A Balancing Act: Parameter Estimation for Biological Models with Steady-state Measurements

Andrzej Mizera[1], Jun Pang[1], Thomas Sauter[2], and Panuwat Trairatphisan[2]

[1] Computer Science and Communications, University of Luxembourg, Luxembourg
[2] Life Sciences Research Unit, University of Luxembourg, Luxembourg

Problem statement. Constructing a computational model for a biological system consists of two main steps: (1) specifying the model structure and (2) determining the values for the parameters of the model. Usually, the model structure is represented in the form of a biochemical reaction network and the parameters are the reaction rate constants. The values of the reaction rates can be determined by fitting the model to experimental data by performing parameter estimation. However, the question remains whether the experimental data allow for unique identification of the parameters. To address this problem, one could perform a number of independent parameter estimations and investigate the range of obtained values among those parameter sets that result in a good fit. From the correlation of the obtained parameter sets one could, e.g., study whether only certain parameters are identifiable. This approach requires an *effective*, *efficient* and *automatic* way of performing estimation. In this study we concentrate on the case of fitting a deterministic mathematical model of a biological process, i.e., expressed in terms of a system of ordinary differential equations (ODEs), and a number of its variants to *multi-experiment steady-state data*. We propose a computational pipeline involving available software packages for achieving this goal while keeping a balance between the optimisation time and accuracy.

Our approach. The number of steady-state measurements may be insufficient to identify all the parameters of a model, especially when the number of parameters is larger than the number of measured species. To address this difficulty, we take into account data of the so called *knockout mutant models*, i.e., variants of the original model obtained by eliminating one or more interactions between species. Since the knockout mutants of real biological systems can be obtained and investigated in experimental practice as well as the physical/chemical properties are common for all variants, the steady-state measurements of the mutants can enrich the set of data available for parameter estimation and make possible the identification of model parameters. In our approach we assume that such multi-experiment, steady-state data are available. The aim is to gather in an automatic way (to the possibly largest extent) a number of parameter sets that result in a good simultaneous fit of all the mutants. The collected parameter sets could further be used to investigate the parameter identifiability question.

We propose MATLAB as a control environment for the task of executing many independent parameter estimation runs. In the proposed computational pipeline the ODE-based knockout model variants are compiled with the

SBTOOLBOX2 for MATLAB to C MEX files for efficiency. Parameter estimation is performed using a pattern search optimisation algorithm provided in the PSwarm global optimisation solver [1], mainly because (1) the solver provides a pattern search algorithm, which assures a local minimum convergence and does not require any information on the gradient of the score function and (2) the search step of the algorithm is implemented as the particle swarm algorithm [2], which makes it to a global optimisation algorithm.

Since the SBTOOLBOX2 does not provide any methods for an efficient and direct computation of the steady-state, we estimate the steady-state values by independently integrating each of the C MEX models to a point where a necessary steady-state condition is satisfied, i.e., the norm of the difference between points on the trajectory is less than a threshold. Since the accuracy depends on this threshold, the main challenge is to find a balance between the computational time and the accuracy of steady state estimation. To reduce the computational time, we first perform model fitting with a threshold that results in a steady state or a state close to a steady state but which can be reached with a relatively small number of integration steps. Next, the obtained parameter values are taken to COPASI for further optimisation with another direct method algorithm, i.e., PRAXIS, and a direct, efficient computation of a steady state.

Preliminary results. We apply the proposed approach to fit an ODE-based model of the PDGF signalling pathway [3] to steady-state experimental data. The model consists of 31 species and 40 reactions. Nine different variants of the model are considered and 31 unknown parameters common for all the variants need to be estimated. The experimental data are on the concentration of two species at steady state. In total the data set consists of 18 measurements. The models are implemented both in MATLAB and COPASI. The simultaneous implementation of all nine mutants in COPASI results in 279 ODEs.

One parameter estimation run in MATLAB lasts for approximately 30 mins and the cost function is evaluated for 40000 times. Then, the resulting parameter values are given to COPASI as a starting point for further optimisation with PRAXIS and 2000 cost function evaluations. This requires up to 2 hours of computational time of four cores. The improvement in the fit score is up to 30% of the score obtained in MATLAB. Our experiments show that this level of fit quality is unreachable in comparable amount of time if parameter estimation is performed from scratch in COPASI alone, even if a combination of COPASI optimisation algorithms is applied.

References

1. Vaz, A.I.F., Vicente, L.N.: A particle swarm pattern search method for bound constrained global optimization. J. Global Optimization **39**(2) (2007) 197–219
2. Kennedy, J., Eberhart, R.: Particle swarm optimization. In: Proc. IEEE International Joint Conference on Neural Networks, IEEE (1995) 1942–1948
3. Yuan, Q., Trairatphisan, P., Pang, J., Mauw, S., Wiesinger, M., Sauter, T.: Probabilistic model checking of the PDGF signaling pathway. T. Comp. Sys. Biology **14** (2012) 151–180