BIOINFORMATICS

Vol. 00 no. 00 2012
Pages 1–3

Data2Dynamics: a modeling environment tailored to parameter estimation in dynamical systems

A. Raue [1,*], B. Steiert[2], M. Schelker[3], C. Kreutz[2], B. Schöberl[1] and J. Timmer [2,4,5]

[1]Merrimack Pharmaceuticals Inc., 02139 Cambridge, MA, USA
[2]University of Freiburg, Institute for Physics, 79104 Freiburg, Germany
[3]Humboldt-Universität zu Berlin, Theoretical Biophysics, 10115 Berlin, Germany
[4]BIOSS Centre for Biological Signalling Studies, University of Freiburg, 79104 Freiburg, Germany
[5]Zentrum für Biosystemanalyse (ZBSA), University of Freiburg, 79104 Freiburg, Germany

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Associate Editor: XXXXXXX

ABSTRACT

Summary: Modeling of dynamical systems using ordinary differential equations is a popular approach in the field of Systems Biology. One of the most critical steps in this approach is to conveniently construct dynamical models of biochemical reaction networks for large data sets and complex experimental conditions and to perform efficient and reliable parameter estimation for model fitting. We present a MATLAB based modeling environment that pioneers these challenges. The numerically expensive parts of the required calculations such as the solving of the differential equations and of the associated sensitivity system are parallelized and automatically compiled into efficient C code. A variety of parameter estimation algorithms as well as frequentist and Bayesian methods for uncertainty analysis have been implemented and used on a range of applications that lead to publications.

Availability and Implementation: The Data2Dynamics modeling environment is a collaborative open source project and freely available at http://www.data2dynamics.org.

Contact: andreas.raue@fdm.uni-freiburg.de

Supplementary information is provided online.

1 INTRODUCTION

For the reconstruction of biochemical interaction networks as they occur in signal transduction or gene regulation, data sets generated under a wide range of experimental conditions have to be analyzed comprehensively. Here a software implementation is presented which is designed for efficient and convenient fitting of ODE models and offers great flexibility for specifying complex relationships between data sets and inferred dynamical model.

2 METHODOLOGY

Our software allows to directly specify the right hand side of the ODE manually, or to automatically generate it by providing a reaction scheme such as $A + B \rightarrow C$ with the respective rate-law like Mass Action or Michaelis-Menten. The resulting ODE system as well as its Jacobian matrix that is calculated automatically by symbolic differentiation are translated to C code and complied together with the ODE solver. The code makes efficient use of pre-calculated reaction fluxes as described in the Supplementary Information 1. Time-varying inputs to the ODE systems can be represented by custom or predefined input functions such as steps, pulses and splines that can depend on unknown parameters (Schelker *et al.*, 2012). The initial concentrations can be considered as functions of unknown parameters as well. The software allows considering multiple different models that can share common parameters and fit them simultaneously to all available data.

A unique feature of the Data2Dynamics software is its ability to consider and estimate the magnitude of measurement errors. Observations are only required for a subset of all dynamic states and experimental data might include additional parameters such as scaling or offset parameters.Another key feature of the Data2Dynamics software is its ability to conveniently and automatically create model variants that represent different experimental conditions. These conditions can be defined directly in the data sheets that contain the measurements and are conveniently parsed and grouped. For instance, a time course experiment with all combinations of two treatment options automatically yields four experimental conditions of the ODE system linked to the respective data. The model simulation will be plotted in the same grouping as well, see trajectories in different color in Fig. 1. For dose response experiments, the software again automatically generate all required model variants and display the simulation results in a dose response plot. For computational efficiency, experimental conditions, and thus model variants, that are shared between different experiments are calculated only once. Since all variants of the original ODE system have to be solved independently, the C code automatically parallelizes the execution of the ODE solver, see Supplementary Information 3 for a performance comparison.

A critical task in modeling of dynamical systems is the efficient and reliable estimation of model parameters, also called model fitting. We implemented a variety of different parameter estimation algorithms (Raue *et al.*, 2013b). The most efficient and reliable algorithm for parameter estimation in our hands is a deterministic trust

*to whom correspondence should be addressed

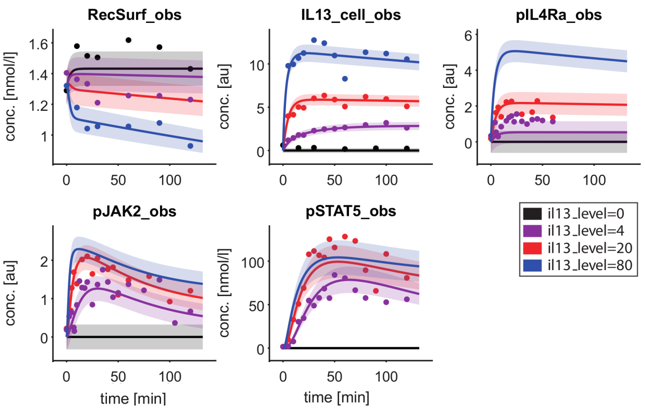

Fig. 1. Raia *et al.* (2011) model fitted to experimental data (dots) representing four different doses of IL-13. The solid lines are the the fitted model trajectories, the shades the estimated experimental noise levels of the data. A detailed step-by-step user guide for this example can be found in the Supplementary Information 2.

region approach combined with multi-start strategy to map out local minima. Parameters can and should be estimated on a log-scale. Prior knowledge about the parameters can be considered as well. If steady state assumptions for the model dynamics are required and the functional relationship to parameters are unknown, steady state constraints can be added to the objective function, including the respective derivatives. A quality control, as proposed in Raue *et al.* (2013b) can be performed to validate robustness of the estimation results.

