BIOINFORMATICS APPLICATIONS NOTE
Vol. 28 no. 9 2012, pages 1290–1291
doi:10.1093/bioinformatics/bts137

Systems biology

Advance Access publication March 25, 2012

ADMIT: a toolbox for guaranteed model invalidation, estimation and qualitative–quantitative modeling

Stefan Streif*, Anton Savchenko, Philipp Rumschinski, Steffen Borchers and Rolf Findeisen

Otto-von-Guericke Universität Magdeburg, Institute for Automation Engineering, Systems Theory and Automatic Control Laboratory, Magdeburg, 39106, Germany

Associate Editor: Jonathan Wren

ABSTRACT

Summary: Often competing hypotheses for biochemical networks exist in the form of different mathematical models with unknown parameters. Considering available experimental data, it is then desired to reject model hypotheses that are inconsistent with the data, or to estimate the unknown parameters. However, these tasks are complicated because experimental data are typically sparse, uncertain, and are frequently only available in form of qualitative *if–then* observations. ADMIT (**A**nalysis, **D**esign and **M**odel **I**nvalidation **T**oolbox) is a MatLab^{TM}-based tool for guaranteed model invalidation, state and parameter estimation. The toolbox allows the integration of quantitative measurement data, a priori knowledge of parameters and states, and qualitative information on the dynamic or steady-state behavior. A constraint satisfaction problem is automatically generated and algorithms are implemented for solving the desired estimation, invalidation or analysis tasks. The implemented methods built on convex relaxation and optimization and therefore provide guaranteed estimation results and certificates for invalidity.

Availability: ADMIT, tutorials and illustrative examples are available free of charge for non-commercial use at http://ifatwww.et.uni-magdeburg.de/syst/ADMIT/

Contact: stefan.streif@ovgu.de

Received on September 30, 2011; revised on March 13, 2012; accepted on March 17, 2012

1 INTRODUCTION

Data-based mathematical modeling can help to improve the understanding of complex biological networks, e.g. by analyzing and identifying core elements, or by predicting the network's behavior. However, usually the data are uncertain and come from heterogeneous sources. This then results in competing model hypotheses and hence an incomplete understanding of the underlying network structure and biological mechanisms. In iterative cycles of mathematical modeling and biological experimentation, model hypotheses are tested and rejected if inconsistent with the experimental data. For the remaining hypotheses the parameters can be estimated, e.g. to allow for quantitative predictions.

Depending on the available data and the assumed uncertainty description, one can distinguish between several approaches for parameter estimation. To obtain optimal parameter estimates (by minimizing a cost function) of nonlinear models and measurements data with statistical uncertainty descriptions, methods using global or local optimization can be considered [e.g. Moles *et al.* (2003)]. Efficient implementations of such methods are available in software tools, [e.g. Maiwald and Timmer (2008); Schmidt and Jirstrand (2006)]. Optimization-based approaches do not necessarily provide information about the precision of the estimates. As a solution, profile likelihood and resampling methods such as Bootstrapping, Jackknife or Monte-Carlo testing have been proposed [e.g. Joshi *et al.* (2006); Kremling *et al.* (2004); Raue *et al.* (2009)]. However, finding a suitable threshold that classifies a model as consistent with the data is challenging (Anderson and Papachristodoulou, 2009).

A complementary approach is based on an unknown-but-bounded (or set-based) uncertainty description [e.g. Milanese and Belforte (2002); Walter and Piet-Lahanier (1990)]. Because entire sets can be directly taken into account (theoretically) definite statements can be made, which therefore allows a rigorous perspective on parameters and model consistency.

We present ADMIT, a MatLab^{TM}-based toolbox that uses a set-based uncertainty description and convex relaxation and optimization framework for model invalidation and parameter estimation (Rumschinski *et al.*, 2010). Besides unknown-but-bounded measurement data, qualitative information such as temporal or causal *if–then* observations or discrete state-variables can be used (Rumschinski *et al.*, 2012). The toolbox automatically constructs a convex constraint satisfaction problem that incorporates all available data and the model equations. Using convex relaxation and optimization methods, outer-bounds of the consistent parameters or states can be determined by solving the constraint satisfaction problem. For this purpose, either the solvers implemented in ADMIT or more efficient external state-of-the-art numerical solvers for mixed-integer linear programs can be used.

2 MAIN FEATURES

Installation and software dependability: ADMIT runs under MatLab^{TM} (requires version $\geq$R2010a and the Symbolic Toolbox), which allows the toolbox to be used with Windows, Linux, Unix and MAC OS. Installation of the toolbox consists of simply unpacking a file to the desired location and running a single install script.

To improve the efficiency and accuracy of the estimation results, a wide range of external solvers (most of which are free of charge for non-commercial research or educational purposes, e.g. SEDUMI, GUROBI and CPLEX) can be used if the freely available YALMIP toolbox (Löfberg, 2004) is installed.

*To whom correspondence should be addressed.

For Monte-Carlo simulations and parallel estimation, MatLab^{TM}'s Optimization and Parallelization toolboxes are required. Installation of the freely available SBToolbox2 (Schmidt and Jirstrand, 2006) and libSBML (http://sbml.org/Software/libSBML) are required for import of SBML and SBToolbox2 models.

Toolbox environment: All functions of the toolbox can be accessed from within the MatLab^{TM} scripting environment. The main tasks and the default behavior can be controlled by options and few simple function calls. For advanced users and more challenging tasks, additional options and functions are available. All functions are explained in detail by help texts and short examples within the MatLab^{TM} help browser.

