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

A. Raue [1],[*] B. Steiert[2], M. Schelker[3], C. Kreutz[2], T. Maiwald[2], H. Hass[2], J. Vanlier[2], C. Tönsing[2], L. Adlung[4], R. Engesser[2], W. Mader[2], T. Heinemann[9,5], J. Hasenauer[6,7], M. Schilling[4], T. Höfer[9,5], E. Klipp[3], F. Theis[6,7], U. Klingmüller[4], B. Schöberl[1] and J. Timmer[2,8]

[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]Systems Biology of Signal Transduction, German Cancer Research Center, 69120 Heidelberg, Germany
[5]BioQuant, University of Heidelberg, 69120 Heidelberg, Germany
[6]Helmholtz Center Munich, 85764 Neuherberg, Germany
[7]Technische Universität München, 85748 Garching, Germany
[8]BIOSS Centre for Biological Signalling Studies, University of Freiburg, 79104 Freiburg, Germany
[9]Divison of Theoretical Systems Biology, German Cancer Research Center, 69120 Heidelberg, Germany

Associate Editor: XXXXXXX

ABSTRACT

Summary: Modeling of dynamical systems using ordinary differential equations is a popular approach in the field of Systems Biology. Two of the most critical steps in this approach are to 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 modeling environment for MATLAB that pioneers these challenges. The numerically expensive parts of the 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 MATLAB based, open source and freely available at http://www.data2dynamics.org.

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

Supplementary information is provided online and contains detailed description of methodology, a user guide and documentation.

For the construction of computational models that are used to analyze signal transduction pathways, gene regulation and cellular decisions, data sets generated under a wide range of experimental conditions have to be analyzed comprehensively. The Data2Dynamics software environment is designed for computationally efficient and user-friendly integration of complex experimental data into models consisting of coupled non-linear ordinary differential equations (ODE).

Our software allows to specify the right hand side of the ODE manually, or to automatically generate it by providing the reaction scheme with the respective rate-law such as Mass Action, Michaelis-Menten or a custom rate equation. 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 into a MATLAB mex-function. The code makes efficient use of pre-calculated reaction fluxes. 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 fitting them simultaneously to all available data.

The software is able to automatically create model variants that represent different experimental conditions. The conditions can be defined directly in the data sheets that contain the measurements and are parsed and grouped. For instance, a data sheet containing four different experimental conditions automatically creates four different variants of the ODE system that are linked to the respective data. The model simulations will be plotted in the same grouping as well, see trajectories in different color in Fig. 1. For dose-response experiments, the software automatically generates all required model variants and displays the simulation results in a

---

[*]to whom correspondence should be addressed
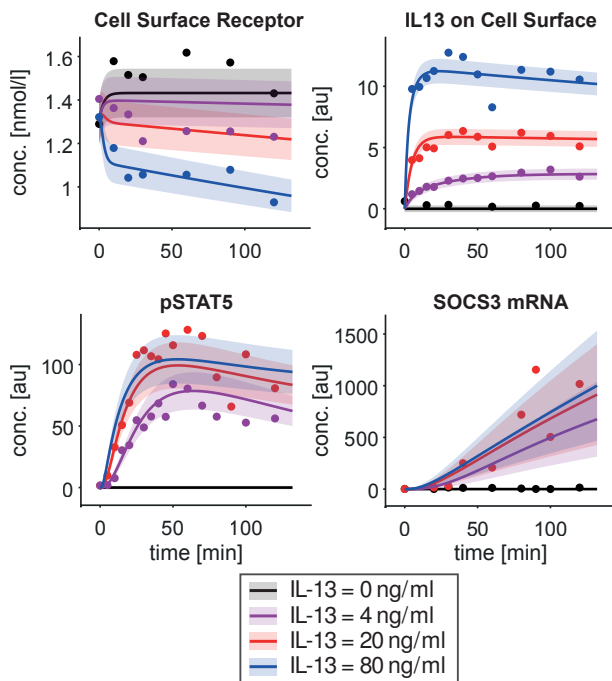
Fig. 1. Raia *et al.* (2011) model fitted to experimental data (dots) representing four different doses of IL-13. The solid lines are the fitted model trajectories, the shades the estimated experimental error of the data.

dose-response plot. For computational efficiency, experimental conditions, and thus model variants, that are shared between different experiments are calculated only once. The C code can automatically parallelize the calculation of the ODE solutions since all model variants can be solved independently. The mapping between experimental data and the simulated model can contain additional parameters that can, for instance, account for unknown offsets or scaling factors. A unique feature of the Data2Dynamics software is its ability to estimate the magnitude of experimental errors together with the model dynamics, see Fig. 1. To this end, a full likelihood function is automatically generated.

A critical task in modeling of dynamical systems is the efficient and reliable estimation of unknown 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 region approach combined with a multi-start strategy to map out local minima. A variety of global optimization algorithms is implemented as well (Kronfeld *et al.*, 2010). Parameters can and should be estimated on a log-scale. Prior knowledge about the parameters can be considered by prior distributions. The respective penalization of the likelihood leads to maximum a posteriori estimation. 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. 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. The sensitivity equations are derived automatically by symbolic differentiation, translated to C code and compiled together with the original ODE system and solver. We showed previously (Raue *et al.*, 2013b) that this approach, which is not implemented in any other software packages, 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. Goodness of fit analysis, Likelihood ratio test, QQ-plots and auto-correlation of the residuals are available.

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). A variety of Markov-chain Monte Carlo approaches are implemented to perform Bayesian inference (Raue *et al.*, 2013a; Hug *et al.*, 2013). 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) and improve prediction accuracy (Kreutz *et al.*, 2013).

The software is developed in a community effort using a web-based code hosting service and a revision control system. A variety of published applications are provided as benchmark examples for further methods development and as guideline for novel applications. For these examples not only the models but also all datasets, 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

REFERENCES

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.

Kronfeld, M., Planatscher, H., and Zell, A. (2010). The EvA2 optimization framework. *Learning and Intelligent Optimization*, **6073**, 247–250.

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.