory response ii. capturing scenarios of repeated endotoxin administration. *Journal of theoretical biology*, 242(1):237–256, 2006.

[3] F. Hussain, R. G. Dutta, S. K. Jha, C. J. Langmead, and S. Jha. Parameter discovery for stochastic biological models against temporal behavioral specifications using an sprt based metric for simulated annealing. In *Proc. of the 2nd International Conference on Computational Advances in Bio and Medical Sciences, Las Vegas, NV*, pages 1–6. IEEE Press, 2012.

[4] A. Solovyev, M. Mikheev, L. Zhou, J. Dutta-Moscato, C. Ziraldo, G. An, Y. Vodovotz, and Q. Mi. SPARK: A Framework for Multi-Scale Agent-Based Biomedical Modeling. *International Journal of Agent Technologies and Systems*, 2(3):18–30, 2010.