Import and export of data, models and problems: Systems biology models developed with the SBToolbox2 or in the SBML format can be imported for further analysis. Measurement data stored in different file formats (plain text, CSV etc.) can also be imported. Preprocessing routines allow the specification of measurement data uncertainties (e.g. relative or absolute errors) and adding qualitative constraints (e.g. monotonicity). Using binary variables, qualitative biological knowledge such as logical or conditional relationships or temporally uncertain data can be specified and considered (Rumschinski *et al.*, 2012). The obtained constraint satisfaction problem can be exported to a text file in a human-readable and intuitive format, which facilitates the exchange of data and models.

Model types and complexity: Both the steady-state or transient behavior of models with nonlinear kinetics, ranging from polynomial (mass-action kinetics) to rational (e.g. Michaelis-Menten or Hill kinetics), can be analyzed. In addition, discrete-valued states, parameters or inputs (e.g. describing stimuli that can only be provided in an on/off way) can be included. Thus, this allows e.g. the analysis of signal transduction and metabolic networks. Currently, dynamic models with up to 10 states and 15 parameters, or steady-state models with up to 250 states and parameters can be analyzed. Further complexity reduction techniques allow even bigger models to be treated (Rumschinski *et al.*, 2010).

Model invalidation, parameter and state estimation: The implemented algorithms automatically reformulate and relax the non-convex constraint satisfaction problems to obtain a convex one. If the solution set of the constraint satisfaction problem is empty, then the corresponding parameter regions or entire model was proved inconsistent with the data. Note that numerical conditioning and round-off errors for models with large uncertainties can pose challenges and may incorrectly classify a model inconsistent. The algorithms have been optimized to reduce such issues.

To solve the constraint satisfaction problem, the toolbox provides own routines or interfaces with external high-end solvers via YALMIP. Two different algorithms are available to obtain an outer approximation of the feasible sets of parameters or states. Using bisectioning, the (possibly non-convex) solution set can be approximated up to a chosen accuracy. Another less accurate but much faster algorithm (outer-bounding by optimization) is particularly suited in case of a large number of variables, e.g. for state estimation. State estimation results can be used to detect discrepancies between models and data (outliers), to assess the quality of the model, or to check qualitative or quantitative constraints. Finally, the toolbox offers simulation of Monte-Carlo samples to compare solutions with the estimated guaranteed bounds.

3 EXAMPLES

Several illustrative examples are available demonstrating different features of the toolbox and how qualitative information and quantitative data can be formulated. All examples are motivated by our current modeling projects with real qualitative or quantitative data. The repository of examples currently include: parameter estimation for a Michaelis–Menten-type reaction network (Rumschinski *et al.*, 2010), parameter estimation and invalidation of an adaptation model based on qualitative and uncertain data (Rumschinski *et al.*, 2012), state estimation and fault diagnosis of a two-tank process. We will also make models and data used in future publications available together with the toolbox.

4 CONCLUSIONS

The toolbox implements new set-based algorithms for modeling and analysis of various types of networks and motifs, and it also makes the algorithms accessible in an easy way. Compared with approaches based on samples, the complementary set-based approach allows definite statements on entire regions in the parameter space. Because only unknown-but-bounded uncertainties are assumed, no assumptions on statistics of measurements have to be made. Additionally, the use of discrete-valued variables allows qualitative data and information to be taken into account. Drawbacks are that set-based approaches can be sensitive to measurement outliers and can be computationally demanding.

In summary, the toolbox complements other approaches and is useful to obtain a better understanding of uncertain systems.

Conflict of Interest: none declared.

REFERENCES

Anderson,J. and Papachristodoulou,A. (2009) On validation and invalidation of biological models. *BMC Bioinformatics*, **10**, 132.

Joshi,M. *et al*. (2006) Exploiting the bootstrap method for quantifying parameter confidence intervals in dynamical systems. *Metab. Eng*, **8**, 447–455.

Kremling,A. *et al*. (2004) A benchmark for methods in reverse engineering and model discrimination: problem formulation and solutions. *Genome Res.*, **14**, 1773–1785.

Löfberg,J. (2004) Yalmip: a toolbox for modeling and optimization in MATLAB. In *Proceedings of the CACSD Conference*, Taipei, Taiwan.

Maiwald,T. and Timmer,J. (2008) Dynamical modeling and multi-experiment fitting with potterswheel. *Bioinformatics*, **24**, 2037–2043.

Milanese,M. and Belforte,G. (2002) Estimation theory and uncertainty intervals evaluation in presence of unknown but bounded errors: linear families of models and estimators. *IEEE T. Automat. Contr.*, **27**, 408–414.

Moles,C. *et al*. (2003) Parameter estimation in biochemical pathways: a comparison of global optimization methods. *Genome Res.*, **13**, 2467.

Raue,A. *et al*. (2009) Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*, **25**, 1923–1929.

Rumschinski,P. *et al*. (2010) Set-base dynamical parameter estimation and model invalidation for biochemical reaction networks. *BMC Syst. Biol.*, **4**, 69.

Rumschinski,P. *et al*. (2012) Combining qualitative information and semi-quantitative data for guaranteed invalidation of biochemical network models. *Int. J. Robust. Nonlin. Control*. (In press). doi: 10.1002/rnc.2793.

Schmidt,H. and Jirstrand,M. (2006) Systems biology toolbox for MATLAB: a computational platform for research in systems biology. *Bioinformatics*, **22**, 514–515.

Walter,E. and Piet-Lahanier,H. (1990). Estimation of parameter bounds from bounded-error data: a survey. *Math. Comput. Simulat.*, **32**, 449–468.