The software implements a sophisticated method to calculate model sensitivities, i.e. the derivatives of the dynamics with respect to model parameters, see Supplementary Information 5 and 6 for details. The sensitivity equations are derived automatically by symbolic differentiation, translated to C code and complied together with the original ODE systems and the solver. We showed previously (Raue *et al.*, 2013b) that this approach is not only about ten times faster but also more precise than the default approach using finite differences. A reliable calculation of these derivatives is key to successful parameter estimation.

In addition to finding the best model fit to a given collection of data, the Data2Dynamics software implements a wide range of algorithms that are able to determine uncertainties in the estimated parameter as well as in the predicted model dynamics. In particular, the frequentist profile likelihood approach for identifiability analysis (Raue *et al.*, 2009), the prediction profile likelihood approach for observability analysis (Kreutz *et al.*, 2012) as well as a variety of Bayesian approaches (Raue *et al.*, 2013a; Hug *et al.*, 2013) that calculate posterior probability distributions are available. Based on the results of such uncertainty analyses, the software allows to design additional experiments (Steiert *et al.*, 2012) that can resolve non-identifiability and non-observability (Raue *et al.*, 2010; Kreutz *et al.*, 2013) and improve prediction accuracy.

3 SUMMARY

We present the Data2Dynamics software, a modeling environment that is especially tailored to parameter estimation and model fitting in dynamical systems. The code is open source and is developed in a community effort using a web-based hosting service and a revision control system. A variety of published applications, e.g. Becker *et al.* (2010); Raia *et al.* (2011); Bachmann *et al.* (2011), that made use of the software are provided as benchmark examples for further methods development and as guide for novel applications. For these examples not only the models but also all datasets and their link to the models as well as all original information used in the parameter estimation and uncertainty analysis are provided. The software was awarded twice as best performer in the Dialogue for Reverse Engineering Assessments and Methods (DREAM, 2011 and 2012).

ACKNOWLEDGEMENT

We thank all academic and industrial collaborators that helped to evolve this modeling environment, in particular the Timmer group at the University of Freiburg, the Klingmüller group and the Höfer group at DKFZ Heidelberg, the Theis group at the German Research Center for Environmental Health, the Bode group at the University Hospital of Düsseldorf, the Klipp group at the Humboldt-University Berlin, Merrimack Pharmaceuticals in Cambridge as well as the Shvartsman lab at Princeton. This work was supported by the German Ministry of Education and Research (LungSys2 0316042G, Virtual Liver Network 0315766).

Conflict of interest: None declared.

REFERENCES

Bachmann, J., Raue, A., Schilling, M., Böhm, M., Kreutz, C., Kaschek, D., Busch, H., Gretz, N., Lehmann, W., Timmer, J., and Klingmüller, U. (2011). Division of labor by dual feedback regulators controls JAK2/STAT5 signaling over broad ligand range. *Molecular Systems Biology*, **7**, 516.

Becker, V., Schilling, M., Bachmann, J., Baumann, U., Raue, A., Maiwald, T., Timmer, J., and Klingmueller, U. (2010). Covering a broad dynamic range: information processing at the erythropoietin receptor. *Science*, **328**(5984), 1404–1408.

Hug, S., Raue, A., Hasenauer, J., Bachmann, J., Klingmüller, U., Timmer, J., and Theis, F. (2013). High-dimensional Bayesian parameter estimation: Case study for a model of JAK2/STAT5 signaling. *Mathematical Biosciences*, **246**(2), 293–304.

Kreutz, C., Raue, A., and Timmer, J. (2012). Likelihood based observability analysis and confidence intervals for predictions of dynamic models. *BMC Systems Biology*, **6**, 120.

Kreutz, C., Raue, A., Kaschek, D., and Timmer, J. (2013). Profile likelihood in systems biology. *FEBS Journal*, **280**(11), 2564–2571.

Raia, V., Schilling, M., Böhm, M., Hahn, B., Kowarsch, A., Raue, A., Sticht, C., Bohl, S., Saile, M., Möller, P., Gretz, N., Timmer, J., Theis, F., Lehmann, W., Lichter, P., and Klingmüller, U. (2011). Dynamic mathematical modeling of IL13-induced signaling in Hodgkin and primary mediastinal B-cell lymphoma allows prediction of therapeutic targets. *Cancer Research*, **71**, 693–704.

Raue, A., Kreutz, C., Maiwald, T., Bachmann, J., Schilling, M., Klingmüller, U., and Timmer, J. (2009). Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*, **25**(15), 1923–1929.

Raue, A., Becker, V., Klingmüller, U., and Timmer, J. (2010). Identifiability and observability analysis for experimental design in non-linear dynamical models. *Chaos*, **20**(4), 045105.

Raue, A., Kreutz, C., Theis, F., and Timmer, J. (2013a). Joining forces of Bayesian and frequentist methodology: A study for inference in the presence of non-identifiability. *Phil. Trans. Roy. Soc. A*, **371**, 20110544.

Raue, A., Schilling, M., Bachmann, J., Matteson, A., Schelker, M., Kaschek, D., Hug, S., Kreutz, C., Harms, B., Theis, F., Klingmüller, U., and Timmer, J. (2013b). Lessons learned from quantitative dynamical modeling in systems biology. *PLOS ONE*, **8**(9), e74335.

Schelker, M., Raue, A., Timmer, J., and Kreutz, C. (2012). Comprehensive estimation of input signals and dynamical parameters in biochemical reaction networks. *Bioinformatics*, **28**(18), i522–i528.

Steiert, B., Raue, A., Timmer, J., and Kreutz, C. (2012). Experimental design for parameter estimation of gene regulatory networks. *PLOS ONE*, **7**(7), e40052.