BIOINFORMATICS

Vol. 00 no. 00 2012
Pages 1–3

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

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

[1]University of Freiburg, Institute for Physics, 79104 Freiburg, Germany
[2]Merrimack Pharmaceuticals Inc., 02139 Cambridge, MA, USA
[3]Humboldt-University, 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 both 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. The code and full documentation are hosted on the website http://www.data2dynamics.org. Participation in the further development is highly welcome!

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

Supplementary information: A supplementary text with mathematical and numerical details is available.

Model Construction The dynamics of cellular processes such as signal transduction pathways or models of cell growth and death can be described by models consisting of ordinary differential equations (ODE). Mathematically, such dynamical systems can be characterised by $\dot{\mathbf{x}}(t,\boldsymbol{\theta}) = \mathbf{f}(\mathbf{x}(t,\boldsymbol{\theta}),\mathbf{u}(t,\boldsymbol{\theta}),\boldsymbol{\theta})$ with $\mathbf{x}(0,\boldsymbol{\theta}) = \mathbf{g}(\boldsymbol{\theta})$. In this case the vector of state variables $\mathbf{x}(t,\boldsymbol{\theta})$ describes the dynamics of molecular components. The function $\mathbf{u}(t,\boldsymbol{\theta})$ represents experimental treatments that are time-varying inputs to the ODE systems. A variety of predefined function such as step functions, pulses, splines or custom defined functions are available that can depend on unknown parameters (Schelker *et al.*, 2012). The initial concentrations $\mathbf{x}(\mathbf{0},\boldsymbol{\theta})$ can be a function $\mathbf{g}(\boldsymbol{\theta})$ of unknown parameters as well. For a specific cell type or biological context, the set of parameters $\boldsymbol{\theta}$ are often not available from literature and have to be estimated from experimental data. The software allows to consider multiple different models that can share common parameters and fit them simultaneously to all available data. Our software allows both to 1) directly specify the right hand side of the ODE manually, or to 2) automatically generate it by providing a reaction scheme such as $A + B \rightarrow C$. In the latter case, for specifying the reaction rate equations one can either choose to apply default Mass Action kinetics, either reversible $\leftrightarrow$ or non-reversible $\rightarrow$, or a custom rate law such as Michaelis-Menten or Hill kinetics. The resulting right hand side of the 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. In case of 2), the code makes efficient use of pre-calculated reaction fluxes throughout the code as described. See in the Supplementary Information 1 for more technical details.

Dataset and Experimental Conditions Experimental data provides information to estimate unknown model parameter by fitting the model. Each possible measurement is mathematically represented by a functional mapping $\mathbf{y}(t,\boldsymbol{\theta}) = \mathbf{h}(\mathbf{x}(t,\boldsymbol{\theta}),\mathbf{u}(t,\boldsymbol{\theta}),\boldsymbol{\theta})$ that might include additional parameters such as scaling or offset parameters and thus increase the dimension of $\boldsymbol{\theta}$. For each model, multiple datasets can and should be considered simultaneously. One key feature of the Data2Dynamics software that is its ability of the conveniently and automatically create model variants that represent different experimental conditions that were used in different experiments. These conditions can directly be defined in the data sheets that contain the experimental data and is conveniently parsed and grouped. For instance, a time course experiment could have been performed with all combinations of two treatment options. The corresponding four experimental setting will be extracted and respective variants of the ODE system are generated and linked to the data. The model simulation will be plotted in the same grouping as well. Another frequently occurring case are dose response experiments at a fixed time points. In this case the software will 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 share between different experiments

*to whom correspondence should be addressed

are calculated only once. Since all variants of the original ODE system have to be solved independently, the C code automatically parallelized the execution of the ODE solver, see Supplementary Information 2 for a performance comparison.

Experimental Noise Another unique feature of the Data2Dynamics software is its ability to consider and estimate uncertainty in experimental data. By implementing a full Maximum Likelihood Estimation framework, see Supplementary Information 3 for details, we demonstrated that this approach provides a more robust and reliable determination of experimental noise (Raue *et al.*, 2013b). The software is able to estimate the amount of experimental noise simultaneously with the unknown parameter that define the dynamics in a statistically more efficient way. A determination of measurement noise by manual inspection or by a preprocessing procedure is not required any more. Therefore, the results that are obtained by our software are independent of the degree of the experimenter's believe in the quality of the experimental data.

Sensitivity Equations The software implements a sophisticated method to calculate model sensitivities, the derivatives of the dynamics with respect to model parameters, see Supplementary Information 4 and 5 for details. The sensitivity equations are derived automatically by symbolic differentiation, are 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 fold faster but also more precise than the default approach. A reliable calculation of these derivatives is key to successful parameter estimation.

Parameter Estimation A critical task in modeling of dynamical systems is the efficient and reliable estimation of model parameters, also called model fitting. We implemented can compared 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 multi-start strategy to map out local minima. Parameters can and should be estimated on a log scale. For experimental data, the software allows to compare model to data on a log scale as well. A very general objective function including all model variant, dataset and experimental conditions is generated based on the Maximum Likelihood Estimation framework. If a steady state assumption $\dot{\mathbf{x}}(t) = 0$ for the model dynamics is required, and its solution $\mathbf{g}(\boldsymbol{\theta})$ is not know, a steady state constraint can added to the objective function, including the respective derivatives. Prior assumption can be added to the objective function as well. A quality control, as proposed in Raue *et al.* (2013b) can be preformed to validate robustness of the estimation results.

Uncertainty Analysis and Experimental Design In addition to finding the best model fit to the collection of experimental data using parameter estimation, the Data2Dynamics software implements a variety of algorithms that determined uncertainty in the estimated parameter as well as in the predicted model dynamics or derived quantities. In particular, the frequentist profile likelihood approach for identifiability analysis (Raue *et al.*, 2009), the predictions 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 the uncertainty analysis, 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.

Visualization and Reports A variety of automatically generated plots and visualizations of all simulated quantities as well as experimental data, model structure and parameter estimation and uncertainty analysis results are available. All information about the model equations, the data set and estimation results as well as generated figure can be automatically complied into a Latex based PDF report.

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 systems. 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 as well as Merrimack Pharmaceuticals in Cambridge.

Funding: 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.