Data driven inference of unmodelled dynamical processes in systems biology models

Maik Kschischo and Matthias Kahm

University of Applied Sciences Koblenz,
Südallee 2, 53424 Remagen, Germany

Abstract. Mathematical models are at the heart of systems biology. They are used to gain a deeper understanding of biological problems and to guide the selection of informative experiments. Model development for a particular process starts with a number of decisions regarding the specification of the systems border, the specification of dynamical variables and assumptions about the interactions between these variables. Processes or quantities believed to be approximately constant at the time scale of interest are then treated as parameters. These decisions require detailed biological knowledge, which is, however, not always available. But even if it is well known that there is crosstalk between the system to be modelled and another biological process, the time course of these interactions is often not known or highly context dependent.

Unmodelled dynamics can have strong effects on the quantitative and qualitative dynamic features of the system. During model development, this might exhibit itself through an unsatisfactory fit of the model to the training data or by unrealistic parameter estimates [1]. The question of where the model should be modified and where the points of application of these unmodelled processes are, is often handled by labour intensive educated guessing or trial and error.

We have devised a data driven strategy, which we call the Reverse Tracking Algorithm (RTA), to handle unmodelled dynamics in ordinary differential equation models. The RTA (i) identifies candidate points of application of the unmodelled dynamical inputs and (ii) estimates the time course of these interfering signals.

An early version of this algorithm has already been successful in predicting regulatory signals in response to potassium starvation in the yeast *Saccharomyces cerevisiae* [2]. However, this early version suffered from numerical instabilities and required careful calibration of the algorithm's parameter. In our presentation we discuss a much more robust version of the RTA. We present the computational method of this improved and now easy to use RTA and demonstrate its application to three additional examples.

The first example illustrates, how the influence of circadian rhythms and other umodelled dynamics can be handled in pharmacokinetic models. In contrast to the standard approach we study the case where the elimination rate of a drug is subject to unknown oscillatory perturbations caused by time-of-day variations. We demonstrate, how the RTA can be applied to infer the time course of the elimination rate oscillations when starting from the standard single compartment model.

The second example is a biochemical reaction cascade with an unmodelled feedback mechanism. The RTA is applied to infer the cascade stage where the feedback applies and the unknown time course of the feedback. A simple phase plane analysis is then sufficient to discriminate between different models for the feedback kinetics, leading to an improved version of the cascade model.
The third example considers a well known model of the JAK/STAT signalling pathway, where the authors compared an initial model to a more accurate model involving rapid nucleocytoplasmic cycles of STAT5 species [4]. Starting from the initial model ignoring these cycles, we used the RTA to identify the unmodelled dynamics as well as the involved species (cycloplasmatic and nuclear STAT5). This was found by Swameye *et al.* using conventional manual model building strategies. We show that the RTA could have provided the same result in a semi-automatic way underpinning its utility for systematic model building and extension.

Keywords: systems biology, inference of biological networks, simulation of biological systems, parameter estimation, dynamic modelling, control theory

References

1. Raue, A., Becker, V., Klingmüller, U. and Timmer, J.: Identifiability and Observability Analysis for Experimental Design in Non-Linear Dynamical Models. Chaos, 20(4), 045105 (2010)
2. Kahm, M., Navarrete, C., Llopis-Torregrosa, V., Herrera, R., Barreto, L., Yenush, L., Arino, J., Ramos, J. and Kschischo, M.: Potassium starvation in yeast: mechanisms of homeostasis revealed by mathematical modeling.PLoS Comput Biol. 8(6):e1002548. doi: 10.1371/journal.pcbi.1002548. Epub 2012 Jun 21 (2012)
3. Liu, Y.Y., Slotine, J.J. , and Barabasi, A.L.: Controllability of complex networks. Nature 473, 167-173, (2011).
4. Swameye, I. and Müller, T. G. and Timmer, J. and Sandra, O. and KlingmÏler, U. Identification of nucleocytoplasmic cycling as a remote sensor in cellular signaling by databased modeling. Proc. Natl. Acad. Sci. USA. 100, (2003).